

# CIENCIA DE DATOS EN ESPAÑA

## COBERTURA DE LA CIENCIA DE DATOS POR LOS MEDIOS DIGITALES EN ESPAÑA

### DATASCIENCE SPAIN



Proyecto financiado por la Fundación Española para la Ciencia y la Tecnología (FECYT) en la Convocatoria de ayudas para el fomento de la cultura científica, tecnológica y de la innovación 2019-2020. [FCT-18-13437]



UNIVERSIDAD  
DE SALAMANCA

## EQUIPO DE TRABAJO

Investigador principal:  
Dr. Carlos Arcila Calderón  
carcila@usal.es

Investigadores:  
Dr. Juan José Igartua Perosanz  
Dr. Félix Ortega Mohedano  
Dra. Beatriz González de Garay Domínguez  
Dr. Francisco Javier Frutos Esteban  
Dra. María Marcos Ramos  
Dra. Lifen Cheng Lee  
Patricia Sánchez Holgado  
Francisco Javier Jiménez Amores  
David Blanco Herrero  
Maximiliano Frías Vázquez

Colaboradores:  
Juan Pablo Latorre Jiménez

## INTRODUCCIÓN

Como se viene destacando, la Ciencia de Datos es uno de los campos de estudio más novedosos, controvertidos y con mayor proyección de la actualidad, y su inmersión en todos los ámbitos de la vida es innegable. No obstante, se viene observando que el nivel de conocimiento por parte de los ciudadanos todavía no es completo, y que los periodistas reclaman esfuerzos también en esta línea.

Debido a esto, conviene abordar también los contenidos mediáticos sobre Ciencia de Datos, una de las principales fuentes de información y conocimiento sobre la materia para los ciudadanos.

Con esto, se aborda un análisis mediático de la cobertura, el tratamiento y la divulgación de la Ciencia de Datos en los medios digitales españoles.

# 03

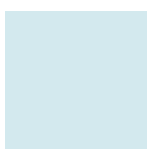
## DESAFÍOS

“¿Cómo es el tratamiento mediático de los contenidos de Ciencia de Datos?”

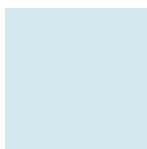


# 03

## FACT SHEET



El tratamiento que se le da al Big Data y la Inteligencia Artificial, aunque con una magnitud débil, es positivo.



Los medios generalistas son los que más frecuentemente cubren el Big Data y la Inteligencia Artificial.



La cobertura de estas materias se realiza en gran medida mediante noticias de economía, destacando también con frecuencia los riesgos de estas tecnologías y su penetración en la esfera privada.

## METODOLOGÍA

Para analizar el tratamiento mediático de la Ciencia de Datos, se diseñó un libro de códigos que, por medio de un análisis de contenido, permitió clasificar las noticias de prensa digital española sobre dicho tema.

## MUESTRA

Para abordar este trabajo se extrajeron todas aquellas noticias digitales publicadas entre enero y febrero de 2020 que mencionaran los conceptos de Big Data e Inteligencia Artificial.

El total de la base de datos contenía 21486 registros de noticias que cumplían esos requisitos, pero para el análisis que aquí se reporta se utilizó una muestra aleatoria de 750 noticias de prensa digital.

Esta clasificación se realizó en la plataforma Doccano, que permite el etiquetado en línea. En este proceso se descartaron 350 noticias que no eran contenidos noticiosos directamente.

Posteriormente, las 400 noticias seleccionadas fueron codificadas en el programa SPSS, según el libro de códigos diseñado *ad hoc*.

## 03

## MEDIDAS

**El libro de códigos recoge 14 variables con el objetivo de conocer y analizar el tratamiento mediático de la Ciencia de Datos.**

Se midió la fiabilidad de todas las variables, obteniendo resultados satisfactorios en el estadístico *Alpha de Krippendorff*.

Algunas variables son generales, de carácter nominal, como el Tipo de Medio que publica la noticia: Medio digital generalista, Medio digital especializado en ciencia y tecnología, Blogs o foros, Medios digitales especializados y Otros. Esta variable se utilizó con la finalidad de analizar si los medios especializados en ciencia y tecnología hablaban más de Ciencia de Datos que los medios generalistas.

La variable Temática trata de la sección o del tema en general de la noticia, en este caso se hizo una división en cuanto a las temáticas más importantes, las opciones de esta variable son: Economía, Política, Ciencia, Tecnología y Salud, Sociedad y Educación y Otros. Esta variable se elaboró con el fin de distinguir qué temáticas se vinculaban con más frecuencia a la Ciencia de Datos en la agenda mediática.

# 03

## MEDIDAS

Otras variables son las de aspectos formales del texto y están conformadas por variables independientes y de control.

La variable Extensión de la Noticia, se emplea para analizar si las noticias de Ciencia de Datos tienen un espacio amplio dentro de la agenda mediática. Esta variable fue medida a través de una escala de 4 ítems: de 10 a 25 líneas, de 26 a 45 líneas, de 45 a 60 líneas y más de 60 líneas.

La variable Objeto, hace referencia si en la noticia se habla de Big Data, Inteligencia Artificial o habla de ambas.

La variable Visión establece si la postura de los medios frente a la Ciencia de Datos es negativa, neutra o positiva. Se trata de una variable que sirve para medir la actitud que se transmite sobre la Ciencia de Datos y la percepción de beneficios y riesgos.

La variable Posición estudia si la noticia está a favor o en contra de la tecnología en cuestión, y está codificada en términos de ataque, defensa, o no toma posición.



## MEDIDAS

También se midieron dos aspectos relativos a las aplicaciones del Big Data y la Inteligencia Artificial cuyos riesgos y beneficios habían sido analizados en el primer estudio. Para ello, se indicaba si cada una de las aplicaciones seleccionadas (8 en el caso del Big Data y 7 en el caso de la Inteligencia Artificial) aparecía o no en el texto; a esto le seguía una variable que codificaba el atributo dado a esta aplicación, en términos de beneficio, riesgo o neutralo.

La variable Comprensibilidad hace referencia a si un codificar humano es capaz de entender o no un texto presentado. Se divide en 4 ítems:

No se entiende. Sin una preparación educativa suficiente en este campo o con ayuda especializada, es casi imposible entender la idea desarrollada.

Se entiende con dificultad. Una persona no puede entender estos textos por sí sola sin recurrir a ayuda o materiales auxiliares.

Se entiende relativamente bien. Una persona puede entender en general la idea principal de la noticia, aunque tenga ciertas dudas o dificultades en la comprensión en algunas partes.

Se entiende muy bien. Una persona puede entender totalmente la idea principal del texto, sin esfuerzo extra.

# 03

## MEDIDAS

La variable Relación con la vida diaria buscaba medir si los contenidos relacionaban la tecnología en cuestión con la vida diaria, y se evaluó a través de la presencia o ausencia en la noticia de menciones a la relación de la Inteligencia Artificial y el Big Data con los siguientes aspectos: "Mi conducta como consumidor y usuario", "Mi comprensión del mundo", "Mis relaciones con otras personas", "Mi profesión o ámbito laboral" y "Mi formación de opiniones políticas y sociales".

Las variables Género y Encuadre, hacen alusión a la franja informativa de las noticias y al tipo de lenguaje periodístico que se usa en ellas.

La variable Género está codificada en tres opciones: Informativo, Interpretativo e Híbrido.

La variable Encuadre se refería al tipo de lenguaje o desarrollo que se le daba a la noticia. Dividida en: información, descripción, narración y cita de acontecimientos.

## RESULTADOS

“¿Cuál es el tratamiento que proponen los medios digitales del Big Data y de la Inteligencia Artificial?”

Los resultados exploratorios evidenciaron que la Ciencia de Datos es un tema que suele tratarse de manera extensa en los medios analizados ( $M=2.27$ ;  $DT=,88$ ), además de hacerlo en conexión con todas la temáticas y aspectos de los diarios digitales.

Por otra parte, la mayoría de las noticias trataban temas de economía y eran emitidas por medios digitales generalistas.

Esto nos daría un primer acercamiento a los dos resultados más importantes de la investigación.



## RESULTADOS

“¿Cuál es el tratamiento que proponen los medios digitales del Big Data y de la Inteligencia Artificial?”

Para responder a la pregunta se hizo una regresión lineal múltiple, donde la variable Visión actuaría como variable independiente y el resto actuarían como variables independientes.

Este análisis permitió observar que los atributos que los medios digitales utilizan para referirse a la Ciencia de Datos tienden a ser positivos cuando las noticias están ligadas a la economía, mientras que es más frecuente que sean negativos cuando están ligadas a la vida privada de los individuos.



## RESULTADOS

“¿En que difiere el tratamiento del Big Data al de la Inteligencia Artificial?”



Aunque los atributos afectivos no mostraron diferencias entre el Big Data y la Inteligencia Artificial, sí se observó que el Big Data es tratado en mayor medida por medio de géneros interpretativos, jugando un papel más importante para los medios digitales, debido a su cantidad de noticias y a la extensión de éstas, y siendo un tema que no se trata o relaciona bien con aspectos de la vida cotidiana o esfera de los individuos.

Por su parte la Inteligencia Artificial goza de menos presencia en la agenda mediática y su tratamiento se hace por medio de géneros y encuadres informativos, más vinculados a la ciencia, la salud y la tecnología.

03

# 44%

*de contenidos se publican en medios generalistas, frente a un 13% en medios especializados en ciencia y tecnología.*

# 53.3%

*La fuente más citada son expertos profesionales y empresas con un 53.3%*

03

**54.8%**

*de los contenidos relacionan el  
BD y la IA con “la profesión o  
el ámbito laboral”*

**72.4%**

*de los textos se clasifican con un  
nivel de comprensibilidad  
bueno*

# 03



# 41%

Hay más volumen de contenidos publicados sobre **Big Data** que sobre **Inteligencia Artificial**.

de los textos hablan o están relacionados con **temas económicos**



**03****COMPRENSIBLE****NO  
COMPRENSIBLE**

Los textos informativos comprensibles:

Están escritos en tercera persona.

Visión objetiva.

No utilizan el humor.

Perspectiva optimista.

Destacan los beneficios.

Los textos informativos no comprensibles:

Presentan largas frases

Usan terminología técnica.

No aportan explicación de los tecnicismos.

No usan descripciones explicativas.

Van dirigidos a personas con conocimientos previos.



Proyecto financiado por la Fundación Española para la Ciencia y la Tecnología (FECYT) en la Convocatoria de ayudas para el fomento de la cultura científica, tecnológica y de la innovación 2019-2020. [FCT-18-13437]

OBSERVATORIO DE LOS CONTENIDOS AUDIOVISUALES  
UNIVERSIDAD DE SALAMANCA

Facultad de Ciencias Sociales. Campus Miguel de  
Unamuno. Edificio F.E.S. Paseo de Francisco Tomás y  
Valiente, s/n. 37007. Salamanca. España  
[www.ocal.es/datasciencespain](http://www.ocal.es/datasciencespain)

Enero 2021. This work is licensed under the Creative  
Commons Attribution-NonCommercial-ShareAlike 4.0  
International License. Diseño base e imágenes de Canva.

