**ENTREGA FINAL**

**Repositorios Github:** <https://github.com/jetabaresj/DespliegueTeam23>

**Vídeo:**

**Problema y su contexto:** La insuficiencia cardíaca es una condición de salud grave que afecta a millones de personas en el mundo, caracterizada por la incapacidad del corazón para bombear suficiente sangre al cuerpo. Este problema se agrava con la presencia de factores como edad avanzada, comorbilidades (diabetes, hipertensión, anemia) y hábitos de vida poco saludables (como el tabaquismo). Estas variables tienen un impacto significativo en la supervivencia del paciente.

En este escenario, surge la necesidad de herramientas predictivas que permitan a los médicos anticipar el riesgo de mortalidad de un paciente basándose en sus características clínicas y de estilo de vida. Sin embargo, identificar y priorizar a los pacientes con mayor riesgo puede ser complejo debido a la gran cantidad de factores interrelacionados que influyen en la insuficiencia cardíaca.

**Pregunta de negocio:** ¿Es posible determinar el riesgo de insuficiencia cardíaca en un paciente al comparar sus características médicas y hábitos de vida con los promedios de la población, analizando además la relación entre las covariables y el momento del evento de muerte del paciente?

**Alcance del proyecto:** Desarrollar un desarrollar un tablero para personal médico especializado para visualizar la información médica de un paciente y hábitos de vida, para comparar con la media de una población y determinar los factores de mayor riesgo y estimar el nivel de riesgo de fallecer mediante la librería de scikit-survival y que de esta manera el médico pueda establecer el nivel de seguimiento del paciente.

**Descripción de conjuntos de datos a emplear:** El conjunto de datos contiene los registros médicos de 299 pacientes con insuficiencia cardíaca, recolectados en dos hospitales de Faisalabad, Pakistán, durante 2015. El dataset incluye 13 características que abarcan información clínica, corporal y de estilo de vida, como la presencia de anemia, hipertensión, diabetes, los niveles de creatinina y sodio en sangre, y la fracción de eyección del corazón.

El propósito principal del dataset es predecir, mediante machine learning, el riesgo de sufrir insuficiencia cardíaca de los pacientes y determinar cuáles son las características más importantes que influyen en este resultado. La variable objetivo es un evento de muerte (si el paciente murió o sobrevivió antes del final del período de seguimiento).

El conjunto de datos contiene diversas variables que son fundamentales para analizar y predecir la probabilidad de supervivencia de pacientes con insuficiencia cardíaca.

El data sheart\_failure\_clinical\_records\_dataset.csv es propocionado por Heart Failure Clinical Records, este conjunto de datos está licenciado bajo una licencia Creative Commons Attribution 4.0 International (CC BY 4.0). Esto permite compartir y adaptar los conjuntos de datos para cualquier propósito, siempre que se otorgue el crédito correspondiente.

**Modelos:**

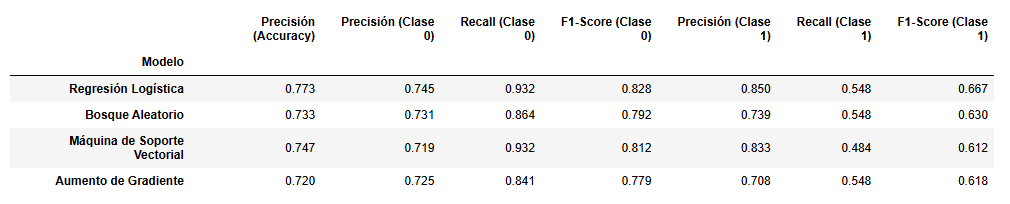
Se realiza un análisis de predicción para eventos de muerte utilizando varios modelos de clasificación: Regresión Logística, Random Forest, SVM y Gradient Boosting .

**Regresión Logística:** Con los mejores hiperparámetros (C=1.0 y solver='liblinear'), obtuvo un accuracy de 77.33%. Su desempeño en términos de métricas muestra un f1-score de 0.83 para la clase 0 (no fallecimiento) y de 0.67 para la clase 1 (fallecimiento). La precisión para la clase 0 fue de 75%, mientras que para la clase 1 fue de 85%, lo que indica que es mejor identificando correctamente los casos de fallecimiento que los no fallecimientos, aunque con menor recall en la clase 1 (55%), sugiriendo que muchos casos positivos fueron clasificados incorrectamente como negativos.

**Random Forest:** Con hiperparámetros optimizados (max\_depth=10 y n\_estimators=100), alcanzó un accuracy de 73.33%. Para la clase 0, el modelo obtuvo un f1-score de 0.79, y para la clase 1 un f1-score de 0.63. A pesar de tener un buen recall para la clase 0 (86%), el recall para la clase 1 fue solo del 55%, similar a la regresión logística, lo que indica que también tiene dificultades en detectar correctamente los casos de fallecimiento.

**Support Vector Machine (SVM)**: Con C=0.1 y kernel='linear' tuvo un accuracy de 74.67%. En términos de precisión, logró un 83% para la clase 1 y 72% para la clase 0. Sin embargo, el recall para la clase 1 fue bajo (48%), lo que significa que, aunque es bastante preciso cuando predice un fallecimiento, no es muy eficaz en detectar todos los casos positivos (fallecimientos). El f1-score para la clase 1 fue de 0.61, lo que refleja un balance moderado entre precisión y recall.

**Gradient Boosting**: Con los parámetros óptimos (learning\_rate=0.1, max\_depth=3, y n\_estimators=100), logró un accuracy de 72%. Tuvo un f1-score de 0.78 para la clase 0 y 0.62 para la clase 1. La precisión para la clase 1 fue de 71%, con un recall de 55%, lo que indica un rendimiento comparable al Random Forest, aunque ligeramente inferior en términos generales.

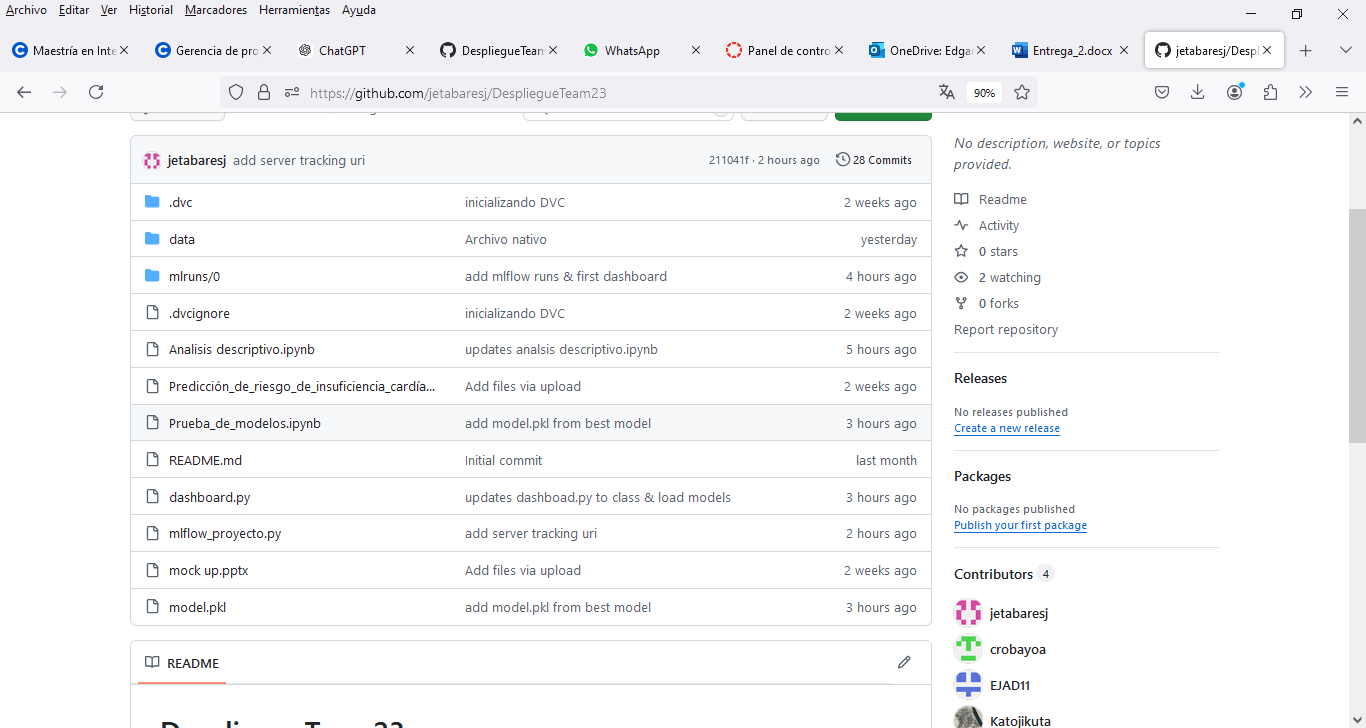


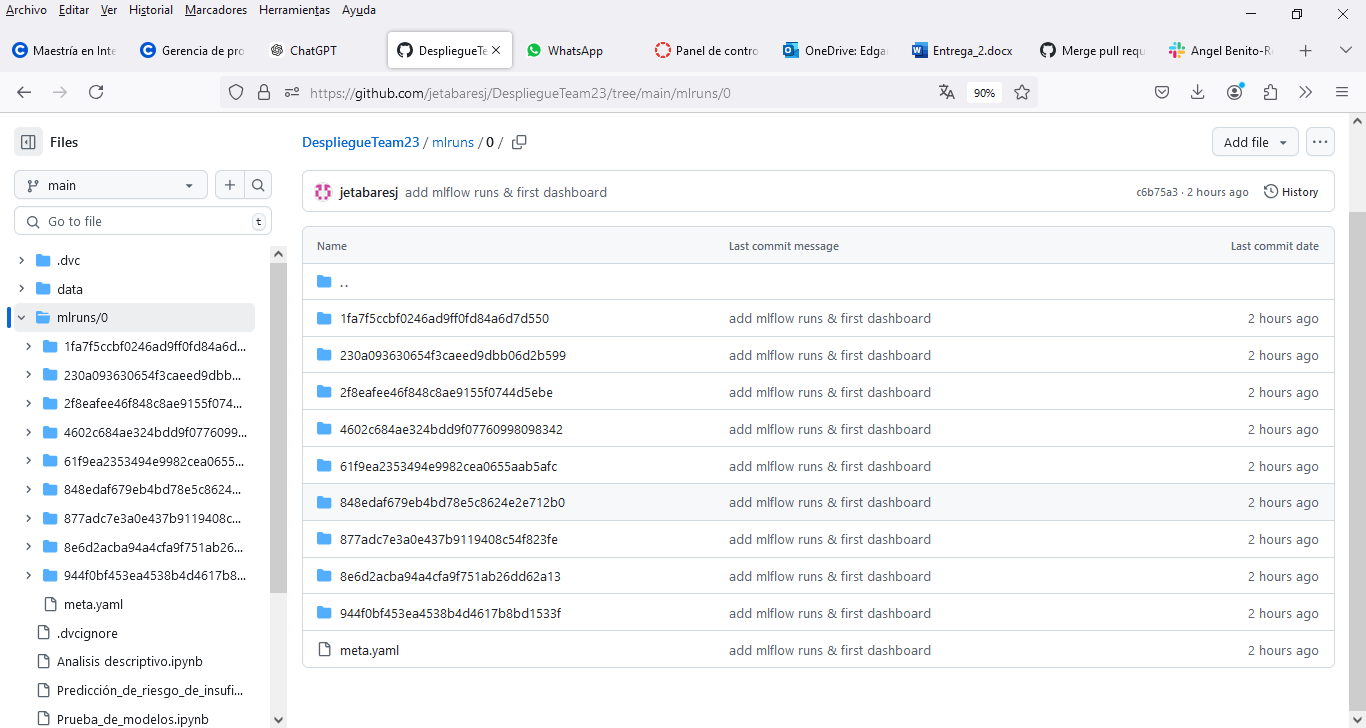
**Comparación General:** El modelo de Regresión Logística tuvo el mejor desempeño general con un accuracy de 77.33%, destacando especialmente en su capacidad para identificar correctamente los casos de fallecimiento (clase 1) con una precisión de 85%.

SVM tuvo un rendimiento ligeramente mejor que Random Forest y Gradient Boosting en términos de accuracy, con un 74.67%, pero su bajo recall en la clase 1 podría limitar su utilidad en un contexto donde es crítico identificar correctamente los casos positivos (fallecimientos). Random Forest y Gradient Boosting obtuvieron resultados similares, con accuracies alrededor del 73% y 72%, respectivamente. Estos modelos parecen ser más equilibrados, aunque no destacan en la detección de la clase minoritaria (fallecimientos).

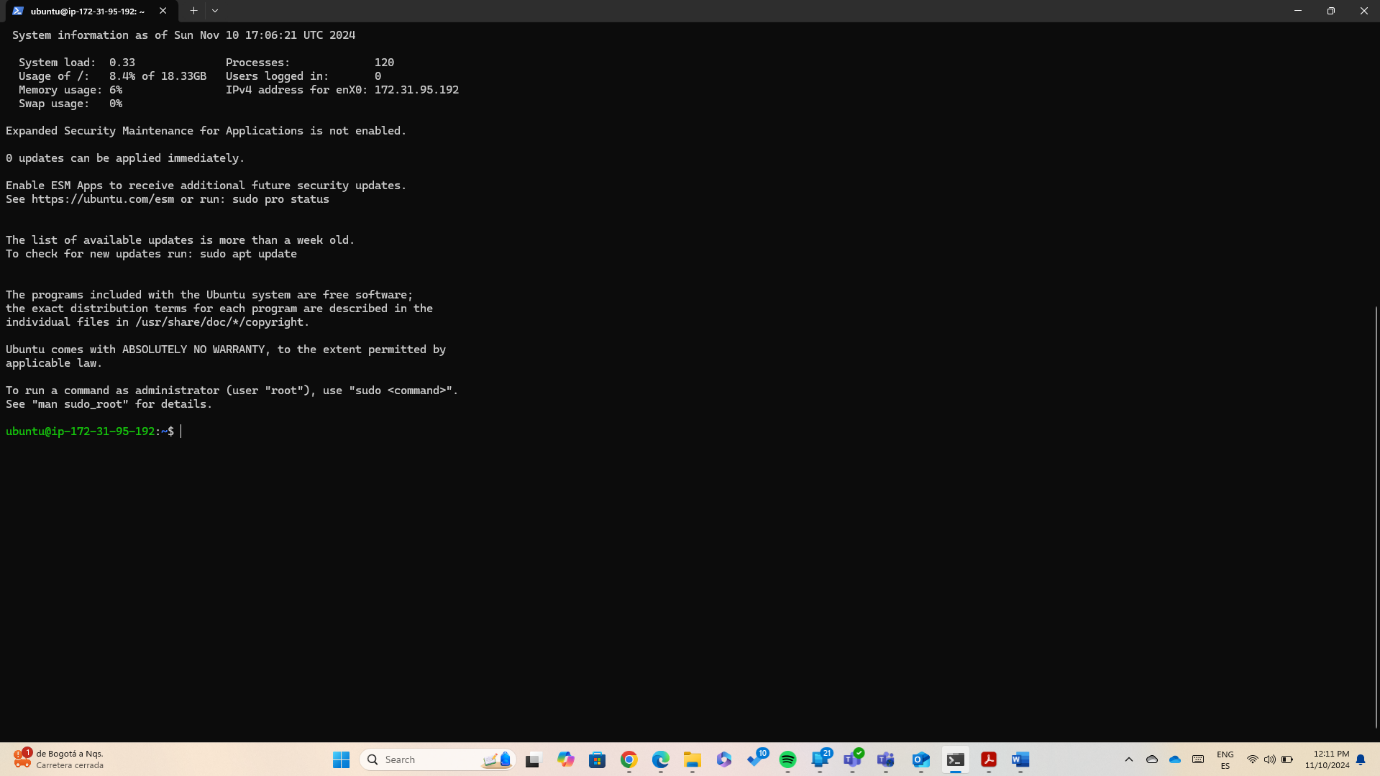
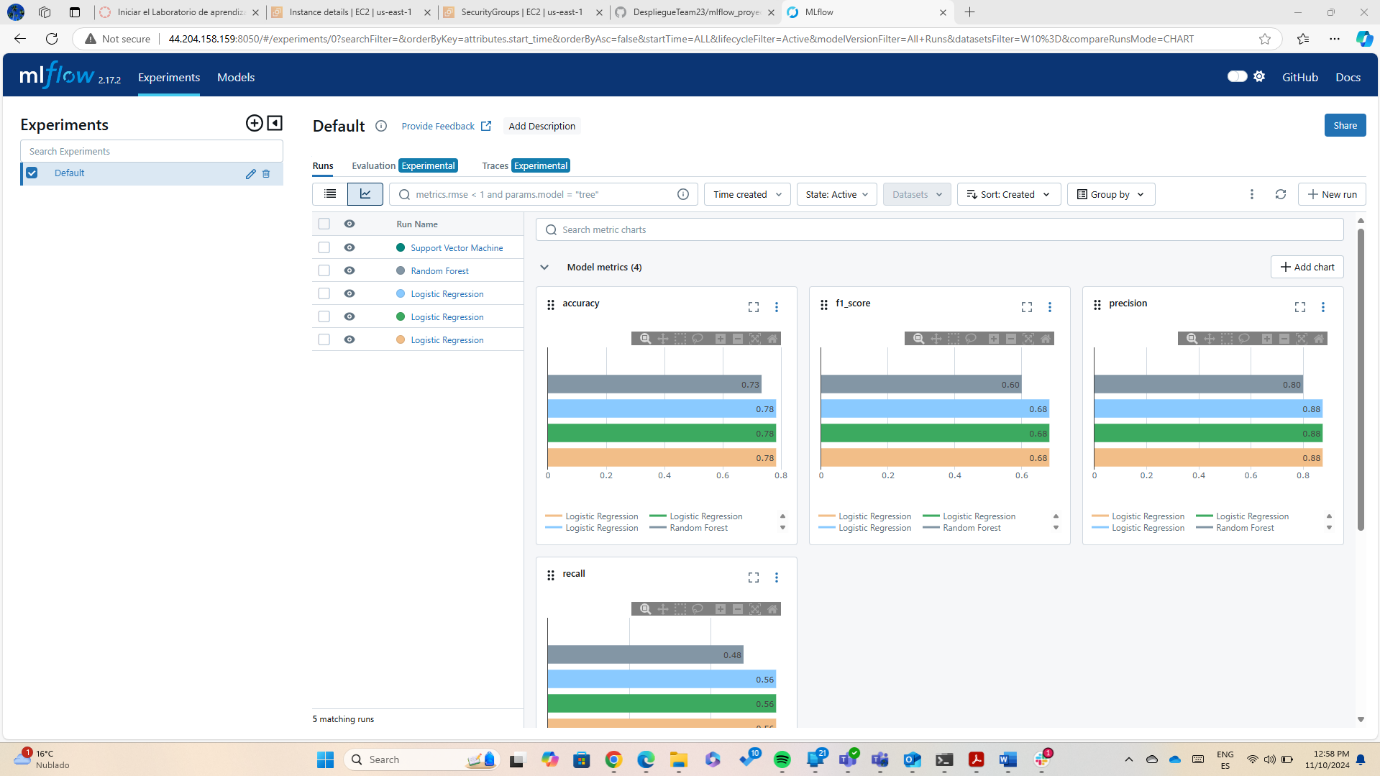
**Conclusión:** Dado que el objetivo es identificar correctamente los casos de fallecimiento, el modelo de Regresión Logística sería la mejor opción en este caso debido a su mayor precisión general y un mejor balance entre precisión y recall para la clase positiva. Sin embargo, se podría explorar un ajuste de hiperparámetros más fino en modelos como SVM o Gradient Boosting para mejorar la sensibilidad hacia la clase minoritaria.

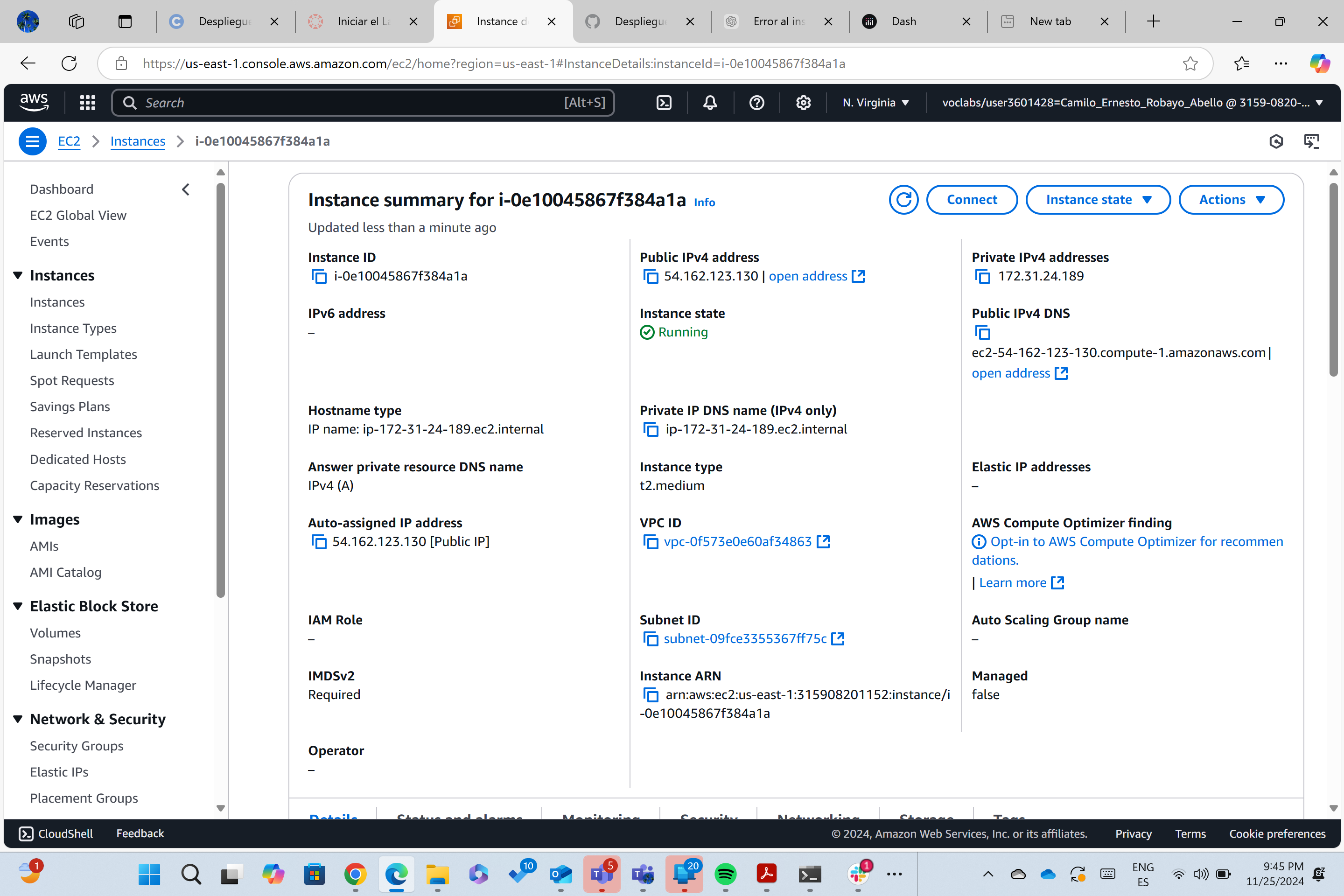
En la carpeta principal se subieron los archivos mlflow\_proyecto.py y dashboard.py los cuales fueron desplegados en la máquina EC2.



En la carpeta **mlruns** quedan los soportes del despliegue de los modelos y el tablero: 

**Despliegue mlflow**

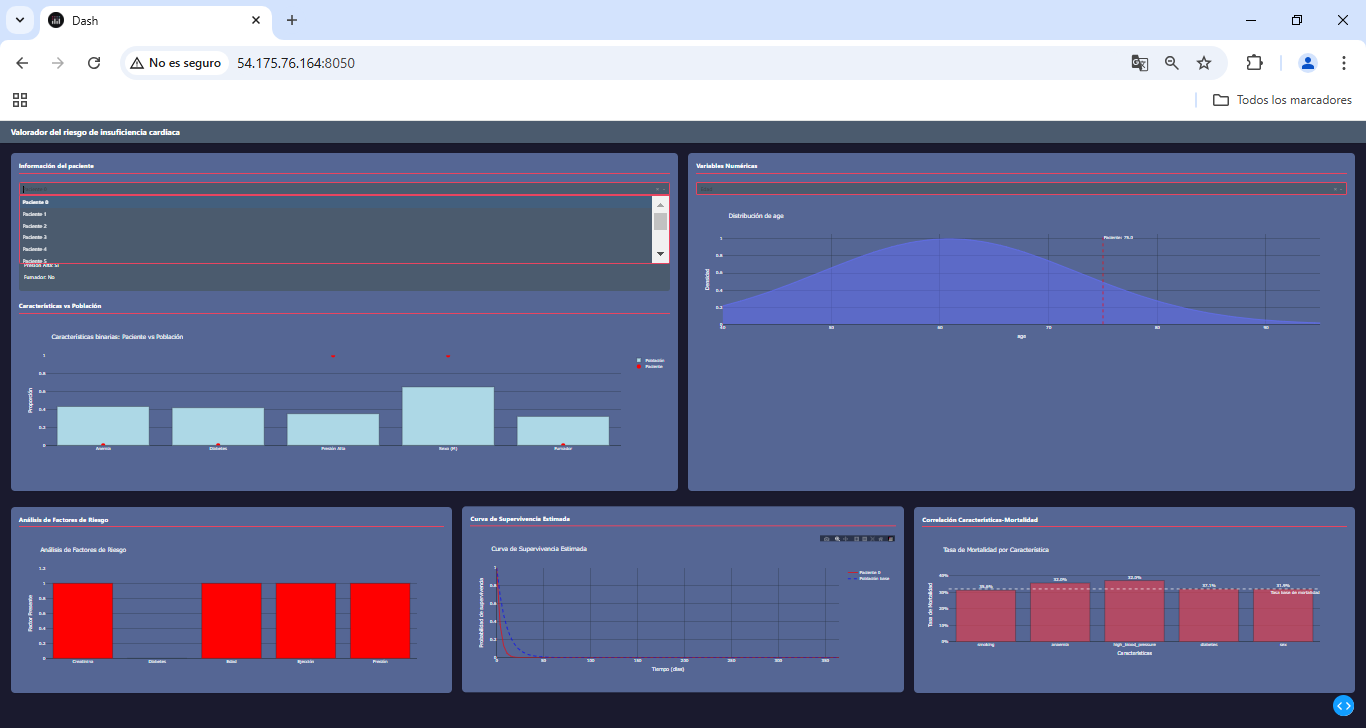
**Conexión máquina virtual****Ejecución modelos**

**Modelos Empaquetados y desplegados:** 

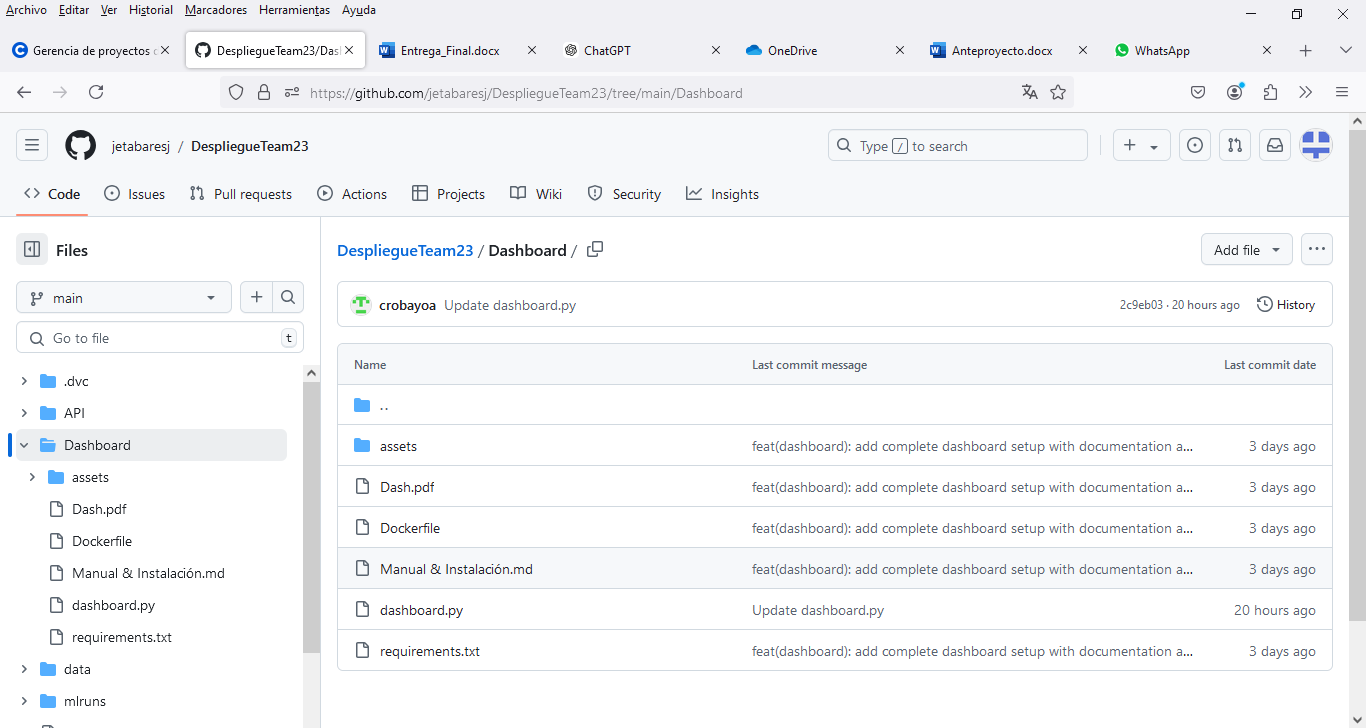
A screenshot of a computer screen

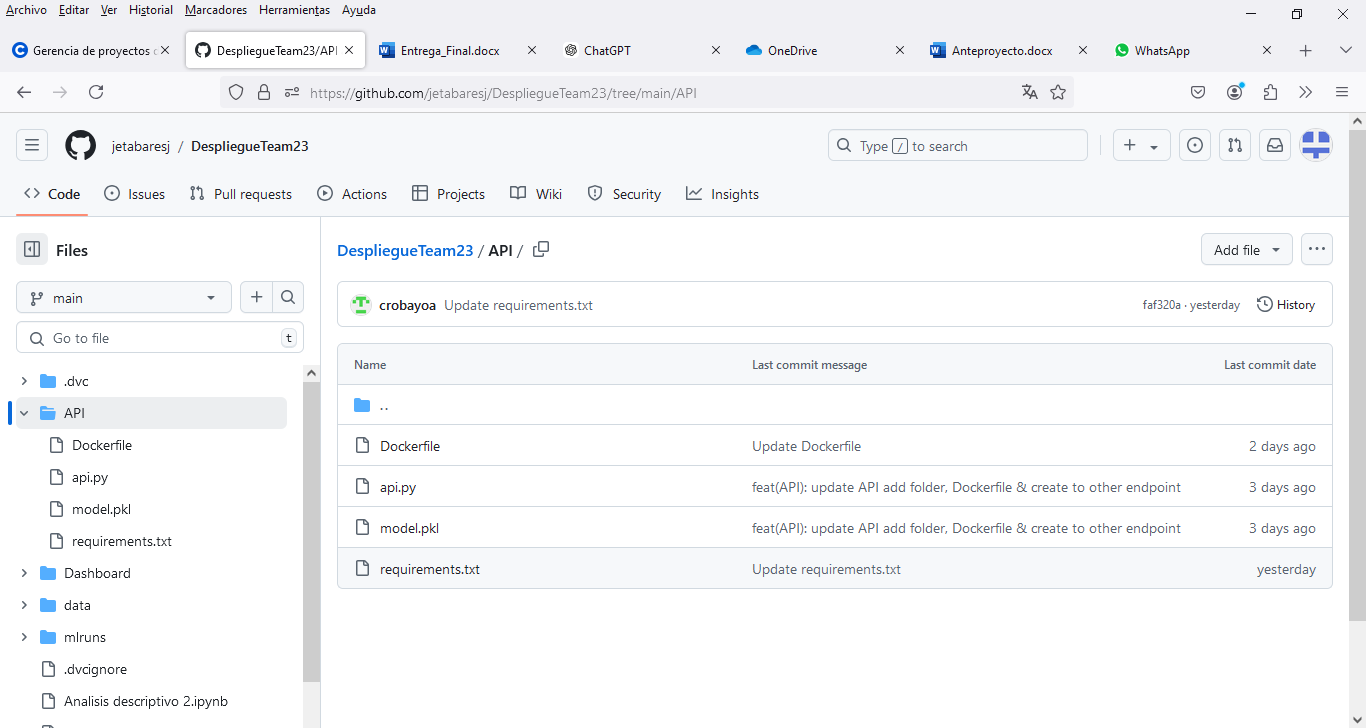
Description automatically generated





**Artefactos para desplegar el tablero y el modelo:**





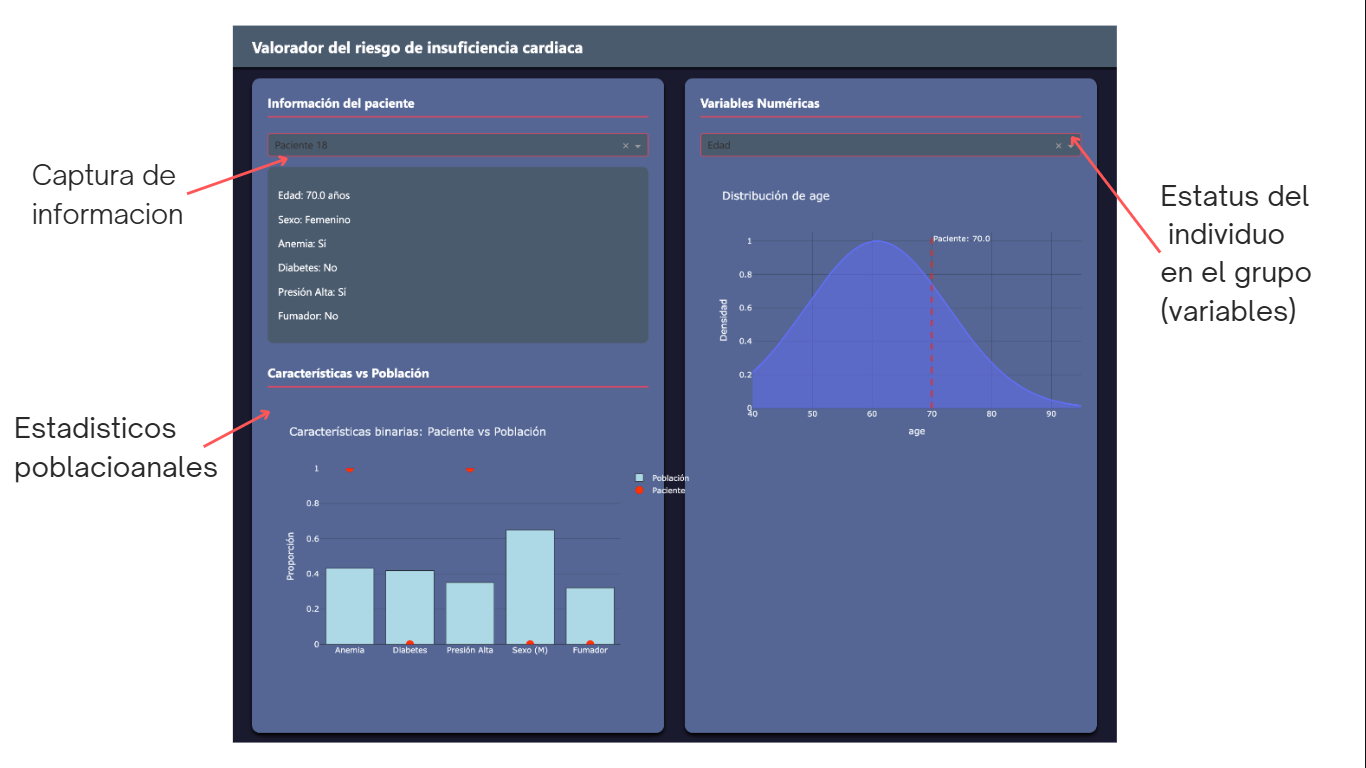
**Manual de usuario del tablero**

1. Introducción

Este dashboard tiene como objetivo evaluar el riesgo de insuficiencia cardíaca en pacientes, empleando un modelo de machine learning y análisis de supervivencia basado en datos clínicos. Su uso principal es brindar un análisis rápido y visual del estado del paciente en relación con una base de población general.

El análisis está fundamentado en datos específicos del paciente, como edad, anemia, diabetes, presión arterial alta y tabaquismo, lo cual permite estimar la probabilidad de supervivencia y entender el impacto de estos factores en la salud del paciente.

2. Descripción de los Componentes del Dashboard



Información del Paciente

Esta sección muestra los datos clínicos del paciente. Los datos visibles incluyen:

- Edad: La edad actual del paciente.

- Sexo: Género del paciente (Femenino o Masculino).

- Anemia: Indica si el paciente tiene o no anemia.

- Diabetes: Muestra si el paciente padece diabetes.

- Presión Alta: Indica si el paciente tiene presión alta.

- Fumador: Muestra si el paciente es fumador o no.

![Información del Paciente](./assets/info.png)

Características vs Población

En esta sección, puedes comparar las características del paciente con las de una base poblacional. Incluye:

- Características Binarias: Un gráfico de barras compara la proporción de características como anemia, diabetes, presión alta, sexo y hábito de fumar entre el paciente y la población general.

- Distribución de Carateristicas Númericas: Un gráfico muestra la distribución en la población general, destacando la edad del paciente para facilitar la comparación.

![Características vs Población](./assets/features\_vs\_pob.png)

![Características vs Población](./assets/features\_vs\_pob1.png)

Análisis de Factores de Riesgo

Este gráfico muestra la presencia de factores de riesgo en el paciente y permite observar su frecuencia en la población. Cada factor (como la creatinina, diabetes, presión arterial, etc.) se compara visualmente para ver su relación con la insuficiencia cardíaca y su relevancia en la población.

![Factores de Riesgo](./assets/factor\_riesgo.png)

Curva de Supervivencia Estimada

La curva de supervivencia proyecta la probabilidad de supervivencia del paciente en el tiempo, comparando sus resultados con una curva de referencia poblacional. Esto ayuda a ver cómo los datos del paciente afectan su probabilidad de supervivencia a lo largo del tiempo.

![Supervivencia Estimada](./assets/survival.png)

Correlación Características-Mortalidad

Este gráfico presenta la tasa de mortalidad asociada con características específicas, tales como fumar o tener presión alta, comparándolas con la tasa base de mortalidad en la población. Ayuda a entender la influencia de cada característica en la mortalidad.

![Características-Mortalidad](./assets/mortalidad.png)

3. Interpretación de Resultados

Probabilidad de Supervivencia

- Interpretación: La curva de supervivencia muestra el porcentaje de probabilidad de que el paciente sobreviva a lo largo del tiempo. Una probabilidad más baja que la de la población general podría indicar un mayor riesgo debido a las condiciones clínicas del paciente.

- Comparación: La línea del paciente en la curva puede compararse con la línea de referencia para ver si está por encima o por debajo de la expectativa de la población.

Tasa de Mortalidad por Característica

- Interpretación: La tasa de mortalidad por característica muestra cómo cada condición (como fumar o tener presión alta) impacta en la mortalidad. Esto permite ver el nivel de riesgo que tiene el paciente en relación con cada característica.

- Comparación: La comparación entre el paciente y la tasa base ayuda a identificar los factores más influyentes en su riesgo de mortalidad.

4. Guía para Navegación y Uso del Dashboard

Seleccionar Paciente

- Cambio de Paciente: Puedes seleccionar entre diferentes pacientes o cargar datos de uno nuevo utilizando el menú de selección en la parte superior del dashboard.

- Entrada de Datos: Los datos pueden ingresarse manualmente o importarse si están disponibles en un archivo compatible con el sistema.

Personalización de Gráficos

- Cambio de Vista: Puedes seleccionar diferentes tipos de gráficos para visualizar los datos, como cambiar entre histogramas y gráficos de barras en la sección de características.

- Filtrar Características: Hay opciones de filtro para ver solo características específicas que pueden ser más relevantes para el análisis de cada paciente.

Exportar Resultados

- Descarga de Datos: Se pueden exportar datos específicos del paciente o gráficas completas en formato PDF o CSV para registros o revisiones posteriores.

5. Recomendaciones de Uso

Consideraciones Clínicas

Los resultados deben ser utilizados por personal médico especializado, ya que este análisis no sustituye un diagnóstico profesional. Los médicos pueden utilizar la información para guiar sus decisiones clínicas, pero deben realizar estudios complementarios para confirmar cualquier conclusión.

Limitaciones del Modelo

- Base de Datos Limitada: El modelo utiliza una base de datos específica para la población de referencia, por lo cual los resultados pueden variar según la procedencia o características de otros pacientes fuera de esta población.

- Factores no Incluidos: Algunos factores de riesgo no incluidos en este análisis podrían influir en la supervivencia del paciente.

6. Preguntas Frecuentes (FAQ)

- ¿Qué hacer si los datos del paciente no son precisos?

- Si notas que hay datos incorrectos o faltantes, puedes actualizar manualmente cada dato ingresando al menú de información del paciente y corrigiendo los valores directamente.

- ¿Puedo usar este análisis para otros tipos de insuficiencia cardíaca?

- Este análisis está específicamente diseñado para insuficiencia cardíaca y puede no ser aplicable a otros tipos de enfermedades cardíacas o condiciones crónicas.

- ¿Cómo se actualiza el modelo de predicción?

- El modelo de predicción se actualiza regularmente para incorporar nuevos datos y mejorar su precisión.

**Manual de instalación del tablero:**

Instalación del Dashboard utilizando Docker

Para facilitar la instalación y despliegue del dashboard, puedes utilizar Docker, que permite ejecutar la aplicación en un contenedor independiente del sistema operativo.

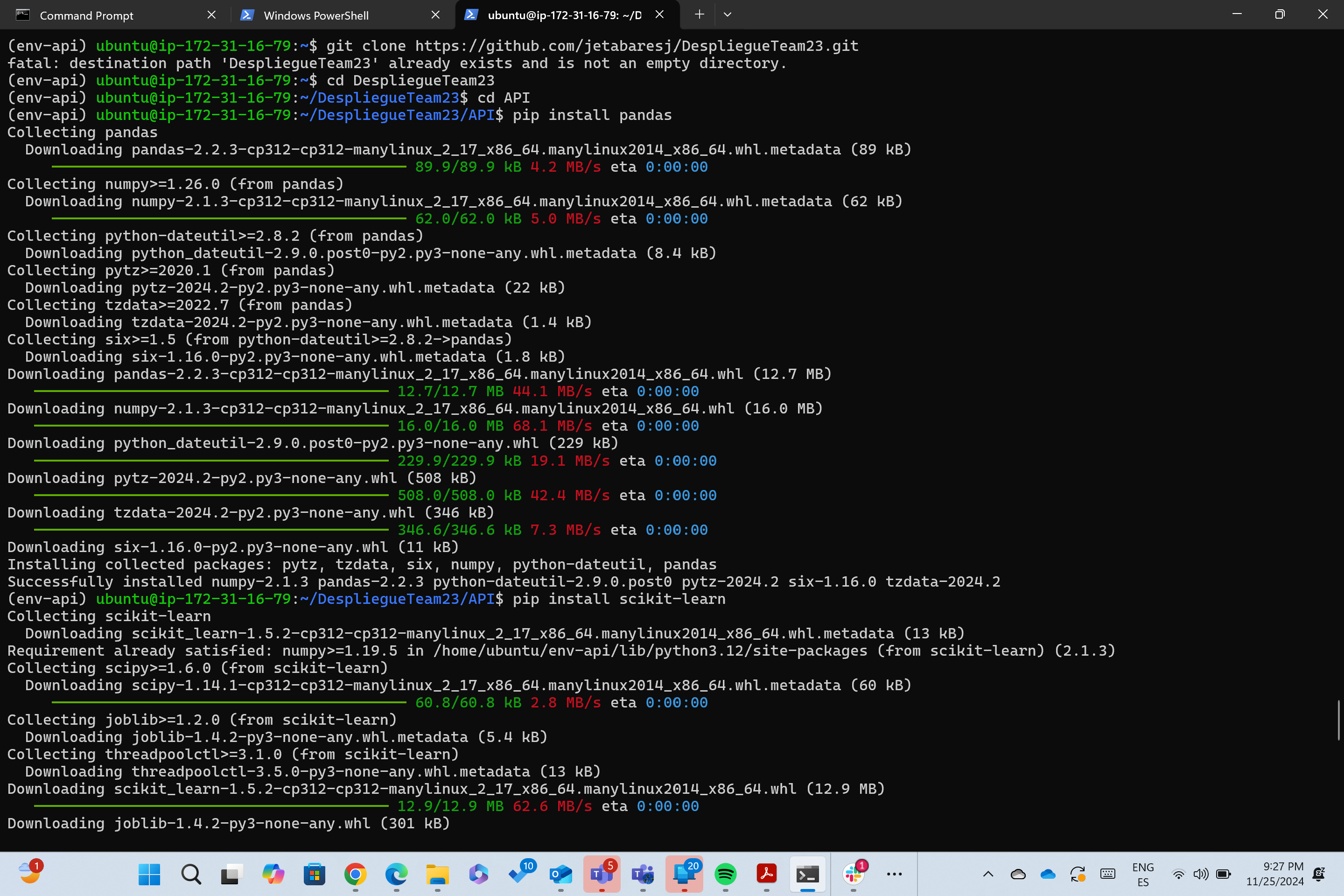
**Requisitos Previos**

1. Docker: Asegúrate de tener Docker instalado en tu sistema. Si no lo tienes, puedes descargarlo e instalarlo desde [https://www.docker.com/get-started](<https://www.docker.com/get-started>).



**Paso a Paso para la Instalación**

1. Clona el repositorio o descarga los archivos necesarios: Asegúrate de tener el archivo `dashboard.py` del código de la aplicación, junto con la carpeta `assets` (si existe) que contiene el archivo CSS para el estilo del dashboard, y el archivo `requirements.txt`.



2. Construye la imagen de Docker: En la terminal, navega hasta el directorio donde tienes los archivos y ejecuta el siguiente comando:

```bash

docker build -t heart\_failure\_dashboard .

```

Este comando construirá una imagen llamada `heart\_failure\_dashboard`.

3. Ejecuta el contenedor: Una vez creada la imagen, ejecuta el siguiente comando para iniciar el contenedor:

```bash

docker run -p 8050:8050 heart\_failure\_dashboard

```

Este comando ejecutará el dashboard y lo expondrá en el puerto `8050` de tu máquina local. Ahora podrás acceder al dashboard en `http://localhost:8050` en tu navegador.

Actualización del Dashboard

**Para realizar cambios en el dashboard**:

1. Modifica el código en los archivos locales.

2. Reconstruye la imagen de Docker con el comando `docker build -t heart\_failure\_dashboard .`.

3. Detén el contenedor anterior y reinícialo con el comando `docker run -p 8050:8050 heart\_failure\_dashboard`.

**Reporte de trabajo en equipo**

En la siguiente tabla se muestra el resumen de las actividades realizadas por cada integrante del equipo en pro de realizar la entrega correspondiente:

|  |  |
| --- | --- |
| Nombre | Actividades |
| Jhon Esteban Tabares Jiménez | -Generación versión final modelos -Generación versión final tablero  -Manual usuario |
| Camilo Robayo Abello | -Despliegue modelos  -Despliegue tablero  -Video presentación |
| Angel Benito Rebollo | -Ajustes presentación  -Ajustes tablero  -Video presentación |
| Javier Abril | -Estructura presentación -Video presentación  -Informe final y carga en Coursera |

Se adicionan también las contribuciones realizadas al repositorio por cada uno de los integrantes del grupo.

