

UNIVERSIDAD FRANCISCO DE VITORIA

**ESCUELA POLITÉCNICA SUPERIOR**

**GRADO EN INGENIERÍA INFORMÁTICA**

PROYECTO FINAL DE GRADO

MODALIDAD INGENIERÍA

**Diagnóstico de TEA en edades infantiles mediante el uso de redes neuronales profundas**

David Molero Peña  
Convocatoria de Julio 2023

**CALIFICACIÓN DEL PROYECTO FINAL DE GRADO**

|  |  |
| --- | --- |
| CUALITATIVA: |  |
| NUMÉRICA: |  |

|  |  |
| --- | --- |
| Conforme Presidente: | Conforme Secretario: |
|  |  |
| Fdo.: | Fdo.: |

|  |  |  |
| --- | --- | --- |
| Conforme Vocal: | Conforme Vocal: | Conforme Vocal: |
|  |  |  |
| Fdo.: | Fdo.: | Fdo.: |

Lugar y fecha: Pozuelo de Alarcón, a \_\_\_\_ de \_\_\_\_\_\_\_\_\_\_\_\_\_\_ de 202\_\_

*Aquí iría una o varias citas o frases célebres que tengan relación con el proyecto (opcional).*

*Si no hubiera, debe eliminarse esta sección.*

*Aquí iría la dedicatoria (opcional).*

*Si no hubiera, debe eliminarse esta sección.*

Agradecimientos

Aquí se incluirían los agradecimientos personales y profesionales. No olvidarse de agradecer la ayuda recibida, especialmente si se ha realizado el proyecto gracias a una beca, contrato o algún tipo de subvención o apoyo.

Este apartado es opcional. Si no hubiera agradecimientos, debe eliminarse esta sección. Sin embargo, pensad bien pues siempre hay a quien dar las gracias por nuestros logros personales.

Resumen

Aquí viene el resumen en español, entre 100 y 200 palabras. Muy importante, esta sección debe ser un resumen de la memoria, es decir, del documento que se presenta, y no solo de la parte técnica del proyecto.

Palabras claves

TEA (Trastorno del Espectro Autista)

Abstract

Keywords

Índice de Contenidos

[1. Introducción 1](#_Toc136804491)

[2. Investigación previa 3](#_Toc136804492)

[3. Objetivos 7](#_Toc136804493)

[3.1. Objetivo general 7](#_Toc136804494)

[3.2. Lista de objetivos específicos 7](#_Toc136804495)

[3.3. Métodos de Validación 8](#_Toc136804496)

[4. Plan de Desarrollo del Proyecto 9](#_Toc136804497)

[4.1. Metodología 9](#_Toc136804498)

[4.2. Tecnologías 10](#_Toc136804499)

[4.3. Plan de desarrollo del proyecto 11](#_Toc136804500)

[4.3.1. PT1: Análisis e investigación previa. 11](#_Toc136804501)

[4.3.2. PT2: Definición de requerimientos. 13](#_Toc136804502)

[4.3.3. PT3: Diseño 16](#_Toc136804503)

[4.3.4. PT4: Implementación 17](#_Toc136804504)

[4.3.5. PT5: Pruebas y validación 20](#_Toc136804505)

[4.4. Plan de Trabajo 22](#_Toc136804506)

[4.4.1. Plan de Trabajo Inicial 22](#_Toc136804507)

[4.4.2. Plan de Trabajo Final 23](#_Toc136804508)

[4.5. Recursos. 24](#_Toc136804509)

[4.6. Costes 24](#_Toc136804510)

[4.7. Condicionantes y Limitaciones 25](#_Toc136804511)

[5. Desarrollo de la Solución Técnica 27](#_Toc136804512)

[5.1. PT1: Análisis e investigación previa. 27](#_Toc136804513)

[5.1.1. PT1-A1: Investigar trastornos del espectro autista (TEA) 27](#_Toc136804514)

[5.1.2. PT1-A2: Analizar fuentes de información médica 30](#_Toc136804515)

[5.1.3. PT1-A3: Investigar técnicas de aprendizaje automático y redes neuronales 33](#_Toc136804516)

[5.2. PT2 33](#_Toc136804517)

[5.2.1. PT2-A1: Definir objetivos específicos y generales 33](#_Toc136804518)

[5.2.2. PT2-A3: Establecer métricas de evaluación 33](#_Toc136804519)

[5.3. PT3 36](#_Toc136804520)

[5.3.1. PT3-A1: Definir el flujo de datos 36](#_Toc136804521)

[5.4. PT4 46](#_Toc136804522)

[5.4.1. PT4-A1: Preparar los datos 46](#_Toc136804523)

[5.4.2. PT4-A2: Desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales 47](#_Toc136804524)

[6. Resultados 53](#_Toc136804525)

[7. Implicaciones Éticas e Impacto Social 55](#_Toc136804526)

[8. Conclusiones 55](#_Toc136804527)

[9. Otros Méritos del Proyecto 57](#_Toc136804528)

[10. Bibliografía 59](#_Toc136804529)

[Anexo A: Planificación detallada 61](#_Toc136804530)

[Anexo B: Requisitos de Usuario 62](#_Toc136804531)

[Anexo B: Diseño del Sistema 63](#_Toc136804532)

[Anexo C: Manuales 65](#_Toc136804533)

[Manual de Usuario 65](#_Toc136804534)

[Manual de Instalación 65](#_Toc136804535)

[Anexo D: Contenido del CD 67](#_Toc136804536)

Índice de Tablas

[Tabla 1. Investigar trastornos del espectro autista (TEA) 11](#_Toc136807395)

[Tabla 2. Analizar fuentes de información médica 12](#_Toc136807396)

[Tabla 3. Investigar técnicas de aprendizaje automático y redes neuronales 13](#_Toc136807397)

[Tabla 4. Definir objetivos específicos y generales 13](#_Toc136807398)

[Tabla 5. Determinar técnicas de aprendizaje automático y redes neuronales 14](#_Toc136807399)

[Tabla 6. Establecer métricas de evaluación 15](#_Toc136807400)

[Tabla 7. Definir el flujo de datos 15](#_Toc136807401)

[Tabla 8. Diseñar la interfaz de usuario 16](#_Toc136807402)

[Tabla 9. Preparar los datos 17](#_Toc136807403)

[Tabla 10. Desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales 17](#_Toc136807404)

Índice de Ilustraciones

[Ilustración 1. Herramienta de tamizaje orientada para niños y adolescentes entre 4 y 17 años con características para alertar al diagnóstico clínico por Saturdays.ai 5](file:////Users/deivddds/Desktop/ASDdiagnosis/documents/memorias/PFG%20Memoria_DavidMoleroPeña_5.0.docx#_Toc136804919)

Lista de Acrónimos

|  |  |
| --- | --- |
| **Acrónimo** | **Significado** |
| TEA | Trastornos del espectro autista |
| AI | Artificial Intelligence |
| PPV | Positive Predictive Value |
| NPV | Negative Predictive Value |
| M-CHAT | Modified Checklist for Autism in Toddlers |
| ADOS | Autism Diagnostic Observation Schedule |
| CSR | Comunicación Social Recíproca |
| CRRE | Comportamientos Restringidos, Repetitivos y Estereotipados |

# Introducción

El enfoque de este proyecto radica en la optimización del diagnóstico temprano y preciso de niños con Trastorno del Espectro Autista (TEA) o síntomas sospechosos. Un diagnóstico oportuno es esencial para proporcionar intervenciones adecuadas y apoyo a estos niños y sus familias.

Al combinar la experiencia de profesionales médicos y psicólogos con avances en inteligencia artificial y aprendizaje automático, el objetivo principal del proyecto es mejorar el proceso de diagnóstico de niños con TEA. Para ello, se desarrolla una plataforma que utiliza redes neuronales convolucionales (CNN) en la realización de diagnósticos precisos y efectivos. Además, la plataforma también facilita la gestión y el mantenimiento de registros de los datos de los niños y sus diagnósticos por parte de los profesionales.

Esta plataforma no solo proporcionará herramientas de diagnóstico basadas en IA, sino que también ofrecerá a los profesionales una forma eficiente de gestionar y mantener un registro de los datos de los niños y sus diagnósticos a lo largo del tiempo. De esta manera, facilitara el trabajo de los profesionales en el campo del TEA, permitiéndoles tomar decisiones informadas y basadas en datos concretos para brindar un mejor apoyo a los niños afectados y sus familias.

Como proyecto de final de grado, abarcará una amplia gama de aspectos, incluida la investigación sobre el TEA y sus características, el desarrollo de algoritmos de aprendizaje profundo y el entrenamiento de modelos de CNN, así como la creación de una interfaz de usuario intuitiva y fácil de usar que permita a los profesionales acceder a las herramientas de diagnóstico y gestión de datos de manera eficiente.

En última instancia, el éxito de este proyecto se medirá por su capacidad para mejorar la precisión y efectividad del diagnóstico del TEA y proporcionar a los profesionales una herramienta valiosa que facilite su trabajo y mejore la calidad de vida de los niños afectados por este trastorno y sus familias.

# Investigación previa

El Trastorno del Espectro Autista (TEA) es un trastorno neurobiológico del desarrollo que afecta la comunicación, el comportamiento y las interacciones sociales de la persona. Las personas con TEA suelen tener dificultades para comunicarse y relacionarse con los demás, así como patrones repetitivos de comportamiento, intereses restringidos y dificultades sensoriales. El TEA es un trastorno del espectro, lo que significa que puede afectar a las personas de diferentes maneras y en diferentes grados.

Los síntomas del TEA pueden manifestarse tempranamente, a menudo antes de los 2 años de edad. Algunos de los signos tempranos de TEA incluyen una falta de interés en las personas, falta de contacto visual, problemas de comunicación, como la falta de balbuceo o la falta de palabras, y patrones de juego repetitivos. Sin embargo, a menudo el diagnóstico del TEA no se realiza hasta los 4 o 5 años de edad.

Los métodos de diagnóstico actuales del TEA incluyen la evaluación clínica, la observación del comportamiento y la entrevista con los padres y cuidadores. Sin embargo, estos métodos tienen limitaciones, ya que pueden ser subjetivos y dependen de la experiencia y el juicio clínico del profesional de la salud mental. Además, el diagnóstico tardío del TEA puede retrasar la intervención temprana y disminuir su efectividad.

La detección temprana del TEA es crucial para lograr una intervención a tiempo y mejorar los resultados a largo plazo para los individuos con TEA. Este diagnóstico temprano puede permitir a los profesionales de la salud y educación identificar las necesidades específicas del niño y diseñar un plan de intervención individualizado. Esto puede ayudar a maximizar el potencial de desarrollo del niño y niña y a mejorar su calidad de vida a largo plazo.

Otro factor importante es que, a medida que los niños con TEA crecen, pueden enfrentar dificultades en las relaciones sociales y la comunicación, lo que puede afectar negativamente su desarrollo académico y su capacidad para desenvolverse en la vida cotidiana. Un diagnóstico temprano puede permitir la implementación de medidas preventivas y tratamientos para abordar estos desafíos y maximizar el éxito del niño en la escuela y en la vida adulta. [1]

Un modelo predictivo para detectar el TEA en edades tempranas puede ser una manera efectiva de abordar las limitaciones de los métodos de diagnóstico actuales y mejorar la detección temprana del TEA.

Los proyectos de Inteligencia Artificial para la detección de TEA incluyen aplicaciones tecnológicas para hacer evaluación y diagnóstico de TEA de forma más precisa y en edades más tempranas**.** Estas aplicaciones utilizan algoritmos automatizados para detectar los detalles que pueden favorecer el diagnóstico temprano y objetivo del autismo.

El Proyecto T-Eye es un sistema de monitorización para niños con Trastorno del Espectro Autista basado en inteligencia artificial y medidas fisiológicas. El proyecto está financiado por el Centro para el Desarrollo Tecnológico Industrial (CDTI) del Ministerio de Economía, Industria y Competitividad. El objetivo del proyecto es investigar el bienestar psico-emocional, las necesidades y el grado de aceptación de los niños con TEA utilizando un nuevo método de diagnóstico para niños con autismo entre 3 y 6 años mediante realidad virtual y biomarcadores fisiológicos. Las sesiones tienen lugar en un entorno amigable llamado T-Room, donde los niños juegan a diferentes juegos con Pablo, un avatar. Los niños usarán una pulsera para medir su actividad electrodermal, un proceso neuropsicológico que hace referencia a las propiedades eléctricas de la piel como respuesta de sudoración ante un determinado estímulo, y gafas ligeras de Eye Tracking para controlar el seguimiento de la mirada. Con técnicas de AI, los investigadores han logrado discriminar entre comportamientos normales y patológicos en los niños con TEA. [2]

Estudios publicados en Nature han probado la eficacia de SaMD (Software as a Medical Device) una aplicación que evalúa el TEA en niños entre 18 y 72 meses. Este estudio probó la precisión de un software basado en inteligencia artificial como un dispositivo médico diseñado para ayudar a los profesionales de atención primaria en el diagnóstico del TEA. El dispositivo combina características conductuales de tres entradas distintas (un cuestionario para cuidadores, el análisis de dos cortos videos caseros y un cuestionario para el profesional) en un algoritmo de aprendizaje automático basado en árboles de decisión incrementales para diagnosticar una salida positiva de TEA, una salida negativa de TEA o una salida indeterminada. Este estudio comparó las salidas del dispositivo con el acuerdo diagnóstico de dos o más especialistas independientes en niños con preocupaciones por retrasos en el desarrollo.

El PPV de la salida del dispositivo para todos los participantes completos del estudio fue del 80.8% y el NPV fue del 98.3%. Para el 31.8% de los participantes que recibieron una salida determinada (positiva o negativa para TEA), la sensibilidad del dispositivo fue del 98.4% y la especificidad fue del 78.9%. La salida indeterminada del dispositivo actúa como una medida de control de riesgos cuando las entradas no son lo suficientemente detalladas para hacer una recomendación determinada con confianza.

Si se eliminara esta medida de control de riesgos, la sensibilidad para todos los participantes del estudio caería al 51.6% (y la especificidad caería al 18.5%. Entre los participantes a los que el dispositivo se abstuvo de proporcionar un resultado, los especialistas identificaron que el 91% tenía uno o más trastornos complejos del neurodesarrollo. El dispositivo muestra promesas para aumentar significativamente el número de niños que pueden ser diagnosticados con TEA en un entorno de atención primaria, lo que podría facilitar una intervención más temprana y un uso más eficiente de los recursos especializados. [3]

Continuando con otros proyectos destinados al diagnóstico de TEA, la empresa Neurometrics ha desarrollado un prototipo de aplicación web que combina la utilización de datos de seguimiento ocular “eyetracking” y algoritmos de Machine Learning. Con esto se logra generar un pre-diagnóstico del Trastorno del Espectro Autista (TEA) en niños de 1 a 6 años. El proyecto, de nombre BRILAB, genera en cuestión de minutos un reporte en línea mediante una prueba estandarizada, con lo cual los niños que así lo requieran pueden ser derivados a un médico especialista oportunamente.

El reporte generado a partir de BRILAB se compone de una serie de indicadores. Estos están basados en los patrones de lectura del niño, que muestran la probabilidad y escala en la que se podría ubicar cada uno en base a un algoritmo de Machine Learning. Algoritmo que fue entrenado para tener la capacidad de distinguir entre un niño neurotípico y uno con autismo. [4]

Además, Saturdays.ai, una organización que capacita a las personas para aprender Inteligencia Artificial de forma colaborativa y basada en proyectos, ha desarrollado una herramienta de tamizaje orientada para niños y adolescentes entre 4 y 17 años con características para alertar al diagnóstico clínico. Estudios piloto han validado el Cuestionario de Comunicación Social (SCQ) como instrumento para la detección del TEA en muestras ecuatorianas[**[3]**](https://saturdays.ai/2021/11/19/inteligencia-artificial-diagnostico-trastornos-desarrollo). Los resultados obtenidos fueron prometedores, mostrando porcentajes aceptables tanto en casos con TEA como sin él. [5]

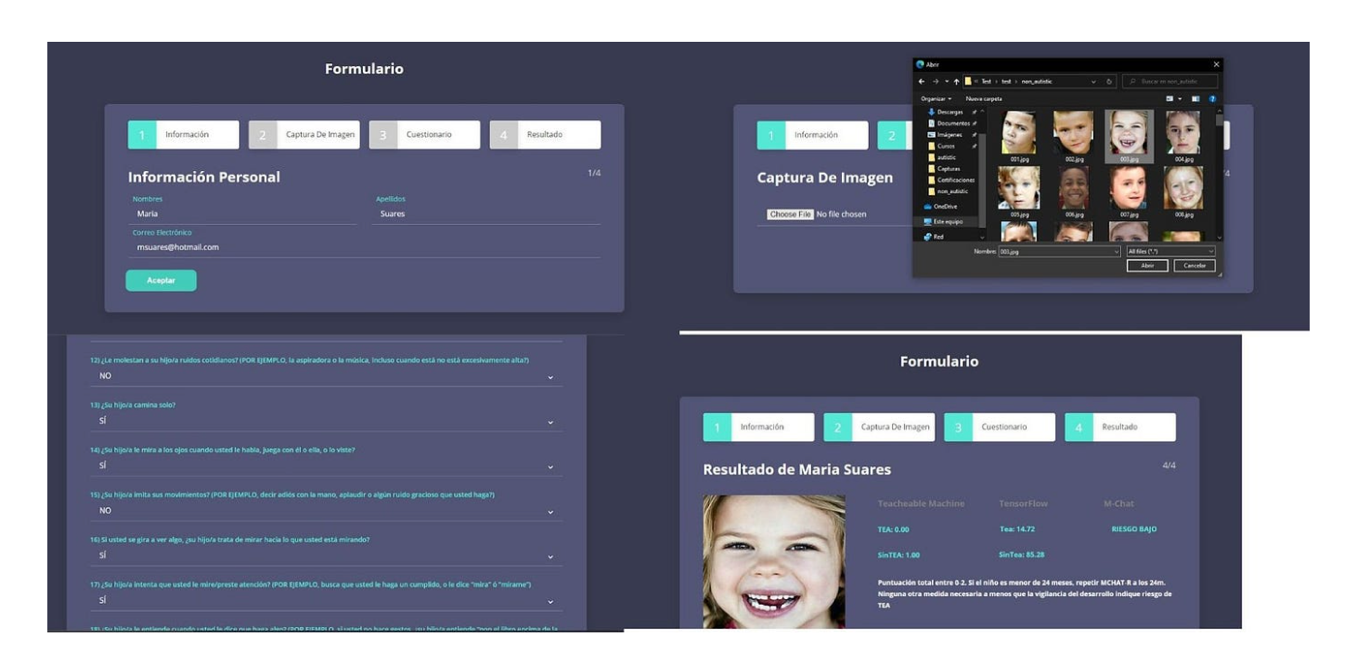


Ilustración 1. Herramienta de tamizaje orientada para niños y adolescentes entre 4 y 17 años con características para alertar al diagnóstico clínico por Saturdays.ai

Streamlit es un framework de código abierto para crear aplicaciones web interactivas de manera rápida y sencilla en Python. Con Streamlit, los desarrolladores pueden crear interfaces de usuario de alta calidad para visualizar datos, presentar modelos de aprendizaje automático y crear prototipos de aplicaciones web. La principal ventaja de Streamlit es su enfoque en la simplicidad. [6]

tf.keras ofrece una interfaz de programación de aplicaciones (API) para construir y entrenar modelos de aprendizaje profundo en TensorFlow. Proporciona una amplia gama de herramientas para la construcción de modelos, el procesamiento de datos, el entrenamiento y la evaluación, y la predicción de resultados en unas pocas líneas de código. tf.keras también admite una variedad de arquitecturas de red neuronal, como redes neuronales convolucionales, redes neuronales recurrentes, y redes neuronales de retroalimentación. [7]

Hay que organizar esto que lo tiene un poco “desparramado” y terminar con alguna frase que nos ligue con el siguiente apartado.

# Objetivos

## Objetivo general

El objetivo principal de este proyecto es mejorar el diagnóstico temprano de TEA en niños, lo que permitiría una intervención temprana y una atención médica adecuada. Se busca utilizar información de los profesionales implicados para desarrollar un sistema de diagnóstico más preciso y efectivo.

## Lista de objetivos específicos

En relación con el objetivo general se han definido los siguientes objetivos SMART:

1. Identificar y evaluar los factores de riesgo y las características clínicas asociadas con el trastorno del espectro autista, con el objetivo de mejorar la identificación temprana y la comprensión de los síntomas.
2. Promover una mayor conciencia y educación sobre los trastornos del espectro autista, tanto entre los profesionales de la salud como en la población en general, para que se pueda fomentar una detección temprana y una atención médica adecuada y fomentar la colaboración entre diferentes disciplinas y profesionales de la salud, como pediatras, neurólogos, psiquiatras y psicólogos, para mejorar la evaluación y el diagnóstico de los trastornos del espectro autista.
3. Desarrollar y mejorar las herramientas y los protocolos de diagnóstico existentes para el trastorno del espectro autista, con el objetivo de mejorar la precisión y la fiabilidad del diagnóstico. Esto podría incluir la revisión de las guías de diagnóstico existentes, la incorporación de nuevas herramientas y pruebas, y la mejora de la capacitación de los profesionales de la salud en la evaluación de los síntomas y la realización del diagnóstico.
4. Promover estrategias de intervención temprana para niños con trastornos del espectro autista, con el objetivo de mejorar su calidad de vida y su capacidad para desarrollarse plenamente.

## Métodos de Validación

Para cada uno de los objetivos específicos identificados se ha definido un método de validación:

1. Para identificar y evaluar los factores de riesgo y las características clínicas asociadas con el trastorno del espectro autista …
2. Para promover una mayor conciencia y educación sobre los trastornos del espectro autista …
3. Para desarrollar y mejorar las herramientas y los protocolos de diagnóstico existentes para el trastorno del espectro autista …
4. Para promover estrategias de intervención temprana para niños con trastornos del espectro autista …

# Plan de Desarrollo del Proyecto

## Metodología

La metodología en cascada es un enfoque de desarrollo de software que sigue una secuencia lineal y rígida de etapas, desde la definición de requerimientos hasta la implementación y mantenimiento. Estas etapas se ejecutan de manera secuencial, y cada etapa debe completarse antes de pasar a la siguiente.

En el contexto de este proyecto, se ha elegido la metodología en cascada por las siguientes razones:

1. **Estructura clara y predecible**: La metodología de cascada proporciona un proceso bien estructurado y fácil de entender que divide el proyecto en etapas claramente definidas. Esto facilita la planificación y el seguimiento del progreso del proyecto.
2. **Enfoque en la documentación**: La metodología de cascada hace hincapié en la documentación detallada en cada etapa del proceso, lo que puede resultar beneficioso para mantener un registro claro de los objetivos, requisitos y decisiones tomadas a lo largo del proyecto.
3. **Adecuada para proyectos con requisitos bien definidos:** Dado que este proyecto tiene un objetivo claro (mejorar el diagnóstico temprano de TEA en niños mediante el uso de redes neuronales profundas y técnicas de aprendizaje automático), la metodología de cascada puede ser adecuada para garantizar que se cumplan todos los requisitos de manera sistemática.
4. **Facilita la estimación de tiempo y recursos**: Como el proyecto se divide en etapas lineales y secuenciales, resulta más fácil estimar el tiempo y los recursos necesarios para cada fase.

Sin embargo, también es importante tener en cuenta las desventajas de la metodología de cascada, especialmente en relación con este proyecto:

1. **Falta de flexibilidad**: La metodología de cascada es rígida en cuanto a la secuencia de etapas y puede ser difícil de adaptar si surgen cambios en los requisitos o nuevos descubrimientos durante el proceso de desarrollo. En el campo de aprendizaje automático, esto podría ser un problema si se requieren ajustes en las técnicas utilizadas o en la selección de datos.
2. **Retrasos en la detección de problemas**: Dado que la etapa de pruebas se realiza después de la implementación, es posible que los problemas no se detecten hasta más tarde en el proceso de desarrollo, lo que podría aumentar los costes y el tiempo requerido para solucionarlos.

## Tecnologías

En el contexto de este proyecto, las tecnologías principales con las que se trabaja serán:

* **TensorFlow**: TensorFlow es una biblioteca de código abierto desarrollada por Google Brain Team para desarrollar y entrenar modelos de aprendizaje automático y redes neuronales profundas. Es especialmente útil para trabajar con redes neuronales y facilita el desarrollo, entrenamiento y validación de modelos de aprendizaje profundo.
* **Keras**: Keras es una interfaz de alto nivel para TensorFlow que facilita la creación y el entrenamiento de modelos de aprendizaje profundo. Keras simplifica el proceso de construcción de redes neuronales al proporcionar bloques de construcción predefinidos y abstraer detalles de implementación de bajo nivel.
* **Scikit-learn**: Scikit-learn es una biblioteca de aprendizaje automático de código abierto para Python que proporciona herramientas simples y eficientes para el análisis de datos y la minería de datos. Ofrece funciones para preprocesamiento de datos, selección de características, ajuste de modelos y evaluación del rendimiento.
* **NumPy y Pandas**: NumPy es una biblioteca de Python para trabajar con matrices y cálculos matemáticos multidimensionales, mientras que Pandas es una biblioteca de manipulación y análisis de datos que proporciona estructuras de datos y funciones necesarias para trabajar con datos estructurados. Ambas bibliotecas son fundamentales para el manejo y procesamiento de datos en este proyecto.
* **OpenCV**: OpenCV (Open Source Computer Vision Library) es una biblioteca de código abierto que incluye varias funciones y algoritmos para el procesamiento de imágenes y visión por computadora. Si se utilizan imágenes médicas como parte de las fuentes de información, OpenCV puede ser útil para preprocesar y analizar dichas imágenes.
* **Streamlit**: Streamlit es un marco de código abierto para crear aplicaciones web interactivas y rápidas utilizando Python. Permite desarrollar interfaces de usuario intuitivas y atractivas con una cantidad mínima de código. Streamlit se utilizará para desarrollar la interfaz del sistema de diagnóstico de TEA.
* **Jupyter Notebook**: Jupyter Notebook es una aplicación web de código abierto que permite la creación y el intercambio de documentos que contienen código en vivo, ecuaciones, visualizaciones y texto narrativo. Esta herramienta es útil para la experimentación, el análisis y la presentación de resultados en el campo del aprendizaje automático y la ciencia de datos.

## Plan de desarrollo del proyecto

### PT1: Análisis e investigación previa.

En esta fase se realizará una investigación exhaustiva de los trastornos del espectro autista (TEA), con el objetivo de profundizar en su conocimiento y entender los desafíos que implica su diagnóstico. Además, se identificarán técnicas de aprendizaje automático y redes neuronales que puedan resultar útiles para abordar el problema. Los resultados obtenidos en esta etapa proporcionarán una sólida base para el diseño y la implementación del sistema.

**PT1-A1: Investigar trastornos del espectro autista (TEA)**

En esta actividad, se investigarán los trastornos del espectro autista y los desafíos del diagnóstico para adquirir un conocimiento sólido sobre el tema. Se revisará la literatura científica y médica relacionada con TEA y se identificarán los desafíos en el diagnóstico de TEA. El resultado será un documento con información relevante sobre TEA y los desafíos del diagnóstico, lo que servirá como base para el resto del proyecto.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT1-A1 | **Nombre** | Investigar trastornos del espectro autista (TEA) |
| **Descripción** | Profundizar en el conocimiento de los trastornos del espectro autista y los desafíos del diagnóstico. | | |
| **Entradas** | Fuentes de información médica | | |
| **Salidas** | Apartado con información relevante sobre TEA y desafíos de diagnóstico | | |
| **Tareas** | * Revisar literatura científica y médica relacionada con TEA * Identificar desafíos en el diagnóstico de TEA | | |
| **Duración** | 1 semana (1-7 de enero de 2023) | | |

Tabla 1. Investigar trastornos del espectro autista (TEA)

**PT1-A2: Analizar fuentes de información médica**

Esta actividad implica investigar y seleccionar las fuentes de información médica más relevantes y útiles para el proyecto. Se investigarán fuentes relacionadas con TEA y se seleccionarán las más adecuadas en función de su relevancia y utilidad. El resultado será un listado de fuentes de información médica seleccionadas que se utilizarán en las siguientes fases del proyecto.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT1-A2 | **Nombre** | Analizar fuentes de información médica. |
| **Descripción** | Investigar y seleccionar las fuentes de información médica más relevantes y útiles para el proyecto. | | |
| **Entradas** | Conocimientos sobre trastornos del espectro autista (TEA). | | |
| **Salidas** | Listado de fuentes de información médica seleccionadas. | | |
| **Tareas** | * Investigar fuentes de información médica relacionadas con TEA * Seleccionar las fuentes de información más relevantes y útiles | | |
| **Duración** |  | | |

Tabla 2. Analizar fuentes de información médica

**PT1-A3: Investigar técnicas de aprendizaje automático y redes neuronales**

En esta actividad, se investigarán y seleccionarán las técnicas de aprendizaje automático y redes neuronales profundas más apropiadas para el diagnóstico de TEA. Se revisará la literatura científica y técnica relacionada con el aprendizaje automático y las redes neuronales aplicadas a diagnósticos médicos. El resultado será un listado de técnicas de aprendizaje automático y redes neuronales seleccionadas que se utilizarán en el diseño y la implementación del sistema de diagnóstico de TEA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT1-A3 | **Nombre** | Investigar técnicas de aprendizaje automático y redes neuronales. |
| **Descripción** | Investigar y seleccionar las técnicas de aprendizaje automático y redes neuronales profundas más apropiadas para el diagnóstico de TEA. | | |
| **Entradas** | Conocimientos en aprendizaje automático y redes neuronales, desafíos de diagnóstico de TEA identificados. | | |
| **Salidas** | Listado de técnicas de aprendizaje automático y redes neuronales seleccionadas. | | |
| **Tareas** | * Revisar literatura científica y técnica relacionada con aprendizaje automático y redes neuronales aplicadas a diagnósticos médicos * Seleccionar las técnicas de aprendizaje automático y redes neuronales más apropiadas para el diagnóstico de TEA | | |
| **Duración** |  | | |

Tabla 3. Investigar técnicas de aprendizaje automático y redes neuronales

### PT2: Definición de requerimientos.

En esta fase se definirán los objetivos específicos y generales del proyecto, así como las métricas de evaluación que se utilizarán para medir el rendimiento del sistema de diagnóstico. También se seleccionarán las técnicas de aprendizaje automático y redes neuronales más adecuadas para el proyecto en función de los objetivos definidos.

**PT2-A1: Definir objetivos específicos y generales**

Esta actividad consiste en establecer los objetivos generales y específicos del proyecto. Se definirá el objetivo general y se establecerán objetivos específicos en función de la investigación previa realizada, la información relevante sobre TEA y los desafíos del diagnóstico, y las técnicas de aprendizaje automático y redes neuronales seleccionadas. El resultado será un listado de objetivos generales y específicos que guiarán el desarrollo del proyecto.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT2-A1 | **Nombre** | Analizar fuentes de información médica. |
| **Descripción** | Establecer objetivos y requerimientos específicos del proyecto. | | |
| **Entradas** | Investigación previa realizada, información relevante sobre TEA y desafíos de diagnóstico | | |
| **Salidas** | Listado de objetivos generales y específicos. | | |
| **Tareas** | * Definir el objetivo general * Definir objetivos específicos | | |
| **Duración** |  | | |

Tabla 4. Definir objetivos específicos y generales

**PT2-A2: Determinar técnicas de aprendizaje automático y redes neuronales**

En este apartado, se evaluarán y seleccionarán las técnicas de aprendizaje automático y redes neuronales más adecuadas y eficientes para el diagnóstico de TEA. Este proceso incluye la revisión de las técnicas investigadas en función de los objetivos específicos y los desafíos de diagnóstico de TEA. Al final de esta actividad, se dispondrá de una lista final de técnicas seleccionadas para implementar en el sistema de diagnóstico.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT2-A2 | **Nombre** | Determinar técnicas de aprendizaje automático y redes neuronales |
| **Descripción** | Seleccionar las técnicas de aprendizaje automático y redes neuronales profundas a utilizar en el proyecto. | | |
| **Entradas** | Objetivos específicos, conocimientos en aprendizaje automático y redes neuronales, listado de técnicas de aprendizaje automático y redes neuronales investigadas. | | |
| **Salidas** | Listado final de técnicas de aprendizaje automático y redes neuronales seleccionadas. | | |
| **Tareas** | * Evaluar las técnicas de aprendizaje automático y redes neuronales investigadas en función de los objetivos específicos y los desafíos de diagnóstico de TEA. * Seleccionar las técnicas de aprendizaje automático y redes neuronales más adecuadas y eficientes para el diagnóstico de TEA. | | |
| **Duración** |  | | |

Tabla 5. Determinar técnicas de aprendizaje automático y redes neuronales

**PT2-A3: Establecer métricas de evaluación**

En esta actividad, se definirán las métricas de evaluación del sistema de diagnóstico. Se investigarán las métricas de evaluación relevantes para el diagnóstico de TEA y se seleccionarán las más adecuadas en función de los objetivos específicos y las técnicas seleccionadas. El resultado será un listado de métricas de evaluación que se utilizarán para evaluar el rendimiento del sistema de diagnóstico de TEA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT2-A3 | **Nombre** | Establecer métricas de evaluación. |
| **Descripción** | Definir las métricas de evaluación del sistema de diagnóstico. | | |
| **Entradas** | Objetivos específicos, técnicas de aprendizaje automático y redes neuronales seleccionadas. | | |
| **Salidas** | Listado de métricas de evaluación. | | |
| **Tareas** | * Investigar métricas de evaluación relevantes para el diagnóstico de TEA * Seleccionar las métricas de evaluación más adecuadas en función de los objetivos específicos y las técnicas seleccionadas | | |
| **Duración** |  | | |

Tabla 6. Establecer métricas de evaluación

### PT3: Diseño

En esta fase se llevará a cabo el diseño de la arquitectura de la red neuronal profunda y se establecerá el flujo de datos para integrar las fuentes de información médica con la red. También se diseñará la interfaz de usuario que permitirá a los profesionales médicos acceder y utilizar el sistema de diagnóstico.

**PT3-A1: Definir el flujo de datos**

El objetivo de esta actividad es establecer cómo se integrarán las distintas fuentes de información con la red neuronal y con la plataforma. Se identificarán los procesos de preprocesamiento de datos necesarios y se diseñará el flujo de datos entre las fuentes de información y la red neuronal. Al final de esta actividad, se dispondrá de un diagrama de flujo de datos que guiará el proceso de preparación y uso de los datos en el sistema de diagnóstico.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT3-A2 | **Nombre** | Definir el flujo de datos. |
| **Descripción** | Establecer cómo se integrarán las distintas fuentes de información con la red neuronal y con la plataforma. | | |
| **Entradas** | Tipo de red neuronal, fuentes de información médica. | | |
| **Salidas** | Diagrama de flujo de datos. | | |
| **Tareas** | * Identificar los procesos de preprocesamiento de datos necesarios * Diseñar el flujo de datos entre las fuentes de información y la red neuronal | | |
| **Duración** |  | | |

Tabla 7. Definir el flujo de datos

**PT3-A2: Diseñar la interfaz de usuario**

El objetivo de esta actividad es crear la interfaz de usuario y definir cómo se presentará la información del diagnóstico a los profesionales médicos. Se investigarán las interfaces de usuario utilizadas en sistemas de diagnóstico médico similares y se diseñará la interfaz de usuario teniendo en cuenta las necesidades de los profesionales médicos y los requerimientos del sistema. Al final de esta actividad, se dispondrá de un diseño de la interfaz de usuario que guiará el proceso de implementación.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT3-A3 | **Nombre** | Diseñar la interfaz de usuario. |
| **Descripción** | Crear la interfaz de usuario y definir cómo se presentará la información del diagnóstico a los profesionales médicos. | | |
| **Entradas** | Arquitectura de la red neuronal, flujo de datos, requerimientos del sistema. | | |
| **Salidas** | Diseño de la interfaz de usuario. | | |
| **Tareas** | * Investigar interfaces de usuario utilizadas en sistemas de diagnóstico médico similares * Diseñar la interfaz de usuario teniendo en cuenta las necesidades de los profesionales médicos y los requerimientos del sistema | | |
| **Duración** |  | | |

Tabla 8. Diseñar la interfaz de usuario

### PT4: Implementación

En esta fase se prepararán los datos y se implementará la red neuronal profunda utilizando las técnicas de aprendizaje automático y optimización seleccionadas. También se codificará la interfaz de usuario diseñada y se integrará con la red neuronal profunda.

**PT4-A1: Preparar los datos**

El objetivo de esta actividad es procesar y preparar las fuentes de datos para su uso en la red neuronal. Se preprocesará el dataset disponible y se dividirá el conjunto de datos en conjuntos de entrenamiento, validación y prueba. Al final de esta actividad, se dispondrá de un conjunto de datos procesados y listos para su uso en la red neuronal.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT4-A1 | **Nombre** | Preparar los datos. |
| **Descripción** | Procesar y preparar las fuentes de información médica para su uso en la red neuronal. | | |
| **Entradas** | Fuentes de información médica, diagrama de flujo de datos. | | |
| **Salidas** | Conjunto de datos procesados y listos para su uso en la red neuronal. | | |
| **Tareas** | * Preprocesar las fuentes de información médica de acuerdo con el flujo de datos establecido * Dividir el conjunto de datos en conjuntos de entrenamiento, validación y prueba | | |
| **Duración** |  | | |

Tabla 9. Preparar los datos

**PT4-A2: Desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales**

En esta actividad, se desarrollará e implementará un algoritmo de búsqueda automatizada que permita explorar y evaluar múltiples arquitecturas de redes neuronales. El objetivo es encontrar la arquitectura que mejor se adapte a los datos y proporcione un rendimiento óptimo en el diagnóstico de TEA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT4-A2 | **Nombre** | Desarrollar e implementar un algoritmo de búsqueda automatizada |
| **Descripción** | Desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales. | | |
| **Entradas** | Técnicas de aprendizaje automático y redes neuronales seleccionadas, conjuntos de datos de entrenamiento y validación | | |
| **Salidas** | Algoritmo de búsqueda automatizada implementado | | |
| **Tareas** | * Diseñar el algoritmo de búsqueda automatizada * Implementar el algoritmo en un entorno de programación * Integrar el algoritmo con los conjuntos de datos de entrenamiento y validación | | |
| **Duración** |  | | |

Tabla 10. Desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales

**PT4-A3: Entrenar y validar las arquitecturas de red neuronal**

Esta actividad se centra en el entrenamiento y validación de las diferentes arquitecturas de red neuronal exploradas y evaluadas en la actividad anterior (PT4-A2). El objetivo es determinar qué arquitectura de red neuronal ofrece el mejor rendimiento en términos de precisión y eficiencia en el diagnóstico de TEA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT4-A3 | **Nombre** | Entrenar y validar las arquitecturas de red neuronal |
| **Descripción** | Entrenar y validar las arquitecturas de red neuronal para determinar cuál ofrece el mejor rendimiento en términos de precisión y eficiencia en el diagnóstico de TEA. | | |
| **Entradas** | Algoritmo de búsqueda automatizada implementado, conjuntos de datos de entrenamiento y validación. | | |
| **Salidas** | Arquitectura de red neuronal óptima. | | |
| **Tareas** | * Entrenar las arquitecturas de red neuronal utilizando el conjunto de datos de entrenamiento I * Validar las arquitecturas de red neuronal utilizando el conjunto de datos de validación * Seleccionar la arquitectura de red neuronal óptima en base a los resultados obtenidos | | |
| **Duración** |  | | |

Tabla 11. Entrenar y validar las arquitecturas de red neuronal

**PT4-A4: Implementar la interfaz de usuario**

El objetivo de esta actividad es codificar la interfaz de usuario diseñada y asegurar su correcta interacción con la red neuronal profunda. Se seleccionará un lenguaje de programación y bibliotecas adecuadas para la implementación de la interfaz de usuario, se codificará la interfaz de usuario siguiendo el diseño establecido y se integrará la interfaz de usuario con la red neuronal profunda. Al final de esta actividad, se dispondrá del código fuente de la interfaz de usuario.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT4-A4 | **Nombre** | Implementar la interfaz de usuario. |
| **Descripción** | Codificar la interfaz de usuario diseñada y asegurar su correcta interacción con la red neuronal profunda | | |
| **Entradas** | Diseño de la interfaz de usuario, código fuente de la red neuronal profunda. | | |
| **Salidas** | Código fuente de la interfaz de usuario. | | |
| **Tareas** | * Codificar la interfaz de usuario siguiendo el diseño establecido * Integrar la interfaz de usuario con la red neuronal profunda | | |
| **Duración** | 1 semana (26 de marzo - 1 de abril de 2023) | | |

Tabla 12. Implementar la interfaz de usuario

### PT5: Pruebas y validación

En esta fase, se realizarán pruebas y validaciones del sistema de diagnóstico de TEA para garantizar su funcionamiento correcto y eficiente. Se evaluará el rendimiento del sistema utilizando las métricas de evaluación establecidas y se realizarán pruebas con usuarios finales para garantizar que la interfaz de usuario cumple con sus necesidades y expectativas.

**PT5-A1: Evaluar el rendimiento del sistema**

El objetivo de esta actividad es evaluar el rendimiento del sistema de diagnóstico utilizando las métricas de evaluación seleccionadas. Se aplicarán las métricas de evaluación al conjunto de datos de prueba y se analizarán los resultados para determinar la eficacia del sistema en la identificación de casos de TEA.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT5-A1 | **Nombre** | Evaluar el rendimiento del sistema. |
| **Descripción** | Evaluar el rendimiento del sistema de diagnóstico utilizando las métricas de evaluación seleccionadas. | | |
| **Entradas** | Arquitectura de red neuronal óptima, conjunto de datos de prueba, métricas de evaluación. | | |
| **Salidas** | Resultados de las métricas de evaluación aplicadas al sistema de diagnóstico. | | |
| **Tareas** | * Aplicar las métricas de evaluación al conjunto de datos de prueba * Analizar los resultados obtenidos y determinar la eficacia del sistema en la identificación de casos de TEA | | |
| **Duración** |  | | |

Tabla 13. Evaluar el rendimiento del sistema

**PT5-A2: Realizar pruebas con usuarios finales**

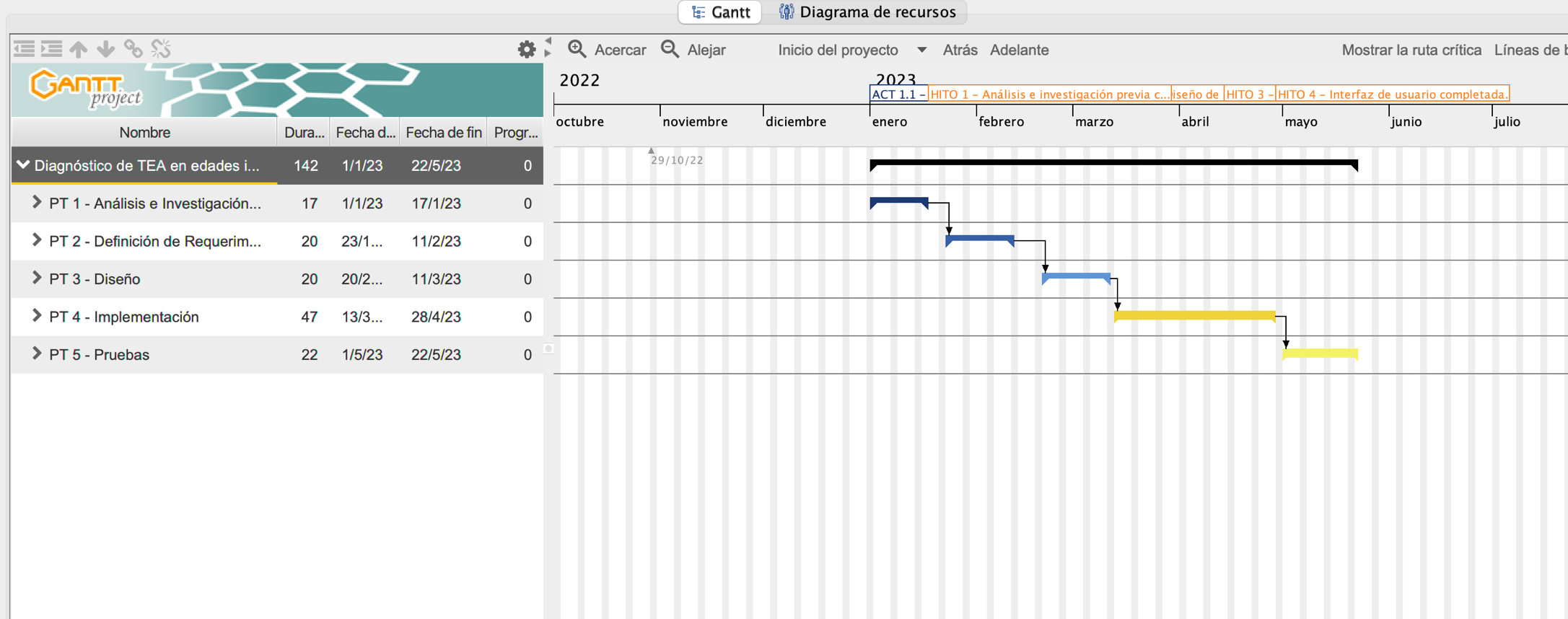
En esta actividad, se llevarán a cabo pruebas con usuarios finales (profesionales médicos) para evaluar la usabilidad y funcionalidad de la interfaz de usuario y recopilar sus opiniones y sugerencias. Estas pruebas permitirán realizar ajustes y mejoras en la interfaz de usuario y en el sistema de diagnóstico en general.

|  |  |  |  |
| --- | --- | --- | --- |
| **Código** | PT5-A2 | **Nombre** | Realizar pruebas con usuarios finales. |
| **Descripción** | Evaluar la usabilidad y funcionalidad de la interfaz de usuario con la ayuda de profesionales médicos. | | |
| **Entradas** | Código fuente de la interfaz de usuario, sistema de diagnóstico implementado. | | |
| **Salidas** | Opiniones y sugerencias de los usuarios finales, ajustes y mejoras en la interfaz de usuario y en el sistema de diagnóstico. | | |
| **Tareas** | * Organizar sesiones de prueba con profesionales médicos * Recopilar opiniones y sugerencias de los usuarios finales * Implementar ajustes y mejoras en la interfaz de usuario y en el sistema de diagnóstico en función de las opiniones y sugerencias recibidas | | |
| **Duración** |  | | |

Tabla 14. Realizar pruebas con usuarios finales

## Plan de Trabajo

### Plan de Trabajo Inicial



### Plan de Trabajo Final

Aquí se pega el Gantt final cuando se ha terminado el proyecto y se comentan las diferencias con el inicial

## Recursos.

En el desarrollo y validación de este proyecto han colaborado diversos profesionales de la salud, incluyendo psicólogos y educadores, con el objetivo de asegurar la precisión y relevancia de los resultados obtenidos.

El ordenador utilizado para el desarrollo del proyecto ha sido un MacBook Pro modelo MacBookPro15,2, con un procesador Intel Core i5 de 4 núcleos a 2.3 GHz y 8 GB de memoria RAM. Este ordenador ha sido suficiente para llevar a cabo todas las tareas necesarias durante el desarrollo del proyecto.

En cuanto a las herramientas y tecnologías utilizadas, se ha empleado TensorFlow y Keras para el desarrollo y entrenamiento de la red neuronal profunda. Scikit-learn ha sido utilizado para la evaluación y validación del sistema de diagnóstico, mientras que NumPy y Pandas han sido fundamentales para el manejo y procesamiento de datos en el proyecto. OpenCV ha sido utilizado en caso de que se empleen imágenes médicas como fuentes de información.

Para el desarrollo de la interfaz de usuario se ha utilizado Streamlit, que ha permitido crear una interfaz intuitiva y atractiva con una cantidad mínima de código. Por último, Jupyter Notebook ha sido una herramienta útil para la experimentación, el análisis y la presentación de resultados en el campo del aprendizaje automático y la ciencia de datos.

## Costes

A continuación, se presenta una estimación de los costes asociados al desarrollo del proyecto. Se han considerado los costes de recursos humanos, hardware, software y otros gastos relacionados.

* Recursos humanos:
* Hardware: El proyecto se ha desarrollado en un MacBook Pro modelo MacBookPro15,2, con un procesador Intel Core i5 de 4 núcleos a 2.3 GHz y 8 GB de memoria RAM. El costo de este equipo es aproximadamente de ..
* Software: Las herramientas y tecnologías utilizadas en el proyecto incluyen TensorFlow, Keras, Scikit-learn, NumPy, Pandas, OpenCV y Streamlit. Todas estas herramientas son de código abierto y gratuitas, por lo que no se han generado costes adicionales en términos de licencias de software.
* Consultoría: Dado que se han colaborado diversos profesionales de la salud, incluyendo psicólogos y educadores, se estima un coste de $3,000 en consultoría y capacitación para asegurar la precisión y relevancia de los resultados obtenidos.
* Gastos generales: Se estima un costo adicional de $1,000 para gastos misceláneos, que incluyen material de oficina, conexión a Internet, electricidad y otros gastos relacionados con el desarrollo del proyecto.

En resumen, la estimación de costes del proyecto es la siguiente:

**Recursos humanos**:

**Hardware**:

**Software**: 0 (software de código abierto y gratuito)

**Consultoría**:

**Gastos generales**:

**Coste total estimado**:

Esta estimación proporciona una visión general de los costes asociados al proyecto. Sin embargo, cabe mencionar que los costes pueden variar en función de factores como la duración del proyecto, los recursos disponibles y las tarifas de consultoría específicas.

## Condicionantes y Limitaciones

# Desarrollo de la Solución Técnica

## PT1: Análisis e investigación previa.

### PT1-A1: Investigar trastornos del espectro autista (TEA)

La primera fase de investigación referente al proyecto fue profundizar en la comprensión de los Trastornos del Espectro Autista (TEA), explorando su naturaleza heterogénea, las dificultades inherentes al diagnóstico y las limitaciones actuales de las herramientas de diagnóstico. Me basé en diversas fuentes de información, incluyendo la literatura científica y médica, informes de organizaciones de salud y bases de datos de investigación.

A medida que avanzaba en la fase de investigación, se enfrentaron varios desafíos. En primer lugar, la diversidad y complejidad de los TEA y su diagnóstico se revelaron como un desafío considerable. Cada caso de TEA es único, y las superposiciones con otros trastornos pueden llevar a diagnósticos erróneos o retrasados, complicando aún más el proceso de diagnóstico.

Con una amplia gama de estudios y fuentes de información, fue necesario dedicar tiempo adicional para discernir las conclusiones más confiables y pertinentes, lo que llevó a una desviación del tiempo inicialmente previsto para esta fase. Sin embargo, considero que este tiempo extra fue valioso para obtener un entendimiento preciso de los TEA y los desafíos del diagnóstico.

A pesar de los desafíos, se han documentado todos los hallazgos en un informe detallado que proporciona una base sólida para las futuras fases del proyecto. Este informe detalla los desafíos inherentes al diagnóstico de los TEA. A continuación, presento el informe que recoge todos los hallazgos y análisis realizados durante esta fase de investigación:

Los Trastornos del Espectro Autista (TEA) son un grupo de trastornos del desarrollo neurológico que afectan la capacidad de una persona para comunicarse y relacionarse con los demás. Los TEA son "trastornos de espectro" porque afectan a los individuos de manera diferente y en diferentes grados (American Psychiatric Association, 2013).

**Características y Síntomas de los Trastornos del Espectro Autista**

Los Trastornos del Espectro Autista (TEA) se caracterizan por un conjunto de síntomas que se manifiestan de manera diversa entre individuos. A pesar de esta variabilidad, los síntomas de los TEA generalmente se agrupan en dos categorías principales: dificultades en la comunicación y la interacción social, y patrones de comportamiento, intereses o actividades restringidos y repetitivos.

2.1 Dificultades en la Comunicación e Interacción Social

Las personas con TEA pueden experimentar una amplia gama de desafíos en su habilidad para comunicarse e interactuar con los demás. Esto puede incluir:

* Dificultades con la comunicación verbal y no verbal: Esto puede manifestarse como retrasos en el desarrollo del lenguaje, uso inusual del lenguaje (por ejemplo, repetición literal de frases o palabras, uso idiosincrásico de palabras o frases), o dificultades con la comunicación no verbal, como hacer y mantener contacto visual, usar gestos para comunicarse o comprender el lenguaje corporal de los demás (American Psychiatric Association, 2013).
* Dificultades en las interacciones sociales: Las personas con TEA pueden tener dificultades para desarrollar y mantener relaciones sociales. Esto puede incluir desafíos en la comprensión y la respuesta a las emociones y las intenciones de los demás, la participación en el juego simbólico, o la adaptación del comportamiento en diferentes contextos sociales (American Psychiatric Association, 2013).

2.2 Patrones de Comportamiento, Intereses o Actividades Restringidos y Repetitivos

Los individuos con TEA también pueden exhibir patrones de comportamiento, intereses o actividades que son restringidos y repetitivos. Esto puede incluir:

* Comportamientos estereotipados o repetitivos: Estos pueden incluir movimientos del cuerpo repetitivos (por ejemplo, aleteo de las manos, balanceo del cuerpo), uso de objetos de manera repetitiva o inusual (por ejemplo, alinear objetos, girar objetos), o adherencia a rutinas o patrones de comportamiento rígidos (American Psychiatric Association, 2013).
* Intereses intensos y restringidos: Las personas con TEA pueden mostrar un interés extremo en un tema o actividad específica a un grado que es inusual o excesivo. Estos intereses pueden ser muy específicos y a menudo son de larga duración (American Psychiatric Association, 2013).

Es importante recordar que los síntomas de los TEA varían considerablemente entre individuos, tanto en términos de los tipos de síntomas que se presentan como en la gravedad de estos síntomas.

**Desafíos en el Diagnóstico de los Trastornos del Espectro Autista**

El diagnóstico de los Trastornos del Espectro Autista (TEA) puede ser un proceso complejo y desafiante debido a una serie de factores, que incluyen la variabilidad de los síntomas, la coexistencia de otros trastornos o las diferencias en la presentación entre grupos de edad y género.

3.1 Variabilidad de los síntomas

Los TEA se caracterizan por una notable variabilidad en los síntomas. Dos personas con un diagnóstico de TEA pueden presentar síntomas muy diferentes y tener grados de gravedad muy diferentes. Esta diversidad puede complicar el proceso de diagnóstico, ya que los profesionales de la salud deben confiar en una evaluación detallada y completa de los comportamientos y habilidades del individuo para hacer un diagnóstico preciso (Zwaigenbaum et al., 2015). Además, los síntomas del TEA pueden cambiar a lo largo del tiempo, lo que puede requerir reevaluaciones y ajustes en el tratamiento (Mayo Clinic, 2018).

3.2 Coexistencia de otros trastornos

Muchas personas con TEA también tienen otras condiciones médicas o psiquiátricas, como trastornos del sueño, trastornos gastrointestinales, trastorno por déficit de atención e hiperactividad (TDAH), trastorno de ansiedad, y trastorno depresivo. Estas condiciones coexistentes pueden complicar el diagnóstico de TEA, ya que algunos de sus síntomas pueden solaparse con los de TEA (Zwaigenbaum et al., 2015). En algunos casos, estos trastornos pueden pasar desapercibidos, ya que los síntomas del TEA pueden ocultar estos trastornos subyacentes (Simonoff et al., 2008).

3.3 Diferencias en la presentación entre grupos de edad y género

Las diferencias en la presentación de los síntomas de TEA entre diferentes grupos de edad y género pueden complicar aún más el diagnóstico. Por ejemplo, las niñas con TEA a menudo se diagnostican más tarde que los niños, ya que pueden ser más capaces de camuflar sus dificultades o pueden mostrar síntomas menos evidentes (Loomes, Hull, & Mandy, 2017). Además, los síntomas de TEA pueden cambiar a medida que las personas crecen y se desarrollan, lo que significa que el diagnóstico puede necesitar ser revisado y ajustado con el tiempo (Zwaigenbaum et al., 2015).

3.4 Retraso en el diagnóstico

A pesar de que los síntomas de TEA suelen comenzar en la primera infancia, a menudo hay un retraso significativo entre la aparición de los síntomas y el diagnóstico formal. Este retraso puede tener un impacto significativo en el acceso al tratamiento y el pronóstico del individuo (Zuckerman, Lindly, & Sinche, 2015). Los esfuerzos para mejorar la detección temprana de los TEA son fundamentales para asegurar que los individuos con TEA reciban el apoyo y el tratamiento que necesitan.

3.5 Dificultades en la evaluación intercultural

Los desafíos culturales y lingüísticos también pueden complicar el diagnóstico de TEA. Los síntomas de TEA pueden interpretarse de manera diferente en diferentes culturas, y los instrumentos de diagnóstico pueden no ser completamente precisos o adecuados para todas las poblaciones culturales y lingüísticas (Norbury & Sparks, 2013).

**Conclusion**

Es evidente que existe una necesidad urgente de mejorar las estrategias de diagnóstico para el TEA. Esto incluye el desarrollo de herramientas y técnicas más objetivas y precisas para el diagnóstico, como biomarcadores fiables, así como la mejora de la formación y la educación de los profesionales de la salud para asegurar una detección y un diagnóstico más tempranos. Además, es crucial tener en cuenta las diferencias culturales y lingüísticas en el diagnóstico del TEA, para asegurar que todos los individuos con TEA tengan acceso a un diagnóstico preciso y a un tratamiento adecuado.

Además, los esfuerzos para mejorar la comprensión de los TEA y su diagnóstico deben ser inclusivos y considerar la diversidad de experiencias y síntomas asociados con estos trastornos. Esto incluye la consideración de las diferencias en la presentación del TEA en diferentes grupos de edad y género, así como en diferentes culturas y contextos lingüísticos.

En conclusión, el diagnóstico del TEA es un proceso complejo y desafiante, que requiere un enfoque individualizado y flexible. A pesar de los desafíos, el diagnóstico temprano y preciso del TEA es crucial para garantizar que los individuos con TEA reciban el apoyo y el tratamiento que necesitan para alcanzar su máximo potencial. Con la investigación y los esfuerzos continuos para mejorar el diagnóstico y la comprensión del TEA, es posible avanzar hacia este objetivo.

### PT1-A2: Analizar fuentes de información médica

Para la identificación de los estudios más relevantes, se seleccionaron las siguientes bases de datos y plataformas de investigación debido a su amplia cobertura de literatura médica y científica:

* **PubMed**. Una base de datos de libre acceso mantenida por la Biblioteca Nacional de Medicina de los EE. UU. que contiene más de 30 millones de citas de literatura biomédica.
* **Web of Science**. Un servicio de información de investigación de alto nivel que cubre diversas disciplinas, incluyendo ciencias, ciencias sociales, artes y humanidades.
* **Embase**. Una base de datos de investigación biomédica que es especialmente útil para investigadores que buscan información sobre medicamentos y dispositivos médicos.
* **ScienceDirect**. Una plataforma de texto completo que ofrece acceso a revistas de ciencias, tecnología y medicina.
* **Nature**. Una plataforma de revistas científicas que cubre una amplia gama de disciplinas.

Estas bases de datos y plataformas proporcionaron un marco sólido para la identificación y selección de los estudios más relevantes para el proyecto. A continuación, se muestra la selección de los estudios identificados como potencialmente relevantes para el proyecto:

* Aldridge, K., et al. (2011). Los investigadores de este estudio llevaron a cabo un análisis pormenorizado de los rasgos faciales de los niños con TEA, utilizando técnicas de morfometría geométrica para identificar correlaciones entre los fenotipos faciales y los fenotipos clínicos. Encontraron que ciertas características faciales, como la distancia interpupilar y la forma de la boca y la mandíbula, estaban asociadas con ciertos subtipos de TEA, lo que sugiere que estos rasgos podrían usarse como biomarcadores faciales para el diagnóstico y la estratificación de los TEA.
* Obafemi-Ajayi, T., et al. (2015). En este estudio, los investigadores utilizaron un enfoque de aprendizaje automático para analizar las características faciales de las personas con TEA. Se utilizaron técnicas de procesamiento de imágenes para extraer características faciales, y luego se entrenó un modelo de aprendizaje automático para clasificar a las personas en subgrupos clínicos de TEA basándose en estas características. Los resultados indican que la estructura facial puede ser un biomarcador útil para distinguir entre diferentes subgrupos de TEA.
* Walsh, P., et al. (2011). Este artículo de revisión proporciona un análisis exhaustivo de los desafíos asociados con la búsqueda de biomarcadores para el autismo. Los autores discuten los desafíos técnicos, como la necesidad de grandes conjuntos de datos y técnicas de análisis sofisticadas, así como los desafíos éticos y sociales, como las implicaciones de etiquetar a las personas con un diagnóstico basado en biomarcadores. A pesar de estos desafíos, los autores argumentan que la búsqueda de biomarcadores para el autismo es crucial para mejorar el diagnóstico y el tratamiento de este trastorno.
* Ozonoff, S., et al. (2010). Este estudio longitudinal siguió a una cohorte de niños desde el nacimiento hasta los 3 años de edad, observando cuidadosamente el desarrollo de los signos conductuales de autismo. Los investigadores encontraron que ciertos comportamientos, como la disminución del contacto visual y la falta de respuesta al nombre, podían predecir un diagnóstico posterior de autismo. Este estudio subraya la importancia de los biomarcadores conductuales en la detección temprana del autismo.
* Shen, M. D., et al. (2013). Este estudio investigó si los cambios cerebrales tempranos, medidos por resonancia magnética, podrían predecir un diagnóstico posterior de TEA. Los investigadores encontraron que los niños que luego fueron diagnosticados con TEA tenían un mayor volumen de fluido extra-axial (un tipo de líquido que rodea el cerebro) a los 6 meses de edad. Este estudio pone de relieve la importancia de los biomarcadores cerebrales en la comprensión del autismo.
* Anzulewicz, A., et al. (2016). Este estudio examinó si los patrones de movimiento durante el juego con tabletas inteligentes podrían usarse como un biomarcador para el autismo. Los investigadores encontraron que los niños con autismo mostraban patrones de movimiento y gestos únicos cuando interactuaban con las tabletas. Estos patrones eran consistentes, lo que sugiere que podrían usarse como un biomarcador motor fiable para el autismo. Los niños con autismo tendían a tener una mayor variabilidad en sus gestos y a usar movimientos más grandes y más lentos en comparación con los niños sin autismo.
* Wang, Y., et al. (2022). Este estudio, realizado por investigadores de la Universidad de California, San Francisco, analizó la microbiota intestinal de 100 niños con autismo y 100 niños sin autismo. Descubrieron que los niños con autismo tenían una composición de microbiota intestinal diferente a la de los niños sin autismo, específicamente, los niños con autismo tenían niveles más bajos de bacterias beneficiosas y niveles más altos de bacterias dañinas. También encontraron que ciertos metabolitos, como los ácidos grasos de cadena corta, estaban asociados con el autismo.
* Zhang, Y., et al. (2022). Investigadores de la Universidad de California, Los Ángeles, analizaron la saliva de 50 niños con autismo y 50 niños sin autismo. Encontraron que ciertos biomarcadores salivales, como los neuropéptidos y el cortisol, estaban asociados con el autismo. Los neuropéptidos participan en la comunicación entre las células nerviosas y el cortisol es una hormona del estrés. Los investigadores sugieren que estos biomarcadores podrían usarse para desarrollar una prueba diagnóstica para el autismo.

Estos estudios abarcan una gama de enfoques en la búsqueda de biomarcadores para el autismo, incluyendo análisis de características faciales, estudios de comportamiento, análisis de estructuras cerebrales, observación de patrones de movimiento y biomarcadores corporales. Sin embargo, los estudios de Aldridge et al. (2011) y Obafemi-Ajayi et al. (2015) demuestran el potencial de los biomarcadores faciales en particular. Estos estudios indican que las características faciales pueden correlacionarse con los subtipos de TEA y que las técnicas de análisis de la estructura facial pueden ser eficaces para distinguir entre estos subtipos. Como resultado, los biomarcadores faciales pueden ser particularmente relevantes y útiles para nuestro proyecto y podrían ser una vía principal de investigación.

### PT1-A3: Investigar técnicas de aprendizaje automático y redes neuronales

TEORIA IA 2

## PT2

### PT2-A1: Definir objetivos específicos y generales

### PT2-A3: Establecer métricas de evaluación

Para definir las métricas que el sistema de diagnóstico almacenará en el contexto del Trastorno del Espectro Autista, comenzamos investigando dos herramientas de diagnóstico bien establecidas: el M-CHAT (Modified Checklist for Autism in Toddlers) y el ADOS (Autism Diagnostic Observation Schedule). Estos instrumentos son ampliamente reconocidos en la comunidad médica y ofrecen una base sólida para nuestro sistema.

**Modified Checklist for Autism in Toddlers**

El M-CHAT es una herramienta de detección de Trastorno del Espectro Autista (TEA) para niños entre 16 y 30 meses. Desarrollado para facilitar la identificación temprana de niños en riesgo de autismo, se basa en preguntas orientadas hacia comportamientos típicos y atípicos en áreas clave del desarrollo.

Este cuestionario consta de 23 preguntas de sí/no diseñadas para ser completadas por los padres o cuidadores del niño. Estas preguntas cubren una amplia gama de comportamientos, incluyendo aquellos relacionados con la interacción social, la comunicación verbal y no verbal (como el uso de gestos y la comprensión de lenguaje) y la flexibilidad de comportamiento (como la respuesta a cambios en la rutina o el entorno).

Las respuestas a estas preguntas se clasifican en dos categorías: "indicativas de riesgo" y "no indicativas de riesgo". Las respuestas "indicativas de riesgo" son aquellas que, según la evidencia existente, podrían señalar un mayor riesgo de TEA. No todas las preguntas tienen el mismo peso; algunas son consideradas "críticas" debido a su fuerte relación con el diagnóstico de TEA.

Después de realizar el cuestionario, se calcula el número total de respuestas "indicativas de riesgo". Si hay un cierto número de respuestas "indicativas de riesgo", se realiza un seguimiento, que implica hacer preguntas adicionales para obtener más detalles sobre los comportamientos de interés. Basándose en el total de respuestas "indicativas de riesgo" y en la información del seguimiento, se genera un resultado final que indica si el niño tiene un riesgo bajo, medio o alto de TEA.

Este resultado es crítico para determinar los próximos pasos. Los niños con riesgo medio o alto son referidos para una evaluación diagnóstica completa, mientras que los niños con riesgo bajo son monitorizados para asegurar que no se presenten comportamientos de interés en el futuro.

Se han decidido almacenar las siguientes métricas de salida para el M-CHAT:

* Número total de respuestas "indicativas de riesgo": Este es un indicador cuantitativo del riesgo percibido de TEA según el M-CHAT.
* Número de respuestas "indicativas de riesgo" en preguntas "críticas": Este número proporciona una visión más detallada del riesgo, ya que estas preguntas están fuertemente asociadas con el TEA.
* Resultado final del M-CHAT: Esta es una interpretación cualitativa del riesgo de TEA, basada en las respuestas "indicativas de riesgo" y el seguimiento.
* Observaciones o comentarios adicionales proporcionados por los padres o profesionales: Estos pueden proporcionar contexto y detalles adicionales que pueden ser útiles para la interpretación del resultado del M-CHAT y para la planificación de los próximos pasos.
* Fecha de realización del cuestionario: Esto nos permitirá rastrear cualquier cambio en el tiempo y responder a tiempo si el riesgo percibido de TEA aumenta.

**Autism Diagnostic Observation Schedule**

El ADOS es una evaluación estandarizada y semi-estructurada de la comunicación, la interacción social, el juego y los comportamientos restrictivos y repetitivos. Esta herramienta de diagnóstico se usa con individuos sospechosos de tener TEA y se puede aplicar a niños, adolescentes y adultos.

El ADOS consta de cinco módulos, cada uno de los cuales se dirige a diferentes grupos de edad y habilidades. Cada módulo tiene una serie de actividades planificadas que permiten al evaluador observar la ocurrencia de comportamientos que son importantes en el diagnóstico de los trastornos del espectro autista.

Los módulos son los siguientes:

* Módulo 1: para niños que no hablan consistentemente con frases de tres palabras, y que son menores de 31 meses.
* Módulo 2: para niños de cualquier edad que usan frases pero no hablan con fluidez.
* Módulo 3: para niños y jóvenes verbalmente fluidos.
* Módulo 4: para adolescentes y adultos verbalmente fluidos.
* Módulo Toddler: para niños de 12 a 30 meses, independientemente de su nivel de lenguaje.

En la evaluación, se observan y codifican las conductas en dos áreas principales: Comunicación Social Recíproca (CSR) y Comportamientos Restringidos, Repetitivos y Estereotipados (CRRE). Los puntajes de estas áreas se suman para formar un puntaje total, el cual es fundamental para determinar si un individuo cumple con los criterios para el TEA según el ADOS.

La Comunicación Social Recíproca incluye aspectos como el uso del contacto visual, la gestualidad, la calidad del juego social, la respuesta a las emociones de los demás, etc. Los Comportamientos Restringidos, Repetitivos y Estereotipados, por otro lado, pueden involucrar movimientos estereotipados y repetitivos, resistencia al cambio, intereses limitados, etc.

Para garantizar una evaluación precisa, se decidieron almacenar las siguientes métricas de salida para el ADOS:

* Módulo utilizado en la evaluación: Este dato nos proporcionará un contexto importante sobre el nivel de habilidad y desarrollo del lenguaje del individuo.
* Puntuaciones totales en Comunicación Social Recíproca y Comportamientos Restringidos, Repetitivos y Estereotipados: Estas son medidas cuantitativas de los síntomas de TEA observados durante la evaluación.
* Puntuación total del ADOS: Este es un indicador global de la presencia y gravedad de los síntomas del TEA.
* Conclusión del evaluador: Esta es la interpretación clínica de los resultados por parte del evaluador. Puede incluir recomendaciones para próximas evaluaciones o intervenciones.
* Fecha de la evaluación: Esto permitirá rastrear cualquier cambio en el tiempo y asegurará que la información esté actualizada.

## PT3

### PT3-A1: Definir el flujo de datos

**ARQUITECTURA DE LA SOLUCIÓN**

Para el desarrollo de la plataforma de diagnóstico de TEA, se ha establecido una arquitectura que integra interfaces de usuario, técnicas de inteligencia artificial mediante redes profundas y bases de datos para proporcionar una herramienta de diagnóstico efectiva para profesionales de la salud y la educación.

La arquitectura de la solución se desglosa en los siguientes componentes principales: la interfaz para profesionales, el sistema de diagnóstico (con una red neuronal profunda integrada), y la base de datos. A continuación, se detalla cómo cada componente contribuye al funcionamiento general de la plataforma, y cómo se integran para crear un sistema de diagnóstico completo y eficaz.

* **Interfaz para profesionales.** Esta es la interfaz de usuario de la aplicación, que es lo que los profesionales ven cuando interactúan con la plataforma. Proporciona todas las funcionalidades necesarias para realizar diagnósticos, introducir información del paciente y revisar resultados de las pruebas.
* **API (Sistema de Diagnóstico).** Esta es la capa que sirve como intermediario entre la interfaz del usuario y los diversos sistemas back-end. Está compuesta por varias sub-APIs, incluyendo la API de la red neuronal, la API de Streamlit y la API para la conexión con la base de datos.
* API de la Red Neuronal Profunda: Esta API es responsable de comunicarse con el modelo de red neuronal para el diagnóstico de TEA. Cuando se solicita un diagnóstico, la imagen del rostro del niño se envía a través de esta API al modelo de red neuronal, que luego devuelve el resultado del diagnóstico.
* API de Streamlit: La API de Streamlit se utilizará para construir la interfaz del usuario y proporcionar la funcionalidad para interactuar con la información almacenada en la base de datos.
* API de conexión con la Base de Datos: Esta API es responsable de todas las interacciones con la base de datos. Esto incluye recuperar y almacenar información de usuario, resultados de pruebas, diagnósticos y detalles del paciente.

**MODELO DE DATOS**

El modelo de datos consta de seis tablas principales: Children, Diagnoses, Professionals, ADOS, MCHAT, y FaceDiagnosis. A continuación, se detallan las tablas y sus correspondientes campos con sus descripciones y tipos de datos:

**Tabla Children**

Esta tabla almacena la información demográfica de cada niño evaluado. Los campos incluyen el nombre, los apellidos, la fecha de nacimiento, el género y cualquier diagnóstico previo de TEA. También se registra si existen antecedentes familiares de trastornos del neurodesarrollo.

* child\_id (Integer, Primary Key): ID del niño.
* professional\_dni (String, Foreign Key): DNI del profesional asociado.
* first\_name (String): Nombre del niño.
* last\_name (String): Apellidos del niño.
* birth\_date (Date): Fecha de nacimiento del niño.
* gender (String): Género del niño.
* family\_history (Boolean): Antecedentes familiares de trastornos del neurodesarrollo.
* previous\_diagnosis (Boolean): Diagnóstico previo.
* observations (String): Observaciones adicionales.

La relación entre la tabla Children y Professionals es de N:1, ya que un profesional puede tener varios niños a su cargo, pero un niño solo puede tener un profesional asignado.

**Tabla Diagnoses**

En la tabla Diagnoses se almacenan los diagnósticos generados para cada niño. Esto incluye una referencia a la imagen de la cara utilizada para la evaluación y los resultados obtenidos de los tests M-CHAT y ADOS, junto con el diagnóstico generado por la red neuronal:

* diagnosis\_id (Integer, Primary Key): ID del diagnóstico.
* child\_id (Integer, Foreign Key): ID del niño asociado.
* diagnosis\_date (Date): Fecha del diagnóstico.
* photo (String): Ruta al archivo de la foto utilizada para el diagnóstico.
* ados\_id (Integer, Foreign Key): ID de la prueba ADOS asociada.
* mchat\_id (Integer, Foreign Key): ID de la prueba M-CHAT asociada.
* face\_diagnosis\_id (Integer, Foreign Key): ID del diagnóstico facial asociado.

La relación entre la tabla Diagnoses y Children es de 1:N, ya que un niño puede tener varios diagnósticos a lo largo del tiempo.

**Tabla Professionals**

La tabla Professionals almacena información sobre los profesionales que utilizan la plataforma. Esto incluye su nombre, apellidos y disciplina profesional. También se almacena el hash de la contraseña para verificar la identidad del profesional al iniciar sesión:

* dni (String, Primary Key): DNI del profesional.
* first\_name (String): Nombre del profesional.
* last\_name (String): Apellidos del profesional.
* discipline (String): Disciplina profesional.
* hashed\_password (String): Contraseña hasheada para el inicio de sesión.

**Tabla ADOS**

La tabla ADOS recoge los resultados de las pruebas ADOS realizadas. Esto incluye el módulo utilizado en la evaluación, las puntuaciones obtenidas en las diferentes categorías, la puntuación total y la conclusión del evaluador:

* ados\_id (Integer, Primary Key): ID de la prueba ADOS.
* module (String): Módulo utilizado en la evaluación.
* reciprocal\_social\_interaction (Integer): Puntuación en Comunicación Social Recíproca.
* restricted\_repetitive\_behaviour (Integer): Puntuación en Comportamientos Restringidos, Repetitivos y Estereotipados.
* total\_score (Integer): Puntuación total.
* conclusion (String): Conclusión del evaluador.
* evaluation\_date (Date): Fecha de la evaluación.

La relación entre la tabla ADOS y Diagnoses es de 1:1, ya que cada diagnóstico se asocia con una única prueba ADOS.

**Tabla MCHAT**

En la tabla MCHAT se almacenan los resultados de las pruebas M-CHAT. Se registra el número total de respuestas que indican un posible riesgo, el número de respuestas de riesgo en preguntas críticas, el resultado final del M-CHAT y cualquier observación adicional:

* mchat\_id (Integer, Primary Key): ID de la prueba M-CHAT.
* total\_risk\_responses (Integer): Número total de respuestas que indican un posible riesgo.
* critical\_risk\_responses (Integer): Número de respuestas que indican un riesgo en preguntas críticas.
* final\_result (String): Resultado final del M-CHAT.
* observations (String): Observaciones.
* date (Date): Fecha de la prueba.

La relación entre la tabla MCHAT y Diagnoses es de 1:1, ya que cada diagnóstico se asocia con una única prueba M-CHAT.

**Tabla FaceDiagnosis**

La tabla FaceDiagnosis almacena los resultados de los diagnósticos generados por la red neuronal basados en las imágenes de la cara del niño. Esto incluye la ruta al archivo de la imagen utilizada y el resultado de la evaluación junto con la probabilidad asociada:

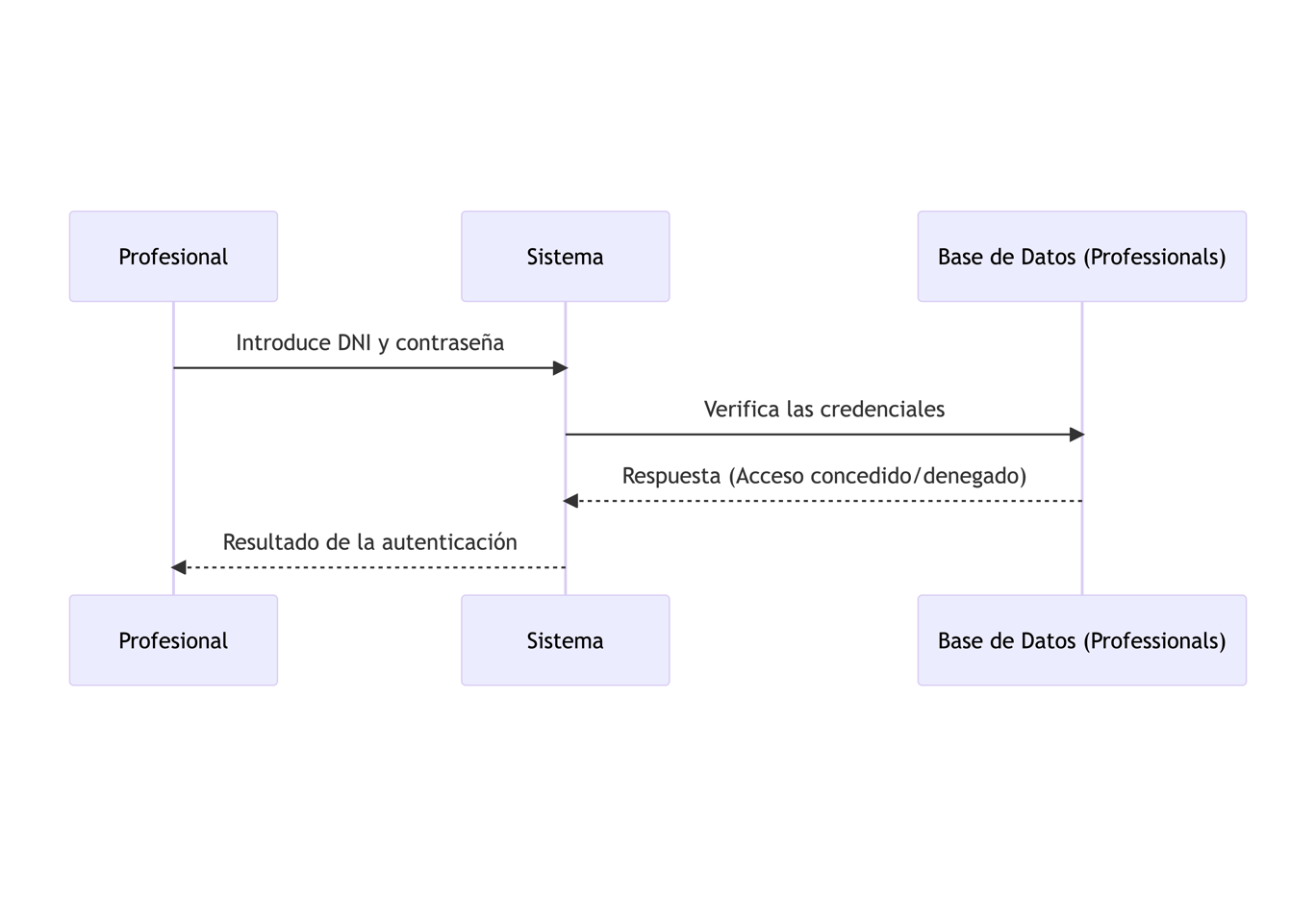
* face\_diagnosis\_id (Integer, Primary Key): ID del diagnóstico facial.
* photo (String): Ruta al archivo de la foto utilizada para el diagnóstico.
* result (String): Resultado del diagnóstico (TEA/No TEA).
* percentage (Decimal): Porcentaje de probabilidad de diagnóstico de TEA.

La relación entre la tabla FaceDiagnosis y Diagnoses es de uno a uno, ya que cada diagnóstico se asocia con un único diagnóstico facial.

**FLUJOS DEL SISTEMA DE DIAGNOSTICO**

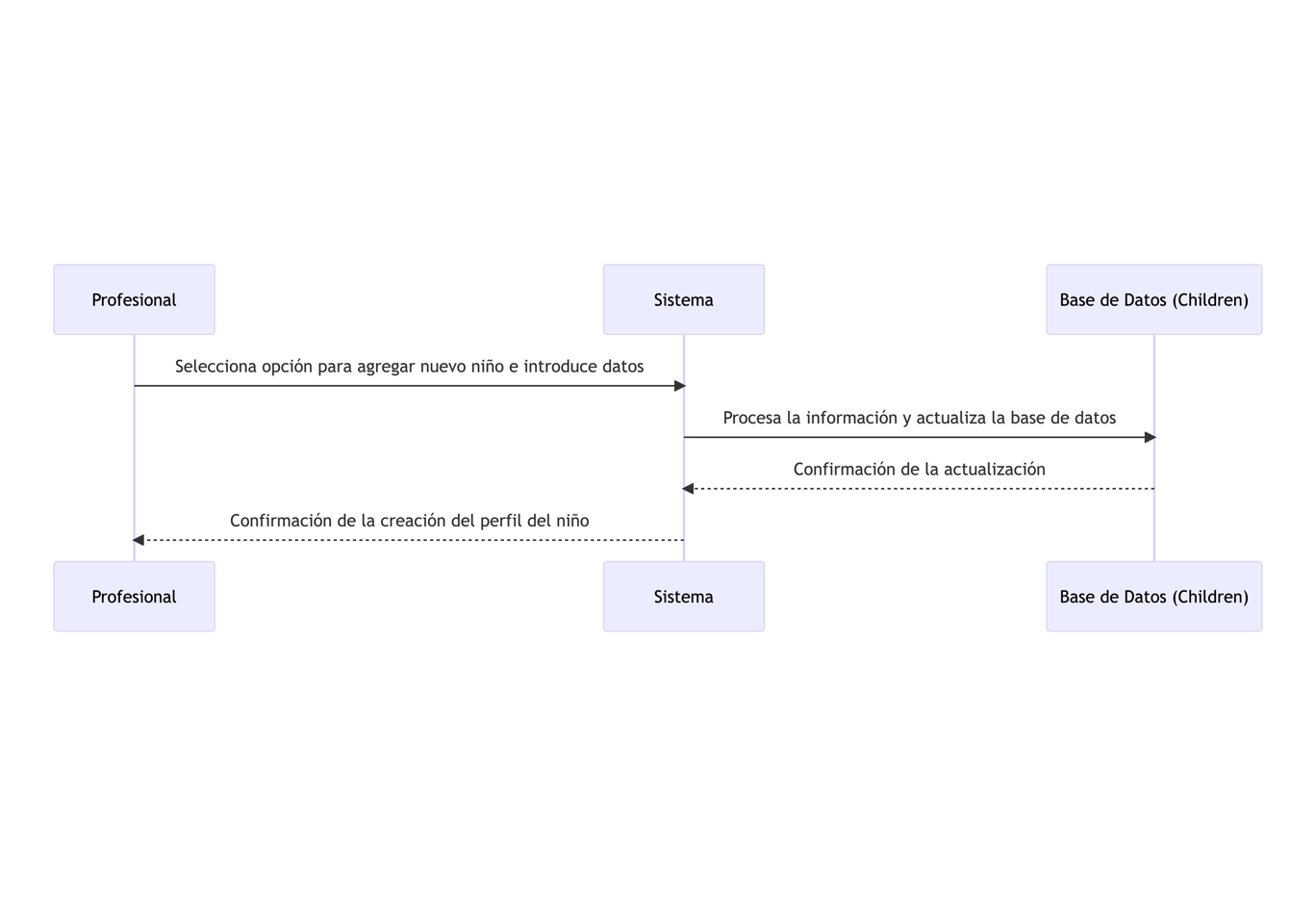
**Autenticación de Usuario**

En este flujo, el profesional abre la plataforma e introduce su DNI y contraseña. Esta información se envía a través del sistema que, tras verificar las credenciales en la tabla "Professionals" de la base de datos, concede o deniega el acceso al profesional.



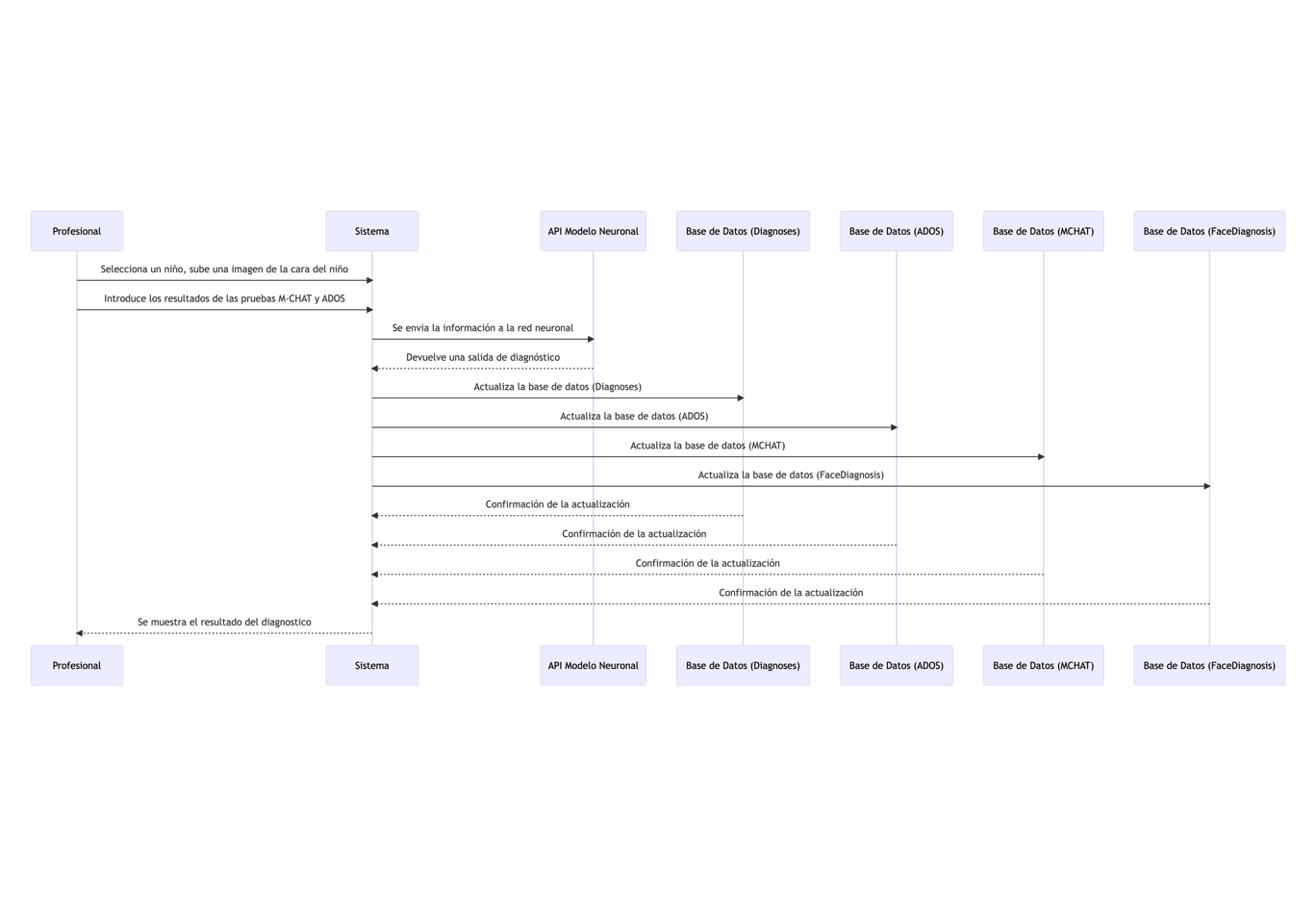
**Gestión de Perfiles de Niños**

Una vez autenticado, el profesional tiene la capacidad de gestionar los perfiles de los niños. Para ello, puede seleccionar la opción para agregar un nuevo niño, introduciendo los datos requeridos en un formulario. Estos datos son recibidos por el sistema, que los procesa y actualiza la tabla "Children" en la base de datos correspondiente.



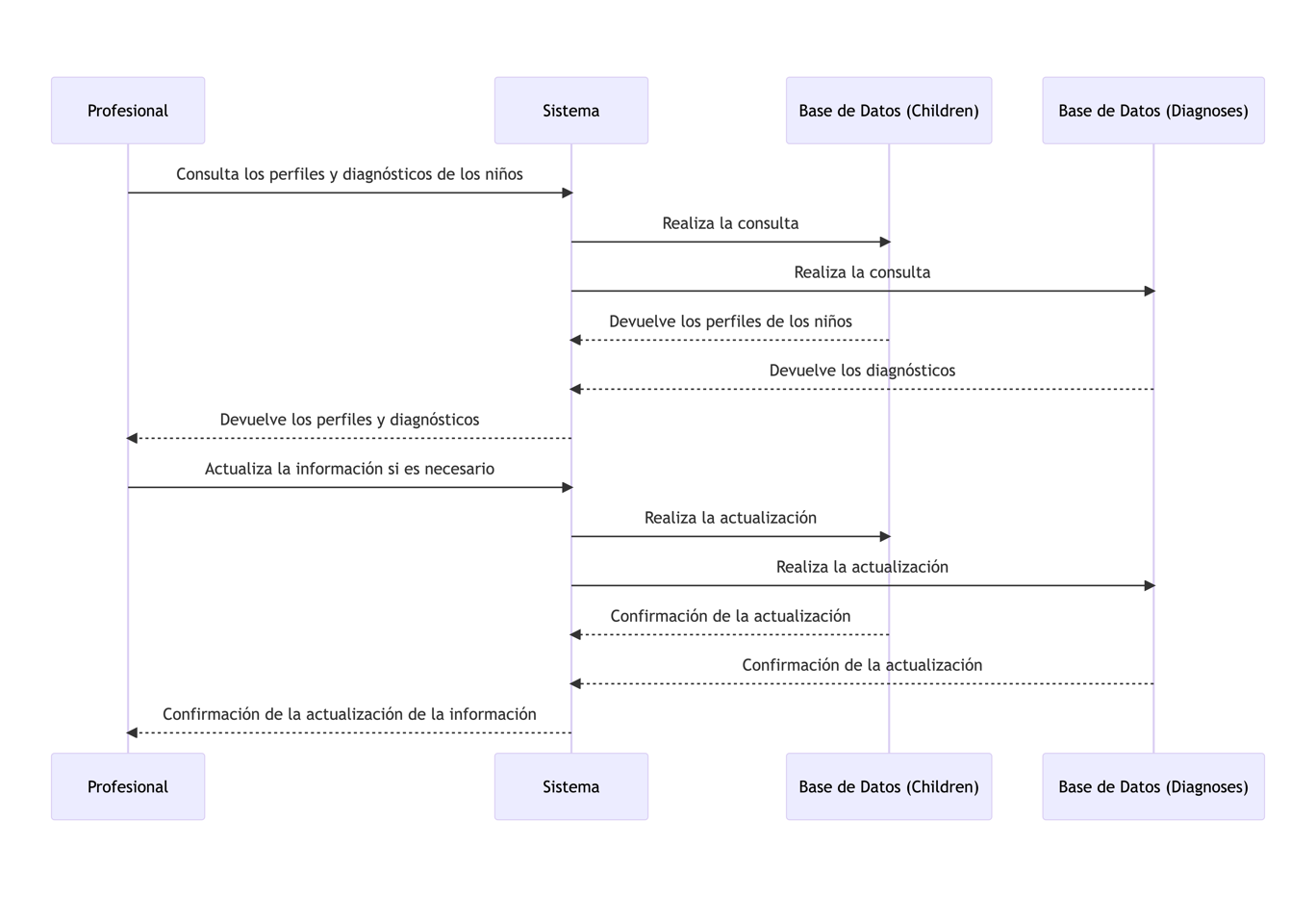
**Diagnóstico de TEA e Ingreso de Resultados de Pruebas M-CHAT y ADOS**

En este flujo, el profesional selecciona a un niño para el diagnóstico. Sube una imagen de la cara del niño a la plataforma y proporciona los resultados de las pruebas M-CHAT y ADOS. La imagen es procesada por la Red Neuronal Profunda, que devuelven una salida de diagnóstico basada en la imagen. Simultáneamente, los resultados de las pruebas M-CHAT y ADOS son procesados por el sistema y almacenados en sus respectivas tablas en la base de datos.



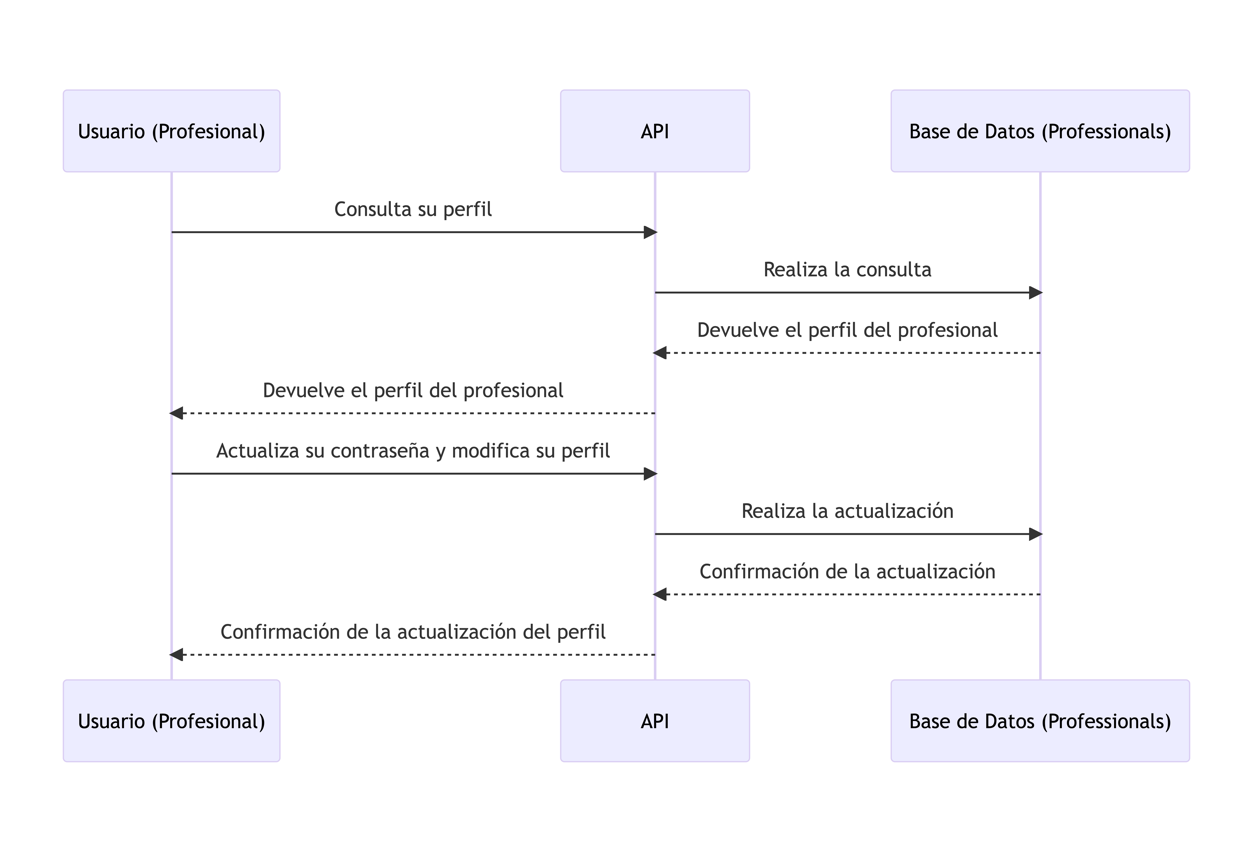
**Consulta y Actualización de Perfiles y Diagnósticos**

El profesional tiene la capacidad de consultar y actualizar la información de los perfiles y diagnósticos de los niños. Puede hacer esto a través de la interfaz de la plataforma, que interactúa con la API para obtener y modificar la información almacenada en las tablas "Children" y "Diagnoses" de la base de datos.



**Gestión de Perfiles de Profesionales**

El profesional puede gestionar su propio perfil. Esto incluye la capacidad de actualizar su contraseña y modificar su información personal. Este flujo de trabajo implica interactuar con la API, que procesa los cambios y actualiza la información en la tabla "Professionals" de la base de datos.



ESBOZO DE LOS REQUISITOS DE LA PLATAFORMA

R1. La plataforma debe soportar cinco tipos de roles profesionales: psicólogos, docentes, terapeutas ocupacionales, logopedas y pediatras.

R2. Cada profesional deberá tener un identificador único que será su DNI.

R3. Cada profesional deberá tener una contraseña única para acceder a la plataforma.

R4. Los profesionales deben poder iniciar sesión en la plataforma utilizando su DNI y contraseña.

R5. El sistema debe verificar las credenciales del profesional durante el inicio de sesión.

R6. Una vez autenticado, el profesional debe tener la opción de agregar nuevos niños.

R7. Al agregar un nuevo niño, se debe proporcionar un formulario para ingresar los datos del niño.

R8. El sistema debe permitir la entrada del nombre del niño.

R9. El sistema debe permitir la entrada de los apellidos del niño.

R10. El sistema debe permitir la entrada de la fecha de nacimiento del niño.

R11. El sistema debe permitir la entrada del género del niño.

R12. El sistema debe permitir la entrada de los antecedentes familiares de trastornos del neurodesarrollo.

R13. El sistema debe permitir la entrada de un diagnóstico previo, si existe.

R14. El sistema debe permitir la entrada de observaciones adicionales.

R15. El sistema debe permitir guardar la información del niño.

R16. El sistema debe asociar la información del niño con el profesional correspondiente.

R17. El sistema debe permitir al profesional ver una lista de los niños que ha registrado.

R18. El sistema debe permitir al profesional seleccionar un niño de la lista para ver más detalles o editar la información.

R19. El sistema debe permitir al profesional realizar un diagnóstico de TEA utilizando la imagen de la cara del niño.

R20. El sistema debe permitir al profesional almacenar los resultados de las pruebas M-CHAT y ADOS.

R21. El diagnóstico de TEA debe ser generado por una red neuronal profunda.

R22. El diagnóstico de TEA debe ser almacenado en la base de datos asociada al niño correspondiente.

R23. El sistema debe permitir al profesional consultar los diagnósticos realizados para cada niño.

R24. El sistema debe permitir al profesional actualizar la información del diagnóstico si es necesario.

R25. El sistema debe permitir al profesional actualizar su contraseña.

R26. El sistema debe permitir al profesional modificar su perfil.

R27. El sistema debe incluir medidas de seguridad para proteger la información almacenada.

R28. El sistema debe garantizar la confidencialidad de los datos de los niños y los profesionales.

## PT4

### PT4-A1: Preparar los datos

El preprocesamiento de los datos se realizó en varias etapas:

Primero, se definieron las rutas a las carpetas que contenían las imágenes de entrenamiento, prueba y validación. Estas carpetas estaban organizadas de tal manera que las imágenes de niños autistas y no autistas estaban separadas en diferentes subcarpetas.

Luego, se cargó cada conjunto de imágenes y sus correspondientes etiquetas utilizando la función **load\_images\_and\_labels**. Esta función recorre todos los archivos de imagen en la carpeta proporcionada y genera una lista de rutas a las imágenes. A continuación, asigna una etiqueta a cada imagen en función de si el nombre del archivo contiene la cadena "Non\_Autistic" o no. Las imágenes de niños no autistas se etiquetan como 0, mientras que las imágenes de niños autistas se etiquetan como 1.

Después de cargar las imágenes y las etiquetas, se preprocesaron las imágenes utilizando la función **preprocess\_image**. Esta función carga cada imagen, la redimensiona a un tamaño de 200x200 píxeles y la convierte en una matriz de numpy. Luego, normaliza los valores de los píxeles de la imagen dividiéndolos por 255.0. La normalización de los valores de los píxeles es un paso común en el preprocesamiento de imágenes para el aprendizaje profundo, ya que ayuda a mejorar la velocidad de convergencia y la estabilidad general del entrenamiento. Además, la normalización asegura que los valores de los píxeles estén en el mismo rango, lo que es importante para que los algoritmos de aprendizaje automático funcionen correctamente.

Finalmente, la función **to\_categorical** de Keras se utilizó para convertir las etiquetas de clase, que son enteros, en una forma categórica. Esto se conoce como codificación one-hot. En la codificación one-hot, cada etiqueta se convierte en un vector que contiene principalmente ceros y un solo uno. La posición del uno en el vector indica la clase de la etiqueta.

La conversión de las etiquetas a un formato categórico es un paso esencial en el preprocesamiento de los datos para la clasificación multiclase en el aprendizaje profundo. En este caso, aunque estábamos ante un problema de clasificación binaria (autista o no autista), el uso de la codificación categórica seguía siendo relevante y beneficioso.

Por ejemplo, en un problema de clasificación binaria como este, las etiquetas 0 y 1 se convertirían en los vectores [1, 0] y [0, 1] respectivamente. En un problema de clasificación con más de dos clases, el tamaño del vector sería igual al número de clases y el uno se colocaría en la posición correspondiente a la clase de la etiqueta.

### PT4-A2: Desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales

Tras el preprocesamiento, se siguieron los siguientes pasos para desarrollar e implementar un algoritmo de búsqueda automatizada para explorar y evaluar múltiples arquitecturas de redes neuronales:

* **Configuración de una arquitectura de prueba:** El primer paso fue configurar una arquitectura de prueba para probar su entrenamiento y verificar que las salidas tuvieran sentido. Se crearon las clases necesarias para el entrenamiento y para mostrar los resultados.
* **Data augmentation:** El siguiente paso fue la implementación de la técnica de data augmentation. Se editaron las funciones necesarias para realizar transformaciones en las imágenes de entrenamiento, como rotaciones, traslaciones, zoom y volteo horizontal. Esta técnica permite aumentar la cantidad de datos de entrenamiento y mejorar la capacidad de la red para generalizar a nuevas imágenes.
* **Ajuste del learning rate y prevención del sobreajuste**: Se desarrolló la clase LR\_ASK para abordar dos desafíos clave en el entrenamiento de redes neuronales profundas: el ajuste del learning rate y la prevención del sobreajuste. Esta clase incluye funciones para ajustar el learning rate durante el entrenamiento y para aplicar técnicas de regularización, como la parada temprana y la regularización L2, que ayudan a prevenir el sobreentrenamiento.
* **Prueba de diferentes arquitecturas de red neuronal:** Se implementó un proceso de búsqueda aleatoria de hiperparámetros para probar diferentes arquitecturas de red neuronal. Este proceso incluye la generación aleatoria de diferentes combinaciones de hiperparámetros, como el número de capas ocultas, el número de neuronas en cada capa, la función de activación, el optimizador y el learning rate. Cada combinación de hiperparámetros se utiliza para entrenar una red neuronal y evaluar su rendimiento en un conjunto de validación. La arquitectura que produce el mejor rendimiento en el conjunto de validación se selecciona como la arquitectura óptima.
* **Graficar y mostrar los resultados de todas las arquitecturas obtenidas:** Finalmente, se grafican y muestran los resultados de todas las arquitecturas obtenidas para comparar su rendimiento. Esto incluye la visualización de las curvas de aprendizaje, la matriz de confusión, las métricas de rendimiento como la precisión, la sensibilidad, la especificidad y el área bajo la curva ROC. Esta comparación permite identificar la arquitectura que mejor se desempeña en la tarea de clasificación.

En los siguientes pasos, iremos desarrollando cada uno de estos puntos en detalle.

**Configuración de una arquitectura de prueba**

En la fase de configuración de una arquitectura de prueba, se estableció una estructura básica de red neuronal que sirvió como punto de partida para la exploración de diferentes arquitecturas. Este paso fue crucial para asegurar que el modelo podía aprender de los datos, aunque la precisión inicial del modelo no fuera la más alta.

Se entrenó y evaluó el modelo utilizando una serie de funciones:

La función **train\_and\_evaluate\_model** se utilizó para entrenar el modelo con los datos de entrenamiento y validación, y luego evaluar su rendimiento con los datos de prueba. Tras el entrenamiento, se calcularon varias métricas de rendimiento, como la precisión, la sensibilidad y el F1 score, y se almacenaron en un diccionario.

La función **plot\_and\_save\_history** se utilizó para visualizar y guardar el historial de entrenamiento del modelo. Esta función genera dos gráficos que muestran cómo cambian la precisión y la pérdida del modelo durante el entrenamiento en los conjuntos de entrenamiento y validación. Esta visualización fue esencial para entender cómo el modelo aprendía de los datos a lo largo del tiempo.

Finalmente, se realizaron predicciones en el conjunto de datos de prueba utilizando el modelo entrenado y se visualizaron los resultados con la función **plot\_and\_display\_results**. Esta función selecciona aleatoriamente un subconjunto de imágenes de prueba, realiza predicciones en estas imágenes y las muestra junto con sus etiquetas verdaderas y predichas y la imagen del niño. Esta etapa fue importante para observar la salida del modelo y entender cómo se comportaba con datos nuevos.

**Data augmentation**

El siguiente paso en el proceso fue la implementación de la técnica de data augmentation. Data augmentation es una estrategia que permite aumentar la cantidad y diversidad de los datos de entrenamiento sin la necesidad de recopilar nuevos datos. Esto se logra aplicando una serie de transformaciones aleatorias a las imágenes existentes, como rotaciones, traslaciones, zoom y volteo horizontal.

Estas transformaciones generan nuevas versiones de las imágenes existentes que, aunque son diferentes en apariencia, todavía representan la misma clase que la imagen original. Por ejemplo, una imagen de un niño autista sigue siendo una imagen de un niño autista incluso después de ser rotada o volteada horizontalmente.

Esta técnica es especialmente útil en el contexto del aprendizaje profundo, ya que los modelos de aprendizaje profundo suelen requerir grandes cantidades de datos de entrenamiento para alcanzar su máximo rendimiento. Al aumentar la cantidad y diversidad de los datos de entrenamiento, la data augmentation puede ayudar a mejorar la capacidad de la red para generalizar a nuevas imágenes, lo que a su vez puede mejorar el rendimiento del modelo en los datos de prueba

Para implementar data augmentation, se creó un generador de imágenes utilizando la clase ImageDataGenerator de Keras. Este generador se configuró para aplicar una serie de transformaciones aleatorias a las imágenes:

La función ImageDataGenerator genera lotes de imágenes con datos en tiempo real. Las imágenes se generan en el momento del entrenamiento. Las transformaciones incluyen rotaciones de hasta 5 grados, traslaciones de hasta el 5% del ancho y la altura de la imagen, zoom de hasta el 5%, y volteo horizontal. El parámetro fill\_mode se establece en 'nearest', lo que significa que los píxeles vacíos que se crean como resultado de las transformaciones se llenan con el valor del píxel más cercano.

Luego, se editó la función train\_and\_evaluate\_model para utilizar el generador de imágenes durante el entrenamiento. En lugar de pasar las imágenes de entrenamiento directamente a la función model.fit, se pasaron a través del generador de imágenes utilizando la función datagen.flow. Esto hizo que el modelo se entrenara en los lotes de imágenes generados por el generador de imágenes, que incluían las transformaciones aleatorias.

De esta manera, se incorporó la data augmentation en el proceso de entrenamiento, lo que permitió aumentar la cantidad de datos de entrenamiento y mejorar la capacidad de la red para generalizar a nuevas imágenes.

**Ajuste del learning rate y prevención del sobreajuste:**

Después de establecer una arquitectura de prueba para la red neuronal y configurar la técnica de data augmentation, se abordaron los dos siguientes desafíos clave en el entrenamiento de redes neuronales profundas: el ajuste del learning rate y la prevención del sobreajuste.

El aprendizaje profundo a menudo implica el entrenamiento de modelos con millones de parámetros en grandes conjuntos de datos. Durante este proceso, el learning rate —la magnitud de los ajustes realizados a los parámetros del modelo durante el entrenamiento— juega un papel crucial. Un learning rate demasiado alto puede hacer que el modelo salte sobre el mínimo global de la función de pérdida, mientras que un learning rate demasiado bajo puede hacer que el entrenamiento sea excesivamente lento o que el modelo se quede atascado en mínimos locales. Por lo tanto, ajustar el learning rate de manera adaptativa puede ayudar a mejorar la eficiencia y la efectividad del entrenamiento.

La clase LR\_ASK aborda este desafío ajustando el learning rate en función del rendimiento del modelo en el conjunto de validación. Si la pérdida de validación aumenta, lo que indica que el modelo puede estar sobreajustando los datos de entrenamiento, LR\_ASK reduce el learning rate y carga los mejores pesos del modelo. Este enfoque permite al modelo hacer ajustes más pequeños a sus parámetros, lo que puede ayudar a prevenir el sobreajuste. Además, si la pérdida de validación disminuye, LR\_ASK guarda los pesos del modelo, lo que permite al modelo retener la configuración de parámetros que produce el mejor rendimiento en el conjunto de validación.

El segundo desafío que LR\_ASK aborda es la prevención del sobreajuste. El sobreajuste ocurre cuando un modelo aprende los datos de entrenamiento tan bien que se desempeña mal en los datos de validación o de prueba. Esto suele ocurrir cuando el modelo es demasiado complejo y tiene demasiados parámetros en relación con la cantidad de datos de entrenamiento disponibles.

LR\_ASK previene el sobreajuste a través de una técnica conocida como "early stopping". En lugar de entrenar el modelo para un número fijo de épocas, LR\_ASK evalúa el rendimiento del modelo en el conjunto de validación después de cada época. Si la mejora en la pérdida de validación es inferior a un umbral mínimo, LR\_ASK detiene el entrenamiento temprano. Esto evita que el modelo se sobreajuste a los datos de entrenamiento al detener el entrenamiento antes de que el modelo comience a aprender ruido irrelevante.

A continuación, se presenta la estructura de la clase:

**Parametros**

- model: El modelo que se está entrenando.

- epochs: El número total de épocas para las que se entrenará el modelo.

- ask\_epoch: La época en la que se evaluará si se debe continuar el entrenamiento.

- dwell: Un booleano que determina si se deben cargar los mejores pesos cuando la pérdida de validación aumenta.

- factor: El factor por el cual se multiplicará el learning rate cuando la pérdida de validación aumente.

- val\_acc\_threshold y val\_loss\_threshold: Los umbrales de precisión y pérdida de validación para ajustar el learning rate.

- highest\_val\_acc: La precisión de validación más alta alcanzada durante el entrenamiento.

**Metodos**

- \_\_init\_\_: Inicializa la clase con los parámetros proporcionados.

- get\_list: Devuelve las listas de mejoras porcentuales en la pérdida de validación y entrenamiento.

- on\_train\_begin: Se ejecuta al inicio del entrenamiento. Establece el número de épocas y verifica si se debe preguntar si continuar el entrenamiento.

- on\_train\_end: Se ejecuta al final del entrenamiento. Carga los mejores pesos y muestra el tiempo de entrenamiento.

- on\_epoch\_end: Se ejecuta al final de cada época. Calcula la mejora en la pérdida de validación y entrenamiento, ajusta el learning rate y decide si se deben cargar los mejores pesos.

El método on\_epoch\_end es el corazón de esta clase. Al final de cada época, calcula la mejora en la pérdida de validación y entrenamiento. Si la pérdida de validación disminuye, guarda los pesos del modelo y ajusta el learning rate. Si la pérdida de validación aumenta, decide si cargar los mejores pesos y ajustar el learning rate en función del parámetro dwell.

Además, si se ha alcanzado la ask\_epoch, el método on\_epoch\_end evalúa si se debe continuar el entrenamiento. Si la mejora en la pérdida de validación es mayor que un umbral mínimo, se añaden más épocas al entrenamiento y se ajusta el learning rate en función de la precisión de validación.

**Prueba de diferentes arquitecturas de red neuronal**

El proceso de prueba de diferentes arquitecturas de red neuronal convolucional (CNN) en este proyecto se realizó mediante la búsqueda aleatoria de hiperparámetros. La búsqueda de hiperparámetros es una técnica comúnmente utilizada para optimizar los modelos de aprendizaje automático y profundo. Consiste en probar diferentes combinaciones de hiperparámetros y seleccionar la combinación que produce el mejor rendimiento del modelo.

El espacio de hiperparámetros fue definido en un diccionario llamado param\_grid. Este diccionario incluye diferentes opciones para los siguientes hiperparámetros:

* dropout\_rate: La tasa de abandono controla la cantidad de neuronas que se "desactivan" durante el entrenamiento para prevenir el sobreajuste. Un valor más alto puede ayudar a prevenir el sobreajuste, pero también puede llevar a un aprendizaje insuficiente si es demasiado alto.
* learning\_rate: La tasa de aprendizaje controla la rapidez con la que el modelo aprende de los datos. Un valor demasiado alto puede hacer que el modelo se salte el mínimo global, mientras que un valor demasiado bajo puede hacer que el aprendizaje sea demasiado lento.
* l2\_reg: La regularización L2 es una técnica que penaliza los pesos grandes en el modelo para prevenir el sobreajuste. Un valor más alto impone una penalización más fuerte y puede ayudar a prevenir el sobreajuste, pero también puede llevar a un aprendizaje insuficiente si es demasiado alto.
* optimizer: El optimizador controla cómo se actualizan los pesos del modelo durante el entrenamiento. Diferentes optimizadores pueden converger a diferentes velocidades y tener diferentes comportamientos en términos de cómo manejan los mínimos locales.
* activation: La función de activación controla la transformación no lineal que se aplica a las salidas de las neuronas. Diferentes funciones de activación pueden tener diferentes propiedades y ser más adecuadas para diferentes tipos de problemas.
* n\_conv\_layers: El número de capas convolucionales controla la profundidad de la red. Una red más profunda puede aprender representaciones más complejas, pero también es más propensa al sobreajuste y requiere más tiempo y recursos para entrenar.
* filters\_per\_layer: El número de filtros por capa controla la anchura de la red. Un número mayor de filtros puede permitir que la red aprenda representaciones más complejas, pero también puede aumentar el riesgo de sobreajuste.
* kernel\_size: El tamaño del núcleo controla el tamaño de la "ventana" que la capa convolucional utiliza para escanear la imagen. Un tamaño de núcleo más grande puede permitir que la red capture características más grandes, pero también puede hacer que la red sea menos sensible a las características más pequeñas.
* dense\_neurons: El número de neuronas en la capa densa controla la capacidad de la red para aprender representaciones complejas en el espacio de características de alto nivel. Un número mayor de neuronas puede permitir que la red aprenda representaciones más complejas, pero también puede aumentar el riesgo de sobreajuste.

Para probar diferentes combinaciones de estos hiperparámetros, se utiliza la clase ParameterSampler de la biblioteca scikit-learn. Esta clase genera una lista de combinaciones de hiperparámetros seleccionadas aleatoriamente a partir del espacio de hiperparámetros definido. Luego, para cada combinación de hiperparámetros, se crea, entrena y evalúa un modelo. Los resultados de cada modelo se guardan y se utilizan para seleccionar la mejor arquitectura de modelo.

Para cada combinación de hiperparámetros, se crea un modelo utilizando la función create\_model con los hiperparámetros seleccionados.

El modelo se entrena utilizando el conjunto de datos de entrenamiento y el generador de aumento de datos, y se evalúa utilizando los conjuntos de datos de validación y prueba. Si la precisión de prueba del modelo es superior al 80%, se guarda el modelo, sus gráficas y sus hiperparámetros.

**PT4-A3: Entrenar y validar las arquitecturas de red neuronal**

ESTA PARTE SON SOLO RESULTADOS

**PT4-A4: Implementar la interfaz de usuario**

# Resultados

# Implicaciones Éticas e Impacto Social

Reflexión acerca de las implicaciones éticas y antropológicas derivadas del proyecto, así como el impacto social del mismo

# Conclusiones

Elaboración de las principales conclusiones que se extraen tras el desarrollo del proyecto. Análisis de las posibilidades de evolución futura del trabajo presentado.

# Otros Méritos del Proyecto

Aquí se podrán describir todos los méritos adicionales del proyecto, es decir, resultados obtenidos no esperados, que aportan un valor adicional al proyecto (disponibilidad pública del sistema o los resultados, sitio web, integración de disciplinas, uso de SW libre, elementos de accesibilidad, etc.)

# Bibliografía

|  |  |
| --- | --- |
| [1] | «¿Por qué es importante el diagnóstico precoz en el autismo?,» 4 3 2020. [En línea]. Available: https://invanep.com/blog\_invanep/por-que-es-importante-el-diagnostico-precoz-en-el-autismo. |
| [2] | Red Cenit | Centro de Desarrollo Cognitivo, «Diagnosticar de manera precoz el autismo mediante realidad virtual e inteligencia artificial: Proyecto T-EYE,» 31 5 2022. [En línea]. Available: https://www.redcenit.com/diagnosticar-de-manera-precoz-el-autismo-mediante-realidad-virtual-e-inteligencia-artificial-proyecto-t-eye/. |
| [3] | «Evaluation of an artificial intelligence-based medical device for diagnosis of autism spectrum disorder,» 5 5 2022. [En línea]. Available: https://www.nature.com/articles/s41746-022-00598-6. |
| [4] | «Usando Machine Learning para detectar autismo - Neurometrics LAB,» 26 5 2022. [En línea]. Available: https://neurometrics.la/dia-mundial-del-autismo-usando-machine-learning-para-detectar-los-tea/. |
| [5] | «DETECTAA-AI: Inteligencia Artificial en el diagnóstico presuntivo de trastornos del desarrollo en niños,» 4 4 2022. [En línea]. Available: https://saturdays.ai/2021/11/19/inteligencia-artificial-diagnostico-trastornos-desarrollo/. |
| [6] | «Streamlit Docs,» [En línea]. Available: https://docs.streamlit.io/. |
| [7] | TensorFlow, «Module: tf.keras  |  TensorFlow v2.11.0,» [En línea]. Available: https://www.tensorflow.org/api\_docs/python/tf/keras.  [8] Aldridge, K., et al., «Análisis detallado de los rasgos faciales de los niños con TEA utilizando técnicas de morfometría geométrica,» 2011. [En línea].  [9] Obafemi-Ajayi, T., et al., «Análisis de características faciales de personas con TEA a través de aprendizaje automático,» 2015. [En línea].  [10] Walsh, P., et al., «Análisis exhaustivo de los desafíos asociados con la búsqueda de biomarcadores para el autismo,» 2011. [En línea].  [11] Ozonoff, S., et al., «Estudio longitudinal sobre el desarrollo de los signos conductuales de autismo desde el nacimiento hasta los 3 años,» 2010. [En línea].  [12] Shen, M. D., et al., «Investigación sobre cambios cerebrales tempranos en niños con TEA a través de resonancia magnética,» 2013. [En línea].  [13] Anzulewicz, A., et al., «Examen de patrones de movimiento en niños con TEA durante el juego con tabletas inteligentes,» 2016. [En línea].  [14] Wang, Y., et al., «Análisis de la microbiota intestinal en niños con y sin autismo,» 2022. [En línea].  [15] Zhang, Y., et al., «Investigación sobre la relación de ciertos biomarcadores salivales con el autismo,» 2022. [En línea]. |

# Anexo A: Planificación detallada

# Anexo B: Requisitos de Usuario

# Anexo B: Diseño del Sistema

# Anexo C: Manuales

## Manual de Usuario

## Manual de Instalación

# Anexo D: Contenido del CD

AÑADIR TODOS LOS APÉNDICES ADICIONALESQUE SEAN NECESARIOS