

# ANÁLISIS SEMÁNTICO

EN SUJETOS CON ENFERMEDAD DE ALZHEIMER, USANDO INTELIGENCIA ARTIFICIAL

## SEMANTIC ANALYSIS IN SUBJECTS WITH ALZHEIMER'S DISEASE USING ARTIFICIAL INTELLIGENCE

Luis Roberto García-Noguez

E-mail: [luis.roberto.garcia@uaq.mx](mailto:luis.roberto.garcia@uaq.mx)

ORCID: <https://orcid.org/0000-0002-6716-9148>

Universidad Autónoma de Querétaro. México.

Cita sugerida (APA, séptima edición)

García-Noguez, L. R. (2025). Análisis semántico en sujetos con enfermedad de Alzheimer, usando Inteligencia Artificial. *Universidad y Sociedad*, 17(1), e4911.

### RESUMEN

En los últimos años, han surgido numerosas investigaciones enfocadas en identificar rasgos de deterioro cognitivo utilizando audio de pruebas neuropsicológicas mediante herramientas de Inteligencia Artificial. Desafortunadamente, los modelos que presentan mejor desempeño requieren una gran cantidad de datos y no son explicables debido a la complejidad de los modelos. Dados estos inconvenientes, en este trabajo se analizan diferentes modelos de vectorización de oraciones pre entrenados en las transcripciones de la base de datos Pitt Corpus con la intención de obtener el mejor modelo para esta tarea. Asimismo, se proponen nuevos elementos para analizar características semánticas en sujetos con enfermedad de Alzheimer. Se obtuvieron mejores resultados: 0.8712 de exactitud, 0.8729 de precisión, 0.8712 de exhaustividad y 0.8709 de puntaje F1 usando RoBERTa como modelo para vectorizar oraciones y SVM como clasificador en conjunto con las cuatro características del lenguaje: palabras clave, circunloquios, complejidad del texto y similitud con texto de ejemplo. Las pruebas muestran que las características del lenguaje propuestas aumentaron los resultados de la clasificación, siendo las palabras clave la más importante de éstas.

**Palabras clave:** Vectorización de oraciones, Deterioro del lenguaje, TabTransformers, Auto-atención.

### ABSTRACT

In recent years, many investigations have focused on identifying cognitive impairment traits using audio from neuropsychological tests using Artificial Intelligence tools. Unfortunately, the best-performing models require a large amount of data and are not explainable due to their complexity. Given these drawbacks, in this work, we analyze different sentence vectorization models pre-trained on the Pitt Corpus database transcripts to obtain the best model for this task. Likewise, new elements are proposed to analyze semantic features in subjects with Alzheimer's disease. Better results were obtained: 0.8712 accuracy, 0.8729 precision, 0.8712 exhaustiveness, and 0.8709 F1 score using RoBERTa as a model to vectorize sentences and SVM as a classifier in conjunction with the four language features: keywords, circumlocutions, text complexity and similarity to example text. Tests show that the proposed language features increased ranking results, with keywords being the most important.

**Keywords:** Word embeddings, Language impairment, TabTransformer, Auto-attention.



## INTRODUCCIÓN

Según la Organización Mundial de la Salud (WHO), alrededor de 55 millones de personas padecen demencia en todo el mundo, siendo la enfermedad de Alzheimer la más común. Estas enfermedades afectan la calidad de vida de quienes las padecen, de sus familias y de su entorno. Los pacientes con demencia pueden sufrir alteraciones del pensamiento, del aprendizaje, de la memoria, del juicio y de la toma de decisiones (Deary et al., 2009).

El deterioro del lenguaje es una característica central de la enfermedad de Alzheimer (EA) y otras demencias (Penfold et al., 2022). Estudios previos han demostrado una relación entre la gravedad de los síntomas de EA y el deterioro de la capacidad del habla y el lenguaje en tareas de descripción de imágenes. Los cambios en el habla y el lenguaje incluyen alteraciones en la frecuencia de las palabras, la velocidad del habla, las dificultades para encontrar palabras y las repeticiones (Hussein et al., 2022).

Dadas las características observadas en el lenguaje de sujetos con EA, el Procesamiento del Lenguaje Natural (PLN) está surgiendo como un método novedoso para evaluar el habla y el lenguaje en personas con trastornos neurológicos y mentales. No obstante, los modelos lingüísticos actuales que mejor funcionan requieren una gran cantidad de datos y son difíciles de interpretar (Nazi & Peng, 2024). Considerando la dificultad que implica generar muestras de audio de sujetos con y sin EA debido al costo y al tiempo necesario, surgen estrategias como el aumento de datos o el uso de embebidos de palabras pre entrenados para generar los clasificadores automáticos. Este trabajo se enfoca particularmente en probar modelos de vectorización de palabras basados en el codificador y decodificador de las redes transformers. Además, se proponen cuatro características de lenguaje observadas en sujetos con enfermedad de Alzheimer para observar el desempeño de las misma en el clasificador e intentar mejorar la explicabilidad de estos clasificadores.

El enfoque asumido y la contribución se pueden resumir de la siguiente manera: Se propone una nueva metodología para analizar características semánticas en el lenguaje de sujetos con y sin enfermedad de Alzheimer; asimismo, se prueban diferentes modelos de vectorización de oraciones para las transcripciones de audios de sujetos en una prueba de lenguaje espontáneo.

## MATERIALES Y MÉTODOS

En esta sección se presenta la metodología propuesta para analizar cuatro características semánticas observadas en sujetos con enfermedad de Alzheimer. Además, se realiza una breve explicación de los modelos de vectorización de oraciones seleccionados y los clasificadores

analizados. Asimismo, se describe el conjunto de datos y el preprocesamiento de texto.

### Conjunto de datos

En este estudio, se usaron datos del Pitt Corpus (Becker et al., 1994), que consta de resultados de la prueba Mini-Mental State Examination (MMSE), datos demográficos, grabaciones de audio y transcripciones recopiladas como parte del protocolo administrado por el Estudio sobre el Alzheimer y otras demencias relacionadas (Alzheimer's and Related Dementias Study) en la Facultad de Medicina de la Universidad de Pittsburgh. De una muestra que incluía sujetos con diferentes diagnósticos de demencia y deterioro cognitivo leve, únicamente se seleccionaron aquellos etiquetados con enfermedad de Alzheimer.

### Preprocesamiento del texto

De la base de datos antes descrita, se seleccionan sujetos de control (243 muestras) y sujetos con enfermedad de Alzheimer (246 muestras). Estos textos son el resultado de la transcripción de audio de una prueba conocida como "el robo de galletas". Cada grabación con la respuesta de los sujetos se transcribe manualmente siguiendo el protocolo TalkBank CHAT (MacWhinney, 2003). Estas transcripciones contienen información breve sobre el sujeto de prueba, incluyendo su identificador, edad, género, puntuación obtenida en la prueba Mini-Mental State Examination (MMSE) y diagnóstico. Se transcriben las instrucciones e intervenciones del investigador, así como la descripción proporcionada durante la prueba por el sujeto. El corpus extraído de la transcripción se generó mediante un filtrado, que solo extrae la respuesta del participante. El preprocesamiento aplicado al corpus consistió en eliminar la simbología CHAT agregada por los investigadores, los signos de puntuación, los caracteres especiales y convertir el texto a minúsculas.

### Representaciones vectoriales de palabras.

Las representaciones vectoriales de las palabras, también conocidas como "word embeddings" o embebidos de palabras, son una forma de representar las palabras como vectores numéricos de alta dimensión. Estos vectores capturan las propiedades semánticas y sintácticas de las palabras, lo que permite que sean utilizadas en una variedad de aplicaciones de Procesamiento de Lenguaje Natural (NLP). Particularmente, para este trabajo se probarán los modelos BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), Llama 3 8B (Dubey, 2024) en la clasificación de sujetos con posible EA. La selección de los modelos de solo encoder (BERT y RoBERTa) es a consecuencia del buen desempeño que han mostrado en trabajos similares (Matošević & Jović, 2022).

Para probar estos modelos se utilizó HuggingFace. Por otro lado, la elección de los modelos de solo decoder fue

para probar su desempeño en la tarea analizada. Para usar estos modelos se empleó Ollama. A continuación, se describen brevemente estos modelos.

### **Modelo BERT y RoBERTa**

El modelo BERT (Devlin et al., 2018) procesa todos los elementos de una oración simultáneamente utilizando un mecanismo llamado autoatención. Dicho modelo se entrenó previamente en dos tareas no supervisadas. La primera consiste en el modelado del lenguaje enmascarado (MLM), que predice las palabras que faltan (enmascaradas) en una oración. La segunda es la predicción de la siguiente oración (NSP), donde el sistema se entrena para predecir si una oración sigue a otra.

El modelo RoBERTa (Liu et al., 2019) se basa en el modelo BERT, pero incorpora algunos cambios que permiten un mejor desempeño en varias tareas del Procesamiento del Lenguaje Natural. RoBERTa utiliza el mismo concepto que BERT, pero elimina la parte de NSP y utiliza tamaños de lote más grandes. Además, este método introduce un enmascaramiento dinámico para el MLM, cuyos tokens enmascarados cambian durante las épocas de entrenamiento.

El modelo RoBERTa mapea las palabras a una dimensión de 768 al igual que BERT. Se genera un vector de oración con el promedio de cada incrustación de palabras. Estos vectores se pueden utilizar para tareas como la agrupación o la búsqueda semántica. En este caso, se utilizó para la tarea de clasificación.

Para el caso de los vectores obtenidos del modelo BERT se utilizaron los mismos pasos que los empleados para el modelo RoBERTa.

### **Modelo Llama 3 8B**

Llama 3 8B es el modelo de lenguaje de código abierto más reciente de Meta AI (Dubey, 2024), diseñado para múltiples casos de uso, como responder preguntas en lenguaje natural, escribir código y generar ideas. Este modelo, se distribuye bajo una licencia permisiva, lo que permite a investigadores, desarrolladores y empresas utilizarlo tanto para investigación como para aplicaciones comerciales.

Además, este modelo de lenguaje de gran tamaño muestra un rendimiento superior en una amplia gama de tareas en comparación con modelos anteriores, gracias a mejoras en su arquitectura y entrenamiento. En este mismo rubro, el modelo tiene una ventana de contexto de 8K tokens, lo que le permite procesar y generar textos más largos y complejos.

### **Modelo Mistral Nemo 12 B**

Mistral NeMo 12B es un modelo de lenguaje de última generación desarrollado por Mistral AI en colaboración con NVIDIA. Este potente modelo ofrece una gran ventana de contexto de hasta 128 000 tokens y se destaca en razonamiento y precisión de codificación. Además, supera a otros modelos de tamaño similar en diversos benchmarks, lo que indica su superioridad en tareas como el razonamiento de sentido común, la comprensión de preguntas y la generación de texto.

Con una ventana de contexto de hasta 128k tokens puede procesar y generar texto más largo y complejo, esto lo hace ideal para tareas como la traducción de documentos largos o la generación de código. Mistral NeMo 12B es un modelo versátil que puede utilizarse para una amplia gama de aplicaciones, desde chatbots y asistentes virtuales hasta herramientas de generación de contenido y sistemas de recomendación.

Un punto interesante de este modelo es que emplea embeddings rotatorios para modelar la posición de las palabras en una secuencia, lo que permite al modelo capturar mejor las relaciones entre las palabras.

**Modelos de clasificación.** Una vez generadas las vectorizaciones de las transcripciones, se utilizaron tres algoritmos de clasificación para probar la combinación con los mejores resultados. Los modelos seleccionados fueron: Regresión Logística, Máquina de Soporte de Vectores (SVM) y TabTransformer (Huang et al., 2020). El motivo para seleccionar estos algoritmos es debido a que tanto Regresión Logística como SVM han tenido buenos desempeños en esta tarea. Por otro lado, se seleccionó TabTransformer para probar el desempeño de un clasificador basado en Redes Transformers en la clasificación de sujetos con EA. A continuación, se describen brevemente cada uno de estos modelos.

### **Clasificador usando el modelo de Regresión Logística**

Este modelo es un algoritmo de aprendizaje automático supervisado que se utiliza para tareas de clasificación. El objetivo es predecir la probabilidad de que un elemento pertenezca a una clase determinada (Das, 2024).

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

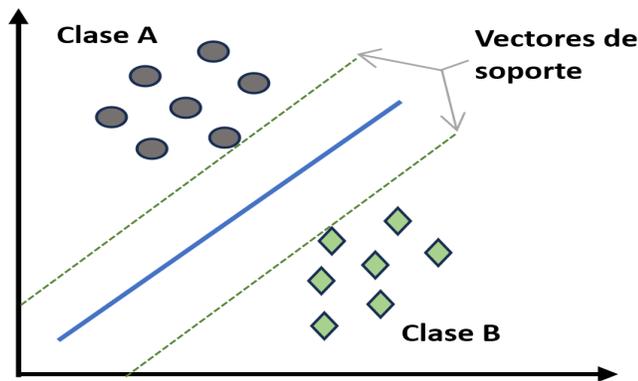
La regresión logística se usa para la clasificación binaria, donde utilizamos la función sigmoidea (ecuación 1), que produce un valor de probabilidad entre 0 y 1.

Este modelo funciona de la siguiente manera: por ejemplo, tenemos la clase 0 (controles) y Clase 1 (casos), si el valor de la función logística para una entrada es mayor que 0.5, entonces pertenece a la Clase 1. Por otro lado, si se obtiene cualquier otro valor, pertenece a la Clase 0. Este algoritmo es una extensión de la regresión lineal. Establezcamos los hiperparámetros 'C' y 'solver' en 0.1 y 'liblinear' respectivamente después del ajuste de los mismos.

**Máquina de soporte de vectores:** La máquina de vectores de soporte (SVM) es un algoritmo de aprendizaje automático supervisado que se utiliza principalmente para tareas de clasificación. También se puede emplear para regresión y detección de valores atípicos.

La idea central detrás de SVM es encontrar el mejor hiperplano posible que separe los puntos de datos en diferentes clases. En 2D, este hiperplano es una línea; en dimensiones superiores, es un plano o una superficie más compleja. El hiperplano óptimo es el que maximiza el margen entre los puntos de datos más cercanos de diferentes clases. Estos puntos más cercanos se denominan vectores de soporte (Figura 1).

Fig 1: Máquina de Soporte de Vectores.



Fuente: Elaboración propia.

Maximizar el margen conduce a un mejor rendimiento de generalización. Un margen más amplio implica que el modelo es menos sensible al ruido y a los valores atípicos, lo que lo hace más robusto.

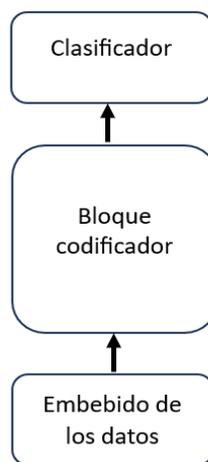
Para la implementación de este clasificador se utilizó la librería Scikit-learn (Pedregosa et al., 2011). No se realizó ningún cambio en los hiper parámetros.

## TabTransformer

TabTransformer (Huang et al., 2020) es un modelo de aprendizaje automático que combina la arquitectura de las Redes Transformers con técnicas de procesamiento de datos tabulares. Está diseñado para analizar y aprender patrones en datos estructurados, como aquellos encontrados en hojas de cálculo, bases de datos o archivos CSV. Este modelo utiliza atención múltiple para enfocarse en diferentes características de los datos, ver figura 2.

Los algoritmos de aprendizaje automático tradicionales, como los árboles de decisión, los bosques aleatorios o los métodos de potenciación de gradiente (por ejemplo, XGBoost, LightGBM) manejan características categóricas a través de la codificación one-hot o la codificación de etiquetas. Si bien estos métodos pueden funcionar bien, tienen limitaciones cuando se trata de capturar interacciones complejas entre características categóricas, especialmente en variables categóricas de alta cardinalidad (características con muchas categorías distintas).

Fig 2: Arquitectura simplificada del modelo TabTransformer.



Fuente: Elaboración propia.

La idea clave de TabTransformer es utilizar una arquitectura de transformador (un tipo de red neuronal diseñada originalmente para tareas de procesamiento de lenguaje natural, como BERT, GPT, etc.) para generar incrustaciones para las variables analizadas (Figura 2).

Se considera una tasa de aprendizaje de  $3 \times 10^{-5}$ , un optimizador Adam y 60 épocas de entrenamiento. Asimismo, se usó un tamaño de embedding de 1024 elementos y se usaron 2 capas de codificador y 2 cabezas.

### Características del lenguaje

Con la finalidad de hacer el modelo más explicativo, se proponen cuatro características de lenguaje observadas

en sujetos con EA. A continuación, se describen cada una de ellas.

### Palabras clave

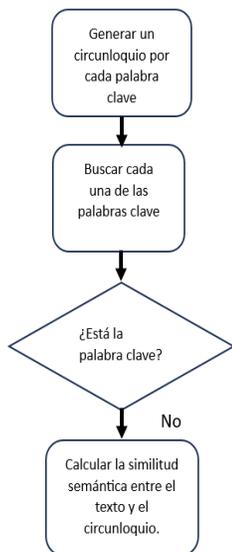
Dado que los sujetos con EA suelen presentar dificultades para encontrar la palabra correcta o mencionan oraciones inconclusas (Can & Kuruoğlu, 2019), se buscaron en los textos una serie de 34 palabras propuestas por Mackenzie et al. (2007). Cabe mencionar que las palabras fueron extraídas con base en investigaciones lingüísticas realizadas por expertos en afasia y geriatría.

El proceso consistió en contar el número de palabras clave encontradas en cada una de las muestras de texto. Se espera que los pacientes con enfermedad de Alzheimer usen, en promedio, una menor cantidad de palabras clave en la descripción de la imagen del robo de galletas.

### Circunloquios

El uso de circunloquios es una característica del lenguaje en sujetos con EA. Esta característica es una figura retórica o técnica lingüística que consiste en expresar una idea de manera indirecta, utilizando más palabras de las necesarias para evitar una expresión directa o sencilla. En lugar de decir algo de forma breve y clara, se utilizan varias palabras para rodear el concepto o idea que se quiere transmitir, ver figura 3.

Fig 3: Diagrama de flujo para el análisis de los circunloquios.



Fuente: Elaboración propia.

El proceso para tratar de detectar esta característica en los sujetos consistió en generar circunloquios para cada una de las palabras clave mostradas en el punto anterior (Figura 3). Una vez generados los circunloquios, se procede a realizar la búsqueda de cada una de las palabras

clave, en caso de que alguna palabra clave no se encuentre, se compara semánticamente la descripción del sujeto con el circunloquio y se calcula la similitud coseno. El modelo utilizado para calcular la similitud semántica fue SBERT (Reimers & Gurevych, 2019).

### Complejidad del texto

Los sujetos con EA tienden a utilizar oraciones con baja complejidad sintáctica. Por esta razón, se analizan cada uno de los textos de los sujetos con y sin deterioro cognitivo usando el puntaje de facilidad de lectura de Flesch. Esta métrica fue desarrollada por el lingüista Rudolf Flesch (1948) y es ampliamente utilizado en la escritura y edición para asegurar que los textos sean accesibles y apropiados para su audiencia objetivo.

El puntaje de facilidad de lectura de Flesch se basa en dos factores: el promedio de sílabas por palabra (PSP) y el promedio de palabras por oración (PPO). A continuación, se describe la fórmula para obtener este valor.

$$\text{Puntaje Flesch} = 206.8 - (1.015 * PPO) - (84.6 * PSP) \quad (2)$$

El puntaje de Flesch varía entre 0 y 100, y cuanto mayor sea el puntaje, más fácil será leer el texto.

### Similitud con texto ejemplo

Como se ha observado en trabajos previos (Can & Kuruoğlu, 2019), los sujetos con EA en etapas tempranas ven afectadas principalmente características semánticas y pragmáticas en el lenguaje. Por ello, se generó una muestra ejemplo de una descripción correcta de la imagen del robo de galletas usando ChatGPT (Achiam et al., 2023) para compararla con cada una de las muestras de los sujetos. Se espera que exista una mayor similitud, en promedio, con los sujetos de control que con los sujetos con enfermedad de Alzheimer. El puntaje de similitud coseno obtenido entre la muestra correcta de ejemplo y cada una de los textos fue el considerado como característica para la clasificación. Al igual que para el caso de los circunloquios, el modelo utilizado para analizar la similitud semántica es SBERT (Reimers & Gurevych, 2019).

## RESULTADOS Y DISCUSIÓN

Se siguen dos estrategias para probar la metodología propuesta. La primera implica probar el clasificador sin considerar las características del lenguaje propuestas, es decir, considerando únicamente la vectorización de las oraciones. Como segunda estrategia, se añade las características propuestas para evaluar el desempeño del modelo.

Se usan las métricas comúnmente utilizadas en estudios similares: exactitud (ecuación 3), precisión (ecuación 4), exhaustividad (ecuación 5) y puntuación F1 (ecuación 6). Estas métricas se definen en términos de los valores estimados correctamente, es decir, los verdaderos positivos

(VP) y los verdaderos negativos (VN), así como las predicciones incorrectas, es decir, los falsos positivos (FP) y los falsos negativos (FN).

$$\text{Exactitud} = \frac{VP+VN}{VP+VN+FP+FN} \quad (3)$$

$$\text{Precisión} = \frac{VP}{VP+FP} \quad (4)$$

$$\text{Exhaustividad} = \frac{VP}{VP+FN} \quad (5)$$

$$\text{Puntaje F1} = 2 * \frac{\text{Precisión} + \text{Exhaustividad}}{\text{Precisión} + \text{Exhaustividad}} \quad (6)$$

También se utiliza validación cruzada k-fold. Para probar la eficacia del clasificador, se considera un valor de k=5. Los datos se dividen en cinco pliegues, dividiéndolos en entrenamiento y prueba, usando 80% de los datos para entrenamiento y 20% para probar el modelo. Se ejecutan los modelos 10 veces y se reportan los resultados promedio. La Tabla 1 muestra los resultados para la primera estrategia, es decir, considerando únicamente las vectorizaciones de oraciones como característica para realizar la clasificación. La Tabla 2 contiene los resultados de los diferentes clasificadores analizados considerando el modelo de vectorización de palabras que logró un mejor desempeño.

Tabla 1. Resultados obtenidos para diferentes modelos de vectorización de oraciones.

Vectorización	Exactitud	Precisión	Exhaustividad	Puntaje F1
BERT (Devlin et al., 2018) + SVM	0.8527	0.8539	0.8529	0.8527
RoBERTa (Liu et al., 2019) + SVM	0.8650	0.8659	0.86525	0.8649
Llama 3 8B + LR	0.7811	0.7819	0.7808	0.78081
Mistral Nemo + LR	0.8036	0.8042	0.8036	0.8035
TF-IDF+SVM	0.8465	0.8489	0.8468	0.8463

Fuente: Elaboración propia.

Tabla 2. Resultados obtenidos para los tres modelos de clasificación analizados considerando el mejor modelo para la vectorización de los textos (RoBERTa model).

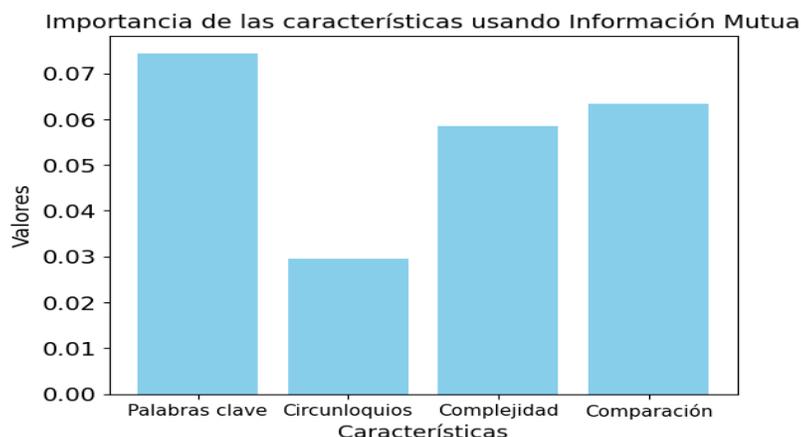
Vectorización	Exactitud	Precisión	Exhaustividad	Puntaje F1
Regresión Logística	0.8507	0.8518	0.8508	0.8506
TabTransformer	0.8527	0.8533	0.8527	0.8527
SVM	0.8650	0.8659	0.86525	0.8649

Fuente: Elaboración propia.

Los resultados de la Tabla 1 revelan que los modelos de solo encoder (BERT y RoBERTa) tienen un mejor desempeño que los modelos de sólo decoder. Para el caso de la Tabla 2, se observa como el clasificador SVM tuvo el mejor desempeño. Asimismo, Regresión Logística y TabTransformer tienen un desempeño similar; no obstante, es preciso señalar que el modelo TabTransformer fue un clasificador que requiere más tiempo en su entrenamiento.

Posteriormente, como segunda estrategia se realiza la clasificación considerando las características propuestas. Antes de ello, se analiza la importancia de cada uno de estos elementos (Figura 4). La importancia de estas características se obtuvo mediante Información Mutua empleando la paquetería Scikit-learn (Pedregosa et al., 2011) en Python.

Fig 4: Análisis de la importancia de las características propuestas usando Información Mutua.



Fuente: Elaboración propia.

Como se observa en la figura anterior (Figura 4), las palabras clave y la comparación con el texto ejemplo son las características más importantes para el clasificador. Ambos elementos tuvieron un puntaje mayor para los sujetos control. Por otro lado, tanto la complejidad del texto como los circunloquios fueron una característica mayor para los sujetos con EA.

Una vez agregadas estas características al clasificador se obtienen mejores resultados (Tabla 3).

Tabla 3. Resultados obtenidos considerando los embebidos de las oraciones más las características de lenguaje propuestas.

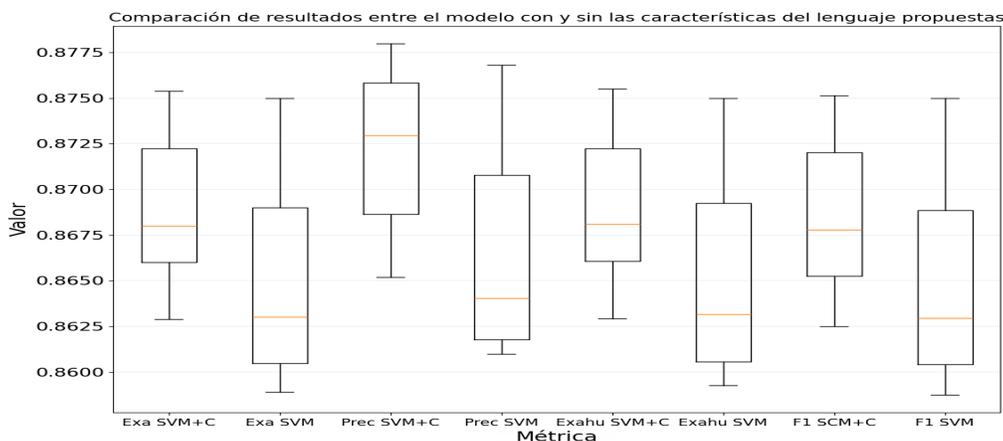
Modelo	Exactitud	Precisión	Exhaustividad	Puntaje F1
RoBERTa +SVM	0.8650	0.8659	0.86525	0.8649
RoBERTa + SVM + Características	0.8712	0.8729	0.8712	0.8709

Fuente: Elaboración propia.

La Tabla 3 muestra una mejora en los resultados una vez que se añaden las características de lenguaje propuestas. Asimismo, se realizan diagramas de caja para cada una de las métricas consideradas para analizar el comportamiento de las mismas en las 10 iteraciones efectuadas para probar los modelos.

La Figura 5 muestra que al incorporar las características del lenguaje propuestas se obtuvo una reducción en la varianza del modelo.

Fig 5: Diagramas de caja para las métricas analizadas.



Fuente: Elaboración propia.

## CONCLUSIONES

El análisis de las características semánticas y sintácticas revela que las palabras clave y la comparación con un texto ejemplo son los factores más determinantes para distinguir entre sujetos sanos y aquellos con enfermedad de Alzheimer. Esto sugiere que las alteraciones en el lenguaje son más evidentes en etapas tempranas o moderadas de la enfermedad.

Los circunloquios y la complejidad del texto muestran poca relevancia en la clasificación, lo que puede indicar que estas características del lenguaje no están tan influenciadas por la enfermedad de Alzheimer y pueden estar más relacionadas con factores como el nivel educativo.

Los modelos de tipo encoder, como RoBERTa, demuestran un mejor rendimiento en tareas de clasificación de texto en comparación con los modelos decoder. El uso de modelos comprimidos para ejecutar Llama 3 y Mistral Nemo localmente pudo haber afectado los resultados, pero aun así muestran que los modelos más pequeños pueden ser efectivos en tareas de clasificación semántica.

Finalmente, como trabajo futuro se espera investigar nuevas características del lenguaje de sujetos, como las ideas incompletas o la incorporación de elementos del audio como las pausas o la duración del discurso. Asimismo, resulta pertinente probar nuevos modelos de vectorización de palabras para observar y analizar su desempeño en la búsqueda de patrones sintácticos y semánticos en los textos de los pacientes que sufren enfermedad de Alzheimer.

## REFERENCIAS BIBLIOGRÁFICAS

- Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & McGrew, B. (2023). Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*. <https://doi.org/10.48550/arXiv.2303.08774>
- Becker, J. T., Boiler, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of neurology*, *51*(6), 585-594. <https://doi.org/10.1001/archneur.1994.00540180063015>
- Can, E., & Kuruoğlu, G. (2019). Language Changes in Late-Onset Alzheimer's Disease. *Psycholinguistics*, *25*(2), 50-68. <https://doi.org/10.31470/2309-1797-2019-25-2-50-68>
- Deary, I. J., Corley, J., Gow, A. J., Harris, S. E., Houlihan, L. M., Marioni, R. E., ... & Starr, J. M. (2009). Age-associated cognitive decline. *British medical journal*, *339*(7728), 135-152. <https://doi.org/10.1093/bmb/ldp041>

- Devlin, J. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. [https://bibbase.org/service/mendeley/bfbfbf840-4c42-3914-a463-19024f50b30c/file/6375d223-e085-74b3-392f-f3fed829cd72/Devlin et al 2019 BERT Pre training of Deep Bidirectional Transform.pdf.pdf](https://bibbase.org/service/mendeley/bfbfbf840-4c42-3914-a463-19024f50b30c/file/6375d223-e085-74b3-392f-f3fed829cd72/Devlin%20et%20al%202019%20BERT%20Pre%20training%20of%20Deep%20Bidirectional%20Transform.pdf.pdf)
- Dubey, A., Jauhri, A., Pandey, A., Kadian, A., Al-Dahle, A., Letman, A., ... & Ganapathy, R. (2024). The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*. <https://doi.org/10.48550/arXiv.2407.21783>
- Flesch, R. (1948). A new readability yardstick. *Journal of applied psychology*, *32*(3), 221-223. <https://psycnet.apa.org/doi/10.1037/h0057532>
- Huang, X., Khetan, A., Cvitkovic, M., & Karnin, Z. (2020). Tabtransformer: Tabular data modeling using contextual embeddings. *arXiv preprint arXiv:2012.06678*. <https://doi.org/10.48550/arXiv.2012.06678>
- Hussein, K. I., Chan, L., Van Vleck, T., Beers, K., Mindt, M. R., Wolf, M., ... & Federman, A. (2022). Natural language processing to identify patients with cognitive impairment. *medRxiv*, 2022-02. <https://doi.org/10.1101/2022.02.16.22271085>
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). RoBERTa: A robustly optimized BERT pretraining approach. *arXiv [Preprint]* (2019). *arXiv preprint arXiv:1907.11692*. <https://doi.org/10.48550/arXiv.1907.11692>
- Mackenzie, C., Brady, M., Norrie, J., & Poedjianto, N. (2007). Picture description in neurologically normal adults: Concepts and topic coherence. *Aphasiology*, *21*(3-4), 340-354. <https://doi.org/10.1080/02687030600911419>
- MacWhinney, B. (2003). Language acquisition. *The handbook of brain theory and neural networks*, 600-603. <https://psyling.talkbank.org/years/2002/arbib.pdf>
- Matošević, L., & Jović, A. (2022). Accurate detection of dementia from speech transcripts using Roberta model. In *2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 1478-1484). IEEE. <https://doi.org/10.23919/MIPRO55190.2022.9803462>
- Nazi, Z. A., & Peng, W. (2024, August). Large language models in healthcare and medical domain: A review. *Informatics*, *11* (3). MDPI. <https://doi.org/10.3390/informatics11030057>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *The Journal of machine Learning research*, *12*, 2825-2830. <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https://>

- Penfold, R. B., Carrell, D. S., Cronkite, D. J., Pabiniak, C., Dodd, T., Glass, A. M., ... & Stang, P. E. (2022). Development of a machine learning model to predict mild cognitive impairment using natural language processing in the absence of screening. *BMC Medical Informatics and Decision Making*, 22(1), 129. <https://doi.org/10.1186/s12911-022-01864-z>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *arXiv preprint arXiv:1908.10084*. <https://doi.org/10.48550/arXiv.1908.10084>