



Universidad del
Rosario

Escuela de Ciencias
e Ingeniería

PREDICCIÓN DEL COMPORTAMIENTO DE UN ÍNDICE BURSÁTIL USANDO DEEP LEARNING Y PROCESAMIENTO DE LENGUAJE NATURAL

Presentado para obtener el título de

PROFESIONAL EN MATEMÁTICAS APLICADAS Y CIENCIAS DE LA COMPUTACIÓN

Samuel David Moreno Vahos

Dirección:

Fabián Sánchez Salazar

Universidad del Rosario

Escuela de Ciencias e Ingeniería

Matemáticas Aplicadas y Ciencias de la Computación

Julio de 2025

DEDICATORIA

A mis padres, por todo su apoyo.

A Pelusita y Kitty, por su compañía.

Y especialmente, a mi Cori, a quien
sigo recordando con mucho cariño.

AGRADECIMIENTOS

Al profesor Fabián Sánchez Salazar, por su paciencia, su disposición para resolver mis dudas y por la confianza que depositó en mi trabajo.

A mi novia, Natalia Martínez Neira, por ser mi apoyo incondicional y por animarme a seguir adelante. Por haber sido mucho más que compañía durante este proceso. Le agradezco también por su valiosa ayuda en la estructuración y presentación de este documento, y por acompañarme tantas tardes en el CRAI y en el Laboratorio de Finanzas.

A mis amigos, Jaider Duván Velasco Díaz y Juan Andrés Alayón Ariza, por sus valiosos comentarios y sugerencias, y por el tiempo que dedicaron a revisar las distintas versiones de esta tesis.

En general, agradezco a todas las personas que, de una u otra manera, me apoyaron durante esta etapa de mi vida.

RESUMEN

En este trabajo de grado se estudió la predicción del comportamiento del índice bursátil colombiano MSCI COLCAP mediante técnicas de aprendizaje profundo y procesamiento de lenguaje natural. Para esto, se entrenaron redes neuronales LSTM con los precios y volúmenes históricos del índice, la tasa de cambio representativa del mercado (TRM) y el sentimiento de noticias económicas y financieras. Con estos datos, se predijeron los cambios en el precio de cierre del COLCAP con 1, 2 y 3 días de anticipación. Al mismo tiempo, se implementaron modelos tanto de regresión como de clasificación, evaluando diferentes combinaciones de variables y ventanas temporales.

Los modelos de regresión no generalizaron bien, ya que presentaron sobreajuste, valores negativos de R^2 en prueba y errores absolutos mayores al cambio promedio del índice. En contraste, al simplificar el problema como una tarea de clasificación binaria (aumento o disminución del COLCAP), los modelos exhibieron métricas de exactitud consistentes. También se destaca que la inclusión del análisis de sentimiento y de la TRM aportó mejoras relevantes en el desempeño, sobre todo para predicciones a uno y tres días en el futuro. Además, al simplificar los datos de entrada (utilizando únicamente sus signos), se redujo el ruido y se mejoró la generalización de los modelos, obteniendo valores de exactitud entre 52% y 59% en datos desconocidos, especialmente con ventanas de 20 a 50 días. Finalmente, se concluye que predecir la dirección del mercado, es decir, si aumenta o disminuye el valor del COLCAP, es más viable que estimar con precisión la variación exacta de este índice, y que el sentimiento de mercado aporta información valiosa para las predicciones.

ABSTRACT

In this undergraduate thesis, the prediction of the behavior of the Colombian stock market index MSCI COLCAP was studied using deep learning techniques and natural language processing. To this end, LSTM neural networks were trained with the historical prices and volumes of the index, the market representative exchange rate (TRM, for its acronym in Spanish), and the sentiment of economic and financial news articles. With these data, changes in the COLCAP's closing price were predicted at 1, 2, and 3 days in advance. At the same time, both regression and classification models were implemented, evaluating different combinations of variables and sliding time windows.

The regression models did not generalize well, as they showed overfitting, negative R^2 values on the test set, and absolute errors greater than the average change of the index. In contrast, by simplifying the problem into a binary classification task (increase or decrease of the COLCAP), the models showed consistent accuracy metrics. It is also highlighted that the inclusion of sentiment analysis and TRM provided relevant improvements in performance, especially for predictions one and three days in the future. Furthermore, by simplifying the input data (using only their signs), noise was reduced and generalization of the models was improved, obtaining accuracy scores between 52% and 59% on unknown data, especially with windows of 20 to 50 days. Finally, it is concluded that predicting the direction of the market, that is, whether the value of the COLCAP increases or decreases, is more feasible than precisely estimating the exact variation of the index, and that market sentiment provides valuable information for predictions.

TABLA DE CONTENIDO

1	INTRODUCCIÓN	1
2	OBJETIVOS	3
2.1	Objetivo general	3
2.2	Objetivos específicos	3
3	PROBLEMA Y JUSTIFICACIÓN	4
4	MARCO TEÓRICO Y ESTADO DEL ARTE	5
4.1	Series de tiempo	5
4.2	Redes neuronales	6
4.3	Indicadores económicos	7
4.4	Análisis técnico	9
4.5	Procesamiento de lenguaje natural	11
4.6	Métricas de evaluación	11
4.7	Revisión de literatura	13
5	METODOLOGÍA	15
5.1	Recolección de datos	15
5.2	Limpieza de datos	16
5.3	Análisis de sentimiento	17
5.4	Análisis de estacionariedad	18
5.5	Modelamiento	19
6	RESULTADOS Y DISCUSIÓN	21
7	CONCLUSIONES Y RECOMENDACIONES	29
	BIBLIOGRAFÍA	30
	ANEXO	34

LISTA DE TABLAS

4.1	Canasta del MSCI COLCAP	8
6.1	R^2 con predicciones de $t + 1$	21
6.2	R^2 con predicciones de $t + 2$	22
6.3	R^2 con predicciones de $t + 3$	22
6.4	MAE con predicciones de $t + 1$	23
6.5	MAE con predicciones de $t + 2$	23
6.6	MAE con predicciones de $t + 3$	24
6.7	Exactitud con predicciones de $t + 1$	25
6.8	Exactitud con predicciones de $t + 2$	25
6.9	Exactitud con predicciones de $t + 3$	26
6.10	Exactitud con predicciones de $t + 1$ (variables simples)	27
6.11	Exactitud con predicciones de $t + 2$ (variables simples)	27
6.12	Exactitud con predicciones de $t + 3$ (variables simples)	28

LISTA DE FIGURAS

4.1	Funcionamiento de una RNN	6
4.2	Composición interna de una celda LSTM	7
4.3	Velas japonesas	10
4.4	Tendencias de mercado	10
5.1	Gráfico de velas del MSCI COLCAP	15
5.2	Valor histórico de la TRM	16
5.3	Nubes de palabras de las noticias	18
5.4	Gráficos ACF de las variables originales	18
5.5	Gráficos ACF de las variables diferenciadas	19

Capítulo 1

INTRODUCCIÓN

La predicción de activos financieros ha sido objeto de estudio durante décadas. Esto se debe a que los inversionistas tendrían mayores ganancias si pudieran conocer la dirección que tomarán los mercados en el futuro. Debido al interés que existe por resolver este problema, en este trabajo se busca implementar un modelo de aprendizaje profundo para predecir el comportamiento del principal índice bursátil de Colombia, el MSCI COLCAP, y mejorar así la información disponible al momento de tomar decisiones de inversión o formular estrategias financieras.

Ahora bien, la literatura existente sobre predicción en el mercado de valores comprende distintos métodos. Los más comunes utilizan variables financieras como precios de activos o indicadores técnicos, mientras que otros acercamientos también han considerado el análisis del sentimiento de mercado a través de noticias o medidas del riesgo de sus respectivos países. La motivación para este trabajo surge de la limitada investigación en Colombia que incluya una variable de sentimiento de mercado en los modelos predictivos. Es por esto que se analizarán los cambios en el valor del índice COLCAP a partir de noticias financieras y un indicador macroeconómico (la TRM) como variables independientes en los modelos.

Múltiples estudios llevados a cabo en los últimos años sugieren que las redes neuronales son más efectivas que otros algoritmos a la hora de predecir movimientos en los mercados. Asimismo, se hallaron trabajos que defienden las ventajas del uso de redes recurrentes, como las LSTM, para abordar este tipo de problemas. De esta manera, este trabajo se enfoca en generar y optimizar un modelo basado en redes neuronales LSTM, utilizando diferentes ventanas de tiempo e intentando predecir el valor de un índice bursátil no sólo con un día sino con hasta tres días de anticipación.

La metodología de este proyecto inicia con la recolección de datos, en donde se detalla la forma en que se obtuvieron los precios y el volumen del COLCAP de los últimos diez años. Al mismo tiempo, se describen los procesos de extracción del valor histórico de la TRM y de los textos de

las noticias económicas necesarias para aplicar el análisis de sentimiento. En la siguiente sección, se listan los cambios que se tuvieron que llevar a cabo para limpiar los datos recogidos.

Luego de esto, se describe el modelo utilizado para clasificar las noticias como positivas o negativas según su contenido y se realiza un pequeño análisis descriptivo de cada categoría. También se revisa la estacionariedad de las series de tiempo que recibirían los modelos, mostrando las transformaciones que fueron necesarias para mejorar la calidad de los datos. Al final se detalla la implementación y el entrenamiento de los modelos predictivos y se explica la forma en que se estructuró la búsqueda de los hiperparámetros. El desempeño de estos modelos se evaluó tanto para un problema de regresión como para uno de clasificación, discutiendo sus implicaciones y la coherencia con los objetivos planteados.

El resto de este documento se organiza de la siguiente manera: en el siguiente capítulo se plantean los objetivos del proyecto; el capítulo 3 comprende la problemática que dio lugar al estudio junto con el contexto y la importancia de la solución propuesta; en el capítulo 4, se resume la teoría necesaria para poder comprender el problema de investigación y se revisan los antecedentes más relevantes; el capítulo 5 está dedicado a describir la metodología empleada durante el desarrollo del trabajo; y los capítulos 6 y 7 contienen los resultados y conclusiones acompañados de recomendaciones para trabajos futuros.

Capítulo 2

OBJETIVOS

2.1 Objetivo general

Predecir el comportamiento del índice bursátil MSCI COLCAP mediante redes neuronales LSTM y análisis de sentimiento de noticias financieras.

2.2 Objetivos específicos

1. Evaluar el rendimiento de un modelo LSTM en la predicción a corto plazo del índice COLCAP a partir de sus datos históricos.
2. Validar si las predicciones mejoran al incluir el sentimiento hacia el mercado colombiano, extraído de noticias financieras y económicas mediante técnicas de procesamiento de lenguaje natural.
3. Analizar el efecto que tiene la inclusión de la TRM como variable adicional sobre el rendimiento de los modelos para predecir el valor del COLCAP.
4. Comparar el rendimiento de los diferentes modelos para identificar la combinación de variables que mejor se ajuste a la predicción del índice COLCAP.

Capítulo 3

PROBLEMA Y JUSTIFICACIÓN

Los índices bursátiles son un importante punto de referencia para los inversionistas, ya que muestran las condiciones de los mercados accionarios en un momento determinado. La predicción de estos índices supone un gran reto y ha sido objeto de interés desde hace varias décadas, ya que, si se lograra hacer buenos pronósticos, habría una mayor oportunidad de obtener ganancias en estos mercados. En particular, el COLCAP es un índice del mercado colombiano que refleja el valor de las acciones más líquidas y robustas del país, por lo que contar con modelos capaces de predecir su evolución sería de gran utilidad para la toma de decisiones financieras y el diseño de estrategias de inversión.

Un acercamiento que no ha sido muy estudiado hasta ahora en Colombia es el uso del sentimiento de mercado como un recurso adicional para predecir tendencias económicas. Es por esto que en el presente trabajo se busca analizar el efecto que tienen las noticias financieras y los cambios en el dólar americano sobre el valor del índice COLCAP. Al mismo tiempo, se revisará si el sentimiento explicado por estas variables puede mejorar la calidad de las predicciones cuando se presentan valores extremos como en el caso de la pandemia. Este es un punto importante, ya que, de ser así, no habría necesidad de excluir dicho período de las bases de datos, como se hace en la mayoría de las investigaciones actualmente, lo cual dejaría una mayor cantidad de información disponible para ajustar mejor los modelos predictivos.

Capítulo 4

MARCO TEÓRICO Y ESTADO DEL ARTE

4.1 Series de tiempo

Una serie de tiempo es una secuencia de datos ordenados cronológicamente [1]. Estas secuencias son muy utilizadas en diversos campos. Un ejemplo es el ámbito económico, donde se pueden observar índices de precios mensuales, volúmenes de ventas trimestrales, precios diarios de acciones, entre otros [2]. Formalmente, una serie de tiempo es un conjunto de observaciones x_t , cada una registrada en un tiempo t específico.

El análisis de series de tiempo se refiere al uso de las observaciones disponibles para inferir la estructura de los datos desconocidos [1]. Ahora bien, una parte importante de este análisis es la selección de un modelo que represente o que se ajuste bien a los datos observados. De aquí surge la necesidad de definir un modelo para una serie de tiempo en términos de las distribuciones conjuntas (frecuentemente expresadas a través de medias y covarianzas) de una secuencia de variables aleatorias, $\{X_t\}$, de la cual $\{x_t\}$ corresponde a una realización.

En este contexto, el estudio de una serie de tiempo no sólo requiere especificar un modelo en términos de sus distribuciones conjuntas, sino también considerar ciertas propiedades estadísticas que faciliten su análisis. Entre ellas, la más relevante es la estacionariedad, ya que muchos modelos predictivos parten de la suposición de que los datos futuros se comportarán de forma similar a los datos pasados. En términos generales, una serie de tiempo discreta $\{X_t\}$ se considera estacionaria si sus propiedades estadísticas son similares a las de la serie rezagada $\{X_{t+h}\}$ para cualquier entero h . Si estas propiedades se restringen a la media y la covarianza, se obtiene la siguiente definición:

$\{X_t\}$ es estacionaria si

- (i) $E(X_t)$ es independiente de t
- (ii) $Cov(X_t, X_{t+h})$ es independiente de t para cualquier h

En términos sencillos, una serie es estacionaria si tiene un promedio constante y la relación entre dos observaciones sólo depende de la distancia temporal entre ellas. Ahora bien, unas de las herramientas que permiten revisar si estas condiciones se cumplen para $\{X_t\}$ son la función de autocovarianza, $\gamma_h = Cov(X_t, X_{t+h})$, y la función de autocorrelación (ACF, por sus siglas en inglés), $\rho_h = \frac{\gamma_h}{\gamma_0} = Cor(X_t, X_{t+h})$, ya que, dependiendo de la forma que tomen las funciones al ser evaluadas con distintos valores de h , es posible determinar si una serie de tiempo ya es estacionaria o si es necesario aplicar alguna transformación sobre los datos, como diferenciaciones, para mejorar la fiabilidad de los modelos predictivos [3].

4.2 Redes neuronales

Las Redes Neuronales Recurrentes (RNN, por sus siglas en inglés) son una familia de arquitecturas especializadas en el procesamiento de secuencias de datos, ya que están compuestas por celdas que utilizan información de entradas anteriores para generar las salidas y comparten los mismos parámetros a través de diferentes partes de las secuencias recibidas (véase la figura 4.1). Las RNN son una generalización de los clásicos modelos lineales para series de tiempo, además de brindar un acercamiento robusto para la predicción de series de tiempo financieras [4].

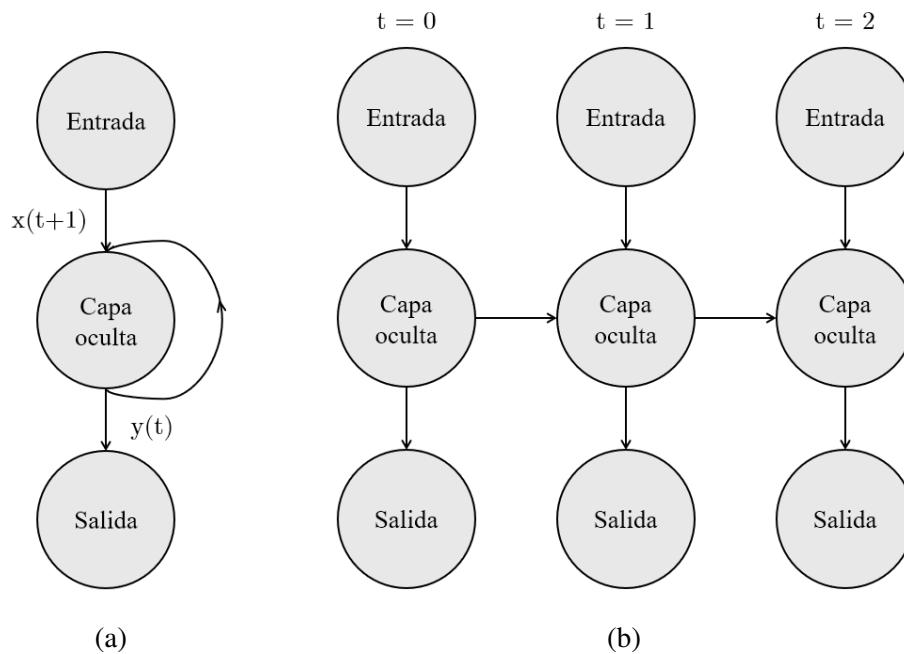


Figura 4.1: Funcionamiento de una RNN
 (a) Diagrama con recurrencia, (b) Diagrama desglosado. (Adaptado de [5])

Dentro del grupo de las RNN, se encuentran las LSTM (Long Short-Term Memory), las cuales sirven para procesar secuencias de datos a largo plazo y aprender patrones temporales de manera robusta y fueron diseñadas para evitar la pérdida de información que se producía con arquitecturas más simples [6].

Las LSTM tienen como base la creación de caminos a través del tiempo donde no hay problemas de desvanecimiento o explosión de gradientes en la red. Las celdas de las LSTM tienen las mismas entradas y salidas que una RNN regular, con la adición de que cada celda contiene puertas de entrada, de olvido y de salida (véase la figura 4.2) que regulan el flujo de información a través de una unidad de memoria: esta última acumula o “recuerda” información si la puerta de entrada lo permite y la borra u “olvida” según el estado de la puerta de olvido, permitiendo a la red aprender qué recordar y qué olvidar a lo largo del tiempo [7], [5].

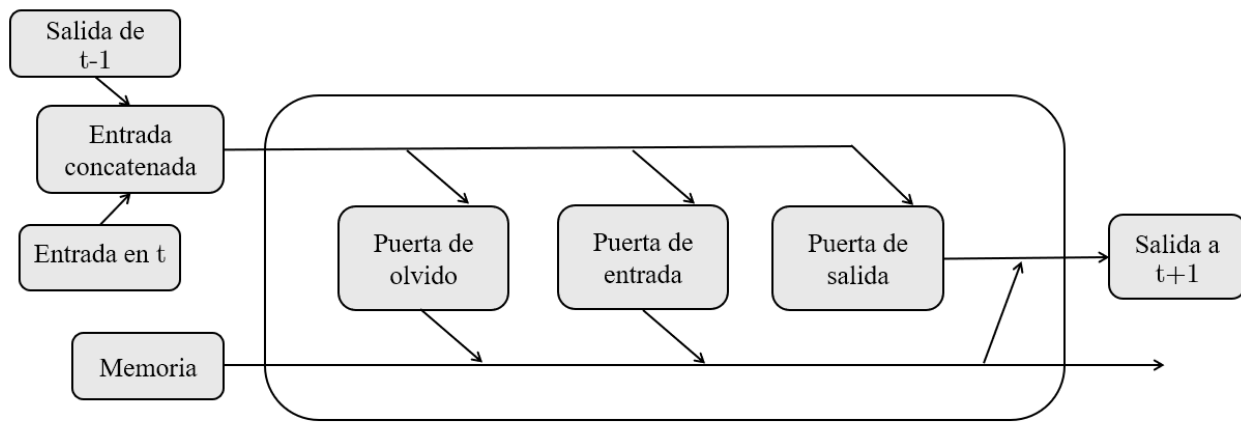


Figura 4.2: Composición interna de una celda LSTM
(Adaptado de [5])

4.3 Indicadores económicos

Un índice bursátil mide el desempeño (en términos financieros) de un grupo de títulos o activos representativos como acciones, bonos de deuda, materias primas, entre otros [8]. Al brindar una medida de referencia que sigue el comportamiento en bolsa de una canasta de empresas, los índices bursátiles permiten comprender mejor la situación económica de un país o un sector de producción específico.

Para calcular el valor numérico de un índice, primero se seleccionan los activos financieros que lo van a componer (acciones, por ejemplo) de acuerdo al agregado económico que se busque

representar. Después, se halla el valor de mercado de cada acción, el cual es igual al producto entre el precio de dicha acción y el número total de acciones en circulación. Luego se define un valor inicial sin unidades para el índice (por convención, casi siempre es 100 o 1000) y, a partir de ahí, el valor sube o baja en función de las variaciones en los precios de las acciones [9]. El peso que tiene la variación de cada acción dentro del índice se define según los valores de mercado o según la importancia que tenga en el sector económico que se esté analizando y, cada cierto tiempo, se pueden hacer ajustes como agregar o quitar empresas del índice.

En particular, Colombia cuenta con varios índices, entre los cuales el MSCI COLCAP es el más representativo del mercado nacional, ya que refleja el valor de las acciones más líquidas del país (véase la tabla 4.1) y su canasta se compone de las 20 empresas más robustas e influyentes de la Bolsa de Valores de Colombia [8].

Cuadro 4.1: Canasta del MSCI COLCAP
 Datos extraídos de [10], el 29 de noviembre del 2024

Nemotécnico	Emisor	Peso (%)
PFBCOLO	Bancolombia S.A.	20.134
BCOLO	Bancolombia S.A.	12.169
ECOPETROL	Ecopetrol S.A.	10.992
ISA	Interconexión Eléctrica S.A. E.S.P.	9.213
GEB	Grupo Energía Bogotá S.A. E.S.P.	8.795
GRUPOARGOS	Grupo Argos S.A.	7.209
CEMARGOS	Cementos Argos S.A.	7.051
PFGRUPOARG	Grupo Argos S.A.	3.453
PFGRUPSURA	Grupo de Inversiones Suramericana	3.247
PEI	Patrimonio Autónomo de Estrategias Inmobiliarias	2.938
PFAVAL	Grupo Aval Acciones y Valores S.A.	2.335
PFDAVVNDA	Banco Davivienda S.A.	2.316
CELSIA	Celsia S.A.	2.114
CORFICOLCF	Corporación Financiera Colombiana S.A.	1.651
GRUPOSURA	Grupo de Inversiones Suramericana	1.492
GRUPOBOLIVAR	Grupo Bolívar S.A.	1.221
PROMIGAS	Promigas S.A. E.S.P.	1.082
BOGOTA	Banco de Bogotá S.A.	0.846
MINEROS	Mineros S.A.	0.626
TERPEL	Organización Terpel S.A.	0.503
CNEC	Canacol Energy Ltd	0.306
PFCORFICOL	Corporación Financiera Colombiana S.A.	0.289
ETB	Empresa de Telecomunicaciones de Bogotá S.A. E.S.P.	0.020

Con respecto al cálculo del COLCAP, la ponderación de las acciones se determina a partir de sus valores de mercado y la recomposición de su canasta se hace cada año en el último día hábil de noviembre. Asimismo, los pesos se actualizan de manera trimestral para reflejar cambios que haya podido tener cada empresa en su capitalización [11].

De forma similar, otra variable importante para Colombia es la tasa de cambio representativa del mercado (TRM), que corresponde a la cantidad de pesos colombianos equivalentes a un dólar de los Estados Unidos. Esta tasa es el promedio ponderado de las operaciones de compra y venta de dólares efectuadas por los intermediarios financieros que transan en el mercado cambiario colombiano en un día determinado [12], [13]. La variación de la TRM afecta el comercio exterior de Colombia, lo cual también cambia variables como el nivel de producción y las inversiones de las empresas del país [14]. Es por eso que en este trabajo se considera el uso de este indicador para predecir los cambios en el COLCAP.

4.4 Análisis técnico

Un gráfico de velas japonesas es un tipo de gráfico financiero que registra los movimientos de precios en el tiempo por medio de cuatro valores diarios: precio de apertura, precio de cierre, precio mínimo y precio máximo [15]. En el gráfico de velas (figura 4.3), una línea vertical muestra el rango que tuvo el precio a lo largo del día, desde el mínimo hasta el máximo, y un rectángulo en el centro de la línea muestra la separación entre la apertura y el cierre [16]. Si el precio de cierre del día es mayor que el de apertura, significa que el precio aumentó a lo largo de esa jornada y el cuerpo de la vela (el rectángulo) es de color claro o verde. En cambio, si el precio cierra más abajo de lo que abrió, se entiende que el precio cayó y el cuerpo de la vela sería de un color oscuro o rojo.

Los cambios de los precios de varios días pueden agregarse para formar tendencias. La tendencia simplemente se refiere a la dirección en que se está moviendo el mercado, pero este no suele moverse en línea recta, es decir, no es común que haya muchas velas verdes o rojas seguidas. Este comportamiento genera una serie de zigzags, como los de la figura 4.4, compuestos por picos y valles, y es la dirección de estos lo que se define como la tendencia del mercado; por ejemplo, una tendencia alcista o ascendente se definiría como una serie de picos y valles sucesivamente más altos.

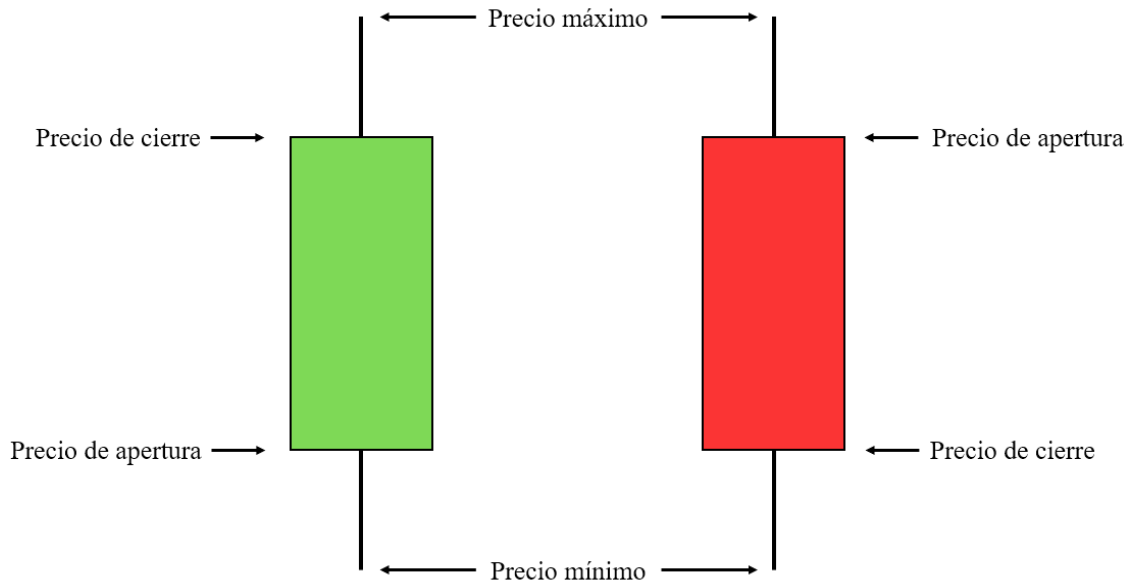


Figura 4.3: Velas japonesas

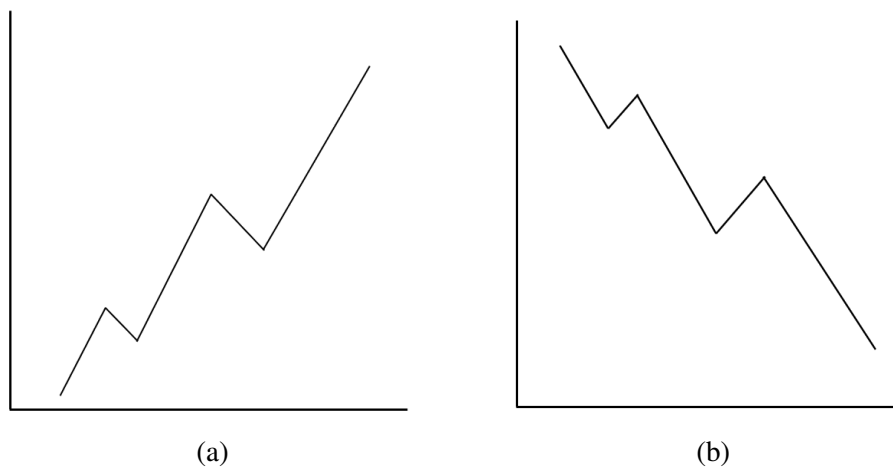


Figura 4.4: Tendencias de mercado

(a) Ejemplo de tendencia alcista, (b) Ejemplo de tendencia bajista. (Adaptado de [16])

Otro elemento importante que suele incluirse en el análisis técnico es el volumen. El volumen representa el total de operaciones ejecutadas en un mercado en un determinado día; esto es, la cantidad de acciones que cambiaron de dueño en un día concreto en el mercado de valores [16]. Aunque el nivel de precios mostrado en el gráfico de velas es el indicador más importante, el volumen sirve como una variable de confirmación, porque mide la intensidad de los movimientos en los precios, o sea que un volumen mayor refleja un grado más alto de intensidad o de certeza en una subida o una caída en el precio, lo que puede traducirse en que el cambio sea mayor o que la tendencia se mantenga durante más días.

4.5 Procesamiento de lenguaje natural

El procesamiento de lenguaje natural es un área de la inteligencia artificial que se enfoca en interpretar el lenguaje natural, es decir, lo que usan los humanos para comunicarse entre sí, y representarlo por medio de datos numéricos que un algoritmo o modelo pueda usar para aprender [5].

Para este trabajo también es necesario conocer BERT (Bidirectional Encoder Representations from Transformers), un modelo implementado con una red neuronal para procesamiento de lenguaje natural diseñado para preentrenar representaciones de texto no etiquetado [17]. Como resultado, esta arquitectura se puede utilizar con una sola capa de salida adicional con el fin de adaptarla para una amplia variedad de tareas como predicción de oraciones, generación de texto, respuesta a preguntas, entre otras [18].

Ahora, el interés del presente trabajo está en la clasificación de textos, ya que se tendrá en cuenta un conjunto de noticias económicas al momento de entrenar los modelos predictivos. No obstante, debido a la incapacidad de las redes neuronales de recibir los textos crudos, es necesario asignar un puntaje de positividad o negatividad a cada texto, proceso conocido como análisis de sentimiento [5].

4.6 Métricas de evaluación

A lo largo de este trabajo se utilizan distintas medidas para cuantificar el desempeño de los modelos implementados. A continuación se enuncian las definiciones de cuatro métricas usadas para evaluar modelos de regresión, en donde y es la etiqueta real, \hat{y} corresponde a la etiqueta predicha e \bar{y} es el promedio de y .

1) Error Cuadrático Medio:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$$

Básicamente, el MSE es la media de los errores elevados al cuadrado. Tomar el cuadrado de los errores permite evaluar diferencias positivas y negativas de la misma manera mientras se penalizan los errores más grandes. Un MSE pequeño indica que, en promedio, los errores para todos los ejemplos son pequeños.

2) Raíz del Error Cuadrático Medio:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} = \sqrt{\text{MSE}}$$

Aunque sea simplemente la raíz del MSE, el RMSE tiene la ventaja de ser interpretable en las unidades originales de los datos.

3) Error Absoluto Medio:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

Es menos sensible a valores atípicos que el RMSE y brinda una interpretación más directa del promedio de los errores.

4) Coeficiente de Determinación:

$$R^2 = 1 - \frac{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

El numerador de la fracción corresponde al MSE, que se puede interpretar como la varianza de los errores, y el denominador es la varianza de las etiquetas reales. Se dice que el R^2 representa la proporción de la varianza de y que ha sido explicada por las variables independientes de un modelo. El mejor valor para el R^2 es 1 y puede tomar valores negativos (para modelos con predicciones peores que simplemente predecir el promedio de los datos).

Por otra parte, también se debe conocer la medida de exactitud o *accuracy* para poder interpretar la evaluación de los modelos de clasificación, ya que esta mide la proporción en que se están etiquetando bien los ejemplos recibidos. La exactitud se define así:

$$\text{Exactitud} = \frac{\text{Predicciones correctas}}{\text{Número de predicciones}}$$

4.7 Revisión de literatura

La literatura sobre predicción en el mercado de valores comprende múltiples métodos. Algunos de estos sólo incluyen datos financieros como precios o indicadores técnicos, mientras que otros también consideran el análisis del sentimiento de mercado a través de noticias o variables macroeconómicas como tasas de cambio o medidores de riesgo. A continuación, se hará una revisión de cada enfoque, describiendo sus fortalezas y limitaciones.

Inicialmente, los estudios se basaban en el uso de algoritmos de aprendizaje automático para predecir el movimiento de los mercados accionarios. En particular, el objetivo de la investigación [19] fue usar Máquinas de Vectores de Soporte y aprovechar la correlación que existe entre los índices bursátiles de Estados Unidos y distintos activos de otras partes del mundo, como euros, yenes, oro, crudo, entre otros. Primero entrenaron un modelo de clasificación para predecir la dirección en que cambia el índice NASDAQ con un día de anticipación y obtuvieron un accuracy del 74.4%; luego mostraron que el modelo generalizó bien al probarlo con otros dos índices bursátiles. En dicho trabajo sólo se utilizó el precio de cierre diario de los índices, pero se halló una forma de medir el sentimiento de forma más objetiva a través del comportamiento de los mercados globales.

Estudios más recientes han revisado la capacidad de predicción de modelos construidos con redes neuronales. Por ejemplo, en [20] se compara una red neuronal clásica con otros modelos de aprendizaje automático para predecir la dirección en que cambiará el precio de cierre semanal de tres acciones chinas y concluyen que la primera sí supera a los otros algoritmos en términos del accuracy, pero este resultado podría no ser tan robusto debido a la poca cantidad de datos disponibles (porque usan datos semanales). Otro ejemplo es [21], donde usaron un modelo LSTM que recibe ventanas de precios de 22 días para predecir el valor del NIFTY 50, el principal índice de la Bolsa de Valores de India. En este caso, probaron diferentes combinaciones de variables tomadas del gráfico de velas (apertura, máximo, mínimo y cierre) y alcanzaron los mejores resultados al incluirlas todas, obteniendo un RMSE de 0.00859 en el conjunto de prueba.

Las conclusiones de [20], aunque no muy robustas, sugieren que el acercamiento de las redes neuronales puede ser mejor que los modelos predictivos clásicos, y los resultados de [21] indican que sí vale la pena incluir los datos de las gráficas de velas en los modelos, en lugar de limitar la información al usar únicamente un precio. Ahora bien, cabe resaltar que estos dos algoritmos siguen siendo relativamente simples, porque solamente usaron los precios de los activos para predecir el mismo, en vez de agregar otras variables que pudieran aportar información valiosa para los modelos.

De la misma forma, el estudio [22] sólo utilizó los precios diarios de las velas de una multinacional estadounidense; sin embargo, su acercamiento se basó en implementar un algoritmo genético para optimizar los hiperparámetros de una red LSTM, lo que resultó en un R^2 de 87 %.

Adicionalmente, se encontraron trabajos apoyados en el análisis de sentimiento de noticias. Para la investigación [23], se usaron modelos LSTM entrenados con los precios de cierre diarios de las cinco empresas más grandes del S&P500, el índice más representativo de la Bolsa de Valores de Nueva York. Su objetivo fue predecir los precios de cierre del siguiente día y luego ajustarlos según el puntaje de sentimiento obtenido para cada empresa, aunque el modelo que se usó para el procesamiento de lenguaje natural clasificó la gran mayoría de los textos como neutrales. Debido a esto, las noticias tuvieron un efecto reducido al momento de ajustar los valores ya predichos por la red neuronal. En [24] también se usó el análisis de sentimiento de noticias económicas para predecir los precios de cierre semanales del S&P500 en un período de 20 años. En este caso, el puntaje de sentimiento se incluyó como una variable más en el modelo, junto con múltiples indicadores técnicos y fundamentales (de las empresas que componen el índice), alcanzando un RMSE de 0.00138.

En Colombia, la literatura relacionada con la predicción de activos financieros sigue siendo bastante limitada. En [25] se desarrolló un estudio reciente, completo y afín al presente trabajo de grado, ya que se enfocó en evaluar varios modelos (incluyendo LSTM) para predecir el precio de cierre diario del COLCAP. De forma similar a [19], en ese trabajo se analizó la relación entre el COLCAP y varios activos e indicadores globales, pero no se tuvo en cuenta el efecto que pueden tener las noticias financieras sobre el índice. También se realizó un análisis de estacionariedad con las series de tiempo para poder transformar las variables y mejorar así la calidad de las predicciones.

Finalmente, la investigación [25] obtuvo un MSE de 4229.38 con el modelo SARIMAX y de 653.77 con la red LSTM, confirmando, al igual que otras de las fuentes descritas, que las redes neuronales recurrentes sí destacan con respecto a otras arquitecturas. Adicionalmente, se obtuvo un MAE de 15.64 en entrenamiento y de 18.90 para prueba, métricas que servirán como referencia al momento de evaluar el rendimiento de los modelos del presente trabajo de grado.

Capítulo 5

METODOLOGÍA

5.1 Recolección de datos

Los datos utilizados en el presente trabajo corresponden a los precios diarios y el volumen del COLCAP. Para asegurar la calidad y fiabilidad de estos, se usó una terminal de Bloomberg [10] para descargar la información contenida en el gráfico de velas (véase la figura 5.1) junto con el volumen del índice desde el 1 de enero del 2015 hasta el 31 de diciembre del 2024. Cabe resaltar que solamente se registran datos en los días en que la Bolsa de Valores está en funcionamiento (días hábiles o laborables), lo cual reduce considerablemente la cantidad de datos disponibles.

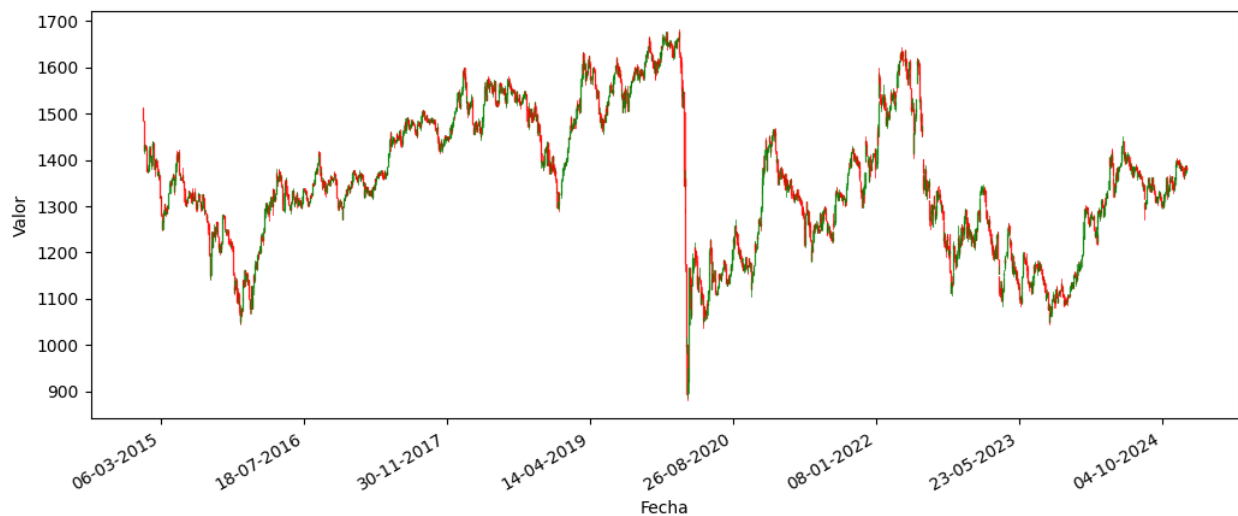


Figura 5.1: Gráfico de velas del MSCI COLCAP

Adicionalmente, se incluyó la TRM en el análisis para poder estudiar su relación con el rendimiento de las principales empresas de Colombia (las que están contenidas en la canasta del COLCAP). La serie de tiempo con el valor diario de este indicador (figura 5.2) se encuentra disponible en [13] y de aquí se extrajo el mismo período que el de los datos de Bloomberg.

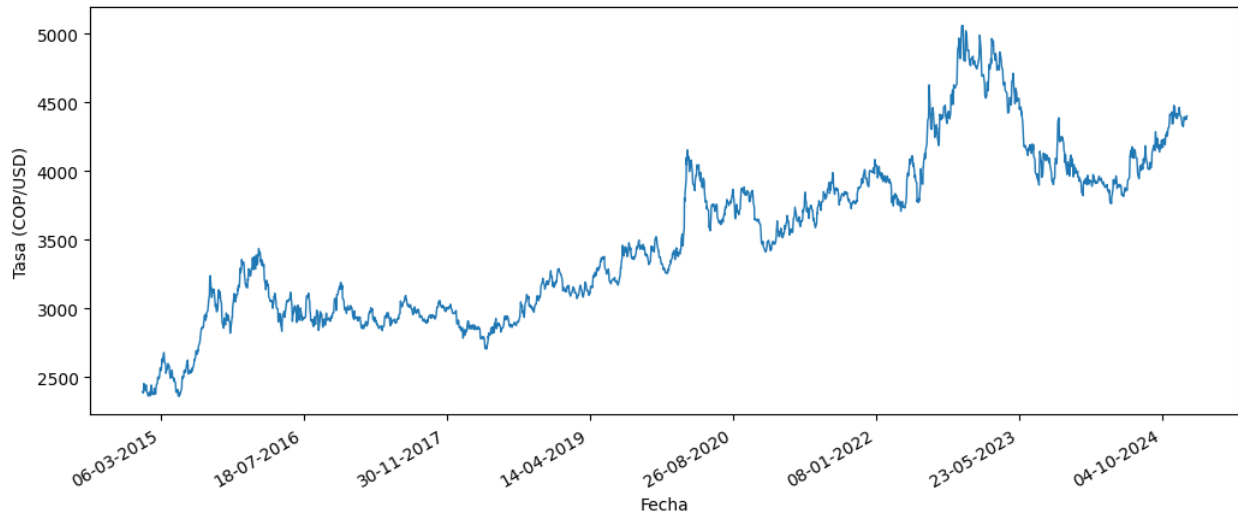


Figura 5.2: Valor histórico de la TRM

Por otro lado, se automatizó la extracción de texto en los sitios web de dos periódicos especializados en noticias económicas: Portafolio [26] y La República [27]. Esto se hizo con el propósito de conseguir el encabezado, el contenido y la fecha de cada una de las noticias disponibles. Para esto, se acotó la búsqueda para que el algoritmo sólo extrajera resultados en el período de interés, es decir, del 2015 al 2024. Además, se filtraron los resultados en ambos periódicos por medio de una lista de palabras clave relacionadas con el COLCAP o con la economía colombiana en general, produciendo así una base de datos con los textos de más de 63000 noticias económicas (un promedio de 26 noticias por día).

5.2 Limpieza de datos

Primero, se estandarizó el formato de las fechas de ambos periódicos para poder reunir las noticias en un solo conjunto de datos. Luego, se modificaron varias fechas, ya que las noticias de Portafolio que salieron el 30 y el 31 de diciembre de 2024 están mal etiquetadas porque los editores del sitio web les pusieron 2025 como año de publicación (fecha que aún no ocurre al momento de la realización de este trabajo de grado).

Al mismo tiempo, se cambiaron las fechas de todas las noticias que salieron los fines de semana. Esto se hizo con el propósito de asegurar que todos los registros de la base de datos se ubicaran entre lunes y viernes, porque el COLCAP solo actualiza su valor entre estos días. Por ejemplo, si una noticia salió un sábado o domingo, se le puso la fecha del viernes inmediatamente anterior. De esta manera, se utilizan todas las noticias del fin de semana para predecir el comportamiento del COLCAP del lunes.

Al final se filtraron los datos de la TRM para que coincidieran con las fechas en las que se tienen registros del COLCAP y se remplazaron los días festivos para las noticias (el tratamiento de estos fue análogo al de los fines de semana). También había cuatro días laborables en los que no se publicó ninguna noticia y se rellenaron con valores nulos para que todos los conjuntos de datos (precios, TRM y noticias) tuvieran las mismas fechas y así poder unirlos más fácilmente al final del preprocesamiento.

Un punto importante acerca de los datos históricos del COLCAP es que aparecen registros para el viernes, 13 de marzo del 2020; sin embargo, ese día no se abrió la sesión de negociaciones de la Bolsa de Valores de Colombia debido a problemas técnicos que se presentaron al inicio de la jornada [28], [29]. En Bloomberg [10] se muestran los mismos valores del día anterior, pero en este trabajo se eliminó el 13 de marzo antes de preparar las fechas de las noticias según los procedimientos descritos anteriormente.

5.3 Análisis de sentimiento

Para el análisis de sentimiento de las noticias recopiladas, se utilizó el modelo Finance Sentiment ES, el cual es una versión de BETO (BERT para textos en español) [30] que fue reentrenada por [31] para clasificar noticias financieras. Con este modelo, se etiquetó cada noticia como positiva, negativa o neutral, asignando puntajes de 1, -1 o 0, respectivamente. Este procedimiento resultó en que se clasificaron 29564 noticias como positivas, 14978 como negativas y 18810 neutrales.

En la figura 5.3a se muestran las palabras más comunes en noticias positivas y se nota que suelen aparecer expresiones como “aumento”, “crecimiento” o “inversión”, mientras que en la figura 5.3b se ve que en las noticias negativas predominan términos como “gobierno” y “caída”. Luego de esto, se agruparon los puntajes obtenidos por el modelo según sus fechas y se calculó el “sentimiento del día” como la moda de los valores de sentimiento de todas las noticias correspondientes a cada fecha. Así, se obtuvieron 1759 días positivos, 269 negativos y 408 neutrales.

Que la ACF decaiga muy lentamente, como en las figuras 5.4a y 5.4c, indica que el promedio de una serie de tiempo no es constante [2], lo cual viola la primera condición para que sea estacionaria. La solución para esto es la diferenciación: se reemplaza una serie $\{X_t\}$ por $\{X_t - X_{t-1}\}$ y esto se repite hasta que su gráfica ACF deje de tener el comportamiento anterior, es decir, que ahora caiga de forma repentina luego de pocos valores y que no vuelva a mostrar valores significativamente altos. En este caso, todas las variables se volvieron estacionarias después de aplicar una sola diferenciación. Las nuevas gráficas son las siguientes:

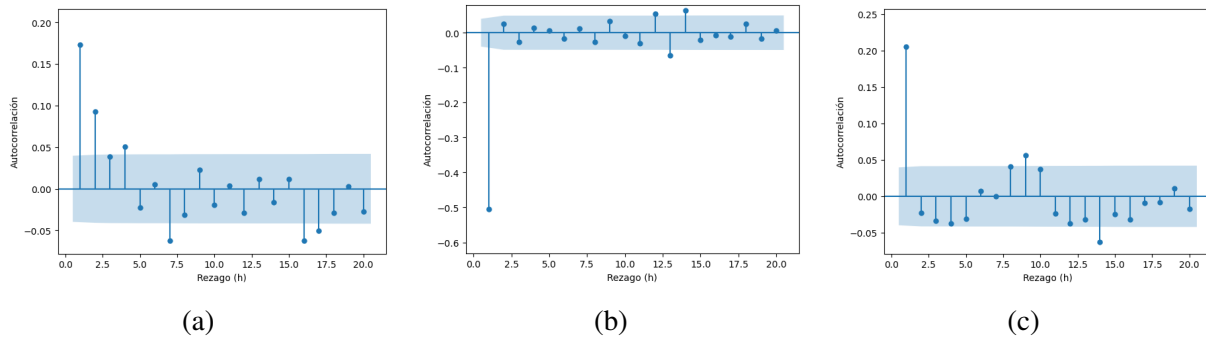


Figura 5.5: Gráficos ACF de las variables diferenciadas
(a) Precio del COLCAP, (b) Volumen del COLCAP, (c) TRM

5.5 Modelamiento

Como la finalidad de este trabajo es predecir el comportamiento del COLCAP, se definió el vector objetivo a partir de la diferencia o cambio en los precios de cierre de este índice. Además, se crearon tres versiones de este vector objetivo, que corresponden al cambio en el precio con uno, dos y tres días de anticipación para analizar la capacidad de las redes LSTM para predecir cambios más distantes en el tiempo.

Para comenzar, se separaron las series de tiempo en tres períodos: entrenamiento, validación y prueba, manteniendo una proporción cercana a 70% – 15% – 15%. Con esto en mente, se tomaron los primeros siete años de la base de datos para el conjunto de entrenamiento; esto es, desde 2015 hasta 2021. Los datos restantes se dividieron entre validación y prueba, resultando en períodos de un año y medio para cada uno. Al mismo tiempo se creó un modelo compuesto por dos capas LSTM de 64 unidades seguidas (cada una) por Dropouts de 30% y una capa densa de 32 unidades con un Dropout de 20% para reducir el sobreajuste de la red. Aparte de eso se utilizó el MSE como función de pérdida para asignarle más peso a los errores grandes.

Además de los tres vectores objetivo, también se probaron diferentes versiones de los datos de entrada. Un primer modelo recibió únicamente los datos del COLCAP para servir como base o punto de partida. Luego se entrenó otro modelo que también recibió los valores de la TRM. Después se incluyeron las clases de sentimiento de las noticias financieras sin la TRM y, por último, se introdujeron ambas variables para entrenar un modelo con el conjunto de datos completo. Otro hiperparámetro sobre el que se realizó una búsqueda fue el tamaño de la ventana temporal que recibirían los modelos, pasándoles entre 10 y 60 días (en múltiplos de 10) con el objetivo de estudiar la cantidad de información del pasado que se necesita para predecir cambios futuros. Con todo esto, se entrenaron, inicialmente, 24 versiones para cada vector objetivo, resultando en un total de 72 modelos.

Al final, se repitió el proceso de entrenamiento de todos los modelos con versiones más simples de los datos para reducir el ruido introducido en la red. Para esto, se tomaron las variables diferenciadas y se reemplazaron los valores por sus signos, es decir, se asignaron valores de 1 y -1 a las diferencias positivas y negativas, respectivamente. Con estos modelos no se buscó resolver un problema de regresión, sino clasificar los cambios del COLCAP según si su tendencia es ascendente o descendente.

Capítulo 6

RESULTADOS Y DISCUSIÓN

Una vez culminadas las etapas de entrenamiento y validación, se obtuvieron los siguientes resultados con los modelos de regresión. Como se puede ver de la tabla 6.1 a la 6.3, siempre se obtuvieron valores negativos de R^2 en el conjunto de prueba, por lo que se entiende que estos modelos no logran explicar la varianza del COLCAP durante dicho período. Lo que esta métrica sí permite determinar es que los modelos que recibieron una ventana de 10 días para predecir los cambios futuros no obtuvieron un buen rendimiento en general, ni siquiera en el período de entrenamiento. Esto sugiere que es necesario usar escalas temporales más amplias, ya que los modelos sí logran aprender las suficientes características de las series de tiempo cuando se utilizan ventanas de 20 días o más.

Cuadro 6.1: R^2 con predicciones de $t + 1$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.3497	0.9804	0.9595	0.9790	0.0007	0.8792
	Con TRM	0.3491	0.9757	0.9731	0.9315	0.9779	0.8441
	Con noticias	0.3485	0.9752	0.9829	0.9457	0.9783	0.9645
	Con TRM y noticias	0.3468	0.9837	0.9424	0.9822	0.9847	0.9845
Prueba	Modelo base	-0.4067	-0.4195	-0.6396	-0.5630	-0.0079	-0.3594
	Con TRM	-0.4074	-0.5732	-0.6214	-1.4965	-0.4110	-0.4946
	Con noticias	-0.4324	-0.4938	-1.0228	-1.0750	-0.3244	-0.3538
	Con TRM y noticias	-0.3940	-0.5787	-0.7106	-0.7077	-0.6803	-0.5301

Cuadro 6.2: R^2 con predicciones de $t + 2$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.3531	0.9824	0.9870	0.9737	0.9842	0.9364
	Con TRM	0.3539	0.9813	0.9680	0.9864	0.9874	0.9449
	Noticias	0.3518	0.9862	0.9618	0.9824	0.9829	0.9809
	Con TRM y noticias	0.3539	0.9828	0.9695	0.9869	0.9756	0.9754
Prueba	Modelo base	-0.6076	-1.3296	-0.8147	-2.0168	-0.4211	-0.9415
	Con TRM	-0.5999	-1.2505	-0.8607	-0.8352	-0.4066	-0.8776
	Con noticias	-0.5099	-1.0188	-1.1397	-0.8199	-0.7102	-0.7984
	Con TRM y noticias	-0.5241	-0.8953	-0.8862	-0.9774	-0.5371	-0.5280

Cuadro 6.3: R^2 con predicciones de $t + 3$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.3790	0.9861	0.9775	0.9821	0.9829	0.9788
	Con TRM	0.3797	0.9877	0.9884	0.9856	0.9774	0.9758
	Con noticias	0.3791	0.9850	0.9855	0.9803	0.9881	0.9867
	Con TRM y noticias	0.3814	0.9809	0.9865	0.9856	0.9862	0.9858
Prueba	Modelo base	-0.7130	-0.7462	-0.8458	-0.4567	-0.3411	-0.9225
	Con TRM	-0.8380	-0.5780	-0.4360	-0.4347	-0.6708	-0.4505
	Con noticias	-0.8895	-3.1671	-0.5153	-0.8844	-1.1681	-0.5542
	Con TRM y noticias	-0.6887	-0.8994	-0.4080	-1.9815	-1.6551	-0.8150

También es posible notar, en términos del MAE, el sobreajuste que se está presentando, ya que los errores suelen ser de alrededor de 2 unidades para los datos de entrenamiento, mientras que en prueba se alcanzan valores mucho más elevados, ubicándose, más o menos, en 10 unidades para la tabla 6.4, 15 en la tabla 6.5 y 18 para la 6.6. Para tener una mejor perspectiva sobre la escala de estos resultados, se calculó el promedio de los cambios registrados en los vectores objetivo para cada período, obteniendo, para los datos de prueba, 8.38, 12.0 y 15.02 unidades para uno, dos y tres

días de anticipación, respectivamente. Con esta información se hace más claro que el MAE siempre es mayor que el cambio diario usual durante el período de prueba, por lo que ningún modelo es capaz de predecir este cambio de forma robusta, reafirmando así que los modelos no pudieron generalizar bien los resultados obtenidos en el período de entrenamiento. No obstante, siempre se alcanzaron mejores métricas que las que se obtuvieron en la investigación [25], expuestas al final de la revisión de literatura.

Cuadro 6.4: MAE con predicciones de $t + 1$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	7.5656	1.4435	2.2608	1.6059	9.5458	3.8611
	Con TRM	7.5875	1.5241	1.8321	2.9042	1.5845	3.6490
	Con noticias	7.5798	1.4517	1.4583	2.6366	1.7030	1.9725
	Con TRM y noticias	7.6197	1.4077	2.6257	1.4279	1.3948	1.3948
Prueba	Modelo base	9.9935	10.0702	10.7728	10.7221	8.2815	9.5939
	Con TRM	9.9805	10.5895	10.6921	11.4617	9.5186	10.1811
	Con noticias	10.0756	10.4444	11.7762	11.1523	9.3705	9.5725
	Con TRM y noticias	9.9485	10.5619	11.1277	11.0212	10.8018	10.1417

Cuadro 6.5: MAE con predicciones de $t + 2$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	11.6660	1.9325	1.8330	2.6231	1.9225	4.4730
	Con TRM	11.5881	2.1298	3.0752	2.0095	1.8786	3.9149
	Con noticias	11.6519	1.9367	3.2441	2.2263	2.2308	2.3987
	Con TRM y noticias	11.7027	2.0141	2.8832	1.9471	2.5866	2.6977
Prueba	Modelo base	15.2479	17.2608	16.7416	18.7925	14.7494	16.7541
	Con TRM	15.3891	17.2856	16.9694	16.7068	14.0857	16.4660
	Con noticias	14.8156	17.2175	16.2648	15.4868	16.0459	16.4783
	Con TRM y noticias	14.9979	15.1925	16.2515	17.0291	15.2278	14.3612

Cuadro 6.6: MAE con predicciones de $t + 3$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	14.4266	2.3651	3.4067	2.9875	2.6339	3.4238
	Con TRM	14.4146	2.2757	2.3557	2.6417	2.9445	3.5851
	Con noticias	14.4273	2.5036	2.4821	3.1759	2.3880	2.5806
	Con TRM y noticias	14.2807	2.9656	2.5987	2.6759	2.5909	2.7507
Prueba	Modelo base	19.6070	20.2030	21.1404	17.6646	17.3521	20.3442
	Con TRM	19.9013	19.5998	18.7863	18.2641	19.6378	17.8668
	Con noticias	20.1948	23.2292	19.0361	21.0036	20.8124	19.1733
	Con TRM y noticias	19.7385	21.6894	18.8099	23.7040	21.7899	21.3049

En las tablas anteriores también se encuentran más pruebas de que las ventanas de 10 días no se desempeñaron bien, incluso en el conjunto de entrenamiento, puesto que el error absoluto fue mayor con esta ventana para todos los modelos entrenados. Además, con las ventanas temporales más grandes se logró predecir mejor los datos de entrenamiento sin aumentar considerablemente el error de prueba.

De acuerdo con la información ya expuesta, los mejores modelos para predecir los cambios con un día de anticipación fueron aquellos que recibieron todas las variables (tanto precios y volumen como TRM y sentimiento). Esto es así con el conjunto de entrenamiento; sin embargo, el error de prueba no aumentó mucho al agregar las variables mencionadas. En el caso de las predicciones con tres días de anticipación, también se evidencian mejores resultados al incluir la TRM, aunque las noticias ya no aportan tanta información como en los casos de uno y dos días.

Ahora bien, como se mencionó al inicio del capítulo, los resultados anteriores corresponden a los modelos de regresión, lo que significa que los modelos no generalizan muy bien cuando se quiere predecir exactamente la medida en que va a cambiar el índice. Es por esto que también se revisó si las predicciones realizadas eran positivas o negativas con el fin de transformar el problema en uno de clasificación y determinar si los modelos podían obtener buenos resultados cuando lo único que se busca es ver si el COLCAP va a aumentar o disminuir.

Como se puede apreciar en la tabla 6.7, hay evidencia clara de que, al predecir un día hacia el futuro, los modelos que incluyeron la TRM y el sentimiento de las noticias obtuvieron mejores

métricas que aquellos que solo consideran los precios y el volumen de negociación del COLCAP. Además, en la tabla 6.9 se observa que, al intentar predecir los cambios con un rezago de tres días, se obtienen mejores resultados de clasificación cuando se incluyen las noticias económicas, lo cual está bien alineado con la justificación de todo este trabajo.

Cuadro 6.7: Exactitud con predicciones de $t + 1$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.7178	0.9495	0.9194	0.9327	0.5260	0.8400
	Con TRM	0.7208	0.9495	0.9229	0.8774	0.9414	0.8607
	Con noticias	0.7161	0.9471	0.9421	0.8984	0.9341	0.9246
	Con TRM y noticias	0.7149	0.9412	0.9032	0.9525	0.9510	0.9398
Prueba	Modelo base	0.5099	0.5536	0.5403	0.5231	0.5175	0.5902
	Con TRM	0.5324	0.5101	0.5015	0.5015	0.5302	0.4918
	Con noticias	0.5155	0.5159	0.5075	0.5138	0.5365	0.4885
	Con TRM y noticias	0.5099	0.5188	0.4955	0.5108	0.4730	0.4918

Cuadro 6.8: Exactitud con predicciones de $t + 2$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.7237	0.9632	0.9624	0.9315	0.9674	0.8905
	Con TRM	0.7279	0.9572	0.9331	0.9567	0.9619	0.9051
	Con noticias	0.7214	0.9578	0.9176	0.9531	0.9480	0.9519
	Con TRM y noticias	0.7184	0.9466	0.9379	0.9555	0.9377	0.9453
Prueba	Modelo base	0.5282	0.5291	0.4880	0.4969	0.5096	0.4901
	Con TRM	0.5028	0.5610	0.5359	0.4568	0.5382	0.5296
	Con noticias	0.5169	0.5116	0.5299	0.5216	0.5159	0.4441
	Con TRM y noticias	0.5311	0.5203	0.5269	0.5031	0.5287	0.5362

Cuadro 6.9: Exactitud con predicciones de $t + 3$

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.7355	0.9567	0.9313	0.9477	0.9522	0.9355
	Con TRM	0.7338	0.9567	0.9564	0.9543	0.9474	0.9270
	Con noticias	0.7338	0.9638	0.9540	0.9459	0.9607	0.9611
	Con TRM y noticias	0.7338	0.9471	0.9612	0.9591	0.9565	0.9556
Prueba	Modelo base	0.5184	0.5452	0.5105	0.5882	0.5815	0.4983
	Con TRM	0.5354	0.5248	0.4745	0.5387	0.5399	0.5347
	Con noticias	0.5071	0.4985	0.5285	0.5232	0.5208	0.4818
	Con TRM y noticias	0.5156	0.4956	0.5345	0.5418	0.5495	0.4620

A pesar del cambio aplicado al objetivo de los modelos, sigue notándose el mismo comportamiento sobre la generalización, ya que para el entrenamiento se suelen obtener valores de exactitud de 95 %, mientras que, en los mejores casos, esta métrica se ubica entre 52 % y 59 % con los datos de prueba. Aun así, como los períodos de validación y prueba comprenden únicamente tres años, no está claro si los modelos que obtuvieron más de un 50 % de predicciones correctas serían capaces de replicar estos resultados de forma consistente.

El sobreajuste de los primeros modelos no se pudo corregir con ningún método de regularización, ni con una búsqueda de hiperparámetros, ni ajustando la cantidad de parámetros entrenables de las redes. Esto podría significar que los datos son demasiado complejos como para que las redes puedan generalizar bien. Por esta razón, se intentó simplificar las entradas de la red de acuerdo al signo de los cambios en cada variable (luego de diferenciarlas para volverlas estacionarias), es decir, para que los modelos sólo recibieran 1 o -1.

Con las entradas simplificadas, solamente se intentó resolver un problema de clasificación, ya que se buscó predecir la dirección que tomaría el COLCAP en el futuro. Como se puede ver en las tablas 6.10, 6.11 y 6.12, sí se alcanzó una buena generalización con estos nuevos datos, ya que las métricas son bastante más consistentes a lo largo de los períodos seleccionados, tanto en los primeros siete años de entrenamiento como con los últimos años de datos desconocidos.

Cuadro 6.10: Exactitud con predicciones de $t + 1$ (variables simples)

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.5496	0.5386	0.5442	0.5469	0.5520	0.5535
	Con TRM	0.5478	0.5362	0.5562	0.5457	0.5490	0.5468
	Con noticias	0.5348	0.5356	0.5329	0.5547	0.5478	0.5481
	Con TRM y noticias	0.5348	0.5333	0.5352	0.5385	0.5435	0.5462
Prueba	Modelo base	0.4930	0.5507	0.5552	0.5508	0.5492	0.5279
	Con TRM	0.4704	0.5391	0.5045	0.5231	0.5016	0.5148
	Con noticias	0.4789	0.5420	0.5612	0.5538	0.5587	0.5410
	Con TRM y noticias	0.4901	0.5507	0.5194	0.5292	0.5270	0.5180

Cuadro 6.11: Exactitud con predicciones de $t + 2$ (variables simples)

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.5567	0.5689	0.5699	0.5595	0.5623	0.5578
	Con TRM	0.5608	0.5463	0.5711	0.5511	0.5550	0.5541
	Con noticias	0.5573	0.5588	0.5657	0.5613	0.5556	0.5572
	Con TRM y noticias	0.5525	0.5433	0.5639	0.5511	0.5586	0.5547
Prueba	Modelo base	0.4802	0.5233	0.5389	0.5247	0.5191	0.5230
	Con TRM	0.4859	0.5174	0.5449	0.5247	0.5255	0.5263
	Con noticias	0.4633	0.5233	0.5210	0.5340	0.5318	0.5230
	Con TRM y noticias	0.4802	0.5145	0.5269	0.5432	0.5318	0.5230

Cuadro 6.12: Exactitud con predicciones de $t + 3$ (variables simples)

		Ventana de tiempo (días)					
		10	20	30	40	50	60
Entrenamiento	Modelo base	0.5602	0.5713	0.5753	0.5649	0.5617	0.5596
	Con TRM	0.5779	0.5493	0.5717	0.5727	0.5665	0.5797
	Con noticias	0.5478	0.5523	0.5585	0.5649	0.5683	0.5602
	Con TRM y noticias	0.5773	0.5523	0.5711	0.5745	0.5689	0.5760
Prueba	Modelo base	0.5552	0.4781	0.5526	0.5139	0.5144	0.5149
	Con TRM	0.5297	0.5015	0.5075	0.5294	0.5208	0.4851
	Con noticias	0.5241	0.5190	0.5165	0.5201	0.5176	0.5182
	Con TRM y noticias	0.5382	0.5219	0.4715	0.4892	0.5144	0.4719

En la tabla 6.10, se observa que los mejores resultados de prueba se obtuvieron al incluir los valores de sentimiento de las noticias y que los peores resultados siguen siendo los de las ventanas de 10 días. Finalmente, se aprecia que, en general, los modelos fueron mejores cuando se usaron ventanas temporales intermedias, ni tan pequeñas como las de 10 días que no aportan suficiente información ni tan amplias como las de 60 días que introducen demasiado ruido durante el entrenamiento.

Capítulo 7

CONCLUSIONES Y RECOMENDACIONES

Los modelos que recibieron los datos más complejos no generalizaron bien. Esto sugiere que formular la predicción del comportamiento de un índice bursátil como un problema de regresión representa una tarea bastante difícil. En cambio, al reducir la complejidad de los datos para clasificar únicamente en función de la dirección que toma el mercado, sí se logró obtener resultados más consistentes en todos los períodos estudiados. Adicionalmente, hay ventanas de tiempo en las que resulta útil incluir el sentimiento de mercado, ya sea a través del sentimiento de noticias o por medio de variables macroeconómicas como la tasa de cambio del dólar, porque parecen aportar información útil en las predicciones.

Con respecto a la TRM, en [14] se aclara que su efecto en la economía nacional no es tan inmediato, sino que suele tomar entre seis meses y un año, por lo que se podría estudiar un posible rezago para esta variable. Otra variable que podría incluirse en futuras investigaciones para explicar el sentimiento de mercado es el EMBI (Emerging Markets Bond Index), el cual es un índice que sirve para medir el riesgo de un país de una forma más directa que las tasas de cambio o la balanza comercial. Si el EMBI tiene un valor más alto, esto hace que el país se perciba como “más riesgoso”, lo cual afecta la capacidad de financiación de las empresas y reduce la eficiencia de estas. En el trabajo [25], se confirmó la fuerte correlación negativa que existe entre este indicador y el rendimiento del COLCAP.

Por otro lado, el sobreajuste de los modelos puede deberse a la poca cantidad de datos disponible, ya que el conjunto de entrenamiento contenía menos de 1700 observaciones. Esta cantidad de datos no suele ser suficiente para entrenar una red neuronal, debido a que estas pueden memorizar los datos con los que se entrenan cuando se enfrentan a una tarea demasiado compleja. Para abordar esta limitación, se podría entrenar un modelo inicial con los registros de las acciones que componen la canasta del COLCAP, ya que así se tendrían muchas más observaciones disponibles. En este caso, se transformarían las diferentes series de tiempo a una escala común y se buscaría un

modelo que logre generalizar a través de varios activos financieros. Teniendo esto, se recomienda reentrenar o ajustar el modelo para que prediga mejor el comportamiento del COLCAP. En línea con esta idea, otro posible acercamiento sería utilizar un modelo que haya sido preentrenado con datos de series de tiempo no financieras y ajustarlo para la tarea más específica que es predecir un índice bursátil.

Otra recomendación se relaciona con el hecho de que, en este proyecto, se realizó una búsqueda inicial de hiperparámetros usando una combinación específica de ventana de tiempo y de variables consideradas en el modelo, y los valores que arrojaron los mejores resultados se usaron para las otras 71 combinaciones (contando las tres variaciones del vector objetivo). Un aspecto que se podría mejorar consiste en realizar una búsqueda de hiperparámetros para cada combinación específica, ya que en cada caso se quiso predecir un objetivo diferente con variables diferentes, por lo que los hiperparámetros óptimos también deberían ser únicos para cada caso.

En cuanto a la calidad de los datos usados, se pudo notar que el promedio de los cambios del COLCAP es más alto desde el 2022 en comparación con las variaciones que se presentaban en años anteriores. Esto significa que la volatilidad pudo haber cambiado, indicando que los conjuntos de validación y prueba no se comportan de una forma tan similar al de entrenamiento. Es por esto que sería útil incluir alguna medida de volatilidad como la desviación estándar o el ATR (Average True Range), el cual es un indicador de análisis técnico que mide la magnitud de las fluctuaciones en los precios de un activo. Al hacer esto, se podrían explicar mejor los cambios en períodos desconocidos para el modelo, mejorando así el R^2 .

Por último, una de las motivaciones de esta investigación consistió en analizar si el sentimiento del mercado es capaz de cubrir la presencia de los datos atípicos de la pandemia. Sobre este punto, es evidente que se presentaron casos en que los modelos que no incluían las noticias, es decir, los modelos base, fueron los mejores predictores. De este modo, se recomienda eliminar estas fechas particulares para trabajos posteriores.

BIBLIOGRAFÍA

- [1] K. Rojas-Jimenez, *Ciencia de Datos para Ciencias Naturales*. Book-down, 2022, ch. Análisis de Series de Tiempo. [Online]. Available: https://bookdown.org/keilor_rojas/CienciaDatos/análisis-de-series-de-tiempo.html
- [2] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods*, 2nd ed. Pearson Education, 2006.
- [3] P. J. Brockwell and R. A. Davis, *Introduction to Time Series and Forecasting*, 2nd ed., ser. Springer Texts in Statistics, G. Casella, S. Fienberg, and I. Olkin, Eds. Springer, 2010.
- [4] M. F. Dixon, I. Halperin, and P. Bilokon, *Machine Learning in Finance: From Theory to Practice*. Springer Nature, 2020, ch. Advanced Neural Networks.
- [5] H. Lane, C. Howard, and H. M. Hapke, *Natural Language Processing in Action: Understanding, analyzing, and generating text with Python*. Shelter Island, NY: Manning Publications, 2019.
- [6] Gamco, “LSTM: Long short-term memory.” [Online]. Available: <https://gamco.es/glosario/lstm-long-short-term-memory/>
- [7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, ch. Sequence Modeling: Recurrent and Recursive Nets.
- [8] Bolsa de Valores de Colombia, “Los elementos más relevantes de los índices.” [Online]. Available: <https://www.bvc.com.co/indices-descripcion-general>
- [9] Banco Santander, “¿Qué es un índice bursátil?” [Online]. Available: <https://www.bancosantander.es/glosario/indice-bursatil>
- [10] Bloomberg, “Bloomberg Terminal.” [Online]. Available: <https://www.bloomberg.com/professional/terminal-overview/>
- [11] Bolsa de Valores de Colombia, “MSCI COLCAP.” [Online]. Available: <https://www.bvc.com.co/msci-colcap>

- [12] Banco de la República | Colombia, “Tasa de cambio o tasa de cambio representativa del mercado (TRM).” [Online]. Available: <https://www.banrep.gov.co/es/glosario/tasa-cambio-trm>
- [13] —, “Estadísticas económicas: Tasa de cambio Representativa del Mercado (TRM).” [Online]. Available: https://suameca.banrep.gov.co/estadisticas-economicas/informacionSerie/1/tasa_cambio_peso_colombiano_trm_dolar_usd
- [14] O. Blanchard and D. Pérez-Enrri, *Macroeconomía: Aplicaciones para Latinoamérica*, 2nd ed. Buenos Aires: Pearson Education, 2011, ch. El mercado de bienes en una economía abierta.
- [15] The Data Visualisation Catalogue, “Gráfico de Velas.” [Online]. Available: https://datavizcatalogue.com/ES/metodos/grafico_de_velas.html
- [16] J. J. Murphy, *Análisis Técnico de los Mercados Financieros*. Barcelona: Ediciones Gestión 2000, 2000.
- [17] Google Research, “TensorFlow code and pre-trained models for BERT.” [Online]. Available: <https://github.com/google-research/bert>
- [18] J. Devlin et al., “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, 2019, pp. 4171–4186. [Online]. Available: <https://api.semanticscholar.org/CorpusID:52967399>
- [19] S. Shen, H. Jiang, and T. Zhang, “Stock Market Forecasting Using Machine Learning Algorithms,” *Department of Electrical Engineering, Stanford University, Stanford, CA*, 2012.
- [20] Y. Song, Y. Zhou, and R. Han, “Neural networks for stock price prediction,” *arXiv preprint*, May 2018, doi: 10.48550/arXiv.1805.11317.
- [21] M. Roondiwala, H. Patel, and S. Varma, “Predicting Stock Prices Using LSTM,” *International Journal of Science and Research (IJSR)*, vol. 6, no. 4, pp. 1754–1756, Apr 2017, doi: 10.21275/ART20172755.
- [22] X. Sha, “Time Series Stock Price Forecasting Based on Genetic Algorithm (GA)-Long Short-Term Memory Network (LSTM) Optimization,” *arXiv preprint*, May 2024, doi: 10.48550/arXiv.2405.03151.
- [23] A. Deep, “A Multifactor Analysis Model for Stock Market Prediction,” *International Journal of Computer Science and Telecommunications*, vol. 14, no. 1, Feb 2023.

- [24] S. Zhong and D. Hitchcock, “S&P 500 Stock Price Prediction Using Technical, Fundamental and Text Data,” *Statistics, Optimization & Information Computing*, vol. 9, no. 4, p. 769–788, Nov 2021, doi: 10.19139/soic-2310-5070-1362.
- [25] D. S. Osorio-Aristizábal, “Comparación de modelos de series temporales ARIMA, SARIMAX y LSTM para la predicción del índice COLCAP,” Master’s thesis, Universidad EAFIT, Medellín, 2024. [Online]. Available: <https://hdl.handle.net/10784/34670>
- [26] Portafolio, “Noticias económicas de Colombia y el Mundo.” [Online]. Available: <https://www.portafolio.co/>
- [27] La República, “Noticias de Economía, Finanzas, Empresas y Negocios de Colombia y el mundo.” [Online]. Available: <https://www.larepublica.co/>
- [28] —, “BVC no abrió sesión de negociación de renta variable por dificultades técnicas.” [Online]. Available: <https://www.larepublica.co/finanzas/operacion-en-la-bvc-no-iniciara-en-su-horario-habitual-por-dificultades-tecnicas-2977227>
- [29] —, “Expectativa por la Bolsa de Valores de Colombia tras una jornada sin negociación.” [Online]. Available: <https://www.larepublica.co/finanzas/expectativa-por-la-bolsa-de-valores-de-colombia-tras-una-jornada-sin-negociacion-2978044>
- [30] J. Cañete et al., “Spanish Pre-Trained BERT Model and Evaluation Data,” in *PML4DC at ICLR 2020*, 2020. [Online]. Available: <https://huggingface.co/dccuchile/bert-base-spanish-wwm-cased>
- [31] bards.ai, “Finance Sentiment ES (base).” [Online]. Available: <https://huggingface.co/bardsai/finance-sentiment-es-base>

ANEXO

Para facilitar la replicación de los procedimientos empleados en la metodología, se construyó un repositorio en GitHub que se actualizó a lo largo de la realización de este trabajo de grado. En este repositorio también se encuentran los datos procesados y los códigos que se utilizaron para visualizar los resultados completos. El proyecto se encuentra en el siguiente enlace:

<https://github.com/Samuel0104/COLCAP-Forecasting>