**Seguimiento del índice nacional de felicidad mediante Twitter para bilingües**

Hoy en día, las redes sociales se han convertido en uno de los métodos de comunicación más importantes que brindan información en tiempo real y rica fuente de información, incluidos sentimientos. Comprender el sentimiento de la población es un objetivo clave para las organizaciones y gobiernos. En los últimos años, se han realizado bastantes investigaciones sobre el análisis de sentimientos en las redes sociales. Sin embargo, todo el trabajo en el **estado del arte** se centra en un **subconjunto específico predefinido de tweets**. Ejemplos: análisis de sentimiento a través de palabras clave, búsqueda de tweets para marcas relevantes, productos, servicios, eventos, etc. La **supervisión** **del sentimiento general a nivel nacional** nivel a través de todo el flujo de redes sociales **no se realiza** debido a los **desafíos** de filtrar información irrelevante para el sentimiento, diversidad de uso de vocabulario en tweets generales sobre diferentes temas, lo que provoca una baja precisión y la necesidad de modelos bilingüe o multilingüe. Este artículo propone un **sistema para monitorear el sentimiento de la población general desde una red social (Twitter), a través de filtros integrales de múltiples niveles, y nuestra propuesta de asignación de Dirichlet latente mejorada (LDA**), método para la clasificación de sentimientos. Los experimentos muestran que nuestra propuesta el LDA mejorado para el análisis de sentimiento produce los mejores resultados y también valida nuestro sistema propuesto para el monitoreo del sentimiento nacional en Abu Dhabi usando Twitter.

**INTRODUCCION**

Las **redes sociales** ahora son reconocidas como una fuente importante de información casi instantánea sobre eventos, noticias, ideas y más importante aún, opiniones y emociones. Representa un método efectivo e inmediato de comunicación entre individuos y comunidades. La naturaleza informal de las redes sociales les ha permitido convertirse en un medio rico para el **intercambio directo de opiniones y la expresión de sentimientos**, por lo que **son un medio eficaz para realizar análisis de sentimientos para monitorear los índices de felicidad casi en tiempo real.**

Comprender el sentimiento general de una determinada población es un objetivo importante de las organizaciones y los gobiernos a nivel mundial. Las redes sociales ofrecen información de primera mano sobre los pensamientos, sentimientos y preocupaciones de la población. Por lo tanto, monitorear las redes sociales permite a las organizaciones y al gobierno medir el grado de felicidad de la población sin implicar costos significativos. Se han llevado a cabo algunas investigaciones sobre el análisis de sentimientos en las redes sociales, pero la mayoría de ellas se centran en el análisis de sentimientos para un servicio o producto/marca en particular y principalmente para idiomas con gramática simple, como el inglés. Sin embargo, **el análisis de sentimiento general a partir de mensajes cortos (tweets) para varios idiomas** ha sido un desafío debido a las siguientes razones:

1. **Información limitada en mensajes muy cortos:** el uso de texto breve con expresiones informales y variaciones de palabras causadas por errores ortográficos, jerga de tweets y abreviaturas.

2. **Gran cantidad de datos ruidosos:** nuestro objetivo es obtener tantos tweets como podamos acceder dentro de una región/país y, por lo tanto, necesitamos filtrar los tweets que sean irrelevantes al sentimiento de la población. Esto es un gran desafío, considerando que muchas organizaciones y entidades utilizan Twitter para sus actividades de marketing, eventos/noticias/anuncios, para agregar anuncios, spam y muchos otros tweets irrelevantes.

3. **El sentimiento es subjetivo**: diferentes personas pueden tener diferentes comprensiones e interpretaciones del mismo tweet en el contexto de la expresión del sentimiento. Por lo tanto, el análisis de sentimientos de los tweets es incluso más desafiante que otras aplicaciones de tweets. Ejemplo: Clasificación de temas a partir de tweets.

4. **El análisis del sentimiento general es más desafiante**: el análisis del sentimiento general es más difícil que el de servicios/productos o eventos particulares, ya que necesita cubrir una enorme cantidad de vocabulario para representar el sentimiento en todas las áreas.

5. **Desafío del análisis de sentimientos de los tweets en el mundo árabe** - Muchos tweets en el mundo árabe están escritos en árabe o en una mezcla de árabe y otro idioma como el inglés. Las técnicas de minería de textos están maduras para trabajar con lenguajes gramaticales relativamente simples y de uso común, como el inglés. Las técnicas y metodologías para la extracción de textos en árabe (especialmente para mensajes cortos con expresiones informales) aún son inmaduras debido a la complejidad inherente del idioma árabe en términos tanto de estructura como de morfología. Por lo tanto, el método y el sistema propuestos deben ser independientes del lenguaje y ser capaces de manejar diferentes lenguajes, incluidos los complejos, sin dejar de lograr una precisión aceptable.

Sin embargo, debido al amplio uso de las redes sociales en los países árabes, existe una gran demanda de un sistema preciso para monitorear el índice de felicidad del sentimiento general en las redes sociales. Un sistema así debe ser al menos bilingüe. Dados los desafíos y la alta demanda mencionados anteriormente, este artículo propone un sistema general para monitorear el índice de felicidad de la población utilizando el análisis de sentimientos de una red social (Twitter). Los aportes de este trabajo se pueden destacar como:

1. **Obtenga un filtro preciso para identificar tweets relevantes para el sentimiento.** Proponemos y utilizamos filtros integrales multinivel. Un filtro preciso es la premisa de un análisis de sentimiento preciso. El filtrado incluye (1) **Bayes Nets** que filtran tweets irrelevantes para el sentimiento por características. Ejemplo: **longitud de los tweets, ya sea que incluyan emoji, menciones @, números de móvil, etc**.; (2) **clasificador LDA mejorado** (aplicado al texto) para **clasificar los tweets como tweets relevantes para el sentimiento o tweets sin sentimiento**.

2. **Lograr una mayor precisión general para el análisis de sentimiento general**. Aplicamos nuestra LDA mejorada propuesta para la clasificación de tweets para lograr una precisión similar o mejor para el análisis de sentimientos que el etiquetado de seres humanos. Se ha demostrado que el método propuesto obtiene mejores resultados en conjuntos de datos de referencia que los métodos más modernos. El método propuesto es independiente del idioma y puede manejar tweets tanto en inglés como en árabe.

3. **Proporcione una selección completa de subsecuencias de tweets que pueda filtrar y luego mostrar el sentimiento por tema o tema de enfoque**. Ejemplos: Tema(s) particular(es), período de tiempo, ubicación(es) y evento(s) específico(s).

4. **Desarrollar un visualizador de tweets integral que sea lo suficientemente flexible como para mostrar el sentimiento del usuari**o, ya sea en general o sobre un tema de interés particular, como series de tiempo, distribuciones de geolocalización y tablas de tweets.

Este artículo está organizado de la siguiente manera: discutimos el trabajo relevante para el análisis de sentimientos de las redes sociales en la Sección. 2 y presentar nuestro sistema propuesto en la Sección. 3. Para demostrar la validez, aplicamos nuestro método propuesto para comparar conjuntos de datos y los resultados y las comparaciones se analizan en la Sección. 4. La Sección 5 presenta los resultados de nuestro método propuesto cuando se utiliza para el análisis del sentimiento de la población general en Abu Dhabi, Emiratos Árabes Unidos.

**Trabajo relacionado**

Se han llevado a cabo diversas investigaciones para investigar cómo **mejorar la precisión del análisis de sentimientos de las redes sociales**. Algunas investigaciones se centraron en la **importancia del preprocesamiento para el análisis de sentimientos en las redes sociales** Singh y Kumari (2016) investigaron la importancia del preprocesamiento y afirman que **el método propuesto es robusto al tamaño de los conjuntos de datos, pero también mejora la precisión**. Algunas investigaciones se realizaron utilizando **ensemble methods**, **eliminando el desequilibrio del conjunto de datos y resolviendo el problema de la falta de datos de entrenamiento**. con el objetivo de mejorar las precisiones generales. Hassan y Abbasi (2013) propusieron un **boostrap emsemble framework para el análisis de sentimiento en Twitter para resolver el problema del desequilibrio de datos y lograron una mayor precisión** (entre 27 y 80% en diferentes conjuntos de datos). Kanakaraj y Guddeti (2015) **analizaron cómo diferentes ensemble methods** **ayudan a mejorar la precisión del análisis de sentimiento de Twitter**. Zhao (2016) utilizó el **emsemble learning** (que representa palabras como vectores) **y combinó la polaridad semántica y previa para impulsar Análisis de sentimiento en Twitter.** Krouska et al. (2017) presentaron un **sistema para el análisis de sentimientos para el lenguaje figurativo (emoción) basado tanto en un enfoque de contenido (texto) como en un enfoque de patrones de emociones**, que demostró que la combinación del análisis de contexto y el análisis de patrones de emociones mejora la precisión general del análisis de sentimientos de las redes sociales. Krouska et al. (2016), Troussas et al. (2016) y Nguyen y Jung (2017) evaluaron cómo el preprocesamiento y el conjunto de datos (Troussas et al. 2016) ayudan a mejorar la precisión del análisis de sentimientos para diferentes métodos de aprendizaje automático, y mostraron mejoras en el conjunto. la precisión de manera significativa. También compararon el rendimiento del uso de diferentes algoritmos para el análisis de sentimientos en conjuntos de datos de referencia (Nguyen y Jung, 2017). Abdelwahab y Bahgat (Abdelwahab et al. 2015) **investigaron la relación entre el tamaño del conjunto de entrenamiento y la precisión**. Sahu y Rout (Sahu et al. 2015) utilizaron un método basado en léxico para el análisis de sentimientos y afirmaron que el análisis de sentimientos en diferentes dominios necesita léxicos diferentes, lo que indica la inflexibilidad de los métodos basados en léxico. Kontopoulos et al. (2013) propusieron un análisis de sentimiento para tweets basado en ontologías. Furini y Montangero (2016) propusieron e investigaron un enfoque de gamificación para el análisis de sentimientos a partir de tweets. Bravo-Márquez et al. (2016) investigaron la transferencia entre palabras en el léxico de sentimientos y los sentimientos de los tweets.

Desde el estado del arte, las técnicas de aprendizaje automático también se aplican al análisis de sentimientos, incluido **naive Bayes**, **máxima entropía** (MAXENT) y **máquina de vectores de soporte** (SVM), así como **redes neuronales artificiales** (ANN). Nakov et al. (2016a, b) analizaron diferentes algoritmos y rendimientos para el análisis de sentimientos a partir de mensajes cortos basados en conjuntos de datos de referencia, SemEval. Su investigación muestra que la mayoría de los sistemas fueron supervisados y utilizaron una variedad de funciones artesanales y codificaciones específicas de Twitter. Chiassi et al. (2013) y Zimbra et al. (2016) propusieron un sistema híbrido que utiliza **n-gram y una red neuronal artificial dinámica** para el análisis del sentimiento de marca basado en un conjunto de léxico definido manualmente y afirmaron que su precisión era superior al 95%. Ajay y Sudhir (AIIT 2016) compararon el rendimiento de seis métodos populares de aprendizaje automático para el análisis de sentimientos a partir de tweets. Shyamasundar y Jhansi (2016) se centraron en la **extracción de características utilizando TF-IDF para mejorar la precisión del análisis de sentimientos a partir de técnicas de aprendizaje automático** (naïve Bayes y SVM) para conjuntos de datos de referencia. Siddiqua et al. (2016) propusieron un **clasificador basado en reglas basado en los léxicos de emoji y sentimiento** para el análisis de sentimiento de Twitter y lo compararon con dos sistemas conocidos (SentiStrength y Semantria) y afirmaron una mayor precisión en tres de cuatro conjuntos de datos de referencia.

Algunos de los métodos propuestos se aplican a varios puntos de referencia o a sus propios conjuntos de datos recopilados y lograron una precisión de entre el 60 y el 85%, según los algoritmos y los conjuntos de datos. Ren et al. (2016) mejoraron el método **word embedding** de palabras para el análisis de sentimiento en Twitter y lograron una precisión del 78,57 % sin conocimiento experto y del 81,02 % con conocimiento experto para conjuntos de datos de referencia. Wu et al (2016) propusieron un **método para extraer conocimiento sobre sentimientos a partir de datos masivos sin etiquetar** (tweets) y lograron una precisión del 85% al 88% al aplicarlo a diferentes conjuntos de datos. Saifa et al. (2016) introdujeron un **enfoque basado en léxico que logró mejores resultados para dos conjuntos de datos, pero peores resultados para el tercero**. Pandey et al. (2017) propusieron una novedosa búsqueda cuco basada en k-media para el análisis de sentimiento de tweets que mostró una precisión del 67% al 84% para diferentes conjuntos de datos de referencia. Zhao y Cao (2015) propusieron un método para utilizar las estadísticas de coocurrencia y las relaciones semánticas contextuales como características para el análisis de sentimientos y produjeron una precisión promedio del 82% para diferentes conjuntos de datos de referencia. Christos y Maria probaron diferentes algoritmos para servicios de análisis de sentimientos.

En este contexto, se han evaluado cinco clasificadores basados en el aprendizaje conocidos (**naïve Bayes, máquina de vectores de soporte, k-vecino más cercano, regresión logística y C4.5)** y un enfoque basado en **léxico** (SentiStrength) basado en matrices de confusión, utilizando tres conjuntos de datos diferentes (**OMD, HCR y STS-Gold**) y dos modelos de prueba (**división porcentual y validación cruzada**). Los resultados demostraron la superioridad de Bayes ingenuo y de la máquina de vectores de soporte, independientemente de los conjuntos de datos y los métodos de prueba.

Algunos métodos se han aplicado a problemas particulares de la vida real. Ejemplos: sentimiento por una marca particular para servicios al cliente, eventos o elecciones políticas. Se han realizado muchas investigaciones para el análisis de sentimientos a partir de tweets de marcas para servicios al cliente (Chiassi et al. 2013; Zimbra et al. 2016; Qaisi y Aljarah 2016). Philander y Zhong (2016) utilizaron un método basado en diccionarios para el análisis de sentimientos de datos de microblogging en redes sociales para operadores hoteleros y utilizaron el casino resort integrado de Las Vegas como caso de uso, que afirmaron que brindaba una confiabilidad razonable para las revisiones, clasificación. Qaisi y Aljarah (2016) utilizaron análisis de sentimiento de tweets para dos proveedores diferentes de servicios en la nube. Gupta y Kohli (2016) utilizaron el análisis de sentimiento de Twitter en el ámbito de la atención sanitaria. Yang y Wang (2014) propusieron un método para analizar los tweets en tiempo real de los fanáticos del fútbol estadounidense durante los partidos de la Copa Mundial de la FIFA 2014 y analizaron los cambios dinámicos de las emociones de los fanáticos a lo largo de los juegos. Schumaker et al. (2016) utilizaron análisis de **sentimiento de las redes sociales para predecir los resultados de los partidos de fútbol** y obtuvieron una precisión del 67%. Peng et al. (2016) utilizaron **análisis de sentimiento en Twitter para incidentes de drogas utilizando paquetes Weka** y la precisión del análisis de sentimiento fue de poco más del 62% al 65%. Ramteke et al. (2016) utilizaron el **análisis del sentimiento de Twitter para predecir los resultados electorales**. Otras aplicaciones incluyeron cómo usar las redes sociales para ayudar a resolver otros problemas. Ejemplo: **Análisis económico y predicción financiera** utilizando las redes sociales como información complementaria. Smailovic et al. (2014) propusieron un **sistema para utilizar las redes sociales para ayudar a predecir los mercados financieros para más comercio** en línea rentable. Porshnev y cols. (2013) utilizaron diferentes diccionarios para el análisis de sentimientos, desde tweets hasta predicciones financieras y bursátiles. Daniel et al. (2017) se centró en el análisis de **sentimiento basado en eventos a partir de tweets y descubrió que el sentimiento de los eventos de tweet influye en los eventos financieros para la comunidad/empresa.**

Aunque se ha realizado mucho trabajo, el análisis de sentimiento en las redes sociales sigue siendo un desafío debido a la información limitada en un solo mensaje de redes sociales, además de expresiones indirectas, así como la información ruidosa dentro de las redes sociales como anuncios y noticias. En todo lo anterior de la técnica anterior, se menciona **la importancia del filtrado de ruido y el preprocesamiento de tweets**. El preprocesamiento incluye una amplia gama de enfoques para reducir el ruido.

Toda la investigación en el estado del arte se centra en un subconjunto de tweets fuera del flujo general de tweets. Ejemplos: **Análisis de sentimiento mediante búsqueda de palabras clave a partir de tweets para marcas/productos relevantes**. Cuando consideramos todo el flujo de tweets para investigar el sentimiento general (general) de la población, debemos considerar toda la información sobre el sentimiento que transmiten los tweets. Necesitamos **filtrar los tweets que no contengan contenido sentimental**. Ejemplos: publicidad, noticias y spam. Como se menciona en el estado de la técnica, el filtrado es muy importante, aunque desafiante. Filtrar el ruido de una subcorriente de tweets relacionados con un tema, evento o marca en particular ha sido difícil; El ruido filtrado de todo el flujo general de tweets es aún más difícil. Por lo tanto, el primer desafío es **cómo filtrar la información sin ningún elemento sentimental del flujo general de tweets.** Después de filtrar todos los tweets que no son de sentimiento, el análisis de sentimiento para tweets generales sigue siendo más desafiante que el análisis de sentimiento para una aplicación específica, esto se debe a la gran diversidad de vocabularios y expresiones utilizadas en diferentes áreas. Se vuelve más complejo cuando el tweet no pertenece a ningún tema definido, por ejemplo, “Buenos días hoy”, y tiene información sobre sentimientos, pero no para un área en particular. Sin embargo, esta información es indicativa de que el individuo es feliz y contribuye al índice general de felicidad de toda la población. Por otro lado, dado el análisis de sentimiento de los tweets generales, obtener sentimientos sobre temas, marcas, eventos, etc. específicos es simplemente una tarea de **subselección de tweets que utiliza palabras clave o técnicas de aprendizaje automático para extraer tweets del flujo general de tweets.**

Se ha realizado una cantidad considerable de trabajo sobre el análisis de sentimientos de las redes sociales, que no podemos enumerar en su totalidad. Sin embargo, como se indica en una encuesta reciente (Yue et al.2018), la mayor parte del trabajo se centra en la **extracción de opiniones de los consumidores y la población para las empresas que realizan encuestas sobre los productos y servicios correspondientes, la seguridad nacional o las opiniones del público sobre eventos como las elecciones.** Se realiza muy poco trabajo sobre el análisis del sentimiento general de la población debido a los desafíos mencionados en la Sección. 1. Este artículo propone un **sistema general para monitorear el índice de felicidad de la población utilizando el sentimiento análisis de datos de redes sociales** (Twitter) que se introduce en la sección 3.

**El sistema propuesto**

Hay **cinco** componentes principales: **recolector inteligente de tweets, filtro de tweets, clasificador de sentimientos, selector de subsecuencias de tweets y visualizador** (como se muestra en la Fig. 1).

**Recolector inteligente**

Estamos utilizando las API REST de Twitter proporcionadas por Twitter para la recolección de tweets. Java se utiliza para el desarrollo del sistema y Twitter4j se utiliza para conectarse a las API REST oficiales de Twitter. Las API REST de Twitter proporcionan 100 \* 180 tweets cada 15 minutos, lo que suma un máximo de 1.728.000 (más de 1 millón) de tweets por día. Sin embargo, sólo podemos acceder a los tweets que hayan sido indexados y puestos a disposición del público. Twitter afirma que no todos los tweets están indexados y que solo se proporciona un subconjunto de todo el flujo de Twitter cuando se utilizan las API REST de Twitter. Twitter también ofrece suscripciones pagas que pueden permitir varios niveles adicionales de acceso. Sin embargo, a nivel de país/nación, una parte del flujo general de Twitter sigue siendo enorme considerando la cantidad total de tweets dentro de ese país/nación. Por lo tanto, siempre que el recolector pueda acceder y recolectar una cantidad constante (porcentaje) del total de tweets, los tweets que recolectamos son representativos de la tendencia/índice general de felicidad de toda la sociedad. Experimentalmente, el recolector obtiene entre 0,3 y 0,4 millones de tweets diarios en los Emiratos Árabes Unidos y alrededor de 50 mil tweets diarios en Abu Dhabi. Las transmisiones son muy estables, como se muestra en la Fig. 2. Todos los tweets adquiridos contienen información sobre el autor, la ubicación, la hora y el contenido del tweet (texto). **La implementación utilizó las API REST de Twitter para solicitar tweets automáticamente y luego enviarlos a Elasticsearch de vez en cuando**. Cabe señalar que, al recopilar tweets, se deben seguir los términos y condiciones de uso de Twitter y se deben tomar medidas para garantizar su cumplimiento.

**Filtrar**

Este es un componente muy importante del sistema propuesto, especialmente cuando se trata de tweets generales y no de un subconjunto particular, como tema, marca o evento. Hay **tres** etapas de filtrado como se muestra en la Fig. 1.

**Nivel 1**

En primer lugar, los tweets se **filtran por autor, mensajes automáticos** (twitts generados automática o semiautomáticamente por un sistema o agencia), **número de móvil, información no válida y oraciones automáticas** (tuits generados automáticamente). Las **cuentas/autores de empresas o medios no se considerarán para el sentimiento de la población** y se etiquetan manualmente como cuentas no personales y, por lo tanto, no contribuyen al sentimiento de la población. Para **identificar estas** **cuentas no personales**, se requiere un proceso manual. Sin embargo, este es un proceso incremental; las nuevas cuentas no personales siempre **se agregarán de manera incremental una vez identificadas**, lo que significa que cada vez se necesita menos participación manual con el tiempo. Los **mensajes automáticos** pueden provenir de cuentas personales, pero se generan automáticamente y tampoco contribuyen al análisis de sentimiento general. Ejemplos. "Un usuario me siguió, dos usuarios dejaron de seguirme".

Captura de pantalla de un computador

Descripción generada automáticamente con confianza media

Además, también observamos que los **tweets con números de móvil tienden a ser anuncios o conversaciones con cuentas de atención al cliente**, que también consideramos que no contienen elementos de sentimiento y se filtran. Los **tweets que no incluyen información válida** (por ejemplo, tweets que solo contienen URL o cadenas sin sentido) también se filtran. En el **mundo árabe**, **muchas personas utilizan servicios para generar automáticamente tweets que tienen enlaces a sitios web de oración**. Consideramos que estos tweets de oración generados automáticamente no contienen información sobre sentimientos; por lo tanto, también serán filtrados en la primera etapa de filtrado.

Diagrama

Descripción generada automáticamente

**Etapa 2**

**Los tweets restantes se ingresarán en la red bayesiana (BN) y en la asignación latente de Dirichlet (LDA) al mismo tiempo para clasificarlos como tweets de sentimiento y tweets sin sentimiento.** Las entradas a BN son las propiedades de contexto de los tweets, como el número de URL/hipervínculos incluidos, menciones (@), hashtags (#) y emoji, así como la longitud válida de los tweets (es decir, la longitud después de eliminarlos). URL y @). Además, ingresamos si los tweets incluyen **números de teléfono/móvil, información del año** y **expresiones informales** (más de 2 caracteres repetidos continuos). **La Figura 3 muestra el modelo generado del filtro de tweets de BN**. El modelo (tanto la estructura como los parámetros) se entrena/optimiza a partir de un lote de tweets etiquetados como datos de entrenamiento. Por lo tanto, **BN examina las propiedades/contexto de un tweet,** mientras que **LDA examina el texto/contenido de un tweet.** Al combinar los resultados de **BN y LDA**, consideramos tanto el contenido/contexto del tweet como otras características de los tweets. Ejemplos: autor, URL, etc. para identificar si un tweet contiene un elemento de sentimiento o necesita ser filtrado. **BN es bueno para tratar con datos estructurados**, es decir, características de los tweets, mientras que **LDA es bueno para tratar con texto, es decir, el contenido de un tweet**. **Los tweets clasificados como tweets sin sentimiento tanto por BN como por LDA se descartarán en esta etapa.**

**Etapa 3**

Ahora, los tweets restantes se ingresan en el **filtro LDA** mejorado para clasificar los tweets **relevantes para el sentimiento y los tweets irrelevantes para el sentimiento** (lo llamamos basura) **y el otro para clasificar positivos, negativos y otros** (neutrales o basura). Se ha demostrado que los despalilladores son una parte importante del preprocesamiento. El número de tweets filtrados por la etapa 1 (filtrados por autores, mensajes automáticos, números de móvil, información no válida y oraciones automáticas) se muestra en la Fig. 4. Podemos ver que un pequeño porcentaje (menos del 5%) están filtrados. durante el filtrado de la etapa 1. Agregar más autores filtrados aumentará este porcentaje, pero debemos equilibrar el esfuerzo manual y el filtrado obtenido en esta etapa. **Debido a que el filtrado de la etapa 1 se basa en reglas, todos los tweets basura de la etapa 1 se definen como basura; por lo tanto, la precisión en esta etapa es del 100%.** La cantidad de tweets filtrados en la etapa 2 se muestra en la Fig. 5. Podemos ver una gran cantidad de tweets filtrados. **Verificamos manualmente una muestra de tweets filtrados en la etapa 2 y demostramos que casi todos los tweets clasificados como basura lo son en realidad (99% de precisión),** lo que significa que perdemos muy poca información al usar el filtrado de la etapa 2. Necesitamos asegurarnos de no perder demasiada información sobre sentimientos (tweets positivos y negativos) al filtrar los tweets basura. Después de los filtros de las etapas 1 y 2, la tarea más difícil es la clasificación de los sentimientos (positivos, neutrales, negativos y basura). Cubriremos los detalles y presentaremos los resultados en las secciones del experimento. **El filtro obtiene los tweets más recientes** (que aún no han sido etiquetados como basura/relevantes) **de Elasticsearch y el modelo generado calcula la probabilidad de que los tweets sean basura y escribe los resultados en Elasticsearch.**

**Clasificador de sentimientos LDA**

**El elemento central de nuestro sistema de monitoreo del índice de felicidad es la clasificación de tweets**, que se utiliza tanto para **el filtrado** (como se mencionó en el último párrafo) como para el **análisis de sentimientos** (para clasificar los tweets relevantes para el sentimiento en positivos, negativos, neutrales y basura). Esto también puede considerarse como una combinación de análisis de sentimiento y una etapa adicional de filtrado. Sin embargo, este es un modelo combinado y no lo separamos como una etapa de filtrado separada. **El LDA (para filtrado y clasificación de sentimientos) necesita un módulo de preprocesamiento**. En el preprocesamiento, además de eliminar la información no válida y normalizar el texto, también se aplica un **stemmer**. La precisión de la clasificación de los tweets es clave. **Aplicamos nuestra versión mejorada de LDA para la clasificación de tweets, que ha logrado una mayor precisión para los problemas generales de clasificación de tweets**. El pseudocódigo del proceso de análisis de sentimiento y filtrado se muestra en la Fig. 6. El algoritmo detallado para nuestro LDA mejorado se incluye en las referencias y el pseudocódigo se incluye en la referencia. Se ha presentado una solicitud de patente sobre el método LDA mejorado, y se encuentran más detalles en la referencia. Nuestra LDA mejorada propuesta también se compara con algoritmos existentes aplicados para comparar conjuntos de datos de sentimiento de tweets, y los resultados se analizan en la Sección. 4. Nuestro sistema propuesto recopila continuamente los últimos tweets (que no han sido etiquetados según el sentimiento) de Elasticsearch **y calcula la probabilidad del sentimiento** (positivo, negativo o neutral) para estos últimos tweets **y almacena los resultados en Elasticsearch.**

**Selector de subsecuencia de tweets**

Además de proporcionar opiniones sobre la secuencia general de tweets, también se pueden proporcionar opiniones sobre temas y eventos de interés limitando la secuencia por selección de temas, selección de hashtags, selección de ubicación y período de tiempo o búsqueda de palabras clave. **El selector de subsecuencia de tweets proporciona una capacidad integral para ver diferentes componentes de sentimiento de diferentes áreas/temas para componer el índice de felicidad general.** En otras palabras, nuestro sistema propuesto **ayuda a comprender no sólo el sentimiento nacional general, sino también los componentes que contribuyen al sentimiento nacional y su efecto,** más precisamente, **las razones subyacentes del sentimiento nacional**. Por lo tanto, proporcionar potencialmente una idea de los factores que influyen en el índice. Ejemplo: La población en general está contenta, pero no tanto, con el transporte público, por lo que mejorar el transporte público podría ayudar a mejorar el índice general de felicidad.

**El selector de subtransmisiones está integrado en el visualizador para seleccionar subtransmisiones de interés mediante la búsqueda desde Elasticsearch.**

Texto

Descripción generada automáticamente con confianza media

Interfaz de usuario gráfica, Texto, Aplicación, Correo electrónico

Descripción generada automáticamente

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

**El visualizador**

El visualizador completo muestra las opiniones de los tweets a lo largo del tiempo y en todas las ubicaciones, tanto para la opinión general como para la subsecuencia de tweets mediante la selección de subsecuencias de tweets, como se muestra en la Fig. 7.

**Clasificador LDA mejorado para el análisis de opiniones para conjuntos de datos de referencia**

El análisis de opiniones de los tweets es un problema difícil de clasificación de textos. **La precisión de la clasificación de texto es la clave para un sistema de monitoreo del índice de felicidad y se utiliza tanto para el proceso de filtrado** (clasificar todos los tweets en tweets relevantes para el sentimiento e irrelevantes para el sentimiento) **como para el análisis de sentimientos** (clasificar los tweets relevantes para el sentimiento en positivos, negativos, neutrales y basura).

Desde el estado del arte, **se aplican varias técnicas de aprendizaje automático para la clasificación de texto al análisis de sentimientos**, incluido **Bayes ingenuo, entropía máxima (MAXENT), máquina de vectores de soporte (SVM), red neuronal artificial (ANN) y asignación latente de Dirichlet (LDA).** LDA (Blei et al. 2003) fue introducido por Blei y **es un modelo estadístico para clasificación basado en la distribución de frecuencia palabra-tema**. Cuando LDA se aplica directamente a los tweets, perdemos precisión debido a la brevedad de los tweets. Propusimos un método, el LDA mejorado, para abordar esta dificultad, que es objeto de una solicitud de patente pendiente. En comparación con otros métodos de aprendizaje automático, **no es necesario definir previamente palabras clave para el LDA en sí**, lo que lo hace más atractivo para minería de texto. Probamos la mayor precisión de nuestra LDA mejorada propuesta sobre SVM y otros algoritmos en nuestro trabajo anterior.

Estamos utilizando nuestro **LDA mejorado para el análisis de sentimiento de tweets, es decir, clasificación de positivos, negativos, neutrales y basura, y como filtro, es decir, clasificación de tweets relevantes e irrelevantes para el sentimiento**. Nuestra LDA mejorada propuesta también se compara con algoritmos de última generación para conjuntos de datos de tweets de referencia.

Para demostrar las ventajas de utilizar nuestro LDA mejorado para el análisis de sentimiento en comparación con los algoritmos de última generación, aplicamos nuestro LDA mejorado a un conjunto de conjuntos de datos de referencia utilizados en la literatura, como se muestra en la Tabla 1.

Primero aplicamos nuestro método propuesto para comparar la base de datos OMD (Wu et al. 2016), STS-Gold (Wu et al. 2016) y la base de datos HRC (Ren et al. 2016; Wu et al. 2016). La comparación de nuestro método propuesto con los métodos existentes y los métodos de referencia se muestran en la Tabla 2 y las Figs. 8, 9 y 10. En comparación con los resultados de (Wu et al. 2016), nuestra **LDA mejorada propuesta supera a todos sus algoritmos para todos los conjuntos de datos, principalmente porque sus métodos están basados en léxico, mientras que nuestra LDA mejorada propuesta** **se basa en el aprendizaje automático**. En comparación con el mejor resultado de otras técnicas de aprendizaje automático con aprendizaje en conjunto, nuestro LDA mejorado propuesto es comparativo con su precisión (0,8579 usando el nuestro y 0,8774 usando métodos basados en conjuntos para el conjunto de datos OMD, 0,8629 usando el nuestro y 0,8510 usando métodos basados en conjuntos para el conjunto de datos HCR y 0.9130 usando el nuestro y 0.8902 usando métodos basados en conjuntos para el conjunto de datos STS-Gold). Su aprendizaje conjunto se basa en los mejores resultados combinados de métodos populares de aprendizaje automático (**NB, SVM, KNN y C4.5**), lo que demuestra que **el aprendizaje conjunto aumenta la precisión de los métodos de aprendizaje automático en general**. Por lo tanto, esperamos que **la precisión de nuestro LDA mejorado propuesto aumente cuando se integre en técnicas de aprendizaje en conjunto**, lo cual es parte de nuestro trabajo futuro. Nuestro LDA mejorado propuesto también es mejor que el método de Zhao (Pandey et al. 2017) para el conjunto de datos STS-Gold.

Tabla

Descripción generada automáticamente

Gráfico, Gráfico de barras

Descripción generada automáticamente

Luego aplicamos nuestro LDA mejorado propuesto a los conjuntos de datos de STS y Sanders para realizar comparaciones de precisión con diferentes sistemas. Aplicamos la misma estrategia de validación cruzada cinco veces para comparar nuestra precisión con el resultado de Ren et al. (2016) y Pandey et al. (2017), como se muestra en la Tabla 3 y la Fig. 11. Podemos ver que nuestro **LDA mejorado propuesto obtiene la mejor precisión**, que en general es un 2% mejor en precisión para ambos conjuntos de datos de STS y Sanders. En comparación con el método de Zhao (Pandey et al. 2017), nuestro método propuesto obtiene una mayor precisión para el conjunto de datos manual STS, y la precisión para el conjunto de datos de Sanders no está disponible en el artículo de Zhao.

Otros conjuntos de datos de tweets que utilizamos para comparar la precisión son: **manual de datos de prueba** (también conocido como prueba manual STS), **Twitter-sander-apple2**, **Twitter-sander-apple3** y **conjunto de datos de Twitter en línea**. Los resultados de precisión se muestran en la Tabla 4 y la Fig. 10. En la Tabla 4, podemos ver que **nuestro LDA mejorado propuesto logra la mejor precisión para los conjuntos de datos del manual de datos de prueba.**

Twitter-sander-apple2 y conjunto de datos de Twitter. Cuando el LDA mejorado propuesto se aplica a Twitter-sander-apple3, logramos una precisión de 0,7836, que es la segunda mejor entre todos los métodos comparados. Twitter-sander-apple3 tiene 3 categorías en lugar de 2 como en los otros conjuntos de datos. Esto es de esperar, ya que **cuando se utilizan clasificadores basados en entrenamiento, la precisión disminuye cuando aumenta el número de categorías.**

Zimbra et al. (2016) probaron el rendimiento de varios algoritmos para el análisis de sentimientos utilizando el conjunto de datos de **Sanders Twitter.** La comparación de exactitud en precisión, recuperación y puntuación F se muestra en la Tabla 5. Nuestro LDA mejorado y nuestro bosque aleatorio propuestos tienden a lograr mejores resultados que otros métodos. El **bosque aleatorio obtiene la mejor precisión, pero un recuerdo muy bajo para las categorías positivas y negativas, lo que hace que el bosque aleatorio sea menos útil aquí porque no logra identificar la mayoría de los tweets positivos y negativos reales**. La puntuación F para las categorías positivas y negativas de Random Forest es muy baja debido a su baja tasa de recuperación. **Nuestra LDA mejorada propuesta logra la mejor puntuación F para las categorías positivas y otras, y la segunda mejor precisión para la categoría negativa.**

**Análisis de sentimiento general**

**Precisión para el análisis de sentimiento general**

Estamos ejecutando un recolector en línea en segundo plano para recolectar continuamente tweets de Abu Dhabi y Emiratos Árabes Unidos. **Estos tweets están prefiltrados utilizando las reglas de autores, mensajes automáticos, información no válida y oraciones automáticas** (como etapa 1 del filtrado). Luego, **los tweets son filtrados por** **Bayes Nets** (como etapa 2 del filtrado). **Luego, la LDA mejorada propuesta se aplica a los tweets que no están prefiltrados como basura/sentimiento irrelevante**. El LDA mejorado propuesto es un método de aprendizaje supervisado que necesita un archivo de entrenamiento para generar el modelo. El archivo de entrenamiento contiene una lista de tweets (después del filtrado) que se etiquetan manualmente. como positivo, negativo, neutral o basura/irrelevante, porque los tweets restantes todavía contienen algunos tweets basura que no se pudieron filtrar en las etapas 1 y 2. El análisis de sentimientos es una tarea difícil de clasificar, y lo es aún más para el análisis de sentimientos generales de todos los tweets. La dificultad surge de la ambigüedad de las expresiones de sentimiento (que son aún más ambiguas en los tweets debido a su naturaleza informal) y la interpretación subjetiva de un individuo a otro.

Para generar los datos de entrenamiento y prueba, se utilizan dos tipos de esquemas de entrenamiento: etiquetado individual y etiquetado grupal. En el etiquetado grupal, se solicita a cada uno de los etiquetadores que etiquete todos los tweets del conjunto, por lo que cada tweet es etiquetado por cada etiquetador. En el etiquetado individual, cada tweet es etiquetado únicamente por un etiquetador. **El etiquetado grupal se utiliza para identificar la coherencia del etiquetado de diferentes etiquetadores, mientras que el etiquetado individual se utiliza para ampliar el tamaño de los datos de entrenamiento.**

Twitter identifica y etiqueta los tweets con el idioma utilizado, pero en muchos países árabes, la mayoría de los tweets son árabe o inglés. Ejemplo: porcentaje de tweets en Abu Dhabi como se muestra en la Fig. 12. En la Fig. 12, **podemos ver que la combinación de tweets en inglés y árabe cubre aproximadamente el 90% del total de tweets en Abu Dhabi, Emiratos Árabes Unidos.**

Sin embargo, algunos tweets en árabe también incluyen ocasionalmente palabras en inglés. Si la mayoría de las palabras de un tweet son palabras árabes, Twitter lo etiqueta como árabe. **Nuestra LDA mejorada propuesta es independiente del idioma y, por lo tanto, puede manejar tanto el árabe como el inglés**. **Generamos modelos separados para árabe e inglés para una mayor precisión usando la misma técnica (nuestro LDA mejorado)** porque los tweets en inglés y árabe usan conjuntos de vocabulario diferentes. Las palabras en inglés que aparecen con frecuencia en los tweets en árabe también se pueden capturar y construir en el modelo árabe generado. En teoría, es posible un modelo único para tratar tanto con árabe como con inglés, pero podría perder algo de precisión debido a una mayor variación en el vocabulario total al combinar vocabularios en árabe e inglés.

**Para los tweets en inglés, entrenamos nuestro modelo LDA mejorado usando 12 mil tweets etiquetados** (conjunto de datos A, de entrenamiento individual) **y lo probamos usando un conjunto diferente de 300 tweets aleatorios con etiquetado grupal** (conjunto de datos B, todos estos 300 tweets están etiquetados por 4 etiquetadores y se utiliza un esquema de votación entre estos cuatro etiquetadores para decidir el etiquetado final para la prueba). **Los 4 etiquetadores son hablantes nativos de inglés de diferentes países de habla inglesa para cubrir los usos generales del inglés.**

Para resaltar la ambigüedad de las expresiones de sentimiento en los tweets y la dificultad del análisis de sentimientos de los tweets, probamos la coherencia de nuestros 4 etiquetadores para los 300 tweets aleatorios (conjunto de datos B) y el porcentaje de coherencia de los 4 etiquetadores es solo del 46 %. La precisión de nuestro LDA está entre 44 y 52 % al cambiar el tamaño de los datos de entrenamiento (600 a 12 k). Por lo tanto, nuestra LDA mejorada propuesta obtiene una precisión comparable o mejor en comparación con el etiquetado manual. La mayoría de los desacuerdos entre etiquetadores están relacionados con neutral versus positivo y neutral versus negativo, lo que indica que el etiquetado en sí depende de la personalidad: pensamiento positivo o pensamiento negativo. Este doble confirma los desafíos del análisis de sentimiento. Además, también existe cierta confusión entre positivo y negativo en el etiquetado manual. Ejemplo: "Te extraño" es etiquetado como positivo por un etiquetador y negativo por otro, y ambos son explicables. Esto muestra aún más los desafíos del análisis de sentimiento general en comparación con el análisis de sentimiento específico de tema/producto/evento/servicio.

Ahora investigamos cómo el tamaño de los datos de entrenamiento influye en la precisión. Entre el conjunto de 12.000 tweets etiquetados muestreados aleatoriamente (conjunto de datos A), tomamos muestras de los tweets etiquetados desde el 5 % (600) hasta el 100 % (12 k) de tweets como conjuntos de entrenamiento para generar el modelo. Probamos el modelo generado utilizando el conjunto de datos B, un conjunto diferente de 300 tweets etiquetados por 4 etiquetadores y un esquema de votación. Ejecutamos el experimento 10 veces y la precisión promedio según el tamaño de los datos de entrenamiento se muestra en la Fig. 13. La línea azul muestra la precisión de cuatro categorías (positiva, negativa, basura y neutral). La mayoría de las veces no necesitamos distinguir entre basura y neutral, porque ninguno de los dos contribuye al cambio de sentimiento. La línea naranja muestra la precisión de tres categorías (positiva, negativa y [basura o neutral]). **Podemos observar la tendencia a aumentar la precisión a medida que aumenta el conjunto de datos de entrenamiento, especialmente al comienzo del aumento del tamaño de los datos de entrenamiento. La precisión se mantiene estable después de 6 k de datos de entrenamiento.**

En el mundo real, para el análisis del sentimiento de la población, ni lo neutral ni lo basura contribuyen al análisis del sentimiento. **Por lo tanto, podemos combinar tweets basura y neutrales como tweets irrelevantes para filtrar, lo que lleva a 3 categorías para el clasificador** (positivo, negativo e irrelevante como una combinación de basura y neutral). **Podemos ver que la precisión es de alrededor del 60% con tres categorías,** lo que nuevamente muestra la dificultad de clasificar los sentimientos de los tweets generales sin restricciones.

La Tabla 6 muestra un conjunto de resultados de pruebas de 300 tweets aleatorios. Las figuras de color verde son las correctamente clasificadas; las cifras de color azul son aquellas mal clasificadas pero se puede considerar que se compensan entre sí para que las estadísticas generales sean correctas; las cifras de color gris son aquellas mal clasificadas entre basura y neutrales que no cambian las estadísticas de cambios en el sentimiento de la población; por lo tanto, pueden ignorarse. **Los errores de clasificación más impactantes que afectan el rendimiento y uso del sistema son la clasificación incorrecta entre neutro y positivo**, **y entre neutro y negativo.** Se necesita más investigación para abordar esta área. Podemos ver que la precisión de la Tabla 6 es del 54,6% para la clasificación de cuatro categorías: positiva, negativa, neutral y basura, y del 63,0% para la clasificación de tres categorías: positiva, negativa y otros (neutral o basura).

**Para los tweets en árabe, utilizamos nuestra LDA mejorada propuesta para los datos de entrenamiento en árabe** (el mismo proceso que para el inglés indicado anteriormente), y la precisión se muestra en la Tabla 7. La precisión para la clasificación de cuatro categorías: positiva, negativa, neutral y basura es del 54,4%, y la de clasificación en tres categorías: positiva, negativa y otras (neutrales o basura) es del 59,2%. Podemos ver que nuestra LDA mejorada propuesta obtiene una precisión similar para el árabe que para el inglés**. La ligera pérdida de precisión se debe a la complejidad del propio idioma árabe.** **Cómo mejorar el preprocesamiento del árabe para mejorar aún más la precisión de su clasificación es parte de nuestro trabajo futuro.**

**Precisión para el análisis de sentimientos generales con emoji**

La gente utiliza muchos emojis en las redes sociales actuales para expresar emociones. **Los emojis son una fuente de información muy valiosa para el análisis de sentimientos y están destinados para proporcionar una representación más reconocible del sentimiento. Incluir emoji en el análisis de sentimientos mejorará enormemente la precisión**. Cuando probamos la coherencia del etiquetado de 4 etiquetadores para otros 300 tweets aleatorios (conjunto de datos C), todos los cuales incluyen emoji, la coherencia del acuerdo para los 4 etiquetadores aumenta al 52% (desde el 44% para el muestreo general de todos los tweets). **Esto muestra que los emojis ayudan a reducir la vaguedad del sentimiento**.

Sin embargo, no existen definiciones claras para el sentimiento de los emojis. Ejemplo: Un corazón negro y pesado puede expresar un sentimiento positivo la mayor parte del tiempo, pero en ocasiones también puede ser negativo. **El sentimiento representado por el emoji también depende en gran medida del contexto** (el texto del mensaje utilizado), lo que significa que un léxico de emoji para el análisis de sentimientos no funcionará bien. La Tabla 8 muestra las frecuencias de diferentes emojis que aparecen en 300 tweets con emoji muestreados al azar. En nuestro método propuesto, como parte del proceso previo, todos los emojis se reemplazan con cadenas semánticas significativas, que se definen en <https://emojipedia.org/>. Ejemplo ✨ se reemplaza por "Destellos". Después de estos reemplazos, todos los emojis se interpretan como cadenas significativas y comprensibles para las máquinas. Ayuda a la máquina a comprender mejor el sentimiento cuando se usan emoji junto con las palabras utilizadas en los tweets.

Nuestro LDA mejorado propuesto funciona mejor cuando se trata de tweets que contienen emoji. Al aplicarlo a los mismos 300 tweets con emoji mencionados anteriormente (conjunto de datos C), obtenemos una precisión del 68% en comparación con la consistencia entre cuatro etiquetadores del 52%. **Esto muestra que nuestra LDA mejorada propuesta obtiene una precisión mucho mayor que el etiquetado manual cuando se trata de tweets que contienen emoji.**

Los emojis junto con el texto dan como resultado un juicio de sentimientos más seguro.

**Monitoreo de sentimiento en tiempo real**

Estamos utilizando los resultados de nuestro sistema para el análisis estadístico de los cambios en el sentimiento de la población a lo largo del tiempo. La Figura 14 muestra los cambios de sentimiento en Abu Dhabi, Emiratos Árabes Unidos, durante dos semanas, del sábado 11 de marzo de 2017 al domingo 26 de marzo de 2017. **En general, la gente tiende a ser más feliz durante los fines de semana, 11 de marzo (sábado), 18 de marzo (sábado), 24 (viernes, que es fin de semana en los Emiratos Árabes Unidos) de marzo de 2017 que los días laborables, lo cual no es ninguna sorpresa**. Además, el lunes 20 de marzo de 2017 (en el cuadro azul) es el día internacional de la felicidad; también podemos ver el pico de sentimiento positivo alrededor de ese día. **Otro hallazgo interesante es que el índice negativo se mantiene estable a lo largo del tiempo y los cambios de sentimiento tienden a ser un movimiento de neutral a positivo durante los fines de semana o períodos de eventos/festivales.** Esto se puede atribuir a que ciertas personas tienden a ser negativas durante períodos prolongados debido a circunstancias personales, ambientales y psicológicas; es más, hay personas que son crónicamente infelices. Por lo tanto, vemos este porcentaje relativamente estable de sentimiento negativo. Por otro lado, la mayoría de las personas tienden a ser más proclives a ser positivas y felices ahorrando factores específicos, como el día de la semana, eventos y factores ambientales. Esta última porción de la población constituye aquellos que muestran sentimientos positivos o neutrales. Si bien reconocemos que hay un cambio estadístico a lo largo del tiempo, al ver a las personas pasar de un grupo a otro, **podemos ver una distribución estable a lo largo del período de los resultados**. **Esto explica mejor los cambios periódicos en el sentimiento de la población y su correlación con los días de la semana (laborables o fines de semana).**

Además, el sistema propuesto proporciona una idea de las razones subyacentes que impulsan el cambio de sentimiento en la población a través del tiempo, especialmente dentro de un tema/área de interés particular. Ejemplo: sentimiento sobre la economía o la educación. Por ejemplo, cuando buscamos tweets relevantes para la educación durante 2019**, podemos ver un pico de tweets negativos relevantes para la educación alrededor del 20 de marzo de 2019** (como se muestra en la Fig. 15). **Cuando miramos el contenido de estos tweets, descubrimos que estamos en el período de exámenes escolares en los EAU y los estudiantes están preocupados o quejándose de los exámenes**. También podemos ver un patrón similar a finales de noviembre. La Tabla 9 muestra ejemplos de tweets relacionados con los exámenes del 24 de marzo como ejemplos.

**Conclusión y trabajo futuro**

En este artículo, propusimos un sistema para monitorear el sentimiento de la población general desde las redes sociales. Destacamos y discutimos los desafíos del análisis de sentimiento a partir de tweets generales en lugar de flujos específicos como marcas, eventos, tweets relacionados, así como los desafíos para el análisis de sentimiento en los países árabes.

Para el análisis de sentimiento general, necesitamos identificar con precisión los tweets con sentimiento entre todos los tweets. Para hacer esto, propusimos y utilizamos filtros integrales de varios niveles que pueden filtrar con precisión los tweets que no son de sentimiento. Un filtro preciso es un requisito previo para un análisis de sentimiento preciso.

Un clasificador de mensajes cortos es clave para un filtrado y un análisis de sentimientos precisos. Utilizamos nuestro LDA mejorado propuesto para tweets (Wang et al. 2017; Wang y AlRubaie 2015; Blei et al. 2003) 7 análisis de sentimiento que puede lograr la mejor precisión para conjuntos de datos de tweets de referencia en comparación con los métodos existentes, como se muestra en la sección. 4. Para el análisis de sentimiento general de todos los tweets, el LDA mejorado propuesto puede lograr una precisión similar o mejor para el análisis de sentimiento que el etiquetado humano, como se muestra en la Sección. 5. También se desarrolló un visualizador de tweets integral para mostrar los cambios en el sentimiento de los usuarios, ya sea en general o por intereses particulares, en distribuciones de ubicación geográfica y a lo largo del tiempo.

Aunque nuestro LDA mejorado muestra una mayor precisión en los conjuntos de datos de referencia, la precisión del análisis de sentimiento general utilizando contenido del mundo real es del 68 % y aún debe mejorarse. El aprendizaje profundo y BERT (Devlin et al. 2019) se están volviendo populares para las tareas de PNL, incluida la clasificación de textos. Sin embargo, BERT comenzó para el inglés y se ha extendido a otros idiomas como el árabe, es decir, AraBERT a través de Hugging Face (Wolf et al. 2020) muy recientemente. Nuestro trabajo futuro será investigar AraBERT para el análisis de sentimiento general y cómo combinarlo con nuestro método propuesto para un mejor rendimiento en precisión.