



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO



FACULTAD DE
**CIENCIAS
ECONÓMICAS**

LICENCIATURA EN ECONOMÍA

Modelos predictivos de años recesivos en Argentina utilizando algoritmos de Machine Learning.

Tesis final de grado

Autor: Rodrigo Dichara

Registro N° 33.116

rodrigo.dichara@fce.uncu.edu.ar

Director: Juan Víctor Diblasi

juan.diblasi@fce.uncu.edu.ar

Mendoza, 2025



Resumen

Las recesiones económicas tienen efectos muy nocivos tanto para la economía como para la sociedad. Es por ello que contar con datos e información oportuna acerca de las condiciones de la economía es muy importante para la toma de decisiones de los distintos agentes económicos y para el diseño de la política económica por parte del policymaker.

El Producto Interno Bruto (PIB) es el principal indicador económico sobre la actividad de una economía. Sin embargo, este indicador cuenta con la gran desventaja de que se publica en forma trimestral y con un retraso mayor a 2 meses, por lo que los datos oficiales del PIB no son nunca oportunos para la toma de decisiones.

El presente estudio desarrolla modelos de machine learning para predecir de manera anticipada si un año terminará siendo recesivo o no. Los modelos reciben como inputs datos de variables relacionadas a la política monetaria y cambiaria: agregado monetario M3, base monetaria, inflación mensual e interanual, tasas de interés de EE.UU. y doméstica, tipo de cambio oficial y reservas internacionales. Los modelos resultantes no sólo podrían ayudar al policymaker a cambiar rápidamente el rumbo de sus políticas ante una alerta de recesión para evitar caer en la misma y evitar así sus efectos nocivos, sino también podrían ayudar a las empresas y demás agentes económicos al brindarles información oportuna sobre el estado de la actividad económica en los próximos meses.

Palabras claves: Economía, Recesión, Machine Learning, Argentina



Índice

1. Introducción	4
2. Antecedentes	9
3. Metodología	12
3.1. Algoritmos con los que se construirán los modelos.	13
3.2. Variables que se utilizarán.....	16
3.3. Estandarización de los datos	18
3.4. Entrenamiento y metodologías de resampling.....	19
3.4.1. Cross-Validation	20
3.4.2. Bootstrapping	20
4. Métricas de evaluación de los modelos	21
4.1. Matriz de confusión	21
4.1.1. Exactitud – Accuracy	22
4.1.2. Precisión – Precision.....	22
4.1.3. La curva ROC.....	22
5. Resultados obtenidos y análisis de métricas	23
5.1. Modelo con algoritmo SVM y con resampling Cross-Validation	23
5.2. Modelo con algoritmo SVM y con resampling Bootstrapping	25
5.3. Modelo con algoritmo k-NN y con resampling Cross-Validation.....	26
5.4. Modelo con algoritmo k-NN y con resampling Bootstrapping.....	28
5.5. Modelo con algoritmo Random Forest y con resampling Cross-Validation	30
5.6. Modelo con algoritmo Random Forest y con resampling Bootstrapping.....	32
6. Incidencia de las variables en las recesiones	34
6.1. Análisis de componentes principales con todos los años recesivos	35
6.2. Análisis de componentes principales excluyendo los años recesivos en los que hay valores atípicos.....	39
7. Predicciones para el presente año 2025	43
8. Conclusiones	44
9. Referencias bibliográficas	48



1. Introducción

Argentina es un país que presenta una historia económica muy interesante: múltiples períodos de muy alta inflación; varios defaults; muchos cambios de moneda; múltiples intentos de planes de estabilización; muchos cambios bruscos en la ideología del partido gobernante y, como consecuencia, cambios bruscos de las políticas económicas ejecutadas; períodos hiperinflacionarios; múltiples períodos recesivos y de crisis, etc.

Particularmente este trabajo de investigación tendrá como columna vertebral los años recesivos ocurridos en Argentina desde principios de la década de 1960 hasta el año finalizado más reciente, 2024.

La gran cantidad de años recesivos que ha tenido Argentina no solo es llamativa para los economistas de este país, sino también para muchas figuras en el mundo. Simon Kuznets, ganador del premio nobel de economía de 1971, dividía a los países en cuatro "clases": los desarrollados, los subdesarrollados, Japón y la Argentina. Con esta clasificación buscaba enfatizar la atipicidad entre Japón y Argentina. Japón era un país sin recursos naturales pero con un alto crecimiento económico, mientras que Argentina tenía abundantes recursos naturales, pero no lograba un crecimiento económico sostenido a lo largo del tiempo ya que los años recesivos eran muy frecuentes.

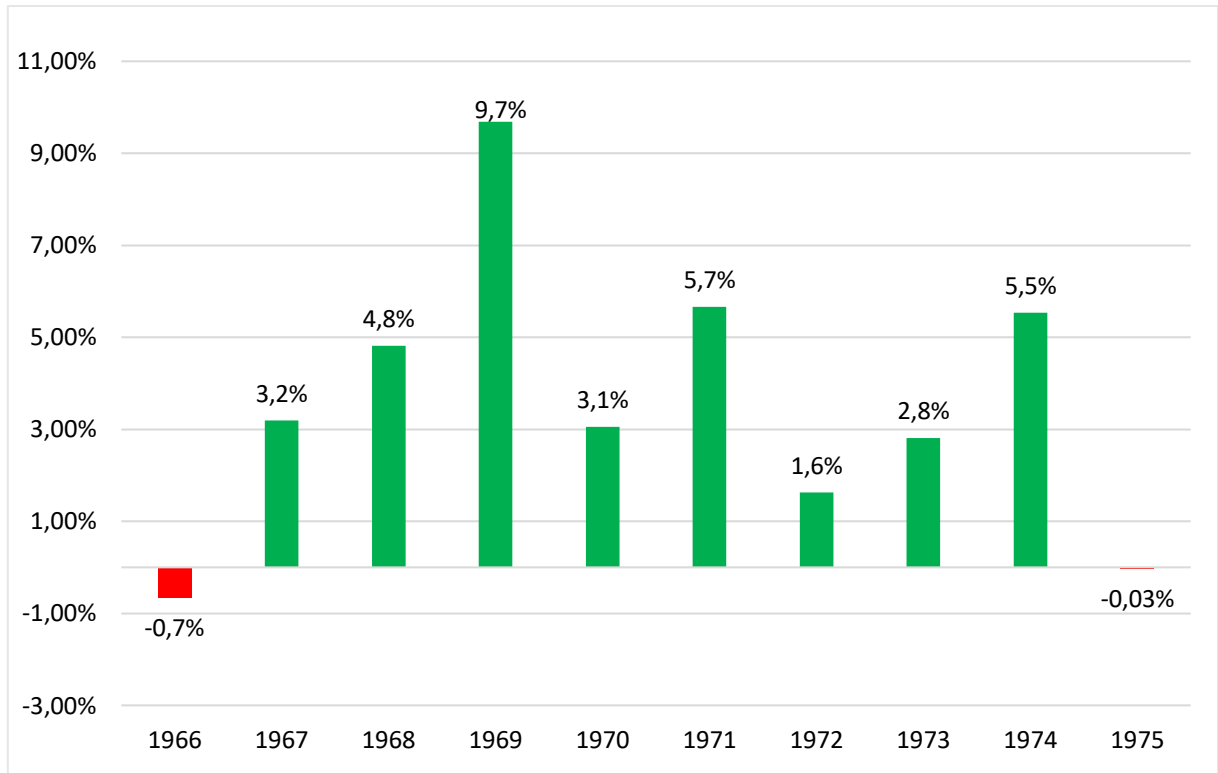
Según el estudio "Conflicto distributivo y crecimiento en Argentina" realizado en 2019 por el economista Martín Rapetti, quien es Director de Desarrollo Económico del Centro de Implementación de Políticas Públicas para la Equidad y el Crecimiento (CIPPEC) y profesor de la Universidad de Buenos Aires, Argentina figuraba, en la comparación a nivel internacional para los años que van desde 1960 a 2019, como el país con mayor número de años con crecimiento negativo, seguido por la República del Congo. A su vez, Argentina formaba parte del 25% de países que menos crecieron durante esos casi 60 años.

Comparando Argentina con los demás países de la región, Rapetti encontraba un panorama similar: para el período que va entre 1960 y 2019, la tasa de crecimiento de la Argentina fue de sólo 2,3% anual, siendo junto con Uruguay la más baja de Latinoamérica. Además, Argentina fue el país de la región que más recesiones tuvo: un total de 14 entre 1960-2018 (en segundo lugar se ubicaba Venezuela con 8 recesiones).

Analizando desde el año 1963 hasta la actualidad, que es la ventana temporal en la cual se desarrollará el presente trabajo, la economía argentina logró crecer por más de cinco años consecutivos sólo en dos períodos: desde 1967 a 1974, y desde 2003 a 2008.

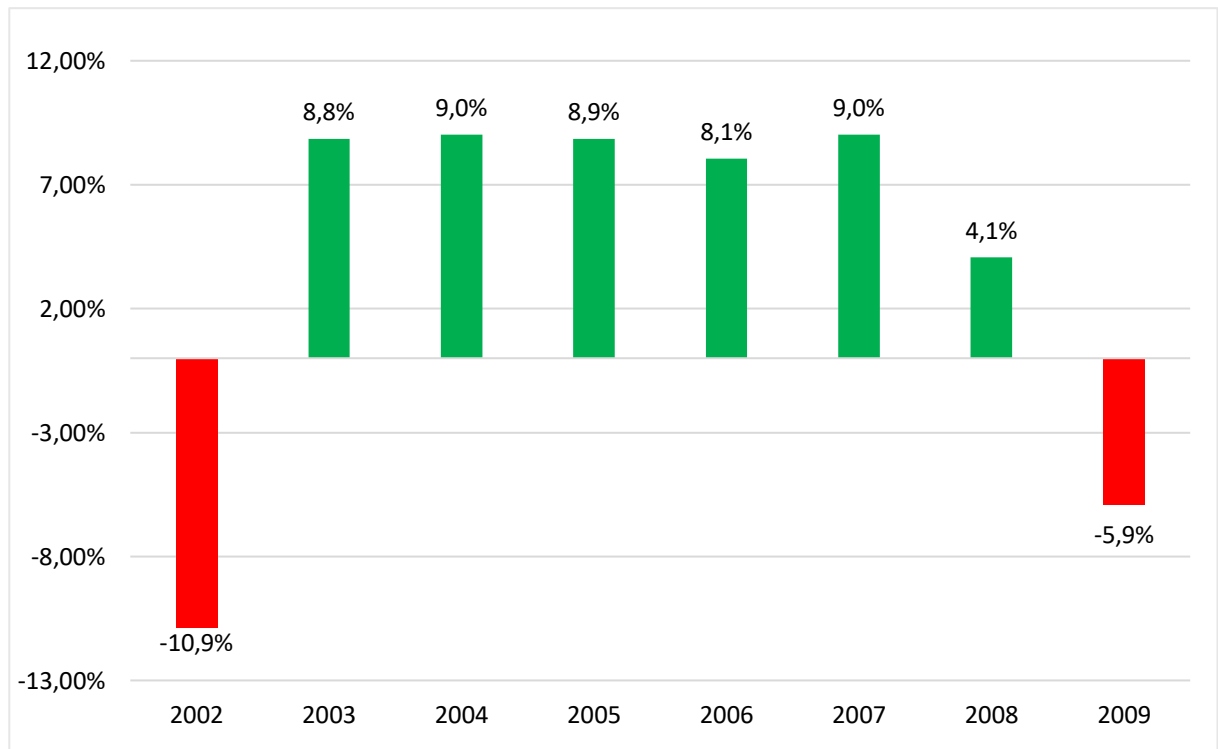


Gráfico 1: Tasa de crecimiento anual del PIB real de Argentina desde 1966 a 1975.



Fuente: Elaboración propia en base a datos extraídos del Banco Mundial, mayo de 2025.

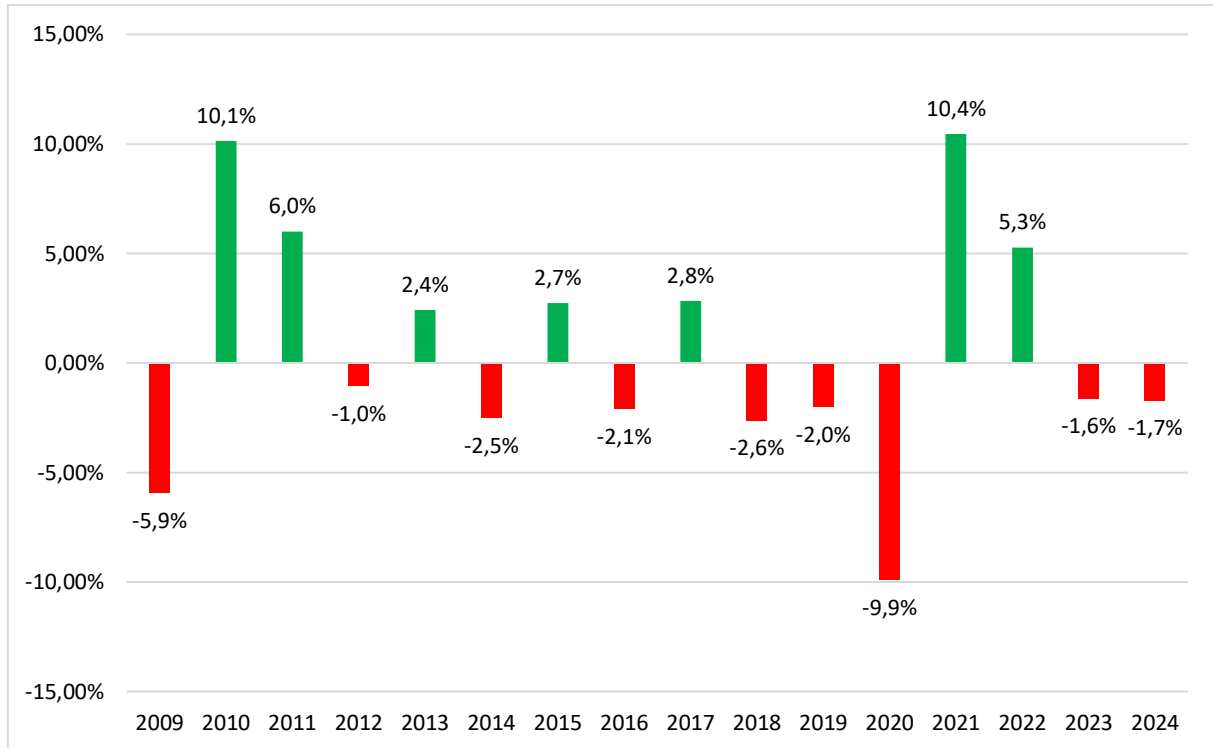
Gráfico 2: Tasa de crecimiento anual del PIB real de Argentina desde 2002 a 2009.



Fuente: Elaboración propia en base a datos extraídos del Banco Mundial, mayo de 2025.

Entre 2009 y 2024, que son los 16 años que le siguieron al último período de crecimiento económico alto y sostenido de la economía argentina, solo hubo 7 años de expansión económica, mientras que los 9 años restantes fueron recesivos.

Gráfico 3: Tasa de crecimiento anual del PIB real de Argentina desde 2009 a 2024.



Fuente: Elaboración propia en base a datos extraídos del Banco Mundial, mayo de 2025.

Está claro que Argentina presenta una gran inestabilidad económica y tiene muy presente el mal de la recesión. Cada vez son más frecuentes los años recesivos, llegando al punto actual donde en la última década y media se cuenta con más años recesivos que de crecimiento económico.

Este historial de Argentina ha sido muy nocivo para el país ya que los períodos de recesión son muy negativos tanto para la economía como para la sociedad. Una recesión no permite que la economía se siga desarrollando, modernizando, ampliando y que sus habitantes continúen mejorando su nivel de vida; por el contrario, en los tiempos de recesión muchos individuos pierden su trabajo, muchas empresas venden menos y, en caso de que la recesión sea profunda y persistente, se genera incluso destrucción de capital en la economía porque varias empresas y fábricas cierran, muchos proyectos de inversión quedan a mitad de realización y nunca se concretan, y la sociedad termina perdiendo recursos. Todo esto genera que el nivel de vida de los individuos empeore, es decir, la sociedad en su conjunto termina siendo más pobre que antes del inicio de la recesión.



Como consecuencia de los importantes efectos de las recesiones sobre la vida de los individuos, contar con datos e información oportuna acerca de las condiciones de la economía es muy importante para la toma de decisiones de los distintos agentes económicos y para el diseño de la política económica por parte del policymaker. El Producto Interno Bruto (PIB) es el principal indicador económico sobre la actividad de una economía. Sin embargo, este indicador cuenta con la gran desventaja de que se publica en forma trimestral y con un retraso mayor a 2 meses, por lo que los datos oficiales del PIB no son nunca oportunos para la toma de decisiones. Existen otros indicadores del nivel de actividad que cuentan con una frecuencia mayor, como por ejemplo el Estimador Mensual de Actividad Económica (EMAE). Sin embargo, similar a lo que ocurre con el PIB, los demás indicadores oficiales de actividad se publican también con un retraso mayor a un mes, por lo que tampoco son oportunos.

Debido a los problemas con los retardos en la publicación de los datos oficiales del gobierno, muchos individuos y organizaciones deciden realizar de manera privada modelos predictivos que les permitan tener información sobre el nivel de actividad en el momento adecuado para la toma de decisiones. Sin embargo, existe un conjunto muy reducido de publicaciones de libre acceso donde se construyen modelos predictivos de la tasa de crecimiento del PIB de Argentina. Si bien en la mayoría de los trabajos encontrados no se presenta explícitamente el objetivo o la intención de predecir recesiones en Argentina, el PIB es el indicador que la mayoría de los profesionales utiliza para determinar si hay recesión o no, por lo que un modelo que predice la tasa de crecimiento trimestral del PIB podría ser utilizado también para predecir recesiones.

Teniendo en cuenta la relevancia de las recesiones tanto para la economía del país como para la sociedad, es menester que se continúen desarrollando muchos más modelos predictivos para poder ayudar/advertir al policymaker acerca del riesgo de recesión y que pueda cambiar a tiempo el rumbo de sus políticas porque como se dijo anteriormente, las recesiones económicas pueden ser demasiado nocivas para la sociedad y para la economía de un país.

En el presente trabajo se busca construir un modelo que permita predecir, antes de llegar a mitad de año y antes de contar con datos oficiales del PIB sobre el año en curso, si la actividad económica en el presente año caerá con respecto al año anterior o no. Detallando aún más, el objetivo principal de este trabajo consiste en determinar cuál de los 3 algoritmos -Support Vector Machines (SVM), k-Nearest Neighbor (kNN) y Random Forest- permite construir un modelo predictivo con un sólido desempeño en la detección temprana de años recesivos en Argentina utilizando exclusivamente las tasas de variaciones de variables económicas nominales. Al hablar de “sólido desempeño” se hace referencia a un umbral mínimo del 85%



en la cantidad total de predicciones correctas realizadas por el modelo y del 80% en la cantidad de años recesivos detectados. Es decir, para que se considere que el modelo tiene un “sólido desempeño”, la métrica “accuracy” del mismo debe ser, como mínimo, del 85% y la métrica “precisión” debe ser, como mínimo, del 80%.

El análisis se circunscribe territorialmente a Argentina y abarca los años dentro de la ventana temporal que va desde 1963 hasta 2024.

Si se logra cumplir con el objetivo del trabajo, entonces el modelo resultante no sólo podría ayudar al policymaker a cambiar rápidamente el rumbo de sus políticas ante una alerta de recesión para evitar caer en la misma y evitar así también los efectos nocivos que tiene una recesión sobre la sociedad, sino también podría ayudar a las empresas y demás agentes económicos. Por ejemplo, saber de manera anticipada que te encuentras en un año en el cual la actividad económica será menor a la del año anterior puede permitirte que realices una mejor planificación empresarial y que no encares proyectos o actividades que, por la situación futura, no terminarán siendo rentables.

Adicionalmente, este trabajo puede aportar un nuevo marco para quienes decidan realizar modelos predictivos de recesiones para Argentina en el futuro ya que en este trabajo se utilizan por primera vez únicamente las variaciones de variables económicas nominales y se implementan, por primera vez también, modelos de clasificación de recesiones con los algoritmos SVM, kNN y Random Forest.

La estructura del trabajo es la siguiente: en el próximo apartado se ofrece una revisión de los antecedentes relacionados a los temas aquí tratados. En el tercer apartado se explican aspectos metodológicos tales como los algoritmos que servirán de base para los modelos, las variables que se utilizarán, las metodologías para entrenar los modelos, etc. Luego, en el cuarto apartado se exponen y explican las métricas que servirán para evaluar y comparar los modelos. En el quinto apartado se exponen los resultados que arrojó cada uno de los modelos y se evalúan las métricas en cada uno de ellos. En el sexto apartado se analiza, mediante análisis de componentes principales, la incidencia de las variables en los años recesivos bajo estudio. En el séptimo apartado se realizan predicciones de recesión para el presente año con los modelos desarrollados en el trabajo. Finalmente, el último apartado ofrece las conclusiones e implicancias socioeconómicas del presente trabajo.



2. Antecedentes

Existe un conjunto muy reducido de publicaciones donde se construyen modelos predictivos de la tasa de crecimiento del PIB de Argentina. A pesar de que en la mayoría de los trabajos encontrados no se presenta explícitamente el objetivo o la intención de predecir recesiones en Argentina, el PIB es el indicador que la mayoría de los profesionales utiliza para determinar si hay recesión o no, por lo que un modelo que predice la tasa de crecimiento trimestral del PIB podría ser utilizado también para predecir recesiones.

D'Amato et al. (2016) señalan que en general los indicadores utilizados para la predicción del PIB son de dos clases: 1) Indicadores hard; e 2) Indicadores soft. En su trabajo emplean un conjunto de 37 indicadores diarios y mensuales del ciclo económico, incluyendo desde datos financieros a recaudación impositiva, información desagregada sobre la producción industrial, encuestas de confianza del consumidor y ventas de automóviles, etc. Mediante esta combinación de indicadores soft y hard, desarrollan un Nowcast para el pronóstico del crecimiento trimestral del PIB en Argentina utilizando dos enfoques: ecuaciones puente y modelo de factores.

Emplean una muestra de estimación que comprende el período que va desde el primer trimestre de 1993 hasta el cuarto trimestre del 2007. Por otro lado, el período de pronóstico va desde el primer trimestre del 2008 hasta el primer trimestre del 2014.

Obtienen que tanto las ecuaciones puente como el modelo de factores superan en capacidad predictiva a un modelo autorregresivo de primer orden (AR(1)). Luego, en una comparación entre los dos enfoques estudiados, obtienen que el Nowcast basado en un modelo de factores supera al de ecuaciones puente. Finalmente, utilizando el test de Giacomini y White (2004) confirman que esas diferencias en la capacidad predictiva de los distintos modelos son estadísticamente significativas.

Martín Llada (2022) construye un panel de datos con 16 indicadores económicos mensuales, algunos hard y otros soft, que cubren desde el primer trimestre de 1998 hasta el segundo trimestre del 2021, y desarrolla luego un modelo con el cual realiza predicciones de la tasa de crecimiento del PIB trimestral de Argentina. Para ello, utiliza un modelo de factores dinámicos. Separa los datos mensuales desde 1998 hasta 2011 como subconjunto para entrenar el modelo, y luego los datos desde 2012 hasta el séptimo mes del 2021 como subconjunto para evaluar el modelo.



El autor evidencia que la capacidad predictiva del modelo de factores dinámicos supera a los pronósticos de los modelos autorregresivos (AR) y a los de un modelo de crecimiento naive estimado recursivamente.

Adicionalmente, para obtener un modelo aún más completo y robusto, el autor utiliza técnicas de procesamiento de lenguaje natural, similar a los realizado por Aguilar et. al. (2021) y Baker et. al. (2016), para construir un indicador del tono sobre el ciclo económico de Argentina. Para construir dicho indicador explota los artículos periodísticos de la sección de economía del diario *Ámbito Financiero*, *La Nación* y *Página 12*.

Blanco et al. (2017) realizan también pronósticos del crecimiento trimestral del PIB de Argentina con el objetivo principal de evaluar y comparar la capacidad predictiva de modelos factoriales y modelos FAVAR (modelo VAR ampliado con factores). También buscan evaluar la diferencia entre usar, por un lado, un gran conjunto de datos que incluye 112 indicadores hard y soft, y por otro, un subconjunto más restringido de 30 indicadores que tengan la mayor correlación contemporánea con el crecimiento del PIB.

Los autores utilizan datos desde el primer trimestre del 2006 hasta el primer trimestre de 2017. A este conjunto de datos lo dividen en 2 subconjuntos: los datos desde 2006 hasta 2010 los utilizan para entrenar el modelo, mientras que los datos desde 2011 en adelante son utilizados para evaluar el modelo. Evidencian que la combinación de modelos de pronóstico supera la performance de modelos individuales, mientras que no encuentran evidencia estadísticamente significativa respecto a que un modelo individual tenga una performance predictiva superior a la de otro modelo.

Camacho et al. (2013) estiman también un modelo de factores dinámico utilizando únicamente 5 indicadores económicos para el período entre enero de 1993 y agosto de 2012. Separa los datos desde 1993 a 2001 como subconjunto para entrenar el modelo, y luego los datos desde 2002 en adelante como subconjunto para evaluar el modelo.

En cuanto al amplio conjunto de indicadores que podrían utilizar en el análisis, los autores eligen aquellos que satisfacen las siguientes cuatro propiedades:

- Deben tener una alta correlación estadística con la tasa de crecimiento del PIB real.
- Si son indicadores con frecuencia trimestral, entonces sus datos deben publicarse antes de que el dato del PIB esté disponible en el trimestre respectivo.
- Deben ser relevantes en el modelo tanto desde el punto de vista teórico como estadístico.



- Deben estar disponibles en al menos un tercio de la muestra.

En base a las propiedades anteriores, los indicadores que eligen para la construcción del modelo son:

- PIB real con frecuencia trimestral
- Índice de Producción Industrial con frecuencia mensual
- Empleo total de la población urbana con frecuencia trimestral
- Indicador Sintético de la Actividad de la Construcción con frecuencia mensual
- Ventas total reales de supermercados con frecuencia mensual

A través de los factores comunes del modelo, los autores pueden estimar las probabilidades de estar en recesión. Durante los períodos de la historia argentina clasificados como recesiones por la literatura de Jorrat y Cerro (2000) y Jorrat (2005,2012), las probabilidades estimadas de estar en recesión que brinda el modelo suelen ser altas. Por ende, los autores logran conseguir un modelo de factores dinámicos que tiene un sólido desempeño como indicador del ciclo económico para Argentina.

Por último, los autores demuestran que la capacidad predictiva del modelo de factores dinámico supera los pronósticos de los modelos bases: un modelo basado en Random Walk y otro modelo autorregresivo de segundo orden (AR(2)).



3. Metodología

Para el desarrollo práctico del trabajo de investigación se construyen distintos modelos de aprendizaje supervisado con los datos de 8 variables explicativas observadas a lo largo de los años 1963-2024. Adicionalmente, se crea una variable dicotómica llamada “Recesión”, la cual tomará el valor “Sí” si el año ha sido recesivo y tomará el valor “No” si el año no ha sido recesivo. Esta variable dicotómica es la variable objetivo del modelo, la cual se intentará predecir utilizando los datos de las demás variables económicas explicativas.

Se trabaja con datos secundarios, los cuales son extraídos de la página oficial del Banco Central de la República Argentina (BCRA) y de la Federal Reserve Economic Data (FRED) del Federal Reserve Bank of St. Louis.

Las páginas nombradas anteriormente ofrecen los datos históricos con distintas periodicidades y cubren una ventana temporal de muchos años, por lo que se pueden encontrar datos desde el año 1960 en adelante sin problema. No hay datos faltantes ni errores en las series, lo cual facilita mucho el trabajo ya que no genera la necesidad de realizar un preprocesamiento de datos avanzado, estar eligiendo técnicas para imputación de datos faltantes, etc.

Dado que en los años que abarca el trabajo nuestro país tuvo una serie de cambios de signos monetarios, la única transformación que hubo que realizar fue sobre aquellas variables expresadas en términos monetarios (tipo de cambio, agregado monetario M3 y base monetaria). Había que convertir la moneda en la que estaban expresados los datos más antiguos a la especie monetaria actual. Sin embargo esto resultó sencillo ya que el BCRA ofrece los datos acompañados por una columna llamada “Equivalencia de la unidad monetaria original con la actual”, y explica sencillamente cómo utilizar esa columna para realizar la conversión.

El presente trabajo se encuadra en un enfoque cuantitativo, ya que su objetivo es estimar las magnitudes de los fenómenos y probar ciertas hipótesis (Sampieri, 2018). Más precisamente, el trabajo busca construir un modelo predictivo con un sólido desempeño (accuracy $\geq 85\%$ y precisión $\geq 80\%$) en la detección temprana de años recesivos en Argentina.

Teniendo en cuenta los tipos de investigaciones que describe Ynoub (2011), este trabajo forma parte de una investigación explicativa ya que como la autora comenta, “en este tipo de investigaciones interesa probar si el comportamiento de una variable (o varias) puede funcionar como factor o causa explicativa del comportamiento de otra (u otras) variable.” (p. 85). Además, el trabajo cuenta con un diseño multivariado (8 variables explicativas) y con observaciones longitudinales (años 1963 a 2024).



3.1. Algoritmos con los que se construirán los modelos.

El aprendizaje automático, más conocido por su nombre en inglés como “machine learning”, es una disciplina en la cual se utilizan distintos algoritmos para construir modelos que aprendan de forma autónoma en base al análisis de grandes cantidades de datos, con el objetivo final de identificar patrones ocultos en los datos y/o realizar predicciones.

Para realizar predicciones existen modelos de regresión y modelos de clasificación. Los modelos de regresión se utilizan cuando la variable dependiente es numérica y, por ende, se quieren predecir valores continuos tales como precios, ingresos, probabilidades, etc. Los modelos de clasificación se utilizan cuando la variable dependiente es categórica y, por ende, se quiere predecir a qué clase corresponde una nueva observación de la variable dependiente.

Siguiendo a Hastie et al. (2008) dentro del campo del machine learning podemos encontrar, a grandes rasgos, dos tipos distintos de modelos. Por un lado tenemos los modelos de aprendizaje supervisado en los cuales el objetivo es predecir el valor de una variable dependiente -la cual puede ser numérica o categórica-, basándose en los valores de una serie de variables explicativas o independientes. Se llama aprendizaje "supervisado" debido a que existen observaciones de la variable dependiente que permiten guiar el proceso de entrenamiento y aprendizaje de los modelos. Por otro lado tenemos los modelos de aprendizaje no supervisado donde no existe una variable dependiente y, por ende, el objetivo es encontrar asociaciones y patrones dentro del conjunto de datos de variables independientes.

Para el desarrollo práctico del presente trabajo se construirán distintos modelos de aprendizaje supervisado utilizando datos desde 1963 hasta 2024.

El primer modelo se construirá en base al algoritmo de aprendizaje supervisado Support Vector Machines (SVM), el cual fue desarrollado por Vladimir Vapnik y su equipo en la década de 1990. Tal como explican desde la empresa International Business Machines Corporation (IBM), este algoritmo se emplea comúnmente en problemas de clasificación binaria. Distingue entre dos clases encontrando el hiperplano óptimo que maximiza la distancia entre los puntos de datos más cercanos de clases opuestas. El número de variables independientes determina si el hiperplano es una línea en un espacio bidimensional o un plano en un espacio n-dimensional. Dado que se pueden encontrar múltiples hiperplanos para diferenciar clases, la maximización de la distancia entre puntos permite al algoritmo encontrar la mejor frontera de decisión entre clases. Esto, a su vez, le permite generalizar bien nuevos datos y hacer predicciones de clasificación precisas.



Debido a que en el presente trabajo habrá 8 variables independientes en la construcción de los modelos, tendremos un espacio multidimensional y por ello uno de los algoritmos elegidos es SVM. Según Vapnik et al. (1997), creadores del algoritmo SVM, “presentamos un nuevo método de estimación de funciones que es especialmente útil para resolver problemas multidimensionales” (p. 286).

El segundo modelo se construirá en base al sencillo y reconocido algoritmo de aprendizaje supervisado k-Nearest Neighbor (kNN). Aunque el algoritmo kNN se puede usar para problemas de regresión o clasificación, se suele utilizar mayoritariamente como un algoritmo de clasificación.

Para comprender mejor el algoritmo kNN, veamos sencillamente su funcionamiento:

1. Se calcula la distancia entre el dato que quiere clasificarse y todos los demás puntos.
2. Se ordenan las distancias de menor a mayor.
3. Se seleccionan los k puntos más cercanos al dato a clasificar (de acá proviene su nombre).
4. Se determina la etiqueta de clase del nuevo dato basándose en la etiqueta que aparece mayoritariamente en los k datos más cercanos.

El tercer modelo se construirá en base al algoritmo de aprendizaje supervisado Random Forest (RF). Este algoritmo se usa tanto para resolver problemas de clasificación como de regresión. El algoritmo Random Forest es un método de ensamble, lo cual quiere decir que combina las predicciones de varios modelos base para mejorar la precisión y la robustez de los resultados. Según la experiencia de Hastie et al. (2008) “los Random Forest funcionan notablemente bien, con muy pocos ajustes necesarios.” (p. 590). Los modelos bases en los cuales se basa el algoritmo RF son los árboles de decisión. En cuanto a la forma de sintetizar los diversos resultados de los distintos árboles de decisión para realizar la predicción, Hastie et al. (2008) explican que “cuando se utiliza para la clasificación, un Random Forest obtiene un voto de clase de cada árbol y luego clasifica mediante el voto mayoritario. Cuando se usa para regresión, las predicciones de cada árbol en un punto objetivo x simplemente se promedian.” (p. 592).

A partir del trabajo de Bates y Granger (1969), la literatura sobre pronósticos ha enfatizado que una combinación de diferentes pronósticos podría llevar a mejores desempeños en comparación a las predicciones de modelos individuales. Pasados 20 años, Clemen (1989) escribió “Los resultados han sido prácticamente unánimes: la combinación de múltiples pronósticos conduce



a una mayor precisión de los pronósticos. En muchos casos, se pueden lograr mejoras espectaculares en el rendimiento simplemente promediando los pronósticos.” (p. 559).

Dentro del campo de la Econometría son muy comunes modelos como las regresiones lineales simples y múltiples, modelos autorregresivos (AR), modelos de medias móviles y modelos Logit y Probit. Estos modelos clásicos son sencillos y ampliamente usados por su interpretabilidad, eficiencia computacional y fundamento estadístico. Además, son transparentes: tienen una forma funcional especificada y, por ende, permiten saber qué variable afecta qué resultado y cómo.

Es por esta razón que en algunos de los antecedentes analizados comparan los nuevos modelos que están construyendo con modelos autorregresivos (AR), éstos son buenos modelos como punto de partida. Los modelos autorregresivos predicen el siguiente valor de una variable utilizando valores pasados de esa misma variable, y se definen por el número de rezagos o momentos de tiempo anteriores que se consideran. Se basan en una regresión lineal donde la variable actual se relaciona con sus propios valores pasados.

La problemática que presentan los modelos clásicos de la Econometría es que para producir buenos resultados predictivos requieren del cumplimiento de una gran cantidad de supuestos, lo cual limita su aplicación.

Por su parte, los algoritmos a tratar en el presente trabajo (SVM, kNN y Random Forest) que forman parte del Machine Learning, no cuentan con una forma funcional específica, lo que presenta ventajas y desventajas al mismo tiempo. Las ventajas son: mayor flexibilidad, no teniendo que suponer relaciones lineales o específicas entre variables; capturar no linealidades y relaciones complejas de manera automática, lo cual es un caso muy recurrente en un análisis multivariado; no hay riesgo de seleccionar una forma funcional incorrecta (lineal, logarítmica, polinomial, etc.); el modelo aprende la estructura de los datos directamente; y, por último, existe un mejor rendimiento predictivo en casos de clasificación o regresión con grandes conjuntos de datos, superando a modelos paramétricos tradicionales. Por otro lado, las desventajas son: baja interpretabilidad por la falta de la forma funcional explícita; difíciles de validar teóricamente; y no se puede realizar inferencia estadística por la imposibilidad de obtener p-valores e intervalos de confianza.



3.2. Variables que se utilizarán.

El profesor Carlos Diego Martínez Cinca (2003) en un análisis sobre los aspectos filosóficos del liberalismo comenta que Milton Friedman y la Escuela Austríaca de Economía “coinciden sin embargo en una tesis capital: la planificación central de la moneda ha creado una enorme inestabilidad económica en el siglo XX.” (p. 5). En el presente trabajo de investigación se sigue a Milton Friedman y a los diversos autores de la Escuela Austríaca de Economía, y se construye un modelo predictivo de años recesivos para Argentina que tiene como variables explicativas únicamente variables económicas nominales. De esta manera se buscará también determinar si las variaciones de las variables económicas nominales son realmente las generadoras de la inestabilidad económica y nos permiten predecir años recesivos en Argentina.

Las variables que funcionan como inputs de los modelos de machine learning son las siguientes:

- Agregado monetario M3: Para obtener la observación anual se toma el promedio de los primeros 5 meses del año y luego se calcula la variación interanual con respecto al promedio de los primeros 5 meses del año anterior.
- Base monetaria: Para obtener la observación anual se toma el promedio de los primeros 5 meses del año y luego se calcula la variación interanual con respecto al promedio de los primeros 5 meses del año anterior.
- Inflación mensual: Para obtener la observación anual se selecciona el máximo dato de los primeros 5 meses del año. Al seleccionar el máximo como dato representativo podremos captar mejor los aumentos de volatilidad en la variable.
- Tasa de interés nominal anual de los plazos fijos de 30 a 59 días: Para obtener la observación anual se selecciona el máximo dato de los primeros 5 meses del año. Al seleccionar el máximo como dato representativo podremos captar mejor las grandes fluctuaciones que generan distorsiones.

Una aclaración importante dentro de los datos de esta variable es que para el primer trimestre de 1990, como consecuencia del plan bonex, se suspendieron las operaciones de plazos fijos y por ende la tasa representativa del primer trimestre de dicho año fue la tasa de interés de las cajas de ahorro. Esto lo podemos encontrar explicado en la página 11 del trabajo denominado “Memorias Anuales 1989-1993” del Banco Central de la República Argentina.

- Tasa de interés anual en Estados Unidos: Para obtener la observación anual se selecciona el máximo dato de los primeros 5 meses del año. Al seleccionar el máximo como dato



representativo podremos captar mejor las grandes fluctuaciones que generan distorsiones.

- Tipo de cambio oficial: Para obtener la observación anual se toma el promedio de los primeros 5 meses del año y luego se calcula la variación interanual con respecto al promedio de los primeros 5 meses del año anterior.
- Inflación interanual: Para obtener la observación anual se selecciona el máximo dato de los primeros 5 meses del año. Al seleccionar el máximo como dato representativo podremos captar mejor las grandes fluctuaciones que generan distorsiones.
- Reservas internacionales en el BCRA: Para obtener la observación anual se toma el promedio de los primeros 5 meses del año y luego se calcula la variación interanual con respecto al promedio de los primeros 5 meses del año anterior.

Cabe aclarar que en el presente trabajo se utilizan datos de los primeros 5 meses del año para obtener las observaciones anuales de cada variable porque hasta el mes 6 no se obtienen datos oficiales del PIB en Argentina para el año en curso por el gran retraso que presenta la elaboración y publicación de dichos datos oficiales. Sin embargo, no es estrictamente necesario que se utilicen datos de los primeros 5 meses del año para obtener las observaciones anuales de las distintas variables, el modelo puede ser utilizado en cualquier momento del año. Se podría, por ejemplo, calcular mes a mes, de la manera en que se explicó anteriormente, la observación anual de cada variable, insertarlas como inputs en el modelo y obtener a lo largo de todo el año predicciones actualizadas sobre el nivel de actividad. Cuantos más datos se incorporen, mayores van a ser las probabilidades de obtener una predicción acertada, ya que se le incorpora mayor cantidad de información al modelo. Es decir realizando predicciones con, por ejemplo, datos de los primeros 9 meses, se obtendrá una predicción más precisa que con los datos de los primeros 3 meses. Es por eso que se recomienda utilizar el modelo en cada mes del año, para que aporte información relevante y oportuna en el proceso de toma de decisiones.

En cuanto a las variables económicas nominales elegidas, éstas están directamente relacionadas con la política monetaria y cambiaria del país. El gobierno tiene poder de decisión sobre una buena parte de las variables incluídas en el modelo, por lo que si el modelo indica que el año será recesivo, el policy maker podrá manipular tales variables e intentar evitar la recesión. Es decir, el accionar del policy maker es directo sobre algunas de las variables incluídas en el modelo. Esto nos permitirá obtener modelos con un alto contenido económico que permitirán predecir la dirección de la actividad económica, comprender las variables que podrían estar generando la situación y actuar en consecuencia. Esta es una gran diferencia con los demás



trabajos existentes relacionados al tema donde se construyen modelos predictivos de la tasa de crecimiento del PIB de Argentina, ya que éstos utilizan muchas variables reales (por ejemplo, indicadores de actividad de mayor frecuencia e índices de distintos sectores de la economía) para sus predicciones, lo cual les permite obtener un buen resultado en cuanto a la predicción de la tasa de crecimiento de la economía, pero pierden contenido económico explicativo y la posibilidad de identificar las variables sobre las que actuar. Es decir, el presente trabajo cambia el foco: no busca solamente predecir, sino también comprender y brindar la posibilidad de intervenir la política monetaria y/o cambiaria para evitar la recesión o una profundización de la misma.

3.3. Estandarización de los datos

Algunos de los algoritmos que se utilizan para construir modelos dentro del campo del machine learning son muy sensibles a la escala de los datos de las distintas variables explicativas. En dichos casos, los modelos pueden verse significativamente afectados por la escala de sus atributos (Zheng y Casari, 2018). Esto lleva a que, por ejemplo, ciertas variables con rangos de valores más amplios puedan influir desproporcionadamente en el modelo y ponderar más en el resultado final simplemente por su escala. Del mismo modo, se les asigna un impacto menor a aquellas variables con rangos de valores más estrechos simplemente por su escala y no por su relevancia explicativa, causal y/o teórica.

Los algoritmos Support Vector Machines (SVM) y k-Nearest Neighbor (kNN), que son los algoritmos que se utilizarán junto al algoritmo Random Forest para construir los modelos, son sensibles a la escala de las variables explicativas debido a que utilizan métodos de optimización que dependen de las distancias entre puntos de datos. Si las variables explicativas no están en la misma escala, el proceso de optimización puede ser más lento o incluso no converger de manera adecuada.

Como consecuencia de lo dicho anteriormente, es indispensable estandarizar los datos para luego obtener resultados más precisos.

En el presente trabajo se decidió aplicar la estandarización Z-score o basada en media y desviación estándar. Este procedimiento consiste en transformar las variables para que tengan una media de 0 y una desviación estándar de 1. La fórmula es la siguiente:

$$Z = \frac{X - \mu}{\sigma}$$



Donde:

- X es el valor original de la variable.
- μ es la media de la variable.
- σ es la desviación estándar de la variable.

3.4. Entrenamiento y metodologías de resampling

En total se cuenta con 62 observaciones (correspondientes a los años que van desde 1963 a 2024), donde aproximadamente el 60% de los años son no recesivos y el 40% restante si lo son.

Para entrenar de manera supervisada los modelos, se dividen aleatoriamente los datos en un subconjunto de entrenamiento con el 80% de las observaciones y en un subconjunto de testeo con el 20% restante de las observaciones. En el subconjunto de entrenamiento quedan 49 observaciones y en el subconjunto de testeo quedan las 13 observaciones restantes. A su vez, en las 13 observaciones del subconjunto de testeo tenemos 8 años no recesivos y 5 años recesivos. De esta manera, se mantiene aproximadamente la misma proporción de años recesivos y no recesivos en los 3 conjuntos: el original, el de entrenamiento y el de testeo.

Durante el proceso de entrenamiento se aplican dos métodos distintos de resampling. Por un lado se aplica la metodología “cross-validation”, y por el otro se aplica “bootstrapping”. Esto se realiza con el objetivo de obtener una mayor cantidad y variedad de modelos, realizar el entrenamiento y validación de manera conjunta y reducir el sobreajuste a los datos de entrenamiento.

De esta manera se desarrollan modelos con 3 algoritmos distintos y con 2 métodos de resampling distintos, por lo que resultan en total 6 modelos:

- Modelo con algoritmo SVM y con resampling Cross-Validation.
- Modelo con algoritmo SVM y con resampling Bootstrapping.
- Modelo con algoritmo k-NN y con resampling Cross-Validation.
- Modelo con algoritmo k-NN y con resampling Bootstrapping.
- Modelo con algoritmo Random Forest y con resampling Cross-Validation.
- Modelo con algoritmo Random Forest y con resampling Bootstrapping.



3.4.1. Cross-Validation

La validación cruzada es una técnica que se utiliza para entrenar y validar un modelo dividiendo el conjunto de datos en varios subconjuntos (o folds). El procedimiento más común es el “k-fold cross-validation”, donde el conjunto de datos se divide en k partes. En cada iteración, el modelo se entrena con k-1 partes del conjunto de datos y se valida con la parte restante. Este proceso se repite k veces, y al final, se promedia el rendimiento de todas las iteraciones.

Al aplicar esta metodología durante el entrenamiento del modelo se obtienen ventajas como:

- Mejorar la capacidad del modelo para generalizar.
- Reducir el sesgo de la estimación que podría ocurrir si solo se usa un conjunto de entrenamiento y uno de prueba.
- Obtener un modelo menos susceptible al azar que con una única partición de entrenamiento y prueba.
- Evitar que el modelo se sobreajuste (overfitting) al conjunto de datos de entrenamiento.

3.4.2. Bootstrapping

El bootstrapping es un método de resampling que implica tomar múltiples muestras con reemplazo del conjunto de datos de entrenamiento. Esto quiere decir que se crean nuevos subconjuntos de entrenamiento a partir del conjunto de entrenamiento original, permitiendo que algunas observaciones sean seleccionadas más de una vez y otras no sean seleccionadas en absoluto.

En cada iteración, el modelo se entrena en uno de estos subconjuntos (llamados bootstrap samples) y se evalúa en el conjunto restante, denominado out-of-bag (OOB). La estimación final del modelo, al igual que en la metodología “cross-validation”, se obtiene promediando los resultados de todas las iteraciones.

Al aplicar esta metodología durante el entrenamiento del modelo se obtienen ventajas como:

- Usar todos los datos disponibles para entrenar y evaluar el modelo.
- Reducir el sobreajuste al entrenar el modelo en diferentes subconjuntos de los datos y evaluarlo en los puntos no seleccionados.
- Obtener un buen resultado pese a trabajar con conjuntos de datos pequeños, ya que el bootstrapping es especialmente útil en estos casos.



4. Métricas de evaluación de los modelos

Siguiendo a Sánchez (2023), la principal característica que tiene que tener un modelo de clasificación para ser considerado óptimo es la capacidad de generalizar buenas clasificaciones sobre datos no vistos anteriormente. Existen distintas métricas que permiten evaluar el rendimiento de un modelo de clasificación y éstas pueden calcularse a partir de la matriz de confusión obtenida de un modelo de clasificación.

4.1. Matriz de confusión

La matriz de confusión ayuda a evaluar el rendimiento del modelo de clasificación comparando los valores predichos por el modelo con los valores reales del conjunto de datos de testeo.

Si el problema es de clasificación y existen solamente dos clases en la variable dependiente, es decir ésta es dicotómica, por lo general las clases hacen referencia a la presencia o no de un cierto atributo. Si el atributo está presente se trata de un caso “positivo”, y si el atributo está ausente se trata de un caso “negativo”.

Lo anterior ocurre en el presente trabajo, donde la variable dependiente “Recesión” indica para cada año si hubo recesión o no, siendo la recesión el atributo que se evalúa. En problemas de este tipo, la matriz de confusión se divide en:

- Verdaderos positivos: el modelo predice presencia del atributo y la predicción es correcta.
- Falsos negativos: el modelo predice ausencia del atributo y la predicción no es correcta.
- Falsos positivos: el modelo predice presencia del atributo y la predicción no es correcta
- Verdaderos negativos: modelo predice ausencia del atributo y la predicción es correcta.

La estructura de una matriz de confusión se muestra en la siguiente imagen:

Verdaderos Positivos	Falsos Positivos
Falsos Negativos	Verdaderos Negativos

A partir de la matriz de confusión, en el presente trabajo se construirán y utilizarán 3 métricas para evaluar los distintos modelos: exactitud/accuracy, precisión/precision y la curva ROC.



4.1.1. Exactitud – Accuracy

Indica la proporción de predicciones correctas respecto al total de predicciones realizadas por el modelo.

4.1.2. Precisión – Precision

Indica la proporción de casos identificados como positivos que son verdaderamente positivos. Esta métrica nos sirve para ver qué tan bueno es nuestro modelo identificando el atributo, que en este caso es la presencia de recesión económica.

4.1.3. La curva ROC

Es una representación gráfica que se construye teniendo en cuenta la proporción de casos positivos que fueron correctamente identificados por el modelo y la proporción de casos negativos que fueron bien identificados por nuestro modelo. Cuanto mayor sea el área por debajo de dicha curva, mejor será nuestro modelo de clasificación binario (un sistema perfecto tendría un área de 1).

Siguiendo nuevamente a Sánchez (2023), de manera general podemos determinar qué tan bueno es nuestro modelo en función del área por debajo de la curva ROC de la siguiente manera:

- ≤ 0.5 : Modelo muy malo
- $[0.5, 0.6)$: Modelo malo
- $[0.6, 0.75)$: Modelo regular
- $[0.75, 0.9)$: Modelo bueno
- $[0.9, 0.97)$: Modelo muy bueno
- $[0.97, 1)$: Modelo excelente

5. Resultados obtenidos y análisis de métricas

En el presente capítulo se presentan los resultados de las predicciones obtenidas en cada uno de los 6 modelos. Adicionalmente, a partir de la matriz de confusión, se analizan las 3 métricas seleccionadas para evaluar y comparar los modelos entre sí: exactitud/accuracy, precisión/precision y la curva ROC.

5.1. Modelo con algoritmo SVM y con resampling Cross-Validation

Una vez realizadas las predicciones con este modelo, las mismas se comparan con el vector de datos de testeo que se había apartado anteriormente y así resulta la matriz de confusión que se observa en la siguiente figura:

Figura 1: Matriz de confusión del modelo con algoritmo SVM y con resampling Cross-Validation.

Predicciones_SVM_CV		
Datos_Reales	No	Si
No	8	0
Si	5	0

Fuente: Elaboración propia, mayo de 2025.

Como se explica en el apartado 3.4., dentro del conjunto de testeo hay 8 años que no fueron recesivos y 5 años que si lo fueron. En la Figura 1 se observa que el modelo predice de manera perfecta los 8 años no recesivos, pero erra fuertemente al predecir años recesivos ya que no clasificó de manera correcta ninguno.

Figura 2: Métricas del modelo con algoritmo SVM y con resampling Cross-Validation.

```
Accuracy : 0.6154
95% CI : (0.3158, 0.8614)
No Information Rate : 1
P-Value [Acc > NIR] : 1.00000

Kappa : 0

McNemar's Test P-Value : 0.07364

Sensitivity : NA
Specificity : 0.6154
Pos Pred Value : NA
Neg Pred Value : NA
Prevalence : 0.0000
Detection Rate : 0.0000
Detection Prevalence : 0.3846
Balanced Accuracy : NA

'Positive' class : Si
```

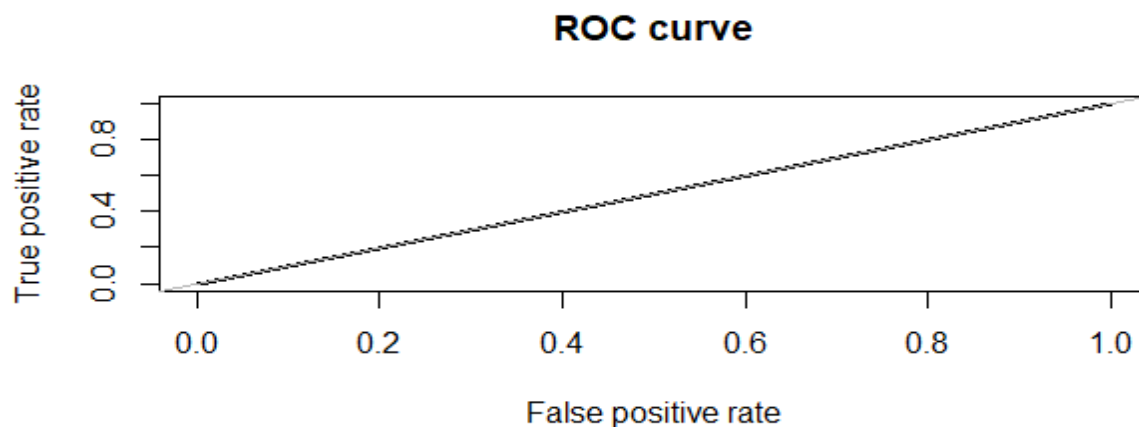
Fuente: Elaboración propia, mayo de 2025.

En la Figura 2 se observa un listado de métricas obtenidas para el modelo bajo análisis. Analizando las métricas elegidas:

- Accuracy: Se observa que en este modelo es de 61,54%, lo cual nos indica que el modelo predice de manera correcta aproximadamente el 62% de todas las predicciones que realiza.
- Pos Pred Value: Es la precisión del modelo. Se observa que para este modelo se obtiene un “NA”, lo cual nos indica que el modelo no pudo clasificar de manera correcta ninguno de los años recesivos.

Ambas métricas indican que este modelo es bastante pobre en cuanto a desempeño ya que no solo no logra clasificar correctamente ni el 63% del total de observaciones, sino que además tiene muchos problemas en la detección de años recesivos, que es lo que verdaderamente nos interesa lograr en el presente trabajo.

Figura 3: Curva ROC del modelo con algoritmo SVM y con resampling Cross-Validation.



Fuente: Elaboración propia, mayo de 2025.

En la Figura 3 se observa la curva ROC del modelo bajo análisis. La curva tiene un área por debajo de 0.5, lo cual indica que es un modelo malo en base a la escala realizada por Sánchez (2023).

Así, la curva ROC se suma a las otras 2 métricas y, en conjunto, las 3 métricas indican que este modelo es malo y no logra cumplir de manera satisfactoria el objetivo de este trabajo.

5.2. Modelo con algoritmo SVM y con resampling Bootstrapping

Una vez realizadas las predicciones con este modelo, las mismas se comparan con el vector de datos de testeo que se había apartado anteriormente y así resulta la matriz de confusión que se observa en la siguiente figura:

Figura 4: Matriz de confusión del modelo con algoritmo SVM y con resampling Bootstrapping.

		Predicciones_SVM_BOOT	
Datos_Reales	No	Si	
No	7	1	
Si	3	2	

Fuente: Elaboración propia, mayo de 2025.

En la Figura 4 se observa que el modelo predice bastante bien años no recesivos, ya que clasifica de manera correcta 7 de estos 8 años. Sin embargo, con este método de resampling el algoritmo SVM permite predecir correctamente 2 de los 5 años recesivos existentes en el conjunto de datos de testeo.

Figura 5: Métricas del modelo con algoritmo SVM y con resampling Bootstrapping.

```
Accuracy : 0.6923
95% CI : (0.3857, 0.9091)
No Information Rate : 0.7692
P-value [Acc > NIR] : 0.8397

Kappa : 0.2973

McNemar's Test P-Value : 0.6171

Sensitivity : 0.6667
Specificity : 0.7000
Pos Pred Value : 0.4000
Neg Pred Value : 0.8750
Prevalence : 0.2308
Detection Rate : 0.1538
Detection Prevalence : 0.3846
Balanced Accuracy : 0.6833

'Positive' class : Si
```

Fuente: Elaboración propia, mayo de 2025.

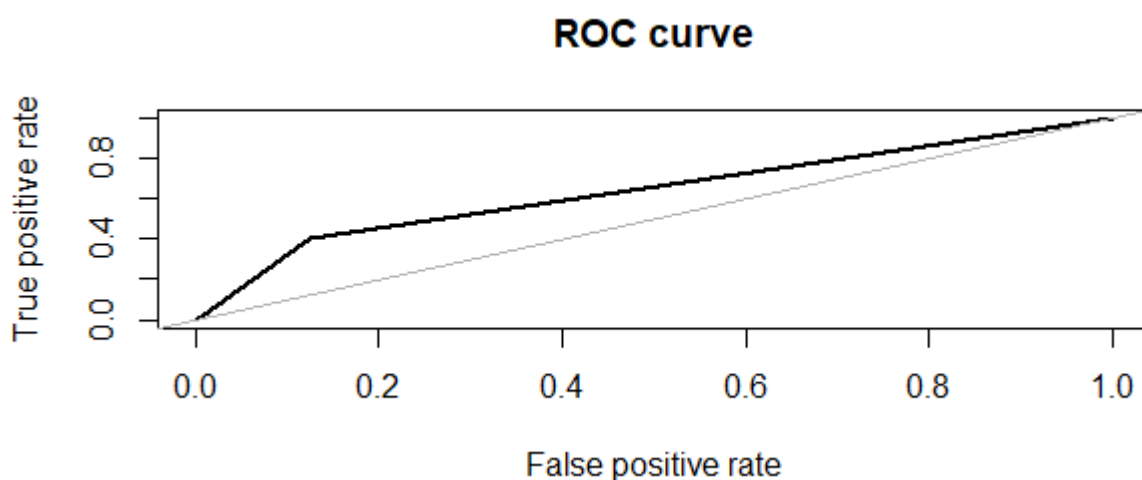
En la Figura 5 se observa un listado de métricas obtenidas para el modelo bajo análisis. Analizando las métricas elegidas:

- Accuracy: Se observa que en este modelo es poco mayor al 69%, lo cual nos indica que el modelo predice de manera correcta aproximadamente el 69% de todas las predicciones que realiza.

- Pos Pred Value: A diferencia del anterior modelo, en este modelo si obtenemos la precisión y la misma es del 40%, lo cual nos indica que el modelo clasifica de manera correcta aproximadamente el 40% de los años recesivos.

Ambas métricas indican que, pese a ser mejor que el anterior, este modelo es también bastante pobre en cuanto a desempeño ya que no logra clasificar correctamente ni el 70% del total de observaciones y además también tiene problemas en la detección de años recesivos.

Figura 6: Curva ROC del modelo con algoritmo SVM y con resampling Bootstrapping.



Fuente: Elaboración propia, mayo de 2025.

En la Figura 6 se observa la curva ROC del modelo bajo análisis. La curva tiene un área por debajo de 0.637, lo cual indica que es un modelo regular en base a la escala realizada por Sánchez (2023).

Así, en este modelo también la curva ROC se suma a las otras 2 métricas y, en conjunto, las 3 métricas indican que este modelo es malo y tampoco logra cumplir de manera satisfactoria el objetivo de este trabajo.

5.3. Modelo con algoritmo k-NN y con resampling Cross-Validation

Una vez realizadas las predicciones con este modelo, las mismas se comparan con el vector de datos de testeo que se había apartado anteriormente y así resulta la matriz de confusión que se observa en la siguiente figura:

Figura 7: Matriz de confusión del modelo con algoritmo k-NN y con resampling Cross-Validation.

		Predicciones_KNN_CV	
Datos_Reales	No	Si	
No	8	0	
Si	1	4	

Fuente: Elaboración propia, mayo de 2025.



En la Figura 7 se observa que el modelo predice a la perfección los años no recesivos, ya que clasifica de manera correcta los 8 años. De manera similar, este modelo es también bastante bueno para predecir años recesivos ya que predijo correctamente 4 de los 5 años, clasificando de manera errónea solamente un año recesivo.

Figura 8: Métricas del modelo con algoritmo k-NN y con resampling Cross-Validation.

```
Accuracy : 0.9231
95% CI : (0.6397, 0.9981)
No Information Rate : 0.6923
P-Value [Acc > NIR] : 0.05688

Kappa : 0.8312

McNemar's Test P-Value : 1.00000

Sensitivity : 1.0000
Specificity : 0.8889
Pos Pred Value : 0.8000
Neg Pred Value : 1.0000
Prevalence : 0.3077
Detection Rate : 0.3077
Detection Prevalence : 0.3846
Balanced Accuracy : 0.9444

'Positive' class : si
```

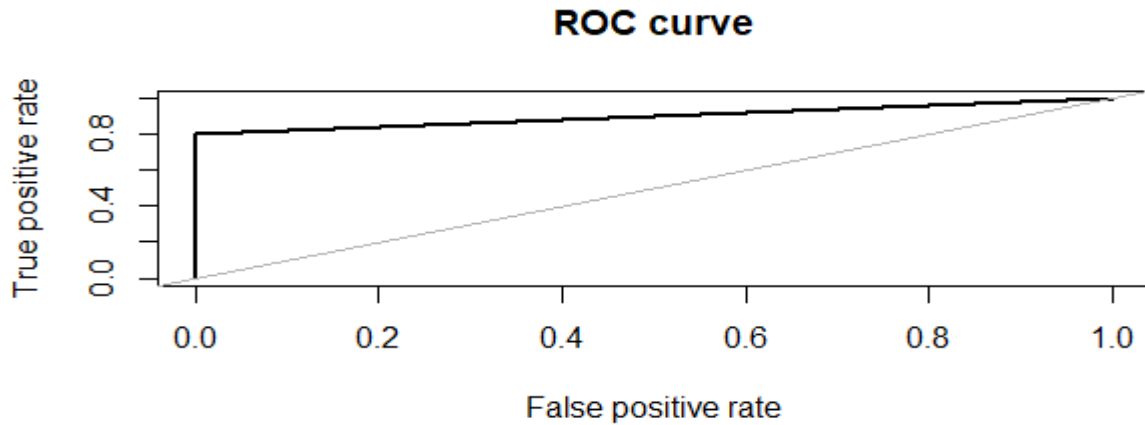
Fuente: Elaboración propia, mayo de 2025.

En la Figura 8 se observa el listado de métricas obtenidas para el modelo bajo análisis. Analizando las métricas elegidas:

- Accuracy: Se observa que en este modelo es poco mayor al 92%, lo cual nos indica que el modelo predice de manera correcta aproximadamente el 92% de todas las predicciones que realiza.
- Pos Pred Value: Es la precisión del modelo. Se observa que en este modelo es del 80%, lo cual nos indica que el modelo clasifica de manera correcta el 80% de los casos positivos de la variable dependiente, es decir, los años recesivos.

Ambas métricas indican que este modelo es considerablemente bueno en cuanto a desempeño ya que no baja de una tasa de acierto del 80% en sus predicciones, ya sea evaluándolo tanto en años recesivos y no recesivos como evaluándolo únicamente en la detección de años recesivos, que es el atributo que se quiere identificar y predecir exitosamente con los modelos.

Figura 9: Curva ROC del modelo con algoritmo k-NN y con resampling Cross-Validation.



Fuente: Elaboración propia, mayo de 2025.

En la Figura 9 se observa la curva ROC del modelo bajo análisis. La curva tiene un área por debajo de 0.9, lo cual indica que es un modelo muy bueno en base a la escala realizada por Sánchez (2023).

Así, la curva ROC se suma a las otras 2 métricas y, en conjunto, las 3 métricas indican que este modelo puede considerarse como satisfactorio para el cumplimiento del objetivo de este trabajo.

5.4. Modelo con algoritmo k-NN y con resampling Bootstrapping

Una vez realizadas las predicciones con este modelo, las mismas se comparan con el vector de datos de testeo que se había apartado anteriormente y así resulta la matriz de confusión que se observa en la siguiente figura:

Figura 10: Matriz de confusión del modelo con algoritmo k-NN y con resampling Bootstrapping.

		Predicciones_KNN_BOOT	
Datos_Reales	No	Si	
No	8	0	
Si	3	2	

Fuente: Elaboración propia, mayo de 2025.

En la Figura 10 se observa que el modelo predice a la perfección los años no recesivos, ya que clasifica de manera correcta los 8 años no recesivos. Sin embargo, este modelo es bastante impreciso para predecir años recesivos ya que solo predijo correctamente 2 de los 5 años, clasificando de manera errónea los restantes 3 años.



Figura 11: Métricas del modelo con algoritmo k-NN y con resampling Bootstrapping.

```
Accuracy : 0.7692
95% CI : (0.4619, 0.9496)
No Information Rate : 0.8462
P-Value [Acc > NIR] : 0.8732

Kappa : 0.4507

Mcnemar's Test P-value : 0.2482

Sensitivity : 1.0000
Specificity : 0.7273
Pos Pred Value : 0.4000
Neg Pred Value : 1.0000
Prevalence : 0.1538
Detection Rate : 0.1538
Detection Prevalence : 0.3846
Balanced Accuracy : 0.8636

'Positive' Class : si
```

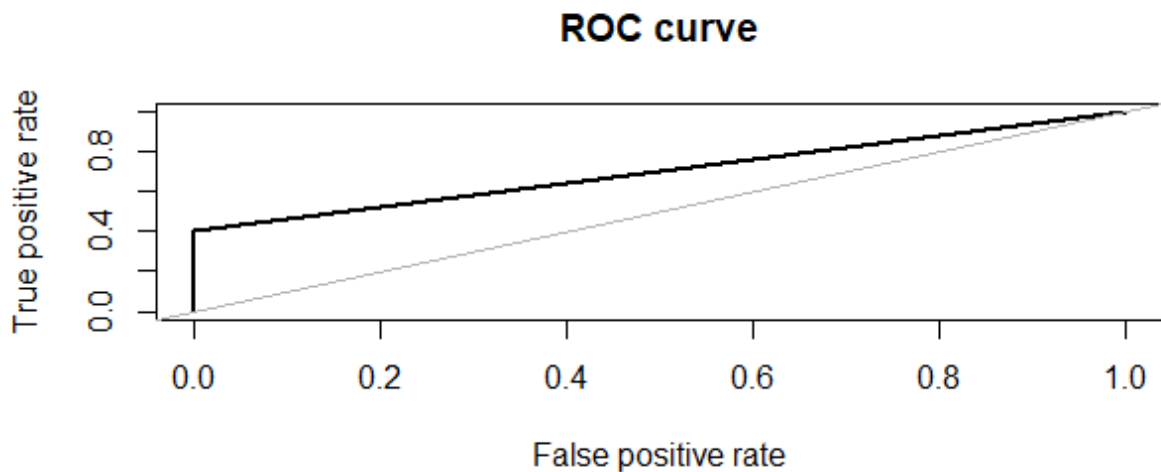
Fuente: Elaboración propia, mayo de 2025.

En la Figura 11 se observa el listado de métricas obtenidas para el modelo bajo análisis. Analizando las métricas elegidas:

- Accuracy: Se observa que en este modelo es de casi el 77%, lo cual nos indica que el modelo predice de manera correcta aproximadamente el 77% de todas las predicciones que realiza.
- Pos Pred Value: En este modelo la precisión es del 40%, lo cual nos indica que el modelo clasifica de manera correcta el 40% de los años recesivos.

La accuracy del modelo no es mala pero tampoco da mucha tranquilidad, está alrededor de un nivel aceptable. Sin embargo, la precisión del modelo si es mala, no es aceptable que el modelo logre detectar únicamente el 40% de los años recesivos.

Figura 12: Curva ROC del modelo con algoritmo k-NN y con resampling Bootstrapping.



Fuente: Elaboración propia, mayo de 2025.

En la Figura 12 se observa la curva ROC del modelo bajo análisis. La curva tiene un área por debajo de 0.7, lo cual indica que es un modelo regular en base a la escala realizada por Sánchez (2023).

Así, la curva ROC se suma a la métrica precisión y, en conjunto, indican que este modelo es regular y no logra cumplir de manera satisfactoria el objetivo de este trabajo. La métrica accuracy, si bien no es mala, no logra ser muy satisfactoria como para compensar los resultados que muestran las otras 2 métricas para este modelo.

5.5. Modelo con algoritmo Random Forest y con resampling Cross-Validation

Una vez realizadas las predicciones con este modelo, las mismas se comparan con el vector de datos de testeo que se había apartado anteriormente y así resulta la matriz de confusión que se observa en la siguiente figura:

Figura 13: Matriz de confusión del modelo con algoritmo Random Forest y con resampling Cross-Validation.

		Predicciones_RF_CV	
Datos_Reales	No	Si	
No	7	1	
Si	1	4	

Fuente: Elaboración propia, mayo de 2025.

En la Figura 13 se observa que el modelo predice bastante bien años no recesivos, ya que clasifica de manera correcta 7 de estos 8 años. De igual manera, este modelo es también bastante preciso para predecir años recesivos ya que predijo correctamente 4 de los 5 años, clasificando de manera errónea solamente un año recesivo.



Figura 14: Métricas del modelo con algoritmo Random Forest y con resampling Cross-Validation.

```
Accuracy : 0.8462
95% CI : (0.5455, 0.9808)
No Information Rate : 0.6154
P-Value [Acc > NIR] : 0.07187

Kappa : 0.675

McNemar's Test P-Value : 1.00000

Sensitivity : 0.8000
Specificity : 0.8750
Pos Pred Value : 0.8000
Neg Pred Value : 0.8750
Prevalence : 0.3846
Detection Rate : 0.3077
Detection Prevalence : 0.3846
Balanced Accuracy : 0.8375

'positive' class : si
```

Fuente: Elaboración propia, mayo de 2025.

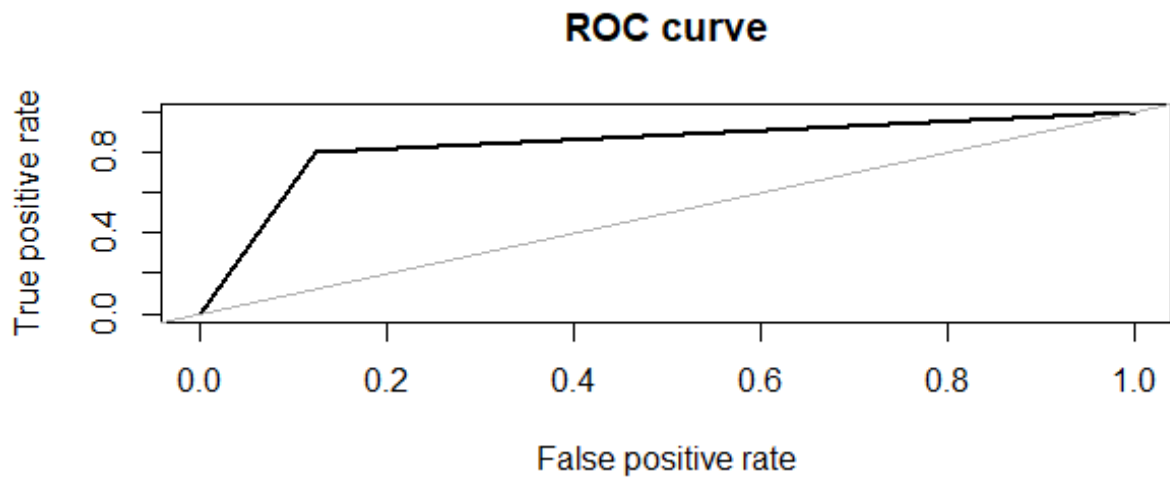
En la Figura 14 se observa un listado de métricas obtenidas para el modelo bajo análisis.

Analizando las métricas elegidas:

- Accuracy: Se observa que en este modelo es poco mayor al 84,6%, lo cual nos indica que el modelo predice de manera correcta aproximadamente el 84,6% de todas las predicciones que realiza.
- Pos Pred Value: Es la precisión del modelo. Se observa que en este modelo es del 80%, lo cual nos indica que el modelo clasifica de manera correcta el 80% de los casos positivos de la variable dependiente, es decir, los años recesivos.

Ambas métricas indican que este modelo es relativamente bueno en cuanto a desempeño ya que no baja de una tasa de acierto del 80% en sus predicciones, ya sea evaluandolo tanto años recesivos y no recesivos como evaluandolo únicamente en la detección de años recesivos, que es el atributo que se quiere identificar y predecir exitosamente con los modelos.

Figura 15: Curva ROC del modelo con algoritmo Random Forest y con resampling Cross-Validation.



Fuente: Elaboración propia, mayo de 2025.

En la Figura 15 se observa la curva ROC del modelo bajo análisis. La curva tiene un área por debajo de 0.838, lo cual indica que es un modelo bueno en base a la escala realizada por Sánchez (2023).

Así, la curva ROC se suma a las otras 2 métricas y, en conjunto, las 3 métricas indican que este modelo puede considerarse como satisfactorio para el cumplimiento del objetivo de este trabajo.

5.6. Modelo con algoritmo Random Forest y con resampling Bootstrapping

Una vez realizadas las predicciones con este modelo, las mismas se comparan con el vector de datos de testeo que se había apartado anteriormente y así resulta la matriz de confusión que se observa en la siguiente figura:

Figura 16: Matriz de confusión del modelo con algoritmo Random Forest y con resampling Bootstrapping.

		Predicciones_RF_BOOT	
Datos_Reales		NO	SI
NO		7	1
SI		2	3

Fuente: Elaboración propia, mayo de 2025.

En la Figura 16 se observa que el modelo predice bastante bien años no recesivos, ya que clasifica de manera correcta 7 de estos 8 años. Por otro lado, tiene un rendimiento regular en la detección de años recesivos ya que logra predecir correctamente 3 de estos 5 años.



Figura 17: Métricas del modelo con algoritmo Random Forest y con resampling Bootstrapping.

```
Accuracy : 0.7692
95% CI : (0.4619, 0.9496)
No Information Rate : 0.6923
P-value [Acc > NIR] : 0.3969

Kappa : 0.4935

Mcnemar's Test P-Value : 1.0000

Sensitivity : 0.7500
Specificity : 0.7778
Pos Pred Value : 0.6000
Neg Pred Value : 0.8750
Prevalence : 0.3077
Detection Rate : 0.2308
Detection Prevalence : 0.3846
Balanced Accuracy : 0.7639

'Positive' Class : si
```

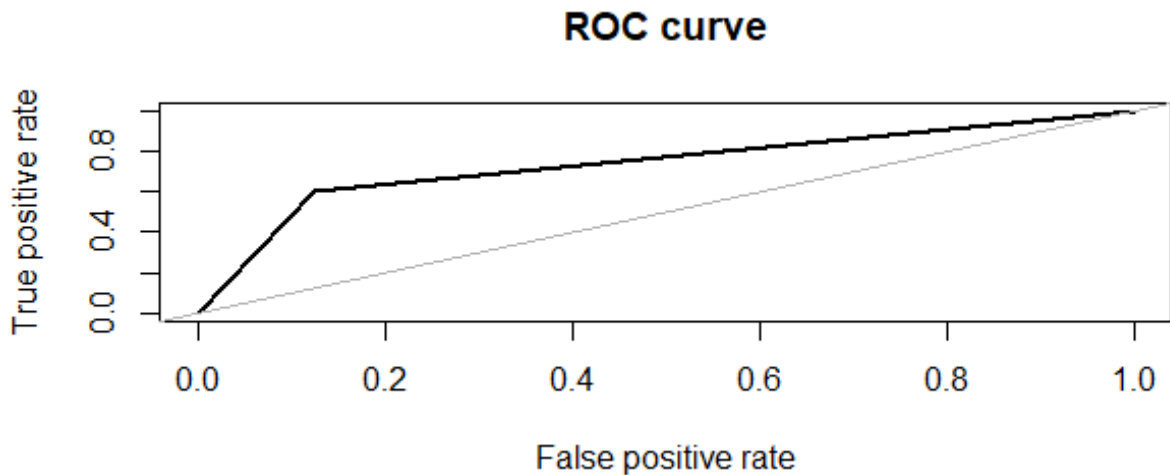
Fuente: Elaboración propia, mayo de 2025.

En la Figura 17 se observa un listado de métricas obtenidas para el modelo bajo análisis. Analizando las métricas elegidas:

- Accuracy: Se observa que en este modelo es de casi el 77%, lo cual nos indica que el modelo predice de manera correcta aproximadamente el 77% de todas las predicciones que realiza.
- Pos Pred Value: Es la precisión del modelo. Se observa que en este modelo es del 60%, lo cual nos indica que el modelo clasifica de manera correcta el 60% de los casos positivos de la variable dependiente, es decir, los años recesivos.

Al igual que lo ocurrido con el modelo con algoritmo k-NN y con resampling Bootstrapping, la accuracy no es mala pero tampoco da mucha tranquilidad, está alrededor de un nivel aceptable. Por su parte, la precisión del modelo es regular ya que si bien clasifica más de la mitad de los años recesivos de manera correcta, el margen de error es bastante grande y no se puede aceptar dicho margen de error.

Figura 18: Curva ROC del modelo con algoritmo Random Forest y con resampling Bootstrapping.



Fuente: Elaboración propia, mayo de 2025.

En la Figura 18 se observa la curva ROC del modelo bajo análisis. La curva tiene un área por debajo de 0.738, lo cual indica que es un modelo regular en base a la escala realizada por Sánchez (2023).

De esta manera, pese a que la accuracy del modelo se encuentra cerca de un nivel aceptable, tanto la curva ROC como la precisión del modelo indican que es un modelo malo y que no logra cumplir de manera satisfactoria el objetivo de este trabajo.

6. Incidencia de las variables en las recesiones

En el presente capítulo se realizará un análisis de componentes principales (PCA) utilizando el conjunto de datos con todos los años recesivos con el objetivo de determinar patrones dentro del conjunto de datos, identificando relaciones entre los distintos años recesivos y entre éstos y las variables bajo estudio.

El análisis de componentes principales transforma un conjunto de datos original, con múltiples variables, en un nuevo conjunto de variables, llamadas componentes principales. Estas nuevas variables son combinaciones lineales de las variables originales y son ordenadas de tal manera que la primera componente principal captura la mayor parte de la variabilidad en los datos, la segunda componente principal captura la segunda mayor cantidad de variabilidad, y así sucesivamente.

Se utilizará la proporción acumulada para determinar la cantidad total de varianza que explican los componentes principales. Se conservarán los componentes principales que expliquen un nivel aceptable de varianza. El nivel aceptable depende de la aplicación específica, pero para



propósitos descriptivos como el de el presente trabajo lo ideal sería explicar como mínimo el 80% de la varianza.

El proceso de PCA identifica aquellas direcciones en las que la varianza es mayor. Como la varianza de una variable se mide en su misma escala elevada al cuadrado, si antes de calcular las componentes no se estandarizan todas las variables para que tengan media 0 y desviación estándar 1, aquellas variables cuya escala sea mayor dominarán al resto.

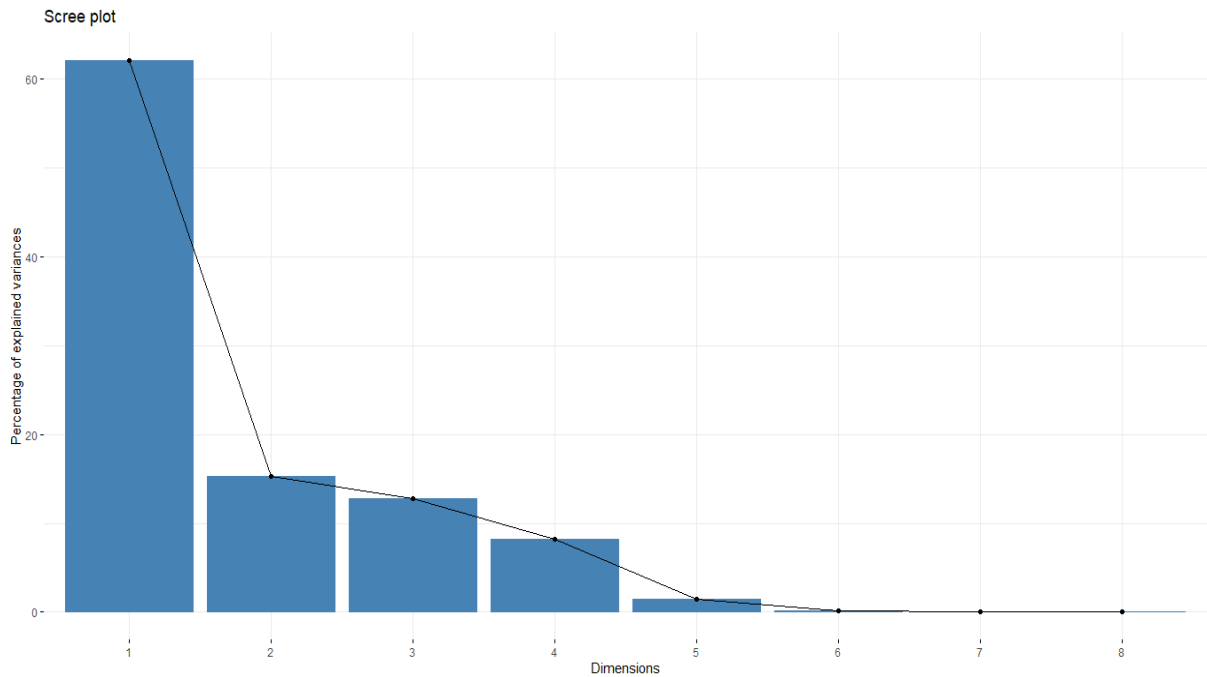
Una vez estandarizadas las variables, se observa que dentro de nuestro conjunto de datos tenemos 4 años recesivos (1976, 1978, 1989 y 1990) en los que algunas variables toman valores muy extremos, es decir, estos años presentan observaciones que son outliers. Teniendo en cuenta que los datos están normalizados, se considera outlier a aquella observación que está a más de 3 desviaciones estándar de la media (regla de las 3 sigmas). Estos outliers pueden llevar a que el análisis de componentes principales brinde un resultado distorsionado. Por ende, vamos a realizar el análisis de componentes principales primero teniendo en cuenta estos años y, luego, excluyendo del análisis a esos años que presentan outliers, y se realizará una comparación de ambos resultados.

Los puntos que se encuentran más cercanos, son puntos que tienen características (valores de las variables) similares. Por otro lado, las variables que tienen una flecha más larga y profunda, son las más relevantes para esos años.

6.1. Análisis de componentes principales con todos los años recesivos

Una vez realizado el análisis de componentes principales con el conjunto de datos completo, lo primero que se analizará es el scree plot. Éste es un gráfico que muestra la importancia de cada componente principal en el PCA y suele utilizarse para determinar cuántos componentes principales conservar. Más precisamente, el scree plot muestra los valores propios, que son medidas de cuánto contribuye cada componente a la varianza total.

Gráfico 4: Scree plot del PCA con todos los años recesivos del conjunto de datos.

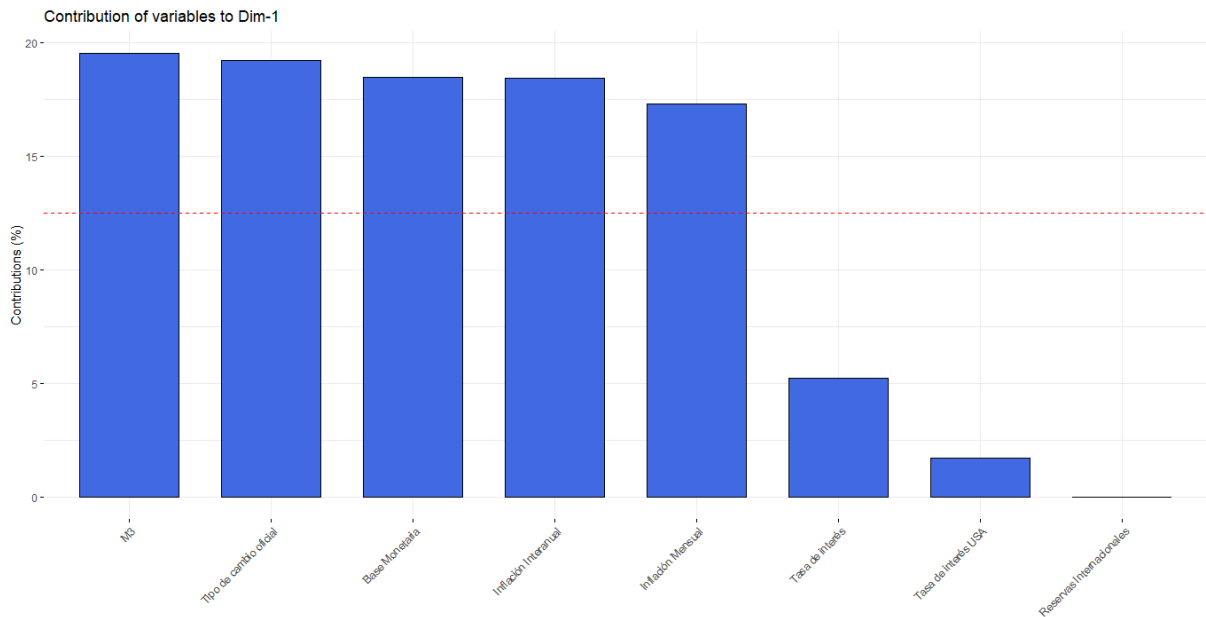


Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 4 se observa que las primeras tres componentes principales contribuyen a la explicación de más del 80% de la varianza total. Sin embargo, el gráfico en 3 dimensiones es más complicado de visualizar y, como consecuencia, es más complicado también extraer conclusiones en base a dicha gráfica, por lo que nos quedaremos con las primeras 2 dimensiones, las cuales explican aproximadamente el 77% de la varianza total y su gráfica es más sencilla de interpretar.

En segundo lugar se analizarán las variables que más contribuyen a la varianza explicada por las dos primeras componentes principales.

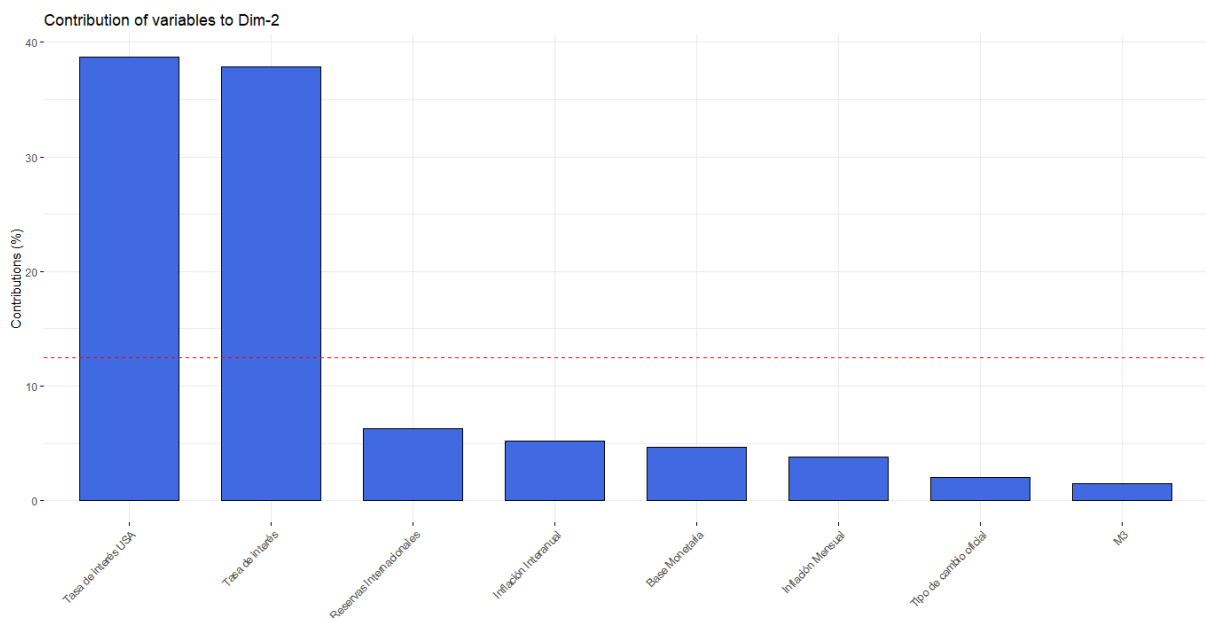
Gráfico 5: Contribución de las variables a la varianza explicada por la primer componente principal.



Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 5 se observa que las variables que más contribuyen en la primer componente principal son el agregado monetario M3, el tipo de cambio oficial, la base monetaria, la inflación interanual y la inflación mensual.

Gráfico 6: Contribución de las variables a la varianza explicada por la segunda componente principal.



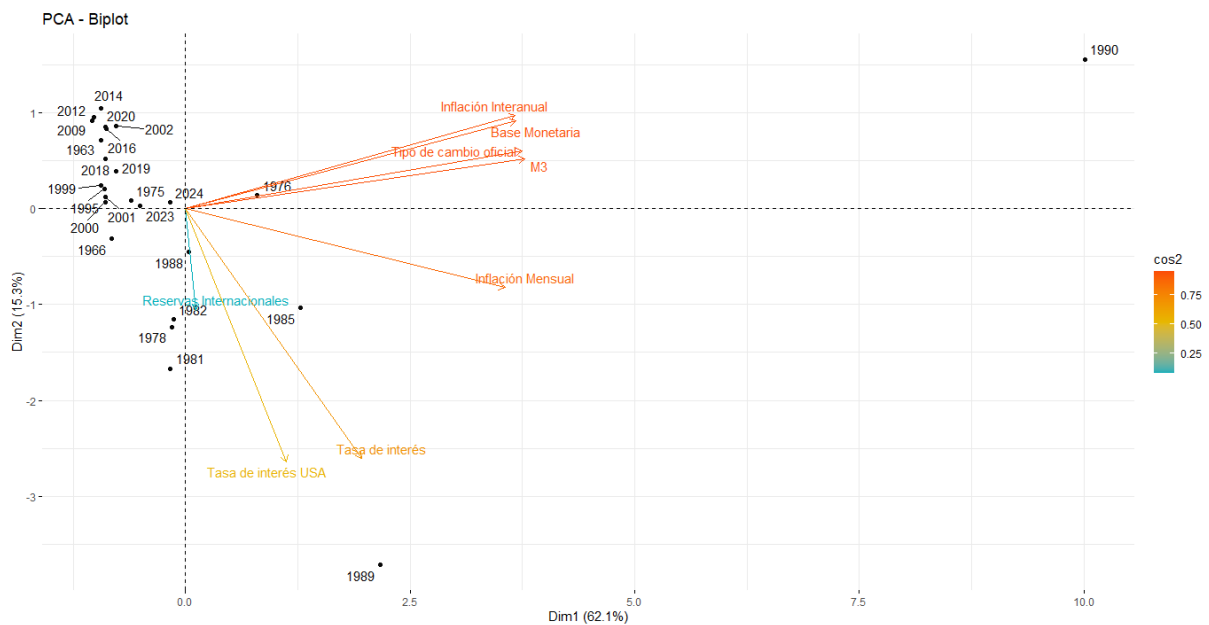
Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 6 se observa que las variables que más contribuyen en la segunda componente principal son la tasa de interés de Estados Unidos y la tasa de interés nacional.

Por las variables que resultan relevantes en cada componente principal, puede pensarse que el resultado de este análisis de componentes principales está muy influenciado por los valores atípicos que se encuentran en los años 1989 y 1990, donde hubo hiperinflación y por ende las variaciones en las variables bajo análisis fueron muy extremas en comparación al resto de años recesivos. En el próximo apartado se realizará el análisis de componentes principales excluyendo los años con valores atípicos y se verá si cambian las variables más relevantes dentro de cada componente principal o si continúan siendo las mismas.

Finalmente analizaremos el biplot del PCA, el cual es un gráfico que representa simultáneamente información sobre las observaciones y las variables.

Gráfico 7: Biplot del PCA con todos los años recesivos del conjunto de datos.



Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 7 se observa que el primer componente principal tiene asociaciones positivas grandes con el agregado monetario M3, el tipo de cambio, la base monetaria, la inflación mensual y la inflación interanual. Como el primer componente principal tiene asociaciones positivas grandes con las variables mencionadas, podríamos decir que este componente mide principalmente el factor inflacionario y devaluatorio generado por la emisión monetaria. El segundo componente tiene asociaciones negativas grandes con las tasas de interés nacional y



de Estados Unidos, podríamos decir entonces que este componente mide principalmente las variables que se usan para contrarrestar la alta inflación y los saltos en el tipo de cambio.

En el biplot cada variable ha sido coloreada según su “cos²”, que es una de las métricas que obtenemos cuando hacemos el análisis de componentes principales. En el contexto del PCA, el cos² refleja la calidad de la representación de las variables originales en el espacio de las componentes principales. Específicamente:

- cos² indica la proporción de la variabilidad de una variable original que se explica por las componentes principales seleccionadas. En otras palabras, mide qué tan bien la información de esa variable se conserva en el nuevo espacio bidimensional.
- Un cos² alto sugiere que la variable está bien representada por las componentes principales y que su información es capturada de manera efectiva.
- Un cos² bajo indica que la variable está mal representada y que su información no se conserva bien en el espacio reducido.

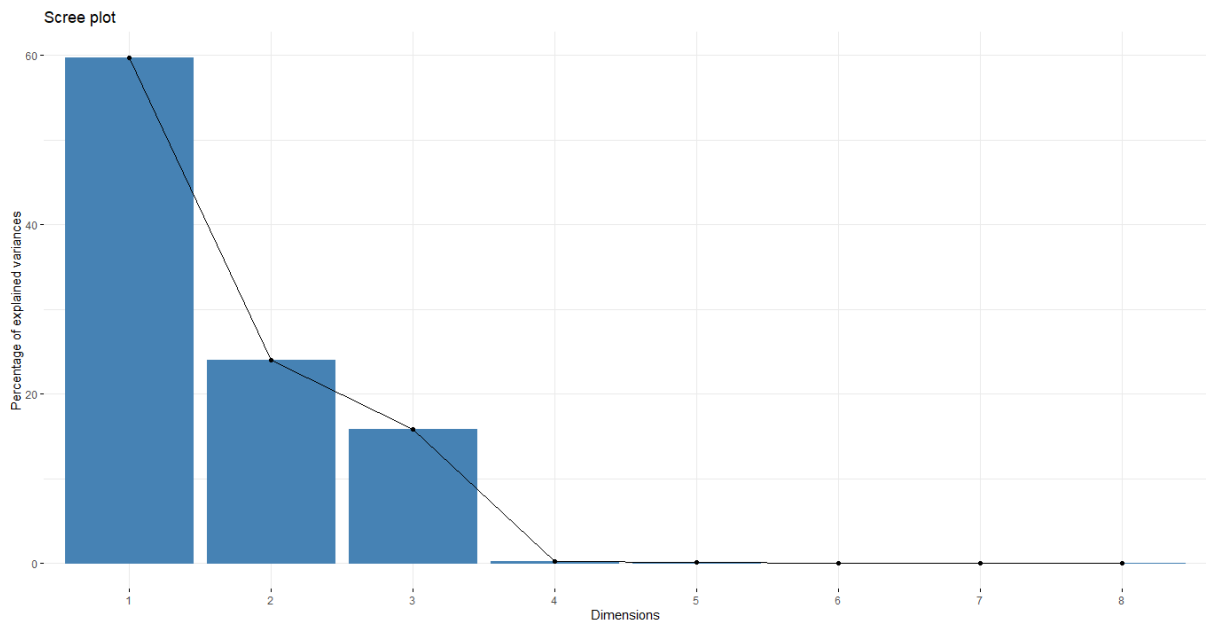
En el Gráfico 7 se observa que la mayoría de las variables están bien representadas, exceptuando las reservas internacionales y la tasa de interés de USA.

Por último, también puede observarse en el Gráfico 7 que destacan los años recesivos con observaciones atípicas que se mencionaron anteriormente. Puede notarse que el año 1990, que tiene la mayor cantidad de outliers en sus observaciones, está generando una gran influencia en el resultado del análisis de componentes principales ya que las variables mejor representadas en el espacio bidimensional son también las variables que fueron más relevantes en dicho año.

6.2. Análisis de componentes principales excluyendo los años recesivos en los que hay valores atípicos.

Una vez realizado el análisis de componentes principales con el conjunto de datos de años recesivos pero excluyendo los 4 años con outliers, se analiza su scree plot.

Gráfico 8: Scree plot del PCA sin los años recesivos con outliers del conjunto de datos.

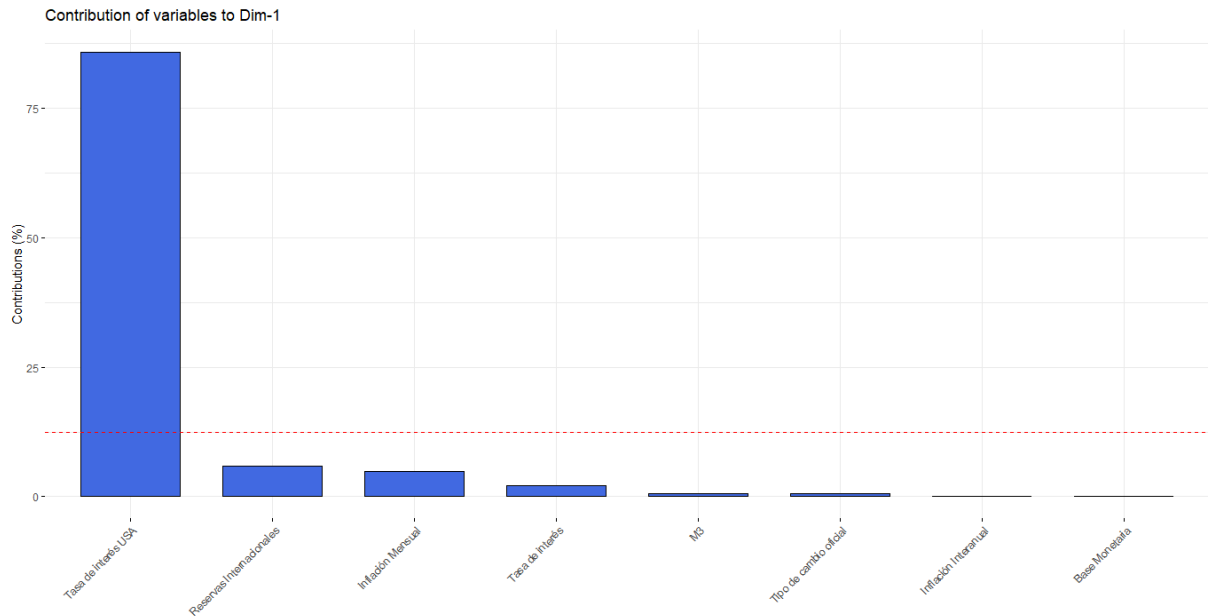


Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 8 se observa que ahora las primeras dos componentes principales contribuyen a la explicación de poco más del 80% de la varianza total. Este es un nivel aceptable para propósitos descriptivos como el del presente trabajo de investigación, por lo que este resultado nos facilita el análisis ya que tenemos un buen porcentaje de varianza explicada en solo 2 dimensiones, lo cual hará que la gráfica bidimensional resultante, además de ser sencilla de interpretar, aporte buena cantidad de información.

En segundo lugar se analizarán las variables que más contribuyen a la varianza explicada por las dos primeras componentes principales.

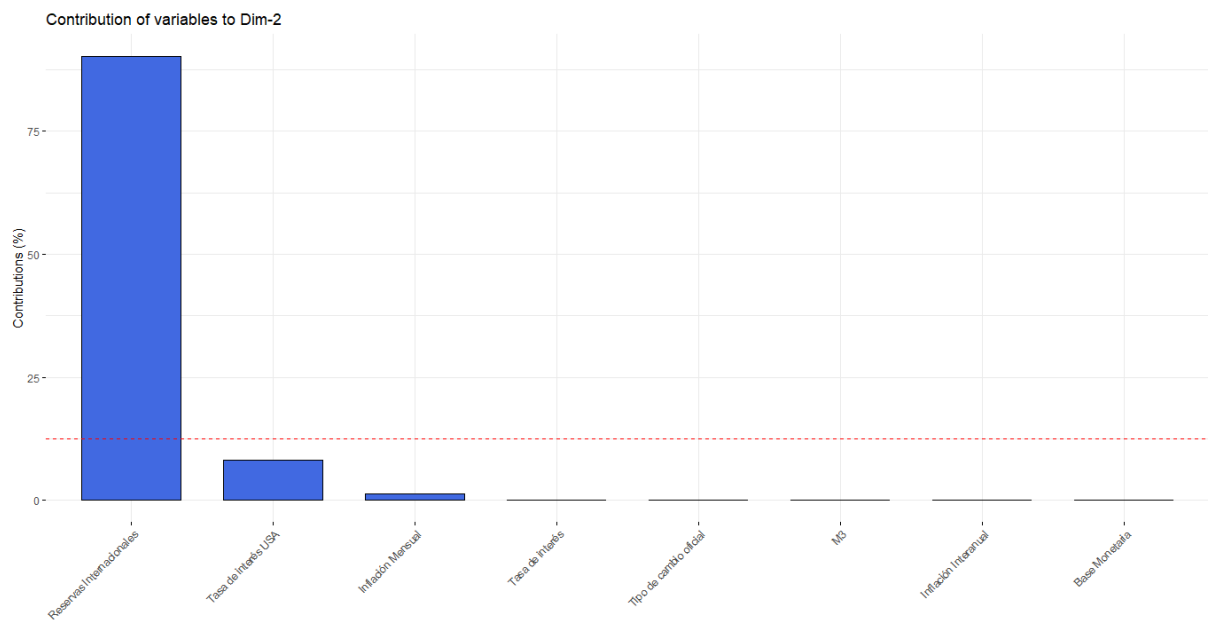
Gráfico 9: Contribución de las variables a la varianza explicada por la primer componente principal.



Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 9 se observa que la variable que más contribuye en la primer componente principal es la tasa de interés de Estados Unidos.

Gráfico 10: Contribución de las variables a la varianza explicada por la segunda componente principal.

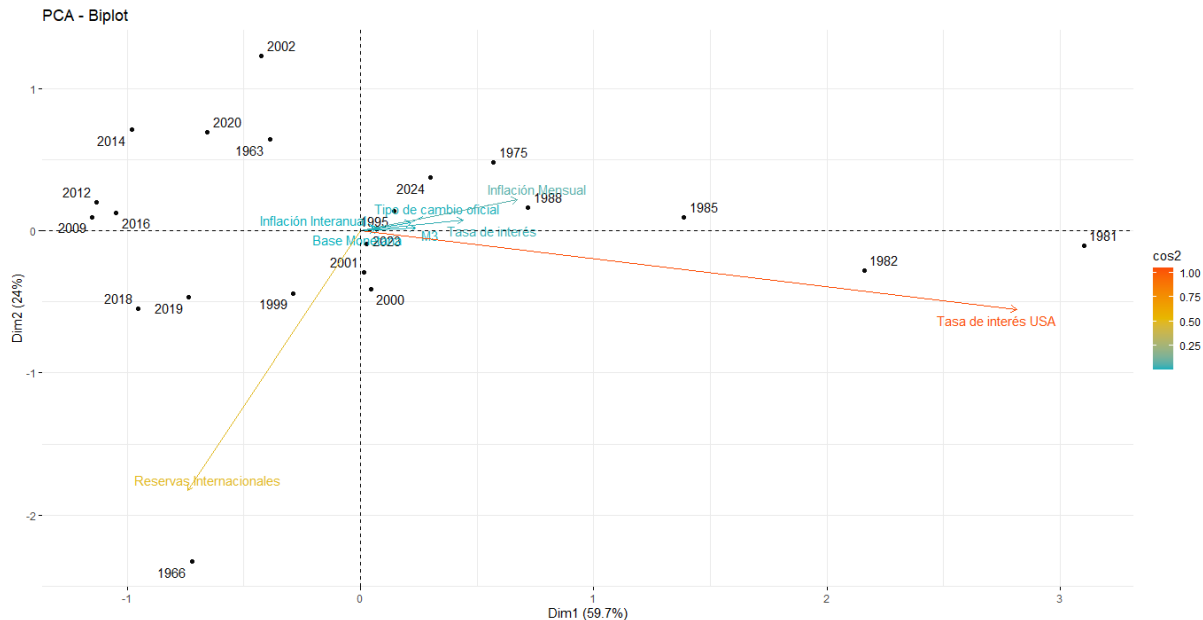


Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 10 se observa que la variable que más contribuye en la segunda componente principal son las reservas internacionales con las que cuenta el Banco Central de la República Argentina.

Finalmente analizaremos el biplot del PCA para analizar simultáneamente la información sobre las observaciones y las variables.

Gráfico 11: Biplot del PCA sin los años recesivos con outliers del conjunto de datos.



Fuente: Elaboración propia, mayo de 2025.

En el Gráfico 11 se observa que el primer componente principal tiene una asociación positiva grande con la tasa de interés de Estados Unidos. Por ende, podríamos decir que este componente mide principalmente la política monetaria de Estados Unidos, la cual parecería haber sido determinante en los años recesivos de Argentina de la década que inició en 1980. El segundo componente tiene una asociación negativa grande con las variaciones en las reservas internacionales. Podríamos decir entonces que este componente está relacionado al saldo de la balanza de pagos de Argentina.

En el Gráfico 11 se observa que la mayoría de las variables no se encuentran bien representadas en dos dimensiones. Las variables que ahora están mejor representadas son las reservas internacionales y la tasa de interés de USA. Estas dos variables eran las que, en el análisis de componentes principales realizado con todos los años recesivos incluidos los años con valores atípicos, resultaban mal representadas. Por ende, en cuanto a las variables representadas en el biplot estamos obteniendo un resultado que es exactamente el opuesto al anterior.



El primer análisis de componentes principales nos brinda la idea de que son muy relevantes las expansiones monetarias bruscas que lleva a cabo el BCRA porque el exceso de oferta de dinero va acompañado de grandes saltos inflacionarios y devaluatorios que generan caídas en el nivel de actividad. El segundo análisis de componentes principales nos brinda la idea de que los aumentos bruscos de tasas de interés en los Estados Unidos llevan a complicaciones para nuestro país como resultado del aumento del costo de la deuda internacional, la salida de capitales y la mayor pérdida de reservas en nuestro BCRA.

7. Predicciones para el presente año 2025

Teniendo en cuenta los datos que brindan el Banco Central de la República Argentina (BCRA) y la Federal Reserve Economic Data (FRED) del Federal Reserve Bank of St. Louis para los primeros meses del año 2025, se pueden utilizar los 6 modelos construidos para intentar predecir si el año 2025 será recesivo o no.

Teniendo en cuenta lo explicado en el apartado 3.2., los valores de las variables son:

- Agregado monetario M3: 95,35%
- Base monetaria: 139,44%
- Inflación mensual: 3,7%
- Tasa de interés nominal anual de los plazos fijos de 30 a 59 días: 33,19%
- Reservas internacionales en el BCRA: 14,04%
- Tasa de interés anual en Estados Unidos: 4,33%
- Tipo de cambio oficial: 26,58%
- Tasa de inflación interanual: 84,50%

Una vez que se les da como inputs los valores anteriores a los distintos modelos, las predicciones que brindan son:

- Modelo con algoritmo SVM y con resampling Cross-Validation: No
- Modelo con algoritmo SVM y con resampling Bootstrapping: No
- Modelo con algoritmo k-NN y con resampling Cross-Validation: No
- Modelo con algoritmo k-NN y con resampling Bootstrapping: No
- Modelo con algoritmo Random Forest y con resampling Cross-Validation: No
- Modelo con algoritmo Random Forest y con resampling Bootstrapping: No



Utilizando los datos de los primeros 5 meses del año 2025, la totalidad de los modelos construidos en el presente trabajo predicen que el año 2025 no será recesivo.

Observando los valores de las variables y analizándolas teóricamente, es razonable suponer que el año no va a terminar siendo recesivo. El agregado monetario M3 y la Base Monetaria se han expandido por encima de la inflación, lo cual en principio podría actuar como estímulo para la actividad económica. La TNA de los plazos fijos es similar a las estimaciones anuales de inflación para el año 2025, por lo que no afectaría de manera negativa a la actividad. La tasa de interés en USA no está en niveles muy altos como para que se produzca una salida de capitales que vaya a afectar a la economía de Argentina. El Banco Central de la República Argentina ha acumulado reservas en comparación con los primeros 5 meses del año anterior, lo cual aporta estabilidad a la economía, y el tipo de cambio oficial también ha evolucionado en la dirección favorable para el nivel de actividad.

Sin embargo, como se dijo anteriormente, sería adecuado continuar realizando predicciones con los modelos al cierre de cada mes cuando se publiquen los nuevos datos de las variables. La economía es muy dinámica y las variables van cambiando rápidamente a medida que el policy maker va aplicando diferentes políticas, por lo que es muy importante la actualización de las predicciones mes a mes para incorporar los últimos movimientos en las variables y sus efectos en la actividad económica.

8. Conclusiones

Ahora que ya se ha desarrollado el trabajo, se pueden responder las preguntas que motivaron su creación, se puede determinar si se cumplieron o no los diversos objetivos de la investigación y puede también confirmarse o rechazarse las hipótesis que se habían planteado de antemano.

La pregunta general del problema era: ¿Cuál de los 3 algoritmos -SVM, KNN ó Random Forest- permite construir un modelo predictivo con un sólido desempeño en la detección temprana de años recesivos en Argentina utilizando exclusivamente las tasas de variaciones de variables económicas nominales?. Para esta pregunta se había establecido la siguiente hipótesis sustantiva: “El algoritmo que permite construir un modelo predictivo con un sólido desempeño en la detección temprana de años recesivos en Argentina utilizando exclusivamente las tasas de variaciones de variables económicas nominales es Random Forest”. Esta hipótesis tuvo que ser rechazada ya que los modelos construidos con el algoritmo Random Forest no fueron los que mejores resultados dieron y, pese a alcanzar el umbral del 80% en la métrica “precisión”, no



llegaron a superar el umbral del 85% en la métrica “accuracy”. El modelo que alcanzó ambos umbrales en las métricas “accuracy” y “precisión” fue el modelo construido con el algoritmo k-NN y con método de resampling Cross-Validation. Este modelo logra predecir de manera correcta más del 92% de todos los años que se le pide que clasifique y, además, clasifica correctamente el 80% de los años que efectivamente son recesivos. Sin embargo, el modelo construido con Random Forest y con método de resampling Cross-Validation presentó buenos resultados y quedó como segundo mejor modelo con una exactitud en sus predicciones del 84,62% y una precisión para detectar años recesivos del 80%. Se había planteado la hipótesis sustantiva de tal manera porque al ser Random Forest un método de ensamble, se creía que su rendimiento iba a ser superior al de los demás modelos creados con los otros 2 algoritmos.

Las preguntas específicas que se buscaba responder a lo largo de la investigación eran:

- ¿Qué precisión tiene el modelo predictivo de años recesivos en Argentina construido con el algoritmo Support Vector Machine?
- ¿Qué precisión tiene el modelo predictivo de años recesivos en Argentina construido con el algoritmo k-Nearest Neighbor?
- ¿Qué precisión tiene el modelo predictivo de años recesivos en Argentina construido con el algoritmo Random Forest?
- ¿En qué grado las tasas de variaciones de las variables económicas nominales, sin tener en cuenta variables reales, permiten predecir años recesivos en Argentina?

Asociados a las anteriores preguntas específicas estaban los siguientes objetivos específicos:

- Determinar la precisión de los modelos construidos con el algoritmo Support Vector Machine para predecir años recesivos en Argentina.
- Determinar la precisión de los modelos construidos con el algoritmo k-Nearest Neighbor para predecir años recesivos en Argentina.
- Determinar la precisión de los modelos construidos con el algoritmo Random Forest para predecir años recesivos en Argentina.
- Determinar el grado en que las tasas de variaciones de las variables económicas nominales permiten predecir años recesivos en Argentina.

Una vez implementados los modelos, se obtuvieron las siguientes respuestas a las preguntas y objetivos anteriores:



- Los modelos construidos con el algoritmo Support Vector Machine arrojaron una precisión del 40% si se los entrenaba con el método de resampling Bootstrapping y del 0% si se los entrenaba con el método de resampling Cross-Validation.
- Los modelos construidos con el algoritmo k-Nearest Neighbor arrojaron una precisión del 80% si se los entrenaba con el método de resampling Cross-Validation y del 40% si se los entrenaba con el método de resampling Bootstrapping.
- Los modelos construidos con el algoritmo Random Forest arrojaron una precisión del 80% si se los entrenaba con el método de resampling Cross-Validation y del 60% si se los entrenaba con el método de resampling Bootstrapping.
- Utilizando únicamente las tasas de variaciones de las variables económicas nominales, y sin tener en cuenta variables reales, se pueden construir modelos que permiten predecir hasta el 80% de los años recesivos en Argentina.

Como ya se ha mencionado, las recesiones económicas tienen efectos muy nocivos para la sociedad: muchas personas pierden su trabajo; muchas empresas tienen que cerrar de manera temporal o, si la recesión es muy profunda, de manera permanente; aumenta la pobreza y la desigualdad, etc. Dado que la mayoría de las variables que se incluyen en los modelos son controladas en gran medida por el gobierno, los modelos aquí realizados son de gran importancia porque si predicen que el año terminará con un menor nivel de actividad que el año anterior, entonces el gobierno podría cambiar a tiempo sus políticas para evitar caer en recesión y así evitar los efectos nocivos sobre la sociedad mencionados anteriormente.

Por otro lado, los modelos también cuentan con una clara relevancia económica, ya que permiten obtener información de manera más oportuna sobre la actividad económica, brindando así a las empresas y demás agentes económicos una herramienta extra para la toma de decisiones. Por ejemplo, saber de manera anticipada que te encuentras en un año en el cual la actividad económica será menor a la del año anterior te puede evitar que te sobre stockees, que inviertas en proyectos que más adelante no encontrarán demanda, que compres a precios que en el futuro serán menores, etc.

Finalmente, este trabajo también tiene buenas implicancias académicas, ya que puede servir como base de un nuevo marco para quienes decidan realizar modelos predictivos de recesiones para Argentina en el futuro, ya que en este trabajo se utilizan por primera vez únicamente las tasas de variaciones de variables económicas nominales y se implementan, por primera vez



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO



FACULTAD DE
**CIENCIAS
ECONÓMICAS**

también, modelos de clasificación de recesiones con los algoritmos SVM, kNN y Random Forest.



9. Referencias bibliográficas

- Bañbura, Marta and Giannone, Domenico and Reichlin, Lucrezia, Nowcasting (November 30, 2010). ECB Working Paper No. 1275.
- Bañbura, M., Giannone, D., Modugno, M. y Reichlin, L. (2013). "Now-casting and the real-time data Flow", European Central Bank, Working Paper N° 1564.
- Bates, J. M., & Granger, C. W. J. (1969). "The combination of forecasts". *Operational Research Quarterly*, Vol. 20, No. 4., pp.451 - 468.
- Blanco, Emilio; D'Amato, Laura; Dogliolo, Fiorella; Garegnani, María Lorena (2017) : Nowcasting GDP in Argentina: Comparing the predictive ability of different models, *Economic Research Working Papers*, No. 74, Banco Central de la República Argentina (BCRA), *Investigaciones Económicas (ie)*, Buenos Aires.
- Clemen, R. (1989). "Combining forecasts: A review and annotated bibliography". *International Journal of Forecasting*, 5(4), 559–583.
- Camacho, M., Dal Bianco, M. y Martínez-Martín, J. (2013). "Short-Run Forecasting of Argentine GDP Growth", BBVA, Working Paper N° 13/14.
- D'Amato, L., Garegnani, L., & Blanco, E. (2024). Nowcasting de PIB: Evaluando las condiciones cíclicas de la economía argentina: Documentos de trabajo| 2015 | N 69. Documentos de trabajo.
- Hastie, T., Tibshirani, R., & Friedman, J. (2008). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). Stanford, CA: Stanford University.
- Llada, M. (2022). FORECASTING DEL PIB ARGENTINO. *Revista de Economía Política de Buenos Aires*, (25). 37-72.
- Sánchez Prieto, Alberto. (2023). Estudio de algoritmos de aprendizaje dinámico y online para aprendizaje profundo aplicado a la detección de tos en pacientes respiratorios. Universidad de Valladolid. Escuela Técnica Superior de Ingenieros de Telecomunicación.
- V. N. Vapnik, S. E. Golowich, and A. J. Smola, "Support vector method for function approximation, regression estimation and signal processing," in *Advances in Neural Information Processing Systems*. Cambridge, MA, USA: MIT Press, 1997, pp. 281–287.



UNCUYO
UNIVERSIDAD
NACIONAL DE CUYO



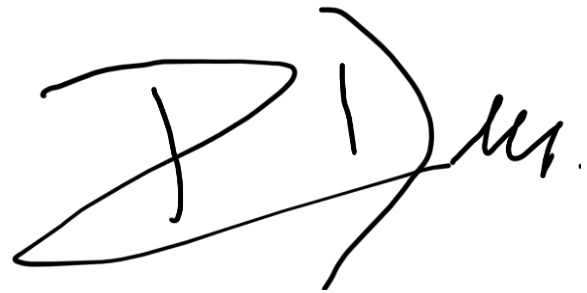
FACULTAD DE
**CIENCIAS
ECONÓMICAS**

Zheng, A. and Casari, A. (2018). Feature Engineering for Machine Learning: Principles and techniques for Data Scientists. O'Reilly Media, Inc., Sebastopol.

DECLARACIÓN JURADA RESOLUCIÓN 212/99 CD

El autor de este trabajo declara que fue elaborado sin utilizar ningún otro material que no haya dado a conocer en las referencias que nunca fue presentado para su evaluación en carreras universitarias y que no transgrede o afecta los derechos de terceros.

Mendoza, 1 de Diciembre de 2025



RODRIGO ALEJANDRO DICHARA

.....
Firma y aclaración

..... 33.116
Número de registro

..... 44.011.490
DNI