|  |  |
| --- | --- |
|  |  |
| Proyecto 4 Riesgo Liquidez |  |
|  |  |
|  | Leonardo Daniel Rosas Ríos  María Hernández Fernández Fernanda  Martha Lizeth Pastrana Basilio  José Domingo Cortes Sandria |

Tabla de contenido

[Introducción 3](#_Toc199703749)

[Productos Bancarios para el Ahorro 3](#_Toc199703750)

[Descripción de la base de datos y variables utilizadas 4](#_Toc199703751)

[**Base de datos** 4](#_Toc199703752)

[**Periodo de análisis** 4](#_Toc199703753)

[**Variables utilizadas** 4](#_Toc199703754)

[Modelos de regresión utilizados 5](#_Toc199703755)

[**A.** **Regresión Lineal Múltiple** 5](#_Toc199703756)

[**B.** **Regresión Lasso (Least Absolute Shrinkage and Selection Operator)** 5](#_Toc199703757)

[**C.** **Regresión Ridge** 6](#_Toc199703758)

[**D.** **Regresión Elastic Net** 6](#_Toc199703759)

[**Estandarización de variables** 6](#_Toc199703760)

[**Validación cruzada** 7](#_Toc199703761)

[**Métricas utilizadas** 7](#_Toc199703762)

[Resultados 8](#_Toc199703763)

[Resultados del análisis comparativo 8](#_Toc199703764)

[Conclusiones 8](#_Toc199703765)

[Aprendizajes obtenidos 8](#_Toc199703766)

# Introducción

Las cuentas de captación son unos de los pilares fundamentales del sistema bancario, esto porque son los recursos que las instituciones financieras reciben de personsas o empresas, y que luego es invertido en diferentes actividades de producción o de consumo.

El objetivo de este proyecto es desarrollar un modelo macroeconómico de regresión múltiple, para analizar y explicar el efecto que diferentes variables macroeconómicas tienen en las cuentas de capatación (Depósito Vista, Depósitos a plazo y Captación tradicional).Con este proyecto se desea identificar las variables que tienen mayor influencia sobre el comportamiento de las cuentas de captación, generando la posibilidad de identificar patrones y predecir cambios en el futuro.

## Productos Bancarios para el Ahorro

Los bancos ofrecen una gran variedad de productos destinados a satisfacer la necesidad de las personas de guardar su dinero de forma segura, generando intereses o rendimientos con bajo riesgo, para poder ser usado en un futuro. Los principales productos para el ahorro son: los pagarés, certifiacos de depósito, fondos de inversión, cuentas de ahorro, cuentas a la vista y seguros de ahorro, etc. Cada uno con diferentes características y beneficios.

Pagares Bancarios

Estos son instrumentos a corto y mediano plazo que ofrecen una tasa fija de interés. Es ideal para ahorradores que buscan una rentabilidad predecible y clara, a cambio de mantener su inversión durante un plazo previamente acordado.

Características

* Tasa fija.
* Plazo fijo (generalmente entre 30 y 365 días).
* Penalización por retiros anticipados.

Certifiados de depósito

Son productos de ahorro a plazo con una tasa de interés fija que asegura ganancias por la permanencia del capital invertido durante un periodo previamente acordado.

Características

* Tasa de interés atractiva.
* Plazos variables, superiores a los pagares.
* Ideal para plazos a largo y mediano plazo.

Fondos de inversión

Es una herramienta que permite a los ahorradores acceder a una cartera de activos administrada por expertos, con diferentes opciones de riesgo y por lo tanto de rendimiento.

Características

* Inversión diversificada.
* Liquidez variables.
* Diferentes opciones de riesgo.

Cuentas de Ahorro

Son una opción de bajo riego que premite depositar y retirar dinero con flexibilidad.

Características

* Mayor seguridad de inversión.
* Baja tasa de interés.
* Alta liquidez con acceso inmediato a los fondos.

# **Descripción de la base de datos y variables utilizadas**

## ****Base de datos****

Se utilizaron dos fuentes principales de datos en formato Excel:

1. df\_series.xlsx  
   Contiene variables macroeconómicas relevantes. Cada fila representa un periodo mensual, con la columna Fecha como clave temporal. Esta base provee indicadores económicos que podrían influir en el comportamiento de los depósitos bancarios.
2. Proyecto 4 Cuentas de Captación 2024.xlsx  
   Contiene la información financiera de captación bancaria para cinco instituciones:
   * TBM
   * Banamex
   * BBVA
   * Santander
   * Banorte

Cada banco está representado en una hoja distinta, y los datos están alineados temporalmente con la base macroeconómica.

## ****Periodo de análisis****

El análisis abarca desde **diciembre de 2019 hasta junio de 2024**, capturando eventos económicos relevantes, incluyendo la pandemia, recuperación económica y fluctuaciones inflacionarias.

## ****Variables utilizadas****

#### **Variables dependientes (a modelar): Estas son las tres formas de captación bancaria, modeladas por separado:**

* **Depósitos a la vista**
* **Depósitos a plazo**
* **Captación tradicional** (suma de los anteriores)

Cada tipo de depósito se analiza individualmente para cada banco.

#### **Variables independientes (predictoras macroeconómicas): Estas provienen del archivo df\_series.xlsx, incluye variables como:**

* "TIIE a 28 días Tasa de interés promedio mensual, en por ciento anual"
* "Tipo de cambio Pesos por dólar E.U.A. Para solventar obligaciones denominadas en moneda extranjera Fecha de determinación (FIX) Cotizaciones promedio"
* "Base Monetaria"
* "Remesas Familiares Total"
* "Índice Nacional de Precios al consumidor Variación mensual"
* "Costo de captación a plazo de pasivos en moneda nacional (CCP) Tasa en por ciento anual"
* "Agregados Monetarios M2 Instrumentos monetarios a plazo en poder de residentes 4/ Captación a plazo En otras instituciones no bancarias Entidades de ahorro y crédito popular 3/"
* "Activos Financieros Internos F1 Instrumentos no monetarios en poder de residentes Fondos de ahorro para la vivienda y el retiro Retiro En Banco de México"

La selección de variables se refina mediante técnicas estadísticas dentro de los modelos.

# **Modelos de regresión utilizados**

El objetivo fue modelar el comportamiento de los depósitos utilizando diferentes técnicas de regresión para evaluar cuál es más adecuada en cada caso. Se emplearon los siguientes modelos:

## ****Regresión Lineal Múltiple****

Modelo base que estima la relación entre una variable dependiente y múltiples variables independientes ​



 : variable objetivo (ej. depósitos)

 ​: variables predictoras (macroeconómicas)

 ​: ​coeficientes a estimar

 : término de error

## ****Regresión Lasso (Least Absolute Shrinkage and Selection Operator)****

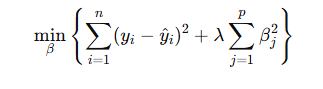
Este modelo introduce una penalización L1 que elimina coeficientes irrelevantes (los lleva a cero):



* : parámetro de regularización
* ​: penalización por el tamaño del coeficiente

## ****Regresión Ridge****

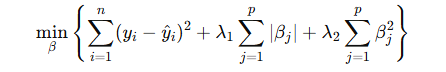
Aplica una penalización L2 que reduce la magnitud de todos los coeficientes, útil ante multicolinealidad:



* La penalización L2 no elimina variables, solo reduce sus efectos.

## ****Regresión Elastic Net****

Combinación de Lasso y Ridge. Introduce ambas penalizaciones:



* : controla el efecto tipo Lasso (selección de variables)
* : controla el efecto tipo Ridge (suavización)

## ****Estandarización de variables****

Dado que las variables macroeconómicas utilizadas tienen distintas escalas (por ejemplo, tasas de interés vs. montos monetarios), fue necesario **estandarizar** los datos antes de ajustar los modelos. Esto implica transformar cada variable para que tenga:

* **Media = 0**
* **Desviación estándar = 1**

La estandarización es especialmente importante para los modelos **Lasso, Ridge y Elastic Net**, ya que las penalizaciones que aplican se ven afectadas por la escala de las variables.

## ****Validación cruzada****

Para asegurar que los modelos tengan buen rendimiento fuera de la muestra (es decir, que no estén sobre ajustados), se empleó la técnica de **validación cruzada**. En particular, se utilizó:

* **Validación cruzada k-fold** con   
  Los datos se dividen en 5 partes iguales. En cada iteración, 4 partes se usan para entrenar el modelo y 1 para validarlo. Esto se repite 5 veces cambiando el conjunto de validación.

Esta técnica permite evaluar el rendimiento promedio del modelo de forma robusta y confiable.

## ****Métricas utilizadas****

Para comparar el desempeño de los modelos se utilizó el coeficiente de determinación:

* :mide qué proporción de la variabilidad de la variable dependiente es explicada por el modelo.
  + Valor cercano a 1: excelente ajuste.
  + Valor cercano a 0: pobre ajuste.
  + Puede ser negativo si el modelo es peor que una predicción constante.

Se emplearon tanto **R** como **Python** para aprovechar las ventajas específicas de cada lenguaje:

#### **Python**

* Se utilizó principalmente para la construcción de modelos de regresión con regularización: **Lasso**, **Ridge** y **Elastic Net**, a través de la librería sklearn.
* Permite una integración eficiente de flujos de trabajo con estandarización, validación cruzada y selección automática de hiperparámetros.
* Buena visualización y automatización de procesos de comparación.

#### **R**

* Se utilizó especialmente para los análisis exploratorios y visualización inicial de las variables.
* También se aplicaron modelos en R para validar los resultados obtenidos en Python y verificar la consistencia.
* Algunas funciones estadísticas avanzadas son más accesibles directamente en R, con menos configuración inicial.

El uso combinado de ambos lenguajes permitió contrastar resultados, validar supuestos y construir un análisis más robusto. Además, refuerza la versatilidad técnica del equipo, al aprovechar lo mejor de ambas herramientas.

# Resultados

## ****Resultados del análisis comparativo****

Se aplicaron cuatro modelos de regresión (Múltiple, Lasso, Ridge, Elastic Net) para predecir el comportamiento de tres tipos de depósitos (Vista, Plazo y Total) en cinco bancos. La comparación se realizó utilizando el coeficiente de determinación para identificar el mejor modelo en cada caso.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Banco | Tipo de Depósito | R² Múltiple | R² Lasso | R² Ridge | R² Elastic Net | Mejor Modelo |
| TBM | Vista | 0.9868 | 0.9867 | 0.9754 | 0.9866 | Múltiple |
| Plazo | 0.959 | 0.9551 | 0.8462 | 0.9584 | Múltiple |
| Total | 0.9829 | 0.9827 | 0.9618 | 0.9827 | Múltiple |
| BBVA | Vista | 0.964 | 0.9652 | 0.9449 | 0.9651 | Lasso |
| Plazo | 0.8771 | 0.8735 | 0.6849 | 0.8818 | Elastic Net |
| Total | 0.9595 | 0.9606 | 0.941 | 0.9613 | Elastic Net |
| Banamex | Vista | 0.8244 | 0.8264 | 0.8082 | 0.8271 | Elastic Net |
| Plazo | 0.7 | 0.6612 | 0.6845 | 0.7034 | Elastic Net |
| Total | 0.7222 | 0.7222 | 0.7106 | 0.7166 | Múltiple |
| Santander | Vista | 0.8799 | 0.8696 | 0.8548 | 0.869 | Múltiple |
| Plazo | 0.8742 | 0.8767 | 0.7738 | 0.8724 | Lasso |
| Total | 0.901 | 0.9084 | 0.8282 | 0.8887 | Lasso |
| Banorte | Vista | 0.932 | 0.925 | 0.922 | 0.9239 | Múltiple |
| Plazo | 0.955 | 0.9549 | 0.6606 | 0.9566 | Elastic Net |
| Total | 0.941 | 0.9451 | 0.8679 | 0.9445 | Lasso |

# ****Conclusiones****

* **Regresión múltiple** fue el mejor modelo en **5 de los 15 casos**, destacando especialmente en **TBM** y en el depósito total de **Banamex** y vista de **Santander**.
* **Elastic Net** resultó ser el modelo más robusto en escenarios con multicolinealidad o relaciones no tan fuertes, siendo el mejor en **5 casos**, incluyendo depósitos en **BBVA**, **Banamex** y **Banorte**.
* **Lasso** fue preferido en **4 casos**, especialmente para **Santander** y **BBVA**, al permitir una mejor selección automática de variables.
* **Ridge** no fue el mejor modelo en ningún caso, lo cual indica que la penalización L2 sola no fue suficiente para superar a los otros enfoques.

## 

## ****Aprendizajes obtenidos****

* La calidad predictiva varía significativamente según el banco y el tipo de depósito, lo que sugiere que no existe un único modelo óptimo universal.
* Modelos con regularización (Lasso, Elastic Net) tienden a mejorar el ajuste cuando hay muchas variables correlacionadas.
* La validación cruzada y la estandarización fueron pasos clave para obtener resultados fiables.
* Usar tanto **R como Python** permitió combinar lo mejor de cada lenguaje, y validar la consistencia entre plataformas de análisis.