**IDENTIFICADOR DE NOTICIAS FALSAS MEDIANTE EL USO DE PROCESAMIENTO DEL LENGUAJE NATURAL Y MAQUINAS DE APRENDIZAJE**

**Santos Da Silveira, Lautaro**

**DOCENTE: Licenciado Sufotinsky, A.**

**Inteligencia Artificial y Ciencia de Datos**

**Procesamiento del Lenguaje Natural**

**Córdoba, Noviembre de 2022**

**Introducción**

Las noticias falsas, conocidas comúnmente como *Fake News*, existen hace tiempo, como es el ejemplo de la propaganda de la obra “La guerra de los mundos” en 1938, donde un locutor generaría terror en los oyentes, ya que se creían cada palabra que el locutor decía, pero a partir de la llegada de internet, estas no han hecho otra cosa que aumentar. El problema de que aumente la difusión de estas noticias es que se dificulta el poder consumir buena calidad de información. “en tiempo de inundación, lo primero que escasea es el agua potable. En esa inundación de señales informativas, lo que escasea es esa agua potable informativa, que ahora es absolutamente vital.” (Gabilondo, 2021)

Xue recalca que la difusión de noticias falsas puede causar efectos negativos masivos, a veces afectando o manipulando eventos públicos importantes (2021). Esto se puede notar con la llegada de la pandemia del 2020, puesto que por internet circulaban una gran cantidad de información errónea que provocaban pánico entre la sociedad, como es el caso con temas de la vacunación o hasta llegando al punto de proponer remedios caseros, provocando que el individuo no asista a una institución médica.

Es por estas razones que se propone la creación automatizada por medios de técnicas de algoritmo de Aprendizaje Automático que pueda identificar cuando una noticia es o no una *Fake News*. Para realizar esta tarea se utilizará el modelo de clasificación de Maquinas Vectores Soporte, como también utilizar técnicas del Procesamiento del Lenguaje Natural, para la detección de características entre las noticias erróneas.

**Marco Teórico:**

**1- Formas de detectar las Fake News (pensar titulo)**

Según Xue, la principal dificultad en la detección de noticias falsas consiste en encontrar las características que las hacen diferentes de las verdaderas (2021). En la búsqueda de los atributos que diferencian las noticias verdaderas de las falsas, Xue observa que las noticias falsas hacen uso de imágenes y textos con una alta orientación sentimental para su rápida propagación (2021). Por otro lado, Mottola analiza la relación de textos fraudulentos con estructuras fijas de expresión y un estilo de narración impactante (títulos en mayúscula, signos de puntuación exclamativos), como así también el uso de imágenes explicitas y llamativas (2020). Zhou y Zafarani proponen realizar la tarea de detectar noticias fraudulentas de dos formas: la primera manualmente, la cual no es eficiente porque se tardaría en procesar el análisis, y la segunda de forma automática con algoritmos de máquinas de aprendizaje (ML, por sus siglas en ingles), y el uso del procesamiento del lenguaje natural (PLN), siendo esta más veloz que la primera (2020).

**2- Procesamiento del Lenguaje Natural y Herramientas**

Según Cortez Vásquez, el Procesamiento del Lenguaje Natural (PLN) es la utilización de un lenguaje natural para comunicarnos con las computadoras, y que esta entienda las oraciones, el uso de los lenguajes naturales y facilitar el desarrollo de programas que realicen tareas relacionadas con el lenguaje, como así también ayudar a comprenderlos. (2008). Lo mismo destacan Deng & Liu expresando que el PLN se encarga de investigar el uso de computadoras para poder procesar o entender el lenguaje natural con el propósito de realizar tareas útiles. Se define como un campo interdisciplinario que combina la lingüística computacional, ciencias de la computación, ciencia cognitiva e Inteligencia Artificial (2018). Ambos autores concuerdan que el uso de PLN se realiza para procesar o comprender el lenguaje natural y utilizar el mismo para realizar futuras tareas. Zapata indica que el PLN debe realizarse de manera sistemática, dividiendo el elemento en partes, agregando elementos gramaticales e identificando elementos interesantes, como así eliminando elementos que no aportan información. Algunos de los elementos comunes en PLN son: Tokenización, Eliminación de palabras vacías (Stop Words), Lematización y TF-IDF (2021).

* 1. **Tokenización**

Baciero Fernández indica que la tokenización es la tarea de dividir en pedazos una secuencia de caracteres. Un token es un elemento de una secuencia de caracteres de un documento, y si estos se agrupan se puede encontrar relaciones para el procesamiento del documento (2020). Diferente a esto, Riaño desarrollan que la tokenización como la transformación de una secuencia de caracteres en una secuencia de palabras que contienen información acerca del lenguaje natural usado (2022).

* 1. **Eliminación de palabras vacías**

Zeokat afirma que el concepto de las palabras vacías (Stop Words) nace en el mundo de la lingüística computacional y se su uso en el PLN. El termino hace referencia a todas aquellas palabras que no tienen un significado por si solas, que suelen ser pronombres, preposiciones, artículos, y demás. (2017)

* 1. **Lematización**

Según Spositto, la lematización sirve para reducir variables morfológicas de una palabra, a la raíz común de esta removiendo el plural, el tiempo, o los atributos de una palabra (2022). Por otro lado, Riaño indica que la lematización consiste en relacionar una palabra flexionada con su palabra en expresión principal, es decir la forma en la que se encuentra en el diccionario (2022).

* 1. **TF-IDF**

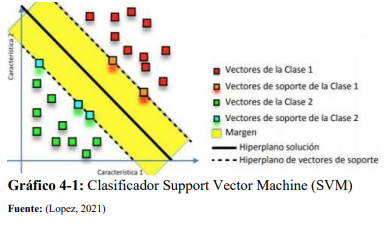
Según Kang el algoritmo TF-IDF se usa para evaluar la importancia de las palabras en el corpus, y esta importancia es proporcional al número de veces que la palabra aparece en el documento, pero a su vez inverso a la frecuencia de palabras que aparecen en el corpus (2016). Chinty expresa que el termino TF representa la frecuencia absoluta de una palabra, e IDF determina la importancia del termino en el conjunto, se obtiene calculando el logaritmo del número total de documentos contenidos en el corpus sobre el número de documentos donde aparece el término a analizar. Para concluir el cálculo, se multiplican los dos términos (TF e IDF) generando así nuestra variable TF-IDF, que determina si una palabra tiene mayor peso cuando esta sea más significativa en el texto (2022).

**3- Algoritmos de Aprendizaje Automatico**

El autor Lugo sostiene que el termino Aprendizaje Automatico (ML según sus siglas en inglés) hace referencia a la ciencia que permite que las computadoras aprendan y actúen como lo hacen los humanos, alimentándolas con datos e información que representan representaciones del mundo real (2020). Por otro lado, ML usa algoritmos de Inteligencia Artificial (IA), por lo cual, según Arbeláez-Campillo si contamos con información cualitativa y cuantitativa, podría alcanzar varios objetivos propuestos (2021). Según Digital55 utiliza algoritmos de ML con la finalidad de determinar los patrones de los datos para luego mediante modelos, realizar predicciones (2020). Cajahuanca dan a conocer los modelos de clasificación de ML, estos son algoritmos que cuentan con IA, que ayudan a aprender a una máquina, produciendo modelos matemáticos capaces de generalizar el comportamiento de los datos

**3.1- Maquina de Vectores de Soporte**

Al-Zoubi define a las Maquinas de Vectores de Soporte (SVM por sus siglas en ingles) como un algoritmo que forma parte del aprendizaje de máquinas, diseñado para trabajos de regresión y clasificación (2018: p. 3). Para entender el funcionamiento del mismo, Escobar explica que SVM proyecta las muestras con las que se entrena en un espacio vectorial, separando las muestras en categorías para luego encontrar las dos muestras opuestas que se encuentren más cercanas, denominados vectores de soporte. Entre medio de estos vectores soportes, se genera un hiperplano que separará las categorías, además de dos hiperplanos mas que se encontrarán sobre nuestros vectores de soporte (2022)



Clasificador Support Vector Machine (SVM)

Fuente: (Lopez, 2021)

**Marco metodológico:**

En los últimos años, es notable el incremento de la circulación de noticias, textos y demás artículos informativos gracias a la implementación de las redes sociales y del Internet, donde las noticias o rumores se difunden a altas velocidades, pero a su vez se vuelve más difícil identificar cuando una noticia es verdadera o no. En internet abundan estas noticias falsas, erróneas, y se las denomina comúnmente como Fake News. El mismo término se ha vuelto muy popular en el año 2017, Gabilondo propone que las Fake News son mentiras amplificadas, pero además es una industria para difundir mentiras, generando así desconfianza de la sociedad hacia lo que leen en los medios (2021). Mientras tanto, Mishra describe a las noticias falsas como cualquier material capaz de hacer creer a los lectores en información que no es real (2020)

Es de público conocimiento el problema que traen estas noticias, Xue recalca que la difusión de noticias falsas puede causar efectos negativos masivos, a veces afectando o manipulando eventos públicos importantes (2021). Es por esto que Castillo llega a la idea de que eliminar las noticias falsas es de gran necesidad para mejorar la calidad de las redes de información y mantener la estabilidad social, evitando su desinformación y manipulación (2011).

Como mencionamos con anterioridad, la divulgación de noticias falsas ha ido incrementando con el pasar del tiempo y, si bien se han desarrollado tecnologías relacionadas al Procesamiento del Lenguaje Natural para su detección, se plantea que siga en aumento, dificultando el nutrirse de información factible. Por estas razones se propone realizar una aplicación que disponga de usos de Procesamiento del Lenguaje Natural y Aprendizaje Automático para así poder identificar si una noticia es verdadera o falsa.

* 1. **Objetivo General**

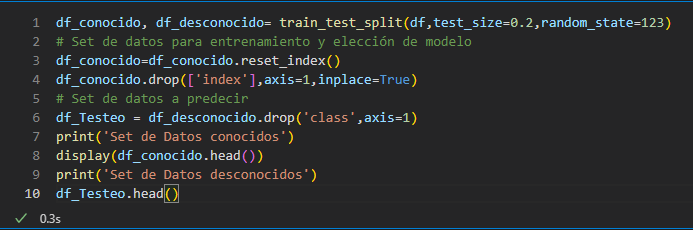
Generar una aplicación que detecte si una noticia es o no una Fake News que cuente con herramientas de Procesamiento del Lenguaje Natural. El mismo trabajo deberá finalizarse antes del 18/11/22

* 1. **Objetivos específicos**
* Caracterizar y analizar los datos que se utilizarán en la aplicación.
* Desarrollar una función de pre-procesamiento aplicando herramientas de PLN
* Desarrollar un algoritmo de Aprendizaje Automático, generando modelos de predicción, cuyo nivel de precisión sea superior al 70%
  1. **Análisis de los datos:**

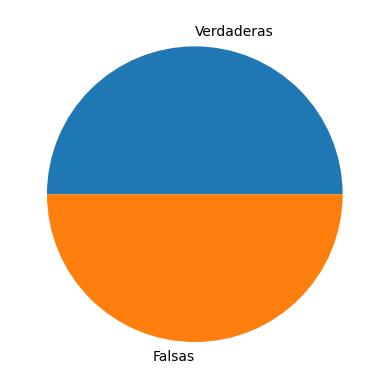
Para la realización de trabajos contamos con 2000 datos, 1000 noticias verdaderas y las restantes son falsas. Se muestran a continuación los primeros 5 datos

|  |  |
| --- | --- |
| Class | Text |
| VERDADERO | Algunas de las voces extremistas más conocidas de EE.UU., cuentas asociadas con al movimiento de la conocida como alt-right o ligadas a grupos antisemitas o xenófobos , han sido eliminadas de un |
| VERDADERO | Después de casi dos años y medio de luchas políticas y judiciales, Donald Trump ha presentado hoy su proyecto de inmigración para Estados Unidos . Es un proyecto elaborado por el yerno y asesor de Donald Trump, Jared Kushner, que sólo concreta una cosa: |
| VERDADERO | Dos periodistas birmanos de la agencia Reuters fueron liberados este martes al recibir un perdón del presidente de su país tras pasar año y medio en prisión por revelar una matanza del Ejército contra la minoría musulmana rohingya. Wa Lone y Kyaw Soe Oo |
| VERDADERO | El Cuerpo Nacional de Policía ha detenido a cuatro personas acusadas de formar parte de un grupo criminal dedicado al tráfico de seres humanos en la frontera francoespañola de Irún (Gipuzkoa). Esta célula captaba extranjeros llegados en pateras y les of |
| VERDADERO | El desfile de la firma en Roma se convierte en una oda a la libertad de las mujeres para decidir sobre sus propios cuerpos. Los Museos Capitolinos de Roma se transformaron la noche del martes en escenario sobre el que presentar la colección crucero 202 |

Se mezclan aleatoriamente los datos por si estos fueron cargados en una forma específica con una función perteneciente a la librería Scikit-Learn llamada shuffle. Del total de mis datos mezclados, con ayuda de la función *train\_test\_split* de Scikit-Learn se apartan 400 de estos para realizar una predicción final, obteniendo así un set de datos con valores conocidos para la creación de los algoritmos con un tamaño de 1600 observaciones. Para facilitar la utilización de los datos al momento de dividirlos, se resetea el índice de mi Set de Datos.



A fines de corroborar si mis datos siguen teniendo una proporción del 50% de los casos para cada observación, se genera un gráfico de proporciones con la librería de Matplotlib, obteniendo el siguiente grafico



Se puede observar que todavía se mantiene la proporción de la mitad de casos para cada una de las observaciones.

Otro análisis que se propone es visualizar cuales son las palabras que más se utilizan, este análisis se genera luego de realizar los procesos de pre-procesamiento, puesto que, si se realiza antes, habría palabras vacías o que tengan el mismo significado. Con la librería *wordcloud* y Matplotlib se pueden realizar y analizar las siguientes graficas

* Palabras que más se repiten en todo mi set de datos:



* Palabras más utilizadas en noticias verdaderas:

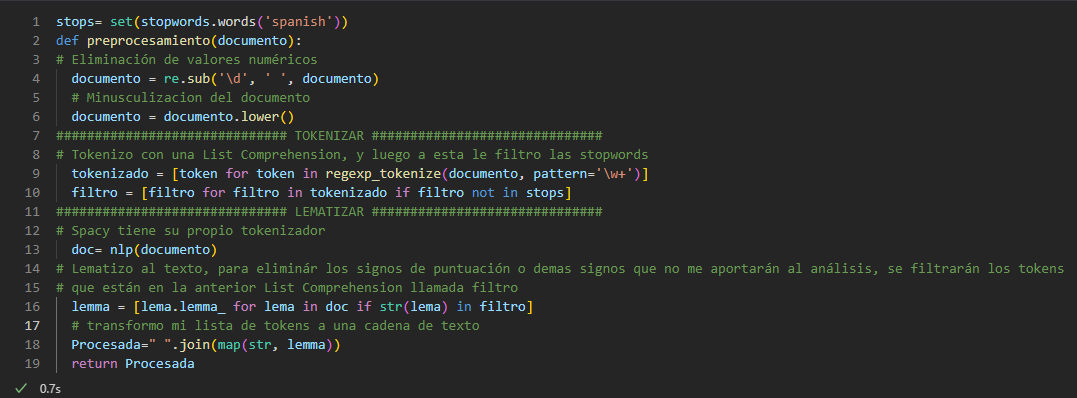


* Palabras más utilizadas en noticias falsas:



* 1. **Función de pre-procesamiento**

Luego de analizar los datos con los que contamos, para que estos puedan ser utilizados en la predicción si una noticia es o no falsa, se creó una función de pre-procesamiento de textos con herramientas de Procesamiento del Lenguaje Natural



La función cuenta con una secuencia de pasos para así poder aplicar correctamente los procesos. Se detallan a continuación:

1. Eliminación de números en el documento a tratar
2. Minusculización del documento para que el análisis sea más certero y no analice palabras iguales como si fuesen diferentes.
3. Tokenizacion del documento, se desarrolla con la librería NLTK, esto eliminará caracteres especiales tales como
4. Eliminación de palabras vacías a fin de reducir los tokens y que el análisis no se vea sesgado por la gran cantidad de palabras vacías que podría haber en el documento, nltk cuenta con su lista de palabras vacías, solo se debe establecer el idioma.
5. Lematización de los tokens, se realizó con la librería Spacy y su diccionario de lemas en español que fueron establecidos con anterioridad, quedándonos la forma base de las palabras y disminuyendo la cantidad de tokens para un mejor análisis

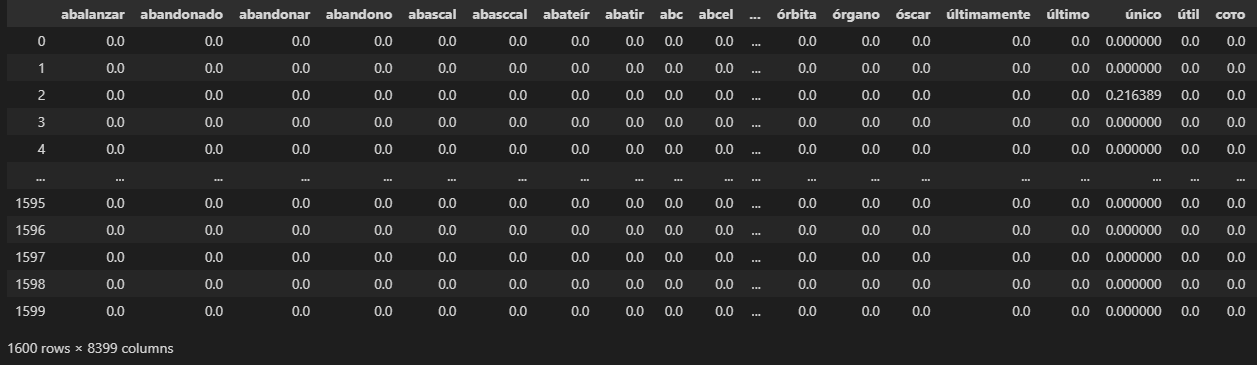
Aplicamos esta función a nuestro Set de Datos perteneciente a nuestras variables que previamente separamos, dejando nuestro valor a predecir de lado. Luego de aplicar la función nos da paso a poder introducirlos en el vectorizador TF-IDF. Podemos visualizar la diferencia entre la noticia original y lo que nos devuelve nuestra función de pre-procesamiento





* 1. **Transformación de los textos para introducirlos en modelos de Aprendizaje Automático**

Aplicada la función de pre-procesamiento, a los datos procesados se los representará de forma numérica, ya que los modelos de Aprendizaje Automático solo toman valores numéricos. Este proceso se realiza mediante la herramienta de TfidfVectorizer perteneciente a la librería de Scikit-Learn. Esta herramienta realiza el calculo de TF-IDF aplicadas a todo mi set de datos, generando así una tabla que tendrá como columnas, todos los tokens aparecidos en la totaldad de los documentos, mientras que las filas serán cada uno de las noticias previamente procesadas. TfidfVectorizer considerará la frecuencia de cada término de un documento con relación a la totalidad de documentos existentes.

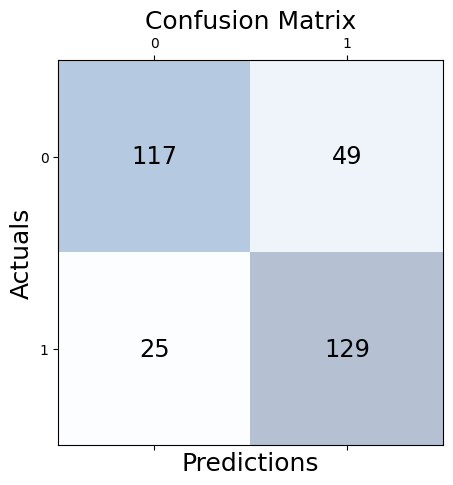


Este set de datos pasará a ser mi x para el proceso de modelado que se realizará a continuación.

* 1. **Modelado de Algoritmo de Predicción**

Una vez obtenido mis valores independientes ya procesados (x) y mis variables dependientes (y), realizo la división de entrenamiento y testeo con la función *train\_test\_split* de Scikit-Learn, obteniendo así mi x\_entreno, x\_testeo, y\_entreno e y\_testeo. Con mis datos de entrenamiento y la librería de GridSerchCV perteneciente a Scikit-Learn se procederá a buscar la mejor combinación de hiperparametros para que el modelo de Maquina de Vectores Soporte tenga un funcionamiento óptimo.

Luego de la búsqueda, se llega a la conclusión de que la mejor combinación de hiperparametros son los parámetros C= 1.33 y gamma= Scaler. Encontrado estos valores, se procede a entrenar el modelo con estos valores, obteniendo así un valor de predicción igual al 76%, superando el umbral propuesto. El funcionamiento de nuestro modelo se puede visualizar en la siguiente matriz de confusión:



Se puede aceptar el funcionamiento del algoritmo, ya que contábamos con 320 casos para realizar el testeo, y el modelo predijo a la perfección 246 casos, tomándose como óptimo.

* 1. **Conclusiones:**

El estudio desarrollado estableció la idea de que es posible detectar noticias falsas mediante usos de algoritmos de Aprendizaje Automático y Procesamiento del Lenguaje Natural. Que este trabajo se haya podido realizar, indica que hay una diferencia en la forma de escritura entre las noticias verdaderas y falsas.

Si bien el modelo solo superó el 70% de precisión, se espera que, si se añaden más datos para la generación del modelo, este irá evolucionando y mejorando su nivel de acertación en las futuras predicciones. Por otra parte, se espera que si se utilizaría un modelo de aprendizaje profundo y más cantidad de variables, como el título de la noticia, la predicción mejoraría notablemente.

**Bibliografía**

Coruña (2021). *Iñaki Gabilondo: “Decir que nosotros hacíamos periodismo de verdad y el resto es basura es una visión cateta”*. Obtenido de <https://www.informacion.es/cultura/2021/11/12/decir-haciamos-periodismo-resto-basura-59454504.html>

Espejel-Rivera et al. (2022). *Detección automática de noticias falsas usando representaciones textuales tradicionales y soluciones basadas en aprendizaje profundo*. Obtenido de <https://repository.uaeh.edu.mx/revistas/index.php/icbi/article/view/9008/9326>

Xue, J. et al. (2021). *Detecting fake news by exploring the consistency of multimodal data*. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0306457321001060#preview-section-snippets>

Arora A. & Choudhary A. (2021). *Linguistic feature based learning model for fake news detection and classification*. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S095741742030909X>

Pérez Álvarez J. (2021). *Clasificador de textos de acuerdo con los Objetivos de Desarrollo Sostenible usando Procesamiento del Lenguaje Natural*.

Zhou X. (2018). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*. Obtenido de <https://www.researchgate.net/publication/329388190_A_Survey_of_Fake_News_Fundamental_Theories_Detection_Methods_and_Opportunities>

Álbarez-Daza N. et al. (2021). *Detección de Noticias Falsas en Redes Sociales Basada en Aprendizaje Automático y Profundo: Una Breve Revisión Sistemática*. Obtenido de <https://www.researchgate.net/publication/349120690_Deteccion_de_Noticias_Falsas_en_Redes_Sociales_Basada_en_Aprendizaje_Automatico_y_Profundo_Una_Breve_Revision_Sistematica>

Zapata García A. (2021). *ANÁLISIS DE TEXTOS MEDIANTE TÉCNICAS NLP PARA LA CATEGORIZACIÓN DE USUARIOS*. Obtenido de <https://repositorio.upct.es/bitstream/handle/10317/9647/tfg-zap-ana.pdf?sequence=1&isAllowed=y>

Castillo C. (2011). *Information Credibility on Twitter*. Obtenido de <https://chato.cl/papers/castillo_mendoza_poblete_2010_twitter_credibility.pdf>

Shukla P & Agarwal R. (2021). *Analyzing Machine Learning Enabled Fake News Detection Techniques for Diversified Datasets*. Obtenido de <https://www.hindawi.com/journals/wcmc/2022/1575365/>

Lugo, A. (2020). *¿Qué es el Machine Learning? INVID*. Obtenido de [¿Qué es el Machine Learning? - INVID (invidgroup.com)](https://invidgroup.com/es/machine-learning-metodos/)

Arbeláez-Campillo D. et al. (2021). *Inteligencia artificial y condición humana: ¿Entidades contrapuestas o fuerzas complementarias?* Obtenido de <https://www.redalyc.org/articulo.oa?id=28066593034>

Al-Zoubi, A et al. (2018). *Evolving Support Vector Machines using Whale Optimization Algorithm for spam profiles detection on online social networks in different lingual contexts.* Obtenido de <https://doi.org/10.1016/j.knosys.2018.04.025>.

Escobar Chávez J. (2022). *DETECCIÓN DE FALLAS EN CAJAS DE ENGRANAJES UTILIZANDO EL MÉTODO DE APRENDIZAJE DE MÁQUINAS SUPPORT VECTOR MACHINE (SVM).* Obtenido de <http://dspace.espoch.edu.ec/bitstream/123456789/16949/1/25T00459.pdf>

Baciero Fernández J. (2020). *ELABORACIÓN DE UN MODELO DE RECONOCIMIETNO DE “ENTIDADES NOMINALES” (NER) PARA SU USO EN APLICACIONES DE PROCESAMIENTO DEL LENGUAJE NATURAL (PLN).* Obtenido de <https://oa.upm.es/62858/1/TFG_JOSE_IGNACIO_BACIERO_FERNANDEZ.pdf>

Cortez Vásquez A. et al. (2009). *Procesamiento de lenguaje natural.* Obtenido de <https://revistasinvestigacion.unmsm.edu.pe/index.php/sistem/article/view/5923/5121>

Vera Guamán K. & Yela García R. (2021). *CLASIFICACIÓN Y ETIQUETADO DE MENSAJES DE TWITTER DE ECUADOR PARA DETERMINAR QUÉ TEMA TRATAN UTILIZANDO UN MODELO TRANSFORMER DE PROCESAMIENTO DE LENGUAJE NATURAL.* Obtenido de <http://repositorio.ug.edu.ec/bitstream/redug/59656/1/B-CISC-PTG%232028-A%c3%b1o%202022%20Vera%20Guam%c3%a1n%20Kerlly%20del%20Pilar%20-%20Yela%20Garc%c3%ada%20Ricardo%20Alberto.pdf>

Zeokat (2014). *¿Qué son las stop words o palabras vacías?.* Obtenido de <https://www.vozidea.com/que-son-las-stop-words-o-palabras-vacias>

Riaño C. et al. (2021). *Propuesta de aplicativo web para el desarrollo de perfiles ocupacionales en el entorno educativo superior, utilizando técnicas para el procesamiento de lenguaje natural.* Obtenido de <http://ojs.incaing.com.mx/index.php/ediciones/article/view/52/propuesta>

Spositto O. et al. (-). *Adecuación de un Sistema de Recuperación de Información para su utilización en un Contexto Jurídico.* Obtenido de <http://sedici.unlp.edu.ar/bitstream/handle/10915/144389/Documento_completo.pdf-PDFA.pdf?sequence=1&isAllowed=y>

Chiny M. et al. (2022). *Netflix Recommendation System based on TF-IDF and Cosine Similarity Algorithms.* Obtenido de <https://www.researchgate.net/publication/360856267_Netflix_Recommendation_System_based_on_TF-IDF_and_Cosine_Similarity_Algorithms>

Valero Cajahuanca J. et al. (2022). *Deserción universitaria: Evaluación de diferentes algoritmos de Machine Learning para su predicción.* Obtenido de <https://www.researchgate.net/publication/360856267_Netflix_Recommendation_System_based_on_TF-IDF_and_Cosine_Similarity_Algorithms>

Anexo