Análisis descriptivo del precio acciones utilizando procesado del lenguaje natural

**ALUMNO 1: David de la Torre Amengual.**

**ALUMNO 1: Henry Ocaña Luna.**

**ALUMNO 1: Óscar Tienda Beteta.**

**ALUMNO 1: Mireya Harillo Gámez.**

**PROGRAMA:**

**MASTER EN BUSSINESS INTELLIGENCE AND DATA SCIENCE**

**NOMBRE DEL PROYECTO:**

Análisis descriptivo del precio de las acciones de GameStop mediante comentarios del subforo de Reddit WallStreetBets

# **CONTENIDO**

[**CONTENIDO** 2](#_Toc82202287)

[**TABLA DE ILUSTRACIONES** 3](#_Toc82202288)

[**RESUMEN** 4](#_Toc82202289)

[**INTRODUCCIÓN** 4](#_Toc82202290)

[**ESTADO DEL ARTE** 5](#_Toc82202291)

[**OBJETIVOS** 7](#_Toc82202292)

[**SOLUCIÓN PLANTEADA** 7](#_Toc82202293)

[**EVALUACIÓN** 8](#_Toc82202294)

[**RESULTADOS** 8](#_Toc82202295)

[**CONCLUSIONES Y TRABAJOS FUTUROS** 8](#_Toc82202296)

[**REFERENCIAS** 8](#_Toc82202297)

# **TABLA DE ILUSTRACIONES**

[Ilustración 1. Evolución de precios GME. 5](#_Toc82202151)

# **RESUMEN**

El problema planteado para el desarrollo de este proyecto es el análisis del comportamiento del precio de las acciones de la compañía GameStop Corporation, que cotiza bajo el símbolo bursátil GME, tras ser el objetivo de inversores de un conocido sub-foro de la plataforma Reddit conocido como WallStreetBets. Con este proyecto se pretende describir cuales fueron las causas del aumento repentino de las acciones de dicha compañía en un determinado periodo de tiempo, mediante el análisis de lenguaje natural que se encuentra en las publicaciones del foro.

Los datos utilizados para el proyecto se obtienen de dos fuentes. Por una parte, los datos de las publicaciones en el foro de Reddit provienen de una base de datos ya extraída de Kaggle. Por otro lado, los datos de los precios de las acciones se han obtenido del sitio web oficial de NASDAQ, National Association of Securities Dealers Automated Quotation.

El objetivo principal del proyecto es analizar si los comentarios del sub-foro WallStreetBets son una fuente fiable de la cual se puede extraer valor para invertir en acciones del mercado de valores, **utilizando técnicas de Deep Learning** cuyos resultados aporten una ventaja competitiva frente a otros inversores.

# **INTRODUCCIÓN**

La introducción debe contener el detalle sobre:

En enero de 2021, se produjo un [estrangulamiento de posiciones cortas](https://es.wikipedia.org/wiki/Estrangulamiento_de_posiciones_cortas) de las acciones del minorista estadounidense de videojuegos [GameStop](https://es.wikipedia.org/wiki/GameStop) y otros valores en varias bolsas de valores, lo que provocó importantes pérdidas financieras para los vendedores en corto, incluidos ciertos [fondos de cobertura](https://es.wikipedia.org/wiki/Fondo_de_cobertura). La contracción corta provocó que su precio alcanzase más de 500 dólares por acción, casi 30 veces el precio de principios de mes. Aproximadamente el 140 por ciento de las acciones de GameStop se habían vendido al descubierto, y la prisa por comprar acciones para cubrir esas posiciones a medida que el precio subía hizo que el precio incrementase aún más. El breve apretón fue inicialmente y principalmente provocado por usuarios del foro de Internet WallStreetBets en [Reddit](https://es.wikipedia.org/wiki/Reddit).

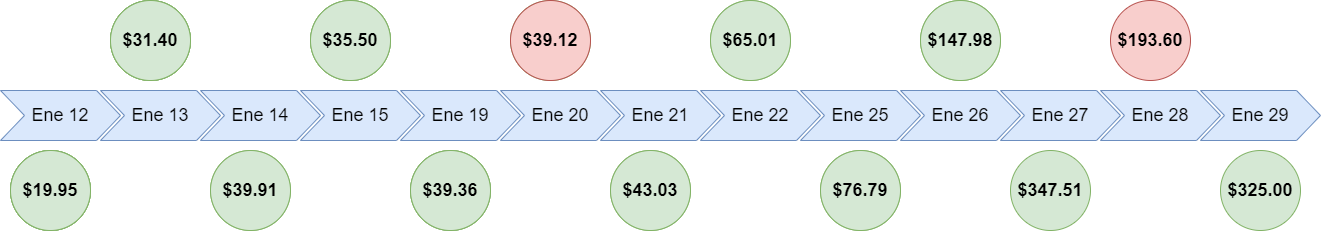


Ilustración 1. Evolución de precios GME

El 28 de enero, varias casas de bolsa, incluida Robinhood, detuvieron la compra de GameStop y otros valores, y luego citaron su incapacidad para depositar garantías suficientes en las cámaras de compensación para ejecutar las órdenes de sus clientes. Esta decisión atrajo críticas y acusaciones de manipulación del mercado por parte de destacados políticos y empresarios de todo el espectro político. Se presentaron demandas colectivas contra Robinhood en los Tribunales de Distrito de EE. UU. para el Distrito Sur de Nueva York y el Distrito Norte de Illinois. Muchos otros valores muy cortos también experimentaron aumentos de precios. En reacción a las casas de bolsa que detuvieron la compra de GameStop y otros valores, la capitalización total del mercado de las criptomonedas también aumentó.​

La Ilustración 1. Evolución de precios GME resume la evolución que sufrió el precio de las acciones GME en el tiempo hasta el bloqueo de transacciones en la plataforma de inversión Robinhood.

La solución que proponemos es innvoadora , ya que no se registra precedente alguno donde se haya tratado de abordar esta coyuntura utilizando Técnicas de Machine Learning. Para la realización del trabajo hemos dividido el desarrollo en 2 partes principales: Análisis Descrptivo - Exploratorio y Análisis Predictivo. En ambos casos se hará uso de técnicas de análisis supervisado como Análisis de Sentimiento y Árboles de Decisiones utilizando como lenguaje de programación Python y apoyándonos en una plataforma de visualización como es el caso de Qliksense.

Como primera medida se ha hecho uso del repositorio GitHub para mantener un orden en todas las versiones de los documentos con los que estamos trabajando. El documento principal de desarrollo es un Jupyter notebook en lenguaje Python. Los archivos de documentación son las plantillas en word y nuestro documento de apoyo es un qvf de Qliksense.

El desarrollo de la jupyter notebook consta de un análisis descriptivo de las principales variables de ambos datasets, un análisis de sentimientos de las publicaciones en el sub-foro de la plataforma Reddit conocido como WallStreetBets y un análisis predictivo entre la relación de sentimientos con los precios de las acciones.

* Resumen brevísimo de los resultados obtenidos.
* Resumen de la estructura de este documento.

Extensión máxima: 2000 palabras.

*Escribe aquí…*

# **ESTADO DEL ARTE**

En este apartado el estudiante deberá reportar al menos 30 artículos, proyectos, o bibliografía disponible en buscadores científicos que describan cómo se ha afrontado el problema que él está tratando de solucionar, justificando con esta literatura la relevancia de la solución que ha planteada.

Un proyecto de master implica que el estudiante se acerque al planteamiento de un problema poco explorado en la literatura y que ofrezca una solución innovadora al mismo.

Por tanto, el estado del arte debe tener un sub-apartado de conclusiones en el cual el estudiante justifique el carácter innovador y de nuevo conocimiento de su TFM.

Extensión máxima: 2000 palabras.

*Escribe aquí…*

# **OBJETIVOS**

En este apartado el estudiante definirá el objetivo general de su Trabajo Final de Máster y máximo 4 objetivos específicos que describan cómo llevó a cabo el objetivo general.

El objetivo general debe dar solución a problema realmente retador en el área del BI o DS, y debe haber quedado bien justificado en el estado del arte.

Objetivo general

Objetivos específicos

1. Objetivo 1
2. Objetivo 2
3. Objetivo 3
4. Objetivo 4

Extensión máxima: 1 página.

*Escribe aquí…*

# **SOLUCIÓN PLANTEADA**

En esta sección el estudiante debe describir la solución planteada, comenzando por la metodología (pasos que siguió) de desarrollo. Posteriormente la descripción del desarrollo de cada etapa seguida.

La metodología que debe utilizar el estudiante de master debe estar validada por la comunidad científica y el estudiante debe justificarlo.

1. METODOLOGÍA
   * Etapa 1
   * Etapa 2
   * Etapa 3
   * Etapa 4
   * Etapa 5
2. DESARROLLO DE CADA ETAPA

Extensión máxima: 30 páginas.

*Escribe aquí…*

# **EVALUACIÓN**

El estudiante en esta sección debe describir cómo realizó la evaluación de la solución planteada.

La evaluación de un TFM debe basarse en una metodología validada y bien establecida que el estudiante deberá justificar.

Extensión máxima: 10 páginas.

*Escribe aquí…*

# **RESULTADOS**

El estudiante en esta sección debe describir los resultados obtenidos en el proceso de evaluación de la solución planteada.

Debe usar notación estándar para presentación de resultados de carácter científico.

Extensión máxima: 10 páginas.

*Escribe aquí…*

# **CONCLUSIONES Y TRABAJOS FUTUROS**

El estudiante plantea las conclusiones de su trabajo, y cómo considera que puede seguir avanzando en el mejoramiento de la solución planteada.

Extensión máxima: 2 páginas

*Escribe aquí…*

# **REFERENCIAS**

Lista de referencias bibliográficas utilizadas, siguiendo las normas APA.