



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

Application of Artificial Intelligence with Natural Language Processing for qualitative research texts in the medical-patient relationship with mental illness through the use of mobile technologies

José Vicente Sancho Escrivá.¹

Universitat Jaume I. España.

al400282@uji.es

Carlos Fanjul Peyró.

Universitat Jaume I. España.

María de la Iglesia Vayá.

Unidad Mixta de Imagen Biomédica FISABIO-CIPF. España.

Joaquín A. Montell.

Centro de Investigación Príncipe Felipe. Unidad Mixta IB. España.

María José Escartí Fabra.

Hospital Clínico Valencia, CIBERSAM Valencia. España.

Resumen

La Inteligencia Artificial (IA) sigue posicionándose en la sociedad como referencia del progreso tecnológico. Dentro de este campo, el Procesamiento de Lenguaje Natural (PLN) alcanza gran aceptación en disciplinas que trabajen con altos volúmenes de datos (Big Data). En este marco queremos ver qué aportan estos algoritmos, pero aplicado a la comunicación en el campo de la salud mental. Establecemos esta metodología con PLN partiendo de observaciones cualitativas previas en textos

¹¹ **Autor para correspondencia:** José Vicente Sancho Escrivá al400282@uji.es: Licenciado en Ciencias de la información, publicidad y RRPP. Profesional de la comunicación digital especialista en tecnología mobile y docente en diferentes universidades españolas en materias relacionadas con la publicidad y la comunicación digital.

transcritos de grupos focales realizados a pacientes con enfermedad mental con el objetivo de entender si la aplicación de esta metodología aporta mejora al análisis de los datos como se ha demostrado en investigaciones previas, pero aplicado novedosamente al campo de la salud mental. Para ello se han ejecutado scripts basados en código Python y se han depurado los textos, clasificando las cadenas de palabras en entidades denominadas tokens y eliminando las palabras vacías. Posteriormente, se ha analizado la frecuencia de palabras y la conexión de frases, obteniendo un conjunto de estructuras donde aplicar técnicas de Machine Learning mediante Word2vec y generando vectores sobre los datos quedando representados con gráficas n-dimensionales en donde se configura un nuevo vocabulario con palabras agrupadas por cercanía. Aplicamos un método que sin el aprendizaje algorítmico se nos escapa en el análisis previo de una investigación cualitativa. Se identifican en el análisis los principales temas encontrados con el análisis cualitativo tradicional, mecanizando el proceso y facilitándolo. Se demuestra además que esta metodología es aplicable en la salud mental como en otros grupos de población.

Palabras clave: inteligencia artificial, procesamiento del lenguaje natural, machine learning, comunicación, ciencias sociales, mHealth, salud mental.

Abstract

Artificial Intelligence (AI) continues to position itself in society as a benchmark for technological progress. Within this field, Natural Language Processing (NLP) reaches great acceptance in disciplines that work with high volumes of data (Big Data). In this framework we want to see what do these algorithms contribute with, but applied to communication in the field of mental health. We establish this methodology with NLP based on previous qualitative observations in transcribed texts of focus groups. These texts were obtained from focus groups carried out on patients with mental illnesses in order to understand whether the application of this methodology contributes to any improvement on the analysis of data, which has been shown in previous researches. However, this research has been applied in a novel way in the field of mental health. To do this, scripts based on Python code have been executed and the texts have been purified, classifying the word strings into entities called tokens and eliminating stopwords. Subsequently, the frequency of words and the connection of sentences have been analyzed, obtaining a set of structures in which to apply Machine Learning techniques using word2vec and generating vectors on the data, which are represented with n-dimensional graphics where a new vocabulary based on proximity words is created. We are applying a method that without algorithmic learning we would be unable to obtain this type of information in the previous analysis of qualitative research. The main themes found with traditional qualitative analysis are identified in the analysis, mechanizing the process and facilitating it. It is also shown that this methodology is applicable in mental health as in other population groups.

Keywords: artificial intelligence, natural language processing, machine learning, communication, social science, mHealth, mental health.

Cómo citar el artículo

Sancho Escrivá, J. V., Fanjul Peyró, C., de la Iglesia Vayá, M, Montell, J. A. y . Escartí Fabra, M. J. (2020). 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*, 10(1), 19-41. doi: [http://doi.org/10.35669/rcys.2020.10\(1\).19-41](http://doi.org/10.35669/rcys.2020.10(1).19-41)

1. INTRODUCCIÓN

La Inteligencia Artificial (IA) sigue posicionándose en todos los ámbitos de la sociedad como referencia del progreso tecnológico. Así queda reflejado en el creciente número de publicaciones dentro de este ámbito en los últimos años (Perrault et al., 2019). Esta disciplina que consiste en la capacidad que tienen las máquinas y los algoritmos para replicar cómo piensa y actúa un ser humano (Aghion et al., 2017) y que parece estar muy lejana a nosotros, en realidad, está más integrada que nunca en nuestras vidas. Un sencillo gesto y tan habitual como puede ser una búsqueda en Google para recoger una información, está basado en la aplicación de la IA que, mediante algoritmos de diversa índole, nos muestra una ingente cantidad de datos mediante enlaces de información basado en las palabras clave de nuestra búsqueda y otras variables como la geolocalización.

En este sentido, hay que decir que la IA como referente de innovación, por sus características, es una oportunidad en el procesamiento de los datos y la aplicación de algoritmos matemáticos sobre textos y palabras aplicable a las metodologías de investigación cualitativa ante la recogida y gestión de un elevado volumen de información y datos (Big Data).

La IA tiene aplicaciones de alto interés en temas de salud y se aplica en muchas áreas biomédicas. Se puede observar que la IA desempeña un papel cada vez más importante en la biomedicina, no sólo por el progreso continuo de la IA en sí, sino también por la compleja naturaleza innata de los problemas biomédicos y la idoneidad de la IA para resolver tales problemas. Las nuevas capacidades de IA proporcionan soluciones novedosas para la biomedicina, y el desarrollo de biomedicina exige nuevos niveles de capacidad de la IA (Rong et al., 2020). Las tecnologías de IA pueden realizar una amplia gama de funciones, tales como ayudar en la orientación diagnóstica y la selección de terapia, hacer predicciones de riesgo y estratificando enfermedades, reduciendo errores médicos y mejorando la productividad (He et al., 2020).

Respecto a la salud mental, las posibles aplicaciones de la IA en la psiquiatría se pueden agrupar en dos amplias categorías. Una categoría se centra en el Procesamiento de Lenguaje Natural (PLN), que permite al mundo de los dispositivos

informáticos comprender, interpretar y manipular el lenguaje humano. La otra categoría es la que se centra en los chatbots, los cuales son agentes de conversación digitales que utilizan métodos de IA a través de texto y/o voz para imitar el comportamiento humano a través de un diálogo en evolución. A los chatbots se les considera un medio para proporcionar atención de salud mental en regiones con bajo acceso a la atención médica o a personas que tienen dificultades para revelar sus sentimientos a un ser humano. Los chatbots han demostrado ser eficaces para reducir los síntomas de la depresión y la ansiedad (Brunn et al., 2020).

Dentro de este amplio campo tecnológico que comprende la IA, la categoría que atiende el PLN tiene cada vez mayor aceptación en disciplinas que trabajen con altos volúmenes de datos, incluyendo, entre ellas, el sector de la salud. Este conjunto de técnicas que comprende el PLN consiste en analizar y representar textos naturales mediante software y algoritmos en uno o diferentes niveles de análisis lingüístico con la finalidad de obtener una apariencia humana en el procesamiento de lenguaje para tareas concretas (Liddy, 2001). En definitiva, las técnicas de PLN consisten en la aplicación de IA para el análisis de datos de comportamiento, las cuales se desarrollan mediante el aprendizaje automático integrado o técnicas de Embedded Machine Learning tras la recolección de los datos (Rong et al., 2020).

Una ventaja que puede tener la aplicación del PLN es que el investigador no tiene que interpretar los textos y son los algoritmos los que mediante aprendizaje con los datos generan resultados. Aunque, por el contrario, la limitación inicial puede ser la falta de comprensión natural del lenguaje por parte del software.

Las técnicas de IA han ido evolucionando en el ámbito del análisis del discurso, hasta que hoy en día se investiga con este tipo de metodologías y herramientas con aplicaciones en el mundo real. Actualmente se trabaja, entre otros ámbitos, en obtener información de salud a partir de la recolección de datos o identificando sentimientos o emociones. En los últimos años, se ha pasado de utilizar métodos más sencillos de análisis de palabras sin identificar la estructura de la oración y el significado del discurso a mejores sistemas que aplican aprendizaje automático a partir de aplicación de software más avanzados o machine learning. Estos avances en IA permiten una mejor comprensión del lenguaje con herramientas y métodos de alto rendimiento que permiten analizar el discurso a partir de los datos, identificando la sintaxis, información semántica y el contexto del propio discurso (Hirschberg & Manning, 2015).

En este marco queremos ver y comprender qué aporta este tipo de algoritmos aplicándolo al ámbito de la comunicación. La comunicación queda circunscrita dentro de las ciencias sociales, en donde se trabaja habitualmente con metodologías cualitativas, algo más cercano a la palabra y a lo descriptivo que las metodologías cuantitativas, más focalizadas en los datos numéricos y lo cuantificable (Taylor & Bogdan, 1987). Se trata de una diferenciación enmarcada entre la subjetividad y la objetividad de los propios investigadores a la hora de enfrentarse a la selección del

método de trabajo. Esta diferenciación ha sido motivo de muchos debates científicos cuestionando a la investigación cualitativa en las ciencias sociales aplicadas a la salud por su aparente falta de validez (Steckler et al., 1992).

En cualquier caso, uno de los argumentos que dota de legitimidad a este tipo de metodologías frente a las cuantitativas en la investigación es la naturalidad que se genera durante la comunicación entre el investigador y los participantes seleccionados a investigar (Calero, 2000), algo muy apreciado en las ciencias sociales. Esto lo podemos observar fundamentalmente en las entrevistas en grupo, donde destaca la técnica de grupo focal, ya que es un procedimiento que congrega a grupos de personas, entre 3 y 12 participantes (Turney & Pocknee, 2005), seleccionadas en base a unos criterios concretos con el objeto de mantener una conversación cercana, natural y lo más horizontal posible (Morgan & Krueger, 1998) mediante un conjunto de preguntas elaboradas con rigor y con un objetivo concreto. Este formato de entrevista, en una siguiente fase, termina transcribiéndose, para posteriormente poder ser codificada, clasificarse y ser analizada (Powell & Single, 1996). Una de las ventajas de este tipo de métodos es que permite obtener una gran cantidad de información y, por tanto, un elevado volumen de datos en poco tiempo (Gibbs, 1997).

La desventaja de este tipo de metodologías que queremos abordar se centra en la gestión de esos grupos de datos y su interpretación, ya que puede ser que se produzca, entre otras limitaciones o errores, un sesgo por algún sujeto investigado que destaque entre el grupo seleccionado o por el propio conductor de la entrevista (Bertoldi et al, 2006).

Internet y el mundo digital es un punto de encuentro para informarse y comunicarse, que facilita la posibilidad de mejorar la relación entre el profesional de la salud y el propio paciente por ser un potencial canal donde mejorar la comunicación de dicho binomio (Lupiañez-Villanueva, 2011).

Los métodos de investigación cualitativa se utilizan cada vez más en todas las disciplinas debido a su capacidad para ayudar a los investigadores a comprender las perspectivas de los participantes en sus propias palabras. Sin embargo, el análisis cualitativo es un proceso laborioso y requiere la intervención de muchos recursos. Para lograr profundidad, los investigadores se limitan a tamaños de muestra más pequeños cuando analizan datos de texto. Un método potencial para abordar este enfoque es el PLN. El análisis de texto cualitativo involucra a investigadores que leen datos, asignan etiquetas de código y desarrollan resultados de forma iterativa. El PLN tiene el potencial de automatizar parte de este proceso. De los estudios que se han centrado en ver las potencialidades del PLN algunos concluyen que este conjunto de técnicas proporciona una base para codificar cualitativamente más rápidamente y un método para validar hallazgos cualitativos (Guetterman et al., 2018, Bustos et al., 2019).

Los métodos cualitativos ofrecen un enorme potencial para contribuir al campo de la investigación de servicios de salud mental, pero tienen como contrapartida el hecho de ser muy laboriosos.

Según nuestra revisión, esta metodología también es novedosa en el campo de la salud mental y en el uso del análisis de información obtenida en grupos focales sobre el uso de nuevas tecnologías de la comunicación.

En este caso partimos de una metodología de recopilación de información con entrevistas a dos grupos focales con una muestra de 5 participantes. Estas muestras estaban formadas por grupos de pacientes que se escogieron en base a estudios preliminares de la investigación con primeros episodios psicóticos los cuales fueron entrevistados por profesionales de la psiquiatría siempre cumpliendo los procedimientos éticos como la obtención de un consentimiento informado firmado voluntariamente. De ahí que, el posible sesgo se produzca por la limitación propia de la enfermedad mental del grupo entrevistado, aunque los conductores del grupo focal conocían muy bien a los pacientes buscando, precisamente, esa naturalidad en la comunicación e intentando motivarlos en la participación.

La diferencia entre ambos grupos de pacientes residía en que el primer grupo se investigó para ver qué grado de adopción a las nuevas tecnologías de la comunicación tenía, si se veían capaces de utilizar Internet y dispositivos smartphone relacionado con temas de salud y de si utilizaban aplicaciones móviles (apps) en este ámbito. Se les preguntó por si entendían que la información que encontraban sobre temas de salud era fiable y si consideraban que les ayudaba en sus casos particulares. Y se les sugirió que si tuvieran una app en su móvil que monitorizara su salud, y que recogiera información de cómo se encuentran y que les recordara tomar la medicación, les pudiera ayudar personalmente y si lo consideraban valioso para su médico. Además de recoger información de si una app de este tipo les daría más autonomía y empoderamiento sobre la enfermedad, y si consideraban que el uso de esta aplicación mejoraría la comunicación con su médico.

Las conclusiones de la primera entrevista al grupo focal establecieron que el uso de Internet y tecnologías de la comunicación e información es similar en los pacientes con enfermedad mental que la población en general y el 100% consideró que las tecnologías propias de dispositivos móviles aplicadas en salud (mHealth) les ayudaría a mejorar la comunicación con su médico y les podría ayudar con la adherencia al tratamiento y toma de medicación. Además, todos los pacientes encuestados afirmaron que una mejor comunicación con su médico hace que se encuentren mejor y más seguros.

El segundo grupo focal estaba diseñado con una muestra de 5 participantes, pero en el que cada uno había pilotado durante un tiempo una app de salud instalada en su smartphone o en el de algún familiar. De los 5 pacientes uno de ellos fue perdido en el

seguimiento. Esta app transfería la información que el usuario cumplimentaba de forma voluntaria cada vez que el software le preguntaba vía notificación, tenía alertas para ayudar en la adherencia al tratamiento de este tipo de enfermedades crónicas y la información quedaba recogida y representada en el software diseñado exclusivamente a los profesionales médicos.

Los resultados estudiados en este grupo focal consideraron que el móvil y las nuevas tecnologías ayudaban a tener una mayor comunicación con su médico en el 75% de los casos. Otro 75% consideraba que las nuevas tecnologías aplicadas al mundo de la salud eran útiles y ayudan a mejorar la adherencia del tratamiento y la toma de medicación diaria. Y por unanimidad, el 100% de la muestra consideró que si la comunicación que tengas con tu médico es mejor hace que tú te encuentres mejor y más seguro.

Con estas investigaciones previas como punto de partida se plantea como nuevo hito qué puede hacer la IA para enriquecer el proceso metodológico de las investigaciones cualitativas en grupos focales y en la relación médico-paciente con enfermedad mental mediante el uso de tecnologías móviles.

2. OBJETIVOS

El objetivo fundamental del presente artículo es investigar las diferencias entre metodologías de trabajo aplicadas a la investigación cualitativa en el campo de la comunicación y la salud.

Se parte de la interpretación subjetiva propia de este campo de investigación basado en metodologías cualitativas obteniendo un tipo de resultados y conclusiones y se plantea el analizar si, con la aplicación de IA y el aprendizaje automático sobre las mismas fuentes de trabajo, los resultados y conclusiones varían o se confirman.

Se pretende realizar una aproximación basada en datos mediante la aplicación de la IA con el objetivo de obtener nuevas conclusiones más empíricas a través de modelos matemáticos de Machine Learning aplicado a investigación cualitativa con la interpretación de textos transcritos para enriquecer el proceso metodológico.

3. METODOLOGÍA

La metodología que se ha seguido en este trabajo de investigación ha consistido en aplicar dentro del contexto de la IA, los algoritmos que trabajaran y nos permiten comprender el lenguaje natural mediante el conjunto de técnicas de PLN.

Con el conjunto de librerías específicas de PLN mediante NLTK (Loper & Bird, 2002) que admite la fácil generación de prototipos y la programación alfabetizada, el pre-procesado de los textos se centró en depurar mediante *scripts* en Python los datos más notables de un texto. Todo ello tras clasificar las cadenas de caracteres, separar

palabras del texto en entidades llamadas *tokens*, en este caso lingüísticas, denominadas palabras, que no necesitan descomponerse en un procesamiento posterior (Webster & Kit, 1992) eliminando las palabras vacías o *stopwords*, es decir, aquellas que acompañan y no tienen significado si no se relacionan con otras palabras.

Tras este pre-procesado, en la siguiente fase (o de exploración inicial) se ha aplicado filtros de limpieza que se suelen usar para poder tratar el texto, mediante la aplicación de expresiones regulares, como, por ejemplo: pasar texto a minúsculas, eliminar signos de puntuación, interrogación, espacios extra, tabulaciones, etc.

Posteriormente se centró en la frecuencia con la que aparecen los caracteres y conexión de frases para obtener resultados que nos permitieran ver cómo, se vectoriza ese conjunto de datos y cómo queda representado a través de una visualización gráfica en el espacio n-dimensional. De esta manera, se pretende identificar las posibles relaciones semánticas y de tipo sintáctico de las palabras o datos procesados.

Se detalla a continuación la metodología propuesta:

3.1 FUENTES Y SOFTWARE

Los textos seleccionados proceden de dos documentos que contienen dos sesiones de análisis cualitativo previo mediante el formato de grupos focales a pacientes con primeros episodios psicóticos. Se trata de dos textos que recogen las transcripciones de esas sesiones de grabadas. Una sesión se centró en pacientes que no disponían de una aplicación móvil. Para centrar el tema, se pretende analizar si el uso de una aplicación móvil con el objetivo de mejorar la comunicación entre pacientes y especialistas de la salud. En el segundo grupo de pacientes se pretende estudiar los efectos del uso del móvil en la mejora de la comunicación médico-paciente.

A partir de ese punto se inicia el proceso y se utiliza *Jupyter Notebook* como interfaz para la realización de los *scripts* en Python. Se trata de un entorno de trabajo de gran aceptación por parte de los científicos de datos, desde su aparición en 2015.

Tras instalar Python 3 (distribución conda), mediante la distribución de Anaconda, se accede a *Jupyter Notebook* para depurar a través de código, qué términos son relevantes en un texto de aquellos que no lo son.

3.2 PRE-PROCESADO DE LOS TEXTOS TRANSCRITOS

En primera instancia, se editó en *Jupyter Notebook* una celda en un nuevo cuaderno *Jupyter*. Se importó el texto al notebook con la codificación *encoding* "UTF-8". Una vez pre-cargada la muestra de pacientes con síntomas psicóticos sin aplicación móvil se elabora un pre-procesado del documento transcrito. Tras ejecutar esta celda se obtuvo

el primer resultado y se pudo pasar a la siguiente celda para avanzar en el procesamiento del lenguaje de los textos transcritos.

Ya iniciado el proceso, el siguiente paso, consistió en trabajar y manipular la cadena de caracteres. Se trata de una de las fases claves y genéricas que se realiza de forma habitual en los procesamientos de texto. El objeto de este procedimiento consiste en ir limpiando el texto original de la transcripción del primer grupo focal. De esta manera, con ayuda de los scripts desarrollados ad hoc, se pudo eliminar del texto importado en bruto, tanto letras mayúsculas como acentos, con el objeto de etiquetar de forma más precisa cada uno de los términos. Para ello, se utilizó la función *lower()*, que permitió presentar en pantalla el valor resultante del texto seleccionado en una cadena de caracteres en minúsculas y sin tildes.

La siguiente fase consistió en negativizar expresiones regulares y patrones de repetición propios de la transcripción de los textos. Para ello se realizó una búsqueda en el texto resultante de un patrón de términos que se repetía por la propia transcripción del grupo focal desde el inicio de la cadena de caracteres que ya teníamos pre-procesados. Estos términos eran los que nombraban a los moderadores de la sesión grupal cada vez que intervenían para hablar con los pacientes.

3.3 TOKENIZACIÓN, ELIMINACIÓN DE PALABRAS VACÍAS Y EXTRACCIÓN DE RAÍCES DE PALABRAS

En un siguiente paso se procedió a tokenizar el texto y así eliminar de la cadena de palabras resultantes los signos de puntuación y espacios, excepto el punto, para no dejar de interpretar bien el conjunto de caracteres. Para ello se utilizó la función *re.sub()* y *replace()*.

Tras este paso se procedió a la fase de limpieza de datos. El objetivo de este procedimiento consiste en eliminar lo que se denomina *stopwords*, o palabras vacías, que no aportan valor al etiquetado de datos, como las preposiciones o los artículos. Este tipo de palabras, como es lógico, se repiten mucho en las transcripciones de textos y no recogen la esencia de las palabras y expresiones que nos ayuden a desarrollar el procesamiento del lenguaje natural. También es habitual en el PLN los procesos de extracción de las raíces de las palabras, conocido como *stemming*, el cual se desestimó al observar que el resultado impedía cierta interpretación de la cadena de caracteres ya que en este contexto determinado la aplicación de este último paso eliminaba sensibilidad y precisión. Posteriormente se consideró fue eliminar cualquier palabra de menos de 3 caracteres con la función *len(w)>3*, de esta manera también se eliminaban palabras vacías y el patrón repetido de las iniciales con dos caracteres que anonimizaban a los pacientes cada vez que intervinieron en la sesión.

Por último, antes de estudiar la frecuencia de caracteres, una vez pre-procesado el texto con el *script*, se presenta el resultado de su ejecución.

3.4 ESTUDIO DE FRECUENCIACIÓN DE PALABRAS

Dentro de las múltiples posibilidades de aplicar las técnicas de PLN se ha ejecutado librerías de código que nos permiten mostrar en un lienzo un conjunto de palabras clave con n-dimensiones diferenciadas dependiendo de la frecuencia analizada. De esta manera, las palabras de mayor repetición en los textos analizados se presentan de forma destacada sobre el resto de los datos de forma que visualmente detectemos las más relevantes en una nube de palabras. Para ello, utilizamos el código propio de la librería *wordcloud* mediante Python que nos permite crear y generar la imagen deseada, mostrarla en pantalla y guardarla para su posterior análisis. Se seleccionó un fondo blanco, tras probar inicialmente un fondo negro, que ayuda a poder identificar visualmente con mayor facilidad los términos destacados de forma más legible, además de asignar con el código unos tamaños máximos de fuente entre las 100 palabras más destacadas.

Se repitió el patrón de PLN para el otro texto de investigación cualitativa a pacientes que venían del programa de primeros episodios psicóticos que sí que utilizaron la aplicación móvil.

3.5 VECTORIZACIÓN DE DATOS

Para mejorar el procesamiento de lenguaje natural, el trabajo de investigación se focalizó en analizar cómo, mediante la aplicación de técnicas de Machine learning, se vectorizaba ese conjunto de datos extraídos y cómo se relacionaban y contextualizaban las diferentes palabras y términos tokenizados.

Para ello, se trabajó las incrustaciones de texto o *Embeddings*, una tecnología que permite representar vectorialmente palabras. La herramienta seleccionada para este cometido fue *Word2vec*, un software de gran aceptación desarrollado por investigadores de Google en 2013 (Mikolov et al., 2013). Con esta tecnología de Inteligencia Artificial se pretende extraer relaciones sintácticas y semánticas entre las palabras ya pre-procesadas. De esta forma, las palabras que comparten más relaciones de cualquier índole entre ellas quedan representadas vectorialmente en dimensiones cercanas mediante puntos en el espacio. La suma de puntos incrustados va aprendiendo y se van adaptando al espacio según se observan asociaciones sintácticas y semánticas. Todo ello con el objeto de obtener patrones o interpretaciones de los datos mediante los algoritmos que proporcionan los ordenadores con estas técnicas de Machine Learning.

A posteriori, con *Jupyter Notebook*, se realizó un nuevo código. En este caso, el procedimiento consistió en elaborar un patrón con el conjunto de caracteres, obteniendo a partir de las frases del corpus de texto, un listado de frases aisladas con espacios.

3.6 LEMATIZACIÓN Y EXTRACCIÓN DE PALABRA RAÍZ

Después de dividir la cadena de caracteres en un listado de frases separadas, se procedió a identificar y mostrar la raíz de las palabras, eliminando cualquier prefijo o sufijo mediante el proceso conocido como lematización o *stemmize*. Para ello se importó la librería *stemmer* en idioma español que nos permitió generar una nueva lista de palabras.

Tras obtener el nuevo conjunto de palabras, se aplicó la librería *Word2vec*. Para ello, se importó a Python la librería *Gensim*, y un conjunto de herramientas que permite convertir las palabras en vectores. Unas palabras que se representan según su contexto en lo que se denomina *embedding* o incrustaciones.

Se generó en primer lugar un archivo de vectores de palabras, un proceso en el que se ejecutó la codificación para la dimensión vectorial donde se incrustaron los puntos, datos o palabras en el espacio. Se trataba de transformar el archivo de trabajo en formato *Word2vec*.

Aplicando el *script gensim.word2vectensor* se extrajeron dos archivos, el primero es *_tensor.tsv*, una archivo en espacio 2D con los vectores de palabras incrustados en su dimensión. y el segundo archivo, *metadata.tsv*, en donde se obtuvieron el conjunto de palabras.

Con las palabras procesadas y preparadas para su incrustación, se procedió a contabilizarlas de forma automática con el código desarrollado.

Posteriormente se procedió a repetir la misma metodología para el segundo conjunto de texto, el de los pacientes que sí que utilizaron la aplicación móvil obteniendo un listado de metadatos diferente al caso anterior para al proceso de representación vectorial en el software *Word2vec*.

3.7 VISUALIZACIÓN DE EMBEDDINGS

La siguiente fase empieza por una visualización de los resultados y se procede a la carga de los archivos generados en un *embedding projector*, en este caso mediante la aplicación web de código abierto <http://projector.tensorflow.org/> (Abadi et al., 2016). A través del panel de datos se seleccionan los archivos analizados con el modelo generado a partir de nuestro dataset y así poder observar los puntos incrustados en su n-dimensión con el panel central o de visualización.

Con esto se consigue mostrar de forma sencilla y automática la visualización de nuestros *embeddings*, en busca de profundizar en las conexiones de palabras en el procesamiento de lenguaje natural.

Se procedió a revisar aquellas palabras que se consideran claves y sus conexiones de datos y asociaciones. La herramienta busca las conexiones más relevantes. Tras probar en un primer momento con la palabra “movil” que se marca en la nube de puntos incrustados porque es clave para su interpretación y centra las conversaciones iniciales de la primera muestra de trabajo. Además, se ha observado que esta palabra, aparecía de forma destacada en la imagen pre-procesada en primera instancia *wordcloud* y en las palabras destacadas del archivo de metadatos en la segunda fase ejecutada del PLN. A partir de esa selección se observaron las relaciones contextuales con palabras como “apps” (software propio de dispositivos móviles), la palabra raíz “medic” que puede derivar en medicina, médico o medicación o la palabra “mal”.

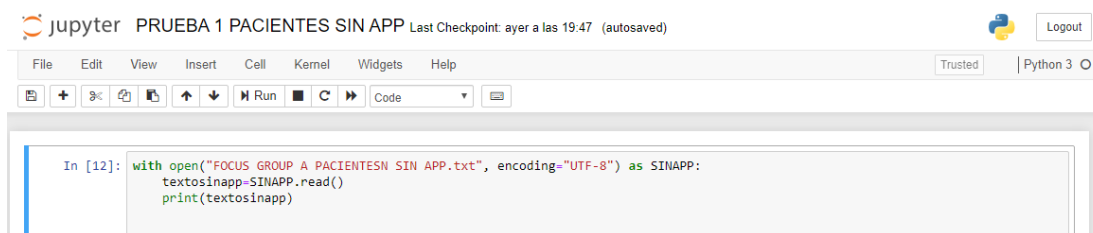
Esto se puede observar en el panel inspector donde se detalla un conjunto de datos vecinos o próximos basados en que todos estos vectores se ubican en un mismo espacio según la similitud coseno.

Tras varias pruebas sobre diferentes vectores, el trabajo se centró en analizar los vectores propios de la comunicación, para intentar ver los resultados que nos aportaba y si podíamos confirmar el objetivo de la investigación. Es decir, para entender si el uso de este tipo de tecnologías en pacientes psicóticos nos aportaba resultados diferenciales en las entrevistas realizadas más allá de la información que ya manejábamos y las interpretaciones cualitativas.

4. RESULTADOS

4.1 PRE-PROCESADO DE LOS TEXTOS TRANSCRITOS

Tras aplicar el método de trabajo con el software Python se obtuvieron los resultados del pre-procesado del documento transcrito desde la fase de investigación cualitativa previa. Una vez ejecutada la primera celda en *Jupyter Notebook* para la importación de texto y así iniciar el PLN con código y scripts de Python se obtuvo el primer resultado y se pudo pasar a la siguiente celda para avanzar en el procesamiento del lenguaje de los textos transcritos (ver ejemplo Figura 1).



The image shows a Jupyter Notebook window titled "PRUEBA 1 PACIENTES SIN APP". The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for file operations, running code, and viewing output. A code cell is active, containing the following Python code:

```
In [12]: with open("FOCUS GROUP A PACIENTES SIN APP.txt", encoding="UTF-8") as SINAPP:
          textosinapp=SINAPP.read()
          print(textosinapp)
```

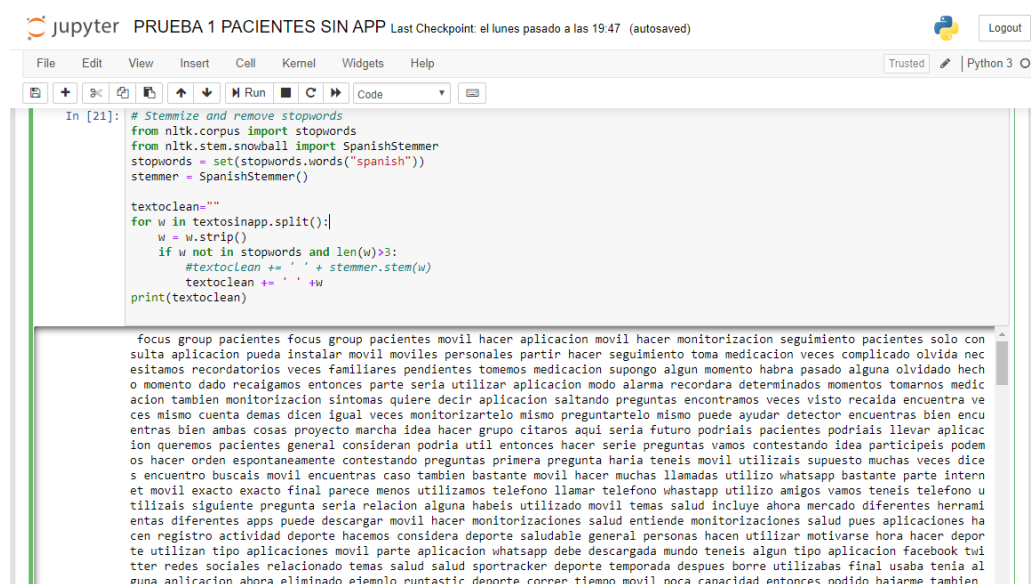
Figura 1. Primera celda del trabajo de importación del texto para procesarlo en PLN.
Fuente: Elaboración propia a partir de importar el texto al software.

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

A partir de este paso se observan los resultados del pre-procesado del documento transcrito desde la fase de investigación cualitativa previa.

4.2 TOKENIZACIÓN, ELIMINACIÓN DE PALABRAS VACÍAS Y EXTRACCIÓN DE RAÍCES DE PALABRAS

El siguiente paso consistió en la *tokenización*, eliminación de palabras vacías y extracción de raíces de palabras obteniendo, tras la ejecución de la nueva celda, un resultado de palabras a partir del texto importado y procesado de los pacientes del grupo focal sin app (ver ejemplo de resultados en Figura 2), algo que después repetimos para el otro grupo focal de pacientes que sí que hicieron uso de la aplicación móvil.



```
In [21]: # Stemming and remove stopwords
from nltk.corpus import stopwords
from nltk.stem.snowball import SpanishStemmer
stopwords = set(stopwords.words("spanish"))
stemmer = SpanishStemmer()

textoclean=""
for w in textosinapp.split():
    w = w.strip()
    if w not in stopwords and len(w)>3:
        #textoclean += ' ' + stemmer.stem(w)
        textoclean += ' ' + w
print(textoclean)

focus group pacientes focus group pacientes movil hacer aplicacion movil hacer monitorizacion seguimiento pacientes solo con
sulta aplicacion pueda instalar movil moviles personales partir hacer seguimiento toma medicacion veces complicado olvida nec
esitamos recordatorios veces familiares pendientes tomemos medicacion supongo algun momento habra pasado alguna olvidado hech
o momento dado recaigamos entonces parte seria utilizar aplicacion modo alarma recordara determinados momentos tomarnos medic
acion tambien monitorizacion sintomas quiere decir aplicacion saltando preguntas encontramos veces visto recaida encuentra ve
ces mismo cuenta demas dicen igual veces monitorizartelo mismo preguntartelo mismo puede ayudar detector encuentras bien encu
entras bien ambas cosas proyecto marcha idea hacer grupo citaras aqui seria futuro podriais pacientes podriais llevar aplicac
ion queremos pacientes general consideran podria util entonces hacer serie preguntas vamos contestando idea participeis podem
os hacer orden espontaneamente contestando preguntas primera pregunta haria teneis movil utilizais supuesto muchas veces dice
s encuentro buskais movil encuentras caso tambien bastante movil hacer muchas llamadas utilizo whatsapp bastante parte intern
et movil exacto exacto final parece menos utilizamos telefono llamar telefono whatsapp utilizo amigos vamos teneis telefono u
tilizais siguiente pregunta seria relacion alguna habeis utilizado movil temas salud incluye ahora mercado diferentes herrami
entas diferentes apps puede descargar movil hacer monitorizaciones salud entiende monitorizaciones salud pues aplicaciones ha
cen registro actividad deporte hacemos considera deporte saludable general personas hacen utilizar motivarse hora hacen depor
te utilizan tipo aplicaciones movil parte aplicacion whatsapp debe descargada mundo teneis algun tipo aplicacion facebook twi
tter redes sociales relacionado temas salud salud sportracker deporte temporada despues borre utilizabas final usaba tenia al
guna aplicacion ahora eliminado ejemplo runtastic deporte correr tiempo movil poca capacidad entonces podido bajarme tambien
```

Figura 2. Tokenización, stemming y stopwords.

Fuente: Elaboración propia a partir de la ejecución del código en Python.

4.3. ESTUDIO DE FRECUENCIACIÓN DE PALABRAS

Tras implementar el código sobre nuestro texto *tokenizado* del primero grupo de pacientes sin app, se trabajó en la cuantificación de palabras y las de mayor repetición en el texto analizado se mostraron de forma destacada sobre el resto de los datos visualmente, obteniendo el siguiente resultado por frecuencia de aparición: “aplicación”, “bien”, “puede”, “medico”, “creo”, “móvil”, “medicación”, “puede”, “mejor” y “hacer”.

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

Además, los metadatos “bien” y “mejor” aparecen en el top 10 de pacientes sin app, al igual que “mejor” coincide también en ese top 10 de metadatos en pacientes con app.

Tabla 1. *Extracto del listado de metadatos generados al aplicar el software para contabilizar la frecuencia de palabras en pacientes sin app y con app.*

Metadatos Pacientes sin app		Metadatos Pacientes con app	
Word	Count	Word	Count
Si	103	si	185
medic	47	pas	45
aplic	45	aplic	44
hac	44	medic	44
cre	38	cre	40
pued	36	clar	40
com	32	mejor	35
bien	32	pregun	32
mal	26	haz	32
mejor	25	pued	30
movil	23	contest	26
inform	21	igual	24
utiliz	21	bien	23

Fuente: *Elaboración propia a partir de ejecutar el código Python.*

4.5 VISUALIZACIÓN DE EMBEDDINGS

Tras ejecutar en la siguientes fase la carga de los archivos generados en el software *embedding projector* para la visualización de datos incrustados se muestra el resultado de investigar entre la suma de datos vectoriales las raíces propias de las palabras clave del top 25 de datos vecinos o cercanos: comunicación, información o desinformación, utilidad, relación y sensación para observar las asociaciones y sentimientos que generan en estos casos, tanto a pacientes que no utilizaron la aplicación de salud móvil como los que sí que la tuvieron instalada en sus dispositivos smartphone (ver ejemplo en la Figura 5).

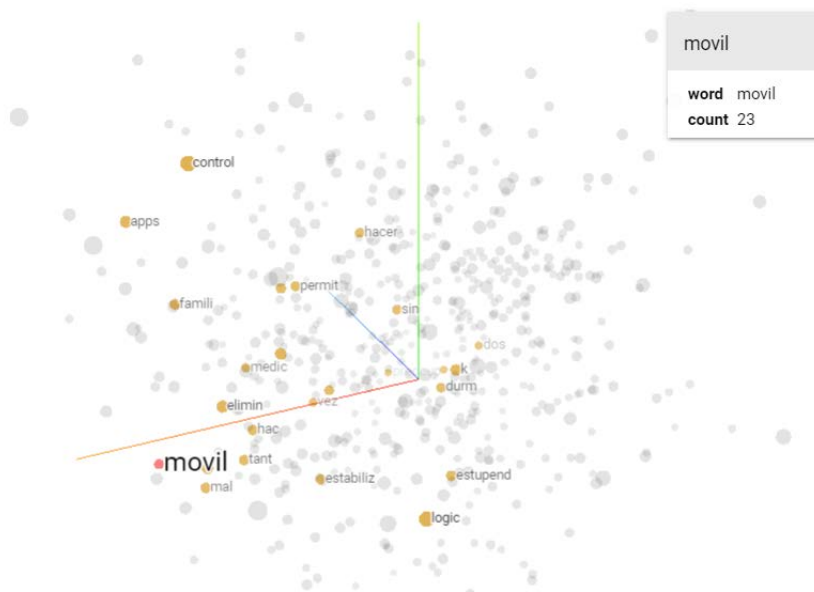


Figura 5. Resultado visual del vector “movil” en la muestra de pacientes sin app.

Fuente: Elaboración propia a partir de ejecutar Word2vec.

4.6 RESULTADOS DE VECTORES EN PRIMER GRUPO FOCAL DE PACIENTES SIN APP

Para poder ver si se cumple el objetivo final de la investigación nos centramos en analizar unos vectores concretos, centrados en las siguientes palabras: comunicación, información o desinformación, utilidad, relación y sensación. Esta muestra de análisis se seleccionó para identificar las asociaciones de datos con ellas.

A continuación, mostramos el resultado del análisis del vector analizado “Inform” en pacientes sin aplicación móvil de salud instalada: “habl”, “frecuent”, “preguntar”, “posit”, “much”, “absurd”, “aplic”, “detect”.

Para el vector “comun” en pacientes sin aplicación móvil de salud instalada el resultado fue: “doctores”, “tecnolog”, “hospital”, “concienci”, “dud”, “favor”, “inconvenient”, “cuent”, “apreci”.

En cuanto al análisis del vector “desinform” en pacientes sin aplicación móvil de salud instalada: “enfermed”, “import”, “diagnostic”, “valor”, “doctor”, “salud”, “estabiliz”, “recaig”, “person”, “mejor”.

El resultado del vector “sentir” en la primera muestra fue: “particul”, “sint”, “absurd”, “selección”, “ver”, “clinic”, “distinta”, “difícil”, “puntual”, “mejor”.

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

Respecto al análisis de los vectores generados en el primer grupo de entrevistas, analizamos el vector “util” con el siguiente resultado: “enfermed”, “seguimient”, “impor”t, “afront”, “sencill”, “app”, “concienci”, “relacion”, “privac”.

Y por último, en cuanto al vector “relacion” en pacientes sin aplicación móvil obtuvimos los siguientes datos: “detección”, “doctor”, “ningun”, “hacer”, “correspond”, “ayud”, “empiez”, “util”, “ningún”.

4.7 RESULTADOS DE VECTORES EN SEGUNDO GRUPO FOCAL DE PACIENTES CON APP

Para el segundo texto del grupo focal de pacientes que tuvieron la app de salud instalada en dispositivo móvil repetimos el mismo conjunto de vectores a analizar. El resultado obtenido, en este caso, del análisis del vector “inform” en pacientes que tuvieron instalada la aplicación móvil de salud fue el siguiente: “mejor”, “igual”, “pregunt”, “anot”, “activ”, “comun”, “cheq”, “ejerc”, “posibil”, “genial”.

Para el vector “comun” en pacientes que tuvieron instalada la aplicación móvil de salud los vectores vecinos fueron: “medic”, “inform”, “preguntart”, “habit”, “bidireccional”, “segur”, “contest”, “coment”, “normal”, “eriquec” (Ver ejemplo Figura 6).

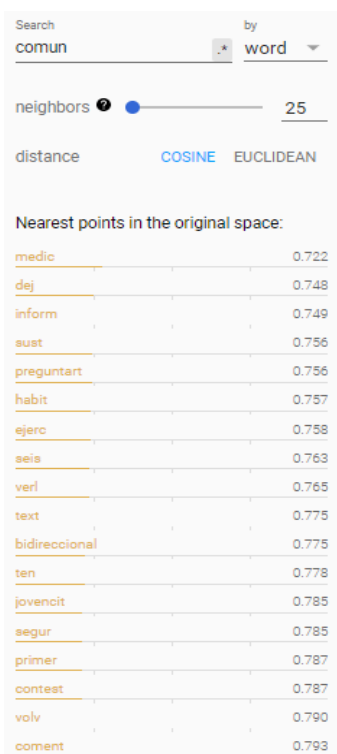


Figura 6. Resultado de vectores vecino para el punto “común”, raíz tokenizada de la palabra clave comunicación.

Fuente: Elaboración propia a partir de ejecutar Word2vec.

En cuanto al análisis del vector “sensacion” en pacientes que tuvieron instalada la aplicación móvil de salud el software mostró los siguientes datos conectados: “sintom”, “psiquiatr”, “nuev”, “bidireccional”, “comod”, “bien”, “util”, “ocasion”, “alarm”.

Para el vector “util” en pacientes que tuvieron instalada la aplicación móvil de salud: “buen”, “sup”, “apoy”, “comprend”, “psicot”, “diari”, “encuentr”, “cuest”, “dorm”, “coñaz”, “quej”, “recordatori”, “cambi”.

Y en este grupo, el análisis del vector “relacion” mostró los siguientes datos procesados: “contig”, “tratamient”, “notif”, “medic”, “complet”, “utiliz”, “resolv”, “demuestr”, “estres”, “descans”.

4.8 RESULTADOS DE LAS DOS TÉCNICAS APLICADAS SOBRE METODOLOGÍAS DE INVESTIGACIÓN CUALITATIVA

Con el conjunto de datos y resultados obtenidos elaboramos una comparativa entre los resultados de los textos cualitativos previos con metodología tradicional y los nuevos una vez aplicamos las técnicas de PLN.

La tabla comparativa de resultados (Tabla 2) nos muestra que se replican datos que reproducen las conclusiones de la primera investigación realizada con metodología tradicional, pero que la nueva metodología aporta matices que se perdían sin la aplicación del PLN.

Se obtuvieron unos resultados tras la aplicación de estas técnicas propias de la IA que confirmaban los resultados de las categorías analizadas en la metodología tradicional de forma más subjetiva. Además, se extrajeron nuevos resultados que no se identificaron previamente en el estudio previo.

Respecto a la categoría analizada si las nuevas tecnologías ayudan a mejorar la comunicación médico-paciente, partíamos de un sí en el 75% de los casos analizados y tras la aplicación de técnicas de PLN el resultado de vectores vecinos “inform”, “genial”, “mejor” entre otras confirma ese sí afirmación inicial. Aparecen además nuevos vectores de información conectados como “habit”, “igual”, “ejerc”.

En cuanto a si son útiles las nuevas tecnologías aplicadas a la salud de la cual partíamos de una afirmación positiva del 75% de los casos, de nuevo los resultados confirman tras el PLN con vectores vecinos como “buen”, “apoy” o “compren”. También aparecen nuevos matices con el resultado de vectores: “psicot”, “diari” y “dorm”.

Por último, el resultado para la categoría de si la comunicación es mejor, el paciente se siente mejor, el cual partía del análisis clásico con una afirmación del 100%, se confirma tras el PLN con vectores obtenidos como: “segur”, “útil”, “bien” y nuevos matices como “alarm” o “psiquiatr”.

Tabla 2. *Comparativa de resultados de las dos técnicas aplicadas sobre metodologías de investigación cualitativa*

Categoría	Conclusiones previas subjetivas	Conclusiones tras aplicación IA: vectores vecinos que confirman	Conclusiones tras aplicación IA: vectores vecinos que aportan nuevos resultados
Las nuevas tecnologías ayudan a mejorar la comunicación médico-paciente	75% sí	“inform”, “enriquec”, “mejor”, “contest”, “genial”, “bidireccional”	“habit”, “igual”, “ejerc”
Las nuevas tecnologías aplicadas a la salud son útiles	75% sí	“buen”, “apoy”, “comprend”, “cambi”	“psicot”, “diari”, “dorm”, “coñaz”
Si la comunicación es mejor el paciente se siente mejor	100% sí	“segur”, “util”, “bien”, “comod”	“alarm”, “sintom”, “psiquiatr”

Fuente: *Elaboración propia*

5. CONCLUSIONES

Tras aplicar las técnicas del PLN sobre los textos de datos analizados previamente sin IA podemos concluir que se producen conexiones que confirman parte de las conclusiones obtenidas con el método tradicional y que la nueva metodología aporta matices que se perdían sin la aplicación del PLN además de que el PLN automatiza el proceso para enfrentarse a bases de datos más amplias.

Fruto de esta investigación, se puede observar similitudes y diferencias de resultados en la nueva metodología sobre el trabajo de investigación cualitativa en el campo de la comunicación y la salud, lo cual nos aporta nuevas conclusiones. Esta aproximación de aplicación de técnicas IA en pacientes con enfermedad mental nos indica que es un camino con mucho potencial y que sigue creciendo a ritmo exponencial en todos los ámbitos de la sociedad.

En este caso en concreto, cuando se ha comparado si las nuevas tecnologías ayudan a mejorar la comunicación entre médico y paciente se ha observado que aplicando PLN se conectan palabras como “genial” o “bidireccional”, algo que enlaza, por el significado de dichas palabras, con las conclusiones previas de que realmente sí lo hacían en el 75% de los casos.

Lo mismo ocurre con el análisis de la categoría sobre si una mejor comunicación repercute en que el paciente se siente mejor, y donde en el estudio tradicional se

concluyó con que así era en el total de los casos, al estudiar las conexiones entre palabras, también se ha observado que valida estos resultados con palabras asociadas como “segur”, “util” y “bien”. Además, aparecen otras conexiones a tener en cuenta como “sintom” o “alarm” que son difíciles de interpretar sin un contexto.

La IA aprende, automatiza y vincula palabras vecinas semánticamente de forma diferente, así que aplicamos un método que sin el aprendizaje algorítmico se nos puede escapar en el análisis previo de una investigación cualitativa para extraer nuevas conclusiones del trabajo, conocido como análisis *data-driven* (Rodriguez et al., 2011).

Con este estudio, confirmamos que la IA es útil para la aplicación en áreas biomédicas y más concretamente en el ámbito que componen la salud mental y comunicación. Concluimos, por tanto, que una de las posibles aplicaciones de la IA en la psiquiatría es el PLN, obtenido en este caso a través de entrevistas focales, y que la IA permite al mundo informático y los algoritmos comprender, interpretar y manipular el lenguaje humano automatizándolo. Observamos como el PLN se posiciona como una técnica válida dentro de la IA para el análisis del discurso, como ya se había demostrado previamente.

En cualquier caso, se detecta en esta investigación que hay diversas limitaciones ya que el procesamiento propio de un software restringe al lenguaje natural y que, aunque se aísle la subjetividad, en los textos cualitativos consideramos que es necesario trabajar con técnicas mixtas para enriquecer las metodologías de investigación cualitativa (Guetterman et al., 2018). El lenguaje está lleno de palabras con diferentes acepciones y matices, que según el contexto puede adquirir diferentes significados, incluso puede variar según la propia intención del comunicador al expresarse y todo esto es algo que de momento no se está solucionando con este tipo de técnicas.

Además, en el campo de la salud mental, la prosodia (estado emocional del hablante) es determinante para entender cómo se encuentra el paciente y en este caso, vemos que, aunque el PLN nos ayuda a automatizar y conectar palabras no acaba de ser autosuficiente en la actualidad para detectar las expresiones de emoción asociadas a las palabras. Otra limitación detectada fue concretamente en las preguntas que realizaron los especialistas médicos o en la actitud de los pacientes en las sesiones de entrevistas realizadas ya que, según los resultados, la palabra “sí” apareció en el mayor número de ocasiones debido a que los pacientes simplemente afirmaron y asintieron lo que se les preguntaba sin poder extraer más palabras a ser analizadas. De esta manera, también se pudo concluir que es necesario incrementar la cantidad de textos y datos para poder extraer conclusiones más fiables y profundas.

En definitiva, el PLN es ya una realidad, un campo abierto y lleno de oportunidades, pero debe seguir avanzando en un futuro inmediato para poder incorporar valor a las investigaciones cualitativas y a los textos que se aplique, siempre que se disponga de bastante material para ser analizado.

En conclusión, para poder entender el verdadero significado de un texto nos encontramos con un desafío y reto indiscutible en donde los algoritmos empiezan a darnos informaciones y datos de interés, pero que todavía necesita seguir evolucionando. Todo ello para que, de forma complementaria a las interpretaciones subjetivas innatas al lenguaje, nos den resultados más determinantes en las investigaciones que realicemos. El camino más prometedor del PLN pasa por entender las palabras procesadas en contextos y ámbitos determinados pudiendo extraer un sentido unificado de las mismas.

5. REFERENCIAS

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... & Ghemawat, S. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*. Disponible en: <https://bit.ly/3bQSEZm>
- Aghion, P., Jones, B. F., & Jones, C. I. (2017). Artificial intelligence and economic growth (No. w23928). *National Bureau of Economic Research*. Disponible en: <https://bit.ly/2VNLUWL>
- Bertoldi, S., Fiorito, M. E., & Álvarez, M. (2006). Grupo Focal y Desarrollo local: aportes para una articulación teórico-metodológica. *Ciencia, docencia y tecnología*, 17(33), 111-131. Disponible en: <https://bit.ly/2yRjRwr>
- Brunn, M., Diefenbacher, A., Courtet, P., & Genieys, W. (2020). The Future is Knocking: How Artificial Intelligence Will Fundamentally Change Psychiatry. *Academic Psychiatry*, Online. Disponible en: <https://bit.ly/3dUvEcQ>
- Bustos, A., Pertusa, A., Salinas, J. M., & de la Iglesia-Vayá, M. (2019). Padchest: A large chest x-ray image dataset with multi-label annotated reports. *arXiv preprint arXiv:1901.07441*. Disponible en: <https://bit.ly/2KLyjbX>
- Calero, J. L. (2000). Investigación cualitativa y cuantitativa. Problemas no resueltos en los debates actuales. *Rev. Cubana Endocrinol*, 11(3), 192-8. Disponible en: <https://go.aws/2yUCgbA>
- Gibbs, A. (1997). Focus groups. *Social research update*, 19(8), 1-8. Disponible en: <https://bit.ly/3bQTdCG>
- Guetterman, T. C., Chang, T., DeJonckheere, M., Basu, T., Scruggs, E., & Vydiswaran, V. (2018). Augmenting Qualitative Text Analysis with Natural Language Processing: Methodological Study. *Journal of medical Internet research*, 20(6), e231. Disponible en: <https://bit.ly/3aP3ded>

- He, J., Baxter, S. L., Xu, J., Xu, J., Zhou, X., & Zhang, K. (2019). The practical implementation of artificial intelligence technologies in medicine. *Nature medicine*, 25(1), 30–36. Disponible en: <https://bit.ly/3h6L4go>
- Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. *Science*, 349(6245), 261-266. Disponible en: <https://stanford.io/3cGRU8J>
- Liddy, E. D. (2001). Natural language processing. Disponible en: <https://bit.ly/2zHhpJp>
- Loper, E., & Bird, S. (2002). NLTK: the natural language toolkit. *arXiv preprint cs/0205028*. Disponible en: <https://bit.ly/2VJV1Yi>
- Lupiáñez-Villanueva, F. (2011). Salud e internet: más allá de la calidad de la información. *Revista española de cardiología*, 64(10), 849-850. Disponible en: <https://bit.ly/3bS3SwM>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems* (pp. 3111-3119). Disponible en: <https://bit.ly/2KI7CVF>
- Morgan, D. L., Krueger, R. A., & Scannell, A. U. (1998). Planning focus groups. Sage. Disponible en: <https://bit.ly/2Soo69A>
- Perrault, R., Shoham Y., Brynjolfsson E., Clark J., Etchemendy J., Grosz, B., Lyons, T., Manyika T., Mishra, S., & Niebles J.C. (2019). “The AI Index 2019 Annual Report”, AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA, December 2019. Disponible en: <https://stanford.io/2KFNLGN>
- Powell, R. A., & Single, H. M. (1996). Focus groups. *International journal for quality in health care*, 8(5), 499-504. Disponible en: <https://bit.ly/3aKAWFt>
- Rodriguez, M., Sivic, J., Laptev, I., & Audibert, J. Y. (2011, November). Data-driven crowd analysis in videos. In 2011 International Conference on Computer Vision (pp. 1235-1242). IEEE. Disponible en: <https://bit.ly/2SfWiV1>
- Rong, G., Mendez, A., Assi, E. B., Zhao, B., & Sawan, M. (2020). Artificial Intelligence in Healthcare: Review and Prediction Case Studies. Engineering. Disponible en: <https://bit.ly/30mAqEO>
- Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., & McCormick, L. (1992). Toward Integrating Qualitative and Quantitative Methods: An Introduction. *Health Education Quarterly*, 19(1), 1–8. Disponible en: <https://bit.ly/2YfU6k7>

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

Taylor, S. J., & Bogdan, R. (1987). Introducción a los métodos cualitativos de investigación (Vol. 1). Barcelona: Paidós. Disponible en: <https://go.aws/2SfsA2b>

Turney, L., & Pocknee, C. (2005). Virtual focus groups: New frontiers in research. *International Journal of Qualitative Methods*, 4(2), 32-43. Disponible en: <https://bit.ly/3bPzByB>

Webster, J. J., & Kit, C. (1992). Tokenization as the initial phase in NLP. In COLING 1992 Volume 4: The 15th International Conference on Computational Linguistics. Disponible en: <https://bit.ly/2SktsCP>