

Minería de opiniones en el comercio electrónico usando n-gramas y algoritmos de aprendizaje automático

Francisco Antonio Castillo Velásquez, Maricarmen Rico Galeana,
Nancy Sánchez Aguilar, José Marcos Zea Pérez,
María del Consuelo Patricia Torres Falcón

Universidad Politécnica de Querétaro,
México

{francisco.castillo, maricarmen.rico,
nancy.sanchez}@upq.edu.mx
{marcos.zea, consuelo.torres}@upq.mx

Resumen. Actualmente hay 320 millones de usuarios de comercio electrónico en América Latina y el Caribe y se espera alcanzar 400 millones en el 2028. Solo las ventas netas de e-Commerce de las 100 principales tiendas online mexicanas representaron alrededor de US\$33,000 millones en 2023. Gran parte de estas transacciones involucran una interacción con el cliente, entre las cuales destaca la posibilidad de redactar comentarios u opiniones sobre el producto o servicio adquirido. Estos datos representan información valiosa para las empresas, ya que con ella pueden tomar decisiones de venta y estrategias con los clientes. Conocer qué y cuántas opiniones de un producto o servicio fueron negativas o positivas resulta prácticamente imposible si se desea hacerlo de forma manual. La minería de opiniones (o análisis de sentimientos) tiene como tarea automatizar la asignación de etiquetas sobre una gran cantidad de textos. El presente trabajo de investigación tuvo como objetivo aplicar la minería de opiniones a un corpus de opiniones de transacciones de e-Commerce usando la técnica de n-gramas y algoritmos de aprendizaje automático. La metodología propone la compilación del corpus, obtención de características basadas en n-gramas, aplicación de algoritmos de aprendizaje automático, generación del modelo computacional y obtención de cifras de clasificación. Los resultados fueron muy alentadores, alcanzando un 88% de clasificación correcta con algoritmos de probabilidad, superando inclusive a otras propuestas del estado del arte, lo que sugiere la factibilidad de aplicación del modelo, con la ventaja adicional de su simplicidad e independencia del lenguaje.

Palabras clave: N-gramas, aprendizaje-automático, comercio-electrónico, análisis-de-sentimientos, minería-de-opiniones.

Opinion Mining in E-Commerce using n-Grams and Machine-Learning Algorithms

Abstract. There are currently 320 million e-commerce users in Latin America and the Caribbean, and it is expected to reach 400 million in 2028. The net e-

Commerce sales of the top 100 Mexican online stores alone represented around US\$33 billion in 2023. Great part of these transactions involves an interaction with the customer, among which the possibility of writing comments or opinions about the product or service purchased stands out. This data represents valuable information for companies, since with it they can make sales decisions and strategies with customers. Knowing which and how many opinions of a product or service were negative or positive is practically impossible if we want to do it manually. Opinion mining (or sentiment analysis) has the task of automating the assignment of labels to many texts. The objective of this research work was to apply opinion mining to a corpus of opinions of e-Commerce transactions using the n-gram technique and machine learning algorithms. The methodology proposes the compilation of the corpus, obtaining characteristics based on n-grams, application of machine learning algorithms, generation of the computational model and obtaining a classification ranking. Very promising results were obtained, reaching 88% correct classification with probability algorithms, even surpassing other state-of-the-art proposals, which suggests the feasibility of applying the model, with the additional advantage of its simplicity and language independence.

Keywords: N-grams, machine-learning, e-commerce, sentiment-analysis, opinion mining.

1. Introducción

La práctica de adquirir bienes y servicios a través de canales electrónicos se está generalizando y ganando cada vez más aceptación entre los consumidores. El comercio electrónico ha hecho posible que los consumidores compren sin salir de la comodidad de sus hogares y con disponibilidad de 24 horas. Es así como esta industria ha experimentado un crecimiento exponencial en el número de sitios web de comercio electrónico [1]. Según la Asociación Mexicana de Ventas Online, las ventas de comercio electrónico en México superaron los US\$33,000 millones en 2023.

De esta forma, el e-Commerce mexicano se posiciona como el segundo mercado online más importante en América Latina, superado únicamente por Brasil. A pesar de estas cifras, la tasa de abandono de las compras online representa un 70%. Para evitar que un futuro cliente desista de la compra se recomienda optimizar el proceso de checkout, ofrecer la posibilidad de finalizar la compra como invitado (muchas personas desisten cuando tienen que crear una cuenta) y tener referencias a través de opiniones de antiguos compradores [2]. El análisis de sentimientos (ADS), también conocido como análisis o minería de opiniones, le proporciona al cliente información acerca del producto o servicio antes de comprarlo

Por lo general, esta información es textual y recae en categorías de comentarios positivos o negativos. Esta información también es importante para las empresas y los especialistas en marketing para así hacer adecuaciones o mejoras a lo que están ofreciendo, satisfaciendo con ello las necesidades del cliente e incrementando sus ventas. Las empresas están integrando cada vez más el ADS en sus aplicaciones y sistemas. Esto se aplica a la atención al cliente, el seguimiento de la reputación en línea, la toma de decisiones basada en datos y mucho más. Los modelos de ADS son una parte importante del Procesamiento del Lenguaje Natural (PLN) y se utilizan para determinar

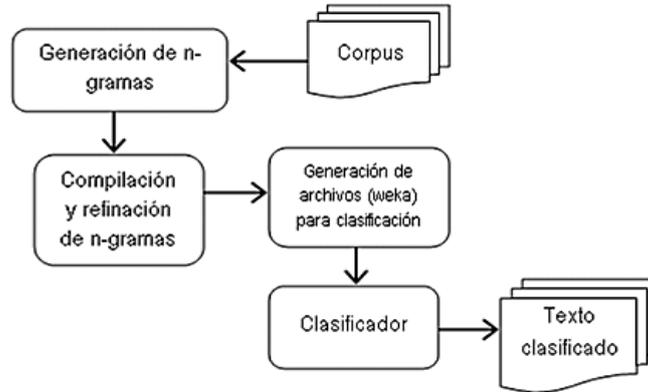


Fig. 1. Modelo de trabajo propuesto.



“Te baja el rizo horrible en cuanto lo aplicas, por el precio no vale la pena.”

Fig. 2. Análisis PoS en FreeLing para la frase.

la actitud o emoción expresada en un texto, ya sea positiva, negativa o neutral. Algunos de los modelos trabajados en este campo son los basados en reglas, en bolsas de palabras, en palabras preentrenadas, aprendizaje profundo, transferencia de aprendizaje y de aprendizaje automático.

Es con este último modelo en el que está basado el presente trabajo de investigación, pero con el soporte de la técnica de n-gramas tradicionales, lo que conlleva la ventaja de ser un modelo más simple, eficiente e independiente del lenguaje.

2. Trabajo relacionado

En los últimos años se ha incrementado el interés en el estudio y aplicación del aprendizaje automático aplicado a la minería de opiniones. A continuación, se detallan algunos trabajos relevantes y actualizados sobre este campo. En [3] se propone un enfoque sintáctico a la minería de opiniones para el español, específicamente aplicando un parser de dependencias para obtener la estructura sintáctica, apoyado en diccionarios semánticos. Aunado a ello, estudiaron de forma especial las construcciones sintácticas de la negación, intensificación y cláusulas subordinadas adversativas.

La investigación de [4] propone un enfoque basado en aspectos y la combinación de dos modelos de aprendizaje profundo. Su trabajo fue orientado a opiniones en español del sector restaurantero, usando un corpus de aproximadamente 2000 opiniones. Un estudio más profundo sobre el análisis basado en aspectos se encuentra en [10]. Una revisión más detallada la llevó a cabo en [15], probando diversos clasificadores supervisados, entre ellos CNN multi-canal, LSTM (BERT) y RMDL (secuencial con cinco capas), alcanzado tasas de validación entre 78% y 80%.

El dataset utilizado fue de 34,660 comentarios en Amazon (en inglés) etiquetados como positivos o negativos. Un trabajo muy interesante lo encontramos en [13], donde hicieron uso de métodos de aprendizaje profundo (RNN, LSTM, BLSTM, CNN) y de

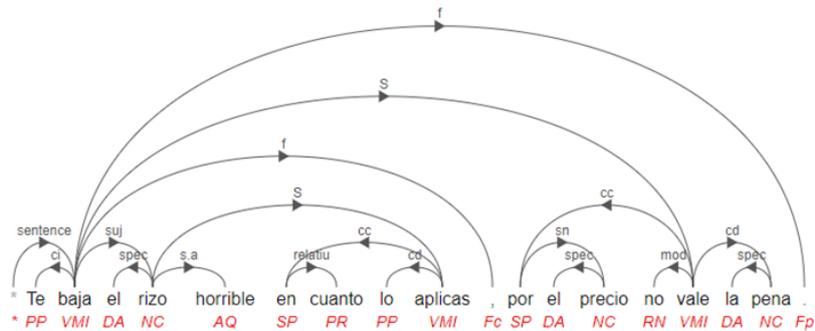


Fig. 3. Árbol de dependencias sintácticas para una oración del corpus.

aprendizaje automático (RF, MNB, SVM) aplicados a un dataset de 100 mil opiniones de comercio electrónico en tres lenguajes (inglés, turco y árabe) con tres categorías de sentimiento (positivo, negativo y neutral). Se usaron modelos pre-entrenados, entre ellos bert-base-multilingua-cased y xlm-roberta-base, obteniendo RNN el mejor desempeño para el inglés (89.5%), mientras que SVM para el árabe un 91.3%. Al tomar en cuenta los 3 lenguajes, SVM alcanzó los mayores valores de precisión. Por último, [9] representa un trabajo con una idea similar a la propuesta en esta investigación, con un enfoque sobre técnicas de bolsas y n-gramas de palabras como extractores de características y SVM, NB y KNN como clasificadores.

3. Metodología

En este apartado se explica el modelo propuesto de trabajo, iniciando con la compilación del corpus, el proceso de obtención de los n-gramas y finalizando con la tarea de clasificación. Los experimentos consistieron en la generación de n-gramas (mediante una herramienta libre de extracción); su compilación, que no es más que el almacenamiento dinámico de los n-gramas únicos (haciendo uso de estructuras matriciales y la implementación de un algoritmo para su manipulación); la refinación, que es la puesta a punto de los n-gramas únicos para que los caracteres no reconocidos por Weka sean detectados y cambiados; la generación de archivos de entrenamiento y clasificación; la aplicación de estadísticas de frecuencias y la aplicación de procesos de clasificación con los modelos de SVM (SMO), NB y J48 para distintas cantidades de n-gramas. Usamos un baseline de etiquetas de partes de oración (PoS). El modelo del trabajo se resume en la figura 1, donde el paso inicial (generación del corpus) se dividió en dos conjuntos: los comentarios originales y las dependencias generadas por el parser sintáctico de Freeling.

3.1. Compilación del corpus

La mayoría de los corpus disponibles en este campo de opiniones se encuentran en inglés, por lo que consideramos pertinente generar uno para el lenguaje español para futuros trabajos y para que esté disponible para la comunidad interesada en el tema. A través de herramientas de Web Scrapping aplicado a uno de los sitios más representativos de comercio electrónico (Amazon), se obtuvo un total de 10 mil

Tabla 1. Gramas de nivel 2,3,4 para las rutas de dependencias de la Fig. 2.

bigramas	trigramas	cuatrigramas
subj→spec	subj→S→cd	S→cc→sn→spec
subj→sa	subj→S→cc	subj→S→cc→relatiu
subj→S	S→cd→spec	
S→cd (2)	S→cc→sn	
S→cc		
S→mod		

Tabla 2. Gramas de nivel 2,3,4 para las partes de oración de la Fig. 2.

bigramas	trigramas	cuatrigramas
PP2CS00→VMIP350	PP2CS00→VMIP350→DA0MS0	PP2CS00→VMIP350→DA0MS0→NCMS000
VMIP350→DA0MS0	VMIP350→DA0MS0→NCMS000	VMIP350→DA0MS0→NCMS000→AQ0CS00
DA0MS0→NCMS000	DA0MS0→NCMS000→AQ0CS00	DA0MS0→NCMS000→AQ0CS00→SP
NCMS000→AQ0CS00	NCMS000→AQ0CS00→SP	NCMS000→AQ0CS00→SP→PR0MS00
...

comentarios (5 mil positivos y 5 mil negativos) vertidos por clientes registrados. Algunos ejemplos de estos textos son los siguientes:

Había oferta y aproveché, el producto es el que siempre he utilizado y me es familiar y agradable el sabor. (positivo).

Precio accesible, me llevo en buenas condiciones y si lo volvería a comprar. (positivo).

No llevaba ni un año y empezó a hacer un ruido extraño, ahora no sirve. (negativo).

Te baja el rizo horrible en cuanto lo aplicas, por el precio no vale la pena. (negativo).

Es importante mencionar que los comentarios no fueron editados, es decir, aquellos que presentan errores ortográficos u otro tipo de omisiones, se dejaron en su forma original.

3.2. Generación de n-gramas

Los n-gramas representan una técnica fundamental en el PLN. Se utilizan para analizar secuencias de palabras o caracteres en un texto al dividirlos en unidades contiguas de n elementos, que generalmente son palabras o caracteres. Estos elementos pueden ser considerados como "tokens" individuales y se utilizan para capturar información sobre la estructura y el contenido del texto. Tomando como ejemplo parte de la segunda opinión arriba mencionada ("*Precio accesible,*") podemos generar los bigramas de caracteres *Pr, re, ec, ci, io, o_, _a, ac, cc, ce, es, si, ib, bl, le y e,*.

De igual forma podemos generar los trigramas *Pre, rec, eci, cio, io_, o_a, _ac, cce, ces, esi, sib, ibl y ble.* Nuestro modelo hace un análisis estadístico de las apariciones de los gramas en cada una de las oraciones. Como parte innovadora, previamente se

Tabla 3. Baseline de la clasificación con gramas de PoS.

n-gramas	Clasificador	Tamaño del n-grama			
		2	3	4	5
500	NB	0.55	0.58	0.72	0.71
	SVM	0.70	0.72	0.74	0.71
	J48	0.70	0.71	0.68	0.68
1000	NB	0.69	0.71	0.70	0.69
	SVM	0.67	0.68	0.71	0.70
	J48	0.66	0.64	0.65	0.67
2000	NB	0.67	0.69	0.70	0.68
	SVM	0.64	0.66	0.65	0.68
	J48	0.61	0.63	0.64	0.63

Tabla 4. Resultados de la clasificación con gramas sintácticos.

n-gramas	Clasificador	Tamaño del n-grama			
		2	3	4	5
500	NBM	0.69	0.69	0.88	0.82
	SVM	0.72	0.72	0.75	0.74
	J48	0.71	0.71	0.71	0.70
1000	NBM	0.70	0.70	0.84	0.74
	SVM	0.70	0.70	0.73	0.69
	J48	0.63	0.63	0.65	0.64
2000	NBM	0.66	0.66	0.79	0.69
	SVM	0.64	0.64	0.68	0.68
	J48	0.60	0.60	0.61	0.60

obtendrá información sintáctica de las oraciones que representan los comentarios, esto para que los gramas incluyan información sobre la estructura de las oraciones.

La implementación de la generación de archivos Weka es un proceso semi-automático a través de un programa en Java, el cual está disponible para la comunidad académica e investigadora, junto con el corpus compilado y los archivos Weka generados.

3.3. Proceso de clasificación

Weka es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Proporciona herramientas para el pre-procesamiento, clasificación, regresión, clustering y visualización de datos. Es un software de código abierto basado en los términos de GNU-GPL. Cada registro de la sección de datos (@data) representa un comentario y cada valor numérico representa las ocurrencias de un grama en

particular en ese texto. El último valor, que en nuestro ejemplo no es numérico, representa la medida de la opinión (positiva o negativa).

El proceso de entrenamiento y clasificación también se llevó a cabo con el software Weka, que proporciona una diversidad de métodos. En particular, fueron usados los clasificadores NB, Optimización Mínima Secuencial (SMO - Support Vector Machines) y árboles de decisión (J48), ya que estos han mostrado buenos resultados en otros trabajos de investigación, como en [13].

En algunas primeras pruebas se llegó a utilizar un archivo de stopwords para el español, pero se notó que no aportó ninguna ventaja en los resultados, de hecho, fue todo lo contrario. Esto nos sugiere que todas las palabras aportan algo de valor en el proceso de clasificación.

3.4. Generación del baseline

El etiquetado de PoS es la tarea dentro del PLN que asigna una etiqueta de categoría gramatical a cada una de sus palabras. Ejemplos de estas categorías son los adjetivos, sustantivos y determinantes. A diferencia de los parsers sintácticos profundos (dependencias, constituyentes), la tarea de un parser sintáctico superficial, como lo es para las PosTags, requiere menos tiempo.

Este proceso se llevó a cabo con la librería de Freeling. Retomando la misma idea de lo mostrado en la tabla 1, podemos aplicarla para obtener gramas para categorías superficiales de PoS, con lo cual obtendríamos el resultado de la tabla 2. De esta manera se logró generar un baseline de trabajo donde notamos que el mayor porcentaje de clasificación correcta fue de 74%, para 500 cuatrigamas y el algoritmo SVM.

4. Resultados experimentales

Los experimentos fueron desarrollados sobre los datos de un corpus de minería de opiniones. Con el corpus de 1000 comentarios se hicieron diversas pruebas, siempre con un 60% de los datos para entrenamiento y el restante para clasificación. En la tabla 3 se muestran los resultados obtenidos. La tarea de clasificación consiste en seleccionar características para construir el modelo de espacio de vectores, algoritmos supervisados de entrenamiento y clasificación – decidir a qué clase pertenece el texto – en nuestro modelo de espacio de vectores.

En este trabajo presentamos resultados para tres clasificadores: SVM (SMO), NB y J48. En los resultados mostrados es interesante notar que con los clasificadores NB y SVM se obtiene una mayor exactitud para cuatrigamas, alcanzando un 88% con el primero de ellos; de hecho, los mejores resultados fueron para NB en prácticamente todas las pruebas. Es importante mencionar que se usó la medida TFD-ID.

Analizando más a profundidad la tabla 4 podemos observar que el clasificador con el porcentaje más alto en todas las pruebas de reseñas fue NaiveBayesMultinomial, logrando hasta un 88% con 500 gramas. Es una cifra similar a la mencionada en [9], pero las diferencias de los estudios es que ellos usaron gramas de palabras y el clasificador NaiveBayes y en nuestro trabajo se usaron gramas sintácticos de dependencias y el clasificador NaiveBayes Multinomial.

Comparado con el baseline, nuestra propuesta mejora en un 10% de clasificación correcta, aunque hay que destacar que SVM logra el porcentaje más alto en las pruebas del baseline. De igual forma, este trabajo corrobora lo mencionado en [11], en el sentido de que el clasificador Naïve Bayes funciona adecuadamente, incluso en situaciones de contexto pesado.

5. Conclusiones

Los resultados de los experimentos demuestran la factibilidad de usar modelos computacionales simples para la tarea de ADS, lo que redundaría en un menor tiempo computacional de procesamiento. La contribución más importante es la simplicidad y la generalidad del modelo (pueden ser aplicados a cualquier lenguaje).

En el comercio electrónico, la minería de opiniones es valiosa para comprender la satisfacción del cliente, identificar problemas con productos o servicios, y tomar decisiones informadas para mejorar la experiencia del cliente. Los ADS basados en n-gramas y algoritmos de aprendizaje automático pueden proporcionar información valiosa sobre tendencias y patrones en las opiniones de los clientes, lo que puede ayudar a las empresas a tomar medidas proactivas, como ajustar precios, mejorar la calidad del producto o el servicio, o desarrollar estrategias de marketing específicas.

Sin embargo, la idea aquí propuesta solo divide a los sentimientos en dos categorías (positivo y negativo), lo que quedaría en desventaja en áreas que necesitaran de más detalles, por lo que un trabajo futuro es trabajar con estas nuevas categorías de clasificación del texto (por ejemplo, “muy negativo”, “muy positivo”), ampliar el corpus de trabajo y analizar la generalización a otros dominios y lenguajes.

Referencias

1. Ayodeji, O.G., Kumar, V.: E-commerce Research Models: A Systematic Review and Identification of the Determinants to Success. *International Journal of Business Information Systems* (2020). DOI: 10.1504/ijbis.2020.10044532.
2. Statista: El comercio electrónico en México. Statista Research Department. <http://es.statista.com/temas/6370/el-comercio-electronico-en-mexico/> (2024)
3. Singh, U., Saraswat, A., Azad, H.K., Abhishek, K., Shitharth, S.: Towards Improving e-Commerce Customer Review Analysis for Sentiment Detection. *Scientific Reports*, vol. 12, no. 1 (2022). DOI: 10.1038/s41598-022-26432-3.
4. Martínez-Seis, B.C., Pichardo-Lagunas, O., Miranda, S., Perez-Cazares, I.J., Rodriguez-Gonzalez, J.A.: Deep Learning Approach for Aspect-based Sentiment Analysis of Restaurants Reviews in Spanish. *Computación y Sistemas*, vol. 26, no. 2, pp. 899–908 (2022). DOI: 10.13053/cys-26-2-4258.
5. Aguilar, D., Sidorov, G., Batyrshin, I.: Caso de estudio de análisis de sentimientos en Twitter: Tratado de libre comercio de América del Norte. *Research in Computing Science*, vol. 147, no. 5, pp. 357–365 (2018). DOI: 10.13053/rcs-147-5-27.
6. Hernández, A., Ramírez, G., Villatoro, E.: Un método para el análisis de sentimientos bajo un enfoque supervisado usando recursos léxicos. *Research in Computing Science*, vol. 147, no. 6, pp. 221–233 (2018)

7. Henríquez, C., Guzmán, J., Salcedo, D.: Minería de opiniones basado en la adaptación al español de ANEW sobre opiniones acerca de hoteles. *Procesamiento del Lenguaje Natural*, vol. 56, pp. 25–32 (2016)
8. Moreno-Sandoval, L.G., Pomares-Quimbaya, A., Cruz-Gutiérrez, C.E., García-Pachón, J.F., Vanegas-Ramírez, D.F.: Comparación de métodos de análisis de sentimientos en comunidades de habla hispana. In: *Encuentro Internacional de Educación en Ingeniería Asociación Colombiana de Facultades de Ingeniería*, pp. 1–12 (2022). DOI: 10.26507/paper.2367.
9. Ali, M., Yasmine, F., Mushtaq, H., Sarwar, A., Idrees, A., Tabassum, S., Hayyat, B., Rehman, K.U.: Customer Opinion Mining by Comments Classification Using Machine Learning. *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5 (2021). DOI: 10.14569/ijacsa.2021.0120547.
10. Nazir, A., Rao, Y., Wu, L., Sun, L.: Issues and Challenges of Aspect-based Sentiment Analysis: A Comprehensive Survey. In: *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 845–863 (2022). DOI: 10.1109/taffc.2020.2970399.
11. López-Condori, J.J., Gonzales-Saji, F.O.: Análisis de sentimiento de comentarios en español en google play store usando BERT. *Ingeniare, Revista Chilena de Ingeniería*, vol. 29, no. 3, pp. 557–563 (2021). DOI: 10.4067/s0718-33052021000300557.
12. Sinnasamy, T., Sjaif, N.N.A.: A Survey on Sentiment Analysis Approaches in e-commerce. In: *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 10, pp. 674–679 (2021). DOI: 10.14569/ijacsa.2021.0121074.
13. Savci, P., Das, B.: Prediction of the Customers’ Interests Using Sentiment Analysis in e-commerce Data for Comparison of Arabic, English, and Turkish Languages. *Journal of King Saud University-Computer and Information Sciences*, vol. 35, no. 3, pp. 227–237 (2023). DOI: 10.1016/j.jksuci.2023.02.017.
14. Agarwal, A., Biadsy, F., McKeown, K.R.: Contextual Phrase-level Polarity Analysis Using Lexical Affect Scoring and Syntactic n-grams. *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics European Chapter of the Association for Computational Linguistics*, pp. 24–32 (2009).
15. Singh, U., Saraswat, A., Azad, H.K., Abhishek, K., Shitharth, S.: Towards Improving e-commerce Customer Review Analysis for Sentiment Detection. *Scientific Reports*, vol. 12, no. 1 (2022). DOI: 10.1038/s41598-022-26432-3.
16. Vilares, D., Alonso, M.A., Gómez-Rodríguez, C.: A Syntactic Approach for Opinion Mining on Spanish Reviews. *Natural Language Engineering*, vol. 21, no. 1, pp. 139–163 (2013). DOI: 10.1017/s1351324913000181.
17. Yang, L., Li, Y., Wang, J., Sherratt, R.S.: Sentiment Analysis for e-commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. *IEEE Access*, vol. 8, pp. 23522–23530 (2020). DOI: 10.1109/access.2020.2969854.