



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE INGENIERÍA DE SISTEMAS

**DESARROLLO DE UNA PLATAFORMA WEB PARA EL ANÁLISIS, CLASIFICACIÓN
Y VISUALIZACIÓN DE LOS COMENTARIOS EN LA RED SOCIAL TWITTER DE LOS
USUARIOS DE LAS EMPRESAS PÚBLICAS DEL CANTÓN CUENCA**

Trabajo de titulación previo a la obtención del
título de Ingeniero de Sistemas

AUTOR: CARLOS HOMERO MOROCHO MEDINA

TUTOR: ING. GABRIEL ALEJANDRO LEÓN PAREDES, PHD.

Cuenca - Ecuador

2022

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Yo, Carlos Homero Morocho Medina con documento de identificación N° 0302699798, manifiesto que:

Soy el autor y responsable del presente trabajo; y, autorizo a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 28 de agosto del 2022

Atentamente,

A handwritten signature in blue ink, consisting of a circular scribble followed by a long horizontal line extending to the right.

Carlos Homero Morocho Medina

0302699798

**CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE
TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA**

Yo, Carlos Homero Morocho Medina con documento de identificación N° 0302699798, expreso mi voluntad y por medio del presente documento cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del Proyecto técnico: “Desarrollo de una plataforma web para el análisis, clasificación y visualización de los comentarios en la red social Twitter de los usuarios de las empresas públicas del cantón Cuenca”, el cual ha sido desarrollado para optar por el título de: Ingeniero de Sistemas, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribo este documento en el momento que hago la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 28 de agosto del 2022

Atentamente,



Carlos Homero Morocho Medina

0302699798

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Gabriel Alejandro León Paredes con documento de identificación N° 0103652186, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **DESARROLLO DE UNA PLATAFORMA WEB PARA EL ANÁLISIS, CLASIFICACIÓN Y VISUALIZACIÓN DE LOS COMENTARIOS EN LA RED SOCIAL TWITTER DE LOS USUARIOS DE LAS EMPRESAS PÚBLICAS DEL CANTÓN CUENCA**, realizado por Carlos Homero Morocho Medina con documento de identificación N° 0302699798, obteniendo como resultado final el trabajo de titulación bajo la opción Proyecto técnico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 28 de agosto del 2022

Atentamente,



Ing. Gabriel Alejandro León Paredes, PhD.

0103652186

DEDICATORIA

El presente trabajo se lo quiero dedicar a mis padres que sin duda fueron una parte esencial en toda mi formación tanto profesional como personal, es increíble como la guía de otra persona puede influir completamente en tu futuro, espero que todo este proceso de aprendizaje y esfuerzo sea recompensado de alguna manera para poder retribuir el tiempo que se ha designado en completar esta carrea que recién empieza.

AGRADECIMIENTO

Quiero empezar agradeciendo a mis padres porque fueron el ejemplo más claro del que eh podido evidenciar que con esfuerzo todo se consigue, agradezco el apoyo económico y emocional que me han brindado todo este tiempo, probablemente sin ellos esto no sería posible, también quisiera agradecer a ciertas personas que formaron parte de esta travesía y que de alguna manera me ayudaron a mejorar mis habilidades, sin duda es algo que nunca podre pagar, gracias.

RESUMEN

Actualmente hemos podido evidenciar el gran impacto que han tenido las redes sociales en nuestro entorno, tanto en lo personal como en lo social pues solo basta mirar a nuestro alrededor para darnos cuenta que todos estamos de alguna u otra forma conectados a través de estas redes sociales, sin duda estamos formando parte de las bases del "mundo moderno" en toda su envergadura, es tal el proceso computacional que hemos alcanzado que incluso se han desarrollado tecnologías como la Inteligencia Artificial (IA) que permiten realizar estas interacciones sin la dependencia de una persona.

Técnicas como el Procesamiento de Lenguaje Natural (PLN) derivadas de implementación de la IA han permitido crear una nueva forma de interacción con las plataformas virtuales dotándolas de un grado mayor de utilidad y funcionalidad, tal es el caso de los Asistentes inteligentes y la Clasificación de texto que es el tema principal de este proyecto, el cual busca enfocarse en el ámbito de las opiniones que publican los usuarios en la red social Twitter para poder convertir esa información en un verdadero set de datos estadísticos para la toma de decisiones dentro de una empresa.

Solo en Cuenca existe un buen número de empresas públicas que proveen de diferentes servicios a toda la comunidad y que manejan su respectiva cuenta en la plataforma Twitter, pero también existen un centenar de cuentas dedicadas a receptar las quejas de los ciudadanos con el fin de que su petición sea resuelta por las entidades correspondientes de la manera más rápida posible, lo que no siempre sucede, por ende el hecho de que dichas empresas cuenten con un administrador de sus redes sociales no implica que este pueda tener una visión clara de los problemas que van reportando los usuarios diariamente, y de ahí se establece el objetivo de este proyecto que es implementar técnicas de PLN que permitirá analizar los comentarios de los usuarios en tiempo real y determinar según el sesgo de negocio de la empresa a que área de servicio pertenece el tweet con el fin de poder visualizar la problemática real de la empresa a través de graficas e información centralizada.

ABSTRACT

Currently we have been able to demonstrate the great impact that social networks have had on our environment, both personally and socially, because it is enough to look around us to realize that we are all connected in one way or another through these social networks. definitely we are forming part of the foundations of the "modern world" in all its magnitude, the computational process that we have achieved is so advanced that technologies such as Artificial Intelligence (AI) have even been developed that allow these interactions to be carried out without depending on a person.

Techniques such as Natural Language Processing (NLP) derived from the implementation of AI have made it possible to create a new way of interacting with virtual platforms, providing them with a greater degree of utility and functionality, such as the case of Intelligent Assistants and the Text Classification that is the main theme of this project, which seeks to focus on the field of opinions published by users on the social network Twitter in order to convert that information into a true set of statistical data for decision making within a company.

Only In Cuenca, there are a good number of public companies that provide different services to the entire community and that manage their respective accounts on the Twitter platform, but there are also a hundred accounts dedicated to receiving complaints from citizens so that your request is resolved by the corresponding entities as quickly as possible, which does not always happen, therefore the fact that these companies have an administrator of their social networks does not imply that they can have a clear vision of the problems that users report daily, and from there the objective of this project is established, which is to implement PLN techniques that will allow analyzing user comments in real time and determining, according to the business bias of the company, to which service area the user belongs. tweet in order to be able to visualize the real problems of the company through graphics and centralized information.

ÍNDICE

I INTRODUCCION	8
1.1 PROBLEMA	10
1.1.1 Antecedentes	10
1.1.2 Importancia y alcances	10
1.1.3 Delimitación	11
1.2 OBJETIVOS	12
1.2.1 General	12
1.2.2 Específicos	12
II FUNDAMENTOS TEORICOS	13
2.1 Inteligencia Artificial	13
2.1.1 Definición	13
2.2 Redes Neuronales	13
2.2.1 Definición	13
2.2.2 Tipos de Redes Neuronales	14
2.3 Procesamiento del Lenguaje Natural	15
2.3.1 Definición	15
2.3.2 Principales aplicaciones de PLN.	16
2.4 Analisis de Sentimientos	16
2.4.1 Definición	16
2.4.2 Aplicaciones	16
2.5 Aprendizaje Automatico	17
2.5.1 Definición	17
2.5.2 Aprendizaje Supervisado	17
2.5.3 Aprendizaje No supervisado	17

2.5.4	Aprendizaje Semi-Supervisado	17
2.6	Redes Sociales	18
2.6.1	Definición	18
2.7	Procesamiento de Datos	19
2.7.1	Definición	19
2.7.2	Visualización de datos	19
2.8	Servicios web	20
2.8.1	Definición	20
2.8.2	Tecnología REST	20
2.9	Streaming	20
2.9.1	Streaming de tweets	21
2.10	Metodología Scrum	21
2.11	Zero-shot Learnig	22
2.11.1	Según la fase de entrenamiento	22
2.11.2	Según la fase de prueba	22
2.12	Trabajos relacionados	23
III MARCO METODOLOGICO		24
3.1	Levantamiento de Requerimientos	25
3.1.1	Personal Involucrado	25
3.1.2	Requisitos Funcionales	25
3.1.3	Requisitos No Funcionales	40
3.2	Arquitectura del sistema	42
3.3	Tecnologías del proyecto	44
3.4	Integración con Twitter	44
3.4.1	Reglas de stream para obtener los Tweets	45
3.4.2	Proceso de Streaming de Tweets	46
3.5	Clasificación de tweets	50
3.5.1	Eliminación de ruido del texto	51

3.5.2	Definición del proceso de clasificación	54
3.5.3	Aplicación del modelo de clasificación	59
3.6	Administración de información	63
3.6.1	Descripción de APIs Rest Full	64
3.6.2	Almacenamiento de información	69
3.6.3	Presentación de información	73
IV	ANÁLISIS DE RESULTADOS	77
4.1	Resultados de clasificación clase "negativo"	77
4.2	Resultados de clasificación clase "servicios de agua"	82
4.3	Resultados de clasificación clase "servicios de internet"	86
4.4	Resultados de clasificación clase "servicios de telefonía"	90
V	CRONOGRAMA DE ACTIVIDADES	93
VI	PRESUPUESTO	96
VII	CONCLUSIONES	97
VIII	RECOMENDACIONES	98
Anexos		101
A	Script de python para el reproceso de tweets	102
B	Interfaces de Usuario	106
2.1	Aplicación Web	106
2.1.1	Inicio de sesión	106
2.1.2	Dashboard de tweets en vivo	106
2.1.3	Detalle del tweet y su clasificación	107
2.1.4	Vista de interacción de usuarios	108
2.1.5	Vista de tweets por tipo	109

2.1.6 Vista de tweets con su clasificacion	109
2.1.7 Visualización de configuracion	111

GLOSARIO

Algoritmo: Conjunto ordenado de operaciones sistemáticas que permiten hacer un cálculo y hallar la solución de un tipo de problemas.

RNA: Artificial Neural Network o Red Neuronal Artificial.

GIHP4C: Grupo de Investigación en Cloud Computing, Smart Cities & High Performance Computing.

Red Social: Medio de comunicación en línea utilizado por la comunidad en Internet para relacionarse entre sí.

Machine Learning: Aprendizaje de la máquina, mediante algoritmos de análisis y procesamiento de información.

Dataset: Conjunto de datos tabulados en cualquier sistema de almacenamiento de datos estructurados.

Procesamiento de Lenguaje Natural (PLN): Es un área de estudio de la informática, que se encarga de analizar la comunicación entre las computadoras y el lenguaje humano.

Análisis Semántico Latente: Es un algoritmo el cual nos permite saber el número de veces que un término o frase de consulta ingresada se repite dentro de una cantidad de textos que tenemos como datos de ingreso.

API, Interfaz de programación de aplicaciones: Es un conjunto de definiciones y protocolos que se utiliza para desarrollar e integrar el software de las aplicaciones.

Twitter: Es un servicio de microblogueo, con sede en San Francisco, California, Estados Unidos, con filiales en San Antonio y Boston.

RESTful: Es un protocolo estándar que define cómo dos objetos en diferentes procesos pueden comunicarse por medio de intercambio de datos JSON.

Capítulo I

INTRODUCCION

Posiblemente el mundo atraviesa por una época donde un canal de YouTube podría tener más visualizaciones que un canal de televisión, una página en Facebook podría tener más interacciones que una locación de uso público, y una publicación en Twitter podría tener más alcance que un periódico local, el hecho es que el mundo cambió, específicamente en el campo de la comunicación pues vemos que ha sido la precursora de una las revoluciones digitales más notorias de los últimos años, y que solo en Ecuador en lo que va del año 2022 hay 14 Millones de usuarios activos interactuando en estas plataformas sociales ubicando a la provincia del Azuay en el tercer lugar dentro del territorio nacional con un total del 12% de personas conectadas diariamente. [Juan Pablo Del Alcázar Ponce \(2022\)](#)

Por tal razón vale la pena mencionar un estudio donde se describe la importancia y el gran impacto que han tenido las redes sociales para las microempresas de la ciudad de Cuenca para darnos cuenta de que bien utilizadas se puede lograr el crecimiento esperado de un negocio e incluso alcanzar el mismo nivel de promoción que se maneja en las grandes empresas. [MAR-QUIDIA BALAREZO \(2015\)](#), sin duda es un claro ejemplo a nivel microeconómico de que los datos digitales siempre son importantes, sin mencionar el nivel macroeconómico donde hemos podido ser partícipes del reconocimiento mundial que han alcanzado ciertas personas (como Luisito Comunica) o empresas (como EnchufeTv) dentro de una red social.

Como todo buen caso de éxito que involucra una gran cantidad de personas y una revolución social de por medio, esto ha desencadenado una serie de problemáticas sociales que estamos viviendo actualmente como es el generar una cultura de libre opinión y lo fácil que es publicarla en una red social, pero del cual hablaremos más a fondo en la siguiente sección ya que es un punto determinante para plantear la problemática de este proyecto, que de alguna u otra manera solo afirma la importancia que actualmente tienen estas plataformas sociales.

Con este contexto las redes sociales generan constantemente información tan valiosa y útil que la mayoría de las organizaciones actualmente aplican diferentes técnicas digitales para el procesamiento inteligente de toda esa información con el único objetivo de generar datos con mayor valor que ayuden a mejorar el estado actual de una empresa [Power Data \(2022\)](#), lo cierto es que todos lo hemos podido evidenciar en algún punto de nuestra vida ya sea viendo un video, viendo un anuncio, viendo una película o simplemente buscando algo en internet, el punto es que son datos tan relevantes que no es coincidencia que cualquier cosa que hagamos en línea este completamente ligado con nuestras preferencias personales, simplemente estamos siendo partícipes de lo que consiste la transformación digital en toda su envergadura.

Plataformas como Twitter donde según la página [Internet Live Stats \(2022\)](#) se publican diariamente más de 500 Millones de tweets permiten realizar integraciones de una forma muy sencilla mediante el uso de sus APIs Rest Full públicas creadas específicamente para desarrolladores que aplican sus conocimientos técnicos en cualquier ámbito profesional relacionado con el análisis de la información, posibilitando desarrollar importantes proyectos tanto profesionales como educativos respecto a la implementación de procesos como la Minería de Datos, Big Data, Procesamiento de Lenguaje Natural, entre otras técnicas enfocadas al procesamiento de información, por lo que, si tomamos en cuenta esta fácil integración se puede aprovechar la información generada por las interacciones de los usuarios en las redes sociales (específicamente en Twitter) con el único objetivo de identificar problemas y mejorar ciertos aspectos de los servicios que brindan las empresas del sector público.

Este proyecto parte de la premisa de que mediante la representación gráfica de la clasificación automática de los tweets que se realizan en la red social Twitter sobre alguna empresa pública en particular se pueda identificar las posibles quejas sobre los diferentes problemas de servicios que están sucediendo en la ciudad de Cuenca, esto mediante el Procesamiento de Lenguaje Natural (PLN) el cual permitirá clasificar los temas de interés para presentarlos de forma centralizada sobre una plataforma web y que las entidades correspondientes puedan dar una pronta solución.

1.1 PROBLEMA

1.1.1 Antecedentes

Hace no mucho tiempo las empresas solían dejar un buzón con un mensaje que decía “Todas tus recomendaciones serán bienvenidas” o algo similar, esto con el fin de poder obtener algún tipo de retroalimentación de sus clientes acerca de sus servicios, sin éxito obviamente el nuevo sistema aunque mejorado sigue teniendo el mismo propósito gracias a las redes sociales, excepto que este último no tiene el aspecto de un barril cerrado con candadito rojo ni mucho menos tiene papelitos, las redes sociales tienen gente tras un teclado que lastimosamente no interactúan en persona pero si lo hacen defendiendo una postura, talvez ocultos o talvez no de la manera correcta, pero sin duda hacen que lanzar una opinión sea mucho más fácil y directo.

Con este principio podemos decir que las empresas públicas no están absueltas de recibir quejas diariamente de los usuarios en redes sociales esto debido al rol que tienen en nuestra sociedad, pues somos nosotros mismos los encargados de hacer que estas entidades cumplan con su trabajo, como toda empresa en la actualidad estas también tienen presencia en línea (Internet) específicamente en la plataforma Twitter que con esta premisa se ha creado un ambiente político bajo el nombre de esta plataforma pues vemos que hasta los mismos funcionarios públicos optan por realizar comunicados oficiales a través de esta red social. [Agenda Pública \(2020\)](#)

1.1.2 Importancia y alcances

Se ha podido evidenciar la gran cantidad de quejas que reportan los usuarios diariamente en las diferentes cuentas oficiales de las empresas públicas de la ciudad de Cuenca que por la falta de respuesta hasta se han creado cuentas destinadas a receptor quejas de los ciudadanos en general para poder tener un mayor alcance con su petición que no siempre es tomada en cuenta, tal es el caso de @QuejasCuenca que como parte de nuestra investigación se determinó que los 17.7K tweets publicados hasta la fecha son en su mayoría criticas directas a la gestión de servicios de las entidades públicas sin profundizar en la cantidad de usuarios que mencionan

esta cuenta cuando tienen algún tipo de reclamo.

Con el avance de la Inteligencia Artificial (IA) en los últimos años se ha tenido un impacto bastante grande en el desarrollo de tecnologías que involucran redes neuronales artificiales para el aprendizaje de máquina, un claro ejemplo de esto es el área del Procesamiento de Lenguaje Natural (PLN) el cual permite a una máquina analizar el contexto de un texto escrito por un humano con el fin de identificar características como el estado de ánimo con la que se escribió el texto, el tema principal del texto o simplemente el significado del texto.

Por lo que, el uso de esta tecnología integrada con las redes sociales permite crear funcionalidades con un nivel superior de análisis que ayuda a identificar problemas significativos sobre la opinión que realizan constantemente los usuarios en las redes sociales como es el caso de la plataforma Twitter, es por esto que, proponemos una solución que permita clasificar los comentarios que se hagan en las cuentas oficiales de las instituciones públicas de la ciudad de Cuenca como Etapa, CentroSur, EMOV, entre otras, para poder mostrar sobre un panel de control la información de la interacción de forma gráfica, clasificada y en tiempo real, además integrando un asistente virtual que permite al administrador de la cuenta realizar preguntas puntuales sobre los eventos sucedidos con las interacciones de los usuarios, con el fin de poder convertir toda esa información analizada en un verdadero set de datos estadísticos que ayuden de alguna manera en el proceso de toma de decisiones de la institución pública.

1.1.3 Delimitación

La herramienta tiene como grupo objetivo ayudar a las empresas públicas de la ciudad de Cuenca a identificar de forma automática cual es el problema real de sus servicios y que de alguna manera se pueda medir la conformidad que tienen sus usuarios, además de que está desplegada en un servidor del Grupo de Investigación en Cloud Computing Smart Cities & High Performance Computing (GIHP4), por lo que la empresa en cuestión podrá visualizar las incidencias que se reporten cuando se haya vinculado su cuenta de Twitter en la plataforma.

1.2 OBJETIVOS

1.2.1 General

Diseñar e implementar una plataforma web basada en servicios para el análisis, clasificación y visualización de los comentarios que se realizan en la red social Twitter de las instituciones públicas proveedoras de servicios del cantón Cuenca.

1.2.2 Específicos

- Realizar un estudio del estado del arte sobre el procesamiento de lenguaje natural, la clasificación de tópicos, análisis de sentimientos en redes sociales y desarrollo de asistentes virtuales.
- Implementar algoritmos de Procesamiento de Lenguaje Natural (PLN) enfocado al análisis, clasificación y visualización de comentarios en redes sociales utilizando como datos de entrada comentarios de la red social Twitter a través de su API.
- Desarrollar y desplegar una aplicación web que gestione los comentarios de los usuarios de Twitter en tiempo real según la cuenta de la institución que inicie sesión en la plataforma.
- Desarrollar e integrar un chatbot interactivo sobre la aplicación web que permita resolver preguntas puntuales de los administradores de la cuenta, usando la información obtenida en el procesamiento de los comentarios.
- Realizar la validación de la aplicación.

Capítulo II

FUNDAMENTOS TEORICOS

2.1 Inteligencia Artificial

2.1.1 Definición

“La Inteligencia Artificial se puede definirse como sistemas informáticos que han sido diseñados para interactuar con el mundo mediante capacidades que solemos considerar humanas” [Giannini \(2021\)](#). Podemos pensar en la IA como la ciencia computacional y de ingeniería con el fin de crear maquinas que se asemejen a los humanos, es decir que puedan razonar, aprender, reconocer y realizar tareas de manera autónoma de forma tal como lo harían los humanos. Un ejemplo claro es el ajedrez, gracias a IBM que construyeron el primer ordenador (Deep Blue) con el objetivo de aprender a jugar ajedrez, este proyecto tuvo éxito en 1997 ganando al campeón de ajedrez Garry Kasparov.

2.2 Redes Neuronales

2.2.1 Definición

“Una red neuronal artificial (RNA), también conocida como red neuronal está formada por la interconexión de muchas unidades de procesamiento llamadas neuronas, formadas por bloques no lineales distribuidas en toda la red neuronal” [Juan Vorobioff \(2022\)](#). Las redes neuronales artificiales son modelos computacionales basado en el proceso que realiza el cerebro humano en interpretar la información y así poder encontrar soluciones a diversos problemas.

2.2.2 Tipos de Redes Neuronales

- **Red neuronal Monocapa – Perceptrón simple:** El perceptrón simple es el modelo neuronal más simple considerado como un modelo unidireccional, está compuesto por dos capas, la primera es de entrada (n neuronas de entrada) y la segunda de salida (m neuronas de salida), utiliza señales binarias (0, 1) y la función de activación tipo signo para la transmisión de la información en ambas capas.

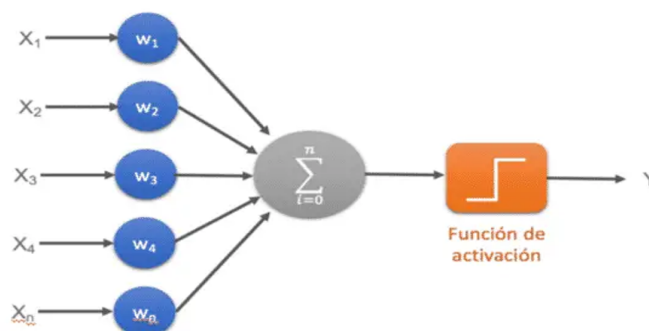


Figura 2.1: Red Neuronal Monocapa Perceptrón Simple
?

- **Red neuronal Multicapa – Perceptrón multicapa:** Esta red neuronal está compuesta por una capa de entrada, una capa de salida y capas ocultas. A diferencia de la monocapa, esta se compone de un conjunto de capas intermedias entre la capa de entrada y la de salida.

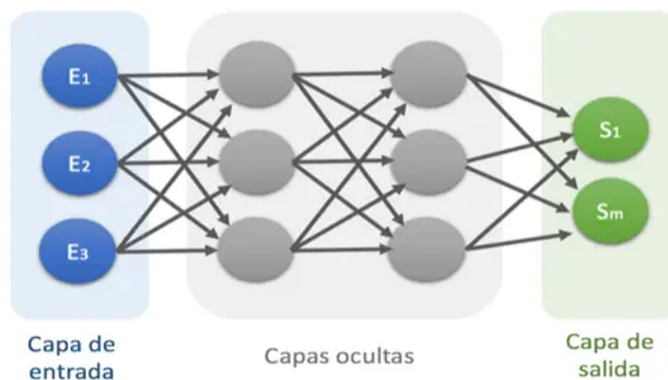


Figura 2.2: Red neuronal Multicapa Perceptrón multicapa
?

- **Red neuronal Convolutacional (CNN):** “Una red neuronal convolutacional es un tipo de red multicapa que consta de varias capas convolutacionales y de pooling (submuestreo) alternadas, y al final tiene una serie de capas full-connected como una red perceptrón multicapa” [Suárez \(2017\)](#). Una red neuronal convolutacional se basa en como la corteza visual del ser humano reconoce alguna característica (imágenes) de su entorno, se puede percibir como filtros en el que una imagen da determinación más alta.

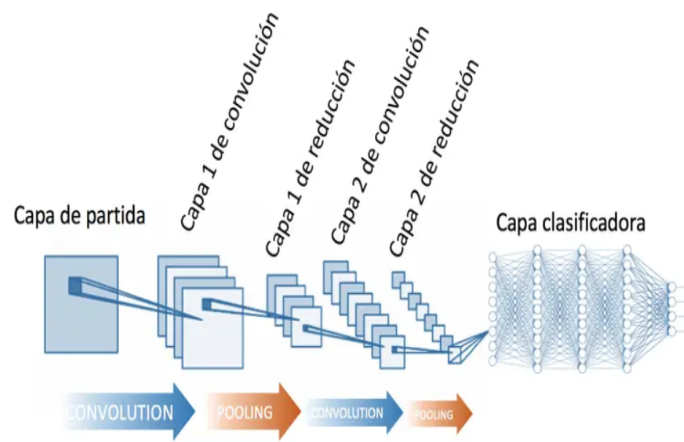


Figura 2.3: Ejemplo de Red neuronal Convolutacional (CNN)
?

2.3 Procesamiento del Lenguaje Natural

2.3.1 Definición

El PLN ayuda a la comprensión del lenguaje de las máquinas para cumplir órdenes o peticiones por voz, texto o ambos con el objetivo de asemejar el análisis que las personas realizan al momento de comprender un lenguaje. “El procesamiento del lenguaje natural consiste en la habilidad de una máquina para procesar información comunicada mediante el uso del lenguaje natural. Crean modelos computacionales del lenguaje suficientemente detallados que permitan escribir programas informativos que realicen distintas órdenes o peticiones donde interviene el lenguaje natural” [Dorys Moreira \(2021\)](#) La utilización de PLN tiene una ventaja, que es la facilidad de analizar grandes cantidades de datos en cortos periodos de tiempo.

2.3.2 Principales aplicaciones de PLN.

- Reconocimiento de patrones de lenguaje.
- Recuperación de información.
- Traducciones automáticas de idiomas.
- Detección de sentimientos y emociones.

2.4 Análisis de Sentimientos

2.4.1 Definición

El análisis de sentimiento es el proceso de determinar si el lenguaje refleja sentimientos positivos, negativos o neutrales combinando técnicas de Machine Learning para ayudar a las grandes empresas a medir la opinión pública, realizar estudios de mercado matizados, supervisar la marca y la reputación de los productos, y comprender las experiencias de los clientes todo esto a través de un algoritmo que hace más fácil la analizar de grandes volúmenes de texto, procesarlos y valorarlos.

2.4.2 Aplicaciones

- Análisis del sentimiento de los documentos de texto.
- Monitorear las opiniones generales sobre los cambios en las políticas y los anuncios de compañías.
- Identificar picos y valles en el sentimiento general de marca o cambios en las actitudes sobre los productos o servicios.
- Recopilar datos sobre las tendencias de los sentimientos y los utilizan para comprender cómo la reputación de su marca está cambiando con el tiempo.

2.5 Aprendizaje Automatico

2.5.1 Definición

Este modelo tecnológico permite a los programas aprender e implementar mejoras por sí mismo según las experiencias, así como los humanos, mediante la recolección de datos un sistema computacional puede analizar e identificar patrones. En la actualidad tenemos muchos ejemplos de la aplicación de esta tecnología como las redes sociales entre otros.

2.5.2 Aprendizaje Supervisado

Es de aprendizaje que se basa en tipo de aprendizaje que se basa en descubrir la relación existente variable de entrada y variables de salida, es decir enseñarles a estos algoritmos cuál es el resultado que quieres obtener para un determinado valor y será capaz de dar un resultado correcto incluso cuando se muestren valores que no haya visto antes.

Si se muestra a uno de estos algoritmos suficientes datos de entrada y de salida y si existe una relación será capaz de aprenderla.

2.5.3 Aprendizaje No supervisado

Es el paradigma que consigue producir conocimientos únicamente de los datos que se proporcionan como entradas sin necesidad en ningún momento de explicarle al sistema qué resultado queremos obtener.

Básicamente el aprendizaje no supervisado es aquel con el objetivo de la máquina aprenda según los patrones y sucesos de su entorno como los humanos. La ventaja de este método es que puede analizar grandes cantidades de daros y procesarlos en un lapso pequeño de tiempo.

2.5.4 Aprendizaje Semi-Supervisado

El aprendizaje Semi-Supervisado es un lenguaje híbrido entre el aprendizaje supervisado y no supervisado, utiliza datos de etiquetados y no etiquetados para el entrenamiento de datos. Se

usa para obtener datos más certeros entre datos etiquetados y datos no etiquetados con el fin de crear modelos predictivos mejorados.

2.6 Redes Sociales

2.6.1 Definición

Las redes sociales se han convertido en la solución para muchos emprendedores, gracias a su facilidad que ofrecen las redes sociales a sus usuarios de compartir intereses en común, como también entablar relación social como laborales. Hoy en día podemos ver millones de ejemplos de personas que emprenden y que han sacado a flote su negociación gracias a las redes sociales y sin hablar de la cantidad de personas que encuentras en una red social, esto último puede ser positivo como negativo, ya que no es noticia que las redes sociales tienen sus puntos de inseguridad que las personas han aprovechado para hacer sus actos ilícitos. Sin embargo, muchas personas recurren a las redes sociales para conocer personas con intereses en común, informarse o simplemente por entretenimiento.

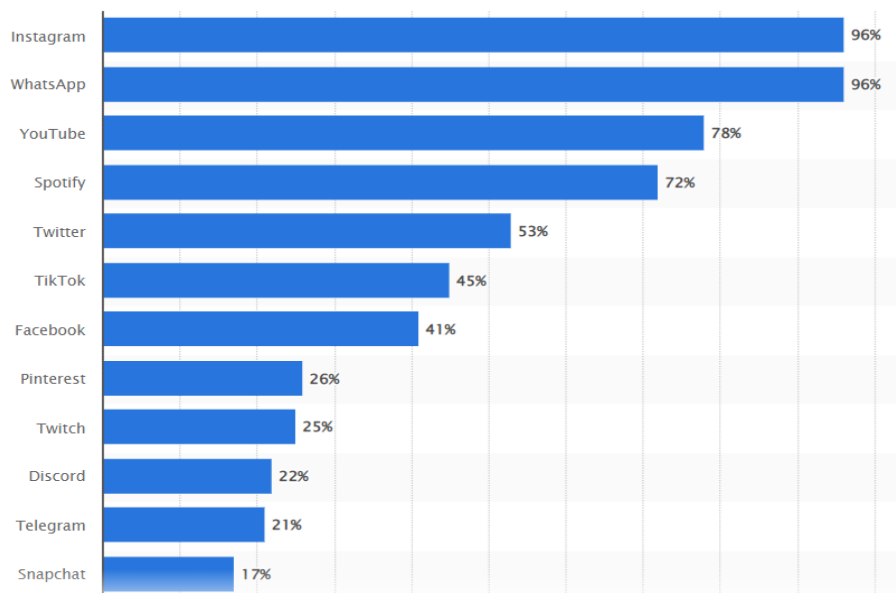


Figura 2.4: Gráfico de Redes sociales más usadas (2022)

Diego Calvo (2019a)

2.7 Procesamiento de Datos

2.7.1 Definición

Es aquel conjunto de medios que aplicamos para llegar a la esencia de una información que se está buscando, El resultado de la búsqueda surge de los datos recolectados que por sí solos son extractos sin sentido, pero el objetivo del procesamiento de datos es mejorar la interpretación de los datos y evitar realizar pasos o pruebas innecesarias como se va a organizar los elementos obtenidos de los datos los cuales son recopilados en la investigación. "El procesamiento de los datos se refiere a todo el proceso que sigue un investigador desde la recolección de datos, hasta la presentación de estos en forma resumida. Tiene básicamente tres etapas: recolección y entrada, procesamiento y presentación" [Hernández R \(2006\)](#) , para así poder reconocer la información que se busca.

2.7.2 Visualización de datos

La visualización de datos precisa de un conjunto de acciones profesionales para componer figuras por medio de imágenes originales, abstractas o figurativas, a partir de unos datos y con ayuda tecnográfica" [Sancho \(2014\)](#). La visualización de datos ayuda a la comprensión de la información a través de la utilización de patrones en formato gráfico que permiten comunicar exactamente lo que representa/significa esa información, es decir, pasar de imágenes saturadas (no comprensible), tabuladas a imágenes más claras, limpias por ende más entendibles para los usuarios.

La visualización de datos se refiere a la representación gráfica de la información y es una herramienta eficaz para comunicar la información. A través de la visualización podemos mostrar aspectos claves de datos complejos y despertar el interés de los usuarios.

2.8 Servicios web

2.8.1 Definición

Un servicio web permite el uso de software de aplicación a través de una red, donde los datos de una aplicación se intercambian a través de consultas de URL. La entrada se procesa y la salida se da a través de HTTP como los datos se intercambian en forma de XML se conoce como Protocolo de Acceso a Objetos Simples (SOAP) o JSON se conoce como Transferencia de Estado Representacional (REST)." [MAMANI \(2022\)](#). Es decir, que son tecnologías que combinan una serie de protocolos y estándares que ayudan a las aplicaciones a intercambiar información a través de la red, esto significa que la comunicación es de máquina a máquina. Estas máquinas tienen la capacidad de soportar todas las interacciones entre ellas.

2.8.2 Tecnología REST

" Es una arquitectura de software para sistemas distribuidos fácil de entender, soportado sobre el protocolo HTTP y sus métodos básicos de mantenimiento, permitiendo que sea fácil codificar y documentar aplicaciones usando servicios REST." [MAMANI \(2022\)](#) Tecnología REST (Transferencia de estados representacional) es una tecnología que ayuda a conectar varias interfaces a través de protocolos HTTP con el fin de obtener y generar datos, realizar operaciones de cualquier tipo y posteriormente devolverlos en formato XML o JSON. REST viene a ser la solución a la problemática que añadía SOAP al momento de diseñar las API, es decir, que REST soluciona el problema haciendo más fácil el diseño de las APIs.

2.9 Streaming

Streaming (significa transmisión) es un contenido (de audio, video y texto) en internet transmitido en tiempo real, es decir, que se envía datos de forma continua para que los usuarios disfruten de los archivos multimedia de su elección como Netflix, Disney + entre otros que nos

permiten ver películas o series, o como Spotify que podemos escuchar música o podcasts, es importante recalcar que el usuario no necesita descargar el archivo de forma completa sino que al momento de reproducir el contenido se enviaran (en formato paquetes) en secuencia.

2.9.1 Streaming de tweets

La plataforma Twitter ha desarrollado un conjunto de APIs Rest Full que permiten a los desarrolladores poder integrar plataformas externas con datos que se generan en esta red social, la forma tradicional de consumir un servicio generalmente trata de petición y respuesta pero las API Stream que maneja Twitter tienen otro propósito, este tipo de servicios permite establecer una petición y respuesta constante de información en este caso de tweets, por lo que podemos decir que el consumo de estos datos no siempre están enfocados a contenido multimedia sino también texto.

2.10 Metodología Scrum

"Scrum es un marco ligero que ayuda a las personas, equipos y organizaciones a generar valor a través de soluciones adaptables para problemas complejos" [JIMENEZ \(2022\)](#) Scrum forma parte de la metodología ágil (tiene que ser rápido y flexible al cambio) para crear software, entre otros, pero scrum es un framework de trabajo que se adapta a los proyectos y se enfoca en desarrollar las funcionalidades de los proyectos, para lo cual scrum divide en roles para trabajar denominados Scrum Team:

- **Product Owner (PO):** Es el puente de comunicación entre los que realizan el proyecto y el cliente, en este se establecerá los tiempos que se tendrá para el sprint (entrega de algo funcional), se establece los backlog (historias de usuarios) y demás, cabe recalcar que product owner no es ni se considera el jefe.
- **Scrum Master:** Facilita el trabajo según las guías de scrum con la finalidad de entregar a tiempo las funcionalidades establecidas.

- **Scrum Team:** Aquí se encuentran todo el equipo de profesionales que actúan en el proyecto, este equipo se requiere que puedan adaptarse al cambio.

2.11 Zero-shot Learnig

"El Aprendizaje sin ejemplos (ZSL, por su denominación en inglés Zero-shot Learning), es un conjunto de problemas de aprendizaje automático, donde en el momento de la prueba se observan muestras de clases que no se observaron durante el entrenamiento, y se necesita predecir la categoría a la que pertenecen" Toledo (2021). Zero-shot learning forma parte del aprendizaje automático. ZSL consiste en crear modelos de aprendizaje profundo en donde el usuario puede observar varios muestras para predecir ciertas informaciones y posteriormente categorizarlos. Estas muestras o clases que no se tienen información se suplirá por la clase vista como la invisible, el espacio que se pretende suplir se denomina espacio auxiliar.

"Es necesario entender que en este tipo de configuración, existen dos tipos de clases. Las vistas, que son todas aquellas que tienen al menos una instancia en los datos de entrenamiento y las invisibles que no tienen ninguna instancia en los datos de entrenamiento" Toledo (2021). Zero-shot Learning se puede dividir en dos categorías según la fase de entrenamiento y de prueba:

2.11.1 Según la fase de entrenamiento

- **zero-shot learning inductivo:** tienen acceso a las clases vistas.
- **zero-shot learning trasductivo:** además de tener acceso a las clases vistas también a las clases no vistas con el fin de que soporte las predicciones para las clases invisibles.

2.11.2 Según la fase de prueba

- **zero-shot learning convencional:** evalúa las clases no vistas.
- **zero-shot learning generalizado:** evalúa las clases vistas como no vistas.

2.12 Trabajos relacionados

El desarrollo de proyectos que involucran técnicas como el Procesamiento del Lenguaje Natural para resolver ciertos problemas de clasificación cada vez más se han ido mejorado conforme las tecnologías van avanzado, en esta sección se describirá trabajos relacionados con el tema principal de este proyecto que va desde la aplicación tradicional de modelos de clasificación hasta los denominados modelos de Disparo Zero (Zero-Shot Clasificación).

Como por ejemplo el proyecto de tesis doctoral denominado “CLASIFICACIÓN AUTOMÁTICA MULTICLASE DE TWEETS Y SU REPRESENTACIÓN GRÁFICA“ desarrollado por Beatriz Hernández Pajares muestra como el uso de la clasificación automática de tweets ayudan a identificar problemáticas sociales que están pasando en una ciudad determinada, esto gracias a la aplicación de modelos de clasificación lineal como el caso de SVM (Maquinas de vectores de soporte), el cual ha permitido establecer un modelo que reconoce temas de interés que han sido identificados anteriormente.

Capítulo III

MARCO METODOLOGICO

En la construcción de este proyecto se pudo implementar una de las metodologías más utilizadas hoy en día por la industria del software, la llamada metodología Scrum que gracias a que permite realizar el proceso de trabajo con gran flexibilidad forma parte de las denominadas metodologías ágiles para la gestión del desarrollo de software dentro de un grupo determinado de personas, y que a breves rasgos consiste en realizar primero la construcción de las funcionalidades (Requerimientos) que tienen mayor prioridad para el cliente y la presentación de los mismos en un rango determinado de tiempo (Sprint), esto con el fin de que el cliente pueda formar parte de todo el proceso de construcción e identificar problemas de forma temprana para realizar los cambios funcionales o de prioridad al inicio de cada iteración.



Figura 3.1: Diagrama de la metodología Scrum
Diego Calvo (2019b)

En la imagen anterior se describe el proceso principal realizado en el desarrollo de cada uno de los requerimientos que detallaremos en las siguientes secciones, este procedimiento se estableció con un sprint de cada 4 semanas (es decir cada mes) con el fin de ajustarnos a los tiempos de cada involucrado.

3.1 Levantamiento de Requerimientos

Para tener una idea clara de la propuesta que se ha desarrollado debemos considerar que todas las funcionalidades que tiene la plataforma parten de las actividades planificadas del proyecto y que permitieron generar el componente principal de esta metodología los requerimientos, de cada actividad se pudieron identificar diferentes requerimientos tanto funcionales como no funcionales y los cuales se detallaran en las siguientes secciones.

3.1.1 Personal Involucrado

Dentro de la planificación del proyecto se encuentran involucradas únicamente dos personas que tienen roles determinados para la correcta ejecución del desarrollo.

Nombre	Carlos Homero Morocho Medina
Categoría Profesional	Estudiante
Rol	Desarrollador
Responsabilidad	Análisis, diseño y programación de la plataforma
Información de contacto	cmorocho1@est.ups.edu.ec

Nombre	Gabriel Alejandro León Paredes
Categoría Profesional	Administrador del proyecto
Rol	Scrum Master
Responsabilidad	Verificar y aprobar los sprint del proyecto
Información de contacto	gleon@ups.edu.ec

Tabla 3.1: Personal involucrado

3.1.2 Requisitos Funcionales

A continuación, se detalla los requerimientos funcionales que se han podido identificar conforme a las actividades que hemos planificado para este proyecto, si bien algunas actividades no representan a un producto final como tal es importante resaltar que estos requerimientos están netamente relacionados al proceso del desarrollo de la plataforma y no a procesos o actividades enfocadas a la investigación que se tuvo que realizar en el transcurso del desarrollo.

- Requisitos funcionales correspondientes a la página de autenticación.

Código	RE01
Nombre	Inicio de sesión.
Descripción	Proceso para ingresar la plataforma.
Entradas	Correo, Contraseña
Proceso	<p>La plataforma deberá implementar un proceso de autenticación de usuario para poder ingresar a visualizar los datos del análisis de los tweets, para esto se debe considerar que cuando el usuario inicie sesión deberá ingresar directamente a la sección del "En vivo" de la cuenta que tenga vinculada, antes de autenticar al usuario se debe validar los siguientes puntos.</p> <ul style="list-style-type: none"> - Se debe validar que el usuario ingrese todos los datos. - Se debe validar que el correo ingresado sea correcto.
Salidas	<ul style="list-style-type: none"> - Datos de sesión del usuario. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.2: Requisito funcional 1.

El siguiente grafico UML representa el proceso del requerimiento RE01.

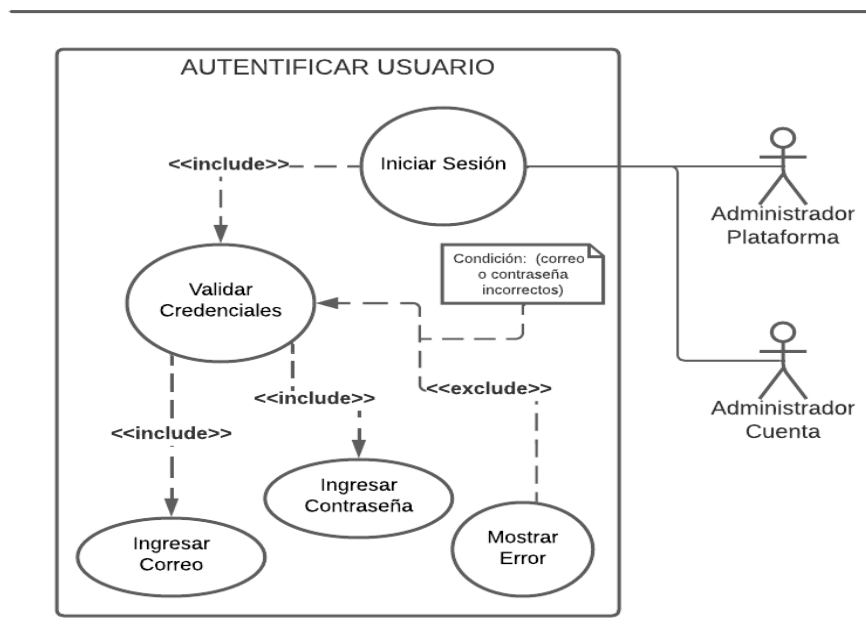


Figura 3.2: Diagrama de caso de uso de "Inicio de session".

- Requisitos funcionales correspondientes a la pagina del Dashboard (Panel de control).

Código	RE02
Nombre	Streaming de tweets.
Descripción	Proceso de recopilación de tweets en tiempo real.
Entradas	Identificador de la cuenta
Proceso	<p>La plataforma deberá implementar la API Streaming de Twitter para recopilar los tweets en tiempo real, es decir cada Mención, Respuesta, Cita o Retweet que se haga a la cuenta de la empresa pública vinculada se debe registrar en la BD y visualizarla en una pantalla donde se simulara el “En vivo” de las interacciones que se vayan generando en el día, para esto se debe considerar que.</p> <ul style="list-style-type: none"> - El usuario podrá redireccionarse a ver el tweet en la plataforma de Twitter. - El usuario podrá ir quitando de la lista de “En vivo” los tweets recopilados. - El usuario podrá ver la informacion completa del Tweet. - La pantalla solo visualizara los últimos 20 Tweets recopilados. <p>Cada que se agregue un nuevo registro en el “En vivo” de tweets, este deberá contar la informacion del usuario (autor del tweet), el tipo de interacción, la fecha de publicación, texto del tweet y la clasificación obtenida.</p>
Salidas	<ul style="list-style-type: none"> - Información del tweet. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.3: Requisito funcional 2.

Código	RE03
Nombre	Informacion completa del Tweet.
Descripción	Proceso de visualización toda la informacion del Tweet.
Entradas	Identificador del tweet
Proceso	<p>La plataforma deberá implementar una pantalla tipo modal (Popup) que se mostrará al dar clic en algún registro de tweet presentado en las diferentes vistas de la plataforma, el cual permitirá visualizar la informacion completa del tweet incluyendo (Informacion del Tweet, Métricas, Entidades y Clasificación).</p>
Salidas	<ul style="list-style-type: none"> - Información completa del tweet. - Mensaje de error al ocurrir algún problema.
Prioridad	Medio

Tabla 3.4: Requisito funcional 3.

Código	RE04
Nombre	Clasificación de tweets.
Descripción	Proceso de clasificación de tweets.
Entradas	Información del tweet, Clases de la cuenta, Identificador de la cuenta
Proceso	<p>La plataforma deberá contar con un proceso donde se analizará y clasificará el texto del tweet obtenido el requerimiento RE02, según la configuración de clases que tenga la cuenta vinculada, para esto se debe considerar los siguientes puntos.</p> <ul style="list-style-type: none"> - El texto debe ser normalizado antes de la predicción. - Las clases deberán pasar unos parámetros de aceptación. - Las clases deberán contar con un color específico.
Salidas	<ul style="list-style-type: none"> - Lista de predicción de clases. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.5: Requisito funcional 4.

El siguiente grafico UML representa el proceso del requerimiento RE02, RE03 y RE04.

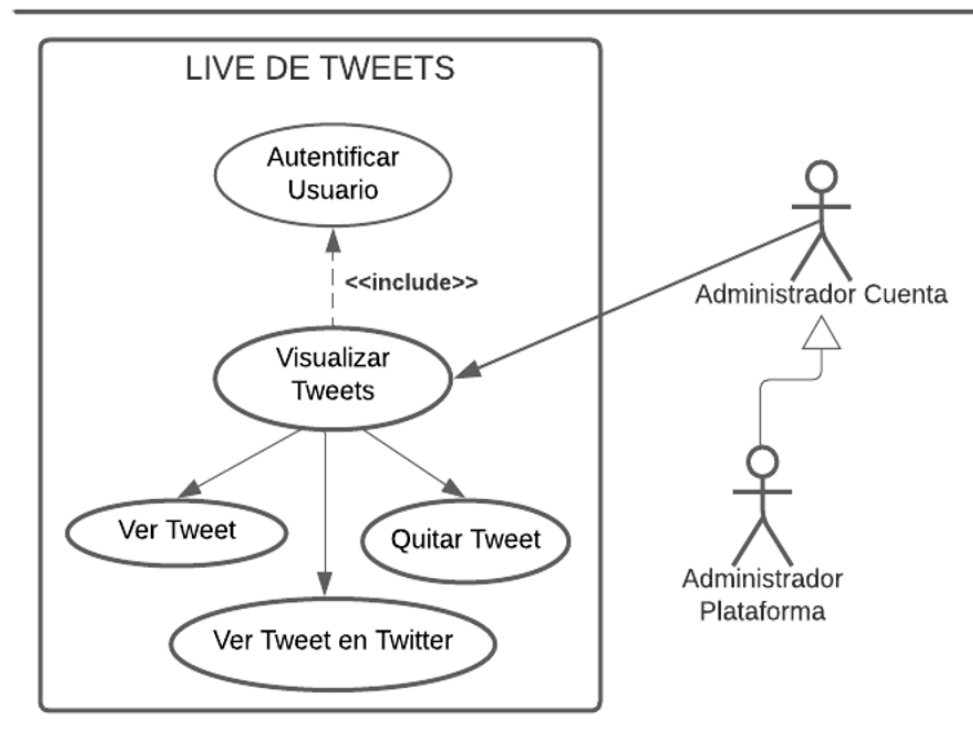


Figura 3.3: Diagrama de caso de uso de "Streaming de tweets".

Código	RE05
Nombre	Live de cuentas de interacción.
Descripción	Proceso del en vivo de las cuentas que interactúan.
Entradas	Identificador de la cuenta
Proceso	<p>La plataforma deberá contar con una sección dentro de la misma pantalla donde se realizar el requerimiento RE02, que permita visualizar los usuarios que interactuaron con la cuenta de la institución pública vinculada, es decir todas cuentas autoras que realizaron alguna Mención, Respuesta, Cita o Retweet en algún punto del “En vivo” de ese día, para esto se debe considerar que.</p> <ul style="list-style-type: none"> - Las cuentas con mayor interacción liderarán la lista. - Se debe detallar la cantidad de tweets por tipo. - El usuario podrá ver el detalle de los tweets.
Salidas	<ul style="list-style-type: none"> - Lista de cuentas de interaccion. - Mensaje de error al ocurrir algún problema.
Prioridad	Medio

Tabla 3.6: Requisito funcional 5.

El siguiente grafico UML representa el proceso del requerimiento RE05.

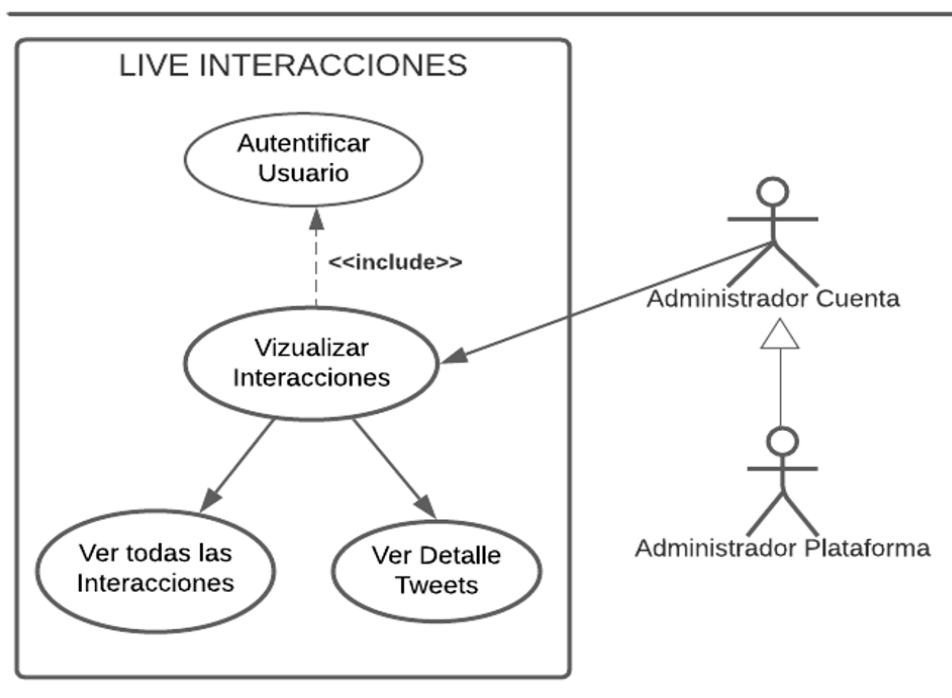


Figura 3.4: Diagrama de caso de uso de "Live de cuentas de interacción".

Código	RE06
Nombre	Detalle de tweets por tipo.
Descripción	Proceso de visualización de tweets por tipo.
Entradas	Filtros de tweets
Proceso	La plataforma deberá implementar una pantalla tipo modal (Popup) que se mostrará al dar clic en algún registro de interacción, esta pantalla mostrara la lista de tweets de forma seccionada por los tipos de interacción (Menciones, Respuestas, Citas y Retweets), es decir cada Tab (pestaña) corresponderá a un tipo de tweet especifico y se cargaran los tweets que correspondan al tipo seleccionado con un límite de 50 tweets, por defecto al cargar la pantalla de debe sebe establecer en la sección de todos sin tomar en cuenta el tipo.
Salidas	- Lista de tweets por tipo. - Mensaje de error al ocurrir algún problema.
Prioridad	Medio

Tabla 3.7: Requisito funcional 6.

El siguiente grafico UML representa el proceso del requerimiento RE06.

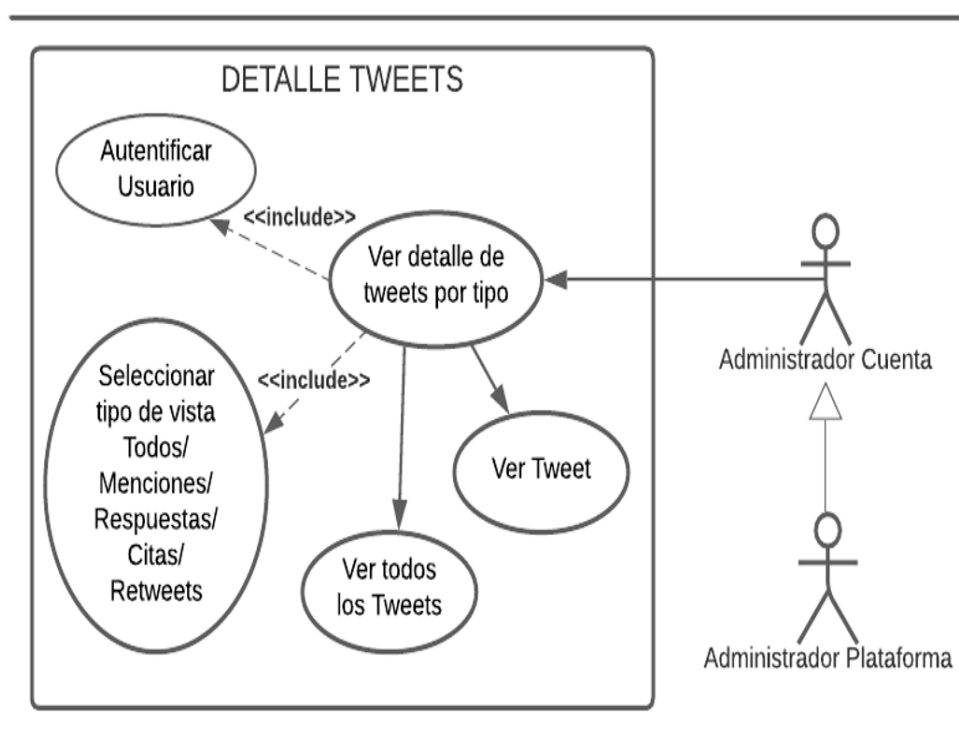


Figura 3.5: Diagrama de caso de uso de "Detalle de tweets".

- Requisitos funcionales correspondientes a la pagina de Interacciones.

Código	RE07
Nombre	Grafica de interacciones de tweets.
Descripción	Proceso de visualización del grafico tipo líneas de interacción.
Entradas	Identificador de la cuenta, Modo de vista
Proceso	<p>La plataforma deberá implementar una pantalla donde se visualice una gráfica de líneas que represente los cuatro tipos de tweets (Mención, Respuesta, Cita y Retweet) que se manejan, esta grafica representara el total de tweets recopilados por fecha según el modo que escoja usuario, los cuales constan de.</p> <ul style="list-style-type: none"> - Modo ultimo 30 días, mostrara el total de tweets por tipo de los ultimos 30 dias. - Modo mensual, mostrara el total de tweets por tipo de todos los meses del año presente. - Modo anual, mostrara el total de tweets por tipo de todos los años recopilados. <p>Además, el usuario podrá visualizar las cuentas de interacción al dar clic en una fecha determinada.</p>
Salidas	<ul style="list-style-type: none"> - Lista de tweets por tipo y fecha. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.8: Requisito funcional 7.

Código	RE08
Nombre	Cuentas de interacción.
Descripción	Proceso de visualización de cuentas de interacción según la fecha.
Entradas	Identificador de la cuenta, fecha desde, fecha hasta
Proceso	<p>La plataforma deberá implementar una pantalla tipo modal (Popup) que se mostrará al dar clic en alguna sección de la gráfica del requerimiento RE07, esta pantalla mostrará la misma informacion del requerimiento RE05 salvo que en este caso será en base a la fecha que estemos evaluando, considerando que.</p> <ul style="list-style-type: none"> - Se mostrara un total de 50 cuentas. - Se podra ver el detalle de tweets de una cuenta.
Salidas	<ul style="list-style-type: none"> - Lista de cuentas de interaccion. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.9: Requisito funcional 8.

El siguiente grafico UML representa el proceso del requerimiento RE07 y RE08.

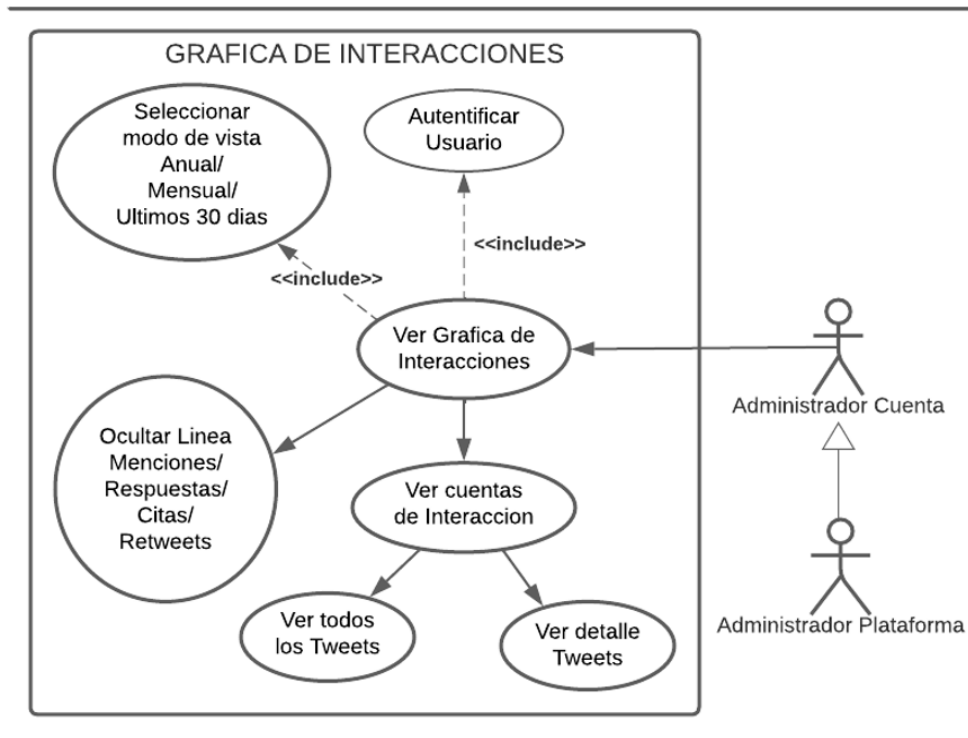


Figura 3.6: Diagrama de caso de uso de "Grafica de interacciones".

- Requisitos funcionales correspondientes a la pagina de Clasificacion.

Código	RE09
Nombre	Selección de filtros de tweets.
Descripción	Proceso de selección de filtros para visualizar los tweets.
Entradas	Texto, Fecha desde, Fecha Hasta, Tipos, Clases, Score min, Score máx, Modo, Autores, limite.
Proceso	La plataforma deberá implementar una pantalla tipo modal (Popup) que se mostrará cada que se ingrese a la pantalla de clasificación de tweets o al dar clic en el botón de filtros, esta pantalla permitirá al usuario seleccionar los diferentes filtros de búsqueda para mostrar los resultados de los requerimientos RE12 - RE13 - RE14 de forma más específica. <ul style="list-style-type: none"> - Se puede no seleccionar ningún filtro. - Limite por defecto será 50
Salidas	Diccionario de filtros seleccionados.
Prioridad	Alta

Tabla 3.10: Requisito funcional 9.

Código	RE10
Nombre	Selección reporte de tweets.
Descripción	Proceso de selección de parámetros para imprimir el reporte de tweets.
Entradas	Campos autor, Campos clasificación, Agregar grafica líneas, Agregar grafica barras, Agregar total por tipo, Título de reporte, Tipo documento
Proceso	<p>La plataforma deberá implementar una pantalla tipo modal (Popup) que solo se mostrará al dar clic en el botón de impresión, esta pantalla permitirá al usuario seleccionar los diferentes parámetros de reportes para por imprimir el documento con la información obtenida según los filtros agregados en el requerimiento RE09.</p> <ul style="list-style-type: none"> - Se puede no seleccionar ningún parámetro. - Campos del tweet por defecto Código, Fecha, Texto, Tipo. - Formato de documento por defecto .pdf
Salidas	Diccionario de reporte seleccionados.
Prioridad	Alta

Tabla 3.11: Requisito funcional 10.

Código	RE11
Nombre	Acción de vista tweets.
Descripción	Proceso de agregar acciones de vista de tweets.
Entradas	Código, Nombre, Descripción, JSON filtros, JSON reporte
Proceso	<p>La plataforma deberá implementar una pantalla tipo modal (Popup) que se mostrará al dar clic en el botón de acción, esta pantalla permitirá al usuario guardar los filtros y parámetros de reporte seleccionados en los requerimientos RE09 Y RE10 dentro de una acción, estas acciones permitirán volver a visualizar la información de tweets en base a una configuración.</p> <ul style="list-style-type: none"> - Se puede especificar acciones favoritas. - EL asistente virtual estará continuamente recomendando estas acciones.
Salidas	- Identificador de la acción. - Mensaje de error al ocurrir algún problema.
Prioridad	Media

Tabla 3.12: Requisito funcional 11.

Código	RE12
Nombre	Sección lista de tweets clasificados.
Descripción	Proceso de listar los tweets clasificados según los filtros.
Entradas	Filtros de tweets
Proceso	<p>La plataforma deberá implementar una sección de lista de tweets donde se mostrará la información de todos los tweets clasificados que pasen los filtros seleccionados en el requerimiento RE09, en esta información se debe mostrar la fecha, el usuario autor, el texto y los resultados de clasificación, considerando que.</p> <ul style="list-style-type: none"> - Los tweets deben mostrarse por páginas. - El límite de tweets obtenidos debe ser de 50. - Los tweets deben mostrarse por fecha de publicación del más actual al más antiguo. - El resultado de clasificación debe mostrarse según el color de la clase y el porcentaje obtenido. <p>Además, que el usuario podrá dar clic en cualquier registro de tweet y visualizar la información completa del tweet.</p>
Salidas	<ul style="list-style-type: none"> - Lista de tweets clasificados. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.13: Requisito funcional 12.

Código	RE13
Nombre	Sección de gráfica (Barras) total de tweets clasificados por clase.
Descripción	Proceso de visualizar la gráfica del total de tweets clasificados según los filtros.
Entradas	Filtros de tweets
Proceso	<p>La plataforma deberá implementar una sección de gráfica de barras donde se mostrará el total de tweets clasificados por clase según el resultado de los tweets que hayan pasado los filtros seleccionados, considerando que.</p> <ul style="list-style-type: none"> - El usuario podrá ocultar las barras de las clases. - El usuario podrá visualizar el detalle de tweets de las clases. - Las barras de la gráfica deberán estar pintadas según el color configurado de la clase.
Salidas	<ul style="list-style-type: none"> - Lista de clasificación. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.14: Requisito funcional 13.

Código	RE14
Nombre	Sección de gráfica (Lineas) total de tweets clasificados por fechas.
Descripción	Proceso de visualizar la gráfica del total de tweets clasificados según los filtros.
Entradas	Filtros de tweets
Proceso	<p>La plataforma deberá implementar una sección de grafica de lineas donde se mostrará el total de tweets clasificados por fecha según el resultado de los tweets que hayan pasado los filtros seleccionados, considerando que.</p> <ul style="list-style-type: none"> - El usuario podrá ocultar las lineas de las clases. - Las lineas de la gráfica deberán estar pintadas según el color configurado de la clase.
Salidas	<ul style="list-style-type: none"> - Lista de clasificación. - Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.15: Requisito funcional 14.

El siguiente grafico UML representa el proceso del requerimiento RE09, RE10, RE11, RE12, RE13 y RE14.

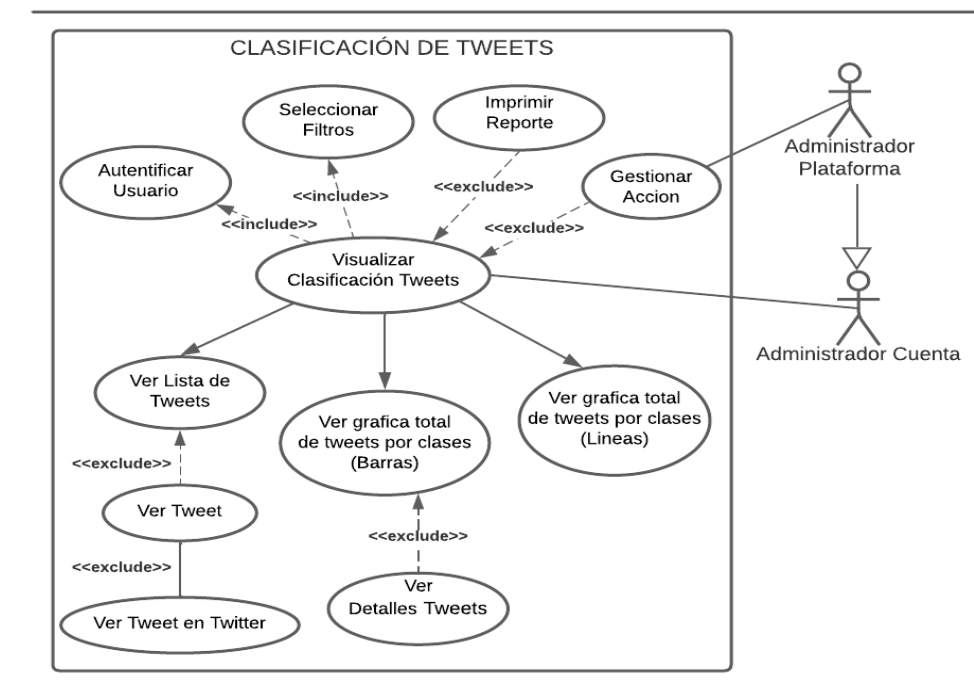


Figura 3.7: Diagrama de caso de uso de "Clasificacion de tweets".

- Requisitos funcionales correspondientes a la pagina de Configuracion.

Código	RE15
Nombre	Vinculación de cuentas
Descripción	Proceso de vinculación de la cuenta de una empresa pública.
Entradas	Nombre de la cuenta
Proceso	<p>La plataforma deberá implementar la API Rest Full de Twitter para buscar la cuenta de la institución pública por el nombre ejemplo "@emov_ep", en el caso de encontrar alguna coincidencia se debe poder vincular la cuenta en la plataforma caso contrario no, además de que se debe considerar lo siguiente.</p> <ul style="list-style-type: none"> - Este proceso solo lo podra realizar el usuario administrador de la plataforma. - Cuando se vincule la cuenta se generara la clase por defecto. - Cuando se vincule la cuenta se agregara la regla de Streaming.
Salidas	- Mensaje de error al ocurrir algún problema.
Prioridad	Alta

Tabla 3.16: Requisito funcional 15.

El siguiente grafico UML representa el proceso del requerimiento RE15.

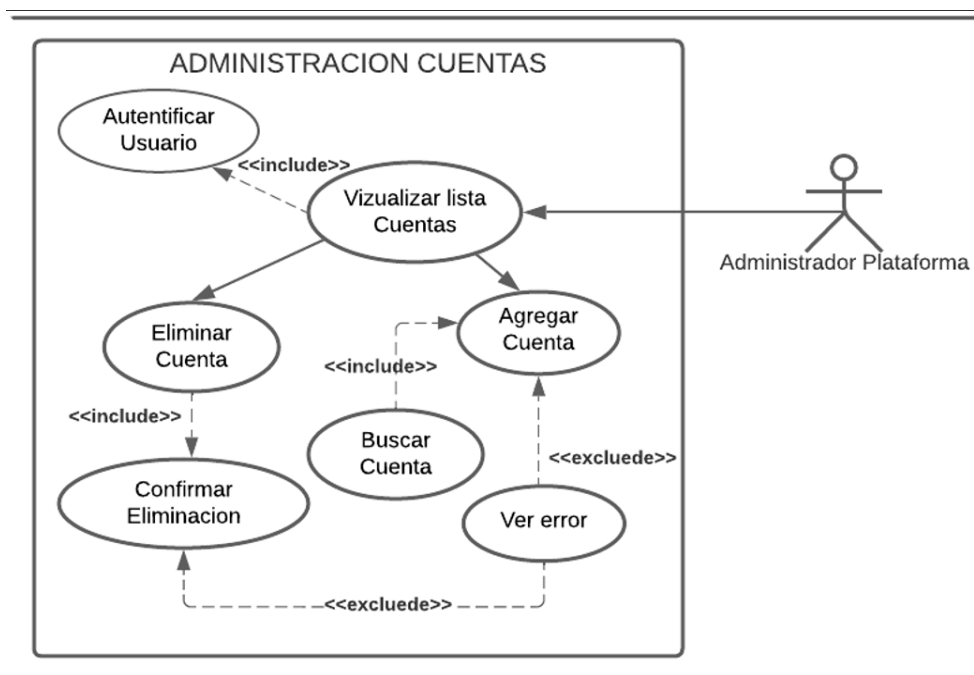


Figura 3.8: Diagrama de caso de uso de "Vinculación de cuentas".

Código	RE16
Nombre	Gestión de clases.
Descripción	Proceso de gestión de las clases de cada cuenta.
Entradas	<p>La plataforma deberá permitir gestionar las clases que se agregan a las cuentas vinculadas, para eso se deberá contar con una sección donde se listen las clases agregadas en la cuenta del usuario que haya iniciado sesión, dentro de esa pantalla se debe contar también con un botón que muestre un modal (Popup) para poder ingresar la información de una nueva clase, y considerando que.</p> <ul style="list-style-type: none"> - Solo el usuario administrador podrá visualizar, agregar, editar y eliminar una clase. - Si el usuario da clic en algún registro de la lista, se abrirá el mismo modal para editar la clase. - Antes de actualizar y/o agregar una clase se debe validar que se haya ingresado toda la información.
Proceso	<ul style="list-style-type: none"> - Identificador de la clase. - Mensaje de error al ocurrir algún error.
Salidas	Mensaje de error al suscitar algún error durante el proceso.
Prioridad	Alta

Tabla 3.17: Requisito funcional 16.

El siguiente grafico UML representa el proceso del requerimiento RE16.

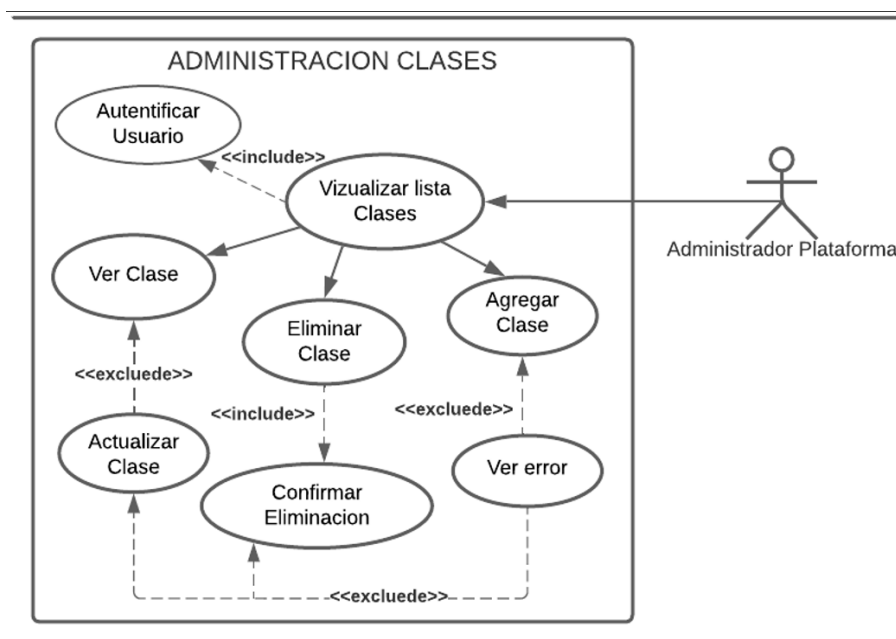


Figura 3.9: Diagrama de caso de uso de "Gestión de clases".

Código	RE17
Nombre	Gestión de usuarios.
Descripción	Proceso de gestión de los usuarios de cada cuenta.
Entradas	<p>La plataforma deberá permitir gestionar los usuarios que se agregan a las cuentas vinculadas, para eso se deberá contar con una sección donde se listen todos los usuarios agregados en la plataforma, dentro de esa pantalla se debe contar también con un botón que muestre un modal (Popup) para poder ingresar la información de un nuevo usuario, y considerando que.</p> <ul style="list-style-type: none"> - Solo el usuario administrador podrá visualizar, agregar, editar y eliminar un usuario. - Si el usuario da clic en algún registro de la lista, se abrirá el mismo modal para editar el usuario. - Antes de actualizar y/o agregar un usuario se debe validar que se haya ingresado toda la información.
Proceso	<ul style="list-style-type: none"> - Identificador del usuario. - Mensaje de error al ocurrir algún error.
Salidas	Mensaje de error al suscitar algún error durante el proceso.
Prioridad	Alta

Tabla 3.18: Requisito funcional 17.

El siguiente grafico UML representa el proceso del requerimiento RE17.

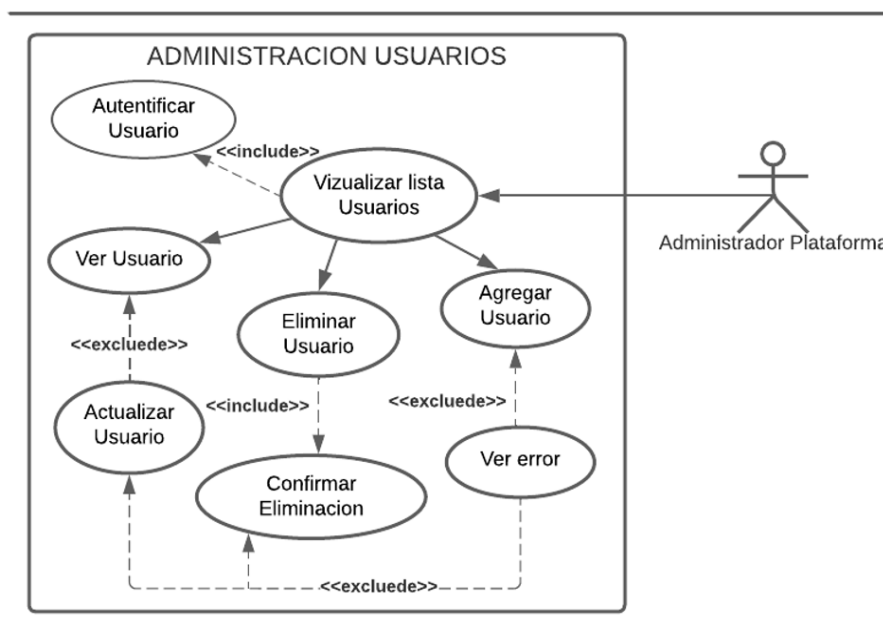


Figura 3.10: Diagrama de caso de uso de "Gestión de usuarios".

Código	RE18
Nombre	Implementación de un Chatbot.
Descripción	Proceso de vinculación de un asistente virtual.
Entradas	Preguntas del usuario
Proceso	La plataforma deberá implementar el asistente virtual de IBM Watson en una sección de la plataforma web como un tipo de Chatbot que permitirá al usuario realizar preguntas puntuales sobre la información recopilada.
Salidas	Respuestas del asistente.
Prioridad	Alta

Tabla 3.19: Requisito funcional 18.

A continuación, se presenta el gráfico UML general que permite identificar los procesos principales de la plataforma web descritos anteriormente junto con los usuarios que intervienen en cada proceso.

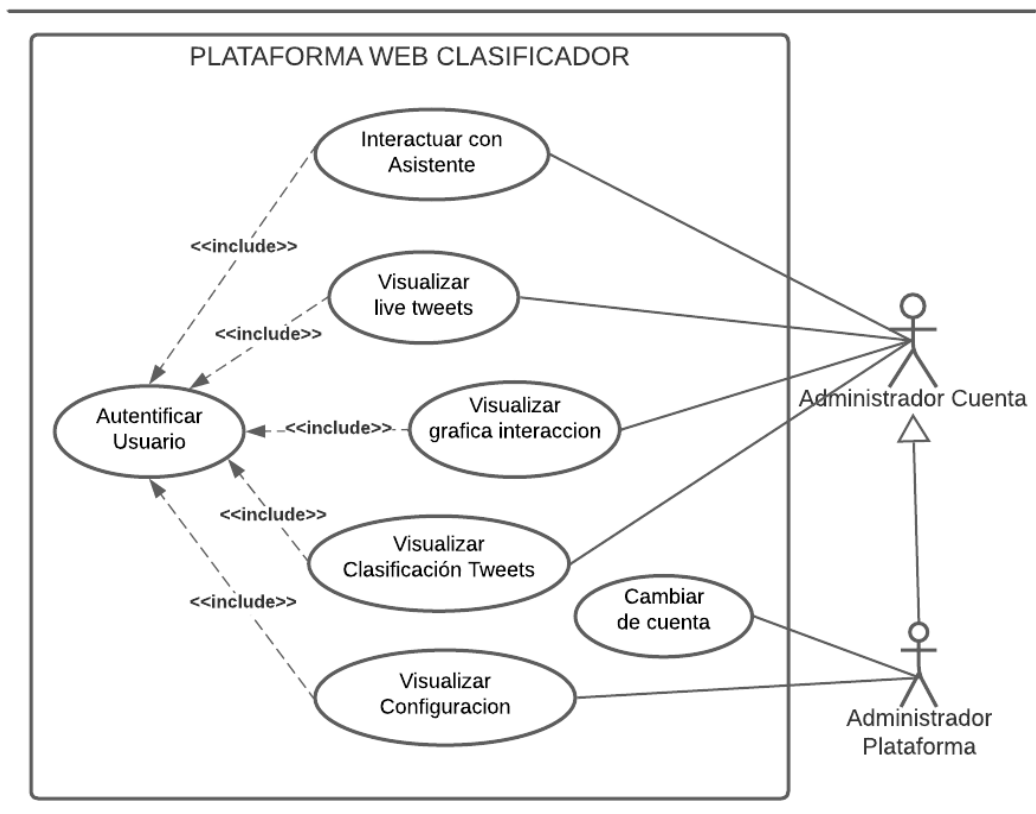


Figura 3.11: Diagrama de caso de uso de "Plataforma web".

3.1.3 Requisitos No Funcionales

Código	RE19
Nombre	Barra de carga.
Descripción	Al momento de buscar según los parámetros agregados en el requerimiento 8 se mostrará una pantalla con una barra de carga antes de visualizar el resultado de la búsqueda.
Entradas	Realizar una consulta.
Proceso	La plataforma buscará los tweets con base en los filtros agregados en el requerimiento 8 y mostrará la cantidad de tweets.
Salidas	Barra de carga.
Prioridad	Baja

Tabla 3.20: Requisito no funcional 19.

Código	RE20
Nombre	Información de la cuenta.
Descripción	La plataforma mostrará la información correspondiente de la cuenta.
Entradas	Id de la cuenta
Proceso	La plataforma buscará las cuentas que correspondan a los filtros agregados en el requerimiento 8 y mostrará la información.
Salidas	Información de la cuenta.
Prioridad	Alta

Tabla 3.21: Requisito no funcional 20.

Código	RE21
Nombre	Total de tweets clasificados.
Descripción	La plataforma mostrará la cantidad de tweets echos por la cuenta.
Entradas	Id de la cuenta
Proceso	La plataforma buscare los tweets en base a los filtros agregados en el requerimiento 8 y mostrará la cantidad de tweets clasificados.
Salidas	Cantidad total de tweets clasificados.
Prioridad	Alta

Tabla 3.22: Requisito no funcional 21.

Código	RE22
Nombre	Total de tweets por tipo.
Descripción	La plataforma mostrará la cantidad de tweets clasificados por tipo echos en las cuenta.
Entradas	Id de la cuenta
Proceso	La plataforma buscara los tweets en base a los filtros agregados en el requerimiento 8 y mostrará la cantidad de tweets clasificados por tipo.
Salidas	Cantidad de tweets clasificados por tipo.
Prioridad	Alta

Tabla 3.23: Requisito no funcional 22.

Código	RE23
Nombre	Paginación de listas de tweets.
Descripción	La plataforma asignará una página a la lista de los tweets filtrados, se asignará una página nueva a cada 50 tweets como máximo, es decir, que se listará 50 tweets en cada página.
Entradas	datos del requerimiento 8
Proceso	La plataforma agregará las páginas a cada 50 tweets.
Salidas	lista de tweets.
Prioridad	Alta

Tabla 3.24: Requisito no funcional 23.

Disponibilidad

La disponibilidad del sistema será en todo momento, es decir, las 24 horas al día de cada semana, garantizando el servicio únicamente si la plataforma social Twitter esta activa y no existan factores externos que puedan interrumpir el uso de la plataforma.

Seguridad

La plataforma fue construida bajo la premisa de servir a una organización publica en particular por lo que la información obtenida de las redes sociales solo podrá ser administrada, visualizada y actualizada por el personal a cargo (administrador).

3.2 Arquitectura del sistema

Todos los requerimientos funcionales dieron como resultado la creación integral de la arquitectura que vemos en la imagen a continuación [3.12](#), donde se observa el uso de las tecnologías necesarias para alcanzar el objetivo principal de este proyecto.

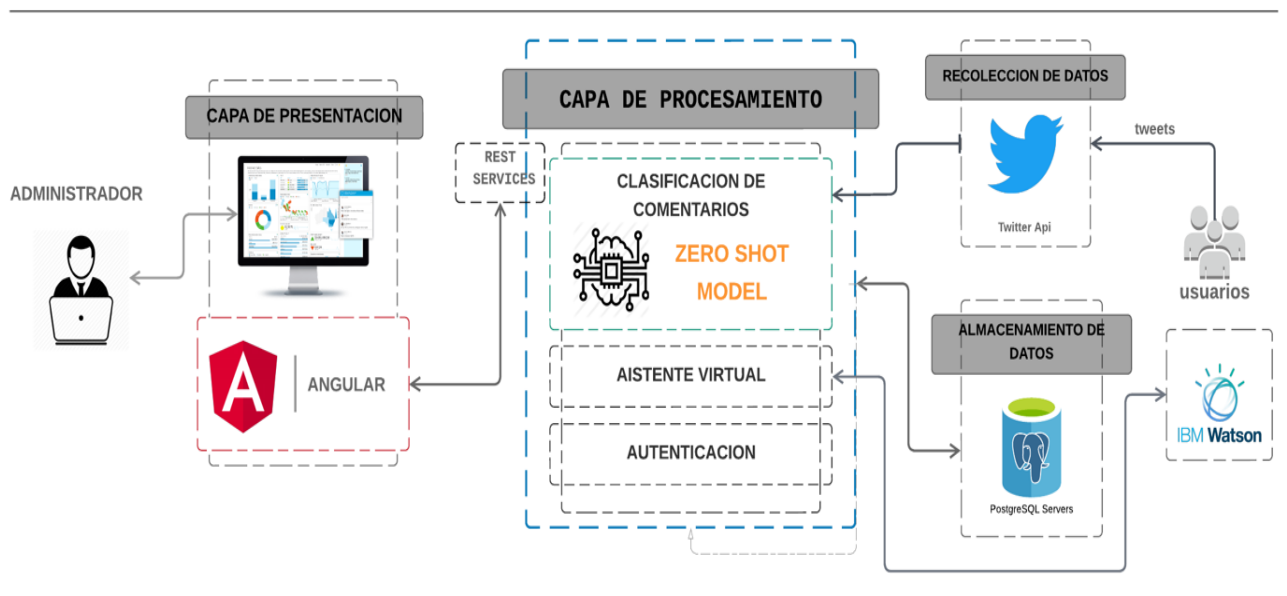


Figura 3.12: Arquitectura de la plataforma.

A continuación, se detalla de forma general el proceso que representa cada una de las capas implementadas en la arquitectura de esta plataforma, como primer punto tenemos la fase denominada Integración con la red social Twitter que mediante el uso de sus APIs Rest Full públicas permite a la plataforma recopilar la información en tiempo real de los tweets publicados por los usuarios finales que interactúen directamente con la cuenta oficial de la institución pública, es decir obtiene cualquier tweet público donde se haga una (Mención, Respuesta, Cita o Retweet) a la cuenta de la institución vinculada en la plataforma esto mediante unas reglas de streaming que estaremos analizando mas detalladamente en la seccion correspondiente.

La segunda fase denominada Clasificación de tweets consiste en realizar el análisis del texto obtenido en la fase de integración haciendo uso del modelo denominado *Zero-Shot-Classification* o en español modelo de Clasificación de Disparo-Zero que básicamente se trata de una técnica de entrenamiento automático basado en un dataset previamente etiquetado el cual permite reconocer clases que no necesariamente hayan formado parte del entrenamiento inicial pero sin embargo comparten ciertas características semánticas que ayudan a establecer una relación con las clases ya etiquetadas.

La tercera fase denominada almacenamiento de información consiste en guardar en una Base de Datos relacionada toda la información obtenida en la fase de integración con Twitter con su respectivo resultado de análisis obtenido en la fase de clasificación, y adicional a eso la información de la administración de los datos de los usuarios, clases y cuentas que se vinculen a la plataforma, todas estas tres fases conforman la capa denominada Procesamiento de Datos que consta de un conjunto de microservicios tipo API Rest Full desplegados sobre un servidor web de Flask (Back-End).

Por último, tenemos la fase de presentación de información donde se integra una aplicación web tipo Front-End construido bajo el framework de Angular y el cual consume todos los servicios desplegados en la capa de procesamiento para poder mostrar al usuario la información del análisis en tiempo real, incluyendo gráficas estadísticas, histórico de tweets clasificados y reportes de incidencias.

3.3 Tecnologías del proyecto

Seccion	Subseccion	Tecnología	Descripción
Back-End	Recopilacion	Request (v2.26.0)	Es una librería que permite consumir servicios tipo Rest Full
	Procesamiento	Pandas (v1.3.4)	Es una librería que proporciona un conjunto de utilitarios para el manejo de datos
		SocketIO (v5.1.0)	Es una librería que proporciona una interfaz para generar sockets de stream
		Transformers (v4.18.0)	Es una librería que proporciona una API para descargar y entrenar fácilmente modelos preentrenados de última generación para PyTorch, TensorFlow y JAX.
		PyTorch (v1.11.0)	Es una biblioteca de aprendizaje automático de código abierto basada en la biblioteca de Torch,
		Flask(v1.1.2)	Es un framework web basado en el lenguaje de programación Python
Almacenamiento	PostgreSQL (v13.5)	Administrador de base de datos relacionales con arquitectura cliente-servidor.	
Front-End	Presentación	Angular (v12.1.2)	Es un framework para aplicaciones web desarrollado en el lenguaje de programación typescript
		Node Js (v14.18.1)	Es un entorno de ejecución multiplataforma de JavaScript para la capa del servidor

3.4 Integración con Twitter

Como ya se había mencionado anteriormente el principio de este proyecto consiste en recopilar la información en tiempo real, por lo que la integración con la plataforma Twitter no estará basado en ningún proceso de minería de datos sino más bien en el consumo constante de ciertas APIs Rest Full públicas que permiten a plataformas externas como la de este proyecto obtener los tweets que se generan constantemente en la interacción de los usuarios finales dentro esta red social, considerando este punto es necesario aclarar que para poder consumir estas APIs se necesita tener una cuenta de Twitter Developer y generar un Token de Autenticación. Se dejara como referencia la siguiente página <https://developer.twitter.com/en/docs/platform-overview> donde se explica el proceso completo de forma más detallada.

3.4.1 Reglas de stream para obtener los Tweets

Como primer punto explicaremos cuales fueron las APIs Rest Full utilizadas para realizar el Streaming de tweets de las cuentas oficiales (Empresas Publicas) que se vinculan a esta plataforma, para esto Twitter API v2 maneja un concepto denominado reglas de stream que básicamente son filtros que permiten segmentar un conjunto de tweets publicados por diferentes usuarios y obtenerlos en tiempo real, el concepto de estos filtros viene dado por los Querys de Twitter que se han venido utilizado para recopilar informacion desde versiones anteriores por lo que no se explicara a fondo su sintaxis sino más bien como fueron estructuradas para crear las reglas de este proyecto, [Twitter \(2022\)](#). Así como se muestra en la siguiente imagen.

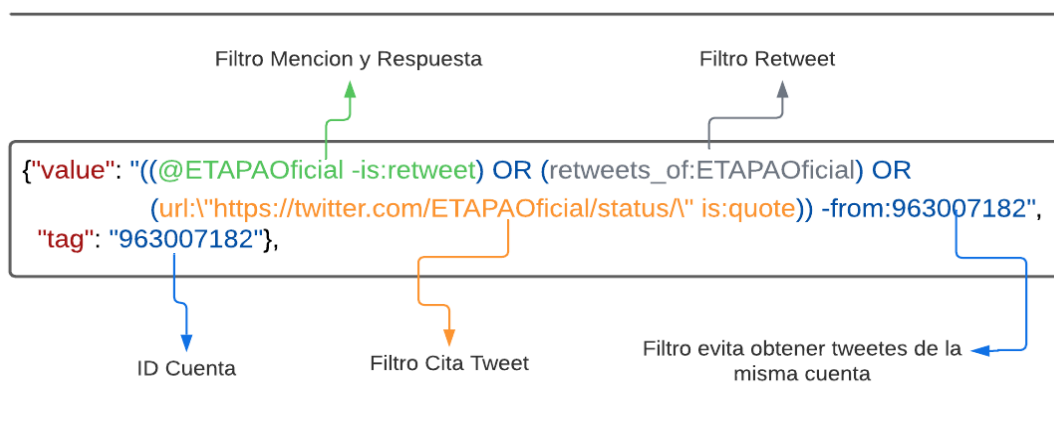


Figura 3.13: Regla de Stream para una cuenta

Cada cuenta que se vincula a la plataforma contara con la misma estructura de regla definida en la imagen anterior, teniendo en cuenta que en la versión Free de Twitter API V2 utilizada en este proyecto solo permite definir un total de 5 reglas se puede estructurar en una misma regla los filtros necesarios para segmentar los tweets de una o más cuentas conforme se vayan vinculando, sin embargo se estableció que cada regla se defina para una cuenta en específico esto debido a que la cantidad máxima de caracteres permitido en una regla es de 512 por lo que en algún punto no sera administrable dentro de la plataforma, además de que se intenta segmentar la informacion lo mas posible de cada una de las empresas publicas.

A continuación, una tabla descriptiva con todas las APIs Rest Full de Twitter utilizadas en este proyecto con el uso específico dentro los procesos de la plataforma.

URL: <https://api.twitter.com>

Método	Endpoints	Descripción
POST	/2/tweets/search/stream/rules	Esta API con el método de consumo POST permite agregar las reglas de stream
GET	/2/tweets/search/stream/rules	Esta API con el método de consumo GET permite obtener las reglas de stream agregadas
GET	/2/tweets/search/stream	Permite crear el proceso de Streaming de tweets
GET	/2/tweets	Permite obtener las métricas de los Tweets mediante el id
GET	/2/users/	Permite buscar una cuenta por su nombre, para poder vincularla a la plataforma
GET	/2/users/by	Busca una cuenta por id, para obtener la información básica de la cuenta

Tabla 3.25: APIs Rest Full de Twitter utilizados

3.4.2 Proceso de Streaming de Tweets

La siguiente grafica ilustra el flujo del proceso de integración con la red social Twitter desde la vinculación de la cuenta que consiste en agregar una regla de stream, hasta el mismo Streaming de tweets que consiste en obtener los datos en tiempo real y almacenarlos en la plataforma para luego presentarlos al usuario.

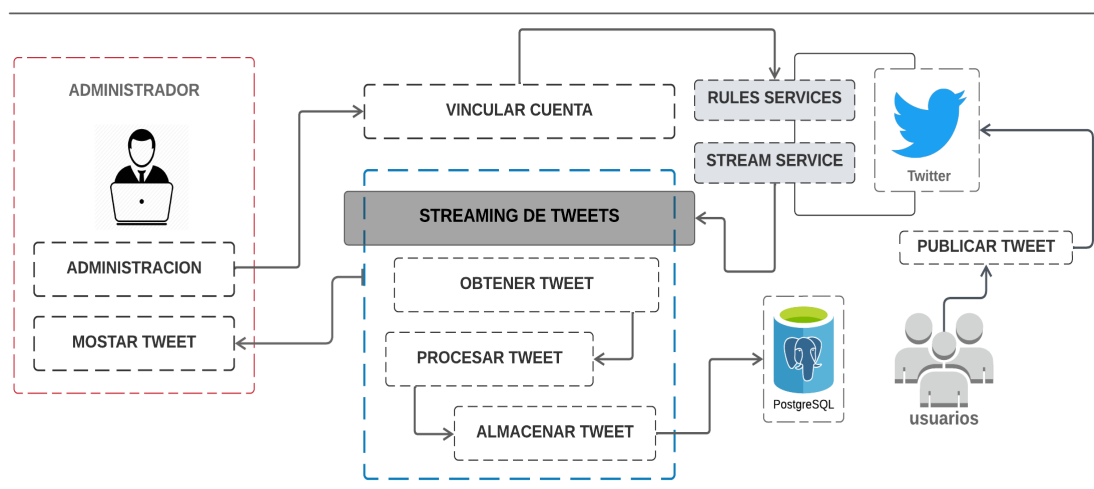


Figura 3.14: Diagrama integración con Twitter

Para este punto ya debemos tener en cuenta que solo el usuario administrador puede vincular una cuenta en la plataforma, cuando se vincule una cuenta de alguna institución pública la plataforma además de guardar la información básica también consume la API Rest que permite agregar una nueva regla de stream al pull de reglas permitidas para la cuenta de Twitter Developer, cada vinculación se hará con el formato de regla especificado en la sección anterior el cual permitirá segmentar las interacciones que tenga dicha cuenta en un tiempo determinado sean estas (Menciones, Respuestas, Citas o Retweets) , en Background la plataforma cuenta con un proceso que constantemente consume el API Streaming de Twitter donde las reglas agregadas en la etapa de vinculación permiten obtener en tiempo real todos los tweets que pasen el filtro de dichas reglas, haciendo que el proceso siempre este obteniendo, procesando y almacenando cualquier interacción (Tweets) que se haga de las cuentas vinculadas, ya sea que los usuarios estén o no conectados a la plataforma.

- **Tweet tipo Mención (M)**

Se considera un tweet tipo mención porque dentro de su texto escrito hace referencia a otra cuenta, esta referencia se representa por el signo "@" al inicio de un nombre, ejemplo @ETAPAOficial.



Figura 3.15: Ejemplo Tweet tipo Mencion

- **Tweet tipo Respuesta (R)**

Se considera un tweet tipo respuesta porque dentro de su texto escrito hace referencia a otra cuenta, así como en las menciones excepto que este tweet cuenta con una referencia

interna que lo enlaza directamente con otro tweet, ejemplo @LeninDelgadoR respondió a un tweet de la cuenta @centrosur_ec.



Figura 3.16: Ejemplo Tweet tipo Respuesta

- **Tweet tipo Cita (QT)**

Se considera un tweet tipo cita porque dentro de su texto escrito hace referencia al link de otro tweet, ejemplo twitter.com/ETAPAOficial/s.. es un hipervínculo a un tweet de la cuenta @ETAPAOficial.

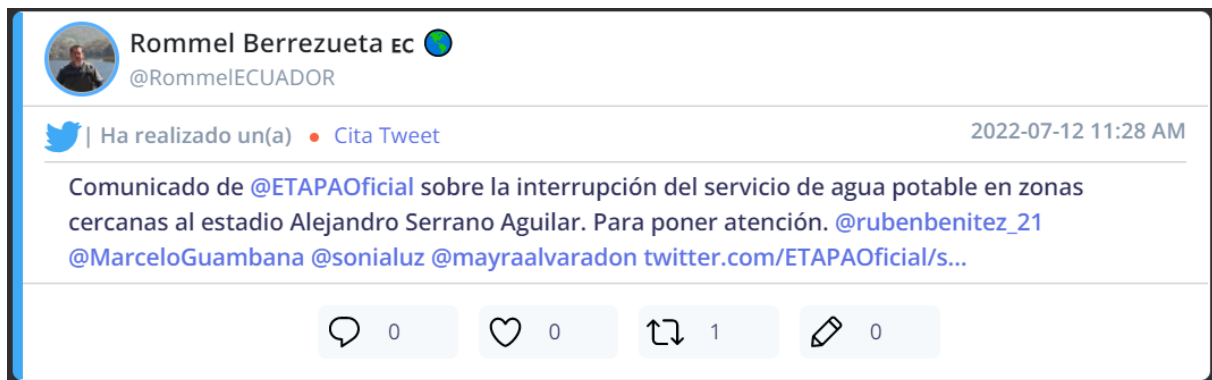


Figura 3.17: Ejemplo Tweet tipo Cita

- **Tweet tipo Retweet (RT)**

Se considera un tweet tipo retweet porque su texto empieza con un RT, lo que indica que simplemente retuiteo un tweet de alguna otra cuenta, ejemplo @jorgepatr retuiteo un tweet de @ETAPAOficial.



Figura 3.18: Ejemplo Tweet tipo Retweet

Es necesario explicar los 4 diferentes tipos de tweets que se maneja en la plataforma para poder describir claramente el importante proceso por el cual debe pasar un tweet que ha sido obtenido en la etapa del Streaming, ya que aquí se debe definir el tipo de interacción que es y a que cuenta vinculada pertenece, si bien este concepto no es muy notorio hemos podido evidenciar que un tweet dependiendo de la cuenta por el cual se lo evalúe va a cambiar su tipo, por ejemplo este tweet que responde a otro tweet de la cuenta @cmorocho4 y a su vez realiza una mención a las cuentas @rubenbenitez_21, @MarceloGuambana, @sonialuz y @mayraalvarado no es necesariamente una simple respuesta ya que para las demás cuentas involucradas es una mención y solo para la cuenta @cmorocho4 si es efectivamente una respuesta, y adicional a eso existe una cita a otro tweet de la cuenta @ETAPAOficial.

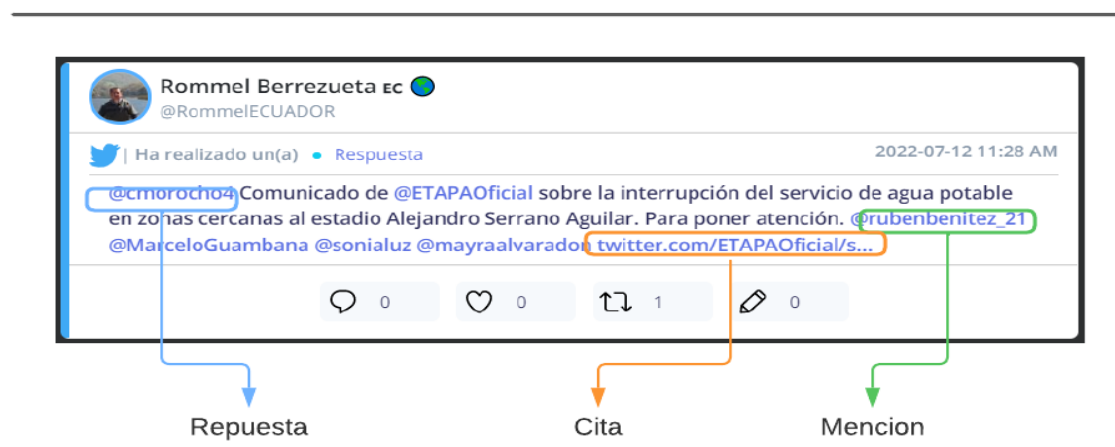


Figura 3.19: Ejemplo Tweet (Respuesta/Mencion/Cita)

3.5 Clasificación de tweets

Esta fase está completamente relacionada con la anterior ya que de hecho forma parte del procesamiento que pasa un tweet después de haber sido obtenido en la etapa del Streaming, para luego ser almacenado en la Base de Datos de la plataforma, como ya se había mencionado anteriormente un tweet está completamente ligado (o al menos para esta plataforma) al contexto de la cuenta con la que se lo evalúe, esto debido a que por lo general en el texto de un solo tweet se puede realizar citas, menciones e incluso respuestas a diferentes cuentas por lo que el tipo de tweet dependerá netamente de la forma con la que se relaciona a una cuenta, de esta misma manera se establecen las clases para cada una de las entidades públicas que se vinculan a la plataforma.

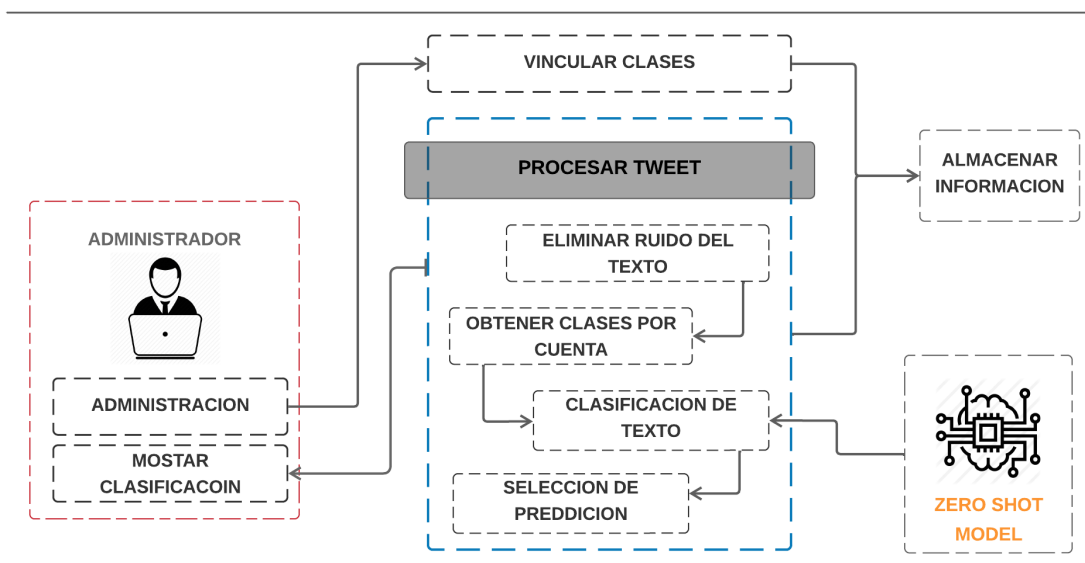


Figura 3.20: Diagrama de procesamiento del tweet

En la imagen anterior se representa el proceso que pasa el texto de un tweet después de haber sido obtenido en la etapa del Streaming, una vez ya definido el tipo de tweet y a que cuenta pertenece se procese a clasificar el texto obtenido en base a ciertos procesos y parámetros que iremos describiendo en las siguientes secciones, con el único fin de poder establecer y almacenar la clasificación correspondiente para cada una de las cuentas involucradas.

3.5.1 Eliminación de ruido del texto

Cuando se trabaja con modelos de predicción lo que se recomienda siempre es tratar de eliminar el mayor ruido posible del texto que vallamos analizar, esto debido a que los caracteres especiales y algunas palabras no aportan significado o sentido alguno para el texto original por lo que puede confundir los resultados de la predicción, en este caso lo que se busco fue eliminar URLs, Hashtags, Menciones o Caracteres especiales, tratando de no modificar el contexto original.

Es decir, debemos considerar ciertas excepciones que ayudan a mejorar el resultado, como es el caso de las Menciones, si bien un tweet puede tener un montón de cuentas mencionadas dentro del mismo texto, hay ocasiones donde esa mención forma parte especifica del sentido con la que se escribió el texto, por ejemplo en la siguiente una imagen se puede observar un Tweet con una sola mención a la cuenta @ETAPAOficial, que si se quitara esa mención prácticamente estriamos alterando el contexto original



Figura 3.21: Procesamiento de datos

En estos casos simplemente se elimina la “@” que tiene al principio para dejar el nombre de la mención intacto, es decir (ETAPAOficial), y lo mismo pasa con algunos caracteres que por lo general si ayudan a darle sentido al texto, pero en otras ocasiones simplemente forman parte de expresiones sin sentido que no aportan nada al contexto original, como por ejemplo “!!!!”, “...”, “,,,”, entre otros juegos de caracteres, en estos casos se eliminan por completo del texto.

Aquí otro ejemplo de un tweet que solo tiene menciones y ningún texto para ser analizado, este proceso es muy importante porque desde aquí podemos establecer los parámetros de aceptación para la predicción, es decir si no hay un texto para analizar ni siquiera debería pasar a la fase de clasificación, o si hay texto, pero sin un mínimo de palabras tampoco debería pasar a la siguiente fase.



Figura 3.22: Procesamiento de datos

A continuación, se presenta una descripción de los métodos aplicados para eliminar el ruido de un texto obtenido del tweet, aplicando técnicas como el REGEX que permitieron manipular el texto para conseguir el propósito establecido.

- **Eliminar URL**

El siguiente regex permite eliminar las urls del texto, sean citas o cualquier tipo de url.

```
r '[( http ( s ) ? ) : \ / \ / ( www \ . ) ? a - z A - Z 0 - 9 @ : % . _ \ + ~ # = ] { 2 , 256 } \ . [ a - z ] { 2 , 6 } \ b ( [ - a - z A - Z 0 - 9 @ : % _ \ + . ~ # ? & / / = ] * ) '
```

- **Eliminar Hashtags/Menciones**

El siguiente regex permite eliminar las menciones o hashtags que tenga el texto considerando los casos explicados anteriormente.

```
r "([@#]\w+)(\s+|$)"
```

- **Eliminar Signos de expresión**

El siguiente regex permite eliminar los signos de expresión que tenga el texto, siempre y cuando formen una secuencia de caracteres.

```
r "(\s * [^\w\s ] + \s *) { 2 , } | [ @ # & ! + / * ~ = ] "
```

- **Eliminar Caracteres especiales**

El siguiente regex permite eliminar los caracteres especiales que tenga el texto como tabulaciones, saltos de línea, entre otros.

```
r ' [\n\t\r\f\s ] + '
```

- **Eliminar emojis**

Para este caso se utilizo una libreria de python, que permite eliminar de un texto todos los caracteres especiales que conforman a un emoji.

```
from cleantext import clean
clean(texto , no_emoji=True , lower=False)
```

En la siguiente tabla se presenta unos ejemplos del texto normalizado, aplicando las diferentes consideraciones que mencionamos.

TEXTO ORIGINAL	TEXTO NORMALIZADO
#Atención @ETAPAOficial @cmt_machangara @pedropalaciosu @rubenbenitez_21 Felicitaciones @ETAPAOficial , felicitaciones @cmt_machangara , somos orgullosos por ser parte de esta "cuenca" que es desarrollo y sobretodo conservación, felicitaciones a su Presidente el Ing. ANTONIO BORRERO !!!.	Felicitaciones ETAPAOficial, felicitaciones cmt_machangara, somos orgullosos por ser parte de esta "cuenca" que es desarrollo y sobretodo conservacion, felicitaciones a su Presidente el Ing. ANTONIO BORRERO.
@QuejasCuenca @MunicipioCuenca Un llamado a @emov_ep en turi no existe control. Los autos estacionados los buses no pueden pasar. pic.twitter.com/zZuV9zh0vm	Un llamado a emov_ep en turi no existe control. Los autos estacionados los buses no pueden pasar.
@ecuarauz @forajida57 @Ashco121 @omaridrovo @uees_ec lo mismo pic.twitter.com/nWoEszJKSk	
#Atención #Cuenca #AguaPotable otra vez @ETAPAOficial con corte de agua twitter.com/ETAPAOficial/s	otra vez ETAPAOficial con corte de agua

Tabla 3.26: Ejemplos textos normalizados

3.5.2 Definición del proceso de clasificación

En este punto se describe el modelo de clasificación que hemos utilizado para este proyecto, si revisamos nuevamente el objetivo principal nos topamos con un tema de segmentación de clasificación, que no debemos pasar por alto debido a que cada cuenta que se vincula a la plataforma constituye a una empresa pública con un rubro de servicio en particular, por lo que, se debe contar con las clases específicas que ayuden a dicha entidad a clasificar los temas de su interés respecto a los servicios que brinda, por ende, se determinó que no se puede entrenar un modelo único de clasificación (al menos no de la forma convencional) para cada entidad sino más bien buscar otro método de entrenamiento que permita de alguna manera concebir en un solo modelo la clasificación de los tweets de todas las cuentas vinculadas en la plataforma, y para esto se utilizó el modelo Zero-Shot-Classification que como ya lo hemos mencionado permite predecir clases que no hayan sido entrenadas previamente dando como resultado la predicción de tweets sin importar el modelo de servicio que maneje la empresa pública.

A continuación se describe con un ejemplo del código desarrollado para implementar el uso del modelo de clasificación dentro de la plataforma.

```
# Importa la libreria
from transformers import pipeline
classifier = pipeline("zero-shot-classification",
                    model="facebook/bart-large-mnli")
# Clasifica el texto
texto = """En Cuenca todos los fines de semana salen a festejar"""
classifier(texto, multi_label=True,
          hypothesis_template="Este_ejemplo_trata_de_{ }.",
          candidate_labels=["politica", "fiestas", "viajes"]
)
# Resultados de prediccion
{'labels': ['fiestas', 'politica', 'viajes'],
 'scores': [0.9850454926, 0.1150539666, 0.0527010895]}
```

Con este punto comprendido, tenemos que aclarar que en esta fase se debe establecer las clases (o temas de interés) para cada cuenta que se vincula en la plataforma, estas clases son las que ayudaran al modelo a clasificar los tweets e identificar algún problema de servicio que tenga la institución pública, por lo que, solo el usuario administrador estará a cargo de establecer las clases necesarias para que la empresa pueda operar en la plataforma y visualizar las incidencias de su cuenta según el sesgo de servicio que maneje, si bien es un proceso que nos obliga a realizar un análisis previo de los tweets que se generan en la red social Twitter sobre alguna empresa pública en particular, esto nos permite establecer de una manera lógica las clases específicas para que el rendimiento de la plataforma sea lo indicado y que los datos estadísticos que se generen a partir de la vinculación sean lo más precisos posible.

A Continuación, se detalla de forma general el proceso de establecer las clases para una cuenta en específico.

1. Clase por defecto

Cuando una cuenta se vincula en la plataforma se crea automáticamente una clase denominada “sin clasificación”, esta clase permite a la plataforma identificar tweets que no hayan pasado los parámetros de aceptación para una coherente clasificación.

- **Es retweet:** Cuando un tweet obtenido es de tipo retweet no se debe tomar en cuenta para la clasificación porque a la final es un texto escrito por la misma empresa y no por otros usuarios.
- **Sin texto:** Cuando después del proceso de normalización de texto, no se obtiene resultado alguno, es decir que todo el texto estuvo conformado por ruido o no paso el mínimo de palabras requeridas.
- **Sin clases:** Cuando después del proceso de clasificación, ninguna clase resultante de la predicción paso el score mínimo permitido.

Cuando alguno de estos puntos se cumple el tweet se guardará automáticamente con esta única clase por defecto, con esto controlamos que la información que se analiza siempre tenga un valor agregado para el resultado de la clasificación.

2. Clase de sentimientos

Definir una clase específica para determinar el estado de ánimo con el que se escribió un tweet, es de alguna manera indispensable para poder identificar las problemáticas reales que tienen los usuarios respecto a los servicios de las empresas públicas que se vinculan en nuestra plataforma.

Si bien esta clase no se crea automáticamente en la vinculación de la cuenta, es responsabilidad del usuario administrador establecer la clase denominada “negativo” que permitirá identificar tweets con un contexto negativo y que podría estar netamente relacionada con una queja de algún servicio.

3. Clases de servicios

Para definir las clases específicas de una cuenta se debe analizar el rubro de servicio que maneja la entidad pública, y luego realizar pruebas de predicción con un set pequeño de datos para determinar la mejor clasificación de los temas de interés que se busca identificar.

Punto 1	Establecer los temas de servicios o interés
Punto 2	Establecer un nombre general que definan a esos temas de interés, ejemplo: internet, trafico, etc
Punto 3	Definir un conjunto pequeño de tweets que permitan realizar una prediccion utilizando las clases definidas en el punto dos.
Punto 4	Validar los resultados obtenidos y redefinir el nombre de la clase en caso de ser necesario.

Tabla 3.27: Puntos para definir las clases específicas

Estos puntos se deben repetir hasta llegar a definir de forma óptima las clases necesarias para cada una de las entidades públicas que se vinculen a la plataforma, una vez establecidas las clases de la institución se deben agregar a la plataforma para que puedan formar parte de la clasificación automática de tweets, para esto se debe tomar en cuenta que cada clase contara con unos parámetros específicos que deben ser considerados al momento de agregarlas a una cuenta vinculada, estos son:

- **Nombre:** Este parámetro establece el nombre de la clase con el cual se realizará la predicción.
- **Color:** Este parámetro permite establecer un color a la clase para poderla identificar de mejor forma en los resultados de la plataforma (listas/gráficas).
- **descripción:** Una descripción sobre el tema de interés que se busca identificar con la clase.
- **Score permitido:** Este parámetro establece un porcentaje mínimo de aceptación para los resultados de predicción, es decir si la predicción es menor a este score no será considerado como una predicción aceptable y por lo tanto no formará parte de los resultados de la clasificación.
- **Score seleccionado:** Este parámetro establece un porcentaje máximo de aceptación para los resultados de predicción, es decir si la predicción es mayor a este score automáticamente será considerado como una predicción válida.

Por último, en el caso de que el score obtenido este en un rango intermedio entre el score permitido y seleccionado, la clase si formara parte de los resultados de clasificación, pero no como una predicción válida sino como una aceptable.

Cuando una predicción es válida (o Seleccionada) la plataforma la mostrara de formas diferentes según la vista en la que se encuentre el usuario de tal manera que se resalte o se muestren los resultados con mayor importancia, por ejemplo.

- **Resultado de la clasificación en el live de Tweets:**

En la siguiente imagen se puede visualizar que el resultado de la predicción para la clase “tráfico” fue válido (Seleccionada) y por tal razón aparece con un componente completamente pintado, en cambio la clase “accidente” paso el score permitido, pero no con un score válido dando como resultado un componente parcialmente pintado.

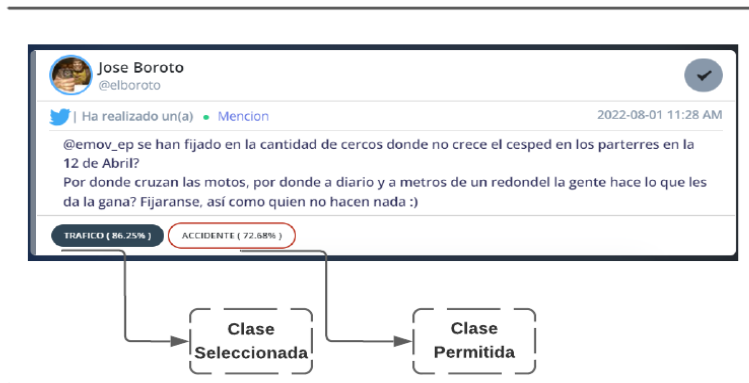


Figura 3.23: Resultado de clasificacion en Live Tweets.

- **Resultado de la clasificación en el detalle de Tweets:**

En la siguiente imagen se puede visualizar que el resultado de la predicción para la clase “servicios de agua” fue valido (Seleccionada) y por tal razón aparece con una línea de porcentaje más grande, a comparación del resultado de la clase “negativo” que si bien paso el score permitido no alcanzo un score valido dando como resultado una línea de porcentaje más delgada.



Figura 3.24: Resultado de clasificación en Detalle Tweets.

- **Resultado de la clasificación en la lista de Tweets:**

En el caso de la lista aplica exactamente la misma representación gráfica que la anterior descrita.

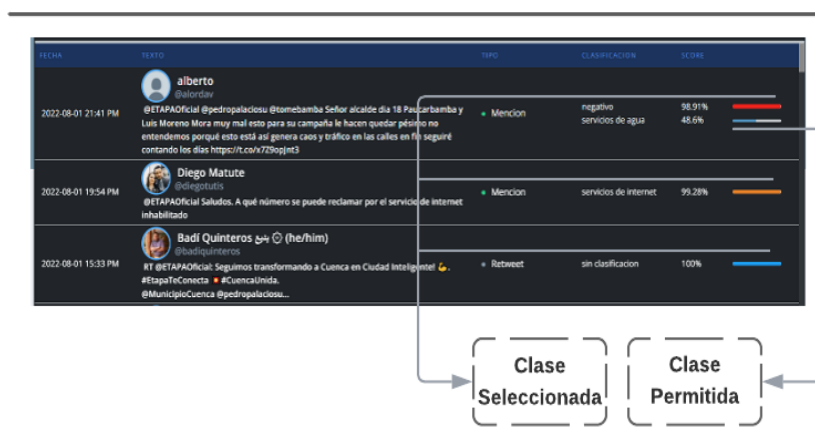


Figura 3.25: Resultado de clasificación en Lista Tweets.

Como podemos observar, el resultado de las clases que pasaron los parámetros de aceptación se muestra con un contexto grafico diferenciador a comparación de las otras clases que solo pasaron el score permitido, si bien es un tema notorio también es importante resaltar el hecho de que el color que se especifica cuando se agrega una clase es utilizado en la visualización de la información de la plataforma con el único objetivo de que el usuario pueda reconocer las clases de forma automática.

3.5.3 Aplicación del modelo de clasificación

Como parte de la ejemplificación de este proceso describiremos como se establecieron las clases para una de las cuentas que hemos gestionamos en este proyecto, la cuenta es @ETAPAOficial, este username de Twitter figura como la cuenta oficial de la empresa pública Etapa que es la encargada de brindar los servicios básicos como el agua potable, internet, telefonía, alcantarillado y saneamiento de aguas residuales, etc, a toda la comunidad cuencana y que actualmente cuenta con un total 67.4k Tweets publicados, con una descripción que dice “Mejorar la calidad de vida de las personas y contribuir al desarrollo de las organizaciones; proporcionando servicios innovadores y sustentables.”

A continuación, se describe una serie de casos que ejemplifican de mejor manera el proceso.

- **Caso de estudio #1:**

El siguiente tweet obtenido, es una ejemplificación exacta de un caso que hace referencia a un problema con el servicio de internet de la empresa pública Etapa, veamos como el modelo clasifica este tweet.

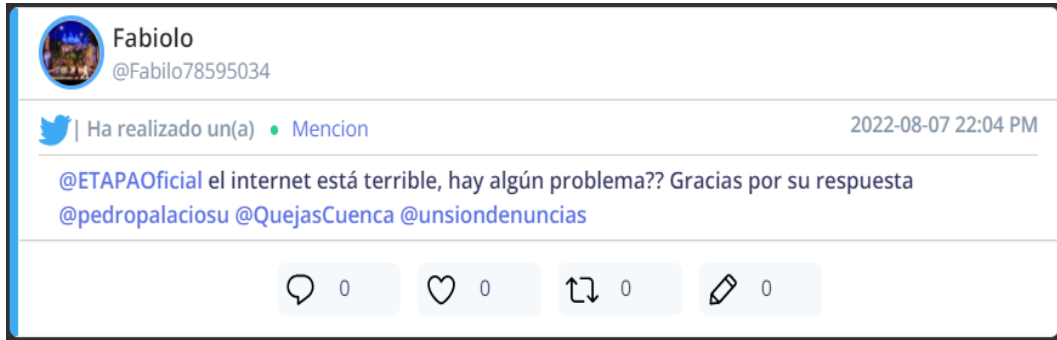


Figura 3.26: Tweet caso de estudio #1

– **Definición del Tweet:**

Cuenta	Tipo	Clases
@ETAPAOficial	Mención	"negativo", "servicios de agua", "servicios de internet", "servicios telefónicos", "sin clasificación"
@pedropalaciosu	Mención	No vinculada
@QuejasCuenca	Mención	No vinculada
@unsiondenuncias	Mención	No vinculada

Tabla 3.28: Cuentas involucradas en el Tweet

– **Normalización del texto:**

TEXTO ORIGINAL	TEXTO NORMALIZADO
@ETAPAOficial el internet está terrible, hay algún problema?? Gracias por su respuesta @pedropalaciosu @QuejasCuenca @unsiondenuncias.	ETAPAOficial el internet esta terrible, hay algún problema Gracias por su respuesta.

Tabla 3.29: Texto del tweet normalizado

– Resultados de predicción:

@ETAPAOficial	Score (P)	Score (S)	Resultado	Aceptación
negativo	10%	50%	90.47%	Seleccionado
servicios de internet	40%	50%	98.95%	Seleccionado
servicios de agua	40%	50%	11.48%	No Permitido
servicios telefonicos	40%	50%	0.06%	No Permitido
sin clasificación	0%	100%	No evaluada	No Permitido

Tabla 3.30: Resultados de predicción para la cuenta @ETAPAOficial

• Caso de estudio #2:

El siguiente tweet obtenido, es una ejemplificación exacta de un caso que hace referencia a un problema con el servicio de agua de la empresa pública Etapa, veamos como el modelo clasifica este tweet.



Figura 3.27: Tweet caso de estudio #2

– Definición del Tweet:

Cuenta	Tipo	Clases
@ETAPAOficial	Respuesta	"negativo", "servicios de agua", "servicios de internet", "servicios telefónicos", "sin clasificación"
@pedropalaciosu	Mención	No vinculada

Tabla 3.31: Cuentas involucradas en el Tweet

– Normalización del texto:

TEXTO ORIGINAL	TEXTO NORMALIZADO
@ETAPAOficial Buenos días Sres de @ETAPAOficial @pedropalaciosu Por favor de nuevo sin servicio de agua potable en el sector del CARMEN. Miércoles, jueves y hoy de nuevo viernes sin agua. Por favor 3 días con cortes.... Avisen por los menos para uno abastecer de agua.	ETAPAOficial Buenos dias Sres de Por favor de nuevo sin servicio de agua potable en el sector del CARMEN. Miercoles, jueves y hoy de nuevo viernes sin agua. Por favor 3 dias con cortes Avisen por los menos para uno abastecer de agua.

Tabla 3.32: Texto del tweet normalizado

– Resultados de predicción:

@ETAPAOficial	Score (P)	Score (S)	Resultado	Aceptacion
servicios de agua	40%	50%	99.9%	Seleccionado
negativo	10%	50%	87.69%	Seleccionado
servicios de internet	40%	50%	0.45%	No Permitido
servicios telefonicos	40%	50%	0.64%	No Permitido
sin clasificación	0%	100%	No evaluada	No Permitido

Tabla 3.33: Resultados de predicción para la cuenta @ETAPAOficial

• Caso de estudio #3:

El siguiente tweet obtenido, es una ejemplificación exacta de un caso que donde se ha retuiteado un tweet publicado por la empresa pública Etapa, veamos como el modelo clasifica este tweet.



Figura 3.28: Tweet caso de estudio #3

– **Definición del Tweet:**

Cuenta	Tipo	Clases
@ETAPAOficial	Retweet	"negativo", "servicios de agua", "servicios de internet", "servicios telefónicos", "sin clasificación"

Tabla 3.34: Cuentas involucradas en el Tweet

– **Normalización del texto:**

No se normaliza el texto cuando es tipo Retweet, porque no es un texto escrito por otro usuario.

– **Resultados de predicción:**

No se clasifica el texto cuando es tipo Retweet, por ende, se pone por defecto la clase "sin clasificación"

3.6 Administración de información

En la figura 3.29 se presenta el flujo base del proceso que maneja la plataforma, de tal manera que se centralizan los procesos descritos en las fases anteriores.

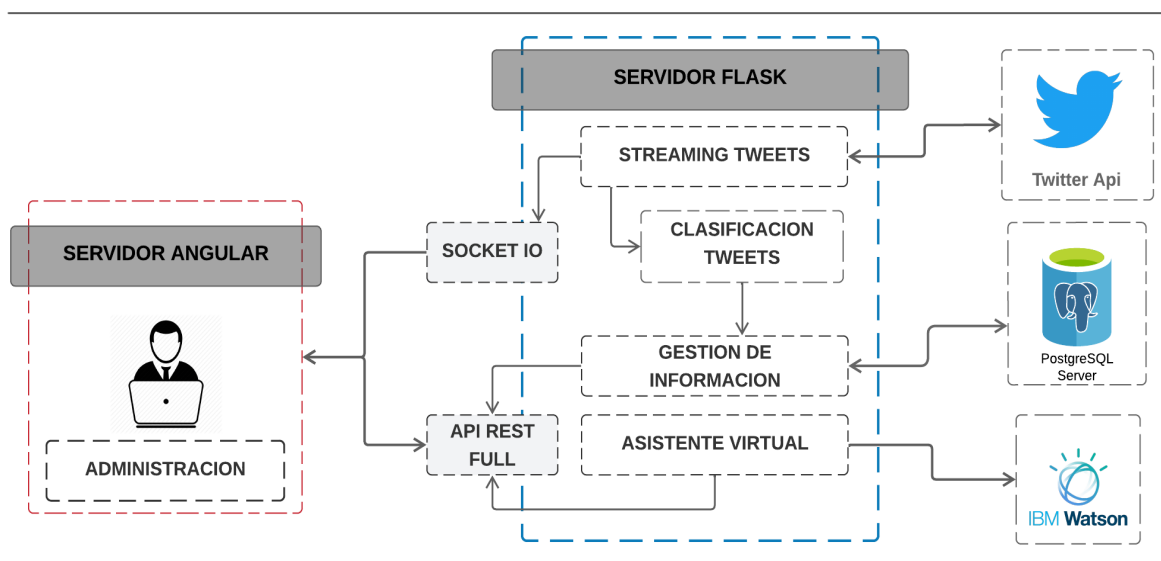


Figura 3.29: Diagrama administración de información

Como se describe en la gráfica anterior esta fase de administración centraliza todos los procesos de gestión y análisis que hemos detallado hasta este punto, dando como resultado la integración de 3 componentes base que conforman toda la plataforma, en primer lugar tenemos un servidor web (Flask) donde se levantaron los servicios necesarios para acceder a los procesos mencionados de la plataforma, el segundo es el servidor de base de datos (PostgreSQL) donde se ha almacenado toda la información obtenida de la plataforma y por último una aplicación basada en JavaScript (Angular) donde se desarrollaron las vistas de los requerimientos detallados en el proyecto implementado el consumo de los servicios desplegados que detallaremos en esta sección.

3.6.1 Descripción de APIs Rest Full

Todos las APIs Rest Full descritas en esta sección devuelve un objeto tipo JSON `"{'data': JSON}"` cuando servicio se ejecuta con normalidad, en el caso de ocurrir algún problema devuelve un mismo objeto tipo JSON `"{'error': string}"` con el error descrito.

- **URL:** <http://cloudcomputing.ups.edu.ec/api/usuario>

Método	Endpoints	Descripción
POST	/autenticar	Autenticar al usuario. – Recibe JSON: {"correo": string, "password": string}
POST	/agregar/cuenta	Vincula una cuenta a la plataforma. – Recibe JSON: {"cuenta_id": string, "username": string}

GET	/obtener/cuentas/{modo}	Obtiene las cuentas vinculadas del usuario. – Recibe URL param: modo:string
GET	/buscar/por/id/{id_cuenta}	Obtiene los datos de una cuenta. – Recibe URL param: id_cuenta:number
GET	/buscar/por/nombre/{username}	Busca una cuenta por nombre. – Recibe URL param: username:string
GET	/obtener/todos	Obtiene todos los usuarios de la plataforma.
POST	/agregar/registro	Agrega un nuevo usuario a la plataforma. – Recibe JSON: {"nombre": string, "password": string, "correo": string, "cuentas": list }
DELETE	/eliminar/{usuario_id}	Elimina un usuario registrado. – Recibe URL param: usuario_id:number
GET	/obtener/clases/{id_cuenta}	Obtiene las clases asociadas a una cuenta. – Recibe URL param: id_cuenta:number
POST	/agregar/clase/{id_cuenta}	Agrega una clase a un cuenta. – Recibe URL param: id_cuenta:number – Recibe JSON: {"nombre": string, "color": string, "descripcion": string, "min_score": number, "max_score": number }

DELETE	/eliminar/clase/{id_clase}	<p>Elimina una clase asociada a una cuenta.</p> <ul style="list-style-type: none"> – Recibe URL param: id_clase:number
GET	/obtener/acciones/{id_cuenta}	<p>Obtiene las acciones asociadas a una cuenta.</p> <ul style="list-style-type: none"> – Recibe URL param: id_cuenta:number
POST	/agregar/accion/{id_cuenta}	<p>Agrega una accioin a un cuenta.</p> <ul style="list-style-type: none"> – Recibe URL param: id_cuenta:number – Recibe JSON: {"codigo": string, "nombre": string, "descripcion": string, "json_filtros": string, json_reporte": string }
GET	/eliminar/accion/{id_accion}	<p>Elimina una accion asociada a una cuenta.</p> <ul style="list-style-type: none"> – Recibe URL param: id_accion:number
GET	/favorito/accion/{id_accion}	<p>Agrega una accion asociada a una cuenta como favorito.</p> <ul style="list-style-type: none"> – Recibe URL param: id_accion:number
POST	/chat/mensaje/{id_cuenta}	<p>Envia un mensaje al asistente virutla de watson.</p> <ul style="list-style-type: none"> – Recibe URL param: id_cuenta:number – Recibe JSON: {"mensaje": string, "nombre": chat_session}

Tabla 3.35: APIs Rest Full de gestion de Usuarios.

- **URL:** <http://cloudcomputing.ups.edu.ec/api/tweets>

Método	Endpoints	Descripción
GET	/total/valores/{id_cuenta}	Obtiene los valores totales de tweets por tipos. – Recibe URL param: id_cuenta:number
GET	/interacciones/{id_cuenta}	Obtiene los datos de interaccion. – Recibe URL param: id_cuenta:string – Recibe Query param: modo:string
POST	/clasificaciones/{id_cuenta}	Obtiene los datos de clasificacion segun los filtros. – Recibe URL param: id_cuenta:number – Recibe Query param: modo:string – Recibe JSON: {"filtros": dict}
POST	/por/filtro/{id_cuenta}/{pagina}	Obtiene los tweets segun los filtros. – Recibe URL param: id_cuenta:number – Recibe URL param: pagina:number – Recibe JSON: {"filtros": dict}
GET	/obtener/data/{id_cuenta}/{id_tweet}	Obtiene la informacion completa de un tweet. – Recibe URL param: id_cuenta:number – Recibe URL param: id_tweet:number

GET	/interacciones/cuentas/{id_cuenta}	Obtiene los datos de las cuentas de interaccion. – Recibe URL param: id_cuenta:number – Recibe Query param: modo:string – Recibe Query param: fecha:date
POST	/por/tipos/{id_cuenta}	Obtiene los tweets por tipos. – Recibe URL param: id_cuenta:number

Tabla 3.36: APIs Rest Full de gestion de Tweets.

- **URL:** <http://cloudcomputing.ups.edu.ec/api/reportes>

Método	Endpoints	Descripción
POST	/generar/por/accion/{id_cuenta}	Obtiene el reporte generado segun los parametros de una accion. – Recibe URL param: id_cuenta:number – Recibe JSON: {"reportes": dict, "filtros": dict}

Tabla 3.37: APIs Rest Full de gestion de Reportes.

Con la descripción de las APIs desarrolladas en este proyecto podemos establecer el contexto general que tiene la plataforma, ya que como describimos anteriormente la gestión de la información y el uso de los procesos se basa netamente en el consumo de los servicios que están desplegados en un servidor web, estos servicios a su vez hacen uso de un servidor de base de datos como contenedor de información del cual explicaremos más a fondo en la siguiente sección, pero que es un punto determinante que permite integrar la fase de presentación y almacenamiento de información dentro de la plataforma.

3.6.2 Almacenamiento de información

El siguiente diagrama entidad relación describe las tablas utilizadas para la gestión de la información dentro este proyecto, si bien el proceso desarrollado esta netamente enfocado al análisis de tweets es importante aclarar que se tuvo que gestionar otro tipo de información para el manejo correcto de la plataforma, como es el caso de los usuarios, cuentas, clases, etc. Sin mencionar el hecho de que todos los tweets recopilados en el Streaming también se guardan en la base de datos, así como se especificó en las secciones anteriores.

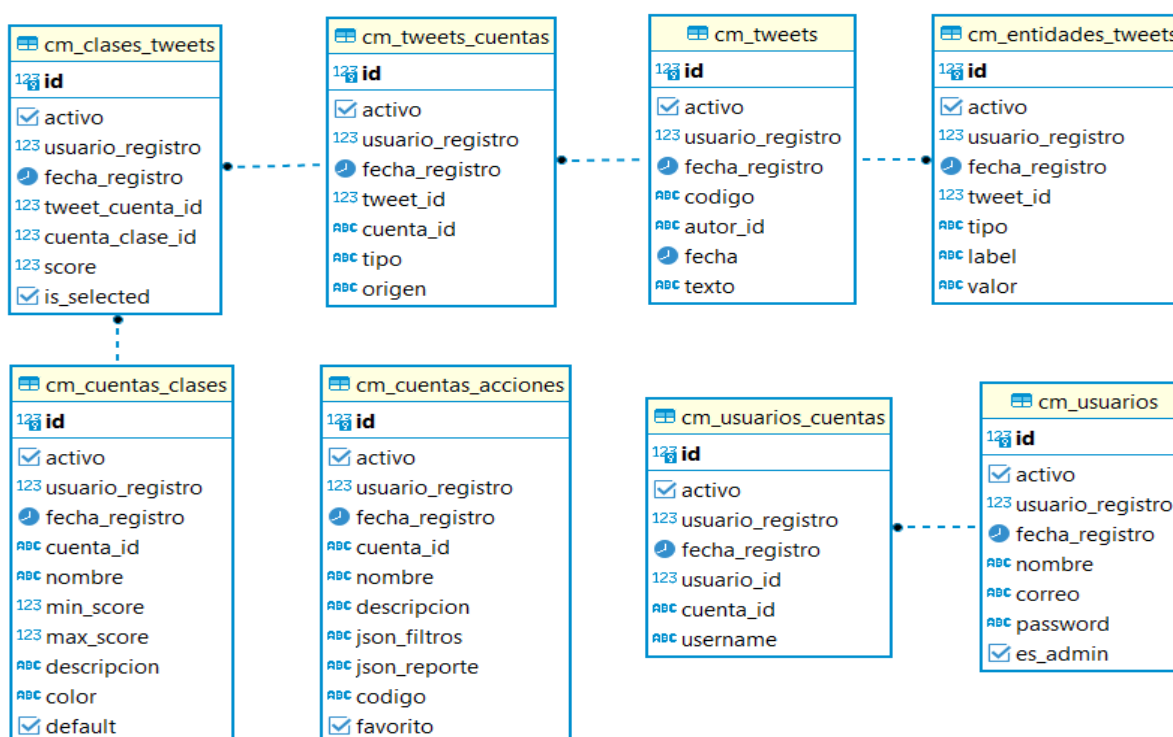


Figura 3.30: Diagrama entidad relacion "Plataforma Web"

Recapitulando el proceso de administración, los servicios desarrollados en esta la plataforma hacen uso constante del componente de base de datos definida en la imagen anterior ya sea para obtener o gestionar la información que se recopila/maneja en la plataforma, permitiendo que el servidor de base de datos solo sea visto como un contenedor de información separando por completo de la lógica del proyecto así como se describe en el diagrama de la arquitectura de la aplicación.

Descripcion de tablas

Descripción de la tabla "cm_usuarios", usada para la gestión de usuarios que usen la plataforma.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
nombre	varchar	not null
correo	varchar	not null
password	varchar	not null
es_admin	boolean	default false

Tabla 3.38: Campos tabla "cm_usuarios".

Descripción de la tabla "cm_usuarios_cuentas", usada para la gestión de cuentas vinculadas a la plataforma.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
usuario_id	number	not null
cuenta_id	varchar	not null
username	varchar	not null

Tabla 3.39: Campos tabla "cm_usuarios_cuentas".

Descripción de la tabla "cm_cuentas_clases", usada para la gestión de clases que manejen las cuentas vinculadas.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
cuenta_id	varchar	not null
nombre	varchar	not null

color	varchar	not null
descripcion	varchar	not null
min_score	number	
max_score	number	
default	boolean	default false

Tabla 3.40: Campos tabla "cm_usuarios_cuentas".

Descripción de la tabla "cm_tweets", usada para la gestión de tweets recopilados.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
codigo	varchar	not null
autor_id	varchar	not null
fecha	datetime	not null
texto	varchar	not null

Tabla 3.41: Campos tabla "cm_tweets".

Descripción de la tabla "cm_entidades_tweets", usada para la gestión de componentes del texto obtenido de los tweets como por ejemplo url, menciones, hastag, lugares, etc.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
tweet_id	number	not null
tipo	varchar	not null
label	varchar	not null
valor	varchar	not null

Tabla 3.42: Campos tabla "cm_entidades_cuentas".

Descripción de la tabla "cm_tweets_cuentas", usada para la gestión de tipos de tweets según la cuenta.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
tweet_id	number	not null
cuenta_id	varchar	not null
tipo	varchar	not null
origen	varchar	not null

Tabla 3.43: Campos tabla "cm_tweets_cuentas".

Descripción de la tabla "cm_clases_tweets", usada para la gestión de los resultados de clasificación según las clases de las cuentas vinculadas.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
tweet_cuenta_id	number	not null
cuenta_clase_id	number	not null
score	number	not null
is_selected	boolean	default false

Tabla 3.44: Campos tabla "cm_clases_cuentas".

Descripción de la tabla "cm_cuentas_acciones", usada para la gestión de acciones que manejen las cuentas vinculadas.

Columna	Tipo	Atributos
id	number	not null, increment
activo	boolean	default true
usuario_registro	number	
fecha_registro	datetime	
cuenta_id	varchar	not null
nombre	varchar	not null
codigo	varchar	not null
descripcion	varchar	not null
json_filtros	varchar	not null

json_reporte	varchar	not null
favorito	boolean	default false

Tabla 3.45: Campos tabla "cm_cuentas_acciones".

Los SQL de las tablas descritas anteriormente estarán adjuntos en la documentación de este proyecto, y en el repositorio oficial que se estará dejando en las referencias.

3.6.3 Presentación de informacion

Por último, tenemos la fase de presentación, esta fase consiste principalmente en la construcción de una aplicación web basada en Angular donde se desarrolló las vistas necesarias para presentar al usuario toda la información obtenida en el proceso de clasificación de tweets, ya sea mediante gráficos o listas este componente permitirá al usuario interactuar con los procesos principales de la plataforma que han sido descritos a lo largo de este documento.

Antes de centrarnos en el proceso del desarrollo de estas interfaces es importante resaltar que aquí se consumen todos los servicios que hemos descrito en la sección anterior, de tal manera que se debe establecer algún método de autorización para poder consumir estos servicios, si bien no hemos descrito o mencionado esta parte de autenticación es importante entender que la plataforma cuenta con la implementación de un Token de autorización que permite consumir los servicios solo a usuarios que hayan pasado el proceso de autenticación dentro de la plataforma.

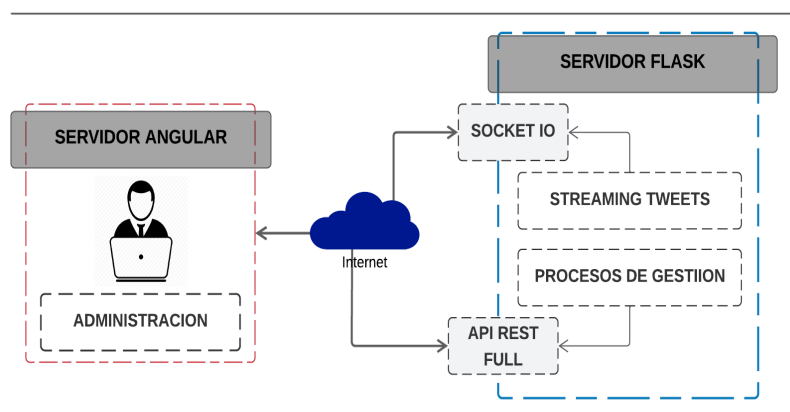


Figura 3.31: Proceso consumo servicios REST.

A continuación, se presenta los prototipos de las pantallas principales desarrolladas en la plataforma y que engloban todos los requerimientos detallados en la sección anterior.

- Prototipo base

Este prototipo muestra los componentes bases que tendrán todas las pantallas de la plataforma, que son menú lateral de opciones, buscador, nombre de página y cuenta vinculada.

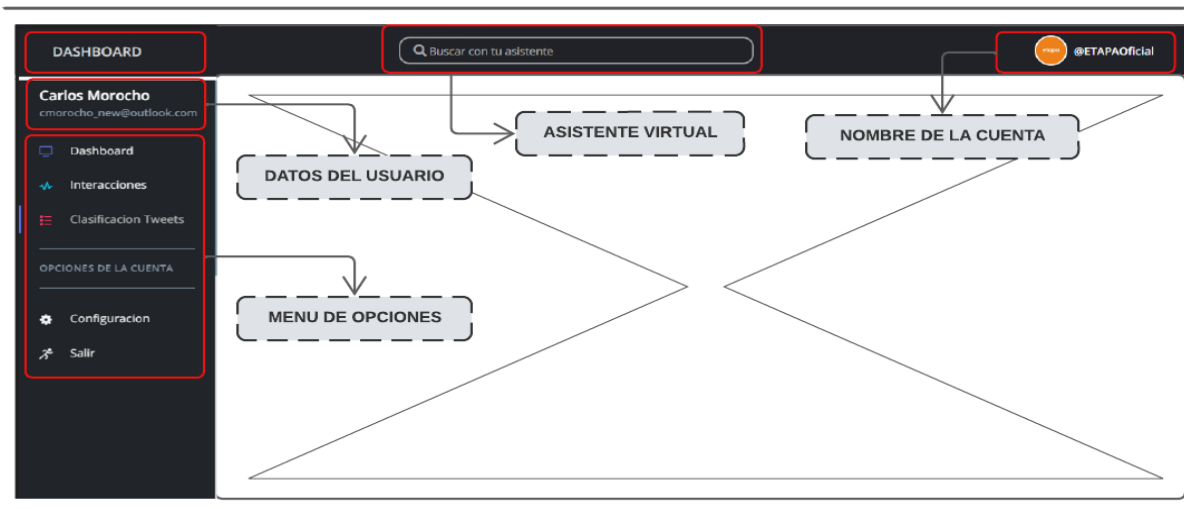


Figura 3.32: Prototipo componentes base.

- Prototipo pantalla Live de Tweets.

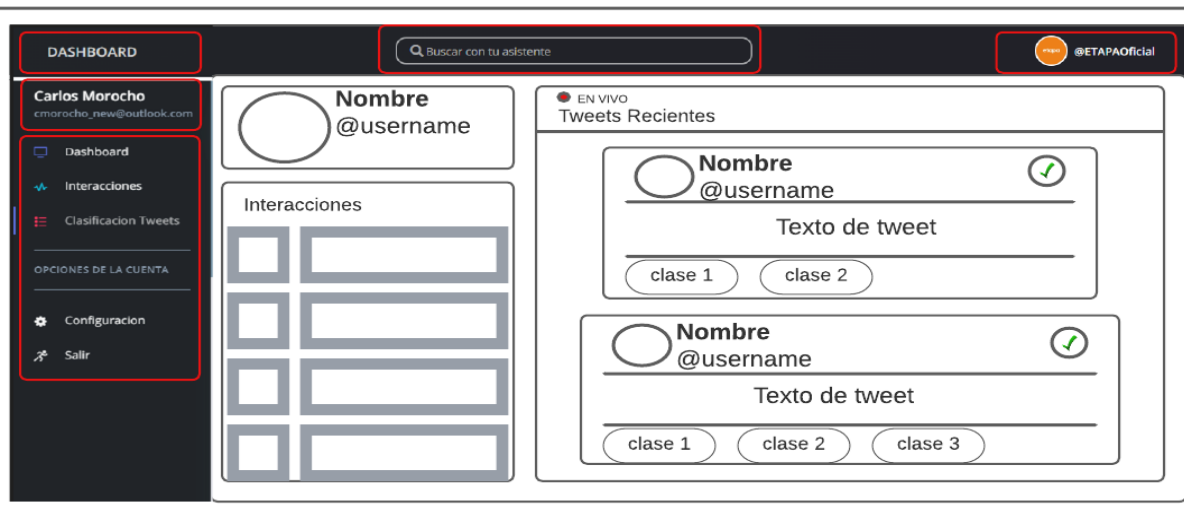


Figura 3.33: Prototipo pantalla Live.

- Prototipo pantalla Interaccion de Tweets.

En este prototipo se representa la gráfica de interacción de tweets que tenga la cuenta vinculada.

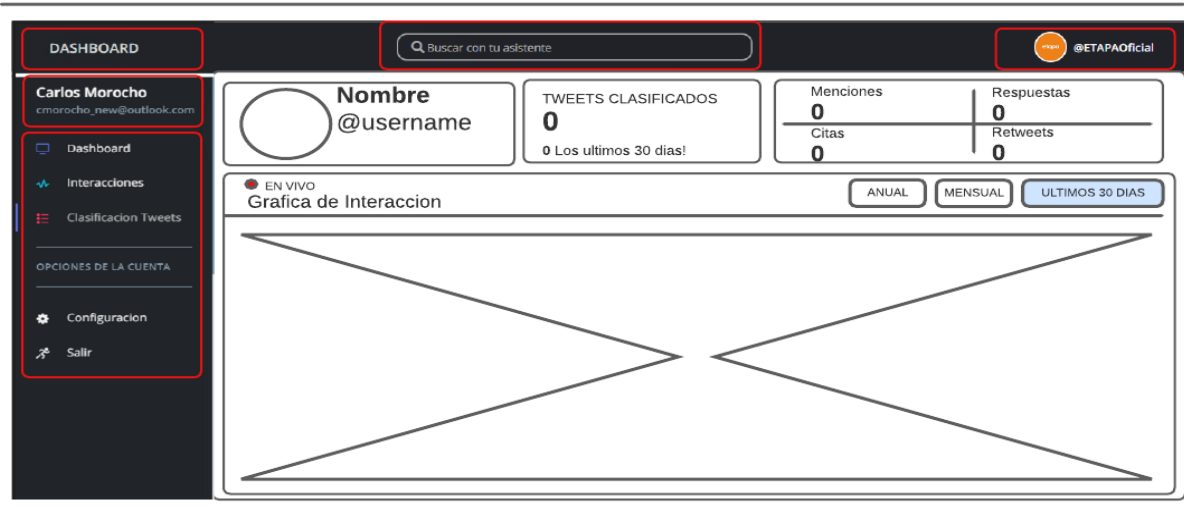


Figura 3.34: Prototipo pantalla Interacciones.

- Prototipo pantalla clasificación de Tweets.

Este prototipo representa la pantalla donde se visualiza el histórico de tweets, es decir todos los tweets recopilados con las respectivas graficas de barras y líneas.

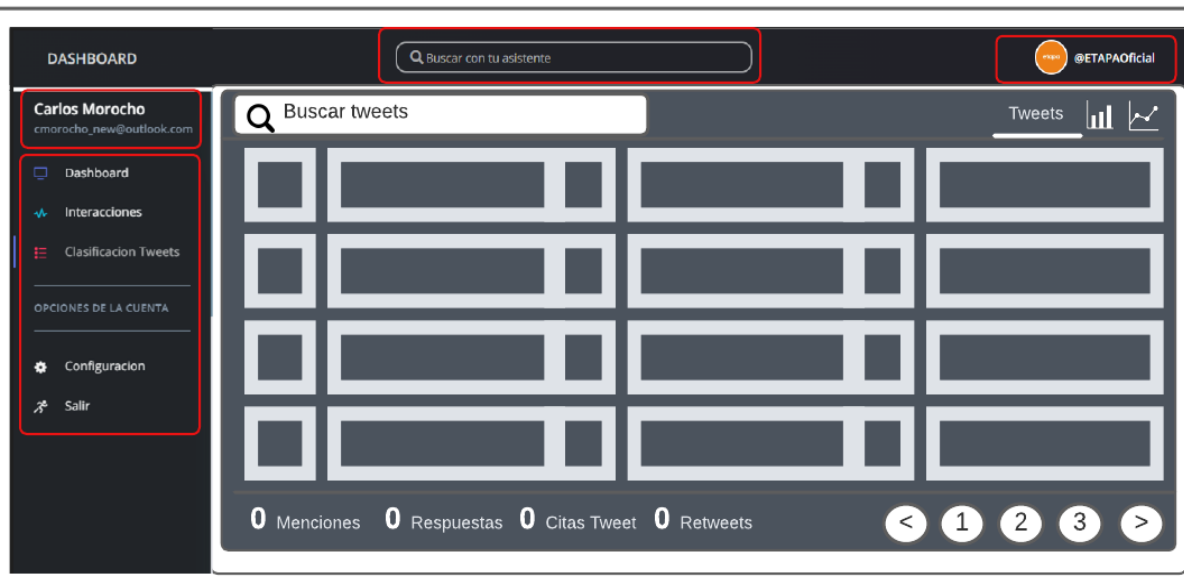


Figura 3.35: Prototipo pantalla Clasificación de Tweets.

Esta sección pertenece a los gráficos de la clasificación, tanto de Barras como de líneas.

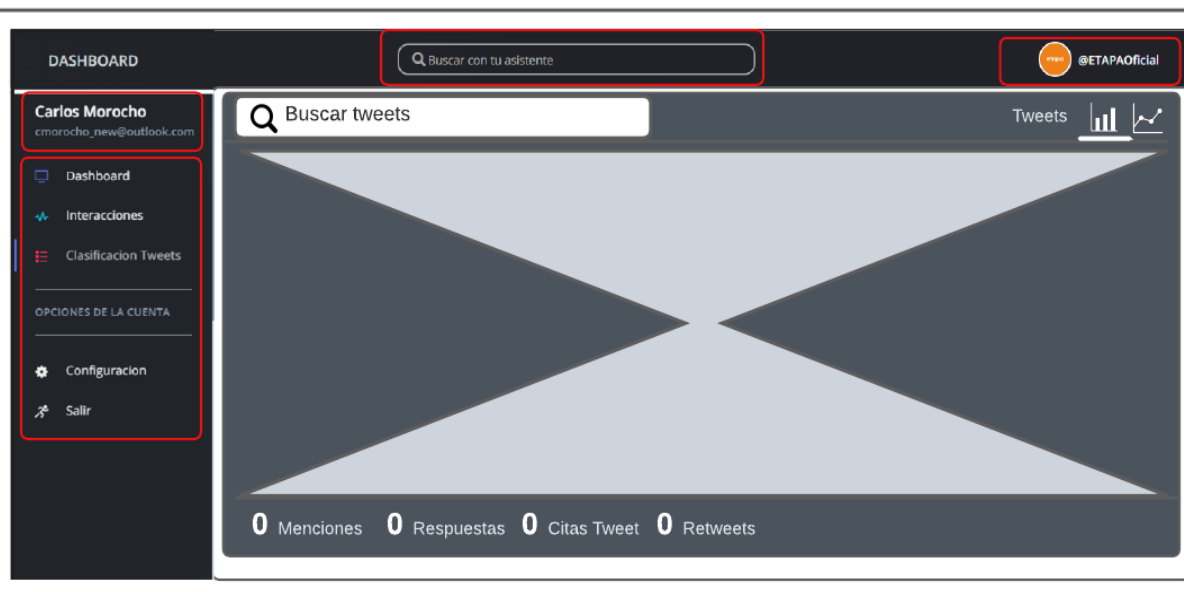


Figura 3.36: Prototipo grafico de clasificación.

Hasta este punto ya hemos descrito de manera general la fase de presentación implementada en esta plataforma por lo que podemos tener un contexto más claro de cómo es que se desarrollaron las diferentes vistas/pantallas para el uso integral de los procesos descritos, si bien en esta sección solo se presentó los prototipos de las pantallas principales se dejara en la sección de anexos una descripción más detallada todas las vistas desarrolladas para este proyecto.

Detalle de vistas desarrolladas

En esta sección se detallara los accesos a las páginas de la plataforma, los cuales son.

URL: http://cloudcomputing.ups.edu.ec/cm_clasificador

URL Acceso	Descripción	Permisos
/login	Pagina de ingreso a la plataforma	ninguno
/dashboard	Pagina de visualización del live de tweets.	autenticación
/interacciones	Pagina de visualización de interacciones.	autenticación
/tweets	Pagina de visualización de tweets clasificados.	autenticación
/configuración	Pagina de configuración de clases, cuentas, usuarios y acciones.	autenticación, administrador

Tabla 3.46: URL de accesos a las páginas de la plataforma.

Capítulo IV

ANÁLISIS DE RESULTADOS

En este capítulo se presentan cada uno de los resultados obtenidos del modelo de clasificación implementado en el análisis de los tweets relacionados con la empresa pública @ETAPAOficial la cual se ha venido utilizando para la ejemplificación del proceso que realiza la plataforma, para poder describir de mejor manera la precisión del modelo alcanzado se ha realizado una clasificación manual y luego se ha evaluado con los resultados de la clasificación automática obtenidos.



CLASES				Agregar
NOMBRE	DESCRIPCION	SCORE (P)	SCORE (S)	
negativo	Clase para identificar tweets con un contexto de servicio	10	50	🗑️
servicios de agua	Clase para etiquetar los tweets del servicio eléctrico	40	50	🗑️
servicios de internet	Clase para etiquetar los tweets referentes al servicio de internet	40	50	🗑️
servicios de telefonía	Clase para etiquetar los tweets del servicio de agua potable	40	50	🗑️
sin clasificacion	Clase para identificar los tweets que no tienen clasificacion	0	100	

Figura 4.1: Clases configuradas de la cuenta @ETAPAOficial.

4.1 Resultados de clasificación clase "negativo"

Tweet	Modelo	Humano
@ETAPAOficial Porque cortan el agua potable, sin ningún aviso, aquí en sector LAS ORQUÍDEAS??	SI	SI

@GeotechTips @pedropalaciosu @ETAPAOficial @estebanfzl @uazuay @upsalesiana con tanta riqueza hidrica en los 4 ríos, nos contemplamos a sólo verlos contaminados y erosionados	SI	SI
Hoy puse el internet de @ETAPAOficial y hoy mismo se rompió el cable de fibra del poste. Comenzamos mal, muy mal.	SI	SI
@ETAPAOficial @rubenbenitez_21 Hasta qué hora se espera señores, todo el día sin agua. Q es chiste	SI	SI
#ATENCIÓN @ETAPAOficial informa que habrán cortes agua hasta el día de mañana 11 de agosto del 2022 a las 09h00 en los sectores alimentados por la Planta de MACHÁNGARA debido a una rotura de tubería por los trabajos de construcción del puente La Compañía #agua #cortedeagua pic.twitter.com/GXa8HF3PdZ	SI	NO
@pedropalaciosu @tomebamba @ETAPAOficial buenas tardes , llegamos al feriado del 10 de agosto y no terminan llegamos al día 29 señor alcalde y así luce la PAUCARBAMBA y LUIS MORENO MORA y nadie da una explicación del porqué está así la calle gracias por la atención a este tweet pic.twitter.com/ImdAtcPH1A	SI	SI
#Cuenca Boletín de @ETAPAOficial dónde a conocer los sectores afectados, tras la rotura de tubería por la construcción del PUENTE LA COMPAÑÍA pic.twitter.com/PEtgdF4Xs4	SI	NO

<p>@pedropalaciosu @ETAPAOOficial Que cuando botan la casa comunal de la Mutualista Azuay.....</p>	SI	NO
<p>@ETAPAOOficial el internet está terrible, hay algún problema?? Gracias por su respuesta @pedropalaciosu @QuejasCuenca @unsiondenuncias</p>	SI	SI
<p>#Atención Tres días se encuentra sin agua potable la comunidad de PUICAY, situada a un kilómetro del centro de CHALLUABAMBA población reclama a @ETAPAOOficial urgente atención a su reclamo. Han recurrido hasta al agua de la lluvia. pic.twitter.com/EMV1CF7z4s</p>	SI	SI
<p>@ETAPAOOficial @pedropalaciosu @tomebamba Señor alcalde día 18 PAUCARBAMBA y LUIS MORENO MORA muy mal esto para su campaña le hacen quedar pésimo no entendemos porqué esto está así genera caos y tráfico en las calles en fin seguiré contando los días pic.twitter.com/x7Z9opJnt3</p>	SI	SI
<p>@ETAPAOOficial @tomebamba @pedropalaciosu Señor alcalde hoy retiraron los montículos de tierra de la paucarbamba el día número 16 pero sigue la calle inhabilitada porque ?? Cual es el problema etapa no se pronuncia no dice nada por favor informen cuál es el problema gracias</p>	SI	SI

Hacemos un llamado urgente a @ETAPAO-ficial @pedropalaciosu @MunicipioCuenca para que retiren los cables de un poste que está a punto de caer en la JUAN LEON MERA y PEDRO CARBO, CORAZÓN DE MARÍA. @tomebamba @elmercurioec @NuevoTiempoCue @GuermanPiedra pic.twitter.com/TIOkdDQdxF	SI	SI
@pedropalaciosu @tomebamba @ETAPAO-ficial Señores etapa por favor indicar cuándo van a terminar de arreglar la calle , producto de una fuga de agua hace 13 días #. Por favor den una respuesta llevo días preguntándoles ave paucarbamba Girasol pic.twitter.com/B4ynG8F1IG	SI	SI
@ETAPAO-ficial sin líquido vital en el sector del parque industrial, ciudadela Uncovia	SI	SI
@rubenbenitez_21 @pedropalaciosu @ETAPAO-ficial Desde hace más de 20 días estamos a la espera de la respuesta sobre el número de empleados con discapacidad que mantiene la institución así también como el cumplimiento a la ordenanza de discapacidad en relación a la compra pública inclusiva sin respuesta	SI	SI

@rubenbenitez_21 @pedropalaciosu @ETA-PAOficial hace más de 20 días entregamos por parte de este observatorio la consulta del empleo de las personas con discapacidad en su institución así también como el cumplimiento de la ordenanza de discapacidad con respecto a las compras públicas incluso no tenemos respuesta	SI	SI
[ATENCIÓN] @ETAPAOficial informa que desde las 13:00 hasta las 20:00 de este sábado 23 de julio de 2022 hasta intermitencias y falta de servicio de agua potable en varios sectores de TOTORACOCHA. Aquí los lugares que tendrán la falta del líquido vital pic.twitter.com/IUG2qunw0n	SI	NO
#Atención @ETAPAOficial informa que hoy sábado de 13H00 A 20H00 habrá interrupción del servicio de agua potable en sectores abastecidos por el Tanque de #Totoracocha por un daño imprevisto ic.twitter.com/II2FPiEigf	SI	NO
@ETAPAOficial comunica falta de servicio de agua potable en sectores abastecidos desde el tanque de Totoracocha. pic.twitter.com/nZj5wh78hU	SI	NO

Tabla 4.1: Tweets clasificados con clase "negativo".

En la siguiente grafica se presenta el flujo de incidencias tipo "negativo" que se ha obtenido de los tweets analizados.

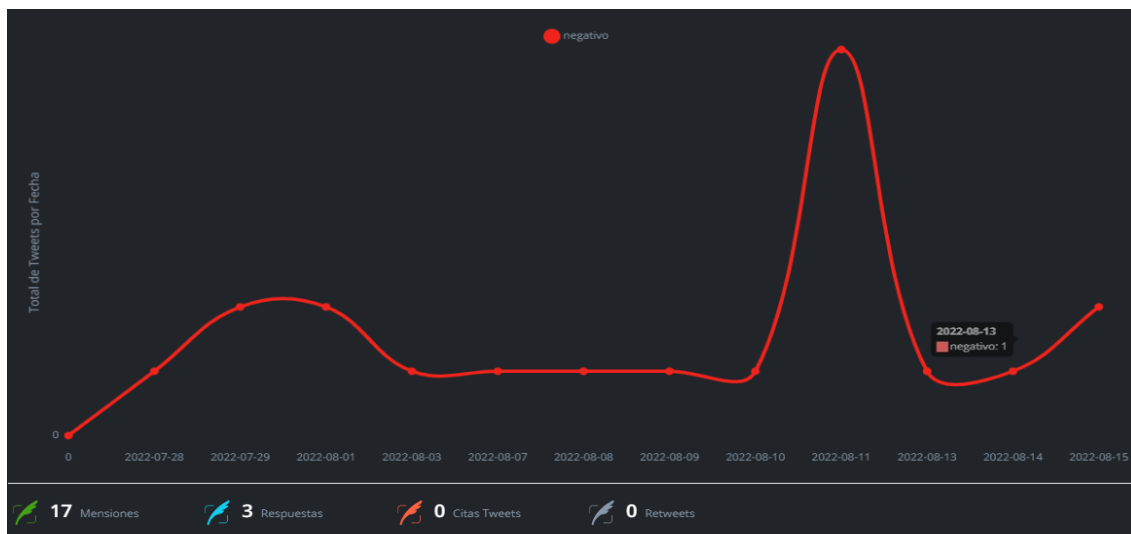


Figura 4.2: Grafica de incidencias "negativo".

4.2 Resultados de clasificación clase "servicios de agua"

Tweet	Modelo	Humano
@elmercurioec Linda su huevada @ETAPAO-ficial	SI	NO
@ETAPAOoficial buenos días, los moradores del sector VIRGEN DE BRONCE estamos sin servicio de agua potable desde el día viernes, sacaron una tubería y dejaron allí botado el trabajo, son las 10 am y todavía no intentan darnos solución @tomebamba	SI	SI
@ETAPAOoficial Porque cortan el agua potable, sin ningún aviso, aquí en sector LAS ORQUÍDEAS??	SI	SI
@ETAPAOoficial @rubenbenitez_21 Por favor den una solución definitiva a este sector, ya que cada rato tenemos esos problemas	SI	NO

@ETAPAOficial @rubenbenitez_21 Hasta qué hora se espera señores, todo el día sin agua. Q es chiste	SI	SI
#ATENCIÓN @ETAPAOficial informa que habrán cortes agua hasta el día de mañana 11 de agosto del 2022 a las 09h00 en los sectores alimentados por la Planta de MACHÁNGARA debido a una rotura de tubería por los trabajos de construcción del puente La Compañía #agua #cortedeagua pic.twitter.com/GXa8HF3PdZ	SI	SI
@ETAPAOficial Por favor el agua de la CIU-DADELA PUMAYUNGA se encuentra en este estado... como se observa en la foto. Que sucede con Etapa. pic.twitter.com/sEi6vB5e6A	SI	SI
@rubenbenitez_21 @ETAPAOficial @pedropalaciosu Muy bien RUBÉN	SI	NO
¡Cuenca unida! Junto al equipo de Agua Potable de @ETAPAOficial, esta tarde, recibí a los moradores de CRUZ LOMA DE LLA-CAO, con quienes dialogamos sobre las obras que ejecutaremos en la zona para fortalecer nuestros servicios de provisión de líquido vital. pic.twitter.com/gMu65HGL16	SI	SI
@ETAPAOficial @MunicipioCuenca @pedropalaciosu Sres pregunté amablemente hace horas que en que momento restablecen el servicio de agua potable por el sector de la empresa eléctrica! Ni un comunicado ni nada. Todo el dia sin agua!	SI	SI

<p>@ETAPAO oficial @MunicipioCuenca @pedropalaciosu Por favor. Sector empresa eléctrica sin agua desde la mañana. A que hora se reestablece el servicio? Gracias</p>	SI	SI
<p>¡Cuidamos el agua de CUENCA! Bajo la visión del Alcalde de Cuenca, @pedropalaciosu, de construir una ciudad amigable con el medio ambiente, hoy nuestra empresa @ETAPAO oficial firmó dos importantes convenios, como parte del COMITÉ DE CONSERVACIÓN DE LA CUENCA DEL YANUNCAY. pic.twitter.com/whlew1hilv</p>	SI	SI
<p>#Atención Tres días se encuentra sin agua potable comunidad de PUICAY, situada a un kilómetro del centro de CHALLUABAMBA a población reclama a @ETAPAO oficial urgente atención a su reclamo. Han recurrido hasta al agua de la lluvia. pic.twitter.com/EMV1CF7z4s</p>	SI	SI
<p>Reunión de trabajo, para sostener un enlace entre los los veedores @ETAPAO oficial @CpccsEc dentro de este proceso de veeduría VIGILAR EL PROCESO DE CONTRATACIÓN LICO-PTARG-001-2020 PARA CONSTRUCCIÓN Y PUESTA EN MARCHA DE LANTA DE TRATAMIENTO DE AGUAS RESIDUALEAS GUANGARCUCO pic.twitter.com/JcRgSYxXpL</p>	SI	NO

<p>@ETAPAOficial Buenas tardes Sres de Etapa. En este momento se acaba de ir el agua en el sector del cruce.... Comenzamos de nuevo con los cortes de agua.</p>	SI	SI
<p>@ETAPAOficial @tomebamba @pedropalaciosu Buenos días hace 14 días hubo una rotura de matriz de agua en la AVENIDA PAUCARBAMBA y LUIS MORENO MORA se hizo el arreglo de la matriz pero la avenida continúa cerrada con un montículo de tierra cual es el motivo señores etapa gracias pic.twitter.com/mooCCWd7mL</p>	SI	SI
<p>@pedropalaciosu @tomebamba @ETAPAOficial Señores etapa por favor indicar cuándo van a terminar de arreglar la calle , producto de una fuga de agua hace 13 días #. Por favor den una respuesta llevo días preguntándoles ave paucarbamba Girasol pic.twitter.com/B4ynG8F1IG</p>	SI	NO
<p>@ETAPAOficial GRACIAS A LA MEJOR EMPRESA DE AGUA POTABLE DEL ECUADOR. pic.twitter.com/71bA8sBvWY</p>	SI	SI
<p>¡Más obras para CUENCA! Junto al Alcalde @pedropalaciosu, esta tarde, como @ETAPAOficial firmamos el contrato para la construcción del sistema de agua potable y alcantarillado, más conexiones domiciliarias, en el sector Quinta Chica Baja - Escudo Nacional - Machángara. pic.twitter.com/icQvkF1dLM</p>	SI	SI

Tabla 4.2: Tweets clasificados con clase "servicios de agua".

En la siguiente grafica se presenta el flujo de incidencias tipo "servicios de agua" que se ha obtenido de los tweets analizados.

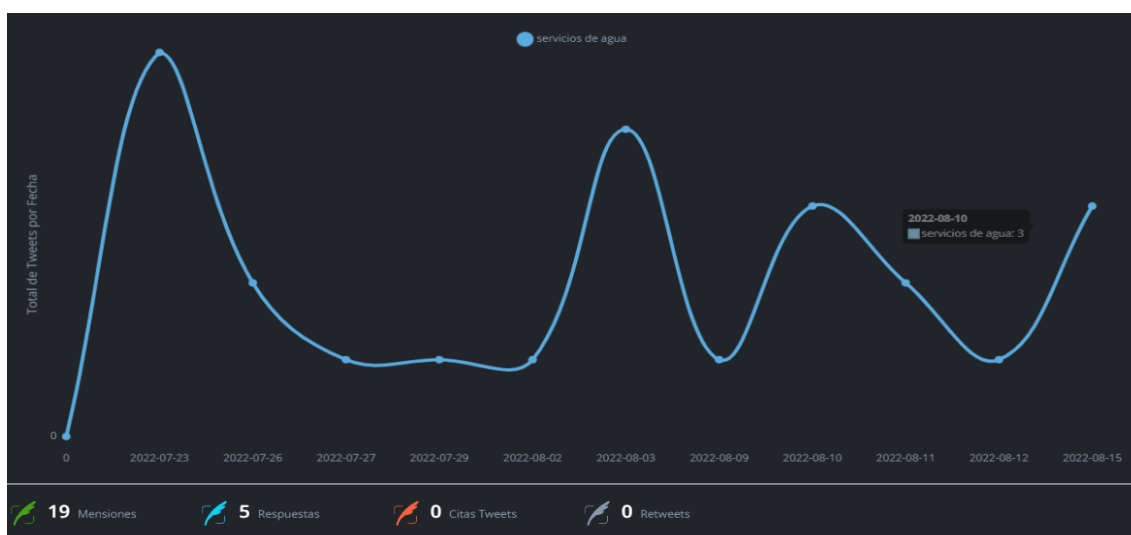


Figura 4.3: Grafica de incidencias "servicios de agua".

4.3 Resultados de clasificación clase "servicios de internet"

Tweet	Modelo	Humano
Hoy puse el internet de @ETAPAOficial y hoy mismo se rompió el cable de fibra del poste. Comenzamos mal, muy mal.	SI	SI
@ETAPAOficial @rubenbenitez_21 El link de drive esta caído por favor actualizar.	SI	SI
@ETAPAOficial el internet está terrible, hay algún problema?? Gracias por su respuesta @pedropalaciosu @QuejasCuenca @unsiondenuncias	SI	SI

@ETAPAOficial @MunicipioCuenca @pedropalaciosu @rubenbenitez_21 Simplificar trámites es hacerlo todo ONLINE !!	SI	SI
@ETAPAOficial Saludos. Xfa su ayuda restaurando el servicio de internet en el barrio Buena Fe. Calle Cantón El Empalme. Un carro del municipio arranco un cable https://t.co/dvNBh1Yn59	SI	SI
@ETAPAOficial Saludos. A qué número se puede reclamar por el servicio de internet inhabilitado	SI	SI
@ETAPAOficial Buenas tardes...por favor si me pueden ayudar el Internet esta demasiado lento...sector parque Iberia...saludos	SI	SI
PUBLICIDAD Conoce tu nuevo @ETAPAOficial Descubre más en el siguiente video https://t.co/1oBihB34XU	SI	NO
Hoy en sesión de trabajo, conjuntamente con el Gerente de Telecomunicaciones de @ETAPAOficial Ing Michael Cabrera y su equipo técnico, para generar proyectos comunicacionales (internet) y mancomunar acciones en beneficio de la ruralidad. #JuntosPorElDesarrollo https://t.co/wGbjllg3iY	SI	SI
@ETAPAOficial Hasta ahora nada que regresa	SI	NO
@ETAPAOficial buenas tardes, estoy sin Internet x el sector del colegio Santa Ana.. Me ayudan con información? O un telefono al cual contesten?	SI	SI

@ETAPAOficial gracias por su servicio, es perfecto https://t.co/NTWpAPpi8F	SI	NO
@NetlifeEcuador @ETAPAOficial @pedropalaciosu @Denunciascuen @unsiondenuncias Realmente el problema es con el internet de @ETAPAOficial que como verán no han respondido, pero muchas gracias por estar atentos	SI	SI
@ETAPAOficial La red ETAPAFi-access se ha convertido en un problema, o la abren al público gratuitamente por un espacio de tiempo o la retiran. La señal de Etapa interfiere con el plan de datos, ya que espera credenciales, y hay que desactivar el Wi Fi, para no perder la comunicación	SI	SI
@ETAPAOficial El pago de \$ 4 es mensual?	SI	NO
PUBLICIDAD Internet fuera de casa gracias a @ETAPAOficial Conoce más en el siguiente video https://t.co/9hLr0Ju1Qc	SI	SI
¡Cuenca unida! Hoy, junto al equipo de @ETAPAOficial, acompañamos al Alcalde de Cuenca, @pedropalaciosu, en la Minga Fest del parque Curiquingue. En este espacio atendimos requerimientos ciudadanos relacionados con la provisión de nuestros servicios de agua e internet. https://t.co/IosyFBhUlm	SI	SI

#Atención @ETAPAO oficial sr de @ETAPAO oficial otra vez problemas con el Internet hasta para pagarles a ustedes de forma online tengo problemas pago mes a mes la tarifa q ustedes me cobran pero no recibo el servicio q ofrecen si les pagaré por el Internet q recibo mi pago seria de 1\$ al mes	SI	NO
@ETAPAO oficial Ya se solucionó.. Gracias	SI	NO
@Observadis @ETAPAO oficial SRES. DEL OBSERVATORIO SOY UN PROFESIONAL CON DISCAPACIDAD, POR FAVOR UNA VEZ RECIBIDA LA INFORMACIÓN SOLICITADA PUBLIQUEN LO RECIBIDO, ADEMÁS TENEMOS MUCHO TRABAJO CON EL SECTOR PRIVADO, A LAS ÓRDENES PARA AYUDARLES.	SI	NO

Tabla 4.3: Tweets clasificados clase "servicios de internet".

En la siguiente grafica se presenta el flujo de incidencias tipo "servicios de internet" que se ha obtenido de los tweets analizados.

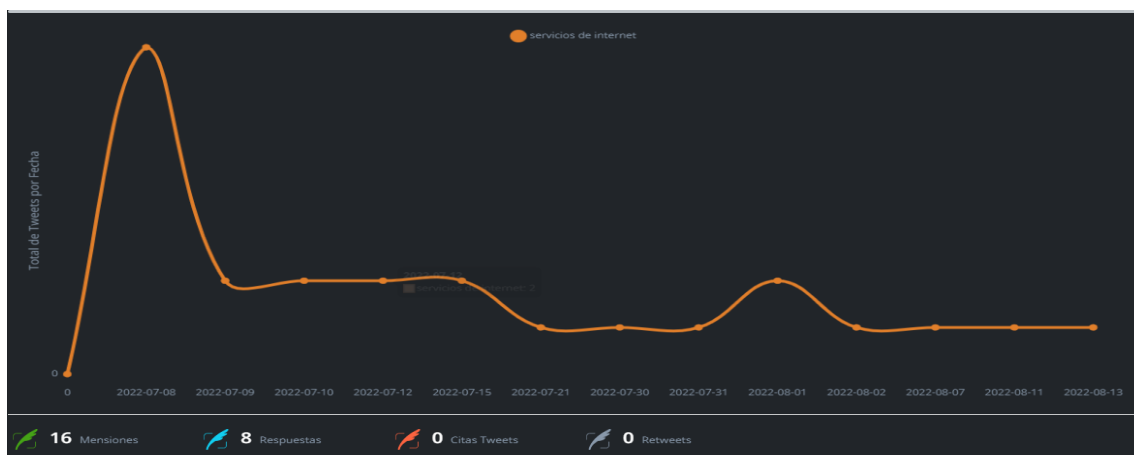


Figura 4.4: Grafica de incidencias "servicios de internet".

4.4 Resultados de clasificación clase "servicios de telefonía"

Tweet	Modelo	Humano
@ETAPAOficial @rubenbenitez_21 El link de drive esta caído por favor actualizar.	SI	NO
@ETAPAOficial buenas tardes, estoy sin Internet x el sector del colegio Santa Ana.. Me ayudan con información? O un telefono al cual contesten?	SI	SI
@ETAPAOficial Ya se está volviendo una costumbre el corte en nuestro sector del Ecu 911, lo peor que no comunican con un tiempo prudente	SI	NO
@ETAPAOficial @pedropalaciosu Buenas noches, mi número es 0997911785	SI	NO
@ETAPAOficial @rubenbenitez_21 <i>Buenasnoches</i> , 0998275228	SI	NO
He estado más de 15 días intentado comunicarme con @ETAPAOficial @EtapaE y no recibo ninguna respuesta, sus teléfonos pasan con un problema, ya sea de sistema o de conmutador. https://t.co/64naakFA8m	SI	SI
@ETAPAOficial 0995093983 o 0995093981	SI	NO
@ETAPAOficial Si, gracias por la gestión	SI	NO
@ETAPAOficial Mi contacto les dejo por mensaje interno	SI	SI
@ETAPAOficial @MunicipioCuenca @pedropalaciosu @rubenbenitez_21 No es cierto! En Challuabamba no hay servicio y en los teléfonos no contestan.	SI	SI

@ETAPAOficial La normalidad!! https://t.co/u7180A62y0	SI	NO
@ETAPAOficial @rubenbenitez_21 Nueva- mente estamos sin agua por challuabamba. La semana pasada igual estuvimos sin agua. No hay como conectarse al 072834700. Por favor su atención	SI	SI
@pedropalaciosu @ETAPAOficial Este men- saje están enviando por WhatsApp. Por favor enviar comunicado oficial. @elmer- curioec @tomebamba @radialvision @an- tenaunofm @super949 @LaSuprema961 @ECU911Austro https://t.co/MaLpre6Q5	SI	NO
@ETAPAOficial @centrosur_ec La dirección es av. Pumapungo entre calle alba calderon y rayoloma mi número es 098-274-2060	SI	NO

Tabla 4.4: Tweets clasificados clase "servicios de telefonía".

En la siguiente grafica se presenta el flujo de incidencias tipo "servicios de telefonía" que se ha obtenido de los tweets analizados.



Figura 4.5: Grafica de incidencias "servicios de telefonía".

En la sección anterior hemos escogido como referencia 20 tweets a la zar, para clasificarlos de forma manual e identificar los falsos positivos de cada clase, dando como resultado la siguiente información.

Clase	VP	FN	Recall
negativo	14	6	0.7
"servicios de agua"	15	5	0.75
"servicios de internet"	13	7	0.65
"servicios de telefonía"	5	9	0.35

Tabla 4.5: Cuentas involucradas en el Tweet

A continuación, una gráfica que representa el total de tweets clasificados desde que se puso en producción la plataforma en los servidores del Grupo de Investigación en Cloud Computing, Smart Cities & High Performance Computing (GIHP4C)

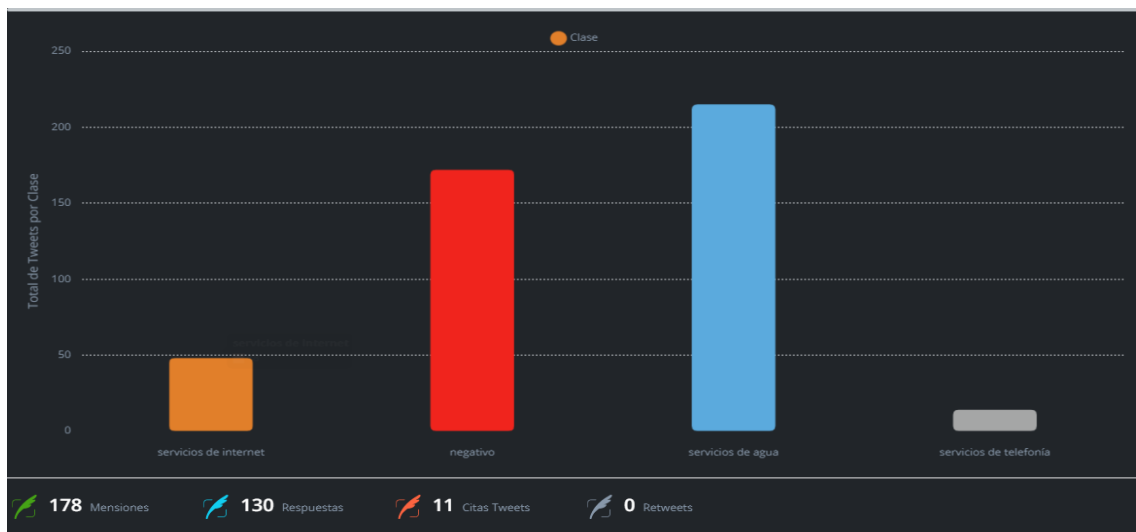


Figura 4.6: Grafica de clasificación de tweets (Histórico).

Capítulo V

CRONOGRAMA DE ACTIVIDADES

Nombre de la tarea	Duración	Comienzo	Fin
Proyecto	1040 horas	26/04/2021	16/08/2022
OE. 1	100 horas	26/04/2021	04/08/2021
ACT 1. Aprender sobre procesamiento de lenguaje natural.	20 horas	26/04/2021	20/07/2021
ACT 2. Investigación sobre cómo obtener la información de la red social Twitter mediante APIs de integración disponibles.	20 horas	01/06/2021	20/06/2021
ACT 3. Investigación de la herramienta FastText y la API de integración proporcionada para manipular el modelo según el lenguaje escogido.	30 horas	01/05/2021	30/07/2021
ACT 4. Investigación sobre una herramienta para desarrollar un chatbot personalizado y poder integrarlo a las tecnologías escogidas.	30 horas	25/06/2021	30/07/2021
OE. 2	180 horas	18/10/2021	10/02/2022

ACT 1. Seleccionar un lenguaje de programación backend e implementar la API de integración para obtener los comentarios de la red social Twitter.	30 horas	18/10/2021	10/02/2022
ACT 2. Desarrollar algoritmos de aprendizaje de máquina para el reconocimiento de tópicos e Implementar el modelo de aprendizaje PLN en el lenguaje seleccionado.	100 horas	18/10/2021	10/02/2022
ACT 3. Comprobar y realizar pruebas de funcionamiento del modelo con datos extraídos en el en la tarea 1 y validar los datos con la clasificación de quejas.	50 horas	18/10/2021	10/02/2022
OE. 3	500 horas	26/05/2021	30/06/2022
ACT 1. Definir requerimientos para el desarrollo de la aplicación web..	40 horas	26/05/2021	10/12/2021
ACT 2. Diseñar y desarrollar la aplicación web para el procesamiento de datos (backend).	220 horas	26/05/2021	30/06/2022
ACT 3. Diseñar y desarrollar la aplicación web para la visualización de datos (frontend).	220 horas	26/05/2021	30/06/2022
ACT 4. Despliegue de las aplicaciones web en los servidores del grupo de investigación.	20 horas	27/06/2022	01/07/2022
OE. 4	160 horas	01/07/2022	22/07/2022

ACT 1. Desarrollar el funcionamiento del asistente virtual en la aplicación backend.	80 horas	01/07/2022	22/07/2022
ACT 2. Entrenar al asistente con la información del procesamiento de comentarios.	20 horas	14/07/2022	22/07/2022
ACT 3. Implementar la interfaz interactiva del chatbot en la aplicación web (frontend).	60 horas	14/07/2022	22/07/2022
OE. 5	100 horas	18/07/2022	10/08/2022
ACT 1. Pruebas de funcionamiento y validación de la aplicación.	20 horas	18/07/2022	21/07/2022
ACT 2. Desarrollo de documentación de la aplicación web.	80 horas	22/07/2022	10/08/2022

Capítulo VI

PRESUPUESTO

DENOMINACIÓN	CANTIDAD	COSTO UNITARIO	COSTO TOTAL
	unidades	dólares	dólares
1. Bienes			
Copias	200	0,05	10,00
Impresiones	500	0,10	50,00
2. Tecnológico			
Laptop	1	1300,00	1300,00
3. Servicios			
Cursos, Libros,etc.	5	50,00	250,00
Servicios de Internet	12	31,00	372,00
5. Otros			
Imprevistos	1	100,00	100,00
Total	719	\$ 1481,15	\$ 1982,00

Capítulo VII

CONCLUSIONES

Los modelos de clasificación de texto basados en el aprendizaje de Disparo Zero han permitido implementar técnicas de PLN con mayor facilidad, considerando que se pudo identificar clases que ni siquiera han sido entrenadas previamente el modelo desarrollado por Facebook "bart-large-mnli" tuvo una gran precisión en los resultados del análisis realizados por la plataforma, sin embargo se tuvo que realizar un proceso de análisis previo para poder definir una clase específica que le permitiera al modelo obtener los resultados deseados.

En este sentido, la principal limitación que se tuvo con este modelo fue poder definir estas clases específicas que pudiera ayudar a identificar una problemática en particular, si bien los resultados descritos en este documento muestran una clara precisión no se considera que sea una solución general para toda empresa pública, en este sentido la plataforma estaría dispuesta a mejorar aplicando diferentes técnica o modelos de clasificación dependiendo la entidad que se vinculese.

Por último, el procesamiento de lenguaje natural aplicado a las redes sociales implica alcanzar un grado mayor de probabilidad al momento de tomar las decisiones, esto debido a que la información que se puede generar con el análisis de clasificación es en todo sentido más precisa y real porque al final del día la verdadera problemática siempre estará en los usuarios, y este proyecto es un claro ejemplo de esta implementación.

Capítulo VIII

RECOMENDACIONES

Lo que se busca con este tipo de proyectos es poder ayudar a mejorar un sector específico de servicio que se gestiona dentro de las empresas públicas, por lo que tratar de generalizar una problemática específica siempre da lugar a otra más complicada, en estos casos se recomienda enfocarse en una empresa o ámbito concreto que permita visualizar la solución propuesta.

Se recomienda analizar bien el modelo de clasificación que se valla a utilizar, ya que si bien existen un numero extenso de modelos basados en el aprendizaje de Disparo-Zero no se considera que todos obtengan buenos resultados, para este proyecto se probó con 5 modelos diferentes y solo uno alcanzo los resultados esperados.

REFERENCIAS

Agenda Pública, g.-i. (2020). ¿para qué usan twitter los partidos en campaña?

Diego Calvo, D. (2019a). Clasificación de redes neuronales artificiales.

Diego Calvo, D. (2019b). Metodología scrum (metodología ágil).

Dorys Moreira¹, Ivan Cruz, K. G. A. Q. (2021). Análisis del estado actual de procesamiento de lenguaje natural.

Giannini, S. (2021). Inteligencia artificial y educación.

Hermández R, P. (2006). Recolección y procesamiento de datos.

Internet Live Stats, N. (2022). Twitter usage statistics - internet live stats.

JIMENEZ, B. L. A. C. P. B. D. A. G. T. B. J. C. M. P. B. B. A. S. (2022). Sistema web utilizando la metodología scrum para mejorar la gestión de los equipos informáticos en la institución educativa 14053-cucungara de cura mori piura-2022.

Juan Pablo Del Alcázar Ponce, M. C. (2022). Estado digital ecuador - abril 2022.

Juan Vorobioff, Santiago Cerrotta, N. E. M. A. A. (2022). Inteligencia artificial y redes neuronales.

MAMANI, B. C. R. Y. (2022). Aplicación móvil con tecnología rest para el monitoreo y gestión de servicio educativo en escuelas de la ugel azángaro 2019.

MARQUIDIA BALAREZO, D. L. (2015). Las redes sociales como nueva opción de promoción de las microempresas en la ciudad de cuenca.

Power Data, n. (2022). Transformación digital.

Sancho, J. L. V. (2014). La visualización de datos.

Suárez, J. D. (2017). Redes neuronales convolucionales en reconocimiento de caracteres escritos a mano.

Toledo, A. H. U. (2021). Detección de objetos en imágenes mediante aprendizaje sin ejemplos.

Twitter, I. (2022). Search Tweets - How to build a query.

Anexos

Anexos A

Script de python para el reproceso de tweets

```
# Importar librerias
```

```
import re
```

```
import psycpg2
```

```
import unidecode
```

```
from cleantext import clean
```

```
from transformers import pipeline
```

```
# Id de la cuenta a procesar
```

```
CUENTA_ID = '963007182'
```

```
# REGEX TO DELETE NOICE
```

```
REGEX_1 = r'[(http(s)?):\/\/(www\.)?a-zA-Z0-9@:%.\_+~#={2,256}\.[a-z]{2,6}\b([-
```

```
# Urls
```

```
REGEX_2 = r"(([@#]\w+)(\s+|$)){2,}"
```

```
REGEX_3 = r"(\s*[\^w\s]+\s*){2,}|[@#&!+/*~#="
```

```
REGEX_4 = r'[\n\t\r\f\s]+'
```

```
# Modelo clasificacion
```

```
classifier = pipeline("zero-shot-classification",
```

```
                        model="facebook/bart-large-mnli")
```

```
# Crea la conexion a la BD
```

```

connection = psycopg2.connect(user="postgres",
                               password="postgres",
                               host="127.0.0.1",
                               port="5432",
                               database="cm_clasificador")

```

Normaliza el texto

```

def normalizar(texto):
    #print(texto)
    texto = clean(texto, no_emoji=True, lower=False)
    texto = re.sub(REGEX_1, "_", texto)
    texto = re.sub(REGEX_2, "_", texto)
    texto = re.sub(REGEX_3, "_", texto)
    texto = re.sub(REGEX_4, "_", texto)
    texto = texto.strip()
    return texto if len(texto.split("_")) > 2 else ''

```

Obtiene el cursor de conexion

```

cursor = connection.cursor()

```

Obtiene las clases de la cuenta

```

cursor.execute("""select c.id, c.nombre, c.min_score, c.max_score, c.default
from cm_cuentas_clases c
where cuenta_id = %s
order by c.default""", (CUENTA_ID, ))
default_clase, labels = None, {}
for c in cursor.fetchall():
    if c[4]:
        default_clase = c[0]
    else:

```

```

labels[c[1]] = {'id': c[0], 'min_score': c[2], 'max_score': c[3]}

# Obtiene los tweets de la cuenta
cursor.execute("""select ctc.id, ctc.tipo, cct.texto from cm_tweets cct
join cm_tweets_cuentas ctc on ctc.tweet_id = cct.id
where ctc.cuenta_id = %s""", (CUENTA_ID, ))
tweets = cursor.fetchall()

# Elimina la clasificacion anterior
cursor.execute("""delete from cm_clases_tweets where tweet_cuenta_id in (select
join cm_tweets_cuentas ctc on ctc.tweet_id = cct.id
where ctc.cuenta_id = %s)""", (CUENTA_ID, ))
connection.commit() # Guarda

def clasificar_texto(texto):
    clasificacion = []
    resultado = classifier(
        texto, hypothesis_template = "Este_ejemplo_trata_de_{}.",
        candidate_labels=list(labels.keys()), multi_label=True
    ) # Predice el texto
    prediccion = dict(zip(resultado['labels'], resultado['scores']))
    print(prediccion)
    for clase in prediccion:
        # Obtiene el score prediction
        score = float('{:.2%}'.format(prediccion[clase]).replace("%", ""))
        # Compara con los scores min y max
        if score >= labels[clase]['min_score']:
            is_selected = score >= labels[clase]['max_score']
            clasificacion.append((labels[clase]['id'], score, is_selected))
    # Retorna la clasificacion

```

```

    return clasificacion

def guardar_clasificacion(id_tweet, clasificacion):
    sql_insert = f"""
    INSERT INTO public.cm_clases_tweets
    (activo, usuario_registro, fecha_registro, tweet_cuenta_id, cuenta_clase_id,
    VALUES(true, 0, now(), {id_tweet}, %s, %s, %s);"""

    for clasif in clasificacion:
        cursor.execute(sql_insert, clasif)
    connection.commit() # Guarda

# Clasifica los tweets
print ("EMPIEZA_LA_CLASIFICACION")
for tweet in tweets:
    clasificacion = [(default_clase, 100, True)]
    if tweet[1] != 'RT':
        texto = normalizar(tweet[2])
        # texto_normalizado = normalizar(texto)
        if texto:
            clasificaciones = clasificar_texto(texto)
            if clasificaciones:
                if len(clasificaciones) == len(labels):
                    clasificacion = [clasificaciones[0]]
                else:
                    clasificacion = clasificaciones
        # Guarda la clasificacion
        guardar_clasificacion(tweet[0], clasificacion)
print ("TERMINA_LA_CLASIFICACION")

```

Anexos B

Interfaces de Usuario

2.1 Aplicación Web

2.1.1 Inicio de sesión

La siguiente imagen nos muestra la pagina para poder iniciar sesión en la plataforma web.

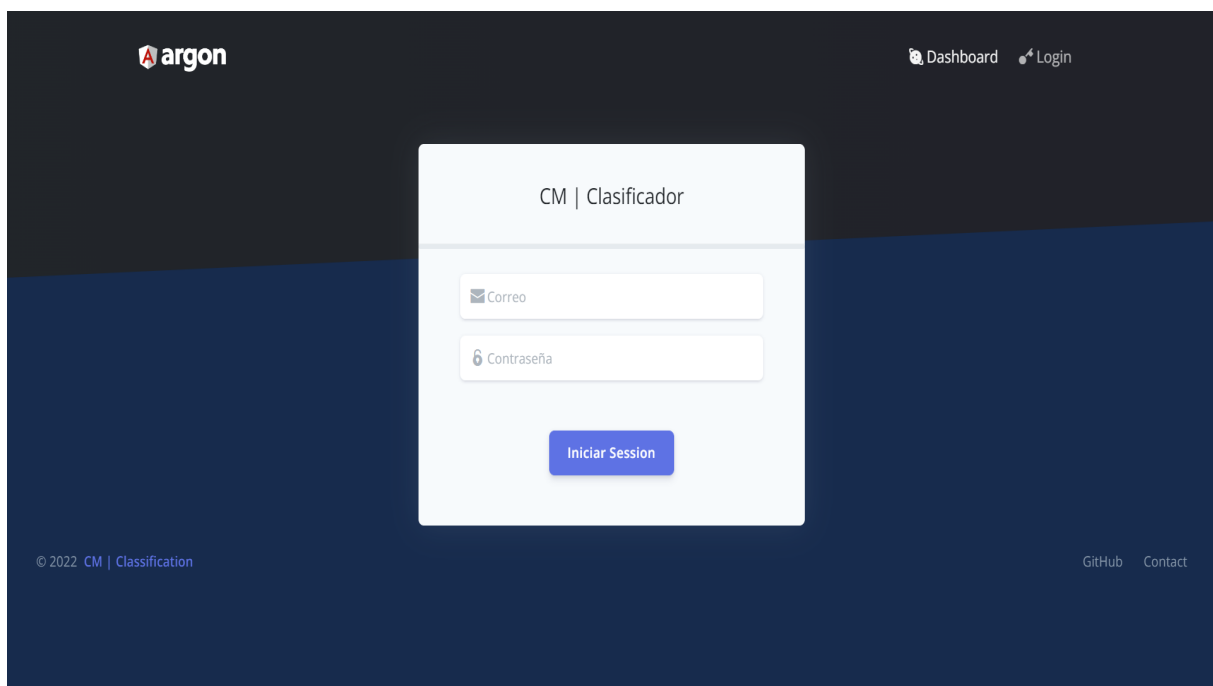


Figura 2.1: Interfaz inicio de sesión.

2.1.2 Dashboard de tweets en vivo

En la siguiente imagen se puede observar el panel donde se representan los tweets recopilados en tiempo real, y en la sección izquierda se puede visualizar las cuentas que interactuaron.

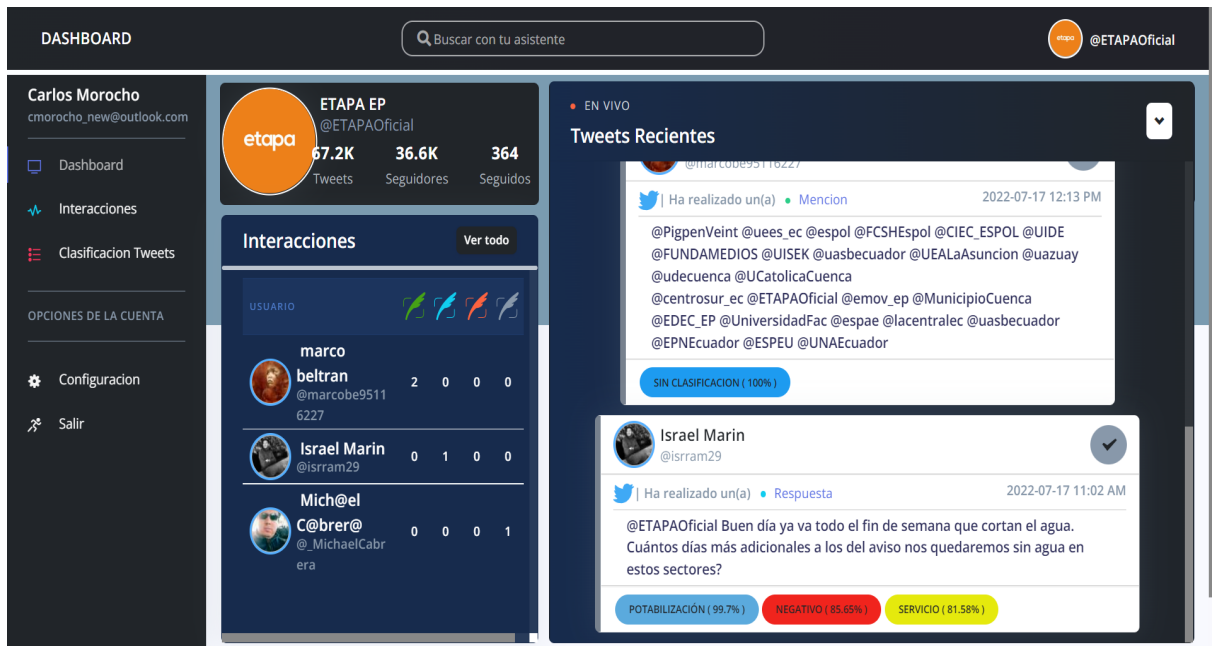


Figura 2.2: Vista Dashboard Live.

2.1.3 Detalle del tweet y su clasificacion

En la imagen [2.3](#), se puede visualizar el modal donde se presenta el detalle completo de un tweet, incluyendo sus metricas, entidades y la clasificacion obtenida.

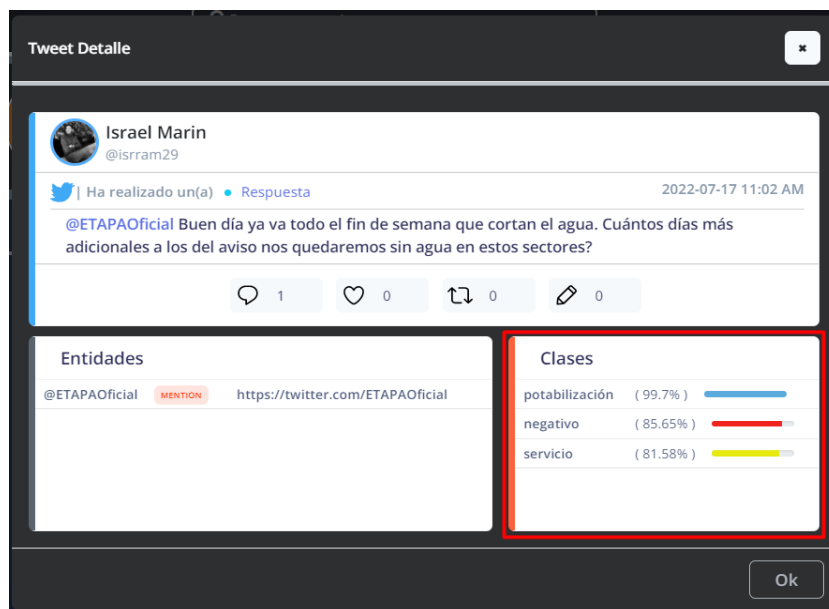


Figura 2.3: Vista detalle de Tweet

2.1.4 Vista de interaccion de usuarios

En la siguiente imagen se observa el grafico de interacciones que representa la cantidad de tweets por tipo (Menciones, Respuestas, Citas, Retweets) que tuvo la cuenta en los ultimos 30 dias, mensual o anual.

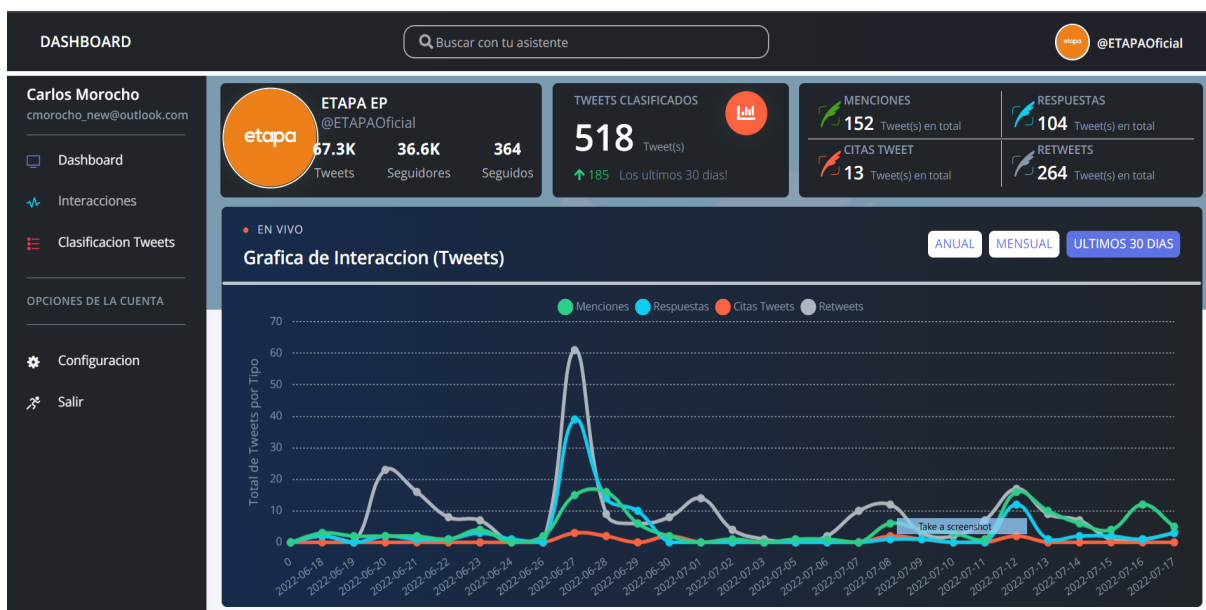


Figura 2.4: Grafico de interacciones

En la siguiente imagen se puede vizualizar los usuarios que interactuaron con la cuneta un dia especifico, esta vista se muestra al dar click sobre la grafica de interaccion.

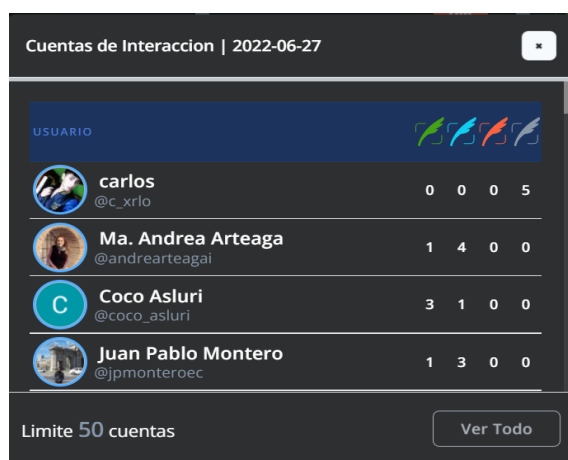


Figura 2.5: Vista cuentas de interaccion

2.1.5 Vista de tweets por tipo

En la imagen siguiente se puede observar la pantalla donde se visualizan el detalle de tweets por tipo, es decir cada pestaña representa a los tipos de tweets.

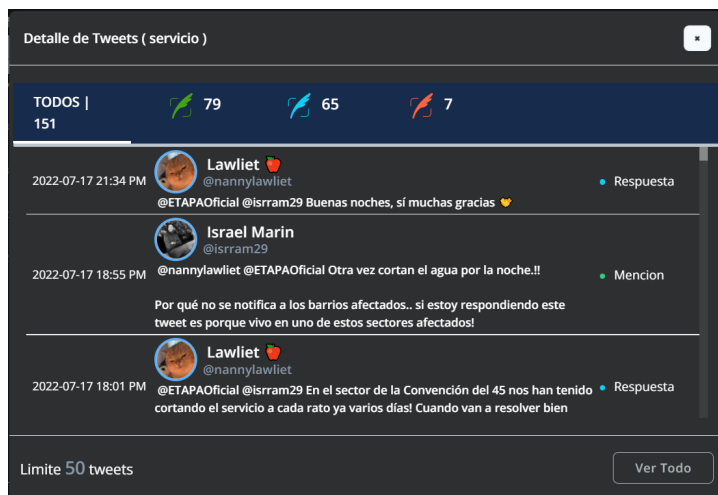


Figura 2.6: Vista de lista de Tweets por tipos

2.1.6 Vista de tweets con su clasificacion

En la imagen siguiente se puede observar la pantalla donde se visualizan los tweets clasificados, esto en base a una serie de filtros que el usuario puede ir modificando.

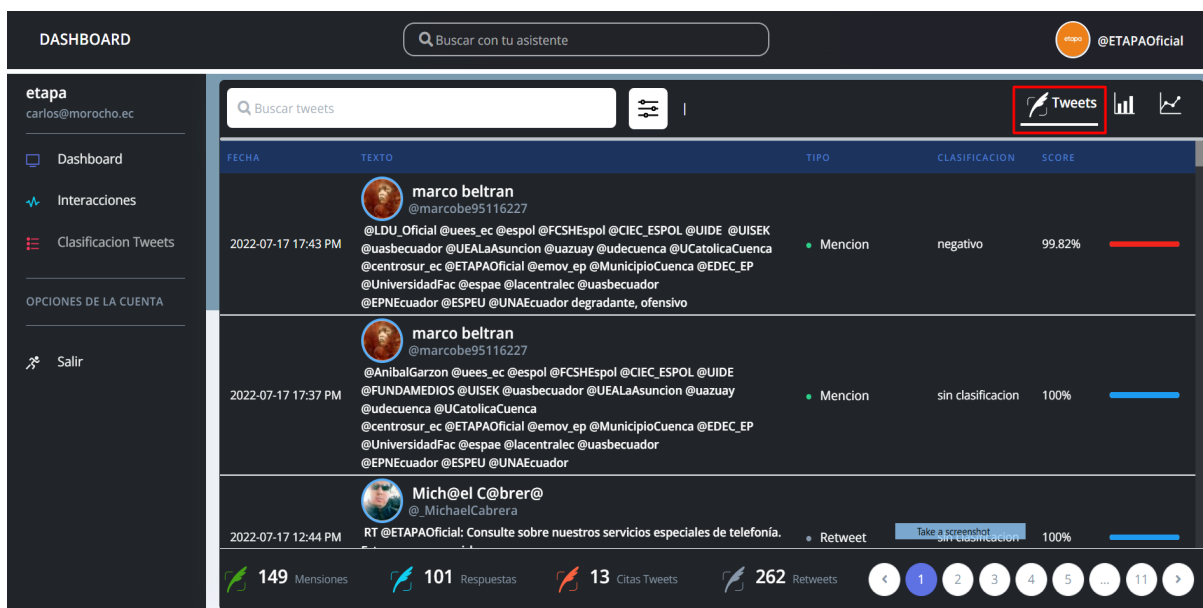


Figura 2.7: Seccion de lista de Tweets

En la imagen siguiente se puede observar la sección de la gráfica de barras donde se representa la cantidad de tweets por clases, según los filtros agregados.

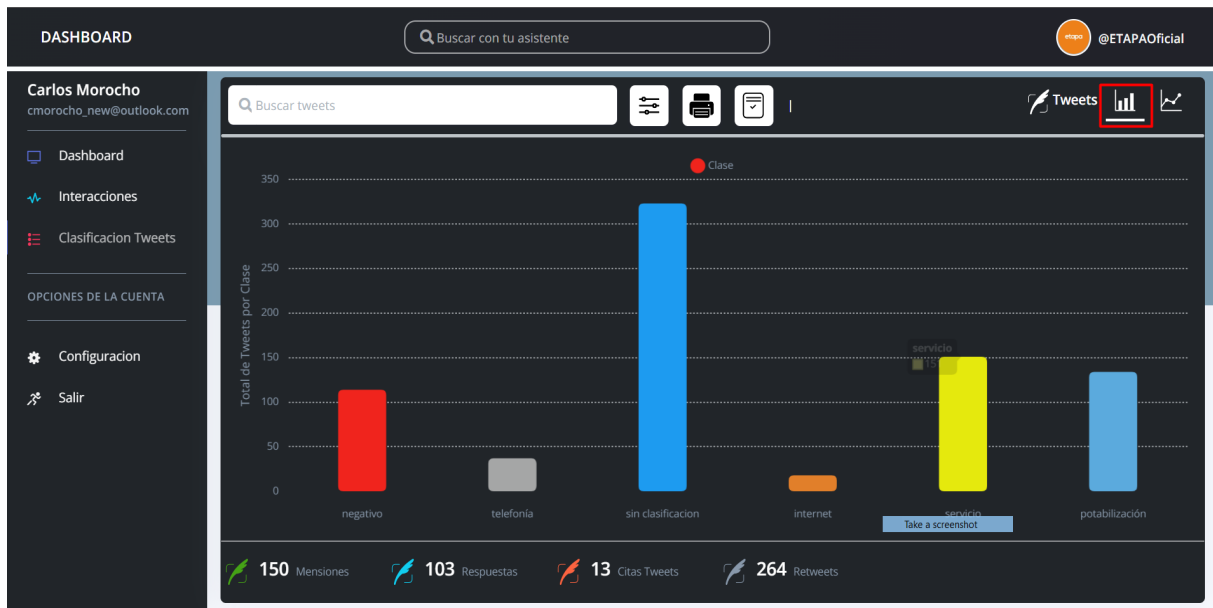


Figura 2.8: Sección de grafica de barras

En la imagen siguiente se puede observar la sección de la gráfica de líneas donde se representa la cantidad de tweets por clases y fechas, según los filtros agregados.

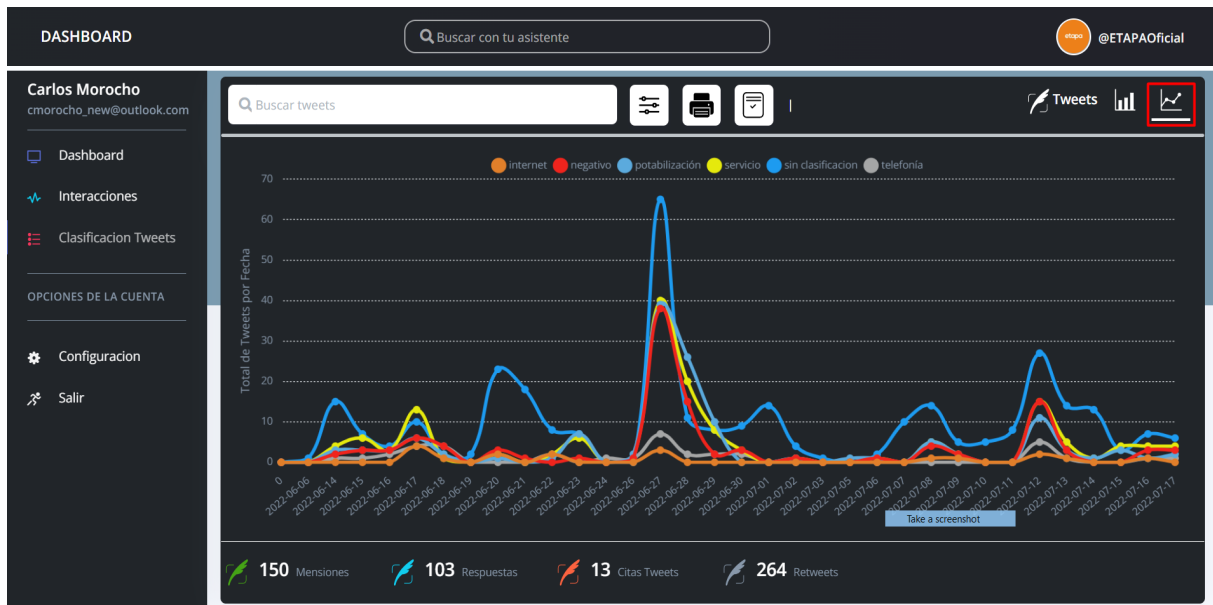


Figura 2.9: Sección de grafica de líneas

2.1.7 Visualización de configuración

En la siguiente imagen se puede visualizar la pantalla de configuración donde se puede gestionar las cuentas, clases, usuarios y acciones.

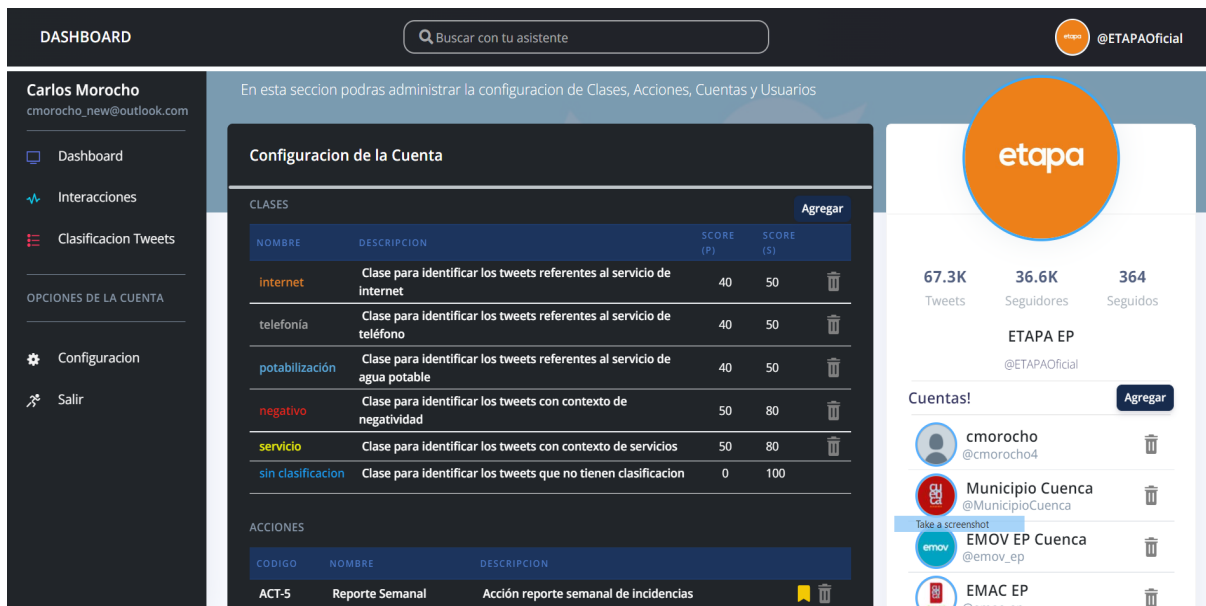


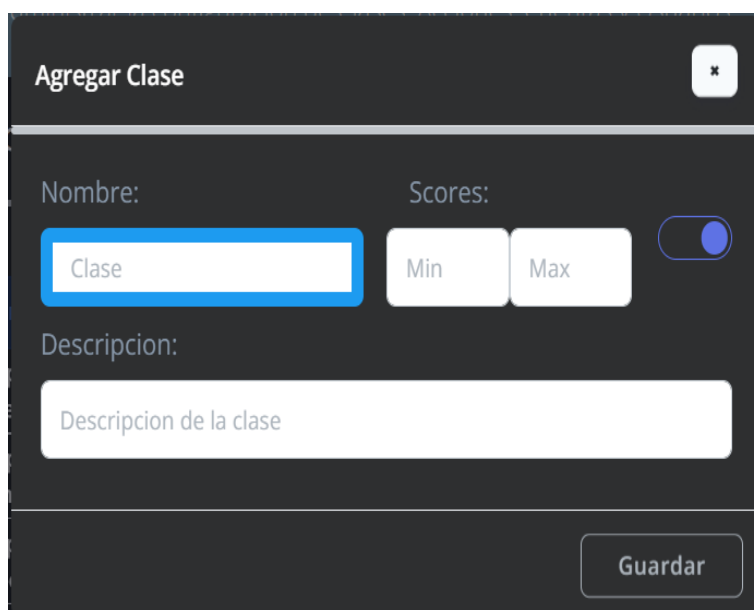
Figura 2.10: Pagina de configuración de cuentas

En la imagen [2.11](#) se puede observar la sección donde se vinculan las cuentas.



Figura 2.11: Modal vincular cuenta en la plataforma.

En la imagen [2.12](#) se puede observar la sección donde se agregan las clases.



Modal titled "Agregar Clase" with a close button (x) in the top right corner. The form contains the following fields and controls:

- Nombre:** A text input field containing the word "Clase".
- Scores:** Two input fields labeled "Min" and "Max", and a toggle switch to the right.
- Descripción:** A larger text input field containing the placeholder text "Descripcion de la clase".
- Guardar:** A button at the bottom right of the modal.

Figura 2.12: Modal agregar clase en la cuenta.

En la imagen [2.13](#) se puede observar la sección donde se agregan los usuarios y se vinculan a la cuenta.



Modal titled "Agregar Usuario" with a close button (x) in the top right corner. The form contains the following fields and controls:

- Nombre:** A text input field containing the placeholder text "Nombre".
- Cuenta:** A dropdown menu showing "@ETAPAOficial x" and a toggle switch to the right.
- Credenciales!** A section header above two input fields: "Correo" and "Contraseña".
- Guardar:** A button at the bottom right of the modal.

Figura 2.13: Modal agregar usuario en la cuenta.

En la imagen 2.14 se puede observar la sección donde se agregan las acciones.

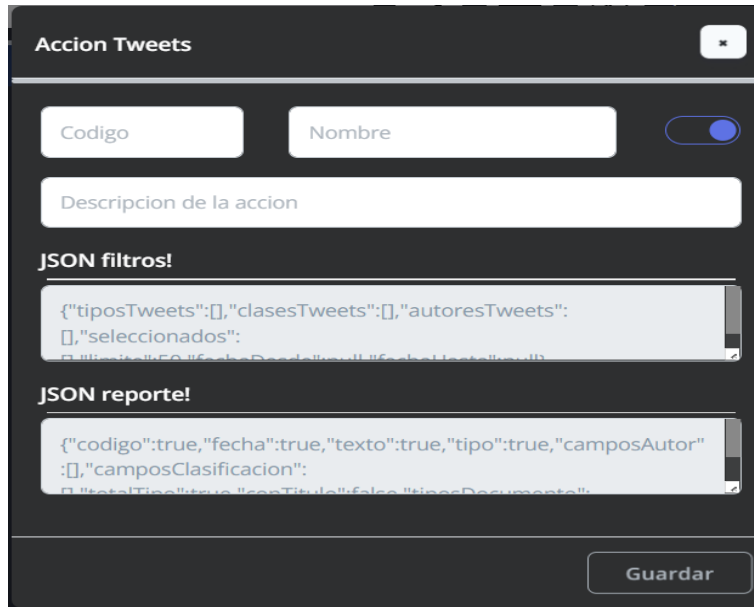


Figura 2.14: Modal agregar accion en la cuenta.

En la imagen 2.15 se puede observar la sección donde se interactura con el asistente virtual.

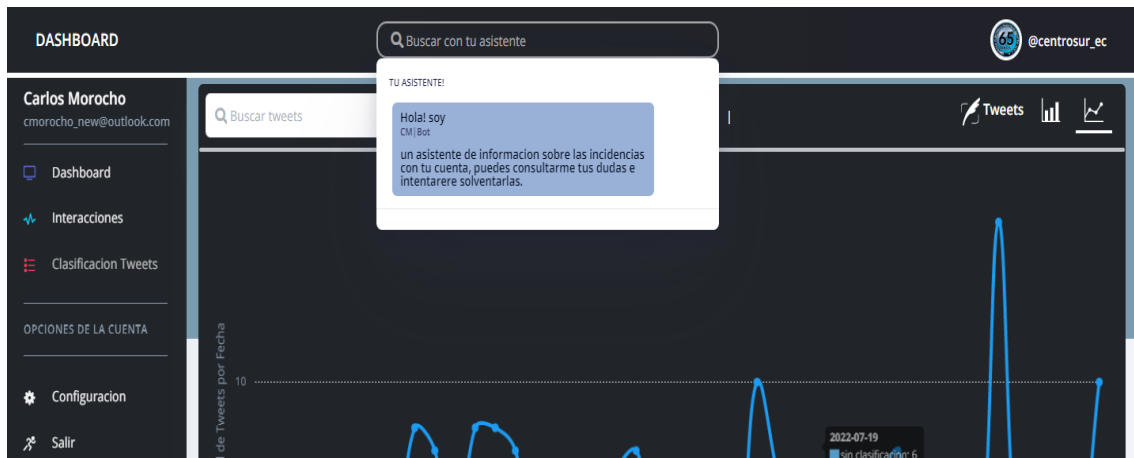


Figura 2.15: Seccion del asistente virtual.