



# Tecnológico de Monterrey

**Instituto Tecnológico y de Estudios Superiores de Monterrey**

**Analítica de datos y herramientas de inteligencia artificial I (Gpo  
101)**

**Actividad 7.2 (Matriz de Confusión)**

Amairany Rodríguez Huerta | A01702927

Luis Pablo Padilla Barbosa | A00572040

Renata Emilia Chávez Martínez | A01351716

5 may 2023

## Matriz de Confusión

Primeramente realizamos una limpieza así como la eliminación de outliers de toda la base de datos para que de esta manera podamos tener bases de datos llenas de información. Y realizamos un modelo de regresión logística en el cual tenemos como variable independiente el TIPO de GASTO y de variable dependiente el TOTAL de gasto que se tuvieron en los 4 años que se tienen registrados por parte de la empresa.

Posteriormente realizamos una matriz de confusión, una matriz de confusión es una herramienta utilizada en el análisis de clasificación de datos para evaluar el rendimiento de un modelo de clasificación. Es una tabla que muestra el número de predicciones correctas e incorrectas hechas por un modelo en comparación con las observaciones reales. En este caso realizamos

La matriz de confusión tiene cuatro elementos principales: verdaderos positivos (TP), falsos positivos (FP), verdaderos negativos (TN) y falsos negativos (FN). Los verdaderos positivos son los casos en los que el modelo predice correctamente la clase positiva, los falsos positivos son los casos en los que el modelo predice incorrectamente la clase positiva, los verdaderos negativos son los casos en los que el modelo predice correctamente la clase negativa y los falsos negativos son los casos en los que el modelo predice incorrectamente la clase negativa.

La matriz de confusión se utiliza para calcular métricas de evaluación de modelos de clasificación, como la precisión, la sensibilidad, la especificidad y la exactitud. La precisión mide la proporción de predicciones correctas que hace el modelo, la sensibilidad mide la proporción de verdaderos positivos respecto al total de casos positivos, la especificidad mide la proporción de verdaderos negativos respecto al total de casos negativos y la exactitud mide la proporción de predicciones correctas en general.

|      | Precisión | Exactitud | Sensibilidad |
|------|-----------|-----------|--------------|
| Años |           |           |              |
| 2020 | 0.598039  | 0.571286  | 0.135556     |
| 2021 | 0.603922  | 0.542261  | 0.327660     |
| 2022 | 0.583333  | 0.532300  | 0.130319     |
| 2023 | 0.666667  | 0.566667  | 0.175439     |

Viendo los resultados del modelo se analiza cómo el modelo tiene un rango de precisión de 50% a 60% que son las variables que nos mide correctas en el modelo, esto dándonos a entender que el modelo es bueno prediciendo los valores positivos en el modelo, y de misma manera en exactitud se tiene un rango de 50% a 60%. Este porcentaje de exactitud nos habla sobre el sesgo de la estimación en relación a los valores verdaderos. Por el otro lado la sensibilidad del modelo logramos ver que esta en un 10%; siendo algo totalmente negativo para el modelo pues nos indica que la proporción de casos positivos que fueron correctamente identificadas por el algoritmo fue baja.