



ORIGINAL

Unsupervised algorithm to classify immigration risk levels

Algoritmo no supervisado para clasificar niveles de riesgo de inmigración

Miguel Valles-Coral¹  , Ulises Lazo-Bartra¹  , Lloy Pinedo¹  , Jorge Raul Navarro-Cabrera¹  , Luis Salazar-Ramírez¹  , Fernando Ruiz-Saavedra¹  , Pierre Vidaurre-Rojas¹  , Segundo Ramírez¹  

¹Universidad Nacional de San Martín. Tarapoto, Perú.

Citar como: Valles-Coral M, Lazo-Bartra U, Pinedo L, Navarro-Cabrera JR, Ruiz-Saavedra F, Ramirez S, et al. Algoritmo no supervisado para clasificar niveles de riesgo de inmigración. Data and Metadata 2023;2:98. <https://doi.org/10.56294/dm202398>.

Enviado: 04-08-2023

Revisado: 20-10-2023

Aceptado: 16-11-2023

Publicado: 17-11-2023

Editor: Prof. Dr. Javier González Argote 

ABSTRACT

Introduction: migration is a social phenomenon that affects the structure and distribution of the population, driven by the search for better opportunities and living conditions. In this regard, irregular migration poses a challenge for host countries, as it involves the entry of individuals without the appropriate documentation, potentially compromising national security and border control.

Objective: to evaluate the application of the unsupervised DBSCAN algorithm to classify foreigners based on the level of risk of irregular immigration at the National Migration Superintendence of Peru.

Methods: we use the DBSCAN algorithm on a dataset from the National Immigration Superintendence, classifying foreigners into clusters according to their level of risk of irregular immigration. In addition, we use the Silhouette, Davies-Bouldin, and Calinski-Harabasz coefficients to evaluate the quality of the classification.

Results: DBSCAN classified foreigners into four clusters based on the level of risk of irregular immigration: high, medium-high, medium-low, and low. The performance of the Silhouette index was 0,5338, the Davies-Bouldin index was 0,6213, and the Calinski-Harabasz index was 3680,2359.

Conclusions: We show that the use of DBSCAN in the National Immigration Superintendence effectively classified foreigners according to the level of risk of irregular immigration. This tool supports informed decisions of immigration inspectors, favoring Peruvian immigration regulation.

Keywords: Unsupervised Algorithm; Migration Quality; Migration Control; Immigration Inspectors; Migration.

RESUMEN

Introducción: la migración es un fenómeno social que afecta la estructura y distribución de la población, siendo motivada por la búsqueda de mejores oportunidades y condiciones de vida. En tal sentido, la migración irregular representa un desafío para los países receptores, dado que conlleva la entrada de individuos sin la documentación correspondiente, pudiendo comprometer la seguridad nacional y el control fronterizo de los países.

Objetivo: evaluar la aplicación del algoritmo no supervisado DBSCAN para clasificar a extranjeros según el nivel de riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones del Perú.

Métodos: empleamos el algoritmo DBSCAN sobre un dataset proveniente de la Superintendencia Nacional de Migraciones, clasificando a extranjeros en clústeres según su nivel de riesgo de inmigración irregular, además, usamos los coeficientes de Silhouette, Davies-Bouldin, y Calinski-Harabasz para evaluar la calidad de la clasificación.

Resultados: DBSCAN clasificó a los extranjeros en cuatro clústeres según el nivel de riesgo de inmigración irregular: alto, medio alto, medio bajo y bajo, donde el desempeño del índice Silhouette fue de 0,5338, el desempeño del índice Davies-Bouldin fue 0,6213 y el desempeño del índice Calinski-Harabasz fue 3680,2359.

Conclusiones: evidenciamos que el uso de DBSCAN en la Superintendencia Nacional de Migraciones clasificó

eficazmente a extranjeros según el nivel de riesgo de inmigración irregular, esta herramienta respalda decisiones informadas de inspectores migratorios, favoreciendo la regulación migratoria peruana.

Palabras clave: Algoritmo No Supervisado; Calidad Migratoria; Control Migratorio; Inspectores Migratorios; Migración.

INTRODUCCIÓN

La migración, como proceso demográfico, es un fenómeno social relacionado a la alteración de la estructura, aumento o distribución de una población, debido principalmente a la búsqueda de mejora de la calidad de vida y situación laboral.^(1,2) Se denomina migrante a la persona, sin importar su género y edad, que toma la decisión de trasladarse interna o externamente; emigrante cuando se traslada e inmigrante cuando se asienta en un espacio geográfico diferente a su residencia natal.^(3,4)

Según el Portal de Datos Mundiales sobre Migración, la estadística de las Naciones Unidas señala que hasta el 2020, el número de migrantes internacionales ascendió a 281 millones, siendo el continente europeo quien presentó mayor proporción (86,7 millones), seguido del continente asiático con 85,6 millones y norteamericano con 58,7 millones, Latinoamérica y El Caribe tuvieron 14,8 millones. Se estima además que el 15 % fueron menores de 20 años, 12 % de 65 años a más, y el 73 % entre 20 a 64 años tuvieron propósitos laborales. En el caso del Perú, al 2019, hubo en total 782 169 inmigrantes, procedentes en gran proporción de Venezuela, Estados Unidos y China.⁽⁵⁾

Aunque la migración se considera una oportunidad de superación y desarrollo cuando se trata de decisiones propias, la literatura documenta que casi siempre está asociada a situaciones forzadas ocasionadas por problemas de seguridad nacional, control de fronteras, entre otros factores.^(6,7,8) Es ahí cuando el ingreso de migrantes de condición irregular genera inseguridades en los países receptores, ya que involucra el ingreso de indocumentados, refugiados, asilo político, trata de personas, etc.^(9,10,11)

Para ingresar a un país en la calidad de inmigrante, el proceso involucra el cumplimiento de normas o protocolos establecidos por cada país, donde las solicitudes están respaldadas por propósitos concretos.⁽¹²⁾ Para el caso del contexto peruano, tomamos la Encuesta Dirigida a la Población Venezolana que Reside en el País (ENPOVE), donde las situaciones de inmigrante otorgadas fueron de Solicitante de Permiso Temporal de Permanencia (PTP) (50,2 %), Permiso Temporal de Permanencia (26,7 %), Visa de Turista (4,1 %), Carnet de Extranjería (4,0 %), Cédula de Identidad (3,9 %), Situación Irregular (3,2 %), etc. No obstante, el 94,7 % manifestaron el propósito de quedarse, siendo reducida la intención de regularización.⁽¹³⁾

La situación de inmigrantes irregulares, fue el problema identificado en el presente estudio, pues la permanencia en calidad migratoria irregular afecta el estatus del estado.⁽¹⁴⁾ En este sentido, la Superintendencia Nacional de Migraciones, organismo del Ministerio del Interior, vela por el control migratorio con el fin de garantizar la seguridad interna, sin embargo, los inspectores encargados de tomar decisiones para otorgar la calidad migratoria, suelen basarse en criterios subjetivos.^(15,16,17,18) Cabe precisar que ser un inmigrante sin riesgo de ser irregular en Perú, involucra cumplir con la situación migratoria otorgada, o buscar la regulación cuando amerite.

Ahora, se reconoce que, pese a la infraestructura tecnológica de soporte para las operaciones de control y regulación migratoria, los datos recopilados y almacenados históricamente son desaprovechados, ya que se desconoce información a partir de su procesamiento que clasifique al extranjero según sus rasgos o características pertinentes para obtener un estatus migratorio.⁽¹⁹⁾ Por tanto, las decisiones de los inspectores están sujetas a corto plazo e información básica de los postulantes (familiares residentes en Perú, propósito de ingreso, etc.),⁽²⁰⁾ reflejando sesgos en la asignación de permisos que, eventualmente, se conviertan en riesgos de inmigraciones irregulares.⁽²¹⁾

Las técnicas de minería de datos y el reconocimiento de patrones, aplicados mediante algoritmos supervisado y no supervisados, son ampliamente usados en este contexto.^(22,23,24,25) Sin embargo, en la Superintendencia Nacional de Migraciones, desconoce su uso y los aportes que pueden generar para la toma de decisiones de los inspectores migratorios, a partir de la formación y visualización de agrupaciones o clústeres de datos similares que descubren el comportamiento de extranjeros en estado de calidad irregular,⁽²⁶⁾ y en base a ello, afronten la decisión de entregar una situación migratoria acertada.

Bajo este panorama, en esta investigación formulamos la pregunta de investigación ¿En qué medida la aplicación de un algoritmo no supervisado permite clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones?; siendo el objetivo general: evaluar la aplicación del algoritmo no supervisado DBSCAN para clasificar a extranjeros según el nivel de riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones del Perú.

MÉTODOS

Tipo, nivel y diseño de la investigación

La investigación fue de tipo básica de nivel de descriptivo, utilizamos un enfoque cuantitativo, donde procesamos datos recopilados de la Superintendencia Nacional de Migraciones y aplicamos técnicas de análisis convencionales y de aprendizaje automático para obtener agrupaciones. El diseño fue no experimental, ya que la investigación se centró en la observación e identificación de las condiciones en las que se llevan a cabo las clasificaciones de extranjeros según el riesgo de inmigración irregular mediante la aplicación de DBSCAN.

Población y muestra

La población estuvo compuesta por 100 000 datos de extranjeros registrados en la base de datos de la Superintendencia Nacional de Migraciones del Perú. Dado que se trató de un estudio en el que aplicamos un algoritmo no supervisado, consideramos que, a mayor cantidad de datos los resultados de las agrupaciones o clústeres serían más confiables y precisas. No obstante, tras aplicar los filtros en el proceso de selección de atributos con mayor relación para identificar extranjeros según riesgo de inmigración irregular, la muestra utilizada fueron 31 625 datos, empleando un muestreo no probabilístico por conveniencia.

Modelo propuesto

Desarrollamos un modelo para clasificar a los extranjeros según el nivel de riesgo de inmigración irregular, basándonos en sus datos y características relevantes. Integrando técnicas de minería de datos y aprendizaje automático, aspiramos a reemplazar los mecanismos convencionales de evaluación en la Superintendencia Nacional de Migraciones del Perú, mejorando así el seguimiento y regulación de la inmigración. La figura 1 ilustra el modelo propuesto que consta de seis etapas.



Figura 1. Modelo propuesto

Preprocesamiento, procesamiento y visualización de datos

Para construir el dataset de clasificación de extranjeros según el riesgo de inmigración irregular reportados por la Superintendencia Nacional de Migraciones, seguimos el siguiente procedimiento:

Paso 1: cargamos el conjunto de datos desde la ubicación proporcionada por la Superintendencia Nacional de Migraciones. El dataset contenía información sobre diversos atributos de los migrantes, incluyendo su país de nacimiento, ocupación, y calidad migratoria, entre otros.

Paso 2: realizamos el proceso de limpieza y preprocesamiento de los datos. Eliminamos los campos de poca o nula relevancia para el análisis, como el número de orden de los registros, la fecha de ingreso, la profesión, el estado civil, el embarazo, la actividad económica, el nivel educativo, el grado superior, y las discapacidades físicas, sensoriales y cognitivas. También descartamos los campos relacionados con los servicios de internet, servicios básicos, tipo de vivienda y situación de salud.

Paso 3: examinamos la distribución de los datos en la columna "País de nacimiento" para comprender la procedencia de los migrantes. Generamos un gráfico de barras que mostraba la cantidad de migrantes por país de nacimiento.

Paso 4: Para la construcción del dataset, establecimos como objetivo identificar el riesgo de inmigración irregular en cinco países de interés: Venezuela, Colombia, Cuba, Ecuador y Uruguay (países con mayor cantidad de migrantes). Se filtraron los registros que no correspondían a la calidad migratoria de turista y se eliminó la columna "Calidad migratoria". Luego, seleccionamos únicamente los migrantes nacidos en los países mencionados (figura 2).

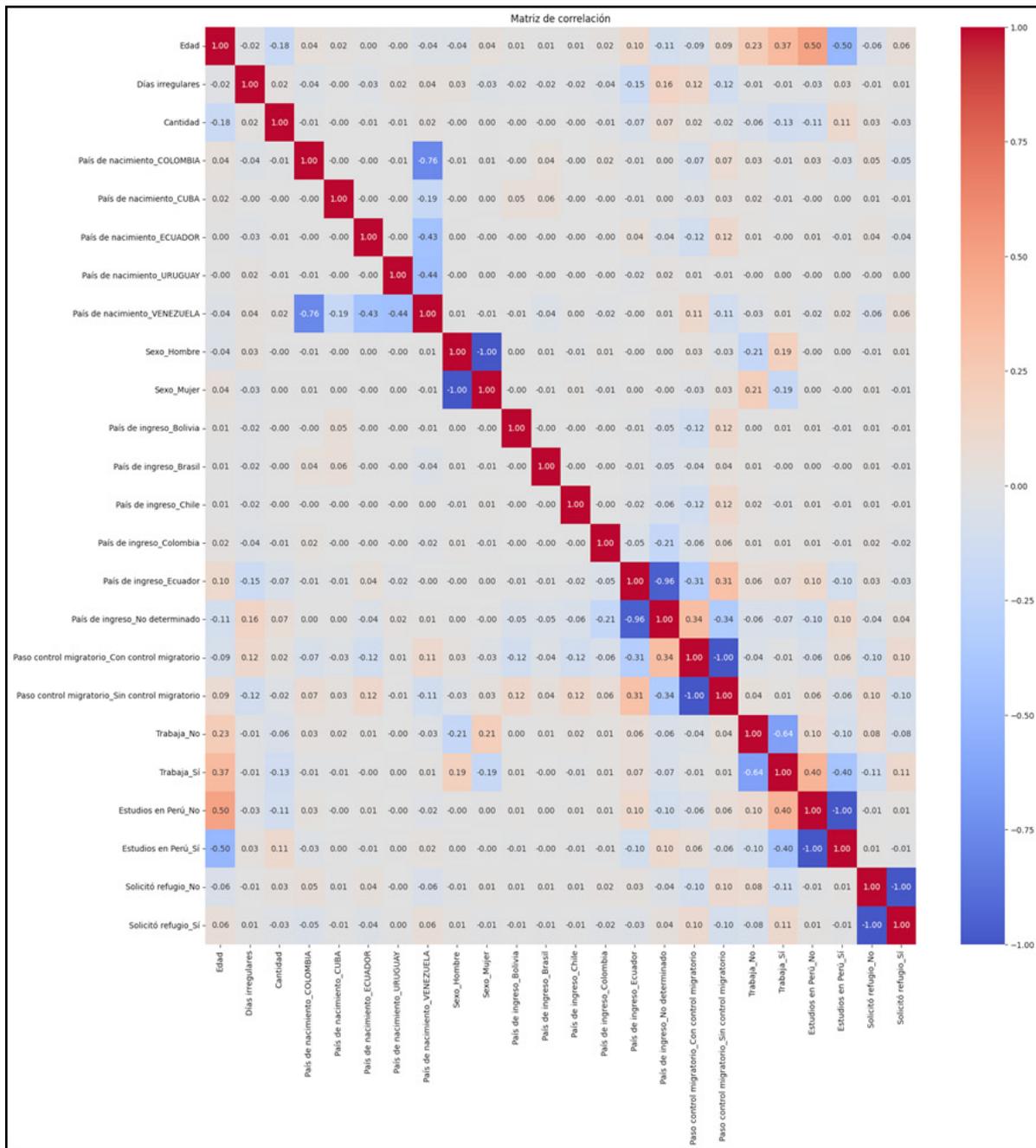


Figura 2. Matriz de correlación de los datos

Paso 5: convertimos las variables categóricas en valores numéricos. Esto permitió representar las características categóricas como variables binarias, lo cual fue necesario para utilizar algoritmos de clasificación.

Paso 6: con el objetivo de mejorar el rendimiento del modelo de clasificación y reducir la dimensionalidad del dataset, aplicamos una técnica de escalamiento de características para estandarizar los valores de los atributos.

Paso 7: utilizamos el análisis de componentes principales (PCA) para reducir aún más la dimensionalidad del dataset de 22 a 10. Esto implicó proyectar los datos en un espacio de menor dimensionalidad que preservara la mayor cantidad de información posible.

RESULTADOS Y DISCUSIONES

Sobre el dataset obtenido en los pasos previos, llevamos cabo la aplicación del algoritmo DBSCAN. Este proceso lo realizamos en dos etapas. La primera etapa, efectuamos iteraciones utilizando diferentes valores de EPS (épsilon) y Min_pts (número mínimo de puntos) para explorar diferentes configuraciones de agrupamiento. Calculamos métricas como el número de clústeres, y los coeficientes Silhouette, Davies-Bouldin, Calinski-Harabasz y la cantidad de puntos considerados como ruido para evaluar la calidad de cada configuración. En

la segunda etapa, seleccionamos los parámetros más apropiados, en este caso, EPS=3,5 y Min_pts=125, dado a la relación entre el número de clústeres (4), la cantidad de puntos de tipo ruido (405) y el coeficiente de Silhouette (0,5338). Los parámetros estimados se muestran en la figura 3.

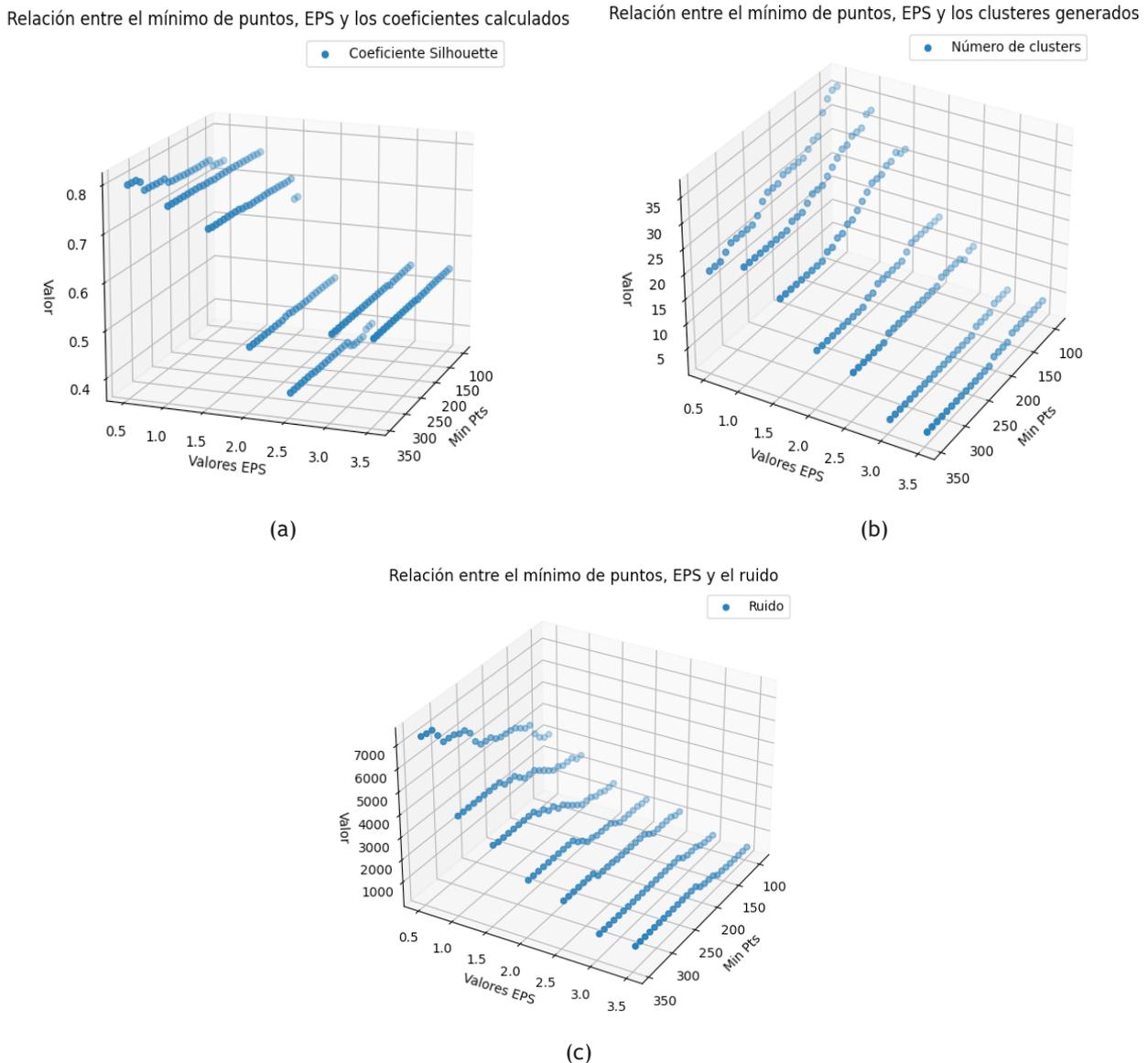


Figura 3. Evaluación del impacto de los parámetros EPS y Min_pts en DBSCAN sobre (a) coeficiente de Silhouette, (b) número de clústeres, y (c) puntos ruido

Tras la selección de los hiperparámetros del modelo, aplicamos nuevamente el algoritmo DBSCAN utilizando estos valores. Los resultados se guardaron en un archivo .xlsx que contenía información clasificada de los extranjeros, incluyendo sus atributos originales y las etiquetas de clúster asignadas por DBSCAN.

Al ejecutar el algoritmo sobre el conjunto de datos propuesto, obtuvimos conjuntos de resultados a los que aplicamos dos técnicas de validación (visual e interna) para verificar la precisión del modelo propuesto.

Validación visual

Para mejorar la visualización de los resultados, aplicamos el análisis de componentes principales (PCA) para reducir la dimensionalidad del conjunto de datos a tres componentes principales. Luego, creamos un gráfico en 3D que representaba los diferentes clústeres identificados por el algoritmo DBSCAN (figura 4).

El enfoque de selección de parámetros durante el presente estudio se alinea fuertemente al enfoque utilizado por Jin-Heng et al.⁽²⁷⁾, quienes proponen un modelo iterativo con el objetivo de mejorar la precisión de la agrupación con el coeficiente de Silhouette y el ruido como criterios significantes.

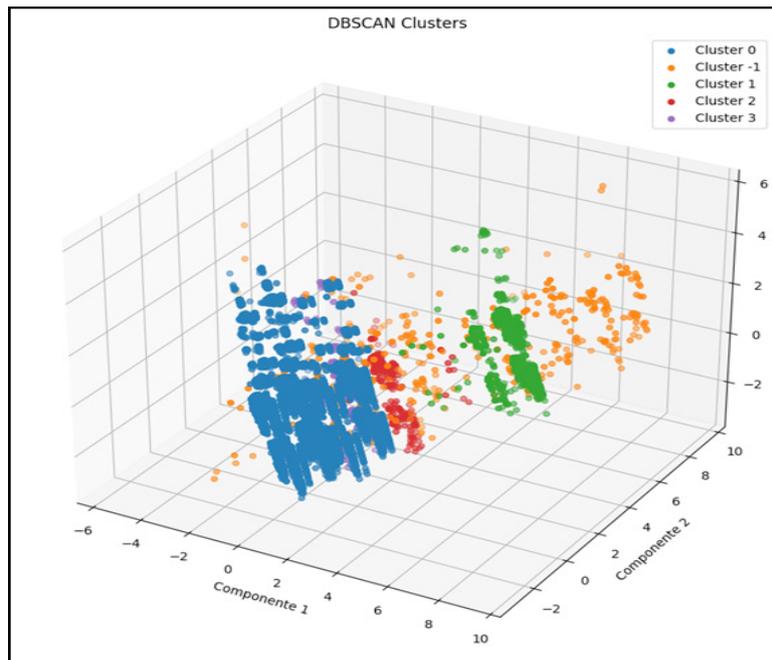


Figura 4. Representación tridimensional de los clústeres generados

Validación interna

Después de haber explorado las diferentes combinaciones de parámetros de valor EPS (épsilon) y Min_pts (puntos mínimos), calculamos las métricas de evaluación: coeficientes de Silhouette, Davies-Bouldin, y Calinski-Harabasz. Dados los parámetros seleccionados, obtuvimos los siguientes coeficientes para el algoritmo DBSCAN: Silhouette de 0,5338, Davies-Bouldin de 0,6213 y Calinski-Harabasz de 3680,2359. Según las características estadísticas de los clústeres y especialmente su coeficiente de Silhouette, podemos afirmar que las agrupaciones formadas tienen una buena cohesión entre los elementos dentro de cada clúster y buena separabilidad respecto a los elementos de los demás, entonces la aplicación de un algoritmo no supervisado clasifica extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones con alto coeficiente de Silhouette. Estos resultados comparten similitud en su comportamiento respecto a estudios como el de Anitha et al.⁽²⁸⁾ el cual alcanzó un coeficiente de Silhouette de 0,3621.

Por otra parte, para poder llevar a cabo el etiquetado de las agrupaciones formadas por el algoritmo de aprendizaje no supervisado DBSCAN, se representaron las medias de cada columna de datos dentro del dataset original para poder interpretar de una manera más clara la distribución de dichos datos dentro de cada clúster.

Tabla 1. Distribución media de las características de inmigrantes irregulares según clúster

Clúster	0	1	2	3
Edad	27,284934	34,4865811	34,6308411	29,8436482
Días irregulares	1007,16256	617,227538	766,495327	851,061889
Cantidad	1,02915354	1,00116686	1	1
País de nacimiento_COLOMBIA	0	0	1	0
País de nacimiento_CUBA	0	0	0	0
País de nacimiento_ECUADOR	0	0	0	0
País de nacimiento_URUGUAY	0	0	0	0
País de nacimiento_VENEZUELA	1	1	0	1
Sexo_Hombre	0,51491187	0,4235706	0,44392523	0,61237785
Sexo_Mujer	0,48508813	0,5764294	0,55607477	0,38762215
País de ingreso_Bolivia	0	0	0	0
País de ingreso_Brasil	0	0	0	0
País de ingreso_Chile	0	0	0	0
País de ingreso_Colombia	0	0	0	1
País de ingreso_Ecuador	0,1798472	0,98483081	0,03738318	0
País de ingreso_No determinado	0,8201528	0,01516919	0,96261682	0

Paso control migratorio_Con control migratorio	1	0	1	1
Trabaja_No	0,30416862	0,40256709	0,44392523	0,35504886
Trabaja_Sí	0,47463307	0,51575263	0,42990654	0,53420195
Estudios en Perú	0,1638295	0,0361727	0,02803738	0,1009772
Solicitó refugio	0,39642115	0,13418903	0,12616822	0,30618893

Basándonos en la distribución media de los datos, convocamos a un equipo de expertos para que clasificaran los clústeres. El objetivo era identificar qué datos indican un mayor riesgo de inmigración irregular, considerando factores como la cantidad de días en situación irregular y si el individuo pasó por un control migratorio. Esta clasificación se llevó a cabo en consonancia con la Ley de Migraciones peruana (Ley N° 1350), lo que nos llevó a obtener los resultados presentados en la tabla 2.

Clúster	Nivel de riesgo de inmigración irregular
0	Alto
1	Bajo
2	Medio bajo
3	Medio alto

Los resultados del análisis permitieron identificar diferentes clústeres:

Clúster 0 (Alto riesgo): este grupo destaca por tener el mayor número de días en situación irregular. Su atención debe ser prioritaria, y las campañas de regularización deberían enfocarse principalmente en ellos. Dada la predominancia de migrantes venezolanos en este clúster, se sugiere la posibilidad de establecer convenios bilaterales entre Perú y Venezuela.

Clúster 1 (Bajo riesgo): a pesar del bajo riesgo asociado a este grupo, ciertas particularidades, como el hecho de que muchos de sus miembros ingresan desde Ecuador sin pasar por un control migratorio, hacen necesario reforzar la cooperación binacional con Ecuador en temas migratorios.

Clúster 2 (Medio bajo riesgo) y Clúster 3 (Medio alto riesgo): estos grupos, situados en un rango intermedio de riesgo, se beneficiarían de campañas informativas que les proporcionen orientación sobre sus derechos y los mecanismos de regularización disponibles.

La clasificación de extranjeros según el riesgo de inmigración irregular desempeña un papel crucial en la gestión migratoria. No sólo guía políticas y estrategias, sino que también se alinea con la realidad legislativa de un país. Como subrayan Czaika et al.⁽¹⁾, entender las motivaciones detrás de la migración es esencial. La identificación de patrones mediante clústeres proporciona pistas sobre factores comunes entre grupos con mayor riesgo de inmigración irregular. En este contexto, resulta imprescindible correlacionar el nivel de riesgo identificado con la ley migratoria peruana vigente. Este análisis profundo nos permite discernir cómo Perú aborda la inmigración irregular y qué medidas legislativas existen para enfrentarla. Se presenta, por tanto, la oportunidad de evaluar si las regulaciones actuales son adecuadas o si se requieren reformas basadas en los datos recabados.

Por otro lado, el vínculo entre sostenibilidad y migración está emergiendo como un área de interés. Franco Gavonel et al.⁽²⁾ destacan el "paradox de migración-sostenibilidad", enfatizando cómo las dinámicas migratorias afectan directamente a la sostenibilidad. La identificación de grupos de alto riesgo es crucial no solo desde la perspectiva de garantizar derechos de inmigrantes, sino también para alinear las intervenciones con objetivos sostenibles. En este sentido, es vital que cualquier estrategia respecto a la inmigración irregular se base no solo en patrones y estadísticas, sino también en el marco legislativo pertinente para garantizar medidas justas y respetuosas de los derechos de los migrantes.

A pesar de sus potenciales aplicaciones, el estudio y el modelo DBSCAN presentan limitaciones. El proceso de selección de hiperparámetros, aunque robusto, tiene un grado de subjetividad. El coeficiente de Silhouette, pese a ser alto, no certifica la relevancia real de los clústeres. DBSCAN, al ser un método no supervisado, podría generar agrupaciones que, aunque matemáticamente sensatas, no tengan un significado práctico claro. Además, su sensibilidad a parámetros y dificultades en ciertos escenarios demandan cautela en su aplicación. Sin embargo, su potencial en la Dirección de Migraciones es evidente, especialmente en el control del tránsito migratorio, aunque requiere de recursos, capacitación y actualización constantes.

Finalmente, la incorporación de DBSCAN en la Dirección de Migraciones se presenta como una herramienta estratégica. Puede desvelar tendencias, identificar segmentos de interés y fomentar la investigación y el desarrollo. La colaboración interinstitucional se ve potenciada por su uso, facilitando una gestión migratoria

coordinada. Por lo tanto, su integración, pese a las limitaciones, se percibe como beneficioso, siempre que se ajuste al cambiante escenario migratorio.

CONCLUSIONES

Logramos evaluar la aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones, demostrando que la aplicación de DBSCAN es estrategia efectiva.

Este estudio aporta una visión valiosa en la clasificación del riesgo de inmigración irregular. A través de técnicas avanzadas de análisis de datos y con el respaldo de investigaciones previas en el campo migratorio, se ha logrado una comprensión más profunda de los patrones asociados con diferentes niveles de riesgo. Sin embargo, es esencial continuar con este tipo de investigaciones, ampliando datasets y considerando más variables que puedan influir en el fenómeno migratorio.

REFERENCIAS BIBLIOGRÁFICAS

1. Czaika M, Reinprecht C. Migration Drivers: Why Do People Migrate? *Introd to Migr Stud.* Springer, Cham; 2022. p. 49-82. https://doi.org/10.1007/978-3-030-92377-8_3.
2. Franco Gavonel M, Adger WN, Safra de Campos R, Boyd E, Carr ER, Fábos A, Fransén S, Jolivet D, Zickgraf C, Codjoe SN, Abu M, Siddiqui T. The migration-sustainability paradox: transformations in mobile worlds. *Curr Opin Environ Sustain.* 2021;49:98-109. <https://doi.org/10.1016/j.cosust.2021.03.013>.
3. Delbosc A, Shafi R. What do we know about immigrants' travel behaviour? A systematic literature review and proposed conceptual framework. *Transp Rev.* 2023;43(5):914-934 <https://doi.org/10.1080/01441647.2023.2179683>.
4. Scholten P, Pisarevskaya A, Levy N. An Introduction to Migration Studies: The Rise and Coming of Age of a Research Field. *Introd to Migr Stud.* Springer, Cham; 2022. p. 3-24. https://doi.org/10.1007/978-3-030-92377-8_1.
5. IOM's GMDAC. Poblaciones de migrantes internacionales. Estadísticas de La Inmigración y Emigración. 2021. <https://www.migrationdataportal.org/es/themes/poblaciones-de-migrantes-internacionales>.
6. Hajro A, Caprar D V., Zikic J, Stahl GK. Global migrants: Understanding the implications for international business and management. *J World Bus.* 2021;56(2):101192. <https://doi.org/10.1016/j.jwb.2021.101192>.
7. Armijos-Orellana AC, Maldonado-Matute JM, González-Calle MJ, Guerrero-Maxi PF. Los motivos de la migración. Una breve revisión bibliográfica. *Universitas.* 2022; 37:223-246. <https://doi.org/10.17163/uni.n37.2022.09>.
8. Lami B, Tarasenko I, Shyshova O, Voropayeva T, Verhun A. International Migration of Human Resources in the Conditions of Social Transformations. *Economies.* 2022;10(9):216. <https://doi.org/10.3390/economies10090216>.
9. Iranzo Á. Sub-Saharan migrants 'in transit': intersections between mobility and immobility and the production of (in)securities. *Mobilities.* 2021;16:739-757. <https://doi.org/10.1080/17450101.2021.1935305>.
10. Zeledon I, Unger JB, West AE, Cruz N, Schwartz SJ. Immigration and cultural stressors and their impact on mental health outcomes. *Encycl Child Adolesc Heal.* Elsevier; 2023. p. 698-710. <https://doi.org/10.1016/B978-0-12-818872-9.00197-7>.
11. Ajzenman N, Aksoy CG, Guriev S. Exposure to transit migration: Public attitudes and entrepreneurship. *J Dev Econ.* 2022;158:102899. <https://doi.org/10.1016/j.jdeveco.2022.102899>.
12. Cormoş VC. The Processes of Adaptation, Assimilation and Integration in the Country of Migration: A Psychosocial Perspective on Place Identity Changes. *Sustainability.* 2022;14(16):10296. <https://doi.org/10.3390/su141610296>.
13. INEI. Situación migratoria. Condiciones Vida la Población Venez que reside en Perú. Instituto Nacional de Estadística e Informática; 2018. p. 35-60. https://www.inei.gob.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1666/libro.pdf.

14. Kaeser A, Tani M. Do immigrants ever oppose immigration? *Eur J Polit Econ.* 2023;102460. <https://doi.org/10.1016/j.ejpoleco.2023.102460>.
15. Guerreiro J, Rebelo S, Teles P. What is the optimal immigration policy? *Migration, jobs, and welfare. J Monet Econ.* 2020;113:61-87. <https://doi.org/10.1016/j.jmoneco.2020.04.009>.
16. Bratu C, Dahlberg M, Engdahl M, Nikolka T. Spillover effects of stricter immigration policies. *J Public Econ.* 2020;190:104239. <https://doi.org/10.1016/j.jpubeco.2020.104239>.
17. Benavides Q, Doshi M, Valentín-Cortés M, Militzer M, Quiñones S, Kraut R, Rion R, Bryce R, Lopez WD, Fleming PJ. Immigration law enforcement, social support, and health for Latino immigrant families in Southeastern Michigan. *Soc Sci Med.* 2021;280:114027. <https://doi.org/10.1016/j.socscimed.2021.114027>.
18. Romero-Chuquital A, Melendres-Velasco JJ. Uso de data Warehouse para la toma de decisiones empresariales: una revisión literaria. *Rev Cient Sist Inform.* 2023;3(2):e543. <https://doi.org/10.51252/rcsi.v3i2.543>.
19. Guevara-Vega EMD, Delgado-Deza JR, Mendoza-de-los-Santos AC. Vulnerabilidades y amenazas en los activos de información. *Rev Cient Sist Inform.* 2023;3(1):e461. <https://doi.org/10.51252/rcsi.v3i1.461>.
20. Linares MD. La Dirección Nacional de Migraciones entre 2003 y 2015: acciones para el cambio de cultura organizacional. *Na mirada “desde adentro.” REMHU Rev Interdiscip da Mobilidade Humana.* 2018;26(53):187-203. <https://doi.org/10.1590/1980-85852503880005312>.
21. Santi S. La externalización de la “matriz de gestión migratoria” de la UE hacia América Latina y el Caribe. *Estud Front.* 2022;23: e109. <https://doi.org/10.21670/ref.2225109>.
22. Udayakumar R, Yogesh Pansambal S, Anbazhagan K, Sugumar R. Real-time Migration Risk Analysis Model for Improved Immigrant Development Using Psychological Factors. *Migr Lett.* 2023;20(4):33-42. <https://migrationletters.com/index.php/ml/article/view/2979>.
23. Ren C, Bloemraad I. New Methods and the Study of Vulnerable Groups: Using Machine Learning to Identify Immigrant-Oriented Nonprofit Organizations. *Socius Sociol Res a Dyn World.* 2022;8:237802312210769. <https://doi.org/10.1177/23780231221076992>.
24. Shao Y, Liu L, Gao H, Xu H, Wang Y, Gong S, Huang H. Clustering Algorithm Based on the Ground-Air Cooperative Architecture in Border Patrol Scenarios. *Electronics.* 2022;11(18):2876. <https://doi.org/10.3390/electronics11182876>.
25. Han S. An analysis of Koreans’ attitudes towards migrants by application of algorithmic approaches. *Heliyon.* 2022;8(8):e10087. <https://doi.org/10.1016/j.heliyon.2022.e10087>.
26. Molnar P. New technologies in migration: human rights impacts. *Forced Migr Rev.* 2019;61:7-9. <https://www.fmreview.org/ethics/molnar>.
27. Jin-Heng G, Jia-Xiang L, Zhen-Chang Z, Han-Yu L. CDBSCAN: Density clustering based on silhouette coefficient constraints. 2022 *Int Conf Comput Eng Artif Intell. IEEE;* 2022. p. 600-605. <https://doi.org/10.1109/ICCEAI55464.2022.00128>.
28. Anitha P, Patil MM. RFM model for customer purchase behavior using K-Means algorithm. *J King Saud Univ - Comput Inf Sci.* 2022;34:1785-1792. <https://doi.org/10.1016/j.jksuci.2019.12.011>.

FINANCIACIÓN

Los autores no recibieron financiación para el desarrollo de la presente investigación.

CONFLICTO DE INTERESES

Los autores declaran que no existe conflicto de intereses.

CONTRIBUCIÓN DE AUTORÍA

Conceptualización: Miguel Valles-Coral, Ulises Lazo-Bartra.

Curación de datos: Miguel Valles-Coral, Lloy Pinedo, Jorge Raul Navarro-Cabrera, Luis Salazar-Ramírez.

Análisis formal: Miguel Valles-Coral, Ulises Lazo-Bartra, Lloy Pinedo, Jorge Raul Navarro-Cabrera, Luis Salazar-Ramírez.

Metodología: Miguel Valles-Coral, Lloy Pinedo, Jorge Raul Navarro-Cabrera, Luis Salazar-Ramírez, Fernando Ruiz-Saavedra.

Supervisión: Miguel Valles-Coral, Luis Salazar-Ramírez.

Validación: Miguel Valles-Coral, Fernando Ruiz-Saavedra, Segundo Ramírez, Pierre Vidaurre-Rojas.

Visualización: Miguel Valles-Coral, Ulises Lazo-Bartra, Fernando Ruiz-Saavedra, Segundo Ramírez.

Redacción - borrador original: Miguel Valles-Coral, Ulises Lazo-Bartra, Lloy Pinedo, Luis Salazar-Ramírez, Pierre Vidaurre-Rojas.

Redacción - revisión y edición: Miguel Valles-Coral, Jorge Raul Navarro-Cabrera, Luis Salazar-Ramírez.