

## Metodología

Los datos que tomamos como base son los siguientes:

	Año	Ventas	Retail	AP_AnC	CAPEX	PIB	Inflación
0	2008	244917	2243854	73701	11316	1109.99	5.13
1	2009	270451	2263837	79265	9735	900.05	5.30
2	2010	335857	2366038	125695	13130	1057.80	4.16
3	2011	380907	2516975	135820	18352	1180.49	3.40
4	2012	412060	2827168	135991	14660	1201.09	4.11
5	2013	423823	2946624	139270	13987	1274.44	3.80
6	2014	440988	3036967	147892	12691	1315.36	4.02
7	2015	475911	3234206	157385	12526	1171.87	2.72
8	2016	532384	3498472	171238	14335	1078.49	2.82
9	2017	573371	3787143	175088	17426	1158.91	6.04
10	2018	616910	4039711	178594	17933	1222.41	4.90
11	2019	646846	4293736	183548	20575	1269.40	3.63
12	2020	701734	4217760	186985	16728	1087.12	3.40
13	2021	736044	4583356	195038	20466	1294.83	5.69

Figura 1: Datos Utilizados para el modelo

Primero decidimos que las características que vamos a usar son las siguientes: Retail, AP\_AnC, CAPEX, PIB y la Inflación, mientras que la variable objetivo será las Ventas, posteriormente separamos en dos grupos los datos, el primero para entrenamiento con el 70 % de las observaciones y el segundo para la evaluación del modelo con el 30 % de los datos. Por otro lado, se decide hacer distintas transformaciones a los datos de entrenamiento para realizar distintos modelos que son las siguientes: la normalización MinMax que lo que hace es pasar todos los valores de las características en el intervalo [0,1], la estandarización que lo que hace es restar la media de las características y se dividen sobre la desviación estándar de estos, otra transformación fue la reducción de dimensiones usando la técnica de PCA usando 3 componentes principales y por último los datos se quedan tal cual vienen en la base de datos, tendiendo así 4 grupos derivados de los datos de entrenamiento que serían los normalizados, los estandarizados, los reducidos por PCA y los datos crudos. Para cada uno de estos grupos se van a realizar 3 tipos de regresión

que son la lineal, la regresión Ridge y la regresión Lasso donde la regresión Ridge busca minimizar  $\min_B(\|y - XB\|_2^2 + \alpha\|B\|_2^2)$  donde  $\alpha$  es un factor de penalización,  $X$  son los datos y  $B$  un vector de pesos, por otro lado la regresión Lasso busca minimizar  $\min_B((2n)^{-1}\|y - XB\|_2^2 + \alpha\|B\|_1)$  donde  $\alpha$  es un factor de penalización,  $X$  son los datos y  $B$  un vector de pesos y  $n$  es el número de las observaciones, entonces se realizaron 3 modelos para cada uno de los 4 grupos con distintos tipos de penalización alfa, teniendo así el mejor modelo con la regresión tipo Ridge sobre los datos estandarizados con  $\alpha = 0,015$ .

## Resultados

En la siguiente imagen se puede ver las cantidades reales y las cantidades predichas por el modelo.

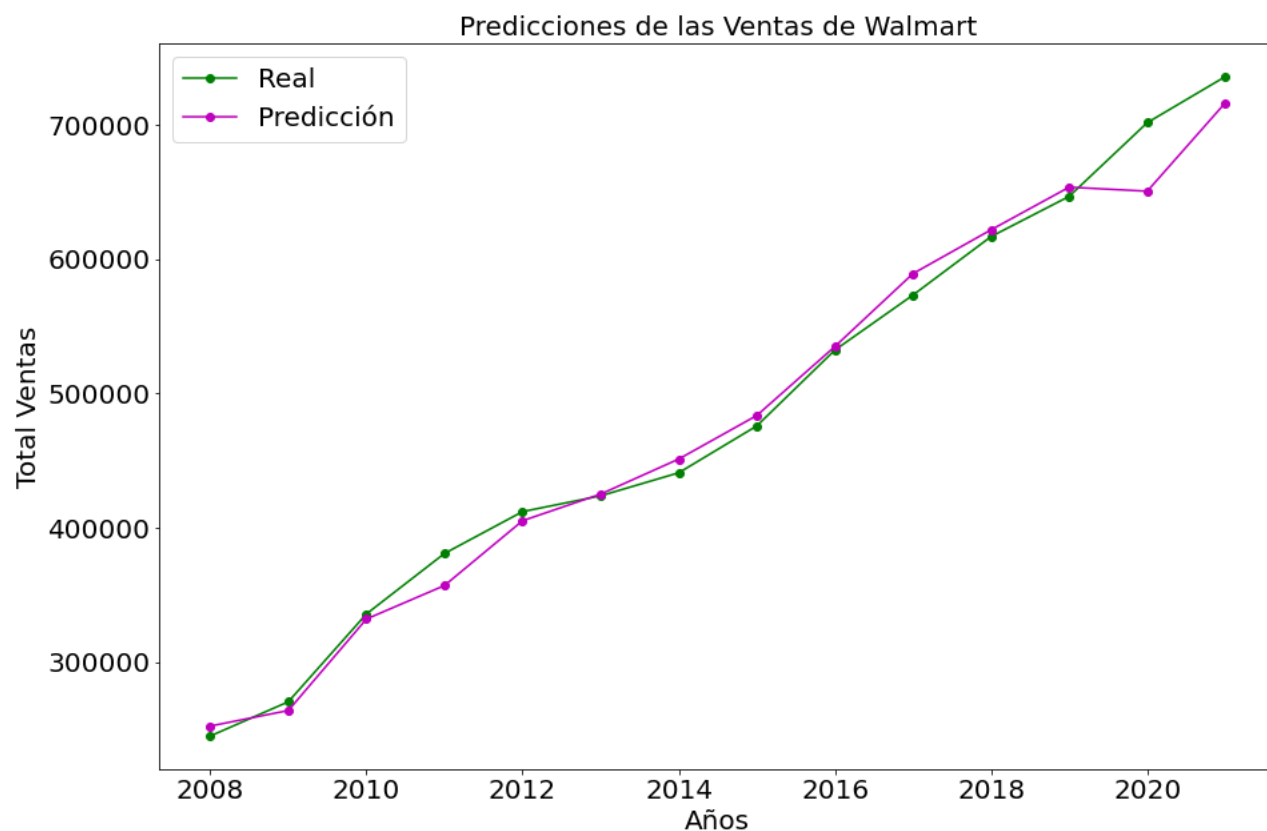


Figura 2: Resultados del modelo

Como se puede ver en la figura 2 el modelo de regresión Ridge seleccionado se ajusta muy bien a las ventas de Walmart, por lo que teniendo una predicción de los datos seleccionados se podría dar una predicción que el modelo ya entrenado.