

# **Análisis Predictivo de la Estabilidad Financiera Argentina (Inflación)**



***Proyecto final de Data Science (CoderHouse)***

**Autor: Romen, Ter Gazarian**

**Mail: [romenter.dev@gmail.com](mailto:romenter.dev@gmail.com)**

**Fecha: 22 ene, 2024**

# Análisis Predictivo de la Estabilidad Financiera Argentina

En este estudio se aborda el desafiante análisis y la predicción de la inflación en la volátil economía de Argentina, que es constantemente marcada por fluctuaciones significativas y una inflación acelerada. Los datos a utilizar provienen de la API del Banco Central de la República Argentina (BCRA). Dentro de nuestro análisis se busca la correlación existente entre las diferentes variables macroeconómicas, las políticas aplicadas, y además, el impacto en la economía al momento de realizarse cambio de entidades gubernamentales u administrativas.

El objetivo primordial de este estudio es diseñar y validar un modelo avanzado de machine learning que explore y prediga la dinámica de la inflación mensual en Argentina, utilizando datos económicos detallados proporcionados por el Banco Central de la República Argentina (BCRA). Este modelo se centrará en analizar cómo diversas variables económicas, como la base monetaria, las reservas internacionales, la caja de ahorro y cuentas corrientes, hechos históricos (como cepo o cambio de entidades gubernamentales) o herramientas macroeconómicas, influyen en la inflación.

## Audiencia

### ***ESTA DIRIGIDO PARA...***

Este estudio busca alcanzar (y ser un apoyo) para el sector práctico de la economía y las finanzas, dirigido a una audiencia especializada, como economistas y analistas políticos. Además permite una lectura rápida para inversores y analistas financieros, ofreciendo un marco analítico que facilita la interpretación de tendencias complejas y la identificación de oportunidades y riesgos en un mercado volátil.

Con la inflación impactando directamente en la inversión, el crédito, y la valoración de activos, la capacidad de predecir y entender estas dinámicas es fundamental para una gestión de riesgos efectiva y toma de decisiones de inversión informadas. De hecho, con algunos simples cambios se puede readaptar la herramienta aquí ofrecida para analizar las demás variables ofrecidas. Tal es la presentación del proyecto, que, reubicando la columna deseada, se podría tener una lectura sobre la evolución (y la predicción) del dólar, de la base monetaria, o las demás variables presentes en el proyecto.

## **NO ESTA DIRIGIDO PARA...**

Es importante mencionar que estudio está orientado hacia una audiencia con conocimientos básicos de economía, política, o finanzas, y no está dirigido al público en general. Viendo y considerando que existe información especializada, de variables con términos específicos, tales como base (por base monetaria), usd, cuentas corrientes, entre otros., se busca un conocimiento básico sobre términos económicos y el funcionamiento general del mismo.

## **Preguntas relevantes**

1. ¿Como es influenciada la inflación por los cambios de entidades gubernamentales?
2. ¿Cómo afectaron los "cepos al dólar" a la brecha entre las cotizaciones del dólar oficial y el mercado libre?
3. ¿Que otra relación existen entre las variables monetarias, tales como las reservas, en relación a la inflación?
4. ¿Teniendo en cuenta una inflación que sufre fluctuaciones considerables, es posible generar un modelo de machine Learning que pueda predecir en base a las variables otorgadas por la API del BCRA u otra información pública? ¿Hasta que punto es fiable?

## **Puntos a tener en cuenta**

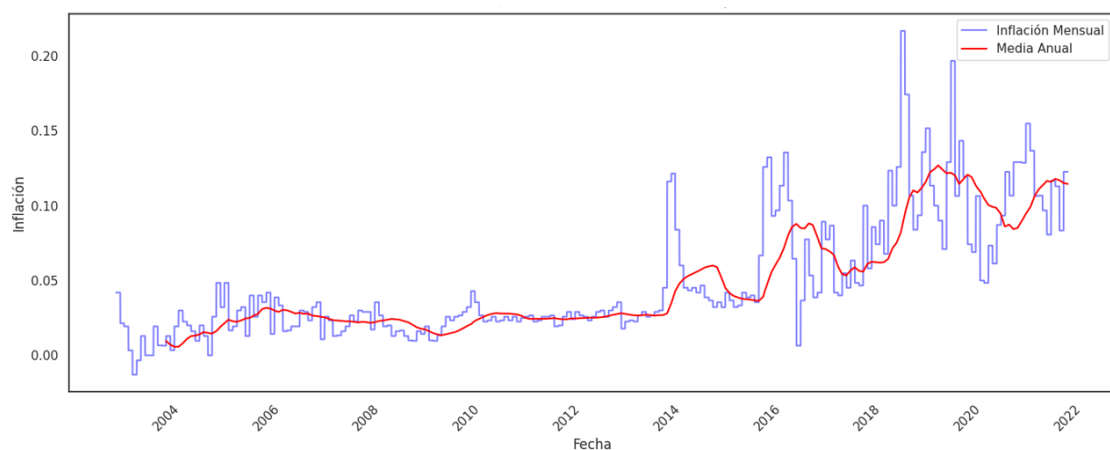
Para realizar este análisis tendremos en cuenta que:

- Se hará en base de variables que perduran a lo largo del tiempo, y otras herramientas que se usad ad hoc. Pueden permanecer (o no), tales como la leliq.
- Teniendo en cuenta que la variable relevante de análisis es la inflación, se tendrá en cuenta la leyenda que advierte manipulación de datos de inflación en la API del bcra, el cual menciona: **"Advertencia: Por favor considere que durante el periodo 2007-2016 el INDEC (Organismo que compila las estadísticas oficiales) fue intervenido y es generalmente aceptado que los datos publicados durante ese periodo sobre inflación no reflejaban correctamente la realidad. Puede leer más aquí: Nota sobre los datos de inflación.**
- Teniendo en cuenta la larga trayectoria de inflación argentina, en diferentes años se utilizan herramientas económicas (tales como las leliq). Se eligen variables que sean constantes a lo largo de los años, de tal manera que se limita a entender las anteriores herramientas con el cambio de entidades gubernamentales, y, hechos históricos económicos (tales como el cepo al dolar). Se utilizaron datos de periodo 01-ene-2003 hasta 31-dic-2021, y las ocupan las siguientes variables: milestones (sucesos coyunturales), usd, usd oficial, base monetaria, depósitos, cuentas corrientes, cajas de ahorro, índice merval, reservas, prestamos, lebacs, indice cer e indice badlar.

- La adquisición de datos es directamente llamada por la API de BCRA.
- Para analizar la inflación se debió utilizar variables de medición diaria. De manera que se haga un análisis que contemple la evolución de la inflación contrastado con estas variables de evolución diaria, la inflación se dividió durante la cantidad de días por mes (según corresponda), y al finalizar de generar el modelo, se volvió a unificar la inflación diaria en mensual, para poder contrastar los resultados.

## Analisis exploratorio

**Análisis sobre la variable de inflación real, comparada por su media anual:** Se analizan grandes fluctuaciones de inflación, que van en crecimiento. Se ameseta levemente entre 2010 y 2014, para luego disparar al alza. La media anual (sobre meses) suaviza las fluctuaciones inflacionarias, y a lo largo del tiempo se nota una consistencia en la inflación. Las intersecciones nos indican un cambio de tendencia, y hasta 2014, pareciera mantener una tendencia estacionaria. Un análisis técnico podría ver un piso (o mesetas) en 4 puntos que conforman un piso, y un techo con 3 picos, con una inflación en tendencia alcista. Aunque se haga un análisis técnico de la inflación, esta es relativa a las herramientas macroeconómicas utilizadas en el país, y la expectativa de precios internacionales.



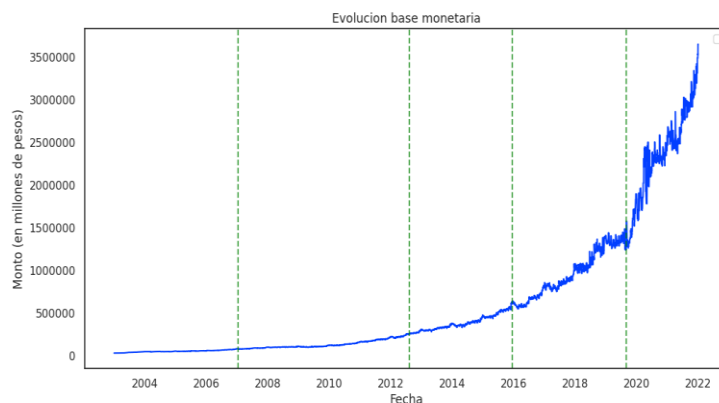
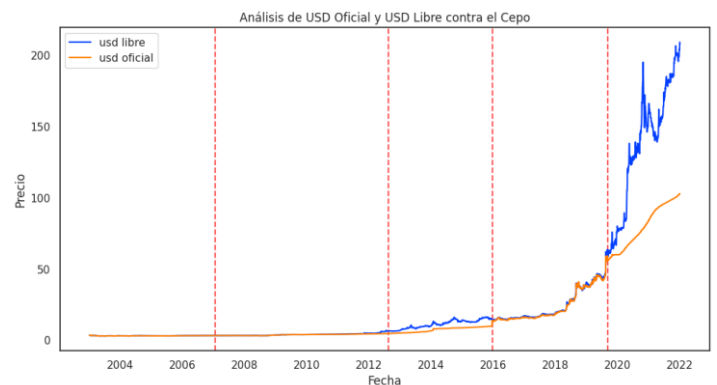
- Coeficiente de variación inflación 83.65294495603581

En el análisis de coeficiente de variación, nos da como resultado un 83.65%. Esto nos sugiere que hay una considerable variabilidad en la inflación mensual (pasado a diario) en comparación con la media. Nuestro error estándar es bastante bajo (0.00049), lo que nos dice que la media de los datos es precisa de la expectativa de la inflación. Esto también puede ser porque al pasar la inflación a diario, existe una seguidilla de fechas, relativa a la cantidad de días por mes, el cual no sufre variaciones.

- Coeficiente de asimetría: 1.3711854921503053
- Kurtosis: 1.405572783082044

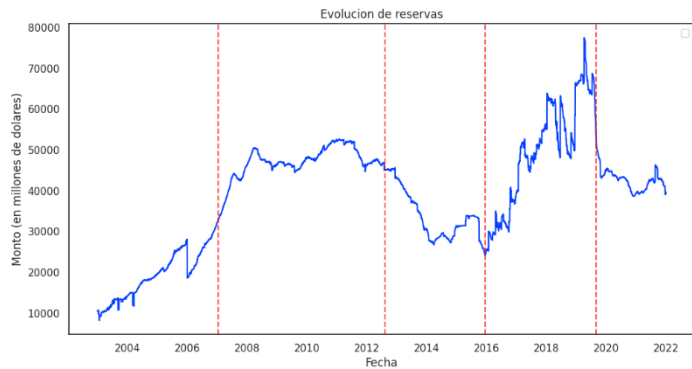
Los datos de inflación diaria tienen una distribución que es sesgada a la derecha (asimetría positiva de 1.37). Nos muestra que hay una cola más larga hacia los valores más altos. La kurtosis es positiva (1.41), lo cual indica que la distribución es leptocúrtica. Esto significa que la distribución tiene un pico más agudo y colas más pesadas en comparación con la distribución normal.

**Análisis de dólar oficial vs paralelo:** La colocación del cepo abre una brecha donde se plasma la diferencia entre el usd oficial y el usd libre. Vemos como ambas siguen una tendencia similar hasta las políticas de cepo implementadas en todas las líneas punteadas verticales. Se observa que el dolar oficial mantiene una tendencia sin sobresaltos, y en fechas puntuales rompe la brecha con el dolar libre alcanzando su cotización. Por otro lado el dolar libre fluctua de manera mas caotica, comportandose como un activo. El despegue del usd libre es verídicamente plausible en la colocación y quita de los cepos.



**Evolución de la base monetaria:** La base monetaria tiene a un aumento constante generado, probablemente, por la inflación. Podemos observar que en el ultimo cepo colocado a fines del año 2019 la base monetaria tiene un aumento exponencial.

**Evolución de las reservas:** Entre el año 2012 y 2016, y entre 2019 y 2022, fechas en donde localizamos el cepo, las reservas internacionales disminuyeron, lo cual amerita un análisis más profundo de las causas. Esto también podría ser el efecto del porque se impuso el cepo y no las



causas. En el año 2019 el cepo se impuso por el aumento de fuga de capitales, lo cual se ve que el cepo es impuesto en el medio salida de capitales, y luego del cepo las reservas permanecen sin grandes fluctuaciones a partir del año 2020.

## Modelado de Machine Learning

La selección de variables más relevantes utilizando el modelado de machine Learning nos da como resultado las siguientes variables: cajas de ahorro, índice merval, préstamos, cer, y todas las variables de hechos históricos ocurridos dentro del periodo (cambio de presidente, ministro de finanzas o economía, colocación o quita de cepo, etc.)

### Variable dependiente: inflación

- R-squared ( $r^2$ ): Es de 0.757, lo que significa que aproximadamente el 75.7% de la variabilidad de 'inflacion\_diaria' se puede explicar por el modelo. Es un valor relativamente alto.
- F-statistic: Un valor de 1348 para la estadística F nos sugiere que el modelo es estadísticamente significativo (la estadística F sugiere que al menos una de las variables independientes en el modelo está relacionada significativamente con la variable dependiente)
- Prob (F-statistic) - 0: El valor de probabilidad cerca de 0 indica que es muy improbable que la variabilidad observada en la respuesta se deba al azar.
- AIC y BIC: Valores AIC y BIC más bajos son generalmente preferidos. Un AIC y BIC negativo (-3.45, -3.44) puede considerarse como modelo indicado, ya que los datos no son innecesariamente complejos entre sí.

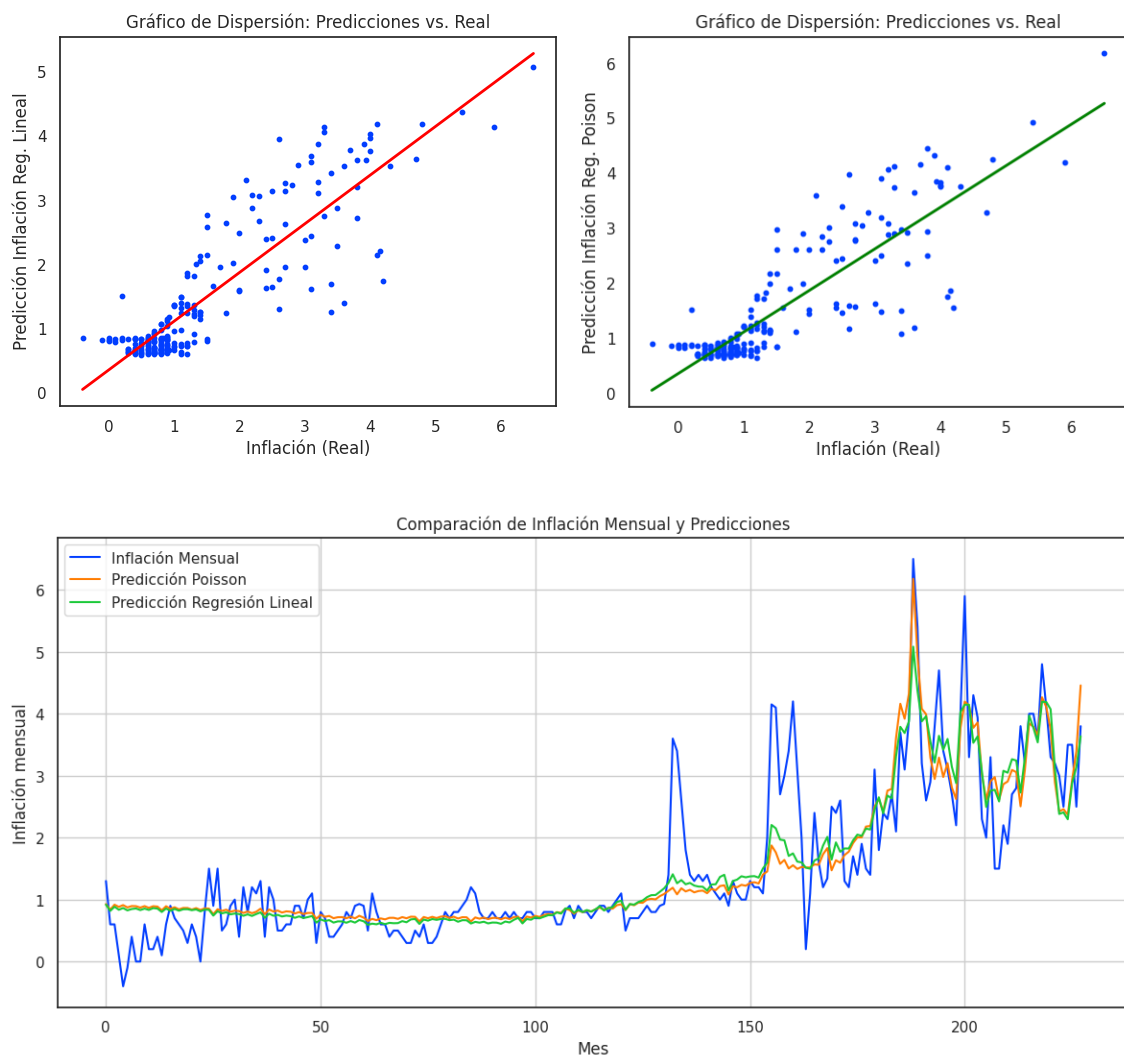
## Variables independientes

- Los coeficientes (coef) representan el cambio esperado en la variable dependiente por cada una unidad de cambio en la variable independiente.
- std err es el error estándar de la estimación del coeficiente.

t es el valor de la estadística t, que se utiliza para probar si el coeficiente es significativamente diferente de cero.

- Los intervalos [0.025 0.975] son los intervalos de confianza del 95% para los coeficientes, lo que significa que hay un 95% de certeza de que el verdadero valor del coeficiente se encuentra dentro de ese rango.

## Regresión lineal y regresión de Poisson



La Regresión Lineal parece realizar un mejor trabajo al predecir la inflación en comparación con la Regresión de Poisson, basándose en la concentración de puntos alrededor de la línea de ajuste en el primer gráfico.

Ambos modelos tienen dificultades para predecir los picos más altos de inflación, especialmente en contextos de alta volatilidad como el argentino. Aun así, los picos logran adaptarse a las subidas repentinas en donde existen subidas repentinas de la inflación real, lo cual podemos pensar que existe una lectura en contextos históricos o aplicación de herramientas económicas coyunturales (tales como el cepo al dólar, o cambio de entidad gubernamental como ministros de economía o presidentes).

Las predicciones parecen seguir la tendencia de la inflación real, lo que indica que los modelos son potencialmente útiles para comprender la dirección general de la inflación, pero pueden no ser tan confiables para predecir cambios extremos o repentinos.

## **Optimización de hiperparametros**

- **'fit\_intercept': True**

La línea de regresión no necesariamente comenzará en el origen (0,0). En este caso, la línea de mejor ajuste que el modelo usa para hacer predicciones no pasa necesariamente por el origen, lo cual es adecuado para los datos.

- **Mejor puntaje (MSE negativo): -0.0003529396135410146**

El MSE es una "función de pérdida" que es mejor cuanto más baja sea. Un MSE "más alto" (menos negativo) indica un modelo mejor. Un MSE de -0.0003529396135410146, cercano a cero (recordando que es un valor negativo), sugiere que el modelo tiene un bajo error cuadrático medio, lo que indica un buen ajuste del modelo.

- **Puntaje en el conjunto de prueba ( $R^2$ ): 0.7929734261001851**

Esto nos indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Entre más cercano sea a 1 el  $r^2$ , significa que el modelo hace un buen trabajo al predecir la inflación diaria. En este caso, el  $r^2$  es de 0.7929734261001851, el cual es bastante efectivo.



## **Conclusiones:**

Se logró desarrollar un análisis predictivo de la estabilidad financiera en Argentina, centrándose en la inflación y su comportamiento a través de diversos escenarios económicos y políticos. Con la inflación como principal variable de interés, se emplearon datos del BCRA para estudiar la interacción entre varias variables macroeconómicas y eventos políticos clave.

El estudio nos revela que la inflación argentina no solo es volátil, sino que también responde de manera compleja a las políticas monetarias, las fluctuaciones del mercado de divisas y los cambios administrativos gubernamentales. Se destacó que, aunque existen períodos de aparente estabilidad, como se observa entre 2010 y 2014, la tendencia general ha sido de un aumento constante, con la media anual mitigando las fluctuaciones mensuales, pero manteniendo una tendencia al alza.

El modelo de machine learning diseñado y validado en este estudio demostró ser eficaz, con un  $R^2$  significativo y un MSE negativo cercano a cero, lo que indica que puede predecir con precisión la variabilidad de la inflación. La regresión lineal ha demostrado ser más efectiva que la regresión de Poisson, especialmente en la adaptación a eventos económicos específicos y cambios de política. Se podrían seguir trabajando contextos puntuales, o hechos (tales como la deflación generada en un mes particular), pero en términos generales el modelo demuestra ser sostenible y trasladable en el tiempo, siempre y cuando se utilicen variables constantes en los periodos.