



Tecnológico de Monterrey

Actividad 4 (Regresión lineal y logística)

Escuela: Instituto Tecnológico y de Estudios Superiores Monterrey

Materia: Desarrollo de proyectos de análisis de datos

Profesor: Alfredo García Suárez

Nivel Académico: Profesional

Grupo: 301

Ciudad: Puebla

Autor

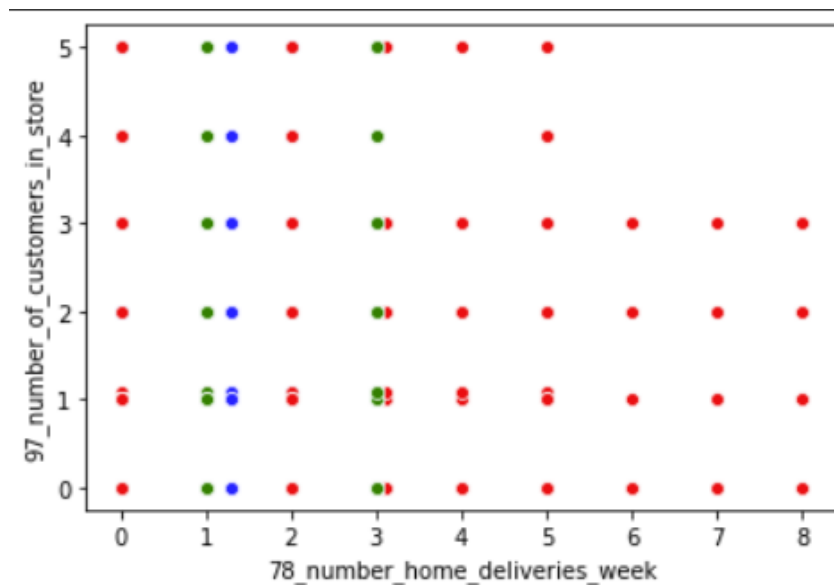
Ulises Nuñez Bautista

A01736379

Fecha de entrega:

27 de Noviembre del 2022

Para realizar el análisis del con el archivo de “microretailer_mit_lift_lab_actualizado.csv” utilizamos los métodos de nulos de backward fill y de forward fill, así que escogimos 4 variables numéricas, 3 de estas van a ser variables independientes (2_current_permanent_employees, 145_number_direct_competitors, 78_number_home_deliveries_week) y la última va a ser nuestra variable dependiente (97_number_of_customers_in_store), así que realizamos un diagrama de dispersión con las variables independientes en la x y nuestra variable dependiente en la y , y nos mostró lo siguiente:



Gráfica 1. Diagrama de dispersión

Posterior a realizar la *Gráfica 1* obtenemos los coeficientes de cada variable independiente para el modelo ajustado.

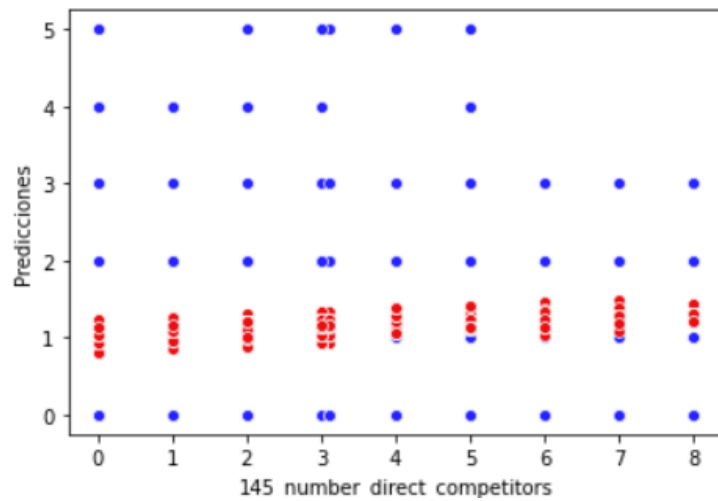
Tabla 1. Coeficientes de las variables independientes

Variable independiente	Coeficiente obtenido
2_current_permanent_employees	0.10392455
145_number_direct_competitors	0.03688242
78_number_home_deliveries_week	0.05415778

Después de obtener los coeficientes, hacemos la evaluación del modelo obtenido con el método de r^2 y el valor que nos resultó fue 0.010436383550985084, este valor nos quiere decir la variabilidad de un factor puede ser causada por su relación con otro factor

relacionado, este siempre va a estar en un rango de 0 a 1, entonces el valor obtenido podemos decir que no modela con exactitud los valores.

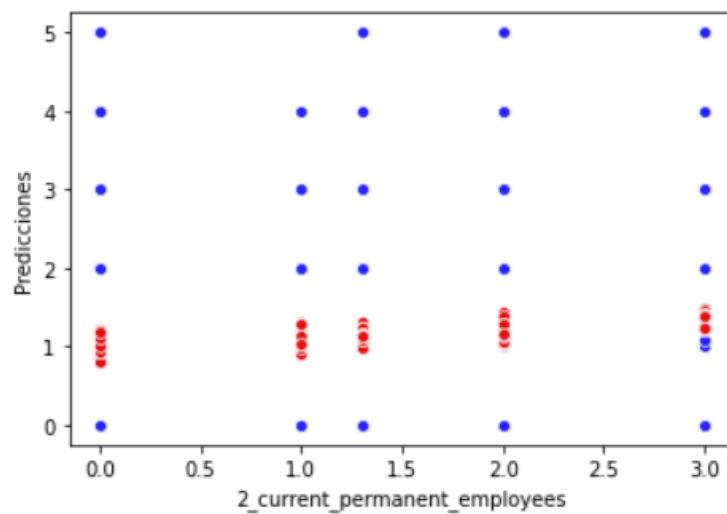
Ya realizado esto hicimos las gráficas con el valor real y el valor que se predice,



Gráfica 2. Número de competidores vs el número de clientes en la tienda

rojo= predicciones

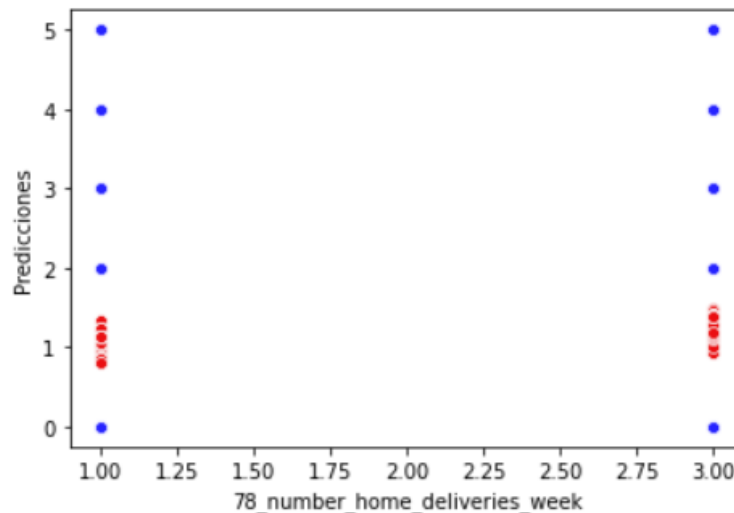
azul= valores reales



Gráfica 3. Número de empleados vs el número de clientes en la tienda

rojo= predicciones

azul= valores reales



Gráfica 4. Número de envíos a domicilio por semana vs el número de clientes en la tienda

rojo= predicciones

azul= valores reales

Ya que tenemos las gráficas sacamos el coeficiente de correlación entre las variables independientes y la dependiente, y el valor que obtuvimos fue 0.10215861956284004 y vemos que tiene correlación positiva pero aún lejos del máximo que es 1.

Ya que realizamos el metodo de regresión lineal, pasamos con regresión logistica en este caso las variables independientes fueron 145_number_direct_competitors', '97_number_of_customers_in_store', '268_number_fridges y la variable dependiente es 310_burnout.

Ya que las tenemos definidas dividimos el conjunto de datos en entrenamiento y prueba, realizamos una predicción que nos salio asi:

```
array(['no', 'no', 'no', 'yes', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no',
      'no', 'no', 'yes', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no',
      'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no',
      'no', 'no', 'yes', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no',
      'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no',
      'yes', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no', 'no'],
      dtype=object)
```

Imagen 1. Predicción logística

Después de la predicción obtenemos la matriz de confusión que nos permite analizar el desempeño de un algoritmo.

```
Matriz de Confusión
[[36  2]
 [26  2]]
```

Imagen 2. Matriz de confusión

Ya después de obtener la matriz se calcularon la precisión, exactitud y la sensibilidad del modelo obtenido.

Tabla 2. Valores de la regresión logística

Tipo	Valor	Descripción
Precisión	0.5	Se refiere a lo cerca que está el resultado de una medición del valor verdadero.
Exactitud	0.5757575757575758	La exactitud se define como la cantidad de predicciones positivas que fueron correctas.
Sensibilidad	0.07142857142857142	Consiste en determinar la magnitud del cambio en la respuesta, al cambio de valores en los componentes del modelo.

Para realizar los otros modelos seguimos el mismo camino, y una tabla con los resultados importantes de las otras regresiones.

Regresiones lineales

Estas variables independientes van con la variable dependiente de 163_number_high_perishable_products

Variable independiente	Coefficiente obtenido	Valor de r^2
272_card_days_receive_money	-0.4682398	0.032500616304659724
145_number_direct_competitors	0.55347023	
4_number_permanent_empl	-0.09026837	

oyees_last_year		
-----------------	--	--

Estas variables independientes van con la variable dependiente de 275_platform_days_receive_money

Variable independiente	Coefficiente obtenido	Valor de r^2
172_supplier_frequency	0.25884573	0.008052919769554245
104_how_many_shelves_does_the_micro_retailer_have	0.00733905	
103_number_own_fridges	-0.05739256	

Regresión logística

Variables independientes:

272_card_days_receive_money', '145_number_direct_competitors', '4_number_permanent_employees_last_year

Variables dependientes: 310_burnout

Tipo	Valor
Matriz de confusión	<pre>[[37 7] [14 8]]</pre>
Precisión	0.5333333333333333
Exactitud	0.6818181818181818
Sensibilidad	0.36363636363636365

Variables independientes:

172_supplier_frequency', '104_how_many_shelves_does_the_micro_retailer_have', '103_number_own_fridges

Variables dependientes: 310_burnout

Tipo	Valor
Matriz de confusión	<pre>Matriz de Confusión [[31 11] [19 5]]</pre>

Precisión	0.3125
Exactitud	0.5454545454545454
Sensibilidad	0.20833333333333334

Concluimos con que ambos modelos obtenidos no son del todo compatibles con las variables seleccionadas, ya que en el primer método el valor de r^2 fue muy lejano al 1, y en el segundo método la exactitud y la precisión fueron ambos cercanos o en 0.5 de acierto y la sensibilidad si quedo muy lejana al 1.