

GRUPO 14



ELEGIBILIDAD DE UN PACIENTE PARA ENSAYOS CLÍNICOS DE CÁNCER

Laura Rodriguez
Jaime Carvajal
John Guzmán



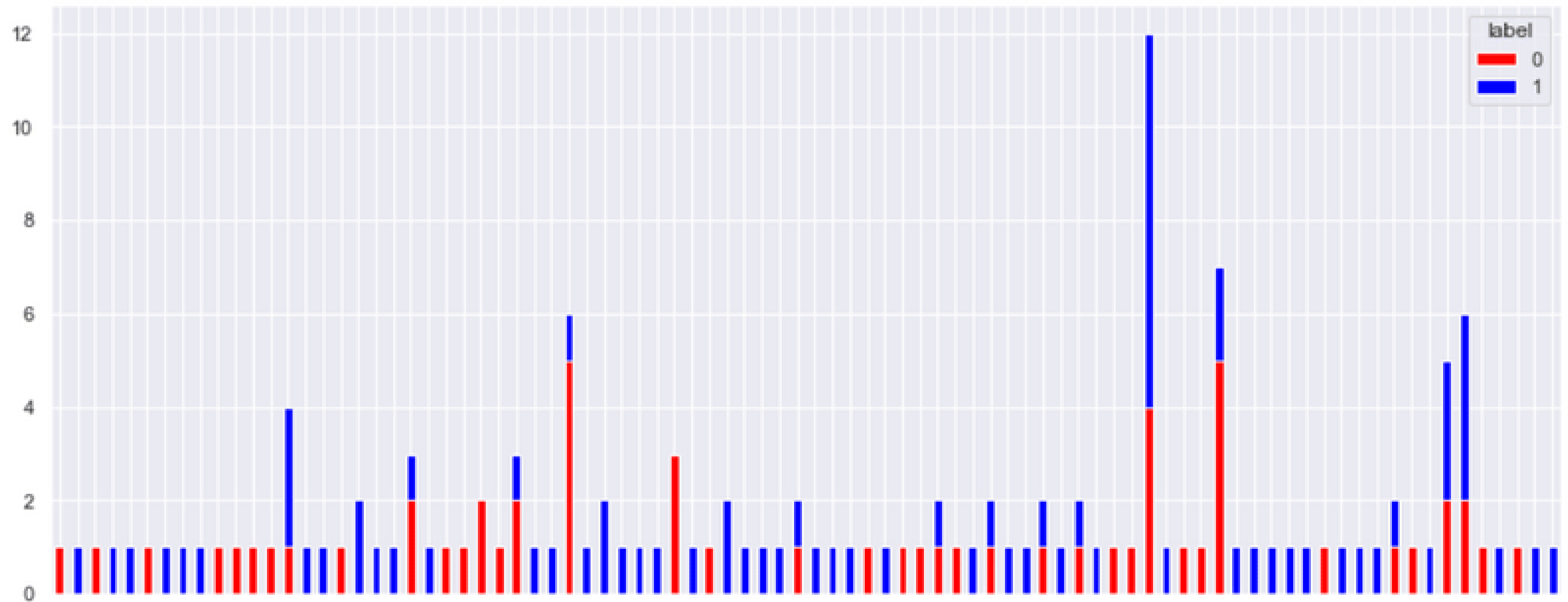
Problema

El objetivo que se está buscando con el análisis de Machine Learning, es determinar la elegibilidad de un paciente para ensayos clínicos de cáncer a partir del texto descriptivo, esto se debe a que se busca optimizar y acompañar el programa de selección de los pacientes, de forma que se ayude a determinar ciertos criterios que ayuden a definir la eficacia y seguridad de nuevos tratamientos para pacientes de estas características.

Por lo tanto, se plantea que este es un problema de aprendizaje supervisado y de clasificación. Esto se debe a que ya de por sí los datos están etiquetados, lo que permite hacer el entrenamiento más fácil y eficiente, y por el otro lado, es un problema de clasificación binaria, pues vamos a definir si un paciente es elegible o no



Comprensión y preparación de los datos



Modelado



Árbol de decisión

Aprendizaje supervisado para
tareas de clasificación



Ayuda a definir las
características más importantes
para redecir



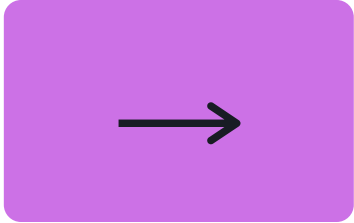
Calculo de hiperparámetros con
validación cruzada usando
GridSearchSV



MEJOR MODELO:

Criterio: 'gini'
máxima profundidad: 6
mínimo de particiones: 5



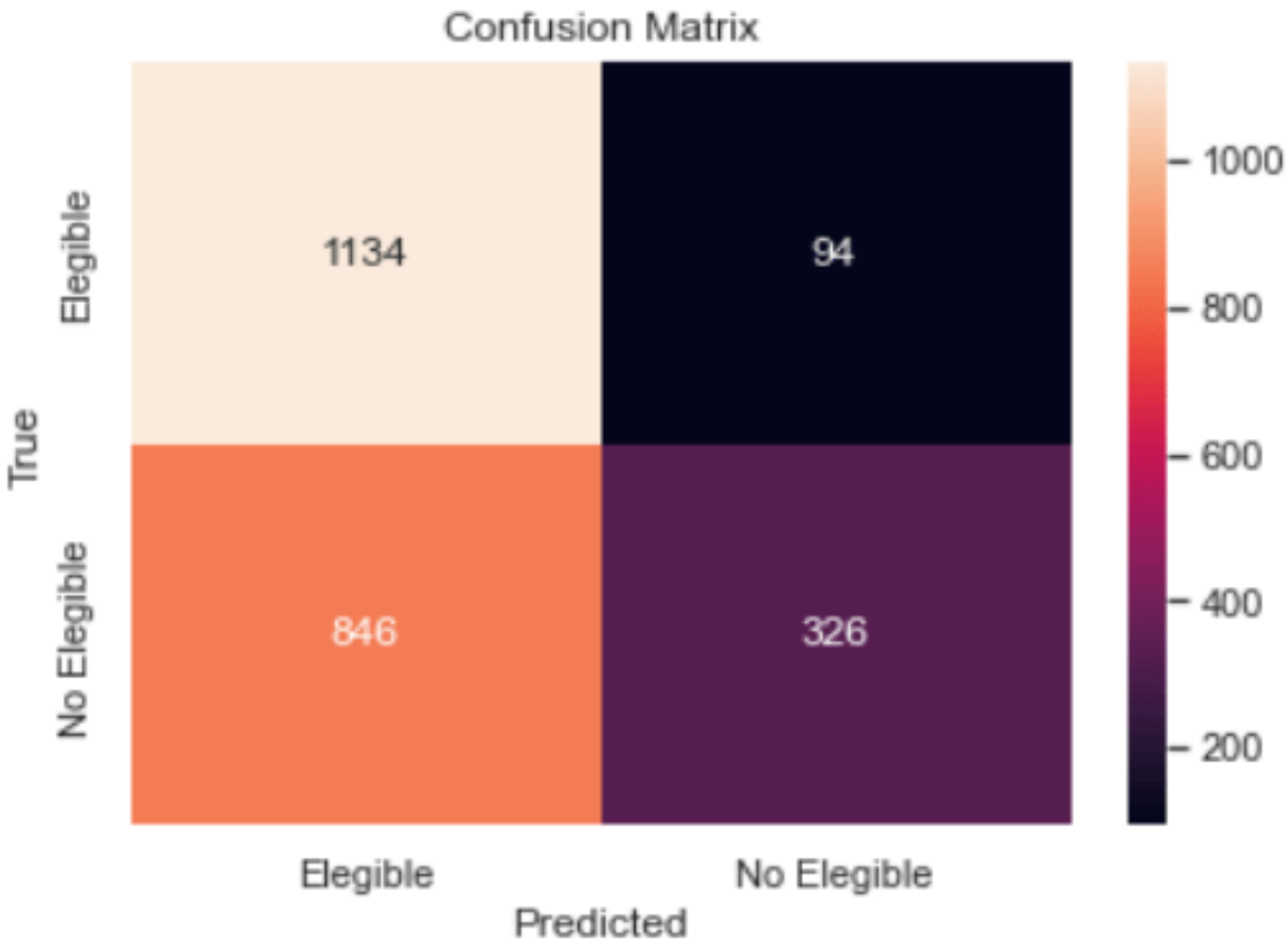


Resultados del modelo

Matriz para datos de PRUEBA

	precision	recall	f1-score	support
0	0.57	0.92	0.71	1228
1	0.78	0.28	0.41	1172
accuracy			0.61	2400
macro avg	0.67	0.60	0.56	2400
weighted avg	0.67	0.61	0.56	2400

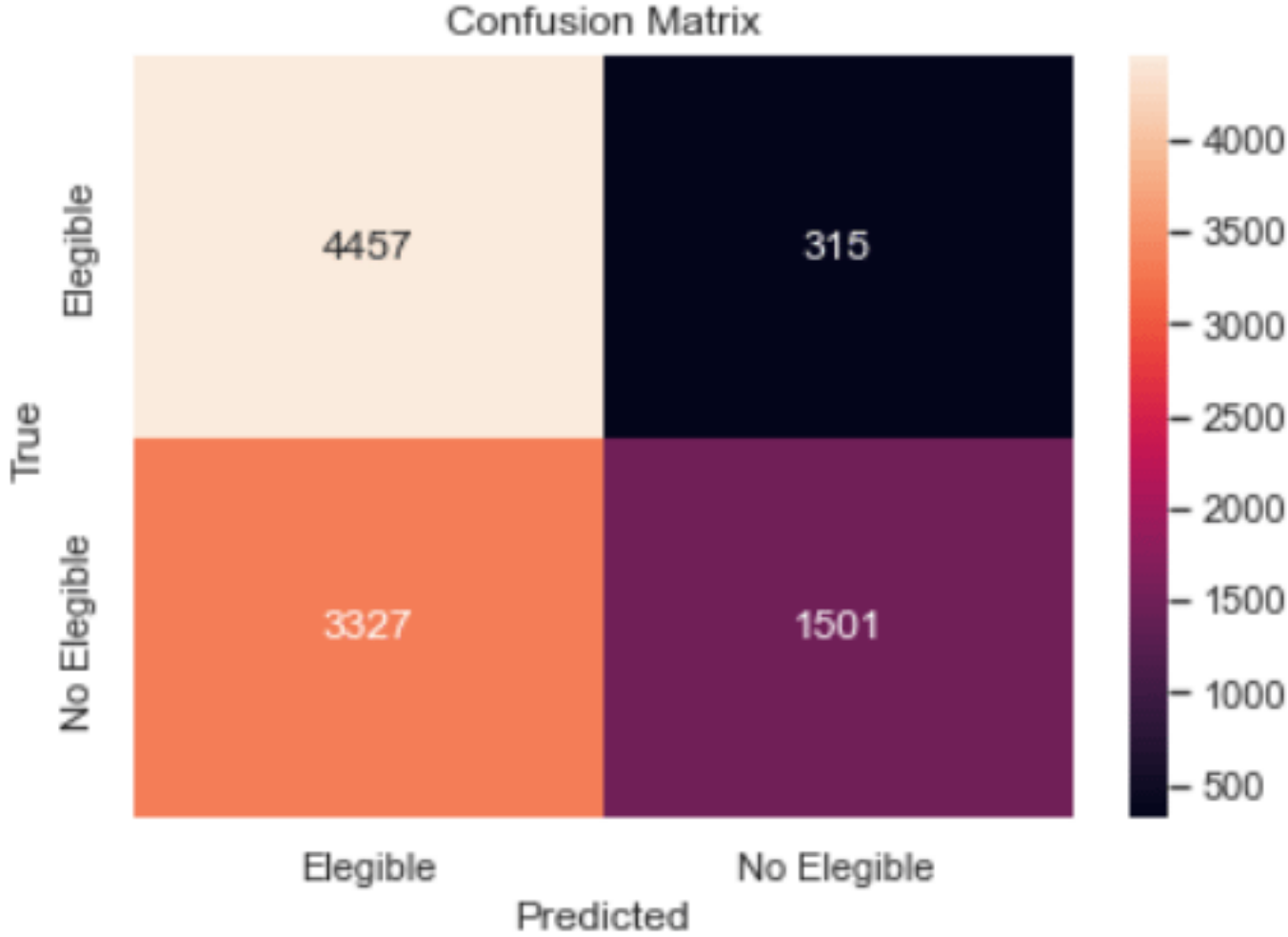
```
[[1134  94]
 [ 846 326]]
```



Matriz para datos de ENTRENAMIENTO

	precision	recall	f1-score	support
0	0.57	0.93	0.71	4772
1	0.83	0.31	0.45	4828
accuracy			0.62	9600
macro avg	0.70	0.62	0.58	9600
weighted avg	0.70	0.62	0.58	9600

```
[[4457 315]
 [3327 1501]]
```



Support Vector Machine

Aprendizaje supervisado para tareas de clasificación



Alta velocidad de entrenamiento y obtención de resultados



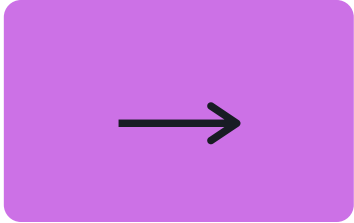
Un único hiperparametro a optimizar



MEJOR MODELO:

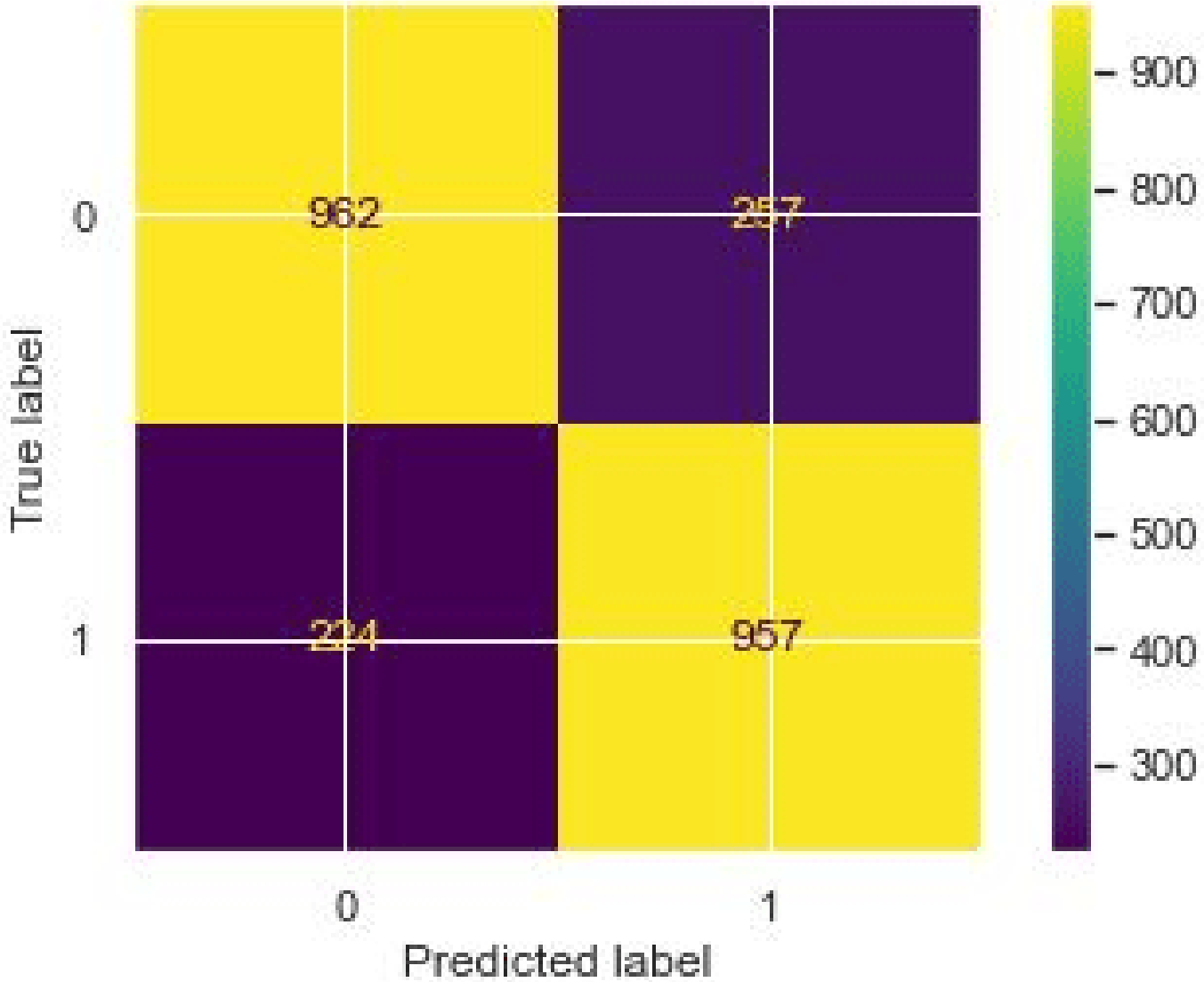
$C = 1$





Resultados del modelo

Resultados de predicciones con datos de prueba				
	precision	recall	f1-score	support
0	0.77	0.78	0.78	1219
1	0.77	0.76	0.77	1181
accuracy			0.77	2400
macro avg	0.77	0.77	0.77	2400
weighted avg	0.77	0.77	0.77	2400



K-Nearest Neighbors

Aprendizaje supervisado para tareas de clasificación



Estimar la probabilidad de que un elemento pertenezca a una clase



Calculo de hiperparámetros con validación cruzada usando GridSearchSV

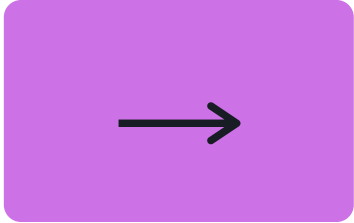


MEJOR MODELO:

k vecinos = 1

p = 1

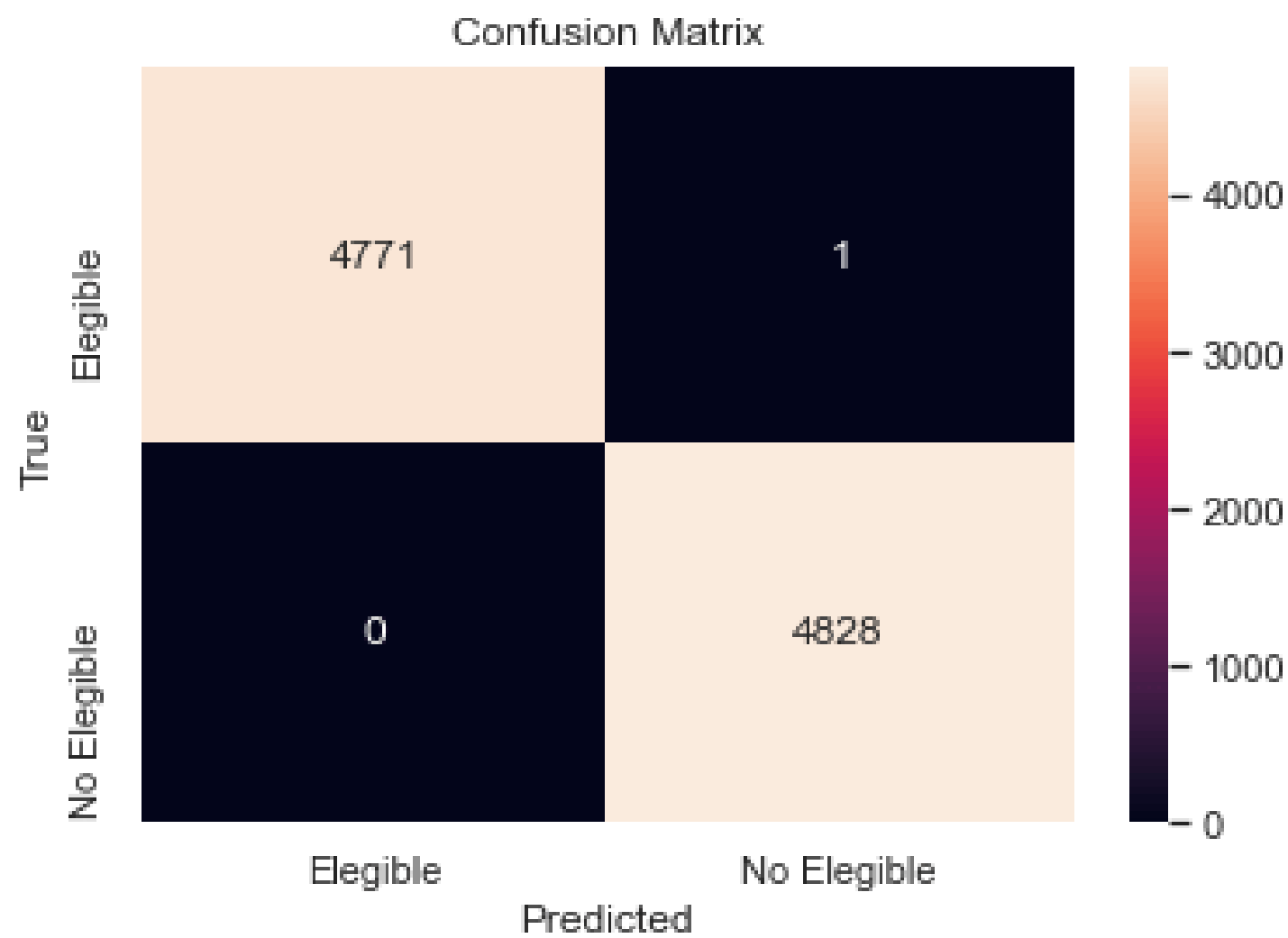




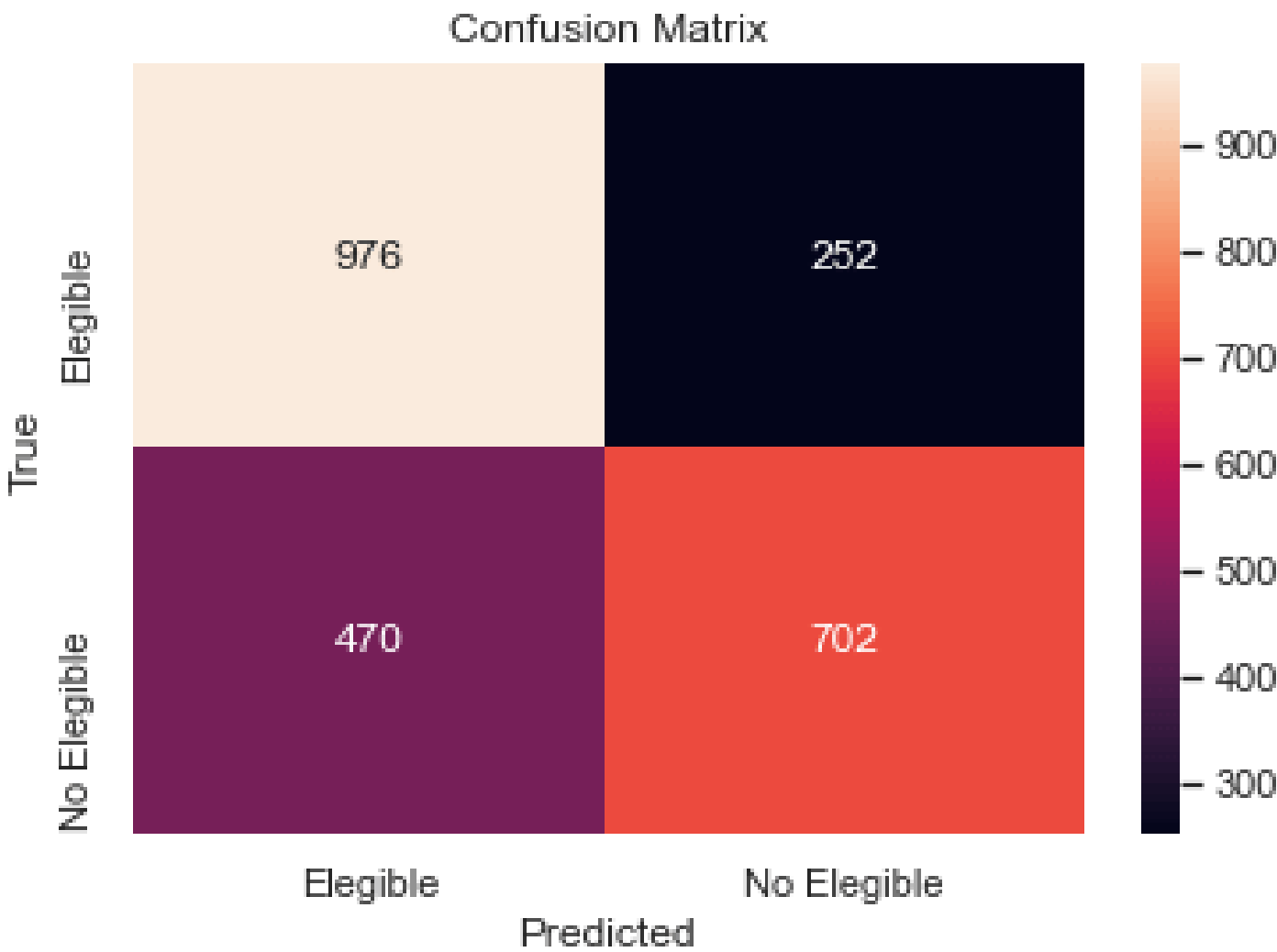
Resultados del modelo

Matriz para datos de PRUEBA

	precision	recall	f1-score	support
0	0.69	0.81	0.74	1228
1	0.75	0.61	0.67	1172
accuracy			0.71	2400
macro avg	0.72	0.71	0.71	2400
weighted avg	0.72	0.71	0.71	2400



Sobre entrenamiento



Regresión Logística

Aprendizaje supervisado para tareas de regresión



Fácil de implementar y se puede usar como base para problemas de clasificación binaria



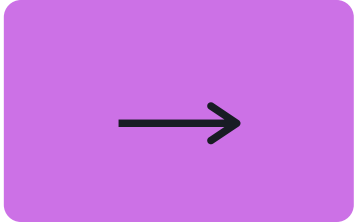
Proporciona una salida constante, lo que es de mayor utilidad para nuestro problema



MEJOR MODELO:

$C = 0.1$
penalización = 12
solver = liblincar



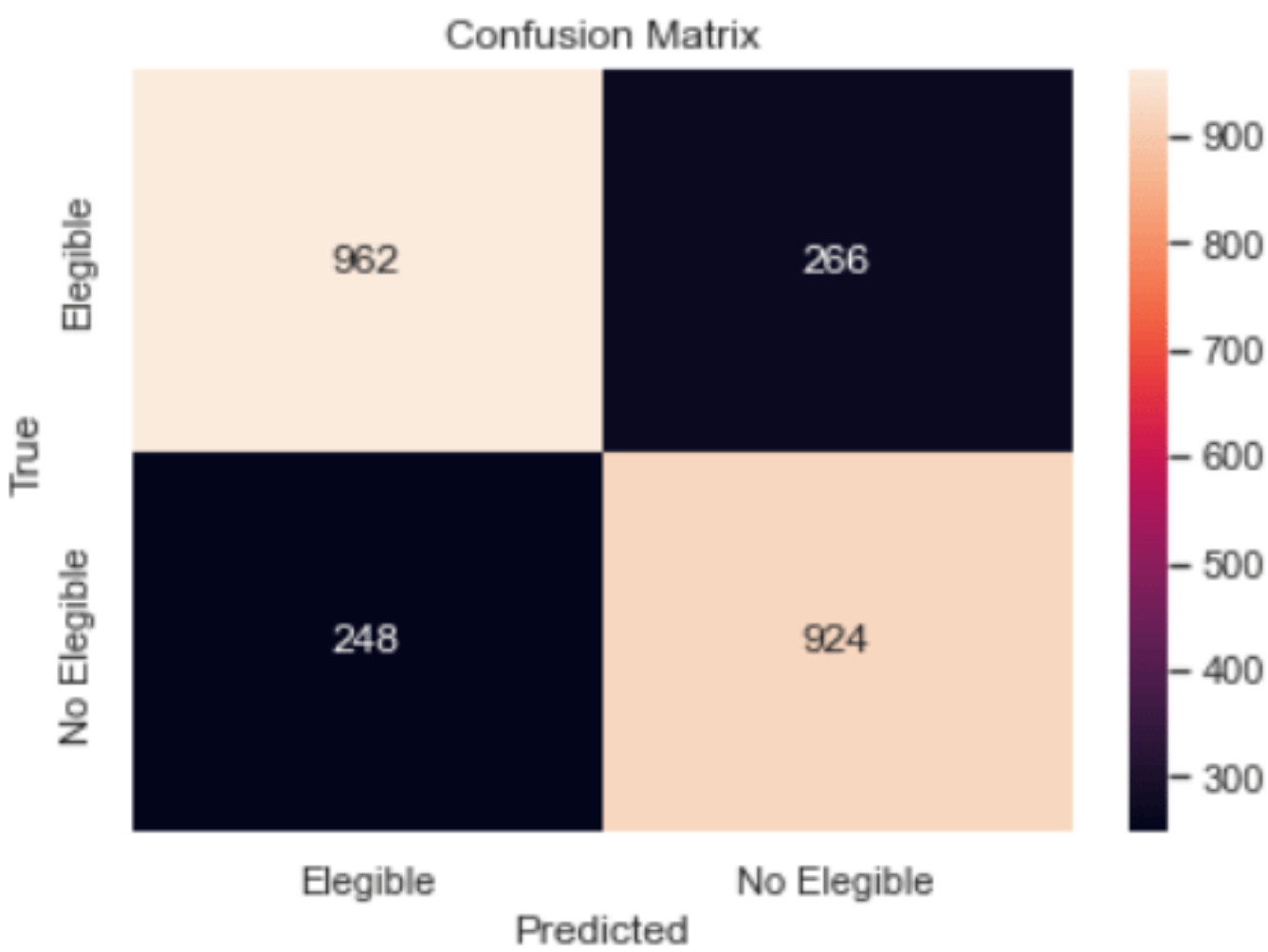


Resultados del modelo

Matriz para datos de PRUEBA

	precision	recall	f1-score	support
0	0.80	0.78	0.79	1228
1	0.78	0.79	0.78	1172
accuracy			0.79	2400
macro avg	0.79	0.79	0.79	2400
weighted avg	0.79	0.79	0.79	2400

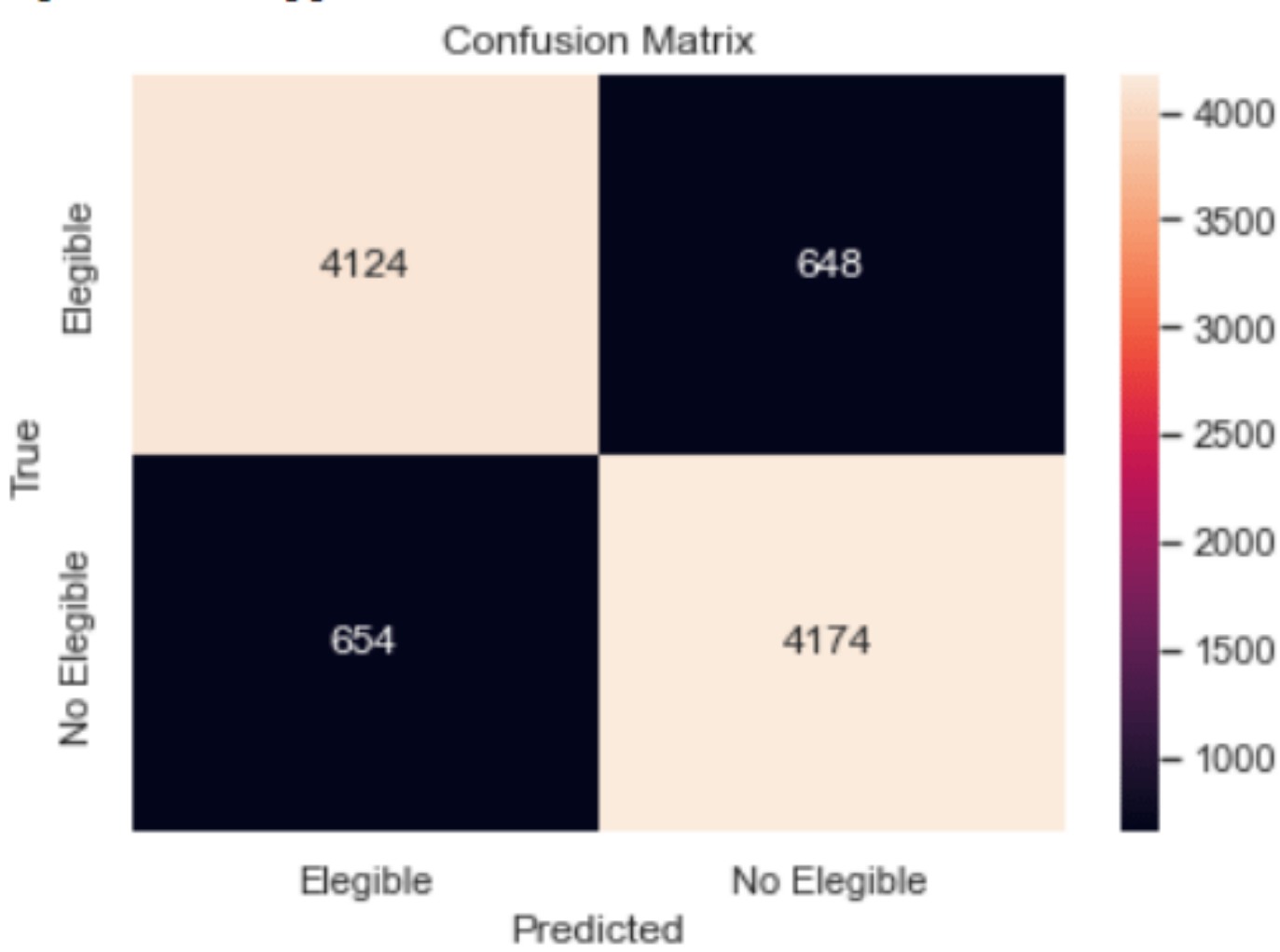
`[[962 266]
[248 924]]`



Matriz para datos de ENTRENAMIENTO

	precision	recall	f1-score	support
0	0.86	0.86	0.86	4772
1	0.87	0.86	0.87	4828
accuracy			0.86	9600
macro avg	0.86	0.86	0.86	9600
weighted avg	0.86	0.86	0.86	9600

`[[4124 648]
[654 4174]]`



Resultados



Conclusiones

El mejor modelo para para hacer la predicción de la elegibilidad de un paciente dado su diagnostico es la **Regresión Logística**

De llevarse a cabo una estandarización en la metodología de escribir los informes médicos, se podría incrementar la precisión y recall del modelo

El modelo representa una herramienta para los doctores, sin embargo, no pueden tomar una **desición final** sobre un paciente solamente con los resultados del modelo.



Gracias!!