

CENTRO DE INVESTIGACIÓN Y DOCENCIAECONÓMICAS, A.C.



**LOS FACTORES DETERMINANTES PARA LA ADOPCIÓN DE LAS
MICROFINANZAS EN MARRUECOS; ANÁLISIS DE HETEROGENEIDAD CON
APRENDIZAJE AUTOMÁTICO**

TESINA

QUE PARA OBTENER EL TÍTULO DE
LICENCIADA EN ECONOMÍA

PRESENTA:

YESOL LEE

DIRECTOR DE LA TESINA:
DR. RUBÉN IRVIN ROJAS VALDEZ

CIUDAD DE MÉXICO

2023

Agradecimientos

Quiero agradecer la elaboración de esta tesina a las siguientes personas:

Primero, a Dios quien da la sabiduría, conocimiento y discernimiento.

Mi familia, a mi padre y a madre quienes me han criado con amor y sacrificio, y a mi esposo Mehni quien me ha apoyado en todo este proceso de aprendizaje.

A los profesores que me acompañaron en el camino de mi carrera, al Dr. Gustavo Del Ángel, Dr. Daniel Ventosa, Maite, Eric, Alejandra, Raciél y Melissa.

A mi director de tesina, Dr. Rubén Rojas por pacientemente caminar conmigo en este largo camino de elaboración de mi tesina.

A mis amigos, con quienes pasé momentos desafiantes, tristes, divertidos y alegres durante la carrera y los cuales me apoyaron durante la misma como Frida, Emilio, Efrén y Luis.

A todo el equipo del CIDE, desde la administración, equipo técnico, comedor, limpieza y la biblioteca puesto que, sin ellos, no hubiera podido cada día continuar mis estudios.

Estoy consciente de que este trabajo ha sido posible por apoyo de todas y todos que realmente no hay nadie a quien no agradecer. Termino este capítulo con la enorme deuda de replicar esta misma bondad en la sociedad.

Resumen

Los programas de microfinanzas presentan características distintivas en comparación con los programas sociales de apoyo, ya que requieren la participación voluntaria de individuos con el objetivo de generar beneficios económicos a través de emprendimientos. El análisis de estudios sobre microfinanzas ha revelado un impacto mixto, aunque con posibles efectos positivos a largo plazo en la vida de los participantes. No obstante, hasta el momento, no se ha realizado un estudio detenido sobre aquellos asignados al grupo de intervención que optan por no tomar el préstamo, un porcentaje que resulta ser significativo.

Con el propósito de diseñar programas de microfinanzas más efectivos en el futuro, es valioso estudiar a aquellos individuos que deciden no aprovechar el apoyo financiero. Esto permitiría identificar las estrategias que podrían aumentar la participación en el programa. En este sentido, este estudio lleva a cabo un análisis de la heterogeneidad entre los grupos que toman y no toman microfinanzas dentro del grupo de tratamiento, utilizando un experimento realizado en Marruecos en 2015 y aplicando los métodos de los tres modelos de aprendizaje automático de heterogeneidad desarrollados por Chernozhukov.

Los resultados indican que existe una diferencia significativa entre los dos grupos mencionados anteriormente. El 20 por ciento de los individuos más afectados en el experimento mostraron, en promedio, un 35 por ciento más de probabilidad de optar por las microfinanzas en comparación con el 20 por ciento de los individuos menos afectados. Además, se identificaron características específicas de los individuos que favorecen una mayor probabilidad de adoptar las microfinanzas, como la situación de las actividades económicas y la experiencia previa en la toma de préstamos. Estos hallazgos proporcionan información valiosa para mejorar el diseño del programa, fomentando una mayor participación en la adopción de microfinanzas.

Palabras Clave: Microfinanzas, Adopción, Heterogeneidad, Efecto de Tratamiento, Aprendizaje Automático, Marruecos,

Índice

Capítulo 1	Introducción.....	1
Capítulo 2	Revisión de Literatura.....	5
Capítulo 3	Contexto y evaluación.....	10
Capítulo 4	Metodología.....	13
4.1	Introducción de la notación y el significado de las variables.....	13
4.2	Mejor Predictor Lineal (BLP en inglés).....	14
4.3	Efectos Promedio de Tratamiento por Grupos Ordenados (GATES).....	15
4.4	Análisis de Clasificación (CLAN).....	15
4.5	Modelos de aprendizaje.....	16
4.6	Algoritmo.....	16
Capítulo 5	Base de datos.....	20
5.1	Resumen estadístico de la base de datos.....	21
Capítulo 6	Resultados.....	24
6.1	Algoritmo de mejor desempeño.....	24
6.2	Resultados del Mejor Predictor Lineal (BLP).....	24
6.3	Resultados de los Efectos Promedio de Tratamiento por Grupos Ordenados (GATES).....	25
6.4	Resultados del Análisis de Clasificación (CLAN).....	26
Capítulo 7	Conclusión.....	34
Capítulo 8	Bibliografía.....	36

Índice de tablas y figuras

Tabla 1. Línea de base estadística.....	22
Tabla 2. Desempeño de los tres modelos de aprendizaje automático.....	24
Tabla 3. Resultados de BLP.....	24
Figura 1. Microfinanzas – GATES.....	26
Figura 2. Microfinanzas – CLAN con la variable “cabeza masculina de hogar”.....	27
Figura 3. Microfinanzas – CLAN con la variable “actividades por mujer”.....	28
Figura 4. Microfinanzas – CLAN con la variable “número de adultos”.....	28
Figura 5. Microfinanzas – CLAN con la variable “número de niños”.....	29
Figura 6. Microfinanzas – CLAN con la variable “gastos totales”.....	30
Figura 7. Microfinanzas – CLAN con la variable “ventas en ganadería”.....	31
Figura 8. Microfinanzas – CLAN con la variable “ventas en agricultura”.....	32
Figura 9. Microfinanzas – CLAN con la variable “monto de préstamo de las sucursales”..	33

1. Introducción

La inclusión financiera se define como la proporción de individuos y empresas que utilizan servicios financieros, lo cual varía ampliamente en todo el mundo (Banco Mundial, 2013). En otras palabras, la inclusión financiera busca ampliar la cobertura de los servicios financieros a los consumidores. Desde 2010, más de 55 países han asumido compromisos con la inclusión financiera, y más de 60 han lanzado o están desarrollando una estrategia nacional al respecto (Banco Mundial, 2022). Un país ejemplar que ha logrado los avances más significativos hacia la inclusión financiera es África Subsahariana. Ellos permitieron el florecimiento de los servicios financieros móviles, por lo que la posesión de cuentas de dinero móvil aumentó del 12 a 21 por ciento. En India, se implementaron las políticas de la identificación digital universal con las cuentas de JDY, lo cual posibilitó cubrir a más de 1.2 mil millones de residentes. Estos casos sirven como el punto de comparación para el contexto mexicano. Por ejemplo, un estudio del Instituto Nacional de Estadística y Geografía reveló que “56.7 millones de personas de 18 a 70 años en México (67.8 por ciento) tenían algún tipo de producto financiero formal, como cuenta de ahorro, crédito, seguro o Afore mientras que para 2018, la cifra era de 54.0 millones de personas” (INEGI, 2021).¹ Estos datos resaltan la existencia de un mercado de crédito con un potencial global excepcional y un crecimiento marginal. En este contexto, resulta imperativo llevar a cabo estudios más amplios sobre la estabilidad financiera desde una perspectiva de desarrollo económico.

En este sentido, es importante destacar el modelo de programas de microfinanzas, ya que estos van más allá del simple acceso financiero para las personas, alentando un mecanismo de sostenibilidad que promueve el emprendimiento entre los individuos. Este modelo transforma el paradigma convencional de la economía del desarrollo, que se centra en el acto de "dar". Además, varios estudios sobre microfinanzas hacen referencia a un impacto positivo en el crecimiento económico de los participantes, pero muestra evidencia mixta (Augsburg et al., 2015; Banerjee et al., 2015). Las evaluaciones formales más rigurosas encuentran que las microfinanzas ayudan a algunos y a otros no. Una razón por la que se explica este resultado mixto es porque hay efectos heterogéneos, es decir, no todos obtienen

¹ AFORE es la abreviatura de la Administradora de Fondos para el Retiro. Es una institución que se encarga de llevar la administración de fondos para el retiro de los trabajadores (Gobierno de México, 2020).

el mismo beneficio. Por ende, este trabajo tiene como objetivo tratar de estudiar los efectos de tratamiento heterogéneos que generan un impacto mixto en un programa de las microfinanzas.

Por ende, es importante notar que, en algunos casos, las intervenciones acercan los programas de microfinanzas a los potenciales clientes, pero estos deben decidir si participan o no. Aunque es valioso examinar el impacto de la asignación a la intervención, también es importante destacar la relevancia de la adopción en los programas de microfinanzas. La adopción se refiere a las personas asignadas al programa que eligen emprenderlo. Por ejemplo, el estudio de Crepón sobre el programa de microfinanzas en Marruecos concluye que hubo un impacto positivo en la mejora de la situación financiera para aquellos que optaron por las microfinanzas. Sin embargo, el estudio señala que solo el 13 por ciento de las personas en el grupo de intervención en el experimento de Marruecos decidieron no tomar las microfinanzas, a pesar de que habría un impacto positivo (Crepón, 2015). Este dato plantea una pregunta evidente: ¿Por qué aproximadamente el 87 por ciento de las personas en el grupo de intervención en el experimento de Marruecos optaron por no utilizar las microfinanzas si se espera un impacto positivo? A partir de la pregunta mencionada, surge una nueva pregunta: ¿cuáles son los factores que determinan la adopción de las microfinanzas o, más bien, si existen características distintivas entre las personas que toman el apoyo financiero y las que no? Este trabajo retoma los datos recabados por Crepón, y a partir de ellos, analiza la heterogeneidad entre los individuos en la adopción del programa y las características que diferencian a quienes lo adoptan de quienes no. Esto podría proporcionar información para aumentar la motivación de los participantes en la elección de las prácticas de educación microfinanciera.

Por lo tanto, el propósito de este trabajo es mostrar la existencia de heterogeneidad entre el grupo que optó por el apoyo financiero y el que no dentro del grupo de tratamiento. El estudio de la heterogeneidad es importante ya que el efecto de tratamiento heterogéneo (HTE por sus siglas en inglés) proporcionaría las razones por las que las microfinanzas son un programa benéfico para algunos y a otros no. Mientras hay individuos asignados al tratamiento que no toman las microfinanzas, el impacto general no sería el óptimo. Por lo tanto, es importante estudiar la heterogeneidad para obtener información de los participantes

para optimizar recursos y mejorar el diseño de estudios que pueden aumentar el porcentaje de individuos que deciden tomar el programa.

El análisis de heterogeneidad requiere identificar si existen diferencias entre las características personales de los individuos en el experimento para aquellos que optaron por las microfinanzas y otros que no. El trabajo se procederá por medio de los tres métodos propuestos por Chernozhukov que son: el mejor predictor lineal, efecto promedio de tratamiento por grupos ordenados y análisis de clasificación (Chernozhukov et al., 2018). Estos métodos permiten analizar la existencia de la heterogeneidad, ordenar los grupos por su efecto, identificar sus diferencias y detectar las características que están más correlacionadas con la heterogeneidad. Estos métodos sobresalen al evaluar el impacto del tratamiento de manera personalizada para cada observación. Estos métodos son apropiados para cumplir el objetivo de este trabajo ya que permiten detectar heterogeneidad en el efecto de tratamiento, computar el efecto de tratamiento por diferentes grupos de observaciones y describir qué covariables están correlacionadas en mayor medida con la heterogeneidad.

El resultado del método de mejor predictor lineal señala que sí existe heterogeneidad entre los individuos que toman las microfinanzas y respecto a los que no. Además, en el momento de ordenar las observaciones en cinco grupos por medio del efecto promedio de tratamiento por grupos ordenados, los dos grupos con mayores y menores afectaciones de tratamiento muestran una diferencia notable en el efecto de tratamiento. Además, el método de análisis de clasificación detecta variables específicas que están mayormente correlacionadas a la heterogeneidad mientras que otras variables no provocan la heterogeneidad. Las variables que muestran la heterogeneidad tienen que ver con las actividades de negocio de los individuos y el historial de préstamo.

El estudio se organiza de la siguiente manera. En primer lugar, se presentan estudios previos sobre microfinanzas, incluyendo investigaciones relacionadas con América Latina. Adicionalmente, se abordará la literatura central en la que se basa este estudio, centrada en el caso de Marruecos. En segundo lugar, se presenta la literatura que sustenta la metodología de análisis de heterogeneidad, la base de datos utilizada y la línea de base de dicha base de datos. En tercer lugar, se expone el trabajo empírico para el análisis de la heterogeneidad y se presentan los resultados correspondientes. Por último, el texto se concluye con las

implicaciones del estudio y posibles recomendaciones para la implementación de los hallazgos en las políticas públicas.

2. Revisión de literatura

En primer lugar, es importante examinar casos previos de microfinanzas con ensayos controlados aleatorios. Un estudio realizado en Bosnia y Herzegovina evaluó el impacto de un programa de microfinanzas a través de un ensayo controlado aleatorio en los clientes que previamente habían sido rechazados para préstamos por una institución microfinanciera (Augsburg et al., 2015). El trabajo se llevó a cabo en colaboración con la institución microfinanciera bosnia y se enfocó en 1,196 individuos del grupo de tratamiento de la población marginada, que en su mayoría consistía en hombres casados con un bajo nivel de educación. El estudio evaluó el impacto de pertenecer al grupo de tratamiento, es decir, el efecto de las microfinanzas, en términos de ingresos, autoempleo, valor del patrimonio, propiedad de inventario y oportunidades laborales para diferentes grupos de edad. El estudio concluyó que no hubo un aumento en los ingresos totales del hogar, pero sí encontró evidencia de que la adopción de las microfinanzas aumenta la proporción de trabajo remunerado, el autoempleo, el valor del patrimonio y las oportunidades laborales para individuos de entre 16 y 19 años, además de reducir los consumos e incrementar las ganancias. Si bien este riguroso ensayo controlado aleatorio se centró en individuos previamente rechazados por la institución microfinanciera, está claro que muestra efectos modestos pero un impacto positivo en los hogares que a largo plazo puede influir en beneficios económicos.

Por otro lado, el estudio de Banerjee presenta resultados que cuestionan el impacto idealizado de las microfinanzas (Banerjee et al., 2015). El programa de microfinanzas, diseñado a través de un ensayo controlado aleatorio en Hyderabad en India, se centró en mujeres de entre 18 y 59 años. El diseño no exigía necesariamente que las participantes tuvieran un plan de negocios al recibir el dinero, dejando la decisión de uso y creación de negocios en manos de los individuos. Banerjee y coautores concluyeron que las microfinanzas en Hyderabad no mostraron un aumento en la salud, la educación ni el empoderamiento a largo plazo de las mujeres, pero sí señaló un incremento en las inversiones y los beneficios de negocios de pequeña escala. Además, las participantes exhibieron un comportamiento modificado, aumentando sus gastos en bienes duraderos y reduciendo el consumo de bienes impulsivos. Está claro que el programa de microfinanzas no produce un impacto drástico. Sin embargo, es importante notar que la mayoría de las veces, el trabajo se

enfoca en individuos no calificados para las condiciones de préstamo, por lo que es poco realista esperar un desempeño significativo desde el principio. Aun así, las microfinanzas claramente fomentan un entorno que permite a individuos no calificados iniciar un negocio y adoptar ciertos comportamientos económicamente más beneficiosos.

Al revisar los estudios realizados en América Latina, es relevante destacar un caso ocurrido en Perú (Aguilar, 2013). Este trabajo analiza el impacto de la expansión de las microfinanzas en 24 regiones peruanas que experimentaron un crecimiento en la actividad económica entre 2001 y 2008. El análisis se basa principalmente en un modelo de datos panel y los resultados demuestran que el programa tuvo un impacto positivo, contribuyendo con un 4 por ciento al crecimiento del PIB per cápita. Esta intervención se destaca por su amplia escala en el ámbito de las microfinanzas, y está claro que su expansión tuvo un efecto beneficioso en el crecimiento económico de las regiones.

Otro caso relevante es el estudio boliviano (Garcia et al., 2021). Lo distintivo de este estudio es que evalúa tanto el impacto del servicio no financiero como el de los microcréditos dentro del programa. El programa llamado *Bolivian MFI* involucra a 600 granjeros en el altiplano boliviano, proporcionándoles apoyo técnico que incluye asesoramiento agronómico, veterinario y entrenamiento para el emprendimiento. En este análisis, los autores optaron por utilizar la regresión transversal con datos de panel. La conclusión del estudio es que el apoyo técnico representa un beneficio adicional en términos de ganancias mensuales y producción lechera.

A continuación, hay estudio que pone en cuestión los efectos de los estudios de las microfinanzas. A pesar de contar con pruebas de múltiples evaluaciones aleatorias de microcréditos, la falta de consenso sobre los resultados ha sido obstaculizada por interrogantes acerca de la validez externa (Meager, 2019). En su análisis, simultáneamente estima el efecto promedio y la heterogeneidad de los efectos en siete estudios utilizando modelos bayesianos jerárquicos. Se concluye que el impacto en las variables de negocios y consumo en los hogares es poco probable que sea significativamente transformador y podría carecer de importancia. Además, se encuentra una validez externa razonable, indicando que la verdadera heterogeneidad en los efectos es moderada y alrededor del 60 por ciento de la heterogeneidad observada se atribuye a variaciones en el muestreo. Los hogares con

experiencia previa en negocios experimentan efectos más notables, pero también más heterogéneos. Las características económicas de las intervenciones de microcrédito demuestran ser mejores predictores de la variación en los efectos que los protocolos de evaluación utilizados en los estudios.

A pesar de los resultados positivos y modestos obtenidos en trabajos previos, es importante señalar que no siempre es posible medir su impacto, ya que los individuos en el grupo de tratamiento pueden optar por no utilizar las microfinanzas. En este sentido, vale presentar el estudio de la intervención de las microfinanzas en Marruecos y las implicaciones del porcentaje de la adopción al resultado del experimento (Crepón et al., 2015). Este servirá como la literatura principal del análisis de este trabajo. El estudio de Crepón y coautores evaluó la intervención en Marruecos que constituye en un programa de microfinanzas en 2006 en áreas rurales, utilizando un diseño de ensayo controlado aleatorio. Los resultados empíricos revelaron un aumento en los beneficios económicos y la inversión en activos utilizados para el autoempleo entre los participantes. Además, el análisis mostró una disminución de los ingresos derivados de trabajos informales, como empleos temporales o de medio tiempo. Aunque el estudio no indicó un aumento en los ingresos o el consumo, los efectos evidenciaron el potencial de transformación del empleo informal en formal y el aumento de las ganancias futuras atribuidas al aumento de la inversión en este sector. A pesar de estos resultados prometedores, es importante destacar que solo el 13 por ciento de los individuos asignados al grupo de intervención optaron por las microfinanzas. Este bajo porcentaje de adopción de las microfinanzas, es decir, la aceptación de préstamos en el grupo de tratamiento, desempeña un papel significativo en el diseño del programa. Esto hace que el impacto sea menos probable de ser significativo cuando el porcentaje es tan bajo, como en el caso de Marruecos. Sin embargo, esta es una característica común en el ámbito de las microfinanzas, donde la mayoría de los estudios demuestran una baja proporción de adopción de préstamos en el grupo de tratamiento. Esto plantea una cuestión importante: explorar las diferencias entre los individuos que adoptan las microfinanzas y aquellos que no lo.

En el ámbito del aprendizaje automático, recientes investigaciones se han centrado en las técnicas de inferencia causal, que permiten descubrir los factores que explican el efecto diferenciado de una intervención entre distintos individuos. El trabajo de Chernozhukov se

enfoca principalmente en los tres métodos: Mejor Predictor Lineal (BLP), Efectos Promedio de Tratamiento por Grupos Ordenados (GATES) y Análisis de Clasificación (CLAN) para evaluar heterogeneidad (Chernozhukov et al., 2018). Los autores hacen una aplicación del método para analizar el efecto del tratamiento heterogéneo en un ensayo controlado aleatorio relacionado con la inmunización en la India. Estos métodos de análisis presentados en la literatura son relevantes para esta tesis, ya que el enfoque fundamental de este trabajo es verificar la existencia de heterogeneidad entre dos grupos: aquellos que optaron por las microfinanzas y aquellos que no lo hicieron.

Basado en esta metodología, Baiardi expone las ventajas de la inferencia causal y los métodos del aprendizaje automático presentados en el trabajo de Chernozhukov et al. (2018) (Baiardi et al., 2021). Baiardi y coautores mencionan que los modelos del aprendizaje automático son útiles, ya que permiten detectar la heterogeneidad en el efecto del tratamiento, calcular el tratamiento para diferentes grupos y describir qué covariables están más correlacionadas con la heterogeneidad. Además, el trabajo de análisis se basa en los métodos presentados para el experimento sobre el desempeño escolar según la intervención del desarrollo profesional de docencia en China. Esta revisión del experimento en China confirma el resultado original que no hubo un efecto significativo, pero también demuestra que existe heterogeneidad en el efecto del tratamiento según las características de los profesores, además de señalar que la educación de Licenciatura fue una covariable más correlacionada con la heterogeneidad.

Por lo tanto, resulta valioso analizar la heterogeneidad de los efectos del tratamiento en el caso de Marruecos. Más allá de la información que se puede obtener al estudiar el efecto de la intervención del experimento, este enfoque puede proporcionar una mayor comprensión para futuros diseños de programas de microfinanzas. En este estudio, es relevante investigar los factores que distinguen entre los individuos que adoptan las microfinanzas y aquellos que no lo hacen, con el objetivo de contribuir a las políticas que aumenten la probabilidad de adopción. Para lograrlo, este trabajo se basará en el análisis de Crepón en Marruecos, ofreciendo un enfoque distinto para analizar la heterogeneidad entre los dos grupos de adopción y estudiar qué factores explican la adopción de las microfinanzas. En este análisis,

utilizaremos las técnicas de aprendizaje automático presentadas en el trabajo de Chernozhukov como metodología.

3. Contexto y Evaluación

A continuación, se presenta el contexto y evaluación del experimento realizado en Marruecos en 2015:

Al Amana es la institución microfinanciera más grande de Marruecos. Desde su comienzo en 2000, Al Amana se ha expandido desde áreas urbanas hasta áreas rurales. Entre 2006 y 2007, la dicha institución abrió alrededor de 60 nuevas sucursales en las áreas no densamente pobladas. El producto principal del Al Amana en áreas rurales es préstamo de responsabilidad colectiva. Se formaba un grupo por tres o cuatro integrantes que se ponen del acuerdo de garantizar mutuamente el reembolso de sus préstamos. La cantidad de préstamo varía desde 1,000 hasta 15,000 dirham marroquí (MAD, 124-1,855 dólares estadounidenses) por individuo. El reembolso de los préstamos puede tardar entre 3 y 18 meses, mediante pagos semanales, bimestrales o mensuales. Para las actividades ganaderas se concede un período de gracia de dos meses. Los tipos de interés de los préstamos rurales oscilan entre el 12,5% y 14,5% en el momento del estudio (es decir, entre 2006 y 2009). Para poder optar a un préstamo colectivo de responsabilidad civil, el solicitante debe tener entre 18 y 70 años, ser titular de un documento nacional de identidad, tener un certificado de residencia y haber ejercido una actividad económica distinta de la agricultura no ganadera durante al menos 12 meses. A diferencia de la mayoría de las instituciones microfinancieras del mundo, Al Amana no restringe sus préstamos exclusivamente a las mujeres, pero sí suele exigir que los agentes de crédito tengan al menos un 35% de mujeres entre sus clientes. Sin embargo, este requisito se suprimió primero entre las sucursales participantes en el estudio y después en todas las sucursales. A partir de marzo de 2008, los préstamos individuales para vivienda y negocios no agrícolas en las zonas rurales. Estos préstamos eran de mayoría cuantía (hasta 48,000 MAD, o uno 6,000 dólares estadounidenses), tenían una serie de requisitos adicionales y estaban dirigidos a clientes que pudieran aportar algún tipo de garantía. Durante el periodo de estudio, los hogares casi sólo solicitaban préstamos de responsabilidad colectiva, por lo que este estudio es principalmente una evaluación de ese producto (Crepon et al., 2015. p.127-130.).

El diseño del estudio de Crepón siguió la expansión de Al Amana en zonas no densamente pobladas entre 2006 y 2007. Antes de la apertura de casa sucursal, se recopilaron datos de al menos seis aldeas situadas en la periferia de las zonas de intervención que podrían haber sido incluidos o excluidos de la zona de captación de la sucursal. Se excluyeron las aldeas cercanas a un centro de población rural o situadas a lo largo de una ruta hacia otras zonas atendidas por la sucursal, ya que ello habría perturbado el desarrollo de Al Amana. También, se excluyó un número reducido de pueblos en los que había otras instituciones microfinancieras (alrededor del 2%). A continuación, los pueblos seleccionados se emparejaron por parejas en función de las características observables (número de hogares, accesibilidad al centro de la comunidad, infraestructuras existentes, tipo de actividades realizadas por los hogares, tipo de actividades agrícolas). Se conservan dos parejas por rama para la evaluación. En cada par, se asignó aleatoriamente un pueblo al tratamiento y el otro al control. En total, se incluyeron en la evaluación 81 parejas pertenecientes a 47 sucursales. Entre 2006 y 2007, Al Amana abrió nuevas sucursales en seis fases. Estas, a efectos de la evaluación de este experimento, dividieron esta expansión en cuatro periodos y realizaron la encuesta de referencia en cuatro oleadas de operaciones de campo entre abril de 2006 y diciembre de 2007. La estrategia de muestro siguió un enfoque novedoso para maximizar el poder de la evaluación para detectar tanto los efectos directos como los efectos a nivel de población del acceso a la microfinanza. En concreto, seleccionaron dos muestras de hogares: una con los que tenían más probabilidades de convertirse en clientes de la institución microfinanciera y otra con una selección aleatoria de hogares del resto de la población. El uso de la primera muestra aumenta la probabilidad de detectar un efecto sobre lo que tienen más probabilidades de convertirse en clientes, si es que lo hay. La utilización conjunta de ambas muestras, con las ponderaciones adecuadas, permite medir el efecto sobre el conjunto de la población de la oferta de acceso a los servicios microfinancieros (Crepon et al., 2015. p.127-130.).

Para ello, en cada una de las 14 aldeas de la primera oleada, muestrearon 100 hogares a los que administraron una encuesta de referencia completa. En los pueblos con menos de 100 hogares, los encuestaron a todos. Esta fase tuvo lugar en abril-mayo de

2006, seis meses antes del inicio previsto de la segunda fase. Utilizaron los datos de esta encuesta y los datos administrativos sobre la concesión de créditos en las aldeas tratadas durante los seis primeros meses (comunicados semanalmente por los agentes de crédito) para estimar un modelo de predicción de la probabilidad de obtener un préstamo de cada hogar (Crepon et al., 2015. p.127-130.).

Dos años después del inicio de cada oleada de la intervención de Al Amana, realizaron una encuesta final de hogares, basada en el mismo instrumento, en los mismos 81 pares de aldeas (mayo de 2008-enero de 2010 y 4,465 hogares entrevistados en la línea de base fueron muestreados para la línea de base final. De ello, el 92 por ciento (4,118 hogares) fue entrevistado de nuevo. Para maximizar la potencia, se seleccionaron otros 1,433 hogares (también de crédito y los datos de la encuesta abreviada). Para seleccionar estos hogares adicionales, volvieron a estimar el modelo para predecir la probabilidad de obtener un préstamo para cada hogar utilizando datos administrativos sobre quién había obtenido un préstamo en el momento de finalizar el estudio. Los datos administrativos sobre los prestatarios en el momento de la encuesta final, emparejados con datos recogidos en la encuesta abreviada antes de la introducción del microcrédito (y, por tanto, no afectados por la introducción), se actualizaron las variables dependientes de dos años, y se volvieron a estimar los coeficientes del modelo. Esto nos permitió identificar mucho mejor a los posibles prestatarios, en total, se realizó a 5,551 hogares (Crepon et al., 2015. p.127-130.).

4. Metodología

Como anteriormente fue mencionado, el análisis principal de este trabajo es revisar el impacto de diversos factores que caracterizan a los individuos sobre la probabilidad de la adoptar las microfinanzas. Para ello, la metodología principal utilizada será las técnicas del aprendizaje automático genérico. El enfoque de aprendizaje automático genérico propuesto por Chernozhukov et al. (2018) se centra en la generación de predicciones relativas a los efectos del tratamiento a nivel individual, y utiliza estas predicciones como base para realizar inferencias sobre las propiedades subyacentes por medio de repetición de la división de datos (Chernozhukov et al., 2018). Este método se destaca por la capacidad de estimar el impacto del tratamiento para cada observación de forma personalizada. Además, las estimaciones y predicciones basadas en los datos divididos de forma repetida facilitan evitar los posibles riesgos de sobreajuste (*overfitting*) y entender la complejidad de la heterogeneidad en el efecto del tratamiento en comparación con una división de muestra tradicional. Este último promueve comprender qué factores sirven para distinguir entre los grupos más afectados o menos afectados. El método de Chernozhukov y coautores logra analizar la heterogeneidad a través de las tres técnicas como las siguientes: BLP, GATES y CLAN. Por medio de estos métodos, los objetivos se analizan de forma siguiente. Primero, analizar si existe heterogeneidad entre los grupos de control y tratamiento según las características personales por medio de mejor predictor lineal. Segundo, evaluar cuál es el nivel de impacto de las variables para distintos subgrupos, lo cual sería cuantificar el impacto de los factores a la toma de préstamo por efectos promedio de tratamiento por grupos ordenados. Por último, este método utiliza análisis de clasificación para lograr identificar qué factores son los que atribuyen a la dicha heterogeneidad dado que no necesariamente todas las variables generan un impacto a la dicha distinción en mayor medida.

4.1 Introducción de la notación y el significado de las variables.

Para presentar el trabajo en práctica, el trabajo considera la variable Y como variable de interés de este análisis que explica la probabilidad de aceptar las microfinanzas. D que es una variable binaria que implica pertenecer al grupo de tratamiento asignado. Z que es el vector de covariables que definen las características personales de cada individuo como especificaciones de hogar o características del negocio de cada hogar. Además, se definen

$b_0(Z) = E[(Y(0)|Z]$ como el efecto promedio de línea base, por lo que sería el efecto de cada individuo cuando aún no hubo la intervención y $s_0(Z) = E[(Y(1)|Z] - E[(Y(0)|Z]$ como efecto promedio condicional de tratamiento (CATE por sus siglas en inglés). Este último explica el efecto promedio de observaciones al calcular el efecto al pertenecer al grupo de tratamiento asignado menos el efecto al no pertenecer al grupo de tratamiento. Utilizando la muestra auxiliar, se obtiene los estimadores del aprendizaje automático para el promedio condicional de línea base y el efecto promedio condicional de tratamiento. Después, procede en computar los efectos predichos de la línea base, $b(Z)$ y los efectos predichos del tratamiento $S(Z)$ al calcular la diferencia entre las predicciones para los grupos de tratamiento y control.

A continuación, se procede en describir los objetivos y funcionamiento de los tres métodos planteados por Chernozhukov y cabe precisar los como los siguientes (Chernozhukov, 2018; Baiardi, 2021; Kitzmüller, 2021):

4.2 Mejor Predictor Lineal (BLP en inglés)

Este método sirve para demostrar la evidencia de heterogeneidad basada en las características que poseen los individuos. Los parámetros son obtenidos utilizando la muestra principal por medio de estimar las regresiones de Mínimos Cuadrados Ordinarios. La ecuación que analizar es la siguiente:

$$Y = \alpha_0'X_1 + \alpha_1(D - p(Z)) + \alpha_2(D - p(Z))(S(Z) - \overline{S(Z)}) + \varepsilon \quad (1)$$

donde X_1 es un conjunto de covariables que incluye los efectos predichos del control; $p(Z) = P[D = 1|Z]$ es la puntuación de propensión y $\overline{S(Z)}$ es el promedio de las estimaciones del efecto del tratamiento previsto en la muestra principal. Se puede notar que el componente $(D - p(Z))$ es parte del regresor $(D - p(Z))(S(Z) - \overline{S(Z)})$, por lo que el componente es ortogonal con el regresor a todas las covariables que pertenece a Z . El coeficiente α_1 refleja el efecto promedio de tratamiento (ATE por sus siglas en inglés) mientras α_2 cuantifica qué tan bien el predictor de proxy aproxima la heterogeneidad del tratamiento. Por el último, Chernozhukov et al. (2018) muestra que si el coeficiente α_2 es igual a cero, esto implicaría la ausencia de heterogeneidad o que predicciones para $S(Z)$ fallan en capturar la heterogeneidad.

Por lo tanto, el primero paso de utilizar este método sería revisar si α_2 es estadísticamente significativamente diferente de cero.

4.3 Efectos Promedio de Tratamiento por Grupos Ordenados (GATES)

Cuando el primer método logra probar que existe el efecto heterogéneo de tratamiento, el método GATES sirve para dividir las observaciones desde la muestra principal en k grupos: G_1, G_2, \dots, G_K basado en los efectos de tratamiento. La técnica, posterior de partir la muestra a quintiles, procede estimar la siguiente regresión ponderada:

$$Y = \alpha_0' X_1 + \sum_k^K \alpha_k [(D - p(Z))] 1(G_K) + v \quad (2)$$

donde $1(G_K)$ es un indicador de función de que si una observación pertenece al grupo k y donde los pesos son iguales a la ecuación (1). El parámetro α_k muestra el resultado de los GATES. Esto permite analizar las observaciones para grupo G_1 con los efectos del tratamiento más bajos y para grupo G_K con los efectos más altos. Por lo tanto, si la diferencia $\alpha_k - \alpha_1$ representa un valor significativamente diferente de cero, confirmaría de nuevo el efecto de heterogeneidad entre los grupos más afectados y menos afectados. De esta forma, el método señala la variación de sus efectos entre los distintos grupos.

4.4 Análisis de Clasificación (CLAN)

Por último, la técnica de análisis de clasificación tiene como propósito de analizar las propiedades o características están asociados en mayor medida a los grupos más afectados y menos afectados. Aquí, los dos grupos anteriormente mencionados representan G_5 y G_1 , respectivamente. Sea $g(Y, D, Z)$ un vector de las características de una unidad observacional; el análisis de clasificación es la comparación de promedio de las características de los grupos más y menos afectados:

$$\delta_1: E[g(Y, D, Z)|G_1] \text{ y } \delta_K: E[g(Y, D, Z)|G_K] \quad (3)$$

donde δ_1 y δ_K se identifican debido a que son los promedios de las variables que son directamente observadas. Este método cuantifica las diferencias entre los grupos más y menos afectamos y señala las covariables que se asocian a la heterogeneidad de efecto promedio condicional de tratamiento.

4.5 Modelos de aprendizaje

Este estudio utilizará modelos de aprendizaje para revisar su desempeño y posteriormente optar por uno de ellos para analizar la heterogeneidad. Un modelo de aprendizaje se refiere a la representación del proceso del mundo real (Géron, 2019) Estos modelos son entrenados basados en los datos históricos para hacer predicciones o decisiones sin ser explícitamente programados. Diferentes tipos de modelos incluyen regresiones lineales, bosques aleatorios, máquina de vectores de soporte, redes neuronales y más. Entre ellos, este estudio toma los tres modelos como bosque aleatorio, red neuronal y vecinos más cercanos para comprobar su rendimiento y optar por uno de ellos para el análisis de heterogeneidad. Primero, los bosques aleatorios son una combinación de predictores de árboles de decisión de manera que cada árbol depende de los valores de un vector aleatorio muestreado de manera independiente y con la misma distribución para todos los árboles en el bosque (Breiman, 2001). Las estimaciones internas supervisan el error, la fortaleza y la correlación, y se utilizan para mostrar la respuesta al aumentar el número de características utilizadas en la división. Las estimaciones internas también se utilizan para medir la importancia de las variables. Estas ideas también son aplicables a la regresión. Segundo, una red neuronal se define como un modelo computacional inspirado en el cerebro humano que consiste en capas interconectadas de nodos o neuronas (Goodfellow, 2016). Cada conexión entre nodos tiene un peso que se ajusta durante el entrenamiento del modelo para aprender patrones y representaciones útiles en los datos. El libro de Goodfellow describe las redes neuronales como sistemas capaces de realizar tareas complejas de aprendizaje a partir de datos, destacando su capacidad para aprender representaciones jerárquicas profundas. Por último, el algoritmo de vecinos más cercanos se define como un método de aprendizaje supervisado que se utiliza tanto para clasificación como para regresión (Muller et al., 2017). La esencia de este modelo radia en asignar una etiqueta o predecir un valor para una nueva instancia basándose en las etiquetas de las instancias vecinas más cercanas en el espacio de características. La elección del número de vecinos, representado por “k”, y la métrica de distancia son factores críticos que afectan el rendimiento del algoritmo.

4.6 Algoritmo

Como se mencionó anteriormente, el experimento a través de un ensayo controlado aleatorio sobre las microfinanzas en Marruecos reveló un impacto positivo en la capacidad económica de los individuos. Sin embargo, este estudio no proporciona más información que el impacto de la adopción de las microfinanzas. No ofrece un análisis de la heterogeneidad en el efecto del tratamiento, como, por ejemplo, qué diferencias existen entre los grupos que toman el préstamo y aquellos que no, o qué factores están influyendo en mayor medida en la decisión de adoptar las microfinanzas, o qué grupos son más afectados por el tratamiento. Resolver estas preguntas previas es relevante, ya que ayudaría en la selección adecuada de los participantes, identificando las variables que tienen un mayor impacto, y permitiría una asignación óptima de recursos al identificar el nivel de impacto diferenciado entre los grupos.

En este sentido, el uso de las técnicas del aprendizaje automático genérico ayuda a estudiar resultados más diversos que solo el impacto de las microfinanzas a partir del experimento en Marruecos y permitiría utilizar los resultados para el mejor diseño del programa de microfinanzas. Adicionalmente, esta metodología complementa la limitación de un ensayo controlado aleatorio al estimar el impacto de ambos casos para cada individuo asignado en el grupo de tratamiento o control (Chernozhukov, 2018).

Para lograr el objetivo de análisis de heterogeneidad planteado por Chernozhukov, este trabajo utiliza los tres métodos previamente introducidos (Chernozhukov, 2018; Baiardi, 2021; Kitzmüller, 2021). Para ello, el enfoque se basa en la partición aleatoria de los datos en una muestra principal y auxiliar que tienen aproximadamente el mismo tamaño. A partir de la muestra auxiliar, se construye un estimador de aprendizaje automático llamado "predicador de proxy" que se utiliza para estimar el efecto promedio condicional del tratamiento. En este punto, se pueden utilizar diversos modelos de aprendizaje automático genérico, como redes elásticas, bosques aleatorios, redes neuronales, entre otros. Es importante destacar que los predictores de proxy pueden ser potencialmente sesgados, y la consistencia no es un requisito (Baiardi, 2021). El análisis toma estos predictores como una aproximación y los utiliza para estimar e inferir las características del efecto promedio condicional del tratamiento. Basándose en la muestra principal y los predictores de proxy, se pueden calcular los estimadores para las técnicas de Chernozhukov y coautores y,

posteriormente, se realiza la inferencia en varias particiones de los datos en las muestras auxiliar y principal.

Habiendo presentado las funciones de los tres métodos de Chernozhukov que se utilizarán en este análisis, se procede al algoritmo (Chernozhukov, 2018; Baiardi, 2021; Kitzmüller, 2021). El trabajo empírico está realizado en el lenguaje de programación R con el uso de paquete del aprendizaje automático genérico llamado *GenericML()*. Este trabajo toma la base de datos del trabajo del experimento en Marruecos que tendrá las especificaciones a detalle en la sección de datos. El primer paso es la partición aleatoria de las observaciones a dos grupos del mismo tamaño que serán una muestra principal y otra muestra auxiliar. Después, se selecciona un set de variables como un vector Z que explican las características de individuos. Con el apoyo de los modelos del aprendizaje automático genérico, el trabajo fijará el número de iteraciones como 1000 y observará los resultados. En la generación de los datos, se construyen nuevos valores al imponer la condición de pertenecer al tratamiento ($D = 1$) una vez e imponer la condición de no pertenecer al tratamiento ($D = 0$) que sería $b(Z)$. Con estos datos y el vector Z , procede predecir los valores de la variable de Y para la muestra principal con el entrenamiento realizado anteriormente con la muestra auxiliar. La diferencia de los valores predichos de Y para la imposición del tratamiento sería el efecto promedio condicional del tratamiento $S(Z)$.

Posteriormente, el trabajo calcula la puntuación del pareamiento mediante un modelo logit de la variable de tratamiento en función del vector Z y obtiene los valores ajustados llamados $p(Z)$ para la muestra principal (Chernozhukov, 2018; Baiardi, 2021; Kitzmüller, 2021). Luego, se calcula la media muestral de $S(Z)$ y la diferencia de las probabilidades entre el grupo asignado al tratamiento y el grupo no asignado para la muestra principal. Como siguiente paso, se genera el vector de los valores $S(Z) - \overline{S(Z)}$ como la diferencia de cada $S(Z)$ predicho para cada observación y la media muestral de $S(Z)$. Adicionalmente, se crean otros vectores de los valores como $(D - p(Z))$: la variable indicadora de tratamiento menos la puntuación del pareamiento, $p(Z) * S(Z)$ y $(D - p(Z)) * S(Z) - \overline{S(Z)}$.

En seguida, se estima la regresión (1) (Chernozhukov, 2018; Baiardi, 2021; Kitzmüller, 2021). Para que el resultado tenga sentido para el análisis de heterogeneidad, primero, el coeficiente de $(D - p(Z))$ debe ser similar al porcentaje de adopción en el grupo de

tratamiento en el experimento en Marruecos, 13 por ciento. Segundo, el coeficiente de $(D - p(Z)) * S(Z) - \overline{S(Z)}$ debe ser distinto a cero con un valor estadísticamente significativo para mostrar el indicativo de heterogeneidad. Esto concluiría el análisis del mejor predictor lineal.

A continuación, este trabajo elige dividir las observaciones en 5 grupos y ordenar por el efecto del tratamiento de individuos en quintiles (Chernozhukov, 2018; Baiardi, 2021; Kitzmüller, 2021). Con el uso de paquete del aprendizaje automático genérico, se graficarán las funciones para visualizar los grupos con los efectos del tratamiento más altos y bajos en el 20 por ciento. Esto sería el uso del método de efectos promedio de tratamiento por grupos ordenados y se podrá analizar la diferencia de los dos grupos para el efecto de tratamiento. Finalmente, el trabajo intentará utilizar cada variable del del vector Z y graficar sus efectos por los grupos para identificar qué variables visualizan mayor efecto de heterogeneidad. Esto sería el análisis de clasificación.

5. Base de datos

Los datos de este análisis provienen de un ensayo controlado aleatorio llevado a cabo como parte del programa llamado *Al Amana* en Marruecos, que ofrece microfinanzas como intervención (Crepón et al., 2015). El diseño del estudio monitoreó la expansión del programa en áreas poco densamente pobladas entre 2006 y 2007. Se excluyeron de este experimento los pueblos donde se habían recopilado datos antes de la apertura de sucursales en las áreas de intervención, así como los pueblos cercanos al centro de población rural o cubiertos por otras sucursales y aquellos donde ya existía una institución de microfinanzas. La muestra consistió en 4934 individuos de entre 18 y 70 años en seis pueblos poco densamente poblados durante el período mencionado. Estas observaciones se dividieron en dos grupos: el grupo de tratamiento, que tenía derecho a acceder a las microfinanzas, y el grupo de control, que no recibió la intervención.

Dado que se trata de un ensayo controlado aleatorio, la partición se realizó de forma aleatoria, emparejando a los individuos en pares según características observables, como datos personales y actividades económicas del hogar. De esta manera, se seleccionó un pueblo para recibir la intervención de microfinanzas (grupo de tratamiento), mientras que otro se mantuvo como grupo de control para comparar los efectos. Debido a la aleatorización a nivel de pueblos, los datos muestran agrupación (cluster) y los errores estándar se calcularon a nivel de pueblo. En total, el experimento incluyó 81 pares de pueblos con 47 sucursales. Esta base de datos, debido a las características particulares del programa de microfinanzas, incluye una variable binaria llamada 'adopción' que indica si los individuos del grupo de tratamiento decidieron o no tomar las microfinanzas. Es importante destacar que la probabilidad de adopción en las observaciones del grupo de tratamiento fue del 13 por ciento. En otras palabras, solo el 13 por ciento de los individuos en el grupo de tratamiento optó por las microfinanzas.

Por lo tanto, la variable principal de interés en este análisis es la 'adopción', una variable binaria que indica si los individuos asignados al grupo de tratamiento optaron por tomar las microfinanzas o no. La variable D representa la pertenencia al grupo de tratamiento o intervención, como se explicó anteriormente, y el vector Z incluye las características de los individuos en las observaciones. La selección de un conjunto adecuado de variables en el

vector Z es crucial para capturar la heterogeneidad. En este sentido, se eligieron variables que se espera que tengan un impacto significativo en la probabilidad de adopción.

Es evidente que las variables demográficas y económicas desempeñan un papel importante en la diferenciación entre el grupo que optó por el apoyo financiero y el que no lo hizo (Crepón et al., 2015). En cuanto a la información sobre préstamos previos, el estudio de Crepón y coautores proporcionó información valiosa. Como se ha destacado en la literatura, el acceso previo a cualquier tipo de préstamo es un factor relevante en la decisión de optar por las microfinanzas. Sin embargo, los autores mencionan que el monto del préstamo no es una variable crucial una vez que se ha tenido acceso a un préstamo. Basándonos en esta información, en este análisis excluirémos los montos de los préstamos de *Al Amana* y de otras instituciones, tanto formales como informales, y consideraremos únicamente si los individuos tuvieron acceso a préstamos.

Por lo tanto, el vector Z se construyó a partir de 43 características personales seleccionadas, que incluyen: características del hogar, como el número de miembros en la residencia, el número de adultos y niños en el hogar, si el hogar se considera pobre, el número de actividades realizadas por el hogar, el gasto total del hogar; características del jefe del hogar, como su edad, una variable ficticia que indica si es un hombre, si no tiene educación formal; variables relacionadas el acceso previo a préstamos de *Al Amana*, instituciones financieras formales e informales, sucursales, el monto de préstamos anteriores de *Al Amana*, instituciones financieras formales, instituciones financieras informales y sucursales, así como actividades de negocio, el número de actividades económicas, si los individuos tienen un autoempleo, venta, ahorro, gasto e inversión en agricultura, si son empleados o tienen autoempleo en agricultura, venta, ahorro, gasto e inversión en ganadería, si son empleados o tienen autoempleo en ganadería, el monto de consumo del hogar, el consumo de productos duraderos y no duraderos, el número de actividades realizadas por las mujeres del hogar, la distancia al souk (mercado), el riesgo de pérdida del 50 por ciento de cosechas o ganado, el riesgo de pérdida de ganado en los últimos 12 meses, si algún miembro del hogar tiene alguna enfermedad, si el hogar depende de trabajos temporales, e ingresos de otras actividades (Crepón et al., 2015).

5.1 Resumen estadístico de la base de datos

Dado que se han mencionado las características del ensayo controlado aleatorio en la introducción, es necesario confirmar que los datos recopilados entre el grupo de tratamiento y el grupo de control muestren homogeneidad en promedio (Crepón, 2015). Por lo tanto, se presenta la línea de base estadística para la comparación entre los dos grupos en la Tabla 1.

Tabla 1: Línea de base estadística

	Grupo de control				Tratamiento-Control	
	Obs.	Obs.	Media	DE	Coeff.	Valor p
<i>Panel A. Muestra inicial de Hogares</i>						
<i>Composición de hogar:</i>						
Núm. Residentes	4,465	2,266	5.14	2.70	0.04	0.583
Núm. Adultos (\geq mayor que 16 años)	4,465	2,266	3.45	1.99	0.03	0.564
Núm. Niños (\geq mayor que 16 años)	4,465	2,266	1.68	1.65	0.01	0.859
Cabeza masculina	4,465	2,266	0.935	0.246	0.001	0.813
Edad del cabeza de hogar	4,465	2,266	48	16	1**	0.012
Gastos totales	4,465	2,266	21,501	69,645	4,046	0.063
<i>Acceso a crédito:</i>						
Préstamo de Al Amana	4,465	2,266	0.007	0.084	-0.003	0.425
Préstamo de institución formal	4,465	2,266	0.060	0.238	0.030**	0.023
Préstamo informal	4,465	2,266	0.068	0.251	0.023***	0.006
Préstamo de sucursales	4,465	2,266	0.156	0.363	0.013	0.523
<i>Monto prestado desde:</i>						
Al Amana	4,465	2,266	34	460	-13	0.534
Institución formal	4,465	2,266	355	2340	92	0.188
Institución informal	4,465	2,266	248	2248	-8	0.880
Sucursales	4,465	2,266	528	1370	22	0.758
<i>Actividades de autoempleo:</i>						
Núm. Actividades	4,465	2,266	1.6	1.2	0.0	0.435
Agricultura	4,465	2,266	0.599	0.490	0.017	0.321
Inversión	4,465	2,266	13	72	0	0.775
Ventas	4,465	2,266	9335	36,981	-392	0.665
Gastos	4,465	2,266	3369	8,428	266	0.241
Ahorros	4,465	2,266	1271	3,505	-77	0.433
Empleado	4,465	2,266	22	95	-1	0.477
Autoempleado	4,465	2,266	61	102	5	0.122
Ganadería	4,465	2,266	0.533	0.499	0.042**	0.027
Inversión	4,465	2,266	397	1,912	67	0.2
Ventas	4,465	2,266	3,444	8,831	339	0.184
Gastos	4,465	2,266	4,111	10,891	386	0.206
Ahorros	4,465	2,266	10,249	17,032	1,066*	0.050

Empleado	4,465	2,266	7	49	-1	0.272
Autoempleado	4,465	2,266	111	158	7	0.215
Tiene negocio no granjero	4,465	2,266	0.217	0.412	-0.034**	0.011
Núm. Actividades por mujer	4,465	2,266	0.218	0.585	0.004	0.075
Distancia de souk (mercado)	4,465	2,077	20.1	25.2	0.2	0.87
<i>Ingreso desde:</i>						
Actividades de autoempleo	4,465	2,266	0.780	0.414	-0.016	0.163
Salarios diarios	4,465	2,266	0.580	0.494	-0.016	0.194
<i>Riesgos:</i>						
Pérdida de más del 50% de cosechas	4,465	2,077	0.106	0.308	0.004	0.642
Pérdida de más del 50% de ganadería	4,465	2,077	0.030	0.172	0.003	0.606
Si ha perdido ganadería en los últimos 12 meses	4,465	2,266	0.189	0.392	0.029**	0.012
Enfermedad de integrante de hogar	4,465	2,266	0.218	0.413	0.013	0.168
<i>Consumos:</i>						
Consumo total	4,465	2,266	2,272	1,349	28	0.440
Productos durables	4,465	2,266	2,227	1,295	20	0.559
Productos no durables	4,465	2,266	45	236	8	0.231
Si hogar es pobre.	4,465	2,266	0.247	0.431	0.002	0.858
<i>Panel B: Atrición</i>						
No encuestado/a en la línea final	4,465	2,266	0.068	0.252	0.017**	0.018

Notas: Unidad de observación: hogar. Panel A y B: muestra incluye todos los hogares encuestados en el inicio.

*** Significativo al nivel del 1 por ciento. ** Significativo al nivel del 5 por ciento* Significativo al nivel del 10 por ciento.

Fuente: (Crépon et al., 2015)

Como se muestra en la tabla, de las 4,465 personas que respondieron a la encuesta, el grupo se dividió en dos categorías: el grupo de control, que no recibió la intervención y consta de 2266 individuos, y el grupo de tratamiento, que tuvo la oportunidad de acceder a las microfinanzas. Al evaluar el valor de p en la última columna donde se expone la diferencia entre grupo de tratamiento y control, la mayoría de los valores son mayores de 0.05, lo cual implica un dato estadísticamente poco significativo. En conclusión, es posible afirmar que las variables contengan las características homogéneas. Esta evidencia estadística respalda la partición de la muestra está correctamente realizada en este ensayo controlado aleatorio.

6. Resultados

6.1 Algoritmo de mejor desempeño

Posterior de la construcción de la variable, es necesario elegir uno de los tres modelos de aprendizaje automático con el mejor desempeño antes de calcular los resultados. La tabla abajo expone los coeficientes de los tres modelos; red neuronal, vecinos más cercanos y bosque aleatorio:

Tabla 2: Desempeño de los tres modelos de aprendizaje automático

	CATES	GATES
Red Neuronal	0.001	0.029
Vecinos más cercanos	0.017	0.039
Bosque Aleatorio	0.038	0.049

Nota: CATES por sus siglas en inglés para efecto promedio condicional del tratamiento, GATES por sus siglas en inglés para los efectos promedio de tratamiento por grupos ordenados.

Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

El criterio de selección del modelo de aprendizaje automático óptimo está basado en que se obtengan mejores puntajes en CATES y GATES. La tabla muestra que el modelo bosque aleatorio posee el mejor desempeño, por lo que el análisis de heterogeneidad de este trabajo se procederá con bosque aleatorio.

6.2 Resultados del Mejor Predictor Lineal (BLP)

Una vez que el estudio identificó el modelo de Bosque Aleatorio como el más adecuado para el análisis de los objetivos, procedemos a observar los resultados basados en los tres métodos propuestos por Chernozhukov et al (2018). El primer paso consiste en analizar si existe heterogeneidad en la adopción de las microfinanzas en función de las características individuales. A continuación, la siguiente tabla presenta los resultados del BLP:

Tabla3: Resultados de BLP

	L.I. de I.C.	L.S.	Pr(< z)	ajdo.	Pr(> z)	ajdo.
Efecto promedio de tratamiento	0.168	0.148	0.187	1		0
Cuantificación del Predictor de Proxy para	1.240	1.115	1.363	1		0

Medir la Aproximación de la Heterogeneidad

Nota: L.I. significa límite inferior, I.C. intervalo de confianza, L.S. límite superior y ajdo. ajustado.

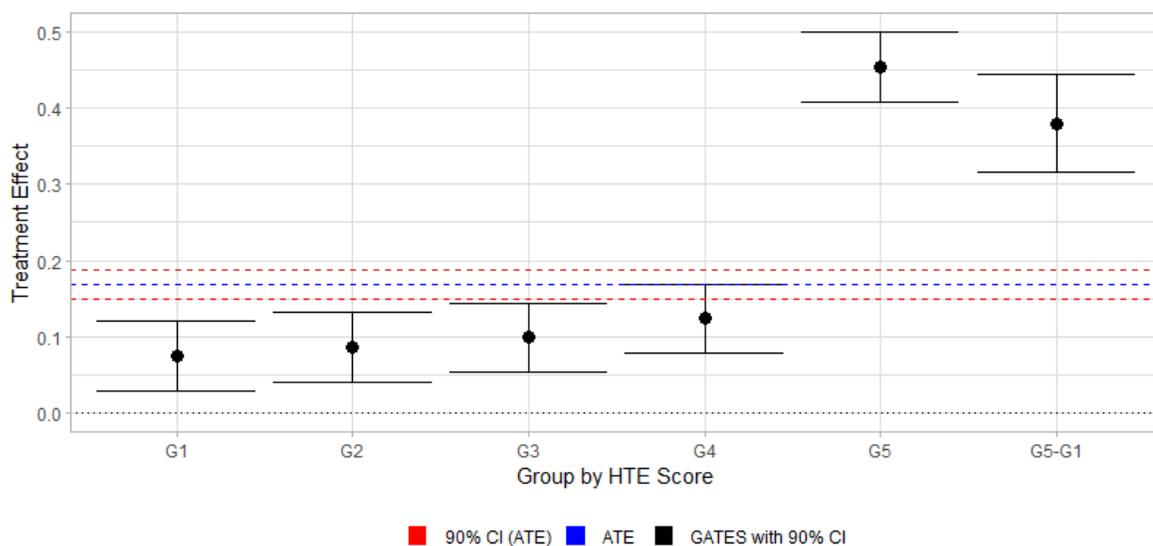
Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

La Tabla 3 presenta los siguientes resultados: el Efecto Promedio de Tratamiento (ATE en inglés) y el estimador que cuantifica qué tan bien el predictor de proxy se aproxima a la heterogeneidad del tratamiento. En primer lugar, el coeficiente del efecto promedio de tratamiento muestra un valor del 16 por ciento, lo cual es muy similar al porcentaje de adopción de las microfinanzas en el caso de Marruecos. Esto sugiere que este análisis se asemeja al modelo principal de análisis de la literatura en el caso de Marruecos. Por otro lado, el coeficiente α_2 es distinto de cero, con un valor de 1.24 y una significancia estadística alta, representada por un valor de $p < 0.001$. Este resultado indica que las predicciones logran capturar la heterogeneidad, lo que implica que existen diferencias sustanciales entre el grupo que optó por el apoyo financiero y el grupo que no lo hizo. Además, el nivel de significancia respalda que el efecto no se debe a una variación aleatoria. Por lo tanto, se puede afirmar con confianza que la cifra observada refleja una relación genuina entre los datos analizados, respaldada por su alta significancia estadística.

6.3 Resultados de los Efectos Promedio de Tratamiento por Grupos Ordenados (GATES)

Después de confirmar la heterogeneidad en el efecto del tratamiento, el trabajo procede a analizar el resultado de la segunda herramienta de Chernozhukov: GATES. Esto contribuye a examinar la diferencia en el efecto entre el grupo más afectado y el menos afectado por el tratamiento. Como se definió previamente en la sección de metodología, las observaciones se dividen en 5 grupos, lo que nos permite observar las diferencias en los efectos del tratamiento entre los grupos con los efectos más altos y bajos, es decir, G5 y G1, respectivamente. La siguiente figura representa la diferencia en los efectos entre los grupos ordenados:

Figura 1: Microfinanzas - GATES



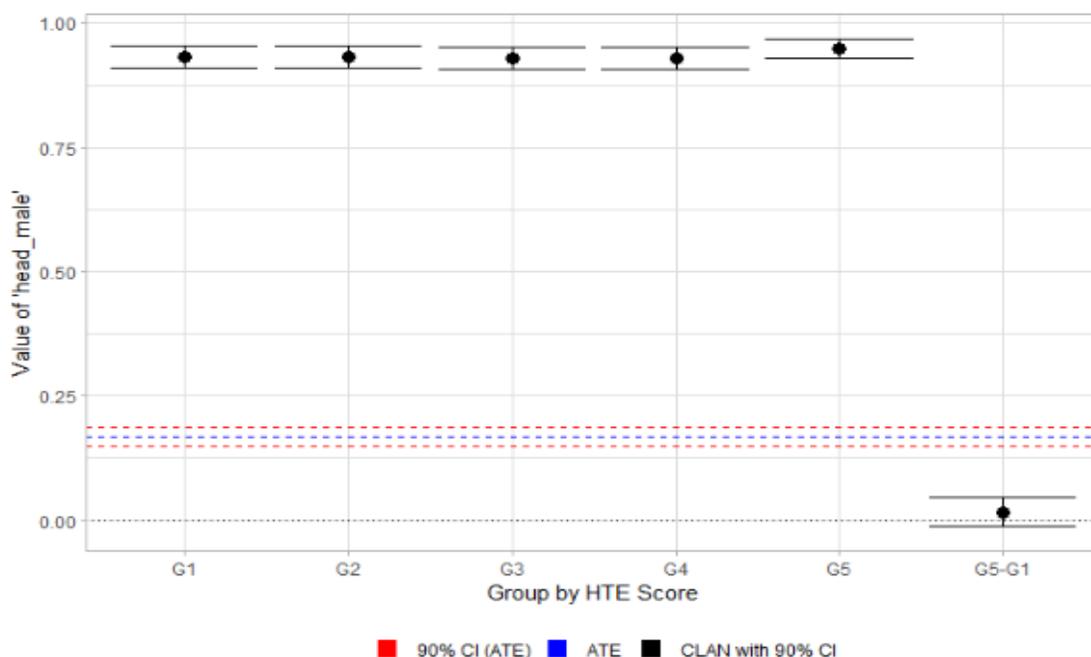
Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

La Figura 1 indica claramente que existen diferencias en los efectos entre los grupos. Se puede observar que el efecto de G5-G1 revela que el 20 por ciento de los individuos más afectados en el experimento tuvieron, en promedio, un 35 por ciento más de probabilidad de optar por las microfinanzas en comparación con el 20 por ciento de los individuos menos afectados. Esta diferencia es significativa, ya que es sustancialmente mayor que el efecto observado en los primeros cuatro grupos. Esto sugiere que ciertas variables tienen un fuerte impacto en la decisión de aceptar los servicios de microfinanzas, lo que resulta en una diferencia palpable en el efecto entre los diferentes grupos.

6.4 Resultados del Análisis de Clasificación (CLAN)

Por último, la herramienta de CLAN se utiliza para evaluar qué variables están más correlacionadas con la decisión de aceptar las microfinanzas. Incluso con el conjunto de variables seleccionadas que confirman la heterogeneidad, es posible que algunas variables influyan más que otras en el porcentaje de adopción, mientras que otras tengan un impacto menor. Por ejemplo, la siguiente figura muestra que la variable "cabeza masculina de hogar" tiene un impacto relativamente bajo en la decisión:

Figura 2: Microfinanzas – CLAN con la variable “cabeza masculina de hogar”

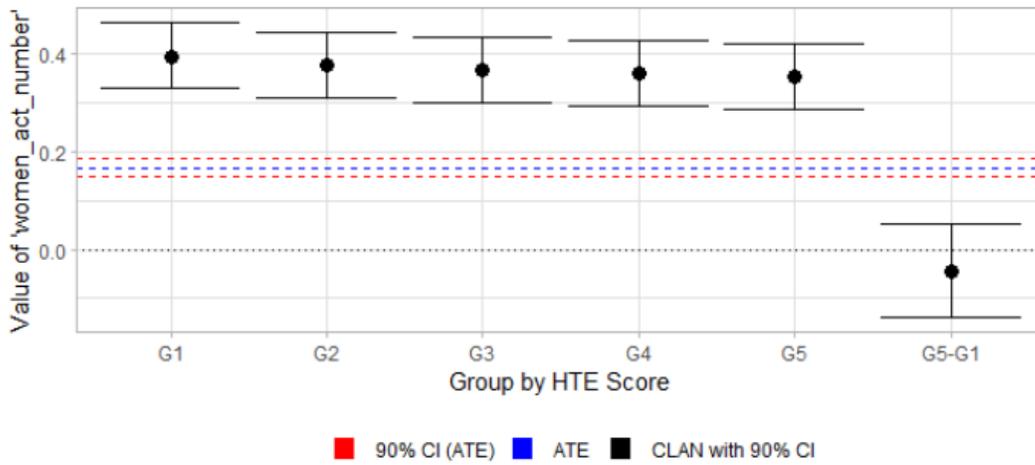


Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

Esta gráfica ilustra que la variable "cabeza masculina de hogar" no parece provocar una diferencia significativa en el efecto del tratamiento en los cinco grupos ordenados. Como se puede observar en la última barra, la diferencia en su efecto entre el grupo G1 y el grupo G5 es prácticamente nula. Esto sugiere que tener un hombre como cabeza masculina del hogar es una variable fuertemente correlacionada con el porcentaje de adopción de los servicios de microfinanzas. Este análisis confirma que el consumo de productos no durables, el autoempleo en ganadería, el número de actividades realizadas por mujeres y el riesgo de pérdida de más del 50% de la cosecha tampoco tuvieron un efecto significativo en la variabilidad entre los grupos ordenados.

Otra variable que muestra una tendencia similar entre los grupos ordenados es “actividades por mujer”. Esta variable mide las actividades organizadas por una mujer en un hogar. La siguiente figura prueba que la elección de optar por las microfinanzas no depende de que si las actividades están organizadas por mujer.

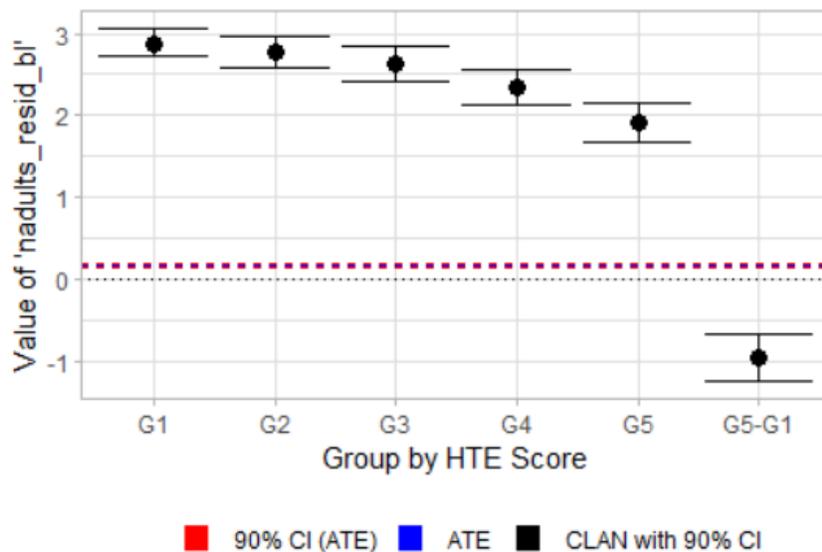
Figura 3: Microfinanzas – CLAN con la variable “actividades por mujer”



Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

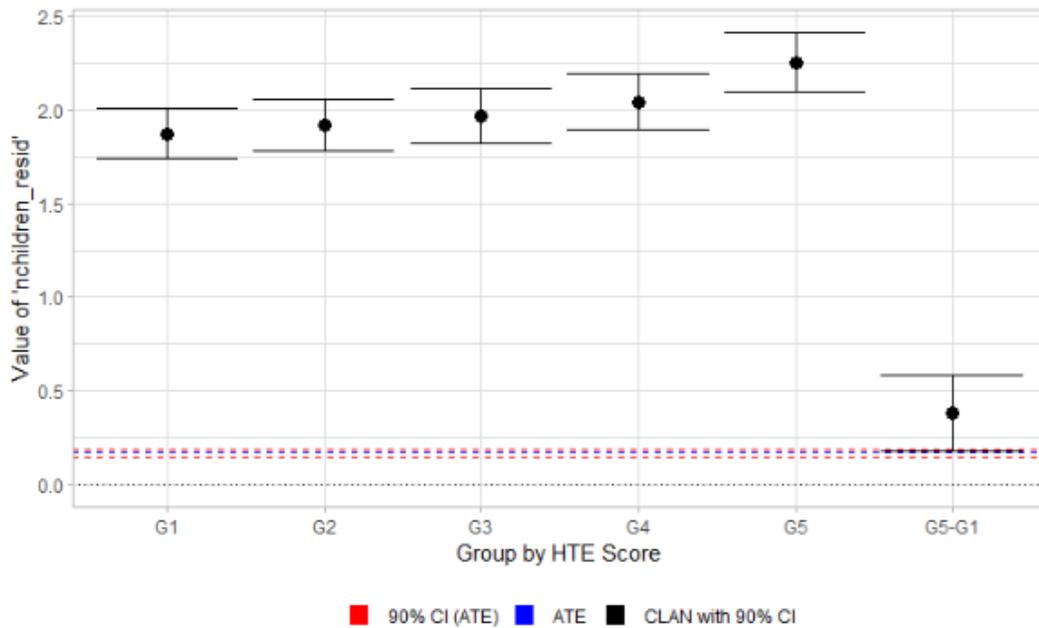
Esto señala que la heterogeneidad no está necesariamente correlacionada con el género de la dinámica de un hogar. A continuación, el estudio revisa otras variables que explican la información demográfica de un hogar como “número de niños” o “número de adultos” de un hogar como las siguientes figuras:

Figura 4: Microfinanzas –CLAN con la variable “número de adultos”



Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

Figura 5: Microfinanzas – CLAN con la variable “número de niños”

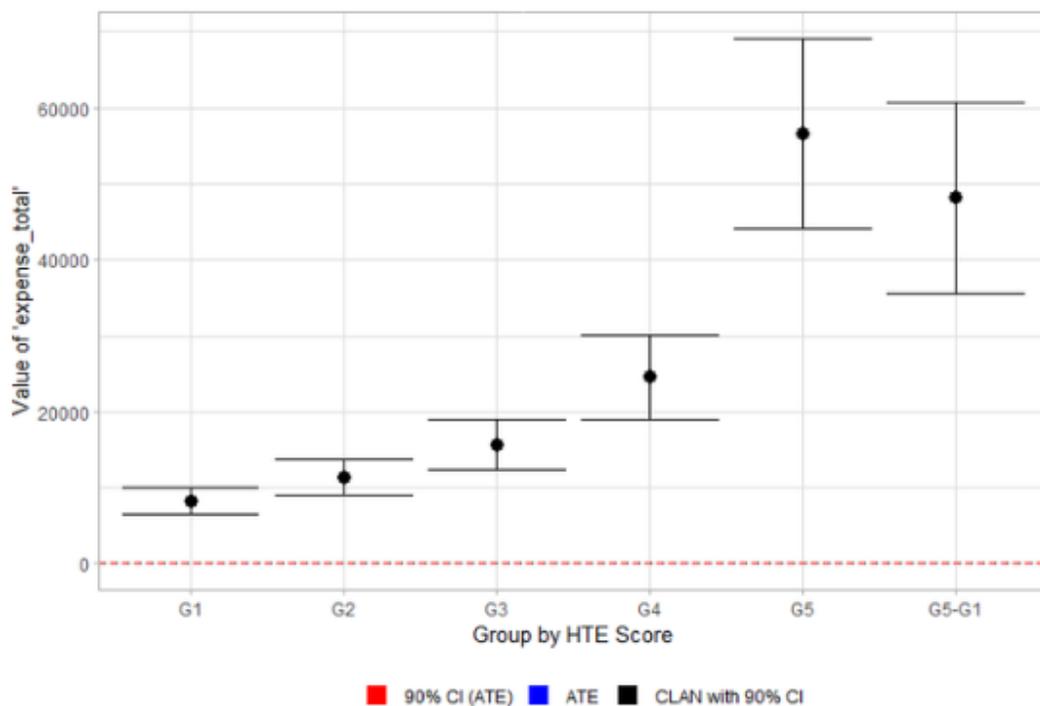


Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

El resultado de análisis de clasificación de estas variables indica que las características demográficas de un hogar no necesariamente provocan una decisión de optar por las microfinanzas.

Por otro lado, existen variables que caracterizan a los individuos en el experimento y muestran una notoria heterogeneidad entre los grupos ordenados, como se aprecia en la variable de "gastos totales" en la siguiente gráfica:

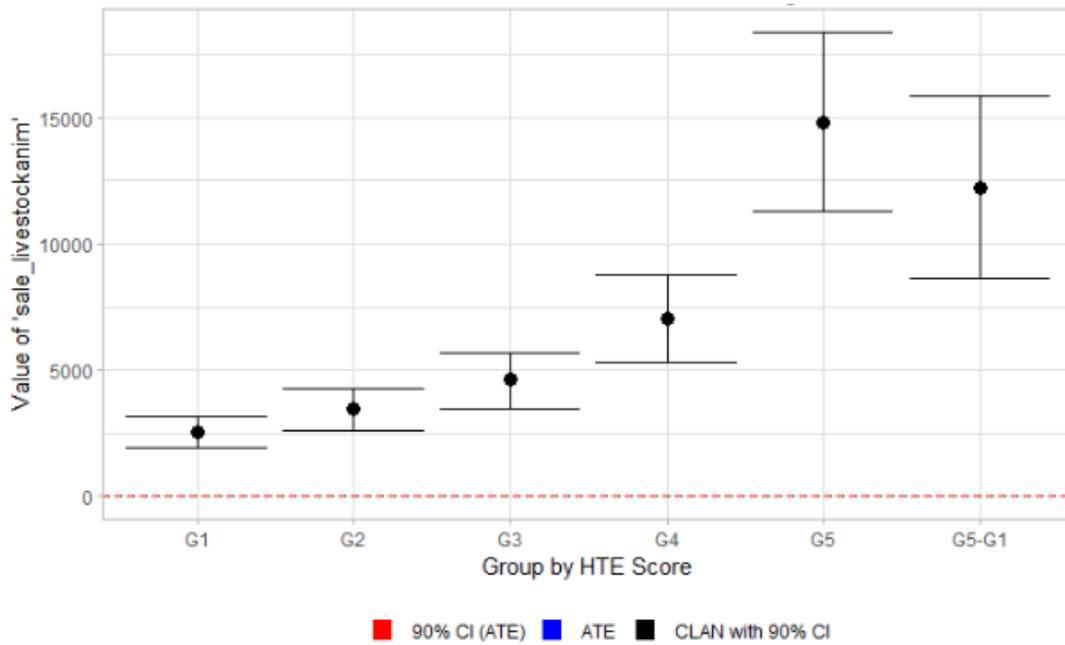
Figura 6: Microfinanzas – CLAN con la variable “gastos totales”



Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

Esta figura indica que los individuos con gastos más altos presentaron un mayor porcentaje de aceptación de las microfinanzas, mientras que aquellos con gastos más bajos, como el grupo G1, mostraron una notoria diferencia en un menor porcentaje de adopción, es decir, optaron por no utilizar los servicios de microfinanzas. Además de la variable anterior, vale prestar atención al patrón de las variables que explican la heterogeneidad. La siguiente figura muestra el resultado de la variable “ventas en ganadería”.

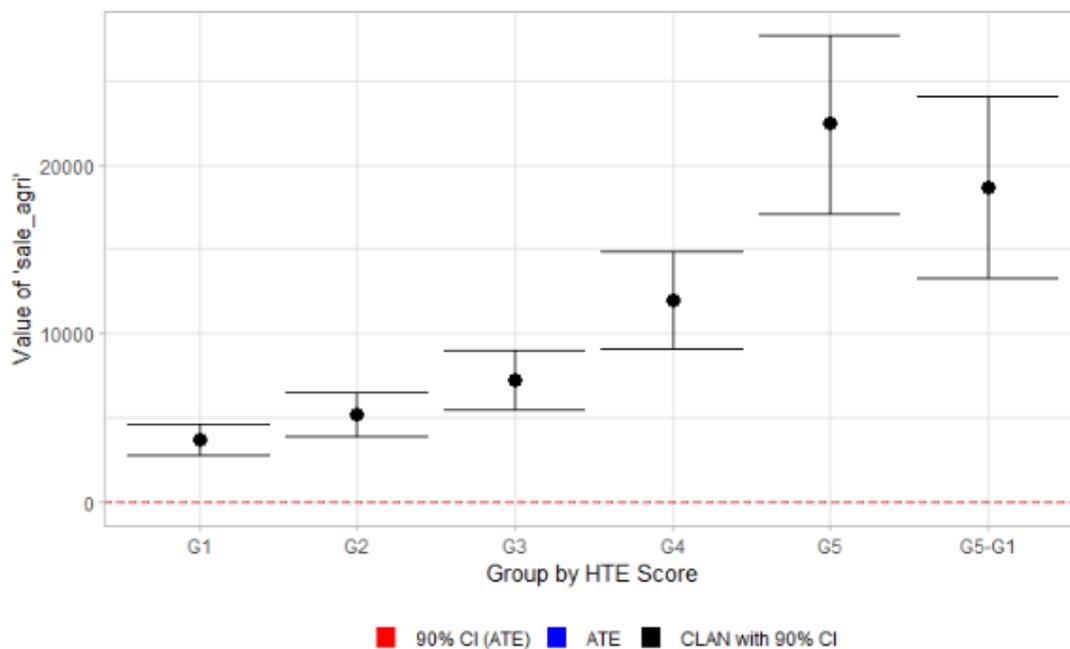
Figura 7: Microfinanzas – CLAN con la variable “ventas en ganadería”



Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

Como se puede observar, existe una heterogeneidad notable entre los cinco grupos ordenados. Otra variable que demuestra la heterogeneidad es “ventas en agricultura”.

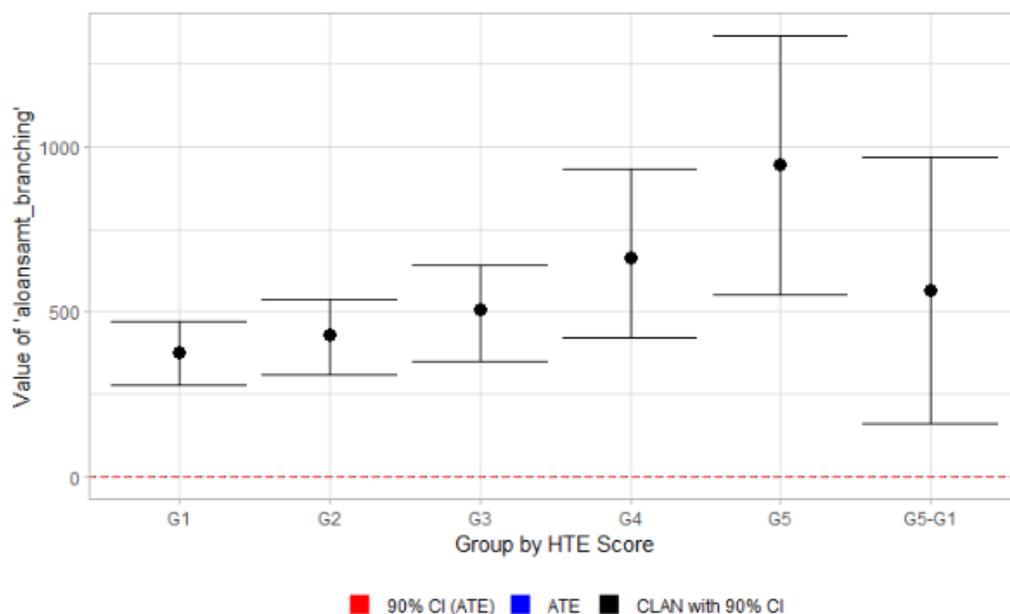
Figura 8: Microfinanzas – CLAN con la variable “ventas en agricultura”



Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

Estas variables señalan que lo que define la heterogeneidad tuvo que ver con las características del patrón de gastos y negocio de un hogar. Otra variable que demuestra una diferencia notable entre los grupos ordenados fue “monto de préstamo de las sucursales” como la siguiente figura:

Figura 9: Microfinanzas – CLAN con la variable “monto de préstamo de las sucursales”



Fuente: elaboración propia a partir de los resultados obtenidos usando el paquete GenericML en R (Chernozhukov et al., 2019).

Como Crepón expuso en su estudio en Marruecos que la posibilidad de adoptar las microfinanzas dependía de que si un individuo había tomado las microfinanzas previamente. Este resultado apunta que no solamente es el hecho de que haya tomado el servicio microfinanciero, sino el monto de préstamo que está también correlacionado con la heterogeneidad entre los grupos ordenados.

7. Conclusión

Este estudio basado en el experimento en el programa de microfinanzas en Marruecos busca analizar la heterogeneidad entre los grupos que toman las microfinanzas y las que no. Mientras que la adopción de las microfinanzas tuvo un impacto positivo a los individuos, solamente el 13 por ciento de los individuos en el grupo de tratamiento optaron por tomar el servicio microfinanciero. Tras la revisión de la literatura y su influencia en las intervenciones, queda claro que, para lograr un impacto significativo, es necesario aumentar la tasa de adopción. Aunque existen múltiples razones por las que un individuo puede optar por no solicitar apoyo financiero, es importante resaltar los resultados obtenidos en el análisis de esta tesis. Por ende, el enfoque de análisis fue que si existen la diferencia entre los individuos que toman y los que no toman. En el caso de encontrar la heterogeneidad, este trabajo continúa revisando que si la heterogeneidad es notable al ordenar las observaciones por los cinco grupos. Por último, el trabajo se enfoca en detectar qué variables, en específico, están correlacionadas en demostrar la heterogeneidad.

Para este análisis, el trabajo toma la base de datos del estudio de Crepón y parte en dos grupos para entregar al algoritmo con el modelo de aprendizaje automático bosque aleatorio. Al entregar, el trabajo utiliza los tres métodos presentados por Chernozhukov: BLP, GATES y CLAN para observar las evidencias de la heterogeneidad. En primer lugar, se observan diferencias en las características individuales entre el grupo que adopta las microfinanzas y las que no. En segundo lugar, esta diversidad genera un efecto notable cuando se ordenan las observaciones por grupos. Esto respalda la noción de que la heterogeneidad no es un incidente en el contexto del estudio en Marruecos.

Por último, es relevante analizar cuáles variables presentan una mayor correlación con la heterogeneidad. Se descubrió que variables como “el género del jefe de hogar” o “el número de actividades realizadas por las mujeres” no generaron heterogeneidad entre los grupos. Además, “el número de niños o adultos” de un hogar no provocaron la heterogeneidad entre los grupos ordenados. Esto indica que las características demográficas de un hogar no son factores que explican la heterogeneidad. En cambio, “gastos totales”, “las ventas en agricultura o ganadería” y “el monto de préstamo de las sucursales” mostraron una mayor correlación con la heterogeneidad entre los grupos. Esto sugiere que los individuos

con mayor gastos o ventas en su negocio previo a la selección del programa tienen mayor probabilidad de optar por las microfinanzas. Además, un individuo con la experiencia previa de tomar un préstamo y su monto pueden definir su decisión de tomar el servicio microfinanciero. En resumen, la clave no está en la información demográfica, sino en las variables que explican la dinámica de negocio o experiencia previa de toma de préstamo.

Es importante destacar que cada experimento en el ámbito de las microfinanzas involucra factores y contextos específicos. En este caso, el programa se implementó en zonas rurales, lo que hizo que el consumo en la producción agrícola y ganadera fuese de particular importancia. Para diseñar políticas efectivas que incrementen la tasa de adopción, es esencial realizar un análisis del contexto y las variables que caracterizan las observaciones. Además, los casos latinoamericanos en el campo de las microfinanzas sugieren que la capacitación, tanto financiera como no financiera, puede desempeñar un papel crucial en aumentar el impacto de dichas microfinanzas. En resumen, sería valioso replicar un estudio similar sobre la heterogeneidad en el contexto de las microfinanzas en América Latina para determinar si existen diferencias entre los grupos y, en caso afirmativo, identificar los factores que influyen en la adopción de los servicios financieros.

8. Bibliografías

- Andía, G. A. (2013). Microfinanzas y crecimiento regional en el Perú. *Economía*, 36(72), 143-173.
- Augsburg, B., De Haas, R., Harmgart, H., & Meghir, C. (2015). The impacts of microcredit: Evidence from Bosnia and Herzegovina. *American Economic Journal: Applied Economics*, 7(1), 183-203.
- Baiardi, A., & Naghi, A. A. (2021). The value added of machine learning to causal inference: Evidence from revisited studies. arXiv preprint arXiv:2101.00878.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Banerjee, A., Duflo, E., Glennerster, R., & Kinnan, C. (2015). The miracle of microfinance? Evidence from a randomized evaluation. *American economic journal: Applied economics*, 7(1), 22-53.
- Chernozhukov, V., Demirer, M., Duflo, E., & Fernandez-Val, I. (2018). *Generic machine learning inference on heterogeneous treatment effects in randomized experiments, with an application to immunization in India* (No. w24678). National Bureau of Economic Research.
- Crépon, B., Devoto, F., Duflo, E., & Parienté, W. (2015). Estimating the impact of microcredit on those who take it up: Evidence from a randomized experiment in Morocco. *American Economic Journal: Applied Economics*, 7(1), 123-150.
- García, A., Cecchi, F., Eriksen, S., & Lensink, R. (2022). The plus in credit-plus-technical assistance: evidence from a rural microcredit programme in Bolivia. *The Journal of Development Studies*, 58(2), 275-291.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media.
- Gobierno de México. (2020). ¿Qué es una AFORE? <https://www.gob.mx/pensionisste/articulos/que-es-una-afore-245550?idiom=es>
- Higgins, P. (2013). From sustainable development to carbon control: urban transformation in Hong Kong and London. *Journal of cleaner production*, 50, 56-67.

- INEGI. (2023, May 22). Encuesta Nacional de Inclusión Financiera 2021. <https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2022/enif/ENIF21.pdf>
- Kitzmüller, L. (2021, October 14). Not merely averages: Using machine learning to estimate heterogeneous treatment effects. *Medium*. <https://towardsdatascience.com/not-merely-averages-using-machine-learning-to-estimate-heterogeneous-treatment-effects-573bf7376a73>
- Meager, R. (2022). Aggregating Distributional Treatment Effects: A Bayesian Hierarchical Analysis of the Microcredit Literature. *American Economic Review*, 112(6), 1818-1847. DOI: 10.1257/aer.20181811.
- Müller, A. C., & Guido, S. (2017). *Introduction to Machine Learning with Python*. O'Reilly Media.
- World Bank. (2013). *Global financial development report 2014: Financial inclusion* (Número de informe 82556, Vol. 1). <https://documents1.worldbank.org/curated/en/225251468330270218/pdf/Global-financial-development-report-2014-financial-inclusion.pdf>
- World Bank. (2022). Financial Inclusion. <https://www.worldbank.org/en/topic/financialinclusion/overview#1>