

Al realizar la Actividad 7, comenzamos por realizar una limpieza de valores nulos y outliers, la cual fue muy similar a la Actividad 6, pero se realizaron diferentes tratamientos, se mantuvieron otras variables como marca y modelo, ya que en la entrega anterior no se ocupan una gran cantidad de variables categóricas, pero para poder realizar una regresión logística, necesitamos variables categóricas, específicamente con valores dicotómicos, o únicamente dos opciones como “1” o “0”; por ende, se remplazan los valores nulos o faltantes por la moda, para diferentes variables como: edad_cliente, marca, modelo, etc. Posteriormente al realizar esta limpieza, se procede a tratar los valores atípicos, por medio, de la desviación estándar y con esto normalizar nuestra distribución. Posteriormente se realiza una modificación para dos variables: “status” y “marca”, con lo cual logramos obtener más variables categóricas para poder realizar la regresión logística. Por último, se realiza la definición de las variables para cada uno de los modelos, tanto las variables independientes como las variables dependientes, y se realiza el escalamiento de cada variable correspondiente a los modelos en sus instancias de entrenamiento, seguido de esto se realiza el modelo de regresión logística; y por último, se realiza la matriz de confusión, sensibilidad, exactitud, precisión y el puntaje F1, utilizando sklearn en sus respectivas librerías, y finalmente se hizo una comparación entre los resultados de los valores anteriormente utilizados.

	Precision	Exactitud	Sensibilidad
Modelo 1	0.000000	0.928749	0.000000
Modelo 2	0.395437	0.755021	0.064356
Modelo 3	0.000000	0.965108	0.000000
Modelo 4	0.773888	0.751210	0.949139
Modelo 5	0.368613	0.750330	0.061925
Modelo 6	0.000000	0.766603	0.000000
Modelo 7	0.661218	0.630553	0.670051
Modelo 8	0.631405	0.630406	0.770412
Modelo 9	0.000000	0.964375	0.000000
Modelo 10	0.414286	0.760006	0.054341

El mejor resultado el obtenido el modelo 4, el cual tiene las siguientes variables independientes: “pagos_realizados”, “riesgo” y “limite_credito”; y la variable dependiente se define como “status”. Proporcionando los mejores valores para precision: 0.774, exactitud: 0.750 y sensibilidad: 0.949. Por ende, es el mejor resultado en comparación a todos basado en los valores.