En este proyecto, nos embarcamos en la tarea de construir y evaluar un modelo de Regresión Logística para pronosticar el desenlace fatal en pacientes con problemas cardíacos, utilizando el dataset "Heart\_failure\_clinical\_records\_dataset.csv". El dataset consta de 299 observaciones y 13 categorías, reflejando los resultados de exámenes médicos en pacientes cardíacos, con la variable objetivo siendo 'DEATH\_EVENT'. A continuación, se detallan las etapas clave y hallazgos del proyecto.

Carga de Librerías Básicas: Inicialmente, se cargaron las librerías esenciales de Python para manipulación de datos y análisis estadístico, como pandas, numpy y matplotlib. A medida que avanzaba el proyecto, se fueron incorporando librerías adicionales según la necesidad.

Carga del Dataset: Se cargó el dataset 'Heart\_failure\_clinical\_records\_dataset.csv' utilizando pandas, una herramienta fundamental para el manejo de datos.

Exploración de Datos: Se utilizó .shape para entender la estructura del dataset, .columns para listar las variables, .info() para obtener un resumen y .describe() para estadísticas descriptivas. Esta exploración inicial fue crucial para entender la naturaleza de los datos.

Análisis de Correlación y Mapa de Calor: Se calculó la correlación entre las variables utilizando la función de correlación de pandas y se visualizó mediante un mapa de calor usando seaborn. Esto permitió identificar relaciones potencialmente significativas entre las variables.

Conclusión Inicial: Del análisis de correlación, se extrajeron insights valiosos sobre qué variables podrían ser predictores significativos para el evento de muerte.

Separación del Dataset: Se dividió el dataset en dos conjuntos: uno para entrenamiento y otro para pruebas, utilizando un ratio estándar de división.

Instancia del Modelo de Regresión Logística: Se utilizó Scikit-Learn para instanciar un modelo de Regresión Logística, una elección adecuada para este tipo de problemas de clasificación binaria.

Ajuste del Modelo: Se ajustó el modelo con los datos de entrenamiento, una etapa crítica para el aprendizaje del modelo.

Pronóstico con Datos de Prueba: Se emplearon los datos de prueba para hacer predicciones con el modelo ajustado.

Evaluación del Desempeño: Se evaluó la precisión (accuracy) del modelo tanto en el conjunto de entrenamiento como en el de prueba, lo que proporcionó una medida inicial de la efectividad del modelo.

Selección de Variables Significativas: Se utilizó un método de selección de variables, como podría ser la eliminación hacia atrás, para identificar aquellas con mayor poder predictivo.

Entrenamiento con Variables Optimizadas: El modelo se reentrenó, esta vez solo con las variables identificadas como más significativas.

Pronóstico con Variables Optimizadas: Se realizaron nuevas predicciones, ahora con un modelo más afinado.

Evaluación Final del Desempeño: Se compararon las precisiones de entrenamiento y prueba para evaluar la mejora en el rendimiento del modelo tras la optimización de variables.

Conclusión Final: Las conclusiones finales giraron en torno a la efectividad del modelo optimizado, la importancia de la selección de variables, y las implicaciones prácticas de los resultados en un contexto clínico.

En resumen, este proyecto demostró la utilidad de la Regresión Logística en el ámbito médico, particularmente para predecir eventos críticos en pacientes con condiciones cardíacas, destacando la importancia del análisis de datos y la selección de variables en la construcción de modelos predictivos eficaces.