



Tecnológico de Monterrey

Campus Puebla

José Antonio Villanueva Arenas A01737550

Pablo García von Westarp A01737407

Diego Matias Rossano A01737509

José Manuel Flores Pérez A01733838

Analítica de datos y herramientas de inteligencia artificial II

Actividad 4.2 (Regresión Logística)

23 Abril 2025

Contexto General

Utilizamos la regresión logística para analizar la relación entre las variables dicotómicas, evaluando el comportamiento de los usuarios con diferentes variables . Después, convertimos las variables categóricas relevantes en variables binarias para que pudieran servir para el modelo. Evaluamos las combinaciones de variables utilizando tres métricas principales. La precisión que mide el porcentaje de predicciones positivas correctas, la exactitud para poder ver el porcentaje total de predicciones correctas y por último la sensibilidad que nos mostró la capacidad para detectar correctamente los casos positivos reales.

Variables Dependientes	Variables independientes	Categorías analizadas .	Precisión	Exactitud (igual para los dos)	Sensibilidad	Interpretación
botón correcto	tiempo de interacción, número de interacción, tiempo de lección	Correcto, incorrecto.	0.59, 0.79	0.69	0.80, 0.62	Buen equilibrio en general, mejor sensibilidad cuando es correcto.
juego	usuario, dificultad, mini juego	astros, cadetes	0.73, 0.67	0.71	0.88, 0.42	Se predice mejor la categoría astros que cadetes.
auto push	Juego, usuario, tiempo de interacción	Correcto, incorrecto.	0, 0.99	0.99	0.00, 0.99	El modelo sólo predice correctamente lo incorrecto.
Rojo (color)	tiempo de interacción, número de interacción, tiempo de lección	rojo, otros colores	0, 0.99	0.99	0, 0.99	Alta exactitud (0.99).
Amarillo (color)	tiempo de interacción, número de interacción,	amarillo, otros colores	0, 0.78	0.78	0, 1	Es el color con mejor desempeño, obtuvo una sensibilidad de 1

	tiempo de lección					y una precisión de 0.78, lo que indica que sí detectó algunos positivos
Verde (color)	tiempo de interacción, número de interacción, tiempo de lección	verde, otros colores	0, 0.78	0.77	0.00, 1	Se predicen correctamente los otros colores, pero no verde.
Violeta (color)	tiempo de interacción, número de interacción, tiempo de lección	violeta, otros colores	1, 0.78	0.81	0.12, 1	Algo mejor predicción de violeta que otros colores.
Azul (color)	tiempo de interacción, número de interacción, tiempo de lección	azul, otros colores	0, 0.79	0.79	0.00, 1	Predice bien otros colores, más no el azul.
Dificultad	Juego, tiempo de sesión, usuario	Dificultad 1, otras dificultades	1, 0.61	0.58	0.001, 0.99	Alta precisión para dificultad 1, pobre para otras dificultades.
número de interacción	tiempo de interacción, botón correcto, color presionado	Menor de 128s, Mayor de 128.5s	0.98, 0.92	0.99	0.99, 0.4	Excelente predicción para interacción corta, menor para larga.
Asteroides (minijuego)	tiempo de sesión, botón correcto, color presionado	asteroides, otros mini juegos	0, 0.80	0.81	0.00, 1	Solo se predice correctamente los otros mini juegos.
Restaurante (minijuego)	tiempo de sesión, botón correcto, color presionado	restaurante, otros mini juegos	0, 0.85	0.85	0.00, 1	El modelo predice correctamente "otros juegos" pero no restaurante.

Leonardo (usuario)	Administrador, Juego, auto push	Leonardo, otros usuarios	0, 0.85	0.84	0.00, 1	El modelo tiene buenas predicciones
Nicolas (usuario)	Administrador, Juego, auto push	Nicolas, otros usuarios	0, 0.90	0.9	0.00, 1	No se logra predecir bien a Nicolás, pero sí a otros usuarios.
Ma Del Rosario (usuario)	Administrador, Juego, auto push	Ma del Rodario, otros usuarios	0, 0.98	0.98	0.00, 1	Predice muy bien el modelo, no reconoce casos positivos
Rene (usuario)	Administrador, Juego, auto push	Rene, otros usuarios	0, 0.99	0.99	0.00, 1	El modelo no captura los casos positivos, pero acierta muy bien
Sergio Angel (usuario)	Administrador, Juego, auto push	Sergio Angel, otros usuarios	0, 0.95	0.95	0.0, 1	Alta exactitud, predice muy bien "otros usuarios".

Comparaciones

Existe una exactitud bastante buena ya que en casi todos los modelos se rebasa el 70% de exactitud, lamentablemente la precisión es muy baja lo que nos dice que el modelo tiene dificultad del modelo para identificar correctamente los positivos. Al igual que la precisión, la sensibilidad de los modelos es extremadamente baja en muchos casos, lo que nos dice que los modelos no detectan bien los verdaderos positivos. Claramente las mejores combinaciones tomando como referencia la sensibilidad fueron las relacionadas a tiempo de interacción y número de interacción. Por otro lado, el desempeño más débil fue en los modelos asociados a usuarios específicos y colores.

Conclusión

Aunque la exactitud de los modelos fue bastante alta, la baja precisión y la baja sensibilidad en la mayoría de los análisis nos dejan ver que los modelos no identifican correctamente los casos positivos. Esto nos dice que hay un desequilibrio de clases en los datos y las variables binarias no

logran capturar de buena manera las diferencias entre categorías. Para mejorar el modelo necesitamos una base de datos más amplia que tenga las clases equilibradas y con las categorías más diferenciadas.