

Trabajo 3 ICD: Modelos Avanzados

Expectativas económicas Empresariales

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

Introducción a Ciencia de Datos y Machine Learning con Python

19 de Noviembre de 2025

- 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?

- **Motivación:**

- Ampliar el análisis incluyendo nuevas variables (rezagos, variaciones, transformaciones e interacciones), aplicamos PCA para entender mejor la estructura de los datos
- Obtener un modelo sólido, bien justificado y reproducible, que permita entender y predecir el comportamiento de las expectativas económicas en el tiempo.

Metodología:

- Modelo anterior elegido: Logit, nos permite estimar directamente la probabilidad de que el indicador de expectativas sea mayor a 50.
- Incluimos nuevas variables: Rezagos, variaciones, transformaciones e interacciones.
- PCA para entender mejor la estructura de los datos y comparamos tres modelos de clasificación: Logit, Random Forest y XGBoost.
- Tratamos el desbalanceo de clases, evaluamos las métricas clave y utilizamos costos para falsos positivos y falsos negativos con el fin de definir un umbral de decisión mas adecuado.

Nuevas Variables

1. Rezagos (Lag 1)

- Tipo de cambio, IPC, PBI y tasa de referencia.
- Capturan efectos dinámicos y evitan simultaneidad.

2. Variaciones mensuales

$$\Delta X_t = X_t - X_{t-1}$$

- Permiten identificar shocks y cambios recientes.

3. Interacciones

TC × IPC: Tipo de cambio vs Inflación.

TC × Tasa: Política monetaria y tipo de cambio.

4. Transformaciones logarítmicas de TC, PBI e IPC

- Reducen heteroscedasticidad.
- Facilitan interpretación como elasticidades.

PCA: Resultados Principales

- Se estandarizaron las variables y se extrajeron 2 componentes.
- **PC1: 53.8 % de la varianza**
 - Captura el nivel general de estabilidad macroeconómica.
- **PC2: 13.7 % de la varianza**
 - Captura shocks o variaciones recientes.
- **Varianza total explicada: 67.4 %**
- La separación entre optimistas y pesimistas sugiere que las expectativas responden a patrones macroeconómicos estructurales.
- El PCA valida la existencia de ciclos económicos visibles en los datos.

PCA: Resultados Principales

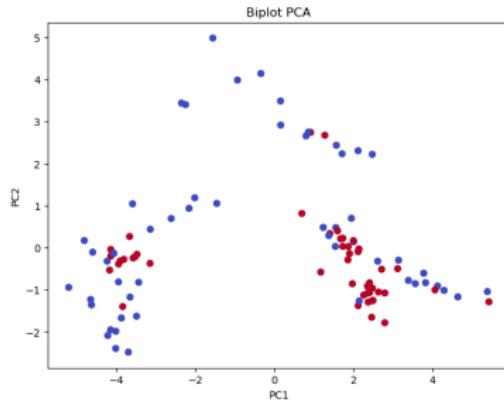


Figura: Grafica Biplot

El PCA sí encuentra cierta separación entre optimistas y pesimistas. Esto indica que las variaciones en el tipo de cambio, el IPC, el PBI y la tasa de referencia sí generan patrones económicos distinguibles.

No es perfecta (porque economía = ruido), pero hay estructura: PC1 diferencia “estabilidad económica” vs “deterioro económico”. PC2 captura cambios/variaciones macro recientes.

Modelo Logit

Resultados:

- F1-score: 0.75
- AUC: 0.79

Interpretación:

- Sugiere que las variables macroeconómicas incluidas permiten capturar parte de la dinámica que determina si las expectativas son optimistas o no.
- Sin embargo, el modelo lineal no logra capturar totalmente las relaciones no lineales y las interacciones típicas de la economía peruana.

Random Forest

- Se realizó GridSearch con validación cruzada.
- Parámetros óptimos:

$$n_estimators = 200, \quad max_depth = 7$$

- AUC 0.92 (mejor modelo).

Interpretación económica:

- Detecta efectos no lineales típicos: umbrales, reacciones abruptas, fuera de rango.
- Permite inferir que los agentes forman expectativas considerando interacciones complejas.

- Hiperparámetros óptimos:

$$n_estimators = 200, \quad max_depth = 5, \quad learning_rate = 0,1$$

- AUC 0.91 (similar a Random Forest).

Interpretación:

- Modelo muy potente, especialmente para patrones no lineales frecuentes en series macroeconómicas.
- Confirma que la formación de expectativas no sigue una sola tendencia lineal.
- Modelos más flexibles permiten predecir con mayor precisión los momentos de optimismo empresarial.

Comparación de Modelos

Modelo	AUC
Logit	0.79
Random Forest	0.92
XGBoost	0.91

- **Random Forest** es el mejor modelo.
- Diferencia importante respecto al Logit: la macroeconomía contiene interacciones complejas.

Umbral Óptimo con Costos FP/FN

- Se definieron costos:

$$C_{FP} = 3, \quad C_{FN} = 1$$

- Se evaluaron 500 posibles umbrales.
- Umbral óptimo:

$$t^* = 0,33$$

Interpretación:

- Clasificar falsamente un estado como optimista es más costoso.
- Un umbral más conservador evita sobreestimar optimismo en períodos inciertos.
- Esto es coherente con el contexto macroeconómico: es preferible ser prudente ante señales débiles antes que asumir optimismo.

Gráfica de Costos

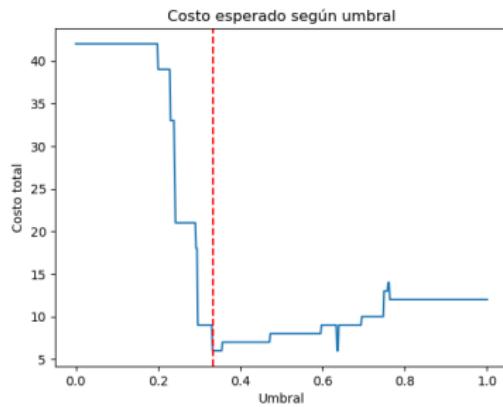


Figura: Gráfica de Costos

- Cuando el umbral es muy bajo o muy alto, el costo total aumenta debido a un exceso de errores. El punto mínimo se encuentra cerca de 0.33, lo que representa el equilibrio entre no cometer demasiados falsos optimismos y no exagerar el pesimismo. En términos económicos, este umbral refleja una estrategia prudente para interpretar la señal de las expectativas.

Conclusiones Generales

- Las expectativas económicas pueden predecirse razonablemente bien con variables macroeconómicas.
- El PCA revela estructura económica clara entre optimismo y pesimismo.
- El análisis de costos indica que los falsos positivos son más costosos económicamente
- Logit funciona, pero modelos no lineales capturan mejor la complejidad macroeconómica.
- Random Forest es el mejor modelo (AUC 0.92).
- Ajustar el umbral permite controlar el riesgo de interpretar señales débiles como optimismo.

Aportes del Trabajo

- Integración de ML con análisis macroeconómico.
- Justificación económica de cada transformación y modelo.
- Metodología reproducible y adaptable.
- Incorporación de costos para decisiones de política económica o empresarial.