

Algoritmos de aprendizaje profundo en un corpus digital con expresiones depresivas

César Jesús Núñez-Prado, Aline Anahi Juárez-Maldonado,
Liliana Chanona-Hernández, Grigori Sidorov

Resumen—La depresión es uno de los principales trastornos mentales que aquejan a la población mundial y cuando no se detecta a tiempo puede tener consecuencias letales. Las redes sociales ofrecen un espacio donde las personas con este trastorno se sienten libres de expresarse. En este contexto, se emplearon técnicas de procesamiento de lenguaje natural y algoritmos de aprendizaje profundo para detectar los patrones e identificar el discurso depresivo sobre un corpus en español de mensajes extraídos de la red social X (antes Twitter) previamente etiquetado, al cual se le aplicaron técnicas de preprocesamiento y fue analizado a través de 2 algoritmos de aprendizaje profundo; una Red Neuronal Recurrente y el modelo BERT (Bidirectional Encoder Representations from Transformers). Los resultados mostraron una exactitud en las predicciones del 79,19% y 81,51% en la clasificación de textos, respectivamente.

Palabras clave—Depresión, procesamiento de lenguaje natural, aprendizaje profundo, red neuronal recurrente, BERT.

Deep Learning Algorithms in a Digital Corpus with Depressive Expressions

Abstract—Depression is one of the main mental disorders afflicting the world's population and when is not detected in time can have lethal consequences. Social networks offer a space where people with this disorder feel free to express themselves. In this context, natural language processing techniques and deep learning algorithms were used to detect patterns and identify depressive speech on a corpora in Spanish of messages extracted from the social network X (formerly Twitter) previously labeled, to which preprocessing techniques were applied and analyzed through 2 deep learning algorithms; a Recurrent Neural Network and the Bidirectional Encoder Representations from Transformers model. The results showed a prediction accuracy of 79.19% y 81.51% for text classification, respectively.

Index Terms—Depression, natural language processing, deep learning, recurrent neural network, BERT.

I. INTRODUCCIÓN

De acuerdo con la Organización Mundial de la Salud (OMS), la depresión es un trastorno común de la salud mental en donde se estima que al menos el 5% de la población adulta en el mundo la padecen. Este tipo de trastornos están ligados con

síntomas comunes, tales como la magnificación de la percepción del dolor, problemas cardiacos y principalmente la pérdida de interés en algunas actividades cotidianas, según se informa en [1]. La Organización Panamericana de las Salud (OPS) señala en [2] que casi 300 millones de personas en el mundo sufren depresión y algunas de las causas están asociadas con la pobreza, el desempleo, acontecimientos vitales (pérdida de un ser querido, ruptura amorosa, afección física o enfermedades en fase terminal) o por consumo de drogas. De acuerdo con su reporte, las mujeres entre los 15 y 80 años son más propensas que los hombres a padecer depresión. Entre las consecuencias de no recibir un tratamiento adecuado se encuentran: la posibilidad de la adicción a sustancias lícitas e ilícitas, comportamiento temerario, problemas de salud, problemas para socializar y autolesionarse.

Las autolesiones pueden empezar desde hacer pequeños cortes en las muñecas, quemaduras intencionales (generalmente con cigarrillos), dejar de comer, producirse el vómito o dejar de hidratarse, pero puede escalar hasta que los individuos que padecen este trastorno pueden colocarse en situaciones de extremo riesgo y culminar con el intento de suicidio o llevar a cabo el suicidio.

Según el informe del Instituto Nacional de Estadística y Geografía (INEGI) [3], en México durante el año 2022 se registraron 819,448 defunciones, de las cuales el 10% se relacionaron con causas externas (accidentes, homicidios y suicidios). Como presunto suicidio se reconocieron 8,237 casos, esta cifra representa el 9,82% de las defunciones contabilizadas por causas externas y es aproximadamente el 10% del total de defunciones registradas.

En ocasiones, aquellos afectados por trastornos mentales encuentran refugio en las redes sociales para expresar sus sentimientos desde un lugar seguro, donde se sienten resguardados de juicios. Este fenómeno ha motivado investigaciones científicas con la finalidad de detectar a tiempo el posible discurso depresivo o suicida con la intención de reducir el número de casos letales que se presentan. Para llevar a cabo dichas investigaciones es necesario contar con bases de datos que sean capaces de reflejar de manera clara los comentarios que denotan el discurso de depresión, después, a estas bases de datos se les puede aplicar métodos de procesamiento de lenguaje natural con diferentes algoritmos de inteligencia artificial buscando identificar patrones en los diferentes textos.

En esta investigación, nuestro objetivo es aplicar dos algoritmos de aprendizaje profundo, específicamente una Red

Manuscript received on 11/05/2024, accepted for publication on 14/07/2024. C.J. Núñez Prado and Grigori Sidorov are with the Instituto Politécnico Nacional (IPN), Centro de Investigación en Computación (CIC), México (cnunezp@ipn.mx, sidorov@cic.ipn.mx).

A.A. Juárez-Maldonado, L. Chanona Hernández are with the Instituto Politécnico Nacional (IPN), Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME), México ({alineajuarezmalodon, lchanona}@gmail.com).

Neuronal Recurrente y BERT con el propósito de clasificar textos en dos categorías: aquellos con contenido depresivo y aquellos sin contenido depresivo, implementando técnicas de preprocesamiento y procesamiento de lenguaje natural con la finalidad de detectar el discurso depresivo en un corpus generado a partir de publicaciones en la red social X.

La estructura de esta investigación se expone de la siguiente manera: en la sección 2 se presentan las investigaciones actuales y a fin a nuestro tema central, en la sección 3 se describe el banco de datos utilizado, en la sección 4 se muestra la metodología y los modelos aplicados, en la sección 5 se presentan los resultados y finalmente, en la sección 6 se exponen las conclusiones.

II. TRABAJOS RELACIONADOS

La aplicación de algoritmos de inteligencia artificial y el procesamiento de lenguaje natural sobre textos con la intención de encontrar patrones es ampliamente usada desde los últimos años. En [4] realizan el análisis de textos extraídos de tecnologías móviles de pacientes con enfermedades mentales. A los textos aplicaron tokenización, eliminación de palabras vacías y extracción de raíces. Vectorizaron los datos utilizando Word2Vec y realizaron el estudio de las frecuencias de las palabras y relaciones entre las mismas palabras. Los resultados de esta investigación proporcionan un conjunto de palabras que están fuertemente relacionadas con una frecuencia mayor y son específicas de la enfermedad mental referida por cada paciente.

La identificación del discurso de depresión en redes sociales se aborda en [5], donde se aplicaron diferentes algoritmos de aprendizaje automático sobre un corpus con expresiones depresivas. Sobre los textos realizaron un preprocesamiento, el cual incluyó la tokenización, lematización, etiquetado de las palabras y eliminación de palabras vacías. En cuanto a los modelos de aprendizaje automático aplicaron los algoritmos de similitud coseno, k vecinos más cercanos y perceptrón multicapa y probaron estos algoritmos haciendo divisiones del banco de datos en conjuntos de entrenamiento y prueba considerando las métricas 90-10, 80-20 y 70-30 respectivamente. El mejor resultado que obtuvieron fue con el modelo perceptrón multicapa con una exactitud de 84,22%, con la división en conjuntos de entrenamiento y prueba 90-10, la función de activación Relu y el optimizador SGD (Stochastic Gradient Descent - Descenso de Gradiente Estocástico).

En [6] analizan textos escritos por personas que experimentan la depresión y buscan identificar los aspectos lingüísticos, sintácticos y semánticos. Los textos analizados fueron extraídos de la plataforma social Reddit en el idioma inglés y cuentan con dos etiquetas: depresivo y no depresivo. Los métodos aplicados en esta investigación fueron Naive Bayes multinomial, regresión logística, máquina de vectores de soporte y árboles aleatorios, utilizando la bolsa de palabras y vectores incrustados como representación numérica de los textos.

Una de las tareas abordadas en el procesamiento de lenguaje natural es la identificación de palabras claves como parte de los patrones en la escritura. En [7] utilizan un algoritmo de frecuencia ponderada para encontrar palabras clave en el campo de sospecha diagnóstica en la lista de espera de consulta de especialidades en Chile. Para ello utilizaron el algoritmo de

procesamiento de lenguaje natural TF-IDF (Term Frequency - Inverse Document Frequency) a través del cual pueden recuperar palabras clave en datos no estructurados de textos clínicos. Dentro de sus resultados crearon nubes de palabras (representación visual) para 4 especialidades con la frecuencia buscada más alta.

En [8] utilizan como modelo de clasificación a BERT para detectar la depresión en Tweets, el conjunto de datos con el que trabajaron fue etiquetado gramaticalmente utilizando la herramienta Freeling Natural Language Processing y además realizaron el etiquetado manual para confirmar las etiquetas asignadas por la herramienta. El conjunto de datos se clasificó manualmente como texto depresivo o texto no depresivo y fue tokenizado por BERT para realizar la clasificación, obteniendo una precisión en las predicciones del 88,06%.

III. BANCO DE DATOS

El conjunto de datos utilizado en esta investigación fue desarrollado y proporcionado por [9] el cual consta de 2,955 mensajes descargados de la red social X (antes Twitter) y debido a las restricciones en cuanto a las publicaciones en esta red social, el tamaño máximo de estos mensajes no excede los 250 caracteres. Cada uno de los mensajes fueron etiquetados manualmente por 5 evaluadoras con el siguiente perfil:

- Profesionistas en Psicología.
- Experiencia actual y mínima de 5 años ininterrumpidos.
- Atención directa a pacientes con depresión o tendencias suicidas.
- Manejo y conocimiento de redes sociales.

La tarea de las evaluadoras consistió en leer cada uno de los mensajes descargados y colocar la etiqueta 1 (uno) si ellas consideraban que el mensaje contenía relación con depresión o tendencia suicida o la etiqueta 0 (cero) en caso contrario, de acuerdo con la experiencia adquirida en su campo laboral. En la Tabla 1 aparece una muestra del conjunto de datos.

Una vez que el conjunto de datos fue evaluado en su totalidad por las 5 psicólogas, se realizó un conteo de las etiquetas colocadas en cada mensaje. Si el mensaje fue etiquetado 3 o más veces con la misma etiqueta, esa etiqueta se asignó como etiqueta final del mensaje.

Las estadísticas finales del corpus se muestran en la Tabla 2. El corpus final cuenta con 2,955 textos con solo dos etiquetas y se encuentra en formato de texto plano.

IV. METODOLOGÍA

El preprocesamiento del texto tiene como objetivo el homogeneizar, limpiar el corpus y reducir el tiempo de procesamiento, por lo cual; se realizó sobre cada mensaje de texto:

- Tokenización: proceso por el cual se realiza la separación en unidades mínimas las cuales incluyen división en párrafos, oraciones, palabras o caracteres. Para esta investigación se escogió como unidad mínima a cada palabra, es decir; unigramas.
- Conversión a minúsculas: para las computadoras la palabra “Hola” y “hola” son consideradas como dos cadenas de

TABLA I
MUESTRA DEL CONJUNTO DE DATOS

# Mensaje	Texto	Clase
1	Que triste es tu vida que estás echándole mierda a nuestro jugador por llevarse bien con su compañero de selección, qué idiotas son.	0
2	Es triste sentirse un idiota por hacer las cosas bien. Nunca aprendo y siempre salgo llorando.	1

TABLA II
DISTRIBUCIÓN DE LAS CLASES EN EL CONJUNTO DE DATOS

Mensajes	Clase
2,421	0 (cero)
534	1 (uno)

caracteres diferentes, por lo que se realiza esta conversión con la finalidad de normalizar el texto.

- Eliminación de valores numéricos: los datos numéricos como fechas o cantidades no son tan relevantes en el análisis semántico de las palabras.
- Eliminación de palabras vacías (conocidas como stop-words en inglés): es un conjunto de palabras que incluyen a los artículos determinados, artículos indeterminados, conjunciones y preposiciones que no aportan contenido semántico a los textos.
- Eliminación de emoticones: estos elementos son representaciones gráficas asociadas con una acción o sentimiento. El presente trabajo se centra únicamente en el análisis del texto por lo cual se decidió prescindir de estos elementos.
- Eliminación de signos de puntuación: al no aportar contenido semántico al análisis, se decidió eliminar cualquier signo de puntuación contenido en los mensajes a procesar.
- Eliminación de enlaces a internet: cuando un mensaje es publicado con una imagen o con un enlace a alguna página web, este aparece en forma de texto y también se decidió realizar la eliminación de cualquier enlace.
- Lematización: proceso mediante el cual se encuentra a la palabra sin flexionar, por ejemplo; el lema del verbo “corriendo” es el verbo en su forma infinitiva, es decir; “correr”.
- Una vez que se realizó el preprocesamiento de los datos, se eliminó cualquier mensaje repetido.

Una fase que antecede a los modelos de clasificación en el paradigma supervisado es la división del conjunto de datos en dos subconjuntos, por una parte, se requiere del conjunto de datos para entrenar el modelo y en nuestro caso, escogimos asignar el 80% de los datos del corpus para esta tarea y el conjunto de datos de prueba con el 20% restante.

A. Red Neuronal Recurrente

La elección de implementar una Red Neuronal Recurrente (RNN Recurrent Neuronal Network) en este estudio se fundamenta en su capacidad única para modelar secuencias de datos, lo cual es esencial al trabajar con datos textuales. Las RNNs son particularmente efectivas para capturar dependencias temporales y contextuales en el lenguaje, permitiendo un procesamiento más profundo y comprensivo de la información en comparación con modelos estáticos.

Este tipo de redes no son capaces de procesar palabras, por lo cual; es necesario encontrar la representación de las palabras en un espacio vectorial [10] y a esto se le conoce como vector incrustado (*word embedding*). La generación de este vector se lleva a cabo a partir de los mensajes del corpus, encontrando primero el vocabulario único que incluya todas las palabras únicas del conjunto de datos.

Una vez que se cuenta con dichos vectores, estos se toman como secuencia de entrada para la Red Neuronal Recurrente y son procesados de manera secuencial tomando la entrada actual más la información oculta (*hidden state*) de la iteración anterior para calcular la salida. La información oculta simula una memoria que captura información de las iteraciones anteriores e influye en las predicciones actuales. Otros parámetros importantes dentro de este tipo de redes son las funciones de pérdida y optimización, las cuales miden la discrepancia de los resultados y ajustan los pesos de la red, respectivamente.

Para el manejo de vectores como matrices utilizamos la biblioteca *Numpy* de *Python* y para la división en conjunto de entrenamiento y prueba se utilizó *Scikit Learn*.

El modelo de RNN aplicado utiliza la interfaz de modelado secuencial de *Keras*, conocida como *Sequential*, y se configura con una capa LSTM. La configuración de la RNN fue la siguiente:

Tokenización y preprocesamiento:

- Número máximo de palabras = 10,000.
- Longitud máxima de la secuencia del texto = 100.
- Función de tokenización de *Keras*.
- Función *padding* para rellenar con ceros a la derecha los vectores que no tuvieran la misma dimensión.

Arquitectura:

- Capa de entrada con dimensiones iguales al número máximo de palabras.
- Capa de salida = 128 dimensiones.
- 64 unidades LSTM (*Long Short Term Memory*).
- Función de activación = *sigmoid*.
- Configuración en entrenamiento:
- Función de optimización = *Adam*.
- Función de pérdida = *binary crossentropy*.
- Número de épocas = 10.
- Monitoreo de la pérdida para la detención temprana mediante el *callback EarlyStopping*.

Esta configuración se escogió cuidadosamente para abordar las complejidades de la tarea de clasificación de textos,

TABLA III
RESULTADOS RNN

Época	Pérdida-Entrena	Exactitud-Entrena	Pérdida-Prueba	Exactitud-Prueba
1	0.5135	0.8071	0.4553	0.8270
2	0.4099	0.8217	0.4461	0.7975
3	0.2475	0.8918	0.4863	0.8143
4	0.1545	0.9426	0.5971	0.7890
5	0.1012	0.9628	0.6547	0.7806
6	0.4470	0.7919	-	-

TABLA IV
RESULTADOS BERT

Época	Pérdida-Entrena	Exactitud-Entrena	Pérdida-Prueba	Exactitud-Prueba
1	0.5729	0.8147	0.5299	0.827
2	0.5342	0.8151	0.5020	0.827
3	0.5143	0.8151	0.4854	0.827
4	0.5008	0.8151	0.4757	0.827
5	0.4931	0.8151	0.4704	0.827
6	0.4895	0.8151	0.4676	0.827
7	0.4868	0.8151	0.4658	0.827
8	0.4839	0.8151	0.4648	0.827
9	0.4830	0.8151	0.4644	0.827
10	0.4824	0.8151	0.4639	0.827

optimizando la red para aprender patrones temporales y contextuales en los datos textuales proporcionados por el corpus.

B. BERT

La elección de implementar BERT (Bidirectional Encoder Representations from Transformers) en este estudio se basa en su capacidad para capturar relaciones semánticas y contextuales en datos textuales. A diferencia de las RNNs tradicionales, BERT aborda la limitación de procesar palabras de manera secuencial al aprovechar la bidireccionalidad y la atención global. BERT es un modelo de lenguaje preentrenado que utiliza la arquitectura de Transformers, permitiéndole capturar dependencias complejas y relaciones entre palabras en ambos sentidos. Esto proporciona representaciones contextualizadas más ricas y mejoradas en comparación con los enfoques tradicionales. La arquitectura de BERT se ha destacado en tareas diversas, como la comprensión del lenguaje natural, la traducción automática y la extracción de información, convirtiéndolo en una elección sólida para nuestro estudio. La configuración de los hiperparámetros en BERT fue la siguiente:

Tokenización y preprocesamiento:

- División del conjunto de datos en entrenamiento (80%) y prueba (20%).
- Tokenizador de BERT ('bert-base-uncased') de la biblioteca Transformers.
- Padding para asegurar que todas las secuencias tengan la misma longitud.

Arquitectura del Modelo BERT:

- Capa de entrada para las secuencias tokenizadas.
- Capa BERT preentrenada ('bert-base-uncased') como capa de TensorFlow.
- Capa de pooling global para reducir la dimensionalidad.
- Capa densa con función de activación sigmoide para la clasificación binaria.

Configuración de entranamiento:

- Función de optimización: pooling global.
- Función de pérdida: entropía cruzada binaria.
- Callback de EarlyStopping para detener el entrenamiento si la pérdida en el conjunto de validación deja de mejorar después de un cierto número de épocas. Esta elección se hizo para evitar el sobreajuste y mejorar la eficiencia del entrenamiento.

Esta configuración se eligió para optimizar el rendimiento de BERT en la tarea de clasificación de textos.

V.RESULTADOS

El modelo RNN fue configurado inicialmente para un máximo de 10 épocas, pero gracias al uso del mecanismo de detención temprana, solo fueron necesarias 6 épocas para lograr una convergencia efectiva. Este enfoque no solo mejoró la eficiencia computacional, sino que también mejoró la eficiencia general del entrenamiento.

A continuación, se presentan los resultados detallados de cada época en la Tabla 3, los valores de las columnas de Pérdida-Entrena y Exactitud-Entrena se refieren a los valores obtenidos en el entrenamiento, mientras que las columnas de Pérdida-Prueba y Exactitud-Prueba se refieren a los valores obtenidos con el conjunto de prueba.

La pérdida disminuyó a lo largo de las primeras cinco épocas, indicando una mejora en la capacidad del modelo para generalizar patrones en el conjunto de entrenamiento. Sin embargo, en la sexta época, se observa un aumento en la pérdida, lo que puede indicar una posible sobreajuste. La exactitud, por otro lado, alcanzó un nivel alto, llegando a 96,28% en la quinta época.

En el caso de BERT, el modelo se entrenó durante 10 épocas, y los resultados detallados se presentan en la Tabla 4, los valores de Pérdida E y Exactitud E se refieren a los valores obtenidos en el entrenamiento, mientras que Pérdida V y Exactitud V se refieren a los valores obtenidos durante la validación.

Se puede observar que la pérdida y la exactitud en el entrenamiento se estabilizan después de la quinta época, indicando que el modelo ha alcanzado su convergencia. La exactitud se mantiene constante en un nivel del 81.51% desde la segunda hasta la décima época.

VI. CONCLUSIONES

Se aplicaron con éxito los modelos de Red Neuronal Recurrente y BERT sobre el banco de datos con expresiones depresivas. La exactitud alcanzada por la RNN fue de 96,28% y 78,06% en el entrenamiento y prueba respectivamente. Por

parte de BERT, la exactitud obtenida fue de 81,51% y 82,7% en entrenamiento y prueba respectivamente.

La RNN demostró ser eficaz en la clasificación de textos, logrando una alta exactitud después de un entrenamiento de tan sólo 6 épocas con la ayuda del mecanismo de detención temprana. Sin embargo, se observó un aumento en la pérdida en la sexta época, resaltando la relevancia de técnicas como la detención temprana para mejorar la generalización del modelo.

BERT, con su capacidad para capturar relaciones semánticas complejas, exhibió una estabilidad y convergencia notables. La exactitud se mantuvo constante después de la segunda época, indicando una rápida convergencia. La eficiencia computacional de BERT se tradujo en un tiempo de entrenamiento más prolongado en comparación con la RNN, pero los resultados estabilizados sugieren una robustez mayor en el modelo.

Al explorar la aplicación de la inteligencia artificial para abordar problemas sociales, se puede contribuir significativamente a la reducción de consecuencias letales en pacientes con este tipo de enfermedades.

AGRADECIMIENTO

El trabajo fue realizado con apoyo parcial del Gobierno Mexicano a través de la beca A1-S-47854 del CONACYT, México, las becas 20241816, 20241819, y 20240951 de la Secretaría de Investigación y Posgrado del Instituto Politécnico Nacional, México. Los autores agradecen al CONACYT por los recursos de cómputo puestos a su disposición a través de la Plataforma de Aprendizaje Profundo para Tecnologías del Lenguaje del Laboratorio de Supercómputo del INAOE, México y reconocen el apoyo de Microsoft a través del Premio de Doctorado Microsoft Latinoamérica.

REFERENCIAS

- [1]. S. Aguilar-Navarro, J.A. Avila-Funesa, “La depresión: particularidades clínicas y consecuencias en el adulto mayor,” *Gac Med Mex*, vol. 143, no. 2, pp. 141–148, 2007.
- [2]. Organización Panamericana de la Salud, “Depresión y otros trastornos mentales comunes,” *Estimaciones sanitarias mundiales*, Washington, D.C., 2017.
- [3]. Instituto Nacional de Estadística y Geografía, *Estadísticas de defunciones registradas (EDR) 2022 (Preliminar)*, Comunicado de prensa, vol. 419, no. 23, 2023.
- [4]. J.V. Sancho Escrivá, C. Fanjul-Peyró, M. de la Iglesia Vayá, J.A. Montell, and M.J. EscartíFabra, “Aplicación de la inteligencia artificial con procesamiento del lenguaje natural para textos de investigación cualitativa en la relación médico-paciente con enfermedad mental mediante el uso de tecnologías móviles,” *Revista de Comunicación y Salud*, vol. 10, no. 1, pp. 19–41 (2020).
- [5]. C.J. Núñez-Prado, C. Talavera-Ortega, L. Chanona-Hernández, and G. Sidorov, “Aplicación de algoritmos de aprendizaje automático sobre un corpus depresivo digital,” *Research in Computing Science*, vol. 152, no. 8, pp. 89–98 (2023).
- [6]. M.J. Garcíarena-Ucelay, L.C. Cagnina and M.L. Errecalde, “Análisis de rasgos lingüísticos con técnicas de procesamiento de lenguaje natural en la detección temprana de depresión,” *Anales de lingüística*. vol. 7, pp. 89–116, 2021.
- [7]. F. Villena and J. Dunstan, “Obtención automática de palabras clave en textos clínicos: Una aplicación de procesamiento del lenguaje natural a datos masivos de sospecha diagnóstica en Chile,” *Revista Médica de Chile*, vol. 147, no. 10, 2019.
- [8]. G.J. Aleman-Zambrano, M.I. Del Carpio-Lazo, D.G. Mendiguri-Chávez, D.C. VilchezSilva, and F. Tejada-Toledo, “Modelo de clasificación de depresión en Tweets usando BERT,” *Revista Innovación y Software*, vol. 4, no. 2, pp. 6–24, 2023. DOI: 10.48168/ innosoft.s12.a89.
- [9]. C.J. Núñez-Prado, C. Talavera-Ortega, L. Chanona-Hernández, and G. Sidorov, “Generación de un corpus lingüístico digital en español enfocado a la depresión,” *Research in Computing Science*, vol. 152, no. 7, pp. 5–12, 2023.
- [10]. G. Sidorov, *Construcción no lineal de n-gramas en la lingüística computacional. N-gramas sintácticos, filtrados y generalizados*, Sociedad Mexicana de Inteligencia Artificial, 2013.
- [11]. J. Devlin, M.W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training on DeepBidirectional Transformers for Language Understanding,” in *CoRR*, 2018.