



UNIVERSIDAD  
PRIVADA  
DEL NORTE

# FACULTAD DE INGENIERÍA

Carrera de Ingeniería de Sistemas Computacionales

“SOFTWARE DE ANÁLISIS DE SENTIMIENTO EN  
LA EVALUACIÓN DE LA SATISFACCIÓN DE LOS  
VISITANTES DEL SECTOR TURISMO DE  
TRUJILLO, 2018”

Tesis para optar el título profesional de:

Ingeniero de Sistemas Computacionales y  
Ingeniero de Sistemas

Autores:

Bach. Luis Alonzo Alvarez Linch  
Bach. Miguel Angel Rios Chacon

Asesor:

Ing. Pedro Gilmer Castillo Domínguez

Trujillo - Perú

2019

## DEDICATORIA

A Illuvar.  
A Luis Alberto, Eloisa y Diego Alvarez, por enseñarme las cosas más importantes,  
aquellas que no se aprenden en un aula.  
A Facundo, por cambiar la forma en la que veo todo.  
A Melisa, ¡Por más 達磨 con ojos completos!  
A Blanca y Daniel, por estar siempre ahí.

Luis Alonzo Alvarez Linch

A mis padres, Carlos Ríos y Rosa Chacón,  
por su entera confianza y disposición en esta etapa de mi vida.  
A Mirtha y Joel Arista, por su excelente apoyo incondicional.

Miguel Ángel Rios Chacon

## **AGRADECIMIENTO**

Al G.. A.. D.. U.., por darnos la vida, la salud, la tolerancia y la fortaleza para seguir adelante día a día acompañándonos a afrontar cada uno de los retos que nos presenta la vida.

Y en especial, por la colaboración directa en el desarrollo de este proyecto; a nuestro amigo y apoyo, el Mg. Geancarlo Leiva Vía, por brindarnos las herramientas necesarias para el desarrollo exitoso del mismo; por su exigencia, disposición, tiempo y amistad, y quien con sus conocimientos y paciencia ha sabido guiar todo el proceso de desarrollo de este trabajo.

## Tabla de contenidos

<b>DEDICATORIA .....</b>	<b>2</b>
<b>AGRADECIMIENTO.....</b>	<b>3</b>
<b>ÍNDICE DE TABLAS .....</b>	<b>5</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>6</b>
<b>CAPÍTULO I. INTRODUCCIÓN .....</b>	<b>9</b>
<b>CAPÍTULO II. METODOLOGÍA .....</b>	<b>61</b>
<b>CAPÍTULO III. RESULTADOS .....</b>	<b>78</b>
<b>CAPÍTULO IV. DISCUSIÓN Y CONCLUSIONES.....</b>	<b>83</b>
<b>REFERENCIAS .....</b>	<b>85</b>
<b>ANEXOS .....</b>	<b>89</b>
<b>1.1. Anexo 1: Presupuesto .....</b>	<b>89</b>
<b>1.2. Anexo 2: Política ambiental. ....</b>	<b>93</b>

## ÍNDICE DE TABLAS

<b>Tabla 1: Cuadro comparativo: Metodologías clásicas VS Metodologías ágiles.....</b>	<b>34</b>
<b>Tabla 2: Declaración del problema.....</b>	<b>62</b>
<b>Tabla 3: Caso de Uso 01 - Mantenedor Usuario.....</b>	<b>64</b>
<b>Tabla 4: Caso de Uso 02 - Revisar Reporte Mensual.....</b>	<b>67</b>
<b>Tabla 5: Caso de Uso 03 - Revisar Reporte Diario.....</b>	<b>67</b>
<b>Tabla 6: Descripción de la Tabla Calificar. ....</b>	<b>69</b>
<b>Tabla 7: Descripción de la Tabla Diccionario.....</b>	<b>69</b>
<b>Tabla 8: Descripción de la Tabla Polaridad. ....</b>	<b>70</b>
<b>Tabla 9: Descripción de la Tabla Usuario.....</b>	<b>70</b>
<b>Tabla 10: Descripción de la Tabla Reporte.....</b>	<b>71</b>
<b>Tabla 11: Polarización de publicaciones .....</b>	<b>83</b>
<b>Tabla 12: Presupuesto de Recursos Humanos.....</b>	<b>89</b>
<b>Tabla 13: Presupuesto Recursos Materiales.....</b>	<b>89</b>
<b>Tabla 14: Presupuesto de Servicios .....</b>	<b>90</b>
<b>Tabla 15: Presupuesto de Equipos.....</b>	<b>90</b>
<b>Tabla 16: Presupuesto de Software .....</b>	<b>91</b>
<b>Tabla 17: Consolidado del Presupuesto. ....</b>	<b>91</b>

## ÍNDICE DE FIGURAS

<b>Ilustración 1: Ciclo SCRUM .....</b>	<b>35</b>
<b>Ilustración 2: Modelo de caso de uso - Administrador .....</b>	<b>63</b>
<b>Ilustración 3: Modelo de Casos de Uso - Sistema .....</b>	<b>64</b>
<b>Ilustración 4: Diagrama Físico de la Base de Datos: AnalisisSentimiento.....</b>	<b>68</b>
<b>Ilustración 5: Diagrama de Arquitectura de Software .....</b>	<b>72</b>
<b>Ilustración 6: Proceso de Análisis de post.....</b>	<b>73</b>
<b>Ilustración 7: Ejemplo enlace post.....</b>	<b>74</b>
<b>Ilustración 8: Función SepararPalabras .....</b>	<b>74</b>
<b>Ilustración 9: Prototipo de Login.....</b>	<b>75</b>
<b>Ilustración 10: Prototipo de Página Principal. ....</b>	<b>76</b>
<b>Ilustración 11: Prototipo de Mantenedor Usuario. ....</b>	<b>76</b>
<b>Ilustración 12: Prototipo de Analizar Comentario .....</b>	<b>77</b>
<b>Ilustración 13: Publicación #1.....</b>	<b>78</b>
<b>Ilustración 14: Resultados obtenidos de la publicación #1. ....</b>	<b>79</b>
<b>Ilustración 15: Publicación #2.....</b>	<b>79</b>
<b>Ilustración 16 : Resultados obtenidos de la publicación #2. ....</b>	<b>80</b>
<b>Ilustración 17: Publicación #3.....</b>	<b>81</b>
<b>Ilustración 18: Resultados obtenidos de la publicación #3. ....</b>	<b>82</b>

## RESUMEN

El turismo es un gran eje para la economía mundial, esto no es diferente en el Perú y menos en la ciudad de Trujillo, que a particularmente cuenta con una ubicación geográfica privilegiada ya que a sus alrededores fueron ubicados los centros urbanos de 2 grandes culturas preincaicas. Además de esto, cuenta con una gran cantidad de museos dentro del centro histórico de la ciudad. Esta investigación está desarrollada en el ámbito del desarrollo de software, orientada a la mejora del proceso de retroalimentación para los centros turísticos de la ciudad de Trujillo realizando la captación y análisis de los comentarios vertidos en una Fan Page de Facebook por los visitantes. El presente trabajo de investigación se realizó con el objetivo de analizar la satisfacción de los visitantes del sector turismo en la ciudad de Trujillo en el 2018. El tipo de estudio fue del tipo trasversal. La recolección de datos se realizó a través de una Fanpage de Facebook y para el análisis de los mismos se utilizó el sistema de análisis de sentimientos. Esto permitirá conocer, en porcentajes, la satisfacción de los visitantes a los centros turísticos. Para culminar, aplicando el S.A.S se podrá conocer las fortalezas, oportunidades, debilidades y amenazas que posee cada centro turístico. Además, se logran aplicar políticas ambientalistas, eliminando la impresión de encuestas físicas logrando un ahorro en los ingresos económicos mensuales.

**Palabras clave:** Turismo, Análisis de sentimiento, Facebook API, Fan Page.

## ABSTRACT

Tourism is a great axis for the world's economy, this is no different in Peru and less in the city of Trujillo, which has a privileged geographical location and it is close from two of the largest pre-Incan culture's urban centers. In addition to this, Trujillo's historic downtown has a large number of museums within. This research is developed in the field of software development, aimed at improving the feedback process for the tourist centers of the city of Trujillo by capturing and analyzing the comments published on a Facebook Fan Page by the visitors. The present research work was carried out with the objective of analyzing the satisfaction of the visitors of the tourism sector in the city of Trujillo in 2018. The type of study is transversal. The data collection was carried out through a Facebook Fanpage and for the analysis of them the Sentiment Analysis System (S.A.S.), was used. This allows to know, in percentages, the satisfaction of the visitors to the tourist centers.

In conclusion, by applying the S.A.S you can know the strengths, opportunities, weaknesses and threats that each tourist center has. In addition, environmental policies are applied, eliminating the impression of physical surveys achieving savings in monthly economic income.

**Key words:** Tourism, Sentiment Analysis, Facebook API, Fan Page.



## CAPÍTULO I. INTRODUCCIÓN

### 1.1. Realidad problemática

El desarrollo turístico sustentable es un sector que se ha vuelto preponderante en el desarrollo económico de los países, donde este desarrollo debe cumplir con tres criterios: la mejora de las condiciones de vida de las comunidades receptoras, la satisfacción de los turistas, y el cuidado del territorio, tanto del medio ambiente natural como del patrimonio cultural, que construyeron las comunidades que allí han vivido. Esto considerado en términos económicos, debe ser un negocio, en el cual todos los actores se beneficien en función de su aportación al proceso. (Orgaz Agüera & Moral Cuadra, 2016).

La importancia que tiene el sector turismo como eje económico para los países en el mundo, el cual es un contribuyente muy alto al porcentaje de la economía mundial, y el sector turismo funciona como cualquier empresa, cuidando lo que ofrece y deseando tener siempre a sus clientes satisfechos. De acuerdo con Shirley Vindas (2005), en su informe El sector turismo: Su aporte a la economía de, en la actualidad, el mayor atractivo de los turistas ha sido visitar culturas ancestrales de otros países, cada país del mundo tiene algo distinto que mostrar. El turismo es conocido como el fenómeno social, cultural y económico ya que beneficia y enriquece en grandes magnitudes estos aspectos. (Vindas, 2005)

Actualmente, la OMT, mostró que, hasta la fecha, los resultados de 2018 están en consonancia con las previsiones de crecimiento para el año, que oscilaban entre el +4% y el +5. La OMT estima que los destinos de todo el mundo recibieron 1.083 millones de llegadas internacionales hasta septiembre, es decir, 56 millones más que durante el mismo período en 2017. (Organización Mundial de Turismo, 2018)

En el artículo publicado en el semanario Semana de Colombia, Edilma Pereira nos indica que el potencial de Colombia es enorme en materia de turismo. El año pasado fue el segundo renglón en generación de divisas para el país, con 5.787 millones de dólares, por encima de las remesas de colombianos en el exterior, que alcanzaron los 5.495 millones de dólares. Pero para la ministra de Comercio, Industria y Turismo, María Lorena Gutiérrez, el turismo puede ser el primer renglón de la economía. “Es un sector que genera mucho empleo, cerca de 1,9 millones en estos años”. (Pereira, 2018).

En el Perú, el turismo constituye una de las principales industrias del país, junto con la pesca y la minería, según un estudio del MINCETUR el PBI Turístico representa el 3,9% del PBI total del país, (Ministerio de comercio exterior y turismo, 2016)

El turismo ha sido y es uno de los rubros con gran rentabilidad en Trujillo y esto es gracias a la ubicación geográfica de nuestra ciudad. Este gran movimiento de personas extranjeras y nacionales, hacia nuestra ciudad, ha logrado la puesta en valor de muchos monumentos históricos, logrando descubrimientos que suman importancia y a la vez interés en los turistas.

El Análisis de sentimientos, extrae efectivamente información subjetiva, opiniones de personas en datos no estructurado, del medio de preferencia para, posteriormente, analizar esta información y mostrar datos estadísticos de ellos. Utilizando técnicas de análisis de sentimientos podremos obtener y analizar las opiniones de los visitantes del sector turismo de la ciudad de Trujillo.

A pesar de la existencia de muchos proyectos arqueológicos, y que éstos son subvencionados por los usuarios visitantes, muy poco interés existe al recopilar información sobre sugerencias de los mismos usuarios, esto se debe a la falta de herramientas tecnológicas, software especializado, que permita la fácil recopilación y proceso de estas opiniones.

Se han considerado los siguientes estudios, trabajos e investigaciones como antecedentes:

El autor Neftalí de Jesús Calderón (2006), en la investigación, Minería de datos una herramienta para la toma de decisiones, Calderón Méndez nos indica que el objetivo fue evaluar el uso de la Minería de Datos como una herramienta que sirva para la toma de decisiones a nivel gerencial. Método de investigación explicativa donde llegó a la conclusión de que se determinó que la minería de datos es el conjunto de herramientas y técnicas de análisis de datos que permiten crear escenarios, de los cuales se puede obtener información útil para la toma de decisiones a nivel gerencial. Las técnicas que utiliza la minería de datos para la exploración consisten la identificación de patrones. El proceso de la minería de datos genera conocimiento por medio de la depuración, enriquecimiento y transformación de datos que sirve para la creación de un modelo en el que se evalúa un conjunto de casos. (Calderón Méndez, 2006).

Los autores Antonio Moreno Ortiz, Francisco Pineda Castillo y Rodrigo Hidalgo García (2010), en el informe Análisis de Valoraciones de Usuario de Hoteles con Sentitext\*: un sistema de análisis de sentimiento independiente del dominio se indica que, Sentitext es una aplicación desarrollada en C++ bajo la arquitectura

cliente-servidor con el objetivo de facilitar la comunicación con el analizador morfológico, el cuál es nutrido de las siguientes fuentes de datos fundamentales:

1. Léxico de palabras individuales.
2. Léxico de frases.
3. Reglas de contexto.

Para esto, Sentitext utiliza una base de datos MySQL editable a través de una plataforma web, para la adecuación de nuevas palabras o una edición de la polaridad de las palabras contenidas. Esta base de datos contiene una fuente de palabras polarizadas denominada “Lexicón”, el cual cuenta con 10.500 palabras polarizadas con uno de estos 4 valores: -2 (muy negativa), -1 (negativa), 1 (positiva), 2 (muy positiva). Esta clasificación fue realizada de manera consensuada por un equipo de lexicógrafos de tecno lengua, quienes se basaron en rasgos semánticos y recurriendo a un corpus de lengua general. Uno de los puntos que resaltan los desarrolladores es la ausencia del dominio al momento de realizar el análisis de los comentarios, por dominio se refieren a el contexto que puede brindar una alteración en la polarización de ciertas palabras, por ejemplo, en el rubro hotelero mencionar una habitación pequeña puede tener una connotación negativa que en otros dominios puede ser inocua. (Moreno Ortiz, Pineda Castillo, & Hidalgo García, 2010).

La autora Luciana Dubiau (2013), en su informe; Procesamiento de Lenguaje Natural en Sistemas de Análisis de Sentimientos, tiene como objetivo la investigación, análisis y comparación de técnicas de procesamiento de lenguaje natural para la clasificación de documentos a partir de la identificación y extracción de información subjetiva como opiniones, sentimientos y emociones. Método de

investigación experimental donde llegó a la conclusión de que las experiencias presentadas muestran el impacto en la performance de los clasificadores ante la variación de parámetros de entrada como transformaciones de textos, tipos de atributos extraídos, conjunto de datos de entrenamiento, conjunto de datos de prueba, tamaño de corpus utilizado y proporción de documentos de cada clase. Hallamos que los modelos supervisados de Naive Bayes, MaxEnt y SVM resultan muy apropiados para clasificar información subjetiva en idioma español alcanzando hasta un 95% de accuracy para los tamaños de corpus más grandes mientras que técnicas basadas en arboles de decisión no proveen buenos resultados. Además, adaptamos la técnica de clasificación no supervisada de Turney al idioma español y obtuvimos excelentes resultados considerando que no se requiere información previa ni un conjunto de datos de entrenamiento. Mostramos que la aplicación de reprocesamientos influye considerablemente. (Dubiau, 2013).

El autor Isidro Peñalver Martínez (2015), en el informe de Minería de opiniones basada en características guiadas por ontologías, tiene como objetivo mostrar la distinción entre las 2 metodologías principales en la minería de datos. La primera tiene como objetivo clasificar documentos enteros con una polaridad que puede ser negativa o positiva. La segunda aproximación se encarga de distinguir frases con algún sentimiento dentro de los textos de opinión del usuario.

Se resalta el impacto generado y la importancia de la interacción de los usuarios en la Web 2.0, mientras que los usuarios de la Web 1.0 sólo eran consumidores del contenido compartido por las compañías. Desde los inicios de la Web 2.0 se ha obtenido contenido en forma de opiniones que no sólo son útiles para los proveedores sino también para los mismos consumidores. Estas opiniones son

importantes cuando alguien desea obtener información para tomar una decisión, esto ha logrado que la minería de opinión, que es una subdisciplina que trata de detectar los sentimientos expresados por el usuario. Esta información es obtenida de comentarios en foros online, blogs o redes sociales. De acuerdo con lo propuesto, se describe una nueva metodología que utiliza una ontología del dominio para identificar características. Se aporta un algoritmo basado en análisis vectorial en el espacio  $R^3$  para realizar el proceso de análisis de sentimiento de los comentarios. (Peñalver Martínez, 2015).

Los autores Javi Fernández, Yoan Gutiérrez, José M. Gómez y Patricio Martínez-Barco (2015), en su informe titulado, Social Rankings: análisis visual de sentimientos en redes sociales, es un informe técnico que tiene objetivo explicar las características de la aplicación web Social Rankings. Esta aplicación ha sido desarrollada en conjunto con la Sociedad Española para el Procesamiento de Lenguaje Natural. Social Rankings permite detectar y analizar opiniones sobre entidades mediante técnicas de análisis de sentimiento en comentarios o publicaciones realizadas en redes sociales previamente registradas, luego se deben establecer las entidades y palabras clave. Este producto software puede ser aplicado únicamente en Twitter, debido a las facilidades de manipulación y disponibilidad que otorga la API de esta red social. Para realizar el análisis Social Rankings utiliza una combinación de dos sistemas, el primero es basado en aproximaciones de (Balahur Dobrescu, 2011), este sistema usa diccionarios de palabras que contienen sentimientos obtenidos de recursos de opinión, tales como; WordNet Affect o SentiWorldNet. La polaridad del texto será obtenida del paso de las palabras examinadas. El segundo sistema está basado en la aproximación híbrida de

(Fernández, Gómez, & Martínez-Barco, 2014) . En esta aproximación, un diccionario es generado a partir de un corpus, que contiene palabras con pesos de positividad y negatividad. (Fernández, Gutiérrez, Gómez, & Martínez-Barco, 2015)

Los autores, Sonia Jaramillo Valbuena, Sergio Augusto Cardona y Alejandro Fernández (2015), en su informe Minería de datos sobre *streams* de redes sociales, una herramienta al servicio de la Bibliotecología tiene como objetivo el desarrollo de un algoritmo que permita buscar ciertas palabras en los grupos de Facebook seleccionados. La primera búsqueda fue realizada en el 2008 y se obtuvieron los siguientes datos con las palabras escogidas. Los resultados fueron los siguientes: bibliotecología (28), documentación (145), archivística (16), ciencia de la información (77) y biblioteconomía (11). La misma evaluación fue realizada en abril del 2015 obteniendo los siguientes datos: bibliotecología (127), documentación (240), archivística (62), ciencia de la información (120) y biblioteconomía (108). Esta información obtenida tiene coherencia con la tendencia marcada por la Web.2.0 de compartir información. Las interacciones de los usuarios de las redes sociales generan grandes flujos de datos no estructurados denominados *streams*, los cuales tras procesos de minería de datos y de análisis pueden producir conocimiento con aspectos como influencia, relaciones de cercanía, opinión y generación de recomendaciones. Se comprueba que, a partir de texto no estructurado, se puede obtener conocimiento, que de acuerdo con la aplicación de un análisis puede considerarse como positivo o negativo para que, en base a ello, puedan aplicarse acciones correctivas. (Jaramillo Valbuena, Cardona, & Fernández, 2015)

El autor, Martín Becerra (2016), en su paper titulado: Análisis de sentimiento en Twitter: El bueno, el malo y el >:(, nos indica que su objetivo es identificar opiniones positivas o negativas en textos generados por usuarios y sobre qué entidad o aspecto de la misma se han realizado. Método de investigación experimental donde llegó a la conclusión de que, en este trabajo, hemos estudiado un problema complejo como es el de la minería de opiniones en Twitter y los desafíos asociados al análisis exploratorio de los resultados de dicho proceso. Si bien Twitter tiene ciertas limitaciones a la hora de proveer información, es más que suficiente para llevar a cabo un análisis exploratorio con la finalidad de comprender mejor un mercado o un acontecimiento. Para esto nos hemos centrado en diseñar un pipeline en el cual adquirimos, pre-procesamos, hacemos clustering y finalmente vemos representados de forma visual un conjunto de tweets sobre un determinado tópico. En cada etapa de este pipeline identificamos conjuntos de problemas y diferentes soluciones que se podían llevar adelante para cada uno de ellos. Si bien este trabajo propone una implementación concreta para este flujo, existe la posibilidad de continuar investigando en mayor profundidad de manera independiente cada etapa. (Becerra, 2016)

El autor Hugo Jacobo Librado (2017), en su informe, Análisis automático de opiniones de productos en redes, tiene como objetivo el de desarrollar una aplicación de software capaz de analizar textos cortos del idioma español extraídos de la red social Twitter y determinar la polaridad de cada documento. Método de investigación experimental donde llegó a la conclusión de que el trabajo desarrollado aporta una aproximación diferente en cuanto a la metodología para la selección de características y el establecimiento de pesos a cada termino, ya que la



mayoría de las aproximaciones considera a las palabras como unipolares, es decir, que solo aportan información hacia una dirección (positiva o negativa), en diferentes intensidades y generalmente invertida cuando se topan con una negación. Sin embargo, nuestra propuesta considera a las palabras multipolares o mutilase, por lo que una palabra positiva también puede aportar información negativa, excluyendo el uso de negaciones que pueden invertir polaridad. Por otra parte, la metodología para calificar los textos se basa en un conjunto de reglas de clasificación que básicamente conforman un árbol de decisión, considerando que los términos que se involucran son multipolares. (Librado Jacobo, 2017)

Los autores, Carlos Henríquez Miranda, Jaime Guzmán y Dixon Salcedo (2016) , en el informe, Minería de Opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles, tiene como objetivo principal, mostrar la importancia que tiene los comentarios u opiniones publicadas en internet por los usuarios de diferentes ámbitos para los administradores o responsables de negocios y empresas. Se habla, además, del grado de dificultad y dedicación que tomaría realizar estos análisis de manera manual y esto se debe a las dimensiones que pueden tener estas opiniones, tales como: valencia, excitación y dominio. El equipo de trabajo resalta el interés que se tiene en generar productos que realicen minería de opinión o analizadores de opinión, términos que denotan el mismo campo de estudio, debido a que la mayoría de los trabajos de investigación sobre el tema son realizados en el idioma inglés. El producto software desarrollado es una adaptación de ANEW, un analizador de opiniones en inglés, esta adaptación consistió en la traducción al español de 1034 palabras realizadas por un filólogo profesional especializado en el idioma inglés. La valencia de estas palabras fue basada en el

informe de (Salas-Zárate , Paredes-Valverde , Limon-Romero , Tlapa , & Baez-Lopez , 2016) donde se demuestra la eficacia de la clasificación en cinco categorías: muy positiva, muy negativa, positiva, negativa y neutra. La metodología utilizada para realizar el procesamiento de los textos no estructurados consiste en un modelo de 4 fases: extracción, preprocesamiento, identificación de sentimientos y clasificación de la opinión. (Henriquez Miranda, Guzmán, & Salcedo , 2016)

La autora, María Narváez (2017), en su informe, Aplicación de técnicas de minería de datos para el análisis de sentimientos en las redes sociales sobre productos de la marca Fideos Cayambe, tiene como objetivo realizar un estudio de opiniones mediante técnicas de minería de datos adecuadas para la explotación de cierta información, por ejemplo, la almacenada en las redes sociales. En este proyecto, precisamente, se trata de analizar la información recolectadas de Facebook a través de la aplicación de la metodología CRISP-DM (CRoss Industry Estándar Process for Data Mining) y además utilizando técnicas como: clustering, regresión lineal y árboles de clasificación. Luego, se realiza una comparación de los medios utilizados, no hace mucho, para publicitar productos, por ejemplo, la televisión o los carteles, a través de estos medios se llegaba al público consumidor, pero se desconocían las opiniones de éstos o de qué manera es recibida esta publicidad. Luego se resalta la importancia que tienen actualmente las redes sociales y la oportunidad de implementar nuevas estrategias de marketing tras un análisis de los criterios y opiniones de los consumidores que se encuentran almacenadas en los comentarios de estas redes, estas opiniones pueden ser positivas, negativas o neutrales. Se concluye que utilizando las técnicas de árboles para la toma de decisiones y clustering se obtiene un menor porcentaje de error al momento de

polarizar y clasificar los comentarios de Facebook. Tras este análisis se determina las fortalezas y debilidades que posee el producto a la vista u opinión del usuario. (Narváez Montoya , 2017)

Los autores, Manuel Vilares, Elena Sánchez Trigo, Carlos Gómez-Rodríguez y Miguel A. Alonso (2017), en el informe Tecnologías de la lengua para análisis de opiniones en redes sociales, tiene como objetivo el resaltar la importancia de estos tiempos en cuestiones de a interacción de los usuarios con las páginas web. Las opiniones o comentarios que vierten los usuarios nos solo contienen lo que el usuario piensa sobre un producto sino también se realizan comparaciones entre productos similares de otras empresas esto permite una visión importante para el ámbito comercial y administrativo de las empresas ya que conocerán las fortalezas que deben ser potenciadas y debilidades en las que deben trabajar. La Minería de opinión (MO) permite que se conozca la polaridad de ciertas palabras, frases o textos completos, pero a la vez se encuentra muy limitada por su escaso recurso de tecnologías de la lengua ya que se basa en un proceso de análisis superficial sin tener en cuenta la relación sintáctica de las palabras y sus roles semánticos en las oraciones. Además, debido a la centralización de las empresas en países angloparlantes, la mayoría de estas soluciones web se encuentran desarrolladas y orientadas al inglés como idioma base. En este contexto, se desarrolla la solución TELEPARES (Tecnología de la lengua para análisis de opiniones en redes sociales), un proyecto elaborado en conjunto con diferentes instituciones como: La ONG Grupo COLE de la Universidad de Vigo, La ONG Grupo LYS de la Universidad de La Coruña y del CITIUS de la Universidad de Santiago de Compostela; quienes han obtenido un financiamiento del Ministerio de Economía y

Competitividad dentro del programa estatal de I+D+i. Los riesgos que corre el equipo de investigación son descritos en el informe *How People Use Twitter in Different Languages* (Weerkamp, Carter, & Tsagkias, 2011), y son los siguientes:

(1) El uso de micro textos carentes de contexto lingüístico. (2) El ruido generado por los usuarios al momento de utilizar escrituras no convencionales, uso irregular de mayúsculas y minúsculas, abreviaciones idiosincrásicas, uso irregular de signos de puntuación, eliminación o uso de palabras abreviadas para no superar el tamaño máximo permitido en un tuit (micro texto de Twitter), uso de emoticones, uso de ironía. (3) El multiengüismo, ya que al menos el 50% del contenido de Twitter se encuentra escrito en inglés, con una presencia relevante y creciente del español, portugués y japonés. (Vilares, Sánchez Trigo, Gómez-Rodríguez, & Alonso, 2017).

La autora Meysi García Ruesta (2011), en el informe *Medición de la satisfacción del cliente en una empresa de retail* tiene como objetivo el medir y controlar el nivel de satisfacción de los clientes de una empresa de retail, exactamente en el área electro. Método de investigación explicativa donde llegó a la conclusión de que la satisfacción general. En cuanto a este artículo de satisfacción, la cual se realizó para analizar la apreciación general de la calidad del servicio brindado por la empresa, se concluye que no se puede hablar de un control bueno de la satisfacción del cliente, a pesar de que los valores se encuentran en un proceso bajo control, en los límites entre niveles de ni satisfacción ni insatisfacción y satisfacción. Estos límites no son los deseados para decir que la empresa mantiene un estándar de la calidad del servicio que brinda. Esto se confirma con el análisis de la variabilidad de las muestras; por tanto, se concluye que los clientes no se

encuentran satisfechos con el servicio brindado por la empresa. (García Ruesta, 2011).

El autor, Miguel Angel Grández Márquez (2017), en el informe; Aplicación de la minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos, es un informe que tiene como objetivo encontrar reglas de consumo que determinen patrones de consumo futuro en una distribuidora de suplementos nutricionales con el fin implementar políticas que incrementen el nivel de ventas. (Grández Márquez, 2017).

La autora Juanita Blanco Pineros (2009), en el informe Medición de la satisfacción del cliente del restaurante museo Taurino y formulación de estrategias de servicio para la creación de valor, se tiene como objetivo principal el medir la satisfacción del cliente. Para lograr este objetivo se trazan objetivos específicos como: Analizar la situación actual del restaurante en cuanto al servicio. Diseñar instrumentos para medir la satisfacción actual de los clientes. Definir y evaluar el ciclo del servicio a través de los momentos de verdad. Formular estrategias de servicio para la creación de valor. La metodología empleada en esta investigación se dividió en dos partes, una la observación, que se empleó para analizar el tipo de clientes que frecuentan el restaurante y los tiempos que demoran en ser atendidos. Posteriormente, la metodología descriptiva, la cual permitió evaluar características fundamentales del restaurante por medio de una encuesta aplicada al cliente y obtener resultados finales que muestran cómo es que el cliente se siente con respecto a la atención. Se

realizaron 62 encuestas durante los días lunes, martes, miércoles, jueves y viernes, siendo jueves el día en el que más clientes asisten. Los resultados fueron favorables para el restaurante en cuanto a Rapidez de atención, comodidad con el ambiente, limpieza y eficiencia del personal, siendo sólo en promedio un 8% de respuestas negativas. (Blanco Pineros, 2009).

La autora, Susan Camila Ballón Cervantes (2016), en el informe Evaluación en la calidad de servicio y satisfacción del cliente en los hoteles de dos y tres estrellas de la ciudad de Abancay, tiene como principal objetivo el analizar las variables de la calidad de servicio que influyen en la satisfacción al cliente. Se cuenta con los objetivos específicos de identificar las variables de la calidad de servicio que influyen en la satisfacción al cliente en los hoteles de dos y tres estrellas de la ciudad de Abancay. Determinar las variables de la calidad de servicio que tienen mayor influencia en la satisfacción al cliente en los hoteles de dos y tres estrellas de la ciudad de Abancay. Se tomó una muestra de 431 turistas, a los cuales se les aplicó una breve encuesta, en los cuales los resultados fueron altamente positivos en cuanto atención al cliente afirmando las hipótesis de dicha investigación. (Ballón Cervantes, 2016).

En el informe del autor Hernán Toniut (2013), titulado, La medición de la satisfacción del cliente en supermercados de la ciudad de Mar del Plata, se muestra como el objetivo principal la identificación de las variables que el cliente usa para tomar una decisión sobre un producto y además con la medición de la percepción del cliente en relación con el negocio en el que realizan las compras. La investigación exploratoria se realizó por medio de entrevistas a personas con más de 5 años de experiencia de trabajo en el formato de negocio sujeto a análisis. El

objetivo de este análisis fue identificar las variables consideradas por el cliente al realizar una compra. La investigación descriptiva fue dividida en dos partes, en primer término, se hizo un relevamiento de las expectativas de los clientes que permitió valorar la importancia de las variables que influyen en la decisión. En segundo lugar, los clientes en relación a los supermercados donde realizan sus compras. Como resultado se obtuvo que las variables que influyen en la decisión de los clientes son el precio, las ofertas, horarios, calidad de atención y métodos de pago. Los clientes creen que es importante la medición de la satisfacción, sin embargo, actualmente no se mide de forma sistemática, sólo a través de elementos genéricos como lo son las quejas, libro de reclamaciones, etc. Para la muestra se tomaron a 270 clientes de supermercados mayores de 18 años. Como conclusión, se determinó la preferencia de los clientes al realizar sus compras en un supermercado, donde un 67.7% de los casos prefieren comprar sus alimentos por este medio, mientras que en otro rubro como la perfumería el porcentaje fue de 56.38%, en productos de limpieza 77.58% y lácteos y fiambre 32.73%. (Toniut, 2013).

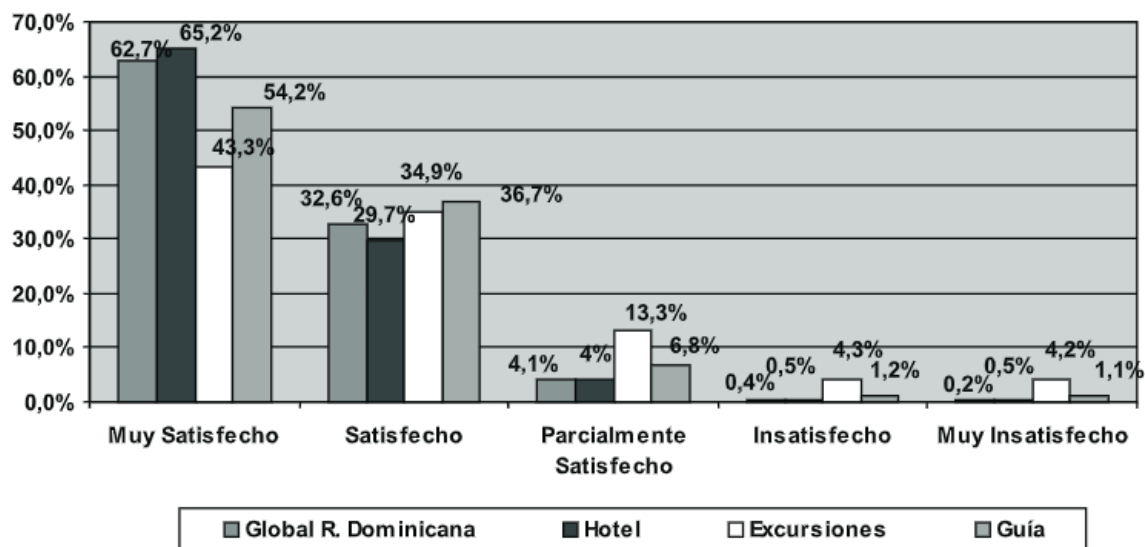
Los autores, José Amelio Medina-Merodio, Carmen de Pablos-Heredero, María Lourdes Jiménez-Rodríguez, Luis de Marcos Ortega, Roberto Barchino-Plata, Daniel Rodríguez-García y Daniel Gómez-Aguado (2014), en su investigación titulada, Análisis de la satisfacción de cliente mediante el uso de cuestionarios con preguntas abiertas, se tiene como objetivo el analizar la satisfacción del cliente mediante el uso de cuestionarios con preguntas abiertas. Para esta investigación se ha diseñado un cuestionario para medir la satisfacción del cliente según la ISO 9001, en el que las preguntas son de respuestas abiertas, en donde el cliente no se limita a responder mediante opciones. Todo el proceso para analizar cada

contestación de cada cliente no tendría ningún valor si no pudiéramos convertir esos datos en información. Con ese fin se ha generado un cuadro de mando como herramienta que permita la toma de decisiones y se han implementado gráficas que facilitan su comprensión de una manera rápida. Para ello nos hemos basado en la norma UNE 66175. (Medina-Merodio, y otros, 2014).

Los autores Sergio Moreno Gil, Daniel F Celis Sosa, Teresa Aguiar Quintana (2002), en su investigación, Análisis de la satisfacción del turista de paquetes turísticos respecto a las actividades de ocio en el destino: El caso de República Dominicana, tiene como objetivo el analizar el nivel de satisfacción de los clientes que viajan por medio de un paquete turístico ofertado por un turoperador. Para esta investigación se realizó un cuestionario estructurado, mediante el que se solicitó al cliente que, contrastando sus expectativas con el servicio recibido, puntuara en una escala Likert su nivel de satisfacción con valores entre 1 (nada satisfecho) y 5 (muy satisfecho). El número de encuestas realizadas fue de 956, lo que para un nivel de confianza del 95.5%. (Moreno Gil, Celis Sosa, & Aguiar Quintana, 2002).



### SATISFACCIÓN GENERAL DEL CLIENTE



*Ilustración 1: Satisfacción general del cliente*

**Fuente: Análisis de la satisfacción del turista de paquetes turísticos respecto a las actividades de ocio en el destino: El caso de República Dominicana.** (Moreno Gil, Celis Sosa, & Aguiar Quintana, 2002)

La presente investigación se realiza con el propósito de crear una base para futuras investigaciones con respecto a los algoritmos utilizados para realizar un análisis de sentimientos y facilitar la evaluación de las opiniones de los visitantes del sector turismo en el ámbito local, usando el análisis de sentimiento mediante herramientas B.I., con los cuales se ejecutará acción (o no) ante ello. Asimismo, con su aplicación se facilita la disminución del uso de papel para realizar encuestas. Esto es un paso importante en la aplicación de políticas ambientalistas, generando un ahorro de aproximadamente 4,000.00 soles, los cuales podrían ser destinados a otros fines. El proyecto muestra su impacto social a través de incremento del interés por el bienestar y la satisfacción de los turistas en la ciudad de Trujillo, contribuyendo de manera indirecta en la mejora de la atención de los mismos y de

los servicios otorgados, siendo posible su aprovechamiento, no solo para los turistas, sino para los residentes de las zonas aledañas. Este incremento en la afluencia de turistas generará mayor inversión en los negocios cercanos.

Conjuntamente, en este trabajo de investigación recogemos conceptos como:

### **Análisis de sentimientos**

En el informe Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español de (Cruz, Troyano, Enriquez, & Ortega, 2008) se afirma que, en los últimos años, se ha producido un alto interés de parte de las empresas, gobiernos, entre otros, por el procesamiento automático de las opiniones contenidas en documentos de texto, como consecuencia del gran aumento de contenidos de usuario en la web, para detectar automáticamente las opiniones vertidas por sus clientes o ciudadanos.

Si bien el análisis computacional de sentimientos permanece aún alejado de utilizar las teorías cognitivas de la emoción o el afecto para tomar decisiones, de acuerdo a (Maas, y otros, 2011) se ha llegado a aceptar que estas llegan a ser muy útiles para identificar similitudes de los sentimientos o su polaridad (positiva, negativa, neutra).

De acuerdo con (Martínez-Cámara, Perea-Ortega, Martín-Valdivia, & Ureña López, 2011), existe una gran cantidad de textos en el que el contenido subjetivo es lo más relevante, y cuyo procesamiento no debería limitarse a aplicar únicamente las técnicas de la clasificación de documentos. Ante esta necesidad de clasificar la orientación, o la opinión que se expresan en los documentos, surge el área análisis de sentimientos, o también denominada en la bibliografía como minería de

opiniones, o en inglés, sentiment analysis u opinion mining. El análisis de sentimientos trata de clasificar los documentos en función de la polaridad de la opinión que expresa su autor. Esta nueva área que combina PLN y minería de textos incluye una gran cantidad de tareas que han sido tratadas en mayor o menor medida. Existen principalmente dos formas distintas de enfrentarse a este problema: aplicando aprendizaje automático o aplicando un enfoque semántico. Dos son las aplicaciones más importantes: determinar la polaridad de las opiniones a nivel de documento, frase o característica, y determinar si un documento contiene opiniones

En el informe Procesamiento de Lenguaje Natural en sistema de análisis de sentimientos (Dubiau, 2013), se describe y separa en grupos, las técnicas más utilizadas en la clasificación subjetiva de textos.

El grupo uno, denominado, Técnicas de Clasificación Supervisada contiene las siguientes técnicas:

**Metodología de Naive Bayes**, se basa en el teorema de Bayes y en la premisa de independencia de los atributos para obtener la probabilidad de que un documento pertenezca a una determinada clase como se indica en la ecuación que sigue:

$$P(C_i|D) \propto P(C_i) \prod_{k=1}^n P(f_k | C_i) \quad (1)$$

Donde  $f_k$  son los atributos del documento,  $C_i$  es la clase y  $P(f_k | C_i)$  es la probabilidad de ocurrencia del atributo en la clase dada. La clase seleccionada por el clasificador será la que maximice la probabilidad anterior. Las implementaciones del algoritmo de Naive Bayes difieren principalmente en la aproximación de

$P(f_k | C_i)$  y las técnicas de smoothing utilizadas para el tratamiento de probabilidades bajas o nulas.

**Modelo de Máxima Entropía**, es un método de clasificación discriminativo donde los documentos del conjunto de datos son descriptos a partir de una lista de atributos, siendo cada uno una restricción del modelo. Este método se basa en seleccionar la distribución de probabilidad que satisfaga todas las restricciones del modelo y maximice la entropía. Esto apunta a preservar la incertidumbre tanto como sea posible. En modelos de máxima entropía la probabilidad de que un documento pertenezca a una clase se define como sigue:

$$P(c|x) = \frac{\exp(\sum_{i=0}^N w_{ci} f_i)}{\sum_{c' \in C} \exp(\sum_{i=0}^N w_{c'i} f_i)} \quad (2)$$

Donde,  $c$  es la clase que se desea evaluar,  $x$  es el documento,  $f_i$  es cada atributo,  $w_{ci}$  es el peso de ese atributo para la clase  $c$  que se está evaluando y  $w_{c'i}$  es el peso del atributo en cada una de las posibles clases. El cálculo de pesos es un problema complejo conocido como convex optimization que busca maximizar la verosimilitud del modelo.

**Support Vector Machines**, es un método supervisado de clasificación binaria en el cual el entrenamiento consiste en encontrar un hiperplano que separe los vectores de atributos que representan los documentos del conjunto de datos en dos grupos, siendo esta separación la más grande posible. Aquellos vectores que definen los márgenes de la máxima separación entre las clases se conocen como support vectors. Para la predicción de la clase utilizando este modelo se define la ecuación que sigue:

$$f(x) = \text{sign}\left(\sum_i \alpha_i x_i * x + b\right) \quad (3)$$

Siendo,  $x$  el vector de atributos del documento a clasificar,  $\alpha_i$  cada uno de los pesos que ponderan los vectores de atributos identificados como support features,  $x_i$  cada uno de los support features y  $b$  el termino independiente. Un valor de  $-1$  indicará que el documento pertenece a una clase y un valor de  $+1$  a la otra, lo que representa de qué lado del hiperplano se encuentra  $x$ .

**Los Árboles de Decisión**, son un método de clasificación supervisado en el que el entrenamiento consiste en la construcción de un árbol de decisión de múltiples caminos en el que para cada nodo se busca el atributo que provee mayor ganancia de información para la clase. El árbol crece hasta su tamaño máximo y luego es acotado para mejorar su capacidad de generalización para los datos que no ocurren en el conjunto de datos de entrenamiento. A partir de este modelo se infieren reglas de decisión sobre las cuales se basa la clasificación. En este trabajo utilizamos una implementación optimizada del algoritmo de CART (Classification And Regression Trees) basado en arboles de decisión.

El segundo grupo, denominado, Técnicas de Clasificación No Supervisada contiene las siguientes técnicas:

**Polaridad de Turney**, es un algoritmo no supervisado de clasificación subjetiva de textos que se basa en predecir la polaridad de un documento a partir de su orientación semántica. Este método consiste en extraer de los documentos aquellos bigramas que cumplen con determinados patrones de opinion y luego calcular la orientación semántica a partir de la distancia de cada bigrama a los términos del

idioma inglés, “excelent” y “poor” que sugieren referencias positivas y negativas respectivamente según la siguiente ecuación:

$$SO(\textit{phrase}) = \log_2 \left[ \frac{\textit{hits}(\textit{phrase NEAR "excellent"}) \textit{hits} ("poor")}{\textit{hits}(\textit{phrase NEAR "poor"}) \textit{hits} ("excellent")} \right] \quad (4)$$

El último grupo se denomina Procesamiento de Textos y tiene por objetivo, el mejorar los resultados del clasificador y dependiendo de la tarea de procesamiento de texto que se esté realizando y el idioma en el que esté escrito el texto, pueden requerirse algunas de las siguientes transformaciones en el corpus de datos de entrada antes de su procesamiento.

**Normalización:** Consiste en unificar términos que representan la misma información y pueden ser escritos en distintos formatos. Por ejemplo, “restaurant”, “restaurant”, “restorán”, “restó”.

**Tokenización:** Separación de oraciones y palabras de un documento a partir de tokens, o caracteres especiales, que indican el fin de una oración o palabra y el comienzo de la que sigue.

**Stemming:** Los algoritmos de Stemming (o Stemmers) permiten obtener la raíz o stem de una palabra eliminando terminaciones, con el objetivo de unificar aquellos términos que aportan la misma información al clasificador. Por ejemplo, los términos “recomendable, recomendamos, recomendar, recomendación” son reemplazados por su stem: “recomend”. El stem de una palabra no necesariamente será un término válido del vocabulario.

**Lematización:** Otra forma de unificar los términos que aportan la misma información al clasificador es reemplazando cada palabra por su lema. El lema de

una palabra es un término válido del vocabulario que por convención es la representación de todas las formas flexionadas de la palabra, es decir, para hallar el lema de un término se eliminan todas las flexiones (conjugaciones, grado, persona, género, número, etc.). Por ejemplo, lema(“pésimo”) = “malo”; lema(“empieza”) = “empezar”; lema(“primeras”) = “primero”; lema(“estas”) = “este”.

**Tratamiento de Negaciones:** Con el objetivo de distinguir aquellos términos que siguen a una negación suele ser de utilidad indicar que la palabra aparece -negada agregando un prefijo a todos los términos que siguen a la negación hasta el siguiente signo de puntuación.

### **Metodologías ágiles.**

(Letelier Torres & Sánchez López, 2003). En una reunión celebrada en febrero de 2001 en Utah - EE. UU., nace el término “ágil” aplicado al desarrollo de software. En esta reunión participan un grupo de 17 expertos de la industria del software, incluyendo algunos de los creadores o impulsores de metodologías de software. Su objetivo fue esbozar los valores y principios que deberían permitir a los equipos desarrollar software rápidamente y respondiendo a los cambios que puedan surgir a lo largo del proyecto. Se pretendía ofrecer una alternativa a los procesos de desarrollo de software tradicionales, caracterizados por ser rígidos y dirigidos por la documentación que se genera en cada una de las actividades desarrolladas. Varias de las denominadas metodologías ágiles ya estaban siendo utilizadas con éxito en

proyectos reales, pero les faltaba una mayor difusión y reconocimiento. Tras esta reunión se creó The Agile Alliance, una organización, sin ánimo de lucro, dedicada a promover los conceptos relacionados con el desarrollo ágil de software y ayudar a las organizaciones para que adopten dichos conceptos. El punto de partida es fue el Manifiesto Ágil, un documento que resume la filosofía “ágil”.

### **El Manifiesto ágil**

Según el Manifiesto se valora:

**Al individuo y las interacciones del equipo de desarrollo sobre el proceso y las herramientas.** La gente es el principal factor de éxito de un proyecto software. Es más importante construir un buen equipo que construir el entorno. Muchas veces se comete el error de construir primero el entorno y esperar que el equipo se adapte automáticamente. Es mejor crear el equipo y que éste configure su propio entorno de desarrollo en base a sus necesidades.

**Desarrollar software que funciona más que conseguir una buena documentación.** La regla por seguir es “no producir documentos a menos que sean necesarios de forma inmediata para tomar una decisión importante”. Estos documentos deben ser cortos y centrarse en lo fundamental. La colaboración con el cliente más que la negociación de un contrato. Se propone que exista una interacción constante entre el cliente y el equipo de desarrollo. Esta colaboración entre ambos será la que marque la marcha del proyecto y asegure su éxito.

**Responder a los cambios más que seguir estrictamente un plan.** La habilidad de responder a los cambios que puedan surgir a lo largo del proyecto (cambios en los



requisitos, en la tecnología, en el equipo, etc.) determina también el éxito o fracaso de este. Por lo tanto, la planificación no debe ser estricta sino flexible y abierta.

### **Principales metodologías ágiles**

Tal como se indica en (Ruiz Larrocha, 2017), en las últimas dos décadas han proliferado multitud de metodologías. A continuación, se enumeran las que se encuentran inscritas en la organización Agile Alliance.

- AD, Agile Database Techniques.
- AM, Agile Modeling.
- ASD, Adaptative Software Development.
- AUP, Agile Unified Process.
- Crystal.
- FDD, Feature Driven Development.
- DSDM, Dynamic System Development Method.
- Lean Software Development.
- SCRUM.
- TDD, Test-Driven Design.
- Xbreed
- XP, eXtreme Programming.

### **Características de las metodologías ágiles**

En el siguiente cuadro se contraponen, a forma de resumen, las características de más metodologías ágiles frente a las clásicas:

*Tabla 1: Cuadro comparativo: Metodologías clásicas VS Metodologías ágiles*

Metodologías clásicas	Metodologías ágiles
Requisitos exhaustivos al inicio del proyecto.	Admisión de incertidumbres en requisitos.
Planificaciones largas y detalladas.	Planificaciones cortas y solapadas.
Procesos muy definidos y estructurados.	Proceso simple.
Equipos de desarrollo especializados y muy jerarquizados.	Equipo multidisciplinarios y auto-organizados.
Documentación exhaustiva.	Poca documentación.
Rigidez frente a los cambios de requisitos	Abiertas al cambio de requisitos.

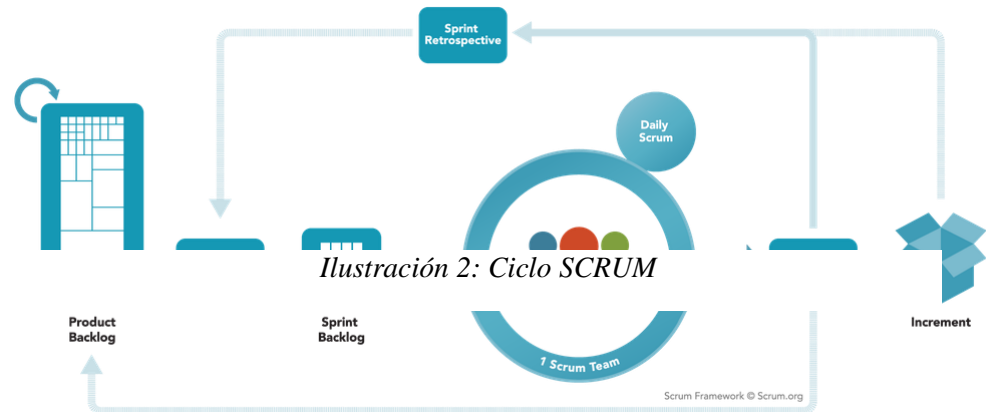
**(Ruiz Larrocha, 2017)**

## SCRUM

(Francia, scrum.org, 2017). Scrum es un proceso de gestión que reduce la complejidad en el desarrollo de productos para satisfacer las necesidades de los clientes. La gerencia y los equipos de Scrum trabajan juntos alrededor de requisitos y tecnologías para entregar productos funcionando de manera incremental usando el empirismo.

Scrum es un marco de trabajo simple que promueve la colaboración en los equipos para lograr desarrollar productos complejos.

El siguiente gráfico representa Scrum como lo describen Ken Schwaber y Jeff Sutherland en su libro Software en 30 días.



(Schwaber & Sutherland, 2016)

Según (Schwaber & Sutherland, 2016), los eventos de Scrum se utilizan para minimizar la necesidad de reuniones no definidas en Scrum y establecer una cadencia que permita al equipo fomentar la comunicación y colaboración reduciendo el tiempo en reuniones extensas además de reducir los procesos restrictivos y predictivos. Todos los eventos tienen una caja de tiempo o “TimeBox”. Una vez que se inicia un Sprint este tiene una duración fija y no se puede acortar o alargar. Los siguientes eventos pueden terminar siempre que se logre el propósito del evento, pero dentro de la caja de tiempo y asegurando el fomento de la transparencia. Los eventos de Scrum son:

**El Sprint**, es el corazón de Scrum, es un bloque de tiempo (time-box) de un mes o menos durante el cual se crea un incremento de producto “Terminado” utilizable y potencialmente desplegable. Es más conveniente si la duración de los Sprints es consistente a lo largo del esfuerzo de desarrollo. Cada nuevo Sprint comienza inmediatamente después de la finalización del Sprint anterior.

**Sprint Planning**, es el trabajo por realizar durante el Sprint. Este plan se crea mediante el trabajo colaborativo del Equipo Scrum completo. La Planificación de

Sprint tiene un máximo de duración de ocho horas para un Sprint de un mes. Para Sprints más cortos el evento es usualmente más corto. El Scrum Máster se asegura de que el evento se lleve a cabo y que los asistentes entiendan su propósito. El Scrum Máster enseña al Equipo Scrum a mantenerse dentro del bloque de tiempo.

**Daily Scrum**, es una reunión con un bloque de tiempo de 15 minutos para que el Equipo de Desarrollo sincronice sus actividades y cree un plan para las siguientes 24 horas. Esto se lleva a cabo inspeccionando el trabajo avanzado desde el último Scrum Diario y haciendo una proyección acerca del trabajo que podría completarse antes del siguiente.

**Sprint Review**, al final del Sprint se lleva a cabo una Revisión de Sprint para inspeccionar el Incremento y adaptar la Lista de Producto si fuese necesario. Durante la Revisión de Sprint, el Equipo Scrum y los interesados colaboran acerca de lo que se hizo durante el Sprint. Basándose en esto y en cualquier cambio a la Lista de Producto durante el Sprint, los asistentes colaboran para determinar las siguientes cosas que podrían hacerse para optimizar el valor. Se trata de una reunión informal, no una reunión de seguimiento, y la presentación del Incremento tiene como objetivo facilitar la retroalimentación de información y fomentar la colaboración. Se trata de una reunión restringida a un bloque de tiempo de cuatro horas para Sprints de un mes. Para Sprints más cortos, se reserva un tiempo usualmente más corto. El Scrum Máster se asegura de que el evento se lleve a cabo y que los asistentes entiendan su propósito. El Scrum Máster enseña a todos a mantener el evento dentro del bloque de tiempo fijado.

**Sprint Retrospective**, es una oportunidad para el Equipo Scrum de inspeccionarse a sí mismo y de crear un plan de mejoras que sean abordadas durante el siguiente Sprint. La Retrospectiva de Sprint tiene lugar después de la Revisión de Sprint y antes de la siguiente Planificación de Sprint. Se trata de una reunión restringida a un bloque de tiempo de tres horas para Sprints de un mes. Para Sprints más cortos se reserva un tiempo usualmente más corto. El Scrum Máster se asegura de que el evento se lleve a cabo y que los asistentes entiendan su propósito. El Scrum Máster enseña a todos a mantener el evento dentro del bloque de tiempo fijado. El Scrum Máster participa en la reunión como un miembro del equipo ya que la responsabilidad del proceso Scrum recae sobre él.

### **eXtreme Programing (XP)**

(Robles & Ferrer, 2002). La programación extrema se basa en una serie de reglas y principios que se han ido gestando a lo largo de toda la historia de la ingeniería del software. Usadas conjuntamente proporcionan una nueva metodología de desarrollo software que se puede englobar dentro de las metodologías ligeras, que son aquellas en la que se da prioridad a las tareas que dan resultados directos y que reducen la burocracia que hay alrededor tanto como sea posible (pero no más) (Fowler, 2004). La programación extrema, dentro de las metodologías ágiles, se puede clasificar dentro de las evolutivas (Harrison). Una de las características de eXtreme Programming es que muchos de, si no todos, sus ingredientes son de sobra conocidos dentro de la rama de la ingeniería del software desde hace tiempo, incluso desde sus comienzos. Los autores de han seleccionado los que han considerados como los mejores y han profundizado en sus relaciones y en cómo se refuerzan unos a otros. El resultado ha sido una metodología única y compacta. Por eso, aunque se pueda alegar

que la programación extrema no se base en principios nada nuevos, se ha de aclarar que, en conjunto, es una nueva forma de ver el desarrollo de software. Aunque, como ya se ha comentado, la programación extrema se basa en unos valores, unos principios fundamentales y unas prácticas, en este artículo no se van a enumerar así de primeras, ya que el autor considera que no es la mejor forma de presentarlos. Los principios y prácticas no se han hecho a priori o porque sí, sino que tienen un porqué a partir de una forma global de desarrollar software que, al menos en teoría, parece ser más eficiente. Por tanto, en este artículo se presentará la programación extrema desde un punto de vista práctico para luego dar paso a enunciar los valores y principios que se han extraído y las prácticas que hacen que se lleven a buen fin. La idea es seguir que el lector pueda seguir en los siguientes párrafos un proceso de desarrollo extremo tal y como debería darse en un equipo de desarrollo que siguiera la metodología XP. De esta forma se irán detallando y explicando las diferentes técnicas utilizadas, así como su razón de ser. Una vez que hayamos visto el proceso de desarrollo extremo, los valores, principios y prácticas serán evidentes y no requerirán mucho detenimiento.

### **Proceso de desarrollo**

La programación extrema parte del caso habitual de una compañía que desarrolla software, generalmente software a medida, en la que hay diferentes roles: un equipo de gestión, un equipo de desarrolladores y los clientes. La relación con el cliente es totalmente diferente a lo que se ha venido haciendo en las metodologías tradicionales que se basan fundamentalmente en una fase de captura de requisitos previa al desarrollo y una fase de validación posterior al mismo.

a) **Interacción con el cliente.** En la programación extrema al cliente no sólo se le pide que apoye al equipo de desarrollo, en realidad podríamos decir que es parte de él. Su importancia es capital a la hora de abordar las historias de los usuarios y las reuniones de planificación, como veremos más adelante. Además, será tarea suya realimentar al equipo de desarrolladores después de cada iteración con los problemas con los que se ha encontrado, mostrando sus prioridades, expresando sus sensaciones. El proceso de captura de requisitos de XP gira en torno a una lista de características que el cliente desea que existan en el sistema final. Cada una de estas características recibe el nombre de historias de usuarios y su definición consta de dos fases:

En la primera fase el cliente describe con sus propias palabras las características y el responsable del equipo de desarrollo le informa de la dificultad técnica de cada una de ellas y por lo tanto de su coste. A través del diálogo resultante el cliente deja por escrito un conjunto de historias y las ordena en función de la prioridad que tienen para él. En este momento ya es posible definir unos hitos y unas fechas aproximadas para ellos.

La segunda fase consiste en coger las primeras historias que serán implementadas (primera iteración) y dividir las en las tareas necesarias para llevarlas a cabo. El cliente también participa, pero hay más peso del equipo de desarrollo, que dará como resultado una planificación más exacta.

En cada iteración se repetirá esta segunda fase para las historias planificadas para ella. Este proceso es una de las principales diferencias

con las metodologías tradicionales. Aunque las historias de usuarios guardan cierta relación con otras técnicas como los casos de uso de UML, su proceso de creación es muy diferente. En lo que al cliente se refiere no se le exige que especifique exactamente lo que quiere al principio con un documento de requisitos de usuario. La parte que se mantiene con este documento es que es el cliente el que tiene que escribir lo que quiere, no se permite que alguien del equipo de desarrolladores lo escriba por él.

**b) Planificación del proyecto.** Es probablemente en este punto donde nos debamos enfrentar a la planificación de entregas (release planning) donde planificaremos las distintas iteraciones. Para ello existen una serie de reglas que hay que seguir para que las tres partes implicadas en este proceso (equipo de gestión, equipo de desarrollo y cliente) tengan voz y se sientan parte de la decisión tomada, que al fin y al cabo debe contentar a todos. La primordial es que las entregas se hagan cuanto antes y que con cada iteración el cliente reciba una nueva versión. Cuanto más tiempo se tarde en introducir una parte esencial, menos tiempo habrá para trabajar en ella posteriormente. Se aconsejan muchas entregas y muy frecuentes. De esta forma, un error en una parte esencial del sistema se encontrará pronto y, por tanto, se podrá arreglar antes.

Diversas prácticas como las pruebas unitarias, la integración continua o el juego de la planificación permiten eliminar los principales motivos por los que suele ser necesario trabajar muchas horas extra. Pero lo mejor de todo es que a la hora de planificar uno se puede equivocar. Es más, todos sabemos que lo común es equivocarse y por ello la metodología ya tiene



previsto mecanismos de revisión. Por tanto, es normal que cada 3 a 5 iteraciones se tengan que revisar las historias de los usuarios y renegociar nuevamente la planificación. Hemos visto que al principio del proyecto se hace una planificación en iteraciones que debe ser retocada al cabo de unas cuantas iteraciones. A esto hay que añadir que en cada iteración también hay que realizar la planificación de esta, lo que ha venido a llamarse planificación iterativa.

A raíz de lo anterior, podemos entender el siguiente consejo: optimiza al final. El eslogan subyacente es "make it work, make it right and then make it fast" (haz que funcione, hazlo bien y entonces haz que sea rápido). Y es que nunca se sabe a priori dónde puede estar el verdadero cuello de botella, así que lo mejor es no añadir funcionalidad demasiado temprano y concentrarnos completamente en lo que es necesario hoy

**c) Diseño, desarrollo y pruebas.** El desarrollo es la pieza clave de todo el proceso de programación extrema. Todas las tareas tienen como objetivo que se desarrolle a la máxima velocidad, sin interrupciones y siempre en la dirección correcta. También se otorga una gran importancia al diseño y establece que éste debe ser revisado y mejorado de forma continua según se van añadiendo funcionalidades al sistema.

## **SCRUM vs XP**

Según (XP vs Scrum, s.f.), las diferencias entre las metodologías de Scrum y XP son las siguientes:

- a) Los Equipos de Scrum normalmente trabajan en iteraciones (llamadas sprints) con una duración de 2 semanas a un mes. Los Equipos de XP normalmente trabajan con iteraciones muy cortas (1 o 2 semanas).
- b) Los equipos de Scrum no permiten cambios en sus sprints. Una vez que el sprint planning meeting ha concluido y se llegó a un acuerdo para elegir los product backlog items, estos no pueden recibir cambios, por lo menos hasta el final del sprint. Los equipos XP aceptan los cambios con más facilidad en sus iteraciones. Siempre y cuando el equipo no haya empezado ya a trabajar en un requerimiento en particular, un nuevo requerimiento de un tamaño similar puede ser cambiado por otro equivalente en la iteración.
- c) Los equipos de XP trabajan en un orden estricto. Los requerimientos que serán desarrollados son priorizados por el cliente (Product Owner en Scrum) y el equipo debe desarrollar los requerimientos en ese orden específico. En contraste, el product owner de Scrum prioriza los product backlog, pero el equipo determina el orden en el que serán desarrollados.
- d) Scrum no recomienda ninguna buena práctica de desarrollo. XP si lo hace. Las buenas prácticas de XP no es necesarios obligar a los equipos a realizarlas pues descubren sus virtudes por cuenta propia.

### **Lenguaje de programación.**

De acuerdo con (Gortázar Bellas, Martínez Unanue, & Fresno Fernández, 2016), señala que Sir Charles Antony Richard Hoare, inventor del algoritmo QuickSort, dijo en una ocasión:

*“El propósito principal de un lenguaje de programación es ayudar a  
programador en la práctica de su arte”.*

Para comprender los principios de diseño de los modernos lenguajes de programación, es necesario conocer algo de su historia. Esta historia comienza realmente en el siglo XIX, cuando Charles Babbage idea la máquina analítica. Esta máquina estaba diseñada para funcionar utilizando tarjetas perforadas. Estas tarjetas ya se conocían puesto que eran las utilizadas en el telar de Jacquard, una máquina inventada por Joseph Marie Jacquard que era capaz de crear telas a partir de patrones grabados en tarjeta de este tipo. Aunque Babbage no terminó nunca de construirla, su máquina analítica se considera la primera computadora. Una matemática contemporánea de Babbage, llamada Ada Lovelace, llegó incluso a crear programas para esta máquina analítica, aunque evidentemente nunca pudo comprobar si funcionaban. Por ello a Babbage se le considera el padre de la computadora y a Ada la primera programadora de la historia (el lenguaje Ada, creado para el ministerio de defensa estadounidense se llamó así en su honor).

Mucho más recientemente, en los años 30 del siglo XX se construyeron las primeras máquinas capaces de realizar ciertas operaciones, si bien normalmente estaban orientadas a realizar algún tipo concreto de cálculo científico. Existe cierto acuerdo en que el ENIAC, desarrollado en el año 1946, puede considerarse el primero computador realmente a propósito general. El ENIAC no se programaba utilizando un lenguaje de programación, sino cableando directamente sus circuitos.

Por esa época, John Von Neumann propuso una arquitectura diferente para las computadoras, donde el programa se almacena en la máquina antes de ejecutarse.

Esto permitió a partir de entonces evitar tener que cablear todos los componentes para cada nuevo programa. Es entonces donde realmente empieza la historia de los lenguajes de programación.

Desde un punto de vista coloquial, un lenguaje de programación es una notación para comunicarle a una computadora lo que queremos que haga. Desde un punto de vista formal, podemos definirlo como un sistema notacional para describir computaciones en una forma legible tanto para la máquina como para el ser humano.

Esta definición forma e la que ha estado guiando la evolución de los lenguajes de programación. En esta evolución podemos distinguir cinco generaciones de lenguajes de programación.

**Primera generación.** A esta generación pertenece el lenguaje de máquina. El lenguaje de máquina consiste exclusivamente en secuencias de ceros y unos y es el único lenguaje que entienden las computadoras modernas. Cada computadora dispone de lo que se denomina un conjunto de instrucciones que son las operaciones que la computadora entiende (sumar, restar, cargar memoria, guardar memoria, etc.). Estas operaciones se indican mediante secuencias de ceros y unos. El principal inconveniente del lenguaje de máquina es que es difícil de entender por un humano. Los programas escritos en lenguaje de máquina generalmente sólo funcionan en un modelo de máquina específico, dado que cada computadora define su propio código de secuencias de ceros y unos.

**Segunda generación.** Pertenecen a esta generación los lenguajes ensambladores. Estos lenguajes establecen una serie de reglas mnemotécnicas que hacen más

sencilla la lectura y escritura de programas. Estas reglas mnemotécnicas consisten simplemente en asociar nombres legibles a cada una de las instrucciones soportadas por la máquina (ADD, SUB, LOAD, STORE, etc.). Los programas escritos en lenguaje ensamblador no son directamente ejecutables por la máquina, puesto que ésta sólo entiende instrucciones codificadas como ceros y unos. Pero es muy sencillo traducirlos a lenguaje de máquina. El lenguaje ensamblador aún se utiliza hoy en día para programar drivers para dispositivos o determinadas partes de los sistemas operativos.

**Tercera generación.** A esta generación pertenecen los lenguajes como C, FORTRAN o Java. Estos lenguajes se denominan lenguajes de alto nivel, porque están bastante alejados del lenguaje de máquina y son mucho más legibles por el hombre. Para convertir los programas escritos en estos lenguajes de alto nivel en código máquina entendible por la computadora empiezan a hacer falta complejos programas que los traduzcan. Estos traductores se denominan compiladores. Los lenguajes de alto nivel son necesarios para programar grandes sistemas software, como sistemas operativos (Linux, Windows), aplicaciones para la web (Facebook, Tuenti, Twitter) o aplicaciones para móviles. Estos lenguajes facilitan el mantenimiento y la evolución del software. Los primeros lenguajes de tercera generación fueron FORTRAN, LISP, ALGOL, COBOL. Los cuatro surgieron a finales de los 50.

**Cuarta generación.** Los lenguajes de la cuarta generación son lenguajes de propósito específico, como SQL, NATURAL o ABAP. Estos lenguajes no están diseñados para programar aplicaciones complejas, sino que fueron diseñados para solucionar problemas muy concretos. Por ejemplo, SQL es un lenguaje empleado

para describir consultas, inserciones o modificaciones de base de datos. El lenguaje está específicamente diseñado para ello. Otro ejemplo son los lenguajes que incluyen paquetes estadísticos como SPSS que permiten manipular grandes cantidades de datos con fines estadísticos.

**Quinta generación.** Los lenguajes de quinta generación son los utilizados principalmente en el área de la inteligencia artificial. Se trata de lenguajes que permiten especificar restricciones que se le indican al sistema, que resuelve un determinado problema sujeto a estas restricciones. Algunos ejemplos de lenguaje de quinta generación son Prolog o Mercury.

Cada generación ha tratado de resolver los problemas detectados en la generación anterior. Así, los lenguajes de segunda generación surgieron a partir de la necesidad de evitar tener que recordar las instrucciones de una computadora concreta como una secuencia de ceros y unos.

Sin embargo, los lenguajes de segunda generación eran poco convenientes para crear programas complejos. A medida que el software se hacía más complejo era más difícil mantenerlo debido a lo poco estructurado que es el lenguaje ensamblador (...). Esto llevó a la aparición de los lenguajes de tercera generación.

### C.

De acuerdo con (Joyanes Aguilar & Zahonero Martínez, 2010), C es el lenguaje de programación de propósito general asociado, de modo universal, al sistema operativo UNIX. Sin embargo, la popularidad, eficacia y potencia de C, se ha producido porque este lenguaje no está prácticamente asociado a ningún sistema

operativo, ni a ninguna máquina en especial. Esta es la razón fundamental, por la cual C, es conocido como el lenguaje de programación de sistemas, por excelencia.

C es una evolución de los lenguajes BCPL – desarrollado por Martin Richards – y B – desarrollado por Ken Thompson en 1970 – para el primitivo UNIX de la computadora DEC PDP-7.

C nació realmente en 1978, con la publicación de *The C Programming Language*, (Kernighan & Ritchie, 1978). Desde su nacimiento, C fue creciendo en popularidad y los sucesivos cambios en el lenguaje, a lo largo de los años, junto a la creación de compiladores por grupos no involucrados en su diseño, hicieron necesario pensar en la estandarización de la definición del lenguaje C.

Así, en 1983, el American National Standard Institute (ANSI), una organización internacional de estandarización creó un comité (el denominado X3J11) cuya tarea fundamental consistía en hacer “una definición no ambigua del lenguaje C, e independiente de la máquina”. Había nacido el estándar ANSI del lenguaje C. Con esta definición C se asegura que cualquier fabricante de software que vende un compilador ANSI C incorpora todas las características del lenguaje, especificadas por el estándar. Esto significa también que los programadores que escriban programas en C estándar tendrán la seguridad de que correrán sus modificaciones en cualquier sistema que tenga un compilador C.

C es un lenguaje de alto nivel, que permite programar con instrucciones de lenguaje de propósito general. También, C se define como un lenguaje de programación estructurado de propósito general; aunque en su diseño primó el hecho de que fuera

especificado como un lenguaje de programación de Sistemas, lo que proporciona una enorme cantidad de potencia y flexibilidad.

El estándar ANSI formaliza construcciones no propuestas en la primera versión de C, en especial, asignación de estructuras y enumeraciones. Entre otras aportaciones, se definió esencialmente, una nueva forma de declaraciones de funciones (prototipos). Pero, otra de las grandes aportaciones es la biblioteca estándar de funciones.

Hoy, en el siglo XXI, C sigue siendo uno de los lenguajes de programación más utilizados en la industria del software, así como en institutos tecnológicos, escuelas de ingeniería y universidades. Prácticamente todos los fabricantes de sistemas operativos, UNIX, Linux, MacOS, Solaris, etc. Soportan diferentes tipos de compiladores de lenguaje C.

### **Características técnicas de C.**

Hay numerosas características que diferencian a C de otros lenguajes y lo hacen eficiente y potente a la vez

- Una nueva sintaxis para declarar funciones. Una declaración de función puede añadir una descripción de los argumentos de la función. Esta información sirve para que los compiladores detecten más fácilmente los errores causados por argumentos que no coinciden.
- Asignación de estructuras (registros) y enumeraciones.
- Procesador más sofisticado.



- Una nueva definición de la biblioteca que acompaña a C. Entre otras funciones se incluyen: acceso al sistema operativo (por ejemplo, lectura y escritura de archivos), entrada y salida con formato, asignación dinámica de memoria, manejo de cadenas de caracteres.
- Una colección de cabeceras estándar que proporcionan acceso uniforme a las declaraciones de funciones y tipos de datos.

## C#.

De acuerdo con (Deitel & Deitel, 2008), C# evolucionó de C. el cual a su vez evolucionó de dos lenguajes de programación anteriores, BCLP y B.

A principios de la década de los ochenta, Bjarne Stroustrup desarrolló una extensión una extensión de C en los laboratorios Bell: C++. Este lenguaje proporciona un conjunto de características que “pulen” al lenguaje C, pero, lo más importante es que proporciona la capacidad de una programación orientada a objetos. C++ también se estandarizó mediante los comités ANSI e ISO.

(Ceballos Sierra, 2008) nos comenta que para entender lo que es C# e imprescindible decir antes lo que es Microsoft .NET Framework o abreviadamente .NET. Se trata de un entorno de desarrollo multilenguaje diseñado por Microsoft para simplificar la construcción, distribución y ejecución de aplicaciones para internet. Tiene fundamentalmente tres componentes: una máquina virtual que

procesa código escrito en un lenguaje intermedio, una biblioteca de clases y ASP.NET que proporciona los servicios necesarios para crear aplicaciones Web.

### **Java.**

(Deitel & Deitel, 2008) mencionan que tal vez la contribución más importante a la fecha, por parte de la revolución del microprocesador, es que hizo posible el desarrollo de las computadoras personales, que ahora suman cientos de millones a nivel mundial. Las computadoras personales han tenido un profundo impacto en la vida de las personas, y en la manera en que las empresas realizan y administran su negocio.

Muchas personas creen que la siguiente área importante en la que los microprocesadores tendrán un profundo impacto es en los dispositivos electrónicos para uso doméstico. Al reconocer esto, Sun Microsystems patrocinó en 1991 un proyecto interno de investigación denominado Green. El proyecto desembocó en el desarrollo de un lenguaje basado en C++ al que su creador, James Gosling, llamó Oak debido a un roble que tenía a la vista desde su ventana en las oficinas de Sun. Posteriormente se descubrió que ya existía un lenguaje de programación con el mismo nombre. Cuando un grupo de gente de un visitó una cafetería local, sugirieron el nombre Java y así quedo.

Pero el proyecto Green tuvo muchas dificultades. El mercado para los dispositivos electrónicos inteligentes de uso doméstico no se desarrollaba tan rápido como Sun había anticipado. Peor aún, un contrato importante por el que Sun había competido se le otorgó a otra empresa de manera que el proyecto corría riesgo de cancelarse. Pero para su buena fortuna, la popularidad de World Wide Web explotó en 1993 y

la gente de Sun se dio cuenta inmediatamente del potencial de Java para agregar contenido dinámico y animaciones a las páginas Web. Esto trajo nueva vida al proyecto.

Sun anunció formalmente a java en una conferencia importante que tuvo lugar en mayo de 1995. Por lo general, un evento como éste no habría generado mucha atención. Sin embargo, Java generó un interés inmediato en la comunidad de negocios, debido al fenomenal interés en World Wide Web. En la actualidad, Java se utiliza para desarrollar aplicaciones empresariales a gran escala, para mejorar la funcionalidad de los servidores de World Wide Web (las computadoras que proporcionan el contenido que vemos en nuestros navegadores Web), para proporcionar aplicaciones para los dispositivos domésticos (como teléfonos celulares, radiolocalizadores y asistentes digitales personales) y para mucho otros propósitos.

### **Gestores de base de datos.**

De acuerdo con el portal (Powerdata, 2016), el gestor de datos es un sistema de software invisible para el usuario final, compuesto por un lenguaje de definición de datos, un lenguaje de manipulación y de consulta, que puede trabajar a distintos niveles. Tanto almacenar, modificar y acceder a la información como realizar consultas y hacer análisis para generar informes.

A su vez, el SGBD puede entenderse como una colección de datos relacionados entre sí, estructurados y organizados dentro del ecosistema conformado por ese conjunto de programas que acceden a ellos y facilitan su gestión.

Básicamente, el gestor de datos controla cualquier operación ejecutada por el usuario contra la base de datos. Para ello, se utilizan herramientas específicas, como sistemas de búsqueda y de generación de informes, además de distintas aplicaciones.

Igualmente, permite la definición de bases de datos a distintos niveles de abstracción, así como su manipulación, garantizando tanto la seguridad como la integridad y consistencia de estos. Siempre de forma separada a los programas o aplicaciones que los gestionan, sin tener que modificar éstos.

### **MySQL.**

De acuerdo con lo presentado por (Thibaud, 2006), MySQL es un sistema de administración de bases de datos relacionales rápido, robusto y fácil de usar. Se adapta bien a la administración de datos en un entorno de red, especialmente en arquitecturas cliente/ servidor. Se proporciona con muchas herramientas y es compatible con muchos lenguajes de programación. Es el más célebre Administrador de bases de datos relacionales del mundo Open Source, en particular gracias a su compatibilidad con el servidor de páginas Web Apache y el lenguaje de páginas Web dinámicas PHP.

### **Características del servidor MySQL.**

La lista siguiente presenta las principales características del servidor MySQL:

- Está escrito en C/C++ y probado con numerosos compiladores
- Funciona con muchos sistemas operativos
- Soporta muchas API como C, C++, PHP, Perl, Eiffel, Python, Java, Tcl.
- Dispone de un driver ODBC (Open DataBase Connectivity) para Windows lo que le aporta compatibilidad con la mayoría de los lenguajes disponible en este sistema operativo.
- Es totalmente multithread, lo que aumenta enormemente el rendimiento del servidor. MySQL está previsto para funcionar en entorno multiprocesador.
- Dispone de un sistema de administración de la memoria extremadamente potente. La cantidad de memoria tiene una incidencia directa sobre el rendimiento de MySQL. Si un servidor presenta fallos de rendimiento, suele bastar con añadir memoria RAM para memorar sentiblemente su rendimiento.
- Soporta muchos tipos de columnas de tipos numéricos, fecha y hora, cadena de texto. MySQL soporta también columnas de gran tamaño como las columnas de tipos texto y binario.
- Dispone de un sistema potente de administración de usuarios y sus privilegios.
- Es capaz de administrar bases de datos de gran tamaño. Ciertas empresas lo utilizan actualmente con bases de datos de unas 60 000 tablas y alrededor de 5 000 000 000 de registros.
- Una tabla puede contener hasta 32 índices y cada índice puede estar constituido por un máximo de 16 columnas.

- Es posible instalar varias instancias de servidor MySQL. Cada instancia es completamente independiente y gestiona sus propias bases de datos.

### **SQL Server 2014.**

(Rodríguez, 2012), nos comenta que SQL Server es un sistema de gestión de bases de datos relacionales (RDBMS) de Microsoft que está diseñado para el entorno empresarial. SQL Server se ejecuta en T-SQL (Transact -SQL), un conjunto de extensiones de programación de Sybase y Microsoft que añaden varias características a SQL estándar, incluyendo control de transacciones, excepción y manejo de errores, procesamiento fila, así como variables declaradas.

Bajo el nombre código Yukon en su etapa de desarrollo, SQL Server 2005 fue lanzado en noviembre de 2005. Se dice que el producto 2005 proporcionó una mayor flexibilidad, escalabilidad, confiabilidad y seguridad a las aplicaciones de base de datos, y permitió que fueran más fáciles de crear y desplegar, lo que reduce la complejidad y el tedio involucrado en la gestión de bases de datos. SQL Server 2005 también incluía más soporte administrativo.

El código original de SQL Server ha sido desarrollado por Sybase; a finales de 1980, Microsoft, Sybase y Ashton-Tate colaboraron para producir la primera versión del producto, SQL Server 4.2 para OS/2. Posteriormente, tanto Sybase como Microsoft ofrecieron productos de SQL Server. Sybase cambió después el nombre de su producto a Adaptive Server Enterprise.

### **Satisfacción del cliente.**

En su libro John Tschohl, (Tschohl, 2007), señala que la mayoría de las empresas no ven realmente que el servicio al cliente es en realidad una acción de venta, puesto que estimula a los clientes a regresar con más frecuencia debido al servicio.

De acuerdo con los autores (Denove & Power, 2006), para alcanzar la excelencia en el nivel de satisfacción que las empresas ofrecen a sus clientes es necesario disponer de una información precisa sobre ellos, la infraestructura necesaria para analizarla, proporcionarla a las personas que saben cómo utilizarla y actuar de acuerdo con ello. Estos datos y la infraestructura que permite su análisis son la auténtica voz del cliente, que es la mejor guía con que cualquier empresa puede contar.

De acuerdo a (Thompson, 2006), que por ese motivo, resulta de vital importancia que tanto mercadólogos, como todas las personas que trabajan en una empresa u organización, conozcan cuáles son los beneficios de lograr la satisfacción del cliente, cómo definirla, cuáles son los niveles de satisfacción, cómo se forman las expectativas en los clientes y en qué consiste el rendimiento percibido, para que de esa manera, estén mejor capacitadas para coadyuvar activamente con todas las tareas que apuntan a lograr la tan anhelada satisfacción del cliente.

Según (Gonzalez, 2014), una definición del concepto de "Satisfacción del cliente" es posible encontrarla en la norma ISO 9000:2015 "Sistemas de gestión de la calidad — Fundamentos y vocabulario", que la define como la "percepción del cliente sobre el grado en que se han cumplido sus requisitos", aclarando además un aspecto muy importante sobre las quejas de los clientes: su existencia es un claro indicador de una baja satisfacción, pero su ausencia no implica necesariamente una

elevada satisfacción del cliente, ya que también podría estar indicando que son inadecuados los métodos de comunicación entre el cliente y la empresa, o que las quejas se realizan pero no se registran adecuadamente, o que simplemente el cliente insatisfecho, en silencio, cambia de proveedor.

También resulta aceptable definir a la satisfacción del cliente como el resultado de la comparación que de forma inevitable se realiza entre las expectativas previas del cliente puestas en los productos y/o servicios y en los procesos e imagen de la empresa, con respecto al valor percibido al finalizar la relación comercial.

Al considerar seriamente cualquiera de las dos definiciones, surge con nitidez la importancia fundamental que tiene para una organización conocer la opinión de sus clientes, lo que le permitirá posteriormente establecer acciones de mejora en la organización.

En resumen, puede definirse rápidamente el grado de satisfacción del cliente mediante la siguiente ecuación:

$$\text{Valor Percibido} - \text{Expectativas} = \text{Grado de satisfacción} \quad (5)$$

El Valor Percibido es aquel valor que el cliente considera haber obtenido a través del desempeño del producto o servicio que adquirió. Los aspectos más relevantes del Valor Percibido son los siguientes:

- Los determina el cliente, no la empresa.
- Se basa en los resultados que el cliente obtiene con el producto o servicio.



- Se sustenta en las percepciones del cliente, y no necesariamente en la realidad.
- Sufre el impacto de las opiniones de otras personas que influyen en el cliente.
- Depende del estado de ánimo del cliente y de sus razonamientos.

Dada su complejidad, el Valor Percibido puede ser determinado luego de una exhaustiva investigación que comienza y termina en el cliente.

Las Expectativas, en cambio, conforman las esperanzas que los clientes tienen por conseguir algo, y se producen por el efecto de una o más de estas cuatro situaciones:

- Promesas que hace la misma empresa acerca de los beneficios que brinda el producto o servicio.
- Experiencias de compras anteriores.
- Opiniones de amistades, familiares, conocidos y líderes de opinión (p.ej.: deportistas famosos).
- Promesas que ofrecen los competidores.

Ante los aspectos que dependen directamente de la empresa, ésta se encuentra con el dilema de establecer el nivel correcto de expectativas, ya que las expectativas demasiado bajas no atraen suficientes clientes; pero las expectativas demasiado altas generan clientes decepcionados luego de la compra.

Un detalle muy interesante a tener en cuenta es que la disminución en los índices de satisfacción del cliente no siempre está asociada a una disminución en la calidad de

los productos o servicios; en muchos casos, es el resultado de un aumento en las expectativas del cliente, algo atribuible a las actividades de marketing.

Alcanzar la satisfacción de los clientes es uno de los factores críticos para toda organización que pretenda diferenciarse de sus competidores y obtener mejores resultados en el negocio.

Seguidamente se presenta un decálogo de ideas para meditar:

- a) Compromiso con un servicio de calidad: toda persona de la organización tiene casi la obligación de crear una experiencia positiva para los clientes.
- b) Conocimiento del propio producto o servicio: transmitir claramente conocimientos de interés para el cliente acerca de las características del propio producto o servicio ayuda a ganar la confianza de éste.
- c) Conocimiento de los propios clientes: aprender todo lo que resulte posible acerca de los clientes para que se pueda enfocar el producto o servicio a sus necesidades y requerimientos.
- d) Tratar a las personas con respeto y cortesía: cada contacto con el cliente deja una impresión, sin importar el medio utilizado (e-mail, teléfono, personalmente, etc.). Esta impresión debe denotar siempre corrección en el trato.
- e) Nunca se debe discutir con un cliente: desde luego que no siempre tiene la razón, pero el esfuerzo debe orientarse a recomponer la situación. Muchos estudios demuestran que 7 de cada 10 clientes harían negocios nuevamente con las empresas que resuelven un problema a su favor.

- f) No hacer esperar a un cliente: los clientes buscan una rápida respuesta, por lo que es fundamental tratar ágilmente las comunicaciones y la toma de decisiones.
- g) Dar siempre lo prometido: una falla en este aspecto genera pérdida de credibilidad y de clientes. Si resulta inevitable un incumplimiento, se debe pedir disculpas y ofrecer alguna compensación.
- h) Asumir que los clientes dicen la verdad: aunque en ciertas ocasiones parezca que los clientes están mintiendo, siempre se les debe dar el beneficio de la duda.
- i) Enfocarse en hacer clientes antes que ventas: mantener un cliente es más importante que cerrar una venta. Está comprobado que cuesta seis veces más generar un cliente nuevo que mantener a los existentes.
- j) La compra debe ser fácil: la experiencia de compra en el punto de venta, página web, catálogo, etc., debe ser lo más fácil posible, y debe ayudar a los clientes a encontrar lo que estás buscando.

## 1.2. Formulación del problema

¿De qué manera el desarrollo de un software de análisis de sentimiento puede facilitar la evaluación de la satisfacción de los visitantes del sector turismo en Trujillo, 2018?

### **1.3. Objetivos**

#### **1.3.1. Objetivo general**

Implementar el Software de Análisis de Sentimientos para la satisfacción de los visitantes del sector turismo en la ciudad de Trujillo en base a la polaridad de comentarios en español en un fan page de Facebook, 2018.

#### **1.3.2. Objetivos específicos**

- Recolectar comentarios de los visitantes del sector turismo a través de las herramientas para desarrolladores de Facebook.
- Crear un algoritmo capaz de realizar técnicas de limpieza de elementos no necesarios en los comentarios. (pronombres, conectores, etc.),
- Mediante un algoritmo realizar el proceso de normalización del texto. Este proceso se divide en los siguientes subprocesos: Tokenización, lematización, stemming y tratamiento de negaciones.
- Calcular en porcentajes, tras el análisis de los comentarios, la polaridad de la opinión de los usuarios de Facebook visitantes del sector turismo.

### **1.4. Hipótesis**

El desarrollo de un software de análisis de sentimientos sí permite la evaluación de la satisfacción de los visitantes del Sector Turismo en la ciudad de Trujillo.

## CAPÍTULO II. METODOLOGÍA

### 2.1. Tipo de investigación

La presente investigación es de tipo transversal, ya que se determinará la polaridad de la satisfacción de los visitantes del sector turismo.

Para el desarrollo se aplicará el sistema de Análisis de sentimiento en una Fan Page de Facebook de un centro turístico de la ciudad de Trujillo y se analizarán los comentarios en español vertidos por los usuarios.

### 2.2. Población y muestra

Nuestra población serán los usuarios de Facebook visitantes del sector turismo durante el periodo de un mes (30 días aproximadamente). Ver Anexo 1.

El muestreo para población infinita, donde  $z = 0,95$ ;  $e = 0,05$ .

$$n = \frac{(z^2) * P * Q}{(e^2)}$$

$n$  = Cantidad teórica de elementos de la muestra.

$z$  = Valor estandarizado en función del grado de confiabilidad de la muestra calculada.

99% -----  $z = 2,58$

95% -----  $z = 1,96$

90% -----  $z = 1,64$

$e$  = Error asumido en el cálculo.

Para  $3 \leq n \leq 10$  -----  $e = 0.1$  (un error del 10%)

Para  $n > 10$  -----  $e = 0.05$  (un error del

5%)

$Q$  = Probabilidad de la población que no presenta las características.

$P$  = Probabilidad de la población que presenta las características.

$$n = \frac{(1.96^2) * \frac{0.5}{0.5}}{(0.1^2)} = 96.04$$

La muestra se tomará a 100 usuarios de Facebook del sector turismo de la ciudad de Trujillo

## 2.3. Procedimiento

### 2.3.1. Oportunidad de negocio

Este producto software facilitará a los negocios, en este caso en especial, a los Proyectos del sector turismo de la ciudad de Trujillo en el análisis de los comentarios registrados por los usuarios en las respectivas Fan Pages de Facebook y así conocer la polaridad de las opiniones y tomar medidas correctivas o fortalecer ciertos procesos.

### 2.3.2. Declaración del problema.

*Tabla 2: Declaración del problema*

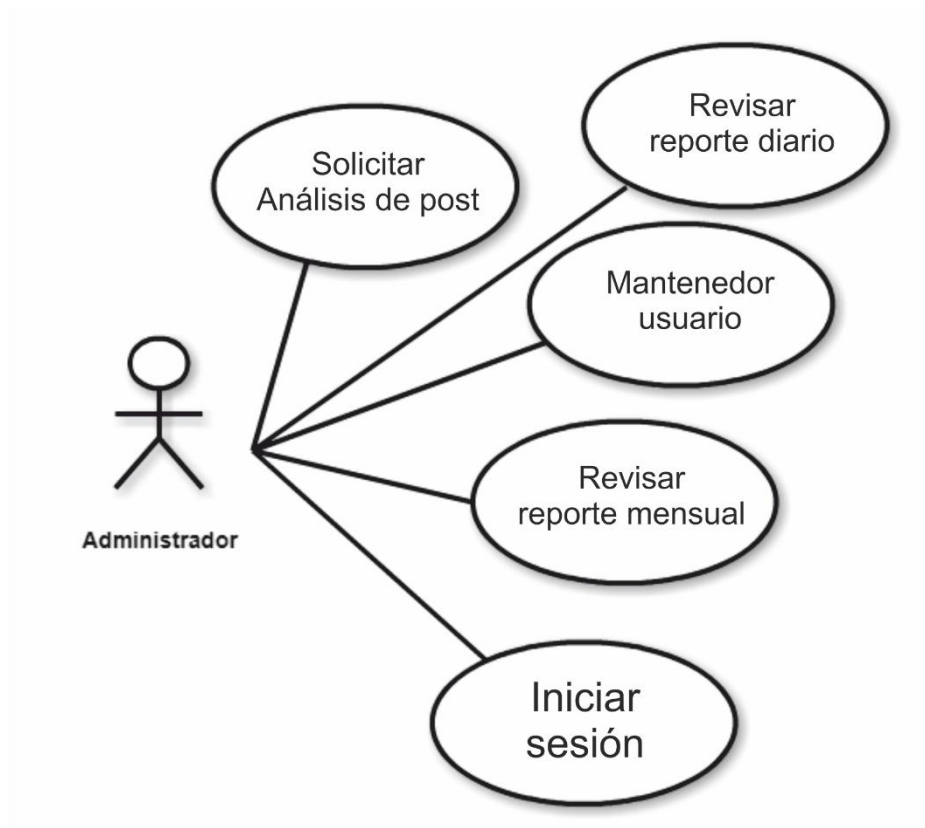
EL PROBLEMA	No se cuenta con una herramienta que realice el análisis de la polaridad y los sentimientos en los comentarios de las Fan Page de Facebook de los centros turísticos.
APECTA A	Centro turístico - visitantes
EL IMPACTO	Proceso lento

<p>ADECUADA  SOLUCIÓN SERÍA</p>	<p>Implementar un software que realice el análisis de los comentarios de las Fan Page de Facebook.</p>
---	--

*Fuente: Elaboración propia.*

### 2.3.3. Propuesta de desarrollo.

Se propone el desarrollo de un producto software especializado en cumplir con ciertas funcionalidades establecidas previamente, se tomó en cuenta para el proceso de desarrollo la utilización de la metodología XP, ya que contiene en su estructura ágil, estrategias, herramientas y documentos necesarios para lograr satisfactoriamente el desarrollo del proyecto.



### 2.3.4. Casos de uso.

Fuente: Elaboración propia.



*Ilustración 4: Modelo de Casos de Uso - Sistema*

Fuente: Elaboración propia.

### 2.3.5. Especificación de modelo de casos de uso.

*Tabla 3: Caso de Uso 01 - Mantenedor Usuario*



<b>IDENTIFICADOR:</b> CUS01		<b>NOMBRE:</b> Mantenedor Usuario	
<b>CATEGORÍA:</b> Mantenedor		<b>COMPLEJIDAD:</b> Media	<b>PRIORIDAD:</b> Media
<b>ACTORES:</b> Administrador			
<b>PROPÓSITO:</b> Administrar las credenciales de Usuario.			
<b>PRECONDICIÓN:</b> Haber iniciado sesión.			
<p><b>FLUJO BÁSICO:</b></p> <p>B1. El administrador selecciona la opción “<b>Mantenedor &gt; Usuario</b>” del Menú Principal.</p> <p>B2. El sistema muestra la interfaz “<b>Usuario</b>” con 2 opciones:</p> <p style="padding-left: 40px;">B2.1. El listado con los Usuarios existentes; con los siguientes campos de búsqueda:</p> <p style="padding-left: 80px;">Nombres, apellidos, nombre de usuario, teléfono, dirección, DNI y estado</p> <p style="padding-left: 40px;">B2.2. Botón para Registrar Usuario Nuevo.</p> <p>B3. El Administrador ingresa los datos solicitados para la búsqueda.</p> <p>B4. El Sistema filtra hasta mostrar el Usuario indicado.</p> <p>B5. El administrador Selecciona otra opción y el CUS finaliza.</p> <p><b>SUBFLUJOS:</b></p> <p><b>1. Registrar Usuario</b></p> <p>1.1. El Administrador selecciona la opción “<b>Registrar Usuario</b>”</p> <p>1.2. El sistema muestra una pantalla para registrar un nuevo Usuario:</p> <p style="padding-left: 40px;">Nombres, Apellidos, Fecha Nacimiento, DNI, Dirección, Teléfono, Celular y Activo. Además, habilita las opciones: Guardar y Cancelar.</p> <p>1.3. El Administrador ingresa los datos del Usuario.</p> <p>1.4. El Administrador selecciona la opción “Guardar”</p> <p>1.5. El sistema Valida los Datos Ingresados.</p> <p>1.6. El sistema genera un nuevo código del Usuario</p> <p>1.7. El sistema graba un nuevo registro de Usuario y muestra el Mensaje</p>			

“Se Creó el usuario Exitosamente.”

1.8. El sistema carga el Nuevo usuario en la Lista de Usuarios y el CUS finaliza.

**POSCONDICIÓN:**

- a. En el sistema quedará registrado el nuevo usuario.
- b. En el sistema quedará actualizado el registro del usuario.

**FLUJOS ALTERNATIVOS:**

Si los datos ingresados son nulos o inválidos, en los 2 sub-flujos, el sistema muestra un mensaje “Este Campo es obligatorio” en el campo faltante y los sub-flujos continúan en el paso 3.

*Fuente: Elaboración propia.*

Tabla 4: Caso de Uso 02 - Revisar Reporte Mensual

<b>IDENTIFICADOR:</b> CUS02	<b>NOMBRE:</b> Revisar Reporte Mensual	
<b>CATEGORÍA:</b> Reporte	<b>COMPLEJIDAD:</b> Media	<b>PRIORIDAD:</b> Media
<b>ACTORES:</b> Administrador		
<b>PROPÓSITO:</b> Visualizar un informe mensual de la satisfacción de los usuarios.		
<b>PRECONDICIÓN:</b> Haber iniciado sesión.		
<b>FLUJO BÁSICO:</b> B1. El Administrador revisa la gráfica de satisfacción mensual mostrada en la pantalla de inicio.		
<b>POSCONDICION:</b> El administrador puede solicitar un Informe diario de cualquier fecha (fecha actual, o fecha pasada*)  <i>*Posterior a la fecha de inicio de funcionamiento del Sistema de Análisis de Satisfacción de Usuarios.</i>		
<b>FLUJOS ALTERNATIVOS:</b> <b>A1. Sin resultados</b> <b>A1.1.</b> El primer día de funcionamiento de Sistema, no mostrará ninguna gráfica.		

*Fuente: Elaboración propia.*

Tabla 5: Caso de Uso 03 - Revisar Reporte Diario

<b>IDENTIFICADOR:</b> CUS03	<b>NOMBRE:</b> Revisar Reporte Diario	
<b>CATEGORÍA:</b> Reporte	<b>COMPLEJIDAD:</b> Media	<b>PRIORIDAD:</b> Media
<b>ACTORES:</b> Administrador		
<b>PROPÓSITO:</b> Visualizar un informe diario de la satisfacción de los usuarios.		
<b>PRECONDICIÓN:</b> Haber iniciado sesión.		

**FLUJO BÁSICO:**

- B1. El Administrador se encuentra en la página principal.
- B2. El sistema muestra un TextBox y un botón Buscar, donde solicita un único dato de entrada, Fecha, para iniciar una búsqueda de la satisfacción diaria del visitante.
- B3. El Administrador ingresa el dato solicitado (Fecha) y selecciona el botón buscar.
- B4. El sistema una gráfica del porcentaje de satisfacción del visitante en la fecha dada.

**POSCONDICION:**

Se visualiza la gráfica de satisfacción del día solicitado por el usuario mediante la fecha.

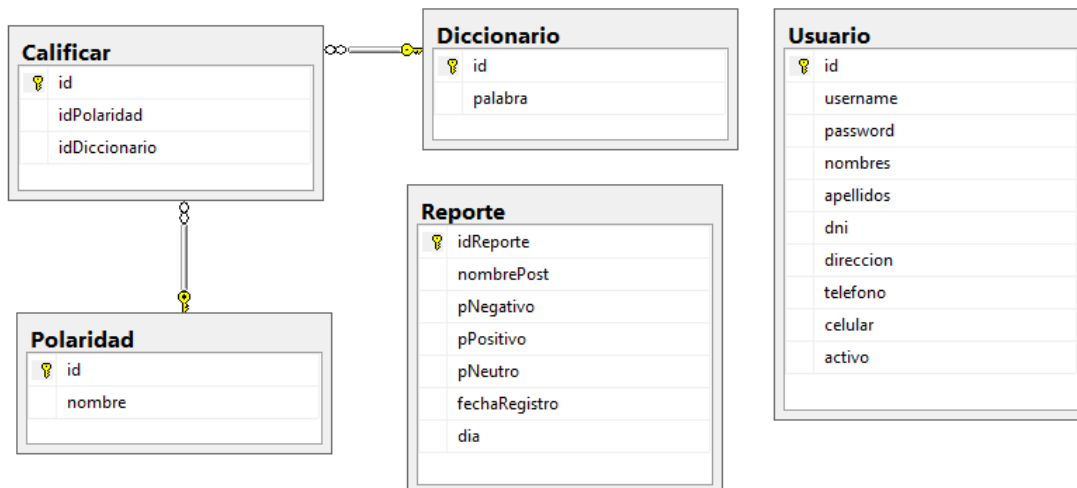
**FLUJOS ALTERNATIVOS:**

**A1. Sin resultados**

**A1.1.** Luego del paso B3 del flujo básico, si la fecha brindada por el Administrador no existe, el sistema mostrará un mensaje: No se encontraron resultados para la fecha proporcionada.

*Fuente: Elaboración propia.*

**2.3.6. Modelo físico de la base datos**



*Ilustración 5: Diagrama Físico de la Base de Datos: AnalisisSentimiento*

*Fuente: Elaboración propia.*

### 2.3.6.1. Diccionario de datos

Nombre de la Base de Datos: AnalisisSentimiento.

#### a. Nombre de Tabla: Calificar.

Tabla 6: Descripción de la Tabla Calificar.

Nombre de Campo	Descripción	Tipo	Capacidad	Requerido	Restricción
id	Llave primaria	Autonumérico	-	Sí	Clave primaria y único.
idPolaridad	Id de la polaridad	Entero	-	Sí	Llave foránea
idDiccionario	Id de la palabra	Entero	-	Sí	Llave foránea

*Fuente: Elaboración propia.*

#### b. Nombre de Tabla: Diccionario.

Tabla 7: Descripción de la Tabla Diccionario.

Nombre de Campo	Descripción	Tipo	Capacidad	Requerido	Restricción
id	Llave primaria	Autonumérico	-	Sí	Clave primaria y único.
palabra	Palabras componentes del diccionario	Cadena	50	Sí	Sólo letras

*Fuente: Elaboración propia.*

**c. Nombre de Tabla: Polaridad.**

*Tabla 8: Descripción de la Tabla Polaridad.*

Nombre de Campo	Descripción	Tipo	Capacidad	Requerido	Restricción
id	Llave primaria	Autonumérico		Sí	Clave primaria y único.
nombre	Nombre de la polaridad	Cadena	50	Sí	Sólo letras

*Fuente: Elaboración propia.*

**d. Nombre de Tabla: Usuario.**

*Tabla 9: Descripción de la Tabla Usuario.*

Nombre de Campo	Descripción	Tipo	Capacidad	Requerido	Restricción
id	Llave primaria	Autonumérico		Sí	Clave primaria y único.
username	Nombre de usuario	Cadena	50	Sí	-
password	Contraseña del usuario	Cadena	50	Sí	-
nombres	Nombre del usuario	Cadena	50	No	-
apellidos	Apellidos del usuario	Cadena	50	No	-
dni	DNI del usuario	Cadena	8	No	Sólo números
direccion	Dirección del usuario	Cadena	50	No	-

telefono	Teléfono del usuario	Cadena	9	No	-
celular	Celular del usuario	Cadena	9	No	-
activo	Indicador de actividad del registro	bit	-	Sí	-

*Fuente: Elaboración propia.*

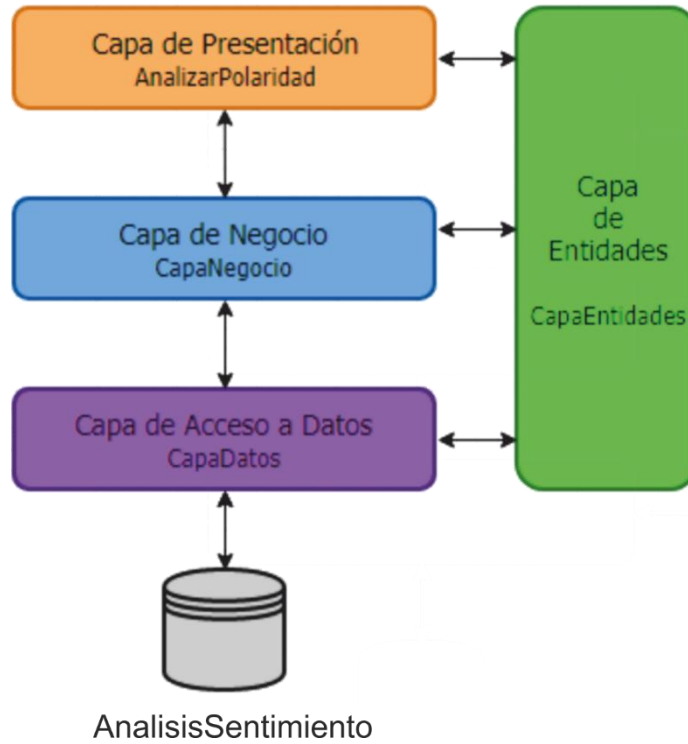
**e. Nombre de Tabla: Reporte.**

*Tabla 10: Descripción de la Tabla Reporte.*

Nombre de Campo	Descripción	Tipo	Capacidad	Requerido	Restricción
idReporte	Llave primaria	Autonumérico	-	Sí	Clave primaria y único.
nombrePost	Nombre de post	Cadena	50	Sí	-
pNegativo	Porcentaje de post o comentarios negativos	int	-	Sí	Sólo números
pPositivo	Porcentaje de post o comentarios positivos	int	-	Sí	Sólo números
pNeutro	porcentaje de post o comentarios neutros	int	-	Sí	Sólo números
fechaRegistro	Fecha del post	date	-	Sí	Sólo fecha
dia	Resumen de la polaridad del día	varchar (1)	-	Sí	Sólo caracteres

*Fuente: Elaboración propia.*

### 2.3.7. Arquitectura de software



*Ilustración 6: Diagrama de Arquitectura de Software*

*Fuente: Elaboración propia.*



### 3.1.1. Atributo del sistema

#### a) Usabilidad

El sistema debe ser sencillo, fácil de utilizar e intuitivo. El usuario debe ser capaz de comprender todas las funciones del sistema, con una capacitación básica.

#### b) Fiabilidad

El sistema debe ser tolerante a fallos, notificando mediante alertas algún tipo de error sucedido mientras esté en uso. Así mismo, el sistema deberá seguir funcionando correctamente luego de ser atendida la alerta.

#### c) Disponibilidad

El sistema debe estar disponible 24 horas al día, durante todo el año. El uso de políticas de respaldo es esencial, tales como, replicando base de datos y recuperación ante fallos.

### 3.1.2. Bussiness core

El proceso Core del S.A.S. es el que realiza la recopilación y análisis de los comentarios de un post.

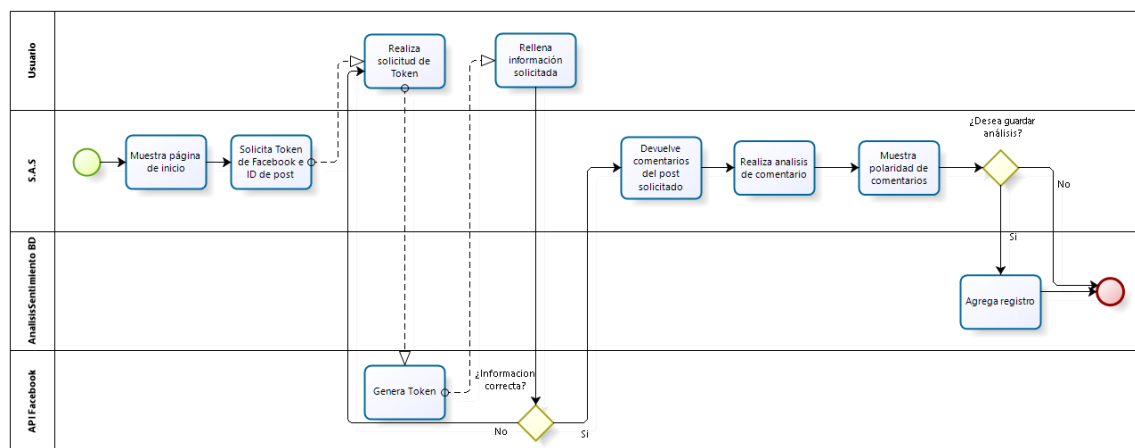


Ilustración 7: Proceso de Análisis de post

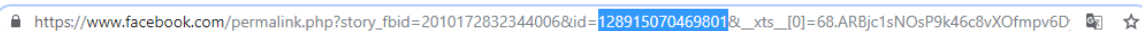
Fuente: Elaboración propia.

El proceso Core del Sistema de Análisis de Sentimiento (S.A.S) inicia cuando el administrador del sistema realiza una solicitud de análisis de post. El sistema solicita 2 parámetros: un Token de acceso y el ID del post que se requiere analizar.

- a) Para obtener el Token de acceso se debe ingresar al explorador gráfico de la Api de Facebook en la siguiente dirección:

[https://developers.facebook.com/tools/explorer/145634995501895/?  
method=GET&path=10154774241137151&version=v2.4.](https://developers.facebook.com/tools/explorer/145634995501895/?method=GET&path=10154774241137151&version=v2.4)

- b) Para obtener el ID del post de Facebook se debe hacer lo siguiente:
- Ingresar al administrador de páginas y dirigirse a la sección Posts.
  - Seleccionar el post requerido.
  - Hacer clic en la fecha y hora del post.
  - De la barra de direcciones, se debe seleccionar los 15 números del ID, tal como lo indica la siguiente imagen.



[https://www.facebook.com/permalink.php?story\\_fbid=2010172832344006&id=128915070469801&\\_xts\\_\\_\[0\]=68.ARBjc1sNOsP9k46c8vXOfmpv6D](https://www.facebook.com/permalink.php?story_fbid=2010172832344006&id=128915070469801&_xts__[0]=68.ARBjc1sNOsP9k46c8vXOfmpv6D)

*Ilustración 8: Ejemplo enlace post*

***Fuente: Elaboración propia.***

Luego de recibir los comentarios del post el S.A.S. realizará el análisis de los comentarios bajo el siguiente algoritmo:

```
List<string> separarPalabras(string cadena)
{
    string[] superlativos = {"esimo", "azo", "aso", "!", "?", ".", "(", ")", ";", "#"};

    for(int i=0; i<superlativos.Length;i++){
        cadena = cadena.Replace(superlativos[i], "");
    }
    return cadena.Split(' ').ToList<string>();
}
```

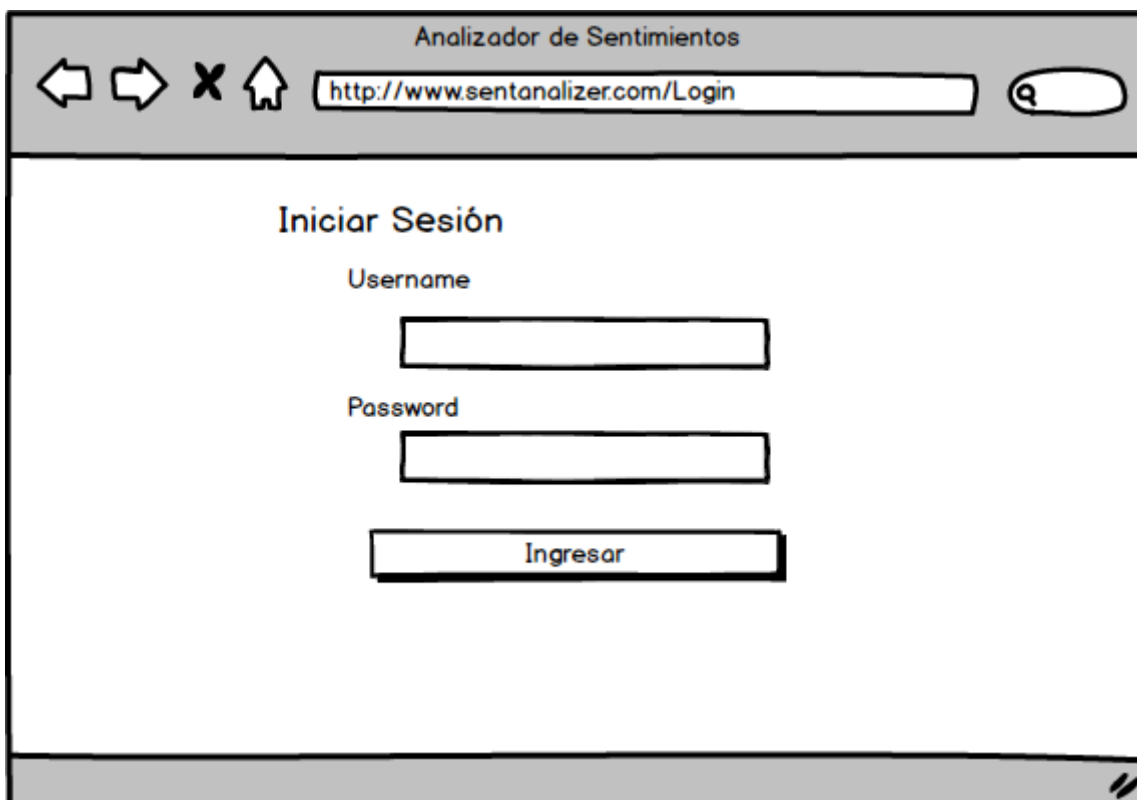
*Ilustración 9: Función SepararPalabras*

***Fuente: Elaboración propia.***

### 3.1.3. Prototipos

Las interfases del S.A.S serán las siguientes:

#### a. Prototipo 01: Login

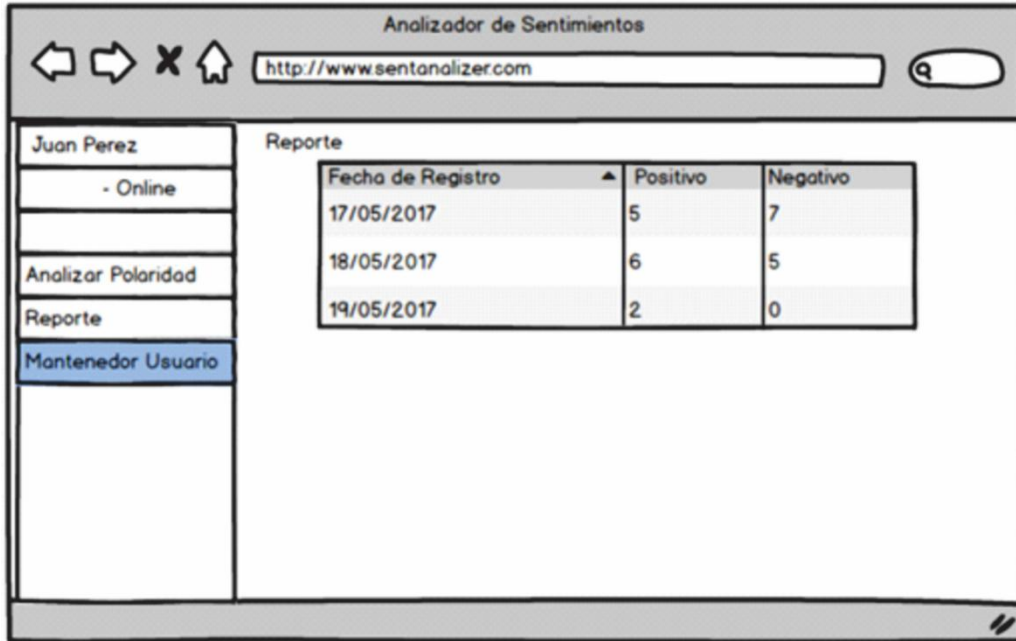


The image shows a wireframe of a login page for a web application. At the top, there is a browser window header with the title "Analizador de Sentimientos". Below the title is a navigation bar with icons for back, forward, close, and home, followed by a search bar containing the URL "http://www.sentanalizer.com/Login". The main content area is titled "Iniciar Sesión" and contains three input fields: "Username", "Password", and a button labeled "Ingresar". The "Ingresar" button is highlighted with a thick border. The entire page is enclosed in a rectangular frame representing a browser window.

*Ilustración 10: Prototipo de Login.*

*Fuente: Elaboración propia.*

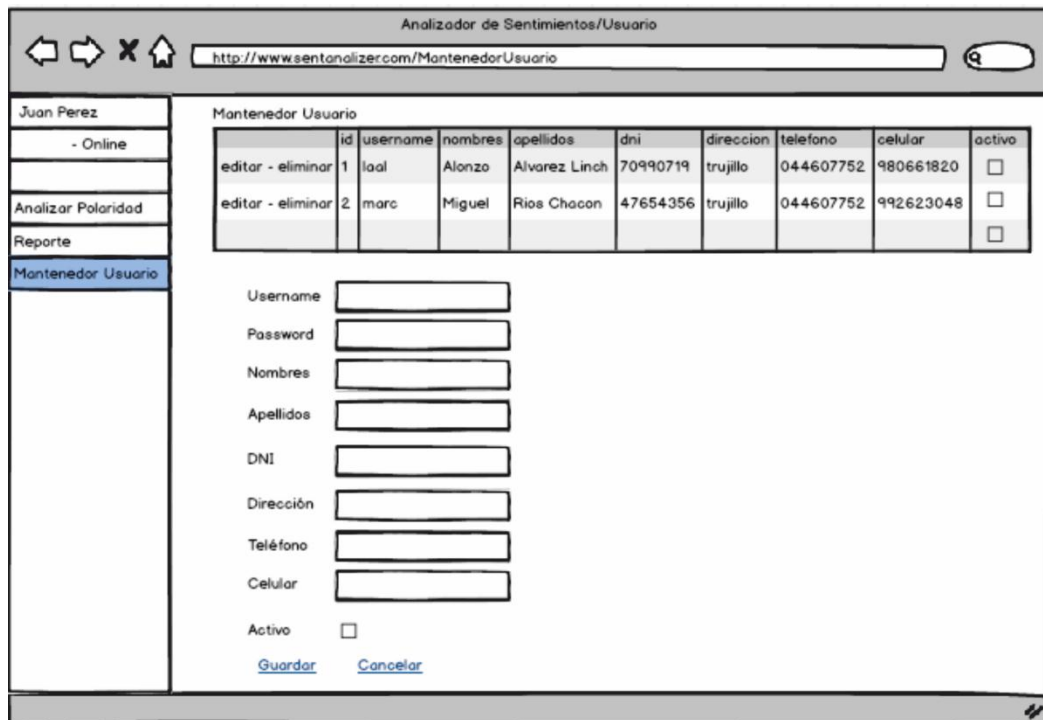
**b. Prototipo 02: Página Principal**



*Ilustración 11: Prototipo de Página Principal.*

*Fuente: Elaboración propia.*

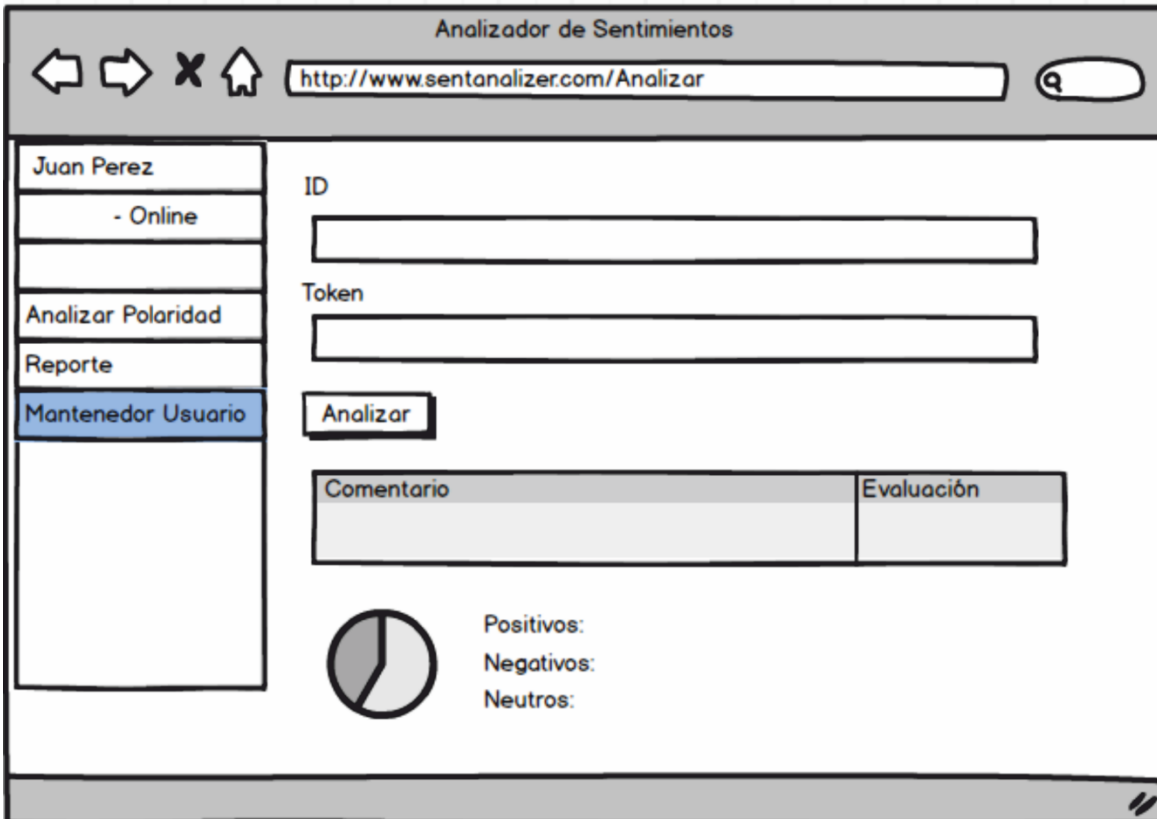
**c. Prototipo 03: Mantenedor Usuario**



*Ilustración 12: Prototipo de Mantenedor Usuario.*

*Fuente: Elaboración propia.*

**d. Prototipo 04: Analizar Comentario**



Analizador de Sentimientos

http://www.sentanalizer.com/Analizar

Juan Perez

- Online

Analizar Polaridad

Reporte

Mantenedor Usuario

ID

Token

Analizar

Comentario	Evaluación
------------	------------

Positivos:

Negativos:

Neutros:

*Ilustración 13: Prototipo de Analizar Comentario*

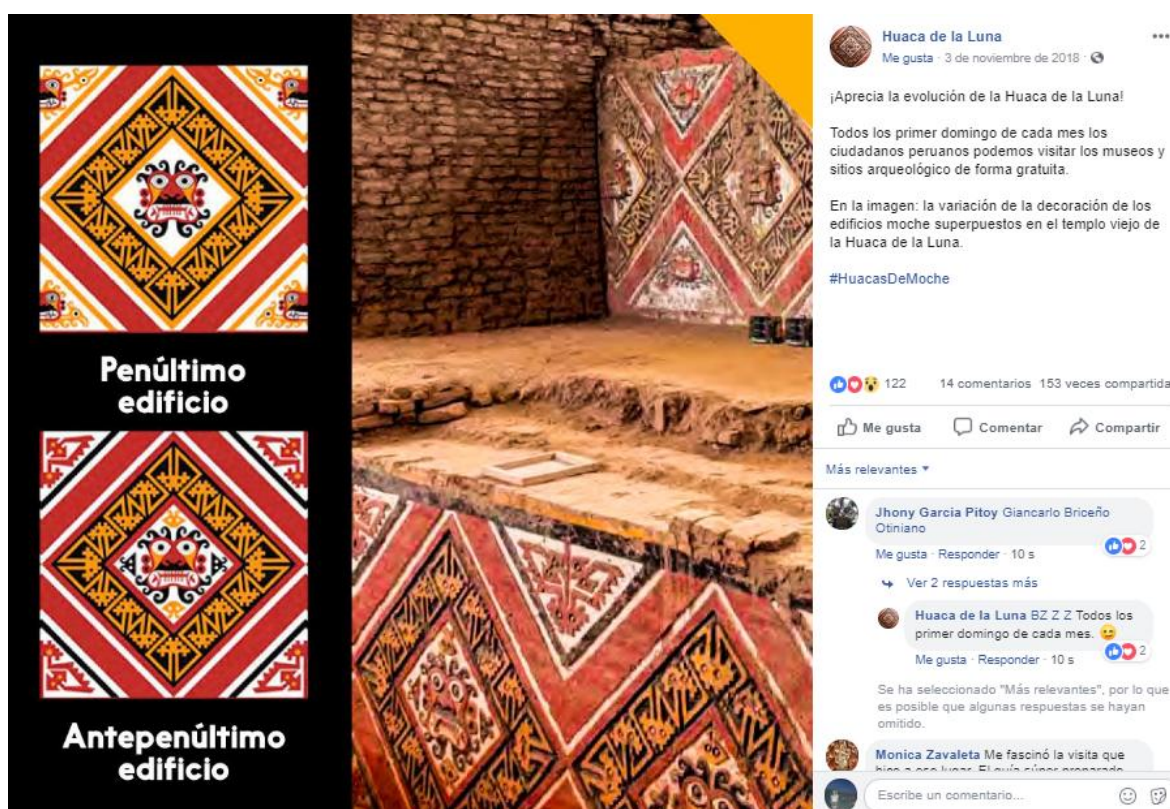
*Fuente: Elaboración propia.*

### CAPÍTULO III. RESULTADOS

Los siguientes datos fueron obtenidos de la Fan Page de Facebook del Proyecto Huaca de la Luna (<https://www.facebook.com/huacasdemoche/>).

Se analizaron los comentarios de las publicaciones realizadas durante el 3 de noviembre al 10 de diciembre (38 días). Los resultados son los siguientes:

#### a) **Publicación #1:** Publicación realizada el 3 de noviembre del 2018



*Ilustración 14: Publicación #1.*

**Fuente:**

<https://www.facebook.com/huacasdemoche/photos/a.395169847211212/2045232185538295/?type=3&theater>



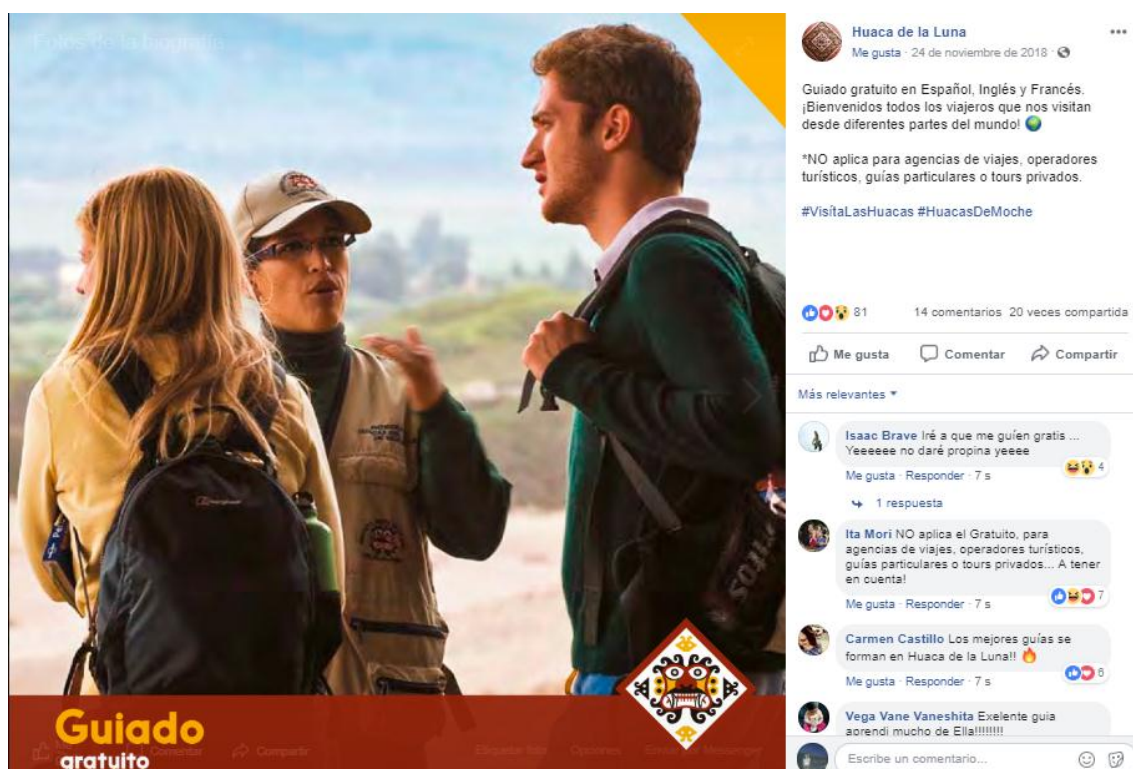
“SOFTWARE DE ANÁLISIS DE SENTIMIENTO EN LA  
EVALUACIÓN DE LA SATISFACCIÓN DE LOS VISITANTES  
DEL SECTOR TURISMO DE TRUJILLO, 2018”

Comentario	Likes	Evaluación
Excelente , asombroso, super recomendable e imperdible!!! Deberían difundir mucho mas todo lo bello q tiene Peru y estoy descubriendo en cada viaje. 😊	0	Neutro
Me fascinó la visita que hice a ese lugar. El guía súper preparado, lástima que no anoté su nombre. 🍌	0	Positivo
Maritere V. Lovatón Juan CC vamos mañana?	0	Neutro
Giancarlo Briceño Otiniano	0	Neutro
<a href="https://media3.giphy.com/media/5YwVkvCW0MxTG/giphy.gif?cid=4bf119fc5bddc1f07631737a732ddea9">https://media3.giphy.com/media/5YwVkvCW0MxTG/giphy.gif?cid=4bf119fc5bddc1f07631737a732ddea9</a>	0	Neutro
Todo esta bien pero los arqueólogos NO podemos guiar ...jjj cosas de locos., ocurre solo en el PERU...!	0	Positivo
M	0	Neutro
Positivos : 2 Negativos : 0 Neutros : 5		

Ilustración 15: Resultados obtenidos de la publicación #1.

**Fuente: Resultados obtenidos del Software de Análisis de Sentimiento**

**b) Publicación #2: Publicación realizada el 24 de noviembre del 2018**



**Huaca de la Luna**  
Me gusta · 24 de noviembre de 2018 · 🌐

Guiado gratuito en Español, Inglés y Francés.  
¡Bienvenidos todos los viajeros que nos visitan desde diferentes partes del mundo! 🌍

\*NO aplica para agencias de viajes, operadores turísticos, guías particulares o tours privados.

#VisitaLasHuacas #HuacasDeMoche

👍❤️👍 81 14 comentarios 20 veces compartida

Me gusta Comentar Compartir

Más relevantes ▾

Isaac Brave Iré a que me guíen gratis ...  
Yeaaaa no daré propina yeaaa  
Me gusta · Responder · 7 s 🙄🙄 4  
1 respuesta

Ita Mori NO aplica el Gratuito, para agencias de viajes, operadores turísticos, guías particulares o tours privados... A tener en cuenta!  
Me gusta · Responder · 7 s 🙄🙄 7

Carmen Castillo Los mejores guías se forman en Huaca de la Luna!! 🙄  
Me gusta · Responder · 7 s 🙄🙄 8

Vega Vane Vaneshita Excelente guia aorendi mucho de Ella!!!!!!  
Escribe un comentario...

Ilustración 16: Publicación #2

**Fuente:**  
<https://www.facebook.com/huacasdemoche/photos/a.395169847211212/2070602626334584/?type=3&theater>

Comentario	Likes	Evaluación
NO aplica el Gratuito, para agencias de viajes, operadores turísticos, guías particulares o tours privados... A tener en cuenta!	0	Neutro
En todo el mundo, cuando uno viaja y queda contento con el servicio que ofrece un hotel, restaurante o guía de turismo se acostumbra dejar una contribución voluntaria por el buen servicio prestado, ya depende de la persona o visitante que valore o no al Guía de turismo, no son improvisados, han estudiado una profesión y siempre están capacitándose y con sus propinas pueden estudiar otros idiomas y mejorar su servicio.	0	Positivo
Exelente guía aprendí mucho de Ella!!!!!!!	0	Neutro
Endora, sos vos!!! Jajaja esa guía pone de cabeza la huaca, una excelente guía Ita Mori	0	Positivo
Mí flaca!!! Huaca de la Luna💖	0	Neutro
Creo que conozco a la guía 🤔🤔 Ita Mori	0	Neutro
Hey Ita Mori 🤔🤔	0	Neutro
for free!	0	Neutro
Ita Mori	0	Neutro
Los mejores guías se forman en Huaca de la Luna!! 🔥	0	Neutro
La propina es opcional ..pero hay q tenerlo muyyyy en cuenta cuando se da un buen servicio. La Huaca de la Luna da gusto verla cada año dando un mejor servicio al turista.	0	Positivo
Iré a que me guíen gratis ... Yeeeeee no daré propina yeeee	0	Positivo
Positivos : 4 Negativos : 0 Neutros : 8		

*Ilustración 17 : Resultados obtenidos de la publicación #2.*

**Fuente: Resultados obtenidos del Software de Análisis de Sentimiento**



c) **Publicación #3:** Publicación realizada el 10 de diciembre del 2018



Ilustración 18: Publicación #3

Fuente:

<https://www.facebook.com/huacasdemoche/photos/a.395169847211212/2094058677322312/?type=3&theater>

Comentario	Likes	Evaluación
Tuve la suerte de conocer Museo Tumbas Reales, Huaca Rajada, El Brujo, Chan Chan y Huacas del Sol y la Luna!! Uno más maravilloso que el otro!! Pero mi favorito sin dudas es El Brujo <3 Saludos desde Argentina!	0	Positivo
Conozco todo los lugares que están en el mapa, son sitios que vale la pena visitarlos las veces que se puedan ;)	0	Positivo
Nunca llegaron a Chile felizmente, por tanto, deberían devolver los sureños algunas piezas.	0	Neutro
Me falta conocer Tumbas reales del Señor de Sipán, Chan Chan y El Brujo 😊	0	Neutro
Ya conosco la parte de La Libertad, me falta Lambayeque 🤔🤔	0	Neutro
Seis 😊 hay que regresar, la riqueza cultural de Perú es un tema apasionante.	0	Positivo
Listo ya conocí todo 🙌 Me encantó! Como no amar el norte ❤️	0	Neutro
Que rico norte!	0	Neutro
Visiten Pakatnamu en Guadalupe 😊	0	Neutro
Carlos Insulza Evelyn Castro Poveda Laurita Paz para el viaje	0	Neutro
Christian amor...	0	Neutro
Alexander :3	0	Neutro
Angie Kriztinna	0	Neutro
Mery Palomares	0	Neutro
Daniel	0	Neutro
Moche es de best culture	0	Neutro
Positivos : 3 Negativos : 0 Neutros : 13		

*Ilustración 19: Resultados obtenidos de la publicación #3.*

**Fuente: Resultados obtenidos del Software de Análisis de Sentimiento**

## CAPÍTULO IV. DISCUSIÓN Y CONCLUSIONES

### 4.1 Discusión

El presente documento tiene como propósito realizar un análisis de sentimientos en base a los comentarios vertidos en una FanPage de Facebook. En una primera etapa, los comentarios fueron extraídos del post realizado a través de la API de Facebook, seguidamente se realizó un proceso de limpieza de elementos no relevantes. Los elementos relevantes fueron contrastados y polarizados basándose en el corpus construido. Del total de elementos polarizados en las tres publicaciones realizadas se obtuvo el siguiente el porcentaje de polaridad:

*Tabla 11: Polarización de publicaciones*

Polarización de publicaciones			
	Publicación #1	Publicación #2	Publicación #3
Positivo	28.6%	33.4%	18.75%
Neutro	71.4%	66.6%	81.25%
Negativo	0%	0%	0%

*Fuente: Elaboración propia*

Una vez que fueron obtenidos los comentarios a través de la API de Facebook, el texto fue contrastado con el listado de palabras polarizadas contenido en el Corpus, la polaridad de cada palabra es guardada. Este procedimiento es también utilizado por (Kogan , Roger, Buccella, & Mattar, 2014), y su agente analizador, el cuál se encarga de asignar la polaridad correcta a cada una de las palabras encontradas.

En nuestro análisis no se consideró el dominio, el cual puede lograr afectar en la polaridad de ciertas palabras, En el informe de (Moreno Ortiz, Pineda Castillo, & Hidalgo García, 2010), se resaltó la falta de uso de un dominio, sin embargo, se consideró una clasificación de polaridades diferente, cuatro polaridades en total.

## 4.2 Conclusiones

Al finalizar el presente trabajo de investigación, se llegaron a las siguientes conclusiones:

- a) Con el uso del Software de análisis de sentimientos, se logra conocer la satisfacción de los asistentes a los establecimientos turísticos y se permite conocer las oportunidad, fortalezas y debilidades de los establecimientos turísticos.
- b) Aplicando el software de análisis de sentimientos, se logra reducir el tiempo de análisis de encuestas físicas.
- c) Aplicando el software de análisis de sentimientos, se logra reducir el costo de impresión de encuestas físicas.
- d) Se logran implementar políticas ambientalistas, asignándose una nueva meta de consumir 60% menos papel que en años anteriores. *Ver Anexo 2.*

## REFERENCIAS

- Aarli, G., & Martínez López, J. (2008). *Elementos del entorno en la lematización de las locuciones*. Alicante: Universidad de Bergen y Universidad Noruega de Ciencias Empresariales.
- Beck, K., & Gamma, E. (2000). *Extreme Programming Explained: Embrace Change*. Addison-Wesley Professional.
- Carreras Montoto, O. (15 de Marzo de 2012). *Estándares formales de usabilidad y su aplicación práctica en una evaluación heurística*. Obtenido de <https://olgacarreras.blogspot.com/2012/03/estandares-formales-de-usabilidad-y-su.html>
- Ceballos, J. (2008). *Microsoft C#. Lenguaje y aplicaciones, 2da Edición*. México: Alfaomega Grupo Editor.
- Deitel, H., & Deitel, P. J. (2004). *Como programar en Java*. México: Pearson Education.
- Denove, C., & Power, J. (2006). *La Satisfacción del Cliente*. Portfolio.
- Dubiau, L. (2013). *Procesamiento de Lenguaje Natural en sistema de análisis de sentimientos*. Buenos Aires.
- Fermín L. Cruz, J. A. (2008). *Clasificación de documentos basada en la opinión y experimentos*. Sevilla.
- Fernández, J., Boldrini, E., & Gómez, J. (2011). *Análisis de Sentimientos y Minería de Opiniones: el corpus EmotiBlog*. Alicante: Universidad de Alicante.
- Fowler, M. (Mayo de 2004). *Is Design Dead?* Obtenido de <http://www.martinfowler.com/articles/designDead.html>
- Francia, J. (25 de Setiembre de 2017). *Scrum.org*. Obtenido de <https://www.scrum.org/resources/blog/que-es-scrum>
- García Ruesta, M. (26 de Mayo de 2011). *Medición de la satisfacción del cliente en una empresa de retail*. Piura, Piura, Perú: Repositorio institucional PIRHUA – Universidad de Piura.
- González Morcillo, C. (2011). *Lógica Difusa Una introducción práctica Técnicas de Softcomputing*.

- Gonzalez, H. (15 de Setiembre de 2011). *COMO MEDIR LA SATISFACCIÓN DEL CLIENTE*. Obtenido de <https://calidadgestion.wordpress.com/tag/satisfaccion-del-cliente/>
- Gortazar Bellas, F., Martínez Unanue, R., & Fresno Fernández, V. (2016). *LENGUAJES DE PROGRAMACIÓN Y PROCESADORES*. Madrid: Editorial Universitaria Ramón Areces.
- Harrison, P. (s.f.). *Evolutionary Programming*. Obtenido de <http://www.devcentre.org/research/evoprogramming.htm>
- Javi Fernández, E. B.-B. (2011). *Análisis de Sentimientos y Minería de Opiniones: el corpus EmotiBlog*. Sevilla.
- Joyanes Aguilar, L., & Zahonero Martinez, I. (2005). *Programación en C. Metodología, algoritmos y estructura de datos*. Madrid: McGraw-Hill / Interamericana España, S.A.U.
- Kernighan, B. W., & Ritchie, D. M. (1978). *The C Programming Language*. Prentice Hall.
- Letelier Torres, P., & Sánchez López, E. (2003). *Metodologías Ágiles en el Desarrollo de Software, Alicante. España*. Alicante: Grupo ISSI.
- Librado Jacobo, H. (Enero de 2016). Análisis automático de opiniones de productos en redes sociales. Enero, Ciudad de México, México.
- Maas, A., Ng, A., & Potts, C. (2011). *Multi-dimensional sentiment analysis with learned representations*. Standford: Stanford Computer Science and Stanford Linguistics.
- Martínez-Cámara, E., Martín-Valdivia, M., & Ureña-López, L. (2015). *Polarity classification for Spanish tweets using the COST corpus*. Jaén: Universidad de Jaén.
- Ministerio de Comercio Exterior y Turismo. (2016). *Medición económica del Turismo*. Obtenido de [https://www.mincetur.gob.pe/wp-content/uploads/documentos/turismo/publicaciones/MEDICION\\_ECONOMICA\\_TURISMO\\_ALTA.pdf](https://www.mincetur.gob.pe/wp-content/uploads/documentos/turismo/publicaciones/MEDICION_ECONOMICA_TURISMO_ALTA.pdf)
- OMT. (28 de Noviembre de 2018). *UNWTO Organización Mundial de Turismo*. Obtenido de <http://www2.unwto.org/es/press-release/2018-11-28/el-turismo-lider-entre-los-principales-sectores-economicos-mundiales>

- Parasuraman, A., Zeithaml, V., & L., L. (1985). *A Conceptual Model of Service Quality and Its Implications for Future Research*. *Journal of Marketing*. American Marketing Association.
- Pereira, E. (31 de Julio de 2018). *El turismo puede ser el principal renglón de la economía*. Obtenido de Semana: <https://www.semana.com/contenidos-editoriales/colombiamencantado/articulo/el-turismo-puede-ser-el-principal-renglon-de-la-economia/577409>
- Perez Araujo, S. (2016). Análisis de Sentimientos. En S. Perez Araujo. Obtenido de Perez, S. Araujo, M. (2016). Análisis de Sentimientos.
- Potts, C. (2011). *Sentiment Analysis Symposium*.
- PowerData. (26 de Mayo de 2016). *PowerData*. Obtenido de <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/que-es-un-gestor-de-datos-y-para-que-sirve>
- Robles, G., & Ferrer, J. (2002). *Programación eXtrema y Software*. Madrid: Spalinux.
- Rodriguez, J. (4 de Octubre de 2012). *Procesamiento de Lenguaje Natural*. Obtenido de <http://pdln.blogspot.com/2012/10/normalizacion-del-texto-normalizacion.html>
- Rouse, M. (Enero de 2015). *TechTarget*. Obtenido de <https://searchdatacenter.techtarget.com/es/definicion/SQL-Server>
- Ruiz Larrocha, E. (2017). *NUEVAS TENDENCIAS EN LOS SISTEMAS DE INFORMACION*. Madrid: EDITORIAL UNIVERSITARIA RAMON ARECES.
- Santana, O., Pérez, J., Carreras, F., & Duque, J. (2014). *FLANOM: Flexionador y lematizador automático de formas*. Las Palmas: Universidad de las Palmas de Gran Canaria.
- Schwaber, K., & Sutherland, J. (2016). *MANAGEMENT BUSINESS & MANAGEMENT SPECIAL TOPICS*. ANAGEMENT BUSINESS & MANAGEMENT SPECIAL TOPICS: John Wiley & Sons.
- Scrum vs XP*. (s.f.). Obtenido de <https://iswugxp.wordpress.com/xp-vs-scrum/>
- Thibaud, C. (2006). *MySQL 5*. Barcelona: Ediciones ENI.
- Thompson, I. (2006). *La Satisfacción del Cliente*.
- Tschohl, J. (2007). *Achieving Excellence Through Customer Service*. Minneapolis: Best Sellers Publishing.

*Turismo, el nuevo motor de la economía colombiana.* (04 de Noviembre de 2015).

Obtenido de Semana: <https://www.semana.com/economia/articulo/turismo-el-nuevo-motor-de-la-economia-colombiana/423615-3>

Turney, P. (2002). *Proceedings of the Association for Computational Linguistics.*

Vindas, S. B. (27 de Octubre de 2005). *Economía y sociedad.* Obtenido de <http://www.revistas.una.ac.cr/index.php/economia/article/view/635/570>



## ANEXOS

### 1.1. Anexo 1: Presupuesto

Tabla

12:

Recursos Humanos	
Nombres	Remuneración
Alvarez Linch, Luis Alonzo Estudiante de Ing. De Sistemas Computacionales	S/. 2.000,00
Ríos Chacón, Miguel Ángel Estudiante de Ing. De Sistemas Computacionales	S/. 2.000,00
Leiva Via, Geancarlo Asesor, Ing. Computación y Sistemas	S/. 2.300,00
<b>Total</b>	<b>S/. 6.300,00</b>

*Presupuesto de Recursos Humanos.*

*Fuente: Elaboración propia.*

Tabla 13: Presupuesto Recursos Materiales.

Recursos Materiales				
Materiales	Cantidad	Unidad	Costo Unitario	Sub-Total
Lapiceros	3	Unidad	S/. 1,00	S/. 3,00

Papel Bond A4 80g 500h	1	Millar	S/. 12,00	S/. 24,00
Folders Manila	20	Unidad	S/. 0,50	S/. 10,00
Resaltador	2	Unidad	S/. 2,00	S/. 4,00
<b>Total</b>				<b>S/. 41,00</b>

*Fuente: Elaboración propia.*

*Tabla 14: Presupuesto de Servicios*

<b>Servicios</b>			
<b>Descripción</b>	<b>Tiempo (Meses)</b>	<b>Costo Mensual</b>	<b>Sub-Total</b>
Teléfono + Internet 15 Mbps	12	S/. 179,00	S/. 2.148,00
Movilidad	Variado	S/. 300,00	S/. 300,00
Impresiones	Variado	S/. 0,50	S/. 80,00
Energía Eléctrica	12	S/. 80,00	S/. 960,00
<b>Total</b>			<b>S/. 3.488,00</b>

*Fuente: Elaboración propia.*

*Tabla 15:*

<b>Equipos</b>			
<b>Descripción</b>	<b>Cantidad</b>	<b>Costo</b>	<b>Sub-Total</b>
Laptop Core i3	1	S/. 1.600,00	S/. 1.600,00
Laptop Core i5	1	S/. 3.200,00	S/. 3.200,00
Memoria USB 32GB	2	S/. 75,00	S/. 150,00
<b>Total</b>			<b>S/. 4.950,00</b>

*Presupuesto de Equipos*

*Fuente: Elaboración propia.*

*Tabla 16: Presupuesto de Software*

<b>Software</b>		
<b>Descripción</b>	<b>Costo</b>	<b>Sub-Total</b>
MS Project 2013	Gratuito	Gratuito
MS Office 2013	Gratuito	Gratuito
Windows 10	Gratuito	Gratuito
<b>Total</b>		<b>S/. 0,00</b>

*Fuente: Elaboración propia.*

*Tabla*

<b>Consolidado de Presupuesto</b>	
<b>Recursos</b>	<b>Sub-Total</b>
Recursos Humanos	S/. 6.300,00
Recursos Materiales	S/. 41,00
Servicios	S/. 3.488,00
Equipos	S/. 4.950,00
Software	S/. 0,00
<b>Total</b>	<b>S/. 14.779,00</b>

*17: Consolidado del Presupuesto.*

*Fuente: Elaboración propia.*

## 1.2. Anexo 2: Política ambiental.

### POLÍTICA AMBIENTAL

Nosotros, consideramos que la preservación y cuidado del medio ambiente es uno de los pilares sobre los cuales nuestros procesos y lineamientos de sostenibilidad se mantienen, por lo tanto, nos comprometemos a:

1. Fomentar, promover y desarrollar conciencia ambientalista en nuestros colaboradores y visitantes mediante capacitaciones y sensibilizaciones continuas.
2. Velar por la protección del medio ambiente y la prevención de contaminación, reduciendo y/o eliminando los impactos ambientales en nuestros procesos.
3. Establecer un proceso de gestión de residuos sólidos orientados hacia la eliminación de estos y el reciclaje de estos.
4. Cumplir con la normativa legal y otros requisitos que sean aplicables a la gestión ambiental en nuestros procesos.
5. Documentar las medidas adoptadas y los resultados obtenidos en nuestro sistema de gestión ambiental, para su difusión y aplicación.
6. Mantener vigentes nuestra política ambiental, y asegurar su disponibilidad para intereses internos o externos.