

# GLOBAL PROJECT

**ALUMNO 1: ANTONIO MARTÍNEZ VICENTE**

**ALUMNO 2: JHAIRO JOSÉ GARCÍA**

**ALUMNO 3: MARIA ALEJANDRA MORA**

**PROGRAMA:**

**MASTER EN DATA SCIENCE Y BIG DATA**

**NOMBRE DEL PROYECTO: “Optimización de la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia”**

**Link al video de presentación:**

**[https://drive.google.com/file/d/1ivdZSJStHtenBb4\\_c7KMj2HM6yOtxvZE/view?ts=651ca5ee](https://drive.google.com/file/d/1ivdZSJStHtenBb4_c7KMj2HM6yOtxvZE/view?ts=651ca5ee)**

## Contenido

<b>RESUMEN</b>	<b>3</b>
<b>INTRODUCCIÓN</b>	<b>4</b>
<b>ESTADO DEL ARTE</b>	<b>7</b>
<b>OBJETIVOS</b>	<b>10</b>
<b>SOLUCIÓN PLANTEADA</b>	<b>11</b>
<b>EVALUACIÓN</b>	<b>21</b>
<b>RESULTADOS</b>	<b>27</b>
<b>CONCLUSIONES Y TRABAJOS FUTUROS</b>	<b>35</b>
<b>REFERENCIAS</b>	<b>36</b>

## RESUMEN

El proyecto aborda la optimización de la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia, cuyos objetivos son la promoción de la sostenibilidad alimentaria y el apoyo a los negocios locales.

La solución planteada involucra una aproximación analítica que incluye la creación de una base de datos consolidada, sistemas de recomendación, modelos predictivos y análisis de sentimientos. Estos logros están en consonancia con el objetivo general de nuestro proyecto: la optimización de la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia.

La justificación de esta solución radica en su capacidad para mejorar la eficiencia en las actividades clave de la asociación. Los sistemas de recomendación personalizan los menús y ofrecen sugerencias a los usuarios, los modelos predictivos permiten anticipar la demanda y el análisis de sentimientos evalúa la satisfacción de los socios. Además, la base de datos centralizada brinda una visión integral de toda la información.

A pesar de algunos desafíos, como la falta de acceso a ciertos datos, el proyecto logró avanzar gracias a la redefinición de objetivos basados en la información disponible. El problema inicial se relacionaba con una discrepancia entre los objetivos formulados en la propuesta del proyecto y la disponibilidad limitada de datos necesarios para abordar estos objetivos.

Mediante una metodología iterativa y un enfoque analítico se busca optimizar las operaciones de una cocina comunitaria y mejorar la satisfacción de sus usuarios. Se han logrado avances significativos en la consecución de los objetivos del proyecto, sentando las bases para mejoras continuas en el futuro.

## INTRODUCCIÓN

El presente proyecto se realiza sobre un caso real: la asociación Sukalde, una cocina comunitaria ubicada en Donostia (San Sebastián), cuyos objetivos incluyen promover la sostenibilidad alimentaria mediante la elaboración de comidas asequibles y el apoyo a los negocios locales.

La asociación Sukalde se encuentra ante la necesidad de optimizar alguna de sus operaciones como cocina comunitaria, donde la planificación de menús y la satisfacción de los miembros de la asociación son esenciales para su funcionamiento. Con recursos humanos limitados y una historia relativamente corta de registros, se enfrenta a desafíos operativos que suponen una carga de tiempo y trabajo que les impide abordar con eficiencia otras áreas del trabajo diario de la asociación. De igual modo, la recogida de datos que se ha realizado hasta la fecha no se encuentra en el formato adecuado para un análisis efectivo, por lo que se dificulta la extracción de información de interés para los miembros de la asociación.

Hasta la fecha, la asociación Sukalde carecía de soluciones efectivas para optimizar sus operaciones. La planificación de menús y la previsión de ingredientes se realizaba manualmente, lo que generaba una carga de trabajo significativa. Además, no se aprovechaba eficazmente la información recopilada ni se realizaba un análisis integral de la satisfacción de los usuarios.

El proyecto busca abordar estas problemáticas mediante la creación de las soluciones siguientes:

- Se propone la creación de una base de datos que permita una gestión eficiente de la recolección de los datos que ya se venían realizando semanalmente, de manera que les permita extraer conocimiento de la cantidad de información que tenían disponible. Además, se cargará la base de datos en la plataforma de AWS para que por un lado, facilite la gestión y el análisis por parte del equipo del Global Project y por otro lado, les permita tras la finalización de este trabajo, continuar actualizando la información de la base de datos y seguir utilizando las herramientas desarrolladas en este proyecto.
- Por otro lado, se crearán cuatro sistemas de recomendación divididos en dos grupos. El primer grupo está destinado al equipo de cocina de la asociación y busca ayudar en la planificación de menús. El segundo grupo se enfoca en los usuarios, proporcionando recomendaciones personalizadas basadas en sus preferencias.
- En cuanto a la previsión de ingredientes, se proponen dos enfoques: el primero está basado en los patrones observados de los históricos de consumo del ingrediente per se, y el segundo, en los patrones generalizables, altamente influenciados por los ingredientes que están presentes en los platos de mayor consumo.
- Con respecto al nivel de satisfacción del cliente se propone un análisis de sentimientos que les permita evaluar tanto la satisfacción actual de los usuarios, como su satisfacción futura que será determinada a través de encuestas posteriores. Se

generará una arquitectura automatizada que muestre los resultados de nuevas encuestas en un panel visual para que presente dichos resultados de manera comprensible para las partes interesadas.

La solución propuesta se considera adecuada e innovadora debido a su capacidad para automatizar y mejorar los procesos de planificación de menús, la previsión de ingredientes y la personalización de las sugerencias de platos. Se contribuye a reducir el desperdicio de comida al adquirir y cocinar únicamente lo que se requiere, garantizando al mismo tiempo que los ingredientes estén a mano cuando se precisen. Por otro lado, se pueden identificar problemas potenciales, como quejas frecuentes sobre ciertos platos, lo que permite a los cocineros actuar proactivamente para mejorar la calidad de los platos. Y por último, la información detallada sobre los platos ofrecidos, recetas, ingredientes o las preferencias de los clientes con sus historiales de pedidos, entre otros, estarán contenidos en una base de datos en la nube, brindando una visión integral de todo el conjunto de información de la asociación.

En cuanto a la ejecución de nuestro Global Project hemos adoptado una metodología iterativa enfocada en la repetición de los diferentes ciclos de trabajo y revisiones a lo largo del proceso. Aunque no hemos seguido de manera integral alguno de los diferentes enfoques populares, nuestra metodología guarda similitudes con la gestión visual de tareas de Kanban, en donde nos hemos servido de la plataforma de Trello para la gestión de las labores.

Hemos enfrentado dificultades para enfrentar algunos de los objetivos iniciales propuestos en la *“Propuesta del proyecto”* en base a la información que teníamos disponible. Ciertos datos cruciales fueron inaccesibles, así que para evitar discrepancia entre nuestros objetivos iniciales y la realidad, y al mismo tiempo, para tratar de evitar posibles desviaciones en la efectividad y fiabilidad de nuestras soluciones, realizamos algunos ajustes en los objetivos del proyecto final. Este es el caso de la optimización de costos, en donde al no disponer de un historial de pedidos se limitaba nuestra capacidad para resolverlo de manera precisa.

Nuestra metodología se ha centrado en las siguientes etapas:

- Business understanding / Identificación y conocimiento del problema
- Aproximación analítica
- Recolección y preparación de los datos
- Elaboración de los modelos

A través de este enfoque, basado en una revisión y retroalimentación iterativa, hemos alcanzado los objetivos planteados exitosamente:

- Se conformó la base de datos utilizando la información disponible en la asociación. Las tablas que registran los pedidos y las raciones de los usuarios se actualizan semanalmente y la carga de la base de datos en AWS se realiza de manera efectiva.

- Se desarrollaron los sistemas de recomendación que abordaban las necesidades del equipo de cocina y de los usuarios. Estos sistemas se basan en la popularidad por familias de platos, aleatoriedad por familias, la similitud por usuarios y la similitud de platos por ingredientes. La primera versión del sistema de recomendación basado en similitud entre usuarios se considera el más preciso.
- Se implementaron exitosamente enfoques univariantes y multivariantes para predecir el comportamiento de los ingredientes. Los modelos univariantes revelaron que el algoritmo Histogram-Based Boosting y ARIMA obtuvieron los errores mínimos más bajos, aunque las predicciones mostraron una variabilidad limitada debido a la escasez de datos. En contraste, los modelos multivariantes, como KNNeighbors y VAR, permitieron un mejor ajuste al considerar patrones y tendencias de múltiples ingredientes.
- Se realizó un estudio exitoso de la opinión de los miembros de Sukalde, generando resultados concluyentes a pesar de una respuesta limitada. Se crearon modelos para analizar futuras encuestas y se desarrolló un panel de control con Google Looker para proporcionar valiosos insights a la asociación.

## ESTADO DEL ARTE

En los últimos años, se ha prestado una atención creciente al uso de Machine Learning (ML) supervisado para prever el inventario en el sector de alimentos y bebidas. En España, este sector representa el 2,6% del Producto Interno Bruto (PIB). Sin embargo, en el ámbito de la distribución de alimentos y bebidas y restauración, no se han tenido a disposición herramientas de software que utilicen modelos avanzados para la previsión de insumos en la gestión de inventarios, sistemas de recomendación y el análisis de sentimiento para la evaluación de la satisfacción de los usuarios.

Dos estudios han demostrado que los modelos de Extreme Gradient Boosting (XGB) y Perceptrón Multicapa (tipo de redes neuronales), superan en rendimiento a los modelos estadísticos tradicionales, como el Modelo de Suavizado Exponencial (MSE). A diferencia de estos modelos tradicionales, que se basan en relaciones predefinidas, los modelos como el XGB y Perceptrón Multicapa aprenden patrones directamente de la demanda histórica en la previsión de insumos (Praveen et al 2020 y Garrido, 2021).

Con el fin de mejorar la capacidad de la empresa para predecir la demanda de sus productos utilizando algoritmos de ML, se realizó un estudio que compara dos modelos de pronóstico para productos de confitería de una de las mayores empresas de distribución en Arabia Saudita. Para lograr este objetivo, se utilizaron los algoritmos de Máquina de Vectores de Soporte (SVM, por sus siglas en inglés) y Memoria a Corto y Largo Plazo (LSTM, por sus siglas en inglés) dependiendo del tipo de productos y sin considerar patrones en conjunto. Ambos algoritmos proporcionaron resultados sólidos al medirse con el modelo de pronóstico de demanda (Nassibi, 2023).

La mayoría de los estudios existentes se han enfocado en buscar algoritmos avanzados que tengan en cuenta la estacionalidad y tendencia en los datos históricos de un insumo específico. Sin embargo, se presta poca atención a la interrelación de este insumo con otros, lo cual podría ayudar a identificar patrones generalizados que el grupo de insumos podría seguir a lo largo del tiempo. Nuestro enfoque se diferencia de estos estudios al proponer modelos predictivos que consideran estos patrones generalizados para prever los ingredientes que la cocina comunitaria Sukalde utiliza para la elaboración de sus platos. Por otro lado, este enfoque nos permite descubrir patrones que no son fácilmente identificables al analizar ingredientes en forma individual, especialmente debido a la limitación actual de los datos temporales.

En cuanto a los sistemas de recomendación, existen multitud de trabajos en el área para la recomendación de platos y recetas. En los resultados de *Google Académico* se pueden encontrar un gran número de resultados sobre sistemas de recomendación de alimentos, recetas, ingredientes, etc.

En la elaboración de este proyecto se ha seguido en mayor parte una implementación propia de estos sistemas debido a la naturaleza de la información disponible. Las restricciones en la calidad de nuestros datos impidieron llevar a cabo investigaciones

con la misma utilidad que algunos de los principales estudios efectuados en este campo.

La mayoría de los trabajos sobre los sistemas de recomendación en el mundo de la gastronomía pasan por la recomendación de recetas en cuanto a sus componentes nutricionales para apoyar una dieta saludable. Una aproximación muy interesante a nuestro caso de estudio es el sistema de recomendación que se encuentra en la plataforma de *Cookidoo*<sup>1</sup>, en donde se generan recomendaciones de recetas basadas en un filtrado colaborativo por usuarios, parecido al “*Sistema de Recomendación 3*” de nuestro proyecto, con la salvedad de la calidad de los datos en estas recetas, puesto que se realizan recomendaciones utilizando también los componentes nutricionales de los ingredientes.

En la creación de nuestro “*Sistema de Recomendación 4*” se han tomado dos estudios como referencia: Yum\_me - Yang et al., 2017 y Trang Tran et al., 2018. Sin embargo, como se señalaba, nuestros datos tienen ciertas limitaciones, puesto que no se dispone de un perfil de usuario tan desarrollado como en esos modelos, ni los componentes nutricionales de las recetas.

En el sector de la alimentación se busca constantemente mejorar la selección y el consumo de alimentos con la finalidad de hacer que el consumidor tenga la mejor experiencia posible. En la última década, ha habido un fuerte interés por incluir las emociones en los estudios de análisis sensorial e investigaciones sobre las opiniones de los consumidores (García, 2021). En la industria de restaurantes y servicios de alimentos, la presentación de los platos, los sabores, el servicio del personal y las experiencias culinarias están diseñados para crear emociones positivas y memorables en los comensales. Un plato bien presentado acompañado de un buen servicio puede evocar emociones de satisfacción y alegría, mientras que una experiencia culinaria decepcionante y un mal servicio puede generar emociones negativas y afectar la percepción de la marca del restaurante. Además, las emociones también están vinculadas a las tendencias alimenticias. Por ejemplo, el interés creciente en la comida saludable y sostenible está impulsado en parte por las emociones asociadas a la preocupación por la salud personal y el medio ambiente (Universidad de Navarra, 2022)

El procesamiento del lenguaje natural (NLP, por sus siglas en inglés) desempeña un papel fundamental en la comprensión de los sentimientos de los clientes al abandonar un restaurante. Al utilizar el NLP, se pueden comprender, procesar y analizar los datos del lenguaje humano, lo que a su vez ayuda a generar información valiosa y conocimientos profundos sobre la experiencia del consumidor (Sanchez, 2023). El proceso automatizado conocido como “análisis de sentimiento”, que forma parte del NLP, permite capturar no solo las palabras, sino también las emociones detrás de ellas, lo que brinda una comprensión completa de cómo se siente el cliente y cómo mejorar su experiencia en el restaurante.

---

<sup>1</sup> Cookidoo - <https://cookidoo.es/foundation/es-ES>



Los datos de las encuestas, cuando se incorporan a los modelos de análisis de sentimientos, pueden ayudar a comprender qué características o rasgos del restaurante ayudaron a ganar la lealtad y confianza de los clientes o incluso comprender qué experiencias negativas llevaron a un aumento de las tasas de abandono (Thomas, 2022).

Un aspecto innovador de nuestro enfoque es el de buscar agilizar el proceso de análisis de las respuestas de las encuestas, a través de una arquitectura que permita automatizar el proceso que lleva desde la recogida de la información hasta la visualización de los resultados, pasando por un análisis intermedio con nuestros modelos de análisis de sentimientos. Se proporcionará así información valiosa para que Sukalde se adapte a las necesidades de su comunidad con un ahorro en costo y tiempo mucho mayor.

Un conjunto de sistemas de recomendaciones tanto para el personal de Sukalde como para los socios, un modelo de previsión de ingredientes capaz de encontrar patrones generalizados para todos los ingredientes y un sistema automatizado de análisis de sentimiento (todo ello en una única fuente de datos centralizada en la nube) son los instrumentos claves que proponemos para optimizar la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia.

## OBJETIVOS

Objetivo general: **Optimización de la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia**

Objetivos específicos y actividades necesarias para alcanzar dichos resultados:

**1. Crear una base de datos consolidada**

- Creación de la base de datos
- Carga de la base de datos en la infraestructura AWS

**2. Automatizar la planificación de los menús y personalizar las sugerencias de los platos.**

- Creación de un sistema de recomendación basado en la popularidad de platos por familia
- Creación de un sistema de recomendación basado en la aleatoriedad de platos por familia
- Creación de un sistema de recomendación basado en similitud entre usuarios
- Creación de un sistema de recomendación basado en similitud de platos por ingredientes.

**3. Mantener los niveles óptimos de insumos a través de la generación de modelos de series temporales.**

- Generación de históricos detallados de las cantidades utilizadas de cada ingrediente basados en las órdenes y platos elaborados semanalmente.
- Análisis de la posible existencia de patrones estacionales y tendencias en los históricos de los ingredientes.
- Generación de pronósticos para cada ingrediente basados en los patrones y tendencias del mismo ingrediente.
- Generación de pronósticos para cada ingrediente basados en los patrones y tendencias generalizadas para todos los ingredientes.
- Selección de los modelos más efectivos basándose en métricas de regresión y rendimiento.

**4. Automatizar el análisis de las encuestas mediante la implementación de un sistema de análisis de sentimientos**

- Análisis y preprocesamiento de las respuestas para la implementación de modelos de ML.
- Clasificación de las respuestas para la evaluación de los modelos de ML.
- Implementación de técnicas y algoritmos de procesamiento de lenguaje natural para la clasificación de sentimientos.
- Evaluación del modelo óptimo para el análisis de sentimientos.
- Creación de una interfaz de visualización.

## SOLUCIÓN PLANTEADA

A pesar de no seguir ninguno de los enfoques populares de manera integral, hemos desarrollado una metodología iterativa para el desarrollo de nuestro Global Project, centrándonos en la repetición de los ciclos de trabajo y la revisión a lo largo del proceso. En ciertos aspectos, se podría asemejar al flujo de trabajo y la gestión de tareas de manera visual que se encuentran en la metodología Kanban, ya que hemos estado utilizando el software Trello para llevar a cabo nuestras labores.

Sin embargo, antes de entrar en profundidad en la metodología que hemos seguido para la consecución del proyecto final es necesario comentar algunas dificultades que hemos tenido que superar para conseguir una coherencia entre los objetivos planteados inicialmente en la *“Propuesta del proyecto”* y la información disponible a la que teníamos acceso para la consecución de dichos objetivos.

A raíz de la metodología iterativa seguida en nuestro proyecto final, hemos podido detectar algunas inconveniencias para una aproximación más adecuada hacia algunos de los objetivos presentados en la propuesta inicial. Las dificultades que hemos enfrentado al avanzar en nuestro proyecto se deben a las circunstancias que rodearon nuestra fase inicial de aproximación, lo que ha dificultado la coherencia en nuestra propuesta original. Por tanto, existen ligeros ajustes en nuestro proyecto final con respecto a nuestra propuesta inicial en base a la metodología establecida.

Hemos detectado en primera instancia, en las fases iniciales de nuestro proyecto, que podíamos incurrir en el error de saltar directamente a las soluciones si eludimos la metodología que nos habíamos propuesto para la elaboración del mismo. En su momento, cuando se entregó la propuesta inicial, no se habían establecido los contactos necesarios con el equipo de Sukalde y a pesar de que existían unas necesidades iniciales, que presentamos precisamente en esa propuesta, hemos detectado y redefinido posteriormente los problemas que realmente podríamos resolver en función de los datos que disponíamos. Si hubiéramos seguido por ese camino, tratando de dar respuesta a cada uno de los objetivos que planteamos en la propuesta inicial, habría existido incluso una desviación significativa a la hora de resolver los problemas reales que enfrentaba Sukalde.

El elemento clave de dicho ajuste se ha producido en la fase de recolección de la información. Para completar algunos de los objetivos mencionados en la propuesta inicial era necesario tener acceso a algunos datos para poder enfrentar las diferentes problemáticas. Si bien se ha realizado un enorme trabajo para digitalizar mucha de la información que la asociación disponía en forma no estructurada, no ha sido posible acceder a información del histórico de compras de los insumos (más allá de los platos consumidos por los usuarios), el coste actualizado de los mismos, informes económicos mensuales y trimestrales, cierta información de los proveedores y una información más detallada de los usuarios. Esto ha supuesto que no se pudieran realizar algunos

modelos que de una manera más afinada optimizaran los costes de la asociación de una manera más directa, más allá de la eficiencia a la hora de realizar las compras y planificar los pedidos.

Nuestro proyecto se centra en un caso real de una asociación basada en una cocina comunitaria que tiene como objetivos: promover la sostenibilidad alimentaria elaborando comidas preparadas de bajo costo y a su vez, apoyar a los negocios locales con la compra de insumos; tal y como se señalaba en nuestra propuesta de proyecto. Esta asociación que está ubicada en Donostia (San Sebastián) está formada por un conjunto de socios que se diferencian en dos tipos: un reducido grupo de socios que se encarga de la ejecución y de la operativa de las principales actividades (compra de los recursos, elaboración y distribución de los menús, administración, etc.) y del resto de miembros de la asociación que de manera clientelar consumen los menús semanalmente y tienen acceso a platos preparados para toda la semana con ingredientes de productores locales y de negocios de la zona.

Como hemos desarrollado nuestro proyecto basándonos en un caso real, hemos tratado de adoptar un enfoque profesional y realista con el objetivo de asegurar la calidad de los resultados. Nos hemos esforzado por garantizar que estos resultados satisfagan las necesidades del cliente y sean útiles de manera efectiva. Hemos realizado incluso un modelo descriptivo<sup>2</sup> respondiendo una petición específica de nuestro cliente.

Nuestra metodología consta de las siguientes etapas:

- **Business understanding** – Entendimiento del negocio/asociación y conocimiento de los procesos que llevan a cabo nuestro cliente objetivo. Una comprensión inicial para obtener claridad en torno a los problemas que se pueden resolver y determinar qué datos se utilizarán para resolver las diferentes cuestiones.

Se realizó una entrevista directa con nuestro cliente objetivo para definir los objetivos e identificar las actividades necesarias que apoyan la consecución de dichos objetivos<sup>3</sup>: ¿Reducir los costes de la actividad? ¿Mejorar la eficiencia? ¿Aumentar la satisfacción de los socios?

Se establecieron nuevos contactos con los interesados para aclarar algunas preguntas y definir las actividades necesarias para la consecución de dichos objetivos. Tras dedicar un tiempo para definir los objetivos adecuadamente se prioriza como área efectiva de revisión:

### **Optimización de la eficiencia operativa y la experiencia de los usuarios**

---

<sup>2</sup> Archivo - "SUKALDE\_Base\_de\_datos.ipynb" en la sección "Generacion de Insights" - Carpeta "Objetivo 1"

<sup>3</sup> "18 de MAYO - Primera entrevista con SUKALDE.pdf" – Carpeta "Anexos".

Al desglosar este problema podemos ver algunas de las dificultades que enfrentaba la cocina comunitaria de Sukalde:

Al ser una asociación con pocos años en funcionamiento y con todavía, un equipo reducido de miembros encargados de la ejecución de las actividades principales de la asociación, necesitaban optimizar y, en caso de que fuera posible, automatizar algunos de los procesos para mejorar el rendimiento y reducir el tiempo que dedicaban a algunas de las tareas.

En este sentido, la asociación quería reducir las cargas de trabajo que supone la elaboración de los menús semanalmente, la planificación y compra de los ingredientes necesarios para la elaboración de dichos menús, la detección de pérdidas y desperdicios y la predicción de necesidades futuras. Por otro lado, no había una gestión eficiente de la recolección de los datos que realizaban semanalmente y por lo tanto tampoco podían extraer conocimiento de la cantidad de información que tenían disponible. Y, por último, se necesitaba tratar de comprender de una forma global la satisfacción de los comensales y sus preferencias culinarias, así como las opiniones y actitudes expresadas por los miembros para tomar decisiones dentro de la asociación.

- **Aproximación analítica** – Una vez definidos los problemas a resolver se realizaron debates estructurados por parte de los tres miembros del proyecto para organizar y planificar cómo abordar los diferentes problemas. Antes incluso de empezar a recolectar los datos, se plantearon diferentes métodos desde enfoques “*data science*” que pudieran resolver los problemas reales. Para ello se propuso desplegar soluciones para abordar los siguientes objetivos:

1 – Crear una base de datos consolidada – Para contener todos los datos en una misma ubicación para realizar los análisis posteriores.

2 – Automatizar la planificación de los menús y personalizar las sugerencias de los platos – Creando para ello diferentes sistemas de recomendación.

3 – Mantener los niveles óptimos de insumos a través de la generación de modelos de series temporales – Se desarrollarán modelos predictivos y simuladores para predecir rendimientos futuros del negocio. Se obtendrá un modelo predictivo basado en un análisis de series temporales usando diferentes algoritmos de ML. Aunque este modelo queda fuera de los contenidos desarrollados en éste Máster nos resultaba un acercamiento bastante acertado para resolver el problema que queremos abordar.

4 – Automatizar el análisis de las encuestas mediante la implementación de un sistema de análisis de sentimientos - Analizar y entender estas opiniones y comentarios de manera efectiva donde se pueda automatizar y agilizar este proceso, proporcionando insights valiosos que permitan a Sukalde adaptarse y evolucionar de acuerdo con las necesidades y deseos de su comunidad.

- **Recolección y preparación de la información** – En la propuesta de proyecto se señalaron las diferentes fuentes de datos disponibles siguientes. Se hará una breve descripción para entender cómo cada una de las fuentes eran necesarias para resolver los problemas a los que queríamos dar respuesta:

---

1. Histórico de pedidos de los clientes - Contiene datos detallados sobre pedidos pasados, fechas, preferencias y detalles de socios asociados, es esencial para analizar patrones de consumo, personalizar recomendaciones y prever la demanda futura.
2. Escandallos con las fichas técnicas de los platos - Los escandallos detallan las fichas técnicas de los platos del menú, incluyendo ingredientes, cantidades, recetas y métodos de preparación. Son esenciales para analizar costos, gestionar el inventario y estimar rendimientos, y se encuentran en documentos y hojas de Excel.
3. Historial de compras semanales - Estos datos incluyen información sobre proveedores, productos adquiridos, cantidades compradas y precios.
4. Informes económicos mensuales y trimestrales – Proporcionan datos sobre ingresos, costos y márgenes de beneficio, fundamentales para el análisis financiero y la toma de decisiones en términos de eficiencia y rentabilidad.
5. Feedbacks de mejoras por parte de los asociados - Feedbacks y sugerencias de mejora proporcionados por los asociados de la asociación. Serán utilizados para el análisis de sentimientos y la implementación de mejoras continuas en el servicio ofrecido.
6. Historial de menús semanales - Incluye detalles de platos, ingredientes y preferencias temáticas. Utilizaremos estos datos para analizar tendencias de consumo, personalizar recomendaciones y planificar estratégicamente futuros menús.

---

Como se señalaba al inicio del texto, no fue posible acceder a los puntos 3 y 4 de las “Fuentes de datos disponibles” de la Propuesta del Proyecto. Si bien existía alguna información con respecto a un historial de compras realizadas, había mucha información faltante, por lo que se determinó no utilizar el historial de compras semanales como una fuente fiable para la extracción de información. Algo similar ocurría con la fuente de datos número 4, los informes económicos trimestrales y mensuales. Se tenía acceso a un informe económico trimestral, pero el resto de informes no estaban completados.

Así que basándonos en esta circunstancia, se realizaron algunos ajustes en cuanto a los objetivos finales que podría alcanzar nuestro proyecto, dejando a un lado cualquier modelo que permitiera la optimización de costos en cualquiera de las áreas o procesos que se realizaban en la cocina comunitaria.

La etapa de recolección y preparación de la información ha sido sin ninguna duda la parte del proyecto que ha demandado más tiempo y esfuerzo. La transformación de los datos para su posterior empleo en los análisis y construcción de modelos ha requerido, en algunos casos, un contacto permanente con el equipo de la cocina comunitaria. Incluso, en fases posteriores se ha revelado la necesidad de volver a la fase de recolección y transformación de los datos para construir una definición más completa de las fuentes de información.

Los escandallos<sup>4</sup> con las fichas técnicas de los platos no estaban completos, de alrededor de un total de 120 elaboraciones, sólo se disponía de 30 fichas técnicas válidas. Tras estudiar diferentes acciones a llevar a cabo, se determinó necesario completar el resto de platos restantes. Así que, de manera conjunta con el equipo de cocina de Sukalde, se elaboraron las 90 fichas técnicas restantes para completar la información.

Para agregar información más valiosa en los estudios de análisis de sentimiento se realizó una encuesta para que fuera enviada a los miembros de la asociación<sup>5</sup>. No obstante, al final, no fue posible enviar esta encuesta que habíamos preparado y se optó por utilizar una encuesta previa que ya había realizado la asociación. La razón por la que no se pudo utilizar la encuesta que nosotros habíamos realizado se debió a que la carga de trabajo diaria de la asociación y la dificultad de obtener las respuestas de los miembros en el plazo requerido imposibilitaba la tarea. De todas formas, la información disponible en la asociación fue suficiente para poder realizar nuestros análisis y para poder extraer diferentes conclusiones sobre el estado de satisfacción de los usuarios.

Por lo tanto, aunque no se pudo utilizar la totalidad de la información que se había previsto desde un inicio, en esta etapa se trató de garantizar que la información existente tuviera el formato correcto, estuviera completa y nos permitiera disponer de los datos en un estado óptimo para el funcionamiento de los modelos y la consecución de los objetivos. Así que, y como parte del proceso iterativo, se redefinieron algunos de los objetivos iniciales en función de la información disponible y se dispusieron de manera definitiva las herramientas de aprendizaje automático clave para la consecución de dichos objetivos.

- **Elaboración de modelos** – Como se mencionaba en la etapa de “*Aproximación analítica*” se establecerá una base de datos en la nube que centralizará los datos en un único lugar, facilitando así futuros análisis y brindando a la asociación la posibilidad de actualizar la información y utilizar las herramientas desarrolladas en el futuro, se realizarán unos sistemas de recomendación, se desarrollarán modelos predictivos y se realizará un análisis de sentimientos en donde se creará además un panel visual para mostrar los resultados de los análisis.

---

<sup>4</sup> “platos\_final.xlsx” - Carpeta “ANEXOS”

<sup>5</sup> “Encuesta de Satisfacción realizada por el equipo del Global Project” - Carpeta “ANEXOS”

Los notebooks desarrollados para cada objetivo contienen los análisis completos, con todo detalle y de manera ordenada. Conforme vamos avanzando en el desarrollo de los modelos, elaboramos conclusiones coherentes siguiendo nuestra metodología establecida. Por cuestiones de espacio y debido a la necesidad de sintetizar información, no se incluyen todos los análisis, justificaciones y conclusiones obtenidas en los notebooks.

## 1 Crear una base de datos consolidada

La creación de la base de datos consta de dos etapas, una primera fase en donde se contienen todos los datos en una misma ubicación, se crean las tablas y se establecen las relaciones entre las mismas. Y una segunda etapa en donde se sube dicha base de datos a la nube para facilitar el análisis del equipo de trabajo del Global Project y en donde, el equipo de Sukalde podrá actualizar la información de la base de datos y seguir utilizando las herramientas que desarrollemos.

Se adjunta en la carpeta de proyecto 5 archivos<sup>6</sup>: una copia de la base de datos elaborada, una imagen con el sistema relacional entre las tablas, unas diapositivas de cómo se subió la base de datos a AWS y dos notebooks con una pequeña muestra de las tablas disponibles y con algunos análisis descriptivos que muestran ciertas informaciones acerca de los datos disponibles.

## 2 Automatizar la planificación de los menús y personalizar las sugerencias de los platos<sup>7</sup>

Se crearán un total de 4 sistemas de recomendación que abordan las problemáticas identificadas y que supondrá al equipo de cocina de Sukalde una ayuda adicional que se traducirá en una carga de trabajo y tiempo mucho menor, y al mismo tiempo, un aumento en la satisfacción de los usuarios:

- El primer sistema de recomendación se centrará en ofrecer los platos más populares por familia. A través de esta herramienta el equipo de cocina podrá identificar rápidamente las preferencias de los usuarios y simplificar la creación de los menús. Supondrá una planificación más eficiente y una reducción del tiempo dedicado a esta tarea.
- En segundo lugar, el siguiente sistema proporcionará listas aleatorias de platos por familias, permitiendo una mayor variabilidad en la oferta de menús. Se creará una segunda versión de este sistema de recomendación en donde se introducirá un nuevo plato que todavía no se haya cocinado en la asociación. De esta forma se introduce un elemento sorpresa a la hora de crear los menús y elegir los platos por parte del usuario. Se mantiene el interés y la satisfacción de los usuarios a lo largo del tiempo.

---

<sup>6</sup> Carpeta - "Objetivo 1"

<sup>7</sup> "SUKALDE\_Sistemas\_de\_recomendación.ipynb" - Carpeta - "Objetivo 2/Notebooks/"



- A continuación, se creará el tercer sistema de recomendación que estará enfocado en sugerir platos a los usuarios en función de sus preferencias y los gustos similares de otros usuarios. Esto permitirá a los clientes recibir recomendaciones personalizadas. Se consigue así aumentar su satisfacción al encontrar platos que todavía no ha probado y que se ajusten a sus preferencias individuales. Se creará una segunda versión en donde se utilizará la biblioteca Surprise con los algoritmos de SVD y KNN.
- Por último, crearemos un sistema de recomendación a través de los ingredientes que contienen los platos, de esta forma podemos sugerir a los usuarios platos similares en función de los ingredientes de los platos que más les gustan.

Para facilitar la comprensión del trabajo se ha considerado relevante establecer dos grupos diferenciados dentro de estos 4 sistemas de recomendación en función de su uso y respuesta de la necesidad: 2 para el equipo de cocina y 2 para los usuarios.

- Sistemas de recomendación para el equipo de cocina - Los dos primeros sistemas de recomendación irán destinados al equipo de cocina de la asociación, de manera que les ayude a la creación de nuevos menús semana a semana.

- Sistemas de recomendación para los usuarios - Los dos últimos sistemas de recomendación irán destinados a los clientes que realizan los pedidos semanalmente para que obtengan sugerencias en función de sus gustos:

En algunos de estos sistemas se creará una segunda versión que, o bien añade información nueva, o bien se aproxima desde un enfoque diferente.

### 3 Mantener los niveles óptimos de insumos a través de la generación de modelos de series temporales <sup>8</sup>

Se generarán modelos predictivos que consideren de manera intrínseca la demanda esperada de ingredientes según los platos más solicitados, así como la disponibilidad de productos en función de la temporada.

A continuación presentamos dos enfoques para realizar la previsión de insumos:

A) Desarrollo de múltiples modelos de series temporales univariantes: Se realizarán pronósticos de los ingredientes basados en los patrones y tendencias individuales, considerando que cada ingrediente se trata de manera independiente de los demás.

Los modelos regresores univariantes, que se adaptan mejor a estos datos estacionarios<sup>9</sup> y que se utilizaran en este apartado, son:

---

<sup>8</sup> "SUKALDE\_Modelos\_Predictivos\_fina.ipynb" - Carpeta - "Objetivo 3/Notebooks/"

<sup>9</sup> Diagnóstico obtenido desde de realizar la evaluación de la estacionalidad y la interdependencia entre los históricos de ingredientes contenida en el notebook "SUKALDE\_Modelos\_Predictivos\_final".

- Los modelos de Autoregression, Integración y Media Móvil (ARIMA)
- Los modelos de Regresión Univariante y Ensemble: Entre los cuales se incluyen:
  - Linear Regression
  - KNeighbors
  - Decision Tree
  - Random Forest
  - Histogram-Based Gradient Boosting
  - XGB

En los modelos de series temporales es común referirse al "backtesting" o prueba retrospectiva. Este término se refiere a la práctica de evaluar un modelo utilizando los datos históricos del conjunto "test" o de validación para ver cómo se habría comportado el modelo en el pasado (Amat & Escobar, J., 2023).

B) Desarrollo de modelos de series temporales multivariantes: Se realizarán pronósticos de los ingredientes, teniendo en cuenta tanto los patrones y tendencias individuales como el grado de dependencia entre los diferentes ingredientes.

A diferencia de los modelos de series temporales univariantes, que se centran en un solo ingrediente a lo largo del tiempo, los modelos de series temporales multivariantes pueden involucrar la interacción y la dependencia entre múltiples ingredientes, lo que puede proporcionar información más completa sobre el sistema o fenómeno que se está estudiando.

Los modelos regresores multivariantes, que se adaptan mejor a este tipo de datos estacionarios y que se utilizarán en este apartado, son:

- Los modelos de Vectores Autorregresivos (VAR)
- Los modelos de Regresión Multivariante y Ensemble: Entre los cuales se incluyen:
  - Linear Regression
  - KNeighbors
  - Decision Tree
  - Random Forest
  - Histogram-Based Gradient Boosting
  - XGB

#### 4 Automatizar el análisis de las encuestas mediante la implementación de un sistema de análisis de sentimientos<sup>10</sup>

Es vital comprender las percepciones, necesidades y expectativas de sus miembros. La respuesta emocional de los usuarios hacia los platos o hacia el trabajo de la asociación puede proporcionar información muy valiosa cuya correcta interpretación puede

---

<sup>10</sup> "SUKALDE\_Análisis\_de\_Sentimiento.ipynb" - Carpeta - "Objetivo 4/Notebooks/"

mejorar significativamente la predicción de la elección de los mismos<sup>11</sup>. La recopilación de opiniones a través de encuestas es fundamental, pero como comentamos, el desafío reside en analizar eficazmente esta información variada y extensa.

Esto es especialmente relevante para Sukalde, ya que identificar y abordar las críticas y preocupaciones de los usuarios es fundamental para mejorar y adaptar el servicio ofrecido.

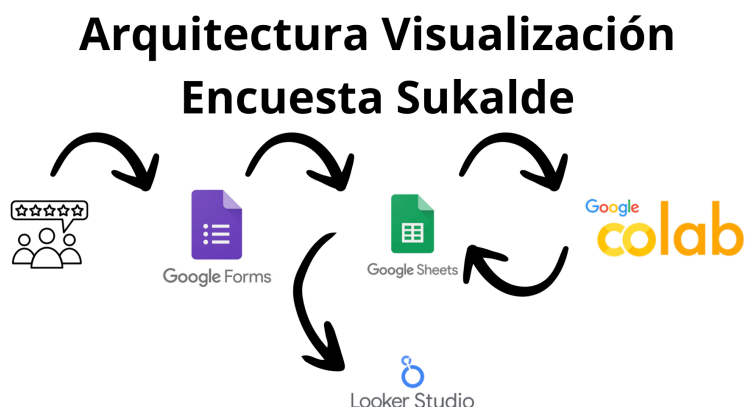
La solución propuesta busca automatizar y agilizar el proceso de análisis de las respuestas de las encuestas, proporcionando información valiosa para que Sukalde se adapte a las necesidades de su comunidad.

Se implementará un sistema de análisis de sentimientos para evaluar las respuestas de las encuestas de los miembros de Sukalde. Esto implica una serie de actividades clave, como el análisis detallado de las respuestas, el preprocesamiento de los datos para su análisis con modelos de ML, la clasificación de las respuestas y la implementación de técnicas de procesamiento de lenguaje natural.

Se realizó un modelo personalizado y se utilizaron además modelos como la Regresión Logística y SVM para lograr una precisión óptima en la clasificación de opiniones.

La solución que hemos propuesto también incluye la presentación de manera visual de resultados de manera comprensible para las partes interesadas. Se realizará a través de una interfaz y herramientas que permitan a los miembros de Sukalde acceder a los análisis y visualizaciones de manera intuitiva y en tiempo real.

A continuación, mostramos la arquitectura diseñada para las siguientes encuestas de la asociación:



La recolección de datos se realizará mediante Google Forms. Es muy fácil de usar y resulta familiar para los miembros de Sukalde. La encuesta que ellos realizaron fue a

<sup>11</sup> García, G. (2021). Las emociones contribuyen a la elección y consumo de alimentos. THE FOOD TECH - Medio de noticias líder en la Industria de Alimentos y Bebidas.  
<https://thefoodtech.com/marketing/las-emociones-contribuyen-a-la-eleccion-y-consumo-de-alimentos/>

través de esta misma plataforma. Posteriormente, las respuestas pasarán automáticamente a Google Sheets, lo que permitirá un acceso sencillo y una manipulación eficiente de los datos.

A continuación, se utilizará el contenido de nuestro notebook en donde se realiza un análisis más profundo de las respuestas recogidas, generando un análisis de sentimiento avanzado.

Los resultados de este análisis se volverán a almacenar en Google Sheets y se presentarán finalmente de manera visual con la ayuda de Looker Studio, que facilitará su interpretación tanto en una vista general como en desgloses por periodos específicos<sup>12</sup>.

Un aspecto destacable de esta solución es la automatización de la tarea en base a la arquitectura diseñada, ya que una vez que lo hemos configurado, el proceso desde la recopilación hasta la visualización de datos es en gran medida automático, lo que ahorra tiempo y esfuerzo. Además, es altamente flexible y escalable, adaptándose a medida que la organización crece o sus requisitos cambian. Y de fácil accesibilidad, ya que está basada en la nube, lo que permite a las partes interesadas acceder a las visualizaciones y análisis en tiempo real desde cualquier dispositivo con conexión a internet.

Se considera la importancia de que el personal técnico de Sukalde esté familiarizado con las herramientas propuestas para garantizar el mantenimiento, la ejecución y la mejora continua del sistema. Hemos sugerido la realización de sesiones de formación para las partes interesadas, con el fin de que puedan interpretar y actuar en función de las sutilezas proporcionadas por el análisis. Estos elementos son esenciales para aprovechar al máximo el potencial de esta solución tecnológica en el contexto de Sukalde.

---

<sup>12</sup> "SUKALDE\_-\_Satisfacción\_Clientes.pdf" - Muestra del Dashboard de Looker Studio en la carpeta - "Objetivo4/Archivos/"

## EVALUACIÓN

A continuación, describimos la evaluación de las 4 soluciones planteadas para alcanzar nuestro objetivo de ***“Optimización de la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia”***:

### **Objetivo 1: Crear una base de datos consolidada**

Para la creación de la base de datos de la asociación Sukalde, se llevó a cabo un proceso de consolidación de la información contenida en diversos archivos, que incluían registros semanales de usuarios y pedidos, listas de platos y el histórico de órdenes de pedido de los usuarios, entre otros. La estructuración de las tablas y sus relaciones se efectuó utilizando la plataforma MariaDB.

Una vez completada la construcción de la base de datos, junto con su conjunto de tablas relacionales, se procedió a la implementación de servicios en la nube en AWS, que dispone de una capa gratuita anual que nos permitía usar sus servicios sin costo alguno (se espera que los miembros de la asociación utilicen su propia cuenta a la finalización del proyecto). Esto permitió que los miembros del equipo Global Project pudieran utilizar la base de datos como un recurso central para el desarrollo de sus modelos, así como un repositorio actualizable de manera semanal por parte del equipo de Sukalde. La información que se ha utilizado para el desarrollo del proyecto abarca desde octubre de 2022 hasta agosto de 2023.

En este contexto, se optó por trabajar con una copia de la base de datos que reflejara el estado de la información hasta la fecha mencionada. Esta decisión se basó en la consideración de que, aunque era posible mantener la base de datos en constante actualización, dicha práctica generaba inconvenientes en la ejecución de los modelos del proyecto, derivados de las dificultades surgidas durante la carga de datos por parte del equipo de Sukalde.

Hemos realizado dos notebooks diferentes<sup>13</sup> para mostrar el contenido de la base de datos y para realizar algunos análisis exploratorios con algunas representaciones gráficas que ayuden a comprender mejor el contenido de las tablas. En el archivo *“Sukalde\_Base\_de\_datos.ipynb”* se muestran algunos gráficos a raíz de una solicitud por parte del equipo de Sukalde al equipo del Global Project, en donde necesitaban algunos *insights* entre los meses de Marzo de 2023 y Agosto de 2023 que utilizarían en la justificación de una subvención que ellos habían recibido.

### **Objetivo 2: Automatizar la planificación de los menús y personalizar las sugerencias de los platos.**

El proceso de creación de los sistemas de recomendación ha contado con diferentes etapas bien establecidas que han culminado en la obtención de los sistemas deseados con resultados acordes a nuestras expectativas.

---

<sup>13</sup> Los dos archivos con extensión *ipynb* se encuentran en la carpeta “Objetivo 1/Notebooks/”

La división que hemos realizado de los sistemas de recomendación en dos grupos refleja la versatilidad de la solución para abordar tanto las necesidades internas de la asociación (derivadas de la planificación y producción de los menús), como las necesidades de los usuarios finales que buscan opciones de menús que se adapten mejor a sus gustos personales.

En algunos de estos sistemas se ha realizado una segunda versión, en la que se añadía nueva información (Sistema de Recomendación 2 - Versión 2) o en la que se probaban nuevos algoritmos (Sistema de Recomendación 3 - Versión 2).

Para garantizar que los resultados sean útiles y coherentes en la construcción de los sistemas de recomendación ha sido necesario aplicar una serie de transformaciones y correcciones previas. Dentro de la base de datos que tenemos alojada en AWS necesitábamos algunas de las tablas que contenían la información necesaria con los usuarios y los pedidos. Hemos aplicado las correcciones necesarias y nos hemos quedado con la información precisa que necesitábamos para la construcción de los sistemas. Posteriormente ha sido necesario utilizar otros documentos adicionales con las fichas técnicas de los platos, con sugerencias de nuevos platos para los usuarios y con los ingredientes de los platos<sup>14</sup>.

A la hora de comenzar con nuestro trabajo para crear los sistemas de recomendación se estuvo analizando la información de la que disponíamos para ver de qué manera podríamos usar la biblioteca Surprise para realizar los sistemas.

La biblioteca Surprise está diseñada para la creación y evaluación de sistemas de recomendación, sobre todo para el filtrado colaborativo. De manera que, en lugar de componer nuestro algoritmo de manera manual desde cero, podríamos simplificar la tarea y usar diferentes algoritmos para evaluar su comportamiento. Una vez establecidos los 4 sistemas de recomendación necesarios para nuestro proyecto, nos dimos cuenta que donde mejor funciona Surprise era a la hora de elaborar el sistema de recomendación número 3, basado en filtrado colaborativo.

Sin embargo, los datos que teníamos para establecer las recomendaciones no se adaptan a la forma en que esta biblioteca trabaja. Por ejemplo, a la hora de diseñar este "*Sistema de recomendación 3*", teníamos el problema de que necesitábamos las calificaciones de los usuarios para poder implementar los diferentes algoritmos. Como no disponíamos de esta información, adaptamos nuestro dataset para ofrecer calificaciones ficticias.

No hemos encontrado trabajos parecidos en donde se utilizara Surprise del mismo modo en el que nosotros lo queríamos implementar. Así que probamos a generar una calificación en base a esta información de la que disponíamos.

Una aproximación a esta solución era convertir las frecuencias de consumo en calificaciones ficticias, asignando una calificación más alta a los platos que se consumieron con mayor frecuencia. De tal forma que aunque no representan una

---

<sup>14</sup> Los tres archivos se encuentran en la carpeta - "Objetivo 2/Archivos/"

calificación real, podíamos tratarlos como calificaciones en el sistema de recomendación. Se consideró utilizar un sistema binario de calificación (1 si había consumido el plato y 0 si no lo había consumido) pero había una gran pérdida de información.

Al final, se realizaron dos modelos: uno con SVD y otro con KNN en la versión 2 de nuestro "Sistema de recomendación 3". Analizaremos los resultados en la siguiente sección.

Por último, cabe mencionar que en principio habíamos contemplado un único sistema de recomendación para los usuarios (el "Sistema de recomendación 3" con sus dos versiones, basado en el filtrado colaborativo). Sin embargo, siguiendo algunos de los principales trabajos realizados en el área (Yum\_me - Yang et al., 2017 y Trang Tran et al., 2018) en donde se generan sistemas para la recomendación de recetas y en donde se analizan los ingredientes de recetas junto a la preferencia de los usuarios, decidimos agregar un nuevo sistema debido a que disponíamos de la información necesaria (las fichas técnicas de los platos con los ingredientes). Lo cierto es que en comparación con los otros estudios nuestros datos tenían bastantes limitaciones, puesto que no disponíamos de un perfil del usuario tan desarrollado como en esos modelos, ni los componentes nutricionales de las recetas. De tal forma, que tuvimos que ajustar y preparar nuestro algoritmo desde cero, ajustándonos a nuestra capacidad de implementación, porque no tendríamos los parámetros necesarios para ejecutar el algoritmo de forma eficiente.

### **Objetivo 3: Mantener los niveles óptimos de insumos a través de la generación de modelos de series temporales.**

- Generación de históricos por ingredientes.

Una de las complejidades añadidas radicó en la diversidad de unidades de medida que se encontraban en los registros originales. Fue esencial estandarizar estas cantidades de ingredientes en un único sistema de unidades, ya sea en términos de peso y volumen, para asegurar coherencia y precisión en nuestro análisis.

Además, para completar el panorama, se hizo necesario recopilar información sobre los escandallos de los platos<sup>15</sup>. A pesar de contar con un número limitado de ellos, esta información resultó invaluable para comprender la estructura interna de los platos y la combinación de los ingredientes en las preparaciones culinarias. También, se utilizó un archivo con información relevante de los proveedores de productos, naturaleza perecedera, capacidad de almacenamiento, disponibilidad y entrega<sup>16</sup> por cada ingrediente para lograr tener un mejor *insight* de los mismos.

- Evaluación de la estacionalidad y la interdependencia en los históricos de los ingredientes.

---

<sup>15</sup> "platos\_final.xlsx" - en la carpeta "Objetivo3/Archivo/" .

<sup>16</sup> "ingredientes\_final.xlsx" - en la carpeta "Objetivo3/Archivo/"

Para elegir los algoritmos adecuados para nuestros datos, fue esencial realizar una evaluación para identificar tendencias, estacionalidades, ciclicidades y estacionalidades. Esta tarea no fue sencilla, especialmente cuando no se trata de evaluar los datos de un solo ingrediente, sino de más de 150 ingredientes que se utilizan para preparar los platos que se ofrecen semanalmente a los socios de Sukalde.

En los datos de series temporales, las observaciones son secuenciales y siguen un orden cronológico. A diferencia de los conjuntos de datos estándar para el aprendizaje automático, donde todas las observaciones se tratan por igual, las series temporales capturaron esta secuencialidad en el tiempo. La suposición fundamental en los datos de series temporales es que los patrones y comportamientos pasados se mantendrán en el futuro.

Antes de llevar a cabo esta evaluación, fue crucial preparar un dataframe de manera que las fechas se conviertan en índices dispuestos en secuencia semanal (cada 7 días). Además, se debían crear tantas columnas como ingredientes haya en el historial. Por último, se definió el conjunto de training y test, cuya proporción fue de 70:30, para poder realizar un análisis completo y detallado a pesar de las limitaciones de datos históricos.

- Desarrollo de múltiples modelos de series temporales univariantes.

Dentro de este enfoque, cada uno de los ingredientes corresponde a una serie temporal distinta por lo que se generó un modelo distinto para cada ingrediente con el fin de capturar los patrones en datos secuenciales a lo largo del tiempo de ese ingrediente en particular.

Según la evaluación de la estacionalidad aplicada en el apartado previo, la mayor parte de los ingredientes (casi un 80%) siguen un comportamiento estacionario, es decir, no presentan estacionalidad, ciclicidad, o tendencia. La interdependencia se tendría que asumir como irrelevante ya que uno de los supuestos de este tipo modelaje es que las series son independientes.

Para ejecutar todos estos modelos, se utilizó "SKForecast" que es una librería de Python que facilita el uso de regresores de scikit-learn como pronosticadores de series temporales univariantes y multivariantes. Además la librería SKForecast proporciona un conjunto completo de herramientas para entrenamiento, validación y predicción en una variedad de escenarios comúnmente encontrados al trabajar con series temporales.

Se utilizó una versión equivalente de ARIMA en la librería SKForecast llamada ARIMA-SARIMAX. Esta versión funciona solo como ARIMA si no se activan los componentes de estacionalidad. El modelo completo de SARIMAX si toma en cuenta los componentes de estacionalidad pero no fue utilizada en los modelos..

- Desarrollo de modelos de series temporales multivariantes.



En este enfoque, la cantidad de consumo de algunos de los ingredientes en el histórico podría tener relación con la de otros ingredientes debido al hecho de que podrían estar contenidos en la elaboración de un mismo plato o no cuenta con mucha información y los pronósticos se basarían en los patrones de los ingredientes de los platos más pedidos .

El modelo Vector Autoregression (VAR) no es compatible con la librería SKforecast por lo que se tuvo que hacer un procedimiento distinto al de los demás regresores. Se empleó el módulo de Python llamado "statsmodels" para importar o llamar al modelo VAR (Masum, 2021 y Singh, 2023). Para ejecutar el resto de los modelos, se utilizó la librería "SKForecast".

- Generación y comparación de las métricas de rendimiento y selección del modelo predictivo final.

Se determinaron los modelos que tenían el error cuadrático medio más bajo para cada ingrediente y se obtuvieron los porcentajes de frecuencia de los modelos univariantes con el MSE más bajo para seleccionar los dos modelos que se ajustaban mejor a los datos históricos.

Finalmente, se graficaron los datos del backtesting y de los pronósticos a 20 semanas para comparar los resultados de los mejores modelos con respecto a los históricos de un ingrediente en particular.

#### **Objetivo 4: Automatizar el análisis de las encuestas mediante la implementación de un sistema de análisis de sentimientos**

Realizamos un análisis descriptivo de las respuestas a las encuestas. En ellas fuimos aportando gráficos en donde se van mostrando los resultados pregunta a pregunta y en donde sacamos las conclusiones en función de dichos resultados. Posteriormente, determinamos las frecuencias de palabras por un conjunto de preguntas seleccionadas (aquellas que recogían texto en forma natural) y generamos las nubes de palabras o Wordcloud.

Debido a que debemos hacer nuestros análisis sobre texto en castellano teníamos algunas limitaciones al usar la biblioteca *nltk*. Se decidió utilizar *spaCy* por ser una biblioteca más adecuada para el procesamiento de lenguaje natural en español. En cuanto a cuestiones de rendimiento y precisión *spaCy* ofrece mejores puntuaciones que *nltk*, además es bastante fácil de utilizar, por lo que finalmente nos decantamos por el uso de esta biblioteca. Se realizó una función de limpieza y preprocesamiento de texto con *spaCy* donde lemmatizamos, normalizamos y eliminamos caracteres especiales.

A continuación, preparamos un etiquetado manual, asignando categorías como "*positivas*" o "*negativas*" a palabras en el código en forma subjetiva. Este criterio subjetivo se basó en la opinión del evaluador en la interpretación contextual de la carga emocional de las palabras. Este etiquetado manual se realizó para lograr una evaluación binaria del sentimiento en las respuestas.

Dentro del conjunto de preguntas que podríamos seleccionar para este trabajo, se realizó el etiquetado sólo para las preguntas 3, 11, 27, y 29 de la encuesta, debido que forman parte del grupo de preguntas que nos ofrecían opiniones sobre los aspectos generales de la comunidad Sukalde. No se consideraron las preguntas 30, 31, y 32 debido a que eran preguntas muy específicas acerca de cuál era su plato favorito o aquel que no le ha gustado, por lo que no se podía clasificar como una sensación positiva o negativa.

A continuación, evaluamos el modelo aplicando el etiquetado sólo a las respuestas de las preguntas que se referían al desempeño de la asociación Sukalde en el trabajo de la cocina comunitaria. Nuevamente, no consideramos las respuestas a las preguntas vinculadas a las preferencias culinarias específicas, aquellas que no se podían categorizar en términos de sentimiento “positivo” o “negativo”.

El modelado se hizo utilizando los algoritmos de regresión logística y Support Vector Classifier (SVC), a los que se aplicó un “*grid search*” previo para encontrar los parámetros óptimos. Generamos las métricas de desempeño del modelo. Las métricas utilizadas fueron: *precision*, *recall*, *f1-score*, *support*, *macro average* y *weighted average*.

Por último, se realizó el trabajo para generar una arquitectura que provea una interfaz visual que presente los resultados del análisis de sentimientos de manera comprensible para las partes interesadas.

Se utiliza Google Forms para poder recoger las respuestas de los usuarios, estas respuestas se van almacenando automáticamente en Google Sheets de manera que posteriormente se facilita su análisis a través del notebook desarrollado en nuestro proyecto. De esta forma, podemos utilizar las técnicas de procesamiento de lenguaje natural que hemos desarrollado a través de la biblioteca *spaCy* e implementar las técnicas y algoritmos para la clasificación de sentimientos. Por último, se presentan los datos de una manera de fácil interpretación a través de Looker Studio ofreciendo representaciones gráficas.

Se consigue así, una arquitectura automatizada de fácil acceso que permite realizar futuras encuestas y que permite a los miembros de Sukalde obtener las visualizaciones y los análisis directamente, pudiendo tomar decisiones en función de los insights proporcionados por las diferentes herramientas.

## RESULTADOS

A continuación, describimos los resultados obtenidos en las 4 soluciones planteadas para alcanzar nuestro objetivo de ***“Optimización de la eficiencia operativa y la experiencia de los usuarios en la cocina comunitaria Sukalde de Donostia”***:

### **Objetivo 1: Crear una base de datos consolidada**

Los resultados obtenidos en la consecución del primer objetivo específico del proyecto cumplen en su totalidad lo esperado por el equipo del Global Project. Se conformó la base de datos con la información disponible en la asociación y las tablas que registran los pedidos y las raciones de los usuarios se van actualizando adecuadamente semana a semana.

La carga de la base de datos en AWS se llevó a cabo exitosamente. Aunque el equipo del Global Project reconoce sus limitaciones en lo que respecta a una configuración precisa de una cuenta en AWS, particularmente en términos de seguridad, cuya configuración adecuada demanda un conocimiento técnico avanzado (que involucra configuración de la VPC, grupos de seguridad y usuarios IAM), se logró utilizar la base de datos de manera efectiva. Para los fines específicos de este proyecto, se consideró que la configuración existente resultaba adecuada y satisfactoria.

### **Objetivo 2: Automatizar la planificación de los menús y personalizar las sugerencias de los platos.**

#### **- Sistema de recomendación 1:**

Este sistema está basado en la popularidad por familias de platos. Es un algoritmo sencillo que funciona de manera efectiva para proporcionar los platos populares dentro de cada categoría de familia de platos. Responde a la demanda del equipo de cocina en cuanto a la planificación de los menús, reducción del tiempo de creación de los menús y seguimiento de los platos más populares por los miembros de la asociación.

A medida que la base de datos se vaya actualizando semana a semana, los resultados de estas tablas se irán modificando y los miembros del equipo de cocina pueden ir viendo la evolución de los platos más populares dentro de cada familia para ir adaptando los menús en función de sus necesidades.

#### **- Sistema de recomendación 2:**

Para este sistema de recomendación hemos generado dos versiones.

Versión 1 - Es un sistema basado en la aleatoriedad por familias. Nuevamente la primera versión de este sistema es un algoritmo muy sencillo que únicamente muestra de manera aleatoria los platos por familia. Responde también a las necesidades del equipo de cocina a la hora de considerar una mayor variabilidad en la oferta de los menús.

Versión 2 - Introduce nuevos platos no cocinados previamente en la lista anterior. Responde a la necesidad del equipo de cocina de Sukalde porque añade un elemento sorpresa en la selección de platos por los usuarios y puede conducir a mantener el interés y la satisfacción de los mismos al ofrecer variedad en la selección de platos.

- Sistema de recomendación 3:

En este sistema de recomendación hemos generado también dos versiones y en la segunda versión hemos probado dos algoritmos.

Versión 1 - Estamos bastante satisfechos con los resultados que nos ofrece esta primera versión del sistema de recomendación por similitud entre los usuarios. Es una versión que hemos diseñado casi al completo, funciona de manera efectiva y mantiene coherencia en los resultados al basarse en la matriz de similitudes entre usuarios.

A través de la matriz de similitud, cuyo empleo del coseno ha sido debidamente justificado en el notebook, se detectan los usuarios similares y en función de ellos se establecen las recomendaciones, sugiriendo platos que el usuario nunca ha consumido.

Los resultados que ofrecen las diferentes recomendaciones nos parecen precisos y con fundamentación matemática válida.

Versión 2 - La segunda versión de nuestro sistema de recomendación nos genera más dudas al respecto. A pesar de que hemos conseguido adaptar nuestros datos para implementar la biblioteca Surprise y utilizar diferentes algoritmos, no consideramos la métrica que hemos utilizado como válida para aceptar los resultados.

Obtenemos buenos resultados de RMSE y MAE en ambos algoritmos.

KNN		SVD	
RMSE: 0.0184		RMSE: 0.0477	
MAE: 0.0088		MAE: 0.0297	

Tienen valores bajos, con lo que nuestros modelos son precisos en sus predicciones y el modelo realiza recomendaciones que son muy similares a las preferencias reales de los usuarios. Sin embargo, consideramos que obtenemos buenos resultados bajo una premisa equivocada.

La aproximación que hemos utilizado como calificación de los usuarios condiciona estos resultados. Que un usuario haya elegido un plato el doble de veces que otro, no debería suponer que uno tenga una calificación 2 veces superior. Pudiera ser que una semana ese plato no podría ofertarse, con lo que las veces que ha elegido ese plato al final ha sido menor.

Por esta razón, aún a pesar de las buenas métricas obtenidas para los algoritmos de SVD y KNN en la versión 2 de nuestro tercer sistema de recomendación, recomendamos usar la primera versión. Sus resultados son más precisos y respaldados con mejor fundamentación matemática.

- Sistema de recomendación 4:

El último sistema de recomendación estaba basado en recomendación por ingredientes.

Hemos justificado el uso de la métrica Jaccard en este último sistema, ya que tratamos de calcular similitud entre conjuntos de ingredientes y para esto, es la métrica adecuada.

Como podemos ver, los resultados obtenidos en el usuario de prueba tienen un valor de similitud no demasiado elevado:

```
5 platos recomendados para el usuario 47 basados en los ingredientes de sus platos favoritos:
-----
Plato: Arroz integral c/ curry amarillo, Puntuación de Similitud: 7.53%
Plato: Fideuá con pollo y hongos:, Puntuación de Similitud: 7.53%
Plato: Guisantes, alcachofas y jamon, Puntuación de Similitud: 7.53%
Plato: Corvina en salsa verde, Puntuación de Similitud: 7.53%
Plato: Trinxat de la cerdanya, Puntuación de Similitud: 7.53%
```

Por un lado podemos observar que todos los platos tienen la misma similitud y puede ser indicativo de que podemos estar cometiendo algún error, sin embargo podemos probar solicitando un número mayor de platos y comprobar que funciona bien:

```
Plato: Osobuco de ternera con setas, Puntuación de Similitud: 5.73%
Plato: Salsa bolognesa, Puntuación de Similitud: 5.02%
Plato: Bizcocho de chocolate y zanahor, Puntuación de Similitud: 3.23%
Plato: Ensalada alubia y bacalao, Puntuación de Similitud: 3.23%
Plato: Ensalada de garbanzo, Puntuación de Similitud: 3.23%
```

Y por otro lado, podemos ver el puntaje de similitud del plato en función de los ingredientes que contienen sus platos favoritos. Como se justificaba al final del sistema de recomendación 4, para la gran variedad de ingredientes de que disponemos, no disponemos en contraparte de una muestra suficiente de platos que permita al algoritmo encontrar platos similares, de tal forma que obtenemos valores con un puntaje porcentual algo bajo.

En definitiva, al igual que en la primera versión del último sistema de recomendación, recomendamos como sistema de recomendación para usuarios la primera versión del "Sistema de recomendación 3" puesto que en este caso no disponemos de un conjunto de datos con una información suficiente que nos pueda ofrecer unas recomendaciones significativas.

**Objetivo 3: Mantener los niveles óptimos de insumos a través de la generación de modelos de series temporales.**

Debido a las limitaciones en los datos históricos de pedidos (menor a 11 meses) de la cocina comunitaria Sukalde, hemos adoptado dos enfoques en el análisis del modelo

predictivo de ingredientes. En el primer enfoque, generamos modelos individuales para cada ingrediente capaces de realizar pronósticos teniendo en cuenta únicamente los patrones y tendencias del ingrediente en cuestión (múltiples modelos de series temporales univariantes). Por otro lado, en el segundo enfoque, desarrollamos un modelo capaz de hacer predicciones para todos los ingredientes, basándose en los patrones y tendencias no solo de un ingrediente, sino también de los otros ingredientes, asumiendo que estos podrían depender los unos con los otros (modelos de series temporales multivariantes).

De la evaluación de estacionalidad e interdependencia, se ha concluido que en la mayoría de los ingredientes no existen patrones estacionales ni períodos de ciclos repetidos, por tanto, los datos son predominantemente 'estacionarios'.

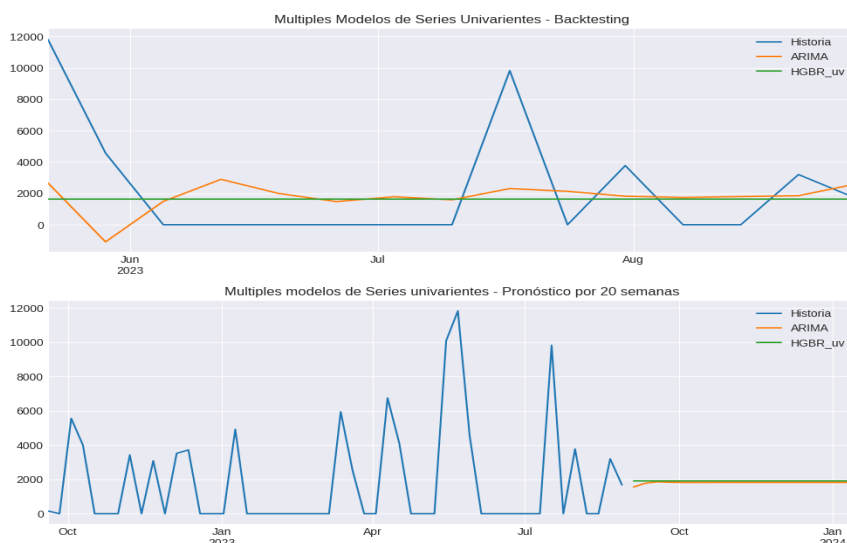
#### - Múltiples Modelos de Series Temporales Univariantes:

Se evaluaron un total de 7 algoritmos diferentes para este enfoque. Para cada uno de estos algoritmos, se creó un modelo individual para cada ingrediente, permitiendo así prever el comportamiento futuro de cada ingrediente. Estos algoritmos son: **Autoregressive Integrated Moving Average (ARIMA), Linear Regression, KNeighbors, Decision Tree, Random Forest y Histogram-Based Gradient Boosting XGB**

Los porcentajes de frecuencia de los modelos univariantes con el MSE más bajo son:

mse_HGBR_uv	32.500
mse_ARIMA	26.250
mse_KNN_uv	18.125
mse_RF_uv	9.375
mse_XGBR_uv	6.875
mse_DT_uv	3.750
mse_LR_uv	3.125
Name: mse_minimo, dtype: float64	

Esto indica que el modelo Histogram-Based Boosting univariante tiene el mayor porcentaje (33%) en tener el MSE más bajo con respecto a los otros modelos. En segunda posición, está el algoritmo ARIMA, con un 26%.



Al graficar el backtesting utilizando los modelos univariantes con los errores mínimos (HGBR, ARIMA) para el ingrediente “Arroz”, se observa casi un valor constante a pesar de que el modelo de ARIMA muestra una curva con muy poca variación que pueda atribuirse a un cierto tipo de ruido. El mismo resultado se obtiene al graficar las predicciones para las próximas 20 semanas. Esto se debe al hecho de que hay muy pocos puntos de datos en el historial de ese ingrediente como para capturar patrones significativos. En su lugar, el modelo produce predicciones casi constantes para minimizar el error, ya que no hay suficiente variabilidad en los datos para hacer predicciones más precisas.

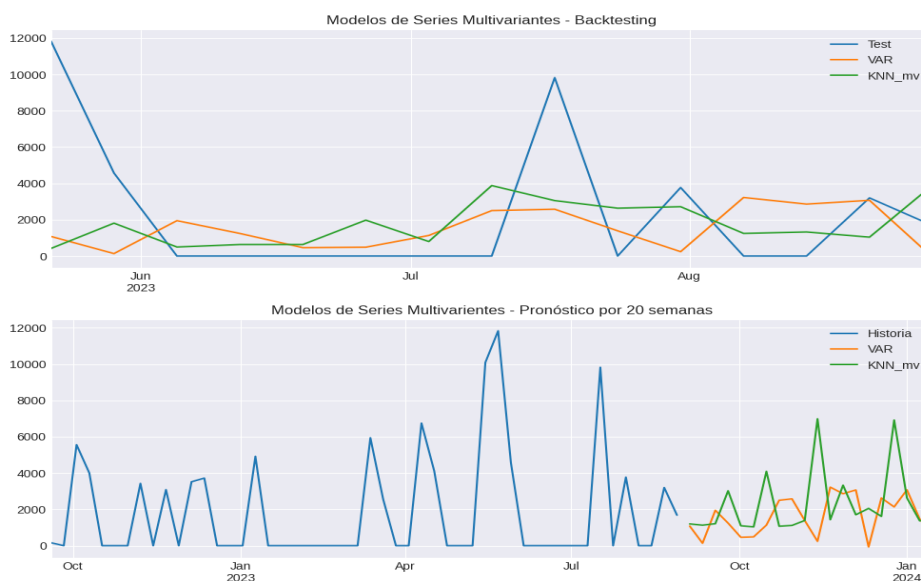
#### - Modelos de Series Temporales Multivariantes:

Se evaluaron un total de 7 algoritmos diferentes en este enfoque. Para cada uno de estos algoritmos, se creó un único modelo capaz de prever los comportamientos futuros de cada ingrediente. Estos algoritmos son: **Vector Autoregression (VAR)**, **Linear Regression**, **KNeighbors**, **Decision Tree**, **Random Forest** y **Histogram-Based Gradient Boosting XGB**.

Los porcentajes de frecuencia de los modelos multivariantes con el MSE más bajo son:

mse_KNN_mv	20.000
mse_VAR	19.375
mse_LR_mv	17.500
mse_RF_mv	16.250
mse_DT_mv	10.625
mse_HGBR_mv	10.000
mse_XGBR_mv	6.250
Name: mse_minimo, dtype: float64	

Esto indica que el modelo de KNeighbors multivariante tiene el mayor porcentaje (20%) en tener el MSE más bajo con respecto a los otros modelos multivariantes. En segunda posición está el modelo VAR con un 19%.



A diferencia del modelo univariante, al graficar el backtesting utilizando el modelo multivariante con los errores mínimos (KNN y ARIMA) para el ingrediente "Arroz", se observan variaciones en la curva. Esto se debe al hecho de que el modelo, además de considerar los patrones y tendencias del historial del ingrediente en cuestión, también generaliza los patrones y tendencias de otros ingredientes. Esta consideración de patrones múltiples permite ajustar el backtesting de manera más ajustada, lo que se traduce en predicciones más consistentes con la historia. Al incluir información de varios ingredientes, el modelo tiene una mayor capacidad para adaptarse a las complejidades de los datos y capturar patrones que un enfoque univariante no podría identificar por sí solo.

#### **Objetivo 4: Automatizar el análisis de las encuestas mediante la implementación de un sistema de análisis de sentimientos**

Para nuestro último objetivo vamos a dividir el análisis de los resultados en 3 partes: los resultados de la encuesta realizada por la asociación, los resultados que ofrecen los modelos que hemos generado y el resultado de la creación de la arquitectura para automatizar el proceso de análisis para futuras encuestas.

La mayor parte de las preguntas de la encuesta de Sukalde van dirigidas a responder cuestiones acerca de la satisfacción de los miembros con respecto a los platos, la eficiencia en el desempeño de la asociación en el trabajo de la cocina comunitaria, valoraciones en cuanto al precio, calidad o la variedad de los platos, y la percepción de los usuarios en cuanto a su ahorro en coste, su reducción a la hora de generar desperdicio o la sensación de aumento de tiempo libre desde su participación en el proyecto, entre otras.

En el notebook se muestra el análisis completo con las conclusiones obtenidas pregunta a pregunta, y en términos generales podemos resaltar que las respuestas proporcionadas en su conjunto muestran una sensación positiva con respecto a cada una de las cuestiones.

A continuación, analizaremos los resultados generados por nuestros modelos:

- Resultados de nuestro Modelo Personalizado:

	precision	recall	f1-score	support
<b>Negativo</b>	1.000000	0.416667	0.588235	12.0
<b>Positivo</b>	0.892308	1.000000	0.943089	58.0
<b>accuracy</b>	0.900000	0.900000	0.900000	0.9
<b>macro avg</b>	0.946154	0.708333	0.765662	70.0
<b>weighted avg</b>	0.910769	0.900000	0.882257	70.0

El modelo personalizado acierta un 90% en los valores positivos y un 100% en los valores negativos lo que quiere decir que el modelo tienen un rendimiento aceptable pero tiene un dificultad de clasificar correctamente las respuestas negativas en un 41%



- Resultados modelo de regresión logística:

	precision	recall	f1-score	support
Negativo	0.00	0.00	0.00	5
Positivo	0.76	1.00	0.86	16
accuracy			0.76	21
macro avg	0.38	0.50	0.43	21
weighted avg	0.58	0.76	0.66	21

La regresión logística genera un 76% de accuracy y el modelo tiene dificultad para predecir los comentarios negativos, clasificándolos positivos.

- Resultados SVC

	precision	recall	f1-score	support
Negativo	0.00	0.00	0.00	3
Positivo	0.85	0.94	0.89	18
accuracy			0.81	21
macro avg	0.42	0.47	0.45	21
weighted avg	0.73	0.81	0.77	21

Como vemos, tampoco obtenemos muy buenos resultados.

En definitiva, a través de nuestro análisis de sentimientos, hemos identificado que el modelo personalizado presenta un nivel de precisión considerable, alcanzando un accuracy del 90%. Sin embargo, se evidencia una dificultad significativa en la predicción precisa de comentarios negativos en los otros modelos. Esta limitación puede atribuirse en gran medida al desequilibrio en el conjunto de datos, donde las respuestas negativas son notablemente menos frecuentes que las positivas.

Este hallazgo resalta la importancia crítica de la diversidad y el equilibrio en la recolección de datos para el aprendizaje automático. Un conjunto de datos robusto y bien equilibrado no solo mejora la precisión del modelo, sino que también asegura que el modelo sea equitativo y no esté sesgado hacia una clase particular. En el contexto del proyecto de Sukalde, esto es esencial para obtener una representación precisa de la percepción del servicio por parte de los usuarios.

El modelo personalizado demostró ser el más prometedor. A pesar de su alta precisión, es crucial destacar la necesidad de mejorar la detección de comentarios negativos, ya que el modelo mostró un recall del 41% en esta categoría. Este aspecto es esencial para Sukalde, ya que identificar y abordar las críticas y preocupaciones de los usuarios es fundamental para mejorar y adaptar el servicio ofrecido.

En resumen, mientras que los modelos actuales proporcionan insights valiosos, hay espacio para mejoras, especialmente en la detección de sentimientos negativos. Continuar con la recolección de datos y la iteración del modelo será crucial para

ofrecer una herramienta de análisis de sentimientos más precisa y efectiva para Sukalde.

Por último, con respecto a la arquitectura para automatizar el proceso a futuras encuestas, se ha conseguido diseñar la solución integral eficientemente. Está alineada con las necesidades actuales del equipo de Sukalde. Se ha priorizado la simplicidad y la accesibilidad, garantizando que la interfaz sea intuitiva y fácil de usar. Además, la estructura propuesta no solo es económica, evitando gastos adicionales, sino que también es escalable, ofreciendo flexibilidad para adaptarse y crecer según las demandas cambiantes de la comunidad.

## CONCLUSIONES Y TRABAJOS FUTUROS

Las soluciones implementadas aportan un potencial de mejora en la eficiencia operativa, la satisfacción del usuario y en la gestión de los recursos.

Este proyecto ha demostrado la versatilidad de los sistemas de recomendación al abordar las necesidades tanto internas de la asociación Sukalde (relacionadas con la planificación y producción de menús) como las necesidades de los usuarios finales que buscan opciones de menús personalizadas. Los resultados anticipados de la implementación de estos sistemas incluyen mejoras en la eficiencia operativa de Sukalde mediante una reducción del tiempo dedicado a estas tareas, así como una mayor satisfacción de los usuarios al recibir sugerencias personalizadas y una oferta de menús más adaptada a sus gustos.

Los modelos univariantes y multivariantes en la previsión de ingredientes permiten explorar tanto características individuales como complejas relaciones entre ingredientes, siendo útil en datos limitados. Analizar patrones individuales e interdependencias genera predicciones más precisas, aunque los modelos multivariantes simplifican el mantenimiento al ajustar todas las variables con un solo modelo, lo que puede ser más eficiente. Optimizar hiperparámetros usando grid search es crucial para mejorar la precisión. Los modelos multivariantes permiten controlar los pesos, personalizando el aprendizaje, aunque algoritmos como HGBR y KNN enfrentan desafíos en datos históricos complejos. Para superar esto, se recomienda el uso de métodos avanzados como SARIMA, Holt-Winters y Redes Neuronales Recurrentes (RNN) y LSTM para capturar elementos de estacionalidad y tendencias así como patrones complejos y dependencias a largo plazo en los datos de Sukalde.

El modelado de análisis de sentimientos y la arquitectura automatizada desarrollada, permite a los líderes de la comunidad Sukalde tomar decisiones informadas basadas en el feedback de sus miembros. Se brinda una solución integral que se inicia con la recopilación de datos a través de Google Forms, sigue con el procesamiento y análisis en Google Colab, y culmina con la visualización de los resultados en Looker Studio. Esta arquitectura, utiliza a Google Sheets como medio de almacenamiento, es no solo eficiente sino también rentable.

Hemos alcanzado importantes progresos en la realización de los objetivos del proyecto, lo que establece una sólida plataforma para futuras mejoras en el futuro. Se identifican espacios de mejora que pasan por una recogida con datos que muestren más información y por un mejor ajuste y mejor desarrollo de los algoritmos. Para ello, la información disponible de los platos y las preferencias o restricciones de los usuarios deben ser también más elaboradas. Una muestra de datos más amplia y de mayor calidad potenciarán los modelos, ofreciendo insights más precisos y confiables.

## REFERENCIAS

- Pérez, S., Cuecuecha, M. C., Ramírez, J. F., & Hernández, J. C. (2018). Análisis, diseño y desarrollo de un sistema de recomendación basado en datos restauraneros de TripAdvisor y Foursquare. *Res. Comput. Sci.*, 147(5), 209-224.
- Trang Tran, T.N., Atas, M., Felfernig, A. et al. An overview of recommender systems in the healthy food domain. *J Intell Inf Syst* 50, 501–526 (2018).  
<https://doi.org/10.1007/s10844-017-0469-0>
- Yang, L., Hsieh, C. K., Yang, H., Pollak, J. P., Dell, N., Belongie, S., ... & Estrin, D. (2017). Yum-me: a personalized nutrient-based meal recommender system. *ACM Transactions on Information Systems (TOIS)*, 36(1), 1-31.
- Méndez López, F. (2018). Sistemas de recomendación en el contexto gastronómico: elaboración y enriquecimiento de un dataset de recetas de cocina.
- Forbes, P., & Zhu, M. (2011, October). Content-boosted matrix factorization for recommender systems: experiments with recipe recommendation. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 261-264).
- Freyne, J. and Berkovsky, S. (2010). Intelligent food planning: Personalized recipe recommendation. In *Proceedings of the 15th International Conference on Intelligent User Interfaces*, pages 321–324.
- Amat, J & Escobar, J. (2023). Multi-series forecasting con Python y SKForecast. <https://cienciadedatos.net/documentos/py44-multi-series-forecasting-skforecast-espa%C3%B1ol.html>
- Singh, A. (2023). Multivariate time series analysis with Python for forecasting and modeling (Updated 2023). Analytics Vidhya. [https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/#Univariate\\_Vs.\\_Multivariate\\_Time\\_Series\\_Forecasting\\_Python](https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/#Univariate_Vs._Multivariate_Time_Series_Forecasting_Python)
- Masum. (2021, 15 diciembre). A Real-World Application of Vector Autoregressive (VAR) model. Medium. <https://medium.com/p/456ace675971>
- Garrido Marín, C. (2021). Previsión de stock con técnicas de Inteligencia Artificial para distribuidoras de alimentación y bebidas - Trabajo de Fin de Grado. Alicante, España. Universidad de Alicante - Escuela Politécnica Superior. Páginas 1-52.
- Praveen, K., Prateek, J., Pradyumna, K., Pragathi, G., Madhuri, J. (2020). Inventory Management using Machine Learning. Bengaluru, India. *International Journal of Engineering Research & Technology (IJERT)*. Páginas 866-869.
- Nassibi, N., Fasihudding, H., Hsairi, L. (2023). Demand Forecasting Models for Food Industry by Utilizing Machine Learning Approaches. Saudi Arabia. *International Journal of Advanced Computer Science and Applications*.

- Molina González, M. D., Martínez Cámara, E., & Martín Valdivia, M. T. (2015). CRISOL: Base de conocimiento de opiniones para el español. Procesamiento del Lenguaje Natural, (55), 143-150. Sociedad Española para el Procesamiento del Lenguaje Natural.
- Calvo Madurga, A. (año de publicación). Análisis de sentimientos y emociones en redes sociales usando ML. Trabajo de grado, Universidad de Valladolid. Tutelado por: Valentín Cardeñoso Payo.
- Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media, Inc.
- Battocchia, M. (2020, 23 de julio). Preprocesamiento de texto para NLP (parte 1) (=^·ω·^=) Limpieza, tokenización y más. Recuperado de [enlace del repositorio GitHub](#).
- García, G. (2021). Las emociones contribuyen a la elección y consumo de alimentos. THE FOOD TECH - Medio De Noticias Líder En La Industria De Alimentos Y Bebidas. <https://thefoodtech.com/marketing/las-emociones-contribuyen-a-la-eleccion-y-consumo-de-alimentos/>
- Universidad De Navarra (2022) La cocina, además de ingredientes, está hecha de emociones y sentimientos, y por ello, llevo esa pasión en el ámbito en el que me encuentre. <https://www.unav.edu/noticias/-/contents/08/07/2022/la-cocina-ademas-de-ingredientes-esta-hecha-de-emociones-y-sentimientos-y-por-ello-llevo-esa-pasion-en-el-ambito-en-el-que-me-encuentre/content/lovPblW1fC70/39612266>
- Sánchez, C. R. (2023, July 21). ¿Cómo puede ayudar la Inteligencia Artificial al sector de la restauración? Restauración News. <https://restauracionnews.com/2023/07/inteligencia-artificial-restauracion/>
- Thomas, C. A. (2022). Análisis de sentimiento: ¿cómo escuchar a tus clientes? Branch Agencia. <https://branch.com.co/marketing-digital/analisis-de-sentimiento-como-escuchar-a-tus-clientes/>