UNIVERSIDAD EAFIT

DISCRIMINACIÓN ÉTNICA EN PRESTAMOS HIPOTECARIOS EN ESTADOS UNIDOS: UN ANÁLISIS PREDICTIVO CON MÉTODOS CAUSALES Y DE APRENDIZAJE AUTOMÁTICO

JUAN PABLO GALEANO NARANJO

Tesis

Modalidad Profundización

Asesoras

Paula María Almonacid Hurtado, Ph.D.

Pilar Beatriz Álvarez Franco, Ph.D.

Vivian Cruz Castañeda, Ph.D.

UNIVERSIDAD EAFIT

ESCUELA DE CIENCIAS APLICADAS E INGENIERÍA

MAESTRÍA EN CIENCIAS DE LOS DATOS Y ANALÍTICA

MEDELLÍN

2025

Agradecimiento

Agradezco profundamente a mis profesoras y asesoras de tesis por su guía, exigencia y generosidad intelectual durante todo este proceso. A mi padre, por su apoyo incondicional y ejemplo de perseverancia. Y a mi compañero de maestría, Hoover Arbeláez, por su amistad y el apoyo mutuo que hizo más llevadero cada reto del camino.

CONTENIDO

LISTA DE TABLAS

LISTA DE FIGURAS

Lista de ecuaciones

Términos y Abreviaciones

A continuación se presenta la lista de símbolos, variables y abreviaciones utilizados a lo largo del documento, con el fin de facilitar su comprensión. También se aclaran ciertas convenciones terminológicas empleadas en el análisis.

- **BNH** — Sigla utilizada para referirse a individuos **Blancos No Hispanos**. En el documento, los términos "blanco", "no hispano" y "BNH" se utilizan de forma intercambiable.
- **Latino / Hispano** — Hace referencia a individuos que se identifican como blancos hispanos o latinos. En este trabajo se utilizan como sinónimos.
- **PSM** — *Propensity Score Matching*: Técnica estadística para emparejar individuos con características similares en variables observadas, a fin de estimar efectos causales reduciendo sesgos de selección.
- **ASMD** — *Absolute Standardized Mean Difference*: Diferencia estandarizada absoluta entre medias o proporciones de dos grupos (tratamiento y control) para una determinada variable. Es usada como métrica para evaluar el balance luego del emparejamiento.
- **SMD** — *Standardized Mean Difference*: Similar a ASMD, se utiliza para variables continuas. En este documento, los valores de SMD para variables categóricas también son llamados ASMD por convención.
- **Odds Ratio** — Medida estadística que expresa la razón entre la probabilidad de ocurrencia de un evento frente a la probabilidad de que no ocurra. En este estudio se usa para comparar la probabilidad de aprobación entre latinos y blancos no hispanos.
- **SHAP** — *SHapley Additive exPlanations*: Método de interpretación basado en teoría de juegos para explicar el impacto de cada variable en la predicción de un modelo de machine learning.
- **Caliper** —En el contexto de técnicas de emparejamiento para inferencia causal, el **caliper** se refiere a un umbral máximo de distancia permitido entre unidades tratadas y de control

para que puedan ser emparejadas. Específicamente, en métodos como el *propensity score matching*, un caliper restringe el emparejamiento a aquellas observaciones cuya diferencia en el puntaje de propensión es menor o igual a dicho umbral. En este estudio, el caliper controla la **proximidad entre pares**, limitando los sesgos por emparejamientos con unidades poco similares. Un caliper demasiado pequeño puede descartar muchas unidades, mientras que uno muy grande puede introducir emparejamientos imprecisos.

Variables del Dataset HMDA 2023

A continuación se presenta una descripción en español de las variables utilizadas del dataset HMDA 2023, según el diccionario provisto:

- **income** — Ingreso anual del solicitante (en miles de dólares).
- **loan_amount** — Monto del préstamo hipotecario solicitado.
- **tract_minority_population_percent** — Porcentaje de población minoritaria en el área censal (*census tract*) del solicitante.
- **tract_to_msa_income_percentage** — Porcentaje del ingreso promedio del tracto en relación con el ingreso medio del área metropolitana.
- **loan_to_value_ratio** — Relación entre el monto del préstamo y el valor de la propiedad.
- **total_loan_costs** — Costos totales del préstamo (intereses, cargos, etc.).
- **origination_charges** — Cargos cobrados por originar el préstamo.
- **loan_term** — Plazo del préstamo en meses.
- **prepayment_penalty_term** — Duración (en meses) de penalización por pago anticipado.
- **tract_population** — Población total del tracto censal.

- **tract_owner_occupied_units** — Unidades de vivienda ocupadas por sus propietarios en el tracto.
- **interest_rate** — Tasa de interés anual del préstamo.
- **rate_spread** — Diferencia entre la tasa de interés del préstamo y una tasa de referencia comparable.
- **debt_to_income_ratio** — Relación entre la deuda del solicitante y su ingreso bruto.
- **applicant_age** — Rango de edad del solicitante (ej. 35–44).
- **loan_purpose** — Propósito del préstamo (ej. compra de vivienda).
- **applicant_credit_score_type** — Tipo de modelo de puntuación crediticia usado para el solicitante.
- **co_applicant_credit_score_type** — Tipo de puntuación crediticia usada para el co-solicitante.
- **occupancy_type** — Tipo de ocupación de la vivienda (residencia principal, secundaria, inversión).
- **state_code** — Código del estado donde se origina la solicitud.

Convenciones Terminológicas

- En este documento, cuando se habla de "rechazos discriminatorios" se hace referencia a los casos en los cuales un solicitante latino fue rechazado mientras su contraparte blanca no hispana —emparejada mediante PSM con características similares— fue aprobada.
- Las comparaciones estadísticas entre grupos se evalúan utilizando ASMD (para evaluar balance) y *odds ratios* (para evaluar diferencia en resultados).
- El año 2023 se utilizó exclusivamente como marco temporal para evitar sesgos derivados de cambios estructurales o regulatorios entre años.

Resumen

En Estados Unidos, el proceso de solicitud y aprobación de préstamos hipotecarios implica la evaluación de variables como ingresos, historial crediticio, nivel de endeudamiento y características de la propiedad, ya sea mediante revisión manual o sistemas automatizados. Aunque este procedimiento busca objetividad, puede estar influenciado por sesgos estructurales, especialmente de tipo étnico. Esta tesis examina el fenómeno de la discriminación en el acceso al crédito hipotecario mediante una estrategia metodológica que combina técnicas de evaluación causal y aprendizaje automático explicable. A partir del caso de los blancos latinos en Estados Unidos —un grupo cuya ambigua categorización étnica lo convierte en un caso clave—, se observa cómo decisiones crediticias aparentemente neutras pueden reproducir patrones históricos de exclusión al incorporar variables geográficas o demográficas estrechamente correlacionadas con la etnicidad.

La principal contribución metodológica consiste en integrar el emparejamiento por puntaje de propensión (PSM) con modelos de clasificación supervisada (Gradient Boosting, Random Forest), auditados mediante SHAP (SHapley Additive Explanations). Esta combinación permite detectar disparidades en la aprobación de créditos no atribuibles a variables observables, revelando formas indirectas de discriminación algorítmica. Aunque los modelos no presentan una alta capacidad predictiva en términos absolutos, los patrones identificados son consistentes y estadísticamente significativos. A nivel empírico, se encuentra que los solicitantes latinos enfrentan tasas de aprobación más bajas, incluso con perfiles crediticios similares a los de sus contrapartes blancos no hispanos. Variables como la proporción de minorías en el vecindario y el estado de residencia influyen de manera considerable en las predicciones, lo que sugiere la presencia de sesgos geoespaciales en los sistemas de originación de crédito. Finalmente, esta investigación ofrece un marco replicable para auditar decisiones algorítmicas y promover una evaluación crediticia más justa. Sus hallazgos son aplicables a contextos como el colombiano, donde persisten brechas de acceso para mujeres, migrantes y poblaciones étnicas. Así, la tesis contribuye tanto al debate académico como al diseño de políticas públicas y modelos institucionales orientados a la inclusión financiera con enfoque diferencial.

Palabras clave: Discriminación, PSM, Etnia, Crédito Hipotecario, Aprendizaje Automático, Latinos, HMDA

0. INTRODUCCIÓN

El acceso equitativo al crédito constituye un pilar fundamental para la inclusión financiera y el desarrollo económico sostenible. Entre las distintas modalidades crediticias, el crédito hipotecario representa una de las decisiones más trascendentales para los hogares, no solo por su magnitud económica, sino por su rol en la estabilidad patrimonial y la movilidad social. En países como Estados Unidos, donde la propiedad de vivienda está estrechamente vinculada con la acumulación de riqueza, la equidad en el acceso a estos créditos es esencial para contrarrestar desigualdades históricas y promover oportunidades reales. En Estados Unidos, el proceso de asignación de préstamos hipotecarios se basa en un sistema mixto que combina criterios financieros tradicionales con modelos de evaluación automatizada. Las entidades crediticias utilizan información proporcionada por los solicitantes —como ingresos, historial laboral, nivel de endeudamiento y puntaje crediticio— para estimar su capacidad de pago y riesgo asociado. Paralelamente, se incorporan algoritmos que predicen la probabilidad de incumplimiento a partir de bases de datos históricas, los cuales pueden incluir variables como la ubicación de la vivienda, el tipo de vecindario y patrones de comportamiento financiero previos.

Si bien estos mecanismos buscan objetividad y eficiencia, su aplicación no está exenta de sesgos, especialmente cuando los algoritmos replican patrones históricos de discriminación presentes en los datos de entrenamiento, estudios recientes han evidenciado que persisten disparidades significativas en las decisiones de aprobación hipotecaria ([1], [2]), aun cuando se controla rigurosamente por factores financieros relevantes como el ingreso, el historial crediticio y el nivel de endeudamiento, lo que sugiere la presencia de sesgos no atribuibles a la solvencia del solicitante.

Las disparidades en el acceso al crédito hipotecario entre diferentes grupos étnicos en Estados Unidos han sido documentadas de forma consistente en las últimas décadas. [1] señalan que las diferencias en las tasas de aprobación de hipotecas entre blancos y minorías apenas han disminuido desde 1970, y que los costos del crédito (como tasas de interés) siguen siendo más altos para prestatarios latinos y afroamericanos, incluso al controlar por ingresos, puntuación crediticia y características del préstamo, persisten brechas inexplicadas. [2] muestran que los solicitantes latinos

pagan tasas de interés más altas que sus contrapartes blancos con perfiles crediticios equivalentes. Este sobreprecio representa cientos de millones de dólares anuales en pagos adicionales.

La heterogeneidad dentro del grupo latino también es significativa. [3] muestran que los latinos que se identifican como blancos tienen mejores resultados en acceso y condiciones de vivienda que aquellos que se identifican como afro-latinos o indígenas. Este fenómeno se vincula con dinámicas de racialización internas que afectan el trato recibido por diferentes subgrupos, asimismo, [4] plantean que el estatus migratorio percibido influye en la interacción con instituciones financieras. En particular, los latinos percibidos como inmigrantes recientes enfrentan barreras adicionales por estereotipos sobre legalidad, estabilidad laboral o dominio del idioma.Finalmente, [5] documentan que la segregación residencial ha sido clave para entender la exposición desproporcionada de las comunidades latinas a hipotecas de alto riesgo, lo cual aumentó su vulnerabilidad durante la crisis subprime.

Estas desigualdades afectan de manera desproporcionada a ciertos grupos étnicos y raciales, entre ellos los blancos latinos, quienes, aunque son categorizados racialmente como blancos, pueden ser objeto de un trato diferencial por su identificación étnica como latinos ([3]). Este subgrupo ofrece un caso singular para estudiar cómo los sesgos estructurales pueden operar de forma sutil pero sistemática en sistemas de decisión crediticia, especialmente cuando se utilizan modelos automatizados.

En este contexto, esta tesis propone un enfoque metodológico innovador que combina técnicas clásicas de evaluación causal, como el emparejamiento por puntaje de propensión (Propensity Score Matching, PSM), con modelos de aprendizaje automático supervisado e interpretabilidad algorítmica, con el fin de identificar y caracterizar rechazos crediticios potencialmente discriminatorios. En particular, se implementan clasificadores como *Gradient Boosting* y *Random Forest*, entrenados sobre datos nacionales submuestreados, para predecir la probabilidad de rechazo crediticio y evaluar el rol de variables sensibles mediante técnicas como SHAP (SHapley Additive Explanations).

El uso de estas herramientas de interpretación permite auditar de manera transparente el fun-

cionamiento de modelos complejos, revelando cómo características aparentemente neutras —como la localización geográfica o la composición demográfica del vecindario— pueden actuar como proxies de variables étnicas, reproduciendo así patrones de exclusión histórica en entornos de decisión automatizada. Esta aproximación no solo permite detectar sesgos indirectos, sino que ofrece una base replicable para construir marcos de auditoría algorítmica aplicables en otros contextos financieros.

Esta tesis se estructura en ocho secciones. La sección 1 plantea el problema objeto de estudio, en la sección 2 y 3 se encuentran la justificación y los objetivos respectivamente, la sección 4 está dedicada a todo el marco teórico en el cual se abordan investigaciones previas relevantes, la sección 5 contiene la metodología utilizada, la sección 6 contiene información acerca de los métodos utilizados, en la sección 7 se encuentran documentados los resultados obtenidos y finalmente en la sección 8 se abordan las conclusiones del presente estudio, así como recomendaciones para futuras investigaciones del tema.

1. PLANTEAMIENTO DEL PROBLEMA

El acceso al crédito hipotecario en los Estados Unidos ha sido históricamente una vía crítica para la acumulación de capital y el ascenso socioeconómico de los hogares [5]. No obstante, una extensa literatura ha documentado disparidades sistemáticas en las decisiones de aprobación y condiciones del crédito según la raza o etnia del solicitante [1]. Incluso después de controlar por ingresos, puntuación crediticia y características del préstamo, estas diferencias persisten de forma significativa [2].

Particularmente, la población latina enfrenta barreras estructurales en el acceso a servicios financieros, muchas veces asociadas a estigmas sobre migración, estabilidad laboral o idioma [4]. Sin embargo, dentro del grupo latino existe una marcada heterogeneidad: los individuos que se identifican como blancos latinos pueden presentar una posición intermedia entre blancos no hispanos y otras minorías en términos de acceso y condiciones crediticias [3]. El caso de la población denominada como blancos hispanos resulta especialmente interesante: a pesar de estar categorizados racialmente como blancos, su autoidentificación étnica como latinos podría influir en la percepción institucional y en las decisiones de aprobación crediticia.

El interrogante central que guía esta investigación es: ¿existe evidencia de discriminación en las decisiones de otorgamiento de crédito hipotecario hacia personas identificadas como blancos latinos en Estados Unidos, aun cuando presentan perfiles financieros equivalentes a los de blancos no hispanos? Para responder esta pregunta, se utilizará una combinación metodológica de emparejamiento por puntaje de propensión (PSM) [6], [7] y técnicas de aprendizaje automático orientadas a identificar patrones de decisión no observables directamente en los datos estructurados [8], [9].

La disponibilidad de bases de datos como la del Home Mortgage Disclosure Act (HMDA) brinda una oportunidad única para estudiar este fenómeno a escala nacional, con suficiente granularidad y diversidad geográfica para realizar comparaciones rigurosas. Los hallazgos podrían contribuir no solo a la literatura académica sobre discriminación financiera, sino también a los marcos regulatorios que buscan garantizar prácticas equitativas en el sistema crediticio estadounidense [7].

2. JUSTIFICACIÓN

Este trabajo se justifica por la necesidad de profundizar en la comprensión de formas sutiles de discriminación que pueden surgir incluso cuando las decisiones de crédito se basan en datos aparentemente neutros. Como han advertido [10] y [11], los modelos estadísticos y algorítmicos pueden reproducir o incluso amplificar sesgos históricos si las variables utilizadas están correlacionadas con categorías protegidas, como la etnia.

El uso combinado de técnicas estadísticas como el PSM y herramientas de auditoría algorítmica (como SHAP) permite detectar diferencias en el tratamiento de grupos demográficos que no se explican por características financieras observadas [9], [12]. Esta aproximación permite aproximarse a los estándares del derecho antidiscriminación, especialmente el criterio de impacto dispar, que no requiere demostrar intención sino impacto desigual [7].

Desde una perspectiva científica, este estudio contribuye a la literatura emergente sobre justicia algorítmica y ética en ciencia de datos, al aplicar metodologías de evaluación crítica sobre modelos en contextos reales y con implicaciones regulatorias. En términos prácticos, los resultados podrían ser útiles para agencias como la Consumer Financial Protection Bureau (CFPB), desarrolladores de modelos crediticios, y defensores de los derechos civiles interesados en reducir las brechas estructurales en el acceso al crédito.

Finalmente, el enfoque sobre los blancos latinos permite enriquecer el análisis de la discriminación étnica al evidenciar que las identidades sociales no son binarias ni homogéneas, y que los mecanismos de exclusión pueden operar incluso dentro de categorías racialmente privilegiadas. Este matiz resulta fundamental en un contexto de creciente diversidad y complejidad identitaria en Estados Unidos. Desde una perspectiva aplicada, el enfoque metodológico propuesto busca ofrecer una herramienta replicable para la auditoría algorítmica de decisiones crediticias, permitiendo identificar patrones de discriminación indirecta. Así mismo, provee insumos técnicos que pueden ser utilizados por entidades reguladoras y financieras en el diseño de procesos más justos, éticos y transparentes, particularmente en entornos mediados por inteligencia artificial.

3. OBJETIVOS

3.1. GENERAL

Predecir posibles patrones de discriminación étnica contra personas blancas latinas en el acceso al crédito hipotecario en los Estados Unidos mediante el uso de emparejamiento por puntaje de propensión (PSM) y modelos de aprendizaje automático aplicados a los datos del Home Mortgage Disclosure Act (HMDA).

3.2. ESPECÍFICOS

- Explorar y analizar las diferencias observadas en las tasas de aprobación y condiciones crediticias entre solicitantes blancos no hispanos y blancos hispanos, con base en variables socioeconómicas y características del préstamo.
- Aplicar una metodología de emparejamiento por puntaje de propensión (PSM) para comparar individuos con características similares entre ambos grupos y aislar el efecto de la variable de etnia.
- Entrenar modelos de aprendizaje automático que permitan identificar patrones predictivos asociados a rechazos injustificados de crédito, evaluando si es posible detectar indicios de discriminación algorítmica.
- Evaluar la presencia de variables proxy y su influencia en los modelos, así como el posible impacto de variables estructurales no observables.
- Determinar la viabilidad del uso de modelos auxiliares o auditorías algorítmicas externas para apoyar procesos regulatorios en la detección de discriminación financiera.

4. MARCO TEÓRICO

4.1. Proceso de aprobación de crédito hipotecario en Estados Unidos

El proceso de aprobación o rechazo de un crédito hipotecario en Estados Unidos sigue una lógica estructurada basada en la evaluación del riesgo crediticio del solicitante. Las instituciones financieras utilizan modelos internos o algoritmos automatizados para analizar una serie de variables que permiten estimar la probabilidad de incumplimiento y la capacidad de pago del solicitante.

Entre las variables más comúnmente consideradas se encuentran: el ingreso anual declarado, la relación deuda-ingreso (*Debt-to-Income ratio, DTI*), el puntaje de crédito (*credit score*), el historial laboral, la relación préstamo-valor (*Loan-to-Value ratio, LTV*), el tipo de propiedad solicitada, y el estatus de ocupación (si será vivienda principal o inversión). Además, se tienen en cuenta variables adicionales como el número de cuentas abiertas, historial de pagos, y eventos negativos como bancarrotas o moras recientes.

El proceso puede seguir un enfoque manual o automatizado. En el primero, un oficial de crédito revisa los documentos y decide con base en políticas internas; en el segundo, sistemas automatizados de decisión (*automated underwriting systems*, AUS) como el *Desktop Underwriter* de Fannie Mae o el *Loan Product Advisor* de Freddie Mac, emiten una aprobación o negación preliminar basada en reglas y umbrales preestablecidos. Este tipo de sistemas han sido adoptados ampliamente por su eficiencia, pero han generado preocupaciones sobre la transparencia y el sesgo algorítmico.

Cabe destacar que aunque la ley impone restricciones sobre el uso de variables sensibles como raza, etnia o género, muchas variables correlacionadas indirectamente (como la localización geográfica o el tipo de institución crediticia) pueden introducir sesgos si no se controlan adecuadamente [2].

4.2. Uso de la base de datos HMDA en el estudio de la discriminación crediticia

La HMDA (Home Mortgage Disclosure Act) es una ley federal promulgada en 1975 que obliga a la mayoría de las instituciones financieras que otorgan créditos hipotecarios a reportar anualmente información detallada sobre las solicitudes, aprobaciones y condiciones de dichos préstamos. Su propósito central es promover la transparencia y garantizar la equidad en el acceso al crédito, especialmente entre comunidades históricamente marginadas. La base de datos resultante —disponible a través del Bureau de Protección Financiera del Consumidor (CFPB, por sus siglas en inglés)— incluye campos detallados sobre características del solicitante (edad, sexo, raza, etnia, ingreso declarado), del préstamo (tipo de producto, monto, tasa de interés, resultado de la solicitud) y de la propiedad (ubicación geográfica, tipo de vivienda, ocupación prevista). La cobertura de la HMDA abarca la mayoría de instituciones reguladas a nivel federal, aunque ciertas cooperativas de ahorro y préstamo o prestamistas pequeños pueden estar exentas. Entre sus limitaciones más relevantes se encuentra la ausencia de información granular sobre historial crediticio o puntaje FICO individual, así como posibles inconsistencias en la categorización étnica cuando esta es auto-reportada. No obstante, la HMDA representa una de las fuentes más completas y accesibles para el estudio de patrones de aprobación crediticia en Estados Unidos.

La base de datos del Home Mortgage Disclosure Act (HMDA) ha sido una fuente fundamental para investigar patrones de discriminación en el otorgamiento de créditos hipotecarios en Estados Unidos. Antes de su expansión en 2018, los estudios se limitaban a analizar diferencias crudas entre grupos raciales, ya que la información sobre factores crediticios esenciales como el puntaje de crédito, la relación deuda-ingreso o la razón préstamo-valor no estaba disponible. Esta carencia dificultaba la capacidad de los investigadores para controlar por riesgo crediticio al evaluar disparidades en aprobaciones o tasas de interés [13], [14].

La incorporación de variables adicionales a partir de 2018 permitió un análisis mas robusto. Por ejemplo, en [15] utilizó los datos expandidos de HMDA para el año 2020 y demostró que, incluso después de controlar por puntaje de crédito, relación deuda-ingreso y razón préstamo-valor, persisten diferencias significativas en las tasas de aprobación y en las condiciones del préstamo entre prestatarios hispanos, afroamericanos y blancos no hispanos [15]. El estudio encuentra que los

solicitantes hispanos enfrentan tasas de denegación mas altas y pagan tasas de interés ligeramente superiores, así como mayores costos totales de los préstamos en comparación con prestatarios blancos con perfiles crediticios similares.

Estos hallazgos han sido respaldados por otras investigaciones que, al vincular datos HMDA con otras fuentes como Optimal Blue o mediante el uso de modelos de ecuaciones simultaneas, han revelado que la discriminación puede manifestarse tanto en la etapa de decisión de aprobación como en la fijación de precios del crédito [16], [17]. De hecho, en [16], [17] se muestra que los prestatarios afroamericanos e hispanos con perfiles crediticios similares a los de prestatarios blancos reciben sistemáticamente peores opciones de tasas y beneficios, especialmente en segmentos con menor puntaje de crédito.

En conjunto, esta literatura resalta la utilidad de la base de datos HMDA como herramienta para monitorear posibles inequidades en el mercado hipotecario. A pesar de sus limitaciones, como la ausencia de datos sobre verificación de ingresos o historial de empleo, los estudios coinciden en que las disparidades observadas no pueden explicarse completamente por factores crediticios observables, lo que refuerza la necesidad de escrutinio continuo en las practicas de otorgamiento de créditos.

4.3. Emparejamiento por puntaje de propensión (PSM) en estudios de discriminación

El emparejamiento por puntaje de propensión (PSM, por sus siglas en inglés) es una técnica estadística que permite estimar efectos causales en estudios observacionales, al equilibrar covariables relevantes entre grupos comparados. Fue formalizado por [6], y desde entonces ha sido ampliamente utilizado en investigación económica y social.

En contextos de discriminación financiera, el PSM permite comparar individuos de grupos demográficos distintos que tienen una probabilidad estadística similar de recibir un tratamiento, controlando por características observadas como ingreso, monto del préstamo, historial crediticio, entre otras. Por ejemplo, [2] y [18] aplican enfoques de emparejamiento para identificar disparidades

en las tasas de aprobación o en los términos del préstamo entre blancos y minorías, evidenciando diferencias que no se explican por variables financieras; para el presente trabajo la utilidad del PSM radica en que permite construir un subconjunto balanceado de datos en el cual se puedan comparar individuos de diferentes grupos étnicos con características estadísticamente similares. Si tras el emparejamiento persisten diferencias en las decisiones crediticias, ello sugiere evidencia de trato desigual que podría atribuirse a discriminación.

Además, [7] destacan cómo la metodología de PSM se alinea con el estándar legal de *disparate impact*, utilizado por reguladores como la CFPB y el Departamento de Justicia de EE.UU. para evaluar prácticas discriminatorias en servicios financieros. Esta convergencia entre método estadístico y criterio legal refuerza su relevancia en análisis de equidad en crédito.

4.4. Modelos de aprendizaje automático y sesgos algorítmicos

El uso de algoritmos de aprendizaje automático (*machine learning*, ML) en servicios financieros ha crecido rápidamente en la última década, impulsado por su capacidad para procesar grandes volúmenes de datos y capturar patrones complejos. Sin embargo, diversos estudios han advertido que estos modelos pueden reproducir y amplificar sesgos históricos, generando resultados injustos para grupos protegidos [10], [19]; en particular, [8] encontraron que modelos más flexibles como *gradient boosting* pueden aumentar las disparidades en tasas de aprobación de crédito entre blancos y minorías, ya que explotan correlaciones presentes en los datos que reflejan desigualdades estructurales. A pesar de mejorar la eficiencia predictiva general, estos modelos pueden deteriorar la equidad si no se introducen mecanismos de control.

Dado que muchas variables tradicionalmente utilizadas para evaluar riesgo crediticio están correlacionadas con características demográficas (como ingresos o zona de residencia), los algoritmos pueden aprender reglas que penalizan indirectamente a minorías [11], [20], esto ha motivado el desarrollo de métodos para evaluar y mitigar la discriminación algorítmica. Algunas estrategias se enfocan en la etapa de preprocesamiento (eliminación de sesgos en los datos), otras durante el entrenamiento (regularizaciones adversariales, aprendizaje equitativo), y otras en el posprocesamiento

(ajustes sobre las predicciones para lograr paridad) [12].

Además, herramientas de interpretabilidad como SHAP (*SHapley Additive exPlanations*) permiten analizar el peso de cada variable en las predicciones de modelos complejos. [9] proponen este enfoque como una vía para auditar modelos y detectar si ciertas variables —por ejemplo, proxies geográficos de raza están siendo utilizadas de forma sistemática en contra de determinados grupos.

El enfoque de auditoría externa ha sido promovido por autores como [21], quienes proponen el uso de modelos independientes para evaluar el comportamiento de los sistemas algorítmicos y su impacto en diferentes poblaciones. En el caso de modelos crediticios, entrenar algoritmos "a ciegas" respecto a la variable de etnia no garantiza la equidad. De hecho, [10] argumentan que ignorar variables sensibles puede impedir la detección de sesgos. Algunos estudios han recomendado incluirlas durante el entrenamiento para cuantificar su efecto y, eventualmente, controlarlo aunque su uso en producción suele estar prohibido [22].

Una de las nociones más debatidas en el campo de la equidad algorítmica es la de *paridad demográfica* (demographic parity), que exige que la probabilidad de una predicción positiva sea independiente del grupo sensible al que pertenece un individuo [23]. Sin embargo, este criterio puede entrar en conflicto con otras nociones de equidad, como la *igualdad de oportunidades* (equal opportunity), que solo exige igualdad en la tasa de verdaderos positivos entre grupos [20]. La elección entre estas definiciones no es trivial, pues cada una implica diferentes compromisos entre equidad y precisión, y su aplicación puede depender del contexto normativo y ético de cada dominio.

Otra área en crecimiento es el desarrollo de métricas personalizadas de sesgo algorítmico, como el *disparate impact* (impacto dispar), definido como la razón entre las tasas de resultado positivo para grupos protegidos y no protegidos. Un valor inferior al umbral legal del 80 % (conocido como *four-fifths rule*) puede ser considerado evidencia de discriminación en contextos regulatorios como el de EE.UU. [24]. Esta métrica ha sido utilizada en análisis empíricos sobre modelos crediticios y ha mostrado su utilidad para identificar patrones sistemáticos de exclusión que podrían no ser evidentes en métricas agregadas de desempeño.

Disparate Impact (Impacto Dispar):

$$DI = \frac{P(\hat{Y} = 1 \mid A = \text{minoría})}{P(\hat{Y} = 1 \mid A = \text{mayoría})} \tag{1}$$

DI**:** Medida de impacto dispar (*Disparate Impact*).

$\hat{Y}$**:** Predicción del modelo; $\hat{Y} = 1$ indica una predicción positiva (por ejemplo, aprobación de crédito).

A**:** Atributo sensible (por ejemplo, grupo étnico o racial).

$P(\hat{Y} = 1 \mid A = \textbf{minoría})$**:** Probabilidad de que un individuo del grupo minoritario reciba una predicción positiva.

$P(\hat{Y} = 1 \mid A = \textbf{mayoría})$**:** Probabilidad de que un individuo del grupo mayoritario reciba una predicción positiva.

donde A representa el atributo sensible (como grupo racial o étnico). Un valor de $DI < 0.8$ puede ser interpretado como evidencia de impacto dispar bajo la regla del cuatro quintos.

Equal Opportunity Difference (Diferencia de Igualdad de Oportunidades):

$$\Delta_{EO} = TPR_{\text{minoría}} - TPR_{\text{mayoría}} = P(\hat{Y} = 1 \mid Y = 1, A = \text{minoría}) - P(\hat{Y} = 1 \mid Y = 1, A = \text{mayoría}) \tag{2}$$

Δ_{EO}**:** Diferencia de igualdad de oportunidades entre grupos sensibles.

$TPR_{\textbf{minoría}}$**:** Tasa de verdaderos positivos (True Positive Rate) para el grupo minoritario.

$TPR_{\textbf{mayoría}}$**:** Tasa de verdaderos positivos para el grupo mayoritario.

Y**:** Clase verdadera; $Y = 1$ indica que el individuo cumple con los criterios para un resultado positivo (por ejemplo, ser elegible para crédito).

$\hat{Y}$**:** Predicción del modelo; $\hat{Y} = 1$ implica una predicción positiva.

A**:** Atributo sensible (por ejemplo, etnia, género).

$P(\hat{Y} = 1 \mid Y = 1, A =$ **minoría**$)$**:** Probabilidad de que un individuo del grupo minoritario reciba una predicción positiva, dado que pertenece a la clase positiva.

$P(\hat{Y} = 1 \mid Y = 1, A =$ **mayoría**$)$**:** Probabilidad equivalente para el grupo mayoritario.

Un valor distinto de cero en Δ_{EO} indica una brecha en la oportunidad de obtener un resultado positivo entre grupos, incluso cuando los individuos son igualmente elegibles.

En resumen, la literatura reciente en fairness y ML advierte sobre los riesgos de confiar exclusivamente en métricas de precisión sin considerar cómo los errores del modelo se distribuyen entre distintos grupos demográficos. La equidad en modelos de crédito requiere no solo evaluar su rendimiento global, sino también su impacto distributivo.

4.5. Variables proxy y discriminación indirecta

Una de las mayores preocupaciones en el uso de modelos automatizados de decisión en el crédito hipotecario es la posibilidad de incurrir en discriminación indirecta a través de variables proxy. Estas son variables que, aunque no representan directamente una categoría protegida (como la raza o la etnia), están altamente correlacionadas con ella y pueden inducir sesgos [10], [25], por ejemplo, el código postal o el porcentaje de población minoritaria en el área donde reside el solicitante pueden reflejar composiciones raciales históricamente segregadas [26]. Aunque estas variables pueden ser útiles para capturar riesgos contextuales (como la depreciación del valor del inmueble en ciertas zonas), su uso también puede conducir a exclusión de comunidades enteras, particularmente latinas y afroamericanas [8].

Desde el punto de vista legal, la *Equal Credit Opportunity Act* y el reglamento B prohíben no solo la discriminación intencional, sino también aquella que resulte en un impacto dispar (*disparate*

impact). Esto significa que una práctica puede considerarse discriminatoria si afecta desproporcionadamente a un grupo protegido, incluso si no hubo intención de discriminar [7].

El debate en la literatura gira en torno a cómo balancear la precisión predictiva con la equidad. Algunos proponen eliminar completamente las variables sensibles o sus proxies, mientras que otros sugieren que incluirlas —al menos durante la fase de entrenamiento— puede facilitar la auditoría y permitir ajustes para asegurar paridad [22]. Existen métodos que buscan minimizar la correlación entre las predicciones del modelo y la variable sensible, como la regresión adversarial o técnicas de desbiasing. Sin embargo, estos métodos pueden reducir el rendimiento predictivo y deben implementarse con cautela, especialmente en contextos regulados como el financiero [12], [20].

En nuestro estudio, esta problemática es especialmente relevante, dado que variables como *tract_minority_population_percent* podrían estar actuando como proxies de la etnia, influyendo en la probabilidad de rechazo sin que la variable étnica sea explícitamente usada por el modelo.

5. METODOLOGÍA

La metodología propuesta para esta investigación se estructura en tres grandes etapas: (1) preprocesamiento y análisis exploratorio de datos, (2) aplicación del emparejamiento por puntaje de propensión (PSM), y (3) modelación con algoritmos de aprendizaje automático. Estas etapas se diseñan para identificar y analizar posibles patrones de discriminación étnica hacia personas identificadas como blancos latinos en el proceso de otorgamiento de créditos hipotecarios en Estados Unidos, utilizando la base de datos HMDA del año 2023.

5.1. Preprocesamiento y análisis exploratorio de datos

En la primera etapa se lleva a cabo un proceso de limpieza, depuración y transformación de los datos obtenidos del Home Mortgage Disclosure Act (HMDA) correspondiente al año 2023. Este análisis se limitara a un solo año con el fin de evitar posibles sesgos derivados de cambios económicos o legislativos interanuales. Se define una selección de variables relevantes con base en dos criterios principales: (i) la pertinencia teórica y empírica según la literatura revisada, y (ii) la calidad de los datos en términos de disponibilidad y consistencia. Este conjunto de variables sirve tanto para el emparejamiento por puntaje de propensión como para alimentar los modelos de aprendizaje automático en etapas posteriores.

En esta fase también se separaran dos subpoblaciones de análisis: personas identificadas como blancos hispanos y personas identificadas como blancos no hispanos. La selección de los blancos no hispanos como grupo base de comparación responde a su condición de grupo mayoritario en la población estadounidense y a su uso frecuente como grupo de referencia en estudios de disparidades étnicas. Esta distinción permite focalizar el análisis en el posible efecto discriminatorio asociado a la condición latina, aislando en lo posible otras fuentes de heterogeneidad racial. Adicionalmente, se calculan estadísticas descriptivas (medias, desviaciones estándar, porcentajes de valores nulos, etc.) por grupo poblacional para cada variable, de modo que se puedan identificar diferencias sistemáticas entre las dos poblaciones y anticipar posibles retos para el emparejamiento

y la modelación posterior. Los resultados de esta exploración permiten ajustar la selección de variables finales y definir estrategias de tratamiento para valores atípicos, datos faltantes y escalado de variables si fuera necesario.

5.2. Emparejamiento por puntaje de propensión y marcación de casos

En la segunda etapa de la metodología, se aplicará un enfoque de emparejamiento por puntaje de propensión (PSM, por sus siglas en inglés) con el objetivo de identificar posibles diferencias de trato entre blancos hispanos y blancos no hispanos en la aprobación de créditos hipotecarios. El análisis se enfocará exclusivamente en el año 2023, con el fin de evitar posibles sesgos derivados de variaciones macroeconómicas o normativas entre diferentes periodos.

El primer paso consiste en construir un modelo logístico para estimar la probabilidad de que un individuo pertenezca al grupo de blancos hispanos, en función de variables sociodemográficas y crediticias seleccionadas previamente. Este modelo permitirá asignar un puntaje de propensión a cada observación. Posteriormente, se realizará un emparejamiento uno a uno sin reemplazo ("one-to-one matching"), utilizando diferentes valores de *caliper* para identificar el umbral óptimo. La selección del caliper se basará en un criterio conjunto que considere tanto la cantidad de emparejamientos obtenidos como el equilibrio logrado, medido a través de la mediana de las diferencias estandarizadas de medias (SMD) entre grupos. Además, se realizará el emparejamiento dentro de cada estado de residencia del solicitante, con el fin de controlar por factores legislativos o institucionales locales que puedan afectar las decisiones crediticias. Se excluirán del análisis aquellos estados en los que no se logren suficientes emparejamientos o donde persistan desequilibrios significativos en las covariables tras el PSM.

Una vez conformadas las parejas comparables de blancos hispanos y blancos no hispanos, se procederá a analizar la diferencia en las tasas de aprobación de crédito entre los grupos. Para ello, se aplicará una prueba de chi-cuadrado sobre las tablas de contingencia de aprobación por grupo, tanto a nivel nacional como desagregado por estado. De manera complementaria, se analizará la relación entre la intensidad de la discriminación observada y el grado de rigidez regulatoria en cada

estado. Para esta tarea, se utilizarán los índices de regulación financiera y legislativa recopilados por el proyecto QuantGov[1]. Se espera con esto determinar si existe una correlación entre el entorno institucional y la probabilidad de que un solicitante latino sea desfavorecido frente a su par blanco no hispano.

Finalmente, se identificarán los casos en los que, dentro de una pareja emparejada, el solicitante blanco no hispano fue aprobado y el blanco hispano fue rechazado. Estas observaciones se marcarán como posibles instancias de rechazo discriminatorio. Esta información servirá como base para la fase posterior de modelación, en la que se buscará entender qué características están asociadas con estos rechazos potencialmente injustificados.

5.3. Modelado de rechazos discriminatorios

Tras la marcación de los casos de rechazo discriminatorio mediante emparejamiento por puntaje de propensión, se propone una fase de modelado orientada a identificar patrones y características asociadas a este tipo de rechazos dentro de la población latina. El objetivo es entrenar un modelo de clasificación que, dado un conjunto de datos de solicitantes latinos que fueron rechazados, sea capaz de predecir si ese rechazo corresponde a un caso discriminatorio (según la definición basada en el PSM).

Para esta tarea se plantea la utilización de un modelo basado en *gradient boosting*, particularmente XGBoost, dada su eficacia documentada en la literatura para tareas de clasificación con variables heterogéneas y conjuntos de datos estructurados [27]. Se complementa con la prueba de otros modelos base como *random forest*, regresión logística, etc, buscando contrastar su rendimiento frente a diferentes métricas.

La evaluación de los modelos considera varias métricas, incluyendo AUC-ROC, F1-score, precisión y estadístico Kappa. Adicionalmente, se utiliza SHAP (*SHapley Additive exPlanations*) para interpretar los resultados del modelo final, permitiendo descomponer las predicciones individuales y

[1] https://www.quantgov.org

cuantificar la contribución de cada variable [9]. Esto permite identificar posibles factores indirectos o proxies que estén asociados a la probabilidad de sufrir discriminación.

Aunque el objetivo de esta fase no es construir un clasificador para uso en producción, sino explorar la posibilidad de caracterizar patrones en decisiones injustificadas, los resultados obtenidos pueden aportar a la discusión sobre equidad algorítmica y guiar futuras intervenciones regulatorias o de auditoría en sistemas automatizados de decisión financiera.

6. MÉTODOS

Esta sección describe formalmente las metodologías estadísticas y computacionales utilizadas en el estudio. Se incluyen las definiciones y ecuaciones necesarias para la comprensión del emparejamiento por puntaje de propensión, el cálculo del *Standardized Mean Difference* (SMD), la prueba de chi-cuadrado, y los modelos de clasificación utilizados para la predicción de discriminación en rechazos crediticios.

6.1. Emparejamiento por Puntaje de Propensión (PSM)

El emparejamiento por puntaje de propensión (extitPropensity Score Matching) permite estimar el efecto de un tratamiento (en este caso, ser blanco latino) controlando por covariables observadas [6]. Sea $D_i \in \{0, 1\}$ la variable indicadora de tratamiento (1 si el individuo es blanco latino, 0 si es blanco no hispano) y X_i el vector de covariables. El puntaje de propensión se define como:

$$e(X_i) = P(D_i = 1 \mid X_i) \tag{3}$$

Este puntaje se estima mediante regresión logística o modelos más flexibles como *gradient boosting*. Posteriormente, se realiza el emparejamiento uno a uno sin reemplazo usando un *caliper* δ, de forma que un individuo tratado i se empareja con un individuo control j tal que:

$$|e(X_i) - e(X_j)| < \delta \tag{4}$$

Para garantizar comparabilidad interestatal, se restringió el emparejamiento a dentro del mismo código de estado.

6.2. Balanceo mediante SMD

El balance entre los grupos se evaluó mediante el *Standardized Mean Difference* (SMD), definido para una covariable continua X como:

$$\text{SMD} = \frac{\bar{X}_T - \bar{X}_C}{\sqrt{(s_T^2 + s_C^2)/2}} \tag{5}$$

donde $\bar{X}_T$ y $\bar{X}_C$ son las medias en los grupos tratado y control, y s_T^2, s_C^2 sus varianzas. Un SMD menor a 0.1 indica balance aceptable [28].

Para variables categóricas, se usó:

$$\text{ASMD} = \frac{p_T - p_C}{\sqrt{(p_T(1-p_T) + p_C(1-p_C))/2}} \tag{6}$$

donde p_T y p_C representan proporciones en cada grupo.

6.3. Prueba de independencia: Chi-Cuadrado

La significancia de la diferencia en tasas de aprobación entre pares emparejados se evaluó mediante la prueba de chi-cuadrado para tablas 2x2, con estadístico:

$$\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \tag{7}$$

donde O_{ij} son las frecuencias observadas y E_{ij} las esperadas bajo independencia. Esta prueba se realizó para cada estado y de forma agregada.

6.4. Modelos de Clasificación y Análisis SHAP

Para detectar patrones asociados a rechazos discriminatorios en blancos latinos, se entrenó un modelo de clasificación tipo *XGBoost*, entrenado sobre los casos etiquetados como discriminados o no discriminados tras el PSM. Se evaluó el rendimiento con las siguientes métricas:

Para evaluar el desempeño de los modelos clasificadores utilizados en esta tesis, se emplean diversas métricas que permiten cuantificar la calidad de las predicciones desde múltiples perspectivas. Estas métricas se construyen a partir de los siguientes componentes de la matriz de confusión:

- **True Positives (TP)**: Casos en los que el modelo predice correctamente la clase positiva.
- **True Negatives (TN)**: Casos en los que el modelo predice correctamente la clase negativa.
- **False Positives (FP)**: Casos en los que el modelo predice la clase positiva, pero en realidad pertenecen a la clase negativa.
- **False Negatives (FN)**: Casos en los que el modelo predice la clase negativa, pero en realidad pertenecen a la clase positiva.

A partir de estas cantidades, se definen las siguientes métricas:

- **Precisión** (Precision): Mide la proporción de verdaderos positivos entre todas las observaciones que el modelo clasificó como positivas. Evalúa la capacidad del modelo para evitar falsos positivos. Se calcula como:
$$\text{Precisión} = \frac{TP}{TP + FP}$$
- **Recall** (Sensibilidad o Tasa de Verdaderos Positivos): Mide la proporción de verdaderos positivos que fueron correctamente identificados por el modelo sobre el total de positivos reales. Evalúa la capacidad del modelo para detectar todos los casos positivos. Se calcula como:
$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1 Score**: Es la media armónica entre la precisión y el recall. Ofrece un balance entre ambos, especialmente útil cuando existe desbalance de clases. Se calcula como:

$$F1 = 2 \cdot \frac{\text{Precisión} \cdot \text{Recall}}{\text{Precisión} + \text{Recall}}$$

- **AUC-ROC** (Area Under the Receiver Operating Characteristic Curve): Representa el área bajo la curva ROC, la cual grafica la tasa de verdaderos positivos (Recall) frente a la tasa de falsos positivos (FPR = $\frac{FP}{FP+TN}$). Un valor de AUC cercano a 1 indica una alta capacidad del modelo para discriminar entre clases.

- **Kappa de Cohen**: Mide el grado de concordancia entre las predicciones del modelo y las clases reales, ajustando por el acuerdo que podría esperarse por azar. Un valor de 1 indica concordancia perfecta, mientras que 0 indica acuerdo equivalente al azar.

- **Exactitud** (Accuracy): Indica la proporción de predicciones correctas (positivas y negativas) sobre el total de observaciones. Es una medida general de desempeño, aunque puede ser engañosa en contextos de clases desbalanceadas. Se calcula como:

$$\text{Exactitud} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Log Loss** (Logaritmic Loss o Binary Cross-Entropy): Mide el error en las predicciones de probabilidad del modelo. Penaliza fuertemente las predicciones con alta confianza que resultan ser incorrectas. Cuanto menor sea su valor, mejor es el ajuste probabilístico del modelo.

- **MCC** (Matthews Correlation Coefficient): Es un coeficiente de correlación que toma en cuenta todos los elementos de la matriz de confusión. Es especialmente útil en problemas con clases desbalanceadas. Su valor varía entre -1 (predicción totalmente errónea), 0 (no mejor que el azar) y 1 (predicción perfecta). Se calcula como:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

- **Brier Score**: Calcula el error cuadrático medio entre las probabilidades predichas y los valores reales (0 o 1). Evalúa la calibración del modelo, es decir, qué tan bien se alinean las probabilidades predichas con las frecuencias observadas. Un valor más cercano a 0 indica mejor calibración.

Adicionalmente, se aplicó el método SHAP (SHapley Additive exPlanations) para interpretar la contribución de cada variable en la predicción de discriminación [9].

6.5. Correlación con Regulaciones Estatales

Finalmente, se evaluó la asociación entre el grado de protección legal en cada estado (según los datos regulatorios obtenidos de *QuantGov*) y el nivel de discriminación observada. Se utilizó el coeficiente de correlación de Spearman (ρ) para captar relaciones monótonas no lineales:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \tag{8}$$

donde d_i es la diferencia entre los rangos de cada variable y n el número de estados considerados.

7. RESULTADOS

En la fase inicial de preprocesamiento se realizó un análisis descriptivo de las variables encontradas en el Dataset HMDA 2023 así como una limpieza de valores extremos para evitar posible ruido en el análisis posterior.

Tras estas filtraciones, la muestra de trabajo consiste de decenas de miles de solicitudes (el número exacto depende de los filtros; HMDA 2023 en bruto contiene millones de registros, pero la mayoría de solicitantes blancos no son hispanos; los blancos latinos son una fracción menor, típicamente concentrada en estados como California, Texas, Florida, etc.). En términos de **preprocesamiento**, se realizan las siguientes tareas: *Limpieza y transformación de variables*: Por ejemplo, la variable de ingreso viene en bruto; la convertimos a una escala de miles para facilitar su lectura. Las variables categóricas (como el tipo de puntaje crediticio, p. ej. "FICO 04", "VantageScore 3.0", etc.) se convierten a indicadores binarios (*one-hot encoding*). *Manejo de valores faltantes*: HMDA 2023, en su versión pública, puede tener datos faltantes para ciertos campos sensibles, en estos casos considerando la importancia de la variable y si la naturaleza del nulo implicaba información adicional o no se toma la decisión de eliminar la variable (cuando el porcentaje de nulos era muy elevado) o de eliminar los registros específicos en caso de que la variable estuviese poblada en su generalidad.

Normalización/estandarización: Variables continuas como ingreso, monto del préstamo, etc., se escalan (estandarización Z-score) cuando se usan en modelos, aunque para la intuición de los resultados reportamos en unidades originales en la discusión cuando sea relevante. Los algoritmos basados en árboles (XGBoost) no requieren estrictamente normalización, pero los basados en distancia (por ejemplo, en la etapa de matching) sí pueden beneficiarse de escalas comparables.

A partir del análisis exploratorio realizado previamente y considerando la literatura existente sobre el uso del dataset de HMDA para estudiar patrones de discriminación —por ejemplo, trabajos como [29], [2] y [30]— se seleccionaron tanto variables numéricas (ver tabla 1) como categóricas (ver tabla 2) que fuesen estadísticamente relevantes y conceptualmente justificadas para realizar

Tabla 1. Resumen de variables numéricas

Variable	Mean_Hisp	Std_Hisp	Nulls_Hisp (%)	Mean_BNH	Std_BNH	Nulls_BNH (%)
income	0.10	0.08	0.00	0.12	0.10	0.00
loan_amount	237.40	166.98	0.00	223.39	181.08	0.00
tract_minority_population_percent	55.90	28.50	0.00	25.66	19.06	0.00
tract_to_msa_income_percentage	96.41	42.90	0.00	108.68	41.27	0.00
loan_to_value_ratio	78.77	21.74	28.10	71.68	23.06	26.56
total_loan_costs	8.76	5.67	57.58	6.23	4.68	56.91
origination_charges	3.92	3.56	56.67	2.97	3.03	56.64
loan_term	329.24	67.00	2.72	318.00	77.91	5.14
prepayment_penalty_term	33.35	5.58	96.25	31.53	7.51	95.25
tract_population	5011.04	2644.03	0.00	4536.02	2038.10	0.00
tract_owner_occupied_units	1101.92	613.80	0.00	1241.10	576.41	0.00
interest_rate	6.96	1.45	44.52	7.05	1.44	37.57
rate_spread	0.63	1.29	50.36	0.52	1.29	42.92

Fuente: elaboración, propia con base en HMDA 2023.

Tabla 2. Resumen de variables categóricas

Variable	Moda_Hisp	Moda %_Hisp	Nulls_Hisp (%)	Moda_BNH	Moda %_BNH	Nulls_BNH (%)
Debt-to-income ratio	50 %-60 %	15.67	22.71	20 %-30 %	15.41	20.05
Applicant age	35–44	26.58	0.00	35–44	21.88	0.00
Loan purpose	Home purchase	63.38	0.00	Home purchase	52.09	0.00
Applicant credit score type	Not applicable	28.38	3.10	Not applicable	26.57	3.79
Co-applicant credit score type	No co-applicant	50.74	0.58	No co-applicant	44.16	1.19
Occupancy type	Principal residence	93.44	0.00	Principal residence	93.84	0.00

Fuente: elaboración, propia con base en HMDA 2023.

el emparejamiento mediante Propensity Score Matching (PSM). Esta selección permitió reducir sesgos y asegurar que las comparaciones entre individuos fueran realizadas sobre casos similares en cuanto a sus características crediticias y sociodemográficas, minimizando así el efecto de variables confusoras.

Se utilizó el dataset del año 2023 de HMDA, filtrando a los individuos blancos no hispanos y blancos hispanos, con el fin de mantener la variable de raza constante y aislar la dimensión étnica. La elección de un solo año busca evitar introducir ruido por variaciones económicas, políticas o legislativas entre periodos.

7.1. Emparejamiento por Estado y Selección de Caliper Óptimo

El emparejamiento se realizó de manera estratificada por estado, considerando que las regulaciones bancarias y las condiciones económicas pueden variar significativamente entre jurisdicciones en los Estados Unidos. Esta decisión se alinea con hallazgos previos que sugieren que los patrones de aprobación de créditos pueden estar influidos por marcos normativos locales [26].

Se utilizó un modelo logístico regularizado para estimar el propensity score basado en las variables seleccionadas. Posteriormente, se realizó un emparejamiento 1-a-1 sin reemplazo utilizando la distancia de menor diferencia en el propensity score, bajo una grilla de calipers que va desde 0.01 hasta 0.2. Para cada valor de caliper, se evaluó la mediana del SMD penalizado (penalización por menor cantidad de pares) y se seleccionó aquel que maximizaba el balance estadístico sin reducir drásticamente el número de emparejamientos. Se descartaron aquellos estados donde:

- El número de pares resultante era insuficiente.
- La mediana de SMD superaba el umbral de 0.1.
- Alguna variable individual presentaba un SMD o ASMD mayor a 0.1.

En el presente análisis se utilizó regresión logística con regularización L2 (Ridge) como modelo

base para estimar los puntajes de propensión (Propensity Scores) en el proceso de emparejamiento. Esta elección se debe a su capacidad para reducir la varianza en presencia de multicolinealidad, manteniendo todos los coeficientes en el modelo y evitando la eliminación de variables relevantes, lo cual es especialmente importante cuando se incluyen variables dummificadas provenientes de categóricas con múltiples niveles. Esta técnica fue implementada mediante el estimador `LogisticRegression` de la biblioteca `scikit-learn`, configurando explícitamente el parámetro `penalty='l2'`. La constante de regularización fue fijada en $C = 0.2$, equivalente a un parámetro $\lambda = \frac{1}{C} = 5$, con el fin de controlar la magnitud de los coeficientes y mitigar posibles efectos de multicolinealidad. Este valor fue seleccionado con base en el criterio de optimización del balance estadístico entre grupos (mediana del SMD penalizado) a lo largo del proceso de emparejamiento iterativo en cada estado, maximizando simultáneamente el número de pares válidos dentro de los límites de caliper evaluados (0.01 a 0.2). No se aplicó selección automática por LASSO o Elastic Net debido a la intención de mantener consistencia con la literatura sobre variables explicativas en crédito. Sin embargo, esto se reconoce como una limitación."

Los resultados de cada estado válido se almacenaron, incluyendo el caliper seleccionado, el número de pares, la mediana de los SMDs, el *odds ratio* de aprobación para hispanos frente a blancos no hispanos, y su respectivo valor p.

Tabla 3. Detalle de resultados por estado válido

Estado	Caliper	Pares	Odds Ratio	p-valor	Rate (Appv) Hispano	Rate (Appv) BNH	Mediana SMD
TX	0.01	108 733	0.747 899	$1.532\,821 \times 10^{-177}$	0.744 034	0.795 343	0.023 24
NY	0.10	12 496	0.636 023	$7.338\,918 \times 10^{-58}$	0.669 414	0.761 044	0.032 87
FL	0.10	93 037	0.699 178	$9.337\,446 \times 10^{-260}$	0.687 447	0.758 881	0.026 34
IL	0.01	21 324	0.755 848	$2.827\,252 \times 10^{-31}$	0.772 697	0.818 186	0.025 34
CO	0.01	13 588	0.716 543	$7.105\,152 \times 10^{-30}$	0.748 381	0.805 785	0.014 28
CA	0.10	73 855	0.969 059	$8.165\,606 \times 10^{-3}$	0.742 170	0.748 182	0.023 35

7.2. Disparidad en tasas de aprobación tras emparejamiento

Luego de aplicar el PSM, contamos con un conjunto de 323 mil pares emparejados de solicitantes (blanco latino vs. blanco no latino) con características similares (ver tabla 3). Al comparar

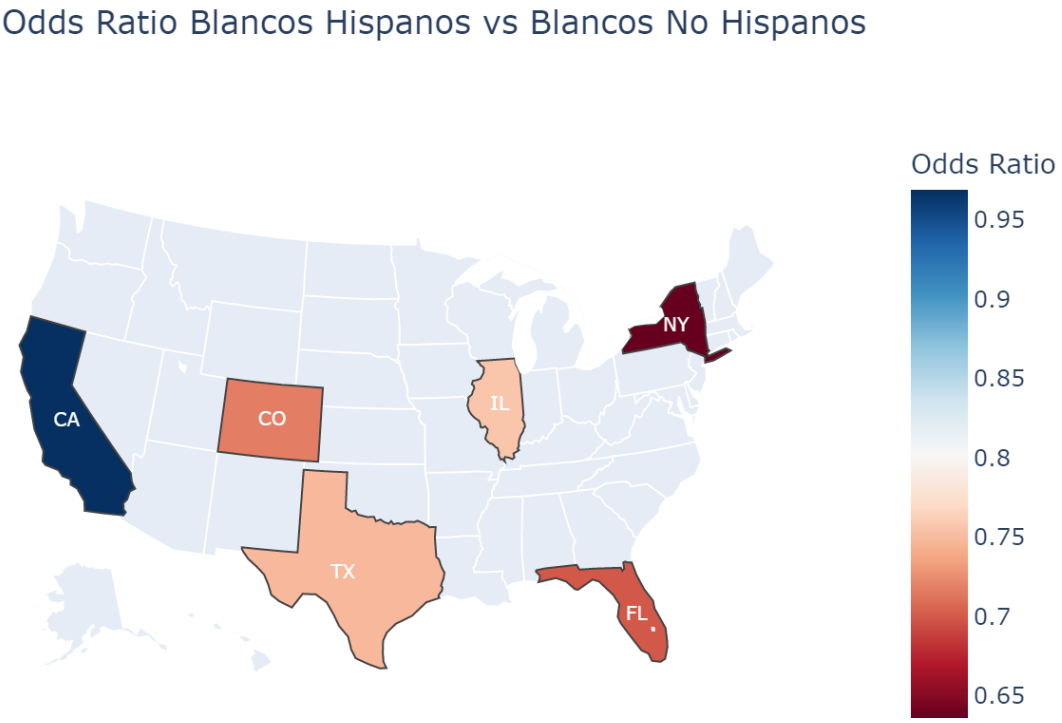

Fig. 1: Mapa de calor del odds ratio de aprobación para blancos hispanos en comparación con blancos no hispanos por estado. Valores por debajo de 1 indican menor probabilidad relativa de aprobación para los hispanos, lo cual puede señalar disparidad en el trato.

la proporción de aprobaciones entre estos dos grupos emparejados, se observa una diferencia consistente: El odds ratio de 0.773 a nivel nacional indica que, controlando por las demás variables del modelo, un solicitante hispano tiene un 22.7 % menos de probabilidades relativas (odds) de ser aprobado para un crédito hipotecario en comparación con un solicitante blanco. En otras palabras, los odds (posibilidades relativas, no probabilidades directas) de aprobación para un hispano son aproximadamente el 77.3 % de los de un blanco, lo que sugiere una desventaja sistemática para el grupo hispano en el proceso de aprobación. Este diferencial es estadísticamente significativo ($p < 0.01$ en prueba de diferencia de proporciones). En términos relativos, implica que los blancos latinos en la muestra, pese a tener perfiles crediticios equivalentes a los blancos no latinos, sufrieron una tasa de negación más alta. Este resultado es indicativo de una potencial discriminación étnica: la variable "ser latino" estaría asociada a un menor chance de aprobación, una vez igualadas otras condiciones. Adicionalmente, exploramos la disparidad por subgrupos geográficos. En la figura 1 es posible visualizar que la brecha de aprobación varía entre estados: en California, la diferencia emparejada fue menor, mientras que en Florida y Texas alcanzó valores mayores. Esto podría indicar que en algunos mercados la discriminación étnica hacia latinos blancos es más pronunciada que en otros, posiblemente correlacionado con factores culturales o nivel de supervisión regulatoria. No obstante, el tamaño de muestra por estado emparejada se reduce, así que estas cifras se deben interpretar con precaución.

7.3. Agregación Nacional y Estimación Global del Efecto

Una vez identificados los estados válidos, se concatenaron todos los pares emparejados de estos estados para realizar una estimación a nivel nacional. Se calcularon nuevamente los SMD y ASMD a nivel agregado, obteniéndose un *odds ratio* global y una prueba *chi-cuadrado* de independencia entre grupo étnico y aprobación crediticia (referirse a figuras 2, 3). Este resultado representa el efecto promedio de la étnia en las decisiones de aprobación de crédito para individuos comparables según sus características.

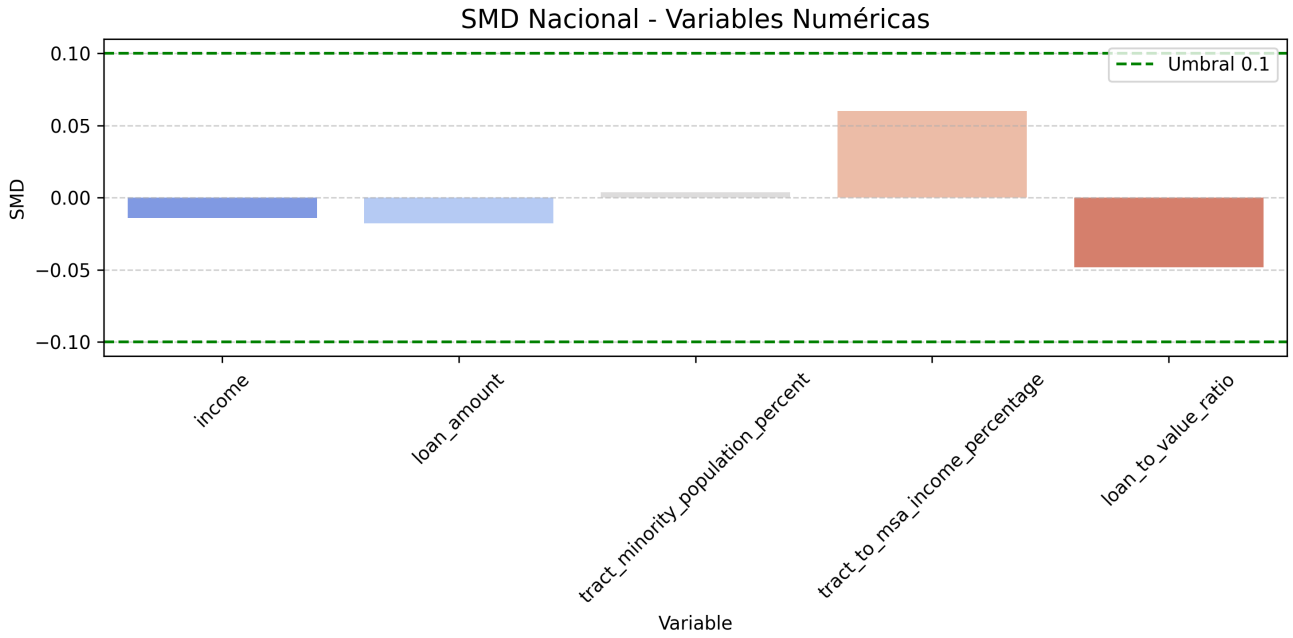

Fig. 2: SMD Nacional

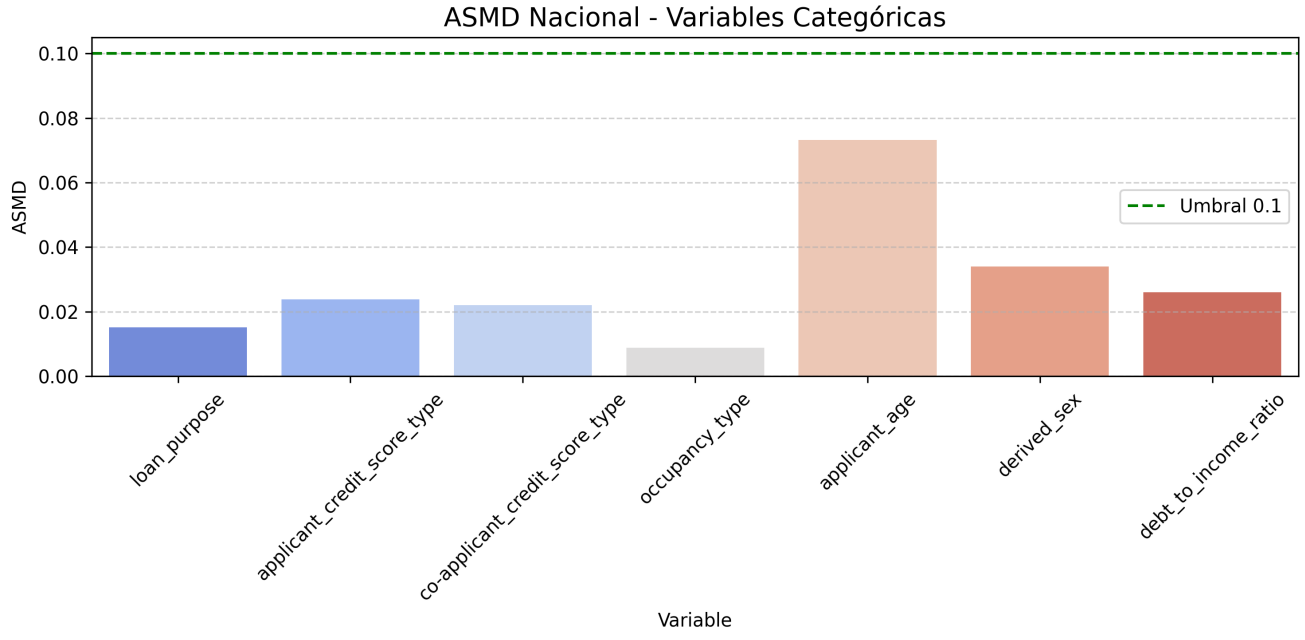

Fig. 3: ASMD Nacional

7.4. Relación con el Marco Regulatorio

Se exploró la hipótesis de que un mayor nivel de regulación financiera podría asociarse con menores niveles de discriminación. Para ello, se recolectó información desde QuantGov sobre el número de estatutos y restricciones regulatorias en el ámbito de banca, seguros y valores por estado (últimos tres años). Se construyó un mapa de calor y una gráfica de dispersión que relaciona estas medidas con el *odds ratio* obtenido en el PSM por estado. Este análisis no obtuvo resultados concluyentes (lo cual se soporta en los resultados obtenidos al calcular el coeficiente de correlación de pearson, ver figura 5) el cual puede observarse gráficamente en la figura 4; de igual forma es importante considerar el hecho de que solo unos pocos estados se utilizaron para realizar esta comparación debido al filtrado que se hizo en el paso anterior (para garantizar que las poblaciones fueran comparables).

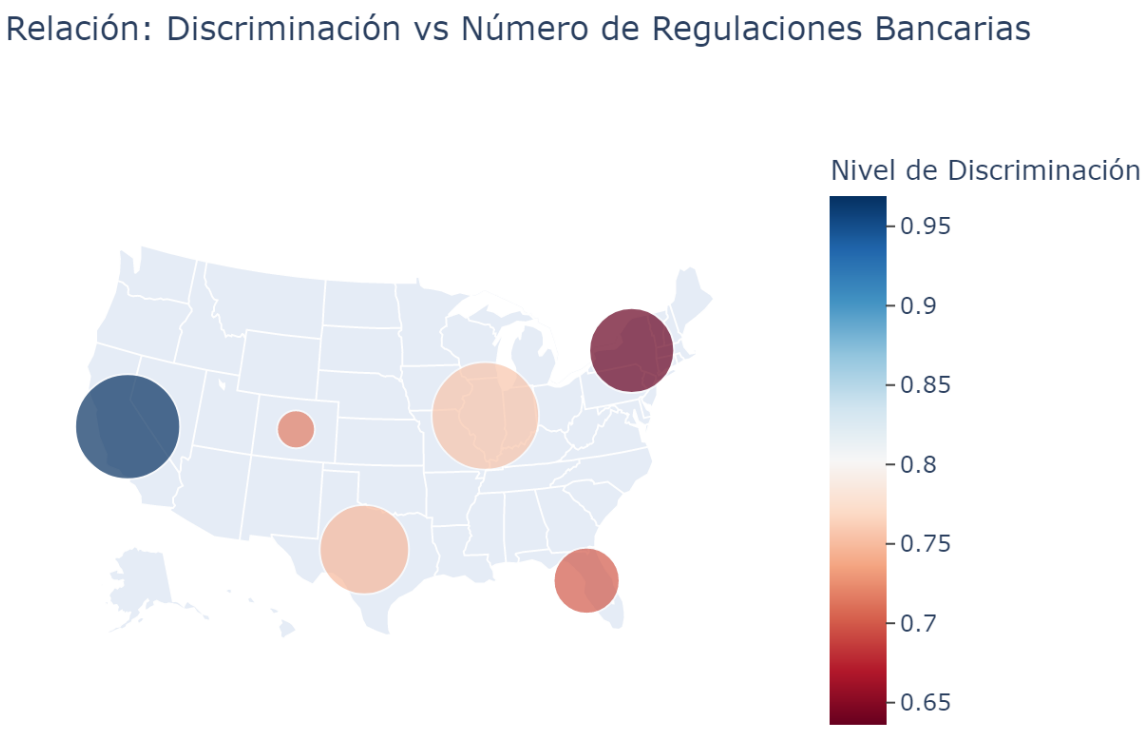

Fig. 4: Número de regulaciones bancarias incluidas en los últimos tres años vs Odds Ratio

Fig. 5: Número de regulaciones bancarias incluidas en los últimos tres años vs Odds Ratio

7.5. Etiquetado de Casos de Discriminación

Se identificaron aquellos pares en los que el individuo hispano fue rechazado mientras que su par blanco no hispano fue aprobado. A estos casos se les asignó una etiqueta binaria que identifica presuntas situaciones de discriminación. Estos registros fueron exportados para ser utilizados en la siguiente etapa de modelación predictiva del fenómeno discriminatorio (ver tabla 4).

Tabla 4. Cantidad de rechazos a solicitantes latinos etiquetados como posiblemente discriminatorios y no discriminatorios, a nivel nacional y por estado

Región	**Rechazos Discriminatorios**	**Rechazos No Discriminatorios**
Nacional	68 251	20 099
Texas (TX)	22 155	5677
Nueva York (NY)	3134	997
Florida (FL)	22 039	7040
Illinois (IL)	3931	916
Colorado (CO)	2740	679
California (CA)	14 252	4790

7.6. Modelación

Un caso se considera potencialmente discriminatorio si, tras el emparejamiento por PSM, el solicitante latino fue rechazado mientras su par blanco no hispano fue aprobado, con condiciones crediticias equivalentes. Estas etiquetas fueron utilizadas como variable objetivo para entrenar los modelos de clasificación.

La idea detrás del entrenamiento de un modelo de clasificación binaria (rechazo discriminatorio vs no discriminatorio) se centra en el hecho de que un éxito en la capacidad del modelo para generalizar podría arrojar indicios de que la probabilidad de ser discriminado por el factor étnico (blanco latino) se encuentra asociada a factores financieros o geográficos del solicitante, por ejemplo

la hipótesis de que los blancos hispanos con menores ingresos tengan una mayor probabilidad de ser discriminados que su contraparte con mayores ingresos.

Como consecuencia del desbalance de clases presente, se utilizaron dos técnicas de balanceo de clases (en diferentes iteraciones respectivamente), las técnicas utilizadas fueron *downsampling* y SMOTE (SMOTENC para variables categóricas), con la ventaja relativa de evitar la creación de datos sintéticos y de evitar la pérdida de datos (para cada una respectivamente). Adicionalmente se iteró la creación y testeo de los modelos tanto a nivel nacional como por estado (por la cantidad de registros se seleccionaron Texas y Florida); adicionalmente se realizó un análisis de correlación entre variables con el fin de descartar posible ruido en el modelo, sin embargo no se encontraron correlaciones significativas entre variables.

Para asegurar robustez en los resultados, se empleó **validación cruzada estratificada de 5 pliegues** a través de `GridSearchCV`, optimizando el `ROC-AUC` como métrica de selección. Se exploró un espacio de hiperparámetros que incluía: `n_estimators` (100 y 200), `max_depth` (4, 6, 8), `learning_rate` (0.01, 0.05, 0.1), y tasas de muestreo (`subsample` y `colsample_bytree`). Se exploraron diversas técnicas de modelación para el clasificador binario (XGBoost, LightGBM, etc) en ejercicios tanto a nivel nacional (agregado total) como a nivel estatal, sin embargo los resultados obtenidos no son concluyentes por lo cual se toma la decisión de presentar toda la documentación respectiva a los diferentes modelos en los anexos del presente trabajo de tesis. De cara a futuras investigaciones se sugiere explorar limitantes del dataset utilizado, otros datasets disponibles y diferentes enfoques de modelación como podrían ser los modelos de regresión.

8. CONCLUSIONES

Los hallazgos presentados en esta investigación refuerzan la hipótesis de que la discriminación en las decisiones de crédito hipotecario no puede explicarse unicamente por factores observables en las condiciones crediticias individuales, sino que responde a un fenómeno complejo, multifactorial y parcialmente estructural, donde la etnia del solicitante opera como un factor transversal e indirecto a través de variables correlacionadas. Al emplear un enfoque riguroso basado en emparejamiento por puntuación de propensión (PSM) y modelación algorítmica supervisada, fue posible aislar el efecto de la etnia latina sobre la probabilidad de rechazo de un crédito, mostrando que incluso en condiciones crediticias similares, los latinos presentan una menor tasa de aprobación que sus contrapartes blancos no hispanos.

Uno de los resultados mas reveladores surge del análisis de interpretabilidad del modelo de Gradient Boosting Classifier entrenado con los datos nacionales submuestreados (downsample), donde se evidencia que variables como la proporción de población minoritaria en el sector censal (extttttract_minority_population_percent) y el estado de residencia tienen una influencia considerable sobre la probabilidad de que un caso sea clasificado como discriminatorio. Este hallazgo apunta a un mecanismo de discriminación indirecta donde el sistema algorítmico, al incorporar variables geográficas o socio-espaciales, termina reflejando patrones segregativos históricos que penalizan de forma sistemática a ciertos grupos [26].

El desempeño general de los modelos, tanto a nivel nacional como en la desagregación por el estado de Texas, sugiere que, aunque la capacidad predictiva de los algoritmos no es sobresaliente en términos absolutos, los patrones identificados son consistentes. El bajo valor de métricas como el Kappa de Cohen en la mayoría de los modelos entrenados indica la dificultad inherente a clasificar de forma precisa los casos discriminatorios debido a la complejidad del fenómeno y a las limitaciones del dataset. Estas limitaciones incluyen la ausencia de variables como el estatus migratorio, la generación migrante o factores subjetivos como la percepción de riesgo de los oficiales de crédito.

Para la población blanca hispana, estos hallazgos aportan evidencia cuantitativa de una posi-

ble desventaja sistemática en el acceso al crédito, incluso cuando se controla por características económicas, crediticias y demográficas. En términos regulatorios, se plantea una oportunidad urgente para revisar las variables permitidas en los modelos de *scoring* y fortalecer los mecanismos de auditoria algorítmica que actualmente no logran identificar o mitigar estos sesgos. Las agencias regulatorias podrían considerar exigencias de *explainability* y auditorias ex ante y ex post a modelos de crédito empleados por entidades financieras, especialmente aquellos que integran datos geoespaciales.

Finalmente, este trabajo sienta las bases para futuras investigaciones que deseen profundizar en la causalidad del fenómeno discriminatorio en créditos hipotecarios. Seria valioso incorporar encuestas cualitativas, experimentos aleatorizados o datasets que incluyan mas dimensiones de vulnerabilidad (estatus migratorio, generación, idioma predominante, etc.). Asimismo, resulta clave explorar las implicaciones distributivas de los algoritmos de decisión automática y su papel en la reproducción de desigualdades estructurales en el acceso a bienes esenciales como la vivienda.

REFERENCIAS

[1] L. Quillian, J. J. Lee y A. Honoré, «Racial Discrimination in Housing: A Review of the Audit-Based Literature and the Challenges Ahead,» *Annual Review of Sociology*, vol. 46, págs. 261-280, 2020.

[2] R. Bartlett, A. Morse, R. Stanton y N. Wallace, «Consumer-lending discrimination in the fintech era,» *Journal of Financial Economics*, vol. 143, n.º 1, págs. 30-56, 2022.

[3] L. A. Martinez, Z. Valdez e Y. Padilla, «How racialized perceptions shape Latinos' experiences with discrimination,» *Socius*, vol. 7, págs. 1-11, 2021.

[4] G. R. Sanchez, E. D. Vargas y M. H. Lopez, «Hispanic immigrants' perceptions of discrimination and support for immigration policy,» *Social Science Quarterly*, vol. 102, n.º 1, págs. 232-247, 2021.

[5] J. S. Rugh y D. S. Massey, «Racial segregation and the American foreclosure crisis,» *American Sociological Review*, vol. 75, n.º 5, págs. 629-651, 2010.

[6] P. R. Rosenbaum y D. B. Rubin, «The central role of the propensity score in observational studies for causal effects,» *Biometrika*, vol. 70, n.º 1, págs. 41-55, 1983.

[7] R. Bartlett, A. Morse, R. Stanton y N. Wallace, «Consumer protection in an age of algorithms,» *Journal of Financial Regulation*, vol. 5, n.º 1, págs. 1-36, 2019.

[8] A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai y A. Walther, «Predictably unequal? The effects of machine learning on credit markets,» *Journal of Finance*, vol. 77, n.º 1, págs. 5-47, 2022.

[9] S. M. Lundberg y S.-I. Lee, «A unified approach to interpreting model predictions,» en *Advances in Neural Information Processing Systems*, vol. 30, 2017, págs. 4765-4774.

[10] S. Barocas y A. D. Selbst, «Big Data's Disparate Impact,» *California Law Review*, vol. 104, n.º 3, págs. 671-732, 2016.

[11] A. Chouldechova, «Fair prediction with disparate impact: A study of bias in recidivism prediction instruments,» *Big data*, vol. 5, n.º 2, págs. 153-163, 2017.

[12] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman y A. Galstyan, «A survey on bias and fairness in machine learning,» *ACM Computing Surveys (CSUR)*, vol. 54, n.º 6, págs. 1-35, 2021.

[13] J. Martinez, S. L. Ross y J. Yinger, «The Secret Sharer: Measuring Racial Disparities in Mortgage Lending Using HMDA Data,» *Cityscape*, vol. 23, n.º 1, págs. 159-186, 2021.

[14] A. Glantz, «Kept Out: For people of color, banks are shutting the door to homeownership,» *Center for Investigative Reporting*, 2018, Available at: https://revealnews.org/article/for-people-of-color-banks-are-shutting-the-door-to-homeownership/.

[15] S. Popick, «Racial Disparities in Mortgage Lending after the 2018 HMDA Expansion,» *Consumer Financial Protection Bureau Working Paper*, n.º 2022-05, 2022.

[16] N. Bhutta y A. Hizmo, «Do Minorities Pay More for Mortgages?» *Review of Financial Studies*, vol. 34, n.º 2, págs. 763-789, 2021.

[17] L. Zhang y P. Willen, «Do Lenders Discriminate in the Mortgage Market? Evidence from New Data,» *Federal Reserve Bank of Boston Working Paper*, n.º 21-4, 2021.

[18] A. Hanson y Z. Hawley, «Discrimination in Mortgage Lending: Evidence from a Correspondence Experiment,» *Journal of Urban Economics*, vol. 93, págs. 48-65, 2016.

[19] H. Suresh y J. V. Guttag, «A framework for understanding unintended consequences of machine learning,» *Communications of the ACM*, vol. 64, n.º 11, págs. 62-71, 2021.

[20] M. Hardt, E. Price y N. Srebro, «Equality of Opportunity in Supervised Learning,» *Advances in Neural Information Processing Systems*, vol. 29, 2016. dirección: `https://proceedings.neurips.cc/paper_files/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf`.

[21] I. D. Raji y J. Buolamwini, «Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products,» en *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 2019, págs. 429-435.

[22] R. Binns, M. Veale, M. Van Kleek y N. Shadbolt, «Apparent unfairness: Perceptions of algorithmic decisions,» *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, págs. 1-14, 2020.

[23] C. Dwork, M. Hardt, T. Pitassi, O. Reingold y R. Zemel, «Fairness through awareness,» *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, págs. 214-226, 2012.

[24] M. Feldman, S. A. Friedler, J. Moeller, C. Scheidegger y S. Venkatasubramanian, «Certifying and removing disparate impact,» en *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, págs. 259-268.

[25] B. Kim, A. Ghorbani y J. Zou, «Model multiplicity and rationalizability: On the value of multiple explanations for high-stakes decisions,» *Advances in Neural Information Processing Systems*, vol. 34, págs. 21 297-21 309, 2021.

[26] J. S. Rugh, «Race, space, and cumulative disadvantage: A case study of the subprime lending collapse,» *Social Problems*, vol. 62, n.º 2, págs. 186-218, 2015.

[27] T. Chen y C. Guestrin, «XGBoost: A scalable tree boosting system,» en *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, págs. 785-794.

[28] P. C. Austin, «Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples,» *Statistics in Medicine*, vol. 28, n.º 25, págs. 3083-3107, 2009.

[29] N. Bhutta y A. Hizmo, «How much do we know about racial disparities in mortgage lending?» *The Review of Financial Studies*, vol. 32, n.º 11, págs. 3947-3991, 2019.

[30] P. Bayer, F. Ferreira y S. L. Ross, «Race, ethnicity, and high-cost mortgage lending,» *American Economic Journal: Economic Policy*, vol. 10, n.º 1, págs. 1-39, 2018.

[31] R. Berk, H. Heidari, S. Jabbari, M. Kearns y A. Roth, «Fairness in criminal justice risk assessments: The state of the art,» *Sociological Methods & Research*, vol. 50, n.º 1, págs. 3-44, 2021.

[32] D. Madras, E. Creager, T. Pitassi y R. Zemel, «Predict responsibly: Improving fairness and accuracy by learning to defer,» en *Advances in Neural Information Processing Systems*, vol. 31, 2018, págs. 6147-6157. dirección: `https://proceedings.neurips.cc/paper_files/paper/2018/file/a4d8f8e4ee0e8a2d46c58fc5a5e0f9d4-Paper.pdf`.

[33] G. D. Squires y J. Chadwick, «Linguistic Isolation, Residential Segregation, and Disparities in Mortgage Lending,» *Journal of Housing Research*, vol. 15, n.º 1, págs. 49-68, 2006.

ANEXOS

Anexo A. Variables sintéticas y resultados de modelación

Variables sintéticas propuestas

A continuación se describen variables sintéticas derivadas, creadas con el objetivo de aportar información relevante para un modelo de clasificación binaria que busca identificar posibles rechazos discriminatorios en solicitudes de crédito.

- **Ingreso relativo al ingreso medio del área**:

$$\text{income\_to\_tract\_median} = \frac{\text{income}}{\text{tract\_to\_msa\_income\_percentage}} \tag{A.1}$$

  Esta variable refleja la posición relativa del ingreso del solicitante en relación con el ingreso promedio del área metropolitana donde reside. Un ingreso significativamente menor podría estar asociado a un mayor riesgo percibido por las entidades, pero su interacción con la variable étnica puede revelar sesgos en el proceso de aprobación.

- **Índice de vulnerabilidad del área**:

$$\text{area\_vulnerability\_index} = \text{tract\_minority\_population\_percent} \cdot (1 - \text{tract\_to\_msa\_income\_percentage}) \tag{A.2}$$

  Esta variable captura la composición racial y el nivel socioeconómico del vecindario. Se asume que áreas con alta proporción de minorías y bajos ingresos relativos pueden estar sujetas a mayor riesgo percibido o sesgos sistemáticos.

- **Riesgo social combinado**:

$$\text{riesgo\_social} = \text{tract\_minority\_population\_percent} \cdot \text{loan\_to\_value\_ratio} \tag{A.3}$$

  Esta variable combina el nivel de concentración de minorías con el ratio préstamo-valor del inmueble, lo cual puede capturar situaciones donde solicitudes con características similares reciben distinto tratamiento según el perfil demográfico del área.

- **Indicador de ingreso medio en zona vulnerable**:

$$\text{income\_vs\_tract\_low} = \begin{cases} 1 & \text{si income} > \text{mediana(income) y tract\_to\_msa\_income\_percentage} < 80 \\ 0 & \text{en otro caso} \end{cases} \tag{A.4}$$

  Este indicador binario busca detectar si los solicitantes con ingresos relativamente altos están siendo rechazados por vivir en zonas de bajo ingreso, lo que puede señalar sesgos geográficos encubiertos.

- **Indicador de alto ratio préstamo-valor**:

$$\text{high\_ltv\_flag} = \begin{cases} 1 & \text{si loan\_to\_value\_ratio} > 0.8 \\ 0 & \text{en otro caso} \end{cases} \tag{A.5}$$

  El ratio préstamo-valor es un factor tradicional de riesgo. Esta variable binaria permite evaluar si los solicitantes hispanos con ratios altos están siendo penalizados de manera sistemática en comparación con otros grupos.

Modelado con XGBoost

Para las diferentes iteraciones (entiéndase por iteración el entrenamiento y testeo de un modelo con diferentes técnicas de balanceo, des-agregación por estado o agrupación nacional, eliminación o inclusión de variables, etc) se utilizaron diferentes técnicas de modelación, entre las cuales se le prestó especial atención al modelado por medio de XGBoost por su robustez, capacidad para manejar datos tabulares y rendimiento superior en tareas de clasificación con desbalance [27]. Su habilidad para capturar interacciones no lineales sin requerir un preprocesamiento intensivo lo hace ideal para datasets estructurados como los de crédito hipotecario. Esta elección también es coherente con otros estudios que han utilizado XGBoost para detectar sesgos algorítmicos en contextos sensibles como justicia penal o decisiones bancarias [31], [32]. Además su interpretabilidad mediante análisis de importancia de características permite comunicar los factores que más contribuyen a la clasificación de un caso como potencialmente discriminatorio. Adicionalmente se agregó el estado (geográfico) como variable categórica en las iteraciones con la población nacional como una forma de capturar sesgos geográficos como los observados en el emparejamiento por PSM.

Se aplicó validación cruzada estratificada de 5 pliegues con `GridSearchCV`, optimizando el `ROC-AUC`. El espacio de hiperparámetros incluyó `n_estimators` (100, 200), `max_depth` (4, 6, 8), `learning_rate` (0.01, 0.05, 0.1), y tasas de muestreo.

En este caso se presentan los resultados de dos de las iteraciones realizadas por medio de un XGBoost, ambos modelos balanceados a través de *downsampling*, uno a nivel nacional y otro a nivel del estado de Florida (ver tabla A.1).

Tabla A.1. Comparación de métricas entre modelos XGBoost con downsampling

Métrica	**XGBoost Nacional**	**XGBoost Estatal (FL)**
Accuracy	0.5184	0.5149
ROC-AUC	0.5313	0.5262
F1 Score	0.5286	0.5165
Precision	0.5176	0.5146
Recall	0.5400	0.5184
Log Loss	0.6919	0.6921
Matthews Corr. Coef.	0.0369	0.0298
Brier Score	0.2494	0.2495
Cohen's Kappa	0.0368	0.0298

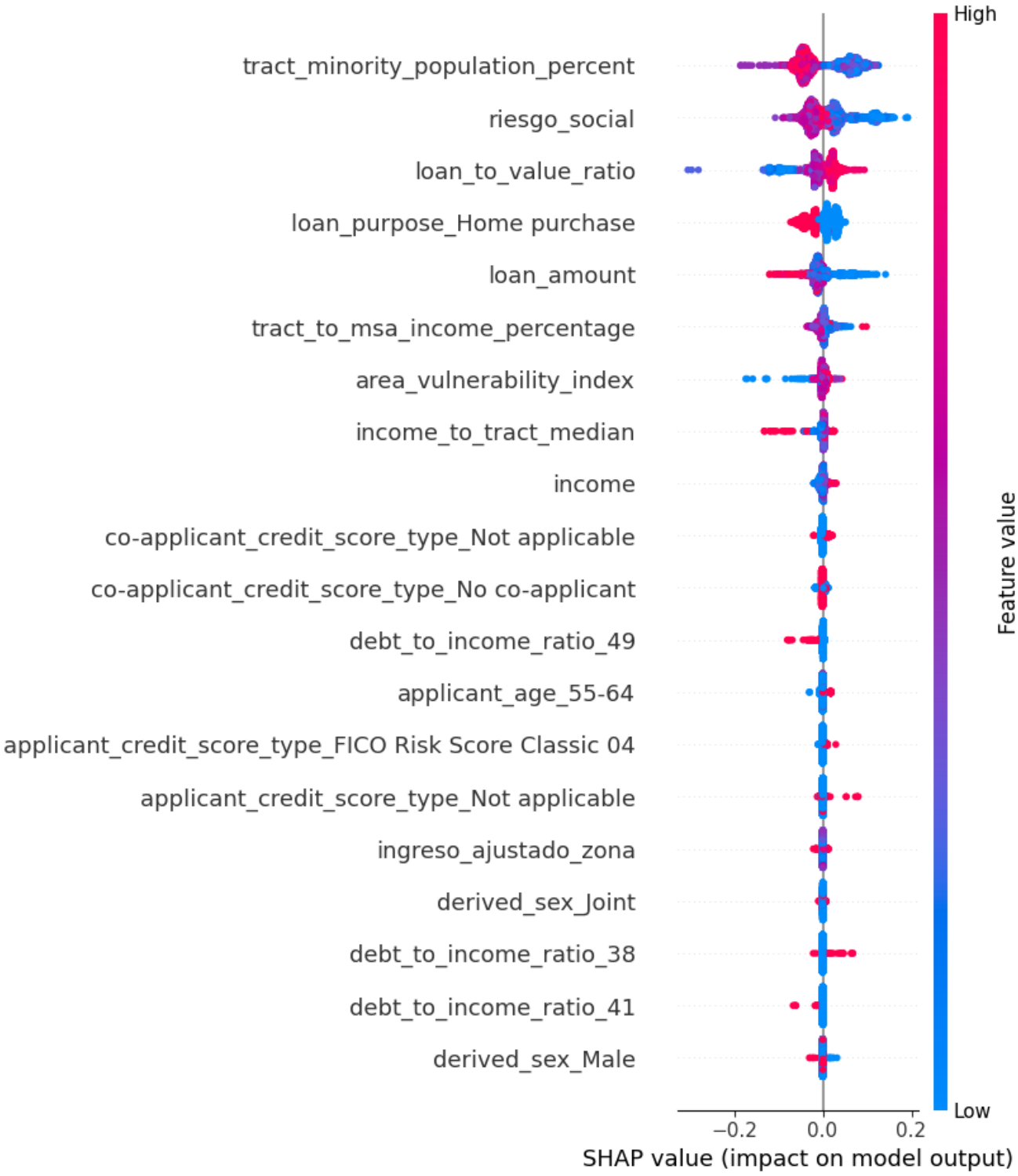

Fig. A.1: SHAP XGBoost FL Downsampled

Discusión de resultados

Los modelos muestran desempeño apenas superior al azar. La variable más relevante en ambos fue `tract_minority_population_percent` , lo que sugiere un fuerte componente geoespacial en las predicciones discriminatorias. Esto apoya la hipótesis de un sesgo estructural no explicado del todo por las variables observadas. Los modelos XGBoost, tanto en su versión nacional como estatal (Florida), presentan métricas de desempeño apenas superiores al azar, con valores de ROC-AUC de 0.5313 y 0.5262 respectivamente (Tabla A.1). Esta baja capacidad predictiva se refleja también en otras métricas clave como el F1 Score (0.5286 y 0.5165) y Cohen's Kappa (0.0368 y 0.0298), lo que sugiere una limitada generalización en la clasificación de los casos marcados como discriminatorios. Estos resultados pueden interpretarse desde dos posibles perspectivas no excluyentes:

- La discriminación no está completamente contenida en las variables observadas del dataset HMDA.
- La discriminación puede ser atribuida en gran medida al hecho mismo de ser latino, más allá de características objetivas como ingreso, crédito o ubicación, lo cual apunta a la existencia de un sesgo sistémico.

Este tipo de discriminación latente ha sido documentado previamente en estudios sobre sesgo algorítmico. En [10] se destaca que los modelos algorítmicos pueden reproducir y reforzar sesgos sociales existentes incluso si las variables explícitamente sensibles (como etnia) no son utilizadas directamente [10]. Asimismo, en [20] se advierte que incluso métricas aparentemente justas pueden enmascarar discriminación sistemática cuando el comportamiento del sistema difiere estructuralmente entre grupos [20].

Un hallazgo consistente en ambos modelos es que la variable con mayor capacidad explicativa según el análisis SHAP (ver figuras A.1, A.2) es tract_minority_population_percent, que mide el porcentaje de minorías étnicas presentes en el área censal del solicitante. Esta variable no solo ocupa el primer lugar en importancia relativa, sino que presenta un patrón claro: valores altos (zonas con más población minoritaria) tienden a aumentar la probabilidad de que el modelo prediga un caso

como discriminatorio (colores rojos hacia valores SHAP positivos en la Figura). Se reconoce que SHAP no considera dependencias entre variables; sin embargo, fue elegido por su amplia adopción y facilidad interpretativa. A futuro, proponemos complementar este enfoque con métodos como Knockoffs o Ghost Variables para mitigar problemas de colinealidad.

Este resultado no es casual ni neutro. De hecho, investigaciones previas han documentado que las zonas con alta concentración de minorías suelen recibir menos acceso a crédito, aun controlando por condiciones económicas [33]. Así, el alto peso predictivo de esta variable podría estar capturando una dimensión geoespacial del sesgo racial un efecto indirecto pero real de discriminación basada en etnia.

Dado que otras variables tradicionalmente asociadas al riesgo crediticio (como income, ltvr o loan_amount) tienen un impacto SHAP considerablemente menor y no logran generar modelos predictivos robustos, se refuerza la hipótesis de que el fenómeno de la discriminación no responde de forma determinista a las variables observables, sino que podría estar reflejando un sesgo estructural o institucional, difícil de capturar completamente con datos administrativos como los del HMDA.

Evaluación de otros modelos

Además de las múltiples iteraciones que se hicieron utilizando modelación a través de XGBoost, con el afán de evitar aislar los posibles resultados del experimento a un modelo específico se presentan algunos resultados obtenidos con diferentes modelos sobre los cuales también se realizaron numerosas iteraciones utilizando la librería pycaret (ver tablas A.2, A.3, A.4).

Tabla A.2. Comparación de modelos a nivel nacional con downsampling

Modelo	**Accuracy**	**ROC-AUC**	**F1 Score**	**Cohen's Kappa**
Gradient Boosting Classifier	0.5344	0.5423	0.5386	0.0495
Ada Boost Classifier	0.5307	0.5415	0.5295	0.0538
Logistic Regression	0.5250	0.5448	0.5186	0.0549
Ridge Classifier	0.5245	0.5450	0.5167	0.0566
Linear Discriminant Analysis	0.5236	0.5451	0.5163	0.0541
LightGBM	0.5230	0.5376	0.5194	0.0461
Extra Trees Classifier	0.5108	0.5267	0.5042	0.0314
Decision Tree Classifier	0.5076	0.5052	0.5096	0.0087
Random Forest Classifier	0.5072	0.5263	0.4984	0.0298
K Neighbors Classifier	0.5065	0.5107	0.5039	0.0161
SVM - Linear Kernel	0.4995	0.5125	0.4909	0.0166
Quadratic Discriminant Analysis	0.4765	0.4991	0.4558	0.0025
Naive Bayes	0.4165	0.5245	0.3428	0.0084
Dummy Classifier	0.2277	0.5000	0.0000	0.0000

Tabla A.3. Comparación de modelos a nivel nacional con SMOTE

Modelo	**Accuracy**	**ROC-AUC**	**F1 Score**	**Cohen's Kappa**
Gradient Boosting Classifier	0.7723	0.5432	1.0000	0.0000
LightGBM	0.7722	0.5431	0.9998	0.0079
Ada Boost Classifier	0.7714	0.5121	0.9984	-0.0005
SVM - Linear Kernel	0.7691	0.5242	0.9935	0.0051
Random Forest Classifier	0.7677	0.5259	0.9898	0.0160
Extra Trees Classifier	0.7465	0.5213	0.9500	0.0124
Logistic Regression	0.7413	0.5338	0.9370	0.0249
Ridge Classifier	0.7318	0.5317	0.9188	0.0248
Linear Discriminant Analysis	0.7318	0.5317	0.9188	0.0248
K Neighbors Classifier	0.6432	0.5147	0.7574	0.0130
Decision Tree Classifier	0.6370	0.5023	0.7496	0.0044
Quadratic Discriminant Analysis	0.5274	0.5007	0.5514	-0.0031
Naive Bayes	0.2406	0.5111	0.0232	0.0000
Dummy Classifier	0.2277	0.5000	0.0000	0.0000

Tabla A.4. Comparación de modelos con downsampling en Texas

Modelo	**Accuracy**	**ROC-AUC**	**F1 Score**	**Cohen's Kappa**
Gradient Boosting Classifier	0.5202	0.5109	0.6372	0.0069
Logistic Regression	0.5187	0.5153	0.6333	0.0169
LightGBM	0.5168	0.5151	0.6322	0.0118
Linear Discriminant Analysis	0.5140	0.5170	0.6288	0.0110
Ada Boost Classifier	0.5134	0.5017	0.6296	0.0040
Ridge Classifier	0.5090	0.5154	0.6234	0.0074
K Neighbors Classifier	0.5087	0.5080	0.6230	0.0080
SVM - Linear Kernel	0.5086	0.5074	0.6199	0.0107
Extra Trees Classifier	0.5074	0.5148	0.6193	0.0162
Decision Tree Classifier	0.5065	0.5096	0.6190	0.0126
Random Forest Classifier	0.5064	0.5223	0.6172	0.0193
Quadratic Discriminant Analysis	0.4388	0.4875	0.4996	-0.0152
Naive Bayes	0.3203	0.5034	0.2121	0.0002
Dummy Classifier	0.2047	0.5000	0.0000	0.0000

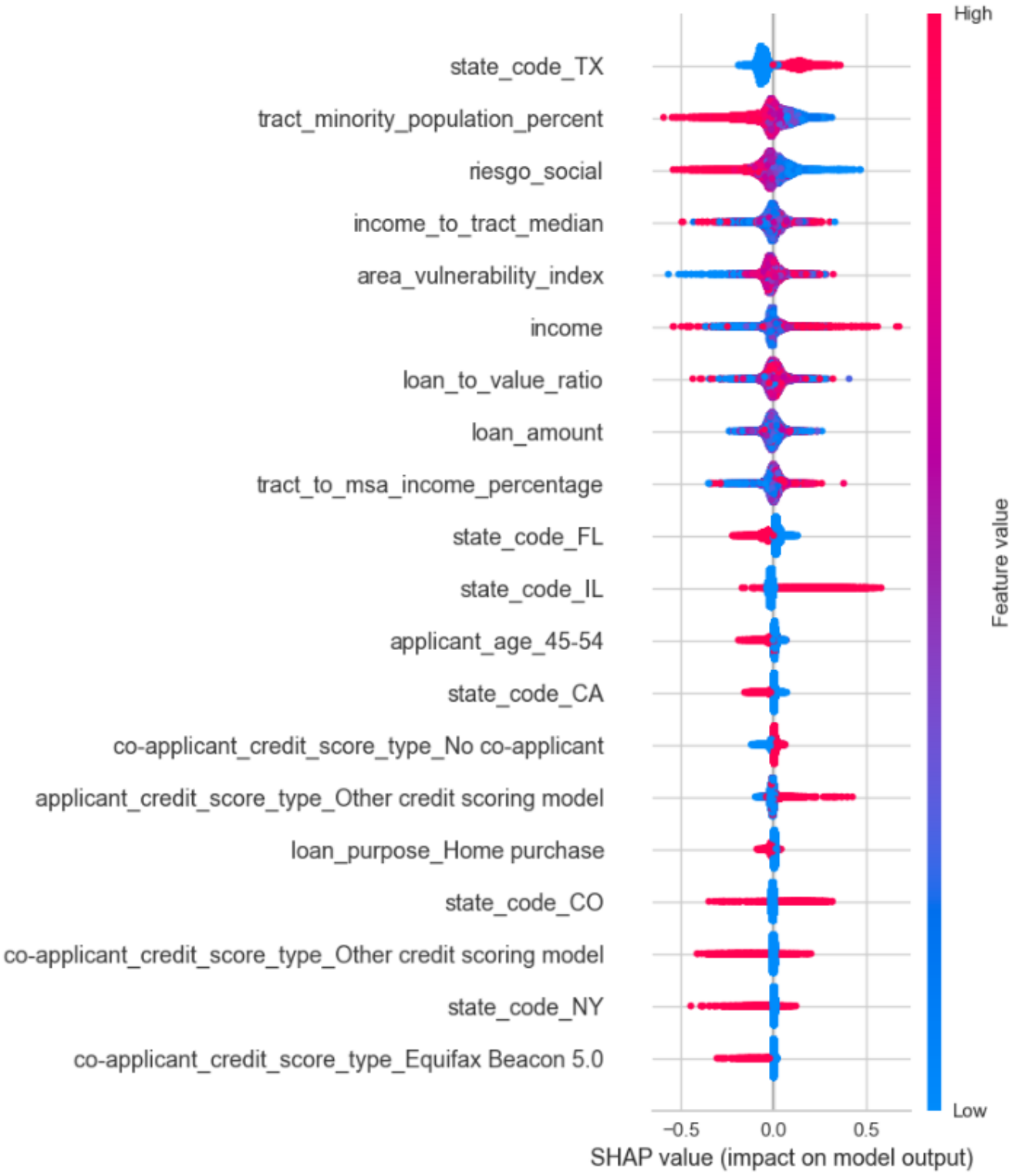

Fig. A.2: SHAP GBC Nacional Downsample