

**UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA**

CARRERA DE COMPUTACIÓN

*Trabajo de titulación previo a la
obtención del título de Ingeniero
en Ciencias de la Computación*

PROYECTO TÉCNICO:

**“DISEÑO Y DESARROLLO DE UNA HERRAMIENTA INTELIGENTE PARA EL
ANÁLISIS DE LA PERSPECTIVA DE LOS CONSUMIDORES SOBRE PRODUCTOS
ESPECÍFICOS DE EMPRESAS DENTRO DE SUS REDES SOCIALES”**

AUTOR:

EDUARDO JAVIER VAZQUEZ ARIAS

TUTOR:

ING. GABRIEL ALEJANDRO LEÓN PAREDES, Ph.D.

CUENCA - ECUADOR

2021

CESIÓN DE DERECHOS DE AUTOR

Yo, Eduardo Javier Vazquez Arias con documento de identificación N° 1401277767, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación: **“DISEÑO Y DESARROLLO DE UNA HERRAMIENTA INTELIGENTE PARA EL ANÁLISIS DE LA PERSPECTIVA DE LOS CONSUMIDORES SOBRE PRODUCTOS ESPECÍFICOS DE EMPRESAS DENTRO DE SUS REDES SOCIALES”**, mismo que ha sido desarrollado para optar por el título de: *Ingeniero en Ciencias de la Computación*, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, octubre de 2021.



Eduardo Javier Vazquez Arias

C.I. 1401277767

CERTIFICACIÓN

Yo, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: **“DISEÑO Y DESARROLLO DE UNA HERRAMIENTA INTELIGENTE PARA EL ANÁLISIS DE LA PERSPECTIVA DE LOS CONSUMIDORES SOBRE PRODUCTOS ESPECÍFICOS DE EMPRESAS DENTRO DE SUS REDES SOCIALES”**, realizado por Eduardo Javier Vazquez Arias, obteniendo el *Proyecto Técnico*, que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana.

Cuenca, octubre de 2021.

A handwritten signature in blue ink, appearing to read 'Gabriel', with a horizontal line extending to the right.

Gabriel Alejandro León Paredes, Ph.D.

C.I. 0103652186

DECLARATORIA DE RESPONSABILIDAD

Yo, Eduardo Javier Vazquez Arias con documento de identificación N° 1401277767, autor del trabajo de titulación: **“DISEÑO Y DESARROLLO DE UNA HERRAMIENTA INTELIGENTE PARA EL ANÁLISIS DE LA PERSPECTIVA DE LOS CONSUMIDORES SOBRE PRODUCTOS ESPECÍFICOS DE EMPRESAS DENTRO DE SUS REDES SOCIALES”**, certifico que el total contenido del *Proyecto Técnico*, es de mi exclusiva responsabilidad y autoría.

Cuenca, octubre de 2021.



Eduardo Javier Vazquez Arias

C.I. 1401277767

DEDICATORIA

Dedico este trabajo a mi madre Mayra Vazquez y a mi abuelita María Arias por todo el esfuerzo, apoyo incondicional y sus consejos que me permitieron seguir día a día para poder alcanzar esta meta tan importante en mi vida, me siento muy orgulloso por tenerlas a mi lado.

A mi tía María Vazquez que me aconsejaba y me apoyaba en todas las decisiones que tomaba, la cual la considero como mi segunda madre.

A mi tío Ángel Vazquez por ser un ejemplo para seguir, el cual me motiva a nunca rendirme y

siempre seguir adelante y quererme como si fuera su hermano.

A mis hermanos Marco Mogro y Kevin Mogro por estar siempre ahí con sus ocurrencias que me alegraban el día.

A mis compañeros quienes sin esperar nada a cambio compartieron su conocimiento, alegrías y tristezas en este arduo camino.

Por ultimo y menos importante dedico esto a mis amigos, Ana Gómez y Richard Ramírez por haber estado a mi lado en todo momento tanto de mi vida personal como profesional, gracias por todo ese apoyo brindado siempre los llevare en mi corazón.

Eduardo Javier Vazquez Arias

AGRADECIMIENTO

Agradezco primeramente a Dios por permitirme cumplir una de las metas más importantes de mi vida, al PhD. Gabriel León por compartirme sus conocimientos y ser una guía durante el desarrollo de este trabajo de titulación.

Agradezco a mi madre y a mi abuelita por todo su cariño, sus consejos y su apoyo incondicional que me han permitido llegar al final de esta etapa de mi vida.

Agradezco también a toda mi familia que siempre han estado ahí pendientes de mi vida personal y profesional motivándome cada día a sacar la mejor versión de mí.

A todo mis compañeros y amigos que hemos compartido un sin número de buenos momentos durante todo este trayecto.

Por último, agradezco todos los docentes de la Universidad Politécnica Salesiana que supieron compartir sus conocimientos y que día a día estuvieron ahí cuando los necesitaba.

Eduardo Javier Vazquez

Resumen

El gran avance tecnológico ha acarreado un sin número de utilidades para la sociedad, como lo es la conectividad que tienen las personas entre ellas en tiempo real, sin importar su ubicación geográfica. Todo esto en gran parte por el surgimiento de las redes sociales que nos permiten esa interacción, y poder mantenernos al tanto con información destacada.

Este avance ha permitido el aumento de las plataformas de redes sociales tales como Facebook, Twitter, Instagram y otras las cuales sirven como canales donde las personas interactúan, debaten y comparten opiniones. De esta forma generando un nuevo universo en el cual las personas comparten todo tipo de información tanto cosas como que les gusta y las que no, de lo que hacen o puedan hacer en un futuro, de las cosas como productos o servicios, y un sin número de cosas.

Actualmente la plataforma con mayor cantidad de usuarios activos es Facebook, haciéndolo una de las redes más utilizadas, dentro de esta red social los usuarios tienden a contar más sobre su vida personal, y mostrar interés por gustos específicos donde podríamos encontrar a nuestro público objetivo.

Estos impactos que tiene actualmente los cambios tecnológicos nos han llevado a ver de una forma diferente las relaciones comerciales entre las empresas y consumidores, destacando el papel importante que cumplen tanto las soluciones de Machine Learning e Inteligencia Artificial en esta nueva era. En el cual la mayor habilidad que deben tener las empresas es identificar aquellos productos o servicios que están teniendo gran acogida por los clientes o consumidores.

Este proyecto tiene como objetivo ayudar a las empresas con una herramienta inteligente que les permita analizar los comentarios sobre uno de sus productos específicos, de esta manera poder identificar los temas o tópicos sobre los que más se habla acerca de dicho producto. Identificando también cuál es el sentimiento del usuario que quiso expresar sobre la calidad del producto, permitiendo de esta forma a la empresa a tomar decisiones de acuerdo con los objetivos que se quieran lograr.

Abstract

The great technological advancement has brought a number of benefits to society, such as the connectivity that people have with each other in real time, regardless of their geographic location. All this in large part due to the emergence of social networks that allow us to interact and keep up to date with important information.

This advance has allowed the rise of social networking platforms such as Facebook, Twitter, Instagram, and others which serve as channels where people interact, debates and share opinions. In this way generating a new universe in which people share all kinds of information such as likes and dislikes, what they do or may do in the future, things like products or services, and countless other things.

Currently the platform with the largest number of active users is Facebook, making it one of the most used networks, within this social network users tend to talk more about their personal life, and show interest in specific tastes where we could find our target audience.

These impacts that currently has the technological changes has led us to see in a different way the commercial relationships between companies and consumers, highlighting the important role that both Machine Learning and Artificial Intelligence solutions play in this new era. In which the greatest skill that companies must have is to identify those products or services that are having a great reception by customers or consumers.

This project aims to help companies with an intelligent tool that allows them to analyze the comments on one of their specific products, thus being able to identify the themes or topics that are most talked about that product. Identifying also what is the user's feeling about the quality of the product, thus allowing the company to make decisions according to the objectives to be achieved

ÍNDICE

I	Introducción	12
II	PROBLEMA	14
2.1	Antecedentes	14
2.2	Importancia	15
2.3	Alcance	15
III	OBJETIVOS	17
3.1	General.....	17
3.2	Específicos	17
IV	MARCO TEORICO	19
4.1	Inteligencia Artificial	19
4.1.1	Definición.....	19
4.1.2	Áreas de la Inteligencia artificial	19
4.2	Redes Neuronales.....	20
4.2.1	Definición.....	20
4.2.2	Estructura de una red neuronal artificial	21
4.2.3	Ventajas y Desventajas	22
4.2.4	Tipos de Redes Neuronales	23
4.2.5	Principales Aplicaciones	24

4.3	Deep Learning	25
4.3.1	Definición	25
4.3.2	Procesamiento de Lenguaje Natural	25
4.3.3	Análisis de Sentimientos	26
4.4	Aprendizaje Automático	27
4.4.1	Definición	27
4.4.2	Aprendizaje Supervisado	27
4.4.3	Aprendizaje No supervisado	28
4.4.4	Aprendizaje Semi-Supervisado	29
4.5	Redes Sociales	30
4.5.1	Definición	30
4.5.2	Clasificación	30
4.6	Minería de Datos	31
4.6.1	Definición	31
4.6.2	Minería de Datos en Redes Sociales	31

V METODOLOGÍA 33

5.1	Levantamiento de Requerimientos	33
5.1.1	Alcance.....	33
5.1.2	Personal Involucrado.....	33
5.1.3	Requisitos Funcionales.....	34
5.1.4	Requisitos No Funcionales.....	41
5.2	Diagramas para la aplicación Web	42
5.2.1	Diagramas de actividades UML.....	42
5.3	Arquitectura del sistema	45
5.3.1	Tecnologías del proyecto	46
5.3.2	Fase extracción de la información.....	47
5.3.3	Fase análisis de datos	48

5.3.4	Fase de Presentación	54
VI	ANÁLISIS DE RESULTADOS	57
6.1	Resultados de los algoritmos desarrollados	57
6.1.1	Clusterización de tópicos.	57
6.1.2	Detección de sentimientos.....	60
6.2	Resultados del análisis de la perspectiva de los consumidores acerca de un pro- ducto específico en una red social.....	62
6.2.1	Resultados	62
6.3	Análisis Encuestas	77
VII	CRONOGRAMA DE ACTIVIDADES	94
VIII	PRESUPUESTO	98
IX	CONCLUSIONES	100
X	RECOMENDACIONES	102
Anexos		106
A	Formato de la encuesta realizada para la validación de la aplicación web	107
B	Interfaces de Usuario	112
2.1	Aplicación Web	112
2.1.1	Página de Inicio.....	112
2.1.2	Inicio de sesión.....	113
2.1.3	Registro de Usuarios	114
2.1.4	Página principal.....	115
2.1.5	Perfil de Usuario	117
2.1.6	Resultados de análisis	118

2.1.7 Equipo de Desarrollo.....122

Índice de tablas

5.1	Personal involucrado en el desarrollo del proyecto.....	34
5.2	Requisito funcional 1.	34
5.3	Requisito funcional 2.	35
5.4	Requisito funcional 3.	36
5.5	Requisito funcional 4.	37
5.6	Requisito funcional 5.	38
5.7	Requisito funcional 6.	38
5.8	Requisito funcional 7.	39
5.9	Requisito funcional 8.	39
5.10	Requisito funcional 9.	40
5.11	Requisito funcional 10.	40
5.12	Requisito no funcional 1.	41
5.13	Requisito no funcional 2.	41
5.14	Requisito no funcional 3.	42
5.15	Parámetros óptimos modelo LDA.....	51
5.16	Parámetros de prueba, proceso Fine-tunning.	52
5.17	Topología del modelo de red neuronal.....	53
6.1	Validación del modelo de red neuronal recurrente.....	61
6.2	Pregunta 2. ¿Qué tipo de productos vende su empresa?	78

6.3	Pregunta 3. Cuales redes sociales usa con más frecuencia para promocionar sus productos.....	79
6.4	Pregunta 4. Con que frecuencia publica nuevos productos en las redes sociales.	80
6.5	Pregunta 5. ¿Qué le pareció la facilidad de uso de esta herramienta?	81
6.6	Pregunta 6. ¿Qué tan conforme esta con la experiencia al registrar y analizar productos en la herramienta?	82
6.7	Pregunta 7. ¿Qué tan satisfecho(a) esta con la seguridad de la aplicación?.....	83
6.8	Pregunta 8. ¿Es comprensible la manera en la que se presenta los resultados de los análisis en la herramienta?	84
6.9	Pregunta 9. ¿Qué tan conforme esta con la confiabilidad de esta herramienta? .	85
6.10	Pregunta 10. Está de acuerdo que la herramienta cumple con el objetivo de ayudar a las empresas a detectar sentimientos de los usuarios y los tópicos que más se hablan acerca de un producto específico	86
6.11	Pregunta 11. La herramienta capturo la mayor parte de los comentarios de la publicación.....	87
6.12	Pregunta 12. Usted cree que la relación de las palabras dentro de un tópico tienen concordancia.	88
6.13	Pregunta 13. Su empresa tomaría en cuenta los resultados arrojados por la herramienta para mejorar sus estrategias empresariales.....	89
6.14	Pregunta 14. Utilizaría nuestra herramienta de manera cotidiana para realizar un análisis de la acogida de sus productos dentro de las redes sociales.....	90
6.15	Pregunta 16. ¿Qué tan probable es que usted recomiende esta herramienta a un(a) amigo(a) o algún miembro de la familia?.....	92

Índice de figuras

4.1	Capas de una red neuronal artificial.....	21
4.2	Red Neuronal Monocapa	23
4.3	Red Neuronal Multicapa	23
4.4	Red Neuronal Recurrente.....	24
5.1	Diagrama de Actividades, Registro y Autenticación.	43
5.2	Diagrama de Actividades, Modificación del perfil de usuario.	44
5.3	Diagrama de Actividades, Análisis de publicación.....	44
5.4	Arquitectura General del Sistema.	45
5.5	Fase extracción de la información.....	47
5.6	Fase Análisis de Datos.	48
5.7	Algoritmo detección de tópicos.	50
5.8	Modelo propuesto.	51
5.9	Proceso para consumir el modelo de red neuronal.....	53
5.10	Fase de presentación.	54
6.1	Tópicos detectados.....	58
6.2	Gráfico interactivo modelo LDA.	59
6.3	Coherencia modelo LDA.	60
6.4	Entrenamiento y validación del modelo propuesto.	60
6.5	Análisis #1, Resultados análisis de comentarios de producto específico.	63

6.6	Análisis #1, Resultados análisis de comentarios de producto específico parte 2.	63
6.7	Análisis #1, Representación porcentual de la perspectiva de los consumidores.	64
6.8	Análisis #1, Representación distribución palabras claves por tópico.....	65
6.9	Análisis #1, Representación visual tópicos del modelo LDA mediante pyL- DAvis.	66
6.10	Análisis #1, Representación visual distribución de comentarios por tópico.....	67
6.11	Análisis #1, Representación reacciones y cantidad de comentarios por tópico.	67
6.12	Análisis #2, Resultados análisis de comentarios de producto específico.	68
6.13	Análisis #2, Representación porcentual de la perspectiva de los consumidores.	69
6.14	Análisis #2, Representación distribución palabras claves por tópico.....	70
6.15	Análisis #2, Representación visual tópicos del modelo LDA mediante pyL- DAvis.	70
6.16	Análisis #2, Representación visual distribución de comentarios por tópico.....	71
6.17	Análisis #2, Representación reacciones y cantidad de comentarios por tópico.	72
6.18	Análisis #3, Resultados análisis de comentarios de producto específico.	73
6.19	Análisis #3, Representación porcentual de la perspectiva de los consumidores.	74
6.20	Análisis #3, Representación distribución palabras claves por tópico.....	74
6.21	Análisis #3, Representación visual tópicos del modelo LDA mediante pyL- DAvis.	75
6.22	Análisis #3, Representación visual distribución de comentarios por tópico.....	76
6.23	Análisis #3, Representación reacciones y cantidad de comentarios por tópico.	77
6.24	Representación porcentual para la pregunta 1.	78
6.25	Representación porcentual para la pregunta 2.	79
6.26	Representación porcentual para la pregunta 3.	80
6.27	Representación porcentual para la pregunta 4.	81
6.28	Representación porcentual para la pregunta 5.	82
6.29	Representación porcentual para la pregunta 6.	83

6.30	Representación porcentual para la pregunta 7.	84
6.31	Representación porcentual para la pregunta 8.	85
6.32	Representación porcentual para la pregunta 9.	86
6.33	Representación porcentual para la pregunta 10.	87
6.34	Representación porcentual para la pregunta 11.	88
6.35	Representación porcentual para la pregunta 12.	89
6.36	Representación porcentual para la pregunta 13.	90
6.37	Representación porcentual para la pregunta 14.	91
6.38	Representación porcentual para la pregunta 16.	92
2.1	Página de inicio.	113
2.2	Inicio de sesión de la aplicación web.	114
2.3	Formulario Registro Usuario.	115
2.4	Primer inicio de sesión en la aplicación web.	116
2.5	Confirmación análisis nueva publicación.	116
2.6	Lista de publicaciones analizadas.	117
2.7	Modificar datos de usuario.	118
2.8	Análisis de sentimientos y tópicos de los comentarios.	119
2.9	Porcentaje acogida de usuarios.	119
2.10	Distribución de las palabras por tópicos.	120
2.11	Relación entre los tópicos.	120
2.12	Distribución de los comentarios por grupos de tópicos.	121
2.13	Distribución cantidad de reacciones y comentarios por tópicos.	121
2.14	Equipo de Desarrollo.	122

Capítulo I

Introducción

Hoy en día el auge de los avances tecnológicos ha ido tomando un gran peso en nuestras vidas, principalmente en las actividades que realizamos día a día. Facilitándonos resolver muchos de los problemas que hace algunos años se consideraba algo imposible.

De este auge tecnológico surgen las redes sociales que nos permiten interactuar con personas ya sean familiares, amigos, compañeros de trabajo, e inclusive interactuar con personas desconocidas pero que comparten gustos en común como lo es música, fotografía, deportes, etc. Esto ha abierto un nuevo campo a un sin número de investigaciones donde se tratan diferentes áreas, como lo es la interacción entre los usuarios, la difusión de información por estos medios, y lo que más nos interesa en este proyecto es conocer la opinión respecto a un producto, la reputación o la valoración que está teniendo.

Según (Pirri et al., 2020) manifiesta que las perspectivas y las opiniones de los integrantes de la comunidad expresada mediante estos medios sociales representan una buena aproximación a los sentimientos y actitudes que pueden contribuir a la toma de decisiones de otros usuarios. Identificar con la mayor precisión posible el contenido de estos sentimientos mejoraría el desarrollo de una estrategia que se adapte de mejor manera a cubrir las necesidades de los consumidores.

Con el crecimiento de las reseñas, recomendación, valoraciones y otras formas de expresión

de opinión en línea, hay un creciente interés por las técnicas de extracción automática de la información que contienen. De estos, dos de los problemas que se han planteado resolver en este proyecto son: la clusterización de tópicos (temas) y el análisis de sentimientos. Por tanto, en este proyecto se ofrece una herramienta que facilita el análisis de productos específicos a través de un servicio web que ofrece simplicidad y usabilidad. La herramienta incluye la generación de gráficos y reportes que permite a las empresas analizar de mejor manera la acogida que ha tenido su producto por parte de los consumidores.

Haciendo uso de las últimas tecnologías como son los bots, los cuales nos permiten extraer información de cualquier página web, especialmente este caso para redes sociales, podemos realizar este hallazgo de tópicos juntamente con la detección de sentimientos los cuales quiso expresar el usuario.

Capítulo II

PROBLEMA

2.1 Antecedentes

Según (Bobadilla et al., 2018) durante los últimos años la minería de datos y los sistemas de recomendación han demostrado ser un apoyo a las empresas con el fin de tratar la gran cantidad de información, analizar datos y crear estrategias de negocio, tanto para ganar clientes, como para mejorar sus ingresos.

Actualmente las redes sociales permiten a sus usuarios intercambiar información de forma dinámica. El conocimiento extraído de las redes sociales se ha demostrado ser de gran importancia en varias aplicaciones una de ellas es la minería de sentimientos en las publicaciones dentro de una red social se ha demostrado tener éxito en la predicción del movimiento del mercado financiero. (Bollen et al., 2011)

Hoy en día en Ecuador existe más de 12 millones de usuarios activos entre las diferentes redes sociales, un equivalente al 69% de la población ecuatoriana, de lo cual el 97% acceden mediante dispositivos móviles (Shum, 2021). Lo cual está permitiendo a las empresas publiquen sus servicios o productos dentro de las redes sociales y que los consumidores puedan a su vez interactuar con las empresas para dar sus opiniones sobre diferentes aspectos como la calidad, experiencia y precios de los productos.

Onur Kucuktunc estudia la influencia de factores como el género, la edad, el nivel educativo, el tema de discusión y la hora del día en la variación del sentimiento en Yahoo! Answers. A menudo en las redes sociales se crea perspectivas que pueden impactar a los consumidores, analizando primero el sentimiento en las reseñas de los consumidores para descubrir si es positivo, negativo o neutro. Esto después puede ayudar a saber los productos que tuvieron mayor acogida por parte de los consumidores. (Kucuktunc et al., 2012)

2.2 Importancia

El problema se da cuando las empresas realizan publicaciones de sus productos dentro de las redes sociales, existe un gran abanico de reseñas por parte de los consumidores, el analizar cada una de estas reseñas tomaría tiempo y mano de obra de parte de las empresas, en donde muchas de ellas no cuentan con un área específica que se dedique a estas nuevas tecnologías que son las redes sociales.

Por lo tanto no existe como tal un análisis completo si un producto en específico está teniendo o no la acogida esperada por parte de los consumidores en las redes sociales, para solucionar esta problemática se propone diseñar y desarrollar una herramienta inteligente que nos ayudará a realizar el análisis de si un producto está teniendo una buena aceptación por parte de los consumidores, de esta forma mediante el procesamiento del lenguaje natural podremos identificar las reseñas tanto positivas como negativas de los consumidores logrando así que las empresas puedan conocer sobre la perspectiva de los consumidores respecto a sus productos y de esta manera perfeccionar las estrategias de marketing dentro de estas plataformas de redes sociales con respecto a sus productos.

2.3 Alcance

La herramienta facilita a las empresas poder identificar cual ha sido la acogida de cada uno de los productos que han sido publicados en la red social de Facebook, esto incluye diferentes

industrias como por ejemplo: tecnológicas, motocicletas, deportivas, de hogar, entre otras. Es una herramienta que no exige personal técnico puesto que se ofrece una GUI (Interfaz Gráfica de Usuario) simple.

El sistema usa un servidor del Grupo de Investigación en Cloud Computing Smart Cities & High Performance Computing (GIHP4), por tanto, cualquier empresa del Ecuador pueda acceder a ella, registrarse y poder realizar el análisis de sus productos. Sin embargo, para los casos de prueba se los realizo con empresas de la ciudad de Cuenca.

Capítulo III

OBJETIVOS

3.1 General

Diseño y desarrollo de una herramienta inteligente para el análisis de la perspectiva de los consumidores sobre productos específicos de empresas publicados dentro de sus redes sociales en base a las reseñas realizadas por parte de los consumidores.

3.2 Específicos

- Estudiar y conocer técnicas a) sobre los enfoques del procesamiento del lenguaje natural y técnicas de clasificación de tópicos, b) para identificar los sentimientos positivos, negativos y neutros de los usuarios.
- Diseñar y desarrollar modelos de clasificación de tópicos de las reseñas de los consumidores utilizando técnica Latent Dirichlet Allocation (LDA).
- Diseñar y desarrollar modelos de red neuronal para el análisis de sentimientos utilizando procesamiento lenguaje natural.
- Diseñar, desarrollar e implementar una herramienta web que permita analizar la perspectiva de un producto específico de una empresa.

- Diseñar y ejecutar un plan de experimentación que posibilite validar la herramienta de análisis de la perspectiva de los productos.

Capítulo IV

MARCO TEORICO

4.1 Inteligencia Artificial

4.1.1 Definición

Los investigadores creen que la IA se puede entender cómo funciona el cerebro humano imitando aquellas cualidades del cerebro humano que tiene para aprender, actuar y pensar. Su principal tarea es construir la teoría del procesamiento inteligente y luego diseñar algunos sistemas informáticos que puedan simular algunos actos de la inteligencia humana.

La IA es una forma de hacer que las máquinas piensen y se comporten de forma inteligente. Es una ciencia que consiste en encontrar teorías y metodologías que pueden ayudar a las máquinas a entender el mundo en consecuencia a situaciones de la misma manera en que lo hacen los humanos. (Joshi, 2017)

4.1.2 Áreas de la Inteligencia artificial

- **Visión por Computador**

Sistemas que tratan datos visuales tanto de imágenes como videos, el cual comprenden extraer la información en función del caso de uso. Algunos ejemplos son el reconocimiento

de personas u objetos, la navegación en un entorno y en la robótica como cuando un robot agarra un objeto.

- **Procesamiento del Lenguaje Natural**

El procesamiento del lenguaje natural es la relación entre el lenguaje natural humano y las computadoras, específicamente en el procesamiento para la comprensión, manipulación y el análisis del lenguaje natural por computadora. Permitiendo interactuar con una maquina escribiendo frases en lenguaje natural y que esta sea capaz de comprender brindándonos la información necesaria.

- **Sistemas Expertos**

Estos sistemas utilizan técnicas de IA para poder asesorar o tomar decisiones, los cuales suelen utilizar base de datos de diferentes áreas de expertas de conocimiento, como pueden ser medicina, finanzas, marketing, etc. Permitiéndonos poder resolver problemas que únicamente podían resolver los expertos de los diferentes dominios.

- **Robótica**

La IA en la robótica permite que las maquinas sean mucho más eficientes con la capacidad de autoaprendizaje para reconocer nuevos objetos, aunque actualmente se utiliza más con fines industriales ya que permite realizar diversas tareas con alta precisión y mayor eficiencia que los humanos

4.2 Redes Neuronales

4.2.1 Definición

Es un sistema de aprendizaje computacional que utiliza una red de funciones para comprender e interpretar una entrada de datos de una forma a una salida deseada. Este concepto de

redes neuronales se inspiró básicamente en cómo las neuronas del cerebro funcionan al trabajar juntas para comprender las entradas de los sentidos humanos.

Las redes neuronales actualmente se pueden aplicar a una amplia gama de problemas pudiendo evaluar muchos tipos de entradas de datos como imágenes, base de datos, archivos y mucho más. Que nos pueden ayudar al reconocimiento de imágenes, patrones, reconocimiento facial, filtrado de contenido esencial e inclusive se utiliza mucho en marketing.

4.2.2 Estructura de una red neuronal artificial

Las redes neuronales generalmente se organizan en capas, donde cada capa se compone de muchos nodos interconectados las cuales contienen una función de activación. Se compone de las siguientes capas:

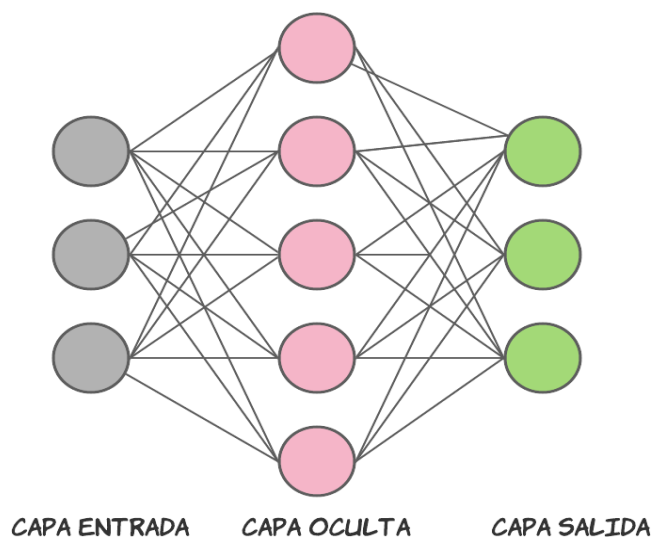


Figura 4.1: Capas de una red neuronal artificial.

- **Capa de Entrada:** Esta capa se encarga de recibir como entrada los datos, en el que el de variables es igual al número de nodos. Estos nodos son pasivos, debido a que no cambian los datos.

- **Capa Oculta:** En estas capas ocultas se aplican transformación a los valores de la capa anterior (capa de entrada) aquí se realizan todos los cálculos matemáticos requeridos. Puede haber una o más capas ocultas.
- **Capa de Salida:** Recibe las conexiones desde las capas ocultas o de la capa de entrada, devolviendo un valor de salida que corresponde a la predicción de la variable de respuesta.

4.2.3 Ventajas y Desventajas

(Rivera et al., 2015) indica varias de las ventajas y desventajas de estas redes neuronales.

Ventajas

- Aprendizaje adaptativo donde aprende a realizar actividades a partir de un conjunto de datos.
- Operación en tiempo real.
- Tolerancia a fallos parciales.
- Auto organización de la información recibida pues las redes neuronales organizan la información utilizando diferentes métodos matemáticos.

Desventajas

- Complejidad de aprendizaje para tareas grandes.
- Enormes cantidades de datos para el entrenamiento.
- Necesidad de un desarrollador o científico de datos para la interpretación de los resultados que produce.

4.2.4 Tipos de Redes Neuronales

- **Red Neuronal Monocapa**

Consta de una sola capa, cuyo objetivo es lanzar la capa de entrada en dirección a una capa de salida, realizando todos los cálculos necesarios.

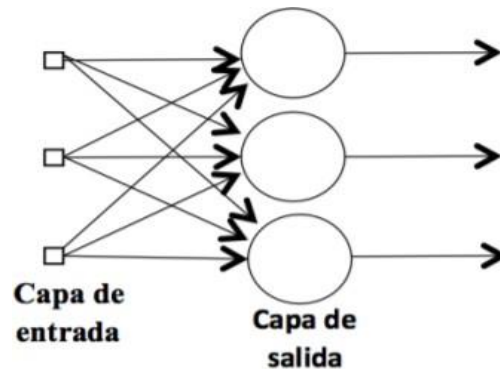


Figura 4.2: Red Neuronal Monocapa
(Asanza and Olivo, 2018)

- **Redes Neuronales Multicapa**

Existe un conjunto de capas ocultas (intermedias) la cual puede estar total o parcialmente conectada. Las cuales reciben señales de entrada de otra capa anterior, y los datos de salida de esta capa son la entrada de la siguiente capa oculta o hacia la capa de salida. Son muy útiles para el reconocimiento o clasificación de patrones.

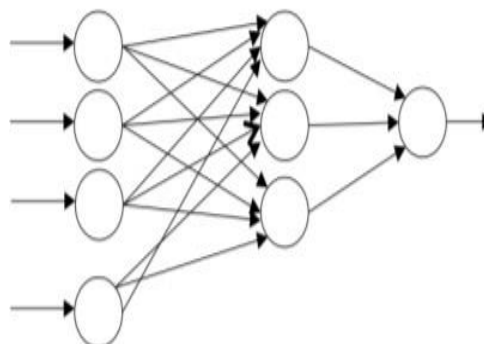


Figura 4.3: Red Neuronal Multicapa
(Asanza and Olivo, 2018)

- **Red Neuronal Recurrente**

Las Redes Neuronales Recurrentes (RNN) son una familia de redes neuronales para procesar datos secuenciales (Goodfellow et al., 2016). Este tipo de redes se ajustan bien a entradas secuenciales es la capacidad de condicionar la próxima capa de procesamiento a las observaciones anteriores. Su característica principal es el permitir compartir parámetros en diferentes partes del modelo, dentro de PLN esto permite encontrarcuando una palabra es relevante sin importar en que posición se encuentre dentro de la oración. (Muñiz Cuza, 2018)

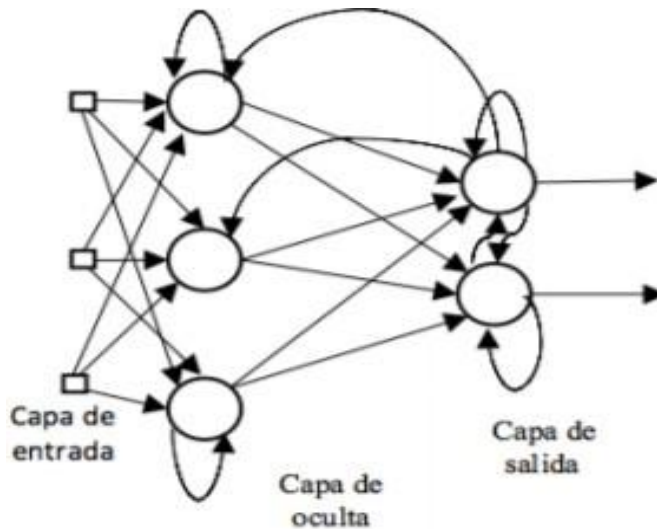


Figura 4.4: Red Neuronal Recurrente
(Asanza and Olivo, 2018)

4.2.5 Principales Aplicaciones

- **Compresión de Imágenes:** Identificación de caracteres: proceso en la que se realiza la interpretación visual y de clasificación de símbolos. (Olabe, 1998)
- **Predicción:** Para poder predecir fenómenos muchas de las veces no basamos en una colección de datos temporales, los mismo que nos permiten modelar su comportamiento. (Olabe, 1998)
- **Filtro de ruido:** Mayormente utilizadas para eliminar el ruido de una señal.

- **Modelos económicos y financieros:** Este es uno de los usos más importantes que se les da a las redes neuronales, la creación de modelos económicos que permitan predecir la cantidad de ventas, el interés de las cuentas, etc. (Olabe, 1998)

4.3 Deep Learning

4.3.1 Definición

Es un modelo computacional en el que consiste en una red neuronal donde se cuenta con una capa de entrada, múltiples capas ocultas y una última capa de salida. Cada uno de estos nodos de las diferentes capas pueden estar o no conectados o todos o solo algunos de los nodos de la siguiente capa. Mientras mayor sea el número de capas ocultas, más profunda será la red. (Zhang et al., 2018)

4.3.2 Procesamiento de Lenguaje Natural

Definición

El procesamiento del lenguaje natural es una forma en que la inteligencia artificial brinda a las computadoras la facultad de poder comprender e interpretar el lenguaje humano. Ayudando de esta forma a las computadoras a determinar que partes del lenguaje humano son importantes.

A través de la IA, campos como el aprendizaje automático y el aprendizaje profundo están abriendo a los científicos de datos a un mundo de todas las posibilidades. Esto se utiliza cada vez más en el análisis de datos para dar sentido a los macrodatos.

El procesamiento del lenguaje natural es un conjunto de técnicas computacionales para analizar y representar textos naturales, en uno o varios niveles lingüísticos con el fin de lograr un procesamiento para una variedad de tareas o aplicaciones. (Liddy, 2001)

Principales usos

- Los motores de búsqueda lo utilizan para generar resultados relevantes basados en un comportamiento de búsqueda similar.
- Aplicaciones de respuesta de voz interactiva, esto generalmente se utiliza en centros de llamadas (Call Center) para atender a las inquietudes de los clientes.
- En el monitoreo de las redes sociales no permite conocer lo que dicen los clientes en las redes sociales sobre una marca puede ayudar a las empresas a seguir ofreciendo excelentes productos o servicios.
- La creación de bots para paginas comerciales o dentro de redes sociales que no solo permiten publicitar un producto o servicio, sino que puedan interactuar con los clientes y así poder brindar una experiencia única.
- Los motores de búsqueda lo utilizan para generar resultados relevantes basados en un comportamiento de búsqueda similar.

Conclusión

El procesamiento del lenguaje natural desempeña un papel importante en el apoyo para las interacciones humano-maquina. Es un campo que se encuentra en constantes investigaciones y esperamos ver avances significativos que harán que las maquinas sean aún más inteligentes para poder reconocer e interpretar el lenguaje humano.

4.3.3 Análisis de Sentimientos

Definición

Consiste en utilizar la tecnología del Big data y Machine Learning para poder identificar las diferentes respuestas emocionales de las personas acerca de un producto o servicio. Mediante esta utilización las empresas u organizaciones puedan crear o mejorar sus estrategias.

Actualmente con el gran auge de las redes sociales, donde se encuentran publicaciones de todo tipo y junto a estos comentarios de todo tipo, la posibilidad de tener herramientas que permitan monitorizar y evaluar las diferentes opiniones vertidas dentro de las redes sociales.

Aplicaciones

- Evitar tener una multitud de datos que carecen de valor para la toma de decisiones.
- Desarrollar mejores estrategias empresariales.
- Identificar clientes muy insatisfechos para poder ofrecer una atención especializada.
- Procesar comentario de los empleados de una organización grande.

4.4 Aprendizaje Automático

4.4.1 Definición

Es un proceso computacional que utiliza datos de entrada para lograr una tarea deseada sin estar programado para producir un resultado específico. Este tipo de aprendizaje se adapta automáticamente su arquitectura a través de un entrenamiento en el que se proporcionan muestras de datos de entrada junto con los resultados deseado. (El Naqa and Murphy, 2015)

4.4.2 Aprendizaje Supervisado

Este proceso de aprendizaje parte de un conjunto de datos donde se considera a un supervisor que se encarga del proceso de aprendizaje. En el cual ya se conocen las respuestas correctas, y el algoritmo trata de hacer predicciones iterativas sobre los datos de entrenamiento y los cuales son corregidos por el supervisor

Este proceso consta de un conjunto de datos de características y etiquetas, donde su tarea principal es construir un estimado capaz de predecir la etiqueta de un objeto nuevo. Aquí el

algoritmo de aprendizaje recibe un conjunto de características junto con las salidas o etiquetas correctas y de esta forma aprende a comparar su salida real con las salidas corregidas para encontrar errores. (Nasteski, 2017)

Lo cual significa que algunos datos ya están etiquetados con la respuesta correcta, mas adelante la maquina recibe un nuevo conjunto de datos para que el este algoritmo de aprendizaje supervisado analice estos nuevos datos y los etiquete (clasifique) de acuerdo con las etiquetas con las que se entrenó previamente al algoritmo.

Se monitorea o supervisa en el sentido que ya conocemos la salida y este cada vez se corrige para poder optimizar sus resultados. Este cada vez se entrena sobre el conjunto de datos y se modifica hasta alcanzar el nivel de rendimiento ideal.

Existen varios tipos de algoritmos de aprendizaje supervisados como:

- Regresión
- Clasificación
- Árboles de decisión
- Máquinas de soporte vectorial (SVM)

4.4.3 Aprendizaje No supervisado

Definición

Según Flórez and Fernández (2008) afirma que el aprendizaje no supervisado o auto supervisado es cuando el proceso de aprendizaje de las redes neuronales no es inspeccionado ni controlado por ningún agente externo.

En este enfoque se desconoce totalmente la salida y tenemos únicamente la variable de entrada, donde el algoritmo aprende por si mismo y va descubriendo nuevas estructuras en los datos. Aquí se necesita de gran cantidad de información de entrada, de manera que la red dispondrá de suficientes datos para generar patrones y poder obtener resultados de calidad.

LDA

Latent Dirichlet Allocation nos permite dentro de un documento identificar aquellas palabras que pertenecen a un tema. El objetivo de LDA es encontrar los temas a los que pertenece un documento, basándose en las palabras que contiene.

La idea inicial de LDA era la suposición que el documento era considerado una combinación de varios temas, donde la característica de un tema estaba determinada por la distribución de las palabras. Vista una manera de reducir el conjunto de características es utilizar un modelo LDA para la reducción de dimensionalidad, dado que LDA reduce cualquier documento a un conjunto fijo de características de valor real. (Blei et al., 2003)

Dado que es un método no supervisado, los temas no se especifican por adelantado y no se garantiza que se acoplen con la forma en que un humano puede clasificar los documentos por temas de forma natural. Los temas se aprenden como una distribución de probabilidad sobre las palabras que aparecen en cada documento.

4.4.4 Aprendizaje Semi-Supervisado

Donde parte de los datos están etiquetados y otros no. En la que parte de los datos etiquetados pueda ayudar al aprendizaje automático de la parte no etiquetada.

Aquí el primer paso es agrupar datos similares con la ayuda de un algoritmo de aprendizaje no supervisado, después de agrupar estos datos no etiquetados se debe tratar de etiquetar utilizando características de los datos etiquetados disponibles. Para finalmente utilizar algoritmos de aprendizaje supervisado para poder resolver el problema planteado.

Este tipo de aprendizaje es similar a los procesos de la naturaleza y trata de emular como los humanos desarrollan sus habilidades.

4.5 Redes Sociales

4.5.1 Definición

Es una plataforma en línea la cual permite a los usuarios crear perfiles públicos o privados e interactuar con otros usuarios. Generalmente permiten a que un usuario proporcione una lista de personas que comparten una conexión y luego permitir que las personas acepten o nieguen la conexión.

"Las redes sociales son lugares en Internet donde las personas publican y comparten todo tipo de información, personal y profesional, con terceras personas, conocidos y absolutos desconocidos." (Celaya, 2011)

4.5.2 Clasificación

Existen diversos tipos de redes sociales, de tal forma que se han tomado en consideración para su clasificación aquellas que presentan una mayor cantidad de usuarios. (Hutt Herrera, 2012)

Estas son:

- **Redes Generalistas:** Este tipo de redes sociales se enfoca mas en fomentar las relaciones entre las personas, donde se ingresa únicamente información personal básica y nuestros gustos, con la finalidad de contactarse con otras personas para compartir estos gustos que los une. Dentro de este grupo no existe mucho control sobre la información, por lo que es muy fácil encontrarnos con perfiles o noticias falsas. Estas son: Facebook, Instagram, Twitter, WhatsApp.
- **Redes Profesionales:** Son plataformas que se enfocan en fortalecer el “networking” creando a través de ellas conexiones laborales que pueden funcionar como bolsas de trabajo, como buscadores de inversionistas e inclusive socios de negocio, esto en base a

su hoja de vida donde se forma estos enlaces en base a sus gustos, expectativas o preferencias. Las más representativas son: LinkedIn, Viadeo, Xing.

- **Redes Especializadas:** Este tipo están enfocadas en una determinada actividad social, económica, deportes o académicas. Permitiendo satisfacer las necesidades del ser humano de formar parte de un grupo con intereses específicos. Por ejemplo: Ediciona, eBuga, CinemaVIP.

4.6 Minería de Datos

4.6.1 Definición

Minería de datos o también llamada Knowledge Discovery in Databases (KDD) traducido al español “extracción del conocimiento” es el descubrimiento de información novedosa y de gran utilidad a partir de grandes cantidades de datos. De manera ordenada desarrollar una estructura de tal forma que sea comprensible y poder tomar decisiones a partir de esta información crucial que se encuentre. (Friedman, 1998)

La necesidad de la minería de datos está motivada por los desafíos planteados por la gran cantidad de datos disponibles en la actualidad. La cual puede ser una herramienta muy poderosa para extraer información útil entre millones de datos, pero así mismo se puede extraer con la misma facilidad información inútil si no se utiliza correctamente.

4.6.2 Minería de Datos en Redes Sociales

Actualmente la minería de datos en las redes sociales se está convirtiendo en toda una estrategia inevitable para poder comprender las tendencias y los nuevos negocios en línea actuales. Todo esto se debe a que el mundo de las redes sociales se ha convertido en un mar de datos que está en continuo crecimiento, en el que diariamente se publican millones de tweets, blogs y artículos.

Esta gran cantidad de datos ha abierto la puerta a diferentes organizaciones públicas y privadas para promover, analizar y mejorar sus estrategias organizacionales en base a los datos proporcionados en las redes sociales.

Valoraciones y reseñas de productos

Actualmente la dependencia de internet especialmente las redes sociales para informarse a la hora de elegir productos o servicios ha aumentado la necesidad de investigar el boca a boca electrónico, donde estas valoraciones y reseñas de los productos o servicios suelen contener expresiones de sentimiento. (Gamon et al., 2005)

Opiniones de los usuarios

De acuerdo con Technorati cerca de 75 mil nuevos blogs y cerca de 1.2 millones de nuevas publicaciones que opinan sobre productos y servicios se generan cada día. (Kim, 2006)

Para (Kaschesky et al., 2011) las opiniones de los usuarios dentro de las redes sociales pueden ser utilizadas para hacer un descubrimiento y reconocimiento de expresiones positivas o negativas sobre diversos temas de interés. Dichas opiniones suelen ser convincentes y sus indicadores pueden ser utilizadas como una motivación hacia los demás usuarios al hacer compras de algunos productos o servicios, incluso en el respaldo del candidato político.

Permitiendo de esta forma a los posibles clientes tener acceso a información de primera mano sobre los productos antes de realizar la compra, donde los productos mal valorados tienden a no atraer a ningún comprador.

Capítulo V

METODOLOGÍA

5.1 Levantamiento de Requerimientos

Es de gran importancia considerar que para todo proyecto de software es imprescindible los requerimientos o requisitos pues es una parte primordial ya que nos instituyen el punto de partida para las actividades. Permittiéndonos de esta manera tener un concepto claro de lo que se pretende cumplir con el sistema.

5.1.1 Alcance

Este software será utilizado por empresas, emprendedores, pequeños negocios, entre otros, que necesiten utilizar la herramienta para poder analizar la perspectiva de los usuarios hacia un producto específico publicado en las redes sociales.

5.1.2 Personal Involucrado

Dentro de este proceso se encuentran involucradas varias personas las cuales todos ellos representan diversos intereses y roles dentro de la planificación del proyecto.

Nombre	Eduardo Javier Vazquez Arias
Categoría Profesional	Estudiante
Rol	Desarrollador
Responsabilidad	Análisis, diseño y programación del software del proyecto
Información de contacto	evazqueza@est.ups.edu.ec

Nombre	Gabriel Alejandro León Paredes
Categoría Profesional	Administrador del proyecto
Rol	Ingeniero de Sistemas
Responsabilidad	Verificación y aprobación de los sprint de software del proyecto
Información de contacto	gleon@ups.edu.ec

Tabla 5.1: Personal involucrado en el desarrollo del proyecto

5.1.3 Requisitos Funcionales

- Requisitos funcionales correspondientes al manejo de usuarios.

Código	RE01
Nombre	Autenticación de Usuario
Descripción	La aplicación deberá posibilitar realizar un inicio de sesión de manera obligatoria para poder acceder a sus funcionalidades.
Entradas	Correo, Contraseña
Proceso	Se contará con una plantilla donde se solicitará los datos correspondientes al usuario, posteriormente realizar una verificación de la información ingresada con la información disponible en la base de datos, en caso de cumplir con esta verificación se procederá a la siguiente interfaz, caso contrario el usuario debe recordar sus credenciales e intentar nuevamente.
Salidas	Las salidas van dirigidas a: Usuario. Mensaje de error por no haber completado todos los campos solicitados. Mensaje de error por no haber ingresado correctamente la información.
Prioridad	Alta

Tabla 5.2: Requisito funcional 1.

Código	RE02
Nombre	Registro de Usuario
Descripción	Se registrar en el sistema toda la información necesaria para realizar el proceso de registro de un usuario.
Entradas	Nombre, Dirección, Correo electrónico, Contraseña
Proceso	Deberá contar con un botón registrarme tanto en la plantilla de autenticación como en el menú de navegación al momento de acceder por primera vez al sistema, donde el usuario tendrá que proporcionar los datos solicitados por el sistema. El sistema guardara esta información en su base de datos, y posteriormente ofrece un acceso di-recto a sus funcionalidades.
Salidas	Las salidas van dirigidas a: Usuario. Registro correcto de un usuario. Mensaje de error en el caso de no haber completado todos los campos solicitados.
Prioridad	Alta

Tabla 5.3: Requisito funcional 2.

- Requisitos funcionales de la pantalla principal.

Código	RE03
Nombre	Analizar publicación
Descripción	El sistema deberá permitir realizar el análisis de nuevas publicaciones de productos publicados en una red social.
Entradas	URL de la publicación, Numero de Tópicos
Proceso	Para cumplir con este requerimiento el usuario primero deberá autenticarse con éxito, y se deberá mostrar una plantilla con una sección donde el usuario proporcionará la información solicitada por el sistema para realizar el análisis.
Salidas	Las salidas van dirigidas a: Usuario. Mensaje de error porque la URL no pertenece a una red social. Mensaje de error debido a que no lleno los campos solicitados. Mensaje de aviso que se encuentra analizando la publicación. Mensaje de error en el caso de que llegase a ocurrir alguna falla con los servicios web.
Prioridad	Alta

Tabla 5.4: Requisito funcional 3.

Código	RE04
Nombre	Registro de las publicaciones analizadas
Descripción	El sistema deberá tener un registro de las publicaciones analizadas correspondientes a cada usuario.
Entradas	Autenticación de usuario.
Proceso	Para cumplir con este requerimiento deberá existir una plantilla o una sección donde se visualice una lista de todas las publicaciones que se han analizado con dicha cuenta de usuario, ordenadas por fecha y hora de análisis. Deberá también permitir ordenar por criterios y re-alizar una búsqueda de la información presentada. Debe presentar información esencial de cada publicación como nombre, cantidad de comentarios, aceptación, etc.
Salidas	Las salidas van dirigidas a: Usuario. Información detallada de las publicaciones.
Prioridad	Alta

Tabla 5.5: Requisito funcional 4.

- Requisitos funcionales sobre el perfil de usuario.

Código	RE05
Nombre	Modificar datos de Usuario
Descripción	El sistema deberá poder cambiar la información personal del usuario ingresada al momento del registro.
Entradas	Nombre, Dirección, Correo Electrónico, Contraseña
Proceso	Para cumplir con este requerimiento, una vez realizado la autenticación en el menú de navegación deberá existir un apartado de Perfil e ingresar la información que se desea cambiar y guardar. El sistema deberá actualizar esta nueva información en su base de datos.
Salidas	Las salidas van dirigidas a: Usuario. Mensaje de éxito actualizado correctamente los datos. Mensaje de error en el caso que el usuario deje en blanco uno de los campos de texto.
Prioridad	Alta

Tabla 5.6: Requisito funcional 5.

- Requisitos funcionales correspondientes a los resultados del análisis de una publicación.

Código	RE06
Nombre	Detección de sentimientos de los comentarios de una publicación.
Descripción	El sistema deberá poder reconocer los sentimientos que los usuarios quisieron expresar mediante comentarios.
Entradas	Selección de una publicación.
Proceso	Para cumplir con este requerimiento se deberá mostrar una plantilla donde se encuentre una sección con los comentarios que los usuarios realizaron en dicha publicación, con el sentimiento detectado.
Salidas	Las salidas van dirigidas a: Usuario. Detección de sentimiento de cada uno de los comentarios.
Prioridad	Media

Tabla 5.7: Requisito funcional 6.

Código	RE07
Nombre	Detección tópicos.
Descripción	El sistema deberá poder identificar tópicos representativos de una publicación.
Entradas	Selección de una publicación.
Proceso	Para cumplir con este requerimiento se deberá mostrar una plantilla donde se encuentre una sección donde se muestre los tópicos de los que más se hablan dentro de una publicación. Deberá poder identificar a que tópico pertenece cada uno de los comentarios.
Salidas	Las salidas van dirigidas a: Usuario. Detección de tópicos de cada uno de los comentarios.
Prioridad	Media

Tabla 5.8: Requisito funcional 7.

Código	RE08
Nombre	Graficas de los sentimientos.
Descripción	El sistema deberá mostrar la cantidad de sentimientos de la publicación.
Entradas	Selección de una publicación.
Proceso	Para cumplir con este requerimiento se deberá mostrar una plantilla donde se encuentre una sección donde se mostrar ya sea por gráficos de barras, pie o dona indicando la cantidad de sentimientos detectados en los comentarios.
Salidas	Las salidas van dirigidas a: Usuario. Gráfico iterativo con los porcentajes de cada sección.
Prioridad	Media

Tabla 5.9: Requisito funcional 8.

Código	RE09
Nombre	Graficas de los tópicos.
Descripción	El sistema deberá mostrar la cantidad de sentimientos de la publicación.
Entradas	Selección de una publicación.
Proceso	Para cumplir con este requerimiento se deberá mostrar una plantilla donde se encuentre una sección donde se mostrar ya sea por gráficos de barras, pie o dona la distribución de los comentarios para cada uno de los tópicos que se obtuvieron de la publicación.
Salidas	Las salidas van dirigidas a: Usuario. Gráfico iterativo con la distribución de los comentarios por tópicos.
Prioridad	Media

Tabla 5.10: Requisito funcional 9.

Código	RE10
Nombre	Palabras de los tópicos.
Descripción	El sistema deberá mostrar cuales son las palabras claves por cada uno de los tópicos.
Entradas	Selección de una publicación.
Proceso	Para cumplir con este requerimiento se deberá mostrar una plantilla donde se encuentre una sección donde se muestre las palabras pertenecientes a cada uno de los tópicos encontrados en la publicación.
Salidas	Las salidas van dirigidas a: Usuario. Imagen con distribución por tópicos.
Prioridad	Media

Tabla 5.11: Requisito funcional 10.

5.1.4 Requisitos No Funcionales

Código	RE11
Nombre	Pantalla de carga.
Descripción	El sistema muestra una pantalla de carga al momento de realizar consultas
Entradas	Acción de consulta.
Proceso	El sistema podrá mostrar una pantalla de carga cada vez que haga una consulta alguna de los servicios web.
Salidas	Las salidas van dirigidas a: Usuario. Pantalla de carga animada.
Prioridad	Baja

Tabla 5.12: Requisito no funcional 1.

Código	RE12
Nombre	Vista previa de los resultados.
Descripción	Mostrar una vista previa al usuario sobre el análisis realizado en la publicación.
Entradas	Acción de consulta.
Proceso	El sistema podrá mostrar una pequeña vista previa de los resultados obtenidos del análisis de la publicación como pueden ser los comentarios, sentimientos o tópicos.
Salidas	Las salidas van dirigidas a: Usuario. Comentarios.
Prioridad	Baja

Tabla 5.13: Requisito no funcional 2.

Código	RE13
Nombre	Gráfico de las reacciones de la publicación.
Descripción	El sistema podrá mostrar grafico distribución de las reacciones.
Entradas	Selección de publicación.
Proceso	El sistema podrá mostrar un gráfico interactivo para poder saber cuántas y cuáles son las diferentes reacciones que a tenido por parte de los usuarios.
Salidas	Las salidas van dirigidas a: Usuario. Gráfico iterativo con las reacciones que ha tenido la publicación.
Prioridad	Baja

Tabla 5.14: Requisito no funcional 3.

Disponibilidad

El sistema deberá estar disponible todo el tiempo, 24 horas al día, los 7 días a la semana y garantizado un servicio ante cualquier falla en el sistema.

Seguridad

Garantizar la correcta administración de la información personal de los usuarios, tales como el uso de contraseñas encriptadas. Únicamente podrá ser visualizada y actualizada por el usuario.

5.2 Diagramas para la aplicación Web

5.2.1 Diagramas de actividades UML

En esta sección mediante diagramas de actividades se representa el flujo de control desde el punto de partida hasta el punto de finalización, mostrando diferentes vías de decisión mientras se finaliza la actividad. Donde podemos también representar tanto el procesamiento secuencial como el procesamiento concurrente de las actividades.

En la figura 5.1 se puede observar el diagrama de actividades tanto para el registro como para la autenticación de los usuarios dentro del sistema.

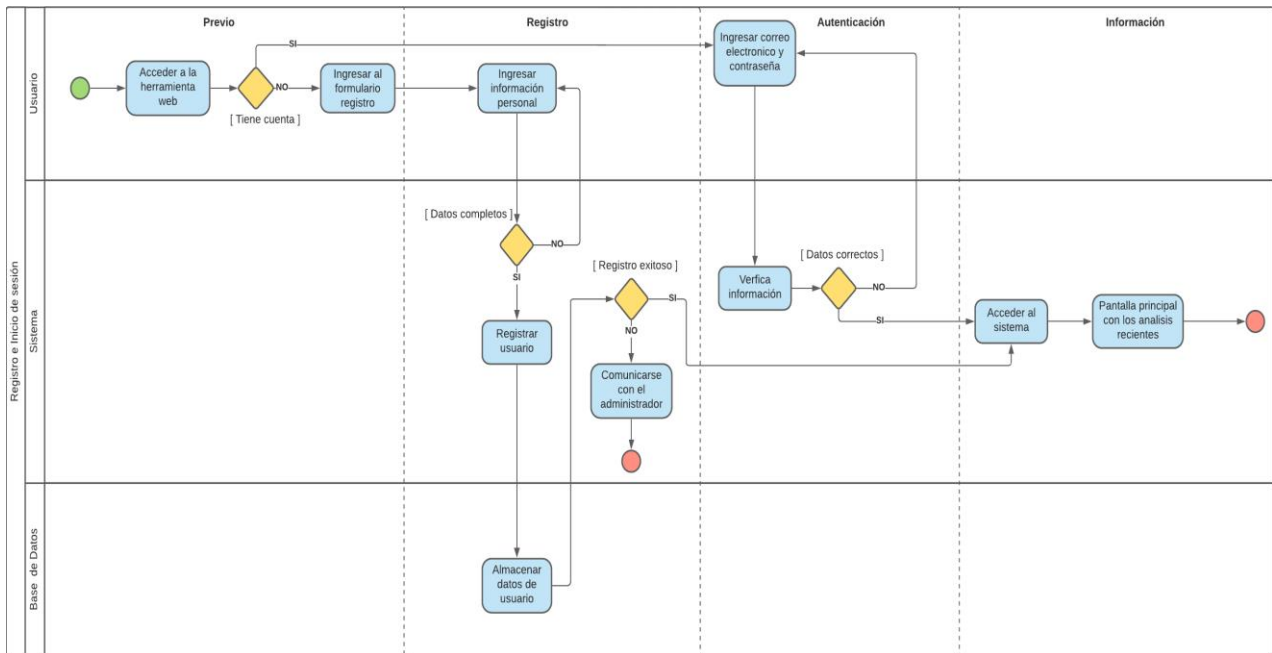


Figura 5.1: Diagrama de Actividades, Registro y Autenticación.

En la figura 5.2 se observa el diagrama de actividades para la modificación de los datos del usuario (Perfil de Usuario).

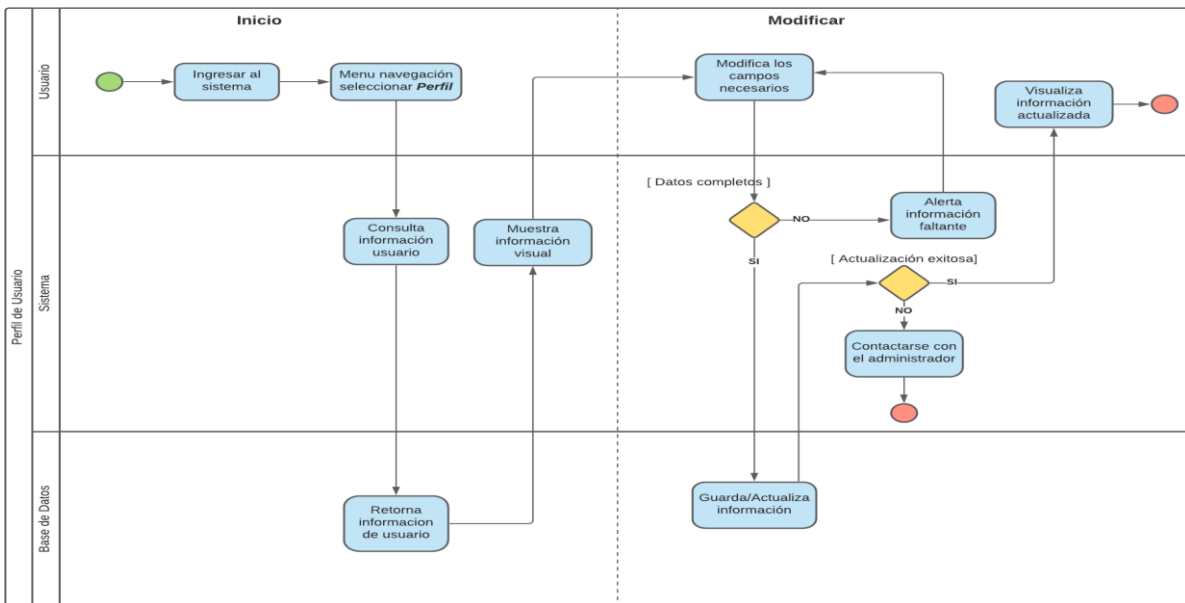


Figura 5.2: Diagrama de Actividades, Modificación del perfil de usuario.

En la figura 5.3 se contempla el diagrama de actividades para realizar un nuevo análisis de una publicación como también acceder a los resultados de una publicación previamente analizada.

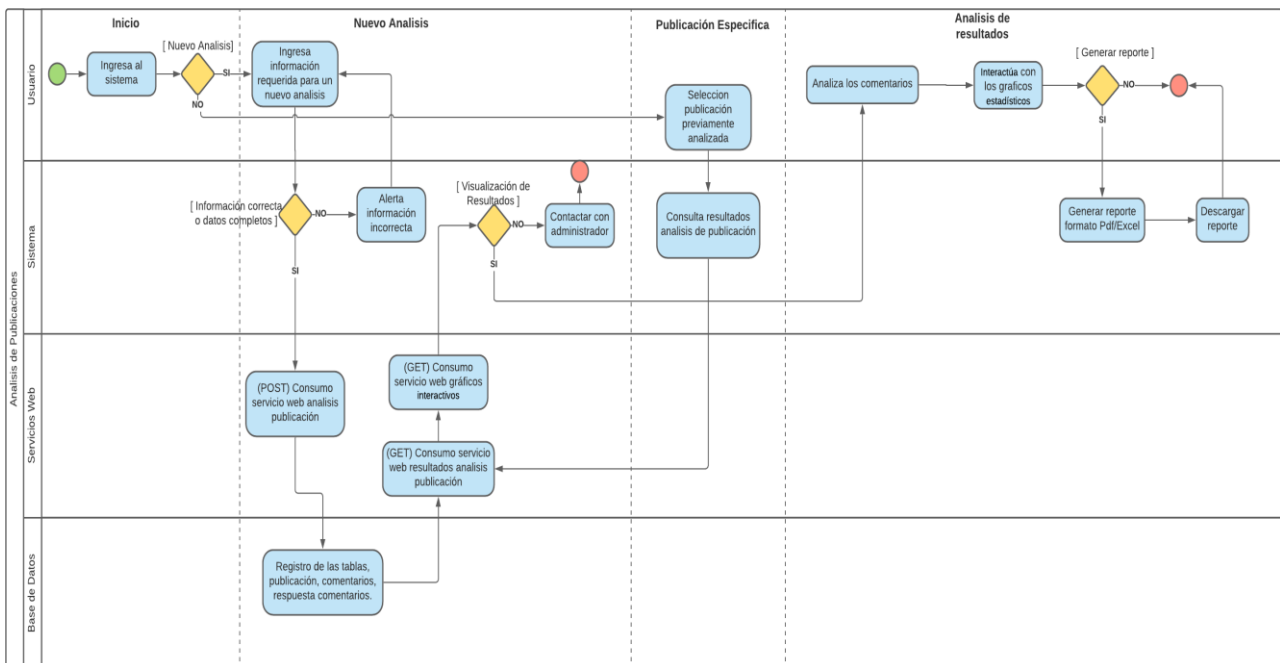


Figura 5.3: Diagrama de Actividades, Analisis de publicación.

5.3 Arquitectura del sistema

La arquitectura del sistema tiene como objetivo reflejar el funcionamiento, estructura y la interacción de las diferentes partes del software. Pudiendo observar en la figura 5.4 el esquema de la arquitectura general del sistema.

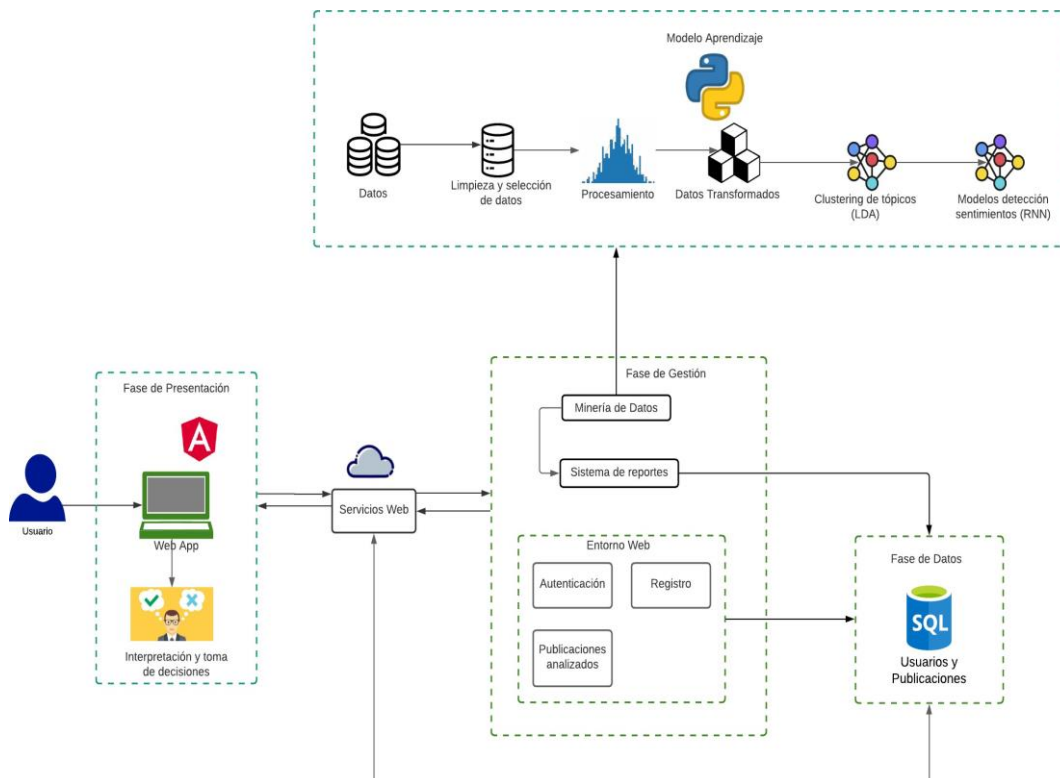


Figura 5.4: Arquitectura General del Sistema.

El sistema está compuesto por varias fases, en las cuales existe una fase de datos la cual contiene un gestor de base de datos MySQL con toda la información de los usuarios y de las publicaciones analizadas dentro de la aplicación.

Una segunda fase de gestión la cual contiene la lógica de funcionamiento del sistema, en la cual primero existe una minería de datos sobre una red social para posteriormente pasar al modelo de aprendizaje donde se realiza todo el procesamiento del lenguaje natural y mediante algoritmos inteligentes realizar la clusterización de tópicos y detección de sentimientos.

Esta misma fase de gestión maneja operaciones relacionadas con la fase de presentación, la cual consta de un control de autenticación, registros y publicaciones analizadas, realizadas hacia la fase de datos donde se encuentra toda la información de los usuarios y los resultados de las publicaciones que fueron analizadas.

Finalmente, la fase de presentación la cual es la encargada de la interfaz de la aplicación web desarrollada con el framework de Angular y mediante el consumo de servicios web se comunica con la fase de gestión previamente mencionada y la fase de datos para proporcionar al usuario la información necesaria y poder realizar una toma de decisiones.

5.3.1 Tecnologías del proyecto

Fase	SubFase	Tecnología	Descripción
Gestión	Fase extracción de la información	Beautiful Soup	Es una biblioteca de Python para extraer datos de archivos HTML y XML.
		Selenium	Permite automatizar las pruebas realizadas en los navegadores web.
	Fase análisis de datos	NLTK	Es un kit de herramientas de lenguaje natural con modelos estadísticos para el PLN.
		spaCy	Cuenta con modelos estadísticos para el PLN en Python incluyendo varios idiomas entre ellos el Español.
		Python	Lenguaje de programación interpretado y multiparadigma.
Datos	Base de Datos	MySQL	Administrador de base de datos relacionales con arquitectura cliente-servidor.
Presentación	Aplicación Web	Angular	Angular es un framework de aplicaciones web de código abierto JavaScript (basado en Type-Script) de Google.
		Node Js	Entorno de ejecución de JavaScript de código abierto y multiplataforma orientado a eventos asíncronos, diseñado para crear aplicaciones escalables.
		JSON Web Token	Es un estándar que se utiliza para compartir información de seguridad entre cliente y servidor
		Servicios Web	Tecnología utilizada para intercambiar datos entre las aplicaciones mediante un conjunto de protocolos y estándares.

5.3.2 Fase extracción de la información

En la figura 5.5 se presenta las diversas tecnologías y herramientas de software empleadas en la extracción de la información (crawler).

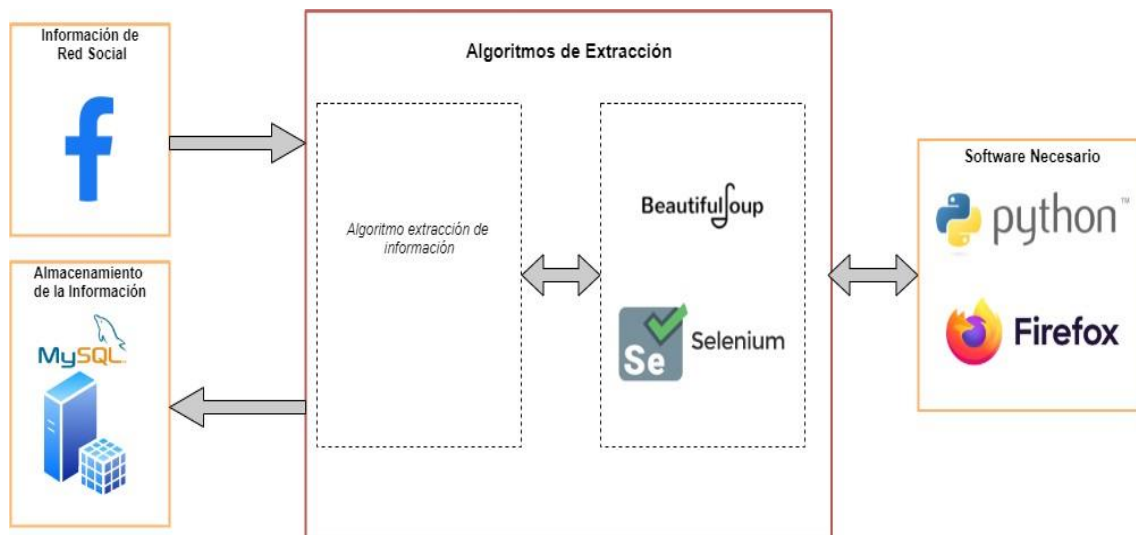


Figura 5.5: Fase extracción de la información.

Para el desarrollo de este crawler, primero debemos contar con una cuenta activa dentro de la red social de Facebook para podernos autenticarnos, de esta forma podremos obtener mayor cantidad de información; ya que al no contar con esta autenticación la red social limita la cantidad de información a la cual podemos acceder.

Posteriormente mediante la herramienta de Selenium la cual permite realizar automatizaciones de tareas en navegadores web, podremos realizar una autenticación dentro de la redsocial antes mencionada, para después acceder a la URL proporcionada por el usuario desde la fase de presentación y navegar hacia las diferentes secciones con las que cuenta la publicación. Mediante BeautifulSoup se realiza un análisis sobre cómo se encuentra estructurada la información dentro del documento HTML. Al ser una red social la cual varía constantemente sus etiquetas HTML y por su puesto también sus atributos, se realizaron varios análisis de tal forma que se logra parametrizar automáticamente estos cambios en sus atributos y poder acceder a la información requerida para su análisis. Por último, toda esta información extraída

es almacenada dentro de una base de datos MySQL.

5.3.3 Fase análisis de datos

Como se puede observar en la siguiente figura 5.6 se presenta el esquema de las tecnologías y herramientas utilizadas para el Análisis de los Datos. Utilizando el lenguaje programación Python, y mediante la ayuda de varias bibliotecas especializadas en el PLN se realizó un procesamiento especializado de la información.

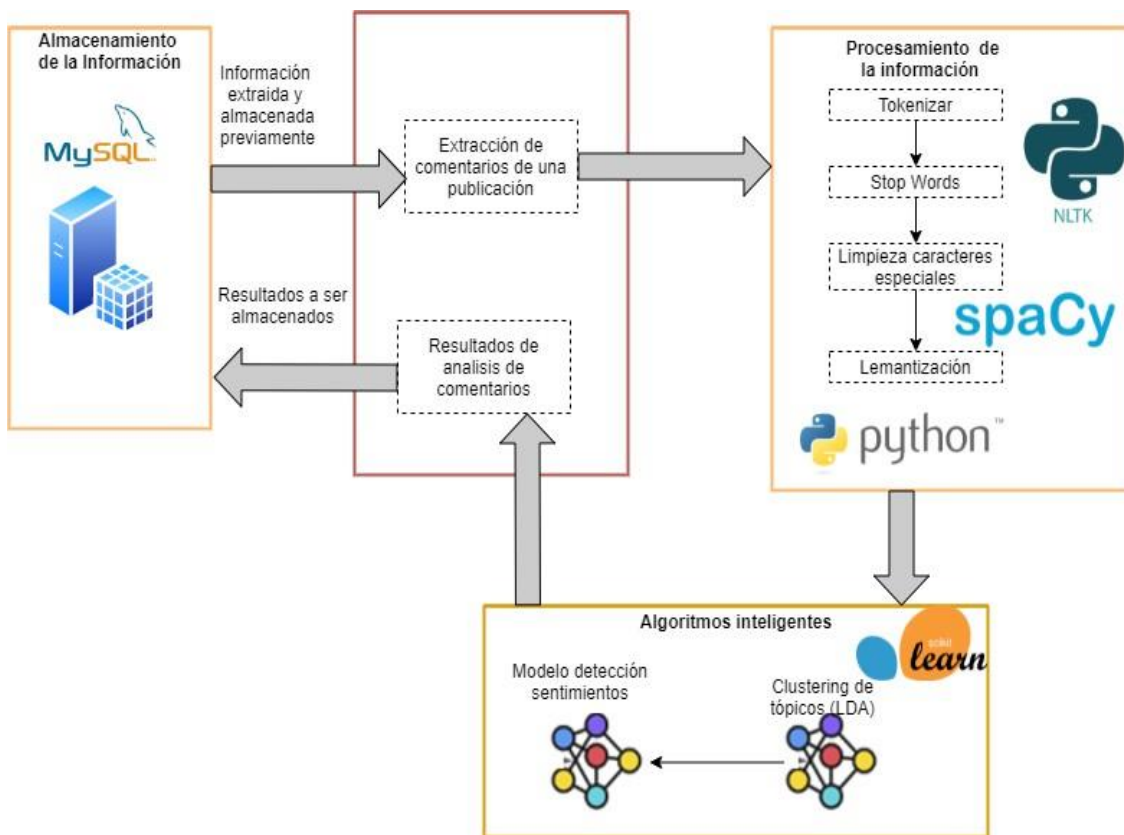


Figura 5.6: Fase Análisis de Datos.

En esta sección, primero se realiza una petición de la información la cual fue extraída y almacenada previamente mediante el crawler de la figura 5.5, de la cual necesitaremos los comentarios pertenecientes a la publicación mismos que son los datos de entrada hacia el procesamiento de la información.

Dentro del procesamiento de la información, mediante NLTK la cual es una biblioteca de Python de modo que permite a los usuarios centrarse en el desarrollo de componentes y sistemas de PLN. Se realiza una tokenización del texto de entrada, proceso que consta en separar un texto en unidades más pequeñas llamadas tokens.

Después de haber tokenizado el texto el siguiente paso es remover los “stop words” en dicho proceso se realiza la eliminación de palabras que no aportan ningún significado al texto analizado. Para luego proceder a realizar una limpieza de caracteres especiales como tildes, emojis, etiquetas HTML y números.

Luego de realizar la limpieza de texto, gracias a la biblioteca spaCy la cual cuenta con modelos estadísticos para varios idiomas entre ellos el Español, este se adapta de mejor manera para realizar procesos de lematización de textos en la que cada palabra es convertida en su forma canónica o de diccionario

Clusterización de tópicos

Para la generación del algoritmo LDA necesitaremos el texto previamente procesado (PLN) estos servirán como datos de entrada para la generación de un diccionario de palabras, que posteriormente permitirá la creación de un corpus de todos los comentarios obtenidos previamente.

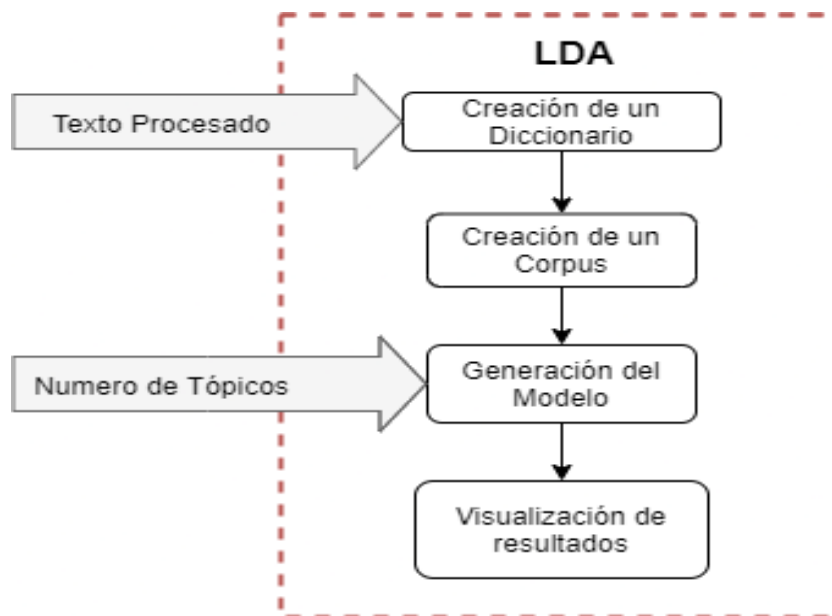


Figura 5.7: Algoritmo detección de tópicos.

Como se observa en la figura 5.7 para la generación del modelo después de realizar varios análisis durante el desarrollo se definieron los siguientes parámetros como los más importantes a configurar:

- **Cantidad de tópicos:** dato de entrada principal que el algoritmo necesita para poder realizar el análisis, mismo que es propuesto por el usuario.
- **Alfa y eta:** estos afectan a la densidad de los tópicos, el cual por defecto es de $1/\text{num_topics}$.
- **Passes:** definida como la cantidad de veces que pasa por el corpus durante el entrenamiento.
- **Chunksize:** número de documentos que se toman en cuenta durante cada pasada del entrenamiento.

Pudiendo observar en la tabla 5.15 cuales fueron los parámetros óptimos para la generación del modelo propuesto.

Modelo	Parámetros				
	Num_Topics	Passes	Chunksize	Iterations	Alfa y eta
LDA	X	30	5	250	1/X

Tabla 5.15: Parámetros óptimos modelo LDA.

Para finalizar y poder contar con una visualización interactiva de los tópicos y sus palabras claves, se hace uso de librerías como pyLDAvis que permite una forma simple de analizar los resultados generados por nuestro modelo.

Modelo detección de sentimientos.

El método propuesto para este trabajo como se puede apreciar en la figura 5.8 consta de 4 fases importantes: 1) Procesamiento de los datos, 2) Diseño y optimización, 3) Predicciones, 4) Evaluación de las predicciones.



Figura 5.8: Modelo propuesto.

1. Procesamiento de los datos

Aquí se busca realizar una limpieza al conjunto de datos como: PLN (limpieza de puntuación, números, eliminación stopwords, lematización), normalizar los valores de

salida en valores de 0 y 1. Con la finalidad de contar con un conjunto de datos estructurados y normalizados.

2. Diseño y Optimización

Esta fase es de suma importancia ya que aquí se realiza el diseño inicial de un modelo de red neuronal recurrente lo que más adelante pasaría a un proceso de Fine-tuning.

- **Aprendizaje mediante la red neuronal:** Se diseña un modelo inicial de la red neuronal recurrente con la finalidad de adquirir un valor de referencia y mejorar este valor mediante un proceso de optimización de Fine Tuning.
- **Optimización y Fine-tuning:** En este proceso se busca obtener los mejores parámetros para los valores de: tamaño de batch, número de veces que se ejecutara el modelo durante el entrenamiento (epochs), numero de neuronas para las diferentes capas de nuestro modelo y finalmente regular el valor de dropout con la finalidad de evitar unsobre entrenamiento. Como se observa en la tabla 5.16 se muestran los parámetros de prueba utilizados dentro de este proceso de Finne-Tuning.

Parámetros de Diseño y Optimización				
Cantidad de neuronas	Valores de dropout	Tamaño de bacth	Epochs	Optimizadores
40, 60, 128, 256	0.1, 0.2, 0.3, 0.4	42, 64, 128, 256, 512	2, 3, 5, 7, 9	adam, softmax, rmsprop

Tabla 5.16: Parámetros de prueba, proceso Fine-tuning.

- **Entrenamiento con el modelo de red neuronal optimizado:** Una vez obtenido el modelo de red neuronal recurrente optimizado se procede a realizar el proceso de entrenamiento con el nuevo modelo RNR optimizado con los valores de la tabla 5.17 .

Topología	Compilación		
	Epochs	Tamaño de batch	Optimizador
1 capa LSTM con 128 neuronas 1 capa de dropout con valor 0.3 1 capa de salida con 1 neurona con función de activación softmax	3	256	rsmprop

Tabla 5.17: Topología del modelo de red neuronal.

3. Predicciones

Con nuestro modelo optimizado y entrenado, procedemos a realizar las predicciones con el conjunto de datos de prueba. Donde cada comentario nuevo debe realizar el mismo procesamiento de texto que se hizo con la red neuronal de tal manera que tengan el mismo formato de entrada, y poder así realizar la predicción de sentimiento de los comentarios. Como se observa en la figura 5.9 el proceso para consumir el modelo de detección de sentimientos.



Figura 5.9: Proceso para consumir el modelo de red neuronal.

4. Evaluación de las predicciones

Con las predicciones obtenidas en la fase anterior, procedemos a evaluar los resultados con la finalidad de conocer la calidad de nuestro modelo desarrollado.

5.3.4 Fase de Presentación

Como podemos observar en la figura 5.10 se presenta el esquema de las tecnologías empleadas para la construcción de la fase de presentación. Donde se puede apreciar un sistema distribuido en la cual principalmente los usuarios interactúan con la aplicación web, desarrollada en Angular uno de los framework mas populares para aplicaciones web y mediante servicios web realizar el consumo de los diferentes servicios expuestos en la red.



Figura 5.10: Fase de presentación.

JSON Web Token

Mediante JSON Web Token (JWT) podemos dar cumplimiento al requerimiento funcional de la tabla 5.2 para mantener una forma segura de autenticar usuarios. Esto permitirá compartir

información de seguridad entre el cliente y el servidor, donde cada JWT contiene objetos JSON codificados necesarios para el funcionamiento del sistema. Donde el servidor utiliza esta información para verificar si el usuario cuenta con permisos para realizar la operación que solicita.

Servicios Web

Utilizando servicios web con arquitectura REST se realiza la validación de los datos, integración del análisis de datos, información del usuario e integración de los gráficos estadísticos. Gracias a WS se puede realizar esta incorporación de una manera simple, evitando que sea desarrollado todo en un mismo lenguaje de programación y sobre todo que no se encuentre todo dentro de una misma aplicación.

Interacción de la aplicación web con el servicio web de la fase de gestión

Para la interacción entre la aplicación web y la fase de gestión, se desarrolló un servicio web en Python mediante el framework “Django” el cual es consumido desde la aplicación web desarrollada en Angular al realizar un nuevo análisis de una publicación, este servicio internamente realiza:

- Crawler para obtener la información requerida de cada publicación los cuales sirven como datos de ingreso para el modelo de aprendizaje.
- Procesamiento del lenguaje natural.
- Clusterización de tópicos.
- Detección de sentimientos.

Comunicación de la aplicación web con la base de datos

Para esta comunicación se encuentra desarrollado servicios web en “Node Js” que son los encargados de mantener dicha comunicación con la base de datos, dichos servicios permiten

realizar tareas de CRUD (Create Read Update Delete) pertinentes a la administración de usuarios y de las publicaciones. Dentro de la aplicación web se presentan varias pantallas donde se puede visualizar el consumo de estos servicios web como son para el registro de usuarios, la actualización de los datos personales del usuario y los resultados de la publicación analizada como los comentarios y gráficos interactivos.

Generación de reportes

Para este caso se realizó el consumo de servicios web realizados en “Node Js” el cual retorna la información requerida, tomando también la información del servidor web “Apache” de la cual podemos acceder a diferentes gráficos para permitirnos construir el reporte mismo que es descargable para cada usuario.

Capítulo VI

ANÁLISIS DE RESULTADOS

En este capítulo se presentan los resultados de los algoritmos desarrollados tanto para la clusterización como para la detección de sentimientos, también encontramos las diferentes pruebas efectuadas a la aplicación web, donde se contó con diferentes usuarios los cuales son los administradores de las redes sociales de las empresas de diferentes ciudades, que realizan uso continuo de las redes sociales para promocionar sus productos.

Utilizando la plataforma de zoom se pudo explicar el objetivo de la herramienta y posteriormente los usuarios procedieron a experimentar con ella, para finalmente proporcionarle una encuesta con varias preguntas que validaran el cumplimiento de los objetivos planteados.

6.1 Resultados de los algoritmos desarrollados

6.1.1 Clusterización de tópicos.

Para el análisis de la clusterización de tópicos se hace uso de los comentarios obtenidos de la figura 6.5 y figura 6.6 pertenecientes al del Análisis #1 de la sección Resultados a diferencia de que se realiza para la detección de 3 tópicos.

```

[(0,
  '0.085*"buen" + 0.042*"bateria" + 0.036*"telefono" + 0.030*"camara" + '
  '0.022*"duro" + 0.021*"celular" + 0.019*"bastante" + 0.018*"duracion" + '
  '0.017*"malo" + 0.017*"hora"'),
(1,
  '0.062*"mejor" + 0.051*"samsung" + 0.051*"existir" + 0.051*"alternativa" + '
  '0.026*"telefono" + 0.023*"funcionar" + 0.016*"correctamente" + '
  '0.013*"perfeccion" + 0.012*"huella" + 0.012*"problema"'),
(2,
  '0.024*"aplicación" + 0.022*"bueno" + 0.019*"digital" + '
  '0.019*"almacenamiento" + 0.019*"huella" + 0.017*"celular" + 0.017*"ram" + '
  '0.015*"excelente" + 0.015*"recomeir" + 0.014*"reconocer"')]

```

Figura 6.1: Tópicos detectados.

Mostrando por cada uno cuales fueron todas las palabras claves perteneciente a dicho tópico, indicando como primera palabra la que mayor otorga al tópico detectado. Al realizar un análisis se identifica en el primer tópico hablar de la batería del teléfono, el segundo tópico menciona la marca a la que pertenece el teléfono y finalmente en el último tópico menciona las características del teléfono entre ellas las que se relacionan como “huella digital” y “almacenamiento y ram”.

En la figura 6.2 se visualiza de una manera interactiva nuestro modelo generado, en el cual cada una de las burbujas de la de la izquierda representa cada tópico detectado mientras que las palabras de la parte derecha son las palabras claves más importantes que conforman el tópico escogido. Donde un buen modelo es aquel que cuenta con burbujas grandes dispersas en todo el grafico y que no se solapen.

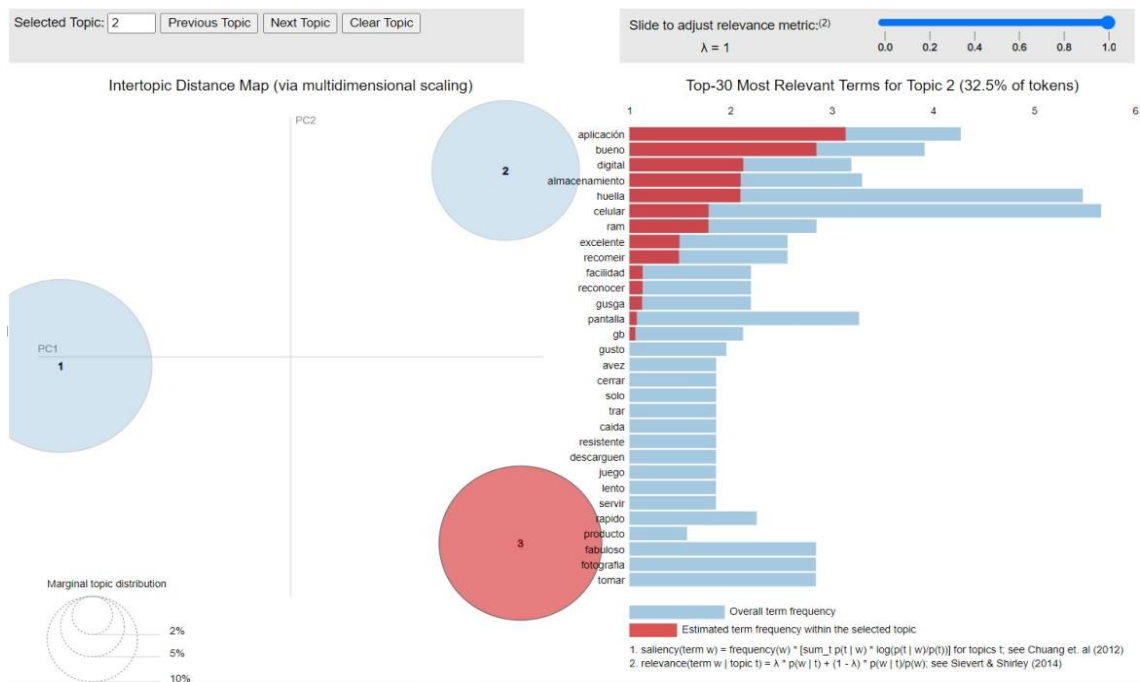


Figura 6.2: Grafico interactivo modelo LDA.

La burbuja numero 3 claramente muestra las palabras que están relacionadas entre ellas como “huella digital” y “almacenamiento y ram” mismas que fueron detectadas por el modelo en la figura 6.1 para el tópico #2, de igual manera para la burbuja 2 la cual habla de la marca del fabricante del teléfono perteneciente al tópico #1, finalmente la burbuja 1 que menciona las mismas características de batería del teléfono, pertenecientes al tópico #0.

Para conocer el nivel de precisión de nuestro modelo, se desarrolla el fragmento de código de la figura 6.3 la cual indica cuan coherente son los tópicos detectados. Esta medida utiliza cosine similarity permitiendo medir la similitud entre dos vectores en un espacio. En este caso la coherencia es de un 65.21%, cabe recalcar que a mayor cantidad de documentos (comentarios) el modelo mejora y por ende la coherencia de los tópicos aumenta.

```
from gensim.models.coherencemodel import CoherenceModel
coherence_model_lda = CoherenceModel(model=lda_model, texts=leman, dictionary=diccionario_LDA, coherence='c_v')
coherence_lda = coherence_model_lda.get_coherence()
print('Coherencia: ', coherence_lda*100)
```

Coherencia: 65.211507324680625

Figura 6.3: Coherencia modelo LDA.

6.1.2 Detección de sentimientos.

Con la optimización realizada ya podemos obtener el nuevo modelo de red neuronal recurrente optimizado, por lo cual se procede a realizar el entrenamiento de está utilizando los parámetros expuestos en la tabla 5.17.

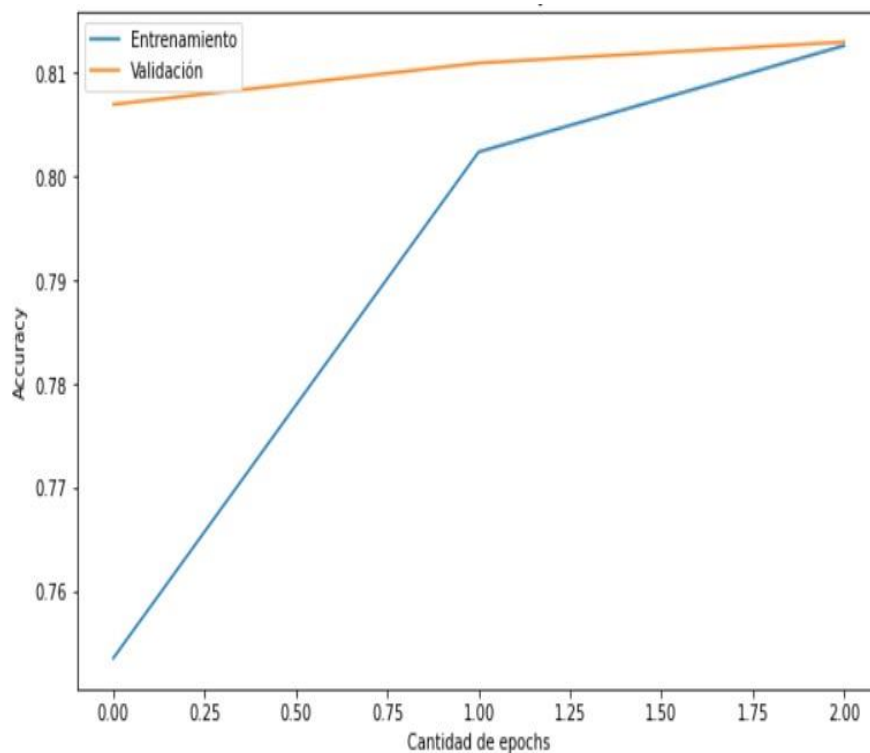


Figura 6.4: Entrenamiento y validación del modelo propuesto.

Como se presenta en la figura el entrenamiento realizado por la red neuronal recurrente optimizada, donde el valor de entrenamiento al inicio es bajo, pero a medida que avanza en la cantidad de epochs este valor tiende a aumentar. Sucede lo mismo en la validación, a diferencia

de que en el primer epoch el nivel de precisión no es tan bajo. Se puede denotar tanto en el entrenamiento como en la validación tienden a aumentar, de tal forma que en el último epoch se tiene el nivel de precisión alto tanto para el entrenamiento como la validación, indicando así cuán importante es el proceso de optimización.

Validación

Para realizar las validaciones del correcto funcionamiento de nuestra red neuronal recurrente, como se observa en la tabla 6.1 se la realiza con varios comentarios extraídos del conjunto de datos con el cual fue entrenado el modelo e inclusive con nuevos comentarios relacionados a la temática tratada.

Comentario	Valor Porcentaje Texto Positivo	Valor Porcentaje Texto Negativo	Sentimiento
Nada bueno se me fue ka pantalla en menos de 8 meses y no he recibido respuesta del fabricante	13.872	86.127	Negativo
Pésimo sonido y bajo y conectividad nula,	3.877	96.122	Negativo
Calidad horrenda. Los tuve que tirar	26.345	73.654	Negativo
No me gusta nada este protector no pega en los bordes queda fatal no lo recomiendo,	7.815	92.814	Negativo
No sirve para nada, no se puede pegar bien y no detecta la pantalla táctil lo tuve que sacar de inmediato,	24.681	75.318	Negativo
Peor calidad de la esperada,	21.055	78.944	Negativo
Bueno muy bien tal como es en la foto,	98.015	1.984	Positivo
Tiene una excelente pantalla, una muy buena cámara y sobre todo la batería me dura todo el día	95,258	4.742	Positivo
La calidad es equivalente al precio. Si necesitas algo sencillo y barato, es perfecto.	98.015	1.984	Positivo
Es precioso. Mi reloj preferido ahora. Me encanta!! Recomiendo al vendedor y el producto. Es grande como me gustan los relojes	92.724	7.275	Positivo
Cumple con sus expectativas, de momento va muy bien	94.617	9.461	Positivo

Tabla 6.1: Validación del modelo de red neuronal recurrente.

6.2 Resultados del análisis de la perspectiva de los consumidores acerca de un producto específico en una red social.

En esta sección se presenta los resultados de una de las cuentas registradas en la aplicación, donde se realizaron análisis de varios productos.

Primero se realiza la extracción de la información (comentarios) de la publicación, posteriormente se realiza el análisis de los datos donde se detectaron tópicos y sentimientos para finalmente presentarlos en la interfaz web juntamente con los gráficos estadísticos de este análisis.

6.2.1 Resultados

- **Análisis #1**

El primer análisis realizado fue de un teléfono celular, específicamente de un celular A70 donde podemos observar en la figura 6.5 y figura 6.6 se presenta los comentarios extraídos de la publicación, mismos que fueron analizados mediante los algoritmos inteligentes determinando cual comentario pertenece al número de tópico analizado, y de la misma manera se realiza la detección de sentimiento que quiso expresar el usuario.

Comentario ↑↓	Sentimiento ↑↓	Topico ↑↓
tiene una muy buena batería, me a durado todo un día usandolo al maximo	Positivo	1
no me gusto el reconocimiento de la huella digital, siempre tiene fallos no es su punto fuerte	Positivo	3
lo recomiendo mucho, la fluidez me encanto mucho para correr las aplicaciones actuales, super buena y rapida	Positivo	4
Un celular que por el precio realmente ofrece muchas características, recomendado para el uso diario básico	Negativo	4
El A70 luce una impresionante pantalla de 6.7 pulgadas; La pantalla es brillante con muy alto contraste y excelente definición. Tiene 128 GB de almacenamiento integrado, un procesador de 8 núcleos, 6 GB de RAM y puede usar una tarjeta micro SD para almacenamiento adicional.	Positivo	1
El teléfono tiene una batería que dura hasta 104 horas, y mi experiencia hasta ahora es que es increíblemente buena. La mayoría de los teléfonos con muy buena duración de la batería tienen entre 80 y 90 horas, por lo que este viene muy por encima de eso.	Negativo	2
El A70 tiene reconocimiento facial y un lector de huellas dactilares en pantalla, y ambos funcionan a la perfección. Este teléfono es, con mucho, el mejor teléfono que he tenido. La gente cuestiona la cubierta de plástico frente a una cubierta de aluminio, pero, sinceramente, nunca lo sabrías si no te lo dijeran.	Positivo	2
Este teléfono no fue desbloqueado correctamente y no pude configurarlo para ponerlo		

Figura 6.5: Análisis #1, Resultados análisis de comentarios de producto específico.

Comentario ↑↓	Sentimiento ↑↓	Topico ↑↓
Este teléfono no fue desbloqueado correctamente y no pude configurarlo para ponerlo en mi red. Después de aproximadamente 6 horas tratando de resolver el problema, me di por vencido. Xfinity era el operador anterior y me dijeron que no podían desbloquearlo porque no había una cuenta activa. Así que ten cuidado de que es un gran teléfono, pero si el vendedor no lo desbloquea correctamente, enfrentarás un problema muy grave.	Negativo	2
El teléfono no abre el menú al desplazar para arriba. Los botones de apagado o reinicio no funcionan correctamente. Ni siquiera ha pasado una semana. El lector de huellas dactilares funcionó durante unos días, pero ahora no. Espero que el vendedor corrija esto rápidamente.	Negativo	2
Este teléfono tiene una buena cámara, carga rápida, buen almacenamiento	Negativo	4
He usado bastante el teléfono y estoy muy impresionado. Toma fotografías fabulosas. La duración de la batería también parece ser bastante precisa, pero eso depende de lo que esté haciendo.	Positivo	3
Recomiendo mucho este producto. ¡Es fabuloso!	Positivo	4
Buen celular con una muy buena cámara	Positivo	4
Algunas aplicaciones que se instalan no funcionan a la perfección	Negativo	4
toma fotografías excelentes	Positivo	1

Figura 6.6: Análisis #1, Resultados análisis de comentarios de producto específico parte 2.

A continuación, en la interfaz web se presenta el porcentaje de perspectiva que a tenido el producto por parte de los consumidores, en este caso como se observa en la figura

6.7 a tenido un 71% de una perspectiva positiva con una cantidad de 22 consumidores expresando sentimientos positivos, por ende, tendríamos un 29% no está conforme con este producto un total de 9 consumidores que expresan un sentimiento negativo.

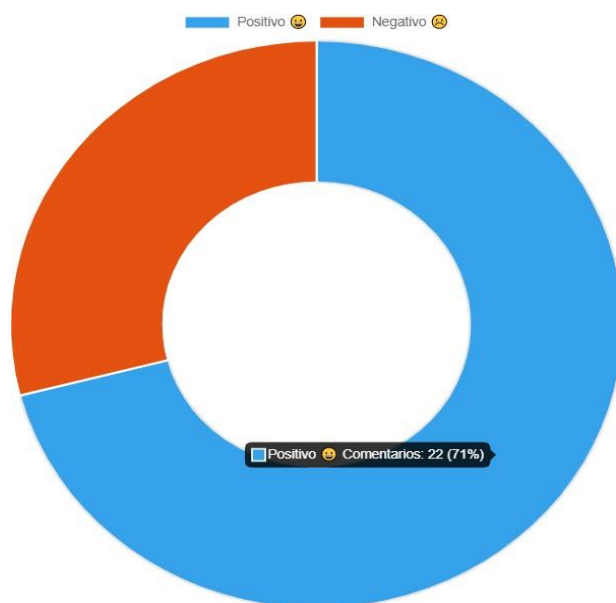


Figura 6.7: Análisis #1, Representación porcentual de la perspectiva de los consumidores.

Seguidamente en la aplicación web como la figura 6.8 muestra cuales han sido las palabras claves por cada uno de los tópicos encontrados (4 tópicos), donde se puede observar que las palabras de mayor tamaño son las principales dentro del tópico.



Distribucion Palabras Topico

Figura 6.8: Análisis #1, Representación distribución palabras claves por tópico.

Como es en el tópico número 1, donde se puede ver que existe una relación entre las palabras: almacenamiento, ram, y batería; esto se podría considerar como “Características del teléfono”. En el tópico número 2, menciona las palabras: Samsung, mejor, teléfono, existe; todo este tópico da a entender que la marca Samsung tiene muy buenos teléfonos. En el tópico número 3 claramente se observa cómo se relaciona aquí las palabras de huella y digital que es característica única de una persona para desbloquear su teléfono celular. Finalmente, en el tópico número 4 las palabras principales son buen, celular, cámara, teléfono lo que se puede determinar también como “Características del teléfono”.

Como se observa en la figura 6.9 se cuenta con una representación visual de los tópicos generados por el modelo LDA generado, donde se puede pasar el cursor sobre los diferentes círculos los cuales representan el número de tópico, este automáticamente en la parte derecha se actualiza cuáles son las palabras claves ordenadas por aquellas que representan mayor peso de cada tópico. En el cual se puede observar que las palabras del tópico seleccionado #3 son las mismas palabras representadas en el tópico # 3 de la figura

6.8.

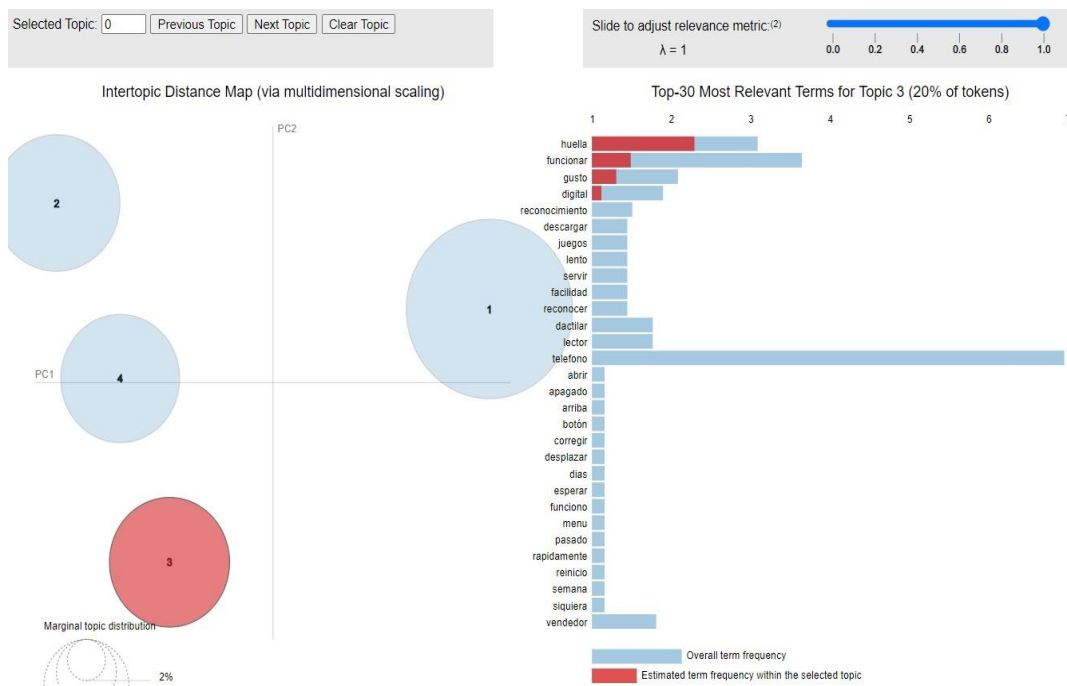


Figura 6.9: Análisis #1, Representación visual tópicos del modelo LDA mediante pyLDAvis.

En la figura 6.10 se observa como los comentarios están agrupados de acuerdo con el número de tópico perteneciente. De igual manera es un gráfico interactivo sobre el cual puede pasar el mouse sobre cada punto de cualquier color y le brindara información detallada del comentario y al tópico perteneciente. Comprobando que el comentario “Un celular que por el precio realmente ofrece muchas características, recomendado para el uso diario básico” pertenece al tópico #4 tal y como se observa en la figura 6.5



Figura 6.10: Análisis #1, Representación visual distribución de comentarios por tópico.

Finalmente, en la figura 6.11 se observa las reacciones que ha tenido la publicación, en este caso únicamente a tenido reacciones “Me gusta”, juntamente se observa la cantidad de comentarios que existe por cada tópico analizado.

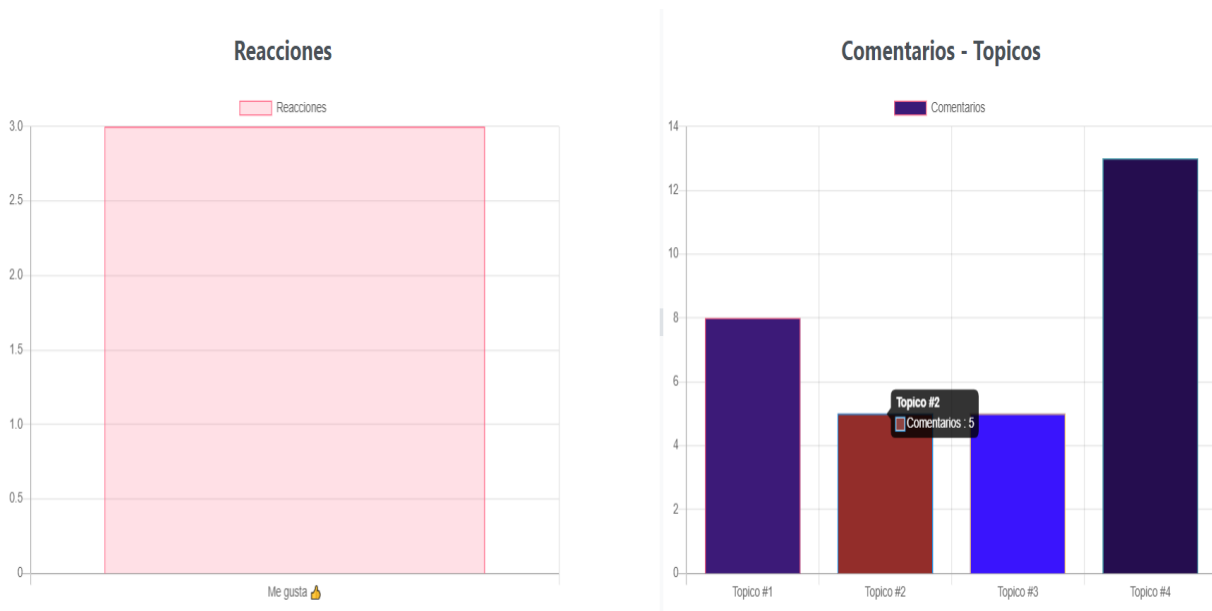


Figura 6.11: Análisis #1, Representación reacciones y cantidad de comentarios por tópico.

- **Análisis #2**

El segundo análisis realizado fue sobre una laptop HP, y como se puede notar en la figura 6.12 se encuentran los comentarios extraídos de la publicación, mismo que fueron analizados por los algoritmos inteligentes tanto para la detección de tópicos como para la detección de sentimientos.

Comentario ↑↓	Sentimiento ↑↓	Topico ↑↓
Cuenta con una excelente batería, tiene una larga duracion	Negativo	1
su pantalla es muy buena, tiene una buen brillo	Positivo	2
No tiene un disco duro lo suficientemente grande. Las actualizaciones de primera mano casi eliminan todo el espacio disponible.	Negativo	3
No recomendado para gente que programa, ya que el tamaño de la pantalla suele causar incomodidad	Positivo	1
Buena computadora pero con un bajo de ram no son tan rápidas para entrar a google	Positivo	2
El disco duro es una desventaja, recomendaría mejor un SSD.	Positivo	3
La computadora tiene unas especificaciones muy básicas. Es un equipo solamente para trabajo de oficina.	Positivo	2
El sistema operativo windows 10 es muy bueno	Negativo	3
Realmente he tenido mala experiencias con esa marca, no la recomiendo para nada.	Positivo	3
El sistema operativo windows es muy bueno ya que es compatible con la mayoría de aplicaciones	Negativo	3

Figura 6.12: Análisis #2, Resultados análisis de comentarios de producto específico.

Esta publicación tal y como se puede apreciar en la figura 6.13 ha tenido una perspectiva positiva del 64% con 7 comentarios expresando un sentimiento positivo, mientras tiene una perspectiva negativa del 36% con 4 comentarios que expresan un sentimiento negativo.

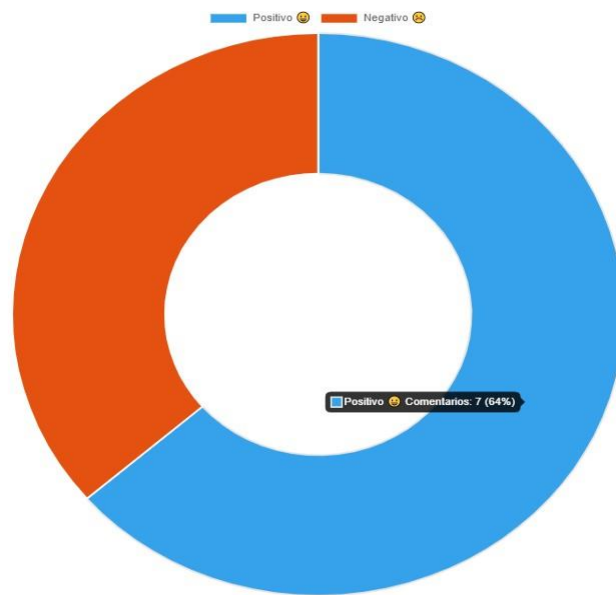


Figura 6.13: Análisis #2, Representación porcentual de la perspectiva de los consumidores.

En la figura 6.14 se visualiza que en el tópico numero 1 existe una relación entre las palabras: gustar, juegos, Windows, recomendado lo que representaría a un tema general como “Windows es recomendado para juegos”. En el tópico número 2, se visualizan las palabras de mayor peso: buen, computadora, trabajar, básica lo que se puede determinar como “Una computadora buena para trabajar”. Finalmente, en el tópico número 3, donde existe claramente una relación entre las palabras: operativo, Windows, disco, compatible se identifica como “Características de la computadora”.

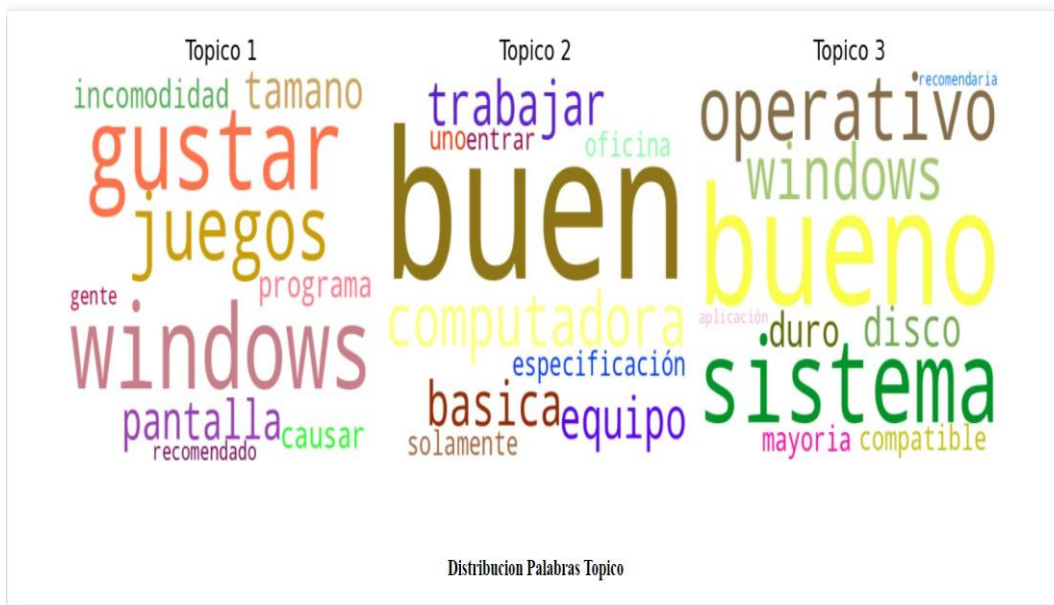


Figura 6.14: Análisis #2, Representación distribución palabras claves por tópico.

Como se observa en la figura 6.15 se cuenta con una representación visual de los tópicos generados por el modelo LDA generado. Donde se observa que las palabras del tópico seleccionado #2 son las mismas palabras representadas en el tópico #2 de la figura 6.14

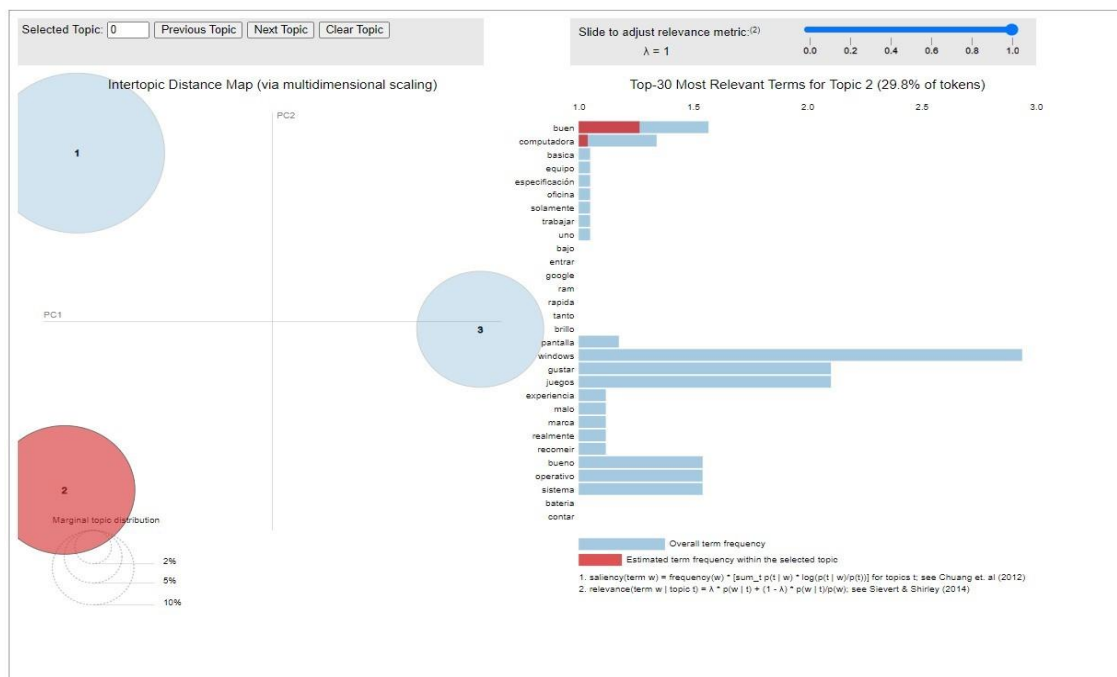


Figura 6.15: Análisis #2, Representación visual tópicos del modelo LDA mediante pyLDAvis.

En la figura 6.16 se observa como los comentarios están agrupados de acuerdo con el número de tópico perteneciente. Brindando información detallada del comentario y al tópico perteneciente. Pudiendo comprobar que el comentario “Realmente he tenido mala experiencias con esa marca, no la recomiendo para nada.” pertenece al tópico #3 tal y como se visualiza en la figura 6.14.



Figura 6.16: Análisis #2, Representación visual distribución de comentarios por tópico.

Finalmente, en la figura 6.17 se presenta las reacciones que ha tenido la publicación, en este caso únicamente a tenido reacciones “Me gusta”, juntamente se observa la cantidad de comentarios que existe por cada tópico analizado.

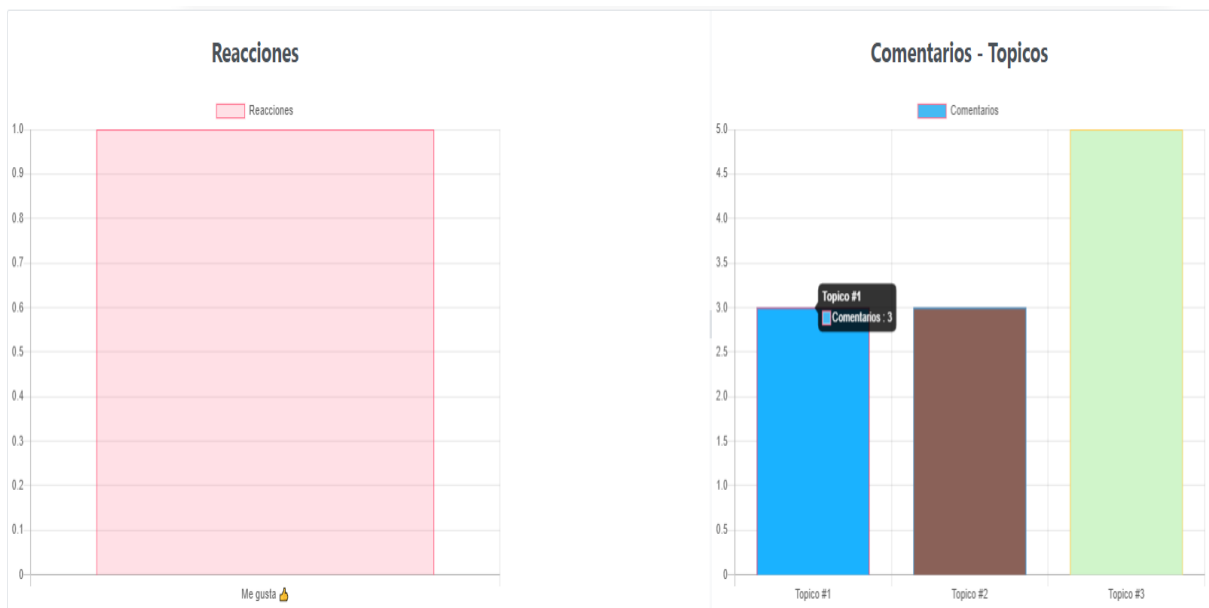


Figura 6.17: Análisis #2, Representación reacciones y cantidad de comentarios por tópico.

- **Análisis #3**

El tercer análisis realizado fue sobre una motocicleta, donde se expone en la figura 6.18 los comentarios extraídos de la publicación, mismos que fueron analizados mediante los algoritmos inteligentes determinando cual comentario pertenece al número de tópico analizado, y de la misma manera se realiza la detección de sentimiento que quiso expresar el usuario.

Comentario ↑↓	Sentimiento ↑↓	Topico ↑↓
precio en quito?	Positivo	2
Stan Gumn ¡Hola! Te enviamos un mensaje para más... More	Negativo	1
Por favor más información para la compra gracias	Positivo	4
Sebas Cortez ¡Hola! Te enviamos un mensaje para... More	Negativo	1
Precio por favor	Positivo	4
Kevin Valles ¡Hola! Te enviamos un mensaje para... More	Positivo	1
Buenas noches infmacion y pecio por favor	Positivo	3
Luisn Ernesto Olivo Martinez ¡Hola! Te enviamos un... More	Positivo	3
PRECIO CONTADO	Positivo	4
David E. Pintag ¡Hola! Te enviamos un mensaje para... More	Positivo	1

Figura 6.18: Análisis #3, Resultados análisis de comentarios de producto específico.

A continuación, se presenta el porcentaje de perspectiva que ha tenido el producto por parte de los consumidores, en este caso como se observa en la figura 6.19 ha tenido un 20% de una perspectiva negativa con una cantidad de 2 consumidores expresando sentimientos negativos, por ende, se tendría un 80% los cuales están conformes con el producto un total de 8 comentarios que expresan sentimientos positivos.

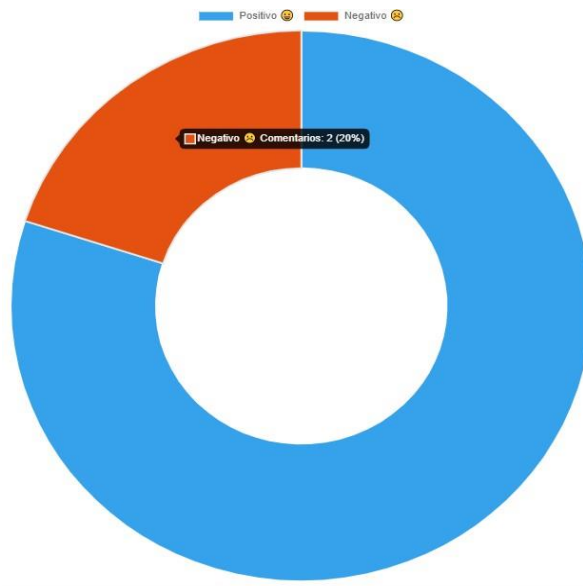


Figura 6.19: Análisis #3, Representación porcentual de la perspectiva de los consumidores.

Seguidamente en la aplicación web como se observa en la figura 6.20 muestra cuales han sido las palabras claves por cada uno de los tópicos encontrados, donde se puede observar que las palabras de mayor tamaño son las principales dentro del tópico.



Distribucion Palabras Topico

Figura 6.20: Análisis #3, Representación distribución palabras claves por topico.

Como resultados, en el tópico #1 las palabras: hola, enviar, y mensaje ayudaría a definir

que este t3pico trata acerca de "Enviar mensajes a los consumidores con informaci3n del producto". En el t3pico #2 con las palabras: quito, precio, enviar y mensaje da a entender que los consumidores necesitan "Informaci3n acerca del producto en una locacion especifica". En el t3pico #3 con las palabras: precio e informaci3n expresa que los consumidores quieren conocer de "Detalles de los precios a trav3s de diferentes m3todos depago". Finalmente, las palabras de mayor peso en el t3pico #4 que son: precio, favor y contado en la que se deduce que los consumidores necesitan "Informaci3n detallada del precio al cantado del producto".

A continuaci3n, en la figura 6.21 se muestra una representaci3n visual de los t3picos generados por el modelo LDA generado, de tal forma que se puede pasar el cursor sobre los diferentes c3rculos los cuales representan el n3mero de t3pico. Pudiendo observar que las palabras del t3pico seleccionado #1 son las mismas palabras representadas en el t3pico #1 de la figura 6.20.

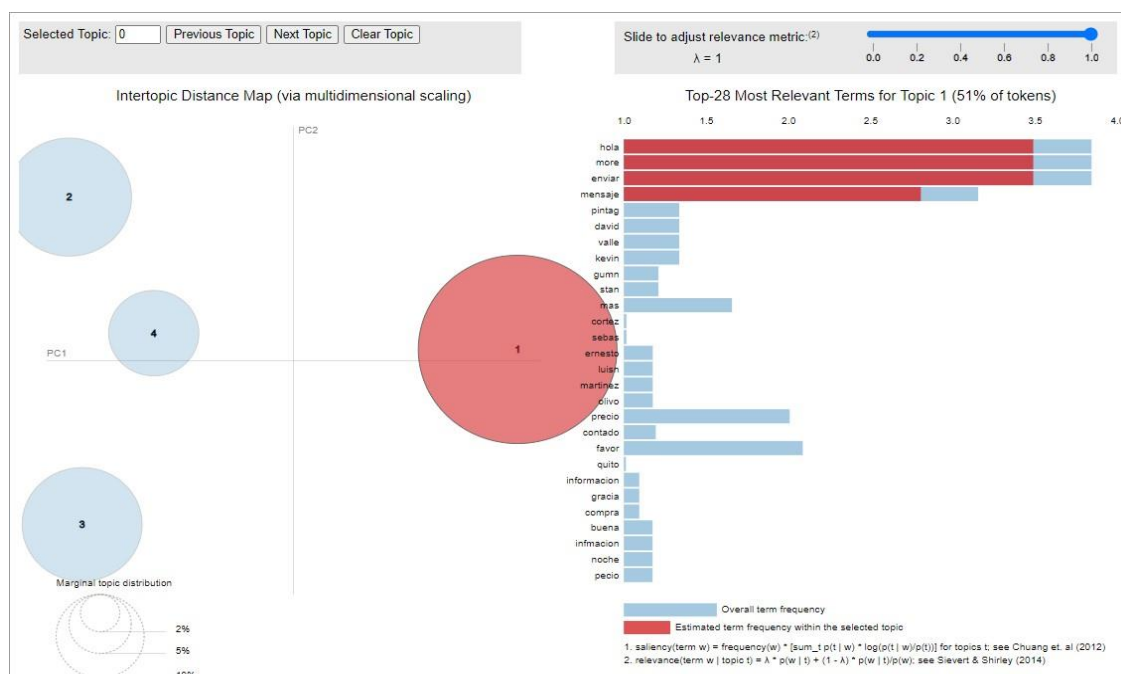


Figura 6.21: An3lisis #3, Representaci3n visual t3picos del modelo LDA mediante pyLDAvis.

En la figura 6.22 se observa como los comentarios están agrupados de acuerdo con el número de tópico perteneciente, de igual manera es un gráfico interactivo sobre el cual puede pasar el mouse sobre cada punto de cualquier color y le brindara información detallada del comentario y al tópico perteneciente. Donde se presenta al comentario perteneciente al tópico #3 tal y comprobando con la figura 6.18 que especifica que pertenece al tópico #3.

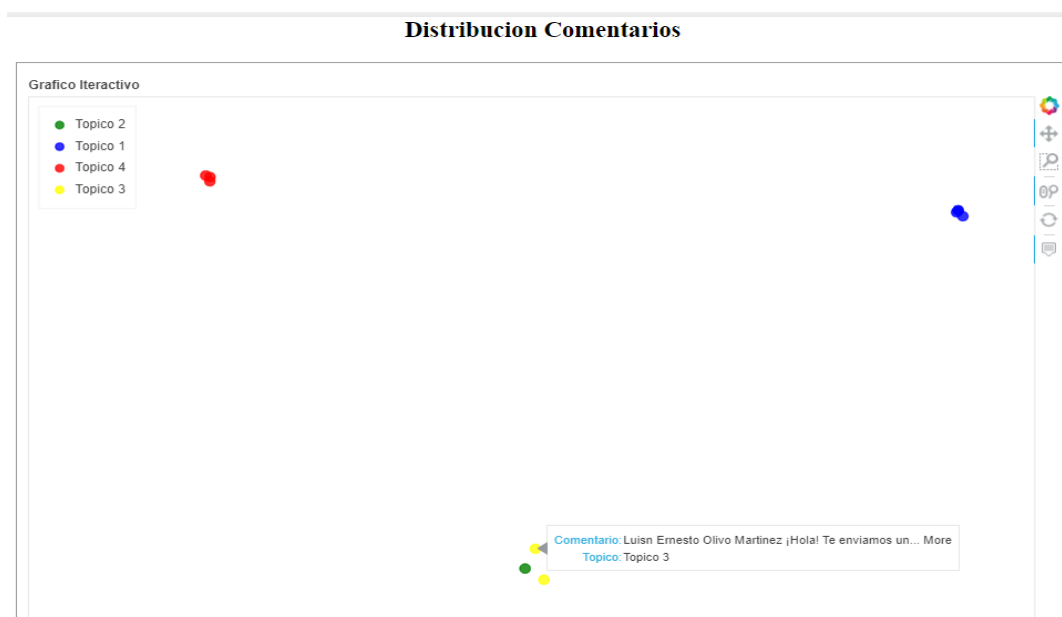


Figura 6.22: Análisis #3, Representación visual distribución de comentarios por tópico.

Finalmente, en la figura 6.23 podemos observar las reacciones que ha tenido la publicación, en este caso ha tenido varias reacciones como “Me gusta”, “Me Encanta” y “MeImporta”; juntamente se observa la cantidad de comentarios que existe por cada tópico analizado.

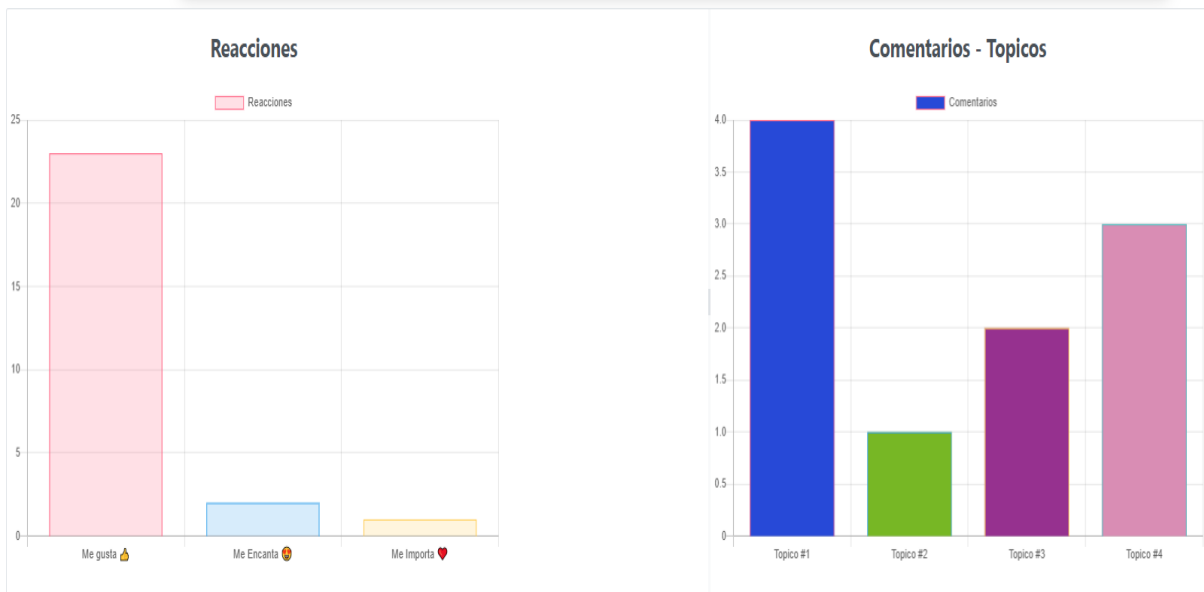


Figura 6.23: Análisis #3, Representación reacciones y cantidad de comentarios por tópico.

6.3 Análisis Encuestas

Se procede a ejecutar una tabulación de las encuestas realizadas, a las empresas de las diferentes ciudades de la provincia de Azuay y Morona Santiago, el cuestionario de las preguntas realizadas se puede hallar Anexo A. Con la finalidad de corroborar el funcionamiento con usuarios reales (administradores redes sociales). Fueron encuestados 11 personas en donde el uso del sistema fue de gran importancia para poder comprender las preguntas y responderlas adecuadamente.

Una vez procesada toda la información de las encuestas, se obtuvieron los siguientes resultados:

- **Pregunta 1. Ciudad donde se encuentra su empresa.**

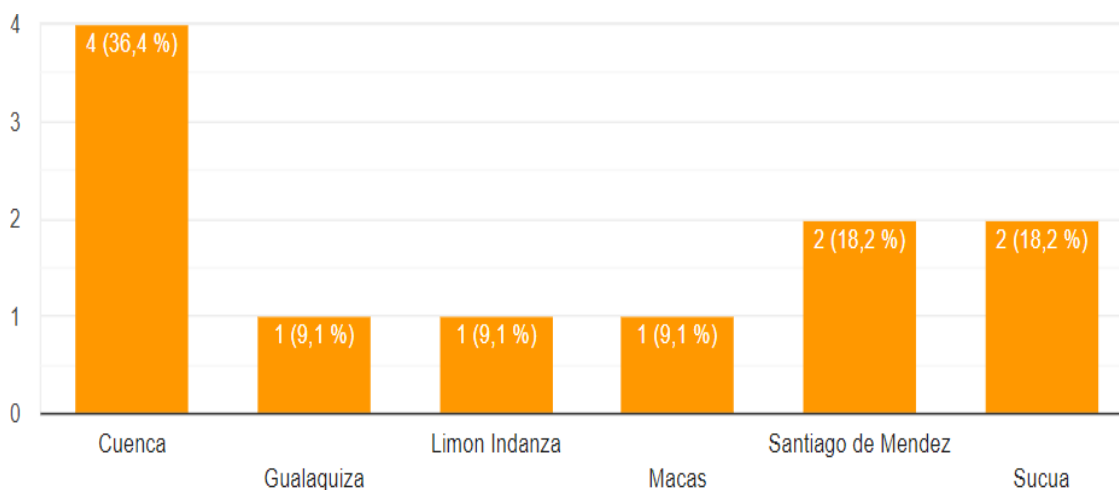


Figura 6.24: Representación porcentual para la pregunta 1.

Interpretación:

Del 100% de usuarios encuestados, el 36.4% pertenecen a la ciudad de Cuenca, el 18.2% de Santiago de Méndez, el 18.2% son de Sucua, el 9.1% es de Gualaquiza, el 9.1% es de Limón Indanza y finalmente el 9.1% pertenece a Macas.

• **Pregunta 2. ¿Qué tipo de productos vende su empresa?**

Alternativas	Respuesta	Porcentaje
Computación	6	54.5%
Hogar y muebles	1	9.1%
Autos, motos u otros	2	18.2%
Celeulares y telefonía	7	63.6%
Deportes y Fitness	0	0%
Electrónica, audio y video	6	54.5%
Ropa y accesorios	0	0%
Otro	2	18.2%

Tabla 6.2: Pregunta 2. ¿Qué tipo de productos vende su empresa?

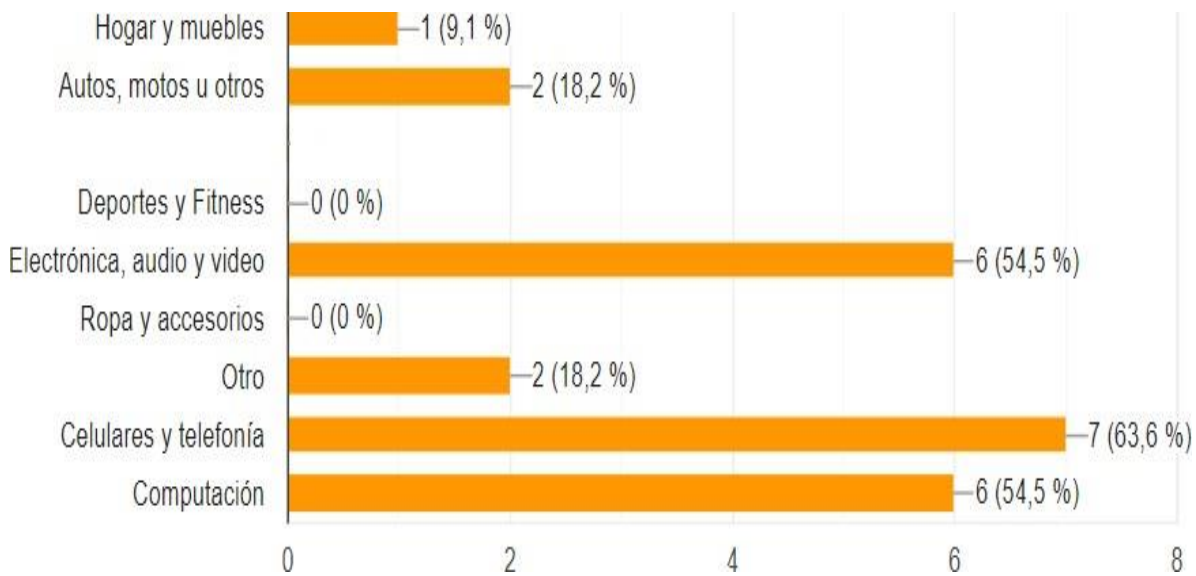


Figura 6.25: Representación porcentual para la pregunta 2.

Interpretación:

Del 100% de usuarios encuestados el 63.6% vende productos de Celulares y telefonía, el 54.5% venden productos de Computación y de Electrónica, audio y video, el 18.2% venden productos de Autos, motos y Otros, finalmente el 9.1% venden productos de Hogar y muebles

- **Pregunta 3. Cuales redes sociales usa con más frecuencia para promocionar sus productos.**

Alternativas	Respuesta	Porcentaje
Facebook	11	100%
Instagram	2	36.4%
Twitter	4	18.2%
WhatsApp	5	45.5%
LinkedIn	0	0%
Otra	0	0%

Tabla 6.3: Pregunta 3. Cuales redes sociales usa con más frecuencia para promocionar sus productos.

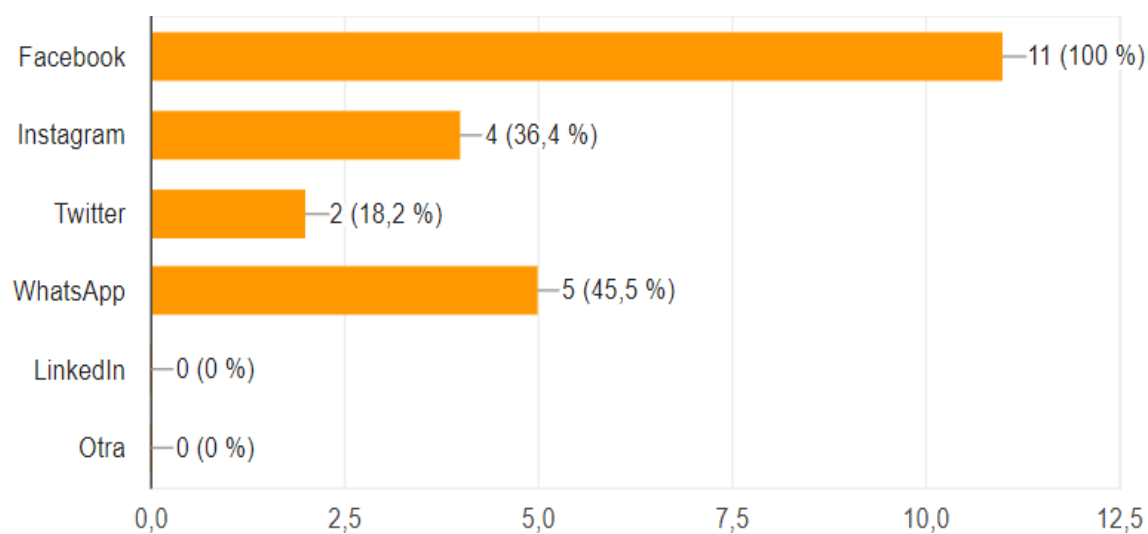


Figura 6.26: Representación porcentual para la pregunta 3.

Interpretación:

Del 100% de usuarios encuestados todos usan la red social de Facebook para promocionar sus productos, el 45.5% utilizan WhatsApp, el 36.4% usa Instagram, el 18.2% usa Twitter, mientras que ninguno de los encuestados utiliza LinkedIn u Otra red social.

• **Pregunta 4. Con que frecuencia publica nuevos productos en las redes sociales.**

Alternativas	Respuesta	Porcentaje
1 vez a la semana	4	36.4%
2 veces a la semana	5	45.5%
3 veces a la semana	2	18.2%
4 o mas veces a la semana	0	0%

Tabla 6.4: Pregunta 4. Con que frecuencia publica nuevos productos en las redes sociales.

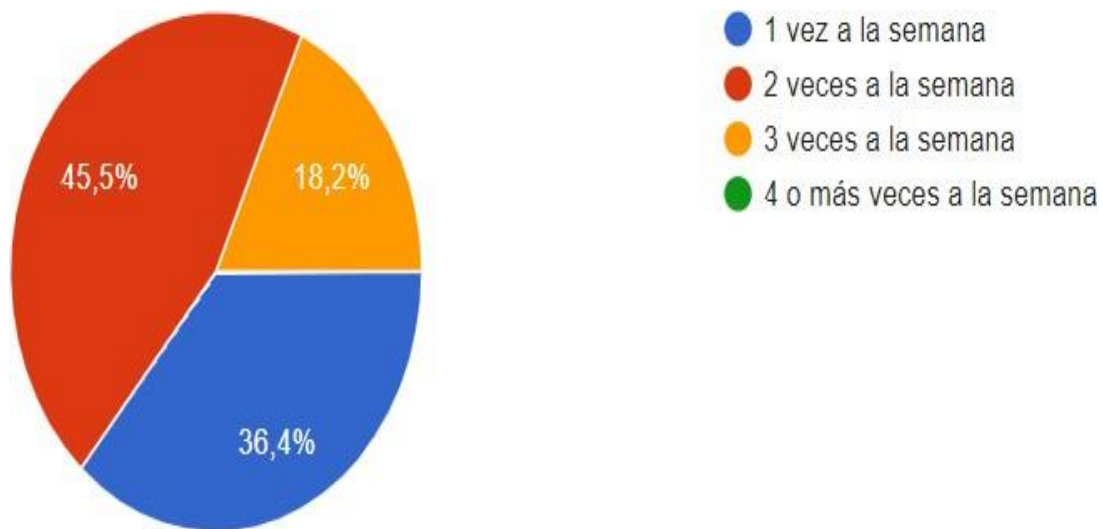


Figura 6.27: Representación porcentual para la pregunta 4.

Interpretación:

Del 100% de usuarios encuestados el 45.5% publica productos en sus redes sociales una vez a la semana, el 36.4% publica 2 veces a la semana, el 18.2% publica 3 veces a la semana y ninguno de los encuestados publica mas de 4 veces a la semana.

• **Pregunta 5. ¿Qué le pareció la facilidad de uso de esta herramienta?**

Alternativas	Respuesta	Porcentaje
Fácil e intuitiva de entender	6	54.5%
Fácil de utilizar	5	45.5%
Dificulta localizar los elementos	0	0%

Tabla 6.5: Pregunta 5. ¿Qué le pareció la facilidad de uso de esta herramienta?

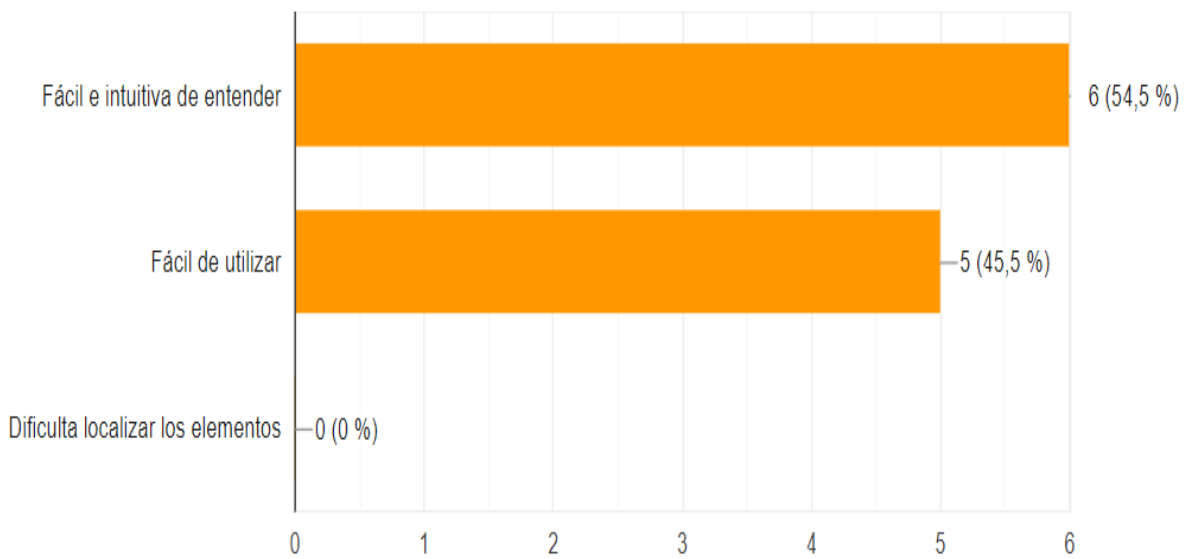


Figura 6.28: Representación porcentual para la pregunta 5.

Interpretación:

Del 100% de usuarios encuestados el 54.5% les pareció una herramienta fácil e intuitiva de entender, el 45.5% piensa que es una herramienta fácil de utilizar mientras que ningún usuario manifiesta que sea difícil de localizar los elementos.

- **Pregunta 6. ¿Qué tan conforme esta con la experiencia al registrar y analizar productos en la herramienta?**

Alternativas	Respuesta	Porcentaje
Muy Fácil	10	90.9%
Fácil	1	9.1%
Difícil	0	0%

Tabla 6.6: Pregunta 6. ¿Qué tan conforme esta con la experiencia al registrar y analizar productos en la herramienta?

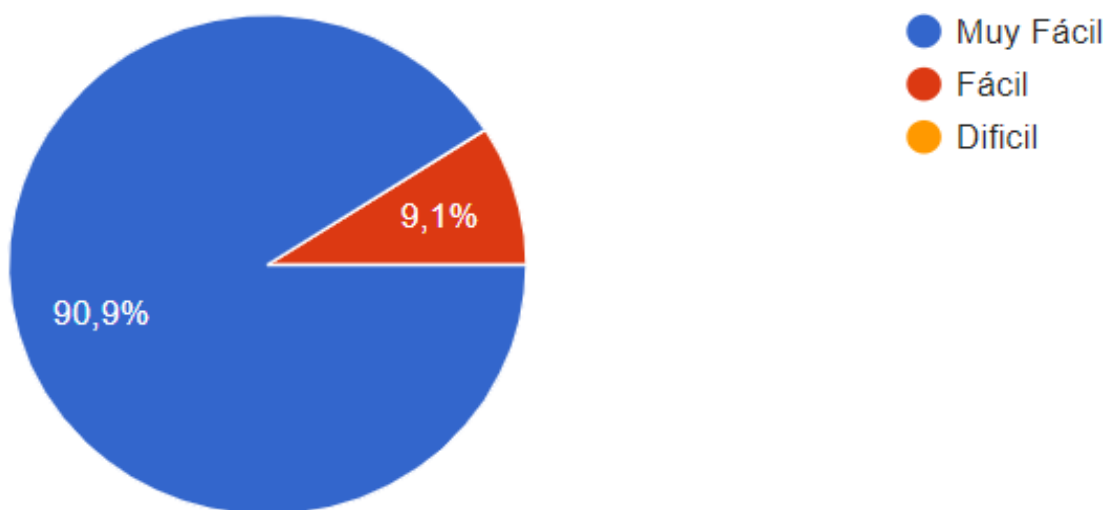


Figura 6.29: Representación porcentual para la pregunta 6.

Interpretación:

Del 100% de usuarios encuestados el 90.9% manifiesta que es Muy Fácil el registrar y analizar productos, el 9.1% piensa que es Fácil mientras que el 0% que es difícil.

• **Pregunta 7. ¿Qué tan satisfecho(a) esta con la seguridad de la aplicación?**

Alternativas	Respuesta	Porcentaje
Excelente	11	100%
Regular	0	0%
Mala	0	0%

Tabla 6.7: Pregunta 7. ¿Qué tan satisfecho(a) esta con la seguridad de la aplicación?

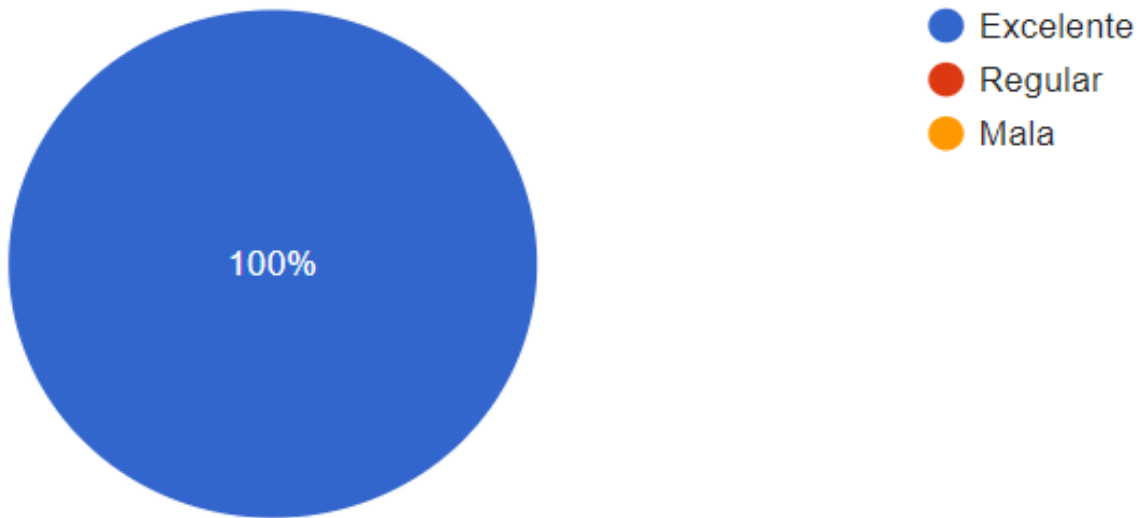


Figura 6.30: Representación porcentual para la pregunta 7.

Interpretación:

Del 100% de los usuarios encuestados todos expresaron que la seguridad manejada dentro de la aplicación es excelente, mientras que el 0% piensa que es regular o mala.

- **Pregunta 8. ¿Es comprensible la manera en la que se presenta los resultados de los análisis en la herramienta?**

Alternativas	Respuesta	Porcentaje
Absolutamente comprensible	6	54.5%
Fácil de comprender	5	45.5%
Absolutamente difícil de comprender	0	0%

Tabla 6.8: Pregunta 8. ¿Es comprensible la manera en la que se presenta los resultados de los análisis en la herramienta?

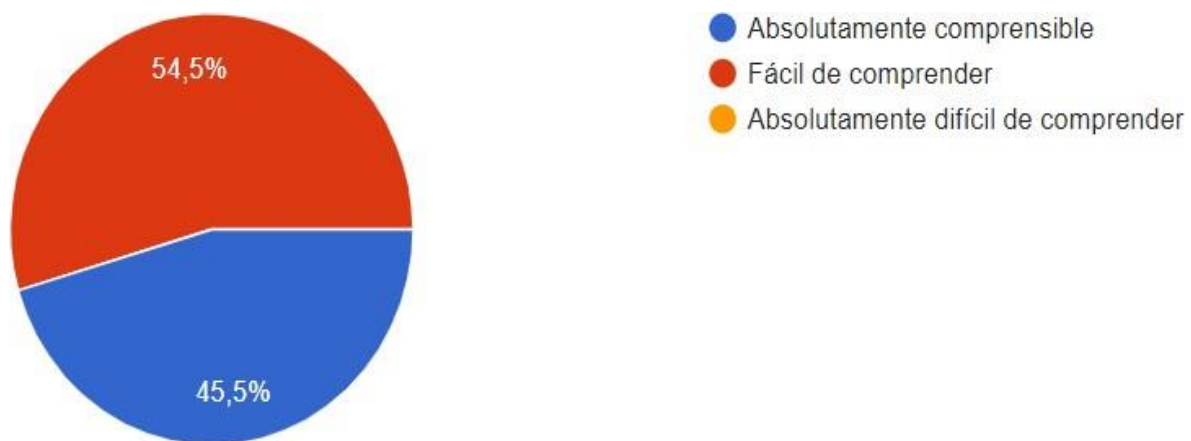


Figura 6.31: Representación porcentual para la pregunta 8.

Interpretación:

Del 100% de usuarios encuestados el 54.5% manifiestan que son absolutamente comprensible los resultados del análisis, el 45.5% dice que es fácil de comprender y el 0% que es absolutamente difícil de comprender.

• **Pregunta 9. ¿Qué tan conforme esta con la confiabilidad de esta herramienta?**

Alternativas	Respuesta	Porcentaje
Muy confiable	11	100%
Poco confiable	0	0%
No es confiable	0	0%

Tabla 6.9: Pregunta 9. ¿Qué tan conforme esta con la confiabilidad de esta herramienta?

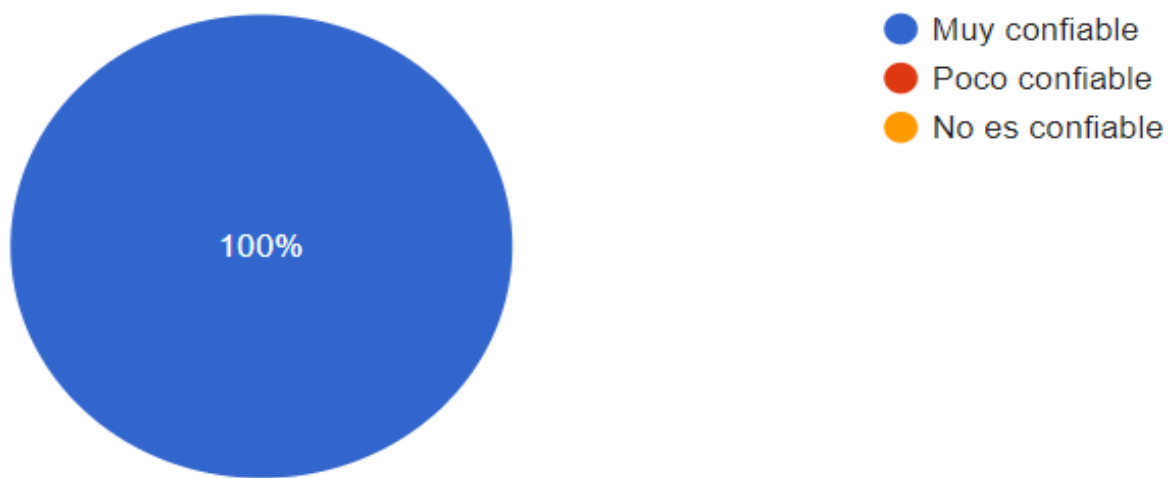


Figura 6.32: Representación porcentual para la pregunta 9.

Interpretación:

Del 100% de usuarios encuestados el 100% se muestran confiables con el uso de la herramienta, mientras que el 0% expresa que sea poco confiable o nada confiable.

- **Pregunta 10. Está de acuerdo que la herramienta cumple con el objetivo de ayudar a las empresas a detectar sentimientos de los usuarios y los tópicos que más se hablan acerca de un producto específico.**

Alternativas	Respuesta	Porcentaje
Totalmente de acuerdo	6	54.5%
De acuerdo	5	45.5%
Regular	0	0%
Muy en desacuerdo	0	0%

Tabla 6.10: Pregunta 10. Está de acuerdo que la herramienta cumple con el objetivo de ayudara las empresas a detectar sentimientos de los usuarios y los tópicos que más se hablan acerca deun producto específico

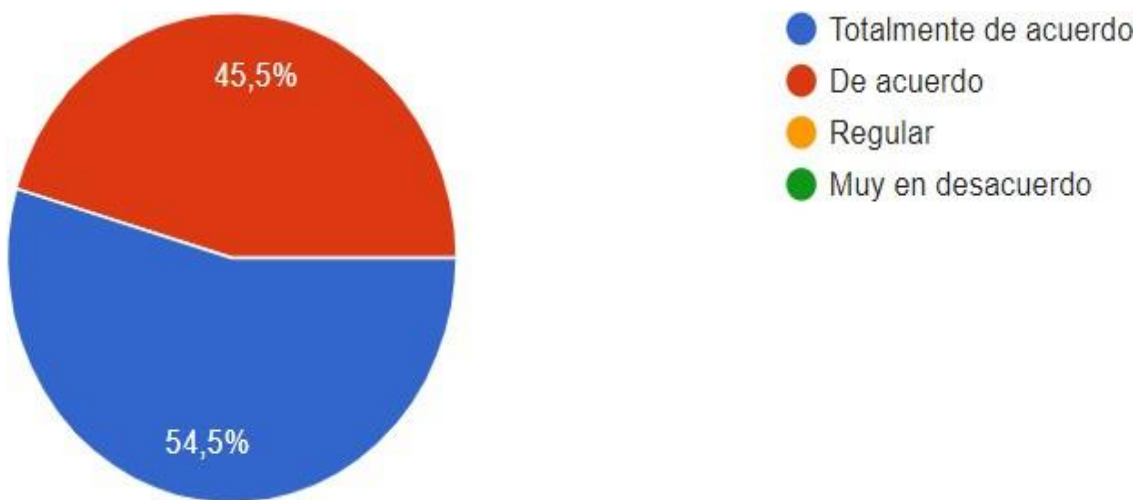


Figura 6.33: Representación porcentual para la pregunta 10.

Interpretación:

Del 100% de usuarios encuestados el 54.5% está de acuerdo en que la herramienta si cumple con los objetivos propuestos de ayudar a las empresas con una herramienta para detectar los sentimientos y tópicos acerca de un producto específico, el 45.5% está de acuerdo en que se cumplen con los objetivos, el 0% esta piensa que regularmente se cumplen con los objetivos, finalmente el 0% está muy en desacuerdo en que se cumplen los objetivos propuestos.

- **Pregunta 11. La herramienta capturo la mayor parte de los comentarios de la publicación.**

Alternativas	Respuesta	Porcentaje
Si	10	90.9%
No estoy seguro	1	9.1%
No	0	0%

Tabla 6.11: Pregunta 11. La herramienta capturo la mayor parte de los comentarios de la publicación.

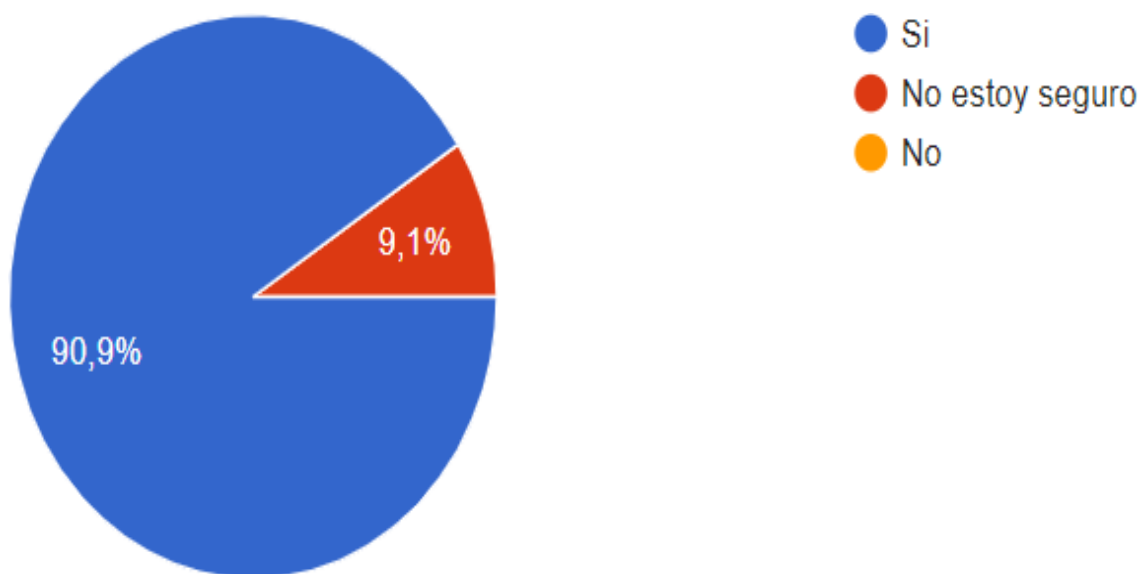


Figura 6.34: Representación porcentual para la pregunta 11.

Interpretación:

Del 100% de usuarios encuestados el 90.9% expreso que la herramienta si capturo la mayor parte de los comentarios de la publicación, el 9.1% no está seguro de que haya capturado todos los comentarios, y el 0% expreso que no capturo la mayor parte de comentarios.

- **Pregunta 12. Usted cree que la relación de las palabras dentro de un tópico tiene concordancia.**

Alternativas	Respuesta	Porcentaje
Si	10	90.9%
Un poco	1	9.1%
No	0	0%

Tabla 6.12: Pregunta 12. Usted cree que la relación de las palabras dentro de un tópico tiene concordancia.

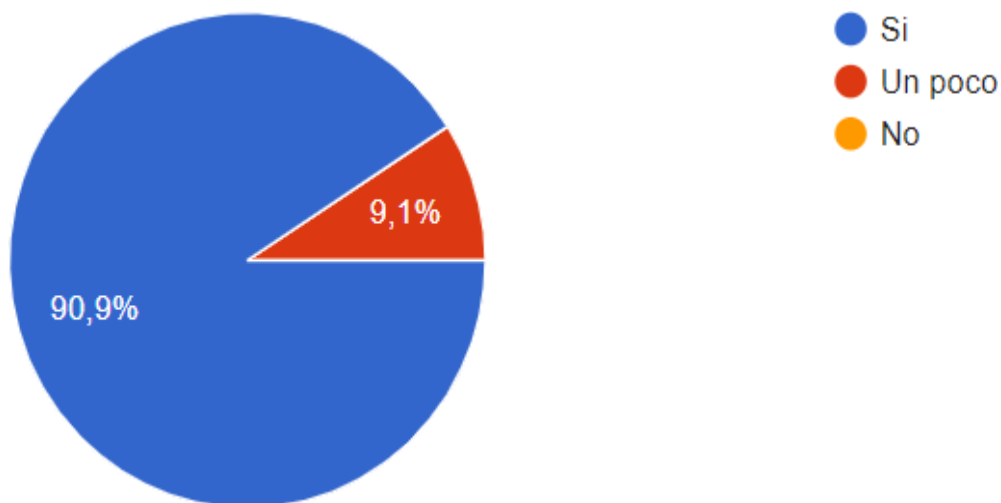


Figura 6.35: Representación porcentual para la pregunta 12.

Interpretación:

Del 100% de usuarios encuestados el 90.9% indican que, si tienen relación las palabras dentro de los tópicos analizados, el 9.1% dice que un poco tienen relación estas palabras, y el 0% expreso que no tienen relación.

- **Pregunta 13. Su empresa tomaría en cuenta los resultados arrojados por la herramienta para mejorar sus estrategias empresariales.**

Alternativas	Respuesta	Porcentaje
Si	8	72.7%
Tal vez	3	27.3%
No	0	0%

Tabla 6.13: Pregunta 13. Su empresa tomaría en cuenta los resultados arrojados por la herramienta para mejorar sus estrategias empresariales.

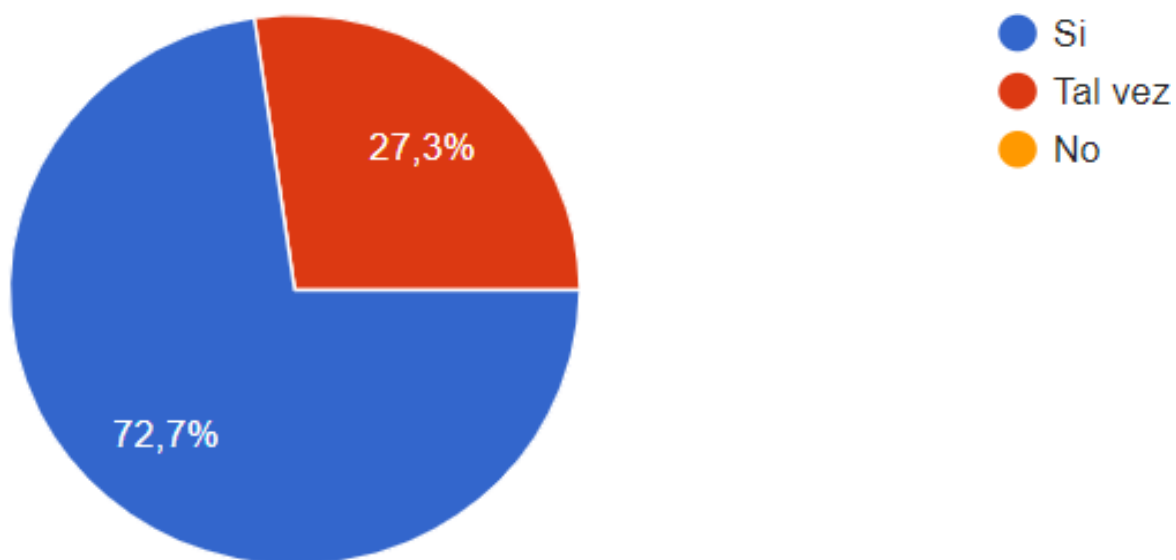


Figura 6.36: Representación porcentual para la pregunta 13.

Interpretación:

Del 100% de usuarios encuestados el 72.7% tomaría en cuenta los resultados arrojados por la herramienta para mejorar sus estrategias empresariales, el 27.3% tal vez tomaría en cuenta estos resultados y el 0% no los tomaría en cuenta.

- **Pregunta 14. Utilizaría nuestra herramienta de manera cotidiana para realizar un análisis de la acogida de sus productos dentro de las redes sociales.**

Alternativas	Respuesta	Porcentaje
Si	7	63.6%
Tal vez	4	36.4%
No	0	0%

Tabla 6.14: Pregunta 14. Utilizaría nuestra herramienta de manera cotidiana para realizar un análisis de la acogida de sus productos dentro de las redes sociales.

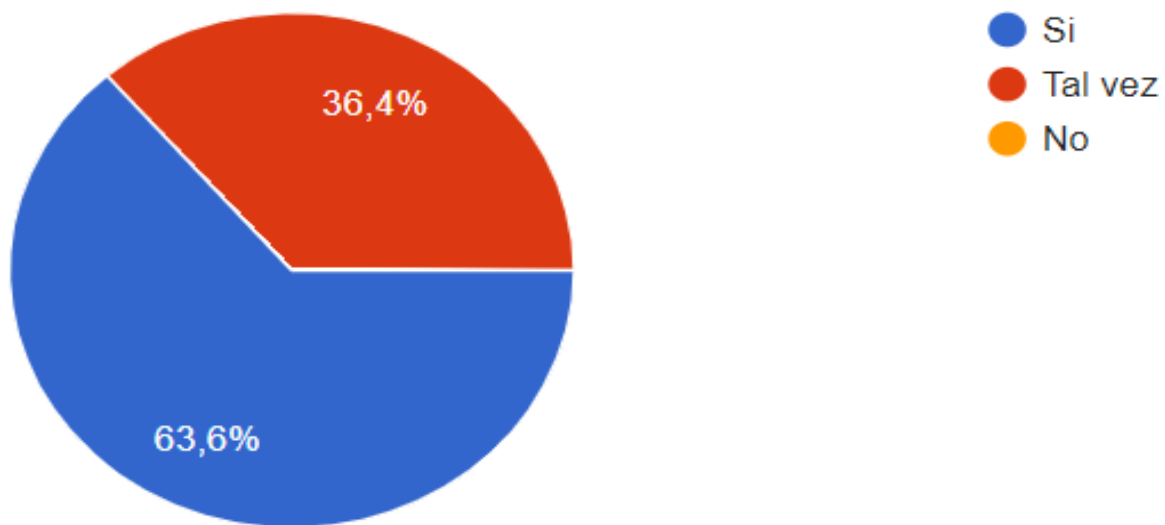


Figura 6.37: Representación porcentual para la pregunta 14.

Interpretación:

Del 100% de usuarios encuestados el 63.6% utilizarían cotidianamente esta herramienta para poder mejorar sus estrategias empresariales, el 36.4% tal vez lo utilizaría de manera cotidiana y el 0% no lo utilizaría.

- **Pregunta 15. Si su respuesta es No en la pregunta anterior, podría indicar una razón para no utilizar esta herramienta.**

En la pregunta anterior ningún usuario respondió que No por ende no existe alguna respuesta textual por parte de los encuestados.

- **Pregunta 16. ¿Qué tan probable es que usted recomiende esta herramienta a un(a) amigo(a) o algún miembro de la familia?**

Alternativas	Respuesta	Porcentaje
Muy probable	10	90.9%
Poco probable	1	9.1%
Nada probable	0	0%

Tabla 6.15: Pregunta 16. ¿Qué tan probable es que usted recomiende esta herramienta a un(a) amigo(a) o algún miembro de la familia?

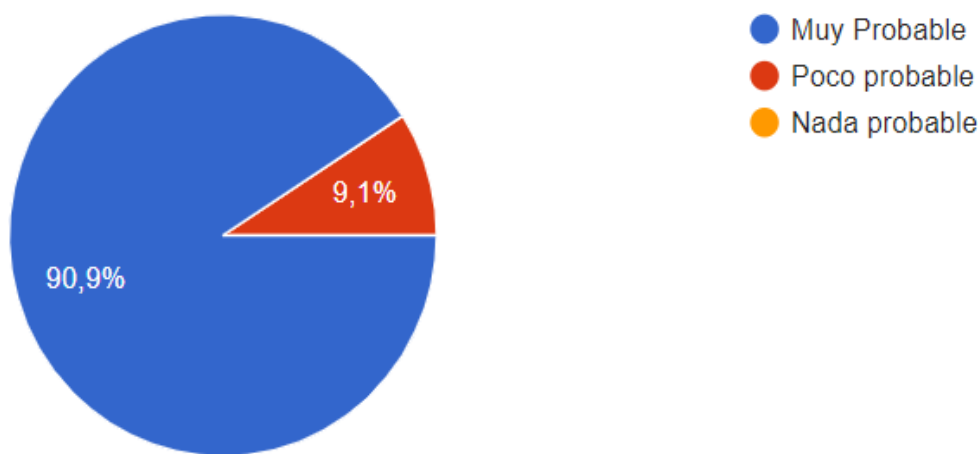


Figura 6.38: Representación porcentual para la pregunta 16.

Interpretación:

Del 100% de encuestados el 90.9% es muy probable que recomienden esta herramienta a algún amigo o miembro de la familia, el 9.1% es poco probable que lo recomiende y el 0% es nada probable de recomendar la herramienta.

• **Pregunta 17. ¿Tiene comentarios o sugerencias que ayuden a mejorar nuestro servicio? (Opcional)**

Al ser una pregunta opcional no todos los encuestados respondieron esta pregunta, aunque si hubo algunos de ellos que la respondieron y expusieron las siguientes sugerencias:

- Mostrar los emojis en los comentarios.

- Generación de reportes con gráficos en Excel.
- Agregar más detalles al reporte generado.
- Permita mayor personalización al usuario.
- Poder realizar análisis de otras redes sociales.
- Implementar una aplicación móvil.

Capítulo VII

CRONOGRAMA DE ACTIVIDADES

Nombre de la tarea	Duración	Comienzo	Fin
Proyecto	300 horas	Lun 26/4/21	Sab 24/7/21
OE. 1	14 horas	Lun 26/4/21	Jue 29/4/21
ACT 1. Estudio de los enfoques del procesamiento del lenguaje natural.	4 horas	Lun 26/4/21	Lun 26/4/21
ACT 2. Estudio de las técnicas de clasificación de tópicos.	5 horas	Mar 27/4/21	Mie 28/4/21
ACT 3. Estudio de análisis de sentimientos y RNN.	5 horas	Mie 28/4/21	Jue 29/4/21
OE. 2	70 horas	Vie 30/4/21	Sab 22/5/21
ACT 1. Construcción del dataset en relación con un producto específico.	8 horas	Vie 30/4/21	Dom 2/5/21
ACT 2. Preparación de los datos esto limpieza incluye eliminación de caracteres raros, stop words, letras repetidas.	10 horas	Lun 3/5/21	Mie 5/5/21
ACT 3. Procesamiento del dataset, tokenización, vectorización, etc.	8 horas	Jue 6/5/21	Vie 7/5/21
ACT 4. Preparación de los datos para el análisis de LDA.	6 horas	Lun 10/5/21	Mar 11/5/21
ACT 5. Construcción y entrenamiento del modelo LDA.	20 horas	Mie 12/5/21	Lun 17/5/21
ACT 6. Análisis e interpretación de los resultados del modelo LDA.	8 horas	Mar 18/5/21	Mie 19/5/21
ACT 7. Revisión / Corrección.	10	Jue 20/5/21	Sab 22/5/21
OE. 3	50 horas	Lun 24/5/21	Sab 5/6/21
ACT 1. Preparación de los datos obtenidos de la clusterización con LDA.	4 horas	Lun 24/5/21	Lun 24/5/21
ACT 2. Diseño y construcción del modelo RNN(LSTM).	20 horas	Mar 25/5/21	Sab 29/5/21
ACT 3. Entrenamiento y optimización.	8 horas	Lun 31/5/21	Mar 1/6/21

ACT 4. Análisis y entrenamiento con nuevos comentarios.	8 horas	Mie 2/6/21	Jue 3/6/21
ACT 5. Revisión / Corrección.	10	Vie 4/6/21	Sab 5/6/21
OE. 4	144 horas	Lun 7/6/21	Sab 17/7/21
ACT 1. Analizar los requerimientos para el diseño y la construcción de aplicación web.	4 horas	Lun 7/6/21	Lun 7/6/21
ACT 2. Diseño y desarrollo de la interfaz web.	30 horas	Mar 8/6/21	Mar 15/6/21
ACT 3. Diseño y desarrollo de la fase de datos. Creación servidor para base de datos para los usuarios y análisis realizados.	18 horas	Mie 16/5/21	Sab 18/6/21
ACT 4. Construcción de API Rest para la comunicación desde la fase de presentación con la fase de gestión.	12 horas	Lun 21/6/21	Mie 23/6/21
ACT 5. Diseño, desarrollo e implementación de reportes y graficas de resultado.	30 horas	Jue 24/6/21	Vie 2/7/21
ACT 6. Despliegue de la aplicación dentro de los servidores de GIHP4C.	20 horas	Sab 3/6/21	Jue 8/7/21
ACT 7. Ingreso, funcionalidad, validaciones, pruebas de usuario.	20 horas	Vie 9/7/21	Mie 14/7/21
ACT 8. Revisión / Corrección.	10 horas	Jue 15/7/21	Sab 17/7/21
OE. 5	22 horas	Lun 19/7/21	Sab 24/7/21
ACT 1. Diseño de un plan de pruebas unitarias y de integración.	8 horas	Lun 19/7/21	Mar 20/7/21
ACT 2. Evaluar diferentes aspectos de los datos procesados; coherencia, apego a la realidad.	8 horas	Mie 21/7/21	Jueves 22/7/21

ACT 3. Analizar y verificar los resultados obtenidos.	6 horas	Vie 23/7/21	Sab 24/7/21
---	---------	-------------	-------------

Capítulo VIII

PRESUPUESTO

DENOMINACION	CANTIDAD	COSTO UNITARIO	COSTO TOTAL
	unidades	dólares	dólares
1. Bienes			
Copias	50	0,02	1
Impresiones	150	0,05	7,5
Empastados	1	20	20
2. Tecnológico			
Computador, Portátil	1	1200	1200
3. Servicios			
Cursos, Libros, etc.	5	50	250
Servicios de Internet	3	40	120
Material Bibliográfico	15	10	150
4. Recursos Humanos			
Desarrollador	1	1	1800
5. Otros			
Imprevistos	1	250	250
Total	526	\$ 1.575,07	\$ 3.798,50

Capítulo IX

CONCLUSIONES

La aplicación desarrollada tuvo como objetivo identificar cual es la perspectiva de los consumidores sobre un producto específico publicado en la red social de una empresa registrada en la aplicación a través de la minería de datos y PLN.

Seguidamente, se a través de aprendizaje automático se realizó el análisis de detección de tópicos y análisis de sentimientos, permitiéndonos procesar una gran cantidad de datos y poder de esta forma detectar nuevos tópicos o temas mencionados por los usuarios juntamente con el sentimiento que expreso el usuario mediante su comentario. Todo esto mediante la creación de algoritmos inteligentes.

La herramienta al trabajar con especialistas en el tema puede mejorar en la precisión de los resultados ya que se cuenta con una coherencia media del 65% de los tópicos detectados en los diferentes análisis realizados juntamente con un 85% de precisión en la detección de sentimientos.

Todo este conocimiento brindado, permite al usuario (empresa) conocer cuáles fueron sus puntos fuertes y los problemas de los servicios o productos que brinda. Permitiendo mejorar la atención a los clientes y sobre todo que puedan generar nuevas estrategias que permitan un crecimiento continuo.

Cabe mencionar que existe trabajo a futuro para esta herramienta y tiene que ver con agregar

la funcionalidad para realizar el análisis de publicaciones de diferentes redes sociales como lo es Twitter o Instagram. Mejorar también el control de las publicaciones realizadas por una cuenta, permitiendo agregar el perfil de la red social y obtener las publicaciones más recientes realizadas por dicha cuenta. Desarrollar también una aplicación móvil la cual cuente con las mismas funcionalidades que la aplicación web permitiendo una mayor portabilidad y usabilidad de la aplicación.

Capítulo X

RECOMENDACIONES

Para mantener la integridad de su cuenta personal es de suma importancia tener cuidado con las credenciales y no compartirlas con terceros evitando así tanto la divulgación de su información privada como la de sus análisis realizados.

Es muy importante contar con una conexión a internet estable ya que la aplicación web necesita comunicarse con servicios web y necesita también cargar imágenes, de esta manera evitar largas esperas de carga de la información y mantener una comunicación estable con el servidor donde se encuentra hospedada la aplicación web.

Tanto el modelo PLN que trabaja con los comentarios de las publicaciones como el modelo LDA deben ser desarrollados en base a los objetivos que se desean cumplir, pues varía mucho los resultados al trabajar de una manera general.

REFERENCIAS

- Asanza, W. R. and Olivo, B. M. (2018). Redes neuronales artificiales aplicadas al reconocimiento de patrones. *Editorial UTMACH*.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research*, 3:993–1022.
- Bobadilla, J., Bojorque, R., Hernando Esteban, A., and Hurtado, R. (2018). Recommender systems clustering using bayesian non negative matrix factorization. *IEEE Access*, 6:3549–3564.
- Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1):1–8.
- Celaya, J. (2011). La empresa en la web 2.0. page 15.
- El Naqa, I. and Murphy, M. J. (2015). What Is Machine Learning? In El Naqa, I., Li, R., and Murphy, M. J., editors, *Machine Learning in Radiation Oncology*, pages 3–11. Springer International Publishing, Cham.
- Flórez, R. and Fernández, J. M. (2008). Las redes neuronales artificiales, fundamentos teóricos y aplicaciones prácticas. *La Coruña, Oleiros, España: Netbiblo*.
- Friedman, J. H. (1998). Data mining and statistics: What’s the connection? *Computing science and statistics*, 29(1):3–9.

- Gamon, M., Aue, A., Corston-Oliver, S., and Ringger, E. (2005). Pulse: Mining customer opinions from free text. In *international symposium on intelligent data analysis*, pages 121–132. Springer.
- Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.
- Hutt Herrera, H. (2012). LAS REDES SOCIALES: UNA NUEVA HERRAMIENTA DE DIFUSIÓN. *Revista Reflexiones*, 91(2).
- Joshi, P. (2017). *Artificial intelligence with Python: build real-world artificial intelligence applications with Python to intelligently interact with the world around you*. Packt, Birmingham Mumbai.
- Kaschesky, M., Sobkowicz, P., and Bouchard, G. (2011). Opinion mining in social media: modeling, simulating, and visualizing political opinion formation in the web. In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, pages 317–326.
- Kim, P. (2006). The forrester wave: Brand monitoring, q3 2006. *Forrester Wave (white paper)*.
- Kucuktunc, O., Cambazoglu, B. B., Weber, I., and Ferhatosmanoglu, H. (2012). A large-scale sentiment analysis for yahoo! answers. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining, WSDM '12*, page 633–642, New York, NY, USA. Association for Computing Machinery.
- Liddy, E. D. (2001). Natural language processing.
- Muñiz Cuza, C. (2018). Deep learning en el análisis de sentimientos.
- Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons*. b, 4:51–62.
- Olabe, X. B. (1998). Redes neuronales artificiales y sus aplicaciones. *Publicaciones de la Escuela de Ingenieros*.

Pirri, S., Lorenzoni, V., Andreozzi, G., Mosca, M., and Turchetti, G. (2020). Topic modeling and user network analysis on twitter during world lupus awareness day. *International Journal of Environmental Research and Public Health*, 17(15).

Rivera, E. et al. (2015). Introducción a las redes neuronales artificiales.

Shum, Y. M. (2021). Situación digital, Internet y redes sociales Ecuador 2020.

Zhang, L., Wang, S., and Liu, B. (2018). Deep learning for sentiment analysis:A survey. *WIREs Data Mining and Knowledge Discovery*, 8(4):e1253. _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/widm.1253>.

Anexos

Anexos A

Formato de la encuesta realizada para la validación de la aplicación web

Encuesta Sistema Inteligente de Análisis de Sentimientos y tópicos

Esta encuesta tiene como objetivo determinar la percepción de los usuarios acerca del uso de la aplicación web que tiene como finalidad ayudar a determinar la acogida de un producto específico publicado en una red social mediante la clusterización de tópicos y sentimientos.

Información de la empresa

1. Ciudad donde se encuentra su empresa.

2. ¿Qué tipo de productos vende su empresa?

Q Computación

Q Hogar y muebles

Q Autos, motos u otros

Q Celulares y telefonía

- Q Deportes y Fitness
- Q Electrónica, audio y video
- Q Otro

3. Cuales redes sociales usa con más frecuencia para promocionar sus productos

- Q Facebook
- Q Instagram
- Q Twitter
- Q WhatsApp
- Q LinkedIn
- Q Otra

4. Con que frecuencia publica nuevos productos en las redes sociales

- Q 1 vez a la semana
- Q 2 veces a la semana
- Q 3 veces a la semana
- Q 4 o más veces a la semana

Aplicación web

5. ¿Qué le pareció la facilidad de uso de esta herramienta?

- Q Fácil e intuitiva de entender
- Q Fácil de utilizar
- Q Dificulta localizar los elementos

6. ¿Qué tan conforme esta con la experiencia al registrar y analizar productos en la herramienta?

Q Muy Fácil

Q Fácil

Q Difícil

7. ¿Qué tan satisfecho(a) esta con la seguridad de la aplicación?

Q Excelente

Q Regular

Q Mala

8. ¿Es comprensible la manera en la que se presenta los resultados de los análisis en la herramienta?

Q Absolutamente comprensible

Q Fácil de comprender

Q Absolutamente difícil de comprender

9. ¿Qué tan conforme esta con la confiabilidad de esta herramienta?

Q Muy confiable

Q Poco confiable

Q No es confiable

10. Está de acuerdo que la herramienta cumple con el objetivo de ayudar alas empresas a detectar sentimientos de los usuarios y los tópicos que más se hablan acerca de un producto específico.

Q Totalmente de acuerdo

Q De acuerdo

Q Regular

Q Muy en desacuerdo

11. La herramienta capturo la mayor parte de los comentarios de la publicación.

Q Si

Q No estoy seguro

Q No

12. Usted cree que la relación de las palabras dentro de un tópico tiene concordancia.

Q Si

Q Un poco

Q No

13. Su empresa tomaría en cuenta los resultados arrojados por la herramienta para mejorar sus estrategias empresariales.

Q Si

Q Tal vez

Q No

14. Utilizaría nuestra herramienta de manera cotidiana para realizar un análisis de la acogida de sus productos dentro de las redes sociales.

Q Si

Q Tal vez

Q No

15. Si su respuesta es No en la pregunta anterior, podría indicar la razón por la cual no utilizaría esta herramienta.

16. ¿Qué tan probable es que usted recomienda esta herramienta a un(a) amigo(a) o algún miembro de la familia?

Q Muy probable

Q Poco probable

Q Nada probable

17. ¿Tiene comentarios o sugerencias que ayuden a mejorar nuestro servicio? (Opcional)

Anexos B

Interfaces de Usuario

2.1 Aplicación Web

2.1.1 Página de Inicio

En la figura se puede observar la página de inicio cuando usted ingresa a <https://cloudcomputing.ups.edu.ec/siast/> en donde se brinda información acerca de cuál es el objetivo del sistema.

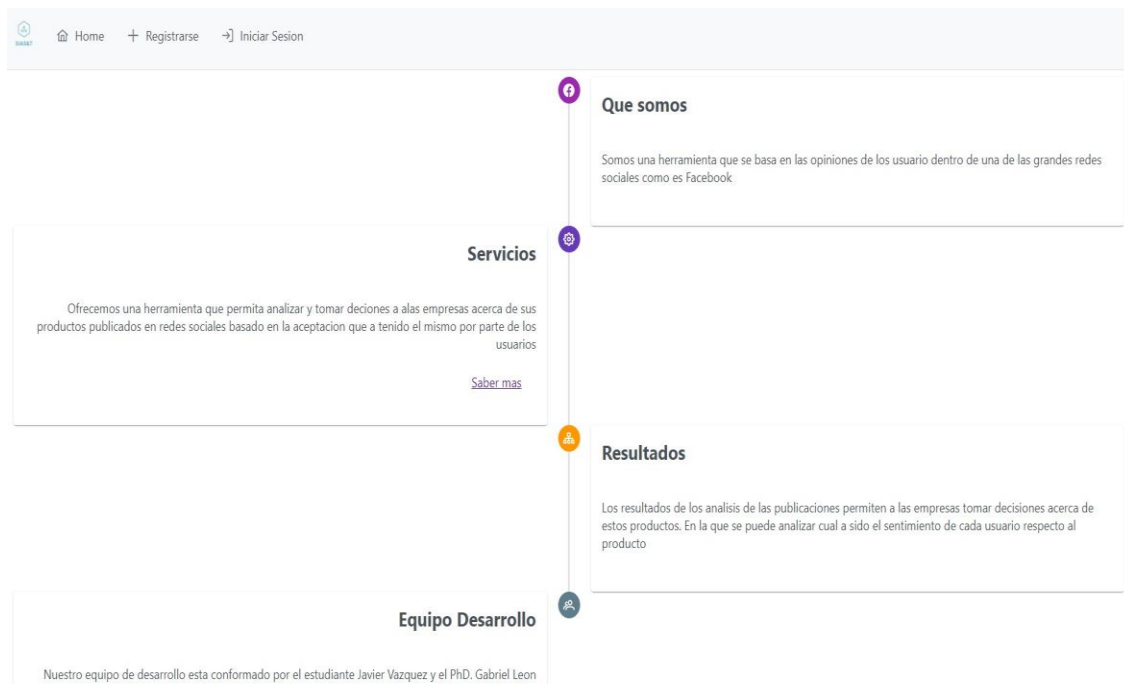


Figura 2.1: Página de inicio.

2.1.2 Inicio de sesión

Como se observa en la figura 2.2 SIAST cuenta con un formulario para poder iniciar sesión, la cual sirve para el ingreso a las funcionalidades del sistema web.

The image shows a web application login interface. At the top, there is a dark blue header with a white logo consisting of a hexagon containing three interlocking circles, with the text 'SIAS&T' below it. Below the header, the word 'Login' is written in a bold, black font. There are two input fields stacked vertically. The first field has a person icon on the left and the text 'Correo' inside. The second field has a lock icon on the left, the text 'Contraseña' inside, and a small eye icon on the right to toggle password visibility. Below the input fields, there are two buttons: a purple button with the text 'Registrarse' and a blue button with the text 'Login'.

Figura 2.2: Inicio de sesión de la aplicación web.

2.1.3 Registro de Usuarios

En la figura 2.3 podemos observar el formulario para los usuarios nuevos, donde tendrán que proporcionar la información requerida por el sistema.

The image shows a user registration form for SIAS&T. At the top, there is a dark blue header with a white hexagonal logo containing a network diagram and the text "SIAS&T" below it. The main content area is white and titled "Registro" in bold, with the subtitle "Registro usuarios" below it. The form consists of four input fields, each with a grey icon on the left: "Nombre" (person icon), "Dirreccion" (house icon), "Correo" (envelope icon), and "Contraseña" (lock icon). The password field has a toggle eye icon on the right. Below the fields is a blue rounded button labeled "Registrarse". At the bottom, there is a purple underlined link that says "Ya tengo cuenta".

Figura 2.3: Formulario Registro Usuario.

2.1.4 Página principal

En la figura 2.4 se podrá observar cuando un usuario nuevo inicia sesión, el sistema desplegara mensaje de bienvenida y posteriormente indicando cual es el objetivo principal del sistema.

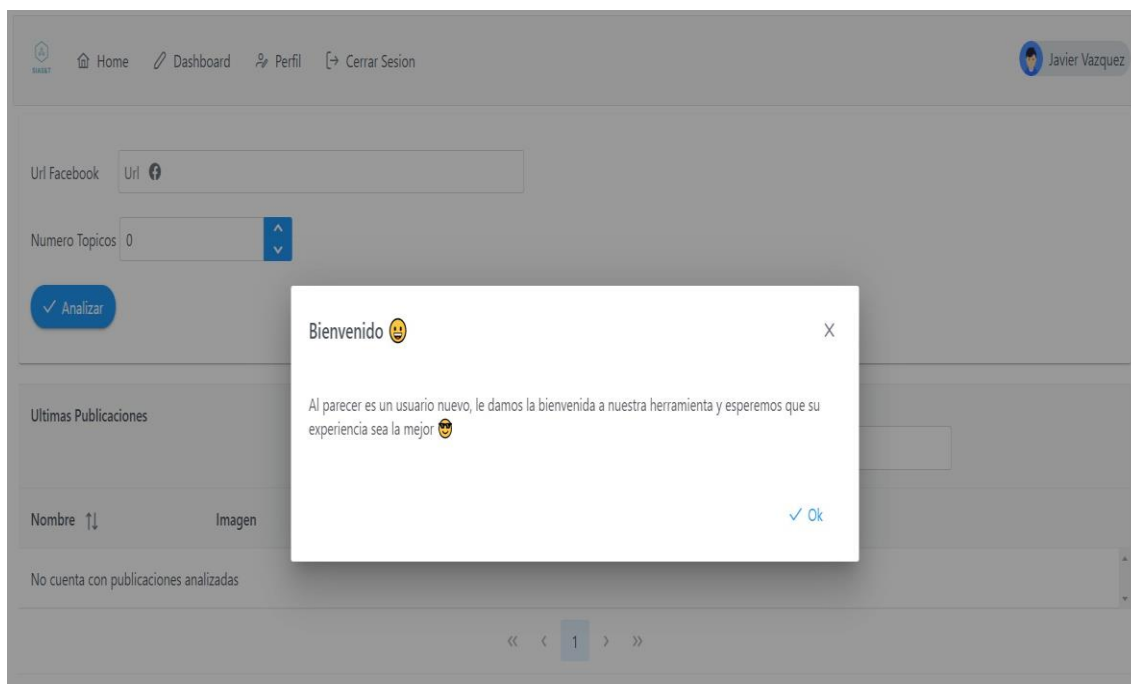


Figura 2.4: Primer inicio de sesión en la aplicación web.

En nuestra página principal, en el primer apartado se encuentra la sección donde podremos ingresar los datos necesarios para realizar un nuevo análisis, en este caso sería la URL de la publicación de Facebook y la cantidad de tópicos que quisiéramos analizar, posteriormente el sistema pediría una confirmación como se muestra en la figura 2.5.

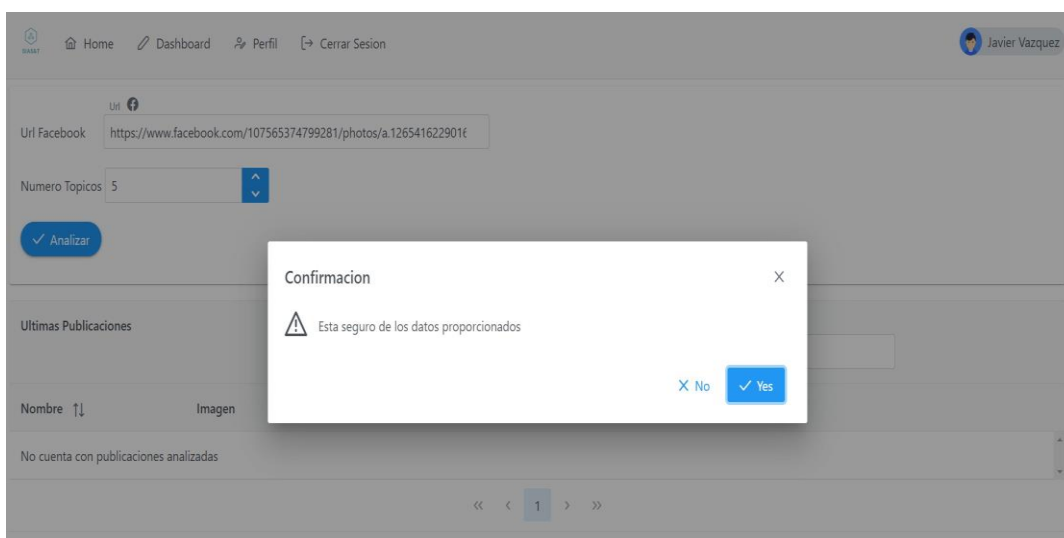


Figura 2.5: Confirmación análisis nueva publicación.

Una vez que ya tengamos realizado análisis dentro del sistema, en nuestra página principal se presentara este listado de publicaciones analizadas resaltando información principal, como el nombre de la publicación, la imagen que incluye está dentro de la red social, la cantidad de comentarios y reacciones que ha tenido la publicación y también la calidad lo cual representa el nivel de aceptación que a tenido por parte de los usuarios como se muestra en la figura 2.6.

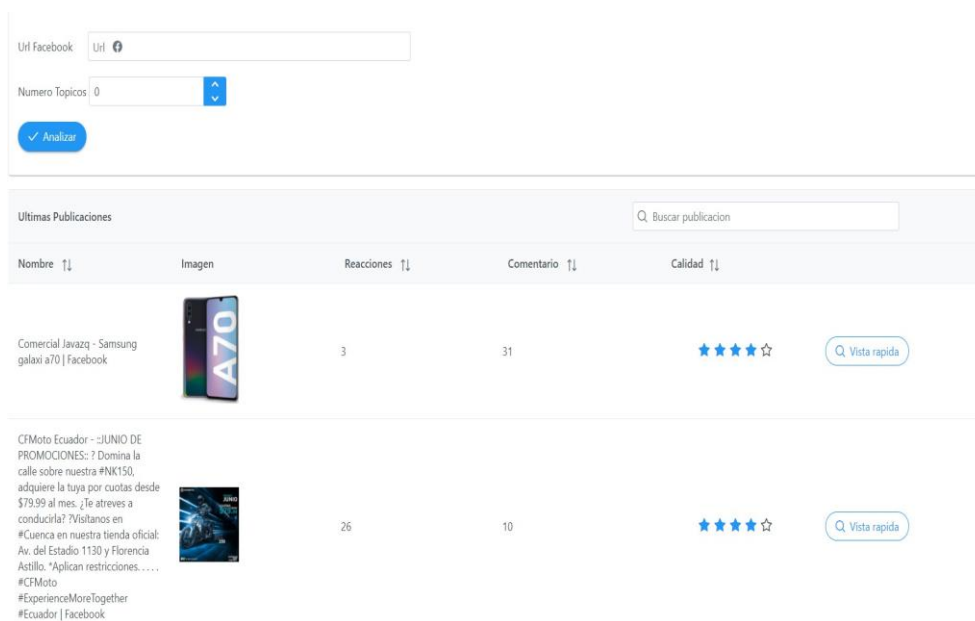


Figura 2.6: Lista de publicaciones analizadas.

2.1.5 Perfil de Usuario

Como se muestra en la figura 2.7 en el menú de navegación una vez que inicia sesión se muestra el apartado de Perfil el cual permite ingresar a modificar nuestros datos personales.

Home Dashboard Perfil Cerrar Sesión Javier Vazquez

Perfil Usuario
Datos Personales

Nombre
Javier Vazquez

Direccion
casique duma y calle del retorno

Correo
dato@hotmail.com

Contraseña

Guardar Cancelar

Figura 2.7: Modificar datos de usuario.

2.1.6 Resultados de análisis

Una vez que realicemos un análisis, el sistema enviara a una nueva página donde por secciones se presentaran los resultados. Como muestra la figura 2.8 la primera sección expone cuales han sido los comentarios de que ha tenido la publicación juntamente con los sentimientos y tópicos detectados por cada uno de los comentarios.

Comentario	Sentimiento	Topico
bastante precisa, pero eso depende de lo que esté haciendo.		
Recomiendo mucho este producto. ¡Es fabuloso!	Positivo	4
Buen celular con una muy buena cámara	Positivo	4
Algunas aplicaciones que se instalan no funcionan a la perfección	Positivo	4
toma fotografías excelentes	Negativo	1
el teléfono es de mala calidad pero buen precio	Negativo	4
Los juegos que se descarguen funcionan lento no sirve	Positivo	4
A veces se cierran solá las aplicaciones y se trava mucho.	Positivo	4
Celular malo. iPhone es mejor!	Negativo	4
Resistente a las caídas	Positivo	4
El Ram del celular es muy bueno	Positivo	1

Figura 2.8: Análisis de sentimientos y tópicos de los comentarios.

En la figura 2.9 se puede analizar cuál ha sido la acogida del producto por parte de los usuarios, mediante un gráfico que indica el porcentaje de los comentarios positivos como negativos.

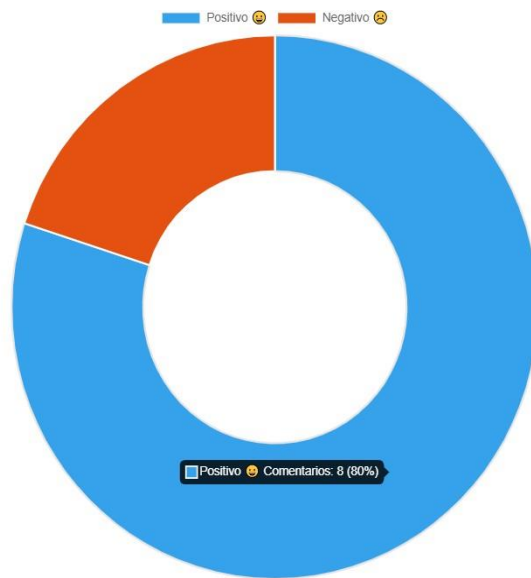


Figura 2.9: Porcentaje acogida de usuarios.

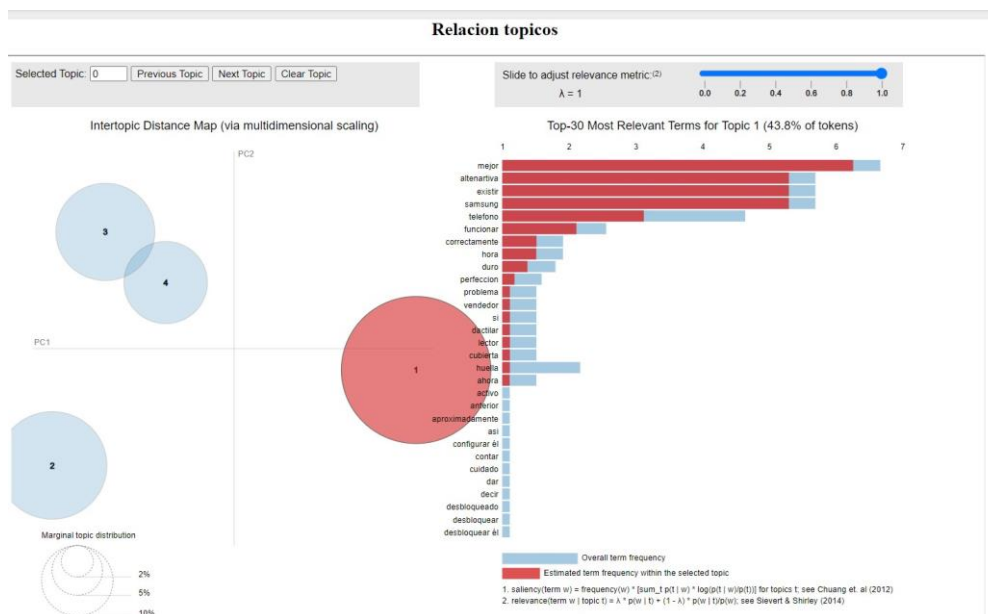
En la siguiente sección mediante una imagen se exhiben cuáles han sido las palabras claves detectadas por cada uno de los tópicos. Indicando por mayor tamaño a la palabra de mayor

relevancia de dicho t3pico como se observa en la figura 2.10.



Figura 2.10: Distribuci3n de las palabras por t3picos.

Como se observa en la figura 2.11 se muestra un gr3fico interactivo que muestra la distribuci3n de las palabras del modelo LDA generado, donde un buen modelo de t3picos es aquel que cuenta con burbujas bastante grandes y que no se solapan. En el cual si movemos el cursor sobre cada una de las burbujas las palabras y las barras del grafico se actualizan



autom3ticamente.

Figura 2.11: Relaci3n entre los t3picos.

En la figura 2.12 se muestra un gr3fico interactivo que con informaci3n relevante a la

distribución de cada comentario que pertenece a dicho tópico.

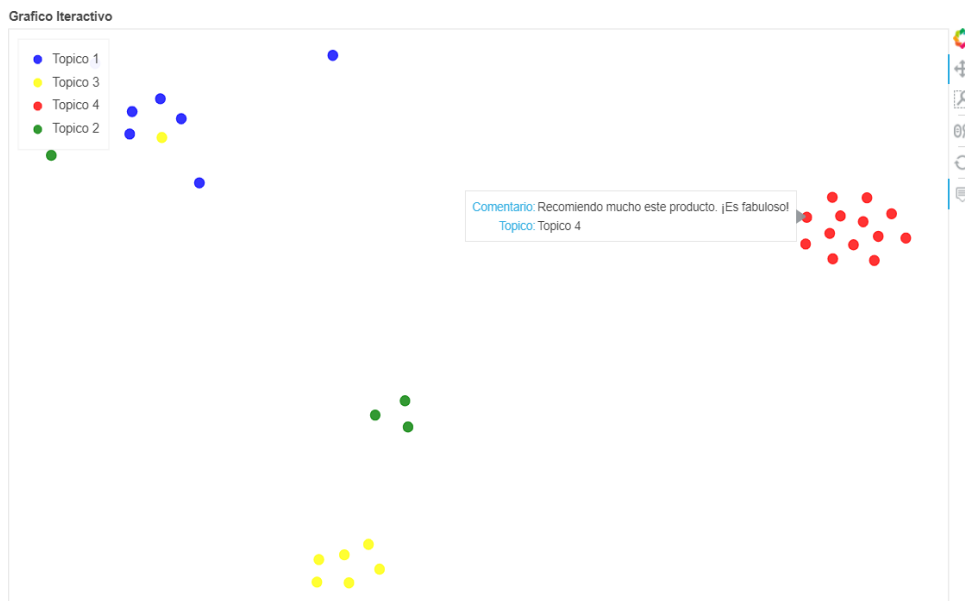


Figura 2.12: Distribución de los comentarios por grupos de tópicos.

Finalmente, en la figura 2.13 podemos observar tanto la distribución de la cantidad de los comentarios por cada uno de los tópicos como la cantidad de reacciones que ha tenido la publicación.

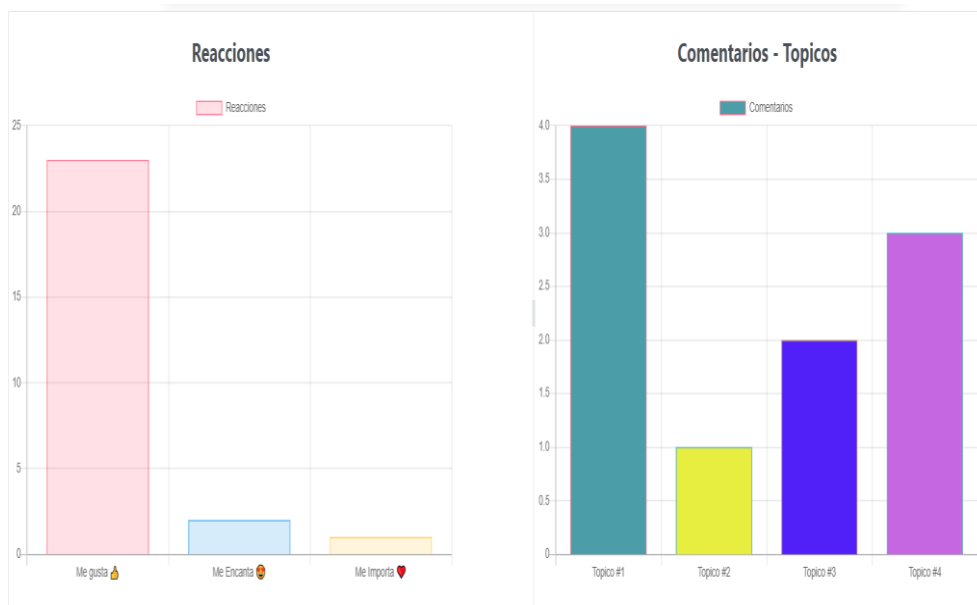


Figura 2.13: Distribución cantidad de reacciones y comentarios por tópicos.

2.1.7 Equipo de Desarrollo

Cuenta con una página en donde se encuentra información de los miembros del equipo de desarrollo, tal y como se muestra en la figura 2.14

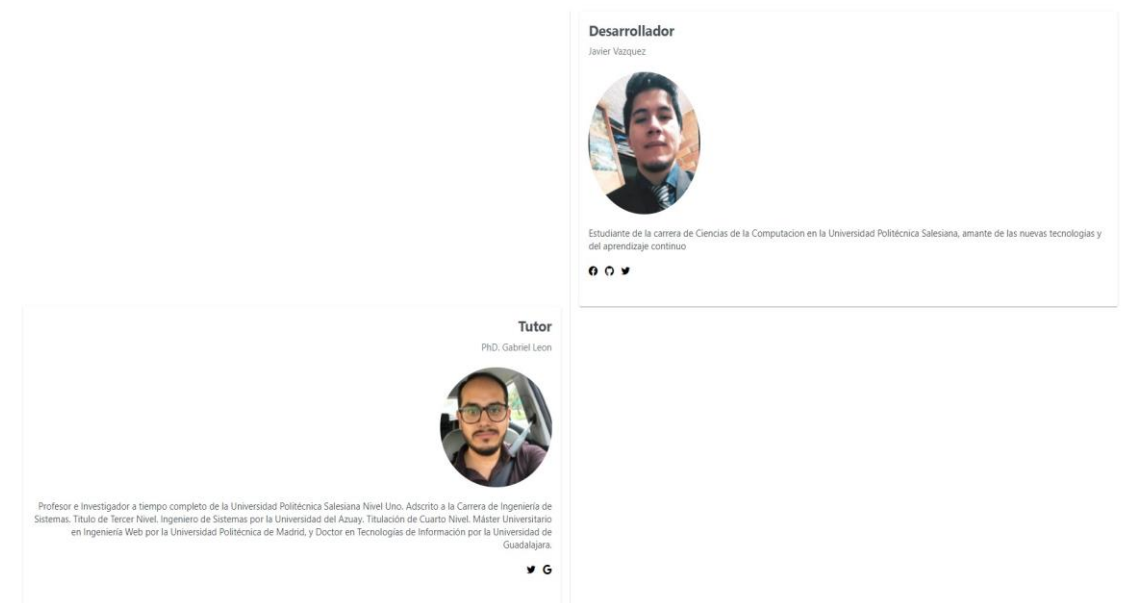


Figura 2.14: Equipo de Desarrollo.