**UNIVERSIDAD NACIONAL AUTÓNOMA DE HONDURAS**

**(VALLE DE SULA)**  
**Catedrático:**

Juan Enrique Alvarenga Rodas

**Clase:**   
Inteligencia Artificial

**Sección:**   
1800  
**Integrantes:**  
Andrea Estefanía Andrade

20142003006

Carlos Eduardo Argueta Rivera  
20172001120  
  
Rafael Ernesto Morales Díaz  
20202001873  
**Tema:**Formato de Propuesta de Proyecto  
**Fecha:**27-7-2025

**Resumen del Proyecto**

El presente proyecto tiene como finalidad desarrollar un modelo de inteligencia artificial capaz de predecir si un paciente diagnosticado con COVID-19 está vivo o ha fallecido, en función de variables clínicas contenidas en el conjunto de datos **Covid Data.csv** obtenido desde la plataforma **Kaggle**. Este archivo incluye registros de miles de pacientes con información relevante como edad, enfermedades preexistentes, síntomas, historial médico, entre otros.

El enfoque del proyecto se basa exclusivamente en el uso de aprendizaje automático supervisado, mediante algoritmos como **árboles de decisión**, entrenados con datos balanceados usando la técnica **SMOTE**, debido al desbalance natural entre pacientes fallecidos y sobrevivientes. El análisis se realiza en entornos como **Google Colab o Visual Studio Code**, sin necesidad de implementar una aplicación con entrada manual de datos ni interfaces de usuario.

**Alcances**

El proyecto contempla los siguientes componentes:

* **Carga y análisis del dataset Covid Data.csv** desde Kaggle, como única fuente de datos del proyecto.
* **Limpieza y transformación de datos**, incluyendo el tratamiento de valores atípicos, reemplazo de datos faltantes y codificación de variables.
* **Balanceo de clases** con la técnica **SMOTE** para mejorar la representatividad de los datos de pacientes fallecidos.
* **Entrenamiento de modelos** de aprendizaje automático (como árboles de decisión) para clasificar pacientes según su desenlace clínico (vivo o fallecido).
* **Evaluación del modelo** con métricas como exactitud, matriz de confusión, recall, precisión y F1-score.
* **Visualización de los resultados**, incluyendo gráficos del árbol de decisión y matrices de confusión.

Supuestos y Limitantes

**Supuestos:**

* Se utilizará el dataset Covid Data.csv de Kaggle, el cual contiene variables relevantes y una columna que indica si el paciente está vivo o muerto (a partir de DATE\_DIED).
* El equipo tiene acceso a herramientas como Python, pandas, scikit-learn, matplotlib, seaborn e imblearn, y cuenta con los conocimientos técnicos para su implementación.
* El análisis se realizará en entorno de desarrollo como Google Colab o Visual Studio Code.

**Limitantes:**

* El modelo no recibirá entradas en tiempo real ni contará con interfaz gráfica.
* No se contará con datos clínicos en tiempo real; todo se basa en datos históricos del dataset.
* El modelo no será validado clínicamente ni se implementará como aplicación para uso médico; su uso es únicamente académico y exploratorio.