Anales de Lingüística. Segunda época
n.7 – OCT 2021 – MAR 2022. Mendoza, Argentina
ISSN 0325-3597 (impreso) - ISSN 2684-0669 (digital)
https://revistas.uncu.edu.ar/ojs3/index.php/analeslinguistica
pp. 89-116

Análisis de rasgos lingüísticos con técnicas de procesamiento del lenguaje natural en la detección temprana de depresión

Analysis of language traits with natural language processing techniques in early detection of depression

María José Garciarena Ucelay[1] https://orcid.org/0000-0001-9158-3763
Universidad Nacional de San Luis, Argentina
mjgarciarenaucelay@gmail.com

Leticia Cecilia Cagnina https://orcid.org/0000-0001-7825-2927
Universidad Nacional de San Luis, Argentina
lcagnina@gmail.com

Marcelo Luis Errecalde https://orcid.org/0000-0001-5605-8963
Universidad Nacional de San Luis, Argentina
merrecalde@gmail.com

Resumen

El desarrollo de métodos computacionales que utilizan información de la Web para la detección temprana de riesgos es un área de investigación socialmente relevante, científicamente atractiva y actualmente en pleno crecimiento. La depresión es uno de los trastornos mentales más frecuentes a nivel mundial y con alta incidencia de suicidio en los casos más severos. Por lo tanto, su detección temprana podría derivar en un tratamiento a tiempo e incluso salvar vidas. En este trabajo, se analiza la relación que existe entre los modelos computacionales que permiten la detección automática de depresión y las propiedades lingüísticas del

[1] Recibido: 30.11.2020 | Aceptado: 07.01.2021

texto escrito por personas que experimentan la enfermedad. Se utilizan representaciones textuales que forman parte del estado del arte en clasificación de documentos y que cubren aspectos lingüísticos, sintácticos y semánticos. Los resultados obtenidos con clasificadores estándares indican que las incrustaciones de palabras capturan información precisa para detectar indicios de depresión de forma rápida y segura.

Palabras clave: detección temprana de depresión, representación de documentos, incrustaciones de palabras, métrica ERDE.

Abstract

The development of computational methods using information from the Web for early detection of risks is a socially relevant, scientifically attractive and currently a growing area of research. Depression is one of the most frequent mental disorders in the world and with high incidence of suicide in the most severe cases. Therefore, early detection of this illness could lead to a timely treatment and to save lives. This paper analyzes the relationship between computational models that allow the automatic detection of depression and the linguistic properties of the text written by people who experience the disease. State-of-the-art text representations in document classification are used, covering linguistic, syntactic and semantic aspects. The results obtained with standard classifiers indicate that word embeddings capture precise information to detect quickly and safely signs of depression.

Keywords: early depression detection, document representations, word embeddings, ERDE metric.

Introducción

La recopilación y el análisis de datos extraídos de medios sociales, en combinación con técnicas de procesamiento del lenguaje natural (PLN) y algoritmos de aprendizaje automático, han contribuido exitosamente a la detección de trastornos mentales (Calvo y otros 2017:65, Chandran y otros 2019, Low y otros 2020). Un ejemplo de este tipo de trastorno es la depresión.

Desde hace varias décadas, la depresión constituye uno de los trastornos mentales más comunes, y en situaciones pandémicas como la causada por

la enfermedad del coronavirus (COVID-19), su frecuencia y gravedad han podido llegar a triplicarse (Boston University School of Medicine 2020). Factores como el confinamiento social, cambios de hábitos y disminución o pérdida de ingresos experimentados durante la pandemia han potenciado el agravamiento de la depresión. Por otro lado, dichos factores han propiciado el incremento del uso de Internet (Nguyen y otros 2020), siendo en muchos casos la única forma posible de trabajar, disfrutar de entretenimientos e incluso tener alguna forma de contacto social con familiares y amigos. Según reporta Kemp (2020), más de la mitad de la población mundial actualmente usa redes sociales, lo que significa un incremento del 10,5% respecto del año anterior en el número de usuarios de este tipo de plataformas de comunicación.

Sistemas como Instagram, Twitter, Facebook y Reddit son algunos de los más elegidos para compartir comentarios, preferencias, actividades, emociones y comportamientos. Habitualmente, muchas personas utilizan esta forma de comunicación para expresar dolencias, enfermedades o sentimientos, en busca de respuestas, consejos, o simplemente atención. Es así que la información que se puede extraer de estas plataformas puede ser muy útil para detectar los primeros signos de depresión y evitar el empeoramiento de la misma, sugiriendo un tratamiento temprano.

En ese marco, el objetivo del presente trabajo es analizar cómo diferentes representaciones de textos permiten encontrar modelos computacionales eficaces, para detectar tempranamente signos de depresión, con base en marcadores lingüísticos presentes en escritos informales. Se trabajará con la hipótesis de que un mayor entendimiento (identificación de rasgos distintivos) de la estructura del lenguaje permite obtener mejores modelos computacionales. Un recurso importante para poder demostrar esta hipótesis es la utilización de datos lingüísticos. En particular, se emplearán los textos provenientes de *posts* del sitio web Reddit (Losada y otros 2018). Estos *posts* están escritos en lengua inglesa y fueron etiquetados como pertenecientes a usuarios “depresivos” (los cuales explícitamente

indicaron que fueron diagnosticados con el trastorno) y otros "no depresivos".

El primer paso para efectuar el análisis es la identificación de características relevantes para la tarea. En este caso, se extrajo información lingüística a nivel léxico, semántico y sintáctico, utilizando características como las provistas por LIWC (*Linguistic Inquiry and Word Count*) (Pennebaker y otros 2015), la clásica bolsa de palabras o *bag-of-words (BoW),* que puede incluir palabras y también caracteres (Salton y McGill 1983:59, Zhang y otros 2010:45), k-TVT (*k-Temporal Variation of Terms*) (Cagnina y otros 2019:550) y las incrustaciones de palabras o *word embeddings* (Jurafsky y James 2020).

La herramienta LIWC permite evaluar texto a través del uso de un diccionario, en tres dimensiones: aspectos lingüísticos, procesos psicológicos e intereses personales. La bolsa de palabras es una de las formas más simples de representar texto y utiliza la frecuencia de ocurrencia de cada término[1] en el documento para formar el vector de características que lo representa. En particular, se consideró la bolsa de palabras que sólo tiene en cuenta las palabras completas y trigramas de caracteres que permiten modelizar información espacial de los caracteres presentes en el documento. K-TVT es un método para la detección anticipada de riesgos que puede ajustarse dependiendo del nivel de urgencia con la que se necesite realizar la predicción. Además, maneja adecuadamente el desequilibrio de las clases (depresivos y no depresivos) que habitualmente está presente en este tipo de tareas, y dada su buena prestación en el laboratorio eRisk 2018[2], se considera en el presente estudio. Finalmente, las incrustaciones de palabras han tomado protagonismo en los últimos años por la eficiencia que poseen en modelizar las palabras y las interacciones que se producen entre ellas en el

[1] Con "término" nos referimos a palabras, secuencias de dos o más palabras o trigramas de caracteres. Nótese que en el resto del artículo se usará "término" como hiperónimo de palabra, secuencias u otra unidad sintáctica que pueda considerarse.
[2] https://erisk.irlab.org/2018/index.html

texto. A través del uso de vectores pre-entrenados, se obtienen representaciones de cada una de las palabras presentes en el documento, con el fin de modelizar propiedades semánticas del lenguaje como la polisemia y la sinonimia.

Posteriormente, se obtuvieron modelos con cada una de las representaciones elegidas utilizando clasificadores estándares como *Support Vector Machine* (Cortes y Vapnik 1995) y *Random Forest* (Breiman 2001). La calidad de los resultados indica que la depresión puede ser detectada de forma rápida con los modelos analizados a pesar de ser una tarea muy desafiante. En particular, con la utilización de las incrustaciones de palabras, focalizando las características a un contexto lingüístico, se obtienen los mejores valores de la métrica ERDE (*Early Risk Detection Error*) (Losada y Crestani 2016:34), mostrando así su eficacia en la detección temprana de la enfermedad.

El escrito está organizado de la siguiente manera. En la Sección 1 se enuncia la tarea de detección temprana de depresión, se describe en detalle el conjunto de datos utilizado en los experimentos y se presentan brevemente las métricas usadas para evaluar los modelos. En la Sección 2 se muestra el estudio experimental, comenzando con una descripción de las representaciones de textos analizadas. Luego, se describen los resultados obtenidos y finalmente se provee un análisis de las representaciones en el contexto de la extracción de información relevante que permite visualizar rasgos de depresión en textos. Por último, las conclusiones y los trabajos futuros son presentados al final del escrito.

1. Detección temprana de depresión

La depresión es una enfermedad frecuente en todo el mundo y se caracteriza por la presencia de sentimientos negativos, culpa, tristeza, baja autoestima, falta de sueño y apetito, cansancio e insomnio, entre otros (Acosta-Hernández y otros 2011). A principios del año 2020, la Organización Mundial de la Salud (OMS) manifestaba que más de 300 millones de personas padecían este trastorno mental en todo el mundo.

Algunos casos de depresión son leves, pero otros llegan a ser tan graves que terminan en suicidios. La OMS estima que casi una cuarta parte de los afectados llegan hasta dicha instancia, de allí la importancia de la detección temprana del trastorno. Así, advertir las manifestaciones de los primeros síntomas podría permitir un tratamiento adecuado a tiempo, evitar el empeoramiento de los síntomas, e incluso salvar vidas.

Desde hace varios años el laboratorio eRisk[1] forma parte del *Conference and Labs of the Evaluation Forum*[2] y organiza eventos anuales en donde se propicia la investigación interdisciplinaria de problemáticas relacionadas con la seguridad y la salud. Los laboratorios invitan a la presentación de propuestas para la resolución de tareas específicas como la detección temprana de la depresión. Uno de los valiosos aportes de estos eventos es el conjunto de datos etiquetados, el cual es utilizado en este trabajo.

1.1. Descripción de los datos

Para analizar los rasgos lingüísticos relacionados con la depresión, se utilizó una colección de *posts* extraídos de la plataforma social Reddit. Estos datos fueron provistos en la edición 2018 del laboratorio eRisk (Losada y otros 2018) para la predicción temprana de riesgo en la Web.

La colección, en lengua inglesa, está dividida en dos grupos y cada documento se corresponde con un conjunto de *posts* por usuario. Los usuarios del grupo “depresivo” (o “positivo”) explícitamente mencionan haber sido diagnosticados con tal trastorno. Los individuos del grupo “no-depresivo” (o “negativo”) constituyen el grupo de control. Estos usuarios fueron seleccionados aleatoriamente, con lo cual en sus *posts* hablan de diversos temas. No obstante, varios de ellos mencionan la enfermedad o tienen algún familiar afectado por la depresión, pero no se puede determinar con certeza que ellos mismos no sean depresivos. Por ejemplo, un *post* de un usuario no depresivo (clase negativa) expresa:

[1] https://erisk.irlab.org
[2] http://www.clef-initiative.eu

> *"**I suffer** from **anxiety** and **clinical depression**. **I'd** fucking love to see a **therapist**, but the truth is **I can't**. Admitting you're **depressed** isn't a simple thing. **I'm afraid my parents** would **feel sorry** for **me**, and they're so old-fashioned they wouldn't understand it anyway. (…)**I'm** gonna have to **keep my depression** to **myself**, at least until **I'm** on **my own**. (…) **I** may tell them **I'm depressed**, that **I** really do have a **mental illness** and that **I** wanna go see a **doctor** (…)"*

Sin embargo, en la clase positiva existen patrones lingüísticos similares. Por ejemplo, el post de un usuario etiquetado como depresivo dice:

> *"(…) Please don't tell me **I'm not depressed because I'm not** acting the same way you were when **depressed**. Ye, **I'm clinically depressed** and have been since **my diagnosis** 1.5 years ago (…)"*

Ambos fragmentos presentan un alto grado de solapamiento de vocabulario y, sin embargo, pertenecen a categorías distintas. Esta es una circunstancia que, naturalmente, dificulta aún más la tarea de predicción.

El conjunto de textos de un usuario consiste en una secuencia de *posts* ordenados en forma cronológica, organizada en 10 partes. A su vez, los datos se distribuyeron en dos conjuntos, uno para el entrenamiento de los métodos predictivos y otro para la prueba de los mismos. El conjunto de entrenamiento consta de 135 usuarios positivos y 752 usuarios negativos. El conjunto de prueba consta de 79 usuarios depresivos y 741 no depresivos. Así, en total hay un 13% de usuarios con depresión contra un 87% de individuos del grupo de control. En consecuencia, las clases están muy desequilibradas, lo cual acerca este conjunto de datos a un escenario realista, pero representa otro factor de dificultad para el aprendizaje del modelo.

La **figura 1** muestra dos diagramas de cajas que resumen información sobre algunas estadísticas de la colección de entrenamiento. En la gráfica de la izquierda se puede observar la distribución de *posts* por usuario que se obtiene en base al total de 49.557 (9%) *posts* de la clase "depresivos" y 481.837 (91%) de la clase "no-depresivos". Este gráfico ilustra que la

cantidad mínima de *posts* por usuario es de 10 para ambas clases y la cantidad máxima para los negativos es de 2.000. Por otro lado, hay dos usuarios positivos que tienen entre 1.600 y 1.850 *posts*, ilustrados por los dos puntos negros que representan valores atípicos, aunque la mayoría de usuarios no supera los 1.400 *posts*. Debido a esta disparidad entre valores extremos, considerar la mediana (línea horizontal dentro de cada caja) es más representativo ya que indica que en general la cantidad de *posts* por usuario oscila entre los 154 y los 375 *posts* para la clase positiva y negativa, respectivamente.

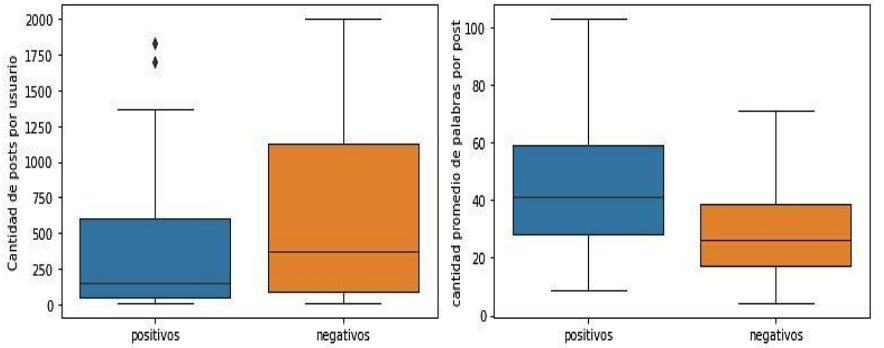

Figura 1. Estadísticas de la colección de textos por clase

Por otra parte, la cantidad total de palabras para la clase depresiva es de 14.846, una suma bastante menor a la de 18.961 palabras totales de los no-depresivos. Si se divide la cantidad de palabras de cada usuario por su número total de *posts*, se puede obtener la cantidad promedio de palabras por *post* para cada una de las clases. Esta información queda ilustrada en los diagramas de caja de la **figura 1** (derecha). La gráfica muestra que en promedio la clase positiva tiene *posts* de 41 palabras. En cambio, los *posts* de la clase negativa son más cortos, ya que están compuestos de 26 palabras aproximadamente.

Por otro lado, se puede obtener información útil sobre la longitud de las palabras. Haciendo un estudio sintáctico de los *post* se determinó que en

ambas clases la longitud promedio de palabras es de 5 caracteres y su desviación estándar oscila entre 0,3 para los positivos y 0,7 para los negativos. Un examen visual de los *posts* con las palabras más largas demostró que contenían ruido, es decir, caracteres repetidos, códigos de programación, texto en español, símbolos, etc.

1.2. Medidas de evaluación

Una medida estándar para evaluar la prestación de un clasificador es la conocida medida F1. Esta provee la media armónica entre la precisión y la cobertura del método evaluado. No obstante, a diferencia de la métrica ERDE, que también se reporta en este trabajo, la medida F1 no toma en cuenta el factor "tiempo/urgencia" con la que se realiza la clasificación.

ERDE (*Early Risk Detection Error*) fue propuesta por Losada y Crestani (2016) para evaluar el error del método en la detección temprana de usuarios con signos de depresión. Esta medida considera la cantidad de información que el clasificador necesita para tomar la decisión. La demora en la respuesta se mide luego de leer cierta cantidad *c* de información (*posts*, en este caso). Dos versiones de esta métrica son ERDE5 y ERDE50, en las cuales difiere el valor tomado como referencia para que la decisión sea tomada (clasificación), es decir, $c>5$ para el primer caso y $c>50$ en el segundo.

2. Estudio experimental

A continuación se describen brevemente las representaciones de textos empleadas en la tarea de clasificación temprana para la detección de depresión. Luego, se establecen las bases para la experimentación y, finalmente, se presentan y analizan los resultados.

2.1. Representaciones de texto

Una de las tareas fundamentales del procesamiento de lenguaje natural es la representación de los textos. Esta representación luego se emplea como entrada para los algoritmos de aprendizaje automático. En esta primera etapa, el texto se considera simplemente como una secuencia de símbolos

de entrada sin interpretación alguna. El objetivo es transformar dicha secuencia en un conjunto de números que agrupen distintas características de esas secuencias, con la intención de introducir una representación del significado del texto en los modelos computacionales.

En vista de ello, los diversos enfoques para representar textos codifican características diferentes. De forma general, se pueden identificar aquellos enfoques que consideran a los textos como bolsas de palabras, en donde el orden de las unidades léxicas no es relevante y, por el otro lado, como conjuntos de secuencias (n-gramas de palabras o de caracteres), en donde se preserva de cierta manera el contexto y por ello, la semántica cobra más importancia. Asimismo, las representaciones se pueden clasificar en función de cómo se generan las características o atributos a emplear. Así se tienen representaciones con características estáticas, dinámicas, distribucionales o aprendidas.

En las representaciones con atributos estáticos, las características se especifican antes del procesamiento de los textos y luego se calculan por cada documento; generalmente con base en frecuencias o cantidades. Una posibilidad es que estas sean relativas a la lengua en sí, como cantidad de palabras totales, longitud promedio de las palabras, cantidad de verbos conjugados en pasado, etc. Alternativamente, pueden estar relacionadas con propiedades del problema o de la colección de textos, es decir, con metadatos como la cantidad de comentarios, información del usuario, el horario en el que fueron publicados los *posts*, etc.

Un ejemplo de representación con atributos estáticos es el obtenido con la herramienta LIWC, mencionada previamente. Este software obtiene 94 atributos con base en un diccionario que contiene 6.400 palabras, raíces de palabras y emoticones categorizados de forma jerárquica. Estas categorías abarcan componentes estructurales, cognitivos y emocionales. Además, incluyen cuatro variables que resumen características del lenguaje como la presencia de pensamiento analítico, la autenticidad, el tono emocional y el nivel de confianza que transmite un texto.

Por otro lado, dentro de las representaciones con atributos dinámicos, que son aquellos que se obtienen como parte del procesamiento de los textos, están las clásicas bolsas de palabras. En este caso, las características son los términos que aparecen en la colección, y un documento está representado por los términos presentes en el texto del mismo. Los valores que pueden tomar estas características dependerán de la importancia o peso que se le otorgue a cada una. La ponderación más sencilla es la binaria (asigna un 1 si el término está presente en el documento o 0, en caso contrario). Otra posibilidad es aquella cuyo valor se incrementa proporcionalmente al número de veces que aparece el término en el documento (*tf*, por sus siglas en inglés). Por último, es posible utilizar la ponderación que, además de considerar la frecuencia, también penaliza la participación del término en los otros documentos (*tf-idf*) (Spärck Jones 1972). Este último es el que se utilizó en el estudio experimental.

A diferencia de las representaciones con características estáticas, en las que usan características dinámicas no se sabe de antemano qué dimensión tendrá el vector de términos, es decir, cuál será el tamaño del vocabulario. Generalmente, los vectores resultantes suelen tener una dimensionalidad alta y ser ralos. Por ello, a menudo no se trabaja con el vocabulario total sino que se realiza la selección de las unidades léxicas más relevantes.

Al considerar unigramas de palabras, la representación se considera una bolsa de palabras y el orden de los términos en las oraciones se pierde completamente. Sin embargo, al considerar bigramas o trigramas, se puede preservar parcialmente el contexto a costa de un incremento en el tamaño del vocabulario.

Si los términos son considerados secuencias de *n* caracteres, la representación obtenida se denomina n-gramas de caracteres. Especialmente para el inglés, los trigramas de caracteres son los que mejor prestación han demostrado porque entre otros, pueden capturar sufijos significativos como *“ly_”*, *“ing”*, *“ed_”*. Asimismo, los trigramas de caracteres son útiles para colecciones de textos informales en donde

usualmente pueden aparecer palabras mal escritas o con repeticiones de caracteres. Una palabra con error ortográfico, a nivel de unigrama de palabras se considera como una palabra nueva o distinta (introduciendo ruido a la representación y por ende luego, al modelo) pero desde el punto de vista de los n-gramas de caracteres, como algunas partes de la palabra coincidirán con n-gramas ya presentes en el vocabulario, el nivel de ruido será menor, favoreciendo el aprendizaje del modelo.

Como se puede apreciar, BoW principalmente se centra en la forma ortográfica de las palabras e ignora parte de la información semántica de las mismas. Si dos palabras son sinónimas, serán contadas por separado; similarmente si una palabra posee distintos significados dependiendo del contexto, se tratará como la misma, es decir, la polisemia y la sinonimia no pueden ser tratadas correctamente con este enfoque.

Así, las representaciones distribucionales (Turney y Pantel 2010:148) surgieron como una mejora a las desventajas antes mencionadas. En este enfoque la atención se centra en las palabras y en los contextos en que éstas ocurren. Se hace hincapié en medir la similitud de las palabras con base en la hipótesis distribucional, la cual establece que las palabras que ocurren en contextos similares tienden a tener significados similares (Harris 1954). En estos enfoques primero se representan los términos (mediante vectores) en función de los distintos contextos en los que pueden ocurrir, y luego se generan los vectores de los documentos a partir de esas representaciones. Cada componente del vector se deriva de la ocurrencia del término en distintos contextos: otras palabras, frases, oraciones, párrafos, capítulos, documentos, etc.

En particular, cada contexto puede ser modelizado mediante elementos semánticos llamados conceptos (Li y otros 2011) y así, representar las palabras y los documentos con una combinación de conceptos. El conjunto de conceptos asociados a cada palabra puede ser visualizado como una bolsa de conceptos y luego el documento estará representado por los conceptos asociados a las palabras de dicho documento. Una forma simple

de determinar los conceptos es utilizar las etiquetas de las clases en una tarea de clasificación; esta es la idea detrás de la representación k-TVT. En esta representación, dos palabras están relacionadas si sus distribuciones de frecuencia relativa en los documentos de las distintas clases son similares. Cuanto más frecuente es una palabra en los documentos que pertenecen a una clase, mayor es su membresía a dicha clase. Además de la buena prestación que han demostrado los modelos que utilizan k-TVT, otras ventajas de esta representación distribucional son la baja dimensionalidad de los vectores que representan a cada documento y el equilibrio que realiza de la clase minoritaria (o positiva) respecto de la mayoritaria (o negativa). Funez y otros (2018) comprobaron que existe una relación directa entre el parámetro *k* (es decir, el número de partes de la clase positiva con la que se aumenta la misma) y el umbral de clasificación *c* empleado en la medida ERDE para evaluar la decisión. Para *c*=5 se requiere una decisión rápida y precisa, por lo que *k*=0 presenta un buen equilibrio entre ambos requerimientos. Para *c*=50 un valor de *k*=4 es más adecuado porque maximiza la exhaustividad del método, beneficiando a la medida F.

Por último, surgen las representaciones aprendidas, en donde la idea es extender el aprendizaje automático, usualmente empleado en un paso posterior para generar el modelo de clasificación, a la representación de los documentos. Para el aprendizaje de representaciones se utilizan las incrustaciones de palabras. Básicamente, las incrustaciones son representaciones distribuidas de palabras basadas en vectores densos y de longitud fija, que se obtienen a partir de estadísticas de co-ocurrencia de las palabras según la hipótesis distribucional. En la literatura específica se puede distinguir entre las representaciones derivadas de enfoques basados en conteo como GloVe (del inglés, *Global Vectors*), (Pennington y otros 2014:1533), y los que surgen de métodos de aprendizaje neuronales predictivos como *word2vec* (Mikolov y otros 2013) y BERT (*Bidirectional Encoder Representations from Transformers*) (Devling y otros 2018), entre otros.

Generalmente, en los enfoques predictivos se utilizan redes neuronales con muchas unidades y se alimentan con extensas colecciones de textos de forma no supervisada, lo que habilita que las representaciones aprendan también conceptos generales del lenguaje. Así, las incrustaciones de palabras capturan relaciones sintácticas y semánticas muy interesantes de las palabras, como por ejemplo los significados relacionales. El ejemplo más famoso presentado junto con la representación *word2vec* en donde se muestra que las analogías de palabras se pueden resolver con aritmética es:

"vector(rey) – vector(hombre) + vector(mujer) ≈ vector(reina)"

Este ejemplo muestra que si se opera aritméticamente sobre los vectores "rey", "hombre" y "mujer", se obtiene un vector similar al de la palabra "reina".

El algoritmo *word2vec* fue el que popularizó las incrustaciones de palabras. Básicamente, los autores propusieron aprender un clasificador con la finalidad de obtener como efecto colateral las representaciones de las palabras. Las incrustaciones se extraen de la capa previa a la de salida de un clasificador neuronal. Este enfoque, como busca también codificar el contexto en el que se encuentra la palabra, tiene como parámetro el tamaño de la ventana de contexto. Para ventanas pequeñas, las palabras más similares a la palabra objetivo serán semánticamente parecidas y tendrán la misma categoría gramatical. En cambio, en ventanas más grandes las palabras más similares a la palabra objetivo estarán relacionadas semánticamente sin llegar por eso a ser necesariamente similares.

El parámetro de ventana de contexto local en combinación con la factorización de matrices globales determina la representación GloVe. Los autores ilustran que el cociente de las probabilidades de co-ocurrencia de dos palabras es lo que contiene la información relevante. En primer lugar, se calcula la matriz de co-ocurrencia entre todas las palabras del vocabulario de la colección de textos. Luego en base a esa matriz, el

algoritmo aprende dos vectores por cada palabra: uno que es la representación de la palabra en sí y otro en el que se modeliza el contexto. Finalmente, se pueden promediar ambos vectores para obtener la representación vectorial de cada palabra y usarlos para representar un documento. Por esta razón, los vectores de palabras resultantes funcionan muy bien en tareas de analogías. La distancia euclidiana o el coseno entre dos vectores de palabras pueden medir la similitud semántica de las mismas. Estas métricas de distancia devuelven un solo número que cuantifica la relación entre las dos palabras, pero, en realidad, dos palabras se pueden relacionar de varias maneras. Por ejemplo, “hombre” puede considerarse similar a “mujer” en el sentido de que ambas palabras describen a los seres humanos, pero por otro lado también se consideran opuestas. Para que el modelo sea capaz de realizar esta distinción es necesario que se asocie más de un número al par de palabras. Así, GloVe está diseñado para que tales diferencias vectoriales capturen tanto como sea posible el significado especificado por la yuxtaposición de dos palabras.

En los últimos años, el estado del arte en el campo del procesamiento del lenguaje natural se ha focalizado en enfoques predictivos basados en *transformers* (Vaswani y otros 2017:6001). Los *transformers* son un tipo de arquitectura de red neuronal profunda que incluye mecanismo de atención. Estos mecanismos codifican cada palabra de una frase en función del resto de la secuencia, permitiendo así introducir el contexto en la representación. Antes del surgimiento de los *transformers*, la entrada a la red neuronal era una palabra y la ventana de contexto era la principal limitante (debido a costos computacionales). Con esta arquitectura, ahora se alimenta a la red con frases completas, permitiendo codificar también la posición de las palabras en la misma. De esta manera, se generan incrustaciones contextuales.

Una de las representaciones más conocida que utiliza *tranformers* es BERT que toma como entrada dos frases y entrena un modelo a partir de dos tareas. La primera consiste en eliminar un porcentaje de las palabras en una frase y pedirle a la red que complete el espacio vacío con las palabras

correctas. La segunda, en cambio, toma como entrada dos frases y se le pregunta si la segunda de ellas viene a continuación de la primera. De esta forma, BERT procesa todas las palabras de las frases en forma simultánea y puede analizar el contexto a izquierda y derecha.

Lo interesante de estos últimos métodos es que se entrenan sobre tareas generales de forma auto-supervisada. Esto resulta relevante ya que significa que el aprendizaje no depende de dominios ni de tareas particulares.

En resumen, tanto para las representaciones distribucionales como para las aprendidas, usualmente primero se genera una representación a nivel de términos u oraciones y, a continuación, se utiliza dicha representación para generar los vectores de los documentos. Estos vectores son los que finalmente servirán para entrenar el modelo que se construirá con el fin de resolver el problema. La forma más sencilla de crear estos vectores es a partir de la concatenación de los vectores de términos u oraciones que aparecen en el documento, o bien mediante algún método de agregación como la suma (ponderada o no) y el promedio.

En la próxima subsección se especificarán con detalle las configuraciones particulares de las representaciones seleccionadas.

2.2. Resultados

En este estudio experimental se comparan las representaciones descritas en la subsección anterior sobre el conjunto de textos detallado en la Sección 1, con la división original de la colección (conjunto de entrenamiento y de prueba). Se presentan los valores obtenidos para las medidas de evaluación F1, precisión (π) y cobertura (ρ) de la clase positiva, ERDE5 (E_5) y ERDE50 (E_50) previamente mencionadas. La programación de las distintas rutinas para el procesamiento de los datos, la generación

de los modelos y su posterior evaluación se llevó a cabo con el lenguaje Python[1] (versión 3.6).

En pruebas iniciales se consideraron cuatro métodos de aprendizaje, *Multinomial Naïve Bayes*, *Logistic Regression*, *Support Vector Machine* (SVM) y *Random Forest* (RF), y cuatro umbrales de clasificación, 0,9, 0,8, 0,7 y 0,6 para resolver la detección de depresión. Sin embargo, sólo se informan los resultados de los algoritmos de clasificación (AC) RF y SVM con umbral igual a 0,6 por ser éstos los mejores valores obtenidos. En todos los casos probados se utilizaron los parámetros por defecto de cada clasificador con el fin de hacer énfasis sólo en las representaciones.

En la **tabla 1** se resume la experimentación resaltando en negrita los mejores resultados. A continuación, se detallan las particularidades consideradas para cada caso.

Para las representaciones BoW y k-TVT se realizaron estudios preliminares en donde se agrupó el conjunto de entrenamiento y de prueba, y se realizó una validación cruzada de cuatro pliegues. Estos experimentos se ejecutaron con la versión original de los datos pero también se obtuvo una versión *pre-procesada* de los mismos. La tarea de pre-procesamiento consistió en eliminar enlaces y extender las contracciones del inglés, por ejemplo, *can't* se transformó en *can not*. Este pre-procesamiento en general ayuda a normalizar los textos debido a que una misma expresión puede aparecer escrita de distintas formas aunque su semántica es similar. A su vez, la normalización permite descubrir otros patrones o relaciones ya que, por ejemplo, al transformar las terminaciones *~n't* en *not,* la negación pasa a tener más protagonismo dentro del modelo obtenido. Los resultados de la validación determinaron la elección de los siguientes parámetros específicos para cada representación. Para BoW se escogieron los 5.000 unigramas de palabras (1gp) más frecuentes con ponderación *tf-idf* sobre los datos pre-procesados y los 5.000 trigramas de caracteres (3gc)

[1] https://www.python.org/

más relevantes con ponderación *tf-idf,* sobre los datos sin procesar. Con respecto a k-TVT, se probó con *k* variando en el rango de 0 a 5, seleccionando finalmente la configuración *k*=0 con los textos pre-procesados y *k*=4 con datos sin pre-procesar.

Para LIWC, se entrenó cada clasificador con todas las características que provee la herramienta y luego se consideraron tres métodos de selección de atributos con el fin de determinar la relevancia de cada característica. Los mejores resultados se obtuvieron con el enfoque empotrado, el cual consiste en entrenar un algoritmo de aprendizaje (como árboles de decisión o regresores logísticos) y utilizar los coeficientes generados en ese proceso para la selección de los atributos. Finalmente, se decidió informar los resultados obtenidos con el método de árboles de decisión denominado *Extra Tree Classifier* (ETC), que determinó que 27 características eran las más relevantes, a diferencia del método *Logistic Regression* (LR) y regularización L2 que fueron 40 las elegidas. En ambos casos se puede observar una reducción de más de la mitad de los 94 atributos totales que devuelve LIWC.

En cuanto a los modelos de incrustaciones de palabras, para todos los casos probados se decidió utilizar los vectores pre-entrenados dado el esfuerzo computacional requerido para obtener nuevos. Si bien se realizaron algunas pruebas preliminares entrenando los vectores desde cero con el conjunto de entrenamiento, los resultados fueron similares a los obtenidos con los pre-entrenados y se requirió bastante tiempo de cómputo en la elaboración de los vectores. Además del ahorro de tiempo de cómputo, otra ventaja de utilizar los vectores pre-entrenados es que al haber sido creados con grandes cantidades de texto, comprimen un mayor conocimiento, lo que permite que el clasificador pueda generalizar mejor, evitando el sesgo que se produciría si es entrenado sólo con la colección de depresión.

Los datos de entrenamiento para GloVe fueron particionados en palabras de al menos un caracter sin considerar la puntuación. En el caso de BERT,

los textos de los usuarios fueron divididos en oraciones, ya que por la forma en la que funciona el método de representación, la palabra tiene que estar en su contexto. Además, como se empleó la librería *bert-as-service*[1] para extraer los vectores del modelo, fue necesario fijar una longitud máxima de oración, en este caso fue de 40 palabras. Esta herramienta por defecto devuelve el vector de la oración, condensando la información desde la segunda a la última capa oculta de la red neuronal. Finalmente, para GloVe, los vectores de los documentos se crearon haciendo un promedio de las incrustaciones de las palabras que aparecen en cada uno, mientras que para BERT, el vector de los documentos se generó promediando las incrustaciones de las oraciones del mismo.

A modo informativo vale la pena mencionar que los vectores de GloVe utilizados fueron entrenados con 2 billones de tuits; el vocabulario contiene 1,2 millones de términos y cada vector posee 200 dimensiones. Los mismos están disponibles en la página oficial del grupo de NLP de Stanford[2] desde 2014. Se decidió emplear este modelo ya que como la colección de depresión fue recolectada a partir de una red social, el vocabulario sería más similar al de Twitter, es decir, incluiría palabras correspondientes a un discurso más informal.

Por otro lado, los vectores pre-entrenados de BERT son más grandes, tienen 768 características, aunque se corresponden con la versión *Base* (existe una versión con 1.024 atributos) y son más recientes (están disponibles desde 2019). Estos fueron entrenados con colecciones de libros (800 millones de palabras) y la Wikipedia en inglés (2,5 billones de palabras). BERT-Base está disponible para su descarga en el repositorio oficial de Google[3].

Asimismo, con la finalidad de valorar los aportes de las representaciones estudiadas en este artículo, se establecieron valores de referencia para su

[1] https://github.com/hanxiao/bert-as-service
[2] https://nlp.stanford.edu/projects/glove/
[3] https://github.com/google-research/bert

comparación. Estos valores se obtuvieron simplemente mediante la intersección del vocabulario entre los documentos de las dos clases. Esto se implementó a partir del cómputo de una bolsa de palabras con ponderación binaria de unigramas de palabras (1gp bin).

Analizando la **tabla 1**, se puede observar el buen desempeño de los modelos que usan los enfoques de representación más modernos. GloVe es el que obtiene los valores más bajos (y por tanto, los mejores) para las métricas de tiempo. Esto es, 8,60 y 6,62 para ERDE 5 y ERDE 50 respectivamente. Por otro lado, BERT alcanza la mayor F1 (0,52) y el mejor valor de precisión (0,65).

Tabla 1. Resultados del estudio experimental

Representación		AC	E_5	E_50	F1	Π	ρ
Valores de Referencia	1gp bin	SVM	9,84	7,41	0,48	0,63	0,39
BoW	1gp *tf-idf*	SVM	9,25	7,14	0,43	0,53	0,37
	3gc *tf-idf*	SVM	9,01	6,93	0,47	0,53	0,42
k-TVT	k=0	RF	8,94	6,78	0,48	0,42	0,54
	k=4	SVM	9,16	6,70	0,43	0,35	**0,56**
LIWC	ETC	RF	8,88	7,37	0,35	0,34	0,37
	LR - L2	SVM	9,38	7,45	0,47	0,59	0,39
Incrustaciones	GloVe	SVM	**8,60**	**6,62**	0,46	0,50	0,43
	BERT	RF	8,90	6,80	**0,52**	**0,65**	0,43

En cuanto a k-TVT, resulta evidente su eficiencia para las tareas de clasificación temprana ya que obtuvo valores de ERDE muy similares a los de las incrustaciones. En este punto es importante remarcar que el tamaño

y costo de obtención de las representaciones k-TVT son bastante menores que los de los modelos de incrustaciones de palabras, lo que significa una importante ventaja a la hora de seleccionar un método de representación eficiente en la métrica ERDE. Además, con k-TVT se obtuvo la segunda mejor F1 (0,48) y los dos mejores valores de cobertura (0,54 y 0,56).

Al contrario de lo esperado, los modelos que utilizaron LIWC no alcanzaron las expectativas, ya que si bien para ERDE 5 LIWC-ETC obtuvo el segundo mejor valor (8,88), la F1 es demasiado baja (0,35) y el ERDE 50 fue uno de los valores más altos (7,37) por detrás de los valores de referencia. Si bien con LIWC se recuperó valiosa información relacionada con el uso de componentes lingüísticos, procesos psicológicos y gustos personales de los *posts*, dichas variables no fueron suficientes para que el modelo sea capaz de extraer rasgos lingüísticos del grupo depresivo. No obstante, observando la prestación del modelo LR-L2 y considerando la baja dimensionalidad de los vectores de representación de los textos, se puede concluir que podría combinarse con otro tipo de características y aumentar así su poder informativo.

Los resultados del enfoque más clásico siguen evidenciando buena competitividad con respecto a la medida F1 ya que BoW con trigramas de caracteres y ponderación *tf-idf* queda en tercera posición (0,47) siendo levemente menor que lo obtenido con la bolsa binaria de palabras (el valor de referencia) y k-TVT con k=0 (0,48). La misma situación se puede observar con la precisión (Π), siendo el segundo mejor valor (0,63) apenas por debajo del mejor obtenido con BERT (0,65). Sin embargo, desde el punto de vista temporal, BoW necesita analizar muchas partes antes de poder tomar una decisión y por ello los valores de ERDE son mayores que los de otros modelos.

Finalmente podemos concluir que, así como se mencionó anteriormente, el estado del arte en representaciones de textos está dominado por los enfoques de incrustaciones de palabras y los resultados obtenidos dejan vislumbrar los motivos. En este trabajo se comprobó la potencialidad de las

incrustaciones y cabe destacar que los resultados del modelo que utiliza BERT con RF logran un buen equilibrio entre las métricas evaluadas, maximizando F1, precisión y cobertura, y minimizando las medidas ERDE.

2.3. Análisis de las características

A continuación, analizaremos algunos de los rasgos lingüísticos recuperados por dos de los modelos propuestos en el estudio experimental: BoW y GloVe con SVM para la detección temprana de signos de depresión en medios sociales.

La **figura 2** presenta un ranking de los 25 términos más relevantes para BoW, según las configuraciones detalladas en la subsección anterior. En las gráficas se puede observar que en la clase positiva (barras azules) aparecen términos relacionados con la primera persona del singular como *“i”*, *“am”*, *“was”*, *“my”*, *“me”* para los unigramas de palabras, mientras que “__i” y “._I” se encuentran dentro de las primeras posiciones del ranking de trigramas de caracteres. Esto se corresponde con estudios psicológicos que afirman que las personas con depresión tienden a centrar la atención en sí mismas (Al-Mosaiwi y Johnstone 2018:538). En cambio, en la clase negativa, algunos de estos términos aparecen también, pero en posiciones del ranking más bajas por tener menor valor de *tf-idf*. Por otro lado, es notorio que *“not”* y *“but”* son más relevantes para la clase positiva, ya que la negación es otro de los rasgos lingüístico presentes en las personas que están cursando la enfermedad (Al-Mosaiwi y Johnstone 2018:538). Los trigramas como *“ed_”*, *“ing”* y *“ng_”*, que se corresponden con terminaciones de tiempos verbales pasados y con el presente continuo, también es otra característica del discurso de las personas con depresión.

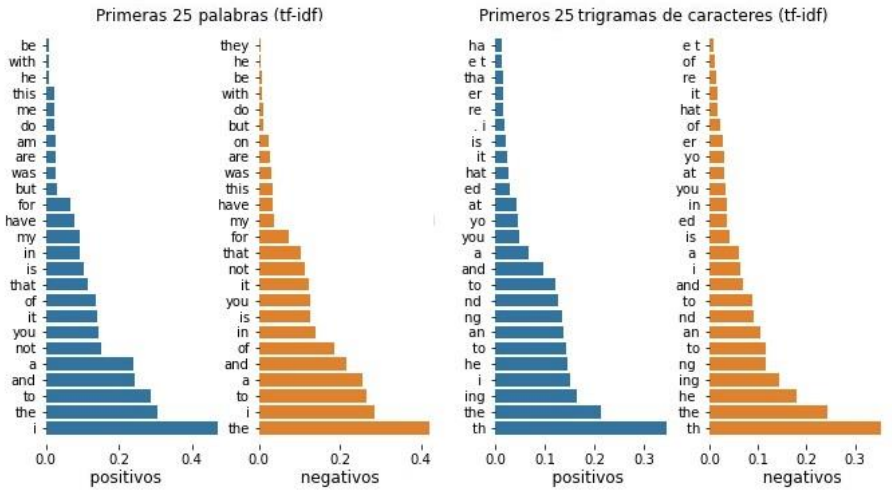

Figura 2. Los 25 términos más relevantes de BoW según *tf-idf* para depresivos (positivos) versus no depresivos (negativos).

Por otro lado, en la **figura 3** se puede visualizar una representación en dos dimensiones de las incrustaciones de GloVe de un fragmento de un caso positivo. Para cada palabra del ejemplo (punto azul y letra en color negro) se calcularon las palabras más cercanas (punto y letra en color gris claro). Además, se agregaron dos ampliaciones de las partes más aglomeradas para su mejor visualización. En la ampliación del extremo izquierdo aparecen agrupados muchos de los términos relevantes vistos en el análisis de la representación BoW.

Si se observa la figura completa, sin embargo, no sólo se puede percibir la similitud ortográfica entre las palabras, como sucede con *"therapy"* y *"therapist",* que son derivados y por lo tanto están próximas, sino que se puede observar claramente el contenido semántico que poseen las incrustaciones. Esto es, las palabras próximas aritméticamente, calculadas por la distancia coseno de sus vectores, también guardan relación con sus significados. Por ejemplo, *"childhood"* (palabra objetivo) y *"parent"*

(palabra cercana). Otro ejemplo es *"disorders"* que aparece entre *"depression"* y *"therapy"*.

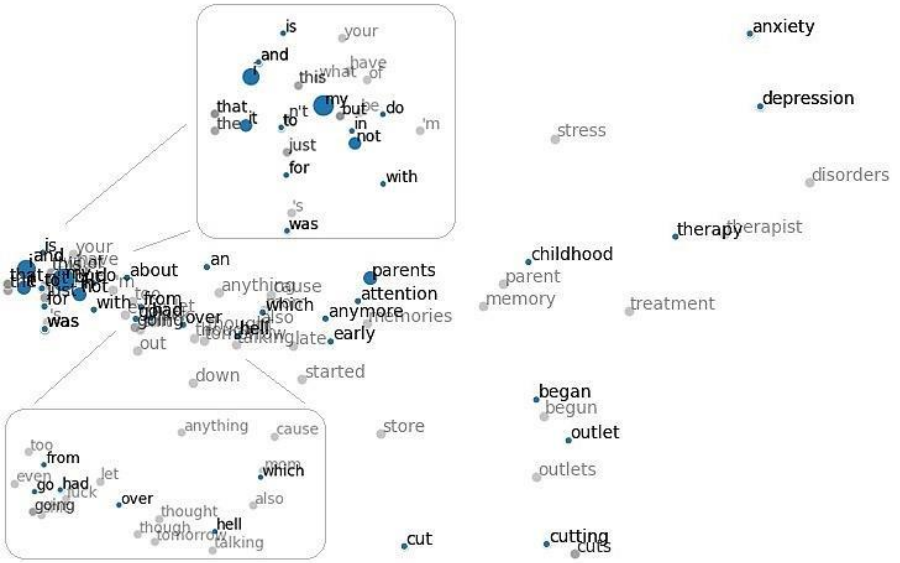

Figura 3. Representación en dos dimensiones de los vectores pre-entrenados de GloVe para un ejemplo positivo (depresivo).

Además de visualizar algunos rasgos lingüísticos capturados por las incrustaciones de palabras, la **figura 3** permite notar la robustez de la representación gracias al contenido semántico que posee. Esto último se puede concluir ya que si por ejemplo en la frase no apareciera la palabra *"depression",* pero sí *"disorders"*, el vector del documento no variará significativamente.

Conclusiones

En este trabajo se analizaron diferentes representaciones de textos en la generación de modelos de aprendizaje automático para la detección temprana de depresión. La hipótesis que se trabajó fue que un mayor

entendimiento de la estructura del lenguaje permite obtener mejores modelos computacionales.

Para ello, en primer lugar se analizó la colección de textos relacionados con el trastorno de la depresión y se comprobó que es una tarea que representa un verdadero desafío, ya que se visualiza solapamiento de vocabulario entre las clases positiva y negativa. Sin embargo, este vocabulario en común presenta un escenario real y sugiere nuevos interrogantes como los relacionados con la forma en la cual distintas representaciones de textos pueden capturar rasgos lingüísticos prevalentes en personas depresivas.

Es así que, en segundo lugar, se realizó el análisis de las representaciones seleccionadas. Se concluyó que enfoques clásicos como la bolsa de palabras aún siguen siendo competitivos en comparación con los surgidos recientemente y sirven como una buena primera aproximación ya que se pueden calcular de forma rápida.

Sin embargo, con respecto a los nuevos enfoques, se estudiaron modelos con incrustaciones de palabras y se ratifica que el contenido aprendido en función de otros textos (modelos pre-entrenados), introduce información implícita en el clasificador que beneficia la clasificación temprana (visualizado en los valores de ERDE para GloVe y en la medida F1 obtenida con BERT que también posee muy buena precisión). Además, el contexto se vuelve indispensable para tratar el problema del solapamiento mencionado anteriormente, dadas las distintas acepciones de las palabras.

Como trabajo a futuro se pretende continuar el estudio de representaciones aprendidas, profundizar el análisis de BERT e incluso probar con las versiones más grandes de vectores pre-entrenados, para comprobar si efectivamente mejora la clasificación al contar con el doble de características.

Referencias bibliográficas

Acosta-Hernández M. E. et al. (2011). Depresión en la infancia y adolescencia: enfermedad de nuestro tiempo. *Archivos de Neurociencia*, 16(3), 156–161.

Al-Mosaiwi M. y Johnstone T. (2018). In an Absolute State: Elevated Use of Absolutist Words Is a Marker Specific to Anxiety, Depression, and Suicidal Ideation. *Clinical Psychological Science*, 6(4), 529–542.

Boston University School of Medicine. (2 de septiembre de 2020). *COVID-19 has likely tripled depression rate, study finds*. ScienceDaily. Accedido el 24 de Noviembre de 2020: www.sciencedaily.com/releases/2020/09/200902152202.htm

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5–32.

Cagnina, L. C. et al. (2019). k-TVT: a flexible and effective method for early depression detection. En: *XXV Congreso Argentino de Ciencias de la Computación CACIC 2019. Libro de Actas*, 547–556.

Calvo, R. A. et al. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, *23*(5), 649–685.

Chandran, D. et al. (2019). Use of Natural Language Processing to identify Obsessive Compulsive Symptoms in patients with schizophrenia, schizoaffective disorder or bipolar disorder. *Scientific Reports*, *9*(1), 1–7.

Cortes, C. y Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273–297.

Devlin, J. et al. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. En: *Proceedings of the 2019 Conference of the North American Chapter of the ACL: Human Language Technologies*, 1, 4171–4186.

Funez, D. G. et al. (2018). UNSL's participation at eRisk 2018 Lab. En: *Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum*, 2125.

Harris, Z. S. (1954). Distributional structure. Word, 10(2-3), 146-162.

Jurafsky, D. y Martin, J. H. (2020). Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition. Tercera Edición. En prensa. Borrador accedido el 5 de Marzo de 2021: https://web.stanford.edu/~jurafsky/slp3/ed3book_dec302020.pdf

Kemp, S. *More than half of the people on earth now use social media*. Datareportal. Accedido el 26 de Noviembre de 2020: https://datareportal.com/reports/more-than-half-the-world-now-uses-social-media

Li, Z. et al. (2011) Fast text categorization using concise semantic analysis. *Pattern Recognition Letters*. 32(3), 441–448.

Losada, D. E. y Crestani, F. (2016). A test collection for research on depression and language use. En: *Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2016. LNCS*, 9822, 28–39.

Losada, D. E., Crestani, F. y Parapar, J. (2018). Overview of erisk: early risk prediction on the internet. En: *Experimental IR Meets Multilinguality, Multimodality, and Interaction. CLEF 2018. LNSC,* 11018, 343–361.

Low, D. M. et al. (2020). Natural Language Processing Reveals Vulnerable Mental Health Support Groups and Heightened Health Anxiety on Reddit During COVID-19: Observational Study. *Journal of medical Internet research*, *22*(10), e22635.

Mikolov, T. et al. (2013). Efficient estimation of word representations in vector space. En *Proceedings of Workshop at International Conference on Learning Representations (ICLR).*

Nguyen, M. H. et al. (2020). Changes in Digital Communication During the COVID-19 Global Pandemic: Implications for Digital Inequality and Future Research. *Social Media + Society*, *6*(3), 1–6.

Pennebaker, J. W. et al. (2015). *Linguistic Inquiry and Word Count: LIWC2015*. Austin, TX: Pennebaker Conglomerates.

Pennington, J., Socher, R. y Manning, C. D. (2014). GloVe: Global vectors for word representation. En *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.

Salton, G. y McGill, M. J. (1983). *Introduction to modern information retrieval*. New York, NY: McGraw-Hill.

Spärck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28, 11–21.

Turney, P. D. y Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research (JAIR)*, 37, 141–188.

Vaswani, A. et al. (2017). Attention is all you need. En *Advances in Neural Information Processing Systems*, 30, 5998–6008.

Zhang, Y., Jin, R. y Zhou, Z. H. (2010). Understanding bag-of-words model: a statistical framework. *International Journal of Machine Learning and Cybernetics*, *1*(1–4), 43–52.

Notas biográficas

María José Garciarena Ucelay

La Licenciada María José Garciarena Ucelay forma parte del Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de

la Universidad Nacional de San Luis (UNSL, Argentina), donde participa del proyecto "Aprendizaje automático y toma de decisiones en Sistemas Inteligentes para la Web". Su campo de investigación abarca el Procesamiento del Lenguaje Natural aplicado a la Minería de Textos. Actualmente está terminando su trabajo final para optar a la Maestría en Ciencias de la Computación. Su tema de investigación versa sobre distintos métodos de representación de textos para la detección temprana de riesgos en la Web.

Leticia Cecilia Cagnina

La Doctora Leticia Cecilia Cagnina es investigadora del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y trabaja en el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional en la Universidad Nacional de San Luis (UNSL). Recibió su título de Doctora en Ciencias de la Computación en la UNSL y es profesora en la misma universidad. Es directora de línea en el proyecto denominado Aprendizaje automático y toma de decisiones en sistemas inteligentes para la Web en la UNSL. Sus temas de interés están relacionados con las técnicas de Procesamiento del Lenguaje Natural aplicadas especialmente a la minería de textos (clasificación, agrupamiento, extracción de características, métodos de representación, etc.) y a la detección de riesgos en la web. Es co-autora de varios capítulos en importantes libros y ha publicado varios artículos en revistas y conferencias nacionales e internacionales.

Marcelo Luis Errecalde

Marcelo Luis Errecalde es Doctor en Ciencias de la Computación y Profesor en la Universidad Nacional de San Luis y en la Universidad Nacional de La Patagonia Austral (Argentina). Dicta cursos de posgrado en la Universidad Nacional de La Plata y actualmente es el Director de LIDIC, el Laboratorio de Investigación y Desarrollo en Inteligencia Computacional de la Universidad Nacional de San Luis.