



Facultad de Ciencias Económicas y Empresariales (ICADE),  
Universidad Pontificia de Comillas

# **MANOS UNIDAS. Análisis cuantitativo y cualitativo de los mensajes recibidos del buzón de Manos Unidas.**

Clave: 201911115

## **RESUMEN:**

Estudio realizado para Manos Unidas con el objetivo de optimizar la gestión de los recursos disponibles y de aumentar el impacto de las campañas. Se realiza un análisis exhaustivo de los mensajes recibidos a través de un buzón en la página web de la organización. Se implementaron diferentes técnicas avanzadas de procesamiento de datos y análisis de texto para prevenir los mensajes de spam mediante un modelo de Random Forest. También se particionó la población según el tipo de mensajes mediante el método de clustering, analizando los sentimientos de los donantes e identificando patrones y tendencias para cada grupo de mensajes (o cluster).

Este estudio ha podido identificar los períodos de alta actividad de mensaje, para poder realizar una mejor planificación y asignación de los recursos disponibles, y también mejorar su relación con los socios y colaboradores.

## **ABSTRACT:**

Study conducted for Manos Unidas with the aim to optimize the management of available resources and increase the impact of campaigns. An exhaustive analysis of messages received through a mailbox on the organization's website was carried out. Various advanced data processing and text analysis techniques were implemented to prevent spam messages using a Random Forest model. Additionally, population partitions were made based on the type of messages using the clustering method, analyzing the sentiments of donors and identifying patterns and trends for each group of messages (or cluster).

This study has been able to identify periods of high message activity, allowing for better planning and allocation of available resources, as well as improving the relationship with members and collaborators.

## **PALABRAS CLAVE:**

Manos Unidas, ONG, mensajes, análisis de clustering, cluster, detección de spam, análisis de sentimiento, transformador BERT, polaridad, series temporales, PCA, bosque aleatorio.

## **KEY WORDS:**

Manos Unidas, NGO, messages, clustering analysis, cluster, spam detection, sentiment análisis, BERT transformer, polarity, time series, PCA, Random Forest.

# ÍNDICE

CAPITULO I: MARCO TEÓRICO.....	4
1.1. Visión general de Manos Unidas .....	4
1.2. Rol de las ONGs en el Desarrollo Global y su Evolución .....	5
CAPITULO II: OBJETIVOS .....	8
2.1. Objetivo General .....	8
2.2. Objetivos Específicos.....	8
CAPITULO III: METODOLOGÍA .....	10
3.1 Antecedentes .....	11
3.2. Extracción y limpieza de la muestra .....	12
3.3. Modelo de Detección de Spam.....	14
3.4. Análisis de clustering .....	15
3.5. Análisis del sentimiento de los mensajes .....	18
3.6. Análisis de series temporales .....	20
CAPITULO IV: ANÁLISIS DE RESULTADOS OBTENIDOS.....	23
4.1. Resultados Modelo Detección de Spam.....	23
4.2. Resultados Análisis Clustering.....	25
4.3. Resultados Sentiment Analysis .....	34
4.4. Resultados Serie Temporal.....	37
CAPITULO V: CONCLUSIONES.....	43
CAPITULO VI: FUTURAS LÍNEAS DE INVESTIGACIÓN .....	46
CAPÍTULO VII: ANEXOS .....	48
7.1. Declaración de Uso de Herramientas de Inteligencia Artificial en el Trabajo de Fin de Grado.....	48
7.2. Bibliografía .....	50

## ÍNDICE FIGURAS.

Figura 1: Representación Muestra de ejemplo con PCA.....	16
Figura 2: Matriz de Confusión .....	23
Figura 3: Curva ROC .....	25
Figura 4: Representación PCA de datos NO filtrados.....	26
Figura 5: Representación PCA de datos filtrados manualmente .....	26
Figura 6: Representación PCA de datos filtrados .....	27
Figura 7: Gráfico método del Codo para k óptimo .....	28
Figura 8: Análisis de Silueta para k óptimo .....	29
Figura 9: Partición Clustering Analysis (k=3) .....	30
Figura 10: Palabras más frecuentes del CLUSTER 1 .....	31
Figura 11: Palabras más frecuentes del CLUSTER 2 .....	32
Figura 12: Palabras más frecuentes del CLUSTER 3 .....	33
Figura 13: Polaridad Mensajes del Cluster 1.....	34
Figura 14: Polaridad Mensajes del Cluster 2.....	35
Figura 15: Polaridad Mensajes del Cluster 3.....	36
Figura 16: Componentes de la Serie Temporal del Cluster 1.....	37
Figura 17: Predicción de ARIMA (2,0,0)(2,0,0)[12] para el Cluster 1 .....	37
Figura 18: Componentes de la Serie Temporal del Cluster 2.....	38
Figura 19: Predicción de ARIMA (0,1,1) para el Cluster 2 .....	39
Figura 20: Componentes de la Serie Temporal del Cluster 3.....	40
Figura 21: Predicción de ARIMA (0,1,1) para el Cluster 3 .....	40
Figura 22: Frecuencia Mensajes Mensuales por Año .....	41
Figura 23: Frecuencia Mensajes Mensuales por Año (Representación Polar) (Elaboración propia).....	42

# **CAPITULO I: MARCO TEÓRICO**

## **1.1. Visión general de Manos Unidas**

Desde el surgimiento de las organizaciones no gubernamentales, estas entidades han ido evolucionando significativamente, adaptándose a los cambios sociales y tecnológicos para maximizar su impacto y eficiencia. A través de campañas de sensibilización, proyectos de desarrollo y acciones humanitarias, las ONGs buscan mejorar la calidad de vida de las personas más vulnerables. Estas organizaciones han sido uno de los factores clave en la configuración de las respuestas humanitarias y de desarrollo desde el siglo XX. Inicialmente, surgieron como entidades de asistencia directa y de manera voluntaria, proporcionando ayuda en contextos atípicos de extrema emergencia y crisis. No obstante, con el transcurso del tiempo, su papel se ha expandido para incluir una amplia gama de movimientos, desde la lucha por los derechos humanos hasta la implementación de proyectos de contención del cambio climático.

En el caso de Manos Unidas, es una Organización No Gubernamental para el Desarrollo (ONGD) española, fundada en 1960 y perteneciente a la Iglesia Católica, que se dedica a la lucha contra el hambre, la pobreza, y la desigualdad en el mundo. Y, además, no solo emplea todos sus esfuerzos en poner fin a estos problemas sociales, sino que busca sus causas en los países más desfavorecidos del mundo. La visión de esta institución ha sido y sigue siendo concisa pero ambiciosa, “Acabar con el hambre es posible” (Manos Unidas, 2023). Así pues, para conseguir esta meta se llevan a cabo dos tipos de propuestas, la primera está enfocada en dar a conocer y sensibilizar la presencia del hambre y la pobreza no solo en los países no desarrollados sino también en la sociedad española. Además de ello, investiga acerca de los factores que pueden causar dicha pobreza y propone posibles soluciones que provoquen una mayor concienciación sobre este problema social tan importante y así aumentar el compromiso en erradicar el hambre y la pobreza.

Por otra parte, el segundo tipo de propuesta realizada por Manos Unidas es la toma de acción y desarrollo para poner fin al hambre y la pobreza mediante la recaudación de fondos económicos. Para que con estos presupuestos, se puedan financiar todas las misiones y proyectos de desarrollo integral, para poder atender las necesidades de las personas más desamparadas, con faltas de alimento y recursos económicos.

A lo largo de los años, ha implementado numerosas campañas y programas para mejorar las condiciones de vida de millones de personas. Tan solo en 2022 se consiguió destinar en torno a 34 millones de euros para proporcionar apoyo a más de un millón y medio de personas alrededor del mundo que se hallaban en condiciones de extrema pobreza y con riesgo de exclusión (Manos Unidas, 2023). En 2022 también se dieron luz verde a 474 nuevos proyectos para cubrir las necesidades más recientes. Sin embargo, ese año no es un caso aislado, en el año previo a la pandemia del COVID-19, es decir, en 2019, se llevaron a cabo 540 proyectos nuevos en 57 países. En 2019 la organización logró una recaudación de más de 43 millones de euros, de los cuales el 90% fue destinado a cumplir los dos objetivos de la asociación no gubernamental (Tufani, 2020).

En cuanto a la localización geográfica donde Manos Unidas presta servicio, según la memoria que publicó la organización el año pasado sobre el ejercicio del 2022, los proyectos se despliegan mayoritariamente en los continentes del sur. La distribución de misiones está distribuida de tal forma que el 43% de éstas se realizan en el continente africano, mientras que el 33% y 24% restantes se desarrollan en el Asia y en América latina, correspondientemente. Agregado a lo anterior, el informe también notifica las áreas o industrias donde la organización tiene un mayor enfoque justificado mediante el número de proyectos e inversión destinada. Estos sectores son mayoritariamente: la alimentación y medios de vida para acabar con el hambre; la educación universal para ayudar a los jóvenes a salir de la pobreza; y la salud básica en las personas más desfavorecidas (*Memoria 2022 - Nuestra indiferencia los condena al olvido*, 2023). No obstante, existen otras problemáticas que también trata Manos Unidas, como viene siendo el agua y saneamiento en las zonas con más sequía, la lucha por la igualdad y los derechos de las mujeres, la protección y aseguramiento de los derechos humanos y de la sociedad civil, o por ejemplo el cuidado del medio ambiente y la concienciación del cambio climático.

## **1.2. Rol de las ONGs en el Desarrollo Global y su Evolución**

En el camino de la sociedad hacia el desarrollo global más justo, los actores principales siempre son los gobiernos centrales, y sobre todo los de los países occidentales, que son los que cuentan con más recursos económicos y políticos. Sin embargo, las ONGs también han jugado un papel fundamental para el desarrollo y mejoría de la calidad de

vida de las personas en el tercer mundo, ya que existen muchas áreas donde los gobiernos no logran cubrir todas las necesidades de la población. Es tal este rol que según Alejandro Segovia y Melissa González, “las ONGs desempeñan un papel cada vez más importante en la promoción de la cooperación global” (Segovia-Guerrero & González-Cantos, 2024). Lo que puede significar que la suma de todos los esfuerzos de las diferentes organizaciones no gubernamentales puede iniciar una colaboración entre diversos países del mundo con el fin de solucionar una injusticia social y política.

El rol de las organizaciones voluntarias y no lucrativas puede contribuir a hacer una gran labor en la ayuda y rescate de personas que están en situaciones de necesidad. A pesar de esto, es probable que una ONG por si sola aporte una ayuda insuficiente o escasa debido a la limitación de sus recursos, los cuales provienen mayoritariamente de donaciones particulares y recaudación de fondos privados. Ahora bien, si todas las instituciones y organismos cooperan entre ellas para conseguir un mismo fin, el resultado de la ayuda proporcionada puede ser muy diferente. En el caso de España, las principales ONGs del país han desarrollado diferentes estrategias de colaboración que les permiten una maximización del impacto de sus intervenciones en los lugares de necesidad, pudiendo así alcanzar los objetivos de ayuda y servicio comunes de manera más efectiva. Entre estos organismos destacan Cáritas, CEAR, Cruz Roja, FUNDESO, y por supuesto, Manos Unidas (Carramiñana et al., 2013). Estas organizaciones han ido desarrollando su servicio en el sector desde hace años, han podido entablar un mayor número de colaboraciones con otras ONGs o incluso instituciones públicas, así pudiendo provocar un mayor cambio positivo en sus campañas.

Las ONGs han evolucionado significativamente desde su aparición en el siglo XX, adaptándose a los cambios sociales, políticos y económicos. Tal ha sido la evolución, que el origen y el desarrollo de la educación para el progreso está estrechamente relacionado con la aparición de las ONGs. Éstas han colaborado de manera crucial en la sensibilización y posterior movilización de la sociedad española y occidental para tratar de poner fin a las injusticias que suceden constantemente alrededor del globo (Mesa, 2014).

Originalmente, las organizaciones no lucrativas tenían el enfoque centrado en la caridad y la asistencia social. No obstante, las ONGs con un origen más moderno han adoptado unas perspectivas y visiones más sostenibles para el desarrollo social. La globalización y el acceso a nuevas tecnologías han permitido a estas organizaciones expandir su alcance

y mejorar la eficiencia de sus proyectos. Donde como indica Rafael Marfil-Carmona, con todas las posibilidades que ofrece Internet en la actualidad, las ONGs necesitan buscar y captar nuevos recursos, ya sean económicos o a través del voluntariado (Marfil-Carmona, 2014).

Algunos de estos recursos mencionados que están siendo más prolíferos son las herramientas de marketing empleadas para las campañas emitidas en diferentes canales de difusión. Es así que en el estudio conducido por Paredes Saavedra y Sonia Aurora concluyeron que “existe una relación significativa entre el Marketing Relacional y Fidelización de clientes” (Saavedra & Aurora, 2019). Esto significa que en los proyectos que tienen un marketing relacional deficiente, los clientes potenciales no responden con su fidelización, sin embargo, con un marketing relacional tanto efectivo como eficiente, la fidelización de los clientes aumenta de manera satisfactoria con los clientes.

En conclusión, Manos Unidas es una ONG que aporta mucho valor a la sensibilización y el desarrollo en los lugares donde aún existen injusticias sociales. Sin embargo, aún hay mucho potencial para gestionar los recursos tanto económicos como administrativos de una manera más óptima, y consecuentemente, realizar campañas que tengan un mayor impacto social.

## CAPITULO II: OBJETIVOS

### 2.1. Objetivo General

El objetivo general del estudio es la mejora del impacto y la eficiencia de la organización de Manos Unidas mediante un **análisis exhaustivo en base a los datos disponibles**, con el objetivo de **optimizar la gestión de los recursos** disponibles para poder redistribuir un porcentaje mayor de las recaudaciones y aumentar el impacto las campañas. Esta investigación se fundamenta en una base de datos estructurada y anonimizada, proporcionada por Manos Unidas, en la cual se pretende implementar un análisis detallado y profundo de los mensajes recibidos por la organización. Asimismo, el propósito de este análisis exhaustivo es **identificar patrones y tendencias** que le permitan a la organización responder de la manera más eficiente, para poder, no solo atender a las necesidades y expectativas de sus beneficiarios y donantes, sino también proporcionar asistencia más completa y rápida a todas las peticiones y consultas enviadas a través del buzón de su página web.

### 2.2. Objetivos Específicos

1. **Extraer la información e identificar mensajes clave.** Extraer la información en tablas de la base de datos de Manos Unidas para identificar los mensajes clave. Y posteriormente, implementar técnicas de procesamiento y filtrado de datos para tener unos datos de mensajes limpios y libres de spam. De esta forma se habrá garantizado que los mensajes considerados para el posterior análisis son los relevantes.
2. **Prevenir de futuros mensajes Spam.** Filtrar con las palabras que sean dudosas de spam. Añadir modelo ML de predicción de spam con el cual reducir el trabajo manual del primer filtrado para no solo eliminar los mensajes de Spam una vez enviados, sino también prevenir el trabajo de eliminación mediante una automatización estudiada por el modelo de ML.
3. **Clustering.** Mediante un análisis de clustering, encontrar los grupos de mensajes y de socios o colaboradores que dan valor a la organización para intentar responder mejor sus cuestiones y entender los problemas que les surgen.
4. **Evaluar el sentimiento de los mensajes.** Utilizando los resultados obtenidos del anterior análisis de clustering, efectuar un análisis del sentimiento, o sentiment

analysis, para evaluar de manera cualitativa las emociones y la polaridad de los mensajes enviados para cada clúster.

- 5. Realizar un análisis temporal de los mensajes.** Desarrollar una serie temporal que analice la frecuencia y patrones temporales de los mensajes de cada clúster, para así poder identificar los periodos con mayor actividad y dar soporte a las inquietudes de los remitentes.
- 6. Analizar los resultados obtenidos y realizar recomendaciones.** Realizar un análisis de los resultados obtenidos para encontrar las mejores estrategias que mejoren la gestión de recursos y así maximizar el impacto de las campañas de Manos Unidas. Posteriormente, formular recomendaciones basadas en los hallazgos del estudio para lograr un mayor desarrollo.

## CAPITULO III: METODOLOGÍA

Para la metodología de este estudio, se ha partido de la base de datos relacional proporcionada por Manos Unidas, esta información se ha conseguido obtener de primera mano gracias a la aprobación de un **convenio de colaboración entre Manos Unidas y la Universidad Pontificia de Comillas**, donde la ONG ha proporcionado los datos que han sido tratados de manera anonimizada durante todo el estudio.

Este Trabajo de Fin de Grado pretende **buscar oportunidades** donde, mediante un análisis exhaustivo de datos, poder conseguir que la organización sea más eficiente en algún ámbito. Estudiando las posibles propuestas, se decidió realizar una implementación de análisis de mensajes recibidos del buzón de la página web de la organización. Este enfoque de investigación puede proporcionar una comprensión más profunda sobre cómo los donantes y beneficiarios perciben a la organización y sus iniciativas de sensibilización y desarrollo.

Implementar una investigación completa y detallada de los mensajes recibidos por Manos Unidas puede generar un grande impacto en la **mejora de la comunicación**, ya que ayuda a la organización a comprender cómo sus mensajes son percibidos, lo cual permite a la ONG aprender para desarrollar una comunicación más adecuada y resonante. Además, el empleo de este análisis de los mensajes ayudará de manera directa a **optimizar los recursos** de la institución sin ánimo de lucro, para más adelante, identificar las posibles áreas que requieran más atención y también para una asignación de recursos administrativos más eficiente. Y, por último, este análisis contribuirá a **incrementar el impacto de las misiones** que está llevando a cabo. Esto se debe a que, basándose en un entendimiento más profundo de las necesidades y opiniones de sus audiencias, se pueden adaptar las campañas y los proyectos que se empleen para aumentar el impacto positivo de la organización en el desarrollo social.

Por lo tanto, para enfrentar el desafío constante de maximizar el impacto de Manos Unidas, al igual que muchas otras ONGs, en un mundo de cambio constante, ésta tiene que **implementar diferentes análisis** siguientes. Algunos de los ejemplos que se estudiarán más en detalle son el clustering de mensajes, donde poder agrupar toda la base de datos de mensajes en diferentes grupos, otro ejemplo es el análisis del sentimiento para cada clúster con el fin de ver cómo es la emoción reflejada en cada grupo de los mensajes.

Y el último ejemplo de estudio que se tratará es un análisis temporal donde poder observar patrones que se repitan a lo largo del año. En definitiva, todas estas pruebas permiten a la organización a comprender mejor a sus donantes y beneficiarios, además de poder realizar una mejor gestión de los recursos económicos y administrativos, y consiguiendo así obtener un mayor impacto en sus campañas. En otras palabras, estos análisis que se van a emplear ayudarán a que Manos Unidas prospere en su misión de luchar contra el hambre y la pobreza en el mundo.

### **3.1 Antecedentes**

En primer lugar, se procede a realizar una revisión crítica de los resultados más importantes que ya han sido publicados relacionados con el tema de este TFG. Es decir, serán mencionados los artículos o estudios públicos que mencionen el uso de cualquier tipo de análisis de datos complejo para una ONG o alguna institución pública. Además, se tendrá también en cuenta el resto de las oportunidades que tiene Manos Unidas para poder sacarle el mayor partido a este tipo de metodología basada en los datos y resultados, y así poder mejorar la eficiencia operativa y generar el mayor impacto posible en sus campañas.

Con respecto a la labor de divulgación de Manos Unidas, gracias a la globalización y la digitalización se puede acceder a un público mucho más amplio y gestionar sus recursos de manera óptima aun contactando con millones de personas alrededor del mundo. Uno de los principales causantes de la globalización que existe hoy en día son las redes sociales, las cuales son percibidas por las ONGs como una “herramienta idónea” (Arroyo Almaraz et al., 2013) para el desarrollo de una comunicación más participativa y directa, con el fin de acceder y contactar con la sociedad de una manera más personal.

De la misma manera que el marketing relacional y las redes sociales ayudan a las ONGs a contactar con más personas, así capturando más clientes y aumentando la recaudación, hay muchas otras herramientas que están surgiendo en los últimos años y que aún no se están implementando en este tipo de entidades, a pesar de tener un potencial muy prometedor. Con la aparición disruptiva de inteligencias artificiales, tales como el famoso ChatGPT (Diego Olite et al., 2023), se abren muchas posibilidades de poder mejorar procesos o aumentar la recaudación.

Si bien es cierto que las ONGs no están aprovechando todo el potencial que brinda la ciencia de datos y los enfoques basados en estadísticos, ya hay algunas instituciones sin ánimo de lucro que han empezado a implementar algunos análisis complejos. Uno de los ejemplos es UNICEF Perú, el cual ha utilizado la herramienta de ciencia de datos para realizar un análisis de una campaña, donde se extrajeron diferentes insights para examinar los resultados y evaluar la performance de la campaña televisiva (Olivera Taboada et al., 2020).

Otro ejemplo de implementación de proyectos de la mano de la ciencia de datos fue llevado a cabo por la Universidad Politécnica de Madrid (UPM). Se trata de una propuesta del diseño de una plataforma cognitiva que, mediante la utilización de técnicas y herramientas de Machine Learning y otros análisis de datos, gestiona de manera inteligente y eficiente la documentación disponible en el Centro de Innovación en Tecnología para el Desarrollo humano (idtUPM). Con la puesta en marcha de esta plataforma por medio de múltiples análisis de datos, se consiguió mejorar la eficiencia en la gestión de grandes volúmenes de información, ya que el idtUPM efectúa proyectos con distintos colaboradores como ONGs, corporaciones, instituciones públicas, etc (Río Ciriza, 2018).

En conclusión, para mejorar su impacto, es fundamental que las ONGs adopten un enfoque basado en datos numéricos y que éstas realicen análisis complejos de ciencia de datos en base a la información que reciben y generan. Estos análisis son capaces de proporcionar resultados y conclusiones muy valiosos para optimizar las campañas y los proyectos. Pero como Manos Unidas no empiece cuanto antes, pueden comenzar a recaudar menos para sus campañas, lo que puede provocar que la organización empiece cada vez a generar menos impacto en el desarrollo mientras las injusticias sociales sigan aumentando.

### **3.2. Extracción y limpieza de la muestra**

El primer paso de nuestro análisis consiste en la **extracción de los datos** a partir de la base de datos de Manos Unidas. En ella se ha desarrollado un script, o código, en Python para poder acceder a las tablas de datos, que estaban disponibles en la plataforma de gestión de la organización. El script emplea bastantes bibliotecas, donde la principal es la biblioteca requests, que se utiliza para realizar una solicitud HTTP al API de Manos

Unidas. Mediante una autenticación con credenciales de usuario proporcionadas por la organización, se puede iniciar una sesión para su posterior descarga de datos.

El código de Python está desarrollado para que los **datos sean descargados en series** de 10,000 observaciones o registros, con el fin de manipular de manera eficiente grandes volúmenes de información. Cada lote se va guardando en un archivo CSV, agregándose finalmente al archivo principal de forma iterativa, la cual en la primera descarga se llegaron a registrar entorno a cuarenta y dos mil registros. Sin embargo, estos no estaban completos, por lo que, empleando su correspondiente filtración a través de importación en un Excel a partir de otro CSV, nos quedamos con más de seis mil observaciones completas y aparentemente listas para ser analizadas.

Algunas de las variables más importantes de la base de datos de los mensajes son las siguientes:

- idMensaje: Es el identificador único para cada mensaje, esta variable es la central de la tabla de datos que está reflejada en otras tablas para poder así reflejar toda la información de esta tabla de mensajes.
- Fecha: refleja el día y la hora que se envió el mensaje al buzón de Manos Unidas. Es otro campo crucial en este estudio, sobre todo para la posterior serie temporal que se llevará a cabo.
- Asunto: es el título de la inquietud o petición que se explica en el mensaje.
- Texto: es el mensaje recibido por Manos Unidas, donde los remitentes piden o reclaman lo que necesiten a través de la solicitud o consulta, es decir, mediante el mensaje.

El segundo paso en este procesamiento de los datos es la **limpieza** de estos mismos. Una vez realizada la extracción a partir del servidor de Manos Unidas y de la importación de los datos completos a un Excel, se procedió a la limpieza de la muestra. Este paso es imprescindible para asegurar la calidad y la precisión, no solo del análisis posterior, sino sobre todo de los resultados obtenidos de la investigación.

Durante esta limpieza se efectuaron diferentes procesos, el primero y más sencillo fue rellenar las **observaciones vacías**, es decir, que no tenían información en la variable llamada 'texto'. Para esos registros se decidió poner en la variable 'texto' la misma información que tenían en la variable 'asunto', esto es debido a que todas esas consultas

habían escrito la petición en el espacio para escribir el asunto, en vez de donde correspondía.

Para el segundo proceso en esta limpieza de datos, y el más complejo, se descubrió que había muchos mensajes que resultaron ser **spam**. Un poco menos de la mitad de los mensajes completos recibidos resultaron no ser válidos, puesto que trataban diferente información irrelevante para la ONG de manos unidas. Algunos ejemplo de estos son mensajes con el contenido de ofertar diferentes cursos para mejorar el tráfico en tu página web, anuncios para invertir en criptomonedas y otros activos financieros, o incluso ofertas en páginas de compraventa de productos. Por lo tanto, con el fin de poder trabajar con tan solo los mensajes valiosos para este estudio, se realizó otro código de Python en el cual se **eliminan** todos los mensajes que no mencionen a Manos Unidas o que contengan algunas **palabras clave** (o también llamadas keywords) relacionadas con propaganda de cualquier tipo. Para la identificación de las palabras clave se estudió detalladamente muchos ejemplos de spam, o mensajes no valiosos, de la muestra original con el fin de hallar las palabras comunes de estos anuncios. Algunos ejemplos de keywords son “gratis”, “oferta”, “criptomoneda”, “investment”, “https”, y otras muchas más.

### **3.3. Modelo de Detección de Spam**

Una vez filtrados y depurados los datos, ya están listos para ser analizados en detalle y extraer información muy valiosa que sea aplicable para el estudio. No obstante, el código ejecutado anteriormente para eliminar el spam en base a las keywords declaradas es un proceso de **filtrado especialmente manual**, lo que significa que no solo es ineficiente, sino que también consume tiempo y recursos. Para mejorar esta herramienta de limpiado de datos, se debe realizar un modelo que automatice el proceso de identificación de los mensajes de spam. Por esta misma razón, después del último filtrado se va a implementar un **modelo predictivo de detección de spam** para que así se elimine automáticamente, y así ahorrar en tiempo y recursos humanos significativos. Este modelo también se desarrolló y se implementó mediante con el lenguaje de programación de Python debido a que tiene muchas bibliotecas muy completas y permite realizar análisis robustos.

Este modelo de Machine Learning también **mejora la gestión de los mensajes**, ya que, al filtrar automáticamente, el equipo puede centrarse en responder a los mensajes de relevancia y urgencia. Asimismo, al eliminar el spam de manera automática, los datos de

los mensajes son y van a seguir siendo **representativos y relevantes** durante y después del estudio, lo que ayudará a que los siguientes análisis de clustering, sentimiento y de serie temporal sean más precisos.

Para este análisis de detección de spam se emplearon diferentes pasos, el primero fue la **identificación de mensajes de spam**. Se disponía de dos conjuntos de datos, uno con los mensajes crudos, es decir, incluyendo el spam, y otra tabla filtrada sin estos mensajes no valiosos. Se compararon ambos conjuntos, creándose una nueva **variable binaria** de estudio la cual se etiqueta como spam (1) todos los mensajes que aparecen en el Excel de los mensajes crudos, pero no en el de los mensajes filtrados, y se etiqueta como no spam (0) si aparece en ambas tablas, ya que significa que pasó los procesos de filtrado del análisis anterior.

Una vez definida esta variable de estudio nueva, se procesó el texto asegurándose de tener todas las letras en minúsculas, eliminando los símbolos, números y signos de puntuación y demás. A continuación, se produjo la división en dos conjuntos, uno de **entrenamiento** (con el 80% de la población) y otro de **prueba** (con el otro 20% restante) para validar las predicciones posteriormente. Más adelante, se empleó el balanceo de clases a través del método de **SMOTE (Synthetic Minority Over-sampling Technique)** para un posterior entrenamiento del modelo mediante un clasificador de **Random Forest**, o también llamado conjunto de árboles de decisión. En este análisis, cada árbol se entrena con una muestra aleatoria de los datos, y su predicción final es el voto mayoritario de todos los árboles. Este tipo de métodos de análisis predictivo es muy útil puesto que es robusto frente al sobreajuste y maneja bien datos con alta dimensionalidad sin necesidad de realizar un PCA, el cual se pierde variabilidad y por tanto información. Por último, se evalúan los resultados del modelo mediante una **matriz de confusión** y los índices derivados de ésta.

### **3.4. Análisis de clustering**

Una vez extraídos, procesados y filtrados los datos quedándonos únicamente los mensajes que aportan valor a esta investigación, da comienzo al primero de los análisis complejos. Éste es el análisis de clustering, que desde una visión estratégica permite a las ONGs **segmentar sus datos recopilados**, y en este caso dividir los mensajes en grupos

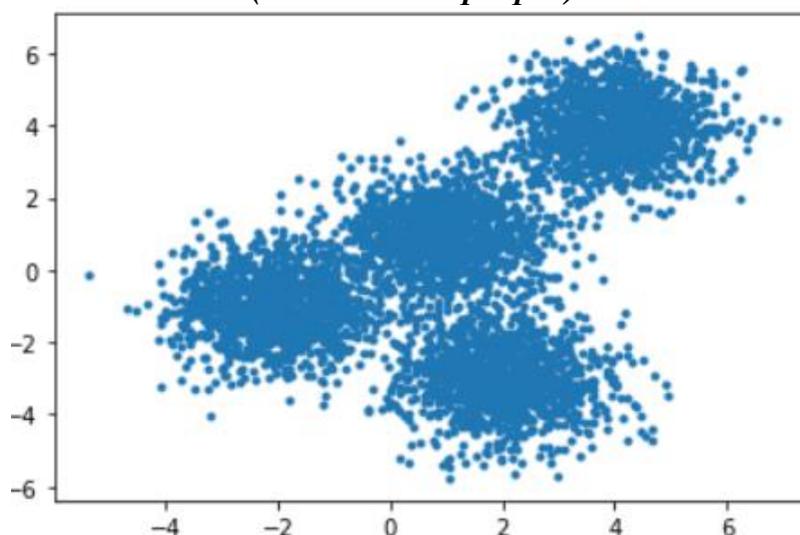
homogéneos, para facilitar la posterior personalización y mejora de sus métodos de comunicación y recaudación de fondos.

Este tipo de análisis tiene mucho potencial puesto que consigue **identificar los posibles patrones** en las donaciones, pero sobre todo ayuda a entender el comportamiento de los donantes, dividiendo la población en segmentos. Esta separación permite conocer más detalladamente las inquietudes y valores de cada grupo de clientes, para así diseñar campañas más efectivas. Un ejemplo de su utilización, Manos Unidas podría utilizar el método de clustering para segmentar a sus donantes según sus intereses y comportamientos, optimizando así sus esfuerzos de marketing. En cambio, como nuestra base de datos trata mensajes de interesados con la organización sin ánimo de lucro, el análisis de clustering tratará de identificar los diferentes tipos de peticiones que se han hecho en el buzón de la ONG.

La primera parte de este análisis, antes de aplicar el método de clustering, es la representación de los datos para saber cómo dividir la población. Esta representación se emplea utilizando el **Análisis de Componentes Principales (PCA)**, que básicamente reduce la dimensionalidad de los datos para facilitar su visualización en dos dimensiones. El PCA transforma los datos originales, que tenían más de diez variables, en tan solo dos. Estas dos son unas variables sintéticas representan a las variables originales mediante una combinación lineal de éstas. Donde representan la mayor variabilidad posible de todos los campos de la población.

Un ejemplo de representación de datos mediante PCA es el siguiente:

***Figura 1: Representación Muestra de ejemplo con PCA  
(Elaboración propia)***



Podemos observar en la ilustración 1 como una población está representada por dos variables sintéticas (o PCAs) en los ejes de abscisas y de ordenadas. En este ejemplo, nos ayuda tanto que no necesitaríamos realizar un análisis para averiguar en cuántos clusters se debería dividir esta población.

Esta transformación de las características originales en dos dimensiones principales permite la **visualización clara y simplificada** de los datos desde un plano bidimensional, el cual resulta mucho más entendible que una representación de quince variables.

A continuación, procedemos a la segunda parte del análisis de clustering, la cual consiste en **determinar del número (k) óptimo de clusters**. Para ejecutar el algoritmo del análisis de clustering es necesario especificar de manera previa el número de k clusters a formar. Además, muy pocas veces está muy claro en cuantos grupos se debe dividir una población o muestra al representarla mediante PCA. Por lo que para realizar una partición que nos proporcione los mejores resultados para su posterior interpretación, se necesita analizar de manera objetiva todas las posibles separaciones y elegir la mejor opción.

Para identificar el número óptimo de clusters, se pueden utilizar dos métodos:

1. Método del Codo: Este primer método consiste en calcular la suma de las distancias cuadradas de las observaciones dentro del cluster para los diferentes números de clusters que se decidan estudiar. Esta variable de suma de distancias cuadradas se denomina inercia o variación total intra-cluster Y al representar en un gráfico esta variable para cada número de k particiones, la inercia disminuye conforme más divisiones hay, ya que existen más centroides y las distancias son cada vez menores. Dentro del gráfico, el punto donde la disminución de la inercia comienza a desacelerarse notablemente se identifica como el "codo", y se considera como el número óptimo de k clusters. Sin embargo, la determinación de número de divisiones en base a este índice es muy susceptible de interpretación del gráfico, es por eso que muchas veces se acude al segundo método para confirmar el número óptimo de clusters.
2. Silhouette Score: El índice de Silhouette mide cuán similares son los datos dentro de un cluster en comparación con los de otros clusters. Se calculan valores de Silhouette para diferentes números de clusters, representados también en un gráfico, y se elige el número de partición que maximice el índice. Este índice no

da margen a la interpretación de los resultados, por lo que este segundo cálculo ayuda a despejar a aclarar la elección de número de centroides.

Para la partición de la población en k clusters se emplea el **algoritmo K-Means**, este algoritmo asigna cada observación a uno de los k clusters con la restricción de que los clusters no se solapan. Este método busca minimizar la dispersión dentro de cada cluster, buscando formar grupos lo más homogéneos posible. O, en otras palabras, particiona los datos en K clusters, donde cada punto pertenece al cluster con el centroide más cercano.

El algoritmo realiza numerosas iteraciones, hasta un máximo de 100, donde primero **asigna de forma aleatoria un cluster a cada observación**. Después, para cada uno de los clusters calcula el centroide, para posteriormente asignar cada observación al cluster cuyo centroide se encuentre más cerca. Donde a continuación, vuelve a calcular el nuevo centroide para cada cluster y vuelve a asignar cada observación al cluster con el centroide más cercano, y así sucesivamente hasta que alcance su tamaño máximo de iteraciones o hasta que la asignación de las observaciones a los clusters no cambie. Dado que los resultados obtenidos dependerán de la asignación aleatoria inicial de cada observación a cada cluster, el algoritmo se ejecutará varias veces utilizando diferentes asignaciones iniciales aleatorias, en el estudio hemos elegido que se correrá 10 veces.

A la hora de hacer el **cálculo de distancias** del algoritmo de K-means, se utiliza la distancia euclídea como medida de similitud. La cual es importante que se estandaricen las variables ya que puede haber una variable tenga valores mucho más grandes que otra variable.

### **3.5. Análisis del sentimiento de los mensajes**

Primeramente, el análisis de sentimiento es un método de procesamiento de lenguaje natural (Natural Language Processing, NLP) que se emplea para averiguar cuál es el **tono emocional de un texto**. En el contexto de este estudio dirigido para Manos Unidas, este análisis se ha llevado a cabo con el propósito de evaluar los mensajes enviados al buzón de la organización. El objetivo principal de este análisis es proporcionar una **visión más cualitativa** al estudio sobre el estado emocional y las preocupaciones de los remitentes a través de los mensajes. Por lo que, combinando el análisis cuantitativo de clustering, esta evaluación cualitativa de los sentimientos proporciona a la investigación una visión más completa y rica de los resultados obtenidos, además de enriquecer el primer análisis de clustering.

Para cada cluster identificado, se realizó un análisis de sentimientos con el fin de **evaluar la polaridad de los mensajes** (positiva, negativa o neutral). Lo que significa que este modelo se ha ejecutado **una vez para cada cluster**, para así poder ver el significado y la emoción de cada tipo de mensajes, y así comprobar que cada cluster está expresando lo que puedes intuir según las palabras más frecuentes de cada cluster. Este enfoque ayuda a la verificación de los clusters, es decir, que los resultados de este análisis de sentimientos ayudan a validar y refinar los clusters identificados, donde provoca que se asegure que las particiones realizadas por el análisis anterior sean coherentes desde una perspectiva emocional.

El estudio de los sentimientos proporciona a la investigación una comprensión profunda de las emociones y preocupaciones de los mensajes enviados. Esta comprensión puede reportar **enormes beneficios** para Manos Unidas, como por ejemplo la **mejora de la comunicación**. Esta brinda la posibilidad de ajustar el tono y contenido de las respuestas a una manera más empática para facilitar la personalización de los mensajes de seguimiento, agradecimiento o información con el fin de mejorar la relación con los donantes y beneficiarios. Otro beneficio para la institución es la **identificación de problemas y oportunidades**, ya que, gracias a este buzón, ayuda a detectar tempranamente cualquier problema o preocupación recurrente a los involucrados con la ONG. Y con una identificación rápida de los problemas, surgen oportunidades para mejorar los servicios proporcionados. Y el último ejemplo, también puede funcionar como un proveedor de **evaluación continua**, donde poder monitorear y examinar el estado emocional de los donantes con respecto a las campañas y los procedimientos de Manos Unidas.

El proceso del análisis de sentimientos fue desarrollado en un script de Python, el cual se llevaron a cabo 6 pasos diferentes y bien marcados:

1. Carga y preparación de los datos. Los datos fueron importados desde un Excel con los resultados del análisis de clustering anterior para cada observación. Y para este segundo análisis, las únicas columnas que fueron seleccionadas para el estudio fueron la del texto y la del cluster al que pertenecen.
2. División de los datos en conjuntos. La tabla de los datos fue dividida en dos grupos, un conjunto de entrenamiento con el 80% del total, y el 20% restante dedicado adjudicado al conjunto de validación.

3. Tokenización. Para convertir el texto en tokens manejables se empleó el tokenizador BERT. BERT (que sus siglas significan Bidirectional Encoder Representations from Transformers) es un modelo de lenguaje desarrollado por Google que se emplea en este tipo de análisis muy revolucionario debido a la capacidad que tiene para comprender el contexto bidireccional del texto. En otras palabras, el modelo de BERT tiene la facultad de interpretar el contexto a la izquierda y a la derecha de la palabra que está siendo analizada.
4. Entrenamiento del Modelo. Se entrena el modelo y se configuran los argumentos del entrenamiento, con el fin de realizar, posteriormente, un análisis de los resultados.
5. Análisis de Polaridad. Gracias al modelo entrenado, se calcula la polaridad, determinándose las puntuaciones del sentimiento (en este caso es la polaridad de las palabras) para cada mensaje de los clusters. El resultado de este análisis es una clasificación de cada uno de los mensajes como positivo, negativo o neutro.
6. Evaluación de precisión del Modelo. Y, por último, se han de evaluar los resultados del modelo, para ello se comprueba la eficacia del análisis generado y certificar la precisión de las clasificaciones.

### **3.6. Análisis de series temporales**

Una serie temporal es una sucesión de observaciones de una variable tomadas en varios instantes de tiempo. Y el análisis de series temporales es una técnica estadística que se utiliza para analizar secuencias de datos recolectados a lo largo del tiempo. Este tipo de análisis es fundamental para identificar patrones, tendencias y estacionalidades en los datos, así como para realizar predicciones futuras. En este estudio, se aplicó un análisis de series temporales a los mensajes recibidos por Manos Unidas para cada uno de los clusters identificados previamente.

El análisis de series temporales tiene la capacidad de proporcionar múltiples beneficios a Manos Unidas, uno de ellos es la posibilidad de **identificar patrones y tendencias** que puedan repetirse a lo largo del año solar. Lo cual se enlace con el segundo beneficio directo que es la mejora de la gestión de recursos. Esto se debe a que, al descubrir tendencias, se puede **predecir la carga de trabajo** futura donde a lo mejor mandan

muchas consultas en una época del año, por lo que para futuras ocasiones se puede ajustar la asignación de recursos mediante una predicción de las necesidades futuras de nuestros donantes y beneficiarios. Además, es posible emplear una optimización de las campañas, puesto que se pueden reconocer las épocas de mayor actividad tanto de bajas como de altas de donantes. Lo cual significa que se pueden llevar a cabo campañas para tanto captar nuevos donantes como para retener los existentes. Y el último beneficio identificado puede ser una **reducción de costes** operativos, si evitamos la sobreasignación de los recursos disponibles en los periodos de baja actividad. Mientras que, por otro lado, en los momentos del año donde se experimenta una alta actividad, al predecir estos picos de recepción de mensajes, se puede responder de manera rápida y efectiva.

Generalmente, las series temporales se descomponen en 3 componentes principales:

1. Tendencia. Refleja la dirección general a la que avanza la serie temporal a lo largo del tiempo. Ésta puede ser ascendente, descendente o constante en el tiempo.
2. Estacionalidad: representa los posibles patrones que se repiten en intervalos regulares en el tiempo, ya sea días, meses o incluso años. Estos patrones son de extrema importancia para identificar variaciones periódicas y así planificar de manera adelantada la temporada alta de mensajes.
3. Ruido. Es la parte residual de la serie temporal, que representa variaciones aleatorias e impredecibles en los datos debido a que no se puede explicar mediante la tendencia o la estacionalidad.

Para este análisis de series temporales emplearemos el modelo de ARIMA (AutoRegressive Integrated Moving Average), el cual es ampliamente utilizado para el modelaje y la predicción de series temporales puesto que tiene una flexibilidad y capacidad para interpretar diferentes tipos de patrones dentro de los datos. El modelo ARIMA se descompone en las siguientes 3 partes.

- AutoRegresivo (AR): esta componente analiza las dependencias entre una observación y un número específico de observaciones anteriores. Es representada con el parámetro  $p$ , indicando el número de retardos del modelo.
- Integrado (I): diferencia las observaciones de la serie temporal para la estacionaria, o, en otras palabras, elimina la tendencia y las estacionalidades a largo plazo. Se representa con el parámetro  $d$  indicando el número de veces que se han de diferenciar los datos.

- Media Móvil (MA): estudia la dependencia entre una observación y un error residual de un número determinado de observaciones anteriores. Se representa con el parámetro  $q$ .

El modelo ARIMA se denota como ARIMA( $p,d,q$ ), donde  $p$  representa el orden del componente auto regresivo,  $d$  muestra el número de diferenciaciones necesarias para hacer la serie estacionaria, y  $q$  el orden del componente de media móvil.

La implementación de la serie temporal se lleva a cabo mediante los siguientes pasos:

1. Identificar los patrones temporales. Se identifican las tendencias generales, las estacionalidades y cualquier ciclo que pueda influir en la variabilidad de toda la serie temporal.
2. Componentes. Se descompone la serie temporal en sus componentes principales, el cual permite una mejor comprensión de los factores que provocan las variaciones en los datos.
3. Predecir comportamientos futuros. Mediante el modelo de ARIMA mencionado anteriormente, se trata de realizar predicciones sobre el volumen de los futuros mensajes. Esta predicción es fundamental se lleva a cabo validando con el conjunto de validación, las predicciones estudiadas mediante con el conjunto de entrenamiento, lo cual ayuda a la preparación previa de futuras demandas.

La combinación de estos elementos permite una **comprensión profunda** de la dinámica temporal de los mensajes recibidos por Manos Unidas y facilita la predicción de futuros comportamientos. Donde se desarrollará y se implementará el código para este análisis de series temporales para cada cluster mediante la herramienta Rstudio.

En conclusión, toda la metodología descrita en este capítulo IV combina técnicas avanzadas de análisis de datos para proporcionar una visión integral y profunda de los mensajes recibidos por Manos Unidas a lo largo del tiempo. Cada paso del proceso, desde la extracción y limpieza de datos hasta los análisis de clustering, sentimientos y series temporales, está diseñado para optimizar la gestión y comunicación de la organización, permitiendo una respuesta más efectiva y eficiente a las necesidades de sus beneficiarios y donantes.

## CAPITULO IV: ANÁLISIS DE RESULTADOS OBTENIDOS

A continuación, vamos a proceder con la evaluación de los resultados que hemos obtenido con todos los diferentes análisis desarrollados e implementados que han sido explicados en la metodología.

### 4.1. Resultados Modelo Detección de Spam.

Como ya hemos explicado anteriormente este análisis realizado tiene como propósito la automatización de la detección de los mensajes de spam que hay en la base de datos obtenida por Manos Unidas.

En cuanto a los resultados del modelo en el conjunto de prueba son los siguientes, expresados en una matriz de confusión:

**Figura 2: Matriz de Confusión  
(Elaboración propia)**

	Predicted No Spam	Predicted Spam
Actual No Spam	578	16
Actual Spam	58	625

Esto significa que, de los 594 mensajes considerados como no spam, fueron correctamente clasificados unos 578 (verdaderos positivos) dejando tan solo 16 observaciones incorrectamente clasificadas como spam, o falsos positivos. Por el contrario, de los 683 mensajes clasificados como spam, 625 (verdaderos negativos) de ellos fueron correctamente identificados y tan solo 58 fueron falsos negativos.

A partir de esta matriz de confusión se han calculado los siguientes índices para comprobar el éxito del análisis empleado.

- Exactitud (Accuracy): 0.942

- Precision: 0.975
- Sensibilidad (Recall): 0.915
- Especificidad (Specificity): 0.973
- F1 Score: 0.944

El modelo manifiesta una precisión de identificación correcta de los mensajes de spam muy elevada (del 97,5%), lo cual significa que la mayoría de los mensajes clasificados como spam son efectivamente spam.

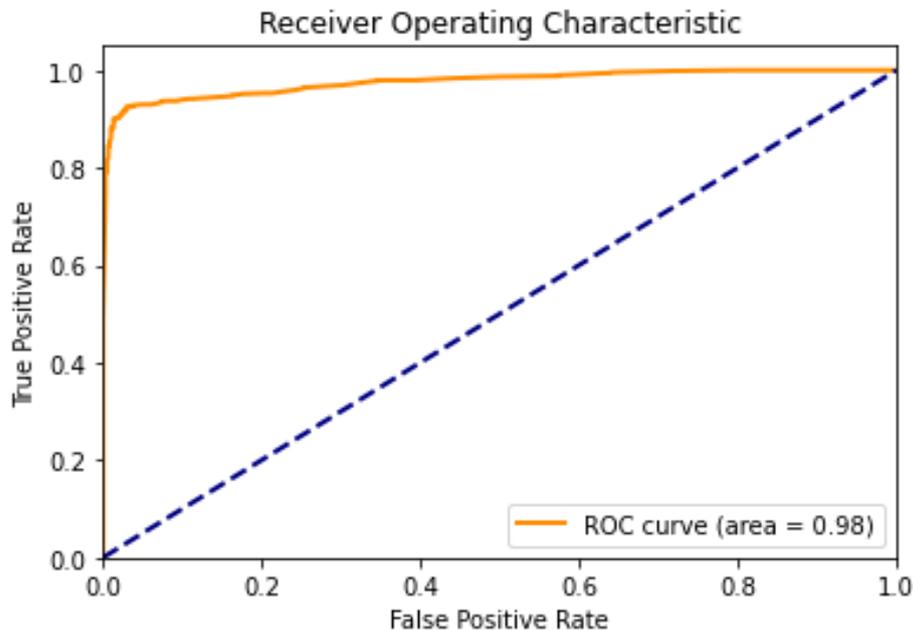
En cuanto a la sensibilidad del modelo, también es alta, con un valor del 92% demuestra que el modelo es capaz de identificar la mayoría de los mensajes de spam.

Para el F1 Score, este índice representa el equilibrio entre la precisión y la sensibilidad halladas en el modelo, y con un F1 Score del 0,94 está muy bien balanceado.

Los resultados del modelo también muestran una precisión global alta con un 94%, lo que confirma que el modelo es confiable y puede ser utilizado para automatizar esta detección de spam ya que tiene una muy alta tasa de éxito.

Y en última estancia, la alta especificidad, con un 0,973, demuestra que los mensajes con valor o no spam, son correctamente identificados, minimizando así los falsos positivos, que en este estudio pueden hacer mucho daño debido a que de estos mensajes válidos sacamos la información de qué y cuantos grupos tenemos que clasificar la población y su sentimiento. Por lo que en este estudio tienen mucho más impacto los falsos positivos frente a los falsos negativos (debido a que los falsos positivos serían mensajes con consultas e inquietudes incorrectamente eliminadas).

**Figura 3: Curva ROC  
(Elaboración propia)**

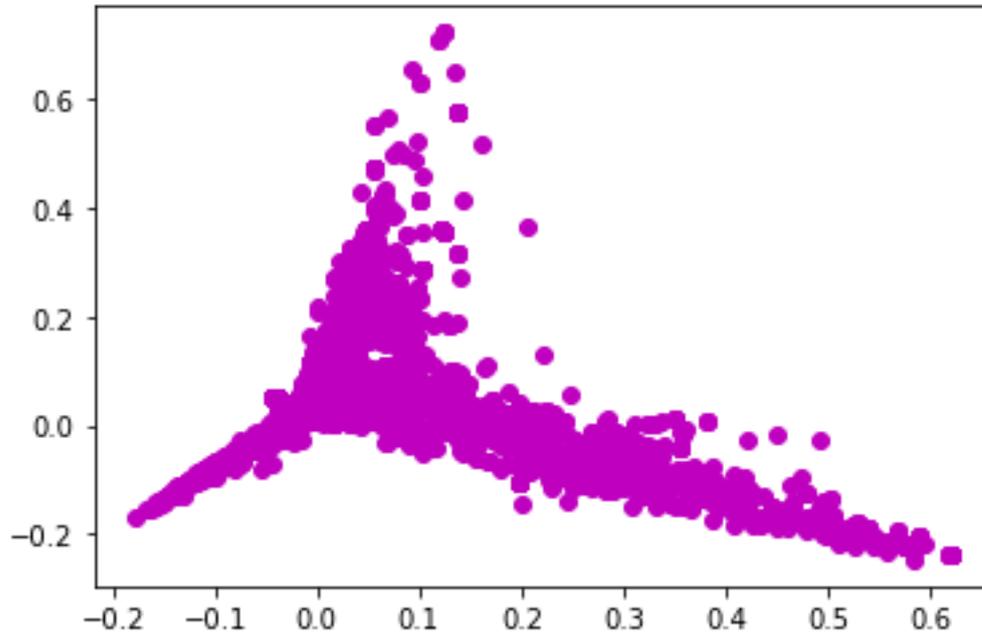


La figura 3, es una representación de la curva ROC obtenida en base a los resultados del modelo, donde muestra el área debajo de la curva (AUC) de 0,98, lo que significa que el modelo de distinción entre mensajes de spam y no spam presenta un rendimiento excelente por no decir casi inigualable, puesto que la perfección en estos modelos predictivos es prácticamente imposible e incluso a veces muestra de algún error el tener un AUC de 1.

#### **4.2. Resultados Análisis Clustering.**

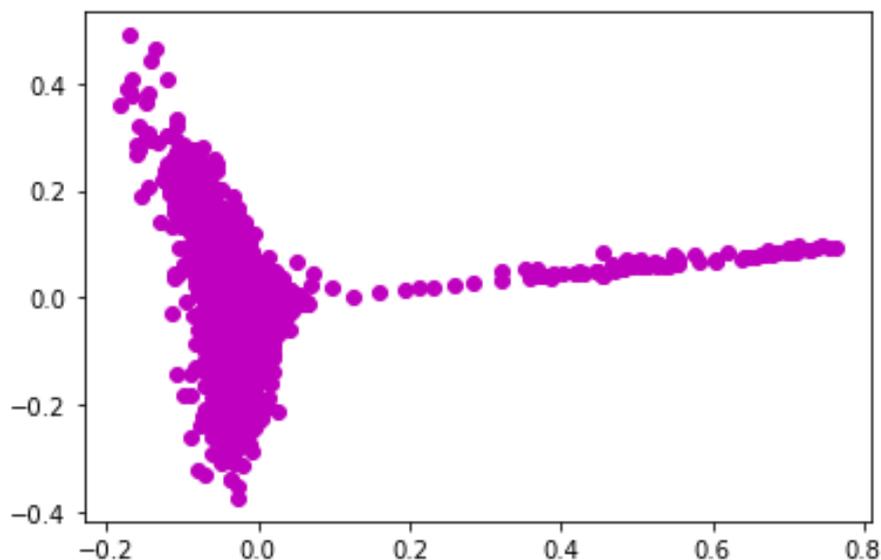
Al realizar el primer análisis de PCA reduciendo a dos dimensiones para representar la población antes de averiguar que había muchos mensajes de spam que no contenían información relevante, sino todo lo contrario.

**Figura 4: Representación PCA de datos NO filtrados  
(Elaboración propia)**



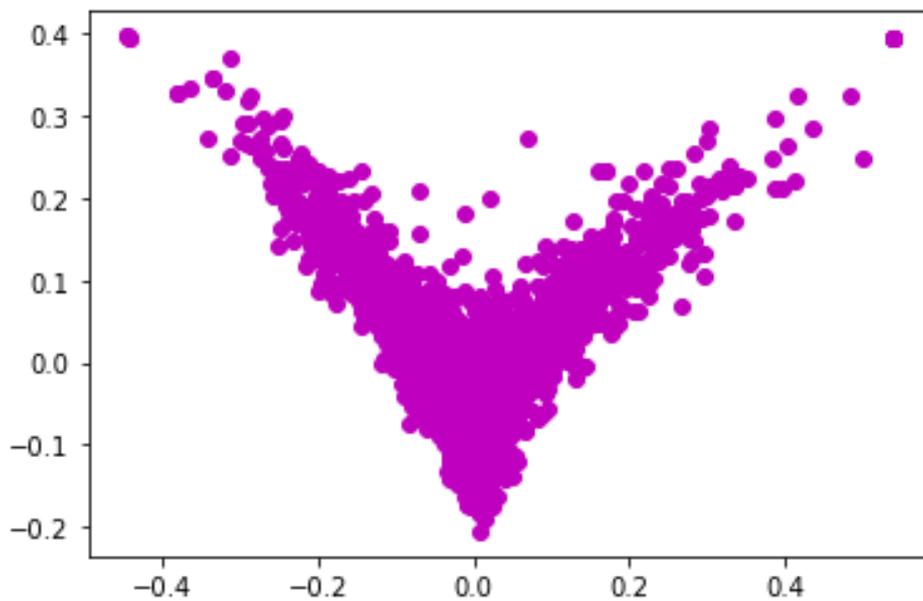
Antes de realizar la limpieza de spam, los datos representados mediante PCA en la Figura 4 mostraban una gran dispersión en algunas zonas mientras que también había concentraciones en otros puntos, lo que indicaba la presencia de ruido y posibles mensajes irrelevantes en la base de datos. Esta representación inicial revelaba que los datos contenían una mezcla significativa de información útil y no útil, siendo complicando el análisis efectivo para una posterior partición en clusters.

**Figura 5: Representación PCA de datos filtrados manualmente  
(Elaboración propia)**



Tras realizar una limpieza inicial de manera manual para eliminar mensajes de spam, la representación de la población de estudio mediante PCA (dos dimensiones sintéticas que expresan la mayor varianza posible de la población) representada en la Figura 5, mostró una reducción en la dispersión y una concentración más definida de los datos. Sin embargo, aún se podían observar agrupaciones densas que sugerían la necesidad de una limpieza más rigurosa para mejorar la calidad de los datos.

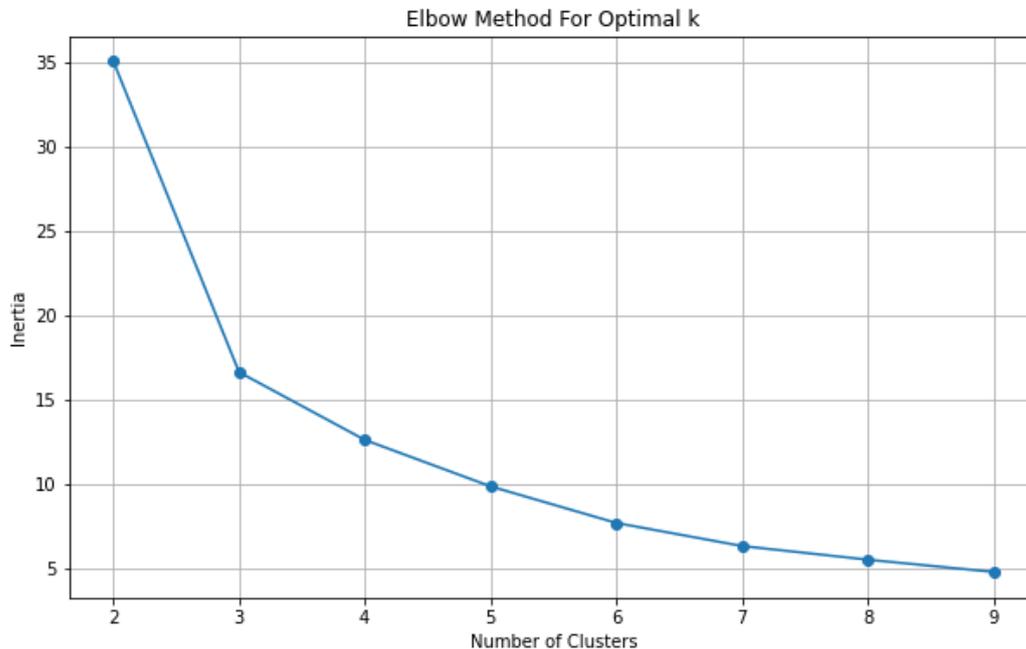
***Figura 6: Representación PCA de datos filtrados  
(Elaboración propia)***



Finalmente, tras aplicar una limpieza más exhaustiva mediante el código de Python mencionado anteriormente, se obtuvo la representación en la Imagen 3. Aquí, los datos aparecen más organizados y menos dispersos, indicando una mejora significativa en la calidad y la relevancia de los datos para el análisis posterior. Esta limpieza fue crucial para asegurar que los datos utilizados en el análisis de clustering fueran representativos y útiles.

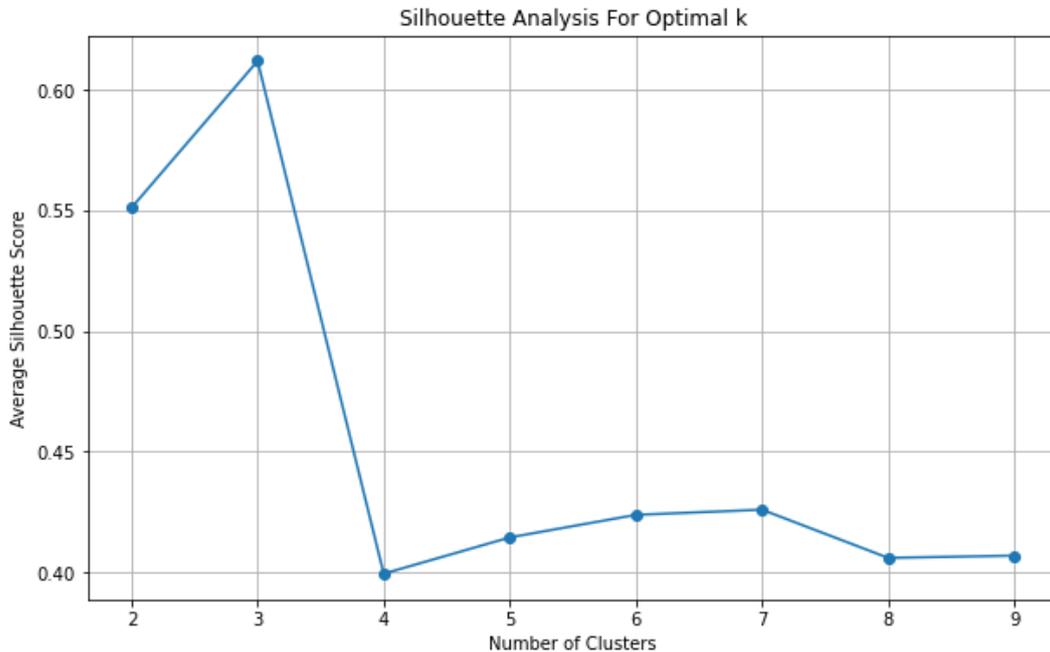
## Determinar el Número de Clusters:

*Figura 7: Gráfico método del Codo para k óptimo  
(Elaboración propia)*



En la Figura 7, gráfico del método del codo muestra, como ya se ha mencionado anteriormente, en el eje Y, la suma de las distancias cuadradas dentro del cluster (inercia) o, en otras palabras, muestra la **variación total intra-cluster**. Y en el eje X muestra los diferentes números de clusters en los que se está barajando dividir la población. El criterio para elegir el número óptimo de k clusters es el número donde al añadir otro cluster más, no mejora mucho la variación total intra-cluster. En el gráfico, se observa ese "codo" claro en el punto **k=3**, ya que para k=4 no disminuye tanto la variación. Lo que sugiere que una partición en tres clusters puede ser una buena elección para segmentar los datos. Este método ayuda a determinar el número óptimo de clusters minimizando la variabilidad dentro de cada cluster y asegurando que los clusters sean distintos entre sí. Sin embargo, es mejor comprobar los resultados de este gráfico se alinean con los del gráfico del método del Average Silhouette.

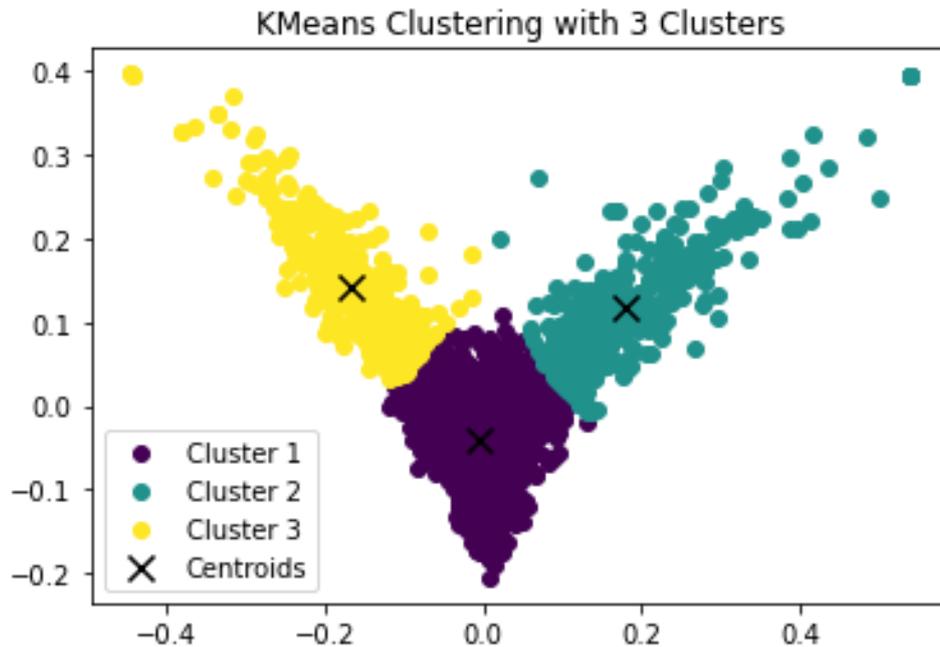
**Figura 8: Análisis de Silueta para  $k$  óptimo  
(Elaboración propia)**



En esta Figura 8, está representado el análisis de silueta, que mide la calidad de la agrupación para diferentes números de clusters. Donde para el eje de ordenadas se muestra la puntuación de la media de los Silhouette coeficients, o índices de silueta. El índice de silueta cuantifica cómo de buena es la asignación que se ha hecho de una observación a un cluster, por lo que **cuanto más alta sea la media de estos coeficients, mejor es la partición  $k$ .**

El valor más alto de índice de silueta se puede observar que se registra para  **$k=3$** , lo cual respalda la elección de tres clusters que habíamos estudiado en el gráfico anterior, indicando que esta partición maximiza la cohesión interna y la separación entre clusters. Por lo tanto, ejecutaremos un fraccionamiento de la población en **3 clusters diferentes**. Este análisis es esencial para la validación de la elección del número óptimos de clusters, y así garantizar que la segmentación sea significativa y efectiva para que las observaciones de los clusters muestren unas características parecidas entre sí, pero diferentes a las del resto de clusters.

**Figura 9: Partición Clustering Analysis ( $k=3$ )  
(Elaboración propia)**

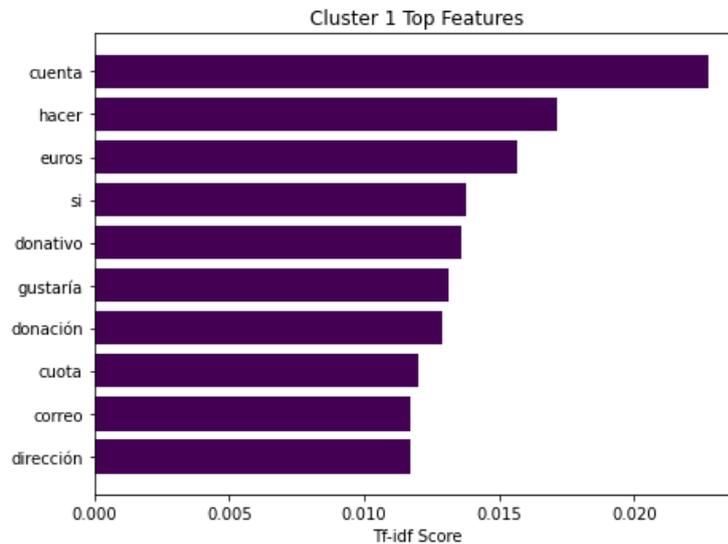


Al realizar una partición de la población en tres clusters con  $k=3$ , se obtuvo la representación mostrada en la Figura 9. Cada punto representa un mensaje (en dos dimensiones PCA), y los diferentes colores indican los 3 clusters diferentes, con sus centroides marcados con una X. Esta representación visual muestra claramente cómo se agrupan los datos en tres segmentos distintos, y cada uno tiene una probabilidad alta de presentar características únicas.

A continuación, se ha realizado un plot muestra las palabras más frecuentes en cada cluster, representadas con sus respectivos puntajes de Tf-idf. Con este análisis de términos se proporciona una visión clara de los temas predominantes en cada cluster a través de las palabras más utilizadas.

- **Cluster 1:**

**Figura 10: Palabras más frecuentes del CLUSTER 1  
(Elaboración propia)**



**Palabras Frecuentes:** "cuenta", "hacer", "euros", "sí", "donativo", "gustaría", "donación", "cuota", "correo", "dirección".

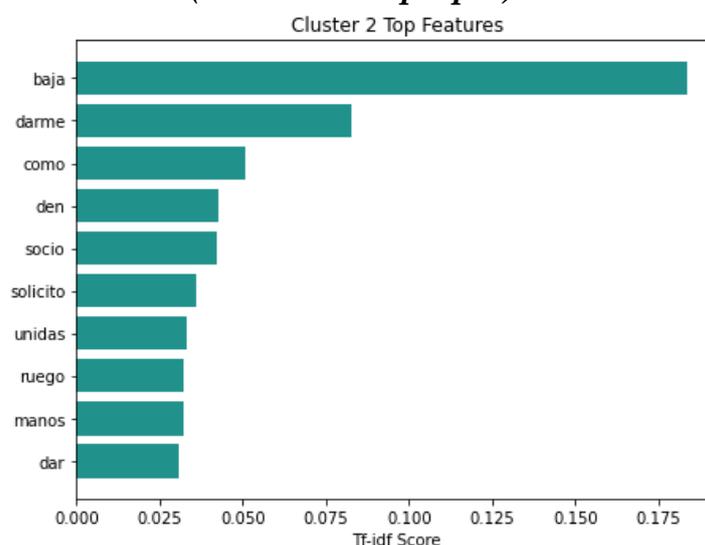
**Tipo de Consulta:** Según las palabras más utilizadas en los mensajes, este cluster se centra en cuestiones muy diversas, relacionadas con donaciones y transacciones financieras, ya sea por ejemplo abrir cuentas, realizar donativos y cuotas. Los mensajes en este cluster reflejan preocupaciones y consultas sobre cómo gestionar las cuentas de donación, o también solicitud de cambio de los datos registrados, ya sea que requieren cambiar su cuenta bancaria, dirección o correo electrónico. Al estar presente la palabra “gustaría” da a entender fácilmente que todo son solicitudes.

Estos mensajes del cluster 1 indican una fuerte preocupación por la realización de donativos y, sobre todo, la gestión de cuentas. Asimismo, el primer grupo averiguado también representa los mensajes con propuestas sobre nuevos proyectos o búsqueda de trabajo. Esto se ve reflejado en las palabras “hacer” y “gustaría”.

**Recomendación.** Para que Manos Unidas mejore su portal de donaciones, la organización puede ofrecer una guía clara y detallada que explique el procedimiento de realización de donativos, gestión de cuentas y de su información personal y comprensión de las cuotas. Además, proporcionar una respuesta asistida rápida y eficiente especialmente para las cuestiones financieras es probable que aumente considerablemente la satisfacción y el compromiso de los donantes.

- **Cluster 2:**

**Figura 11: Palabras más frecuentes del CLUSTER 2  
(Elaboración propia)**



**Palabras Frecuentes:** "baja", "darme", "cómo", "den", "socio", "solicito", "unidas", "ruego", "manos", "dar".

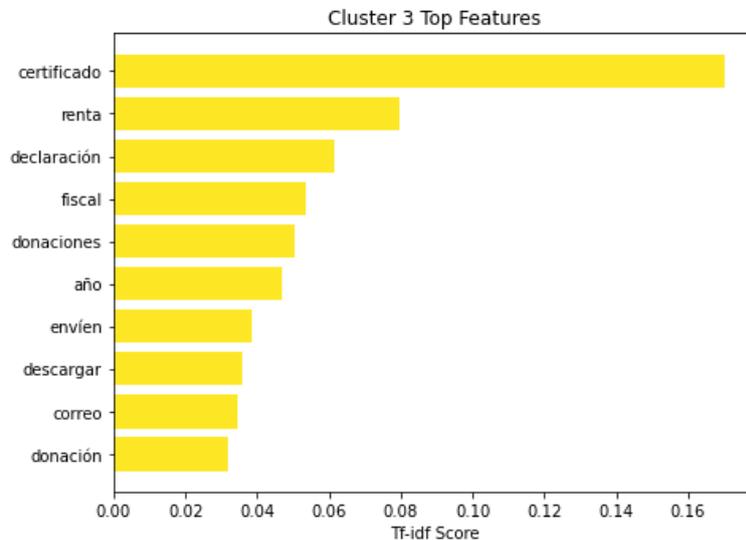
**Tipo de consulta:** Este cluster está claramente diferenciado del resto ya que agrupa mensajes de remitentes que desean darse de baja o retirar suscripciones, así como en peticiones para ser eliminados de listas de socios. Las palabras clave sugieren que estos usuarios buscan información o están realizando una petición para dejar de ser socios y desvincularse de la organización. Incluso se puede concatenar algunas de las palabras más frecuentes para formar una frase de petición para “darse de baja como socio”.

Además, leyendo más en detalle los mensajes enviados con el tag de este segundo cluster, el texto de las observaciones muestra una frecuencia muy elevada de solicitudes para darse de baja. Lo que significa que este cluster debe ser cuanto más pequeño mejor para que no afecte a los recursos económicos disponibles de la institución sin ánimo de lucro.

**Recomendación:** Es esencial que Manos Unidas facilite un proceso sencillo y transparente para que los usuarios puedan darse de baja o modificar su estatus de socio. Implementar un sistema automatizado para gestionar estas solicitudes puede mejorar la experiencia del usuario y reducir la carga administrativa. Además de esto, es importante analizar más en detalle, los mensajes de este cluster ya que muchos de ellos han expuesto las causas de su decisión de darse de baja, lo cual puede proporcionar un feedback valioso y significativo para Manos Unidas con respecto a la gestión y trato con sus socios.

- **Cluster 3:**

**Figura 12: Palabras más frecuentes del CLUSTER 3  
(Elaboración propia)**



**Palabras Frecuentes:** "certificado", "renta", "declaración", "fiscal", "donaciones", "año", "envíen", "descargar", "correo", "donación".

**Tipo de consulta:** En este último cluster se hayan las solicitudes relacionadas con certificados y documentación fiscal, con el fin de obtener información sobre certificados de donaciones para realizar una posterior de declaración de la renta. Los mensajes en este cluster indican una necesidad de asistencia administrativa y documental relacionada con las donaciones realizadas.

Los mensajes en este cluster no indican una petición con noticias negativas para la organización como en el cluster 2, ni una solicitud más urgente como cambio de cuenta bancaria como en el cluster 1. Sino que indican una necesidad clara de asistencia administrativa con la documentación fiscal y certificados de donación.

**Recomendación:** Manos Unidas debería crear una sección en su sitio web dedicada de manera exclusiva a la obtención y descarga de certificados y otra documentación fiscal, además de incluir tutoriales y FAQs (Frequently Asked Questions) para guiar a los usuarios en este proceso. Asimismo, ofrecer soporte dedicado para estas consultas puede mejorar la eficiencia y satisfacción del usuario mediante una derivación a un equipo o departamento fiscal específico para que puedan acceder de manera casi inmediata a la información financiera sobre las solicitudes de los certificados de donaciones.

### 4.3. Resultados Sentiment Analysis

En las próximas 3 figuras (Figura 13, 14 y 15) se van a mostrar todos los mensajes de los clusters 1, 2 y 3, respectivamente. En estos gráficos se podrá observar que en el eje Y está representado el conteo del número de mensajes, y el eje X muestra los 3 posibles resultados del análisis de sentimiento para cada mensaje. Donde cada número significa lo siguiente:

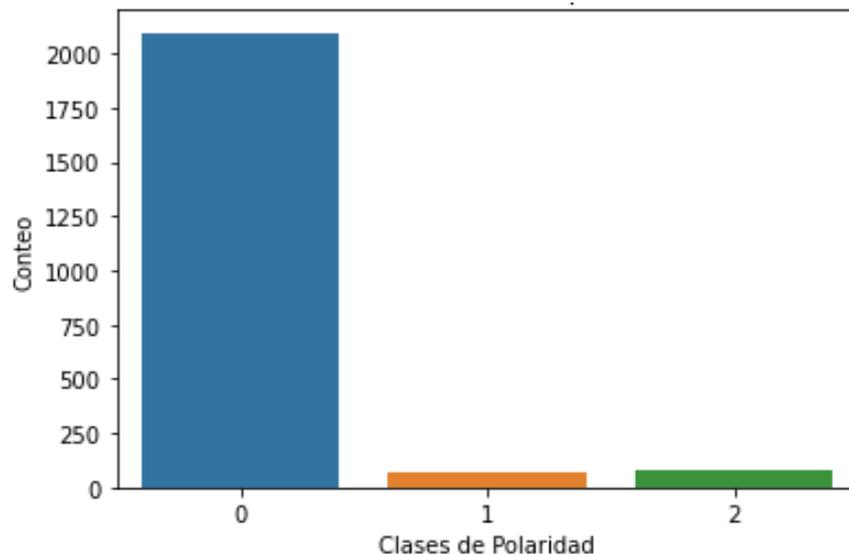
- 0: Mensaje Neutro
- 1: Mensaje negativo
- 2: Mensaje positivo

A continuación, se muestran los resultados de la predicción del sentimiento de cada cluster y su predicción de éxito, y después un análisis de éstos enfocándolo al estudio.

#### Resultados CLUSTER 1

- Precision: 0.95436
- Recall: 0.93738
- F1 Score: 0.94580

*Figura 13: Polaridad Mensajes del Cluster 1  
(Elaboración propia)*



La gran mayoría de los mensajes de este cluster son clasificados como **neutros**. Esto logra sentido al saber previamente que las interacciones relacionadas con donaciones y transacciones financieras son principalmente consultas que no evocan ninguna emoción

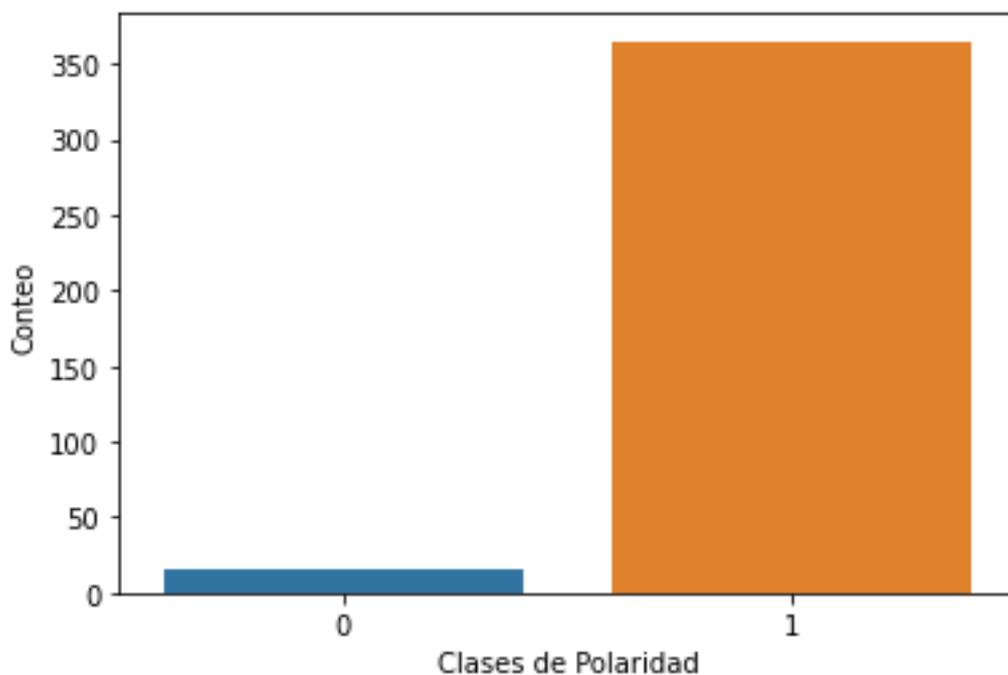
fuerte que pueda resultar positiva o negativa. Los mensajes positivos muestran un agradecimiento por alguna labor o algún proyecto en concreto, mientras que los escasos valores negativos son mensajes en protesta por los episodios recientes de conflicto geopolítico entre Palestina e Israel. Mientras que otros mensajes de este cluster buscan hacer propuestas para nuevos proyectos o campañas, además de búsqueda de prácticas por parte de algunos remitentes.

La precisión, la sensibilidad (Recall) y el F1 Score muestran que los resultados son fiables ya que su porcentaje de éxito y exactitud es muy alto.

### Resultados CLUSTER 2

- Precision: 0.94829
- Recall: 0.95800
- F1 Score: 0.95312

**Figura 14: Polaridad Mensajes del Cluster 2**  
(Elaboración propia)



En el cluster 2 podemos observar que los mensajes con polaridad neutra tienen muy baja representación, los mensajes positivos no tienen ninguna, mientras que los mensajes **negativos** predominan significativamente. Este cluster está clasificado mayoritariamente con una polaridad negativa, lo cual es consistente con la naturaleza de su contenido debido a que sus remitentes expresan la intención y deseo de darse de baja de sus suscripciones

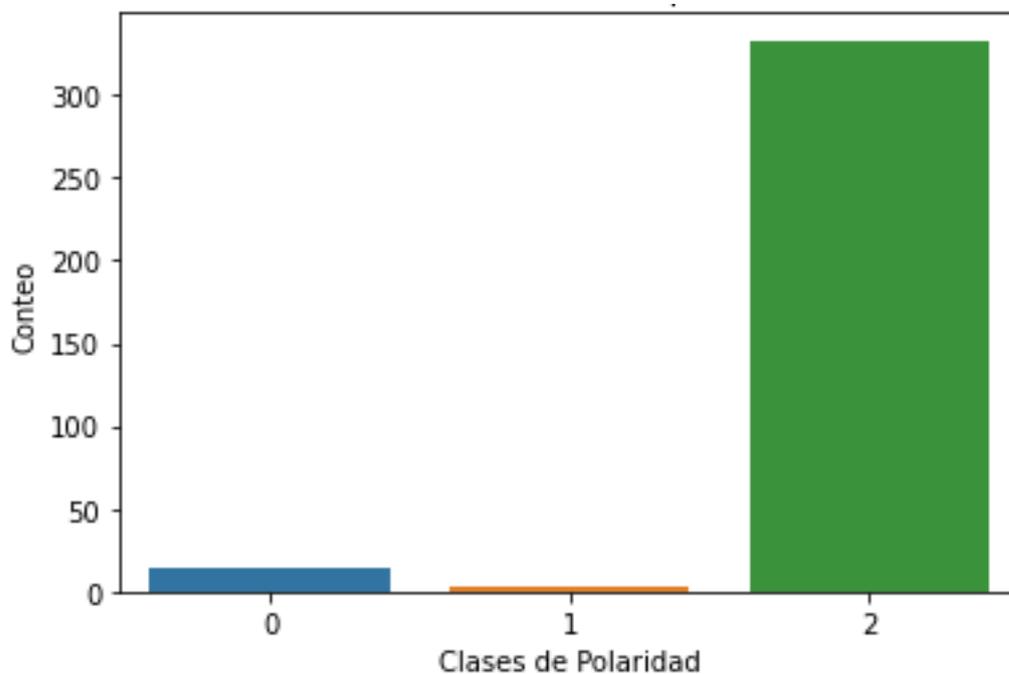
como socios. Esto demuestra que los mensajes están mostrando posibles quejas y lamentaciones por tener que darse de baja.

El alto rendimiento del modelo y sus calificaciones demuestra que las clasificaciones de la polaridad en este cluster también son confiables.

### Resultados CLUSTER 3

- Precision: 0.93881
- Recall: 0.95142
- F1 Score: 0.94507

*Figura 15: Polaridad Mensajes del Cluster 3  
(Elaboración propia)*



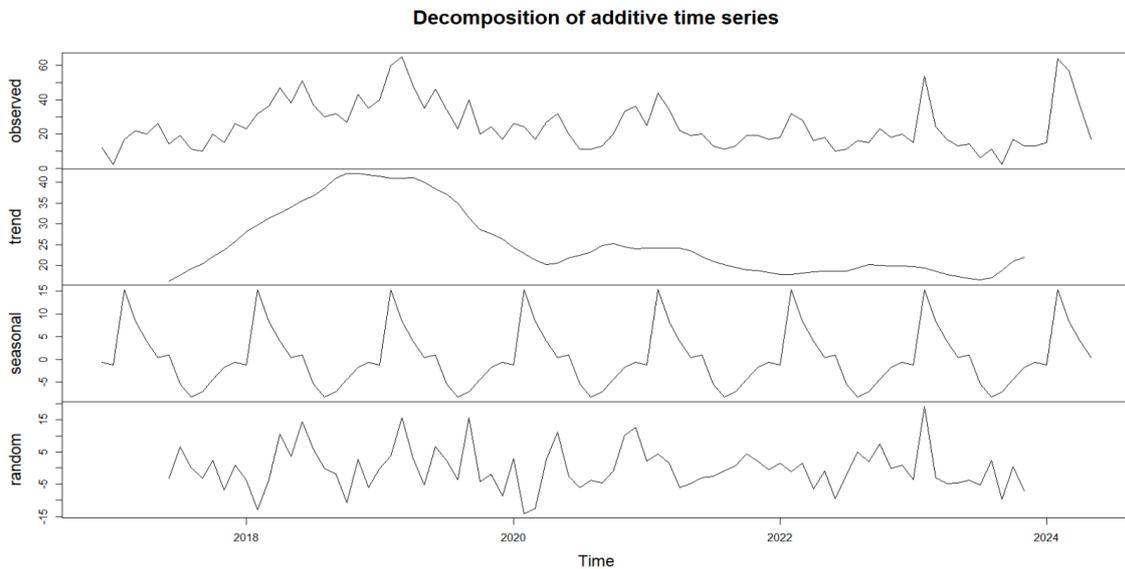
La gran mayoría de los mensajes de este tercer cluster, el cual refería a peticiones de certificados de donaciones para asuntos fiscales muestra una polaridad absolutamente **positiva**. Es posible que como la mayoría son donaciones que han realizado y están solicitando justificantes, al tener la palabra “donación”, seguido de un “muchas gracias”, puede provocar que estas peticiones las clasifique como positivas. Además, también de que los usuarios estén satisfechos con la rapidez y eficiencia de las consultas.

Este cluster también cuenta con una predicción y Recall muy elevadas, lo que significa que las predicciones de este grupo de mensajes también son significativamente fiables.

#### 4.4. Resultados Serie Temporal

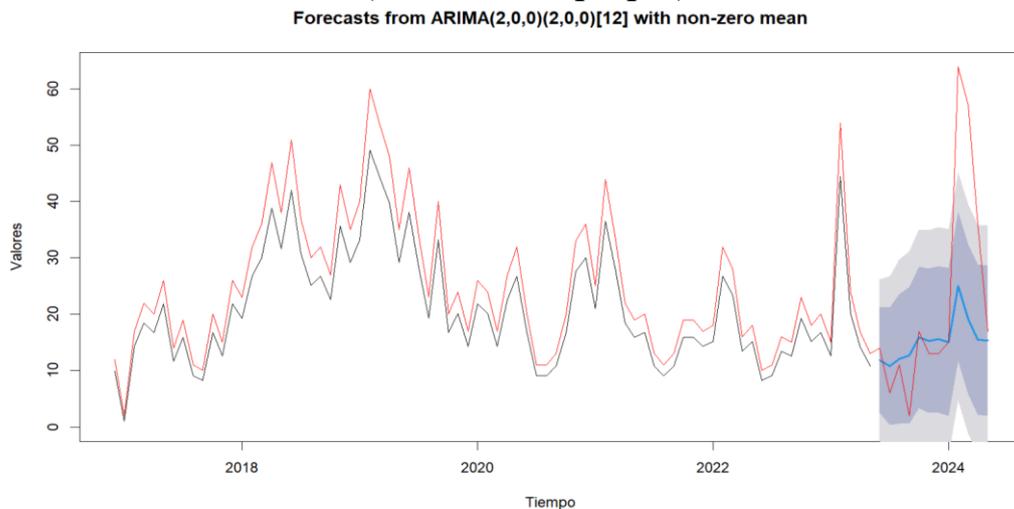
- **CLUSTER 1:**

**Figura 16: Componentes de la Serie Temporal del Cluster 1  
(Elaboración propia)**



En este gráfico podemos observar la serie temporal original y sus diferentes componentes mencionadas en la metodología que son la tendencia, estacionalidad y ruido para el cluster 1. En la tendencia (o trend) se muestra un aumento gradual en la frecuencia de los mensajes hasta el 2020 que empieza a disminuir hasta estabilizarse. Para la componente estacional se puede observar un patrón anual.

**Figura 17: Predicción de ARIMA (2,0,0)(2,0,0)[12] para el Cluster 1  
(Elaboración propia)**



Para las predicciones de los mensajes futuros del cluster 1 con el modelo ARIMA Se ha empleado la **ARIMA(2,0,0)(2,0,0)[12]**, la cual muestra una concordancia con la serie temporal ajustada. No obstante, existe cierta incertidumbre en las predicciones futuras.

El p-valor del modelo para el cluster 1 es de  $p\text{-value} = 0.05253$ , lo cual sugiere que los residuos del modelo se desvían un poco pero no de manera significativa de la normalidad.

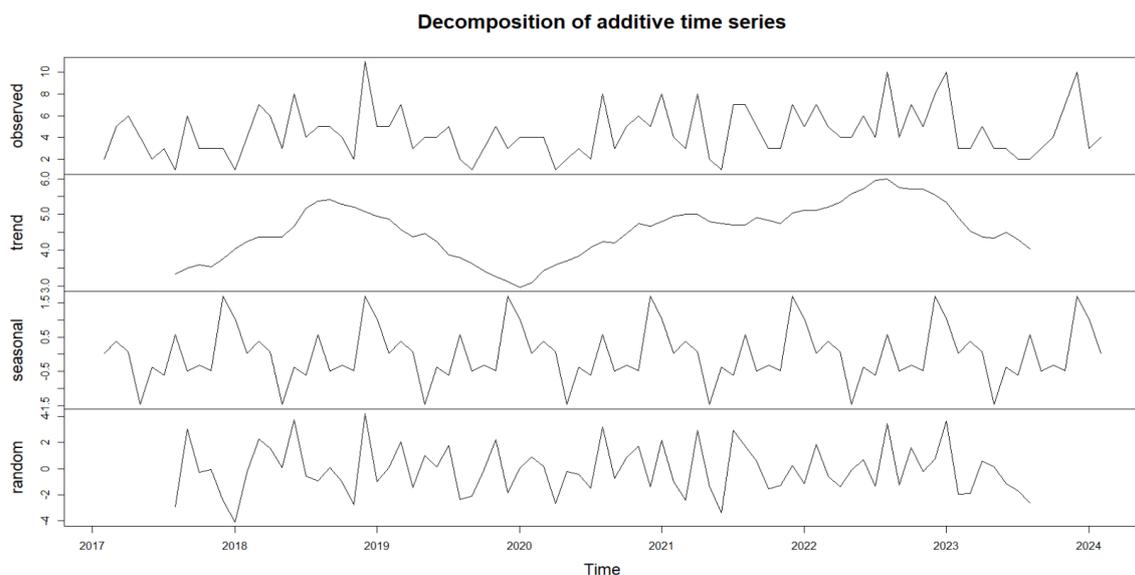
Diferencia porcentual en el RMSE: 141.935315772142

Las métricas de error indican que el modelo tiene un **rendimiento** bastante razonable en el conjunto de entrenamiento, no obstante, hay una variabilidad muy notoria en el conjunto de validación que provoca cierta incertidumbre para futuras predicciones.

En resumidas cuentas, el análisis de la serie temporal para el cluster 1 ha proporcionado información de valor para Manos Unidas y para el estudio ya que se ha detectado un aumento considerable en el 2020, posiblemente provocado por la pandemia, lo cual provocó cambios en las actividades de donaciones. Además, este tipo de mensajes observa patrones con **picos y valles** bastante predecibles, aunque las predicciones de la ARIMA generan un poco de incertidumbre, especialmente en el último año 2024, con el conjunto de validación.

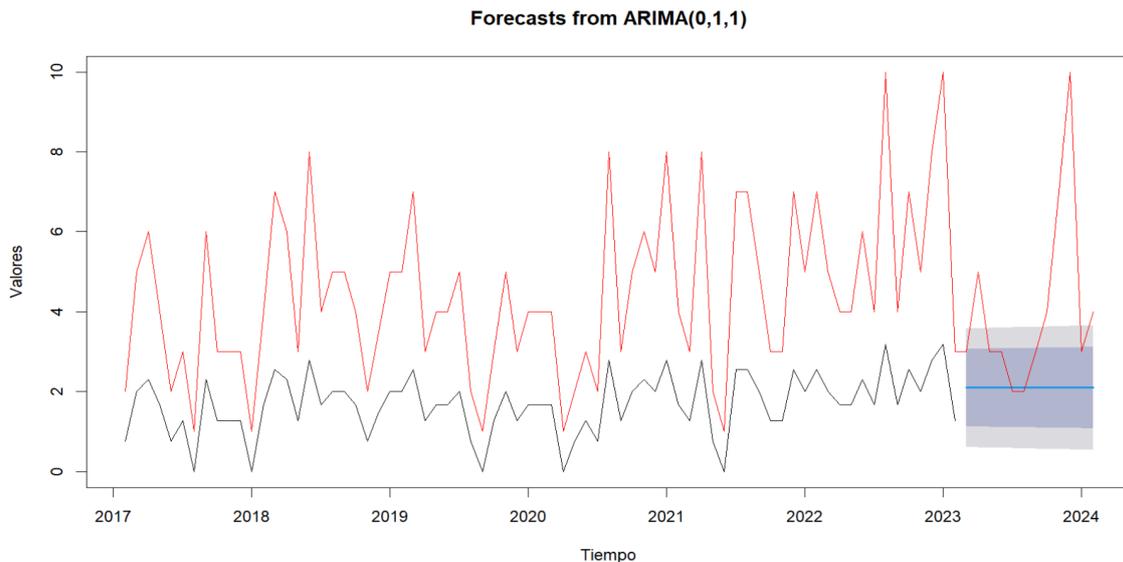
- **CLUSTER 2:**

***Figura 18: Componentes de la Serie Temporal del Cluster 2  
(Elaboración propia)***



Para la descomposición de la serie temporal de este cluster 2, primero observamos la serie temporal original. En la tendencia vemos una disminución en el 2020 bastante notable acompañada de un aumento de los mensajes antes y después. Para la estacionalidad, la serie revela unas fluctuaciones mensuales bastante consistentes que indican un patrón de repetición a nivel anual.

**Figura 19: Predicción de ARIMA (0,1,1) para el Cluster 2 (Elaboración propia)**

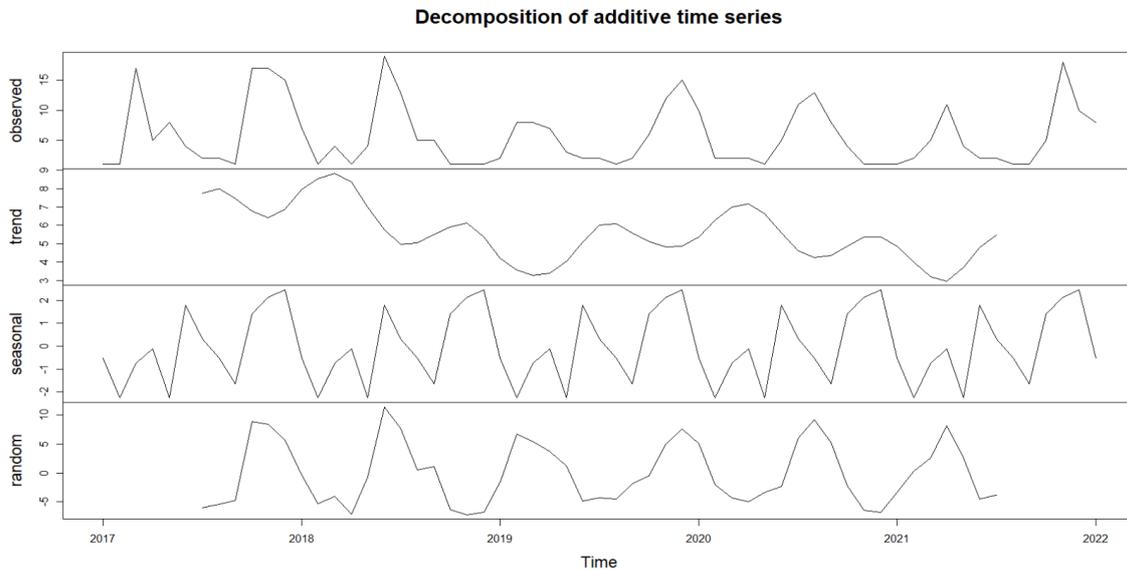


Para la predicción de la serie, se ha implementado, a través de la herramienta Rstudio, el código de `arima.fit` utilizando la **ARIMA (0,1,1)**, lo cual indica que con tan solo un término de media móvil es suficiente para capturar la dinámica de la serie temporal. No obstante, parece que no ha sido suficiente ya que las predicciones que se muestran en la Figura 19 presentan una incertidumbre muy alta ya que no acierta en el conjunto de validación. Además, este modelo ha obtenido un RMSE (2.977) significativamente mayor para el conjunto de prueba.

En resumen, este análisis de serie temporal para el cluster 2 tiene un aumento inicial en el número de mensajes en 2019, para posteriormente disminuir considerablemente y al siguiente año volver a aumentar. Asimismo, se encuentran **patrones anuales** consistentes de temporadas tanto altas como bajas de consultas al buzón de Manos Unidas. Y, por último, las predicciones del modelo ARIMA muestran demasiada **incertidumbre** y variabilidad como para fiarse de sus resultados, especialmente en el conjunto de validación, por lo que se deben interpretar con cierta cautela.

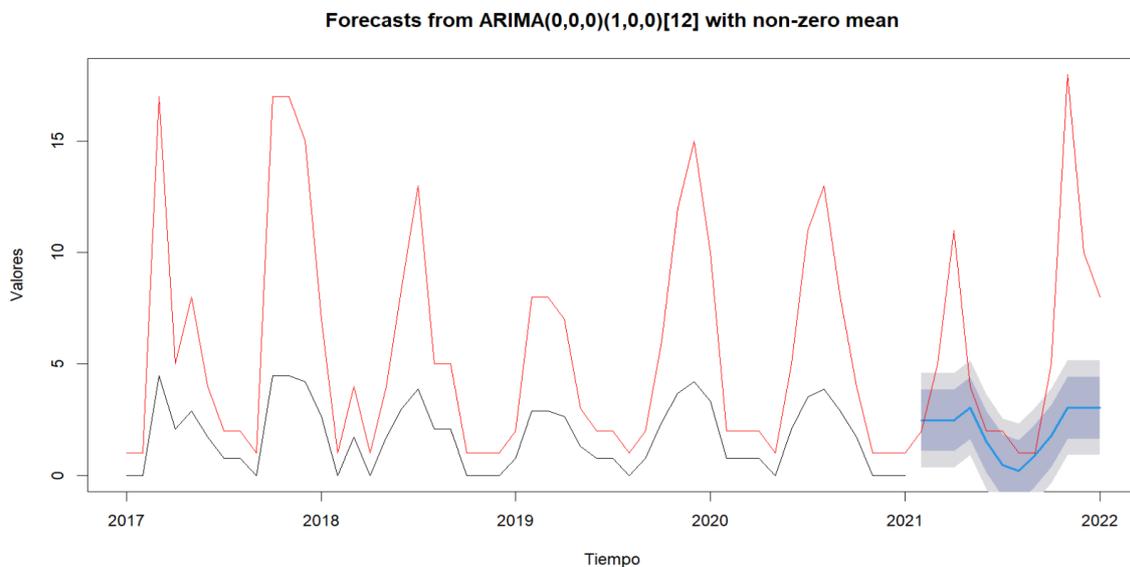
- **CLUSTER 3:**

**Figura 20: Componentes de la Serie Temporal del Cluster 3  
(Elaboración propia)**



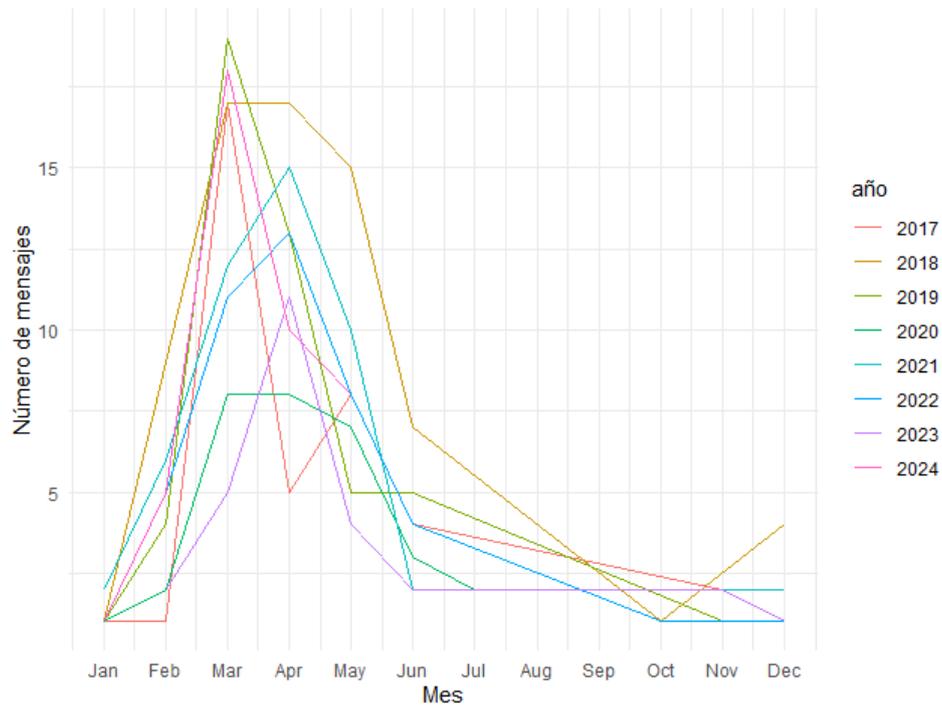
A la hora de analizar las componentes observadas de la serie temporal de este tercer cluster, se puede observar con facilidad que la serie tiene picos y valles pronunciados, lo que indica que hay fluctuaciones significativas a lo largo del tiempo. La tendencia observada tiene también un patrón ascendente y descendente, lo que explica también la estacionalidad tan clara y marcada por esos incrementos y descensos tan repetitivos en intervalos regulares.

**Figura 21: Predicción de ARIMA (0,1,1) para el Cluster 3  
(Elaboración propia)**



Para esta serie el modelo de **ARIMA (0,0,1)(1,0,0)[12]** fue ajustado. Donde las predicciones muestran una banda que si bien es verdad que acierta en los puntos de inflexión, la propia serie presenta mucha más variabilidad que las predicciones. A pesar de ello, con este análisis se puede implementar una mejor preparación para cuando los mensajes empiecen las temporadas altas.

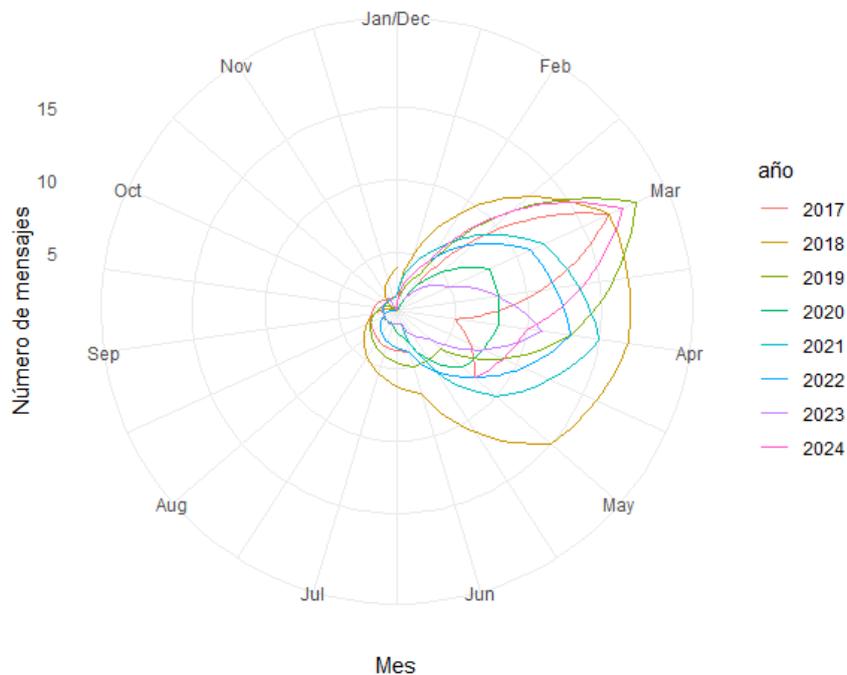
**Figura 22: Frecuencia Mensajes Mensuales por Año (Elaboración propia)**



Este gráfico de la Figura 22, muestra la **frecuencia agregada** de mensajes mensuales recibidos del cluster 3, donde en el eje X encontramos los meses del año y el eje Y el número de mensajes, y además hay una leyenda que indica el color de la línea que representa cada año registrado en esta serie temporal. Como se puede observar existe una estacionalidad muy marcada y clara, esto se debe a que, como las consultas de este tipo de mensajes son relacionadas con temas fiscales y administrativos. Por lo que el período para realizar la declaración de la renta son los meses donde más frecuencia de mensajes hay, es decir, entre marzo y junio.

**Figura 23: Frecuencia Mensajes Mensuales por Año (Representación Polar) (Elaboración propia)**

Mensajes mensuales por año (Formato Polar)



Lo mismo sucede con este gráfico, donde el mes de **marzo y abril** son los que concentran casi toda la frecuencia de consultas comparadas con otros meses. Esta representación está basada en los mismos datos que la anterior (la Figura 22), solo que, al representarse en formato polar, nos podemos dar más cuenta de la diferencia entre la temporada alta y la baja en un mismo año. Por lo tanto, los administrativos que se encarguen de responder a este tipo de cuestiones van a tener que estar bien preparados en los meses donde se realiza la declaración de la renta y se cierran los ejercicios del año anterior.

## CAPITULO V: CONCLUSIONES

Una vez realizado el análisis exhaustivo y detallado de los resultados obtenidos a lo largo del presente estudio, se han obtenido unas **conclusiones muy enriquecedoras** que ofrecen diferentes perspectivas de los análisis realizados. Estas conclusiones pretenden responder a los objetivos establecidos en el inicio de la investigación, además de resaltar tanto los puntos fuertes y aprendizajes positivos, como las debilidades y aspectos a mejorar de cada uno de los siguientes análisis partiendo del conocimiento extraído en este Trabajo de Fin de Grado.

En cuanto al objetivo general de la investigación, éste cumple de manera **exitosa** las expectativas debido a que se ha obtenido una **visión muy clara y profunda** del significado de los mensajes recibidos por Manos Unidas. Los análisis que se han realizado anteriormente ayudan a proporcionar una base sólida que permite no solo una **optimización de la gestión de recursos**, sino también el **aumento del impacto en las campañas**.

En el primer análisis de todo el estudio, trataba de extraer información de la base de datos e identificar los mensajes clave de la base de datos. Tras la implementación de proceso se ha conseguido **eliminar el spam**, garantizando así que solo se estudien los mensajes relevantes para los posteriores análisis. La implementación de técnicas de procesamiento y filtrado mediante las keywords han sido efectivas, no obstante, al principio fue un trabajo poco eficiente ya se tuvo que buscar manualmente en los datos muchas palabras que se encontrasen en los mensajes de spam y no en los relevantes. Por lo tanto, a pesar de no haber eliminado los mensajes de spam de una manera más rápida, este análisis ha sido **estrictamente necesario para obtener resultados** con información coherente y valiosa.

En base a los resultados conseguidos procedentes al segundo análisis, el modelo de Machine Learning implementado para la **detección de mensajes de spam**, se pueden extraer conclusiones muy útiles y aplicables. Para este modelo, se redujo de forma significativa la carga de trabajo manual encargada de leer los mensajes y no borrar los que no aportaban valor. Además, con los **excelentes resultados** que mostraba el modelo de clasificación de spam basado en el método de **Random Forest**, con este análisis se puede automatizar la detección de mensajes no valiosos, lo cual ayuda a reducir aún más carga de trabajo. Asimismo, como recomendación, el modelo podría ser mejorado y,

sobre todo, actualizado periódicamente con el fin de adaptarse a las nuevas tendencias, para que sus resultados continúen siendo igual más acertados.

Con respecto al tercer análisis del estudio, donde se realizaba un **clustering** para segmentar la población de los mensajes de Manos Unidas en diferentes grupos según sus características, se identificaron **tres grupos**, o clusters, diferentes. Estos clusters han permitido a esta investigación obtener una mejor comprensión de las preocupaciones de los socios y colaboradores de la organización. Cada cluster tenía palabras frecuentes totalmente diferentes al resto de clusters, pero muy parecidas entre las palabras del mismo cluster, lo cual ha permitido diferenciar rápidamente el **significado de cada grupo** de mensajes. Esto puede ayudar a la institución sin ánimo de lucro de muchas formas, la primera es para obtener una mejor gestión de sus recursos administrativos, puesto que pueden crear diferentes departamentos especializado para un tipo de consulta determinada. Asimismo, también se puede desarrollar específicas estrategias de respuesta para cada cluster, donde se enfoque en las principales inquietudes y necesidades identificadas de cada grupo.

Para el método de **análisis del sentimiento de los mensajes** de cada cluster empleando transformador de BERT, se ha obtenido resultados importantes a la hora de entender mejor la información de cada grupo. La **evaluación cualitativa** de las emociones y, concretamente, de la **polaridad** de los mensajes ha ayudado a la obtención de una visión clara del sentimiento y tono de los remitentes. El empleo de este sentiment analysis puede ayudar a una mayor personalización de las respuestas a estas inquietudes dependiendo del cluster al que pertenezca cada mensaje. Este tipo de análisis también ayuda a poner el foco en los tipos de mensajes que tienen una polaridad negativa, ya que podemos conocer de manera objetiva la insatisfacción de los socios y colaboradores para esos mensajes.

Finalmente, el último análisis empleado, la **serie temporal** que analiza la frecuencia además de los patrones de los mensajes en cada cluster, consiguió identificar de manera mayoritariamente acertada los patrones de mayor actividad de consultas a la organización para cada cluster. Estos resultados ayudan a la ONG a **preparar mejor los períodos de mayor frecuencia** de mensajes, relacionados con cada tipo de solicitudes que enviaban los remitentes. Por lo que, por ejemplo, Manos Unidas deberá implementar un método para descargar fácilmente los certificados de donaciones en los meses de primavera, como consecuencia de ser los meses destinados a realizar la declaración de la renta. Aunque aun así deberá asegurarse de que haya suficiente personal y recursos disponibles para

proporcionar el mejor servicio de asistencia posible durante estos críticos meses. Con las predicciones diseñadas, brinda la posibilidad de ajustar los recursos y la planificación en función de los picos de actividad de los mensajes.

Una vez terminadas las conclusiones y recomendaciones específicas para cada análisis, se va a identificar tanto las debilidades y puntos a mejorar, como los puntos fuertes y aprendizajes durante la investigación y el empleo de este Proyecto de Fin de Grado.

Por una parte, para las **debilidades** y para los aspectos de mejora, a pesar de los esfuerzos de limpieza empleados, pueden seguir **mejorándose la calidad y el volumen** de los datos para asegurar que la base de datos sea más robusta y completa, y así poder realizar análisis posteriores más precisos. Añadido a esto, los modelos de Machine Learning y de análisis de sentimiento **requieren unas actualizaciones** con cierta frecuencia, ya que el tipo de mensajes y sus expresiones pueden cambiar en los próximos años. Y la última oportunidad para mejorar futuros análisis es una **unificación de herramientas** más avanzadas que puedan automatizar el procedimiento de los diferentes análisis para reducir tiempo de trabajo.

Y, por otra parte, se han encontrado tres **aprendizajes positivos** que han ayudado a mejorar el estudio. El primero de ellos muestra que la metodología aplicada ha permitido que el análisis de los mensajes sea muy **exhaustivo y detallado**, proporcionando así una comprensión profunda de los patrones y las tendencias para cada cluster. El segundo punto fuerte ha sido la utilización de unas **técnicas muy avanzadas de Machine Learning** y de Natural Language Processing (NLP), que ha permitido aumentar la eficiencia de gestión de mensajes. Y, por último, la **identificación de los periodos de alta actividad** mediante las series temporales brinda la oportunidad a Manos Unidas de poder realizar una mejor planificación y asignación de los recursos disponibles para mejorar el trato con los remitentes de los mensajes.

En resumen, este estudio ha conseguido cumplir con los objetivos declarados al principio de la investigación, proporcionando también herramientas y análisis de mucho valor. Estos análisis han conducido a **recomendaciones detalladas y acertadas** para mejorar la eficiencia de la gestión de recursos y el impacto de las campañas de Manos Unidas. Consecuentemente, la implementación de las conclusiones y recomendaciones a Manos Unidas permite que la organización pueda **mejorar su relación con los socios y colaboradores**. No obstante, la última instancia la tiene la propia organización.

## CAPITULO VI: FUTURAS LÍNEAS DE INVESTIGACIÓN

El conjunto de diferentes análisis que se han llevado a cabo en esta investigación ha proporcionado una visión clara y una base muy sólida para la comprensión de los mensajes recibidos por Manos Unidas, donde ahora van a poder comprender de una mejor manera las preocupaciones y necesidades de sus socios y colaboradores. A pesar de ello, este estudio puede ser tan solo el punto de salida hacia un mayor entendimiento de las inquietudes de las personas que están en contacto con la institución. Esto se debe a que existen numerosos análisis enriquecedores que no se han empleado en este estudio, pero que pueden ser considerados como futuras líneas de investigación para expandir los descubrimientos realizados en el Trabajo de Fin de Grado presente.

Algunos de los ejemplos de diferentes análisis más experimentales, pero con mucho potencial pueden ser los siguientes:

- Un **meta-análisis con transformadores**. Este tipo de análisis se basa en la utilización de múltiples modelos transformadores que, al combinar sus resultados entre sí, mejoran significativamente la precisión y la calidad de las predicciones. Así los resultados obtenidos por un transformador pueden ayudar a enriquecer los resultados de otro más complejo.
- **Redes neuronales de KAN** (Kolmogorov-Arnold Neural Networks). Este tipo de redes neuronales muestran una estructura totalmente diferente a las consolidadas en el ámbito de Deep Learning. A pesar de que todavía estén en la fase de desarrollo y no proporcionen resultados excelentes, se predice que tienen mucho potencial para provocar mucho impacto en los análisis de text mining.
- **Redes sociales y Grafos**. Otro tipo de análisis que de momento se encuentra en fases experimentales y de desarrollo es el análisis de redes sociales y grafos. Este tipo de análisis ofrece una mejor comprensión de los mensajes mediante la búsqueda de conexión entre los diferentes mensajes. Mediante las redes sociales y grafos se puede identificar diferentes comunidades dentro de la red de mensajes de Manos Unidas para así poder segmentar campañas y adaptarlas a las necesidades de cada grupo diferente.

En conclusión, este estudio de los mensajes abre muchas puertas a posibles futuras investigaciones más técnicas, para sacar conclusiones más detalladas o para confirmar las obtenidas anteriormente. Y como cada vez se están desarrollando y descubriendo nuevos

y mejores métodos para el procesamiento de texto, aún existen muchas posibilidades de realizar más análisis que proporcionen una información única y valiosa para Manos Unidas.

## CAPÍTULO VII: ANEXOS

### 7.1. Declaración de Uso de Herramientas de Inteligencia Artificial en el Trabajo de Fin de Grado

**ADVERTENCIA:** Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Sergio Íñiguez Rodríguez, estudiante de ADE y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado MANOS UNIDAS. Análisis cuantitativo y cualitativo de los mensajes recibidos del buzón de Manos Unidas., declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Referencias:** Usado juntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Interpretador de código:** Para realizar análisis de datos preliminares.
4. **Constructor de plantillas:** Para diseñar formatos específicos para secciones del trabajo.
5. **Corrector de estilo literario y de lenguaje:** Para mejorar la calidad lingüística y estilística del texto.
6. **Sintetizador y divulgador de libros complicados:** Para resumir y comprender literatura compleja.
7. **Revisor:** Para recibir sugerencias sobre cómo mejorar y perfeccionar el trabajo con diferentes niveles de exigencia.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: [03/06/2024]

Firma: Sergio Íñiguez Rodríguez

## 7.2. Bibliografía

Arroyo Almaraz, I., Baladrón Pazo, A. J., & Martín Nieto, R. (2013). La comunicación en redes sociales: Percepciones y usos de las ONG españolas. *Cuadernos.info*, 32, 77-88. <https://doi.org/10.7764/cdi.32.497>

Carramiñana, E. N., García del Junco, J., Rafael de Reyna Zaballa, & Rafael Robina Ramírez. (2013). *Los roles de colaboración entre las principales ONG españolas* (79, 87-114.). <http://www.redalyc.org/articulo.oa?id=17429940004>

Diego Olite, F. M., Morales Suárez, I. del R., Vidal Ledo, M. J., Diego Olite, F. M., Morales Suárez, I. del R., & Vidal Ledo, M. J. (2023). Chat GPT: Origen, evolución, retos e impactos en la educación. *Educación Médica Superior*, 37(2). [http://scielo.sld.cu/scielo.php?script=sci\\_abstract&pid=S0864-21412023000200016&lng=es&nrm=iso&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_abstract&pid=S0864-21412023000200016&lng=es&nrm=iso&tlng=es)

Manos Unidas. (2022). *Manos Unidas: Memoria de actividades 2022*.

Manos Unidas. (2023, marzo 5). *Manos Unidas. Acabar con el hambre es posible*. ONG Manos Unidas. <https://www.manosunidas.org/que-hacemos>

Marfil-Carmona, R. (2014). Interactividad digital y estrategias narrativas en la publicidad audiovisual de Manos Unidas y Unicef. *Historia y Comunicación Social*, 18(0), 169-181. [https://doi.org/10.5209/rev\\_HICS.2013.v18.43957](https://doi.org/10.5209/rev_HICS.2013.v18.43957)

Marfil-Carmona, R. (2018). *La representación de la salud en la publicidad audiovisual de Manos Unidas. Análisis de contenido de las campañas institucionales de 1995 a 2017*. <https://doi.org/10.14198/MEDCOM2018.9.2.19>

*Memoria 2022—Nuestra indiferencia los condena al olvido*. (2023, mayo 18). ONG Manos Unidas. <https://www.manosunidas.org/memoria-2022>

Mesa, M. (2014). *PRECEDENTES Y EVOLUCIÓN DE LA EDUCACIÓN PARA EL DESARROLLO: UN MODELO DE CINCO GENERACIONES*.

Olivera Taboada, L. A., Sialer Puelles, M. A., & Velarde Gonzales, J. J. (2020). Aplicación de Data Science para el análisis de una campaña de respuesta directa televisiva de UNICEF Perú. *Universidad Peruana de Ciencias Aplicadas (UPC)*. <https://repositorioacademico.upc.edu.pe/handle/10757/655986>

Río Ciriza, J. del. (2018). *Diseño de una plataforma cognitiva para la gestión inteligente de la documentación del itdUPM* [Info:eu-repo/semantics/bachelorThesis]. E.T.S.I. Telecomunicación (UPM). <https://oa.upm.es/51994/>

Ruiz Simón, E. (2020). *Historia de la Cooperación al Desarrollo: El caso de Manos Unidas*. <https://www.educacion.gob.es/teseo/imprimirFicheroTesis.do?idFichero=bi0TshQVyh w%3D>

Saavedra, P., & Aurora, S. (2019). *MARKETING RELACIONAL Y SU RELACIÓN CON LA FIDELIZACIÓN DE CLIENTES DE LA EMPRESA TIKA MANOS UNIDAS SMP. LIMA AÑO 2018*. 80.

Segovia-Guerrero, A. D., & González-Cantos, M. C. (2024). El papel de las Organizaciones Internacionales en la Cooperación Global. 92, 9(4), 2127-2140. <https://doi.org/10.23857/pc.v9i4.7047>

Tao, J., Zhou, L., & Feeney, C. (2019). *I Understand What You Are Saying: Leveraging Deep Learning Techniques for Aspect Based Sentiment Analysis*. Hawaii International Conference on System Sciences. <https://doi.org/10.24251/HICSS.2019.057>

Tufani, A. (2020, julio 17). *Manos Unidas: 57 países y 6 millones de personas beneficiadas en 2019—Vatican News*. <https://www.vaticannews.va/es/iglesia/news/2020-07/manos-unidas-memoria-actividades-2019-personas-beneficiadas.html>

Xavier, F. C. (1979). MANOS UNIDAS.