# ANÁLISIS Y REPORTE SOBRE EL DESEMPEÑO DEL MODELO

**1. INTRODUCCIÓN**

Durante el módulo de aprendizaje máquina se pudieron analizar diversos tipos de modelos para el análisis de comportamiento de algún sistema, en este caso se usó el set de datos del Titanic para el comportamiento de la supervivencia de una persona en el Titanic con relación a su edad y clase en la que se encontraba.

Para el momento de retroalimentación 1 (MomentoRetroalimentacion.py) se realizó la actividad con la implementación manual de una regresión logística; mientras que para el segundo momento de retroalimentación se realizó la implementación con uso de librerías, de manera que se pudieran comparar los modelos.

Con ambos casos implementados, se realizó una comparación del nivel de precisión del sistema, y para poder mejorar el desempeño del modelo se debieron de modificar los hiperparámetros del mismo.

**2. ANÁLISIS DEL MODELO**

Para realizar el análisis se usaron distintos indicadores como el ajuste, sesgo y precisión de los modelos con base en el tipo de implementación que implica cada uno.

Para la separación de los datos en una proporción 80 % de entrenamiento y 20% de prueba, así como los resultados preliminares.

|  |  |
| --- | --- |
| Momento de retroalimentación 1  (sin framework) | Momento de retroalimentación 2  (con framework) |
| En este caso la división de los datos se hace de manera manual:    Para hacer el entrenamiento de y la prueba del modelo se hizo uso de las funciones creadas por la clase RegresionLogistica creada.    Al correr esta primera prueba se obtuvo como resultado una precisión de: | En este caso se pudo usar la función de train\_test\_split de sklearn.    Hacer la prueba del mismo tipo de rergresión con la librería de sklearn se utilizaron los conjuntos obtenidos como se muestra:    En este caso se obtuvieron los datos correspondientes al ajuste del modelo con respecto a los datos de entrenamiento y de los datos de prueba. |

Como primer análisis podemos observar que el modelo propuesto por la librería utilizada tiene una mejor precisión, donde la diferencia es de 0.682 a 0.642.

Por otro lado, también se puede obtener el sesgo

|  |  |
| --- | --- |
| Momento de retroalimentación 1  (sin framework) | Momento de retroalimentación 2  (con framework) |
| El bias obtenido es de | En este caso los resultados son |

En este caso podemos determinar que tenemos un sesgo positivo, lo que indica que el modelo mide por encima del valor real.

También se puede observar como el nivel de exhaustividad es muy bajo, lo que indica que, si bien el modelo puede de representar el fenómeno, necesita de ser mejorado para poder ajustarse de mejor manera, o simplemente no es la alternativa más adecuada para este sistema.

**2. TÉCNICAS DE REGURALIZACIÓN Y AJUSTE DE PARÁMETROS**

Considerando de que se tiene una precisión baja para el modelo con ambas propuestas, es necesario el poder modificar los hiperparámetros de la regresión logística, para lo cual se deben de saber cuales son estos y por ende poder modificarlos.

Los hiperparámetros de el modelo de regresión logística son:

Texto

Descripción generada automáticamente

Con estos podemos de modificarlos para poder tener un mejor ajuste del modelo , podemos de obtener los mejores valores para los hiperparámetros a través de la búsqueda con un GridSearchCV y con eso poder modificar los adecuados de la mejor forma posible.

Lo anterior, nos da una modificación con los siguientes hiperparámetros y un resultado del desempeño del modelo:



De manera que se configuró el modelo con dichos parámetros:

Texto

Descripción generada automáticamente

Una vez que se tienen este modelo, podemos volver a comparar el rendimiento con los nuevos parámetros:

Imagen que contiene Texto

Descripción generada automáticamente

Con esto podemos de concluir que aún teniendo los hiperparámetros de mejor ajuste del modelo, el sistema no es eficiente para modelar el nivel de supervivencia de una persona en el Titanic, basándonos en su edad y clase.