

Actividad 3.2 (Regresión No Lineal)

Analítica de datos y herramientas de inteligencia artificial II (Gpo 501)

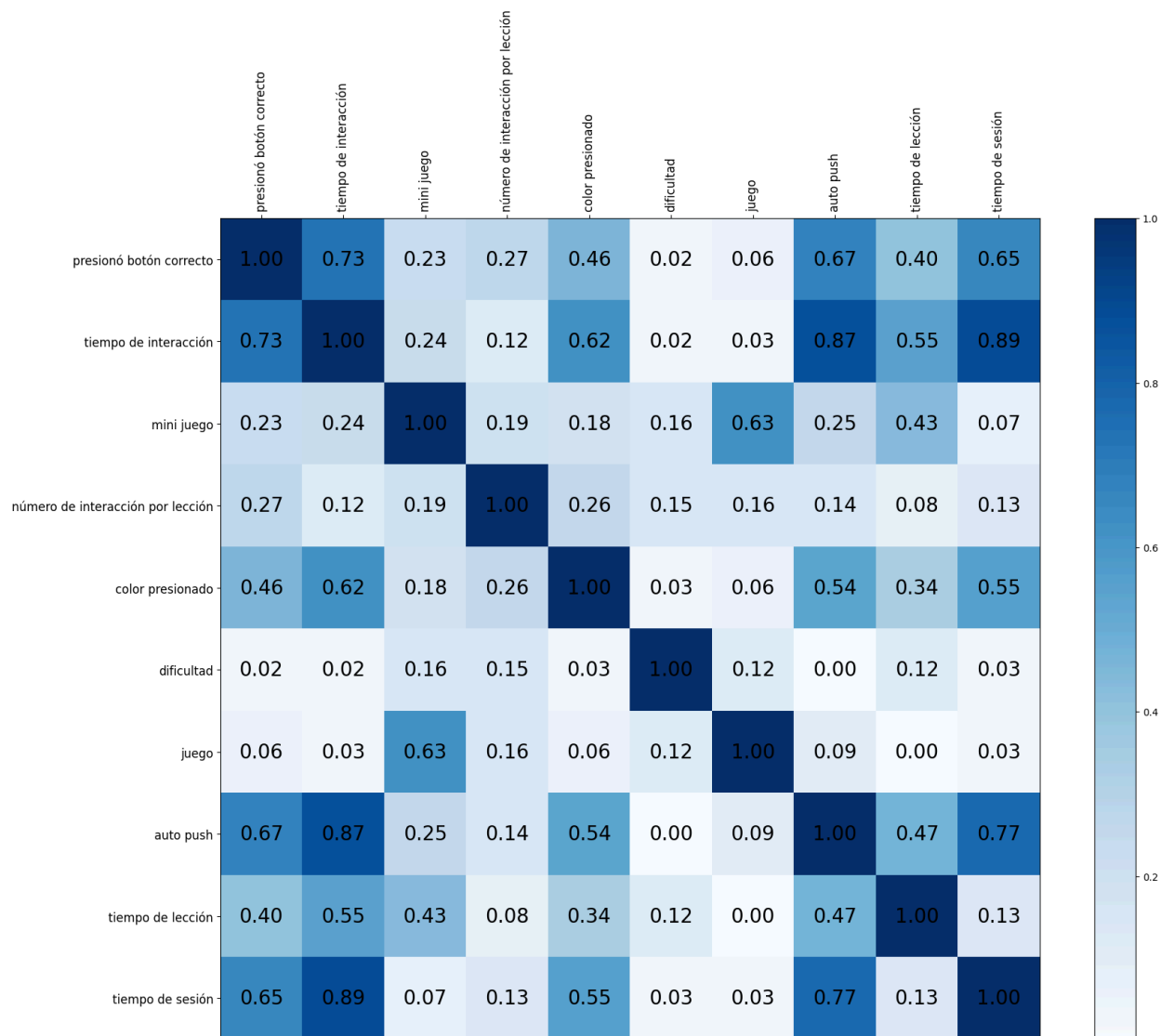
Alfredo García Suarez

EQUIPO 2

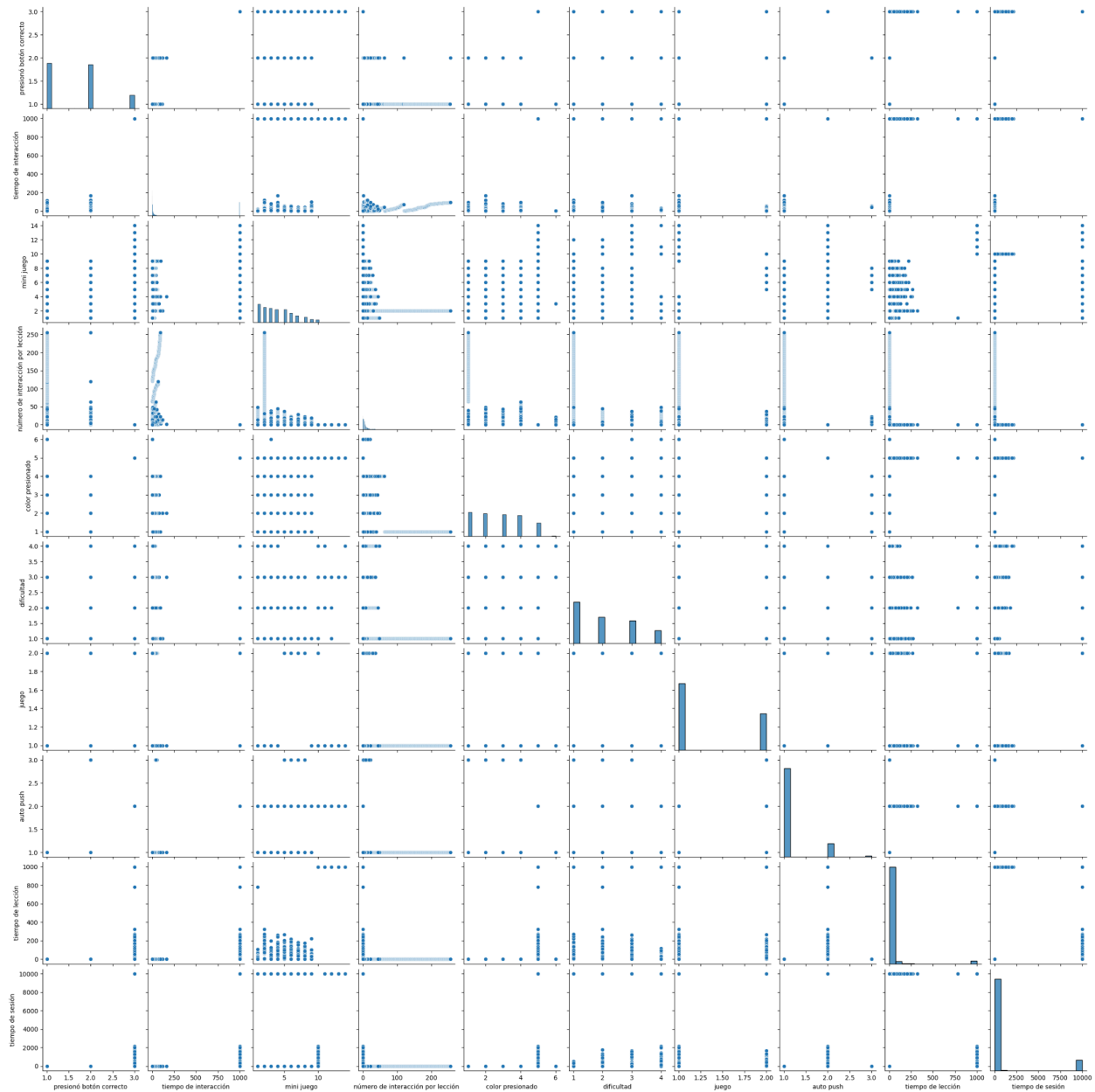
André Calmus González	A017333529
Diego Sánchez Márquez	A01734778
Emilio Rugerio Pastrana	A01737819
Ximena Italya Jimenez Huerta	A01277667

12 de Abril del 2025

Para poder realizar un análisis no lineal de los datos de Wuppi, primero tuvimos que eliminar las columnas que no eran realmente necesarias para el estudio (fecha, administrador y usuario). Se seleccionó la columna objetivo que es la de más importancia de acuerdo con estudios lineales previos que es 'presionó botón correcto', sacamos la correlación de la información adquirida por el socio anteriormente, lo que ayudó a la elaboración del Heatmap.



De esta misma información ‘limpia’ obtuvimos modelos lineales de la dispersión entre todas las variables:



Con esto ya sabemos que todas las columnas son numéricas; modificadas anteriormente.

Obtuvimos el modelo lineal simple que dió como resultado lo siguiente:

La mejor variable correlacionada con presionó botón correcto es tiempo de interacción con una correlación de 0.73

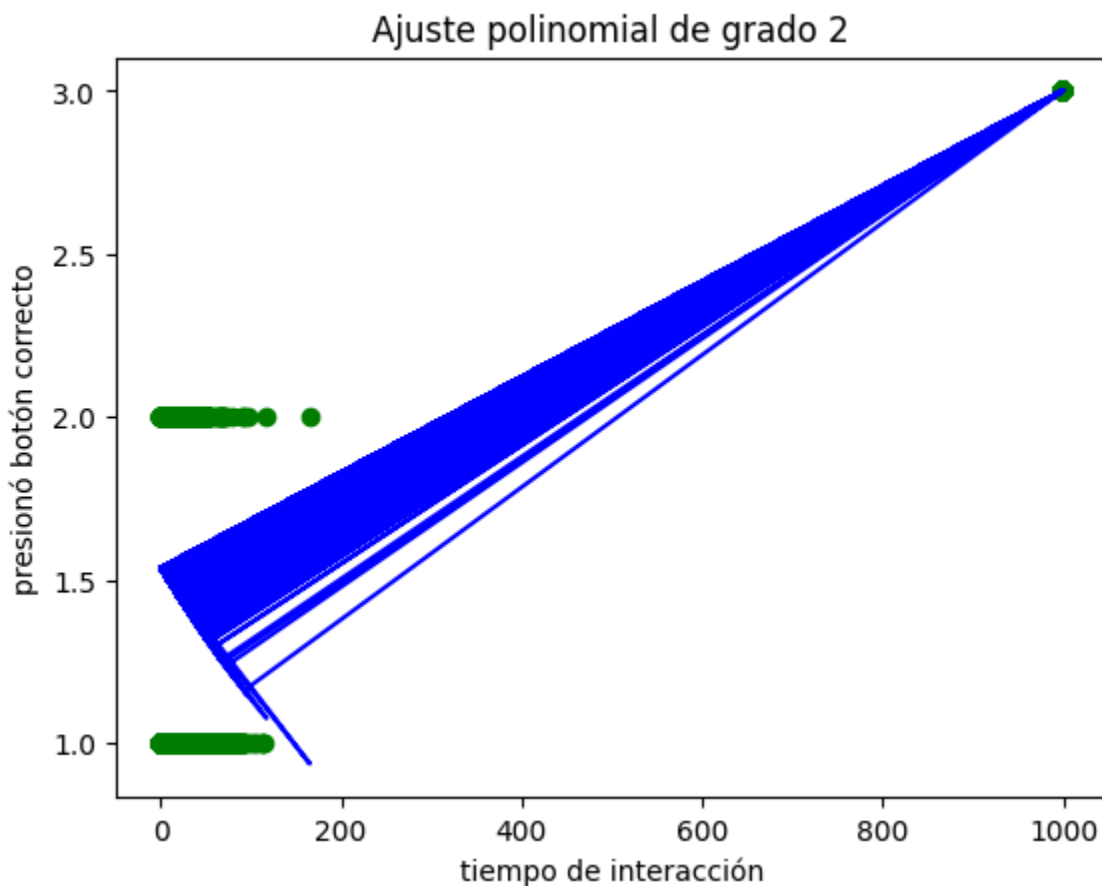
También se calculó un modelo no lineal:

R^2 del modelo polinomial de grado 2: 0.55

R^2 de la correlación lineal simple: 0.54

Lo que demuestra que el modelo no lineal supera por .01 el modelo lineal; por lo tanto, la relación entre 'tiempo de interacción' y 'presionó botón correcto' no es perfectamente lineal.

Arrojando la siguiente gráfica:



La forma de abanico indica que la varianza cambia con el tiempo de interacción. Al inicio hay muchos datos agrupados; conforme el tiempo aumenta, los datos se dispersan más.

Como conclusión, se observó que la correlación lineal simple entre "tiempo de interacción" y "presión botón correcto" presentó un R^2 de 0.54. Posteriormente, al ajustar un modelo polinomial de grado 2, se obtuvo un R^2 de 0.55, evidenciando una ligera mejora en la capacidad de ajuste del modelo no lineal respecto al lineal.

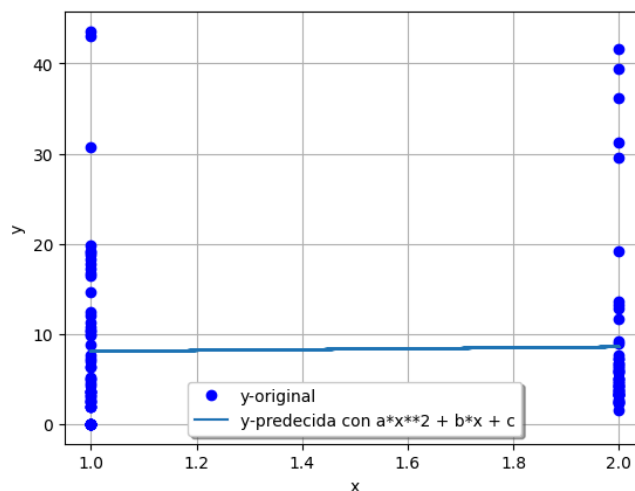
La dispersión de los datos sugiere una relación no estrictamente lineal, apoyada visualmente por la forma de abanico en la gráfica, lo cual justifica el pequeño incremento en el desempeño al emplear un modelo de mayor complejidad.

Después pudimos hacer filtros por usuarios para tener una análisis más profundo y personalizado. Usando como ejemplo a Erick Osvaldo usando:

$$y = ax^2 + bx + c \quad (\text{"Función cuadrática"})$$

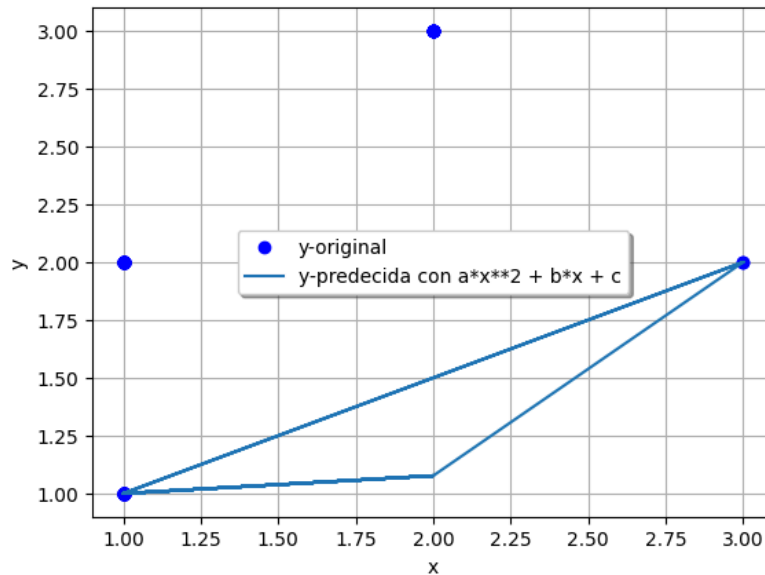
Propusimos 4 modelos con la misma función para conocer la correlación que existe entre las columnas objetivo.

Modelo 1 $x = \text{preionó botón correcto (datos nulos eliminados)}$ | $y = \text{tiempo de interacción}$



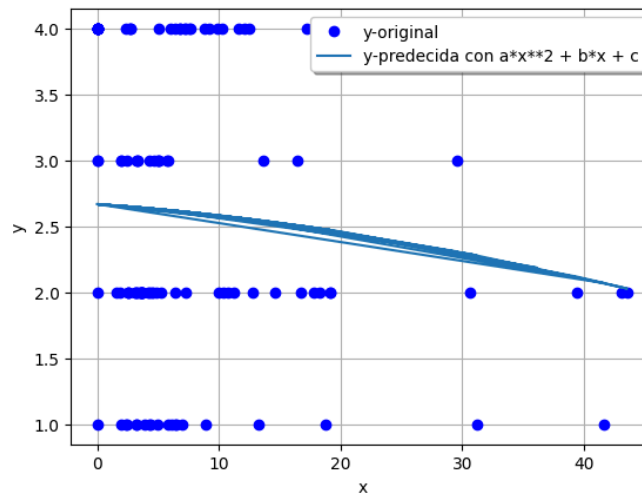
El coeficiente de correlación es 0.02

Modelo 2 $x = \text{auto push}$ (datos nulos eliminados) | $y = \text{preionó botón correcto}$



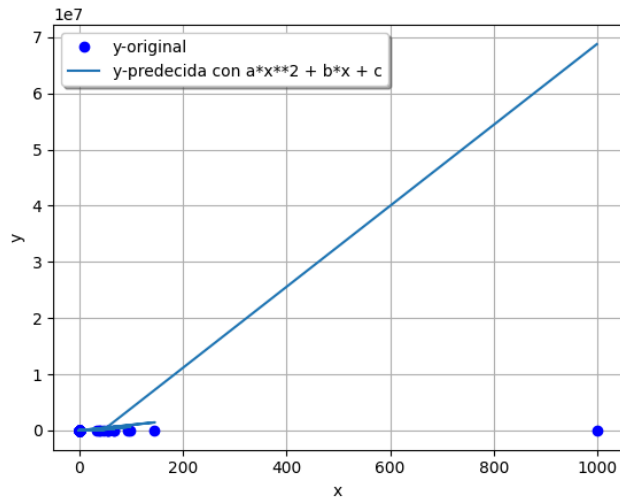
El coeficiente de correlación fue nulo, lo que significa que no existe una dependencia entre los datos para su comportamiento.

Modelo 3 $x = \text{tiempo de interacción}$ | $y = \text{colore presionado}$ (en ambos se eliminaron los datos nulos)



Con un coeficiente de correlación de 0.11

Modelo 4 $x = \text{tiempo de lección}$ | $y = \text{dificultad}$ (Estos datos nulos no fueron eliminados)



El coeficiente de correlación es nulo.

En conclusión, cuando los datos nulos son eliminados el coeficiente de correlación incrementa levemente, sin embargo las columnas usadas para el estudio no están correlacionadas. Por lo que deberían hacerse más filtros por columnas y calcular las correlaciones e identificar con diferentes usuarios para conocer cuál es la correlación más alta.