



# REPORTE DE CONVERSACIÓN DE MIGRACIÓN Y XENOFOBIA

**COSTA RICA**

1 DE ABRIL 2020 A 1 DE ABRIL 2021



**#XENOFOBIAZERO**  
ALZA TU VOZ CONTRA LA DISCRIMINACIÓN



ÍNDICE

1. INTRODUCCIÓN .....	3
2. DATOS Y METODOLOGÍA.....	4
3. LA CONVERSACIÓN DE XENOFOBIA .....	4
4. CLASIFICACIÓN DE LA CONVERSACIÓN DE XENOFOBIA .....	11
5. CUANTIFICACIÓN DE XENOFOBIA POR NACIONALIDAD .....	16
6. CONCLUSIONES Y RECOMENDACIONES DE POLÍTICA .....	18
ANEXO 1. NOTICIAS Y XENOFOBIA .....	18
ANEXO 2. ESTIMACIÓN A PARTIR DE LDA .....	21

1. INTRODUCCIÓN

El presente informe muestra un análisis de la conversación en línea de xenofobia y lenguaje de odio hacia la población migrante en Costa Rica. Para realizarlo, se construyó una base de datos que recoge todas las publicaciones de xenofobia y lenguaje de odio en línea hacia las población migrante salvadoreña, venezolana, mexicana y nicaragüense entre el 1 de abril de 2020 y el 1 de abril de 2021. Durante este periodo se generaron un total de 4.307 mensajes de este tipo que tuvieron un alcance de alrededor de 6.397.190 usuarios. Esto incluye retweets y compartidos.

El mes con mayor cantidad de publicaciones con contenido xenófobo fue junio, debido principalmente a un aumento de la migración nicaragüense en contexto de pandemia por el Coronavirus. Al separar las publicaciones de xenofobia por la nacionalidad a que iban dirigidas, se encuentra que 39.4% iba dirigida a población migrante nicaragüense, 19.9% a población migrante venezolana, 14.2% hacia población migrante mexicana y 27% se refiere a migrantes en general.

Durante los brotes más importantes de xenofobia en el período analizado, las publicaciones con contenido de xenofobia se agruparon principalmente en una temática, la cual corresponde a la asociación de la población migrante nicaragüense con la expansión del Covid en Costa Rica.

Para estudiar a profundidad los distintos componentes de la conversación de xenofobia, se utilizó un modelo de clasificación de Machine Learning (LDA). Este permite clasificar las principales temáticas dentro de todas las publicaciones de xenofobia. Los resultados indican que hubo tres temáticas principales que se diferenciaron del resto. 1. El estatus migratorio de la población nicaragüense que está llegando a Costa Rica. La mayoría de estas publicaciones mostraban un sentimiento de rechazo e injusticia con el hecho de que la población migrante nicaragüense esté llegando de manera ilegal al país y se estén destinando recursos para la atención de esta población. 2. Un rechazo marcado en contra de las personas migrantes mexicanas y venezolanas, así como una asociación de estas con hechos criminales. 3. El descontento con el supuesto papel que están cumpliendo las personas migrantes en el agravamiento de la situación producto del covid-19. Estas publicaciones, además, expresaban un descontento por el hecho de que se les estuviera brindando asistencia y seguro médico al llegar al país.

Por último, a partir de los hallazgos de este informe se sugieren las siguientes recomendaciones de política:

1. Atender la asociación entre migración y contagio por covid. Además de generar estigmatización, este tipo de posturas puede obstaculizar el acceso de la población migrante a servicios de salud durante la crisis sanitaria. Es necesario combatir estos prejuicios para disminuir la discriminación en línea.
2. Proveer herramientas a la ciudadanía para reaccionar ante el aumento del discurso de odio en línea.
3. Tener en cuenta que las publicaciones de discriminación fueron orientadas principalmente a la población migrante proveniente de Nicaragua (40)%. Esto incluye los mensajes más violentos dentro de la conversación de xenofobia.

El resto del informe sigue de la siguiente manera: En la sección 2 se describen los datos y la metodología utilizada para procesarlos. En la sección 3 se presentan los eventos más importantes en la conversación de xenofobia durante el periodo. En la sección 4 se clasifican y analizan los tópicos dentro de la conversación de xenofobia. En la sección 5 se estudia la nacionalidad a la que van dirigidas las publicaciones de xenofobia. En la sección 6 se concluye y proponen recomendaciones de política.

2. DATOS Y METODOLOGÍA

Los datos utilizados en este informe provienen de la información pública de las redes sociales Facebook, Twitter, Instagram y Youtube en Costa Rica, entre abril de 2020 - abril de 2021. Para obtener y filtrar los datos se realizó un ejercicio de Web Scraping a través de la plataforma Meltwater. En este sentido, se organizaron y filtraron todos los posts públicos asociados a migración en Costa Rica. La organización se llevó a cabo a través de una selección de palabras claves, frases y asociaciones que permiten capturar la conversación de migración. Estas palabras y frases clave captan las conversaciones sobre distintas poblaciones migrantes y fueron construidas en conjunto con el equipo local de la OIM en el país y pueden encontrarse en el Anexo 1. En el caso de Costa Rica, las poblaciones migrantes incluidas en el análisis fueron la salvadoreña, mexicana, venezolana y nicaragüensae, por ser las mayores comunidades de migrantes en el país.

Una vez capturada y delimitada la conversación de migración, el segundo paso fue cualificar los mensajes para poder identificar las publicaciones de discriminación hacia las poblaciones migrantes. Para esto, se capturó la conversación de discriminación a partir de palabras, frases y expresiones que mostraban una actitud xenófoba hacia la población migrante. Estas palabras, frases y expresiones también fueron trabajadas con el equipo local de OIM. Además, para asegurarse de capturar solo publicaciones de discriminación, se eliminaron todas las publicaciones que denunciaban el comportamiento xenófobo. Los datos finales corresponden a la totalidad de publicaciones en línea que muestran actitudes de discriminación, rechazo o lenguaje de odio hacia alguna de las poblaciones migrantes en Costa Rica durante el período de estudio. En las siguientes secciones se procede a analizar los contenidos de estas publicaciones.

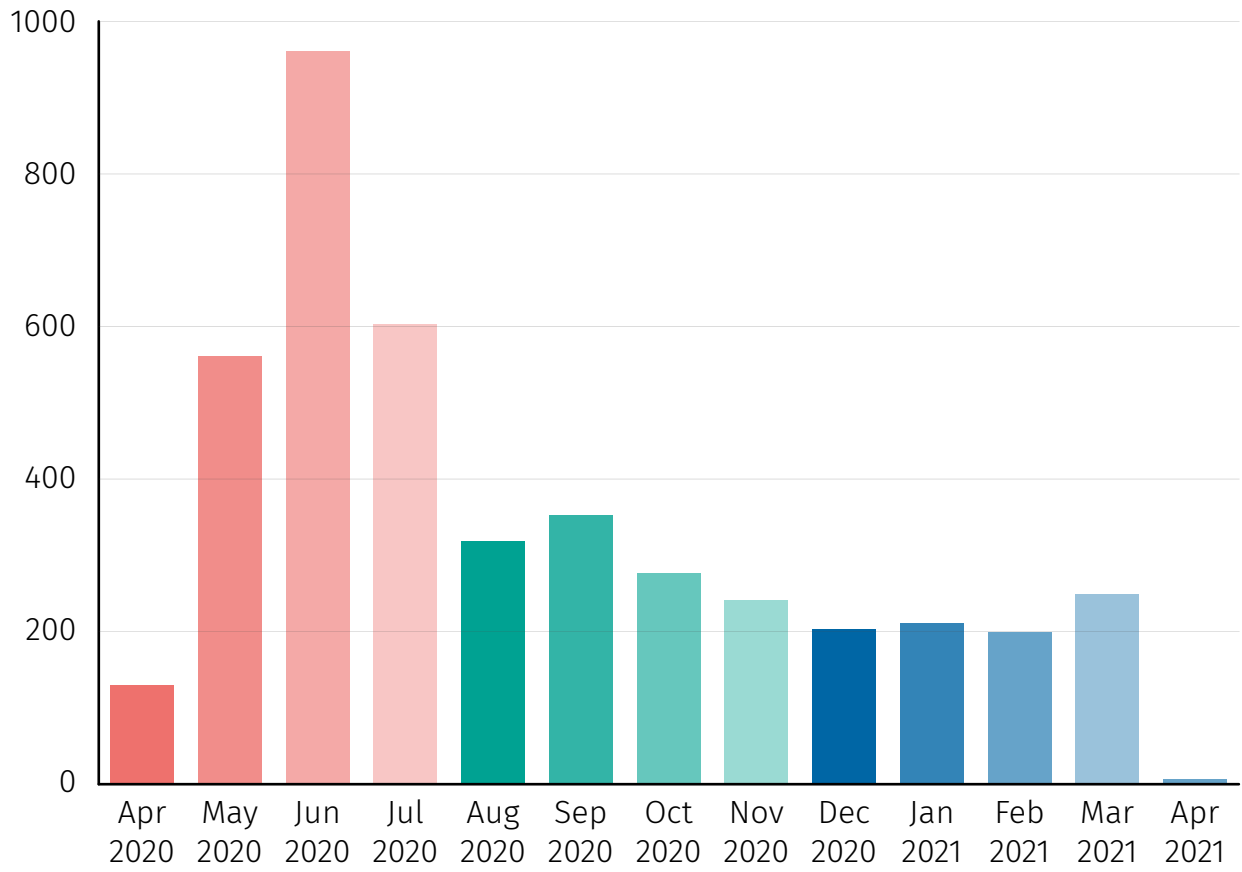
3. LA CONVERSACIÓN DE XENOFOBIA

En esta sección se estudia el comportamiento de la conversación de xenofobia en línea entre el 1 de abril de 2020 y el 1 de abril de 2021. Durante este periodo se generaron un total de 4.307 mensajes de xenofobia en línea. Con el propósito de explicar este comportamiento, primero se discuten los principales eventos que incidieron en esta conversación; segundo, se estudian los tópicos más relevantes dentro de esta conversación; y tercero, se discute la composición por nacionalidad dentro de la conversación de xenofobia.

3.1 Eventos más relevantes en la conversación de xenofobia

Una forma de entender los impulsores de las publicaciones de xenofobia es a partir de los eventos en donde más se generaron ese tipo de publicaciones. Durante el año de estudio, se encontró que hay tres meses que contienen publicaciones de xenofobia por encima del promedio del período. El gráfico 3.1 muestra que los meses con mayor incidencia de xenofobia en línea fueron mayo, junio y julio de 2020. A continuación, se hace una descripción de la conversación de xenofobia durante estos meses.

Gráfico 3.1 Volumen de Publicaciones de Xenofobia por Mes.



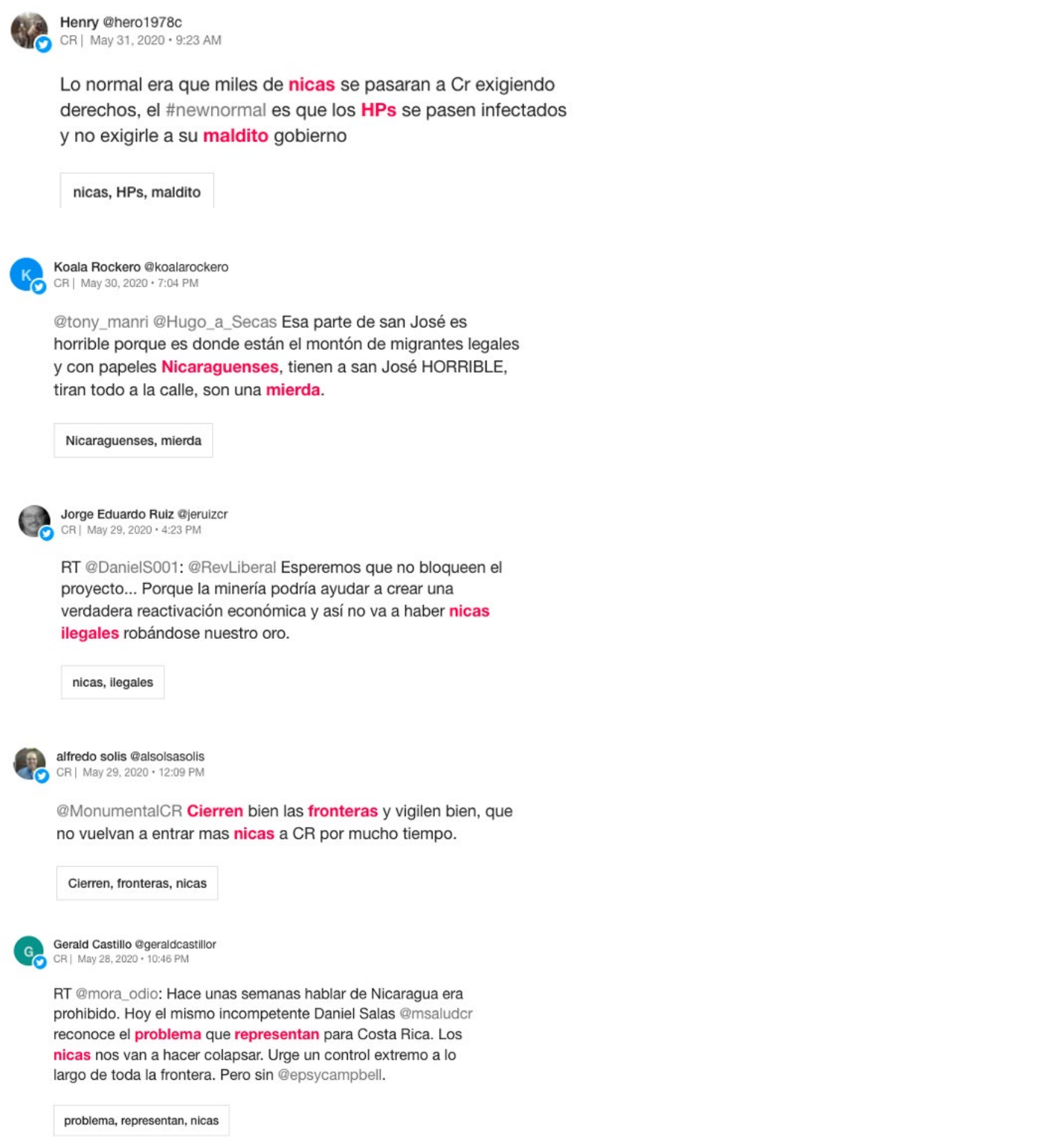
Meltwater (Construcción Propia), 2021.

Mayo 2020

El mes de mayo estuvo marcado por mensajes con contenido xenofóbico dirigidos a migrantes de nacionalidad nicaragüense. En la primera imagen de la gráfica 3.2 se puede apreciar que se asocia la entrada de personas de Nicaragua con los casos de Covid en Costa Rica. También se observa que se relaciona a las personas migrantes con el desorden en San José, como se muestra en la segunda imagen.

Además de lo anterior, en las publicaciones es posible evidenciar un llamado explícito al cierre de las fronteras de Costa Rica, mensajes que se dirigen con énfasis a la migración desde Nicaragua. Sumado a lo anterior, se observa que muchos de estos comentarios que buscan despreciar a las personas migrantes nicaragüenses, van acompañados de palabras ofensivas o de comentarios estigmatizantes hacia esta población.

Gráfico 3.2 Ejemplo Mensajes Junio de 2020

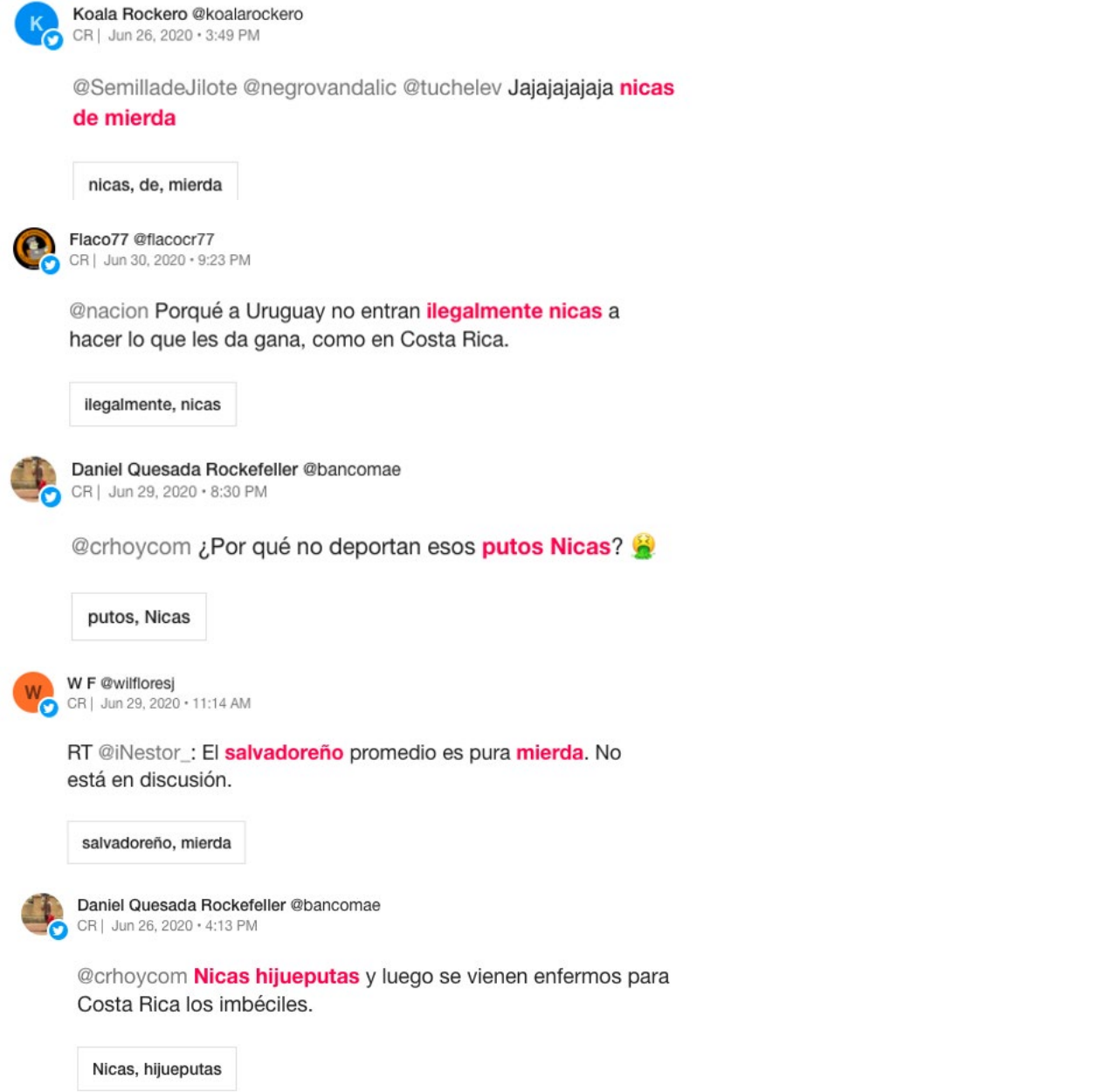


Junio 2020

Durante el mes de junio se produjeron un total de 962 mensajes xenofóbicos, siendo este mes el período con mayor número de menciones en contra de migrantes. En este sentido, los mensajes dirigidos hacia personas provenientes de Nicaragua fueron los que más abundaron en el tiempo analizado. Al igual que en el mes anterior, la principal temática apunta a la entrada ilegal de nicaragüenses en territorio de Costa Rica, en donde los usuarios hacen llamados a deportarles. Esto puede observarse en la gráfica 3.3, en la segunda, tercera y quinta imagen.

También se observa que se asocia la migración de nicaragüenses a la llegada de enfermedades a Costa Rica, como se aprecia en el último mensaje. Además, se puede ver cómo estos discursos de xenofobia van acompañados de insultos en la mayoría de los casos.

Gráfico 3.3 Ejemplo Mensajes Junio de 2020



Meltwater, 2021.

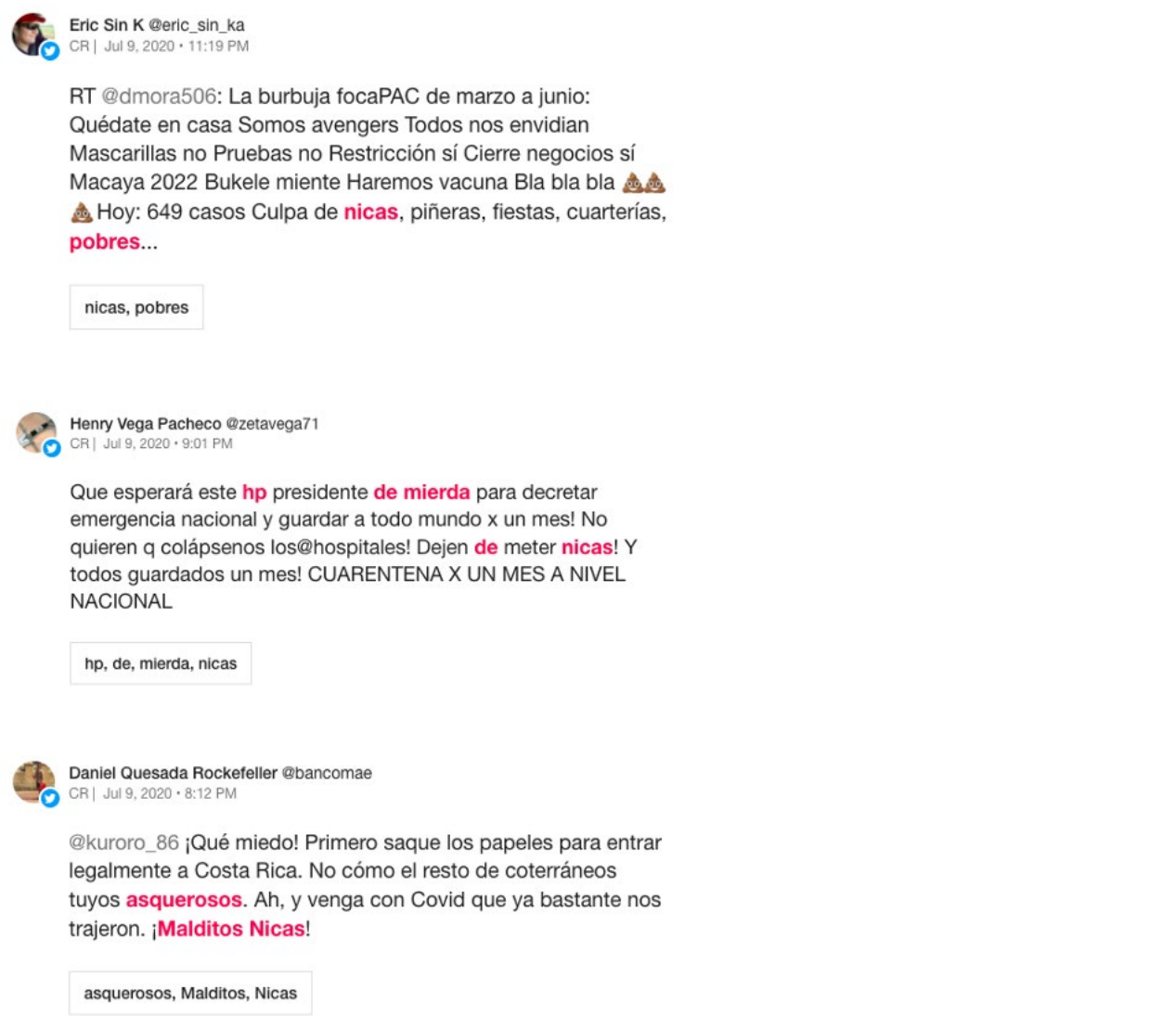
Si bien la mayoría de los mensajes xenofóbicos son dirigidos en contra de migrantes nicaragüenses, existen publicaciones hacia otras nacionalidades, como se puede evidencia en el cuarto post en donde se nombra a las personas migrantes provenientes de El Salvador.

Julio 2020

Durante el mes de julio, la conversación de xenofobia estuvo concentrada principalmente en contra de las personas migrantes nicaragüenses, al igual que en el mes de mayo y junio. Sin embargo, a diferencia del mes pasado, los mensajes apuntan al contexto sanitario debido a la pandemia, igual que en el mes de mayo. En este sentido, las publicaciones en contra de personas provenientes de Nicaragua apuntaron a que estas llevaron el coronavirus a Costa Rica, como puede observarse en las imágenes del gráfico 3.4.

En los mensajes se observa que los argumentos que se utilizan a la hora de emitir publicaciones xenofóbicas, está el hecho de que las personas migrantes nicaragüenses son asquerosas y que viven en barrios marginales a donde “llevaron el virus”.

Gráfico 3.4 Ejemplo Mensajes Julio de 2020



Meltwater, 2021.

Xenofobia y COVID-19

Un posible foco de discriminación hacia la población migrante es la asociación de esta con temáticas relacionadas a la crisis sanitaria originada por el COVID-19. Existen varias razones por las que la crisis sanitaria puede estar relacionada a una postura de rechazo hacia las personas migrantes. Por ejemplo, se puede tener una percepción de que traen consigo la enfermedad desde su país de origen o, por otro lado, una percepción de escasez por la crisis sanitaria y competencia de recursos. En total, los mensajes de discriminación que vincularon a la población migrante con temáticas relacionadas al COVID-19 o la crisis sanitaria representaron un 5% del total de mensajes de xenofobia. Si bien este es un porcentaje relativamente bajo, se requiere más investigación para entender cómo surge y qué forma tiene el vínculo entre los procesos migratorios y esta coyuntura.

Publicaciones de Integración

A lo largo del período de estudio también se generaron publicaciones que denunciaron el comportamiento xenófobo. Estas publicaciones mostraron dinámicas de integración hacia la población migrante y rechazaron su estigmatización. Durante el período se generaron 37 mil publicaciones de integración. Esto representa más de ocho veces la cantidad de publicaciones con contenido xenófobo. El gráfico 3.5 muestra ejemplos representativos de las publicaciones de integración.

Gráfico 3.5 Ejemplo de publicaciones de integración



4. CLASIFICACIÓN DE LA CONVERSACIÓN DE XENOFOBIA

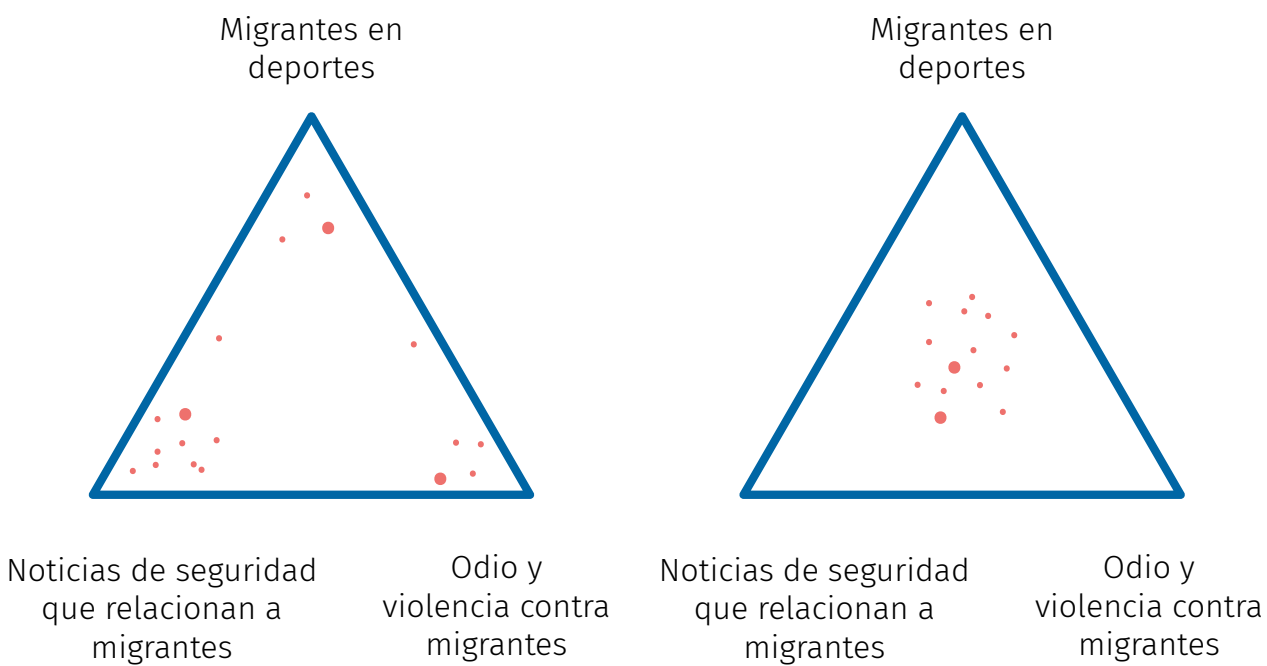
En la sección anterior se discutieron los momentos más importantes de la conversación de xenofobia, así como los impulsores de estos. Sin embargo, otra aproximación para entender el contenido dentro de las publicaciones de xenofobia es estudiando los tópicos o temas utilizados dentro de esta conversación. Existen modelos de aprendizaje de máquinas que permiten identificar las temáticas utilizadas dentro de una conversación. En esta sección se sugiere la utilización de un modelo LDA (Latent Dirichlet Allocation) y se muestran los resultados para la clasificación de la conversación de xenofobia en Costa Rica.

4.1 Modelo LDA

Para poder estudiar las temáticas más frecuentes dentro de la conversación de xenofobia en Costa Rica, se utiliza un modelo de tópicos. Este modelo también es conocido bajo el acrónimo LDA (Latent Dirichlet Allocation) en honor a la distribución de Dirichlet, la cual permite estimar la probabilidad de que cada publicación o documento pertenezca a un tópico determinado. La intuición detrás de este modelo resulta de tomar una colección de publicaciones (corpus de documentos) en la que pueden existir varios tópicos que se repiten entre ellas. En este sentido, se espera que las palabras que aparecen frecuentemente juntas tienen significados parecidos. A estos significados parecidos se les llama tópicos. Cada tópico es una combinación de diferentes palabras y cada documento es una combinación de diferentes tópicos.

El LDA es un modelo que intenta determinar el mecanismo subyacente que genera las publicaciones de xenofobia y sus tópicos. Una manera de interpretar el modelo es imaginando que todas las publicaciones de xenofobia son generadas por una máquina, pero no podemos observar dicha máquina ni sus criterios, solo lo que ella produce. LDA simula un conjunto de máquinas con diferentes criterios y selecciona la que mejor se acomoda a los resultados de las publicaciones de xenofobia evidenciados. El gráfico 4.1 muestra la forma en que se pueden separar los tópicos dentro de una conversación con dos ejemplos. En uno los tópicos están separados entre sí (izquierda) y en otro los tópicos se encuentran más juntos (derecha). Entre más separados los tópicos más fácil es analizar su contenido como una temática única.

Gráfico 4.1: Ejemplo de la separabilidad entre tópicos

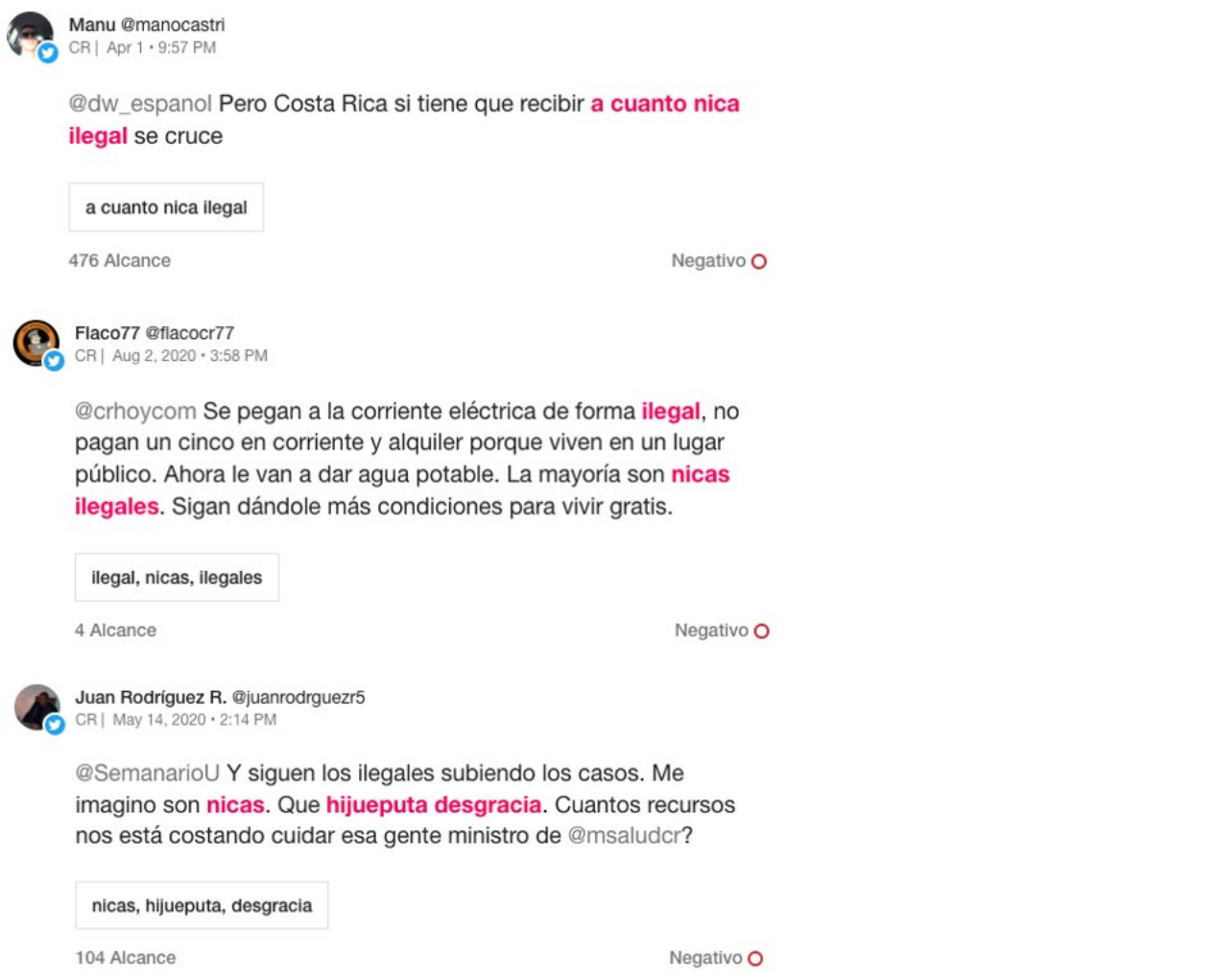


4.2 Resultados

Los resultados de la metodología propuesta sugieren que la conversación de xenofobia estuvo organizada en ocho tópicos diferentes. El primer tópico estuvo relacionado con el estatus migratorio de la población nicaragüense que estaba llegando al país. El segundo tópico tuvo que ver con un rechazo marcado en contra de las personas migrantes mexicanas y venezolanas, así como una asociación de estas con hechos criminales. El tercer tópico tuvo que ver con la idea de que las personas migrantes estaban llegando al país exacerbando la problemática surgida a causa del coronavirus. El contenido del resto de los tópicos no se diferencia entre sí e incluye temas varios como participación de migrantes en crimen, burlas u ofensas a la población migrante El gráfico 4.6 del anexo muestra la forma en que se distribuyen los tópicos dentro de esta conversación.

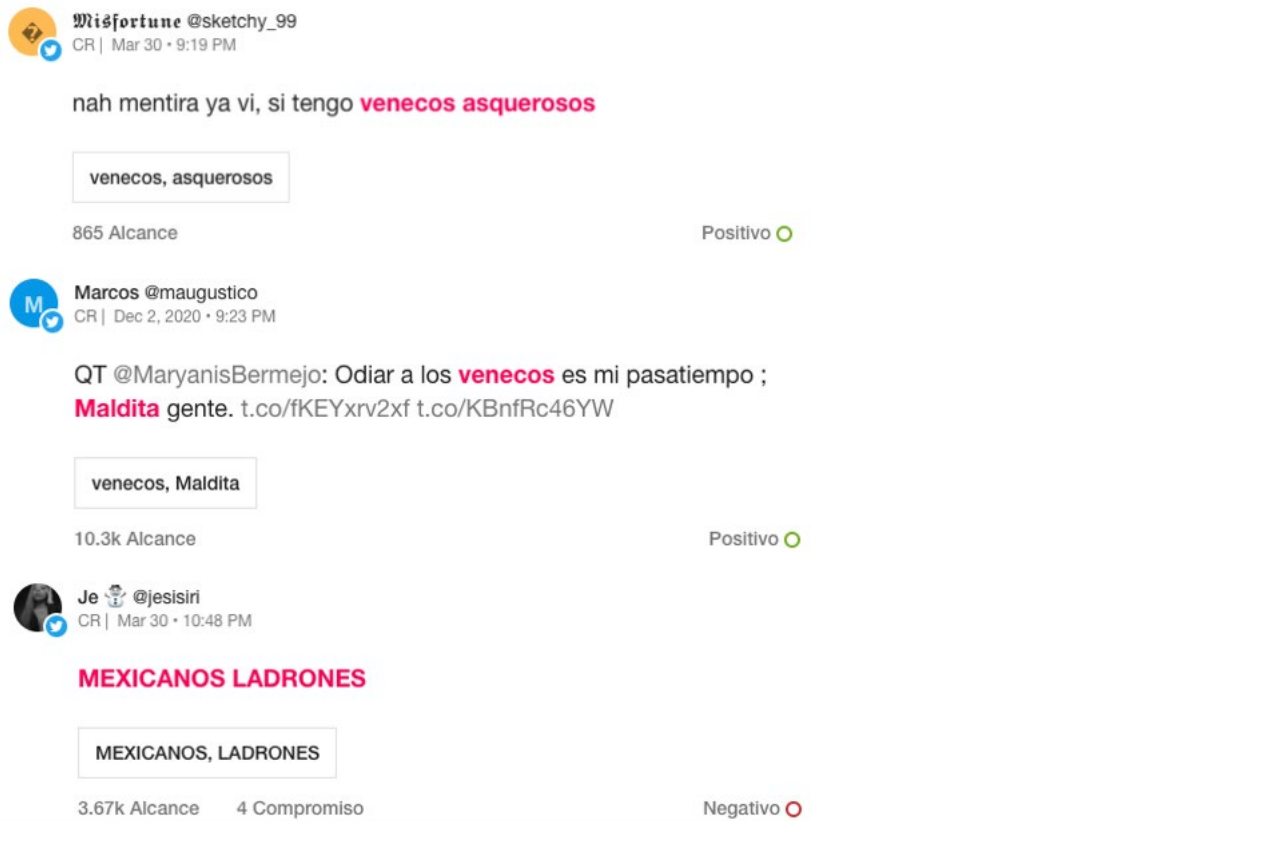
El tópico que más agrupó publicaciones de xenofobia (tópico 1) tuvo que ver con el estatus migratorio de las pesonas migrantes, principalmente nicaragüenses. La mayoría de estas publicaciones mostraban un sentimiento de rechazo e injusticia con el hecho de que la población migrante nicaragüense esté llegando de manera ilegal al país y se estén destinando recursos para la atención de esta población. Un hecho interesante es que gran parte de esta conversación surgió a propósito de noticias relacionadas con la atención a migrantes en temas de servicios públicos dispuesta por el Gobierno Costarricense. En el gráfico 4.7 se pueden observar los principales términos que contiene el tópico en cuestión.

Gráfico 4.7: Distribución general de términos en el tópico 1



El segundo tópico que más agrupó publicaciones consiste de distintas conversaciones que muestran rechazo y odio en contra de la población migrante mexicana y venezolana (tópico 2). Los mensajes aquí observados se caracterizan por ser ofensas en contra de las dos poblaciones migrantes ya mencionadas, algunos ejemplos de esto son palabras como “pobres”, “asquerosos” o “basura” que son utilizadas para referirse tanto a personas venezolanas como mexicanas. Ahora bien, es interesante notar que en este tópico se agrupan también los mensajes que asocian a migrantes con crímenes como el hurto. En el gráfico 4.8 se pueden observar los principales términos que contiene el tópico en cuestión.

Gráfico 4.8: Distribución general de términos en el tópico 2



Por último, otro tópico interesante mostró un descontento con la idea de que la población estaba generando un impacto negativo en las repercusiones que tuvo la pandemia en Costa Rica (tópico 3). Este tipo de mensajes tuvo dos características principales. Primero, denuncias de presuntas violaciones a las medidas tomadas por el Gobierno de Costa Rica para combatir el coronavirus. Segundo, se observa que algunos autores muestran una inconformidad por el hecho de que se le esté brindando un seguro médico a la población migrante. En el gráfico 4.9 se pueden observar los principales términos que contiene el tópico en cuestión.

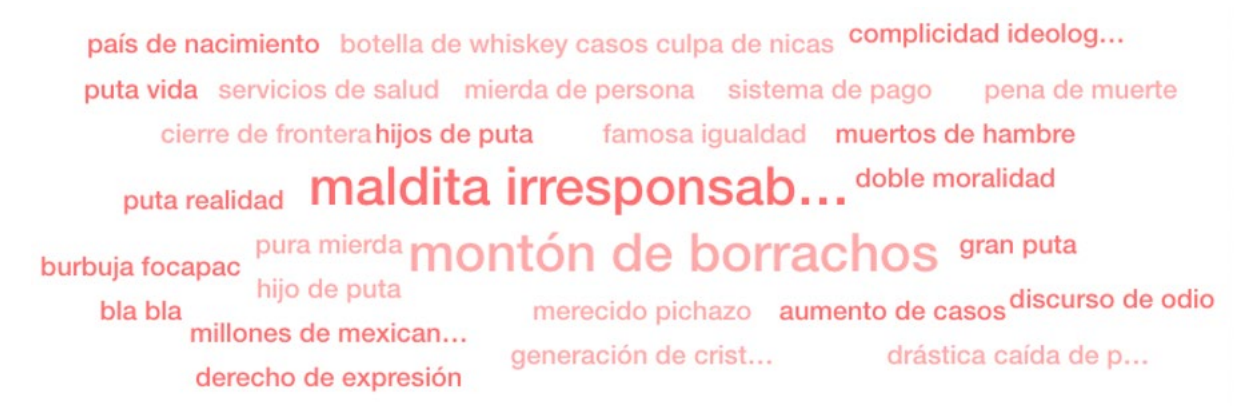
Gráfico 4.9: Distribución general de términos en el tópico 3



Palabras y frases más frecuentes en las publicaciones de xenofobia

Además de los tópicos dentro de la conversación de xenofobia, un aspecto interesante a detallar es la forma que tienen los comentarios de xenofobia. Es decir, si existen algunas expresiones o palabras que sean usadas recurretemente dentro del lenguaje de odio y discriminación. El gráfico 4.10 es una nube de palabras que clasifica las palabras según la frecuencia con la que fueron usadas a lo largo del periodo de análisis. Entre mayor sea el tamaño de la frase o palabra dentro de la nube, fue más alto el número de veces que se usó dentro de la conversación.

Gráfico 4.10 Nube de palabras de la Conversación de Xenofobia



Meltwater, 2021.

Las frases más utilizadas fueron “maldita irresponsabilidad”, “montón de borrachos” y “merecido pichazo”. Esto muestra la forma en que se mueve el discurso de odio hacia la población migrante. Un hecho interesante es que la mayoría de los mensajes encontrados van contra la población migrante nicaragüense y se desarrollan sobre dos temáticas principales. En primer lugar, se genera la idea de que la propagación del coronavirus en Costa Rica se debe a la llegada de migrantes nicaragüenses “irresponsables” que cruzan la frontera y generan un aumento en los casos presentados del virus. Ahora bien, como se observa en la tercera frase, hay mensajes que enaltecen la violencia contra las personas migrantes bajo la premisa de que estas merecen ser golpeadas, notando además que esta conversación se da en un contexto donde la población migrante es asociados con algún tipo de crimen.

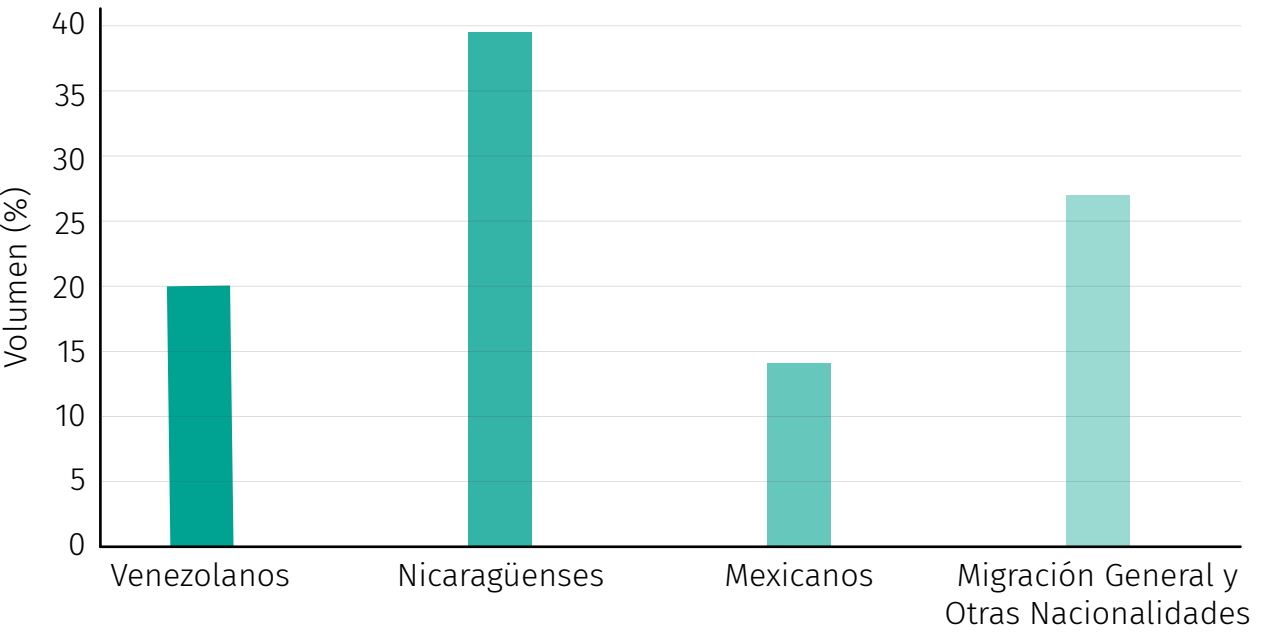
Es importante destacar la relación que los autores de estos mensajes hacen entre las persons migrantes y los problemas del país. En particular, la frase “aumento de casos” enmarca estas publicaciones, ampliamente utilizadas y que generan la percepción de que son las personas nicaragüenses quienes tienen la culpa de que se agrave la situación vivida por el país en términos de salud. Sumado a esto, la frase “muertos de hambre” alude a la idea de que la población migrante llega a Costa Rica en condiciones de pobreza, lo cual genera un deterioro social y peores condiciones de vida. Ahora bien, se encuentra que “cierre de frontera” fue una frase ampliamente usada a lo largo de la conversación. Esta frase refleja la posición de algunos autores según los cuales el cierre de la frontera es la mejor solución para las problemáticas sociales y sanitarias a las cuales se ha enfrentado el país durante el periodo de análisis.

5. CUANTIFICACIÓN DE XENOFOBIA POR NACIONALIDAD

Por último, un elemento importante en el análisis de la conversación de xenofobia en línea es entender la población a la que van dirigidos. En específico, dentro de Costa Rica existen migrantes de distintas nacionalidades, siendo las poblaciones migrantes nicaragüense, venezolana y mexicana las más importantes. En esta sección se estudia la distribución de las publicaciones de xenofobia por nacionalidad a la que van dirigidos.

Mediante la aplicación de análisis de texto y, particularmente expresiones regulares, es posible establecer la nacionalidad a la que van dirigidas las publicaciones de xenofobia. La forma de hacer esto es mediante la búsqueda de palabras o frases que denoten nacionalidad dentro del mensaje de xenofobia. Cabe resaltar que algunas publicaciones discriminatorias no expresan directamente la población, es decir, están orientadas a la población migrante en general; estas son clasificadas como “Migración General y Otras Nacionalidades”. En este sentido, la distribución calculada es una cota inferior de los mensajes de discriminación publicados hacia cada nacionalidad.

Gráfico 5.1 Volumen de Xenofobia por Nacionalidad



Meltwater (Construcción Propia), 2021.

El gráfico 5.1 muestra el porcentaje de mensajes que se refirieron a migrantes de nacionalidad nicaragüense, venezolana o mexicana. Salta a la vista que los mensajes xenófobos hacia la población nicaragüense corresponden a un 39.4%, siendo la nacionalidad con la mayor participación entre las incluidas en el análisis y superando lo observado para el caso de los mensajes en contra de migrantes venezolanos, los cuales corresponden al 19.9% de la conversación. Adicionalmente, el 14.2% de los mensajes de xenofobia fueron en contra de población mexicana, mientras que el 27% de los mensajes no se definen dentro de ninguna de las nacionalidades anteriormente descritas.

Gráfico 5.2 Ejemplo Mensajes de Xenofobia por Nacionalidad



Meltwater, 2021.

Un ejemplo de los resultados aquí vistos se observan en el gráfico 5.2, así como en la sección 3.1 donde es posible identificar que los mensajes producidos durante los picos de mayor volumen en la conversación se encuentran dirigidos en su mayoría en contra de migrantes nicaragüenses, por sobre el resto de nacionalidades.

6. CONCLUSIONES Y RECOMENDACIONES DE POLÍTICA

El análisis de las publicaciones de xenofobia y lenguaje de odio en línea arroja varios hallazgos que pueden ser utilizados para mitigar los riesgos asociados al incremento de la xenofobia. Primero, existen coyunturas en las que las publicaciones de odio hacia la población migrante incrementan significativamente. Durante el período de estudio estas coyunturas estuvieron asociadas a la publicación de noticias que relacionaban a migrantes con crímenes violentos y al anuncio de políticas públicas, como la atención gratuita el parto a las mujeres migrantes. Segundo, el rechazo hacia la población migrante parece estar agrupado en prejuicios y mitos. Además del rechazo hacia la atención del parto de las migrantes, otros prejuicios incluyen la percepción de la población migrante haitiana como una amenaza para la seguridad de la nación y el supuesto obstáculo que representa la comunidad internacional para tomar políticas de mano dura. Para mitigar el aumento de la discriminación hay que tomar en cuenta estos prejuicios. Teniendo en cuenta lo anterior, se proponen las siguientes recomendaciones de política.

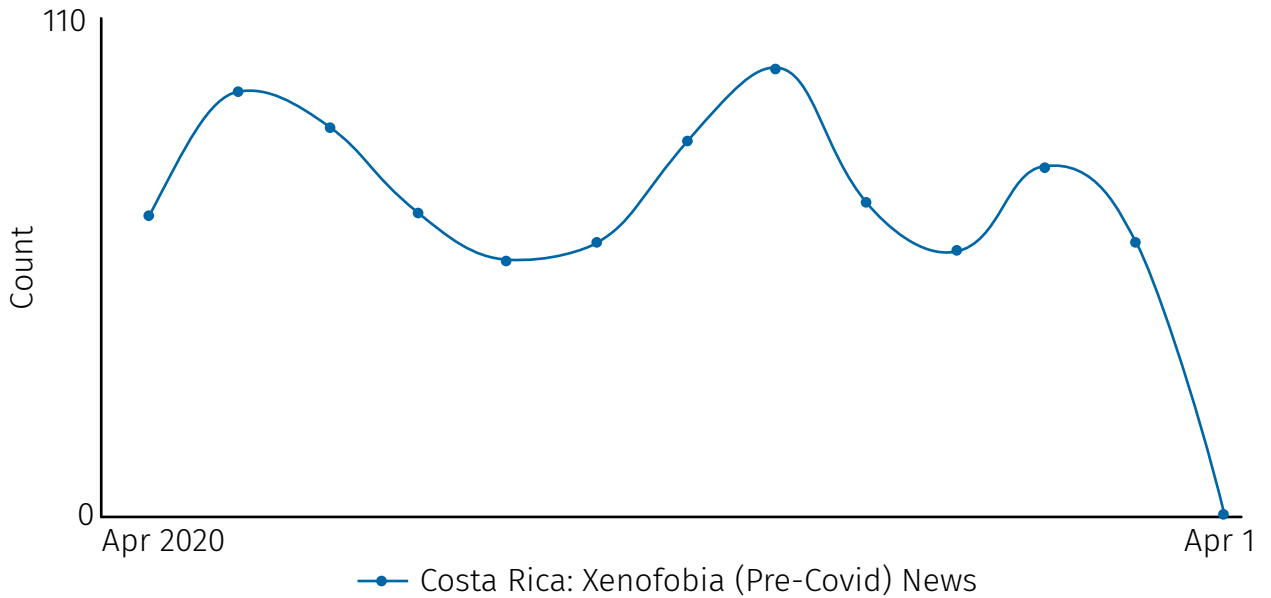
Recomendaciones

- Una de las principales ideas alrededor de los comentarios de xenofobia tiene que ver con la asociación entre las personas migrantes nicaragüenses y el aumento en el contagio del COVID-19. Estas publicaciones, además, solicitaban medidas como el cierre de la frontera para contener la migración y el contagio. Además de generar estigmatización, este tipo de posturas puede obstaculizar el acceso de la población migrante a servicios de salud durante la crisis sanitaria. Es necesario combatir estos prejuicios para disminuir la discriminación en línea.
- Las publicaciones de discriminación fueron orientadas principalmente a la población migrante proveniente de Nicaragua (40)%. Esto incluye los mensajes más violentos dentro de la conversación de xenofobia. Esto debo ser tenido en cuenta en cualquier esfuerzo orientado a disminuir la discriminación en línea.
- Deben proveer herramientas a la ciudadanía para reaccionar ante el aumento del discurso de odio en línea. Es importante evitar un escalamiento en la violencia en línea que polarice aún más la conversación sobre migración.

ANEXO 1. NOTICIAS Y XENOFOBIA

Durante el periodo estudiado, pudimos evidenciar que se han generado 20.057 publicaciones sobre migrantes, con un promedio de 55 publicaciones diarias. Sin embargo, no todas estas son catalogadas como xenófobas. Luego de realizar una nueva regla, se pudo apreciar que solo un 4.6% del total de publicaciones fueron con rasgos xenófobos. Es decir, 887 publicaciones, lo que implica un promedio de 2 publicaciones diarias.

Gráfico de volumen por mes, publicaciones xenófobas en medios.



Meltwater, 2021.

El peak de mayo se genera por el ingreso a través de un paso irregular de un ciudadano de Nicaragua contagiado de Covid. Otros ejemplos de noticias con rasgos xenófobos son:

Nicaragüense asalta negocio y policías lo detienen

Tras gritos de auxilio de la víctima, Heredia Enviar noticia por correo electrónico pidiendo ayuda llevaron a los oficiales de Fuerza Pública a **detener** un **nicaragüense** como sospechosos de asaltar un negocio donde labora la



Migración captura a coyote que por 30 rojos trasladaba a nicaragüenses ilegales

Extranjeros fueron devueltos a su país siendo transportados hoy”, explicó Migración. El supuesto **coyote**, quien también es **nicaragüense**, fue llevado a la Fiscalía de Flagrancia de

Operativo busca frenar entrada ilegal de nicas

En Sarapiquí y San Carlos se instala cordón policial y ríos utilizados por **coyotes** para el tráfico de personas indocumentadas provenientes mayoritariamente de **Nicaragua**. Cualquier informe

Conozca el fuerte control que ejerce Nandayure contra los ilegales nicaragüenses

Redacción – El alcalde de Nandayure, Giovanni Jiménez, ejerce un fuerte control contra los ilegales nicaragüenses que llegan al cantón ...  
 , Giovanni Jiménez, ejerce un fuerte control contra los **ilegales nicaragüenses** que llegan al cantón guanacasteco a trabajar. Nandayure



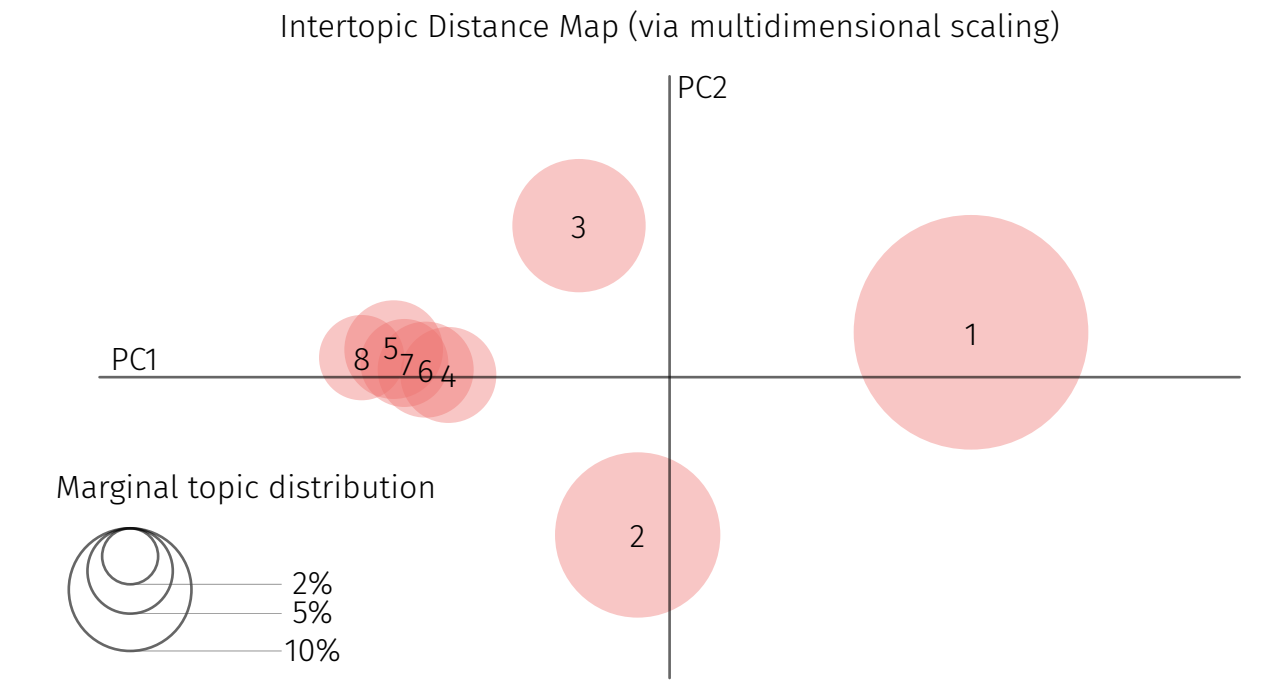
Baquianos muestran puntos ciegos a policía

Un grupo de vecinos de pueblos en la frontera norte se unió y pide una reunión con las autoridades de seguridad con el fin de llegar a los ...  
 con el fin de llegar a los puntos ciegos por donde entran **nicaragüenses ilegales**. Juan Carlos Bolaños asegura ser el representante de esta



ANEXO 2. ESTIMACIÓN A PARTIR DE LDA

Gráfico 4.6: Distribución general de tópicos dentro de la conversación de xenofobia

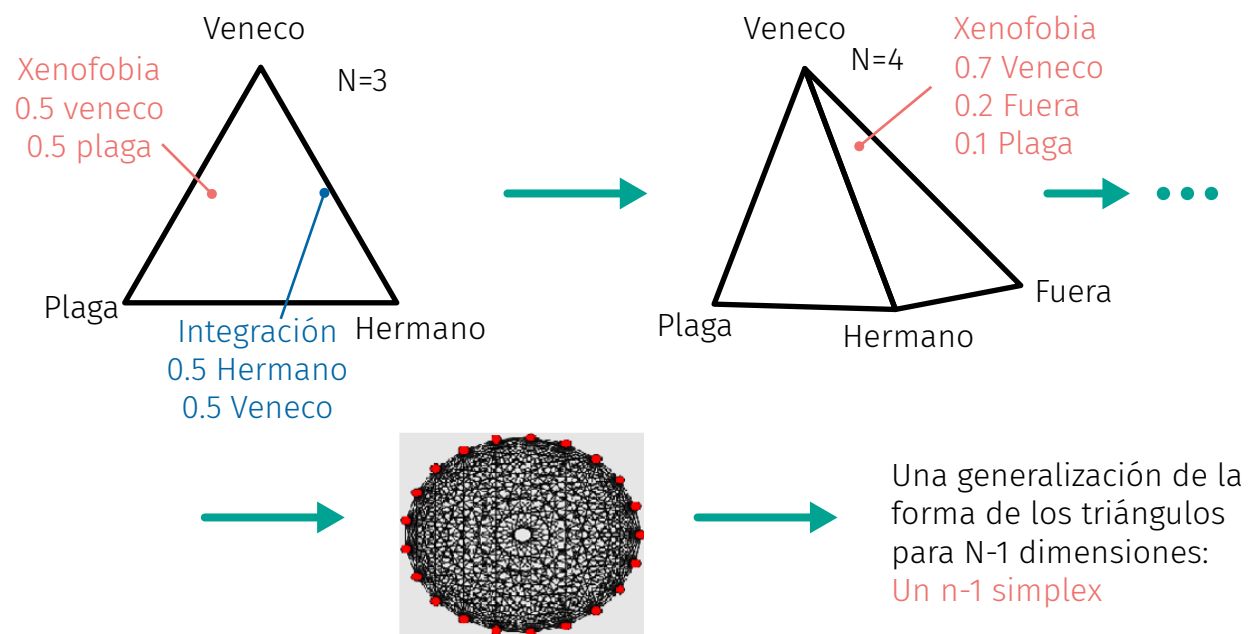


Fuente: Creación propia usando pyLDAvis. Los números representan el orden de los tópicos en términos de número de publicaciones. Es decir, el tópico 1 es el que agrupó más publicaciones y el tópico 8 el que menos. La interpretación del contenido de cada tópico es subjetiva al investigador. Esto quiere decir que, si bien el modelo propone la organización de unos tópicos basadas en las palabras que contienen cada Tweet, es trabajo del investigador estudiar las publicaciones que entraron dentro de cada grupo para determinar de qué se trata el tópico.

Una vez identificados los tópicos se puede calcular qué tan diferentes son entre ellos a partir del prior de Dirichlet. Esto es una métrica que nos permite entender qué tan separados o mixtos están los tópicos entre ellos usando como referencia los documentos en su interior. La distribución de Dirichlet suele representarse como un triángulo equilátero.

Como muestra la gráfica 4.1, a la derecha se encuentra un triángulo en donde cada uno de sus vértices es un tópico que puede aparecer en la conversación capturada a través de Meltwater. Los puntos que están al interior de este triángulo son los documentos o publicaciones que resultan de esta captura. Se puede observar como estos puntos están agrupados en el centro del triángulo, esto quiere decir las publicaciones están distribuidas de forma similar entre tópicos y que entre ellas existen diferencias reducidas (A esto le llamamos un alto prior de Dirichlet;  $\alpha > 1$ ). Por su parte, el triángulo ubicado a la izquierda describe una situación en donde las publicaciones pertenecen con mayor probabilidad a un tópico específico y están separadas fuertemente entre ellas (un bajo prior de Dirichlet;  $\alpha < 1$ ). Existen entonces, por su parte, palabras que pertenecen a varios tópicos (se conoce como el prior de Dirichlet sobre los términos;  $\beta$ ); LDA permite establecer cuales son las palabras más usadas por tópico y con esa información entender la clasificación que está realizando sobre las publicaciones. Sin embargo, esto genera un problema de visualización, la idea de poseer una combinación lineal de tópicos por palabra supone que, si el diccionario es lo suficientemente grande, dejemos de visualizar ese triángulo equilátero en donde cada esquina es una palabra (véase gráfico 4.2).

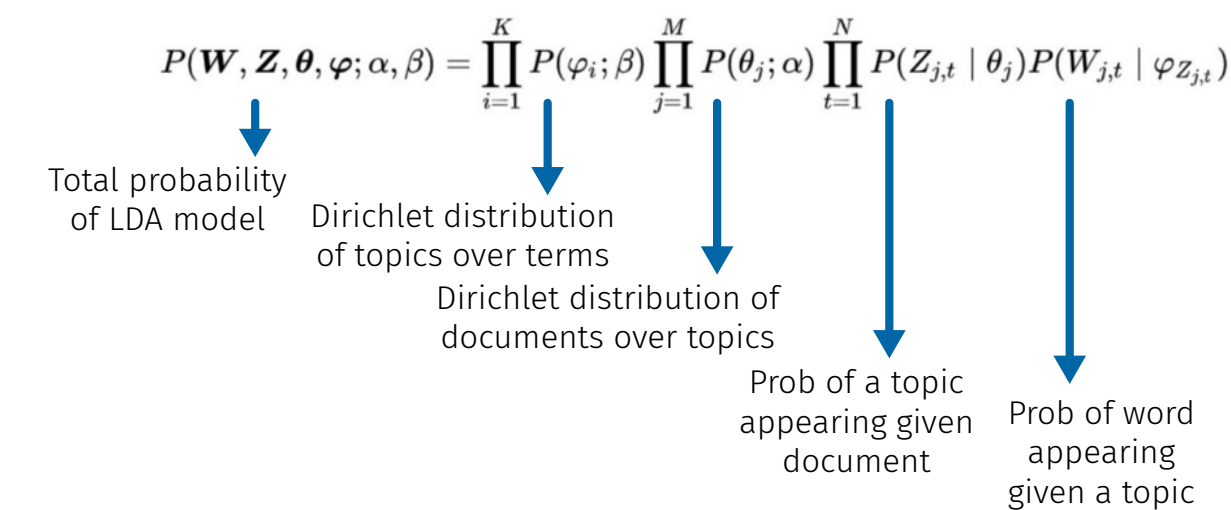
Gráfico 4.2: Clasificación de tópicos con un diccionario creciente



El gráfico 4.2 muestra que sucede con la clasificación de los tópicos una vez el número de palabras totales que constituyen el diccionario crecen. En él se puede encontrar como cada tópico se sitúa en un espacio de términos y como la combinación lineal de esos términos genera un tópico. Por ejemplo, el tópico de xenofobia cuando N = 3, se compone de forma equitativa por las publicaciones de 3 palabras que contienen solo “veneco” y “plaga”, una vez el número de palabras incrementa, la complejidad de las combinaciones lineales a su vez incrementa. El concepto de simplex resume esta idea del prior de Dirichlet para N-1 dimensiones (en este caso palabras), esto es gracias a que podemos asegurar una generalización de la noción de triángulo o tetraedro para dimensiones arbitrarias.

El modelo LDA en su forma base es una combinación de probabilidades en donde cada una sigue un grupo de distribuciones y un conjunto específico de parámetros que dependen de su jerarquía.

Gráfico 4.3: Ecuación de la probabilidad total modelo LDA



Fuente:https://towardsdatascience.com/latent-dirichlet-allocation-intuition-math-implementation-and-visualisation-63ccb616e094

De izquierda a derecha, se tiene la probabilidad total del modelo, la distribución de Dirichlet de los tópicos sobre los términos (note el beta en el paréntesis), la distribución de los documentos sobre los tópicos (note el alpha en el paréntesis), la probabilidad de que un tópico aparezca en un documento dado y la probabilidad de que una palabra aparezca dado un tópico determinado.

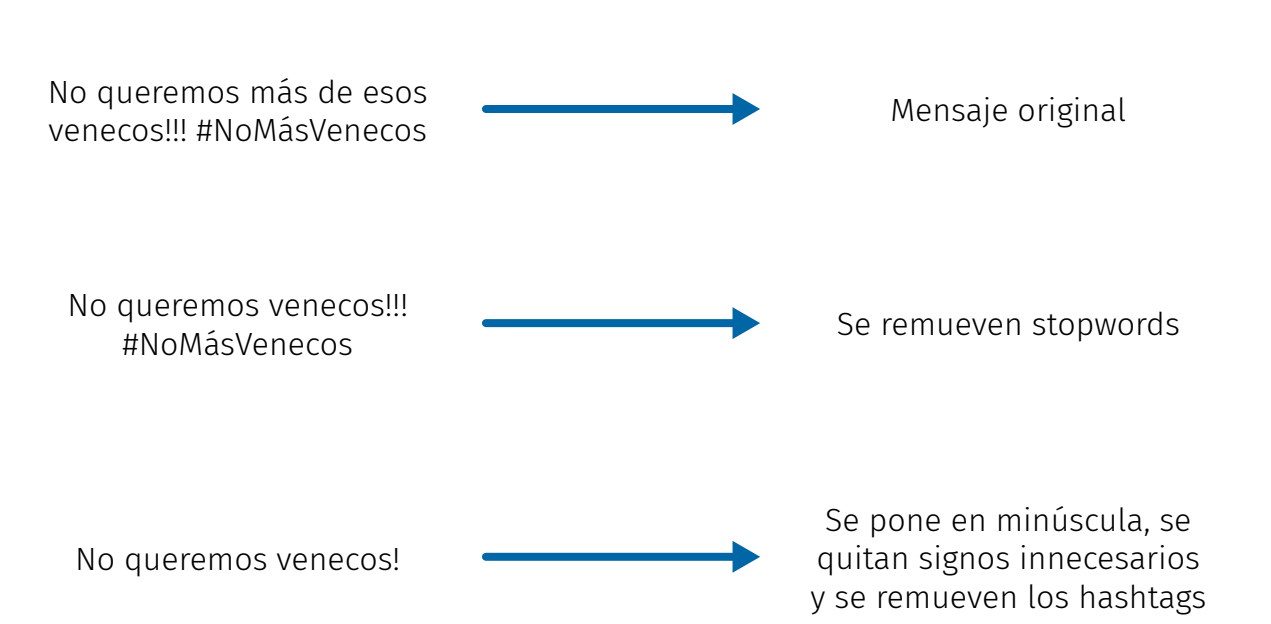
Una vez establecido el modelo se comienzan a optimizar los hiper parámetros del mismo, la idea es encontrar una combinación de beta, alpha y el número de tópicos que maximice las métricas de desempeño del modelo, para este estudio, se usarán las métricas coherence score y perplexity score que son usuales en la literatura para evaluar el desempeño de los modelos LDA y así obtener una combinación óptima orientada a una clasificación eficiente. Una vez realizada la optimización, se procede a interpretar los resultados del modelo y a establecer una marcación de tópicos por documento.

Procesamiento de Datos

Una vez obtenida la base de datos con cada una de las publicaciones de los usuarios, el objetivo es analizar el contenido de estas. Una forma de realizar el análisis es a partir de técnicas de inteligencia artificial y modelos de clasificación que permiten entender la forma en que se relacionan los mensajes. Para poder utilizar estos modelos, es necesario procesar el contenido de las publicaciones con el fin de hacerlas comparables y eliminar el ruido. Este proceso se llama vectorización de las publicaciones, y tiene como objetivo obtener una representación única y coherente de cada mensaje.

El primer paso de la vectorización busca eliminar palabras que no interesan para definir relaciones entre publicaciones. Consiste de distintas técnicas de limpieza como la eliminación de OOV (palabras fuera del vocabulario o Out-Of-Vocabulary words), stopwords (conectores lexicográficos) y las OUW (palabras demasiado usadas o Over-Used-Words). El segundo paso consiste en eliminar elementos dentro de las publicaciones que sean diferentes a contenido de léxico, ya que esto puede ensuciar la clasificación o las relaciones que encontremos entre los documentos. Siguiendo esta idea, se eliminan links (HTTP, HTTPS), hashtags, signos de puntuación innecesarios, dobles espacios, letras repetidas y letras mayúsculas. A continuación, se presentan ejemplos de cómo se verían los mensajes luego de esta limpieza y de algunas palabras que recaen en estas categorías de eliminación.

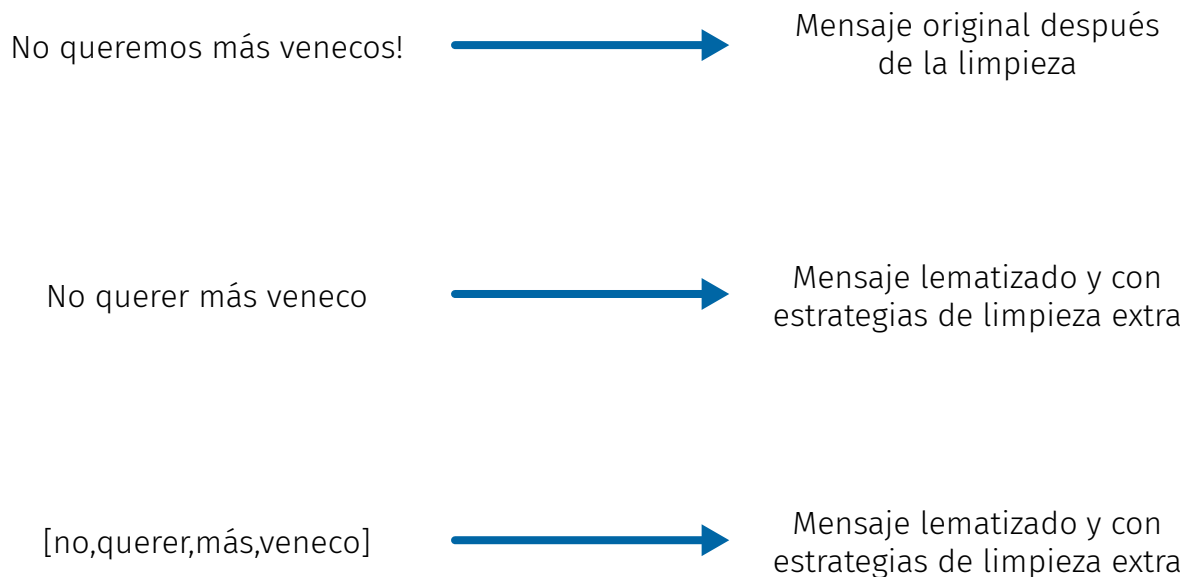
Gráfico 4.4: Ejemplo de limpieza de un mensaje de xenofobia



# REPORTE DE CONVERSACIÓN DE MIGRACIÓN Y XENOFOBIA

Una vez realizada esta primera limpieza, se propone construir una especie de diccionario con las palabras que contiene cada texto. Para esto, debemos introducir el concepto de Bag-Of-Words (BOW o bolsa de palabras), el cual sugiere que cada palabra puede ser representada con un índice que la posiciona de forma única en la bolsa (diccionario). Previo a esto, se requiere convertir palabras en forma flexionada a su lema correspondiente, este proceso lingüístico se conoce como lematización y es ampliamente usado en el procesamiento del lenguaje natural para interpretar y representar textos. De igual manera, la técnica conocida como “Stemming” permite reducir las palabras a su raíz correspondiente, la cual nos da la oportunidad de encontrar mejores relaciones entre textos, reduciendo palabras como “bibliotecario” y “biblioteca” a una raíz conocida equivalente “biblioteca”.

## Gráfico 4.5: Ejemplo de vectorización de mensajes



Una vez construido el diccionario, se procede a construir los documentos que contienen las palabras modificadas. Un documento es un conjunto de vectores que comprenden la estructura numérica de las palabras contenidas en el mismo. Un uso importante de estos documentos es que permiten evidenciar n-gramas (subsecuencias de palabras) de las publicaciones por cada uno de los autores y así obtener diferentes representaciones de un mensaje. En la siguiente sección veremos cómo vamos a usar estos documentos para clasificar los mensajes basados en sus relaciones, para este informe, lo llamaremos modelamiento de tópicos.