

## **Análisis de sentimiento de comentarios en español en Google Play Store usando BERT**

*Sentiment analysis for spanish reviews on Google Play Store using BERT*

Juan José López Condori<sup>1\*</sup>      Freddy Orlando Gonzales Saji<sup>2</sup>

Recibido 6 de Enero de 2021, aceptado 10 de Mayo de 2021

*Received: January 6, 2021      Accepted: May 10, 2021*

### **RESUMEN**

Las tiendas de aplicaciones móviles como Google Play Store tienen una amplia gama de aplicaciones dirigidas hacia diferentes clientes. Estas plataformas digitales proporcionan un mecanismo de calificación para que los usuarios califiquen las aplicaciones alojadas y dejen sus comentarios. Los comentarios de los usuarios contienen información valiosa que tiene impacto inevitable sobre el éxito de una aplicación. Debido a la gran cantidad de datos generados en esta plataforma, el procesamiento del lenguaje natural se ha vuelto más frecuente. El análisis de sentimientos es una de las aplicaciones de clasificación de texto que se puede utilizar para analizar estos comentarios. El análisis de sentimientos se ha llevado a cabo en inglés, mientras que, en otros idiomas como el español, no ha habido muchos intentos.

En este trabajo se usa un modelo pre-entrenado BERT para el análisis de sentimiento de comentarios en español de la tienda de aplicaciones de Google Play Store, los resultados experimentales indican que después del pre-procesamiento adecuado puede mostrar resultados prometedores en nuestro conjunto de datos en español, utilizando el modelo BERT se puede alcanzar una precisión de 0.81 en promedio, incluso con la limitación de datos.

Palabras clave: Google Play Store, análisis de Sentimiento, BERT.

### **ABSTRACT**

*Mobile application stores, such as Google Play Store, have a wide range of applications (apps) target to different customers. These digital platforms provide a rating mechanism for users to rate hosted apps and leave their reviews. User feedback contains valuable information that has an inevitable impact on the success of an application. Due to the large amount of data generated on this platform, natural language processing techniques have become applied more frequently. Sentiment analysis is a text classification task that can be used to analyze these reviews. Sentiment analysis has been conducted mainly for the English language, while in other languages like Spanish, there are just few attempts.*

*In this work, we use a pre-trained BERT model to perform sentiment analysis for Spanish reviews on Google Play Store. Experiments using the BERT model show that after proper preprocessing steps it can achieve promising results in a Spanish dataset. An accuracy of 0.81 can be reached on average, even with limited data.*

*Keywords: Google play store, sentiment analysis, BERT.*

---

<sup>1</sup> Universidad Nacional de San Agustín de Arequipa. Ciencia de la Computación. Arequipa, Perú. E-mail: jlopezco@unsa.edu.pe

<sup>2</sup> Universidad Nacional de San Agustín de Arequipa. Ingeniería de Sistemas. Arequipa, Perú. E-mail: fgonzales@unsa.edu.pe

\* Autor de correspondencia: jlopezco@unsa.edu.pe

## INTRODUCCIÓN

La tecnología está creciendo rápidamente, así como las aplicaciones en teléfonos inteligentes basados en Android, Windows o sistemas operativos iOS. El rápido crecimiento de la tecnología ha generado muchos datos de información valiosa que será útil cuando los datos se procesen y usen [1].

Una de las plataformas que acumula mucha información es Google Play, es un servicio de contenido digital que pertenece a Google que consiste en tiendas de productos en línea de música/canciones, libros, aplicaciones, juegos, o un reproductor multimedia basado en la nube. Google Play se puede acceder a través de la web, aplicaciones de Android (Play Store) y Google TV. Google Play Store se ha convertido en un mercado inmensamente competitivo para desarrolladores de aplicaciones debido al rápido aumento en el número de aplicaciones móviles y usuarios de teléfonos inteligentes [2]. Cada página de aplicación en Google Play Store tiene una sección de comentarios donde los usuarios transmiten sus críticas constructivas. Una medida del éxito del servicio puede verse en los comentarios de los consumidores.

La cantidad de comentarios publicados en esta plataforma es demasiado grande para ser procesado manualmente. Por lo tanto, se necesita un método o técnicas especiales que pueden categorizar las revisiones automáticamente. El análisis de sentimientos y la minería de opiniones es el campo de estudio que analiza las opiniones de las personas, sentimientos, evaluaciones, actitudes y emociones del lenguaje escrito [3]. Es una de las áreas de investigación más activas en el procesamiento del lenguaje natural y también se estudia ampliamente en minería de datos, minería web, y minería de texto.

Los resultados de las revisiones de los consumidores pueden proporcionar resultados concisos para facilitar a los nuevos usuarios la toma de decisiones [4] [5]. Si se obtiene una mayor investigación del comentario, se generará un sentimiento que, si se recopila, podrá realizar conclusiones como la experiencia del usuario de una determinada aplicación.

La tarea básica en el análisis de sentimientos es clasificar la polaridad del texto presente en los documentos, oraciones u opiniones. Uno de los

mayores desafíos en el procesamiento del lenguaje natural (PLN) es la escasez de datos de capacitación. Debido a que PLN es un campo diversificado con muchas tareas distintas, la mayoría de los conjuntos de datos específicos de tareas contienen solo unos pocos miles o unos cientos de miles de ejemplos de entrenamiento etiquetados por humanos [6].

Existen estudios que analizan distintos enfoques del análisis de sentimiento. En [7] describe la importancia de la fuente de datos y concatenación de modelos de aprendizaje profundo que utilizan la capa funcional de Keras para combinar vectores de texto y columnas numéricas de datos para realizar clasificación, también podemos ver que en [8] se discuten las etapas para hacer el proceso de análisis de sentimientos para comentarios comenzando desde la etapa de preprocesamiento, hasta la etapa análisis de sentimientos con el clasificador Naive Bayes y el algoritmo Optimización por enjambre de partículas (PSO, por sus siglas en inglés Particle Swarm Optimization), otros estudios aplican métodos como árboles de decisión y Support Vector Machine(SVM) para el análisis de sentimiento de los comentarios positivos o negativos [9]. Existen estudios que se encargan de analizar los sentimientos de los comentarios de una aplicación en particular para mejorar el servicio de los usuarios [10]. Investigaciones utilizan el modelo BERT también para solucionar subtareas del análisis de sentimiento como es el caso del análisis de sentimientos basado en aspectos, que tiene como objetivo identificar la polaridad de opinión detallada hacia un aspecto específico, como se detalla en [11] donde se logran nuevos resultados de vanguardia.

Los modelos de PLN basados en el aprendizaje ven los beneficios de cantidades mucho mayores de datos, mejorando cuando se entrena en millones o miles de millones de ejemplos de entrenamiento anotados. Para ayudar a cerrar esta brecha en los datos, los investigadores han desarrollado una variedad de técnicas para entrenar modelos de representación de lenguaje de propósito general utilizando la enorme cantidad de texto no anotado en la web (conocido como pre-entrenamiento) [12].

El modelo previamente entrenado se puede ajustar en tareas de PLN de datos pequeños como la respuesta a preguntas, completar oraciones y el análisis de sentimientos, lo que resulta en mejoras de precisión

sustanciales en comparación con el entrenamiento en estos conjuntos de datos desde cero [3].

Este documento muestra el proceso seguido para realizar con éxito el análisis de sentimientos de comentario en español obtenidos de Google Play Store mediante un modelo pre-entrenado BERT (Bidirectional Encoder Representations from Transformers) [12]. BERT está pre-entrenado en un gran corpus de texto sin etiquetar que incluye toda la Wikipedia (2,500 millones de palabras). BERT utiliza Transformer [13], un mecanismo de atención que aprende las relaciones contextuales entre palabras (o subpalabras) en un texto. Es un modelo profundamente bidireccional. Bidireccional significa que BERT aprende información tanto del lado izquierdo como del lado derecho del contexto de un token durante la fase de entrenamiento. La bidireccionalidad de un modelo es importante para comprender verdaderamente el significado de un lenguaje.

### PREPARACIÓN DE DATOS

Google Play Store representa diferentes tipos de aplicaciones de Android, la mayoría de las cuales son gratuitas. Así que toda la sección de comentarios es un reflejo de las opiniones de personas de diferentes gustos y mentalidades. Se construyó un scraper web (extraer información de páginas web de forma automatizada) para el propósito de recopilación de datos que proporcionó información crítica de revisión: nombre de usuario, comentario y calificación de una respectiva aplicación. La Figura 1 muestra el funcionamiento de los procesos seguidos para recopilar el conjunto de datos.

### Recopilación de datos

En el proceso de recopilación de datos se extrae los comentarios en español de las 15 aplicaciones más descargadas de la categoría educación de Google Play Store: Google Classroom, Cake Learn English for free, Brainly, Duolingo, Photomath, Mathway, Simply Piano by JoyTunes, Aprendo en casa, Lingokids, Symbolab-Math Solver, Kahoot, moodle, Pinkfong Baby Shark, Babbel - Learn Language y U-Dictionary: Oxford, para reconocer el identificador (ID) de las aplicaciones utilizamos AppAnnie.

Los datos se obtienen utilizando el paquete Google play Scraper que proporciona una API(Interfaz de Programación de Aplicaciones.) para Python, los datos obtenidos son guardado en un archivo con formato JSON (JavaScript Object Notation).

### Pre-procesamiento de datos

En esta investigación se trató con datos extraídos sin procesar, el objetivo principal era limpiarlos para evitar cualquier discrepancia en nuestros resultados.

El siguiente paso es preparar los datos que se han obtenido, quitando las palabras innecesarias que generan ruido en los datos obtenidos [14], se omitió las palabras con una longitud menor a 2, errores ortográficos, variaciones léxicas, emoticones, además de filtrar algunos comentarios que estaban escritos en otros idiomas que no sea el español, este proceso se realizó de manera automatizada utilizando la librería NLTK (Natural language toolkit) de Python.

### Normalización de datos para BERT

Una vez que los datos están preparados, se debe tener en consideración, que para poder entrenar

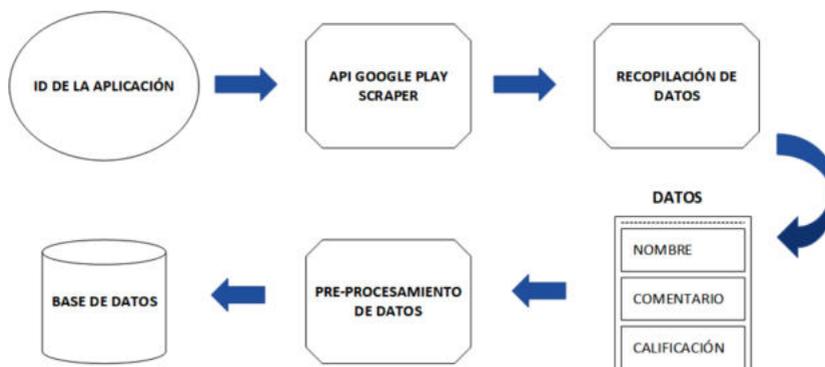


Figura 1. Proceso para la extracción de datos.

el modelo BERT con los datos obtenidos, lo siguiente:

- 1) Agregar tokens especiales al comienzo y final del comentario, estos tokens son los siguientes:
  - a. [SEP]: token especial que se agrega al final de cada oración.
  - b. [CLS]: Para las tareas de clasificación, debemos anteponer el [CLS] token especial al comienzo de cada oración.

Esta ficha tiene un significado especial. El modelo BERT consta de 12 capas de transformadores. Cada transformador toma una lista de incrustaciones de tokens y produce la misma cantidad de incrustaciones en la salida (pero con los valores de la característica cambiados). En la salida del transformador final (12vo), el clasificador solo utiliza la primera incrustación (correspondiente al token [CLS]).

En la Figura 2 se muestra cómo se insertan los tokens especiales en un comentario.

- 2) Rellenar y truncar los comentarios a una sola longitud constante.
 

Las oraciones en el conjunto de datos tienen diferentes longitudes, entonces BERT tiene dos restricciones:

  - a. Todas las oraciones deben tener la misma longitud por ello son rellenadas o truncadas en una sola longitud fija.

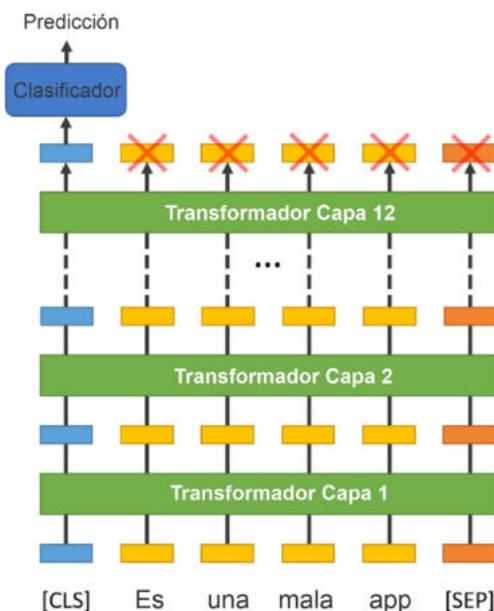


Figura 2. Proceso de inserción de tokens especiales.

- b. La longitud máxima de la oración es de 512 tokens.

El relleno se realiza con un token especial denominado: [PAD].

- 3) Diferenciar explícitamente los tokens de relleno:
 

La “Máscara de atención” es simplemente una matriz de 1s y 0s que indica qué tokens están rellenos y cuáles no. Esta máscara le dice al mecanismo de “auto-atención” en BERT que no incorpore estos tokens [PAD] en su interpretación de la oración.

El paso 2 y 3 se muestran en la Figura 3.

### CONSTRUCCIÓN DEL MODELO

Después de preparar los datos, se construye el modelo, se carga el modelo BERT pre entrenado y se ajusta, creando una nueva capa única para la tarea de análisis de sentimiento. En esta capa de deserción, se aplica la función Softmax para obtener las probabilidades predichas de nuestro modelo entrenado, a las salidas del modelo BERT y para ayudar a prevenir el sobreajuste. Finalmente convertimos las etiquetas en codificación de una sola vez, entonces se tiene un modelo entrenado en comentarios con etiqueta negativa, positiva o neutral y podemos predecir la etiqueta del nuevo comentario como negativa, positiva o neutral.

La Figura 4 muestra la arquitectura del modelo propuesto en esta investigación, para la tarea de análisis de sentimiento en comentarios en español de Google Play Store usando BERT.

### ENTRENAMIENTO

Para entrenar el modelo propuesto es necesario definir diferentes parámetros, para esta presente investigación, luego de pruebas y de la recomendación dada en [12], los parámetros para el ajuste fino son:

1. Tamaño del lote = 16
2. Tasa de aprendizaje =  $2e-5$
3. Longitud máxima de secuencia = 160
4. Épocas = 10

El rendimiento de la etapa de entrenamiento del modelo se muestra en la Figura 5, con los datos de entrenamiento y validación.

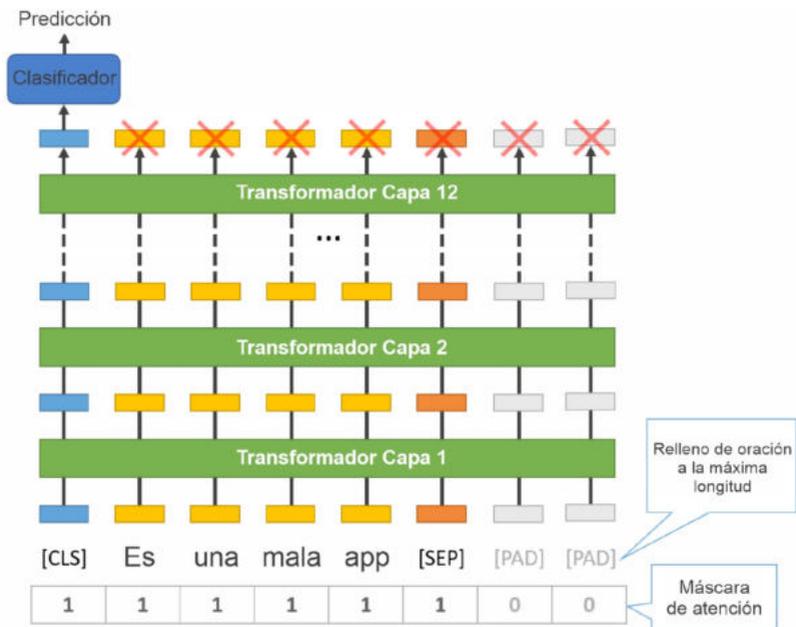


Figura 3. Paso 1 y 2 de normalización de datos para BERT.

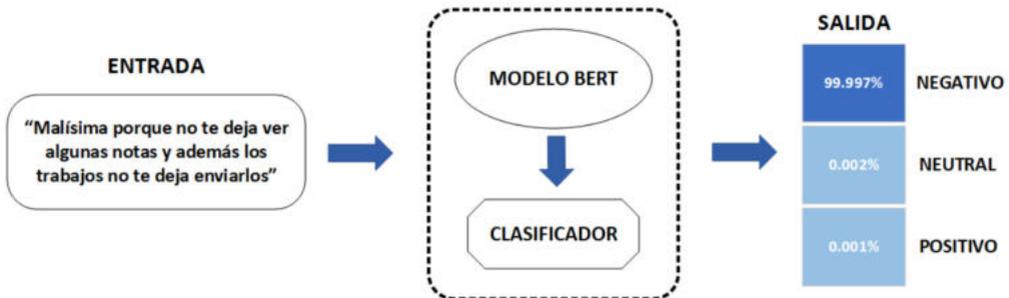


Figura 4. Arquitectura del modelo propuesto.

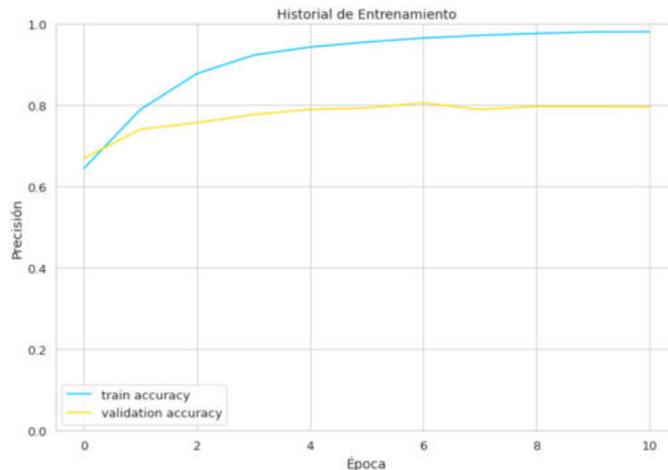


Figura 5. Entrenamiento del modelo propuesto.

## RESULTADOS

Nuestro conjunto de datos contiene 15985 comentarios de los cuales 5300 comentarios son negativos, 5300 son neutrales y 5385 son comentarios positivos en español, estos datos fueron divididos de la siguiente manera: 14386 fueron utilizados para el entrenamiento, 799 para la validación y 800 para las pruebas.

Se comparó el modelo propuesto con diferentes algoritmos: Clasificador Naïve Bayes utilizando TF-IDF (Term frequency – Inverse document frequency) para vectorizar los comentarios y SVM (Support Vector Machine). En la Tabla 1 se muestra los resultados obtenidos.

En la Tabla 1 se puede ver que, aunque el clasificador Naïve Bayes no proporciona los mejores resultados, ciertamente están cerca de los otros algoritmos, esto indica que el clasificador Naïve Bayes funciona bien incluso en situaciones de contexto pesado. El algoritmo SVM muestra mejores resultados que el clasificador Naïve Bayes. El texto a menudo es linealmente separable, por lo que este método obtuvo buenos resultados. El modelo propuesto supera a los métodos anteriores, porque, a diferencia de ellos, es el primer sistema no supervisado (el modelo aprende a adaptarse basada en las experiencias recogidas de los patrones de entrenamiento anteriores), contextual y profundamente bidireccional.

Tabla 1. Precisión en promedio del Modelo.

Método	Precisión	Recall	F1-SCORE
TF-IDF- Naïve Bayes	0.71	0.72	0.72
SVM	0.73	0.74	0.73
BERT	0.81	0.81	0.80

Tabla 2. Precisión por sentimiento.

Sentimiento	Precisión	RECALL	F1-SCORE
Negativo	0.86	0.85	0.85
Neutral	0.73	0.75	0.74
Positivo	0.83	0.81	0.82

En la Tabla 2 se muestra los resultados obtenidos de nuestro modelo por sentimiento.

En la Figura 6 se muestra la matriz de confusión, esto confirma que nuestro modelo tiene una alta precisión al identificar comentarios positivos y negativos, en el caso de los comentarios neutrales existe una cierta confusión por la existencia de ambigüedad.

## CONCLUSIONES

En este artículo, utilizamos el modelo pre entrenado BERT para el análisis de sentimiento en los

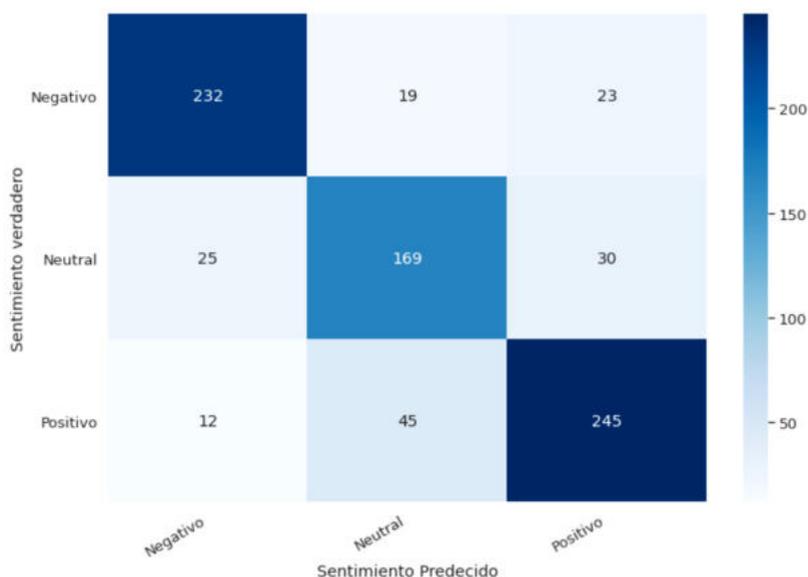


Figura 6. Matriz de confusión.

comentarios en español de Google Play Store, basado en los resultados de la precisión, el modelo BERT se comparó con el clasificador Naïve Bayes y la máquina de vectores de soporte (SVM), el modelo registró un nivel de precisión mejor y más alto al predecir el sentimiento en un comentario en español de Google Play Store.

BERT es un método de representaciones lingüísticas previas al entrenamiento que supera a los métodos anteriores porque es el primer sistema no supervisado, contextual y profundamente bidireccional para la formación previa de PNL.

Con los resultados de nuestros experimentos logramos resultados prometedores. En el futuro, investigaremos más información sobre BERT y buscaremos aplicarlo en un estudio de redes sociales para encontrar patrones de sentimientos de las personas frente a los sucesos que se han presentado respecto a la pandemia COVID-19.

## REFERENCIAS

- [1] S. Sulaiman, N. Rahim and A. Pranolo. “Generated rules for AIDS and e-learning classifier using rough set approach”. *International Journal of Advances in Intelligent Informatics*. Vol. 2, pp. 103-122. 2016. ISSN: 2548-3161. DOI:10.26555/ijain.v2i2.74.
- [2] E. Noei and K. Lyons. “A survey of utilizing user-reviews posted on google play store”. *Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*, pp. 54-63. 2019. DOI: 10.5555/3370272.3370278.
- [3] M. Farahani, M. Gharachorloo, M. Farahani & M. Manthouri. “ParsBERT: Transformer-based model for persian language understanding”, pp. 1-10. 2020.
- [4] M. Thawakar, S. Phiroj & M Atique. “An intelligent opinion mining for customer reviews”. *International Journal Of Computer Science And Applications*. Vol. 8, pp. 20-25. 2015. ISSN:0974-1011.
- [5] W. Songpan. “The analysis and prediction of customer review rating using opinion mining”. *IEEE Fifteenth International Conference on Software Engineering Research, Management and Applications (SERA)*. DOI: 10.1109/SERA.2017.7965709.
- [6] M. Soumik, S. Farhavi, E. Farzana, T. Sinha and M. Shafiul. “Employing machine learning techniques on sentiment analysis of google play store bangla reviews”. *Twenty-second International Conference on Computer and Information Technology (ICCIT)*. 2019. DOI: 10.1109/ICCIT48885.2019.9038348.
- [7] S. Venkatakrishnan, A. Kaushik and J. Kumar. “Sentiment analysis on google play store data using deep learning”. *Applications of Machine Learning*, pp. 103-122.2020. ISBN: 978-981-15-3356-3 .DOI:10.1007/978-981-15-3357-0\_2.
- [8] S. Saputra, D. Rosiyadi, W. Gata and S. Maulana. “Analisis sentimen E-Wallet pada google play menggunakan algoritma naive bayes berbasis particle swarm optimization”. *Jurnal RESTI*. Vol. 3, pp. 377-382.2019. ISSN: 2580-0760.
- [9] Nurfaizah, T. Hariguna and & I. Romadon. “The accuracy comparison of vector support machine and decision tree methods in sentiment analysis”. *Journal of Physics: Conference Series*. Vol. 1367.2019. DOI: 10.1088/1742-6596/1367/1/012025.
- [10] S. Wahyu, D. Saputra, Hasirun, R. Mega and G. Ramadhan. “Sentiment analysis for Go-Jek on google play store”. *IOP Conf. Series: Journal of Physics*. Vol. 1196. 2019. DOI: 10.1088/1742-6596/1196/1/012032.
- [11] C. Sun, L. Huang and X. Qiu. “Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence”. *ArXiv*.2019.
- [12] J. Devlin, M. Chang, K. Lee and K. Toutanova. “BERT: Pre-training of deep bidirectional transformers for language understanding”. *arXiv*. 2019.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, N. Gomez, L. Kaiser and I. Polosukhin. “Attention is all you need”. *ArXiv*.2017.
- [14] R. Amir, M. Talha, S. Aslam, M. Farhan, F. Ijaz and A. Karim. “Data scraping from google play store and visualization of its content for analytics”. *Second International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*, pp. 1-8. 2019. DOI: 10.1109/ICOMET.2019.8673523.