



# Minería de Opinión no supervisada en Twitter

J. Ángel Díaz-García  
Dpto. Ciencias de la Computación e I.A.  
Universidad de Granada

M. Dolores Ruiz  
Dpto. Ingeniería Informática  
Universidad de Cádiz

Maria J. Martín-Bautista  
Dpto. Ciencias de la Computación e I.A.  
Universidad de Granada

**Resumen**—La minería de opiniones ha sido ampliamente estudiada en la última década dado su gran interés en el ámbito de investigación e innumerables aplicaciones al mundo real. En este trabajo proponemos un sistema que combina reglas de asociación, generalización de reglas y análisis de sentimientos para catalogar y descubrir tendencias de opinión en la red social Twitter. A diferencia de lo extendido, se usa el análisis de sentimientos para favorecer la generalización de las reglas de asociación. Para ello, primeramente mediante minería de textos se resume un conjunto inicial de 1.7 millones de tuits captados de manera no dirigida en un conjunto de entrada para los algoritmos de reglas y análisis de sentimientos de 140718 tuits. Sobre este último conjunto se obtienen sets de reglas, estándar y generalizadas, fácilmente interpretables sobre personajes que el propio sistema revelará como interesantes.

**Index Terms**—Minería de opiniones, reglas de asociación, análisis de sentimientos, análisis de tendencias, aprendizaje no supervisado

## I. INTRODUCCIÓN

Las técnicas de minería de datos están presentes en casi todas las vertientes de estudio y desarrollo con las que los seres humanos actualmente trabajan. Hay ciertos problemas en los que estas técnicas destacan notablemente influenciados por los nuevos paradigmas económicos y sociales, donde las redes sociales han tomado un papel relevante. Es en este último punto de las redes sociales donde surge lo que conocemos como análisis de tendencias o minería de opiniones en la que se utilizan las técnicas de minería de datos para el análisis de opiniones. Objeto de estudio en el que se trata de comprender o analizar comportamientos, actividades y opiniones, por ejemplo, de consumidores de cierto producto o usuarios de cierta red social.

En este trabajo, se propone un novedoso enfoque no supervisado para el problema del análisis de tendencias y la minería de opiniones cuya base nace de la unión de dos técnicas bien diferenciadas dentro del ámbito de la minería de datos, las reglas de asociación y la minería de opiniones o análisis de sentimientos, aunque diferenciando la supervisada y no supervisada. Estas técnicas por separado, han sido ampliamente estudiadas en la literatura, donde se ha constatado la potencia de las reglas de asociación para resumir y descubrir conocimiento de un gran conjunto de datos, así como la gran utilidad del análisis de sentimientos para la realización de un análisis subjetivo de los problemas o dominios donde se aplican estas técnicas. El presente proyecto, propone por tanto combinar las tres vertientes utilizando el análisis de sentimientos para enriquecer el proceso posterior de obtención

de reglas de asociación, cuya finalidad será descubrir patrones de opinión, algo que difiere de la práctica totalidad de la literatura, donde se aplican las reglas de asociación para mejorar el paso posterior de análisis de sentimientos, polarizando entidades textuales, como por ejemplo tuits en buenos, neutros o malos sin obtención de patrones sobre aquellos factores que implicarán esos resultados, así como su interpretación.

La aportación del estudio al estado del arte en la materia de la minería de opiniones y el análisis de tendencias está por tanto en el diseño de un sistema que es capaz de trabajar con datos sin filtrado que puedan ser obtenidos de Twitter, por ejemplo de un país o lengua en concreto durante un período temporal que podrá variar. Sobre estos datos, el sistema será capaz de resumir cientos de miles de tuits, en un conjunto de reglas fácilmente interpretable sobre aquella temática o personaje que al analista interese en un determinado momento. Posteriormente se podrá generalizar y visualizar estas reglas en función de los sentimientos generados, de manera que pasaremos de un conjunto de datos desestructurado a conjuntos muy reducidos de reglas que representarán los sentimientos y las tendencias, con las que podremos categorizar las opiniones sobre personajes que el sistema obtendrá de manera no dirigida.

La metodología seguida por el sistema para conseguir este objetivo es la que podemos ver en la figura 1. Partiendo de un gran conjunto de tuits, el sistema propuesto limpia en primera instancia los datos para posteriormente realizar una selección de muestras basada en la localización de personas relevantes, tras esto el sistema polariza en paralelo los términos de los tuits y obtiene reglas de asociación sobre los mismos de manera que puede combinar la polarización de sentimientos y las reglas en un enfoque jerárquico, ofreciendo en la salida, para un mismo personaje, una nube de términos con los patrones y tendencias de opinión y un set de reglas generalizadas que nos ofrecen un nivel más de obtención de información sobre esas tendencias de opinión.

Para validar el buen funcionamiento del sistema propuesto, se han escogido dos conocidos políticos de EEUU, Donald Trump y Hillary Clinton. El motivo de escoger estos personajes, entre todos los que el sistema descubrió como relevantes en la red social Twitter, es que podemos validar acorde a los actos acaecidos en el último año las opiniones, reglas y reglas basadas en sentimientos que el modelo sea capaz de minar sobre los mismos. Cabe destacar, que podríamos haber escogido cualquier otro personaje y el sistema habría funcionado también correctamente.

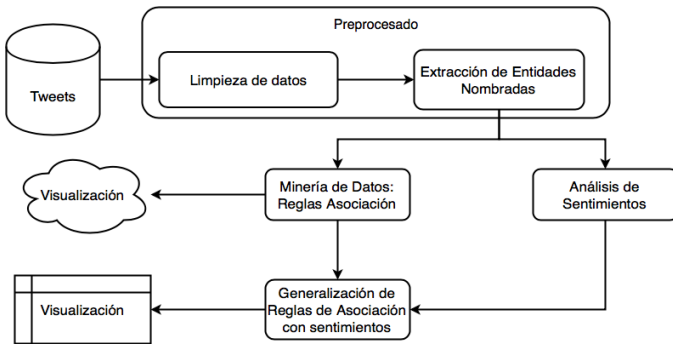


Figura 1. Flujo de la metodología.

## II. TRABAJOS RELACIONADOS

En esta sección, se estudian los trabajos relacionados con el presente estudio en función de las técnicas que se usan.

### *Reglas de asociación y minería de redes sociales.*

Uno de los principales trabajos que encontramos en el campo de las reglas de asociación es el propuesto en el 2000 por Silverstein et al. [1], donde se usan técnicas de reglas de asociación para el conocido problema de las cestas de la compra, que relaciona la compra de un determinado producto con la posibilidad de comprar otro distinto. Posteriormente, el campo ha sido ampliamente estudiado con artículos de gran interés aunque no será hasta 2010 cuando aparezca uno de los principales artículos que relacione el uso de reglas de asociación de causalidad en medios sociales. El artículo en cuestión, es propuesto por Oktay et al. [2], y en él, se estudia la relación entre la aparición de ciertos términos en preguntas de la web *Stack Overflow*, con la aparición de ciertos términos en las respuestas a estas preguntas, lo cual podría relacionarse con nuestro estudio, en el que tratamos de obtener e interpretar a relación de unos términos con otros, aunque las técnicas y el dominio difieren por completo.

Si nos centramos en nuestro dominio, Twitter, es a raíz de la publicación del trabajo de Pak and Paroubek [3], donde denotan la potencia de los datos obtenidos de Twitter para la minería de opiniones, cuando comienzan a aparecer multitud de estudios que se centran en esta red social como origen de datos sobre los cuales aplicar distintas técnicas de minería de datos, entre las que se encuentran, aunque en menor medida que otros enfoques, las reglas de asociación. En este área, encontramos trabajos como el propuesto por Cagliero y Fiori [4] o el propuesto por Erlandsson et al. [5]. En el primer estudio los autores usan reglas de asociación dinámicas, es decir, en el que las medidas de confianza y soporte cambian a lo largo del tiempo, con el fin de obtener datos sobre los hábitos y comportamientos de los usuarios en Twitter. En el segundo, se propone un análisis basado en reglas de asociación para encontrar *influencers* en Twitter, estudio que aborda el **estudio de influencias** dentro de las distintas vertientes de la minería de medios sociales. Nuestro trabajo podría verse por tanto como un híbrido entre estos dos estudios, donde se creará un modelo capaz de obtener tendencias de opinión o

relaciones acorde a personajes influyentes que a muy seguro aparecerán en el proceso. Esto se diferencia de otros trabajos como el de Meduru et al. [6] o el de Abascal-Mena et al. [7] entre otros, donde se realiza un filtrado previo de los tuits sobre los que se aplicará el procesado. En nuestro estudio, no se realiza ningún filtrado previo de los datos de Twitter.

### *Reglas de asociación y análisis de sentimientos.*

En cuanto al campo del análisis de sentimientos y las reglas de asociación, hay pocos trabajos relacionados debido al predominio de los métodos de clasificación en este área, pero, cabe mencionar el reciente interés en esta temática donde encontramos estudios como el de Hai et al. [8] donde se aplica un enfoque basado en reglas de asociación, co-ocurrencias de palabras y clustering, para obtener las características más comunes respecto a determinados grupos de palabras que puedan representar una opinión. El fin del estudio es ofrecer una vuelta de tuerca al proceso de análisis de sentimientos, que simplemente polariza una opinión, para poder refinarlo de manera que no solo se polarice ésta sino que se pueda ver acorde a qué palabras o características de opinión se ha llevado a cabo esta polarización. En este enfoque de análisis de sentimientos y aprendizaje no supervisado encontramos también el trabajo de Yuan et al. [9], en el que los autores proponen una nueva medida para la discriminación de términos frecuentes sin orientación aparente de las opiniones, lo que favorece el proceso de análisis de sentimientos posterior. El nexo de unión de estos estudios a grandes rasgos es el uso de las reglas de asociación e itemsets frecuentes para mejorar el proceso de análisis de sentimientos, esto difiere de nuestro estudio en que una vez minadas e interpretadas las reglas, se aplicará un enfoque jerárquico de las mismas basado en sentimientos, o lo que es lo mismo, usaremos el análisis de sentimientos para mejorar la interpretación de las reglas de asociación.

### *Reglas de asociación generalizadas*

Las aproximaciones jerárquicas al proceso de minado de reglas de asociación, están siendo estudiadas últimamente, debido en gran medida a la necesidad de condensar la información que estas representan de cara por ejemplo a mejorar los procesos de visualización. Un reciente ejemplo de este uso, es propuesto por Hahsler y Karpienko [10], donde se propone una visualización basada en matrices que hace uso de una simplificación jerárquica de los ítems que aparecen en las reglas de asociación. En el presente estudio, también se usa el enfoque jerárquico para simplificar las reglas, pero en lugar de hacer esto por categorías de ítems, lo haremos por sentimientos. Es en este punto, donde encontramos uno de los trabajos más íntimamente ligados al presente estudio, donde, como hemos introducido anteriormente, se usa el análisis de sentimientos para mejorar la información aportada por las reglas de asociación. Este trabajo realizado por Dehkharghani et al. [11], propone el uso de reglas de asociación para relacionar la co-ocurrencia de términos en tuits, a los que



posteriormente se clasifica en función de los sentimientos de estos términos que forman las reglas obtenidas.

Nuestro estudio, está ligado al anterior trabajo en la interpretación de correlación que hacemos de la co-ocurrencia de términos en un determinado tuit y en el uso de análisis de sentimientos para interpretar desde un enfoque subjetivo las reglas de asociación obtenidas. Nuestra propuesta se diferencia de la anterior en que realiza el análisis sin filtrar los datos de entrada, las palabras y términos son polarizados antes de la obtención de las reglas en un enfoque jerárquico basado en sentimientos tras lo que conseguimos un resumen de las reglas pudiendo, además de obtener opiniones y tendencias, categorizar los sentimientos suscitados por un personaje (también podría ser una marca, un producto o incluso un lugar) en un conjunto de tuits recopilados durante un período de tiempo.

### III. METODOLOGÍA PROPUESTA

En esta sección estudiaremos la metodología propuesta.

#### III-A. Preprocesado

Atendiendo a la naturaleza y cantidad de los datos con los que vamos a trabajar donde prácticamente cada uno de los de tuits contiene algún elemento que hace que sea totalmente distinto de los demás, se hace estrictamente necesaria una etapa previa de preprocesado de datos. Las técnicas usadas han sido las siguientes:

1. Eliminación de palabras vacías en inglés. A estas se le ha añadido la palabra *via*, que podemos considerar vacía en el ámbito que nos incumbe.
2. Eliminación de enlaces, eliminación de signos de puntuación, caracteres no alfanuméricos y valores perdidos (tuits vacíos).
3. Términos poco comunes. Se eliminan las palabras cuya frecuencia de aparición sea inferior a 30 ocurrencias, además de aquellas palabras que a pesar de tener más de 30 ocurrencias, tienen una longitud mayor de 13 letras. Esto último indicará que provienen de hashtags o unión de palabras que no tienen significado real o importancia en nuestro proceso.
4. Extracción de Entidades Nombradas: Se ha realizado un proceso de selección de instancias en la que mantendremos solo aquellas tuits que hablen de personas. Esto se ha realizado usando la técnica de *Name Entity Recognition* [12], de ahora en adelante NER, propuesta por la Universidad de Stanford y que se incluye en el *coreNLP* propuesto por Manning et al. [13] en la misma universidad. Tras la ejecución del proceso NER, obtenemos resultados bastante aceptables donde se localizan 140.718 tuits que hacen referencia a personas. Sobre este conjunto de tuits, vuelven a aplicarse las técnicas anteriores de preprocesado para refinar el proceso.
5. Paso a minúsculas: Se pasa todo el contenido a letras minúsculas para evitar discordancias. Pese a que el paso a minúsculas es uno de los principales pasos en minería de textos, en nuestro caso se aplica tras el proceso de

NER, ya que el uso de mayúsculas en los nombres propios facilita y mejora los resultados del proceso.

6. 2-gramas: Cabe esperar la posibilidad de obtener nombres compuestos por dos términos cuyo análisis en conjunto sea mucho más interesante y evite la aparición de reglas de asociación redundantes. La idea es relacionar o fusionar términos como **donald** seguido de **trump** en una sola palabra del tipo **donald-trump**. Para evitar trabajo innecesario, utilizamos un estudio que nos permita discernir si esta premisa de nombres compuestos es acertada y si sus frecuencias son tales como para ser útil esta unión. Mediante un estudio de los bigramas más comunes, damos por cierta la premisa de que hay ciertas palabras que cuya unión evitará redundancia en las reglas. Tras identificar estos términos se realiza mediante un proceso iterativo la unión de los mismos mediante un guión. Por otro lado con el análisis de comenzamos a obtener información sobre el dominio de los datos y las conversaciones en la red social que nos permitirán guiar los procesos de obtención de información posteriores.
7. Por último transformaremos los tuits en transacciones, donde cada palabra será un ítem.

Hemos obviado el proceso de *steaming* debido a que se podría perder interpretabilidad de cara a los procesos posteriores.

#### III-B. Obtención de reglas de asociación

La etapa de minería de datos ha sido llevada a cabo usando reglas de asociación mediante el algoritmo Apriori [14] sobre el que hemos aplicado valores de soporte de 0.001 y 0.0001 junto con un valor de confianza de 0.7. De esta manera obtendremos un número considerable de reglas pero cohesionadas y fuertes.

#### III-C. Polarización de términos basada en sentimientos

La motivación en este estudio para realizar análisis de sentimientos es el tener la posibilidad de polarizar las reglas de asociación que hemos obtenido en la anterior etapa de la metodología en función de los términos que aparezcan en el antecedente o consecuente de las mismas. Se ha propuesto un proceso iterativo en el que se recorre cada uno de los tuits obteniendo el sentimiento asociado a cada palabra en ese tuit, para esto usamos el paquete *syuzhet* [16] el cual hace uso del diccionario de sentimientos de mismo nombre creado por el laboratorio de literatura de Nebraska y que tiene en cuenta las 8 emociones básicas más extendidas propuestas por el psicólogo Plutchik [17]. Haciendo esto se genera una estructura de datos, en la que para cada palabra se obtiene cuantas ocurrencias tiene para cada sentimiento. Algunas de estas palabras y su sentimiento pueden verse en la figura 2.

#### III-D. Generalización de reglas basada en sentimientos

El último paso de la metodología pasa por combinar la etapa vista en el punto III-B de obtención de reglas y el punto III-C de polarización por sentimientos de los términos. Una de las diferencias más radicales del estudio frente a





Cuadro I  
REGLAS INTERESANTES SOBRE DONALD TRUMP.

Antecedente	Consecuente	Sop	Conf	Lift
{military,people,trans}	{donald-trump}	3.5e-04	0.71	68.79
{bans,serving,trans}	{donald-trump}	8.5e-05	0.92	88.90
{ignored,rape}	{donald-trump}	9.9e-05	1	96.31
{child,rape}	{donald-trump}	9.9e-05	0.93	89.89
{caucus,lead}	{donald-trump}	8.5e-05	0.85	82.55

*IV-B1. Donald Trump:* Para Donald Trump se ha generado un conjunto de 156 reglas. Tras el estudio de las mismas, se han resumido algunas interesantes en el cuadro I.

Analizando las primeras reglas del cuadro anterior podemos constatar una tendencia clara en cuanto a las políticas de Trump con las personas transgénero y su posibilidad de servir en el ejército de los Estados Unidos. Concretamente la regla {bans,serving,transgender} → {donald-trump} nos deja entrever que el actual presidente tenía muy claro que prohibiría el servicio de estas personas en el ejército, algo que ya por 2016 se venía barajando y que fue confirmado en 2017. Otra tendencia interesante puede ser marcada por las dos siguientes reglas, {ignored,rape} → {donald-trump} y {child,rape} → {donald-trump}, donde por medio de la minería de datos hemos obtenido una tendencia en Twitter durante la primera mitad del año 2016, donde el por aquel entonces candidato a ocupar la Casa Blanca, se vio involucrado en ciertos escándalos relacionados con violaciones o la no condena de estas. Por último, encontramos también una regla interesante en {caucus,lead} → {donald-trump} que nos constata el hecho comprobado de que todas las encuestas consideraban a este candidato líder en el caucus donde los partidos deciden a que candidato presentar a las elecciones.

Dado que tratamos de representar y obtener tendencias en Twitter, en la figura 4 se ha obtenido una representación en forma de nube de palabras, que representa en función del tamaño las palabras más usadas en co-ocurrencia con nuestro objetivo. De esta manera hemos condensado toda la información de las reglas sobre **Trump** en un solo gráfico.

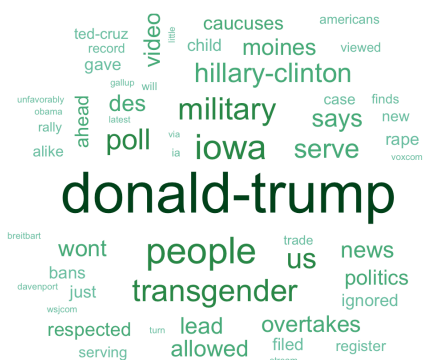


Figura 4. Nube de palabras de las reglas para Donald Trump.

Si atendemos por tanto a la representación de las reglas con nube de términos, hasta una persona sin conocimientos

sobre la temática podría deducir de qué se está hablando en Twitter y cuáles son las tendencias en relación con el candidato. Por ejemplo encontramos las palabras *transgender*, *rape*, *child* sobre las cuales anteriormente en nuestro proceso manual hemos podido obtener tendencias.

*IV-B2. Hillary Clinton:* Para Hillary Clinton se han obtenido un total de 93 reglas de asociación. Estudiando manualmente el conjunto de las reglas obtenidas podríamos acotar las del cuadro II como algunas de las más interesantes.

Cuadro II  
REGLAS INTERESANTES SOBRE HILLARY CLINTON.

Antecedente	Consecuente	Sop	Conf	Lift
{musician,squad}	{hillary-clinton}	3.83e-04	1	273.77
{musician,support}	{hillary-clinton}	3.83e-04	1	88.90
{emails,repblicans}	{hillary-clinton}	1.56e-04	1	273.77
{attack,emails}	{hillary-clinton}	1.56e-04	1	273.77

Si realizamos un enfoque de interpretación por grupos, podríamos definir claramente dos tendencias y grupos de opiniones en los tuits relacionados con Hillary Clinton:

1. El compromiso del mundo del espectáculo con su candidatura: Las primeras reglas, hacen referencia al apoyo recibido por la candidata por parte de grandes estrellas del mundo del espectáculo.
2. El escándalo de los mails: Las dos últimas reglas hacen referencia al escándalo de los mails filtrados y al uso como ataque que se dio de ellos.

El gráfico de etiquetas sobre las reglas elaborado en este caso puede verse en la figura 5.

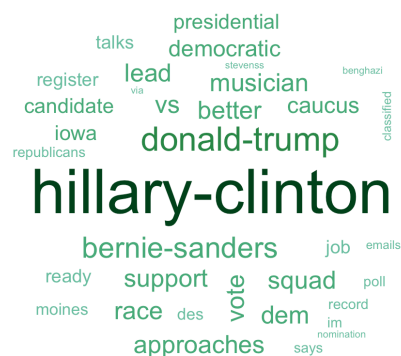


Figura 5. Nube de palabras de las reglas para Hillary Clinton.

*IV-B3. Interpretación generalizada basada en sentimientos:* Dado que tras el proceso de análisis de sentimientos tenemos para cada palabra su sentimiento mayoritario asociado, podemos intercambiarlas para generalizar. Para ello podremos filtrar los tuits que hacen referencia a los personajes que venimos estudiando en esta sección y cambiaremos las palabras de estos tuits por el sentimiento mayoritario asociado a cada una de ellas. Posteriormente, se vuelve a obtener las reglas de asociación sobre este conjunto de datos. Los resultados para Donald Trump pueden verse en el cuadro III mientras que los resultados obtenidos para Hillary Clinton podemos verlos en el cuadro IV.

Cuadro III  
REGLAS POR SENTIMIENTOS SOBRE DONALD TRUMP

Antecedente	Consecuente	Sop	Conf	Lift
{trust}	=>{donald-trump}	0.94592745	1	1
{anticipation}	=>{donald-trump}	0.59411362	1	1
{surprise}	=>{donald-trump}	0.42505133	1	1
{anger}	=>{donald-trump}	0.34565366	1	1
{fear}	=>{donald-trump}	0.29500342	1	1
{joy}	=>{donald-trump}	0.22655715	1	1
{disgust}	=>{donald-trump}	0.11293634	1	1
{sadness}	=>{donald-trump}	0.07460643	1	1

Cuadro IV  
REGLAS POR SENTIMIENTOS SOBRE HILLARY CLINTON

Antecedente	Consecuente	Sop	Conf	Lift
{trust}	=>{hillary-clinton}	0.93968872	1	1
{anger}	=>{hillary-clinton}	0.49221790	1	1
{anticipation}	=>{hillary-clinton}	0.48638132	1	1
{fear}	=>{hillary-clinton}	0.29961089	1	1
{surprise}	=>{hillary-clinton}	0.20038911	1	1
{joy}	=>{hillary-clinton}	0.14591440	1	1
{sadness}	=>{hillary-clinton}	0.07976654	1	1
{disgust}	=>{hillary-clinton}	0.07782101	1	1

Viendo la reglas se obtiene un ranking de los sentimientos que identifican a cada una de las personas estudiadas. Lo primero que sale a la vista y que podríamos concluir es que en Twitter se han emitido más tuits de apoyo y respaldo contra ambos candidatos que de otro tipo de sentimiento. Una interpretación muy interesante es la que podemos hacer del sentimiento *anger*, donde vemos cómo el 50% de los tuits que hablan de Hillary Clinton, tienen a su vez relacionados este sentimiento, por contra, Trump, tiene un 20% menos de este sentimiento, por lo que parece que la sociedad americana, a pesar de lo que parecía en Europa, estaba más en contra de Hillary Clinton que de Trump. Esto posteriormente se vería confirmado al vencer el republicano.

## V. CONCLUSIÓN Y TRABAJO FUTURO

Se ha desarrollado un modelo que es capaz de obtener patrones de comportamiento en la red social Twitter que podrían ser catalogados como tendencia y que llevados a un enfoque basado en *data streaming* podría ser incluso utilizado para categorizar las opiniones y sentimientos mayoritarios de un determinado lugar sobre un personaje en tiempo real. Se ha constatado la potencia de los métodos no dirigidos y su fácil interpretabilidad, muy cercana al lenguaje natural, en problemas similares pese a la actual hegemonía de los métodos dirigidos. Un resultado similar al obtenido con las reglas de asociación podría obtenerse mediante un estudio de frecuencias, pero estos serían computacionalmente costosos de obtener para la cantidad de datos barajada, eliminarían en cierta medida el componente no dirigido que ofrecen las reglas, al tener que decidir sobre qué palabras obtener las relaciones con otras palabras, y por último ofrecerían métricas distintas de las que pueden obtenerse de algoritmos como Apriori. De igual modo resulta interesante una comparativa entre ambos resultados y se deja la misma como trabajo futuro

que complementará al actual. Por último es necesario destacar la infinidad de temas distintos que pueden ser tratados en Twitter, haciendo que los dataset obtenidos de este sean muy interesantes y ruidosos a la par.

## AGRADECIMIENTOS

La investigación de este trabajo ha sido parcialmente apoyada por el Ministerio de Economía y Competitividad del Gobierno de España bajo el proyecto TIN2015-64776-C3-1-R.

## REFERENCIAS

- [1] C. Silverstein, S. Brin, R. Motwani, J. Ullman. Scalable techniques for mining causal structures. *Data Mining and Knowledge Discovery*, 4 (2-3) (2000), pp. 163-192
- [2] H. Oktay, B.J. Taylor, D.D. Jensen Causal discovery in social media using quasi-experimental designs. Proceedings of the first workshop on social media analytics, ACM (2010), pp. 1-9
- [3] Pak, A., & Parubek. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*
- [4] L. Cagliero and A. Fiori. Analyzing Twitter User Behaviors and Topic Trends by Exploiting Dynamic Rules. *Behavior Computing: Modeling, Analysis, Mining and Decision*. Springer, 2012 pp. 267-287.
- [5] F Erlandsson, P Bródka, A Borg, H Johnson: Finding Influential Users in Social Media Using Association Rule Learning. *Entropy* 18: 164 (2016).
- [6] Meduru, M., Mahimkar, A., Subramanian, K., Y. Padiya, P., & N. Gunjgur, P. (2017). Opinion Mining Using Twitter Feeds for Political Analysis. *International Journal Of Computer (IJC)*, 25(1), 116-123.
- [7] Abascal-Mena R., López-Ornelas E., Zepeda-Hernández J.S. User Generated Content: An Analysis of User Behavior by Mining Political Tweets. In: Ozok A.A., Zaphiris P. *Online Communities and Social Computing, OCSC 2013*. Lecture Notes in Computer Science, vol 8029. Springer, Berlin, Heidelberg
- [8] Hai Z., Chang K., Kim J. (2011) Implicit Feature Identification via Co-occurrence Association Rule Mining. In: Gelbukh A.F. (eds) *Computational Linguistics and Intelligent Text Processing, CICLing 2011*. Lecture Notes in Computer Science, vol 6608. Springer, Berlin, Heidelberg
- [9] Yuan M., Ouyang Y., Xiong Z., Sheng H. (2013) Sentiment Classification of Web Review Using Association Rules. In: Ozok A.A., Zaphiris P. (eds) *Online Communities and Social Computing, OCSC 2013*. Lecture Notes in Computer Science, vol 8029. Springer, Berlin, Heidelberg
- [10] Hahsler, M., Karpienko, R. Visualizing association rules in hierarchical groups. (2017) *Journal of Business Economics*, 87, pp. 317-335.
- [11] R. Dehkharghani, H. Mercan, A. Javeed, Y. Saygin: Sentimental causal rule discovery from Twitter. *Expert Syst. Appl.* 41(10): 4950-4958 (2014).
- [12] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating Non-local Information into Information Extraction Systems by Gibbs Sampling. *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL 2005)*, pp. 363-370.
- [13] Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. 2014. The Stanford CoreNLP Natural Language Processing Toolkit In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 55-60.
- [14] R. Agrawal and R. Srikant Fast algorithms for mining association rules in large databases. 1994. *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, pp. 487-499.
- [15] B Liu, L Zhang . A survey of opinion mining and sentiment analysis. *Mining text data*, 2012. Springer.
- [16] Jockers, M. (2017). Package 'syuzhet'. URL: <https://cran.r-project.org/web/packages/syuzhet>.
- [17] Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American scientist*, 89(4), 344-350.
- [18] Matei Zaharia, Reynold S. Xin, Patrick Wendell, Tathagata Das, Michael Armbrust, Ankur Dave, Xiangrui Meng, Josh Rosen, Shivaram Venkataraman, Michael J. Franklin, Ali Ghodsi, Joseph Gonzalez, Scott Shenker, and Ion Stoica. 2016. Apache Spark: a unified engine for big data processing. *Commun. ACM* 59, 11 (2016), 56-65.