

Análisis de sentimientos para predecir el movimiento del mercado de acciones farmacéutico hindú a partir de noticias en Twitter

Nicho Galagarza, Jorge
Computer Science
201810205 100%

Morales Panitz, Alexander
Computer Science
202020195 100%

Vásquez Auqui, Fabrizio
Computer Science
201810423 100%

Abstract—La influencia de las noticias en la predicción de la compra y venta de acciones se ha convertido en un factor crítico en los mercados financieros. Las noticias desempeñan un papel fundamental en la transmisión de información sobre indicadores económicos, resultados empresariales, políticas gubernamentales y acontecimientos mundiales. Las noticias positivas pueden generar optimismo entre los inversores y provocar una subida de las cotizaciones bursátiles, mientras que las noticias negativas pueden desencadenar ventas masivas y provocar caídas de las cotizaciones. La evaluación oportuna y precisa de el sentimiento provocado por estas se ha convertido en un factor crucial para comprender y predecir los movimientos bursátiles. Al incorporar el análisis del sentimiento de las noticias a los modelos de predicción bursátil, los inversores pueden comprender mejor la dinámica del mercado y tomar decisiones de inversión más informadas. En este trabajo se subraya la importancia de tener en cuenta el sentimiento de las noticias en la predicción bursátil y destaca el valor de un análisis preciso de las noticias para optimizar las estrategias de inversión y mitigar los riesgos financieros. De esta manera, dentro del presente documento se recupera una base de datos de noticias en tweets que impactan el mercado farmacéutico, se extraen las características necesarias, y se analizan los efectos de los sentimientos de estas noticias en el mercado de acciones. Nuestra experimentación parte del desarrollo de tres modelos de análisis de sentimientos, y finaliza con la comparación entre los modelos: GRU, Transformers, y BERT. Dentro de nuestros resultados logramos un *accuracy* del 78,1% en la predicción de las tendencias en el movimiento del precio de las acciones a corto plazo, utilizando un modelo GRU. Paralelamente, usando un modelo de Transformers, logramos un *accuracy* del 81,4%. Mientras que, con un modelo BERT, alcanzamos un *accuracy* del 82,5%.

Keywords— mercado de acciones, noticias en tweets, GRU, Transformers, BERT, análisis de sentimientos

I. INTRODUCCION

El mercado de acciones es un escenario en constante movimiento y transformación, donde inversores, empresas y personas comunes convergen en busca de oportunidades financieras y de crecimiento. Este mercado, que ha evolucionado a lo largo de los años, ofrece un espacio en el cual diversos actores tienen la posibilidad de invertir y aprovechar los beneficios que brinda.

En primer lugar están los inversores profesionales. Estos individuos o entidades destinan su capital con el objetivo de obtener ganancias a través de la compra y venta de acciones,

ya sea a corto o largo plazo. Su participación en el mercado impulsa la liquidez y la dinámica de los mercados, generando movimientos y tendencias que reflejan las expectativas y decisiones estos. En segundo lugar están las empresas, que también encuentran en el mercado de acciones una fuente de financiamiento esencial. Al cotizar en el mercado de acciones, estas pueden emitir acciones o bonos y captar capital de los inversores interesados en su negocio. Esta financiación les permite expandirse, realizar inversiones estratégicas, innovar y cumplir con sus objetivos de crecimiento [1]. Por último, y gracias tanto los avances tecnológicos como a la aparición de plataformas de inversión accesibles, las personas comunes también pueden participar en el mercado de acciones. Esta democratización de la inversión ha permitido que individuos con diferentes niveles de conocimiento financiero y recursos económicos puedan incursionar en el mercado de acciones y beneficiarse de las oportunidades que brinda. Esta interacción entre los inversionistas, empresas, y personas comunes crea un ambiente dinámico y competitivo que impulsa el desarrollo económico y financiero [1].

Durante mucho años los inversores han buscado obtener una visión más completa de las oportunidades y riesgos en el mercado de acciones, y mejorar sus posibilidades de éxito [1]. Un factor importante para este análisis es revisar constantemente las últimas noticias relacionadas a la empresa en la que se desea invertir, ya que la percepción del mercado a partir de esta puede impactar de manera positiva o negativa, y las redes sociales como *Twitter* son medios rápidos y masivos para las últimas tendencias. El análisis fundamental y técnico son dos enfoques utilizados para predecir los movimientos en el mercado de acciones y tomar decisiones de inversión informadas. El análisis fundamental se basa en el estudio de los fundamentos financieros de una empresa y los factores externos que pueden afectar su valor, mientras que el análisis técnico se centra en el estudio de los patrones y tendencias de precios para predecir futuros movimientos en el mercado. Sin embargo, en las últimas décadas la inteligencia artificial ha demostrado ser una herramienta revolucionaria en diversas áreas de la vida moderna, y los mercados financieros no son la excepción [2].

En particular, el uso de *machine learning* para predecir el movimiento del mercado de acciones ha despertado un gran interés en la comunidad financiera. La capacidad de anticipar las fluctuaciones de los precios de las acciones, las tendencias del mercado y los eventos económicos clave puede marcar la diferencia entre el éxito y el fracaso en el mundo de las inversiones. En esta era de enormes cantidades de datos y avances tecnológicos, las técnicas de *deep learning* y el procesamiento del lenguaje natural se han convertido en aliados poderosos para los inversores y los analistas financieros [2].

A partir de aquí existe el constante interés en desarrollar y aplicar nuevos métodos que permitan predecir el comportamiento del mercado de acciones. Por lo tanto, en este documento se plantea comparar la eficacia de tres modelos de *machine learning* que buscan predecir el sentimiento en usuarios a partir de noticias en *Twitter* relacionadas a empresas que operan en el mercado de acciones, y como estas impactan en las cotización de acciones. Los modelos elegidos son: GRU, Transformers, y BERT. Lo que resta del documento se divide en las siguientes secciones: Sección 2, enfocada en el marco teórico, en donde se explica la arquitectura de cada modelo. La sección 3 muestra la experimentación, y la sección 4 se muestran los resultados a partir del análisis.

II. MARCO TEÓRICO

1) *Modelo GRU*: Las Unidades Recurrentes Aseguradas (GRU, por sus siglas en inglés de Gated Recurrent Units) son una variante de las redes neuronales recurrentes (RNN), diseñadas para mejorar la eficacia de las RNN en el manejo de secuencias de datos [3]. Las RNN son útiles para procesar secuencias de datos porque pueden mantener información en 'memoria' a lo largo del tiempo. Sin embargo, tienen dificultades con las secuencias largas debido al problema del gradiente que se desvanece, que es cuando los gradientes tienden a acercarse a cero durante el entrenamiento, lo que hace que la red sea incapaz de aprender conexiones a largo plazo.

Las GRU, al igual que las unidades de memoria a largo y corto plazo (LSTM), abordan este problema de una manera bastante ingeniosa. Ambos tipos de redes neuronales introducen mecanismos de "compuertas" que regulan el flujo de información a través de la red, pero las GRU lo hacen de una manera más simplificada que las LSTM [3].

Las GRU tienen dos tipos de compuertas: una "compuerta de actualización" y una "compuerta de reinicio". La compuerta de actualización determina qué parte de la información anterior se conservará para el siguiente paso. La compuerta de reinicio, por otro lado, decide cuánta información del estado anterior se debe olvidar [3]. En otras palabras, estas compuertas permiten a la GRU decidir qué información es relevante para mantener a lo largo del tiempo y qué información puede descartarse.

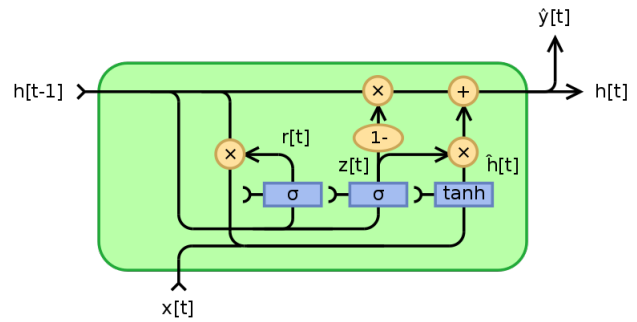


Fig. 1. Arquitectura GRU [3]

Esto es similar a tener un portero que decide qué información es relevante para el futuro y cuál no. Este mecanismo permite a las GRU manejar eficazmente las dependencias a largo plazo en los datos de secuencia, superando así el problema del gradiente que se desvanece que afecta a las RNN estándar [3].

En resumen, las GRU son una mejora de las RNN que permiten un manejo más eficaz de las secuencias de datos, especialmente aquellas con dependencias a largo plazo. Esto las hace útiles en una variedad de aplicaciones, incluyendo el procesamiento del lenguaje natural, la predicción de series temporales, y más.

2) *Modelo Transformer*: Los *Transformers* son un tipo de red neuronal que se diferencia significativamente de las redes neuronales recurrentes (RNN). A diferencia de las RNN, que procesan secuencias de datos paso a paso, los *Transformers* son capaces de procesar toda la secuencia de datos simultáneamente [4]. Esta capacidad de procesamiento paralelo es similar a tener una vista aérea de una situación, lo que permite una visión global de la secuencia de datos en lugar de tener que recorrerla paso a paso.

Una característica distintiva de los *Transformers* es su mecanismo de "atención", que permite al modelo ponderar la importancia relativa de cada parte de la secuencia de datos. Este mecanismo de atención permite a los *Transformers* decidir qué partes de la secuencia son más relevantes para la tarea en cuestión [4]. Es similar a la forma en que, al leer un libro, uno puede resaltar las partes más importantes para comprender la historia, ignorando o dando menos importancia a las partes menos relevantes.

El mecanismo de atención no sólo permite a los *Transformers* manejar eficazmente las dependencias a largo plazo en los datos de secuencia, sino que también les permite manejar secuencias de longitud variable y centrarse en diferentes partes de la secuencia para diferentes tareas. Esto ha demostrado ser especialmente útil en tareas de procesamiento del lenguaje natural, como la traducción automática, donde diferentes partes de la secuencia de entrada

pueden ser relevantes para diferentes partes de la secuencia de salida [4].

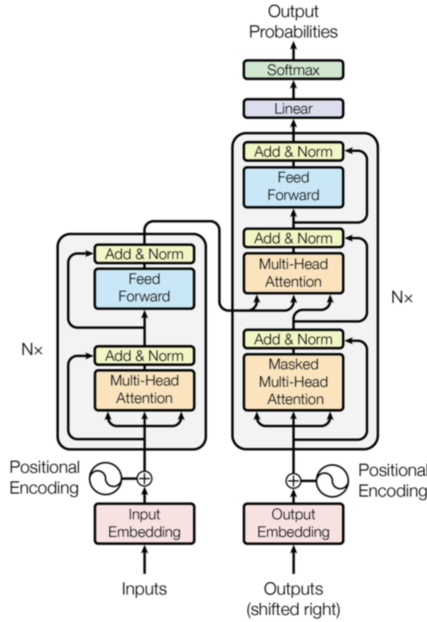


Fig. 2. Arquitectura Transformer [4]

Los *Transformers* han demostrado ser muy eficaces en una serie de tareas de procesamiento del lenguaje natural, superando a menudo a las RNN y a las GRU en términos de rendimiento. Además de la traducción automática, los *Transformers* también han sido exitosos en tareas como el resumen de texto, la generación de texto y la respuesta a preguntas, entre otras [4].

En resumen, los *Transformers* son un tipo de red neuronal que utiliza un mecanismo de atención para procesar eficazmente las secuencias de datos, lo que los hace especialmente útiles para tareas de procesamiento del lenguaje natural.

3) *Modelo BERT*: BERT, acrónimo de Bidirectional Encoder Representations from Transformers, es una evolución de los modelos *Transformer* que introduce una importante mejora: la capacidad de analizar el contexto de una palabra en ambas direcciones [5]. A diferencia de los modelos anteriores que solo consideraban el contexto a la izquierda o a la derecha de una palabra para entender su significado, BERT examina el contexto en ambas direcciones. Esto es similar a leer una oración no solo de izquierda a derecha, sino también de derecha a izquierda para obtener una comprensión más completa de cada palabra [5].

Lo que distingue a BERT es su capacidad para ser pre-entrenado en un gran corpus de texto y luego afinado para tareas específicas. Este pre-entrenamiento implica aprender representaciones de palabras a partir de millones de oraciones,

lo que permite a BERT capturar una gran cantidad de información sobre el lenguaje [5]. Luego, este modelo pre-entrenado puede ser afinado para tareas específicas, como la predicción de sentimientos, añadiendo una capa de salida adicional y entrenando el modelo en un conjunto de datos específico para esa tarea [5].

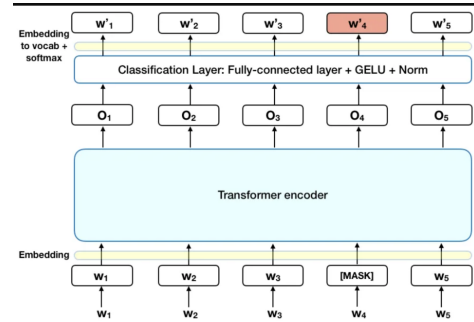


Fig. 3. Arquitectura BERT [5]

En el caso que nos ocupa, estamos utilizando BERT para entender los sentimientos expresados en los *tweets* y cómo estos podrían afectar las inversiones en el mercado de acciones. BERT es capaz de capturar el matiz y el contexto de cada tweet, lo que permite un análisis de sentimiento más preciso que los métodos tradicionales. Al entender los sentimientos expresados en los *tweets*, podemos obtener una visión más profunda de las percepciones del mercado y, potencialmente, predecir las tendencias del mercado de acciones.

En resumen, BERT es un modelo de procesamiento del lenguaje natural altamente avanzado que utiliza *Transformers* bidireccionales para entender el contexto de las palabras desde ambas direcciones. Su capacidad para ser pre-entrenado y luego afinado para tareas específicas lo hace extremadamente versátil y poderoso para una amplia gama de tareas de procesamiento del lenguaje natural.

A. Métricas

1) *Accuracy*: Es un métrica que brinda información sobre cuanto es el número de predicciones correctas, sobre las predicciones totales [6]. *Accuracy* responde a la pregunta "De todas las predicciones que hicimos, ¿cuántas fueron ciertas?".

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

2) *Precision*: Es una métrica que brinda la proporción de verdaderos positivos en base a la cantidad de positivos totales que predice el modelo [6]. *Precision* responde a la pregunta "De todas las predicciones positivas que hicimos, ¿cuántas fueron ciertas?".

$$Precision = \frac{TP}{TP + FP}$$

3) *Recall*: Se enfoca en qué tan bueno es el modelo para encontrar todos los aspectos positivos. La recuperación también se denomina tasa de verdaderos positivos [6] y responde a la pregunta "De todos los puntos de datos que

deberían predecirse como verdaderos, ¿cuántos predijimos correctamente como verdaderos?”.

$$Recall = \frac{TP}{TP + FN}$$

4) *F1 Score*: Es una medida que combina memoria y precisión. Como hemos visto, existe una compensación entre *precision* y *recall*, por lo que *F1 score* se puede usar para medir la eficacia con la que nuestros modelos logran esa compensación [6].

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

5) *MCC*: Es una tasa estadística confiable que produce una puntuación alta solo si la predicción obtuvo buenos resultados en las cuatro categorías de la matriz de confusión (verdaderos positivos, falsos negativos, verdaderos negativos y falsos positivos) proporcionalmente tanto al tamaño de los elementos positivos como al tamaño de los elementos negativos en el conjunto de datos [7].

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

6) *Kappa*: En los modelos de clasificación binaria, uno de los evaluadores asume el rol del modelo de clasificación, mientras que el otro evaluador asume el rol de un observador del mundo real que conoce las verdaderas clasificaciones de cada registro o conjunto de datos. El *Kappa de Cohen* tiene en cuenta la frecuencia con la que los evaluadores están de acuerdo (positivos y negativos reales) y en desacuerdo (falsos positivos y falsos negativos). Después de considerar el azar, *Cohen* y *Kappa* pueden determinar acuerdo total y acuerdo. La puntuación *Kappa de Cohen* se puede definir como la métrica utilizada para medir el rendimiento de los modelos de clasificación de aprendizaje automático basados en la evaluación del acuerdo perfecto y el acuerdo al azar entre los dos evaluadores (un observador del mundo real y el modelo de clasificación) [8].

$$Kappa(k) = \frac{P_0 - P_e}{1 - P_e}$$

III. EXPERIMENTACIÓN

A. Dataset

Los datos utilizados para entrenar y probar los modelos consisten en un conjunto de 5000 *tweets* relacionados con empresas del sector farmacéutico hindú, también empleamos el mismo criterio para una base de datos adicional sobre sentimientos en general a manera de validar lo anterior. Estos dos conjunto de datos se compone de dos columnas (previamente preprocesado), como se muestra en la Tabla II: una columna que contiene los *tweets* y otra columna que indica el sentimiento asociado, siendo 1 para los positivos y -1 para los negativos. Estos datos se dividieron en un 80% para el entrenamiento y un 20% para la prueba.

Adicionalmente mostramos la distribución del tamaño de palabras.

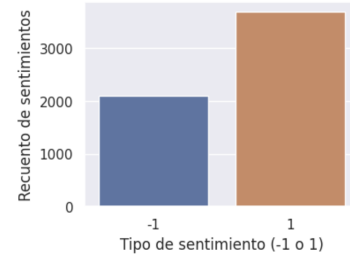


Fig. 4. Conjunto de datos de tweets - Base de datos farmaceutica

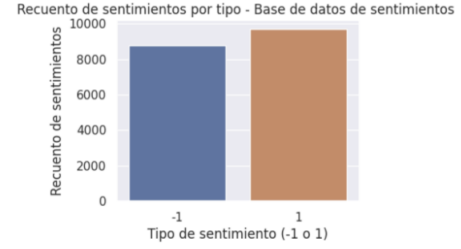


Fig. 5. Conjunto de datos de tweets - Base de datos sentimiento general

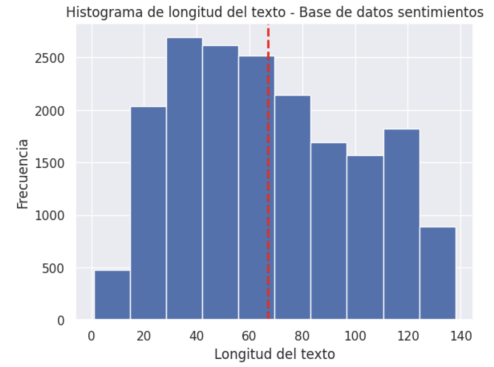


Fig. 6. Conjunto de datos de tweets - Base de datos sentimiento general

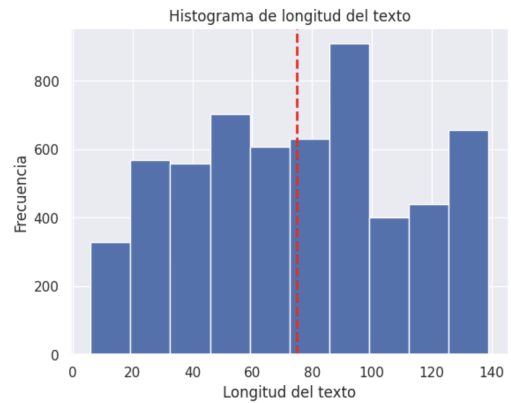


Fig. 7. Distribución del la longitud de los *tweets* de la base de datos farmaceutica

Tweet	Sentimiento
Just watched the new movie. It was absolutely fantastic! Highly recommended.	1
I'm really disappointed with the service at this restaurant. The food took forever to arrive and it was cold. Never coming back again.	-1
The latest product from this company is a game-changer. It has exceeded all my expectations. Great job!	1
I had high hopes for this concert, but it turned out to be a complete disaster. The sound quality was terrible, and the performers seemed unprepared.	-1

TABLE I
MUESTRA DE CONJUNTO DE DATOS DE TWEETS

Tipodesentimiento	Porcentaje
-1	36.36%
1	63.62%

TABLE II
DISTRIBUCIÓN DE DATOS POR EL TIPO DE SENTIMIENTO PARA LA BASE DE DATOS DE TWEETS DE FARMACEÚTICA.

Tipodesentimiento	Porcentaje
-1	47.55%
1	52.44%

TABLE III
DISTRIBUCIÓN DE DATOS POR EL TIPO DE SENTIMIENTO PARA LA BASE DE DATOS DE TWEETS SENTIMIENTOS EN GENERAL.

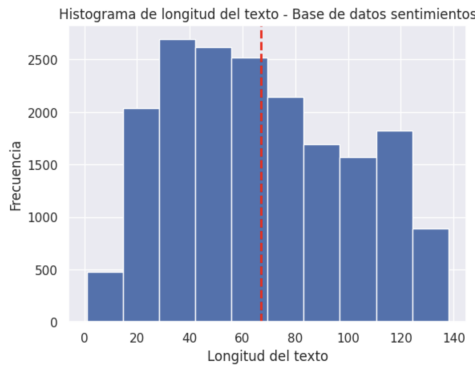


Fig. 8. Distribución de la longitud de los tweets de la base de datos de sentimiento en general

B. Tokenización

Para el proceso de tokenización se utilizó Tokenizer de la librería Tensorflow. Como el objetivo del análisis de sentimiento es poder detectar si es un sentimiento positivo o negativo (binario), se utilizó esta función con la finalidad de tokenizar, es decir generar un índice de palabras basado en el texto proporcionado, asignando un número a cada palabra. El proceso que se abordó para todos los experimentos fue el siguiente:

- Solo consideramos las 5000 palabras más frecuentes, el objetivo es crear las secuencias de *tokens*.
- Luego generamos un índice único para cada palabra basado en el texto.

- Cada secuencia que pasamos representa un número entero.
- Luego acotamos para que todos los *tokens* para que tengan la misma longitud. Completamos esto mediante una función que agrega ceros al principio o final de la secuencia.

Además, para el *target* o sentimiento se utilizó *One-Hot-Encoding*, el objetivo es que soporte un análisis de sentimiento multiclase.

C. Resultados

Los resultados, como se observa en la Tabla I, indican que el modelo BERT supera al modelo *Transformer* en términos de precisión, mientras que ambos son superiores al modelo GRU. El elevado nivel de precisión del modelo BERT se debe a su capacidad para analizar el texto de manera bidireccional. Por otro lado, el modelo *Transformer* es superior al GRU debido a su acceso a información contextual previa al token analizado. En cambio, el modelo GRU no realiza esta tarea. La diferencia de 2.5% entre BERT y *Transformer*, así como el 3% entre *Transformer* y GRU, podría atribuirse al tamaño del tweet analizado. Es posible que para *tweets* o textos más extensos, estas diferencias sean aún más pronunciadas.

Para el caso del modelo de la GRU podemos ver como disminuye la pérdida conforme incrementan las épocas en la siguiente imagen para la primer y segunda base de datos, por lo que llegamos a la conclusiones que bastaba con 8 épocas para no sobreentrenar el modelo.

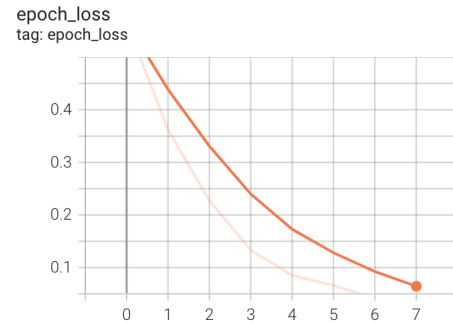


Fig. 9. Función de pérdida *train* y *test* para modelo GRU en base de datos farmacéutica

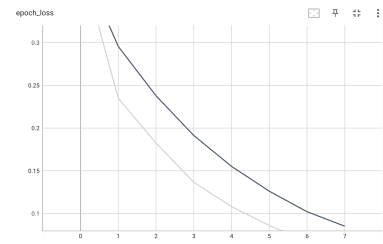


Fig. 10. Función de pérdida *train* y *test* para modelo GRU en base de datos farmacéutica

En cuanto a la precisión, también confirmamos que el modelo BERT supera a los dos modelos anteriores en un

Modelo	Transformers	Transformer	BERT
Accuracy	0.744607	0.805004	0.80673
Recall	0.744607	0.742243	1
Precision	0.744607	0.724942	1
F1	0.744607	0.733491	1
Kappa	0.244607	0.579883	1
MCC	0.226414	0.579783	1

TABLE IV
MÉTRICAS PARA BASE DE DATOS FARMACEÚTICA

Modelo	GRU	Transformer	BERT
Accuracy	0.876827	0.929616	0.932593
Recall	0.876827	0.942509	1
Precision	0.876827	0.909753	1
F1	0.876827	0.925841	1
Kappa	0.324307	0.858905	1
MCC	0.376827	0.859392	1

TABLE V
MÉTRICAS PARA BASE DE DATOS DE SENTIMIENTOS EN GENERAL

4.89%, como se puede apreciar en la Fig. 12. En ese sentido, observamos un patrón similar en el resto de las métricas evaluadas, excepto en *recall*, la precisión y F1, donde el modelo GRU muestra un rendimiento ligeramente superior al modelo *Transformer*, pero no logra superar al modelo BERT.

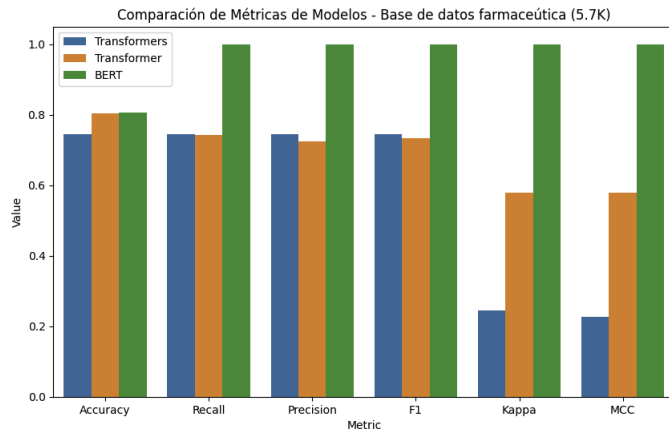


Fig. 11. Al evaluar los modelos GRU, Transformer y BERT mediante diversas métricas, podemos realizar una comparación exhaustiva de sus atributos y rendimiento para la base de datos de farmacéutica.

IV. CONCLUSIONES

- Usando la base de datos farmacéutica: El modelo BERT supera con un 0.1% de *accuracy* al modelo Transformer que presenta un 80.5% de *accuracy*. Y este último supera a GRU por una diferencia de aproximadamente 6% de *accuracy*.
- En la base de datos de sentimientos en general: El modelo BERT supera al modelo Transformer en un 1% aproximadamente y casi en un 6% a GRU.
- BERT es superior a otras arquitecturas modelos. Pese a que parezca que existe una poca diferencia, en un mercado de acciones, un pequeño margen puede significar grandes ganancias.
- Cuanto se tiene mayor data y clases mas balanceadas, se obtienen mejores resultados.

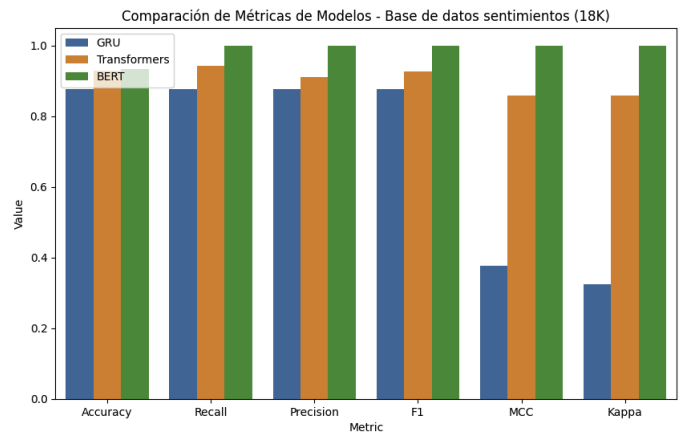


Fig. 12. Al evaluar los modelos GRU, Transformer y BERT mediante diversas métricas, podemos realizar una comparación exhaustiva de sus atributos y rendimiento para la base de datos de sentimientos.

V. REPOSITORIO

Pueden encontrar la implementación con *PyTorch* de los tres modelos en el siguiente link de github: [GitHub](#).

REFERENCES

- [1] E. B. Boukherouaa, K. AlAjmi, J. Deodoro, A. Farias, and R. Ravikumar, "Powering the digital economy: Opportunities and risks of artificial intelligence in finance," *Departmental Papers*, vol. 2021, no. 024, p. A001, 2021.
- [2] I. Parmar, N. Agarwal, S. Saxena, R. Arora, S. Gupta, H. Dhiman, and L. Chouhan, "Stock market prediction using machine learning," in *2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC)*, pp. 574–576, 2018.
- [3] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," 2014.
- [4] T. Lin, Y. Wang, X. Liu, and X. Qiu, "A survey of transformers," 2021.
- [5] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2019.
- [6] R. Yacoubby and D. Axman, "Probabilistic extension of precision, recall, and f1 score for more thorough evaluation of classification models," in *Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems*, (Online), pp. 79–91, Association for Computational Linguistics, Nov. 2020.
- [7] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, 01 2020.
- [8] S. M. Vieira, U. Kaymak, and J. M. C. Sousa, "Cohen's kappa coefficient as a performance measure for feature selection," in *International Conference on Fuzzy Systems*, pp. 1–8, 2010.