

## Informe

**Nombres:** Andrés Felipe Gómez Pacheco

- Jose Manuel Giraldo Pérez

### 1. Introducción

En este informe, se comparan dos modelos de aprendizaje automático utilizados para predecir la variable de interés en una base de datos que contiene información relacionada con el comportamiento de los consumidores. La base de datos incluye columnas como destination, passenger, weather, temperature, coupon, entre otras, y la variable objetivo es Y, que representa si el usuario ha utilizado un cupón o no.

El objetivo es elegir el modelo más adecuado para predecir la probabilidad de que un usuario utilice un cupón en función de las características proporcionadas.

### 2. Modelo de Regresión Logística

#### 2.1 Justificación del Modelo

La **regresión logística** es adecuada cuando la variable dependiente (en este caso, Y) es binaria, es decir, tiene dos posibles resultados (por ejemplo, 0 para no utilizar el cupón, y 1 para utilizarlo). Este modelo es capaz de predecir probabilidades y clasificar las observaciones en dos categorías (binarias), lo que lo convierte en una opción natural para problemas de clasificación binaria.

#### 2.2 Resultados de la Regresión Logística

Los resultados de la regresión logística muestran las métricas de **precisión** obtenidas en los conjuntos de entrenamiento y prueba. A continuación se presentan las métricas obtenidas:

- **Training Accuracy:** 58.31%
- **Testing Accuracy:** 58.57%

Aunque la precisión no es extremadamente alta, se puede ver que el modelo es razonablemente consistente en el conjunto de entrenamiento y prueba, lo que sugiere que la regresión logística ha aprendido bien la relación entre las características y la variable objetivo.

#### 2.3 Conclusión sobre el Modelo de Regresión Logística

La regresión logística es adecuada para problemas de clasificación binaria, como el presente. Aunque la precisión es moderada, el modelo es relativamente bueno para

prever el uso de cupones basándose en las características del usuario. La pequeña diferencia entre la precisión en el conjunto de entrenamiento y el conjunto de prueba sugiere que no hay sobreajuste significativo.

### 3. Modelo de Regresión Lineal

#### 3.1 Justificación del Modelo

La **regresión lineal** es utilizada comúnmente para predecir valores continuos. A pesar de no ser la opción más adecuada, la regresión lineal puede ser útil para comparar el rendimiento de modelos simples con otros más complejos como la regresión logística.

#### 3.2 Resultados de la Regresión Lineal

A continuación, se presentan las métricas obtenidas con el modelo de regresión lineal en el conjunto de prueba y el conjunto de entrenamiento:

- **Test Set Evaluation:**
  - **MAE (Mean Absolute Error):** 0.49
  - **MSE (Mean Squared Error):** 0.24
  - **RMSE (Root Mean Squared Error):** 0.49
  - **R2 (R-squared):** 0.01
- **Train Set Evaluation:**
  - **MAE:** 0.49
  - **MSE:** 0.24
  - **RMSE:** 0.49
  - **R2:** 0.00

Los resultados del modelo de regresión lineal muestran un **R2** muy bajo (cercano a 0), lo que indica que el modelo no explica la variabilidad de la variable Y de manera significativa. Además, el MAE, MSE y RMSE son relativamente altos, lo que sugiere que el modelo no es preciso.

#### 3.3 Conclusión sobre el Modelo de Regresión Lineal

El modelo de regresión lineal no es adecuado para este problema de clasificación binaria, ya que no logra predecir bien el comportamiento de la variable objetivo. El bajo valor de  $R^2$  indica que el modelo tiene un rendimiento deficiente en comparación con la regresión logística.

#### 4. Comparación de los Modelos

- **Precisión:** El modelo de **regresión logística** muestra una precisión significativamente mejor que la regresión lineal en ambas métricas de entrenamiento y prueba. La regresión logística tiene un **58.31% de precisión en entrenamiento** y **58.57% en prueba**, mientras que la regresión lineal muestra una **precisión mucho más baja** en términos de errores cuadráticos.
- **Métricas de error:** El modelo de **regresión lineal** muestra un  **$R^2$**  muy bajo, lo que indica que el modelo no es adecuado para este tipo de tarea. La **regresión logística**, aunque no alcanza una precisión muy alta, es claramente más adecuada debido a su naturaleza de clasificación binaria y su capacidad para modelar probabilidades.

#### 5. Conclusiones

Con base en los resultados obtenidos, se concluye que el **modelo de regresión logística** es más adecuado para este problema de clasificación binaria. Aunque ambos modelos tienen limitaciones, la regresión logística ofrece un mejor rendimiento en términos de precisión y es la opción preferida para este tipo de predicciones.