

---

ESTUDIOS / RESEARCH STUDIES

---

## Método para recomendar factores de posicionamiento personalizados en el motor de búsqueda de Google

Richard Injante\*, David Mauricio \*\*

\* Universidad Nacional Mayor de San Marcos, Perú.

Universidad Nacional de San Martín - Tarapoto, Perú.

Correo-e: richard@unsm.edu.pe | ORCID iD: <https://orcid.org/0000-0002-2449-8937>

\*\* Universidad Nacional Mayor de San Marcos, Perú.

Correo-e: dmauricios@unmsm.edu.pe | ORCID iD: <https://orcid.org/0000-0001-9262-626X>

Recibido: 26-11-2018; 2ª versión: 29-04-2019; Aceptado: 15-05-2019.

**Cómo citar este artículo/Citation:** Injante, R.; Mauricio, D. (2020). Método para recomendar factores de posicionamiento personalizados en el motor de búsqueda de Google. *Revista Española de Documentación Científica*, 43 (1), e253. <https://doi.org/10.3989/redc.2020.1.1628>

**Resumen:** El considerable aumento de sitios web en Internet con temáticas de diversa índole ha hecho que los usuarios utilicen este medio para buscar y conseguir información. De todos los motores empleados para esta tarea, la mayoría de personas emplea Google como su motor de búsqueda preferido. Teniendo esto en consideración, se vuelve fundamental alcanzar las mejores posiciones en los resultados de búsqueda para poder promocionar un sitio web. Este trabajo ofrece un método basado en 6 fases para recomendar factores de posicionamiento personalizados a los propietarios de páginas web con el fin de que mejoren la clasificación de sus páginas web en el buscador de Google. El método se aplicó en una página web y se logró alcanzar mejoras considerables en su posicionamiento.

**Palabras clave:** factores de posicionamiento; motor de búsqueda; buscador de Google; optimización en los motores de búsqueda.

### Method to Recommend Personalized Ranking Factors in the Google Search Engine

**Abstract:** The considerable increase of websites on the Internet with different themes has prompted users to use this medium to search for and obtain information. Of all the search engines used for this task, Google is the one most people use as their preferred search engine. With this in mind, it becomes critical to reach the best positions in the search results in order to promote a website. This work offers a 6 phases based method to recommend personalized positioning factors to the owners of web pages, in order to improve the ranking of their web pages in the Google search engine. The method was applied in a test web page and a considerable improvement in its positioning was obtained.

**Keywords:** ranking factor; search engine; Google search; search engine optimization.

## 1. INTRODUCCIÓN

Los motores de búsqueda, tal como lo señalan Pérez-Montoro y Codina (2017), son los sistemas de información documental más importantes de nuestro tiempo. De hecho, cuesta concebir una mejor manera de explotar la inmensa riqueza de la Web que con un motor de búsqueda, cuyas funciones no han dejado de aumentar o sufrir cambios desde su aparición en los años noventa.

Como es sabido (Mangles, 2018), existen diferentes motores de búsqueda, entre ellos Google, Bing, Yahoo, Ask, entre otros, siendo el primero el más popular y usado, convirtiéndose en la herramienta de búsqueda más importante del mundo.

El principal problema para todos los propietarios de sitios web con respecto a estos motores de búsqueda es el bajo posicionamiento, el bajo tráfico y la falta de visibilidad en ellos, así que para mejorar su posicionamiento tienen que utilizar diferentes técnicas de posicionamiento; de este modo, se incurre muchas veces en malas prácticas o en el uso inadecuado de estas técnicas, lo que genera entonces penalizaciones del sitio web y perjudica gravemente su posicionamiento. Muchos estudios aportan diferentes factores de posicionamiento a utilizar; sin embargo, debido al cambio frecuente del algoritmo de ranking de Google, los propietarios de sitios web corren el riesgo de aplicar factores que podrían ya estar prohibidos o desfasados, por tanto, requieren de un método que les permita identificar los factores más convenientes para su sitio web, esto con el fin de que utilicen estrategias adecuadas y alcancen mejores posiciones en Google. Por tal motivo, surge la necesidad de desarrollar un método que permita a los propietarios utilizar los factores de posicionamiento adecuados y, sobre todo, personalizados para su sitio web.

Este trabajo propone un método para recomendar a un sitio web qué factores de posicionamiento debe aplicar para mejorar su clasificación en los resultados del motor de búsqueda de Google, con el apoyo de una palabra o frase clave en particular. Además, debido al constante cambio del algoritmo de Google, la propuesta permitirá mostrar los valores de los factores de posicionamiento vigentes y adecuados para el sitio web cada vez que se ejecute el método, el cual llega a ser más efectivo que otros trabajos sobre optimización en los motores de búsqueda (SEO), pues éstos solo tienen información captada en el momento puntual de su estudio.

En general, el método propuesto consiste en hacer un rastreo del contenido de los resultados de búsqueda de Google bajo determinadas palabras clave y al de los sitios web indexados. Esta

información servirá para obtener los valores de los factores de posicionamiento de los sitios web, para luego ser procesados mediante una técnica de aprendizaje de máquina y obtener las reglas que posicionen a un sitio web. Para finalizar, se comparan las reglas con los valores de los factores del sitio web que se desee posicionar para que de esta manera se recomiende la regla más similar y los factores de posicionamiento que se debe cambiar en la página web.

El presente trabajo está organizado en 6 secciones. En la sección 2, generamos el contexto acerca del motor de búsqueda de Google y SEO. Luego, en la sección 3, presentamos la literatura sobre los factores de posicionamiento. En la sección 4, describimos el método propuesto para posicionar un sitio web, el cual es validado, con un caso de estudio, en la sección 5. Finalmente, las conclusiones se consignan en la sección 6.

## 2. ANTECEDENTES

### 2.1. Motor de Búsqueda de Google

Google fue creado en 1997 como parte de un proyecto de investigación en la Universidad de Stanford y fue un prototipo de un motor de búsqueda a gran escala que hace un uso intensivo de la estructura hipertextual, diseñado para rastrear e indexar la Web de manera eficiente, produciendo resultados de búsqueda mucho más satisfactorios que los sistemas existentes (Brin y Page, 2012).

Google proporciona un servicio para buscar miles de millones de páginas web indexadas de forma gratuita. Su funcionamiento básico se puede describir de la siguiente forma: cuando el usuario ejecuta una consulta en el buscador de Google, éste presenta una lista ordenada de páginas web según su grado de relevancia; la visualización de los resultados para cada consulta de búsqueda enviada se compone de páginas web gratuitas y de pago (anuncios). Las páginas web ordenadas naturalmente, también conocidas como búsqueda orgánica, son páginas web determinadas por su complejo y secreto algoritmo matemático que toma en cuenta una serie de factores internos y externos de las páginas web.

Google sigue siendo el motor de búsqueda más popular en la Internet (Allen, 2017) y, debido a su popularidad, este trabajo estudia los factores internos que podría considerar el motor de búsqueda.

### 2.2. Optimización del motor de búsqueda (SEO)

Existen diferentes definiciones sobre SEO: según el Equipo de Calidad de la Búsqueda de Goo-

gle (2018), éste consiste en hacer pequeñas modificaciones en el sitio web (contenido y código). Estos cambios, a su vez, pueden mejorar el posicionamiento. Por su parte, Duklan y otros (2015) indican que es el proceso de mejorar la visibilidad de un sitio web en los resultados de búsqueda sobre una consulta de palabras clave; Anderson (2017) nos dice que es un proceso técnico, analítico y creativo para mejorar la visibilidad de un sitio web en los motores de búsqueda, mientras que su función principal es impulsar más visitas a un sitio. En síntesis, SEO es el uso de técnicas para posicionar un sitio web en los resultados del motor de búsqueda.

El arte de hacer SEO radica en la comprensión de cómo los usuarios de sitios web buscan temas de interés y en entender qué tipo de resultados debe mostrárseles. Se trata de reunir muchas cosas para buscar oportunidades. Un buen optimizador tiene una comprensión de cómo los motores de búsqueda generan sus resultados naturales para satisfacer las consultas de navegación, informativas y transaccionales de los usuarios.

Según Gudivada y otros (2015), existen dos tipos de técnicas: SEO Sombrero Blanco y Sombrero Negro. El primero se lleva a cabo a través de un enfoque formal para la optimización del sitio web, mientras que el segundo usa métodos no convencionales para lograr o engañar la mejora del efecto de la clasificación del sitio; por lo general son prácticas no éticas y pueden generar la penalización del sitio web. Rehman y Khan (2013) mencionan que las técnicas de SEO también implican dos procesos importantes: la optimización interna y externa. La optimización interna se aplica cuando el propietario de sitios web tiene control directo del contenido del sitio web, mientras que en la externa los factores que influyen en el posicionamiento son ajenos a éste.

En los medios en línea, el contenido del sitio web es muy importante y se debe asegurar que se cumplan los factores de clasificación adecuados. En este sentido, se debe asegurar que el contenido esté diseñado de una forma que mejore el posicionamiento del sitio web en el motor de búsqueda de Google.

### 3. TRABAJOS PREVIOS

Diferentes estudios se han abocado a identificar cuáles son los factores de posicionamiento que más influyen para que un sitio web se posicione mejor en los resultados del motor de búsqueda. Egri y Bayrak (2014) utilizaron herramientas como PageSpeedInsights y Pingdom para medir el tiempo de carga, velocidad, tasa de rebote, vistas de página y diseño de la página para mantener al usuario en

el sitio. Además, efectuaron un análisis con Google Analytics e identificaron que el factor de posicionamiento más importante es el tiempo de duración del usuario en el sitio y éste está influenciado directamente con la velocidad de carga de la página web. Por otro lado, Lin y Chi (2014) propusieron un método que utiliza la frecuencia de término-frecuencia inversa de documento (TF-IDF) y K-means para identificar la combinación de palabras clave que beneficiarán la optimización del motor de búsqueda; como resultado, el sitio web de su estudio recibió un importante impacto reflejado en diversos indicadores, entre los que destaca su mejora en el ranking Alexa, y en factores, como el número de backlinks. Hussien (2014) llevó a cabo una investigación empírica mediante la ponderación de 20 factores basados en una ecuación propuesta, recomendando el uso de guiones en el localizador de recursos uniforme (URL) del sitio, minimizar los errores ortográficos, utilizar encabezados H1, una meta descripción adecuada y uso de sustantivos en la página. Moráguez y Cancio (2014), a diferencia de los otros autores, emplearon una encuesta para identificar los factores que influyen en un bajo posicionamiento web en algunos sitios de especialidades médicas, los cuales fueron el uso inadecuado de las palabras clave en las etiquetas Meta, enlaces internos que no facilitan la navegación del usuario, poca actualización de documentos y contenidos del sitio muy diseminados. Duklan y otros (2015) identificaron factores que tienen máximo impacto en el ranking. Para este propósito, usaron el análisis de clúster de k-means para agrupar los factores externos: así, se obtuvo que el intercambio de enlaces, Metatags, publicación en directorios, seguimiento de sitios web, cumplimiento de normas W3C y la presentación de marcadores sociales, mejora el posicionamiento de un sitio web. Krrabaj y otros (2017) estudiaron diferentes factores gracias al uso de herramientas proporcionadas por Google para evaluar su sitio. En este caso, el análisis fue dirigido a un sitio web educativo, donde identificaron que los factores de mayor impacto son la presencia de la palabra clave en el título del artículo y en la URL del sitio web. Además, las palabras clave deben aparecer al menos tres veces en el contenido principal. Finalmente, Eswarawaka y otros (2017) revisaron diferentes factores de posicionamiento y proponen un método para hacer SEO, apoyado en la metodología Sigsixma, en la que experimentan con las palabras clave dentro de la etiqueta título y el documento, donde concluyen que el contenido debe estar directamente relacionado con lo que el usuario está buscando, es decir, que la palabra clave aparezca con mayor frecuencia en el documento.

#### 4. MÉTODO PROPUESTO

Según la revisión de la literatura, muchos estudios revelan los factores que pueden influir en el posicionamiento de sitios web, pero no se ha encontrado un estudio que brinde un método que ayude a identificarlos de forma personalizada y que pueda ejecutarse en diferentes períodos con el fin de encontrar nuevos cambios que ayuden a las páginas web a posicionarse. El objetivo de este estudio es aportar un método para la selección de factores de posicionamiento de manera personalizada, el cual evite el abuso de SEO y de donde se obtengan los cambios necesarios que debe realizar el editor de contenidos web basado en la estructura de su sitio web.

El método propuesto es empírico y tiene como base el proceso de extracción de conocimiento llamado Knowledge Discovery in Databases (KDD) y está dividido en 6 fases. El método inicia con la selección de factores (1) y palabras clave de una temática determinada (2), para luego extraer, mediante un rastreador, las URL de los resultados de Google y también los valores de los factores de posicionamiento de los sitios web indexados (3); posteriormente, normalizamos los datos obtenidos (4) y aplicamos una técnica de Machine Learning (5) con el fin de obtener reglas de decisión que luego serán comparadas con los valores de los factores del sitio web a posicionar, para después obtener las recomendaciones que permitirán mejorar el posicionamiento en el motor de búsqueda Google (6).

A continuación se detalla cada fase y los procedimientos a seguir.

##### 4.1. Fase 1. Selección de los factores de posicionamiento

Los factores de posicionamiento son aquellos que utilizan los motores de búsqueda para evaluar el or-

den de relevancia de una página web cuando alguien la busca basándose en una palabra clave, por lo cual es de presumir que los factores de clasificación tienen un peso diferente asignado y que éstos son utilizados por el algoritmo de Google (Aregay, 2014). En esta fase se deben seleccionar los factores de posicionamiento a considerar para la ejecución del método, con el fin de generar una lista que se usará en la fase 3. La tabla I muestra las entradas, herramientas y técnicas, así como las salidas de esta fase.

##### Procedimiento:

Los factores internos más relevantes deben ser tomados mediante una exhaustiva revisión literaria de libros y artículos. También se pueden considerar aquellos recomendados por expertos y la propia experiencia. Es importante que el peso o valor de los factores pueda obtenerse mediante el rastreo de documentos HTML.

La lista debe contener las cabeceras que aparecen en la tabla II.

##### 4.2. Fase 2. Selección de las palabras clave

Miles de millones de personas en todo el mundo realizan búsquedas todos los días mediante el envío de palabras clave. Una palabra clave es la consulta textual enviada por los usuarios a los motores de búsqueda con la finalidad de obtener páginas web relevantes a su consulta (Aregay, 2014).

En esta fase se seleccionan las palabras clave más relevantes; para ello se debe identificar la temática o nicho relacionado al sitio web que se desea posicionar y elegir una palabra clave objetivo para ejecutarla en la herramienta de la fase 3. La tabla III muestra las entradas, herramientas y técnicas, así como las salidas de esta fase.

**Tabla I.** Fase 1: Selección de los factores de posicionamiento

Entrada	Herramientas y Técnicas	Salida
- Artículos - Libros	- Revisión Literaria - Juicio de expertos	- Lista Factores

**Tabla II.** Ejemplo de lista de factores

ID	Factor	Descripción	Tipo de Dato	Valores

**Tabla III.** Fase 2: Selección de las palabras clave

Entrada	Herramientas y Técnicas	Salida
- Palabra Clave Objetivo	- Keyword Planner de Adwords	- Lista de Palabras Clave

**Procedimiento:**

Se debe elegir una palabra clave objetivo en un único idioma y consultarla en la herramienta de palabras clave de Google (Keyword Planner de Adwords). Esta herramienta mostrará una lista de palabras clave relacionadas y ordenadas por volumen de búsquedas. Es recomendable que la lista de palabras clave tenga la mayor cantidad posible de aquellas que se le relacionen.

**4.3. Fase 3. Rastreo de contenidos**

Los rastreadores web o las arañas se utilizan principalmente para recopilar diferentes tipos de información de las páginas web para su posterior procesamiento. También se pueden usar para automatizar las tareas de mantenimiento en un sitio web, como verificar enlaces, validar el código HTML o extraer textos específicos (Lavania y otros, 2013).

En esta fase se deben extraer los valores de los factores de posicionamiento de las páginas web. Para ello, se requiere de dos herramientas que serán ejecutadas en dos etapas: la primera se encarga de rastrear el contenido de resultados del motor de búsqueda por cada palabra clave, para después obtener las URL de las páginas indexadas; la segunda rastrea el contenido de las páginas web obtenidos en la primera etapa, esto con el fin de obtener los valores de cada factor interno. La tabla IV muestra las entradas, herramientas y técnicas, así como las salidas de esta fase.

**Procedimiento:**

En esta fase se debe ejecutar la herramienta para rastrear el contenido de los resultados del

motor de búsqueda. Esta herramienta debe tomar cada palabra clave de la lista inicial, realizar la consulta en el motor de búsqueda, tomar las URL de las tres primeras posiciones de la primera página de resultados y tres de la quinta o mayor página de resultados. Para Ochoa (2012), los usuarios de Google solo visitan los primeros resultados antes de cambiar sus consultas, solo un 16 % de ellos pasa a la segunda página de resultados y menos del 1 % llegan hasta la cuarta página, por lo tanto, la justificación para tomar estas URL se debe a que los sitios web con mejor visibilidad y posicionamiento se encuentran en las 3 primeras posiciones; y las que se encuentran en la quinta página en adelante son las que prácticamente no tienen visibilidad y se puede considerar como páginas web no posicionadas. A mayor distancia entre las páginas posicionadas y no posicionadas, mayor será la diferencia de los valores de los factores. Las URL a tomar únicamente deben ser documentos HTML. También debe considerar que para ejecutar el rastreador la sesión de usuario de la cuenta de Google debe estar cerrada, se debe ejecutar en un solo dominio de Google (p. ej., www.google.com.pe) y tomar únicamente las URL de los resultados orgánicos.

En la segunda etapa del rastreo del contenido de las páginas web indexadas, se debe tomar las URL obtenidas en la primera etapa y extraer los valores de cada factor de las páginas web, según la lista de factores obtenida en la fase 1. Las cabeceras del conjunto de datos de factores deben ser como aparecen en la tabla V.

Se recomienda ejecutar esta fase durante un período superior a siete días con el fin de manejar una gran cantidad de datos.

**Tabla IV.** Fase 3: Rastreo de contenidos

Entrada	Herramientas y Técnicas	Salida
<ul style="list-style-type: none"> <li>- Lista de palabras clave</li> <li>- Lista de factores</li> </ul>	<ul style="list-style-type: none"> <li>- Rastrear el contenido de los resultados del motor de búsqueda</li> <li>- Rastrear el contenido de las páginas web indexadas</li> </ul>	<ul style="list-style-type: none"> <li>- Conjunto de datos de factores</li> </ul>

**Tabla V.** Ejemplo del conjunto de datos de factores

Posición	Factor1	Factor2	Factor3	...	FactorN

#### 4.4. Fase 4. Preparación de datos

En esa fase se prepara el conjunto de datos de los factores obtenidos en la fase 3, esto con el fin de tener datos completos y sin duplicidad. También se deben numerar normalizar o discretizar buscando adaptar los datos a las necesidades de los algoritmos. La tabla VI muestra las entradas, herramientas y técnicas, así como las salidas de esta fase.

##### **Procedimiento:**

- Eliminar registros duplicados e incompletos.
- Transformar todos los datos numéricos a escalas de ordinales como "Muy Alto, Alto, Normal, Bajo, Muy Bajo".
- Transformar los valores booleanos a verdadero (V) y Falso (F).
- Transformar los valores del campo "Posición" en "Posicionado" para las 3 primeras posiciones y "NoPosicionado" para las siguientes 3 posiciones según el conjunto de datos de factores.

#### 4.5. Fase 5. Aplicación de técnica de Machine Learning

Esta es la etapa en la que se genera el modelo de conocimiento aplicando una técnica de Machine Learning con base en el conjunto de datos de factores normalizados obtenidos en la fase 4, con el fin de extraer reglas de decisión que nos permi-

tan proponer factores importantes para posicionar un sitio web. La tabla VII muestra las entradas, herramientas y técnicas, así como las salidas de esta fase.

##### **Procedimiento:**

- Seleccionar una herramienta para Machine Learning.
- Aplicar una técnica de Machine Learning que genere reglas de decisión.
- Las reglas generadas por la herramienta deben ser trasladadas a un conjunto de datos en forma de vector.
- Los factores de cada regla que no tengan valor se pueden reemplazar con una "X", este representa a cualquier valor, es decir, que su valor no influye en la decisión final, pero es recomendable aplicarlas.
- Considerar solo las reglas con la clasificación "Posicionado".

El conjunto de reglas generado se muestra en la tabla VIII.

#### 4.6. Fase 6: Recomendación de factores

El conjunto de reglas generado en la fase 6 contiene todas las reglas útiles para posicionar un sitio web. Según el sitio web a posicionar, se

**Tabla VI.** Fase 4: Preparación de datos

Entrada	Herramientas y Técnicas	Salida
- Conjunto de datos de factores	- Limpieza de datos - Transformación de datos	- Conjunto de datos de factores normalizados

**Tabla VII.** Fase 5: Aplicación de técnica de Machine Learning

Entrada	Herramientas y Técnicas	Salida
- Conjunto de datos de factores normalizados	- Herramienta para Machine Learning - Técnica de Machine Learning	- Conjunto de reglas

**Tabla VIII.** Ejemplo del conjunto de reglas

Posición	Factor1	Factor2	Factor3	...	Factor N
Posicionado	V	F	X	...	Alto
Posicionado	X	F	V	...	Bajo

debe realizar un proceso de comparación hasta obtener la regla ideal y personalizada para el sitio web. La tabla IX muestra las entradas, herramientas y técnicas, así como las salidas de esta fase.

#### **Procedimiento:**

- Mediante un rastreador de documentos, extraer los valores de los factores del sitio web que se desea posicionar.
- Comparar cada registro del conjunto de reglas con los valores de los factores del sitio web hasta encontrar el más similar, empezando con los factores de la página web que tengan el valor "Verdadero" y el menor número de registros filtrados, hasta llegar al mínimo de reglas.
- En caso de que hubiese más reglas, comparar con el tamaño del documento, número de enlaces externos y número de enlaces internos (el orden es con base en el menor número de registros filtrados).
- Seleccionar la regla que tenga menos factores a cambiar.
- Finalmente, realizar los cambios respectivos, según los factores recomendados, en la página web.

## **5. VALIDACIÓN**

### **5.1 Caso de estudio**

Para validar el método propuesto, se seleccionó el sitio web *bambutravels.com*, página web que pertenece a una agencia de viajes peruana dedicada a la venta de vuelos y paquetes turísticos en la ciudad de Tarapoto, Perú. Sus principales clientes son turistas nacionales.

La palabra clave seleccionada, con la que se busca posicionar la página web, es "paquetes turísticos a Tarapoto". Esta palabra clave cuenta con 112,000 resultados en *google.com.pe* y, según la herramienta de Keyword Planner de Adwords, mantiene

un promedio de 120 búsquedas mensuales con un nivel de competencia alto. No se considera el uso de la tilde puesto que los usuarios no suelen utilizar los acentos en sus consultas.

Antes de aplicar el método, la página web se encontraba en la posición 46 en los resultados de búsqueda. Según Ochoa (2012), los usuarios no llegan hasta la cuarta página de los resultados de búsqueda, lo cual implica que la página web no tiene visibilidad en el motor de búsqueda de Google.

### **5.2 Aplicación del método**

A continuación se detalla la aplicación del método propuesto y resultados obtenidos.

#### **Fase 1. Selección de los factores de posicionamiento**

Para este caso de estudio se han considerado 18 factores de posicionamiento internos que fueron tomados por medio de una revisión literaria y de la experiencia de los autores. Los factores seleccionados se muestran en la tabla X.

#### **Fase 2. Selección de las palabras clave**

Debido a que el sitio web a posicionar está dentro del sector turismo, la palabra clave objetivo elegida fue "paquetes turísticos". A partir de esto y con el uso de la herramienta Keyword Planner de Adwords, se tomaron 599 palabras clave relacionadas con la palabra clave objetivo. Con ello se generó la lista de 600 palabras clave.

#### **Fase 3. Rastreo de contenidos**

Se desarrolló una herramienta para rastrear los resultados del motor de búsqueda de Google por cada palabra clave consultada (lista de palabras clave) y, posteriormente, se rastrearon los documentos HTML con el fin de extraer los valores de cada factor interno, según la lista de factores definida en la primera fase. La figura 1 muestra el proceso de rastreo que realizó la herramienta.

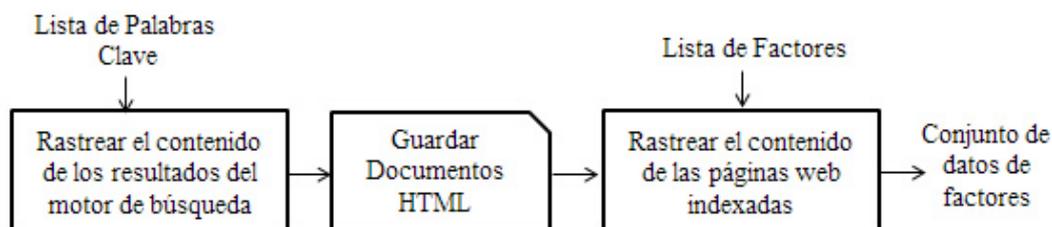
**Tabla IX.** Fase 6: Recomendación de factores

<b>Entrada</b>	<b>Herramientas y Técnicas</b>	<b>Salida</b>
<ul style="list-style-type: none"> <li>- Conjunto de Reglas</li> <li>- Página Web a posicionar</li> </ul>	<ul style="list-style-type: none"> <li>- Extraer valores de los factores de la página web a posicionar</li> <li>- Comparar las reglas con los valores del sitio web</li> </ul>	<ul style="list-style-type: none"> <li>- Lista de factores recomendados</li> </ul>

**Tabla X.** Lista de factores seleccionados

ID	Factor	Descripción	Tipo de Dato	Valores
1	Palabra clave en la URL	La palabra clave se encuentra en la URL de la página	Booleano	V o F
2	Palabra clave en el Meta Keyword	La palabra clave se encuentra dentro de la meta etiqueta "Keyword"	Booleano	V o F
3	Palabra clave en la etiqueta título	La palabra clave se encuentra dentro de la meta etiqueta "title"	Booleano	V o F
4	Palabra clave en la etiqueta H1	La palabra clave se encuentra dentro de la meta etiqueta "h1"	Booleano	V o F
5	Palabra clave en la etiqueta H2	La palabra clave se encuentra dentro de la meta etiqueta "h2"	Booleano	V o F
6	Palabra clave en la etiqueta H3	La palabra clave se encuentra dentro de la meta etiqueta "h3"	Booleano	V o F
7	Palabra clave en el Meta Descripción	La palabra clave se encuentra dentro de la meta etiqueta "description"	Booleano	V o F
8	Meta etiqueta Facebook	Hace uso de las meta etiquetas para Facebook	Booleano	V o F
9	Meta etiqueta UTF8	Hace uso de la meta etiqueta utf8 para la codificación de pagina	Booleano	V o F
10	Meta etiqueta Twitter	Hace uso de las metas etiquetas para Twitter	Booleano	V o F
11	Uso de HTML5	El sitio web está en HTML	Booleano	V o F
12	Uso del protocolo https	El sitio web usa un certificado de seguridad SSL	Booleano	V o F
13	Idioma de la página	Se especifica el idioma de la página	Booleano	V o F
14	Dominio sin WWW	Si el dominio está indexado sin "www"	Booleano	V o F
16	Página responsiva	El sitio web es responsivo	Booleano	V o F
16	Número de enlaces externos	Número total de enlaces externos	Numérico	>=0
17	Número de enlaces internos	Número total de enlaces internos	Numérico	>=0
18	Tamaño del documento	Tamaño del documento HTML de la página web	Numérico	>0

V= verdadero: F= falso

**Figura 1.** Proceso de rastreo

Durante determinado período se procedió a extraer las URL de los resultados del motor de búsqueda de Google por palabra clave enviada, para ello, se han considerado los siguientes filtros:

- Solo búsquedas en idioma español
- Búsquedas sin la sesión activa de la Cuenta de Google
- Búsquedas al dominio Google.com.pe
- Solo se consideran los resultados orgánicos
- Se tomaron las URL de las posiciones 1,2,3 y 51,52,53

La solicitud ejecutada por la herramienta fue a la siguiente URL: [https://www.google.com.pe/search?q={Palabra Clave}&lr=lang\\_es](https://www.google.com.pe/search?q={Palabra Clave}&lr=lang_es)

La herramienta ejecutó la lista de palabras clave en el motor de búsqueda de Google durante 7 días. En total se extrajeron 3.594 URL por día, lo cual hizo un total de 25.158 URL almacenados en la base de datos. Posteriormente, se almacenaron en un directorio todos los documentos por cada URL válida (no se incluyeron documentos diferentes a HTML), por lo que se obtuvo un Dataset de factores con 19.385 registros.

#### Fase 4. Preparación de datos

En esa fase se normalizaron los valores de los factores "Número de enlaces externos", "Número enlaces internos" y "Tamaño del documento" a escala ordinal basados en la distribución de Gauss (tabla XI).

También se reemplazaron los valores "1, 2, 3" del campo posición por "Posicionado" y los valores "51, 52, 53" por "NoPosicionado". Al hacer este cambio se generaron registros duplicados, los cuales fueron eliminados haciendo un conjunto de datos de 1576 registros.

#### Fase 5. Aplicación de técnica de Machine Learning

Para la validación, se utilizó la técnica de Random Forest y la herramienta WEKA 3.8.2 con las siguientes configuraciones:

- Dimensión: 18 Atributos
- Conjunto de entrenamiento: 1572 instancias
- Número de árboles: 100
- Regla de clasificación: Posicionado, NoPosicionado

Los resultados de WEKA mostraron una precisión promedio de 84,8 %. La tabla XII muestra el resumen de los resultados:

**Tabla XII.** Resumen de resultados con WEKA.

	Posicionado	NoPosicionado
Precisión	82.6 %	85.5 %
Reglas	30911	33358

- Instancias correctamente clasificadas: 1.337
- Instancias incorrectamente clasificadas: 235
- Media de precisión: 84,8 %
- Número de reglas generadas: 64.269

Matriz Confusión

```
a      b <-- classified as
219 189 | a = Posicionado
46  1118 | b = NoPosicionado
```

Finalmente, se eliminaron las reglas duplicadas y se generó una lista de reglas únicamente con la regla de clasificación "Posicionado", lo cual hizo un total de 30911 reglas.

#### Fase 6. Recomendación de factores

Como se mencionó anteriormente, la palabra clave con la que desea posicionar la página web bambustravels.com es "paquetes turísticos a Tarapoto"; cabe mencionar que los resultados del motor de búsqueda de Google varían con el uso de la tilde, pero en este estudio no se ha considerado el uso de la tilde en la palabra clave debido a que los usuarios no suelen utilizarlas con frecuencia en sus consultas.

**Tabla XI.** Factores normalizados

Enlaces Externos	Escala	Enlaces Internos	Escala	Tamaño del documento	Escala
[ >125 ]	Muy Alto	[ >200 ]	Muy Alto	[ >200 ]	Muy Alto
[75 - 125]	Alto	[100 - 200]	Alto	[100 - 200]	Alto
[50 - 75]	Normal	[50 - 100]	Normal	[50 - 100]	Normal
[25 - 50]	Bajo	[25 - 50]	Bajo	[25 - 50]	Bajo
[0 - 25]	Muy Bajo	[0 - 25]	Muy Bajo	[0 - 25]	Muy Bajo

Se rastreó la página principal de bambutravels.com para obtener los valores de sus factores internos. Posteriormente, se compararon estos valores de la página web con la regla más similar que nos dio el método de la siguiente manera:

- El primer filtro: "Página Responsiva" con 7609 reglas
- El segundo filtro: "Uso de HTML5" con 2313 reglas
- El tercer filtro: "Tamaño del documento" con 464 reglas
- El cuarto filtro: "Número de enlaces externos" con 59 reglas
- El quinto filtro: "Número de enlaces internos" con 18 reglas

Debido a que el quinto filtro generó 18 reglas, se seleccionó la regla que menos cambios recomienda. Así se obtuvo la tabla XIII.

Según la tabla XIII, el método reveló obligatoriamente el "Uso del protocolo https" y no usar "Palabra clave en la etiqueta H1", además, no cambiar los factores "Uso de HTML5", "Página Responsiva",

"Número de enlaces externos", "Número de enlaces internos", "Tamaño del documento". La regla quedó de la siguiente manera:

Página Responsiva = V

| Uso de HTML5 = V

| | Tamaño del documento = Normal

| | | Número de enlaces externos = Normal

| | | | Número de enlaces internos = Muy Bajo

| | | | | Número de enlaces internos = Muy Bajo

| | | | | Uso del protocolo https = V

| | | | | | Palabra clave en la etiqueta H1 = F:  
Posicionado

Algunos factores con valor X fueron cambiados de manera opcional, no influyen en la regla, pero se recomienda utilizarlas (tabla XIV).

Posteriormente, después de los cambios realizados, mediante la herramienta Google Search Console, se solicitó la re-indexación de la página web.

**Tabla XIII.** Comparación de los factores de la página web y regla válida

ID	Factor	Valor	Regla	¿Requiere cambio?
1	Palabra clave en la URL	F	X	Opcional
2	Palabra clave en la Meta etiqueta Keyword	F	X	Opcional
3	Palabra clave en la etiqueta Title	F	X	Opcional
4	Palabra clave en la etiqueta H1	F	F	No
5	Palabra clave en la etiqueta H2	F	X	Opcional
6	Palabra clave en la etiqueta H3	F	X	Opcional
7	Palabra clave en la Meta etiqueta Description	F	X	Opcional
8	Meta etiqueta Facebook	F	X	Opcional
9	Meta etiqueta UTF8	F	X	Opcional
10	Meta etiqueta Twitter	F	X	Opcional
11	Uso de HTML5	V	V	No
12	Uso del protocolo https	F	V	Si
13	Idioma de la página	F	X	Opcional
14	Dominio sin WWW	F	X	Opcional
15	Página responsiva	V	V	No
16	Número de enlaces externos	Muy Bajo (3)	Bajo	No
17	Número de enlaces internos	Muy Bajo (21)	Muy Bajo	Opcional
18	Tamaño del documento	Normal (58.7kb)	Normal	No

**Tabla XIV.** Cambios efectuados en la página web

Factor	Cambios realizados
Palabra clave en la etiqueta Title (sin tilde)	<title>Paquetes Turisticos a Tarapoto 2019 - Agencia de Viajes en Tarapoto</title>
Palabra clave en el Meta Description (sin tilde)	<meta name='description' content='Ofrecemos Paquetes Turisticos a Tarapoto, tenemos paquetes todo incluido, tours, hoteles, vuelos. Disfruta semana Santa, fiestas patrias, semana larga' >"
Palabra clave en la Meta etiqueta Keyword (sin tilde)	<meta name='keywords' content='paquetes turisticos a tarapoto, paquetes turisticos, paquetes turisticos tarapoto, paquetes turisticos baratos, tours en tarapoto, agencia de viajes, operador turistico, viajes Tarapoto, paquete a Tarapoto' >
Uso del protocolo https	Instalación de un certificado digital https://bambutravels.com

### 5.3 Resultados

Para el seguimiento del posicionamiento de la página web <https://bambutravels.com/>, se utilizó la herramienta de rendimiento de Google Search Console. Esta herramienta provee a los creadores de sitios web la oportunidad de comprobar el estado de indexación en el buscador y optimizar su visibilidad.

Después de los cambios realizados con base en las recomendaciones del método, la figura 2 muestra la evolución del posicionamiento con la palabra clave "paquetes turisticos a tarapoto" durante 3 meses, de este modo, logra alcanzar la posición media máxima de 3.5 en los resultados del motor de búsqueda Google.com.pe

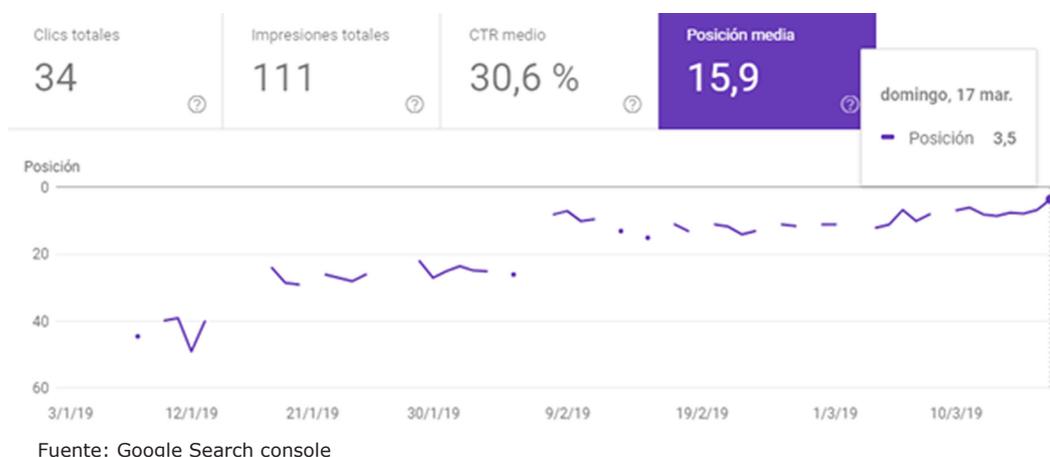
En general, sin considerar palabras clave específicas o zona geográfica, la figura 3 muestra que la posición media de la página web <https://bambutravels.com/> fue de 16.8, siendo 98.4 la posición mínima y 5.6 la posición máxima alcanzada, lo cual indica que también existieron mejoras en el posicionamiento con otras palabras clave, las

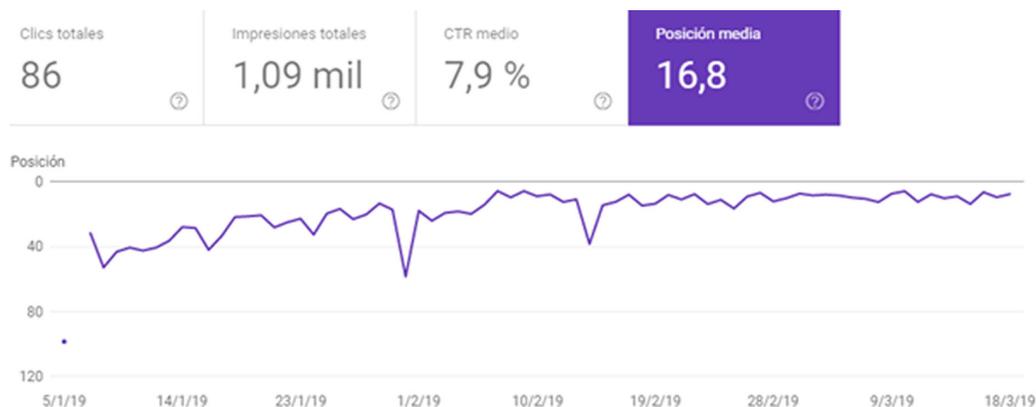
palabras clave más comunes fueron "tours tarapoto 2019", "agencia de viajes bambu", "bambu tours", "paquetes turisticos tarapoto 2019", "tarapoto paquetes turísticos" (con tilde) y "tours en tarapoto precios".

### 6. CONCLUSIONES

El trabajo permitió elaborar un método recomendado para lograr factores de posicionamiento personalizados. Mediante la técnica de minería de datos se ha logrado analizar estos factores de posicionamiento de 19.385 páginas web con la finalidad de obtener reglas que sugieran a una página web cuáles son los factores que requieren cambio para mejorar su posicionamiento. El constante cambio del algoritmo de Google no afecta al método, ya que puede ser ejecutado en tiempos diferentes y desde éste se puede obtener reglas actualizadas.

La página web del caso de estudio inicialmente no se encontraba visible en los resultados de búsqueda, pero, luego de los cambios recomendados

**Figura 2.** Evolución del posicionamiento con la palabra clave "paquetes turisticos a Tarapoto" en *Google.com.pe*

**Figura 3.** Evolución general del posicionamiento.

Fuente: Google Search console

por el método, ésta logró mejoras significativas en su posicionamiento; así, alcanzó la primera página de los resultados del motor de búsqueda de Google con una posición media máxima de 3,5. Si bien es cierto la página web no alcanzó las tres primeras posiciones, los resultados conseguidos son optimistas, ya que se incrementó la visibilidad en el motor de búsqueda. Además de las palabras clave ya expuestas, también se mejoró su posicionamiento con otras palabras clave.

Como se apreció en los trabajos previos, se utilizan diferentes métodos y herramientas para identificar los factores que posicionen a un sitio web; sin embargo, ninguno propone un método

sistemático que permita identificar, de una lista de factores, cuáles de estos son necesarios para implementarse en un sitio web de forma personalizada, es decir, recomendar a los propietarios de sitios web los factores idóneos que le sirvan para posicionarse. Esto demuestra que no es necesario aplicar todos los factores que las literaturas recomiendan, sino aplicar los necesarios para hacer SEO, lo cual guarda relación con la definición sobre SEO por parte del Equipo de Calidad de la Búsqueda de Google (2018), donde mencionan que SEO es hacer *pequeñas modificaciones* en el sitio web (contenido y código) y que estos cambios pueden mejorar el posicionamiento.

## 7. REFERENCIAS

- Allen, R. (2017). Search Engine Statistics 2017. *Smart Insights*. Disponible en: <http://www.smartinsights.com/search-engine-marketing/search-engine-statistics/> [Fecha de consulta: 12/06/17].
- Anderson, S. (2017). SEO Tutorial For Beginners in 2017. Disponible en: <https://www.hobo-web.co.uk/seo-tutorial/#what-is-seo-> [Fecha de consulta: 12/10/17].
- Aregay, T. (2014). *Ranking Factors for Web Search: Case Study in the Netherlands* (Tesis de Máster). Enschede, Netherlands: University of Twente.
- Brin, S.; Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, 56 (18), 3825–3833. <https://doi.org/10.1016/j.comnet.2012.10.007>
- Duklan, N.; Mourya, D.; Bahuguna, H. (2015). Classification of search engine optimization techniques: A data mining approach. *4<sup>th</sup> International Conference on System Modeling & Advancement in Research Trends*, pp. 275–280. Moradabad, India: Teerthanker Mahaveer University.
- Egri, G.; Bayrak, C. (2014). The Role of Search Engine Optimization on Keeping the User on the Site. *Procedia Computer Science*, 36, 335–342. <https://doi.org/10.1016/j.procs.2014.09.102>
- Equipo de Calidad de la Búsqueda de Google. (2018). Guía de optimización en buscadores (SEO) para principiantes - Ayuda de Search Console. Disponible en: <https://support.google.com/webmasters/answer/7451184?hl=es> [Fecha de consulta: 08/06/18]
- Eswarawaka, R.; Kudikala, S. K.; Kuchi, S. C.; Verma K, V. (2017). The analysis on search engine optimization supported by six sigma methodology. *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, pp. 653–658. Bangalore, India: IEEE. <https://doi.org/10.1109/ICIMIA.2017.7975544>
- Gudivada, V.; Rao, D.; Paris, J. (2015). Understanding Search Engine Optimization. *IEEE Computer*, 48(10), 43–52. <https://doi.org/10.1109/MC.2015.297>

- Hussien, A. S. (2014). Factors Affect Search Engine Optimization. *IJCSNS International Journal of Computer Science and Network Security*, 14 (9), 28-33.
- Krrabaj, S.; Baxhaku, F.; Sadrijaj, D. (2017). Investigating search engine optimization techniques for effective ranking: A case study of an educational site. *6th Mediterranean Conference on Embedded Computing (MECO)*, pp. 1-4. Bar, Montenegro: IEEE. <https://doi.org/10.1109/MECO.2017.7977137>
- Lavania, K. K., Jain, S., Gupta, M. K., Sharma, N. (2013), Google: A Case Study (Web Searching and Crawling), *International Journal of Computer Theory and Engineering*, 5 (2), 337-340. <https://doi.org/10.7763/IJCTE.2013.V5.705>
- Lin, T. F.; Chi, Y. P. (2014). Application of Webpage Optimization for Clustering System on Search Engine V Google Study. *2014 International Symposium on Computer, Consumer and Control*, pp. 698-701. Taichung, Taiwan: IEEE. <https://doi.org/10.1109/IS3C.2014.186>
- Mangles, C. (2018). Search Engine Statistics 2018. *Smart Insights*. Disponible en <https://smartinsights.com/search-engine-marketing/search-engine-statistics/> [Fecha de consulta: 08/03/18].
- Morález, M.; Cancio, L. (2014). Propuesta de factores a considerar en el posicionamiento de los sitios web de salud. *Revista Internacional de Gestión del Conocimiento y la Tecnología*, 2 (1), 10-30.
- Ochoa, E. (2012). *An Analysis of the Application of Selected Search Engine Optimization (SEO) Techniques and Their Effectiveness on Google's Search Ranking Algorithm* (Tesis de maestría inédita). California State University, Northridge, EE. UU.
- Pérez-Montoro, M.; Codina, L. (2017). *Navigation design and seo for content intensive websites: A Guide for an Efficient Digital Communication*. Oxford: Elsevier. <https://doi.org/10.1016/C2015-0-00739-0>
- Rehman, K.; Khan, M. A. (2013). The Foremost Guidelines for Achieving Higher Ranking in Search Results through Search Engine Optimization. *International Journal of Advanced Science and Technology*, 52, 101-110.