



**UNIVERSIDAD DE GUAYAQUIL  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA DE INGENIERÍA EN TELEINFORMÁTICA**

**TRABAJO DE TITULACIÓN  
PREVIO A LA OBTENCIÓN DEL TÍTULO DE  
INGENIERA EN TELEINFORMÁTICA**

**ÁREA  
TECNOLOGÍA DE LOS ORDENADORES**

**TEMA  
“DESARROLLO DE UN PROTOTIPO WEB PARA  
GENERACIÓN DE SINOPSIS DE NOTICIAS  
RELACIONADAS AL COVID-19 MEDIANTE ALGORITMO  
NLP Y ENFOQUE PRAGMÁTICO, FCI 010-2021”**

**AUTORA  
TAYNYS AMANCHA DANIELA ALEJANDRA**

**DIRECTOR DEL TRABAJO  
ING. COMP. ACOSTA GUZMÁN IVÁN LEONEL, MSIG.**

**GUAYAQUIL, SEPTIEMBRE 2022**



## ANEXO XI.- FICHA DE REGISTRO DE TRABAJO DE TITULACIÓN

### FACULTAD DE INGENIERÍA INDUSTRIAL CARRERA INGENIERÍA EN TELEINFORMÁTICA



<b>REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA</b>			
<b>FICHA DE REGISTRO DE TRABAJO DE TITULACIÓN</b>			
<b>TÍTULO Y SUBTÍTULO:</b>	Desarrollo de un prototipo web para generación de sinopsis de noticias relacionadas al covid-19 mediante algoritmo NLP y enfoque pragmático, FCI 010-2021		
<b>AUTOR(ES)</b> (apellidos/nombres):	Taynys Amancha Daniela Alejandra		
<b>TUTOR /REVISOR</b> (apellidos/nombres):	Ing. Comp. Acosta Guzmán Iván Leonel, MSIG. / Ing. Telecom. Veintimilla Andrade Miguel Ángel, Mg.		
<b>INSTITUCIÓN:</b>	Universidad de Guayaquil		
<b>UNIDAD/FACULTAD:</b>	Facultad Ingeniería Industrial		
<b>MAESTRÍA/ESPECIALIDAD:</b>			
<b>GRADO OBTENIDO:</b>	Ingeniería en Teleinformática		
<b>FECHA DE PUBLICACIÓN:</b>	28 de septiembre del 2022	<b>No. DE PÁGINAS:</b>	182
<b>ÁREAS TEMÁTICAS:</b>	Tecnología de los Ordenadores		
<b>PALABRAS CLAVES/ KEYWORDS:</b>	Prototipo Web, raspado web, NPL, IA. Web Prototyping, web scraping, NPL, AI.		
<b>RESUMEN:</b>			
<p>El presente estudio se construyó a partir de la situación problemática en cuanto a la proliferación de fake news del Covid-19 generados en la comunidad. Por ello, este estudio tiene como principal objetivo desarrollar un prototipo web para generación de sinopsis de noticias relacionadas al Covid-19 mediante algoritmo NLP y enfoque pragmático, se realizó un análisis de herramientas tecnológicas más usadas para escoger el más adecuado en el desarrollo del prototipo. Adicional, se emplearon las metodologías cualitativa y cuantitativa, utilizando como instrumento de investigación la entrevista y encuesta basados en datos estadísticos, la metodología del prototipado para el desarrollo web y metodología exploratoria. Llegando a la conclusión de que a través del algoritmo LUNH se obtiene la validación de funcionalidad en el prototipo dando los siguientes resultados rouge-1: 53,846% recuerdo(r), 100% precisión(p), 70% medida(f); rouge-2: 51,61% r, 100% p, 68,09% f; rouge-l: 53,85% r, 100% p, 70% f. Una vez finalizado el desarrollo del prototipo web y etapas de pruebas se obtuvo una interfaz interactiva para el usuario haciendo la presentación de resultados esperados.</p>			

## ABSTRACT

This study was built from the problematic situation regarding the proliferation of Covid-19 fake news generated in the community. Therefore, the main objective of this study is to develop a web prototype to generate synopses of news related to Covid-19 using NLP algorithm and pragmatic approach, an analysis of the most used technological tools was performed to choose the most appropriate in the development of the prototype. Additionally, qualitative and quantitative methodologies were used, using as research instrument the interview and survey based on statistical data, the prototyping methodology for web development and exploratory methodology. The conclusion was that through the LUNH algorithm the validation of functionality in the prototype was obtained, giving the following results rouge-1: 53.846% recall(r), 100% precision(p), 70% measure(f); rouge-2: 51.61% r, 100% p, 68.09% f; rouge-l: 53.85% r, 100% p, 70% f. Once the development of the web prototype and testing stages were completed, an interactive interface was obtained for the user, making the presentation of expected results.

ADJUNTO PDF:	SI <input checked="" type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Teléfono: 593-979459189	E-mail: danielaataynysamancha@outlook.es
CONTACTO CON LA INSTITUCIÓN:	Nombre: Ing. Ramón Maquilón Nicola, Mg.	
	Teléfono: 593- 2658128	
	E-mail: direcciónTi @ug.edu.ec	



## ANEXO XII DECLARACIÓN DE AUTORÍA Y DE AUTORIZACIÓN DE LICENCIA GRATUITA



**INTRANSFERIBLE Y NO EXCLUSIVA PARA EL USO NO  
COMERCIAL DE LA OBRA CON FINES NO ACADÉMICOS**

**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

---

LICENCIA GRATUITA INTRANSFERIBLE Y NO COMERCIAL DE LA OBRA  
CON FINES NO ACADÉMICOS

Yo, **TAYNYS AMANCHA DANIELA ALEJANDRA**, con C.C. No. **0972456638**, certifico que los contenidos desarrollados en este trabajo de titulación, cuyo título es **“DESARROLLO DE UN PROTOTIPO WEB PARA GENERACIÓN DE SINOPSIS DE NOTICIAS RELACIONADAS AL COVID-19 MEDIANTE ALGORITMO NLP Y ENFOQUE PRAGMÁTICO, FCI 010-2021”** son de mi absoluta propiedad y responsabilidad, en conformidad al Artículo 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN\*, autorizo la utilización de una licencia gratuita intransferible, para el uso no comercial de la presente obra a favor de la Universidad de Guayaquil.

*Daniela Taynys f*

**TAYNYS AMANCHA DANIELA ALEJANDRA**  
C.C.No. 0952476638



**ANEXO VII.- CERTIFICADO PORCENTAJE DE SIMILITUD**



**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

Habiendo sido nombrado **ING. COMP. ACOSTA GUZMÁN IVÁN LEONEL, MSIG**, tutor del trabajo de titulación certifico que el presente trabajo de titulación ha sido elaborado por **TAYNYS AMANCHA DANIELA ALEJANDRA, C.C.: 0952476638**, con mi respectiva supervisión como requerimiento parcial para la obtención del título de INGENIERA EN TELEINFORMÁTICA.

Se informa que el trabajo de titulación: **DESARROLLO DE UN PROTOTIPO WEB PARA GENERACIÓN DE SINOPSIS DE NOTICIAS RELACIONADAS AL COVID-19 MEDIANTE ALGORITMO NLP Y ENFOQUE PRAGMÁTICO, FCI 010-2021**, ha sido orientado durante todo el periodo de ejecución en el programa Antiplagio TURNITIN quedando el 9 % de coincidencia.

[https://ev.turnitin.com/app/carta/en\\_us/?lang=en\\_us&o=1898562374&s=1&u=1133714199](https://ev.turnitin.com/app/carta/en_us/?lang=en_us&o=1898562374&s=1&u=1133714199)

QUICK SUBMIT | NOW VIEWING: NEW PAPERS ▼

Submit						
<input type="checkbox"/>	AUTHOR	TITLE	SIMILARITY	FILE	PAPER ID	DATE
<input type="checkbox"/>	Daniela Alexandra Ta...	Tesis Daniela Taynys - Versión 16	9%		1898562374	12-Sep-2022

**Tesis Daniela Taynys - Versión 16**

**ORIGINALITY REPORT**

<b>9%</b>	<b>9%</b>	<b>1%</b>	<b>3%</b>
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS



Firmado electrónicamente por:  
**IVAN LEONEL  
ACOSTA GUZMAN**

**ING. COMP. ACOSTA GUZMÁN IVÁN LEONEL, MSIG.**  
**DOCENTE TUTOR**  
**C.C. 0914940812**  
**FECHA: LUNES 12 SEPTIEMBRE 2022**



**ANEXO VI. - CERTIFICADO DEL DOCENTE-TUTOR DEL**



**TRABAJO DE TITULACIÓN  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

Guayaquil, 12 de septiembre del 2022

Sr (a).

**Ing. Annabelle Lizarzaburu Mora, MG.**

Director (a) de Carrera Ingeniería en Telemática / Telemática

**FACULTAD DE INGENIERÍA INDUSTRIAL DE LA UNIVERSIDAD DE  
GUAYAQUIL**

Ciudad. -

De mis consideraciones:

Envío a Ud. el Informe correspondiente a la tutoría realizada al Trabajo de Titulación **DESARROLLO DE UN PROTOTIPO WEB PARA GENERACIÓN DE SINOPSIS DE NOTICIAS RELACIONADAS AL COVID-19 MEDIANTE ALGORITMO NLP Y ENFOQUE PRAGMÁTICO, FCI 010-2021**, de la estudiante TAYNYS AMANCHA DANIELA ALEJANDRA, indicando que ha cumplido con todos los parámetros establecidos en la normativa vigente:

- El trabajo es el resultado de una investigación.
- El estudiante demuestra conocimiento profesional integral.
- El trabajo presenta una propuesta en el área de conocimiento.
- El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se adjunta el certificado de porcentaje de similitud y la valoración del trabajo de titulación con la respectiva calificación.

Dando por concluida esta tutoría de trabajo de titulación, **CERTIFICO**, para los fines pertinentes, que la estudiante está apta para continuar con el proceso de revisión final.

Atentamente,



Firmado electrónicamente por:  
**IVAN LEONEL  
ACOSTA GUZMAN**

---

**ING. COMP. ACOSTA GUZMÁN IVÁN LEONEL, MSIG.  
DOCENTE TUTOR  
C.C. 0914940812  
FECHA: LUNES 12 SEPTIEMBRE 2022**



**ANEXO VIII.- INFORME DEL DOCENTE REVISOR**



**FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

Guayaquil, 23 de septiembre del 2022

Sr (a).

**Ing. Annabelle Lizarzaburu Mora, MG.**

Director (a) de Carrera Ingeniería en Teleinformática / Telemática

**FACULTAD DE INGENIERÍA INDUSTRIAL DE LA UNIVERSIDAD DE  
GUAYAQUIL**

Ciudad. -

De mis consideraciones:

Envío a Ud. el informe correspondiente a la REVISIÓN FINAL del Trabajo de Titulación: **“DESARROLLO DE UN PROTOTIPO WEB PARA GENERACIÓN DE SINOPSIS DE NOTICIAS RELACIONADAS AL COVID-19 MEDIANTE ALGORITMO NLP Y ENFOQUE PRAGMÁTICO, FCI 010-2021”** de la estudiante **TAYNYS AMANCHA DANIELA ALEJANDRA**. Las gestiones realizadas me permiten indicar que el trabajo fue revisado considerando todos los parámetros establecidos en las normativas vigentes, en el cumplimiento de los siguientes aspectos:

Cumplimiento de requisitos de forma:

El título tiene un máximo de 22 palabras.

La memoria escrita se ajusta a la estructura establecida.

El documento se ajusta a las normas de escritura científica seleccionadas por la Facultad.

La investigación es pertinente con la línea y sublíneas de investigación de la carrera.

Los soportes teóricos son de máximo 5 años.

La propuesta presentada es pertinente.

Cumplimiento con el Reglamento de Régimen Académico:

El trabajo es el resultado de una investigación.

El estudiante demuestra conocimiento profesional integral.

El trabajo presenta una propuesta en el área de conocimiento.

El nivel de argumentación es coherente con el campo de conocimiento.

Adicionalmente, se indica que fue revisado, el certificado de porcentaje de similitud, la valoración del tutor, así como de las páginas preliminares solicitadas, lo cual indica el que el trabajo de investigación cumple con los requisitos exigidos.

Una vez concluida esta revisión, considero que el estudiante está apto para continuar el proceso de titulación. Particular que comunicamos a usted para los fines pertinentes.

Atentamente,



firmado electrónicamente por:  
**MIGUEL ANGEL  
VEINTIMILLA  
ANDRADE**

**ING. TELEC. VEINTIMILLA ANDRADE MIGUEL ÁNGEL, MG**

**C.C: 0922668017**

**FECHA: 23 SEPTIEMBRE DEL 2022**

## **Dedicatoria**

Este trabajo no sería posible sin el apoyo incondicional primordialmente de Dios quien me ha iluminado proporcionándome grandes experiencias, guiándome hacia el camino correcto llevando al logro de obtener esta meta.

A mi esposo, quien me ha apoyado en todo aspecto, así como: la ayuda de sostenimiento en cada proyecto o herramientas a utilizar en cada semestre concurrido, dándome alientos de no darme por vencida en momentos difíciles, por la paciencia que tuvo en comprender mis estudios y dejar que mis sueños se cumplieran para así llevar un excelente ejemplo a seguir en mi familia.

A mi hija, quien ha sido una gran parte fundamental para lograr este paso, por permitir luchar por el bienestar de ella, por ser mi fuente de motivación e inspiración para superarme cada día más y así obtener un futuro mejor en nuestras vidas.

A mis docentes de la carrera por sus enseñanzas y lograr formarme una persona de buen perfil y ética profesional, quienes dieron sus experiencias para llenarme de sabiduría y así poder lograr llegar hasta el punto en el que me encuentro.

A mi tutor de tesis quien ha sido una de las personas que me ha guiado de forma correcta para el proceso de tesis, en base a su sabiduría ha logrado direccionar mis conocimientos hacia la formalización de este proyecto.

Ellos son quienes dieron fortalezas y enseñanzas día a día de no dejarse vencer ante las adversidades y con ello dieron paso a que logre la meta propuesta inicialmente.



## **Agradecimiento**

Agradezco primeramente a Dios por inicialmente darme esta oportunidad de estar en la una de las mejores universidades y de permitir no desaprovechar el tiempo, avanzando en cada ciclo semestral, a pesar de las barreras que se tuvo que pasar para ser una gran profesional.

Gracias a mi abuela quien ha estado incondicionalmente junto a mí dándome fortalezas en mi meta profesional para obtención de un futuro mejor, dando un perfil ejemplar y así poder desarrollar muchos objetivos más a planes futuros, sin embargo, debo reconocer que sin su apoyo hubiese sido imposible mi finalización de esta meta tan esperada.

Debo agradecer de manera especial y sincera al Ing. Comp. Acosta Guzmán Iván Leonel, Msig. por su aceptación en realizar esta tesis bajo su dirección, debo recalcar que gracias a su apoyo y confianza supo acaparar el desarrollo de mis conocimientos e implementarlo, a su vez debo reconocer que me ha sabido guiar de manera correcta en este proceso, así como también su aporte de tiempos extras que generaba junto a mí para el logro de mejores resultados y por darme los medios suficientes para llevar a cabo las actividades en el tiempo previamente planificado.

## **Índice general**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
	Introducción	1

## **Capítulo I**

### **El problema**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
1.1.	Planteamiento del problema	3
1.2.	Formulación del problema	3
1.3.	Sistematización del problema	3
1.4.	Objetivos	4
1.4.1.	Objetivo General	4
1.4.2.	Objetivos Específicos	4
1.5.	Justificación	4
1.6.	Delimitación del problema	5
1.7.	Alcance	5
1.8.	Premisa de la investigación	6
1.9.	Operacionalización	6
1.9.1.	Variable Independiente	6
1.9.2.	Variable Dependiente	6

## **Capítulo II**

### **Marco teórico**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
2.1.	Antecedentes del estudio	8
2.2.	Fundamentación teórica	11
2.2.1.	Orígenes de la inteligencia artificial	12
2.2.2.	Inteligencia Artificial	13
2.2.3.	Técnicas de la Inteligencia Artificial	13
2.2.4.	Aprendizaje Automático (Machine Learning)	14
2.2.4.1.	Ciclo de vida del Machine Learning	14
2.2.4.2.	Niveles de madurez de Machine Learning	15
2.2.4.3.	Librerías usadas de Machine Learning	15
2.2.4.4.	Tipos de Aprendizajes de Machine Learning	15

2.2.4.4.1.	Aprendizaje Supervisado	16
2.2.4.4.2.	Aprendizaje No Supervisado	16
2.2.4.4.3.	Aprendizaje por Refuerzo	16
2.2.4.4.4.	Aprendizaje Profundo (Deep Learning)	16
2.2.4.5.	Algoritmos de Aprendizaje No Supervisado	17
2.2.4.5.1.	Lineales	17
2.2.4.5.1.1.	Clustering	17
2.2.4.5.1.2.	K-Medias	17
2.2.4.5.1.3.	Análisis factorial	17
2.2.4.5.1.4.	Análisis de componentes principales (PCA)	18
2.2.4.5.2.	No Lineales	18
2.2.4.5.2.1.	Escala multidimensional (MDS)	18
2.2.4.5.2.2.	Mapeo de características isométricas (ISOMAP)	18
2.2.4.5.2.3.	Incrustación localmente lineal (LLE)	19
2.2.4.5.2.4.	Mapas Heissien (HLLE)	19
2.2.4.5.2.5.	Incrustación espectral	19
2.2.4.5.2.6.	Incrustación de vecinos estocásticos distribuidos en t (t-SNE)	19
2.2.5.	Lenguaje Natural	19
2.2.6.	Natural Language Processing (NLP)	19
2.2.6.1.	Librerías utilizadas en NLP	19
2.2.6.1.1.	NLTK	19
2.2.6.1.2.	spaCY	20
2.2.6.1.3.	JSON	20
2.2.6.1.4.	ReGex o re	20
2.2.6.1.5.	Scikit Learn (SKLEARN)	20
2.2.6.2.	Pre-Procesamiento de datos en NLP	20
2.2.6.2.1.	Técnicas para el Pre-Procesamiento de datos	20
2.2.6.2.2.	Tipos de variables en el Pre-Procesamiento de datos	21
2.2.6.3.	Técnica de procesamiento de datos en NLP	21
2.2.6.3.1.	Algoritmos de Stemming o derivación	22
2.2.6.4.	Clasificación en NLP	23
2.2.6.5.	Tipos de generación de resúmenes de textos en NLP	23
2.2.6.5.1.	Resúmenes extractivos	24
2.2.6.5.2.	Resúmenes abstractivos	24

2.2.6.5.3.	Resúmenes extractivos vs. Resumen abstractivos	24
2.2.6.5.4.	Modelos para generación de resúmenes de textos extractivos	24
2.2.6.5.4.1.	Basado en matriz de similitud	24
2.2.6.5.4.2.	Basado en frecuencia de las palabras por el índice de TF	25
2.2.6.5.4.3.	Basado en frecuencia de las palabras por el índice de TFIDF	25
2.2.6.5.4.4.	Basado en métrica de LUHN	26
2.2.6.6.	Arquitectura de Procesamiento de Lenguaje Natural	26
2.2.6.6.1.	Pragmático	26
2.2.6.6.2.	Elementos de la pragmática	27
2.2.6.6.3.	Importancias que incluye el componente pragmático	27
2.2.7.	Web Scraping	28
2.2.7.1.	Éticas del Web Scraping	28
2.2.8.	Prototipo	28
2.2.9.	La web	28
2.2.10.	Internet	29
2.2.11.	Desarrollo web	29
2.2.12.	Lenguajes de Programación	29
2.2.12.1.	Python	32
2.2.12.2.	C	32
2.2.12.3.	Java	33
2.2.12.4.	C ++	34
2.2.13.	Frameworks	35
2.2.14.	Entorno de Desarrollo Integrado (IDE)	39
2.2.14.1.	Entorno de desarrollo integrado (IDE) de escritorio	39
2.2.14.1.1.	PyCharm	39
2.2.14.1.2.	Visual Studio Code	39
2.2.14.1.3.	SPyder	40
2.2.14.1.4.	Jupyter Notebook	41
2.2.14.1.5.	Sublime Text	41
2.2.14.2.	Entorno de desarrollo integrado (IDE) en la nube	42
2.2.14.2.1.	Google colab (Google Colaboratory)	42
2.2.14.2.2.	IBM Watson Studio	43
2.2.14.2.3.	Azure Machine Learning	43
2.2.14.2.4.	Amazon Machine Learning (AWS)	44

2.2.15.	Fundamentación legal	44
2.2.15.1.	Constitución de la República del Ecuador	44
2.2.15.2.	Reglamento de Régimen Académico Consejo de Educación Superior	45
2.2.15.3.	Ley Orgánica de Comunicación	45
2.2.15.4.	Ley Orgánica de Telecomunicaciones	45
2.2.15.5.	Consultas de Normas Internacionales del uso de la IA	45
2.2.15.5.1.	Resolución del Parlamento Europeo sobre uso policial de la Inteligencia Artificial	46

### **Capítulo III**

#### **Metodología**

<b>N°</b>	<b>Descripción</b>	<b>Pág.</b>
3.1.	Diseño de la investigación	45
3.2.	Propuesta tecnológica	45
3.2.1.	Descripción del proceso metodológico	45
3.2.2.	Etapas de la investigación científica	45
3.3.	Tipos de investigación	46
3.4.	Metodología de investigación	47
3.4.1.	Metodología Bibliográfica	47
3.4.1.1.	Fases de la metodología bibliográfica	47
3.4.2.	Metodología Exploratoria	48
3.4.3.	Metodología Cualitativa	49
3.4.4.	Metodología Cuantitativa	49
3.4.5.	Metodología Mixta	50
3.4.6.	Metodología del Prototipado	50
3.5.	Técnicas de investigación	52
3.5.1	Encuesta	52
3.5.2	Entrevista	53
3.6.	Descripción del procedimiento metodológico	53
3.6.1.	Población	53
3.6.2.	Muestra	53
3.7.	Análisis de las encuestas	54
3.8.	Resumen de la entrevista	67
3.9.	Construcción del sitio web	69

3.9.1	Prototipo de baja fidelidad para ingreso del sitio web	69
3.9.2	Prototipo de baja fidelidad para acceso a categorías de páginas oficiales del sitio web	70
3.9.3	Prototipo de baja fidelidad con acceso a página oficial a artículos noticiarios.	70
3.9.4	Prototipo de baja fidelidad al realizar una petición en el buscador	71
3.9.5	Prototipo de baja fidelidad redireccionando a la página original del contenido de interés escogido por el usuario	71
3.9.6	Prototipo de baja fidelidad para acceso a Correo Electrónico	72
3.9.7	Prototipo de baja fidelidad para acceso a Facebook	72
3.9.8	Prototipo de baja fidelidad para acceso a Twitter	73
3.9.9	Prototipo de baja fidelidad para acceso a Instagram	73
3.9.10	Prototipo de baja fidelidad para acceso a Youtube	74
3.9.11	Prototipo de baja fidelidad con acceso a la Google Maps	74
3.9.12	Modelo Arquitectónico de prototipo web de baja fidelidad y media fidelidad Diseño de la investigación	75
3.9.13	Prototipo de alta fidelidad para ingreso del sitio web	76
3.9.14	Prototipo de alta fidelidad para acceso a categorías de páginas oficiales del sitio web	76
3.9.15	Prototipo de alta fidelidad con acceso a página oficial a artículos noticiarios	77
3.9.16	Prototipo de alta fidelidad al realizar una petición en el buscador	77
3.9.17	Prototipo de alta fidelidad redireccionando a la página original del contenido de interés escogido por el usuario	78
3.9.18	Prototipo de alta fidelidad para acceso a Correo Electrónico	78
3.9.19	Prototipo de alta fidelidad para acceso a Facebook	79
3.9.20	Prototipo de alta fidelidad para acceso a Twitter	79
3.9.21	Prototipo de alta fidelidad para acceso a Instagram	80
3.9.22	Prototipo de alta fidelidad para acceso a Youtube	80
3.9.23	Prototipo de alta fidelidad con acceso a la Google Maps	81
3.10.	Construcción de estructura para la ejecutar la técnica de Web Scraping	81
3.10.1.	Importar librerías	81
3.10.2.	Definir la abstracción de items	82

3.10.3.	Definir spider	82
3.10.4.	Definir dominios permitidos y url semilla	82
3.10.5.	Definir reglas	82
3.10.6.	Definir la función parse	83
3.10.7.	Analizar el árbol HTML	83
3.10.8.	Construcción lógica para descargar la información y llenado de ítems	83
3.10.9.	Retorno de los ítems llenos	84
3.10.10.	Probar el código	84
3.10.11.	Resultado utilizando la técnica de Web Scraping	84
3.11.	Construcción del modelo de Machine Learning para resúmenes de textos	84
3.11.1.	Resultados de métricas ROUGE para evaluación de los funcionamientos en los algoritmos de NLP para resúmenes de textos automáticos tipo extractivos	85
3.11.2.	Técnicas de preprocesamiento	86
3.11.2.1.	Instalación de librerías	86
3.11.2.2.	Cargar la data	86
3.11.2.3.	Explorar la data	87
3.11.2.4.	Imputar la data	88
3.11.2.5.	Normalizar los datos	88
3.11.2.6.	Seleccionar los datos	89
3.11.3.	Técnicas de Procesamiento	89
3.11.3.1.	Definición de función para preprocesamiento de la data	89
3.11.3.2.	Aplicación de función para el preprocesamiento de la data	90
3.11.3.3.	Definición de función para el cálculo de frecuencia de palabras en la data	90
3.11.3.4.	Aplicación de función para el cálculo de frecuencia de palabras en la data	91
3.11.3.5.	Definición de función para ponderación de palabras frecuentes en la data	91
3.11.3.6.	Definición de función para resúmenes de textos de la data	93
3.11.3.7.	Resultados finales	95
3.11.3.8.	Guardado de la data en formato JSON	96
3.11.3.9.	Importación al prototipo web	97

3.11.3.10.	Diagrama Arquitectónico para la construcción del modelo NLP de generación de resúmenes de textos tipo extractivo	98
3.12.	Conclusiones	99
3.13.	Recomendaciones	100
	Anexos	101
	Bibliografía	138



<b>Nº</b>	<b>Descripción</b>	<b>Pág.</b>
1	Delimitación del problema	5
2	Operacionalización	7
3	Elementos de la pragmática utilizados en el estudio	27
4	Características de Python	32
5	Características de C	33
6	Características de Java	33
7	Características de C++	34
8	Comparativa de Frameworks utilizados en Python	36
9	Características de PyCharm	39
10	Característica de Visual Studio Code	40
11	Características de Spyder	40
12	Características de Jupyter	41
13	Característica de Sublime Text	41
14	Especificaciones de Google Colab	42
15	Características de Google Colab	42
16	Características de IBM Watson Studio	43
17	Características de Azure Machine Learning	43
18	Características de Amazon Machine Learning	44
19	Tipos de Investigación utilizados en el estudio	46
20	Edad	54
21	Género	55
22	Lugar que reside de la zona 8 del Ecuador	56
23	Medios que la población utiliza para informarse sobre noticias del Covid-19	56
24	Nivel de dificultad al buscar información en los portales de información	58
25	Nivel de aceptabilidad en la construcción de un sitio web para centralizar noticias del covid-19	58
26	Nivel de consideración al contar con un sitio web que permita verificar las noticias oficiales del covid-19	59
27	Nivel de estimación al contar con un sitio web que permita evitar la compartición de noticias falsas	60

28	Nivel de consideración de importancia al contar con un sitio web que permita centralizar las noticias del covid-19 y que la misma redireccione al sitio oficial	61
29	Tiempo disponible para noticias referentes al covid-19 en dispositivos digitales	62
30	Encuestados que tengan al menos un dispositivo	63
31	Encuestados que tengan al menos un dispositivo	64
32	Encuestados indicando su tipo de ancho de banda	65
33	Cantidad de personas conectados al mismo plan de internet	66
34	Resultados de las Métricas de ROUGE	86

## Índice de figuras

<b>Nº</b>	<b>Descripción</b>	<b>Pág.</b>
1	Historia de la IA	12
2	Clasificación de la IA	13
3	Ciclo de vida de ML	14
4	Niveles de madurez de Machine Learning basada en 750 empresas	15
5	Clasificación de los tipos de aprendizajes	16
6	Mapa mental del método factorial	17
7	Pasos de la función del algoritmo PCA	18
8	Técnicas de preprocesamiento de datos en NLP	20
9	Tipos de variables existentes para el pre-procesamiento de datos en NLP	21
10	Mapa mental de la Clasificación de NLP	23
11	Tipos de generación de resúmenes	23
12	Modelo arquitectónico de modelo basado en matriz de similitud	25
13	Modelo arquitectónico de modelo basado en frecuencia de palabras	25
14	Sistema Arquitectónico NPL	26
15	Variantes importantes del componente pragmático en el lenguaje natural	27
16	Personalización para generar el espectro IEEE.	30
17	Espectro IEEE evaluando los mejores lenguajes de programación	30
18	Análisis estadístico de los mejores lenguajes de programación	30
19	Datos de calificaciones del índice de la comunidad de programación	31
20	Proceso de la investigación científica	46
21	Características de la metodología mixta	50
22	¿Para qué sirve prototipar?	52
23	Edad de los encuestados	55
24.	Género de los encuestados	55
25.	Residentes de la Zona 8	56
26	Medios que la población utiliza para informarse sobre noticias del Covid-19	57
27	Nivel de dificultad de los portales que ofrece información reciente del Covid-19	58

28	Nivel de aceptabilidad en la creación de un sitio web para centralización de noticias oficiales referentes al covid-19	59
29	Nivel de consideración al contar con un sitio web que permita verificar las noticias oficiales del covid-19	60
30	Nivel de consideración sobre la existencia de un sitio web que permita centralizar las noticias de múltiples paginas oficiales para evitar los Fake News	61
31	Importancia sobre la existencia de un sitio web que permita centralizar las noticias del covid-19 y que la misma redireccione al sitio oficial	62
32	Tiempo de disponibilidad para revisión de noticias de covid-19 en dispositivos digitales. Información obtenida.	63
33	Encuestados que tengan al menos un dispositivo digital	64
34	Tipo de enlace que emplea habitualmente	65
35	Ancho de banda que utiliza	66
36	Cantidad de personas conectados al mismo plan de internet	67
37	Prototipo de baja fidelidad ingreso al sitio web	69
38	Prototipo de baja fidelidad despliegue de páginas oficiales	70
39	Prototipo de baja fidelidad con acceso a noticias oficiales referentes al covid-19	70
40	Prototipo de baja fidelidad con resultados de búsqueda ingresada por el usuario	71
41	Prototipo de baja fidelidad redireccionando a la página original del contenido de interés escogido por el usuario	71
42	Prototipo de baja fidelidad para acceso a correo electrónico	72
43	Prototipo de baja fidelidad para acceso a Facebook	72
44	Prototipo de baja fidelidad para acceso a Twitter	73
45	Prototipo de baja fidelidad para acceso a Instagram	73
46	Prototipo de baja fidelidad para acceso a Youtube	74
47	Prototipo de baja fidelidad con acceso a Google Maps	74
48	Modelo Arquitectónico del prototipo web de alta y media fidelidad	75
49	Prototipo de alta fidelidad ingreso al sitio web	76
50	Prototipo de alta fidelidad despliegue de páginas oficiales	76

51	Prototipo de alta fidelidad con acceso a noticias oficiales referentes al covid-19	77
52	Prototipo de alta fidelidad con resultados de búsqueda del texto ingresado por el usuario	77
53	Prototipo de alta fidelidad redireccionando a la página original del contenido de interés escogido por el usuario	78
54	Prototipo de alta fidelidad para acceso a correo electrónico	78
55	Prototipo de alta fidelidad para acceso a Facebook	79
56	Prototipo de alta fidelidad para acceso a Twitter	79
57	Prototipo de alta fidelidad para acceso a Instagram	80
58	Prototipo de alta fidelidad para acceso a Youtube	80
59	Prototipo de alta fidelidad con acceso a Google Maps	81
60	Importación de librerías a utilizar para web scraping	81
61	Definición de ítems para web scraping	82
62	Definición de spider para web scraping	82
63	Definición de dominios permitidos y url semilla para web scraping	82
64	Definición de reglas para web scraping	83
65	Definición de función parse para web scraping	83
66	Análisis de árbol HTML para web scraping	83
67	Construcción lógica para descargar la información hacia el web scraping	84
68	Retorno de los ítems llenos para web scraping	84
69	Prueba de la estructura de código para web scraping	84
70	Visualización de resultados utilizando la técnica de web scraping	84
71	Resultados de métricas Rouge a fin de evaluar los funcionamientos en los algoritmos de NLP para resúmenes de textos tipo extractivo	85
72	Resultados de métricas Rouge a fin de evaluar el algoritmo de NLP para resúmenes de textos tipo abstractivo	85
73	Instalación de librerías para el algoritmo NLP aplicando técnicas de preprocesamiento	86
74	Importación de la data a Google Colab aplicando técnicas de preprocesamiento	87
75	Carga de la data aplicando técnicas de preprocesamiento	87
76	Explorar la data aplicando técnicas de preprocesamiento	87

77	Consulta de valores perdidos aplicando técnicas de preprocesamiento	88
78	Gráfico de valores perdidos aplicando técnicas de preprocesamiento	88
79	Imputación de la data aplicando técnicas de preprocesamiento	88
80	Verificación de valores perdidos aplicando técnicas de preprocesamiento	88
81	Normalizar la data aplicando técnicas de preprocesamiento	89
82	Visualización del contenido crudo y procesado de la data aplicando técnicas de preprocesamiento	89
83	Seleccionar columna de la data aplicando técnicas de preprocesamiento	89
84	Definición de función para el algoritmo aplicando técnicas de procesamiento	90
85	Aplicación de función para el preprocesamiento del algoritmo empleando técnicas de procesamiento	90
86	Definición de función para cálculo de frecuencia de palabras del algoritmo empleando técnicas de procesamiento	90
87	Resultados de gráficos estadísticos aplicando función para cálculo de frecuencia de palabras del algoritmo empleando técnicas de procesamiento	91
88	Definición de función para ponderación de oraciones en la data aplicando técnicas de procesamiento	93
89	Definición de función para resúmenes de textos en la data aplicando técnicas de procesamiento	94
90	Visualización del Dataframe con funciones implementadas aplicando técnicas de procesamiento	95
91	Aplicación de conteo de palabras de una columna en la data aplicando técnicas de procesamiento	95
92	Verificación de la dimensionalidad inicial y final del contenido aplicando técnicas de procesamiento	95
93	Aplicación de métricas ROUGE de toda la data para resultados por cada fila	96
94	Resultados de cada fila aplicando métricas ROUGE	96

95	Resultados del valor promedio de la data aplicando las métricas de ROUGE	96
96	Data lista para descarga en formato Json	96
97	Descarga en formato Json	97
98	Importación de archivo al prototipo web	97
99	Visualización previa de datos a importar al prototipo web	97
100	Modelo arquitectónico general del prototipo	97
101	Diagrama arquitectónico para la construcción del modelo NLP de generación de resúmenes de textos tipo extractivo	98

<b>Índice de Anexos</b>		
<b>Nº</b>	<b>Descripción</b>	<b>Pág.</b>
1	Entrevista realizada a profesionales de Desarrollo web	102
2	Entrevista realizada a profesionales de Inteligencia Artificial	108
3	Modelo de la encuesta realizada a los ciudadanos de la Zona 8 de la Provincia del Guayas	112
4	Manual de usuario del Prototipo Web	117
5	Justificación legal	133





**ANEXO XIII.- RESUMEN DEL TRABAJO DE  
TITULACIÓN (ESPAÑOL)  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**



---

**“DESARROLLO DE UN PROTOTIPO WEB PARA GENERACIÓN DE  
SINOPSIS DE NOTICIAS RELACIONADAS AL COVID-19 MEDIANTE  
ALGORITMO NLP Y ENFOQUE PRAGMÁTICO, FCI 010-2021”**

**Autor:** Taynys Amancha Daniela Alejandra.

**Tutor:** Ing. Comp. Acosta Guzmán Iván Leonel. MSIG.

### **Resumen**

El presente estudio se construyó a partir de la situación problemática en cuanto a la proliferación de fake news del Covid-19 generados en la comunidad. Por ello, este estudio tiene como principal objetivo desarrollar un prototipo web para generación de sinopsis de noticias relacionadas al Covid-19 mediante algoritmo NLP y enfoque pragmático, se realizó un análisis de herramientas tecnológicas más usadas para escoger el más adecuado en el desarrollo del prototipo. Adicional, se emplearon las metodologías cualitativa y cuantitativa, utilizando como instrumento de investigación la entrevista y encuesta basados en datos estadísticos, la metodología del prototipado para el desarrollo web y metodología exploratoria. Llegando a la conclusión de que a través del algoritmo LUNH se obtiene la validación de funcionalidad en el prototipo dando los siguientes resultados rouge-1: 53,846% recuerdo(r), 100% precisión(p), 70% medida(f); rouge-2: 51,61% r, 100% p, 68,09% f; rouge-l: 53,85% r, 100% p, 70% f. Una vez finalizado el desarrollo del prototipo web y etapas de pruebas se obtuvo una interfaz interactiva para el usuario haciendo la presentación de resultados esperados.

**Palabras claves:** Prototipo Web, raspado web, NPL, IA.



**ANEXO XIV.- RESUMEN DEL TRABAJO DE  
TITULACIÓN (INGLÉS)  
FACULTAD DE INGENIERÍA INDUSTRIAL  
CARRERA INGENIERÍA EN TELEINFORMÁTICA**

---



**“DEVELOPMENT OF A WEB PROTOTYPE TO GENERATE NEWS  
SYNOPSSES RELATED TO COVID-19 USING NLP ALGORITHM AND  
PRAGMATIC APPROACH, FCI 010-2021”**

**Author:** Taynys Amancha Daniela Alejandra

**Advisor:** Ing. Comp. Acosta Guzmán Iván Leonel. MSIG.

**Abstract**

This study was built from the problematic situation regarding the proliferation of Covid-19 fake news generated in the community. Therefore, the main objective of this study is to develop a web prototype to generate synopses of news related to Covid-19 using NLP algorithm and pragmatic approach, an analysis of the most used technological tools was performed to choose the most appropriate in the development of the prototype. Additionally, qualitative and quantitative methodologies were used, using as research instrument the interview and survey based on statistical data, the prototyping methodology for web development and exploratory methodology. The conclusion was that through the LUNH algorithm the validation of functionality in the prototype was obtained, giving the following results rouge-1: 53.846% recall(r), 100% precision(p), 70% measure(f); rouge-2: 51.61% r, 100% p, 68.09% f; rouge-l: 53.85% r, 100% p, 70% f. Once the development of the web prototype and testing stages were completed, an interactive interface was obtained for the user, making the presentation of expected results.

**Keywords:** Web Prototyping, web scraping, NPL, AI.

## **Introducción**

Este estudio toma en cuenta minimizar la transmisión de noticias falsas conocidos como fake news acerca del covid-19 en la población de la zona-8 (Guayaquil, Samborondón y Durán), se debe tomar en cuenta que en la actualidad el método de transmisión de información crece constantemente, a través, de medios digitales por medio de mecanismos tecnológicos como el internet de las cosas(IoT) y está muy ligado a la Inteligencia Artificial (IA), dado que son herramientas que consideran múltiples funciones y una considerar en el proyecto es la técnica del Procesamiento del Lenguaje Natural (NPL) con enfoque pragmático implementado en un prototipo web.

La IA está causando sensación en el IoT. Gartner establece que en el 2022 más del 80% de las planificaciones IoT implementarán la IA comparado con el 10% que se tenía en los años anteriores. Por lo que, considera que los pronósticos a partir de estas tecnologías son capaces de reducir un tiempo estimado de 20% al 50%, permitiendo aumentar el tiempo de actividad, en cuanto a la disponibilidad que tiene un ordenador estaría en un 10% al 20% y con ello se reducen los costos de mantenimiento de un 5% al 10%. Por último, considera la predicción basada en el aprendizaje automático que está dentro de la IA. (Deloitte, 2022).

Según datos de la Universidad Tecnológica de Pereira, El NLP es el área de conocimiento de la IA y la lingüística computacional cuyo principal objetivo es entender, interpretar y procesar de manera inteligente la comunicación entre máquinas y el lenguaje humano, mediante estas herramientas muchos desarrolladores pueden complementar tareas tales como: resumen automático, corrección de textos, recuperación de la información, sistemas inteligentes demostrativos, etc. (Tintinago et al., 2018).

Talamé et al. (2020) argumenta que el componente pragmático de NLP incluye el estudio de un contexto para lograr exponer su uso a la representación final. El mismo influye el entendimiento del contexto en donde las oraciones son emitidas para otorgar un significado llevando un paso más allá de la estructura del lenguaje, cabe resaltar que para alcanzar este objetivo es indispensable procesar la información de manera que se normalicen los textos y poder estructurarlos y analizarlos de manera accesible.

Para poder alcanzar dicho fin se tiene en consideración dos preceptos principales de la IA que son sus fundamentos y sus ramas; Dentro de los fundamentos se encuentran la lingüística computacional sirviendo como fuente en NLP y unas de las ramificaciones en cuanto a las ciencias de la informática es la IA, la misma ha generado mayor impacto a nivel mundial y uno de los motivos es por la pandemia del covid-19.

Un prototipo web es un modelo inicial a partir de un desarrollo original del diseño de un sitio de web, siendo un método sencillo y económico por lo que es caracterizado en tomar un proceso de etapas hasta lograr el objetivo propuesto, creando una referencia visual, a través, de su contenido organizando de mejor manera la información a nivel de página web (Barbosa de Lira et al., 2020).

La iniciativa se basa en proporcionar a la población la accesibilidad de un sitio web para centralizar resúmenes de noticias y validar la veracidad de las mismas, minimizando las transmisiones de fake news generadas comúnmente en medios no confiables. Lo más importante es que actualmente se cuenta con mecanismos y herramientas tecnológicas que brinda una manera más estable con respecto al almacenamiento, procesamiento y difusión de grandes cantidades de datos.

La presente investigación está estructurada por el capítulo I, II, III y basa en lo siguientes componentes:

Capítulo I: Comprende el planteamiento de la problemática encontrada, formulación y realización de preguntas de investigación con respecto al problema, descripción de los objetivos de la investigación tanto general como específicos, justificación, delimitaciones, premisas correspondientes a la investigación y la medición de características del estudio por medio de la operacionalización de variables.

Capítulo II: Engloba antecedentes del estudio, marco teórico mostrando las definiciones básicas utilizadas durante el desarrollo del proyecto y marco jurídico en el que se cubrirá el presente estudio.

Capítulo III: Contiene el diseño y desarrollo de propuesta de la investigación, tipo de investigación llevados a cabo a partir de la obtención de datos con gráficos estadísticos, técnicas e instrumentos aplicados en encuestas y entrevistas, diseño de la propuesta desarrollando aspectos metodológicos empleados finalizando con las conclusiones y recomendaciones del trabajo de titulación obtenidas a partir de todos los procedimientos dados.

## **Capítulo I**

### **El problema**

#### **1.1. Planteamiento del problema**

En la comunidad de la zona-8, se presentan caos informativos con respecto a la difusión de noticias falsas conocidos como fake news mediante medios de comunicaciones, redes sociales o personal no profesional, lo cual se genera pánico y confusión en la población acerca del Covid-19 volviéndose un alto impacto en el ámbito de la salud, económico, psicológica, entre otros aspectos. Por otro lado, esta comunidad no cuenta con un instrumento informático que le facilite el acceso de manera rápida con información fiable y no extensa, tomando en cuenta que muchos cibernautas pretenden llegar a una información de su interés, por lo que evitan lecturas de textos amplios.

Bayas (2021), describe que en la pandemia de Covid-19 se surgieron no solo problemas de sistemas de salud, sino también la proliferación de desinformación sobre temas como tratamientos, síntomas, orígenes, formas de prevención, entre otros aspectos. Por lo que es necesario cubrir más frentes en la comunicación y así poder transmitir información con eficiencia este fenómeno. Esto permitió reconocer que durante el primer semestre Ecuador contaba con una calificación cualitativa de bajos criterios, generando la necesidad de analizar las informaciones engañosas que circulan en varios medios de transmisión comunicativa como: redes sociales, sitios web, entre otros.

#### **1.2. Formulación del problema**

¿Es posible mediante el uso de la IA centralizar las noticias relevantes acerca del covid-19, por medio de algoritmos para lectura rápida de textos amplios?

#### **1.3. Sistematización del problema**

En la población de la zona-8 conformada por tres cantones Guayaquil, Samborondón y Durán, se realizó un análisis del estado de recepción de noticias sobre el Covid-19, llegando a una determinada conclusión en que se pueden utilizar nuevos mecanismos tecnológicos para minimizar las fake news y así elevar conocimientos de páginas confiables que generen transmisiones de noticias referentes al Covid-19.

En base al estudio se determina contestar las siguientes preguntas:

- ¿Cuáles serían las plataformas más adecuadas para el uso de la IA?
- ¿Qué algoritmo es necesario para la generación de resúmenes de textos amplios?
- ¿Qué metodología de desarrollo de software es la más adecuada para elaborar un prototipo?

- ¿Qué grado de aceptación existen en los pobladores de la zona-8 referente a contar con una solución tecnológica que centralice las noticias relacionadas al covid-19?

## **1.4. Objetivos**

### **1.4.1. Objetivo General**

Desarrollar un prototipo web para generación de sinopsis de noticias relacionadas al covid-19 mediante algoritmo NLP y enfoque pragmático, dirigido a los habitantes de la zona-8.

### **1.4.2. Objetivos Específicos**

- Levantar información bibliográfica para comprender la funcionalidad de la estructura gramatical del nivel pragmático de NLP por medio de una investigación científica.
- Mediante el uso de la técnica de Web Scraping extraer informaciones relacionadas a noticias del Covid-19.
- Identificar el algoritmo NLP más adecuado para la generación de resúmenes de textos.
- Diseñar un diagrama arquitectónico para la construcción del modelo NLP de generación de resúmenes de textos.
- Validar el funcionamiento del prototipo por medio de una métrica para la evaluación del modelo NLP.

## **1.5. Justificación**

En la actualidad el mundo entero está pasando por el transcurso de experimentar cambios en todo ámbito como: personas, procesos y entornos tecnológicos, llevando a cabo el surgimiento de la necesidad de suministrar soluciones con herramientas informáticas para cualquier internauta.

En el presente proyecto de investigación se enfoca en la construcción de un prototipo web que centralicen las noticias confiables publicadas en múltiples medios de comunicación tanto nacional e internacional, logrando obtener textos principales para que el usuario logre determinar un tema de su interés y redirija al sitio oficial donde se aloja la noticia completa, esto permitirá que la comunidad de la zona-8, cuente con un medio de acceso a la información que apoye a reducir el nivel de ansiedad, preocupación o estrés permitiendo obtener respuestas confiables en situaciones del covid-19.

### 1.6. Delimitación del problema

En la siguiente Tabla 1 se expone el campo, área y aspecto del tema que se efectuará en el presente proyecto.

**Tabla 1** Delimitación del problema

<b>Campo</b>	Aplicación de tecnología de la información
<b>Área</b>	Tecnología de los ordenadores
<b>Aspecto</b>	Construcción de sitio web para transmisiones de noticias confiables.
<b>Tema</b>	Desarrollo de un prototipo web para generación de sinopsis de noticias relacionadas al covid-19 mediante algoritmo NLP y enfoque pragmático, FCI 010-2021.

---

*Información tomada de la investigación directa. Elaborado por: Taynys Amancha Daniela*

El desarrollo del proyecto comprende los beneficios que muestran las herramientas de hardware y software que se utilizaron para la creación del prototipo web siendo un entorno de prueba, permitiendo una interacción amigable con el usuario, incentivando a la comunidad a minimizar riesgos provocados por los fake news, generando un favorecimiento a la comunidad de la zona-8 que deseen tener una información fiable sobre noticias referentes al covid-19, formando un ágil manejo y una interfaz fácil, manejable y entendible ante cualquier usuario de esta población que acceda al sitio web, teniendo la ventaja de que a planificaciones futuras se pueda implementar más módulos que ayuden a mejorar el prototipo web.

### 1.7. Alcance

Para el desarrollo del presente trabajo se va a tomar en consideración lo siguiente:

- La comparativa de diversos lenguajes de programación, frameworks, editores de códigos que será vital para elección del más adecuado a implementar.
- Se determinará una cantidad para rastrear páginas web en noticias oficiales relacionadas al covid-19 entre nacionales e internacionales.
- Se establece una fecha para la extracción de información de las páginas web en noticias oficiales del covid-19 del presente año 2022.
- No se considera implementar el pase en producción debido a que en la Facultad de Ingeniería Industrial no posee servidor web disponible.

### **1.8. Premisa de la investigación**

El desarrollo de un prototipo web poniendo en disponibilidad noticias referentes al covid-19 aplicando un enfoque pragmático, permitiendo que los usuarios puedan visualizar e interactuar con el sitio de manera fácil, rápida y entendible, dejando puesta en práctica la habilidad de manejo en herramientas tecnológicas como la IA proporcionando una utilización ideal de contenidos en la interfaz web en la cual brindará una mejor recepción de páginas oficiales en la zona-8 reduciendo la problemática inicialmente dada en esta comunidad.

### **1.9. Operacionalización**

Según Espinoza (2020), operacionalizar una variable, consiste en definir de manera clara como se observará y medirá cada característica del estudio.

En este proyecto de investigación se identificaron las siguientes variables y son:

#### **1.9.1. Variable Independiente:**

- Accesibilidad a la información relacionada al covid-19.

#### **1.9.2. Variable Dependiente:**

- Sitio web amigable para consultas de noticias.
- Disponibilidad de acceso a Internet.



**Tabla 2** Operacionalización

<b>Variables</b>	<b>Dimensión</b>	<b>Indicadores</b>	<b>Técnicas y/o instrumentos</b>
<b>Independiente:</b>			
<b>Accesibilidad a la información relacionada al covid-19</b>	<ul style="list-style-type: none"> <li>• Accesibilidad a noticias.</li> </ul>	<ul style="list-style-type: none"> <li>• Facilidad de hallazgo de noticias de interés.</li> <li>• Accesibilidad con dispositivos digitales para revisión de noticia.</li> </ul>	<ul style="list-style-type: none"> <li>• Encuesta a pobladores.</li> <li>• Encuesta de número de horas al día que tiene acceso a plataformas para revisión de noticias.</li> </ul>
<b>Dependiente:</b>			
<b>Sitio web amigable para consulta de noticias.</b>	<ul style="list-style-type: none"> <li>• Facilidad de uso de sitio web.</li> </ul>	<ul style="list-style-type: none"> <li>• Nivel de dificultad de uso de sitios web.</li> </ul>	<ul style="list-style-type: none"> <li>• Encuesta a pobladores.</li> </ul>
<b>Disponibilidad de acceso a Internet.</b>	<ul style="list-style-type: none"> <li>• Acceso a internet.</li> </ul>	<ul style="list-style-type: none"> <li>• Tipos de accesibilidad a internet.</li> <li>• Capacidad de ancho de banda del acceso a internet.</li> </ul>	<ul style="list-style-type: none"> <li>• Encuestas a habitantes para detectar con qué acceso cuentan.</li> <li>• Encuestas a pobladores o estadística de censo poblacional del INEC.</li> </ul>

Elaborado por: Taynys Amancha Daniela Alejandra

## **Capítulo II**

### **Marco teórico**

#### **2.1. Antecedentes del estudio**

Para el desarrollo de este trabajo se efectuó el levantamiento de datos con respecto al conocimiento de conllevar un entorno pragmático en distintas herramientas tecnológicas como: el Procesamiento de Lenguaje Natural (NPL - Natural Language Process) en el campo que comprende la Inteligencia Artificial (IA) y para la implementación del prototipo web.

Como indica Pissinati et al. (2019), en su artículo científico destaca que un prototipo web permite crear un ambiente interactivo entre el ser humano y ordenador, en el cual el usuario se siente activo al interaccionar en los procesos por medio de distintas vistas, destacando especificaciones de manera clara y sencilla para el entendimiento del público en general, además describe el uso de la investigación metodológica aplicada y basada en modelo del prototipado creando surgimientos de pasos tales como: comunicar, planear, crear, probar y la consolidación del software web. Los estudiantes en el transcurso académico del área de tecnología en la actualidad cuentan con herramientas que permiten impulsar el desarrollo rápido de prototipos, empleando Python y Django es posible construir sitios web estáticos o dinámicos que almacenan información de las transacciones de manera permanente en base de datos, logrando con ello construir soluciones que ante los usuarios son fáciles de manejar (Silva, 2021).

Según Condori (2012), en su artículo de desarrollo web indica que el framework Django es muy potente siendo utilizado en grandes compañías a nivel mundial, de manera que con esta herramienta tecnológica se establezca un entorno de desarrollo simple, eficiente y rápido. Django establece un desarrollo de código limpio por lo que promueve buenas prácticas de desarrollo web utilizando la arquitectura MTV(Model-Template-View), logrando que esta forma de trabajar sea pragmática. Uno de los ejemplos de lanzamiento en el framework es el Campus Party de Colombia en el 2010, por lo que dio a conocer que esta herramienta es novedosa y de última generación pretendiendo alcanzar un nivel muy alto no solo en Latino América si no a nivel mundial.

Según Paradavisual (2019), describe que la combinación en torno al internet de las cosas (IoT) e IA son unos de los progresos tecnológicos en alcanzar a la industria de manera mundial, los mismos garantizan una experiencia totalmente enlazada al usuario, dotándose de conexión a internet e inteligencia software, permitiendo procesar un ecosistema de servicios a través de grandes cantidades de datos y automatizando sus funcionalidades, generando la comunicación hombre-máquina por medio de visualizaciones de datos en

tiempo real mediante dispositivos IoT (Teléfonos inteligentes, tabletas, asistentes inteligentes, etc.).

Por otro lado, según Noticias Parlamento Europeo (2021), describe que la IA es la destreza de que un ordenador presente las mismas capacidades que los seres humanos tales como: ser creativos, razonar, aprender, y planificar. Permitiendo que sus procedimientos tecnológicos perciban su entorno, de manera que tengan conexión con él, resuelvan diversas problemáticas y actúen con según su finalidad específica, a través, de su respectivo entrenamiento, es decir, la máquina recibe datos ya preparados de manera organizada, los procesa y da como resultado una respuesta al usuario que utilice dicha tecnología, esto permite destacar algunos tipos de IA como: asistentes virtuales, motores de búsqueda, etc.

González (2020), menciona que existen varios tipos de aplicaciones en la Inteligencia Artificial y las más destacadas son:

- **Procesamiento de imágenes:** Sistemas inteligentes automatizados.
- **Procesamiento de lenguaje natural:** Comunicación entre hombre-máquina.
- **Sistematización en la industria:** Aplicación en grandes sectores para automatización.
- **Incremento de robots no físicos:** Interacción entre máquina-máquina.

Para esta investigación se va a realizar uso de una de las aplicaciones de la IA catalogado como NPL.

Según Echeverri y Manjarrés (2020), destacan que NLP:

Es un área científica correspondiente a la informática, IA y la lingüística en donde estudia las interrelaciones entre los ordenadores y el lenguaje del ser humano, a través de la descomposición sintáctica, semántica, pragmática y morfológica; Se redactan reglas de reconocimiento buscando patrones estructurados, empleando formalidad gramatical de manera precisa. Estas reglas combinadas con los datos de información que son previamente almacenados en diccionarios informáticos, definen los modelos de reconocimiento en una letra, palabra u oración. Dando paso al NLP pudiendo reconocer imágenes, texto y voz.

En un artículo publicado por Camacho y Navarro (2020), menciona en su investigación de NLP con Django que la arquitectura está ligada por 5 niveles y son:

- **Fonológico:** Son unidades textuales solicitadas para el estudio de la jerga en el ser humano.
- **Morfológico:** Permite estudiar las formas de palabras, examinando sus significados en cuanto a la gramática o el vocabulario de la humanidad.

- **Sintáctico:** Examina la composición que tiene una oración y las palabras para unificarlas y formar una oración. Esta funcionalidad, conexión de las oraciones y palabras son a través de sus construcciones.
- **Semántico:** Permite descubrir todo lo que se refiere al significado de palabras que otorgan sentido a una o varias frases.
- **Pragmático:** Genera la forma de visualizar cada oración, permitiendo conocer en el estudio el significado de un contexto.

Alfonseca y Rodríguez (2004), indica que la pragmática parte de una anáfora dado a que se refiere a un antecedente y ambas partes tienen la misma referencia en el mundo real y las aplicaciones que se pueden dar son:

- Extracción de información
- Resúmenes automáticos
- Resúmenes Multi-documento
- Sistemas de traducción automática
- Búsquedas de respuestas a preguntas

Khder (2021), publicó en un artículo su investigación sobre la utilización de la técnica Web Scraping dando una perspectiva en el estado del arte, técnicas, enfoques y aplicación, implementada en el lenguaje de programación Python debido a su potente sintaxis y su accesibilidad en la curva de aprendizaje, favorable para la extracción de información usando tres fragmentos tales como: obtención de enlaces deseados en la web, extraer información para obtener datos deseados de los enlaces de origen y finalmente el guardado de información en un formato pudiendo ser: json, csv, entre otros. Por otro lado, una de las ventajas es que también pueden integrar un análisis visual informático y el NLP para replicar la forma en que los usuarios visualizan la información en la web.

Valle et al. (2019), publicaron en una revista la implementación en la generación de resúmenes tipos extractivos, por lo que radica en simplificar un texto de manera corta la información más destacada que puede contener distintos alojamientos de información pudiendo ser en la web o documentos, haciendo una utilización de un algoritmo por agrupación que se ubica en un método no supervisado y para la evaluación se utilizó las métricas de calidad ROUGE para evaluación del mismo.

Una tesina en licenciatura de informática perteneciente a la Universidad Nacional de la Plata impulso un proyecto en la realización de resúmenes extractivos de documentos y menciona que un resumen extractivo es creado a partir de la selección de unidades textuales

del original, mientras que un resumen abstractivo es reescribir lo que se entiende de la lectura, por lo que por medio de una comparativa determina que el adecuado es de tipo extractivo (Corvi, 2019).

## **2.2. Fundamentación teórica**

En la Universidad Técnica de Babahoyo un estudiante describe en su propuesta de tesis que, para tener excelencia y calidad administrativa y pedagógica, se debe manejar una atención directa con el usuario, utilizando un medio para interactuar estas pueden ser: páginas web, mensajerías instantáneas, chatbot, etc. Haciendo uso de la IA para que se desarrolle consultas en el área de atención por cada usuario de la institución (Urrutia, 2020).

En un trabajo de fin de grado de la Universidad de la Laguna un alumno propuso de tesis trabajar con un entorno colaborativo con cloud computing, Google colab, Neo4J Aura y Python para obtener los datos de la categoría gramatical para el texto procesado por la herramienta spaCy, buscando ejecutar análisis de NLP como: morfológico en el apartado del depurado y preprocesamiento de los textos, luego el análisis sintáctico para determinar las estructuras sintácticas, y por último el análisis semántico para comprender el significado de la frase, aprovechando así extraer las posibles relaciones y datos valiosos de los textos (Delgado, 2021).

La Universidad de Guadalajara publicó en la revista de Filosofía, Letras y Humanidades un trabajo de aproximación en aspectos de la Lingüística computacional que es vinculada a la informática, la IA, entre otros campos tecnológicos, simulando una capacidad humana con el computador, consistiendo en un modelo representativo poniendo como alcance maximizar la sintaxis y semántica a la pragmática, pudiendo hacer un análisis intencional de las locuciones (Casado, 2021).

En la Universidad Politécnica de Madrid un estudiante propuso la determinación de expresiones de odio y ofensivas, por lo que llega con un enfoque pragmático basándose en unigramas y patrones, utilizando un algoritmo de Inteligencia Artificial, el mismo es entrenado usando 20.000 tweets, por lo que comienzan con un preprocesamiento de los tweets, extracción y tokenización. Esta propuesta conlleva a una precisión del 87,4 % para tweets en ofensivos y no ofensivos y una precisión del 78,4 % para tweets en expresiones de odios, ofensivos y limpios (Marquinez, 2021).

### 2.2.1. Orígenes de la inteligencia artificial

En los años 300 a.C., Aristóteles expone un descubrimiento denominado Silogismos en el que destaca el manejo del pensamiento humano generando conclusiones razonables a partir de afirmaciones. Ctesibio de Alejandría a partir de los 250 a.C., efectuó la construcción del primer artefacto autocontrolado (Redacción España, 2019).

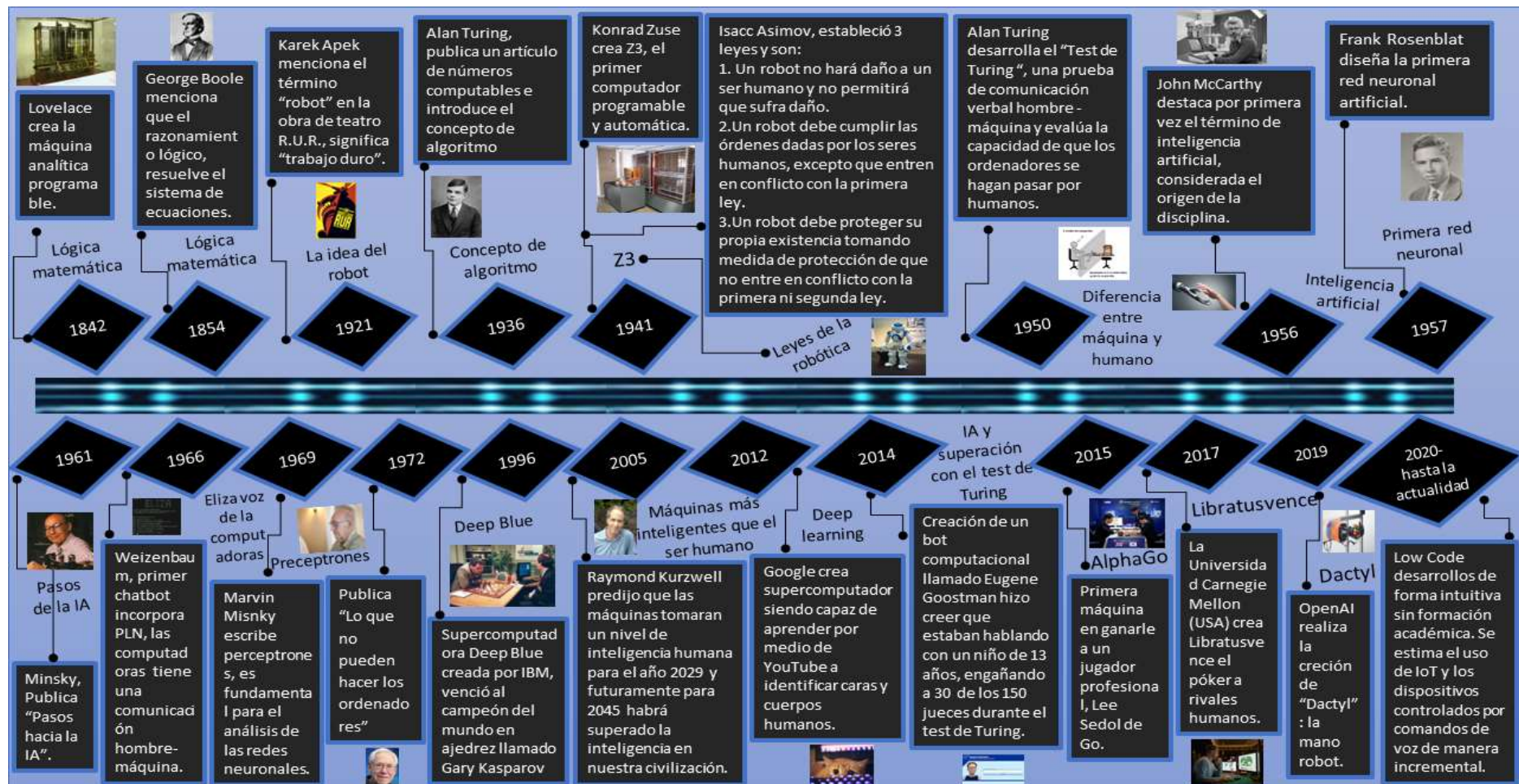


Figura 1. Historia de la IA. Información tomada de revistas (National Geographic, 2020) (Alva, 2020) (IT USER TECH & BUSSINES, 2020). Elaborado por Daniela Taynys.

### 2.2.2. Inteligencia Artificial

La IA es la capacidad que posee un ordenador en demostrar que tiene las mismas habilidades de un ser humano, dado que tiene la facultad de razonar, lleva un aprendizaje fluido, es creativo y planea de forma óptima determinadas tareas. Permitiendo que exista una comunicación con las áreas tecnológicas existentes, accediendo a la resolución de problemas para un fin específico. Está capacitado en poder ajustar su conducta en distinta medida para así analizar los resultados de hechos de manera anticipada y trabaja en forma autónoma. Por otro lado, no necesita ser indistinguible del comportamiento humano, o capaz de simular el razonamiento humano, lo que a veces se define como "IA fuerte" en el campo de la informática. Solo necesitamos un sistema informático para realizar tareas simples que tradicionalmente realizaban los seres humanos, como reconocer imágenes o procesar el lenguaje natural. Esto es lo que tradicionalmente define la "IA débil" (Verganti et al., 2020).



Figura 2. Clasificación de la IA. Información tomada de (Samart, 2017). Elaborado por Daniela Taynys.

### 2.2.3. Técnicas de la Inteligencia Artificial

La IA se compone de varias técnicas en función de satisfacer varias necesidades:

- Machine learning o aprendizaje automático (ML)
- Lógica difusa
- Vida artificial
- Sistemas expertos
- Sistemas basados en reglas
- Sistemas reactivos
- Minería de datos
- Redes bayesianas
- Ingeniería del conocimiento
- Redes neuronales artificiales
- Razonamiento basado en casos
- Técnicas de representación de conocimiento



- Redes semánticas
- Procesamiento del lenguaje natural. (Redacción APD, 2021).
- Lingüística computacional

### 2.2.4. Aprendizaje Automático (Machine Learning)

Raschka et al. (2020), expresan que Machine Learning (ML) es el estudio y desarrollo de enfoques que automatizan la toma de decisiones complejas, esto implica el uso de los ordenadores para analizar hechos pasados y construir un modelo que pueda descubrir reglas predictivas a partir de datos, dando paso a que las computadoras actualicen los algoritmos para predecir resultados óptimos de nuevos acontecimientos.

#### 2.2.4.1. Ciclo de vida del Machine Learning

Es una sucesión cíclica para el desarrollo de un proceso eficiente. Lo más importante es entender el problema y conocer el propósito del mismo. Por lo tanto, un buen resultado depende de la mejor comprensión del problema y para su resolución se debe crear un sistema de ML llamado modelo, el mismo es creado por medio de un entrenamiento para optimización de resultados y evaluación del modelo de su conjunto de datos (CENTRO EUROPEO DE POSTGRADO, 2021).



Figura 3. Ciclo de vida de ML. Información tomada de (CENTRO EUROPEO DE POSTGRADO, 2021) (Chhabra, 2022). Elaborado por Daniela Taynys.



#### 2.2.4.2. Niveles de madurez de Machine Learning

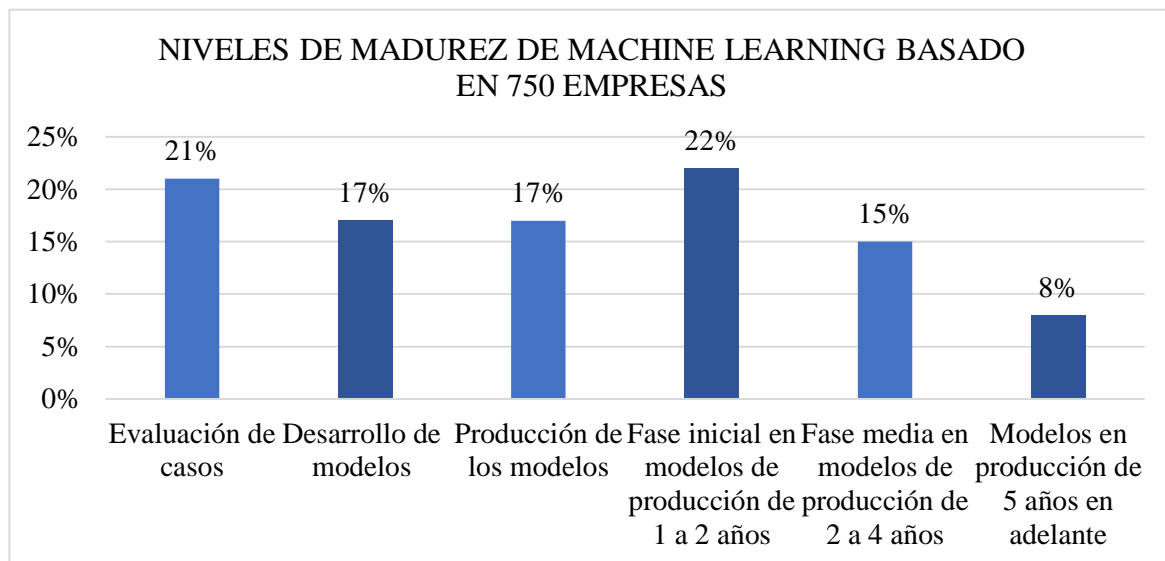


Figura 4. Niveles de madurez de Machine Learning basada en 750 empresas. Información tomada de artículo por (Hofesmann, 2021). Elaborado por Daniela Taynys.

De acuerdo con el gráfico estadístico No. 4 se puede observar que para un modelo sea implementado se sigue un procedimiento inicial como: evaluación de casos, desarrollo del modelo a implementar según la problemática, ejecución del modelo en términos de prueba y las evoluciones de fases en producción para que el modelo se siga entrenando y sea más efectivo.

#### 2.2.4.3. Librerías usadas de Machine Learning

##### 2.2.4.3.1. Pandas

Es un acrónimo que significa “Panel de datos” permitiendo cargar y manipular datos con el objetivo de trabajar con datos tabulares para la conversión de informaciones no estructuradas a estructuradas, depende del paquete NumPy. Por otro lado, contiene dos objetos principales que son: DataFrame (tabla de datos) y la Serie (columna o fila de datos). Proporcionando estructuras de información de manera ágil, con alto rendimiento, confiables y flexibles (Anand et al., 2020).

##### 2.2.4.3.2. Numpy

Según Moreno (2020), es un paquete de funciones matemáticas avanzadas conocido como “Numerical Python”, que permite implementar vectores o matrices multidimensionales con extensa capacidad y estructura de datos. Utilizado comúnmente para: el desarrollo de algoritmos de ML, matrices N-dimensionales, contiene integrales, generadores de números totalmente aleatorios, transformadas de Fourier, etc.

#### 2.2.4.4. Tipos de Aprendizajes de Machine Learning

Los métodos de ML se clasifican por los siguientes tipos:



Figura 5. Clasificación de los tipos de aprendizajes. Información tomada por la Investigación Directa. Elaborado por Daniela Taynys.

#### 2.2.4.4.1. Aprendizaje Supervisado

Es un método de enseñanza al algoritmo de lo que se desea y lo que debe aprender por sí mismo, donde se requiere la intervención humana para generar esta retroalimentación. En este aprendizaje los algoritmos utilizan datos organizados que se generan para entrenamiento, constando de variables de entrada (independientes) y salida (dependientes), pudiendo minimizar errores de mejor manera posible, algunos ejemplos son: regresión, clasificación, predicción, entre otros (Martínez, 2021).

#### 2.2.4.4.2. Aprendizaje No Supervisado

Es aquel que clasifica datos basado en el agrupamiento de patrones o clústeres y reducción de la dimensionalidad. El agrupamiento se encarga de relacionar los datos de entrada a varios grupos. Por otro lado, la reducción de la dimensionalidad encuentra aquellas variables de entrada que mantengan la mejor versión de la información de los datos. algunos ejemplos son: segmentación texto, identificación de correos no deseados, etc. (Martínez, 2021).

#### 2.2.4.4.3. Aprendizaje por Refuerzo

Permite que el algoritmo tenga una retroalimentación por medio de análisis de datos adquiriendo un mejor resultado. Permite tener un modelo de aprendizaje a través de prueba y error. Por lo que, a medida que transcurre el tiempo dicho algoritmo tomará una formación automatizada incrementándose de manera eficiente, un ejemplo claro es la observación de consecuencias (Ossa y Jaramillo, 2021).

#### 2.2.4.4.4. Aprendizaje Profundo (Deep Learning)

Ossa y Jaramillo (2021), indican que es aquella que utiliza redes neuronales en diferentes capas consecutivas para el aprendizaje de datos en forma iterativa, muy utilizado en buscar patrones de datos no estructurados. Este tipo de aprendizaje busca ser similar al

funcionamiento del cerebro humano, con la finalidad de que los ordenadores puedan ser entrenados y realizar abstracciones. Se utilizan mucho para reconocimiento de voz e imágenes.

#### 2.2.4.5. *Algoritmos de Aprendizaje No Supervisado*

##### 2.2.4.5.1. *Lineales*

###### 2.2.4.5.1.1. *Clustering*

Su objetivo es la agrupación, es decir encontrar diversos grupos de elementos en los datos según su calificación. Por lo cual, permite encontrar dentro de la estructura de datos la similitud de los elementos del mismo clúster, y así poderlos agrupar con los objetos del mismo grupo determinándolos lo más similares posibles. Las agrupaciones más comunes utilizadas son: K-Medias, Clusterización Jerárquica, Agrupación de exploración basada en la densidad (DBSCAN), Modelo de Agrupamiento Gaussiano (Roman, 2019).

###### 2.2.4.5.1.2. *K-Medias*

Encuentra y agrupa por clases los segmentos de datos que tienen alta similitud entre sí, es decir, que cuando más aproximado estén los puntos de datos tendrás más probabilidades de pertenecer al mismo clúster o grupo (Roman, 2019).

###### 2.2.4.5.1.3. *Análisis factorial*

Es una técnica que busca el número mínimo de dimensiones que demuestra la explicación de la extensa información contenida en los datos. Su funcionamiento consiste en la definición de la estructuración subyacente en una matriz de datos, considerando todas las variables y dimensiones (Fernández, 2011).

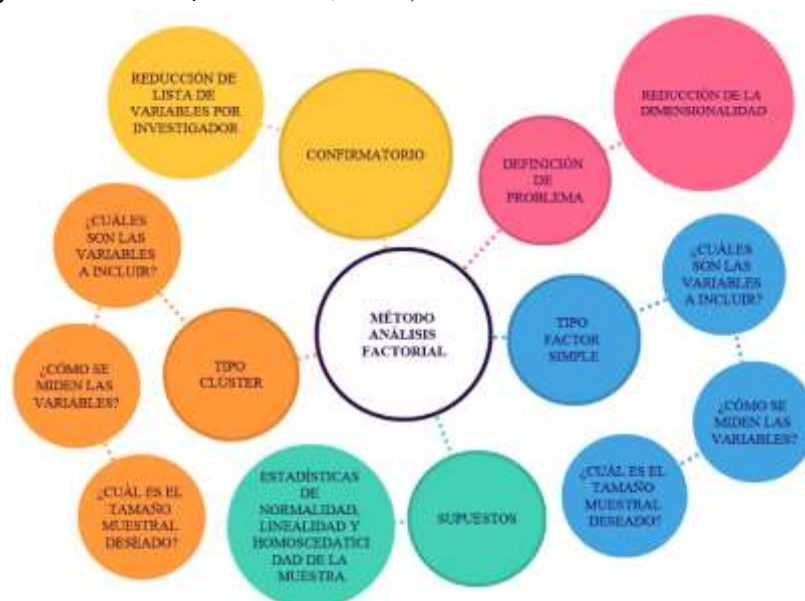


Figura 6. Mapa mental del método factorial. Información tomada de artículo (Fernández, 2011). Elaborado por Daniela Taynys.

Se puede destacar que para el método de análisis factorial se tiene en consideración como principal paso la definición del problema que contiene la reducción de la dimensionalidad, su segundo punto es la confirmación que es por medio del investigador y detalla un listado de variables a analizar, tercer punto se divide en dos tipos tales como: clúster y factor simple unificando preguntas importantes para medición e importancia de lo que se desea llegar y como finalización se detallan los supuestos implicando estadísticas de normalidad de la muestra escogida inicialmente.

#### 2.2.4.5.1.4. *Análisis de componentes principales (PCA)*

Es una forma con visión estadística en la que no existe correlación entre las características rotadas, pero permiten conservar la mayoría de cantidad en la varianza de datos originales. Realiza una transformación donde selecciona nuevos conjuntos de datos de  $n$  dimensiones a un subespacio de  $m$  dimensiones, para proceder a la reducción de redundancia de los datos, por lo que  $m$  tendrá menor dimensionalidad en el texto (Mamani, 2019).

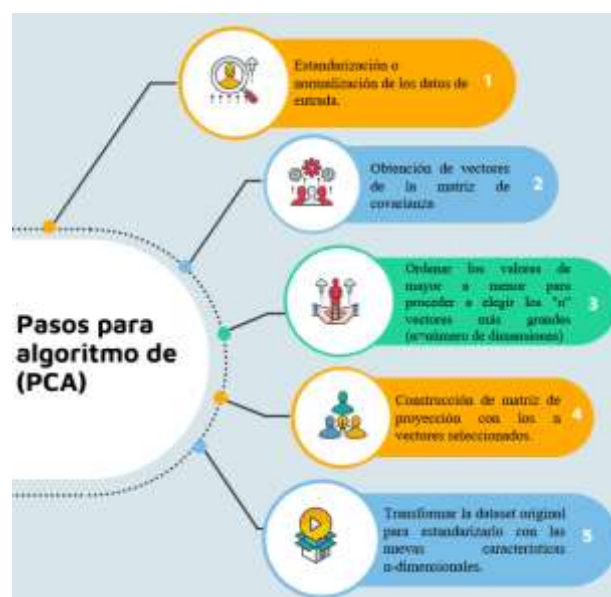


Figura 7. Pasos de la función del algoritmo PCA. Información tomada de artículo (Mamani, 2019). Elaborado por Daniela Taynys.

#### 2.2.4.5.2. *No Lineales*

##### 2.2.4.5.2.1. *Escala multidimensional (MDS)*

Es un método muy utilizado en el análisis de similitud entre datos como distancias (Mamani, 2019).

##### 2.2.4.5.2.2. *Mapeo de características isométricas (ISOMAP)*

Permite proyectar las informaciones a un tamaño menor al tiempo preservado la distancia, esta distancia permite ver el punto más corto entre los dos puntos (López y Ortiz, 2019).

#### 2.2.4.5.2.3. *Incrustación localmente lineal (LLE)*

Realiza el restablecimiento global no lineal de los acomodamientos, por lo que cada cubrimiento podrá escribirse como la suma tanto lineal como ponderada, pertenecientes a sus vecinos con información competente (BIGDATA, 2020).

#### 2.2.4.5.2.4. *Mapas Heissien (HLLE)*

Se proyecta los datos a determinada dimensión baja utilizando el operador de Hessian y mantiene el vecindario local como LLE para resultados óptimos (Mamani, 2019).

#### 2.2.4.5.2.5. *Incrustación espectral*

Esta técnica preserva la localidad utilizando métodos espectrales para disminuir la dimensionalidad por medio de estructuras en entradas cercanas pasando a salidas cercanas (BIGDATA, 2020).

#### 2.2.4.5.2.6. *Incrustación de vecinos estocásticos distribuidos en $t$ (t-SNE)*

Describe BIGDATA (2020), que podría usarse en datos de alta dimensionalidad y luego de ello esas dimensiones se transforman en insumos para utilización de otro modelo.

### 2.2.5. **Lenguaje Natural**

Se comprende como la lengua que utilizan los seres humanos para interaccionar por medio de alguna forma de comunicación pudiendo ser escrita, verbal o no verbal. El fundamento para entender el lenguaje natural es a través del idioma que se emplea usualmente mediante el habla (castellano, inglés, chino, entre otros) (Gómez, 2021).

#### 2.2.6. **Natural Language Processing (NLP)**

Es un campo de la investigación académica y tecnológica que comprende varias técnicas en la computación científica, catalogado como una subdisciplina de la Inteligencia artificial, acogido a la rama de la ingeniería lingüística computacional y estudia la comunicación entre un ordenador y el lenguaje humano, mediante métodos sintáctico, semántico, pragmático y morfológico, formando una estructura gramatical de manera concreta, permitiendo el proceso que genera el lenguaje natural por vía de reconocimiento de imágenes, texto y verbal. Por lo que, el computador debe de llevar un entrenamiento eficaz para que sea capaz de admitir información, entenderla y transmitirla al lenguaje humano (Chowdhary, 2020).

##### 2.2.6.1. *Librerías utilizadas en NLP*

###### 2.2.6.1.1. *NLTK*

Es un paquete conformado por bibliotecas y programas dirigido para el procesamiento simbólico y estadístico de NLP. Desarrollado por Steven Bird y Edward Loper de la

Universidad de Pensilvania. Esta librería es utilizada con la finalidad de apoyar la investigación educativa o áreas como: lingüística empírica, ciencia cognitiva, IA, la recuperación de información y el aprendizaje automático. Por otro lado, existen alrededor de 25 países que utilizan esta herramienta para: clasificación, la tokenización, el stemming, el tagging análisis sintáctico, razonamiento semántico, etc. (Yao, 2019).

#### 2.2.6.1.2. *spaCY*

Según Delgado (2021), es una librería avanzada para trabajar con NLP en Python muy utilizada para: lematizar, etiquetado, segmentación, clasificación, etc.

#### 2.2.6.1.3. *JSON*

Es un acrónimo de JavaScript Object Notation con formato de texto ágil para realizar múltiples cambios, utilizado actualmente por muchos proyectos en distintos lenguajes de programación como: Python, C, Java, etc. Ideal para crear y almacenar estructuras de datos con formato entendible por el humano, gracias a su sintaxis. Esta librería permite: parsear archivos o strings, convierte diccionarios a listas, etc. (Rahman, 2022).

#### 2.2.6.1.4. *ReGex o re*

Es denominado como una herramienta útil para trabajar con cadenas de textos en expresiones regulares. Estas utilizan patrones de coincidencia como: extraer, comparar, reemplazar o dividir segmentos de textos a otras cadenas (Fosado, 2019).

#### 2.2.6.1.5. *Scikit Learn (SKLEARN)*

Es una librería gratuita de aprendizaje automático en Python, que contiene múltiples funcionalidades para el tratamiento de datos como: algoritmos de clasificación, clustering, etc. Contiene variedades de módulos que son: datasets, preprocessing, metrics, model\_selection, feature\_selection, clustering, multiclass, naive\_bayes, neighbors, neural\_network, svm, tree, ensemble y utils (Messina, 2018).

### 2.2.6.2. **Pre-Procesamiento de datos en NLP**

#### 2.2.6.2.1. *Técnicas para el Pre-Procesamiento de datos*



Figura 8. Técnicas de preprocesamiento de datos en NLP. Información tomada de (Battocchia, 2020). Elaborado por Daniela Taynys.

El preprocesamiento son etapas que recorre un conjunto de datos para eliminar el ruido dentro del texto, es decir, que elimina muchas variantes como: espacios en blanco, conversión en caracteres acentuados en caracteres ASCII, elimina caracteres especiales, palabras vacías, datos numéricos, conversión del texto de mayúscula a minúscula, etc. Estos procedimientos se lo realizan con la finalidad de que los modelos de ML se desempeñen de mejor manera (Romero, 2021).

#### 2.2.6.2.2. Tipos de variables en el Pre-Procesamiento de datos

Existen variables categóricas y numéricas, en este tipo de investigación se centrará en el análisis de datos tipo categóricas para la depuración de información y posteriormente pasar por el procesamiento en donde se obtendrá los resúmenes correspondientes.

Destaca IMB (2021), que las variables categóricas son datos con número limitado de valores, por ejemplo: genero, religión, pudiendo ser a su vez alfanuméricas utilizando códigos para representación de lo que desea que capte el ordenador, ejemplo: hombre=0 y mujer=1, haciendo referencia a datos cualitativos.

Estas variables se clasifican en dos componentes tales como: Nominal denominados como valores sin clasificación intrínseca ejemplo: hombre, verde, etc. y Ordinal que consta de valores con clasificación intrínseca ejemplo: pequeño, grande, entre otros.



Figura 9. Tipos de variables existentes para el pre-procesamiento de datos en NLP. Elaborado por Daniela Taynys.

#### 2.2.6.3. Técnica de procesamiento de datos en NLP

De acuerdo con Aprende Machine Learning (2018), describe las siguientes técnicas comunes para NLP y son:

- Tokenización: división de párrafo en palabras usando `word_tokenize(palabras)`, `sent_tokenize(oraciones)`, etc.
- Stop Words: eliminación de palabras de parada (vacías).
- Normalización: incluye muchos cambios como: de mayúsculas a minúscula, eliminación de caracteres, signos de puntuación, retira palabras, etc.

- Lematización: Conversión de palabras a forma estándar, incluye lematizador en inglés y stemmers para español.
- POS tag: Integra etiquetado que parte del discurso PoS, etiquetado en partes de la oración, sustantivo, verbo, adjetivo, entre otros.
- Análisis superficial / Trozos: utilizado para entendimiento de la gramática en oraciones, realizando un recorrido de cada uno de los tokens y Pos tag, en donde se elabora un árbol de su determinada estructura.
- Significado de las palabras: integra semántica léxica y desambiguación del sentido de las palabras.
- Pragmatic Analysis: detección en cuanto a determinar cómo se dicen las cosas por medio de la comunicación produciendo: ironía, sarcasmo, intencionalidad, etc.
- Bolsa de palabras: representación del léxico que se utilizará en el algoritmo consistiendo en la creación de una matriz que tendrá la función de contabilizar las veces que se manifieste en cuanto al token de cada oración haciendo una representación por fila.
- Word2vec: recurso muy utilizado para aprender a leer grandes cantidades de informaciones y memoriza las palabras de similitud en distintos contextos.
- Stemming o derivación: Proceso de reducción de inflexiones de una palabra a la raíz, pudiendo ser realizada por la librería NLTK en donde encuentra distintos escenarios al momento de procesar las tareas.

#### 2.2.6.3.1. Algoritmos de Stemming o derivación

Indica Molina (2018), que los principales algoritmos son:

- Truncado: toma los primeros n caracteres y el resto procede a eliminarlos.
- Lovins: elimina el sufijo más extenso de la palabra.
- Posters: contiene cinco procedimientos donde se aplican reglas y a cada una de ellas realiza una verificación si la regla es admitida el sufijo será eliminado accediendo a los siguientes pasos.
- Paice/Husk: incluye 120 reglas volviéndose un núcleo iterativo y por cada última letra del sufijo hace la comprobación para al finalizar eliminar o cambiar si la regla es inválida dando paso a que el algoritmo quede terminado.
- Dawson: integra una lista aproximada de 1200 sufijos viajando a través de corpus una sola vez, los mismos permiten un almacenamiento de forma invertida a su longitud y última letra.



- N-gramas: es la descomposición de la información como el texto, documentos, publicaciones, etc. Su funcionamiento implica que N elementos se tomarán en consideración, haciendo uso de varias funcionalidades tales como el reconocimiento de voz, detección de errores ortográficos, etc. (Jayasuriya et al, 2021).

#### 2.2.6.4. Clasificación en NLP

De acuerdo con McKenzie & Adams (2021), el NLP es una rama de ML y la IA que supera la amplia brecha que existía en la comunicación entre el ser humano y la comprensión computacional. Por otro lado, destaca los subcampos comunes como: respuesta a preguntas (motores de búsqueda), reconocimiento de voz (Siri), traducción automática (Google Translate), extracción de información (información importante, noticias relevantes, etc.), Análisis de sentimiento, detectar la actitud (positiva, negativa, neutral) de un texto.



Figura 10. Mapa mental de la Clasificación de NLP. Información tomada por la Investigación Directa. Elaborado por Daniela Taynys

#### 2.2.6.5. Tipos de generación de resúmenes de textos en NLP

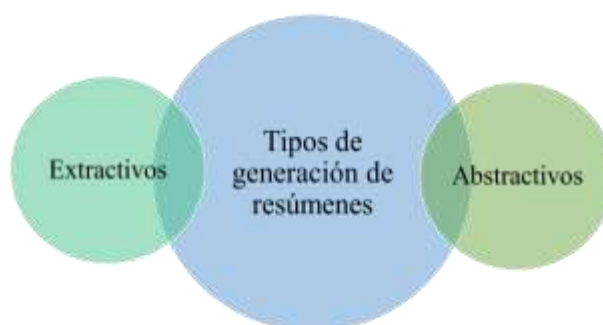


Figura 11. Tipos de generación de resúmenes. Información tomada por la investigación directa. Elaborado por Daniela Taynys.

Describe Liu et al. (2018), que para resúmenes de una gran cantidad de datos se surge la necesidad de reducción en la dimensionalidad textual para lograr manejar información con pocos datos, logrando la optimización de búsquedas en la web. Es importante destacar

que la idea primordial es obtener la información con poca dimensionalidad sin acceder a grandes cantidades de contenidos, no obstante, si el usuario desea ingresar a un contenido que le interese este podrá tener acceso a la misma en cuanto lo desee. Existen dos tipos de generación de resúmenes son: Resúmenes extractivos y abstractivo.

#### *2.2.6.5.1. Resúmenes extractivos*

Es creado a partir de la selección de varias unidades textuales de la información original donde se copia el texto con datos dando como prioridad la generación del resumen. Este proceso de selección puede ser por palabras, oraciones, o párrafos implicando una valoración en el peso que posee cada una de ellas asignando un puntaje y realizando una ponderación de unidades frecuentes que aparecen en el texto. Es importante estos tipos de resúmenes porque resalta la opinión del autor en cuanto a la información que se encuentre en dicho medio informativo (Grusky, Naaman & Artzi, 2018).

#### *2.2.6.5.2. Resúmenes abstractivos*

Determina que oraciones son significativas dentro de un contexto para proceder a ser parte del proceso de resumen. Para obtener un resumen óptimo en calidad de información y lingüística se debe aplicar este método. Su funcionalidad consiste en generar nuevos resúmenes a partir de lo que se considera importante y lo reescribe en una nueva versión (Li, Xu, Li, & Gao, 2018).

#### *2.2.6.5.3. Resúmenes extractivos vs. Resumen abstractivos*

Describe Liu et al. (2018), que cuando se resume de manera extractiva es como seleccionar varias partes del contexto considerando sus puntuaciones más altas como importantes para generar la sinopsis, en cambio el resumen abstractivo es equivalente a la decisión entre oraciones de contenido o reescritura de lo que se pretendió decir en la lectura. Se considera que el resumen extractivo da mejores resultados ya que el abstractivo tiene muchos inconvenientes al momento de la representación del lenguaje natural.

Se concluye que para este proyecto se basará en resúmenes automáticos tipo extractivos para no alterar tanto el contexto de la estructura gramatical de las fuentes de noticias originales y luego mostrar las sinopsis en el prototipo web.

#### *2.2.6.5.4. Modelos para generación de resúmenes de textos extractivos*

##### *2.2.6.5.4.1. Basado en matriz de similitud*

Esta matriz es empleada por el algoritmo de PageRank y proporciona la selección de frases para representación de ideas principales, su proceso consiste en eliminar conectores

inicialmente más conocidas como palabras de paradas para posteriormente determinar la importancia o peso de cada una de las frases. Por lo que, este método permite la selección de N frases con el peso más alto para la construcción de resumen (Elbarougy et al., 2020).



Figura 12. Modelo arquitectónico de modelo basado en matriz de similitud. Información tomada de artículo por (Elbarougy et al, 2021). Elaborado por Daniela Taynys

#### 2.2.6.5.4.2. Basado en frecuencia de las palabras por el índice de TF

López & Torres (2021), indican que son algoritmos por lo que se calcula el puntaje de cada frase de acuerdo con la frecuencia de palabras (TF) que la componen, por lo que estos puntajes son calculados por medio de frecuencia de término y frecuencia inversa (TF-IDF).

Para el cálculo de puntuación se obtiene mediante la formula:

$$TF = \frac{nx}{n}$$

nx: número de veces de frecuencia de palabra que aparece en el documento

n: número total de palabras que existen en el documento

#### 2.2.6.5.4.3. Basado en frecuencia de las palabras por el índice de TFIDF

Es similar al índice basado en frecuencia de TF pero le da un paso extra al cálculo mediante lo siguiente:

$$TF = \frac{nx}{n}$$

$$TF = \log_e \frac{m}{mx}$$

$$TFIDF = TF * IDF$$

m: número total de documentos considerados

n: número de documentos que incluyen el término x



Figura 13. Modelo arquitectónico de modelo basado en frecuencia de palabras. Información tomada de artículo por (Lopez & Torres, 2021). Elaborado por Daniela Taynys

#### 2.2.6.5.4.4. Basado en métrica de LUHN

Este método toma como factor significativo el valor de una sentencia a la frecuencia y establece la posición relativa de las palabras con más importancias que se encuentran dentro de una oración. Se destaca que la forma de medir este método es muy útil ya que lo realiza a través de frecuencia de aparición por palabras. Para ello se deben obtener las palabras de mayor importancia adquiriéndolas por medio de ponderaciones de palabras dentro del contexto, por lo que propone puntuar cada sentencia basándose en la ubicación y la cantidad de las palabras que contengan la oración (Corvi, 2019).

#### 2.2.6.6. Arquitectura de Procesamiento de Lenguaje Natural

Como indica Zhao et al. (2020) el NLP es una agrupación de técnicas, herramientas y recursos informáticos para análisis y representación de texto natural, llevado a cabo en uno o varios niveles lingüísticos pudiendo ser: fonológico, morfológico, léxico, sintáctico, semántico, discursivo y pragmático, con la finalidad de procesar información similar al ser humano para ejecuciones de varias tareas.

Los diversos niveles que conforman el NLP se entienden como modelos de gramática, con la finalidad de poder desarrollar un análisis de las reglas, a su vez los principios que conducen al uso de la lengua en el funcionamiento, ayudando a que su descripción sea de manera formal en lenguas naturales. (Moreira et al., 2020).

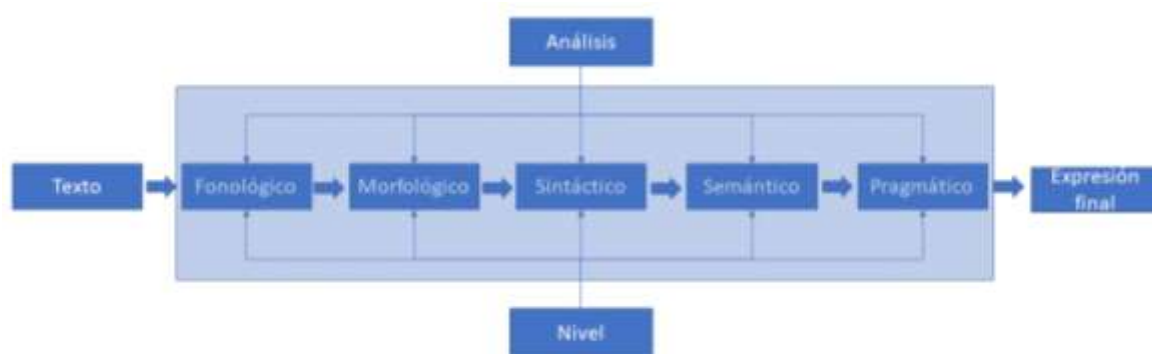


Figura 14. Sistema Arquitectónico NPL. Información tomada de artículo *Cómputo Aplicado de* (Camacho y Navarro, 2020). Elaborado por Daniela Taynys.

##### 2.2.6.6.1. Pragmático

Se establece como el uso social e interactivo del lenguaje como medio de comunicación, seleccionando conductas en el contexto cognitivo o social tanto en el receptor como el uso de ajustar el lenguaje al contexto ya que por medio de una misma oración puede tener varios sentidos en diferentes contextos, siendo referente a la práctica, la ejecución de hechos y no en base a la teoría o especulación (Brasdefer, 2018).

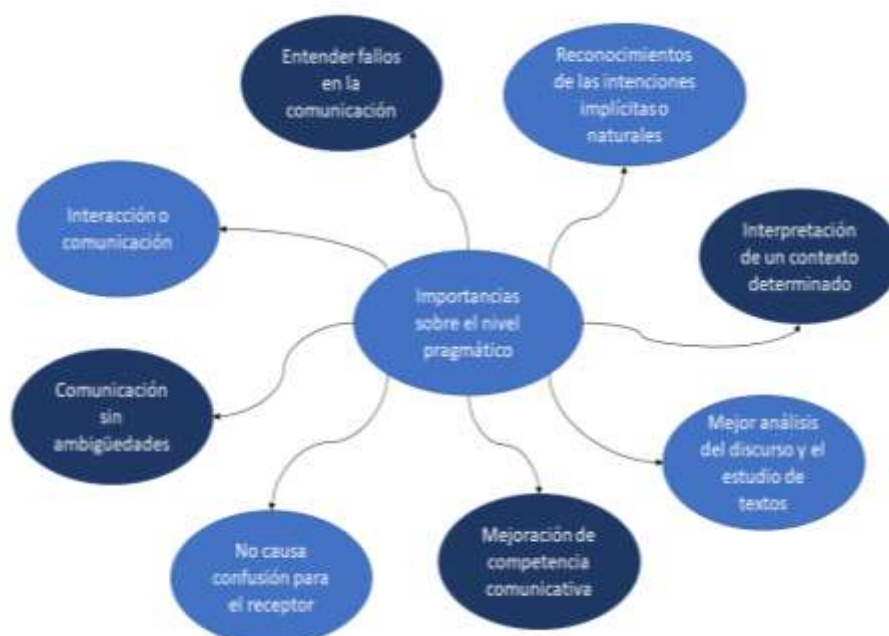
### 2.2.6.6.2. Elementos de la pragmática

**Tabla 3** Elementos de la pragmática utilizados en el estudio

Denominación	Interpretación
Intencionalidad	Propósito de la comunicación.
Contacto visual	Herramientas de la comunicación no verbal.
Representación de intenciones	Proyecta comportamientos para comunicar de manera no verbal, basándose en respuestas emocionales pudiendo ser verdaderas o falsas.
Aclaraciones y reparaciones	Darle la información precisa al receptor.
Inferencias lingüísticas	Interpreta enunciados indirectos, ambiguos, mentiras, ironía, etc.

*Información tomada Cornago, (2018). Elaborado por Daniela Taynys*

### 2.2.6.6.3. Importancias que incluye el componente pragmático



*Figura 15. Variantes importantes del componente pragmático en el lenguaje natural. Información tomada del Libro (Ridao, 2020). Elaborado por Daniela Taynys*

### **2.2.7. Web Scraping**

Es una técnica muy utilizada para realizar raspado web donde permite recolectar datos e información de manera automatizada, por lo que se transforma a minería de datos en la web. Esta aplicación permite recuperar distintos tipos de información tales como: noticias, marketing competitivo, comentarios, artículos deseados, etc. (Thomas & Mathur, 2019).

#### **2.2.7.1. Éticas del Web Scraping**

Indica Krotov et al. (2020), que antes de inicializar un proyecto de minería de textos como el raspado web, los profesionales tienen que registrarse sobre el grado de legalidad y ética con respecto al uso de los datos de la web. Se destaca que han existido distintos marcos jurídicos aplicando casos judiciales como: la accesibilidad y uso ilegal de los datos, derechos de autor, invasión de privacidad, secretos judiciales, etc. Al aplicar esta técnica hay que tener en cuenta de las posibles consecuencias éticas relacionadas con la información existente en la web. Se describe que para hacer uso de esta técnica se debe realizar las siguientes preguntas:

- ¿Puedo tomar estos datos?
- ¿Estoy violando alguna reglamentación local?
- ¿Estoy violando los “Términos y Condiciones” del sitio?
- ¿Estoy accediendo a lugares no autorizados?
- ¿Puedo tomar estos datos?
- ¿Es legal el uso que le voy a dar a estos datos?
- ¿Para qué puedo usar estos datos?

### **2.2.8. Prototipo**

Es un primer modelo que sirve para concretar un sistema o parte del mismo, estableciendo una representación o simulación del producto final para comprender su previa funcionalidad, a su vez, permite verificar el diseño sobre aspectos diversos del software (Barbosa de Lira et al., 2020).

### **2.2.9. La web**

La World Wide Web son diversos documentos conectados a través de protocolos de transferencias hipertextos (HTTP) y su funcionamiento es por medio de internet (EDTeam, 2020).

### 2.2.10. Internet

Se entiende como el conjunto de redes en comunicaciones que están interconectadas entre sí, por lo que hacen la utilización del vínculo de protocolos TCP/IP estableciendo una conexión y el intercambio de datos entre ordenadores, a su vez, indican normas de cómo se debe viajar la información (REVISTA ESPAÑOLA DE ELECTRÓNICA, 2022).

### 2.2.11. Desarrollo web

Según Alarcon (2022), es una herramienta tecnológica que accede a crear páginas web, sitios web o aplicaciones web de manera que se pueda compartir información de acuerdo con las demandas de diversos requerimientos de la sociedad. Sus diferencias son:

**Página web:** Es una documentación en cuanto a la información electrónica normalmente encontrada en formato HTML por lo que proporciona acceso a otras páginas web mediante enlaces de hipertexto, el mismo puede contener texto imágenes, enlaces, etc.

**Sitio web:** Es una agrupación de páginas web relacionada a un dominio o subdominio en la World Wide Web dentro de Internet. Se destaca por referenciarse a un tema en particular o propósito en específico, son escritos en código HTML y se aloja a un servidor web o HTTP.

**Aplicaciones web:** Es una herramienta en el que los usuarios acceden a un servidor por medio de internet a través de un navegador. Su objetivo principal es que el usuario realice una determinada tarea o acciones.

### 2.2.12. Lenguajes de Programación

Según estudios de Diakopoulos et al. (2014), a través del espectro de la IEEE permite la clasificación de decenas de lenguajes de programación en donde puede ser filtrado, combinado y ponderado por métricas de medio de 11 fuentes tales como: Stack Overflow, Hacker News, etc. Permitiendo personalizar la clasificación de forma dependiente de se desea buscar.

Fuente	Ponderación (0-100)
Búsqueda de Google	50
Tendencias de Google	50
Github (activo)	50
Github (creado)	30
Desbordamiento de pila (Preguntas)	30
Reddit	30
Noticias de piratas informáticos	20
Constructor de la carrera	5
Gorjeo	50
Empleos de IEEE	50

Botones: Guardar mi clasificación, Cancelar

ponderado por métricas de como: GitHub, IEEE, Google, Permitiendo personalizar la clasificación de forma dependiente de se desea buscar.

Figura 16. Personalización para generar el espectro IEEE. Información tomada de <https://spectrum.ieee.org/top-programming-languages/>. Elaborado por IEEE Spectrum.

Rank	Language	Type	Score
1	Python	Scripting	100.0
2	Java	General	95.4
3	C++	General	94.7
4	C#	General	92.4
5	JavaScript	Scripting	88.1
6	R	Scripting	82.4
7	Go	Scripting	81.7
8	PHP	Scripting	77.7
9	HTML	Scripting	75.4

Figura 17. Espectro IEEE evaluando los mejores lenguajes de programación. Información tomada de <https://spectrum.ieee.org/top-programming-languages/>. Elaborado por IEEE Spectrum.

Por otro lado, en un estudio científico por Rytikova & Medvedev (2021), establece que los lenguajes programación más populares son: JavaScript, Java, C#, Python, destacando como los porcentajes altos.

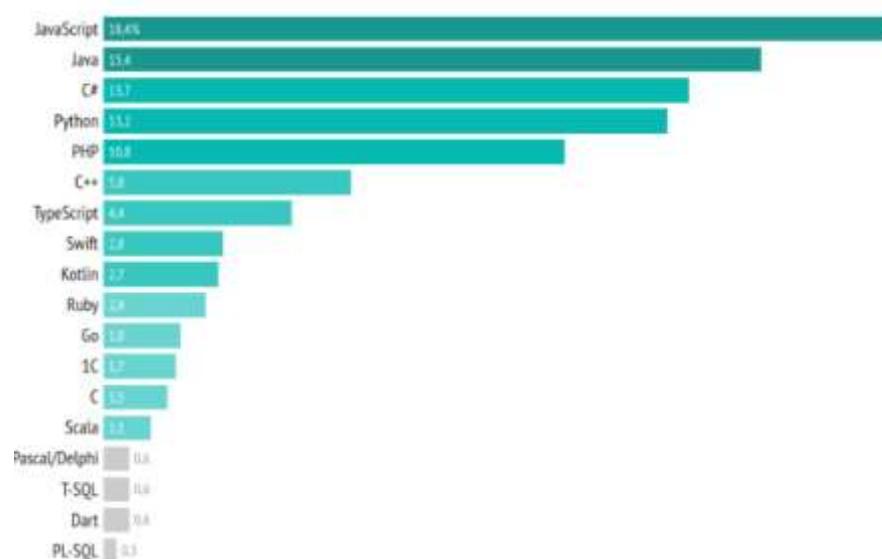


Figura 18. Análisis estadístico de los mejores lenguajes de programación. Información tomada de <https://aip.scitation.org/doi/pdf/10.1063/5.0041762>. Elaborado por (Rytikova & Medvedev, 2021).



TIOBE (2022) menciona que, mediante su indicador respecto a la aceptación de los desarrolladores en los diversos lenguajes de programación, permite compartir su estadística porcentual a la comunidad, este tiende a actualizarse una vez por cada mes. En cuanto a sus puntuaciones están basadas a los desarrolladores de todo el mundo. Algunos buscadores de internet muy conocidos como: YouTube, Wikipedia, Google, Amazon, entre otros, se utilizan para calcular las ponderaciones. Es fundamental señalar que TIOBE considera que la medición que se da no trata de conocer el potente lenguaje de programación o el que se han escrito la mayoría de las líneas de código.

Dado a la estadística del índice de TIOBE se puede observar que principalmente se encuentra Python ocupando el primer puesto, subiendo de rango en este 2022 a diferencia del año 2021 que ocupaba el segundo puesto con respecto al lenguaje de programación C, sin embargo, se puede destacar que R está en el onceavo lugar perdiendo su popularidad.



Figura 19. Datos de calificaciones del índice de la comunidad de programación TIOBE. Información tomada de <https://www.tiobe.com/tiobe-index/>. Elaborado por TIOBE.

Dado a los análisis descritos anteriormente se pudo visualizar distintos tipos de estadísticas estableciendo una mínima variabilidad en la escala de los lenguajes de programación más populares. Descrito dichas estadísticas analizadas por el transcurso de los años se procede a escoger los cuatro mejores lenguajes de programación con más tendencia en utilización y así evaluar sus indicadores determinando el más recomendado para el desarrollo del proyecto.

Es importante destacar sus características primordiales para concluir la diferencia que existe entre cada uno de ellos y así escoger el más apropiado. Entre los cuatro principales escogidos para la evaluación es: Python, C, Java, C++.

### 2.2.12.1. Python

Según Python (2022), describe que es un lenguaje de programación orientado a objetos, explicativo e interactivo. Permite añadir diversos modelos, excepciones, clases, tipo dinámicos normal y de muy alto nivel. Recopila variedades de paradigmas buscando un camino más profundo del modelo en cuanto al problema de destino en sus programas siendo una programación procedimental y funcional, a su vez coordina una sintaxis muy clara. En cuanto a interfaces está disponible para muchas bibliotecas y llamadas al sistema, pudiendo ser extensible de manera adaptada para los que requieran de una interface programable, así como en C o C++. Por último, una de sus cualidades es que es portátil ya que permite su ejecución tanto para Windows como Unix que incluye Linux y macOS.

**Tabla 4** Características de Python



- Sintaxis fácil de leer, escribir, aprender y comprender.
- Código abierto y disponible para la comunidad de forma gratuita.
- Simple y legible para creaciones de prototipos.
- Expresivo e interpretado.
- Multiplataforma y portátil.
- Lenguaje de nivel alto con tipado diligente.
- Orientado a objetos compatible con la programación funcional.
- Escritura dinámica, extensible e integrable.
- Apoyo popular y de la comunidad grande.

---

*Información tomada de TechVidvan por (Cohen, 2022). Elaborado por Daniela Taynys*

### 2.2.12.2. C

Es un lenguaje de programación que es examinado como uno de los más calificados en la actualidad, con el cual se pueden desarrollar tanto sistemas operativos como aplicaciones, a la vez formando alineamientos en cimiento con respecto a otros lenguajes existentes en la actualidad tales como: C++, Java, C#, etc. Se caracteriza por su uso de manera global, teniendo una estructura de sintaxis compactada y de extensa portabilidad, la misma implementa componentes esenciales para cualquier ordenador tales como: utilización de asignación de memoria, direcciones, caracteres, números, etc. (Lucas, 2019).

**Tabla 5** Características de C

- Programación estructurada de nivel medio.
- Tiene una programación a nivel mediano.
- No es dependiente del hardware y tiene la capacidad de migrar a otros sistemas.
- No es un lenguaje para una labor específicamente, es decir, que puede permitir desarrollar diversas funcionalidades como: juegos, sistema operativo, etc.
- Lleva una inspección total del computador.
- Organiza el trabajo en su totalidad.
- Los programas rinden de manera ágil y potente.
- Excelente en manejo de tipo de operadores, datos, y variables.

*Información tomada de OpenWebinars por (Lucas, 2019). Elaborado por Daniela Taynys.*

### **2.2.12.3. Java**

Es un lenguaje de programación de elevado nivel orientado a objeto y fue expuesta por Sun Microsystem actualmente es propiedad de Oracle y su creador fue James Gosling. Es un lenguaje de propósito general, obteniendo una implementación del compilador original y máquina virtual (Yucra, 2021).

**Tabla 6** Características de Java

- Lenguaje simple.
- Orientado a objeto.
- Distribuido para una serie de clases para uso en diversos ordenadores conectados entre sí por medio de una red.
- Interpretado y compilado a la vez.
- Robusto y altamente fiable.
- Indiferente a la arquitectura.
- Tiene portabilidad.
- Cuenta con Máquina Virtual Java (JVM), corresponde a una elevada productividad y multi-hebra.
- Resiste la sincronización de numerosos hilos ejecutados por medio del lenguaje.
- Es dinámico en la fase de enlazado.

*Información tomada de Slideplayer Lenguaje de programación estructurado por (Yucra, 2021). Elaborado por Daniela Taynys*

Según Martínez (2021), es conocido como el más popular del universo, potente, versátil, estable, estándar, multiplataforma y orientada a objeto. Permite el arranque en distintos sistemas operativos existentes en la actualidad, su código es independiente y abierto. Es mayormente utilizado en la interface para la programación, a su vez accede de manera muy práctica ejecutándolo a cualquier dispositivo indispensablemente de la marca o modelo, de manera segura y completa. Este lenguaje permitirá escribir el programa en una plataforma y ejecutarla virtualmente en otra.



- Lenguaje de uso popular
- Procedimental
- Programación estructurada
- Gestión de memoria
- Puntero apuntando a la memoria de una variable
- Cuenta con tipado fuerte
- Compatible para compilar en C
- Sintaxis del lenguaje C
- Lenguaje compilado

#### 2.2.12.4. C ++

Es un lenguaje de programación de versión mejorada y extendida del lenguaje de programación C, fue desarrollado por Bjarne Stroustrup en 1979. Es refinado, eficaz y es compatible con la programación orientada a objetos con un formato libre y compilado (Corvo, 2020).

C++ no reemplaza al lenguaje de programación C, pero se lo considera como superconjunto tomando en cuenta de que en la actualidad cuenta con nuevas características, sin embargo, no procura forzarse a cambios drásticos ni mucho menos terminar con lo que ya se comprendía. Establece nuevas definiciones tales como: objetos y clases. Hay 3 conjuntos relacionados los mismos se establecen como: programación orientada a objetos, características ampliadas y características clásicas de C (Rico, 2019).

#### **Tabla 7** Características de C++

*Información tomada de Lifeder por (Corvo, 2020). Elaborado por Daniela Taynys*

Analizando los cuatro lenguajes de programación más populares referente el análisis comparativo de distintas estadísticas, se concluye como el más adecuado Python para realizar el prototipo web, debido a que es uno de los más rápido e interpretado que contiene tipado dinámico orientado a objetos contando con un gran estándar de biblioteca permitiendo

que sea extendible e incrustable. Python es un lenguaje de fácil aprendizaje con alta estabilidad en desarrollos, contando con código abierto y gratuito.

### **2.2.13. Frameworks**

Establece Ritesh (2021), que son herramientas con varios componentes para su utilización, este puede incluir programas de soporte, librerías, bibliotecas de códigos, etc. Se utiliza en los desarrollos de proyectos, dando un funcionamiento de manera rápida y eficiente. Siempre se actualizan pudiendo contar con las siguientes características:

- Funcionalidad: operatividad necesaria para el proyecto.
- Consistencia: apoyo para equipamientos grandes.
- Documentación: es necesario que brinde implementación y capacitación.
- Comunidad activa: marco bien establecido con una base de usuarios activos.

**Tabla 8** Comparativa de Frameworks utilizados en Python.

<b>Framework</b>	<b>Características</b>	<b>Ventajas</b>	<b>Desventajas</b>
<b>Django</b>	Variedades de bibliotecas listas para usar con soportes en autenticación, migraciones de bases de datos, mapeador relacional de objetos (ORM), soporte para servidores web, motor de plantillas, enrutamiento de URL, ofrece formularios model-based.	El ORM es sencillo, sistemas de permisos, panel de administración incluido, múltiples paquetes, autenticación aprobada.	Es un monolito, lento, su ORM no es asincrónico, curva de aprendizaje moderada, no construido para proyectos pequeños, problema de solicitudes múltiples.
<b>Flask</b>	Depurador rápido incorporado, manejo de solicitudes HTTP, servidor web de desarrollo incorporado, plantillas Jinja2, envío de solicitudes RESTful, soporte para conectar cualquier ORM, admite cookies seguras para establecer sesiones del lado del cliente, basado en Unicode, soporte de pruebas unitarias, cumplimiento con WSGI (Web Server Gateway Interface-comunicación entre servidor web y aplicación web) 1.0, diseñado para aplicación web en poco tiempo.	Flexible, fácil de entender, buenas herramientas de prueba, micro framework, depurador y soporte integrado, manejo de rutas.	Al ser un micro framework el desarrollador puede romper las buenas prácticas y los estándares, su representación de plugins no es extensa, complicaciones con pruebas unitarias o migraciones.

<b>CherryPy</b>	Cuenta con series de herramientas listas para usar autenticación, almacenamiento en caché, codificación, etc., sistema de complemento integrado flexible, servidor web de subprocesos WSGI compatible con HTTP/1.1, soporte para cobertura, creación de perfiles y pruebas, simplicidad para múltiples servidores HTTP.	Marco de aplicación web orientado a objeto, desarrollo rápido, inicialización en cualquier entorno compatible con WSGI.	Nivel bajo, no realiza creación de plantillas para la representación de salida o el acceso de back-end, el marco es extensible con filtros a una solicitud/respuesta.
<b>Bootle</b>	Soporte de adaptador para motores de plantillas de terceros y servidores WSGI/HTTP, acceso simple desde cookies, datos, cargas, etc., compatibilidad con complementos para diferentes bases de datos, proporciona rutas de envío de solicitudes que admiten parámetros de URL.	Proceso de distribución en un solo archivo, fácil de compartir o subir aplicación, flexible, enrutamiento, biblioteca básica en Python es de fácil aprendizaje.	Proporciona herramientas mínimas al desarrollador, se distribuye como un módulo de archivo único, no tiene más dependencias que la biblioteca estándar de Python.

---

*Información tomada de artículos (Affde, 2021) (Ankush, 2022) (Epitech, 2021) (Presta, 2022) (Rodríguez, 2019) y (Zepeda, 2021). Elaborado por Daniela Taynys*

Se concluye como el más adecuado a implementar Django ya que permite tener las buenas prácticas de programación, una de sus ventajas a utilizar para el desarrollo del prototipo web es el panel de administración donde permitirá almacenar todo el contenido que se requiera, a través de

migraciones de bases de datos. Al ser un prototipo web es necesario inicializar con bases fuertes para futuramente implementar varios módulos de manera incremental. Se considera que no sea futuramente un prototipo por el contrario sea un producto final a producción.



### 2.2.14. Entorno de Desarrollo Integrado (IDE)


Un IDE es un ambiente digital implementado en la programación para desarrollar aplicaciones, juegos, prototipos web, es decir, cualquier tipo de software con el objetivo de acelerar todos los procedimientos de diseño de un software. Se torna indispensable tanto para el ámbito de construcción de aplicaciones web (DAW) como en la construcción de aplicaciones multiplataforma (DAM). Permite que las tareas de los programadores sean más fáciles y manejables por medio de herramientas como: compiladores, depuradores o bibliotecas conformando un aumento en la productividad (UNIR, 2021).

#### 2.2.14.1. Entorno de Desarrollo Integrado de Escritorio

##### 2.2.14.1.1. IDE PyCharm

Es muy considerado para el lenguaje Python en torno a los desarrolladores, siendo uno de los más complementados y fue construido inicialmente por JetBrains brindando todas las herramientas funcionales de manera precisa para Python en un solo sitio, a su vez tiene una amplia serie de herramientas para sus distintos desarrollos (Prasad, 2021).

**Tabla 9** Características de PyCharm

	<ul style="list-style-type: none"> <li>• Disponible en versión profesional(paga) y Community (Open source) para multiplataforma.</li> <li>• Resaltado de sintaxis y múltiples sugerencias autocompletado.</li> <li>• Interactivo con sangría automática.</li> <li>• IDE's más completos para Python.</li> <li>• Crear y gestionar proyectos.</li> <li>• Tarda en cargar al iniciar y requiere al menos 4GB de RAM.</li> <li>• Adecuado para desarrollo de aplicaciones web, IA, ML.</li> <li>• Integración con frameworks y herramientas de análisis para Python.</li> </ul>
---	--


*Información obtenida de YouTube por (Universidad de Antioquia, 2020). Elaborado por Daniela Tainys*

##### 2.2.14.1.2. Visual Studio Code (VS Code)

Es el mejor de las creaciones de Microsoft codificando de forma más inteligente con un mejor rendimiento, permitiendo redactar variables de manera concreta y veloz, por medio de recomendaciones en pequeños fragmentos de código. Mejora la estructura del código mediante bombillas y el mismo sugiere determinadas acciones tales como: añadir parámetros, renombramiento de funciones, entre otros. Está desarrollado bajo el paradigma

Open Source, en el que integra GitHub añadiendo la inspección de versiones, adicionalmente de plataforma como: Windows, Mac OS y Linux, compatible con muchos lenguajes como: Python, Java, C++, JavaScript, entre otros. Manteniendo la complejidad y superando complejidades en el transcurso de su navegación hacia un archivo, miembro, tipo o símbolo (Prasad, 2021).

**Tabla 10** Característica de Visual Studio Code


	<ul style="list-style-type: none"> <li>• Open Source y multiplataforma.</li> <li>• Resaltado de sintaxis y múltiples sugerencias autocompletado.</li> <li>• Código libre y sangría automática.</li> <li>• Consola interactiva.</li> <li>• Depurador con puntos de interrupción y pasos.</li> <li>• Instalación extensiones para mayor personalización.</li> <li>• Crear y gestionar proyecto.</li> </ul>
---	--

*Información obtenida de YouTube por (Universidad de Antioquia, 2020). Elaborado por Daniela Taynys*

#### 2.2.14.1.3. SPyder

Su uso es de manera gratuita y con disposición de código para todo el mundo, construido para establecerse en un marco científico siendo compatible para python. Proporciona un elevado nivel de funcionalidades en su edición, depuración y explorador de datos. Permite manejar la ciencia de datos, por ser ágil y ejecutar las sucesiones de comandos dificultosos en python. Por otro lado, está diseñado para desarrolladores, ingenieros y científicos de datos, permitiendo la utilización de clases, funciones, análisis de código, definición de acceso teniendo una división vertical y horizontal (Prasad, 2021).

**Tabla 11** Características de Spyder

	<ul style="list-style-type: none"> <li>• Open Source y multiplataforma.</li> <li>• Consola interactiva.</li> <li>• Instalación de extensiones.</li> <li>• Incorpora librerías como: Numpy, Matplotlib, SciPy, etc.</li> <li>• Es un poderoso IDE para campos de ingeniería, científico, análisis de datos e IA.</li> </ul>
---	--

*Información obtenida de YouTube por (Universidad de Antioquia, 2020) Elaborado por Daniela Taynys.*

#### 2.2.14.1.4. Jupyter Notebook

Es una interfaz web de código abierto, que incorpora textos, sonidos, ilustraciones, a su vez permite ejecutar el desarrollo por medio del navegador o en entornos de escritorios compatible con varios lenguajes (Prasad, 2021).

**Tabla 12** Características de Jupyter



- Gratuito y de código abierto
- Open source, versión web y de escritorio.
- Incorpora librerías científicas como: Numpy, Matplotlib, SciPy, etc.
- Widgets interactivos
- IDE diseñado para campos de Big Data, ML e IA.
- Recepta hasta 40 idiomas incluyendo ML.
- Fácil aprendizaje y útil para realizar presentaciones educativas profesionales.

*Información obtenida de YouTube por (Universidad de Antioquia, 2020). Elaborado por Daniela Taynys.*

#### 2.2.14.1.5. Sublime Text

Sirve para componer códigos siendo compatible con cualquier formato de archivo y es fabricado primordialmente para escrituras de códigos sin distracciones. Ofrece un entorno de desarrollo en donde las líneas de código que se escriba resalten de manera que el desarrollador pueda centrarse exclusivamente en ellas. Tiene una versión de pago, a su vez, cuenta con una versionamiento evaluativo sin fecha limitada, es descargable desde su página oficial, contando con su versión actual denominado como Sublime text 3 y está disponible para macOS, Windows, y Linux (Sarabia, 2021).

**Tabla 13** Característica de Sublime Text



- Paleta de comandos para accesos en todo el menú de configuración.
- Cambio de archivo de forma rápida.
- Ir a Símbolos para enlistarlos y fácil de encontrarlos.
- Multi-Edición e interfaz muy sencilla.
- Atajos de teclado.
- Interfaz de usuario muy atractiva.
- Plugins para ayuda en el desarrollo de los proyectos.
- Compatibilidad con varios lenguajes de programación.

*Información tomada de (Web Design, 2022). Elaborado por Daniela Taynys*

De acuerdo a el levantamiento de información con respecto a los entornos de desarrollo de integrado IDE's de escritorio y a los recursos que se necesitan para el prototipo web, se concluye que la mejor opción para la elaboración es por medio del IDE Visual Studio Code, ya que es gratuito bajo el paradigma de Open Source y permite la agilización de programar de una manera más cómoda agradable y sencilla, contando que en la actualidad es el más reconocido y utilizado por los desarrolladores.

#### **2.2.14.2. Entorno de Desarrollo Integrado (IDE) en la nube**

##### **2.2.14.2.1. Google Colaboratory (Google colab)**

Describe Fuente (2019), es un ambiente con disponibilidad en línea de manera gratuita de Jupyter Notebook, no requiere de configuración adicional y su ejecución es completamente en la nube por lo que permite entrenar módulos de Deep Learning y ML en varias instancias como: CPU, GPU, TPU. Como tiempo máximo de ejecución continua es de 12 horas, luego de ello borra todo archivo importado y ejecutado en el entorno. Por otro lado, también hay duración de hasta 30 horas, pero se debe acceder una configuración extra.

**Tabla 14** Especificaciones de Google Colaboratory

CPU	GPU	TPU
Intel Xeon Processor with two cores @2.30Ghz and 13 GB RAM	Up to Tesla K80 with 12 GB of GDDR5 VRAM, Intel Xeon Processor with two cores @ 2.20 GHz and 13 GB RAM	Cloud TPU with 180 teraflops of computation, Intel Xeon Processor with cores @ 2.30 GHz and 13 GB RAM

*Información tomada de artículo por (Sharma, 2020). Elaborado por Daniela Taynys.*

A continuación, se destaca las principales características que lo identifican como uno mejores entornos de máquinas virtuales.

**Tabla 15** Características de Google Colab



- No requiere configuración adicional.
- Compatible para las bibliotecas de Python para analizar o visualizar los proyectos.
- Acceso al Hardware de Google de forma gratuita.
- Opciones de guardar y compartir desde la nube los notebooks enlazando desde Google Drive.
- Ahorra tiempo con atajos de teclado.
- Vincula Colab a Github

*Información tomada de sitio web por (Sandoval, 2022). Elaborado por Daniela Taynys.*

#### 2.2.14.2.2. IBM Watson Studio

Indica IBM (2020), que es uno de los principales creadores y escaladores confiables para trabajar en la nube, donde permite que cree, ejecute, gestione y optimice modelos de la IA tomando decisiones en cualquier lugar de IBM Cloud Pak® for Data

**Tabla 16** Características de IBM Watson Studio



- AutoAI para una experimentación más rápida.
- Refinería de datos avanzada.
- Compatibilidad con Notebook de código abierto.
- Herramientas visuales integradas.
- Desarrollo y entrenamiento de modelos.
- Amplios marcos de código abierto.
- Optimización de decisiones incorporada.
- Gestión de supervisión de modelos.
- Gestión de riesgos de modelos.

Información tomada de (IBM, 2020). Elaborado por Daniela Taynys.

#### 2.2.14.2.3. Azure Machine Learning

Es un servicio que ofrece Microsoft y es basado en la nube de código abierto para utilización interactiva con componentes que permiten acelerar y administrar el ciclo de vida de los proyectos de ML. Su uso está destinada a profesionales de ML, científicos de datos e ingenieros para flujos de trabajos como: entrenar e implementar modelos y organización de ML Ops por lo que ayudan a inspeccionar, restituir el entrenamiento y retornar a la integración de los modelos. Este entorno no almacena ni procesa los datos fuera de la región donde se realiza la implementación. Esta plataforma es compatible con Pytorch, TensorFlow o Scikit-learn (Microsoft, 2022)

**Tabla 17** Características de Azure Machine Learning



- Abierto e interoperable.
- Caracterización automatizada y selección de algoritmos (AutoML).
- Optimización de hiperparámetros.
- Entrenamiento distribuido de varios nodos.
- Implementación de modelos.
- Puntuación por lotes y en tiempo real (inferencia).

Información obtenida de artículo por (Microsoft, 2022). Elaborado por Daniela Taynys.

#### 2.2.14.2.4. *Amazon Machine Learning (AWS)*

Describe Amazon Machine Learning (2022), es un servicio potente con una infraestructura soportada en la nube, por lo que está disponible para cualquier Desarrollador, Ingenieros, Profesionales de ciencias de datos con aprovechamiento de recursos de ML contando con cualquier nivel de cualificación. En la actualidad este servicio ya no será actualizado, ni se aceptará usuarios.

**Tabla 18** Características de Amazon Machine Learning



- Integra modelos de ML.
- Medición de calidad de ML
- Predicción por lotes.
- Entrenamiento distribuido de varios nodos.
- Implementación de modelos.
- Puntuación por lotes y en tiempo real (inferencia).

*Información obtenida de sitio web por (Amazon Machine Learning, 2022). Elaborado por Daniela Taynys.*

De acuerdo al análisis y a los medios que se poseen para la presente investigación, se toma en consideración la conclusión de que la mejor opción para la elaboración del modelo de ML-NLP es trabajar por medio Google Colab, tomando su característica de gratuidad y permitiendo el guardado en línea por lo que se puede futuramente se continúa manipulando en cualquier momento sin la necesidad de instalar algún software extra.

### 2.2.15. Fundamentación legal

#### 2.2.15.1. *Constitución de la República del Ecuador*

En este apartado se detallan en un listado los artículos de la Constitución de la Republica del Ecuador considerados en la presente investigación. Por otro lado, estarán descritos de manera completa en el anexo 5 y son:

- Art 262: establece las distintivas políticas tanto de innovación así como de estudio científico para los saberes esenciales que se desarrollan en las tecnologías de manera incremental según sus actualizaciones.
- Art. 277: menciona que para continuar con el término del buen vivir deberán ser generados desde el estado, a través de fomentar e impulsar las diversas tecnologías existentes, la ciencia evolutiva, la asociación, la cooperación y la privación del mismo.
- Art. 350: especifica que desde el 2020 en adelante el régimen de satisfacción de necesidades que integra la humanidad tiene como objetivo la educación superior teniendo de beneficio una formación de conocimientos en cuanto a profesionalismo, científico y

humanista, por lo que promueve la expansión de saberes, innovación, entre otros. Esto contribuye las distintivas soluciones a problemáticas que se pueden generar en el país relacionados con los objetivos descritos en el presente régimen.

- Art. 385: este artículo explica que, en la sección octava de la ciencia, innovación y tecnología, promueve que para el 2020 en adelante se podrán generar, difundir y adaptar diversos conocimientos en la tecnología, entre otros relacionados con ella para mejoramiento en cuanto a la calidad de vida.

- Art 386: especifica que durante el 2020 en adelante diversas instituciones podrán realizar actividades investigativas para estudios en desarrollos de tecnologías e innovación para beneficio de la comunidad comprendiendo los programas, políticas, diversos recursos, etc.

#### ***2.2.15.2. Reglamento de Régimen Académico Consejo de Educación Superior***

Para esta sección se detallará el artículo considerado dentro del Reglamento de Régimen Académico Consejo de Educación Superior destacando el:

- Art 78: establece que los diversos programas académicos que se generen en las planificaciones académicas son de totalidad pertinencia.

#### ***2.2.15.3. Ley Orgánica de Comunicación***

En esta parte se detalla que dentro de la Ley Orgánica de Comunicación se encuentra el artículo 35 considerado en la presente investigación ya que describe los derechos de igualdad en cuanto al acceso universal de diversas tecnologías existentes proporcionando la comunicación e información, para más detalles sobre el presente artículo véase el anexo 5.

#### ***2.2.15.4. Ley Orgánica de Telecomunicaciones***

En esta sección se describe que el Art 88 es considerado en el presente estudio siendo útil su beneficio que genera para la humanidad en cuanto a la información promovido por las tecnologías que se abarca en todo el territorio nacional apoyando las tecnológicamente las ciencias científicas y facilitando su uso de manera adecuada en los estudios que se puedan generar. Para visualizar la descripción completa del artículo véase el anexo 5 del presente proyecto.

#### ***2.2.15.5. Consultas de Normas Internacionales del uso de la IA***

De acuerdo a la normativa con respecto al manejo, entrenamiento y uso de la IA en el Estado Ecuatoriano aún no se encuentra disponible debido a que no se han desarrollado las leyes correspondientes. A continuación, se verifican leyes de vigencia en otros Estados:

#### *2.2.15.5.1. Resolución del Parlamento Europeo sobre uso policial de la Inteligencia Artificial*

Describe Adaya y Pardo (2021), que el pasado 6 de Octubre del 2021, se estableció la aprobación de la Resolución del Parlamento Europeo sobre uso policial de la IA con 377 votos a su favor, 248 en contra y 62 personas tomaron la decisión de no ejercer su derecho al voto. El presente reglamento recalca la proliferación del uso de la IA, por lo que advierte distintos riesgos sobre el uso de esta tecnología al momento de aplicarla

En este apartado se detallan en un listado los artículos Resolución del Parlamento Europeo propuestos para el 2021 en adelante sobre uso policial de la Inteligencia Artificial considerados en la presente investigación, dicho esto se complementa su descripción en el anexo 5 de la presente investigación.

- Art 6: describe que los sistemas que promueven la IA son de alto riesgo cuando no cumplen con el debido proceso en cuanto a la seguridad en el momento de finalización del producto, a su vez el mismo tiene que someterse a pruebas pertinentes de distintiva organización dispuesta en el mercado o servicio.

- Art 10: explica que los sistemas que aplican diversas técnicas para el posterior entrenamiento de modelos se desarrollarían partiendo de una o varias bases de datos que promoverán su respectiva validación y pruebas para el respectivo criterio de calidad. Para mas detalles del presente articulo véase anexo 5.



## **Capítulo III**

### **Metodología**

#### **3.1. Diseño de la investigación**

Es el conjunto de métodos y procedimientos en el que un investigador emplea la recolección y la distinción de medidas de las variables descritas en la investigación del problema. Esta medida incorpora el tipo de investigación, objetivo, técnicas empleadas, población, análisis de datos y los pasos para realizar el estudio. Por otro lado, el diseño de la investigación determina lo que incluye o no, estableciendo parámetros de cualquier investigación y se realiza la definición de los criterios para evaluar los resultados (Ayala, 2020).

#### **3.2. Propuesta tecnológica**

Este capítulo está conformado con respecto a la metódica que se utilizó en el diseño e implementación del prototipo web para la transmisión de sinopsis de sitios oficiales de noticias referentes al covid-19, por medio de algoritmo NLP para los habitantes de la zona-8 proveniente del Guayas respectivamente Guayaquil, Durán y Samborondón. Por este medio se entiende la determinada investigación que se empleó conformado por la población y muestra. El estudio utilizado fue exploratorio aplicando el método de revisión bibliográfica y metodología del prototipado.

##### ***3.2.1. Descripción del proceso metodológico***

Según Zita (2021), menciona que la investigación científica es una transformación ordenada y sistemática de total averiguación en el cual, por medio de la adaptación rígida de un agrupamiento de procedimientos y criterios, se busca el estudio, investigación o averiguación de lo que engloba un determinado tema, con la finalidad de superar el conocimiento del mismo. El objetivo es realizar búsquedas de soluciones a problemáticas tales como: explicación de fenómenos, desarrollo de teorías, ampliación de conocimientos, principios, reformulación de planteamientos, etc.; Permitiendo conocer el porqué de las cosas y su funcionamiento para adaptarlos a su existencia para mejorar una cualidad de vida.

##### ***3.2.2. Etapas de la investigación científica***

A continuación, se presentarán los pasos que se debe realizar para que un proyecto sea efectivo y de calidad.

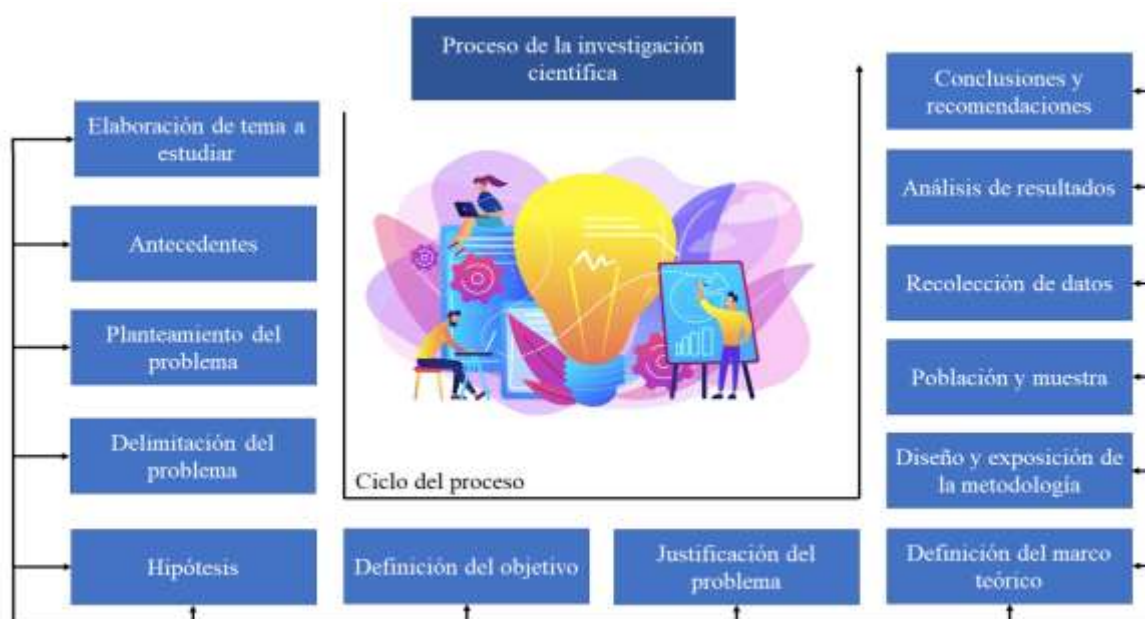


Figura 20. Proceso de la investigación científica. Información obtenida de libro por (Soliz, 2019). Elaborado por Daniela Taynys.

### 3.3. Tipos de investigación

Son las que pueden aglomerar según el objetivo que se persigue, su altitud de profundizar, la manipulación de variables, el modo de inferir estadísticamente, el tipo de datos en el transcurso del tiempo de estudio (Arias, 2020).

Los tipos de investigación se detallan por sus particularidades según su grado de investigación y distintivos tipos de estudio mencionados a continuación:

**Tabla 19** Tipos de Investigación utilizados en el estudio

Tipos de investigación	
<b>Por los medios para obtener los datos o fuente</b>	<ul style="list-style-type: none"> <li>Documental: se realiza por medio de consultas de documentos, con la finalidad de indagar en libros, revistas, periódicos, etc.</li> </ul>
<b>En función de su propósito o finalidad</b>	<ul style="list-style-type: none"> <li>Básica, pura o teórica: se utilizan para bases teóricas pudiéndose plantear con alcances exploratorios, descriptivos, etc.</li> <li>Aplicada: mediante la teoría se encarga de resolución de problemas prácticos basándose en los descubrimientos y soluciones, los alcances que se pueden dar son predictivos o explicativos.</li> </ul>
<b>Por su nivel de profundidad o alcance</b>	<ul style="list-style-type: none"> <li>Exploratoria: estos estudios se realizan cuando el investigador quiere familiarizarse con un</li> </ul>

fenómeno desconocido. Están conformados por variable cuantitativa y categórica.

- Descriptiva: permite especificar las características de cualquier objeto o fenómeno.
- Explicativa: se establece como causa – efecto entre sus variables.
- Cualitativa: pueden ser investigación de acción e investigación participativa y cuenta con estudio de: Casos, Etnográfico, Histórico.
- Cuantitativa: tiene datos directos y secundarios

#### **Por la naturaleza de los datos y la información**

---

*Información obtenida de libro por (Arias, 2021). Elaborado por Daniela Taynys.*

### **3.4. Metodología de investigación**

Según Cortés y Iglesias (2019), es la ciencia que permite conducir el proceso de un modo eficiente y tiende a la excelencia en el proceso de la investigación, por medio de una serie de principios, conceptos y leyes. Esta metodología también es llamada como Metodología Científica. Se enfoca en el objeto de estudio definido como el proceso, el cual está conformado por un conjunto de procedimientos de manera estructurada y relacionados entre sí.

A continuación, se redactan las metodologías de investigación que se implementarán en el presente estudio:

#### **3.4.1. Metodología Bibliográfica**

También conocido como método documental, gabinete de biblioteca, de la literatura, etc. Utiliza la recopilación de definiciones con la finalidad de obtener un conocimiento sistematizado de una investigación propia. Se deben utilizar herramientas que permitan hallar y seleccionar información útil en todas las documentaciones recolectadas.

##### ***3.4.1.1. Fases de la metodología bibliográfica***

Destaca Gómez et al. (2014), estos tipos de metodología se componen en tres fases:

**Definición del problema:** Tendrá un aporte sumamente amplio permitiendo la retroalimentación de la investigación, teniendo el problema bien definido para proceder a una revisión bibliográfica.

**Búsqueda de la información:** Se procede a la búsqueda de material informativo desde una perspectiva estructurada y profesional. Este material tiene que ser reconocido, es decir, que no contengan informaciones no fiables o sin mayores referencias. Estas se presentarán

en variedades de formatos, unos más reconocidos académicamente y mayormente valorados que otros.

**Organización de la información:** Consiste en ordenar de manera sistemática la documentación hallada. Estas informaciones pueden ser ordenadas en hojas de cálculos o carpetas, siendo implementadas por el propio investigador de forma manual. Por otro lado, también se puede hacer usos de herramientas informáticas tales como: Mendeley Desktop, JabRef, Reference, etc. Permitiendo organizar la debida información obteniendo finalmente la bibliografía para el informe final.

**Distinción de la información:** finalmente se realiza un análisis de la información ya organizada, seleccionando los documentos más útiles para el tema de estudio.

La metodología bibliográfica es una fase de la investigación perteneciente a la ciencia donde se examina la ejecución de la sociedad académica sobre un determinado tema. Por este método se obtiene un agrupamiento de acciones predispuestas a ubicar documentos relacionados con un asunto. Adicional permite reconocer el estado del arte de lo que se indagando determinando el punto de partida en el que debe basarse cualquier tarea científica (Universidad De La República Uruguay, 2020)

Se realizará la utilización de la metodología bibliográfica para el levantamiento de información y así poder reforzar varios tipos de conocimientos de manera sistemática. Esta metodología se realizará, a través, de la recopilación de artículos, investigaciones, blogs científicos, entre otras, accediendo a bases de datos científicas como: Scielo, Dialnet, Redalyc, Sciencedirect, Google académico con datos confiables para la implementación de este proyecto, permitiendo llevar una buena documentación y comprensión de la misma.

### **3.4.2. Metodología Exploratoria**

Es un estudio preliminar con el objetivo de familiarizarse con un fenómeno a investigar, muy utilizado para desarrollar o refinar hipótesis, de modo que conlleve a una mejor comprensión y precisión del estudio. Permite elegir las técnicas más adecuadas en la investigación. Existen dos razones básicas, primero el investigador puede ser curioso y desear una comprensión sobre el fenómeno de interés y segundo para estimación de viabilidad y costo de emprender un proyecto más riguroso. Por lo que, es prudente que se haga una exploración más pequeña del estudio primordialmente (Piovesan & Temporini, 1995).

Es aquella que aborda un asunto poco estudiado, por lo que se tiene en consideración una visión aproximada de dicho tema, es decir, que permitirá recopilar información para

tener una mejor comprensión general del tema de investigación, a través, de antecedentes, localización de aspectos relevantes y relaciones entre tendencias y variables, que profundizan otras investigaciones. Por otro lado, esta metodología no explica el objeto de la investigación (Alava & Díaz, 2022).

Se aplicará la metodología exploratoria para abordar conocimientos con distintas herramientas informáticas como un tema innovador, es decir, recién implementado siendo un prototipo de inicialización para precisar las preguntas de investigación implementadas inicialmente. Por otro lado, al recopilar distintas fuentes de información sobre el tema se obtendrá una comprensión en el funcionamiento del prototipo de una manera estructurada, conllevando una mejor precisión en el estudio.

### **3.4.3. Metodología Cualitativa**

Es una agrupación de técnicas de investigación utilizados para obtención de una visión general del comportamiento y la percepción sobre el tema tratado. Ayuda a entender como es percibido un problema, por lo que, es muy utilizada en las principales fases de los trabajos de estudio científico (Latinoamérica, 2019).

Esta metodología también es conocida como naturalista fenomenológica, interpretativa o etnográfica, en la que utiliza como técnica fundamental la observación de realidades subjetivas y permite facilitar el conocimiento de diferentes culturas promoviendo al investigador distintas representaciones en cuanto a la exploración del conocimiento y en la forma en que los colaboradores comparten sus experiencias (Piza et al., 2018).

En este proyecto se implementará esta metodología para la realización de entrevistas a profesionales que aborden conocimientos de desarrollos de NLP con enfoque pragmático y construcción del prototipo web, para poder seguir lineamientos adecuados correspondientes al proyecto.

### **3.4.4. Metodología Cuantitativa**

Es una metodología que se basa en la resolución de las investigaciones preservando los aspectos fundamentales de los fenómenos, por lo que consiste en contrastar las teorías ya existentes por medio de variedades de hipótesis surgidas por la misma, para ello es necesario la obtención de una muestra, ya sea de forma aleatoria o discriminada. Este método se obtiene mediante la forma representativa los datos recogidos tanto poblacional como del fenómeno del objeto de estudio. Es indispensable poder contar con una teoría principal previamente construida, por lo que este método científico es deductivo (Delgado et al., 2018).

Además, se implementará la metodología cuantitativa para determinar las variables a medir por medio de la implementación de encuestas, las mismas ayudaran a verificar los datos corroborando que en la investigación se recolectaron datos oficiales, determinando la necesidad existente de la población de la zona-8, con la finalidad de presentar la propuesta en cuanto a la creación de un prototipo web que centralice las noticias oficiales referentes al Covid-19.

### 3.4.5. Metodología Mixta

Según Jiménez, et al. (2019) es una combinación de metodologías como cualitativa y cuantitativa, siendo la más adecuadas para realizar un estudio científico debido a esta implementación mixta. Por otro lado, permite tener un enfoque pragmático en el mundo de la investigación empírica, por lo que está basada en la recolección de datos desde diferentes perspectivas (cualitativas y cuantitativas) proporcionando un mejor entendimiento del fenómeno del estudio.

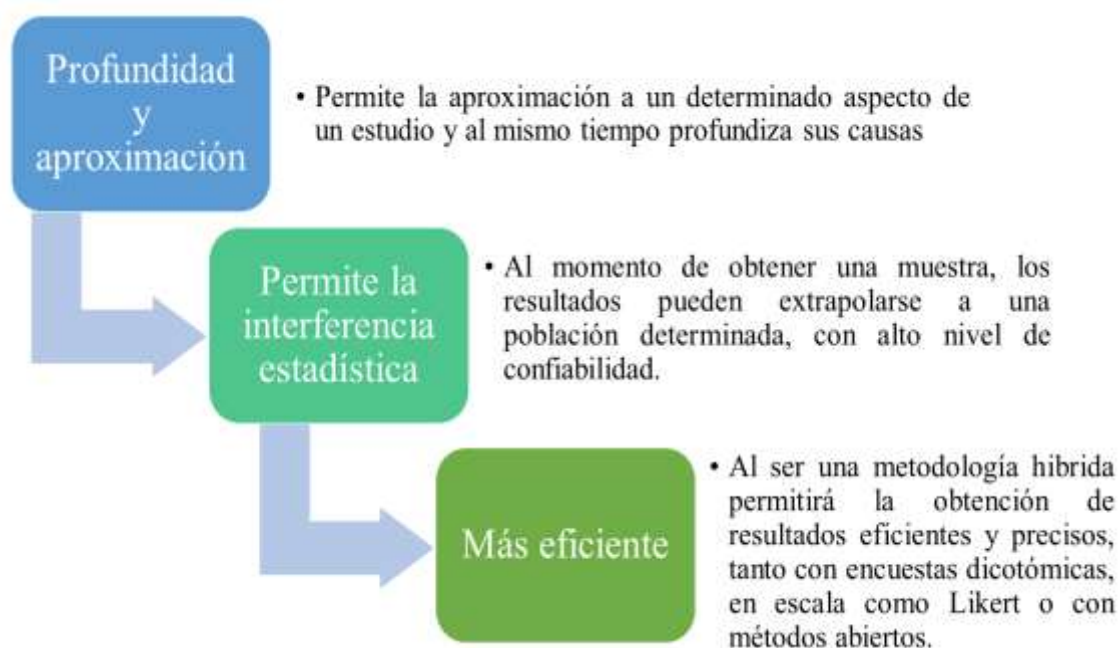


Figura 21. Características de la metodología mixta. Información tomada de artículo Lifeder por (Ayala, 2021). Elaborado por Daniela Taynys.

Se hará uso de la metodología mixta para poder obtener una mejor comprensión de la problemática del proyecto, por lo que enriquecerá la investigación con mayor profundidad, amplitud, interpretación y comprensión, permitiendo justificar la necesidad, descubriendo los problemas, relacionándolos y por último cuantificarlos.

### 3.4.6. Metodología del Prototipado

Es aquella que permite llevar a cabo la realización y materialización de diferentes ideas para soluciones propuestas en un proyecto en cuanto al diseño o rediseño de productos o

servicios y está relacionado en la representación de escenarios en cuanto a los resultados finales involucrando la experiencia de los usuarios (Sánchez y Gereá, 2021).

Es una metodología de evaluación donde se destaca el procedimiento a seguir para establecer la aceptabilidad de un prototipo a ser usado para determinado propósito, de tal forma que se acomode a la necesidad que requiere el usuario, por lo que, el mismo debe ser ajustado de una manera fácil y económica para mitigar riesgos que se puedan dar en producción (Sarraipa et al., 2019).

Describe Guamán & Merino (2019), que el uso de la metodología del prototipado define la fidelidad determinando el grado de funcionalidad que establece un prototipo, por lo que se adapta a diferentes fases de desarrollo dependiendo del grado. A lo largo del tiempo se han establecido tres prototipos los cuales son:

**Baja fidelidad:** Se utilizan los requerimientos previos para la creación de bosquejos determinándose el nivel más bajo y con menos costo donde se pueden desarrollar múltiples requerimientos de los usuarios con grandes cantidades de cambios, llevando mejoras a medida que avanza el proyecto.

**Media fidelidad:** Se basa en el prototipado de baja fidelidad previamente aprobado, por lo que se aumenta funcionalidades entre pantallas, tomando en consideración que aún no se establece conexión con bases de datos, siendo más atractiva visualmente para los internautas.

**Alta fidelidad:** Para este tipo de prototipado se establece todas las funcionalidades del sistema, incorporando las conexiones a datos e interfaces que cumplen las diversas necesidades, por lo que en esta etapa es difícil y costoso la realización de cambios del proyecto.

Indica Pérez & Codina (2019), que existen los denominados Blueprints y Wireframes. Blueprints también conocido como plano o mapa, es un esquema de representación arquitectónica de la página, poniéndose en relación las páginas existentes con la página principal y los contenidos que la conforman. Por otro lado, los Wireframes o maqueta realizan una representación arquitectónica de cada una de las vistas que conforman el sitio web, las mismas incluyen funciones, estructura y contenido. Permite visualizar una situación estática donde no hay cabida a la interacción de los usuarios, es decir que representa un momento en concreto a todo el proceso que el internauta llevará a cabo dentro del sitio web.



Figura 22. ¿Para qué sirve prototipar? Información tomada de (Sánchez y Gereá, 2021). Elaborado por Daniela Taynys.

Por último, se implementará la metodología del prototipado donde primordialmente se hará uso de los tres tipos de prototipos tales como: baja fidelidad, media fidelidad y alta fidelidad, encajando cada uno de los pasos dando como resultado las funcionalidades pertinentes del presente proyecto. Por otro lado, se hará uso de la herramienta Marvel App para la creación rápida del prototipo, siendo interactivo de forma digital, por medio de la creación de Wireframes y realizando la función del prototipado con transiciones animadas entre diferentes tipos de pantalla.

### 3.5. Técnicas de investigación

#### 3.5.1 Encuesta

Según Torres et al (2019), la encuesta es aquella que es realizada a través de una muestra de sujeto representativa, por lo que establece la expresión del medio ya sea por la examinación o la experimentación. Utiliza procesos estandarizados de interrogación con la finalidad de cuantificar una gran cantidad de variedad en cuanto a características tanto objetivas como subjetivas de una determinada población.

Esta técnica de recolección se realizó debido a la necesidad de probar hipótesis surgidas de soluciones referenciadas a la problemática inicialmente dada, por lo que se identifica e interpreta un conjunto de opiniones con la finalidad de servir con un propósito, basándose en datos y no en suposiciones.

La encuesta fue realizada en la población de la zona-8 realizando una recolección de 575 ciudadanos, teniendo como finalidad de apoyo en cuando a impulsar la propuesta en la



creación de un prototipo web de centralización de noticias oficiales referentes al Covid-19. La herramienta tecnológica que se utilizó es Google Forms, debido a que es útil para las tabulaciones de las contestaciones del formulario siendo transferible, por medio del envío a través de medios comunicativos, con ello se permite validación en torno a la problemática en la muestra correspondiente a una población infinita.

### **3.5.2 Entrevista**

Es la recolección de información de manera significativa sobre un tema de interés, por lo que se requiere experiencia del entrevistador brindando información de forma directa sobre los conocimientos relevantes, obteniendo un método de validación útil para la verificación de fiabilidad por medio de juicios de expertos (Marín, 2019).

Se realizó entrevistas como técnicas de investigación para la recolección de datos en la que se hacía variedades de preguntas con la finalidad de conocer sugerencias de especialistas en cuanto a tecnologías con conocimientos en IA y Desarrollo de Web, por lo que se llevó a cabo una muestra aleatoria en cuanto a expertos propuestos a compartir sus experiencias y conocimientos.

## **3.6. Descripción del procedimiento metodológico**

### **3.6.1. Población**

La población tomada para este estudio está compuesta por los moradores que habitan en la zona 8 de la provincia del Guayas, la cual se encuentra distribuida por los cantones de Guayaquil, Samborondón y Durán. Como lo indico el Instituto Nacional de Estadísticas y Censos (INEC), (2010), es una institución encargada de la transmisión de información estadística en el Ecuador con un alcance demográfico, económico, de agricultura, educativo, salud y bienestar, entre otros. Por lo que incluye en su base de datos (BD) para el año 2010 el cantón Guayaquil consto de 2.350.915, el cantón de Samborondón consto de 67.590 y el cantón de Durán consto de 235.769 habitantes, por lo que en total se suma 2.654.274 ciudadanos pertenecientes a la zona 8.

En las correspondientes encuestas se muestra la variable principal denominado como el tipo de estudio cualitativo siendo proporcional al fenómeno estudiado en la población de referencia. Por otro lado, se considera que la población infinita es cuando no se conoce el total de unidades de inspección o cuando la población supera a los 10,000 (Castro, 2019).

### **3.6.2. Muestra**

Es aquella que se divide en dos categorías tales como: probabilísticos y no probabilísticos, por lo que los probabilísticos son aquellos que están basados en la igualdad

de probabilidad, es decir, que tenga la misma posibilidad de ser seleccionado para representarla y formen parte de la muestra, soliendo ser los más utilizados ya que buscan una mayor representatividad. En cambio, los no probabilísticos se utilizan cuidadosamente con criterios específicos en la población de manera que sean lo más representativos posibles (Ávila, 2019).

De esta manera se empleó las técnicas estadísticas de población infinita que son encargadas de obtener datos respectivos de una población. por lo que se procede a calcular la muestra a través de la siguiente formula:

$$n = \frac{Z_{\alpha}^2 p * q}{d^2}$$

$$n = \frac{1.96_{\alpha}^2 0.5 * 0.5}{0,041^2}$$

$$n = \frac{3.8416 * 0.5 * 0.5}{0,001681}$$

$$n = \frac{0,9604}{0,001681}$$

$$n = 571$$

Donde:

n= tamaño de la muestra

NC = nivel de confianza deseado = 95%

Z= se define según el NC = 1.96

p = proporción de la población deseada (éxito) = 0.50

q = proporción de la población no deseada (fracaso) 1-p (1-0.05) =0.50

d= nivel de error disponible 4.1% = 0.041

### 3.7. Análisis de las encuestas

#### Pregunta N° 1.1- Seleccione la edad

**Tabla 20** Edad

Opciones de respuestas	Encuestados	Porcentaje
De 14 a 24 años	385	67%
De 25 a 34 años	123	21%
De 35 a 44 años	40	7%
De 45 a 54 años	20	3%
De 55 a 65 años	7	1%
<b>Total</b>	<b>575</b>	<b>100%</b>

*Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*

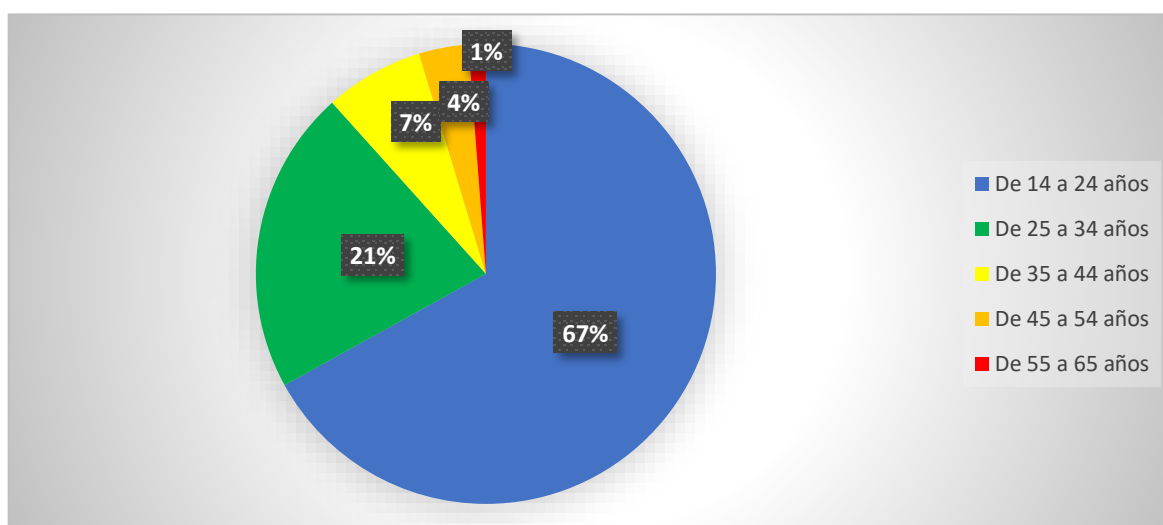


Figura 23. Edad de los encuestados. Información tomada de formulario de Google. Elaborado por Daniela Taynys

De acuerdo con los resultados del gráfico estadístico N°23 se puede visualizar que de la muestra de 575 ciudadanos pertenecientes a la zona 8 de la provincia del Guayas, el rango de edades entre 14 a 34 años presenta una participación del 88%, mientras que edades entre 35 a 54 años contiene un porcentaje del 11% y el porcentaje más bajo esta entre 55 a 65 años de edad establecido como el 1% de los datos estadísticos.

#### Pregunta N° 1.2.- Seleccione el Género con el que se identifica

**Tabla 21** Género

Opciones de respuestas	Encuestados	Porcentaje
Femenino	318	55%
Masculino	257	45%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

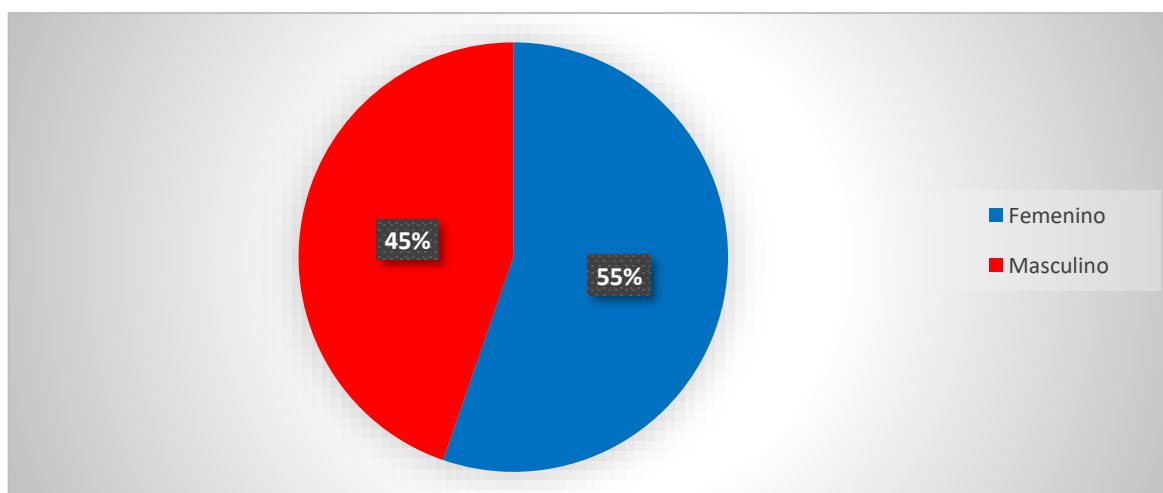


Figura 24. Género de los encuestados. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Conforme al gráfico estadístico N°24 se observa que de la muestra de 575 de los encuestados correspondientes a la zona 8 de la provincia del Guayas, se identifica que existe una diferencia mínima del 5% entre ambos géneros, por lo que el género femenino tiene un mayor porcentaje constando con el 55% de la población lográndose una participación sustancial de los dos géneros.

### Pregunta N° 1.3.- Lugar que reside de la zona 8 del Ecuador

**Tabla 22** Lugar que reside de la zona 8 del Ecuador

Opciones de respuestas	Encuestados	Porcentaje
Guayaquil	519	90%
Durán	40	7%
Samboorondón	16	3%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

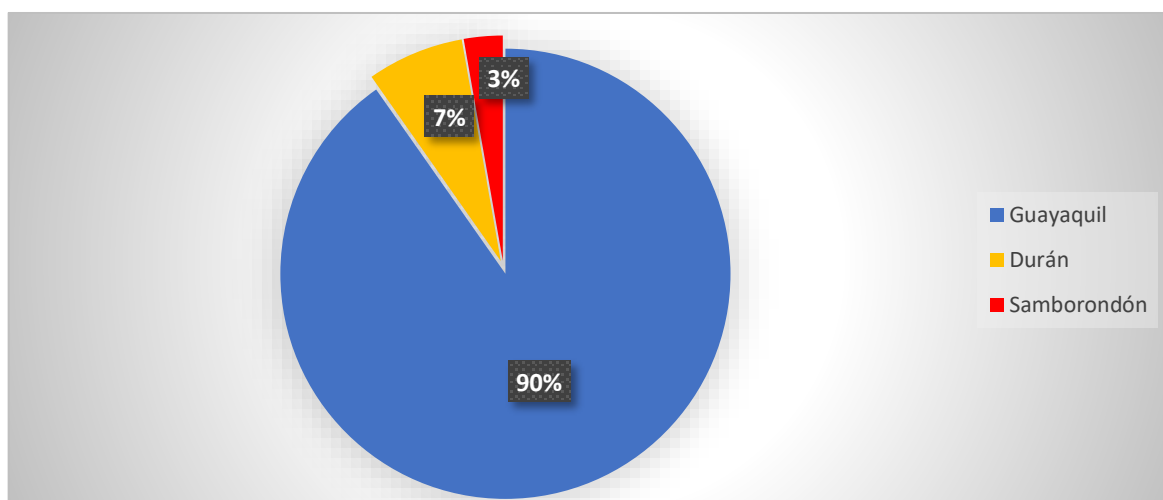


Figura 25. Residentes de la Zona 8. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Respecto al gráfico estadístico N°25 se observa que de la muestra de n=575 en relación con el lugar donde reside cada encuestado, se determina que la mayor cantidad de ciudadanos se encuentra en el Cantón de Guayaquil alcanzando el 90%, por otro lado, el porcentaje más bajo que conlleva el 3% de la población encuestada es de Samboorondón.

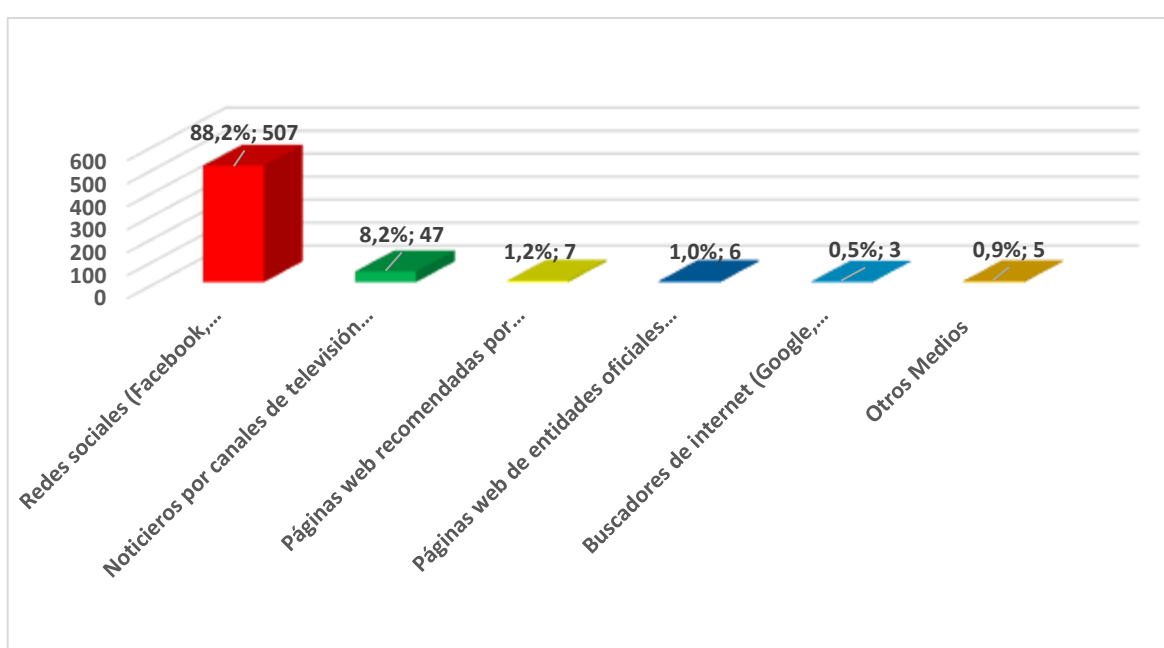
### Pregunta N° 2.1.- ¿A través de que medios de comunicación usted se informa sobre noticias del covid-19?

**Tabla 23** Medios que la población utiliza para informarse sobre noticias del Covid-19

Opciones de respuestas	Encuestados	Porcentaje
Redes sociales (Facebook, Instagram, Twitter, YouTube,etc)	507	88,2%

Noticieros por canales de televisión o radio	47	8,2%
Páginas web recomendadas por conocidos	7	1,2%
Páginas web de entidades oficiales (Ministerio de Salud, Organización Mundial de la Salud, entre otros)	6	1,0%
Buscadores de internet (Google, Bing, Yahoo!, Ask, etc)	3	0,5%
Otros Medios	5	0,9%
Total	575	100%

*Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*



*Figura 26. Medios que la población utiliza para informarse sobre noticias del Covid-19. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*

De acuerdo con el gráfico estadístico N°26 se puede visualizar que de la muestra de encuestados  $n=575$ , el 88,2% de la población utiliza redes sociales por lo que para este estudio toma en cuenta dicha problemática en la utilización de estos medios comunicativos informales sobre noticias del Covid-19, mientras que el 8,2% se informa por canales de televisión o radio, el 2,7% engloba páginas web recomendadas por conocidos, entidades oficiales y buscadores de internet, por último el 0,9% usa otros medios.

**Pregunta N° 2.2.- Cuando usted busca información reciente relacionada con el covid-19, de los portales que ofrece la información. ¿Cuál es su nivel de dificultad que usted ha encontrado al navegar o interactuar en ellos para encontrar la noticia de su interés?**

**Tabla 24** Nivel de dificultad al buscar información en los portales de información

Opciones de respuestas	Encuestados	Porcentaje
Extremadamente fácil	122	21,2%
Muy fácil	237	41,2%
Ni fácil ni difícil	187	32,5%
Medianamente difícil	28	4,9%
Extremadamente difícil	1	0,2%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

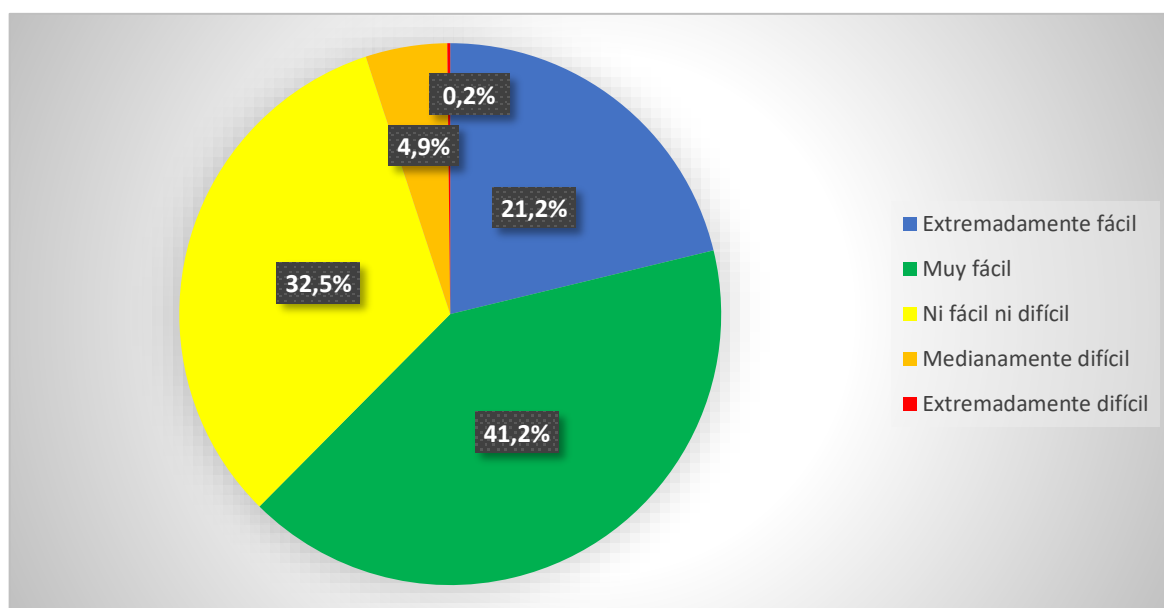


Figura 27. Nivel de dificultad de los portales que ofrece información reciente del Covid-19 Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Conforme al gráfico N°27 de la muestra de 575 ciudadanos encuestados de la zona 8 de la provincia del Guayas, el 62,4% destaca que es extremadamente fácil y muy fácil interactuar con portales que ofrece información acerca del covid-19, mientras que el 37,4% indica que es ni fácil ni difícil y medianamente difícil, finalmente el 0,2% opina que es extremadamente difícil.

**Pregunta N° 2.3.- ¿Está usted de acuerdo en que se impulse un proyecto que construya un sitio web que centralice las noticias relacionadas con el covid-19 las cuales se encuentran dispersas en múltiples páginas oficiales?**

**Tabla 25** Nivel de aceptabilidad en la construcción de un sitio web para centralizar noticias del covid-19

Opciones de respuestas	Encuestados	Porcentaje
Totalmente de acuerdo	398	69,2%
Parcialmente de acuerdo	117	20,3%

Ni de acuerdo ni desacuerdo	51	8,9%
Parcialmente en desacuerdo	5	0,9%
Totalmente en desacuerdo	4	0,7%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

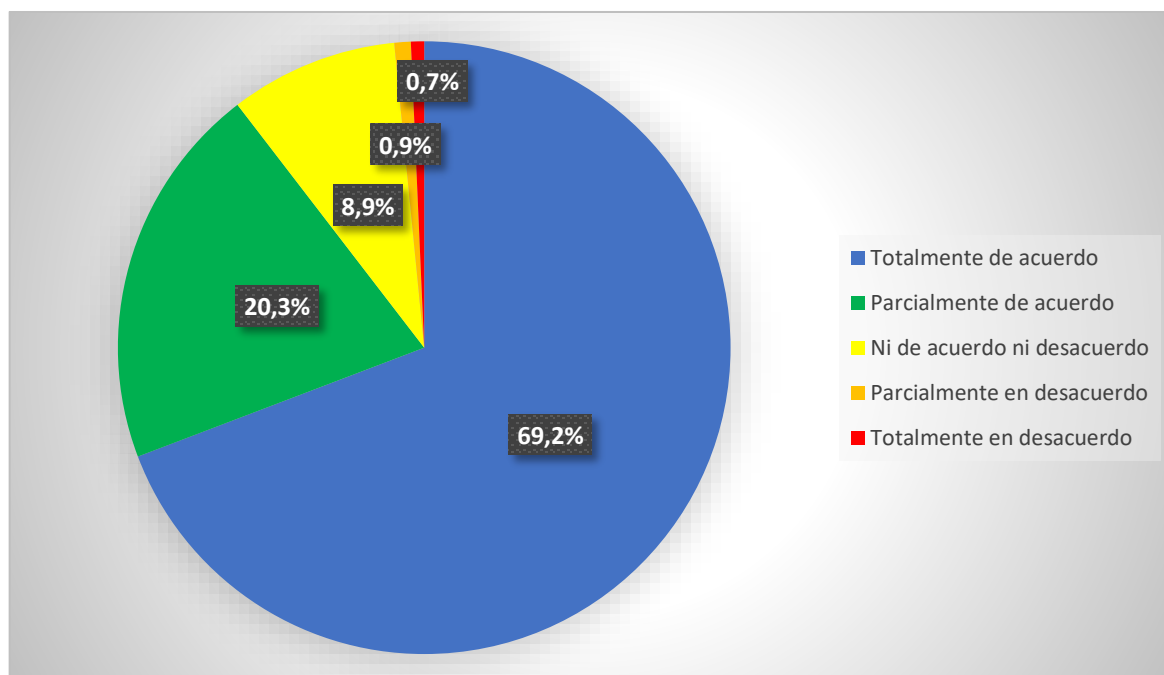


Figura 28. Nivel de aceptabilidad en la creación de un sitio web para centralización de noticias oficiales referentes al covid-19. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

De acuerdo con el gráfico estadístico N°28 de la muestra de 575 ciudadanos indica que está totalmente de acuerdo y parcialmente de acuerdo en cuanto a la creación de un sitio web que centralice las noticias del covid-19, llevando un nivel de aceptabilidad del 89,5%, mientras que el 9,8% está ni de acuerdo ni en desacuerdo y parcialmente de desacuerdo, por último, el 0,7% de la población están en total desacuerdo.

**Pregunta N° 2.4.- ¿Al contar con un sitio web el cual centralicen las noticias de múltiples páginas oficiales del covid-19, considera usted que su nivel de preocupación se disminuiría al poder verificar con la ayuda de este portal la veracidad de noticias que llegan por redes sociales u otros medios informales?**

**Tabla 26** Nivel de consideración al contar con un sitio web que permita verificar las noticias oficiales del covid-19

Opciones de respuestas	Encuestados	Porcentaje
Totalmente de acuerdo	304	52,870%
Parcialmente de acuerdo	182	31,652%

Ni de acuerdo ni desacuerdo	78	13,565%
Parcialmente en desacuerdo	7	1,217%
Totalmente en desacuerdo	4	0,696%
Total	575	100%

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

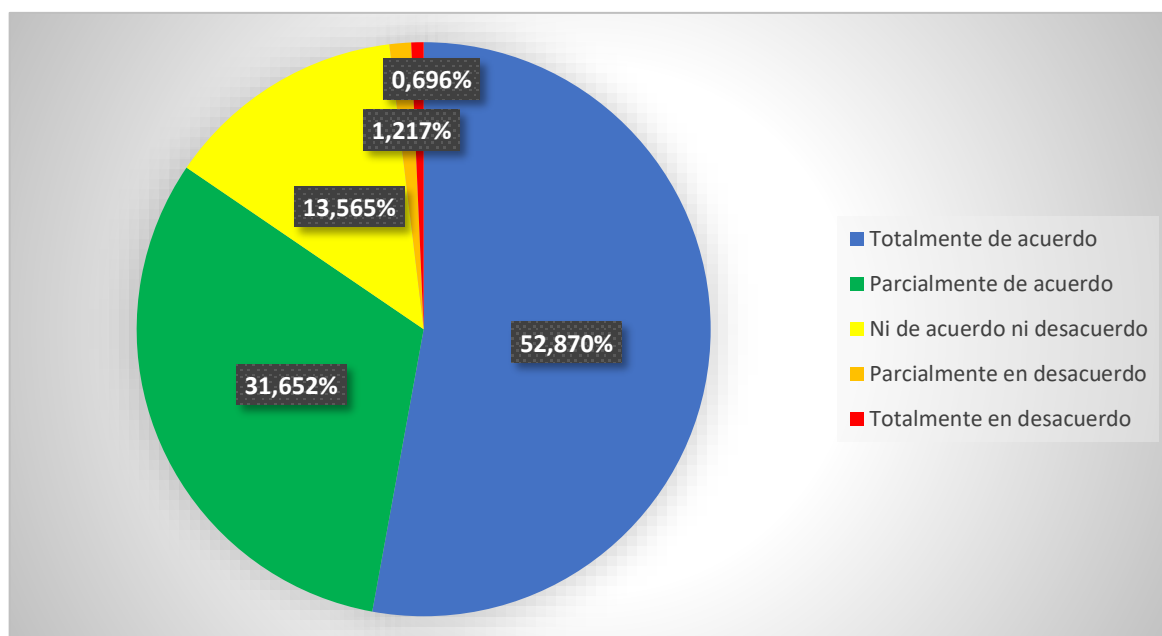


Figura 29. Nivel de consideración al contar con un sitio web que permita verificar las noticias oficiales del covid-19. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Respecto al gráfico N° 29 de un total de 575 habitantes encuestados de la zona 8 de la provincia del Guayas, indican su nivel de consideración entre totalmente de acuerdo y parcialmente de acuerdo con un 84,522% de los ciudadanos en que exista un portal web donde centralice las noticias relacionadas al covid-19, mientras que el 14,782% está ni de acuerdo ni en desacuerdo y parcialmente desacuerdo, por último, el 0,696% de la población está totalmente en desacuerdo.

**Pregunta N° 2.5.- ¿Considera usted que el contar con un sitio web que centralicen las noticias de múltiples páginas oficiales del covid-19 le ayudarían a usted a evitar el uso y/o compartición de noticias falsas (Fake News)?**

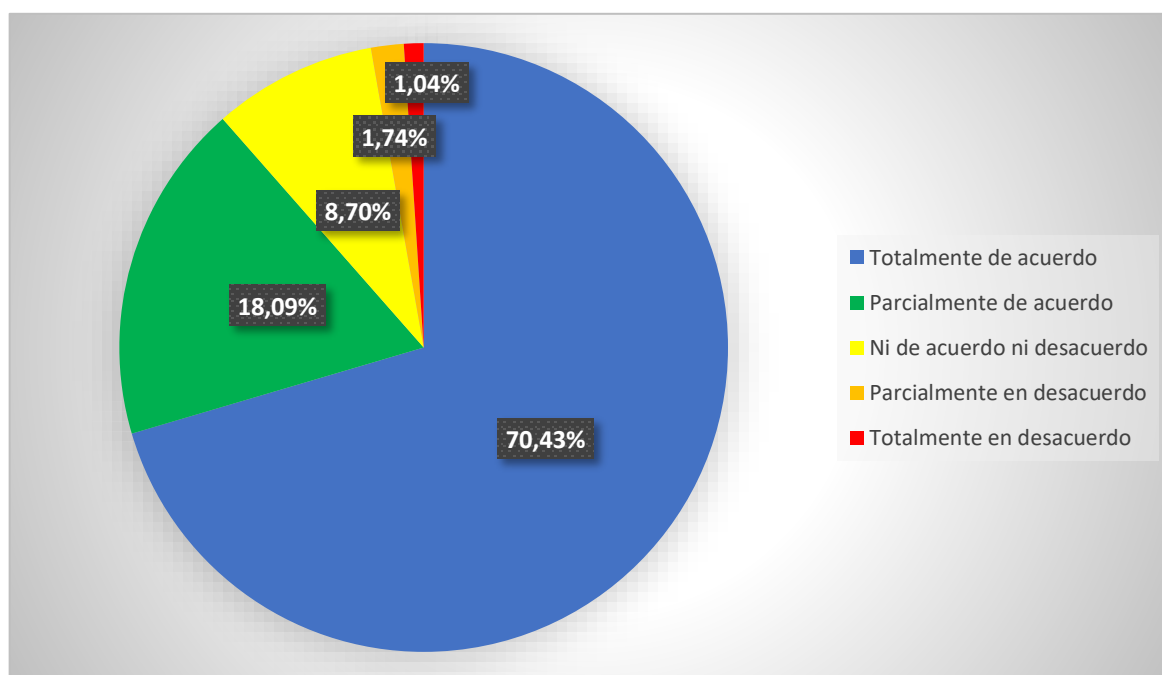
**Tabla 27** Nivel de estimación al contar con un sitio web que permita evitar la compartición de noticias falsas

Opciones de respuestas	Encuestados	Porcentaje
Totalmente de acuerdo	405	70,4%
Parcialmente de acuerdo	104	18,1%



Ni de acuerdo ni desacuerdo	50	8,7%
Parcialmente en desacuerdo	10	1,7%
Totalmente en desacuerdo	6	1%
<b>Total</b>	<b>575</b>	<b>100%</b>

*Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*



*Figura 30. Nivel de consideración sobre la existencia de un sitio web que permita centralizar las noticias de múltiples páginas oficiales para evitar los Fake News. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*

Conforme al gráfico N°30 establece que de la muestra  $n=575$  el 88,52% de la población conlleva un nivel de aceptabilidad entre totalmente de acuerdo y parcialmente de acuerdo en que exista un sitio web que centralice las noticias referentes al covid-19 para evitar la generación de Fake News, el 10,44% se encuentran ni de acuerdo ni desacuerdo y parcialmente en desacuerdo, finalmente el 1,04% está totalmente en desacuerdo.

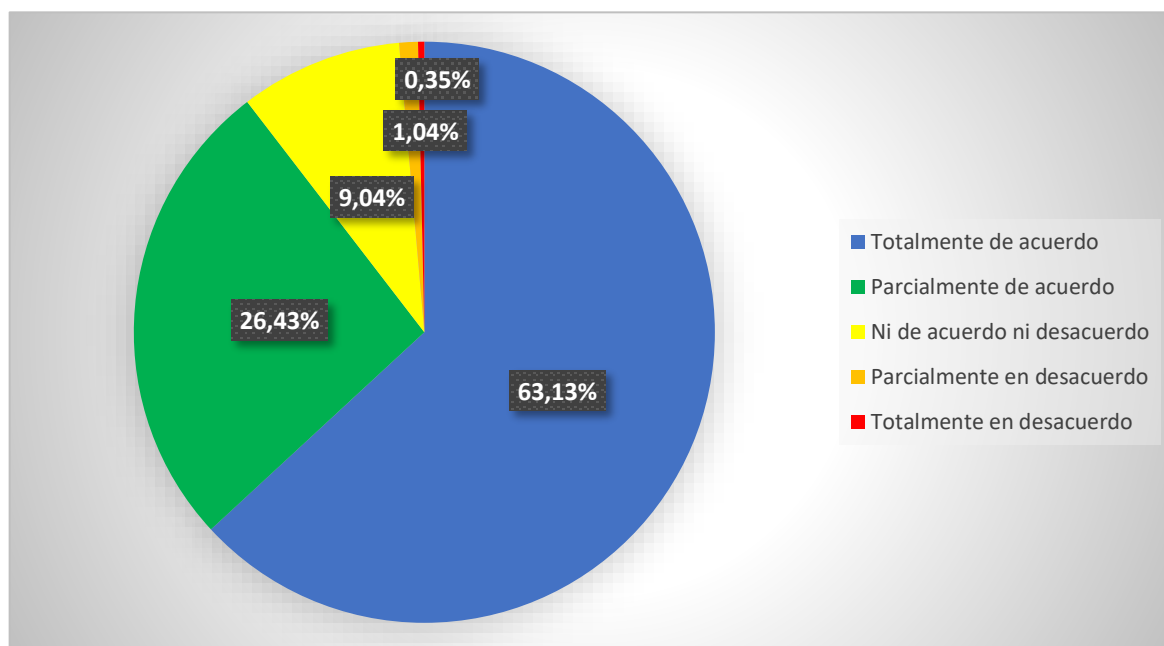
**Pregunta N° 2.6.- ¿Considera usted importante contar con esta herramienta tecnológica de centralización de noticias del covid-19 y que la misma le permita llegar al detalle de la noticia de interés en el respectivo sitio oficial?**

**Tabla 28** Nivel de consideración de importancia al contar con un sitio web que permita centralizar las noticias del covid-19 y que la misma redirija al sitio oficial

Opciones de respuestas	Encuestados	Porcentaje
Totalmente de acuerdo	363	63,13%
Parcialmente de acuerdo	152	26,43%

Ni de acuerdo ni desacuerdo	52	9,04%
Parcialmente en desacuerdo	6	1,04%
Totalmente en desacuerdo	2	0,35%
<b>Total</b>	<b>575</b>	<b>100%</b>

*Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*



*Figura 31. Importancia sobre la existencia de un sitio web que permita centralizar las noticias del covid-19 y que la misma redireccione al sitio oficial. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.*

De acuerdo con el gráfico N°31 acogido a una muestra de n=575, el 89,56% está totalmente de acuerdo y parcialmente de acuerdo en que es importante contar con la herramienta tecnológica para centralizar noticias del covid-19 y que la misma permita redireccionar a la página oficial, el 10,08% están ni de acuerdo ni en desacuerdo y parcialmente en desacuerdo, por último, el 0,35% estima que está totalmente en desacuerdo.

**Pregunta N° 2.7.- ¿Cuál es el tiempo que tiene disponible en el día para revisión de noticias referentes al covid-19 a través de dispositivos digitales (Celular, Laptop, PC)?**

**Tabla 29** Tiempo disponible para noticias referentes al covid-19 en dispositivos digitales

Opciones de respuestas	Encuestados	Porcentaje
Alto tiempo de disponibilidad (entre 3 a 5 horas)	83	14%
Medio tiempo de disponibilidad (entre 2 a 3 horas)	139	24%
Poco tiempo de disponibilidad (entre 1 a 2 horas)	163	28%
Muy poco tiempo de disponibilidad (Menos de 1 hora)	181	31%

No poseo tiempo disponible	9	2%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

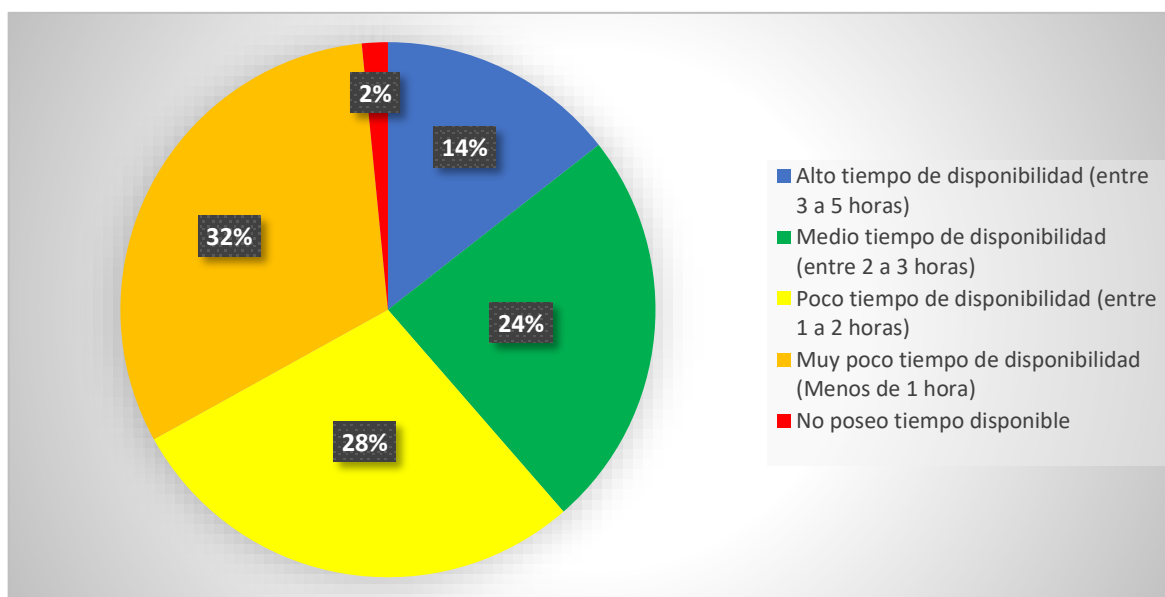


Figura 32. Tiempo de disponibilidad para revisión de noticias de covid-19 en dispositivos digitales. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Respecto al gráfico N°32 de un total de 575 encuestados pertenecientes a la zona 8 de la provincia del Guayas, el 38% de la población posee medio y alto tiempo de disponibilidad entre el rango de 5 a 2 horas, el 60% tiene poco y muy poco tiempo de disponibilidad entre 2 y menos de 1 hora, por último, el 2% de la población no posee tiempo disponible.

### Pregunta N° 3.1.- ¿Usted cuenta con al menos un dispositivo digital?

**Tabla 30** Encuestados que tengan al menos un dispositivo

Opciones de respuestas	Encuestados	Porcentaje
Si cuento con un Smart Phone(Celular) de Gama Media	260	45,2%
Si cuento con un Smart Phone(Celular) de Gama Alta	156	27,1%
Si cuento con una Laptop	62	10,8%
Si cuento con un Smart Phone(Celular) de Gama Baja	34	5,9%
Si cuento con un Ordenador (PC de escritorio)	27	4,7%
Si cuento con un Smart TV	22	3,8%
Si cuento con una Tabled	12	2,1%
No cuento con ningún dispositivo digital	2	0,3%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

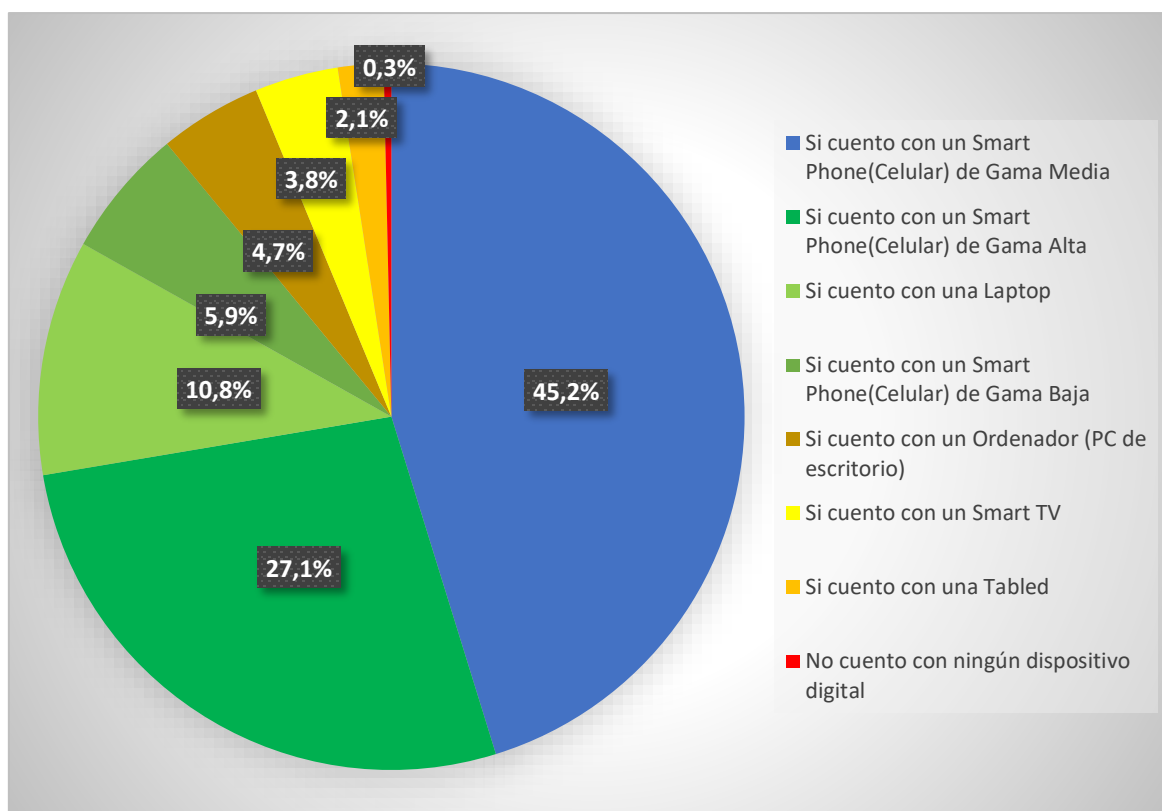


Figura 33. Encuestados que tengan al menos un dispositivo digital. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

De acuerdo con el gráfico estadístico N°33 de la muestra de 575 pobladores que habitan en la zona 8, el 72,3% indica que posee con al menos un dispositivo digital Smart Phone de Gama Media o Gama alta, el 16,7% cuenta por lo menos con una Laptop o un Smart Phone de Gama baja, el 8,5% cuenta con al menos un Ordenador o Smart Tv, el 2,1% establece que cuenta con al menos una tabled y el 0,3% no cuenta con ningún dispositivo digital.

**Pregunta N° 3.2.- ¿Cuál es el tipo de enlace a internet que emplea habitualmente?**

**Tabla 31** Encuestados que tengan al menos un dispositivo

Opciones de respuestas	Encuestados	Porcentaje
Acceso por Fibra Óptica	220	38%
Acceso Inalámbrico (Acceso internet público,WIFI)	150	26%
Acceso Telefónico (Teléfono convencional)	60	10%
Acceso por Teléfono Móvil (Megas)	57	10%
Acceso por Cable modem (TV CABLE)	48	8%
Acceso por Banda Ancha ADSL	27	5%
Otros medios	13	2%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

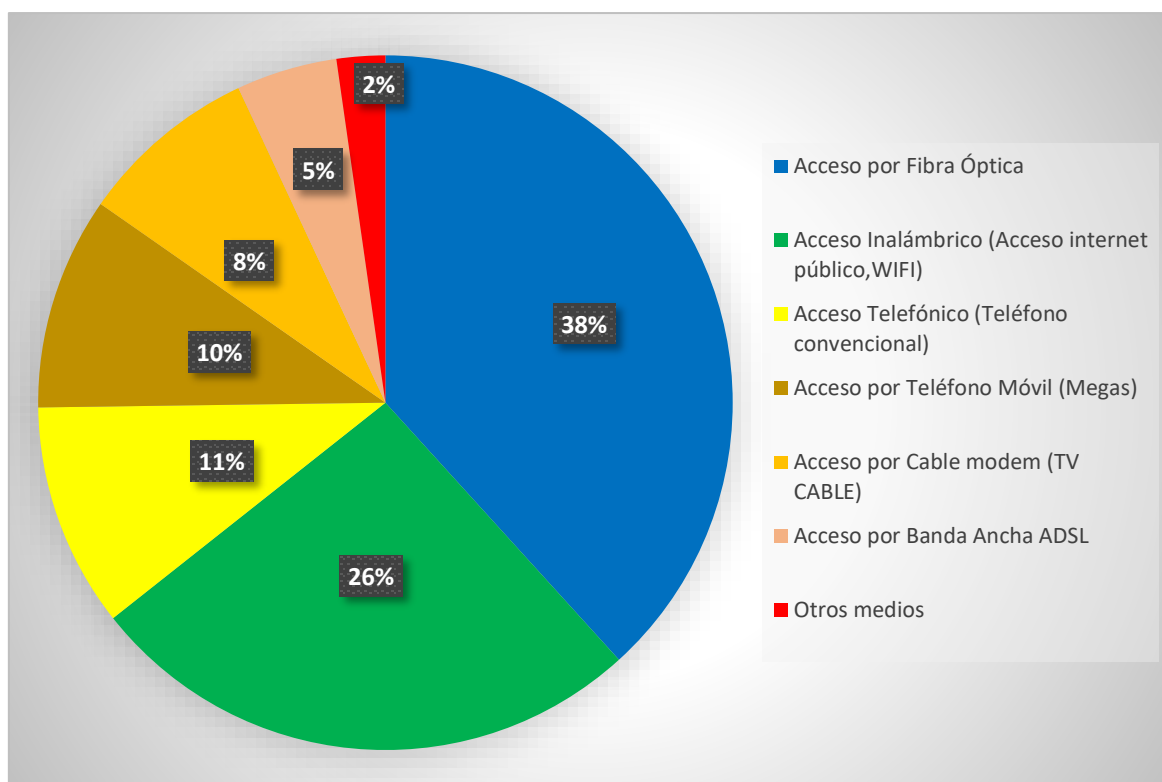


Figura 34. Tipo de enlace que emplea habitualmente. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Conforme al gráfico N°34 establece que de la muestra  $n=575$ , el 64% de la población cuenta con acceso por fibra óptica o inalámbrico, el 21% establece que tiene acceso por vía telefónica o megas, el 13% de la población dispone de acceso por cable modem y banda ancha, por último, el 2% de la población indica que utiliza otros medios.

### Pregunta N° 3.2.- ¿Cuál es el ancho de banda del enlace que usted utiliza?

**Tabla 32** Encuestados indicando su tipo de ancho de banda

Opciones de respuestas	Encuestados	Porcentaje
1 Mbps a 25 Mbps	107	18,6%
26 Mbps a 50 Mbps	185	32,2%
51 Mbps a 75 Mbps	114	19,8%
76 Mbps a 100 Mbps	81	14,1%
Otra velocidad de internet	10	1,7%
Actualmente no cuento con esa información	77	13,4%
Actualmente no poseo conectividad de internet	1	0,2%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

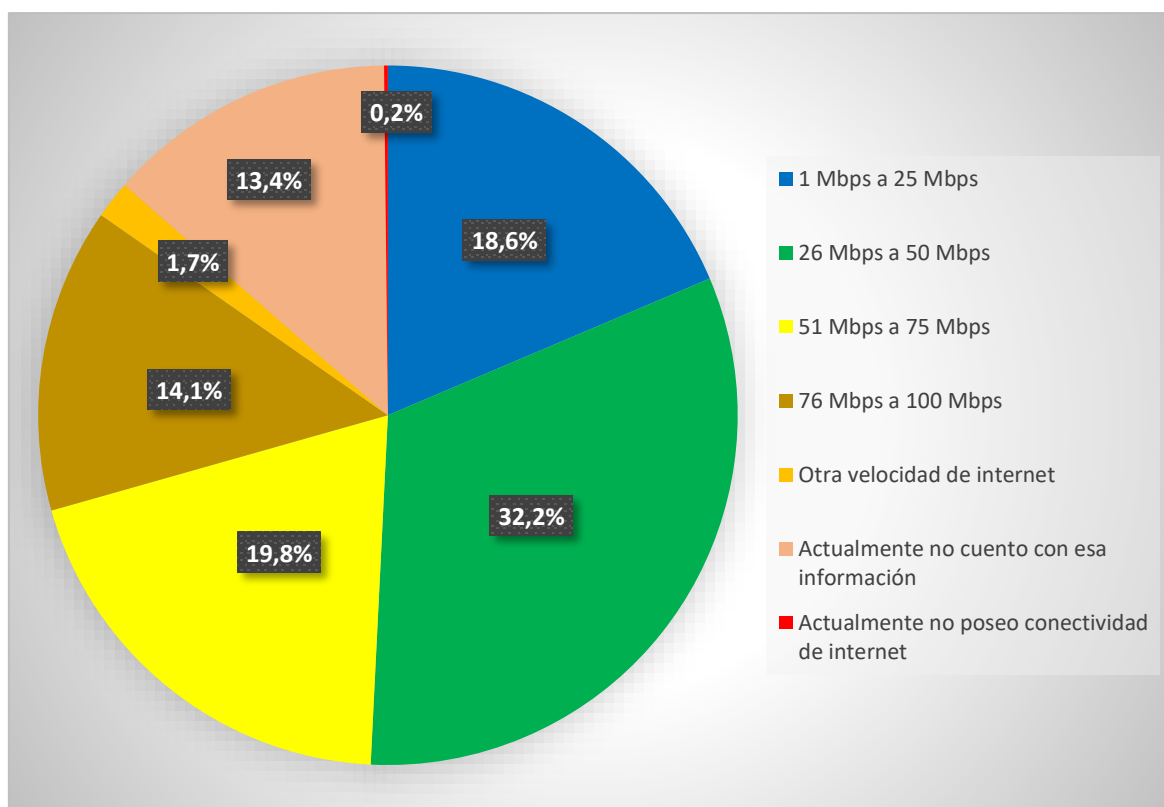


Figura 35. Ancho de banda que utiliza. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Respecto al gráfico N°35 de un total de 575 encuestados pertenecientes a la zona 8 de la provincia del Guayas, el 52% de la población posee un ancho de banda en el rango de 26Mbps a 75Mbps, el 32,7% de 1 Mbps a 25 Mbps o 76Mbps a 100Mbps, el 1,7% posee otra velocidad de internet, mientras que el 13,4% de los ciudadanos no poseen información sobre su ancho de banda y el 0,2% de la población no posee conectividad de internet.

**Pregunta N° 3.4- ¿Cuántas personas están conectadas al mismo plan de internet que usted utiliza?**

**Tabla 33** Cantidad de personas conectados al mismo plan de internet

Opciones de respuestas	Encuestados	Porcentaje
1 persona se conecta al mismo plan de internet	50	9%
2 personas se conectan al mismo plan de internet	48	8%
3 personas se conectan al mismo plan de internet	117	20%
4 personas se conectan al mismo plan de internet	176	31%
5 o más personas se conectan al mismo plan de internet	179	31%
No cuento con ningún acceso de internet	5	1%
<b>Total</b>	<b>575</b>	<b>100%</b>

Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys

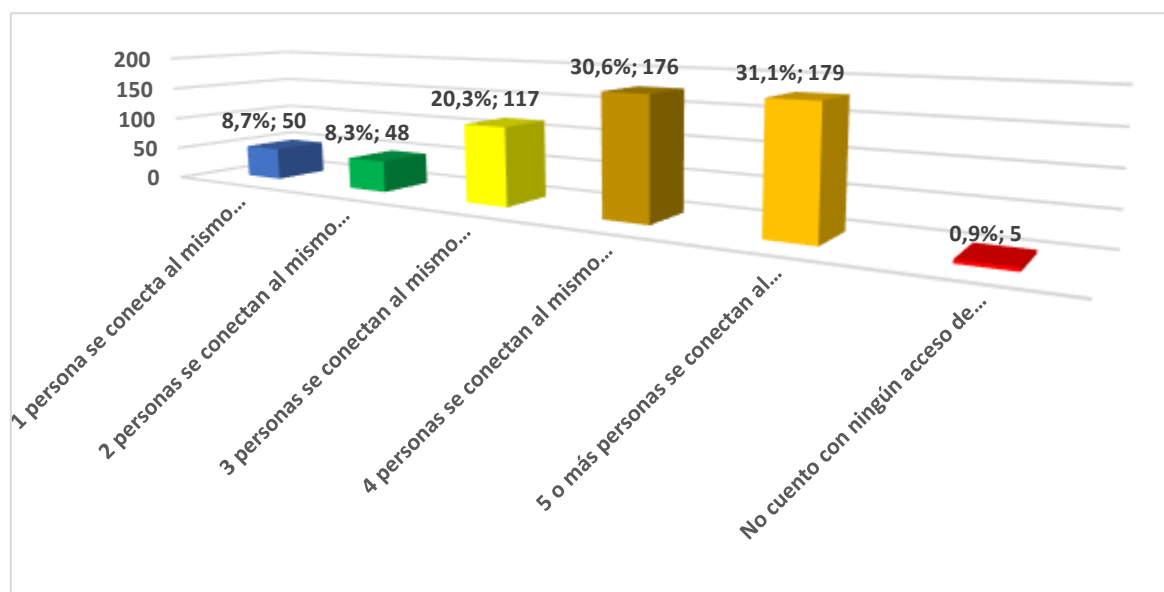


Figura 36. Cantidad de personas conectados al mismo plan de internet. Información obtenida de formulario de Google y de la investigación directa. Elaborado por Daniela Taynys.

Conforme al gráfico N°36 establece que de la muestra  $n=575$ , el 17% de la población se conectan 1 o 2 personas al mismo plan de internet, por otro lado, el 50,9% indican que se enlazan entre 3 o 4 personas, el 31,1% tienen conectividad 5 o más personas al mismo plan de internet y el 0,9% de los ciudadanos indican que no cuentan con ningún acceso a internet.

### 3.8. Resumen de la entrevista

Se realizó la entrevista a 5 profesionales de la especialidad de Magíster en sistemas de Información gerencial Msg. Omar Francisco Gutiérrez Pizarro, Ingeniero en sistemas de información Ing. Johnny José Cruz Baquerizo, Ingeniero en sistemas administrativos computarizados Ing. Christian Calero, Magíster en Business Analytics Msg. Ariosto Cabrera, Magíster en ciencias Estadísticas MSc. Néstor Montaña, llevan entre 3 a 12 años trabajando en proyectos de Desarrollo Web y/o IA.

Los 3 especialistas de Desarrollo web Msg. Omar Francisco Gutiérrez Pizarro, Ing. Johnny José Cruz Baquerizo e Ing. Christian Calero describen que poseen conocimientos muy altos con respecto a desarrollos e indican que en su experiencia siempre tienen que estar al tanto de todos los avances técnicos y teóricos que se actualizan a diario para generar procesos e implementaciones, por lo que es una iniciativa muy importante la implementación de estas herramientas ya que ayudan mucho en los medios digitales sin muchas complicaciones considerando una alta importancia en las empresas, por lo que detallan que si no lo consideran pueden tener riesgos muy altos de perder ante la competencia con criterios tecnológicos. Los lenguajes de programación que recomendarían para creación de plataformas web es Python, C#, JavaScript, PHP. Por otro lado, los frameworks que

recomiendan son: Angular, Laravel, Django, Spring, Net Core; Indicando que poseen conocimiento entre medio, alto y muy alto en desarrollos con el framework web de Django, por lo que describen que el patrón MVC y MTV es parcialmente similar, pero como expertos destacan que los manejos de los llamados controladores son demasiados tediosos y explican que las ventajas que tiene Django son: Funcional, Mantenible, Escalable, Código Limpio, Trabajo de forma pragmática. Con respecto a los frameworks de CSS que consideran adecuados son: Bootstrap, Foundation, Semantic UI, Materialize, a su vez describen que las clases y las identificaciones en el CSS permiten aplicar de manera rápida modificaciones con respecto a la estructura definida y también se puede trabajar de una manera escalable separando el contenido del estilo de una página web.

Los 2 especialistas de Inteligencia Artificial Msc. Ariosto Cabrera y MSc. Néstor Montaña, indican que ha tenido oportunidad de participar en diferentes proyectos de Data Warehouse, análisis de redes sociales a través de NLP, aplicación de Machine Learning en diferentes proyectos, ejercicios de reconocimiento de imágenes y proyectos con datos tabulares de imágenes o texto, por lo que poseen altos conocimientos en IA, ML, y NLP. Por lo tanto, ambos especialistas consideran importante el uso de la IA y NLP para soluciones de problemáticas referenciadas a la pandemia y para la automatización en general, para lo cual aplica muy bien en análisis de Covid-19 y la técnica a usar dependerá del problema a resolver, a su vez describen que están parcialmente de acuerdo en aplicar el nivel pragmático en el proceso de generación de resúmenes de textos escritos en español permitiendo obtener información útil para el lector. Destaca que el módulo de ML para una arquitectura de NLP creado para resúmenes es considerable utilizar Escala multidimensional (MDS), Mapeo de características isométricas (ISOMAP), Incrustación localmente lineal (LLE), Incrustación de vecinos estocásticos distribuidos en t (t-SNE) y para el algoritmo de generación de resúmenes tipos extractivos recomiendan utilizar los basados en el peso de las palabras (LUHN) y Basado en Matriz de Similitud. Detallan que las técnicas de preprocesamiento adecuados a utilizar son: Incluir librerías, Carga de la data, Explorar la data, Imputar datos, Codificar datos, Escalar datos, Normalizar datos, Seleccionar datos y Extraer datos. Por otro lado, establecen que para el Procesamiento del algoritmo son: Conversión de texto de mayúscula a minúscula, Eliminación de palabras de parada, Eliminación de string, Cálculo de frecuencia de palabras, Peso de cada una de las palabras, Tokenización, Escalamiento de datos, Cálculo de ponderación en oraciones, Orden de oraciones. Para el entorno de desarrollo consideran adecuados utilizar Google Colab, Azure Machine Learning, Amazon Machine Learning, a su vez indican que para el método de



extracción de información de páginas noticiarias es la utilización de la Técnica de Web Scraping utilizando desarrollo de código o Raspado web usando API de datos.

### 3.9. Construcción del sitio web

Antes de comenzar con la codificación se crearán interfaces por medio de la metodología del prototipado utilizando como herramientas: Microsoft Word y Marvel, por lo que realizará creaciones rápidas con pruebas y transferencia, por medio de los Wireframe potenciando el flujo de trabajo permitiendo llevar una baja, media y alta fidelidad.

#### 3.9.1 Prototipo de baja fidelidad para ingreso del sitio web

En esta sección el usuario podrá tener acceso libremente al sitio web.

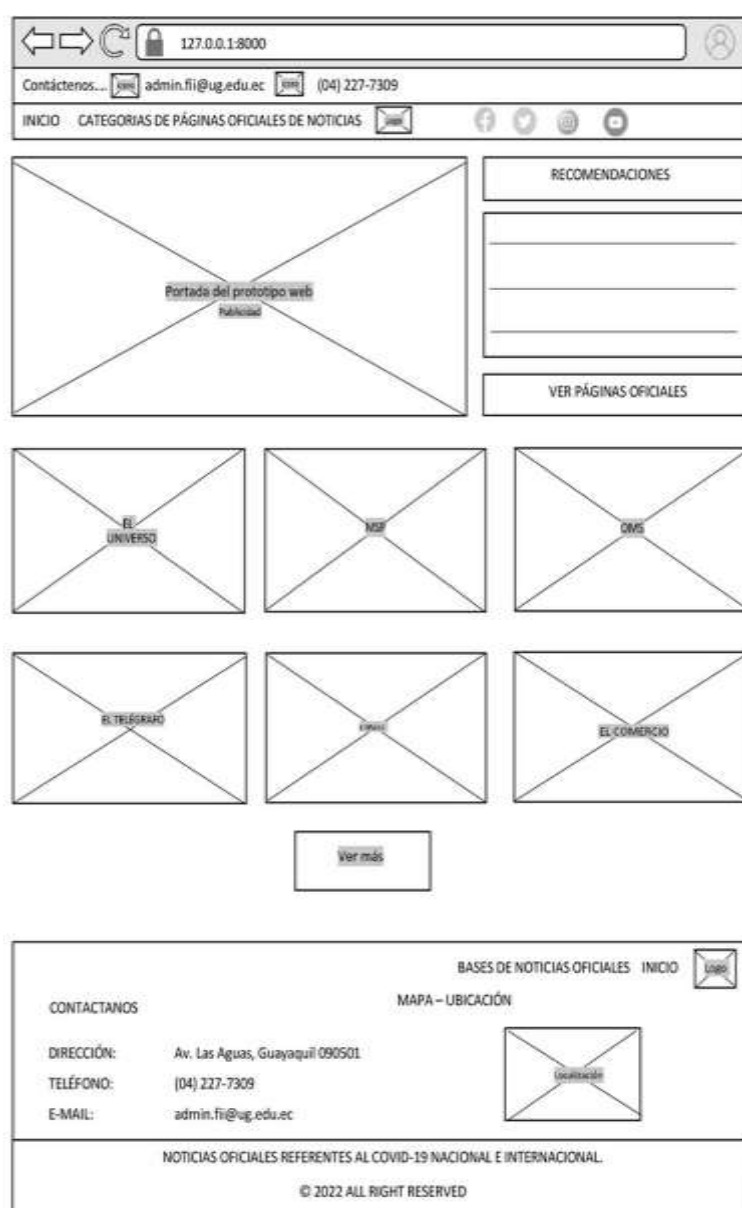


Figura 37. Prototipo de baja fidelidad ingreso al sitio web. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.2 Prototipo de baja fidelidad para acceso a categorías de páginas oficiales del sitio web

En este apartado se despliega las categorías de página oficiales.

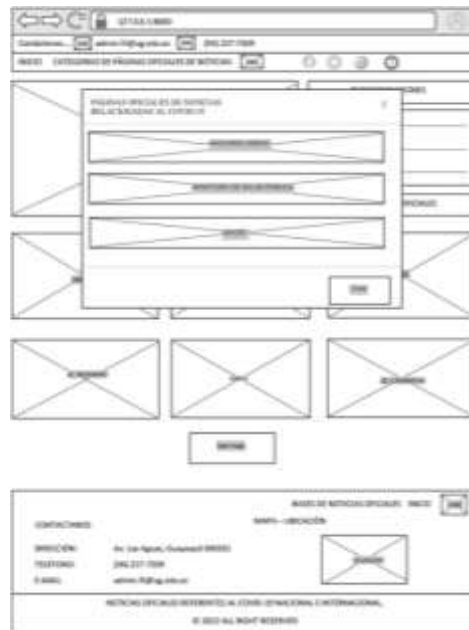


Figura 38. Prototipo de baja fidelidad despliegue de páginas oficiales. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.3 Prototipo de baja fidelidad con acceso a página oficial a artículos noticiarios.

En este apartado detalla la vista donde está implementado el texto procesado por el algoritmo NLP de resúmenes de textos adicionando fecha de publicación, actualización e imágenes importándolo al sitio web. También integra un buscador y paginador para despliegue de noticias.

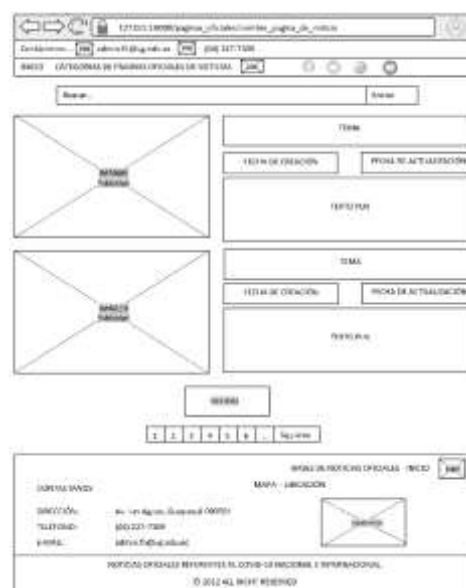


Figura 39. Prototipo de baja fidelidad con acceso a noticias oficiales referentes al covid-19. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.4 Prototipo de baja fidelidad al realizar una petición en el buscador

En este apartado se visualiza la vista que genera el ingreso de texto de interés del usuario en la barra del buscador.

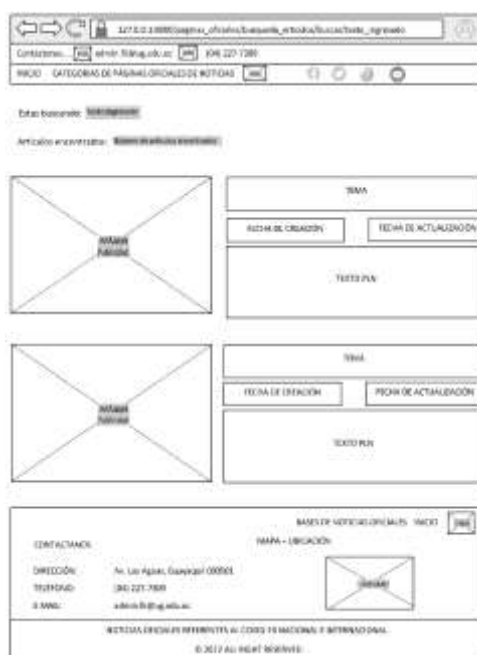


Figura 40. Prototipo de baja fidelidad con resultados de búsqueda ingresada por el usuario. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.5 Prototipo de baja fidelidad redireccionando a la página original del contenido de interés escogido por el usuario

En esta sección permite el redireccionamiento hacia la página original del contenido de interés escogido por el usuario.



Figura 41. Prototipo de baja fidelidad redireccionando a la página original del contenido de interés escogido por el usuario. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.6 Prototipo de baja fidelidad para acceso a Correo Electrónico

En esta sección permite el acceso al correo electrónico.



Figura 42. Prototipo de baja fidelidad para acceso a correo electrónico. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.7 Prototipo de baja fidelidad para acceso a Facebook

En este apartado permite el acceso a la bandeja de Facebook.

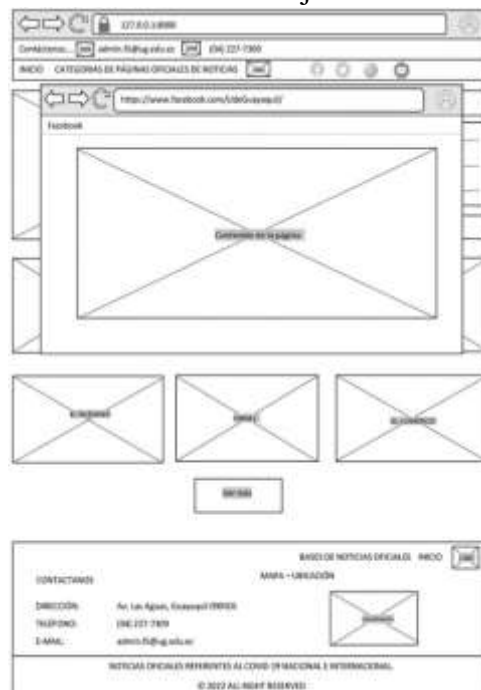


Figura 43. Prototipo de baja fidelidad para acceso a Facebook. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.8 Prototipo de baja fidelidad para acceso a Twitter

En esta sección permite el acceso a Twitter.

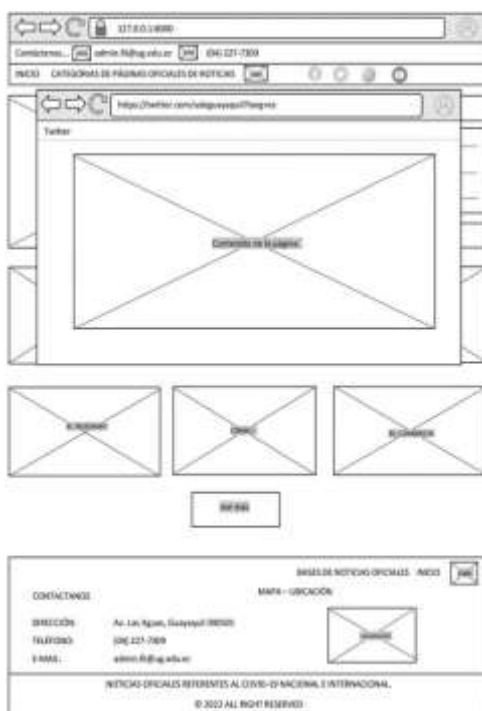


Figura 44. Prototipo de baja fidelidad para acceso a Twitter. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.9 Prototipo de baja fidelidad para acceso a Instagram

En esta sección permite el acceso a Instagram.

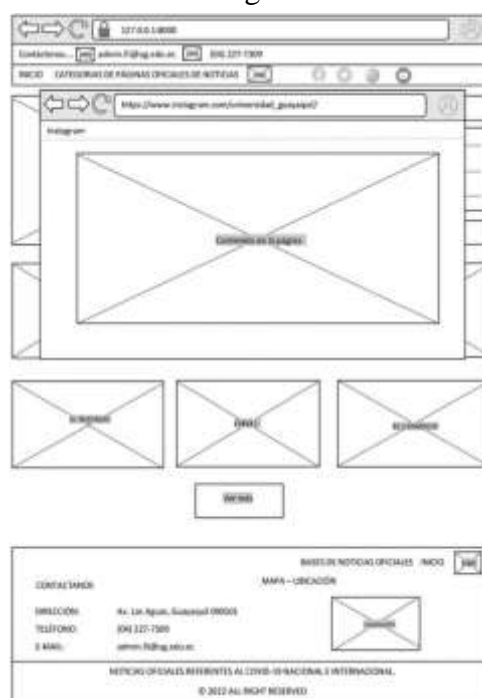


Figura 45. Prototipo de baja fidelidad para acceso a Instagram. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.10 Prototipo de baja fidelidad para acceso a Youtube

En esta sección permite el acceso a Youtube.

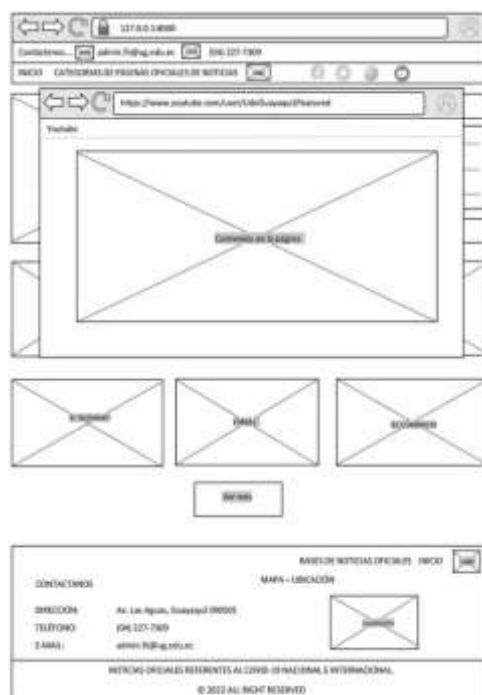


Figura 46. Prototipo de baja fidelidad para acceso a Youtube. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.11 Prototipo de baja fidelidad con acceso a la Google Maps

En este apartado se visualiza la vista que genera al dar clic en Localización.

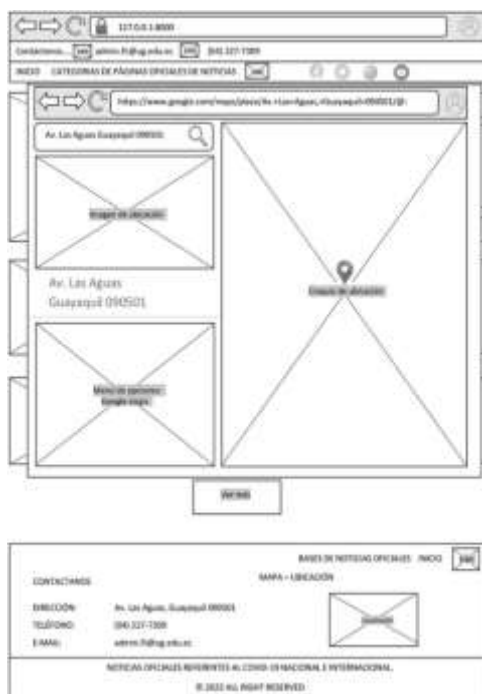


Figura 47. Prototipo de baja fidelidad con acceso a Google Maps. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.12 Modelo Arquitectónico de prototipo web de baja fidelidad y media fidelidad Diseño de la investigación

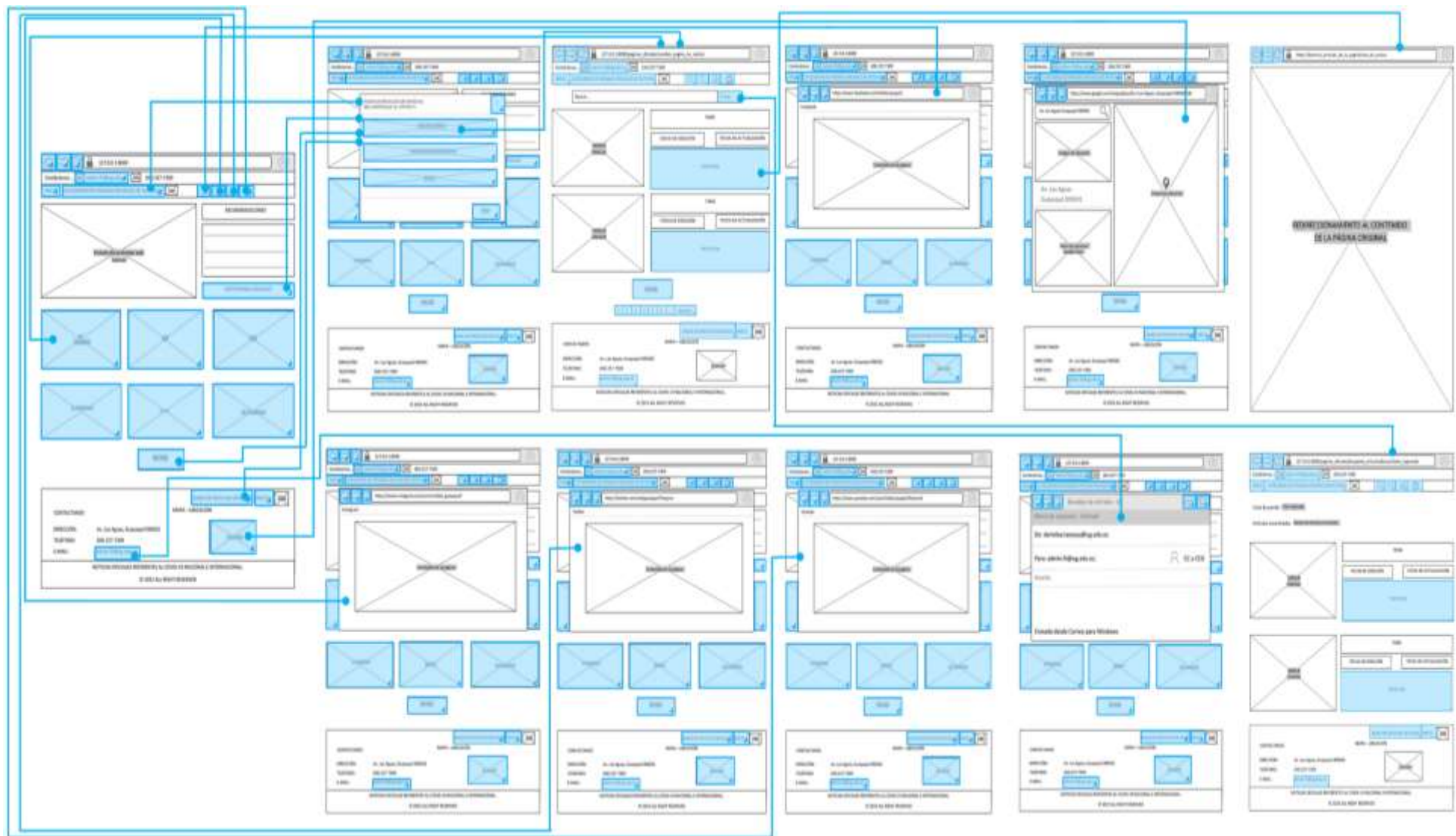


Figura 48. Modelo Arquitectónico del prototipo web de alta y media fidelidad. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.13 Prototipo de alta fidelidad para ingreso del sitio web

Se avanza a la tercera fase del prototipo web, para ello se usará herramientas tecnológicas como: Visual Studio Code, Bootstrap, Python, Django, gestor de base de datos (DB) Browser SQLite, etc. Permitiendo ejercer navegabilidad y funcionalidad. Para más detalles con respecto al manual de usuario véase Anexo 4.

En esta sección se despliega la interfaz inicial del usuario ingresando al sitio web.



Figura 49. Prototipo de alta fidelidad ingreso al sitio web. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.14 Prototipo de alta fidelidad para acceso a categorías de páginas oficiales del sitio web

En este apartado se despliega la interfaz de las categorías de página oficiales.

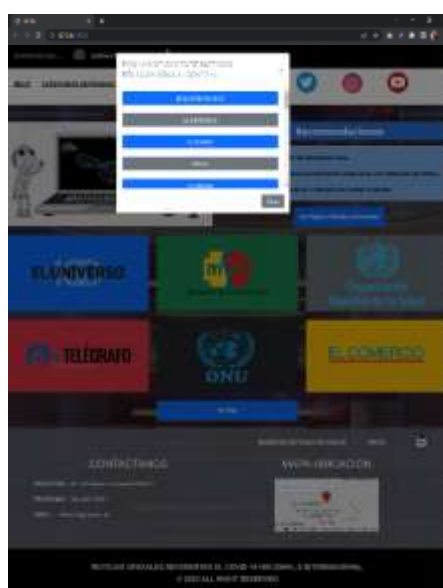


Figura 50. Prototipo de alta fidelidad despliegue de páginas oficiales. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



### 3.9.15 Prototipo de alta fidelidad con acceso a página oficial a artículos noticiarios

En este apartado detalla la interfaz donde esta implementado el texto procesado por el algoritmo NLP de resúmenes de textos adicionando fecha de publicación, actualización e imágenes extraídas usando la técnica de Web Scraping e importándolo al sitio web. También integra un buscador y paginador para despliegue de noticias.



Figura 51. Prototipo de alta fidelidad con acceso a noticias oficiales referentes al covid-19. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.16 Prototipo de alta fidelidad al realizar una petición en el buscador

En este apartado se visualiza la vista que genera el ingreso de texto de interés del usuario en la barra del buscador.



Figura 52. Prototipo de alta fidelidad con resultados de búsqueda del texto ingresado por el usuario. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.17 Prototipo de alta fidelidad redireccionando a la página original del contenido de interés escogido por el usuario

En esta sección permite la visualización del redireccionamiento hacia la página original del contenido de interés escogido por el usuario.



Figura 53. Prototipo de alta fidelidad redireccionando a la página original del contenido de interés escogido por el usuario. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.18 Prototipo de alta fidelidad para acceso a Correo Electrónico

En esta sección permite visualizar la interfaz del acceso al correo electrónico.

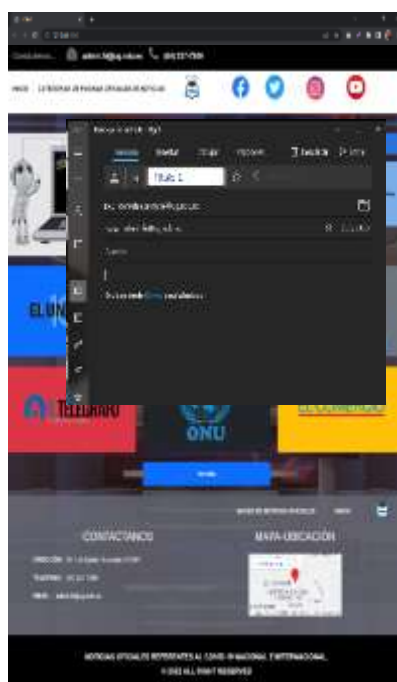


Figura 54. Prototipo de alta fidelidad para acceso a correo electrónico. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.19 Prototipo de alta fidelidad para acceso a Facebook

En este apartado permite visualizar la interfaz del acceso a Facebook.

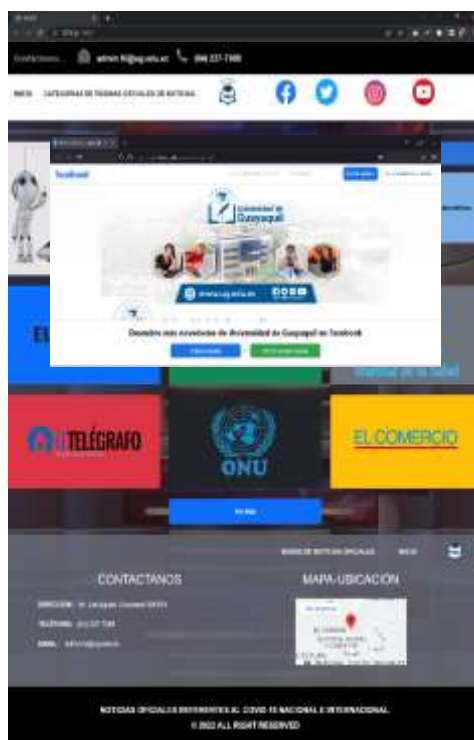


Figura 55. Prototipo de alta fidelidad para acceso a Facebook. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.20 Prototipo de alta fidelidad para acceso a Twitter

En esta sección se permite ver la interfaz de acceso a Twitter.

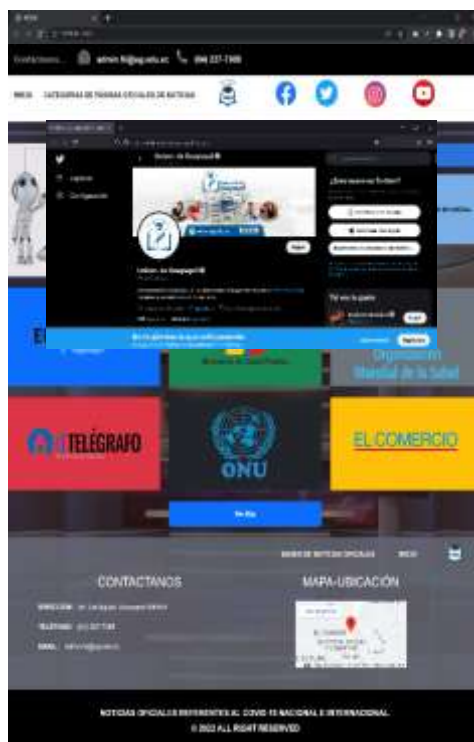


Figura 56. Prototipo de alta fidelidad para acceso a Twitter. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.21 Prototipo de alta fidelidad para acceso a Instagram

En esta sección permite visualizar la interfaz en el acceso a Instagram.

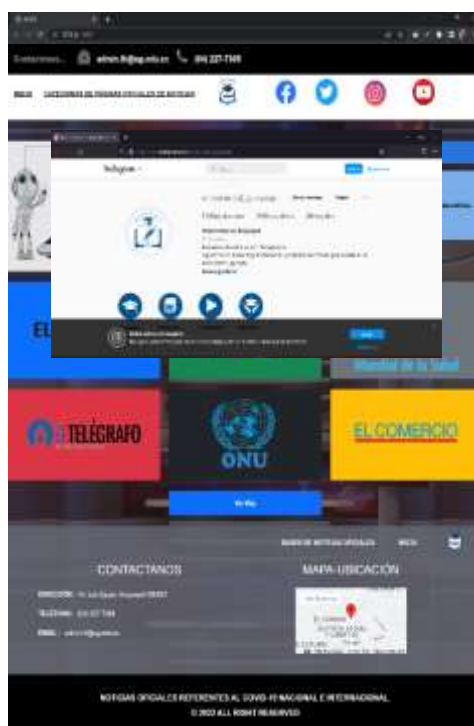


Figura 57. Prototipo de alta fidelidad para acceso a Instagram. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.22 Prototipo de alta fidelidad para acceso a Youtube

En esta sección permite visualizar la interfaz del acceso a Youtube.

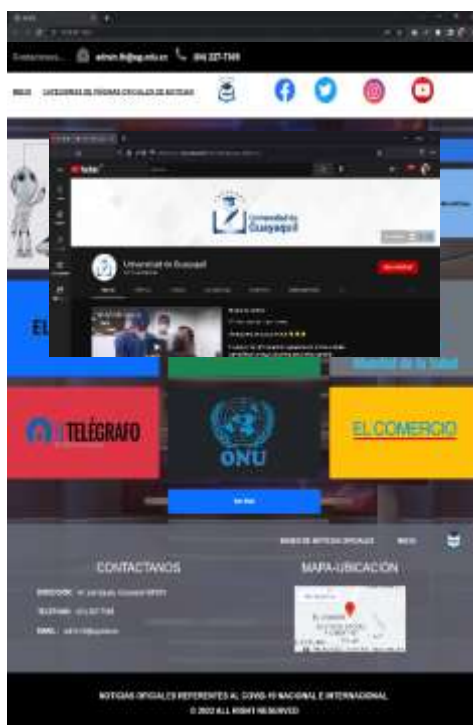


Figura 58. Prototipo de alta fidelidad para acceso a Youtube. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.9.23 Prototipo de alta fidelidad con acceso a la Google Maps

En este apartado se visualiza la vista que genera al dar clic en Localización.

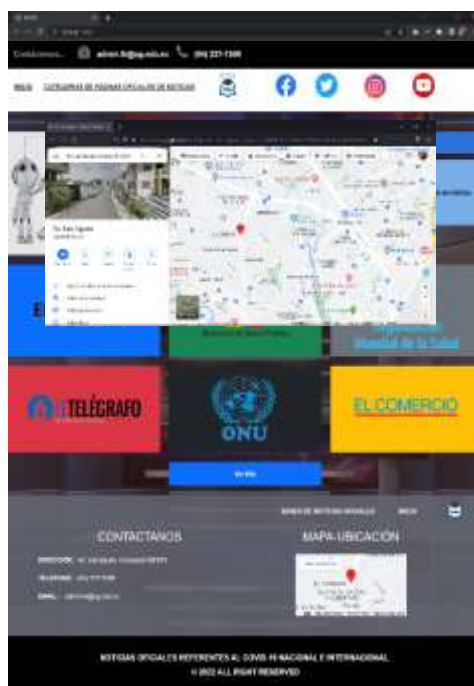


Figura 59. Prototipo de alta fidelidad con acceso a Google Maps. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10. Construcción de estructura para la ejecutar la técnica de Web Scraping

Para la extraer contenido de la web se desarrolló estructuras de códigos utilizando Visual Studio Code por lo que se realizaron 22 raspados web de noticiarios tanto nacionales e internacionales. Esta técnica se la considero para extraer contenidos de noticias y realizar el posterior entrenamiento del modelo de ML con la finalidad de reducir dimensionalidad al texto siendo considerable que estos contenidos no alteren los derechos del autor, luego de ello será importado al prototipo web. Se estableció un calendario de procesamiento de raspados para las diferentes fuentes recuperando información correspondientes a cada página noticiaría y se consideró guardar los datos en un archivo formato .json almacenándose en el ordenador.

#### 3.10.1. Importar librerías

Se realiza la importación de librerías en el archivo creado de Visual Studio Code.

```
import scrapy
from scrapy.item import Item, Field
from scrapy.spiders import CrawlSpider, Rule
from scrapy.linkextractors import LinkExtractor
from scrapy.loader import ItemLoader
from scrapy.selector import Selector
from scrapy.loader.processors import Join
from scrapy.loader.processors import MapCompose
from scrapy.crawler import CrawlerProcess
```

Figura 60. Importación de librerías a utilizar para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.2. Definir la abstracción de ítems

Se realiza la definición de los ítems considerados a realizar el raspado web.

```
class Noticia(Item):
    titulo = Field()
    contenido = Field()
    imagen = Field()
    url = Field()
    created = Field()
    updated = Field()
```

Figura 61. Definición de ítems para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.3. Definir spider

Se realiza la definición de spider para los requerimientos establecidos en la url semilla, por lo que la url será la página web de noticias oficiales de Covid-19. Hay que considerar que cada página funciona de manera diferente.

```
class EcuadorEnVivoCrawler(CrawlSpider):
    name = 'EcuadorEnVivoCrawler'
    custom_settings = {
        'USER_AGENT': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/80.0.3987.149 Safari/537.36',
        'FEED_EXPORT_FIELDS': ['titulo', 'contenido', 'imagen', 'url', 'created', 'updated'], # Ordenar las columnas del archivo
        'CLOSESPIDER_ITEMCOUNT': 2000,
        'FEED_EXPORT_ENCODING': 'utf-8'
    }
    download_delay = 1 #segundo de espera en cada interacción
```

Figura 62. Definición de spider para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.4. Definir dominios permitidos y url semilla

Toda página debe ser analizada de manera adecuada por lo que se destaca su comportamiento en las urls, verificando protocolo de internet que sea un sitio seguro tales como: dominio seguro (https://), dominio principal (ecuadorenvivo.com), ruta de acceso (busqueda/itemlist/search?searchword=covid&x=0&y=0&Itemid=405).

```
allowed_domains = ['ecuadorenvivo.com'] #dominios permitidos
start_urls = ['https://ecuadorenvivo.com/index.php/busqueda/itemlist/search?searchword=covid&x=0&y=0&Itemid=405']
```

Figura 63. Definición de dominios permitidos y url semilla para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.5. Definir reglas

Como cada página web funciona de manera diferente, es considerable analizar que tipo de información se quiere adquirir y si es preciso entrar a otros enlaces. En este caso se muestra un ejemplo de la definición de reglas para extracción Vertical y Horizontal.



```
rules = {
    Rule(
        LinkExtractor(allow="//index.php/buscqueda/itemlist/search?searchword=conlita-04y-04start=0"), follow=True, parse_items=True
        """
        por tipo de adquisición se ha creado callback ya que aquí es muy a
        extraer datos.
        """,
        Rule(LinkExtractor(allow="*/ovid/"), follow=True, callback="parse_items")
        """
        por facilidad de paginación en cada tipo se ha creado callback ya que aquí
        es a extraer datos.
        """,
    )
}
```

*Figura 64. Definición de reglas para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.*

### 3.10.6. Definir la función parse

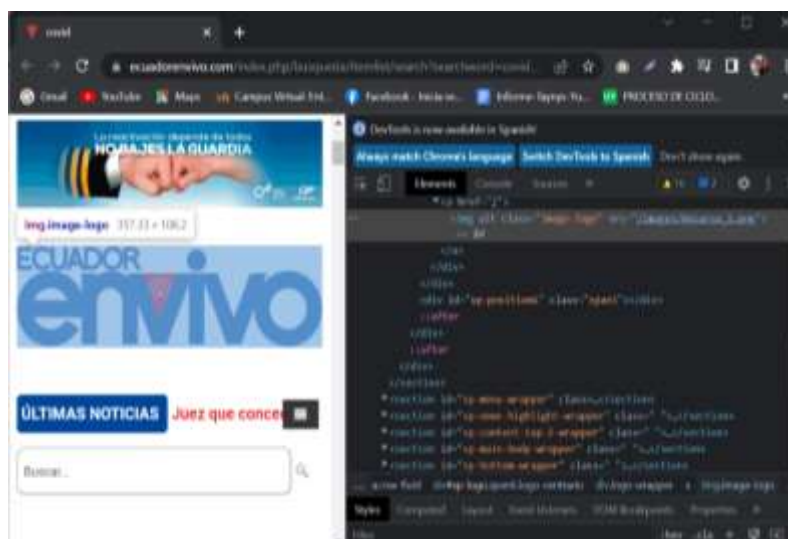
Se realiza la definición de esta función que tendrá como parámetro self porque viene de la clase y response que es la respuesta del árbol HTML. En este proceso se establece un objeto que llamara los ítems vacíos y el árbol HTML.

```
def parse_items(self, response):
    item = scrapy.loader.ItemLoader(Noticia(), response)
```

*Figura 65. Definición de función parse para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.*

### 3.10.7. Analizar el árbol HTML

En esta sección se visualiza un ejemplo de análisis para árbol HTML, es muy considerado este paso ya que permite visualizar fragmento que se utilizara para hacer el raspado web.



*Figura 66. Análisis de árbol HTML para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.*

### 3.10.8. Construcción lógica para descargar la información y llenado de ítems

Para este proceso se analizó anteriormente el HTML y se utilizó un lenguaje que permite la construcción de expresiones que recorrerán y procesarán un documento XML.

```
def parse_items(self, response):
    item = scrapy.loader.ItemLoader(Noticia(), response)
    item.add_xpath('titulo', '//h2[@class="itemTitle"]/text()', MapCompose(
        self.limpia_text))
    item.add_xpath('contenido', '//div[@class="itemFullText"]//p/text()',
        MapCompose(self.limpia_text))
    item.add_xpath('created', '//span[@class="sp_date_day"]/text()')
    # Recordar texto de fecha
    item.add_value('updated', '30 julio, 2022') # fecha de extracción para
    actualización de resumen
    item.add_xpath('imagen', '//div[@class="itemImageBlock"]//span[
        @class="itemImage"]//img/@src', MapCompose(self.Obtenerimg))
    item.add_value('url', response.url)
```

Figura 67. Construcción lógica para descargar la información hacia el web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.9. Retorno de los ítems llenos

Llegando a este proceso se procede a retornar los ítems llenos.

```
yield item.load_item()
```

Figura 68. Retorno de los ítems llenos para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.10. Probar el código

Por último, se realiza la comprobación de la estructura de código definida. El proceso de guardado se lo realiza con el formato .json recorriendo el Crawler.

```
process = CrawlerProcess({
    'FEED_FORMAT': 'json',
    'FEED_URI': 'ecuadorenvivo.json'
})

process.crawl(EcuadorenVivoCrawler)
process.start()
```

Figura 69. Prueba de la estructura de código para web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.10.11. Resultado utilizando la técnica de Web Scraping

Visualización de resultados una vez ejecutado el desarrollo del código.



Figura 70. Visualización de resultados utilizando la técnica de web scraping. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

## 3.11. Construcción del modelo de Machine Learning para resúmenes de textos

Para realización del siguiente estudio se trabajará con el entorno de desarrollo de Google Colab, por lo que se requiere de una dataset, la misma proporcionará información en cuanto a la realización de reducir la dimensionalidad de textos. Para ello se utilizó la técnica de



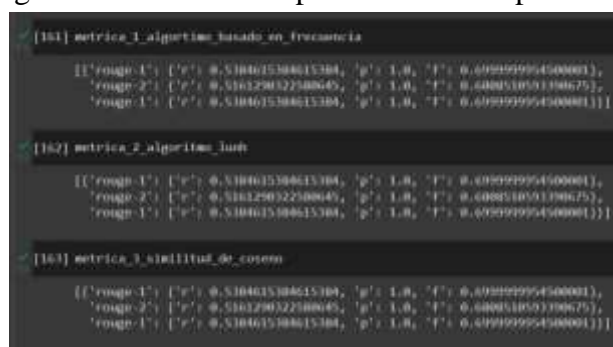
Web Scraping con la finalidad de obtener datos publicados en páginas noticiarias tanto nacionales e internacionales. Cada base de datos fue usada en el año 2022, contiene datos estructurados, el cual es un modelo de data con formato tabular, en documento JSON con extensión .json.

Inicialmente se realizó un análisis de tres algoritmos de NLP para resúmenes de textos automáticos y por medio de métricas de calidad de ROUGE comprobar el funcionamiento determinando el más adecuado a implementar en este proyecto. Este paquete de métricas es diseñado específicamente para la evaluación de resúmenes textos automáticos.

Se efectuará los resúmenes de textos midiendo la dimensionalidad inicial y final, cálculo de frecuencias de palabras para el modelo de NLP y ML, se utilizarán técnicas de preprocesamiento y luego procesamiento del algoritmo, la visualización de resultados de las mejores oraciones y prueba de validación de las métricas de ROUGE en la data resultante.

### 3.11.1. Resultados de métricas ROUGE para evaluación de los funcionamientos en los algoritmos de NLP para resúmenes de textos automáticos tipo extractivos

En este apartado se visualiza las métricas de Rouge implementados en tres algoritmos de resúmenes de textos tipo extractivo y uno de tipo abstractivo, verificando su funcionamiento del modelo de NLP de resúmenes de textos, por lo que se puede verificar que al usar cualquiera de los 3 métodos de algoritmo tipo extractivos la precisión (p) es de 100%, el recuerdo (r) está entre el 51% a 53% y la medida (f) entre 68% a 69%. Por otro lado, al verificar un algoritmo de resumen tipo abstractivo la precisión baja su calidad.



```
[151] métrica_1.algoritmo basado en frecuencia
[{'rouge-1': {'p': 0.5104615384615384, 'r': 1.0, 'f': 0.6999999994500001},
 'rouge-2': {'p': 0.5163290322388649, 'r': 1.0, 'f': 0.68851053398675},
 'rouge-1': {'p': 0.5104615384615384, 'r': 1.0, 'f': 0.6999999994500001}}]

[152] métrica_2.algoritmo_luhn
[{'rouge-1': {'p': 0.5104615384615384, 'r': 1.0, 'f': 0.6999999994500001},
 'rouge-2': {'p': 0.5163290322388649, 'r': 1.0, 'f': 0.68851053398675},
 'rouge-1': {'p': 0.5104615384615384, 'r': 1.0, 'f': 0.6999999994500001}}]

[153] métrica_3.similitud_de_coseno
[{'rouge-1': {'p': 0.5104615384615384, 'r': 1.0, 'f': 0.6999999994500001},
 'rouge-2': {'p': 0.5163290322388649, 'r': 1.0, 'f': 0.68851053398675},
 'rouge-1': {'p': 0.5104615384615384, 'r': 1.0, 'f': 0.6999999994500001}}]
```

Figura 71. Resultados de métricas Rouge a fin de evaluar los funcionamientos en los algoritmos de NLP para resúmenes de textos tipo extractivo. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



```
[94] métrica_4.resumen_abstractivo = r.get_scores(resumen,texto_original_abo)
métrica_4.resumen_abstractivo
[{'rouge-1': {'p': 0.542572881555122,
 'r': 0.9411764705882353,
 'f': 0.688728383728641},
 'rouge-2': {'p': 0.5542257142857142,
 'r': 0.923636373673211,
 'f': 0.6687584541288236},
 'rouge-1': {'p': 0.542572881555122,
 'r': 0.9411764705882353,
 'f': 0.688728383728641}}]
```

Figura 72. Resultados de métricas Rouge a fin de evaluar el algoritmo de NLP para resúmenes de textos tipo abstractivo. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

**Tabla 34** Resultados de las Métricas de ROUGE

Métricas	Medidas	Algoritmo basado en frecuencia	Algoritmo LUNH	Similitud de coseno	Algoritmo Transformers de summarization
rouge-1	r	53,846%	53,846%	53,846%	54,237%
	p	100%	100%	100%	94,12%
	f	70,00%	70,00%	70,00%	68,82%
rouge-2	r	51,61%	51,61%	51,61%	51,43%
	p	100%	100%	100%	92,31%
	f	68,09%	68,09%	68,09%	66,06%
rouge-l	r	53,85%	53,85%	53,85%	54,24%
	p	100%	100%	100%	94,12%
	f	70,00%	70,00%	70,00%	68,82%

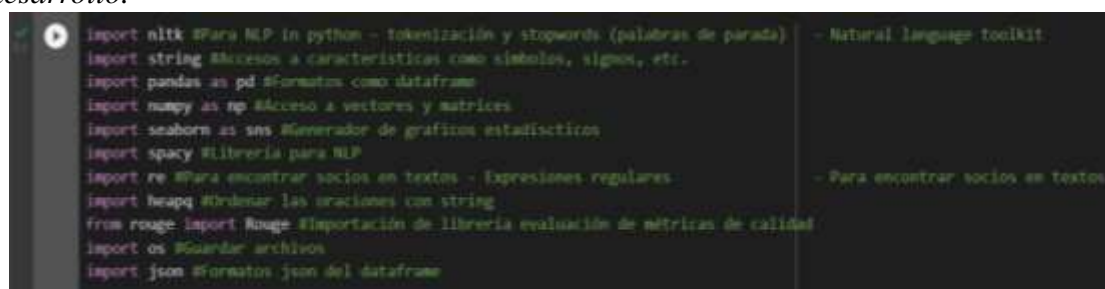
Información obtenida de la investigación directa. Elaborado por Daniela Taynys.

Se puede visualizar por medio de la tabla N°34 se logra obtener resultados óptimos para su utilización en generar sinopsis referentes al covid-19, por tal motivo se consideró que es adecuado usar el Algoritmo de tipo extractivo LUNH ya que por medio de este método se pueden realizar previas configuraciones en cuanto al ingreso de la cantidad de n-gramas, oraciones y porcentaje que se deseen.

### 3.11.2. Técnicas de preprocesamiento

#### 3.11.2.1. Instalación de librerías

Inicialmente se debe cargar e instalar las librerías necesarias y dejar listo el entorno de desarrollo.



```

import nltk #Para NLP in python - tokenización y stopwords (palabras de parada) - Natural language toolkit
import string #Acceso a características como símbolos, signos, etc.
import pandas as pd #Formatos como dataframe
import numpy as np #Acceso a vectores y matrices
import seaborn as sns #Generador de gráficos estadísticos
import spacy #Librería para NLP
import re #Para encontrar socios en textos - Expresiones regulares - Para encontrar socios en textos
import heapq #Ordenar las oraciones con string
from rouge import Rouge #Importación de librería evaluación de métricas de calidad
import os #Guardar archivos
import json #Formatos json del dataframe

```

Figura 73. Instalación de librerías para el algoritmo NLP aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

#### 3.11.2.2. Cargar la data

Para cargar la data primordialmente se debe importar la dataset desde la carpeta del proyecto de NLP donde se aplicó la técnica de Web Scraping, ubicado en el ordenador hacia el entorno IDE de Google Colab.

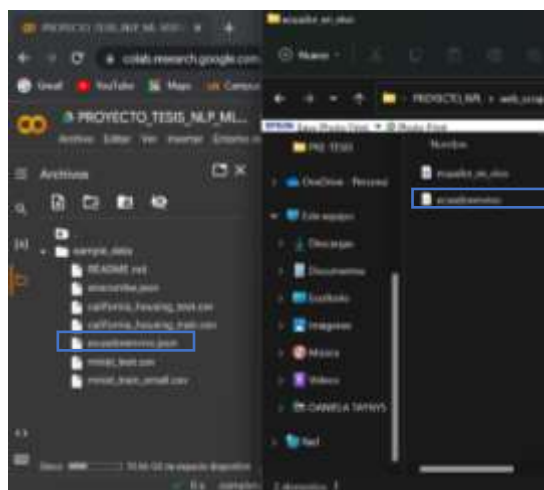


Figura 74. Importación de la data a Google Colab aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Luego de ello se procede a realizar la carga de la dataset haciendo el llamado de la librería pandas donde inicialmente fue renombrada como pd, luego se crea un objeto que va a tener el método DataFrame(df) siendo una estructura de datos, similar a una hoja de cálculo por lo que contiene dos dimensiones denominado como filas y columnas.

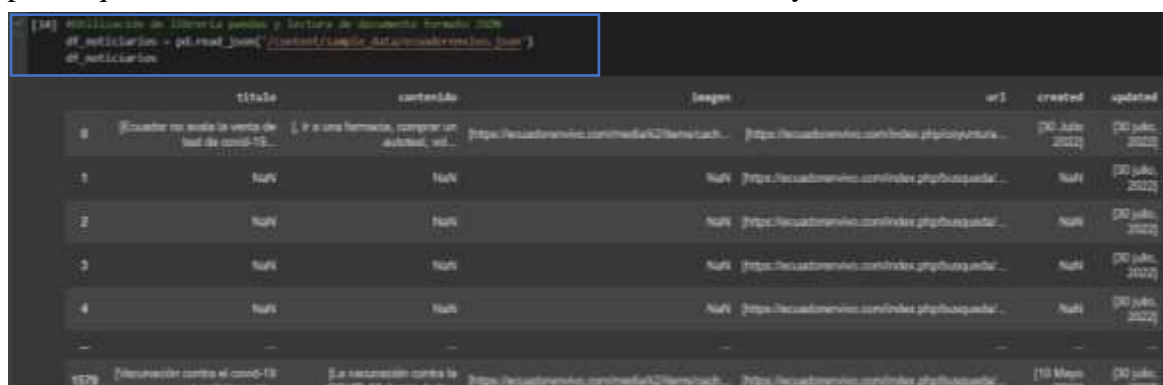


Figura 75. Carga de la data aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.2.3. Explorar la data

Luego se explora la data verificando su estructura interna, detallando sus tipos y la contabilización de filas que incluye cada columna, .head permite visualizar la cabecera de la data mostrando las primeras cinco filas del df, .tail se muestra las últimas cinco filas y con .shape[:] devuelve una tupla representando la dimensionalidad.

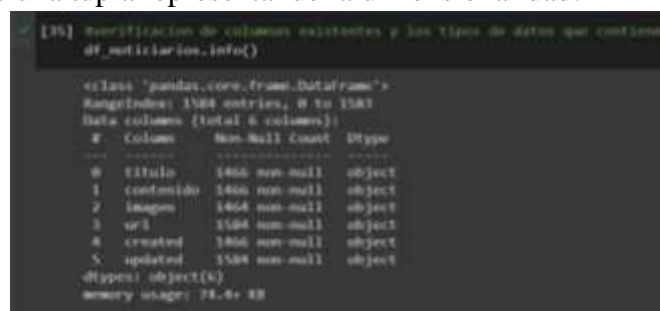


Figura 76. Explorar la data aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.2.4. Imputar la data

Una vez explorada la data se procede a verificar cuantas filas tienen valores perdidos por cada columna.

```
[41] #Numero total de NaN en la Dataset.
df_noticiarios.isnull().sum()

titulo      118
contenido   118
imagen      120
url          0
created     118
updated      0
dtype: int64
```

Figura 77. Consulta de valores perdidos aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

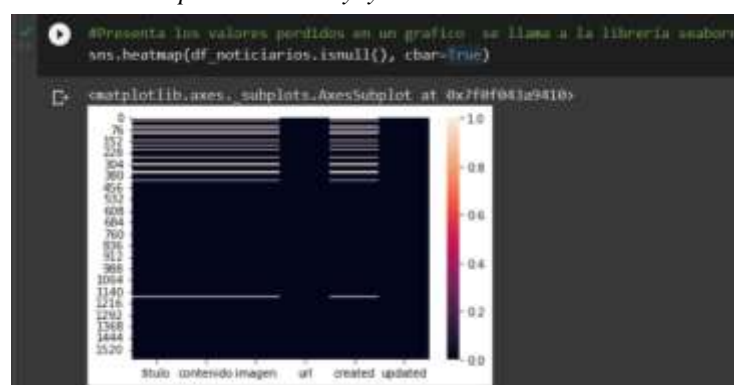


Figura 78. Gráfico de valores perdidos aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Se realiza el retiro de filas con valores nulos, en otras ocasiones se puede reemplazar datos para no perder información valiosa.

```
#Retiro las filas que esten vacias
df_noticiarios_imp_primerio = df_noticiarios.loc[df_noticiarios['titulo'].notnull()]
```

Figura 79. Imputación de la data aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Se verifica si no hay valores nulos en la data, esto se lo realiza con la finalidad de importarlo al prototipo web teniendo la misma dimensionalidad de filas y columnas.

```
[46] #Verificación de filas retiradas
df_noticiarios_imp.isnull().sum()

titulo      0
contenido   0
imagen      0
url          0
created     0
updated     0
dtype: int64
```

Figura 80. Verificación de valores perdidos aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.2.5. Normalizar los datos

Se procede con la normalización de los datos para ello se va a escoger la columna en la que se va a realizar el procesamiento por lo que recorre cada fila del df y lo convierte sin formato de lista.

```
[45] df_noticiarios_imp['contenido_sin_formato_lista_inicio'] = df_noticiarios_imp.apply(lambda x: ' '.join(x['contenido']), axis=1)

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is being set on a copy of a slice from a DataFrame.
Try using .loc[row_index,col_index] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/10min/10min#setting-a-slice-with-a-copy
***Try print for launching an IPython kernel.
```

Figura 81. Normalizar la data aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.2.6. Seleccionar los datos

Luego se realiza la selección de los datos visualizando inicialmente el contenido de datos crudos y el contenido preprocesado.

	contenido	contenido_sin_formato_lista_inicio
0	[, Ir a una farmacia, comprar un autotest, vol...	Ir a una farmacia, comprar un autotest, vol...
11	[, Por COVID prolongado, long COVID o síndrome...	Por COVID prolongado, long COVID o síndrome p...
12	[, El segundo mandatario, Dr. Alfredo Borrero ...	El segundo mandatario, Dr. Alfredo Borrero Ve...

Figura 82. Visualización del contenido crudo y procesado de la data aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Se procede con la selección de los datos a utilizar para el procesamiento, por lo que esta columna es del contenido obtenido una vez preprocesado.

	contenido_sin_formato_lista_inicio
0	Ir a una farmacia, comprar un autotest, vol...
11	Por COVID prolongado, long COVID o síndrome p...
12	El segundo mandatario, Dr. Alfredo Borrero Ve...

Figura 83. Seleccionar columna de la data aplicando técnicas de preprocesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3. Técnicas de Procesamiento

#### 3.11.3.1. Definición de función para preprocesamiento de la data

Como primer paso para aplicación del procesamiento en el algoritmo de NLP y ML para resúmenes de textos automáticos tipo extractivo se procede a definir una función denominado como preprocess, debido a que en la aplicación del algoritmo hay un preprocesamiento que pasará como parámetro un texto, el mismo será el contenido preprocesado que se realizó en pasos anteriores. Dentro de la función incluirá algunas técnicas comunes de preprocesamiento como: el cambio de palabras de mayúsculas a minúscula, la tokenización (separará el texto en palabras) agregándolas a una lista y por último crear condición que en el listado de la tokenización excluya palabras de stopwords (palabras de parada) y todo tipo de string (cadena de caracteres). Se crea un objeto donde recorrerá elemento por elemento de la lista de tokenización y los convertirá en un solo texto y por último retornar el texto con todos los procesos vistos aplicados.



```
[58] def preprocess(text): #Recibe como parámetro el texto ingresado
    formatted_text = text.lower() #conversión a minúscula
    #formatted_text = text #Texto
    #Se crea una variable con el formato de una lista
    tokens = []
    #Implementación de un bucle que pasa por cada una de las palabras tokenizadas
    for token in nltk.word_tokenize(formatted_text):
        tokens.append(token) #Aplicación de la tokenización y agregar a la lista
    #print(tokens)
    #para definir un bucle para palabras en conversaciones" el tokens es para cada vez que se ejecuta
    # "not in stopwords" selecciona las palabras que no están dentro de estas listas
    tokens = [word for word in tokens if word not in stopwords and word not in string.punctuation]
    #el texto es igual y unir todos los aceros con un espacio vacío e implementar 4 elementos
    #tokens = [word for word in tokens] #NO Eliminar las palabras de paradas usando "if word not in st
    #el texto es igual y unir todos los aceros con un espacio vacío e implementar 4 elementos
    formatted_text = ' '.join(element for element in tokens)

    #return tokens #aquí se ve el listado de palabras tokenizadas
    return formatted_text #Devuelve el texto preprocesado o formateado, extrarto del texto en minúscula
```

Figura 84. Definición de función para el algoritmo aplicando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.2. Aplicación de función para el preprocesamiento de la data

Se visualiza la creación de un nuevo objeto creando una columna adicional al df denominado como contenido\_preprocess, por lo que dentro del objeto creara el método de aplicación de la función preprocess a la columna llamada contenido\_sin\_formato\_lista\_inicio.

```
x_inicio['contenido_preprocess'] = x_inicio.apply(lambda x: preprocess(x['contenido_sin_formato_lista_inicio']), axis=1)
x_inicio['contenido_preprocess']

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead
```

Figura 85. Aplicación de función para el preprocesamiento del algoritmo empleando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.3. Definición de función para el cálculo de frecuencia de palabras en la data

Se procede a calcular la frecuencia de palabras del df seleccionado para el procesamiento y por cada una de las filas que conforman la columna a utilizar la aplicabilidad de esta función, se hará la representación de las frecuencias de palabras en gráficos estadísticos.

```
def frecuencia_palabras_grafico(text): #Recibe como parámetro el texto ingresado
    formatted_text = text.lower() #conversión a minúscula
    #formatted_text = text #Texto
    #Se crea una variable con el formato de una lista
    tokens = []
    #Implementación de un bucle que pasa por cada una de las palabras tokenizadas
    for token in nltk.word_tokenize(formatted_text):
        tokens.append(token) #Aplicación de la tokenización y agregar a la lista
    #print(tokens)
    #para definir un bucle para palabras en conversaciones" el tokens es para cada vez que se ejecuta
    # "not in stopwords" selecciona las palabras que no están dentro de estas listas
    tokens = [word for word in tokens if word not in stopwords and word not in string.punctuation]
    #el texto es igual y unir todos los aceros con un espacio vacío e implementar 4 elementos
    #tokens = [word for word in tokens] #NO Eliminar las palabras de paradas usando "if word not in st
    #el texto es igual y unir todos los aceros con un espacio vacío e implementar 4 elementos
    formatted_text = ' '.join(element for element in tokens)

    freq = nltk.FreqDist(tokens)
    for key, val in freq.items():
        #print(str(key) + ':' + str(val))
    #m.set()
    freq.plot(10, cumulative=True) #Gráfico estadístico de frecuencia de palabras

    return freq #aquí se muestra un diccionario de cada fila con su valor y frecuencia
```

Figura 86. Definición de función para cálculo de frecuencia de palabras del algoritmo empleando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.4. Aplicación de función para el cálculo de frecuencia de palabras en la data

Una vez aplicado esta función se obtendrá resultados gráficos estadísticos de cada fila que conforman la columna denominada contenido\_sin\_formato\_lista\_inicio.

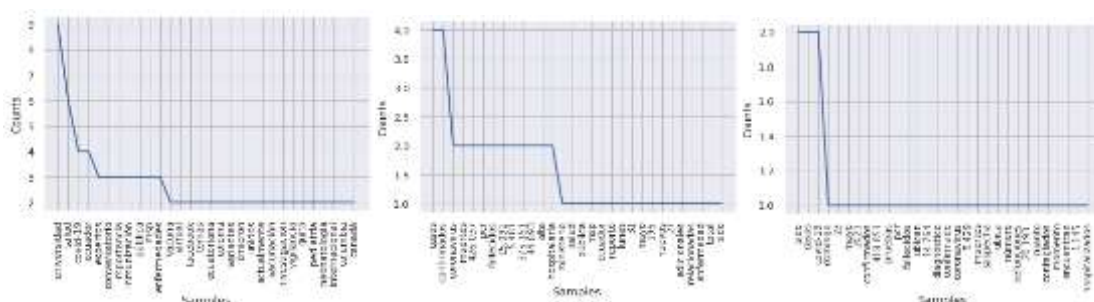


Figura 87. Resultados de gráficos estadísticos aplicando función para cálculo de frecuencia de palabras del algoritmo empleando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.5. Definición de función para ponderación de palabras frecuentes en la data

Se procede a realizar la ponderación de oraciones la misma pasa como parámetros: las oraciones, las palabras importantes y la distancia. Como retorno de la función se va a obtener las puntuaciones de las oraciones tokenizadas determinando las palabras importantes y haciendo una suma de palabras por palabras de manera que se obtenga una ponderación de las oraciones. Se considera palabras importantes las que tienen mayor frecuencia dentro de la data analizando cada una de las filas. Una vez realizado la definición de función se procede a detallar el funcionamiento del mismo:

- `scores = []`: se establece un objeto que tendrá un método con formato de lista vacía para alojamiento de puntuación hacia cada una de las oraciones.
- `sentece_index = 0`: se crea un objeto que tendrá una inicialización de cero.
- `for sentece in [nltk.word_tokenize(sentece) for sentece in senteces]`: recorre la lista de oraciones formateadas tokenizando las palabras.
- `word_index = []`: creación de objeto con método de lista para índice de palabras con mayor frecuencia.
- `for word in important_words`: recorriendo un bucle según el número de palabras ingresadas siendo consideradas como importantes
- `try: word_index.append(sentence.index(word))`: bloque de sentencia para agregar a las oraciones formateadas y considere el índice de palabras importantes hacia la lista vacía de `word_index`.
- `except ValueError: word_index.sort()`: si no encuentra el índice de palabras importantes en cada oración, puede continuar con el algoritmo, a su vez realiza el ordenamiento del índice de palabras de la lista.

- `if len(word_index) == 0`: continue: condicional si el valor de número de palabras es igual a 0 continua el algoritmo.
- `groups_list = []`: Creación de objeto agregando un método de formato lista vacía para grupos de palabras por oración.
- `group = [word_index[0]]`: objeto que aloja el índice de palabras.
- `i=1` método de incremento empezando desde 1
- `while i < len(word_index)`: ciclo para el índice de palabras con el objeto de incremento.
- `if word_index[i] - word_index[i - 1] < distance`: `group.append(word_index[i])`: condicional de palabras determinando si el índice es menor que distancia se agregara al objeto group.
- `else`: `groups_list.append(group[:])` `group = [word_index[i]]` `i += 1`: caso contrario accede a todos los elementos y los agrega al objeto de `groups_list`, por último el objeto group va a ser incremental.
- `groups_list.append(group)`: añade a la lista `group_list` las resultantes de ciclo while.
- `max_group_score = 0`: creación de un objeto inicializando en cero
- `for g in groups_list`: recorre lista por separado.
- `important_words_in_group = len(g)`: recuento de palabras importantes del recorrido de la lista por separado.
- `total_words_in_group = g[-1] - g[0] + 1`: objeto creado para el realizar el método del total de palabras importantes y el número total de palabras de posición últimas y primeras.
- `score = 1.0 * important_words_in_group**2 / total_words_in_group`: creación de objeto con el método para el cálculo de puntuación multiplicación de  $1.0 * \text{raíz cuadrada del recuento de palabras importantes dividido para el total de palabras en el grupo}$ .
- `if score > max_group_score`: `max_group_score = score`: condicional para obtener una mejor puntuación (si la puntuación es mayor que la puntuación máxima del grupo), entonces el metodo `max_group_score` será igual a `score`.
- `scores.append((max_group_score, sentence_index))`: añade las máximas puntuaciones y el número de índice de la oración.
- `sentence_index += 1`: incremento del objeto.
- `return scores`: retorno de función de las puntuaciones resultantes.



```

def ponderacion_oraciones_luhn(sentences, important_words, distance): #cálculo de oraciones cuadradas recibe como parámetro(oraciones, palabras impor
scores = [] #Creación de lista vacía para alojamiento de puntuación para cada una de las oraciones
sentence_index = 0 #Creación de un índice para cada una de las frases empezando desde 0

for sentence in [nltk.word_tokenize(sentence) for sentence in sentences]: #Bucle recorriendo una lista de oraciones
    #print("-----") #Prueba de palabras tokenizadas por oraciones
    #print(sentence) #Direcciónar para prueba a la función del paso 1

    word_index = [] #Creación de lista para índice de palabras
    for word in important_words: #Bucle para las palabras importantes según la cantidad de palabras ingresadas
        #print(word) # Se tiene cada palabra de las palabras importantes en cada oración
        try:
            word_index.append(sentence.index(word))
        except ValueError:
            pass #Si no encuentra oraciones pasará a la siguiente fase para que no se salga el mensaje de error
    #Dirigirse a -- summarize(original_text, 5) #saldrá como resultado las palabras importantes y casos de que ninguna palabra se haya tomado en c
    #Luego ir a Paso 3 e implementar el parámetro de distancia
    word_index.sort() #Ordenamiento de índice de palabras en lista
    #print(word_index)

    if len(word_index) == 0: #Condición si el índice de palabras es igual a cero
        continue #Podrá continuar el algoritmo

    groups_list = [] #Creación de lista vacía
    group = [word_index[0]] #variable que aloja el índice de palabras importantes
    i = 1 #variable que sea igual a 1 para que se inicialice el bucle siguiente
    while i < len(word_index): # - ciclo while para el índice de palabras controlandola
        # primera ejecución: 1 - 0 = 1
        # segunda ejecución: 2 - 1 = 1
        if word_index[i] - word_index[i - 1] < distance: #Palabras accedidas con índice de posición menor que distancia - cálculo de distancia entre pa
            group.append(word_index[i]) #añadiendo a grupos con el índice de posición
            #print('group', group) #valor de índice de las palabras importantes con mejores valoraciones
        else:
            groups_list.append(group[:]) #añadiendo a todos los elementos del grupo
            group = [word_index[i]] #nuevo número lista
            #print('group', group)
            i += 1
    groups_list.append(group) #añadiendo resultados a la lista de grupos
    #print('all groups', groups_list) #índice de las posiciones de palabras importantes a considerar en la oración y valor total objetos de la lista

    max_group_score = 0 #creación de una variable de máxima puntuación inicializada por cero
    for g in groups_list: #recorrer los grupos por separado
        #print(g)
        important_words_in_group = len(g) #cuántas palabras se tienen en el grupo
        total_words_in_group = g[-1] - g[0] + 1 #creación de variable para contar las palabras importantes y el número total de palabras últimas y prim
        score = 1.0 * important_words_in_group**2 / total_words_in_group #Cálculo de puntuación
        #print('group score', score) #puntuación visualizando solo las más alta

        if score > max_group_score: #Condición if para obtener una mejor puntuación (si la puntuación es mayor que la puntuación máxima del grupo)
            max_group_score = score #ese valor será asignado a la variable inicial dada

    scores.append((max_group_score, sentence_index)) #añadiendo en las puntuaciones la máxima puntuación y el número de índice de la oración
    sentence_index += 1 #Incremento de la primera variable que se está ejecutando «1»

    #print('final scores', scores) #valida igual a la puntuación por el número de la oración.***
    return scores

```

Figura 88. Definición de función para ponderación de oraciones en la data aplicando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.6. Definición de función para resúmenes de textos de la data

Una vez realizado la aplicabilidad de funciones se procede a definir una nueva función que permita realizar los resúmenes de textos tipo extractivos utilizando NLP Y ML, dicho esto se procede a la explicación del funcionamiento del algoritmo:

- original\_sentences: recorre las oraciones originales y las tokeniza.

- `formatted_senteces`: realiza el formateado de las oraciones pasando la función del preprocesamiento a las oraciones originales tokenizadas.
- `words`: adquiere las palabras del texto formateado representado en una lista.
- `frequency`: cálculo de frecuencia de las palabras tokenizadas del método anterior.
- `top_n_words`: recuentos del cálculo de frecuencia de palabras desde el más común hasta el menos común.
- `ponderacion_oraciones_lunh`: realiza el llamado de la función antes detallada y pasa como parámetro el formateado de oraciones, los recuentos de los cálculos de frecuencias de palabras y la distancia que tomara entre oraciones.
- `senteces_score_lunh`: se realiza la creación de un método para obtener la `ponderacion_oraciones_lunh`
- `if porcentaje > 0`: explica que se crea una condicional realizando la pregunta “si el porcentaje es mayor a 0”, entonces se crea un método que tendrá el llamado de una librería `heapq.nlargest` la misma devolverá la lista de `n` elementos más grandes midiendo la longitud de las oraciones formateadas, el porcentaje y la ponderación de oraciones.
- `else`: caso contrario realiza el ordenamiento de oraciones importantes dependiendo del parámetro ingresado y la lista obteniendo las puntuaciones de cada una de las oraciones.
- `best_senteces_lunh`: método que obtendrá las mejores oraciones.

```
def summarize_lunh(text, top_n_words, distance, number_of_sentences, percentage = 0):
    """
    @param text: texto a resumir
    @param top_n_words: número de palabras más importantes del texto
    @param distance: distancia entre oraciones
    @param number_of_sentences: número de oraciones a incluir en el resumen
    @param percentage: porcentaje de palabras a incluir en el resumen
    """
    original_sentences = [sentence for sentence in nltk.sent_tokenize(text)]
    formatted_sentences = [preprocess(sentence) for sentence in original_sentences]
    words = [word for sentence in formatted_sentences for word in nltk.word_tokenize(sentence)]
    frequency = nltk.FreqDist(words)
    top_n_words = [word for word in frequency.most_common(top_n_words)]
    ponderacion_oraciones_lunh(formatted_sentences, top_n_words, distance)
    senteces_score_lunh = ponderacion_oraciones_lunh(formatted_sentences, top_n_words, distance)
    if percentage > 0:
        best_senteces_lunh = heapq.nlargest(number_of_sentences, senteces_score_lunh)
    else:
        best_senteces_lunh = heapq.nlargest(number_of_sentences, senteces_score_lunh)
    return [original_sentences[i] for (score, i) in best_senteces_lunh]
```

Figura 89. Definición de función para resúmenes de textos en la data aplicando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys

Se observa resultados de funciones aplicadas en pasos anteriores, de tal forma que se estime la diferencia de cada una de las columnas.

	contenido_sin_formato_lista_inicio	contenido_preprocess	contenido_frecuencia_grafico	contenido_con_formato_lista_final
0	Ir a una farmacia, comprar un autotest, volar...	e farmacia comprar autotest volver casa aplo...	({"f": 2, "farmacia": 1, "comprar": 1, "autite...	(Son test de alta sensibilidad, superior al 95...
11	Por COVID prolongado, long COVID o síndrome...	covid prolongado long covid síndrome post covid...	(covid: 10, "prolongado": 3, "long": 2, "sín...	(Puede durar una media de dos semanas y como...
12	El segundo mandatario, Dr. Alfredo Romero Ve...	segundo mandatario dr. alfredo romero vega...	(segundo: 1, "mandatario": 1, "dr": 1, "alf...	(Esta es una de las resoluciones anunciadas po...
13	Investigadores de los Institutos Gladstone, U...	investigadores institutos gladstone uc berkele...	(investigadores: 3, "institutos": 1, "gladst...	(Por ejemplo, si una partícula similar a un ví...
14	"Éssoy muy agradecido con el personal que ha ...	"agradecido personal venido voluntarios demue...	({"": 1, "agradecido": 1, "personal": 2, "veni...	(El personal de la Dirección Distrital (DDCO) d...
...	...	...	...	...
1579	La vacunación contra la COVID-19 dentro de la ...	vacunación covid-19 dentro fase 1 plan vacun...	(vacunación: 3, "covid-19": 2, "dentro": 2, ...	(El presidente Moreno también anunció, durante...

Figura 90. Visualización del Dataframe con funciones implementadas aplicando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys

### 3.11.3.7. Resultados finales

En este apartado principalmente se realiza la contabilización de palabras de la columna en la que se seleccionó inicialmente para realizar el procesamiento y también el conteo de palabras en la columna resultante donde se implementó la función del algoritmo de NLP y ML.

```
[87] #Se crea una columna en la que se hace un conteo de palabras se tiene en la fila descrita
x_inicio["Numero_palabras_inicio"] = x_inicio.apply(lambda x: len(x['contenido_sin_formato_lista_inicio']), axis=1)

/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

Figura 91. Aplicación de conteo de palabras de una columna en la data aplicando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Se visualiza que en la columna denominada x\_num\_palabra\_al\_inicio tiene una dimensionalidad extensa, mientras que en la columna y\_num\_palabra\_al\_final se reduce la dimensionalidad aplicando el algoritmo de NLP y ML de resúmenes de textos tipo extractivos.

x_num_palabra_al_inicio = x_inicio.iloc[ : , [5]]	y_num_palabra_al_final = x_inicio.iloc[ : , [6]]
Numero_palabras_inicio	Numero_palabras_fin
0	86
11	130
12	248
13	259
14	202
...	...
1579	113

Figura 92. Verificación de la dimensionalidad inicial y final del contenido aplicando técnicas de procesamiento. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Para validación del funcionamiento de toda la data se debe pasar por las métricas de ROUGE, por lo que se lo realizó la prueba correspondiente generando resultados por cada fila de la data y un promedio general de los datos pasando como parámetro la columna del texto original y el texto generado por el algoritmo NLP de resúmenes de textos tipo extractivo.

```
[78] hyps, refs= map(list, zip(*[[d['contenido'], d['contenido_sin_formato_lista_inicio']] for d in data]))
rouge = Rouge()
scores = rouge.get_scores(hyps, refs, avg=False)
```

Figura 93. Aplicación de métricas ROUGE de toda la data para resultados por cada fila. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Se visualiza los resultados previos de cada fila generados por las métricas de ROUGE

```
[{'rouge-1': {'r': 0.049342105263157895, 'p': 1.0, 'f': 0.09404388625111783},
 'rouge-2': {'r': 0.028747433264887063, 'p': 1.0, 'f': 0.055888223009629444},
 'rouge-1': {'r': 0.049342105263157895, 'p': 1.0, 'f': 0.09404388625111783},
 'rouge-1': {'r': 0.046052631578947366, 'p': 1.0, 'f': 0.08805031362366995},
 'rouge-2': {'r': 0.025240384615384616, 'p': 1.0, 'f': 0.04923798310708087},
 'rouge-1': {'r': 0.046052631578947366, 'p': 1.0, 'f': 0.08805031362366995},
 'rouge-1': {'r': 0.19540229885057472, 'p': 1.0, 'f': 0.3269230741882397},
 'rouge-2': {'r': 0.15555555555555556, 'p': 1.0, 'f': 0.26923076690088765},
 'rouge-1': {'r': 0.19540229885057472, 'p': 1.0, 'f': 0.3269230741882397},
 'rouge-1': {'r': 0.09392265193370165, 'p': 1.0, 'f': 0.17171717014743393},
 'rouge-2': {'r': 0.05934718100890208, 'p': 1.0, 'f': 0.11204481686949291},
 'rouge-1': {'r': 0.09392265193370165, 'p': 1.0, 'f': 0.17171717014743393},
 'rouge-1': {'r': 0.2109375, 'p': 1.0, 'f': 0.34838709389719047},
```

Figura 94. Resultados de cada fila aplicando métricas ROUGE. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Aplicación de la métrica de ROUGE para generar resultados del valor promedio de toda la data, verificando su funcionalidad del algoritmo.

```
[82] scores
{'rouge-1': {'r': 0.20112828585549833, 'p': 1.0, 'f': 0.3182409094794257},
 'rouge-2': {'r': 0.16169358821620866, 'p': 1.0, 'f': 0.2631349590518316},
 'rouge-1': {'r': 0.20112828585549833, 'p': 1.0, 'f': 0.3182409094794257}}
```

Figura 95. Resultados del valor promedio de la data aplicando las métricas de ROUGE. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.8. Guardado de la data en formato JSON

En este último apartado detalla los pasos a realizar para el correcto guardado de la data unificando las columnas de interés para importarlo al prototipo web.

[83] df_new_noticias_Final								
	id	titulo	contenido	imagenes	url	created	updated	
0	1	Escuador no avisa la venta de test de covid-19...	San test de alta sensibilidad, superior al 95%...	https://lacuadrenuevo.com/media/52/imagenes/cach...	https://lacuadrenuevo.com/index.php/queyuna/...	30 Jun 2022	30 Jun 2022	
11	2	Covid prolongado: los problemas en la pol se...	Desde el día que media de dos semanas y, como a...	https://lacuadrenuevo.com/media/52/imagenes/cach...	https://lacuadrenuevo.com/index.php/queyuna/...	29 Jun 2022	30 Jun 2022	
12	3	COE Nacional analiza situación para fortalecer v...	Esta es una de las resoluciones estructuradas por...	https://lacuadrenuevo.com/media/52/imagenes/cach...	https://lacuadrenuevo.com/index.php/queyuna/...	29 Jun 2022	30 Jun 2022	
13	4	¿Por qué Orizón es más infecciosa que las ot...	Por ejemplo, si una persona sufre a un oc...	https://lacuadrenuevo.com/media/52/imagenes/cach...	https://lacuadrenuevo.com/index.php/queyuna/...	29 Jun 2022	30 Jun 2022	
14	5	PPLA reciben dosis de refuerzo contra la COVID...	El personal de la Dirección Distrital DIOH de...	https://lacuadrenuevo.com/media/52/imagenes/cach...	https://lacuadrenuevo.com/index.php/queyuna/...	30 Jun 2022	30 Jun 2022	

Figura 96. Data lista para descarga en formato Json. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



Se realiza el llamado del df listo y se aplica la extensión .json, como parámetro tendrá el nombre del archivo a descargar y configuraciones adicionales del formato de interés.

```
[93] df_new_noticiarios_final.to_json('datos_ecuadorenvivo_resumes.json', orient = 'records', compression = 'infer', index = 'false')
```

Figura 97. Descarga en formato Json. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.9. Importación al prototipo web

Se visualiza la interfaz del panel de administración de Django para importar el archivo al prototipo web con formato seleccionado json.

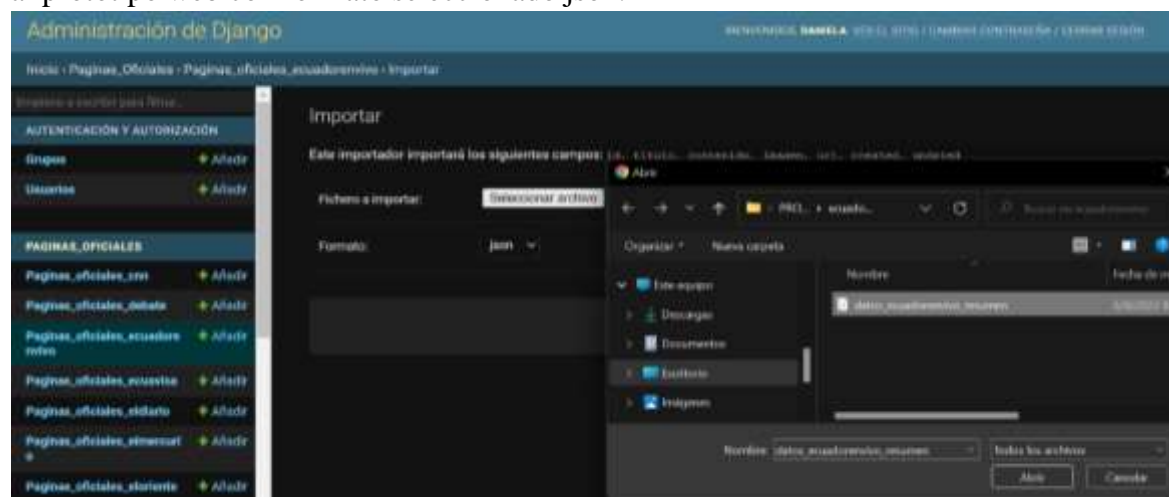


Figura 98. Importación de archivo al prototipo web. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista previa

ID	TÍTULO	CONTENIDO	IMAGEN	URL
Nuevo: 1	Ecuador no analiza la venta de test de covid 19 en boticas	San José de alta sensibilidad superó al 95%, por lo que sus resultados son fiables	<a href="https://ecuadorenvivo.com/media/92/items/cache/318c0442ac8718e070f74ef029f03c011.jpg">https://ecuadorenvivo.com/media/92/items/cache/318c0442ac8718e070f74ef029f03c011.jpg</a>	<a href="https://ecuadorenvivo.com/index.php/comunicacion/ecuador-no-analiza-la-venta-de-test-de-covid-19-en-boticas">https://ecuadorenvivo.com/index.php/comunicacion/ecuador-no-analiza-la-venta-de-test-de-covid-19-en-boticas</a>
Nuevo: 2	COVID prolongando los problemas en la piel se sumaron a los síntomas	Señala durar una media de dos semanas y como el resto de manifestaciones del virus en la piel, requiere un	<a href="https://ecuadorenvivo.com/media/92/items/cache/777a8f718ec38292054c4d13306cd7981.jpg">https://ecuadorenvivo.com/media/92/items/cache/777a8f718ec38292054c4d13306cd7981.jpg</a>	<a href="https://ecuadorenvivo.com/index.php/salud/covid-19-prolongando-los-problemas-en-la-piel-se-sumaron-a-los-sintomas">https://ecuadorenvivo.com/index.php/salud/covid-19-prolongando-los-problemas-en-la-piel-se-sumaron-a-los-sintomas</a>

Figura 99. Visualización previa de datos a importar al prototipo web. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

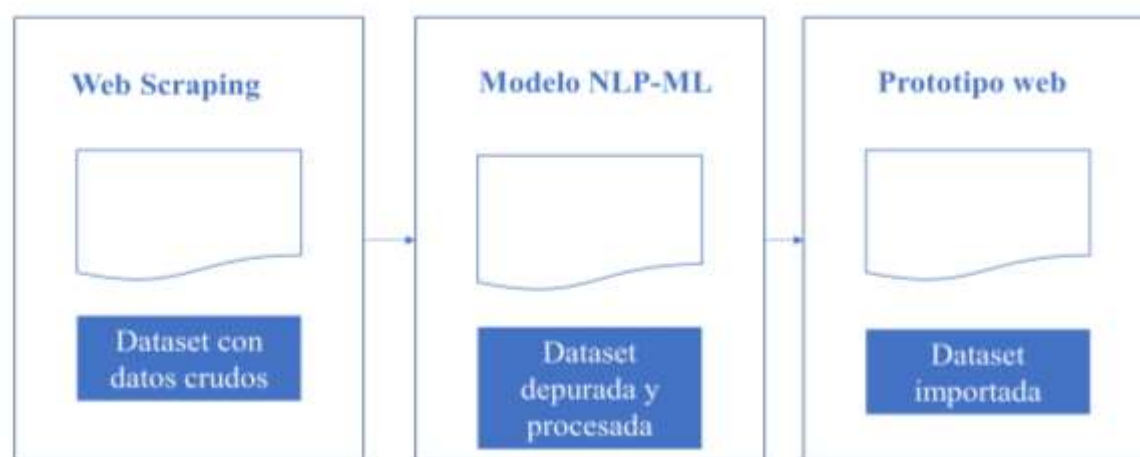


Figura 100. Modelo arquitectónico general del prototipo. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

### 3.11.3.10. Diagrama Arquitectónico para la construcción del modelo NLP de generación de resúmenes de textos tipo extractivo

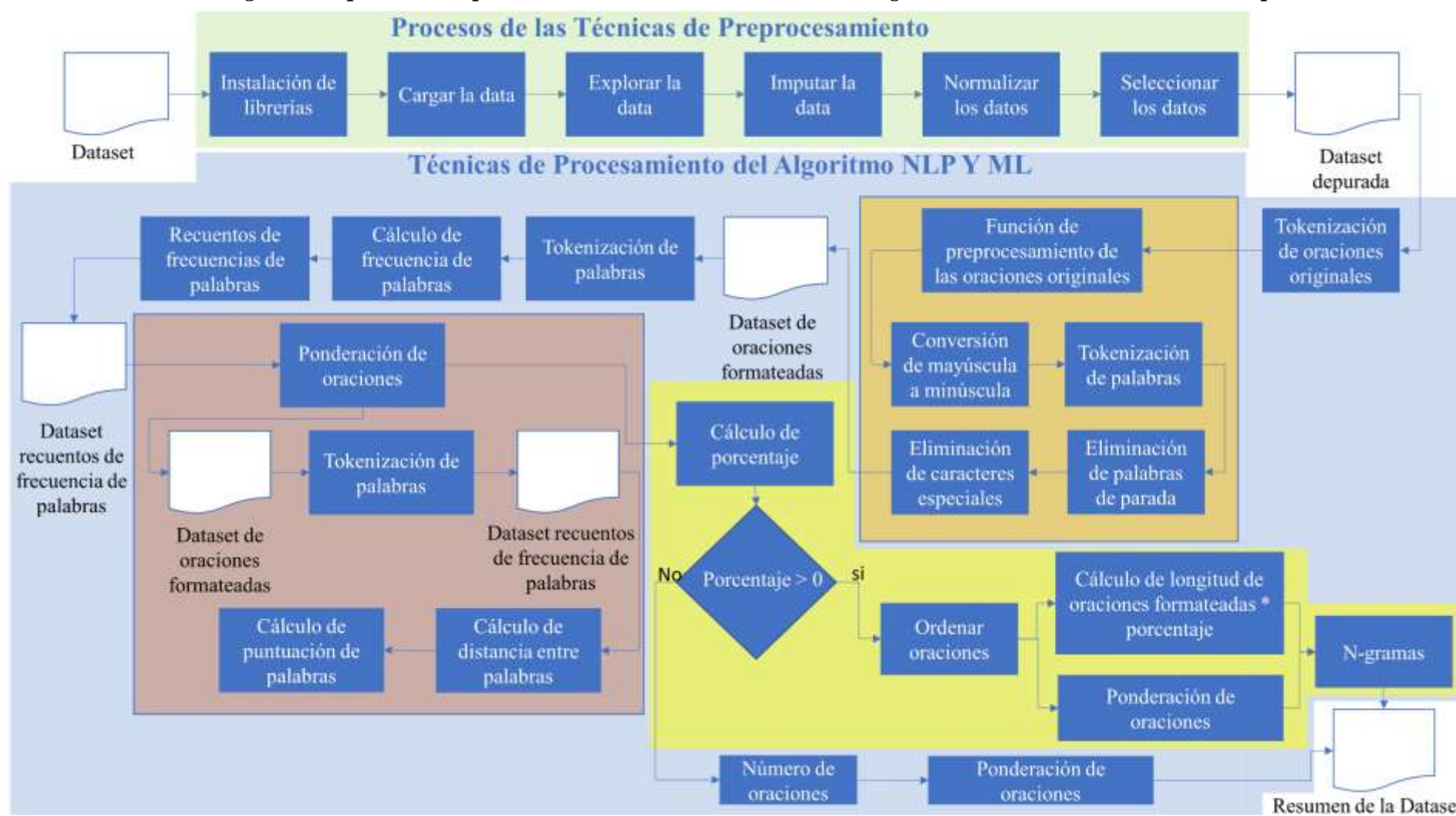


Figura 101. Diagrama arquitectónico para la construcción del modelo NLP de generación de resúmenes de textos tipo extractivo. Información tomada de la investigación directa. Elaborado por Daniela Taynys

### 3.12. Conclusiones

- Se realizó el levantamiento de información bibliográfica de 147 fuentes destacando papers, tesis doctorales, tesis universitarias, documentales, revistas, relacionados con las tecnologías de la IA, ML, algoritmos de NLP, etc.
- Se efectuó el desarrollo para el uso de la técnica Web Scraping logrando extraer 23.925 datos crudos relacionados al Covid-19.
- Se identificó el algoritmo NLP más adecuado para la generación de resúmenes de textos.
- Se realizó el diseño del diagrama arquitectónico para la construcción del modelo NLP de generación de resúmenes de textos.
- Por medio de las métricas de Rouge se validó el funcionamiento del prototipo haciendo una evaluación de tres algoritmos NLP de resúmenes de textos tipo extractivo y un algoritmo tipo abstractivo, pudiendo verificar que el más adecuado a implementar es el algoritmo de LUNH para la generación de sinopsis de noticias relacionados al Covid-19, dando los siguientes resultados rouge-1: 53,846% recuerdo(r), 100% precisión(p), 70% medida(f); rouge-2: 51,61% r, 100% p, 68,09% f; rouge-l: 53,85% r, 100% p, 70% f.

### 3.13. Recomendaciones

A partir de esta investigación se recomienda lo siguiente:

- Para futuras investigaciones se analice el uso de algoritmos de NLP en cuanto al desarrollo de resúmenes de textos tipo abstractivos y así establecer una comparación de los resultados con el algoritmo ya probado.
- Se sugiere impulsar tutorías a futuros tesisistas para trabajos de investigación y con ello seguir implementando más funcionalidades al prototipo web ya que cuenta con herramientas tecnológicas soportables.
- Se recomienda implementar el prototipo web a producción ya que permite poner a disposición el trabajo realizado a los usuarios y así obtener un feedback para validar el impacto del sitio web.



# Anexos

## **Anexo 1**

### **Entrevista realizada a profesionales de Desarrollo web**



- 1. Fecha de la Entrevista:** 11/8/2022
- 2. Nombre:** Omar Francisco Gutiérrez Pizarro
- 3. Grado de educación:** Magíster
- 4. Título de educación:** Magíster en sistemas de Información gerencial
- 5. Edad:**  
35 años
- 6. Género**  
Masculino
- 7. Lugar de Residencia**  
Guayaquil
- 8. ¿Posee conocimientos de Desarrollos Web?**  
Muy Alto
- 9. Indique los años de experiencia trabajando en temas o proyectos de Desarrollos Web.**  
8 años
- 10. Describa brevemente su experiencia o trayectoria laboral en el ámbito de Desarrollos Web.**  
Desarrollador web senior.
- 11. ¿Qué tan importante usted considera sobre el uso de los Sitios Web para difusión de información confiable o para la disponibilización de soluciones que ayuden a superar las problemáticas referenciadas a la pandemia del Covid-19?**  
Es una iniciativa importante ya que nos ayudan los medios digitales sin muchas complicaciones.
- 12. Como experto ¿Cuáles son los lenguajes de programación que usted recomendaría para el desarrollo de plataformas web?**  
Python, JavaScript y PHP.

**13. Como experto ¿Cuáles son los frameworks que usted recomendaría para el desarrollo de plataformas web?**

Angular, Laravel y Django.

**14. ¿Posee conocimientos sobre el lenguaje de programación de Python?**

Alto

**15. ¿Posee conocimientos sobre el Framework Web de Python denominado como Django?**

Alto

**16. ¿Está de acuerdo que el patrón MVC (Modelo Vista Controlador) es similar MTV (Modelo Plantilla Vista)?**

Parcialmente de acuerdo

**16.b. De haber indicado que está totalmente en desacuerdo, indique su motivo.**

El manejo de las llamadas de los controladores es demasiado tedioso.

**17. ¿Cuáles ventajas usted considera que utiliza el patrón MTV (Modelo Plantilla Vista)?**

Funcional, Escalable y Código Limpio.

**18. ¿Cuál marco o framework CSS considera adecuado para la implementación en el sitio web?**

Bootstrap, Semantic UI y Materialize.

**19. Como experto en desarrollo web detalle: ¿Cuál es la diferencia entre las clases y las identificaciones en el CSS?**

Puedo separar el contenido del estilo de la página web.



- 1. Fecha de la Entrevista:** 13/8/2022
- 2. Nombre:** Johnny José Cruz Baquerizo
- 3. Grado de educación:** Ingeniero
- 4. Título de educación:** Ingeniero en sistemas de información
- 5. Edad:**  
24 años
- 6. Género**  
Masculino
- 7. Lugar de Residencia**  
Guayaquil
- 8. ¿Posee conocimientos de Desarrollos Web?**  
Muy Alto
- 9. Indique los años de experiencia trabajando en temas o proyectos de Desarrollos Web.**  
3 años
- 10. Describa brevemente su experiencia o trayectoria laboral en el ámbito de Desarrollos Web.**  
Estar al tanto de todos los avances técnicos y teóricos que se actualizan a diario.
- 11. ¿Qué tan importante usted considera sobre el uso de los Sitios Web para difusión de información confiable o para la disponibilización de soluciones que ayuden a superar las problemáticas referenciadas a la pandemia del Covid-19?**  
Muy importante
- 12. Como experto ¿Cuáles son los lenguajes de programación que usted recomendaría para el desarrollo de plataformas web?**  
Python, C# y PHP.
- 13. Como experto ¿Cuáles son los frameworks que usted recomendaría para el desarrollo de plataformas web?**  
Angular, Django y Spring Framework.
- 14. ¿Posee conocimientos sobre el lenguaje de programación de Python?**  
Muy Alto

**15. ¿Posee conocimientos sobre el Framework Web de Python denominado como Django?**

Muy Alto

**16. ¿Está de acuerdo que el patrón MVC (Modelo Vista Controlador) es similar MTV(Modelo Plantilla Vista)?**

Ni de acuerdo ni en desacuerdo

**17. ¿Cuáles ventajas usted considera que utiliza el patrón MTV(Modelo Plantilla Vista)?**

Mantenible, Escalable y Trabajo de forma pragmática (Práctica).

**18. ¿Cuál marco o framework CSS considera adecuado para la implementación en el sitio web?**

Bootstrap, Semantic UI y Materialize

**19. Como experto en desarrollo web detalle: ¿Cuál es la diferencia entre las clases y las identificaciones en el CSS?**

Se puede trabajar de una manera más escalable.



- 1. Fecha de la Entrevista:** 14/8/2022
- 2. Nombre:** Christian Calero
- 3. Grado de educación:** Master en curso 10 de 12 módulos
- 4. Título de educación:** Ingeniero en sistemas administrativos computarizados
- 5. Edad:**  
39 años
- 6. Género**  
Masculino
- 7. Lugar de Residencia**  
Guayaquil
- 8. ¿Posee conocimientos de Desarrollos Web?**  
Medio
- 9. Indique los años de experiencia trabajando en temas o proyectos de Desarrollos Web.**  
6 años
- 10. Describa brevemente su experiencia o trayectoria laboral en el ámbito de Desarrollos Web.**  
Procesos e implementaciones.
- 11. ¿Qué tan importante usted considera sobre el uso de los Sitios Web para difusión de información confiable o para la disponibilización de soluciones que ayuden a superar las problemáticas referenciadas a la pandemia del Covid-19?**  
Importancia alta, empresas que no lo consideren tienen riesgos muy altos de perder ante competencia con criterios tecnológicos.
- 12. Como experto ¿Cuáles son los lenguajes de programación que usted recomendaría para el desarrollo de plataformas web?**  
Python, C# y PHP.
- 13. Como experto ¿Cuáles son los frameworks que usted recomendaría para el desarrollo de plataformas web?**  
Laravel, Django y Net core.

**14. ¿Posee conocimientos sobre el lenguaje de programación de Python?**

Medio

**15. ¿Posee conocimientos sobre el Framework Web de Python denominado como Django?**

Medio

**16. ¿Está de acuerdo que el patrón MVC (Modelo Vista Controlador) es similar MTV (Modelo Plantilla Vista)?**

Parcialmente de acuerdo

**16.b. De haber indicado que está totalmente en desacuerdo, indique su motivo.**

No son iguales. Pero coinciden en parte.

**17. ¿Cuáles ventajas usted considera que utiliza el patrón MTV (Modelo Plantilla Vista)?**

Funcional, Mantenible y Trabajo de forma pragmática (Práctica).

**18. ¿Cuál marco o framework CSS considera adecuado para la implementación en el sitio web?**

Bootstrap, Foundation y Materialize.

**19. Como experto en desarrollo web detalle: ¿Cuál es la diferencia entre las clases y las identificaciones en el CSS?**

Permite aplicar de manera rápida modificaciones respetando la estructura definida.

**Anexo 2****Entrevista realizada a profesionales de Inteligencia Artificial**

**1. Fecha de la Entrevista:** 14/8/2022

**2. Nombre:** Ariosto Cabrera

**3. Grado de educación:** Magíster

**4. Título de educación:** Master in Business Analytics

**5. Describa brevemente su trayectoria o experiencia laboral en Inteligencia Artificial:**

He tenido oportunidad de participar en diferentes proyectos de Data Warehouse, análisis de redes sociales a través de NLP, aplicación de Machine Learning en diferentes proyectos y ejercicios de reconocimiento de imágenes.

**6. Edad**

53 años

**7. Género**

Masculino

**8. Lugar de Residencia**

Guayaquil

**9. Escoja cuál sería su nivel de conocimiento en Inteligencia Artificial**

Alto

**10. Indique los años de experiencia trabajando en temas o proyectos de Inteligencia Artificial.**

5 años

**11. ¿Posee conocimientos de la rama de Inteligencia Artificial denominada como Machine Learning (Aprendizaje Automático)?**

Alto

**12. ¿Posee conocimientos de la rama de la Inteligencia Artificial denominada Procesamiento de Lenguaje Natural (NLP)?**

Alto

**13. ¿Qué tan importante usted considera sobre el uso de la Inteligencia Artificial y NLP para soluciones de problemáticas referenciadas a la pandemia del Covid-19?**



La inteligencia artificial y NLP puede aplicar muy bien en análisis de Covid, la técnica a usar dependerá del problema a resolver.

**14. ¿Ha trabajado en proyectos de generación de resúmenes automáticos mediante algoritmos de NLP?**

Ni de acuerdo ni en desacuerdo

**15. ¿Considera usted que el aplicar el nivel pragmático en el proceso de generación de resúmenes de textos escritos en español, permitirá obtener resúmenes con información de mayor utilidad para el lector?**

Parcialmente de acuerdo

**16. ¿Qué módulo de Machine Learning (Aprendizaje No Supervisado) considera usted adecuado para usarlo en una arquitectura NLP creada para resúmenes automáticos de textos de noticias oficiales de covid-19?**

Escala multidimensional (MDS), Mapeo de características isométricas (ISOMAP) e Incrustación localmente lineal (LLE).

**17. De la lista de algoritmos de NLP ¿Cuál considera usted que sería el más adecuado para la generación de resúmenes de textos tipo extractivos y transpórtalo a un portal web destinado a una población, combatiendo los Fake News del Covid-19?**

Basado en el peso de palabras (LUHN) y Basado en Matriz de Similitud.

**18. ¿De acuerdo a sus conocimientos cuál de las técnicas de preprocesamiento de texto considera usted los más adecuados para utilizar en un proyecto de generación de resúmenes de textos?**

Incluir librerías, Carga de la data, Explorar la data y Extraer datos.

**19. De la lista de procesamiento del algoritmo NLP ¿Cuál considera usted los más adecuados para utilizar en un proyecto de generación de resúmenes de textos?**

Conversión de texto de mayúscula a minúscula, Eliminación de palabras de parada, Peso de cada una de las palabras, Tokenización, Cálculo de ponderación en oraciones y Orden de oraciones.

**20. De la lista de IDE's para desarrollos en la nube ¿Cuál considera usted que sería el más adecuado para la implementación del algoritmo de generación de resúmenes automáticos de NLP?**

Azure Machine Learning y Amazon Machine Learning.

**21. ¿Cuál método para extracción de noticias cree usted que se deba aplicar para poderlas procesar en un algoritmo NLP de generación de resúmenes y transportarlo a un sitio donde se alojen solo noticias relacionadas al Covid-19?**

Técnica de Web Scraping utilizando desarrollo de código y Raspado web usando API de datos.



**1. Fecha de la Entrevista:** 18/8/2022

**2. Nombre:** Néstor Montaña

**3. Grado de educación:** Magíster

**4. Título de educación:** Master en Ciencias Estadísticas

**5. Describa brevemente su trayectoria o experiencia laboral en Inteligencia Artificial:**

He sido consultor en proyectos de Ciencia de Datos por más de 10 años, participando de diversos proyectos, proyectos con datos tabulares, imágenes o texto.

**6. Edad**

36 años

**7. Género**

Masculino

**8. Lugar de Residencia**

Guayaquil

**9. Escoja cuál sería su nivel de conocimiento en Inteligencia Artificial**

Alto

**10. Indique los años de experiencia trabajando en temas o proyectos de Inteligencia Artificial.**

12 años

**11. ¿Posee conocimientos de la rama de Inteligencia Artificial denominada como Machine Learning (Aprendizaje Automático)?**

Alto

**12. ¿Posee conocimientos de la rama de la Inteligencia Artificial denominada Procesamiento de Lenguaje Natural (NLP)?**

Alto

**13. ¿Qué tan importante usted considera sobre el uso de la Inteligencia Artificial y NLP para soluciones de problemáticas referenciadas a la pandemia del Covid-19?**

Considero que es muy importante la automatización en general, permite aumentar el alcance de las actividades, sea detección de una enfermedad, esquematización de síntomas o responder a preguntas frecuentes.

**14. ¿Ha trabajado en proyectos de generación de resúmenes automáticos mediante algoritmos de NLP?**

Parcialmente en desacuerdo

**15. ¿Considera usted que el aplicar el nivel pragmático en el proceso de generación de resúmenes de textos escritos en español, permitirá obtener resúmenes con información de mayor utilidad para el lector?**

Parcialmente de acuerdo

**16. ¿Qué módulo de Machine Learning (Aprendizaje No Supervisado) considera usted adecuado para usarlo en una arquitectura NLP creada para resúmenes automáticos de textos de noticias oficiales de covid-19?**

Incrustación de vecinos estocásticos distribuidos en t (t-SNE)

**17. De la lista de algoritmos de NLP ¿Cuál considera usted que sería el más adecuado para la generación de resúmenes de textos tipo extractivos y transpórtalo a un portal web destinado a una población, combatiendo los Fake News del Covid-19?**

Basado en Matriz de Similitud

**18. ¿De acuerdo a sus conocimientos cuál de las técnicas de preprocesamiento de texto considera usted los más adecuados para utilizar en un proyecto de generación de resúmenes de textos?**

Incluir librerías, Carga de la data, Explorar la data, Imputar datos, Codificar datos, Escalar datos, Normalizar datos, Seleccionar datos y Extraer datos.

**19. De la lista de procesamiento del algoritmo NLP ¿Cuál considera usted los más adecuados para utilizar en un proyecto de generación de resúmenes de textos?**

Conversión de texto de mayúscula a minúscula, Eliminación de palabras de parada, Eliminación de string, Cálculo de frecuencia de palabras, Peso de cada una de las palabras, Tokenización, Escalamiento de datos, Cálculo de ponderación en oraciones, Orden de oraciones.

**20. De la lista de IDE's para desarrollos en la nube ¿Cuál considera usted que sería el más adecuado para la implementación del algoritmo de generación de resúmenes automáticos de NLP?**

Google Colab.

**21. ¿Cuál método para extracción de noticias cree usted que se deba aplicar para poderlas procesar en un algoritmo NLP de generación de resúmenes y transportarlo a un sitio donde se alojen solo noticias relacionadas al Covid-19?**

Técnica de Web Scraping utilizando desarrollo de código.

### Anexo 3

#### **Modelo de la encuesta realizada a los ciudadanos de la Zona 8 de la Provincia del Guayas**

##### ENCUESTA PARA EL DESARROLLO DE UN SITIO WEB QUE CENTRALICE NOTICIAS OFICIALES RELACIONADAS AL COVID-19

La Universidad de Guayaquil a través de sus investigadores impulsa la creación de soluciones tecnológicas siendo así que la intencionalidad sobre esta encuesta es entender las diversas situaciones que se encuentran en la población de la zona-8.

Por lo cual, se requiere conocer el nivel de aceptación de un canal informativo que centralice las noticias relacionadas con el Covid-19 o futuros virus que afecten a la comunidad de la zona-8.

Considere las siguientes premisas previo a llenar la encuesta

- La encuesta es anónima.
- Lea las instrucciones antes de contestar.
- Contestar cada una de las preguntas.

##### PARTE 1: DATOS INFORMATIVOS

###### 1.1 Indique su Edad (en años)

---

###### 1.2 Seleccione el Género con el que se identifica

- Femenino
- Masculino

###### 1.3 Lugar que reside de la zona 8 del Ecuador

- Guayaquil
- Durán
- Samborondón
- Otra ciudad del Ecuador
- Otro País

###### 1.3.b De haber indicado que reside en otra ciudad del Ecuador, indique cual.

---

1.3.c De haber indicado que reside en otro País, indique cual.

---

## PARTE 2: MEDIOS INFORMATIVOS UTILIZADOS PARA REVISAR NOTICIAS DE CORONA VIRUS (COVID-19)

2.1 ¿A través de que medios de comunicación usted se informa sobre noticias del covid-19?

- Redes sociales (Facebook, Instagram, Twitter, YouTube, etc)
- Noticieros por canales de televisión o radio
- Páginas web recomendadas por conocidos
- Buscadores de internet (Google, Bing, Yahoo!, Ask, etc)
- Publicidades en medios electrónicos (imágenes, audios digitales, etc)
- Referencias verbales
- Páginas web de entidades oficiales (Ministerio de Salud, Organización Mundial de la Salud, entre otros)
- Ningún medio
- Otros Medios

2.1.b En caso de haber seleccionado que utiliza otros medios, indique cuáles.

---

2.2 Cuando usted busca información reciente relacionada con el covid-19, de los portales que ofrece la información. ¿Cuál es su nivel de dificultad que usted ha encontrado al navegar o interactuar en ellos para encontrar la noticia de su interés?

- Extremadamente fácil
- Muy fácil
- Ni de fácil ni difícil
- Medianamente difícil
- Extremadamente difícil

2.3 ¿Está usted de acuerdo en que se impulse un proyecto que construya un sitio web que centralice las noticias relacionadas con el covid-19 las cuales se encuentran dispersas en múltiples páginas oficiales?

- Totalmente de acuerdo
- Parcialmente de acuerdo
- Ni de acuerdo ni en desacuerdo

- Parcialmente en desacuerdo
- Totalmente en desacuerdo

2.4 ¿Al contar con un sitio web el cual centralicen las noticias de múltiples páginas oficiales del covid-19, considera usted que su nivel de preocupación se disminuiría al poder verificar con la ayuda de este portal la veracidad de noticias que llegan por redes sociales u otros medios informales?

- Totalmente de acuerdo
- Parcialmente de acuerdo
- Ni de acuerdo ni en desacuerdo
- Parcialmente en desacuerdo
- Totalmente en desacuerdo

2.5 ¿Considera usted que el contar con un sitio web que centralicen las noticias de múltiples páginas oficiales del covid-19 le ayudarían a usted a evitar el uso y/o compartición de noticias falsas (Fake News)?

- Totalmente de acuerdo
- Parcialmente de acuerdo
- Ni de acuerdo ni en desacuerdo
- Parcialmente en desacuerdo
- Totalmente en desacuerdo

2.6 ¿Considera usted importante contar con esta herramienta tecnológica de centralización de noticias del covid-19 y que la misma le permita llegar al detalle de la noticia de interés en el respectivo sitio oficial?

- Totalmente de acuerdo
- Parcialmente de acuerdo
- Ni de acuerdo ni en desacuerdo
- Parcialmente en desacuerdo
- Totalmente en desacuerdo

2.7 ¿Cuál es el tiempo que tiene disponible en el día para revisión de noticias referentes al covid-19 a través de dispositivos digitales (Celular, Laptop, PC)?

- Alto tiempo de disponibilidad (entre 3 a 5 horas)
- Medio tiempo de disponibilidad (entre 2 a 3 horas)

- Poco tiempo de disponibilidad (entre 1 a 2 horas)
- Muy poco tiempo de disponibilidad (Menos de 1 hora)
- No poseo tiempo disponible

### PARTE 3: ASISTENTES VIRTUALES

Se considera asistente virtual, la comunicación entre el ser humano y un dispositivo inteligente llevando una interacción de hombre-máquina.

#### 3.1 ¿Usted cuenta con al menos un dispositivo digital?

- Si cuento con un Smart Phone(Celular) de Gama Alta
- Si cuento con un Smart Phone(Celular) de Gama Media
- Si cuento con un Smart Phone(Celular) de Gama Baja
- Si cuento con un Smart TV
- Si cuento con un Ordenador (PC de escritorio)
- Si cuento con una Laptop
- Si cuento con una Tabled
- No cuento con ningún dispositivo digital, pero pienso adquirir uno a corto plazo.
- No cuento con ningún dispositivo digital y no planeo adquirir uno.

#### 3.2 ¿Cuál es el tipo de enlace a internet que emplea habitualmente?

- Acceso Telefónico (Teléfono convencional)
- Acceso por Banda Ancha ADSL
- Acceso por Cable modem (TV CABLE)
- Acceso por Teléfono Móvil (Megas)
- Acceso por Fibra Óptica
- Acceso Inalámbrico (Acceso internet público)
- Acceso Satelital
- Acceso por Línea Eléctrica
- No cuento con ningún acceso a internet, pero pienso contratarlo en el corto plazo
- Actualmente no cuento con ningún acceso a internet y no planeo contratarlo

#### 3.3 ¿Cuál es el ancho de banda del enlace que usted utiliza?

- 1 Mbps a 25 Mbps
- 26 Mbps a 50 Mbps

- 51 Mbps a 75 Mbps
- 76 Mbps a 100 Mbps
- Otra velocidad de internet
- Actualmente no cuento con esa información
- Actualmente no poseo conectividad de internet

3.3.b En caso de haber seleccionado que tiene otra velocidad de internet, indique cual.

---

3.4 ¿Cuántas personas están conectadas al mismo plan de internet que usted utiliza?

- 1 persona se conecta al mismo plan de internet
- 2 personas se conectan al mismo plan de internet
- 3 personas se conectan al mismo plan de internet
- 4 personas se conectan al mismo plan de internet
- 5 o más personas se conectan al mismo plan de internet
- No cuento con ningún acceso de internet

3.4. b En caso de haber seleccionado que 5 o más personas se conectan al mismo plan de internet, indique cuántas.

---



**Anexo 4**

# **Manual de usuario del Prototipo Web**

## ÍNDICE DE FIGURAS

<b>Nº</b>	<b>Descripción</b>	<b>Pág.</b>
1	Vista principal del usuario con acceso libre al prototipo web	120
2	Acceso a categorías de páginas noticiarias sobre el Covid-19 por medio de una modal	121
3	Indicación para cerrar la modal de listado de categorías de páginas noticiarias	122
4	Guía para acceder al servicio de Correo Electrónico	123
5	Vista del servicio de Correo Electrónico al dar clic en el icono	123
6	Indicación para acceder a la red social de Facebook	124
7	Vista de red social Facebook al dar clic en el icono	124
8	Para acceder a la red social de Twitter	125
9	Vista de red social Twitter al dar clic en el icono	125
10	Guía para acceder a la red social de Instagram	126
11	Vista de red social de Instagram al dar clic en el icono de identificación	126
13	Vista de red social de YouTube al dar clic en su icono de identificación	127
14	Guía para acceder al servicio de Google Maps	128
15	Vista al servicio de Google Maps dando clic en vista previa de ubicación	128
16	Indicación para acceso a categorías de páginas noticiarias sobre el Covid-19 por medio de una modal	129
17	Vista al dar clic en una categoría de página noticiaria referente al Covid-19 e indicación para acceder a la página oficial de noticia de interés	130
18	Vista del detalle completo de noticia de interés dando clic en cualquier bloque de artículo de sinopsis desplegado en el prototipo web	132

## **Presentación**

Durante el presente manual se podrá validar el correcto funcionamiento de cada uno de los componentes integrados en el prototipo web, por lo que se presentará las funcionalidades de cada vista tales como: dirigirse a las secciones donde se despliegan el listado de páginas noticiarias pudiendo tener accesibilidad de visualizar las noticias referentes al Covid-19 y su redireccionamiento a la página oficial de noticia de su interés, por otro lado se podrá observar el acceso que ofrece desde la vista principal a Redes Sociales más usados Facebook, Twitter, Instagram, Youtube, tomando como iniciativa la interacción con el usuario que pueda desplegarse de manera amigable ya que en la actualidad la comunicación por estos medios es muy amplia. De esta manera el usuario tendrá la disponibilidad de verificar por medio de este prototipo web la veracidad de noticias referentes al Covid-19.

## **Objetivo General**

Describir el manejo del prototipo web para tener una visualización de la funcionalidad y mostrando una mejor perspectiva por lo que fue creado.

Vista principal del usuario con accesibilidad libre al prototipo web.

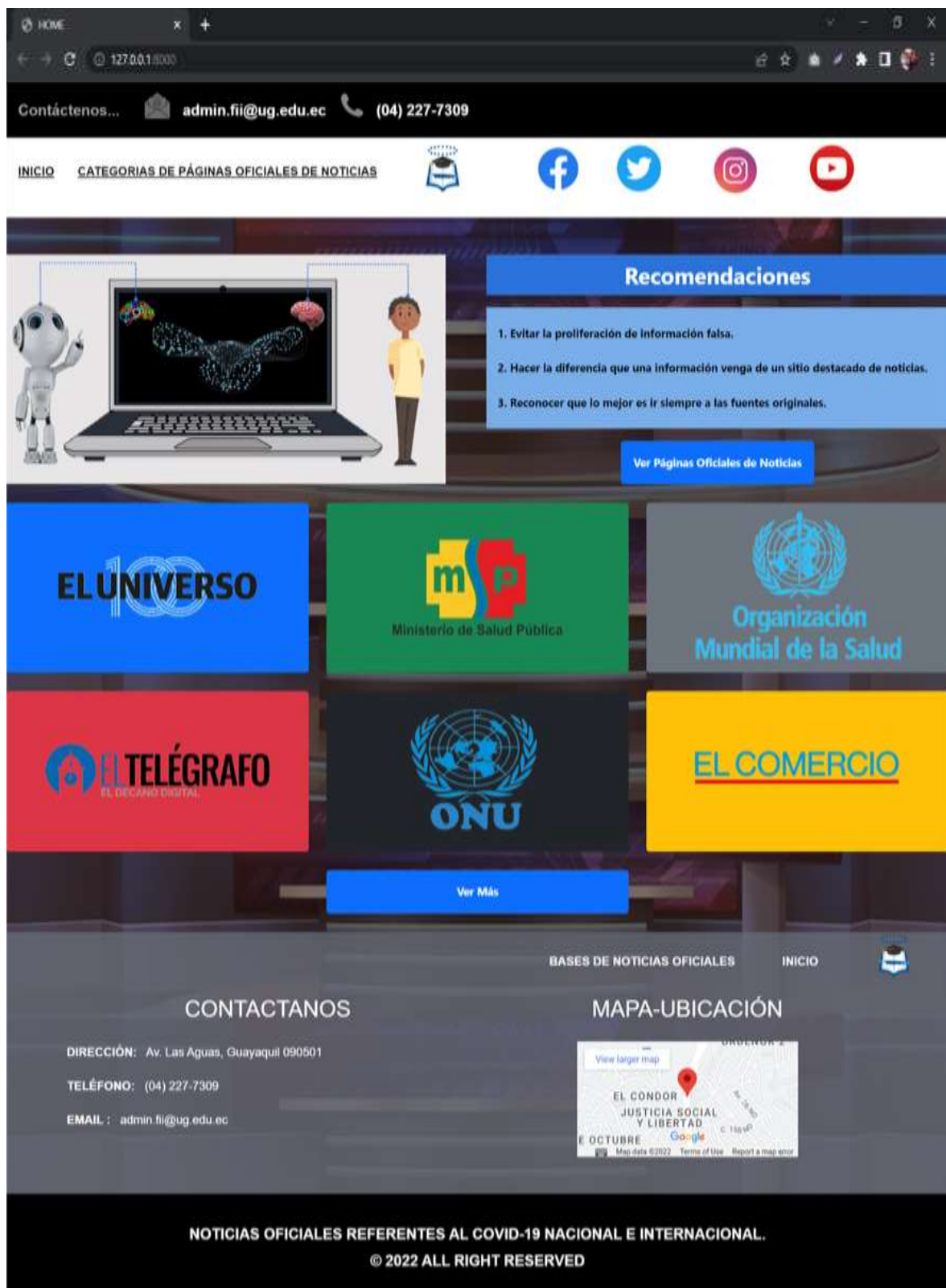


Figura 1. Vista principal del usuario con acceso libre al prototipo web. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Acceso desde la vista principal del usuario al despliegue de categorías de páginas noticiarias combinadas entre nacionales e internacionales por medio de una modal, pudiendo tener 4 métodos para accesibilidad.

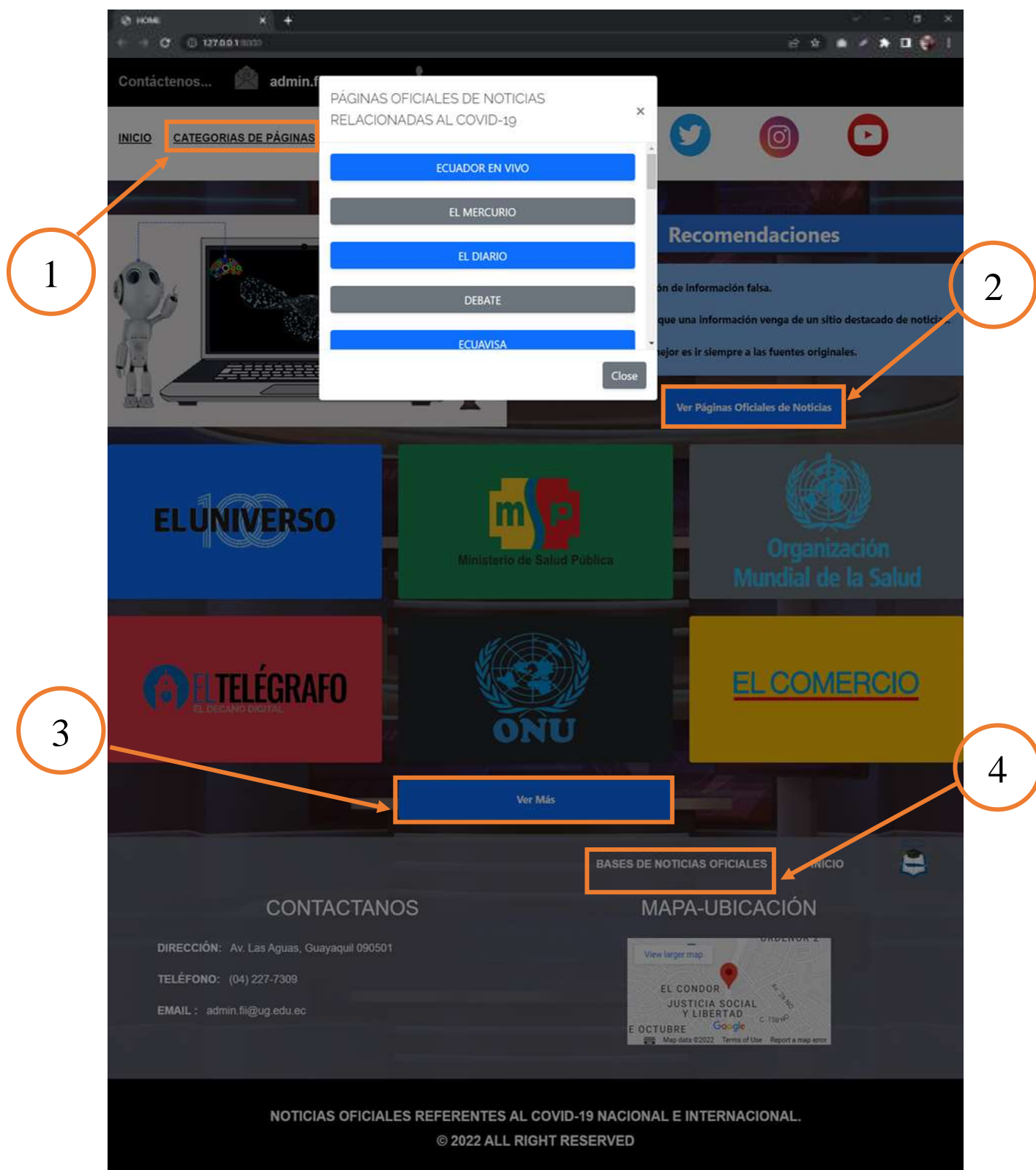


Figura 2. Acceso a categorías de páginas noticiarias sobre el Covid-19 por medio de una modal. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Para cerrar la modal del despliegue de categorías de páginas noticiarias combinadas entre nacionales e internacionales, para lo siguiente se tiene 2 métodos.

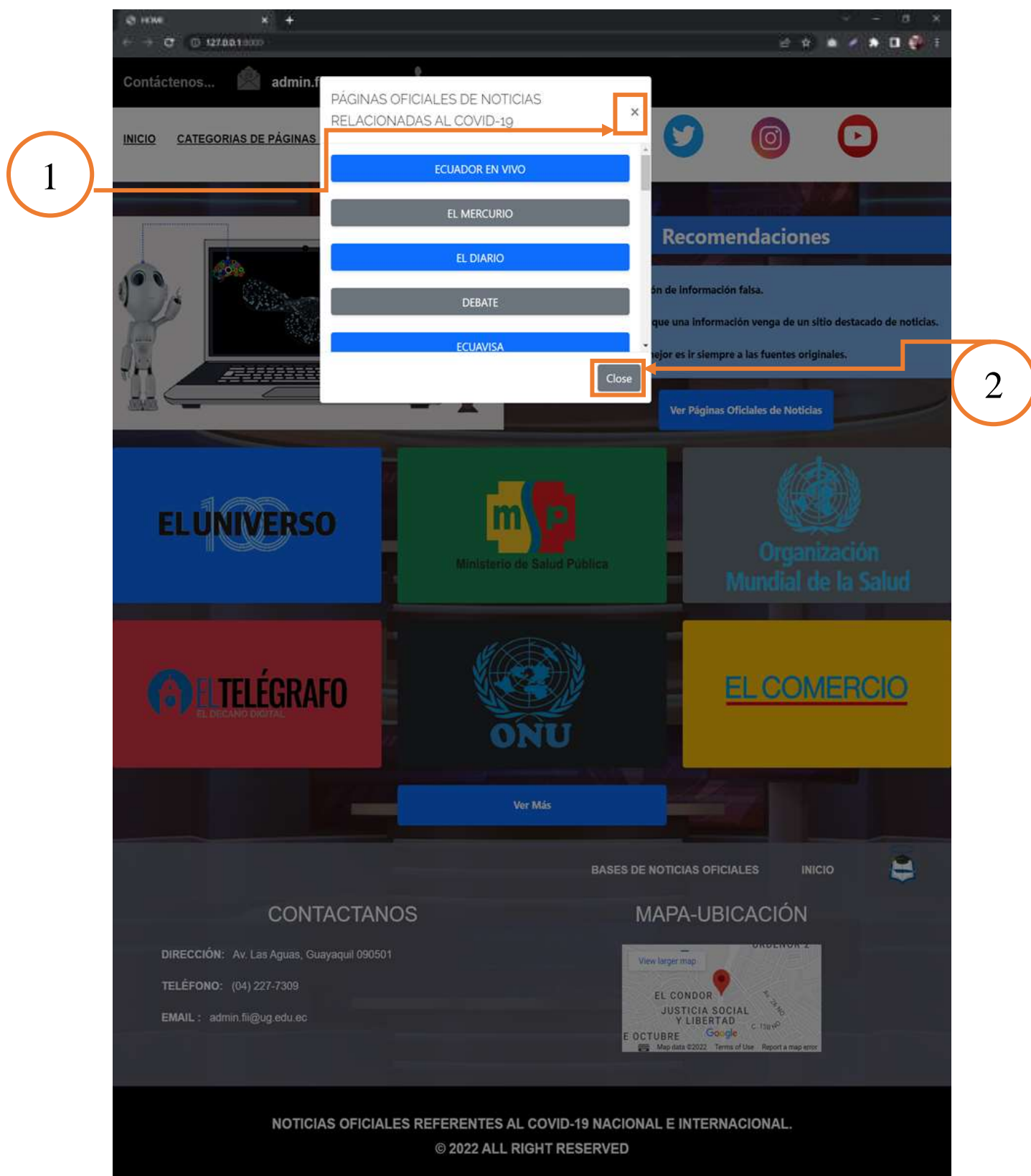


Figura 3. Indicación para cerrar la modal de listado de categorías de páginas noticiarias. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Para dirigirse al servicio de Correo Electrónico dar clic en su icono de identificación de la siguiente manera.



Figura 4. Guía para acceder al servicio de Correo Electrónico. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista al dar clic en el icono del servicio de Correo Electrónico.

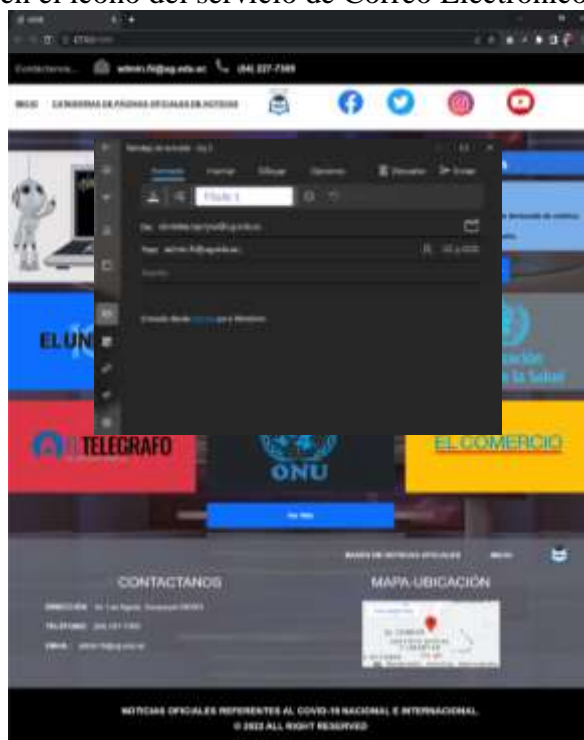


Figura 5. Vista del servicio de Correo Electrónico al dar clic en el icono. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



Para dirigirse a la Red social de Facebook dar clic en su icono de identificación de la siguiente manera.

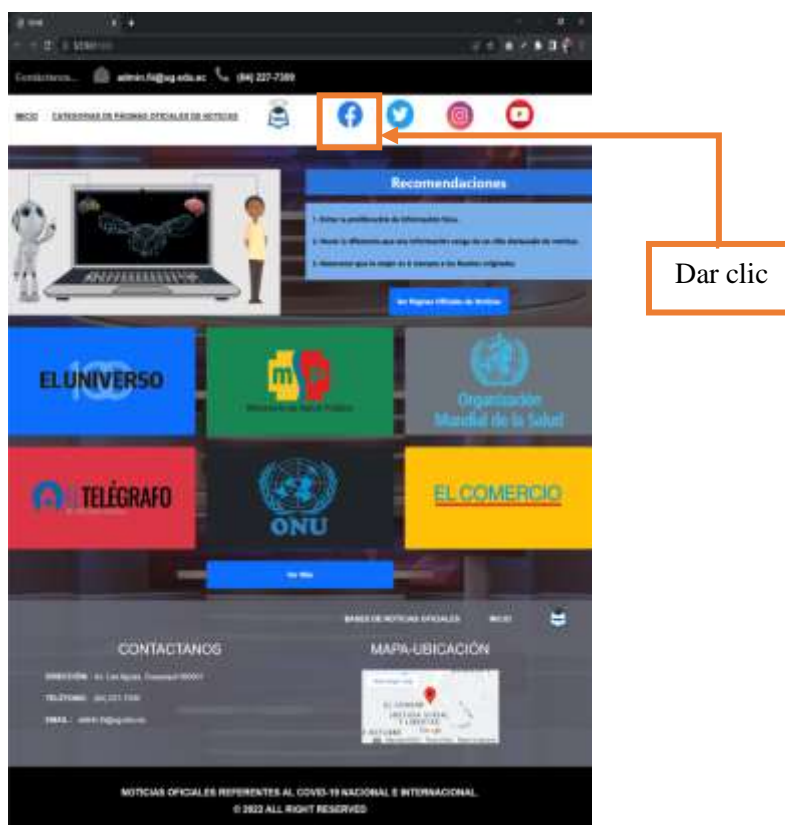


Figura 6. Indicación para acceder a la red social de Facebook. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista al dar clic en el icono de Facebook.



Figura 7. Vista de red social Facebook al dar clic en el icono. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



Para dirigirse a la Red social de Twitter, dar clic en su icono de identificación de la siguiente manera.

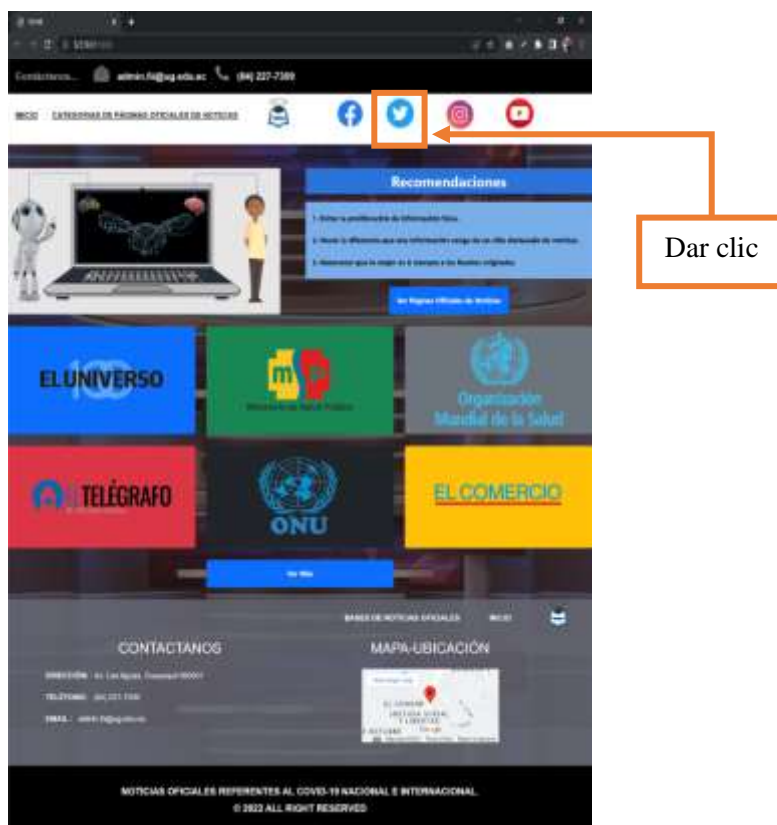


Figura 8. Para acceder a la red social de Twitter. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista al dar clic en el icono de Twitter.



Figura 9. Vista de red social Twitter al dar clic en el icono. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Para dirigirse a la Red social de Instagram dar clic en su icono de identificación de la siguiente manera.

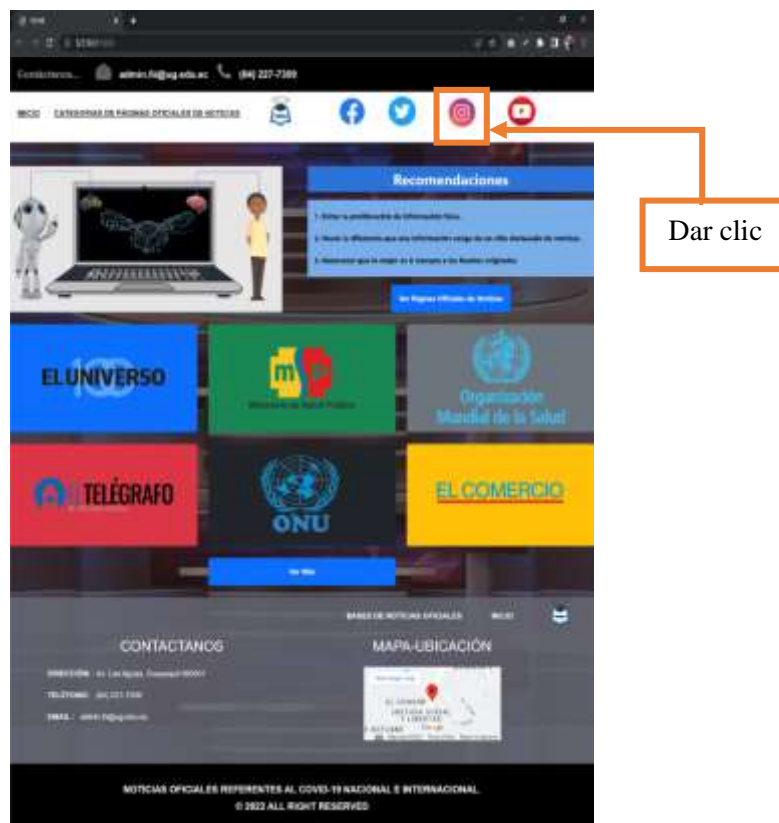


Figura 10. Guía para acceder a la red social de Instagram. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista al dar clic en el icono de Instagram.

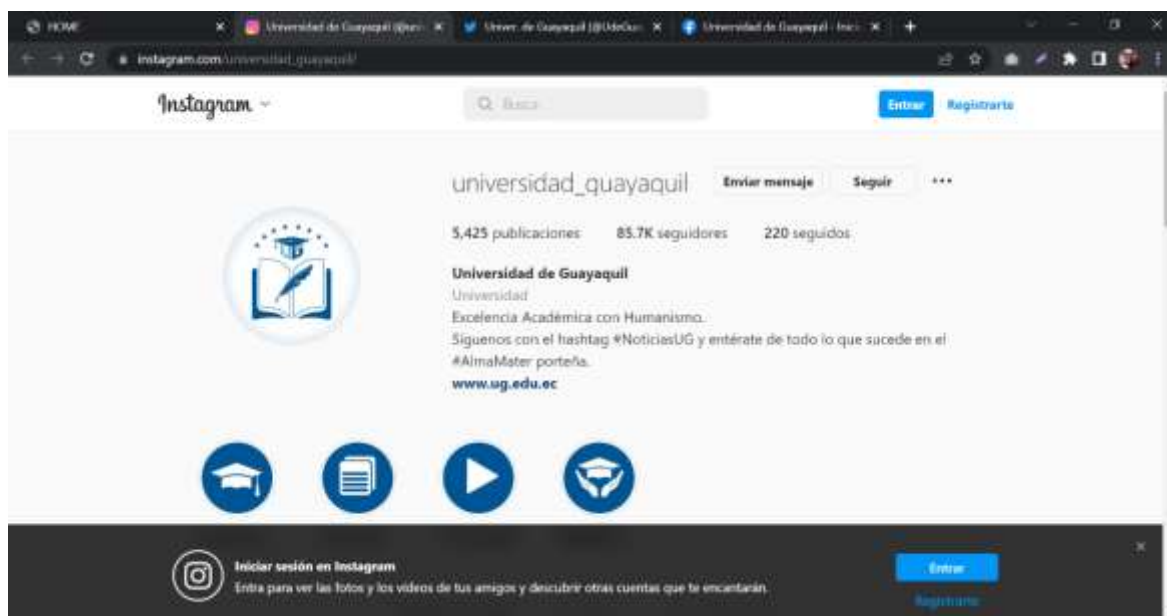


Figura 11. Vista de red social de Instagram al dar clic en el icono de identificación. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Para dirigirse a la Red social de YouTube dar clic en su icono de identificación de la siguiente manera.

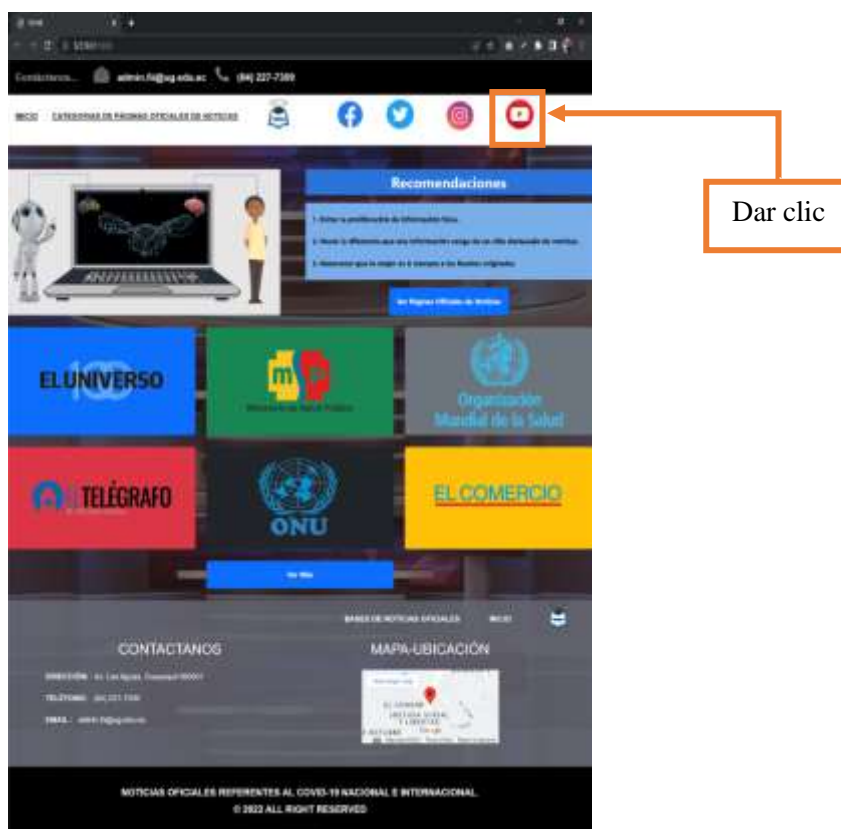


Figura 12. Indicación para acceder a la red social de YouTube. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista al dar clic en el icono de YouTube.



Figura 13. Vista de red social de YouTube al dar clic en su icono de identificación. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Para dirigirse a Google Maps dar clic en la vista previa del servicio de Google Maps de la siguiente manera.



Figura 14. Guía para dar clic en la vista previa del servicio de Google Maps. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista al dar clic en la vista previa del servicio de Google Maps.

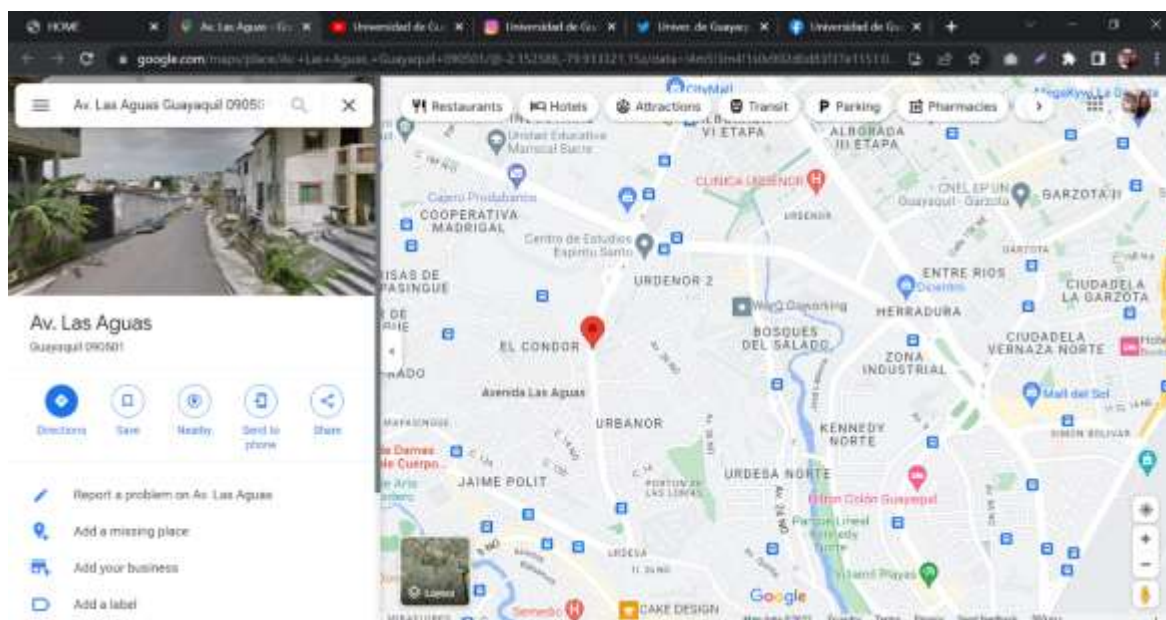


Figura 15. Vista al servicio de Google Maps dando clic en vista previa de ubicación. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



Acceso a una página de interés para lecturas rápidas de noticias sobre la terminología de Covid-19 accediendo desde la vista principal del usuario desplegando las categorías de páginas noticiarias combinadas entre nacionales e internacionales por medio de una modal.

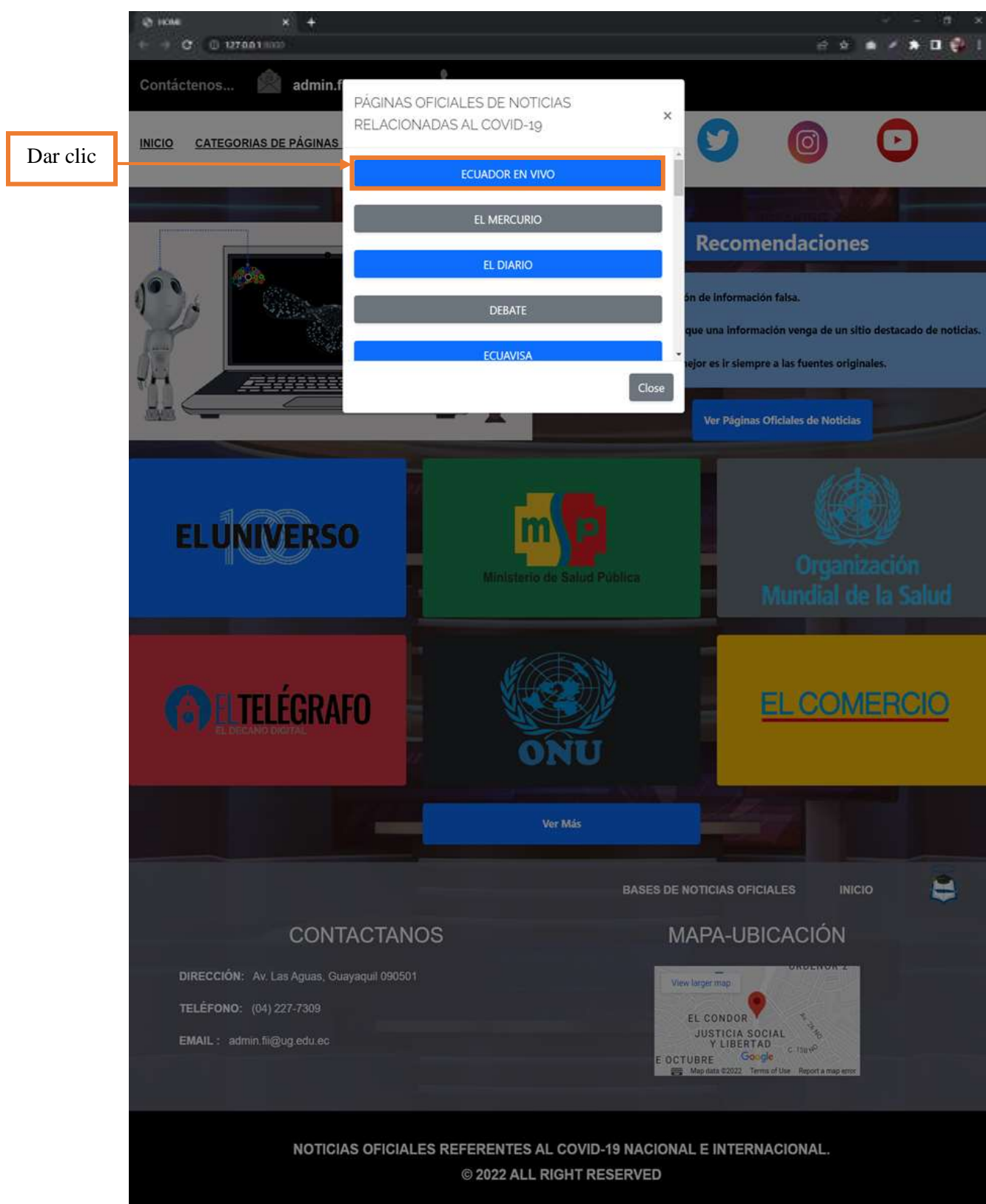


Figura 16. Indicación para acceso a categorías de páginas noticiarias sobre el Covid-19 por medio de una modal. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Vista de sinopsis de noticias al dar clic en una de las categorías de interés. Para redireccionar a la página oficial del artículo de noticias se debe dar clic en cualquiera de los bloques desplegados en referencia a su noticia de interés.

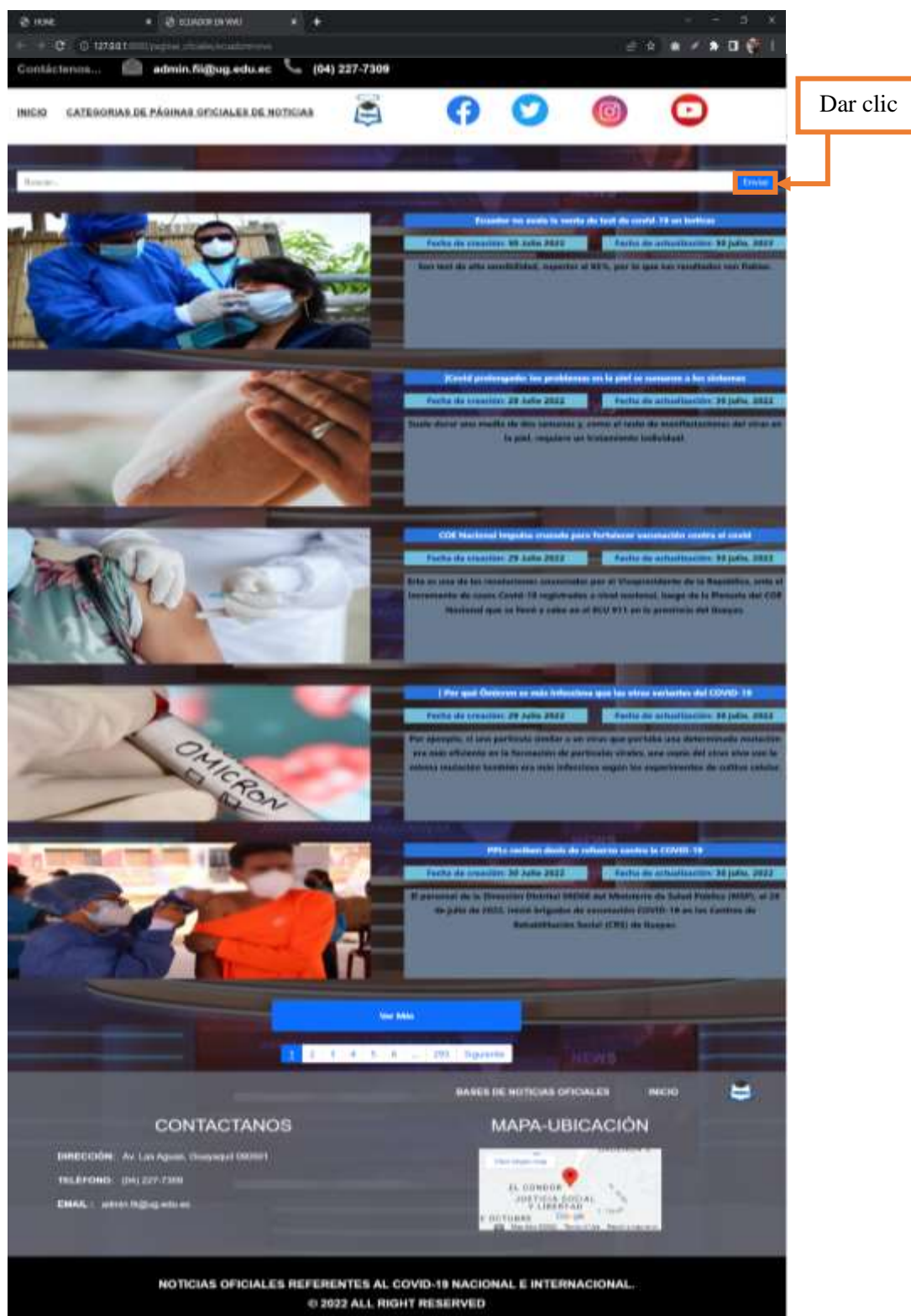


Figura 17. Vista al dar clic en una categoría de página noticiaría referente al Covid-19 e indicación para acceder a la página oficial de noticia de interés. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Al dar clic en enviar proporcionando inicialmente el ingreso de un texto en el buscador se desplegará el listado de noticias correspondiente al enunciado ingresado, visualizando la cantidad de artículos encontrados.

Texto de interés y números de artículos

Estás buscando: COE  
Artículos encontrados: 3 artículos

COE Nacional impulsa estado para fortalecer vacunación contra el covid  
Fecha de creación: 29 Julio 2022 Fecha de actualización: 30 Julio, 2022  
Este es uno de las resoluciones avanzadas por el Vicepresidente de la República, ante el incremento de casos Covid-19 registrados a nivel nacional, luego de la Planaria del COE Nacional que se llevó a cabo en el SCU 911 en la provincia del Guayas.

COE nacional analiza este sábado situación del COVID-19 en Ecuador  
Fecha de creación: 21 Enero 2022 Fecha de actualización: 30 Julio, 2022  
Inicialmente la sesión iba a realizarse esta mañana, pero Juan Zapata, presidente del COE nacional, indicó que por pedido del Ministerio de Salud Pública (MSP) será el sábado para tener más datos de cómo está la situación.

COE Nacional evalúa medidas focalizadas en El Oro y Guayaquil contra el covid-19  
Fecha de creación: 19 Julio 2021 Fecha de actualización: 30 Julio, 2022  
Estas acciones, tomadas en días pasados, se enfocaron en el monitoreo constante de la

Contenidos de interés dar clic.

CONTACTANOS  
DIRECCIÓN: Av. Las Aguas, Guayaquil 090501  
TELÉFONO: (04) 227-7309  
EMAIL: admin.fis@ug.edu.ec

MAPA-UBICACIÓN  
EL CONDOR  
JUSTICIA SOCIAL  
LIBERTAD  
5 OCTUBRE

NOTICIAS OFICIALES REFERENTES AL COVID-19 NACIONAL E INTERNACIONAL.  
© 2022 ALL RIGHT RESERVED

Figura 18. Vista al dar clic previo a un texto ingresado en la barra del buscador. Información tomada de la investigación directa. Elaborado por Daniela Taynys.

Al dar clic en cualquiera de los contenedores de noticias va a redireccionar a la página original referente a noticia de interés del Covid-19.



Figura 19. Vista del detalle completo de noticia de interés dando clic en cualquier bloque de artículos de sinopsis desplegado en el prototipo web. Información tomada de la investigación directa. Elaborado por Daniela Taynys.



**Anexo 5**

# **Justificación legal**

### ***Constitución de la República del Ecuador***

Según Art 262 de la Constitución de la República del Ecuador (2020) en el “Régimen de competencias” establece que: “Los gobiernos regionales autónomos tendrán las siguientes competencias exclusivas, sin perjuicio de las otras que determine la ley que regule el sistema nacional de competencias:

Como indica el inciso 6: Determinar las políticas de investigación e innovación del conocimiento, desarrollo y transferencia de tecnologías, necesarias para el desarrollo regional, en el marco de la planificación nacional.”

Según Art. 277 de la Constitución de la República del Ecuador (2020) en el “Régimen de Desarrollo” establece que: “Para la consecución del buen vivir, serán deberes generales del Estado:

Como indica el inciso 6: Promover e impulsar la ciencia, la tecnología, las artes, los saberes ancestrales y en general las actividades de la iniciativa creativa comunitaria, asociativa, cooperativa y privada.”

Según Art. 350 de la Constitución de la República del Ecuador (2020) en el “Régimen del buen vivir” establece que: "El sistema de educación superior tiene como finalidad la formación académica y profesional con visión científica y humanista; la investigación científica y tecnológica; la innovación, promoción, desarrollo y difusión de los saberes y las culturas; la construcción de soluciones para los problemas del país, en relación con los objetivos del régimen de desarrollo."

Según Art. 385 de la Constitución de la República del Ecuador (2020) en la “Sección Octava de Ciencia, tecnología, innovación y saberes ancestrales” establece que: “El sistema nacional de ciencia, tecnología, innovación y saberes ancestrales, en el marco del respeto al ambiente, la naturaleza, la vida, las culturas y la soberanía, tendrá como finalidad:

Como indica el inciso 1: Generar, adaptar y difundir conocimientos científicos y tecnológicos. [..]

Como indica el inciso 3: Desarrollar tecnologías e innovaciones que impulsen la producción nacional, eleven la eficiencia y productividad, mejoren la calidad de vida y contribuyan a la realización del buen vivir.”

Según Art 386 de la Constitución de la República del Ecuador (2020) en la “Sección Octava de Ciencia, tecnología, innovación y saberes ancestrales” establece que: “El sistema comprenderá programas, políticas, recursos, acciones, e incorporará a instituciones del Estado, universidades y escuelas politécnicas, institutos de investigación públicos y

particulares, empresas públicas y privadas, organismos no gubernamentales y personas naturales o jurídicas, en tanto realizan actividades de investigación, desarrollo tecnológico, innovación y aquellas ligadas a los saberes ancestrales.”

### ***Reglamento de Régimen Académico Consejo de Educación Superior***

Según Art. 78 del Reglamento de Régimen Académico Consejo Educación Superior (2017) en el “Capítulo I de Pertinencia” establece que: “Se entenderá como pertinencia de carreras y programas académicos a la articulación de la oferta formativa, de investigación y de vinculación con la sociedad, con el régimen constitucional del Buen Vivir, el Plan Nacional de Desarrollo, los planes regionales y locales los requerimientos sociales en cada nivel territorial y las corrientes internacionales científicas y humanísticas de pensamiento.”

### ***Ley Orgánica de Comunicación***

Según Art. 35 de la Ley Orgánica de Comunicación (2019) en el “Capítulo II de la Sección II de Derechos de igualdad e interculturalidad” establece que: “Derecho al acceso universal a las tecnologías de la información y comunicación. - Todas las personas tienen derecho a acceder, capacitarse y usar las tecnologías de información y comunicación para potenciar el disfrute de sus derechos y oportunidades de desarrollo.”

### ***Ley Orgánica de Telecomunicaciones***

Según Art. 88 de la Ley Orgánica de Telecomunicaciones (2016) en el “Capítulo Único de Promoción de la Sociedad de la Información y Prestación del Servicio Universal” establece que: “Promoción de la Sociedad de la Información y del Conocimiento.

Como indica el inciso 5: Promover el desarrollo y masificación del uso de las tecnologías de información y comunicación en todo el territorio nacional.

Como indica el inciso 6: Apoyar la educación de la población en materia de informática y tecnologías de la información, a fin de facilitar el uso adecuado de los servicios o equipos.”

### ***Consultas de Normas Internacionales del uso de la IA***

De acuerdo a la normativa con respecto al manejo, entrenamiento y uso de la IA en el Estado Ecuatoriano aún no se encuentra disponible debido a que no se han desarrollado las leyes correspondientes. A continuación, se verifican leyes de vigencia en otros Estados:

*Resolución del Parlamento Europeo sobre uso policial de la Inteligencia Artificial*

Describe Adaya y Pardo (2021), que el pasado 6 de Octubre del 2021, se estableció la aprobación de la Resolución del Parlamento Europeo sobre uso policial de la IA con 377 votos a su favor, 248 en contra y 62 personas tomaron la decisión de no ejercer su derecho al voto. El presente reglamento recalca la proliferación del uso de la IA, por lo que advierte distintos riesgos sobre el uso de esta tecnología al momento de aplicarla

Según Art. 6 del Reglamento del Parlamento Europeo y del Consejo (2021) en el “Capítulo I de clasificación de los sistemas de IA como sistemas de alto riesgo” establece que: “Un sistema de IA se considerará de alto riesgo cuando reúna las dos condiciones que se indican a continuación, con independencia de si se ha introducido en el mercado o se ha puesto en servicio sin estar integrado en los productos que se mencionan en las letras a) y b):

Como indica el inciso a): el sistema de IA está destinado a ser utilizado como componente de seguridad de uno de los productos contemplados en la legislación de armonización de la Unión que se indica en el anexo II, o es en sí mismo uno de dichos productos;

Como indica el inciso b): conforme a la legislación de armonización de la Unión que se indica en el anexo II, el producto del que el sistema de IA es componente de seguridad, o el propio sistema de IA como producto, debe someterse a una evaluación de la conformidad realizada por un organismo independiente para su introducción en el mercado o puesta en servicio.”

Según Art. 10 del Reglamento del Parlamento Europeo y del Consejo (2021) en el “Capítulo II de Requisitos para los sistemas de IA de alto riesgo” establece que:

1. “Los sistemas de IA de alto riesgo que utilizan técnicas que implican el entrenamiento de modelos con datos se desarrollarán a partir de conjuntos de datos de entrenamiento, validación y prueba que cumplan los criterios de calidad”

2. “Los conjuntos de datos de entrenamiento, validación y prueba se someterán a prácticas adecuadas de gobernanza y gestión de datos. Dichas prácticas se centrarán, en particular, en:

- b) la recopilación de datos;
- c) las operaciones de tratamiento oportunas para la preparación de los datos, como la anotación, el etiquetado, la depuración, el enriquecimiento y la agregación;

- d) la formulación de los supuestos pertinentes, fundamentalmente en lo que respecta a la información que, ateniéndose a ellos, los datos miden y representan;
- e) la evaluación previa de la disponibilidad, la cantidad y la adecuación de los conjuntos de datos necesarios;
- g) la detección de posibles lagunas o deficiencias en los datos y la forma de subsanarlas.”

4. “Los conjuntos de datos de entrenamiento, validación y prueba tendrán en cuenta, en la medida necesaria en función de su finalidad prevista, las características o elementos particulares del contexto geográfico, conductual o funcional específico en el que se pretende utilizar el sistema de IA de alto riesgo.”

6. “Se emplearán prácticas adecuadas de gobernanza y gestión de datos para desarrollar sistemas de IA de alto riesgo distintos de aquellos que utilizan técnicas que implican el entrenamiento de modelos, con vistas a garantizar que dichos sistemas de IA de alto riesgo cumplan lo dispuesto.”

## **Bibliografía**

- Adaya, E. y Pardo, I. (21 de Octubre del 2021). El Parlamento Europeo pone de nuevo el foco en el uso policial de la Inteligencia Artificial en su última resolución. Recuperado el 26 de Julio del 2022, de: [https://www.cuatrecasas.com/es/latam/articulo/parlamento\\_europeo\\_pone\\_foco\\_uso-policial\\_inteligencia\\_artificial\\_ultima\\_resolucion](https://www.cuatrecasas.com/es/latam/articulo/parlamento_europeo_pone_foco_uso-policial_inteligencia_artificial_ultima_resolucion)
- Affde (4 de Septiembre de 2021). affde. Pros y contras de Django Web Framework para el desarrollo de aplicaciones. Recuperado el 26 de Febrero de 2022, de affde: <https://www.affde.com/es/pros-and-cons-of-django-web-framework-for-app-development.html>
- Alarcon L., R.(25 de Julio de 2022) Plan de asignatura y secuencia didáctica, DESARROLLO WEB. 2022. <http://ddigital.umss.edu.bo:8080/jspui/handle/123456789/31902>
- Alava, C. F. C., y Díaz, A. H. (2022). Recursos para el desarrollo de la lectoescritura en estudiantes de la Educación General Básica. Aplicación de metodología exploratoria. Polo del Conocimiento: Revista científico-profesional, 7(4), 16. Recuperado el 2 de Agosto del 2022, de: <https://dialnet.unirioja.es/servlet/articulo?codigo=8482973>
- Alfonseca, E. y Rodriguez, P. (2012). Procesamiento de Lenguaje Natural Tema 4 Análisis Semántico y Pragmático. Artículo. Recuperado el 15 de Junio del 2022, de: <http://alfonseca.org/nlp/tema4.pdf>
- Alva, A. (2020). Inteligencia artificial: Camino a un nuevo esquema del mundo. Revistas. SCIENDO, 23(4), 299-308. Epub 2 Nov 2020. Recuperado el 1 de Marzo de 2022, de: <https://revistas.unitru.edu.pe/index.php/SCIENDO/article/view/3203/pdf>
- Amazon Machine Learning (2022). ¿Qué es Amazon Machine Learning?. Recuperado el 25 de Julio del 2022, de: [https://webcache.googleusercontent.com/search?q=cache:utWzaGGv3mYJ:https://docs.aws.amazon.com/es\\_es/machine-learning/latest/dg/what-is-amazon-machine-learning.html+&cd=12&hl=es-419&ct=clnk&gl=ec](https://webcache.googleusercontent.com/search?q=cache:utWzaGGv3mYJ:https://docs.aws.amazon.com/es_es/machine-learning/latest/dg/what-is-amazon-machine-learning.html+&cd=12&hl=es-419&ct=clnk&gl=ec)
- Ana Zita (26 de Febrero del 2021). significados. “Investigación científica”. Recuperado el 21 de marzo de 2022, en: <https://www.significados.com/investigacion-cientifica/>
- Anand, V., Bochkay, K., Chychyla, R., & Leone, A. (2020). Using python for text analysis in accounting research. Foundations and Trends in Accounting, 14(3–4), 128–359. Recuperado el 13 de Julio del 2022. Epub 3 Dic 2020, de nowpublishers: <https://doi.org/10.1561/14000000062>

- Ankush (21 de Marzo del 2021). geekflare. 9 mejores marcos de Python para crear aplicaciones pequeñas y empresariales. Recuperado el 26 de Febrero de 2022, de geekflare: <https://geekflare.com/es/best-python-frameworks/>
- Aprende Machine Learning, (2018). Procesamiento del Lenguaje Natural (NLP) | Aprende Machine Learning. Recuperado el 22 de Julio del 2022. Epub 27 de Diciembre del 2018, de: <https://www.aprendemachinelearning.com/procesamiento-del-lenguaje-natural-nlp/>
- Arias Gonzáles, J. L. (2021). Diseño y metodología de la investigación. Arequipa-Perú, Perú: Enfoques Consulting EIRL. Recuperado el 9 de Marzo de 2022, de: <http://repositorio.concytec.gob.pe/handle/20.500.12390/2260>
- Arias, E. R. (5 de Diciembre de 2020). Tipos de Investigación. *Economipedia*. (Economipedia, Editor) Recuperado el 22 de Marzo de 2022, de Economipedia: <https://economipedia.com/definiciones/tipos-de-investigacion.html>
- Ávila, C. E. (24 de Abril de 2019). Introducción a los tipos de muestreo. ALERTA Revista científica del Instituto Nacional de Salud. Recuperado el 8 de Agosto del 2022, de: <https://alerta.salud.gob.sv/introduccion-a-los-tipos-de-muestreo/>
- Ayala, M. (4 de Septiembre de 2020). Diseño de investigación: características, cómo se hace, ejemplo. *lifeder*. Recuperado el 21 de Marzo de 2022, de lifeder: <https://www.lifeder.com/diseno-de-investigacion/>
- Ayala, M. (4 de Septiembre del 2020). Diseño de investigación: características, cómo se hace, ejemplo. *lifeder*. Recuperado el 21 de Marzo de 2022, de lifeder: <https://www.lifeder.com/diseno-de-investigacion/>
- Ayala, Maite. (5 de Agosto de 2021). Investigación mixta. Lifeder. Recuperado el 9 de Agosto del 2021, de: <https://www.lifeder.com/investigacion-mixta/>
- Barbosa de Lira, T., Viana, F., Almeida, C., Amorim, F., y Viana, L. (2020). Desarrollo y evaluación de un prototipo de aplicación para cuidadores de ancianos. *Enfermería Global*, 19(59), 389-421. Recuperado el 2 de Junio del 2022, de: <https://dx.doi.org/10.6018/eglobal.396671>
- Battocchia, M. (2020). Preprocesamiento de texto para NLP. Recuperado el 22 de Julio del 2022. Epub 23 de Julio del 2020, de: <https://matiasbattocchia.github.io/datos/Preprocesamiento-de-texto-para-NLP-parte-1.html#El-pre-procesamiento-hasta-ahora>

- Bayas Ramírez, K. (2021). Comunicación y desinformación en tiempos de COVID-19 en Ecuador. URU, Revista De Comunicación Y Cultura, (4), 178–193. Recuperado el 2 de Julio del 2022, de: <https://doi.org/10.32719/26312514.2021.4.5>
- BIGDATA (2020). Algoritmos de reducción de la dimensionalidad con Python. Recuperado el 18 de Julio del 2022. Epub 9 de Julio del 2020, de: <https://topbigdata.es/6-algoritmos-de-reduccion-de-la-dimensionalidad-con-python/>
- Brasdefer, J. C. (2018). Pragmática del español: contexto, uso y variación. Routledge, de <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315443201&type=googlepdf>
- Camacho, M. y Navarro, E. (2020). Procesamiento del lenguaje natural con Python. Revista de Cómputo Aplicado, 4(13), 24-28, 10.35429/JCA.2020. Epub Mar 2020. Recuperado el 3 de Marzo del 2022. Obtenido de: [https://www.researchgate.net/publication/346152605\\_Procesamiento\\_del\\_lenguaje\\_natural\\_con\\_Python](https://www.researchgate.net/publication/346152605_Procesamiento_del_lenguaje_natural_con_Python)
- Camino Valle, O. D., Simón-Cuevas, A., Valladares-Valdés, E., Olivas, J. Á., & Romero, F. P. (2019). Generación de resúmenes extractivos de múltiples documentos usando grafos semánticos. Recuperado el 16 de Marzo del 2022, de: <https://rua.ua.es/dspace/handle/10045/96600>
- Casado, M. (2021). Una aproximación a la lingüística computacional. Epub 14 Nov 2021. e-ISSN: 1562-384X. DOI: 10.32870/sincronia.axxvi.n81.34\*22. 746-761. Recuperado el 16 Marzo de 2022, de: [http://sincronia.cucsh.udg.mx/pdf/81/746\\_761\\_2022a.pdf](http://sincronia.cucsh.udg.mx/pdf/81/746_761_2022a.pdf)
- Castro, E. M. (2019). Bioestadística aplicada en investigación clínica: conceptos básicos. Revista médica clínica las Condes, 30(1), 50-65. Recuperado el 8 de Agosto del 2022, de: <https://reader.elsevier.com/reader/sd/pii/S0716864019300045?token=03021F3193C79547D00308FB27920EC640AA20F3DD874FDB7D1BE114FFED70FF3863ABF3FB51719A4CBB6BF4389DC9D4&originRegion=us-east-1&originCreation=20220808162913>
- CENTRO EUROPEO DE POSTGRADO (2021). CEUPE. Epub 24 Ago 2021. Recuperado el 4 de Julio del 2022, de: <https://masterdatascienceceupe.com/machine-learning-ciclo-de-vida-y-sus-tipos/>
- Chhabra, Y. (2022). Slideteam. Recuperado el 4 de Julio del 2022. Epub 4 Junio 2022, de slideteam: <https://www.slideteam.net/machine-learning-life-cycle-for-model-data.html>



- Chocho, E., Mora R. (Marzo de 2022), Prototipo Web para la gestión de colecciones científicas de las plantas vasculares del herbario de la Facultad de Ciencias Naturales. 2022. Tesis de Licenciatura. Universidad de Guayaquil. Facultad de Ciencias Matemáticas y Físicas. Carrera de Ingeniería en Sistemas Computacionales, de: <http://repositorio.ug.edu.ec/handle/redug/59910>
- Chowdhary, K. R. (2020). Fundamentals of artificial intelligence (pp. 603-649). New Delhi: Springer India. Epub 5 Abr 2020. Recuperado el 29 de Junio del 2022, de: <https://link.springer.com/book/10.1007/978-81-322-3972-7?noAccess=true>
- Cohen, B. (2020). Python Tutorial – Set the pace, learn Python from the base. Recuperado el 4 de abril del 2022, de TechVidvan: <https://techvidvan.com/tutorials/python-tutorial/>
- Condori Ayala, J. L. (2012). Python-DjangoFramework de desarrollo web para perfeccionistas Basado en el Modelo MTV. Revista de información, tecnología y sociedad, 36. Recuperado el 15 de Junio del 2022, de: [http://www.revistasbolivianas.ciencia.bo/scielo.php?script=sci\\_arttext&pid=S1997-40442012000200016&lng=en&nrm=iso&tlng=es](http://www.revistasbolivianas.ciencia.bo/scielo.php?script=sci_arttext&pid=S1997-40442012000200016&lng=en&nrm=iso&tlng=es)
- Constitución de la República del Ecuador (2021). defensa. Constitución de la República del Ecuador 2008. Recuperado el 26 de Julio del 2022, de: [https://www.defensa.gob.ec/wp-content/uploads/downloads/2021/02/Constitucion-de-la-Republica-del-Ecuador\\_act\\_ene-2021.pdf](https://www.defensa.gob.ec/wp-content/uploads/downloads/2021/02/Constitucion-de-la-Republica-del-Ecuador_act_ene-2021.pdf)
- Cornago (2018), Qué es y cómo se trabaja la pragmática. El sonido de la hierba el crecer. Blog Académico. Recuperado el 18 de Marzo del 2022, de <https://elsonidodelahierbaelcrecer.blogspot.com/2018/07/que-es-y-como-trabajar-la-pragmatica.html#ixzz7NsbCgcQt>
- Cornago, A. (7 de Diciembre de 2018). ¿Qué es y cómo trabajar la pragmática? *elsonidodelahierbaelcrecer*. Recuperado el 2022 de Marzo de 18, de [elsonidodelahierbaelcrecer](https://elsonidodelahierbaelcrecer.blogspot.com/2018/07/que-es-y-como-trabajar-la-pragmatica.html#ixzz7NsbCgcQt): <https://elsonidodelahierbaelcrecer.blogspot.com/2018/07/que-es-y-como-trabajar-la-pragmatica.html#ixzz7NsbCgcQt>
- Cortés M. y Iglesias M. (2019). Studocu. Generalidades sobre la Metodología de la Investigación. Universidad Nacional de Huancavelica. Estrella RT. Recuperado el 1 de Agosto del 2022, de: <https://www.studocu.com/pe/document/universidad-nacional-de-huancavelica/analisis-matematico/libro-generalidades-sobre-metodologia-de-la-investigacion-manuel-e-cortes-cortes/20907428>

- Corvi, J. P. (2019). Resumen extractivo de documentos: Un análisis comparativo de técnicas de puntuación (Doctoral dissertation, Universidad Nacional de La Plata). Recuperado el 16 de Marzo del 2022, de: <https://oatd.org/oatd/record?record=oai:sedici.unlp.edu.ar:10915%2F81471&q=id%3A%22oai%3A%22sedici.unlp.edu.ar%3A10915%2F81471%22>
- Corvo, H. S. (31 de Marzo de 2020). C++: historia, características, tipos de datos, ejemplos. *lifeder*. Recuperado el 5 de Abril de 2022, de lifeder: <https://www.lifeder.com/c-mas-mas/>
- Corvo, S., H. (31 de marzo de 2020). C++: historia, características, tipos de datos, ejemplos. *Lifeder*. Recuperado el 24 de Julio del 2022, de lifeder: <https://www.lifeder.com/c-mas-mas/>.
- Cualitativa, 135–171. Recuperado el 11 de Agosto del 2022, de: <https://doi.org/10.2307/j.ctvdf06h7.9>
- Delgado, K., Gadea, W., y Vera, S. (2018). Rompiendo barreras en la investigación. Ediciones UTMACH. Recuperado el 2 de Agosto del 2022, de: <https://dialnet.unirioja.es/servlet/libro?codigo=778887>
- Delgado, L., S. (2021). Procesamiento de Lenguaje Natural sobre textos antiguos. Epub 8 Ago 2021. Recuperado el 16 Marzo de 2022, de: <https://riull.ull.es/xmlui/bitstream/handle/915/25424/Procesamiento%20de%20Lenguaje%20Natural%20sobre%20textos%20antiguos.pdf?sequence=1>
- Deloitte. (2022). Deloitte. La Inteligencia Artificial, clave para el futuro del IoT. *Deloitte*. Recuperado el 15 de Marzo de 2022, de deloitte: <https://www2.deloitte.com/es/es/pages/operations/articles/inteligencia-artificial-futuro-internet-of-things.html>
- Deloitte. (2022). Deloitte. Recuperado el 15 de Marzo del 2022, de: <https://www2.deloitte.com/es/es/pages/operations/articles/inteligencia-artificial-futuro-internet-of-things.html>
- Diakopoulos, N., Cass, S., & Romero, J. (2014). Data-driven rankings: the design and development of the IEEE Top Programming Languages news app. In Proceedings of the Symposium on Computation+ Journalism. Recuperado el 23 de Julio del 2022. Epub 2014, de aclanthology: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.699.9092&rep=rep1&type=pdf>

- Echeverri, M., y Manjarrés, R. (2020). Asistente virtual académico utilizando tecnologías cognitivas de procesamiento de lenguaje natural. Epub Jun 2020. Revista Politécnica, 16(31),85-96. ISSN: 1900-2351. Recuperado el 22 de Junio del 2022, de: <https://www.redalyc.org/articulo.oa?id=607863449007>
- EDTeam. (11 de Noviembre de 2020). La ruta del desarrollador web, de EDTeam: <https://ed.team/cursos/web>
- Elbarougy, R., Behery, G., & El Khatib, A. (2020). Extractive Arabic text summarization using modified PageRank algorithm. Egyptian informatics journal, 21(2), 73-81. Recuperado el 15 de Agosto del 2022, de: <https://www.sciencedirect.com/science/article/pii/S1110866519301355>
- Epitech (8 de Julio de 2021). epitech. Qué es Flask (Python) y cuáles son sus principales ventajas. Recuperado el 26 de Febrero de 2022, de epitech: <https://www.epitech-it.es/flask-python/>
- Espinoza Freire, Eudaldo Enrique. (2019). Las variables y su operacionalización en la investigación educativa. Segunda parte. Conrado, 15(69), 171-180. Epub 02 de Sep de 2019. Recuperado el 11 de Junio del 2022, de: [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1990-86442019000400171&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1990-86442019000400171&lng=es&tlng=es).
- Fernández, S. (2011). Análisis Factorial. Recuperado el 18 de Julio del 2022. Epub 2011, de libro: <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/GuiaSPSS/20factor.pdf>
- Flores, W. G. (2021). Programación Básica en Lenguaje C. Recuperado el 4 de abril del 2022, de Cinvestav: <https://www.tamps.cinvestav.mx/~wgomez/lenguajeC/notas.pdf>
- Flores, W. G. (Mayo de 2021). Programación Básica en Lenguaje C. *tamps*. Recuperado el 5 de Abril de 2022, de tamps: <https://www.tamps.cinvestav.mx/~wgomez/lenguajeC/notas.pdf>
- Fosado, G. (2019). NLP lo importante es empezar: ReGex en Python. Recuperado el 15 de Julio del 2022. Epub 14 Julio 2022, de medium: <https://yeralway1.medium.com/nlp-lo-importante-es-empezar-regex-en-python-efd34d82d66e>
- Fuente, S., Ó. (4 de Junio del 2019). Adictosaltrabajo. Google Colab: Python y Machine Learning en la nube. Recuperado el 25 de Julio del 2022, de: <https://www.adictosaltrabajo.com/2019/06/04/google-colab-python-y-machine-learning-en-la-nube/>

- Gómez J. (2021). Lenguaje Natural Definición de la tecnología. Elipse. Epub 4 Jun 2021. Recuperado 17 de Marzo de 2022, de: <https://elipse.ai/blog/definicion-de-la-tecnologia-lenguaje-natural/>
- Gómez, J. (4 de Junio de 2021). Lenguaje Natural: Definición de la tecnología. *elipse*. Recuperado el 17 de Marzo de 2022, de elipse: <https://elipse.ai/blog/definicion-de-la-tecnologia-lenguaje-natural/>
- Gómez-Luna, E., Fernando-Navas, D., Aponte-Mayor, G., & Betancourt-Buitrago, L. A. (2014). Metodología para la revisión bibliográfica y la gestión de información de temas científicos, a través de su estructuración y sistematización. *Dyna*, 81(184), 158-163. Recuperado el 1 de Agosto del 2022, de: <https://www.redalyc.org/pdf/496/49630405022.pdf>
- González, C. (15 de Octubre de 2020). Conoce 4 aplicaciones de la Inteligencia Artificial en la vida diaria. *Tecnológico de Monterrey*. Recuperado el 10 de Marzo de 2022, de *Tecnológico de Monterrey*: <https://tec.mx/es/noticias/guadalajara/investigacion/conoce-4-aplicaciones-de-la-inteligencia-artificial-en-la-vida>
- González, C. (2020). Conoce 4 aplicaciones de la Inteligencia Artificial en la vida diaria. Conecta. Epub 15 Oct 2020. Recuperado el 10 de Marzo del 2022, de: <https://conecta.tec.mx/es/noticias/guadalajara/investigacion/conoce-4-aplicaciones-de-la-inteligencia-artificial-en-la-vida>
- Grusky, M., Naaman, M., & Artzi, Y. (2018). Newsroom: A dataset of 1.3 million summaries with diverse extractive strategies. arXiv preprint arXiv:1804.11283. Recuperado el 23 de Julio del 2022. Epub 2018, de arxiv: <https://arxiv.org/pdf/1804.11283.pdf>
- Guamán Pullas, D. M., & Merino Córdova, S. R. (2020). bibdigital. Desarrollo de una aplicación web para la gestión del centro de especialidades médicas Vacarí. Recuperado el 10 de Agosto del 2022, de: <https://bibdigital.epn.edu.ec/handle/15000/20762>
- Hardt, M., Chen, X., Cheng, X., Donini, M., Gelman, J., Gollaprolu, S., He, J., Larroy, P., Liu, X., McCarthy, N., Rathi, A., Rees, S., Siva, A., Tsai, E., Vasist, K., Yilmaz, P., Bilal, M., Das, S., Haas, K., Hill, T. & Kenthapadi, K. (2021). Amazon sagemaker clarify: Machine learning bias detection and explainability in the cloud. arXiv preprint arXiv:2109.03285. Recuperado el 3 de Julio del 2022. Epub 7 Sep 2021, de Arxiv: <https://arxiv.org/pdf/2109.03285.pdf>

- Hofesmann, E. (2021). Towards Data Science. The Machine Learning Lifecycle in 2021. Recuperado el 4 de Julio del 2022. Epub 21 Junio 2021, de slideteam: <https://towardsdatascience.com/the-machine-learning-lifecycle-in-2021-473717c633bc>
- IBM (2020). IBM Watson Studio. Retrieved July 25, 2022 de: <https://www.ibm.com/pe-es/cloud/watson-studio>
- IBM, (2021). Tipos de variables. Recuperado el 22 de Julio del 2022. Epub 10 de Junio del 2021, de: <https://www.ibm.com/docs/es/spss-statistics/SaaS?topic=charts-variable-types>.
- Instituto Nacional de Estadísticas y Censos (INEC) (2010). Transmisión de información estadística en el Ecuador. Recuperado el 2 de Agosto del 2022, de: <https://www.ecuadorencifras.gob.ec/institucional/home/>
- IT USER TECH & BUSSINES (2020). RPA automatización en pro de agilidad corporativa. Revistas. It User (pp. 603-649). Recuperado el 30 de Junio del 2022. Epub Dic 2020, de: <https://www.ituser.es/whitepapers/content-download/9c9ba193-e5e1-4cb5-ab03-00fa7f4a5354/it-user-62.pdf>
- Jiménez, M. B., Pacheco, G. F., & García, A. C. (Abril de 2019). RESEARCHGATE. Obtenido de RESEARCHGATE: [https://www.researchgate.net/publication/340818034\\_Metodologia\\_mixta\\_estudios\\_de\\_caso](https://www.researchgate.net/publication/340818034_Metodologia_mixta_estudios_de_caso)
- Jiménez, M., Fernández, P., G. y Cuervo G., A. (2019). Metodología mixta; estudios de caso. Recuperado el 8 de Agosto del 2022, de: [https://www.researchgate.net/publication/340818034\\_Metodologia\\_mixta\\_estudios\\_de\\_caso](https://www.researchgate.net/publication/340818034_Metodologia_mixta_estudios_de_caso)
- Khder, M. A. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. International Journal of Advances in Soft Computing & Its Applications, 13(3). Epub Nov 2021. Recuperado el 23 de Junio del 2022, de: <http://www.i-csrs.org/Volumes/ijasca/2021.3.11.pdf>
- Krotov, V., Johnson, L., & Silva, L. (2020). Tutorial: Legality and ethics of web scraping. Communications of the Association for Information Systems, 47(1), 539–563, de <https://doi.org/10.17705/1CAIS.04724>
- Latinoamérica, E. (21 de Noviembre de 2019). ¿Qué es la Investigación Cualitativa? QuestionPro, de: <https://www.questionpro.com/es/investigacion-cualitativa.html>

- Ley Orgánica de Comunicación(2019).telecomunicaciones.LEY ORGANICA DE COMUNICACIÓN. Recuperado el 26 de Julio del 2022, de: <https://www.telecomunicaciones.gob.ec/wp-content/uploads/2020/01/Ley-Organica-de-Comunicaci%C3%B3n.pdf>
- Ley Orgánica de Telecomunicaciones (2016). telecomunicaciones. LEY ORGANICA DE TELECOMUNICACIONES. Recuperado el 26 de Julio del 2022, de: <https://www.telecomunicaciones.gob.ec/wp-content/uploads/downloads/2016/05/Ley-Org%C3%A1nica-de-Telecomunicaciones.pdf>
- Li, C., Xu, W., Li, S., & Gao, S. (2018, June). Guiding generation for abstractive text summarization based on key information guide network. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers) (pp. 55-60). Recuperado el 23 de Julio del 2022. Epub June 2018, de aclanthology: <https://aclanthology.org/N18-2009.pdf>
- Liu, P. J., Saleh, M., Pot, E., Goodrich, B., Sepassi, R., Kaiser, L., & Shazeer, N. (2018). Generating wikipedia by summarizing long sequences. arXiv preprint arXiv:1801.10198. Recuperado el 23 de Julio del 2022. Epub 30 de Junio del 2018, de arxiv: <https://arxiv.org/pdf/1801.10198.pdf>
- López, E. y Ortiz, A. (2019). Identificación de impactos en el fuselaje de un avión utilizando algoritmos de aprendizaje de máquinas. Recuperado el 18 de Julio del 2022. Epub 2019, de repositorio uchile: [https://repositorio.uchile.cl/bitstream/handle/2250/174430/cf-pinto\\_ma.pdf?sequence=1](https://repositorio.uchile.cl/bitstream/handle/2250/174430/cf-pinto_ma.pdf?sequence=1)
- López-Trujillo, S., & Torres-Madroño, M. C. (2021). Comparación de algoritmos de resumen de texto para el procesamiento de editoriales y noticias en español. TecnoLógicas, 24(51), 120-132. Recuperado el 15 de Agosto del 2022, de: <https://revistas.itm.edu.co/index.php/tecnologicas/article/view/1816>
- Lucas, J. (2019). Qué es C: Características y sintaxis. Publicado en 04 de septiembre de 2019. Recuperado el 4 de abril del 2022, de Openwebinars: <https://openwebinars.net/blog/que-es-c/>
- Lucas, J. (5 de Septiembre de 2019). Características de C. *openwebinars*. Recuperado el 5 de Abril de 2022, de openwebinars: <https://openwebinars.net/blog/que-es-c/>
- Mamani, F. (2019). Aprendizaje no supervisado: Reducción de dimensiones. Recuperado el 18 de Julio del 2022.Epub el 15 de Febrero del 2021, de artículo:

- <https://medium.com/analyst-ryma/aprendizaje-no-supervisado-reducci%C3%B3n-de-dimensiones-67d3aa2385f3>
- Marín, M. E. G. (2019). Técnicas de Investigación. Estrategias de Investigación Social
- Marquinez Burgos, David (2021). Análisis de la polarización del discurso de comunidades en redes sociales y blogs. Proyecto Fin de Carrera / Trabajo Fin de Grado, E.T.S.I. de Sistemas Informáticos (UPM), Madrid, de: <https://oa.upm.es/68317/>
- Martínez Martínez, L. (2021). Comparativa de modelos de Machine Learning para la estimación de parámetros de interés empleando datos de la European Soil Database. Recuperado 5 de Julio del 2022. Epub 28 Septiembre 2021, de Universidad de Valladolid: <https://uvadoc.uva.es/handle/10324/50055>
- Martínez, A. (2021). Estudio comparativo de las mejoras del lenguaje de programación Kotlin y el lenguaje Java en el desarrollo de aplicaciones Android. Epub Nov 2021. Recuperado el 5 de abril del 2022, desde dspace: <http://dspace.utb.edu.ec/bitstream/handle/49000/10535/E-UTB-FAFI-SIST-000244.pdf?sequence=1&isAllowed=y>.
- Martínez, D. (Noviembre de 2021). Estudio comparativo de las mejoras del lenguaje de programación Kotlin y el lenguaje Java en el desarrollo de aplicaciones android. *dspace*. Recuperado el 5 de Abril de 2022, de dspace: <http://dspace.utb.edu.ec/bitstream/handle/49000/10535/E-UTB-FAFI-SIST-000244.pdf?sequence=1&isAllowed=y>
- McKenzie, G. D., & Adams, B. (2021). Natural Language Processing in GIScience Applications. The Geographic Information Science & Technology Body of Knowledge. Data iku. (pp. 1-16). Canadá. Epub 18 Nov 2021. Recuperado el 1 de Julio de 2022, de: <https://content.dataiku.com/nlp-basics>
- Messina, V., C. (2018). Librería de métodos de poda en conjuntos de clasificadores para Scikit-Learn. UAM\_Biblioteca, 21-71. Recuperado el 15 de Julio del 2022. Epub 2018, de repositorio: <https://repositorio.uam.es/handle/10486/68829>
- Microsoft (19 de Julio del 2022). Microsoft. ¿Qué es Azure Machine Learning? Recuperado el 25 de Julio del 2022, de: <https://docs.microsoft.com/es-es/azure/machine-learning/overview-what-is-azure-machine-learning>
- Molina, G. N. (2018). Infraestructura para la evaluación intrínseca de algoritmos de Stemming. Repositorio de la Universidad de Matanzas Cuba, 25-70. Recuperado el 22 de Julio del 2022. Epub 2018, de: <http://cict.umcc.cu/repositorio/tesis/Trabajos%20de%20Diploma/Ingenier%C3%A>

- Da%20Inform%C3%A1tica/2018/Infraestructura%20para%20la%20evaluaci%C3%B3n%20intr%C3%ADnseca%20de%20algoritmos%20de%20stemming%20(Noe1%20Molina%20Gonz%C3%A1lez).pdf
- Moreira, D., Cruz, I., Gonzalez, K., Quirumbay, A., Magallan, C., Teresa, G., Andrade, A., Castillo, C. (2020). Análisis del Estado Actual de Procesamiento de Lenguaje Natural. Proquest(E42), 126-136. Epub 7 Dic 2020. Recuperado el 2 de Marzo de 2022, de: <https://www.proquest.com/openview/a44d67c88cfaada2563dc16f94ccd3c8/1?pq-origsite=gscholar&cbl=1006393>
- Moreno, F., S. (2020). Herramienta de Reconocimiento de Imágenes en Python. Trabajo Fin de Grado Inédito, Universidad de Sevilla, Sevilla., 24-88. Recuperado el 13 de Julio del 2022. Epub 2020, de idus: <https://idus.us.es/bitstream/handle/11441/102090/TFG-2877-MORENO%20FERNANDEZ.pdf?sequence=1&isAllowed=y>
- National Geographic. (02 de Diciembre de 2020). Breve historia visual de la inteligencia artificial. *nationalgeographic*, de *nationalgeographic*: [https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial\\_14419/5](https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial_14419/5)
- National Geographic. (2020). Breve historia visual de la inteligencia artificial. Epub 2 Dic 2020. Recuperado el 1 de Marzo de 2022. National Geographic, de: [https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial\\_14419/5](https://www.nationalgeographic.com.es/ciencia/breve-historia-visual-inteligencia-artificial_14419/5)
- Noticias Parlamento Europeo (2021). ¿Qué es la inteligencia artificial y cómo se usa? Epub 26 Mar 2021. Recuperado el 25 de Febrero del 2022, de: <https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa>
- Ossa, G. y Jaramillo, V. (2021). Machine Learning para la estimación del riesgo de crédito en una cartera de consumo (Doctoral dissertation, Universidad EAFIT). Recuperado el 5 de Julio del 2022. Epub 2021, de Universidad EAFIT: [https://repository.eafit.edu.co/bitstream/handle/10784/29589/Wbeimar\\_OssaGiraldo\\_Veronica\\_JaramilloMarin\\_2021.pdf?sequence=2&isAllowed=y](https://repository.eafit.edu.co/bitstream/handle/10784/29589/Wbeimar_OssaGiraldo_Veronica_JaramilloMarin_2021.pdf?sequence=2&isAllowed=y)
- P. Jayasuriya, R. Munasinghe and S. Thelijagoda, “Sentiment Classification of Sinhala Content in Social Media: A Comparison between Stemmers and N-gram Features,” 2021 IEEE 16th International Conference on Industrial and Information Systems (ICIIS), 2021, pp. 134-139, doi: 10.1109/ICIIS53135.2021.9660711. Recuperado el 22



- de Julio del 2022. Epub 2021 de:  
<https://ieeexplore.ieee.org/abstract/document/9660711>
- ParadaVisual (2019). IoT e Inteligencia Artificial para la industria hotelera. Epub 21 Dic 2019. Recuperado el 15 de Marzo del 2022, de: <https://www.paradavisual.com/lot-e-inteligencia-artificial-para-industria-hotelera/>
- Paradavisual. (21 de Diciembre de 2019). IoT e Inteligencia Artificial para la Industria Hotelera. *paradavisual*. Recuperado el 15 de Marzo de 2022, de [paradavisual: https://www.paradavisual.com/lot-e-inteligencia-artificial-para-industria-hotelera/](https://www.paradavisual.com/lot-e-inteligencia-artificial-para-industria-hotelera/)
- Parlamento Europeo. (26 de Marzo de 2021). ¿Qué es la Inteligencia Artificial y como se usa? *europarl*. Recuperado el 25 de Febrero de 2022, de [europarl: https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa](https://www.europarl.europa.eu/news/es/headlines/society/20200827STO85804/que-es-la-inteligencia-artificial-y-como-se-usa)
- Pérez-Montoro, M., & Codina, L. (2010). Software de prototipado para la arquitectura de la información: funcionalidad y evaluación. *Profesional De La información*, 19(4), 417–424. Recuperado el 10 de Agosto del 2022, de: <https://doi.org/10.3145/epi.2010.jul.12>
- Piovesan, A., & Temporini, E. R. (1995). Pesquisa exploratória: procedimento metodológico para o estudo de fatores humanos no campo da saúde pública. *Revista de saúde pública*, 29, 318-325. Recuperado el 2 de Agosto del 2022, de: <https://www.scielo.br/j/rsp/a/ff44L9rmXt8PVYLNvphJgTd/?fo>
- Pissinati, P., Martinez, Y., Rossaneis, M., Gvozdz, R., Souza, M. y Fernandez, M. (2019). Desarrollo de un prototipo de web software de apoyo a la planificación de la jubilación. Presentada en la Universidad de Estadual de Maringá, Maringá, PR, Brasil. *Revista Latino-Americana de Enfermagem* [online]. 2019, v. 27. Epub 19 Ago 2019. ISSN 1518-8345. Recuperado el 15 de Junio del 2022, de: <https://doi.org/10.1590/1518-8345.3024.3169>
- Piza Burgos, N. D., Amaiquema Márquez, F. A., & Beltrán Baquerizo, G. E. (2019). Métodos y técnicas en la investigación cualitativa. Algunas precisiones necesarias. *Conrado*, 15(70), 455-459. Recuperado el 2 de Agosto del 2022, de: <http://scielo.sld.cu/pdf/rc/v15n70/1990-8644-rc-15-70-455.pdf>
- Prasad, D. (5 de Noviembre de 2021). Geekflare. Los 11 mejores IDE de Python para potenciar el desarrollo y la depuración. Recuperado el 9 de Abril de 2022, de [geekflare: https://geekflare.com/es/best-python-ide/#:~:text=aumentar%20su%20productividad.-,PyCharm,Python%20en%20un%20solo%20lugar](https://geekflare.com/es/best-python-ide/#:~:text=aumentar%20su%20productividad.-,PyCharm,Python%20en%20un%20solo%20lugar)

- Prasad, D. (5 de Noviembre de 2021). Mejores IDE de Python para potenciar el desarrollo y la depuración. *geekflare*. Recuperado el 9 de Abril de 2022, de *geekflare*: <https://geekflare.com/es/best-python-ide/#:~:text=aumentar%20su%20productividad.-,PyCharm,Python%20en%20un%20solo%20lugar.>
- Presta, M. (13 de Julio del 2022). Googleusercontent. Los 10 mejores marcos de trabajo de Python. Recuperado el 26 de Febrero de 2022, de *googleusercontent*: [https://webcache.googleusercontent.com/search?q=cache:wmvd6h\\_ktT8J:https://blog.back4app.com/es/los-10-mejores-marcos-de-trabajo-de-python/+&cd=15&hl=es&ct=clnk&gl=ec&client=firefox-b-d](https://webcache.googleusercontent.com/search?q=cache:wmvd6h_ktT8J:https://blog.back4app.com/es/los-10-mejores-marcos-de-trabajo-de-python/+&cd=15&hl=es&ct=clnk&gl=ec&client=firefox-b-d)
- Python (2022). General Python FAQ. Recuperado el 4 de Abril del 2022, de Python: <https://docs.python.org/3/faq/general.html#what-is-python>
- Rahman, A. (2022). How to Work With JSON Data Using Python. Recuperado el 15 de Julio del 2022. Epub 30 Junio 2022, de *tutsplus*: [https://programacion.net/articulo/como\\_trabajar\\_con\\_datos\\_json\\_utilizando\\_python\\_1403](https://programacion.net/articulo/como_trabajar_con_datos_json_utilizando_python_1403)
- Raschka, S., Patterson, J., Nolet, C. (2020). Machine Learning in Python: Main Developments and Technology Trends in Data Science, Machine Learning, and Artificial Intelligence. *Information* 2020, 11, 193. Epub 4 Abr 2020. Recuperado el 2 de Julio de 2022, de: <https://doi.org/10.3390/info11040193>
- Redacción APD. (2021). Métodos y técnicas de inteligencia artificial. Redacción APD. Epub 7 Ene 2021. Recuperado el 25 de Febrero de 2022, de: <https://www.apd.es/tecnicas-de-la-inteligencia-artificial-cuales-son-y-para-que-se-utilizan/>
- Redacción APD. (7 de Enero de 2021). Métodos y técnicas de inteligencia artificial. *REDACCIÓN APD*. Recuperado el 25 de Febrero de 2022, de *apd*: <https://www.apd.es/tecnicas-de-la-inteligencia-artificial-cuales-son-y-para-que-se-utilizan/>
- Redacción España (2019). Origen del concepto de Inteligencia Artificial. Epub 11 Nov 2019. Recuperado el 1 de Marzo de 2022, de: <https://agenciab12.com/noticia/origen-concepto-inteligencia-artificial>
- Redacción España. (11 de Noviembre de 2019). Origen del concepto de Inteligencia Artificial. *agenciab12*. Recuperado el 2 de Marzo de 2022, de *agenciab12*: <https://agenciab12.com/noticia/origen-concepto-inteligencia-artificial>
- Reglamento de Régimen Académico Consejo de Educación Superior (2017). *ces*. REGLAMENTO DE REGIMEN ACADEMICO CONSEJO EDUCACION

- SUPERIOR. Recuperado el 26 de Julio del 2022, de:  
<https://www.ces.gob.ec/lotaip/2017/Diciembre/Anexos%20Procu/An-lit-a2-Reglamento%20de%20R%C3%A9gimen%20Acad%C3%A9mico.pdf>
- Reglamento del Parlamento Europeo y del Consejo (2021). Europa. REGLAMENTO DEL PARLAMENTO EUROPEO Y DEL CONSEJO. Recuperado el 26 de Julio del 2022, de:  
<https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>
- REVISTA ESPAÑOLA DE ELECTRÓNICA (17 de Enero de 2022). ¿Cómo funciona internet?.  
<https://www.redeweb.com/actualidad/como-funciona-internet/#:~:text=Definimos%20como%20internet%20al%20conjunto,conexiones%20de%20fibra%20%C3%B3ptica%2C%20cable%E2%80%A6>
- Ridao, S. Fundamentos de pragmática. ed. Albolote (Granada): Editorial Comares, 2020. 1-113. Recuperado el 28 de Junio del 2022, de:  
<https://elibro.net/es/ereader/uguayaquil/158245>
- Ritesh, R. (2021). What is a Framework in Programming & Why You Should Use One. Retrieved July 24, 2022, de netsolutions: <https://www.netsolutions.com/insights/what-is-a-framework-in-programming/>
- Rodríguez, X. (21 de Agosto de 2019). openwebinars. Django vs Flask. Recuperado el 26 de Febrero de 2022, de openwebinars: <https://openwebinars.net/blog/django-vs-flask/>
- Roman, V. (2019). Aprendizaje No Supervisado en Machine Learning: Agrupación. Recuperado el 18 de Julio del 2022. Epub el 12 de Julio del 2019, de artículo: <https://medium.com/datos-y-ciencia/aprendizaje-no-supervisado-en-machine-learning-agrupaci%C3%B3n-bb8f25813edc>
- Romero, A. (2021). Procesamiento del Lenguaje Natural para el Apoyo en el Diagnóstico de Tuberculosis. Recuperado el 22 de Julio del 2022. Epub 2021, de:  
<https://repository.urosario.edu.co/handle/10336/31562>
- Samrat K. (2017). AI Mind Map. Machine Learning And Artificial Intelligence Study Group. Epub 22 Mar 2017. Recuperado el 30 de Junio del 2022, de:  
<https://medium.com/ml-ai-study-group/ai-mind-map-a70dafcf5a48>
- Sánchez Molina, A. A., & Murillo Garza, A. (2021). Enfoques metodológicos en la investigación histórica: cuantitativa, cualitativa y comparativa. Debates por la Historia, 9(2), 147-181. Recuperado el 2 de Agosto del 2022, de:  
<https://vocero.uach.mx/index.php/debates-por-la-historia/article/view/792/993>

- Sánchez, D. y Gereá, C. (12 de Marzo del 2021). Recuperado el 9 de Agosto del 2022, de: <https://freed.tools/blogs/ux-cx/prototipo#:~:text=cumple%20o%20no.-,Metodolog%C3%ADa%20de%20prototipo%20o%20prototipado,medir%20y%20ajustar%20un%20plan>
- Sandoval (25 de Julio del 2022). webcache. Google Colaboratory. Recuperado el 25 de Julio del 2022, de: <https://webcache.googleusercontent.com/search?q=cache:dYGT8-BwcGsJ:https://internetpasoapaso.com/google-colaboratory/+&cd=12&hl=es-419&ct=clnk&gl=ec>
- Sarabia, J. (Junio de 2021). Sangabrielriobamba. Diseño e implementación de una aplicación web para el seguimiento y publicación de proyectos del gadpr comput perteneciente al cantón chunchi, desarrollado en php con framework laravel y gestor de datos mysql. Recuperado el 11 de Abril de 2022, de sangabrielriobamba: <https://www.sangabrielriobamba.edu.ec/tesis/sistemas/tesis26.pdf>
- Sarabia, J. (Junio de 2021). *sangabrielriobamba*. Recuperado el 11 de Abril de 2022, de sangabrielriobamba: <https://www.sangabrielriobamba.edu.ec/tesis/sistemas/tesis26.pdf>
- Sarraipa, J., Artíficie, A. y Jiménez, H. (2019). Metodología De Evaluación De Prototipo Innovador. Recuperado el 10 de Agosto del 2022, de: <https://acacia.red/wp-content/uploads/2019/07/Gu%C3%ADa-Metodologi%C3%A1-de-evaluaci%C3%B3n-de-prototipo-innovador.pdf>
- Sharma, A. (March 23, 2020). Analyticsvidhya. Free GPUs for Everyone! Get Started with Google Colab for Machine Learning and Deep Learning. Retrieved July 25, 2022, de: <https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning/>
- Silva, C., Ortiz, A., Serrano, J., y Rubio, J. M. (2021). Experiencia académica en desarrollo rápido de sistemas de información web con Python y Django. Formación universitaria, 14(5), 85-94. Epub Oct 2021. Recuperado el 16 de Junio del 2022, de: <https://scielo.conicyt.cl/pdf/formuniv/v14n5/0718-5006-formuniv-14-05-85.pdf>
- Soliz, D. J. S. (2019). Cómo hacer un perfil proyecto de investigación científica. Palibrio. Recuperado el 10 de Agosto del 2022, de: <https://books.google.es/books?hl=es&lr=&id=Q-GCDwAAQBAJ&oi=fnd&pg=PT6&dq=ciclo+del+proceso+de+la+investigaci%C3%B3n+cientifica+&ots=bJ-YHcxfoX&sig=MHb04ecq1G5EO5qRIK0hhSoFnOk#v=onepage&q&f=false>

- Talamé, L., Cardoso, A., y Amor, M. (2019). Comparación de herramientas de procesamiento de textos en español extraídos de una red social para Python. In XX Simposio Argentino de Inteligencia Artificial (ASAI 2019)-JAIIO 48 (Salta), de: [https://repositoriosdigitales.mincyt.gob.ar/vufind/Record/SEDICI\\_f270c502d594ef82dad2d2b07c448c65](https://repositoriosdigitales.mincyt.gob.ar/vufind/Record/SEDICI_f270c502d594ef82dad2d2b07c448c65)
- Thomas, D., & Mathur, S. (2019). Data Analysis by Web Scraping using Python. Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019, 450–454. <https://doi.org/10.1109/ICECA.2019.8822022>
- Tintinago, A., Muñoz, Y., Uribe, G. y Álvarez, P. (2018). Etiquetado asistido de documentos de investigación mediante procesamiento de lenguaje natural y tecnologías de la web semántica. *Scientia Et Technica*, 23(4), 528-537. ISSN: 0122-1701. Recuperado el 2 de Junio del 2022, de: <https://www.redalyc.org/articulo.oa?id=84959055012>
- TIOBE. (2022). Tiobe. Recuperado el 29 de Marzo de 2022, de Tiobe: <https://www.tiobe.com/tiobe-index/>
- Torres, M., Salazar, F. G., y Paz, K. (2019). Métodos de recolección de datos para una investigación. Recuperado el 10 de Agosto del 2022, de: [https://fgsalazar.net/LANDIVAR/ING-PRIMERO/boletin03/URL\\_03\\_BAS01.pdf](https://fgsalazar.net/LANDIVAR/ING-PRIMERO/boletin03/URL_03_BAS01.pdf)
- UNIR. (6 de Julio de 2021). ¿Qué es un IDE en programación?. *Unir*. Recuperado el 7 de Abril de 2022, de Unir: <https://www.unir.net/ingenieria/revista/ide-programacion/>
- Universidad de Antioquia (2020). Youtube. Ambientes de desarrollo para python. Recuperado el: 7 de abril del 2022, de: <https://www.youtube.com/watch?v=hXkZqxe2HE8&t=117s>.
- Universidad De La República Uruguay (2020). fenf. ETAPAS DE LA INVESTIGACIÓN BIBLIOGRÁFICA. Recuperado el 23 de Marzo del 2022, de: <https://www.fenf.edu.uy/wp-content/uploads/2020/12/14dediciembrede2020Etapasdela-investigacionbibliografica-1.pdf>
- Urrutia, G. G. (2020). Dspace. Estudio de los beneficios del uso de chatbots en los procesos de atención a usuarios del sistema académico integral de la universidad técnica de Babahoyo. Epub 2020. Recuperado el 15 de Marzo de 2022, de: <http://dspace.utb.edu.ec/bitstream/handle/49000/7708/URRUTIA%20ORTIZ.pdf?sequence=1&isAllowed=y>

- Verganti, R. V. (19 de Marzo de 2020). Innovation and Design in the Age of Artificial Intelligence. *onlinelibrary*. Recuperado el 1 de Marzo de 2020, de onlinelibrary: <https://onlinelibrary.wiley.com/doi/full/10.1111/jpim.12523>
- Verganti, R., Vendraminelli, L. & Iansiti, M. (2020). Innovation and Design in the Age of Artificial Intelligence. *Onlinelibrary. Journal of Product Innovation Management*, 37(3), 212-227. Epub 19 Mar 2020. Recuperado el 1 de Marzo de 2022, de: <https://onlinelibrary.wiley.com/doi/full/10.1111/jpim.12523>
- Victoria O. Rytikova and Maxim A. Medvedev. Statistical data and programming languages analysis for developing a chatbot in telegram. *AIP Conference Proceedings* 2333, 100005 (2021). Recuperado el 23 de Julio del 2022, de: <https://doi.org/10.1063/5.0041762>
- Web Design (17 de Julio del 2022). *webdesigncusco*. Principales Características de Sublime Text 3. Recuperado el 12 de Abril del 2022, de *webdesigncusco*: <https://webdesigncusco.com/principales-caracteristicas-de-sublime-text-3/>
- Yao, J. (2019, April). Automated sentiment analysis of text data with NLTK. In *Journal of Physics: Conference Series* (Vol. 1187, No. 5, p. 052020). IOP Publishing. Recuperado el 15 de Julio del 2022. Epub 2019, de *iopscience*: <https://iopscience.iop.org/article/10.1088/1742-6596/1187/5/052020/pdf>
- Yucra D. A. (2021) Lenguaje de programación estructurado. Epub Nov 2021. Recuperado el 5 de abril del 2022, de *Dspace*: <http://dspace.utb.edu.ec/bitstream/handle/49000/10535/E-UTB-FAFI-SIST-000244.pdf?sequence=1&isAllowed=y>
- Zepeda, E. (24 de Marzo de 2021). *coffeebytes*. ¿Por qué deberías usar Django Framework?. Recuperado el 26 de Febrero de 2022, de *coffeebytes*: <https://coffeebytes.dev/por-que-deberias-usar-django-framework/>
- Zhao, L., Alhoshan, W., Ferrari, A., Letsholo, K., Ajagbe, M., Chioasca, E., & Riza, B. (2020). Natural Language Processing (NLP) for Requirements Engineering: A Systematic Mapping Study. 5. Recuperado el 1 de Julio de 2022, de: <https://arxiv.org/ftp/arxiv/papers/2004/2004.01099.pdf>
- Zita, A. (26 de Febrero de 2021). ¿Qué es la investigación científica?. *significados*. Recuperado el 2022 de Marzo de 21, de *significados*: <https://www.significados.com/investigacion-cientifica/>