

Análisis de sentimientos en críticas de cine utilizando SVM y combinación de n-gramas

Cesar Alexis Estrada Palacios, José Luis Tapia Fabela

Universidad Autónoma del Estado de México,
Estado de México,
México

{kirdrazler, joseluis.fabela}@gmail.com

Resumen. El análisis de sentimientos en críticas de cine es una tarea importante en la industria cinematográfica y en la investigación académica. En este trabajo se propone un nuevo método para el análisis de sentimientos en críticas de cine utilizando SVM y la combinación de n-gramas. El método fue evaluado en el corpus Muchocine y se comparó con los métodos del estado del arte. Los resultados muestran que el método propuesto logra un mejor rendimiento en términos de *precision*, *recall* y *f-measure* en comparación con los métodos del estado del arte, demostrando que la combinación de n-gramas como modelo de representación mejora el rendimiento del clasificador. Las aportaciones de este trabajo incluyen el nuevo método propuesto, la demostración de la eficacia de la combinación de n-gramas y la comparación con los métodos del estado del arte.

Palabras clave: Análisis de sentimientos, críticas de cine, SVM, n-gramas.

Sentiment Analysis in Movie Reviews Using SVM and Combination of N-grams

Abstract. Sentiment analysis in film reviews is an important task in the film industry and in academic research. In this paper, a new method for the analysis of sentiments in film reviews is proposed using SVM and the combination of n-grams. The method was evaluated in the Muchocine corpus and compared with state-of-the-art methods. The results show that the proposed method achieves a better performance in terms of precision, recall and f-measure compared to the state of the art methods, demonstrating that the combination of n-grams as a representation model improves the performance of the classifier. The contributions of this work include the new proposed method, the demonstration of the efficiency of the combination of n-grams and the comparison with the methods of the state of the art.

Keywords: Sentiment analysis, movie reviews, SVM, n-grams.

1. Introducción

La creciente cantidad de información disponible en línea ha llevado a una mayor necesidad de herramientas y técnicas que permitan el análisis y la clasificación de

grandes volúmenes de datos textuales. La clasificación de sentimientos es un tema relevante en el análisis de opiniones y evaluaciones de productos y servicios. Un caso en particular es el análisis de sentimientos en críticas de cine el cual es de gran interés en la industria del entretenimiento.

A lo largo de los años, se han propuesto varios métodos para abordar este problema, incluyendo enfoques basados en reglas, aprendizaje automático y redes neuronales. A pesar de los avances, todavía existen desafíos en la identificación precisa de la polaridad de los sentimientos.

En este contexto, el presente estudio propone un nuevo método basado en aprendizaje automático para la clasificación de sentimientos en críticas de cine en español. El método se evaluó en el corpus Muchocine y se comparó con los métodos del estado del arte. Se utilizó una combinación de características basadas en n-gramas y la técnica de clasificación *SVM primal*. En línea con estudios previos, se reconoce la importancia de la selección de características y del algoritmo de clasificación en el rendimiento del modelo [1, 2]. Por otra parte, se investigó el efecto del *stemming* en la *precision* de la clasificación y se encontró que empeora los resultados.

Los resultados del presente estudio indican que el método propuesto logra una *precision*, *recall* y *f-measure* superiores a los resultados expuestos en el estado del arte [2, 3]. Además, se encontró que la combinación de características basadas en n-gramas mejora el rendimiento del clasificador. Estos hallazgos sugieren que el método propuesto podría ser útil en la industria del entretenimiento para la evaluación de críticas de cine.

En cuanto a trabajos futuros, se sugiere la exploración de modelos basados en redes neuronales recurrentes y la utilización de modelos pre-entrenados como Bert para la clasificación de sentimientos en críticas de cine en español. También se propone la optimización de *SVM* mediante algoritmos genéticos y la evaluación del modelo utilizando *word2vec*. Asimismo, se sugiere la aplicación y evaluación del método propuesto en otros corpus de críticas de cine en español.

En resumen, el presente estudio aborda el problema de la clasificación de sentimientos en críticas de cine en español y propone un nuevo método basado en aprendizaje automático. Los resultados indican que el método propuesto supera los métodos del estado del arte y que la combinación de características basadas en n-gramas mejora el rendimiento del clasificador. Estos hallazgos podrían ser útiles en la industria del entretenimiento para la evaluación de críticas de cine en español.

2. Materiales y método

2.1. Materiales

Los materiales utilizados para los experimentos fueron:

- Corpus de críticas de cine: El corpus utilizado en este estudio es Muchocine, que consta de 5500 críticas de cine en español, previamente etiquetadas como positivas, negativas o neutrales. Estas críticas fueron recopiladas de diferentes fuentes en línea, como blogs y sitios web de reseñas de películas. Este corpus ha sido ampliamente utilizado en la literatura para evaluar el rendimiento de diferentes métodos de análisis de sentimientos en español [2,

3]. La importancia de este corpus radica en el hecho de que es el único corpus en español que se ha utilizado específicamente para el ámbito de críticas de cine. Además, se utilizaron algunas herramientas de procesamiento de lenguaje natural como la biblioteca *Natural Language Toolkit* (NLTK) [4] y la herramienta *Porter Stemmer* [5] para el preprocesamiento de los textos.

- Procesamiento de texto: Se utilizó la biblioteca NLTK para preprocesar el corpus de críticas de cine. Este preprocesamiento incluyó la eliminación de puntuación, la conversión a minúsculas, la eliminación de stopwords y la lematización.
- Características de texto: Para representar las críticas de cine, se utilizó la técnica de representación de n-gramas, que ha demostrado ser efectiva en la detección de polaridad en textos [6], pero en este trabajo se propone la combinación de n-gramas de palabras (unigramas, bigramas y trigramas) con la finalidad de conocer si el combinar n-gramas mejora los resultados obtenidos, tal como se ha visto en el trabajo de [7] para la tarea de atribución de autoría donde se logran grandes resultados al combinar n-gramas.
- Clasificadores de texto: Se evaluaron diferentes implementaciones de SVM, como SVM lineal y SVM primal, para conocer si se pueden mejorar aún más los resultados obtenidos [8].
- Evaluación del modelo: Para evaluar el rendimiento del modelo, se utilizaron las métricas estándar de precisión, recall y f-measure [2]. También se utilizó la matriz de confusión para evaluar la capacidad del modelo para distinguir entre las críticas positivas y negativas.

2.2. Método

En esta sección se describe detalladamente el proceso seguido para desarrollar el método propuesto en este trabajo de investigación. El objetivo principal de este método es el análisis de sentimientos en críticas de cine en español, utilizando como corpus de evaluación el conjunto de datos Muchocine.

Se utilizó una metodología basada en el proceso de minería de texto, que se divide en las siguientes etapas:

1. Preprocesamiento de datos: En esta etapa, se aplicaron diversas técnicas para limpiar y normalizar los datos. Se eliminaron signos de puntuación, caracteres especiales y se convirtieron todas las letras a minúsculas tal como se ha hecho en otros trabajos de análisis de sentimientos en español [9, 10].
2. Extracción de características: En esta etapa, se construyeron los modelos de representación de las críticas, a partir de las características que se consideran relevantes para la clasificación. Para ello se utilizaron los n-gramas, que han demostrado ser una técnica efectiva en el análisis de sentimientos en español [6]. Además, se utilizó la técnica de representación de bolsa de palabras (*bag of words*) para la construcción de las características [11]. Se utilizaron tres tipos de características: palabras, bigramas y trigramas. La combinación de estos n-gramas se convierte en el modelo de representación de cada crítica.

3. Entrenamiento del clasificador: En esta etapa, se entrenó el clasificador para que pudiera clasificar las críticas en positivas o negativas. Para el entrenamiento del clasificador se utilizó *SVM*. Este clasificador ha sido ampliamente utilizado en el análisis de sentimientos en español. Además, se utilizó la técnica de validación cruzada para evaluar el rendimiento de cada clasificador, tal como se ha hecho en otros trabajos de análisis de sentimientos en español.
4. Evaluación del rendimiento del clasificador: Se utilizaron diversas métricas, como *precision*, el *recall* y el *f-measure* [12]. Estas métricas han sido ampliamente utilizadas en la literatura para la evaluación del rendimiento de clasificadores en el análisis de sentimientos en español. Además, se realizaron pruebas de significación estadística para comparar los resultados obtenidos con los métodos del estado del arte, tal como se ha hecho en otros trabajos de análisis de sentimientos en español.
5. Análisis de resultados: En esta etapa, se analizaron los resultados obtenidos para determinar cuál es el mejor método de clasificación de sentimientos en críticas de cine en español en base a los resultados obtenidos.

3. Estado del arte

El análisis de sentimiento en críticas de películas ha sido un área de investigación popular en los últimos años, y se ha explorado ampliamente el uso de técnicas de aprendizaje automático para realizar esta tarea. En un estudio reciente realizado por [13] propone el uso de análisis de sentimiento para predecir el éxito o fracaso de una película a través de reseñas en línea utilizando minería de texto y la técnica de aprendizaje automático de *SVM*.

La principal dificultad en el análisis de sentimiento es la selección de palabras que indican sentimiento y la forma en que las personas expresan su opinión. El artículo también señala que los algoritmos de análisis de sentimiento pueden ser limitados debido a factores culturales, matices lingüísticos y contextos diferentes. Los resultados de la investigación anterior muestran que los métodos de aprendizaje automático, como *SVM* y *Naive Bayes*, han mejorado significativamente la precisión en la clasificación de las revisiones de películas en positivas o negativas.

El artículo "Análisis de sentimientos en Twitter para español" presenta un enfoque de *SVM* para el análisis de sentimientos en tweets en español. Los autores participaron en el taller TASS2013, que se centró en el idioma español, y utilizaron un corpus de tweets en español anotados con polaridad.

El artículo destaca los desafíos del análisis de sentimientos en Twitter, incluyendo el uso de frases no gramaticales, emoticonos, abreviaturas y jerga. En general, proporciona información sobre los desafíos del análisis de sentimientos en Twitter en español, y presenta un enfoque de *SVM* que logró los mejores resultados para las tareas de análisis de sentimientos del taller TASS2013 [14].

El artículo [15] presenta un enfoque de análisis de sentimientos de revisiones de películas utilizando *SVM* con el método de Ganancia de Información para la selección de características. El objetivo del estudio es mejorar la precisión de la clasificación de opiniones positivas y negativas en revisiones de películas. Los resultados muestran que *SVM* basado en Ganancia de Información logró una mayor precisión que el *SVM*

convencional en dos conjuntos de datos de revisiones de películas, alcanzando un aumento de precisión del 2.6% en el conjunto de datos de Cornell y un aumento de precisión del 0.166% en el conjunto de datos de Stanford.

El artículo también discute la importancia de la selección de características en el rendimiento de la clasificación de texto, y cómo la Ganancia de Información ha demostrado ser el mejor método de selección de características en comparación con otros algoritmos. Los autores creen que el estudio puede ayudar a los usuarios a tomar decisiones informadas sobre la calidad de las películas, y contribuir al desarrollo de la teoría relacionada con el análisis de sentimientos y la clasificación de texto.

En el campo de análisis de sentimientos en críticas de cine en español, el trabajo de Cruz [3] fue pionero. Para superar la falta de un corpus adecuado, los autores propusieron el corpus Muchocine, que consistía en 3878 críticas de cine extraídas del sitio web. Sin embargo, debido a las limitaciones en tiempo y hardware disponible en ese momento, el corpus no fue utilizado en su totalidad. El autor utilizó 400 críticas en total para experimentar con diferentes métodos, dividiéndolas en 200 críticas positivas y 200 críticas negativas, y dejando a un lado las críticas neutrales.

Utilizó bigramas de palabras, pero no realizó ninguna combinación como proponemos nosotros. Además, no utilizó un clasificador como SVM o redes neuronales, ya que no tenía los recursos para procesar todas las críticas. El autor logró obtener un resultado del 77%, lo que proporciona una base para comparar los resultados con el método propuesto. Aunque el autor solo experimentó con un conjunto reducido de críticas, su trabajo sigue siendo valioso para entender los desafíos y oportunidades del análisis de sentimientos en críticas de cine en español.

Este artículo [2] se enfoca en el análisis de sentimientos en textos en español, que es una tarea desafiante relacionada con la minería de texto y el procesamiento del lenguaje natural. Aunque hay trabajos actuales, la mayoría se centran en textos en inglés. La presencia del español en la web está aumentando, por lo que se han llevado a cabo algunos experimentos sobre un corpus de reseñas de cine en español.

En este trabajo se presentan varios experimentos utilizando cinco algoritmos de clasificación (*SVM*, *Naive Bayes*, *BBR*, *KNN*, *C4.5*). Los resultados obtenidos son muy prometedores y alentadores para continuar investigando en esta línea. Este artículo describe los experimentos realizados sobre un corpus de reseñas de cine en español, y se comparan los resultados con otros trabajos que han utilizado el enfoque semántico.

4. Experimentos y resultados

En esta sección comparamos los mejores resultados obtenidos por el método propuesto contra los del estado del arte, posteriormente mostramos una serie de experimentos que nos condujeron a lograr a estos resultados.

4.1. Comparación con los métodos del estado del arte

Se realizó una comparación entre los mejores resultados obtenidos por el método propuesto y los métodos del estado del arte previamente aplicados al corpus Muchocine, con el objetivo de evaluar su rendimiento. Los resultados se presentan en

Tabla 1. Comparación de los resultados del método propuesto (Estrada) y los métodos de Martínez [2] y Cruz [3].

Autor	Clasificador	Precision	Recall	F-measure
Estrada	SVM primal	91.94	89.06	90.48
Martinez	SVM Lineal	87.73	87.69	87.71
Martinez	Naive Bayes	84.08	84.01	84.04
Cruz	SOv1 - S	77.50	N/A	N/A
Cruz	SOv2 - NS	69.05	N/A	N/A

*N/A No se cuenta con la información, ya que no se evaluó con esta métrica.

la tabla 1. Se puede observar que el método propuesto logra una *precision*, *recall* y *f-measure* superior a los métodos del estado del arte [2, 3], demostrando que el método propuesto es una mejor opción para el análisis de sentimientos en críticas de cine.

Específicamente, el método propuesto logra una *precision* del 91.94%, un *recall* del 89.06% y una *f-measure* del 90.48%, lo que representa una mejora significativa en comparación con los resultados obtenidos por otros autores que utilizaron clasificadores *SVM lineal* y *Naive Bayes*.

Además, el método propuesto también supera a los resultados obtenidos por Cruz utilizando los clasificadores *SOv1 - S* y *SOv2 - NS*. Se concluye que la combinación de n-gramas es la clave para lograr un mejor rendimiento en la clasificación de textos, y que la elección del clasificador se vuelve secundaria cuando se utiliza una combinación de características en el modelo de representación.

A pesar de los resultados prometedores, todavía hay margen de mejora para alcanzar la meta de una precisión del 99.9%. Por lo tanto, se espera que este trabajo sirva como punto de partida para futuros estudios que busquen optimizar la precisión y aplicarlo en la vida real.

Es importante mencionar que, para llegar a estos resultados, se testearon diferentes variantes del método, como fue probar distintos *SVM*, tamaños de n-gramas, combinación de n-gramas y el uso de *stemming*. A continuación, se presentan estos experimentos y se comparan los resultados con las distintas variantes evaluadas.

4.2. Experimento 1: Comparación entre SVM primal, lineal y polinomial

En este experimento, se evaluaron las variantes de *SVM lineal*, *primal* y *polinomial* como clasificadores en el método propuesto, utilizando la combinación de unigramas + bigramas + trigramas como base de comparación. Los parámetros $c=1$ y $\epsilon=0.001$ para cada variante. Los resultados obtenidos se presentan en la tabla 2.

Los resultados mostrados en la tabla 2 indican que el clasificador *SVM Primal* tuvo el mejor rendimiento, con una *precision* del 90.87% y un *f-measure* del 90.48%. El *SVM lineal* también obtuvo buenos resultados, con una *precision* del 89.73% y un *f-measure* del 89.24%. Por otro lado, el *SVM polinomial* tuvo un rendimiento inferior, con una *precision* del 84.79% y un *f-measure* del 83.74%. Estos resultados sugieren que el clasificador *SVM Primal* es la mejor opción para la clasificación de textos en este corpus específico.

Tabla 2 Comparación del rendimiento de las variantes de SVM.

Clasificador	Accuracy	Precision	Recall	F-measure
SVM Primal	90.87	91.94	89.06	90.48
SVM lineal	89.73	91.06	87.50	89.24
SVM polinomial	84.79	87.29	80.47	83.74

Tabla 3 Comparación entre diferentes tamaños de n-gramas de palabra.

Característica	Accuracy	Precision	Recall	F-measure
Unigramas	88.21	88.19	87.5	87.84
Bigramas	80.61	79.84	80.47	80.16
Trigramas	79.09	78.29	78.91	78.6

Tabla 4 Comparación de la combinación de las bolsas n-gramas.

Característica	Accuracy	Precision	Recall	F-measure
Unigramas+bigramas	88.97	88.98	88.28	88.63
Unigramas+trigramas	88.97	85.6	83.59	84.58
Bigramas+trigramas	85.17	79.23	80.47	79.84
Unigramas+bigramas +trigramas	90.87	91.94	89.06	90.48

4.3. Experimento 2: Comparación de rendimiento entre diferentes tamaños de n-gramas de palabra

En este experimento, se evaluó el rendimiento de tres modelos diferentes de n-gramas: unigramas, bigramas y trigramas, para identificar con cual se obtienen mejores resultados. La table 3 muestra los resultados en términos de las métricas *accuracy*, *precision*, *recall* y *f-measure*.

De acuerdo con los resultados, unigramas presenta una mayor *accuracy* con un 88.21%, seguido de bigramas con 80.61% y trigramas con 79.09%. En cuanto a la métrica *precision*, unigramas también presenta el mayor valor con 88.19%, seguido de bigramas con 79.84% y trigramas con 78.29%.

En cuanto al *recall*, unigramas presenta un valor superior a bigramas y trigramas con 87.5% en comparación con 80.47% y 78.91% respectivamente. Respecto a la métrica *f-measure*, unigramas nuevamente presenta un valor superior con 87.84%, seguido de bigramas con 80.16% y trigramas con 78.6%. En conclusión, los resultados indican que en general, el tamaño de unigramas es el que presenta un mejor rendimiento en las diferentes métricas evaluadas.

4.4. Experimento 3: Combinación de n-gramas como modelo de representación

En este experimento se combinan diferentes tamaños de n-gramas para crear una sola bolsa de n-gramas. En la tabla 4 se muestran los resultados de las diferentes

Tabla 5 Aplicación de *stemming* en la etapa de preprocesamiento.

Característica	Accuracy	Precision	Recall	F-measure
Sin stemming	90.87	91.94	89.06	90.48
Con stemming	87.45	88.62	85.16	86.85

combinaciones posibles entre unigramas, bigramas y trigramas. Con base en los resultados obtenidos, se puede afirmar que la combinación de unigramas + bigramas + trigramas es la que proporciona el mejor rendimiento, con una *accuracy* del 90.87%, una *precision* del 91.94%, un *recall* del 89.06% y un *f-measure* del 90.48%. Estos resultados demuestran que la utilización de tres tamaños de n-gramas es más efectiva que la combinación de solo dos tamaños.

En segundo lugar, se encuentra la combinación de unigramas + bigramas, con una *accuracy* del 88.97%, una *precision* del 88.98%, un *recall* del 88.28% y un *f-measure* del 88.63%. Aunque estos resultados son buenos, no superan a los obtenidos por la combinación de tres tamaños de n-gramas.

Las combinaciones de unigramas y trigramas, y bigramas y trigramas, presentaron resultados similares, pero aún están por debajo de la mejor combinación de n-gramas. Esto indica que la combinación de tres tamaños de n-gramas es la que proporciona el mejor rendimiento en este corpus.

En conclusión, la combinación de tres tamaños de n-gramas puede ser de gran utilidad para mejorar la *precision* de modelos de lenguaje en el futuro. Los resultados obtenidos en este estudio proporcionan una guía útil para la elección de la combinación de n-gramas en la construcción de modelos de clasificación de texto.

4.5. Experimento 4: Impacto de la técnica de stemming

En este experimento se evaluó el impacto de la técnica de *stemming* en el rendimiento del método propuesto. Para llevar a cabo la comparación, se utilizó el modelo de n-gramas combinado que previamente había obtenido los mejores resultados, es decir, la combinación de unigramas, bigramas y trigramas de palabra. Los resultados obtenidos utilizando *stemming* en la etapa de preprocesamiento se compararon con los resultados obtenidos sin utilizar esta técnica.

Los resultados obtenidos se presentan en la tabla 5. Se puede observar que la técnica de *stemming* no mejora los resultados del clasificador, al contrario, se obtuvieron peores resultados en todas las métricas. Además, la técnica de *stemming* conlleva más tiempo de ejecución en la etapa de preprocesamiento. Es importante destacar que esta comparación se limitó únicamente al modelo de n-gramas combinado que previamente había obtenido los mejores resultados.

5. Conclusiones y trabajo a futuro

La discusión de los resultados obtenidos en este estudio revela varias conclusiones importantes. En primer lugar, se pudo demostrar que el método propuesto en este trabajo logró un desempeño superior en términos de *accuracy*, *precision*, *recall* y *f-*

measure en comparación con los métodos del estado del arte evaluados en el corpus Muchocine.

Esto indica que la combinación de características basadas en n-gramas puede mejorar significativamente la *precision* del análisis de sentimientos en críticas de cine. Además, se observó que el rendimiento del clasificador *SVM Primal* fue ligeramente superior al del clasificador *SVM lineal* en todas las métricas evaluadas.

Otra conclusión importante que se puede extraer de los resultados obtenidos es que el uso de *stemming* en la etapa de preprocesamiento no mejoró los resultados del clasificador, sino que en realidad empeoró los resultados. Este hallazgo es consistente con los estudios del estado del arte y sugiere que el *stemming* no es una técnica efectiva para mejorar el análisis de sentimientos en críticas de cine.

Además, se observó que las diferencias en los resultados obtenidos para cada métrica de evaluación (*precision*, *recall* y *f-measure*) fueron relativamente pequeñas. Esto se debe en parte a la cantidad similar de críticas positivas y negativas evaluadas. Sin embargo, estos resultados también sugieren que las métricas de evaluación pueden ser insuficientes para medir la efectividad del análisis de sentimientos en críticas de cine, y que pueden ser necesarias medidas adicionales de desempeño.

En cuanto a las aportaciones de este estudio, se propone un nuevo método para el análisis de sentimientos en críticas de cine, que logra mejores resultados que los métodos del estado del arte evaluados en el corpus Muchocine. Además, se demuestra que la combinación de n-gramas es mejor que no combinar n-gramas, lo que puede abrir el camino a nuevas investigaciones donde se puedan combinar distintos tipos de características.

Finalmente, es importante destacar que este estudio tiene algunas limitaciones. Por ejemplo, se utilizó un único corpus de datos Muchocine y se limitó el análisis a las críticas de cine en español. Por lo tanto, es posible que los resultados no sean generalizables a otros idiomas o géneros de películas.

Además, se utilizó un conjunto de características relativamente limitado, lo que puede limitar la *precision* del análisis de sentimientos. Se necesitan investigaciones adicionales para evaluar la generalización de los resultados obtenidos en este estudio y para explorar la utilidad de otros conjuntos de características para el análisis de sentimientos en críticas de cine.

Referencias

1. Turney, P.: Thumbs Up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Philadelphia, Pennsylvania, USA, pp. 417–424 (2002) doi.org/10.3115/1073083.1073153
2. Martínez-Cámara, E., Martín-Valdivia, M. T., Ureña-López, L. A.: Opinion classification techniques applied to a Spanish corpus. In: Muñoz, R., Montoyo, A., Métails, E. (eds.) Natural Language Processing and Information Systems, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 169–176 (2011) doi: 10.1007/978-3-642-22327-3_17
3. Cruz, F. L., Troyano, J. A., Enriquez, F., Ortega, J.: Clasificación de documentos basada en la opinión: experimentos con un corpus de críticas de cine en español. *Procesamiento de Lenguaje Natural*, no. 41, pp. 73–80 (2008)
4. Bird, S., Klein, E., Loper, E.: Natural language processing with Python: Analyzing text with the natural language toolkit. O'Reilly Media, Inc. (2009)

5. Porter, M. F.: An algorithm for suffix stripping. *Program: Electronic Library and Information Systems*, vol. 14, no. 3, pp. 130–137 (1980) doi: 10.1108/eb046814
6. Gamon, M., Aue, A., Corston-Oliver, S., Ringger, E.: Pulse: Mining customer opinions from free text. In: Famili, A. F., Kok, J. N., Peña, J. M., Siebes, A., Feelders, A. (eds.) *Advances in Intelligent Data Analysis VI*, Springer Berlin Heidelberg, pp. 121–132 (2005) doi: 10.1007/11552253_12
7. Sari, Y., Stevenson, M., Vlachos, A.: Topic or Style? Exploring the most useful features for authorship attribution. In: *Proceedings of the 27th International Conference on Computational Linguistics*, Association for Computational Linguistics, pp. 343–353. (2018) doi: 10.17863/CAM.78746
8. Machova, K., Mach, M., Vasilko, M.: Comparison of machine learning and sentiment analysis in detection of suspicious online reviewers on different type of data. *Sensors*, vol. 22, no. 1, pp. 155 (2021) doi: 10.3390/s22010155
9. Pérez-Rosas, V., Banea, C., Mihalcea, R.: Learning sentiment lexicons in Spanish. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, vol. 12, pp. 73 (2012)
10. Tellez, E. S., Miranda-Jiménez, S., Graff, M., Moctezuma, D., Siordia, O. S., Villaseñor, E. A.: A case study of Spanish text transformations for twitter sentiment analysis. *Expert Systems with Applications*, vol. 81, pp. 457–471 (2017) doi: 10.1016/j.eswa.2017.03.071
11. Qader, W. A., Ameen, M. M., Ahmed, B. I.: An overview of bag of words: Importance, implementation, applications, and challenges. In: *2019 International Engineering Conference (IEC)*, pp. 200–2004 (2019) doi: 10.1109/IEC47844.2019.8950616
12. Sebastiani, F.: Machine learning in automated text categorization. *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47 (2002) doi: 10.1145/505282.505283
13. Priya, B.: Sentiment analysis for online movie reviews using SVM. *International Journal of Scientific research and Review*, vol. 7, no. 5, pp. 804–811(2019)
14. Pla, F., Hurtado, L. F.: Sentiment analysis in twitter for Spanish. In: Métais, E., Roche, M., and Teisseire, M. (eds.) *Natural Language Processing and Information Systems*, Springer International Publishing, Cham, pp. 208–213 (2014) doi: 10.1007/978-3-319-07983-7_27
15. Maulana, R., Rahayuningsih, P. A., Irmayani, W., Saputra, D., Jayanti, W. E.: Improved accuracy of sentiment analysis movie review using support vector machine based information gain. *Journal of Physics: Conference Series*, vol. 1641 (2020). doi: 10.1088/1742-6596/1641/1/012060