



## Artículo

# Un método de aprendizaje automático para la predicción del mercado de valores Uso de datos de Twitter en tiempo real

Saleh Albahli <sup>1,\*</sup>, Aun Irtaza <sup>2,3</sup>, Tahira Nazir <sup>2</sup>, Awais Mehmood <sup>2</sup>, Ali Alkalifah <sup>1</sup> y Waleed Albattah <sup>1</sup>

<sup>1</sup> Department of Information Technology, College of Computer, Qassim University,

Buraydah 52571, Saudi Arabia Department of Computer Science, University of

<sup>2</sup> Engineering and Technology, Taxila 39161, Pakistan Department of Computer and Electrical

<sup>3</sup> Engineering, University of Michigan, Dearborn, MI 48128, EE . UU . \* Correspondencia: salbahli@qu.edu.sa

**Resumen:** Las finanzas representan uno de los requisitos fundamentales para realizar cualquier actividad útil para la humanidad. Los mercados financieros, por ejemplo, las bolsas de valores, divisas y bolsas mercantiles, etc., brindan la oportunidad a cualquiera de invertir y generar finanzas. Sin embargo, para obtener los máximos beneficios de estos mercados financieros, se requiere una toma de decisiones eficaz para identificar las direcciones comerciales, por ejemplo, ir en largo/corto analizando todos los factores influyentes, por ejemplo, la acción del precio, las políticas económicas y la estimación de oferta/demanda, en de manera oportuna. En este sentido, el análisis de las noticias financieras y las publicaciones de Twitter juegan un papel importante para predecir el comportamiento futuro de los mercados financieros, la estimación del sentimiento público y la estimación del riesgo sistemático/idiosincrásico. En este artículo, nuestro trabajo propuesto tiene como objetivo analizar las publicaciones de Twitter y los datos de Google Finance para predecir el comportamiento futuro de los mercados bursátiles (uno de los mercados financieros clave) en un marco de tiempo particular, es decir, por hora, día, semana, etc. , a través de un novedoso modelo StockSentiWordNet (SSWN). El modelo SSWN propuesto amplía el léxico de opinión estándar llamado SentiWordNet (SWN) a través de los términos relacionados específicamente con los mercados de valores para entrenar la máquina de aprendizaje extremo (ELM) y la red neuronal recurrente (RNN) para la predicción del precio de las acciones. Los experimentos se realizan en dos conjuntos de datos, es decir, los conjuntos de datos Sentiment140 y Twitter, y lograron un valor de precisión del 86,06 %. Los hallazgos muestran que nuestro trabajo supera los enfoques más avanzados con respecto a la precisión general. En el futuro, planeamos mejorar la capacidad de nuestro método agregando otras redes sociales populares, por ejemplo, Facebook y Google News, etc.

**Palabras clave:** aprendizaje automático; SentiWordNet; predicción de existencias; análisis de los sentimientos



Cita: Albahli, S.; Irtaza, A.; Nazir,

t.; Mehmood, A.; Alkalifah, A.;

Albattah, W. Aprendizaje automático

Método para la Predicción de Stock

Mercado utilizando datos de Twitter en tiempo real.

Electrónica **2022**, *11*, 3414.

[https://doi.org/10.3390/](https://doi.org/10.3390/electronics11203414)

electronics11203414

Editores académicos: Yanhui Guo,

Deepika Koundal y Rashid Amin

Recibido: 19 Septiembre 2022

Aceptado: 4 de octubre de 2022

Publicado: 21 de octubre de 2022

**Nota del editor:** MDPI se mantiene neutral con respecto a reclamos jurisdiccionales en mapas publicados y afiliación institucional acciones.



**Copyright:** © 2022 por los autores.

Licenciario MDPI, Basilea, Suiza.

Este artículo es un artículo de acceso abierto distribuido bajo los términos y condiciones de Creative Commons

Licencia de atribución (CC BY) (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introducción

La fluctuación del precio de las acciones significa las tendencias existentes del mercado y la evolución de la empresa que se pueden medir para vender o comprar acciones. Una estimación del mercado de valores ha sido considerada como una de las tareas más desafiantes y esenciales debido a su comportamiento no lineal o dinámico [1].

Los precios de las acciones suben y bajan cada minuto o incluso cada segundo debido a las variaciones en la oferta y la demanda. Si un grupo de individuos quiere comprar una acción específica, su precio aumentará. Mientras que, cuando la mayoría de las personas que poseen una acción específica quieren venderla, su precio de mercado disminuirá. Esta asociación entre la oferta y la demanda está ligada a las noticias, los blogs y el análisis de sentimientos (SA), etc. La predicción del mercado de valores que utiliza SA se ocupa del rendimiento automático [2] del mercado de valores. En este sentido, Twitter es la plataforma más popular que se puede utilizar para predecir la opinión pública, por lo que puede ser útil para pronosticar el precio del mercado de valores [3].

Hoy en día, ha habido un debate sobre la efectividad de los sentimientos transmitidos a través de las redes sociales para pronosticar el cambio en el mercado de valores. Varios investigadores han revelado que los sentimientos pueden influir en el movimiento del mercado de valores y actuar como predictores potenciales de resultados de compensación [4,5]. Además, los diferentes métodos de minería de sentimientos se pueden emplear de manera diferente en numerosas circunstancias bursátiles [6]. En otras palabras,

Hay muchas responsabilidades involucradas en la evaluación de opiniones sobre los rasgos y características de las acciones. [7,8]. Sin embargo, las técnicas existentes no sugieren una confianza absoluta en el número de tweets por unidad de tiempo. La cantidad de datos recopilados y analizados durante los estudios existentes siguen siendo inadecuados, lo que provoca predicciones con baja precisión [9,10].

Aunque la comunidad de investigación ha presentado técnicas extensas para la predicción del mercado de valores, estos enfoques tienen algunas limitaciones potenciales. Los métodos existentes no son sólidos para abordar la naturaleza versátil de las acciones. Además, el tamaño masivo de los datos requiere métodos que puedan aprender un conjunto más confiable de características para demostrar mejor los comportamientos variables de las acciones a lo largo del tiempo. Por lo tanto, existe la necesidad de mejorar el rendimiento tanto para la precisión de la predicción de tendencias de stock como para la complejidad del tiempo.

Para hacer frente a los problemas de los enfoques actuales, proponemos la técnica a saber, el clasificador SSWN con ELM para la predicción del mercado de valores. El método presentado comprende tres pasos principales que son la recopilación de datos, el cálculo de sentimientos junto con el entrenamiento del modelo y, finalmente, el módulo de predicción del mercado de valores. De manera más descriptiva, estas son las contribuciones de este documento: 1. Se propone un marco eficiente, a saber, SSWN, con clasificadores ELM y RNN para la predicción del comportamiento del mercado de valores.

2. Utilización de SA para la predicción del mercado de valores y modificación de SWN mediante la introducción de nuevos términos relacionados con el mercado de valores.
3. Asignación de puntajes de sentimiento a términos relacionados con el mercado de valores recientemente introducidos mediante la aplicación del método de ganancia de información, lo que resulta en el desarrollo de un nuevo léxico de sentimiento SSWN.
4. Realizar análisis comparativos con otros métodos para mostrar la efectividad del método propuesto.

El resto de este documento está estructurado de la siguiente manera: la Sección 2 muestra el trabajo relacionado. El método propuesto se presenta en la Sección 3. Los experimentos y los resultados se describen en la Sección 4, mientras que la Sección 5 concluye nuestro trabajo.

## 2. Trabajo relacionado

Se han exhibido numerosos estudios [11–21] sobre el uso de conocimientos electrónicos para pronosticar tendencias de existencias. Por ejemplo, Zhang et al. [22] propusieron un método basado en LSTM para estimar la tendencia del mercado de valores. En el primer paso, la entrada se divide en tres partes: espacio de opinión abierto, transacciones bursátiles y datos de transacciones de mercado. El LSTM de una capa se empleó para preparar la memoria larga en el espacio de la opinión pública, mientras que se aplicaron dos capas de LSTM para entrenar la memoria corta en series bursátiles y de mercado. Después de esto, los datos se combinaron utilizando la capa fusionada y se utilizó una capa lineal para mejorar los resultados del modelo. El método predice el comportamiento del mercado y evalúa la relación entre las emociones de los inversores y los datos de las transacciones. Sin embargo, el método necesita más mejoras en la técnica de abstracción de emociones. Xu et al. [23] presentó un método para el pronóstico del mercado de valores mediante la introducción de la SA. Inicialmente, el conjunto de datos se recopila mediante un proceso heurístico medio-fin y luego se identifican los sentimientos a partir de los datos adquiridos. SA se combinó con el estudio de eventos y el resultado se utilizó como entrada del análisis de componentes principales (PCA), que se utilizó para análisis posteriores. El método predice el comportamiento del mercado utilizando SA con una precisión del 84,89%. Sin embargo, el método enfrenta problemas relacionados con la estabilidad y existe una desigualdad entre el pronóstico y los valores reales.

Wu et al. [24] propusieron un método de aprendizaje profundo (DL) para la predicción de los sentimientos de excitación de valencia dimensional de las acciones en el mercado de valores. El método utilizó el título, las palabras clave y la descripción general de los mensajes relacionados con el mercado de valores para la estimación de todos los vectores utilizando el enfoque de atención hierática. El método logró el éxito, produciendo mejores resultados. Sin embargo, no puede identificar las palabras con múltiples significados y también necesita algunas mejoras de estabilidad. Similar a la técnica antes mencionada, se empleó un método basado en DL en [22] para la extrapolación del mercado de valores usando análisis de sentimiento. El modelo

se basa en técnicas RNN y LSTM que luego se utilizan para definir los sentimientos en clases positivas y negativas. El aumento o la disminución de los precios de las acciones se pronostica a partir del análisis de sentimientos. Rene et al. [15] presentó un marco para la predicción mediante el examen de los sentimientos de los inversores. Inicialmente, el contenido revisado financiero se recopiló de dos sitios, a saber, Sina Finance y Eastmoney. Luego, el SVM fue entrenado sobre los datos financieros para predecir un índice esencial en China, a saber, el índice SSE 50, mediante la aplicación de una técnica de validación cruzada de cinco veces. El método confirmó que fusionar los puntos clave de sentimiento con los datos del mercado de valores puede obtener resultados sólidos en comparación con utilizar solo datos del mercado de valores para estimar la dirección del movimiento. Sin embargo, esta técnica no es robusta para analizar grandes datos en tiempo real. Bouktif et al. [14] introdujo un enfoque para predecir las direcciones futuras del mercado de valores. Inicialmente, los datos bursátiles se recopilan de recursos en línea junto con tweets públicos. En el segundo paso, se aplicó el enfoque NLP para calcular los puntos clave informativos de los tweets. Luego, se entrenaron varios métodos basados en ML, a saber, Naive Bayes, regresión logística, SVM, ANN, random forest y XGBoost, para clasificar los datos. La técnica necesita mejoras adicionales para características textuales complejas.

Kelotra et al. [13] ofreció una técnica basada en DL, a saber, la optimización de la mariposa monarca Rider (MBO), basada en el marco ConvLSTM para la predicción del mercado de valores. En el primer paso, los datos de entrada se recopilaron del mercado de ganado que se pasó al proceso de cálculo de puntos clave para calcular el conjunto representativo de características basado en indicadores técnicos. En el siguiente paso, se empleó la técnica de agrupamiento, a saber, C-means dispersas y difusas (FCM), sobre los puntos clave extraídos para agruparlos. Después de esto, los puntos clave más importantes se pasaron a la red Deep-ConvLSTM basada en RiderMBO presentada para realizar la predicción. Otro enfoque de predicción del mercado de valores basado en el análisis de sentimientos se presentó en [12], que hace uso de características profundas textuales computadas. Después de recopilar los datos del mercado de valores, se emplearon CNN y RNN para calcular las características profundas. Después de esto, se aplicaron los algoritmos PCA y LDA para extraer el conjunto significativo de características. Finalmente, se entrenó el clasificador SVM sobre las características calculadas para la predicción de movimientos bursátiles. El modelo funciona bien para la predicción del mercado de valores, pero es posible que no muestre un mejor rendimiento en escenarios del mundo real. De manera similar, en [11], se presentó un marco basado en DL que emplea análisis de sentimiento para la predicción del mercado de valores. El modelo LSTM se utilizó para pronosticar los valores de cierre futuros de un mercado de valores. Al admitir los tweets solo en inglés, este método es robusto para calcular los movimientos del

Las respuestas de los usuarios de artículos históricos se pueden emplear para predecir los comportamientos de los consumidores con el tiempo. Uno de estos métodos se presentó en [25] utilizando un enfoque de CNN dual con comportamientos de usuario para incorporar tanto la información semántica como la estructural de los artículos de texto.

Otro enfoque que emplea información divulgada del Pilar 3 se presentó en [26] y se centró en la investigación de los intereses y el comportamiento de los usuarios de depósitos utilizando información de sitios web que estaban profundamente arraigados en las divulgaciones de los bancos comerciales. El marco regulatorio del Pilar 3 tenía como objetivo fortalecer la estabilidad de precios garantizando la rendición de cuentas y mejorando las divulgaciones públicas de las instituciones financieras. El trabajo [26] funciona bien para analizar el comportamiento del consumidor. Sin embargo, el modelo necesita evaluación en un conjunto de datos estándar.

### 3. Metodología propuesta

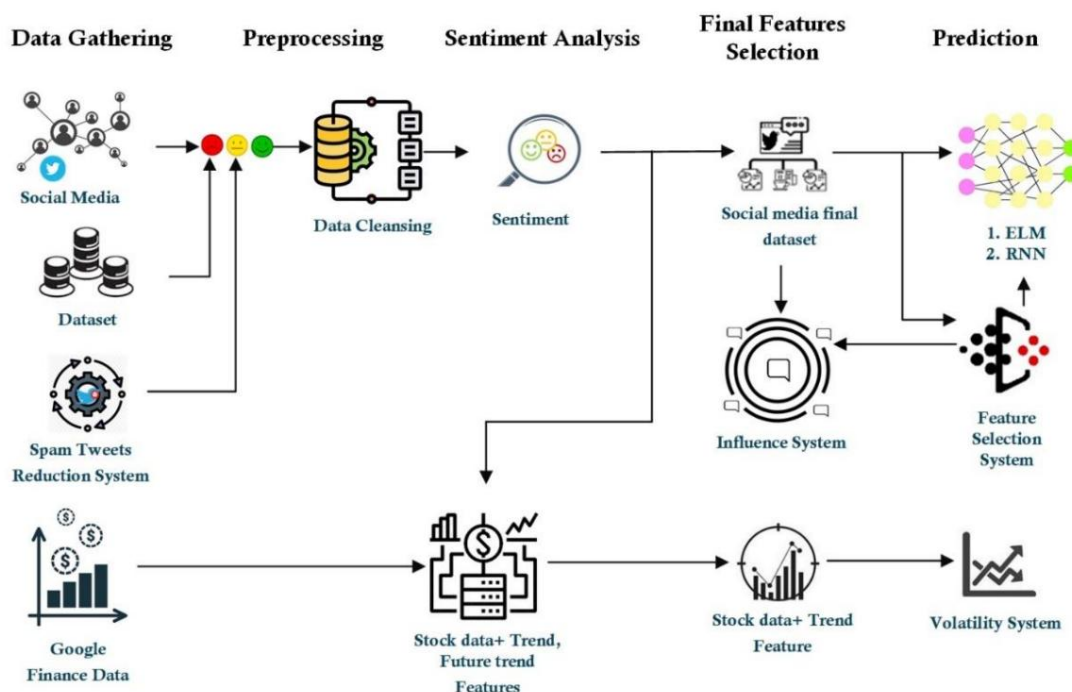
Nuestra técnica propuesta abarca tres pasos: recopilación de datos, extracción de sentimientos, capacitación y predicción del mercado de valores.

#### 3.1. Recopilación y limpieza de datos

Primero, recopilamos datos de Twitter. Esta plataforma de redes sociales se selecciona debido a su concisión. Además de los datos de tweets extraídos directamente de Twitter, hemos utilizado el conjunto de datos de última generación llamado Sentiment140 [27]. Después de la adquisición de datos, limpiamos estos datos recopilados mediante la eliminación de spam, tweets redundantes, sin sentido o irrelevantes mediante el uso de un sistema de reducción. El paso de preprocesamiento incluye además lo siguiente: • Conversión de tweets en tokens de palabras usando bigramas, lo que significa que el modelo evalúa dos tokens/palabras al mismo tiempo. Esto significa que si un tweet describe

- Conversión de tweets en tokens de palabras mediante el uso de bigramas, lo que significa que el modelo evalúa dos tokens/palabras al mismo tiempo. Esto significa que si un tweet describe algo como "no bueno", eso se considerará como un comentario negativo, en lugar de positivo solo porque contiene la palabra "bueno".
  - Eliminación de etiquetas de autor (@). Estas etiquetas deben eliminarse porque son positivas solo porque contienen el nombre de la persona.
  - Eliminación de hashtags como la etiqueta de autor (@). Estas etiquetas deben eliminarse porque no contienen conocimientos valiosos para obtener sentimientos.
  - Eliminación de stopwords. Las palabras vacías existen con frecuencia en los tweets como an, is, are, the, de las palabras vacías, más palabras para las palabras de alta frecuencia en tweets como an, is, are, the, etc., y se eliminan para reducir el ruido.
  - Conversión de palabras repetidas en una sola palabra, como la eliminación de palabras.
  - Eliminación de palabras duplicadas.
  - Eliminación de tweets duplicados.

5. Eliminación de tweets duplicados. Después del preprocesamiento, el conjunto de datos limpio se usa para la extracción de características y la opinión identificada. Después de eliminar los tweets duplicados, el conjunto de datos limpio se transforma en un conjunto de datos de opinión. Este proceso transforma los datos brutos de Twitter en un conjunto de datos estándar que contiene un conjunto de funciones y tweets con sus opiniones previstas, es decir, conjunto **Positivo, Negativo y Neutro** indicados por 1, -1 y 0, respectivamente. Además, **neutral** los tweets pueden causar un desequilibrio en el proceso de entrenamiento o se puede degradar el rendimiento. Los tweets pueden causar un desequilibrio en el proceso de entrenamiento si un algoritmo simple que identifica el clasificador. Para eliminar los tweets neutrales, usamos un algoritmo simple que identifica y elimina los tweets neutrales (es decir, 0) y los filtra fuera del conjunto de datos, lo que resultó en que los refirió por su etiqueta (es decir, 0) y los filtró fuera del conjunto de datos, lo que resultó en la reducción de la versión reducida del conjunto de datos sin tweets neutrales. El conjunto de datos se reduce aún más por la versión del conjunto de datos sin tweets neutrales. El conjunto de datos se reduce aún más al eliminar eliminando los tweets neutrales ya que no juegan ningún papel en el proceso de predicción. Los tweets neutrales ya que no juegan ningún papel en el proceso de predicción, la eliminación de los tweets neutrales es necesario por dos razones: (i) los tweets neutrales no contienen ningún tweet neutral es necesario por dos razones: (i) los tweets neutrales no contienen ninguna opinión polaridad de opinión o sentimiento; por lo tanto, no juegan ningún papel significativo en la polaridad de opinión o sentimiento, por lo tanto, no juegan ningún papel significativo en la minería de opiniones, y (ii) la inclusión de un conjunto neutral de tweets genera un conjunto de datos más grande, lo que resulta en (ii) la inclusión de un conjunto neutral de tweets genera un conjunto de datos más grande, lo que resulta en la sobrecarga adicional e innecesaria para el clasificador durante el entrenamiento del modelo [28–30]. La sobrecarga innecesaria para el clasificador durante el entrenamiento del modelo [28–30]. El general toda la arquitectura se muestra en la Figura 1.



En segundo lugar, también hacemos uso de los datos bursátiles proporcionados en Google Finance, donde Los datos de stock históricos globales están disponibles. Se seleccionan los datos de precios de las acciones seleccionadas y descargado del proveedor de servicios en un archivo CSV. Los datos recogidos mantienen siete características nombradas: fecha, apertura, alto, bajo, cierre, volumen y cierre ajustado. Estas características indicar fecha de negociación, precio de apertura, precio más alto para negociar, precio más bajo para negociar, precio al cierre, las acciones negociadas y el precio de cierre de las acciones cuando los inversionistas reciben sus dividendos, respectivamente. Estos datos también se preprocesan agregando algunos valores calculados basados en características existentes (es decir, diferencia de precio de 5 días, diferencia de precio de 10 días, precios extrapolados

durante las vacaciones y el retorno del mercado (RM)), y eliminando algunas columnas que incluyen el precio de cierre ajustado, el volumen y el precio de apertura. Las razones para agregar esos valores calculados son las siguientes: la diferencia de precios de 5 y 10 días proporciona un breve comportamiento pasado de la acción en discusión. Los precios de cierre del fin de semana se han extrapolado para completar la cronología del conjunto de datos, lo que puede resultar en una precisión general mejorada del modelo. El rendimiento del mercado (RM) se calcula para proporcionar al inversor una idea probabilística del riesgo frente a la ganancia esperada.

Una vez completada la etapa de preprocesamiento para ambas fuentes de datos, el siguiente paso es el entrenamiento del modelo y la predicción de existencias. Se entrenó un modelo basado en ELM y RNN utilizando las funciones extraídas de los conjuntos de datos de Twitter y Google Finance. Ambos conjuntos de datos se distribuyen en dos subconjuntos; el primer 70% se reserva para formación y el segundo 30% para pruebas/validación. Se han proporcionado más detalles sobre los conjuntos de datos incorporados en la sección de resultados y discusiones.

### 3.2. Extracción de características

Una vez que los datos vuelven a pasar de la etapa de preprocesamiento, se envían a la etapa de extracción de características donde se realiza un procesamiento de datos adicional. Por esta razón, hemos propuesto un enfoque novedoso, a saber, el SSWN. En las secciones siguientes se proporciona una descripción detallada del enfoque propuesto.

#### 3.2.1. SWN

Varios recursos léxicos son muy utilizados en diversas investigaciones. En la Tabla 1 se proporciona un resumen de los activos altamente aplicados. El primer recurso léxico mencionado en la tabla llamado SenticNet es un recurso semántico que está disponible públicamente y se utiliza para realizar SA a nivel de concepto. No utiliza las técnicas estándar de minería de gráficos, sino que utiliza como concepto personalizado "flujos de energía" para representaciones de conocimiento de sentido común. Por otro lado, AFINN es uno de los léxicos más simples y populares que contiene cientos de synsets y palabras asociadas con un puntaje de polaridad que va de -5 a 5. De manera similar, SO-CAL también es un recurso léxico que contiene más de seis mil Synsets al asignar cada una palabra una puntuación de polaridad que va de -5 a 5. Otro recurso léxico popular es WordNet, que es una semejanza superficial del diccionario de sinónimos, que agrupa las palabras según sus significados. Es una gran base de datos léxica de libre acceso que agrupa sustantivos, verbos, adverbios y adjetivos en synsets, también conocidos como sinónimos cognitivos. Además, WordNet Affect amplía los dominios de WordNet al incluir aún más un subconjunto de sinónimos cognitivos (synsets) que son apropiados para representar los conceptos afectivos en una correlación con palabras afectivas. Hay varias aplicaciones de SWN en SA que se pueden emplear para predecir el mercado de valores, ya que la estructura de sus puntos clave es conveniente para realizar el modelado matemático. SWN es un recurso léxico para minería de opinión [23], en el que se nombra cada synset de WordNet, un triple de puntajes de polaridad, es decir, un puntaje de positividad, negatividad y objetividad. SWN se ha establecido rutinariamente al implicar una mezcla de clasificadores lingüísticos y estadísticos. Se ha empleado en varias misiones relacionadas con la opinión, es decir, para el análisis de sesgos y AS con hallazgos alentadores.

#### 3.2.2. SSWN

Para predecir las tendencias futuras del mercado de valores, presentamos SSWN, que se basa en SWN 3.0 y contiene un conjunto de palabras características específicamente útiles para identificar y puntuar tweets relacionados únicamente con el mercado de valores. El procedimiento de creación de SSWN comienza con dos conjuntos de semillas. El primer grupo comprende términos positivos mientras que el otro contiene términos negativos. Los grupos semilla se amplían combinando todos los synsets de SWN relacionados con las palabras semilla. Se elige un valor particular del radio para la expansión de la semilla. También se introduce otro conjunto, a saber, la palabra objetiva. En el segundo paso, las semillas calculadas se utilizan para clasificar los synsets SSWN en clases positivas y negativas. En el enfoque presentado, hemos empleado clasificadores junto con cuatro opciones de radio = 0, 2, 4, 6. Las salidas de todos los clasificadores se promedian para decidir el valor final del synset. Tab

describe una muestra de SSWN en la que cada tupla de SSWN especifica un synset compuesto por datos de diálogo, un identificador que vincula el synset con WordNet, partituras y una glosa que mantiene la denotación junto con el uso de los valores disponibles en cada synset. Todos las palabras/tokens en cada fila de los datos limpios se reemplazan con las puntuaciones calculadas, dando como resultado una matriz de características que está alineada/estandarizada con los requisitos de entrada de el clasificador ELM. La puntuación objetiva (OS) se puede calcular como:

$$OS = 1 / (PS + NS) \quad (1)$$

donde PS es la puntuación positiva mientras que NS es negativa. La puntuación de sentimiento (SS) se puede calculado utilizando la Ecuación (2):

$$SS = PS - NS \quad (2)$$

La fuerza del sentimiento (ST) se puede encontrar a través de la Ecuación (3), en la que  $r$  es la rango de la función.

$$ST = \frac{1}{r} \sum_{r=1}^r SS(r) \quad (3)$$

La Tabla 3 demuestra la relación entre un término  $t$  y una clase  $c$ .

**Tabla 1.** Resumen de recursos léxicos.

Recurso	Características totales		Rango de puntuación		TPV
	Sinsets	Palabras	mínimo	máx.	
SenticNet [31]	15,143	N / A	1	1	1
AFINN [32]	2477	N / A	5	5	5
SO-CAL [33]	6306	N / A	5	5	5
Léxico de subjetividad [34]	8221	N / A	N / A	N / A	1
Léxico de opinión [35]	6786	N / A	N / A	N / A	1
Investigador general [36]	11.789	N/A	N / A	N / A	1
SentiSense [37]	2190	5496	N/A 0	N/D 1	1
Micro-WNOp [38]	1105	1960			1
WordNet [39]	117.659	155,287	N / A	N / A	1
WordNet-Afect [40]	2874	4787	N/A 0	N/D 1	1
SentiWordNet [41]	117.659	155,287			1

**Tabla 2.** Una muestra del recurso léxico denominado SSWN.

TPV		Puntuación pos.	Puntuación negativa	Términos de Synset
a	2098	0	0.75	incapaz#1
	37006	0.625	0	obra maestra#2
r	5453	0.375	0	descaradamente#1
v	18813	0.375	0	despertar # 1 despertar_up # 1 despertar # 5 despertar#4 despertar#1 despertar#5

**Tabla 3.** Asociación entre  $t$  y  $c$ .

	Presencia de un Término $t$	Ausencia de un Término $t$
Presencia de una clase $c$	A	C
Ausencia de una clase $c$	B	D

Ganancia de información (IG)

IG, también denominada información mutua esperada, es una técnica basada en ML que es empleado para calcular el término bondad para una técnica determinada [23]. Funciona por computación los bits de información basados en la existencia o ausencia de una palabra en un archivo. Por ejemplo, la colección de grupos en un espacio de destino está representada por [30]  $i = 1, \dots, m$ . Entonces, el IG para un término  $t$  se calcula usando la fórmula en la Ecuación (4).



$$G(t) = \sum_{y=1}^m \Pr(c_i) \log \Pr(c_i) + \Pr(t) \sum_{y=1}^m \Pr(c_i/t) \log \Pr(c_i/t) + \Pr(t) \sum_{y=1}^m \Pr(c_i/t) \log \Pr(c_i/t) \quad (4)$$

Es un tipo simplificado de categorización binaria [21] ya que los enfoques de categorización de texto suelen utilizar un espacio de clasificación de matriz  $n$ , es decir, el rango de  $n$  puede ser de hasta decenas de miles. Además, la bondad de un valor se calcula universalmente de acuerdo con todas las clases en promedio. El valor IG se calcula para cada término distintivo de un corpus específico. Además, se define un umbral frente a la puntuación IG en función de qué términos se eliminan del corpus. La complejidad de cálculo para IG es  $O(Vn)$ , donde  $V$  es el tamaño del vocabulario y  $n$  es la categorización de  $n$  matrices. Al emplear la tabla de correlación, el valor de IG se calcula a través de la Ecuación (5). Cuanto mayor sea el valor de IG, mejor será la unión.

$$IG(t, c) = B \times N \times \log \frac{B}{(B + D) \times (A + B)} \quad (5)$$

Procedimiento de generación de la base de conocimiento de opiniones (SKB)

Para producir el SKB, el enfoque presentado sigue los siguientes pasos:

1. Tome todas las filas de SSWN una por una.
2. Calcular synset de cada fila seleccionada.
3. Calcule la orientación del sentimiento (SO) para cada synset.
4. Si se encuentra que el SO calculado es subjetivo, continúe con el paso 5; de lo contrario, elimine el conjunto de sincronización seleccionado y salte al paso 1 nuevamente.
5. Para cada synset subjetivo, ubique y calcule las porciones de su información de voz.
6. Encuentra palabras específicas del synset.
7. Calcule el vector de características combinando todos los términos individuales junto con fragmentos de voz diferenciado con un hash, es decir,  $\text{term}\#\text{POS}$ .
8. Guarde los puntos clave calculados en la lista de características nominadas.
9. Repita los pasos 1 a 8 para todas las filas.
10. La misma característica puede tener registros replicados con diferente polaridad y puntajes de opinión en la lista de puntos clave debido a su uso basado en la clasificación de sentido. Entonces, este paso se realiza para localizar las características distintivas.
11. Las ocurrencias positivas y negativas se calculan para todas las características detectadas en paso 10.
12. Según la puntuación de conteo calculada en el paso 11, se emplea IG para producir puntuaciones de opinión.
13. Finalmente, se asigna un identificador distintivo (ID) a cada característica.

Los SKB producidos a través de este procedimiento son independientes del dominio, ya que la fuerza de la opinión se calcula empleando un léxico de opinión genérico que no requiere la capacitación de un dominio específico. Los SKB presentados son capaces de lidiar con el problema de la ausencia de datos y la diversidad de datos. Además, estos SKB pueden ubicar fácilmente la orientación del sentimiento, el peso y el sentido de las palabras en función de su uso. Estos recursos de sentimiento se utilizan en la técnica introducida para mejorar SA específicamente para la predicción del mercado de valores y para SA en general. La Tabla 4 muestra una muestra del recurso léxico propuesto SS. Otro problema desafiante para una SA efectiva es la aparición constante de nuevas palabras u oraciones. Por lo tanto, existe la necesidad de un método de este tipo que pueda tratar con una base de datos que comprenda palabras frecuentes fuera del vocabulario (OOV). En el procesamiento del lenguaje natural, las palabras que están presentes en el conjunto de datos de prueba/real pero que no están disponibles en el conjunto de datos de entrenamiento se llaman palabras fuera del vocabulario (OOV). El problema principal es que el modelo asigna erróneamente una probabilidad cero a las palabras OOV, lo que da como resultado una probabilidad de una palabra igual a 0. Este problema común normalmente ocurre cuando el modelo se entrena en un conjunto de datos más grande. Existen múltiples soluciones para resolver este problema, incluida la tokenización, la técnica de suavizado y las representaciones semánticas [42,43]. Como los términos OOV pertenecen a un dominio específico, se necesita mucha información del dominio para especificar su fuerza. Para hacer frente a este problema, por lo general, se emplea el aprendizaje activo en el que se calcula una puntuación de polaridad a través de humanos. para evadir el

sesgo, hemos elegido sólo aquellas palabras OOV por las que han votado al menos diez personas. los La puntuación de sentimiento final se calcula tomando el valor promedio de las diez puntuaciones.

**Tabla 4.** Una muestra del recurso léxico propuesto SSWN.

TPV	identificador	Pos.	negativo	Términos de Synset
A	2772347	0.125	0.625	volatilidad#4
norte	2772317	0.375	0.125	acciones_blue_chip_#1
				stock_blue_chip_#1
V	2772344	0.125	0.625	venta_corta#2 venta_corta#2
A	2772348	0.375	0.125	venta_corta#2
norte	2772336	0	0 0	volumen #7
				abierto#1
V	2772314	0.125		promedio_abajo#1
	2772453	0.425	0.125	toro#1
A	2772494	0.125	0.625	alcista#
norte	2772458	0.625	0.125	oso # 1
				bajista#2
norte	2772455	0.125	0	Fugarse
norte	2772457	0.125	0.125	Gorra
				Piso
A	2772501	0.425	0.785	codicia#1
A	2772450	0.125	0.625	codicioso#2
A	2772660			Miedo
A	2772561	0	0.625	Ganar
				Pérdida
A	2772462	0.125	0.625	entrada_tarde#1
A	2772470	0.425	0.125	entrada_posterior#2
norte	2772468	0.625	0.125	entrada temprana
				estrella de la mañana
norte	2772465	0.125	0.625	Estrella de la tarde
norte	2772483	0.125	0.375	aumento_índice_tasa
norte	2772481	0.375	0.125	tasa_de_índice_descendente
				verde#1
A	2772463	0.625	0	fichas_verdes#2
				azul#1
A	2772610	0.625	0.125	fichas_azules#2
				rojo#1
A	2772623	0	0.425	fichas_rojas#2
A	2772612	0.425	0.125	riesgo bajo
A	2772503	0.125	0.625	alto riesgo
V	2772473	0.125	0.625	escape_comprador
V	2772474	0.425	0	vendedor_de escape

### 3.3. Fase de predicción

El vínculo entre las acciones y los sentimientos definitivamente no es lineal. Por lo tanto, después de descubrir una asociación de causalidad entre los estados de ánimo de los últimos 3 días y las acciones actuales precios, intentamos dos técnicas (ELM y RNN) para descubrir y examinar el definitivo asociación [44], y los mercados financieros a menudo siguen tendencias no lineales. Como se discutió anteriormente, la técnica propuesta incorpora dos conjuntos de datos, es decir, datos extraídos de Twitter y un conjunto de datos de última generación llamado Sentiment140. Las características extraídas de los datos de Twitter por utilizando SSWN se han incorporado para predecir las tendencias de existencias mediante el uso de los últimos tres Días de datos de acciones extraídos de Google Finance. Estas características extraídas luego se utilizan para predecir las tendencias bursátiles del día actual de un conjunto de marcas específicas.

#### 3.3.1. Máquina de aprendizaje extremo

Las características importantes de la clasificación de texto incluyen un gran número de entrenamiento muestras y alta dimensionalidad del texto. Se aumenta la dimensión alta de los resultados del texto. carga computacional para el ELM. Un método tradicional y eficaz para resolver este problema



es reducir la dimensionalidad del texto mediante el uso de algunas representaciones de texto que ayudan a aumentar la precisión de la aclaración. Los investigadores suelen utilizar el modelo de espacio vectorial (VSM) para la representación de texto en la clasificación de texto. En comparación con otros métodos de representación de texto, la representación de vectores de palabras ha demostrado tener una mejor capacidad de representación de texto. El vector de palabras se ocupa del problema de la dimensionalidad mediante el mapeo de cada término (una palabra distinta en el conjunto de datos textuales) con un vector real con una dimensión baja mediante el entrenamiento del corpus no etiquetado. Hemos considerado la apertura, el máximo y el mínimo como entrada para el ELM y el precio de cierre como salida del ELM. En el enfoque propuesto, el clasificador ELM [45] se introdujo inicialmente para una red neuronal feed-forward con una sola capa oculta sin necesidad de ajustarla. La salida con L nodos ocultos para el conjunto de entrenamiento se explica en la Ecuación (6):

$$F_L(x) = \sum_{y=1}^L \tilde{y}_i h_i(x) = h(x) \tilde{y} \quad (6)$$

Aquí,  $\tilde{y} = \{\tilde{y}_1, \dots, \tilde{y}_L\}^T$  presenta pesos de salida entre los nodos de la capa oculta y de salida mientras que  $h(x) = \{h_1(x), \dots, h_L(x)\}$  es el vector de salida. El método de decisión para el clasificador ELM se da como:

$$F_L(x) = \text{signo}(h(x)\tilde{y}) \quad (7)$$

Para obtener un desempeño robusto, el ELM tiene como objetivo lidiar con el error de entrenamiento más bajo y alcanzar la norma mínima de los pesos resultantes al reducir la función objetivo dada:

$$\text{Minimizar: } \|H\tilde{y} - T\|_2 \text{ y } \|\tilde{y}\| \quad (8)$$

Aquí, H muestra la matriz de salida de las capas ocultas.

$$H = \begin{matrix} & h(x_1) & & h_1(x_1) \dots h_L(x_1) & \\ \tilde{y} & \vdots & \tilde{y} & \vdots & \vdots & \vdots & \tilde{y} \\ & \vdots & & \vdots & & \vdots & \\ \tilde{y}h(x_N) & \tilde{y} & \tilde{y}h_1(x_N) & \vdots & h_L(x_N) & \tilde{y} \end{matrix} = \begin{matrix} h(x_1) & \dots & h_L(x_1) \\ \vdots & & \vdots \\ h(x_N) & \dots & h_L(x_N) \end{matrix} \quad (9)$$

Para reducir la norma de los pesos de salida, ELM dibuja un hiperplano óptimo para clasificar las muestras en diferentes clases maximizando el margen:  $2/\|\tilde{y}\|$  empleando el método de mínimos cuadrados nominales como:

$$\tilde{y} = H^\dagger T \quad (10)$$

Aquí,  $H^\dagger$  presenta la inversa generalizada de Moore-Penrose de la matriz que se calcula utilizando los enfoques de ortogonalización, proyección ortogonal y descomposición en valores singulares.

La salida deseada de ELM es:

$$\text{Prueba } T = \tilde{y} H^\dagger \quad (11)$$

### 3.3.2. Red neuronal recurrente

La red neuronal recurrente, también conocida como RNN, es adecuada en los problemas en los que debemos tratar con una secuencia de datos. Muchos investigadores recomiendan utilizar RNN para el análisis de series temporales [8–10]. En este tipo de trabajo, el modelo aprende de su observación actual, también conocida como memoria a corto plazo de la red, asemejándose al lóbulo frontal del cerebro. La razón para usar RNN cuando vamos a trabajar con datos secuenciales es que el modelo usa su memoria a corto plazo para predecir los próximos datos con mayor precisión. En lugar de utilizar una fecha límite fija para eliminar los datos anteriores, los pesos asignados a los datos anteriores determinan el tiempo durante el cual estos datos se mantendrán en la memoria. Por lo tanto, RNN es más adecuado en el caso de problemas, como el etiquetado de secuencias, el análisis de sentimientos y el etiquetado de voz, etc. [46,47].

El análisis de series de tiempo es generalmente un problema importante que puede resolverse utilizando RNN. En este problema, necesitamos trabajar con datos que están en orden secuencial. Tal

Los trabajos implican aprender de las observaciones más recientes, alternativamente llamadas memoria a corto plazo. Este trabajo se centra principalmente en la clasificación de textos. Entonces, RNN en esta investigación se usa para clasificar los datos de Twitter. Proponemos un modelo para predecir el precio de cierre del mercado de valores.

Los datos de Twitter no están en un formato uniforme, lo que significa que la cantidad de palabras en un tweet puede variar de 3 a 5 palabras a 17 a 20 palabras, por ejemplo. Sin embargo, nuestra red neuronal no acepta entradas en este formulario. Necesitamos convertir estos datos en un formato uniforme. La solución más adecuada a este problema puede ser incrustar y rellenar las filas de datos. El proceso de incrustación implica representar las palabras con vectores utilizando el procedimiento mencionado en la discusión relacionada con ELM. Se determina la posición de un término o palabra en un espacio vectorial y se representa en el vector de características. Los datos incrustados deben tener una longitud uniforme, por lo que rellenamos los datos con ceros.

RNN [48] emplea enlaces entre nodos para construir un gráfico dirigido en un marco de tiempo. Esto le permite mostrar un comportamiento vibrante secuencial. RNN utiliza su memoria para manipular las secuencias de entradas de longitud variable, lo que lo hace apropiado para la predicción de existencias. Cada unidad de procesamiento en un RNN consta de una activación de valor real arbitrario basada en el tiempo y un peso adaptable que se genera empleando el mismo conjunto de pesos en un bucle sobre una estructura similar a un gráfico. La ecuación (12) se utiliza para especificar los valores de las unidades ocultas.

$$H^t = f(h^{t-1}, x; \tilde{y}) \quad (12)$$

En RNN, el tamaño de la entrada sigue siendo el mismo para cada modelo aprendido, ya que se indica en forma de cambio de un estado a otro. Además, la estructura emplea la misma función de transición que tiene los mismos parámetros para cada paso de tiempo. RNN almacena la salida de las capas anteriores para hacer predicciones, lo que le permite trabajar con datos secuenciales. En este trabajo, hemos probado el RNN para la predicción del comportamiento del mercado de valores.

#### 4. Resultados

**experimentales** Esta sección describe la demografía de los conjuntos de datos utilizados, una descripción general de las métricas de evaluación y una discusión exhaustiva de los resultados obtenidos junto con una comparación con las técnicas más modernas.

##### 4.1. Configuración

experimental El banco de pruebas consta de una estación de trabajo equipada con una CPU x64 Intel Core i7-6700 a 3,40 GHz con 16 GB de RAM DDR4 y 4 GB de tarjeta gráfica NVIDIA GeForce. La capacidad de almacenamiento es de 1 TB HDD y 256 GB de SSD. El sistema operativo de 64 bits es Microsoft Windows 10 Professional, que está instalado en el SSD. Los conjuntos de datos y los entornos de trabajo se almacenan en SSD para evitar el retraso mecánico causado por el HDD y acelerar el proceso de prueba y entrenamiento del modelo.

La versión 3.7.15 de Python junto con las bibliotecas necesarias como NLTK, Stanford NER Tagger y BeautifulSoup, Numpy, Scikit-learn, etc. están instaladas en el entorno de Anaconda. Hemos utilizado la función de activación Relu y la tasa de aprendizaje es 0.001 para nuestro modelo de entrenamiento. Para la evaluación del desempeño, hemos empleado diferentes métricas, es decir, exactitud, precisión, recuperación y medida F.

##### 4.2. conjuntos de datos

Como se describió en las secciones anteriores, incorporamos dos conjuntos de datos, es decir, Sentiment140, que es un conjunto de datos de última generación ampliamente utilizado para tareas relacionadas con SA, y el otro conjunto de datos se recopila directamente de la plataforma de Twitter mediante una API de Twitter, es decir, Tiburón.

##### 4.2.1. El conjunto de datos Sentiment140

El conjunto de datos Sentiment140 contiene un total de 1,6 millones de tweets extraídos mediante una API de Twitter [49]. Todos los tweets han sido anotados como negativo = 0, neutral = 2 y positivo = 4

y se utilizan para descubrir sus sentimientos. El conjunto de datos contiene seis columnas descritas en Tabla 5. Puede encontrar una descripción detallada del conjunto de datos Sentiment140 aquí [27].

Tabla 5. Descripción del conjunto de datos de Sentiment140.

Columna	Descripción
indicador de polaridad	Polaridad del tweet (negativo = 0, neutral = 2 y positivo = 4)
fecha de publicación	ID de tweet (p. ej., 2088)
ID de tweet	La hora en que se publicó el tuit. (p. ej., domingo 17 de mayo a las 22:57:44 UTC de 2008)
objetivo	La consulta. El indicador NO_QUERY significa que no hay consulta.
usuario	ID del usuario que publicó este tweet. (por ejemplo, el compañero de Twitter)
texto	Cuerpo del tuit. (por ejemplo, las acciones de #AAPL se han mantenido estables durante una semana).

Filtramos los tuits que mencionaban una de las marcas especificadas en el tuit. cuerpo. Esta filtración dio como resultado un nuevo subconjunto del conjunto de datos Sentiment140 que consiste en total 56 mil tuits. El conjunto de tweets neutrales ha sido ignorado y sustraído del conjunto de datos como los tweets neutrales no juegan ningún papel significativo en la predicción de acciones.

4.2.2. Datos directos de Twitter

Este conjunto de datos se recopila mediante un código personalizado que utiliza una API de Twitter. Los tweets que mencionan las marcas se muestran en la Tabla 6 y se publicaron entre el 1 y el 21 de marzo de 2021. se han extraído/descargado usando una biblioteca de Python llamada Tweepy. Después de realizar los pasos de preprocesamiento y limpieza mencionados en las secciones anteriores, los datos recopilados los datos finalmente están en condiciones de ser procesados y utilizados para predecir el mercado de valores valor de marcas específicas. La Tabla 6 demuestra la demografía de los datos recopilados directamente de Twitter.

Tabla 6. Detalle de datos extraídos directamente de Twitter.

Valores Mercado	Símbolo	Número de tuits (Antes del preprocesamiento)				Número de tuits (Después del preprocesamiento)		
		Positivo	Negativo	Neutral	Total	Positivo	Negativo	Total
Manzana	APLICAR	14.400	10.500	5940	30.840	12.384	9135	21,519
tesla	TSLA	40.050	31.290	80,916	152.256	33.642	27,222	60,864
microsoft	MSFT	26.100	20.730	21,963	68.793	22.446	17,828	40,274
Walmart WMT		13.200	11.400	17,222	41.822	11.220	9918	21,138
Paypal	PPPL	10.500	5670	30,182	46.352	9030	4820	13,850
nvidia	NVDA	6150	4800	31,140	42.090	5351	3984	9335
Intel	INTC	3900	3360	31,750	39.010	3315	2789	6104
Facebook	—	15,150	10,500	28,912	54.562	13,181	8820	22,001
Gorjeo	TWTR	10,650	10,350	28,170	49.170	9053	9005	18,057
Amazon AMZN		6630	6300	23,247	36.177	5636	5418	11,054
Gran total		146,730	114,900	299,442	561,072	125,256	98,938	224,194

Similar a los pasos realizados en el conjunto de datos setntiment140, descargamos el tweets mencionando una de las marcas en estudio utilizando el mencionado código personalizado. Esto resultó en un nuevo conjunto de datos que consta de aproximadamente 506 K tweets. Además, también calculamos el término frecuencia. La figura 2 muestra una nube de palabras que muestra palabras de uso frecuente en el conjunto de datos. El conjunto de tweets neutrales fue ignorado y excluido del conjunto de datos como tweets neutrales. no juegan ningún papel significativo en la predicción de existencias. Después de realizar el preprocesamiento y la resta de los tweets neutrales, el conjunto de datos se redujo aún más a un total de 224,2 K tweets pertenecientes a clases positivas y negativas solamente.

12 de 20



Electronics 2022, 11, x PARA REVISIÓN

metodo en términos de precisión, recuperación y medida f. La figura mencionada representa que el



Para evaluar aún más nuestro método, hemos trazado las precisiones de todas las marcas en un diagrama de caja diagrama de caja que se puede ver en la Figura 4. La Figura 4a describe los resultados de la clasificación ELM. La precisión de predicción de todas las marcas, es decir, APPL, TSLA, MSFT, WMT, PYPL, NVDA, INTC, FB, la precisión de predicción de todas las marcas, es decir, APPL, TSLA, MSFT, WMT, PYPL, NVDA, INTC, FB, FB, TWTR y AMZN es 90,3 %, 85,01 %, 88,21 %, 85,18 %, 84,716 %, 87,35 %, 80,733 %, 79,25%, 91,05% y 88,78% respectivamente. Por lo tanto, la precisión promedio de nuestra técnica propuesta es del 86,06 %, lo cual es impresionante y se puede usar para predecir con precisión el comportamiento del mercado de valores. La Figura 4b muestra los resultados del clasificador RNN para todas las marcas. Aquí, nuestro método

Machine Translated by Google

la precisión, la recuperación y la medida f son 0,8603, 0,811 y 0,8537, respectivamente. El gráfico de columnas muestra los resultados de marca de nuestro método, en el que las barras azul, roja y gris muestran precisión, recuperación y medida f, respectivamente. Entonces, podemos decir que nuestro método puede predecir con precisión el comportamiento del mercado de valores de todas las marcas.

Para evaluar aún más nuestro método, hemos trazado las precisiones de todas las marcas en un diagrama de barras de 13 de 20 que se puede ver en la Figura 4. La Figura 4a describe los resultados de la clasificación ELM. La precisión de predicción de todas las marcas, es decir, APPL, TSLA, MSFT, WMT, PYPL, NVDA, INTC, FB, TWTR y AMZN es 90,3 %, 85,01 %, 88,21 %, 86,733 %, 89,736 %, 87,35 %, 80,733 %, TWTR y AMZN es 90,3 %, 85,01 %, 88,21 %, 85,18 %, 84,716 %, 87,35 %, 79,857 %, 91,05 % y 88,78 %, respectivamente. Entonces, la precisión promedio de nuestra tecnología propuesta es 91.05% que es 86,06% lo cual es impresionante para que se pueda predecir con precisión el mercado de valores 86,06%, lo cual es impresionante. La Figura 4b muestra los resultados de la clasificación RNN basados en las marcas. Aquí, nuestro método logró la precisión que se puede ver en el gráfico que es 81,4 %, lo que demuestra la precisión lograda por ELM la precisión promedio del 81,4 %, clasificados. Según los resultados, podemos decir que nuestro método es muy preciso con mayor precisión. Según los resultados, podemos decir que nuestro método es muy preciso con mayor precisión. Según los resultados, podemos decir que nuestro método es muy preciso con mayor precisión.

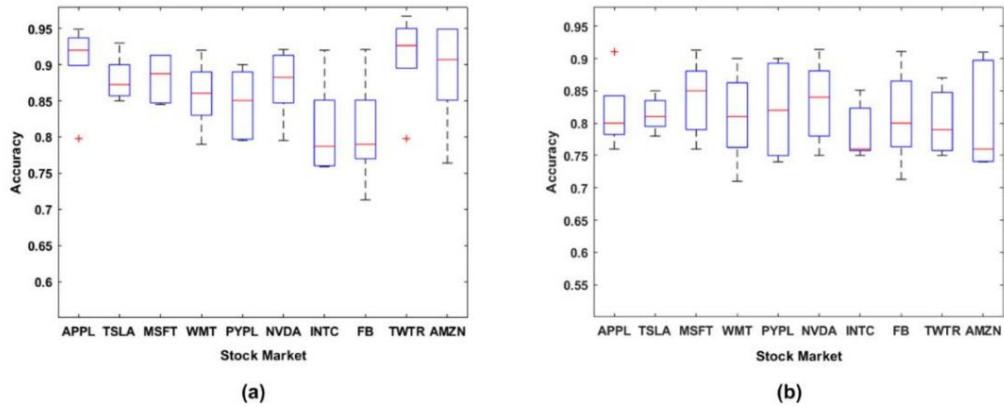


Figura 4. Precisión de todas las marcas usando (a) ELM y (b) RNN.

4.2.4. Evaluación del Desempeño de los Clasificadores

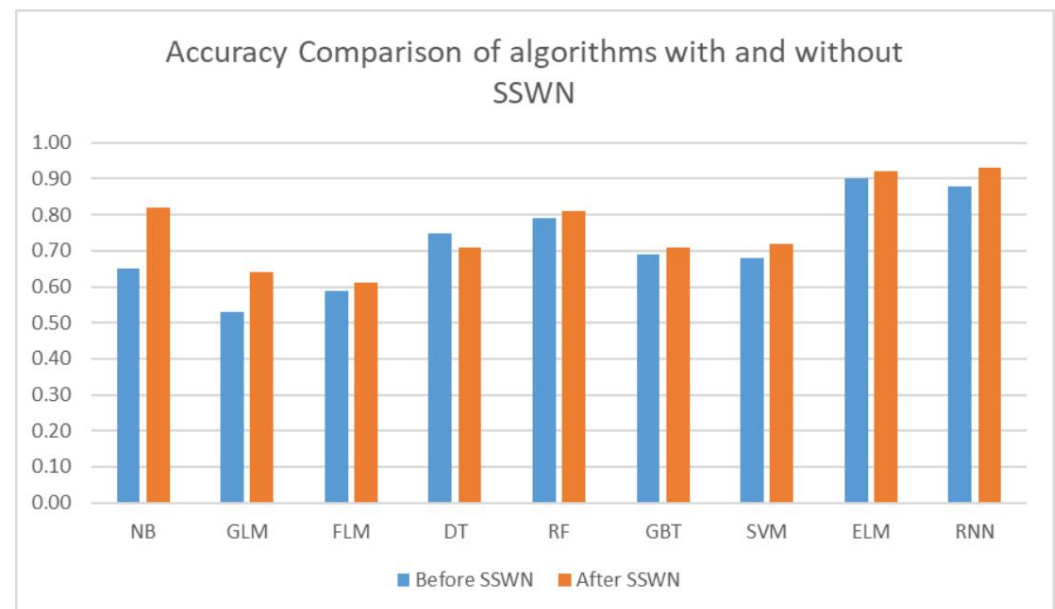
Seleccionamos nueve algoritmos de ML y comparamos su rendimiento con respecto al exactitud de la predicción. Entrenamos estos algoritmos y luego los probamos en ambos conjuntos de datos. para predecir las tendencias futuras del mercado de valores. Antes de aplicar estos algoritmos de ML, dividimos el conjuntos de datos finales en dos porciones, es decir, el 70% de las muestras como datos de entrenamiento y el resto 30% como datos de prueba. El entrenamiento y la prueba de los algoritmos se realizan utilizando un Biblioteca Python para ML llamada Scikit-learn [31]. La Tabla 7 proporciona una lista de algoritmos de ML utilizados en esta experimentación junto con sus parámetros óptimos.

Tabla 7. Algoritmos de ML seleccionados con sus valores de parámetros óptimos.

No.	Algoritmo	Abreviatura	Conjunto de parámetros óptimo
1	bayesiana ingenua	núcleo lineal	N / A
2	Modelo lineal generalizado	GLM	núcleo = rbf, C: 0.6
3	Gran margen rápido	FLM	Solucionador = L2, C = 0.5, épsilon = 0.25, class_weights = 1, use_bias = false
4	Árbol de decisión	DT	criterio: 'entropía', divisor = 'mejor', profundidad máxima = 8, min_samples_split = 2, min_samples_leaf = 1, min_weight_fraction_leaf = 0.5, preordenar = 'verdadero'
5	Bosque aleatorio	RF	n_trabajos = 1, min_muestras_hoja: 2, n_estimadores: 25, estado_aleatorio: 125, criterio: gini, min_samples_split: 4
6	Árboles potenciados por gradiente	GBT	min_samples_split = 2500, min_samples_leaf = 50, profundidad_máxima = 8, características_máximas = 'sqrt', submuestra = 0.8, estado_aleatorio = 8
7	Máquinas de vectores soporte	MVS	núcleo = 'rbf', C = 10, gamma = automático,
8	Aprendizaje extremo Máquina	OLMO	capas_ocultas = 20, pesos = [1, 1], activación_función = 'sigmoide'
9	Red neuronal recurrente	RNN	init = 'glorot_uniform', inner_init = 'ortogonal', activación = 'tanh', w_regularizador = ninguno, u_regularizador = ninguno, abandono_w = 0.1, abandono_u = 0.02

## 4.2.5. Desempeño de Algoritmos antes y después de SSWN

Evaluamos el desempeño de las técnicas elegidas con SSWN y sin SSWN usándolo, es decir, empleando el SWN estándar. La figura 5 muestra un aumento general en la precisión de todos los algoritmos después de emplear SSWN.



**Figura 5.** Comparación de precisión de algoritmos antes y después de SSWN.

## 4.3. Rendimiento de algoritmos en ambos conjuntos de datos

Como en otras comparaciones, también comparamos el desempeño de algoritmos seleccionados de datos con otras estándar Sentiment140, como se muestra en la Tabla 8.

**Tabla 8.** Rendimiento de los algoritmos sobre el conjunto de datos de sentiment140.

Modelo	Precisión	Precisión 0.65	Recordar	Medida F
GLM	0.82	0.53	0.75	0.66
GLM	0.64	0.59	0.71	0.60
FLM	0.61	0.59	0.95	0.60
FLM	0.61	0.75	0.66	0.50
DT	0.67	0.79	0.66	0.53
DT	0.71	0.69	0.20	0.53
RF	0.69	0.69	0.30	0.51
RF	0.71	0.69	0.30	0.51
GBT	0.61	0.80	0.20	0.27
GBT	0.71	0.68	0.77	0.27
OLMO	0.81			0.82
MVS	0.61			0.29
RNN	0.86			0.85
OLMO	0.81			0.82

Es evidente a partir de la Tabla 8 que el clasificador ELM supera a otros algoritmos en términos de exactitud y precisión. Sin embargo, la recuperación y la medida F de ELM no pueden permanecer en la parte superior. Es evidente en la Tabla 8 que el clasificador ELM supera a otros algoritmos en términos de exactitud y precisión. Sin embargo, la recuperación y la medida F de ELM no pueden seguir siendo términos de exactitud y precisión, mientras que el resto se basa en términos de medida F.

En la parte superior, también se puede observar en la tabla que RNN muestra el segundo mejor desempeño. La Tabla 8 demuestra el desempeño de estos algoritmos en el conjunto de datos de Twitter.

La Tabla 8 demuestra la mejora significativa de rendimiento obtenida al utilizar SSWN. La mayoría de los clasificadores mejoraron con SSWN, y el rendimiento de todos los algoritmos excepto para NB DT aumentó, es decir, estos dos algoritmos no mostraron ninguna mejora significativa en su desempeño. Mientras que RNN aquí, nuevamente, muestra el segundo mejor desempeño en términos de exactitud y precisión y el mejor rendimiento en términos de medida F.

**Tabla 9.** Rendimiento de los algoritmos sobre el conjunto de datos extraído de Twitter.

Modelo	Exactitud	Precisión	Recuperación	Medida F
GLM	0.67	0.56	0.75	0.63
FLM	0.64	0.62	0.79	0.53
DT	0.69	0.71	0.69	0.57
RF	0.72	0.83	0.32	0.11
GBT	0.75	0.72	0.21	0.28
MVS	0.64	0.71	0.21	0.31



Electrónica 2022, 11, x PARA REVISIÓN POR PARES

Tabla 9. Rendimiento de los algoritmos sobre el conjunto de datos extraído de Twitter.

16 de 21

Modelo	Precisión	Precisión	Recuerdo	Medida F
Naïve Bayes	0.81	0.67	0.72	0.64
RNN	0.89	0.90	0.85	0.86
FLM	0.67	0.56	0.75	0.63
DT	0.69	0.71	0.69	0.57
La Tabla 9 demuestra la mejora significativa del rendimiento obtenida utilizando un ma	0.72	0.83	0.32	0.11
mayoría de los clasificadores. Al incorporar SSWN, el rendimiento de todos los algoritmos excepto NB DT aumentó, es decir,	0.73	0.72	0.29	0.28
estos dos algoritmos no mostraron ninguna mejora significativa en su rendimiento. Mientras que RNN aquí, nuevamente,	0.84	0.84	0.84	0.84
muestra el segundo mejor desempeño en términos de exactitud y precisión y el mejor desempeño en términos de medida F.	0.85	0.84	0.84	0.86
RNN	0.89	0.90	0.85	0.89

4.4.4. Complejidad del tiempo

El rendimiento de todos los algoritmos seleccionados también se comparó con respecto al tiempo que tardaron los modelos en entrenar y asignar las puntuaciones de sentimiento. La Figura 6 muestra una comparación detallada de la complejidad del tiempo de los algoritmos en segundos. de tiempo en segundos para el entrenamiento y la puntuación. NB y DT fueron los más rápidos en el entrenamiento, pero FLM fue el más rápido FLM fue el más rápido mientras que el sentimiento. En general, ELM y RNN fueron los más rápidos en la puntuación, pero RNN fue el más rápido en la puntuación.

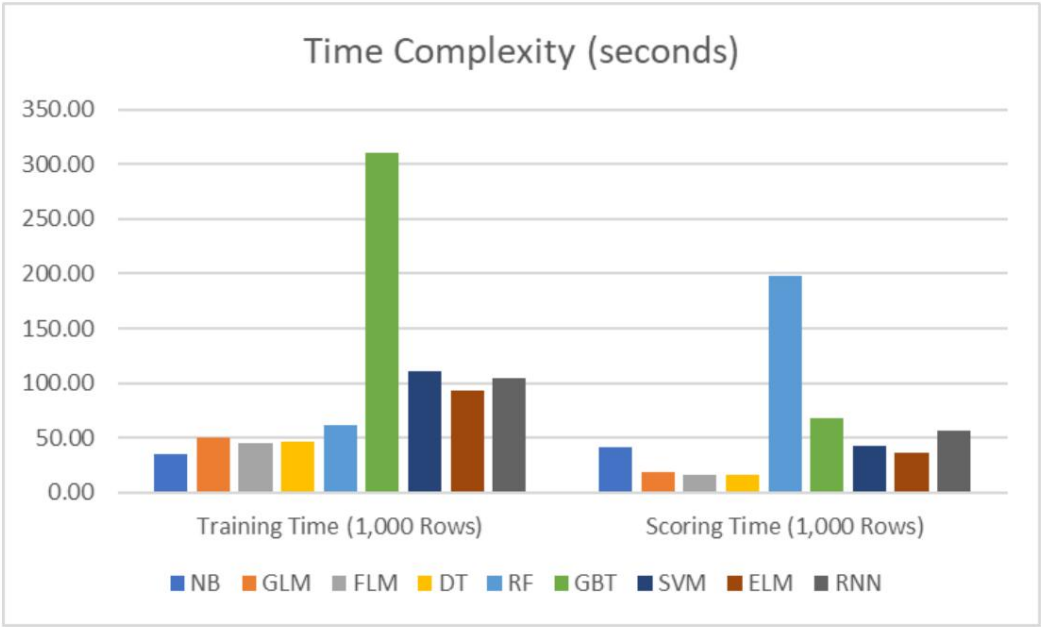


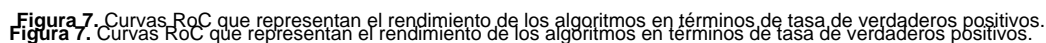
Figura 6. Comparación de rendimiento de algoritmos en términos de tiempo empleado.

4.5. Desempeño de Clasificación de los Algoritmos Seleccionados

La figura 7 muestra una comparación de rendimiento de los algoritmos seleccionados en el formato de gráfico de curva ROC, que muestra que ELM supera a otros en términos de clasificación correcta de las muestras de entrada.

4.6. Comparación con técnicas de vanguardia

Varios investigadores han presentado trabajos basados en ML para predecir las tendencias futuras de el mercado de valores. Por lo tanto, en esta sección, para evaluar la robustez de predicción de nuestro enfoque, realizamos un análisis comparativo de nuestro marco con los últimos enfoques basados en ML. Este análisis se evalúa en términos de técnica empleada, datos, así como como exactitud y precisión obtenidas.



4.6. **Comparación de resultados comparativos en la Tabla 10.** Zhou et al. [33] presentó un enfoque para la predicción del mercado de valores mediante el uso de las emociones en línea utilizadas por las personas para evaluar sus comportamientos. El método [33] empleó la SVM para realizar la clasificación y alcanzó una precisión del 64.15%. Nguyen et al. [34] introdujo un marco para el mercado de valores y realizó un análisis comparativo de nuestro marco con la última predicción basada en ML utilizando los sentimientos relacionados con un tema específico de la empresa y empleamos enfoques. Este análisis se evaluó en términos de técnica empleada, datos, así como el clasificador SVM para la predicción. Este método [34] mostró una precisión promedio de 54.41% exactitud y precisión obtenidas.

El trabajo en [35] presentó una técnica de minería de datos para la predicción del mercado de valores con los resultados comparativos se informan en la Tabla 10. Zhou et al. [33] presentó una precisión obtenida por de 66.48%, Khan et al. [25] introdujo un marco mediante el uso de la técnica para la predicción del mercado de valores mediante el uso de las emociones en línea utilizadas por las personas en sitios junto con eventos políticos para la predicción del mercado de valores y alcanzó una precisión de 75.38%. De manera similar, la técnica en [36] usó el mismo concepto con el clasificador ANN y alcanzó una precisión del 64.15%. Nguyen et al. [34] introdujo un marco para stock mar alcanzó una precisión promedio de 77.12%. Khan et al. [37] presentó otro enfoque utilizando el prediccion utilizando los sentimientos relacionados con un tema específico de la empresa y em datos de noticias financieras y obtuvo una precisión del 80.6%. La técnica en [38] empleada utilizó el clasificador SVM para la predicción. Este método [34] mostró una precisión promedio de los sentimientos de las personas de los sitios sociales junto con el clasificador ingenio Bayes y SVM y alcanzó una precisión del 54.41%. El trabajo en [35] presentó una técnica de minería de datos para la predicción del mercado de valores y mostró la mejor precisión promedio de 80.6%. De la Tabla 10, se puede observar que la con una precisión obtenida del 66.48%. Khan et al. [25] introdujo un marco mediante el uso del enfoque presentado mostró un valor de precisión del 85.7%, que es más alto que todos los sitios sociales de com junto con eventos políticos para la predicción del mercado de valores y logró métodos acumulativos. Además, los métodos comparativos alcanzaron un valor de precisión promedio de 75.38%. De manera similar, la técnica en [36] usó el mismo concepto con las clases ANN de 71.24%, que es 85.7% en nuestro caso, por lo que nuestro método obtuvo una ganancia de rendimiento de 14.46%.

El trabajo en [35] presentó una técnica de minería de datos para la predicción del mercado de valores y mostró la mejor precisión promedio de 77.12%. Khan et al. [37] presentó otro enfoque utilizando datos de noticias financieras y obtuvo una precisión del 80.6%. La técnica en [38]

Técnica	Precisión (%)	Precisión
Propuesta (ELM)	86,06	0,88
Khan et al. [27]	75,38	0,67
Zhou et al. [50]	64,15	0,65

Los valores informados demuestran que el enfoque introducido supera las técnicas comparativas [25,33–39], al introducir el léxico de sentimiento SSWN que ayuda a seleccionar un conjunto más representativo de características relacionadas con el mercado de valores. Además, los métodos en [25,33–39] son computacionalmente más costosos y pueden resultar en un problema de sobreajuste. Sin embargo, en nuestro caso, la solidez del clasificador ELM para manejar los datos de entrenamiento sobreajustados ayuda a lograr una precisión eficiente con menos tiempo de procesamiento. Por lo tanto, el método propuesto puede describirse como más eficaz y eficiente para la predicción del mercado de valores.

#### 4.7. Discusión

La predicción de los precios del mercado de valores es un tema interesante de investigación y es una tarea desafiante debido a la volatilidad, diversidad y comportamiento dinámico del mercado de valores. Investigaciones recientes han revelado que los sentimientos y las noticias pueden influir en el movimiento del mercado de valores y actuar como predictores potenciales de los resultados de las compensaciones. Por lo tanto, las plataformas de redes sociales pueden considerarse una fuente importante de información para extraer fragmentos importantes de información de las publicaciones en redes sociales ya publicadas por los usuarios. En este sentido, Twitter se convierte en una fuente de información más adecuada debido al carácter conciso de los tuits que allí se publican. Sin embargo, esta concisión también hace que el trabajo sea más desafiante debido al uso de palabras abreviadas, duplicación y diferentes tipos de ruido que residen en los tweets. Combinados con el poder del aprendizaje automático, los tweets pueden ser importantes para la predicción de los precios del mercado de valores. En este trabajo, presentamos un enfoque novedoso para la predicción de los precios de las acciones mediante el uso de SA. Para este propósito, implementamos dos clasificadores distintos, es decir, RNN y ELM, junto con otros populares que se basan en el léxico de sentimientos propuesto denominado SSWN y dos conjuntos de datos, es decir, datos adquiridos directamente de Twitter y un conjunto de datos estándar denominado Sentiment140. Realizamos la experimentación con diez datos de acciones del mercado estadounidense obtenidos de Google Finance. En primer lugar, comparamos y evaluamos el rendimiento de nueve algoritmos de aprendizaje automático diferentes en dichos datos bursátiles, donde el rendimiento de ELM se mantuvo en la cima. En segundo lugar, comparamos nuestro trabajo con el más avanzado mientras logramos una precisión general superior debido al uso de un léxico de sentimiento dedicado especialmente propuesto para la predicción del mercado de valores. El alcance y el funcionamiento de la técnica propuesta se pueden mejorar aún más considerando otros enfoques basados en DL.

#### 5. Conclusiones

Las personas usan las redes sociales para compartir sus ideas y opiniones personales con respecto a una marca, entidad, persona o asunto. Twitter es una plataforma de redes sociales moderna y reconocida a nivel mundial para compartir ideas y opiniones de una manera muy concisa. Usando el poder de SA ML, las publicaciones en las redes sociales, como los tweets, pueden desempeñar un papel importante en la predicción del comportamiento del mercado de valores. Este trabajo presenta un enfoque novedoso para la predicción del mercado de valores utilizando SA. El modelo se basa en el léxico de sentimiento SSWN propuesto junto con los clasificadores RNN y ELM. Hemos utilizado datos de Twitter y el conjunto de datos Sentiment140 para la evaluación del rendimiento de los modelos de ML considerando diez marcas diferentes para la predicción del mercado de valores. Logramos una precisión promedio de 81,40 % para RNN y 86,06 % para el clasificador ELM. Comparamos nuestro enfoque con varios modelos de ML, así como con métodos de última generación y logramos resultados notables. En el futuro, planeamos mejorar la capacidad y la cobertura de este enfoque agregando otras plataformas populares de redes sociales, por ejemplo, Facebook y Google News. Además, podemos evaluar el enfoque propuesto sobre otros conjuntos de datos desafiantes mientras consideramos también más acciones.

**Contribuciones de los autores:** conceptualización, análisis formal, análisis de datos, interpretación de datos, búsqueda bibliográfica, adquisición de fondos, administración de proyectos, SA; conceptualización, software, recursos, metodología, redacción—borrador original, TN; validación, visualización, redacción—borrador original, AM; supervisión, validación, redacción—revisión y edición, IA; búsqueda bibliográfica, investigación, validación, AA; conceptualización, supervisión, redacción—revisión y edición, WA. Todos los autores han leído y están de acuerdo con la versión publicada del manuscrito.

**Financiamiento:** Esta investigación fue financiada por el Decanato de Investigación Científica de la Universidad de Qassim.

**Agradecimientos:** Los investigadores desean agradecer al Decanato de Investigación Científica de la Universidad de Qassim por financiar la publicación de este proyecto.

**Conflictos de interés:** Los autores declaran que no existe ningún conflicto de interés entre los autores.

## Referencias

- Li, X.; Wu, P.; Wang, W. Incorporación de precios de acciones y sentimientos de noticias para la predicción del mercado de valores: un caso de Hong Kong. *información Proceso. Administrar* **2020**, *57*, 102212. [\[Referencia cruzada\]](#)
- Roy, SS; Mittal, D.; Basu, A.; Abraham, A. Pronóstico del mercado de valores utilizando el modelo de regresión lineal LASSO. en *afroeuropo Conferencia para el Avance Industrial*; Springer: Cham, Suiza, 2015; págs. 371–381.
- Ruán, Y.; Durrezi, A.; Alfantoukh, L. Uso de la red de confianza de Twitter para el análisis del mercado de valores. *Sistema basado en el conocimiento*. **2018**, *145*, 207–218. [\[Referencia cruzada\]](#)
- Oliveira, N.; Cortés, P.; Areal, N. El impacto de los datos de microblogging para la predicción del mercado de valores: uso de Twitter para predecir rendimientos, volatilidad, volumen de operaciones e índices de sentimiento de encuestas. *Sistema experto aplicación* **2017**, *73*, 125–144. [\[Referencia cruzada\]](#)
- Bosé, A.; Hsu, C.-H.; Roy, SS; Lee, KC; Mohammadi-Ivatloo, B.; Abimannan, S. Pronóstico del precio de las acciones mediante un modelo híbrido de splines de regresión adaptativa multivariante en cascada y redes neuronales profundas. *computar eléctrico Ing.* **2021**, *95*, 107405. [\[Referencia cruzada\]](#)
- Gite, S.; Khatavkar, H.; Kotecha, K.; Srivastava, S.; Maheshwari, P.; Pandey, N. Predicción explicable de precios de acciones a partir de artículos de noticias financieras utilizando análisis de sentimiento. *Cómputo entre pares. ciencia* **2021**, *7*, e340. [\[Referencia cruzada\]](#)
- Patel, J.; Sha, S.; Thakkar, P.; Kotecha, K. Predicción del índice bursátil mediante la fusión de técnicas de aprendizaje automático. *Sistema experto aplicación* **2015**, *42*, 2162–2172. [\[Referencia cruzada\]](#)
- Patel, J.; Sha, S.; Thakkar, P.; Kotecha, K. Predicción del movimiento del índice bursátil y del precio de las acciones mediante la preparación de datos deterministas de tendencia y técnicas de aprendizaje automático. *Sistema experto aplicación* **2015**, *42*, 259–268. [\[Referencia cruzada\]](#)
- Derakhshan, A.; Beigy, H. Análisis de sentimiento en las redes sociales bursátiles para la predicción del movimiento del precio de las acciones. *Ing. aplicación Artefacto Intel.* **2019**, *85*, 569–578. [\[Referencia cruzada\]](#)
- Pai, P.-F.; Liu, C.-H. Predicción de ventas de vehículos mediante análisis de sentimiento de datos de Twitter y valores bursátiles. *Acceso IEEE* **2018**, *6*, 57655–57662. [\[Referencia cruzada\]](#)
- Panday, H.; Vijayarajan, V.; Mahendran, A.; Krishnamoorthy, A.; Prasath, V. Predicción bursátil mediante análisis de sentimiento y memoria a largo plazo. *EUR. J. Mol. clin. Medicina*. **2020**, *7*, 5060–5069.
- Shi, Y.; Zheng, Y.; Guo, K.; Ren, X. Predicción del movimiento de acciones con análisis de sentimiento basado en redes de aprendizaje profundo. *De acuerdo. computar Pr. Exp.* **2020**, *33*, e6076. [\[Referencia cruzada\]](#)
- Kelotra, A.; Pandey, P. Predicción del mercado de valores usando un modelo optimizado de conversión profunda. *Grandes datos* **2020**, *8*, 5–24. [\[Referencia cruzada\]](#) [\[PubMed\]](#)
- Bouktif, S.; Fiaz, A.; Awad, M. Predicción del mercado de valores basada en características textuales aumentadas. *Acceso IEEE* **2020**, *8*, 40269–40282. [\[Referencia cruzada\]](#)
- Ren, R.; Wu, DD; Liu, T. Pronóstico de la dirección del movimiento del mercado de valores mediante el análisis de sentimientos y la máquina de vectores de soporte. *Sistema IEEE J.* **2018**, *13*, 760–770. [\[Referencia cruzada\]](#)
- Deveikyte, J.; Germán, H.; Piccari, C.; Provetti, A. Un enfoque de análisis de sentimiento para la predicción de la volatilidad del mercado. *arXiv* **2020**, arXiv:2012.05906.
- Mudinas, A.; Zhang, D.; Levene, M. Predicción de tendencias de mercado mediante análisis de sentimientos: lecciones aprendidas y caminos a seguir. *arXiv* **2019**, arXiv:1903.05440.
- Pimprikar, R.; Ramachandran, S.; Senthilkumar, K. Uso de algoritmos de aprendizaje automático y análisis de sentimientos de Twitter para la predicción del mercado de valores. En *t. J. Aplicación pura. Matemáticas*. **2017**, *115*, 521–526.
- Kilimci, ZH Análisis del sentimiento financiero con Deep Ensemble Models (DEM) para la predicción del mercado de valores. *Fac. J. Ing. Archit. Universidad de Gazi*. **2020**, *35*, 635–650.
- Al Kubaisi, GAA; Kamaruddin, SS; Husni, H. Modelo de clasificación del mercado de valores utilizando análisis de sentimiento en Twitter basado en clasificadores híbridos Naive Bayes. *computar información ciencia* **2017**, *11*, 52–64. [\[Referencia cruzada\]](#)
- Al-mashhadani, MI; Hussein, KM; Khudir, ET Análisis de sentimiento utilizando conjuntos de características optimizadas en diferentes facebook/twitter dominios de conjuntos de datos utilizando big data. *Iraqi J. Comput. ciencia Matemáticas*. **2022**, *3*, 64–70. [\[Referencia cruzada\]](#)
- Zhang, G.; Xu, L.; Xue, Y. Modele y pronostique el comportamiento del mercado de valores integrando el análisis del sentimiento de los inversores y los datos de transacciones. *grupo. computar* **2017**, *20*, 789–803. [\[Referencia cruzada\]](#)
- Xu, Q.; Chang, V.; Hsu, C.-H. Estudio de eventos y análisis de componentes principales basado en el análisis de sentimiento: una metodología combinada para estudiar el mercado de valores con un estudio empírico. *información sist. Frente*. **2020**, *22*, 1021–1037. [\[Referencia cruzada\]](#)
- Wu, J.-L.; Huang, MT.-T.; Yang, C.-S.; Liu, K.-H. Análisis de sentimiento de los mercados bursátiles utilizando un enfoque novedoso de excitación de valencia dimensional. *Cómputo suave*. **2021**, *25*, 4433–4450. [\[Referencia cruzada\]](#)
- Qian, F.; Gong, C.; Sharma, K.; Liu, Y. Generador de respuesta de usuario neuronal: Detección de noticias falsas con inteligencia de usuario colectiva. *IJCAI* **2018**, *18*, 3834–3840.
- Munk, M.; Pilikova, A.; Benko, L.; Blazekova, P.; Svec, P. Análisis de uso de la web del Pilar 3 información divulgada por clientes de depósito en tiempos turbulentos. *Sistema experto aplicación* **2021**, *185*, 115503. [\[Referencia cruzada\]](#)

27. Kan, W.; Malik, U.; Ghazanfar, MA; Azam, MA; Ayoubi, KH; Alfakheh, AS Predecir las tendencias del mercado de valores utilizando algoritmos de aprendizaje automático a través del sentimiento público y el análisis de la situación política. *Cómputo suave*. **2020**, *24*, 11019–11043. [\[Referencia cruzada\]](#)
28. Agrawal, A.; Hamling, T. Análisis de sentimiento de los tuits para obtener información sobre las elecciones estadounidenses de 2016. *Licenciatura en Colombia. ciencia j* **2021**, *11*. [\[Referencia cruzada\]](#)
29. Ding, X.; Liu, B.; Yu, PS Un enfoque holístico basado en el léxico para la extracción de opiniones. En *Actas de la Conferencia Internacional de 2008 on Web Search and Data Mining*, Palo Alto, CA, EE. UU., 11 y 12 de febrero de 2008; págs. 231–240.
30. Singh, T.; Nayyar, A.; Solanki, A. Sistema de recomendación de películas de minería de opinión multilingüe utilizando RNN. En *Actas de la Primera Conferencia Internacional sobre Informática, Comunicaciones y Seguridad Cibernética (IC4S 2019)*; Springer: Singapur, 2020; págs. 589–605.
31. Cambria, E.; Speer, R.; Havasi, C.; Hussain, A. Senticnet: un recurso semántico disponible públicamente para la extracción de opiniones. En *Commonsense Knowledge: Documentos del Simposio de Otoño AAAI*; AAAI Press: Menlo Park, CA, EE. UU., 2010.
32. Aung, KZ; Myo, NN Análisis de sentimiento de los comentarios de los estudiantes utilizando un enfoque basado en el léxico. En *Actas de la 16.ª Conferencia Internacional sobre Informática y Ciencias de la Información (ICIS) de IEEE/ACIS de 2017*, Wuhan, China, 24–26 de mayo de 2017; págs. 149–154.
33. Taboada, M.; Brooke, J.; Tofiloski, M.; Voll, K.; Stede, M. Métodos basados en Lexicon para el análisis de sentimientos. *computar Lingüista*. **2011**, *37*, 267–307. [\[Referencia cruzada\]](#)
34. de Smedt, T.; Daelemans, W. "Vreselijk mooi!" (terriblemente hermoso): un léxico de subjetividad para adjetivos holandeses. En *Actas de la Octava Conferencia Internacional sobre Evaluación y Recursos Lingüísticos (LREC'12)*; Asociación Europea de Recursos Lingüísticos (ELRA): Luxemburgo, 2012; págs. 3568–3572.
35. Bravo-Márquez, F.; franco, e.; Pfahringer, B. Creación de un léxico de opinión de Twitter a partir de tweets con anotaciones automáticas. *Sistema basado en el conocimiento*. **2016**, *108*, 65–78. [\[Referencia cruzada\]](#)
36. Rao, D.; Ravichandran, D. Inducción de léxico de polaridad semisupervisada. En *Actas de la 12ª Conferencia de la Unión Europea Capítulo de la ACL (EACL 2009)*, Atenas, Grecia, 30 de marzo–3 de abril de 2009; págs. 675–682.
37. de Albornoz, JC; Plaza, L.; Gervás, P. SentiSense: Un léxico afectivo basado en conceptos fácilmente escalable para el análisis de sentimientos. En *Actas de la Octava Conferencia Internacional sobre Evaluación y Recursos Lingüísticos (LREC'12)*; Asociación Europea de Recursos Lingüísticos (ELRA): Luxemburgo, 2012; págs. 3562–3567.
38. Cerini, S.; Compagnoni, V.; Demontis, A.; Formentelli, M.; Gandini, C. Micro-WNOp. *Recursos Lingüísticos y Teoría Lingüística*; Franco Angeli: Milán, Italia, 2007; pags. 200.
39. Miller, GA WordNet: Una base de datos léxica para inglés. *común ACM* **1995**, *38*, 39–41. [\[Referencia cruzada\]](#)
40. Strapparava, C.; Valitutti, A. Wordnet afecto: Una extensión afectiva de wordnet. En *Actas de la Cuarta Conferencia Internacional sobre Evaluación y Recursos Lingüísticos (LREC'04)*; Asociación Europea de Recursos Lingüísticos (ELRA): Luxemburgo, 2004; Volumen 4, pág. 40
41. Esuli, A.; Sebastiani, F. Sentiwordnet: Un recurso léxico disponible públicamente para la minería de opiniones. En *Actas de la Quinta Conferencia Internacional sobre Evaluación y Recursos Lingüísticos (LREC'06)*; Asociación Europea de Recursos Lingüísticos (ELRA): Luxemburgo, 2006.
42. Schuster, M.; Nakajima, K. Búsqueda por voz en japonés y coreano. En *Actas de la Conferencia Internacional IEEE 2012 sobre Procesamiento de Acústica, Habla y Señales (ICASSP)*, Kioto, Japón, 25–30 de marzo de 2012; págs. 5149–5152.
43. Sennrich, R.; Haddow, B.; Birch, A. Traducción automática neuronal de palabras raras con unidades de subpalabras. *arXiv* **2015**, arXiv:1508.07909.
44. Lewis, DD; Ringuette, M. Una comparación de dos algoritmos de aprendizaje para la categorización de texto. En *Tercer Simposio Anual de Análisis de Documentos y Recuperación de Información*; Universidad de Nevada: Reno, NV, EE. UU., 1994; Volumen 33, págs. 81–93.
45. Ding, S.; Zhao, H.; Zhang, Y.; Xu, X.; Nie, R. Máquina de aprendizaje extremo: algoritmo, teoría y aplicaciones. *Artefacto Intel. Rvdo*. **2015**, *44*, 103–115. [\[Referencia cruzada\]](#)
46. Bodapati, S.; Bandarupally, H.; Shaw, enfermero registrado; Ghosh, A. Comparación y análisis de RNN-LSTM y CNN para la clasificación de reseñas sociales. En *Avances en Aplicaciones de la Computación Basada en Datos*; Springer: Berlín/Heidelberg, Alemania, 2021; págs. 49–59.
47. Wei, D.; Wang, B.; Lin, G.; Liu, D.; Dong, Z.; Liu, H.; Liu, Y. Investigación sobre minería de datos de texto no estructurado y clasificación de fallas Basado en RNN-LSTM con informe de inspección de mal funcionamiento. *Energías* **2017**, *10*, 406. [\[Referencia cruzada\]](#)
48. Williams, G.; Baxter, R.; El, H.; Hawkins, S.; Gu, L. Un estudio comparativo de RNN para la detección de valores atípicos en la minería de datos. En *Actas de la Conferencia Internacional IEEE de 2002 sobre Minería de Datos*, Ciudad de Maebashi, Japón, 9–12 de diciembre de 2002; págs. 709–712.
49. Khan, FH; Qamar, U.; Bashir, S. Un enfoque semisupervisado para el análisis de sentimientos utilizando la fuerza del sentimiento revisada basada en SentiWordNet. *Saber información sist*. **2017**, *51*, 851–872. [\[Referencia cruzada\]](#)
50. Zhao, H.; Sol, M.; Deng, W.; Yang, X. Un nuevo método de extracción de características basado en EEMD y entropía difusa multiescala para Rodamiento de motores. *Entropía* **2016**, *19*, 14. [\[Referencia cruzada\]](#)
51. Nguyen, TH; Shirai, K.; Velcin, J. Análisis de sentimiento en las redes sociales para la predicción del movimiento de acciones. *Sistema experto aplicación* **2015**, *42*, 9603–9611. [\[Referencia cruzada\]](#)
52. Bing, L.; Chan, KC; Ou, C. Análisis de la opinión pública en los datos de Twitter para predecir los movimientos del precio de las acciones de una empresa. En *las actas de la 11.ª Conferencia internacional sobre ingeniería de comercio electrónico del IEEE de 2014*, Guangzhou, China, del 5 al 7 de noviembre de 2014; págs. 232–239.
53. Nti, IK; Adekoya, AF; Weyori, BA Predicción del movimiento del precio del mercado de valores utilizando el análisis de sentimiento: evidencia de Ghana. *aplicación computar sist*. **2020**, *25*, 33–42. [\[Referencia cruzada\]](#)

- 
54. Kan, W.; Ghazanfar, MA; Azam, MA; Karami, A.; Ayoubi, KH; Alfakheh, AS Predicción del mercado de valores utilizando clasificadores de aprendizaje automático y redes sociales, noticias. J. Inteligencia ambiental. Humanizar computar **2020**, *13*, 3433–3456. [\[Referencia cruzada\]](#)
55. Kordonis, J.; Symeonidis, S.; Arampatzis, A. Pronóstico del precio de las acciones a través del análisis de sentimiento en Twitter. En Actas del 20 Conferencia Panhelénica sobre Informática, Patras, Grecia, 10 a 12 de noviembre de 2016; págs. 1 a 6.
56. Jing, N.; Wu, Z.; Wang, H. Un modelo híbrido que integra el aprendizaje profundo con el análisis del sentimiento de los inversores para la predicción del precio de las acciones. Sistema experto aplicación **2021**, *178*, 115019. [\[Referencia cruzada\]](#)