

Conclusiones

- El modelo parece funcionar razonablemente bien, especialmente en la identificación de la clase "220+" (alto recall).
- La precisión en el conjunto de prueba es ligeramente menor que en el entrenamiento, lo que sugiere una posible sobreestimación de la precisión durante el entrenamiento o un pequeño ajuste del modelo para el conjunto de prueba.
- El F1 Score es consistente entre el entrenamiento y la prueba, lo que indica un equilibrio razonable entre precisión y recall en ambos conjuntos.