



REPORTE DE CONVERSACIÓN DE MIGRACIÓN Y XENOFOBIA

REPÚBLICA DOMINICANA

1 DE ABRIL 2020 A 1 DE ABRIL 2021



#XENOFOBIAZERO
ALZA TU VOZ CONTRA LA DISCRIMINACIÓN



ÍNDICE

1. RESUMEN EJECUTIVO	3
2. DATOS Y METODOLOGÍA.....	4
3. LA CONVERSACIÓN DE XENOFOBIA	4
3.1 EVENTOS MÁS RELEVANTES EN LA CONVERSACIÓN DE XENOFOBIA	4
4. CLASIFICACIÓN DE LA CONVERSACIÓN DE XENOFOBIA	9
4.1 MODELO LDA	9
4.2 RESULTADOS	10
5. CUANTIFICACIÓN DE XENOFOBIA POR NACIONALIDAD	12
6. CONCLUSIONES Y RECOMENDACIONES DE POLÍTICA	13
ANEXO 1. NOTICIAS Y XENOFOBIA	14
ANEXO 2. ESTIMACIÓN A PARTIR DE LDA	16

1. RESUMEN EJECUTIVO

El presente informe muestra un análisis de la conversación en línea de xenofobia y lenguaje de odio hacia la población migrante en República Dominicana. 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 haitiana, venezolana y nicaragüense entre el 1 de abril de 2020 y el 1 de abril de 2021. Durante este periodo se generaron un total de 14.588 mensajes de este tipo que tuvieron un alcance de 1.523.761.911 de usuarios. Esto incluye retweets y compartidos.

El mes con mayor cantidad de publicaciones con contenido xenófobo fue septiembre, producto de la noticia en que se informaba que hombres presuntamente de nacionalidad haitiana violaron a una adolescente. De hecho, la mayoría de las publicaciones de xenofobia en línea estuvieron dirigidas hacia la población migrante haitiana. Al separar las publicaciones de xenofobia por la nacionalidad a que iban dirigidas, se encuentra que el 40% estaban orientadas a población migrante haitiana, 20% a población migrante venezolana y 40% no indicaba directamente una nacionalidad.

Durante los brotes más importantes de xenofobia en el período analizado, las publicaciones con contenido de xenofobia se agruparon en tres conversaciones principales. Primero, la asociación de la migración haitiana con hechos de crimen violento. Segundo, la idea del supuesto colapso del sistema de salud causado por la llegada de migrantes al país para tener hijos. Tercero, la afirmación de que la población migrante posee privilegios en el acceso a servicios en comparación a la población local y que, en este sentido, el gobierno tiene incentivos de recibir a más migrantes para obtener recursos de la comunidad internacional.

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 toda las publicaciones de xenofobia. Los resultados indican que hubo tres temáticas principales que se diferenciaron del resto. 1. El rechazo de la llegada de mujeres haitianas a dar a luz en los hospitales de República Dominicana. La mayoría de estas publicaciones mostraban un sentimiento de rechazo e injusticia con el hecho de que la población migrante haitiana utilizará los servicios públicos nacionales y se referían a las mujeres haitianas como parturientas. 2. Conversaciones que muestran a la migración haitiana como una amenaza para la nación. Un ejemplo de esto fue una conversación que estuvo centrada en la idea de construir un muro en la frontera para proteger a República Dominicana de la amenaza haitiana. 3., El descontento con el involucramiento de la comunidad internacional en el manejo de la situación migratoria en el país. Estas publicaciones mostraban la necesidad de tomar políticas de mano dura para manejar la situación migratoria y afirmaban que la comunidad internacional era el principal obstáculo para tomar ese tipo de políticas.

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

1. **La identificación oportuna de brotes de xenofobia** para alertar a las autoridades y a la población.
2. **Proveer herramientas a la ciudadanía** para reaccionar ante el aumento del discurso de odio en línea.
3. **Tener en cuenta los prejuicios identificados** a la hora de diseñar estrategias de comunicación.
4. **Desmentir la percepción de competencia de recursos o posición de privilegio** de la población migrante en el acceso a servicios.

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 se proponen unas 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 República Dominicana 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 República Dominicana. 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 República Dominicana, las poblaciones migrantes que fueron incluidas en el análisis fueron la haitiana, venezolana y nicaragüense, 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 República Dominicana durante el período de estudio. En las siguientes secciones se procede a analizar los contenidos de estas publicaciones.

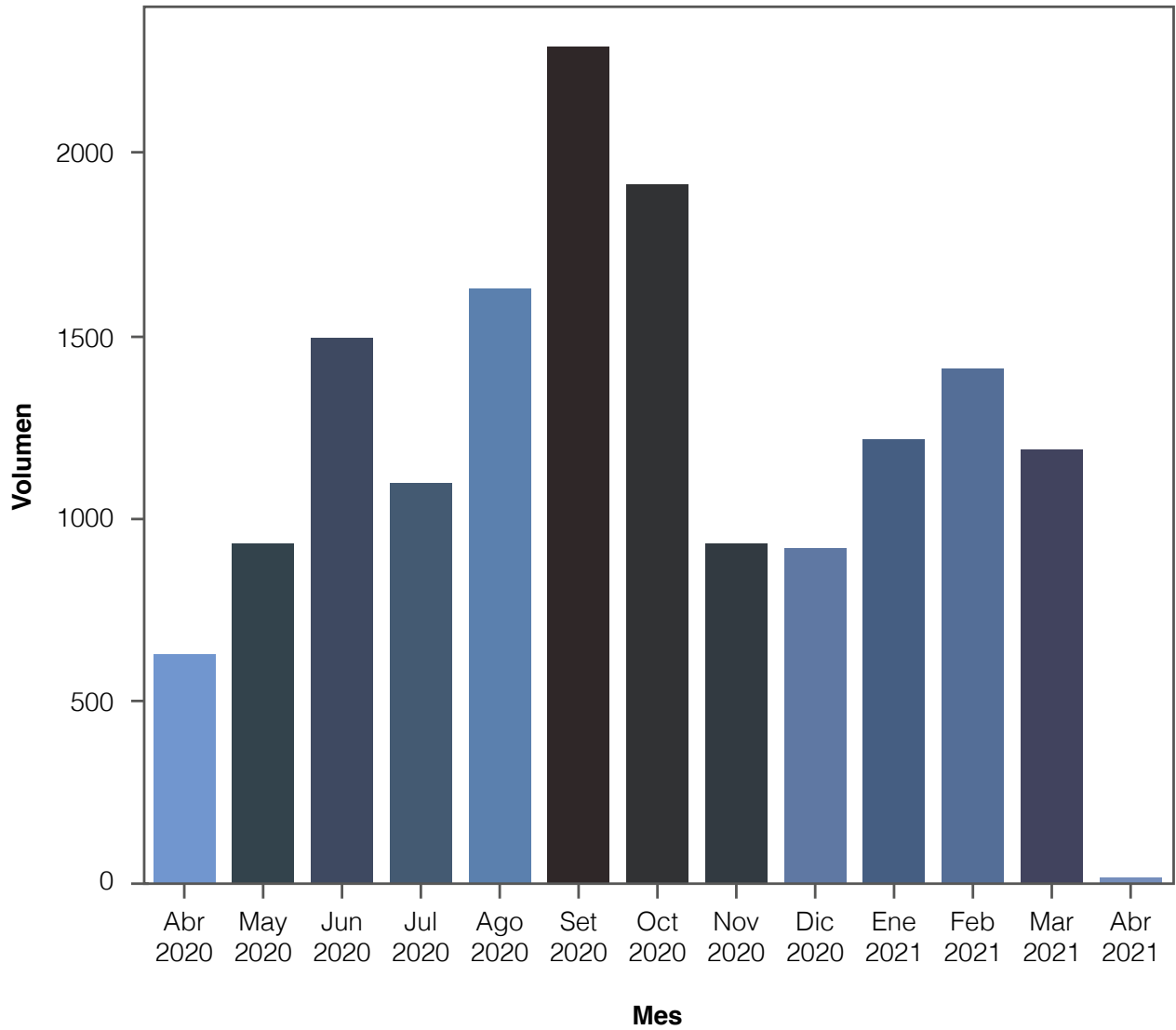
3. LA CONVERSACIÓN DE XENOFOBIA

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 14.588 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 la 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 junio, septiembre y octubre 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.

Agosto 2020

Durante el mes de agosto, la conversación de xenofobia estuvo agrupada en la idea de una supuesta invasión por parte de los migrantes haitianos. Las publicaciones encontradas en esta conversación critican que haya personas defendiendo la “invasión haitiana” y, como consecuencia, la destrucción del país. En este sentido, se les relaciona al incremento de los hechos delictivos y se hace un llamado a que vuelvan a su país. Esto se puede ver ejemplificado en los ejemplos del gráfico 3.3.

Algunos de estos mensajes hacían un llamado a las autoridades a tomar acciones en contra de la población migrante haitiana. Más aún, mostraban un descontento con las autoridades estatales por percibir poca

gestión en el manejo de la migración. Como lo muestran algunos ejemplos del gráfico 3.3 las ‘acciones’ para controlar la migración hacen especial énfasis en la condición de regularidad de los migrantes.

Otra temática que aparece en este mes es el de la ocupación de hospitales por parte de mujeres haitianas embarazadas. En este sentido, se hace alusión a que si estas están indocumentadas, deberían “ir a parir” Haití.

Gráfico 3.3 Ejemplo Mensajes Agosto de 2020



Septiembre 2020

Septiembre de 2020 fue el mes en el que se reportó el máximo volumen en la conversación de xenofobia, con un total de 1.764 mensajes. Estas publicaciones estuvieron motivadas por una noticia en la que se informaba que hombres presuntamente de nacionalidad haitiana violaron a una adolescente. La gráfica 3.3 muestra los mensajes con mayor difusión durante el mes y que resumen las principales temáticas de las que se habló, siendo la violencia y el sentimiento de invasión y deterioro los principales asuntos que se trataron.

Gráfico 3.4 Ejemplo Mensajes de Xenofobia Septiembre de 2020



Meltwater, 2021.

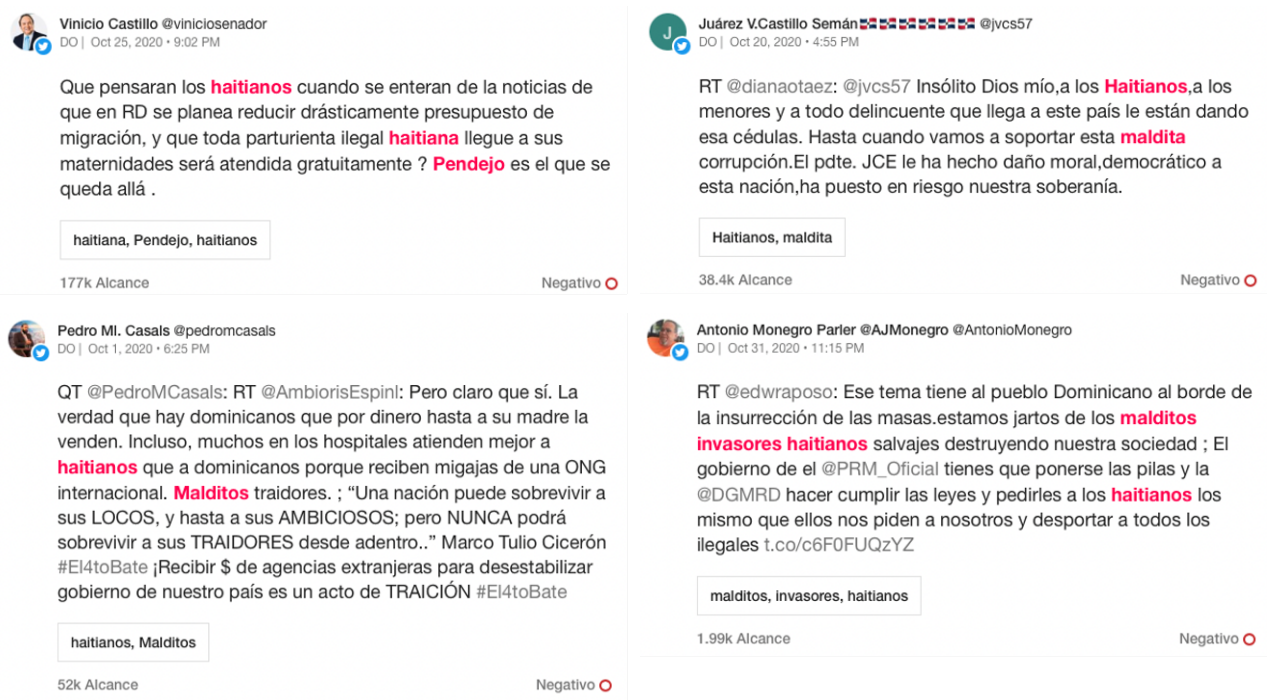
Sobre los temas abordados durante la conversación en este periodo, los ejemplos expuestos en el primer, segundo y quinto mensaje conectan a la población migrante con actos violentos y criminales. A diferencia de lo visto en el mes de junio, la cantidad de mensajes que se refieren a las personas migrantes haitianas como ladronas, asesinas y violadoras, aumentó. Adicionalmente, se observa que estas publicaciones ahora incluyen comentarios respecto a organizaciones internacionales, indicando que estas conocen la problemática que generan la población migrante y que es por estas que no es posible expulsarles del país.

En este mes se observa de nuevo que entre los temas de conversación se incluyen ofensas hacia la población venezolana. Sin embargo, en este caso se le suma a las ofensas vistas en el mes de junio, el relacionamiento de la migración venezolana a problemas sociales (pobreza, delincuencia y terrorismo) en el país de destino, como se observa en el tercer mensaje. Es de resaltar que los mensajes de xenofobia contra población venezolana en este mes no superan los dirigidos hacia personas migrantes haitianas.

Octubre 2020

La conversación sobre xenofobia en este mes tuvo una connotación diferente respecto a lo observado en meses anteriores. Como se observa en el primer mensaje del gráfico 3.4, en octubre de 2020 se dio una noticia según la cual se reduce el presupuesto destinado a migración y, además, se anunció la atención gratuita el parto a las mujeres migrantes. Esta noticia impulsó la conversación en esta dirección y generó una serie de mensajes en contra de esta política, aduciendo que esto sería un apoyo por parte del gobierno de República Dominicana para la invasión haitiana.

Gráfico 3.5 Ejemplo Mensajes de Xenofobia Octubre de 2020



Meltwater, 2021.

Se debe mencionar también que la conversación sigue incluyendo la idea de que las organizaciones internacionales y ONG son actores que facilitan la llegada de la población haitiana al país, haciendo que el sistema de salud, por ejemplo, prefiera atender personas haitinas por un pago de comisiones. El segundo mensaje es un ejemplo de esto, puesto que explícitamente se habla de que el personal médico en República Dominicana prefiere prestar servicios a migrantes con el fin de recibir “migajas” por parte de organizaciones internacionales.

Xenofobia y COVID-19

Un posible foco de discriminación hacia la población migrante es la asociación de esta población con temáticas relacionadas a la crisis sanitaria originada por el COVID-19. Existen varias razones por las que la crisis sanitaria puede estar asociada 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 4% 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 42 mil publicaciones de integración. Esto representa casi tres veces más 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

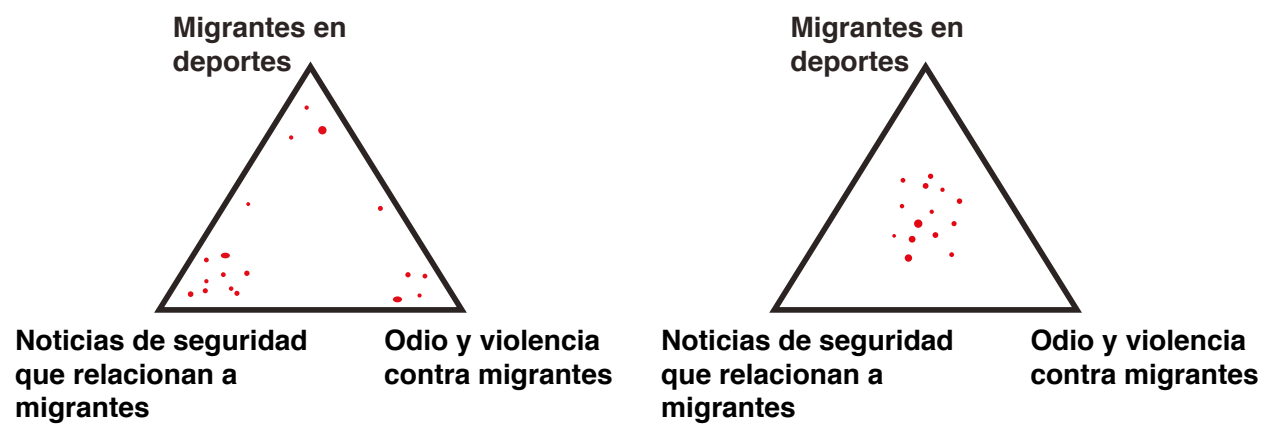
En la sección anterior se discutieron los momentos más importantes de la conversación de xenofobia, así como los impulsores de estos. Sin embargo, otra aproximación para entender el contenido dentro de las publicaciones de xenofobia es estudiando los tópicos o temas utilizados dentro de esta conversación.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

4.1 MODELO LDA

Para poder estudiar las temáticas más frecuentes dentro de la conversación de xenofobia en República Dominicana, se hace uso de 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 temáticas de salud y migración. El segundo tópico tuvo que ver con la migración haitiana como amenaza a la seguridad nacional. El tercer tópico tuvo que ver con el involucramiento de la comunidad internacional en la forma en que se atiende el fenómeno migratorio. 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 muestra la forma en que se distribuyen los tópicos dentro de esta conversación.

El tópico que más agrupó publicaciones de xenofobia (Salud y Migración) tuvo que ver con la llegada de mujeres haitianas a dar a luz en los hospitales de República Dominicana. La mayoría de estas publicaciones mostraban un sentimiento de rechazo e injusticia con el hecho de que la población migrante haitiana utilizara los servicios públicos nacionales y se referían a las mujeres haitianas como *parturientas*. Un hecho interesante es que gran parte de esta conversación surgió a propósito de las exigencias por legalizar el aborto en República Dominicana. En el gráfico 4.7 del anexo se pueden observar los principales términos que contiene el tópico en cuestión.

El segundo tópico (Percepción de Amenaza) que más agrupó publicaciones consiste de distintas conversaciones que muestran a la migración haitiana como una amenaza para la nación. Un ejemplo de esto fue una conversación que estuvo centrada en la idea de construir un muro en la frontera para proteger a República Dominicana de la *amenaza haitiana*. Junto con esto, este tópico mostró un rechazo hacia políticos

que tuvieran una posición de integración de personas migrantes haitianas, denunciando el supuesto peligro que representaba esta población para la seguridad de la comunidad. En el gráfico 4.8 se pueden observar los principales términos que contiene el tópico en cuestión.

Por último, otro tópico interesante mostró un descontento con el involucramiento de la comunidad internacional en el manejo de la situación migratoria en el país (tópico 3). Este tipo de mensajes tuvo dos características principales. Primero, la exigencia de políticas de mano dura en el manejo de la situación migratoria, como el cierre de la frontera. Segundo, se afirmaba que la comunidad internacional era el principal obstáculo para tomar ese tipo de políticas. En el gráfico 4.9 se pueden observar los principales términos que contiene el tópico en cuestión.

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 recurrentemente 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 “malditos invasores”, “malditos haitianos” e “ilegales haitianos”. 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 haitiana y se desarrollan en torno al hecho de que arriban a República Dominicana como invasores ilegales que destruyen el país.

Es importante destacar la relación que los autores de estos mensajes hacen entre la población migrante y los problemas del país. En particular, la frase “verdaderos culpables” enmarca estas publicaciones, ampliamente utilizadas y que generan la percepción de que son las personas haitianas quienes tienen la culpa de los problemas sociales y de seguridad que puede estar experimentando el país. Lo anterior está ligado también a la frase “malas costumbres”, según la cual es la forma de actuar de la comunidad migrante y sus costumbres las que están teniendo un impacto negativo. Este concepto engloba mensajes que asocian de

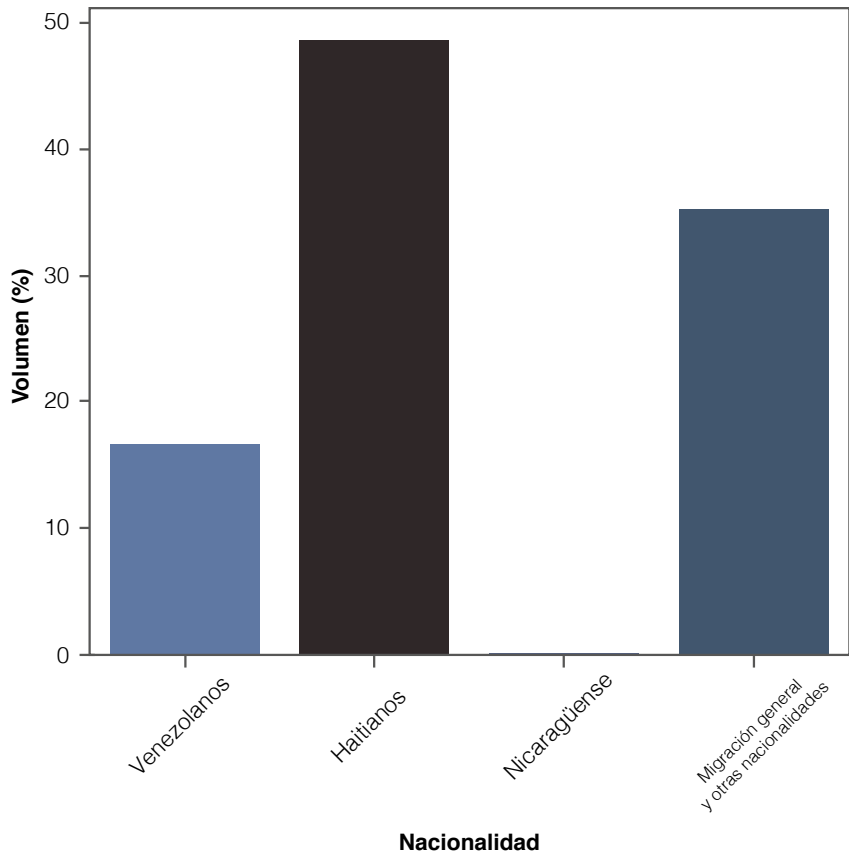
manera despectiva a migrantes con pobreza, suciedad o hacinamiento. Por otro lado, la frase “deportación de haitianos” expone la idea de algunos autores de que las personas migrantes haitianas deben ser expulsadas del país al quitarles oportunidades y ser focos de criminalidad.

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 República Dominicana existen migrantes de distintas nacionalidades, siendo las poblaciones migrantes haitiana, nicaragüense y venezolana 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 a que van dirigidas, es decir, están orientadas a la población migrante en general, o van dirigidas a nacionalidades que no son relevantes dentro del estudio.

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 haitiana, venezolana o nicaragüense. Salta a la vista que los mensajes xenófobos hacia la población nicaragüense corresponden a un 0.05%, siendo la nacionalidad con la menor participación entre las incluidas en el análisis. Por el contrario, se encuentra que el 48.5% de los mensajes de xenofobia van dirigidos a la población migrante haitiana, siendo esta la nacionalidad que recibió una mayor cantidad de mensajes de xenofobia entre abril de 2020 y 2021. Adicionalmente, el 16.5% de los mensajes de xenofobia fueron en contra de población venezolana y cerca del 35% 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 personas migrantes haitianas, 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.

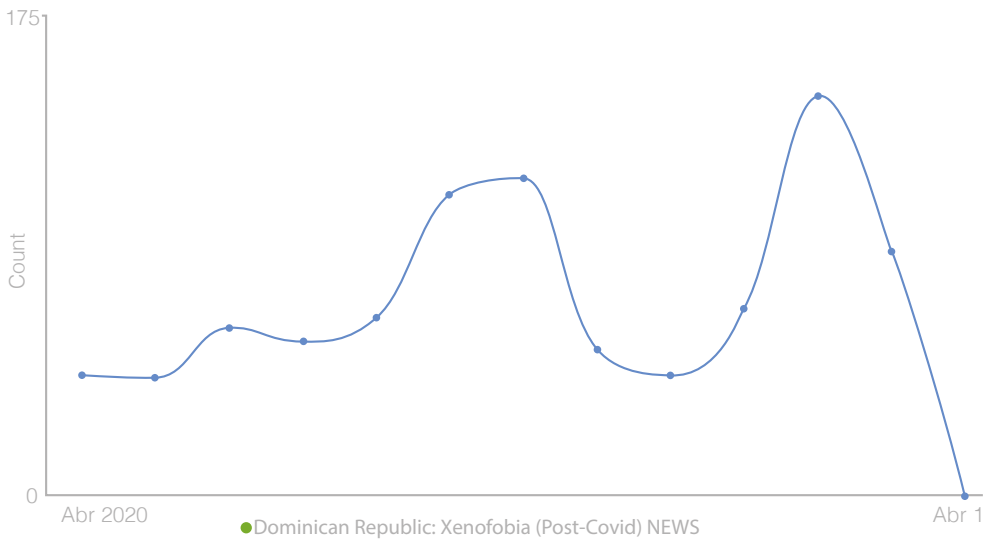
Recomendaciones

1. En coyunturas de brotes de xenofobia, como el ocurrido en septiembre, aumenta la probabilidad de que se vulneren los derechos de la población migrante. Es importante que se identifiquen estos brotes oportunamente para poder alertar a las autoridades y a la población en general sobre los aumentos en el discurso de odio y sus riesgos.
2. 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.
3. En caso de diseñar estrategias de comunicación para disminuir los niveles de rechazo hacia la población migrante, es importante que se tomen en cuenta los prejuicios identificados.
4. Una parte importante de las publicaciones de xenofobia tienen que ver con la percepción de competencia por recursos públicos. A esta idea se suma la sensación de que la población migrante posee más privilegios que la comunidad de acogida. Es primordial que estas preocupaciones sean atendidas para disminuir la percepción de injusticia en la atención del fenómeno migratorio.

ANEXO 1. NOTICIAS Y XENOFOBIA

Durante el periodo estudiado, pudimos evidenciar que se han generado 19.272 publicaciones sobre migrantes, con un promedio de 53 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, 892 publicaciones, lo que implica un promedio de 2 publicaciones diarias.

Gráfico de volumen por mes, publicaciones xenófobas en medios.



Meltwater, 2021.

El peak que se genera en el mes de febrero responde a la “necesidad” de cobrar a mujeres haitianas por dar a luz en hospitales dominicanos. Otros ejemplos de noticias con rasgos xenófobos son:

Hospitales del noroeste saturados de parturientas haitianas y con poco personal para atenderles

El número de parturientas haitianas se ha triplicado en diez años, llegando a 30.322 nacimientos en 2020 natural entre ambos países, nueve de cada diez parturientas son haitianas, la mayor proporción en toda República Dominicana. HOSPITALES



Advierten que nuevo canciller impulsaría agenda que amenaza la dominicanidad

Luis Abinader anunció el pasado viernes que Álvarez será designado como nuevo canciller de RD un gobierno se pretenda nacionalizar a la gran cantidad de hijos de haitianos ilegales que hay en la República Dominicana, irrespetando la



“Debemos evitar que sigan viniendo a parir a nuestros hospitales extranjeras ilegales”, dice presidente de sociedad médica

Santo Domingo.- Para el doctor, Jorge Vargas, presidente de la Sociedad Dominicana de Obstetricia y Ginecología, la alta cantidad de partos ... , considera que lo primordial es, “evitar que sigan viniendo a parir a nuestros hospitales extranjeras ilegales”, enfatizó. Añadió que el



Prohibición del aborto terapéutico y colocación de muro en frontera contribuyen a elevar costos

Prohibición del aborto terapéutico y colocación de muro en frontera contribuyen a elevar costos Por Danilo Cruz Pichardo “El derecho a la ... que los militares dominicanos cobran mil pesos por la entrada de cada haitiano. A partir del levantamiento del muro entonces cobrarían cinco



Cónsul dominicano en Juana Méndez ve atinado construcción verja fronteriza

JUANA MÉNDEZ.- El cónsul dominicano en esta ciudad fronteriza, José Valenzuela, consideró “como muy atinada” la construcción de una verja ... detener el proceso migratorio de nacionales haitianos hacia el territorio dominicano. “Decir que los haitianos harán túneles para venir al

La desbordante migración masiva haitiana

Por: PORFIRIO ROMERO Por más de 50 años la República Dominicana ha tenido que soportar la embestida de una migración ilegal haitiana ... haitianos es como volver a permitirle a Boyer invadimos como lo hicieron en 1822. Con esta migración silenciosa los vecinos haitianos

ANEXO 2. ESTIMACIÓN A PARTIR DE LDA

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



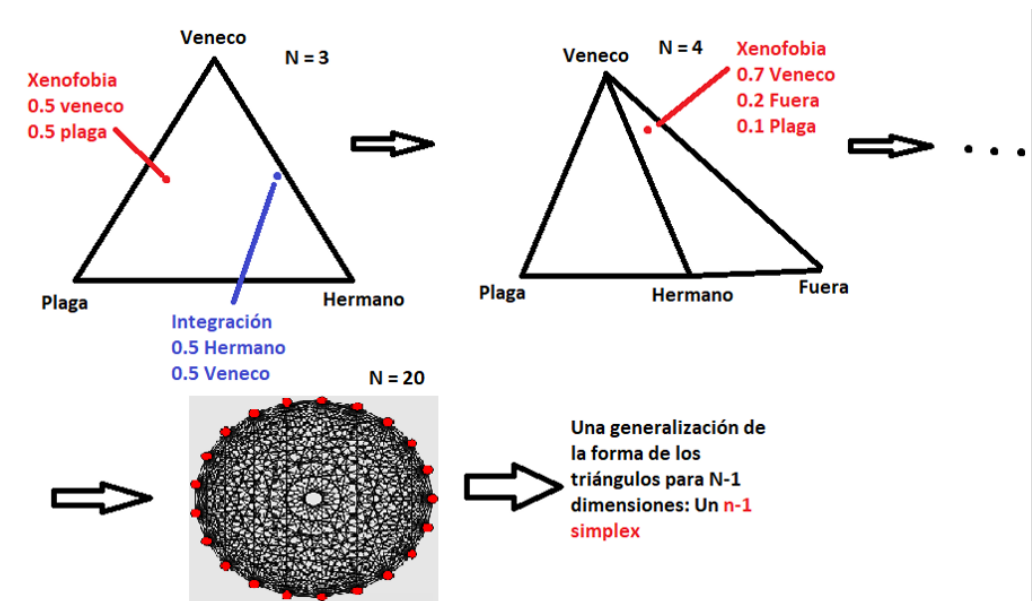
Fuente: Creación propia usando pyLDAvis. Los números representan el orden de los tópicos en términos de número de publicaciones. Es decir, el tópico 1 es el que agrupó más publicaciones y el tópico 7 el que menos. La interpretación del contenido de cada tópico es subjetiva al investigador. Esto quiere 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; β); 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).

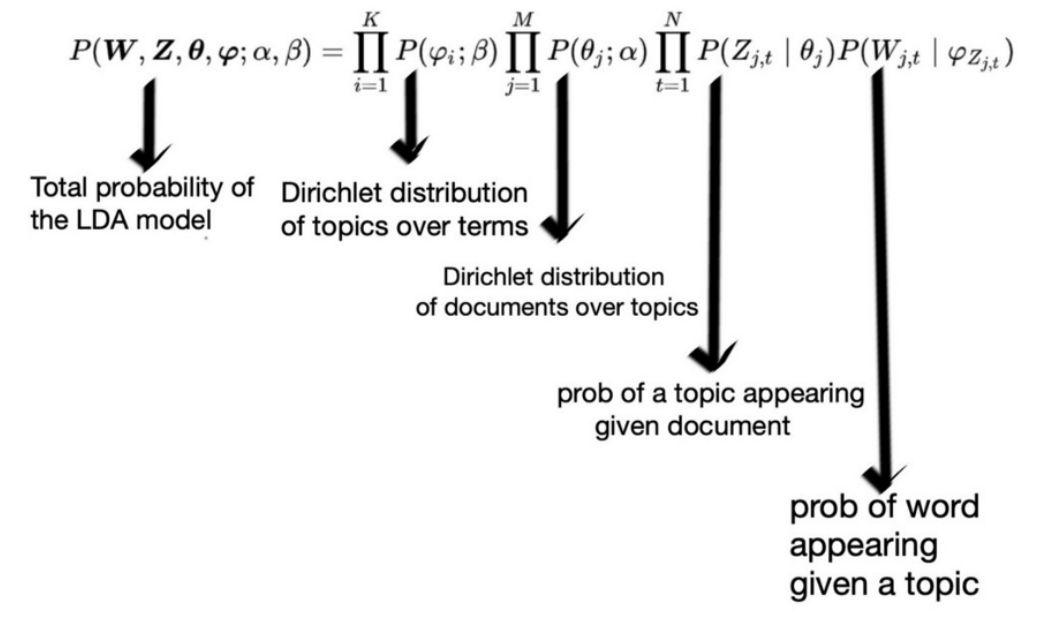
Gráfico 4.2: Clasificación de tópicos con un diccionario creciente



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.

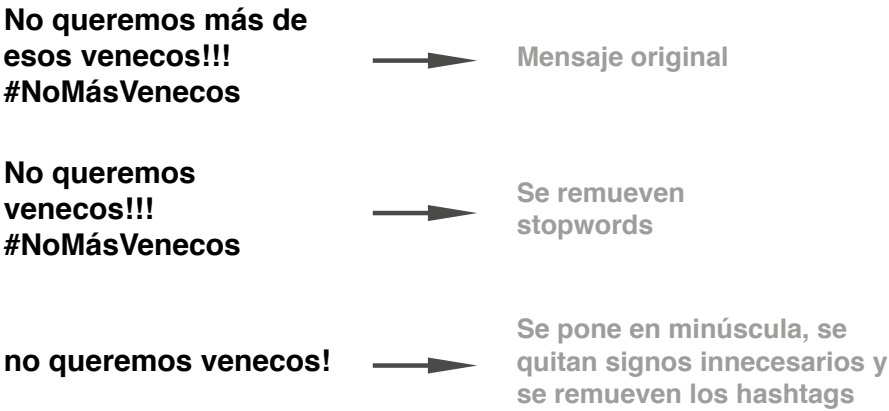
Procesamiento de Datos

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

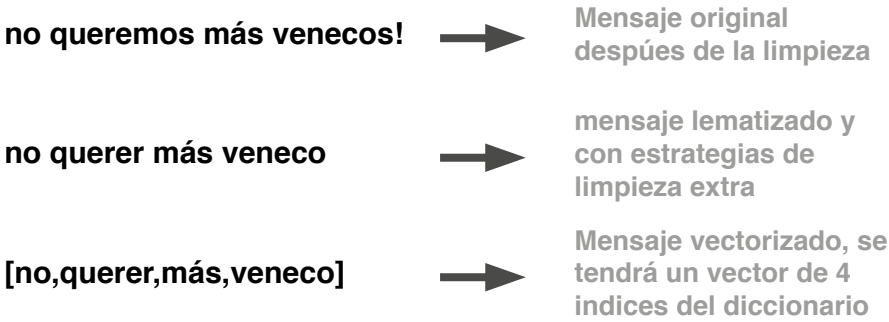
continuación, se presentan ejemplos de cómo se verían los mensajes luego de esta limpieza y de algunas palabras que recaen en estas categorías de eliminación.

Gráfico 4.4: Ejemplo de limpieza de un mensaje de xenofobia



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.