






Detección de rasgos en estudiantes con tendencia suicida en Internet aplicando Minería Web

Detection of traits in students with suicidal tendencies on Internet applying Web Mining

-  Dr. Iván Castillo-Zúñiga. Profesor Titular, Departamento de Sistemas y Computación, TecNM-Instituto Tecnológico el Llano Aguascalientes (México) (ivan.cz@llano.tecnm.mx) (<https://orcid.org/0000-0001-8017-5908>)
-  Dr. Francisco-Javier Luna-Rosas. Profesor Titular, Departamento de Sistemas y Computación, TecNM-Instituto Tecnológico de Aguascalientes (México) (fcoluna2000@yahoo.com.mx) (<https://orcid.org/0000-0001-6821-4046>)
-  Dr. Jaime-Iván López-Veyna. Profesor Titular, Departamento de Sistemas y Computación, TecNM-Instituto Tecnológico de Zacatecas (México) (ivanlopezveyna@zacatecas.tecnm.mx) (<https://orcid.org/0000-0002-9225-3202>)

RESUMEN

Este artículo presenta un modelo de análisis de datos en Internet basado en Minería Web con el objetivo de encontrar conocimiento sobre grandes cantidades de datos en el ciberespacio. A fin de probar el método propuesto, se analizaron páginas web sobre el suicidio como caso de estudio con la intención de identificar y detectar rasgos en estudiantes con tendencias suicidas. El procedimiento considera un Web Scraper para localizar y descargar información de Internet, así como técnicas de Procesamiento de Lenguaje Natural para la recuperación de los vocablos. Con el propósito de explorar la información, se construyó un conjunto de datos basado en Tablas Dinámicas y Ontologías Semánticas, especificando las variables predictivas en jóvenes con inclinación suicida. Por último, para evaluar la eficiencia del modelo se utilizaron algoritmos de Aprendizaje de Máquina y Aprendizaje Profundo. Cabe destacar que se optimizaron los procedimientos para la construcción del dataset (utilizando Algoritmos Genéticos) y obtención del conocimiento empleando Cómputo Paralelo y Aceleración con Unidades de Procesamientos de Gráfico (GPU). Los resultados revelan una precisión del 96,28% sobre la detección de las características en adolescentes con tendencia suicida, alcanzando el mejor resultado a través de una Red Neuronal Recurrente con un 98% de precisión. De donde se infiere que el modelo es viable para establecer bases sobre mecanismos de actuación y prevención de comportamientos suicidas, que pueden ser implementados en instituciones educativas o distintos actores de la sociedad.

ABSTRACT

This article presents an Internet data analysis model based on Web Mining with the aim to find knowledge about large amounts of data in cyberspace. To test the proposed method, suicide web pages were analyzed as a study case to identify and detect traits in students with suicidal tendencies. The procedure considers a Web Scraper to locate and download information from the Internet, as well as Natural Language Processing techniques to retrieve the words. To explore the information, a dataset based on Dynamic Tables and Semantic Ontologies was constructed, specifying the predictive variables in young people with suicidal inclination. Finally, to evaluate the efficiency of the model, Machine Learning and Deep Learning algorithms were used. It should be noticed that the procedures for the construction of the dataset (using Genetic Algorithms) and obtaining the knowledge (using Parallel Computing and Acceleration with GPU) were optimized. The results reveal an accuracy of 96.28% on the detection of characteristics in adolescents with suicidal tendencies, reaching the best result through a Recurrent Neural Network with 98% accuracy. It is inferred that the model is viable to establish bases on mechanisms of action and prevention of suicidal behaviors, which can be implemented in educational institutions or different social actors.

PALABRAS CLAVE | KEYWORDS

Conducta suicida, cibersuicidio, minería web, aprendizaje de máquina, aprendizaje profundo, redes neuronales recurrentes.

Suicidal behavior, cybersuicide, web mining, machine learning, deep learning, recurrent neural networks.

1. Introducción

Los beneficios que las nuevas tecnologías han traído a nuestras vidas no se pueden negar, sin embargo, también han generado problemas que antes eran desconocidos, uno de ellos es lo que se denomina como ciber-suicidio. Según López-Martínez (2020), es un fenómeno que se refiere a la influencia de la información que circula por Internet e incita a cometer suicidio. El suicidio es la segunda causa de muerte en la población mundial de 10 a 24 años, lo que representa 100.000 adolescentes muertos al año por esta causa. El suicidio es un comportamiento complejo que se construye en el tiempo y depende de múltiples factores biológicos, familiares, sociales, educativos, entre otros.

Internet y las redes sociales han transformado de forma rápida y sustancial el modo como los adolescentes y personas se comunican y tienen acceso a información relacionada con el suicidio, esta gran cantidad de sitios y volúmenes de información se conoce comúnmente como Big Data. Molina y Restrepo (2018) mencionan que los usuarios no solo buscan páginas web que brindan información relacionada con métodos y formas para realizar suicidio, sino que también buscan ayuda, apoyo y orientación frente al sufrimiento que experimentan derivado de los pensamientos suicidas, la tristeza, soledad y ansiedad. Estas razones hacen necesaria una detección temprana y una respuesta inmediata ante este problema de salud pública, en donde la información es esencial para su análisis, y un medio útil para educar y prevenir el comportamiento suicida. Las instituciones educativas juegan un papel muy importante en la promoción de estilos de vida saludable y a la hora de brindar apoyo temprano a jóvenes en riesgo.

Desde diferentes enfoques se han realizado muchos esfuerzos para estudiar los grandes volúmenes de información que se generan en Internet aplicando técnicas automáticas, con el fin de analizar y predecir eventos que afecten a la sociedad, como el suicidio. Con esta perspectiva existen programas, desarrollos, algoritmos y procesos en evolución que siguen siendo estudiados por los investigadores. Según Nalini y Sheela (2014), comprender la relación entre la capacidad de análisis y las características de un evento delictivo puede ayudar a los investigadores a utilizar esas técnicas de forma más eficiente para identificar tendencias y patrones, abordar distintas problemáticas e incluso predecir delitos. Por estas razones, esta investigación propone un nuevo enfoque a los procesos de Minería Web, integrando técnicas de Aprendizaje Automático y Aprendizaje Profundo empleando Cómputo Paralelo Acelerado con GPU para procesar grandes volúmenes de datos con el objetivo de identificar y detectar rasgos en adolescentes con tendencias suicidas.

Investigaciones como las de Bonami et al. (2020), Denia (2020), Kim y Chung (2019), Anggraini et al. (2018) y Roy et al. (2017) han tenido como objetivo el tratamiento de grandes volúmenes de datos, el análisis de información con técnicas de Inteligencia Artificial y el Procesamiento del Lenguaje Natural (PLN), que se resumen en la Tabla 1.

Tabla 1. Características de estudios relacionados al Big Data e Inteligencia Artificial

Trabajos Relacionados	Big Data	Inteligencia Artificial					Enfoque de la Investigación
	Fuente de Datos	Minería de Datos	PLN	Cómputo			
				Secuencial	Paralelo	Aceleración con GPU	
Bonami, Piazzentini & Dala-Possa (2020)	Sensor	Sí	No	Sí	No	No	Educativa
Denia (2020)	Red Social	No	Sí	Sí	No	No	Impacto social de discursos científicos en Twitter
Kim & Chung (2019)	Sitios Web	Sí	Sí	Sí	No	No	Salud
Anggraini, Sucipto & Indriati (2018)	Red Social	Sí	Sí	Sí	No	No	Ciberbullying
Roy et al. (2017)	Sitios Web	Sí	Sí	Sí	No	No	Cibercrimen Detección de intrusos
Actual investigación	Sitios Web	Sí	Sí	Sí	Sí	Sí	Suicidio

Algunos de estos proyectos de investigación presentan métodos orientados a la búsqueda de conocimiento dirigidos a la minería de datos (Bonami et al., 2020; Kim & Chung, 2019; Anggraini et al., 2018; Roy et al., 2017), otros estudios se han enfocado en la recuperación de información utilizando técnicas de Procesamiento de Lenguaje Natural (Denia, 2020; Kim & Chung, 2019; Anggraini et al.,

2018). Una de las complejidades de esta investigación radica en que la información se encuentra en formato no estructurado, en grandes cantidades de datos dispersos en sitios web con diferentes protocolos de seguridad y comunicación, además los textos presentan errores ortográficos y abreviaturas. Para iniciar el proceso de detección de las características de estudiantes con tendencia suicida es necesario transformar la información en datos estructurados, prosiguiendo con el análisis empleando técnicas de Aprendizaje de Máquina para la obtención de patrones que se conviertan en conocimiento y valor agregado.

La presente investigación se basa en la combinación de técnicas de Big Data e Inteligencia Artificial con una estrategia genética utilizando Cómputo Paralelo, adaptando procesos de la Web Semántica con procedimientos basados en Ontologías Semánticas, Vocabularios y Tablas Dinámicas, integrando métodos de PLN para el procesamiento de datos, como: limpieza, separación de información del código computacional y eliminación de palabras sin significado, tokenización (separación de palabras), sinónimos, lematización (raíz de palabras) y frecuencia de término con el propósito de generar información valiosa y de valor agregado, obtenido del análisis de un gran número de páginas con contenido suicida que se descargan de la Web. Cabe resaltar que, dentro del proceso mencionado, construimos una Ontología Semántica que clasifica la información a través de conceptos validados con técnicas de Aprendizaje de Máquina y Aprendizaje Profundo, aplicando Cómputo Paralelo Acelerado con GPU.

El motivo principal de la investigación es explorar los beneficios que se pueden obtener de la información que circula en la web, buscando descubrir patrones, correlaciones desconocidas, información adicional que pueda ser de gran utilidad para dar soporte al proceso de toma de decisiones en la solución de problemas. Los objetivos del estudio contemplan la detección de rasgos en estudiantes con tendencias suicidas, así como, llevar a la práctica técnicas de recuperación de páginas web, pre-procesar, analizar y clasificar datos, finalmente, crear una categorización del corpus lingüístico del suicidio. La principal aportación del artículo se centra en un modelo de estudio sobre datos relacionados con problemas de ciberacoso en Internet, concretamente el cibersuicidio. Dentro del marco del estudio se tiene como límite específicamente la detección de vocablos de Cibersuicidio en Internet por medio de ontologías semánticas haciendo uso del análisis predictivo en un computador; cabe señalar que no se considera dentro del alcance el análisis descriptivo de los vocablos.

1.1. Cibersuicidio y su impacto en estudiantes

López-Martínez (2020) menciona que el suicidio es un fenómeno de naturaleza multifactorial, y que, en la actualidad, con la aparición de las nuevas Tecnologías de la Información y la Comunicación (TIC), supone un nuevo escenario y con él, un nuevo problema para la prevención de la conducta suicida. En este contexto, nace un nuevo concepto, el cibersuicidio, que hace referencia a la acción de quitarse la vida, motivado por la influencia de páginas prosuicida, foros y salas de chat en Internet, entre otras variables. Dentro del mismo marco, Olivares (2019) señala que el cibersuicidio se refiere a la influencia de la información que circula en Internet, así como la incitación que hay en esos medios para ejercerlo. Aunado a la situación, Moreno y Blanco (2012) indican que es un fenómeno que se propaga con rapidez por todo el planeta, aumentando los casos de suicidio año tras año.

Durkheim (2008) señala que el suicidio es todo caso de muerte que resulte, directa o indirectamente, de un acto positivo o negativo, realizado por la víctima misma, sabiendo ella que debía producir este resultado; desde otra perspectiva, Berengueras (2018) menciona que el suicidio es cuando un individuo atenta contra las leyes de su propia naturaleza al tomar la decisión de acabar con su vida, en un acto profundamente expresivo y esencial de ser comprendido y escuchado en sus manifestaciones previas. Por su parte, Marchiori (2015) establece que los instrumentos más frecuentes para cometer suicidio son las armas de fuego, armas blancas, cuerdas, alambres, telas para el ahorcamiento, fármacos, drogas, arrojarse de puentes, edificios, a las vías del tren o paso de automóviles, asfixia por inmersión, venenos, combustibles (gas, carbón, keroseno, nafta), entre otros.

Investigaciones como las de Arevalos (2020), SeGob (2021), Luna y Dávila (2018), Sánchez- García et al. (2018), Blanco (2019), Healy (2019) y la Organización Mundial de la Salud (OMS) (2019) revelan que el suicidio es un problema a nivel mundial, el cual se puede producir a cualquier edad. No obstante, es prevenible mediante intervenciones oportunas, en donde el sector educativo juega un

factor importante, debido a que se reporta que en 2016 fue la segunda causa de muerte para los jóvenes de edades entre los 15 y 19 años. En Argentina se obtuvieron testimonios de jóvenes escolarizados en escuelas secundarias urbanas periféricas relacionados con la conducta suicida, en donde destacan los siguientes rasgos: menosprecio, humillación, bullying, ninguneo, obesidad, maltrato familiar y falta de apoyo, divorcio de padres, violación, abuso, complejos con su cuerpo o imagen, problemas de noviazgo, drogas, alcohol, indiferencia social y visualizan pocas oportunidades en el futuro (Arevalos, 2020). En México, la secretaría de gobernación reporta el impacto de la pandemia de COVID-19 en niñas y niños con respecto al suicidio, en donde en 2021 presentó una cifra record de 1.150 suicidios, incrementándose en un 37% en niños entre 10 y 14 años y un 12% en mujeres adolescentes entre 15 y 19 años (SeGob, 2021), en el mismo sentido se identifica que la agresión, violencia familiar, el rezago educativo, el consumo de alcohol o tabaco son factores de riesgo asociados con el intento de suicidio en adolescentes sobre todo en las mujeres más jóvenes (Luna & Dávila, 2018). En España se revela que el 7,7% de adolescentes entre 12 y 19 años mostraron dificultades en el ajuste emocional y mayor ideación suicida derivada de conductas relacionadas con el acoso escolar y el consumo de tabaco y cannabis (Sánchez-García et al., 2018), en el mismo sentido los registros del Instituto Nacional de Estadística reportaron 3.679 muertes por suicidio en el año 2017 con un 74% para varones y un 26% para mujeres, con una media de 10 suicidios al día (Blanco, 2019). En Estados Unidos el suicidio cobró la vida de 5.016 hombres y 1.225 mujeres de entre 15 y 24 años de edad en el año de 2017, con un índice de suicidio juvenil de 14,6 por cada 100.000 habitantes (Healy, 2019). Por su parte, la OMS (2019) revela que, en el mundo, cerca de 800.000 personas se quitan la vida y muchas más intentan hacerlo, en donde el 79% de los suicidios tuvieron lugar en países de ingresos bajos y medianos. La ingestión de plaguicidas, el ahorcamiento y las armas de fuego son los métodos más comunes para quitarse la vida.

1.2. Conducta suicida en adolescentes

Beaven-Ciapara et al. (2018) describen un estudio sobre los factores de riesgo asociados a la conducta suicida en jóvenes de 13 a 18 años en la comunidad de Guaymas, Sonora, México. Los resultados indican que el factor psicosocial que se está presentando son las familias disfuncionales, causando depresión y baja autoestima, en donde el modelo utilizado relaciona la conducta suicida con este hallazgo en un 37%, presentándose con mayor frecuencia en mujeres. El estudio fue realizado a 120 estudiantes de secundaria y preparatoria (41% hombres y 59% mujeres), y su análisis estadístico fue fundamentado con el software SPSS versión 21.0. A su vez, Mosquera (2016) presenta una revisión no sistemática de la literatura sobre conducta suicida infantil, revelando que dentro de los factores de riesgo más destacados se encuentran: ser hombre, tener intentos de suicidios previos, exclusión social, conflicto sentimental. Además, se observa una alta comorbilidad con trastornos depresivos, trastorno bipolar y esquizofrenia. Entre los tratamientos más eficaces destacan terapia dialecto-conductual y cognitivo-conductual. Por otro lado, Carballo-Belloso y Gómez-Peñalver (2017) encuentran en su investigación una fuerte asociación causal entre factores de vulnerabilidad individual y estresores, como experiencias de bullying en la infancia, y el posterior desarrollo de pensamientos y/o conductas auto lesivas, lo que subraya la importancia de una adecuada detección de ese factor de riesgo potencialmente modificable.

1.3. Análisis del suicidio con Inteligencia Artificial

La Tabla 2 muestra un comparativo de trabajos relacionados para el análisis y predicción del suicidio con técnicas de Aprendizaje de Máquina, en donde Ramírez-López et al. (2021) buscan predecir posibles casos de suicidio en la ciudad de Aguascalientes, México, utilizando una base de datos SQL Server del servicio de emergencias 911 que registra los suicidios consumados. En sus pruebas implementan un análisis geoespacial con el método EBK (Empirical Bayesian Kriging) de Weka para ubicar los lugares en donde se han realizado los suicidios. Los resultados revelan un 99,22% de predicción con un clasificador bayesiano en Matlab identificando gráficamente las áreas de probabilidad en donde ocurra un suicidio y en donde no. Gen-Min et al. (2020) investigan cómo predecir ideas suicidas en el personal del ejército, ya que se encuentran en un mayor estrés psicológico y un mayor riesgo de intento de suicidio en comparación con la población general. El análisis utiliza técnicas de Aprendizaje Automático que

incluyen Regresión Logística, Árboles de Decisión, Bosques Aleatorios, Arbol de Regresión, Máquinas de Soporte Vectorial y Perceptrón Multicapa, considerando cinco dominios psicopatológicos (BSRS-5), ansiedad, depresión, hostilidad, sensibilidad interpersonal e insomnio. Sus resultados superan el 98% de precisión en la clasificación utilizando un dataset basado en cuestionarios de 3.546 personas.

Por su parte, Pérez-Martínez et al. (2020) describen un análisis en Twitter donde se debate en varios países el acoso escolar, violación y suicidio relacionados con la serie de Netflix «13 Reasons Why», recuperando 154.470 tuits para su exploración. Los resultados revelan que el 51% de los tuits fueron sobre suicidio, 24% de acoso escolar y 23% de violación, en donde Estados Unidos y España fueron los países con mayor participación. Se utilizó la técnica de hashtag #13ReasonsWhy y #PorTreceRazones para recuperar los tuits, eliminando tuits duplicados y conformando 3 muestras con palabras en español, inglés, francés, portugués e italiano. Las proyecciones y cálculos se realizaron con análisis estadístico del software SPSS. Chiroma et al. (2018) identifican texto de suicidio en Twitter con Árboles de Decisión, Bayes, Bosques Aleatorios y Máquinas de Soporte Vectorial, obteniendo una precisión de entre el 34,6% y el 77,8%, alcanzando mejores resultados con el primero. Por otro lado, Du et al. (2018) extraen factores estresantes psiquiátricos de los datos de Twitter relacionados con el suicidio utilizando un enfoque basado en Aprendizaje Profundo y una estrategia de Aprendizaje por Transferencia. En donde se obtiene un 78% de precisión con Redes Neuronales Convulsionales y un 67,94% con Redes Neuronales Recurrentes.

Por último, Hermsillo-De-la-Torre et al. (2015) muestran la relación de los síntomas depresivos, la desesperanza y los recursos psicológicos sobre la tentativa de suicidio en una muestra de 96 adolescentes en Aguascalientes, México. A través del software SPSS, aplicaron estadísticos descriptivos para el análisis de proporciones y estimación de parámetros poblacionales, y estadísticos no paramétricos para las comparaciones de grupos de estudio. Posteriormente, implementaron el estadístico Rho de Spearman para conocer la manera en que las variables se asociaron y la regresión lineal para observar la relación entre ellas. Los resultados muestran que el desarrollo de la capacidad para manejar adecuadamente la tristeza, es uno de los factores a considerar como medida de prevención de suicidio para fomentar y desarrollar en los adolescentes.

Tabla 2. Comparativo de técnicas aplicadas al análisis de datos para detección de suicidio								
Trabajos Relacionados	Fuente de Datos	Técnicas de Análisis de Datos						Software usado
		Estadística	Aprendizaje de Máquina	Aprendizaje Profundo	Pre-procesamiento de Datos			
					Web Scraper	PLN	Ontología Semántica	
Ramírez-López et al. (2021)	SQL Server	No	Sí	No	No	Sí	No	Weka y MatLab
Gen-Min et al. (2020)	Cuestionario	Sí	Sí	No	No	No	No	No específica
Pérez-Martínez et al. (2020)	Twitter	Sí	No	No	No	No	No	SPSS
Chiroma et al. (2018)	Twitter	No	Sí	No	No	Sí	No	No específica
Du et al. (2018)	Twitter	No	Sí	Sí	No	Sí	No	No específica
Hermsillo-De-la-Torre et al. (2015)	Escalas recursos psicológicos	Sí	No	No	No	No	No	SPSS
Actual investigación	Sitios web	Sí	Sí	Sí	Sí	Sí	Sí	Desarrollo con Java y Python

2. Material y métodos

2.1. Metodología para ayudar a detectar rasgos en estudiantes con tendencia suicida en sitios web

El método aplicado en esta investigación se inicia con la obtención de datos en el ciberespacio y finaliza con la detección de rasgos de estudiantes con tendencia suicida. El procedimiento está formado por 3 etapas que integran técnicas de Big Data Analytics, Procesamiento de Lenguaje Natural, Web Semántica e Inteligencia Artificial, las cuales se describen a continuación.

2.1.1. Localización y descarga de páginas web de suicidio (etapa 1)

Para la localización y descarga de sitios web con contenido suicida, se desarrolló un Web Scraper basado en código abierto con las librerías JSOUP del lenguaje de programación Java, seguido de técnicas

de scraping para obtener el nombre de la página web, la dirección de Internet (URL) y realizar una copia del archivo en el disco duro del computador. Dicha información se almacena en bases de datos evitando duplicidad, unida a campos de control que permiten el pre-procesamiento de los documentos.

2.1.2. Construcción de conjuntos de datos para las pruebas (etapa 2)

Una de las partes más determinantes y complejas para la detección de rasgos de estudiantes con tendencia suicida es la construcción del conjunto de datos (dataset), ya que es la parte principal donde se realizan las pruebas de clasificación y predicción a través de técnicas de Inteligencia Artificial.

Para representar diferentes conjuntos de características suicidas en los adolescentes, se utilizó un enfoque basado en Ontologías Semánticas que nos permiten asociar conceptos a través de técnicas orientadas a objetos (clases-objetos-atributos), facilitando la agrupación de diferentes condiciones de suicidio como señales de suicidio, formas de llevarlo a cabo, tipos de suicidio, factores de riesgo, prevención, influencias y sinónimos. Donde la clase simboliza el tema principal como el suicidio, los objetos son los subtemas como señales de suicidio, y los atributos son las características, como euforia, angustia, sueño, despedirse, aislarse entre otros.

El dataset es el resultado de la transformación de datos no estructurados a datos estructurados, el cual se inicia con la eliminación del código computacional del texto, seguido de la separación de palabras (tokenización), suprimiendo las palabras sin significado (stop word), como preposiciones, pronombres, artículos, adverbios, conjunciones y algunos verbos. Finalmente, los términos que formarán parte del corpus lingüístico del suicidio se almacenan en una base de datos. Para indicar la importancia de cada rasgo suicida y obtener una mejor precisión de su análisis, se ajustaron las técnicas de Frecuencia de Término (TF) y Frecuencia Inversa del Documento (IDF) sustituyendo el término por la raíz de la palabra, la cual se obtiene con el algoritmo de Porter (2006) mediante la técnica de Lematización (stemming), en donde se considera en la búsqueda de un término como suicidio, las palabras suicidio, suicidios, suicida y suicidarse, generando un resultado más exacto.

Posteriormente, se construyen Tablas Dinámicas en MySQL utilizando como metadatos las características definidas en la Ontología Semántica que se enlazan con el corpus lingüístico generando el dataset. Es importante mencionar que el proceso de construcción del dataset se optimizó con Cómputo Paralelo a través de una estrategia genética para distribuir con igualdad la transformación de las páginas web estableciendo un clúster con los núcleos del procesador simulando un cromosoma y sus genes. El proceso del Algoritmo Genético evoluciona hasta alcanzar el óptimo en la distribución de las páginas web y cumpla el criterio de terminación con una función de adaptación basado en la media.

La evolución de la población se fundamenta con elitismo, selección por torneo, cruce de un punto y mutación con remplazo aleatorio. Por último, se establece la variable objetivo de tipo binario, con el nombre «correcta» y los valores «sí/no» que determinan la respuesta que debe generar el algoritmo de Inteligencia Artificial a partir de las variables predictoras definidas en la Ontología Semántica. Para llevar a cabo este experimento, la variable a predecir es determinada con el valor «sí», cuando la frecuencia de grupos (señales de suicidio, influencias, tipos de suicidio, sinónimos, formas de llevarlo a cabo, prevención, y factores de riesgo) es mayor que cero, y «no» en caso contrario.

2.1.3. Inteligencia Artificial aplicada a la detección de rasgos de estudiantes con tendencia suicida (etapa 3)

Con el fin de evaluar el dataset, se seleccionaron las técnicas de Inteligencia Artificial: Bosques Aleatorios, Red Neuronal, Árbol de Decisión y Regresión Logística, debido a que presentan mayor similitud con los trabajos relacionados (Gen-Min et. al., 2020, Chiroma et al., 2018). En la construcción de los algoritmos se utilizó el lenguaje de programación Python, empleando técnicas de Aprendizaje de Máquina con Cómputo Secuencial basado en la API Sklearn, Pandas y Numpy.

Dentro del mismo marco de pruebas, se consideró optimizar las estrategias mencionadas utilizando Cómputo Paralelo acelerado con GPU, aplicando la tecnología escalable de Apple que permite el desarrollo de modelos personalizados mediante la API Turicreate con el objeto (SFrame) marco de datos, que puede mutar y escalar a Big Data, considerando los algoritmos Bosques Aleatorios, Regresión Logística

y Árbol de Decisión. Por último, se realizaron pruebas con Aprendizaje Profundo utilizando Redes Neuronales Recurrentes RNN (del inglés Recurrent Neural Networks), en donde a mayor cantidad de capas y neuronas, mayor es la profundidad de la red y mayor la capacidad de aprendizaje. El modelo RNN propuesto integra una secuencia de 4 capas (una capa de entrada, una capa oculta recurrente con memoria a corto plazo LSTM del inglés Long Short Term Memory, una capa oculta y una capa de salida) que están densamente conectadas, en donde todas las neuronas de una capa están conectadas con todas las neuronas de la capa siguiente. Dentro de los algoritmos de optimización se consideraron el algoritmo de Adam, el método del Descenso del Gradiente Estocástico SGD, y la técnica de Aprendizaje Incremental RMSProp. El algoritmo de Adam, combina las ventajas de los algoritmos AdaGrad y RMSProp, donde calcula la tasa de aprendizaje de parámetros adaptativos en función del valor medio del primer momento y hace uso completo del valor medio del segundo momento del gradiente basado en la varianza no centrada. El método SGD, mantiene una tasa de aprendizaje única para actualizar todos los pesos durante todo el entrenamiento. Por último, la técnica de RMSProp considera un factor de entrenamiento diferente para cada dimensión, en donde el escalado del factor de entrenamiento se realiza dividiéndolo por la media del declive exponencial del cuadrado del gradiente.

El algoritmo RNN planteado está basado en el paralelismo de datos síncrono de Keras y TensorFlow, donde cada capa es representada por un tensor con la información de manera global (llamado lote global) y se divide en sublotes de acuerdo al número de GPU (denominados lotes locales), en donde se realizan los cálculos del gradiente tomando en cuenta la pérdida del modelo. Posteriormente, las actualizaciones originadas en los gradientes locales se fusionan con el resto de las replicas, permaneciendo de esta forma sincronizado el proceso.

2.2. Materiales

Los sistemas operativos implementados en las distintas pruebas incluyen MacOS Big Sur, Linux Ubuntu 20.04 y Windows 10. Las herramientas de software integran el lenguaje de programación Java, edición 8.0 y el gestor de bases de datos MySQL, edición 5.7 utilizados para la localización y descarga de páginas web (etapa 1). Así también, para la construcción del dataset a través del Algoritmo Genético, técnicas de PLN, construcción del Vocabulario, y el diseño de las Ontologías Semánticas basadas en objetos (etapa 2). Para los procesos de Inteligencia Artificial, se utilizó el lenguaje de programación Python, edición 3.8, con las librerías Sklearn, Pandas, Numpy, Turicreate, Keras y TensorFlow, con Cómputo Paralelo y Aceleración con GPU (etapa 3). Por último, el equipo de cómputo empleado en el presente estudio fue un MacBook Pro con procesador Intel Core i9 de 2,4 GHz, 16 GB DDR4 de memoria, disco duro Flash de 500 GB, GPU Intel UHD Graphics 630, y GPU Radeon Pro 560X, conectado a un servicio de Internet de 100MB para la localización y descarga de las páginas web.

3. Pruebas y resultados

3.1. Procedimiento de pruebas

El procedimiento de pruebas formulado permite el análisis de conjuntos de datos pequeños y grandes, donde se procesan de miles a millones de registros y se lleva a cabo según el procedimiento que se describe a continuación:

Etapa 1:

- Se localiza un conjunto de páginas web relacionadas al suicidio y se realiza una copia a disco duro para su análisis.

Etapa 2:

- Se establece una Ontología Semántica sobre los rasgos de estudiantes con tendencia suicida basada en un conjunto de libros y se procesa mediante la técnica de clases-objetos-atributos.
- Se selecciona un conjunto de palabras sin significado basadas en un estándar del SEO de Google (Landaeta, 2014).
- Se genera un corpus lingüístico del suicidio empleando técnicas de PLN y Web Semántica, optimizando con Cómputo Paralelo a través de Algoritmos Genéticos para balancear la carga.

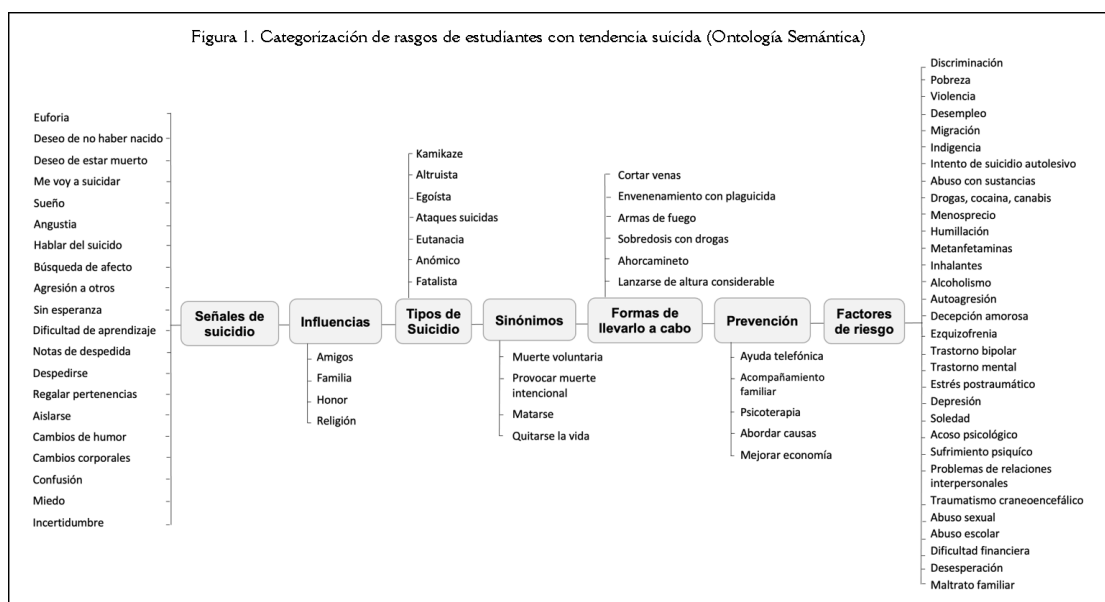
- Se construye el dataset de ciber-suicidio para las pruebas enlazando la Ontología Semántica con el corpus lingüístico y se establece la variable objetivo de tipo binario.

Etapa 3:

- Para la construcción de los programas con Cómputo Secuencial, se utilizó la librería Sklearn, específicamente con los algoritmos: Bosques Aleatorios (RandomForestClassifier), Red Neuronal (MLPClassifier), Árboles de Decisión (DecisionTreeClassifier) y Regresión Logística (LogisticRegression).
- Para la optimización de los programas con Cómputo Paralelo y Aceleración con GPU, se emplearon los algoritmos basados en la librería Turicreate, concretamente: Bosques Aleatorios (tc.random_forest_classifier), Árboles de Decisión (tc.decision_tree_classifier) y Regresión Logística (tc.logistic_classifier).
- Finalmente, para la construcción de Redes Neuronales Recurrentes con Aprendizaje Profundo, se consideró utilizar un algoritmo que conecta varias capas. La capa de entrada (inputLayer) que recibe 97 rasgos suicidas, la capa oculta recurrente con memoria a corto plazo (RecurrentLayer LSTM) definida con 97 neuronas, que consideran las posibles combinaciones de los rasgos suicidas, donde los parámetros obtenidos son el resultado de las transformaciones lineales y no lineales calculando la predicción actual tomando en cuenta el resultado anterior de los pesos y sesgos, antes de usar la función de activación (recurriendo a lo aprendido), lo que permite tener una memoria a corto plazo, la capa oculta Dense (HiddenLayer) con 48 neuronas, reduciendo los cálculos de las transformaciones y filtrando los resultados óptimos con el fin de interpretar la salida de la capa oculta LSTM, finalmente la capa de salida (outputLayer_sigmoid) de 1 neurona compatible con predicciones binarias (outputLayer_sigmoid), la cual devuelve la predicción final.

3.2. Análisis y resultados

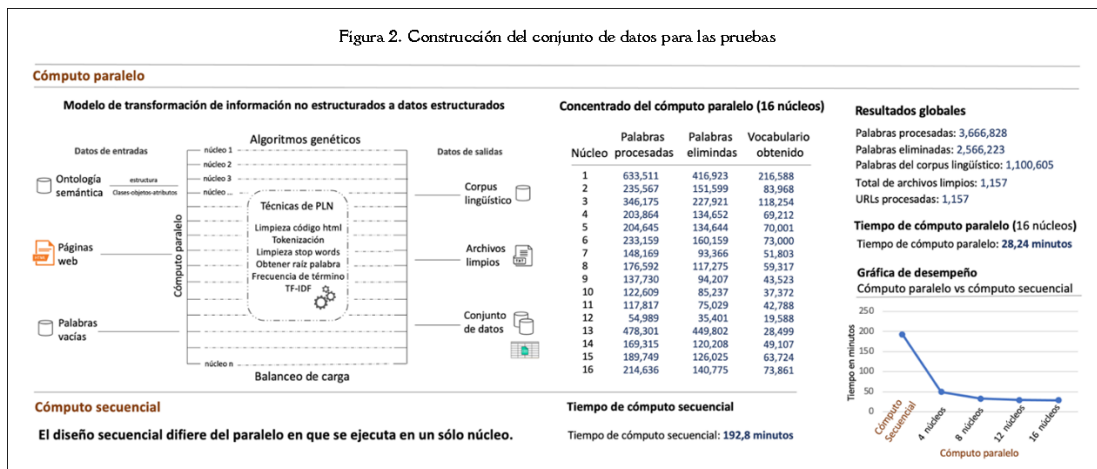
Con referencia a las páginas web relacionadas al tema de suicidio para las pruebas de detección, se localizaron 1.157 sitios web a través del Web Scraper y se realizó una copia de seguridad al disco duro. En relación a la Ontología Semántica, el resultado se presenta en la Figura 1, donde se describen los rasgos de estudiantes con tendencia suicida, los cuales son organizados por agrupaciones que reflejan los factores de riesgo, modos de prevención, formas de llevarlo a cabo, semejanzas de suicidio, tipos de suicidio, factores que tienen influencia para realizarlo, y las señales que indican tendencia suicida.



Conviene subrayar que la información se encuentra fundamentada en los libros: El suicidio (Durkheim, 2008), El pensamiento del suicidio en la adolescencia (Villardón-Gallego, 2013), El suicidio enfoque criminológico (Marchiori, 2015), Cuando nada tiene sentido: reflexiones sobre el suicidio desde la logoterapia (Rocamora, 2017), Suicidio la insoportable necesidad de ser otro (Berengueras, 2018), La huella de la desesperanza: estrategias de prevención y afrontamiento del suicidio (Urra, 2019), y el Suicidio: una mirada integral e integradora (García-Peña, 2020).

Asimismo, se presentan los resultados globales del proceso, donde se extrajeron 3.666.828 palabras, de las cuales se eliminaron 2.566.223 palabras vacías, obteniendo un corpus lingüístico de 1.100.605 palabras de ciber-suicidio y 1.157 archivos limpios con información extraída de las páginas web. Para finalizar, se incluyen los resultados de tiempo del Cómputo Secuencial.

Es preciso señalar que en la ejecución de las pruebas se obtiene un mejor desempeño con el Cómputo Paralelo (16 núcleos) sobre el Cómputo Secuencial, obteniendo una optimización del 682%, con un tiempo de 28 minutos 24 segundos, contra 192 minutos 8 segundos. Obteniendo el mejor tiempo de respuesta al utilizar un cromosoma de 16 núcleos, como se aprecia en la gráfica de desempeño. Del mismo modo, se puede apreciar, que el tiempo de cómputo se comienza a estabilizar a partir de 12 núcleos.



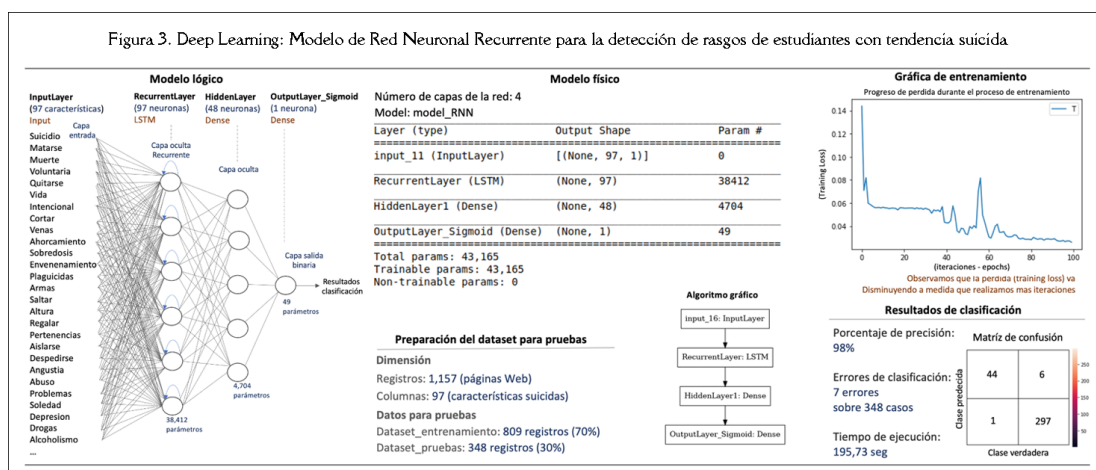
Por otro lado, los resultados de clasificación de rasgos en estudiantes con tendencia suicida, se dan a conocer en la Tabla 3. Los cuales se organizaron por técnicas de Aprendizaje de Máquina (Cómputo Secuencial y Cómputo Paralelo) y Aprendizaje Profundo. Especificando el algoritmo empleado, los valores de la matriz de confusión (verdaderos positivos, verdaderos negativos, falso positivo y falso negativo), el porcentaje de precisión y el tiempo de ejecución.

Para el Aprendizaje de Máquina con Cómputo Secuencial (librería Sklearn), el algoritmo que obtuvo el mejor resultado de clasificación fue Regresión Logística con 97,41% de precisión y un tiempo de procesamiento de 0,06 segundos. Los Bosques Aleatorios, Red Neuronal y Árbol de Decisión mostraron una estabilidad arriba del 96,55% muy cercano al primero. El mejor tiempo se obtuvo con el algoritmo Árbol de Decisión con 0,03 segundos.

Con respecto al Aprendizaje de Máquina con Cómputo Paralelo acelerado con GPU (librería Turicreate) de Apple. El algoritmo que obtuvo el mejor resultado de clasificación fue Árbol de Decisión con un 95,11% de precisión y un tiempo de respuesta de 0,01 segundos. Los algoritmos Bosques Aleatorios y Regresión Logística obtuvieron un porcentaje estable cercano al 95%. De acuerdo a las pruebas realizadas, los algoritmos de la API Turicreate no operan en Sistemas Operativos Windows, además, que no se encontró soporte para Redes Neuronales en el sitio oficial de la librería.

Tabla 3. Resultados de predicción de características en estudiantes con tendencia suicida								
Tipos de Aprendizaje	Técnicas de Aprendizaje con Python	Algoritmo	Evaluación del Aprendizaje				Precisión (%)	Tiempo Ejecución (Segundos)
			Matriz de Confusión					
			Verdadero Positivo	Verdadero Negativo	Falso Positivo	Falso Negativo		
Aprendizaje de Máquina	Cómputo Secuencial (Sklearn)	Bosques Aleatorios	293	44	6	5	96,84%	0,58 seg
		Red Neuronal	294	43	6	5	96,84%	2,09 seg
		Árbol de Decisión	290	46	8	4	96,55%	0,03 seg
		Regresión Logística	293	46	6	3	97,41%	0,06 seg
	Cómputo Paralelo (Turicoreate) CPU + GPU AMD	Bosques Aleatorios	289	52	14	4	94,98%	0,02 seg
		Regresión Logística	281	47	10	9	94,52%	0,07 seg
Aprendizaje Profundo	Cómputo Paralelo (Tensorflow y Keras) CPU	Árbol de Decisión	288	43	10	7	95,11%	0,01 seg
		Red Neuronal Recurrente	297	44	6	1	98,00%	195,73 seg

La Figura 3 presenta las soluciones del modelo RNN para la detección de rasgos de estudiantes con tendencia suicida, mostrando el modelo lógico y físico con las salidas generadas (capas), el algoritmo gráfico, la preparación del dataset para las pruebas, y los resultados de entrenamiento y clasificación.



Los resultados de la RNN revelan un 98% de predicción y 195,73 segundos de cómputo, obteniendo el mejor resultado de clasificación comparado con el resto de los algoritmos. El procedimiento examina el dataset que representa las 1.157 páginas web y las 97 características de la Ontología Semántica, empleando un 70% para el entrenamiento y un 30% para las pruebas. Cabe resaltar que, para optimizar el modelo, se implementó el método basado en el ritmo de Aprendizaje de Adam, ya que obtuvo un mejor desempeño en las pruebas, que el método SGD y RMSProp. La exactitud del modelo se aprecia en la gráfica de entrenamiento, en la cual se observa que la pérdida va disminuyendo a medida que se realizan más iteraciones del algoritmo. En el mismo sentido, la matriz de confusión devela la precisión de la clasificación reportando siete errores y 341 aciertos. Cabe mencionar que se generaron 43.165 soluciones en las diferentes capas del modelo RNN para obtener la solución óptima.

4. Discusión

Primero, es importante enfatizar que la combinación de distintas técnicas y procedimientos tiene un mayor alcance tecnológico y metodológico, lo cual se aprecia en los excelentes resultados. Donde llevar a la práctica técnicas de Big Data Analytics en los procesos de recuperación y transformación de información, así como, el Aprendizaje de Máquina para el descubrimiento de conocimiento permite agilizar el análisis y clasificación de páginas web para el usuario con un tiempo de respuesta aceptable.

Uno de los hallazgos más importantes en el estudio es que la muestra descargada de 1.157 urls representa a las páginas web sobre suicidio en español, debido a que las búsquedas del rastreador web

mostraban urls repetidas en su mayoría después de las 1.000. En cuanto al tiempo de cómputo relacionado a los procesos de Inteligencia Artificial, el algoritmo que mostró el mejor desempeño fue Árbol de Decisión empleando Cómputo Paralelo Acelerado con GPU y, en contraste, el mayor tiempo, lo obtuvo el algoritmo RNN, asociado a la misma naturaleza de recurrencia del algoritmo. Por su parte, en la transformación de los datos no estructurados a datos estructurados, se logró optimizar el tiempo en un 682% aplicando Cómputo Paralelo comparado con Cómputo Secuencial.

Respecto al objetivo de la investigación (detectar rasgos o características en estudiantes con tendencia suicida en sitios web), la metodología y arquitectura propuesta alcanzan un 98% de precisión en la clasificación, identificando patrones relacionados con: señales de suicidio, factores de riesgo, formas de llevarlo a cabo, personas que influyen en la tendencia suicida, tipos de suicidio y formas de prevención. Lo que permite establecer fundamentos para mecanismos de actuación y prevención de comportamientos suicidas, que pueden ser implementados por distintos actores en la sociedad, incluyendo instituciones educativas, organismos gubernamentales y asociaciones pro-suicida.

En relación con otros autores, como Gen-Min et al. (2020), Chiroma et al. (2018), coincidimos en que las técnicas de Aprendizaje de Máquina empleadas (Red Neuronal, Árbol de Decisión, Regresión Logística y Bosques Aleatorios) son altamente efectivas en la clasificación y predicción del suicidio. Con respecto al Aprendizaje Profundo con RNN, la presente investigación obtuvo mejores resultados de clasificación que Du et al. (2018), con 98% y 67,94% de precisión respectivamente, derivado de aplicar distintas técnicas para la construcción del dataset. Por último, se realizaron pruebas de funcionalidad con los algoritmos de Inteligencia Artificial en distintos sistemas operativos, encontrando, que el Sistema Operativo Ubuntu 20,04 de Linux, es ideal para trabajar con las distintas tecnologías tanto de hardware como de software.

5. Conclusiones

El ciberespacio se convierte en un ecosistema compartido de información en sitios web, redes sociales, y personas que comentan experiencias suicidas basado en el anonimato, en el que interactúan y expresan una opinión o información concreta sin necesidad de exponer su identidad. El comportamiento suicida en los estudiantes pone en evidencia las grietas que nuestra sociedad contemporánea está sufriendo y confronta a las comunidades académicas porque generan frustración, impotencia, culpa de no haber hecho lo que se requería en su momento, esto redundando en un cuestionamiento, a veces severo del sistema educativo en vigencia. Es necesario comprender a la educación como un proceso trascendental de la sociedad, ya que es el punto de partida de socialización fuera del núcleo familiar y en el que se incuba a la ciudadanía. Por tanto, concierne un lugar privilegiado para formar a las nuevas generaciones en los valores humanos, que posibilitan el sostén y recreación cultural de una sociedad en específico y puedan prevenir problemáticas como la ocurrencia de eventos suicidas.

Cabe resaltar que el modelo propuesto en esta investigación puede representar una interesante contribución al análisis de datos en el ciberespacio sobre tendencias suicidas en estudiantes y adolescentes, en donde a medida que es expuesto a nuevos sitios web, puede adaptarse de forma independiente, para identificar patrones relacionados con señales de suicidio, formas de llevarlo a cabo, factores de riesgo, formas de prevenirlo, e influencias (vocablos definidos en la ontología semántica). Con lo cual, se pueden establecer fundamentos para la elaboración de protocolos (pro-suicida) en instituciones educativas, que permitan prevenir conductas suicidas mediante información oportuna y sensibilización tanto en los estudiantes como en padres de familia.

Dentro de los resultados, se reportan porcentajes elevados en la detección de rasgos en estudiantes con tendencia suicida en páginas web, asociado a las distintas técnicas de construcción de dataset y Aprendizaje de Máquina empleadas. Los resultados muestran una mejora del tiempo de transformación de los datos no estructurados a datos estructurados empleando técnicas de Cómputo Paralelo con Algoritmos Genéticos, obteniendo un 682% de ahorro de tiempo comparado con el Cómputo Secuencial. Asimismo, se obtiene una media de precisión del 96,28%, alcanzando un valor óptimo del 98% de precisión con el algoritmo RNN. Por tanto, se infiere que una Red Neuronal Recurrente es una arquitectura robusta para tratar análisis de texto, en la que la salida del estado anterior es retroalimentación para preservar la memoria de la red a lo largo del tiempo o secuencia de palabras. Del mismo modo, se concluye que la metodología

y arquitectura propuesta son viables para identificar y clasificar señales suicidas en los estudiantes con información de la web.

En continuidad a los resultados de la investigación, se ha considerado explorar técnicas de análisis descriptivo sobre los rasgos suicidas obtenidos en el estudio con el propósito de establecer el grado de asociación entre los grupos y sus variables. Como trabajo futuro, se plantea explorar el comportamiento suicida en adolescentes basado en textos cortos (tuits) de la red social Twitter con el objetivo de ayudar en la detección de incidentes suicidas, y establecer un modelo de análisis de datos híbrido que combine técnicas de Análisis de Sentimientos, Ontologías Semánticas y Procesamiento de Lenguaje Natural, integrando Redes Neuronales Convolucionales con Aprendizaje Profundo para la clasificación de texto y búsqueda de patrones.

Contribución de Autores

Idea, I.C.Z.; Revisión de literatura (estado del arte), I.C.Z.; Metodología, I.C.Z., F.J.L.R.; Análisis de datos, I.C.Z., F.J.L.R.; Resultados, I.C.Z., J.I.L.V.; Discusión y conclusiones, I.C.Z., J.I.L.V.; Redacción (borrador original), I.C.Z.; Revisiones finales, I.C.Z., F.J.L.R., J.I.L.V.; Diseño del Proyecto y patrocinio, I.C.Z.

Apoyos

Tecnológico Nacional de México, campus Instituto Tecnológico el Llano Aguascalientes.

Referencias

- Anggraini, I.Y., Sucipto, S., & Indriati, R. (2018). Cyberbullying detection modelling at Twitter social networking. *Jurnal Informatika*, 6(2), 113-118. <https://doi.org/10.30595/juita.v6i2.3350>
- Arevalos, D.H. (2020). El sentido de la vida y las prácticas ligadas al suicidio. Testimonio de jóvenes escolarizados. *Revista Latinoamericana de Estudios sobre cuerpos, emociones y sociedad. RELACES*, 32, 52-63. <https://bit.ly/3pXSVC8>
- Beaven-Ciapara, N.I., Campa-Álvarez, R.A., Valenzuela, B.A., & Guillen-Lúgigo, M. (2018). Inclusión educativa: Factores psicosociales asociados a conductas suicidas en adolescentes. *Prisma Social*, 23, 185-207. <https://bit.ly/3GRIO9X>
- Berengueras, M. (2018). *Suicidio la insoportable necesidad de ser otro*. Universidad Autónoma del Estado de Morelos. <https://bit.ly/3F7a8iW>
- Blanco, C. (2019). El suicidio en España, respuesta institucional y social. *Revista de Ciencias Sociales*, 33(46), 79-106. <https://doi.org/10.26489/rvs.v33i46.5>
- Bonami, B., Piazentini, L., & Dala-Possa, A. (2020). Education, Big Data and Artificial Intelligence: Mixed methods in digital platforms. [Educación, Big Data e Inteligencia Artificial: Metodologías mixtas en plataformas digitales]. *Comunicar*, 65, 43-52. <https://doi.org/10.3916/C65-2020-04>
- Carballo-Belloso, J.J., & Gómez-Peñalver, J. (2017). Relación entre bullying, autolesiones, ideación suicida e intentos autolíticos en niños y adolescentes. *Revista de estudios de Juventud*, 115, 207-218. <https://doi.org/10.3916/C65-2020-04>
- Chiroma, F., Liu, H., & Cocea, M. (2018). Text Classification For Suicide Related Tweets. In *International conference on Machine Learning and Cybernetics (ICMLC)* (pp. 587-592). <https://doi.org/10.1109/ICMLC.2018.8527039>
- Denia, E. (2020). The impact of science communication on Twitter: The case of Neil deGrasse Tyson. [El impacto del discurso científico en Twitter: El caso de Neil deGrasse Tyson]. *Comunicar*, 65, 21-30. <https://doi.org/10.3916/C65-2020-02>
- Du, J., Zhang, Y., Luo, J., Jia, Y., Wei, Q., Tao, C., & Xu, H. (2018). Extracting psychiatric stressors for suicide from social media using deep learning. *BMC Medical Informatics & Decision Making*, 18(43), 77-87. <https://doi.org/10.1186/s12911-018-0632-8>
- Durkheim, E. (2008). *El suicidio*. Grupo Editorial Éxodo. <https://bit.ly/3p6C8h7>
- García-Peña, J.J. (2020). *El suicidio: Una mirada integral e integradora*. Universidad Católica Luis Amigó. <https://doi.org/10.21501/9789588943619>
- Gen-Min, L., Szu-Nian, Y., & Yueh-Ming, T. (2020). Machine Learning based suicide ideation prediction for military personnel. *IEEE Journal of Biomedical and Health Informatics*, 24(7), 1907-1916. <https://doi.org/10.1109/JBHI.2020.2988393>
- Healy, M. (2019). Alcanzan máximo histórico los índices de suicidio de adolescentes y adultos jóvenes en EE.UU. *Los Angeles Times*. <https://lat.ms/3IYhebO>
- Hermosillo-De-La-Torre, A.E., Vacío-Muro, M.A., Méndez-Sánchez, C., Palacios-Salas, P., & Sahagún-Padilla, A. (2015). Sintomatología depresiva, desesperanza y recursos psicológicos: una relación con tentativa de suicidio en una muestra de adolescentes mexicanos. *Acta Universitaria*, 25(2), 52-56. <https://doi.org/10.15174/au.2015.900>
- Kim, J., & Chung, K. (2019). Associative feature information extraction using text mining from health big data. *Wireless Pers Commun*, 105, 691-707. <https://doi.org/10.1007/s11277-018-5722-5>
- Landaeta, G. (2014). *Lista de stop words o palabras vacías en español*. SEO para Google. <https://bit.ly/3p2ysg3>
- López-Martínez, L.F. (2020). Suicidio, adolescencia, redes sociales e Internet. *Norte de Salud Mental*, 17, 25-36. <https://bit.ly/3sg25g2>

- Luna, M., & Dávila, A. (2018). Adolescentes en riesgo: factores asociados con el intento de suicidio en México. *Revista Gerencia y Política de Salud*, 17(34), 1-14. <https://doi.org/10.11144/Javeriana.rgsp17-34.arfa>
- Marchiori, H. (2015). *El suicidio enfoque criminológico*. Editorial Porrúa. <https://bit.ly/3sfMYDu>
- Molina, M.J., & Restrepo, D. (2018). Internet y comportamiento suicida en adolescentes: ¿Cuál es la conexión? *Revista Pediatría*, 51, 30-39. <https://doi.org/10.14295/pediatr.v51i2.109>
- Moreno, P., & Blanco, C. (2012). Suicidio e Internet. Medidas preventivas y de actuación. *Revista Psiquiatría.com*, 16(18). <https://bit.ly/3E110Lb>
- Mosquera, L. (2016). Conducta suicida en la infancia: Una revisión crítica. *Revista de Psicología Clínica con Niños y Adolescentes*, 3, 9-18. <https://bit.ly/3p5kG8>
- Nalini, K., & Sheela, L. (2014). A survey on datamining in cyber bullying. *International Journal on Recent and Innovation Trends in Computing and Communication*, 2(7). <https://bit.ly/3q7GHqt>
- Olivares, S. (2019). Uso de Internet y conductas suicidas en adolescentes de 14 a 18 años en México. *Visión Criminológica-Criminalística*. (pp. 6-21). <https://bit.ly/3p7uKSF>
- Organización Mundial de la Salud (Ed.) (2019). *Suicidio. Información obtenida el 6 de abril de 2021 en la dirección de Internet*. OMS. <https://bit.ly/3q8TkBG>
- Pérez-Martínez, V.M., Aparicio-Vinacua, B., & Rodríguez-González, M.D. (2020). Acoso escolar, violación y suicidio en Twitter: Segunda temporada de «Por trece razones». *Vivat Academia*, 153, 137-168. <https://doi.org/10.15178/va.2020.153.137-168>
- Porter, M.F. (2006). An Algorithm for suffix stripping. *Program: Electronic Library and Information Systems*, 40, 211-218. <https://doi.org/10.1108/00330330610681286>
- Ramírez-López, C.M., Montes, M., Ochoa-Zezzatti, A., Ponce-Gallegos, J.C., & Guzmán-Mendoza, J.E. (2021). Identification of possible suicide cases using a Bayesian Classifier with the database the Emergency Service 911 of Aguascalientes. *International Journal of Combinatorial Optimization Problems and Informatics*, 12(1), 43-57. <https://bit.ly/32gamVvt>
- Rocamora, A. (2017). *Cuando nada tiene sentido: Reflexiones sobre el suicidio desde la logoterapia*. Editorial Desclée de Brouwer. <https://bit.ly/3F9TxLa>
- Roy, S.S., Mallik, A., Gulati, R., Obaidat, M.S., & Krishna, P.V. (2017). A deep learning based artificial neural network approach for intrusion detection. In D. Giri, R. Mohapatra, H. Begehr, & M. Obaidat (Eds.), *Mathematics and Computing. ICMC 2017. Communications in Computer and Information Science* (pp. 44-53). Springer. https://doi.org/10.1007/978-981-10-4642-1_5
- Sánchez-García, M.A., Pérez-De-Albéniz, A., Paño, M., & Fonseca, P. (2018). Ajuste emocional y comportamental en una muestra de adolescentes españoles. *Actas españolas de psiquiatría*, 46, 205-216. <https://bit.ly/3yAOv8a>
- SeGob (Ed.) (2021). *Impacto de la pandemia en niñas y niños*. Secretaría de Gobernación de México. <https://bit.ly/3J1bsGg>
- Urra, J. (2019). *La huella de la desesperanza: Estrategias de prevención y afrontamiento del suicidio*. Ediciones Morata. <https://bit.ly/30H1Qiq>
- Villardón-Gallego, L. (2013). *El pensamiento de suicidio en la adolescencia*. Publicaciones de la Universidad de Deusto. <https://bit.ly/33FwFV>