

Tipificación de frases relevantes de entrevistas de valoración en el marco de la evaluación psicológica

Typification of relevant phrases from assessment interviews within the framework of psychological assessment
Tipificação de frases relevantes em entrevistas de valoraçãõ no marco da avaliação psicológica



Elio Leonardo **Pardo Jaime**
Jorge Mario **Calvo Londoño**

New Africa

Photo By/Foto:

Rip
16²

Volumen 16 #2 may-ago
16 Años

ID: [10.33881/2027-1786.rip.16203](https://doi.org/10.33881/2027-1786.rip.16203)

Title: Typification of relevant phrases from assessment interviews within the framework of psychological assessment

Título: Tipificação de frases relevantes em entrevistas de avaliação no marco da avaliação psicológica

Título: Tipificación de frases relevantes de entrevistas de valoración en el marco de la evaluación psicológica

Alt Title / Título alternativo:

[en]: Typification of relevant phrases from assessment interviews within the framework of psychological assessment

[pt]: Tipificação de frases relevantes em entrevistas de avaliação no marco da avaliação psicológica

[es]: Tipificación de frases relevantes de entrevistas de valoración en el marco de la evaluación psicológica

Author (s) / Autor (es):

Pardo Jaime & Calvo Londoño

Keywords / Palabras Clave:

[en]: Psychological assessment, interview, machine learning, support vector machines, Naive Bayes

[pt]: Avaliação psicológica, entrevista, aprendizado de máquina, máquinas de vetores de suporte, Naive Bayes

[es]: Evaluación psicológica, entrevista, aprendizaje de máquina, máquinas de soporte vectorial, Naive Bayes

Submitted: 2022-06-23

Accepted: 2023-01-13

Resumen

La evaluación se considera como una disciplina científica dentro del quehacer del profesional en psicología, donde la entrevista es fundamental para aumentar la probabilidad de éxito en este proceso; no obstante, este análisis depende del criterio del profesional, lo cual tiende a generar un sesgo en el análisis de la información recolectada. Por lo cual, el objetivo de esta investigación fue tipificar frases relevantes para el análisis de entrevistas de valoración. Para esto se realizó un juicio de expertos para identificar las frases relevantes en 10 entrevistas, que se obtuvieron al comparar la selección de los tres jueces, estableciendo como parámetro una similitud del 75%, luego se clasificó las oraciones de los textos originales en relevantes y no relevantes para poder entrenar dos modelos de aprendizaje. Los resultados reflejaron que el modelo Naive Bayes presenta mejores indicadores para la tipificación de frases en comparación con el modelo de máquinas de soporte vectorial (SVM)., además, se planteó un proceso metodológico para analizar entrevistas utilizando algoritmos de machine learning.

Resumo

A avaliação é considerada uma disciplina científica dentro do trabalho do profissional em psicologia, onde a entrevista é fundamental para aumentar a probabilidade de sucesso neste processo; porém, essa análise depende do critério do profissional, o que tende a gerar um viés na análise das informações coletadas. Portanto, o objetivo desta pesquisa foi classificar frases relevantes para análise de entrevistas de avaliação. Para isso, foi realizado um parecer pericial para identificar as sentenças relevantes em 10 entrevistas, as quais foram obtidas comparando a seleção dos três juizes, estabelecendo como parâmetro uma similaridade de 75%, em seguida as sentenças dos textos originais foram classificadas como relevante e não relevante para poder treinar dois modelos de aprendizagem. Os resultados mostraram que o modelo Naive Bayes apresenta melhores indicadores para digitação de sentenças comparado ao modelo Support Vector Machine (SVM), além disso, foi proposto um processo metodológico para análise de entrevistas utilizando algoritmos de aprendizado de máquina

Abstract

The evaluation is considered as a scientific discipline within the work of the professional in psychology, where the interview is fundamental to increase the probability of success in this process; however, this analysis depends on the professional's criteria, which tends to generate a bias in the analysis of the information collected. Therefore, the objective of this research was to classify relevant phrases for the analysis of assessment interviews. For this, an expert judgment was carried out to identify the relevant phrases in 10 interviews, which were obtained by comparing the selection of the three judges, establishing a similarity of 75% as a parameter, then the sentences of the original texts were classified as relevant and not relevant to be able to train two learning models. The results showed that the Naive Bayes model presents better indicators for sentence typing compared to the support vector machine (SVM) model. In addition, a methodological process was proposed to analyze interviews using machine learning algorithms.

Citar como:

Pardo Jaime, E. L. & Calvo Londoño, J. M. (2023). Tipificación de frases relevantes de entrevistas de valoración en el marco de la evaluación psicológica. *Revista Iberoamericana de Psicología*, 16 (2), 23-32. Obtenido de: <https://reviberopsicologia.iberu.edu.co/article/view/2420>

Elio Leonardo **Pardo Jaime**, MA Psi
ORCID: [0000-0002-5250-4812](https://orcid.org/0000-0002-5250-4812)

Source | Filiación:
Universidad Central

BIO:
Psicólogo, Magister en analítica de datos

City | Ciudad:
Bogotá DC [co]

e-mail:
epardoj@ucentral.edu.co

Dr Jorge Mario **Calvo Londoño**, Dr MAEsp Ing

Source | Filiación:
Universidad Central

BIO:
Ingeniero de sistemas, Especialista en telemática, Magister en ingeniería de sistemas, Doctor en informática

City | Ciudad:
Bogotá DC [co]

e-mail:
jcalvoc@ucentral.edu.co

Tipificación de frases relevantes de entrevistas de valoración en el marco de la evaluación psicológica

Typification of relevant phrases from assessment interviews within the framework of psychological assessment

Tipificação de frases relevantes em entrevistas de avaliação no marco da avaliação psicológica

Elio Leonardo **Pardo Jaime**
Jorge Mario **Calvo Londoño**

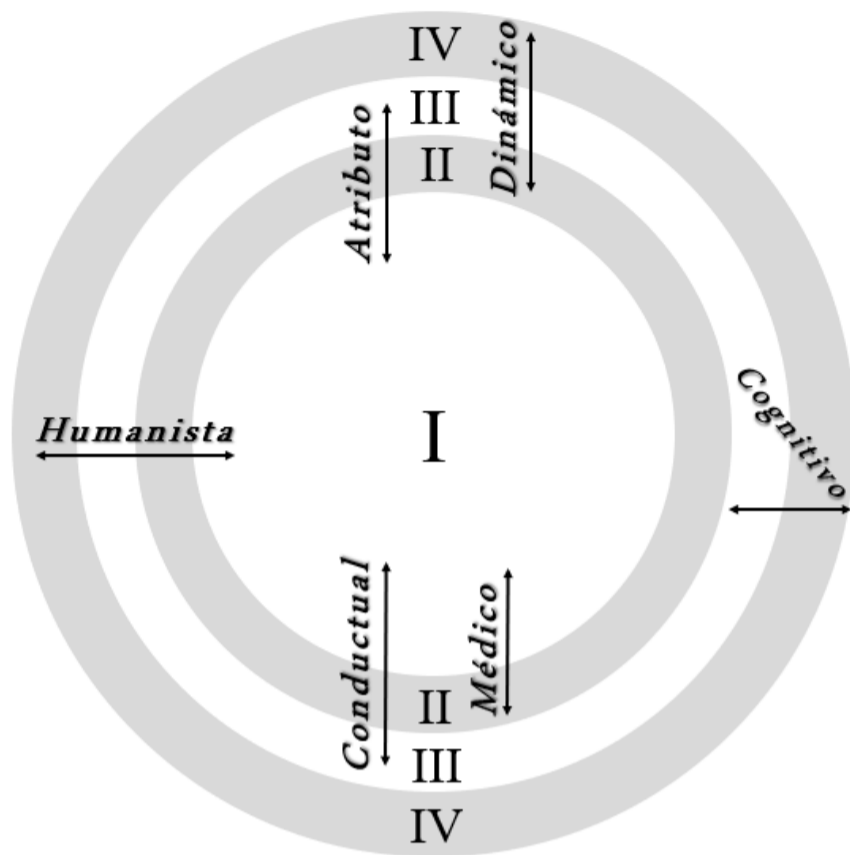
Marco Teórico

Evaluación Psicológica

La evaluación se considera como una disciplina científica dentro del quehacer del profesional en psicología, cuyo objetivo es establecer la posible predicción de un comportamiento y la planificación de una intervención, fundamentada en la comprobación de hipótesis, si bien ésta suele relacionarse con las actividades desarrolladas en el campo clínico, se considera como una función integral para todos los profesionales, la cual se despliega en tres grandes fases: recolección de información preliminar, desarrollo y contraste de hipótesis, y programación de la intervención (Jorge, 2018; Muñoz et al., 2019).

Ahora bien, este proceso se gestiona desde varios paradigmas de la psicología; específicamente suele implementarse con base en 6 modelos de evaluación (atributo, dinámico, médico, conductual, cognitivo y humanista) lo cual implica que las hipótesis y los objetivos del proceso varíen según el constructo seleccionado. Otra característica de esta variabilidad consiste en los niveles de inferencia de sus procesos, donde algunos se centran más en la conducta manifiesta (nivel 1), y otros en una explicación especulativa de la conducta desde variables intrapsíquicas (nivel 4) (véase figura 1; Fernández-Ballesteros, 2013)

Figura 1. Distribución de los modelos de evaluación por los niveles de inferencia



Nota. Las flechas que están bajo los modelos de evaluación abordan los niveles de inferencia que se puede obtener en cada uno

Entrevista y procesamiento de lenguaje natural (NLP)

Por otra parte, el punto de partida de la evaluación es la entrevista, definida como el proceso donde el entrevistador – en este caso el psicólogo – le realiza preguntas al consultante (el entrevistado) para obtener información que le permita comprender la situación a evaluar. Es importante resaltar que este proceso no es aleatorio, por el contrario, es el entrevistador quien dirige las preguntas, controla los comportamientos (verbales y no verbales), la frecuencia y latencia de las intervenciones y los temas que se abordan, buscando que el entrevistado exponga sus emociones y las situaciones relevantes a los problemas que lo aquejan (Perpiña, 2012; Keats, 2009; Morrison, 2015)

Dado que en la entrevista es donde se establece la relación terapéutica, los objetivos, prioridades del tratamiento y se toman las decisiones clínicas relevantes, se considera el proceso más importante de la evaluación; no obstante, el análisis de la información recolectada depende del paradigma que practique el psicólogo; y es esto, junto con la precariedad de los sistemas de información clínica, son los principales obstáculos para obtener los datos más relevantes de cada caso con el menor sesgo posible (Bellack y Hersen, 1990; Camacho et al., 2013; Llavona, 2011)

Ahora bien, estas dificultades pueden ser solucionadas al utilizar el procesamiento de lenguaje natural, considerada como una rama de la inteligencia artificial (IA) que permite a la máquina comprender textos redactados en lenguaje coloquial (natural) para así poder analizarlos posteriormente. Su utilidad se ve reflejada en diversas aplicaciones tales como el procesamiento de textos, la traducción automática y las interfases de lenguaje natural (Borja, 2020; Cortez et al. 2009; De la Calle, 2014)

El procesamiento de lenguaje natural (NLP por sus siglas en inglés) suele trabajar en distintos componentes para analizar los textos, con el propósito de facilitar las tareas de reconocimiento y extracción de la información relacionada. El primero corresponde al análisis léxico, el cual examina la morfología de los párrafos o las frases redactadas, intentando establecer los separadores o su categoría gramatical (verbo, sustantivo, entre otros); el análisis sintáctico establece si se cumple una estructura gramatical; el análisis semántico pretende establecer el significado de las secuencias generadas en el análisis sintáctico; y por último, el análisis pragmático evalúa la relación entre palabras y el contexto donde son utilizadas (Cortez et al., 2009; De la Calle, 2014; Kadlaskar, 2021; Pajaro, 2018)

En cuanto a la eficacia de la aplicación de procesamiento de lenguaje natural en salud mental se ha encontrado su implementación en investigación cualitativa para determinar la relación médico paciente con enfermedad mental, en la recuperación de registros de asesoría psiquiátrica, para hacer aplicaciones que generen una retroalimentación a pacientes, en sistemas conversacionales y para extraer información de publicaciones científicas, evidenciando usos útiles para obtener datos que, de forma tradicional, no sería posible (Aparicio et al. 2011; Babativa, 2021; Camacho et al., 2013; de la Calle, 2014; Iglesias et al., 2008; Sancho et al., 2020)

Ahora bien, para los propósitos de esta investigación – en el cual se desarrollará un ejercicio de clasificación – algunos autores reportan que los algoritmos más utilizados para este ejercicio son el Naive Bayes y la máquina de soporte vectorial (SMV por sus siglas en inglés) donde, algunos indican que el modelo más efectivo es el primero (Florián, 2013); mientras que otros han reportado el segundo como el de mayor efectividad (Páramo y Espitia, 2018; Singh, 2021) Tomando en cuenta la utilidad del procesamiento de lenguaje natural para el análisis de textos ¿Cómo implementar el procesamiento de lenguaje natural en la recolección de información basado en entrevistas de la evaluación psicológica? Es así como esta investigación pretendió tipificar las frases relevantes para el análisis de entrevistas de valoración clínica.

Método

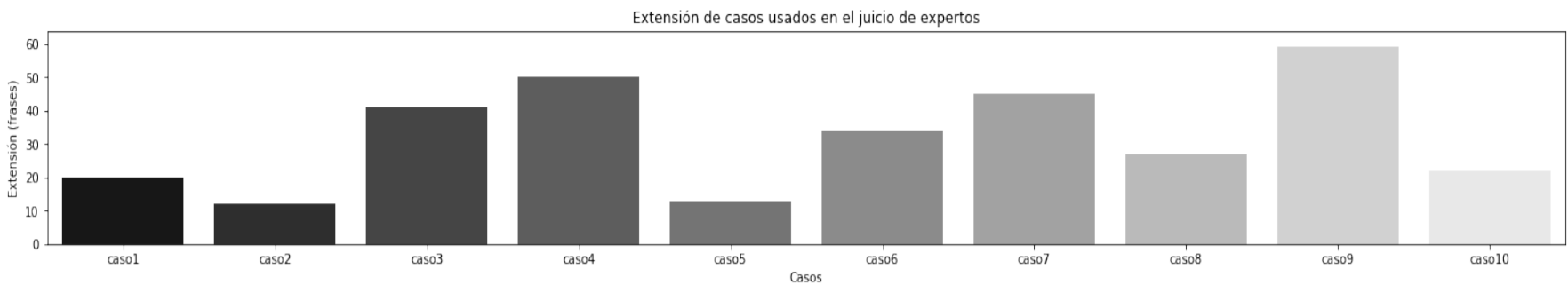
Tipo De Estudio.

La presente investigación se estableció de corte instrumental, esto debido a que se busca adaptar una herramienta que permita tipificar frases relevantes para analizar entrevistas de valoración en un contexto de evaluación psicológica (Montero, & León, 2005).

Muestra.

El desarrollo de esta investigación contó con un muestreo no aleatorio a conveniencia, donde se eligieron 10 entrevistas publicadas en un artículo y un libro, cuya temática es informar trabajos relacionados con la psicología clínica y de la salud. Dado que el análisis se realizó con base en las frases que componen el texto (de punto a punto), se estableció que las 10 entrevistas contaban con 323 frases, donde la más corta tiene 12 frases y la más larga 59 (véase figura 2), además, se utilizaron 3 entrevistas (107 frases que no fueron usadas en el entrenamiento) para la verificación del modelo. Todos los archivos se encuentran disponible en Pardo (2023)

Figura 2.
Frecuencia de frases de casos clínicos.



Procedimiento.

El desarrollo de esta investigación fue realizado en tres fases: la extracción de información, el juicio de expertos y entrenamiento de modelos.

Extracción De Información.

Uno de los primeros pasos fue extraer las entrevistas; para lo cual se realizó la búsqueda de artículos de caso único publicados en revistas indexadas entre el 2010 y el 2020; donde se revisó de forma manual cada uno para identificar los apartados que contenían las entrevistas iniciales, y luego copiar y pegar éstas en un archivo de Excel.

Juicio De Expertos.

Para establecer la relevancia de las frases, se solicitó a tres jueces expertos (un psicólogo clínico, un psicólogo jurídico y un psicólogo educativo) que leyeran y extrajeran las frases que consideraran relevantes en cada caso; cuyo registro se realizó en un Google forms donde se presentó el caso y se habilitó un espacio para registrar las frases. Una vez hecha esa evaluación, se verificó la concordancia de las frases seleccionadas por cada juez, haciendo la limpieza, tokenización y lematización de los textos (Alghani, 2020; Ma, 2018; Teja, 2020) y valorando la similitud entre las frases a través del método similarity de la librería Spacy, donde se consideró que una similitud igual o mayor al 75% correspondería a un acuerdo entre los jueces, esto basado en el criterio de interpretación utilizado en la V de Aiken; donde se considera que un valor mayor a 0.7 representa un nivel de concordancia riguroso para la evaluación por jueces (Charter, 2003; citado por Merino y Livia, 2009). Finalmente, se compararon las frases que tuvieron concordancia con las originales, donde se estableció que una similitud igual o mayor a 75% correspondía a una frase relevante; en caso contrario, se clasificó como irrelevante (véase figura 3)

Entrenamiento De Modelos.

Una vez obtenidas las frases relevantes, se clasificaron las oraciones de las 10 entrevistas, obteniendo 108 relevantes y 215 irrelevantes. Dado que los datos eran pocos, se utilizó la librería nlpaug, la cual permite generar datos sintéticos basados en sinónimos, y así aumentar las frases disponibles a 756 frases relevantes y 1505 frases irrelevantes. Con el 75% de los datos (1695 frases) se entrenaron dos modelos de aprendizaje supervisado sin modificación de hiperparámetros (Naive Bayes Multinomial y máquinas de soporte vectorial (SVM); véase tabla 1), vectorizándolos con el método CountVectorizer de la librería

Scikit-Learn, y con el 25% restante (565 frases) se extrajo la matriz de confusión y el reporte de clasificación para comprobar los indicadores de calidad de los modelos. Por último, con las 107 frases se verificó la precisión del modelo. Para el análisis de información se usaron librerías del lenguaje de programación Python 3.8.8 a través del entorno de desarrollo integrado (IDE) Jupyter Notebook.

Figura 3.
Proceso de clasificación de frases relevantes e irrelevantes.

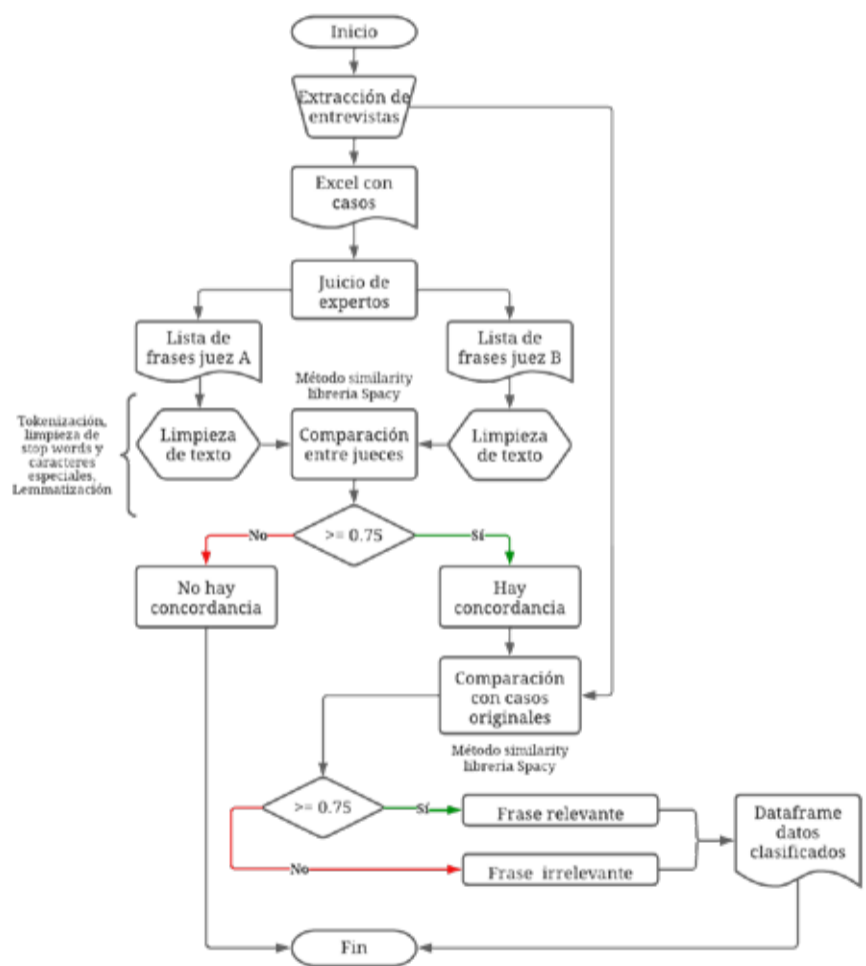


Tabla 1.
Valores de hiperparámetros de los modelos Naive Bayes y SVM.

Naive Bayes	SVM	
alpha: 1.0	C: 1.0,	max_iter: 1000,
class_prior: None	class_weight: balanced,	multi_class: ovr,
fit_prior: True	dual: True,	penalty: l2,
	fit_intercept: True,	random_state: None,
	intercept_scaling: 1,	tol: 0.0001,
	loss: squared_hinge,	verbose: 0

Consideraciones Éticas.

Para el desarrollo de esta investigación se basó en la resolución 8430 de 1993, la cual establece que es una investigación sin riesgo al utilizar documentos retrospectivos, como lo son las historias clínicas y las entrevistas; y donde no hay ninguna manipulación de variables biológicas. Por su parte, el artículo 14 parágrafo 1 de la resolución 1995 de 1999 el cual indica que debe mantenerse la reserva legal en el momento de acceder a la historia clínica; el ítem h del artículo 4 de la ley 1581 de 2012 que estipula el principio de confidencialidad como principio para realizar el tratamiento de datos personales.

También se consideró la ley 1090 del 2006 artículo 2 ítem 5 que habla del principio de confidencialidad, ítem 9 y el artículo 50 que estipula el principio al respeto de la dignidad y el bienestar de las personas, por el código de ética de la asociación americana de psicología (APA, por sus siglas en ingles) en el apartado 4.07 en la que se indica que los psicólogos no revelarán información individualmente identificable relativa los destinatarios de sus servicios y la doctrina N° 1 del colegio colombiano de psicólogos que dictamina el uso de la historia clínica

como instrumento de investigación siempre que se cumplan con las condiciones establecidas en la ley 1090 y el código de ética de la APA.

Resultados

Juicio De Expertos

Para establecer la concordancia de los 10 casos se compararon las frases extraídas por los tres jueces en pares (juez 1 y juez 2; juez 1 y juez 3; juez 2 y juez 3) obteniendo una matriz de porcentajes de similitud por cada caso, como la observada en la tabla 2, 3 y 4. Una vez obtenidas la concordancia entre las frases (=> 75%), se compararon con los casos originales, obteniendo 108 frases relevantes y 215 frases irrelevantes. Algunos ejemplos pueden observarse en la tabla 4.

Tabla 2.
Matriz de comparación entre el juez 1 y 2 (caso 1).

Similitud de jueces	Frases juez 2											
	0	1	2	3	4	5	6	7	8	9	10	
Frases juez 1	0	1	0.36	0.29	0.76	0.47	0.41	0.65	0.47	0.37	0.66	0.44
	1	0.49	0.29	0.19	0.58	0.25	0.19	0.51	0.23	0.38	0.54	0.44
	2	0.36	1	0.48	0.41	0.63	0.24	0.55	0.15	0.46	0.68	0.18
	3	0.39	0.49	0.17	0.53	0.33	0.24	0.59	0.21	0.66	0.54	0.16
	4	0.12	0.25	0.18	0.22	0.08	0.08	0.16	0.1	0.05	0.13	0.13
	5	0.13	0.14	0.07	0.05	0.03	-0.01	0.1	0.22	0.14	0.1	0.09
	6	0.5	0.4	0.25	0.52	0.39	0.2	0.55	0.37	0.43	0.48	0.43
	7	0.56	0.52	0.33	0.66	0.68	0.25	0.62	0.43	0.5	0.57	0.46
	8	0.69	0.42	0.38	0.55	0.62	0.28	0.61	0.41	0.27	0.55	0.44
	9	0.44	0.18	0.12	0.48	0.31	0.04	0.32	0.26	0.19	0.32	1

Tabla 3.
Matriz de comparación entre el juez 1 y 3 (caso 1).

Similitud de jueces	Frases juez 3																
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Frases juez 1	0	0.36	0.29	0.25	0.48	0.76	0.4	0.47	0.4	0.59	0.27	0.69	0.44	0.52	0.61	0.49	0.59
	1	0.29	0.19	0.24	0.34	0.58	0.18	0.23	0.28	0.34	0.15	0.37	0.44	0.52	0.54	0.35	0.48
	2	1	0.48	0.38	0.37	0.41	0.6	0.15	0.5	0.31	0.42	0.42	0.18	0.62	0.51	0.45	0.58
	3	0.49	0.17	0.39	0.39	0.53	0.22	0.21	0.25	0.36	0.28	0.3	0.16	0.69	0.58	0.46	0.61
	4	0.25	0.18	0.14	0.11	0.22	0.05	0.1	0.38	0.14	0.06	0.08	0.13	0.11	0.08	0.13	0.06
	5	0.14	0.07	0.1	0.16	0.05	0.01	0.22	0.11	0.16	0.01	0.12	0.09	0.13	0.16	0.14	0.1
	6	0.4	0.25	0.26	0.53	0.52	0.33	0.37	0.4	0.47	0.23	0.47	0.43	0.59	0.61	0.46	0.52
	7	0.52	0.33	0.36	0.6	0.66	0.65	0.43	0.55	0.53	0.3	0.61	0.46	0.65	0.65	0.61	0.59
	8	0.42	0.38	0.27	0.52	0.55	0.59	0.41	0.42	0.59	0.27	1	0.44	0.47	0.57	0.53	0.72
	9	0.18	0.12	0.17	0.45	0.48	0.27	0.26	0.32	0.44	0.08	0.44	1	0.31	0.41	0.38	0.25

Tabla 4.
Matriz de comparación entre el juez 2 y 3 (caso 1).

Similitud de jueces	Frases juez 3															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
0	0.36	0.29	0.25	0.48	0.76	0.4	0.47	0.4	0.59	0.27	0.69	0.44	0.52	0.61	0.49	0.59
1	1	0.48	0.38	0.37	0.41	0.6	0.15	0.5	0.31	0.42	0.42	0.18	0.62	0.51	0.45	0.58
2	0.48	1	0.12	0.2	0.25	0.36	0.14	0.35	0.27	0.19	0.38	0.12	0.26	0.21	0.25	0.38
3	0.41	0.25	0.3	0.56	1	0.39	0.26	0.34	0.46	0.32	0.55	0.48	0.67	0.65	0.58	0.62
4	0.63	0.39	0.23	0.49	0.5	0.97	0.23	0.33	0.36	0.43	0.62	0.31	0.59	0.54	0.59	0.7
5	0.24	0.24	0.21	0.2	0.4	0.25	0.12	0.17	0.26	0.34	0.28	0.04	0.28	0.25	0.12	0.32
6	0.55	0.37	0.3	0.52	0.67	0.52	0.39	0.43	0.55	0.42	0.61	0.32	0.7	0.65	0.52	0.73
7	0.15	0.14	0.19	0.34	0.26	0.24	1	0.32	0.56	0.13	0.41	0.26	0.32	0.37	0.35	0.28
8	0.46	0.07	0.39	0.47	0.49	0.33	0.23	0.25	0.29	0.29	0.27	0.19	0.69	0.7	0.46	0.49
9	0.68	0.41	0.36	0.52	0.71	0.57	0.24	0.35	0.4	0.42	0.55	0.32	0.71	0.71	0.48	0.71
10	0.18	0.12	0.17	0.45	0.48	0.27	0.26	0.32	0.44	0.08	0.44	1	0.31	0.41	0.38	0.25

Tabla 5.
Ejemplos de frases clasificadas como relevantes e irrelevantes.

caso	frases	relevancia
0	Fátima es una chica de 19 años, de origen marroquí.	Irrelevante
0	sin desaparecer la sensación de fortaleza que transmite, en esos momentos del «¡no sé!» también comunica que puede sentirse perdida y desorientada.	Relevante
1	no se quita la chaqueta en toda la entrevista.	Irrelevante
1	pienso que algo malo me va a pasar y si me corto pues ya me ha pasado algo malo... o sea que ya está justificado que tenga ese miedo, y que entonces ya no me va a pasar nada más».	Relevante

Evaluando los tokens de las oraciones categorizadas como irrelevantes, se evidencia que la mayoría de las oraciones usan las siguientes palabras: año, tiempo, pareja, querer y tenía (véase figura 4); mientras

que las relevantes muestran que las palabras él, madre, padre, sentir, pensar, llorar, pasar, dejar y explicar son las más utilizadas (véase figura 5).

Figura 4.
Nube de palabras de frases irrelevantes.

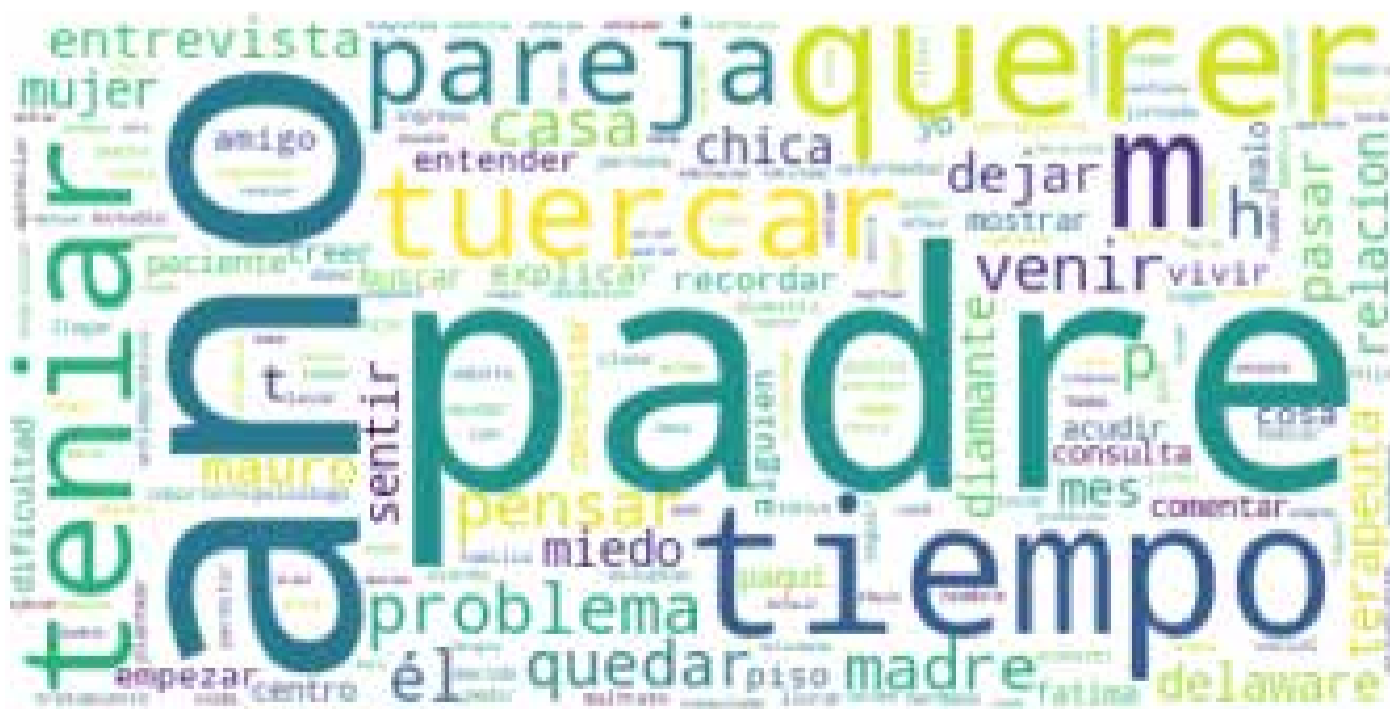
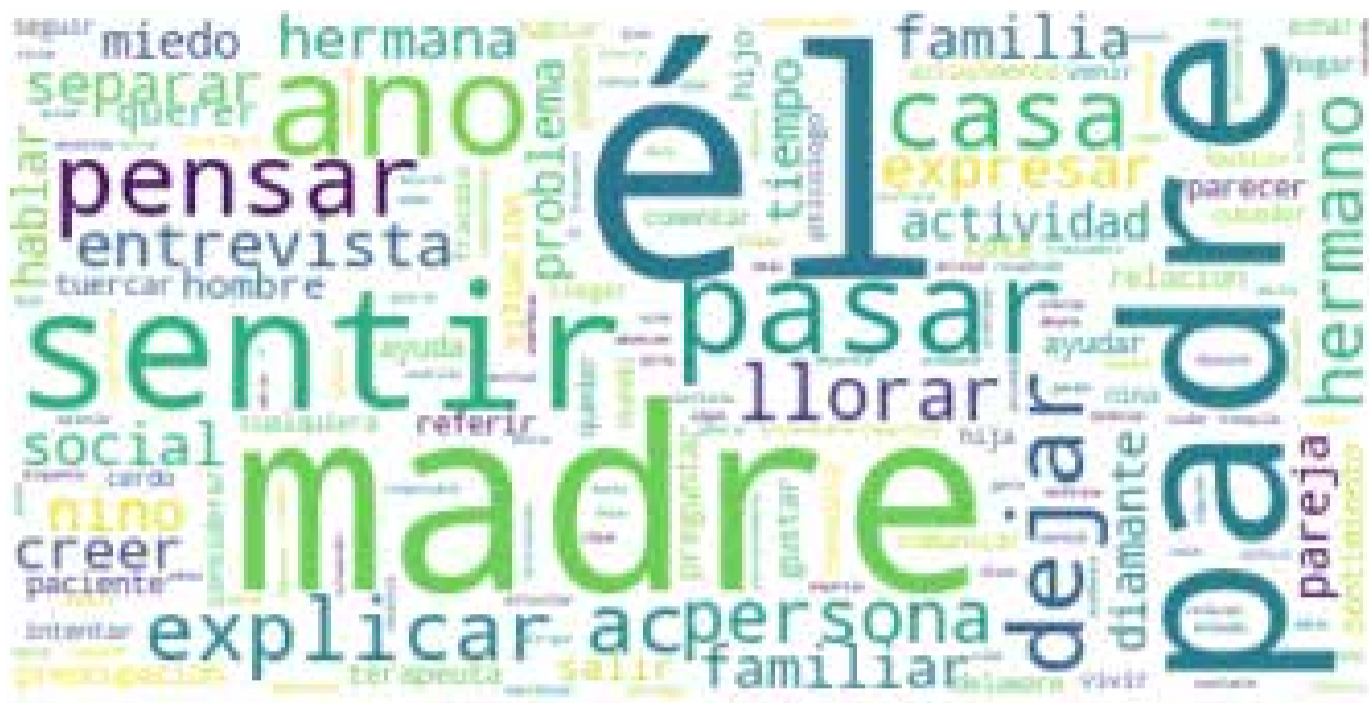


Figura 5.
Nube de palabras de frases relevantes.



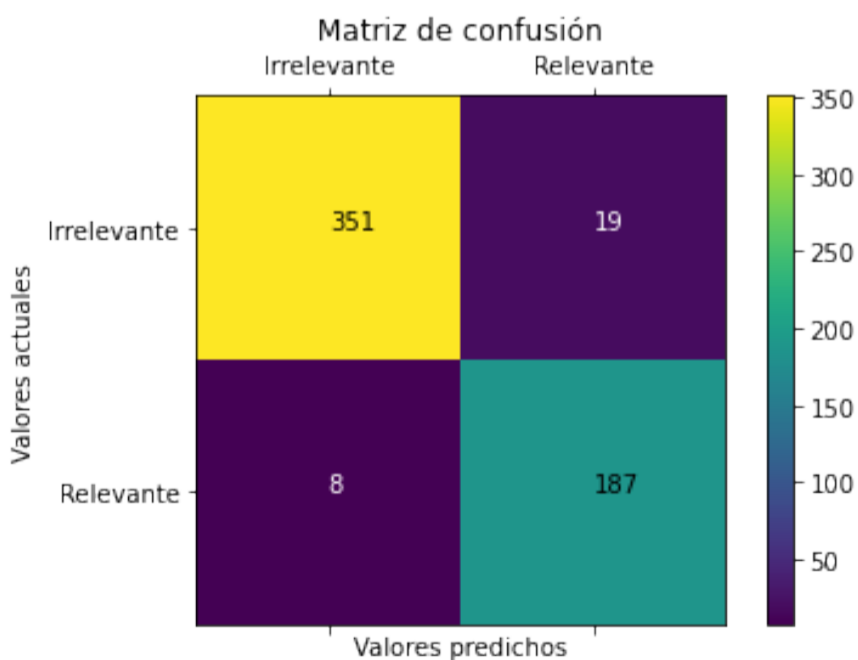
Entrenamiento de modelos

Modelo Naive Bayes

Matriz De Confusión.

Al extraer la matriz de confusión se obtuvo que el modelo identificó 538 frases verdaderas (diagonal de la matriz) y 27 frases falsas. Estos resultados pueden observarse en la figura 6.

Figura 6.
Matriz de confusión modelo Naive Bayes.



Reporte De Clasificación.

De acuerdo con lo observado en la tabla 6, el indicador de precisión para las frases irrelevantes es del 97.78%, la sensibilidad es del 94.86%, obteniendo una media armónica (f1-score) del 96.30%. Por otra parte, las frases relevantes evidenciaron una precisión del 90.78%, una sensibilidad del 95.90% y una media armónica del 93.27%. Finalmente, la precisión general del modelo (accuracy) fue del 95.22%.

Tabla 6.
Reporte de clasificación del modelo Naive Bayes.

	precision	recall	f1-score	support
Irrelevante	0.9777	0.9486	0.9630	370.0000
Relevante	0.9078	0.9590	0.9327	195.0000
accuracy	0.9522	0.9522	0.9522	0.9522
macro avg	0.9427	0.9538	0.9478	565.0000
weighted avg	0.9536	0.9522	0.9525	565.0000

Modelo de máquinas de soporte vectorial (SVM)

Matriz De Confusión.

Verificando esta matriz se evidencia que detecta 563 frases verdaderas y 2 falsas (véase figura 7).

Figura 7.
Matriz de confusión modelo SVM

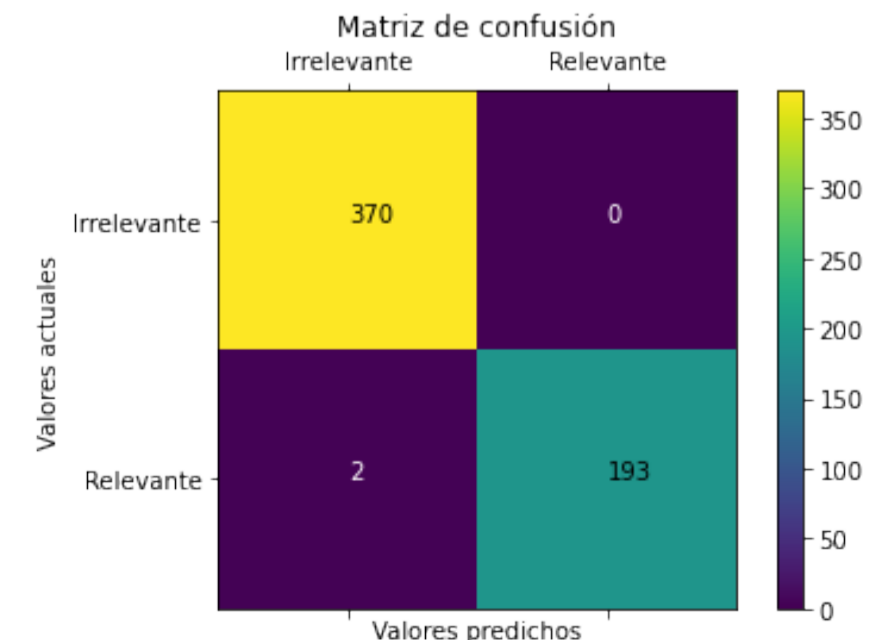


Figura 9.
Matriz de confusión en la verificación del modelo SVM

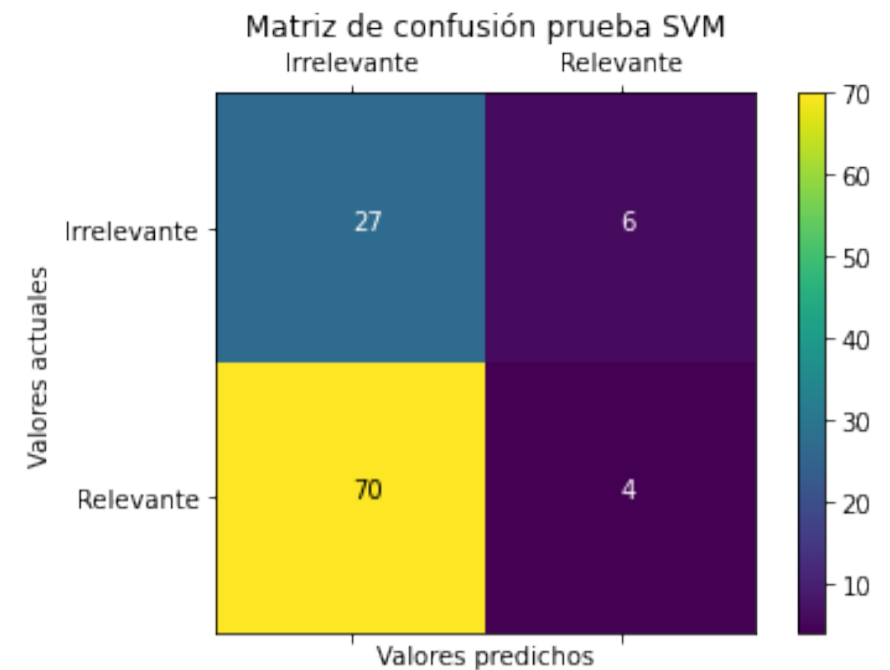


Tabla 8.
Reporte de clasificación en la verificación de los modelos Naive Bayes y SVM.

Modelo Naive Bayes				
	precision	recall	f1-score	support
Irrelevante	0.3065	0.5758	0.4000	33.0000
Relevante	0.6889	0.4189	0.5210	74.0000
accuracy	0.4673	0.4673	0.4673	0.4673
macro avg	0.4977	0.4973	0.4605	107.0000
weighted avg	0.5709	0.4673	0.4837	107.0000

Modelo SVM				
	precision	recall	f1-score	support
Irrelevante	0.2784	0.8182	0.4154	33.0000
Relevante	0.4000	0.0541	0.0952	74.0000
accuracy	0.2897	0.2897	0.2897	0.2897
macro avg	0.3392	0.4361	0.2553	107.0000
weighted avg	0.3625	0.2897	0.1940	107.0000

Conclusiones

Inicialmente, el juicio de expertos arrojó un total de 30 matrices que permitió identificar 185 frases relevantes en los 10 casos; y al compararlas con los textos originales, se identificó que el 33.4% (108 frases) de las frases son relevantes, las cuales tienden a tener relación con la mención de sí mismo, de emociones, sentimientos y acciones que se mencionan en las entrevistas. Por su parte, el 66.6% (215 frases) corresponden a frases irrelevantes, y estas parecen contener palabras relacionadas con eventos contextuales o temporales, que suelen acompañar los relatos del entrevistado.

En cuanto al entrenamiento de los modelos, con base en la matriz de confusión y el reporte de clasificación, el modelo Naive Bayes presenta un grado de exactitud bastante alto, logrando identificar en un 95.22% las frases que son relevantes y aquellas que son irrelevantes; Por su parte, el modelo de máquinas de soporte vectorial (SVM) arrojó una exactitud del 99.64%; no obstante, al verificar los modelos con datos nuevos, Naive Bayes obtuvo una precisión del 46.73% y SVM obtuvo

Reporte De Clasificación.

Para este modelo se obtuvo una precisión para las frases irrelevantes del 99.46%, una sensibilidad del 100%, una media armónica del 99.73%; en las frases relevantes se evidenció una precisión del 100%, una sensibilidad del 98.97% y un f1-score del 99.48%. Con estos datos el modelo presenta una precisión general de 99.65% (véase tabla 7).

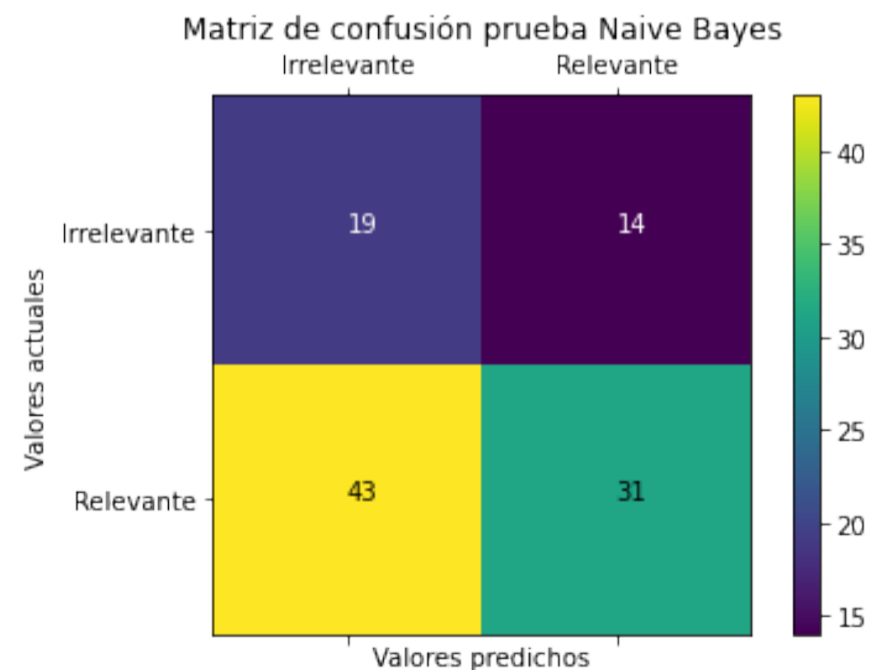
Tabla 7.
Reporte de clasificación del modelo SVM.

	precision	recall	f1-score	support
Irrelevante	0.9946	1.0000	0.9973	370.0000
Relevante	1.0000	0.9897	0.9948	195.0000
accuracy	0.9965	0.9965	0.9965	0.9965
macro avg	0.9973	0.9949	0.9961	565.0000
weighted avg	0.9965	0.9965	0.9965	565.0000

Verificación De Los Modelos.

Una vez entrenados los modelos, al utilizarlo en datos nuevos se evidenció que la precisión del modelo Naive Bayes fue del 46.73%; donde los indicadores de este demuestran que el modelo clasifica mejor las frases relevantes que las irrelevantes (véase figura 8 y tabla 8)

Figura 8.
Matriz de confusión en la verificación del modelo Naive Bayes.



Por su parte, el modelo SVM tuvo una precisión del del 28.97% y, al igual que el modelo anterior, tiende a clasificar mejor las frases relevantes que las irrelevantes (véase figura 9 y tabla 8)

un 28.97% de precisión, lo que indicaría que el primer modelo es más adecuado para esta tarea, coincidiendo con lo reportado por Florián (2013).

Si bien la precisión de los modelos es baja, estos hallazgos pueden ser evidencia del uso de procesamiento de lenguaje natural como una herramienta eficaz y eficiente para tipificar frases que permitan analizar entrevistas; donde el modelo Naive Bayes parece presentar indicadores más estables en este proceso dentro del marco de la evaluación psicológica. Por otra parte, el desarrollo de esta investigación abre la puerta a tres beneficios: el primero, desde la primera entrevista se posibilita extraer información oportuna que guíe el proceso; la segunda, los tiempos para establecer hipótesis podrían verse disminuidos, lo cual conlleva a reducir los tiempos de evaluación; y finalmente, se plantea una metodología – desde la ciencia de datos – para evaluar entrevistas de valoración e identificar frases relevantes.

Por último, se considera que las limitantes de esta investigación corresponden, en primer lugar, a los pocos datos utilizados para entrenar los modelos, dado que estos funcionan mejor cuando se cuenta con una gran cantidad de información; segundo, la fuente de las entrevistas, si bien los textos se obtuvieron de fuentes primarias, estos presentan la información desde la experiencia del evaluador; y tercero, es necesario probar si los modelos presentan sobre ajuste, condición que no pudo ser corroborada debido a la primera limitante. Por lo cual, en futuras investigaciones se recomienda acceder a entrevistas originales que permitan ofrecer un relato más inmediato, poner a prueba los modelos de aprendizaje referidos para evidenciar si hay sobre ajuste e implementar otros modelos (ej. redes neuronales) que permitan mejorar el proceso de clasificación.

Referencias.

- Alghani, I. (11 de octubre de 2020) Cleaning text data with Python. Towards data science. <https://towardsdatascience.com/cleaning-text-data-with-python-b69b47b97b76>
- Aparicio, F., De Buenaga, M., Rubio, M., Hernando, M., Gachet, D. Puertas, E. y Giráldez, I. (2011). TMT: una herramienta para guiar a los usuarios en la búsqueda de información sobre textos clínicos. *Procesamiento de lenguaje natural*, 46, 27 – 34.
- Babativa, D. (2021). Modelo de aprendizaje automático para la clasificación temprana de flujos de texto aplicado a la detección de desórdenes psicológicos. [Tesis de maestría, Universidad Nacional de Colombia]. Repositorio institucional – Universidad Nacional.
- Borja, M. (2020) Análisis de las herramientas de procesamiento de lenguaje natural para estructurar textos médicos [Tesis de maestría, Universidad de Navarra]. Repositorio académico digital Universidad de Navarra (dadun).
- Camacho, J., Moreno, S., Suarez-Obando, F., Carlos Puyana, J., & Gomez-Restrepo, C. (2013). El procesamiento de lenguaje natural y su relación con la investigación en salud mental. *Revista Colombiana de Psiquiatría*, 42(2), 227–233. [https://doi.org/10.1016/S0034-7450\(13\)70011-8](https://doi.org/10.1016/S0034-7450(13)70011-8)
- Cortez, A., Vega, H. y Pariona, J. (2009) Procesamiento de lenguaje natural. *Revista de ingeniería de sistemas e informática*. 6(2), 45 – 54.
- De la Calle, G. (2014). Modelo basado en técnicas de procesamiento de lenguaje natural para extraer y anotar información de publicaciones científicas. [Tesis de doctorado, Universidad Politécnica de Madrid]. http://oa.upm.es/30856/1/GUILLERMO_DE_LA_CALLE_VELASCO.pdf
- Fernández-Ballesteros, R. (2013) Evaluación psicológica. Conceptos, métodos y estudio de casos. Ediciones pirámide.
- Florián, J. (2013) Categorización de texto usando técnicas de machine learning aplicado a la clasificación de reclamos en los procesos de la Universidad Tecnológica de Bolívar. [Tesis de pregrado] <https://hdl.handle.net/20.500.12585/3589>
- Jorge, E. (2018) La evaluación clínica como función integral del psicólogo. *Perspectivas en psicología*. 15(1), 98 – 107.
- Iglesias, A., Castro, E., Pérez, R., Castaño, L. Martínez, P., Gómez-Pérez, J. Kohler, S. y Melero, R. (2008). MOSTAS: un etiquetador morfo-semántico, anonimizador y corrector de historiales clínicos. *Procesamiento de lenguaje natural*, 41, 299 – 300.
- Keats, D. (2009) Entrevista. Guía práctica para estudiantes y profesionales. McGraw Hill.
- Ley 1090 del 2006. por la cual se reglamenta el ejercicio de la profesión de Psicología, se dicta el Código Deontológico y Bioético y otras disposiciones. 06 de septiembre de 2006.
- Ley 1581 de 2012. Por la cual se dictan disposiciones generales para la protección de datos personales. 17 de octubre de 2012.
- Llavona, L. (2011) Entrevista en F. J. Labrador (Ed.), *Técnicas de modificación de conducta* (pp.103 – 120). Ediciones pirámide.
- Ma, E. (27 de mayo de 2018) NLP pipeline: lemmatization (part 3). Medium. <https://medium.com/@makcedward/nlp-pipeline-lemmatization-part-3-4bfd7304957>
- Merino, C y Livia, J. (2009). Intervalos de confianza asimétricos para el índice la validez de contenido: un programa de Visual Basic para la V de Aiken. *Anales de psicología*. 25(1), 169 – 171. <https://revistas.um.es/analeps/article/view/71631>
- Montero, I., & León, O. (2005) Sistema de clasificación del método en los informes de investigación en Psicología. *International Journal of Clinical and Health Psychology*, 5 (1), 115-127. <http://www.redalyc.org/articulo.oa?id=33701007>
- Morrison, J. (2015) La entrevista psicológica. Manual Moderno.
- Muñoz, M., Ausín, B., y Panadero, S. (2019) Manual práctico de Evaluación psicológica clínica (2ª ed.) . Síntesis.
- Pajaro, J. (2018) Procesamiento de lenguaje natural para la evaluación de problemas sociales. [Tesis de maestría, pontificia universidad Javeriana] <https://repository.javeriana.edu.co/handle/10554/40774>
- Páramo, J., y Espitia, C. (2018). Aplicación del aprendizaje automático en la clasificación de textos cortos: un caso de estudio en el conflicto armado colombiano [Tesis de pregrado] <https://repository.ucatolica.edu.co/server/api/core/bitstreams/3bdfb760-c23a-41ef-84e5-dc3bc3694d89/content>
- Pardo, E. (23 de agosto de 2023). Tipificacion_frases. Github. https://github.com/epardoj/tipificacion_frases
- Perpiña, C. (Ed.) (2012) Manual de la entrevista psicológica. Saber escuchar, saber preguntar. Ediciones pirámide.
- Resolución N° 008430. Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud. Colombia, 04 de octubre de 1993.
- Resolución 1995 de 1999 [Ministerio de salud] Por la cual se establecen normas para el manejo de la Historia Clínica. 8 de Julio de 1999.
- Sancho, J., Fanjul, C., De la Iglesia, M., Montell, J., & Escartí, M. (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. [https://doi.org/10.35669/rcys.2020.10\(1\).19-41](https://doi.org/10.35669/rcys.2020.10(1).19-41)
- Singh, S. (2021) Clasificación de textos mediante algoritmos de Machine Learning [Tesis de pregrado] http://dspace.umh.es/bitstream/11000/26781/1/SinghKaur_Sukhwinder.pdf
- Teja, S. (10 de junio de 2020). Stop words in NLP. Medium. <https://medium.com/@saitejaponugoti/stop-words-in-nlp-5b248dad47>