

Análisis de sentimientos en críticas de películas mediante transformadores BERT

Informe de experimento realizado en críticas de películas
en la plataforma IMBD

Cristian Camilo Segura Morales

A210366

Matemáticas para las Ciencias de la Computación

Maestría para las Ciencias de la Computación

Instituto Politécnico Nacional

1 de noviembre de 2021

Abstract

El reconocimiento de emociones es una tarea de gran importancia en las tareas de procesamiento de lenguaje natural con diferentes aplicaciones en varias áreas del conocimiento. La detección implícita de emociones en los textos principalmente en aquellas interacciones en las redes sociales es el problema más desafiante a resolver debido a que las emociones a transmitir y expresar se encuentran ocultas en el texto y encontrarlas requiere comprender el contexto de los diferentes tipos de texto. En el presente documento se realizará la implementación del análisis inicial de emociones de críticas de películas en la plataforma IMDB para realizar la clasificación entre críticas positivas y negativas y se mostrarán los resultados obtenidos del experimento realizado y las respectivas conclusiones.

Índice general

	Page
1. Justificación	3
1.1. Objetivos	3
1.2. Contribuciones	4
2. Estado del arte	5
2.1. Medidas de evaluación	5
2.2. Resumen estado del arte	6
3. Fundamentos	9
3.1. Herramientas tecnológicas	9
4. Solución Propuesta	11
5. Resultados	13
5.1. Resultados con datos de prueba	13
5.2. Predicciones	14
6. Conclusiones	19
7. Trabajo Futuro	20

1 Justificación

El procesamiento de lenguaje natural es uno de los campos más extensos en la interacción que puede establecer una inteligencia artificial con una persona. Abarca conceptos de computación, lingüística y la implementación de sistemas de Inteligencia Artificial. Debido a que el tema de tesis seleccionado para optar al título de master en Ciencias de la Computación consiste en la identificación de sentimientos en redes sociales, el desarrollo del presente proyecto permitirá realizar un acercamiento al estado del arte de técnicas actuales utilizadas para el análisis de los sentimientos permitiendo realizar una verificación documental y la implementación técnica de los modelos existentes para comprender los sistemas actuales y comparar su desempeño para con ello contar con un fundamento fuerte en el desarrollo de la tesis de grado.

1.1. Objetivos

1.1.1. Objetivo General

Implementar un sistema de análisis de sentimientos en textos compartidos en redes sociales mediante técnicas de aprendizaje profundo supervisado para la correcta clasificación de emociones.

1.1.2. Objetivos Específicos

Identificar el estado del arte de trabajos realizados sobre análisis de sentimientos en las redes sociales

Verificar los diferentes conjuntos de datos disponibles y realizar la selección del conjunto para utilizar como insumo en la técnica de análisis supervisado.

Seleccionar la técnica de aprendizaje profundo a implementar de acuerdo con el conjunto de datos seleccionado.

Realizar el preprocesamiento de los datos utilizados aplicando procesos de tokenization, eliminación de palabras gramaticales, lematización, POS Tagging y análisis de dependencias.

Realizar entrenamiento del modelo seleccionado con los datos disponibles para este concepto.

Realizar predicciones con los datos de prueba para verificar el funcionamiento del modelo.

Aplicar medidas de evaluación precisión, recall, accuracy y F-score para verificar el comportamiento del modelo y sus tasas de clasificación.

1.2. Contribuciones

Mediante la implementación de modelo de aprendizaje automático con transformadores se hará manejo de un sistema con el que no se había trabajado anteriormente y que cuenta con la definición de ser uno de los mejores sistemas que existen actualmente para realizar el proceso de clasificación.

Por ello conocerlo, identificar su arquitectura y su forma de procesamiento contribuirá a contar con un mayor conocimiento en este sistema y ampliar la gama de opciones disponibles para realizar la implementación en el proyecto de tesis de maestría.

2 Estado del arte

2.1. Medidas de evaluación

Cada uno de los trabajos identificados en el estado del arte realizan la medición de su desempeño mediante diferentes técnicas, para así realizar la medición de la similitud entre los conjuntos pronosticados contra los datos reales y también identificar el nivel de precisión en el pronóstico. Algunas medidas identificadas en los documentos son las siguientes.

Presición

$$\mathbf{P}_e = \frac{VP_e}{VP_e + FP_e} \quad (2.1)$$

Donde e correponde a la etiqueta de emoción, VP a Verdadero Positivo y FP a Falso Positivo.

Recall

$$\mathbf{R}_e = \frac{VP_e}{VP_e + FN_e} \quad (2.2)$$

Donde e correponde a la etiqueta de emoción, VP a Verdadero Positivo y FN a Falso Negativo.

F_{Score}

$$\mathbf{F}_1 = \frac{P_e}{R_e} \quad (2.3)$$

Donde P_e corresponde al valor de la presición y R_e al valor del recall calculados previamente.

Acuracy

$$\mathbf{A} = \frac{\sum_{e \in E} VP + \sum_{e \in E} VN}{\sum_{e \in E} VP + \sum_{e \in E} VN + \sum_{e \in E} FP + \sum_{e \in E} FN} \quad (2.4)$$

Donde E corresponde al conjunto de etiquetas de emoción, VP corresponde al número de verdaderos positivos, VN es el número de verdaderos negativos, FP es el número de Falsos Positivos y FN es el número de falsos negativos.

2.2. Resumen estado del arte

En el Cuadro 1 se presenta el resumen del estado del arte revisado realizando una comparación entre la bibliografía mediante una tabla donde se indica el artículo mencionado, el idioma de los conjuntos de datos utilizados, el enfoque utilizado, la herramienta y/o algoritmo utilizados en la implementación, el nombre del corpus utilizado y el desempeño del modelo implementado.

Referencia	Idioma	Método	Herramienta - Algoritmo	Dataset	Desempeño	
					Medida	Valor (%)
Bandhakavi et al. (2017)	Inglés	Aprendizaje clasico	Máquinas de vectores de soporte	SemEval-07	F-Score	40.99
				Twitter		49.55
				Blogs		58.32
				ISEAR		31.21
(Yasmina et al., 2016)	Inglés	Aprendizaje clásico	Máquinas de vectores de soporte WEKA	Comentarios en YouTube	Presición	92.75
					Recall	72.72
					Accuracy	68.82
(Wu et al., 2018)	Inglés	Aprendizaje Profundo	Redes Neuronales Convolucionales	SemEval2018	Pearson	81.00
			Redes de Memoria a corto plazo			
(Tao, 2020)	Chino	Aprendizaje Profundo	LDA (Latent Dirichlet Allocation)	Reseñas Hoteles	Presición	88.47
			Redes de Memoria a corto plazo		Recall	79.88
					Accuracy	87.51
(Ghanbari-Adivi and Mosleh, 2019)	Inglés	Hibrido:	Máquinas de vectores de soporte	ISEAR	Presición	82.00
		Reglas y Aprendizaje clásico	Vecino más cercano		Recall	92.88
			Perceptrón multicapa		Accuracy	88.49
			Árboles de decisión		F-Score	81.00

Referencia	Idioma	Método	Herramienta - Algoritmo	Dataset	Desempeño	
					Medida	Valor (%)
(Sailunaz and Alhaji, 2019)	Inglés	Aprendizaje clásico	Naive Bayes	ISEAR	Presición	82.00
			Máquinas de vectores de soporte	Twitter	Accuracy	66.86
			Bosques aleatorios		Accuracy	23.32
(Ibrahiem et al., 2020)	Inglés	Aprendizaje clásico	Clasificación binaria	SemEval2018	Presición	63.00
			Redes Neuronales Convolucionales	NRC Emotion	Recall	82-00
				NRC Hashtag	F-Score	69.00
(Baziotis et al., 2018)	Inglés	Aprendizaje Profundo	Clasificación binaria	SemEval2017	Jaccard	57.88
			Redes Neuronales Recurrentes			
			Redes de Memoria a Corto Plazo			
(Ge et al., 2019)	Inglés	Aprendizaje Profundo	Redes Neuronales Convolucionales	SemEval2019	Presición	75.98
					Recall	74.88
					F-Score	75.42
(Patwa et al., 2020)	Hinglish	Aprendizaje	Redes Neuronales Convolucionales	SemEval2020	F-Score	80.6
	Spanglish	Profundo				

Cuadro 2.1: Resumen de trabajos sobre reconocimiento de emociones en textos en las redes sociales

3 Fundamentos

Dentro de las herramientas y los conceptos teóricos utilizados en el desarrollo de proyecto se encuentran los siguientes.

3.1. Herramientas tecnológicas

3.1.1. Pandas.

Librería de código abierto que ofrece estructuras de datos flexibles para la fácil manipulación y tratamiento de datos. La librería permite la carga, modelamiento, análisis, manipulación y preparación de datos.

3.1.2. Matplotlib.

Librería con un conjunto de funciones que permite la generación de gráficos a partir de datos almacenados en diferentes tipos de estructuras.

3.1.3. Numpy.

Librería que se especializa en el cálculo numérico y análisis de datos para aquellos que cuentan con un gran volumen. Representa colecciones de datos de un mismo tipo en diferentes dimensiones.

3.1.4. TensorFlow.

Librería para tareas de aprendizaje automático y profundo desarrollado por Google con objetivo de construcción y entrenamiento de diferentes modelos de Machine Learning. Los modelos de TF han mostrado un rendimiento 10 veces mayor en tareas de aprendizaje automático que en los sistemas tradicionales.

3.1.5. Keras.

Librería utilizada para el manejo de redes neuronales. Puede ejecutarse sobre diferentes plataformas como TensorFlow o Theano. Está diseñado para la construcción por bloques de la arquitectura de las

redes neuronales, incluyendo redes convolucionales y recurrentes que permiten entrenar modelos deep learning.

3.1.6. Bidirectional Encoder Representations from Transformers (BERT)

Modelo de aprendizaje automático de última generación utilizado para tareas de Procesamiento de Lenguaje Natural. Desarrollado por el equipo de Jacob Devlin en Google en 2018. El modelo BERT fue capacitado con artículos de Wikipedia en inglés (2500 millones de palabras) y BooksCorpus (800 millones de palabras) y lograron las mejores precisiones para algunas de las tareas de PNL. Existen actualmente dos Variaciones generales de BERT previamente entrenadas: el modelo base es una arquitectura de red neuronal de 12 capas, 768 ocultos, 12 cabezas, 110 millones de parámetros, mientras que el modelo grande es un parámetro de 24 capas, 1024 ocultos, 16 cabezas y 340 millones de parámetros.

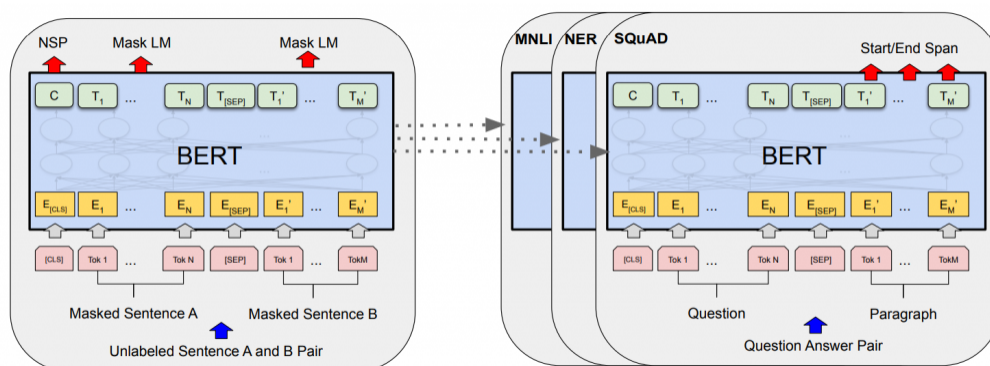


Figura 3.1: Procedimiento de entrenamiento y afinamiento BERT

4 Solución Propuesta

Internet Movie Database (IMDb) es un repositorio en línea que cuenta con información relacionada con películas, programas de televisión, eventos en vivo, entrega de premios y especiales. Los usuarios son partícipes en la creciente riqueza de información del sitio clasificando películas en una escala del 1 al 10 además de incluir una crítica personal hacia la película o espectáculo que se está calificando. Cuenta con una división entre reseñas de usuarios y de expertos críticos de cine.

El conjunto de datos seleccionados para el modelo implementado contiene reseñas de películas junto con la clase de clasificación de cada una de las reseñas (positiva o negativa). Se cuenta con un total de 50.000 revisiones divididas uniformemente en dos grupos de 25.000 reseñas, donde un grupo es utilizado para entrenamiento y el otro para prueba. La distribución general de las etiquetas de reseña se encuentra equilibrada 25.000 reseñas positivas y negativas. En toda la colección, no se permiten más de 30 reseñas para la misma película debido a que las críticas de la misma película tienden a correlacionarse. Una revisión es considerada negativa cuando la crítica tiene una puntuación menor o igual a 4 sobre 10, y una reseña es positiva cuando tiene una puntuación mayor o igual a 7 sobre 10. En ambos conjuntos las reseñas con clasificaciones neutrales (5 y 6) no son incluidas en el repositorio de prueba.

En la implementación realizada, se utilizó el modelo base de 110 millones de parámetros configurables como se evidencia en la imagen 2 del resumen del modelo creado.

Layer (type)	Output Shape	Param #
bert (TFBertMainLayer)	multiple	109482240
dropout_37 (Dropout)	multiple	0
classifier (Dense)	multiple	1538
Total params: 109,483,778		
Trainable params: 109,483,778		
Non-trainable params: 0		

Figura 4.1: Procedimiento de entrenamiento y afinamiento BERT

El total de 25.000 críticas positivas y negativas fue dividido en 2 grupos principales de 20.000, las cuales corresponden al grupo de datos utilizados en el proceso de entrenamiento y 2 grupos de 5.000 reseñas las cuales corresponden al grupo de prueba del modelo entrenado. Posteriormente, se realizó el gargue de los datos de entrenamiento y los de prueba en un dataset de tipo Pandas. La previsualización se evidencia en la tabla 4.3. Y luego con el procedimiento de tokenización incluido con el modelo BERT

DATA_COLUMN	LABEL_COLUMN
Canadian director Vincenzo Natali took the art-...	1
I gave this film 10 not because it is a superbl...	1
I admit to being somewhat jaded about the movie...	1
For a long time, 'The Menagerie' was my favorit...	1
A truly frightening film. Feels as if it were m...	0

Cuadro 4.1: Previsualización de crítica y etiqueta de asignación

se convierte el dataframe en los tipos de datos de entrada que entiende el modelo para su entrenamiento.

La etapa de entrenamiento es la más larga del proceso, debido a que únicamente se realizaron 2 épocas, pero el tiempo de entrenamiento para cada una de ellas es superior a siete horas por la cantidad de parámetros configurables. En la tabla 4.2 se presenta el resumen del tiempo de entrenamiento para las dos épocas realizadas.

Época	Pasos	Tiempo	Pérdida	Accuracy
1	1250	30024 s / 8.34 h	0.2658	0.8891
2	1250	26113 s / 7.25 h	0.0827	0.9725
3	2500	56137 s / 15.59 h	0.0827	0.9725

Cuadro 4.2: Resumen entrenamiento modelo BERT

5 Resultados

5.1. Resultados con datos de prueba

En la tabla 5.1 se presenta el reporte de clasificación en donde la medida de recall corresponde a la media de cuantos elementos de cada clase fueron encontrados sobre el total de elementos de cada una de las clases. La precisión corresponde al número total de elementos correctamente clasificados en las dos clases, el f1-score corresponde a la media armónica entre precision y recall. El valor support corresponde a las ocurrencias de una clase dada de acuerdo con el conjunto de datos de entrada, es decir, que tan bien balanceados se encuentran los datos para cada uno de los tipos de reseña, positiva o negativa.

	presicion	recall	f1-score	support
Negative	0.87	0.89	0.88	2443
Positive	0.89	0.88	0.88	2557
accuracy			0.88	5000
macro avg	0.88	0.88	0.88	5000
weighted avg	0.88	0.88	0.88	5000

Cuadro 5.1: Reporte de resultados de predicciones con set de prueba

Posteriormente realizando el análisis sobre la matriz de confusión para los 5000 datos de prueba el cual se presenta en la figura 5.1, se identifica que para la diagonal principal un total de 2171 (89 %) datos fueron clasificados de manera correcta teniendo etiqueta de calificación de crítica negativa y siendo clasificados como tal y un total de 2241 (88 %) fueron clasificados como críticas positivas de manera correcta.

Con respecto a los falsos positivos y los falsos negativos, 316 (12 %) de reseñas marcadas originalmente como positivas fueron clasificadas como negativas (falsos negativos), y 271 (11 %) de reseñas marcadas de manera original como negativas fueron clasificadas como positivas (falsos positivos) generando así los datos de la matriz de confusión.

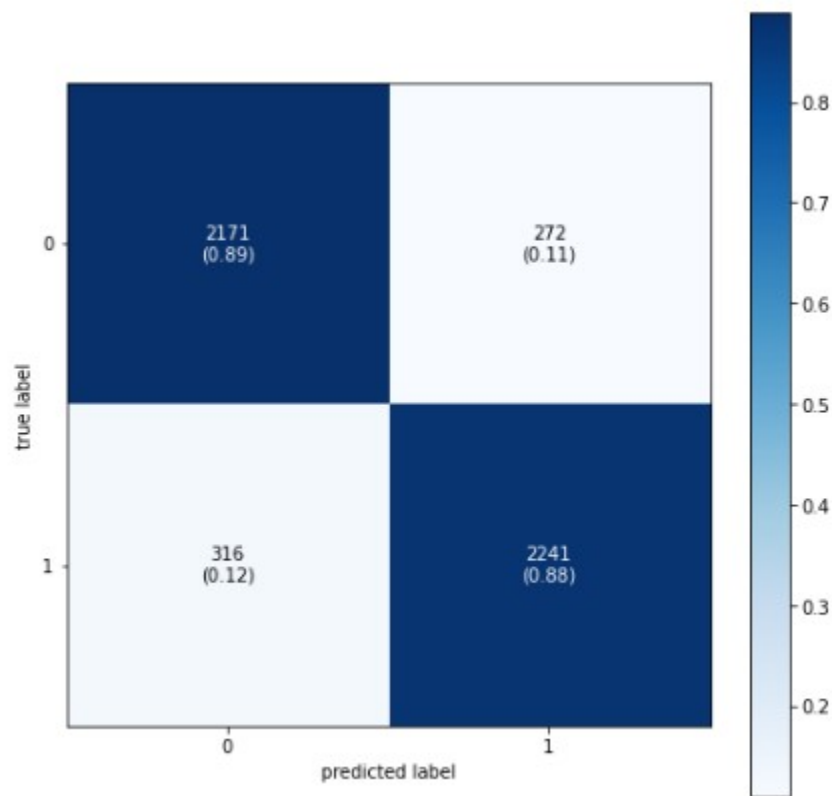


Figura 5.1: Matriz de confusión

5.2. Predicciones

Con el objetivo de realizar la generación de predicciones para verificar el funcionamiento del modelo resultante entrenado, se realizaron diferentes predicciones con datos reales de la nueva versión de la película Dune la cual ha tenido una acogida en su mayoría positiva y cuanta actualmente con una calificación de 8.3 / 10. Se tomaron dos referencias para cada uno de los dos grupos de asignación, donde las reseñas negativas corresponden a aquellas con calificación de 1 a 4, y reseñas positivas con calificación de 7 a 10.

En la tabla 5.2 se presenta el consolidado de las 16 predicciones, donde en la primera columna se presenta la crítica de cada usuario, en la segunda la calificación en estrellas correspondiente a la reseña, en la tercera y cuarta los porcentajes de asignación a cada una de las clases realizada por el resultado de la predicción y en la quinta columna la clasificación realizada a la crítica.

Reseña	Calificación	% Negativa	% Positiva	Resultado
They missed it again. Massive visual effects (that are really good) , combined with dour acting within a confusing script. This made for a long , depressing , and boring movie . Perhaps Dune can never be properly made? The book never ends up on the screen . They end this movie in mid stride it'll be a couple of years to see part 2 , if it gets made at all .	1★	99.90	0.095	Negativa
DISAPPOINTING... So glad I have the Original, cause this one STINKS! And to think it's only part one. Please let them cancel part 2 and beyond. Since they seem to be reviving the franchise, let's hope and pray it doesn't get to the Children of Dune.	1★	99.40	0.59	Negativa
I really like Denis Villeneuve as a director, and was truly excited to see his version of Dune. However, this is a drab, morose, self-serious misery. The film is completely lacking in joy or spectacle. Every CGI set-piece is obscure and dark and devoid of detail. Every performance is pompous and sour. The film is two and a half hours long because two hours of it is filmed in slow-motion. Movies like this are supposed to be fun. This was cinematic torture.	2★	99.92	0.072	Negativa
Sorry, but what was that all about? I got lost right at the start and the story (if there was one) went downhill from there! Just who was fighting who? Who are the goodies and who are the baddies? This film was just way too obscure to be serious! Still not sure what Spice is or does! Might be an ok film for those who know the story, but if you don't it's just one big mess! Lucky I didn't pay to this this!	2★	99.91	0.089	Negativa
Dullsville. Rambling meandering walk through bland desert. How in the hell do you make this so dull? Perhaps to stretch out to unnecessary sequels or books of Denis Villeneuve still hasn't graduated film school. Pretentious, self-serving snore. Here's an idea, instead of dead soul dialogue vomiting try...a story of some kind.	3★	99.90	0.090	Negativa

Wanted something to happen. 15 minutes nothing. Half hour nothing The end nothing. Huge snoozefest. Not sure if I watched the se movie as most. If there is a part two I won't be watching.	3★	99.82	0.170	Negativa
No character development didn't explain mostly anything they talked about and not excited to wait two years if they even make a sequel I won't even remember anything that I saw in the first one because it's not a movie that you would rewatch again with all the other choices that we have these days! I would be shocked if they kept making more of these movies! Oh yeah and the girl that plays Mary Jane in Spider-Man, I think might say a few words in this movie at the very end that's it! This movie was the biggest disappointment I have seen I a while. Waste of money and time.	4★	99.90	0.098	Negativa
I haven't read the book nor seen the Lynch version. I found this film empty and way too long. Too much emphasis on the visuals at the expense of the story. Maybe if I had read the book this would have been more enjoyable. I love sci-fi and loved Villeneuve's Arrival and Blade Runner, but this was a chore to get through. A lot of posing and emoting instead of actual storytelling.	4★	99.87	0.122	Negativa
Well i really loved the music and soundtrack of the movie but the storay took too long for a 2h+ movie and even not finished they could cut on a lot of sences from the movie and it would be even too understandable.	7★	98.00	1.991	Negativa
Dune has a handful of remarkable visuals and does a solid job of basically telling the book page by page (with some exceptions.). That, in and of itself, is a huge accomplishment given how Dune shaped so much of pop-culture. However, my big gripe with this film was the worms. They were very underwhelming. My second grips is that the acting sucks. The cast isn't very good. So, I don't love Dune but I don't hate it - and I'd watch a sequel.	7★	99.56	0.433	Negativa

I've never read Frank Herbert's novel or seen any other adaptation of Dune; so I have to take Denis Villeneuve's adaptation as the guide. It's an impressive movie. Not just the visuals, but in the story's complexity. Timothée Chalamet continues to reaffirm himself as one of the greatest actors of his generation. I'm eager to see part 2, as well as David Lynch's adaptation.	8★	0.072	99.92	Positiva
I don't get all the hype for Timothy Chalamet and Zendaya, both are adequate actors but by no means phenomenal and it shows here if actors with more screen presence and charisma you'd have a 9 or even 10. Great cinematography, phenomenal special effects, loved the whole film. Also Dave Bautista made it sound like this is a break out role for him he's in it for 5 minutes and was just average nothing the Rock couldn't have done as good if not better. Loved we got to watch it on HBO for free.. THANKS WARNER BROS!!	8★	0.184	99.81	Positiva
Dune is such an impressive feat of technical filmmaking; I partly believe they filmed this on-location in space. I wish it had a bit more levity, and I wish I was more personally invested in the individual characters, who have very few distinguishing characteristics (other than being portrayed by whatever famous person plays them). Nit-picking aside, this is giant, bold, gorgeous, epic, compelling, exciting, capital-C Cinema at its finest.	9★	0.098	99.90	Positiva
Modern master Denis Villeneuve delivers a masterful, spectacular adaptation of Frank Herbert's sci-fi classic. Dune takes its time to build and present its universe, but Villeneuve makes the most out of his canvas, using score, sound, and design to tell stories, and building the atmosphere with Greig Fraser's cinematography. The performances are uniformly excellent with Timothée Chalamet holds his own in his first blockbuster leading role. An immersive cinematic experience.	9★	0.094	99.90	Positiva

I read the books at least 40 yrs ago...and have waited for a depiction that does it justice...this is it...everything about it...the casting was very good... the photography...the way the thopters were portrayed...and groups that are given their import...it's like watching the book come to life...thank you film makers...	10★	0.093	99.90	Positiva
This is one of the best IMAX experiences I've had with an exhilarating soundtrack that flowed so well! I'm glad I got to see this with a good friend named Joe he's great to be around. Anyway Denis Villeneuve directed such a wild ride that enthralls you from start to finish and I would say there's no complaints Dune has everything you'd want in it! I think if anything I'd say I don't understand why so many people get up during the movie there was a big audience a portion of which got up often I'd much prefer being prepared to see everything. I loved Dune worth every penny.	10★	0.099	99.90	Positiva

Cuadro 5.2: Predicciones con reseñas de usuarios de la película Dune

6 Conclusiones

El modelo configurado cuenta con una alta precisión aproximadamente del 97 % (de acuerdo con el resultado de la fase de entrenamiento) al momento de realizar la distinción entre una reseña positiva y una negativa. Pero realizando la verificación con el resultado de la matriz de decisión se puede identificar que con respecto a las etiquetas originales hay un total de 588 reseñas que no son clasificadas de manera correcta y generan un elevado número de falsos positivos y falsos negativos los cuales quisieran ser reducidos para contar con un modelo más exacto.

Al realizar el procedimiento de predicción con las críticas reales de la película Dune se puede evidenciar que a pesar de que las reseñas de 7 estrellas pueden ser consideradas subjetivamente positiva, de acuerdo con la clasificación realizada por el modelo se evidencia que las dos reseñas quedaron clasificadas como negativas a pesar de contar con una clasificación inicial positiva. Realizando una lectura subjetiva de ambas críticas, es inevitable notar que efectivamente a pesar de que la calificación en número de estrellas se considera como alta, los comentarios realizados tienen más una opinión negativa que una positiva; mientras que todas las reseñas de 8, 9 y 10 estrellas fueron clasificadas de forma positiva. Verificando el conjunto de datos de prueba del dataset completo, hay un total de 2496 reseñas que cuentan con un total de 7 estrellas, las cuales son las que podrían estar generando esta diferencia al realizar la clasificación errónea en algunas de las críticas.

Con este resultado es posible concluir que una clasificación de 7 estrellas podría aún no ser considerada dentro de la categoría de críticas positivas, pero nos daría también la posibilidad de considerar el rango de las críticas negativas en algunos puntos más, ya que a pesar que una película pueda contar con un número de estrellas elevado, de acuerdo a diferentes estudios en las ciencias cognitivas suelen concluir que el cerebro de las personas cuentan con un método de defensa en el que aproximadamente se asocian los diferentes elementos del día a día en una relación 70 – 30 donde el primer valor corresponde a los aspectos negativos contra el segundo que corresponden a los positivos. Por lo que este factor puede ser el que ocasione que una clasificación con un valor de 7 pueda contarse como una crítica negativa en lugar de una positiva.

7 Trabajo Futuro

En primera instancia, como trabajo futuro puede considerarse realizar modificación en el conjunto de datos de entrenamiento, en el que al realizar la modificación del conjunto de datos inicial, realizando la exclusión de las calificaciones de 7 estrellas como clasificaciones positivas para así volver a realizar el entrenamiento completo del modelo y verificar nuevamente con la matriz de confusión si hubo una mejora al momento de realizar las clasificaciones posterior a esta exclusión y verificar si la precisión del momento aumenta con este ajuste.

Como posterior implementación, puede realizarse la implementación de otros dataset con un mayor número de etiquetas de emoción, debido a que con las críticas solo se cuenta con valores positivos y negativos, pero los diferentes modelos de emociones manejan un mayor número de clasificaciones emocionales en diferentes rangos, de acuerdo con la intensidad para así realzar el cambio de un modelo binario a un modelo con un mayor rango de emociones a clasificar y así verificar si el modelo BERT llega a contar con una mejor, menor o igual tasa de exactitud.

Bibliografía

- Bandhakavi, A., Wiratunga, N., Padmanabhan, D., and Massie, S. (2017). Lexicon based feature extraction for emotion text classification. *Pattern Recognition Letters*, 93:133–142.
- Baziotis, C., Nikolaos, A., Chronopoulou, A., Kolovou, A., Paraskevopoulos, G., Ellinas, N., Narayanan, S., and Potamianos, A. (2018). NTUA-SLP at SemEval-2018 Task 1: Predicting Affective Content in Tweets with Deep Attentive RNNs and Transfer Learning. pages 245–255.
- Ge, S., Qi, T., Wu, C., and Huang, Y. (2019). THU_NGN at SemEval-2019 Task 3: Dialog Emotion Classification using Attentional LSTM-CNN. pages 340–344.
- Ghanbari-Adivi, F. and Mosleh, M. (2019). Text emotion detection in social networks using a novel ensemble classifier based on Parzen Tree Estimator (TPE). *Neural Computing and Applications*, 31(12):8971–8983.
- Ibrahiem, S. S., Ismail, S. S., Bahnasy, K. A., and Aref, M. M. (2020). A Case Study in Multi-Emotion Classification via Twitter. *2020 12th International Conference on Electrical Engineering, ICEENG 2020*, pages 115–120.
- Patwa, P., Aguilar, G., Kar, S., Pandey, S., PYKL, S., Gambäck, B., Chakraborty, T., Solorio, T., and Das, A. (2020). SemEval-2020 Task 9: Overview of Sentiment Analysis of Code-Mixed Tweets.
- Sailunaz, K. and Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. *Journal of Computational Science*, 36:101003.
- Tao, W. (2020). Emotional classification model of review text based on topic and sentiment characteristics. *Proceedings - 2020 2nd International Conference on Economic Management and Model Engineering, ICEMME 2020*, pages 1057–1061.
- Wu, C., Wu, F., Liu, J., Yuan, Z., Wu, S., and Huang, Y. (2018). THU_NGN at SemEval-2018 Task 1: Fine-grained Tweet Sentiment Intensity Analysis with Attention CNN-LSTM. pages 186–192.
- Yasmina, D., Hajar, M., and Hassan, A. M. (2016). Using YouTube Comments for Text-based Emotion Recognition. *Procedia Computer Science*, 83(Ant):292–299.