**UNIVERSIDAD AUTONOMA DE MADRID**

**ESCUELA POLITECNICA SUPERIOR**

** **

**Grado en Ingeniería Informática**

**TRABAJO FIN DE GRADO**

**Sistema para la Detección**

**de Indicios de Delitos de Odio.**

**Juan Carlos Pereira Kohatsu**

**Tutor:**

**Ponente:**

**JUNIO 2017**

Resumen.

Las llamadas *redes sociales* constituidas por plataformas tales como **Facebook**™**,** **Twitter**™ que operan sobre Internet constituyen el soporte de los *medios (de comunicación) sociales* que facilitan el intercambio y la discusión de información, experiencias y opiniones entre individuos de manera rápida y masiva, nunca antes vista en la historia de la humanidad.

El abanico de medios sociales abiertos al uso público es variadísimo y creciente y sus usos son múltiples:

* artículos en *wikis,*
* opiniones sobre la calidad de hoteles, restaurantes ([tripadvisor](http://www.tripadvisor.es), [yelp](https://www.yelp.es)),
* contactos sociales y profesionales ([facebook](https://es-es.facebook.com/), [linkedin](https://es.linkedin.com/)),
* *blogs* o bitácoras *web* ([wordpress](https://es.wordpress.com)).

Ciertamente, como todo lo nuevo, la explosión de los medios sociales ha tenido consecuencias que han sido valoradas tanto positiva como negativamente para el conjunto de la sociedad.

Entre los efectos generalmente considerados como negativos, los medios sociales han hecho persistentemente *‘visibles’* algunas actitudes de ciertos grupos sociales que, hasta la fecha, solo se mostraban de una manera velada y/o esporádica. Entre ellas destacan las que se traducen en ataques a personas o colectivos en razón de su pertenencia a un determinados grupos definidos por características de nacionalidad, preferencias sexuales, raza, religión…

Este fenómeno junto con un cambio de actitud frente a ciertas conductas o grupos sociales ha motivado que, en muchos países, surja una nueva categoría delictiva: los llamados *delitos de odio* que, en España han sido regulados en 2015 mediante modificación del Código Penal, que, en su artículo 510 dice:

(Ley Orgánica 1/2015, 2015)

1. Serán castigados con una pena de prisión de uno a cuatro años y multa de seis a doce meses:

a) Quienes *públicamente* *fomenten*, *promuevan* o *inciten* directa o indirectamente al *odio*, hostilidad, *discriminación* o *violencia* contra un grupo, una parte del mismo o contra una persona determinada por razón de su pertenencia a aquél, por motivos racistas, antisemitas u otros referentes a la ideología, religión o creencias, situación familiar, la pertenencia de sus miembros a una etnia, raza o nación, su origen nacional, su sexo, orientación o identidad sexual, por razones de género, enfermedad o discapacidad.

Desde el trabajo de Gary Becker (Becker, 1968) sobre crimen y castigo sabemos que los resultados que se derivan de los modelos indican que un *incremento en la probabilidad de sanción o arresto*, sin importar la disposición al riesgo del infractor, *tiene un efecto negativo sobre la oferta de delito*. Sin embargo, *el efecto de un incremento de la pena* *es indeterminado* o ambiguo sin más suposiciones y las suposiciones que se haga sobre la posición frente al riego son la clave del efecto de una mayor severidad sobre el crimen. Este es incierto para los amantes al riego, mientras para los adversos al riego, un incremento en la severidad de la pena reduce el delito.

De manera que los medios sociales:

1. si somos capaces de analizar masiva y automáticamente mensajes y detectar aquellos que puedan constituir delito de odio;
2. facilitarán enormemente la identificación de los infractores de las leyes y la obtención de pruebas.

Por lo tanto, como corolario (Becker, 1968) aumentará la probabilidad de sanción al infractor y, por consiguiente, disminuirá la frecuencia de este tipo de delitos.

Cabe, incluso, ir más allá y plantearse a futuro la aplicación de la *justicia maquinal* mediante la cual, el infractor es incluso *juzgado* por un sistema informático que utiliza las herramientas desarrolladas en el campo de la inteligencia artificial para determinar su inocencia o culpabilidad, al menos como fase previa a la iniciación de un proceso legal convencional (Goodman, 2016).

Este problema de clasificación de tuits presenta un claro desequilibrio de clases, ya que la constituida por los tuits sospechosos de delito de odio, son muchos menos que los neutros (V. García, 2007). Este tipo de problemas se denominan ‘*la aguja en el pajar*’

*Palabras clave:* redes sociales, medios sociales, grupos sociales, colectivos, etnia, raza, nacionalidad, religión, orientación sexual, odio, discriminación, violencia, delitos de odio, desequilibrio entre clases.

Abstract.

The so-called “social networks” built-up by platforms such as **Facebook**™**,** **Twitter**™ which operate on the Internet underpin the *social media* that facilitate quick and mass exchange and discussion of information, experiences and opinions between individuals in a way never before seen in human history.

The range of social media available to the public is varied and growing and can be used for multiple purposes:

* articles in *wikis,*
* opinions about quality in hotels, restaurants ([tripadvisor](http://www.tripadvisor.es), [yelp](https://www.yelp.es)),
* social and professional contacts ([facebook](https://es-es.facebook.com/), [linkedin](https://es.linkedin.com/)),
* web blogs([wordpress](https://es.wordpress.com)).

Like anything new, the explosion of social media has had consequences that have been valued both positively and negatively for society as a whole.

Among the effects generally considered as negative, social media have persistently made 'visible' some attitudes of certain social groups that, to date, were present only in a veiled and/or sporadic way. Prominent amongst them are those which result in attacks on individuals or groups because their affiliation to certain groups defined by characteristics of nationality, sexual preferences, race, religion ..

This phenomenon, coupled with a change of attitude towards certain social behaviors or groups, has led to the emergence of a new criminal category in many countries: the so-called *hate crimes.*

Since Gary Becker’s ‘*Crime and Punishment*‘ (Becker, 1968) we know that the punishment of criminals is probabilistic. The offender may escape detection or apprehension, or be apprehended but not convicted. His economic theory of crime states that some criminal justice variables, are much more effective than others. Increasing *arrest rates*, followed by increasing the likelihood of *being convicted* have the largest impact. On the contrary, increasing the penalties beyond current levels has an uncertain effect on the crime rate. From this theory, it is clear that public authorities should focus more attention on strategies that increase the risk of arrest and less on strategies that increase the severity of punishment.

So, the social media

1. if we are capable of analyzing automatically mass messages in social media in so that we can detect those  that can constitute a hate crime
2. will enormously facilitate the identification of offenders and the collection of evidence.

Therefore, as a corollary, both the likelihood of being arrested and that of being convicted will raise for the infringer and, consequentially, the frequency of hate crimes will be reduced.

It is even possible to go further and consider in the future the application of *machine justice* by which the offender is tried by a computer system that uses tools developed in the field of artificial intelligence to determine their innocence or guilt, at least as a phase prior to the start of a conventional legal process (Goodman, 2016).

This classification problem for tweets presents a clear imbalance of classes, since the one formed by messages suspected of hate crime, are much less than its complementary class (V. García, 2007). These types of problems are often called *'the needle in a haystack'.*

1. Introducción.

Los *delitos de odio* son un tipo de delito cuyo motivo principal es la existencia de *prejuicios* respecto a la víctima del mismo y tienen lugar cuando el perpetrador del delito elige a su víctima en base a su pertenencia a un cierto *grupo*.

Los atributos principales que definen el grupo de pertenencia de la víctima suelen ser el sexo , la etnicidad o raza, la nacionalidad, el idioma, la orientación sexual, la religión, la discapacidad, la apariencia física o la identidad de género, entre otros.

Existen evidencias de que tales delitos basados de odio están influidos por eventos *singulares* de *amplia difusión* (Pete Burnap , Matthew L. Williams, 2015) (atentados terroristas, migración incontrolada, manifestaciones, revueltas, …). Este tipo de sucesos suelen actuar como detonadores de manera que la frecuencia de este tipo de delitos aumenta espectacularmente tras ellos. Por ello, parece razonable dotar a los responsables de mantener la seguridad de herramientas que permitan evaluar la probabilidad del alza de tales delitos y, si es posible, su localización geográfica y temporal.

Los medios sociales de comunicación juegan un importante papel en la comisión de estos delitos en tanto en cuanto las redes sociales se llenan de mensajes de individuos afines a los perpetradores que incitan a castigar al grupo elegido como diana (King, R.D., G.M. Sutton, 2013) que, recogidos a lo largo de un periodo temporal posterior al incidente detonante, pueden servir para analizar la evolución de la amenaza: escalada, estabilización, duración y descenso.

Tal es la importancia de estos medios que, en muchos países, se han tipificado recientemente[[1]](#footnote-1) también como pertenecientes a la categoría de delito de odio aquellas *manifestaciones públicas* que puedan considerarse una incitación al odio hacia ciertos colectivos.

Uno de los servicios más utilizados para realizar manifestaciones abiertas mediante la publicación de *microblogs* es Twitter™, motivo por el cual este servicio se ha seleccionado como fuente básica de datos para el desarrollo de un *Sistema para la Detección de Indicios de Delitos de Odio*.

Como en todo proyecto relacionado con la *Ciencia de los Datos*, es evidente que, antes de trabajar con datos es preciso capturarlos, lo que se hará mediante la utilización de la API de Twitter.

A continuación, se realiza un análisis exploratorio de datos que servirá de base para la depuración de los mismos, su formateo y modelización.

A partir de los datos depurados, se procederá a su análisis mediante técnicas de *Procesado de Lenguaje Natural* (NLP) para extraer patrones y atributos de los textos para, finalmente, clasificar los mensajes mediante técnicas de *Inteligencia Artificial* (AI) como positivos (que indican una mayor *oferta* de delitos[[2]](#footnote-2)) o negativos/neutros que no aportan pistas al respecto.

El elemento clave para tal clasificación es el *contenido del mensaje* en e*l* que el redactor del mismo[[3]](#footnote-3) –en este caso del t*weet* - manifiesta su sentimiento u opinión respecto a una *entidad* o aspecto de la misma.

El caso que nos ocupa, se enmarca dentro de un grupo de problemas de clasificación binaria - dos clases: contenido de odio o neutro - caracterizado por un ***desequilibrio*** muy pronunciado entre el número de instancias en cada clase[[4]](#footnote-4), cuya proporción puede alcanzar valores de 1:1000.

Este tipo de situaciones es de importancia en el mundo real en situaciones en que el coste de una clasificación errónea de las instancias de la clase *minoritaria* es muy elevado. Como ejemplos citaremos el diagnóstico de enfermedades o la detección de fraudes en tarjetas de crédito. En el primer caso, unos pocos píxeles del conjunto que constituye una imagen son la base del diagnóstico y en el segundo la proporción de fraudes sobre el total puede ser inferior al 1%.

Esta situación presenta dos dificultades importantes:

* 1. dificulta la aplicación de *métodos de clasificación* *supervisada* ya que para el etiquetado manual de unos cientos de casos, se requiere el examen de cientos de miles de tuits lo que alarga y encarece el etiquetado.
  2. Por otro lado, el desequilibrio entre las clases provoca que el algoritmo que entrenamos sobre un conjunto de con muy pocas instancias de la clase minoritaria sea, con frecuencia, incapaz de generalizar el comportamiento de esta clase y, por tanto, el algoritmo tiene una escasa capacidad predictiva.

Ambos problemas se tratan en este trabajo.

## Estructura.

La memoria explicativa del proyecto se estructura de la siguiente manera:

* 1. Comenzamos revisando algunos conceptos básicos referidos a la clasificación de en conjuntos de desequilibrio y revisaremos el *estado del arte* al respecto
  2. Seguidamente, en el apartado *descripción del proyecto* se detallan aspectos fundamentales del mismo tales como
     + Especificación del producto,
     + Herramientas utilizadas
       1. Metodología de gestión de proyectos utilizada
       2. Fuentes de datos,
       3. Software y hardware,
       4. Control de versiones
  3. A continuación se expone el diseño del proyecto

## El desequilibrio de clases en problemas de clasificación.

### Conjuntos de datos.

Lo ideal es recoger *varios conjuntos de* *datos independientes*, si ello no es posible, debemos conformarnos con un solo conjunto de datos que habremos de dividir en dos o tres conjuntos.



Fig. : División del Conjunto de Datos

La estrategia consistente en usar

1. un *conjunto de entrenamiento* para aprender y estimar los parámetros del modelo;
2. un *conjunto de validación* para evaluar modelos y seleccionar uno de ellos y
3. un *conjunto de prueba* o test para valorar la capacidad de predicción de los modelos.

Existen múltiples métodos de división de datos, de los cuales el más simple es el *método de retención* (holdout) que consiste en dividir aleatoriamente el conjunto original en dos subconjuntos ( o para entrenamiento y el resto para prueba). Si el conjunto original no es lo bastante grande, el método es ineficiente.

En *aprendizaje estadístico* se utiliza como supuesto básico que *tanto el conjunto de entrenamiento como el de prueba se extraen de una misma distribución subyacente* constituida por la combinación de las distribuciones de la clase mayoritaria y minoritaria. Si se llevan a cabo modificaciones en el conjunto de entrenamiento para tratar de reequilibrar las clases, el *conjunto de entrenamiento y el de prueba tendrán distribuciones diferentes* violando este supuesto.

### Medida del rendimiento de un clasificador.

La *matriz de confusión*, elemento básico de evaluación de clasificadores, se expresa en el caso de que solo existan dos clases como:



Tabla : Matriz de Confusión

Las medidas más inmediatas de evaluación del modelo que se nos ocurren son:

* *Exactitud*:
* *Tasa de error*:

Sin embargo, cuando *existe un claro desequilibrio entre clases*, como es el caso de los tuits de odio que pueden ser un 2 ‰ del total, es posible obtener una exactitud enorme aun clasificando todos los tuits (erróneamente) como negativos (no de odio), y equivocándonos, por tanto, en todos los positivos (Tomas Borovicka, 2012):



Tabla : Matriz de Confusión con desequilibrio de clases

Lo que nos daría unos valores de exactitud del 99,8%:



Por ello, son más útiles los indicadores:

1. ***Precisión*** (*p*recision)
2. ***Exhaustividad*** (*r*ecall)
3. el ***ROC****[[5]](#footnote-5)* gráfico que presenta varios puntos de interés (Fawcett, 2004)*.*

* *Precisión* (p) es el porcentaje de los *tuits* realmente pertenecientes a una clase que se asignan a la misma (aciertos) sobre el total de los asignados a dicha clase por el clasificador:

Es decir, el porcentaje de predicciones que se acierta.

* *Exhaustividad* (r) es el porcentaje de los *tuits* que han sido clasificados como pertenecientes una clase sobre el total de miembros de dicha clase (porcentaje de instancias de la clase bien clasificadas):

Se utilizan también combinaciones de p y r, tales como la media geométrica:

Y la armónica:

Un valor p=1 nos dice que todos los elementos recuperados como relevantes, lo son, pero no nos dice nada acerca de si hemos recuperado todos los documentos relevantes (r).

Otro indicador que se está utilizando cada vez más es el *ROC* habitual en Medicina y Biología para hablar de la detección de falsos positivos y negativos.

Ahora a la exhaustividad - - se la denomina ***sensibilidad****.* Como se ve*,* es la

y se introduce la ***especificidad****:*

De manera que

Y es

Por el teorema de la probabilidad total sabemos que:

Si dibujamos el gráfico que relaciona ambas magnitudes, obtenemos la Figura 4 en que el *ROC* es el *área bajo la curva* que puede tomar valores entre 0 (no acierta nunca) y 1 (la predicción acierta siempre).

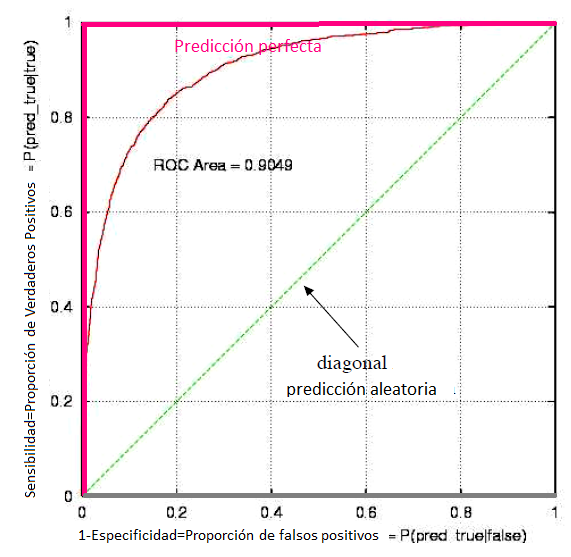


Fig. : ROC

En el ejemplo de los tuits (Tabla 2), los valores tanto de p como de r son nulos:



### Desequilibrio de clases y etiquetado y clasificación.

En aprendizaje supervisado, tenemos un conjunto - muestra - de datos S sobre el cual queremos construir un modelo de clasificación binario. La primera suposición que haremos es que las instancias del conjunto positivas y negativas observadas (S+ y S-) se extraen de dos distribuciones diferentes P y G. Las instancias positivas son las minoritarias. En estas circunstancias es fácil ver que, por una parte,

1. el etiquetado manual de las instancias resulta engorroso puesto que para encontrar un tuit – la *aguja* - de la clase minoritaria se precisa examinar una cantidad ingente de tuits de la otra clase – la *paja*. Y, por otra, se explica
2. el motivo por el cual un modelo clasificador sobre el conjunto produce una baja exhaustividad: la distribución positiva está subrepresentada y los valores *atípicos de G* - de mucha mayor cardinalidad - aunque sean una pequeña fracción, influirán en el clasificador ya que se considerarán por el clasificador como pertenecientes a S+ y el clasificador inducido estará sesgado hacia la clase minoritaria, es decir más cercano a los puntos de esta de lo que debiera, produciendo un rendimiento bajo del clasificador[[6]](#footnote-6) (Fig. 3 (A)).



Fig. : Errores de Clasificación en Muestras Desequilibradas

En la Fig. 3 (B) se muestra otro ejemplo cuando se usa 1-NN como clasificador en un conjunto desequilibrado. Lo mismo puede deducirse de la clasificación bayesiana.

## Estado del arte.

El proyecto se mueve, por una parte en el terreno problemático de los conjuntos no equilibrados, tema sobre el que existe una abundante literatura que se refiere fundamentalmente a:

* cómo simplificar el etiquetado de las instancias y
* cómo paliar los efectos de la asimetría entre clases.

Por otro lado, el proyecto de clasificación de tuits utiliza herramientas del campo del *Procesamiento del Lenguaje Natural* (PNL), rama de la Inteligencia Artificial que tiene sus orígenes en los años 50 del pasado siglo cuando aparecieron las primeras computadoras. Su desarrollo ha venido acompasado a los avances en la capacidad de cómputo (*Ley de Moore*) y en los algoritmos de *aprendizaje estadístico.*

## Etiquetado de instancias.

## Clases con probabilidades *a priori* no equilibradas.

## Herramientas disponibles.

Dado el enfoque práctico del proyecto, nos interesaremos más por las herramientas existentes a disposición del analista que por los fundamentos teóricos en que estas se apoyan. Tales herramientas se refieren tanto al campo del PNL propiamente dicho como al de los algoritmos de clasificación y recuperación de información que son necesarios para la clasificación de los tuits.

La escasez del tiempo disponible nos habrá hecho, sin duda, olvidar mencionar algunas de las plataformas existentes que, por otra parte, se encuentran en permanente evolución.

## Plataformas de código abierto.

* Basadas en Python:
  + Natural Language Tool Kit (NLTK) (Dan Garrette, 2009)
  + gensim (Řehůřek, 2010)
  + Scikit-learn (Gael Varoquaux, 2011)
* Basadas en Java:
  + CoreNLP (Christopher D. Manning)
  + MALLET (McCallum A. , s.f.) programas para modelizado temático, es y clasificación de textos.
  + LingPipe (Carpenter, 2007)
  + Weka (Ian H. Witten, 2007)
  + yTextMiner (yTextMiner, s.f.) : plataforma desarrollada en la universidad de Yonsei, Corea integrando parte de los modelos anteriores con otras librerías tanto para modelización temática como clasificación de textos y análisis de sentimiento.
  + Apache Lucene (https://lucene.apache.org/core/, s.f.)

## Plataformas comerciales.

Probablemente la más interesante es la que forman los diferentes paquetes del Sistema *SAS™*, más concretamente:

* *SAS™ Enterprise Miner*: paquete de aprendizaje automático y minería de datos enfocado a *business analytics*
* *SAS™ Text Miner*: utiliza la técnica SVD (la misma que gensim) para simplificar y acelerar la clasificación de textos y extracción de información.

## Descripción del proyecto.

* 1. **Especificaciones.**

El objetivo del proyecto es seleccionar y desarrollar un método para clasificar un conjunto de tuits ya recogidos de Twitter en dos categorías:

* Próximos a mensaje de odio,
* Neutros.
  1. **Etapas.**

Comenzaremos por seleccionar los campos que nos interesan de los tuits que son:

* + - Identidad del tuit,
    - Texto del *tuit* (‘documento’)

Los documentos han de ser depurados

## Diseño.

## Desarrollo.

## Pruebas.

## Resultados.

## Referencias.

1. Contenido

[*Palabras clave:* 3](#_Toc471990161)

[*Keywords:* 3](#_Toc471990162)

[2. Estructura. 4](#_Toc471990163)

[3. Estado del arte. 4](#_Toc471990164)

[4. Diseño. 4](#_Toc471990165)

[5. Desarrollo. 4](#_Toc471990166)

[6. Pruebas. 4](#_Toc471990167)

[7. Resultados. 4](#_Toc471990168)

[8. Referencias. 4](#_Toc471990169)

[Bibliografía 6](#_Toc471990170)

# Bibliografía

Becker, G. S. (Marzo-Abril de 1968). Crime and Punishment: An Economic Approach. *Journal of Political Economy, 76*(2), 169-217.

*BOE*. (30 de Marzo de 2015). Obtenido de https://www.boe.es/boe/dias/2015/03/31/pdfs/BOE-A-2015-3439.pdf

Carpenter, B. (2007). Bob Carpenter. *Proceedings of the 2nd BioCreative workshop.* Valencia,Spain.

Christopher D. Manning, M. S. (s.f.). *The Stanford CoreNLP Natural Language Processing Toolkit.* Obtenido de http://nlp.stanford.edu/pubs/StanfordCoreNlp2014.pdf

Dan Garrette, E. K. (2009). An Extensible Toolkit for Computational Semantics. *Proceedings of the Eighth International Conference on Computational Semantics.* Tilburg University, Netherlands.

Gael Varoquaux, A. G. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 2825-2830.

Goodman, J. (2016). *Robots in Law: How Artificial Intelligence is Transforming Legal Services.* Londres: Ark Group.

Ian H. Witten, E. F. (2007). Weka: Practical Machine Learning Tools and Techniques. *Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems.*, (págs. 192-196).

King, R.D., G.M. Sutton. (2013). High Times for Hate Crime: Explaining the Temporal Clustering. *Criminology , 51*(4), 871–94.

McCallum, A. (s.f.). *MALLET: A Machine Learning for Language Toolkit*. Obtenido de https://people.cs.umass.edu/~mccallum/mallet/

Pete Burnap , Matthew L. Williams. (2015). Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making. *Policy & Internet, 7*(2), 223-242,.

Řehůřek, R. a. (2010). Software Framework for Topic Modelling with Large Corpora. *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*, (págs. 46--50). Valetta, Malta.

Tomas Borovicka, M. J. (2012). Selecting Representative Data Sets. En E. b. Karahoca, *Advances in Data Mining Knowledge Discovery and Applications.* INTECH.

*yTextMiner*. (s.f.). Obtenido de http://informatics.yonsei.ac.kr/yTextMiner

A

AI · *Véase* Inteligencia Artificial.

análisis exploratorio de datos

Proceso al que se someten los datos antes de su modelización. Suele consistir en resumir sus características principales, con frecuencia usando métodos gráficos. Su objetivo es permitir una planificación más adecuada tanto del proceso de recolección como de su tratamiento posterior. · 5

aprendizaje estadístico

Rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. Se trata de crear programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos. · 10

B

*business analytics*

Conjunto de habilidades, técnicas y prácticas para la exploración iterativa del desempeño pasado de una empresa a fin de obtener una mejor comprensión de su funcionamiento y su desarrollo futuro. Utiliza tanto la minería de datos como la inteligencia artificial. · 10

C

*Ciencia de los Datos*

Campo interdisciplinario que comprende los procesos y sistemas para extraer conocimiento ,,de grandes volúmenes de datos en sus diferentes formas (estructurados o no estructurados) y formatos (.txt, .dat, .doc, .jpg, etcétera · 5

clasificación de textos

Tarea consistente en asignar un documento a una categoría determinada. · 10

*clasificación* *supervisada*

Este tipo de clasificación cuenta con un conocimiento *a priori*, es decir para la tarea de clasificar una instancia dentro de una categoría contamos con modelos ya clasificados (instancias agrupadas que tienen características comunes) · 6

*conjunto de entrenamiento*

Es el histórico de datos que se usa para entrenar al sistema que detecta los patrones. El conjunto de datos se compone de instancias, y las instancias de factores, características o propiedades · 7

*conjunto de prueba*

Es el usado para evaluar la bondad d las predicciones del modelo. · 7

*conjunto de validación*

Es el usado para evaluar y seleccionar modelos *entrenados* sobre el conjunto de entrenamiento · 7

D

*delitos de odio*

Aquellos motivados por prejuicios respecto a la víctima del mismo y tienen lugar cuando el perpetrador del delito elige a la víctima en base a su pertenencia a un cierto grupo. · 5

E

*entidad*

Producto, persona, evento, organización o tópico. · 6

estado del arte

Una de las primeras etapas dentro de un proyecto es la construcción de su estado del arte, ya que permite determinar la forma como ha sido tratado el tema, cómo se encuentra el avance de su conocimiento en el momento de realizar una investigación y cuáles son las tendencias existentes, en ese momento, para el desarrollo de proyectos en el mismo campo. · 7

etnicidad

etnicidad

Una etnia es un conjunto de personas que tienen en común rasgos culturales, idioma, religión, vestimenta, nexos históricos, tipo de alimentación, y, muchas veces, un territorio. Dichas comunidades, a veces, reclaman para sí una estructura política y el dominio de un territorio · 5

*Exactitud*

Medida de desempeño de un clasificador. Es el porcentaje de instancias que se clasifican en su clase real. · 8

*Exhaustividad*

Porcentaje de instancias clasificadas en su clase real sobre el total de las existentes en dicha clase. · 9

I

identidad de género

Percepción subjetiva que un individuo tiene sobre sí mismo en cuanto a sentirse hombre, mujer, o de un género no binario, sin considerar características físicas o biológicas. · 5

*Inteligencia Artificial*

*ciencia de hacer máquinas que actúan racionalmente. • Racional es todo agente que busca alcanzar unos objetivos de manera tal que optimiza el valor de una función de utilidad .* · 6

La Inteligencia Artificial es la ciencia de hacer máquinas que actúan racionalmente. · 10

L

*Ley de Moore*

La ley de Moore expresa que aproximadamente cada dos años se duplica el número de transistores en un microprocesador. · 10

M

*matriz de confusión*

Herramienta que permite la visualización del desempeño de un algoritmo que se emplea en *aprendizaje supervisado*. Cada columna de la matriz representa el número de predicciones de cada clase, mientras que cada fila representa las instancias en la clase real. · 8

*microblogs*

Servicio que permite a sus usuarios enviar y publicar mensajes breves, generalmente solo de texto. · 5

modelizado temático

(*topic model*) Modelo de aprendizaje sin supervisión que permite detectar el tema o asunto del que trata un documento. · 10

N

NLP · *Véase* Procesado de Lenguaje Natural

O

*orientación*

Tipo de opinión respecto a una entidad

favorable, desfavorable o inexistente. · 2, 3, 5

orientación sexual

orientación sexual

Patrón de atracción sexual, erótica, emocional o amorosa a determinado grupo de personas definidas por su sexo. · 3, 5

P

*Precisión*

Porcentaje de instancias clasificadas en su clase real sobre el total de las clasificadas en dicha clase. · 8

Procesado de Lenguaje Natural

Técnicas para conseguir que los ordenadores lleven a cabo tareas que involucran el uso del habla humana, tales como comunicación por voz entre hombre y máquina, procesamientomde texto o de voz. · 6

S

SAS

Acrónimo de Statistical Analysis Systems

Lenguaje de programación desarrollado por SAS Institute a finales de los años sesenta. · 10

SVD

*D*escomposición en *V*alores *S*ingulares de la matriz término-documento que permite simplificar las tareas de minería de datos. · 11

T

*Tasa de error*

Medida de desempeño de un clasificador. Es el porcentaje de instancias que se clasifican en una clase equivocada · 8

1. En España en 2015 (Ley Orgánica 1/2015, 2015) [↑](#footnote-ref-1)
2. Utilizamos la terminología económica de G. Becker (Becker, 1968) [↑](#footnote-ref-2)
3. O *fuente de opinión.* [↑](#footnote-ref-3)
4. Es decir, diferencias significativas entre las probabilidades *a priori*. [↑](#footnote-ref-4)
5. Receiver Operator Characteristic. [↑](#footnote-ref-5)
6. A veces se denomina el problema de la *aguja en el pajar* (“needle in the haystack”). [↑](#footnote-ref-6)