



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 3.2 regresión no lineal

11 Abril 2025

Introducción

El presente reporte se centra en el proyecto con el socio formador WUUPI, una herramienta interactiva utilizada por usuarios bajo la supervisión de administradores, el objetivo principal fue comprender el comportamiento de los usuarios dentro de la plataforma, utilizando la variable "tiempo de interacción" como punto de partida, ya que mostró las correlaciones más altas con otras variables del conjunto de datos. Para ello, se aplicaron modelos de regresión no lineal, con el fin de identificar patrones complejos que no pueden ser explicados mediante correlaciones lineales simples.

Metodología:

Se inició el análisis identificando las correlaciones más relevantes entre variables, encontrando valores particularmente altos entre el tiempo de interacción y variables como botón correcto (0.90), número de interacción (0.80), y color presionado (0.90). A partir de esto, se aplicaron diversos modelos no lineales, como funciones cuadráticas, logarítmicas, de valor absoluto, tangenciales y polinomiales inversas, para observar cómo mejoraba el ajuste de los datos frente a la relación tradicional lineal.

Además del análisis general, se exploró el comportamiento por usuario individual (ej. Leonardo, María del Rosario, Nicolás, René y Sergio Ángel), ajustando modelos personalizados para cada uno. Por ejemplo, para María del Rosario, el modelo logarítmico aplicado al "botón correcto" alcanzó una correlación no lineal de 0.99, superior a la correlación lineal de 0.98. De manera similar, Nicolás mostró un ajuste excelente con una función cuadrática inversa aplicada al número de interacciones; también se elaboraron heatmaps comparativos para visualizar rápidamente los niveles de correlación entre variables, facilitando la detección de relaciones fuertes o áreas de mejora.

Resultados

El uso de modelos no lineales mejoró en algunos niveles de correlación en casi todos los casos analizados, destacando el modelo cuadrático como el más aplicado, especialmente en los análisis de usuarios como Leonardo y Nicolás. Los mejores ajustes se lograron con la función logarítmica en el caso de María del Rosario ($R \approx 0.99$) y con la función cuadrática inversa en los casos de Sergio Ángel y Nicolás ($R \approx 0.99$).

Tabla de comparación por usuario:

Leonardo:

variable dependiente	variable independiente	modelo utilizado	Correlación simple	Correlación del modelo no lineal
tiempo de interacción	botón correcto	función cuadrática	0.73	0.76
botón correcto	tiempo de interacción	función cuadrática	0.73	0.99
color presionado	tiempo de interacción	función cuadrática	0.72	0.86
dificultad	mini juego	función cuadrática	0.96	0.98

Para superar las correlaciones simples de Leonardo, se utilizó la función cuadrática para todas las variables objetivos, ya que coincidió ser el mejor modelo para cada una de ellas; obteniendo 0.76 para 'tiempo de interacción', 0.99 para 'botón correcto', 0.86 para 'color presionado' y 0.98 para 'dificultad'.

María de los Ángeles:

variable dependiente	variable independiente	modelo utilizado	Correlación simple	Correlación del modelo no lineal
tiempo de interacción	botón correcto	función cuadrática	0.98	0.99
botón correcto	tiempo de interacción	función cuadrática	0.98	0.99
color presionado	tiempo de interacción	función cuadrática	0.98	0.99
dificultad	tiempo de lección	función tangencial	0.10	0.14

María del Rosario, al tener correlaciones altas, se buscó “aproximar” las correlaciones para cada variable objetivo en el modelo no lineal, sin embargo, utilizando la función cuadrática y tangencial se lograron superar las correlaciones con valores de 0.99 para ‘tiempo de interacción’, ‘botón correcto’ y ‘color presionado’, mientras que la dificultad se superó con 0.14 ya que la correlación simple era muy baja inicialmente.

Nicolás:

variable dependiente	variable independiente	modelo utilizado	Correlación simple	Correlación del modelo no lineal
tiempo de interacción	botón correcto	función cuadrática	0.93	0.84
botón correcto	número de interacción	función cuadrática inversa	0.99	0.998
color presionado	número de interacción	función cuadrática	0.99	0.999
dificultad	administrador	función polinomial inversa	0.52	0.53

Nicolás, al igual que María del Rosario, tiene correlaciones muy altas, pero mismo caso, se pudieron superar casi todas las variables objetivo utilizando la función cuadrática, cuadrática inversa y polinomial inversa; la única variable que no pudo superarse fue ‘tiempo de interacción’, pero la función cuadrática fue la que tuvo una correlación más alta, con 0.84.

René:

variable dependiente	variable independiente	modelo utilizado	Correlación simple	Correlación del modelo no lineal
tiempo de interacción	botón correcto	función cuadrática	0.98	0.87
botón correcto	número de interacción	función cuadrática inversa	0.98	0.998

color presionado	número de interacción	función cuadrática inversa	0.98	0.99
dificultad	NA	NA	NA	NA

En el caso de René la variable ‘tiempo de interacción’, sin embargo, no se pudo superar con el modelo no lineal, pero la función cuadrática fue con la que se obtuvo la correlación más alta en comparación con las demás. Para las variables ‘botón correcto’ y ‘color presionado’ se utilizó la función cuadrática inversa, y ambas obtuvieron un valor de 0.99 o superior. En cuanto a ‘dificultad’, no existe correlación.

Sergio Angel:

variable dependiente	variable independiente	modelo utilizado	Correlación simple	Correlación del modelo no lineal
tiempo de interacción	botón correcto	función cuadrática	0.9676	0.9691
botón correcto	número de interacción	función cuadrática inversa	0.96	0.99
color presionado	tiempo de interacción	función cuadrática inversa	0.96	0.98
dificultad	tiempo de sesión	función polinomial inversa	0.13	NAN

Para Sergio Ángel se utilizó la función cuadrática para ‘tiempo de interacción’ y la función cuadrática inversa para ‘botón correcto’, ambas superando las correlaciones simples, con 0.99 y 0.98 respectivamente. En cuanto a ‘dificultad’ no se encontró ningún modelo que funcione para esta variable.

Conclusiones

El análisis del comportamiento de los usuarios de WUUPI nos muestra que los modelos de regresión no lineal tienen un enfoque más preciso para entender las relaciones complejas entre variables, superando las bajas correlaciones del modelo lineal. El modelo cuadrático se destaca por que se puede aplicar de una manera exitosa en casi todas las pruebas dentro de este análisis, sobre todo los modelos como el logarítmico o el cuadrático inverso dieron como resultados ajustes muy superiores en casos individuales, alcanzando correlaciones cercanas a 0.99. Estos resultados nos confirman que los enfoques no lineales son de gran utilidad para tener mejores patrones de interacción y que sean más personalizados, lo que representa una base sólida para futuras mejoras en el diseño y la adaptabilidad de la plataforma.