



Universidad de San Andrés

Departamento de Economía

Licenciatura en Economía

Tesis de grado

***Análisis y predicción de la dinámica
de la pobreza en Argentina***

Autor: Santiago Peci

Legajo: 30346

Mentor: Walter Sosa Escudero

Diciembre de 2022



Universidad de San Andrés

DEPARTAMENTO DE ECONOMÍA

ANÁLISIS Y PREDICCIÓN DE LA DINÁMICA
DE LA POBREZA EN ARGENTINA

Tesis de grado

Autor:
Santiago Peci

Legajo: 30346

Mentor:
Walter Sosa Escudero

Diciembre de 2022

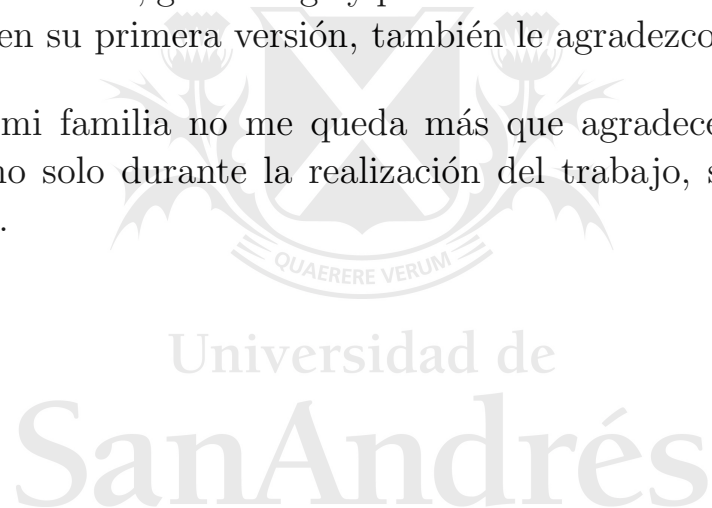
Agradecimientos

En primer lugar, me gustaría agradecer a mi mentor Walter Sosa Escudero. Este trabajo no habría sido posible sin su aporte, desde todos los conocimientos que me brindó en su curso de Big Data y Minería de Datos, hasta el seguimiento continuo en el armado del trabajo en estos últimos meses.

También me gustaría agradecer a todos los integrantes del Departamento de Economía de la Universidad de San Andrés. En especial a Mariano Tomassi y María Edo, quienes impulsaron mi interés por los estudios de la pobreza en la Argentina, y a Facundo Pernigotti y Joaquín Campabadal, quienes brindaron seguimiento dentro de la etapa inicial del proyecto.

A Felipe García Vasallo, gran amigo y parte fundamental del desarrollo técnico del trabajo en su primera versión, también le agradezco por su aporte.

Por último, a mi familia no me queda más que agradecerle por su apoyo incondicional no solo durante la realización del trabajo, sino a lo largo de toda la carrera.



Abstract

En este trabajo realizamos un análisis y predicción de la dinámica de la pobreza en Argentina durante los años 2016 y 2017. Se utiliza la Encuesta Permanente de Hogares (EPH) del Instituto Nacional de Estadística y Censos (INDEC).

En la primera parte del trabajo nos centramos en el análisis de la dinámica. Realizamos estadística descriptiva sobre la movilidad social de los individuos entre cada uno de los cuatro períodos de la muestra y observamos la correlación con el nivel educativo.

En la segunda parte del trabajo, predecimos la dinámica de la pobreza a nivel individuo con la utilización de modelos de Machine Learning. Luego generamos un análisis comparativo sobre las variables que los modelos toman como de mayor relevancia al realizar la predicción y las que se mencionan en la literatura econométrica como significativas.

Por último, realizamos otras dos predicciones excluyendo las variables directa e indirectamente relacionadas a los niveles de ingreso actuales de los individuos. Esto lo hacemos con el objetivo de analizar la importancia de las variables que no son de ingreso corriente (como lo pueden ser la composición del hogar y las características habitacionales) en la transición de la pobreza.

Índice

Abstract	2
1. Introducción	5
2. Literatura	7
3. Análisis preliminar	11
3.1. Armado de base de datos	11
3.2. Estadística descriptiva	12
4. Metodología de las predicciones:	16
5. Resultados	22
6. Conclusión	31
A. Anexo	34



Universidad de

Índice de figuras

San Andrés

1. Representación base de datos	11
2. Tasa de pobreza por trimestre (2016 - 2017)	12
3. Dinámica de la pobreza	13
4. Dinámica de la pobreza 2	14
5. Densidad de grupos	14
6. Nivel educativo por grupo	15
7. Tabla de predicciones	22
8. Matrices de confusión	23
9. Curvas ROC	24

10. Coeficientes Ridge	25
11. Tercer predicción	27
12. Tercer curva ROC	28
13. Árbol CART con variables ingreso	29
14. Árbol CART sin variables ingreso	30



Universidad de
San Andrés

1. Introducción

Este trabajo se centra en la dinámica de la pobreza en la Argentina. En primer lugar, realizamos un análisis general de la variación de la pobreza a lo largo del 2016 y 2017. Luego utilizamos una serie temporal con cuatro relevamientos por persona a lo largo de estos dos años para entrenar modelos que permitan predecir la dinámica de la pobreza a nivel individuo. Es decir, los modelos utilizan las características de cada persona en su segundo relevamiento de la base para predecir si van a encontrarse en situación de pobreza de ingresos dentro de un año a futuro (en su cuarto relevamiento). Finalmente, realizamos un análisis ex-post sobre la eficiencia de la predicción, la transición de las personas y las variables que los modelos toman como de mayor relevancia a la hora de predecir la situación de ingresos futura. Las predicciones son efectuadas con metodologías de machine learning (detalladas más adelante).

El estudio es de gran relevancia en varias dimensiones. En primer lugar, la clasificación de la vulnerabilidad de una persona puede ser mucho más precisa utilizando una serie temporal a mediano plazo. En este caso, al contar con 4 relevamientos por individuo en un plazo de 21 meses, podemos analizar la vulnerabilidad de cada persona según en cuántos de los 4 intervalos de tiempo estuvieron por debajo de la línea de la pobreza. Más allá de que seguimos teniendo el problema de la unidimensionalidad (solo se tiene en cuenta el ingreso), podemos generar un índice más informativo de la vulnerabilidad que no dependa tanto de situaciones contextuales relacionadas al ingreso.

En segundo lugar, en el caso de que la predicción de la dinámica sea efectiva, podemos realizar un análisis ex-post de los modelos predictores para generar un aporte significativo y desde una metodología distinta al debate sobre qué características de las personas u hogares son relevantes a la hora de predecir las probabilidades de salida y entrada a la pobreza.

Además, realizamos un ejercicio comparativo entre 3 predicciones. Primero utilizamos todas las variables incluidas en la encuesta permanente de hogares del INDEC. Luego, excluimos las variables que denotan directa e indirectamente un ingreso. Esto lo haremos con el objetivo de medir en cuánto varía la eficacia de la predicción y analizar que tanta dependencia tiene la dinámica de la pobreza en el mediano plazo sobre las variables de ingreso presentes

(en comparación con las de educación y características del hogar). Es decir, la hipótesis es la de que si la precisión de los modelos cae abruptamente al excluir las variables de ingreso corriente, éstas serían de suma importancia en la transición ya que los modelos no logran anticipar de manera precisa la dinámica de la pobreza sin utilizarlas.

Por último, el hecho de generar un modelo predictor que sea efectivo es relevante de por sí. Debido al alto poder predictivo que tienen los modelos de machine learning por fuera de la muestra de entrenamiento, se podrían extrapolar a series temporales que contengan los mismos datos pero de otros años y así predecir la transición de la pobreza en el mediano plazo y cuáles son los individuos con alta probabilidad de caer en la pobreza (y sus características).

Estas tres dimensiones mencionadas pueden ser de gran utilidad a la hora de realizar lineamientos de políticas públicas, no solo ayudando a generar un diagnóstico más acertado sobre el verdadero número de personas vulnerables en la Argentina y logrando una predicción efectiva de la transición de la pobreza, sino también detectando que variables son determinantes en esta transición.

2. Literatura

Existe un gran esfuerzo en la literatura que trata la transición de la pobreza y las causas de la pobreza crónica en la Argentina. Sin embargo, la gran mayoría de los estudios del tema están dados a partir de la econometría. Encontramos un único trabajo realizado por Lucchetti (2018) [5], en el cual se logra una predicción con machine learning (metodología LASSO) de la dinámica de la pobreza entre 2014 y 2015 para algunos países de América Latina. Nuestro objetivo es ampliar algunas líneas de este trabajo, buscando no solo realizar varias predicciones para años posteriores que incluyan diversas metodologías y distintos clusters de variables, sino también un análisis ex-post completo e informativo que pueda introducir conclusiones de características relevantes en la transición de la pobreza a la literatura.

Alejo y Garganta (2014) [6], por ejemplo, estudian la dinámica de la pobreza a partir de paneles sucesivos para el período 1997-2012, lo que les permite descomponer la pobreza (variabilidad del ingreso) en componentes permanentes/crónicos y transitorios. Terminan encontrando que “el factor transitorio está asociado principalmente con las características laborales, mientras que las cualidades estructurales y demográficas del hogar, y la educación del jefe y cónyuge del mismo, poseen incidencia fundamental sobre el componente crónico”. En principio, la idea del trabajo es analizar si logramos conclusiones similares a partir del análisis a posteriori.

Paz (2002) [3], por su parte, analiza la dinámica de la pobreza en el periodo 1998-2000 mediante la utilización de la EPH en forma de panel. Se clasifica la población observada según si se está por debajo de la línea de pobreza y su persistencia. Se pueden identificar lo que en el paper llama “pobres persistentes o pobres estructurales”, los cuales son todos aquellos que comenzaron y terminaron siendo pobres sin salir de la pobreza en ninguno de los periodos observados. Estos representan el 14,2% del total de hogares de la muestra y 50% de los alguna vez pobres. En términos de individuos, representan el 20,9% del total de individuos de la muestra y el 68,4% de los individuos alguna vez pobres. También se observa que el 43% de los hogares estuvo por lo menos una vez por debajo de la línea de pobreza. En este trabajo buscamos realizar un análisis similar para el periodo de nuestra muestra.

Por último, Paz (2002) también marca que las variables que mejor explican entrada y salida de la pobreza son la edad, la educación del jefe de hogar y la desocupación del jefe de hogar. La probabilidad de ser pobre está fuertemente relacionada con si el hogar o individuo fue pobre alguna vez y aún mayormente si fue en un periodo anterior inmediato. Además de las variables ya nombradas, la cantidad de perceptores de ingreso en el hogar aumenta la probabilidad de salir de la pobreza o no entrar en ella y en sentido contrario que el hogar se encuentre en NEA o NOA. En una continuación del trabajo a futuro nos gustaría incluir variables como la mencionada a las bases y ver si varían significativamente la eficacia de la predicción.

Por otro lado, Cruces y Wodon (2006) [2] estudian como varía el efecto de shocks macroeconómicos negativos según la vulnerabilidad del hogar. Utilizando el periodo 1995-2002 mediante la EPH, encuentran que los efectos de los shocks en la variabilidad del ingreso son heterogéneos dependiendo a qué quintil pertenece. Es decir, cuanto más rico sea el hogar, menor es el efecto negativo. Los más afectados son los hogares con mayor vulnerabilidad, como pueden ser los que contienen trabajadores informales, desempleados, migrantes recientes o inactivos. Además, observan un mayor impacto negativo en situaciones de crisis, aumentando los efectos en la pobreza. Es por esto que nuestra idea es utilizar la EPH en dos años en los que la variación interanual de la actividad económica no haya sido significativa, para que la transición de la pobreza detectada tenga que ver lo menos posible con shocks macroeconómicos temporales.

De esta forma, podemos observar que los avances de la literatura Argentina sobre la dinámica y la transición de la pobreza sigue un hilo conductor que es la utilización de la EPH en forma de datos de panel con el objetivo de realizar un seguimiento de un mismo individuo/hogar a lo largo del tiempo. Teniendo en cuenta que la EPH no es ideal para los estudios longitudinales, los autores buscan adaptarse a la escasez de datos y utilizan esta base como un second best. De esta forma, la estructura de datos que utilizan es similar a la que buscamos lograr, pero en la literatura se la aproxima a partir de una diferente metodología (no existe demasiada aproximación al problema a partir del machine learning para Argentina).

Aparte del estudio de la dinámica de la pobreza, también revisamos la lite-

ratura Argentina sobre la pobreza crónica y su aproximación académica. El concepto de pobreza crónica sirve para caracterizar un grupo demográfico que sufre de privaciones en múltiples planos de manera persistente, lo que les deja en una situación de alta vulnerabilidad. La salida de la pobreza en estos casos es compleja y generalmente no es suficiente el crecimiento económico, ya que las características de los individuos en estas condiciones impiden aprovechar en su totalidad los shocks positivos y las políticas públicas inclusivas.

Para Hulme y Shepherd (2003)[1], un pobre crónico es aquel individuo que padece privaciones de capacidad significativas, de por lo menos cinco años. Agregan que para la comprensión de esta definición se deben tener en cuenta ciertos aspectos. En primer lugar, la cantidad de períodos bajo la línea de pobreza. A mayor cantidad, menor es la probabilidad de salida en el próximo periodo y mayor es la probabilidad de una transmisión intergeneracional. En Yaqub (2000) [7], se encuentra que si un individuo ya padeció cinco años de pobreza es muy probable que lo siga siendo por el resto de su vida.

En segundo lugar, la privación de ingreso o consumo temporal no es suficiente para determinar un pobre crónico. Un indicador de pobreza multidimensional suele ser más informativo en el estudio de la vulnerabilidad. Esto es debido a que el ingreso o consumo es un factor fluctuante en el corto plazo. A mayor cantidad de dimensiones que un individuo tenga carencias, menor probabilidad de salir de la pobreza.

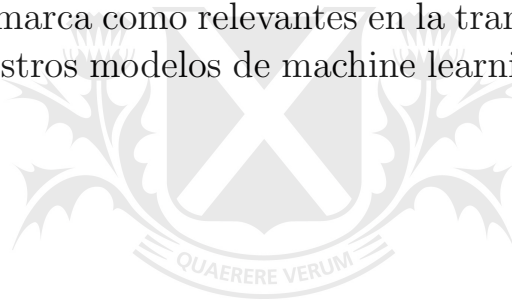
Por último, tener en cuenta que el estudio de pobreza crónica es distinto en términos relativos o absolutos. Este último se usa sobre todo en países en desarrollo. En Yaqub (2003) [8], se argumenta que los pobres crónicos medidos como los del último quintil por ingresos tienen igual de difícil la salida de la pobreza o hasta más compleja que los pobres crónicos medidos de manera absoluta. Por ejemplo, Gasparini et al. (2019) [4] utiliza, dada la limitación de datos en Argentina, un concepto de pobreza crónica relativa tomando como pobres crónicos al 10% más vulnerable de la población.

Para la salida de la pobreza crónica en la literatura de países en desarrollo se encuentra una correlación con la adquisición de terrenos, nivel de educación y nivel al inicio de escolaridad. Para la movilidad hacia abajo hay una alta correlación con el aumento en la cantidad de personas en un hogar y su número de dependientes. También analizaremos si hay relevancia de estas

variables en nuestros modelos predictores.

En la práctica se tienen dos alternativas para estimar la pobreza crónica, el “components approach” y el “spells approach”. El primero se fija en la variable ingreso o consumo enfocándose en la profundidad de la pobreza. El segundo se centra en la entrada y salida de la pobreza orientando el concepto de pobreza crónica al de pobreza persistente. El análisis descriptivo que realizaremos está más enfocado en este último.

En resumen, primero realizaremos un “spells approach” de manera similar a la metodología utilizada en Jorge Paz (2002) [3]. Es decir, llevaremos a cabo un análisis descriptivo de la transición de la pobreza en Argentina a lo largo de 2016 y 2017 para caracterizar a los individuos según su persistencia. Luego abordaremos un análisis comparativo entre las variables que la literatura econométrica citada marca como relevantes en la transición de la pobreza Argentina y los que nuestros modelos de machine learning toman como mejores predictores.



Universidad de
San Andrés

3. Análisis preliminar

3.1. Armado de base de datos

Para realizar el análisis preliminar mencionado, armamos nuestra base en panel a partir de las Encuestas Permanentes de Hogares (EPH-C) realizadas por el INDEC. Esta base va a contener datos de dos clusters de individuos en un plazo total de 21 meses. El primer cluster va a tener observaciones del segundo y tercer trimestre de 2016 y 2017, y el segundo cluster del tercer y cuarto trimestre de estos mismos años. Esto es debido a que la EPH renueva el 25 % de la muestra en cada onda trimestral, por lo que te permite mantener el seguimiento de un mismo individuo u hogar en el plazo máximo consecutivo de cuatro trimestres. Además, en la EPH-Continua se utilizan datos para hogares en dos trimestres consecutivos, luego se retiran temporalmente en los dos trimestres siguientes, y finalmente se vuelven a incorporar en dos trimestres adicionales sucesivos.

La idea era buscar dos años en los que la variación interanual de la actividad económica no haya sido significativa, para que la transición de la pobreza detectada tenga que ver lo menos posible con shocks macroeconómicos temporales, y a su vez evitar el periodo en el que los datos del INDEC no eran del todo confiables. Es por esto que nos decidimos por el periodo que va desde el segundo trimestre del 2016 al cuarto trimestre del 2017.

A continuación, un diagrama de como queda la base de datos en panel:

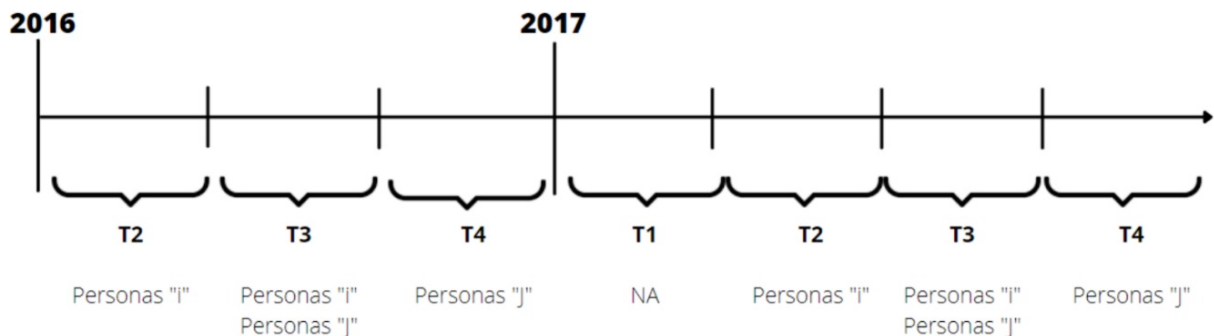


Figura 1: Representación base de datos

Luego, una vez que tenemos la base con los datos de individuos que reportaron ingreso en los cuatro periodos, realizamos una limpieza y ciertas

modificaciones con el objetivo de realizar la predicción. En primer lugar, reemplazamos ciertos valores sin sentido para que sigan siendo utilizables. Por ejemplo, cambiamos las edades e ingresos negativos por el valor promedio de las personas que cumplen el mismo rol en el hogar. Luego agregamos columnas que indican si los individuos son o no pobres en cada periodo analizando la canasta básica promedio trimestral, el ITF y la tabla de equivalencias. Finalmente, realizamos una selección preliminar de variables explicativas que creemos a priori puedan ser relevantes en la predicción de pobreza a futuro (aproximadamente 70 variables en total).

Luego de realizar esta limpieza, controlamos que la base de datos siga siendo representativa (y no hayamos incurrido en algún sesgo). Para esto calculamos la tasa de pobreza de hogares por trimestre y la comparamos con la reportada por el INDEC. En la figura 2 observamos que los resultados son muy similares:

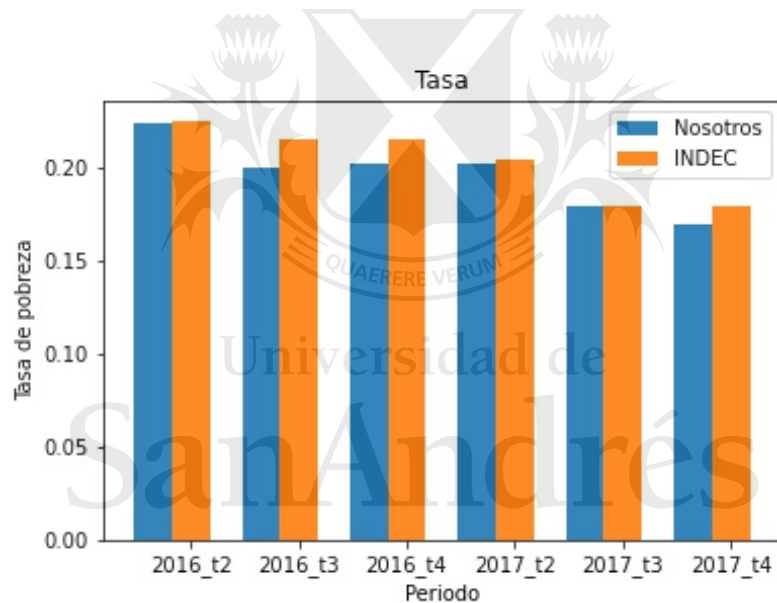


Figura 2: Tasa de pobreza por trimestre (2016 - 2017)

3.2. Estadística descriptiva

Una vez que ya tenemos nuestra base preliminar armada, y comprobamos que sea representativa, comenzamos con la estadística descriptiva. En primer lugar, en la figura 3 realizamos una primera aproximación a la dinámica de la pobreza. En azul se mide el porcentaje de los pobres en un trimestre que logró salir de la pobreza al periodo siguiente. En naranja el porcentaje de no

pobres que cayó por debajo de la línea de la pobreza al trimestre siguiente.

En la primera barra azul observamos que el 22 % de los individuos que eran pobres en el segundo trimestre de 2016, no lo fueron en el tercero. Por otro lado, la primer barra naranja denota que el 10 % de los no pobres en el segundo trimestre de 2016 terminó cayendo en la pobreza al semestre siguiente. Notamos que la mayor salida de la pobreza se logró entre el segundo y el tercer trimestre del 2017, logrando que más de un 25 % de los pobres logren ubicarse por encima de la línea de pobreza de ingresos al periodo siguiente.

En promedio, entre el 20 y el 25 % de los pobres logran salir de la pobreza entre períodos, mientras que entre el 10 y 15 % de los no pobres cae en la pobreza.

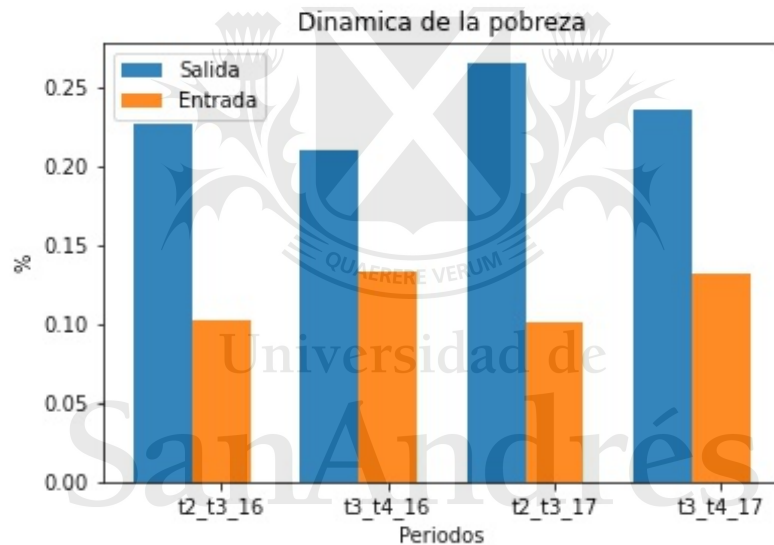


Figura 3: Dinámica de la pobreza

De todas formas, dado que bastante menos del 50 % de los individuos eran pobres entre 2016 y 2017, este gráfico no muestra necesariamente que la tasa de pobreza de individuos disminuye considerablemente trimestre a trimestre. Dado que esta representación puede derivar en esta confusión, en la figura 4 también representamos la dinámica entre periodos en valores absolutos.

En la primera barra azul se marca que el 9 % de la población total era pobre en el segundo trimestre de 2016, y no lo fue en el tercero (salió de la pobreza). La barra naranja marca que el 7 % cayó en la pobreza entre estos periodos. Se puede observar que del tercer al cuarto trimestre de ambos años hubo

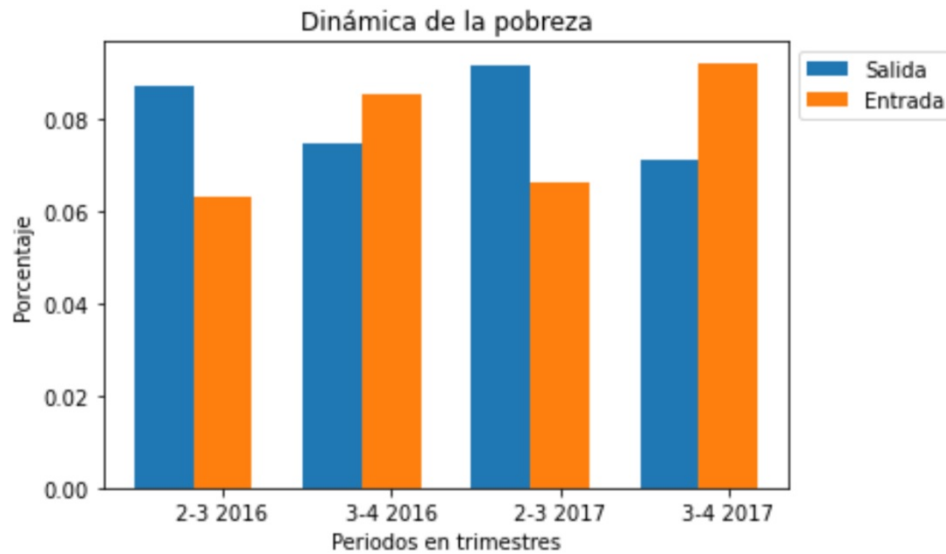


Figura 4: Dinámica de la pobreza 2

una mayor cantidad de personas que entraron a la pobreza, mientras que del segundo al tercero fue mayor la cantidad de los que salieron.

Luego, realizamos dos gráficos más. La figura 5 clasifica a los individuos como crónicos, muy vulnerables, vulnerables o no vulnerables según si estuvieron 4, 3 o 2, 1 o 0 períodos por debajo de la línea de la pobreza respectivamente. La figura 6, por su parte, discrimina el nivel educativo entre estos grupos formados.

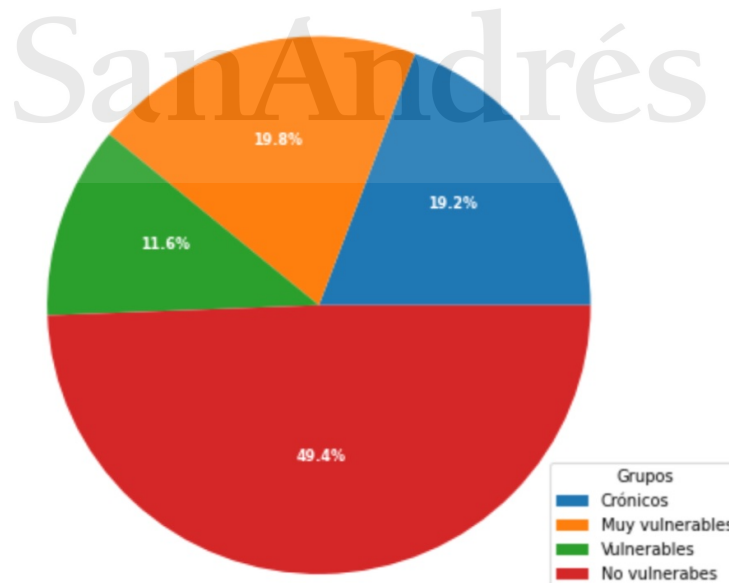


Figura 5: Densidad de grupos

En la figura 5 notamos que más del 50% de los individuos fue pobre en al

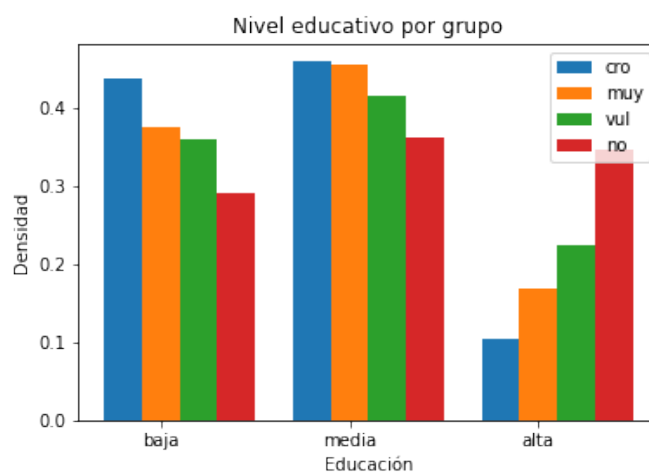


Figura 6: Nivel educativo por grupo

menos uno de los 4 periodos, y que casi el 20 % estuvo por debajo de la línea de la pobreza en todos ellos. En la figura 6 lo que hacemos es discriminar qué porcentaje dentro de cada uno de los grupos formados corresponde a los distintos niveles educativos. Es decir, dentro del 19,2 % de la población total que se mantuvo en situación de pobreza en los cuatro periodos (pobres crónicos o persistentes), 43 % tiene nivel educativo bajo, 47 % medio y 10 % alto (barras azules en la Figura 6). De manera contraria, dentro del 49,4 % que nunca cayó en la pobreza, 28 % tiene nivel educativo bajo, 37 % medio y un 35 % alto (barras rojas en Figura 6).

Formamos cada nivel educativo a partir de una división arbitraria sobre el nivel educativo más alto que el individuo cursa actualmente o alguna vez cursó:

- **Bajo:** Jardín/preescolar, Primario 3 o EGB
- **Medio:** Secundario o Polimodal
- **Alto:** Terciario, Universitario, Posgrado universitario o Educación especial.

Este gráfico nos pareció muy interesante e informativo ya que muestra muy bien como la vulnerabilidad de las personas según la clasificación de densidad realizada tiene una correlación negativa muy marcada con el nivel educativo.

4. Metodología de las predicciones:

En esta parte del trabajo realizamos las 3 predicciones interanuales para cada individuo, las cuales utilizan las características de su segunda observación para predecir si van a encontrarse en situación de pobreza de ingresos o no en su última observación. Por ejemplo, para el caso de un individuo tipo “i”, con sus características del trimestre 3 de 2016 se busca predecir si va a ser pobre o no en el trimestre 3 de 2017 (ver Figura 1 para mayor comprensión).

Como ya comentamos, estamos lanzando tres predicciones. En la primera incluimos todas las variables de la EPH y la dummy generada de pobreza actual. En la segunda excluimos las variables definidas como “variables ingreso” en el Diseño de Registro realizado por el INDEC. La idea es, a partir de la variación de la eficacia de la predicción, determinar en que medida la transición de las personas en el mediano plazo es producto del nivel de ingreso presente, y en que medida de las variables que pasan a tomar relevancia en las nuevas predicciones.

La hipótesis es la que de que si al eliminar las variables de ingreso perdemos todo el poder predictivo, podríamos concluir que estas son fundamentales en la transición de la pobreza interanual de un individuo, ya que no se logra predecir esta dinámica de manera precisa al excluirlas.

Como la segunda predicción sigue teniendo ciertas variables que de manera indirecta denotan ingresos, realizamos la tercer y última predicción con variables que hacen referencia exclusivamente a las características habitacionales, nivel educativo y composición del grupo familiar.

Vale aclarar que estas últimas dos predicciones son con el fin exclusivo de realizar un análisis ex post informativo en términos comparativos de eficacia de los modelos y coeficientes relevantes. Es decir, el modelo predictor no es relevante de por sí ya que no se va a ganar poder predictivo.

Entonces, las variables predictoras serían todas las que contiene la EPH y seleccionamos como relevantes al limpiar la base de datos (primero con y luego quitando las variables de ingreso), y la variable a predecir sería una dummy de si el individuo va a ser o no pobre en su última observación (un

año a futuro). Es decir, no se predice el nivel de ingreso futuro y luego se lo clasifica como pobre o no pobre, sino que se predice directamente si el individuo va a ser pobre o no. ($Y_i = \{0, 1\}$)

Los métodos son los más utilizados en la literatura de machine learning: Lasso, Ridge, Elastic Net, Cart, Bagging, Boosting, Random Forest, Knn, Logit, Discriminante Lineal y SVM. Vamos a definir por cross-validation los parámetros óptimos para cada uno, y luego seleccionar el modelo que logra una mayor tasa de aciertos en la predicción de la dinámica.

Incluiremos una breve explicación simplificada de que es lo que hace intuitivamente cada método.

1. **Lasso:**

$$R_l(\beta) = \sum_{i=1}^n (y_i - x_i\beta)^2 + \lambda \sum_{s=2}^p |\beta_s| \quad , \quad \lambda > 0 \quad (1)$$

La ecuación 1 es la función objetivo que el modelo Lasso busca minimizar. Notamos que el primer término penaliza la falta de ajuste (igual que el método MCO), pero ahora se le agrega un segundo término que penaliza la cantidad de variables predictoras. Esto logra que ciertos coeficientes sean llevados hacia cero (con respecto a las de MCO), generando en ellos un sesgo, pero haciendo que el modelo termine ganando en materia predictiva dada la disminución de la varianza.

El parámetro de penalización (λ o Shrinkage) óptimo que minimice el error de pronóstico es elegido por la metodología Cross-Validation. Notamos que esta penalización, la elección del parámetro y el cómputo de error por cross validation son cruciales, ya que evitan que se genere el problema de overfit del modelo y logran que este mantenga una buena performance por fuera de la muestra (lo que cumple nuestro objetivo de que sea utilizable para predecir pobreza futura).

2. **Ridge:**

$$R_r(\beta) = \sum_{i=1}^n (y_i - x_i\beta)^2 + \lambda \sum_{s=2}^p (\beta_s)^2 \quad , \quad \lambda > 0 \quad (2)$$

La lógica es la misma que en lasso pero ahora la penalización está ele-

vando al cuadrado, lo que genera que las soluciones de los coeficientes tiendan a ser interiores y no necesariamente de esquina. Es decir, ahora no sucede que muchos coeficientes toman valores iguales o muy cercanos a cero, por lo que genera un mayor interpretabilidad con respecto a Lasso. Se puede demostrar que existe λ tal que ridge predice mejor que MCO.

3. Elastic Net:

$$R_{en}(\beta) = \sum_{i=1}^n (y_i - x_i\beta)^2 + \lambda_1 \sum_{s=2}^p (\beta_s)^2 + \lambda_2 \sum_{s=2}^p |\beta_s| \quad , \quad \lambda > 0 \quad (3)$$

En la ecuación 3 podemos ver que elastic net utiliza el termino MCO, el segundo termino de lasso (ecuación 1) y el segundo termino de ridge (ecuación 2). Es decir, lo que hace es tomar cada uno de los modelos. De lasso utiliza la elección de variables predictoras significativas y la reducción del modelo y de ridge la resolución de un problema técnico de inestabilidad por agrupamiento muy común en bases como la eph que cuentan con variables muy correlacionadas. Tiene gran poder predictivo.

4. Vecinos Cercanos (KNN):

Es un método muy sencillo que va a pedir si la persona va a ser pobre o no según si las K observaciones más similares son pobres o no. La cantidad de vecinos a utilizar (K) se elige por cross-validación. En bases de datos con muchas variables como la EPH puede tener problemas prácticos al definir la cercanía entre observaciones.

5. CART:

Cart es un método muy utilizado para capturar efectos no lineales de las variables predictoras en la variable a predecir. En particular, se generan predicciones a partir del armado de un árbol de decisión que contiene en cada uno de sus nodos particiones generadas a partir de valores de las variables predictivas. En los nodos terminales se encuentran las predicciones efectivas. Notamos que el árbol tiende a poner en los primeros nodos a las variables más relevantes, por lo que podemos realizar un análisis ex

post de la predicción y ver que variables fueron detectadas por el modelo como las más relevantes en términos de minimizar error de pronóstico (y cómo se relacionan entre ellas).

Es un método muy intuitivo (que nos va a permitir interpretar qué variables son relevantes), pero no suele ser de los mejores predictores debido a que puede llegar a tener una alta varianza. Bagging, Random Forest y Boosting buscan resolver ese problema (más allá de que no son tan interpretables o intuitivos).

6. **BAGGING:**

Modelo que sigue la misma lógica que CART, pero aplican un método llamado Bootstrap que predice el árbol “B” veces pero tomando muestra con reemplazo de las observaciones, lo que potencialmente disminuye la varianza. Esto no sucede cuando hay un predictor muy fuerte, por lo que los árboles ponderados en la predicción final siguen siendo muy parecidos.

7. **Random Forest:**

Lo que hace es el mismo procedimiento de Bagging pero en cada una de las “B” iteraciones elegir “ $m < p$ ” predictores, lo que genera que se resuelva el problema de alta varianza incluso en situaciones en las que haya un predictor muy fuerte.

8. **Boosting:**

Es un algoritmo que genera “M” modelos predictores débiles a partir de árboles CART que tengan pocas ramas, pero ponderando en mayor medida las observaciones que fueron mal predichas en la iteración anterior. Esto permite derivar una predicción a partir del promedio ponderado de las predicciones de los M árboles (ponderado por el error de pronóstico de cada modelo). Suele ser una técnica de gran performance predictiva, pero con baja interpretabilidad).

9. SVM:

Este método es un algoritmo que sitúa a las observaciones en un plano de acuerdo a sus variables predictoras y luego genera un hiperplano separador que realiza la predicción según si las observaciones quedaron de un u otro lado. Es muy bueno en términos predictivos, pero por supuesto que es imposible de interpretar un plano con tantas variables (como lo sería uno que utilice las de la eph).

10. Logit:

Utiliza la siguiente función:

$$P(Y = 1/X) = \frac{e^{x\beta}}{1 + e^{x\beta}} \quad (4)$$

donde β es el vector de los “p” coeficientes y es estimado por máxima verosimilitud. Luego con esto computa la probabilidad de que un individuo sea pobre dentro de un año dadas las características presentes, y utiliza el clasificador de Bayes:

$$P(Y = 1/X) \geq 1/2 \rightarrow \hat{Y}_i = 1$$
$$P(Y = 1/X) < 1/2 \rightarrow \hat{Y}_i = 0$$

Sabemos de antemano que algunos de los modelos pueden incurrir en ciertos problemas prácticos. Por ejemplo, cuando la cantidad de observaciones es mayor a la cantidad de variables y además hay una alta correlación entre predictores, Ridge tiende a funcionar mejor que Lasso (debido a que este último tiende a eliminar varias de estas variables quedándose solo con unas pocas). Elastic net, a su vez, suele predecir bien bajo estas circunstancias, reduciendo la dimensionalidad y eligiendo correctamente los grupos de variables significativas. KNN y CART también tienen sus limitaciones prácticas en esta predicción. Vecinos Cercanos tiende a perder eficiencia cuando el problema tiene tantas dimensiones debido a que comienza a tener dificultades para definir la cercanía. Por su parte, Cart, tiene un problema de alta varianza que puede ser corregido por modelos como Random Forest y Bagging (como ya mencionamos).

Luego de realizar la predicción, vamos a analizar la eficacia de los modelos y detectar cual y con qué parámetros predice mejor comparando su accuracy (i.e.: precisión). Elegimos esta medida ya que representa la fracción de predicciones que el modelo realizó correctamente. Como nuestra prioridad es predecir bien, sin tener aversión a cierto tipo de errores, la accuracy resulta una buena medida para jerarquizar que modelo predice mejor. En otras palabras, penalizamos igual que un modelo prediga que alguien va a caer en la pobreza y que no lo termine haciendo, a que prediga que alguien va a salir de la pobreza y tampoco lo haga. Además, como es calculada a partir de observaciones no utilizadas en la muestra de entrenamiento, indica ajuste por fuera del modelo entrenado (no manifiesta overfitting).

Vale aclarar que mientras más equilibrados estén los datos a clasificar, más informativa será la medida de accuracy. Es decir, en un país que cuente con tasas de pobreza menores al 5%, lograr una accuracy por encima del 95% sería tan sencillo como clasificar a todos los individuos como no pobres. En Argentina, por su parte, no es para nada sencillo lograr una precisión por encima del 90% (ya que cuenta con niveles de pobreza bastante más altos que el 10%).

En el caso de que la predicción del mejor modelo haya sido efectiva, no solo tendríamos un modelo que pueda predecir la pobreza interanual, sino que también vamos a poder realizar un análisis a posteriori de algunos de los modelos mencionados que resulte informativo y detectar si efectivamente encontramos las correlaciones que la literatura existente marca como relevantes en la transición y la dinámica de la pobreza.

En Lasso, Ridge y Elastic Net, buscaremos analizar y describir qué coeficientes no parecen ser significativos, cuales sí lo son, y qué signo (incidencia) tienen. En los modelos CART, analizar las ramas generadas y el orden de los nodos.

5. Resultados

Como habíamos mencionado, entrenamos a cada uno de los modelos con una serie distinta de parámetros y luego calculamos el ECM promedio de cada iteración por cross validation. En la figura 7a observamos cual es el mejor parámetro para cada modelo y que error de pronóstico contiene cuando se utilizan todas las variables (incluidas las de ingreso). Luego, en la figura 7b repetimos el procedimiento pero excluyendo las variables de ingreso. Recordamos que tanto el ECM (Error Cuadrático Medio) como la accuracy (ACC) representan ajuste por fuera del modelo entrenado.

	modelo	ecm	parámetro	accuracy
0	Lineal	0.230699	0.4	0.769301
0	KNN	0.199818	3.0	0.800182
0	SVM	0.244021	3.0	0.755979
0	Bagging	0.212534	3.0	0.787466
0	Logit	0.174992	10000.0	0.825008
0	CART	0.085377	20.0	0.914623
0	RandomForest	0.099606	20.0	0.900394
0	Boosting	0.033000	10.0	0.967000
0	Lasso	0.162277	100000.0	0.837723
0	Ridge	0.159552	1000.0	0.840448

(a) Predicciones con variables ingreso

	modelo	ecm	parámetro	accuracy
0	Lineal	0.217984	0.01	0.782016
0	KNN	0.270360	17.00	0.729640
0	SVM	0.238874	1.00	0.761126
0	Bagging	0.227066	3.00	0.772934
0	Logit	0.212534	1000.00	0.787466
0	CART	0.175295	20.00	0.824705
0	RandomForest	0.158644	20.00	0.841356
0	Boosting	0.093854	10.00	0.906146
0	Lasso	0.220103	1000.00	0.779897
0	Ridge	0.210718	0.10	0.789282

(b) Predicciones sin variables ingreso

Figura 7: Tabla de predicciones

En primer lugar, notamos que en ambas predicciones todos los modelos cuentan con una accuracy por encima del 70 %, teniendo como promedio un 83.5 % cuando se utilizan las variables de ingreso, y un 79.2 % cuando no. Es decir, en promedio no parecería caer significativamente la efectividad de los modelos cuando no se utilizan las variables de ingreso. Más allá de que la accuracy promedio es bastante estándar (no muy elevada), observamos que hay mucha diferencia predictiva entre modelos.

El que mejor predice en ambos casos es el de Boosting con parámetro igual a 10. Con las variables de ingreso logra una accuracy extremadamente alta del 96 %, con una tasa de falso pronóstico de pobreza (fp) del 5 % y de falsos pronóstico de no pobreza (fn) del 2 % (ver en figura 8a la matriz de confusión). Este modelo resulta de alta relevancia, ya que si se cuenta con

toda la información proporcionada por la EPH, se puede pronosticar la tasa de pobreza interanual con una eficacia promedio mayor al 95 %.

En cuanto a la predicción sin las variables de ingreso, se logra una accuracy del 90 %, con una tasa de falso pronóstico de pobreza (fp) del 15 % y de falsos pronóstico de no pobreza del 8 % (ver en la figura 8b la matriz de confusión). Esto resulta también muy relevante, ya que la predicción continúa teniendo una eficacia promedio mayor al 90 %, lo que respalda la literatura econométrica que afirma que la probabilidad de salida o entrada a la pobreza en el mediano plazo depende no sólo del monto de ingresos actual, sino también de otras dimensiones que vamos a estar analizando más adelante.

	Negativos	Positivos
Negativos	1921	47
Positivos	76	1256

(a) Matriz de confusión con ingreso

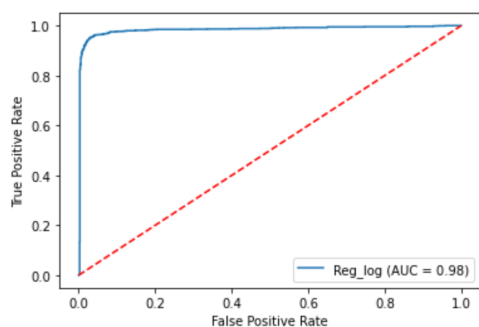
	Negativos	Positivos
Negativos	1807	159
Positivos	203	1134

(b) Matriz de confusión sin ingreso

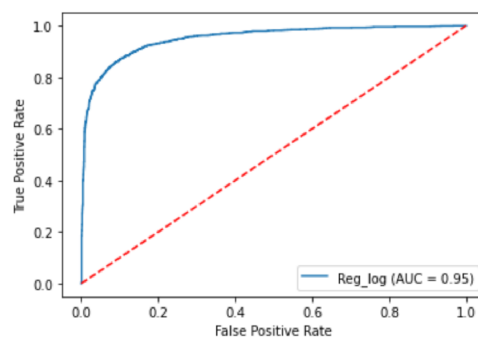
Figura 8: Matrices de confusión

Notamos que las filas de la matriz de confusión denotan el pronóstico realizado, y las columnas la verdadera realización. Es decir, en la figura 8a, el cuadrante de abajo a la izquierda marca que 76 individuos fueron pronosticados como pobres, y no lo terminaron siendo. El cuadrante de abajo a la derecha, por su parte, marca que 1256 personas fueron pronosticadas como pobres y sí lo terminaron siendo.

Para obtener una interpretación gráfica de estos resultados, utilizamos la curva ROC, la cual muestra como varía la tasa de verdaderos positivos en relación a la tasa de falsos positivos a medida que va variando el umbral de decisión. En los casos extremos de la curva, todas las observaciones son pronosticadas como 0 o 1, dependiendo de sí se penaliza infinitamente el error de tipo 1 o de tipo 2. En los puntos intermedios, se puede notar que la concavidad de la curva muestra como crece en mayor medida la tasa de verdaderos positivos con respecto a la de falsos positivos (caso contrario el modelo sería peor que lanzar una moneda). Entonces, a mayor área por debajo de la curva (AUC), mejor es la predicción.



(a) Curvas ROC con ingreso



(b) Curvas ROC sin ingreso

Figura 9: Curvas ROC

Observando las figuras 9a y 9b, notamos que en el gráfico de las curvas ROC con ingreso la AUC es de 0.98, mientras que en el de la curva sin ingreso la AUC es de 0.95 (en ambos casos extremadamente alta).

En el caso de los modelos Ridge, que logran una eficacia del 84% y 78%, buscamos analizar los coeficientes. La Figura 10a muestra los coeficientes de mayor relevancia en la predicción con las variables de ingreso (aquellos que son mayores a 0.10 en valor absoluto), mientras que la Figura 10b muestra los coeficientes más importantes en la predicción sin variables de ingreso (mayores a 0.70 en valor absoluto).

Variable	Coefficiente
Categoría de inactividad	0,1046
Cuantos ambientes/habitaciones tiene la vivienda en total	-0,1104
Combustible utilizado para cocina	0,1525
¿En los últimos tres meses, las personas de este hogar han vivido de subsidio o ayuda social?	-0,1306
Cantidad de miembros del hogar	0,343
Número de decil del ingreso total del hogar del total EPH	-0,1783
Número de decil del ingreso per cápita familiar del total EPH	-0,3381
Dummy de Pobreza	0,2276

(a) Coeficientes Ridge con Ingreso

El detalle de cada variable se encuentra en el anexo del trabajo.

Más allá de que la interpretabilidad de estos coeficientes es limitada, dado que están sesgados por el shrinkage mencionado, podemos desprender algunas conclusiones interesantes ya que las variables de la EPH están estandarizadas.

En primer lugar, notamos que los coeficientes de mayor valor en la predicción con variables de ingreso están relacionados con las características laborales y

Variable	Coefficiente
Condición de actividad laboral	-0,7028
¿Durante el tiempo de desocupación hizo algún trabajo / changa?	-0,7351
¿Su trabajo le proporciona vacaciones pagas?	0,9084
¿En los últimos tres meses, las personas de este hogar han vivido de subsidio o ayuda social?	-0,9696
Material de la cubierta exterior del techo	0,9494
¿Tiene baño / letrina?	2,1031
¿En los últimos tres meses, las personas de este hogar han vivido algún alquiler (por una vivienda, terreno, oficina, etc.) de su propiedad?	0,8288
¿En los últimos tres meses, las personas de este hogar han vivido de ganancias de algún negocio en el que no trabajan?	2,8681
¿En los últimos tres meses, las personas de este hogar han vivido de intereses o renta por plazos fijos o inversiones?	2,2844
¿Tuvieron otros ingresos en efectivo (limosnas, juegos de azar, etc)?	0,8047

(a) Coeficientes Ridge sin ingreso

Figura 10: Coeficientes Ridge

el nivel de ingreso del hogar/individuo. Estas son: 1)El ingreso total familiar, 2)El número de decil del ingreso per cápita familiar, 3)La dummy generada de si el individuo es pobre o no actualmente, 4)La categoría de inactividad del individuo. Este resultado es consistente con la conclusion mencionada en el trabajo de Alejo y Garganta (2014), quienes mencionaban que el factor transitorio está asociado principalmente con las características laborales. También va en línea con Paz (2002), quien marcaba que la probabilidad de ser pobre está fuertemente relacionada con si el hogar o individuo estuvo en la pobreza recientemente. Sin embargo, se destacan algunas variables significativas e interesantes que no se mencionaban en la literatura como relevantes en la transición. Algunas de estas son: 1)Cuántos ambientes/habitaciones tiene la vivienda en total, 2)Qué combustible es utilizado en el hogar para cocinar, 3)La cantidad de miembros en el hogar.

En cuanto a la predicción Ridge sin las variables de ingreso, notamos que ahora se le da una mayor preponderancia a muchas variables que antes no la tenían. Una de las variables de mayor coeficiente es la de si el hogar cuenta con baño o letrina, mientras que se destaca como variables relevante la composición de la cubierta exterior del techo del hogar.

Finalmente, realizamos un análisis sobre las ramas superiores de los árboles de decisión generados por el modelo CART, los cuales lograron una accuracy del 91 % y 82 %. En la figura 13 observamos el árbol del modelo con variables de ingreso, y en la figura 14 el árbol del modelo sin las variables de ingreso.

Como mencionamos en la sección 4, los arboles CART tienen las ventajas de ser de muy intuitivos e interpretables, y las de representar muy bien los efectos no lineales de las variables. La manera de interpretarlos es la siguiente:

1) Se comienza por el nodo superior, donde el modelo indica cual cree que es la variable más relevante. En la Figura 13, por ejemplo, notamos que la variable más relevante es la dummy de pobreza en la actualidad, realizando la partición inicial según si estás en situación de pobreza actualmente o no ($POBRE \leq 0,5$). Vemos que el árbol también aclara cuantas de las personas en la muestra cumplen con el requisito (Rama *True*), es decir que no son pobres, y cuantas no (Rama *False*).

2) La interpretación continua con la misma lógica en los nodos subsiguientes, hasta llegar a un nodo terminal donde se terminan realizando las clasificaciones.

Dado la no linealidad del árbol de decisión, se debe aclarar como y cuando influyen el resto de las variables. Continuamos con el ejemplo del árbol CART con variables de ingreso (Figura 13). En el caso de que el individuo no sea pobre, la segunda variable más relevante es la del número de decil del ingreso per cápita familiar ($DECCFR \leq 3,5$). Caso contrario, la más importante es la que indica si en los últimos 3 meses las personas del hogar han recibido ayuda social o un subsidio ($V5 \leq 0,5$). Luego se podría realizar un análisis más minucioso de la incidencia de todas las variables, pero en definitiva notamos nuevamente la alta preponderancia de las variables de ingreso.

En la figura 14, en cambio, vemos que la variable más destacada en la predicción sin variables de ingreso pasa a ser la de si se recibe en el hogar ayuda social o subsidio del gobierno y/o iglesia ($V5 \leq 0,5$). Luego se incluye la cantidad de miembros del hogar en el caso negativo ($IX_TOT \leq 3,5$), y si tiene cobertura médica o no en caso afirmativo ($CH08 \leq 3,50$). Nuevamente dos variables no mencionadas que parecerían tener relevancia en la transición.

Cabe destacar que el modelo tomó como importantes a un grupo de variables que no estaban en la sección de variables de ingreso en el Diseño de Registro y Estructura realizado por el INDEC, pero que están denotando de forma indirecta un ingreso. Es por esto que, como ya habíamos mencionado, realizamos la tercer predicción que incluya sólo variables de identificación, característi-

cas habitacionales del hogar y características del miembro del hogar. De esta forma intentamos aislar lo máximo posible el efecto ingreso sobre la eficacia de la predicción, y analizar como varía la accuracy. Los resultados se pueden observar en la figura 11.

	modelo	ecm	parámetro	accuracy
0	Lineal	0.230699	0.01	0.769301
0	KNN	0.275507	15.00	0.724493
0	SVM	0.235241	1.00	0.764759
0	Bagging	0.228883	3.00	0.771117
0	Logit	0.234030	0.01	0.765970
0	CART	0.199818	20.00	0.800182
0	RandomForest	0.176809	20.00	0.823191
0	Boosting	0.123524	10.00	0.876476
0	Lasso	0.218286	10.00	0.781714
0	Ridge	0.233121	100000.00	0.766879

Figura 11: Tercer predicción

Notamos que la eficacia de las predicciones se mantiene bastante alta. El mejor modelo predictor fue Boosting, logrando un 87% de accuracy. Esto es apenas 3 puntos porcentuales menos que la predicción anterior, y 9 menos que la predicción que incluye todas las variables (incluido la dummy de pobreza actual). A partir de este resultado se puede apreciar la importancia de las características habitacionales y educativas en la transición de la pobreza.

La limitación de este trabajo surge a partir de que, dado el acotado seguimiento de los individuos en las EPH, el análisis se queda en un punto intermedio entre la transición de la pobreza en el corto plazo y el componente crónico de largo plazo. La hipótesis a evaluar cuando se cuente con una base más prolongada, es la de que a mayor sea el periodo de la transición a predecir, mayor preponderancia tomarían las variables no laborables/ingreso.

A continuación la curva ROC de la tercer predicción:

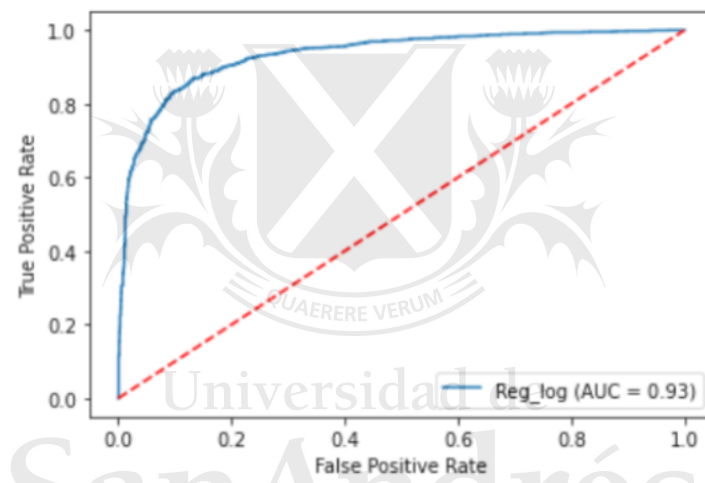


Figura 12: Tercer curva ROC

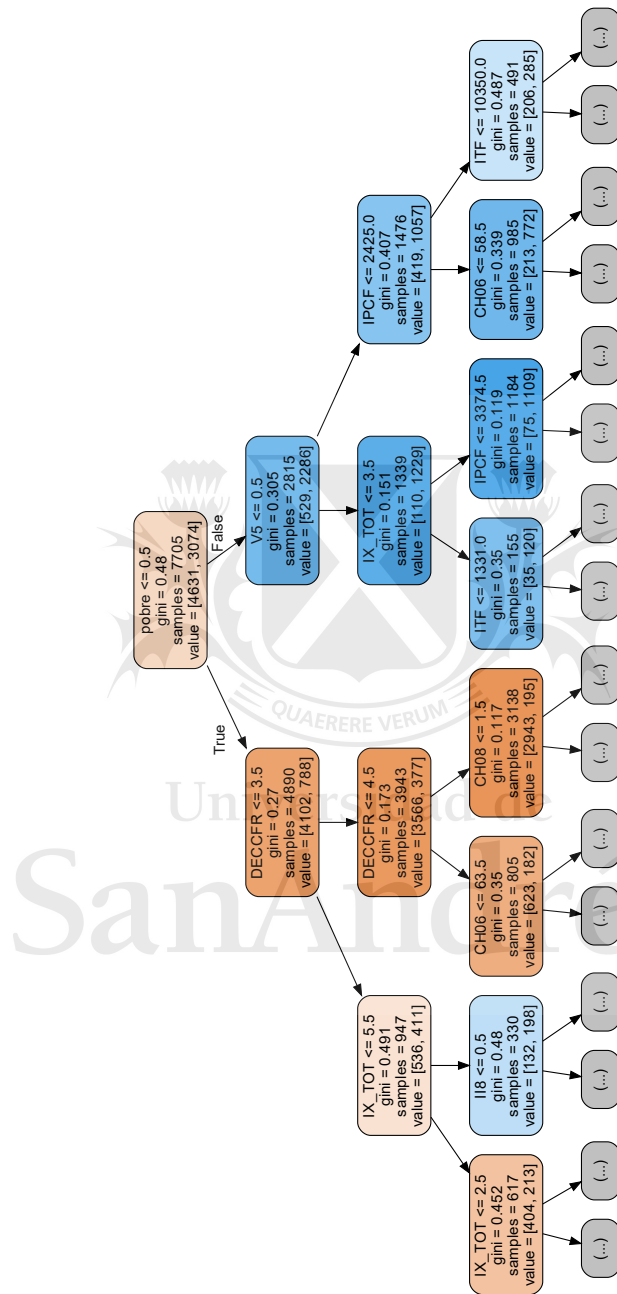


Figura 13: Árbol CART con variables ingreso

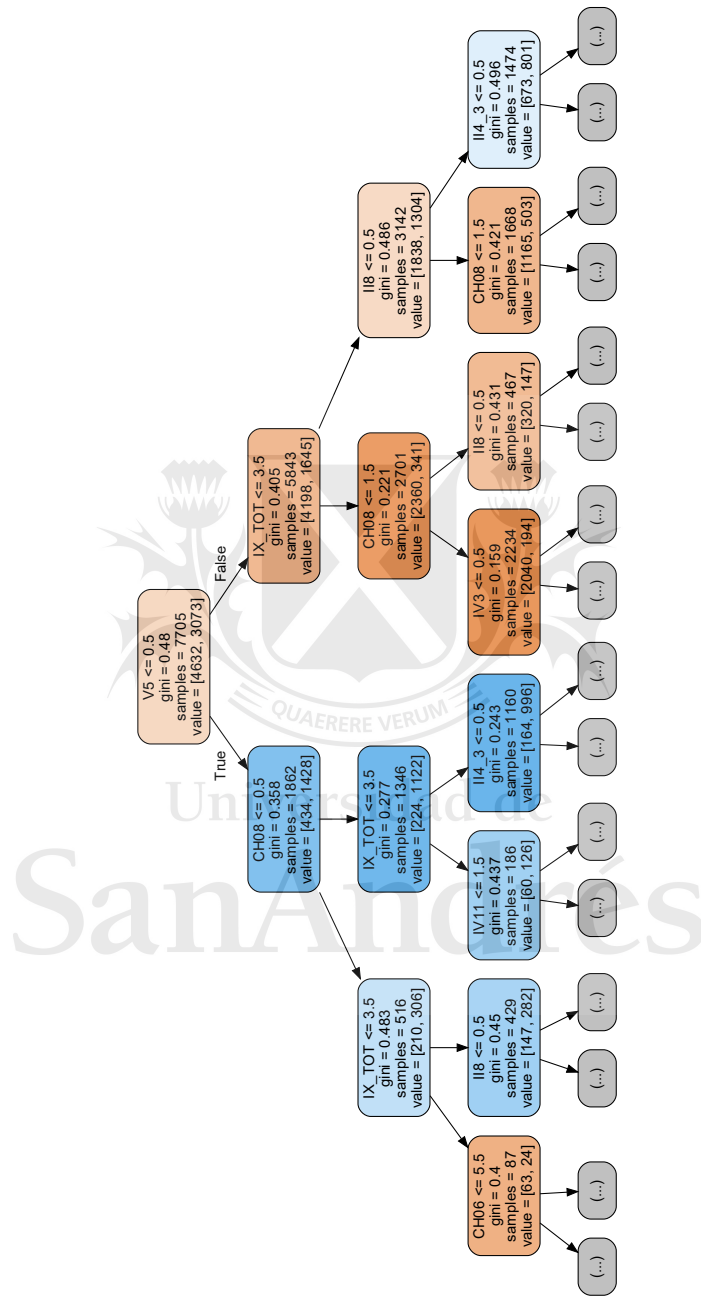


Figura 14: Árbol CART sin variables ingreso

6. Conclusión

A lo largo del trabajo logramos analizar y predecir la dinámica de la pobreza en Argentina para los años 2016 y 2017.

Comenzamos con la estadística descriptiva de la base de datos generada, a partir de la cual obtuvimos algunos resultados interesantes y significativos. Por ejemplo, encontramos que más del 50 % de los individuos habría estado al menos uno de los 4 periodos por debajo de la línea de la pobreza y que 19,2 % fueron pobres en todos los periodos. Estos resultados son similares a los hallados por Paz (2002) para el periodo 1998-2000, donde el 20,9 % de los individuos tampoco habían podido salir de la pobreza en ningún periodo.

Además, también logramos ilustrar con claridad la dinámica de la pobreza entre periodos y la correlación entre los niveles educativos y la vulnerabilidad de las personas. Los resultados de este último gráfico también van en línea y respaldan los trabajos de Paz (2002) y Alejo y Garganta (2014) donde se hace énfasis en la incidencia de la educación tanto en el factor crónico como en el transitorio.

En la segunda parte del trabajo realizamos las predicciones sobre la dinámica de la pobreza a nivel individuo con la utilización de modelos de Machine Learning. Más allá de que la eficacia promedio de los modelos no fue particularmente buena, se logró una precisión extremadamente alta en algunos modelos. Como mencionamos al principio del trabajo, estos resultados son relevantes ya que, dado el poder predictivo de los modelos por fuera de la muestra, se pueden utilizar estas metodologías para datos actuales de la EPH y realizar predicciones altamente eficaces. Es necesario aclarar que se podrían potenciar aún más con la incorporación de variables macroeconómicas externas (y sus proyecciones).

Luego, generamos el análisis comparativo entre las variables relevantes en los modelos de Machine Learning y las mencionadas en la literatura. A partir de los outputs de los modelos, logramos validar la mayor parte de los puntos a los que hace referencia los autores citados, e incorporar algunos puntos o variables puntuales que parecerían tener relevancia en la transición a mediano plazo (interanual). Por ejemplo, tanto en el modelo Ridge como en el CART

(con variables de ingreso), se valida que las variables de mayor relevancia en la transición son las de características laborales y del nivel de ingreso del hogar/individuo (puntos a los que lógicamente se hace mucho énfasis en la literatura).

Más allá de esto, los modelos logran identificar algunas variables significativas que no habían sido encontradas en la literatura. Se destacan las relacionadas con la composición del hogar: cuántos ambientes tiene la vivienda, qué combustible es utilizado para cocinar, y otras como si se tiene cobertura médica o si se recibe alguna ayuda social o subsidio.

Luego, al realizar las predicciones sin las variables que denotan ingreso, logramos validar parcialmente la hipótesis. Es decir, las predicciones de varios modelos se mantuvieron con niveles de precisión relativamente altos, pero se generó una caída en comparación a las predicciones que utilizan todas las variables. Interpretamos esto como que, más allá de haber quitado las variables que se marcaban como de mayor relevancia en la predicción inicial (las de ingreso), las variables restantes (nivel educativo, composición del hogar, etcétera) siguen teniendo una incidencia importante en la transición.

En las predicciones sin variables de ingreso también logramos desprender otras variables relevantes que no habían sido mencionadas: si el hogar cuenta con baño propio, cual es la composición del techo, entre otras.

En conclusión, se puede lograr una predicción extremadamente alta de la pobreza interanual (a nivel individual) utilizando modelos de Machine Learning, y las variables que los modelos marcan como de mayor relevancia son las que denotan directamente un ingreso. Más allá de esto, encontramos que también se encuentran en un segundo plano ciertas variables de composición del hogar y nivel educativo, a partir de las cuales también se puede predecir la pobreza con niveles de precisión relativamente altos.

La hipótesis a analizar en una continuación del trabajo sería la de que, a medida que la predicción sea a mayor longitud entre los periodos sobre los que se predice la transición, las variables de composición del hogar y del nivel educativo pasarían a tener gradualmente mayor importancia, dejando a las de ingreso en segundo plano. Lamentablemente la diagramación actual de la EPH, dada la inexistencia de datos de un mismo individuo en un plazo mayor

al de 4 ondas temporales, impide este análisis.



Universidad de
San Andrés

A. Anexo

Detalle de variables EPH:

ESTADO: Condición de actividad: 0 = Entrevista individual no realizada (no respuesta al cuestionario individual) 1 = Ocupado 2 = Desocupado 3 = Inactivo 4 = Menor de 10 años.

PP10C: ¿Durante el tiempo de desocupación hizo algún trabajo / changa? 1= Sí 2= No

PP07G1: Su trabajo le proporciona vacaciones pagas? 1 = Si 2 = No

V4: La cubierta exterior del techo es de: 1. membrana / cubierta asfáltica 2. baldosa / losa sin cubierta 3. pizarra / teja 4. chapa de metal sin cubierta 5. chapa de fibrocemento / plástico 6. chapa de cartón 7. saña / tabla / paja con barro / paja sola. 9. N/S. Departamento en propiedad horizontal

IV8: ¿Tiene baño / letrina? 1 = Sí 2 = No

V8: ¿En los últimos tres meses, las personas de este hogar han vivido algún alquiler (por una vivienda, terreno, oficina, etc.) de su propiedad? 1 = Sí 2 = No

V9: ¿En los últimos tres meses, las personas de este hogar han vivido ganancias de algún negocio en el que no trabajan? 1 = Sí 2 = No.

V18: Tuvieron otros ingresos en efectivo (limosnas, juegos de azar, etc.) 1 = Sí 2 = No.

CAT_INAC: Categoría de inactividad: 1 = Jubilado / Pensionado 2 = Rentista 3 = Estudiante 4 = Ama de casa 5 = Menor de 6 años 6 = Discapacitado 7 = Otros.

IV2: ¿Cuántos ambientes/habitaciones tiene la vivienda en total? (sin contar baño/s, cocina, pasillo/s, lavadero, garaje).

II8: Combustible utilizado para cocinar: 01 = Gas de red 02 = Gas de tubo / garrafa 03 = Kerosene / leña / carbón

V5: ¿En los últimos tres meses, las personas de este hogar han vivido de subsidio o ayuda social (en dinero) del gobierno, iglesias, etc.? 1 = Sí 2 = No

IX.Tot: Cantidad de miembros del hogar.

DECIFR: Número de decil del ingreso total del hogar del total EPH.

DECCFR: Número de decil del ingreso per cápita familiar del total EPH

CH08: ¿Tiene algún tipo de cobertura médica por la que paga o le descuentan? 1 = Obra social (incluye PAMI) 2 = Mutua / prepaga / servicio de emergencia 3 = Planes y seguros públicos 4 = No paga ni le descuentan 9 = Ns/Nr 12 = Obra social y mutua / prepaga / servicio de emergencia 13 = Obra social y planes y seguros públicos 23 = Mutua / prepaga / servicio de emergencia/ Planes y seguros públicos 123 = obra social, mutua/prepaga/ servicio de emergencia y planes y seguros públicos

Pobre: Dummy generada que indica si el individuo se encuentra en pobreza de ingresos (=1) o no (=0).

Referencias

- [1] Hulme D. and Shepard A. Conceptualizing chronic poverty. *World Development*, (31):403–423, 2003.
- [2] Cruces G.A. and Wodon Q. Risk-adjusted poverty in argentina: measurement and determinants. *Financiamiento para el Desarrollo 182, Naciones Unidas Comisión Económica para América Latina y el Caribe (CEPAL)*, 2006.
- [3] Paz J. A. Una introducción a la dinámica de la pobreza en la argentina. *CEMA Working Papers: Serie Documentos de Trabajo.*, 226, 2002.
- [4] Gasparini L., Tornarolli L., and Gluzmann P. El desafío de la pobreza en argentina: diagnóstico y perspectivas. *CEDLAS, CIPPEC y PNUD.*, pages 143–165, 2019.
- [5] Lucchetti L., Corral P., Ham A., and Garriga S. Lassoing welfare dynamics with cross-sectional data. *Policy Research Working Paper. World Bank.*, 2008.
- [6] Alejo O.J. and Garganta S. Pobreza crónica y transitoria: evidencia para argentina 1997-2012. *Centro de Estudios Distributivos, Laborales y Sociales; Documentos de trabajo (CEDLAS)*, 175:1–36, 12-2014.
- [7] Yaqub S. Intertemporal welfare dynamics: extents and causes. in ‘globalization: new opportunities, new vulnerabilities’ workshop. *Brookings Institution, Carnegie Endowment.*, 2000.
- [8] Yaqub S. Chronic poverty: scrutinising patterns, correlates, and explorations. *CPRC Working Paper 21. Manchester: IDPM, University of Manchester.*, 2003.

