

# Universidad del Valle de Guatemala

Deep Learning y Sistemas Inteligentes

Sección 10



## PROYECTO FINAL

Diego Lemus - 21469

Fabián Juárez - 21440

Guatemala, 11 de Noviembre del 2024

## ÍNDICE

DESCRIPCIÓN DEL PROBLEMA.....	3
ANÁLISIS.....	4
PROPUESTA DE SOLUCIÓN.....	5
DESCRIPCIÓN DE LA SOLUCIÓN.....	7
Análisis de sentimientos.....	7
Stock Prediction.....	9
Stock-Sentiment Prediction.....	12
HERRAMIENTAS APLICADAS.....	14
Predicción de acciones.....	14
Procesamiento de texto.....	15
Análisis de sentimientos.....	15
Otros.....	15
RESULTADOS.....	16
Stock Prediction.....	16
Stock-Sentiment Prediction.....	19
CONCLUSIONES.....	22
RECOMENDACIONES.....	22
REPOSITORIO.....	23
PRESENTACIÓN.....	23
VIDEO DE FUNCIONAMIENTO.....	23
BIBLIOGRAFÍA.....	23

## DESCRIPCIÓN DEL PROBLEMA

La predicción de acciones en el mercado de valores plantea el desafío de anticipar el comportamiento de precios mediante la interpretación de múltiples factores que influyen en el mercado. Tradicionalmente, los modelos predictivos se han basado en análisis de tendencias históricas de precios, patrones de comportamiento y otros datos financieros. Sin embargo, en la actualidad, la opinión pública y la presencia de las empresas en redes sociales representan factores significativos que pueden afectar la percepción de la empresa y, por ende, el valor de sus acciones.

El problema que se aborda en este proyecto consiste en diseñar un modelo de predicción que no solo tome en cuenta la tendencia histórica de las acciones, sino que también integre datos provenientes de redes sociales. Esto implica capturar y analizar la frecuencia de menciones y el sentimiento asociado a la empresa en plataformas como X, anteriormente llamada Twitter, para entender cómo la opinión popular y la exposición en redes sociales pueden influir en el mercado. La principal dificultad reside en combinar de manera efectiva estos dos tipos de datos, financieros y sociales, para obtener una predicción más robusta y sensible a cambios que, en ocasiones, los datos históricos no reflejan de inmediato.

## ANÁLISIS

En este proyecto se propone la predicción de acciones en el mercado de valores mediante la integración de dos tipos de datos: históricos de mercado y datos de redes sociales, representando estos últimos la percepción pública de la empresa. La hipótesis central es que los movimientos en el precio de las acciones pueden verse influenciados no solo por los patrones históricos de precios y volúmenes de negociación, sino también por el sentimiento público reflejado en plataformas de redes sociales.

Para los datos financieros, se seleccionan características estándar del mercado, como el precio de apertura, cierre, volumen y precios ajustados, que son representativas de la actividad y el valor de las acciones. Estos datos ofrecen una visión de las tendencias de mercado a lo largo del tiempo, y, al ser históricos, facilitan la captura de patrones de comportamiento previos que los modelos LSTM pueden procesar de manera efectiva.

Simultáneamente, los datos de redes sociales representan un componente adicional que busca capturar el “sentimiento del mercado” de forma más inmediata. Este análisis se enfoca en tweets sobre Tesla, extrayendo y procesando el contenido textual de cada publicación para calcular puntajes de sentimiento diarios. En el análisis de sentimiento se utilizarán herramientas de procesamiento de lenguaje natural (NLP) que transforman cada tweet en un valor numérico, que luego se promedia diariamente. Este promedio diario de sentimiento se emplea para reflejar el tono general del público y su potencial influencia sobre el precio de las acciones.

La combinación de estos dos tipos de datos nos permitirá explorar la hipótesis de que el sentimiento en redes sociales puede complementar los datos históricos, aportando un componente de “opinión pública” que puede influir en el valor de las acciones. Esto con el fin de identificar si existe una correlación significativa entre el sentimiento público y los movimientos del mercado, y en caso de confirmarse, desarrollar un modelo que aproveche esta relación para realizar predicciones más precisas y sensibles a factores externos a los datos históricos financieros.

## PROPUESTA DE SOLUCIÓN

En primer lugar, se plantea utilizar un modelo LSTM (Long Short-Term Memory) para realizar la predicción de acciones basado en datos históricos, esto dado que son la opción más viable para predecir datos históricos debido a su capacidad única para manejar dependencias a largo plazo y secuencias de datos de manera eficiente. Las LSTM evitan el problema del gradiente desvanecido, lo que permite aprender y recordar patrones a lo largo de períodos extendidos. Además, su arquitectura con puertas de entrada, salida y olvido facilita la retención y el olvido de información específica, mejorando la precisión de las predicciones.

Los datos se obtendrán directamente de la presencia de la empresa, Tesla, en la bolsa de valores. Estos datos incluyen:

- Volumen: Es el total de acciones negociadas durante un período de tiempo determinado. Puede ser un indicador de la liquidez de la acción.
- Open Price (Precio de Apertura): Es el precio al cual una acción comienza a negociarse al inicio del período de trading.
- Close Price (Precio de Cierre): Es el precio al cual una acción termina de negociarse al final del período de trading.
- Low (Precio Mínimo): Es el precio más bajo al que una acción se negoció durante el período de trading.
- High (Precio Máximo): Es el precio más alto al que una acción se negoció durante el período de trading.
- Adj Close Price (Precio de Cierre Ajustado): Es el precio de cierre ajustado para reflejar eventos corporativos como dividendos, divisiones de acciones y emisiones de nuevas acciones.

De los cuáles, se seleccionará el precio de cierre ajustado como la variable objetivo, ya que refleja el valor real y ajustado de una acción, considerando eventos corporativos como dividendos y divisiones de acciones. Esto proporciona una visión más precisa y consistente del rendimiento de la acción a lo largo del tiempo, esencial para el tipo de modelo elegido.

Posteriormente, se llevará a cabo una limpieza de los datos de tweets, en la cual se seleccionará la fecha de creación del tweet, el contenido completo y el ID del mensaje. A estos datos se les aplicará una eliminación de stopwords y elementos que no aporten información relevante para optimizar el análisis de texto. El análisis de sentimientos se realizará utilizando una metodología basada en técnicas de procesamiento de lenguaje natural (NLP), con el fin de transformar el contenido textual en un puntaje numérico de sentimiento.

Este puntaje reflejará la percepción del público sobre Tesla en redes sociales, indicando si las opiniones son positivas, negativas o neutras. Posteriormente, se calculará un promedio diario de estos puntajes de sentimiento para correlacionarlo con el rendimiento de la acción, permitiendo explorar cómo las percepciones del público podrían influir en la valoración de la empresa en la bolsa de valores.

Por último, al obtener los puntajes promedio diarios de sentimiento, estos se integrarán con los datos históricos de precios de las acciones en una única fuente de datos, que contiene tanto información financiera como de percepción pública. Esta fusión de datos permite al modelo LSTM considerar no solo las tendencias de mercado, sino también las emociones y opiniones expresadas por el público en redes sociales. Al entrenar el modelo con esta información, se busca que el LSTM aprenda patrones complejos y relaciones entre el sentimiento del mercado y el precio de cierre ajustado (Adj Close Price) de Tesla. Esta combinación de datos financieros y sentimentales en un solo modelo de predicción ofrecerá un análisis integral que podría ayudar a anticipar cómo cambios en el sentimiento público podrían impactar en el rendimiento de las acciones de la empresa, proporcionando una herramienta útil para la toma de decisiones basada en datos de mercado y opinión pública.

## DESCRIPCIÓN DE LA SOLUCIÓN

### Análisis de sentimientos

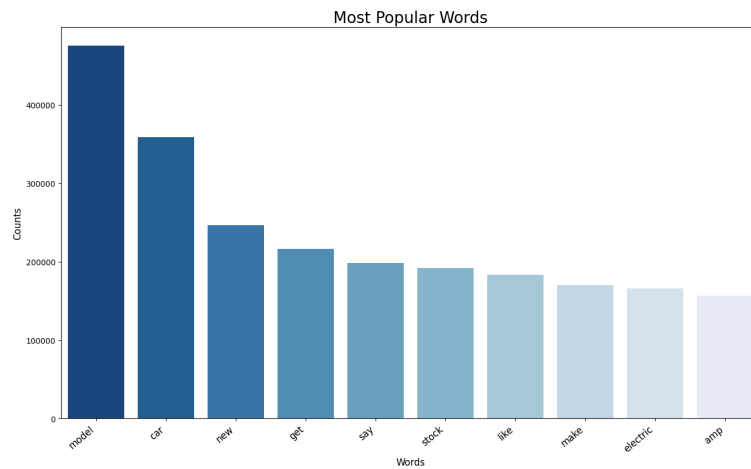
El análisis de sentimientos de tweets comienza con la instalación e importación de bibliotecas para procesamiento de datos, análisis de texto y visualización. Luego, se carga un conjunto de tweets sobre Tesla (TSLA) en un periodo específico (1 de enero de 2018 al 12 de enero de 2020), seleccionando solo las columnas clave (`created_at`, `id_str` y `full_text`) y convirtiendo `created_at` a formato de fecha para permitir la agregación diaria posterior. A continuación, se ejecuta una limpieza de texto con la función `clean()`, que estandariza el formato añadiendo un espacio al final de cada tweet, elimina enlaces URL, etc. Luego, cada tweet se tokeniza en palabras individuales, eliminando stopwords en inglés. Tras esta limpieza, se muestra una muestra de los tweets procesados para verificar los resultados.

Para refinar aún más el texto, se define y ejecuta una función de lematización `lemmatize()` que convierte palabras a su forma raíz (por ejemplo, "running" se convierte en "run"), ayudando a estandarizar términos y mejorar la precisión en el análisis de sentimientos. Esta función asigna una etiqueta gramatical a cada palabra en el tweet, luego convierte las etiquetas a un formato compatible con el lematizador `WordNetLemmatizer`, y finalmente transforma cada palabra en su forma base antes de reensamblar el tweet lematizado.

Posteriormente, se eliminan palabras comunes y específicas del tema, como "tesla," "tsla," "musk" y "elonmusk", mediante la función `final_clean()`. Estas palabras, que no aportan información única para el análisis de sentimientos, se eliminan para enfocar el análisis en términos relevantes. Luego, se combinan todos los tweets en una cadena única y se genera una lista de palabras individuales. Esto permite crear una distribución de frecuencias de palabras, la cual es visualizada mediante una nube de palabras. Esta visualización muestra los términos más frecuentes, ofreciendo una perspectiva general de los temas recurrentes en los tweets.

Adicionalmente, se genera una tabla de frecuencias con los 10 términos más comunes, que luego se representan en un gráfico de barras. Este gráfico permite visualizar la frecuencia de palabras de manera cuantitativa y complementa la nube de palabras, destacando los términos más mencionados en el conjunto de tweets.

Aquí se hace evidentemente la gráfica para poder visualizar los términos más comunes.



*Figura 1. Las 10 palabras más comunes en los tweets.*

Para facilitar el análisis posterior sin repetir la limpieza, el conjunto de datos procesado se guarda en un archivo CSV. Luego, el archivo se recarga, revisando y eliminando valores nulos para asegurar que el conjunto de datos esté completo antes de proceder al cálculo de los sentimientos.

En el análisis de sentimientos, se añade una nueva columna (`sentiment_score`) calculada mediante `TextBlob`, que asigna a cada tweet un puntaje de polaridad entre -1 (muy negativo) y +1 (muy positivo), representando el sentimiento general. Para validar estos puntajes, se evidencia una muestra aleatoria de tweets con sus versiones originales, limpias y el puntaje de sentimiento correspondiente. Este paso asegura que el análisis de sentimientos se alinea con la interpretación esperada.

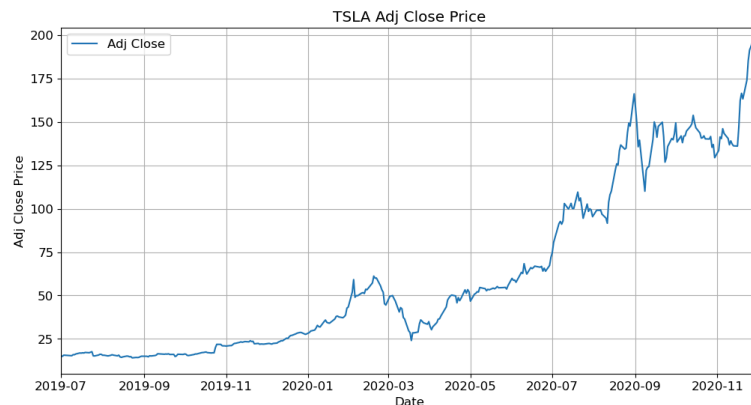
Luego, los puntajes de sentimiento se agrupan en intervalos para visualizar la distribución de los sentimientos en diferentes rangos. Un gráfico de barras muestra el conteo de tweets en cada intervalo de sentimiento, permitiendo observar cómo se distribuyen las opiniones en el tiempo. Finalmente, los datos se agrupan por día, manteniendo solo las columnas de `created_at` y `sentiment_score` para calcular el promedio diario de sentimiento. Este enfoque facilita la identificación de tendencias diarias en el sentimiento del público hacia la empresa.



## Stock Prediction

Se comienza con la importación de bibliotecas esenciales para la manipulación y el análisis de datos, obtención de datos bursátiles, visualización de gráficos, modelado y evaluación, y construcción de redes neuronales. También se incluye la biblioteca optuna para la optimización de hiperparámetros, la cual resulta clave para mejorar el rendimiento del modelo y optimizar su precisión.

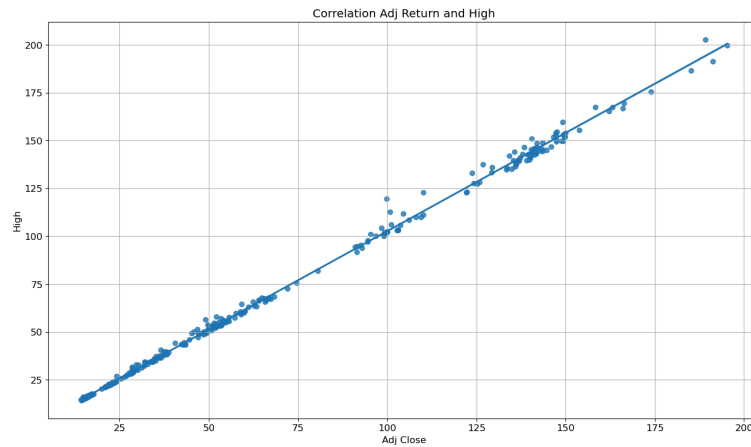
Como primera etapa, el programa descarga datos históricos de precios de Tesla (TSLA) desde Yahoo Finance, seleccionando el periodo entre el 1 de enero de 2018 y el 12 de enero de 2020. Para el análisis, se extraen las variables de interés: precios de apertura (Open), cierre (Close), máximo (High), mínimo (Low), volumen (Volume) y el precio ajustado de cierre (Adj Close). Posteriormente, se realiza una visualización de la serie temporal del precio ajustado de cierre para analizar la evolución histórica del precio de Tesla.



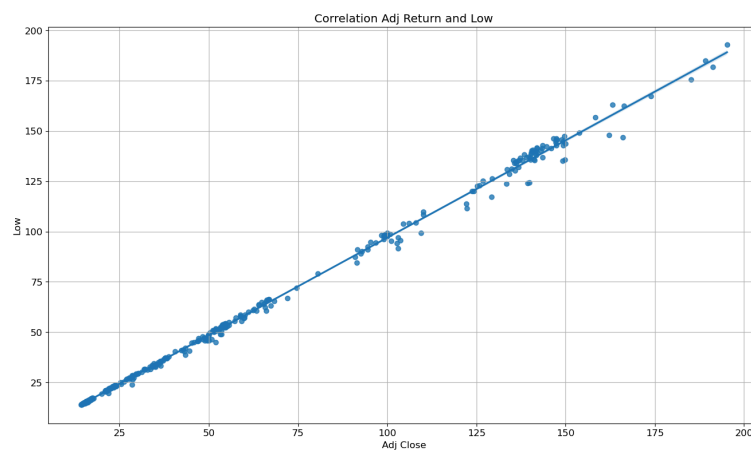
*Figura 2. Serie temporal del precio ajustado de cierre.*

Este gráfico inicial permite observar la tendencia general del activo y facilita la identificación de posibles cambios relevantes en el periodo analizado. Seguido de esta visualización, se procede a la preparación de los datos, separando las variables de entrada (Open, High, Low, Volume) de la variable objetivo (Adj Close) y normalizando todas las variables para que queden en un rango similar. La normalización es importante para estabilizar el proceso de entrenamiento de los modelos de redes neuronales. Además, se estructuran los datos en una secuencia temporal para alimentar a la red neuronal con una ventana de observación de precios previos, lo cual mejora la capacidad del modelo para captar patrones temporales.

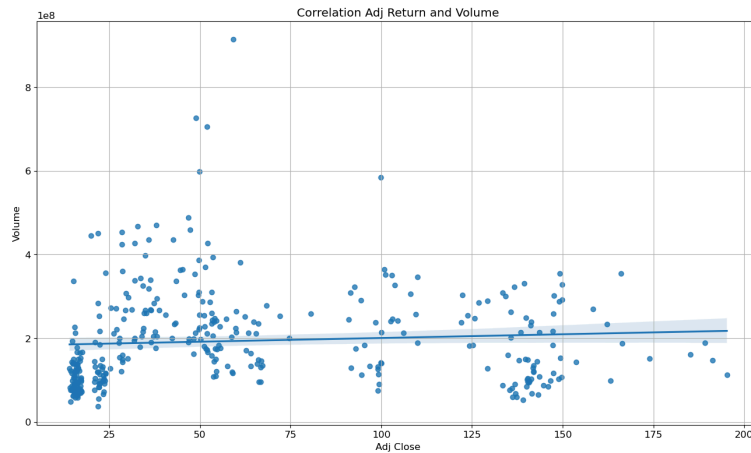
Para comprender mejor las relaciones entre variables, se realiza un análisis de correlación entre la variable objetivo y las variables High, Low y Volume. A través de gráficos de regresión, se visualizan las relaciones lineales entre estas variables y el precio ajustado de cierre, proporcionando una idea de la influencia de cada una sobre la variable objetivo.



*Figura 3. Correlación entre el precio de cierre y precio máximo de la acción.*



*Figura 4. Correlación entre el precio de cierre y precio mínimo de la acción.*



*Figura 5. Correlación entre el precio de cierre y el total de acciones negociadas durante un período de tiempo.*

El conjunto de datos se divide en subconjuntos de entrenamiento y prueba. El entrenamiento permite al modelo aprender patrones históricos, mientras que el conjunto de prueba se utiliza para evaluar el rendimiento del modelo en datos que no ha visto antes. Esta división temporal es crucial en series temporales, ya que asegura que el modelo no aprenda información futura durante su entrenamiento.

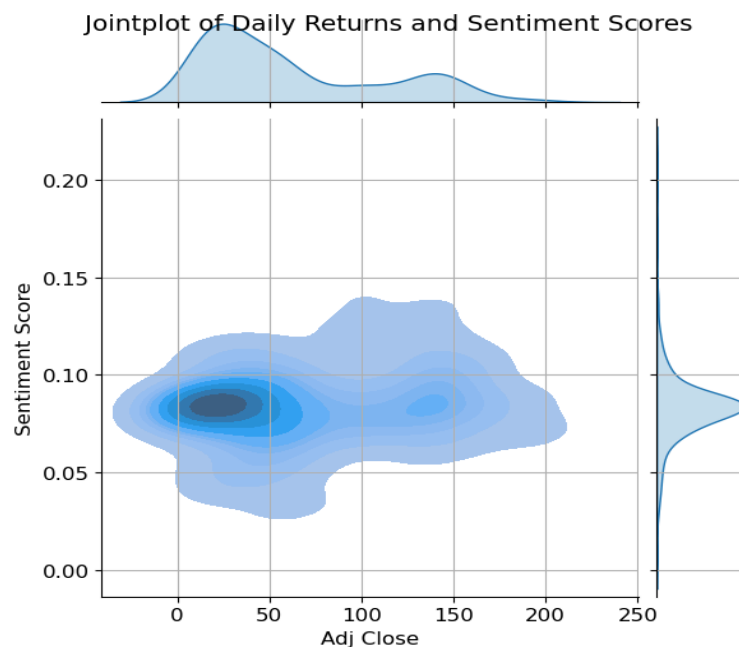
La fase de modelado se ejecuta mediante torch, utilizando una red neuronal recurrente de tipo LSTM, la red se configura con capas ocultas y funciones de activación que permiten captar patrones secuenciales en los precios de manera efectiva. El modelo se entrena minimizando el error cuadrático medio (MSE) entre las predicciones y los valores reales, lo cual ayuda a reducir el error en la predicción de precios. Para mejorar el rendimiento del modelo, Optuna se emplea en la búsqueda de hiperparámetros óptimos, probando diferentes configuraciones de parámetros, tales como el número de unidades en las capas de la red o la tasa de aprendizaje. Esta optimización permite encontrar la mejor configuración de la red neuronal para incrementar la precisión de las predicciones.

Una vez entrenado el modelo, se evalúa su rendimiento en el conjunto de prueba. Las predicciones se comparan con los valores reales de Adj Close y se calcula el error medio absoluto (MAE) y el error cuadrático medio (MSE). Los resultados de esta evaluación se grafican para visualizar el ajuste entre las predicciones y los valores reales del precio de cierre ajustado, permitiendo identificar si el modelo sigue adecuadamente la tendencia general o si existen desviaciones importantes.

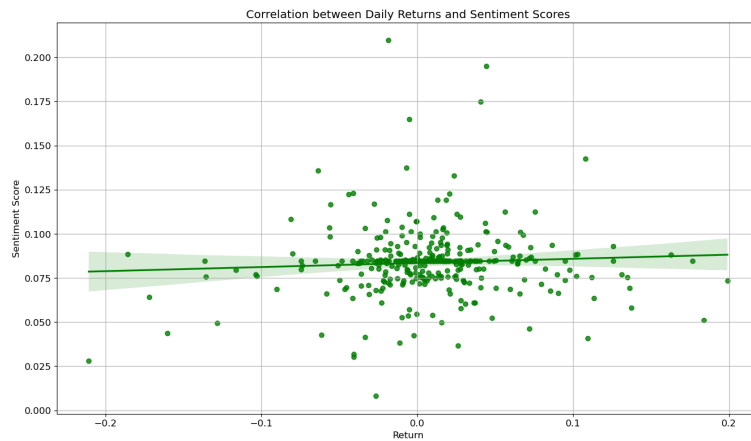
## Stock-Sentiment Prediction

Este programa combina datos históricos de precios de acciones con puntajes de sentimiento para prever el comportamiento futuro de las acciones de Tesla. Al comienzo, se cargan los puntajes de sentimiento diarios desde un archivo CSV, estableciendo el índice como las fechas y asegurándose de que los valores nulos se reemplacen con el promedio de la columna. Esto asegura una continuidad en los datos de sentimiento. Luego, se descargan datos de precios históricos de Tesla entre el 1 de julio de 2019 y el 2 de diciembre de 2020 mediante yfinance. Se seleccionan las variables clave: precios de apertura (Open), cierre (Close), máximo (High), mínimo (Low), volumen (Volume) y el precio ajustado de cierre (Adj Close) que se analizarán junto con los datos de sentimiento.

Para visualizar el comportamiento de las acciones, se traza una gráfica del precio ajustado de cierre de TESLA, permitiendo observar la tendencia de precios y los movimientos significativos en el período. Esta etapa ayuda a contextualizar la serie temporal de precios antes de integrarse con los puntajes de sentimiento, que ofrecen una perspectiva sobre cómo las opiniones en redes sociales pueden relacionarse con las fluctuaciones de precios.



*Figura 6.* Gráfico conjunto de los precios de cierre ajustados y las puntuaciones de sentimiento.



*Figura 7.* Correlación entre el rendimiento diario y el puntaje de sentimientos.

El proceso continúa con la normalización de los datos de precios y sentimiento para garantizar que todas las variables se encuentren en un rango similar, evitando que algunas dominen sobre otras en el entrenamiento del modelo. Se preparan las variables de entrada y salida para el modelo de predicción. El modelo en sí se basa en una red neuronal recurrente LSTM construida con torch, configurada para analizar secuencias de datos y captar patrones temporales. Esta red recibe los datos de precios y de sentimiento como entrada para aprender patrones históricos que puedan indicar movimientos futuros en el precio de las acciones.

Una vez que se construye la red, el modelo se entrena minimizando el error cuadrático medio (MSE) entre las predicciones y los valores reales de precios. Además, se implementa Optuna para optimizar los hiperparámetros clave, tales como el número de capas y la tasa de aprendizaje, asegurando un rendimiento óptimo del modelo. Posteriormente, el modelo se evalúa en un conjunto de prueba separado para verificar su capacidad de predicción en datos no vistos.

Finalmente, se calcula y grafica el error de predicción para visualizar la precisión del modelo. La predicción resultante ofrece una representación visual de la tendencia de precios estimada en comparación con los valores reales, brindando una medida de qué tan bien el modelo sigue los movimientos de mercado. Este flujo de trabajo permite una visión combinada de los factores de mercado y de sentimiento, proporcionando un modelo que podría aplicarse a otras acciones con ajustes en los datos y configuraciones de parámetros.

## HERRAMIENTAS APLICADAS

### Predicción de acciones

- a. yfinance: Utilizada para descargar datos de mercado históricos y actuales desde Yahoo Finance, específicamente implementada en la recuperación de datos de acciones.
- b. Sklearn:
  - Min Max Scaler: Función que escala y traduce cada característica individual a un rango dado, normalmente entre 0 y 1. Utilizada para normalizar los datos antes de entrenar los modelos LSTM, asegurando que todas las características contribuyan uniformemente.
  - Parameter Sampler: Función que genera muestras de parámetros utilizando una distribución especificada. Utilizada en la búsqueda de hiperparámetros para explorar diferentes combinaciones de parámetros y encontrar los mejores para el modelo.
  - Mean Squared Error: Función que calcula el error cuadrático medio entre los valores predichos y los verdaderos. Utilizada para medir la precisión del modelo.
- c. Scipy: Utilizada para generar distribuciones aleatorias uniformes de enteros y números decimales respectivamente. En este caso, aplicada para generar parámetros aleatorios durante la búsqueda de hiperparámetros.
- d. Torch:
  - torch.nn: Un submódulo de PyTorch que proporciona herramientas y clases para construir redes neuronales, específicamente para la arquitectura LSTM, que permite crear su estructura y modificar sus parámetros.
  - torch.optim: Un submódulo de PyTorch que incluye varios algoritmos de optimización para ajustar los parámetros del modelo durante el entrenamiento. Utilizada para definir los métodos de optimización que ajustarán los pesos del modelo para minimizar la función de pérdida.

## Procesamiento de texto

- e. NLTK (Natural Language Toolkit): Es una librería de Python para trabajar con texto. Proporciona herramientas para el preprocesamiento de texto, como la eliminación de stopwords y la lematización. Se utilizó para descargar y aplicar una lista de stopwords en inglés, así como para realizar la lematización de palabras, reduciendo las palabras a su raíz.
- f. Regex (Regular Expressions): Utilizadas para buscar y manipular texto mediante patrones. En este caso, se usaron para limpiar el texto de los tweets.
- g. Lematización (WordNetLemmatizer de NLTK): La lematización reduce las palabras a su forma base, lo cual es útil en análisis de texto para agrupar términos similares.

## Análisis de sentimientos

- h. TextBlob: Es una librería de procesamiento de lenguaje natural (NLP) que simplifica tareas como el análisis de sentimiento y la traducción de texto. Fue utilizada para realizar análisis de sentimiento en los tweets, proporcionando un `sentiment_score` para cada uno.

## Otros

- i. Pandas: Utilizado para cargar, manipular y analizar datos estructurados.
- j. Matplotlib y Seaborn: Utilizados para la visualización de datos. Permite un control detallado de gráficos, y facilita la creación de gráficos estadísticos.

## RESULTADOS

### Stock Prediction

Para las predicciones del precio de cierre ajustado de Tesla en la bolsa, basado únicamente en la información extraída de esta última, se realizó un el primer modelo LSTM (1.0), a este se le dio una estructura estándar con hiperparámetros base predefinidos. Mostró un rendimiento aceptable en términos de seguir tendencias generales, pero con una desviación significativa en la magnitud de las predicciones.

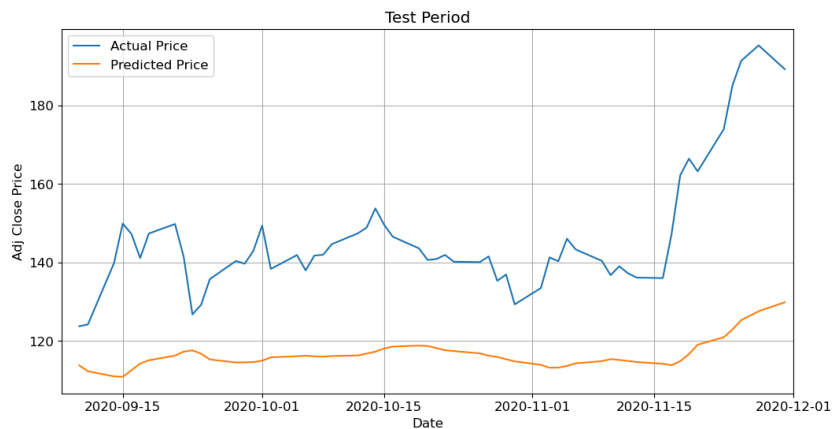


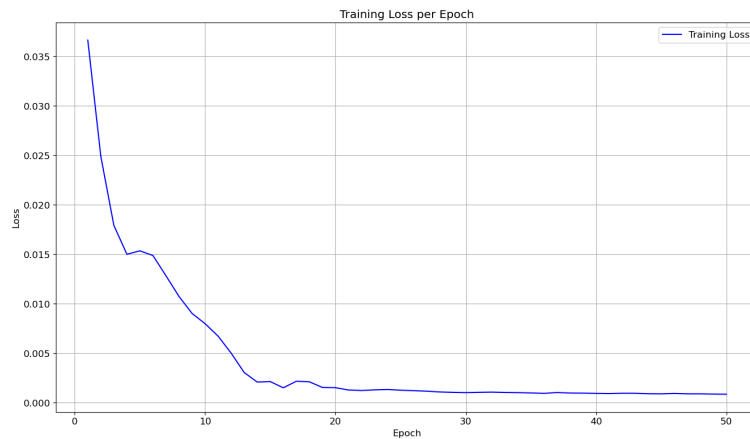
Figura 8. Comparación de precios de cierre ajustados reales y predichos.

MSE	MAE	RMSE
1031.5665	29.4895	32.1180

Tabla 9. Métricas de precisión para la predicción.

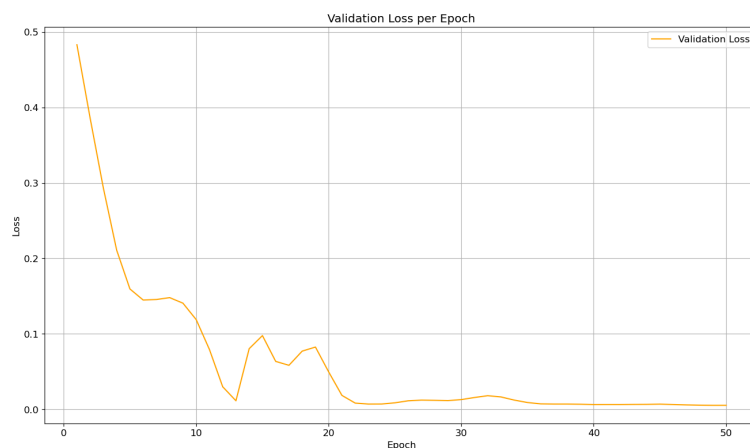
Las métricas de error reflejan esta discrepancia entre los datos reales y predichos, y los gráficos de pérdida sugieren un ajuste que podría beneficiarse de más regularización o ajuste fino de hiperparámetros. Por esto mismo, se definió una etapa de entrenamiento intensiva para encontrar los hiperparámetros más adecuados. De esta fase de ajuste se generó un nuevo modelo LSTM (2.0), con una estructura basada en el original pero mejor condicionado, además, se disminuyó en un 30% la cantidad de épocas de entrenamiento, con tal de evitar que el modelo sufriera de sobreajuste con los datos de entrenamiento y de beneficiar su capacidad de generalización.





*Figura 10. Pérdida de entrenamiento por época.*

La figura 10 demuestra que la pérdida de entrenamiento disminuye de manera constante y se estabiliza en niveles muy bajos, indicando que el modelo efectivamente logró aprender de los datos de entrenamiento sin signos evidentes de sobreajuste o subajuste.

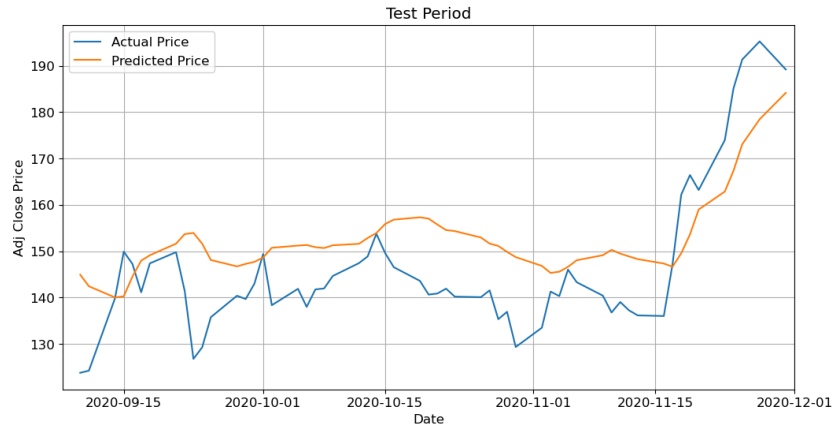


*Figura 11. Pérdida de validación por época.*

En la figura 11 se observa una disminución significativa en la pérdida hasta estabilizarse, indicando que el modelo mejora con cada época al principio, y luego llega a una meseta, lo que sugiere que ha alcanzado su capacidad óptima de ajuste para los datos de validación.

MSE	MAE	RMSE
139.1785	10.1357	11.7974

*Tabla 12. Métricas de precisión para la predicción.*



*Figura 13. Comparación de precios de cierre ajustados reales y predichos.*

Las métricas indican que el modelo tiene un nivel de error moderado. El RMSE de 11.7974 sugiere que, en promedio, la predicción del modelo difiere del valor real en alrededor de 11.8 unidades, aunque estos valores de error son manejables dado el contexto de estudio.

A simple vista, el modelo logra capturar parcialmente la tendencia general, pero presenta desviaciones considerables en ciertos puntos, especialmente hacia finales del período de prueba, donde los valores reales muestran un aumento pronunciado y las predicciones no siguen con la misma rapidez. Sin embargo, se observa que sí es extraíble cierto tipo de información relevante respecto a los precios de cierre ajustados a futuro, denotando que la solución propuesta está encaminada a predecir valores acertados.

## Stock-Sentiment Prediction

Para las predicciones del precio de cierre ajustado de Tesla en la bolsa, agregando los puntajes de sentimientos de tweets relacionados con la empresa, se realizó un modelo LSTM (3.0), al igual que en el primer sistema de predicción, este modelo tiene estructura estándar con hiperparámetros base predefinidos. Se observó que las predicciones se ajustan mucho mejor a los valores reales, especialmente en términos de la tendencia general. Aunque las predicciones aún no capturan a la perfección los cambios abruptos, la línea de predicción sigue de manera más consistente la dirección de los datos reales.

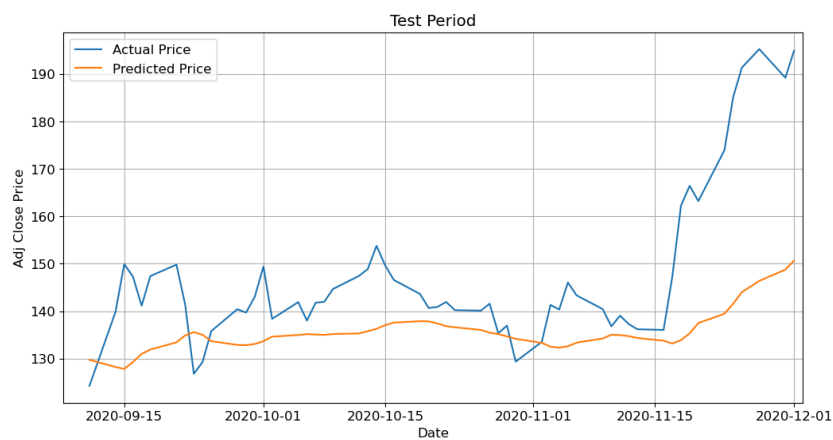
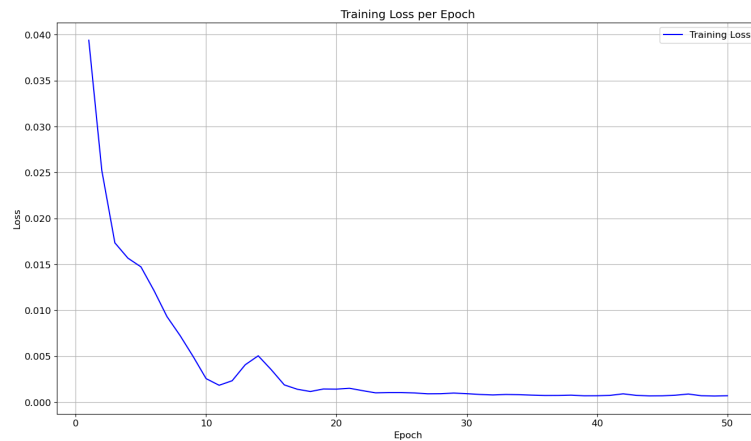


Figura 14. Comparación de precios de cierre ajustados reales y predichos.

MSE	MAE	RMSE
314.6341	12.6178	17.7379

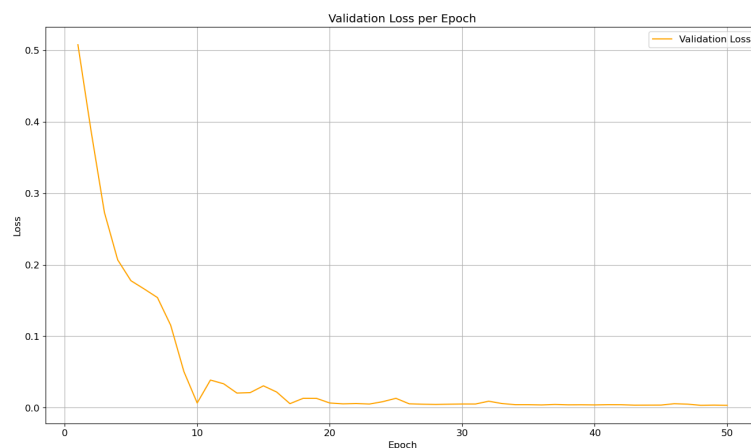
Tabla 15. Métricas de precisión para la predicción.

Las métricas de error, que son considerablemente más bajas en comparación con el modelo de predicción solo basado en las acciones, indicando una mejora notable en la precisión, lo cual sugiere que el análisis adicional de los puntajes de sentimiento aporta valor en la precisión de la predicción. No obstante, se decidió nuevamente continuar desarrollando y optimizando un nuevo modelo LSTM (4.0) explorando el ajuste preciso de los hiperparámetros.



*Figura 16. Pérdida de entrenamiento por época.*

La pérdida de entrenamiento disminuye de forma rápidamente en las primeras épocas, lo que indica que el modelo está aprendiendo los patrones de los datos. Después de alrededor de 10 épocas, la pérdida de entrenamiento se estabiliza en un valor muy bajo y se mantiene así hasta el final.

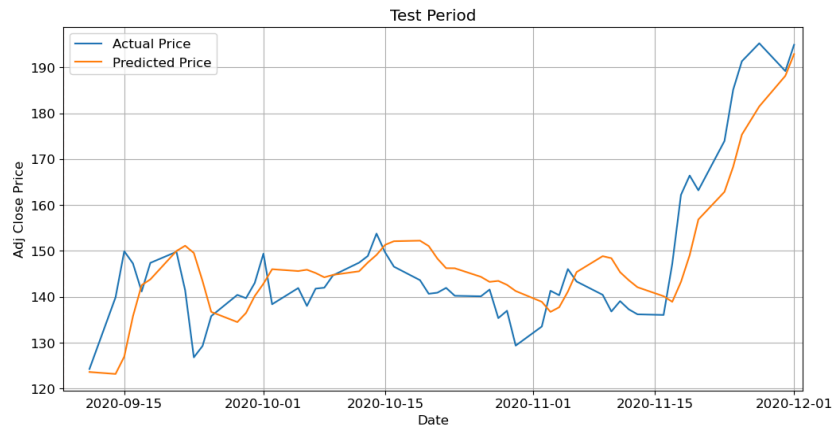


*Figura 17. Pérdida de validación por época.*

La pérdida de validación también muestra una caída pronunciada en las primeras épocas, estabilizandose después de las 10-15 épocas. Demostrando que el modelo no está sobreajustando, dado que la pérdida de validación se comporta de manera similar a la de entrenamiento y alcanza un valor bajo y constante. La presencia de un ligero valle en las primeras 10 épocas indicó que se realizó un ajuste efectivo y una correcta sintonización de hiperparámetros.

MSE	MAE	RMSE
82.0883	7.1163	9.0603

*Tabla 18. Métricas de precisión para la predicción.*



*Figura 19. Comparación de precios de cierre ajustados reales y predichos.*

La línea de predicción sigue de cerca la tendencia general de los valores reales, lo que indica que el modelo está capturando bien el comportamiento del mercado. A diferencia de modelos previos que mostraban desviaciones significativas, este modelo muestra una alineación mucho mejor en términos de tendencia. En los momentos donde hay un incremento o descenso pronunciado (hacia finales del período de prueba), la predicción sigue la dirección, aunque con una ligera desviación de magnitud. La predicción aún tiene dificultades para ajustarse a los picos y valles más abruptos. Sin embargo, logra una mejor aproximación en comparación con modelos previos, lo cual sugiere que la implementación del análisis de puntajes de sentimientos de los tweet y el tuneo de hiperparámetros contribuyeron a mejorar su precisión y robustez, siendo este modelo LSTM (4.0) el más confiable y completo.

## CONCLUSIONES

- La integración de datos de mercado con análisis de sentimientos de redes sociales mejora significativamente la precisión en la predicción de precios de acciones.
- Los modelos LSTM basados únicamente en datos históricos de mercado presentan limitaciones significativas. Si bien, se puede llegar a seguir tendencias generales, los datos de mercado por sí solos no son suficientes para predecir con precisión los precios de las acciones de una empresa.
- El modelo LSTM optimizado que combina datos del mercado y sentimientos de la opinión pública es el más confiable y completo para la predicción de acciones. Esto sugiere que la combinación de datos heterogéneos y un ajuste cuidadoso del modelo conduce a mejoras sustanciales en la predicción de precios de acciones.
- El ajuste de hiperparámetros y la prevención del sobreajuste mejoran el rendimiento del modelo, sin embargo, aún persisten errores moderados y dificultades para capturar cambios abruptos en los precios, reflejando la complejidad inherente del mercado bursátil.

## RECOMENDACIONES

- Incorporar fuentes de datos adicionales para enriquecer el modelo predictivo. Por ejemplo, información como noticias financieras, indicadores económicos, eventos globales y datos macroeconómicos.
- Implementar un proceso continuo de ajuste y actualización del modelo. Dado que las condiciones del mercado financiero son dinámicas, es esencial recalibrar regularmente los hiperparámetros y la arquitectura del modelo para adaptarse a nuevos patrones y tendencias.

## REPOSITORIO

El código desarrollado puede ser encontrado en el [repositorio de GitHub](#).

## PRESENTACIÓN

La presentación puede ser encontrada en el siguiente [enlace](#).

## VIDEO DE FUNCIONAMIENTO

El video del funcionamiento puede ser encontrado en el siguiente [enlace](#).

## BIBLIOGRAFÍA

Hindy. (2020). *Tesla tweets*. Kaggle.  
<https://www.kaggle.com/datasets/hindy51/tesla-tweets>

Sayah, F. (2022). *Stock Market Analysis + Prediction using LSTM*. Kaggle.  
<https://www.kaggle.com/code/faressayah/stock-market-analysis-prediction-using-lstm>

Siddharth (2024). *Stock Price Prediction using LSTM and its Implementation*. Analytics Vidhya.  
<https://www.analyticsvidhya.com/blog/stock-price-prediction-using-lstm/>

Xiao, D., Yue, Z., Ting, L. y Junwen, D. (2015). *Deep Learning for Event-Driven Stock Prediction*. Universidad de Singapur.  
[https://scholar.google.com.gt/scholar\\_url?url=http://www.wins.or.kr/DataPool/Board/4xxxx/455xx/45587/329.pdf](https://scholar.google.com.gt/scholar_url?url=http://www.wins.or.kr/DataPool/Board/4xxxx/455xx/45587/329.pdf)