

Trabajo 2 ICD: Modelo Base

Expectativas económicas Empresariales

Abad Aniceto Anguiela, Navarro Lozada Damaris, Robledo Jimenez
Meyli, Pizarro Otero Alberto

Introducción a Ciencia de Datos y Machine Learning con Python

21 de octubre de 2025

- El objetivo de este trabajo es plantear un modelo base para el análisis de la pregunta de investigación:
- ¿Cómo influyen el tipo de cambio interbancario, la inflación mensual, el PBI mensual y la tasa de referencia del BCRP en el índice de expectativas sobre la economía a tres meses en el Perú durante el periodo 2015–2025?
- Se evaluará distintos modelos (lineales y no lineales) y analizar su capacidad predictiva mediante métricas adecuadas.

Training y Test:

- ➊ **División de datos:** entrenamiento (70 %) y prueba (30 %).
- ➋ **Estandarización:** se utiliza `StandardScaler` para normalizar variables numéricas.
- ➌ **Modelado:** se emplean modelos de regresión lineal y logística según la naturaleza de la variable dependiente.
- ➍ **Validación:** se aplican técnicas de `cross-validation` (`KFold`) y métricas de desempeño.

Modelo de Regresión Lineal

El modelo de regresión lineal se plantea como:

$$\hat{ExpEconómicas} = \beta_0 + \beta_1 TC + \beta_2 TPM + \beta_3 PBI + \beta_4 IPC + \varepsilon$$

- Donde *ExpEconómicas* representa el valor estimado de la variable objetivo.
- **Modelo simple:** El coeficiente de nuestra variable Tipo de Cambio mostró tener efecto estadísticamente significativo y negativo (-20.8211).
- **Modelo complejo:** Las variables independientes - con excepción de TPM- si muestran un efectos estadísticamente significativo. Sin embargo, la validación cruzada nos muestra que en terminos predictivos el modelo no mejora, dado que su MSE y desviación estandar aumentan a 267.2365 y 192.6637 respectivamente.

Modelo de Regresión Lineal

--- Tabla Comparativa de Rendimiento Predictivo ---

	MSE Promedio (CV)	Std Dev (CV)
Modelo		
Baseline	103.202361	0.000000
OLS Simple	107.776252	121.361898
OLS Complejo	267.236451	192.663708

Figura: Tabla Comparativa de Rendimiento Predictivo

El modelo OLS Simple se ajusta mejor a nuestra variable dependiente (Expectativas Económicas a 3 meses), lo que evidencia que el añadir variables no mejora el modelo, o que dichas variables no son las más indicadas.

Modelo de Regresión Logística

Transformamos nuestra variable de Expectativas en una variable categórica, donde:

- 1 = optimismo si > 50
- 0 = Pesimismo si < 50

Cuando la variable objetivo es categórica, se aplicó un modelo logístico:

$$P(\text{ExpEconómicas} = 1|X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 TC + \beta_2 TPM + \beta_3 PBI + \beta_4 IPC)}}$$

Modelo de Regresión Logística

```
Logit Regression Results
=====
Dep. Variable:  Expectativas_Binaria    No. Observations:    95
Model:          Logit                  Df Residuals:        90
Method:         MLE                     Df Model:             4
Date:           Mon, 20 Oct 2025        Pseudo R-squ.:       0.2752
Time:           17:49:27                Log-Likelihood:      -47.538
Converged:      True                    LL-Null:              -65.591
Covariance Type: nonrobust              LLR p-value:         2.752e-07
=====
               coef    std err          z      P>|z|    [0.025    0.975]
-----
const          19.5315     5.453     3.582    0.000     8.844    30.219
Tasa_Referencia_PM -0.2526     0.175    -1.444    0.149    -0.596     0.090
Tipo_Cambio_Promedio -11.4107     2.583    -4.418    0.000    -16.473    -6.348
PBI_Mensual      0.0222     0.029     0.764    0.445    -0.035     0.079
IPC              0.1804     0.062     2.901    0.004     0.059     0.302
=====

--- Odds Ratios ---
const          303,691,284.5258
Tasa_Referencia_PM    0.7767
Tipo_Cambio_Promedio  0.0000
PBI_Mensual          1.0225
IPC                  1.1977
dtype: float64
```

Figura: Modelo Logit

- Solo Tipo de Cambio y el IPC muestran ser predictores significativos del evento de interés.
- Pseudo R-cuadrado: Nos indica que el modelo explica al rededor de un 27.52 % de la variabilidad del evento.
- En conclusión tiene capacidad explicativa moderada y mejora.

Comparación y Decisión del Problema

- Regresión: El modelo OLS simple, a pesar de su mejor ajuste debido a un menor MSE y desviación standar en comparación con el modelo complejo, posee un poder predictivo de solo el 27.7 %, lo que evidencia un posible problema de sobre ajuste.
- Clasificación: En base al Accuracy promedio el modelo acierta en un 60 % de los casos al predecir . El valor ± 0.1886 indica la variabilidad del desempeño entre los distintos pliegues (folds) de la validación cruzada. El F1-Score promedio nos indica un equilibrio moderado (59.72 %).
- Es más relevante tener la capacidad de predecir si las Expectativas Económicas que tienen los empresarios resultan ser positivas o negativas ante la variación de variables económicas como el Tipo de Cambio y el PBI.

Conclusiones

- Se implementaron modelos de regresión lineal y logística para analizar la influencia de variables macroeconómicas (Tipo de Cambio, IPC, PBI y TPM) sobre las expectativas económicas empresariales a 3 meses en Perú.
- Aunque el modelo de regresión lineal simple ofreció un mejor ajuste que el modelo múltiple, su bajo poder predictivo sugiere limitaciones para capturar relaciones complejas o no lineales.
- El modelo de regresión logística resultó más adecuado al transformar la variable dependiente en una categoría binaria (optimismo vs pesimismo). Esto permitió interpretar mejor los efectos direccionales y la probabilidad de eventos económicos relevantes.
- El Tipo de Cambio y el IPC fueron identificados como predictores estadísticamente significativos en el modelo logístico, lo que resalta su impacto directo en la percepción empresarial del entorno económico.