

## **EL 8-M, SEGÚN EL BIG DATA**

**Grupo Gizsare Lab-CPES Arangoya BHIP<sup>1</sup>**

<sup>1</sup>*Ercilla 18, 48009 Bilbao*

### **Resumen**

Las estadísticas de búsqueda de Google pueden ser útiles para predecir acontecimientos futuros. El 6 de marzo de 2018, martes, se quiso hacer un estudio de prospectiva de la magnitud de la convocatoria del “8 de marzo”. El Diario de noticias DEIA publicó, el 7 de marzo del 2018, los resultados de este estudio Big Data sobre el Día Internacional de la Mujer “8 de marzo” en Euskadi, estudio elaborado por el Grupo Gizsare Lab-CPES Arangoya BHIP. Este extracto resume el trabajo.

**Palabras clave:** Data Mining, prospectiva, vigilancia competitiva, IC, gestión de la información, Formación Profesional, Lanbide Heziketa, Matemáticas de Bachillerato.

## **1 INTRODUCCIÓN**

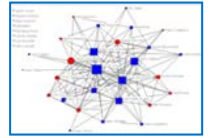
La Ciencia de los Datos, con el Data Mining y el Big Data, nos proporciona información estratégica sobre las audiencias y la visibilidad en Internet. ¿Qué nos dice, dos días antes, el Big Data de Google sobre el “8 de marzo” en la Comunidad Autónoma de Euskadi? El Día Internacional de la Mujer, antes llamado Día de la Mujer Trabajadora, rememora la lucha de la mujer por su igualdad con el hombre en la sociedad/economía y en su desarrollo integral como persona. Pero, ¿qué dice el Big Data? ¿Cuál es el interés que despierta en la Comunidad Autónoma de Euskadi la expresión “8 de marzo” en Google los días anteriores? ¿Con qué términos se asocia en los buscadores y en los medios de comunicación? Estas preguntas son el objeto de este estudio.

## **2 METODOLOGÍA**

El estudio combina un Modelo ARMA con un Análisis de Redes de las palabras clave asociadas a la expresión “8 de marzo”. La extracción de datos se ha realizado del Big Data del buscador Google. Los datos se extrajeron el 6 de marzo del 2018, es decir, dos días antes.

## **3 RESULTADOS**

Este estudio analiza la utilidad de las estadísticas de búsqueda en Google a la hora de explicar y/o predecir diferentes acontecimientos futuros. Se propone un modelo de series temporales ARMA que demuestra que las estadísticas de búsqueda de Google tienen capacidad explicativa y predictiva sobre la evolución de la intención de secundar la convocatoria del 8 de marzo (Gráfico 1, Gráfico 2 y Tabla 1).



014620 IKASTETXEA

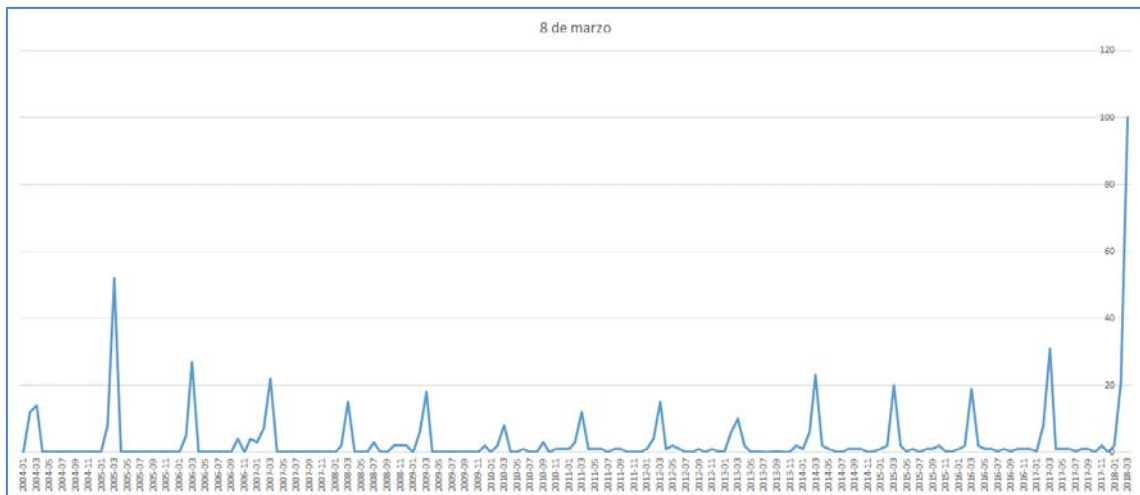


Gráfico 1. Interés de los internautas por la expresión “8 de marzo” según el Big Data de Google: Datos extraídos el 6 de marzo.

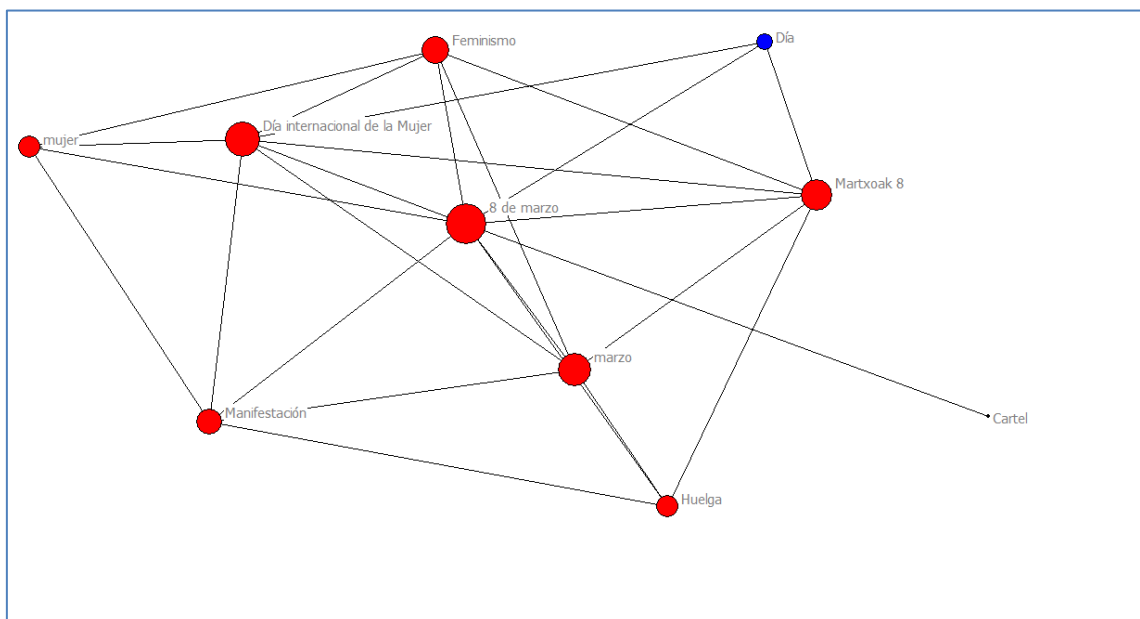
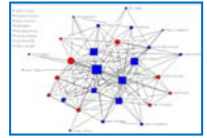


Gráfico 2. Palabras clave asociadas a la expresión “8 de marzo” según el Big Data de Google: Datos extraídos el 6 de marzo.



014620 IKASTETXEA

Tabla 1. Análisis del Patrón Estacional con Modelos ARMA (2004Q1-2018Q1)

2004-2018 Marzo	2005	2006	2007	2008	2010	2011	2012	2013	2014	2017	2018	
Q1	23.75	60.00	32.00	32.00	17.00	10.00	16.00	20.00	16.25	30.00	39.25	123.00
Q2	1.70	-	-	-	-	-	-	-	-	-	-	-
Q3	1.21	-	-	-	-	-	-	-	-	-	-	-
Q4	1.69	-	8.00	-	-	-	-	-	-	-	-	-

Fuente: Elaboración propia. Véase el modelo ARMA en el Anexo I.

#### 4 CONCLUSIONES

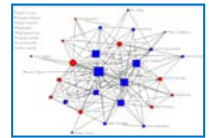
El Día de la Mujer de 2018 concitó un interés previo y un tráfico de información y consultas sin precedentes, muy superior al de otros años, búsquedas y menciones en Internet (Gráfico 1 y Tabla 1). Estas búsquedas guardan relación con el Día Internacional de la Mujer, las reivindicaciones feministas y la propia convocatoria de huelga (Gráfico 2).

#### AGRADECIMIENTOS

Los autores agradecen a la Fundación Art4pax Fundazioa su apoyo financiero al Proyecto eMETRICS (2014-2022).

#### REFERENCIAS

- [1] DEIA (2018) El 8-M, según el Big Data. *Diario de noticias DEIA*, 07.03.2018. URL: [www.deia.eus/2018/03/07/sociedad/euskadi/el-8-m-segun-el-big-data](http://www.deia.eus/2018/03/07/sociedad/euskadi/el-8-m-segun-el-big-data)
- [2] CPES Arangoya BHIP (2018) Proyecto Rutinas de Innovación Tknika RRdI. Arangoya: Noticia #662. URL: <http://arangoya.org/vernoticia.asp?IdNoticia=662>
- [3] Plaza, B. (2011) Google Analytics for measuring website performance. *Tourism Management*, 32(3), 477-481.
- [4] Zuazo, MP *et al* (2014) Methodologies for monitoring website performance: Assessing the effectiveness of AdWords campaigns on a tourist SME website. *Online Information Review*, 38(4), 8-9.

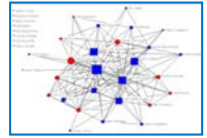


014620 IKASTETXEA

**ANEXO I: MODELO ARMA**

Dependent Variable: OCHODEMARZO  
 Method: Least Squares  
 Date: 03/06/18 Time: 10:55  
 Sample: 2004:1 2018:1  
 Included observations: 57

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	23.75000	0.767176	30.95769	0.0000
DMM_2005_Q1	36.25000	1.715458	21.13139	0.0000
DMM_2006_Q1	8.250000	1.715458	4.809212	0.0000
DMM_2006_Q4	6.307692	1.592272	3.961440	0.0003
DMM_2007_Q1	8.250000	1.715458	4.809212	0.0000
DMM_2008_Q1	-6.750000	1.715458	-3.934810	0.0003
DMM_2010_Q1	-13.75000	1.715458	-8.015353	0.0000
DMM_2011_Q1	-7.750000	1.715458	-4.517744	0.0001
DMM_2012_Q1	-3.750000	1.715458	-2.186005	0.0346
DMM_2013_Q1	-7.500000	1.715458	-4.372011	0.0001
DMM_2014_Q1	6.250000	1.715458	3.643342	0.0007
DMM_2017_Q1	15.50000	1.715458	9.035489	0.0000
DMM_2018_Q1	99.25000	1.715458	57.85627	0.0000
@SEAS(2)	-22.05357	0.869896	-25.35197	0.0000
@SEAS(3)	-22.53571	0.869896	-25.90622	0.0000
@SEAS(4)	-22.05769	0.877299	-25.14272	0.0000
R-squared	0.995514	Mean dependent var	9.846491	
Adjusted R-squared	0.993872	S.D. dependent var	19.60103	
S.E. of regression	1.534352	Akaike info criterion	3.926018	
Sum squared resid	96.52370	Schwarz criterion	4.499506	
Log likelihood	-95.89151	F-statistic	606.5291	
Durbin-Watson stat	1.646865	Prob(F-statistic)	0.000000	



014620 IKASTETXEA

Date: 03/06/18 Time: 10:59

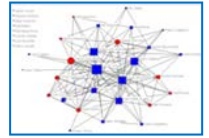
Sample: 2004:1 2018:1

Included observations: 57

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *	. *	1	0.150	0.150	1.3574	0.244
. . .	. * . .	2	-0.036	-0.060	1.4352	0.488
. . .	. . .	3	0.000	0.015	1.4352	0.697
. **	. **	4	0.289	0.292	6.7364	0.150
. * . .	. * . .	5	-0.067	-0.176	7.0266	0.219
. . .	. * . .	6	0.029	0.110	7.0818	0.313
. * . .	. * . .	7	0.191	0.194	9.5329	0.217
. **	. * . .	8	0.266	0.121	14.402	0.072
. . .	. . .	9	-0.037	-0.020	14.499	0.106
. * . .	. * . .	10	0.088	0.112	15.054	0.130
. . .	. * . .	11	0.005	-0.132	15.056	0.180
. . .	. * . .	12	0.001	-0.058	15.056	0.238
. * . .	. . .	13	-0.069	0.005	15.419	0.282
. . .	. . .	14	0.083	-0.037	15.952	0.316
. . .	. * . .	15	-0.049	-0.113	16.142	0.373
. . .	. . .	16	-0.041	-0.039	16.280	0.434
. * . .	. * . .	17	-0.091	-0.109	16.981	0.456
. . .	. * . .	18	-0.010	-0.062	16.989	0.524
. . .	. * . .	19	0.045	0.185	17.165	0.579
. . .	. * . .	20	-0.042	-0.096	17.322	0.632
. . .	. * . .	21	0.018	0.122	17.351	0.690
. . .	. * . .	22	0.037	0.085	17.486	0.736
. . .	. . .	23	0.043	0.034	17.667	0.775
. * . .	. . .	24	-0.115	-0.057	19.018	0.751

Breusch-Godfrey Serial Correlation LM Test:

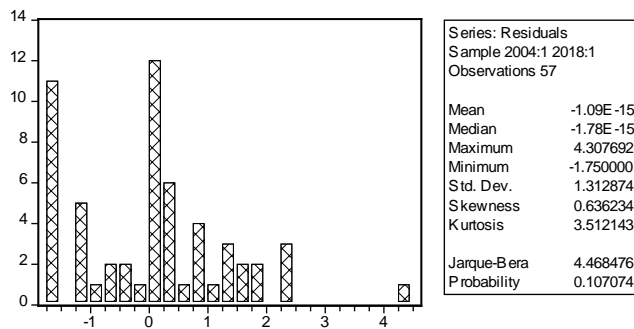
F-statistic	0.622792	Probability	0.541694
Obs*R-squared	1.764126	Probability	0.413928



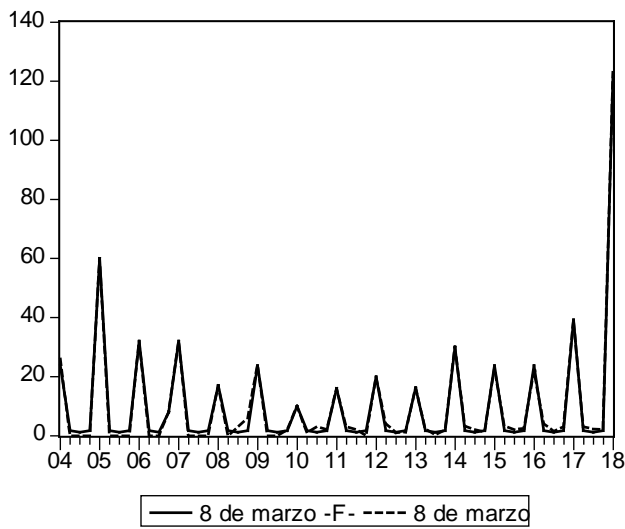
014620 IKASTETXEA

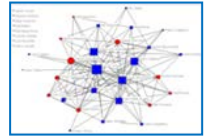
White Heteroskedasticity Test:

F-statistic	0.466775	Probability	0.944112
Obs*R-squared	8.314151	Probability	0.910571



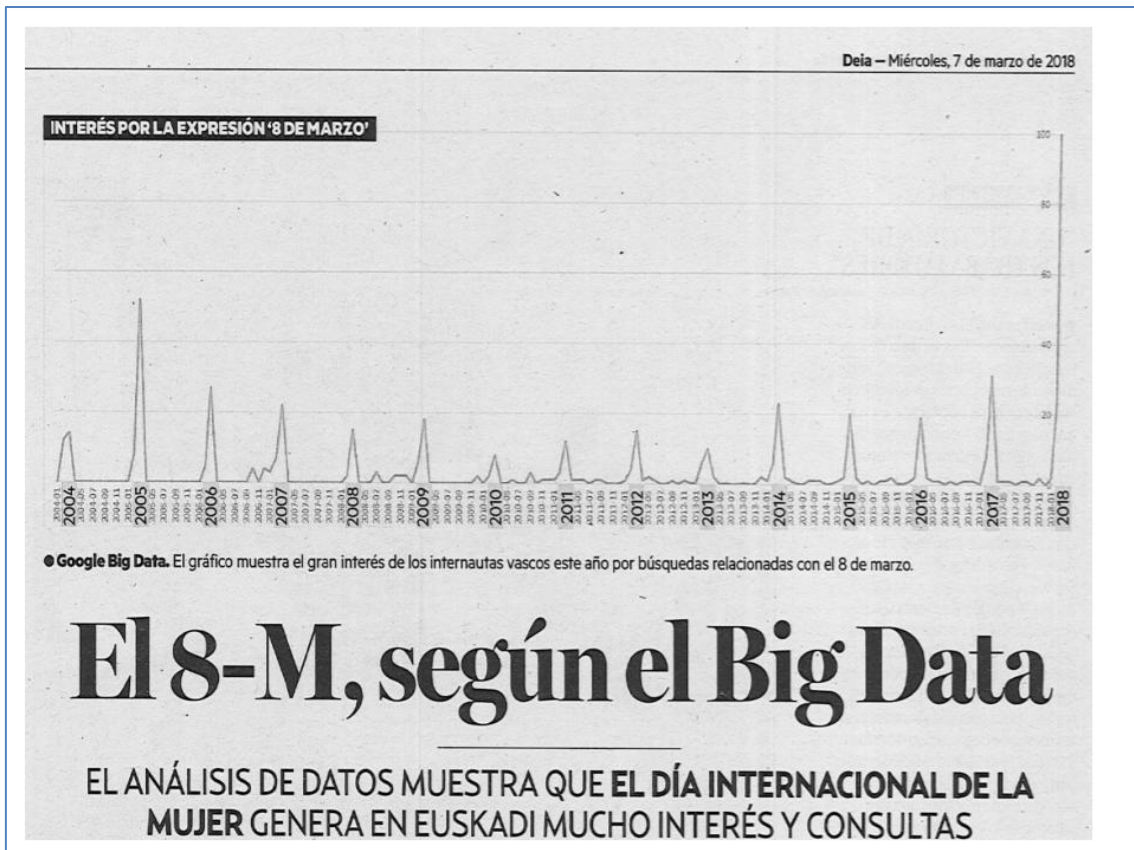
Modelo predictivo según ARMA:





014620 IKASTETXEA

Anexo II: DEIA (2018) El 8-M, según el Big Data. Diario de noticias DEIA, 07.03.2018.



Fuente: DEIA (2018) El 8-M, según el Big Data. Diario de noticias DEIA, 07.03.2018.

URL: [www.deia.eus/2018/03/07/sociedad/euskadi/el-8-m-segun-el-big-data](http://www.deia.eus/2018/03/07/sociedad/euskadi/el-8-m-segun-el-big-data)