

# Modelo automático para la prevención del suicidio mediante la detección precoz de mensajes suicidas en redes sociales

## *Automatic model for suicide prevention by means of early detection of suicide messages in social networks*

Saray Zafra

Dpt. Leng. y Sist. Inf.

Univ. de Alicante

Apdo. de correos, 99

E-03080 Alicante

saray.zafra@ua.es

**Resumen:** Desgraciadamente, el número de suicidios en España parece haber aumentado en los últimos años. Constantemente nos llegan noticias de jóvenes y no tan jóvenes que han intentado suicidarse y que previamente habían mostrado conductas extrañas en los perfiles de sus redes sociales. Con la prevención como telón de fondo, es fundamental poder detectar dichos mensajes mediante técnicas de aprendizaje automático o *Machine Learning* por su utilidad contrastada en otras áreas de investigación. Sin embargo, las mismas requieren de un corpus equilibrado de calidad que cuente con el mayor número de muestras posible y así, poder aprender mediante ejemplos. Por ello, entendemos que el proceso de compilación de un corpus para la prevención del suicidio debe extenderse a todas las redes sociales posibles, con independencia de que estén ubicadas tanto en la *Surface Web* como en la *Deep Web*.

**Palabras clave:** Suicidio, Redes Sociales, Aprendizaje Automático, Compilación, Corpus Equilibrado, Internet Superficial, Web Profunda

**Abstract:** Unfortunately, the number of suicides in Spain is growing more and more. Constantly, news are published about young and not-so-young people who have attempted suicide but, previously, they have shown strange behaviours in their social networks profiles. With the aim of preventing, it is important to be capable of detecting messages by means of using *Machine Learning* techniques due to its utility observed in other researching areas. However, the previous task requires a balanced corpus with a lot of samples in order to learn by examples. That's why the compilation process should be spread out to all of social networks based on *Surface Web* and *Deep Web*.

**Keywords:** Suicide, Social Networks, *Machine Learning*, Compilation, Balanced Corpus, *Surface Web*, *Deep Web*

### **1 Motivación de la investigación**

Según los datos ofrecidos por la Organización Mundial de la Salud en su última nota descriptiva (OMS, 2017), el suicidio sigue siendo la segunda causa de defunción entre los jóvenes de 15 a 29 años. Además, a nivel mundial, el suicidio representa el 50 % de las muertes violentas en los hombres y el 71 % en las mujeres (Navarro-Gómez, 2017). En el plano nacional, el suicidio representa la primera causa de muerte por causas externas, con más de 2.500 muertes en hombres frente a las casi 1.000 muertes en mujeres (Instituto Nacional de Estadística, 2017).

El pensamiento de quitarse la vida requiere

re una previa reflexión por parte del sujeto (Dunlop, More, y Romer, 2011) y puede obedecer a un sinfín de causas (Aranguren, 2009; Brea, 2013) cada cual más compleja y dependiente de una amplia gama de factores de protección y de riesgo que a su vez tienen su origen tanto en factores internos como externos. Es por ello que las estrategias destinadas a su prevención deberían ser lo más variadas posibles (Mansilla, 2010) y centradas en aquellos ámbitos en los que más presentes estén dichas manifestaciones, como por ejemplo, las Redes Sociales.

En una era marcada por los nativos digitales (Prensky, 2010), las redes sociales se

han convertido en un modo de comunicación habitual entre los jóvenes de 14 y 24 años en el 91,3 % de los casos (Instituto Nacional de Estadística, 2014). En la actualidad, Internet es una de las herramientas más empleadas para realizar gestiones habituales y consultar dudas de todo tipo, incluidas las relacionadas con el suicidio (Alao, Yolles, y Armenta, 1999; Trujano Ruiz, Dorantes Segura, y Tovilla Quesada, 2009; Moreno Gea y Blanco Sánchez, 2012; Arroyo Fernández y Bertomeu Ruiz, 2012).

En la *Deep Web* o aquel entorno con motores de búsqueda distintos de los convencionales cuya IP es imposible de rastrear, muchos usuarios han encontrado un entorno ideal para la libre expresión de ideas relacionadas con la temática del suicidio.

Establecido el trinomio *Redes sociales-Jóvenes-Suicidio*, existe la necesidad de desarrollar una herramienta capaz de detectar el nivel de alerta de un mensaje suicida compilado desde las redes sociales (tanto de la *Deep Web* como de la *Surface Web*) con el fin de que el mismo sea derivado a los servicios correspondientes y así, poder prevenir el suicidio.

## 2 Antecedentes y trabajos relacionados

El suicidio es un fenómeno complejo. Además, todavía representa un tabú en la época en la que vivimos (mención aparte requiere la ausencia de datos “contrastados” por parte del INE<sup>1</sup>). Las pocas aproximaciones teóricas existentes (Morton et al., 1995; Leenaars, 2010) aparte de las innumerables clasificaciones basadas en la primigenia categorización realizada por Durkheim (Durkheim, 1897), también dificulta su estudio.

No obstante, en el campo de las *TLH* y mano a mano con la disciplina médica, sí se han realizado varios corpus relacionados con la temática. Estos se encuentran en inglés y centran su ámbito de estudio en ambientes controlados. Nosotros, por el contrario, queremos que una máquina aprenda a detectar el “lenguaje suicida” sin que sepa si dicho texto pertenece o no a un suicida previamente identificado. *Grosso modo*, los corpus relacionados son los siguientes:

1. *The Pak's and Paroubek's Twitter Cor-*

<sup>1</sup>[https://elpais.com/2017/06/12/ciencia/1497291180\\_123865.html](https://elpais.com/2017/06/12/ciencia/1497291180_123865.html)

*pus* (Pak y Paroubek, 2010).

Compila posts de Twitter y los clasifica en función de si presentan (i) *Emociones positivas* (e.g. Felicidad, diversión, alegría, plenitud, etc.), (ii) *Emociones Negativas* (e.g. Tristeza, enfado, decepción, etc.) o son (iii) *Textos Objetivos* (e.g. Hechos fácticos o emociones confusas que los investigadores no podían clasificar en las categorías anteriores).

2. *The Cincinnati Children's Hospital Medical Center Corpus* (Pestian, Matykie-wicz, y Linn-Gust, 2012).

Digitaliza y compila notas de suicidio real recogidas entre los años 1950 y 2012 por dos especialistas en el ámbito de la prevención del suicidio: Edwin Shneidman (considerado el padre de la Suicidología Moderna (Chávez-Hernández y Leenaars, 2010)) y John Perstian.

3. *The Genuine Suicide Notes Corpus* (Schoene y Dethlefs, 2016).

La compilación de textos procede de artículos de revista, corpus existentes y otros recursos académicos (Scheidman y Farberow, 1957), clasificando cada muestra en dos categorías: (i) *Características Sentimentales* (e.g. Miedo, culpa, esperanza, culpa, información, orgullo, amor, etc.) y (ii) *Características Lingüísticas* (e.g. Items tales como la longitud de la nota, el tiempo verbal de las mismas, etc.).

4. *The Corpus of Palestian Suicide Bombers' Farewell Letters* (Cohen, 2016).

Recoge la compilación de notas de suicidio real de hombres bomba palestinos desde el 2000 hasta el 2006 y analiza: (i) La Temática más frecuente en las notas suicidas (e.g. Empleo de mecanismos cognitivos, sociedad, afecto, religión, etc.), (ii) Las palabras más frecuentes. Además realiza dos análisis: (i) Análisis computerizado de textos de los temas más frecuentes (e.g. 27 palabras relacionadas con “MyMother”, 21 con “MyFather”, etc.) y (ii) Análisis de concordancia.

5. *The Mowery's et al. Twitter Corpus* (Mowery et al., 2016).

Los textos aquí compilados proceden de notas suicidas escritas en Twitter y el esquema de anotación seguido se basa en

los criterios de depresión establecidos en el DSM-V (American Psychiatric Association, 2013).

En los ejemplos mostrados previamente, podemos observar tres factores comunes que marcan la diferencia entre dichos corpus y el que queremos desarrollar en este proyecto de investigación: (i) Ambiente controlado, (ii) Conocimiento del historial clínico de muchos de los sujetos que componen la muestra y (iii) Compilación de notas de suicidio confirmadas.

Nuestro objetivo en este plan de investigación es claro: Conseguir que una máquina aprenda acerca de un modo particular de expresión del lenguaje suicida en un entorno concreto (como son las redes sociales) y que, sin la previa elaboración del perfil del usuario del que procede el mensaje o los mensajes, sea posible establecer un nivel de alerta suficiente como para derivar a los servicios que sean necesarios. Somos conscientes de que supone una tarea complicada, pero sería un gran avance en lo que a la prevención del suicidio se refiere.

### 3 Hipótesis y objetivos de investigación

Las hipótesis que guían el proceso de investigación son las siguientes:

1. Los textos extraídos de redes sociales de la *Deep Web* son diferentes a los obtenidos desde la *Surface Web*.
2. El anonimato que ofrece la *Deep Web* dificulta las labores de prevención y/o actuación en el caso concreto.
3. Un corpus de calidad debidamente anotado sobre mensajes suicidas puede mejorar los sistemas automáticos de detección de este tipo de mensajes.

Entre los objetivos planteados en la investigación encontramos los siguientes:

1. Desarrollar un *crawler* capaz de extraer información de *Deep Web* y *Surface Web* de manera automática.
2. Crear un entorno web donde sea posible realizar todas las tareas de anotación y gestión del proceso de anotación en general.

3. Generar conciencia social acerca de que el suicidio es un problema real de salud pública (OMS, 2017) y que el desarrollo de herramientas preventivas es la clave para reducir las cifras (800.000 personas se quitan la vida cada año).
4. Construir un corpus amplio y de calidad valorado mediante medidas objetivas como el acuerdo mutuo para que pueda ser empleado en futuros sistemas de aprendizaje automático.
5. Garantizar el libre acceso al corpus, que irá actualizándose periódicamente.
6. Asegurar el anonimato de los datos presentados en el corpus.
7. Evaluar la utilidad de este corpus probándolo en sistemas de aprendizaje automático.

### 4 Metodología y experimentos propuestos

El objetivo de la elaboración de un corpus de mensajes suicidas basado en redes sociales no es otro que poder crear una herramienta preventiva. Dentro de las *Tecnologías del Lenguaje Humano (TLH)* nosotros hemos seleccionado el Aprendizaje Automático (o *Machine Learning*

Aportando a una máquina el mayor número de ejemplos posible y tras la compilación de mensajes suicidas extraídos de redes sociales, es posible que la misma aprenda sobre un particular modo de expresión de ideas suicidas <sup>2</sup>. Para ello, se requiere la elaboración de los siguientes pasos:

#### 1. Proceso de compilación

Desarrollo de un *crawler* capaz de realizar una búsqueda automática de textos en redes sociales (tanto de la *Deep Web* como *Surface Web*) de perfiles públicos y que no requieran el *login*.

#### 2. Guía de anotación

Indicaciones sobre el período de entrenamiento y su duración, qué anotar, cómo anotar, así como una descripción detallada de las propiedades requeridas para cada texto además del programa con el que se realizará dicha anotación (para lo

---

<sup>2</sup><https://qz.com/1001968/artificial-intelligence-can-now-predict-suicide-with-remarkable-accuracy/>

que se enviará a los anotadores el correspondiente *plugin* y documentos informativos). En el caso de que se consiga crear un entorno web para dicho proceso, también se elaborará una guía de anotación en la que se indique cuáles son los pasos a seguir para realizar el proceso de anotación correctamente.

### 3. Fase de Anotación

A medida que el proyecto se inicie, cada anotador contará con muestras procedentes del *crawler* desarrollado que irá anotando según las categorías especificadas en la guía de anotación.

### 4. *Gold-Standard* o corpus consensuado

Se realizarán reuniones periódicas para analizar desacuerdos entre los anotadores y así poder establecer qué categorías son las más idóneas para cada propiedad. Este proceso será documentado con el fin de observar cuáles han sido las dificultades a lo largo del proceso de anotación y, de ser necesario, proceder a realizar cambios en la guía de anotación.

### 5. Uso del corpus en sistemas de aprendizaje automático.

Para demostrar que el corpus puede ser una herramienta valiosa en la detección del fenómeno suicida, se aplicará el corpus en varias etapas de construcción a sistemas de aprendizaje automático para poder estimar si el corpus puede ser parte de una solución global a este problema.

## 5 Cuestiones de interés a tratar durante el simposio

Por la envergadura que el fenómeno del suicidio representa, nos gustaría plantear las siguientes cuestiones al auditorio:

- ¿Es posible la prevención del suicidio mediante la aplicación del *Machine Learning* a los textos compilados de las redes sociales de ambos entornos?
- ¿Existe realmente un particular modo de expresión de las ideas relacionadas con el suicidio en las *Redes Sociales*?
- ¿Facilita el anonimato que se genera en la *Deep Web* la mayor difusión de contenidos relacionados con las ideaciones suicidas?

- ¿Dificulta la *Deep Web* la tarea de derivar un mensaje a los servicios de salud correspondientes en función del nivel de alerta previamente establecido para el mismo?.

## Agradecimientos

Especial agradecimiento a las **Ayudas Fundación BBVA a equipos de investigación científica** por la financiación del proyecto *Análisis de Sentimientos Aplicado a la Prevención del Suicidio en las Redes Sociales (ASAP)* y a Beatriz Botella Gil, Isabel Moreno Agulló y María de los Ángeles Herrero, compañeras del Departamento de Lenguajes y Sistemas Informáticos de la Universidad de Alicante (*DLSI*).

## Bibliografía

- Alao, A. O., J. C. Yolles, y W. Armenta. 1999. Cybersuicide: The Internet and suicide. *American Journal of Psychiatry*, 156(11):1836–1837.
- American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*. American Psychiatric Association, may.
- Aranguren, M. 2009. Modelos teóricos de la comprensión del suicidio. *Universidad de Buenos Aires*.
- Arroyo Fernández, A. y A. Bertomeu Ruiz. 2012. Métodos suicidas e Internet. 38(4):143–148.
- Brea, J. M. 2013. En torno al suicidio. *Cuadernos de atención primaria*, páginas 168–172.
- Chávez-Hernández, A. M. y A. A. Leenaars. 2010. Edwin S Shneidman y la suicidología moderna. *Salud Mental*, 33(4):355–360.
- Cohen, S. J. 2016. Mapping the minds of suicide Bombers using linguistic methods: The corpus of Palestinian suicide bombers' farewell letters (CoPSBFL). *Studies in Conflict & Terrorism*, 39(7-8):749–780, jul.
- Dunlop, S. M., E. More, y D. Romer. 2011. Where do youth learn about suicides on the Internet, and what influence does this have on suicidal ideation? *Journal of Child Psychology and Psychiatry*, 52(10):1073–1080, oct.

- Durkheim, É. 1897. *El suicidio. Un estudio de sociología*. 2ª edición.
- Instituto Nacional de Estadística. 2014. Encuesta sobre Equipamiento y Uso de Tecnologías de Información y Comunicación en los Hogares. páginas 1–9.
- Instituto Nacional de Estadística. 2017. España en cifras. páginas 17–19.
- Leenaars, A. A. 2010. Edwin S. Shneidman on Suicide. *Suicidology Online*, 1:5–18.
- Mansilla, F. 2010. *Suicidio y Prevención*. Valldargents, 27 bajos, 2010 edición.
- Moreno Gea, P. y C. Blanco Sánchez. 2012. Suicidio e Internet. Medidas preventivas y de actuación. *Moreno Gea P. Psiquiatria.com*, 16:1–12.
- Morton, M., R. W. Suicide, L. Threatening, B. New, y Y. Spring. 1995. The prevention of suicidal behaviors : An overview. 25(Spring):10–21.
- Mowery, D., H. A. Smith, T. Cheney, C. Bryan, y M. Conway. 2016. Identifying Depression-Related Tweets from Twitter for Public Health Monitoring. *Online Journal of Public Health Informatics*, 8(1):2579.
- Navarro-Gómez, N. 2017. El suicidio en jóvenes en España: cifras y posibles causas. Análisis de los últimos datos disponibles. *Clínica y Salud*, 28(1):25–31.
- OMS. 2017. *Nota Descriptiva. Prevención del suicidio*.
- Pak, A. y P. Paroubek. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of the Seventh Conference on International Language Resources and Evaluation*, páginas 1320–1326.
- Pestian, J. P., P. Matykiewicz, y M. Linn-Gust. 2012. What's In a Note: Construction of a Suicide Note Corpus. *Biomedical informatics insights*, 5:1–6.
- Prensky, M. 2010. Nativos e Inmigrantes Digitales. *Cuadrenos SEK 2.0*, (M-24433-2010):21.
- Scheidman, E. y N. Farberow. 1957. Clues to Suicide. *Public Health Report*, 71(Nº 2).
- Schoene, A. M. y N. Dethlefs. 2016. Automatic Identification of Suicide Notes from Linguistic and Sentiment Features. (2010):128–133.
- Trujano Ruiz, P., J. Dorantes Segura, y V. Tovilla Quesada. 2009. Violencia En Internet: Nuevas Víctimas, Nuevos Retos. 15(1):7–19.