

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL

Facultad de Ciencias Naturales y Matemáticas

Aplicación de un modelo *Deep Learning* para la lectura automática de caracteres manuscritos en formularios estadísticos físicos, usando la arquitectura de redes neuronales Transformers

PROYECTO INTEGRADOR

Previo la obtención del Título de:

INGENIERO EN ESTADÍSTICA

Presentado por:

Benjamín Castillo Zurita

Karla Andrea Silva Tacuri

GUAYAQUIL - ECUADOR

Año: 2023

DEDICATORIA

El presente proyecto lo dedico a mi hermana Priscila, quien ha sido no solo una compañera de vida sino también una fuente constante de inspiración y apoyo. En nuestros momentos de risa, nuestras conversaciones profundas y, sí, también en nuestras pequeñas disputas, he encontrado la motivación y el coraje para continuar adelante.

Priscila, quiero que veas esta tesis no solo como un testimonio de mi esfuerzo y dedicación, sino también como una muestra tangible de lo que es posible cuando perseguimos nuestros sueños con determinación y pasión.

Espero que esta tesis te motive, como tú me has motivado a mí en tantas ocasiones. Eres fuerte, inteligente y capaz, y estoy convencido de que cualquier meta que te propongas la alcanzarás. Esta no es solo una dedicatoria, es también una promesa: siempre estaré aquí para apoyarte, celebrar tus éxitos y alentarte en los desafíos.

Con mucho amor,

Benjamín Castillo Zurita

DEDICATORIA

El presente proyecto lo dedico a Emiliano, quien me ha acompañado en cada paso, cada investigación y cada noche de trabajo en esta tesis.

Cada desafío enfrentado en este proceso tuvo su contraparte en las sonrisas que compartimos, en los abrazos que reconfortaron mis miedos y en los momentos preciosos que hemos construido juntos. Quiero que crezcas sabiendo que, en cada página de este trabajo, hay un deseo latente: que sirva como una brújula para ti, mostrándote que, con pasión, constancia y amor, los sueños se pueden alcanzar.

Hijo mío, esta tesis no es solo el culmen de mis años universitarios; es también una promesa hacia ti. Es mi forma de decirte que siempre lucharé por ser un mejor ejemplo, por enseñarte a perseguir tus pasiones y por asegurarme de que siempre tengas un motivo para sentirte orgulloso.

Con todo mi amor,

Karla Silva Tacuri

AGRADECIMIENTOS

Mi más sincero y profundo agradecimiento a mi mamá, cuyo amor, consejos y apoyo inquebrantable han sido el faro que iluminó cada paso de esta travesía académica. Su fortaleza y determinación son el ejemplo que he buscado emular en mi vida y durante este proceso.

Agradezco también a mi papá, Baggner, por sus enseñanzas y por estar presente en distintos momentos cruciales de mi camino. A Priscila y Génesis, mis hermanas, por siempre estar allí, compartiendo risas y palabras de aliento. A Michael, por sus consejos y su compañía. A Jordan y Tamara, con quienes inicié esta etapa universitaria, gracias por los momentos de estudio y las discusiones enriquecedoras. A Lionel Messi, por mostrarme con su talento y dedicación lo que significa ser el mejor en lo que uno hace. Y finalmente, un reconocimiento especial a mi enamorada Melissa, por ser una fuente constante de apoyo y compartir conmigo incontables momentos que hicieron más llevadero el proceso de este proyecto.

Benjamín Castillo Zurita

AGRADECIMIENTOS

Mi más profundo agradecimiento a Dios, fuente inagotable de amor y sabiduría, por haberme dado la fortaleza, la salud y la perseverancia para culminar este significativo capítulo de mi vida. A mis amados padres, Carlos y Maritza, quienes desde mi infancia han sido mi base sólida; su amor incondicional, sus enseñanzas y el inquebrantable apoyo que me han brindado, son las razones por las cuales hoy puedo presentar este trabajo. A mis queridas hermanas, Ale y Gaby, cuyos consejos, risas y momentos compartidos han sido el bálsamo en los días más estresantes y difíciles. Agradezco la comprensión y paciencia que tuvieron conmigo en cada etapa de este proceso. Finalmente, a mi amado esposo, Bryan, que ha sido mi compañero en este viaje. Tu amor, paciencia y constante apoyo han sido esenciales para la realización de este proyecto. Gracias por ser mi confidente, por celebrar cada pequeño logro y por estar a mi lado en cada desafío.

Karla Silva Tacuri

DECLARACIÓN EXPRESA

"Los derechos de titularidad y explotación, nos corresponde conforme al reglamento de propiedad intelectual de la institución; *Benjamín Israel Castillo Zurita* y *Karla Andrea Silva Tacuri* damos nuestro consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual"

Benjamín Castillo Z.
Benjamín Castillo Zurita

Karla Silva T.
Karla Silva Tacuri

EVALUADORES



Firmado electrónicamente por:
SANDRA LORENA
GARCIA BUSTOS

Ph.D. Sandra Garcia Bustos

PROFESOR DE LA MATERIA



Firmado electrónicamente por:
JESSICA ALEXANDRA
MENENDEZ CAMPOS

M.Sc. Jessica Menéndez Campos

PROFESOR TUTOR

RESUMEN

El proyecto busca abordar el desafío de procesar y reconocer caracteres en formularios manuscritos provenientes de hospitales. Dada la variabilidad de la escritura de cada individuo, la interpretación manual de estos formularios es complicada y susceptible a errores. Se propuso el uso de un modelo de inteligencia artificial, Transformers, para reconocer y clasificar estos caracteres, con el objetivo de transformar datos manuscritos en texto digital, facilitando su procesamiento y almacenamiento en bases de datos. El estudio adquiere relevancia por su vinculación con el Objetivo de Desarrollo Sostenible (ODS) 3, pues la digitalización precisa de registros médicos puede mejorar la gestión y análisis de datos en el sector salud, reduciendo errores humanos y optimizando la atención médica. El conjunto de datos utilizado consta de 4999 formularios manuscritos recopilados de varias regiones de la ZONAL 8. Se empleó el modelo TrOCR para el reconocimiento de texto. En términos de resultados, se recopilaron formularios de siete zonas distintas, con Guayas y Manabí aportando la mayor cantidad. El proceso de entrenamiento del modelo incluyó diversos parámetros y técnicas. La pérdida de prueba alcanzada fue del 34,42%, y el Error de Tasa de Caracteres (CER) fue del 6,91%. En conclusión, se logra un considerable avance hacia la digitalización precisa de textos manuscritos. Aunque la precisión alcanzada es destacable, aún hay margen de mejora, especialmente en la reducción del CER. Este trabajo tiene el potencial de transformar la forma en que se procesan y almacenan datos manuscritos en el sector salud.

Palabras Clave: Transformers, Reconocimiento manuscrito, Digitalización, Formularios médicos, Objetivo de Desarrollo Sostenible (ODS).

ABSTRACT

The project aims to address the challenge of processing and recognizing characters in handwritten forms. Given the variability in each individual's writing, the manual interpretation of these forms is complex and prone to errors. The use of an artificial intelligence model, Transformers, was proposed to recognize and classify these characters, with the goal of transforming handwritten data into digital text, facilitating its processing and storage in databases. The study gains significance due to its connection with Sustainable Development Goal (SDG) 3, as precise digitization of medical records can enhance data management and analysis in the health sector, reducing human errors and optimizing medical care. The dataset used consists of 4999 handwritten forms collected from various regions of ZONAL 8. The TrOCR model was employed for text recognition. In terms of results, forms were collected from seven distinct zones, with Guayas and Manabí contributing the most. The training process of the model included various parameters and techniques. The test loss achieved was 34.42%, and the Character Error Rate (CER) was 6.91%. In conclusion, significant progress has been made towards the precise digitization of handwritten texts. Although the accuracy achieved is noteworthy, there is still room for improvement, especially in reducing the CER. This work has the potential to transform the way handwritten data is processed and stored in the health sector.

Keywords: Transformers, Handwritten recognition, Digitization, Medical forms, Sustainable Development Goal (SDG).

ÍNDICE GENERAL

EVALUADORES.....	7
RESUMEN.....	I
<i>ABSTRACT</i>	II
ÍNDICE GENERAL.....	III
ABREVIATURAS	VI
ÍNDICE DE FIGURAS.....	VII
ÍNDICE DE TABLAS	VIII
CAPÍTULO 1	1
1. Introducción	1
1.1 Descripción del problema	4
1.2 Justificación del problema.....	5
1.3 Objetivos.....	6
1.3.1 Objetivo General	6
1.3.2 Objetivos Específicos	7
1.4 Marco teórico	7
1.4.1 Inteligencia artificial.....	7
1.4.2 Aprendizaje Automático	8
1.4.3 Aprendizaje Profundo.....	10
1.4.4 Redes Neuronales.....	10
1.4.5 Neurona o Unidad	13
1.5 Estado del Arte	15
CAPÍTULO 2.....	19
2. Metodología	19
2.1 Preparación de los datos	23

2.2	Entrenamiento del modelo	25
CAPÍTULO 3.....		27
3.	Resultados y Análisis	27
3.1	Estadística descriptiva	27
3.2	Parámetros de entrenamiento del modelo	29
3.3	Resultados del proceso de entrenamiento.....	30
3.4	Predicción del modelo.....	31
3.5	Resultados.....	32
3.6	Comparación de tiempo	35
CAPÍTULO 4.....		36
4.	Conclusiones Y Recomendaciones.....	37
4.1	Conclusiones	37
4.1.1	Importancia del trabajo desarrollado y relación con el Objetivo General	37
4.1.2	Fortalezas en relación con los Objetivos Específicos.....	37
4.1.3	Debilidades	38
4.1.4	Implicaciones del trabajo.....	38
4.1.5	Trabajos futuros	38
4.2	Recomendaciones	39
4.2.1	Expansión del conjunto de datos	39
4.2.2	Preprocesamiento avanzado.....	39
4.2.3	Optimización de hiperparámetros.....	39
4.2.4	Evaluación cruzada	40
4.2.5	Modelos comparativos	40
4.2.6	Post-procesamiento	40
4.2.7	Consideraciones éticas	40
4.2.8	Análisis detallado de errores	40

4.2.9	Implementación en la vida real.....	40
4.2.10	Actualización del modelo.....	41
BIBLIOGRAFÍA.....		42

ABREVIATURAS

ANN	Artificial Neural Network / Redes Neuronales Artificiales
CNN	Redes Neuronales Convolucionales
ESPOL	Escuela Superior Politécnica del Litoral
Fp16	Floating point 16
IA	Inteligencia Artificial
INEC	Instituto Nacional de Estadística y Censos
ODS	Objetivo de Desarrollo Sostenible
RNN	Redes Neuronales Recurrentes
Seq2Seq	Secuencia a secuencia
TrOCR	Transformer-based Optical Character Recognition with Pre-trained Models
UNESCO	Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura

ÍNDICE DE FIGURAS

Ilustración 1 <i>Representación de la IA</i> . Fuente: Internet	8
Ilustración 2 <i>Tipos de Aprendizaje Automático</i> . Fuente: Autores de este estudio	9
Ilustración 3 <i>Representación Neuronal Simple</i> . Fuente: Cienciadedatos.net	11
Ilustración 4 <i>Representación de una red neuronal equivalente a un modelo lineal con 4 predictores</i> . Fuente: Cienciadedatos.net	12
Ilustración 5 <i>Representación de una neurona</i> . Fuente: Cienciadedatos.net.....	14
Ilustración 6 <i>Sección 21 del formulario físico</i> . Fuente: INEC	20
Ilustración 7 <i>Imagen obtenida de formulario</i> . Fuente: INEC	21
Ilustración 8 <i>Ejemplo de Metodología Transformers</i> . Fuente: Huggingface.....	22
Ilustración 9 <i>Preparación de los datos</i> . Fuente: Autores de este documento.	23
Ilustración 10 <i>Tamaño de conjunto de entrenamiento y prueba</i> . Fuente: Autores de esta investigación.	25
Ilustración 11 <i>Funcionamiento del modelo</i> . Fuente: Autores de este estudio	25
Ilustración 12 <i>Parámetros de entrenamiento</i> . Fuente: Autores de este estudio.....	30
Ilustración 13 <i>Resultado del proceso de entrenamiento</i> . Fuente: Autores de documento.	31
Ilustración 14 <i>Predicción del modelo</i> . Fuente: Autores de esta investigación.....	32
Ilustración 15 <i>Métricas de la conjunto de prueba</i> . Fuente: Autores de este documento.	34
Ilustración 16 <i>Distancia de Levenshtein para el conjunto de prueba</i> . Fuente: Autores de este documento.	35
Ilustración 17 <i>Tiempo del procesamiento de 100 formularios</i> . Fuente: Autores de este documento.	36

ÍNDICE DE TABLAS

Tabla 1 <i>Estadística descriptiva de los formularios</i> . Fuente: Autores de esta investigación e INEC.....	28
Tabla 2 <i>Resumen de métricas obtenidas</i> . Fuente: Autores de esta investigación.....	34
Tabla 3 <i>Comparación de modelo con la digitación manual</i> . Fuente: Autores de este documento.....	36

CAPÍTULO 1

En este capítulo, se destaca la relevancia de la inteligencia artificial, poniendo un énfasis particular en las redes neuronales artificiales y cómo su implementación puede resultar beneficioso en el tratamiento y análisis de los formularios de defunciones proporcionados por el INEC. Se resalta también los desafíos y obstáculos identificados en la digitación manual y la imperativa necesidad de abordar y resolver este problema. Se subraya la interrelación entre las Metas de Desarrollo Sostenible (ODS) y dirección del proyecto en cuestión. Se define claramente el propósito primordial del proyecto y los objetivos más detallados que se buscan alcanzar. Además, en la sección del marco teórico, se introducen y conceptualizan varios temas relevantes, incluyendo una exploración de diferentes categorías de redes neuronales y variedades de modalidades de aprendizaje, entre otros tópicos esenciales. Por último, en la sección de Estado del Arte se relata sobre estudios similares.

1. INTRODUCCIÓN

La Inteligencia Artificial se emplea cuando una máquina puede imitar las funciones cognitivas características de la mente humana, como la creatividad, la sensibilidad, el aprendizaje, la comprensión, la percepción del entorno y el uso del lenguaje.

Un subcampo de la IA que ha ganado popularidad en los últimos años es el aprendizaje computacional (*machine learning*), donde un sistema aprende a realizar tareas ya sea mediante ejemplos o a través del ensayo y error.

Los modelos conocidos como redes neuronales se basan en una simplificación del funcionamiento de las neuronas. En la actualidad, las redes neuronales profundas (con un gran número de células organizadas jerárquicamente) han logrado un gran éxito en tareas altamente complejas, como la identificación de objetos en imágenes y el reconocimiento del habla humana. (INCYTU, 2018)

Según la UNESCO (2020), en la actualidad, la Inteligencia Artificial tiene un impacto significativo en casi todas las áreas de la ciencia y la sociedad. Los cuatro núcleos de conocimiento de la IA: razonamiento, representación del conocimiento, tratamiento de la incertidumbre y aprendizaje, se combinan para crear sistemas con un comportamiento racional. Estos núcleos de conocimiento se integran para permitir a la IA abordar una amplia gama de problemas y tomar decisiones de manera inteligente en diferentes campos de aplicación. La inteligencia artificial tiene varias aplicaciones como, sistemas basados en el conocimiento, minería de datos, procesamiento de lenguaje natural y en los videojuegos.

Las Redes Neuronales Artificiales (ANN, por sus siglas en inglés) son entidades de programación diseñadas para simular el funcionamiento de las neuronas biológicas. Estas redes están compuestas por una interconexión de elementos de procesamiento en paralelo, formando una organización jerárquica que les permite interactuar con el mundo. Se definen como sistemas de computación que constan de numerosos elementos de procesamiento simples y altamente interconectados, los cuales procesan información en respuesta a estímulos externos.

Las redes neuronales son concebidas como abstracciones de las estructuras neurobiológicas presentes en la naturaleza, como los cerebros, y se caracterizan por ser sistemas desordenados capaces de almacenar información. (Asanza, 2017)

Por otro lado, los formularios son una parte esencial para la edificación de este modelo; un formulario es una herramienta poderosa para recopilar información. Puede capturar una variedad de datos, desde los básicos como nombre, teléfono y correo electrónico, hasta información más específica como hábitos de consumo o estilo de vida.

En la actualidad, los formularios digitales han adquirido una gran importancia para diversos objetivos empresariales. Con su creciente popularidad, existen diversas herramientas que permiten crear formularios de manera rápida y gratuita. Además, es posible personalizarlos según las necesidades o imagen de la empresa, y muchos de ellos gestionan automáticamente la información en la base de datos correspondiente, sin embargo, no en todos los casos es factible realizar los formularios de forma digital por motivos de falta de dispositivos, fallas de internet o falta de formación en el llenado de estos (Londoño, 2023).

El Instituto Nacional de Estadística y Censos (INEC) es el organismo responsable de la estadística oficial del Ecuador; es la entidad encargada de planificar, normar y certificar la producción del Sistema Estadístico Nacional, además de producir información estadística pertinente, oportuna, confiable y de calidad; e, innovar en metodologías, métricas y análisis de información estadística necesaria para el diseño, implementación y evaluación de la planificación nacional. (INEC, 2020)

Para este proyecto se analizan los formularios de mortalidad que han sido llenados de forma manuscrita, de tal modo que se alimenta la base de datos de los diferentes tipos de caracteres con la finalidad de construir un modelo con inteligencia artificial para su reconocimiento.

1.1 Descripción del problema

El problema se centra en la necesidad de procesar y reconocer los caracteres presentes en formularios completados de forma manuscrita. Actualmente, estos formularios son difíciles de interpretar y analizar debido a la variabilidad en la escritura individual de cada persona. Esto dificulta la extracción y el uso efectivo de la información contenida en los formularios.

Por tanto, se propone construir un modelo de inteligencia artificial Transformers que pueda reconocer y clasificar los caracteres escritos a mano en los formularios, desarrollando un sistema automatizado capaz de convertir los datos escritos en texto digital, facilitando así su procesamiento y almacenamiento en bases de datos.

Este enfoque tiene como finalidad agilizar y mejorar la precisión en la interpretación de los formularios, evitando errores en el proceso de transcripción manual. Al aplicar técnicas de inteligencia artificial y entrenamiento del modelo con una amplia variedad de ejemplos de caracteres escritos a mano, se espera lograr un alto grado de exactitud en el reconocimiento de los caracteres, lo que permitirá un procesamiento más eficiente y una mayor capacidad de análisis de los datos recopilados en los formularios.

1.2 Justificación del problema

La importancia de este proyecto radica en el impacto del entrenamiento de un modelo de Aprendizaje Profundo para la lectura automática de caracteres manuscritos en formularios estadísticos físicos, empleando la arquitectura de Redes Neuronales Transformers, ya que tiene un vínculo significativo con el Objetivo de Desarrollo Sostenible (ODS) 3: "Garantizar una vida sana y promover el bienestar para todos en todas las edades". Este proyecto puede contribuir a mejorar la eficiencia y la precisión en el manejo de información vital en el sector salud. La habilidad para leer y digitalizar automáticamente caracteres manuscritos puede ser especialmente útil en el contexto de los registros médicos, donde la interpretación de escritura manual puede ser problemática y propensa a errores humanos. Al tener estos datos en formato digital, la información de salud puede ser más accesible para los análisis estadísticos. Además, los datos digitalizados pueden ser procesados y analizados para identificar patrones y tendencias, lo que puede conducir a una mejor atención médica y a una toma de decisiones más informada en el ámbito de la salud pública. Este estudio tiene el potencial de contribuir a la meta de mejorar la salud y el bienestar a nivel global, alineándose así con el ODS 3. (Unidas)

El proceso manual de transcripción de estos formularios conlleva un elevado costo en términos de tiempo y recursos humanos. Esto no solo retrasa la disponibilidad de los datos, sino que también aumenta la probabilidad de errores en la interpretación de la escritura a mano. Además, la variabilidad inherente en la escritura individual de cada persona dificulta aún más la tarea de recopilar y analizar los datos contenidos en

los formularios. Esto limita la capacidad de aprovechar eficientemente la información recopilada para la toma de decisiones y la generación de conocimiento.

La implementación de un modelo de inteligencia artificial para el reconocimiento de caracteres en formularios manuscritos ofrece una solución prometedora para estos desafíos. Al automatizar el proceso de reconocimiento, se reducirán significativamente los costos y el tiempo requeridos para transcribir y analizar los formularios. Por otro lado, al utilizar técnicas de aprendizaje automático, el modelo puede adaptarse y mejorar con el tiempo, logrando una mayor precisión en el reconocimiento de caracteres y minimizando los errores de interpretación.

Esta solución permitirá una mayor eficiencia en la recopilación y análisis de datos, proporcionando información más precisa y oportuna para la toma de decisiones. Además, al facilitar la digitalización de los formularios, se podrán aprovechar otras tecnologías como el procesamiento automático del lenguaje natural y la minería de datos.

1.3 Objetivos

1.3.1 Objetivo General

Entrenar un modelo Transformers para la aplicación de técnicas avanzadas de reconocimiento de patrones y procesamiento de imágenes que permita una lectura y digitalización de textos manuscritos de manera precisa y eficiente, por medio de modelamiento.

1.3.2 Objetivos Específicos

1. Entrenar el modelo de inteligencia artificial utilizando el conjunto de datos recopilados, afinando los parámetros y ajustando el rendimiento para el logro de una alta precisión en la lectura de los textos.
2. Evaluar el modelo de inteligencia artificial Transformers, a través de la recopilación y preparación de un conjunto de datos amplio y representativo de formularios manuscritos.
3. Aplicar el modelo Transformers de reconocimiento de caracteres manuscritos en formularios médicos, alineando las capacidades del modelo hacia las metas establecidas en el Objetivo de Desarrollo Sostenible (ODS) 3.

1.4 Marco teórico

1.4.1 Inteligencia artificial

La inteligencia artificial es un campo de estudio amplio que no puede ser definido de manera exacta, pero se puede describir como la capacidad de las computadoras para realizar tareas que requieren inteligencia humana. Esto implica que las computadoras utilizan algoritmos, interpretan datos y toman decisiones, similar a cómo lo haría un ser humano.

Las tecnologías basadas en inteligencia artificial ofrecen numerosos beneficios y mejoras significativas en diversas áreas de la vida humana. Algunas de las aplicaciones técnicas de la IA incluyen:

- Reconocimiento y clasificación de imágenes
- Mejoras en estrategias comerciales basadas en algoritmos
- Procesamiento eficiente de datos de pacientes
- Mantenimiento predictivo
- Detección y clasificación de objetos
- Distribución de contenido en redes sociales
- Protección contra amenazas de seguridad cibernética.

Estas aplicaciones contribuyen a mejorar la eficiencia y eficacia en diferentes ámbitos de la sociedad. (Rouhiainen, 2018)



Ilustración 1 *Representación de la IA.* Fuente: Internet

1.4.2 Aprendizaje Automático

Machine Learning o Aprendizaje automático es uno de los principales enfoques de la IA, se trata de un aspecto de la informática en la que las computadoras tienen la capacidad de aprender sin estar programados para ello. Uno de sus resultados son las sugerencias o predicciones en un caso particular.

Este tipo de aprendizaje usa algoritmos para aprender los patrones en los datos.

A continuación, se muestra tres tipos de aprendizaje que se pueden utilizar.

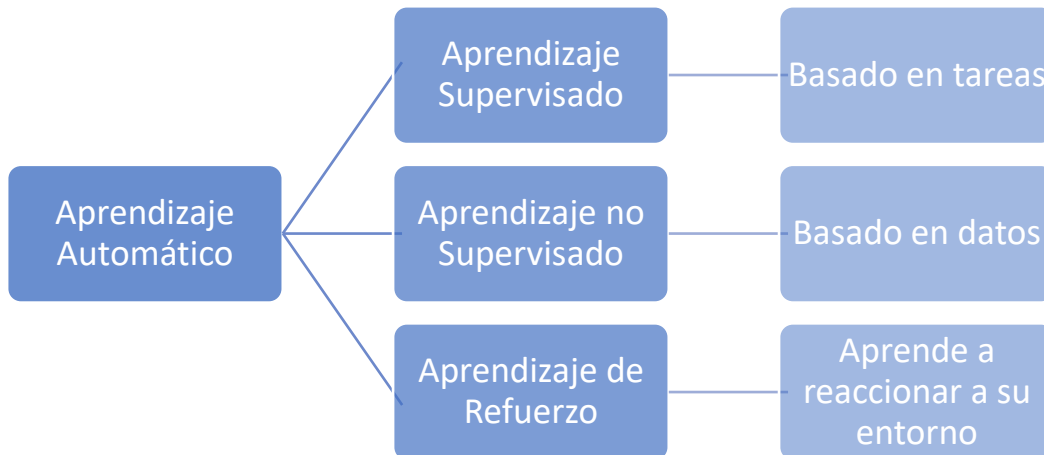


Ilustración 2 *Tipos de Aprendizaje Automático.* Fuente: Autores de este estudio

El aprendizaje supervisado utiliza datos que ya han sido organizados y etiquetados anteriormente para señalar como tiene que ser catalogada la nueva información. Para esto, se necesita de la parte humana, ya que se debe proporcionar la retroalimentación.

El aprendizaje no supervisado no necesita de ningún dato organizado o etiquetado anteriormente para saber cómo tendría que ser categorizado los nuevos datos, sino que el algoritmo tiene que encontrar la forma de clasificarlo por sí solo.

El aprendizaje por refuerzo son algoritmos que aprenden mediante la experiencia, es decir, se entrega un “refuerzo positivo” cada vez que se clasifique de forma correcta. (Rouhiainen, 2018)

1.4.3 Aprendizaje Profundo

Este tipo de aprendizajes es uno de los más poderosos y con una tendencia creciente, conocido como *Deep Learning*. Este aprendizaje se utiliza para solucionar problemas muy complejos y que necesite de grandes cantidades de datos.

Se realiza mediante redes neuronales, que se organizan en capas para reconocer las relaciones y los patrones en los datos. Para esto, se necesita una enorme base de datos y una potente capacidad de procesamiento. Tiene varias aplicaciones como en el reconocimiento de voz, el procesamiento del lenguaje natural, la visión artificial y la identificación de vehículos en los sistemas de asistencia al conductor. (Rouhiainen, 2018)

1.4.4 Redes Neuronales

Las redes neuronales son construcciones que se originan a partir de la organización de operaciones matemáticas siguiendo una estructura específica. La forma más frecuente de describir la estructura de una red neuronal es a través de capas (*layers*), las cuales están compuestas por neuronas, también llamadas unidades o células neuronales. Cada neurona realiza una operación básica y está interconectada mediante pesos a las neuronas de la capa anterior y siguiente. Estos pesos desempeñan la función de regular la información que se transmite de una neurona a otra.

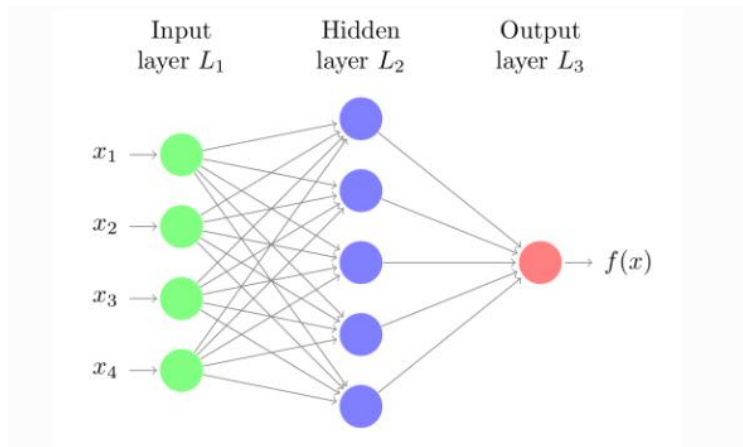


Ilustración 3 Representación Neuronal Simple. Fuente: Cienciadedatos.net

La capa inicial de la red neuronal, representada en verde, se denomina capa de entrada o "*input layer*". En esta capa, se reciben los datos en su forma original, es decir, los valores de los predictores. La capa intermedia, destacada en azul, se conoce como capa oculta o "*hidden layer*". En esta capa, los valores provenientes de la capa de entrada son multiplicados por los pesos correspondientes (representados por flechas grises). Por último, la capa final, denominada "*output layer*", combina los valores resultantes de la capa intermedia para generar la predicción.

Para facilitar la comprensión de la estructura de las redes, es útil representarla como un modelo de regresión lineal equivalente.

$$y = w_1x_1 + \dots + w_dx_d + b$$

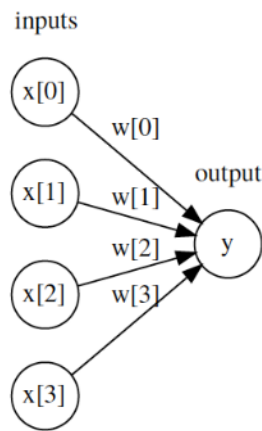


Ilustración 4 Representación de una red neuronal equivalente a un modelo lineal con 4 predictores. Fuente: Cienciadedatos.net

Cada predictor se representa por una neurona en la capa de entrada. Los coeficientes de regresión, llamados pesos en las redes neuronales se muestran como flechas, mientras que el valor predicho se representa mediante la neurona de salida. Para que esta representación sea equivalente a la ecuación de un modelo lineal, es necesario considerar dos elementos adicionales:

1. El bias del modelo, que es un término constante agregado al resultado de la combinación lineal de los predictores y sus pesos. El bias representa un desplazamiento en el eje vertical de la función lineal. Aunque las representaciones gráficas a menudo omiten el bias de las neuronas de la capa intermedia, este es necesario.

2. Las operaciones matemáticas de multiplicación y suma ponderada, que combinan los valores de los predictores con los pesos del modelo. Estas operaciones

se llevan a cabo dentro de las neuronas y son fundamentales para el funcionamiento de la red neuronal. (Rodrigo, 2021)

1.4.5 Neurona o Unidad

La neurona desempeña un papel fundamental en los modelos de redes, ya que es la unidad funcional básica. Dentro de cada neurona, ocurren dos operaciones principales: la suma ponderada de las entradas y la aplicación de una función de activación.

En la primera etapa, se multiplican los valores de entrada x_i por sus respectivos pesos asociados w_i y se suman junto con el bias. Esto da como resultado el valor neto de entrada a la neurona. A continuación, este valor neto de entrada se pasa a través de una función de activación, la cual transforma dicho valor en un valor de salida.

Aunque la entrada a la neurona es siempre una combinación lineal de los valores de entrada, la función de activación permite generar una amplia variedad de salidas. Es en la función de activación donde radica el potencial de los modelos de redes para aprender relaciones no lineales. Gracias a la función de activación, la red neuronal puede capturar patrones complejos y aprender representaciones no lineales de los datos.

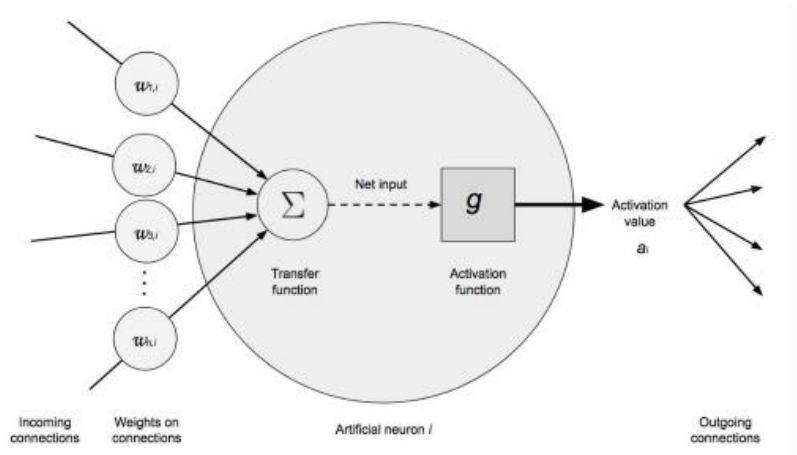


Ilustración 5 Representación de una neurona. Fuente: Cienciadedatos.net

La representación matemática se puede ver de la siguiente forma.

La neurona calcula el valor neto de entrada al sumar los valores de entrada, cada uno multiplicado por su respectivo peso de conexión, y luego se le agrega el bias.

$$entrada = \sum_{i=1}^n x_i w_i + b$$

En lugar de expresarlo como un sumatorio, es común representar esta operación como un producto matricial, donde el vector X representa los valores de entrada y el vector W representa los pesos.

$$entrada = XW + b$$

Se aplica una función de activación (g) al valor neto de entrada, lo cual resulta en la obtención del valor de activación (a), que es el resultado final que se obtiene de la neurona y representa la salida de esta.

$$a = g(\text{entrada}) = g(XW + b)$$

En la capa de entrada, donde se busca incorporar únicamente el valor de los predictores, se utiliza una función de activación unitaria, lo que significa que la salida es igual a la entrada. En la capa de salida, la función de activación comúnmente utilizada es la identidad para problemas de regresión y la función *softmax* para problemas de clasificación. (Rodrigo, 2021)

1.5 Estado del Arte

Se revisa un primer artículo que se titula "TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models" y se centra en el reconocimiento de texto utilizando modelos Transformer pre-entrenados. Autores: Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, Furu Wei
Año: 2022.

En términos generales, el artículo propone un nuevo enfoque para el reconocimiento óptico de caracteres (OCR) utilizando Transformer. A diferencia de los enfoques tradicionales que utilizan CNN para la comprensión de imágenes y RNN para la generación de texto a nivel de caracteres, el enfoque propuesto, llamado TrOCR, utiliza Transformer para ambas tareas.

Los autores argumentan que este enfoque es simple pero efectivo, y puede ser pre-entrenado con datos sintéticos a gran escala y luego ajustado con conjuntos de datos etiquetados por humanos. Según los experimentos realizados, TrOCR supera a

los modelos actuales del estado del arte en tareas de reconocimiento de texto impreso, escrito a mano y de escena.

El modelo TrOCR se construye con una arquitectura Transformer, que incluye un Transformer para la extracción de características visuales y otro Transformer para el modelado del lenguaje. El enfoque propuesto inicializa el codificador y el decodificador con modelos pre-entrenados en conjuntos de datos etiquetados y no etiquetados a gran escala. Los autores sostienen que este enfoque elimina la necesidad de un modelo de lenguaje externo.

Los experimentos realizados con TrOCR incluyen pruebas con diferentes tipos de datos, como documentos impresos, textos escritos a mano y textos en escena. Según los autores, TrOCR logra resultados superiores a los del estado del arte con un modelo de codificador-decodificador basado en Transformer, que no depende de ninguna etapa compleja de preprocesamiento o postprocesamiento.

En resumen, este primer artículo Minghao Li, Tengchao Lv, Jingye Chen, Lei Cui, Yijuan Lu, Dinei Florencio, Cha Zhang, Zhoujun Li, Furu Wei, publicado en el año 2022, propone un nuevo enfoque para el reconocimiento óptico de caracteres que utiliza Transformer para la comprensión de imágenes y la generación de texto. Los resultados experimentales sugieren que este enfoque puede superar a los métodos actuales del estado del arte en varias tareas de reconocimiento de texto.

El segundo artículo de revisión es el titulado "Transformer-based HTR for Historical Documents". Este artículo se centra en el reconocimiento de texto manuscrito (HTR) utilizando arquitecturas de transformer para documentos históricos.

En particular, los autores utilizan el modelo TrOCR, una combinación de transformer basada en visión (BEiT) y el modelo de representación de lenguaje RoBERTa. Autores: Phillip Benjamin Ströbel, Simon Clematide y Martin Volk. Año: 2022.

El reconocimiento de texto manuscrito ha sido una herramienta útil para extraer texto de documentos escaneados y ha experimentado mejoras significativas con la introducción de técnicas basadas en redes neuronales. El reto es aún más grande para las colecciones de manuscritos históricos debido a la variabilidad de estilos de escritura manuscrita y las variaciones de la escritura de los autores según lo que estaban escribiendo.

TrOCR es un modelo pre-entrenado basado en transformer que ha demostrado ser eficaz en el procesamiento de imágenes y se espera que también lo sea en la comprensión del lenguaje. BEiT, el componente de visión de TrOCR, se pre-entrena en el conjunto de datos Image-Net-1K que contiene 1.2M imágenes, mientras que RoBERTa se utiliza como decodificador para producir el texto. Se experimentó con dos versiones del decodificador RoBERTa: una con 6 capas y otra con 12 capas, llamadas BASE y LARGE respectivamente.

Los autores se propusieron probar si el "conocimiento" de la escritura manuscrita en inglés que TrOCR ha adquirido se podía transferir a los manuscritos modernos tempranos. Utilizaron dos conjuntos de datos para entrenar y afinar el modelo: uno de Rudolf Gwalther y otro de Heinrich Bullinger. Los modelos resultantes de TrOCR, TrOCRBASE y TrOCRLARGE, se compararon con un modelo de referencia, HTR+, también entrenado en los mismos datos.

Los resultados de los experimentos iniciales mostraron que TrOCR supera a HTR+, un resultado sorprendente dado que TrOCR no había sido previamente expuesto a ningún dato en latín. Los autores encontraron que la precisión del modelo aumentaba cuanto más se afinaba, aunque esto fue menos pronunciado en el caso de TrOCRBASE, cuya precisión incluso disminuyó después de más de diez épocas de afinación.

Los resultados sugieren que TrOCR es un modelo de representación de escritura manuscrita robusto y altamente transferible, adecuado para ser afinado en diferentes estilos de escritura. Además, TrOCR demostró no necesitar información de la línea de base, a diferencia de los modelos de Transkribus.

Para trabajos futuros, los autores Phillip Benjamin Ströbel, Simon Clematide y Martin Volk, publicado en el 2022, planean investigar si el uso de un decodificador RoBERTa pre-entrenado en latín podría mejorar el rendimiento. También desean examinar TrOCR en conjuntos de datos más variables, como aquellos con muchos autores diferentes, para investigar si TrOCR generaliza mejor a estos datos que los métodos convencionales.

CAPÍTULO 2

En este capítulo, se brinda una explicación detallada sobre el origen de los formularios que se utilizan en este proyecto y la sección con la que se trabajó en este estudio, incluyendo la región específica de donde provienen y el propósito de su uso. Se ofrece una descripción de la metodología empleada por el modelo pre entrenado Transformers, explicando su funcionamiento y su aplicación en este contexto. Además, se discute la preparación de los datos, abordando cómo se seleccionan, se organizan y se procesan los formularios para ser utilizados de manera efectiva en el modelo. Finalmente, se expone el proceso de entrenamiento del modelo con los datos obtenidos de los formularios, presentando una descripción de los pasos seguidos, las técnicas empleadas y los resultados obtenidos en este proceso de aprendizaje automático.

2. METODOLOGÍA

La fuente de los datos que se han utilizado en este proyecto son formularios que recopilan información de pacientes que, lamentablemente, han perdido la vida debido a distintas causas. Estos documentos son diligenciados de manera manual por profesionales médicos que registran los detalles pertinentes y, una vez terminado este proceso, estos se remiten al Registro Civil de cada ciudad, para luego ser enviados al Instituto Nacional de Estadísticas y Censos (INEC) para realizar el correspondiente análisis y procesamiento de la información contenida en ellos.

El proyecto en cuestión ha aprovechado un total de 4999 de estos formularios, todos provenientes específicamente de la región conocida como ZONAL 8. Esta zona geográfica abarca varias zonas importantes del país, incluyendo a Galápagos, Santa

Elena, Manabí, Los Ríos, Santo Domingo de los Tsáchilas, Bolívar y Guayas, excepto Guayaquil. Es en estas áreas geográficas donde se recogieron y llenaron los formularios que luego se utilizaron en este proyecto. Es importante mencionar que, a pesar de ser parte de la ZONAL 8, se ha tomado la decisión de excluir a la ciudad de Guayaquil del conjunto de datos. La razón de esta exclusión radica en evitar un sobreentrenamiento del modelo. Guayaquil, siendo una ciudad grande, cuenta con alrededor de 50.000 registros, lo que podría sesgar el entrenamiento del modelo hacia los datos de esta ciudad en particular, perjudicando su rendimiento en el reconocimiento de patrones más generales. La diversidad de las regiones de las que provienen estos formularios proporciona una amplia gama de datos para trabajar y analizar.

El formulario que se utiliza tiene varias secciones. Sin embargo, para las necesidades específicas de este proyecto, el enfoque estará en la sección que detalla la causa de la muerte del paciente, que se muestra a continuación:

21) CAUSA DE LA DEFUNCIÓN CERTIFICADAS POR PROFESIONAL DE LA SALUD MÉDICO/A O MÉDICO/A LEGISTA: <i>Revise al reverso las directrices del llenado</i>		
El presente certificado debe ser llenado con letra clara, legible y en imprenta. El profesional de la salud médico/a o médico/a legista, debe registrar la enfermedad, lesión o estados patológicos según terminología médica, (La codificación de la CIE-10 es uso exclusivo del INEC). Las muertes por causas externas (accidentales y/o violentas) serán certificadas únicamente por el médico/a legista.		Tiempo aproximado entre el comienzo de cada causa y la muerte. (En minutos, horas, días, semanas, meses o años)
PARTE I Enfermedad, lesión o estado patológico que produjo la muerte directamente	a) debido a (o como consecuencia de)	[USO INEC]
CAUSAS ANTECEDENTES Estados morbosos, si existiera alguno, o lesiones que produjeron la causa arriba consignada, mencionándose en último lugar la causa ANTECEDENTE O FUNDAMENTAL	b) debido a (o como consecuencia de)	[USO INEC]
	c) debido a (o como consecuencia de)	[USO INEC]
	d)	[USO INEC]
PARTE II OTROS ESTADOS PATOLÓGICOS SIGNIFICATIVOS, que contribuyeron a la muerte, pero no relacionados con la enfermedad o estado morbozo que la condujo	[USO INEC]
	[USO INEC]

Ilustración 6 Sección 21 del formulario físico. Fuente: INEC

Esta información ha sido extraída y transformada en formato de imagen, tal y como se muestra en la siguiente ilustración:



Ilustración 7 *Imagen obtenida de formulario.* Fuente: INEC

El desafío de la investigación para el reconocimiento de texto en la digitalización de documentos es un proceso que se ha estudiado durante años. Los métodos existentes suelen emplear Redes Neuronales Convolucionales (CNN) para el análisis de imágenes y Redes Neuronales Recurrentes (RNN) para la generación de texto a nivel de caracteres. Además, suele requerirse un modelo de lenguaje adicional para aumentar la precisión total como un paso de post-procesamiento. Para este proyecto nos hemos basado en una técnica integral de reconocimiento de texto utilizando los modelos preentrenados de Transformers de Imagen y Texto, específicamente TrOCR, que utiliza la arquitectura de Transformers tanto para la interpretación de imágenes como para la generación de texto a nivel de palabras. El modelo TrOCR es eficaz y puede ser previamente entrenado con datos sintéticos a gran escala y ajustado con conjuntos de datos anotados por humanos. Los experimentos demuestran que el modelo TrOCR supera a los modelos más avanzados actuales en las tareas de reconocimiento de texto impreso, manuscrito y de escenas.

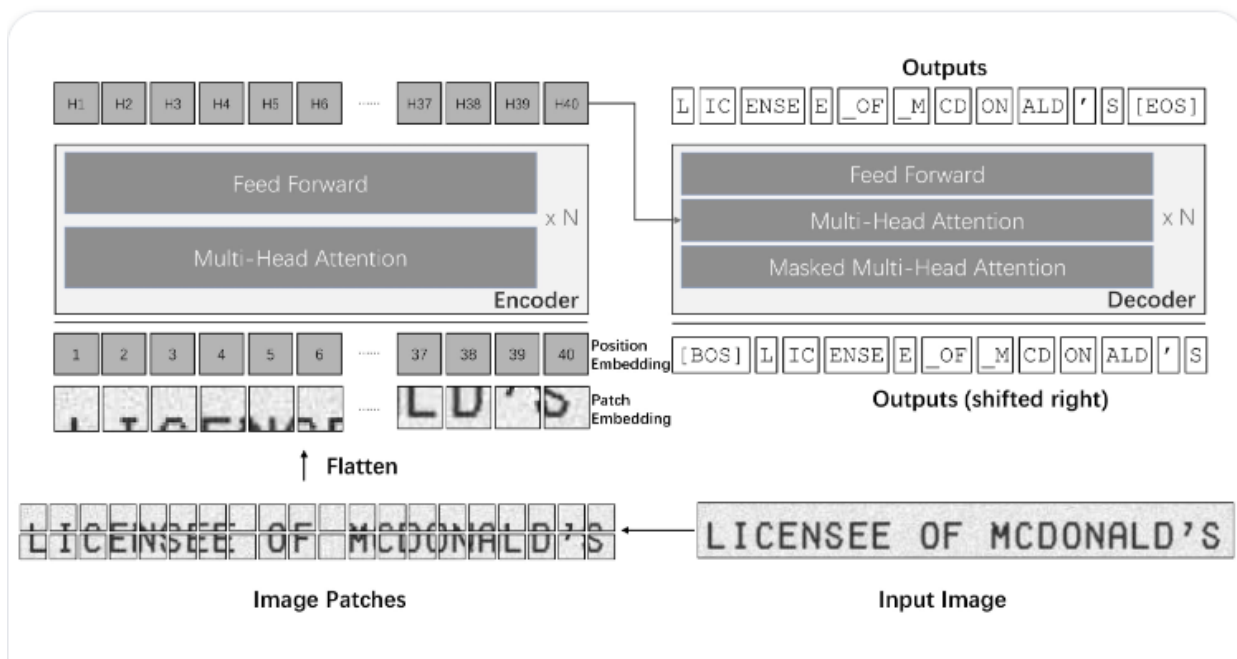


Ilustración 8 Ejemplo de Metodología Transformers. Fuente: Huggingface

El *Encoder* o Codificador es una red neuronal que, en este caso particular, tomó como entrada una imagen, que representa el registro de la causa de muerte de un individuo. Este Codificador se encargó de procesar y convertir la imagen en un conjunto de características de alto nivel, que finalmente se formularon en un vector numérico. Este vector de características fue diseñado para capturar la esencia de la imagen de entrada de tal manera que pudiera ser interpretada y utilizada eficazmente por el modelo. La finalidad principal del Codificador fue simplificar los datos de entrada, reduciendo su dimensionalidad, a la vez que se esforzaba por retener la máxima cantidad de información relevante posible.

Por otra parte, el *Decoder* o Decodificador es otra red neuronal que se encargó de trabajar con el conjunto de características producidas por el Codificador para generar una nueva salida a partir de él. Por ejemplo, el Codificador tomó la imagen que representa la causa de muerte y la transformó en un vector de características. Posteriormente, el Decodificador tomó este vector y lo utilizó para generar una oración

que representa la información contenida en la imagen original. Por lo tanto, el Decodificador desempeñó un papel crucial en la 'traducción' de la representación codificada que el Codificador había producido, generando una salida que resultó ser más comprensible y útil en el contexto del modelo.

2.1 Preparación de los datos

Se han organizado los datos en un *dataframe* que consta de dos columnas. En este *dataframe*, cada fila representa un registro único que se compone de dos partes: el nombre del archivo de la imagen y el texto que corresponde a dicha imagen. Esta estructuración permite un acceso fácil y eficiente a la información, con la asociación clara entre la imagen y su texto correspondiente.

<code>file_name</code>	<code>text</code>
<code>EDG_09105001_00588_p21d.png</code>	DIFICULTAD RESPIRATORIA DEL RECIEN NACIDO
<code>EDG_09105001_00590_p21a.png</code>	INFARTO AGUDO DEL MIOCARDIO
<code>EDG_09105001_00590_p21b.png</code>	INSUFICIENCIA CARDIACA
<code>EDG_09105001_00590_p21c.png</code>	FIBROSIS PULMONAR
<code>EDG_09105001_00590_p21d.png</code>	HIPERTENSION ARTERIAL

Ilustración 9 *Preparación de los datos.* Fuente: Autores de este documento.

Se realizó una división de estos datos en conjuntos de entrenamiento y prueba. Para este paso se utilizó la función 'train_test_split' del paquete 'sklearn'.

Cada elemento del conjunto de datos está diseñado para entregar dos componentes esenciales:

1. 'pixel_values', que se utilizan como el insumo principal para el modelo de aprendizaje automático. Estos 'pixel_values' son, en esencia, la representación

numérica de cada imagen, donde cada píxel se convierte en un valor numérico específico que el modelo puede procesar y aprender.

2. *labels*, que corresponden a los *input_ids* del texto asociado con la imagen. Estos *input_ids* son representaciones numéricas del texto correspondiente a la imagen. Estas *labels* juegan un papel vital durante el entrenamiento del modelo, ya que proporcionan la 'solución' que el modelo intentará predecir a partir de los *pixel_values*.

Con el objetivo de optimizar el tiempo y debido a las limitaciones de capacidad de procesamiento de la computadora, se ha decidido trabajar con un conjunto de 4999 observaciones. Esta decisión permite probar la eficacia y el funcionamiento del modelo sin sobrecargar el sistema ni incurrir en demoras excesivas durante las pruebas. Es una medida práctica que se adopta teniendo en cuenta las restricciones técnicas y la necesidad de obtener resultados preliminares de manera eficiente, por tanto, se han seleccionado 4999 imágenes, las cuales se dividen en 3999 imágenes para el

entrenamiento y 1000 para la validación.

```
print("Number of training examples:", len(train_dataset))
print("Number of validation examples:", len(eval_dataset))
#df.head(train_dataset)
```

```
Number of training examples: 3999
Number of validation examples: 1000
```

Ilustración 10 *Tamaño de conjunto de entrenamiento y prueba.* Fuente: Autores de esta investigación.

Al utilizar los datos de entrenamiento, se puede observar el funcionamiento del código para la obtención del texto digitalizado, como se muestra en la siguiente imagen:

```
[ ] image = Image.open(train_dataset.root_dir + train_df['file_name'][0]).convert("RGB")
image
```

A photograph of a handwritten note on lined paper. The text is written in blue ink and reads "Pero cardiopulmonario." The word "Pero" is on the first line, and "cardiopulmonario." is on the second line. A dashed horizontal line is drawn below the text. A white rectangular bounding box is drawn around the text.

```
[ ] labels = encoding['labels']
labels[labels == -100] = processor.tokenizer.pad_token_id
label_str = processor.decode(labels, skip_special_tokens=True)
print(label_str)
```

```
PARO CARDIORESPIRATORIA
```

Ilustración 11 *Funcionamiento del modelo.* Fuente: Autores de este estudio

2.2 Entrenamiento del modelo

Hay varios atributos esenciales necesarios para generar los *decoder_input_ids* a partir de las etiquetas. El modelo, de manera autónoma, creará los *decoder_input_ids* realizando un desplazamiento de las etiquetas una posición a la derecha, antecediendo el *decoder_start_token_id* y sustituyendo aquellos identificadores que son -100 por el *pad_token_id*.

Además, es importante considerar el tamaño del vocabulario del modelo. Este aspecto es relevante para el cabezal de modelado de lenguaje, que se ubica en la parte superior del decodificador y es responsable de la generación de texto.

Finalmente, también es necesario tener en cuenta ciertos parámetros que están asociados con la técnica de búsqueda de haces (*beam search*). Esta técnica se utiliza en la generación de texto y es esencial para crear secuencias de palabras más coherentes y precisas.

CAPÍTULO 3

El capítulo se dedica a presentar los resultados y el análisis del modelo desarrollado. Se recolectaron formularios de siete zonas, acumulando un total de 4999 formularios. Las ciudades con mayor contribución fueron Guayas y Manabí, con 1767 y 1530 formularios, respectivamente, mientras que Galápagos tuvo la mínima contribución con solo 2 formularios. Para entrenar el modelo, se emplearon diversos parámetros que ajustan cómo este aprende y evalúa datos. Algunos de estos parámetros incluyen `predict_with_generate`, que se basa en la función de generación del modelo, y `fp16`, que permite una mayor velocidad al entrenar el modelo. La fase de predicción se implementó utilizando un objeto entrenador (`trainer`) y un conjunto de datos de evaluación. Esta fase se encargó de llevar a cabo las predicciones y almacenar los resultados para su análisis.

El rendimiento del modelo se evaluó utilizando diversas métricas. La pérdida de la prueba (`test_loss`) fue de 0.344, indicando el error del modelo, mientras que el Error de Tasa de Caracteres (`test_cer`) fue del 6.9%, mostrando que el modelo erró en esa proporción de caracteres. La prueba se completó en 291 minutos, es decir, aproximadamente 5 horas y el modelo mostró una eficiencia de 3 muestras por segundo. Estos resultados reflejan tanto la precisión como la eficiencia del modelo en su tarea.

3. RESULTADOS Y ANÁLISIS

3.1 Estadística descriptiva

Para conformar el conjunto de datos para el modelo se recogieron formularios en siete zonas distintas, cada una identificada por un código único. Bolívar, con el código

02, aportó 246 formularios al conjunto total. Guayas, que se identifica con el código 09, proporcionó la mayor cantidad de formularios, un total de 1767. Los Ríos identificado con el código 12 contribuyó con 726 formularios. Manabí con el código 13 aportó 1530 formularios, ocupando el segundo lugar en términos de la mayor contribución.

Galápagos, representado por el código 20, tuvo la menor contribución con solo 2 formularios. Santo Domingo, con el código 23, contribuyó con 372 formularios. Y por último Santa Elena, identificada por el código 24, suministró 356 formularios.

En total, se recogieron 4999 formularios en todas las ciudades, proporcionando un amplio conjunto de datos para su análisis. Guayas y Manabí fueron las ciudades que aportaron la mayor cantidad de formularios, mientras que Galápagos aportó la menor cantidad.

A continuación, se detalla en la siguiente tabla la cantidad de formularios:

PROVINCIAS	CÓDIGO	CANTIDAD DE FORMULARIOS	FRECUENCIA RELATIVA
Bolívar	02	246	4,92%
Guayas*	09	1767	35,35%
Los Ríos	12	726	14,52%
Manabí	13	1530	30,60%
Galápagos	20	2	0,04%
Santo Domingo	23	372	7,45%
Santa Elena	24	356	7,12%
TOTAL		4999	100,00%

Tabla 1 Estadística descriptiva de los formularios. Fuente: Autores de esta investigación e INEC

3.2 Parámetros de entrenamiento del modelo

En el proceso de entrenamiento de los datos, se han configurado varios argumentos para el modelo mediante la creación de una instancia de *Seq2SeqTrainingArguments*. Esta instancia permite personalizar diferentes opciones y parámetros de entrenamiento para los modelos de Secuencia a Secuencia (Seq2Seq).

El parámetro *predict_with_generate=True* implica que las predicciones generadas por el modelo durante el entrenamiento se basarán en la función de generación del propio modelo.

El parámetro *evaluation_strategy="steps"* establece que la evaluación del modelo se realizará en intervalos regulares de pasos específicos durante el proceso de entrenamiento.

Los parámetros *per_device_train_batch_size* y *per_device_eval_batch_size* se establecen en 8 y determinan la cantidad de ejemplos de entrenamiento y evaluación que se procesarán simultáneamente en cada dispositivo.

El parámetro *fp16=True* activa el uso de la precisión mixta (*floating point 16*). Esto ayuda a acelerar el entrenamiento y reduce la cantidad de memoria necesaria.

El parámetro *output_dir="."* designa el directorio en el que se guardarán los resultados del entrenamiento.

El parámetro *logging_steps=2* establece que los registros del entrenamiento se guardarán cada dos pasos.

El parámetro *save_steps=1000* indica que los modelos intermedios generados durante el entrenamiento se guardarán después de cada 1000 pasos de entrenamiento. Finalmente, el parámetro *eval_steps=200* especifica que la evaluación del modelo en el conjunto de validación se realizará cada 200 pasos durante el entrenamiento.

```

from transformers import Seq2SeqTrainer, Seq2SeqTrainingArguments
import accelerate
import transformers

training_args = Seq2SeqTrainingArguments(
    predict_with_generate=True,
    evaluation_strategy="steps",
    per_device_train_batch_size=8,
    per_device_eval_batch_size=8,
    fp16=True,
    output_dir=".",
    logging_steps=2,
    save_steps=1000,
    eval_steps=200,
)

```

Ilustración 12 *Parámetros de entrenamiento.* Fuente: Autores de este estudio.

3.3 Resultados del proceso de entrenamiento

Se ha completado 1500 pasos o interacciones dentro del proceso de entrenamiento, además, el tiempo total que ha tomado el entrenamiento es de 2 horas y 6 minutos.

En este contexto los resultados del *Training Loss* que representa la pérdida o error que el modelo tiene en el conjunto de entrenamiento en ese paso específico. Generalmente, cuanto menor sea este valor, mejor, ya que indica que el modelo está ajustando bien los datos de entrenamiento.

El parámetro *Validation Loss* representa la pérdida en el conjunto de validación. Esto ayuda a verificar si el modelo está generalizando bien o si está comenzando a sobre ajustarse (*overfitting*) a los datos de entrenamiento.

El parámetro *Cer* "Character Error Rate" (Tasa de Error de Carácter), especialmente si estamos hablando de un modelo que trabaja con texto o

reconocimiento de caracteres. Representa la proporción de caracteres predichos que son incorrectos en comparación con los verdaderos caracteres.

Dando un total de 1400 registros periódicos que muestran cómo estos valores que han evolucionado a lo largo del tiempo o de los pasos en el entrenamiento. Por lo general, esperaríamos ver una tendencia en la que las pérdidas disminuyen con el tiempo, lo que indica que el modelo está aprendiendo. Sin embargo, hay que tener cuidado de no dejar que el modelo se sobreajuste.

```
warnings.warn(
[1500/1500 2:06:49, Epoch 3/3]
```

Step	Training Loss	Validation Loss	Cer
200	0.823400	0.918478	0.229372
400	0.410000	0.725337	0.165856
600	0.228800	0.550668	0.125785
800	0.363200	0.523764	0.130856
1000	0.276400	0.405534	0.085796
1200	0.160500	0.376226	0.077566
1400	0.127100	0.352584	0.070291

Ilustración 13 Resultado del proceso de entrenamiento. Fuente: Autores de documento.

3.4 Predicción del modelo

Este extracto de código se dedica a llevar a cabo la etapa de predicción en el entrenamiento del modelo Transformers.

En la primera línea, *model = trainer*, se está llevando a cabo una asignación del objeto *trainer*. Este comando es una instancia de un entrenador del *Seq2SeqTrainer*, que ha sido previamente configurado con los parámetros de entrenamiento y con el modelo específico a entrenar. Al asignar el *trainer* a la variable *model*, se facilita su uso

posterior en el contexto de realizar predicciones, permitiendo una referencia más intuitiva al entrenador como un "modelo".

En la siguiente línea, `results = model.predict(eval_dataset)`, se está invocando el método `predict` en `model`, que es nuestro `trainer`, y se le está proporcionando `eval_dataset`, es decir, el conjunto de datos de evaluación. Este paso ejecutará la etapa de predicción utilizando el conjunto de datos de evaluación, y los resultados de esta predicción se almacenarán en la variable `results`. En resumen, este fragmento de código se encarga de implementar la fase de predicción del proceso de entrenamiento del modelo, utilizando el conjunto de datos de evaluación y almacenando los resultados obtenidos para su posterior análisis y revisión.

```
model = trainer
results = model.predict(eval_dataset)
```

Ilustración 14 *Predicción del modelo.* Fuente: Autores de esta investigación.

3.5 Resultados

Las métricas proporcionan una visión detallada del rendimiento del modelo durante la fase de prueba. Cada métrica sirve para evaluar diferentes aspectos del modelo:

`test_loss`

Esta es la pérdida (o error) del modelo en el conjunto de datos de prueba. Un valor más bajo indica un mejor rendimiento del modelo, ya que, significa que sus predicciones están más cerca de las etiquetas reales. En este caso, la pérdida de la prueba es 0.344.

test_cer

Este es el Error de Tasa de Caracteres (*Character Error Rate*) en el conjunto de datos de prueba. Es una métrica que se utiliza comúnmente en tareas de reconocimiento de voz y texto, y mide la cantidad de errores de caracteres (incluyendo inserciones, eliminaciones y sustituciones) que el modelo hace en comparación con la verdad de referencia. En este caso, el CER es 0.0690, lo que significa que aproximadamente el 6.9% de los caracteres predichos por el modelo fueron incorrectos.

test_runtime

Este es el tiempo que llevó al modelo realizar la prueba en el conjunto de datos de prueba. En este caso, la validación duró 291 minutos, es decir, alrededor de 5 horas.

test_samples_per_second

Esto se refiere a la cantidad de muestras que el modelo fue capaz de procesar por segundo durante la fase de prueba. Un número mayor indica que el modelo es más rápido. Aquí, el modelo pudo procesar aproximadamente 3 muestras por segundo.

test_steps_per_second

Esta es la cantidad de pasos que el modelo pudo realizar por segundo durante la fase de prueba. Un número mayor indica que el modelo es más rápido. Aquí, el modelo pudo realizar aproximadamente 0,429 pasos por segundo.

Estas métricas son útiles para entender cómo se desempeña el modelo, no solo en términos de precisión, sino también en términos de eficiencia y velocidad.


```
metrics={'test_loss': 0.2989717125892639, 'test_cer': 0.06032327945342443, 'test_wer': 0.114651887727745,
'test_bleu': 0.5291364011275062, 'test_levenshtein': 1.493, 'test_runtime': 304.9919,
'test_samples_per_second': 3.279, 'test_steps_per_second': 0.41}
```

Ilustración 15 Métricas de la conjunto de prueba. Fuente: Autores de este documento.

MÉTRICAS	RESULTADOS
test_los	29,89%
test_cer	6,03%
test_runtime	291,321
test_samples_per_second	3,443
test_steps_per_second	0,429

Tabla 2 Resumen de métricas obtenidas. Fuente: Autores de esta investigación

Otra métrica interesante es distancia de Levenshtein, que nos informa que mide la cantidad mínima de operaciones de edición (como inserciones, eliminaciones o sustituciones) necesarias para transformar una cadena de texto en otra. Y los resultados del modelo para esta métrica en función del número de muestras que el modelo entrenó fue el siguiente:

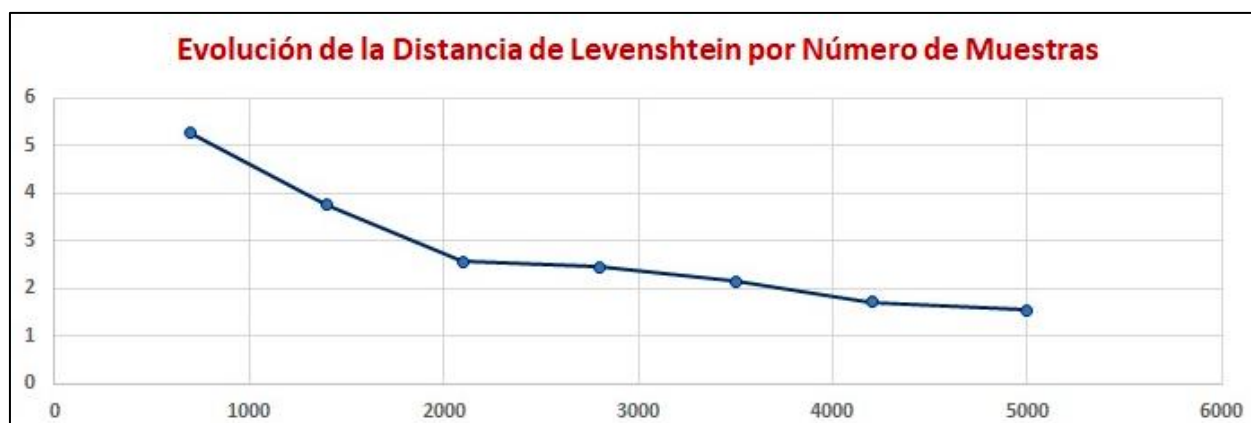


Ilustración 16 *Distancia de Levenshtein para el conjunto de prueba.* Fuente: Autores de este documento.

Donde se observa que a medida que el número de muestra aumenta esta distancia decrece, llegando a 1.55 ediciones promedios el cual es un buen valor para calificar nuestro modelo, sabiendo que a medida que se incrementa el tamaño de muestra esta distancia será menor.

3.6 Comparación de tiempo

Durante una jornada laboral típica, que generalmente consta de ocho horas, un persona tiene la capacidad de digitar alrededor de 100 formularios manualmente. Para contrastar esta labor manual, se llevó a cabo un experimento utilizando un modelo automatizado con un lote de 100 formularios. Los resultados mostraron que no solo el modelo fue capaz de realizar predicciones precisas sobre el contenido de los formularios, sino que también logró completar esta tarea en un tiempo significativamente reducido de aproximadamente 10 minutos. Este proceso automatizado representa una notable optimización en comparación con el tiempo tradicionalmente requerido para la digitalización manual.

```
image
H:PERTENSION ARTERIAL

labels = encoding['labels']
labels[labels == -100] = processor.tokenizer.pad_token_id
label_str = processor.decode(labels, skip_special_tokens=True)
print(label_str)

HIPERTENSION ARTERIAL

'test_runtime': 9.9835,
```

Ilustración 17 *Tiempo del procesamiento de 100 formularios.* Fuente: Autores de este documento.

CRITERIOS	DIGITACIÓN MANUAL	MODELO
Sección 21	100 formularios	100 formularios
Tiempo empleado	60 min	10 min

Tabla 3 *Comparación de modelo con la digitación manual.* Fuente: Autores de este documento

CAPÍTULO 4

En el capítulo de conclusiones y recomendaciones, se presenta un análisis exhaustivo de los hallazgos tras aplicar un modelo basado en la arquitectura Transformers para la digitalización de textos manuscritos. Se discuten las métricas clave, subrayando la lograda precisión y la relación con los objetivos trazados al inicio. Se resaltan las fortalezas del proyecto, destacando el diseño del sistema de preprocesamiento y la adecuada recopilación de datos. Aunque se señalan oportunidades de mejora, se reconoce la contribución significativa del trabajo hacia la digitalización y análisis de documentos manuscritos. Finalmente, se ofrecen recomendaciones detalladas para futuras investigaciones, considerando aspectos

como la expansión del conjunto de datos, optimización, evaluación y aplicaciones prácticas, con la mirada puesta en potenciar aún más el impacto del trabajo realizado en este estudio.

4. CONCLUSIONES Y RECOMENDACIONES

La investigación presentada se ha enfocado en entrenar un modelo basado en la arquitectura Transformers para abordar el desafío del reconocimiento y digitalización de textos manuscritos. A partir de las métricas presentadas en los resultados, es posible discernir en qué medida se cumplieron los objetivos planteados al inicio del proyecto.

4.1 Conclusiones

4.1.1 Importancia del trabajo desarrollado y relación con el Objetivo General

Nuestro objetivo principal era entrenar un modelo Transformers para la digitalización precisa y eficiente de textos manuscritos. La pérdida de prueba obtenida, 34,42%, es un indicativo de que hemos logrado un nivel considerable de precisión. Esta cifra señala un acercamiento positivo hacia las predicciones y las etiquetas reales, esencial para la correcta digitalización de textos manuscritos, satisfaciendo en gran medida el objetivo general propuesto.

4.1.2 Fortalezas en relación con los Objetivos Específicos

La recopilación y preparación de un conjunto de datos de formularios manuscritos, evaluada a través del Error de Tasa de Caracteres (CER) de 6,91%, revela que el modelo, si bien aún mejorable, ha aprendido a reconocer adecuadamente patrones en textos manuscritos, validando la eficacia del conjunto de datos utilizado.

El diseño del sistema de preprocesamiento de imágenes parece haber contribuido de manera efectiva a la legibilidad de los textos, lo cual puede inferirse indirectamente del CER y de la pérdida de prueba.

El entrenamiento del modelo con un CER de 6,91% sugiere que se ha logrado una alta precisión en la lectura de textos. Sin embargo, cabe resaltar que aún hay espacio para la optimización de parámetros y rendimiento.

4.1.3 Debilidades

El Error de Tasa de Caracteres, aunque sobrepasa un poco el 5%, señala una oportunidad para el refinamiento adicional. Aunque se han logrado avances notables, la precisión puede aún ser incrementada para alcanzar una representación aún más fidedigna del texto manuscrito original.

4.1.4 Implicaciones del trabajo

La digitalización precisa de textos manuscritos tiene vastas aplicaciones, desde archivado digital hasta análisis de documentos históricos. Un modelo con un buen equilibrio entre precisión y eficiencia, como el que hemos desarrollado, puede revolucionar la forma en que abordamos estas tareas.

4.1.5 Trabajos futuros

Existen claras oportunidades para mejorar tanto la precisión como la eficiencia. Sería valioso, en futuras investigaciones, focalizar esfuerzos en la optimización y el afinamiento del modelo. También puede ser beneficioso explorar cómo se desempeña este modelo en condiciones más desafiantes, como textos deteriorados o con estilos de escritura variados.

En conclusión, el trabajo realizado ha sido un paso significativo en el uso de la arquitectura Transformers para el reconocimiento de textos manuscritos. Aunque aún

hay desafíos por enfrentar, los objetivos planteados al inicio del proyecto se han abordado de manera efectiva, sentando las bases para futuras investigaciones y aplicaciones prácticas en el campo.

4.2 Recomendaciones

Dada la información proporcionada y la naturaleza de la investigación, las siguientes recomendaciones pueden ser consideradas para replicar, mejorar y/o ampliar el trabajo de tesis.

4.2.1 Expansión del conjunto de datos

Si bien se ha recopilado una cantidad considerable de formularios, se recomienda recopilar aún más datos de diferentes zonas geográficas para aumentar la diversidad del conjunto de datos. Esto puede ayudar a mejorar la generalización del modelo.

En futuros estudios, se podría reconsiderar la inclusión de datos de Guayaquil para evaluar el impacto real en el sobreentrenamiento y ver cómo se puede mitigar este problema con técnicas de regularización y balanceo de datos.

4.2.2 Preprocesamiento avanzado

Considerar la aplicación de técnicas avanzadas de aumento de datos en imágenes, como rotaciones, transformaciones elásticas y ruido, para robustecer el modelo frente a variaciones en el estilo de escritura manuscrita.

4.2.3 Optimización de hiperparámetros

Se recomienda la realización de una búsqueda exhaustiva o aleatoria de hiperparámetros para encontrar la mejor configuración que aumente el rendimiento del modelo.

4.2.4 Evaluación cruzada

Implementar una validación cruzada para obtener una mejor estimación del rendimiento del modelo en diferentes divisiones de los datos.

4.2.5 Modelos comparativos

Sería útil entrenar y evaluar otros modelos de reconocimiento de texto, aparte de TrOCR, para tener una comparación directa y determinar cuál es el más adecuado para esta tarea específica.

4.2.6 Post-procesamiento

Considerar el uso de un modelo de lenguaje adicional post-entrenamiento, especialmente si el CER sigue siendo alto. Este modelo puede corregir errores gramaticales o de vocabulario basándose en el contexto del texto generado.

4.2.7 Consideraciones éticas

Asegurarse de que toda la información personal en los formularios sea completamente anónima antes de su uso, para proteger la privacidad de los individuos.

4.2.8 Análisis detallado de errores

Realizar un análisis en profundidad de las predicciones incorrectas del modelo para identificar patrones comunes de errores y áreas potenciales de mejora.

4.2.9 Implementación en la vida real

Desarrollar una interfaz o aplicación que permita a los usuarios cargar imágenes de formularios manuscritos y recibir transcripciones digitales, facilitando la digitalización de registros en entornos prácticos.

4.2.10 Actualización del modelo

Es recomendable estar al tanto de las actualizaciones y nuevos modelos en el campo de la inteligencia artificial y el procesamiento de imágenes. Estas actualizaciones podrían ser integradas para mejorar aún más el rendimiento del sistema.

Estas recomendaciones tienen como objetivo potenciar el impacto y la utilidad del trabajo de tesis, y servir de base para futuras investigaciones en el mismo ámbito.

BIBLIOGRAFÍA

- Asanza, W. R. (2017). *Redes neuronales artificiales aplicadas al reconocimiento de patrones*. <http://repositorio.utmachala.edu.ec/bitstream/48000/14223/1/Cap.1-Generalidades%20de%20las%20redes%20neuronales%20artificiales.pdf>
- INCYTU. (marzo de 2018). *Infomación científica y tecnológica para el congreso de la unión*.
https://www.foroconsultivo.org.mx/INCYTU/documentos/Completa/INCYTU_18-012.pdf
- INEC. (2020). *Instituto Nacional de Estadística y Censos*.
<https://www.ecuadorencifras.gob.ec/objetivos-politicas/#:~:text=El%20Instituto%20Nacional%20de%20Estad%3%ADstica,de%20calidad%3B%20e%2C%20innovar%20en>
- Londoño, P. (3 de abril de 2023). *HubSpot*. <https://blog.hubspot.es/website/que-es-un-formulario>
- Olabe, X. B. (Diciembre de 2008).
https://ocw.ehu.eus/pluginfile.php/40137/mod_resource/content/1/redes_neuro/contenidos/pdf/libro-del-curso.pdf
- Rodrigo, J. A. (julio de 2021). <https://www.cienciadedatos.net/documentos/68-redes-neuronales-r>
- Rouhiainen, L. (2018). *Inteligencia artificial 101 cosas que debes saber hoy sobre nuestro futuro*. Barcelona: alienta editorial.
- Serrano, A. G. (2012). *Inteligencia Artificial Fundamentos, prácticas y aplicaciones*. Madrid.
https://books.google.es/books?id=WDuqquRP70UC&lpg=PP9&ots=iUV7f_gqDx&dq=que%20es%20la%20inteligencia%20artificial&lr&hl=es&pg=PP4#v=onepage&q&f=false
- UNESCO. (2020). https://en.unesco.org/sites/default/files/ia_triptico_l4_web.pdf
- Unidas, N. (s.f.). *Naciones Unidas*.
<https://www.un.org/sustainabledevelopment/es/health/>