

Especificación de requisitos

ITAPP

Hoja de control

Organismo	Equipo de desarrollo del producto ITAPP		
Proyecto	ITAPP		
Entregable	Especificación de requisitos		
Autor	Haritz Saiz, Xabier Gandiaga, Álvaro Huarte, Xabier Etxezarreta, Onintza Ugarte y Ander Bolumburu		
Versión/Edición	0001	Fecha Versión	30/05/20
Aprobado por		Fecha Aprobación	08/06/20
		Nº Total de Páginas	24

REGISTRO DE CAMBIOS

Versión	Causa del Cambio	Responsable del Cambio	Fecha del Cambio
0100	Versión inicial	Ander Bolumburu	25/05/20

CONTROL DE DISTRIBUCIÓN

Nombre y Apellidos
Ander Bolumburu Casado

Indice

1.	Introducción.....	2
1.1	Alcance	5
1.2	Objetivos.....	5
2.	Información del dominio de problema.....	7
2.1	Introducción al Dominio del Problema	7
2.2	Glosario de Términos.....	8
3.	Descripción de la situación actual	11
4.	Necesidades de negocio.....	12
4.1	Objetivos de negocio.....	12
4.2	Modelo de proceso de negocio a implantar	14
4.2.1	Descripción de los actores de negocio a implantar	14
4.2.2	Descripción de procesos de negocio a implantar.....	14
5.	Clasificación de los subsistemas del sistema a desarrollar	16
6.	Catálogo de requisitos del sistema a desarrollar.....	18
6.1	Requisitos Generales del Sistema	18
6.2	Casos de uso del sistema.....	18
6.3	Requisitos Funcionales del Sistema	18
6.3.1	Requisitos de información del sistema	19
6.3.2	Requisitos de Reglas de Negocio del Sistema	19
6.3.3	Requisitos de Conducta del Sistema.....	19
6.4	Requisitos No Funcionales del Sistema.....	20
6.4.1	Requisitos de fiabilidad.....	20
6.4.2	Requisitos de Usabilidad	20
6.4.3	Requisitos de eficiencia.....	20
6.4.4	Requisitos de Mantenibilidad.....	21
6.4.5	Requisitos de Portabilidad	21
6.4.6	Requisitos de seguridad	22
6.5	Restricciones Técnicas del Sistema	22
6.6	Requisitos de integración de sistema	22

Tabla de contenido

Tabla 1. Conjunto de validación del software.....	12
Tabla 2. Procesamiento del lenguaje natural.....	12
Tabla 3. Realización del entrenamiento distribuido	13
Tabla 4. Etiquetación de versiones y almacenamiento de código en repositorio.....	13
Tabla 5. Community Manager.....	14
Tabla 6. Flujo general de la aplicación.....	14
Tabla 7. Sistema de ingesta y tratamineto de datos	16
Tabla 8. Aplicación web.....	16
Tabla 9. Entrenamiento distribuido del modelo.....	17
Tabla 10. Almacenamiento del código y etiquetado de versiones para el despliegue	17
Tabla 11. Sistema de ficheros distribuido.....	17
Tabla 12. Generaición de texto	18
Tabla 13. Almacenamiento de los datos del sistema de ingesta	19
Tabla 14. Consulta de histórico de publicaciones	19
Tabla 15. Compatibilidad con los navegadores actuales	22
Tabla 16. Integración con las APIs de terceros	23

1. Introducción

Hoy en día vivimos en una sociedad en la que el uso de redes sociales es muy habitual. Tal es el hábito que tiene la gente a utilizarlas, que estas han sido la primera opción de gran parte de la sociedad a la hora de ponerse en contacto con las empresas cuando tienen algún tipo de problema con sus productos.

Es ahí donde cobra importancia el rol del Community Manager (en adelante, CM), ya que es el encargado de coordinar las redes de una empresa y atender las posibles incidencias que puedan tener los clientes con sus productos. Es por ello por lo que la actitud que dé el CM de cara al público puede influir mucho en la imagen de la empresa.

Con el fin de que esa imagen sea atractiva de cara a la clientela, hemos propuesto la creación de ITAPP. Esta aplicación consiste en una *single page application* en la cual el CM introducirá diferentes temáticas de intereses para sus seguidores de la cual le gustaría generar un texto y esta, mediante un modelo de Deep Learning, le proporcionará un texto que posteriormente el usuario elegirá si publicar o no.

Por contextualizarlo un poco, pondremos dos casos de uso:

- Imaginemos que el CM de la cadena de venta de videojuegos GAME, quiere que sus seguidores estén al tanto de la llegada de las nuevas consolas PlayStation 5 y Xbox Series X. Lo que hará será introducir una de esas dos temáticas en la aplicación y esta le proporcionará un texto con el que captar la atención de sus clientes y al mismo tiempo crear un aliciente de compra. Por lo que la empresa sale doblemente beneficiada, por un lado, la imagen de la empresa en las redes sociales mejorará (ya que además de atender a los clientes, anuncia novedades sobre los videojuegos) y por el otro sale beneficiada económicamente ya que, al anunciar estas novedades, al ser un público propenso al consumo de videojuegos, los comprarán en sus tiendas.
- Por otro lado, imaginemos una cuenta que se dedica a publicar noticias de todo tipo. El dueño de la cuenta no tendría que preocuparse en redactar pequeños textos que sean de interés de sus seguidores. Con introducir diferentes temas de actualidad coronavirus, por ejemplo, ITAPP será capaz de proporcionarle un texto publicable en su cuenta. De este modo, poco a poco empezaría a ganar relevancia

en las redes sociales y dependiendo de esta, podría comenzar a explotarla con campañas publicitarias, por ejemplo.

Hablando de este último punto, como estrategia de marketing el uso de Itapp sin fines concretos también es posible. Ya se ha dado el caso de empresas que simplemente han anunciado el que creasen cuentas gestionadas completamente por bots. Esto genera interés que por consiguiente genera tráfico, haciendo que la gente que estuviera interesado por el bot terminase conociendo el nombre de la marca responsable del bot.

Aunque de trágico final, el ejemplo más conocido es el del bot que publicó Microsoft, Tay.ai, con el simple objetivo de tener un bot autónomo gestionando una cuenta, lo cual generó mucho tráfico que hablaba sobre Microsoft.

Se podría intentar lo mismo mediante otros bots o scripts que utilizasen Itapp como base.

1.1 Alcance

El desarrollo del sistema afecta plenamente en toda la sociedad, ya que el contenido que proporciona la aplicación está relacionado con las redes sociales y esta al mismo tiempo por los motivos dados en la sección anterior, a la sociedad.

1.2 Objetivos

Los principales objetivos para cumplir por el desarrollo de este producto son los siguientes:

- El texto proporcionado por el modelo de inteligencia artificial debe tener coherencia respecto a la(s) temática(s) proporcionada(s) por el usuario.
- El entrenamiento del modelo debe ser realizado de forma distribuida (entre varias máquinas).
- El desarrollo debe ser seguro y contemplar la seguridad en todos los apartados del desarrollo.
- Se debe llevar a cabo un control de versiones en el repositorio con el fin de etiquetar las versiones aptas para poner en producción.
- Se deben desarrollar un conjunto de validación de la aplicación con el fin de validar su funcionamiento en cada aporte de código (commit).
- Se deben almacenar tanto el modelo como el conjunto de datos en un sistema de ficheros distribuido.

- Se debe crear un sistema de despliegue automatizado para la aplicación que realice la validación.

2. Información del dominio de problema

El dominio del problema, como se ha especificado en la sección anterior, se concierne al mundo de las redes sociales, en el cual podemos diferenciar diferentes actores:

- **Empresa:** Organización que quiere adaptarse a la sociedad de hoy en día, haciendo uso de las redes sociales, con el fin de obtener beneficio. Este beneficio puede verse desde dos puntos de vista, bien por el aspecto económico o por el de la imagen de cara a sus clientes.
- **Red social:** Una red social es una estructura social compuesta por un conjunto de usuarios (tales como individuos u organizaciones) que están relacionados de acuerdo con algún criterio (relación profesional, amistad, parentesco, entre otras). Hoy en día son muy utilizadas y son un nexo habitual entre una compañía y sus clientes.
- **Community Manager (CM):** Encargado de gestionar las redes sociales de una empresa concreta. Su objetivo es el dar solución a los posibles problemas que puedan tener los clientes y por otro lado generar contenido atractivo para los seguidores con el fin de tener una buena imagen de cara a los mismos.
- **Cliente:** Usuario de la red social que interactúa con otros usuarios o cuentas representantes de empresas con el fin de dar solución a sus problemas u obtener información.
- **Agencia de publicidad:** Una agencia de publicidad es una empresa de carácter independiente, compuesta por un personal de profesionales y creativos, con aptitudes en marketing y publicidad, que desarrolla, prepara y coloca la publicidad, por cuenta de un anunciante que busca encontrar un respectivo consumidor para sus bienes y servicios o difundir sus ideas.

2.1 Introducción al Dominio del Problema

Como se ha especificado en la introducción los cinco actores entran en juego de la siguiente forma:

La empresa desea adaptarse a la sociedad de hoy en día, haciendo uso de las redes sociales, con el fin de obtener un beneficio comentado en el apartado anterior. Para ello, contrata a un Community Manager (CM) con el fin de que encargue de gestionar estas redes (dando solución a los problemas que puedan surgir en el día a día) y generar

contenido atractivo para sus seguidores. Si el contenido es suficientemente atractivo y además genera aliciente de compra, la compañía se va a ver doblemente beneficiada ya que su imagen de cara al público será mejorada y además verá un beneficio económico ya que los clientes consumirán el producto que promocionan.

Es por ello por lo que se ha pensado en la herramienta ITAPP, que, mediante inteligencia artificial, genera contenido relacionado con las temáticas que se introduzcan.

2.2 Glosario de Términos

- **#:** Símbolo tipográfico que en redes sociales es conocido como Hashtag. Es usado para indicar palabras relevantes dentro de un determinado contexto, siendo posible indexarlo en el directorio de búsqueda de redes como el Twitter, Facebook, Google + e Instagram, por ejemplo, en forma de hipervínculo o enlace para búsqueda del contenido marcado.
- **Alcance de una publicación:** Término usado para indicar cuántas personas accedieron a determinada publicación en una red social, en especial en Facebook. Esta es una de las principales métricas evaluadas en la gestión de redes sociales y para la elaboración de nuevas estrategias de alcance y engagement de público.
- **Algoritmo:** Secuencia de reglas o fórmulas usadas para la resolución de un problema. Este concepto es usado con cierta constancia en la definición de herramientas internas como Google o Facebook, por ejemplo, en que la indicación de un algoritmo es el responsable por la realización de búsquedas inteligentes, clasificación de publicaciones, segmentación de públicos, etc.
- **Anuncio:** Los anuncios, o Ads, son publicaciones pagas que ocupan lugares estratégicos dentro de las redes sociales. En Facebook, ellos pueden ser encontrados en la parte derecha de la pantalla. Son organizados y seleccionados de acuerdo con los intereses del usuario. Es una herramienta interesante para la promoción del contenido puntual en la red y por Facebook. Está abierta para cualquier usuario que desee experimentar.
- **API:** Una API comprende patrones de programación que permiten el acceso a determinada aplicación o software, a parte de su integración con otras funciones personalizadas. El Graph API, de Facebook, y Fabric, de Twitter, son los dos APIs más conocidos y replicados de las redes sociales actualmente.

- **Big Data:** De manera resumida, Big Data indica una gran cantidad de datos complejos reunidos en un determinado local. Esas informaciones necesitan ser pulidas y organizadas para que luego puedan ser analizadas con alguna precisión. En redes sociales, es posible aprovechar el potencial del Big Data para la obtención de informaciones privilegiadas sobre su público, tendencias del área y oportunidades de negocio.
- **Block:** La expresión block es más conocido en Twitter, pero es válido para la mayor parte de redes sociales como sinónimo de limitación de acceso de un determinado usuario a tus informaciones. La acción de block garantiza el bloqueo de una cuenta con el objetivo de impedir que existan futuras interacciones con el solicitante, sin limitar las otras funciones.
- **Campaña:** Reunión de uno o más anuncios (o conjunto de anuncios) pagos en el Facebook y todo el proceso referente a su publicación, monitoreo y gestión.
- **Chat:** Conversación, o charla, a distancia e en el momento, basada en el envío de mensajes textuales.
- **Clickbait:** El Clickbait indica todo tipo de enlace formulado para llamar la atención del usuario para una página externa con pocas o ninguna relación con la expectativa generada anteriormente. Un ejemplo común incluye publicaciones con títulos sensacionalistas que son desmentidos por el contenido. Este tipo de publicación generalmente es hecha con el objetivo de generar visitas o impresiones para un determinado enlace.
- **Denunciar como SPAM:** Función semejante a la de proveedores de e-mail, permite que el usuario denuncie al moderador de la red social cualquier contenido que juzgue impropio o que pueda ser considerado como spam. Generalmente, viene acompañado de la opción bloquear el responsable por la publicación.
- **Feed:** Formación de datos en las redes sociales o blogs para proporcionar contenido de forma organizada y cronológica. Puede ser entendido como la línea de tiempo que demuestra las actualizaciones de los contactos en una red, o el RSS de un blog.
- **Followers:** Seguidores. De forma general, son usuarios que se inscriben en tu red social con la intención de recibir tus principales actualizaciones.

- **Influyente/Influencer:** Usuario con grande área de alcance en las redes sociales. Es capaz de influenciar en la decisión de compra o interacción de otros usuarios comunes.
- **Me gusta:** También conocido como “like”, es usado cuando al usuario le gusta o simpatiza con una publicación determinada. En Facebook, es determinado por una mano cerrada con pulgar hacia arriba, mientras que en Instagram el símbolo que representa es un corazón.
- **Meme:** En las redes sociales, memes son imágenes, expresiones, prints, videos, gifs, etc, pasados de manera viral entre usuarios de forma continua.
- **Mutear:** Función que permite silenciar usuarios en Twitter sin que necesites dejar de seguirlos. El usuario sigue dentro de tu red de amigos, pero sus mensajes no son más indexados a tu timeline.
- **Notificación:** Mensaje que avisa sobre alguna novedad o actualización en redes sociales. Si un usuario interactúa con una publicación en Facebook, por ejemplo, recibes una notificación en la red o en el celular (caso habilites la función).
- **Phishing:** Intento fraudulento de obtener informaciones particulares como logins, contraseñas e informaciones de tarjeta de crédito por el intento de replicar layouts de instituciones de confianza, principalmente por e-mails.
- **Post:** Publicación. Entradas de texto (pudiendo ser acompañada por imágenes, audios, gifs, videos, etc) en blogs, sitios y redes sociales, generalmente dispuestos de manera cronológica.
- **Reply:** Replies, en Twitter, indican respuestas directas por medio de la mención de un usuario o Tweet anterior. Implica el uso del botón “reply” o “responder” indicado por una flecha abajo del tweet de tu interés. Se establecerá una relación cronológica de la charla que puede ser exhibida en la línea de tiempo de la red.
- **Retweet:** La función de Retweet es semejante a la función de compartir en carácter automático del Facebook e indica que un usuario replico tu tweet en la propia línea de tiempo de Twitter.
- **Share:** Lo mismo que “compartir”. Es usada en Facebook cuando un usuario decide replicar determinada publicación en un grupo o timeline.
- **Trending Topic (TT):** Expresión popular entre usuarios de Twitter. Indica los principales temas hablados en un espacio de tiempo dentro de la red social. Pueden ser visualizados por ciudad, país o de modo global.

- **Viral:** Término usado para designar cualquier texto, imagen o contenido que se difunda con alta velocidad por Internet (generalmente por el elevado índice de difusión) generando aumento exponencial de visitación cuando comparado con las tasas normales de acceso del negocio.

3. Descripción de la situación actual

Actualmente, este es el primer desarrollo de la empresa y se va a llevar a cabo pensando en que existe una necesidad de negocio como se explicará en la siguiente sección.

Aunque antes que nada creemos oportuno describir cómo se encuentra la situación actual en este contexto.

Como se ha explicado anteriormente, hoy en día la labor del Community Manager se centra en la creación de contenido atractivo para sus seguidores, y existen casos o contextos en el cual la información abunda. En cambio, hay otros en los que la variedad brilla por su ausencia y la labor del CM se dificulta enormemente.

Es por ello por lo que nos ha resultado atractiva la idea de desarrollar una herramienta a la cual, introduciendo unas temáticas concretas, haciendo uso de un modelo de inteligencia artificial, sea capaz de generar un contenido publicable relacionado con las temáticas proporcionadas como input.

Por otra parte, como se ha comentado anteriormente, contamos de ejemplos donde otras empresas han intentado crear un bot propio o han utilizado uno prehecho para generar tráfico y así atraer público hacia su marca.

Finalmente, como actualmente no existe nada desarrollado relacionado con la problemática definida no tenemos pros o contras que valorar y apoyarnos con el fin de mejorar el desarrollo del software establecido.

4. Necesidades de negocio

En esta sección se detallará información sobre los objetivos de negocio de clientes y usuarios, incluyendo los modelos de procesos de negocio a implantar.

4.1 Objetivos de negocio

Mediante esta sección se detallarán los objetivos de negocio que se esperan alcanzar cuando el sistema software a desarrollar esté en producción, especificados mediante las plantillas de objetivos de negocio que se muestran a continuación.

1	Conjunto de validación del software
Dependencias	<ul style="list-style-type: none"> - Desarrollar con herramientas que sean validables. - Si se debe automatizar, debe considerarse una herramienta que lo permita. - La validación constata de tests de funcionalidad y análisis de código estático.
Descripción	Realizar un conjunto de validación del software con el fin de verificar que cada aporte de código no rompe por completo alguna funcionalidad previamente implementada. Este conjunto de validación además debería verificar que el software no contiene vulnerabilidades en cuanto a la seguridad.
Importancia	Alta, ya que la seguridad es una parte fundamental del desarrollo del producto
Prioridad	Alta
Comentarios	Es deseable automatizar su ejecución en cada aporte de código al repositorio.

Tabla 1. Conjunto de validación del software

2	Procesamiento del lenguaje natural
Dependencias	<ul style="list-style-type: none"> - Con el fin de tener unos buenos resultados, es recomendable formar un conjunto de datos de entrenamiento de calidad. - El conjunto de datos de entrenamiento deberá proceder de varias fuentes para asegurar su heterogeneidad. - Se limitará el conjunto de caracteres a predecir para facilitar las predicciones al modelo. Esto significa que el conjunto de datos no los contendrá.
Descripción	El modelo deberá tener en cuenta que la frase predicha cumple con un mínimo de coherencia morfosintáctica, ya que al final es la oración que se va a reflejar en la red social.
Importancia	Alta, ya que la coherencia morfosintáctica puede ser un valor diferencial respecto a los competidores.
Prioridad	Alta
Comentarios	Una posibilidad podría ser que el CM pueda validar si la oración propuesta cumple con sus requisitos antes de publicar nada en sus redes sociales. Además, esto nos puede servir para señalar al modelo donde está cometiendo el error.

Tabla 2. Procesamiento del lenguaje natural

3	Realización del entrenamiento distribuido
Dependencias	Tener configurado previamente un clúster de equipos con el que poder trabajar de forma distribuida.
Descripción	El modelo deberá ser entrenado de forma distribuida con el fin de aprovechar el poder computacional de cada uno de los equipos que compone nuestro sistema.
Importancia	Media, ya que se puede individualizar el entrenamiento del modelo en un solo equipo, aunque tardaría más.
Prioridad	Media
Comentarios	Utilizad el sistema que más se adecúe a vuestro algoritmo de entrenamiento. Tened en cuenta que una buena configuración del entrenamiento puede ahorrar mucho tiempo en el futuro, ya que cada uno de los aportes de código deberá comprobar si la métrica de validación aumenta o disminuye. Si este entrenamiento es distribuido, os hará ahorrar tiempo en cada re-

Tabla 3. Realización del entrenamiento distribuido

4	Etiquetación de versiones y almacenamiento de código en repositorio
Dependencias	Tener definida previamente una política de despliegue acorde al versionado del código del repositorio.
Descripción	El almacenamiento del código deberá ser realizado en un repositorio compartido entre todos los integrantes del equipo de desarrollo. Además, se deben aprovechar las herramientas de automatización con el fin de validar cada uno de los aportes de código y versionar el software en base a su calidad. Una vez versionado el código, se deberá organizar su despliegue haciendo uso de diferentes políticas.
Importancia	Alta, ya que desplegar una versión del código que contenga vulnerabilidades puede ocasionar pérdidas para la empresa.
Prioridad	Alta
Comentarios	El proveedor de repositorio es indiferente, pero hay que asegurarse de que los minutos de ejecución del pipeline se adecúen a nuestros requisitos.

Tabla 4. Etiquetación de versiones y almacenamiento de código en repositorio

4.2 Modelo de proceso de negocio a implantar

En esta sección se abordará cual es el flujo del proceso de la aplicación, valorando tanto los actores implicados como la descripción de procesos.

4.2.1 Descripción de los actores de negocio a implantar

Mediante esta sección se definirá cada uno de los actores que formen parte del flujo de la aplicación:

1	Community Manager
Descripción	Este actor de negocio actual representa al cliente principal de la aplicación. Es quien desea generar un texto relacionado con las temáticas introducidas con el fin de publicarlas en la cuenta de la red social que gestiona.
Comentarios	Deberá valorar si el texto generado es de calidad o no.

Tabla 5. Community Manager

4.2.2 Descripción de procesos de negocio a implantar

A través de esta sección se definirán los procesos de negocio a implantar.

1	Flujo general de la aplicación
Descripción	<p>Al comienzo de la ejecución, el cliente, en este caso un CM de la red social de una empresa, selecciona las temáticas con las que quiere generar el texto.</p> <p>Una vez seleccionadas las categorías, el modelo de inteligencia artificial procedes a la generación de texto relacionado con las temáticas introducidas por el usuario.</p> <p>El usuario, tras comprobar si el texto generado tiene la calidad suficiente como para ser publicado, selecciona publicarlo o en caso contrario, finaliza la ejecución.</p> <p>Se le da la opción al usuario para editar el texto antes de publicarlo si lo desea.</p>
Actores	Community Manager
Comentarios	Todo este proceso se puede comprobar de una forma gráfica en la siguiente ilustración:

Tabla 6. Flujo general de la aplicación

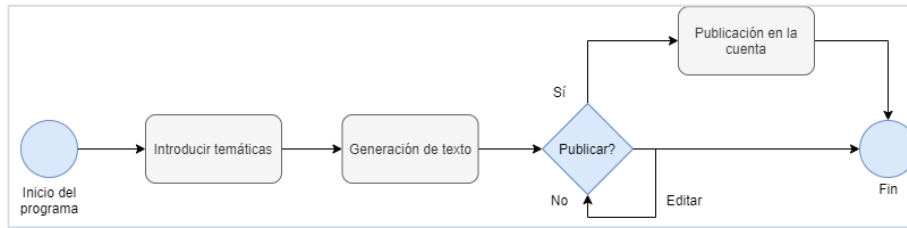


Ilustración 1. Diagrama de la aplicación

5. Clasificación de los subsistemas del sistema a desarrollar

Mediante esta sección se explicarán cada uno de los subsistemas que forman parte del sistema a desarrollar con el fin de facilitar la comprensión de estos.

Tanto los subsistemas como su funcionamiento se explicarán a través de las siguientes plantillas:

1	Sistema de ingesta y tratamiento de datos
Descripción	La labor principal de este subsistema es el de recolectar y tratar datos de diferentes fuentes con el fin de formar el dataset con el que posteriormente se procederá al entrenamiento. También se encarga de anonimizar los datos personales descartando la información no necesaria aplicando así un preproceso a los datos.
Importancia	Alta
Prioridad	Alta, es de los primeros subsistemas a implementar, ya que tanto los datos como su posterior tratamiento son la base de todo el sistema.
Comentarios	Cuanto más modularizado y ampliable sea mejor, ya que no sabemos cuántas fuentes de datos vamos a ser capaces de albergar en un futuro.

Tabla 7. Sistema de ingesta y tratamiento de datos

2	Aplicación web
Descripción	La labor principal de este subsistema es el de hacer de intermediario entre el cliente (Community Manager) y el modelo de inteligencia artificial.
Importancia	Alta, ya que al final es el único punto de contacto entre el cliente y el producto.
Prioridad	Media, ya que es más importante tener sentadas las bases del sistema de ingesta de datos o el entrenamiento del modelo.
Comentarios	-

Tabla 8. Aplicación web

3	Entrenamiento distribuido del modelo
Descripción	La labor principal de este subsistema es el de llevar a cabo el entrenamiento del modelo en base al conjunto de datos recogido por parte del subsistema de ingesta y tratamiento de datos. Cabe destacar, que el entrenamiento se debe realizar de forma distribuida con el fin de que el entrenamiento sea un proceso
Importancia	Alta, ya que habrá muchos reentrenamientos y cuanto más rápido sea el proceso, mejor.

Prioridad	Alta, al igual que el sistema de ingesta, es de los primeros subsistemas a implementar, ya que en cuanto se tenga disposición de estos, se debe empezar a crear un prototipo del futuro modelo.
Comentarios	-

Tabla 9. Entrenamiento distribuido del modelo

4	Almacenamiento del código y etiquetado de versiones para el despliegue
Descripción	La labor principal de este subsistema es el de almacenar el código que forma parte del proyecto de forma centralizada. Además, dependiendo de si el aporte de código ha pasado o no el conjunto de validación, será etiquetado de una forma u otra lo que condicionará su futuro despliegue en base a las políticas
Importancia	Alta, ya que es donde se albergará todo el código fuente de la aplicación.
Prioridad	Alta, es de los primeros subsistemas a implementar, ya que en cuanto se tenga un mínimo de código funcional, se debe analizar en base a los criterios acordados.
Comentarios	-

Tabla 10. Almacenamiento del código y etiquetado de versiones para el despliegue

5	Sistema de ficheros distribuido
Descripción	La labor principal de este es el de almacenar el conjunto de datos recolectado mediante el sistema de recolección y tratamiento en un sistema de ficheros distribuido con el fin de que las consultas a ella sean lo más rápidas posibles.
Importancia	Alta, ya que es donde se albergará todos los datos con los que se entrenará el modelo.
Prioridad	Alta, es de los primeros subsistemas a implementar, ya que en cuanto se tengan unos datos se tienen que poner a disposición del sistema de entrenamiento con el fin de comenzar a implementar el primer prototipo del
Comentarios	-

Tabla 11. Sistema de ficheros distribuido

6. Catálogo de requisitos del sistema a desarrollar

A través de esta sección se definirán de forma estructurada los requisitos del sistema a desarrollar.

6.1 Requisitos Generales del Sistema

En esta sección se detalla el requisito principal de la aplicación que consiste en lo siguiente:

1	Generación de texto
Descripción	El sistema deberá generar un texto relacionado con la temática introducida por el cliente en este caso un CM tal y como se ha especificado a lo largo de todo el documento.
Requisitos hijos	Para su correcto funcionamiento depende de todos los subsistemas mencionados en el apartado 5
Comentarios	-

Tabla 12. Generación de texto

6.2 Casos de uso del sistema

Al ser un proyecto de tamaño pequeño/moderado, la aplicación únicamente contempla un caso de uso y es el señalado en el apartado 4.2.2 de este mismo documento.

Además, por contextualizarlo, existen diferentes escenarios en los que la herramienta puede resultar útil en la introducción de este mismo documento.

Con el fin de detallar de forma más concreta ese flujo cabe destacar lo siguiente:

- No existe ningún tipo de precondition a la hora de hacer uso del sistema, es decir, no es necesario ningún tipo de registro previo.
- En cuanto al rendimiento, se espera que el texto sea generado en un máximo de 10 segundos.
- La postcondición está sujeta a el resultado que, del modelo, ya que, en base a él, el usuario elegirá si publicar el tweet o no.

6.3 Requisitos Funcionales del Sistema

Mediante este apartado se intentarán describir los requisitos funcionales del sistema que se hayan identificado a partir de los requisitos generales. Estos requisitos funcionales

detallan los servicios que prestará el sistema, en la forma en que reaccionará en determinados momentos.

6.3.1 Requisitos de información del sistema

A través de este apartado se especificarán los requisitos de almacenamiento de información que haya identificado, especificado mediante la siguiente plantilla:

1	Almacenamiento de los datos del sistema de ingesta
Dependencias	Sistema de ingesta de datos Sistema de almacenamiento distribuido Solo almacena datos no personales
Descripción	El sistema deberá almacenar la información correspondiente a las fuentes de datos definidas previamente.
Datos específicos	Se guardarán tanto tweets como posts de diferentes redes sociales anonimizados con el fin de entrenar el modelo de inteligencia artificial.
Comentarios	

Tabla 13. Almacenamiento de los datos del sistema de ingesta

6.3.2 Requisitos de Reglas de Negocio del Sistema

Estos requisitos deben especificar qué reglas de negocio debe respetar el sistema, evitando que se incumplan durante su funcionamiento. En el punto de desarrollo en el que estamos, no hemos contemplado ningún tipo de limitación. Aunque es posible que, si en un futuro el producto tiene éxito, podamos plantear modalidades de pago, siendo la gratuita la que contenga algún tipo de limitación en cuanto el uso del producto.

6.3.3 Requisitos de Conducta del Sistema

Estos requisitos deben especificar cualquier otro comportamiento deseado del sistema que no se haya especificado mediante los casos de uso del sistema, como generación de informes, funcionalidades transversales a varios casos de uso del sistema, etc.

1	Consulta de histórico de publicaciones
Descripción	El sistema deberá proporcionar al usuario un histórico de los textos generados y publicados a través de la aplicación.
Comentarios	No es necesario identificar quien ha sido el que haya publicado el tweet, ya que en esta fase del desarrollo no se contempla un registro/logueo de usuarios.

Tabla 14. Consulta de histórico de publicaciones

6.4 Requisitos No Funcionales del Sistema

Mediante este apartado se procederá a detallar diferentes requisitos no funcionales del sistema como pueden ser los de fiabilidad o usabilidad. Estos requisitos se diferencian de los anteriores ya que detallan propiedades del sistema como pueden ser: Rendimiento, seguridad o disponibilidad.

6.4.1 Requisitos de fiabilidad

Estos requisitos deberán establecer, de la manera más objetiva y medible posible, los niveles que debe cumplir el sistema a desarrollar en aspectos como recuperabilidad y tolerancia a fallos.

El sistema, como se detallará en posteriores documentos, será desplegado mediante un orquestador de contenedores de modo que, si en algún momento el sistema cae, la comprobación de disponibilidad avisará al orquestador y este se encargará de lanzar una réplica del servicio.

Mediante este sistema explicado, lo que se consigue es que el servicio se encuentre disponible la mayor parte del tiempo. Además, en caso de que haya que lanzar una actualización del software (bien por querer lanzar una nueva funcionalidad o posible solución de bugs) el orquestador ofrece funcionalidades para que este nuevo despliegue reemplace el anterior sin que el cliente se dé cuenta, lo cual facilita mucho lo que se viene conociendo como despliegue down-time.

6.4.2 Requisitos de Usabilidad

Estos requisitos deberán establecer, de la manera más objetiva y medible posible, los niveles que debe cumplir el sistema a desarrollar en aspectos como facilidad de aprendizaje, comprensión, operatividad y atraktividad.

En este caso se ha propuesto como criterio de aceptación que el sistema deberá permitir en el 100% de las veces que con un máximo de 5 clicks sea suficiente para llegar a la información deseada.

6.4.3 Requisitos de eficiencia

Estos requisitos deberán establecer, de la manera más objetiva y medible posible, los niveles que debe cumplir el sistema a desarrollar en aspectos como tiempo de respuesta.

En nuestro caso, uno de los requisitos es que el sistema deberá tener un tiempo máximo de respuesta de 5 segundos para cualquier operación de consulta/generación de texto.

6.4.4 Requisitos de Mantenibilidad

Estos requisitos deberán establecer, de la manera más objetiva y medible posible, los niveles que debe cumplir el sistema a desarrollar en aspectos como estabilidad, facilidad de análisis, facilidad de cambio, facilidad de pruebas.

En este caso se ha propuesto que absolutamente todo el código fuente que forme parte del repositorio debe pasar por el proceso de análisis automático con el fin de asegurar que no rompe por completo ninguna funcionalidad clave del sistema.

Además, esta misma automatización se encarga de formatear el código en base a una convención configurada en el primer aporte de código.

En caso de querer conocer a fondo en que consiste este proceso de análisis/despliegue/etiquetado del código fuente es recomendable proceder a la lectura del documento de desarrollo seguro.

6.4.5 Requisitos de Portabilidad

Estos requisitos deberán establecer, de la manera más objetiva y medible posible, los niveles que debe cumplir el sistema a desarrollar en aspectos relacionados con la escalabilidad: capacidad de instalación, capacidad de sustitución, adaptabilidad, coexistencia, compatibilidad con hardware o software, etc.

En nuestro caso, la aplicación debe ser contenedorizada de modo que la portabilidad sea máxima. La aplicación debe tener una alta capacidad de instalación, compatibilidad y despliegue.

Además, el hecho de contenedorizarla provoca que, mediante el orquestador, tanto el despliegue inicial como los posteriores (cuando se actualice la aplicación) resulten muy sencillos.

Por otro lado, también provoca que la capacidad de tolerancia a fallos sea muy alta ya que el mismo el orquestador se encarga de que el servicio esté disponible durante gran parte del tiempo (por no decir 24/7).

6.4.6 Requisitos de seguridad

Estos requisitos deberán establecer, de la manera más objetiva y medible posible, los niveles que debe cumplir el sistema a desarrollar en aspectos como accesos al sistema, identificación y autenticación, protección de datos y privacidad.

Por esta parte, el sistema deberá ser capaz de hacer frente a cualquier tipo de ataque basado en inputs malformados. Esta capacidad de afrontamiento será evaluada por el análisis de código/conjunto de validación que se ejecutará cada vez que un aporte de código sea subido al repositorio y un fuzzing manual sobre el entorno de staging. Este apartado se puede observar con más detalle en el documento de política de desarrollo seguro.

Por otro lado, en cuanto al repositorio de código el cual es un activo del sistema, su acceso es restringido por unas credenciales a las cuales solo tiene acceso el equipo de desarrollo.

6.5 Restricciones Técnicas del Sistema

Esta sección debe contener las restricciones técnicas que se imponen al sistema software a desarrollar (tecnología a usar, protocolos de comunicaciones, compatibilidad con navegadores, etc.)

1	Compatibilidad con los navegadores actuales
Descripción	El sistema deberá respetar la siguiente restricción técnica: Deberá ser compatible con gran parte del abanico de navegadores disponible para el público. Al menos para Chrome, Microsoft Edge y Firefox.
Importancia	Alta, ya que cuanto mayor variedad de navegadores soportemos, mayor será el atractivo de cara al usuario.
Comentarios	-

Tabla 15. Compatibilidad con los navegadores actuales

6.6 Requisitos de integración de sistema

Estos requisitos deben identificar aquellos servicios disponibles en el entorno tecnológico de producción o componentes software (por ejemplo, bibliotecas enlazables) cuya funcionalidad sea relevante para el sistema a desarrollar y deban ser utilizados por el mismo.

1	Integración con las APIs de terceros
Descripción	El sistema deberá utilizar el recolector de datos para aquellos aspectos relacionados con generar textos relacionados con los proporcionados por el usuario y que consecuentes con la actualidad.
Importancia	Media, ya que, sin la obtención de nuevos datos, puede funcionar con los
Comentarios	-

Tabla 16. Integración con las APIs de terceros