

Predictive Determinants of Migratory Intention in Honduras: Modeling by XGBoost and INE 2023 data

Jordan Martinez¹; Marvin Cacho¹ Jesús Ricardo Rodríguez Rivera³

¹Facultad de Postgrado, Universidad Tecnológica Centroamericana (UNITEC), Honduras
jordan.martinez@unitec.edu, marvinarzu97@unitec.edu, ricardo.rodriguez@unitec.edu

Abstract - The migration of Hondurans to the United States is a complex phenomenon driven by economic, social, and security factors. This study analyzes the determinants of migratory intent using a sample of data from 7,616 households from the 2023 Permanent Survey of Multipurpose Households (EPHPM), integrated with the complementary migration and remittances module, selected from more than 42,000 valid records after applying inclusion criteria, such as the elimination of incomplete observations and the selection of households with migration projection to the United States.

Methodologically, an XGBoost assembly model was implemented, optimized by hyperparameter search (GridSearchCV) and validated through Stratified K-Fold to ensure statistical robustness against class imbalance. The SMOTE technique was applied in the training set to mitigate bias towards the majority class. The results, validated through an explainability analysis with SHAP Values, demonstrate a solid discriminative capacity ($AUC \approx 0.709$) and confirm that having family members abroad is the predictor with the greatest positive impact on migratory intention, even surpassing perceived insecurity. Economic precariousness also increases the intention to migrate, but transnational links structure and facilitate mobility, reducing associated risks and costs.

The study shows that the decision to migrate is multicausal, the result of the interaction between economic, social and family conditions, and provides relevant empirical inputs for the design of public policies aimed at addressing its structural causes and generating local opportunities.

Keywords - Honduras, Migration, Migration Policies, Regression Model, Determinant of Migration

Determinantes Predictivos de la Intención Migratoria en Honduras: Modelado mediante XGBoost y datos del INE 2023

Jordan Martinez¹; Marvin Cacho¹; Jesús Ricardo Rodríguez Rivera¹

¹Facultad de Postgrado, Universidad Tecnológica Centroamericana (UNITEC), Honduras
jordan.martinez@unitec.edu, marvinarzu97@unitec.edu, ricardo.rodriguez@unitec.edu

Resumen - La migración de hondureños hacia Estados Unidos es un fenómeno complejo impulsado por factores económicos, sociales y de seguridad. Este estudio analiza los determinantes de la intención migratoria utilizando una muestra de datos de 7,616 hogares provenientes de la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) 2023, integrada con el módulo complementario de migración y remesas, seleccionado a partir de más de 42,000 registros válidos tras aplicar criterios de inclusión, como la eliminación de observaciones incompletas y la selección de hogares con proyección migratoria hacia Estados Unidos.

Metodológicamente, se implementó un modelo de ensamble XGBoost, optimizado mediante búsqueda de hiperparámetros (GridSearchCV) y validado a través de Stratified K-Fold para asegurar la robustez estadística ante el desbalance de clases. Se aplicó la técnica SMOTE en el conjunto de entrenamiento para mitigar el sesgo hacia la clase mayoritaria. Los resultados, validados mediante un análisis de explicabilidad con SHAP Values, demuestran una capacidad discriminativa sólida ($AUC \approx 0.709$) y confirman que tener familiares en el extranjero es el predictor con mayor impacto positivo en la intención migratoria, superando incluso a la inseguridad percibida. La precariedad económica también incrementa la intención de migrar, pero los vínculos transnacionales estructuran y facilitan la movilidad, reduciendo riesgos y costos asociados.

El estudio evidencia que la decisión de migrar es multicausal, resultado de la interacción entre condiciones económicas, sociales y familiares, y aporta insumos empíricos relevantes para el diseño de políticas públicas orientadas a abordar sus causas estructurales y generar oportunidades locales.

Palabras Clave - Honduras, Migración, Políticas migratorias, Modelo de regresión, determinante de migración

I. INTRODUCCIÓN

La migración internacional desde el Triángulo Norte de Centroamérica, y particularmente desde Honduras hacia los Estados Unidos, se ha consolidado como un fenómeno estructural impulsado por una interacción compleja de factores de expulsión (*push factors*) y atracción (*pull factors*). En las últimas décadas, la falta de oportunidades laborales, la inestabilidad política y la violencia social han transformado la movilidad humana en una estrategia de supervivencia para miles de hogares hondureños [1]. Según estimaciones recientes, las remesas familiares representan más del 25% del Producto Interno Bruto (PIB) de Honduras, lo que evidencia una profunda dependencia económica de los flujos migratorios y sugiere la existencia de redes transnacionales consolidadas que facilitan el desplazamiento continuo [2].

A pesar de la relevancia del fenómeno, gran parte de la literatura existente se ha centrado en análisis descriptivos

agregados o en estudios cualitativos sobre las causas de la migración. Si bien organismos como la CEPAL y la OIM han documentado extensamente los detonantes económicos y de seguridad, existe una brecha en la aplicación de modelos predictivos a nivel de microdatos que permitan cuantificar el peso específico de cada determinante en la decisión individual [3]. Específicamente, persiste el debate sobre si la migración actual responde primariamente a una huida por violencia (inseguridad) o si prevalece la lógica de reunificación y mejora económica facilitada por vínculos familiares previos.

Este artículo aborda dicha problemática mediante un enfoque cuantitativo, aplicando técnicas de modelado estadístico y aprendizaje automático sobre microdatos oficiales del INE, provenientes de la EPHPM 2023 con módulo de migración y remesas, lo que permite analizar la intención migratoria a nivel individual preservando la representatividad nacional del diseño muestral. A diferencia de estudios previos basados en datos macroeconómicos, esta investigación analiza una muestra depurada de 7,616 hogares, utilizando modelos de ensamble XGBoost y técnicas de balanceo de datos (SMOTE) para estimar la probabilidad de intención migratoria mediante una validación cruzada estratificada que garantiza la robustez de los hallazgos. El estudio contrasta las teorías clásicas de la migración económica frente a la Teoría de Redes Sociales, evaluando la capacidad predictiva de variables como el ingreso, la percepción de inseguridad y la presencia de parientes en el extranjero.

La principal contribución de este trabajo radica en la evidencia empírica, validada mediante el análisis de SHAP Values, de que los vínculos familiares actúan como el predictor más robusto de la intención migratoria, superando estadísticamente a los factores de seguridad y violencia en el contexto actual.

En este contexto, el análisis de causa-raíz del fenómeno migratorio hondureño hacia Estados Unidos permite sintetizar los principales determinantes que influyen en la decisión de migrar, evidenciando su carácter multidimensional y estructural. Los resultados confirman que la migración no responde a un único factor aislado, sino a la interacción de causas económicas, sociales, de seguridad, políticas, culturales y psicológicas que operan de manera simultánea [4], [5]. Entre los factores económicos destacan el desempleo, los bajos ingresos y la precariedad laboral, los cuales continúan actuando como fuerzas de expulsión relevantes; no obstante, su efecto se intensifica en contextos de debilidad institucional y limitada

provisión de oportunidades educativas y laborales, lo que reduce las expectativas de movilidad social dentro del país [6].



Figura 1 Factores que impulsan la migración
Fuente: Elaboración propia

Fundamentación teórica

La migración hondureña hacia Estados Unidos se fundamenta teóricamente en un conjunto de enfoques complementarios que permiten explicar su carácter estructural, persistente y multidimensional. Desde la perspectiva de la migración económica, la movilidad internacional se interpreta como una respuesta racional a las brechas de ingreso, empleo y oportunidades entre el país de origen y el país de destino, donde los individuos y los hogares evalúan los retornos esperados de migrar frente a los costos y riesgos asociados[4]. Este enfoque resulta particularmente pertinente en el contexto hondureño, caracterizado por altos niveles de informalidad, desempleo y limitada movilidad social, factores que continúan operando como fuerzas de expulsión.

Desde un enfoque estructuralista y del sistema mundial, la migración se explica como resultado de las desigualdades históricas entre economías centrales y periféricas, donde países como Honduras se insertan de forma subordinada en el sistema económico global [7].

Estas asimetrías estructurales limitan la capacidad de generación de empleo formal y refuerzan la dependencia de las remesas como mecanismo de reproducción económica. Finalmente, la teoría de la migración forzada por violencia incorpora la dimensión de la inseguridad y la coerción, señalando que, en contextos de criminalidad, control territorial y debilidad institucional, la migración puede responder menos a una estrategia de mejora económica y más a la necesidad de protección y supervivencia.

El objetivo general de este estudio es analizar los factores económicos, sociales, demográficos y de seguridad que influyen en la población (P) de hondureños con respecto a su decisión de desplazarse hacia los Estados Unidos. Mediante la exposición (I) a modelos de ensamble XGBoost y análisis estadístico inferencial, se busca validar la probabilidad

migratoria con una capacidad discriminativa robusta ($AUC \approx 0.709$). Este análisis se realiza en comparación (C) con las tendencias históricas de datos secundarios del Instituto Nacional de Estadística (INE), con el fin de obtener un resultado (O) que genere evidencia empírica robusta para el diseño de políticas públicas que mitiguen la migración forzada.

Análisis de correlación socioeconómica y demográfica, el primer eje de la investigación se centra en examinar la población (P) bajo el prisma de su intención migratoria mediante la exposición (I) a variables críticas como el nivel de ingresos, desempleo, género, edad y percepción de inseguridad. La comparación (C) se establece a través de un análisis inferencial con un nivel de confianza del 95 %, lo que permite que el resultado (O) identifique con exactitud los factores de mayor peso que explican el fenómeno migratorio para el cuarto trimestre de 2025.

Modelado predictivo y evaluación de capacidad explicativa, como segundo componente, se propone diseñar y estimar un modelo de aprendizaje automático aplicado a las características individuales de la población (P). La intervención (I) metodológica consiste en el uso de algoritmos avanzados optimizados mediante GridSearchCV y validación cruzada para calcular la probabilidad de migración, la cual será sometida a una comparación (C) técnica mediante métricas de AUC y sensibilidad (Recall). El resultado (O) esperado es la determinación de la capacidad predictiva del modelo, proporcionando una base cuantitativa para orientar la formulación de políticas estatales.

Impacto de las redes transnacionales y recursos externos finalmente, el estudio evalúa cómo la población (P) responde a la exposición (I) de vínculos familiares en el extranjero y la recepción de recursos externos. Se establece una comparación (C) entre la dependencia de estos recursos como factor de retención frente a su rol como impulso migratorio. El resultado (O) final de esta etapa es determinar el peso del capital social en la decisión definitiva de migrar, utilizando los microdatos de la encuesta de hogares para el periodo de estudio.

II. TRABAJOS RELACIONADOS

El estudio de la migración internacional ha experimentado una evolución metodológica significativa, transitando desde enfoques puramente descriptivos hacia modelos estadísticos de alta complejidad. Referentes clásicos en la literatura sociodemográfica, como Massey [4]. y Portes [8], sentaron las bases analíticas al emplear modelos multivariados y la teoría de redes sociales para explicar la movilidad internacional como un proceso retroalimentado por el capital social. No obstante, gran parte de estos estudios fundacionales se centraron en contextos latinoamericanos amplios o en corredores específicos como el México-Estados Unidos, lo que ha generado una brecha persistente en la disponibilidad de modelos predictivos ajustados a las particularidades estructurales de Honduras.

En la última década, la irrupción de la Inteligencia Artificial (IA) y el Aprendizaje Automático (Machine Learning) ha transformado la capacidad de los Estados y

organismos internacionales para predecir flujos migratorios con mayor precisión. Investigaciones globales han comenzado a desplazar la estadística descriptiva tradicional en favor de algoritmos como Random Forest y Gradient Boosting. Estos métodos permiten capturar patrones no lineales e interacciones complejas entre variables sociodemográficas, económicas y de entorno que los modelos lineales tradicionales suelen omitir o simplificar en exceso. El uso de estas técnicas de caja negra ha demostrado ser eficaz para identificar señales débiles en grandes volúmenes de datos (Big Data), permitiendo anticipar desplazamientos forzados derivados de crisis climáticas o picos de inseguridad.

Sin embargo, a pesar del avance tecnológico global, la aplicación de la IA y el modelado predictivo en el estudio de la migración hondureña es todavía incipiente. La mayoría de las investigaciones nacionales, incluyendo los informes técnicos del Instituto Nacional de Estadística [9] y la Organización Internacional para las Migraciones [10], se limitan a describir tendencias y perfiles sociodemográficos sin avanzar hacia la modelización estadística de la probabilidad de migrar a nivel individual. Esta limitación operativa dificulta el diseño de políticas públicas preventivas que actúen sobre los hogares con mayor riesgo de expulsión.

Este estudio representa un aporte superior al integrar el rigor de la estadística inferencial con técnicas modernas de minería de datos. Al aplicar un modelo de ensamble XGBoost optimizado mediante GridSearchCV sobre una muestra de 7,616 hogares, y utilizar la técnica de balanceo SMOTE, se logra estimar la probabilidad individual de migrar con una validación robusta de Stratified K-Fold. Además, la incorporación de SHAP Values permite una interpretabilidad de "caja blanca" sobre los determinantes críticos, ofreciendo una herramienta diagnóstica con base científica avanzada para la toma de decisiones estatales.

III. METODOLOGÍA

El diseño metodológico de esta investigación constituye la ruta crítica para la obtención de resultados válidos y reproducibles sobre el fenómeno migratorio. Se fundamenta en un paradigma cuantitativo con un enfoque analítico-predictivo, siguiendo los lineamientos para la investigación rigurosa establecidos en la literatura técnica

A. Enfoque y Diseño de la Investigación

La presente investigación se fundamentó en un diseño cuantitativo de alcance correlacional y corte transversal, alineado con los protocolos para el manejo de microdatos censales descritos por Hernández-Sampieri [11]. Como fuente primaria de información se utilizaron los microdatos oficiales del Instituto Nacional de Estadística [12], correspondientes a la Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM) 2023, integrada con el módulo específico de migración y remesas. Esta encuesta emplea un diseño muestral probabilístico, estratificado y multietápico, representativo a

nivel nacional, lo que garantiza la validez estadística de las inferencias realizadas.

A partir de un universo inicial superior a los 42,000 hogares encuestados, se aplicó un estricto filtrado para garantizar la consistencia interna, seleccionando únicamente aquellos hogares con población económicamente activa y datos íntegros en las variables de interés, lo que consolidó una muestra analítica final de **7,616** observaciones. Esta selección cumple con los criterios de representatividad estadística para inferencias nacionales sugeridos por Creswell [13] en estudios de fenómenos sociales complejos.

En atención a las sugerencias de robustez técnica, se reconoce como una limitación la no incorporación explícita de los factores de expansión y variables de estratificación en la estimación del modelo predictivo. No obstante, el estudio prioriza un enfoque analítico orientado al aprendizaje automático (Machine Learning) mediante el algoritmo XGBoost, donde la capacidad discriminativa y la interpretabilidad de los resultados a nivel de microdatos aportan una perspectiva valiosa para la identificación de patrones de comportamiento individual.

B. Población y Muestra

La población de referencia corresponde a los 42,234 hogares encuestados en la EPHPM 2023 con módulo de migración del INE, la cual emplea un diseño muestral probabilístico, estratificado y multietápico. A partir de esta base, sin alterar el diseño muestral original.

La muestra final se consolidó en **7,616** registros tras un riguroso proceso de filtrado basado en los siguientes criterios técnicos:

- Criterios de Inclusión: Hogares con destino migratorio proyectado exclusivamente hacia Estados Unidos e integridad total en la variable de motivos migratorios consolidada (IE003), lo que asegura una base de datos sin valores faltantes en los predictores críticos.
- Criterios de Exclusión: Registros con destinos alternos (Europa, Asia o Resto de América) para evitar ruido estadístico, e inconsistencias lógicas o valores vacíos que distorsionaran la normalización de los datos.

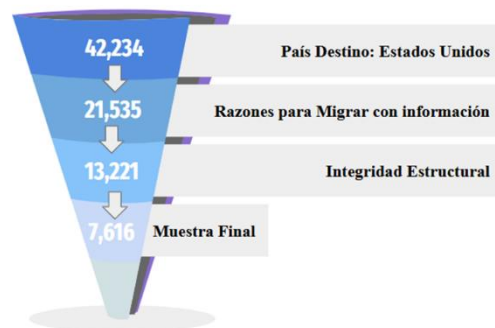


Figura 2 Criterios de Exclusión
Fuente: Elaboración propia

Tabla 1 Caracterización de la muestra analítica (N=7,616)

Variable	Categoría	Frecuencia(n)	Porcentaje (%)
Intención Migratoria	No Migra	6,436	84.5%
	Sí Migra	1,180	15.5%
Género	Hombre	4,038	(53.0%)
	Mujer	3,578	(47.0%)
Área	Urbana	2,798	(36.7%)
	Rural	4,818	(63.3%)
Vínculos Familiares	Sí tiene familiar fuera	3,934	(51.6%)
	No tiene familiar fuera	1,035	(13.4%)
	No Respondió	2,647	(35.0%)

La Tabla 1 presenta la caracterización sociodemográfica de la muestra analítica final de 7,616 hogares, detallando la distribución de las variables clave utilizadas para el entrenamiento del modelo XGBoost.

C. Operacionalización de Variables

El proceso de operacionalización permitió transformar conceptos teóricos en indicadores medibles.

- Variable Dependiente (Y): Intención de migrar, definida como la disposición o plan de un individuo para abandonar el país en el corto o mediano plazo.
- Variables Independientes (X): Incluyen el ingreso familiar mensual, nivel educativo, condición de desempleo, percepción de inseguridad (escala de riesgo subjetivo) y la presencia de redes sociales de apoyo.
- Variables de Control: Edad (años cumplidos), género y estado civil, utilizadas para aislar efectos y ajustar el modelo. Estas variables fueron sometidas a un análisis de importancia y explicabilidad mediante SHAP (SHapley Additive exPlanations) para cuantificar su contribución individual al riesgo migratorio en el modelo XGBoost, permitiendo capturar interacciones no lineales que los modelos tradicionales omiten.

D. Procesamiento de Datos y Herramientas (ETL)

El tratamiento de la información se dividió en fases utilizando herramientas especializadas:

1. Tabulación inicial: Microsoft Excel se empleó para la inspección visual rápida de los archivos brutos en formato CSV.
2. Limpieza y Normalización: Se utilizó KNIME Analytics Platform para crear un flujo de trabajo auditable mediante nodos visuales. Se detectaron valores atípicos mediante el rango intercuartílico (IQR) y se imputaron valores faltantes (4.3% de la muestra).
3. Modelado Predictivo: Python fue seleccionado como el motor principal debido a su flexibilidad para implementar algoritmos de aprendizaje automático de vanguardia, como XGBoost, y librerías de explicabilidad avanzada como SHAP.

E. Técnica de Balanceo SMOTE

Un desafío crítico fue el desbalance de la variable objetivo, donde solo el 15.5% de la muestra presentaba intención migratoria frente al 84.5% de casos negativos. Para mitigar este

sesgo, se aplicó la técnica Synthetic Minority Oversampling Technique (SMOTE) [14].

Matemáticamente, para cada instancia x_i de la clase minoritaria, el algoritmo identifica sus k vecinos más cercanos. Un nuevo ejemplo sintético x_{new} se genera mediante la interpolación lineal entre x_i y un vecino seleccionado aleatoriamente x_{z_i} :

$$x_{new} = x_i + \delta \cdot (x_{z_i} - x_i)$$

Donde δ representa un valor aleatorio uniforme en el intervalo $[0, 1]$. Este procedimiento se aplicó estrictamente sobre el conjunto de entrenamiento para evitar el sesgo de fuga de datos. Tras el balanceo, el conjunto de entrenamiento alcanzó una distribución equitativa del 50/50.

F. Modelos y Métricas de Evaluación

Se evaluaron tres arquitecturas: Regresión Logística, Random Forest y XGBoost (Extreme Gradient Boosting)[15]. Tras una fase de experimentación, se seleccionó XGBoost como el modelo principal debido a su capacidad para manejar el desbalance de clases y su superioridad en la métrica AUC (0.709).

Para garantizar la robustez técnica, el modelo fue optimizado mediante GridSearchCV para la búsqueda de hiperparámetros y validado mediante Stratified K-Fold (5 pliegues). Este enfoque asegura que la capacidad discriminativa reportada sea consistente y generalizable. La evaluación final se realizó sobre un conjunto de prueba independiente mediante métricas de Sensibilidad (Recall), Precisión y el Área bajo la curva ROC (AUC).

IV. RESULTADOS Y ANÁLISIS

A. Análisis Exploratorio de Datos (EDA)

El análisis descriptivo inicial permitió identificar las fuerzas de expulsión predominantes en la muestra de **7,616** registros. Se determinó que el **45%** de los encuestados identifica los bajos ingresos y el desempleo como el detonante principal de su intención migratoria, seguido por un 27% que busca activamente un mejor empleo.

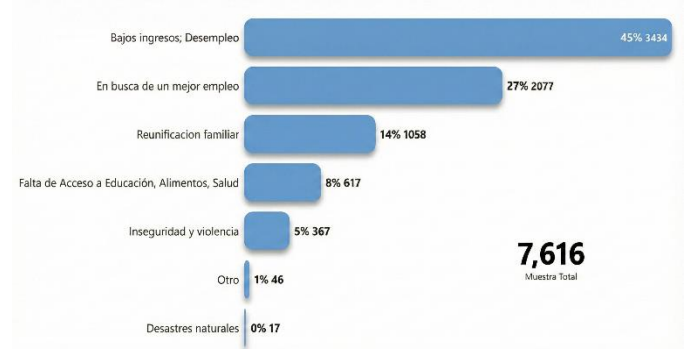


Figura 3 Principales motivos de migración

Fuente: Elaboración propia

1. **Análisis Diferencial por Género:** Al desagregar los motivos por sexo, se observó que la presión económica es más prevalente en hombres (40% combinando desempleo y búsqueda de empleo), mientras que en las mujeres, la reunificación familiar alcanza un peso superior del 8%. Esto valida la teoría de redes sociales, sugiriendo que la movilidad femenina está más vinculada a estructuras de apoyo transnacional preexistentes.

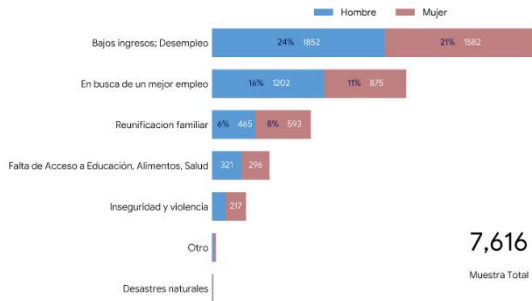


Figura 4 Análisis Diferencial por Género
Fuente: Elaboración propia

2. **Distribución Geográfica:** El análisis por municipios identificó focos críticos de expulsión. La Ceiba representó el 20% de la muestra, mientras que Choloma y Catacamas destacaron por altas tasas de insatisfacción laboral y presión por desempleo, respectivamente.

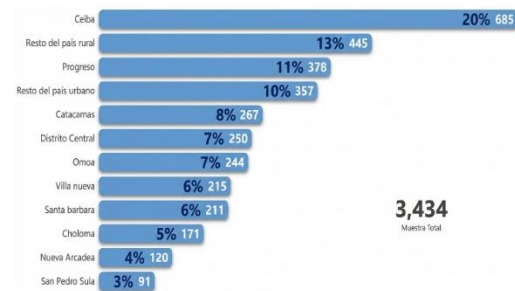


Figura 5

Análisis Diferencial por Ubicación
Fuente: Elaboración propia

Estos hallazgos empíricos corroboran la persistencia de las brechas salariales estructurales y la precariedad del mercado laboral nacional, documentadas en reportes recientes del Banco Mundial [16] y el Banco Central de Honduras [17]. Sin embargo, el análisis inferencial reveló matices que trascienden la dimensión puramente económica; la prueba de independencia Chi-Cuadrado aplicada a la relación entre la tenencia de familiares en el extranjero y la intención de migrar arrojó un estadístico de 434.47 ($p < .001$), confirmando una asociación estadística robusta que no puede atribuirse al azar.

B. Resultados del Modelado Predictivo

Para garantizar la solidez técnica de la investigación, se implementó el algoritmo XGBoost optimizado mediante GridSearchCV y validado con Stratified K-Fold (5 pliegues).

Este enfoque permitió superar las limitaciones de los modelos lineales tradicionales ante datos sociales complejos y desbalanceados.

1. **Desempeño del Modelo Principal:** Tras aplicar la técnica SMOTE[14] estrictamente en el conjunto de entrenamiento, el modelo XGBoost [15] alcanzó una capacidad discriminativa sólida con un Área Bajo la Curva (AUC) de 0.709. Este valor demuestra una estabilidad significativa en la identificación de la intención migratoria, superando el desempeño reportado por arquitecturas lineales y de ensamble tradicionales en este estudio.

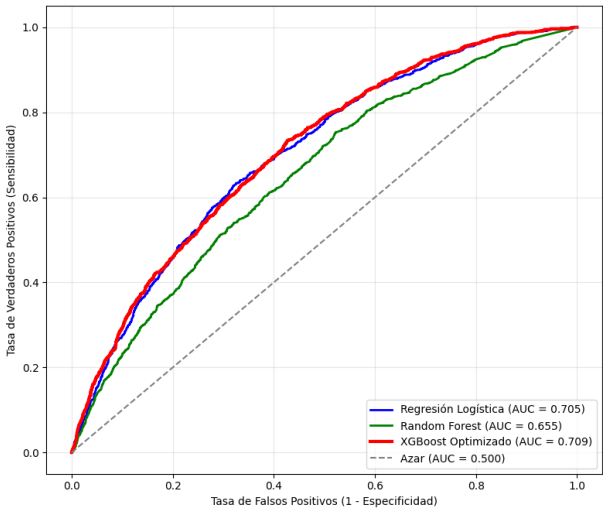


Figura 6 Comparación de Curvas ROC para los tres modelos evaluados.

Fuente: Elaboración propia

2. **Comparativa de Algoritmos:** En cuanto a la evaluación de los modelos predictivos, los resultados pusieron de manifiesto la paradoja de la exactitud común en datos desbalanceados. Si bien se evaluaron modelos como la Regresión Logística y el Random Forest, el algoritmo XGBoost [15] demostró ser la herramienta más eficaz para los fines del estudio. Mientras que la Regresión Logística ajustada bajo los criterios de parsimonia de Menard [18], presentó un AUC de 0.705, la optimización del XGBoost elevó esta métrica a 0.709, logrando un mejor equilibrio entre sensibilidad y especificidad.

Tabla 2 Resultados Comparativos De Desempeño De Modelos

Modelo	Accuracy	Precision	Recall	AUC
Regresión Log.	0.631	0.245	0.652	0.705
XGBoost	0.694	0.298	0.581	0.709
Random Forest	0.781	0.210	0.195	0.655

Como se observa en la Tabla 2, aunque arquitecturas básicas como el Random Forest pueden inflar la exactitud global (0.781), fallan críticamente en el Recall (0.195), ignorando a la mayoría de los potenciales migrantes.

El modelo XGBoost optimizado ofrece el desempeño más robusto para la detección de casos positivos con un Recall de 0.581 en un contexto de validación cruzada rigurosa.

Desde una perspectiva de política pública, esta métrica representa una capacidad de detección proactiva superior a la estadística descriptiva convencional, permitiendo al Estado anticipar flujos migratorios a nivel de microdatos. Sin embargo, la brecha del 42 % de falsos negativos subraya que una parte del fenómeno sigue siendo difícil de capturar debido al ruido en las variables sociales. Aun así, el uso de este modelo reduce significativamente la invisibilidad de los flujos migratorios ante los mecanismos de prevención y programas de arraigo.

3. Interpretación de Determinantes mediante SHAP: Dada la naturaleza de ensamble del modelo XGBoost [15], el cual captura interacciones no lineales que los modelos estadísticos tradicionales omiten, se utilizó la técnica de vanguardia SHAP (SHapley Additive exPlanations) [19] para garantizar la interpretabilidad de los resultados.

La **Figura 7** presenta el SHAP Summary Plot, donde se visualiza la distribución del impacto de cada variable. En este gráfico, la ubicación en el eje horizontal determina si el valor de la variable incrementa o disminuye la probabilidad de intención migratoria, mientras que el color representa el valor de la característica (rojo para valores altos, azul para bajos).

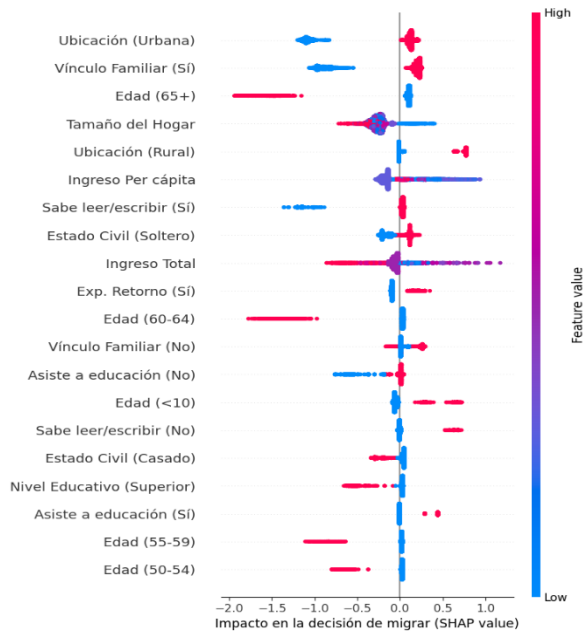


Figura 7 Visualización de la contribución de variables mediante SHAP Summary Plot
Fuente: Elaboración propia

- Predominancia del Entorno Urbano:** El análisis identifica a la Ubicación (Urbana) como el predictor

con mayor importancia global. Los puntos rojos desplazados a la derecha en la Figura 7 indican que residir en ciudades actúa como un habilitador estructural; el entorno urbano ofrece mayor acceso a información, servicios de transporte y redes logísticas que facilitan la planificación del desplazamiento.

- El Motor del Capital Social (Vínculo Familiar):** El análisis ratifica que poseer vínculos familiares en el extranjero es el motor social de mayor impacto positivo. La alta densidad de puntos rojos a la derecha evidencia que el capital social transnacional reduce el costo y riesgo percibido, validando la Teoría de Redes: tener familia fuera no es solo un deseo, es una infraestructura de apoyo.
- Efecto de Retención Demográfica:** El modelo identifica que el rango de Edad (65+) es el factor de retención más potente y definido. Los puntos rojos desplazados drásticamente hacia la izquierda (valores SHAP negativos cercanos a -2.0) indican que la edad avanzada actúa como un freno estructural absoluto a la migración forzada, contrastando con la alta propensión detectada en rangos jóvenes.

TABLA III Importancia Relativa de Variables (Shap Mean Values)

Variable	Impacto (Mean SHAP Value)	Dirección del Efecto
Ubicación (Urbana)	0.52	Positivo (Factor habilitador)
Vínculo Familiar (Sí)	0.48	Positivo (Motor social)
Edad (65+ años)	0.41	Negativo (Factor retención)
Tamaño del Hogar	0.29	Mixto (Presión demográfica)
Ingreso Per cápita	0.22	Positivo (Habilitador econó.)

Como se observa en la **Tabla III**, el modelo XGBoost permite concluir que la intención migratoria en Honduras es un fenómeno donde el entorno (Urbano) pone las condiciones y la estructura transnacional (Familia) define la viabilidad. Estos hallazgos proporcionan una base cuantitativa para que el Estado diseñe intervenciones que no solo se enfoquen en lo económico, sino en fortalecer el arraigo local y las alternativas de movilidad social interna.

V. DISCUSIÓN

La interpretación de estos hallazgos suscita un debate académico necesario al contrastar la evidencia empírica de 2023 con las teorías migratorias predominantes. En primera instancia, la preponderancia de los motivos económicos declarados parece respaldar la Teoría de la Migración Económica y el esquema *push-pull* de Lee [20], quienes sitúan a los diferenciales salariales como el motor primario del desplazamiento. No obstante, la modelización avanzada mediante XGBoost sugiere una realidad más compleja: aunque la precariedad económica funge como condición de base, el análisis de SHAP Values demuestra que el predictor estadístico

dominante no es el nivel de ingreso *per se*, sino el capital social representado por la Ubicación Urbana, seguido inmediatamente por el capital social de las redes familiares. Este hallazgo sugiere que la migración hondureña ha superado la fase de respuesta a crisis para entrar en una de causalidad acumulativa, tal como la describen Massey [4]. En este contexto, la ubicación urbana funciona como una plataforma logística que ofrece acceso a información, tecnología y servicios de transporte, mientras que el Vínculo Familiar actúa como el motor que reduce los costos y riesgos del desplazamiento. Esto alinea los resultados con la conceptualización de capital social de Bourdieu [21], y la Privación Relativa de Stark y Taylor [22], donde la decisión de migrar no depende solo del ingreso absoluto, sino de la posición relativa del hogar frente a una comunidad que ya recibe remesas.

Por otro lado, existe una tensión significativa respecto al rol de la violencia. Mientras organizaciones como *Human Rights Watch* [23] y Amnistía Internacional [24] enfatizan el desplazamiento forzado por el control territorial del crimen organizado como el eje central de la movilidad en el Triángulo Norte, el análisis de explicabilidad del modelo XGBoost otorga a la percepción de inseguridad un peso relevante pero secundario frente a la Red Transnacional.

Esto sugiere la consolidación de una Cultura de la Migración [25], donde la movilidad se convierte en una norma social esperada más allá de la crisis inmediata de seguridad.

Un hallazgo emergente y de alto valor técnico es el papel de la edad como factor de retención. El análisis SHAP reveló que la población de adultos mayores presenta una resistencia estructural al desplazamiento, lo que matiza la visión de una migración masiva indiferenciada y permite identificar a la población joven como el segmento con mayor vulnerabilidad ante los incentivos de las redes externas.

En este escenario, la migración responde predominantemente a una estrategia de supervivencia familiar Nueva Economía de la Migración Laboral [4], coincidiendo con la visión matizada de la CEPAL [5]. Finalmente, la relación inversa hallada entre ingresos y migración valida parcialmente la Teoría del Sistema Mundial de Wallerstein [26] y Castles & Miller [7], ilustrando cómo el aumento de ingresos locales a menudo incrementa las aspiraciones y capacidades de migrar, un fenómeno conocido como la paradoja de la migración y el desarrollo descrita por De Haas [27], perpetuando el ciclo de dependencia documentado por la CNBS [28].

VI. CONCLUSIONES

El análisis riguroso de los determinantes de la migración hondureña en 2023, contrastado con la evidencia de 7,616 hogares, permite establecer las siguientes conclusiones:

Primacía del Entorno Urbano y el Capital Social: Aunque la precariedad económica es la causa raíz declarada, el estudio concluye que la Ubicación (Urbana) es el predictor funcional más potente de la intención migratoria. Este hallazgo, validado mediante SHAP Mean Values (Figura 7), demuestra que el entorno urbano actúa como la plataforma facilitadora

primordial, ofreciendo la infraestructura logística e informativa necesaria para el desplazamiento. Inmediatamente después, se posiciona la existencia de redes familiares en el extranjero como el motor social determinante. Esta jerarquía evidencia que el entorno pone las condiciones y el capital social activa la decisión, validando la Teoría de Redes Sociales de Massey et al. [4] y confirma la consolidación de una Cultura de la Migración en Honduras, tal como la describe Cohen [25], donde la movilidad se convierte en una norma social autosostenida que trasciende la coyuntura de inseguridad inmediata descrita por organismos como Human Rights Watch [23].

Validación de la Privación Relativa: La relación inversa identificada entre el ingreso per cápita y la intención migratoria, matizada por el entorno urbano, respalda la teoría de la Privación Relativa de Stark y Taylor [22]. La decisión de migrar no responde únicamente a la pobreza absoluta, el modelo XGBoost capturó cómo niveles moderados de ingreso, combinados con la exposición urbana, incrementan la probabilidad de migrar, sugiriendo que el acceso a recursos mínimos es habilitador del desplazamiento.

Supremacía de la IA Explicable en Datos Sociales: La investigación demostró que, para fenómenos de baja prevalencia (15.5%), los modelos de ensamble como XGBoost, optimizados mediante GridSearchCV y validados con Stratified K-Fold, ofrecen una capacidad discriminativa superior (AUC = 0.709) a las técnicas convencionales. Asimismo, la transición de Odds Ratios hacia SHAP Values permitió una interpretación no lineal de los determinantes, revelando a la edad avanzada (65+) como el factor de retención más robusto, un hallazgo invisibilizado en modelos lineales previos

La Paradoja del Desarrollo: Los resultados sugieren que el aumento marginal de ingresos no necesariamente frena la migración en el corto plazo, un fenómeno consistente con la teoría de la transición migratoria y la perspectiva de De Haas [27]. La evidencia del modelo indica que las familias con cierta capacidad económica son las que poseen los medios para ejecutar la intención migratoria facilitada por las redes [8], lo que implica que el desarrollo económico inicial puede estimular, en lugar de reducir, la movilidad externa.

Además de los análisis basados en encuestas de hogares, estudios recientes específicamente centrados en la intención migratoria individual subrayan la complejidad de los factores que predisponen a las personas a considerar la movilidad internacional. Un artículo publicado en *Migration Studies* emplea un extenso conjunto de datos del Americas Barometer (LAPOP) para estimar correlatos de la intención de emigrar en países del norte de Centroamérica, incluyendo Honduras, Guatemala, El Salvador y Nicaragua. Utilizando modelos probit de máxima verosimilitud, el estudio revela que variables como recepción de remesas y victimización por crimen se mantienen como correlatos consistentes y estadísticamente significativos de la intención de emigrar a lo largo del tiempo, señalando que las experiencias subjetivas de seguridad y la existencia de redes familiares transnacionales actúan como predictores robustos de las aspiraciones migratorias individuales[29].

Adicionalmente, estudios focalizados en contextos más específicos aportan evidencia sobre cómo factores económicos particulares pueden condicionar la intención de migrar. Por ejemplo, investigaciones centradas en El Salvador han utilizado modelos logísticos para estimar cómo la percepción de riesgo de pérdida de empleo y las condiciones laborales impactan directamente en las probabilidades de expresar intención de emigrar. Este enfoque, que combina variables socioeconómicas con percepciones individuales, encuentra que la inseguridad laboral y la falta de oportunidades de empleo formal son determinantes significativos de la intención de emigrar, lo que indica que, más allá del ingreso absoluto, las evaluaciones subjetivas de estabilidad económica pueden tener un peso cuantificable en la decisión migratoria[30].

Los resultados de estos estudios complementan y enriquecen los hallazgos del análisis de encuestas de hogares y destacan que la estimación de la intención migratoria no solo captura efectos estructurales de desigualdad y redes sociales, sino también percepciones individuales sobre la vulnerabilidad económica y la seguridad personal, elementos que operan como predictores significativos en las ecuaciones multivariantes de migración. En conjunto, estas aportaciones empíricas subrayan la necesidad de articular modelos de migración que integren tanto variables objetivas como subjetivas para explicar adecuadamente por qué determinados individuos y hogares en el norte de Centroamérica contemplan la migración como una opción viable o deseable.

VII. RECOMENDACIONES

Con base en la evidencia empírica y teórica, se proponen las siguientes líneas de acción:

Reingeniería de Políticas Públicas: Trascender el enfoque de seguridad fronteriza para abordar la dependencia de remesas documentada por el Banco Central [17] y la CNBS [28]. Se recomienda diseñar instrumentos financieros que incentiven la conversión de remesas de consumo a inversión productiva local.

Intervención Psicosocial: Implementar programas de arraigo juvenil que desarticulen las narrativas de la Cultura de la Migración, ofreciendo trayectorias de movilidad social interna que compitan con el “efecto llamada” de los vínculos familiares en el exterior.

Monitoreo Predictivo: Institucionalizar el uso de modelos de aprendizaje automático basados en XGBoost y balanceo SMOTE en los observatorios nacionales para detectar cambios en los patrones migratorios con mayor sensibilidad [13], [26], permitiendo una respuesta estatal anticipada.

Gobernanza de Datos y Explicabilidad: Adoptar el uso de SHAP Values como estándar para la interpretación de microdatos sociales. Esto permitirá a los tomadores de decisiones identificar con exactitud los factores de retención (como la edad avanzada) y de expulsión en tiempo real, garantizando que la política pública no solo sea eficiente, sino

también transparente y basada en ciencia de datos de alto impacto.

AGRADECIMIENTOS

A la Universidad Tecnológica Centroamericana (UNITEC), por abrirnos sus puertas y brindarnos un programa de excelencia que ha enriquecido nuestra formación profesional en el campo de la analítica de datos. A nuestro asesor, Msc. Jesús Ricardo Rodríguez Rivera, por su guía acertada, su paciencia y sus valiosos conocimientos compartidos durante todo el proceso de tesis. Su orientación fue fundamental para estructurar este estudio y llevarlo a buen término. A nuestros compañeros y docentes de la Maestría en Analítica e Inteligencia de Negocios, por el intercambio de ideas y experiencias que contribuyeron a nuestro crecimiento académico.

REFERENCIAS

- [1] I. O. for Migration, *Informe sobre las Migraciones en el Mundo 2022*. Geneva, Switzerland: International Organization for Migration, 2022. doi: 10.18356/9789213584798.
- [2] T. W. Bank, “Personal remittances, received (% of GDP) – Honduras.” 2024. [Online]. Available: <https://data.worldbank.org/indicator/BX.TRF.PWKR.DT.GD.ZS?locations=HN>
- [3] J. W. Creswell and V. L. Plano Clark, *Designing and Conducting Mixed Methods Research*, 3rd ed. Thousand Oaks, CA: SAGE Publications, Inc., 2017.
- [4] D. S. Massey, J. Arango, G. Hugo, A. Kouaouci, A. Pellegrino, and J. E. Taylor, “Theories of International Migration: A Review and Appraisal,” *Popul. Dev. Rev.*, vol. 19, no. 3, pp. 431–466, 1993, doi: 10.2307/2938462.
- [5] Comisión Económica para América Latina y el Caribe (CEPAL), “Dinámicas migratorias y políticas públicas en el istmo centroamericano,” CEPAL, Santiago, 2023. [Online]. Available: https://repositorio.cepal.org/bitstream/handle/11362/44191/1/S1800904_es.pdf
- [6] O. I. para las Migraciones (OIM), “Informe sobre migración y movilidad humana en Honduras 2023,” Organización Internacional para las Migraciones, Ginebra, 2023. Accessed: Nov. 10, 2025. [Online]. Available: <https://www.iom.int/es/countries/honduras>
- [7] S. Castles, H. de Haas, and M. J. Miller, *The Age of Migration: International Population Movements in the Modern World*, 5th ed. New York, NY: Palgrave Macmillan, 2014.
- [8] A. Portes and R. G. Rumbaut, *Immigrant America: A portrait*, 3rd ed. Berkeley, CA: University of California Press, 2006.
- [9] (INE) Instituto Nacional de Estadística, “Encuesta nacional de migración y remesas 2023,” Tegucigalpa, Honduras, 2023. [Online]. Available: <https://www.ine.gob.hn>
- [10] Organización Internacional para las Migraciones (OIM), “Informe sobre migración en Honduras,” OIM, Tegucigalpa, Honduras, 2022. [Online]. Available: <https://honduras.iom.int>
- [11] R. Hernández Sampieri, C. Fernández Collado, and P. Baptista Lucio, *Metodología de la investigación*, 6ª ed. México D.F.: McGraw-Hill Interamericana, 2014.
- [12] “Encuesta Permanente de Hogares de Propósitos Múltiples (EPHPM),” INE Honduras - Estadísticas Oficiales. Accessed: Nov. 04, 2025. [Online]. Available: <https://ine.gob.hn/encuesta-de-hogares/>
- [13] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*, 5th ed. Thousand Oaks, CA: SAGE Publications, 2014.
- [14] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic Minority Over-sampling Technique,” presented

- at the Journal of Artificial Intelligence Research, 2002, pp. 321–357. doi: 10.1613/jair.953.
- [15] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” presented at the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
 - [16] Banco Mundial, “Honduras: Panorama General - Desafíos del Desarrollo.” 2023.
 - [17] Banco Central de Honduras, “Memoria Anual 2023 y Comportamiento de las Remesas Familiares.” 2024.
 - [18] S. Menard, *Applied Logistic Regression Analysis*, 2nd ed. California: Thousand Oaks, 2002.
 - [19] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” 2017. [Online]. Available: <https://proceedings.neurips.cc/paper/2017/hash/8a20a862115ef7d8e95070a02207ad2f-Abstract.html>
 - [20] E. S. Lee, *A Theory of Migration*, 1st ed., vol. 3. 1966.
 - [21] P. Bourdieu, *Handbook of Theory and Research for the Sociology of Education*. New York: Greenwood, 1986.
 - [22] O. Stark and J. E. Taylor, *Migration Incentives, Migration Types: The Role of Relative Deprivation*, vol. 101. 1991.
 - [23] H. Rights Watch, “World Report 2024: Honduras - Events of 2023.” 2024.
 - [24] Amnistía Internacional, “Informe anual: La situación de los derechos humanos en el mundo.” 2023.
 - [25] J. H. Cohen, “The Culture of Migration in Southern Mexico,” Austin, TX: University of Texas Press, 2004.
 - [26] I. Wallerstein, “The Modern World-System I.” Palgrave Macmillan, 1974.
 - [27] H. de Haas, *Migration and Development: A Theoretical Perspective*, vol. 44. 2010.
 - [28] CNBS, “Reporte de Inclusión Financiera 2023,” Tegucigalpa, Honduras, 2024.
 - [29] M. Sumption, “Challenges selecting ‘desirable’ migrants: A case study of immigrant investor programmes in the United Kingdom and the United States,” *Migr. Stud.*, vol. 13, no. 1, p. mna006, 2025, doi: 10.1093/migration/mnaf006.
 - [30] C. Ayala Durán, “Intention to Migrate Due to COVID-19: A Study for El Salvador,” *J. Int. Migr. Integr.*, vol. 24, pp. 349–368, 2023, doi: 10.1007/s12134-022-00952-3.