

Using of Transformers Models for Text Classification to Mobile Educational Applications

Anabel Pilicita Garrido and Enrique Barra Arias

Abstract— In Q2 2022, educational apps were the second most popular category on the Google Play Store, accounting for 10.47% of the apps available worldwide. This work explores the application of five BERT-based pre-trained models with the Transformers architecture to classify mobile educational applications. These five models are according to the knowledge field: bert-base-cased, bert-base-uncased, roberta-base, albert-base-v2 and distilbert-base-uncased. This study uses a dataset with educational apps of Google Play, this dataset was enriched with description and category because it lacked this information. In all models, a tokenizer and fine-tuning works were applied for training in the classification task. After training the data, the testing phase was performed in which the models had to go through four training epochs to obtain better results: roberta-base with 81% accuracy, bert-base-uncased with 79% accuracy, bert-base-cased obtained 80% accuracy, albert-base-v2 obtained 78% accuracy and distilbert-base-uncased obtained 76% accuracy.

Index Terms— Natural Language Processing, Multiclass Text Classification, Transformers, Bidirectional Encoder Representations from Transformers.

I. INTRODUCCIÓN

El procesamiento del lenguaje natural (en adelante, PLN) tuvo sus inicios en la década de los 50, siendo un campo de la inteligencia artificial aplicado a la lingüística [1]. El objetivo principal es alcanzar el procesamiento del lenguaje, similar al humano. Por lo que no deja de ser un proceso complejo donde interviene el léxico, la semántica, la sintaxis y la pragmática [1] [2]. El campo de aplicación del PLN es muy diverso y existen varias líneas de investigación como el análisis de sentimientos [3], la clasificación de preguntas [6] o la clasificación de textos [4] [5], este último consiste en categorizar o etiquetar textos en clases organizadas según su contenido.

Por otro lado, se cuenta con más de 6.300 millones de usuarios con teléfonos inteligentes en todo el mundo [7] y una de las plataformas más importante es Google Play con 2,65 millones de aplicaciones, juegos, libros, música o películas que pueden ser descargados en dispositivos Android [8][9]. A nivel mundial presenta una distribución por categorías como juegos, social, entretenimiento, productividad, salud, entre otras [10].

En el segundo trimestre de 2022, las aplicaciones móviles educativas se posicionaron como la segunda categoría más popular de Google Play con él con el 10,47% [11]. Esto significa que existe un gran interés por las aplicaciones educativas, ya que pueden ser utilizadas por estudiantes y docentes.

En este contexto, la búsqueda por categorías según su contenido en Google Play Store no está disponible a diferencia de otras plataformas como Google Imágenes que añade a los resultados de búsqueda categorías e insignias para identificar el tipo de contenido de una imagen [12]. Asimismo, YouTube implementa categorías de búsqueda para los videos según las preferencias de los usuarios.

Por lo tanto, la clasificación de aplicaciones móviles educativas sería importante porque ayudaría a los usuarios a encontrar aplicaciones de calidad y a los docentes a encontrar herramientas útiles para el proceso de enseñanza. En este sentido, el presente trabajo pretende abordar la clasificación de aplicaciones móviles educativas utilizando modelos basados en *Bidirectional Encoder Representations from Transformer* (en adelante, BERT). Frente a este desafío, se planteó los siguientes objetivos:

1. Consolidar un *dataset* de entrenamiento y prueba que permita clasificar aplicaciones móviles educativas.
2. Entrenar los modelos 5 modelos basados en BERT, en base a las descripciones de las aplicaciones móviles educativas, para su posterior clasificación multiclase según un área general de conocimiento.
3. Determinar el modelo lingüístico que con mayor exactitud y precisión asigna una categoría general de conocimiento a las aplicaciones móviles educativas en base a su descripción.

Este es el primer estudio dedicado a este tema y es una investigación que con sus resultados contribuye al estado del arte del PLN. Además, de ser un complemento para ahorrar tiempo de búsqueda inteligente en las plataformas, ordenando los resultados que van acorde a categorías en el ámbito educativo.

El artículo se organiza de la siguiente manera: Sección II presenta los trabajos relacionados; la Sección III la metodología; mientras que la Sección IV muestra los distintos resultados con cada modelo. Por último, la Sección V presenta las conclusiones y algunas reflexiones sobre trabajos futuros.

II. TRABAJOS RELACIONADOS

Por lo general, cualquier entorno que utilice texto es candidato para el PLN y el campo de aplicación es diverso. La

A. P. Garrido es estudiante de doctorado en la Universidad Politécnica de Madrid (a.pilicita@alumnos.upm.es).

Enrique Barra Arias, trabaja en la Universidad Politécnica de Madrid (enrique.barra@upm.es)

clasificación de texto etiqueta o categoriza textos según su contenido [16][17]. Actualmente, es una tarea de investigación abierta debido a la complejidad, a las aplicaciones en el mundo real (detección de spam, análisis de sentimiento, entre otros), o la aparición de nuevos algoritmos de aprendizaje [18]. De hecho, OBOE es un Framework propuesto para la clasificación de texto, donde los algoritmos y las técnicas son intercambiables, lo que significa que se puede personalizar [18].

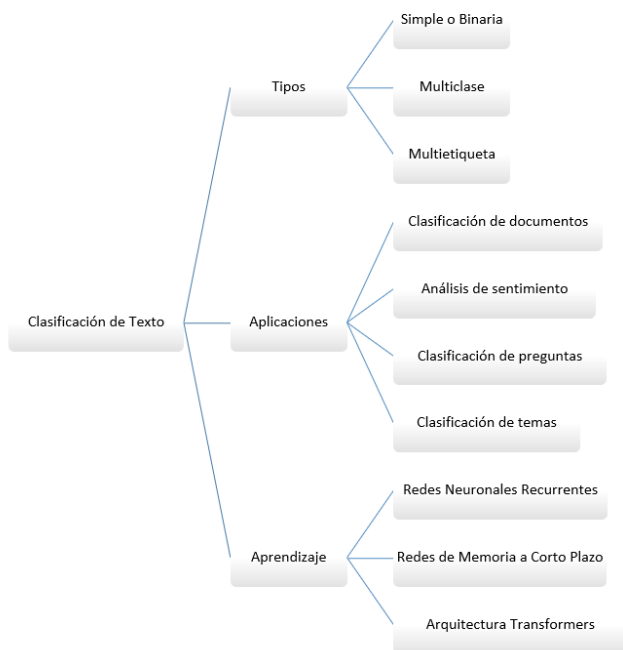


Fig. 1. Resumen de los temas relacionados a la tarea de clasificación de texto en el PLN.

La revisión de la literatura se realizó en base a la clasificación de texto como se muestra en la Figura 1. En primer lugar, el tipo de clasificación puede ser simple, multiclase o multietiqueta como se detalla a continuación [19]:

- la clasificación simple o binaria, utiliza solamente dos categorías para clasificar el texto.
- la clasificación multiclase tiene más de dos categorías, en este caso la cantidad de etiquetas aumenta y se asigna una etiqueta a un texto en particular.
- y la clasificación multietiqueta se diferencia de las anteriores al asignar más de una etiqueta a un texto determinado.

En base al campo de aplicación, encontramos líneas de investigación aplicadas a la clasificación de documentos [23], análisis de sentimientos [3], clasificación de preguntas [6] y clasificación de temas [5]. En ejemplos puntuales, se aplican clasificadores de texto para identificar noticias falsas relacionadas con la política [20]. En el ámbito de la salud, Khadhraoui [21] realiza su investigación para la detección de noticias falsas asociadas al COVID-19 mediante clasificadores multiclase y Gao [22] clasifica textos clínicos de pacientes. En el campo financiero, Arslan [23] realiza clasificación de documentos con un vocabulario adaptado al mundo financiero. De hecho, la clasificación de texto se aplica en distintos

campos, pero no deja de ser un proceso complejo donde intervienen modelos de aprendizaje automático.

Hace años, las Redes Neuronales Recurrentes (del inglés *Recurrent Neural Network*, abreviado RNN) tuvieron gran aceptación y se han aplicado a trabajos para capturar el estado del interlocutor y el estado contextual entre los enunciados [24]. Sin embargo, una desventaja es que estas retienen la información en la memoria del último estado, y a medida que aumenta el número de palabras, el tamaño del vector de contexto disminuye, provocando un problema al no asignar pesos adecuados para las predicciones del algoritmo [25]. Las redes neuronales evolucionaron a Redes de Memoria a Corto Plazo (del inglés *Long Short-Term Memory*, abreviado LSTM), que dependiendo de los pesos del algoritmo almacenan más o menos memoria, resolviendo el problema que presentan las RNN [25]. En este caso, el cálculo que se realiza es en serie sin aprovechar la computación en paralelo provocando que el costo computacional aumente en conjunto con el tiempo de entrenamiento.

En 2017, se propuso una nueva arquitectura con *Transformers*, aplicando computación en paralelo y aprendizaje por transferencia con mecanismo de autoatención [26]. Esta arquitectura revolucionó el campo del PLN al mejorar significativamente las predicciones en tareas como la clasificación de textos [27], la traducción automática [28], el resumen de texto [29], entre otras. Es una arquitectura de vanguardia que dispone de miles de modelos lingüísticos preentrenados para diferentes tareas del PLN disponibles en la plataforma de HuggingFace [30].

ELMo [31], GloVe [32], GTP de OpenAI [33] y BERT [34] son algunos de los modelos más conocidos. BERT, en comparación con los otros modelos citados se basa en un *Transformer* bidireccional multicapa. Se caracteriza por entrenar texto plano para tareas de predicción de palabras enmascaradas y de predicción de frases siguientes. Además, de una fase de "preentrenamiento" sin etiquetas, que le permite adquirir un conocimiento más detallado y proporciona mejores resultados en distintas tareas del PLN. Como elemento importante afronta el tamaño reducido del texto de entrada y la falta de vocabulario. Al ser un modelo que cuenta estas ventajas, no es extraño encontrar modelos basados en BERT tales como RoBERTa [35], ALBERT [14] y DistilBERT [15].

Las investigaciones relacionadas con modelos de clasificación multiclase con *Transformers* demuestran que los modelos BERT y DistilBERT superan alternativas de aprendizaje profundo en clasificación de noticias falsas [36]. Los estudios comparativos en un clasificador de reseñas falsas implementado con RoBERTa superó a los modelos ALBERT y DistilBERT [37]. Asimismo, en estudios para clasificar emociones [38], y para clasificar noticias [39] el modelo RoBERTa presenta mejores resultados en comparación a otras variantes de BERT. Por otra parte, los modelos de *Transformers* multilingües aumentan la precisión en diferentes dominios, incluyendo el análisis de sentimientos, la detección de emociones, la categorización de noticias y la atribución de autoría [40], [41].

Aunque BERT ha logrado resultados sorprendentes en

muchas tareas de comprensión del lenguaje natural, su potencial aún no se ha explorado por completo, tampoco en los modelos RoBERTa, ALBERT y DistilBERT. El presente trabajo tiene como enfoque la clasificación de texto multiclase, mediante los modelos lingüísticos basados en BERT para asignar automáticamente una categoría predefinida a una aplicación educativa móvil a partir de su descripción.

III. METODOLOGÍA

El proceso de asociar una categoría a un determinado texto depende de la información que contiene. Además, la longitud de texto es un elemento importante para considerar en la clasificación [42], [43]. De hecho, la tarea no es fácil al tener una entrada y en función de esta obtener una predicción a la salida con un grado de precisión y exactitud alto. En este trabajo, el método propuesto para la clasificación de las aplicaciones móviles educativas se muestra en la Figura 2.

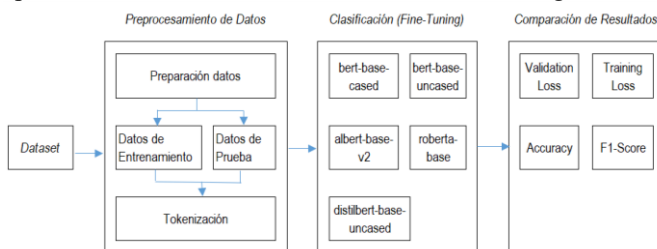


Fig. 2. Método para clasificar aplicaciones móviles educativas.

En general, el proceso desarrollado contempla las siguientes etapas:

- El proceso de creación del *dataset*: Permitió el reordenamiento de datos y creación de campos necesarios para la clasificación.
- Preprocesamiento de datos: Permitió la limpieza de los datos, creación de conjuntos de entrenamiento y prueba que luego fueron preparados para el proceso de aprendizaje.
- Clasificación por medio de un ajuste fino (*Fine-Tuning*): Entrenamiento de 5 modelos basados en BERT para realizar la clasificación de las aplicaciones móviles educativas.
- Comparación de resultados: Los modelos basados en BERT fueron comparados con diferentes métricas para determinar la exactitud y precisión de las predicciones.

La metodología aplicada es flexible tanto para el *dataset* como para los modelos basados en BERT; ya que son intercambiables por otros que se ajusten a la tarea de clasificación de texto, como el Framework OBOE [18].

A. Creación del Conjunto de Datos App-Dat-22

La plataforma de Google Play Store ofrece acceso para usuarios de dispositivos móviles Android en todo el mundo; lo que significa que los datos y las aplicaciones pueden ser utilizados por investigadores de diferentes países y en distintos contextos. Las aplicaciones disponibles en su plataforma incluyen información sobre su uso, calificaciones, opiniones de los usuarios, entre otras. Por lo tanto, estos datos pueden ser utilizados para analizar el uso de las aplicaciones educativas y

para identificar las tendencias y patrones de uso y ofrecer constantes oportunidades de innovación en el ámbito educativo.

Inicialmente, se utilizó un *dataset* en inglés con más de 2,3 millones de aplicaciones Android, tomadas de Google Play en junio de 2021 y disponible en la plataforma de Kaggle [44]. Sin embargo, para nuestra investigación, se delimitó las aplicaciones que sean exclusivamente educativas, con más de mil descargas y una valoración mayor a 3,5 estrellas.

El proceso para validar el *dataset* tomó alrededor de 8 meses debido a la falta de información en la descripción y categoría de las aplicaciones del *dataset* original.

Se ingresó a plataforma de Google Play de donde se obtuvo la descripción de cada una de las aplicaciones móviles educativas para ingresar manualmente toda la información en el *dataset*. A continuación, se debió realizar el proceso de etiquetado a cada una de las aplicaciones. Por lo tanto, para asignar una categoría se tomó como referencia a los campos propuestos por la Organización de las Naciones Unidas para la Educación (UNESCO) [45], [46]. De tal manera, que se consideró 7 campos generales de conocimiento con sus distintas ramas para ser asignados a cada una de las aplicaciones. Es importante mencionar que se halló aplicaciones que no correspondieron a ningún campo, por ejemplo, buscadores de becas y se las etiquetó como “*others*”. La Tabla I presenta los 7 campos generales de conocimiento que representan a las categorías con las que se etiquetó a cada una de las aplicaciones móviles educativas basadas en la UNESCO.

TABLA I
CODIFICACIÓN DE CATEGORÍAS

CATEGORÍA	REGISTROS	DESCRIPCIÓN
<i>linguistics</i>	687	Se interesa por las aplicaciones relacionadas ramas de la lingüística que contemplen a la fonética, fonología, morfología, sintaxis, semántica, lexicología y gramática [47].
<i>applied science</i>	360	Se centra en aplicaciones relacionadas estrictamente a las ciencias aplicadas como Ingeniería civil, ingeniería mecánica, ingeniería eléctrica, ingeniería industrial, ingeniería en sistemas, entre otras [48].
<i>natural sciences</i>	311	Se ocupa de aplicaciones dentro de las ramas de la física, química, astronomía, geología, biología, astronomía y ciencias de la salud [49].
<i>math</i>	305	Se centra en aplicaciones de álgebra, geometría, probabilidad y estadística, aritmética, topología, análisis, trigonometría y nomografía [50].
<i>social sciences</i>	268	Se interesa en aplicaciones enfocadas en la economía, geografía, ciencias políticas, psicología, antropología, sociología, historia y demografía [51].

<i>art</i>	112	Se ocupa en aplicaciones de diseño y arquitectura, audiovisual, danza, teatro, dibujo, creación y composición música [52].
<i>others</i>	1.078	Destacan aplicaciones relacionadas a buscadores de becas, sistemas de gestión de aprendizaje para la creación de aulas virtuales o bibliotecas.

Finalmente, el *dataset* consolidado se denominó App-Dat-22 con 3.121 registros, suficientes para la fase de entrenamiento y prueba. El *dataset*, así como el todo el proyecto está disponible en <https://github.com/Anabel-Pilicita/NLP>.

B. Preprocesamiento de Datos

Antes de iniciar el proceso de aprendizaje por transferencia para el clasificador de aplicaciones, el conjunto de datos en bruto App-Dat-22 se transformó en un formato comprensible. En este caso, se realizó un proceso de limpieza para filtrar los datos inútiles y se aplicó funciones al *dataset* para establecer las etiquetas codificadas. Posteriormente, los datos fueron divididos aleatoriamente con el 80% para el entrenamiento (2.496 registros) y un 20% para la fase de prueba (625 registros).

Finalmente, se realizó un proceso de tokenización para dividir el texto en unidades significativas llamadas tokens. En este caso se utilizó un tokenizador predeterminado de BERT y el resultado se utilizó como entrada para el ajuste fino (*Fine-Tuning*).

C. Ajuste Fino (*Fine-Tuning*)

Para muchas aplicaciones de PLN en las que intervienen modelos *Transformers*, basta con tomar un modelo preentrenado y ajustarlo directamente a los datos [53]. En nuestro caso el *dataset* estaba alejado del corpus inicial con el que se entrenó a todas las variantes de BERT. Debido a esto fue necesario realizar el ajuste fino para mejorar el rendimiento en la clasificación de las aplicaciones educativas.

D. Modelos de Entrenamiento.

El modelo preentrenado BERT lanzado por Google en 2019 es conceptualmente sencillo y potente [28]. BERT está diseñado para un entrenamiento bidireccional profundo con base en *Transformers*. La decisión de explorar el modelo BERT se justifica por sus diferentes ventajas, como mejorar peticiones de los usuarios para proporcionar resultados más relevantes, especialmente para las peticiones formuladas de forma natural.

En esta investigación, se trabajó con variantes del modelo BERT manteniendo en todos los modelos como entrada la columna “*descripcion*” que contiene la información de una aplicación móvil educativa y como salida la columna “*general*” que dispone de las etiquetas para cada una de las categorías descritas en la Tabla I. En este caso fueron seleccionados 5 modelos de la familia de BERT. La selección se basó en la gran cantidad de descargas y la popularidad de los modelos en la plataforma de HuggingFace que se describen a continuación:

- bert-base-cased (BERT): Modelo preentrenado en lengua inglesa con un *Modelado Lingüístico Enmascarado* (MLM). Este modelo distingue entre

mayúsculas y minúsculas [54]. Un modelo descargado más de 6 millones de veces y preentrenado con 110 millones de parámetros.

- bert-base-uncased (BERT): Es un modelo entrenado con 110 millones de parámetros, que elimina los marcadores de acento del texto, y no distingue entre mayúsculas y minúsculas. Modelo que sobrepasa los 33 millones de descargas [55].
- roberta-base (RoBERTa): Es un modelo basado en BERT que optimiza su arquitectura al modificar hiperparámetros claves, eliminando el objetivo de preentrenamiento de la siguiente frase y, entrenando con mini lotes y tasas de aprendizaje mucho más grandes [35]. Este modelo tiene más de 6 millones de descargas [56].
- albert-base-v2 (ALBERT): Una versión más ligera de BERT, en la que se sustituye la predicción de la siguiente frase por la predicción del orden de las frases, y se utilizan técnicas de reducción de parámetros para un menor consumo de memoria y un entrenamiento más rápido [57]. Actualmente, tiene más de 2 millones de descargas [58].
- distilbert-base-uncased (DistilBERT): Es un modelo reducido en un 40% de BERT, conservando el 97% de sus capacidades de comprensión del lenguaje y siendo un 60% más rápido [15]. Hasta la fecha, cuenta con más de 9 millones de descargas [58].

Adicionalmente, se fijaron los hiperparámetros de entrenamiento para los 5 modelos expuestos anteriormente. El tamaño del lote fue de 64; la tasa de aprendizaje (Adam) de $2e-5$. El número de épocas (*Epoch*) para el entrenamiento se fijó en 4 y el valor de la semilla en 42.

IV. RESULTADOS

El conjunto de 2.496 registros se utilizó para entrenar los modelos bert-base-cased, bert-base-uncased, roberta-base, albert-base-v2 y distilbert-base-uncased. Mientras que los 625 registros se utilizaron en la etapa de prueba para calcular el rendimiento y la evaluación de los modelos. El proceso de evaluación inicia utilizando el modelo entrenado para predecir las categorías de las aplicaciones móviles del conjunto de pruebas y las compara con las etiquetas de datos proporcionadas en el *dataset* App-Dat-22. En esta fase de prueba se evaluó el comportamiento de los modelos en base a las siguientes métricas [59]:

- Pérdida de Entrenamiento (*Validation Loss*): Evalúa el error del modelo sobre el conjunto de datos de entrenamiento
- Pérdida de Validación (*Training Loss*): Evalúa el rendimiento de un modelo de aprendizaje profundo en el conjunto de prueba.
- Precisión (*Accuracy*): Es la proporción de predicciones correctas entre el número total de casos procesados; esta métrica revela cuántas de las clases previstas están correctamente etiquetadas.

TABLA II
RESULTADOS DE LOS MODELOS

MODELOS	EPOCH	VALIDATION LOSS	TRAINING LOSS	ACURRACY	F1- SCORE
bert-base- cased	1	1,28	1,18	64%	52%
	2	0,76	0,79	77%	74%
	3	0,58	0,66	79%	76%
	4	0,56	0,63	80%	77%
bert-base- uncased	1	1,25	1,18	61%	50%
	2	0,75	0,83	75%	71%
	3	0,68	0,71	79%	77%
	4	0,58	0,68	79%	77%
roberta-base	1	1,13	1,02	63%	53%
	2	0,64	0,66	80%	72%
	3	0,58	0,61	81%	80%
	4	0,50	0,58	81%	80%
albert-base-v2	1	1,21	1,17	60%	50%
	2	0,75	0,89	72%	67%
	3	0,69	0,76	77%	72%
	4	0,59	0,72	78%	74%
distilbert- base-uncased	1	1,34	1,27	60%	48%
	2	0,88	0,95	72%	66%
	3	0,83	0,83	75%	70%
	4	0,75	0,80	76%	71%

- Exactitud (*F1-Score*): Es la media armónica de la precisión y la recuperación para obtener un valor mucho más objetivo.

En general, el modelo BERT con sus variantes obtuvieron buenos resultados alcanzando una precisión superior al 75% y una exactitud mayor al 70%. En la Tabla II se visualiza que los modelos bert-base-cased y bert-base-uncased presentaron el mismo resultado en exactitud (*F1-Score*) correspondiente al 77%. Por otro lado, distilbert-base-uncased obtuvo 71%, albert-base-v2 con 74%. Finalmente, el modelo que mejor resultado de exactitud fue roberta-base con un 80%.

En el caso de la precisión (*Accuracy*), en todos los modelos este valor es superior al presentado en exactitud, siendo el modelo de roberta-base con el valor más alto de 81%.

En los resultados se observa que los valores de pérdida de entrenamiento y validación son menores. En consecuencia, las métricas de *F1-Score* y *Accuracy* presentan mejores resultados debido a que disminuye la pérdida de datos durante el entrenamiento y prueba de los modelos. Con respecto a los valores más bajos de pérdida de validación y entrenamiento los encontramos en roberta-base con 0,50 y 0,58 respectivamente durante la cuarta época.

En todos los modelos aplicados en la clasificación de aplicaciones móviles educativas la precisión y exactitud es más baja en la primera época; esto se debe a que los modelos presentan mayor error en la pérdida de entrenamiento y prueba durante la primera ronda de entrenamiento. No obstante, los modelos mejoraron sus resultados a medida que se ajustaba en la segunda, tercer y cuarta época. Aunque es importante mencionar que no se recomienda aumentar más épocas porque se puede caer en un sobreajuste innecesario.

Los resultados obtenidos son prometedores dado su rendimiento y su simplicidad. Todos los modelos tienen como base a BERT y son viables para una futura implementación como un servicio de clasificación de aplicaciones educativas por categoría. Adicionalmente, estos resultados presentados

permiten seleccionar el modelo más apropiado para una futura implementación.

Tras entrenar los datos, se realizó la fase de pruebas, en la que los modelos tuvieron que pasar por cuatro épocas de entrenamiento para obtener mejores resultados: roberta-base con 81% de precisión, bert-base-uncased con 79%, bert-base-cased obtuvo un 80% de precisión, albert-base-v2 un 78% de precisión y distilbert-base-uncased un 76% de precisión

V. CONCLUSIONES

La clasificación automática de textos estudia las soluciones existentes para el tratamiento de textos no etiquetados. En este trabajo, se abordó la tarea de clasificar aplicaciones móviles educativas según su campo general de conocimiento. El artículo planteó 3 objetivos que fueron desarrollados en la sección III y IV.

Los resultados muestran claramente que el modelo que obtuvo mejores resultados en el proceso de clasificación de aplicaciones móviles educativas fue roberta-base con un valor de *Accuracy* del 81% y un valor de *F1-Score* del 80%. Este modelo es una variante de BERT, entrenado con estrategias más eficientes y exhaustivas. Además, el modelo tiene una mejor capacidad para manejar palabras nuevas o raras que no se vieron durante el entrenamiento. Otro factor que contribuye a la mejoría es que utiliza más datos de entrenamiento que el modelo original de BERT.

Los resultados obtenidos con los modelos bert-base-cased, bert-base-uncased, albert-base-v2 y distilbert-base-uncased son prometedores ya que superan el 75% en *Accuracy* y el 71% en *F1-Score* para realizar la tarea de clasificación de aplicaciones móviles educativas. Los modelos pertenecen a la familia de BERT motivo por el cual los resultados se aproximan entre los modelos mencionados.

Los resultados obtenidos en esta investigación son referenciales; ya que, en la plataforma de HuggingFace los modelos son actualizados cada cierto tiempo por lo que los puede producirse variaciones a futuro.

En futuros trabajos, se recomienda investigar modelos lingüísticos alternativos para la representación del lenguaje como GPT-3 [60], o su competidor abierto GPT-NEO[61], MarIA [62] o BETO [63] que es una versión en español del modelo BERT.

REFERENCES

- [1] E. D. Liddy, "Natural Language Processing," 2001, [Online]. Available: <https://surface.syr.edu/istpub>
- [2] R. Dale, "HANDBOOK OF NATURAL LANGUAGE PROCESSING SECOND EDITION," Second., N. Indurkha and F. J. Damerau, Eds. Florida, 2010, pp. 3–7.
- [3] T. Shangipour ataei, S. Javdan, and B. Minaei-Bidgoli, "Applying Transformers and Aspect-based Sentiment Analysis approaches on Sarcasm Detection," in *Proceedings of the Second Workshop on Figurative Language Processing*, Jul. 2020, pp. 67–71. doi: 10.18653/v1/2020.figlang-1.9.
- [4] M. Razno, "Machine Learning Text Classification Model with NLP Approach," in *Computational Linguistics and Intelligent Systems*, 2019, pp. 71–77.
- [5] S. Wang and C. Manning, "Baselines and Bigrams: Simple, Good Sentiment and Topic Classification," in *Proceedings of the 50th Annual*

- Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Jul. 2012, pp. 90–94. [Online]. Available: <https://aclanthology.org/P12-2018>
- [6] L. Schmidt, J. Weeds, and J. P. T. Higgins, “Data Mining in Clinical Trial Text: Transformers for Classification and Question Answering Tasks.” [Online]. Available: <https://orcid.org/0000-0002-3831-4019>
- [7] “Estadísticas para este 2022 en la tienda de Google Play! — Apps Rentables.” <https://appsrentables.com/estadisticas-tienda-google-play/> (accessed Sep. 29, 2022).
- [8] “Google Play Store: number of apps 2022 | Statista.” <https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/> (accessed Sep. 29, 2022).
- [9] Google Play, “Aplicaciones de Android en Google Play,” 2022. https://play.google.com/store/games?gl=ES&utm_source=emea_Med&utm_medium=hasem&utm_content=Dec2020&utm_campaign=Evergreen&pcampaignid=MKT-EDR-emea-es-1001280-Med-hasem-py-Evergreen-Dec2020-Text_Search_BKWS-test_ctrl_ca%7CONSEM_kwid_43700008794309700&gclid=CjwKC_Ajw9suYBhBIEiwa7iMhNPNsbashTJ8e69tNRpqtquhiU4pXkG9Ci-XUjlCrajJEf3OESrBwBoC0acQAvD_BwE&gclid=aw.ds (accessed Sep. 02, 2022).
- [10] “Global Google Play app revenue share by category 2025 | Statista.” <https://www.statista.com/statistics/1010710/google-play-app-revenue-share-by-category-worldwide/> (accessed Sep. 29, 2022).
- [11] “Google Play most popular app categories 2022 | Statista.” <https://www.statista.com/statistics/279286/google-play-android-app-categories/#statisticContainer> (accessed Mar. 16, 2023).
- [12] Google, “Novedades de la búsqueda de imágenes - Ayuda de Búsqueda web de Google,” 2022. <https://support.google.com/websearch/answer/9792245?hl=es-419> (accessed Sep. 04, 2022).
- [13] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [14] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soiccut, “ALBERT: A Lite BERT for Self-supervised Learning of Language Representations,” Sep. 2019, doi: 10.48550/1909.11942.
- [15] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.01108>
- [16] M. Zaveri, “Automatic Text Classification: A Technical Review Cite this paper Related papers,” 2011.
- [17] P. M. Lavanya and E. Sasikala, “Deep Learning Techniques on Text Classification Using Natural Language Processing (NLP) In Social Healthcare Network: A Comprehensive Survey,” *2021 3rd International Conference on Signal Processing and Communication (ICSPSC)*, 2021, doi: 10.1109/ICSPSC51351.2021.9451752.
- [18] R. A. del Águila Escobar, M. C. Suárez-Figueroa, and M. Fernández-López, “OBOE: an Explainable Text Classification Framework,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. InPress, no. InPress, p. 1, 2022, doi: 10.9781/ijimai.2022.11.001.
- [19] Z. Shaheen, G. Wohlgenannt, and E. Filtz, “Large Scale Legal Text Classification Using Transformer Models,” Oct. 2020, [Online]. Available: <http://arxiv.org/abs/2010.12871>
- [20] H. Jwa, D. Oh, K. Park, J. M. Kang, and H. Lim, “exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT),” *Applied Sciences*, vol. 9, no. 19, 2019, doi: 10.3390/app9194062.
- [21] M. Khadhraoui, H. Bellaaj, M. Ben Ammar, H. Hamam, and M. Jmaiel, “Survey of BERT-Base Models for Scientific Text Classification: COVID-19 Case Study,” *Applied Sciences (Switzerland)*, vol. 12, no. 6, Mar. 2022, doi: 10.3390/app12062891.
- [22] S. Gao *et al.*, “Limitations of Transformers on Clinical Text Classification,” *IEEE J Biomed Health Inform*, vol. 25, no. 9, pp. 3596–3607, 2021, doi: 10.1109/JBHI.2021.3062322.
- [23] Y. Arslan *et al.*, “A Comparison of Pre-Trained Language Models for Multi-Class Text Classification in the Financial Domain,” in *Companion Proceedings of the Web Conference 2021*, 2021, pp. 260–268. doi: 10.1145/3442442.3451375.
- [24] M. G. Huddar, S. S. Sannakki, and V. S. Rajpurohit, “Attention-based multi-modal sentiment analysis and emotion detection in conversation using rnn,” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 6, pp. 112–121, 2021, doi: 10.9781/ijimai.2020.07.004.
- [25] N. K. Manaswi, “RNN and LSTM,” in *Deep Learning with Applications Using Python: Chatbots and Face, Object, and Speech Recognition With TensorFlow and Keras*, Berkeley, CA: Apress, 2018, pp. 115–126. doi: 10.1007/978-1-4842-3516-4_9.
- [26] A. Vaswani *et al.*, “Attention Is All You Need.”
- [27] O. Dekel and O. Shamir, “Multiclass-Multilabel Classification with More Classes than Examples,” 2010.
- [28] J. Devlin, M.-W. Chang, K. Lee, K. T. Google, and A. I. Language, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” 2019. [Online]. Available: <https://github.com/tensorflow/tensor2tensor>
- [29] J. Pilault, R. Li, S. Subramanian, C. Pal, and P. Montreal, “On Extractive and Abstractive Neural Document Summarization with Transformer Language Models,” 2020.
- [30] T. Wolf *et al.*, “HuggingFace’s Transformers: State-of-the-art Natural Language Processing,” Oct. 2019, doi: 10.48550/1910.03771.
- [31] M. E. Peters *et al.*, “Deep contextualized word representations,” Feb. 2018, doi: 10.48550/1802.05365.
- [32] J. Pennington, R. Socher, and C. D. Manning, “GloVe: Global Vectors for Word Representation.” [Online]. Available: <http://nlp>.
- [33] A. R. Openai, K. N. Openai, T. S. Openai, and I. S. Openai, “Improving Language Understanding by Generative Pre-Training.” [Online]. Available: <https://gluebenchmark.com/leaderboard>
- [34] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” Oct. 2018, doi: 10.48550/1810.04805.
- [35] Y. Liu *et al.*, “RoBERTa: A Robustly Optimized BERT Pretraining Approach,” Jul. 2019, doi: 10.48550/1907.11692.
- [36] C. Blackledge and A. Atapour-Abarghouei, “Transforming Fake News: Robust Generalisable News Classification Using Transformers,” in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 3960–3968. doi: 10.1109/BigData52589.2021.9671970.
- [37] P. Gupta, S. Gandhi, and B. R. Chakravarthi, “Leveraging Transfer learning techniques- BERT, RoBERTa, ALBERT and DistilBERT for Fake Review Detection,” in *ACM International Conference Proceeding Series*, Dec. 2021, pp. 75–82. doi: 10.1145/3503162.3503169.
- [38] I. Staliūnaitė and I. Iacobacci, “Compositional and Lexical Semantics in RoBERTa, BERT and DistilBERT: A Case Study on CoQA,” Sep. 2020, doi: 10.48550/1909.08257.
- [39] A. Pritzkau, O. Blanc, M. Geierhos, and U. Schade, “Nlytics at CheckThat! 2022: Hierarchical multi-class fake news detection of news articles exploiting the topic structure,” 2022. [Online]. Available: <https://www.fkie.fraunhofer.de>
- [40] F. El-Alami, S. Ouatic El Alaoui, and N. En Nahnahi, “Contextual semantic embeddings based on fine-tuned AraBERT model for Arabic text multi-class categorization,” *Journal of King Saud University - Computer and Information Sciences*, 2021, doi: <https://doi.org/10.1016/j.jksuci.2021.02.005>.
- [41] T. Alam, A. Khan, and F. Alam, “Bangla Text Classification using Transformers,” 2020.
- [42] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Inf Process Manag*, vol. 24, no. 5, pp. 513–523, 1988, doi: [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).
- [43] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, “Bag of Tricks for Efficient Text Classification,” in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, Apr. 2017, pp. 427–431. [Online]. Available: <https://aclanthology.org/E17-2068>
- [44] “Google Play Store Apps | Kaggle.” <https://www.kaggle.com/datasets/gauthamp10/google-playstore-apps> (accessed Sep. 24, 2022).
- [45] Instituto de Estadística de la UNESCO, “Campos de Educación y Capacitación 2013 de la CINE,” Montréal, 2014. doi: 10.15220/978-92-9189-157-3-sp.
- [46] Instituto de Estadística de la UNESCO, “Clasificación internacional normalizada de la educación,” 2013.
- [47] B. M. Islas, “Las Ramas de la Lingüística,” *Unidades de Apoyo para el Aprendizaje*, 2018.

- http://132.248.48.64/repositorio/moodle/pluginfile.php/1396/mod_resource/content/1/contenido/index.html (accessed Sep. 24, 2022).
- [48] Inc. Fundación Wikimedia, “Ciencias aplicadas - Wikipedia, la enciclopedia libre,” 2022. https://es.wikipedia.org/wiki/Ciencias_aplicadas (accessed Sep. 24, 2022).
- [49] UNAN-MANAGUA, “Ciencias Naturales,” 2019.
- [50] Pol Bertran Prieto, “Las 20 ramas de las Matemáticas (y qué estudia cada una).” <https://medicoplus.com/ciencia/ramas-matematicas> (accessed Sep. 24, 2022).
- [51] COMECSO, “Ventana a disciplinas de las ciencias sociales.” <https://www.comecso.com/ventana-disciplinas-ciencias-sociales> (accessed Sep. 24, 2022).
- [52] Blog Bellas Artes, “¿Cuáles son las bellas artes?” <https://www.elcorteingles.es/entradas/blog/cuales-son-las-bellas-artes/> (accessed Sep. 24, 2022).
- [53] “Fine-tuning a masked language model - Hugging Face Course.” <https://huggingface.co/course/chapter7/3?fw=tf> (accessed Sep. 29, 2022).
- [54] “bert-base-cased · Hugging Face.” <https://huggingface.co/bert-base-cased> (accessed Sep. 28, 2022).
- [55] “bert-base-uncased · Hugging Face.” <https://huggingface.co/bert-base-uncased> (accessed Sep. 28, 2022).
- [56] “roberta-base · Hugging Face.” <https://huggingface.co/roberta-base> (accessed Sep. 28, 2022).
- [57] Z. Lan *et al.*, “ALBERT: A LITE BERT FOR SELF-SUPERVISED LEARNING OF LANGUAGE REPRESENTATIONS.” [Online]. Available: <https://github.com/google-research/ALBERT>.
- [58] “albert-base-v2 · Hugging Face.” <https://huggingface.co/albert-base-v2> (accessed Sep. 28, 2022).
- [59] “Hugging Face – The AI community building the future.” <https://huggingface.co/metrics> (accessed Nov. 05, 2022).
- [60] R. Dale, “GPT-3: What’s it good for?,” *Nat Lang Eng*, vol. 27, no. 1, pp. 113–118, Jan. 2021, doi: 10.1017/S1351324920000601.
- [61] L. Gao *et al.*, “The Pile: An 800GB Dataset of Diverse Text for Language Modeling,” Dec. 2020, doi: 10.48550/arxiv.2101.00027.
- [62] “Nace MarIA, el primer modelo masivo de inteligencia artificial de la lengua española | TECNOLOGÍA | ComputerWorld.” <https://www.computerworld.es/tecnologia/nace-maria-el-primer-modelo-masivo-de-inteligencia-artificial-de-la-lengua-espanola> (accessed Sep. 28, 2022).
- [63] J. Tiedemann, “Parallel Data, Tools and Interfaces in OPUS,” in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*, May 2012, pp. 2214–2218. [Online]. Available: http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf

incluyen la videoconferencia, los juegos en la educación y las redes sociales en la educación



Anabel Pilicita obtuvo el Máster en Ingeniería de Redes y Servicios Telemáticos en la Universidad Politécnica de Madrid (UPM), en 2016, donde actualmente está cursando el doctorado en Ingeniería de Servicios Telemáticos.

Sus intereses de investigación incluyen la aplicación del procesamiento del lenguaje natural y nuevos modelos de inteligencia artificial.



Enrique Barra se recibió como doctor (Ph.D.) en Ingeniería Telemática con especialización en Aprendizaje Multimedia por la Universidad Politécnica de Madrid (UPM). Ha participado en numerosos proyectos europeos, como GLOBAL, FIWARE y C@R. Actualmente está involucrado en varios proyectos que contribuyen a la generación y distribución de contenidos educativos en entornos TEL. Sus intereses de investigación