

**Arquitectura de Big Data para el análisis de sentimientos multimodales en el sector  
turístico del Departamento de Boyacá, Colombia.**

**Cristian David Algecira Arbelaez**  
**Para optar el título de Ingeniero de Sistemas**

**Director**  
**Julián Alberto Monsalve Pulido**

**Universidad Santo Tomás Sede Tunja**  
**Facultad de Ingeniería Sistemas**  
**2020**

## RESUMEN

El presente proyecto propone el desarrollo de una arquitectura de Big data que permita evaluar la satisfacción de los usuarios del sector turístico del departamento de Boyacá Colombia mediante el análisis de sentimientos multimodales, a través de una serie de parámetros en la extracción de datos a herramientas como Booking, TripAdvisor y twitter, facilitando el análisis de los sentimiento del usuario por medio de variables de polaridad, esto con tal de conocer la opinión de los usuarios frente al sector turístico del departamento de Boyacá. El proyecto atiende a tres puntos fundamentales dentro del desarrollo de la arquitectura: extracción de datos, análisis de datos y visualización.

Dichos puntos establecen un punto de partida dentro del desarrollo de la arquitectura; puesto que los subpuntos de los tres parámetros iniciales buscan el buen funcionamiento y desarrollo de la arquitectura. La extracción comprende el uso de la herramienta webs scraper encargada de la extracción de datos de las herramientas ya mencionadas para posterior uso en el segundo parámetro, el análisis; en este punto se pretende que los datos obtenidos de la extracción se sometan al proceso de polaridad que permita determinar las emociones y opiniones de los usuarios frente a sus visitas; esto para finalmente en el proceso de visualización los datos obtenidos como resultados puedan ser plasmados en gráficas que permiten su entendimiento y comprensión. Finalmente, el proyecto presenta una serie de conclusiones frente a los resultados obtenidos con la arquitectura. La arquitectura está basada en el proyecto de investigación Doctoral de la Pontificia Universidad Javeriana “Modelo multimodal para el análisis de sentimientos en un dominio turístico”.

**Palabras clave:** Análisis de sentimientos, Big Data, Extracción, Turismo, web scraper.

## ABSTRACT

This project proposes the development of a Big data architecture that allows evaluating the satisfaction of users of the tourism sector in the department of Boyacá Colombia through the analysis of multimodal sentiments, through a series of parameters in the extraction of data to tools such as Booking, TripAdvisor and twitter, facilitating the analysis of user sentiment through polarity variables, this in order to know the opinion of users regarding the tourism

sector of the department of Boyacá. The project addresses three fundamental points within the development of the architecture: data extraction, data analysis and visualization.

These points establish a starting point within the development of architecture; since the sub-points of the three initial parameters seek the proper functioning and development of the architecture. The extraction includes the use of the web scraper tool in charge of extracting data from the aforementioned tools for later use in the second parameter, the analysis; At this point, it is intended that the data obtained from the extraction be subjected to the polarity process that allows determining the emotions and opinions of the users regarding their visits; this for finally in the visualization process the data obtained as results can be captured in graphs that allow their understanding and comprehension. Finally, the project presents a series of conclusions regarding the results obtained with architecture based on its objective.

## **LISTADO DE SIGLAS**

**PIB:** Producto Interno Bruto.

**DANE:** Departamento Administrativo Nacional de Estadísticas.

**IA:** Inteligencia Artificial.

**BD:** Big Data.

**CSV:** Comma Separated Values (Valores Separados por Comas).

**GPS:** Global Position Satelital (Posición Global Satelital).

**NLP:** Natural Language Processing (Procesamiento de Lenguaje Natural.)

## **AGRADECIMIENTOS**

Primeramente, agradezco a mi familia por apoyarme en cada decisión y proyecto de mi vida, gracias a la Universidad Santo Tomás Sede Tunja por haberme aceptado ser parte de ella y abierto las puertas de su seno científico para poder estudiar mi carrera, así como también a los diferentes docentes que brindaron sus conocimientos y su apoyo para seguir adelante día a día.

Agradezco también a mi Decano de facultad de Ingeniería de Sistemas el Ingeniero Carlos Andrés Guerrero por acompañarme en mi proceso de realización como profesional a lo largo de mi carrera universitaria.

Agradezco al Ingeniero Julián Alberto Monsalve Pulido por haberme brindado la oportunidad de recurrir a su capacidad y conocimiento profesional, así como también haberme tenido paciencia para guiarme durante todo el desarrollo de mi trabajo de grado.

Agradezco a mis compañeros de clase durante todos los semestres de Universidad ya que gracias al compañerismo, amistad y apoyo moral han aportado a mi desarrollo profesional como ingeniero de sistemas.

No ha sido sencillo el camino hasta ahora, pero gracias a sus aportes, a su amor y a su inmensa bondad, la complejidad de esta meta se ha notado, pero ha servido como enseñanza para mi vida profesional.

## TABLA DE ILUSTRACIONES

<b>Figura 1</b>	<b>Lista de nociones del turista.</b>	<b>16</b>
<b>Figura 2</b>	<b>Niveles de PLN.</b>	<b>17</b>
<b>Figura 3</b>	<b>Análisis de sentimientos por imágenes.</b>	<b>19</b>
<b>Figura 4</b>	<b>Reconocimiento facial.</b>	<b>20</b>
<b>Figura 5</b>	<b>Problemas empresariales de Big data.</b>	<b>22</b>
<b>Figura 6</b>	<b>Tipos de Big data.</b>	<b>23</b>
<b>Figura 7</b>	<b>Clasificación de Big data.</b>	<b>24</b>
<b>Figura 8</b>	<b>Versión preliminar de la arquitectura.</b>	<b>26</b>
<b>Figura 9</b>	<b>Arquitectura Global.</b>	<b>28</b>
<b>Figura 10</b>	<b>Fase de extracción de datos.</b>	<b>30</b>
<b>Figura 11</b>	<b>Fase de análisis de datos.</b>	<b>31</b>
<b>Figura 12</b>	<b>Semáforo de emociones.</b>	<b>32</b>
<b>Figura 13</b>	<b>Fase de visualización de datos.</b>	<b>33</b>
<b>Figura 14</b>	<b>Funcionamiento del análisis de sentimientos.</b>	<b>37</b>
<b>Figura 15</b>	<b>Mapa de Sentimientos.</b>	<b>38</b>
<b>Figura 16</b>	<b>Facial Landmarks de Dlib</b>	<b>40</b>
<b>Figura 17</b>	<b>Facial Landmarks de Dlib en imagen con rostro</b>	<b>40</b>
<b>Figura 18</b>	<b>Reconocimiento de ojos y rostro por medio de región</b>	<b>41</b>
<b>Figura 19</b>	<b>Gráfico de área</b>	<b>42</b>
<b>Figura 20</b>	<b>Gráfico de barras</b>	<b>43</b>
<b>Figura 21</b>	<b>Gráfico de matriz</b>	<b>43</b>
<b>Figura 22</b>	<b>Gráfico de red</b>	<b>44</b>
<b>Figura 23</b>	<b>Gráfico circular</b>	<b>45</b>
<b>Figura 24</b>	<b>Cartograma</b>	<b>45</b>
<b>Figura 25</b>	<b>Histograma</b>	<b>46</b>
<b>Figura 26</b>	<b>Nube de palabras</b>	<b>46</b>
<b>Figura 27</b>	<b>Problemática y estadísticas de Boyacá</b>	<b>49</b>
<b>Figura 28</b>	<b>Modelamiento del sistema</b>	<b>50</b>
<b>Figura 29</b>	<b>Servicios presentados en el servidor, almacenamiento y aplicativo web</b>	<b>50</b>
<b>Figura 30</b>	<b>Categoría de polaridades y nube de palabras</b>	<b>51</b>
<b>Figura 31</b>	<b>Planificador de Viaje de fechas con mayor tráfico.</b>	<b>54</b>
<b>Figura 32</b>	<b>Planificador de Viajes en diagrama de barras de los sitios más visitado por la plataforma de Booking.</b>	<b>55</b>
<b>Figura 33</b>	<b>Planificador de Viajes en diagrama de barras de los sitios más visitado por la plataforma de TripAdvisor.</b>	<b>56</b>
<b>Figura 34</b>	<b>Colección de Booking en mongoDB.</b>	<b>57</b>
<b>Figura 35</b>	<b>Colección de TripAdvisor en mongoDB.</b>	<b>57</b>
<b>Figura 36</b>	<b>Resultado del reconocimiento de rostro y ojos.</b>	<b>60</b>
<b>Figura 37</b>	<b>Reconocimiento de imagen para demostrar la diferente polaridad.</b>	<b>61</b>
<b>Figura 38</b>	<b>Colección de Booking en la base de datos de mongoDB para la utilización en el sitio web.</b>	<b>64</b>

<b>Figura 39 Colección de TripAdvisor en la base de datos de mongoDB para la utilización en el sitio web.</b>	65
<b>Figura 40 Inicio del aplicativo web.</b>	65
<b>Figura 41 Inicio del aplicativo web donde se ingresa a los datos.</b>	66
<b>Figura 42 Inicio del aplicativo web donde se mostrarán los datos extraídos.</b>	66
<b>Figura 43 Muestra la ventana de “Booking” con las estadísticas que se dan por municipio.</b>	67
<b>Figura 44 Muestra la ventana de “Booking” con las estadísticas que se dan por municipio.</b>	67
<b>Figura 45 Muestra la ventana de “Booking” con las estadísticas de un municipio en específico y con las nubes de palabras.</b>	68
<b>Figura 46 Muestra la ventana de “Booking” con las estadísticas que se dan por municipio.</b>	68
<b>Figura 47 Muestra la ventana de “Tripadvisor” con las estadísticas que se dan por municipio.</b>	69
<b>Figura 48 Muestra la ventana de “Tripadvisor” con las estadísticas de un municipio en específico y con las nubes de palabras.</b>	69
<b>Figura 49 Muestra la ventana de “Twitter texto” con las estadísticas de Boyacá y con las nubes de palabras.</b>	70
<b>Figura 50 Muestra la ventana de “Twitter texto” con las estadísticas de Boyacá.</b>	70
<b>Figura 51 Muestra a los participantes del sector empresarial turístico del departamento de Boyacá.</b>	71
<b>Figura 52 Muestra a los participantes del sector académico y científico de la Universidad Santo Tomas sede Tunja.</b>	71

# CONTENIDO

<b>1.INTRODUCCIÓN</b>	8
<b>Descripción del Problema</b>	8
<b>2. OBJETIVOS</b>	11
<b>2.1 Objetivo General.</b>	11
<b>2.2 Objetivos específicos.</b>	11
<b>3. JUSTIFICACIÓN</b>	12
<b>4.MARCO REFERENCIAL</b>	14
<b>4.1 Turismo Experiencial</b>	14
<b>4.2 Análisis de sentimientos</b>	16
<b>4.2.1 Análisis de sentimientos en texto.</b>	18
<b>4.2.2 Análisis de sentimientos en imágenes.</b>	18
<b>4.3 Arquitectura de Big Data</b>	20
<b>5.ARQUITECTURA PROPUESTA</b>	28
<b>5.1 Extracción:</b>	33
<b>5.2 Análisis:</b>	35
<b>5.2.1 Librerías Utilizadas.</b>	38
<b>5.3 Visualización:</b>	41
<b>6.1 Pruebas</b>	49
<b>6.1.1 Fase 1 Estado del arte.</b>	49
<b>6.1.2 Fase 2 Modelamiento del sistema.</b>	49
<b>6.1.3 Fase 3 Implementación.</b>	50
<b>6.1.4 Fase 3 Pruebas.</b>	51
<b>6.2 Resultados</b>	51
<b>7.CONCLUSIONES</b>	71
<b>8.BIBLIOGRAFÍA</b>	73

## 1.INTRODUCCIÓN

### **Descripción del Problema**

Si bien el turismo se presenta como uno de los grandes motores de la economía a nivel mundial. A raíz de la globalización y el cambio de perspectivas sociales y culturales el turismo se convirtió en una de las actividades de mayor interés para la población mundial. Para el caso colombiano, el turismo, en perspectiva de la última década, presentó un aporte al PIB entre el 3,5% y 3,8%, por lo tanto el aporte al PIB del turismo se ha mantenido en una senda de crecimiento entre el 3,5% y 3,8% (La Vanguardia, 2020), de los cuales Boyacá representa el 2.7% a nivel general en el PIB nacional, de acuerdo con el DANE (Departamento Administrativo Nacional de Estadística), seguido de esto según las Cuentas Regionales del 2010 del Departamento Administrativo de Planeación Dirección de Planeación Territorial, el sector turístico tuvo una participación de 223.319 millones de pesos (Boyacá Visión 2019, 2019), evidenciando que el sector del turismo en Boyacá es uno de los sector que tiene gran contribución a la economía del departamento, de esta manera se encuentra la importancia de la implementación de un programa que permita desarrollar un análisis frente a la potencialidad del sector en Boyacá a través del uso de herramientas como el Big data; entendida esta última como el análisis masivo de datos, que a su vez se refiere a las nuevas tecnologías que hacen posible el almacenamiento y procesamiento, además de al uso que se hace de la información obtenida a través de dichas tecnologías (Facultad de Estudios Estadísticos, Universidad Complutense de Madrid, 2020).

Ahora en más, el uso de esta herramienta resulta útil para las empresas ya que proporciona respuestas frente a situaciones que previamente no se formularán. De esta manera, los datos pueden ser usados para que las empresas sean capaces de identificar los problemas de manera más comprensible. La recopilación de los datos y la búsqueda de tendencias dentro de ellos, permiten que estas reaccionen rápidamente ante los problemas (Universidad de Alcalá, 2020).

En relación con el turismo la arquitectura de Big data resulta productiva puesto que ofrece diversos beneficios para todo tipo de compañías del sector turístico, tales como identificar nuevos servicios y productos que se puedan ofrecer a los clientes (Mérida, 2014). De igual modo el uso de esta herramienta busca mejorar las relaciones existentes con el usuario. El procesamiento de todos los datos recopilados es más rápido por lo que de este modo se obtienen respuestas inmediatas frente a las necesidades presentes en el mercado.

Otro punto importante es que el desarrollo de Big Data en la industria del turismo actualmente está siendo impulsada por un grupo importante de empresas. Estas empresas están usando Big Data de diversas maneras, como, por ejemplo, optimizar la administración de las ganancias, personalizar la distribución de los viajes, perfeccionando operaciones internas y mejorando el performance financiero de las empresas, establecen (Mérida, 2014), en su Modelo arquitectónico de información para una plataforma de Big Data para el Sector Turístico.

En la región de Boyacá se concentra un gran número de turistas cada año, este número tiende a aumentar cada vez más, además, el turista promedio que visita Boyacá tiende a planificar sus viajes en los meses más congestionados de la región y fechas importantes como: fiestas, rutas navideñas, eventos, conciertos, etc. Esto propone un gran reto, ya que la prioridad de quienes ofrecen el servicio es que los turistas y visitantes tengan una mayor tranquilidad y satisfacción frente al destino que tienen pensado visitar. Por otro lado, el turista tiende a ver estadísticas por internet sobre el sitio que va a visitar y estos muchas veces son fuentes no confiables, conllevando a que se pierda una cantidad de turistas y de credibilidad.

De esta manera y como impulso hacia el sector se tiene como objetivo la creación de una arquitectura de Big data que permita evaluar la satisfacción de los usuarios del sector turístico en Boyacá mediante el análisis de sentimientos multimodales, brindando la información necesaria a los turistas por medio de estadísticas en forma de gráficos y de sus plataformas creíbles como Booking y TripAdvisor de tal modo que los resultados obtenidos a través de la herramienta puedan ser divulgados frente a la gobernación de Boyacá con tal de proporcionar información pertinente correspondiente a la satisfacción de los turistas en

relación a sus actividades y sitios de interés. Logrando así que el sector pueda ampliarse como foco de atención a nivel turístico.

## **2. OBJETIVOS**

### **2.1 Objetivo General.**

Desarrollar una arquitectura de Big data para evaluar la satisfacción de los usuarios del sector turístico en Boyacá mediante el análisis de sentimientos multimodales.

### **2.2 Objetivos específicos.**

1. Realizar un estado de arte de investigaciones en el área de arquitectura de Big Data en el análisis de sentimientos, para determinar características iniciales investigativas y técnicas para la construcción de la propuesta.
2. Adaptar técnicas y modelos de análisis de sentimientos multimodal que integren fases de extracción, filtrado, análisis, visualización y calidad de información en un dominio turístico.
3. Implementar una arquitectura de Big Data en el sector turístico del Departamento de Boyacá Colombia, para analizar sentimientos multimodales en redes sociales y en planificadores de viaje.
4. Divulgar los resultados de la investigación a la comunidad académica, científica y empresarial del sector turístico del Departamento de Boyacá.

### 3. JUSTIFICACIÓN

Hace varios años desde el sector gubernamental y privado se han realizado actividades encaminadas a impulsar el turismo en Boyacá, de esta manera, el gobierno departamental identificó los lugares sobresalientes en cada uno de los 123 municipios y esto permitió generar herramientas de configuración interna de las provincias para promover clúster turísticos que logren apalancar esta actividad económica con la generación de una encadenamiento productivo de los municipios, así se desarrollaron los 6 anillos turísticos que se trazaron sobre círculos de riqueza de los municipios que tenían un carácter diferenciador, involucrando los distintos productos ya identificados anteriormente.

No obstante, de acuerdo con la ley general del turismo (Ley 300/1996) se relacionan los “Círculos Metropolitanos Turísticos”, los cuales están conformados por municipios que contribuyan a la prestación de servicios turísticos de calidad y asociados con los demás municipios de Boyacá, estos círculos están conformados por: Paipa, Tibasosa, Nobsa, Duitama, Sogamoso, Monguí, Iza, tota, Aquitania, Tunja, Villa de Leyva, Ráquira, Sáchica, Chiquinquirá, Tópaga, Mongua y Gámeza. Para ayudar al desarrollo del turismo en la región la investigación se centrará en la construcción de una arquitectura de Big Data en el área de análisis de sentimientos, basados en el turismo experiencial, ya que es una nueva forma de turismo que se basa en las emociones y experiencias que logra vivir el turista a través de la interacción con el destino, se define como una vivencia personal, que sale de lo cotidiano, donde la persona combina aspectos de carácter tangible que se representa en los productos turísticos, y de carácter intangible como libertad, seguridad, tranquilidad, relajación.

Tal recopilación de datos busca ampliar y/ o estimular la categorización dentro de posibles destinos turísticos, el programa en desarrollo muestra el déficit o eficiencia del turismo en ciertos municipios y departamento en general. Creando una alerta (Notificación), la cual presentará los siguientes datos: el departamento que presentó mayor incremento en la actividad turística, como los municipios concurridos, por su parte, se adicionarán aquellos que mostraron bajas, los municipios menos concurridos buscando lugares que son necesarios implementar mayores estrategias en cuanto al marketing mix, buscando así la promoción de

los lugares menos concurridos con buen potencial turístico, así como, el mantenimiento y alcance de aquellas que ya se encuentran posicionadas en el área turística nata.

El desarrollo de la arquitectura busca optimizar la toma de decisiones de los actores (turista, hoteles, restaurantes, sector gubernamental, entre otros) del sector turístico del departamento de Boyacá, mediante una aplicación disponible en la web, para la consulta de sentimientos multimodales del sector turístico.

## **4.MARCO REFERENCIAL**

### **4.1 Turismo Experiencial**

El sector del Turismo Experiencial está en un proceso de transformación hacia unos productos y servicios centrados en la creación de experiencias y en la interacción con el turista de forma más directa, humana y personalizada. Al integrar aspectos como la cultura y las tradiciones de las gentes del lugar, el turista se siente inmerso en su forma de vida, y sus sentimientos hacia el destino serán positivos (Molina et al. 2007; Troitiño, 1991, 1991).

La experiencia es objeto esencial del viaje para una gran parte de los viajeros actuales, que sienten cada vez más la necesidad de experimentar un destino para optimizar su vivencia del mismo. Esta nueva tendencia, busca establecer una relación emocional entre el turista y el destino. Esta “economía de la experiencia” (Pine & Gilmore, 1998).

(RICHARDS, 2004) afirma que el turismo cultural ha pasado a convertirse en un producto de vivencia en el que la visita se juzga en función de todas las características del lugar, además de su valor cultural. En esta línea, los visitantes buscan cada vez más una experiencia total que incluya ocio, cultura, gastronomía e interacción social (De Rojas y Camarero, 2008; Rivera, 2013, 2013).

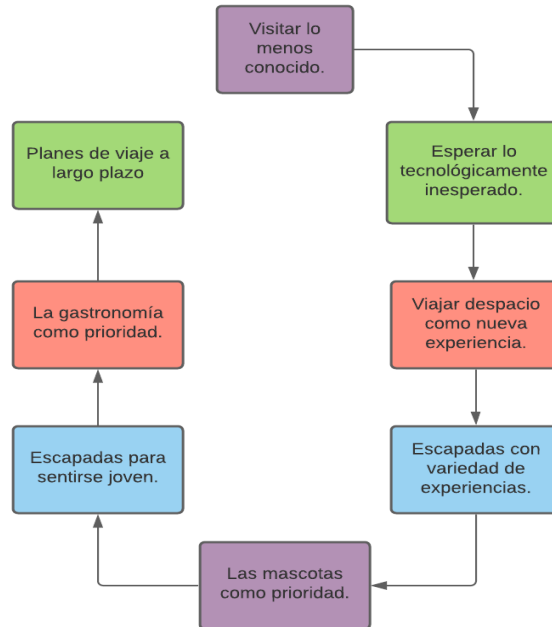
Así mismo el turista busca vincular el lugar de su visita con las vivencias esperadas, es decir la selección de un destino ligado a la expectativa de la elección, por ejemplo: el turismo gastronómico implica para sus “usuarios” un medio de conocer la cultura del lugar que visitan a través de su comida por tanto la experiencia vivida asegura satisfacer la necesidad del turista de conocer la cultura a través de un medio diferente que no sea netamente los lugares enigmáticos propios del lugar. Por otro lado, es importante destacar que la elección de un destino turístico en función a lo experiencial también va ligado a los lugares que afirman destacar y hacer reales dichas experiencias, el turismo oscuro afirma dicha situación puesto que asegura transmitir las emociones relacionadas estos lugares, ejemplo: aquellos sitios, países y ciudades. etc., que tienen rastros de guerra o desastres; por lo cual las personas

buscan viajar a ese tipo de sitios por el mero deseo de sentir aquello que en su momento se vivió en dichos lugares.

El mundo actual, al igual que los viajeros, está en continuo movimiento. Booking.com, que tiene por misión hacer que descubrir el mundo sea más fácil para todos. Prevé que en 2020 los viajes alcanzarán unos niveles sin precedentes, impulsados por la tecnología, el creciente sentido de la responsabilidad y una mayor conexión con la gente y los lugares que visitamos (Gómez, 2020).

La siguiente es una lista de nociones que tiene el turista en cuenta a la hora de realizar “turismo experiencial” además de considerarlos como ejes inspiradores para innovar en turismo.

En la **Figura 1** se muestra una secuencia que representa la lista de nociones que se tiene en cuenta a la hora de realizar el turismo experiencial, en primer lugar, se encuentra la “Visita lo menos conocido”, en segundo lugar, se encuentra “Esperar lo tecnológicamente inesperado”, en tercer lugar, se encuentra “Viajar despacio como nueva experiencia”, en cuarto lugar se encuentra “Escapadas con variedad de experiencias”, en quinto lugar, se encuentra “Las mascotas como prioridad”, en sexto lugar, se encuentra “Escapadas para sentirse joven”, en séptimo lugar, se encuentra “La gastronomía como prioridad”, y por último se encuentra “Planes de viaje a largo plazo”.



**Figura 1 Lista de nociones del turista.**

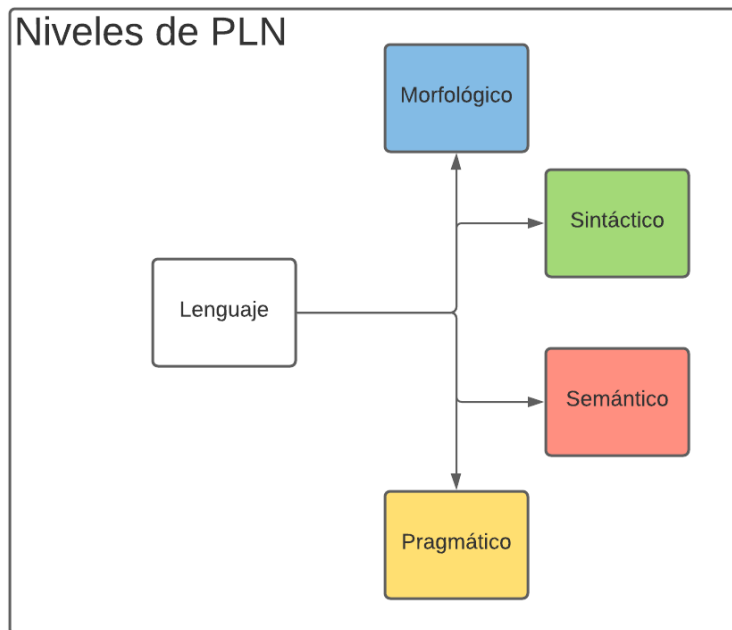
**Fuente:** (Gómez, 2020).

## 4.2 Análisis de sentimientos

El análisis de sentimientos se define como “Conjunto de técnicas computacionales para la extracción, clasificación, comprensión y evaluación de opiniones expresadas en fuentes publicadas en internet, comentarios en portales web y en otros contenidos generados por usuarios” (Erik Cambria, 2012).

Todo sistema de PLN (Procesamiento de Lenguaje Natural) debe llevar a cabo un conjunto de tareas de análisis del lenguaje que faciliten el entendimiento entre el usuario y el propio sistema (Pauli, 2019).

En la **Figura 2** se muestra los niveles de PLN asociados al lenguaje encargado de proporcionar las pautas del que se guiará cada nivel y elegir el correspondiente para su debido proceso.



**Figura 2 Niveles de PLN.**

**Fuente:** (Pauli, 2019).

- **Nivel morfológico:** en este componente se examinan las palabras para extraer raíces, rasgos flexivos, sufijos, prefijos y otros elementos. Su objetivo es entender cómo se construyen las palabras a partir de unidades de significado más pequeñas denominadas morfemas (Pauli, 2019).
- **Nivel sintáctico:** analiza la estructura de las oraciones en base al modelo gramatical empleado con el objetivo de conocer cómo se unen las palabras para crear oraciones (Pauli, 2019).
- **Nivel semántico:** proporciona sentido a las oraciones y les otorga un significado, resolviendo además las ambigüedades léxicas y estructurales que pudieran aparecer (Pauli, 2019).
- **Nivel pragmático:** se encarga del análisis de los textos más allá del de una oración aislada, teniendo en consideración aquella inmediatamente anteriores la relación existente entre ellas y el contexto en el que se producen (Pauli, 2019).

El análisis de sentimientos es un campo de investigación dentro del PLN que trata de extraer de manera automática y mediante técnicas computacionales información subjetiva expresada en el texto de un documento dado y acerca de un determinado tema. Mediante el análisis de sentimientos se puede saber si un texto presenta connotaciones positivas, negativas o neutras (Pauli, 2019).

Sin embargo, el análisis de sentimientos constituye una de las aplicaciones de la lingüística computacional más activa actualmente y en la que las empresas están realizando una importante inversión económica por el impacto que tiene en su funcionamiento y obtención de beneficios el conocer la opinión de los clientes o potenciales clientes acerca de sus productos (Cesteros, 2018)

En definición, es posible deducir cuáles serían los componentes que formarían parte de la definición formal de una opinión, del cual, se define en una cuádrupla (o, s, h, t) en donde o es el objeto de opinión, s el sentimiento, positivo, neutro o negativo, h es la persona que expresa dicha opinión y t el momento en el que lo hace, por ende, la definición, aunque concisa y válida, no permite representar determinados elementos que aparecen en el ejemplo, tales como el sonido, la imagen o el sistema de navegación (Sande, 2018).

#### **4.2.1 Análisis de sentimientos en texto.**

El análisis de sentimientos en texto es una rama de la computación afectiva basada en el procesamiento del lenguaje natural. Estos siguen procedimientos que dependen en gran medida de varios aspectos, como puede ser el tipo de texto de entrada, el tipo de resultados que se desean obtener o que información sobre los sentimientos queremos que nos provean y finalmente qué métodos de clasificación se deseen utilizar (Zambrano, 2017)

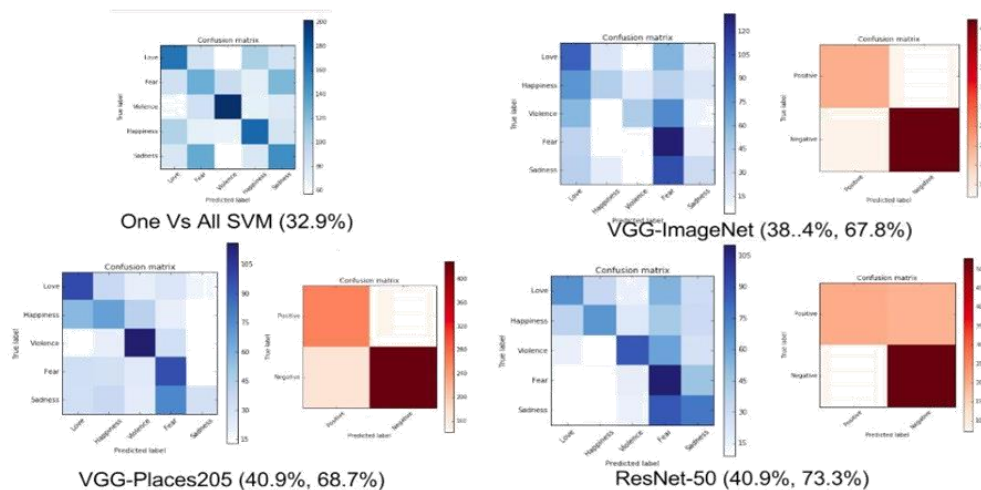
#### **4.2.2 Análisis de sentimientos en imágenes.**

Por otro lado, el análisis de sentimientos en texto particularmente permite clasificar la polaridad de un texto dado, es decir si la opinión expresada en un documento o una oración es positiva, negativa o neutra. También podemos obtener clasificaciones más detalladas que

buscan, por ejemplo, estados emocionales tales como “enfado”, “tristeza” o “felicidad” (Zambrano, 2017)

En el análisis de sentimientos de imágenes, es la forma de clasificar una polaridad por medio de una matriz de confusión, donde se indican que existen emociones positivas (amor, felicidad), negativas (violencia, miedo, tristeza) y neutras (asombro, admiración, éxtasis) (Vasavi Gajarla)

En la **Figura 3** se observa cómo es el análisis de sentimientos en imágenes, colocando referencias puntos en un plano cartesiano en 2D para saber qué tan oscura esta la imagen en esos contornos y saber si es positiva, negativa o neutra, especificar que hace referencia a imágenes con rostro y ojos.

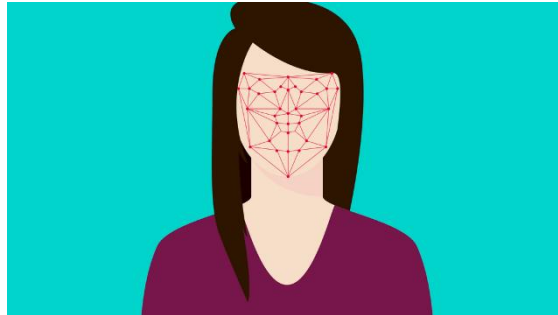


**Figura 3 Análisis de sentimientos por imágenes.**

**Fuente:** (Vasavi Gajarla).

Por el contrario, el reconocimiento facial utiliza algoritmos (analiza cientos de rostros y utiliza un mapeo facial que capta 100 expresiones faciales) para el procesamiento de imágenes, todas las imágenes tienen 50 x 50 píxeles de ancho y alto y una direccionalidad de 2500 píxeles, por lo que su costo computacional va en aumento (José Augusto Cadena Moreano, 2017)

En la **Figura 4** se ve cómo actúa los algoritmos en el reconocimiento facial para poder procesar y reconocer el usuario indicado en el sistema.



**Figura 4 Reconocimiento facial.<sup>1</sup>**

La detección, localiza la región facial (si existe) y la segmentación de la misma del resto de la escena. El acondicionamiento localiza los componentes y la escala a la que encontramos al rostro, mediante transformaciones geométricas. Siguiendo el proceso la normalización, consiste en normalizar las imágenes en la etapa de pre procesado de modo que se atenúen los efectos de los cambios de iluminación, es decir, se iguala en tamaño, intensidad de un rango determinado entre otras y así realizar un escalado y recorte con un rectángulo o elipse (José Augusto Cadena Moreano, 2017).

Se puede decir que el reconocimiento facial es una herramienta muy importante en el medio en tanto permite identificar a través de ciertas características a un individuo, aunque a veces resulta beneficioso el reconocimiento facial, no se debe olvidar que el mal uso del mismo afecta de gran manera al desarrollo del individuo (José Augusto Cadena Moreano, 2017).

### **4.3 Arquitectura de Big Data**

En el artículo de “**Arquitectura y Patrones de Big Data**” publicado por IBM, da a conocer un enfoque simplificado sobre la arquitectura general de Big data, con enfoque a escenarios

---

<sup>1</sup> Obtenido de <https://n9.cl/hiogh>

empresariales que constituyen un problema de Big data, determinando soluciones oportunas para aplicar al desarrollo (Divakar Mysore, 2013).

En la **Figura 5** se observan varios problemas empresariales regados en un cuadro más grande indicando que cada uno es un problema diferente, en primer lugar, están los “Datos generados por maquina” en la casilla de color “azul” que son generados por sensores en tiempo real de maquinaria industrial o vehículos que registran el comportamiento del usuario de modo online, en segundo lugar, están “Las telecomunicaciones: analítica de deserción de clientes” en la casilla de color “morado” que se utiliza para predecir y evaluar las estadísticas de mal servicio, costo/efectividad, etc. En tercer lugar, se encuentra “Marketing: análisis de sentimientos” en la casilla de color “verde aguamarina” que consta de evaluar opiniones públicas sobre un evento o un producto a través del análisis de datos, en cuarto lugar, se encuentra el “Servicio al cliente: supervisión de llamadas” en la casilla de color “rojo”, se encarga en la gestión y control de alguna petición del usuario prestador de servicio, en quinto lugar, “Comercio: mensajería personalizada basada en el reconocimiento facial y las redes sociales” en la casilla de color “verde claro”, se utiliza para él envió, confirmación o verificación de algún cambio importante en la información del usuario, en sexto lugar, está el “Comercio y marketing: datos móviles y focalización basada en la ubicación” en la casilla de color “rosado”, se basa en servicios GPS para poder localizar al usuario y prestar un servicio de ubicación o llegar a un punto de reunión, y por último, está el “FSS, ciudad de la salud: detección de fraudes” en la casilla de color “naranja” ,en este espacio aplica la realización de pruebas como evaluaciones, documentos, etc., que necesitan ser validados para tener un visto bueno.

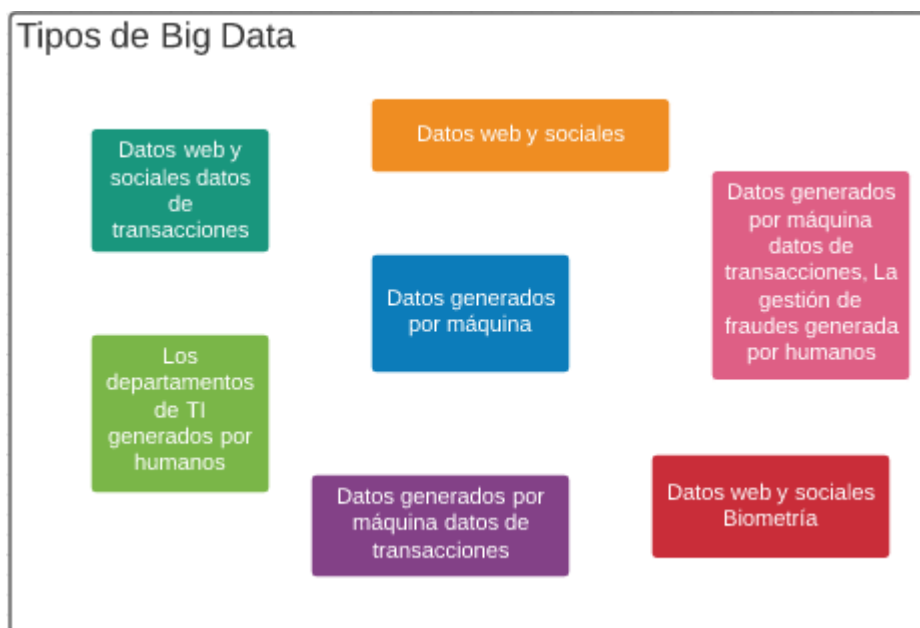


**Figura 5 Problemas empresariales de Big data.**

**Fuente:** (Divakar Mysore, 2013)

En la **Figura 6** se evidencia los tipos de Big data implementados en cada uno de los problemas explicados en la figura anterior, en el caso de “Datos web y sociales datos de transacciones” en la casilla de color “verde aguamarina” está asociado con el problema de “Servicios público: predecir el consumo de energía”, en segundo lugar, esta “Datos web y sociales” en la casilla de color “naranja” que está enfocado al problema de “Marketing: análisis de sentimientos”, en tercer lugar, se da por “Datos generados por maquinas” en la casilla de color “azul” que hace parte del problema de “Servicios públicos: predecir el consumo de energía”, en cuarto lugar, están los “Datos generados por maquina Datos de transacciones la gestión de fraudes generada por humanos” en la casilla de color “rosado” está vinculada con el problema “FSS, cuidado de la salud: detección de fraudes”, en quinto lugar, se encuentra los “Datos web y sociales Biometría” en la casilla de color “rojo” que esta junto con el problema de “Comercio: mensajería personalizada basada en el reconocimiento facial y las redes sociales”, en sexto lugar, esta los “Datos generados por maquina datos de transacciones” en la casilla de color “morado” que se encuentra junto al problema “Comercio y marketing: datos móviles y focalización basada en la ubicación”, y

por último, están “Los departamentos de TI generados por humanos” en la casilla de color “verde claro” que contiene como problema a “Servicio al cliente: supervisión de llamadas”.



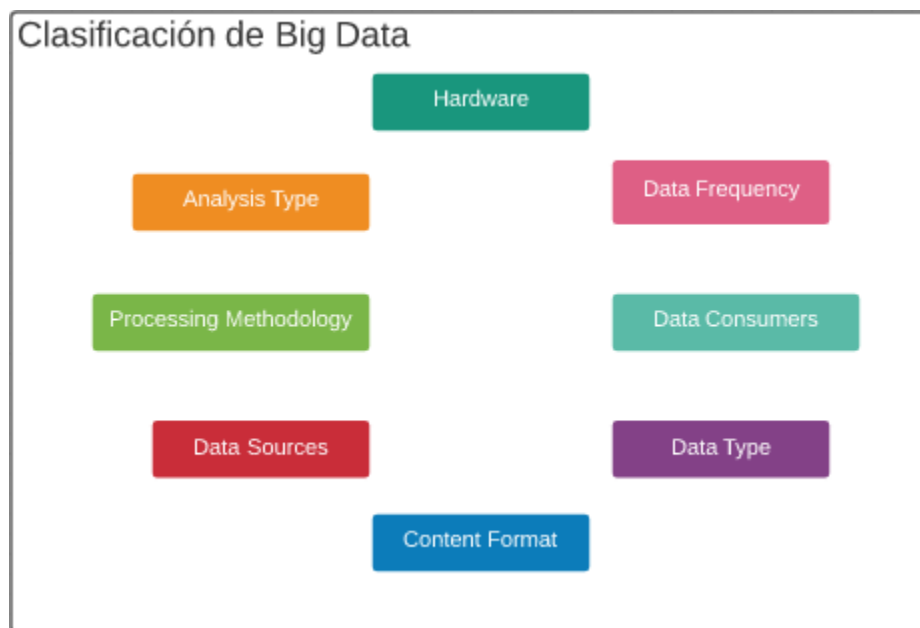
**Figura 6 Tipos de Big data.**

**Fuente:** (Divakar Mysore, 2013)

Con la clasificación de la tabla se logra una facilidad de entender cómo se adquieren los datos, cómo se procesan en el formato adecuado y con qué frecuencia están, con ello, se realiza un filtro en el cual evalúan por tipo de contenido, tipo de datos, la frecuencia con la cual los datos se ponen a disposición, la intención y si el procesamiento debe ser en tiempo real (Divakar Mysore, 2013).

En la **Figura 7** se explica la clasificación de Big Data dependiendo del caso en que se demuestre, en primer lugar, está el “Hardware” en la casilla de color “verde aguamarina” que tiene la comodidad del hardware y el estado del mismo, en segundo lugar, se encuentra el “Data Frequency” en la casilla de color “rosado” encargado de alimentar la demanda, alimentar en tiempo real y en serie de tiempos, en tercer lugar, está el “Data Consumer” en la casilla de color “azul aguamarina” que tienen un repositorio de datos, aplicaciones de

empresas y procesos empresariales, en cuarto lugar, se encuentra el “Data Type” en la casilla de color “morado” es el que contiene los meta datos, el historial y la transaccionalidad, en quinto lugar, “Content Format” en la casilla de color “azul” contiene la estructura del Big Data , en sexto lugar, se encuentra “Data Sources” en la casilla de color “rojo” que tiene el organizador de la vía de datos, los datos biométricos, etc., en séptimo lugar, está’ el “Processing Methodology” en la casilla de color “verde claro” contiene los análisis predictivos, la analítica y reportes de consultas, y en por último, está el “Analysis Type” en la casilla de color “naranja” que es el que monitorea en tiempo real.



**Figura 7 Clasificación de Big data.**

**Fuente:** (Divakar Mysore, 2013)

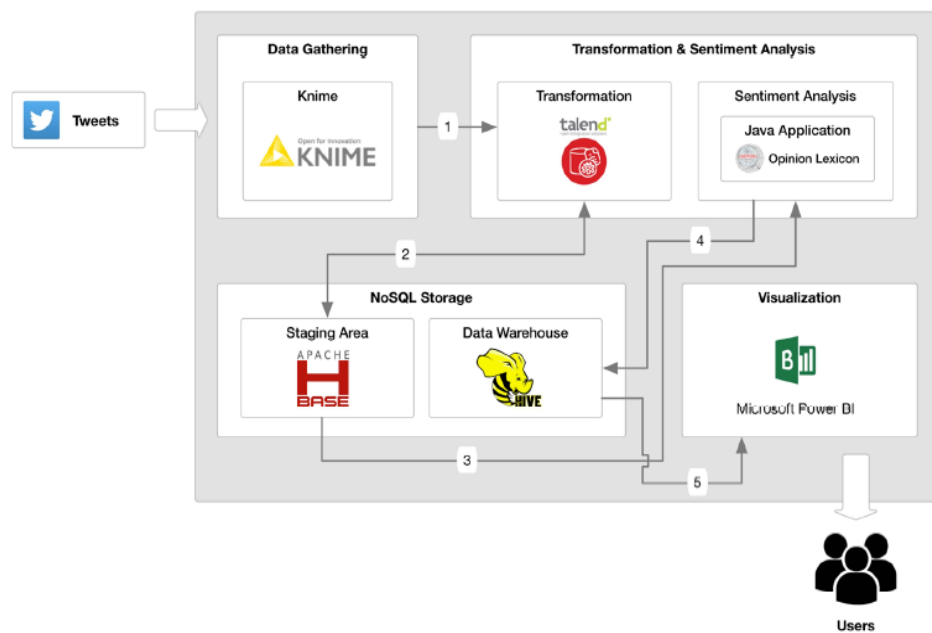
Establece una arquitectura lógica y capas de soluciones diferentes con respecto a Big data, desde el punto de acceso hasta el consumo de Big data. Además, incluye una lista de orígenes de datos y presentaremos los patrones atómicos que se enfocan en cada uno de los aspectos importantes de una solución de Big data (Divakar Mysore, 2013).

Existe un uso del tipo de Big data para clasificar las características de Big data, es útil observar las características de Big data junto con ciertas líneas que se recopilan, analizan y procesan los datos, por ejemplo:

- Tipo de análisis.
- Detección de fraude.
- Análisis de tendencias para decisiones empresariales estratégicas.
- Metodología de procesamiento.
- Frecuencia y tamaño de los datos.
  - Según la demanda.
  - Alimentación continua y tiempo real.
  - Serie de tiempo.
  - Tipos de datos.
  - Formato del contenido.
  - Origen de datos.
  - Consumidores de datos.
  - Procesos empresariales.
  - Usuarios empresariales.
  - Aplicaciones empresariales.
  - Personas individuales en diversos roles empresariales.
  - Parte de los flujos de proceso.
  - Otros repositorios de datos o aplicaciones empresariales.
  - Hardware.

Cuando un usuario compra en línea, generalmente comparte comentarios en el artículo y la tienda, o cuando se une a un evento o va a un restaurante, hotel o una película, él por lo general también hace un comentario al respecto. Todos estos datos pueden ser utilizados por varias partes interesadas en el proceso de toma de decisiones, considerando las opiniones expresadas (Carina Sofía Andrade and Maribel Yasmina Santos, 2017).

En la **Figura 8** se evidencia cómo entran los datos en este caso de Twitter con sus tweets a la primera fase que es la minería de datos por medio de la plataforma de Knime, luego pasa a la transformación que utiliza la herramienta talend para proporcionar una integración de los datos para que estos sean confiables y análisis de sentimientos con la herramienta Opinión Lexicon que es propiedad de Java, además, en el almacenamiento que es la siguiente fase, se utiliza base de datos NoSQL de apache llamada Hbase y para orientarla a un tema en específico se utiliza data warehouse y en la visualización está la última fase ver los datos en business intelligence.



**Figura 8 Versión preliminar de la arquitectura.**

**Fuente: (Carina Sofía Andrade and Maribel Yasmina Santos, 2017) Sentiment Analysis with Text Mining in Contexts of Big Data. International Journal of Technology and Human Interaction.**

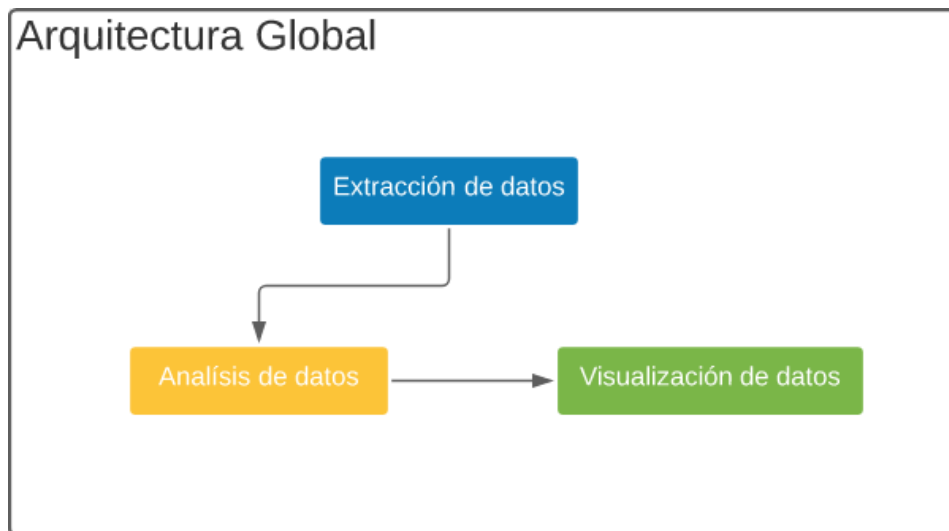
Con el desarrollo de la tecnología de Internet, estamos entrando en una nueva era de la nube y los grandes datos generados por las redes sociales y redes. Las redes sociales se han convertido en una herramienta popular para comunicarse y compartir información con personas en estos días. Social la influencia de los medios continúa creciendo y analizando las

redes sociales los datos se están convirtiendo en un tema importante para la investigación y el desarrollo y servicio para aprender sobre lo que la gente piensa y siente (Hong, 2017).

## 5.ARQUITECTURA PROPUESTA

La arquitectura está basada en el proyecto de investigación Doctoral de la Pontificia Universidad Javeriana “Modelo multimodal para el análisis de sentimientos en un dominio turístico”. La arquitectura consta de 3 fases que conforman el sistema de análisis al sector turístico del departamento de Boyacá: extracción, análisis y visualización. A su vez la arquitectura a implementar procesos de inteligencia artificial encargada de analizar los sentimientos de las imágenes extraídas por medio de reconocimiento facial, así como también de implementar un aplicativo web en la que se muestra los datos de todo el proceso realizado, ilustradas en el siguiente cuadro.

En la **Figura 9** se muestra la arquitectura global del proyecto implementada en un cuadro encerrando con tres fases indicando un lineamiento y orden entre ellas, en la primera fase, se encuentra “Extracción de datos” en la casilla de color “azul”, en la segunda fase, se encuentra “Análisis de datos” en la casilla de color “amarillo”, en la tercera y última fase, se encuentra la “Visualización de datos” en la casilla de color “verde claro”.



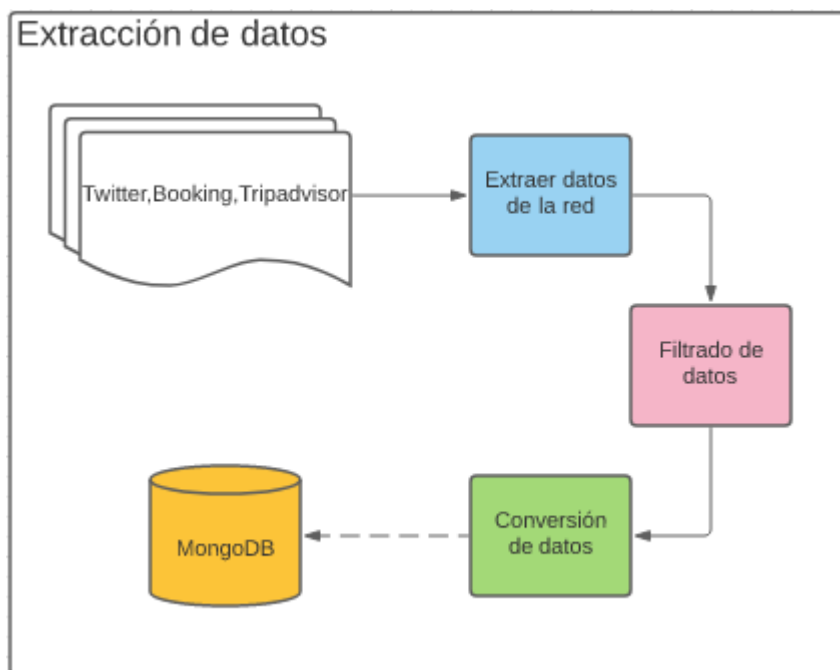
**Figura 9** Arquitectura Global.

**Fuente:** Autor.

En la primera fase se evidencia el proceso de la extracción de datos, esta extracción inicia por un escaneo de datos en las plataformas de Booking, TripAdvisor y twitter, dicha extracción de datos en la red se lleva a cabo con la herramienta previamente establecida “web scraper”.

Posterior a la extracción, la filtración de datos se realiza en Excel, seguido de esto se efectúa la conversión de datos, esto consiste en enlazar los datos, logrando que sea posible realizar consultas a través de esta; por último, los datos son colocados en la base mongoDB.

En la **Figura 10** se demuestra la primera fase que es la extracción de datos con sus respectivos procedimientos, en el primer procedimiento en la casilla de color “blanco”, se ingresa a los sitios web que se piensa extraer la información, en este caso Twitter, Booking y Tripadvisor, en el segundo procedimiento en la casilla de color “azul claro”, se extraen los datos respectivos del sitio web como comentario, calificación, por nombrar algunos, en el tercer procedimiento en la casilla de color “rosado”, se procede al filtrado de datos donde se eliminan comentarios repetidos, se organiza en orden alfabético por municipio, por nombrar algunos, en el cuarto procedimiento en la casilla de color “verde claro”, lo convertimos en archivos csv para que estos puedan ser almacenados, por último esta la base de datos que está en el cilindro de color “amarillo” de mongoDB, que almacena los archivos csv.

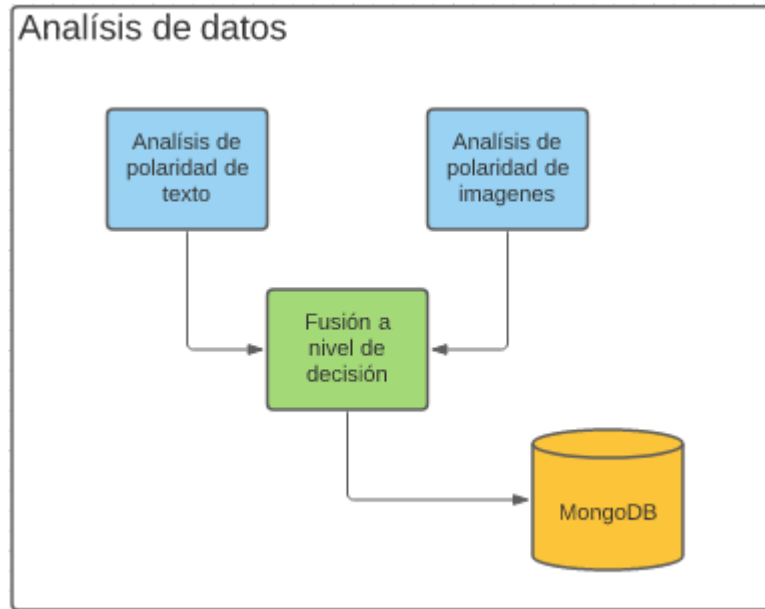


**Figura 10 Fase de extracción de datos.**

**Fuente: Autor.**

En la segunda fase, en el análisis de datos, se establece un algoritmo de clasificación de polaridad a los datos en texto e imagen, esto con tal de entender, en nuestro semáforo de emociones, si el usuario se encuentra a satisfacción o no ante la visita a cualquier sitio turístico en el departamento de Boyacá.

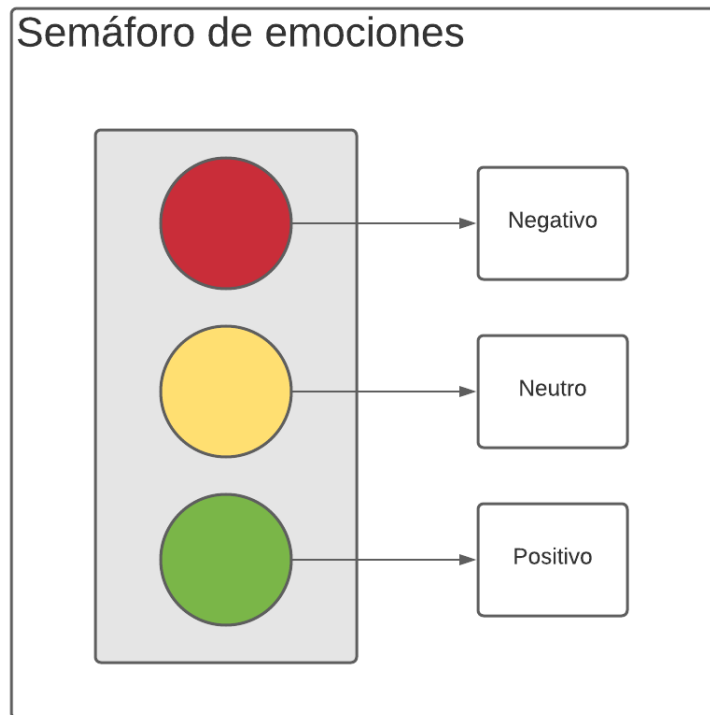
En la **Figura 11** se muestra la fase 2 de la arquitectura que es el análisis de datos, en este caso se tiene dos entradas que hace parte del primer procedimiento en la casilla de color “azul claro”, son el “Análisis de polaridad de texto” y el “Análisis de polaridad de imágenes” ya que produce la polaridad de las dos entradas y arroja un resultado, en el segundo procedimiento en la casilla de color “verde claro”, se hace la fusión de estos resultados arrojados y se colocan en las tres categorías establecidas en el semáforo de emociones, por último, procedimiento se almacenan estos datos en la base de datos que está en el cilindro de color “amarillo” de mongoDB para ser utilizados en el aplicativo web.



**Figura 11 Fase de análisis de datos.**

**Fuente: Autor.**

En la **Figura 12** se muestra el semáforo de emociones de forma ilustrada, es esta parte se demuestra que en el proyecto se utiliza tres tipos de clasificaciones negativo que se encuentra en el círculo de color “rojo”, neutro que se encuentra en el círculo de color “amarillo”, y positivo que se encuentra en el círculo de color “verde”.

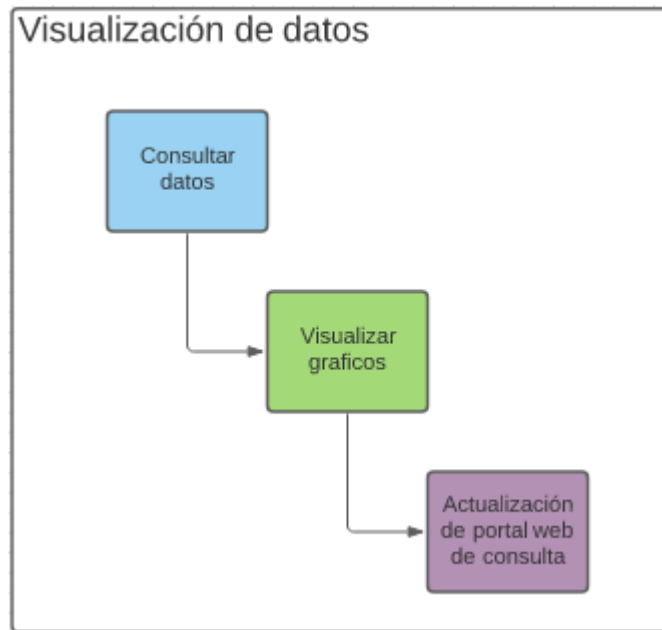


**Figura 12 Semáforo de emociones.**

**Fuente: Autor.**

En la tercera y última etapa, se encuentra la visualización de datos con la consulta a un aplicativo web hecha en un servidor local; dentro de esta se encuentran los datos obtenidos de los demás sitios web usados para la extracción de datos, para posteriormente ser mostrados y analizados.

En la **Figura 13** se muestra la fase 3 de visualización de datos que es la última, en el primer procedimiento en la casilla de color “azul claro”, el usuario entra al aplicativo web y consulta los datos de la plataforma que ofrece, en este caso, Booking, Tripadvisor y Twitter, en el segundo procedimiento en la casilla de color “verde claro”, se muestran los datos de forma gráfica y dinámica como en diagrama de barras y de torta, por último en la casilla de color “morado”, se actualiza el aplicativo web para ofrecer a los usuarios los resultados de finales de semana.



**Figura 13 Fase de visualización de datos.**

**Fuente: Autor.**

### **5.1 Extracción:**

La extracción de datos es definida como el proceso de extraer datos capturados dentro de fuentes semiestructuradas y no estructuradas (Ahmed, 2020), los datos estructurados se refieren a la información que se suele encontrar en la mayoría de bases de datos relacionales (RDBMS). Estos suelen ser archivos de texto que se almacenan en formato tabla, hojas de cálculo o bases de datos relacionales con títulos para cada categoría que permite identificarlos.

Para gestionar esta clase de datos se utiliza un tipo de lenguaje de programación estructurado, conocido como SQL (Structured Query Language) diseñado para administrar y recuperar información de sistemas de gestión de bases de datos relacionales (RDBMS).

Por su parte, los datos se caracterizan por que son generalmente binarios, es decir no poseen una estructura interna identificable. Los datos no estructurados que puede generar el ser se pueden almacenar en múltiples formatos como: documentos PDF, documentos Word, correos

electrónicos; estos habitualmente, se suelen categorizar como datos semiestructurados. No obstante, el campo del mensaje no está estructurado y las herramientas tradicionales de análisis no son capaces de identificarlos.

Por lo cual se hace uso de herramientas que permitan definir parámetros de la identificación de los datos; en este caso (Web scraping), esta herramienta se define como la técnica por la que un equipo de desarrolladores es capaz de scrapear o liberar datos de páginas web de gobiernos, instituciones públicas u organizaciones para acceder a datos privados o públicos que puedan ser publicados o distribuidos en formato abierto (BBVA API\_Market, 2016).

En este punto, se lleva a cabo la extracción necesaria de datos, que aplica la recuperación de información en la world wide web (www) que emplean la devolución de referencias a los documentos en lugar del propio documento utilizando el modelo de extracción de Ext-webservice que compone en la interface 2 servicios Web RestFul.

El primero consta de recibir una url que contiene la dirección del recurso que se va a extraer y devuelve dependiendo la petición en una lista de acciones de un plan o una lista de planes asociados, para ello es necesario apoyarse en 2 modelos descendientes Wrapper-plan y Browse-Cat., componente que recibe un recurso que representa la página HTML que contiene una categoría de planes, a su vez, a partir del ingreso de una página obtiene un conjunto de URL donde se encuentran planes de tarea (Bordignon).

Para efectos de la arquitectura se hizo uso de la herramienta “Web Scraper”, herramienta anclada como extensión en el navegador, esta es la encargada de extraer los datos de la página web tanto de Booking como TripAdvisor que serán útiles en el proceso de análisis.

Los datos extraídos con respecto al hotel visitado son:

<b>Variables</b>	<b>Tipo de datos</b>
Nombre del hotel	Texto
Cantidad de opiniones	Número
Calificación del hotel en general	Número
Dirección	Texto

Los datos extraídos con respecto al usuario visitante son:

<b>Variables</b>	<b>Tipo de datos</b>
Nombre de usuario	Texto
Título de comentario	Texto
Comentario	Texto
Calificación	Número
Ciudad	Texto
Imagen (dependiendo el caso)	Imagen
Fecha del comentario	Fecha

## 5.2 Análisis:

El análisis de datos se hace con el propósito de sacar conclusiones referentes a la información extraída para determinar futuras decisiones o solo para ampliar un catálogo de diversos temas, en este caso se establece consistentemente en someter los datos a la realización de operaciones que ayudan a alcanzar los objetivos del proyecto, por otro lado, este paso puede utilizarse en varios métodos (José Luis García Cué, 2008), (Antonio Manuel Rodríguez García, 2019), QuestionPro<sup>2</sup>:

---

<sup>2</sup> Obtenido de <https://n9.cl/nlo7>

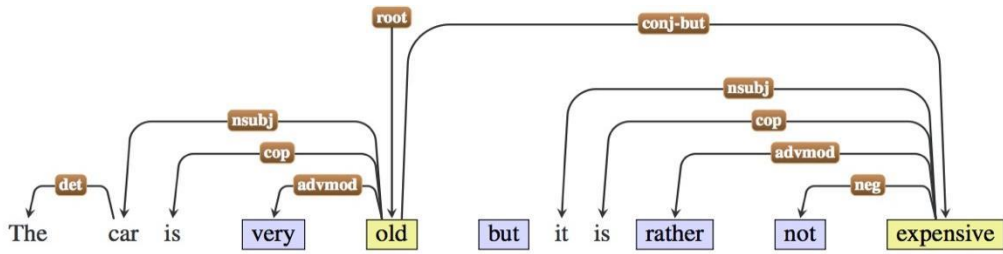
Mercadotecnia: Es usado principalmente para predecir el comportamiento de los consumidores, incluso para poder calificarlos.

Recursos Humanos: Es para saber cómo se mantiene una empresa en su entorno laboral y calificando a los empleados potenciales.

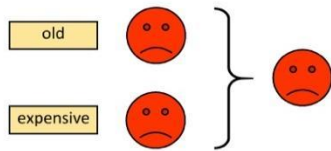
Académicos: Es para ver a los alumnos nuevos, registrarlos, medir el ingreso y el rendimiento de los mismos.

Ya que el procedimiento está basado en el análisis de polaridad por medio de texto e imágenes, se hace uso de la librería SenticNet (Erik Cambria, 2012) que aprovecha la IA subsimbólica (descendiente de un alto nivel de problemas de lógica matemática y de búsqueda) que descubre automáticamente las primitivas conceptuales que pueden generalizar mejor el conocimiento de sentido común, por ello es importante SenticNet gracias a la descomposición que concede expresiones traducidas de varias palabras que transmiten un conjunto específico de emociones y la polaridad de cada una de ellas, si bien, existen maneras innumerables de formas de expresar el mismo concepto en lenguaje natural y tener una lista de todos ellos es casi imposible.

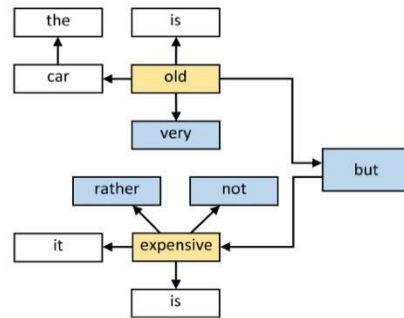
En la **Figura 14** se muestran cinco tipos de diagramas en los que se representan el funcionamiento del análisis de sentimientos, en el primer diagrama (a), evidencia un árbol de sentencias dependientes donde el valor recibido es una oración completa, pero se elimina las palabras que no definen un adjetivo calificativo, en el segundo diagrama (b), muestra a los resultados de la polaridad en las palabras, por ende, deducir cuales son negativas, positivas o neutras para arrojar el resultado final, en el tercer diagrama (c), se evidencia cómo actúa el árbol de sentencias dependientes, haciendo que estas sigan un camino por medio de condiciones que llegan a una conclusión, en el cuarto diagrama (d), se muestra de manera circuito electrónico indicando el curso dependiendo la entrada hasta llegar al resultado final, en el quinto y último diagrama (e), se demuestra cómo en cada condición del circuito electrónico tiene su resultado dependiendo en cual se establezca el resultado.



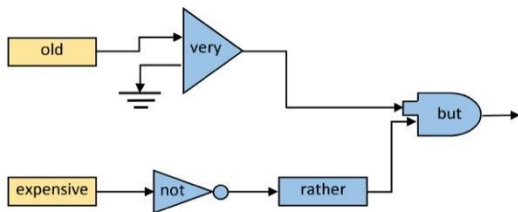
(a) Dependency tree of a sentence.



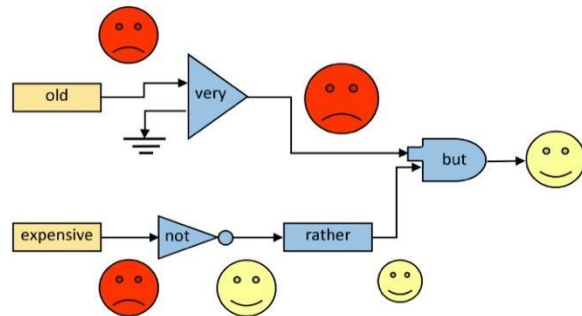
(b) The old way: averaging over a bag of sentiment words. The overall polarity of a sentence is given by the algebraic sum of the polarity values associated with each affect word divided by the total number of words.



(c) The dependency tree of a sentence resembles an electronic circuit: words shown in blue can be thought as a sort of "boolean operations" acting on other words.



(d) The electronic circuit metaphor: sentiment words are "sources" while other words are "elements", e.g., *very* is an amplifier, *not* is a logical complement, *rather* is a resistor, *but* is an OR-like element that gives preference to one of its inputs.

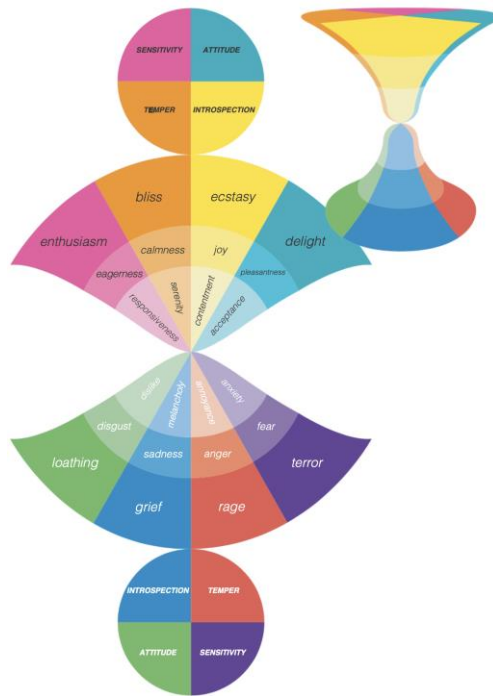


(e) The final sentiment data flow of the "signal" in the "circuit".

**Figura 14 Funcionamiento del análisis de sentimientos.**

**Fuente: (Erik Cambria, 2012)**

En la **Figura 15** se evidencia un mapa de sentimientos que SenticNet utiliza en forma 3D parece más a un reloj de arena, colocando así en la parte de arriba los datos positivos como alegría, serenidad, aprobación, interés, optimismo, entre otros, en la parte del medio, ubica a los datos que se asimilan para saber si pueden ser neutros, como asombro, admiración, éxtasis, aversión, entre otros. Y por último en la parte de abajo están los datos negativos como tristeza, ira, furia, temor, miedo, melancolía, odio, etc.



**Figura 15 Mapa de Sentimientos.**

**Fuente: (Erik Cambria, 2012)**

Por lo cual, las inflexiones semánticas de sinónimos o conceptos relacionados semánticamente deben estar guiadas por la dependencia conceptual y el razonamiento analógico, del mismo modo, en la utilización del proyecto fue fundamental ya que ayuda al archivo encargado del lenguaje español que es utilizado para analizar los patrones en cada comentario (Erik Cambria, 2012).

### **5.2.1 Librerías Utilizadas.**

Por otro lado, esta Sklearn es una data mining (minería de datos) que basa en implementar sus algoritmos en machine learning dando lugar a encontrarse librerías en prácticamente todas la técnicas, sin embargo, el lado negativo es su dificultad de aprendizaje porque cada paquete tiene una sintaxis propia, pero posee la facilidad de unificar las etapas de preprocesador, entrenamiento, optimización y validación de modelos predictivos, es importante en el proyecto porque ayuda en la optimización y el entrenamiento de las

imágenes y de texto (Scikit-Learn <sup>3</sup>), se encuentra también, la librería de pandas es una extensión de Numpy que ofrece la manipulación y análisis de datos en lenguaje de programación de Python, integrando DataFrames de manipulación de datos, alineación y manejo de datos faltantes, cadenas de operaciones e indexación jerárquica de ejes, este manejo es importante en el proyecto en la manipulación de los archivos extraídos (Pandas <sup>4</sup>), otro también, es la librería de Numpy que un escrito en C de muy alta velocidad cuando se refiere a trabajar conjuntos grandes de datos o mejor conocido como Big Data, este incluye operaciones matemáticas, lógicas de ordenador, estadísticas, entrada y salida de ficheros y Data Science, así, es importante en el proyecto porque se utiliza como soporte para pandas a la hora de manejar grandes cantidades de información en ficheros (Numpy <sup>5</sup>), y por último está la librería de Dlib es un modelo de reconocimiento facial capaz de detectar hasta 68 o más puntos característicos del rostro identificando una triangulación precisa y así identificar la emoción por medio de esa triangulación, es importante par proyecto ya que nos ayuda a identificar las imágenes que están en el proceso de análisis de polaridad en imágenes y sacar los resultados necesarios de cada imagen con sus emociones (Dlib <sup>6</sup>).

En la **Figura 16** se muestra como la librería Dlib actúa en el reconocimiento facial, en este caso se muestra en plano 2D donde cada punto referenciado equivale al rostro de una persona donde se ubica los ojos, las cejas, la boca y la nariz, todo para poder decidir en una imagen si tiene o no rostro.

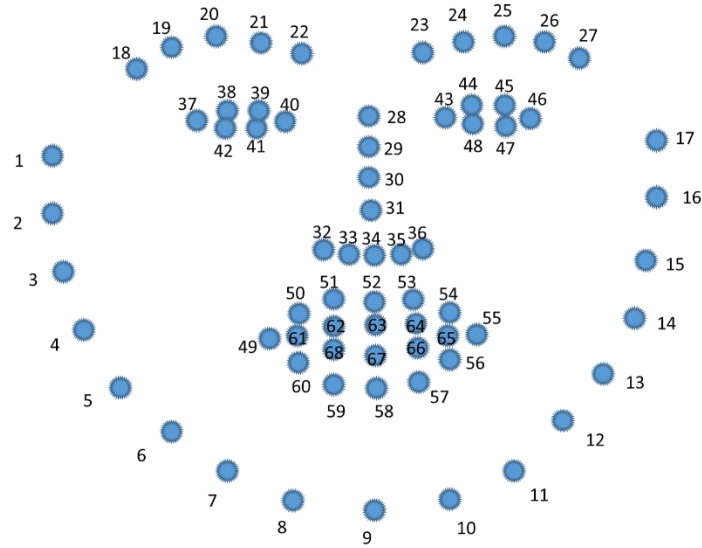
---

<sup>3</sup> Obtenido de <https://n9.cl/rq10h>

<sup>4</sup> Obtenido de <https://n9.cl/sw0i1>

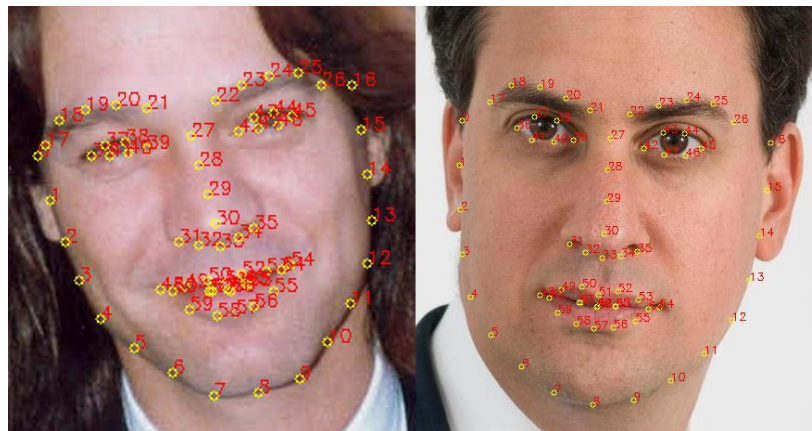
<sup>5</sup> Obtenido de <https://n9.cl/b7qsq>

<sup>6</sup> Obtenido de <https://n9.cl/t7513>



**Figura 16 Facial Landmarks de Dlib<sup>7</sup>**

En la **Figura 17** se muestra una imagen con el funcionamiento de Dlib en general cuando este reconoce un rostro, se observa que saca absolutamente todos los puntos al igual que la figura pasada, pero aplicada en inclinaciones de la propia imagen

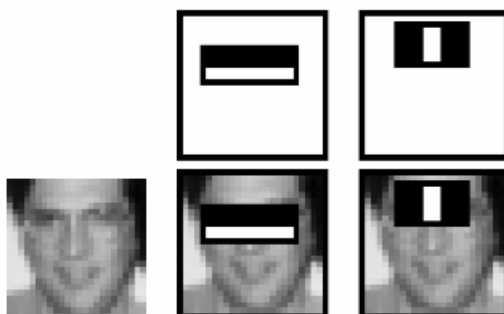


**Figura 17 Facial Landmarks de Dlib en imagen con rostro<sup>8</sup>**

<sup>7</sup> Obtenido de <https://n9.cl/6eh5>

<sup>8</sup> Obtenido de <https://n9.cl/ywryh>

En la **Figura 18** se muestra el funcionamiento de cómo identifica los ojos de la imagen, es el caso, es un cuadro que mide la región de los ojos a través de las mejillas superiores con una diferencia de intensidad, estas son intensidades que se comparan con el puente de la nariz, pero esta característica es relativa, puesto que debe ser insensible al tamaño y la ubicación de la cara a lo que se adapta cuando lo reconoce.



**Figura 18 Reconocimiento de ojos y rostro por medio de región <sup>9</sup>**

En este proceso se categorizan 2 tipos de entradas; una por medio de imágenes y otra por medio de texto, estas a su vez, son llevadas a un proceso de polarización que analiza a qué tipo de categoría pertenecen; las categorías son: positivo, negativo y neutral.

La polarización de imágenes tiene 6 categorías: miedo, tristeza, asco, angustia, felicidad y neutral; el valor obtenido de las seis categorías mencionadas, se ponderan en las 3 categorías mencionadas anteriormente: positivo, negativo y neutral, esto con tal de unir los resultados del texto y así obtener un resultado definitivo.

### **5.3 Visualización:**

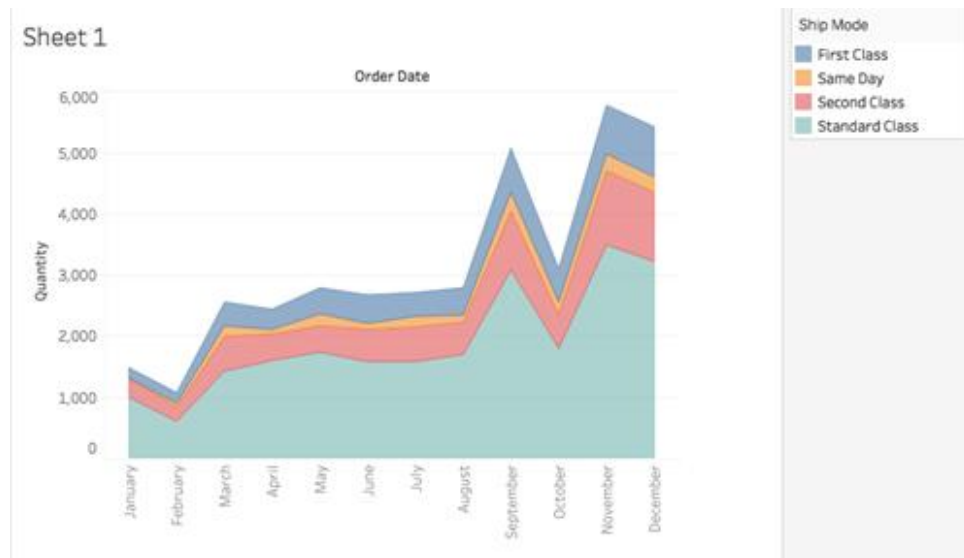
La visualización de datos es importante puesto que nuestros ojos son atraídos por los colores y patrones que se pueden identificar rápidamente, es una forma de arte visual que capta nuestro interés y mantiene nuestros ojos en el mensaje, cuando se enfoca en un gráfico directamente vemos tendencias y valores atípicos y cuanto más llamativos sean estos más eficaz puede ser la visualización (Rober Morales Chaparro, 2012) (Akram Hernandez Vásquez, 2019).

---

<sup>9</sup> Obtenido de <https://n9.cl/uhky>

Existen diferentes tipos de visualización:

- Gráfico de área: Es utilizado para representar los totales acumulados a lo largo del tiempo y son la forma convencional de visualizar líneas aplicadas (Tableau <sup>10</sup>) (Figura 19).



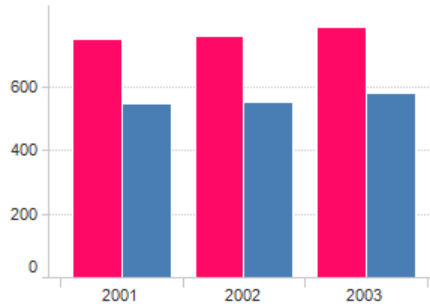
**Figura 19 Gráfico de área <sup>11</sup>**

- Gráfico de barras: Es la forma de resumir un conjunto de datos por categorías. Muestra los datos usando varias barras de la misma anchura, cada una representa una categoría concreta. La altura de cada barra es proporcional a una agregación específica (por ejemplo, la suma de los valores de la categoría que representa). Las categorías podrían ser desde grupos de edad a ubicaciones geográficas (TIBC <sup>12</sup>) (Figura 20).

<sup>10</sup> Obtenido de <https://n9.cl/63d4>

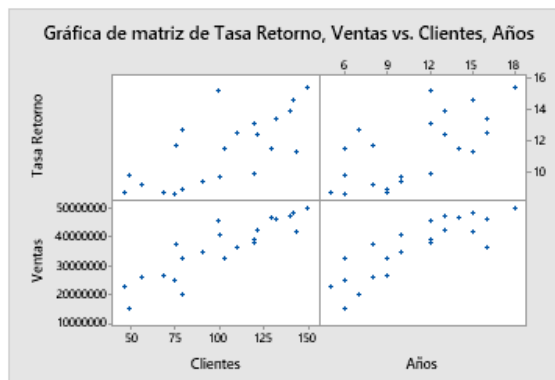
<sup>11</sup> Obtenido de <https://n9.cl/63d4>

<sup>12</sup> Obtenido de <https://n9.cl/8huz>



**Figura 20 Gráfico de barras** <sup>13</sup>

- Gráfico de matriz: Es un tipo de gráfica que acepta hasta 20 variables y crea una gráfica para cada combinación posible. Una matriz de gráficas es efectiva cuando usted tiene muchas variables y desea ver la relación entre los pares de variables (Minitab <sup>14</sup>) (**Figura 21**).



**Figura 21 Gráfico de matriz** <sup>15</sup>

- Gráfico de red: Consiste en una serie de gráficos interconectados a través de los cuales quedan representadas las distintas tareas de un proyecto, así como las dependencias que hay entre ellas y la ruta que ha de seguir un proyecto en su ejecución (ekon <sup>16</sup>) (**Figura 22**).

<sup>13</sup> Obtenido de <https://n9.cl/8huz>

<sup>14</sup> Obtenido de <https://n9.cl/q2g2g>

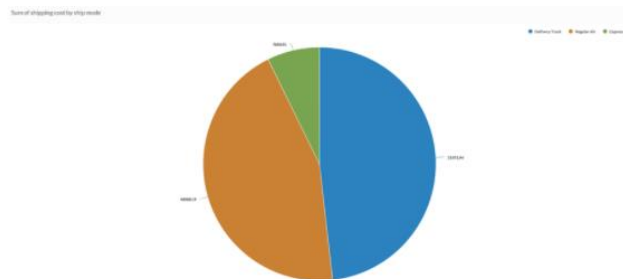
<sup>15</sup> Obtenido de <https://n9.cl/q2g2g>

<sup>16</sup> Obtenido de <https://n9.cl/o0td>



**Figura 22 Gráfico de red <sup>17</sup>**

- Gráfico circular: Muestra a las categorías como una proporción o porcentaje de la totalidad. Utilice gráficos circulares para mostrar la composición de los datos de una categoría en los que cada segmento es proporcional a la cantidad representada (HighBond <sup>18</sup>) (**Figura 23**).



**Figura 23 Gráfico circular <sup>19</sup>**

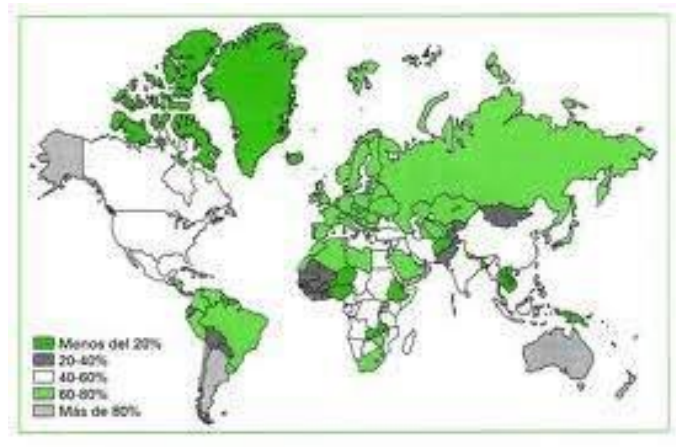
- Cartograma: Son unos gráficos que se realizan de forma de mapa, indicando según sus características las distintas zonas y las cantidades según su representación, es decir que representan las unidades geométricas y sus atributos como son expuestos en el área, solo mostrando los cambios que dependen del incremento, como por

<sup>17</sup> Obtenido de <https://n9.cl/7k4zb>

<sup>18</sup> Obtenido de <https://n9.cl/8o9v0>

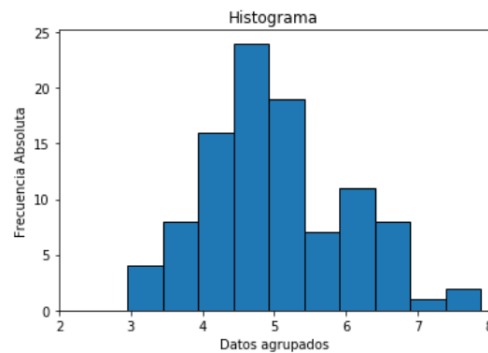
<sup>19</sup> Obtenido de <https://n9.cl/8o9v0>

ejemplo no representar el espacio geográfico sino la cantidad de habitantes que tiene si se trata de un cartograma de un país (ConceptoDefinicion <sup>20</sup>) (**Figura 24**).



**Figura 24 Cartograma <sup>21</sup>**

- Histograma: Un histograma es la representación gráfica en forma de barras, que simboliza la distribución de un conjunto de datos. Sirven para obtener una “primera vista” general, o panorama, de la distribución de la población, o de la muestra, respecto a una característica, cuantitativa y continua (Superprof <sup>22</sup>) (**Figura 25**).



**Figura 25 Histograma <sup>23</sup>**

<sup>20</sup> Obtenido de <https://n9.cl/oep73>

<sup>21</sup> Obtenido de <https://n9.cl/fatr>

<sup>22</sup> Obtenido de <https://n9.cl/3ofpx>

<sup>23</sup> Obtenido de <https://n9.cl/3ofpx>



solicitudes entre el código y al aplicativo web (npm morgan, 2020)<sup>28</sup>, por otro lado, se utiliza la herramienta de mongoose que es básicamente es un modelado de mongoDB para funcionar de manera asíncrona, por lo que admite promesas como devoluciones de llamada, es importante en el proyecto ya que realiza la conexión de la base de datos al aplicativo web (npm mongoose, 2020)<sup>29</sup>.

Se utiliza la extensión de ejs ya que otorga los css del aplicativo web para poder implementar alguna plantilla ya establecida o hacer una desde cero, es importante en el proyecto ya que embellece el aplicativo web y puede ser actualizado el diseño cuando sea requerido (npm ejs, 2020)<sup>30</sup>, y por último, está el nodemon una herramienta que ayuda a desarrollar aplicaciones basadas en node.js al reiniciar automáticamente la aplicación de nodo cuando se detectan cambios de archivo en el directorio, es importante en el proyecto (npm nodemon, 2020)<sup>31</sup>.

Por consecuencia, se determina que el usuario necesita consultar de alguna manera sus datos, para ello se potencia un aplicativo web que se puede consultar por medio de la plataforma y del municipio que desea para así darle control y tranquilidad al usuario a la hora del viaje hacia la región de Boyacá.

---

<sup>28</sup> Obtenido de <https://n9.cl/7m0p>

<sup>29</sup> Obtenido de <https://n9.cl/hwk3>

<sup>30</sup> Obtenido de <https://n9.cl/psnah>

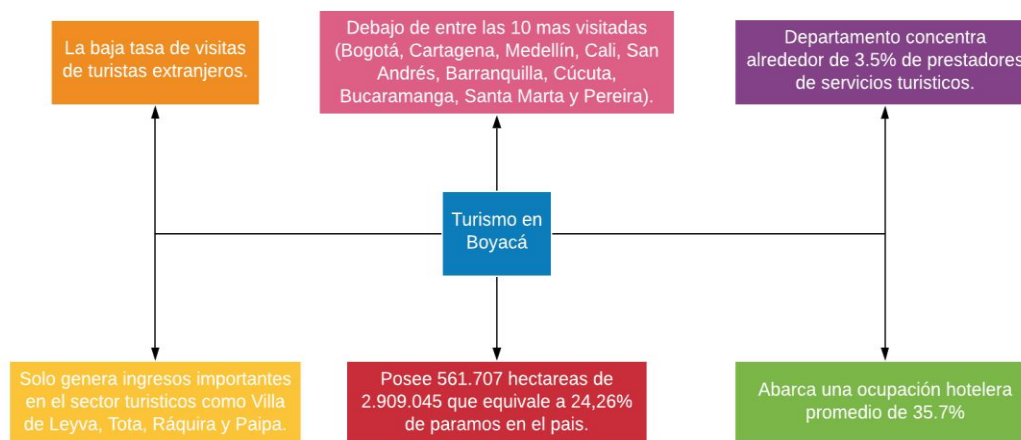
<sup>31</sup> Obtenido de <https://n9.cl/odh0>

## 6. PRUEBAS Y RESULTADOS

### 6.1 Pruebas

#### 6.1.1 Fase 1 Estado del arte.

En la siguiente imagen **Figura 27** se representa la fase 1 del estado del arte y problemática del proyecto, todo esto basados en artículos científicos, artículos psicológicos, noticias gubernamentales y artículos de las herramientas implementadas en el proyecto, todo esto para consolidar las buenas bases y prácticas para el inicio del proyecto.



**Figura 27 Problemática y estadísticas de Boyacá** <sup>32</sup>

#### 6.1.2 Fase 2 Modelamiento del sistema.

En la siguiente imagen **Figura 28** se muestra la fase 2 del proyecto, que consta del modelamiento del sistema en general que utiliza la realización de la extracción de datos, análisis de datos visualización de datos, herramientas de extracción y almacenamiento de datos, además se muestra las herramientas y plataformas que se utiliza en cada tarea correspondiente.

<sup>32</sup> Obtenido de <https://n9.cl/br2px>



**Figura 28 Modelamiento del sistema**

**Fuente: Autor**

### **6.1.3 Fase 3 Implementación.**

En la siguiente imagen **Figura 29** se implementa la fase 3 de implementación, consta de subir el proyecto a un servidor con prestaciones de cPanel y WHM, junto con el almacenamiento NoSQL de la base de datos mongoDB y el aplicativo web de estructura nodejs en la parte de las funciones de consulta.



**Figura 29 Servicios presentados en el servidor, almacenamiento y aplicativo web** <sup>33 34</sup>  
35

**6.1.4 Fase 3 Pruebas.**

En la siguiente imagen **Figura 30** se muestra la fase 3 de pruebas, en el que se verifica los datos junto a la polaridad para saber si están correctamente organizados con el positivo, negativo y neutro, además, se utiliza una nube de palabras que pueden determinar las palabras más usadas por los viajeros.



**Figura 30 Categoría de polaridades y nube de palabras** <sup>36 37</sup>

<sup>33</sup> Obtenido de <https://n9.cl/1w4bu>

<sup>34</sup> Obtenido de <https://n9.cl/xyg7d>

<sup>35</sup> Obtenido de <https://n9.cl/jr8l0>

<sup>36</sup> Obtenido de <https://n9.cl/8n6p2>

<sup>37</sup> Obtenido de <https://n9.cl/stpuo>

## 6.2 Resultados

Se realizó la extracción de los archivos csv con la extensión de “Web Scraper” tanto para Booking como TripAdvisor, se pueden observar después aplicados los filtros de la siguiente manera:

En la **Tabla 1** se visualiza los datos extraídos de la plataforma de TripAdvisor organizados de manera alfabética por municipio detallada en una hoja de cálculo.

<b>Nombre Hotel</b>	<b>Título Comentario</b>	<b>Comentario</b>	<b>Código Departamento</b>	<b>Código Municipio</b>	<b>Municipio</b>
Costa Azul Glamping	Luna de miel mágica	estuvimos en costa azul glamping hospedados por dos días el fin de semana de amor y amistad.	15	047	Aquitania
Hotel Refugio Turístico & Restaurante Dylam	La mejor atención	En todos los hoteles que he visitado nunca uno se ha preocupado tanto por sus huéspedes como este, la atención es la mejor sin duda.	15	047	Aquitania
Hotel Sol de la villa	Delicioso	Lo recomiendo.	15	407	Villa de leyva
Hotel Meson de los virreyes	¡Un buen descanso!	Es excelente para pasar con la familia.	15	407	Villa de leyva

**Tabla 1 Datos de TripAdvisor organizados en orden alfabético aplicados con los filtros.**

**Fuente: Autor.**

En la **Tabla 2** se muestran los datos organizados de la plataforma de Booking en orden alfabético por municipio detallada en una hoja de cálculo.

<b>Nombre Hotel</b>	<b>Título Comentario</b>	<b>Comentario Positivo</b>	<b>Comentario Negativo</b>	<b>Código Departamental</b>	<b>Código Municipal</b>	<b>Municipio</b>
La cabaña del sol	Naturaleza pura			15	047	Aquitania
El camino real			Un poco sucio por el precio que pagamos	15	047	Aquitania
Hotel casa terra			El colchón es un poco blando	15	407	Villa de leyva
Hotel capitán Ricaurte		La atención y el servicio fue muy buena		15	407	Villa de leyva

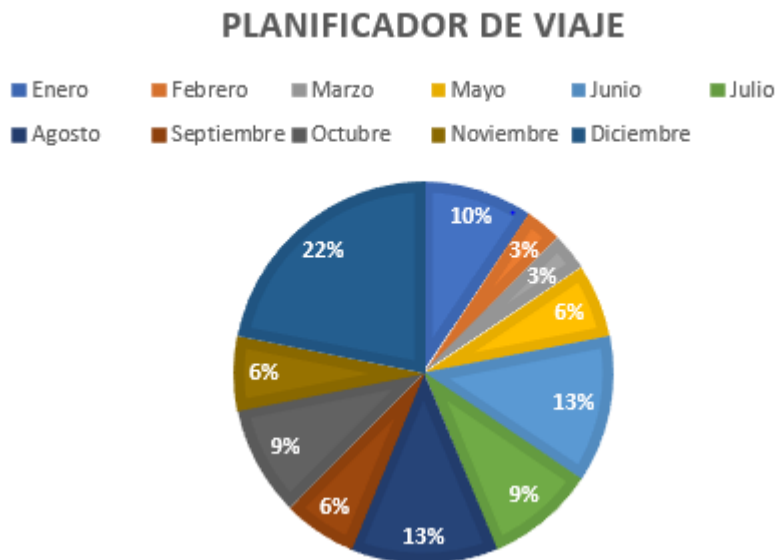
**Tabla 2 Datos de Booking organizados en orden alfabético aplicados con los filtros.**

**Fuente: Autor.**

Parte de la extracción se basa en sustraer las fechas para que posterior al análisis; donde se evidencie el tráfico de turistas por meses, permitiendo comprender en que época del año se encuentra con mayor y con menor tráfico de turistas, a raíz de entender la fecha se podrá inferir que sitios son más concurridos en las épocas del año.

En la **Figura 31** se presenta un diagrama con los valores de cada mes del año, con un 10% para enero, un 3% para febrero, un 3% para marzo, un 6% para mayo, un 13% para junio, un 9% para julio, un 13% para agosto, un 6% para noviembre, un 9% para octubre, un 6% para

noviembre y un 22% para diciembre, esto hace referencia a la cantidad de turistas que se movilizan a la región de Boyacá dependiendo el mes del año.

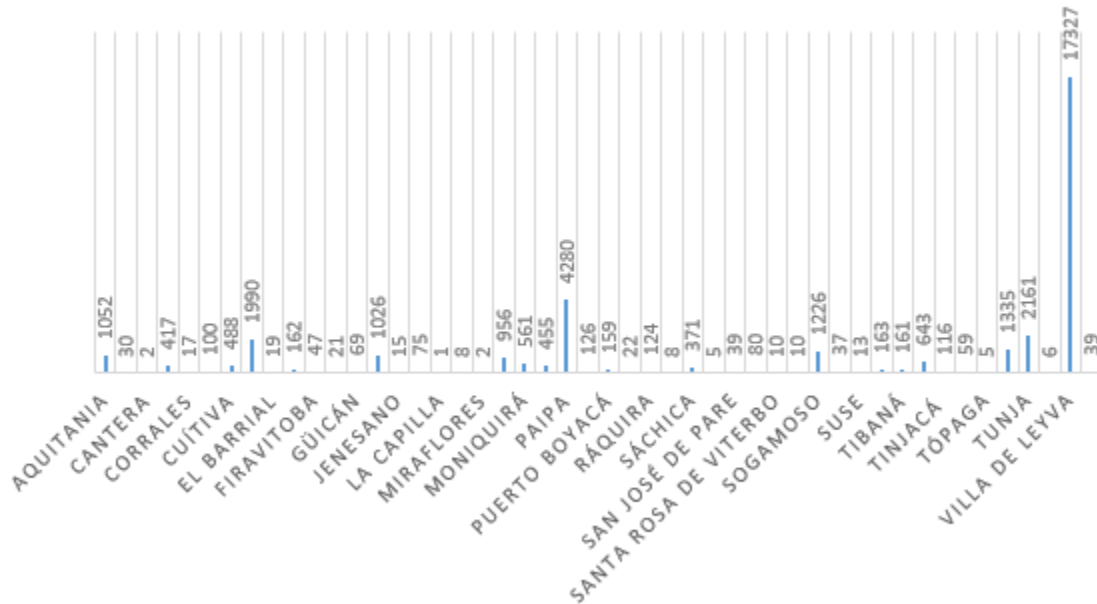


**Figura 31 Planificador de Viaje de fechas con mayor tráfico.**

**Fuente: Autor.**

En la **Figura 32** se muestra cuáles de los municipios es más visitado en la región de Boyacá durante todo el año en la plataforma de Booking, con una gran mayoría para el municipio de Villa de Leyva con casi 18000 visitas registradas aproximadamente, después el municipio de Paipa con un poco más de 4000 visitas registradas aproximadamente, con el municipio de Tunja con un poco más de 2000 visitas registradas aproximadamente y el resto de municipios con menos de 2000 visitas registradas aproximadamente.

## PLANIFICADOR DE VIAJES (BOOKING)

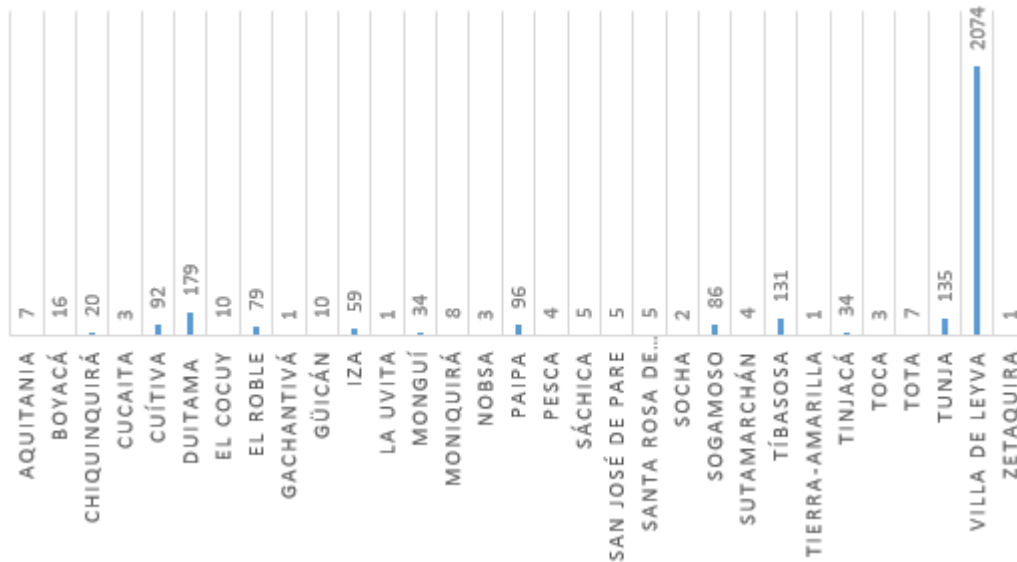


**Figura 32 Planificador de Viajes en diagrama de barras de los sitios más visitado por la plataforma de Booking.**

**Fuente: Autor.**

En la **Figura 33** se muestra cuáles de los municipios es más visitado en la región de Boyacá durante todo el año en la plataforma de TripAdvisor, con una gran mayoría para el municipio de Villa de Leyva con más de 2000 visitas registradas aproximadamente y el resto de municipios con menos de 200 visitas registradas aproximadamente.

## PLANIFICADOR DE VIAJES (TRIPADVISOR)

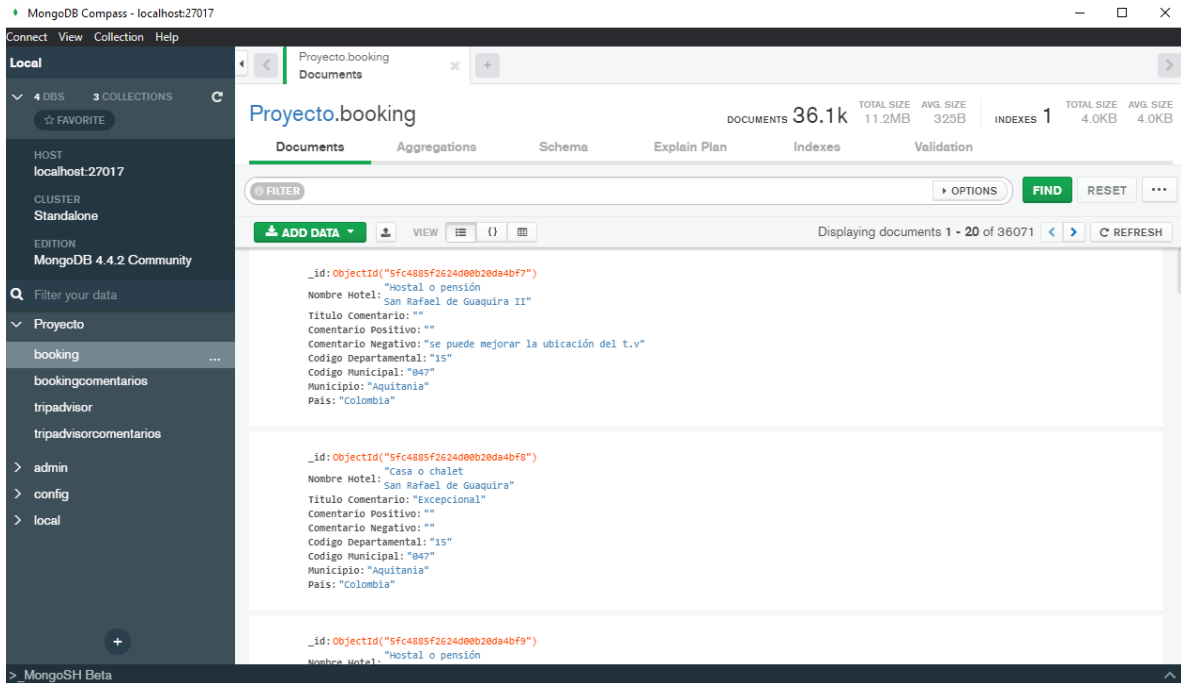


**Figura 33 Planificador de Viajes en diagrama de barras de los sitios más visitado por la plataforma de TripAdvisor.**

**Fuente: Autor.**

Estos datos en archivos csv son almacenados en la base de datos de mongoDB, donde se reúne la información; cabe resaltar que dichos datos no son usados aún ya que para este punto la polaridad no ha sido aplicada, sin embargo, estos datos son tomados como puntos de referencia.

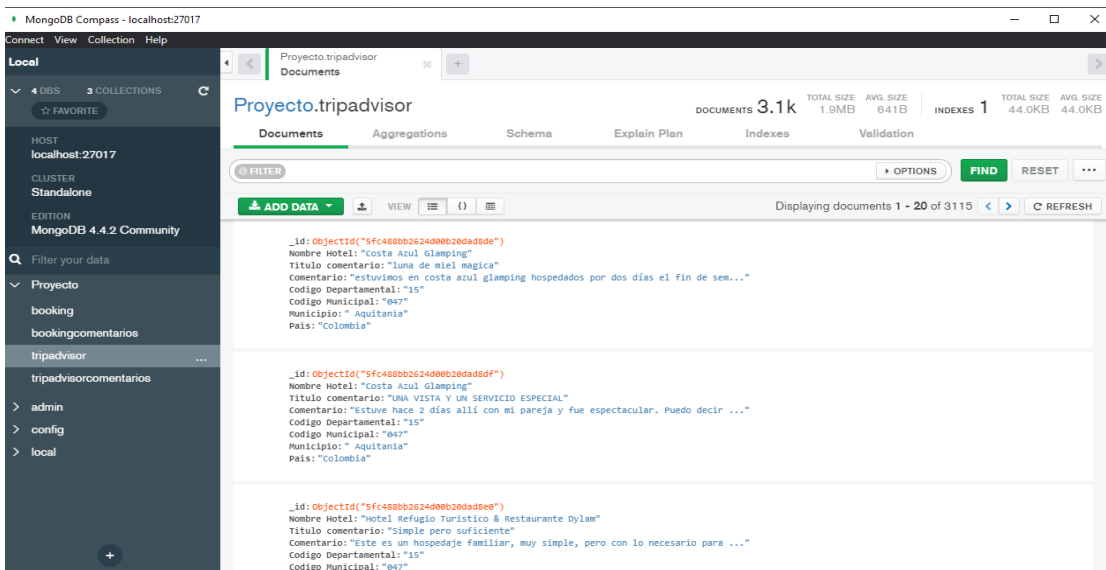
En la **Figura 34** se muestra la base de datos en mongoDB llamada “Proyecto”, mostrando la colección de “booking” donde se almacenan los datos extraídos después del filtrado.



**Figura 34 Colección de Booking en mongoDB.**

**Fuente: Autor.**

En la **Figura 35** se muestra la base de datos en mongoDB llamada “Proyecto”, mostrando la colección de “tripadvisor” donde se almacenan los datos extraídos después del filtrado.



**Figura 35 Colección de TripAdvisor en mongoDB.**

**Fuente: Autor.**

Los resultados del análisis de polaridad de texto, son mostrados en los archivos csv, tanto para la plataforma de TripAdvisor y Booking; solo después de la extracción.

En la **Tabla 3** se muestran los datos extraídos de la plataforma Booking con la polaridad correspondiente al frente del mismo comentario.

<b>Comentario</b>	<b>Polaridad</b>
“La vista. Y David es muy "amable."”1547Aquitania”	Positivo
“COMODA LA CABAÑA, MUY BUENA VISTA AL LAGO Y TIENES TOTAL "PRIVACIDAD."”1547Aquitania”	Neutro
“"Vista hermosa, me gustá mucho pasar una noche con mi esposo en este "lugar."”1547Aquitania”	Positivo
“excelente atención, con una muy bonita vista, acogedor y tranquilo para disfrutar con la "familia."”1547Aquitania”	Positivo

**Tabla 3 Resultados de análisis de polaridad de Booking, pero descargado en formato csv.**

**Fuente: Autor.**

En la **Tabla 4** se muestran los datos extraídos de la plataforma TripAdvisor con la polaridad correspondiente al frente del mismo comentario.

Comentarios	Polaridad
<p>“Situado a unos 4 km al sur-oeste de Aquitania, lago tota. Un edificio moderno, situado en un pequeño pueblo consta de un hotel, un restaurante, un bar y una tienda general en el nivel inferior. Una cálida bienvenida le espera a la llegada pronto seguido de bebidas calientes. El personal es muy amable y servicial. Las habitaciones en la 2<sup>d</sup>a planta son de tamaño pequeño y compacto pero muy limpio y ordenado, con baño en-suite y agua caliente y ducha. Hay una televisión en la mayoría de las habitaciones y Wi-Fi para el acceso a Internet. El bar en la planta 3 está ¡bien amueblada y diseñada. Las grandes ventanas ofrecen una excelente vista del lago. El comedor está ¡en la parte superior del hotel, en el piso 4 con grandes ventanas que ofrecen vistas del lago y el paisaje que le rodea.”</p>	Positivo
<p>“Estuvimos alojados una noche, la calma y tranquilidad que se siente en este hotel es única, su estilo colonial hace alarde de ello. El desayuno que viene incluido con la tarifa es lo necesario. Pensaría volver y disfrutar un poco más de él. Los jardines son únicos.”</p>	Positivo
<p>“El hotel queda bien ubicado si deseas hacer un recorrido por varios pueblos de Boyaca, Nobsa, Mongui, DUITAMA, Sogamoso, el viñedo de Puntalarga etc... Las habitaciones son limpias, amplias, calientes, con una decoración antigua. Tiene Direct tv, parqueadero y restaurante, el personal que trabaja en el hotel es muy amable.”</p>	Positivo
<p>“Definitivamente no podré volver.”</p>	Negativo

**Tabla 4 Resultados de análisis de polaridad de TripAdvisor, pero descargado en formato csv.**

**Fuente: Autor.**

Los resultados del análisis de polaridad de imágenes son dados en forma de cuadros para facilitar el análisis del rostro y los ojos.

En la **Figura 36** se muestra una imagen del resultado que arroja la polaridad de las imágenes que reconoce con rostro y ojos en el que enmarca con azul el rostro y los ojos con rojo.



**Figura 36 Resultado del reconocimiento de rostro y ojos.**

**Fuente: Autor.**

En la **Figura 37** se muestra una imagen del resultado que arroja la polaridad de las imágenes, pero esta vez con los resultados de las emociones que puede identificar en la imagen dependiendo del rostro y de los ojos.



**Figura 37 Reconocimiento de imagen para demostrar la diferente polaridad.**

**Fuente: Autor.**

La unificación de la polaridad de imágenes como de texto se hace por medio archivos csv, estos son archivados y catalogados para posteriormente ser agregados a mongoDB.

En la **Tabla 5** se muestra cómo se organizaron los comentarios con su debida polaridad por orden alfabético de los municipios de la plataforma de Booking.

<b>Comentario</b>	<b>Polaridad</b>	<b>Código Municipio</b>	<b>Municipio</b>
COMODA LA CABAÑA, MUY BUENA VISTA AL LAGO Y TIENES TOTAL PRIVACIDAD.	Positivo	15047	Aquitania
Es un sitio demasiado tranquilo con una vista espectacular	Positivo	15047	Aquitania
Muy tranquilo	Neutro	15466	Monguí
la vía algunos tramos son destapados y en mal estado	Negativo	15407	Villa de Leyva

**Tabla 5 Resultados de la polarización hecha por texto y por imágenes en un archivo csv separadas por municipio de Booking.**

**Fuente: Autor.**

En la **Tabla 6** se muestra cómo se organizaron los comentarios con su debida polaridad por orden alfabético de los municipios de la plataforma de Booking.

<b>Comentario</b>	<b>Polaridad</b>	<b>Código Municipal</b>	<b>Municipio</b>
Es un hotel de excelente calidad a nivel general, cómodas habitaciones con todas las facilidades, excelente atención del personal, muy buena la comida Muy recomendado tanto para viajeros empresariales como familiares.	Positivo	15238	Duitama
El personal del hotel es muy amable, el desayuno es delicioso, las habitaciones son muy confortables, es muy limpio, es nuevo y muy bien ubicado, no hay ruido, parqueadero gratuito, para el precio que pagamos fu una estadía increíble	Positivo	15238	Duitama
Volveré en cuanto pueda.	Neutro	15238	Duitama

Muy costoso en relación a lo que ofrece. Habitaciones con poco espacio. Problemas con el servicio de agua, no hay disponibilidad completa. Pocas posibilidades de descanso y privacidad en la noche, habitaciones con poco aislamiento.	Negativo	15407	Villa de Leyva
---	----------	-------	----------------

**Tabla 6 Resultados de la polarización hecha por texto y por imágenes en un archivo csv separadas por municipio de TripAdvisor.**

**Fuente: Autor.**

En la **Figura 38** se muestra cómo está el nombre en la base de datos de mongoDB la colección de booking para ser mostrada a los usuarios en el aplicativo web.

bookingcomentarios	23,876	216.8 B	4.9 MB	1	236.0 KB
--------------------	--------	---------	--------	---	----------

**Figura 38 Colección de Booking en la base de datos de mongoDB para la utilización en el sitio web.**

**Fuente: Autor.**

En la **Figura 39** se muestra cómo está el nombre en la base de datos de mongoDB la colección de Tripadvisor para ser mostrada a los usuarios en el aplicativo web.



**Figura 39** Colección de TripAdvisor en la base de datos de mongoDB para la utilización en el sitio web.

**Fuente: Autor.**

La visualización de datos está conectada a un aplicativo web que proporciona los datos necesarios a los usuarios para que hagan sus consultas.

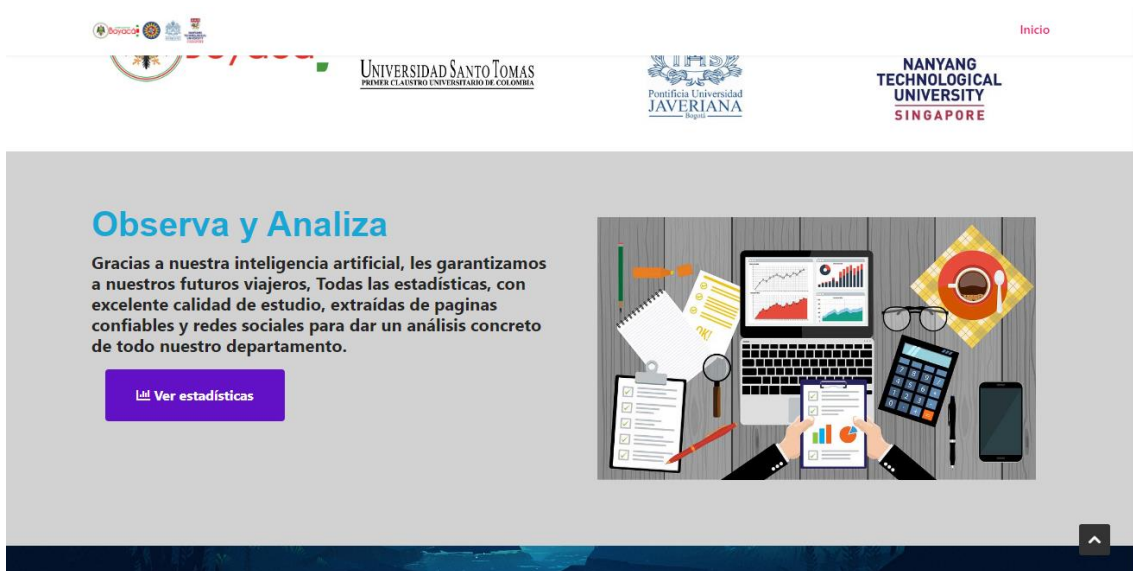
En la **Figura 40** se muestra la ventana inicial del aplicativo web que se va a mostrar al iniciar el servicio.



**Figura 40** Inicio del aplicativo web.

**Fuente: Autor.**

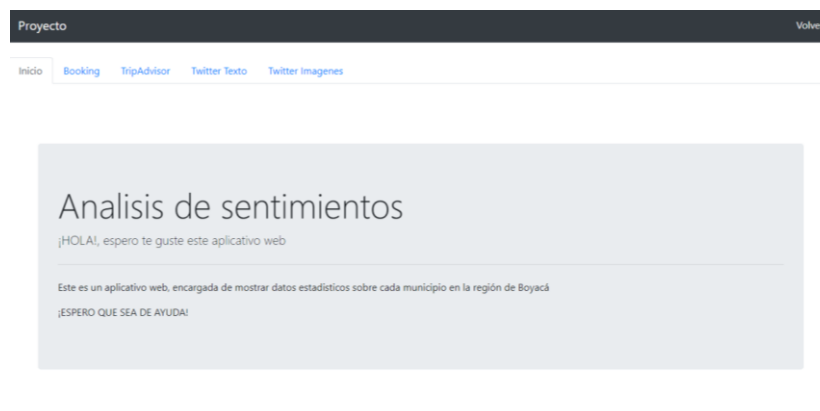
En la **Figura 41** se muestra el botón que dirige al aplicativo web de consultas para ver los resultados de extracción que se encuentra en color “morado” con el nombre de “Ver estadísticas”.



**Figura 41** Inicio del aplicativo web donde se ingresa a los datos.

**Fuente: Autor.**

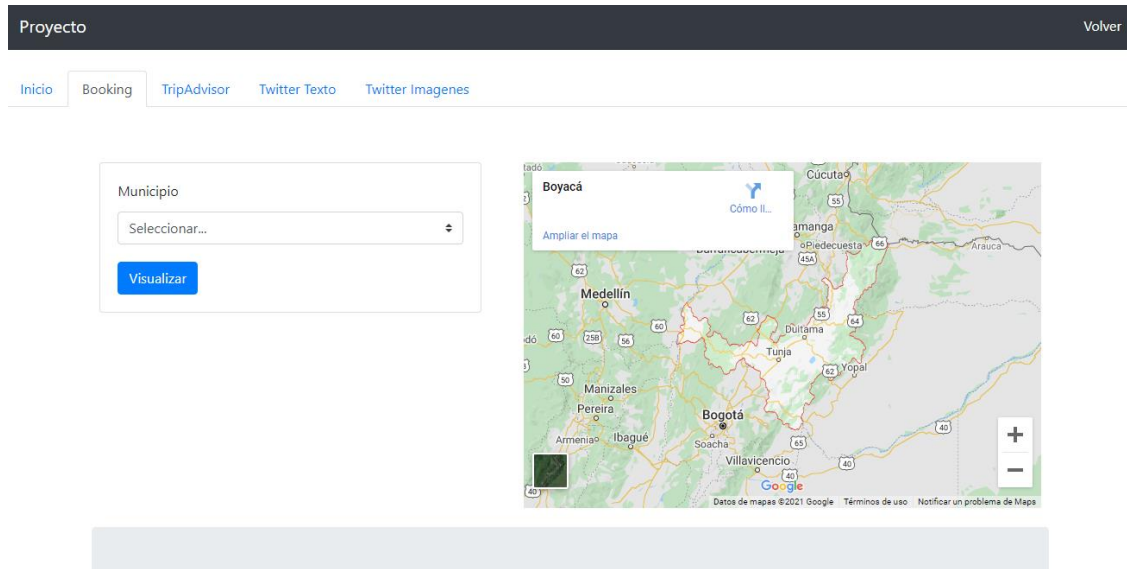
En la **Figura 42** se muestra el inicio del aplicativo web que mostrará el análisis de sentimientos a los usuarios interesados.



**Figura 42** Inicio del aplicativo web donde se mostrarán los datos extraídos.

**Fuente: Autor.**

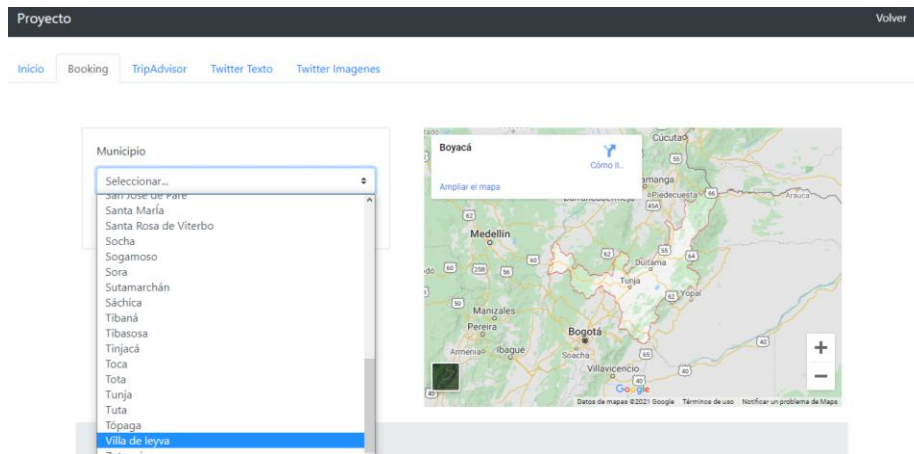
En la **Figura 43** se muestra la ventana de “Booking” que el usuario utiliza para la consulta de las estadísticas de esa plataforma.



**Figura 43 Muestra la ventana de “Booking” con las estadísticas que se dan por municipio.**

**Fuente: Autor.**

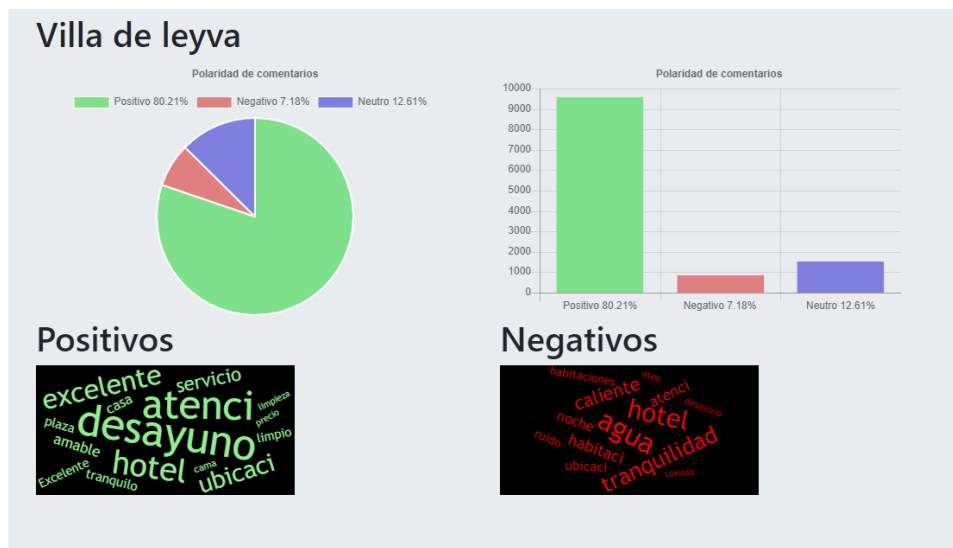
En la **Figura 44** se muestra la ventana de “Booking” que el usuario puede elegir el municipio para saber las estadísticas específicas del sitio.



**Figura 44 Muestra la ventana de “Booking” con las estadísticas que se dan por municipio.**

**Fuente: Autor.**

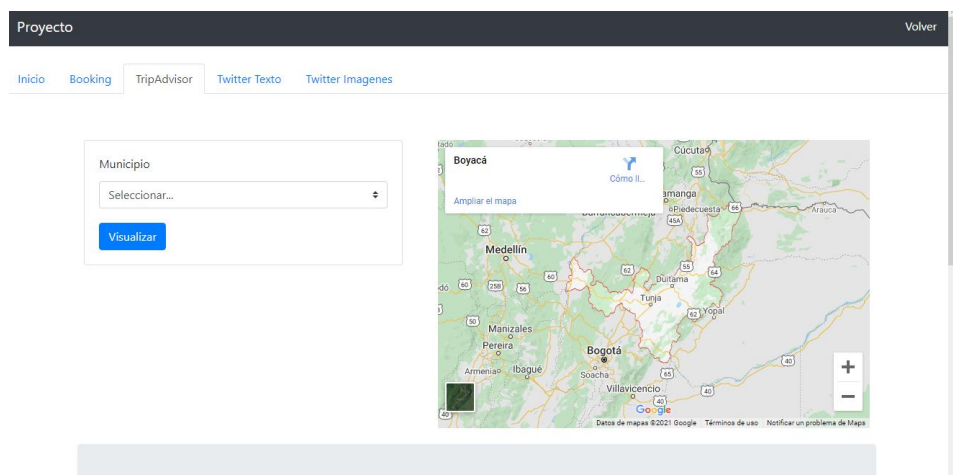
En la **Figura 45** se muestra la ventana de “Booking” con las estadísticas de un sitio seleccionado con las respectivas nubes de palabras positivas como negativas.



**Figura 45** Muestra la ventana de “Booking” con las estadísticas de un municipio en específico y con las nubes de palabras.

**Fuente: Autor.**

En la **Figura 46** se muestra la ventana de “Tripadvisor” que el usuario utiliza para la consulta de las estadísticas de esa plataforma.

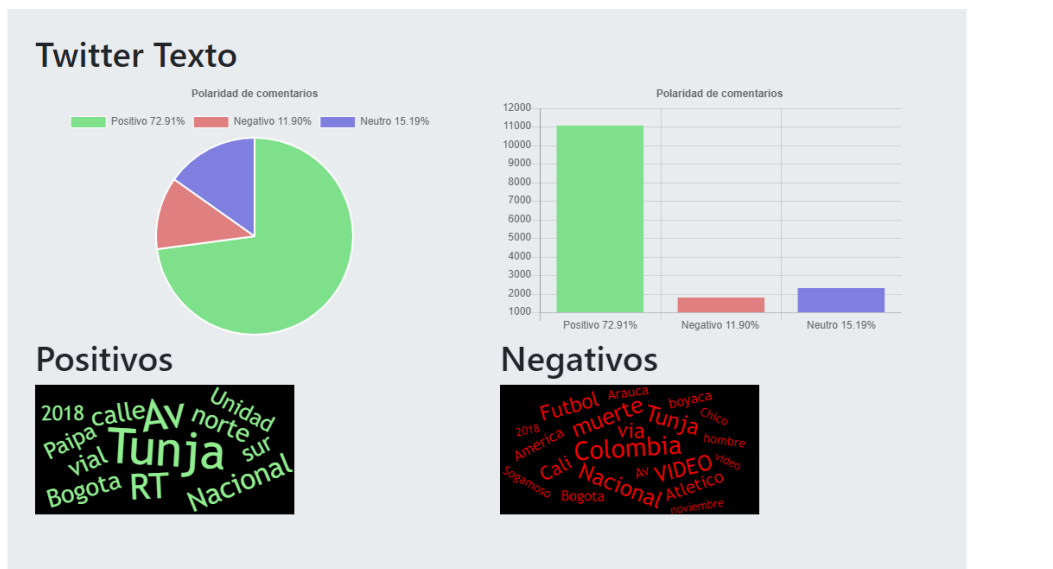


**Figura 46** Muestra la ventana de “Booking” con las estadísticas que se dan por municipio.

**Fuente: Autor.**



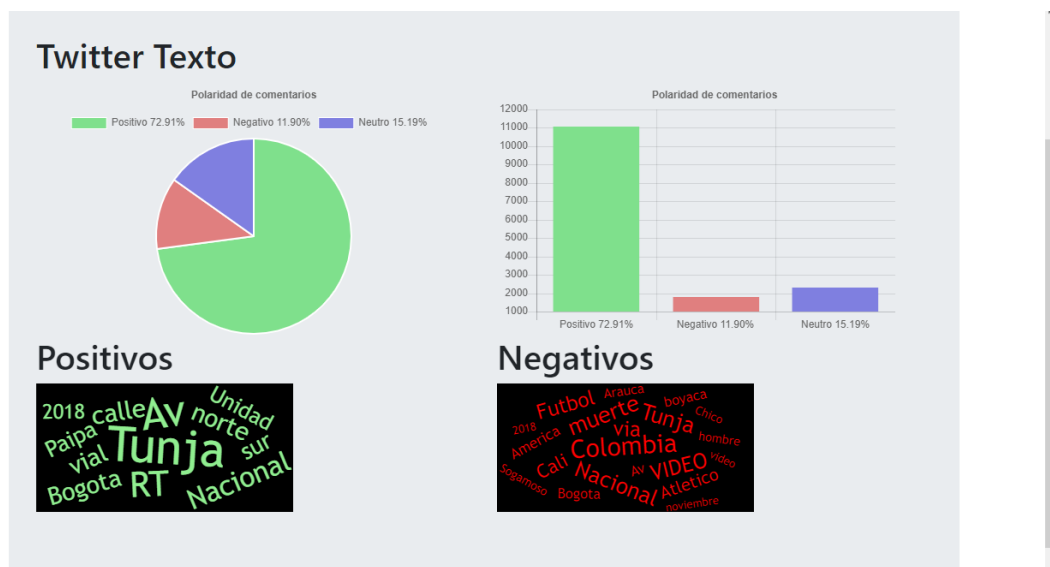
En la **Figura 49** se muestra la ventana de “Twitter Texto” con las estadísticas de Boyacá con las respectivas nubes de palabras positivas como negativas.



**Figura 49** Muestra la ventana de “Twitter texto” con las estadísticas de Boyacá y con las nubes de palabras.

Fuente: Autor.

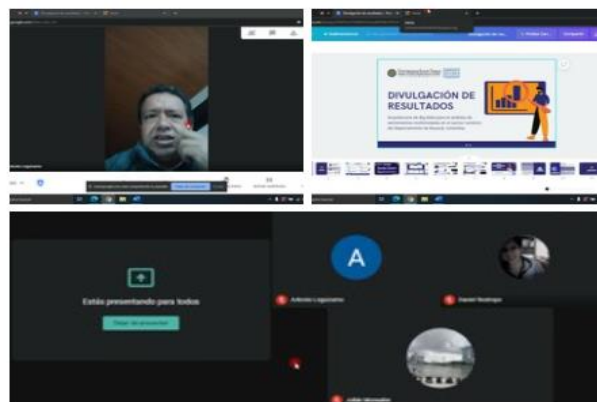
En la **Figura 50** se muestra la ventana de “Twitter Imágenes” con las estadísticas de Boyacá.



**Figura 50** Muestra la ventana de “Twitter texto” con las estadísticas de Boyacá.

Fuente: Autor.

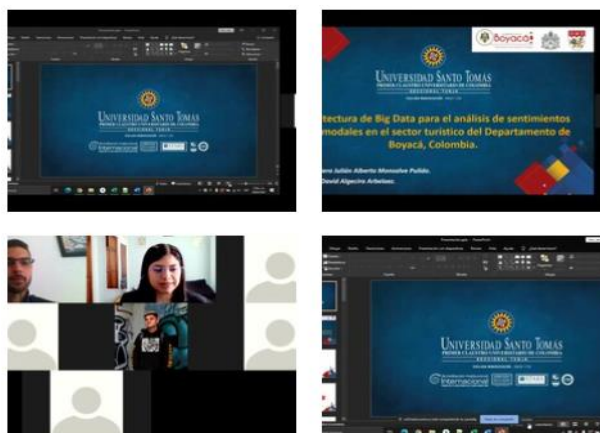
En la **Figura 51** se muestra la evidencia de la socialización en el sector empresarial turístico del departamento de Boyacá, con el acompañamiento del señor secretario de turismo del departamento de Boyacá Antonio Leguizamo Diaz, junto con un invitado y el director del proyecto el ingeniero Julián Alberto Monsalve Pulido.



**Figura 51 Muestra a los participantes del sector empresarial turístico del departamento de Boyacá.**

**Fuente: Autor.**

En la **Figura 52** se muestra la evidencia de la socialización en el sector académico y científico realizado en la Universidad Santo Tomas Sede Tunja, en la facultad de ingeniería de sistemas en la que asistió el ingeniero Iván Fernando Fonseca Barinas junto con un grupo de estudiantes y la ingeniera Jennifer Eliana Correa Ussa junto con un grupo de estudiantes.



**Figura 52 Muestra a los participantes del sector académico y científico de la Universidad Santo Tomas sede Tunja.**

**Fuente: Autor.**

## 7.CONCLUSIONES

En razón al estado del arte, la revisión documental permite esclarecer el rumbo del proyecto ya que como función propia de un estado del arte expone los diferentes puntos de vista de diferentes autores frente el tema a tratar, por lo tanto, este dio a conocer aspectos relevantes frente al análisis de sentimientos, la arquitectura de big data y el turismo experiencial.

Se integra los conocimientos con la revisión documental y el uso de los modelos con tal de que la arquitectura propuesta para el sector turístico del departamento contarán con los parámetros necesarios y específicos que permitan el análisis correcto de los datos suministrados como definitivos, esto a su vez esclarece el amplio flujo y concentración de información puesto que al hacer uso de la filtración y demás procesos como la polaridad de texto e imágenes la información final es concreta y comprensible.

Los resultados evidencian una arquitectura con una serie de estadísticas que permiten el análisis basado en polaridades (negativo, positivo y neutral) demostrando su utilidad frente a una percepción general de los usuarios en relación con las visitas turísticas dentro del departamento de Boyacá; Tres fases propuestas (extracción, análisis y visualización), dichas fases permiten que los resultados finales sean delimitados y comprensibles; así mismo permite organizar la información obtenida para su análisis definitivo.

De esta forma, las herramientas utilizadas para la realización del proyecto son de uso open source, con ello, no tenemos inconvenientes en utilizarlas de manera libre, además, el uso de las herramientas tanto de extracción como de análisis y visualización, fueron cruciales para poder hacer la interacción entre los distintos actores que dan un ordenamiento al proceso correspondiente, para comprenderlo y aprender de la práctica del mismo.

En relación a lo antes expuesto, los resultados de la polaridad en general, de las opiniones de los viajeros sobre Boyacá fueron en su gran mayoría positivas, eso indica que Boyacá puede

llegar a ser una potencia del turismo en Colombia, ya que cuenta con historia, paisajes, páramos, ríos y pueblos históricos que conforman una gran diversidad visual para el viajero.

## 8.BIBLIOGRAFÍA

- IngenioVirtual. (s.f.). *IngenioVirtual*. Obtenido de <https://www.ingeniovirtual.com/tipos-de-graficos-y-diagramas-para-la-visualizacion-de-datos/>
- 3.0, E. (30 de Septiembre de 2020). Obtenido de <https://www.educaciontrespuntocero.com/recursos/crear-una-nube-tags-las-palabras-mas-usadas-texto/>
- 300/1996, L. (1996).
- Ahmed, I. (27 de Noviembre de 2020). *Astera*. Obtenido de <https://www.astera.com/es/type/blog/automated-data-extraction-tools-for-faster-insights/>
- Akram Hernandez Vásquez, H. C. (2019). *MANIPULACIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS DE*. Obtenido de *MANIPULACIÓN, ANÁLISIS Y VISUALIZACIÓN DE DATOS DE*: <https://www.scielo.org/pdf/rpmesp/2019.v36n1/128-133/es>
- Antonio Manuel Rodriguez Garcia, F. R. (2019). *Píxel-BIT Revista de Medios y Educación*. Obtenido de *Píxel-BIT Revista de Medios y Educación*: [https://idus.us.es/bitstream/handle/11441/93761/Rodriguez-Garcia%20-%20competencia%20digital\\_removed.pdf?sequence=1&isAllowed=y](https://idus.us.es/bitstream/handle/11441/93761/Rodriguez-Garcia%20-%20competencia%20digital_removed.pdf?sequence=1&isAllowed=y)
- BBVA API\_Market. (11 de Enero de 2016). Obtenido de <https://www.bbvaapimarket.com/es/mundo-api/herramientas-de-extraccion-de-datos-para-principiantes-y-profesionales/>
- Bordignon, G. H. (s.f.). *Introducción a la Recuperación de Información*. Obtenido de <http://eprints.rclis.org/12243/1/Introduccion-RI-v9f.pdf>
- Boyacá Visión 2019. (2019).
- Carina Sofía Andrade and Maribel Yasmina Santos. (2017).
- Cesteros, A. F.-P. (2018). *La eficacia del análisis de sentimientos para la empresa: el caso de estudio Dell Technologies Inc*. MADRID. Obtenido de [https://www.ucm.es/data/cont/docs/758-2019-01-04-TFG\\_Panico\\_Chicara\\_TFG.pdf](https://www.ucm.es/data/cont/docs/758-2019-01-04-TFG_Panico_Chicara_TFG.pdf)
- clicdata. (s.f.). *clicdata*. Obtenido de <https://www.clicdata.com/es/ejemplos/project-management/>
- conceptodefinicion. (s.f.). *conceptodefinicion*. Obtenido de <https://conceptodefinicion.de/histograma/>
- datavizcatalogue. (s.f.). *datavizcatalogue*. Obtenido de [https://datavizcatalogue.com/ES/metodos/grafica\\_de\\_area.html](https://datavizcatalogue.com/ES/metodos/grafica_de_area.html)
- De Rojas y Camarero, 2008; Rivera,2013. (2013). Visitors experience, mood and satisfaction in A heritage context: Evidence from an interpretation center. *Tourism Management*, 525-537.
- Departamento Administrativo Nacional de Estadística. (s.f.).
- Divakar Mysore. (2013).

Divakar Mysore, S. K. (17 de Septiembre de 2013). *IBM*. Obtenido de <https://developer.ibm.com/es/technologies/analytics/articles/bd-archpatterns1/>

Dlib. (2020). *Dlib*. Obtenido de Dlib: <http://dlib.net/>

Earl, M. (28 de Julio de 2015). Obtenido de <https://matthewearl.github.io/2015/07/28/switching-eds-with-python/>

ecured. (s.f.). *ecured*. Obtenido de <https://www.ecured.cu/Cartograma>

Erik Cambria, A. H. (2012). Sentic Album: Content-, Concept-, and Context-Based Online. *Science+Business Media, LLC 2012*. Obtenido de <https://sentic.net/sentic-album.pdf>

Estadística, P. y. (27 de Mayo de 2012). *Probabilidad y Estadística*. Obtenido de <http://probabilidadyestadisticaitsav.blogspot.com/2012/05/16-diagrama-de-arbol.html>

Facultad de Estudios Estadísticos, Universidad Complutense de Madrid. (2020). Obtenido de <https://www.masterBigdataucm.com/que-es-Big-data/>

freepik. (2018). *freepik*. Obtenido de [https://www.freepik.es/vector-gratis/ilustracion-burbujas-discurso-nube\\_2904317.htm](https://www.freepik.es/vector-gratis/ilustracion-burbujas-discurso-nube_2904317.htm)

gamavision. (01 de Julio de 2020). *gamavision*. Obtenido de <https://www.gamavision.com.ec/pros-y-contras-del-reconocimiento-facial/>

García, S. M. (1997). *Psicosociología del ocio y el turismo*. Málaga: Algibe.

geeksforgeeks. (24 de Mayo de 2020). *geeksforgeeks*. Obtenido de <https://www.geeksforgeeks.org/opencv-facial-landmarks-and-face-detection-using-dlib-and-opencv/>

gisandbeers. (17 de Marzo de 2019). *gisandbeers*. Obtenido de <http://www.gisandbeers.com/crear-mapas-calor-heatmaps/>

Gómez, J. C. (10 de Enero de 2020). Obtenido de <https://josecantero.com/uncategorized/tendencias-en-turismo-experiencial-en-2020-bienvenidos-a-la-decada/>

HAA DnL. (s.f.). *HAA DnL*. Obtenido de <https://sites.google.com/site/haadnl/home/diagrama-radial>

Hong, M. K. (2017). *Unstructured Social Media Data Mining System Based on Emotional Database and Unstructured Information Management Architecture Framework*. ASP.

Jacinto, G. (1995). *Psicología social del turismo. Psicología Social de la Educación y de la Cultura, Ocio, Deporte y Turismo*. Salamanca, 245-272.

José Augusto Cadena Moreano, R. H. (2017). *RECONOCIMIENTO FACIAL CON BASE EN IMÁGENES*. B O L E T Í N V I R T U A L - M A Y O - V O L 6 - 5 I S N N 2 2 6 6 - 1 5 3 6. Obtenido de <file:///C:/Users/Usuario/Downloads/Dialnet-ReconocimientoFacialConBaseEnImagenes-6145639.pdf>

José Luis García Cué, J. A. (Octubre de 2008). *Revista Estilos de Aprendizaje*. Obtenido de <http://revistaestilosdeaprendizaje.com/article/view/852/1540>

Kanuk y Schiffman . (s.f.).

La Vanguardia. (11 de enero de 2020). Obtenido de <https://www.vanguardia.com/economia/nacional/turismo-hara-importante-aporte-en-el-pib-del-pais-XM1860653>

lucidchart. (s.f.). *lucidchart*. Obtenido de <https://www.lucidchart.com/pages/es/que-es-un-diagrama-de-flujo>

Martín, S. (1997).

Martinez, J. L. (08 de Junio de 2014). *slideshare*. Obtenido de <https://es.slideshare.net/joseluis199222/tabla-de-resultados-obtenidos-mecanismos-1>

McCannell. (1976). *The Tourist*. New York.

Mckercher. (1999).

McKercher, B. (1999). A chaos approach to tourism. *Tourism Management*, 425- 434.

Mérida, C. (2014). *Modelo arquitectónico de información para una plataforma de Big Data para el Sector Turístico*. Lima, Perú.

microsoft. (15 de Junio de 2016). *microsoft*. Obtenido de <https://docs.microsoft.com/es-es/sql/reporting-services/tutorial-add-a-pie-chart-to-your-report-report-builder?view=sql-server-ver15>

Molina et al. 2007; Troitiño, 1991. (1991). Segmentación de la demanda turística: Un análisis aplicado a un destino cultural. *Revista de análisis turístico n° 4*, 36-48.

Montaner, J. (1996). *Psicosociología del turismo*. Barcelona: Síntesis.

NodeJS. (2020). *NodeJS*. Obtenido de NodeJS: <https://nodejs.org/es/docs/>

npm ejs. (2020). *npm ejs*. Obtenido de npm ejs: <https://www.npmjs.com/package/ejs>

npm express. (2020). *npm express*. Obtenido de npm express: <https://www.npmjs.com/package/express>

npm mongoose. (2020). *npm mongoose*. Obtenido de npm mongoose: <https://www.npmjs.com/package/mongoose>

npm morgan. (2020). *npm morgan*. Obtenido de npm morgan: <https://www.npmjs.com/package/morgan>

npm nodemon. (2020). *npm nodemon*. Obtenido de npm nodemon: <https://www.npmjs.com/package/nodemon>

Numpy. (2020). *Numpy*. Obtenido de Numpy: <https://numpy.org/doc/>

Pandas. (2019). *Pandas*. Obtenido de Pandas: <https://pandas.pydata.org/docs/>

Paul Viola, M. J. (2001). *Rapid Object Detection using a Boosted Cascade of Simple Features*. Obtenido de <https://web.iitd.ac.in/~sumeet/viola-cvpr-01.pdf>

Pauli, P. A. (2019). *Análisis de sentimiento. Comparación de algoritmos predictivos y métodos utilizando un lexicon español*. BUENOS AIRES: INSTITUTO TECNOLÓGICO DE BUENOS AIRES – ITBA. Obtenido de <https://ri.itba.edu.ar/bitstream/handle/123456789/1782/Proyecto%20Final.pdf?sequence=1&isAllowed=y>

- Pearce, P. (1993). Fundamentals of Tourist Motivation. *Tourism Research: Critiques and Challenges*.
- Pine & Gilmore. (1998). The experience economy: Work is Theater & Every Business a Stage. *Harvard Business School Press*.
- practicobaul. (s.f.). *practicobaul*. Obtenido de <https://sites.google.com/site/unpracticobaul/tablas-en-un-texto>
- QuestionPro. (2020). *QuestionPro*. Obtenido de QuestionPro: <https://www.questionpro.com/es/analisis-de-datos.html#:~:text=El%20an%C3%A1lisis%20de%20datos%20es%20la%20ciencia%20que%20se%20encarga,los%20conocimientos%20sobre%20diversos%20temas.>
- Redacción España. (18 de Noviembre de 2019). *B12*. Obtenido de <https://agenciab12.mx/noticia/diferencia-datos-estructurados-y-datos-no-estructurados>
- Revuelta-Domínguez, F. I. (Diciembre de 2012). *researchgate.net*. Obtenido de [https://www.researchgate.net/figure/Figura-4-Representacion-grafica-de-la-red-segun-el-grado\\_fig1\\_266025222](https://www.researchgate.net/figure/Figura-4-Representacion-grafica-de-la-red-segun-el-grado_fig1_266025222)
- RICHARDS. (2004). *Diputació de Barcelona*. Obtenido de [www.diba.es/cerc/axinterac04/arcem1/richards/dipbarcelona](http://www.diba.es/cerc/axinterac04/arcem1/richards/dipbarcelona)
- Rober Morales Chaparro, J. C. (2012). *Desarrollo dirigido por modelos de visualización*. Obtenido de Desarrollo dirigido por modelos de visualización: <https://biblioteca.sistedes.es/wp-content/uploads/2016/05/JISBD-2012-565-Desarrollo-dirigido-por-modelos-de-visualizaci%C3%B3n-de-datos-para-la-Web.pdf>
- Sande, J. C. (2018). Análisis de sentimientos en Twitter. *UOC, Universitat Oberta de Catalunya*, 99. Obtenido de <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/81435/6/jsobrinotsTFM0618memoria.pdf>
- Schiffman, L. &. (1983). *Consumer behavior*. *Englewood Cliffs: Prentice Hall*.
- Scikit-Learn. (2019). *Scikit-Learn*. Obtenido de Scikit-Learn: <https://scikit-learn.org/0.21/documentation.html>
- SenticNet. (2018). *SenticNet*. Obtenido de SenticNet: <https://sentic.net/>
- support.minitab. (2020). */support.minitab*. Obtenido de <https://support.minitab.com/es-mx/minitab/19/help-and-how-to/graphs/matrix-plot/before-you-start/overview/themexpert.> (s.f.). *themexpert*. Obtenido de <https://www.themexpert.com/docs/quix/elements/polar-area-chart>
- TIBCO. (s.f.). *TIBCO*. Obtenido de [https://docs.tibco.com/pub/spotfire\\_web\\_player/6.0.0-november-2013/es-ES/WebHelp/GUID-6023CECC-E502-4AE1-B5C5-FFE5DAF6FAE2.html](https://docs.tibco.com/pub/spotfire_web_player/6.0.0-november-2013/es-ES/WebHelp/GUID-6023CECC-E502-4AE1-B5C5-FFE5DAF6FAE2.html)
- Tremolosa. (2002). ACTIVIDAD DE OCIO-TURÍSTICO. Universidad Alcalá. (s.f.).

- Universidad de Alcalá. (2020). Obtenido de <https://www.master-Bigdata.com/Big-data-actualmente-tan-importante/>
- Vasavi Gajarla, A. G. (s.f.). *Emotion Detection and Sentiment Analysis of Images*. Obtenido de [https://www.cc.gatech.edu/~hays/7476/projects/Aditi\\_Vasavi.pdf](https://www.cc.gatech.edu/~hays/7476/projects/Aditi_Vasavi.pdf)
- Vieira, D. (06 de Mayo de 2019). *rockcontent*. Obtenido de <https://rockcontent.com/es/blog/que-es-diagrama-de-gantt/>
- Zambrano, D. M. (2017). *Innovación para el Análisis de Sentimientos en Texto, una revisión de la técnica actual aplicando metodologías de crowdsourcing*. researchgate. Obtenido de file:///C:/Users/Usuario/Downloads/340-940-1-PB.pdf