



Tecnológico de Monterrey

Instituto Tecnológico y de Estudios Superiores de Monterrey Campus Puebla

Actividad 4.2 (Regresión Logística)

Equipo 5:

José Manuel Morales Escalante

Josué Ibáñez Santana

Sergio Giles Bracamontes

Ismael Martinez Duran

Materia:

Analítica de datos y herramientas de inteligencia artificial II

Fecha:

27 de abril de 2025

Análisis de Correlaciones y Modelos de Regresión Logística

El presente reporte expone los resultados del análisis de correlaciones y de los modelos de regresión logística desarrollados, cuyo objetivo fue predecir la probabilidad de pertenencia a categorías específicas en distintos escenarios evaluados. A diferencia de la regresión lineal tradicional, la regresión logística estima probabilidades acotadas entre 0 y 1, permitiendo modelar relaciones entre variables predictoras y un resultado categórico.

El análisis de correlaciones reveló que las variables de naturaleza numérica presentaron un mayor grado de significancia frente a las variables convertidas en categorías. No obstante, variables categorizadas como "color_cat y minijuego_cat" mostraron un impacto relevante dentro de ciertos modelos específicos.

En cuanto al modelado de la variable "Botón Correcto - Juego - Auto Push", se evaluaron tres modelos. El primer modelo mostró un desempeño moderado con una precisión del 61.8%, exactitud del 63.3% y una sensibilidad del 73.4%, lo cual indica una capacidad razonable para identificar resultados positivos, aunque con un nivel apreciable de falsos positivos. El segundo modelo mejoró estas métricas, alcanzando una precisión del 65.6% y una sensibilidad perfecta, aunque con la posibilidad de presentar falsos positivos. El tercer modelo presentó resultados sobresalientes, registrando una precisión del 99% y una exactitud del 99.1%, aunque podría estar influenciado por un desbalance de clases.

Respecto a la variable "Color Presionado", se observó que los modelos asociados a los colores rojo, azul, violeta y amarillo no lograron identificar correctamente resultados positivos, evidenciando precisión y sensibilidad nulas, a pesar de presentar altos niveles de exactitud, atribuibles a una predicción sesgada hacia la clase negativa. El modelo relacionado con el color verde destaca ligeramente con una precisión del 87.5%, aunque con una sensibilidad limitada de apenas 12.7%.

En el análisis de "Dificultad - Mini Juego - Número de Interacción", el primer modelo presentó una precisión aceptable del 66%, pero una sensibilidad baja de 15%, indicando una limitada capacidad de detección de casos positivos. Modelos aplicados a juegos específicos como "Asteroideos" y "Restaurante" demostraron deficiencias similares, a excepción del juego "Restaurante", cuyo modelo alcanzó un desempeño sobresaliente, con precisión y sensibilidad del 98.3% y 100% respectivamente.

El análisis por usuario evidenció resultados mixtos. La mayoría de los modelos no detectaron de manera adecuada los verdaderos positivos. Sin embargo, el modelo relacionado con el usuario "Kytzia" se destacó con una precisión del 85% y una sensibilidad del 82.1%, lo que refleja una capacidad sobresaliente de clasificación en comparación con otros usuarios.

	Valor	Precisión Pos Label 1	Sensibilidad Pos Label 1	Exactitud	Precisión Pos Label 2	Sensibilidad Pos Label 2
0	botón correcto	0.608258	0.805333	0.650555	0.728625	0.501921
1	juego	0.642064	1	0.642064	0	0
2	autopush	0.989549	1	0.989549	0	0
3	color rojo	0	0	0.996734	0.996734	1
4	color azul	0	0	0.758328	0.758328	1
5	coloe verde	0	0	0.759634	0.759634	1
6	color violeta	0.954545	0.155556	0.774657	0.766553	0.997336
7	color amarillo	0	0	0.774004	0.774004	1
8	dificultad	0.742188	0.149371	0.625082	0.614398	0.963128
9	minijuego asteroides	0	0	0.813847	0.813847	1
10	minijuego restaurante	0	0	0.85565	0.85565	1
11	numero de interacciones	0.977139	1	0.977139	0	0
12	usuario jose ian	0	0	0.984977	0.984977	1
13	usuario jose ignacio tadeo	0	0	0.984324	0.984324	1

14	usuario jose javier	0	0	0.928152	0.928152	1
15	usuario joshua	0	0	0.988896	0.988896	1
16	usuario kytzia	0.791667	0.791667	0.993468	0.996682	0.996682

Conclusión

A través de este análisis de regresión logística, logramos evaluar la capacidad predictiva de diversos modelos para estimar la probabilidad de pertenencia a categorías específicas en los escenarios planteados. Este enfoque nos permitió identificar qué variables, tanto numéricas como categóricas ejercieron una influencia significativa en los resultados, destacando casos como el modelo del "Botón Correcto - Juego - Auto Push", que alcanzó una precisión del 99%, y el del usuario "Kytzia", con un 85% de precisión y 82.1% de sensibilidad.

Si bien algunos modelos presentaron limitaciones, como baja sensibilidad o incapacidad para detectar verdaderos positivos en ciertas categorías (como los colores rojo, azul, violeta y amarillo), el análisis nos brindó información valiosa para comprender las relaciones entre las variables predictoras y los resultados. Además, nos permitió reconocer posibles sesgos, como el desbalance de clases en ciertos modelos, y áreas donde futuros ajustes podrían mejorar la capacidad predictiva.

En conclusión, este estudio nos sirvió para cuantificar el impacto de distintos factores en los comportamientos analizados, proporcionando una base estadística sólida que puede ser utilizada para la toma de decisiones, el refinamiento de estrategias o el desarrollo de modelos más robustos en futuras investigaciones. Los resultados obtenidos no solo validan la utilidad de la regresión logística en este contexto, sino que también señalan oportunidades de optimización para escenarios con menor desempeño predictivo.