Revista Cubana de Ciencias Informáticas
Vol. 13, No. 4, Mes Octubre-Diciembre, 2019
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 56-66

Tipo de artículo: Artículos originales
Temática: Reconocimiento de patrones
Recibido: 09/08/19 | Aceptado: 20/10/19

Descubrimiento de conocimientos en los comentarios que realizan los usuarios en un sistema de noticias digital

Discovery of knowledge in the comments made by users in a digital news system

Camila Gonzalez[1*], Eliana B. Ril[1], Hector Gonzalez[1], Vladimir Milian[1], Julio Camejo[2]

[1]Universidad de las Ciencias Informáticas (UCI). Km $2^{1/2}$ Autopista La Habana - San Antonio de los Baños La Habana, Cuba. {cgnapoles, ebril, hglez, vmilian}@uci.cu
[2]Universidad de Cienfuegos Carlos Rafael Rodríguez. Km 4, Cuatro Caminos, Cienfuegos, Cuba. jcamejo@ucf.edu.cu

[*]Autor para correspondencia: cgnapoles@uci.cu

Resumen
La minería de opinión es un proceso de extracción de nuevos conocimientos y datos textuales no estructurados mediante los métodos automáticos de detección y extracción de opiniones. Ha mostrado en los últimos años un gran auge, debido a la necesidad de analizar gran cantidad de opiniones y comentarios que se publican diariamente. El principal problema de los datos que se recopilan de un sistema de noticias web, es que se encuentran en forma no estructurada, lo que dificulta la identificación de la opinión pública y a su vez el sentimiento que transmiten en ella. El objetivo de esta investigación se centró en descubrir conocimientos para determinar la intencionalidad de los usuarios a través de sus comentarios sobre artículos de diferentes temáticas de la sociedad. Con el fin de realizar el procesamiento de datos y transformación de los mismos, se creó inicialmente un dataset con los comentarios de 5 artículos seleccionados por mayor nivel de comentarios. Dicha fuente de datos se utilizó como entrenamiento para el aprendizaje automático. Para ello, luego de realizar un estudio sobre los principales algoritmos de procesamiento del lenguaje natural y minería de opinión para el análisis de sentimientos se escogió específicamente máquina de soporte vectorial. Se obtuvo como resultado, mediante los métodos de clasificación, si los usuarios tenían valoraciones positivas, negativas o neutras respecto a los artículos publicados.

Palabras claves: análisis de sentimiento, descubrimiento de conocimiento, minería de opinión, procesamiento del lenguaje natural

Abstract
Opinion mining is a process of extracting new knowledge and unstructured textual data using automatic methods of detecting and extracting opinions. It has shown a great boom in recent years, due to the need to analyze a large number of opinions and comments that are published daily. The main problem with the data collected from a web news system is that they are in unstructured form, which makes it difficult to identify public opinion and, in turn, the feeling it conveys. The objective of this research focused on discovering knowledge to determine the intentionality of users through their comments on articles on different topics of society. In order to carry out data processing and data transformation, a dataset was initially created with the comments of 5

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 4, Mes Octubre-Diciembre, 2019
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 56-66

articles selected by higher level of comments. This data source was used as training for automatic learning. For this purpose, after a study on the main algorithms of natural language processing and opinion mining for the analysis of feelings, a vectorial support machine was specifically chosen. It was obtained as a result, by means of the classification methods, if the users had positive, negative or neutral valuations with respect to the published articles.

Keywords: *sentiment analysis, discovery of knowledge, opinion mining, natural language processing*

Introducción

El avance de la tecnología e internet han proporcionado plataformas para el intercambio de ideas, puntos de vista y sentimientos en todo el mundo. A partir de las redes sociales se ha creado un acceso más fácil y rápido a la información, proporcionando que desde cualquier parte del mundo se pueda acceder a cualquier revista, periódico o sistema de información digital. En dichos sitios, el usuario (lector) es el eje principal, contribuyendo mediante comentarios o posts respecto a cualquier tema que se publique, observándose como influye la información sobre las personas. Estos posts tienen características particulares, tienen formato de tipo texto, por lo que no pueden ser clasificados por categoría y su búsqueda no es sencilla, puesto que no son datos organizados y no están estructurados para ser guardados en una base de datos. Estos posts son datos no estructurados, es la información que no es fácil de buscar como audios, vídeos, imágenes y publicaciones en las redes sociales, ejemplos de ellos correos electrónicos, archivos de textos, pdf, fotos, videos y cuando se exponen los gustos y disgustos de contenidos publicados en las redes sociales. Surge una necesidad de analizar y categorizar estos datos no estructurados, por lo que se crea la ciencia, Minería de opinión (MO), dominio de investigación que trata sobre los métodos automáticos de detección y extracción de opiniones presentes en el texto según (Tran and Phan (2017); Balazs and Velásquez (2016)).

Por otra parte, la expresión de opiniones puede llevar a cabo el análisis de uno o varios sentimientos, puesto que refleja satisfacción o no dependiendo del tipo de información que se analice. Para el estudio de minería de opinión es necesario llevar a la par el término análisis de sentimiento. El análisis de sentimiento(AS) es la tarea de detectar, extraer y clasificar opiniones, sentimientos y actitudes sobre diferentes temas, como expresados en datos textuales según (Medhat et al. (2014); Liu (2012)). Esta definición tiene como objetivo observar el estado de ánimo de un usuario en relación con cualquier tema y así extraer opiniones, identificar los sentimientos expresados y clasificar su polaridad, dígase positivo, negativo y neutro. Análisis de sentimiento y Minería de opinión son términos interrelacionados, pero cumplen diferentes objetivos.

Actualmente, existen diversos métodos y sistemas que categorizan las opiniones mediante los sentimientos, dictando así en un sistema de noticias, si el artículo es positivo, negativo o neutro. Sin embargo, en Cuba,

los sistemas de noticias web aun no permiten descubrir de manera automática información relevante sobre los comentarios de la población. Actualmente, se categorizan los comentarios manualmente, siendo un trabajo no factible, puesto que se deben supervisar una significativa cantidad de comentarios diarios, haciéndose difícil hallar información relevante. Los contenidos (Posts) tienen características muy particulares, por lo general son textos cortos, descritos de manera informal o son términos escritos con errores de tipiado. En los Post puede además el autor del comentario usar un enfoque positivo o a favor del contenido o un enfoque negativo con términos mal intencionados. Teniendo en cuenta lo dicho anteriormente es muy importante para cualquier sitio web de noticias clasificar la intencionalidad del autor cuando emite un comentario.

Metodología computacional

El Descubrimiento de conocimiento (KDD, por sus siglas en inglés) Fayadd en (Fayyad et al. (1996)) lo define como:

Procesos no comunes de identificación de patrones válidos, novedoso, potencialmente útiles y finalmente comprensibles en los datos.

Su propia definición se debe a que comprende los procesos siguientes: Selección de datos, Preprocesamiento de datos, Transformación de datos, Minería de datos, Interpretación y Evaluación del conocimiento descubierto. Por lo que vale puntualizar que se utilizará KDD como metodología de trabajo.

Aunque hay varias definiciones del proceso KDD, se puede observar que todos los autores coinciden en que este es un proceso con el cual es posible obtener información útil de grandes almacenes de datos. A través de los años han variado las etapas de KDD, aunque manteniendo su objetivo y los siguientes procesos (Fayyad et al. (1996)):

- **Selección de datos:** En esta primera etapa del proceso se realiza una selección de las fuentes de datos, estas pueden ser bases de datos y/o archivos. Además, se eliminan los datos inconsistentes y se combinan las diferentes fuentes de datos que fueron seleccionadas en un Data Warehouse.
- **Preprocesamiento y transformación de los datos:** En este paso del proceso se seleccionan los atributos que serán utilizados y son transformados en un formato apropiado para el análisis que será realizado posteriormente con la minería de datos. Las dos primeras etapas del proceso KDD son las etapas en las que se consume más tiempo dado que es aquí donde se debe tener especial cuidado en la limpieza que exista en los datos, ya que sin calidad en ellos no habrá calidad en los resultados obtenidos a través de la minería de datos.

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 4, Mes Octubre-Diciembre, 2019
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 56-66

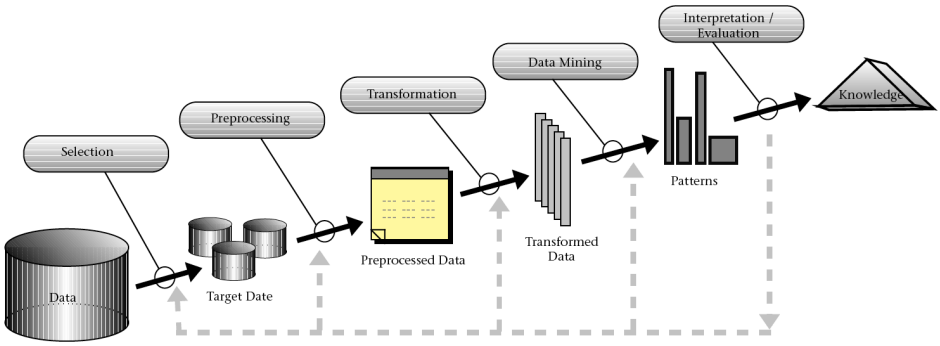

Figura 1. Etapas en la extracción de conocimiento según KDD

- **Minería de datos:** Es la parte medular del proceso KDD puesto que se perfeccionan las técnicas y algoritmos que se encargan de extraer y representar el conocimiento de forma adecuada para la toma de decisiones. Se combinan técnicas potenciando las ventajas de cada una y atenuando sus debilidades. La minería de datos tiene como propósito la identificación de un conocimiento obtenido a partir de las bases de datos que aporten hacia la toma de decisión.
- **Interpretación y evaluación del conocimiento descubierto**: Se procede al análisis de los resultados descubiertos. Incluye a su vez la resolución de posibles inconsistencias con otros conocimientos anteriores a la investigación.

Resultados y discusión

A continuación, se muestra en la figura 2, la forma en que se aplicó el proceso KDD en la presente investigación:

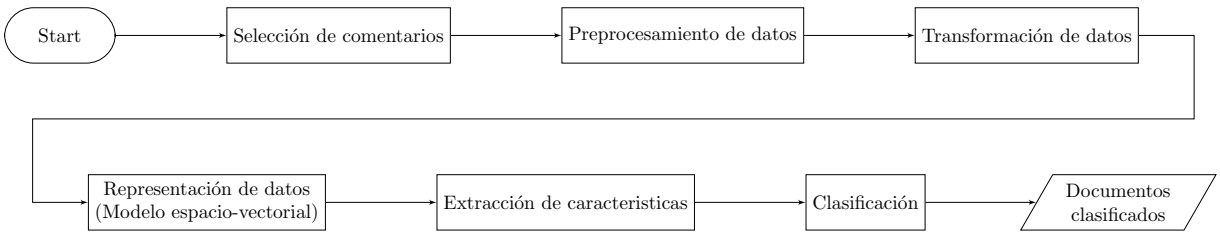

Figura 2. Aplicación de las etapas de KDD en la propuesta de solución

Revista Cubana de Ciencias Informáticas
Vol. 13, No. 4, Mes Octubre-Diciembre, 2019
ISSN: 2227-1899 | RNPS: 2301
http://rcci.uci.cu
Pág. 56-66

En La etapa, **Selección de comentarios**, se escogen los comentarios de artículos publicados en la sección, noticias más comentadas, del sitio web de noticias Cubadebate. Se seleccionan 5 noticias con diferentes temáticas e incluso diferentes cantidades de comentarios, los más polémicos entre sus ramas. Se combinan los comentarios seleccionados en una bolsa de documentos, puesto que los datos son de tipo textos. La primera fuente es el artículo *ETECSA. Internet en el móvil a partir del seis de diciembre*, obteniendo 2013 comentarios. La segunda fuente es el artículo *Esta página es toda tuya Comparte con Cubadebate tu homenaje a Fidel Castro*, donde se realizaron 4945 comentarios. La tercera fuente es el artículo *En su 20 cumpleaños, Mailen regala las primeras fotos tras el accidente aéreo*, obteniendo 737 comentarios. La cuarta fuente es el artículo *Leinier Domínguez se nacionaliza por Estados Unidos-Federación Cubana expresa desacuerdo*, con 584 comentarios. La quinta fuente, el artículo *Si de alimentos se trata. Miradas a la industria nacional* con 420 comentarios.

En la etapa de **Pre-procesamiento y Transformación de los datos** se utilizan elementos clásicos del procesamiento del lenguaje natural los cuales quedan representados en la figura 3. En la primera etapa se emplea como entrada de datos el texto a analizar (comentarios) y tendrá como salida las palabras clasificadas en categorías, o sea datos preprocesados. Esto es un importante paso en cualquier proceso de minería de datos, prácticamente implica transformar los datos en bruto en un comprensible formato de modelos de lenguaje natural. Para el procesamiento de datos y su transformación se utiliza como ejemplo el artículo 1 (Dataset 1) en toda la investigación, aunque se implementó para todos los dataset.

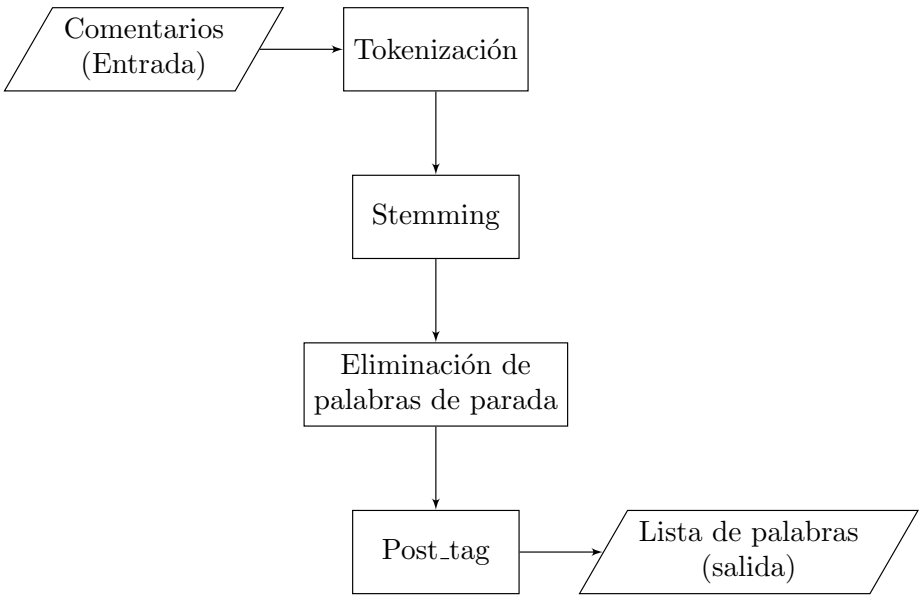

Figura 3. Etapa de pre-procesamiento de datos

Ya aplicado el procesamiento de lenguaje natural se aplica la transformación, mediante modelos de representación vectorial, es un aspecto clave en la categorización de tareas de textos. Este tiene como propósito categorizar documentos en un número fijo de categorías predefinidas. Cada documento puede ser clasificado en múltiples, exactamente una o en ninguna categoría en absoluto. La clasificación de documentos es vista como una tarea de aprendizaje supervisado, el objetivo es utilizar el aprendizaje automático para clasificar automáticamente los documentos en categorías, basadas en documentos previamente etiquetados. La inserción de palabras se refiere a las representaciones numéricas de las palabras. Actualmente existen varios enfoques de integración de palabras, aunque los que se utilizaron en la investigación son: Bolsa de palabras, Word2Vec, Esquema TF-IDF.

Bolsa de palabras se utiliza en la vectorización al transformar una colección de documentos de texto en vectores con características numéricas (Goldberg and Levy (2014); Church (2017)).

Es necesario este método para convertir las palabras en algún conjunto de vectores numéricos para luego utilizarlos en la transformación. La idea subyacente aquí es que las palabras similares tendrán una distancia mínima entre sus vectores. Con bolsa de palabras, el orden de las palabras en la historia no influye la proyección y predice la palabra actual basada en el contexto. Word2Vec usa todos estos tokens para crear internamente un vocabulario (conjunto de palabras únicas). Como resultado muestra que existe un vocabulario de 29 palabras únicas. Tf-idf se utiliza para determinar qué palabras de un corpus pueden ser favorables al uso en función de la frecuencia del documento de cada palabra. El corpus, donde guardan los dataset, se divide en dos conjuntos de datos, entrenamiento y prueba. Los conjuntos de datos de entrenamiento serán usados para ajustar al modelo y las predicciones en los datos de prueba. Los datos de entrenamiento tendrán el 70 % del corpus y los datos de prueba el restante 30 %.

Ya como último paso de la transformación de datos, se debe implementar el algoritmo de representación gráfica y estadística LDA. Se crea un objeto para el modelo LDA y se debe entrenarlo en la matriz de documento-término, se utiliza este algoritmo puesto que trata de una manera de descubrir automáticamente categorías o temas, a partir de una colección de documentos y ordenarlos en lista en función de la similitud que presenten las palabras. En la figura 4 se muestra el proceso de transformación de datos.

Una vez ejecutados los algoritmos, los datos transformados se someten a la clasificación para detectar su polaridad, mediante el algoritmo SVM. Permitirán con la información que se extraerá, a los analistas humanos completar el análisis iniciado por el texto mediante la visualización una herramienta.

En la implementación del algoritmo SVM (Weston et al. (2001); Cortes and Vapnik (1995)) o cualquier otro tipo de clasificador, es necesario introducir un conjunto de datos que ya estén manualmente clasificados. En

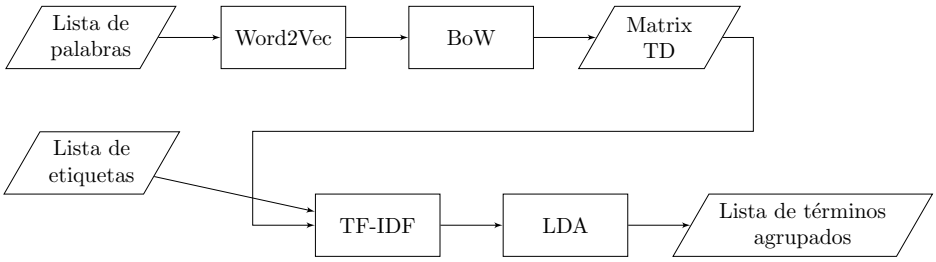

Figura 4. Etapa de transformación de datos

este caso label_data, extrae de un Excel y guarda en un arreglo los comentarios clasificados manualmente en -1(negativo), 0(neutro), 1(positivo), enseñándole al clasificador como debe comportarse. Se muestra en la Ilustración 6 el label_data con la polaridad, observándose un desbalance entre los datos. Dígase que polaridad 1 (positivo los datos), polaridad 0 (neutro los datos), polaridad -1 (negativo los datos). Este problema existe cuando hay una gran diferencia entre los porcientos de los datos, siendo esto un caso peculiar llamado problema de balance.

Para los problemas de desbalance se aplica SGDClassifier (Bottou (2012)), clasificador de máquina de soporte vectorial, este estimador implementa modelos lineales regularizados con aprendizajes de descenso de gradiente estocástico (SGD). Este clasificador encontrará el hiperplano de separación óptimo usando SVC reemplazándolo con SGDClassifier y observando las diferencias para clases que están desbalanceada.

Los resultados de aplicar la clasificación (función que predice la clase dominante) al dataset 1 (Artículo 1), obtiene como categoría dominante el término *felicidades* siendo positiva la polaridad, encontrándose que el dataset 1 (Artículo 1) es positivo.

La evaluación experimental de un clasificador usualmente mide la efectividad, que para este caso particular corresponde a la habilidad de tomar la decisión de clasificación correcta. Para poder evaluar las soluciones de los distintos algoritmos de clasificación que se aplican, es necesario disponer de un conjunto de datos, sobre el cual se ejecutan estos algoritmos, los cinco dataset de entrenamiento utilizados y que cuentan con diferentes particularidades. Las métricas que se emplean dadas por la librería sklearn para evaluar los resultados obtenidos, son:

- **accuracy_score**: mide la precisión que posee algún algoritmo sobre la fuente de datos indicada
- **predict**: da como resultado la predicción del clasificador sobre la fuente de datos entrada.

Para analizar el clasificador escogido se decide compararlo con otros clasificadores, se escoge a *Naive Bayes*, puesto que es el segundo clasificador más utilizado para la categorización de textos y el clasificador *RandomForestClassifier* un enfoque basado en conjuntos para identificar datos relevantes (Khabsa et al. (2016)), aunque como otros clasificadores no está diseñado para tratar con problemas de datos no balanceados. En la figura 5 se observa una ilustración con los valores resultantes de *accuracy* y el algoritmo que mayor valores tiene sobre la fuente de datos es el clasificador RandomForest, por lo que prueba que el algoritmo escogido en la investigación no es el más eficiente sobre la precisión de sus datos. El modelo escogido, la máquina de soporte vectorial tiene una precisión de 4.34 % de precisión que no lo hace viable. Se necesita datos para usar en la demostración,

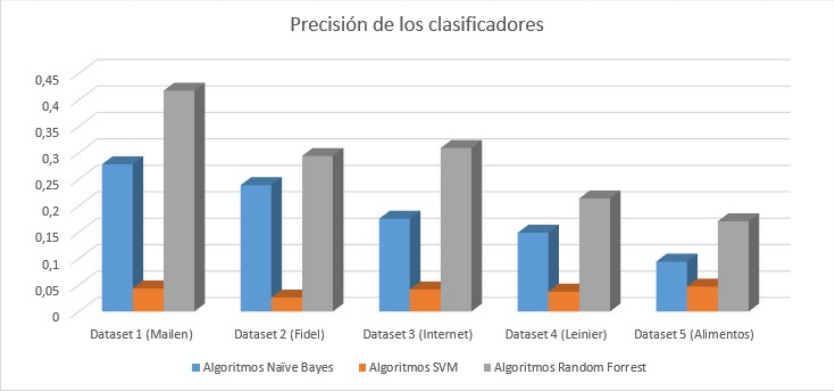

Figura 5. Precisión entre los clasificadores

así que se obtuvo el conjunto de datos de las colecciones de prueba de recuperación de texto. El conjunto de datos se dividirá en dos partes las X, que son datos predictores, siendo las variables de entrada (matriz término-documento) y las Y que son los datos tarjeta (labels). También cada tipo de datos se dividirá en un conjunto de datos para entrenamientos y otro para pruebas. Se muestra a continuación la precisión hallada a los conjuntos de datos divididos y siendo 1 su máximo valor entonces se puede proseguir adelante en los próximos pasos obteniendo una buena precisión en los datos. En la figura 5 se observan los resultados obtenidos a partir de los valores que da como resultado la precisión de cada clasificador escogido sobre la fuente de datos existente.

Como conclusión se demuestra con la métrica *accuracy* los valores altos los proporciona el clasificador *RandomForestClassifier*, superando a *Naive Bayes* y *SVM*, siendo el algoritmo escogido en las 5 dataset probada da entre los valores 0 y 0.1 nunca llegando a este último valor. Por lo que nos da la validación del clasificador como mejor para utilizar *RandomForestClassifier* con la métrica *accuracy*. Para asegurarse de las validación del clasificador, se evaluará con otras medidas de evaluación.

Para validar los resultados se siguieron los siguientes pasos: Primero se carga los conjunto de datos para

Tabla 1. Tabla comparativa entre los clasificadores aplicando precision-recall

	Naive Bayes		SVM		Randon Forest	
	Presicion	Recall	Presicion	Recall	Presicion	Recall
Dataset 1	0.091	0.096	0.000	0.006	**0.187**	**0.192**
Dataset 2	0.047	0.049	0.000	0.002	**0.074**	**0.076**
Dataset 3	0.031	0.034	0.000	0.003	**0.089**	**0.091**
Dataset 4	0.018	0.020	0.000	0.002	**0.042**	**0.044**
Dataset 5	0.008	0.010	0.000	0.002	**0.036**	**0.038**

analizar. Luego se procesan los textos y se construye un espacio con los vectores asignados a cada post.

Se escoge para precision-recall el average tipo macro y micro, se observan que los dos se comportan iguales con respecto a sus valores por lo que se escoge a desarrollar macro también porque este tipo de métrica, calcula las métricas para cada label y encuentra la media no ponderada. Ya evaluados los clasificadores sobre la fuente de datos elegidas, dan los mismos resultados validados arriba sobre la métrica **accuracy**, RandomForestClassifier es el clasificador con mayores valores.

Ya entonces aplicadas las métricas Precision-Recall, sobre los algoritmos para comparar Máquina de soporte vectorial (escogido), Naive Bayes y RandomForest, este ultimo posee los mejores valores para escoger el algoritmo a desarrollar. Obteniendo una precision macro de 18 % y 19 % en Recall, concluyendo es que si un comentario es analizado bajo el clasificador RandomForest posee una probabilidad de un 18 % de ser correcto. Aunque estos valores siguen siendo bajos para ser la probabilidad de escoger el clasificador correcto. Después de aplicar las métricas Accuracy-score y Precision-Recall se concluye que el algoritmo escogido (máquina de soporte vectorial) aunque era uno de los mejores para categorización de texto según (Manek et al. (2017); Basu et al. (2003); Lilleberg et al. (2015); Joachims (1998); Mohammad et al. (2018); Sun et al. (2017); Lan et al. (2005)), resultó ser el de menores valores en la validación.

Realizando una comparación con respecto a las predicciones de los 5 dataset se observa en los Anexos 1, 2, 3, 4 y 5 las predicciones encontradas por cada uno, dando como resultado: Artículo 1 positivo y los Artículos 2, 3, 4 y 5 negativos. Obtenidos estos resultados se puede realizar un análisis social de porque son negativos los artículos señalados y darle solución.

A parte de las métricas también se comparan los algoritmos implementados sobre la base de los dataset. Algoritmos como word2vec, su vocabulario y su puntuación máxima, la puntuación del modelo LDA sobre las fuentes de datos y la matriz resultante de cada uno de los dataset. De esta comparativa se obtiene que el uso de LDA mejora los resultados de clasificación ya que opera sobre un espacio de menor dimensión.

Conclusiones

En base a los resultados obtenidos se arribó a las siguientes conclusiones:

- Al ejecutar los pasos del procesamiento del lenguaje natural se transformó a un lenguaje binario los comentarios seleccionados del sitio web de noticias Cubadebate. A su vez, el algoritmo clasificador de análisis de sentimientos permitió, clasificar el artículo en positivo, negativo o neutro y el modelo de aprendizaje automático y la máquina de soporte vectorial posibilitó el aprendizaje automático para futuras fuentes de datos.
- Se clasificó automáticamente la polaridad a través de los comentarios de 5 artículos seleccionados en el sitio web de noticias Cubadebate, clasificando el artículo 1 como positivo y el artículo 2, 3, 4 y 5 como negativo.

Bibliografía

Jorge A Balazs and Juan D Velásquez. Opinion mining and information fusion: a survey. *Information Fusion*, 27:95–110, 2016.

Atreya Basu, Christine Walters, and M Shepherd. Support vector machines for text categorization. In *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*, pages 7–pp. IEEE, 2003.

Léon Bottou. Stochastic gradient descent tricks. In *Neural networks: Tricks of the trade*, pages 421–436. Springer, 2012.

Kenneth Ward Church. Word2vec. *Natural Language Engineering*, 23(1):155–162, 2017.

Corinna Cortes and Vladimir Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

Usama Fayyad, Gregory Piatetky-Shapiro, and Padhraic Smyth. The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11):27–34, 1996.

Yoav Goldberg and Omer Levy. word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722*, 2014.

Thorsten Joachims. Text categorization with support vector machines: Learning with many relevant features. In *European conference on machine learning*, pages 137–142. Springer, 1998.

Madian Khabsa, Ahmed Elmagarmid, Ihab Ilyas, Hossam Hammady, and Mourad Ouzzani. Learning to identify relevant studies for systematic reviews using random forest and external information. *Machine Learning*, 102(3):465–482, 2016.

Man Lan, Chew-Lim Tan, Hwee-Boon Low, and Sam-Yuan Sung. A comprehensive comparative study on term weighting schemes for text categorization with support vector machines. In *Special interest tracks and posters of the 14th international conference on World Wide Web*, pages 1032–1033. ACM, 2005.

Joseph Lilleberg, Yun Zhu, and Yanqing Zhang. Support vector machines and word2vec for text classification with semantic features. In *2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC)*, pages 136–140. IEEE, 2015.

Bing Liu. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1): 1–167, 2012.

Asha S Manek, P Deepa Shenoy, M Chandra Mohan, and KR Venugopal. Aspect term extraction for sentiment analysis in large movie reviews using gini index feature selection method and svm classifier. *World wide web*, 20(2):135–154, 2017.

Walaa Medhat, Ahmed Hassan, and Hoda Korashy. Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4):1093–1113, 2014.

Adel Hamdan Mohammad, Tariq Alwada'n, and Omar Al-Momani. Arabic text categorization using support vector machine, naïve bayes and neural network. *GSTF Journal on Computing (JoC)*, 5(1), 2018.

Shiliang Sun, Chen Luo, and Junyu Chen. A review of natural language processing techniques for opinion mining systems. *Information fusion*, 36:10–25, 2017.

Thien Khai Tran and Tuoi Thi Phan. Mining opinion targets and opinion words from online reviews. *International Journal of Information Technology*, 9(3):239–249, 2017.

Jason Weston, Sayan Mukherjee, Olivier Chapelle, Massimiliano Pontil, Tomaso Poggio, and Vladimir Vapnik. Feature selection for svms. In *Advances in neural information processing systems*, pages 668–674, 2001.