



**Asignación Universal por Hijo y Trabajo Infantil: Un
análisis predictivo.**

Martín Aberg Cobo

Mentor: Walter Sosa Escudero

Introducción

El trabajo infantil es una realidad para millones de niños/as alrededor del mundo. Actualmente, según la Organización Internacional del Trabajo (OIT) son 152 millones de niños que trabajan incluso antes de aprender a leer o escribir. Es considerada trabajo infantil, según la OIT, toda actividad o trabajo que priva a los niños de su infancia, su dignidad y que sea perjudicial para su desarrollo físico y psicológico. Una de las principales causas del trabajo infantil es la pobreza y los factores asociados a la vulnerabilidad socioeconómica del hogar. Las familias en situación de pobreza no suelen disponer de los recursos necesarios para comprar alimentos, ropa y en algunos casos no pueden acceder a una vivienda digna. Resultando en que los niños se vean obligados a contribuir a la economía familiar para subsistir. El trabajo infantil aparece en numerosos sectores, entre ellos, el sector agropecuario, la construcción, la minería, la pesca, el comercio, la vía pública y el ámbito doméstico.

A nivel internacional, la erradicación del trabajo infantil es uno de los objetivos establecidos en la Agenda 2030 de las Naciones Unidas. Muchos países han adoptado en los últimos años políticas o programas que buscan prevenir y desalentar el trabajo infantil. En Argentina, según la Ley N° 26.390, el trabajo de las personas menores de 16 años está prohibido en todas sus formas, exista o no relación de empleo contractual y más allá de que esté remunerado o no. Siguiendo esta línea, existen diversos programas y transferencias monetarias que buscan la protección social relacionada mediante la seguridad de ingresos con el foco en los niños. Este trabajo se centra en la Asignación Universal por Hijo (AUH), un programa de transferencias monetarias condicionadas en la Argentina.

La siguiente investigación está dividida en dos partes. En la primera, se analiza el rol que cumple la AUH en materia de trabajo infantil a la luz de otras investigaciones. Luego, se establecen los distintos mecanismos por los cuales este programa logra directa e indirectamente reducir el trabajo infantil. En la segunda parte, se busca definir el grupo poblacional el cual debería recibir la AUH para prevenir el trabajo infantil. Para esto, se utilizan dos encuestas distintas realizadas por el INDEC; la “Encuesta de Actividades de Niños, Niñas y Adolescentes 2016-2017” (EANNA) y la

“Encuesta Permanente de Hogares” (EPH) más reciente (2021, tercer trimestre) al momento de la elaboración de esta tesina. Por un lado, la EANNA es la única encuesta en nuestro país que busca captar el trabajo infantil; la cual se realizó en 2004 y en 2016-2017. En cambio, la EPH es una encuesta que se realiza todos los años, pero no incluye esa variable.

El objetivo es predecir los niños, niñas y adolescentes (NNyA) que caerán en lo que se considera trabajo infantil. Se utiliza un modelo de *machine learning* conocido como *Extreme Gradient Boosting* (XGBoost) y un modelo de *deep learning* conocido como *Neural Networks* (redes neuronales) para predecir cuales son los hogares en donde hay trabajo infantil. La gran ventaja de utilizar este tipo de modelos, es que se “entrenan” (concepto que se profundizará más adelante) con una muestra en particular, para poder predecir fuera de esta. En este caso, el modelo se “entrena” con la EANNA, ya que esta contiene la variable de interés, trabajo infantil; para luego realizar una predicción en la EPH. La EANNA la llamaremos entonces muestra de entrenamiento.

De esta manera, se logra identificar el trabajo infantil en una encuesta como la EPH que no logra captarlo en un principio. Este análisis predictivo sería relevante para poder conocer el tipo de hogares que tienen al menos un NNyA que trabaja. Esto sería útil para el Estado, ya que podría definir con mayor exactitud la características e ubicación de los hogares para poder otorgar transferencias monetarias como la AUH con el fin de disminuir el trabajo infantil.

PARTE I

Transferencias Monetarias Condicionadas

La pobreza se entiende como la privación de salud mental, educación, seguridad, aspiraciones, entre otras. Esta investigación se centra en el análisis del trabajo infantil como privación de una niñez adecuada. En particular, la inasistencia a la escuela suele ser muy alta debido a las tareas domésticas y los trabajos infantiles. Realizar tareas del hogar o incluso trabajar desde que uno es muy pequeño no deja tiempo para actividades recreativas que todo infante necesita realizar. Yuningsih Y., Salamah U., Nurwati N. (2018) destacan la importancia de los derechos durante la infancia. En particular, los autores afirman que el desarrollo en la edad temprana de las personas es fundamental y por eso se debe garantizar tanto salud y educación como un espacio seguro en el que los niños y niñas puedan jugar como tales. Establecen que la educación funciona como una forma de protección social que colabora con el objetivo de interrumpir la transferencia intergeneracional de la pobreza. Específicamente, la inversión en educación alivia la pobreza, aumenta la acumulación de capital humano y reduce la transmisión intergeneracional de la pobreza. Por lo tanto, sale a la luz la necesidad de implementar una política integral que busque garantizar el derecho a la educación, tendiente a eliminar el trabajo de menores, interrumpiendo así el círculo de la pobreza.

Banerjee A., Duflo E. (2011) nos brindan una perspectiva diferente a la hora de entender a las personas que viven en situación de pobreza, y la forma en la que toman decisiones. Muchas veces, estos individuos se ven presionados por el ahora, es decir, la necesidad inmediata, los lleva a considerar únicamente el presente, y no son capaces de evaluar las consecuencias a largo plazo. Además, los ganadores del premio Nobel de Economía 2019, plantean que las políticas públicas no deben buscar resolver grandes problemas macroeconómicos, sino cuestiones puntuales. La manera en que presentan las soluciones a las problemáticas que atraviesan la pobreza, parten desde un enfoque personalizado, teniendo en cuenta su manera de vivir, pensar y desde allí tomar decisiones. Siguiendo esta línea de pensamiento, la falta de dinero puede significar en muchos casos, que un jefe de hogar vea la

necesidad de que los menores trabajen. Por ende, la ayuda financiera del Estado resulta ser una intervención prudente.

El Estado suele utilizar las transferencias de dinero como una herramienta para disminuir, entre otras problemáticas, el trabajo infantil. Rosati F., de Hoop J. (2014) ponen en evidencia el efecto de las transferencias monetarias condicionadas y no condicionadas sobre el trabajo de menores de edad. Los autores sostienen que el incremento en la economía del hogar incide positivamente en la reducción del trabajo infantil. Este cambio puede asociarse al aumento en la escolaridad, debido a la mejora en la situación económica familiar. Cabe destacar, que Rosati F., de Hoop J. (2014) argumentan que la asistencia a la escuela y el trabajo infantil no son actividades mutuamente excluyentes. De todos modos, Pholphirul P., Xayavong T. (2018) declaran que la razón más común por la que jóvenes entre 5 y 17 años dejan la escuela es la necesidad de salir a trabajar. Dichos autores enseñan una relación positiva entre la deserción escolar y el trabajo de menores. Por lo tanto, es relevante aclarar que los programas de transferencias monetarias no tienen como objetivo principal evitar el trabajo infantil o adolescente, sino más bien proteger a poblaciones económicamente vulnerables para promover la acumulación de capital humano que imponen las condicionalidades. Por ende, una posible consecuencia de estas condicionalidades es una reducción en el trabajo infantil, por el incremento del ingreso del hogar, así como también la restricción de tiempo que establece ir a la escuela.

Por su parte, Spiga, F. (2019) propone un análisis similar a esta investigación. Su trabajo se enfoca en el impacto de la AUH en el trabajo adolescente específicamente. Es decir, se concentra en personas con edades entre 15 y 17 años. A pesar de que no abarca trabajo infantil, esta investigación es sumamente relevante para nuestro trabajo. Spiga, F. (2019) realiza un estudio empírico basado en tres variables: tasa de empleo, tasa de actividad y horas trabajadas por parte de los adolescentes. Su estimación obtiene que la AUH tiene un efecto estadísticamente significativo para adolescentes entre 15 y 17 años. La transferencia monetaria parecería reducir la tasa de empleo, la de actividad y el promedio de horas trabajadas de los adolescentes.

Sin embargo, su estudio encuentra efectos heterogéneos por género en una de las tres variables, particularmente en el promedio de horas trabajadas. Donde la desagregación por género de la variable, muestra que el impacto negativo de la AUH sobre el promedio de horas trabajadas estaría siendo liderado por las mujeres. Este impacto heterogéneo podría ser explicado por la diferencia entre lo que gana un adolescente varón y un adolescente mujer. Suponiendo un trabajo de 8 a 10 horas de una mujer, que podría competir con el ir a la escuela, las mujeres estarían dejando de trabajar porque ese ingreso lo ven sustituido por la AUH. En cambio, supongamos que los varones ganan lo mismo con 4 horas de trabajo, esto no estaría compitiendo con ir a la escuela.

Asignación Universal por Hijo

La Asignación Universal por Hijo o AUH, es un programa de transferencias condicionadas, ejecutado por la ANSES (Administración Nacional de la Seguridad Social). Este programa se implementó en noviembre de 2009 y está destinado a aquellos niños/as cuyos padres se encuentren desocupados, o bien trabajando en el sector informal o en servicio doméstico, sean monotributistas o cuenten con otros programas sociales. El objetivo de la AUH es incrementar el capital humano de los niños/as, independientemente de la situación laboral de sus padres.

El programa consiste en una transferencia de un monto de dinero, el cuál varía frente a casos particulares (como puede ser durante el embarazo, o con un/a hijo/a con discapacidad). Mensualmente se le otorga el 80% del monto total a la madre, tutor, curador o parientes hasta el tercer grado de consanguinidad del niño/a en cuestión, para asegurar el cumplimiento de las condicionalidades. Estas consisten en la asistencia a la escuela y en la vacunación, entre otros controles correspondientes de salud. Una vez al año se debe presentar la documentación necesaria, para demostrar el cumplimiento de las mismas y recibir el 20% restante del dinero correspondiente. La AUH busca impactar tanto en la pobreza como en la educación y en la salud, tras requerir el cumplimiento de las condiciones que presenta. Edo, M. y Marchionni, M. (2019), exponen que los efectos del programa en la educación varían según el género y la edad de los beneficiarios. En términos generales, hubo una

mejora significativa en la asistencia a clase en niñas entre 12 y 14 años y en niños entre 15 y 17 años. Además, aumentó la graduación de la escuela primaria y disminuyó la participación laboral en menores de 18 años. Consideramos que, como mencionan María Edo y Mariana Marchionni en el texto en cuestión, la AUH busca interrumpir la transferencia intergeneracional de la pobreza, y sus incentivos a la escolarización colaboran altamente con dicho objetivo.

Específicamente, para esta investigación es relevante la relación entre la transferencia monetaria que conlleva la AUH y sus condicionalidades sobre el trabajo de menores. Cabe destacar que la subida en el ingreso del hogar junto con el requerimiento de asistir a la escuela, producen un aumento en el efecto sustitución del trabajo infantil. Es decir, dejar la escuela para trabajar se vuelve relativamente más costoso. En otras palabras, el costo de oportunidad de trabajar es más alto cuando un menor pasa a ser beneficiario de la AUH.

En líneas generales, los resultados de esta política han sido positivos, si los evaluamos en un corto plazo; podemos destacar algunos aspectos en cuanto a su focalización, al aumento y mejora del consumo por parte de las familias beneficiarias, al aumento en las tasas de matriculación y asistencia por parte de los niños/as, al aumento en la utilización de los servicios de salud, a la reducción de la tasa de empleo infantil, entre otros. En cuanto a los resultados de largo plazo, se dificulta aún su medición de impacto. UNICEF (2017) comprueba que el impacto social es significativo, aunque deja de ser determinante en términos de eliminar la pobreza o reducir fuertemente la desigualdad.

Mecanismos

Desde un punto de vista teórico, según Jiménez M., Jiménez M. (2021), los programas de transferencias monetarias condicionadas pueden producir un impacto en el trabajo infantil a través de una serie de mecanismos subyacentes. En primer lugar, el mecanismo quizás más evidente es el efecto ingreso que tienen sobre el hogar. De esta manera, la transferencia monetaria permite a los hogares suavizar su consumo y satisfacer sus necesidades básicas. Sin embargo, si el ingreso del hogar proviene del trabajo remunerado de los niños y niñas; y no es superado por el monto

la transferencia monetaria que otorga el programa, la pérdida de los ingresos provenientes del trabajo infantil podría reducir la magnitud del efecto ingreso que produce la transferencia. Esto podría disminuir los incentivos para su obtención y su cobro. Además, Edo M., Marchionni M., Garganta S. (2015) establecen que en la medida en que la educación se considere un bien normal, el consumo del mismo podría aumentar cuando lo hace el ingreso familiar. Entre los mecanismos a través de los cuales la AUH puede mejorar los resultados educativos, se encuentra la mera existencia de la transferencia monetaria que relaja las limitaciones presupuestarias del hogar.

En segundo lugar, las condicionalidades de estos programas podrían producir una suerte de efecto sustitución en relación al trabajo infantil. En el caso de la AUH, en la medida en que aumentan los costos de oportunidad de la inasistencia escolar. Es decir, si el monto de la transferencia es lo suficientemente grande, entonces el trabajo infantil se verá sustituido por asistencia a la escuela, para poder recibir dicho monto. Por lo tanto, la condición de escolaridad de los niños y niñas supone un aumento de su escolarización, así como también, una reducción del trabajo remunerado principalmente en el caso de los varones, o de actividades domésticas intensas, en especial entre las niñas. Asimismo, Edo M., Marchionni M., Garganta S. (2015) se preguntan qué mecanismos dentro de la AUH son responsables del aumento de tasas de asistencia escolar. Los autores argumentan que es posible que el efecto venga dado por el propio beneficio mensual, o impulsado por la condicionalidad, o ambos mecanismos podrían estar operando al mismo tiempo. También resulta crucial entender los resultados que se generan entorno a la temática de trabajo infantil. En particular, se espera que un aumento en las tasas de asistencia contribuya a reducir el empleo entre las personas de 15 a 17 años.

Por último, los programas de transferencias monetarias pueden producir un efecto sobre el trabajo infantil a través de un tercer mecanismo: el distributivo. Cuando las madres son seleccionadas como receptoras de la transferencia, esto tiene un impacto sobre la asignación y el gasto de los recursos dentro del hogar. Esto podría estar impactando en las decisiones intrafamiliares respecto de la participación en el mercado de trabajo de los miembros del hogar, y todo lo que respecta a las tareas del hogar.

PARTE II:

Datos

En su trabajo, Jiménez M., Jiménez M. (2021) utilizan la EANNA 2016-2017, como fuente de información disponible, la cual resulta apropiada para evaluar el impacto de las políticas de seguridad de ingresos en la niñez sobre el trabajo infantil. Esta encuesta contiene datos sobre los hogares que son perceptores de la AUH y de otros programas sociales. Jiménez M., Jiménez M. (2021) establecen que estas políticas públicas pueden representar un incremento del ingreso para el hogar por lo que es posible que produzcan un impacto indirecto en el trabajo infantil en la medida en que el beneficio recibido sea significativo en términos de satisfacción de sus necesidades básicas.

Cabe destacar que, dentro del grupo de niños, niñas y adolescentes (NNyA), la AUH es el programa de mayor cobertura, por sobre el resto de los programas sociales. En su estudio, los autores resaltan que alrededor del 13% de los NNyA trabajan y un 5,6% de estos viven en un hogar alcanzado por la AUH. Por otro lado, un 2,7% de los NNyA que trabajan viven en un hogar no cubierto por ningún programa social. Es decir, 1 de cada 5 chicos que trabaja no recibe ningún tipo de transferencia monetaria, independientemente del tipo de actividad que realicen. Sin embargo, su estudio se centra en menores de 18 años, por lo cual incluyen a los adolescentes en su muestra.

La EANNA 2016-2017 urbana está compuesta por tres cuestionarios distintos:

- El cuestionario 1 describe los hogares. Muestra las características de la vivienda; ya sean materiales de construcción, acceso a servicios, cantidad de ambientes, etc.
- El cuestionario 2 describe los miembros del hogar mayores de 18 años. Muestra características como nivel de ingresos, nivel educativo, categoría ocupacional y si reciben transferencias monetarias como AUH entre otros.
- El cuestionario 3 describe miembros del hogar menores entre 5 a 17. Muestra si asisten a la escuela, que tipo de actividades realizan y si realizan alguna actividad productiva entre otras.

La Fig. 1 muestra el porcentaje de trabajo infantil según la actividad que realizan en zonas rurales y urbanas. Las actividades productivas se clasifican como: actividad para el mercado, para el autoconsumo y para el hogar (doméstica). Notemos que un 86,4% de los NNyA pertenecen a zonas urbanas, algo completamente lógico teniendo en cuenta la densidad poblacional de estas.

Tal como realiza Chocobar (2021), se utilizan únicamente los cuestionarios realizados en las zonas urbanas. Esto se debe a que es muy común que en zonas rurales se observen a NNyA realizando actividades productivas, particularmente para el autoconsumo por costumbre familiar. Esto se puede ver en la Fig. 1 donde los NNyA de zonas rurales que realizan actividades para el autoconsumo representan un 9,8%; a diferencia de zonas urbanas donde solo el 1,9% realizan actividades productivas de este tipo. Por lo tanto, dado que la base no especifica si los NNyA trabajan por costumbre familiar, no es posible aislar el efecto. Unir ambas bases podría resultar entonces en una reducción predictiva ya que no se estaría considerando la relevancia de este efecto.

Fig. 1: Participación en actividades productivas según ámbito urbano o rural. Niños y niñas de 5 a 15 años.

	Total niños y niñas		Urbano		Rural		Urbano sobre total nacional	Rural sobre total nacional
	Absolutos	%	Absolutos	%	Absolutos	%	%	%
Total	7.648.413	100,0	6.604.464	100,0	1.043.949	100,0	86,4	13,6
Realizan actividad para el mercado (!)	291.335	3,8	218.527	3,3	72.808	7,0	75,0	25,0
Realizan actividad para el autoconsumo (!)	227.497	3,0	125.315	1,9	102.182	9,8	55,1	44,9
Realizan actividad doméstica intensiva (!)	367.569	4,8	283.664	4,3	83.905	8,0	77,2	22,8
Realizan al menos una actividad productiva	763.544	10,0	556.909	8,4	206.635	19,8	72,9	27,1

(!) Cada categoría considera al total de la población que realizó la actividad. De este modo, se incluyen a los que realizaron solo esa actividad y a aquellos que la hicieron junto con otra/s. Por lo tanto, los niveles encontrados no coinciden estrictamente con aquellos referidos a la incidencia de al menos una actividad en el universo de niños y niñas.

Fuente: EANNA Urbana (2016/2017) y EANNA Rural (2017), INDEC y MTEySS.

Por otro lado, el objetivo de esta segunda parte es poder predecir los hogares donde hay trabajo infantil para la Encuesta Permanente de Hogares (EPH, 2021 tercer trimestre). Esta se realiza todos los años, sin embargo, no logra captar el trabajo infantil en su totalidad. Las únicas preguntas realizadas que podrían representar el trabajo infantil son si los menores de 10 años aportaron dinero trabajando o pidiendo dinero. Estas no estarían captando el rango etario completo de los NNyA (menores de 17 años), y tampoco estarían captando actividades no remuneradas que son consideradas como trabajo infantil, por ejemplo; actividades domésticas intensivas y actividades para el autoconsumo.

Resulta útil poder predecir los hogares donde hay trabajo infantil en todas sus dimensiones para la EPH una de las encuestas más utilizadas en nuestro país para fines prácticos y académicos, realizada varias veces por año. Esta encuesta está dividida en dos cuestionarios:

Base Hogar: Muestra las características de la vivienda y habitacionales del hogar, y las estrategias de los residentes.

Base Individuos: Muestra las características de los miembros del hogar y la ocupación principal de los trabajadores independientes.

Elección del Modelo

En *machine learning* se busca predecir fuera de la muestra. Es decir, “entrenar” un algoritmo utilizando una base de datos para poder realizar una predicción fuera de esta. Justamente, por esa razón es que este tipo de modelos son muy adecuados para resolver el problema de predecir el trabajo infantil en la Argentina. Hoy por hoy, únicamente la EANNA 2016-2017 (muestra de entrenamiento) incluye la variable si un NNyA trabaja o no. Esta encuesta a diferencia de la EPH no se realiza todos los años. Entonces, lo que se busca es utilizar un modelo que a partir de la EANNA 2016-2017, realice una predicción para la EPH. Con el objetivo de poder definir los hogares donde habrá al menos un NNyA que trabaje. Esto permitirá también identificar las características de dichos hogares para extraer conclusiones.

La elección de este modelo no es trivial; justamente Chocobar (2021) dedica su investigación a encontrar el mejor modelo para predecir el trabajo infantil utilizando la EANNA 2016-207 como su muestra de entrenamiento. Es decir, utiliza la misma muestra que este trabajo para entrenar su modelo. Para encontrar el mejor modelo compara la performance de modelos de *machine learning* como: *Ridge* y *Lasso*, y tres tipos de árboles; *Random Forest*, *Gradient Boosted model*, *Extreme Gradient Boosting* (XGBoost). A su vez, los compara con *Logit*, un modelo clásico de la econometría tradicional. Para todos los modelos las variables estadísticamente significativas son edad y sexo de NNyA y del jefe del hogar, región a la que pertenece, deseos para la mayoría de edad, cantidad de menores de edad en el hogar, horas destinadas e ingreso percibido por ocupación principal y situación conyugal. Lo interesante que uno de los dos modelos que mejor performance en las distintas métricas, agregó la variable AUH, es decir, si percibe o no este programa. Chocobar (2021) encuentra mediante la comparación de distintas métricas de performance; que el modelo de XGBoost es el que mejor predice trabajo infantil utilizando la misma encuesta para la muestra de entrenamiento que este trabajo, por esa razón, sería prudente utilizar o al menos probar con este modelo para nuestra predicción y verificar su performance.

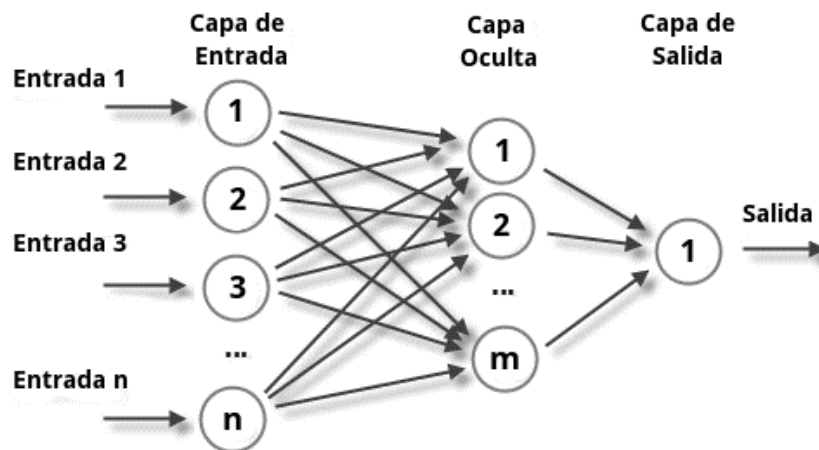
Extreme Gradient Boosting o XGBoost es un modelo predictivo que pertenece a lo que se conoce como modelos de árboles de decisión. Estos logran dividir las observaciones en función de sus características, permitiendo que las distintas variables interactúen entre sí y estas relaciones se asemejan visualmente a las ramificaciones de un árbol. Son fáciles de interpretar y comunicar. Identificar las variables más relevantes resulta sencillo ya que figuran arriba del árbol. Estos modelos pueden utilizar predictores numéricos como categóricos y se utilizan mucho para problemas de clasificación. La forma de operar de estos algoritmos básicamente es; teniendo Y_i como variable dependiente y distintos predictores como X_1 y X_2 , el algoritmo realiza una partición recursiva binaria de los espacios de los predictores, donde en cada espacio, se particiona por cierto punto elegido por el algoritmo. El árbol puede llegar a tener la misma cantidad de particiones que observaciones en la muestra. Sin embargo, en ese caso limite se sobre ajusta los datos, es decir, el modelo predecirá muy bien dentro de la muestra, pero muy mal fuera de la muestra, esta problemática se conoce como *overfitting*.

Por esta razón, es importante elegir el nivel de complejidad óptimo del árbol, teniendo en cuenta que se busca predecir fuera de la muestra el trabajo infantil. La EANNA es la muestra de entrenamiento y predecimos fuera de ella, utilizando la EPH. El XGBoost utiliza aproximaciones más precisas para justamente evitar el problema de sobreajuste y encontrar el mejor árbol. Este modelo tiene la ventaja de poder incluir valores faltantes, algo muy usual en las encuestas a utilizar; por lo que permite incluir más variables que otros modelos.

Por su lado, Libaque-Saenz, C.F; Lazo, J et. al. (2020). Los autores buscan predecir el trabajo infantil en Perú utilizando redes neuronales, mostrando que es más preciso que un modelo logit tradicional. Encuentran que indicadores geográficos, niveles de ingreso, género, composición de la familia y nivel de educación son variables significativas para predecir el trabajo infantil. De manera similar, en este trabajo se utiliza también un modelo de redes neuronales. Se comparan las métricas de performance en la muestra de testeo con el modelo XGBoost con el fin de seleccionar el mejor modelo para realizar la predicción en la muestra de la EPH.

Una red neuronal es un modelo simplificado que simula la forma en que el cerebro humano procesa la información. Normalmente, existen tres partes en una red neuronal: una capa de entrada, esta estaría formada por los campos de entrada, una o varias capas ocultas, donde se realizan las ponderaciones para cada variable; y una capa de salida (campo de destino). Los datos de entrada se presentan en la primera capa, y los valores se propagan desde cada neurona hasta cada neurona de la capa siguiente, hasta llegar a la capa de salida, enviando el resultado final (Ver Fig. 2). Las redes neuronales aprenden examinando los registros individuales, generando una predicción para cada registro y realizando ajustes a las ponderaciones o pesos cuando realiza una predicción incorrecta. Este proceso se repite numerosas veces y la red sigue mejorando sus predicciones hasta haber alcanzado uno o varios criterios de performance.

Fig 2: Esquema red neuronal.



En un principio, todas las ponderaciones son aleatorias y las respuestas que resultan de la red son muy posiblemente erradas. Continuamente se presentan a la red ejemplos para los que se conoce el resultado, y las respuestas que proporciona se comparan con los resultados conocidos. Así se van ajustando las ponderaciones gradualmente. De esta manera, la red se va haciendo cada vez más precisa; al igual que el XGBoost hay que tener cuidado con el problema de *overfitting*. En el caso de las redes neuronales, una arquitectura muy compleja, es decir, una gran cantidad de capas y de neuronas puede resultar en que el modelo ajuste muy bien a la muestra de entrenamiento. Por esta razón, durante el entrenamiento se vigila la función de pérdida de validación con respecto a la de entrenamiento para evitar este tipo de problemas.

Metodología

Como se mencionó anteriormente, se utiliza la base Urbana de la EANNA 2016-2017 como muestra de entrenamiento. La variable dependiente a utilizar es la variable binaria "Trabaja" donde 0 si el NNyA no trabajo y 1 en el caso contrario. Para realizar la predicción es necesario seleccionar las variables similares o iguales entre ambas encuestas (Ver Anexo). Esta es una de las grandes limitaciones de este trabajo. El hecho de que se quiera realizar una predicción en una encuesta distinta a la cual se entrenó el modelo, presenta una serie de complicaciones. Se deben seleccionar las variables que capten la misma información en ambas bases, reduciendo

significativamente la cantidad de variables con la que se entrenará el modelo. Los cuestionarios no realizan exactamente las mismas preguntas ya que sus objetivos son distintos. Sin embargo, ambas bases comparten una misma metodología a la hora de su elaboración. Tanto la EANNA como la EPH realizan un cuestionario para los hogares y los individuos, identificando los últimos con un código que los asocia a cada hogar. Por lo tanto, en el procesamiento de los datos, se agrupan los datos por hogar, una vez seleccionado las variables similares. Se utiliza como estrategia la moda (el valor que más aparece) para la agrupación por hogar de las variables de las bases de individuos ya que la mayoría de estas son variables categóricas y también se utiliza la misma estrategia para remplazar los valores faltantes.

Cabe mencionar que la muestra de entrenamiento esta desbalanceada. Es decir, la proporción de NNyA que trabaja es muy chica en nuestra muestra de entrenamiento. Solo en el 6,7% de los hogares hay al menos un NNyA que trabaja. Esto afecta a los algoritmos en su proceso de generalización de la información y perjudicando a las clases minoritarias en un problema de clasificación como este. Es decir, si al modelo le otorgamos 990 datos de una categoría (“No trabaja”) y sólo 10 de la otra categoría (“Trabaja”), no podemos pretender que logre diferenciar una clase de otra. Lo más probable es que el modelo se limite a clasificar con (“No trabaja”), esto se debe al hecho que así tuvo un acierto del 99% en su fase de entrenamiento. Una forma de resolver dicho problema es utilizar una técnica conocida como SMOTE (*Synthetic Minority Oversampling Technique*), esta realiza un *oversample* de la clase minoritaria, en este caso seria los hogares con trabajo infantil. Es decir, duplica los casos donde hay trabajo infantil, creando de esta manera ejemplos sintéticos de la clase minoritaria y logrando balancear la distribución de los datos sin agregar información adicional al modelo. En este trabajo particularmente, se utiliza la versión SMOTENC que puede procesar variables categóricas.

Se divide entonces la base de la EANNA 2016/17 (agrupada por hogares con las variables seleccionadas) en una muestra de entrenamiento y otra de testeo. Esto sirve para saber si el modelo es capaz de predecir correctamente. Es decir, el modelo se entrena con la muestra de entrenamiento, donde observa los datos de la variable “Trabaja” y predice en la muestra de testeo, pero sin observar esos datos. Para realizar esto, se utiliza una técnica llamada *K-fold-Cross-Validation* para evaluar la

capacidad predictiva del modelo. Esta consiste básicamente en dividir la base de la EANNA en k subconjuntos de la muestra, uno de los subconjuntos se utiliza como datos de testeo y el resto $K - 1$ como datos de entrenamiento. Esto se repite k veces y se computa el *accuracy* o precisión del modelo.

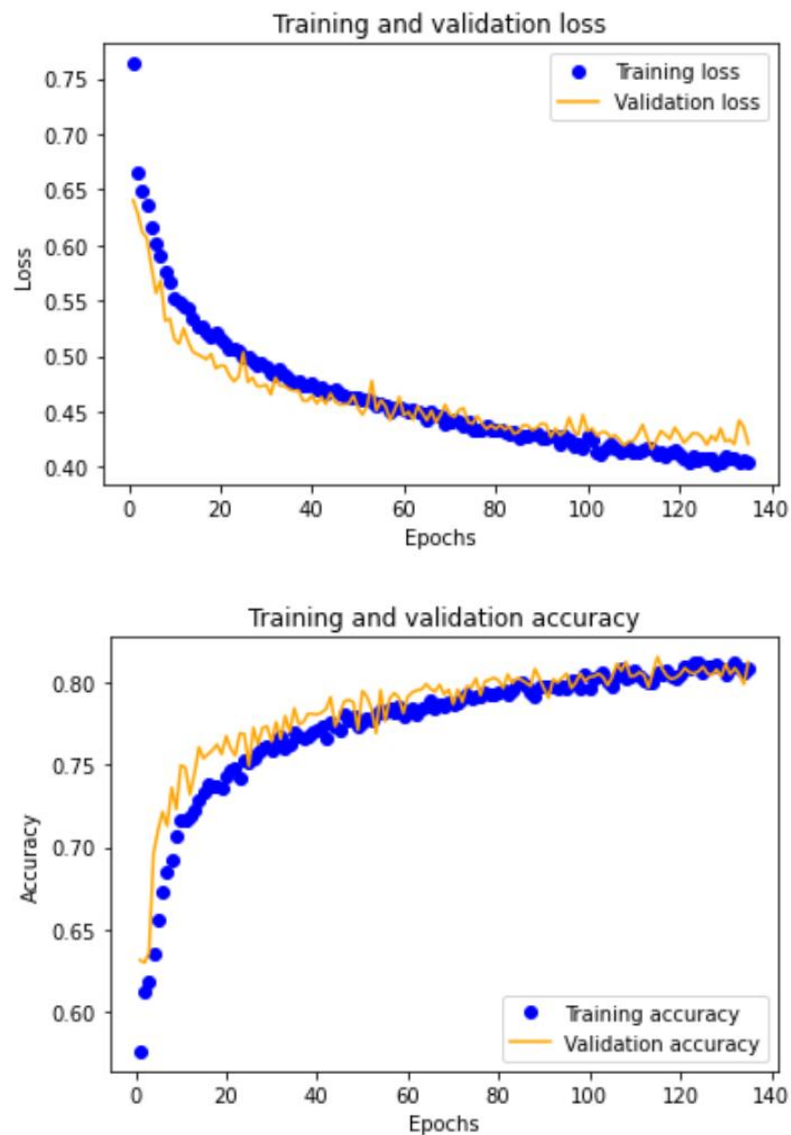
Por último, resulta útil ajustar una serie de parámetros en cada modelo para mejorar su capacidad predictiva para este caso en particular. Pueden ser consideradas como ‘perillas’ del panel de control del algoritmo, las cuales nos permiten “tunear” para mejorar su performance. Estos son conocidos como hiperparámetros, y tienen la característica de ser parámetros que no surgen de los datos. Por ejemplo, la profundidad del árbol en XGBoost. El desafío está en encontrar la mejor configuración de hiperparámetros para cada modelo.

En primer lugar, se buscan los hiperparámetros óptimos para el modelo de XGBoost que minimicen la función de pérdida conocida como *Logloss*, función que penaliza las clasificaciones erróneas. Nuevamente, se utiliza *cross validation* para encontrarlos, partiendo la muestra de entrenamiento en dos; entrenamiento y validación. Esta última se utilizará para probar las distintas configuraciones de hiperparámetros. XGBoost cuenta con numerosos hiperparámetros para configurar. En este trabajo utilizamos algunos de los más importantes, entre ellos, la profundidad del árbol. Una vez hecho esto, se entrena el modelo con los parámetros óptimos y se testea su performance. Previo a la optimización, el modelo había obtenido un score de 0.869. Luego, con los hiperparámetros encontrados logro un score de 0.903; clasificando a los NNyA que no trabajan correctamente con una precisión de 0.93 y con 0.86 a los NNyA que no trabajan.

Similar es metodología para el modelo de redes neuronales. Sin embargo, en este caso es muy importante a la hora de entrenar el modelo seguir de cerca las curvas de pérdida tanto de entrenamiento como de validación. De esta manera, se podrá visualizar muy fácilmente si el algoritmo esta ajustándose demasiado a los datos de entrenamiento, lo que resultaría en un problema de *overfitting* como mencionamos anteriormente. En la Fig 3 a partir del *epoch* o iteración 120 del entrenamiento, la curva de pérdida de validación comienza a ser mayor que la de entrenamiento. Es ahí justamente cuando comienza a aparecer el problema de

overfitting. Nótese también que la curva de *accuracy* o precisión del modelo tiende a estancarse en 0.80. Al evaluar el modelo de redes neuronales, obtuvo un score de 0.81.

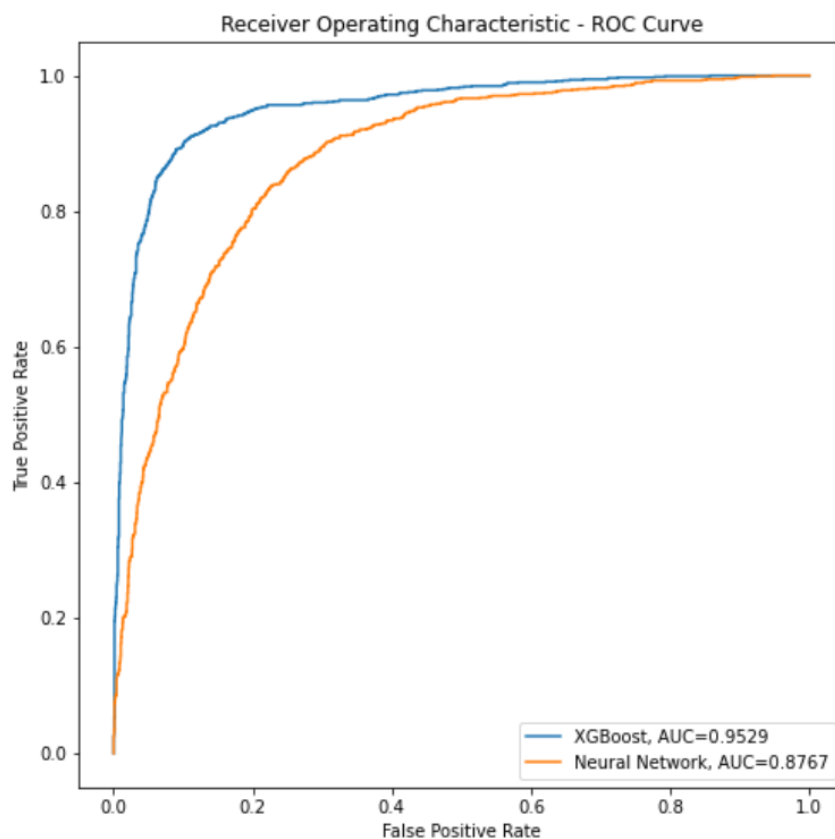
Fig 3: Curvas de Perdida y Precisión de Validación y Entrenamiento



Una forma útil y visual para comparar la performance de varios modelos es utilizando la Curva ROC. Esta muestra la relación entre la tasa de verdaderos positivos (TPR) y la tasa de falsos positivos (FPR) del modelo. TPR describe la tasa a la que el clasificador predice como “positivo” a las observaciones que son

“positivas”. FPR describe la tasa a la que el clasificador predice como “positivo” a las observaciones que en realidad son “negativas”. En este caso, TPR sería la tasa en la cual el modelo predice que en el hogar hay trabajo infantil cuando realmente lo hay; y FPR cuando en realidad no lo hay. En general, cuanto más “arriba y a la izquierda” del diagrama se encuentre la curva ROC, mejor será el modelo clasificador. También, existe una medida relacionada a la curva ROC, llamada AUC. La cual representa el área bajo la curva ROC. Es una medida que sirve para comparar rápidamente clasificadores. Los clasificadores perfectos, es decir, con una AUC 1 tienen una TPR de 1 y una FPR de 0. En ese caso, la curva estará pegada a la esquina superior izquierda del cuadrante. Notemos en la Fig. 4 como el modelo XGBoost clasifico de manera muy acertada obteniendo una puntuación AUC de 0.95. Por su lado, el modelo de redes neuronales logro una AUC de 0.87. Se puede afirmar que ambos modelos son buenos clasificadores, pero XGBoost parecería ser el ganador entre los dos.

Fig. 4: Curva ROC



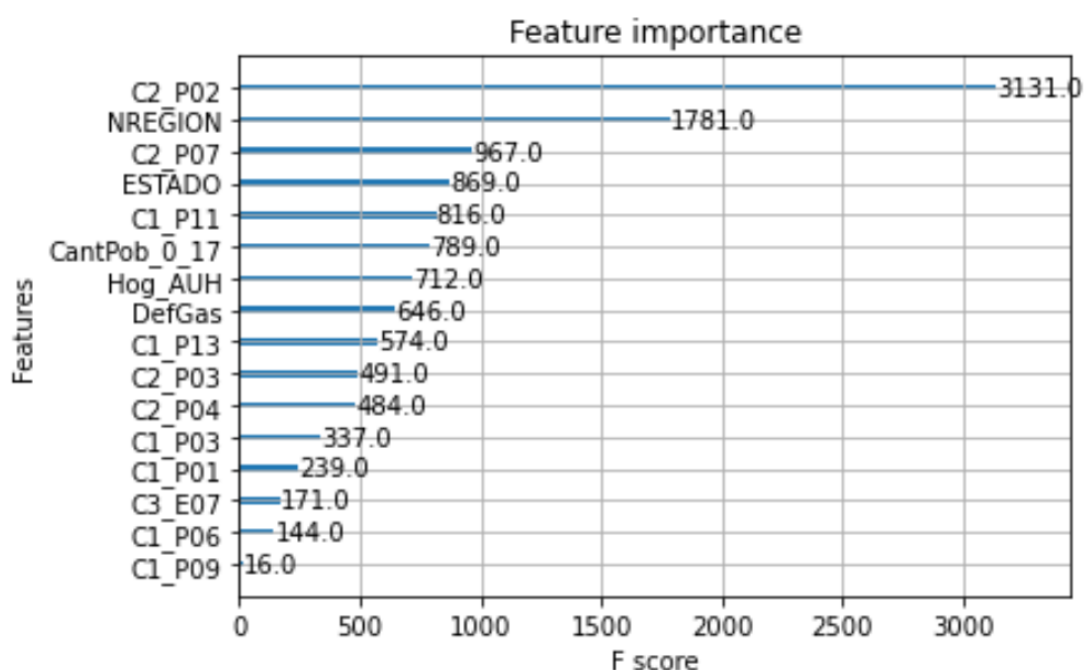
Resultados

Una vez entrenado los modelos, se realiza la predicción para la Encuesta Permanente de Hogares 2021 (EPH) con el fin de poder identificar los hogares donde habita al menos un NNyA que trabaja. Por un lado, el modelo de redes neuronales predijo un 70% de los hogares hay al menos un NNyA que trabaja, mientras que el modelo XGBoost predijo el trabajo infantil en 21% de los hogares. Esta diferencia podría atribuirse a la diferencia de performance de ambos modelos, lo que llevaría a inclinarse por el XGBoost a la hora de extraer conclusiones sobre los resultados.

Una ventaja del modelo XGBoost es que tiene incorporado la posibilidad de rápidamente obtener la relevancia de cada variable explicativa a la hora de predecir si hay trabajo infantil en el hogar. El criterio de importancia está basado en el F score, donde las variables que mayor puntaje tienen, son las que fueron utilizadas para hacer decisiones claves dentro de los árboles del XGBoost. A partir de la Fig. 5, notemos que parecieran haber dos variables que sobresalen por sobre las demás. La más relevante es la cantidad de años cumplidos (C2_P02). Es importante aclarar que esta variable representa a todos las personas que habitan el hogar. Resulta lógico que esta sea una variable de corte principal ya que el algoritmo distingue a la hora de predecir el trabajo infantil que la edad es lo primero que tiene que separar para lograrlo. Es decir, modelo aprende que para predecir correctamente el trabajo infantil debe separar entre adulto e infante en primera instancia.

Luego, la segunda más importante es la región (NREGION) donde reside el hogar. Parecería ser que la región tiene una gran incidencia en el trabajo infantil. Una vez explorado los datos de la EPH junto con la predicción realizada por ambos modelos se identifica que la región con más trabajo infantil es la región pampeana con un 33% de los hogares donde hay al menos un NNyA que trabaja, seguido por la región del NOA. Resulta interesante que los resultados para las regiones con mas trabajo infantil coinciden en ambos modelos.

Fig. 5: Variables relevantes XGBoost.



En tercer lugar, la situación conyugal del jefe/a de hogar (C2_P07) tendría cierta relevancia en el trabajo infantil del hogar. Esto podría atribuirse al hecho de que el nivel ingreso del hogar podría verse reducido al haber madre/padre soltero como jefe del hogar. Como se menciona en la primera parte, esto podría generar la necesidad económica familiar causando que los NNyA se vean obligados a contribuir la economía familiar.

Finalmente, habiendo desarrollado en la primera parte de esta tesina los mecanismos por los cuales las transferencias monetarias condiciones, particularmente la AUH podría reducir el trabajo infantil. No es casualidad que ambos modelos hayan coincidido en que un 70% de los hogares identificados con trabajo infantil no reciben AUH. Este es el resultado principal de este análisis. Es una prueba empírica de los mecanismos teóricos expuestos anteriormente y también información clave para identificar las características específicas de los hogares que deberían recibir el programa de la AUH.

Conclusiones

En la primera parte de esta investigación se expuso la relevancia del trabajo infantil en tiempos actuales. Se analizaron, a la luz de otras investigaciones, los mecanismos por los cuales la AUH podría reducir el trabajo infantil. Dado la falta de datos sobre esta problemática en nuestro país, el principal aporte de esta tesina es brindar información relevante sobre el trabajo infantil en zonas urbanas. Esto se logra utilizando modelos clasificadores binarios como el XGBoost y un modelo de redes neuronales. Se utiliza la única encuesta que logra captar el trabajo infantil conocida como la EANNA 206/17 para entrenar los modelos. Este análisis predictivo sirve de precedente también para probar la eficacia del clasificador XGBoost a la hora de realizar predicciones binarias.

Luego de optimizar ambos modelos y comparar su performance se realiza la predicción para la EPH 2021, una de las encuestas mas importantes de nuestro país y la principal encuesta de hogares la cual se realiza frecuentemente. De esta manera se identifica empíricamente los hogares donde hoy en día tienen trabajo infantil. Además, se logra identificar las regiones donde hay mayor probabilidad de haber trabajo infantil. Finalmente, se demuestra que el 70% de los hogares que habita al menos un NNyA que trabaja no están recibiendo están recibiendo la AUH; por lo que se identifica con precisión el tipo de hogares los cuales el Estado podría otorgar este tipo de programas con el fin de reducir el trabajo infantil.

Existen ciertas limitaciones a la hora de realizar la predicción expuesta. El hecho de utilizar bases de datos distintas limitó la selección de variables ya que ambas deben coincidir. Tanto la muestra de entrenamiento como la de la predicción final deben tener las mismas variables para que el modelo se entrene con las mismas y su predicción sea válida. Por ende, el número de variables seleccionadas fue ampliamente menor a la cantidad de variables disponibles en ambas bases. Sin embargo, cabe mencionar que muchas veces no todas las variables son relevantes al problema, es decir, no colaboran significativamente al objetivo de predecir el trabajo infantil. Quizás en un futuro cercano se cuenten con encuestas actualizadas que capten este tipo de problemáticas, para poder realizar estudios más robustos.

Referencia bibliográfica

Banerjee A., Duflo E. (2011), Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty.

Chocobar, E. (2021), Econometria vs Machine Learning: ¿cual tecnica predice mejor el trabajo infantil? Universidad de San Andres

de Hoop et. al. (2014), “Cash Transfers and Child Labor”. The World Bank Research Observer , August 2014, Vol. 29, No. 2 (August 2014), pp. 202-234.

Edo M., Marchionni M. (2019), “The impact of a conditional cash transfer program beyond school attendance in Argentina”, Journal of Development Effectiveness.

Edo M., Marchionni M., Garganta S. (2015), “Conditional cash transfer programs and enforcement of compulsory education laws. The case of Asignación Universal por Hijo in Argentina”, Universidad Nacional de la Plata. CEDLAS.

Jiménez M., Jiménez M. (2021), “Trabajo infantil y protección social en la niñez en Argentina” Organización Internacional del Trabajo.

Libaque-Saenz, C.F; Lazo, J et. al. (2020), “Predicting child labor in Peru: A comparison of logistic regression and neural networks techniques”. Universidad Católica del Peru.

Pholphirul P., Xayavong T. (2018), “Child Labour and school dropout in least-developed countris: evidence from Lao PDR”. International Journal of Education Economics and Development.

Spiga F. (2019), “Los adolescentes y el mercado laboral: ¿influye la AUH en la decisión de trabajar?” Universidad de San Andrés.

UNICEF (2017), “Análisis y Propuestas de Mejoras para ampliar la AUH”. UNICEF y la Facultad de Ciencias Económicas de la Universidad de Buenos Aires.

Xayavong et. al. (2018), “Child labour and school dropout in least-developed countries: empirical evidence from Lao PDR”. Int. J. Education Economics and Development, Vol. 9, No. 1, 2018

Yuningsih et. al. (2018), “Social Protection for Child Labors”. International Journal of Engineering and Technology, 404-407.

Anexo

Tabla 1: Variables del hogar

Descripción de la variable	Nombre EANNA	Nombre EPH
Cantidad de ambientes	C1_P04	IV2
Régimen de tenencia	C1_P13	II7
¿Tiene agua? (tipo:)	C1_P06	IV6
¿Tiene baño?	C1_P09	IV8
¿Cómo es el desagüe del baño?	C1_P11	IV11
¿El hogar recibe transferencias de dinero (AUH)?	Hog_AUH	V5
Region	NREGION	REGION
Tipo de vivienda	C1_P01	IV1
Material de los pisos	C1_P03	IV3
¿Tiene gas de red?	DefGas	II8
Cantidad habitantes menores de 10 años	CantPob_0_17	IX_Men10
Cantidad miembros hogar	TamHog	IX_Tot

Tabla 2: Variables de los individuos

Descripción de la variable	Nombre EANNA	Nombre EPH
Relación	C2_P04	CH03
Años cumplidos	C2_P02	CH06
Sexo	C2_P03	CH04
Situación conyugal	C2_P07	CH07
Condición de Actividad adultos	ESTADO	ESTADO
Categoría ocupacional	CATOCUP	CAT_OCUP
Escuela pública o privada	C3_E07	CH11

Link al Repositorio Github con el código utilizado:

<https://github.com/martinabergc/EconomicsThesis/blob/main/Child%20Labour%20Prediction.ipynb>