

**CARRERA**

|  |  |
| --- | --- |
|  | |
| NRC - ASIGNATURA: |  |
| PROFESORA: | Ing. Mayra Álvarez |
| PERÍODO ACADÉMICO: | 2024\_50 |
|  | |

**Informe Nº 1**

|  |  |
| --- | --- |
|  | |
| TÍTULO:  **Despliegue de modelo Machine Learning localmente** | |
| **ESTUDIANTE** | |
| Ferrín Josué  Pozo Steven | |
|  | |
|  | |
|  | |
|  | |
| FECHA DE ENTREGA: | 24 / 07/ 2024 |
| CALIFICACIÓN OBTENIDA: |  |
|  | |

1. **Alcance**

El propósito de este taller es acabar de mostrar la diferencia entre una regresión logística y una regresión lineal en base a datos normalizados y no normalizados. Para ello, se hará uso de un archivo CSV con el cual se podrá recopilar información para usarla como DataFrame o vectores, dependiendo las librerías que se usará para su análisis.

1. **Objetivos**

* Implementar datos normalizados y no normalizados en la regresión logística para evaluar su precisión.
* Implementar datos normalizados y no normalizados en la regresión lineal para evaluar su precisión.
* Comparar el mejor modelo para el análisis de los datos mediante su accurrancy score.

1. **Descripción**

En base a los datos de los sobrevivientes del Titanic, se estudian estos datos con regresión Lógica y Regresión Lineal, a continuación, se explica los dos tipos de regresiones

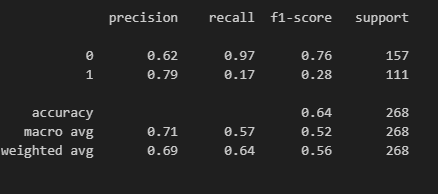
**Regresión Lógica**

En la regresión Lógico se necesita que los datos estén en forma binaria, las métricas de evaluación será la precisión de los datos y el acurracy, en base a la gráfica debe coincidir para la decisión del modelo

**Normalizado**

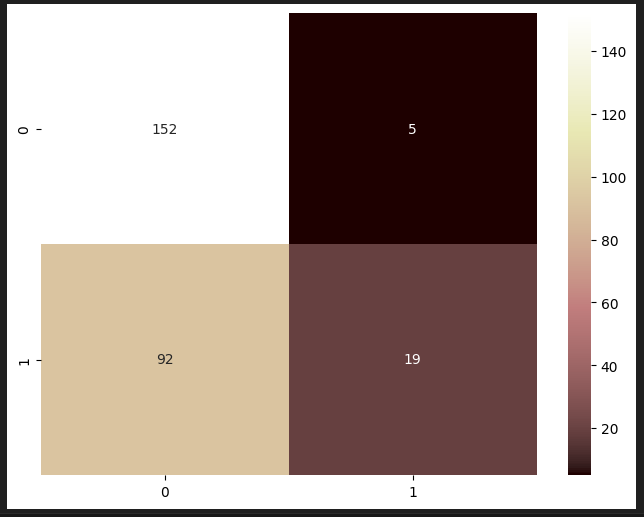
A continuación, se muestra el reporte de los datos con la precision y e accuracy, se sabe que cuando el acurracy es menor al 70% es que el modelo de prueba no es bueno. En la siguiente imagen nos muestra un accuracy 0.64

**Figura 1.** Score del modelo



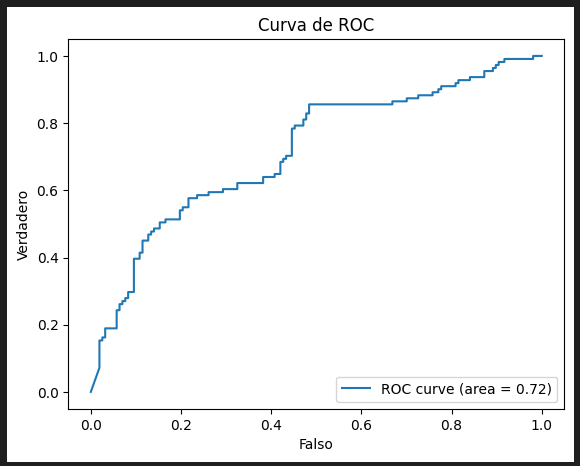
En base a la matriz de confusión que se muestra continuación, se observa que los errores son mayores altos.

**Figura 2.** Gráfica de matriz de confusión



Una de las representaciones visuales para evaluar el modelo es la curva de ROC, esta curva para validar si el modelo es bueno debe seguir con la curva de manera ascendentemente y en la imagen que se observa a continuación la curva muestra que el modelo va aprendiendo y decayendo.

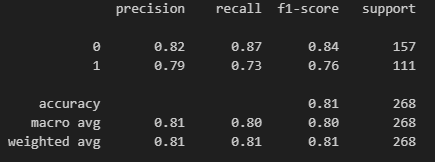
**Figura 3.** Curva de ROC



**No Normalizado**

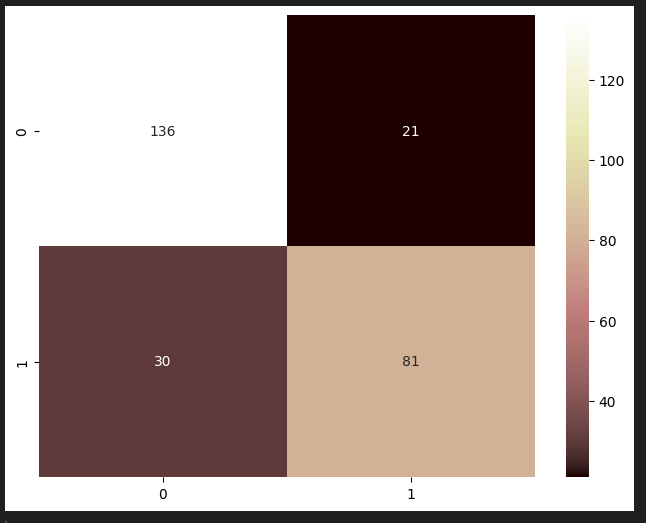
Para la representación de la precisión y del accuracy, normalizando los datos se muestra los siguientes datos, se observa que el modelo tiene un accuracy mayor a 70 con esto el modelo es bueno.

**Figura 4.** Score del modelo no normalizado



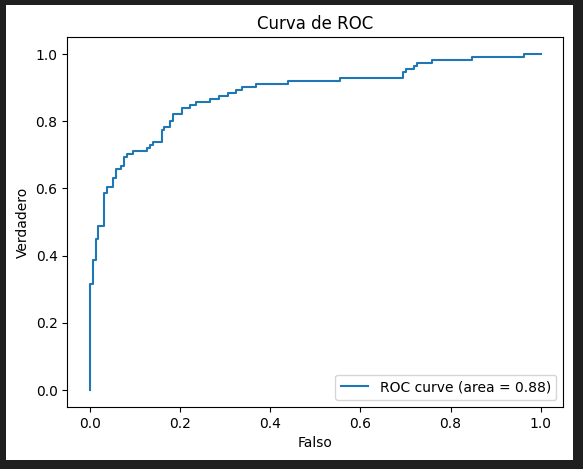
En la Matriz de Confusión se muestra los valores de error de la diagonal son menor haciendo referencia a que el análisis grafico coincide con el reporte

**Figura 5.** Gráfica de Matriz de confusión



A continuación, se muestra la gráfica con una curva de ROC, la cual muestra que el modelo es bueno ya que la curva va de manera ascendente

**Figura 6.** Curva de ROC no normalizado



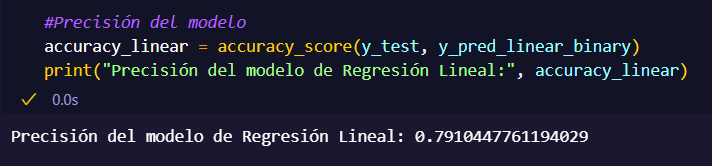
**Regresión Lineal**

**No Normalizado**

* **Precisión:**

La precisión de la regresión lineal es de 0.79104, esto quiere decir que tiene un 79.10% de precisión.

**Figura 7.** Precisión no normalizada



En la gráfica de dispersión de muestra los valores predichos vs los valores reales.

**Figura 8.** Gráfica de dispersión

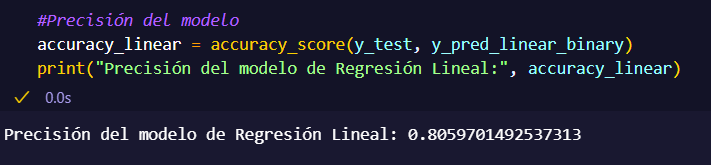
Gráfico, Gráfico de líneas

Descripción generada automáticamente

**Normalizado**

Al no normalizar los datos la precisión es mayor: es de 80,59%

**Figura 9**. Precisión normalizada



Los puntos al no estar cerca de la línea roja, muestra que la predicción no se acerca a los valores reales, al menos no del todo. Esto indica que el modelo de regresión lineal no es adecuado para representar datos binarios

**Figura 10.** Gráfica de dispersión normalizada

Gráfico, Gráfico de líneas

Descripción generada automáticamente

1. **Conclusiones**

En la regresión Logística el mejor modelo en base a los datos de los sobrevivientes en el titanic es el No normalizado porque el acurracy tiene un valor mayor y en los gráficos se observan las mismas predicciones y también porque este modelo es bueno para datos binarios, en este caso sobrevivientes y no.

La regresión lineal es al contrario que el logístico, cuando se normalizan los datos, su precisión sube a 80% y cuando no se lo normaliza es del 79%. No es una diferencia significativa, pero si tiene su pequeña diferencia.

En base a las dos regresiones la logística y la línea, el mejor modelo es el logístico no normalizado por el acurracy el cual es mayor a 70% y también por que el modelo logístico es mejor cuando tenemos valores binarios.