

Mejorando la búsqueda de información de contacto en Facebook analizando emociones

Improving contact information search on Facebook analyzing emotions

Daniel Alejandro Illanes Peredo, Wendoline Arteaga Sabja & Juan Pablo Sandoval Alcocer

Departamento de Ciencias Exactas e Ingeniería, Universidad Católica Boliviana “San Pablo”, Regional Cochabamba, Bolivia

dany.gaara.ds@gmail.com

Resumen: Existen grupos de *Facebook* destinados a la solicitud de información de contacto, en los cuales un usuario consulta por algún producto o servicio y otros usuarios aportan con su conocimiento sobre estos en los comentarios. Si bien los grupos actuales contienen mucha información, la gran cantidad de publicaciones y la redundancia en las mismas hace que el usuario invierta mucho tiempo en encontrar la información que necesita. Para ayudar a encontrar datos en los grupos, *Facebook* provee filtros además de un buscador, los cuales permiten obtener cierto tipo de información, sin embargo, estas opciones son limitadas e insuficientes para quienes quieren encontrar información específica sin invertir mucho tiempo.

En este artículo se propone el uso de análisis de emociones y aprendizaje automático para mejorar la búsqueda de información de contacto de las publicaciones en Facebook. Se desarrolló un prototipo clasificador, en base a redes bayesianas ingenuas, que permite buscar y categorizar la información de tal forma que los usuarios encuentren lo que necesitan rápidamente.

Se realizaron pruebas con 15 usuarios, a quienes se les pidió buscar información de contacto usando el prototipo desarrollado para luego contrastarlo con el buscador que provee Facebook, los resultados mostraron que la mayoría de los usuarios que utilizaron el buscador de Facebook encuentran información pertinente en la 6ta y 7ma posición de la lista de resultados, mientras que utilizando el prototipo desarrollado el usuario encuentra la información deseada en la 1era o 2da posición de la lista.

Palabras clave: red social, aprendizaje automático, red bayesiana ingenua, clasificador, análisis sentimental.

Abstract: There are different Facebook groups to help users find contact information. In these groups users may ask recommendations about a product or service to other users in the same group, where users share their experience through comments. Although these groups contains thousands of contact information, search recommendations in these groups is a tedious and time-consuming activity, mainly because the large amount of information contained in this group and the

redundant post and comments. To help users find post/comments of interest, Facebook provides a search engine that contains different kind of filters. However, these filter options are limited and sometimes insufficient for those who need to find specific information.

This article proposes the use of emotion analysis and machine learning to improve the search for contact information for Facebook publications. This paper also reports an experiment conducted with 15 users, where they asked to search for contact information using the prototype developed and the search engine that provides Facebook. The result shows that most users who used the Facebook search engine find relevant information in the 6th and 7th position of the list of results, while using the prototype developed the user finds the desired information in the 1st or 2nd position of the list.

Key words: social network, machine learning, naive Bayesian network, classifier, sentimental analysis.

1 Introducción

Hoy en día, la forma más usada para compartir información son las redes sociales. La red social *Facebook* alberga un gran número de grupos con enormes cantidades de información en publicaciones y comentarios, el grupo de prueba seleccionado “Alguien sabe? Cbba” es un grupo destinado a la obtención de información de contacto sobre algún producto o servicio a través de publicaciones realizadas. Este grupo permite realizar consultas nuevas y también permite buscar productos o servicios en publicaciones pasadas.

2 Información de contacto en Facebook

Existen varios grupos en las redes sociales donde la gente solicita información de contacto de diferentes profesionales o servicios. El grupo “Alguien sabe? Cbba” recibe en promedio más de 150 consultas al día, fue creado hace más de 4 años (FACEBOOK, 2019) y acumuló en tan solo el mes de abril más de 60 000 publicaciones y comentarios con información de contacto de diferentes profesionales, empresas y consultoras.

Facebook cuenta con un buscador para ayudar a los usuarios a encontrar algún *post* de interés. Este buscador genérico ayuda a buscar cualquier tipo de publicación dentro del grupo. Donde el usuario, proporciona un texto y el buscador selecciona las publicaciones que contengan alguna de las palabras del texto seleccionado. Sin embargo, a pesar que este grupo tiene tanta información disponible, las personas tienen dificultades al momento de buscar la información de contacto. Como consecuencia, los usuarios muchas veces prefieren volver a realizar su consulta a través de un nuevo post. Para entender las dificultades que se presentan al buscar información de contacto se distribuyó una encuesta a 1000 miembros de este grupo. Donde se realizaron diferentes preguntas, como “¿Qué tipo información busca?” o

“¿Qué limitaciones tiene al momento de buscar información de contacto?”. La encuesta se puede encontrar en línea (ILLANES, 2017).

El 85.7% de los encuestados utiliza el buscador que provee Facebook en estos grupos para consultar por información existente sobre algún producto o servicio. Un 58.3% indicó que ha utilizado los filtros que provee el buscador. Según la información que desean encontrar en los comentarios de una publicación, un 57.1% espera obtener números telefónicos y el 78.6% direcciones de referencia y perfiles o páginas de Facebook etiquetados. Con respecto a la descripción de su experiencia con la búsqueda manual por información de contacto en los resultados, el 80% indicó “Me toma mucho tiempo encontrar la información que quiero”.

Basados en la encuesta, se pudo identificar tres dificultades al momento de buscar información en este tipo de grupos:

- **Información semánticamente equivalente.** El buscador del grupo no considera información semánticamente equivalente. Por ejemplo, si un usuario introduce la palabra de búsqueda “zapatos deportivos” en el buscador que proporciona Facebook en el grupo “Alguien Sabe? Cbba” buscará la existencia de al menos una repetición de cada palabra y el buscador devolverá como resultado todas aquellas publicaciones y comentarios que contengan al menos una de las palabras de la frase.
- **Información redundante.** Entre las dificultades que se presentan al momento de realizar una búsqueda manual por todos los comentarios, el 80% respondió que hay muchas publicaciones repetidas y un 60% indicó que hay mensajes que estorban como “jajaja”, “yo no sé” y otros.
- **Filtros.** Una forma de mejorar estas situaciones aplicar los filtros que provee el buscador de Facebook. Por ejemplo, podría filtrar las publicaciones destacadas o más recientes, así como también filtrar las publicaciones de una persona particular, o por fecha de publicación. Sin embargo, estos filtros según la encuesta no facilitan la obtención de información que el usuario necesita. Pues una buena recomendación de un contacto, no necesariamente tiene relación con los filtros antes mencionados. Por ejemplo, una publicación destacada puede ser aquella que tiene muchos comentarios actuales, pero estos pueden ser comentarios negativos o sin información relevante.

3 Aprendizaje automático y análisis de sentimientos al rescate

Este artículo propone mejorar la búsqueda de información de contacto haciendo uso de aprendizaje automático y análisis de sentimientos. Mediante el uso de aprendizaje automático se propone categorizar todas las publicaciones, el objetivo es

agrupar todas las publicaciones semánticamente equivalentes, por ejemplo, todas aquellas donde se consulte por un pediatra. De esta forma, cuando el usuario ingrese un texto en el buscador, primero se determinará a qué categoría pertenece el mismo y se filtrarán solo las publicaciones que pertenezcan a esta categoría. Del mismo modo, una vez filtradas las publicaciones, se propone ordenar la misma en base al positivismo o negatividad de las mismas. Situando al principio de la respuesta a la búsqueda aquellas publicaciones más positivas.

3.1 Categorización automática de publicaciones

El aprendizaje automático enseña a la computadora, mediante métodos computacionales y algoritmos, a que “aprenda” directamente de los datos, sin tener que depender de una ecuación predeterminada que sirva como modelo (MATHWORKS, 2016). Para la categorización de las publicaciones se siguieron los siguientes pasos.

Red bayesiana ingenua. Para la categorización se utilizó una red bayesiana ingenua, que es un clasificador probabilístico fundamentado en el teorema de Bayes e hipótesis simplificadoras adicionales, estas se suelen resumir en la hipótesis de independencia entre las variables predictoras. Dado esto, es que recibe el apelativo de ingenuo. Esta red asume que la presencia o ausencia de una característica particular no está relacionada con la presencia o ausencia de cualquier otra característica, dada la clase variable. (SLOTH'S, 2015), (SUCAR, 2006). Se decidió utilizar una red bayesiana ingenua debido a que:

- Implementan el teorema de Bayes, el cual permite que se vayan adaptando con su uso y puedan combinar datos provenientes de dos o más fuentes diferentes y expresarlos en el grado de probabilidad.
- Son rápidas de entrenar y clasificar.
- No son sensitivas a características poco relevantes.
- Solo se requiere una pequeña cantidad de datos de entrenamiento para estimar los parámetros (las medias y las varianzas de las variables) necesarios para la clasificación.
- Manejan información discreta y subjetiva.
- No tienen problemas manejando flujos de datos continuos.

Conjunto de Datos. Para entrenar y validar la red bayesiana se descargaron más de 66 265 publicaciones del grupo “Alguien Sabe? Cbba”. Para la extracción de datos, se utilizó la página *web Grytics.com* (GRITYCS, 2018), esta permite obtener publicaciones y comentarios de grupos o perfiles de Facebook que se encuentren asociados a una cuenta. Las versiones gratuitas de esta aplicación web permiten

obtener información de los últimos 30 días. La Tabla 1, muestra los campos que se obtuvieron de cada publicación (*post*) del grupo.

Tabla 1. Datos almacenados de cada publicación.

| Dato | Descripción |
|---------------------|--|
| *Comment_id | Número identificador asignado por Facebook al comentario realizado. |
| *Post_id | Número identificador asignado por Facebook a la publicación realizada. |
| Post_content | Contiene el texto de la publicación realizada. |
| *Post_people_name | Nombre del usuario el cual ha realizado la publicación. |
| *Post_people_id | Número identificador del usuario el cual ha realizado la publicación |
| *Parent_comment_id | Número identificador |
| Comment_date | Fecha en la cual se ha realizado el comentario. |
| Comment_text | Contiene el texto del comentario realizado a una publicación. |
| *CommentPeople_name | Nombre del usuario que ha realizado un comentario. |
| *CommentPeople_id | Número identificador del usuario que ha realizado el comentario. |
| Comment_likes | Número que indica la cantidad de "Me gusta" que recibió un comentario. |
| *Comment_comments | Número que indica la cantidad de comentarios que recibió un comentario. |
| *Attachment | Enlace a algún archivo adjunto al comentario como imagen u otros. |
| *Group id | Número de identificador asignado al grupo. |
| *Groupname | Nombre del grupo en el cual se ha realizado la publicación o comentario. |
| + Category | Nombre de la categoría a la que pertenece el comentario. |

| Dato | Descripción |
|--------------------|--|
| + Has_Phone_Number | Número de 0 o 1 que indica si el comentario contiene algún número telefónico. 1 si contiene alguno y 0 en caso contrario. |
| + Has_Direction | Número de 0 o 1 que indica si el comentario contiene alguna posible dirección de referencia. 1 si contiene alguno y 0 en caso contrario. |
| + Has_Profile | Número de 0 o 1 que indica si el comentario contiene algún posible nombre usuario o página etiquetado. 1 si contiene alguno y 0 en caso contrario. |
| + Sentiment_Value | Número entre 0 y 1 que indica el valor obtenido del análisis sentimental aplicado al texto. |

(*) Campos no utilizados

(+) Campos añadidos

Como se puede observar en la anterior tabla, no se utilizaron aquellos campos marcados con un *, esto debido a que tras una demanda que sufrió Facebook este ya no proporciona información en estos campos, en su lugar contienen datos como “*unknown*”, información que no es relevante o no puede ser utilizada para el proyecto. Por otro lado, se añadieron 5 nuevos campos: *Category*, *Has_Phone_Number*, *Has_Direction*, *Has_Profile* y *Sentiment_value*. Los campos agregados se los lista a continuación:

- El campo *Category* se utilizó para almacenar la categoría a la que pertenece el comentario. Este campo fue ingresado manualmente para cada dato en el primer conjunto de entrenamiento del clasificador y de forma automatizada luego del entrenamiento del sistema.
- Los campos *Has_Phone_Number*, *Has_Direction* y *Has_Profile* fueron llenados mediante procesos automatizados los cuales detectan si el comentario contiene un número telefónico, una posible dirección o el nombre de alguna persona con la cual el usuario final se pueda contactar. Principalmente se utilizaron estos campos para aplicar los filtros a los resultados de búsqueda.
- Para el campo de *Sentiment_value* se asignó un valor entre 0 y 1 a cada dato tras la aplicación del análisis sentimental. La siguiente subsección describe como se calculó este valor.

Pre-procesamiento. Posterior a la extracción de los datos, se realizó una limpieza para mejorar la calidad de los datos de entrenamiento para el clasificador.

Este proceso de limpieza corresponde a la etapa de pre-procesamiento de datos para la generación del *corpus* de entrenamiento.

Durante este proceso:

- Se encontraron 9 771 comentarios repetidos que se eliminaron, esto debido a que las repeticiones no aportan un mayor valor al *corpus* de entrenamiento y no se los utiliza para fines estadísticos en el proyecto.
- Se encontraron comentarios vacíos que de igual forma se eliminaron, debido a que no es posible crear comentarios para llenarlos por ser opiniones de personas. Se eliminaron 5 170 campos vacíos.
- Se encontraron más de 100 000 emoticones y caracteres especiales que no son reconocidos por el clasificador, se procedió a eliminarlos debido a que no son de relevancia para la formación del *corpus* de entrenamiento.

Posteriormente, debido a que el modelo clasificador que emplea el teorema de Bayes, asigna un valor probabilístico a cada palabra en las frases de los datos de entrenamiento, aquellas palabras que no brindan un valor sustancial a la formación del *corpus* de entrenamiento fueron eliminadas. De esta forma, las búsquedas serán más precisas porque las probabilidades de ocurrencias de las palabras son más exactas y no disminuyen a causa de otras palabras que no aportan un valor para este proyecto.

De cada frase del *corpus* de entrenamiento se procedió a eliminar:

- Artículos, se los eliminó porque no se requiere identificar el género de la palabra para entrenar el clasificador.
- Signos de puntuación, debido a que no se requiere conocer donde delimita cada frase sino solamente obtener las palabras más importantes.
- Pre-posiciones, debido a que no es necesario conocer las relaciones de los elementos de la frase.

Al finalizar este proceso, la cantidad de datos resultantes es de 51 324.

Clasificación. En esta etapa se aplicaron Redes Bayesianas Ingenuas para clasificar comentarios, con el uso del clasificador de texto que provee la librería de *Textblob* (TEXTBLOB, 2018).

Para evaluar el clasificador se consideró la precisión como métrica, que está definida como la proporción de los casos pronosticados como positivos que son correctos, se la calcula mediante la siguiente fórmula:

$$\text{Precisión} = \text{Positivos_Verdaderos} / (\text{Positivos_Verdaderos} + \text{Falsos_Positivos})$$

Para obtener los datos de entrenamiento se procedió a realizar una clasificación manual de los primeros 200 comentarios en el campo *Category*, asignando la categoría

a la que pertenece cada uno. Se inició clasificando los primeros comentarios manualmente con 2 personas diferentes, y se encontró una similitud del 92% entre ambas clasificaciones, por lo tanto se procedió a trabajar con estos datos de entrenamiento iniciales.

Esta primera iteración dio pie a otras 7 iteraciones para preparar y mejorar los datos de entrenamiento para el clasificador. En cada iteración se incrementaron los datos de entrenamiento hasta llegar a 500 comentarios clasificados en la última iteración, para llegar a esta cantidad de datos de entrenamiento se automatizó su generación en base a datos recopilados en la iteración anterior. En el camino se aplicaron procesos de limpieza, como la eliminación de caracteres especiales no visibles para el ojo humano, permitiendo una precisión final de 84% en la clasificación de los comentarios. Una vez entrenado el modelo con estos 500 datos iniciales se procedió a la clasificación del resto de los más de 50 000 comentarios utilizando Redes Bayesianas Ingenuas.

A continuación se muestra la representación gráfica de la evolución de la precisión del clasificador a través de las iteraciones:

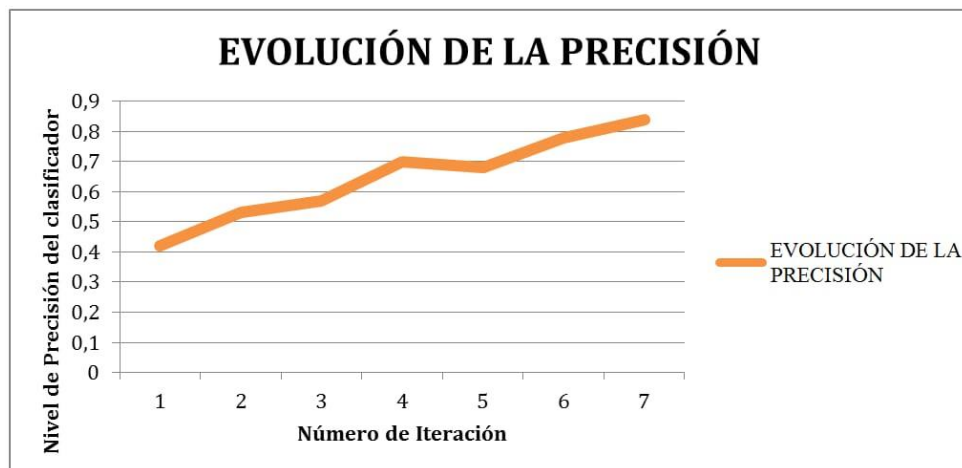


Figura 1: Gráfico de la evolución de la precisión.

Fuente: Elaboración propia, 2018.

3.2 Priorización en base a sentimientos

Una vez obtenidas las publicaciones resultantes de la búsqueda, se priorizan en base al análisis de sentimientos. El análisis de sentimientos, se basa en emplear técnicas computacionales para inspeccionar y determinar la opinión emocional de un texto dado, especialmente para identificar la actitud del escritor del texto como positiva, negativa o neutral. Para este análisis se utilizó la librería *Senti-py*, (GITHUB,

2018). Con esta librería se realizó el análisis sentimental de los comentarios que son devueltos como resultados de la búsqueda.

El análisis sentimental asigna un número entre 0 y 1 al texto analizado. Este número representa si el comentario expresa una opinión positiva, neutral o negativa. Las aproximaciones más cercanas al número 1 sugieren un alto nivel de positivismo expresado en el texto, las aproximaciones al 0 indican un alto nivel de negativismo y los valores cercanos a 0.5 expresan neutralidad en el texto.

Tras obtener los comentarios resultantes de una búsqueda, se realiza el análisis sentimental a cada uno de los resultados para priorizarlos, es decir ordenarlos y mostrar primero aquellos que tengan un nivel más alto de positivismo. La Tabla 2 muestra un ejemplo de los resultados de búsqueda con y sin aplicar la priorización.

Tabla 2. Ejemplo de priorización en base a sentimientos.

| ORDEN DE PRESENTACIÓN | | | |
|---------------------------------|---------------------------------|---|--------------------------|
| Sin Análisis sentimental | Con Análisis sentimental | TEXTO DEL COMENTARIO | VALOR SENTIMENTAL |
| 4to | 1ro | “YO LOS HAGO MUY ESPECIALES!! ESPONJOSOS DELICIOSOS 76910151” | 0.98202 |
| 1ro | 2do | “Claudia Mayorga lo realiza recomendable exquisitos” | 0.95589 |
| 7mo | 3ro | “Los mejores son de Nataly Bazoalto comunicate con ella esos si son alfajores mendocinos tipo argentinos” | 0.95530 |
| 3ro | 4to | “Nancy Franco hace deliciosos!!!” | 0.89077 |
| 2do | 5to | “Dulce Olivia...calle A. Padilla y Potosí” | 0.85710 |
| 6to | 6to | “Sonia Velasco hace los mejores alfajores te los aconsejo y a muy buen precio” | 0.84891 |
| 5to | 7mo | “Isabel Galindo, los de tu amiga los mejores.” | 0.05129 |

Como se puede apreciar en la anterior tabla, a medida que el comentario expresa una opinión más positiva, esta obtiene un valor más elevado en el análisis sentimental. Este valor es el que define el orden de aparición en la lista de resultados.

Primeramente, mostrando aquellos resultados cuyo análisis sentimental sean más altos.

4 Evaluación

4.1 Diseño experimental

- **Prototipo.** Para evaluar el impacto de la estrategia y priorización propuesta. Se desarrolló un prototipo mediante una aplicación móvil que proporciona un buscador que categoriza automáticamente las publicaciones recolectadas con *Grytics* y prioriza los resultados de búsqueda basada en el análisis sentimental. Tal como se describió en la sección anterior.
- **Participantes.** Los participantes para la prueba realizada están conformados por 15 usuarios frecuentes del grupo “Alguien sabe? Cbba”; 12 usuarios con alta experiencia y uso frecuente del grupo en consultas y búsquedas; 3 de los usuarios nunca habían utilizado el buscador en el grupo.
- **Actividades.** Para cada participante se le pidió realizar dos búsquedas sobre el mismo asunto de interés utilizando el buscador que provee Facebook en el grupo y luego utilizando el prototipo.
- **Recolección de Datos.** A cada participante se le pidió ordenar los 10 primeros resultados de la búsqueda, en base a la información de contacto que considere más útil. Tanto con el prototipo como con el buscador de Facebook, posicionando al principio las más útil y al final la que considera menos interesante.
- **Métricas.** Para poder comparar el orden de los resultados de búsqueda (con y sin análisis sentimental) con el orden que espera el usuario al realizar la búsqueda se utilizaron dos métricas.
 - **Primera respuesta pertinente.** La posición donde aparece la opción que el usuario considera más útil.
 - **Distancia Footrule.** La diferencia entre el orden propuesto por las herramientas con el orden esperado por los usuarios. Para cuantificar esta diferencia se utiliza la distancia Footrule, esta es una fórmula que permite comparar el orden entre los elementos de dos listas ordenadas. La distancia se calcula sumando los valores absolutos de las restas de las posiciones de sus ítems, tomando como referencia el orden de la primera lista ordenada (REVOLEDU, 2018). En este proyecto en particular, se toma como referencia el orden esperado por los participantes. Mientras más elevado sea el valor resultante, mayor será la diferencia de las

posiciones de ambas listas. Esta fórmula asume que las dos listas a comparar son dos permutaciones de un mismo conjunto de datos.

4.2 Resultados

Primera respuesta pertinente. La Figura 2; muestra la posición de la primera respuesta útil según la percepción de los participantes, tanto en el prototipo como en Facebook. Como se puede ver el prototipo posiciona la respuesta más útil al principio del resultado de búsqueda. Es decir, los participantes encontraron la información útil al principio de la lista, entre la primera y segunda posición. Por otro lado, usando la aplicación de búsqueda de Facebook, los usuarios encontraron información útil entre la quinta y décima posición.

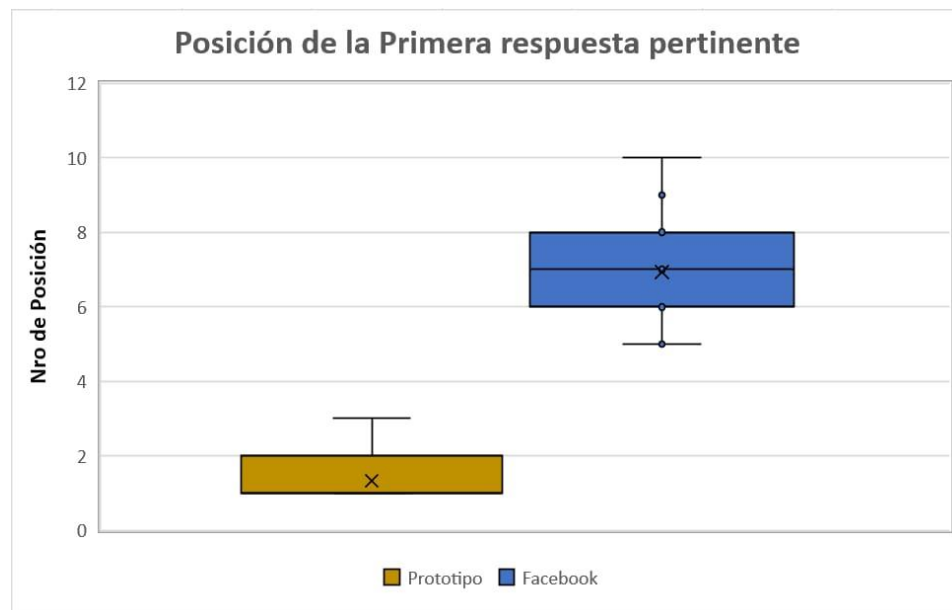


Figura 2: Primera respuesta pertinente entre los 10 primeros comentarios.

Fuente: Elaboración propia, 2018.

Orden de presentación de resultados (*Footrule Distance*). Como se puede apreciar en la tabla siguiente, tras aplicar la fórmula de *Footrule Distance* entre los resultados presentados por la aplicación móvil y el orden designado por los usuarios, se obtuvo un promedio de 1,2. Mientras que el valor obtenido de la comparación del orden de resultados designado por los usuarios y los presentados por Facebook, se obtuvo un valor promedio de 5,67. Esto significa que el orden en que se presentaron los resultados en el prototipo es más cercano (en términos de la distancia *Footrule*) al orden establecido (deseado) por los usuarios, en comparación al orden de los

resultados de *Facebook*. La Figura 3: muestra la distribución de las distancias de *Footrule*, haciendo uso de diagramas de caja.

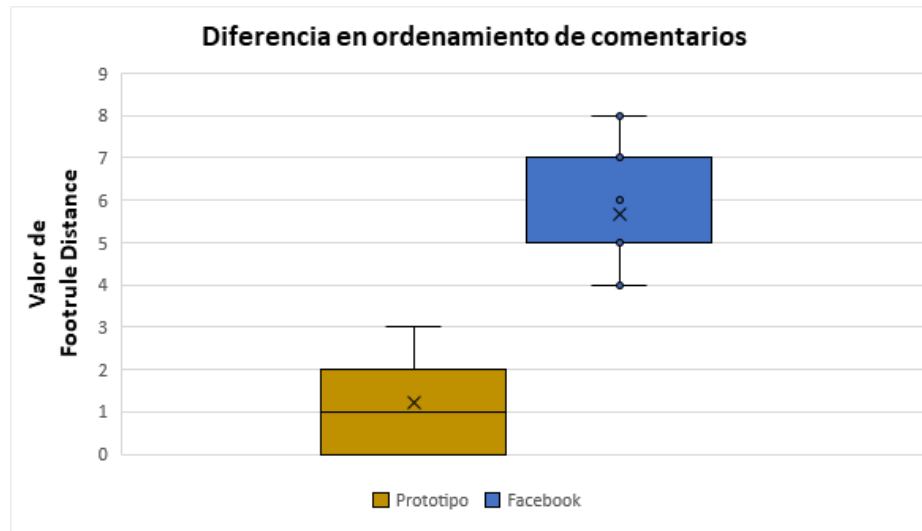


Figura 3: Diferencia de orden de resultados usando Footrule Distance.

Fuente: Elaboración propia, 2018.

5 Discusión y trabajo futuro

El acceso a los datos que proporciona Facebook mediante sus APIs se ha visto limitado luego de una demanda que sufrió en 2018, varias de las funcionalidades y, en especial, la información a la cual se podía acceder mediante la API fueron restringidas. La sección que se encarga del manejo de grupos de la API estuvo sin poder ser consumido por un prolongado tiempo, durante y después de finalizar la demanda a Facebook. La restricción afectó el acceso a los comentarios de grupos, siendo solo posible acceder a ellos tras la elaboración de una aplicación, la cual tiene que pasar por un proceso de aprobación por parte de Facebook. Por lo que el modelo de clasificación solo se basó en los datos proporcionados por *Grytics.com*. A pesar que el modelo tiene una buena precisión, podría ser posible mejorarlo si se tuviera acceso a más datos. Por ejemplo, si se pudieran conocer los datos de los usuarios que señalaron me gusta en un comentario.

Este trabajo presenta una evaluación con 15 participantes. A pesar de que los resultados fueron considerablemente positivos, como trabajo futuro, se planea realizar una herramienta que pueda monitorear el uso del prototipo en el día a día de los usuarios. De esta forma, se puede recolectar información del uso del prototipo por un periodo de tiempo más largo, con mayor cantidad de consultas de información de contacto y abarcando una mayor cantidad de participantes.

Como trabajo futuro también se planea realizar estudios más profundos sobre cómo las personas realizan consultas, haciendo uso de procesamiento del lenguaje natural, y analizar especialmente qué expresiones son comúnmente utilizadas para realizar consultas y expresar opiniones.

6 Conclusiones

El presente trabajo propone el uso de aprendizaje automático y análisis de emociones para mejorar la búsqueda y el orden en que los resultados de esta búsqueda son presentados al usuario. Este artículo presenta un prototipo que: 1) categoriza las publicaciones en Facebook usando redes bayesianas ingenuas y 2) ordena los resultados de búsqueda en base a sentimientos. Se realizó un experimento con 15 usuarios, donde se puede apreciar que el orden en el que el prototipo presenta los resultados de una búsqueda es más cercano a lo que el usuario espera, comparado con el orden que proporciona el buscador de grupos de Facebook.

Referencias Bibliográficas

- [1] FACEBOOK (2019). Alguien sabe? Cbba. Obtenido en: <<https://www.facebook.com/groups/905493379496191/about/>>, (fecha de consulta 10/06/2019).
- [2] GRITYCS (2018). "Sobre nosotros". Obtenido en: <<https://grytics.com/es/sobre-nosotros>>, (fecha de consulta 15/06/2018).
- [3] GITHUB, Senti-py. Obtenido en Github: <<https://github.com/ayllote/senti-py>>, (fecha de consulta 25/05/2018).
- [4] ILLANES, Daniel (2017). Encuesta sobre los grupos "Alguien sabe? Cbba" en Facebook En: <https://docs.google.com/forms/d/e/1FAIpQLSdJstI5Euti33R0hsFi9rbl1E.A1fuxkcEK7ZlqOzAdSjQj-A/viewform>
- [5] MATHWORKS (2016). "Introducing Machine Learning". Obtenido en: <https://la.mathworks.com/content/dam/mathworks/tagteam/Objects/i/88174_92991v00_machine_learning_section1_ebook.pdf>, (fecha de consulta 20/07/2018).
- [6] MATHWORKS (2016). "Applying Supervised Learning". Obtenido en: <https://la.mathworks.com/content/dam/mathworks/tagteam/Objects/i/90221_80827v00_machine_learning_section4_ebook_v03.pdf>, (fecha de consulta 22/07/2018).
- [7] REVOLEDU (2018). "Footrule Distance". En: <<https://people.revoledu.com/kardi/tutorial/Similarity/FootruleDistance.html>>, (fecha de consulta 09/11/2018).

- [8] SLOTH'S LAB (2015). "Clasificador Bayesiano Ingenuo". Obtenido en: <http://www.slotslab.com/python/2015/12/03/clasificador-bayesiano-ingenuo-python.html>. (fecha de consulta 09/11/2018)
- [9] SUCAR, L. E. INAOE (2006). Redes Bayesianas. Obtenido de inaoep. Obtenido en: <https://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>, (fecha de consulta 12/06/2018).
- [10] TEXTBLOB (2018). "Tutorial: Building a Text Classification System". Obtenido en: <https://textblob.readthedocs.io/en/dev/classifiers.html#classifying-text>, (fecha de consulta 10/06/2018).
- [11] TEXTBLOB (2018). Obtenido en: <https://textblob.readthedocs.io/en/dev/>, (fecha de consulta 10/06/2018).