



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
(ICAI)

DOBLE GRADO EN INGENIERÍA INDUSTRIAL Y ADE

**Predicción de precios en mercados
financieros usando la información de
las noticias a través de métodos
estadísticos**

AUTOR: Javier Arturo San Millán Montero

DIRECTOR: Carlos Maté Jiménez

Madrid, Julio de 2020

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título
Predicción de precios en mercados financieros usando la información de las noticias a
través de métodos estadísticos

"Price forecasting in financial markets using news information through statistical methods"

En la ETS de Ingeniería – ICAI de la Universidad Pontificia Comillas en el
Curso académico 2019/20 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos. El Proyecto no es plagio de
otro, ni total ni parcialmente y la información que ha sido tomada de otros
documentos está debidamente referenciada.

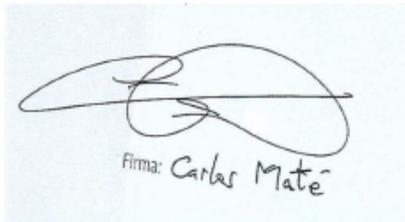
Fdo.: Javier Arturo San Millán Montero

Fecha: 12/07/2018



Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO



Firma: Carlos Maté

Fdo.:

Fecha: 13/ 07/ 2020

Professor Carlos Maté
Department of Industrial Organization
Alberto Aguilera, 25 - 28015 Madrid
Tel. +34 91 542 28 00 Ext. 2430
cmate@comillas.edu;



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
(ICAI)

DOBLE GRADO EN INGENIERÍA INDUSTRIAL Y ADE

**Predicción de precios en mercados
financieros usando la información de
las noticias a través de métodos
estadísticos**

AUTOR: Javier Arturo San Millán Montero

DIRECTOR: Carlos Maté Jiménez

Madrid, Julio de 2020

Predicción de precios en mercados financieros usando la información de las noticias a través de métodos estadísticos

Autor: Javier Arturo San Millán Montero

Director: Carlos Maté Jiménez

Entidad colaboradora: Icai – Universidad Pontificia Comillas

Resumen

Introducción

Cada vez hay más noticias que tratan sobre los mercados financieros, esas noticias pueden tener un efecto en la bolsa. En la actualidad, se están creando modelos que intentan combinar los efectos de las noticias con los modelos tradicionales.

Este TFG quiere obtener una relación entre el valor sentimental de las noticias y el mercado financiero de las acciones.

Metodología

El primer paso en este trabajo fin de grado ha sido la recolección de los datos de empresas del S&P500, la bolsa donde las mejores 500 empresas estadounidenses cotizan, y su procesamiento de los datos recogidos.

Por otro lado, la clasificación del valor sentimental se ha trabajado con un banco de palabras en Excel que se ha creado por cuenta propia.

Por último, con los datos de la clasificación obtenidos se han importado a MATLAB donde finalmente se ha llevado a cabo el análisis.

Resultados y conclusiones

Tras analizar más de 50 empresas, se ha llegado a varias conclusiones.

En primer lugar, no se ha podido rechazar la teoría del camino aleatorio debido a que el

modelo predictivo no fue capaz de predecir los futuros valores de la cotización.

En segundo lugar, el valor sentimental que había sido asignado a una noticia no solía coincidir con el efecto de esa noticia. Además, esto fue apoyado por las series temporales que realiza la herramienta.

Por último, las noticias suelen tener un alto impacto en el volumen de la empresa.

Price forecasting in financial markets using news information through statistical methods

Author: Javier Arturo San Millán Montero

Director: Carlos Maté Jiménez

Collaborating institution: Icai – Universidad Pontificia Comillas

Abstract

Introduction

Every day, there are more and more news about financial markets, these news may have an impact in the stock market. Currently, Models that combine traditional methods with news are being developed.

This end-of-grade project wants to obtain a relationship between the sentimental value of the news and the financial market of the shares.

Methodology

The first step in this end-of-grade project was the collection of company data from the S&P500, the stock exchange where the top 500 U.S. companies are listed, and their processing of the collected data.

On the other hand, the sentimental value classification has been worked with a word bank in Excel that has been created from scratch

Finally, with the classification data obtained they have been imported into MATLAB where the analysis has finally been carried out.

Results and conclusions

After analyzing more than 50 companies, several conclusions have been reached.

First of all, the random path hypothesis could not be rejected because the predictive model was not able to predict future quote values.

Secondly, the sentimental value that had been assigned to a news story did not often match the effect of that news. Moreover, this claim was supported by the time series that the tool performs.

Finally, news often has a high impact on the company's volume.

Agradecimientos

Se quiere hacer una mención a las personas que me apoyaron durante todo el trabajo de fin de grado, en especial a mi novia Carol, a mi hermana Celia, a mis padres y a mi amigo Ángel. Pero no hay que olvidarse del director de este TFG, Carlos Maté, el cual ha sido una fuente de inspiración cuando no sabía por donde seguir. Por último, pero no menos importante, este TFG va dirigido a mi compañero y amigo Marino.

Índice

1. Predicción de precios en los mercados de acciones.....	1
a. Introducción.....	1
b. Teoría del Camino Aleatorio.....	1
c. Hipótesis de los mercados eficientes.....	2
d. Análisis técnico y fundamental.....	3
e. Contrastes de hipótesis.....	3
i. Contraste de pendientes.....	4
ii. Contraste de diferencia de medias de dos normales en muestras emparejadas.....	5
2. Métodos estadísticos en la predicción de precios.....	7
a. Introducción.....	7
b. Regresión lineal.....	7
c. Regresión no lineal.....	8
d. Modelo Holt.....	9
3. Efecto de las noticias en el mercado de acciones. Consideraciones para el S&P500.....	11
a. Introducción.....	11
b. Datos.....	13
c. Valor sentimental.....	14
d. Análisis.....	17
4. Herramienta desarrollada.....	23
a. Introducción.....	23
b. Excel.....	23
c. Matlab.....	27
i. Filtro	27
ii. Conversión de noticias.....	34
iii. Diagrama de cajas.....	35
1. Valor sentimental.....	35
2. Variación diaria.....	39
iv. Variación diaria con noticias.....	42
v. Cotización con medias móviles.....	44
vi. Primera ventana de la herramienta.....	46
vii. Modelo Holt herramienta.....	48
viii. Cotización con análisis sentimental de las noticias.....	53
ix. Efecto de las noticias por medio del contraste de pendientes.....	54
x. Efecto de las noticias por medio del contraste de variables emparejadas.....	58
xi. Segunda ventana de la herramienta.....	62
xii. Volumen	62
1. Volumen con análisis sentimental de las noticias.....	63
2. Efecto de las noticias por medio del contraste de pendientes.....	63
3. Efecto de las noticias por medio del contraste de variables emparejadas.....	64
4. Tercera ventana de la herramienta.....	65
5. Las noticias según el volumen	66
xiii. Series temporales.....	68
5. Conclusiones y futuras líneas de investigación.....	75

a.	Introducción.....	75
b.	Conclusiones.....	75
	i. Efecto de las noticias en la cotización.....	76
	ii. Volumen y las noticias.....	78
	iii. Series temporales.....	81
	iv. Modelo Holt.....	86
c.	Futuras extensiones.....	87
	i. Completar el estudio con los datos del COVID-19.....	88
	ii. Procesamiento del Lenguaje Natural.....	88
	iii. Bot recoge noticias.....	89
6.	Bibliografía.....	90
7.	Anexo ODS.....	92

ÍNDICE DE FIGURAS

Figura 1. Cotización Disney	12
Figura 2 Cotización Monsanto.....	13
Figura 3. Muestra de los datos obtenidos de Baxter International Inc.	24
Figura 4. Muestra del Excel que busca las palabras con carácter negativo.....	25
Figura 5. Muestra del Excel que busca las palabras con carácter positivo.....	25
Figura 6. Ejemplo de inserción de datos.....	26
Figura 7. Ejemplo de visualización de empresas	28
Figura 8. Verificación número de empresa.....	31
Figura 9. Filtro del periodo	31
Figura 10. Verificación del filtro del periodo	31
Figura 11. Filtro por fecha	32
Figura 12. Comprobación fecha final	34
Figura 13. Diagrama de cajas de VS de Avon	36
Figura 14. Diagrama de cajas de VS de AT&T	36
Figura 15. Diagrama de cajas de VS de Goldman Sachs	37
Figura 16. Diagrama de cajas de Coca Cola	38
Figura 17. Diagrama de cajas de la variación diaria de Avon.....	39
Figura 18. Diagrama de cajas de la variación diaria de Chevron.....	40
Figura 19. Diagrama de cajas de la variación diaria de General Electric.....	41
Figura 20. Variación diaria con noticias de Avon	42
Figura 21. Comparativa diagrama de cajas y la variación diaria con noticias de Avon.....	43
Figura 22. Medias móviles de cotización de Avon.....	44
Figura 23. Medias móviles de cotización de Chevron	45
Figura 24. Apariencia de las gráficas en la herramienta	46
Figura 25. Apariencia de las gráficas con filtro fecha	47
Figura 26. Modelo Holt con los mejores estimadores de Google	48
Figura 27. Zoom modelo Holt figura 26	49
Figura 28 Modelo Holt con restricciones.....	49

Figura 29. Zoom modelo Holt figura 28	50
Figura 30. Modelos Holt	51
Figura 31. Zoom figura 30	51
Figura 32. Cotización con noticias	52
Figura 33. Comparación de Cotización con noticias vs. variación diaria	53
Figura 34. Cotización con el VS noticias por el contraste de pendientes	54
Figura 35. Comparativa del VS y el contraste de pendientes	56
Figura 36. Diferencias en las noticias	56
Figura 37. Cotización con el contraste de variables emparejadas	58
Figura 38. Comparativa entre VS y el contraste de variables emparejadas	59
Figura 39. Zoom figura 38	60
Figura 40. Comparativa entre los 3 métodos	61
Figura 41. Volumen con noticias de McDonald's	62
Figura 42. Volumen con noticias evaluadas con el contraste de las pendientes	63
Figura 43. Volumen con noticias evaluadas con el contraste de variables emparejadas	64
Figura 44. Comparativa del volumen con los 3 métodos	65
Figura 45. Volumen con noticias clasificadas según volumen	67
Figura 46. Muestra de todo el periodo de PepsiCo	68
Figura 47. Primera serie temporal de PepsiCo	69
Figura 48. Segunda serie temporal de PepsiCo	70
Figura 49. Tercera serie temporal de PepsiCo	71
Figura 50. Cuarta serie temporal de PepsiCo	72
Figura 51. Quinta serie temporal de PepsiCo	73
Figura 52. Comparativa del efecto de las noticias en la cotización 3M	76
Figura 53. Comparativa del efecto de las noticias en la cotización RF	76
Figura 54. Volumen y noticias en Walmart	77
Figura 55. Volumen y noticias con contraste de pendientes en Walmart	78
Figura 56. Volumen y noticias con contraste variables pareadas en Walmart	79
Figura 57. Volumen y noticias por volumen en Walmart	79
Figura 58. Primera serie temporal de Qualcomm	80
Figura 60. Tercera serie temporal de Qualcomm	82
Figura 61. Cuarta serie temporal de Qualcomm	83
Figura 62. Quinta serie temporal de Qualcomm	84

Índice de Tablas

1. Hipótesis para el contraste de diferencia de medias de dos normales en muestras emparejadas.....	5
2. Notación utilizada en el modelo Holt.....	9
3. Ejemplos de palabras positivas y negativas.....	14
4. Notación utilizada con el valor sentimental.....	15
5. Lista con las noticias de las empresas.....	17
6. Empresas con más noticias positivas.....	18
7. Empresas con menos noticias positivas.....	19
8. Empresas con más noticias negativas.....	19
9. Empresas con menos noticias negativas.....	20
10. Valor sentimental de las noticias a lo largo del tiempo.....	21
11. Empresas y sector.....	30
12. Clasificación de noticias.....	35
13. Noticias según su efecto con el contraste de pendientes.....	55
14. Noticias según su efecto con el contraste de variables emparejadas.....	58
15. Clasificación de noticias según volumen.....	67
16. Notación de las noticias según volumen.....	68

Capítulo 1

Predicción de precios en los mercados de acciones

Introducción

En este capítulo se tratarán los siguientes temas: la teoría del camino aleatorio, la hipótesis de los mercados eficientes, el análisis técnico y fundamental, y los contrastes de hipótesis que se van a utilizar a lo largo de este trabajo de fin de grado.

Este capítulo es importante debido a que se empiezan a introducir conceptos que serán importantes a lo largo de esta memoria, y dan una base al lector.

Teoría del Camino Aleatorio

La teoría del camino aleatorio estipula que los precios no se pueden predecir con exactitud debido a la impredecibilidad y aleatoriedad de los mercados eficientes. Si existiera una tendencia o se pudieran predecir los precios con exactitud, se podría utilizar para obtener beneficios superiores a los que da el mercado. Rodríguez et al (2006) [11]

En este trabajo de fin de grado, se podrá a prueba la teoría del camino aleatorio con algún modelo predictivo que será explicado en los próximos capítulos. En el caso que se pueda predecir con exactitud, se habrá refutado esta teoría, pero si en cambio el modelo no es capaz de predecir, esta teoría seguirá siendo lo que es, una teoría.

Hipótesis de los mercados eficientes

Un mercado eficiente es aquel en el que los precios de los activos que en el cotizan reflejan la información disponible en el mercado, esta idea fue introducida por primera vez por Fama (1970) [4].

Para que un mercado se considere eficiente, Fama (1970) [4] estipula que se necesitan darse varias condiciones:

La primera condición es que toda información tiene que estar disponible y ser gratuita para todos los públicos. Lo cual implica que no pueden existir agentes que vendan información clasificada que pueda dar ventaja a algún inversor sobre el resto.

La segunda condición dice que los precios se ajustan rápidamente a la nueva información. No sería un mercado eficiente si ante cualquier suceso, los precios no variaran ya que no se estarían ajustando a la realidad.

Para terminar, el mercado debe tener muchos participantes que actúen en base a la información que hay disponible. Si la mayoría de los participantes en un mercado actúan al azar, sin tener en cuenta la información disponible, los precios no se ajustarían a la realidad, lo que haría imposible que ese mercado fuera eficiente.

Aparte, Fama (1970) [4] clasifica la eficiencia del mercado en tres formas.

En primer lugar, se encuentra la eficiencia débil, donde el mercado refleja todas las noticias pasadas. Si las noticias llegan de forma aleatoria, la modificación del precio será también aleatoria ya que no se puede prever cuando sucederá una noticia. Esto implica que no pueden existir patrones o tendencias sistemáticas que se mantengan en el tiempo

En segundo lugar, se encuentra la eficiencia semi-fuerte, en la cual toda la información disponible se encuentra reflejada en los precios. Esta forma se basa en los analistas, los cuales intentan conseguir beneficios con sus análisis. El analista que consiga unos mayores beneficios será porque su análisis es mejor. En el caso que ningún analista lo consiga, estaremos ante un caso de competencia perfecta porque nadie ha conseguido tener ventaja con su análisis.

Por último, se encuentra la eficiencia fuerte, en la cual los precios reflejan toda la información disponible, incluyendo información confidencial. Esto es una condición para que el mercado sea considerado eficiente. Además, esta condición implica que nadie pueda tener beneficios superiores a los que el mercado estipule.

Según Malkiel (2003) [7], en la realidad, la mayoría de los mercados tienen una eficiencia semifuerte, y un buen análisis puede dar rendimientos superiores a los que estipula el mercado. Aunque también está la posición de Dunis (2015) [3], el cual admite que hay un escepticismo justificable en la habilidad para ganar dinero prediciendo los movimientos de la bolsa en un mercado eficiente, pero que en la realidad los precios no se adecuan inmediatamente a la nueva información, lo que puede dar ventaja a algún analista si es lo suficientemente rápido.

Análisis técnico y fundamental

El análisis técnico se centra en estudiar y prever los movimientos en un mercado financiero, en especial del precio, usando gráficas como herramienta principal Stasinakis (2014) [13]. Debido al camino aleatorio y los mercados eficientes, esta tarea es bastante complicada. Aún así, el análisis técnico se basa en dos supuestos. El primero es que, al ser un mercado eficiente, se tiene toda la información. El segundo supuesto es que los movimientos de los precios tienden a repetirse a lo largo de la historia.

Según Murphy (1999) [8], la diferencia entre el análisis fundamental y el análisis técnico es que el primero estudia las causas del movimiento del mercado, mientras que el segundo estudia los efectos del movimiento del mercado.

Contrastes de hipótesis

Para poder determinar el efecto de una noticia en la cotización de una empresa, se van a desarrollar dos métodos distintos en la herramienta, de los cuales se explicarán paso a paso en el capítulo 4.

Contraste de pendientes

El primer método es el contraste de pendientes de dos rectas de regresión lineal. Se realizará una regresión lineal para los datos anteriores a la noticia, y otra regresión lineal para los datos posteriores a la noticia. Si las pendientes de las regresiones son diferentes, se podrá concluir que hay un efecto de la noticia en la cotización. Para realizar este contraste, se han seguido las instrucciones de Andrade (2014) [1], pero teniendo consideraciones al trabajo de Zarnoch (2009) [14]. A continuación, aparece una Tabla obtenida de su trabajo, en la cual compra las medidas.

	Means comparison	Slopes comparison
Null hypothesis	$H_0 \equiv \mu_1 = \mu_2$	$H_0 \equiv \beta_1 = \beta_2$
Statistic and its distribution	$\bar{x}_1 - \bar{x}_2 \in N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ \Downarrow $\bar{x}_1 - \bar{x}_2 \in N\left(\mu_1 - \mu_2, \sqrt{\text{var}(\bar{x}_1) + \text{var}(\bar{x}_2)}\right)$ <p style="text-align: center;">being σ_1^2, σ_2^2 the variances of the random variables in each dataset</p>	$b_1 - b_2 \in N\left(\beta_1 - \beta_2, \sqrt{\frac{\sigma_1^2}{\sum (x_{i1} - \bar{x}_1)^2} + \frac{\sigma_2^2}{\sum (x_{i2} - \bar{x}_2)^2}}\right)$ \Downarrow $b_1 - b_2 \in N\left(\beta_1 - \beta_2, \sqrt{\text{var}(b_1) + \text{var}(b_2)}\right)$ <p style="text-align: center;">being σ_1^2, σ_2^2 the variances of the regression model in each group, i.e., the variances of the response variables in each dataset</p>
F-test	$F_{\mu_1 - \mu_2} = \frac{s_1^2}{s_2^2} \in F_{n_1 - 1, n_2 - 1}$	$F_{\beta_1 - \beta_2} = \frac{s_{(y/x)_1}^2}{s_{(y/x)_2}^2} \in F_{n_1 - 2, n_2 - 2}$
t-Test assuming equal variances	$t_{\mu_1 - \mu_2} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\text{pool}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$ with $s_{\text{pool}}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$	$t_{\beta_1 - \beta_2} = \frac{b_1 - b_2}{s_{(y/x), \text{pool}} \sqrt{\frac{1}{\sum (x_{i1} - \bar{x}_1)^2} + \frac{1}{\sum (x_{i2} - \bar{x}_2)^2}}}$ with $s_{(y/x), \text{pool}}^2 = \frac{(n_1 - 2)s_{(y/x)_1}^2 + (n_2 - 2)s_{(y/x)_2}^2}{n_1 + n_2 - 4}$

Como se puede observar, Andrade (2014) [1] presupone que las varianzas, a pesar de ser desconocidas, son varianzas iguales. Pero en este trabajo de fin de grado se analiza el mercado financiero, el cual es impredecible y volátil, no se puede asumir que las varianzas sean iguales, por lo que el estadístico de contraste utilizado es:

$$t_{\beta_1 - \beta_2} = \frac{b_1 - b_2}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}};$$

Contraste de diferencia de medias de dos normales en muestras emparejadas

Las variables emparejadas son 2 variables, en este caso X e Y, dependientes. Se pueden encontrar variables emparejadas cuando se mira el antes y el después de un evento. Las variables emparejadas se utilizan en diversos campos, desde en medicina para comprobar si un medicamento mejora el estado de los enfermos, hasta en finanzas para comprobar si un evento ha supuesto un cambio en la tendencia de unas acciones. Este segundo ejemplo es el que se desarrolla en este trabajo de fin de grado, los eventos en cuestión serán las noticias. Por otro lado, la variable X serán las variaciones diarias antes de la noticia, mientras que la variable Y serán las variaciones diarias después de la noticia.

Lo primero es calcular la diferencia entre el antes y el después, para ello tenemos la variable d.

$$D_k = Y_k - X_k;$$

Siendo k cualquier valor comprendido dentro de la muestra.

Al tener n muestras, tendremos n valores de d:

$$(d_1, d_2, \dots, d_n) = (Y_1 - X_1, Y_2 - X_2, \dots, Y_n - X_n)$$

Una vez definidas las variables, hay que proponer las hipótesis, las cuales se disponen en la Tabla 1.

H_0	$\mu_d = 0$
$H_{1,a}$	$\mu_d > 0$
$H_{1,b}$	$\mu_d < 0$

Tabla 1

Siendo \bar{d} la media de la variable d, n el número de muestras, y t_{n-1} una t-student de n-1 grados de libertad. El contraste será el siguiente:

$$\mu_y - \mu_x = \bar{d} \pm t_{n-1} * \frac{S_d}{\sqrt{n}}$$

Este contraste será implementado en la herramienta como segundo método para determinar el efecto de una noticia, y poder comparar el anterior método del contraste de las pendientes con este. Pero todo eso será explicado en con más profundidad en el capítulo 4.

Capítulo 2

Métodos estadísticos en la predicción de precios

Introducción

En este capítulo se tratarán los siguientes temas: la regresión lineal, la regresión no lineal y el modelo Holt

En este capítulo trata la base de los cálculos que la herramienta es capaz de realizar, y necesarios para entender el capítulo 4.

Regresión lineal

Una regresión lineal es un modelo matemático que relaciona una variable dependiente, en este caso la denominaremos Y, con las variables independientes (X_1, X_2, \dots, X_n) y un error (ε).

De tal forma, que la ecuación de la regresión lineal tendrá la forma:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \dots + \beta_n * X_n + \varepsilon;$$

Siendo β_k el coeficiente que mejor se ajusta para la variable independiente X_k . Para obtener los mejores coeficientes β , se mira el modelo que tenga el mínimo error cuadrático.

La información se ha obtenido gracias a Pearson Education (2010) [9]

Las regresiones lineales se utilizan para tener un modelo matemático que relacionen variables que normalmente no tienen asignada una fórmula preestablecida. Por ejemplo: el tiempo que tarda en llegar al trabajo una persona dependiendo de la distancia de su casa al trabajo, o los ingresos de una persona dependiendo de su antigüedad en la empresa, sus estudios, su género, y su edad.

No todas las variables independientes serán relevantes para el modelo, es trabajo del analista determinar que variables son las más relevantes.

En este trabajo de fin de grado, las regresiones lineales que se han utilizado tienen una única variable independiente, que es el tiempo, y su variable dependiente es la cotización de la acción. Estas regresiones se han utilizado en diferentes partes para la creación de la herramienta. En primer lugar, se han utilizado en un método para averiguar el efecto de las noticias en la cotización de la bolsa. En segundo lugar, se han utilizado las regresiones lineales para separar la cotización en series temporales.

Regresión no lineal

Al igual que la regresión lineal, la regresión no lineal es un modelo matemático que relaciona la variable independiente Y con una función no lineal de variables independientes y un error (ε).

De tal forma, que la ecuación de la regresión no lineal tendrá la forma:

$$Y = f(X_1, X_2, \dots, X_n) + \varepsilon$$

Un ejemplo de función no lineal podría ser $f(X_1) = \beta_0 + \beta_1 * X_1^2$.

El problema de las regresiones no lineales viene en el nombre, que no son lineales, y, por lo tanto, en algunas ocasiones puede ser difícil obtener valores. Es por este motivo que el modelo se intenta linealizar, o al menos linealizar localmente.

En este trabajo de fin de grado se ha considerado que el análisis con regresiones lineales es suficiente, por lo que no se han realizado regresiones no lineales.

Modelo Holt

El modelo Holt es un método predictivo que se basa en los anteriores datos para hacer su predicción. Tiene dos componentes: el nivel y la tendencia. La realización de este modelo ha sido obtenida siguiendo los pasos de Sánchez (2018) [12].

El modelo Holt se calcula de la siguiente forma. A continuación, la Tabla 2 explica la notación utilizada.

Notación	Descripción
t	Periodo t
P	Pendiente
N	Nivel
C	Cotización
α	Estimador del nivel
β	Estimador de la pendiente
F	Predicción (Forecast)
E	Error

Tabla 2

A través de una regresión lineal, la cual ha sido explicada anteriormente, se obtiene una recta del estilo: $y = a * t + b$, con esta recta se obtienen los valores del nivel y tendencia para el primer periodo.

$$T_0 = a ; L_0 = b$$

Para el periodo t, la predicción y el error cometido se calculan de la siguiente forma.

$$E_t = F_t + C_t;$$

$$F_{t+1} = L_t + T_t;$$

A continuación, se obtendrán el nivel y la pendiente para el periodo t+1

$$N_{t+1} = \alpha * C_{t+1} + (1 - \alpha) * (N_t + P_t);$$

$$P_{t+1} = \beta * (N_{t+1} - N_t) + (1 - \beta) * P_t;$$

Una vez obtenidos la pendiente y el nivel para el periodo t+1, se obtendría la predicción para el periodo t+2, y se repetirían los pasos hasta que se terminen los datos.

En el capítulo 4, para obtener el mejor modelo Holt para la cotización, cambiarán los estimadores para obtener el modelo óptimo. Los estimadores α y β varían entre los rangos 0 y 1. Cuanto mayor sea α , el modelo tendrá más en cuenta los datos más recientes, mientras que, si es un valor cercano a 0, el modelo tendrá más en cuenta los datos más antiguos. Por otro lado, cuanto mayor sea β , el modelo responderá más rápidamente a los cambios de la tendencia, mientras que un β cercano a 0 responderá más lentamente a los cambios en la tendencia, lo que implica que el modelo esté más suavizado.

Capítulo 3

Efecto de las noticias en el mercado de acciones. Consideraciones para el S&P500

Introducción

Los datos son titulares de noticias de empresas que cotizan en el S&P500. El S&P500 es uno de los índices más importantes en el cual cotizan 500 de las empresas más grandes de Estados Unidos de América.

Los titulares de las noticias se han recogido de Bloomberg, la cual es la mayor empresa de noticias económicas, seguida de Thomson Reuters y Dow Jones. Bloomberg genera sobre 5.000 noticias diarias y provee noticias a las personas más influyentes en el mundo de las finanzas.

Los datos de la muestra van desde el 24/08/2016 hasta el 15/10/2019. En este periodo que dura 1.114 días, han sido recogidos titulares de noticias de 54 empresas distintas. Estas noticias han sido seleccionadas por la opción de Bloomberg “Eventos clave”, los cuales están representados por la letra H como se puede ver en la Fig. 1. Por otro lado, de han tenido que descartar algunas empresas por falta de noticias, se han descartado aquellas empresas que no llegaban a un mínimo 10 noticias en todo el periodo, y también se han descartado aquellas empresas que no hubieran cotizado todo el periodo.

Empresas como Walt Disney no ha pasado el corte porque tan solo tiene 5 eventos clave, los cuales son representados la letra *H* dentro de un cuadrado azul (ver Fig. 1), a pesar de ser la 24 empresa más grande del mundo y la más grande dedicada a los medios de comunicación, con una capitalización de más de doscientos mil millones de dólares americanos en 2019 según Bloomberg.



Figura 1. Cotización Disney

Por otro lado, empresas como Monsanto, empresa agroquímica que cotizaba en el S&P500, fue absorbida por Bayer, y solo hay datos hasta el 6/6/2018, lo cual supone más de un año sin datos de cotización porque el periodo de las muestras termina el 15/10/2019. Aparte, solo hay 5 eventos clave (ver Fig. 2). Por estos dos motivos, se descarta esta empresa para su análisis.



Figura 2 Cotización Monsanto

DATOS

Se han recogido 1828 titulares para analizar posteriormente con el valor sentimental de la noticia. El valor sentimental de la noticia es un valor numérico que se asigna a cada noticia para poder cuantificar ‘como de positiva’ o ‘como de negativa’ es una noticia.

La clasificación de palabras según su carácter positivo, negativo o neutro es la parte subjetiva de este trabajo de fin de grado, pero a la vez es una parte imprescindible porque se emplea para obtener el valor sentimental de la noticia. Como las noticias fueron recogidas en inglés, las palabras que se incluyeron en el banco de palabras positivas y en el banco de palabras negativas son palabras inglesas. Una palabra con carácter positivo es una palabra que indica si una noticia es favorable para los intereses de la empresa, por el contrario, una palabra con carácter negativo es una palabra que indica si una noticia es desfavorable para la empresa. Para terminar, una palabra neutra es aquella que no indica si una noticia es favorable o desfavorable. Por ejemplo: “Google mejora sus previsiones para el 2021”. En este caso, la palabra “mejora” es una palabra con carácter positivo porque indica que este titular es favorable para la compañía debido a que va a obtener mayores beneficios, mientras que el resto de las palabras son neutras. Por el contrario, si la noticia fuera “Google empeora sus previsiones para el 2021”. La palabra “empeora” es una palabra negativa porque implica que, en este caso, Google va a obtener peores resultados, y el resto de las palabras seguirían siendo neutras.

Para poder calificar las noticias, primero se han creado dos bancos de palabras, un banco de palabras positivas y otro de palabras negativas. Un banco de palabras es un listado de palabras que en este caso comparten alguna característica, en este caso, comparten su carácter, ya sea positivo o negativo. El banco de palabras positivas está compuesto por 125 familias de palabras mientras que el banco de palabras negativas está formado por 141 familias de palabras. El funcionamiento de los bancos de palabras para la asignación del valor sentimental de cada noticia se explicará en profundidad en el siguiente capítulo de Herramienta desarrollada. Debido al gran tamaño que tendría una Tabla completa con todas las familias de palabras, se incluye en la Tabla 3 que se muestra a continuación una muestra de las palabras traducidas que se han considerado positivas y negativas.

Positivas	Negativas
Mejorar	Alertar
Aplaudido	Culpar
Agradar	Confusión
Brillante	Preocupación
Épico	Fallar
Crecer	Perder
Favorable	Escépticos

Tabla 3

VALOR SENTIMENTAL

¿Se pueden clasificar las noticias de forma objetiva? Esta pregunta la abarcó Glasserman (2019) [5], entre otros, ya que no se puede analizar el efecto de las noticias en la cotización sin antes analizar el valor sentimental de estas.

El valor sentimental se podría considerar como una puntuación que indica como de positiva o negativa es una noticia, cuanto mayor sea esa puntuación, más positiva será esa noticia, mientras que cuanto menor sea el valor sentimental, más negativa será la noticia.

Las noticias se han calificado con el valor sentimental (VS), y el valor sentimental de la noticia i se calcula de la siguiente forma:

$$VS_i = \frac{n_i^{pos} - n_i^{neg}}{n_i^{tot}}$$

NOTACIÓN	DESCRIPCIÓN
VS_i	Valor sentimental de la noticia i
n_i^{pos}	Número de palabras positivas en el titular de la noticia i
n_i^{neg}	Número de palabras negativas en el titular de la noticia i
n_i^{tot}	Número de palabras totales en el titular de la noticia i

Tabla 4

La notación utilizada se explica en la Tabla 4.

El valor sentimental de una noticia será un número entre -1 y 1, porque el número total de palabras siempre tiene que ser mayor o igual a la suma del número de palabras positivas y negativas ($n^{tot} = n^{pos} + n^{neg} + n^{neutro}$). Cuanto más cerca de 1 se encuentre el valor sentimental de una noticia, más positiva se considera. Cuanto más cerca de -1 se encuentre el valor sentimental de una noticia, más negativa se considera. Cuanto más cercano a 0 se encuentre una noticia, más neutra se considera.

Un ejemplo de noticia positiva sería “Boeing 4Q core EPS beats highest est.; Shares rise”, en la cual nos dice que las acciones de Boeing suben porque Boeing dará más dividendos en el cuarto trimestre que lo que estimaba la previsión más optimista. Esta noticia se ha calificado con un valor sentimental de 0’555, el cual es uno de los más altos de la muestra recogida.

Un ejemplo de noticia neutra sería “A guide to investing from the richest family on the planet”, de Walmart, la cual se traduce como “Una guía para invertir de la familia más

rica del mundo”. No hay ninguna palabra positiva ni negativa en esta noticia, aparte no hay una aparente repercusión en Walmart.

Un ejemplo de noticia negativa sería “Regional banks slide along with yields, recession worries”, de USBankcorp, la cual nos indica que los bancos regionales empeoran y hay preocupación de una recesión. Esta es una de las noticias más negativas de la muestra con un -0’5 de valor sentimental.

Esta forma de calcular el valor sentimental de una noticia tiene varias limitaciones en cuanto a las frases que puede analizar. En primer lugar, califica de forma errónea las frases negativas. Por ejemplo: “Google no ha presentado unos buenos resultados en 2019”. La palabra “buenos” se ha incluido previamente en el banco de palabras positivas, y, por lo tanto, esta noticia se le asignaría un valor sentimental positivo, pero realmente esta noticia tiene un valor sentimental negativo. Por otro lado, no se puede incluir “no” en el banco de palabras negativas porque “no” es una palabra con carácter neutro que simplemente cambia el sentimiento de la noticia. Si usamos el ejemplo anterior: “Google no ha presentado unos buenos resultados en 2019” y aplicando la fórmula del valor sentimental correctamente, tendría un valor de aproximadamente $-0,111$ ($VS = (-1) * \frac{1-0}{9}$). Por el contrario, si esa misma frase fuera positiva, la frase sería “Google ha presentado unos buenos resultados en 2019”. Aplicando la fórmula del valor sentimental, tendría un valor de $0,125$ ($VS = \frac{1-0}{8}$), el cual es similar al anterior en valor absoluto. No son exactamente iguales porque al añadir “no”, el número de palabras totales aumenta en 1 unidad.

En segundo lugar, las noticias con ironía no pueden ser procesadas correctamente. Se utilizará “ironía” según lo define la Real Academia Española (RAE): “f. Expresión que da a entender algo contrario o diferente de lo que se dice, generalmente como burla disimulada”. En lenguaje oral se pueden identificar por el tono en el que se dice, pero de forma escrita es muy difícil identificarlos correctamente. Por ejemplo: “Google va a presentar unos buenísimos resultados este año”. Si se analiza esta frase, tendría un valor sentimental positivo de aproximadamente $0,111$ ($VS = \frac{1-0}{9}$), pero al no saber el contexto ni la intención del emisor, no se puede saber si esta noticia es verdaderamente positiva.

Por último, los titulares con frases exclamativas o con palabras superlativas no se procesan correctamente porque la fórmula únicamente tiene en cuenta el número de palabras con carácter, no la intensidad que procesan. Por ejemplo: “¡Google ha obtenido los mejores resultados del 2019!” tendría el mismo valor sentimental que “Google ha obtenido unos buenos resultados en 2019!”, aunque “los mejores resultados” sea más positivo que “unos buenos resultados”.

ANÁLISIS

Los resultados de los análisis de noticias por empresas se han representado a continuación en la Tabla 5.

Noticias					Noticias				
Empresa	Totales	Positivas	Neutras	Negativas	Empresa	Totales	Positivas	Neutras	Negativas
AVP	27	14	5	8	MCD	30	12	6	12
AXP	22	7	6	9	MDT	33	10	10	13
BA	47	12	8	27	MMM	29	8	10	11
BAX	27	9	2	16	MRK	37	17	6	14
CSCO	30	14	2	14	MSFT	38	11	13	14
CVX	39	22	6	11	NKE	33	13	4	16
DVN	24	12	4	8	ORCL	29	10	7	12
F	47	16	8	23	PEP	35	16	13	6
FCX	18	4	7	7	PFE	35	9	11	15
FDX	34	9	8	17	PG	28	12	9	7
GD	30	11	4	15	PM	29	7	7	15
GE	71	19	15	37	QCOM	38	6	19	13
GILD	26	11	7	8	RF	24	8	6	10
GOOG	31	10	8	13	RTN	35	11	5	19
GS	28	5	6	17	SLB	41	19	12	10
HAL	36	11	12	13	SO	23	5	8	10
HD	36	13	7	16	T	45	12	15	18
HON	25	11	7	7	TGT	46	23	6	17
HPQ	23	10	6	7	TXN	42	10	7	25
IBM	26	11	6	9	UPS	38	14	8	16
INTC	48	18	14	16	USB	38	12	11	15
JCJ	45	18	10	17	UTX	33	17	9	7
JPM	34	10	8	16	VZ	34	13	10	11
KHC	38	16	5	17	WBA	41	13	9	19
KO	26	10	10	6	WMB	16	7	6	3
LMT	34	10	5	19	WMT	30	8	11	11
MA	47	23	5	19	XOM	39	14	10	15

Tabla 5

En total hay 1838 noticias, de las cuales 653 son positivas, aproximadamente un 35% del total, 429 son neutras, aproximadamente un 24%, y 746 son negativas, aproximadamente un 41%.

Para saber que empresas han tenido más noticias positivas o negativas, se ha hecho un contraste de hipótesis de deferencia de proporciones entre cada una de las empresas con respecto a la proporción total. Los resultados se incluyen a continuación en la Tabla 6, Tabla 7, Tabla 8, y Tabla 9.

Mas noticias positivas		
Empresas	Proporción	p-valor
AVP	0,52	0,03947
CVX	0,56	0,00358
MA	0,49	0,02920
TGT	0,50	0,02163
UTX	0,52	0,02881

Tabla 6

En la Tabla 6 aparecen las empresas que tienen una mayor proporción de noticias positivas que la media con un intervalo de confianza superior al 95%. Estas empresas son: Avon (AVP), empresa centrada en la industria de cosméticos, Chevron (CVX), petrolera estadounidense, Mastercard (MA), multinacional de servicios financieros, Target (TGT), cadena de grandes almacenes, y United Technologies Corporation (UTX), conglomerado multinacional que se dedica al desarrollo de tecnología.

Estas cinco compañías son de ámbitos diversos, por lo tanto, no se puede decir que haya un mismo sector cuyas empresas destaquen por la gran cantidad de noticias positivas en comparación con el resto de los sectores.

Menos noticias positivas		
Empresas	Proporción	p-valor
GS	0,18	0,02605
QCOM	0,16	0,00582

Tabla 7

En la Tabla 7 aparecen las empresas que tienen una menor proporción de noticias positivas que la media con un intervalo de confianza superior al 95%. Estas empresas son Goldman Sachs (GS), dedicados a la banca de inversión y de valores, y Qualcomm (QCOM), centrados en equipos de telecomunicaciones.

Estas dos compañías son de ámbitos diversos, por lo tanto, no se puede decir que haya un mismo sector cuyas empresas destaquen por la pequeña cantidad de noticias positivas en comparación con el resto de los sectores.

Estas son las empresas con menor proporción de noticias positivas, y por tanto, las empresas que en proporción han recibido menos publicidad positiva por parte de los medios.

Más noticias negativas		
Empresas	Proporción	p-valor
BA	0,57	0,01016
BAX	0,59	0,02504
GE	0,52	0,02635
GS	0,61	0,01578
LMT	0,56	0,03612
TXN	0,60	0,00681

Tabla 8

En la Tabla 8 aparecen las empresas que tienen una mayor proporción de noticias negativas que la media con un intervalo de confianza superior al 95%. Estas empresas son: Boeing (BA), centrada en la industria aeroespacial y defensa, Baxter International (BAX), empresa de la industria del equipamiento médico, General Electric (GE), conglomerado multinacional de medios de comunicación, servicios financieros, e infraestructura, Goldman Sachs (Gs), Lockheed Martin (LMT), centrada en la industria aeroespacial y militar, y Texas Instruments (TXN), centrada en la industria tecnológica.

Estas son las empresas con mayor proporción de noticias negativas, y, por tanto, las empresas que en proporción han recibido más publicidad negativa por parte de los medios.

Menos noticias negativas		
Empresas	Proporción	p-valor
KO	0,23	0,03536
PEP	0,17	0,00253
PG	0,25	0,04760
SLB	0,24	0,01823
UTX	0,21	0,01224
WMB	0,19	0,03816

Tabla 9

En la Tabla 9 aparecen las empresas que tienen una menos proporción de noticias negativas que la media con un intervalo de confianza superior al 95%. Estas empresas son Coca-Cola (KO), multinacional de bebidas, PepsiCo (PEP), multinacional de bebidas, Procter & Gamble (PG), multinacional de bienes de consumo, Schlumberger ltd (SLB), servicios a yacimientos petroleros, United Technologies (UTX), Williams Companies (WMB), compañía energética.

Estas son las empresas con menor proporción de noticias negativas, y por tanto, las empresas que en proporción han recibido menos publicidad negativa por parte de los medios.

United Technologies (UTX) es la única empresa con una proporción significativamente superior a la media en cuanto a noticias positivas, pero a la vez tiene una proporción significativamente inferior a la media en cuanto a noticias negativas. Lo cual implica que ha tenido buena publicidad por parte de los medios durante el periodo analizado.

En el otro lado del espectro se encuentra Goldman Sachs (GS). Es la única empresa que cuenta con una proporción significativamente inferior en cuanto a noticias positivas, pero a la vez tiene una proporción significativamente superior a la media en cuanto a noticias negativas, lo cual implica que ha tenido mala publicidad durante este periodo, y posiblemente un mal desempeño.

A continuación, se presentan las noticias segmentadas por trimestres y años en la Tabla 10

Periodo de tiempo	Noticias totales	Noticias positivas	Noticias neutras	Noticias negativas
2016	181	70	46	65
Trim.3	15	7	5	3
Trim.4	166	63	41	62
2017	483	180	109	194
Trim.1	94	30	22	42
Trim.2	112	39	24	49
Trim.3	121	46	30	45
Trim.4	156	65	33	58
2018	737	263	168	306
Trim.1	207	84	54	69
Trim.2	200	69	52	79
Trim.3	128	52	25	51
Trim.4	202	58	37	107
2019	437	140	116	181
Trim.1	145	45	36	64
Trim.2	147	46	36	65
Trim.3	136	48	41	47
Trim.4	9	1	3	5
Total	1838	653	439	746

Tabla 10

Se ha hecho un contraste de diferencias de proporciones entre los periodos observados y el total. Al analizar los datos, no se ha encontrado ningún periodo con una proporción significativamente superior en cuanto a noticias positivas se refiere al 95% de intervalo de confianza, pero si que se ha encontrado que, en el cuarto trimestre del año 2018, la proporción de noticias positivas han sido significativamente más bajas que la media con un p-valor de 0'0268. Por otro lado, en ese mismo trimestre se han publicado una proporción significativamente superior de noticias negativas con respecto a la media con un p-valor de 0'000353, significativo incluso a un intervalo de confianza del 99%. Con pocas noticias positivas y muchas negativas, se podría pensar que el cuarto trimestre fue un periodo de desaceleración del S&P500.

Por otro lado, únicamente en el primer trimestre de 2018 se ha encontrado un periodo con una proporción significativamente inferior de noticias negativas con respecto a la media con un p-valor de 0'0216.

Una característica que se aborda en este trabajo de fin de grado es el verdadero valor sentimental de las noticias. Lo lógico es pensar que después de una noticia positiva, el valor de la acción va a subir, pero no siempre es así. Esto suele deberse a que los grandes inversores y los fondos de cobertura, también conocidos como *hedge funds*, tienen más información que el resto de los inversores minoritarios, por lo cual, una noticia que aparentemente es positiva para el público, para los inversores mayoritarios puede ser negativa. Por este motivo, parte de la herramienta desarrollada se ha centrado en obtener el verdadero valor sentimental con dos métodos estadísticos diferentes, los cuales serán mostrados en el capítulo siguiente.

Capítulo 4

Herramienta desarrollada

Introducción

Este Trabajo de Fin de Grado se ha realizado con la ayuda de tres programas informáticos: la app de Bloomberg, la cual ha proporcionado todas las noticias y los datos financieros, Excel, ha guardado los datos y se ha utilizado para hacer el banco de palabras, y, por último, MATLAB R2018a. Para todos los programas se han utilizado licencias proporcionadas por ICAI – Universidad Pontificia de Comillas.

El código se ha querido desarrollar para que el usuario pueda ver los datos de 54 empresas distintas en el periodo desde el 24/08/2016 hasta el 15/10/2019.

EXCEL

Excel es el programa en el cual están almacenado todos los datos de las empresas. En el caso que el usuario prefiera otras fechas u otras empresas, se explicaran los pasos a seguir para guardar los datos y poder ejecutar MATLAB correctamente.

En debería guardar los datos en un Excel con el formato de la Figura 3 que se muestra a continuación.

Date	Price	Returns	Volume	EPS	News	Sentiment value	Total words	Positive words	Negativa words
30/03/2017	52,25	0,00519431	2530953						
31/03/2017	51,86	-0,00746411	2198880	0,5					
03/04/2017	52,29	0,00829155	2655105						
04/04/2017	52,38	0,00172117	2817434						
05/04/2017	52,55	0,00324551	2839784						
06/04/2017	52,53	-0,00038059	2552893						
07/04/2017	52,79	0,00494955	3417469						
10/04/2017	52,84	0,00094715	4972033						
11/04/2017	52,97	0,00246026	2774515						
12/04/2017	53,53	0,01057202	3741292						
13/04/2017	53,25	-0,00523071	3923786						
17/04/2017	52,82	-0,00807512	3764633						
18/04/2017	52,73	-0,0017039	2933651						
19/04/2017	52,71	-0,00037929	4242968						
20/04/2017	52,89	0,00341491	2998283						
21/04/2017	52,88	-0,00018907	3217472						
24/04/2017	53,45	0,01077912	2664036						
25/04/2017	53,21	-0,00449018	4410491						
26/04/2017	55,59	0,04472843	7921690		BAX boosts FY Adj. EPS View, tops Est.	0,25	8	2	0
27/04/2017	55,42	-0,0030581	6411321						
28/04/2017	55,68	0,00469145	3456634						

Figura 3. Muestra de los datos obtenidos de Baxter International Inc.

Cada categoría en una columna diferente, siguiendo este orden empezando por la primera columna: fechas, cotización, volumen, variación diaria, dividendos, el titular de la noticia, valor sentimental de la noticia, número de palabras en la noticia, número de palabras positivas en la noticia, y número de palabras negativas en la noticia. La fila azul con los nombres de las categorías de la Figura 3 no se debe poner, simplemente se ha incluido con fines ilustrativos.

En el caso que el formato ya esté preparado pero el usuario no tenga los datos, se recomienda el uso de Bloomberg, el cual tiene un complemento para Excel llamado BQL (Bloomberg Query Language) Builder, el cual nos permite obtener las fechas, la cotización, el volumen, la variación diaria, y dividendos de Bloomberg de múltiples empresas a la vez.

BQL Builder no es capaz de insertar los titulares de las noticias, y no se encontró ninguna forma alternativa de obtener los titulares de las noticias de forma eficiente, por lo tanto, se han recogido manualmente. Este problema ha sido desarrollado más en profundidad en el capítulo 5 como una futura extensión.

Una vez obtenidos las noticias, se calcula su valor sentimental en un documento Excel. En una hoja de Excel se encuentran los bancos de palabras positivos y negativos ya explicados en el capítulo anterior. Por otro lado, se añaden las noticias en dos hojas distintas, en una hoja se buscan las palabras positivas de las noticias y en la otra hoja se buscan las palabras negativas de las noticias. Para encontrar las palabras, se utiliza la función “Encontrar”, la cual busca en la celda correspondiente si encuentra exactamente la secuencia de caracteres que se ha introducido y nos devuelve su posición. En este caso,

se han introducido todas las raíces de palabras positivas y negativas. Al buscar secuencias de letras, se han introducido las raíces de las palabras para tener toda la familia y así reducir el número de palabras en los bancos. Una vez encontrada la posición de las palabras, se utiliza la función “Contar” para saber cuántas raíces han sido encontradas y así poder determinar el número de palabras con carácter positivo y negativo.

Debido al gran tamaño del archivo Excel, se incluye una muestra del Excel a continuación en la Figura 4 y en la Figura 5

Fechas	Noticia	Compañía	affect	alert	anti-Psychotic	anxiety	apathy	arrest
16/08/2018	CSCO systems gets a standing ovation from analysts on growth	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
15/11/2018	CSCO price targets lifted as outlook reassures wall street	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
06/12/2018	Huawei CFO arrest could have US chip implications	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	12
07/12/2018	CSCO's errant China-Travel alert shows tech's Huawei alarm	Cisco	#¡VALOR!	28	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
11/12/2018	CSCO is said to be in talks to buy optical chip maker luxtera	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
14/12/2018	CSCO cut at Nomura, 2019 could reveal 'Imperfections' in story	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
14/02/2019	CSCO's breadth of strength prompts analysts to boost targets	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
16/05/2019	CSCO gains as analysts say outlook offsets trade uncertainty	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!
12/06/2019	CSCO slips from multi-year High after blair downgrade	Cisco	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!	#¡VALOR!

Figura 4. Muestra del Excel que busca las palabras con carácter negativo

Fechas	Noticia	Compañía	above	add	allure	ally	applaud	approv	attract	beat
25/07/2018	GD 2Q revenue 1,7% above est.	General Dynamics	20	#¡VALOR!						
24/10/2018	GD 3Q EPS from continuing Operations beats est.	General Dynamics	#¡VALOR!	38						
10/12/2018	Defense stocks prancing a 'wall of worry', Goldman Sachs says	General Dynamics	#¡VALOR!							
17/12/2018	Arms maker warns of cost with Trudeau looking to end Saudi deal	General Dynamics	#¡VALOR!							
21/12/2018	Defense contractors sell off as shutdown looms, Mattis resigns	General Dynamics	#¡VALOR!							
30/01/2019	GD 2019 revenue view below est.; shares down	General Dynamics	#¡VALOR!							
27/03/2019	Navy's \$7,8B destroyer due for delivery 5 years late	General Dynamics	#¡VALOR!							
24/04/2019	GD 1Q EPS beats highest est.	General Dynamics	#¡VALOR!	11						
20/05/2019	GD raised as Goldman sees valuation near through	General Dynamics	#¡VALOR!							

Figura 5. Muestra del Excel que busca las palabras con carácter positivo

En la primera columna de la Figuras 4 y la Figura 5 se encuentra la fecha de la noticia, la segunda columna muestra el titular de la noticia, la tercera columna muestra el nombre de la compañía y las siguientes columnas, llamada conjunto de búsqueda, se utilizan para comprobar si hay alguna palabra con valor sentimental en los titulares. Cada celda del conjunto de búsqueda busca la palabra que le corresponde a su columna en la noticia que comparte su misma fila. Por ejemplo, en la primera fila, cuarta columna de la Figura 5, aparece el número 20. La cuarta columna corresponde con la palabra ‘above’, que se traduce en español a ‘por encima’, y el 20 indica que la palabra empieza en el carácter número 20 del titular de la noticia (Los espacios y signos de puntuación se consideran caracteres). Por otro lado, las celdas con ‘#¡VALOR!’ significa que no se ha encontrado la palabra de su columna en el titular de la noticia.

En la Figura 4, en la fila de la segunda columna se puede encontrar la palabra ‘downgrade’, la cual se traduce como ‘degradar’, y aunque se ha considerado como una palabra negativa, no aparece la columna ‘downgrade’ en la Figura 4. Esto se debe a que la Figura 4 es una pequeña muestra para ilustrar al lector, y lo mismo sucede en la Figura 5.

Una vez asignados todos los valores sentimentales, se ha desarrollado una macro para juntar los datos obtenidos con el BQL y los datos analizados en el banco de palabras, de tal forma que el resultado final el representado en la Figura 6.

	A	B	C	D	E	F	G	H	I	J	K
1	42606	102,2	0,00511408	5178275							CVX
2	42607	101,89	-0,00303327	4490051							
3	42608	101,32	-0,00559427	4776217							
4	42611	102,05	0,0072049	4026893							
5	42612	101,7	-0,00342969	3712813							
6	42613	100,58	-0,01101278	6249070							
7	42614	100,21	-0,00367866	6044999							
8	42615	100,93	0,00718491	4676603							
9	42619	102,43	0,01486179	5428985							
10	42620	102,88	0,00439324	5284271							
11	42621	104,12	0,01205288	5788784							
12	42622	101,27	-0,02737226	7808659							
13	42625	102,25	0,0096771	8149060							
14	42626	99,43	-0,02757946	9300364							
15	42627	98,42	-0,0101579	8201430							
16	42628	99,5	0,01097338	6755332							
17	42629	97,84	-0,01668342	12725620							
18	42632	98,04	0,00204415	6182037							
19	42633	97,7	-0,00346797	5771895							
20	42634	99,63	0,01975435	6361645							
21	42635	99,98	0,003513	5416740							
22	42636	99,22	-0,00760152	7396784							
23	42639	98,78	-0,00443459	5781775							
24	42640	98,98	0,0020247	6476099							
25	42641	102,15	0,03202667	10190924		energy space	0,2	10	2	0	

Figura 6. Ejemplo de inserción de datos

La Figura 6 muestra como están los datos directamente preparados para ser procesados. Las columnas siguen la estructura que había en la Figura 3. La primera columna, es la fecha en formato 'General'. Excel asigna a cada fecha un valor numérico siendo el valor 1 el 01/01/1900, el 2 está asignado al 02/01/1900, el 3 está asignado al 03/01/1900 y así sucesivamente. Matlab tiene la misma forma de procesar las fechas, pero con distinto origen, es por ello por lo que se ha tenido que hacer una conversión en Matlab para que cuadren las fechas.

MATLAB

Se han desarrollado varias funciones con el objetivo de proporcionar al usuario un análisis de una empresa seleccionada por el usuario. La herramienta es capaz de filtrar empresas, filtrar por fechas, hacer diagramas de cajas, medias móviles, representaciones gráficas del precio o del volumen con noticias, crea modelos predictivos y divide la cotización en series temporales.

Filtro

Este programa ha sido pensado para facilitar al usuario el análisis de los datos, por lo tanto, se ha añadido un filtro para seleccionar la empresa y otro filtro para seleccionar las fechas del análisis, los cuales han sido realizados con condicionales.

En primer lugar, el usuario debe seleccionar la empresa que quiere analizar tal y como se muestra a continuación en la Figura 7, y en la Tabla 11 se encuentran el símbolo de la Figura 7, con su nombre y su sector

Para seleccionar una empresa, tan solo escriba el número que tiene al lado

1-AVN
2-AXP
3-BA
4-BAX
5-CSCO
6-CVX
7-DVN
8-F
9-FCX
10-FDX
11-GD
12-GE
13-GILD
14-GOOG
15-GS
16-HAL
17-HD
18-HON
19-HPQ
20-IBM
21-INTC
22-JNJ
23-JPM
24-KHC
30-MMM
31-MRK
32-MSFT
33-NKE
34-ORCL
35-PEP
36-PFE
37-PG
38-PM
39-QCOM
40-RF
41-RTN
42-SLB
43-SO
44-T
45-TGT
46-TXN
47-UPS
48-USB
49-UTX
50-VZ
51-WBA
52-WMB
53-WMT
54-XOM

¿Qué empresa desea ver? |

Figura 7. Ejemplo de visualización de empresas

Símbolo	Nombre	Sector
AVP	Avon Products	Cosméticos
AXP	American Express	Servicios financieros
BA	Boeing	Industria aérea y espacial
BAX	Baxter International	Farmacéutica, tecnología médica
CSCO	Cisco Systems	Telecomunicaciones
CVX	Chevron Corporation	Petróleo y gas
DVN	Devon Energy	Petróleo y gas
EMC	EMC Corporation	Fabricante de electrónica
F	Ford	Automoción
FCX	Freeport-McMoRan	Minería
FDX	FedEx	Logística
GD	General Dynamics	Armamento
GE	General Electric	Variado
GILD	Gilead Sciences	Farmacéutica, biotecnología
GOOG	Google	Servicios de Internet
GS	Goldman Sachs	Banca
HAL	Halliburton	Petróleo y gas
HD	Home Depot	Venta al por menor de materiales de construcción
HON	Honeywell International	Variados
HPQ	Hewlett-Packard	Fabricante de electrónica
IBM	IBM	IT
INTC	Intel	Microelectrónica
JNJ	Johnson & Johnson	Farmacéutica, bienes de consumo
JPM	JP Morgan Chase & Co.	Banca
KHC	Kraft Heinz Company	Alimentación
KO	Coca-Cola Company	Alimentación
LMT	Lockheed Martin	Industria aérea y espacial
MA	MasterCard	Servicios financieros
MCD	McDonald's	Gastronomía
MDT	Medtronic	Tecnología médica
MMM	3M	Diversos
MRK	Merck & Co.	Farmacéutica
MSFT	Microsoft	Software
NKE	Nike	Equipación deportiva
ORCL	Oracle	Software

PEP	PepsiCo	Alimentación
PFE	Pfizer	Farmacéutica
PG	Procter & Gamble	Bienes de consumo
PM	Philip Morris International	Tabaco
QCOM	Qualcomm	Telecomunicaciones
RF	Regions Financial	Banca
RTN	Raytheon	Armamento
SLB	Schlumberger	Petróleo y gas
SO	Southern Company	Energía
T	AT&T	Telecomunicaciones
TGT	Target Corporation	Venta al por menor
TXN	Texas Instruments	Fabricante de electrónica
UPS	United Parcel Service	Logística
USB	U.S. Bancorp	Banca
UTX	United Technologies Corporation	Tecnología
VZ	Verizon Communications	Telecomunicaciones
WAG	Walgreens	Venta al por menor
WMB	Williams Companies	Petróleo y gas
WMT	Wal-Mart	Venta al por menor
XOM	ExxonMobil	Petróleo y gas

Tabla 11

Tal y como se puede ver en la primera línea de la Figura 7, el usuario debe insertar el número para seleccionar la empresa que quiere analizar. En caso de que el usuario se equivoque e inserte un número que no corresponda a ninguna empresa, el programa lo detecta y pide al usuario que lo vuelva a insertar, tal y como se muestra a continuación en la Figura 8.

```
53-WMT
54-XOM
¿Qué empresa desea ver? 55
Por favor, seleccione un valor numérico disponible
Para seleccionar una empresa, tan solo escriba el número que tiene al lado
1-AVN
2-AXP
```

Figura 8. Verificación número de empresa

En el caso de la Figura 8, al seleccionar la empresa 55, la cual no está disponible, aparece un mensaje de error y pide al usuario que seleccione un valor disponible.

Después de seleccionar la empresa, el usuario debe seleccionar el periodo de tiempo que le gustaría que el programa analizase. Los datos recogidos son de un periodo comprendido entre el 24/08/2016 y el 12/10/2019, pudiendo analizar todo el periodo o simplemente una parte, como se puede ver en la Figura 9.

```
¿Quieres fijarte en un periodo en concreto o prefieres ver todos los datos disponibles?
Selecciona la opción numérica que precises
1-filtrar
2-todo
Quiero |
```

Figura 9. Filtro del periodo

Al igual que para seleccionar una empresa, el usuario debe poner el número adyacente a la opción que quiere seleccionar. En caso de que el usuario se equivoque e inserte un número que no corresponda a ninguna opción, el programa lo detecta y pide al usuario que lo vuelva a insertar, tal y como se muestra a continuación en la Figura 10.

```
¿Quieres fijarte en un periodo en concreto o prefieres ver todos los datos disponibles?  
Selecciona la opción numérica que precises  
1-filtrar  
2-todo  
Quiero 3  
Por favor, seleccione una opción disponible  
¿Quieres fijarte en un periodo en concreto o prefieres ver todos los datos disponibles?  
Selecciona la opción numérica que precises  
1-filtrar  
2-todo  
quiero |
```

Figura 10. Verificación del filtro del periodo

En este caso, al seleccionar la opción 3, la cual no está disponible, aparece un mensaje de error y pide al usuario que seleccione un valor disponible.

En el caso que el usuario quiera filtrar por fechas, debe seguir paso a paso las instrucciones que da el programa. En primer lugar, debe introducir el día de inicio del periodo, seguido del mes de inicio del periodo, y por último el año de inicio del periodo. Después, el usuario debe introducir la fecha final tal y como ha escrito la fecha inicial. A continuación, la Figura 11 muestra el procedimiento.

```
¿Quieres fijarte en un periodo en concreto o prefieres ver todos los datos disponibles?  
Selecciona la opción numérica que precises  
1-filtrar  
2-todo  
Quiero 1  
¿Desde qué día?  
31  
Ha seleccionado el día 31  
¿Desde qué mes?  
1.enero 3.marzo  
5.mayo 7.julio 8.agosto  
10.octubre 12.diciembre  
5  
Seleccione un año:  
2017, 2018, 2019  
2017  
La fecha inicial es el 31/5/2017  
¿Es correcta la fecha?  
1. Sí, es correcta 2. No, quiero cambiar  
Pon el valor numérico  
1  
¿Hasta qué día?  
30  
Ha seleccionado el día 30  
¿Hasta qué mes?  
1.enero 3.marzo 4.abril  
5.mayo 6.junio 7.julio 8.agosto  
9.septiembre 10.octubre 11.noviembre 12.diciembre  
11
```

```
Seleccione un año:
2016, 2017, 2018
2018
la fecha final es el 30/11/2018
¿Es correcta la fecha?
1. Si, es correcta 2. No, quiero cambiar
Pon el valor numérico
1
```

Figura 11. Filtro por fecha

La Figura 11 muestra el procedimiento para el filtrado de datos por fechas. El periodo filtrado empieza el 31/05/2017 y termina el 30/11/2018. El día de la fecha inicial es el 31, el programa lo procesa y las opciones que devuelve son únicamente los meses con 31 días. El mes de la fecha inicial es mayo, el programa lo procesa y las opciones que nos devuelve para elegir el año no incluye 2016. Esto se debe a que los datos se empezaron a recoger el 24/08/2016, por lo tanto, no hay datos de mayo en 2016. Por otro lado, el día de la fecha final es el 30, por lo cual, febrero es el único año que no se puede elegir. El mes final escogido es noviembre, por lo tanto, el programa lo procesa y las opciones que nos devuelve para elegir el año no incluye 2019. Esto se debe a que los datos se recogieron hasta el 15/10/2019, por lo tanto, no hay datos de noviembre de 2019.

Si el usuario se equivocara al introducir una fecha, tiene la opción de volver a escribir la fecha sin reiniciar. Esto se debe a que el programa pide la confirmación del usuario al seleccionar la fecha, como se puede ver en la Figura 11.

Además, se ha integrado un proceso por el cual el programa identifica si la fecha final es anterior a la fecha inicial, ya que un periodo negativo no tendría sentido.

```
Seleccione un año:
2017, 2018, 2019
2017
la fecha final es el 15/6/2017
La fecha final debe ser posterior a la fecha inicial
Por favor, vuelva a seleccionar la fecha final
¿Hasta qué día?
```

¿Quieres fijarte en un periodo en concreto o prefieres ver todos los datos disponibles?
Selecciona la opción numérica que precises
1-filtrar
2-todo
Quiero 1
¿Desde qué día?
15
Ha seleccionado el día 15
¿Desde qué mes?
1.enero 2.febrero 3.marzo 4.abril
5.mayo 6.junio 7.julio 8.agosto
9.septiembre 10.octubre 11.noviembre 12.diciembre
6
Seleccione un año:
2017, 2018, 2019
2018
La fecha inicial es el 15/6/2018
¿Es correcta la fecha?
1. Sí, es correcta 2. No, quiero cambiar
Pon el valor numérico
1
¿Hasta qué día?
15
Ha seleccionado el día 15
¿Hasta qué mes?
1.enero 2.febrero 3.marzo 4.abril
5.mayo 6.junio 7.julio 8.agosto
9.septiembre 10.octubre 11.noviembre 12.diciembre
6

Figura 12. Comprobación fecha final

La Figura 12 muestra el caso en el cual el usuario ha seleccionado una fecha inicial que es posterior a la fecha final, la fecha inicial es el 15/06/2018 y la fecha final que ha intentado seleccionar es el 15/06/2017. Al no ser compatibles, el programa directamente pide al usuario que cambie la fecha final.

Conversión de noticias

Para mostrar al usuario las noticias que han sucedido a lo largo del periodo seleccionado, se ha creado un código de colores para clasificar las noticias, tal clasificación aparece a continuación en la Tabla 12.

TIPO DE NOTICIA	VALOR SENTIMENTAL (VS)	COLOR
MUY POSITIVA	$vs \geq 0'25$	Cian
POSITIVA	$0 < vs < 0'25$	Verde
NEUTRA	$vs = 0$	Amarillo
NEGATIVA	$-0'25 < vs < 0$	Rojo
MUY NEGATIVA	$vs \leq -0'25$	Negro

Tabla 12

Diagrama de cajas

El diagrama de caja es una forma de representar grupos numéricos a través de sus cuartiles. Se han representado los valores sentimentales de las noticias de una empresa y la variación diaria para que el usuario pueda hacerse una idea del tipo de noticias que la empresa ha protagonizado y su respuesta habitual en la cotización.

Valor sentimental

En primer lugar, el diagrama de caja de los valores sentimentales de las noticias ayuda al usuario a hacerse una idea rápida de la magnitud de las noticias que ha protagonizado esa empresa en el periodo seleccionado.

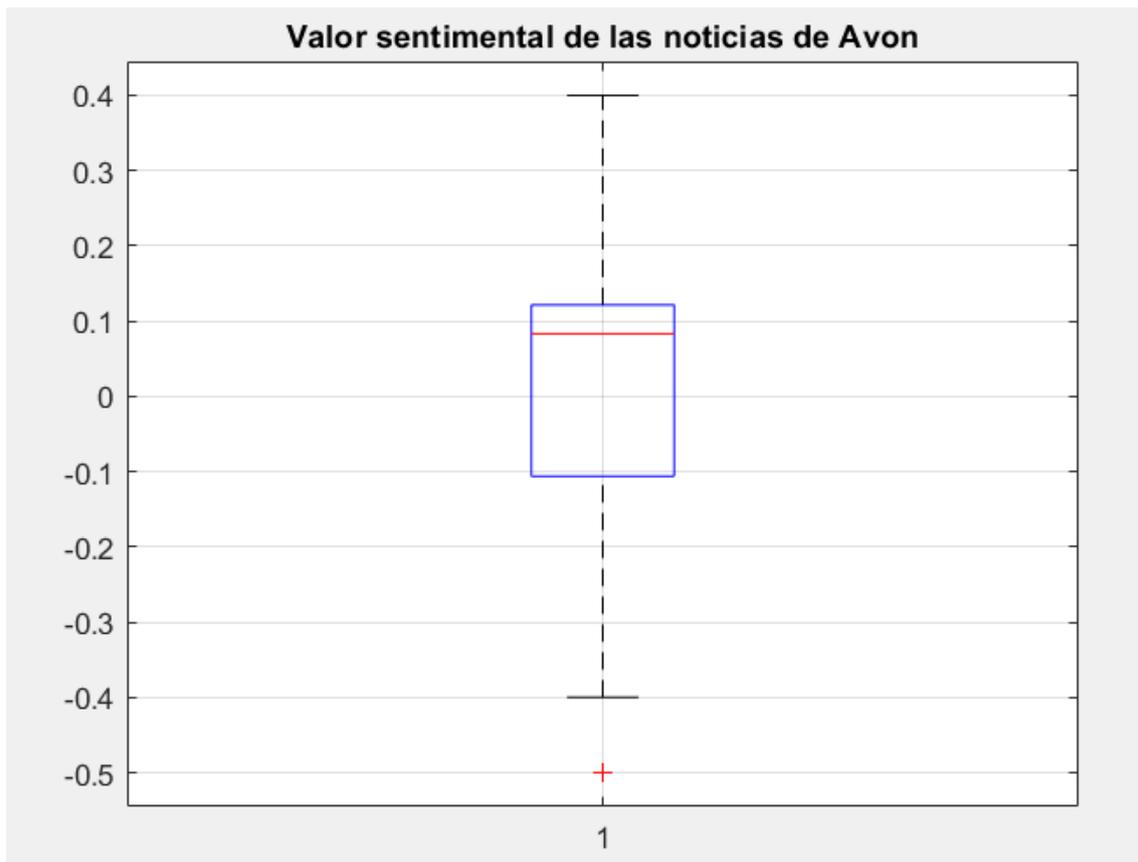


Figura 13. Diagrama de cajas de VS de Avon

La Figura 13 muestra el diagrama de caja del valor sentimental de las noticias de Avon, el cual es un ejemplo de una empresa con un mayor número de noticias positivas que negativas. La línea roja dentro del rectángulo azul representa la mediana, la cual se sitúa en 0'083, lo que significa que más de la mitad de las noticias son positivas. Además, la mediana está más alejada del 0 que del tercer cuartil, representado por el lado superior del rectángulo azul, lo cual indica que el 25% de todas las noticias tienen un valor sentimental muy similar, entre 0'083 y 0'121. Por otro lado, aparece un 'outlier' o valor atípico de una noticia con valor sentimental -0'5, lo que indica que hubo una noticia muy negativa que no corresponde con el valor sentimental del resto de noticias.

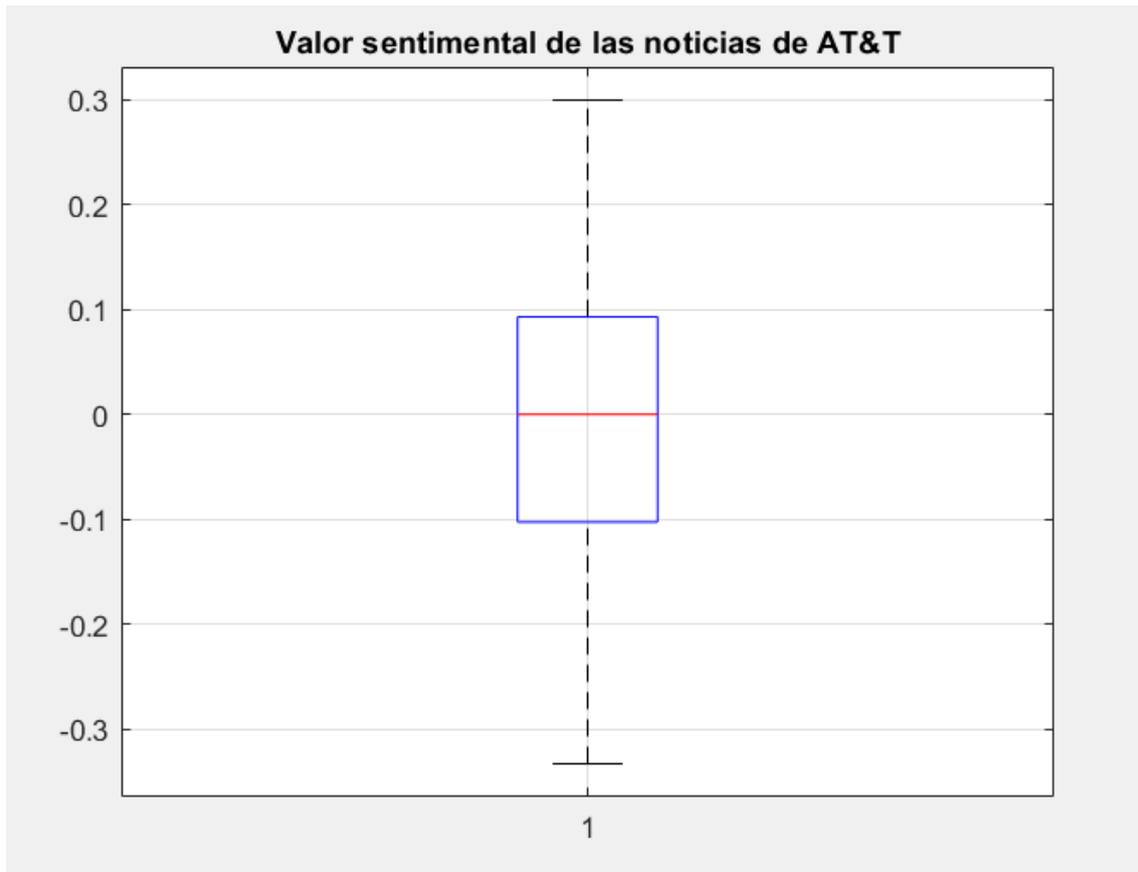


Figura 14. Diagrama de cajas de VS de AT&T

La Figura 14 muestra el diagrama de caja del valor sentimental de las noticias de AT&T, el cual es un ejemplo de una empresa en el que las noticias positivas compensan las noticias negativas. La mediana se encuentra en el 0, no existen valores atípicos y el valor absoluto del primer cuartil tiene un valor similar al tercer cuartil.

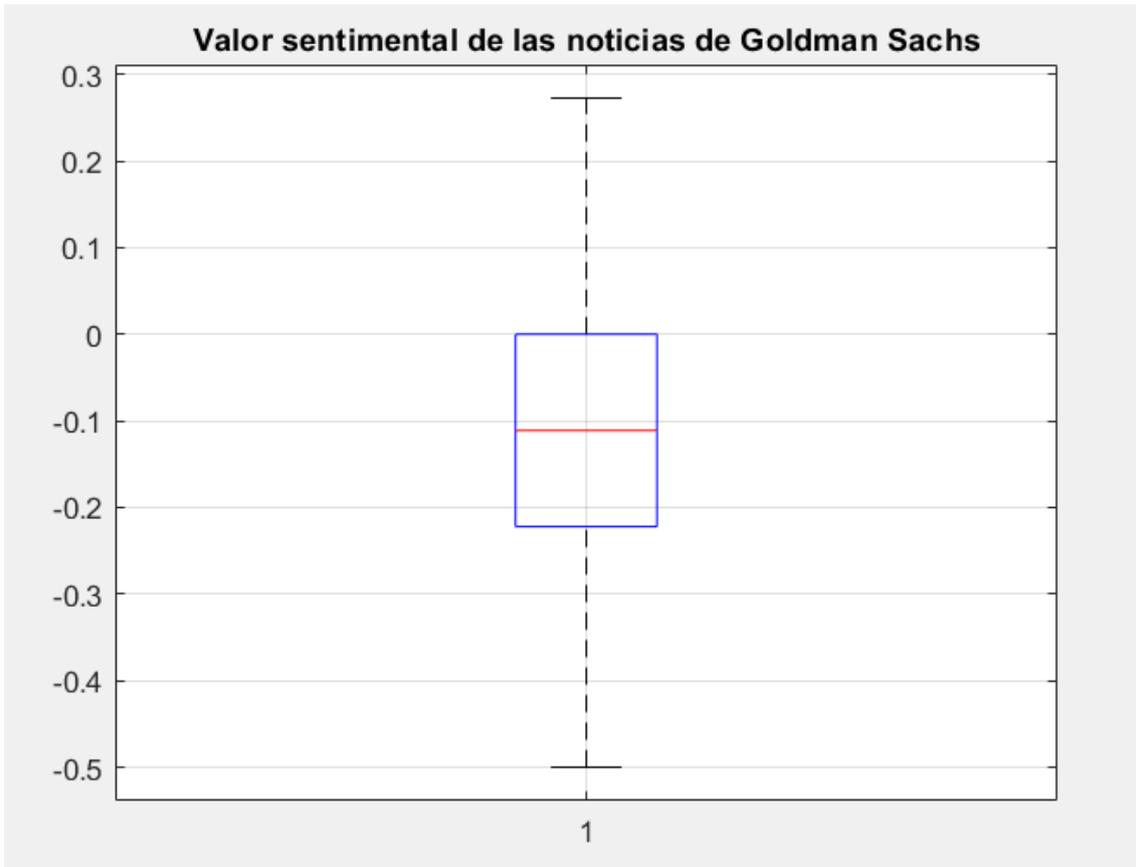


Figura 15. Diagrama de cajas de VS de Goldman Sachs

La Figura 15 muestra el diagrama de caja del valor sentimental de las noticias de Goldman Sachs, el cual es un ejemplo de una empresa con un mayor número de noticias negativas que positivas. La mediana se encuentra en el -0.11 y el tercer cuartil en el 0. Esto significa que más de la mitad de las noticias son negativas y menos del 25% del total de las noticias son positivas. Esto indica que Goldman Sachs no ha tenido buena publicidad durante el periodo.

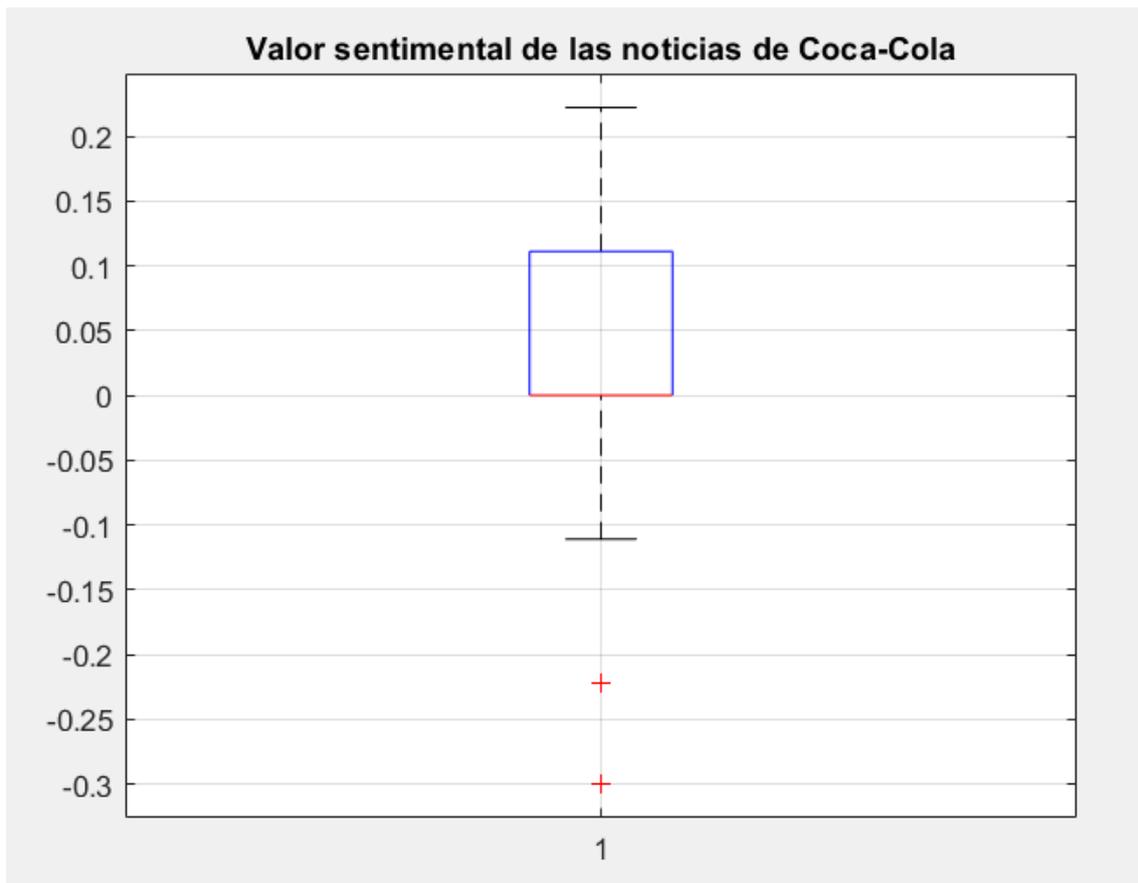


Figura 16. Diagrama de cajas de Coca Cola

La Figura 16 muestra el diagrama de caja del valor sentimental de las noticias de Coca-Cola, el cual es un caso atípico ya que el primer cuartil coincide con la mediana en 0. Esto se debe a la gran cantidad de noticias neutras que ha tenido durante el periodo analizado. Digno de mención son sus dos valores atípicos en -0'22 y -0'3.

Variación diaria

En primer lugar, el diagrama de caja de la variación diaria ayuda al usuario a hacerse una idea rápida de la volatilidad en la cotización de una empresa en el periodo seleccionado. Estos diagramas de cajas tienen más valores atípicos que los diagramas de cajas con el valor sentimental de las noticias.

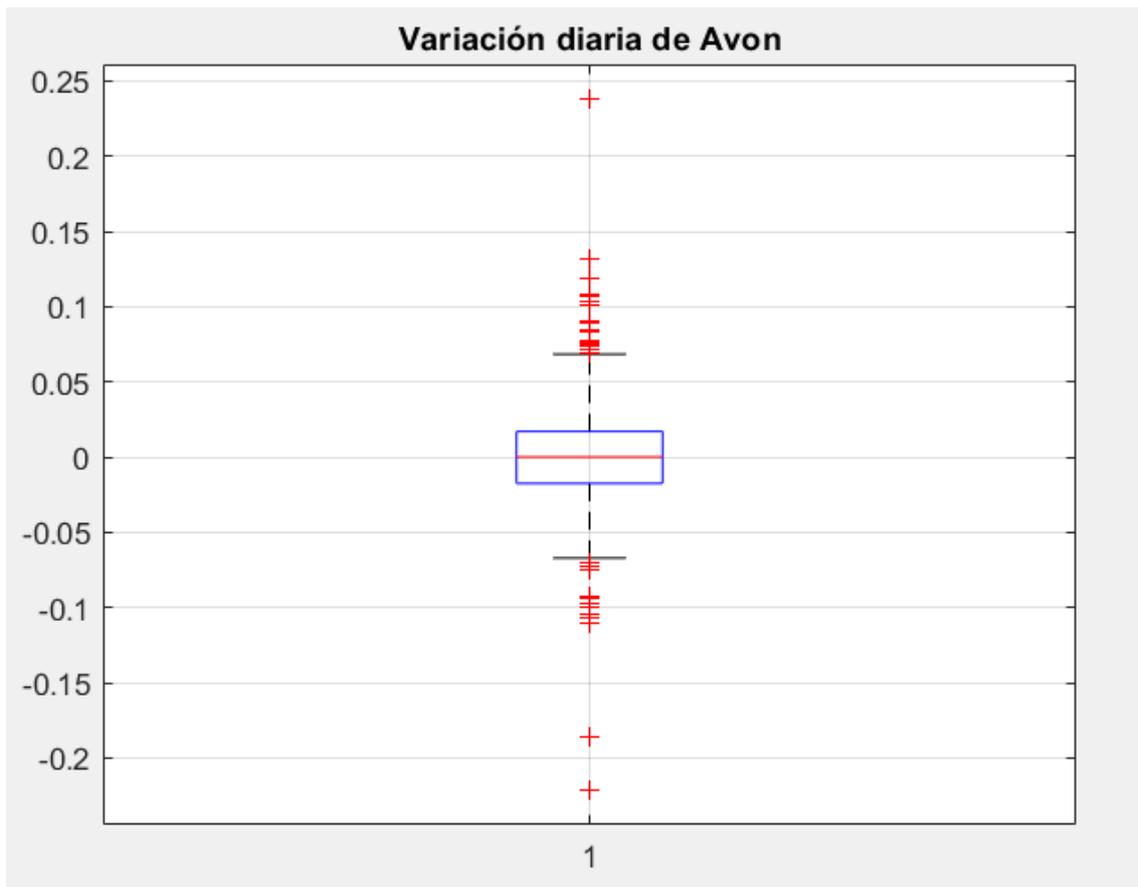


Figura 17. Diagrama de cajas de la variación diaria de Avon

La Figura 17 muestra el diagrama de caja de la variación diaria de Avon. Como en la gran mayoría de empresas analizadas, primer y tercer cuartil están dentro del rango $[-0'02, 0'02]$, en este caso el primer cuartil se sitúa en $-0'0175$ y el tercer cuartil en $0'0170$. Esto indica que el 50% de las variaciones diarias están entre $-1'75\%$ y $1'70\%$. Por otro lado, en el caso de la Figura 17, se pueden observar algunos valores atípicos muy alejados del primer y tercer cuartil, algunos incluso superando el $\pm 20\%$ de variación diaria ($6'3\% < 10 * Q_3$; $-20\% < 10 * Q_1$). Esto indica que, aunque generalmente Avon es bastante estable, puede tener días muy volátiles. Para concluir con la Figura 17, la mediana se sitúa en 0, lo cual indica que durante el periodo analizado, no hay un crecimiento ni decrecimiento notable por parte de Avon.

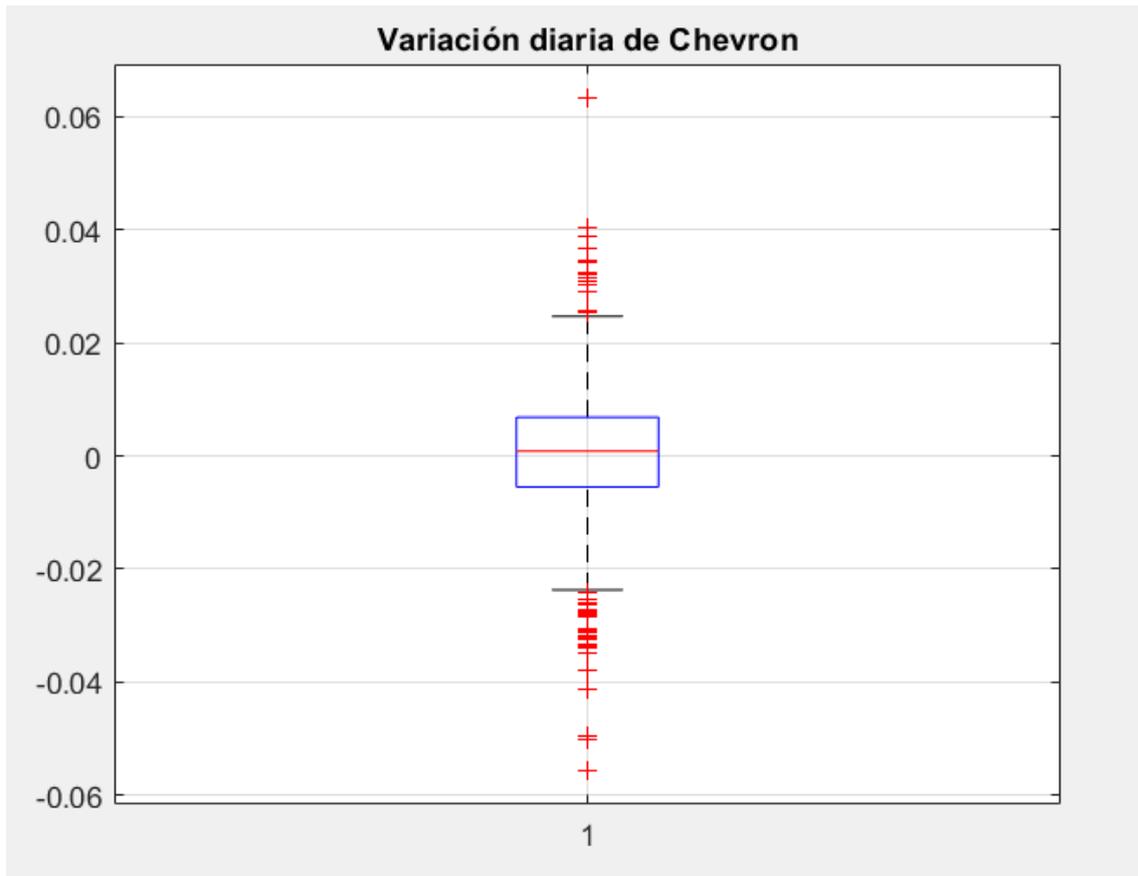


Figura 18. Diagrama de cajas de la variación diaria de Chevron

La Figura 18 muestra el diagrama de caja de la variación diaria de Avon. En este caso el primer cuartil se sitúa en $-0'0055$ y el tercer cuartil en $0'007$. Esto indica que el 50% de las variaciones diarias están entre $-0'55\%$ y $0'70\%$. Por otro lado, en el caso de la Figura 18, se puede observar que todos valores atípicos no están muy alejados del primer y tercer cuartil ($6'3\% > 10 * Q_3$; $-5'4\% < 10 * Q_1$). El conjunto de los cuartiles de Chevron y sus valores atípicos hacen que sea un valor mucho menos volátil que Avon, del cual se ha hablado antes en la Figura 17. Para concluir con la Figura 18, y a diferencia de Avon, la mediana se sitúa en $0'9\%$, lo cual indica que, durante el periodo analizado, hay un crecimiento por parte de Chevron.

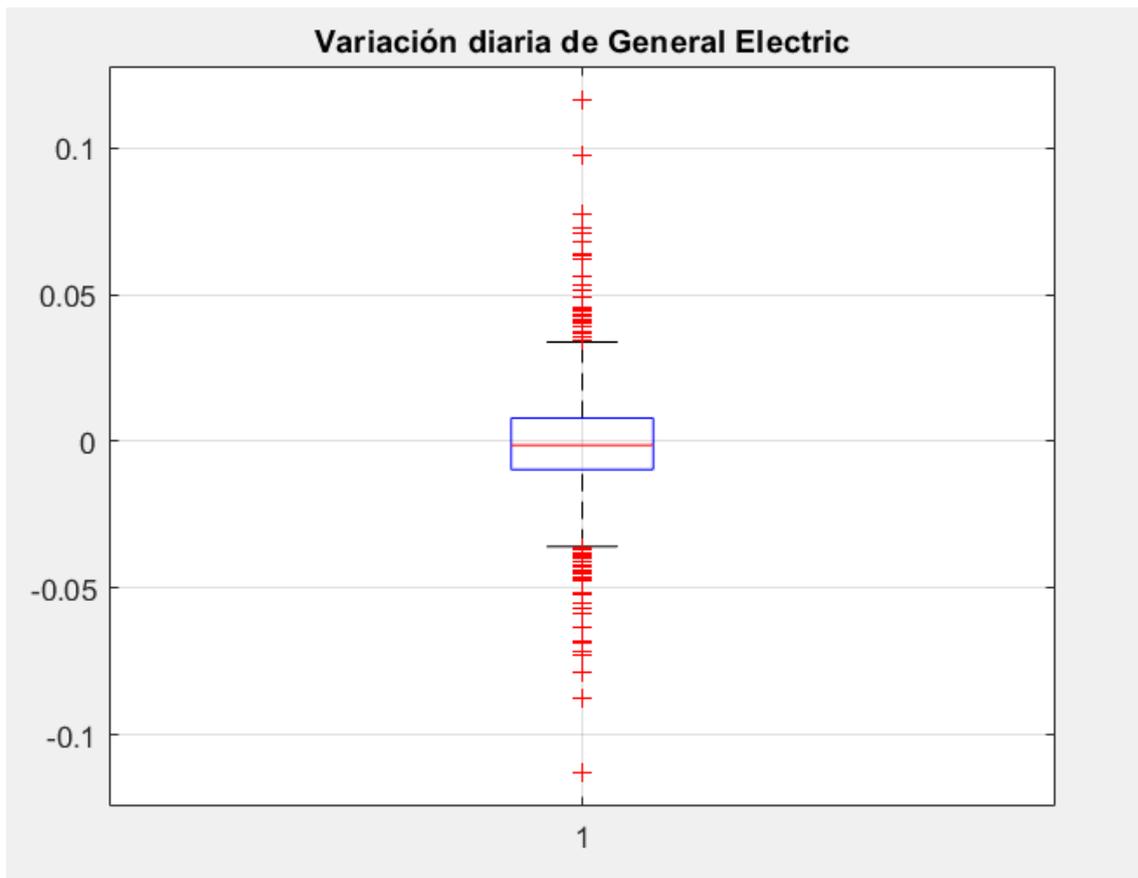


Figura 19. Diagrama de cajas de la variación diaria de General Electric

La Figura 19 muestra el diagrama de caja de la variación diaria de General Electric. En este caso el primer cuartil se sitúa en $-0'0096$ y el tercer cuartil en $0'0079$. Esto indica que el 50% de las variaciones diarias están entre $-0'96\%$ y $0'79\%$. Por otro lado, en el caso de la Figura 19, se pueden observar algún valor atípico que están bastante alejados ($11'6\% > 10 * Q_3$; $-11'3\% < 10 * Q_1$), pero aún así, los valores atípicos no llegan al 20%, cosa que sí ocurría con Avon.

Variación diaria con noticias

Se ha representado la variación diaria con las noticias en el día que sucedían, lo cual sirve para complementar el diagrama de cajas de la variación diaria.

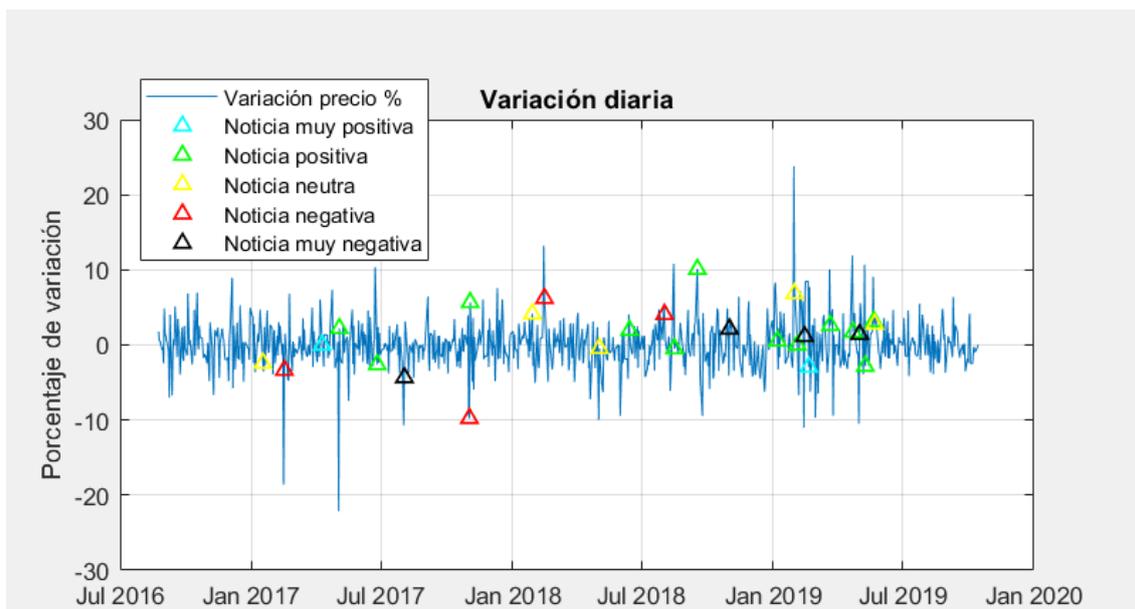


Figura 20. Variación diaria con noticias de Avon

En la Figura 20 aparece la variación diaria de la empresa Avon durante todo el periodo analizado. Si lo contrastamos con la Figura 17 y combinando ambas gráficas, se puede observar que los valores más atípicos sucedieron el 03/05/2017 y el 30/01/2019. Un día después de la variación diaria de -22%, se hizo pública una noticia con valor sentimental positivo, por otro lado, un día después de la variación diaria de 23'75%, se hizo pública una noticia con valor sentimental neutro.

Por este tipo de comparaciones, y para mejorar la experiencia del usuario, la herramienta muestra el diagrama de cajas encima de la variación diaria. A continuación, en la Figura 21, se muestra una parte del resultado de la herramienta donde aparece representadas ambas gráficas, en concreto, son las gráficas de Avon desde el 24/08/2016 hasta el 15/10/2019.

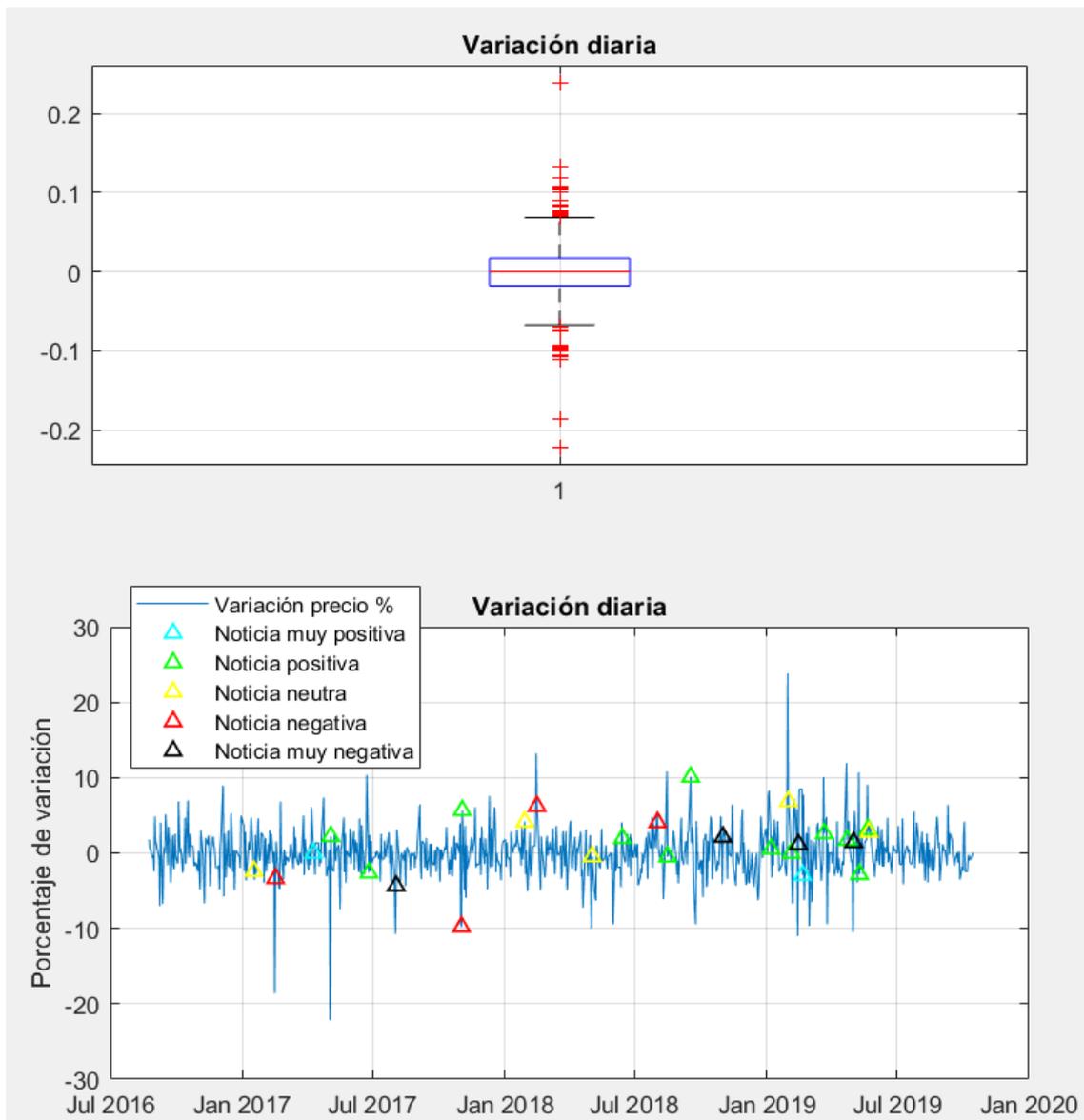


Figura 21. Comparativa del diagrama de cajas y la variación diaria con noticias de Avon

Cotización con medias móviles

Las medias móviles son un recurso muy utilizado que da información al usuario sobre la tendencia de las acciones. Se calcula haciendo la media de los precios de los últimos ‘x’ días. Es por esto por lo que las medias móviles tienen un pequeño retardo, por lo cual,

este indicador no prevé cambios de tendencia, sino que confirma el cambio o continuación de esta.

La media móvil depende de los últimos 'x' días. Un 'x' mayor implica una curva que reacciona menos a la variación diaria y se centra más en la tendencia general, ya que, al tener más muestras, los nuevos datos se diluyen más que si hubiese pocos datos.

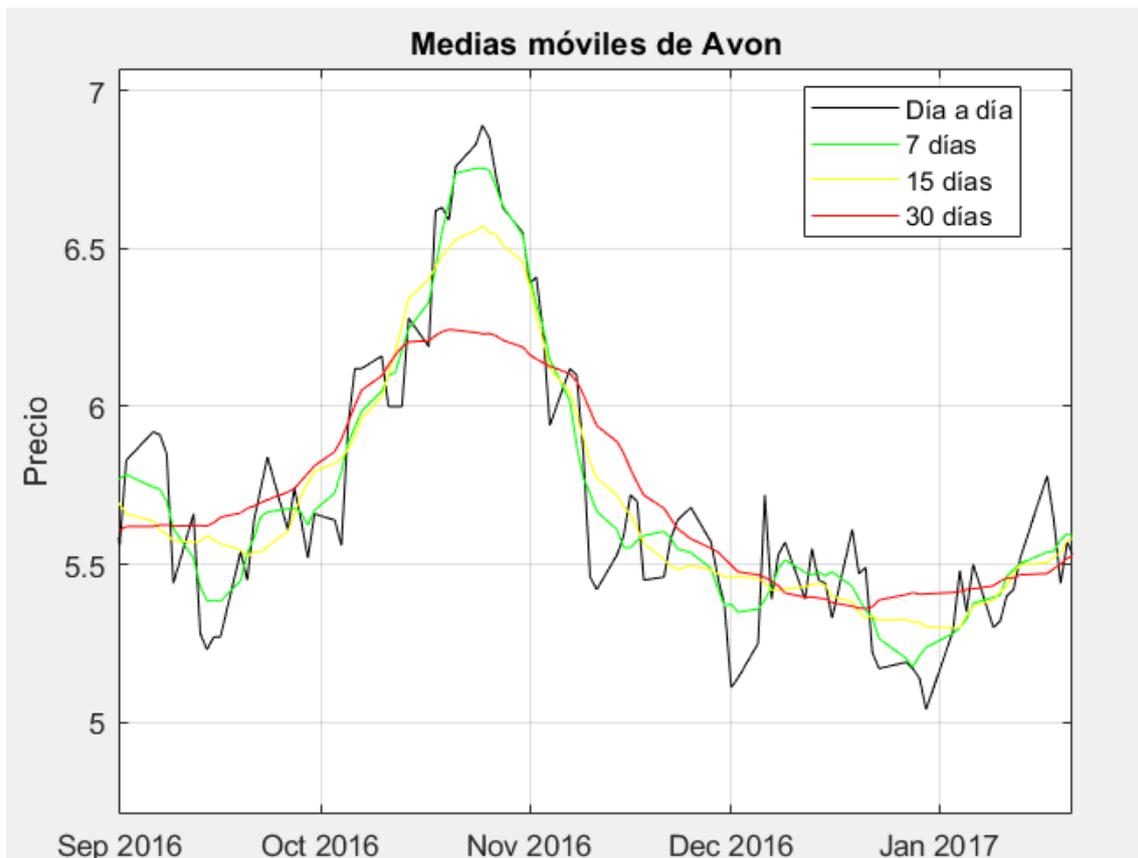


Figura 22. Medias móviles de cotización de Avon

La Figura 22 muestra medias móviles y la cotización de la empresa Avon desde el 01/09/2016 hasta el 24/01/2017. Representado se puede observar que justo debajo del pico que sucede en octubre del 2016 se encuentra la media móvil de 7 días, seguido de la media móvil de 15 días, y por último la media móvil de 30 días, la cual reacciona menos al pico por el motivo explicado anteriormente.

Otra característica que hace de las medias móviles un buen indicador es por su capacidad de indicar la volatilidad de una acción. Cuanto más separadas estén las

medias móviles, más variaciones o picos habrá en la cotización de la acción. La Figura 22 es un ejemplo de ello, en el pico de octubre, las medias móviles están más separadas que en el resto de la muestra, donde el precio se muestra relativamente estable y las medias móviles adoptan valores más similares.

Además, el usuario es capaz de ver donde se cruzan las medias móviles con otras medias móviles y con el precio. Esto puede determinar la compra y venta de acciones por parte del usuario. Se recomienda comprar cuando el precio justo supera el valor de la media móvil, y vender cuando la media móvil supera al precio, pero elegir la media móvil más adecuada es el trabajo del analista.

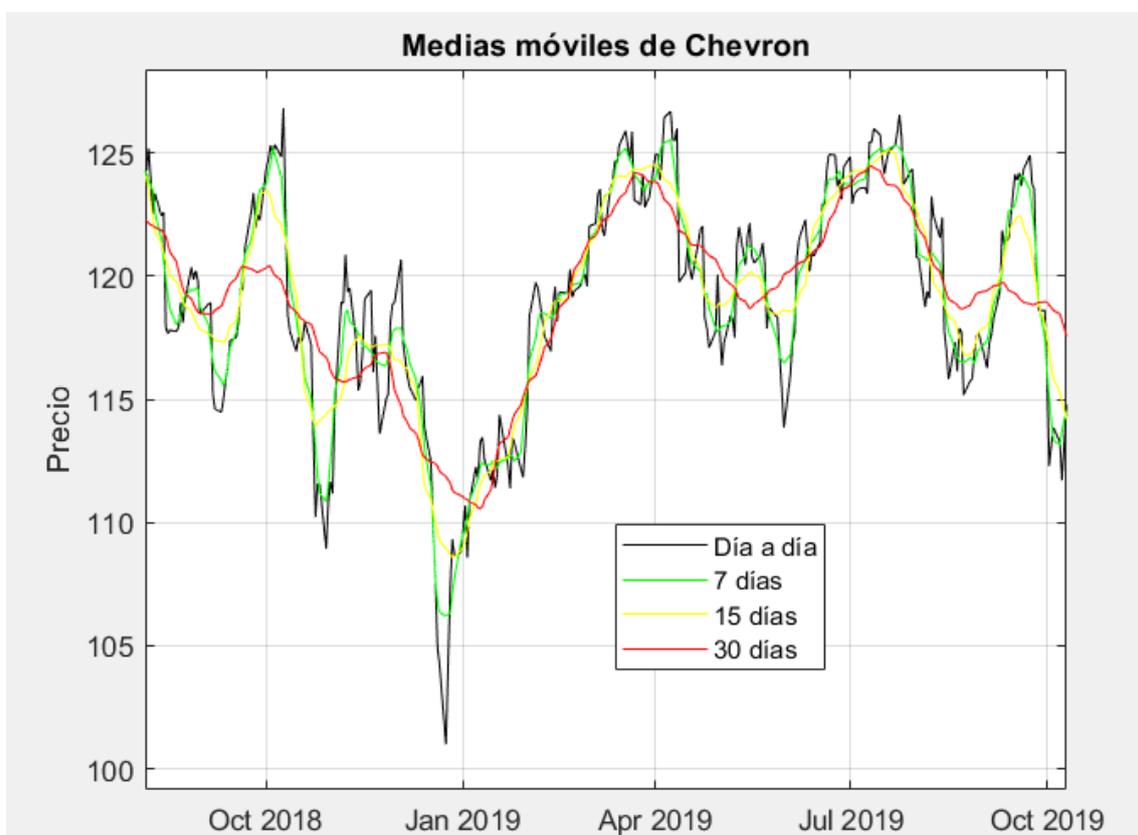


Figura 23. Medias móviles de cotización de Chevron

En el caso de la Figura 23 muestra medias móviles y la cotización de la empresa Chevron, en la cual se puede volver a observar que la media móvil de 7 días es más sensible que la media móvil de 30 días. Por otro lado, se puede ver mejor el retardo de las medias móviles en esta Figura 23, vemos que el mínimo del precio y el mínimo de la media móvil de 7

días se dan el 24/12/2018, mientras que el pico de la media móvil de 15 días se da el 27/12/2018, y, por último, el pico de la media móvil de 30 días se da el 09/01/2019.

Primera ventana de la herramienta

La primera ventana de la herramienta representa todas las gráficas explicadas anteriormente en este capítulo, como se muestra en la Figura 24

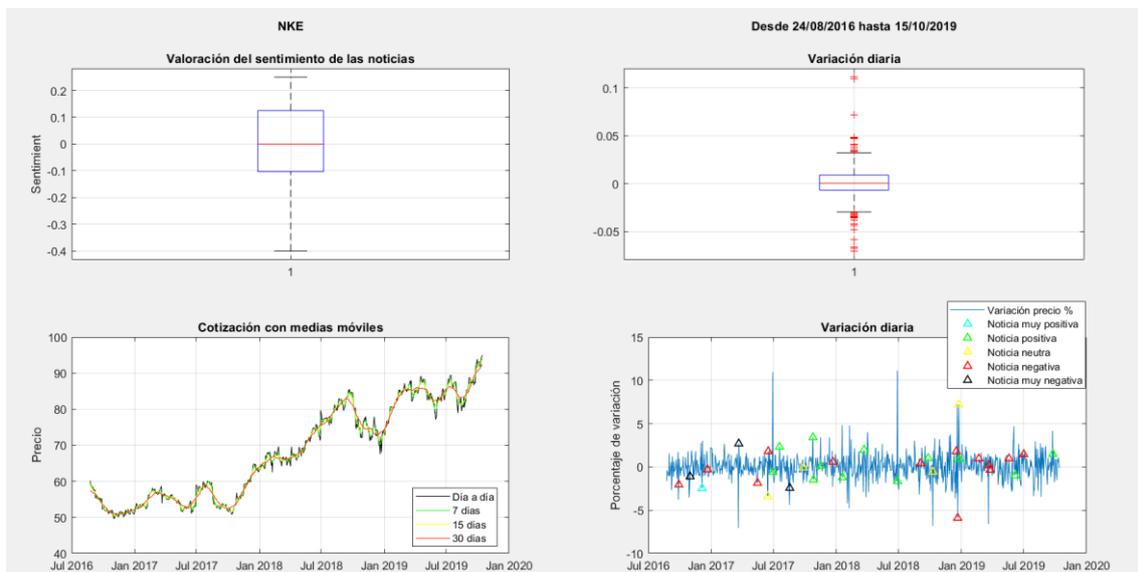


Figura 24. Apariencia de las gráficas en la herramienta

Como se puede observar en la Figura 24, arriba a la izquierda aparece el nombre de la empresa analizada, en este caso se trata de Nike (NKE). Siguiendo la horizontal, encontramos el periodo, en el caso de la Figura 24, el periodo empieza el 24/08/2016 y termina el 15/10/2019. A continuación, aparecen representados los diagramas de cajas, las medias móviles y la variación diaria con las noticias, las cuales son representadas por triángulos con un código de colores.

Al igual que con Nike, el usuario puede repetir este proceso con las otras 53 empresas disponibles y filtrando en el periodo que desee analizar. A continuación, la Figura 25 muestra un ejemplo del resultado que obtendría el usuario si seleccionara el año 2017 para analizar. En este caso, la empresa a analizar es AT&T (T), pero la disposición del resultado es idéntica a la Figura 24

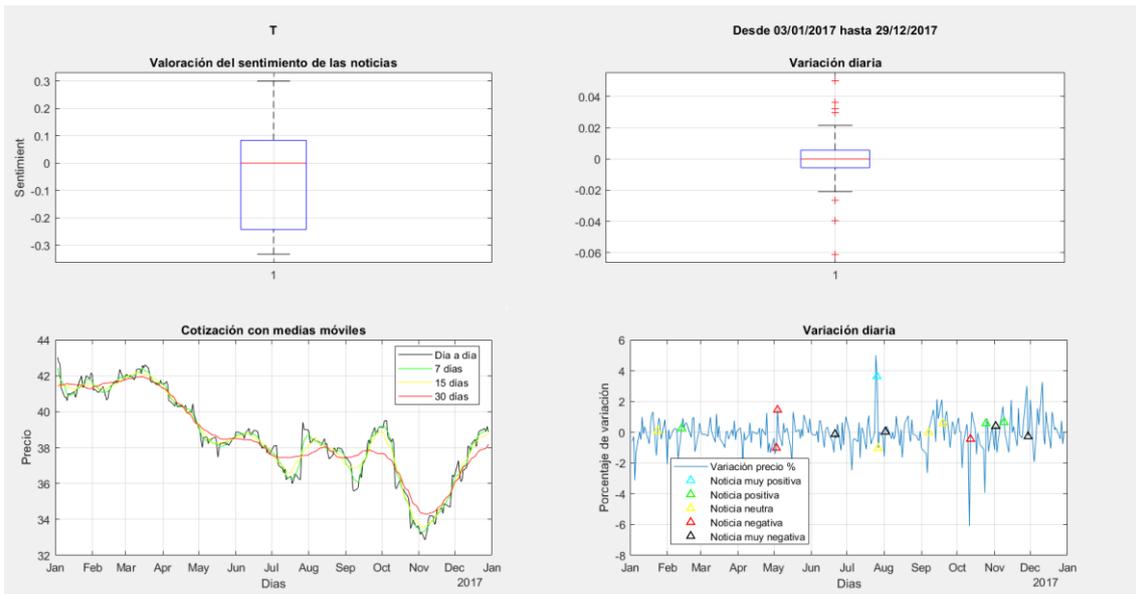


Figura 25. Apariencia de las gráficas con filtro fecha

Modelo Holt herramienta

Se ha creado dos modelos Holt para analizar el periodo. Para determinar los mejores estimadores, la herramienta comprueba 1.000.000 de casos con distintos estimadores. El primer modelo Holt es aquel que tiene el menor error cuadrático, mientras que el segundo modelo Holt es aquel que tiene el menor error cuadrático para un alfa menor o igual a 0'1. Como ya se ha explicado en capítulos anteriores, cuanto menor es el alfa, menos importancia tendrán los últimos datos. A continuación, se incluye un ejemplo de la empresa Google con su óptimo modelo Holt.

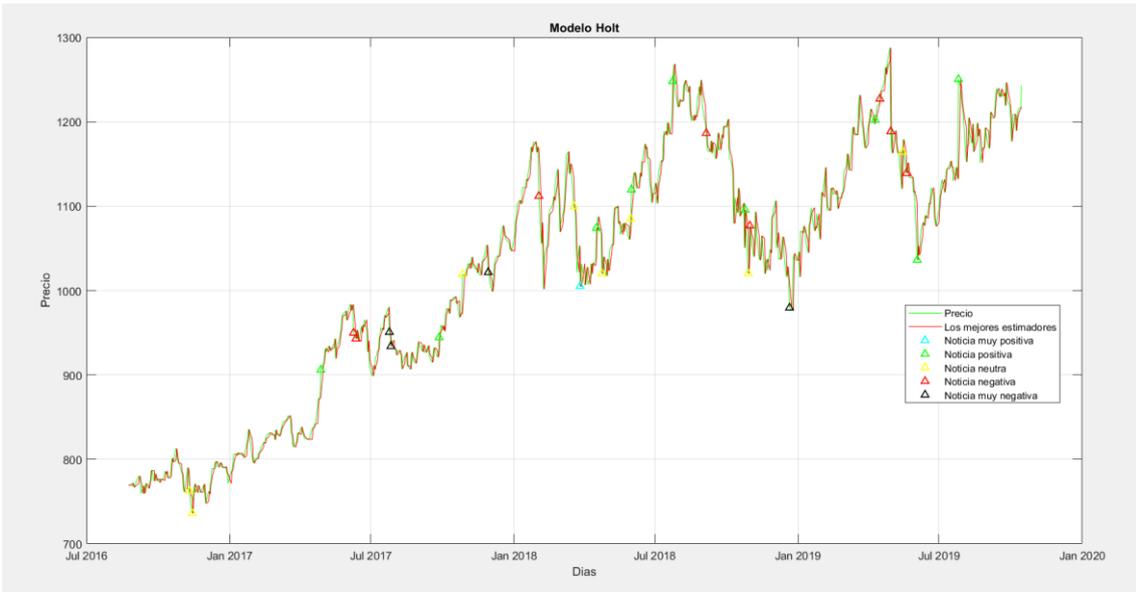


Figura 26. Modelo Holt con los mejores estimadores de Google

La Figura 26 muestra la cotización de la acción (gráfica verde), el modelo Holt con los mejores estimadores (gráfica roja), y las noticias. La Figura 27 que se muestra a continuación, es una ampliación de la Figura 26, en ella se puede observar que ambas gráficas son muy similares pero que el modelo Holt presenta un cierto retardo con respecto al gráfico de la cotización.

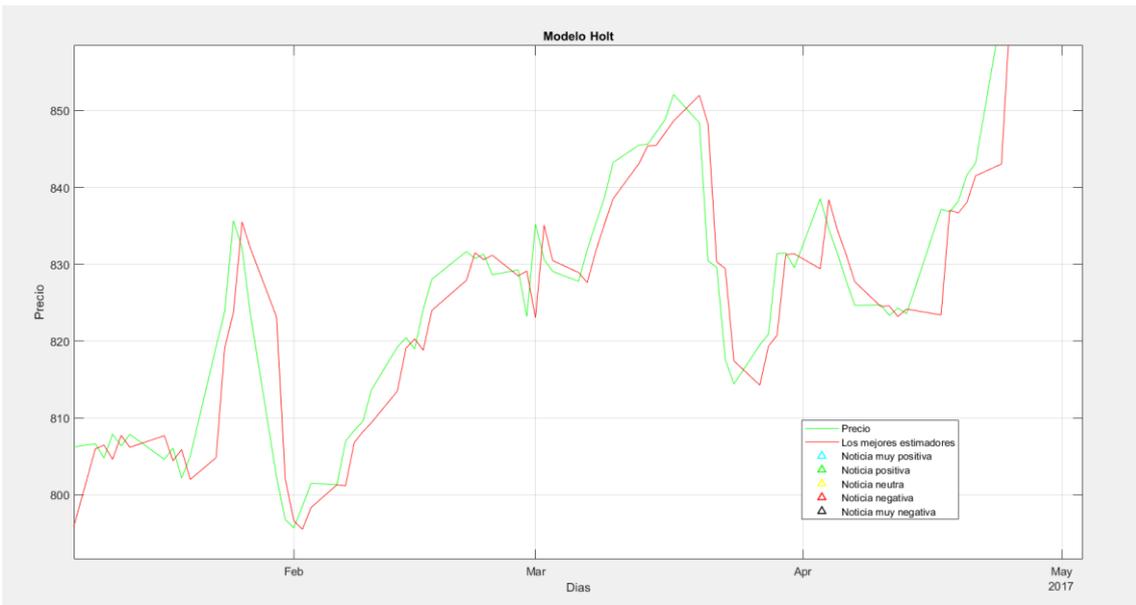


Figura 27. Zoom modelo Holt figura 26

A continuación, la Figura 28, muestra la cotización de la acción (gráfica verde), el modelo Holt con los mejores estimadores con la condición de $\alpha \leq 0.1$ (gráfica roja), y las noticias de la empresa Google. El modelo sigue la tendencia de la cotización, pero no consigue prever los picos que hay en el precio. Además, el retardo con respecto a la cotización sigue siendo apreciable ya que el modelo alcanza sus máximos y mínimos unos días después de que lo haga la cotización.

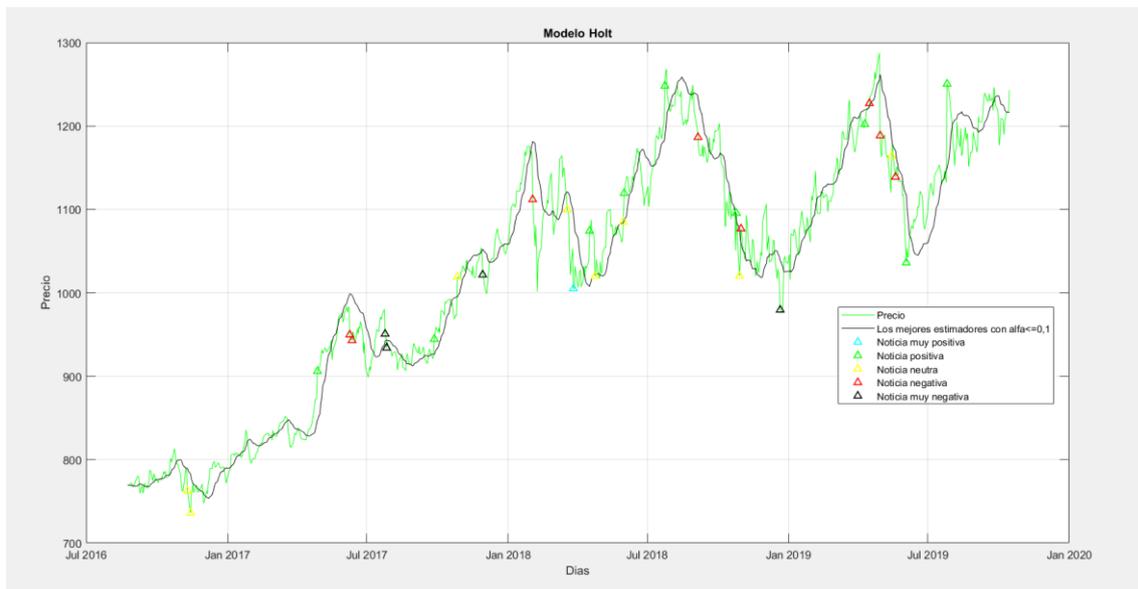


Figura 28 Modelo Holt con restricciones

La Figura 29 que se encuentra a continuación es una ampliación de la Figura 28, en ella se puede observar dos cosas. La primera es que el modelo Holt presenta un cierto retardo con respecto al gráfico de la cotización. La segunda es que este modelo está más suavizado que el anterior, no hay cambios tan bruscos.

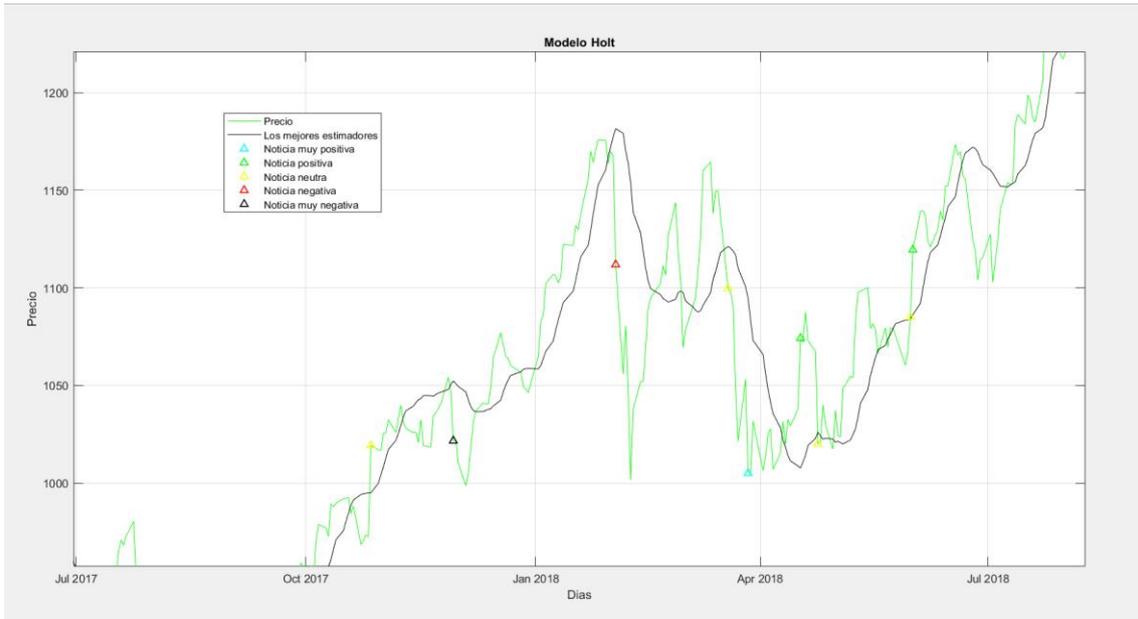


Figura 29. Zoom modelo Holt figura 28

A continuación, se muestra en la Figura 30 una representación conjunta de ambos modelos Holt, la cotización y las noticias.

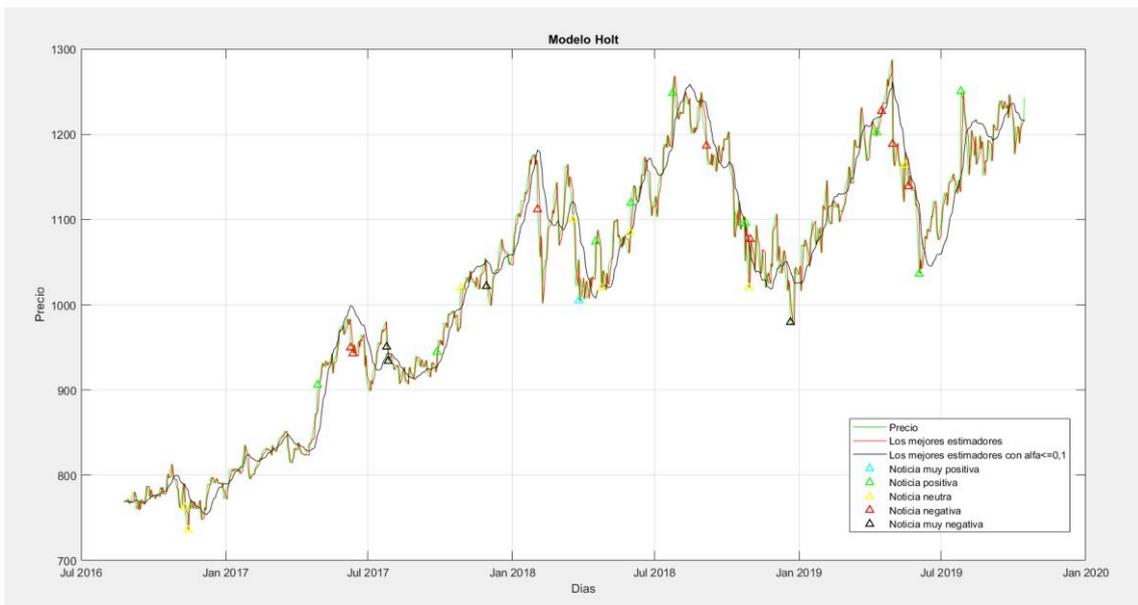


Figura 30. Modelos Holt

La Figura 30 muestra la cotización de la empresa Google con sus noticias y los dos modelos Holt representados por la gráfica roja y por la gráfica negra. El periodo analizado empieza el 24/08/2016 y termina el 15/10/2019. La herramienta dispondrá los datos de las Figuras 26 y 28 superpuestos como en la Figura 31 para mostrar las diferencias que hay entre ambos modelos. A continuación, se explicarán las diferencias entre los modelos y en el capítulo de conclusiones se profundizará aún más.

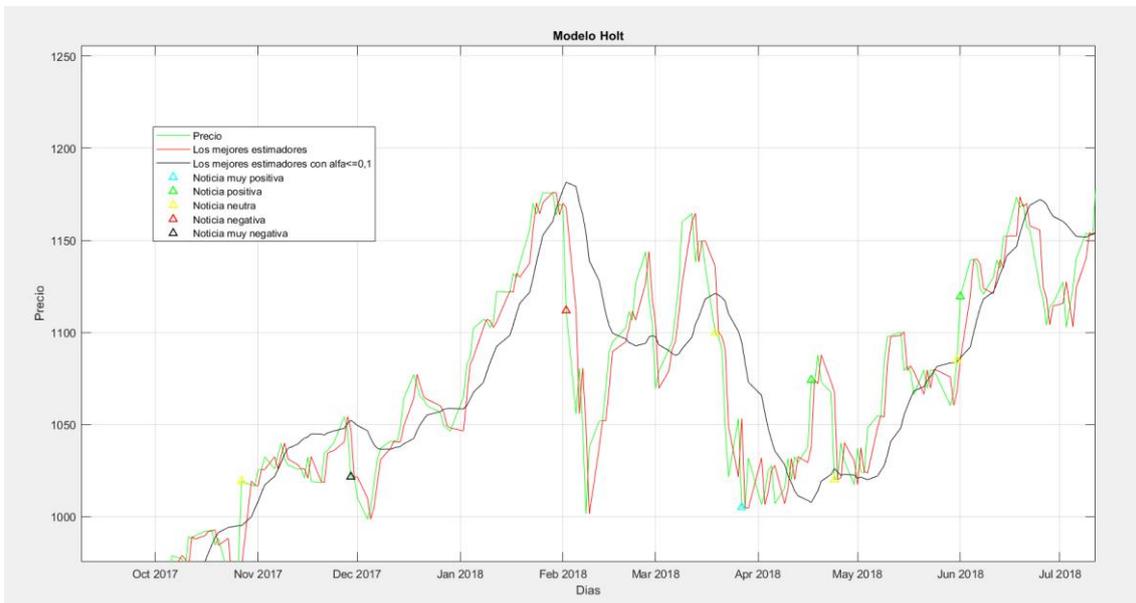


Figura 31. Zoom figura 30

En la Figura 31, se puede observar que las gráficas de la cotización y del modelo Holt sin restricciones son muy similares, pero el modelo Holt con la limitación de $\alpha \leq 0.1$ es una curva que no consigue prever los picos, cosa que no ocurre con el otro modelo. Además, el modelo Holt con la limitación de $\alpha \leq 0.1$ tiene un retardo superior al modelo Holt sin limitaciones, se puede observar cómo es el último gráfico que alcanza los máximos y mínimos. Por otro lado, hay que añadir que el modelo Holt con la limitación de $\alpha \leq 0.1$ es una curva más suavizada y por eso tiene un error cuadrático superior al del modelo Holt sin limitaciones.

Cotización con análisis sentimental de las noticias

Se ha representado la cotización con las noticias que han sucedido a lo largo del periodo. Se trata de un análisis descriptivo para que el usuario tenga una idea de como han afectado las noticias a la cotización de la empresa. Un ejemplo es la Figura 32 que se muestra a continuación.

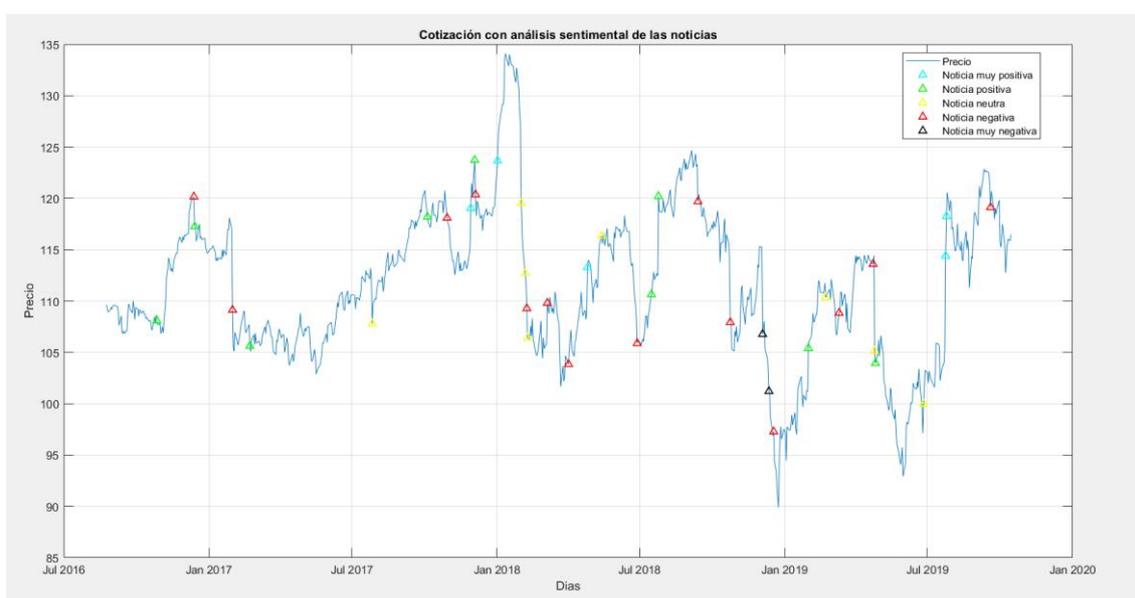


Figura 32. Cotización con noticias

La herramienta ha realizado un cálculo anterior similar al representar la variación diaria con las noticias. Tal y como se muestra a continuación en la Figura 33, el usuario obtiene la información del valor de la acción y cuanto ha variado. Incluso podría ver si esa variación o ese precio se debe a una noticia de los días anteriores.

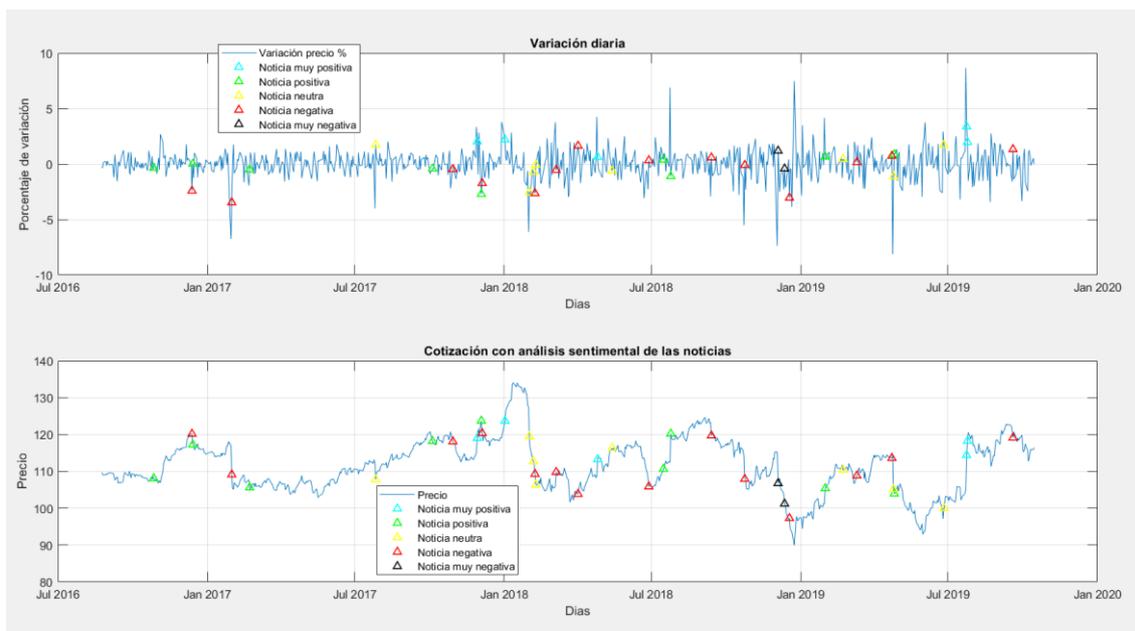


Figura 33. Comparación de Cotización con noticias vs. variación diaria

Efecto de las noticias por medio del contraste de pendientes

Como ya se ha explicado en el capítulo anterior, una noticia positiva no tiene que implicar una subida en el precio de la acción. Es por esto por lo que se ha implementado un contraste de pendientes para determinar el verdadero valor sentimental de las noticias.

En primer lugar, la herramienta detecta donde están las noticias. Según Heston (2017) [6], las noticias solo tienen un impacto durante 7 días después de la noticia. Como la bolsa solo está abierta de lunes a viernes, el programa analiza los 5 días de antes de la noticia y los 5 días después de la noticia, realiza dos regresiones lineales, una para los 5 días anteriores y otra para los 5 días posteriores, y con un contraste de pendientes, determina si la pendiente anterior era diferente a la posterior. Dependiendo del p-valor y de la diferencia de las pendientes de las regresiones lineales, se han clasificado las noticias con su verdadero valor sentimental, tal y como se explica en la Tabla 13.

CLASIFICACIÓN DE LAS NOTICIAS	DIFERENCIA DE PENDIENTES (DESPUÉS - ANTES)	P_VALOR
MUY POSITIVA	>0	$p_valor \leq 0'01$
POSITIVA	>0	$0'1 \leq p_valor < 0'01$
NEUTRA	-	$p_valor > 0'01$
NEGATIVA	<0	$0'1 \leq p_valor < 0'01$
MUY NEGATIVA	<0	$p_valor \leq 0'01$

Tabla 13

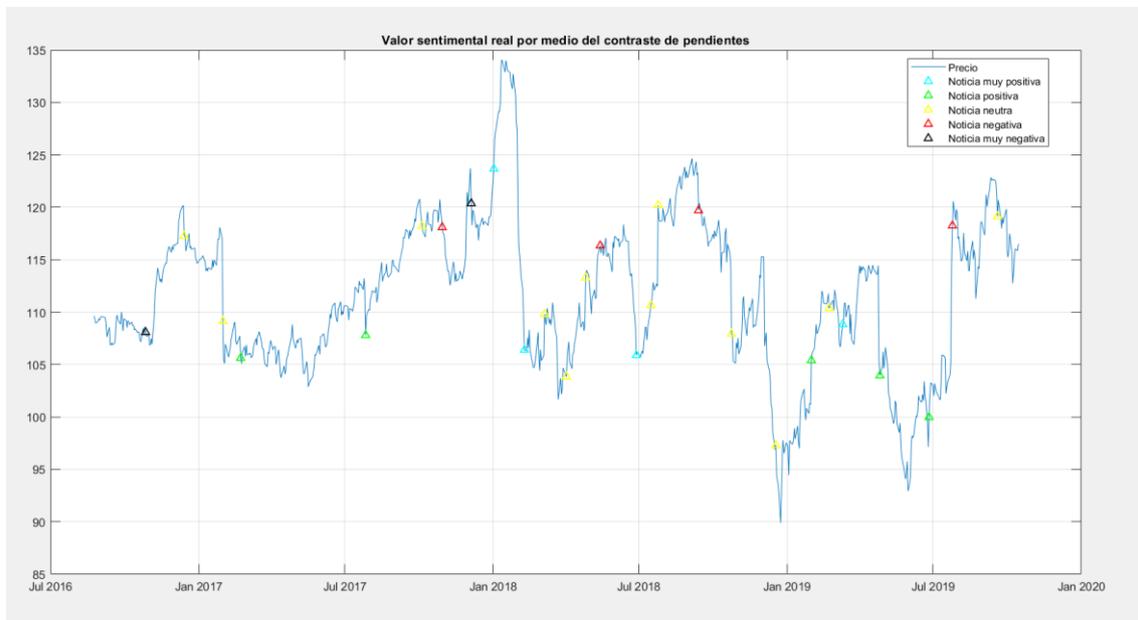


Figura 34. Cotización con el VS noticias por el contraste de pendientes

En la Figura 34, se puede ver un ejemplo del programa analizando la compañía UPS desde el 24/08/2016 hasta el 15/10/2019.

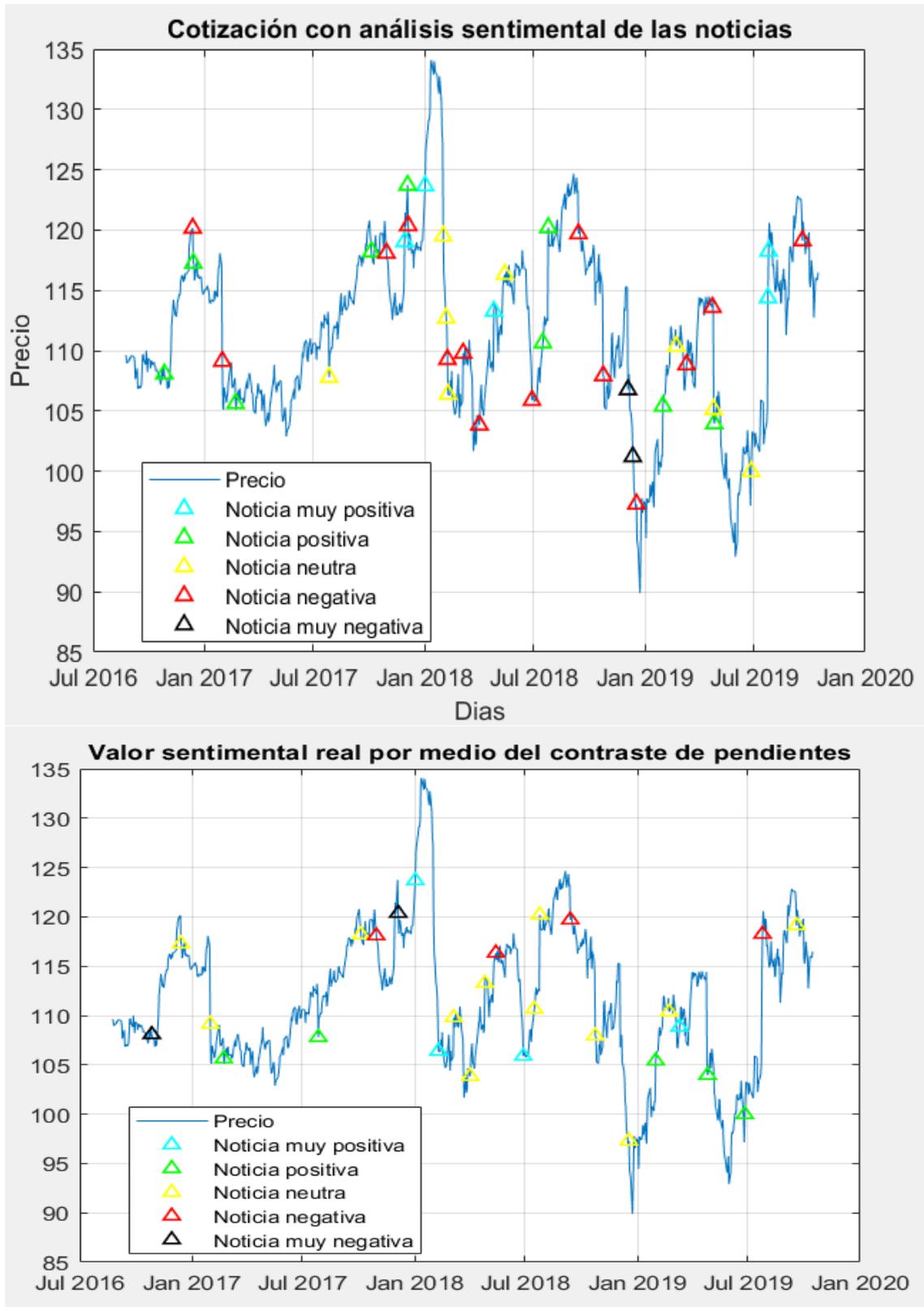


Figura 35. Comparativa del VS y el contraste de pendientes

Como se puede observar en la Figura 35, la gráfica cotización con análisis sentimental de noticias tiene más noticias que la gráfica valor sentimental real por medio del contraste de pendientes. Esto se debe al efecto aditivo que tienen dos noticias al estar cerca en el tiempo (intervalo inferior a 7 días naturales o 5 días hábiles), este efecto ya ha sido explicado en capítulos anteriores. La siguiente diferencia es que no todas las noticias mantienen el mismo valor sentimental, como se puede ver a continuación en la Figura 36

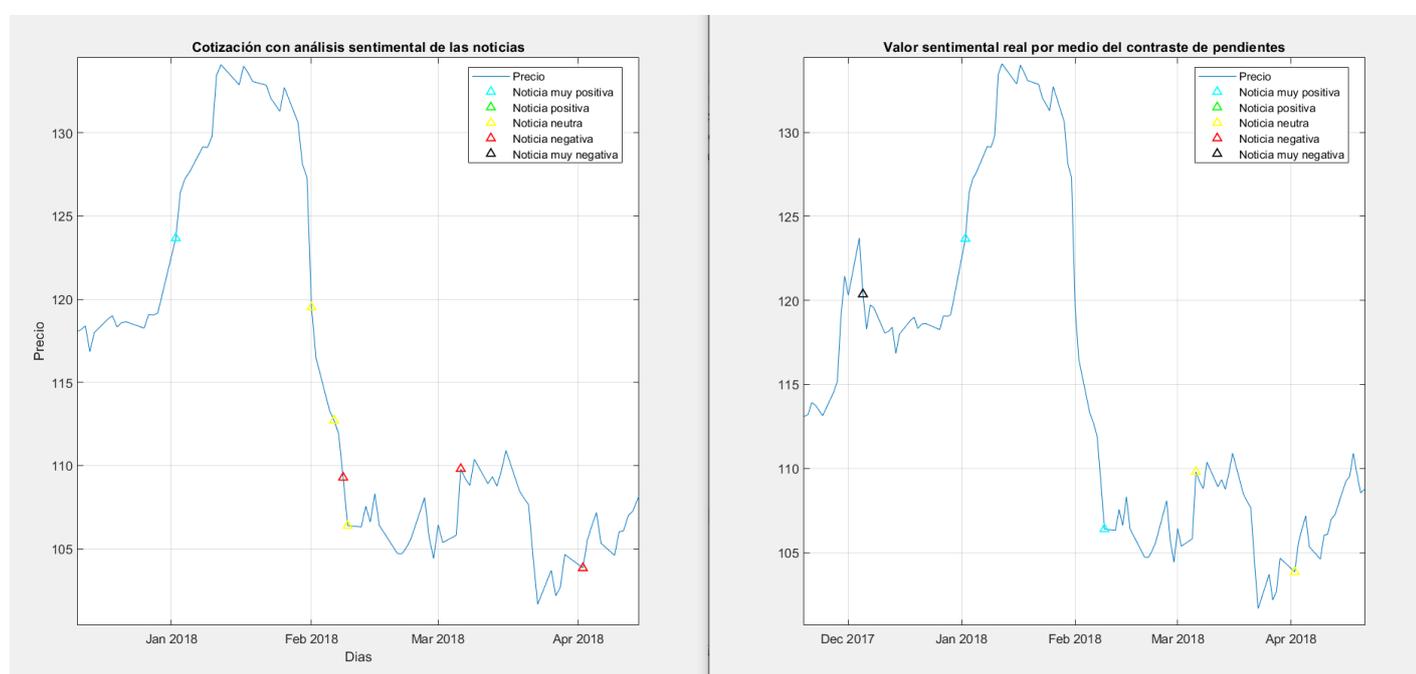


Figura 36. Diferencias en las noticias

En el 09/02/2018 de la Figura 36 se encuentra una noticia que no mantiene el valor sentimental. Justo coincide con el final de una gran caída en el precio de la acción. Al principio, esa noticia tenía un valor sentimental neutro, y al hacer el contraste de hipótesis, se calificó como una noticia muy positiva. Por otro lado, la noticia del 02/01/2018 mantiene su clasificación como muy positiva.

Efecto de las noticias por medio del contraste de variables emparejadas

Otra forma de ver el verdadero valor sentimental de las noticias es por medio de la diferencia de medias de dos normales de muestras emparejadas, en este caso las muestras emparejadas son las variaciones diarias de antes de la noticia y después de esta misma. Para clasificar las noticias, se mira la diferencia de medias y su p-valor. A continuación, en la Tabla 14 aparecen los criterios.

CLASIFICACIÓN DE LAS NOTICIAS	MEDIA DE LAS DIFERENCIAS DE LAS VARIABLES EMPAREJADAS (DESPUÉS - ANTES)	P_VALOR
MUY POSITIVA	>0	$p_valor \leq 0'01$
POSITIVA	>0	$0'1 \leq p_valor < 0'01$
NEUTRA	-	$p_valor > 0'01$
NEGATIVA	<0	$0'1 \leq p_valor < 0'01$
MUY NEGATIVA	<0	$p_valor \leq 0'01$

Tabla 14

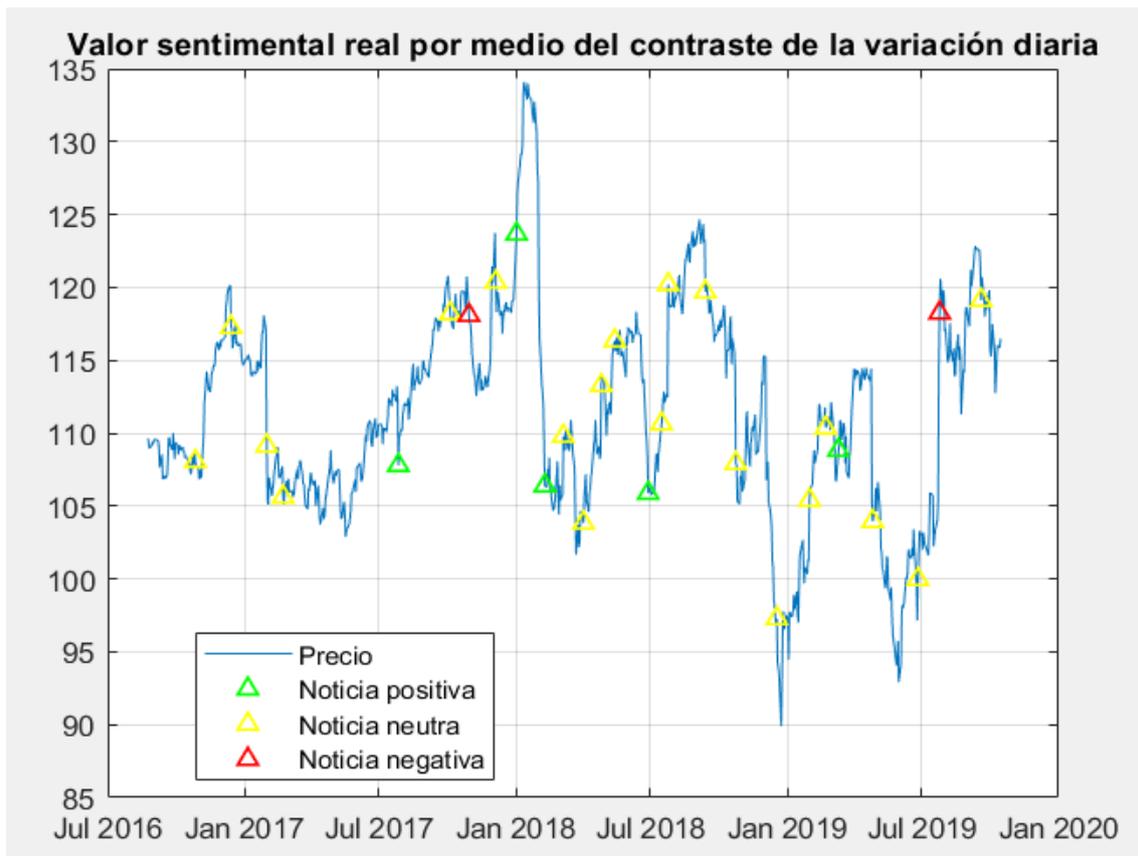


Figura 37. Cotización con el contraste de variables emparejadas

En la Figura 37, se puede ver un ejemplo del programa analizando a la compañía UPS desde el 24/08/2016 hasta el 15/10/2019 por medio del contraste de diferencia de medias de variables emparejadas.

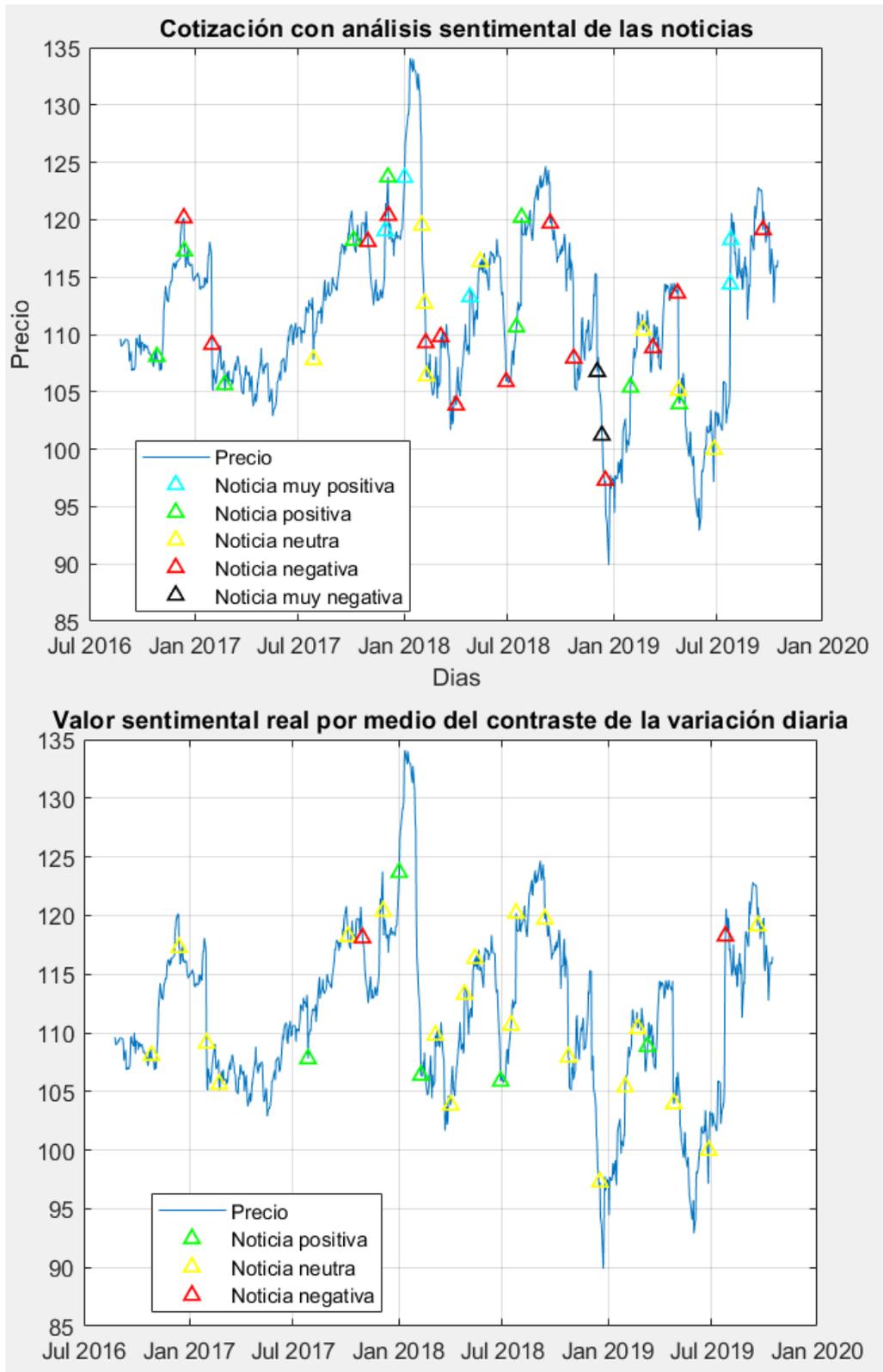


Figura 38. Comparativa entre VS y el contraste de variables emparejadas

Como se puede observar en la Figura 38, al igual que ocurría con el anterior contraste, la gráfica cotización con análisis sentimental de noticias tiene más noticias que la gráfica valor sentimental real por medio del contraste de pendientes. Esto se debe al efecto aditivo que tienen dos noticias al estar cerca en el tiempo (intervalo inferior a 7 días naturales o 5 días hábiles), este efecto ya ha sido explicado en el capítulo anterior. La siguiente diferencia es que no todas las noticias mantienen el mismo valor sentimental, como se puede ver a continuación en la Figura 39

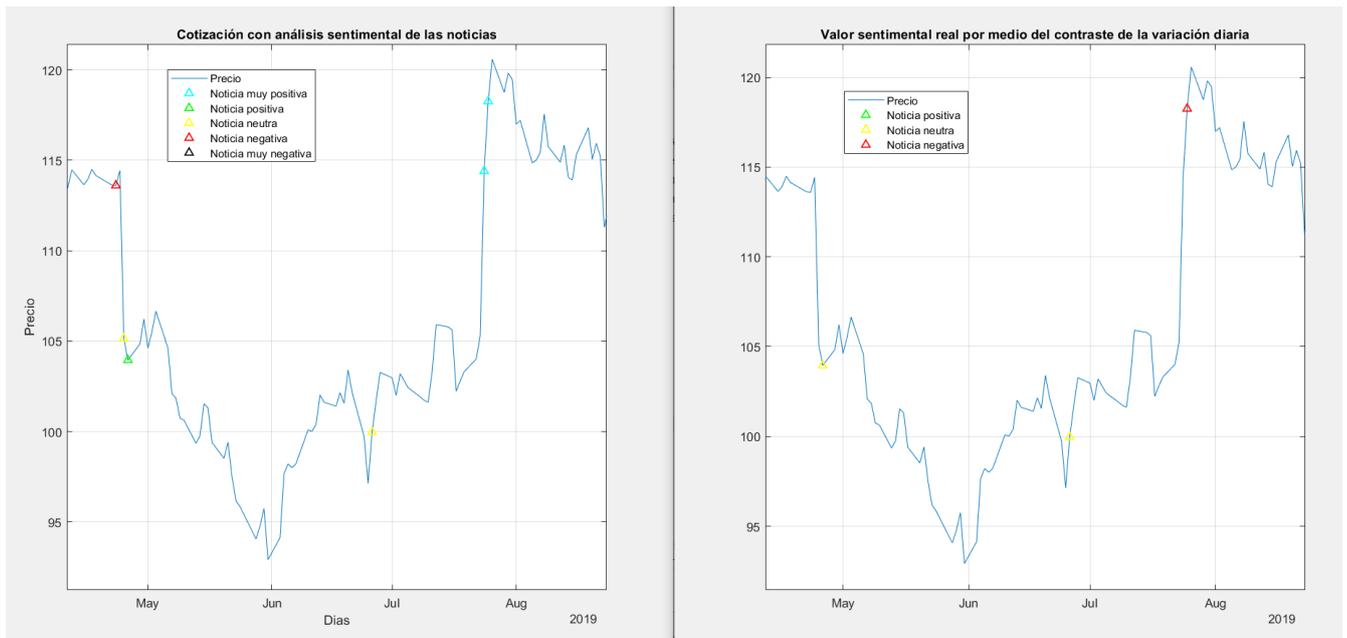


Figura 39. Zoom figura 38

En la Figura 38, solo hay 3 noticias que mantengan el mismo valor sentimental, una de ellas sale representada en Figura 39, el 26/06/2019, la noticia se mantuvo como noticia neutra. Por el contrario, la noticia del 25/07/2019 que fue clasificada como muy positiva y teniendo el efecto aditivo de otra noticia muy positiva fue considerada como negativa.

Segunda ventana de la herramienta

La segunda ventana representa los contrastes de hipótesis y la cotización con las noticias, tal y como se vería en la herramienta. Un ejemplo se muestra en la Figura 40

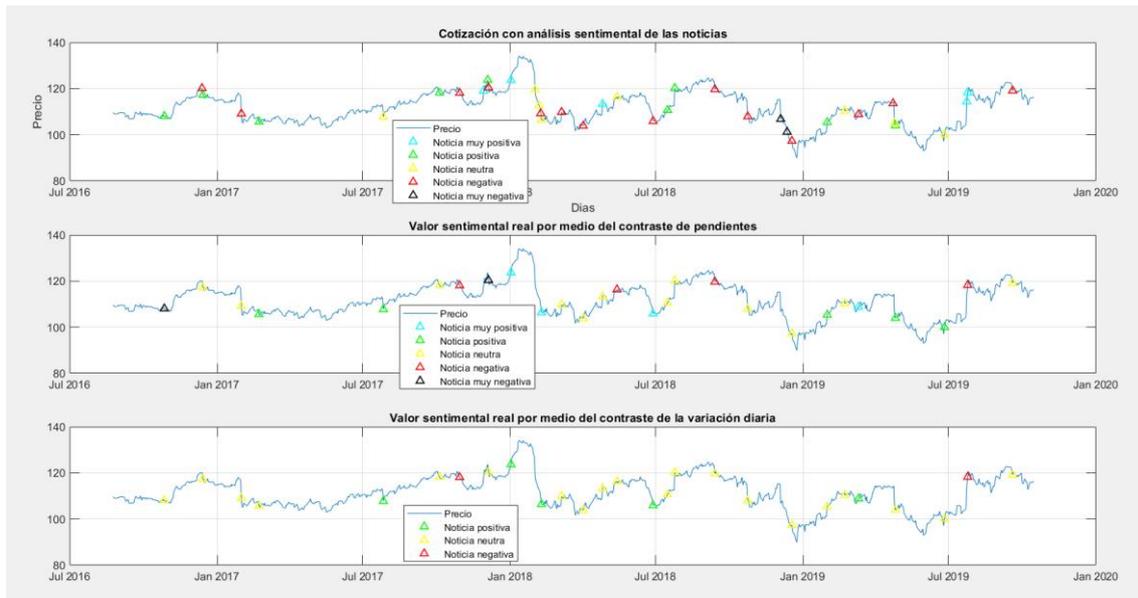


Figura 40. Comparativa entre los 3 métodos

La Figura 40 muestra como el usuario vería los gráficos en la herramienta desarrollada. Se han colocado los gráficos de tal forma que le resulte como al usuario compararlos.

Volumen

Se ha representado la cotización con las noticias que han sucedido a lo largo del periodo. Se trata de un análisis descriptivo para que el usuario tenga una idea de cómo han afectado las noticias a la compraventa de acciones de la empresa.

Volumen con análisis sentimental de las noticias

En este caso, se ha acompañado al volumen con las noticias y su valor sentimental. Como ejemplo se dispone a la empresa McDonald's en la Figura 41. Como se puede observar, una gran parte de las noticias se sitúan en picos del volumen, sin importar el tipo de noticia.

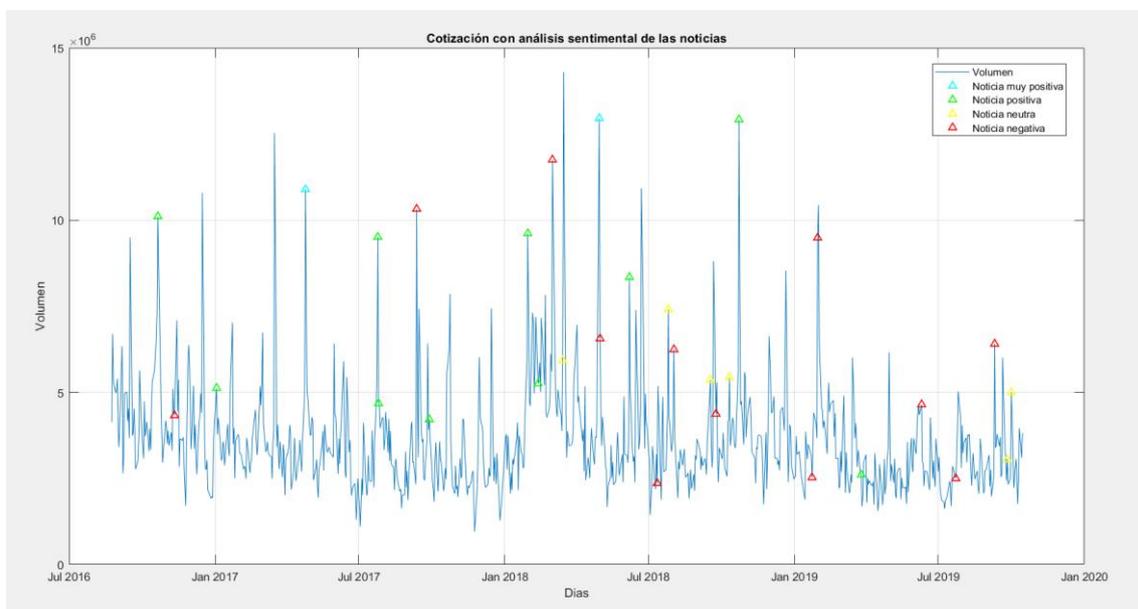


Figura 41. Volumen con noticias de McDonald's

Efecto de las noticias por medio del contraste de pendientes

Para ver si el efecto de las noticias influía en el volumen, se ha hecho lo mismo que en la anterior Figura, pero con las noticias analizadas con el contraste de pendientes. Utilizando la misma empresa, McDonald's, se ha representado en la Figura 42. Esta vez, predominan las noticias negativas en los picos del volumen, pero también hay noticias negativas que no están en picos, y noticias neutras en picos. Esta variedad también ocurre en las distintas empresas analizadas, donde en los picos hay noticias de distinto valor sentimental.

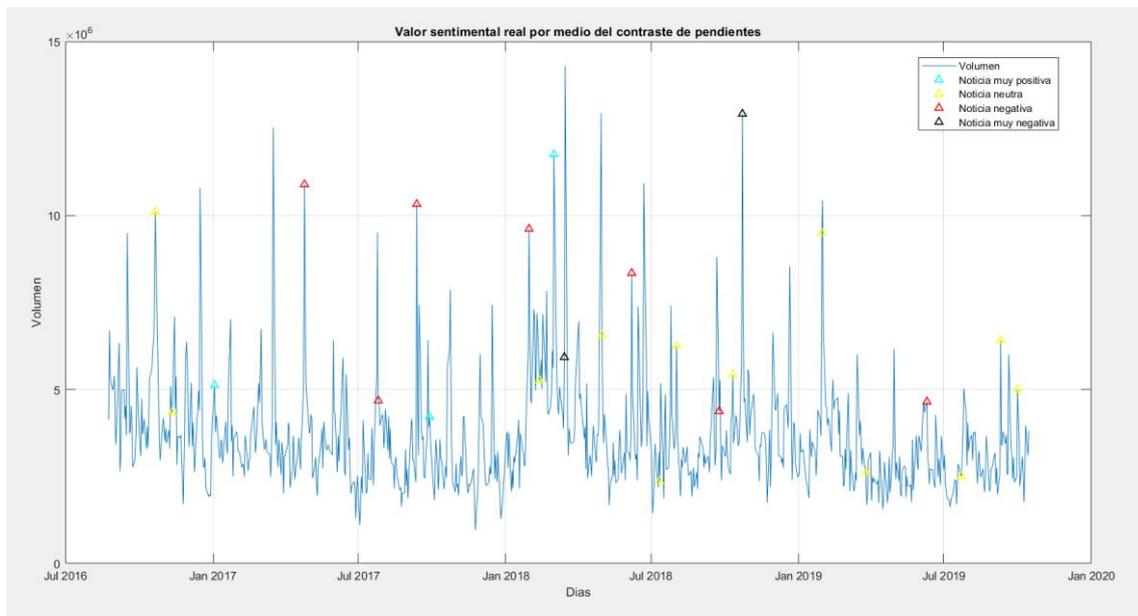


Figura 42. Volumen con noticias evaluadas con el contraste de las pendientes

Efecto de las noticias por medio del contraste de variables emparejadas

En la Figura 43, se ha vuelto a analizar McDonald's con el contraste de variables emparejadas, en el cual predominan las noticias neutras, ya sean picos o no. Esto difiere mucho con el anterior método, ya que consideraba que había más noticias negativas en picos del volumen. Esto se debe a que este método tiende a considerar que las noticias son más bien neutras.

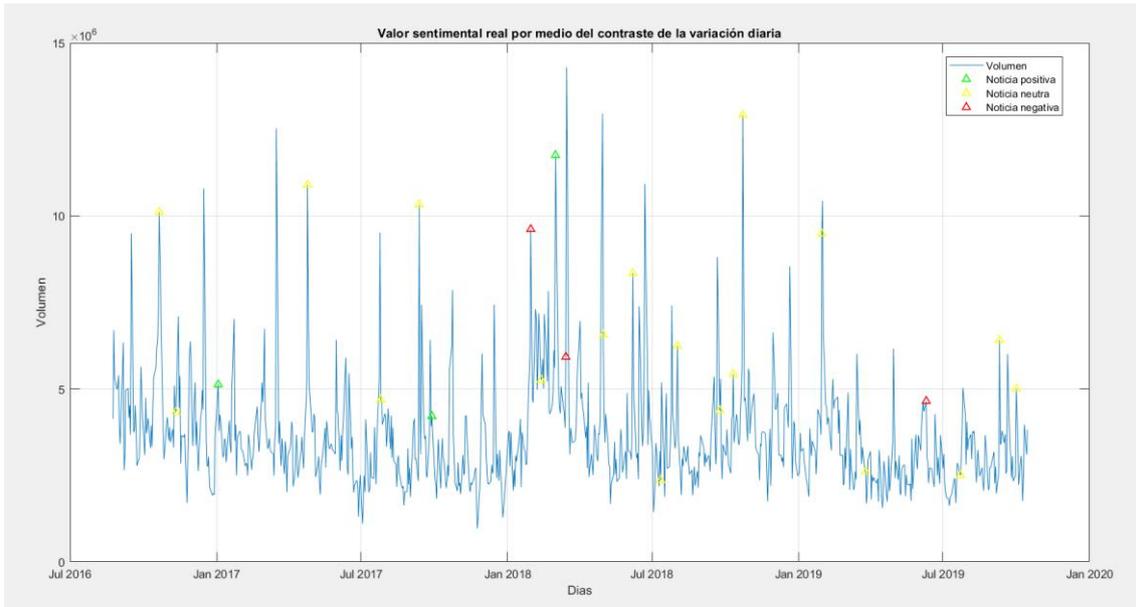


Figura 43. Volumen con noticias evaluadas con el contraste de variables emparejadas

Tercera ventana de la herramienta

La herramienta muestra al usuario las gráficas como se muestran en la Figura 44 a continuación. Se han colocado los gráficos para que al usuario los pueda comparar fácilmente.

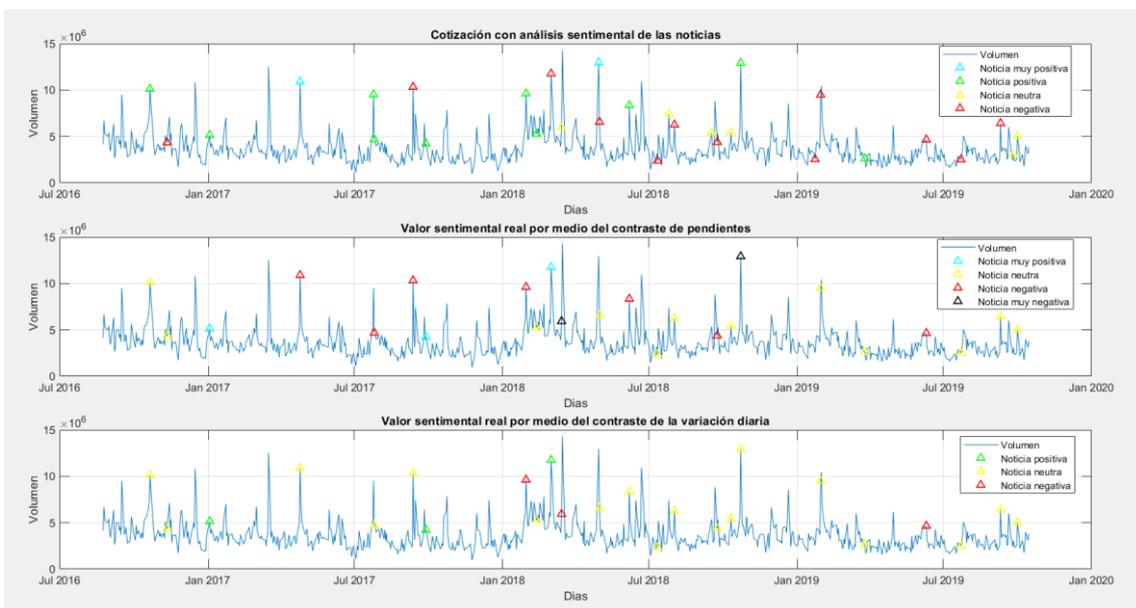


Figura 44. Comparativa del volumen con los 3 métodos

Las noticias según el volumen

Después de analizar las noticias según el precio, surgió la necesidad de ver las noticias en función del volumen.

A diferencia del precio, el volumen no depende de lo que haya sucedido el día anterior. Un día se pueden negociar 10, al siguiente día 1.000 y al siguiente 2.

Para clasificar una noticia, se han cogido los datos de los 15 anteriores días hábiles (3 semanas) a la noticia, y se ha visto a cuantas desviaciones típicas está el valor del volumen el día en el que hay una noticia. En la Tabla 15 se encuentran los límites de los volúmenes, mientras que en la Tabla 16 se encuentra la notación utilizada en la Tabla 16.

NOTACIÓN	DESCRIPCIÓN
V	Volumen el día de la noticia
M15	Media del volumen de los 15 días anteriores a la noticia
DT15	Desviación típica de los 15 días anteriores a la noticia

TABLA 16

NIVEL DE REACCIÓN	VALOR DEL VOLUMEN EL DÍA DE LA NOTICIA	COLOR
MUY ALTO	$V \geq M15 + 3 * DT15$	Cian
ALTO	$M15 + 2 * DT15 \leq V < M15 + 3 * DT15$	Verde
EN LA MEDIA	$M15 - 2 * DT15 < V < M15 + 2 * DT15$	Amarillo
BAJO	$M15 - 3 * DT15 < V \leq M15 - 2 * DT15$	Rojo
MUY BAJO	$V \leq M15 - 3 * DT15$	Negro

Tabla 15

Una vez definido todo, se puede ver en la Figura 45, en la cual se ha analizado la empresa McDonald's, como solo hay noticias con un alto nivel de reacción o con un nivel de reacción en la media.

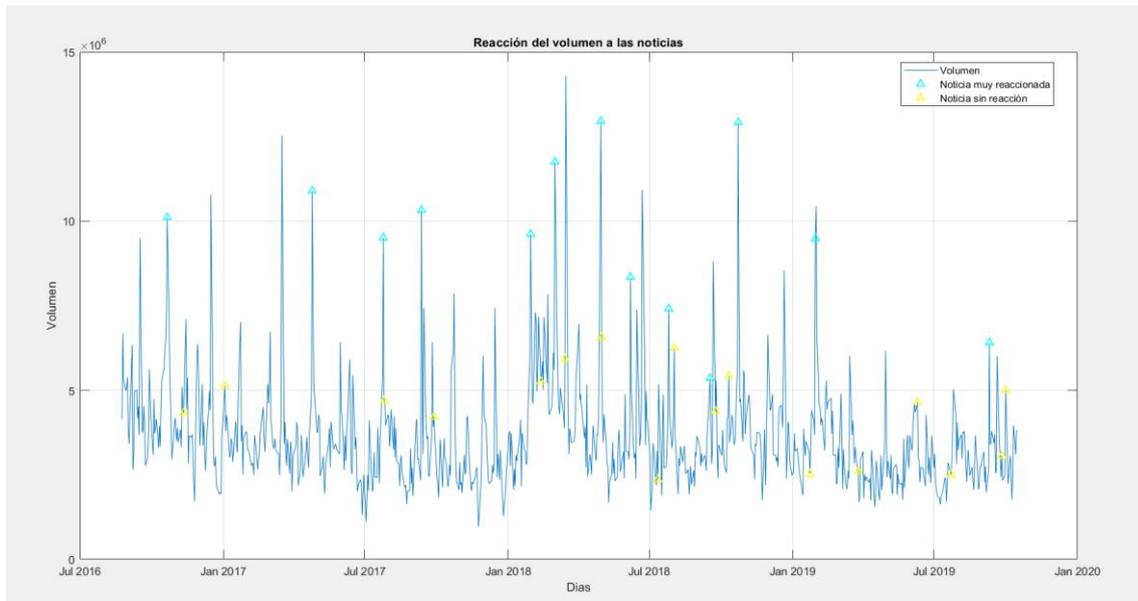


Figura 45. Volumen con noticias clasificadas según volumen

Series temporales

Según Cavalcante (2016) [2], analizar una empresa es más sencillo si separamos la gráfica de la cotización en más gráficas cuyos puntos compartan la misma tendencia. Para saber si la tendencia cambia o se mantiene, se ha hecho el contraste de hipótesis de las pendientes. Con una regresión lineal de los 30 días hábiles anteriores y los 30 días hábiles posteriores, se determina si las pendientes son diferentes al 95% de confianza. En caso de que sean diferentes, se grafica en una nueva Figura y se le añaden las noticias. A continuación, se añade un ejemplo de la compañía PepsiCo siendo analizada por la herramienta.



Figura 46. Muestra de todo el periodo de PepsiCo

La Figura 46 muestra la cotización a lo largo del periodo analizado, empezando el 24/08/2016 y terminando el 15/10/2019. Según la herramienta, esta gráfica se puede dividir en 5 partes.

A continuación, la primera parte está representada en la Figura 47. La gráfica azul representa la cotización, mientras que la recta roja representa la recta de regresión que se ha calculado. Tal y como aparece en el título, el periodo comienza el 24/08/2016 y termina el 21/08/2017. A pesar de que este periodo casi comprende un año, hay que tener en cuenta que la tendencia se mantiene durante 250 días hábiles, ya que el S&P500 cierra los fines de semanas y en algunas otras fiestas. La Figura 47, la cual muestra el primer intervalo, contiene 4 noticias que se han considerado positivas y otras 3 que se han considerado neutras, lo cual podría explicar que la regresión lineal tenga una pendiente positiva.

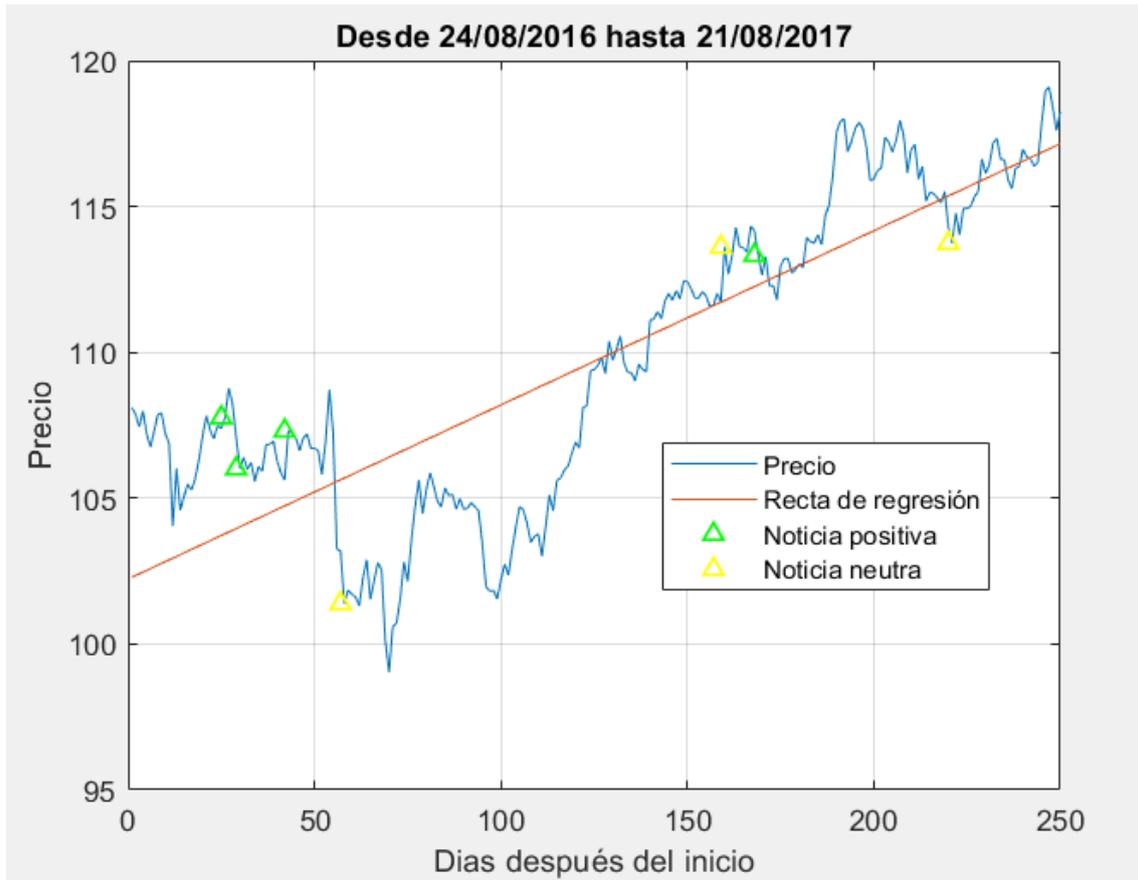


Figura 47. Primera serie temporal de PepsiCo

A continuación, la segunda parte está representada en la Figura 48. La gráfica azul representa la cotización, mientras que la recta roja representa la recta de regresión que se ha calculado. Tal y como aparece en el título, el periodo comienza el 22/08/2017 y termina el 09/05/2018. La tendencia se mantiene durante 180 días hábiles, aunque hayan pasado 260 días naturales, ya que el S&P500 cierra los fines de semanas y en algunas otras fiestas. La Figura 48, la cual muestra el segundo intervalo, contiene 6 noticias que se han considerado positivas, otras 4 que se han considerado neutras, y por primera vez aparecen 4 noticias negativas. Estas condiciones han acompañado a un pendiente de regresión lineal negativa.

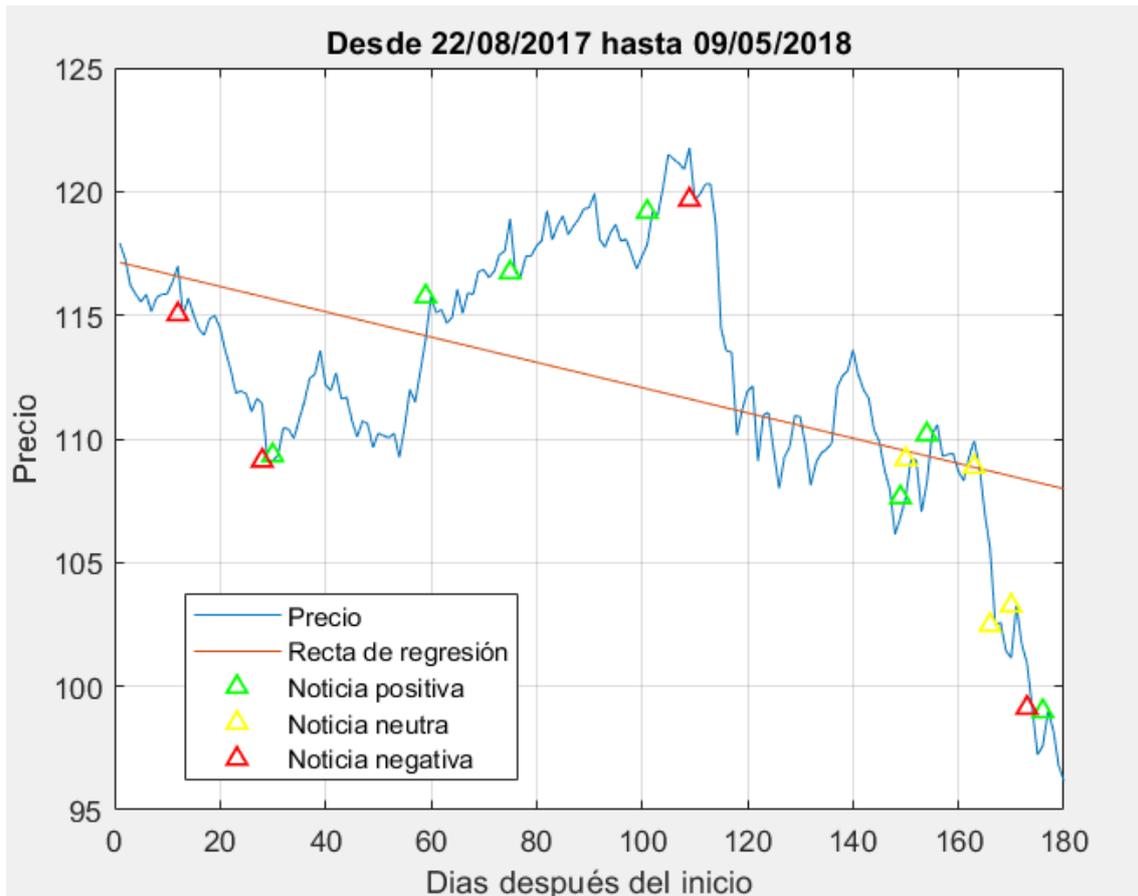


Figura 48. Segunda serie temporal de PepsiCo

A continuación, la tercera parte está representada en la Figura 49. La gráfica azul representa la cotización, mientras que la recta roja representa la recta de regresión que se ha calculado. Tal y como aparece en el título, el periodo comienza el 10/05/2018 y termina el 06/08/2018. La tendencia se mantiene durante 61 días hábiles, aunque hayan pasado 88 días naturales, ya que el S&P500 cierra los fines de semanas y en algunas otras fiestas. La Figura 49, la cual muestra el tercer intervalo, solamente contiene 3 noticias que se han considerado neutras. A pesar de ello, estas condiciones han acompañado a un pendiente de regresión lineal positiva, incluso más positiva que la pendiente en la Figura 46.

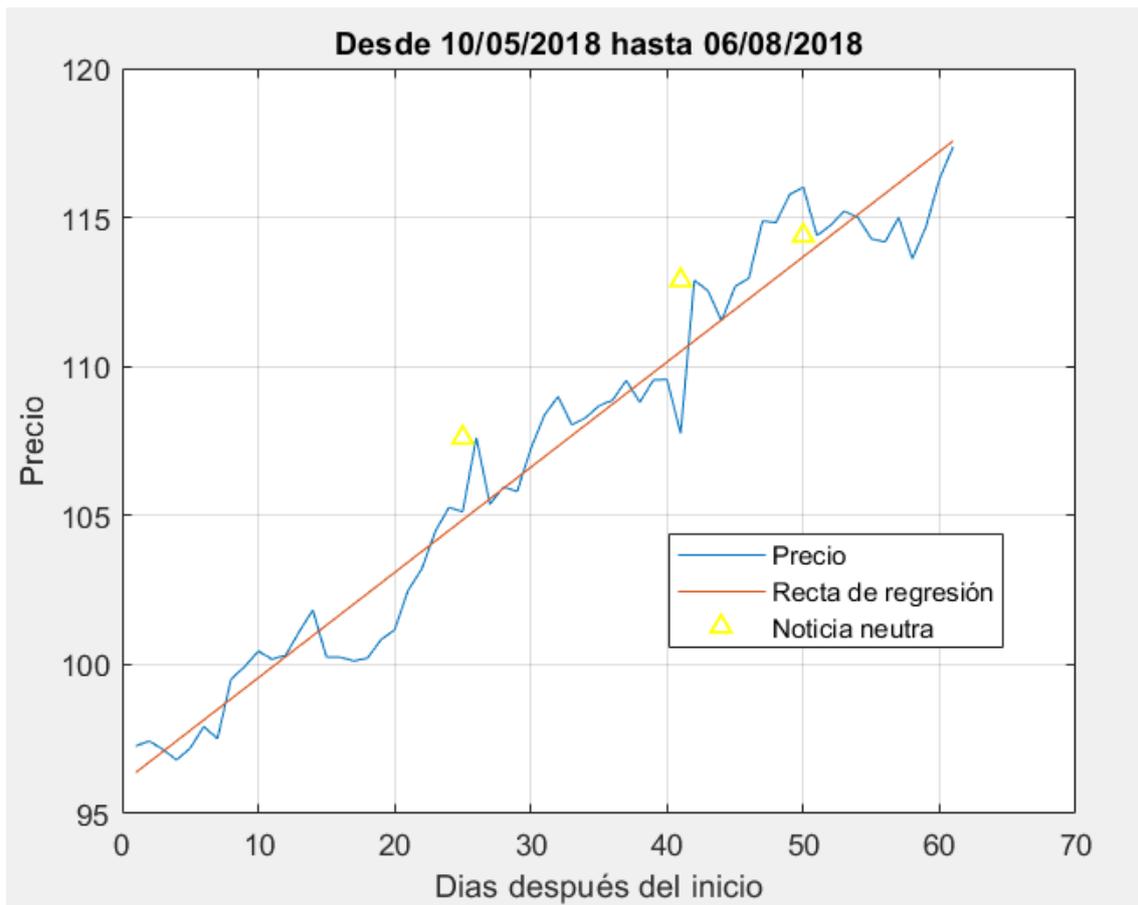


Figura 49. Tercera serie temporal de PepsiCo

A continuación, la cuarta parte está representada en la Figura 50. La gráfica azul representa la cotización, mientras que la recta roja representa la recta de regresión que se ha calculado. Tal y como aparece en el título, el periodo comienza el 07/08/2018 y termina el 14/01/2019. La tendencia se mantiene durante 110 días hábiles, aunque hayan pasado 160 días naturales, ya que el S&P500 cierra los fines de semanas y en algunas otras fiestas. La Figura 50, la cual muestra el cuarto intervalo, contiene 1 noticia que se ha considerado positiva. Estas condiciones han acompañado a un pendiente de regresión lineal negativa, pero bastante horizontal.

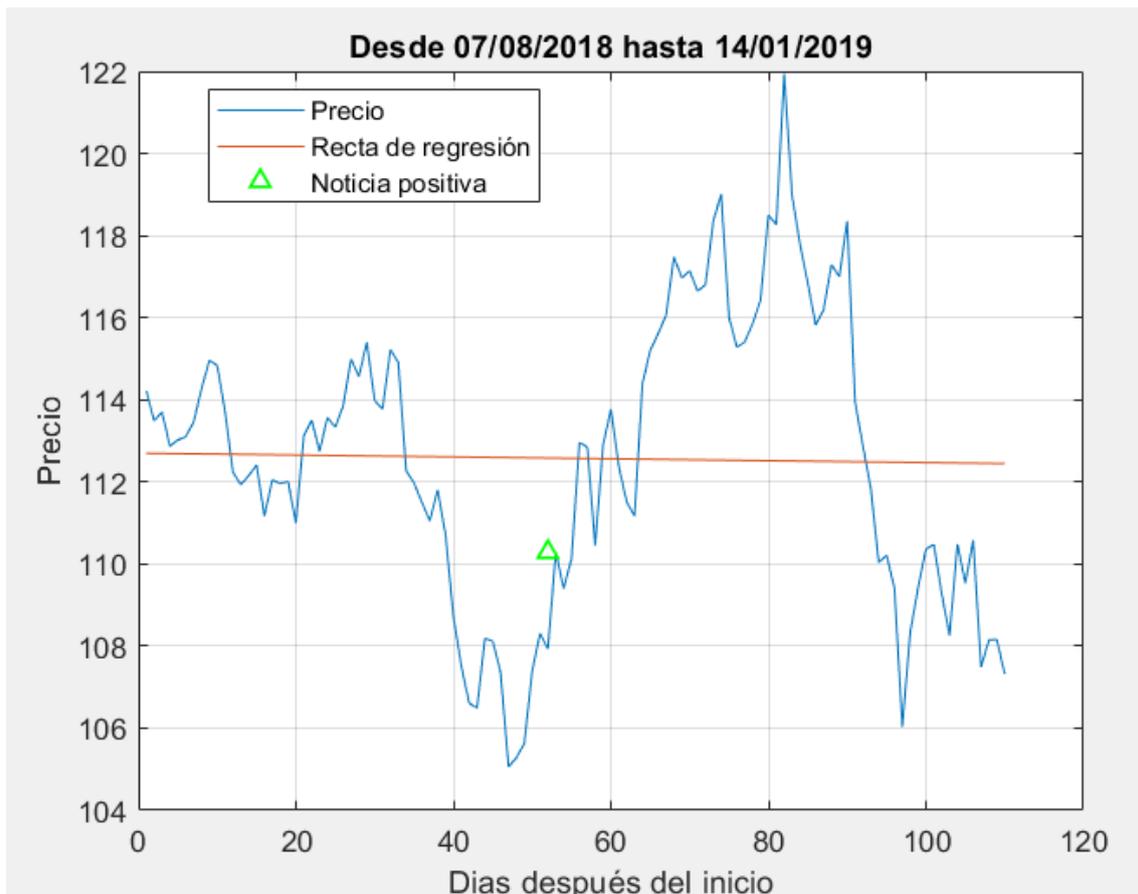


Figura 50. Cuarta serie temporal de PepsiCo

A continuación, la quinta y última parte está representada en la Figura 51. La gráfica azul representa la cotización, mientras que la recta roja representa la recta de regresión que se ha calculado. Tal y como aparece en el título, el periodo comienza el 15/01/2019 y termina el 04/10/2019. La tendencia se mantiene durante 183 días hábiles, aunque hayan pasado 262 días naturales, ya que el S&P500 cierra los fines de semanas y en algunas otras fiestas. La Figura 51, la cual muestra el quinto intervalo, solamente contiene 3 noticias, 2 de ellas se han considerado neutras y la otra se ha considerado positiva. Estas condiciones han acompañado a un pendiente de regresión lineal positiva. A pesar de que el periodo analizado termine el 15/10/2019, la herramienta ha considerado que la recta de regresión se adapta mejor terminando el 04/10/2019.

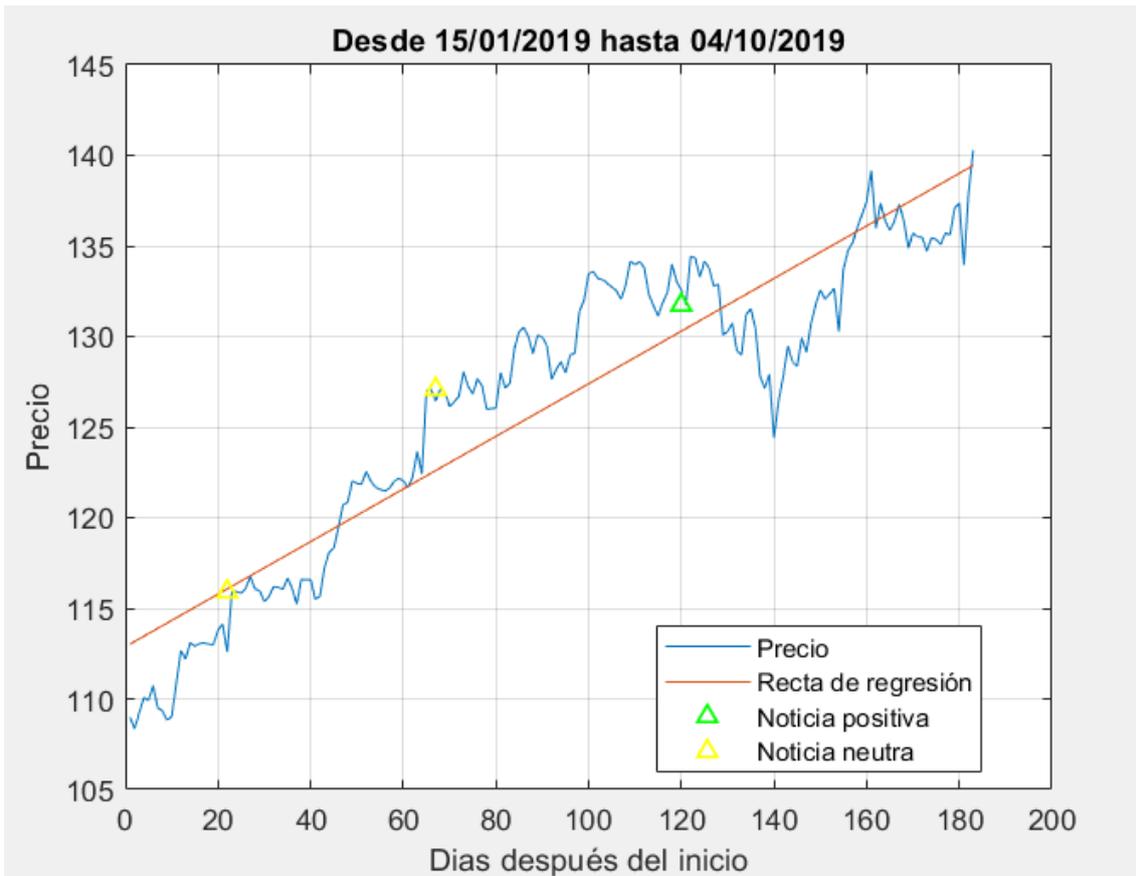


Figura 51. Quinta serie temporal de PepsiCo

Capítulo 5

Conclusiones y futuras líneas de investigación

Introducción

En este proyecto de fin de grado, se ha realizado un análisis descriptivo y en profundidad sobre la relación entre las noticias y la cotización en bolsa de las empresas, considerando en especial el S&P500.

Conclusiones

Se va a concluir en base a los resultados de los análisis de la herramienta: el efecto de las noticias en la cotización, en el volumen, el modelo Holt y las series temporales.

Efecto de las noticias en la cotización

Internet ha permitido comunicaciones casi instantáneas entre el emisor y el resto del mundo. Actualmente, cualquier persona, empresa o corporación puede publicar sus noticias sin mucho coste y llegar a todas. Esta situación ha propiciado un aumento de las noticias y de las noticias falsas, también conocidas como *fake news*. Es por este motivo que el efecto de las noticias es una parte fundamental de este proyecto, para que el usuario pueda comprobar si una noticia es como aparenta ser.

Para conseguirlo, se han propuesto dos métodos distintos: comparación de pendientes y comparación de dos normales en muestras emparejadas.

El primer método, comparación de muestras, se ha realizado siguiendo los pasos de Andrade (2014) [1], pero con una modificación, que las varianzas en ambas regresiones no eran iguales.

El segundo método, se ha hecho una comparación de dos normales en muestras emparejadas.

Los distintos métodos llevaron a distintos resultados dependiendo de la empresa analizada.



Figura 52. Comparativa del efecto de las noticias en la cotización 3M

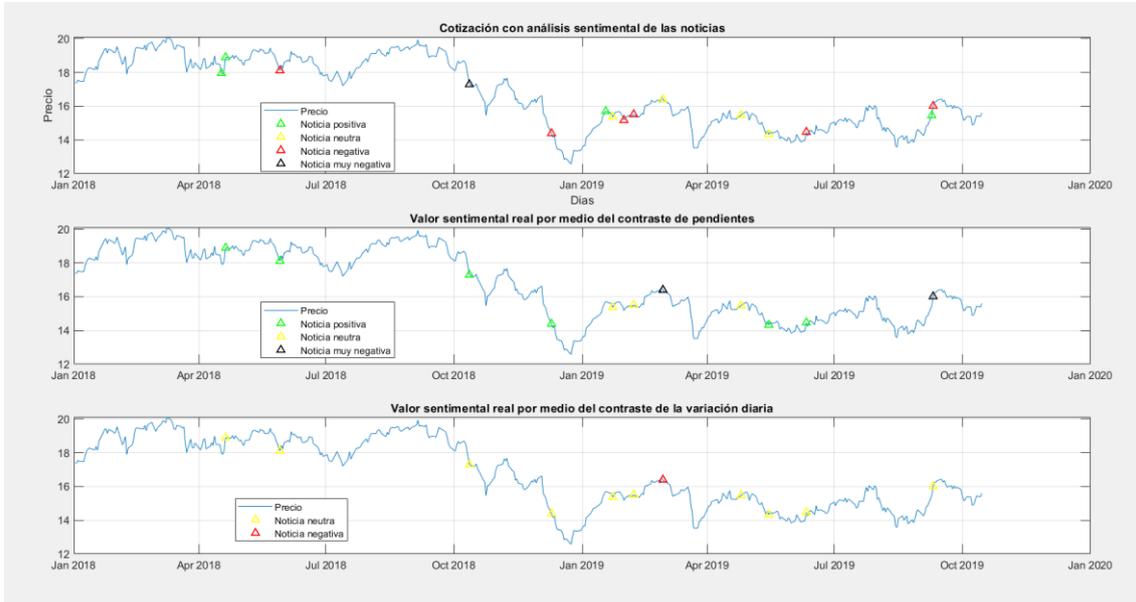


Figura 53. Comparativa del efecto de las noticias en la cotización RF

La Figura 52 corresponden a la empresa 3M, como se puede observar, los dos métodos han clasificado de la misma forma el 68% de las noticias. Por otro lado, el método de las pendientes ha mantenido la clasificación del 35% de las noticias otorgado por el valor sentimental, mientras que el método de las normales con muestras emparejadas ha mantenido la clasificación del 54% de las noticias otorgado por el valor sentimental.

La Figura 53, en la cual están representados los datos de Regions Financial, los dos métodos han clasificado de la misma forma el 27% de las noticias, bastante inferior al 68% obtenido en el anterior ejemplo. Por otro lado, el método de las pendientes ha mantenido la clasificación del 8% de las noticias otorgado por el valor sentimental, mientras que el método de las normales con muestras emparejadas ha mantenido la clasificación del 25% de las noticias otorgado por el valor sentimental.

Al tener resultados tan dispares dependiendo de la empresa que se analice, no se puede sacar una conclusión clara sobre los propios métodos, pero lo que sí se puede decir es que se necesita más información a parte de la proporcionada por el valor sentimental de la noticia para determinar si la cotización probablemente vaya a subir o probablemente vaya a bajar.

Volumen y las noticias

En este trabajo de fin de grado se ha querido ver el efecto de las noticias en la cotización, pero también hay que tener en cuenta el volumen. Por este motivo, se ha graficado el volumen junto a las noticias, para ver si se podría obtener alguna conclusión.

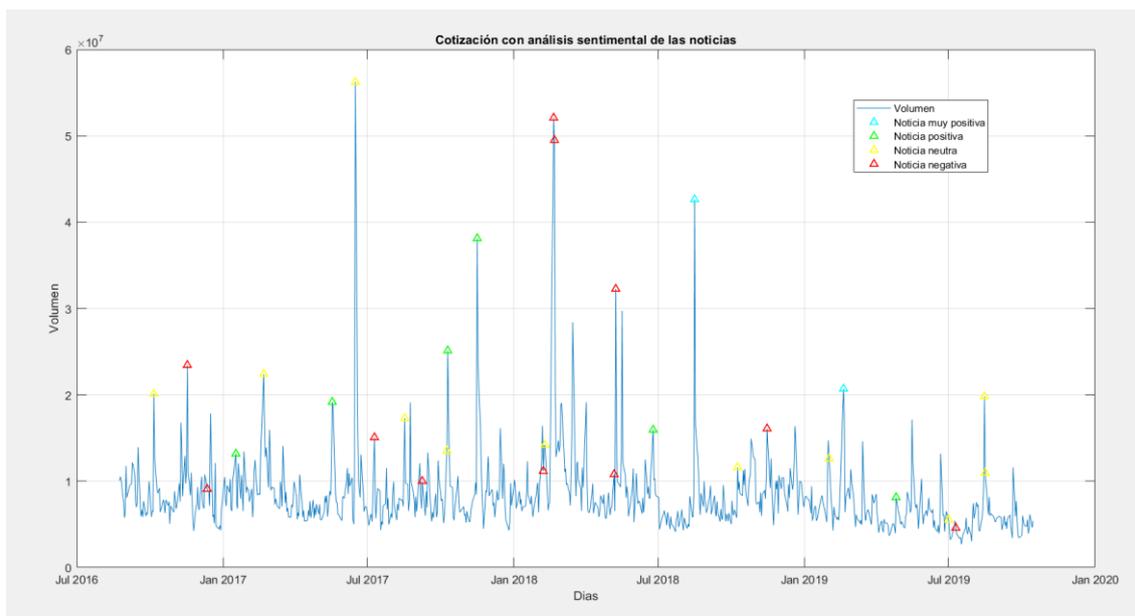


Figura 54. Volumen y noticias en Walmart

La Figura 54 muestra el volumen de la empresa Walmart, en ella se puede observar como las puntas suelen estar acompañadas por alguna noticia. El valor sentimental de estas noticias varía. En la gran mayoría de empresas analizadas no parece que haya una relación entre el valor sentimental y un pico en el volumen. Aunque sí parece que los picos en el volumen están relacionados con la aparición de noticias.

Si analizamos el volumen con el verdadero valor sentimental de las noticias, obtenemos un resultado similar. En la Figura 55, se muestra Walmart, pero las noticias han sido analizadas por medio del contraste de pendientes.

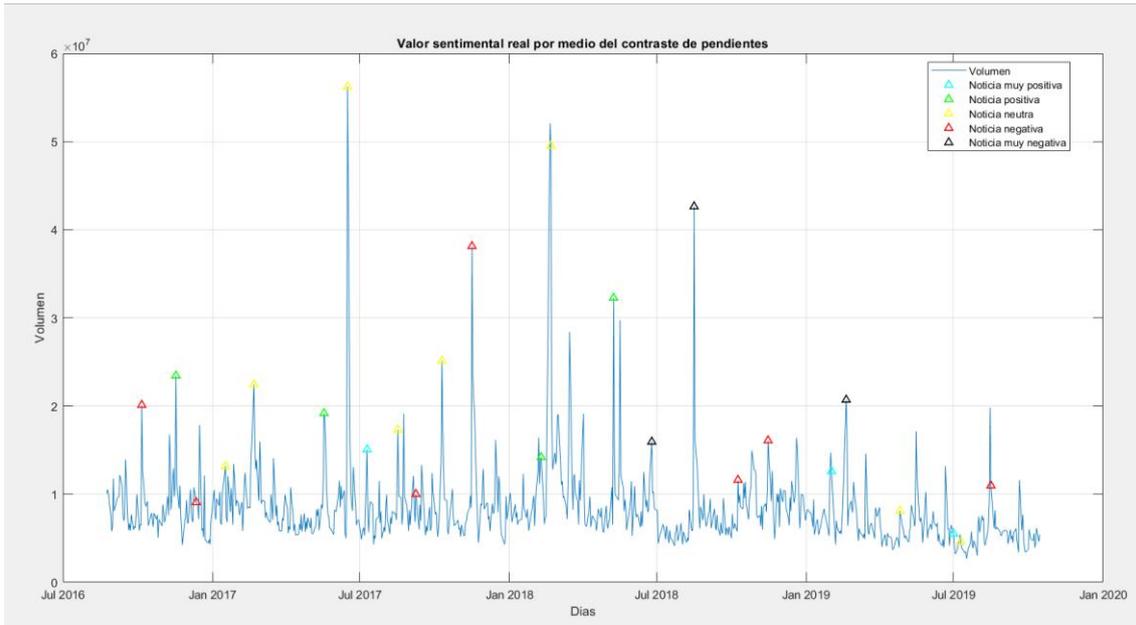


Figura 55. Volumen y noticias con contraste de pendientes en Walmart

Y en la Figura 56, se han analizado por medio del contraste de las variables emparejadas usando la variación diaria, y ocurre lo mismo, aunque en este caso hay más noticias neutras porque este modelo suele considerar más noticias como neutras que el contraste de las pendientes.

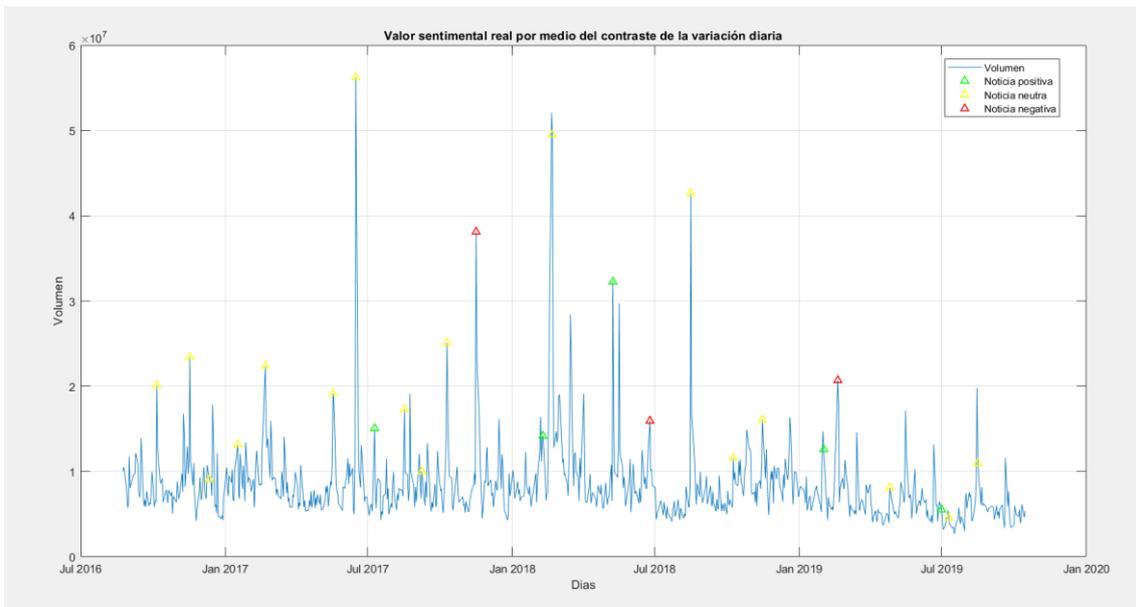


Figura 56. Volumen y noticias con contraste variables pareadas en Walmart

Con los datos analizados se puede concluir que no importa el valor sentimental de la noticia para aumentar el volumen, incluso una noticia aparentemente neutra es capaz de aumentar la cantidad de contratos negociados. Por otro lado, parece que los inversores reaccionan en el mismo día de la noticia, lo cual era una condición indispensable para que se considere un mercado eficiente.

Al concluir que no importa el valor sentimental, se ha decidido ver el efecto de la noticia desde el punto de vista del volumen. En la Figura 57, se muestra el nivel de reacción que ha causado cada noticia de la empresa Walmart. Como se puede, no hay ninguna noticia que haya causado una reacción por debajo de la media. Walmart no es el único en no tener ninguna noticia por debajo de la media, la mayoría de las empresas analizadas no las tienen. Es más, la mayoría de las empresas analizadas muestran una muy alta reacción a cualquier noticia que se presenta. En el caso de Walmart, tiene 19 noticias que han causado una muy alta reacción en el volumen de la empresa, seguido de 9 noticias que no han causado un aumento en el volumen, y por último solamente 2 noticias que han causado una alta reacción. Este esquema se repite en la mayoría de empresas analizadas.

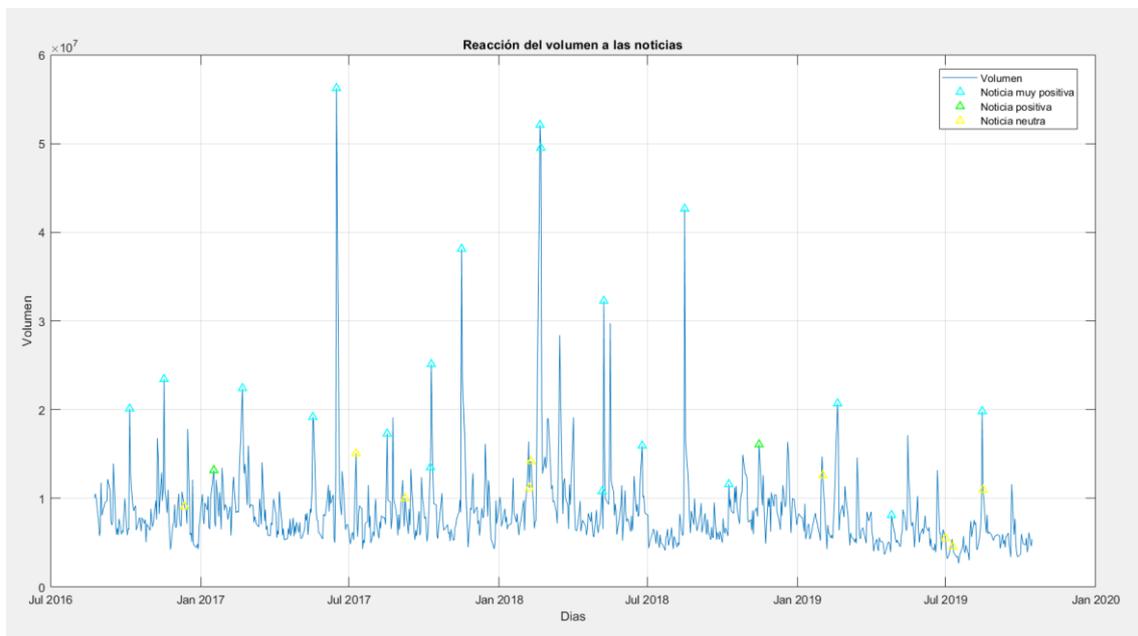


Figura 57. Volumen y noticias por volumen en Walmart

Series temporales

Como se mostró en el capítulo 4, se separó el periodo total de la cotización en partes que compartían la misma tendencia. Analizando los gráficos, se concluye de igual forma, el valor sentimental asignado a la noticia no siempre corresponde con la tendencia adoptada en la realidad. Para apoyar esta afirmación, se incluye a continuación el ejemplo de la empresa Qualcomm, la cual se ha dividido en seis partes.

En la Figura 58 se muestra el periodo desde el 24/08/2016 y termina el 15/03/2017 de la empresa Qualcomm. En este periodo se encuentran 2 noticias negativas, 1 noticia muy negativa, 4 noticias neutras y solamente 1 noticia positiva. Al haber más noticias con valor sentimental negativo, concuerda con una regresión lineal con pendiente negativa.

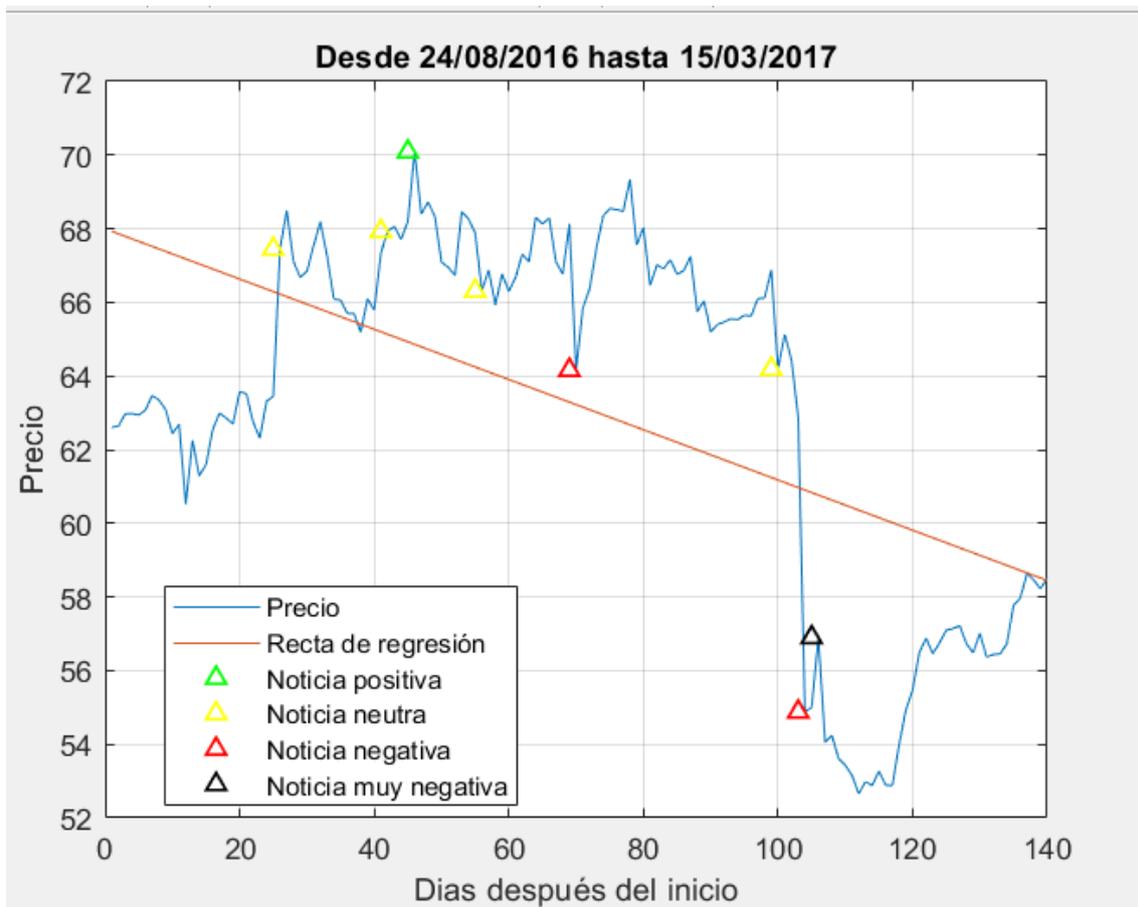


Figura 58. Primera serie temporal de Qualcomm

A continuación, en la Figura 59 se muestra el periodo desde el 16/03/2017 y termina el 05/06/2017 de la empresa Qualcomm. En este periodo no hubo noticias. La ausencia de noticias combinado con las noticias negativas del final del anterior periodo, representado en la Figura 58, pudo aumentar la confianza de los inversores al ver que ya no salían a la luz nuevas noticias que comprometieran a la empresa, lo que ocasionó que la pendiente de la regresión lineal sea positiva.

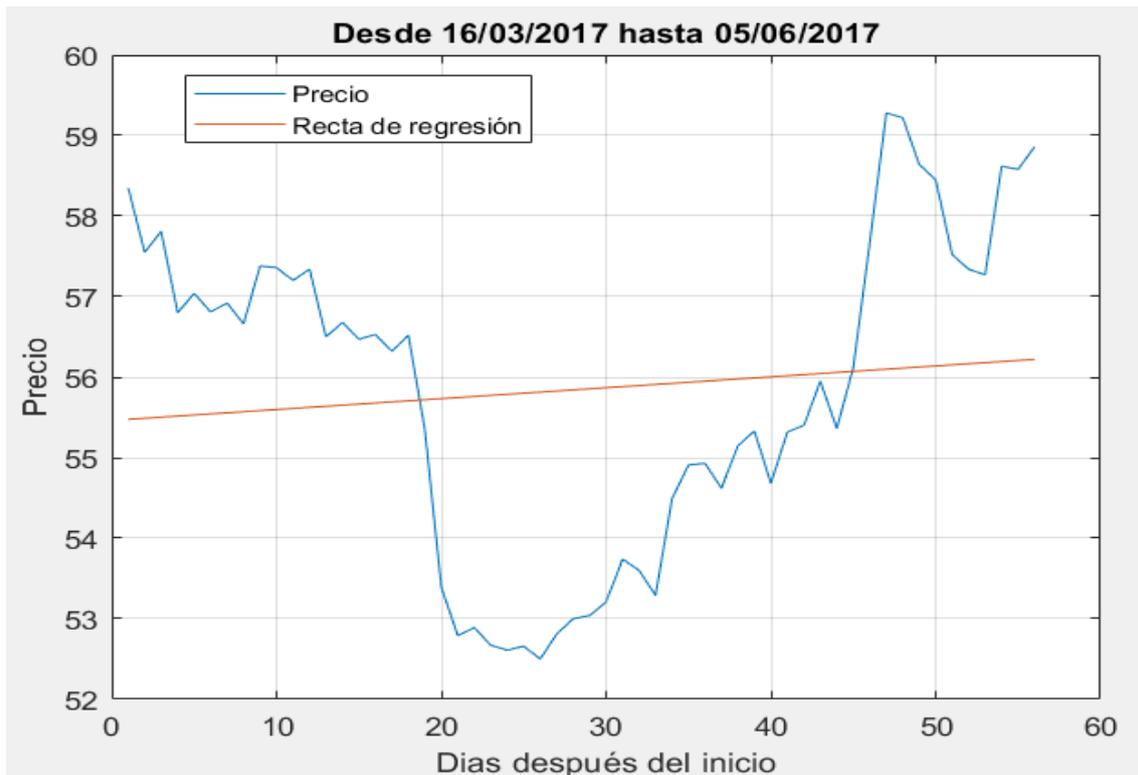


Figura 59. Segunda serie temporal de Qualcomm

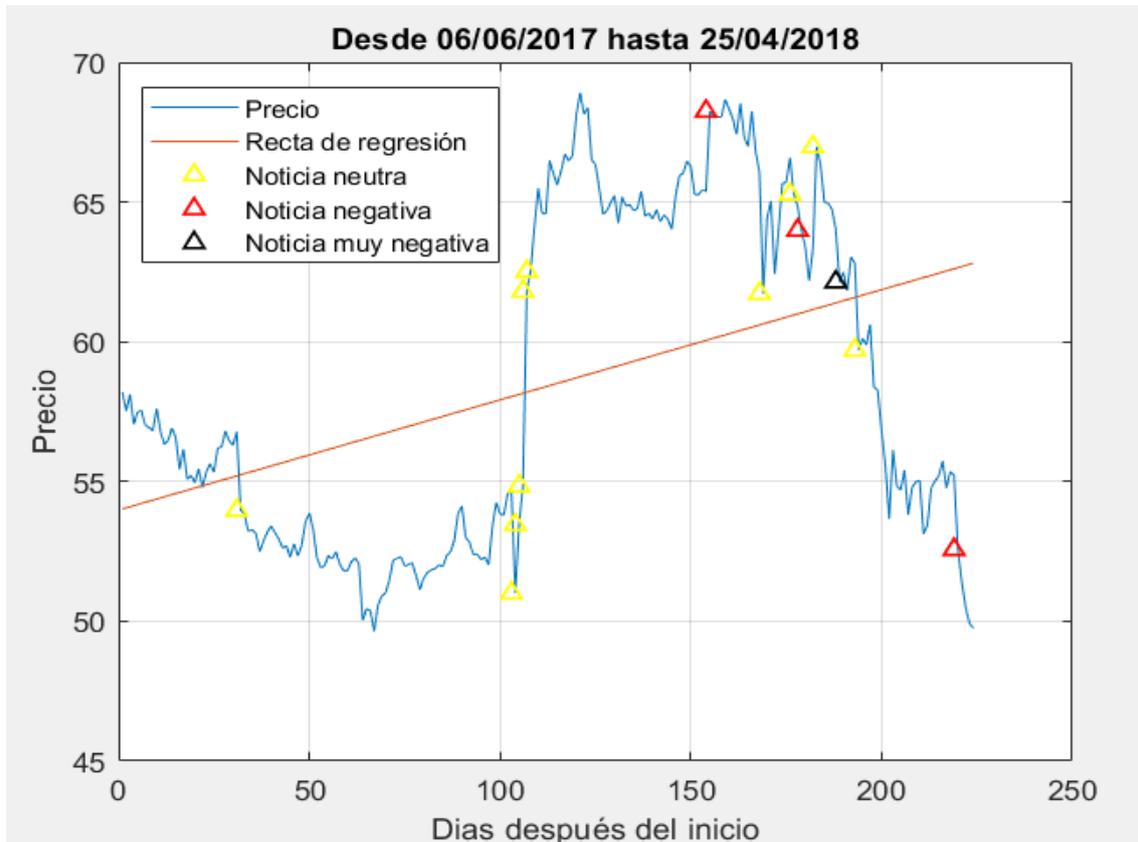


Figura 60. Tercera serie temporal de Qualcomm

En la Figura 60 se muestra el periodo desde el 06/06/2017 y termina el 25/04/2018 de la empresa Qualcomm. En este periodo hubo 10 noticias neutras, 3 negativas y 1 muy negativa. A pesar de la ausencia de noticias positivas y de la cantidad de noticias negativas, la pendiente de la regresión lineal es positiva, lo cual es sorprendente.

A continuación, en la Figura 61 se muestra el periodo desde el 26/04/2018 y termina el 14/09/2018 de la empresa Qualcomm. En este periodo hubo 3 noticias neutras, y 3 positivas. La pendiente de la regresión lineal es positiva, lo cual tiene sentido con las noticias que ha protagonizado Qualcomm.

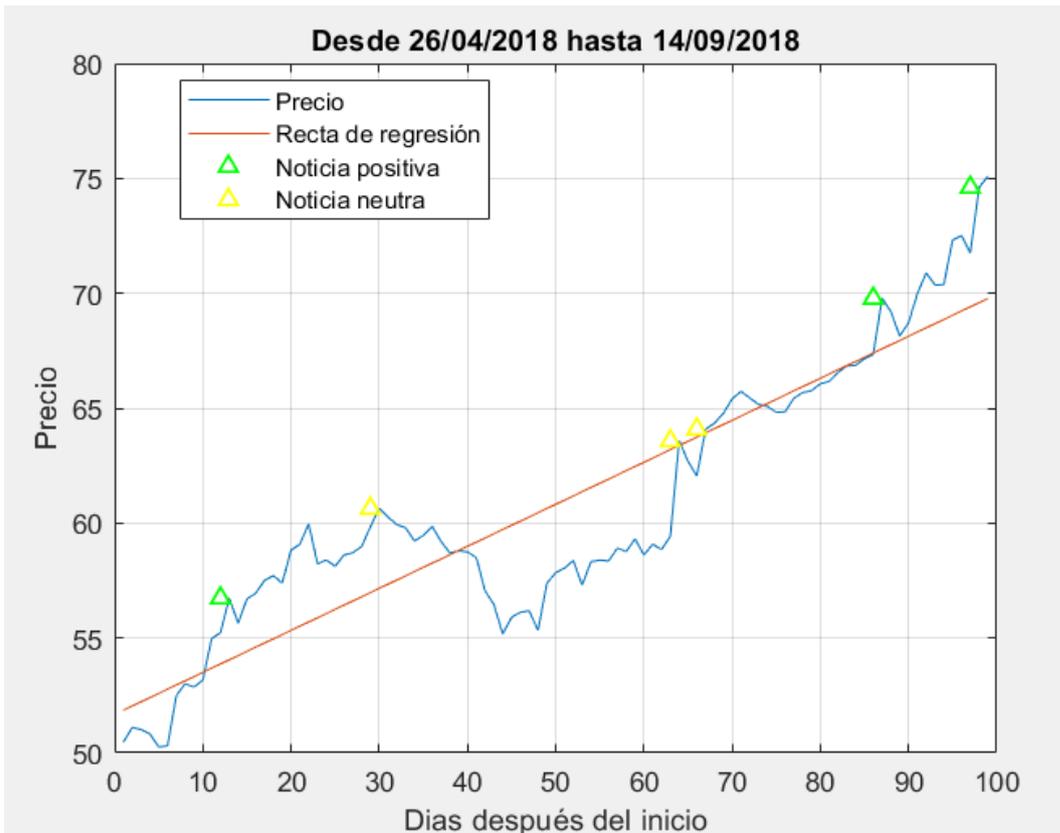


Figura 61. Cuarta serie temporal de Qualcomm

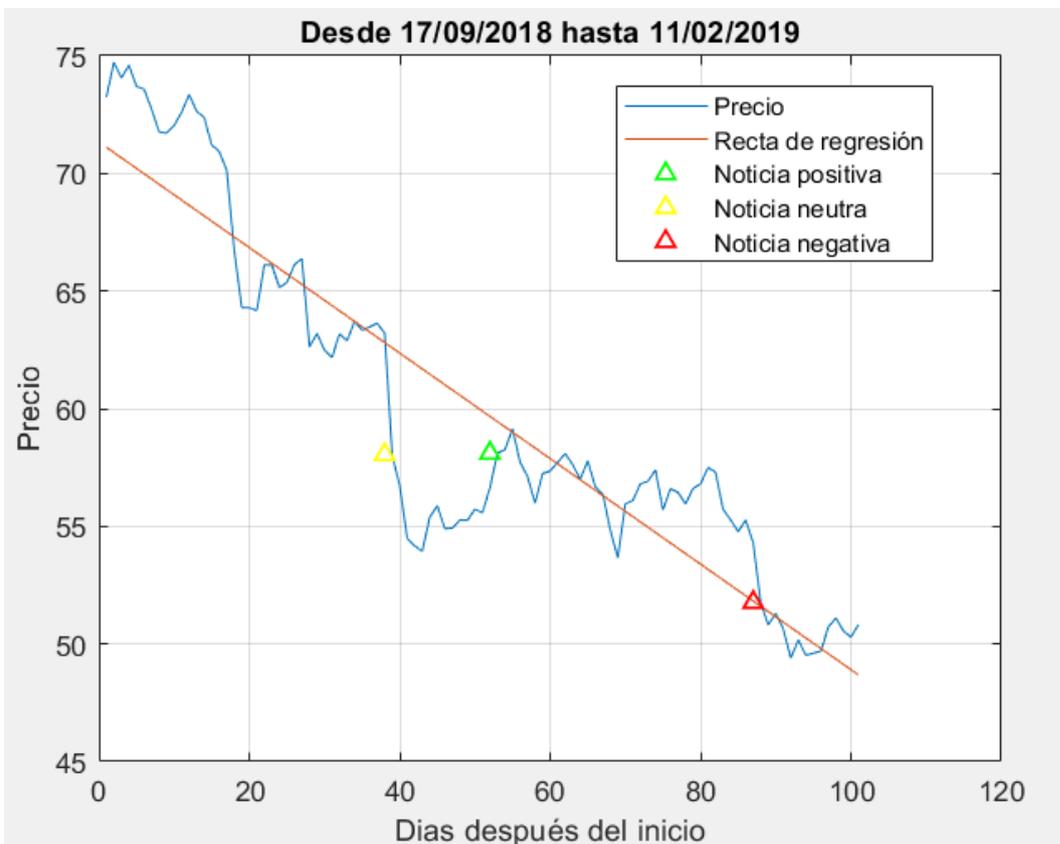


Figura 62. Quinta serie temporal de Qualcomm

En la Figura 62 se muestra el periodo desde el 17/09/2018 y termina el 11/02/2019 de la empresa Qualcomm. En este periodo hubo 1 noticias neutra, 1 negativas y 1 positiva. A pesar de que el global sea un periodo neutro ya que se podría pensar que la noticia negativa anula la positiva, la tendencia es la más negativa de todos los periodos, lo cual no es lógico ya que hay periodos con un mayor número de noticias negativas como el de la Figura 60.

Por último, en la Figura 63 se muestra el periodo desde el 12/02/2019 y termina el 11/09/2019 de la empresa Qualcomm. En este periodo hubo 2 noticias negativas y 1 noticia muy negativa. A pesar de que el periodo solo tenga noticias negativas, la pendiente de la recta de regresión es positiva, con un valor un poco inferior al de la Figura 61.

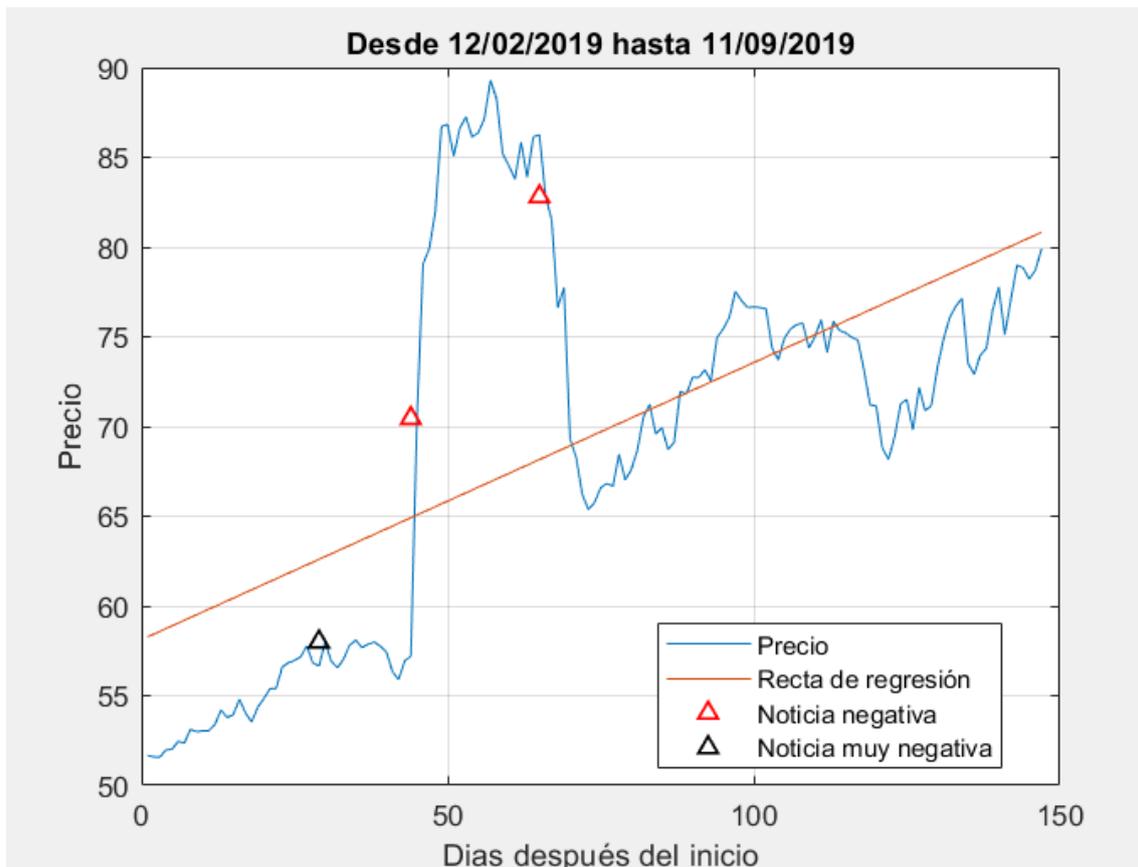


Figura 63. Sexta serie temporal de Qualcomm

Qualcomm tiene 3 periodos en los cuales las noticias sí que concuerdan con la tendencia, mientras que los otros 3 no lo hacen. El ejemplo de Qualcomm no es el único, la gran mayoría de las empresas siguen esta irregularidad, pero por temas de espacio se ha decidido no incluir todas las gráficas de todas las empresas.

Esta conclusión fortalece la conclusión del verdadero valor sentimental de las noticias. El valor sentimental es subjetivo, ya que recae en la clasificación de palabras que se consideran positivas y negativas, añadido a la falta de información y de análisis, una noticia aparentemente positiva puede resultar no ser tan positiva como se creía en un inicio. Es por ello por lo que se recomienda unir un análisis técnico en profundidad con un análisis de las noticias.

Modelo Holt

El modelo Holt no ha sido capaz de predecir correctamente la cotización de la acción en ninguna de las empresas, esto se debe a que, en todas las empresas analizadas, el modelo predictivo con un mínimo error cuadrático seguía al precio en lugar de predecir el valor del precio, y además compartían un valor del alfa superior al 0,9 en el 90,74% de los casos. Es por ello por lo que se decidió restringir el alfa a un valor inferior a 0,1 para ver si podía predecir algo mejor o si por el contrario seguía al precio. Al ver los resultados, se concluyó que el modelo restringiendo el alfa va incluso más retrasado y no predice correctamente los máximos ni mínimos de forma constante, cosa que sí lo hace el modelo sin restringir el alfa

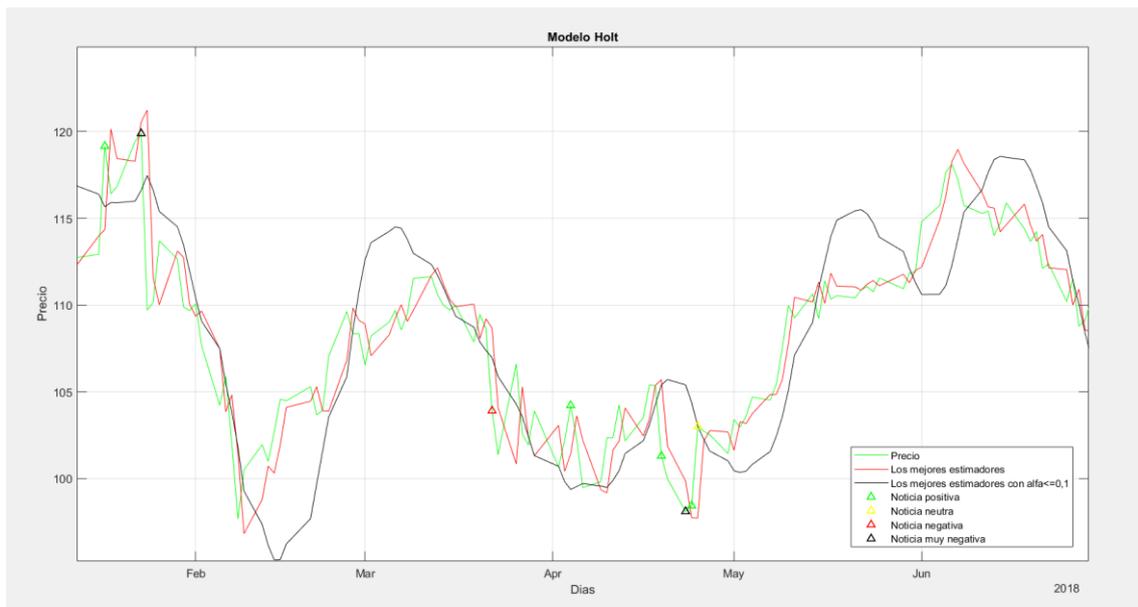


Figura 64. Modelo Holt para Mastercard

En la Figura 64, se trata de los modelos Holt calculados por la herramienta en el primer semestre del año 2017 para la empresa Mastercard. Se puede observar como el modelo con los mejores estimadores, representado por la gráfica roja, sigue en todo momento los movimientos del precio, representado por la gráfica verde. Por otro lado, también se puede observar como el mejor estimador con el alfa limitado no suele acertar con el valor de los picos, y sigue teniendo un retraso con respecto al precio.

Esta conclusión apoya la teoría del camino aleatorio, el cual estipula que no es posible determinar la futura evolución de los mercados debido a su aleatoriedad y eficiencia.

Futuras extensiones

Durante la realización de este trabajo de fin de grado, se han superado diversos obstáculos, sobre todo en la parte de la recogida de datos, es por ellos que se proponen varias futuras extensiones relacionados con este proyecto.

Completar el estudio con los datos del COVID-19

La pandemia ocasionada por el COVID-19 ha sido la causa de una crisis a nivel mundial. Numerosas PYMES han echado el cierre, las bolsas de todo el mundo se desplomaron, se han colapsado los servicios sanitarios y los gobiernos de todo el mundo han tenido que tomar medidas drásticas para evitar la propagación del virus. Por todos estos acontecimientos, se recomienda la extracción de los datos financieros y sus noticias durante este turbulento periodo, y la realización de un análisis de ellos.

El cierre de las universidades impidió la recogida de datos, y por ese motivo, este periodo no ha sido incluido en este trabajo de fin de grado.

Procesamiento del Lenguaje Natural

La parte más subjetiva de este trabajo de fin de grado ha sido la clasificación de las palabras como palabra positiva o negativa. Es por ello por lo que se propone como futura extensión la creación de un bot que procese el lenguaje natural para determinar que palabras son positivas y cuales son negativas. El Procesamiento del Lenguaje Natural (PLN) es uno de los campos que abarca la inteligencia artificial y permite a los ordenadores analizar y comprender el lenguaje humano. Se utiliza en múltiples campos, desde marketing para analizar la satisfacción de los consumidores y así poder vender mejor el producto hasta en la creación de traductores online.

Según Raj (2018) [10], Para la creación del bot, se recomienda utilizar la herramienta Python, la cual tiene una librería específicamente para el procesamiento del lenguaje natural llamada *spyCy*.

El principal problema sería el entrenamiento del bot en noticias financieras, ya que las noticias no tienen puntuación como sí lo tienen las críticas de un producto en internet o las críticas de una película. Para ello, se necesitaría encontrar un sitio web donde clasifiquen las noticias con una puntuación o encontrar a varios expertos dispuestos a clasificar varias noticias.

Bot recoge noticias

Uno de los principales problemas de la herramienta es la actualización de la información. Los datos económicos se pueden recoger y actualizar rápidamente con el complemento BQL Builder de Bloomberg, pero la recolección de las noticias se ha hecho a mano y por el momento no hay una forma rápida de actualizarla. Por este motivo, se propone como futura extensión de este trabajo de fin de grado la creación de un bot que recoja las noticias automáticamente de páginas web.

Para extraer las noticias de una página web hay varias herramientas disponibles, pero se recomienda la elección de Python por ser una herramienta conocida globalmente y por sus librerías. Una de sus librerías se llama ‘scrapy’, la cual permite la extracción de información de páginas web. Con esta librería, una persona que sepa programar podría obtener las noticias de la web de forma automática.

En este trabajo de fin de grado, se ha creado un banco de palabras para asignar el valor sentimental a cada noticia, por lo que se recomienda que el bot termine recogiendo los datos en una hoja Excel, a pesar que los datos extraídos inicialmente estén en un formato .json. Por otro lado, se puede crear un nuevo banco de palabras o utilizar el procesamiento del lenguaje natural en Python para obtener el valor sentimental de las nuevas noticias, y así combinar dos futuras extensiones en un solo trabajo de fin de grado.

Para terminar, se recomienda la creación de un diccionario con el nombre de las empresas que sean interesantes de analizar. De esta forma, si el bot encuentra el nombre de una de las empresas en el titular de la noticia, se puede clasificar como noticia de esa empresa de forma automática. Por el contrario, si no hay ninguna coincidencia, la noticia será descartada.

Bibliografía

[1] **Andrade, Jose & Estévez-Pérez, Graciela. (2014).** Statistical comparison of the slopes of two regression lines: A tutorial. *Analytica Chimica Acta*. 838. 10.1016/j.aca.2014.04.057.

[2] **Cavalcante, Rodolfo & Brasileiro, Rodrigo & Souza, Victor & Nobrega, Jarley & Oliveira, Adriano. (2016).** Computational Intelligence and Financial Markets: A Survey and Future Directions. *Expert Systems with Applications*. 55. 10.1016/j.eswa.2016.02.006.

[3] **Dunis, Christian & Williams, Mark. (2005).** Applications of Advanced Regression Analysis for Trading and Investment. 10.1002/0470013265.ch1.

[4] **Fama, E.F. (1970).** Efficient capital markets: A review of theory and empirical work, *The Journal of Finance*, 25 (2), 383–417.

[5] **Glasserman, Paul & Li, Fulin & Mamaysky, Harry. (2019).** Time Variation in the News>Returns Relationship. *SSRN Electronic Journal*. 10.2139/ssrn.3420981.

[6] **Heston, Steven & Sinha, Nitish. (2017).** News vs. Sentiment: Predicting Stock Returns from News Stories. *Financial Analysts Journal*. 73. 1-17. 10.2469/faj.v73.n3.3.

[7] **Malkiel, Burton. (2003).** The Efficient Market Hypothesis and Its Critics. *Journal of Economic Perspectives*. 17. 59-82. 10.1257/089533003321164958.

[8] **Murphy, J.J (1999),** Technical Analysis of the Financial Markets: A Comprehensive Guide To Trading Methods And Applications. Instituto de finanzas de Nueva York.

[9] **Pearson Education (2010),** Apuntes de asignatura. Estadística Empresarial II

[10] **Raj, Sumit (2018).** Building chatbots with Python. 29-32.

[11] **Rodríguez, Luis Mariano & Fermín, José Simón (2006).** Mercado eficiente y caminata aleatoria en la bolsa de valores de Caracas. 31. 12. 888-890

[12] **Sánchez, Pedro (2018).** Apuntes de asignatura. Logistic Planning: Demand Forecasting. Universidad Pontificia Comillas

[13] **Stasinakis, Charalampos & Sermpinis, Georgios. (2014).** Financial Forecasting and Trading Strategies: A Survey. 10.4324/9780203084984.

[14] **Zarnoch, Stanley J. 2009.** Testing hypotheses for differences between linear regression lines. Res. Note SRS-17. Asheville, NC: U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station. 20 p.

Anexo ODS

Este trabajo de fin de grado intenta ayudar con varios objetivos de desarrollo sostenible. En especial con el objetivo 1, fin de la pobreza, el objetivo 8, un trabajo decente y crecimiento económico y por último el objetivo 10, reducir la desigualdad en el trabajo.

El objetivo 1, fin de la pobreza, es el único objetivo social, mientras que el objetivo 8 y el 10 son objetivos económicos.

En primer lugar, hay que recalcar que, en los mercados financieros, cualquier persona, sin importar su género, ni lugar de procedencia, puede invertir en la bolsa. Además, no se necesita un mínimo capital para invertir, cualquier persona puede invertir la cantidad de dinero con la que se sienta cómodo.

Esta herramienta está orientada en el análisis de la bolsa, intentando hacer que este mundo sea más accesible a cualquier persona que quiera iniciarse. Con este objetivo, y con lo accesible que es la bolsa, se quiere reducir la pobreza intentando informar a los usuarios sobre la cotización, sus movimientos e informando de los peligros.

Por otro lado, y aunque esta herramienta esté pensada para personas que se quieran iniciar debido a su sencillez en la selección de datos, puede ser el inicio que lleve a una persona a vivir de la bolsa, el cual es un trabajo decente.

Por último, cada persona puede hacer lo que desee con sus inversiones, el mejor analista será el que haga un mejor análisis y por tanto consiga más beneficios. Por lo que no importa el género, ni el lugar de procedencia del analista.

Se quiere hacer una distinción con los típicos trabajos de fin de grado orientados a la instalación de paneles solares o a los trabajos de fin de grado con un objetivo solidario. Esos trabajos de fin de grado tienen un impacto directo en la calidad de vida de las personas, mientras que este trabajo de fin de grado, al ser una herramienta, puede llegar a más personas y además, tiene el potencial de ayudar económicamente a personas con problemas financieros para conseguir una vida mejor, en especial a las personas que están sufriendo la actual crisis del COVID-19.