

# **DETECCIÓN AUTOMÁTICA DE NOTICIAS FALSAS EN ESPAÑOL CON TÉCNICAS DE MACHINE LEARNING**

**Juan Alberto Ortega Riveros**

**Diany Yisel Quintero Perozo**

**Trabajo de grado dirigido por:**

**Haydemar María Núñez Castro**

**Universidad de los Andes**

**Facultad de Ingeniería**

**Ingeniería de Sistemas y Computación**

**Bogotá, Colombia**

**Diciembre 2020**

## **RESUMEN**

El trabajo desarrollado en el presente proyecto nació de la necesidad de frenar de alguna manera la creciente propagación de noticias falsas compartidas por la ciudadanía en general en los medios digitales. Es por esto que se propuso desarrollar un producto viable que ayude en la identificación automática de noticias falsas con técnicas de aprendizaje automático.

Para la implementación de esta solución se decidió tener en cuenta atributos alternos a procesar directamente el texto de la noticia que pudieran determinar la veracidad de esta, tales como el porcentaje de mayúsculas en el título o el número de signos de interrogación.

Como resultado de lo descrito, se obtuvo un chatbot implementado en Amazon Lex integrado a Facebook Messenger que recibe el título y el texto de la noticia, seguidamente estos parámetros son procesados por el modelo de aprendizaje automático y finalmente se le responde al usuario si la noticia es verdadera o falsa y con qué probabilidad.

## TABLA DE CONTENIDO

<b>1. INTRODUCCIÓN.....</b>	<b>4</b>
<b>2. DESCRIPCIÓN GENERAL .....</b>	<b>5</b>
<b>2.1    Objetivos .....</b>	<b>5</b>
<b>2.1.1    Objetivo general .....</b>	<b>5</b>
<b>2.1.2    Objetivos específicos .....</b>	<b>5</b>
<b>2.2    Antecedentes .....</b>	<b>5</b>
<b>2.3    Identificación del problema y de su importancia .....</b>	<b>7</b>
<b>3. DISEÑO Y ESPECIFICACIONES .....</b>	<b>7</b>
<b>3.1    Definición del problema computacionalmente .....</b>	<b>7</b>
<b>3.2    Especificaciones .....</b>	<b>7</b>
<b>3.3    Restricciones .....</b>	<b>8</b>
<b>4. DESARROLLO DEL DISEÑO .....</b>	<b>8</b>
<b>4.1    Recolección de información.....</b>	<b>10</b>
<b>5. IMPLEMENTACIÓN .....</b>	<b>11</b>
<b>5.1    Descripción de la implementación .....</b>	<b>11</b>
<b>5.2    Resultados esperados .....</b>	<b>16</b>
<b>6. VALIDACIÓN.....</b>	<b>16</b>
<b>6.1    Métodos .....</b>	<b>16</b>
<b>6.2    Validación de resultados.....</b>	<b>16</b>
<b>7. CONCLUSIONES.....</b>	<b>17</b>
<b>7.1    Discusión .....</b>	<b>17</b>
<b>7.2    Trabajo futuro.....</b>	<b>17</b>
<b>8. REFERENCIAS .....</b>	<b>17</b>

## 1. INTRODUCCIÓN

Las noticias falsas son un tipo de bulo que consiste en un contenido pseudoperiodístico difundido a través de diferentes medios como portales de noticias, prensa, radio, televisión y redes sociales y cuyo objetivo es la desinformación. Actualmente, con la irrupción de las tecnologías de la información, este fenómeno ha adquirido una nueva dimensión principalmente por el alcance universal que tienen los nuevos medios de comunicación social (redes sociales, páginas web, blogs, etc.). Las noticias falsas provienen de sitios ficticios diseñados para parecer fiables que producen títulos llamativos para atraer al lector. La polémica creada por estas noticias maliciosas implica principalmente a las fuentes o portales que las crean, pero también a las redes sociales que ayudan a su propagación viral y a las plataformas de publicidad que las financian.

Detrás de la creación de noticias falsas existe un interés económico, tal como se menciona en el artículo publicado por el Ministerio de Tecnologías de la Información y las Comunicaciones (2020):

“De acuerdo con el medio de comunicación internacional BBC, plataformas de noticias falsas de Estados Unidos han llegado a recaudar hasta US\$10.000 por una noticia. Así, la generación de clics por contenido compartido de información de alto impacto, aunque sea falsa, termina siendo lucrativa para ciberdelincuentes.”

Por esta razón, la difusión de estas noticias representa un riesgo real para nuestra sociedad. Un claro ejemplo de esto es la cantidad de noticias falsas que se han generado en el contexto de la pandemia creando pánico en la ciudadanía en general e incidiendo en su comportamiento. La viralización de información falsa influye directamente en la conciencia colectiva llevando así a los ciudadanos a realizar acciones poco coherentes, como las noticias que afirmaban la inexistencia del Covid-19 que motivaban a muchas personas a ignorar las medidas sanitarias impuestas por el gobierno.

Según el artículo publicado por Herrera Cabello (2020):

“Los medios han tenido que adaptarse a la realidad de la conectividad y a la obligación de liberar contenidos las veinticuatro horas del día. Salvo en los medios impresos como los periódicos, cuyas noticias se reportean durante el día para ser publicadas en la jornada siguiente o en los días posteriores, los contenidos en Internet deben replicarse inmediatamente debido a la competencia en tiempo real con otros portales de noticias y también con los mismos usuarios en las redes sociales.”

Esto provoca que se generen cadenas masivas de desinformación debido al gran alcance que tiene el internet. De acuerdo con lo anterior y teniendo en cuenta que a lo largo de los años la comunidad científica ha estado desarrollando investigaciones para lograr la detección de noticias falsas, tomando principalmente un enfoque lingüístico que comprende la representación de datos, la sintaxis profunda, el análisis semántico y la estructura retórica, es posible proponer métodos alternos a procesar directamente el texto de la noticia para su clasificación como verdadera o falsa. Por ejemplo, a partir de

características que se pueden obtener del texto y título de la noticia como el número de mayúsculas o el número de signos de interrogación, clasificar una noticia como verdadera o falsa.

En conclusión, el presente documento explica detalladamente el proceso que se llevó a cabo desde el entendimiento del problema hasta la implementación del chat-bot propuesto como solución. Cubriendo así las etapas de análisis, diseño, construcción y evaluación del prototipo funcional.

## 2. DESCRIPCIÓN GENERAL

### 2.1 Objetivos

#### 2.1.1 Objetivo general

Proponer, diseñar e implementar una solución tecnológica que permita la automatización en el proceso de detección de noticias falsas para un contexto local en idioma español.

#### 2.1.2 Objetivos específicos

- Comprender el panorama actual respecto a la identificación de noticias falsas.
- Desarrollar habilidades en el uso de herramientas de aprendizaje automático y análisis de datos.
- Consolidar los conocimientos aprendidos durante la carrera para la construcción de una solución tecnológica de calidad.

### 2.2 Antecedentes

El tema de la detección de *fake news* ha sido abordado desde varias perspectivas, entre estas se destacan: (1) *knowledge-based*, (2) *style-based*, (3) *propagation-based* y (4) *source-based* (Zhou and Zafarani, 2020). A continuación, se describen las perspectivas mencionadas.

- **Knowledge-based:** En la detección de noticias desde esta perspectiva es común utilizar un proceso llamado verificación de datos que tiene como objetivo evaluar la autenticidad comparando el conocimiento extraído del contenido de las noticias. Dicha verificación de datos puede realizarse de manera manual o de manera automática con técnicas de recuperación de información (IR), procesamiento del lenguaje natural (NLP) y aprendizaje automático (ML), así como en la teoría de grafos.

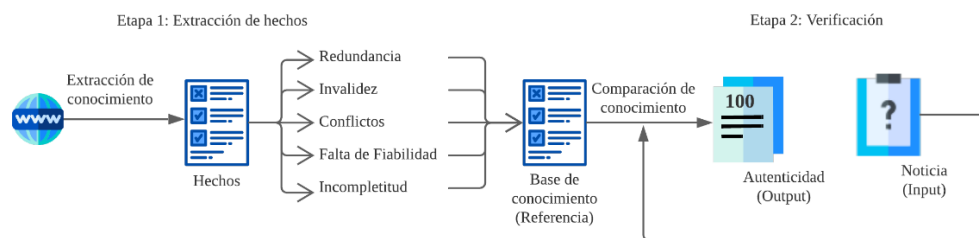


Fig 1 Proceso automático de verificación de noticias.  
Adaptado de: <https://doi.org/10.1145/3395046>.

El proceso automático de verificación de datos se muestra en la Fig. 1. Y se puede dividir en dos etapas: extracción de hechos, que se basa en una fuente comparativamente confiable para extraer conocimiento y verificación de datos, en donde se compara el conocimiento extraído del contenido de las noticias por verificar con los hechos.

- **Style-based:** Esta perspectiva también se enfoca en analizar el contenido de las noticias. Sin embargo, los métodos basados en el estilo buscan evaluar la intención de la noticia, es decir, ¿existe la intención de engañar al público o no? La suposición detrás de estos métodos es que los que originan las noticias falsas las escriben con un estilo “especial” para animar al público a leerlas. En esta perspectiva se define *Fake News Style* como “un conjunto de características cuantificables que pueden representar bien el contenido de noticias falsas y diferenciarlo del contenido de noticias verdaderas” (Zhou and Zafarani, 2020).

Teniendo en cuenta lo anterior, para la detección se utilizan modelos basados en aprendizaje automático (ML) tradicional que se entrenan con las características basadas en el estilo, por ejemplo Random Forest (RF) o XGBoost, y en aprendizaje profundo (DL) para procesamiento de las imágenes, por ejemplos redes neuronales.

- **Propagation-based:** Desde esta perspectiva se utiliza información relacionada con la difusión, por ejemplo, los usuarios que las difunden. “La entrada a un método basado en propagación puede ser (I) cascada de noticias, una representación directa de la propagación de noticias, representada en términos del número de pasos (saltos) que ha recorrido la noticia o las veces que se publicó, o un (II) grafo autodefinido, una representación indirecta que captura información adicional sobre la propagación de noticias” (Zhou and Zafarani, 2020). Al igual que en las perspectivas anteriores, los métodos para la detección se basan en aprendizaje automático tradicional o redes neuronales profundas.
- **Source-based:** Mientras que en las perspectivas anteriores la atención se ha centrado en la noticia, esta perspectiva la atención se centra en cada autor, editor y usuario que podría haber escrito, publicado o difundido las noticias ya que consiste en evaluar la credibilidad de la fuente de las noticias. Este enfoque para detectar noticias falsas puede parecer arbitrario, pero es eficiente (Nørregaard, 2019), ya que la evidencia ha revelado que muchas historias de noticias falsas provienen de sitios web de noticias falsas que solo publican engaños. Sin embargo, existen casos en los cuales una noticia falsa publicada en una fuente no confiable sea re-publicada por una fuente confiable por la necesidad de publicar contenido constantemente.

### 2.3 Identificación del problema y de su importancia

La principal problemática de las noticias falsas es que estas afectan directamente en la conciencia colectiva de la población, creando desinformación que puede llevar a acciones con poco o nulo razonamiento. Este problema cobra aún más relevancia teniendo en cuenta 1) el alcance global que brinda internet, por medio principalmente de sitios web y redes sociales y 2) su rápida propagación debido a que suelen ser títulos llamativos que buscan captar atención masiva.

Sumado a esto, se tiene que la detección manual de las noticias falsas no es factible debido a que se necesita un periodo considerable de tiempo, en el cual esta noticia falsa ya pudo haber ocasionado estragos. Finalmente, se requieren métodos adicionales a la detección de las noticias falsas por su fuente, ya que una práctica común de los sitios web informativos que se consideran como confiables, es replicar el contenido de otros portales que también poseen esta etiqueta, logrando así que una noticia falsa pueda extenderse en portales creíbles.

## 3. DISEÑO Y ESPECIFICACIONES

### 3.1 Definición del problema computacionalmente

Desarrollar un sistema de detección de noticias falsas para el idioma español basado en estilo haciendo uso de técnicas de aprendizaje automático aplicado al procesamiento del lenguaje natural (NLP).

En específico se requiere diseñar un sistema soportado por:

- Un API REST que reciba el título y cuerpo de una noticia en español y genere las variables necesarias para dar una respuesta sobre su validez basado en un modelo pre-entrenado de aprendizaje automático.
- Un Chatbot creado con la infraestructura de Amazon Lex que permita modelar la lógica del procesamiento del input del usuario. Este chatbot será integrado con la plataforma Messenger, de Facebook.
- Una función Lambda que será llamada por Amazon Lex para consultar el API del que se habló anteriormente.

### 3.2 Especificaciones

#### 3.2.1 Requerimientos funcionales

<b>Nombre</b>	R1 – Predecir la veracidad de una noticia (verdadera o falsa).
<b>Resumen</b>	Se desea predecir, basado en el estilo del título y el cuerpo de una noticia escrita en el idioma español si esta es falsa o no.
<b>Entradas</b>	El título de la noticia y el cuerpo de la noticia.
<b>Resultados</b>	Verdadera o falsa.

*Tabla 1 Requerimiento funcional 1*

<b>Nombre</b>	R2 – Predecir la veracidad de una noticia (probabilidad asociada a una etiqueta).
<b>Resumen</b>	Se desea detallar, basado en el estilo del título y el cuerpo de una noticia escrita en el idioma español el porcentaje de probabilidad asociado a la etiqueta predicha en R1.
<b>Entradas</b>	El título de la noticia y el cuerpo de la noticia.
<b>Resultados</b>	Un porcentaje de probabilidad asociado a la etiqueta predicha (Falsa o verdadera).

*Tabla Requerimiento funcional 2*

### 3.2.2 Requerimientos no funcionales

- Se desea que la solución tenga un tiempo de respuesta de máximo un minuto
- Se desea que la solución sea integrada a Facebook Messenger para mejorar la experiencia de usuario
- La solución deberá contar con una arquitectura que soporte los requerimientos funcionales
- El sistema debe soportar mínimo 5 usuarios de forma simultánea

### 3.3 Restricciones

- No se recolectará información del usuario que use el sistema
- No se generarán costos ya que se usarán equipos personales y la capa gratuita de Amazon Web Services.
- No se garantiza la predicción correcta de una noticia en todos los casos. El sistema deberá usarse como una herramienta para permitir emitir mejores juicios de valor.

## 4. DESARROLLO DEL DISEÑO

Durante el desarrollo de la solución se hizo uso de la metodología CRISP-DM, la cual proporciona una descripción normalizada del ciclo de vida de un proyecto estándar de análisis de datos. La metodología contempla el proceso de análisis de datos como un proyecto profesional, estableciendo así un contexto mucho más rico que influye en la elaboración de los modelos. El ciclo de vida del proyecto de minería de datos consiste en seis fases mostradas en la siguiente figura:



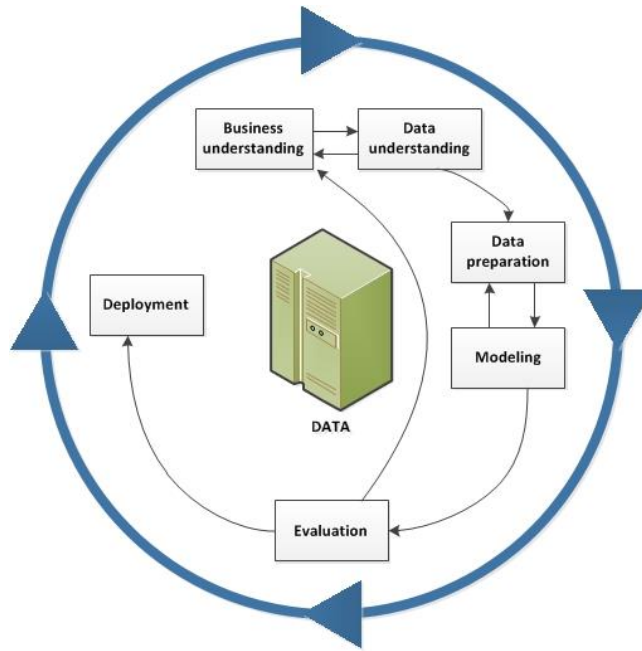


Fig 2 Ciclo de vida de un proyecto de minería de datos. Tomado de:  
[https://www.ibm.com/support/knowledgecenter/es/SS3RA7\\_sub/modeler\\_crispdm\\_ddita/clementine/crisp\\_help/crisp\\_overview.html](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html)

Adicionalmente, se propone la siguiente arquitectura para la implementación de la solución.

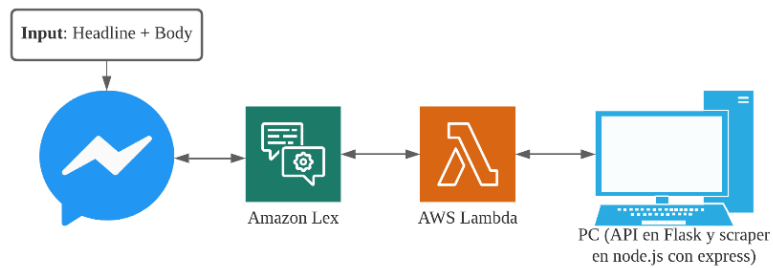


Fig 3 Arquitectura propuesta

La arquitectura consiste en un chat-bot implementado en Amazon Lex y a su vez integrado a Facebook Messenger que recibe como entrada el título y cuerpo de la noticia a validar. El bot consume mediante una función en AWS Lambda el modelo final generado a través de un API basado en Flask que se encuentra corriendo localmente. El API realiza todas las funciones de preprocesamiento, incluido el llamado a otra API en node.js, que también corre localmente, en la que se realizan las funciones de *web scraping*. Finalmente, se realiza la predicción y se le informa al usuario si la noticia es verdadera o falsa y con qué probabilidad. A continuación, se muestra un ejemplo de respuesta.

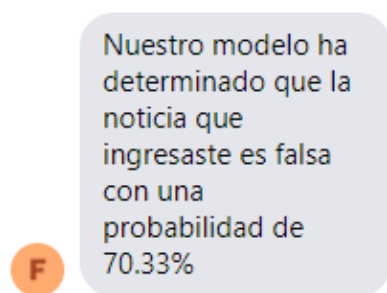


Fig 4 Ejemplo de respuesta del chat-bot

#### 4.1 Recolección de información

Se utilizó un conjunto de datos abierto que contiene 971 noticias etiquetadas divididas en 491 noticias reales y 480 noticias falsas divididas en entrenamiento y test manteniendo una relación 70%-30%. El conjunto cubre noticias de 9 temas diferentes: ciencia, deporte, economía, educación, entretenimiento, política, salud, seguridad y sociedad. A continuación, se describen las columnas que presenta el conjunto de datos:

Columna	Descripción
<b>Id</b>	Asigna un identificador a cada instancia.
<b>Category</b>	Indica la categoría de la noticia (verdadera o falsa).
<b>Topic</b>	Indica el tema relacionado con la noticia.
<b>Source</b>	Indica el nombre de la fuente.
<b>Headline</b>	Contiene el titular de la noticia.
<b>Text</b>	Contiene el texto sin formato de la noticia.
<b>Link</b>	Contiene la URL de la fuente.

Tabla 3 Descripción de columnas del dataset

Adicionalmente el dataset ya había pasado por un proceso previo de normalización que eliminó elementos comunes en la estructura de las noticias y se pueden utilizar como marcadores. Dichos elementos se etiquetaron de la siguiente forma.

- 1) Los números que representan cantidades, horarios o precios fueron etiquetados usando NUMBER;
- 2) Las direcciones de correo electrónico de los autores o editores se etiquetaron usando EMAIL;
- 3) Las URL de referencias se etiquetaron usando URL;
- 4) Los números de teléfono fueron etiquetados usando PHONE;

- 5) Los símbolos del dólar y el euro fueron etiquetados DOL y EUR respectivamente.

## 5. IMPLEMENTACIÓN

### 5.1 Descripción de la implementación

Describir cómo el proyecto se divide en etapas. Presentar detalle de las etapas.

- I. **Entendimiento del negocio:** Se investigó el panorama general de la teoría en la que se soporta la detección de noticias falsas, el impacto que dichas noticias generan en la sociedad y las soluciones existentes. Así como los distintos tipos de aproximaciones que se realizan para detectarlas automáticamente, los cuáles son: (1) *knowledge-based*, (2) *style-based*, (3) *propagation-based* y (4) *source-based* (Zhou and Zafarani, 2020).
- II. **Comprensión de los datos:** La fase de entendimiento de datos consiste en familiarizarse con los datos, identificar los problemas de calidad, descubrir conocimiento preliminar sobre los datos, y/o descubrir subconjuntos interesantes para formar hipótesis en cuanto a la información oculta<sup>5</sup>. Durante esta fase recolectamos los datos (de entrenamiento y testeo) e hicimos el respectivo perfilamiento de estos para así detectar errores o problemas que afecten su calidad.

A continuación, se presentan algunas gráficas de interés realizadas en este paso:

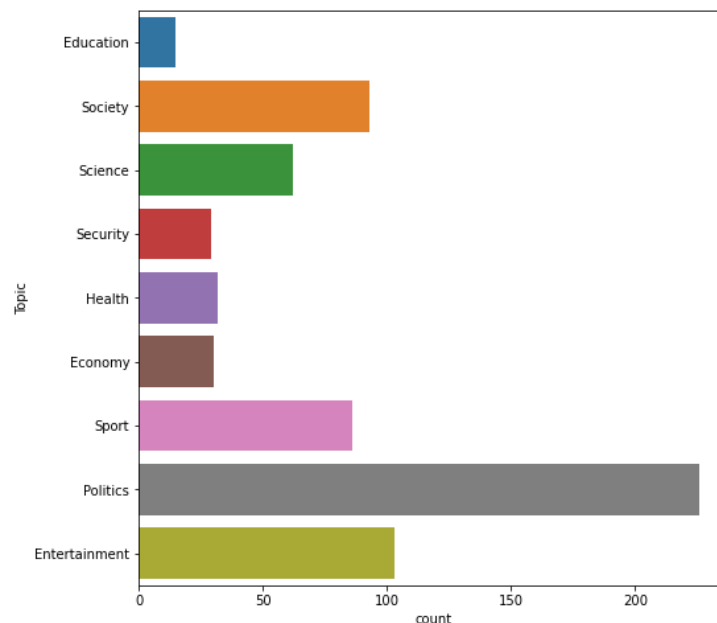


Fig 5 Distribución de noticias por tema.



Como se notó en las figuras 7 y 8, las palabras más comunes tanto para noticias falsas como para noticias verdaderas son similares, lo que indica que los tópicos de los que se hablan en los dos tipos de etiquetas son generalmente los mismos.

**III. Preparación de los datos:** Esta fase cubre todas las actividades necesarias para construir el conjunto final de datos a partir de los datos iniciales. Las tareas incluyen la selección de tablas, registros y atributos, así como transformación y limpieza de datos. En esta fase nos enfocamos específicamente en realizar las transformaciones necesarias a los datos para que puedan ser ingresados correctamente a los modelos propuestos. Sin embargo, hubo un preprocesamiento diferente dependiendo de la aproximación propuesta, de la siguiente manera:

**Para la aproximación por Random Forest y Gradient Boosting usando variables de estilo generadas del texto:**

- Se crearon funciones para generar variables de estilo a partir del texto del título de la noticia y su contenido. Además, se creó un API auxiliar para la generación de las variables para las cuales se necesitaba realizar *web scraping*. A continuación, se aprecian algunas de estas funciones utilizadas:

```
# palabras negativas
def negativeTot(str):
    # break the string into List of words
    str = str.split()
    # Loop till string values present in list str
    count = 0
    for i in range(0, len(str)):
        if str[i] in negative_words:
            count = count + 1;
    return count

def negativeRel(str):
    # break the string into List of words
    str = str.split()
    # Loop till string values present in list str
    count = 0
    for i in range(0, len(str)):
        if str[i] in negative_words:
            count = count + 1
    return count/len(str)

# palabras unicas / palabras totales
def redundancia(str):
    # break the string into List of words
    str = str.split()
    str2 = []
    # Loop till string values present in list str
    for i in range(0, len(str)):
        # checking for the duplicacy
        if str[i] not in str2:
            # insert value in str2
            str2.append(str[i])

    rta = len(str2)/len(str)

    return rta
```

Fig 9 Funciones para la deneración de atributos

- Luego se aplicaron estas funciones tanto al conjunto de datos de entrenamiento como al de prueba. Un ejemplo de una fila generada con las funciones anteriores es el siguiente:

Id	4
Category	1
Topic	Education
Source	EL UNIVERSAL
Headline	UNAM capacitará a maestros para aprobar prueba...
Text	UNAM capacitará a maestros para aprobar prueba...
Link	http://www.eluniversal.com.mx/articulo/nacion/...
%MayusculasHeadLine	0.0980392
#ResultadosGoogle	63500
#ResultadosGoogleNews	10
0ResultadosGoogleNews	0
#PalabrasPositivas	12
%PalabrasPositivas	0.0718563
#PalabrasNegativas	8
%PalabrasNegativas	0.0479042
Palabras unicas/palabras totales	0.568862
#Mistakes	1
#Numeros	0
%Numeros	0
#Comillas	0
%Comillas	0
#SignosInterrogación	0
%SignosInterrogación	0
%SignosExclamación	0
#SignosExclamación	0
Name: 3, dtype: object	

Fig 10 Ejemplo de una fila del dataset después de la generación de atributos

## Para la aproximación por Redes neuronales bidireccionales y LSTM usando bag of words:

- Se removieron las *stopwords* del título y el cuerpo de la noticia, para esto se utilizó la lista de *stopwords* en español proporcionada por la librería de procesamiento de lenguaje natural *nltk*<sup>1</sup>.
- Se realizaron procesos de tokenización y vectorización las palabras, un ejemplo de un cuerpo de noticia pre-procesado y listo para ser aplicado a un modelo es el siguiente:

```
The encoding for document
incluirá palabra lady diccionario idioma español definición mujer problemática
españa presidente real academia española diario villanueva conferencia prensa in
cluirá término lady palabra diccionario idioma español diario lady servirá defin
ir mujer problemática mujer causa problemas palabra utilizan idioma inglés cast
ellano significado contadas palabras idioma inglés utilizan español significado
globalización personas comenzado adoptar términos anglosajones utilizan signifi
cado real caso expresión lady significará idioma original dama inglés usará def
inir mujer problemática acostumbra causar problemas alborotos gente decirle lad
y fémica cause escándalo agresiva provoque tipo problema término dejara conside
rarse palabra exclusiva idioma inglés incluida diccionario lengua española pasa
ra palabra oficial castellano significado distinto confesó villanueva presentó
medios definición oficial aparecerá diccionarios señalando lady anglosajón angl
isc difficilis problematica ferox agresiva number mujer excesivamente problemát
ica number mujer causa problemas alborotos number mujer tiende causar conflicto
s agresiva interesar cholas descubren donitas bimbo comer sirven maquillarse nu
mber mujer guía instintos animales importa crear conflictos presidente miembros
mexicanos propuso incluir palabra meses análisis finalmente aceptada comité térm
ino origen México usará países hispana videos lady circulado sirvieron crear pe
rfecta definición servirá resumir definir hembra problemática diario reveló encu
entran analizando idea incluir término lord diccionario equivalente definición
masculina lady
is : [2257, 589, 1276, 1510, 1001, 159, 3932, 49, 3367, 86, 6, 160, 549, 391,
3933, 7842, 145, 75, 2257, 1277, 1276, 589, 1510, 1001, 159, 3933, 1276, 2036,
1833, 49, 3367, 49, 314, 136, 589, 1378, 1001, 332, 4726, 2037, 11608, 360, 100
1, 332, 1378, 159, 2037, 7843, 10, 2914, 3368, 550, 11609, 1378, 2037, 168, 13,
819, 1276, 11610, 1001, 1278, 193, 332, 2915, 1833, 49, 3367, 5927, 1379, 136,
7844, 22, 2566, 1276, 5928, 7845, 518, 2916, 5929, 92, 154, 1277, 2917, 5930, 5
89, 1511, 1001, 332, 2918, 1510, 2258, 391, 11611, 589, 138, 4726, 2037, 3369,
958, 7842, 437, 41, 3932, 138, 2919, 11612, 669, 1276, 11613, 11614, 11615, 116
16, 11617, 2916, 1, 49, 5931, 3367, 1, 49, 314, 136, 7844, 1, 49, 3934, 1379, 1
179, 2916, 7846, 11618, 1834, 11619, 5932, 1002, 2920, 11620, 1, 49, 3935, 1162
1, 670, 820, 590, 1179, 6, 551, 28, 1835, 1512, 589, 42, 740, 11622, 3936, 107
9, 1277, 420, 2, 2915, 65, 4727, 775, 1276, 11623, 11624, 590, 2921, 3932, 203
6, 11625, 1833, 5933, 3367, 3933, 260, 421, 3370, 125, 1512, 1277, 7847, 1510,
1279, 3932, 7848, 1276]
```

Fig 11 Ejemplo de una noticia tokenizada

## IV. Modelado: En esta fase, se seleccionan y aplican las técnicas de modelado que sean pertinentes al problema, y se calibran sus parámetros a valores óptimos. Durante este paso entrenamos varios modelos de aprendizaje

<sup>1</sup> Librería disponible en <https://www.nltk.org/>

automático usando principalmente tres técnicas: Random Forest, Gradient Boosting y Redes Neuronales Bidireccionales.

**Para la aproximación por Random Forest y Gradient Boosting usando variables de estilo generadas del texto:**

- Se usó la técnica de validación cruzada para escoger los mejores hiperparámetros tanto para Random Forest como para Gradient Boosting, de la siguiente forma:

Para la validación cruzada de Random Forest se usaron los siguientes parámetros:

```
'n_estimators': [40, 60, 100, 150],  
'max_features': [9, 12, 14, 16, 20, 22],  
'max_depth' : [None, 3, 10, 20],  
'criterion' : ['gini', 'entropy']
```

Obteniendo el siguiente resultado

```
-----  
Mejores hiperparámetros encontrados (cv)  
-----  
{'criterion': 'entropy', 'max_depth': 10, 'max_features': 12, 'n_estimators': 100} : 0.7874628753562462 f1
```

*Fig 12 Resultado de mejores hiperparámetros con grid search CV*

Mientras que para GradientBoosting se usaron los siguientes parámetros:

```
'n_estimators': [40, 60, 100, 150],  
'max_features': [9, 12, 14, 16, 20, 22],  
'max_depth' : [None, 3, 10, 20],  
'learning_rate': [0.01, 0.025, 0.05],
```

Obteniendo el siguiente resultado:

```
-----  
Mejores hiperparámetros encontrados (cv)  
-----  
{'learning_rate': 0.05, 'max_depth': 3, 'max_features': 12, 'n_estimators': 40} : 0.7790397959023836 f1
```

*Fig 13 Resultado de mejores hiperparámetros con grid search CV*

- V. Evaluación:** En esta etapa se evaluaron los modelos propuestos con las métricas de precisión y de valor-F. Obteniendo para el mejor modelo un 76.27% de precisión y 76% de valor-F, siendo Random Forest la técnica de aprendizaje automática utilizada.



**VI. Despliegue:** Como se propuso en la arquitectura anteriormente, se logró construir un *chat-bot* usando la tecnología de Amazon Lex<sup>2</sup> y Amazon Lambda<sup>3</sup> para el despliegue en la nube. Además, se montó un API en Flask<sup>4</sup> que recibe los parámetros desde AWS Lambda y calcula las variables de estilo necesarias para ser evaluadas en el modelo final (Random Forest). Finalmente, también se hizo uso de un API auxiliar para calcular las variables que necesitaran técnicas de *web scraping*.

## 5.2 Resultados esperados

Se espera una herramienta que sirva como ayuda en el proceso de detección de noticias falsas en el idioma español por medio de aprendizaje automático. En especial, se espera que la solución sea de fácil uso para el usuario final. Por ello, se propuso realizar una integración con la red social Facebook por medio de su plataforma de mensajería instantánea Messenger, para que se pueda hacer una verificación instantánea sin salir de la Facebook, donde día a día se generan noticias falsas de todo tipo.

## 6. VALIDACIÓN

### 6.1 Métodos

Se utilizaron las siguientes métricas para evaluar el desempeño de los modelos:

- Precisión: esta métrica permite medir el porcentaje de casos en los que el modelo ha acertado.
- Valor-f: esta métrica combina las medidas de precisión, que mide calidad del modelo y exhaustividad, que compara la cantidad de casos clasificados como verdaderos positivos sobre todo lo que realmente era positivo. Se escogió esta métrica ya que es importante mantenerse lejos de falsos positivos y falsos negativos y el valor-f logra llegar a un balance entre la precisión y la exhaustividad.

### 6.2 Validación de resultados

A continuación, se muestran las métricas resultantes para las predicciones realizadas con el conjunto de prueba para cada uno de los modelos construidos:

Modelo	Precisión	Valor-f
Random Forest	76.27 %	76 %
Gradient Boosting	73.89 %	74 %
Deep Learning	58,98 %	53 %

Tabla 4 Comparación de resultados

<sup>2</sup> Más información sobre Amazon Lex en <https://aws.amazon.com/es/lex/>

<sup>3</sup> Más información sobre Amazon Lambda en <https://aws.amazon.com/es/lambda/>

<sup>4</sup> Librería disponible en <https://flask.palletsprojects.com/>



## 7. CONCLUSIONES

### 7.1 Discusión

Se logró construir un prototipo viable para la detección de noticias falsas en el idioma español que podrá ser utilizado mediante la interfaz de Facebook Messenger para mayor usabilidad.

El uso de variables alternas a procesar directamente el texto de la noticia permitió aumentar en un 17% la exactitud del modelo en comparación con realizarlo con aprendizaje profundo basado en redes neuronales bidireccionales y LSTM (59%).

Las variables referentes al número de coincidencias del título de la noticia en Google y en Google News tuvieron alta relevancia, por lo cual las técnicas de web scraping resultaron beneficiosas.

El código fuente, scraper y datos utilizados se encuentran en el siguiente link: <https://github.com/dianyQuintero/FakeNewsDetectorSpanish>

### 7.2 Trabajo futuro

Como trabajo futuro, se sugieren los siguientes añadidos que se considera que aportarían valor a la solución:

- **Más variables para la predicción:** Se propone utilizar el link de la noticia con el fin de generar más variables que permitan aumentar la precisión del modelo, por ejemplo, el número de anuncios publicitarios o estilo de la página web utilizando técnicas de web scraping.
- **Uso de imágenes:** Utilizar las imágenes que contenga la noticia para obtener mejores resultados.
- **Alertas:** Generar alertas que le permitan al usuario emitir juicios de valor sobre la veracidad de la noticia.

## 8. REFERENCIAS

Aragón, M. E., Jarquín, H., Gómez, M. M. Y., Escalante, H. J., Villaseñor-Pineda, L., Gómez-Adorno, H., ... & Posadas-Durán, J. P. (2020, September). Overview of mex-a3t at iberlef 2020: Fake news and aggressiveness analysis in mexican spanish. In Notebook Papers of 2nd SEPLN Workshop on Iberian Languages Evaluation Forum (IberLEF), Malaga, Spain.

Herrera Cabello, J. (2020). La verificación de noticias y el problema de la inmediatez: análisis de cobertura de una noticia falsa en medios digitales chilenos. Anuario

Electrónico de Estudios en Comunicación Social “Disertaciones”, 13(1), 138-158. Doi: <https://doi.org/10.12804/revistas.urosario.edu.co/disertaciones/a.7246>

Ministerio de Tecnologías de la Información y las Comunicaciones. (2020, 8 abril). ¿Por qué las noticias falsas también son un riesgo real? Recuperado de: <https://mintic.gov.co/portal/inicio/Sala-de-Prensa/Noticias/126509:Por-que-las-noticias-falsas-tambien-son-un-riesgo-real>.

Posadas-Durán, J. P., Gómez-Adorno, H., Sidorov, G., & Escobar, J. J. M. (2019). Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent & Fuzzy Systems*, 36(5), 4869-4876.

Villena, J. (2016, 2 agosto). CRISP-DM: La metodología para poner orden en los proyectos. <https://www.sngular.com/>. <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>.

ZHOU, X., & ZAFARANI, R. (2020, julio). A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities (N.o 1). <https://doi.org/10.1145/3395046>.