



ESCUELA DE POSGRADO
UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA
PROGRAMA EN MAESTRÍA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN

Tesis

**Aplicación de un algoritmo no supervisado para
clasificar extranjeros según riesgo de inmigración
irregular en la Superintendencia Nacional de
Migraciones**

Para optar el grado académico de Maestro en Ciencias con Mención
en Tecnología de la Información

Autor:

Ulises Lazo Bartra

<https://orcid.org/0009-0001-2077-9600>

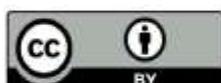
Asesor:

Ing. Dr. Miguel Ángel Valles Coral

<https://orcid.org/0000-0002-8806-2892>

Tarapoto, Perú

2023



Esta obra está bajo una [Licencia Creative Commons Atribución - 4.0 Internacional \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Vea una copia de esta licencia en <https://creativecommons.org/licenses/by/4.0/deed.es>





ESCUELA DE POSGRADO
UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA
PROGRAMA EN MAESTRA EN CIENCIAS CON MENCIÓN EN TECNOLOGIA DE
LA INFORMACIÓN

Tesis

Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

Para optar el grado académico de Maestro en Ciencias con Mención
en Tecnología de la Información

Autor

Ulises Lazo Bartra

<https://orcid.org/0009-0001-2077-9600>

Asesor

Ing. Dr. Miguel Angel Valles Coral

<https://orcid.org/0000-0002-8806-2892>

Tarapoto, Perú

2023



ESCUELA DE POSGRADO

UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA
PROGRAMA DE MAESTRÍA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN

Tesis

**Aplicación de un algoritmo no supervisado para
clasificar extranjeros según riesgo de inmigración
irregular en la Superintendencia Nacional de
Migraciones**

Para optar el grado académico de Maestro en Ciencias con mención en
Tecnología de la Información

Autor:

Ulises Lazo Bartra

<https://orcid.org/0009-0001-2077-9600>

Asesor:

Ing. Dr. Miguel Ángel Valles Coral

<https://orcid.org/0000-0002-8806-2892>

Tarapoto, Perú

2023



ESCUELA DE POSGRADO

UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERÍA DE SISTEMAS E
INFORMÁTICA
PROGRAMA DE MAESTRÍA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA
DE LA INFORMACIÓN

Tesis

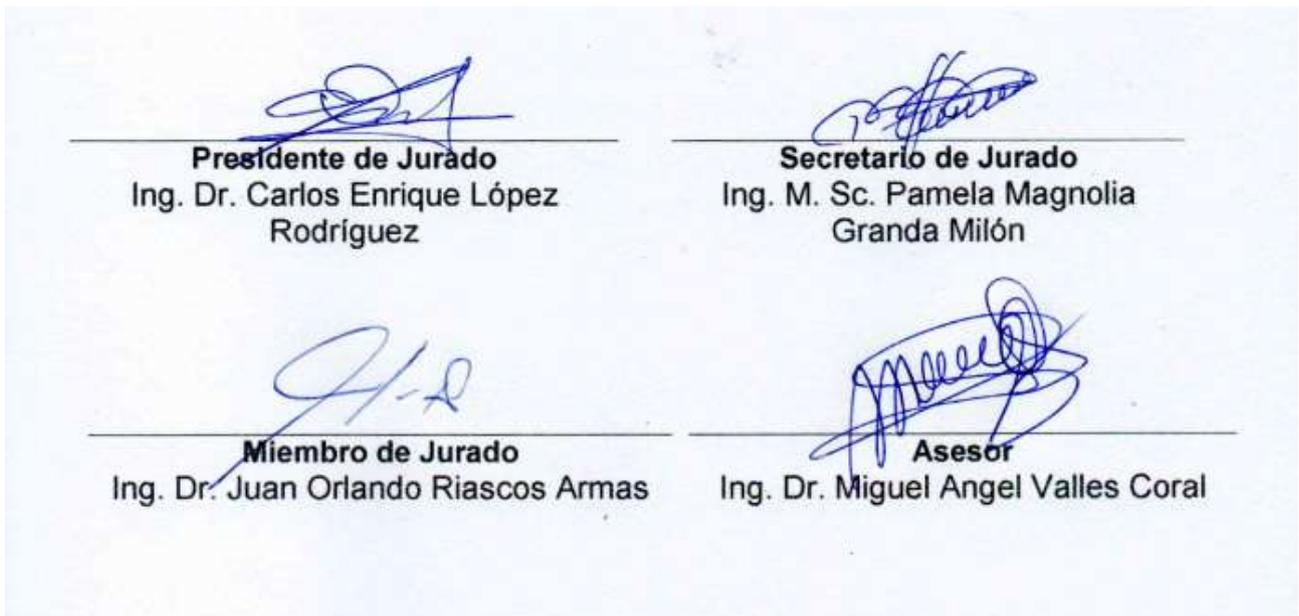
**Aplicación de un algoritmo no supervisado para
clasificar extranjeros según riesgo de inmigración
irregular en la Superintendencia Nacional de
Migraciones**

Para optar el grado académico de Maestro en Ciencias con mención en
Tecnología de la Información

Autor:

Ulises Lazo Bartra

Sustentado y revisado el 25 de agosto de 2023 por los jurados:



Tarapoto, Perú

2023



UNIVERSIDAD NACIONAL DE SAN MARTÍN-T



Escuela de Posgrado

ACTA DE SUSTENTACIÓN DE TESIS

Los Miembros del Jurado que suscriben, reunidos para estudiar y escuchar la sustentación y defensa del Trabajo de Tesis, modo presencial, presentado por:

Bach. Ulises Lazo Bartra

Con el asesoramiento del Ing. Dr. Miguel Ángel Valles Coral.

"Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones"

Teniendo en consideración los méritos del referido trabajo, así como los conocimientos demostrados por el sustentante, lo declaramos:

APROBADO

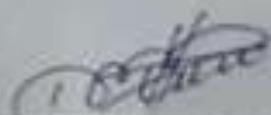
Con el calificativo (*)

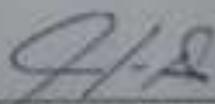
MUY BUENO (18)

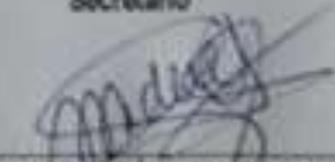
En consecuencia, queda en condición de ser considerado APTO por el Consejo Universitario y recibir el Grado Académico de Maestro, de conformidad con lo estipulado en el Artículo 30° del Reglamento de Tesis de la Escuela de Posgrado de la UNSM.

Tarapoto, 25 de agosto de 2023.


Ing. Dr. Carlos Enrique López Rodríguez
Presidente


Ing. M.Sc. Pamela Magnolia Granda Milón
Secretario


Ing. Mg. Joan Orlando Riascos Armas
Miembro


Ing. Dr. Miguel Ángel Valles Coral
Asesor

(*) De acuerdo con el Artículo 40° del Reglamento General de Ciencia, Tecnología e Innovación (RC - CTS) la Universidad Nacional de San Martín - Tarapoto, estas deberán ser calificadas con términos de: BUENO, MUY BUENO, EXCELENTE, también considerar la nota.



ESCUELA DE POSGRADO

UNIDAD DE POSGRADO DE LA FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
PROGRAMA DE MAESTRIA EN CIENCIAS CON MENCIÓN EN TECNOLOGÍA DE LA INFORMACIÓN

Tesis

Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

Para optar el grado académico de Maestro en Ciencias con mención en
Tecnología de la Información

El suscrito declara que el presente trabajo de tesis es original, en su
contenido y forma

Ulises Lazo Bartra
Autor

Ing. Dr. Miguel Ángel Valles Coral
Asesor

Tarapoto, Perú

2023

Declaratoria de autenticidad

Ulises Lazo Bartra, identificado con DNI N° 43025310, egresado de la Escuela de Posgrado de la Universidad Nacional de San Martín, Unidad de Posgrado de la Facultad de Ingeniería de Sistemas e Informática, Programa de Maestría en Ciencias con Mención en Tecnología de la Información, con la tesis titulada: “Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones”.

Declaro bajo juramento que:

- 1) La tesis presentada es de mi autoría.
- 2) He respetado las normas internacionales de citas y referencias para las fuentes consultadas. Por tanto, la tesis no ha sido plagiada ni total ni parcialmente.
- 3) La tesis no ha sido auto plagiada; es decir, no ha sido publicada ni presentada anteriormente para obtener algún grado académico previo o título profesional.
- 4) Los datos presentados en los resultados son reales, no han sido falseados, ni duplicados, ni copiados y por tanto los resultados que se presenten en la tesis se constituirán en aportes a la realidad investigada.

De considerar que el trabajo cuenta con una falta grave, como el hecho de contar con datos fraudulentos, demostrar indicios y plagio (al no citar la información con sus autores), plagio (al presentar información de otros trabajos como propios), falsificación (al presentar la información e ideas de otras personas de forma falsa), entre otros, asumo las consecuencias y sanciones que de mi acción se deriven, sometiéndome a la normatividad vigente de la Universidad Nacional de San Martín.

Tarapoto, 25 de agosto de 2023.



Ulises Lazo Bartra
DNI N° 43025310

Ficha de identificación

<p>Título del proyecto Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones</p>	<p>Área de investigación: Ciencias Naturales Línea de investigación: Ciencias de la Computación Sublínea de investigación: Inteligencia Artificial y Recuperación de Información Grupo de investigación: Grupo de Investigación de IA Resolución: Resolución N°134-2021-UNSM/FISI/CFT Tipo de investigación: Básica <input checked="" type="checkbox"/> Aplicada <input type="checkbox"/> Desarrollo experimental <input type="checkbox"/></p>
<p>Autora: Ing. Ulises Lazo Bartra</p>	<p>Escuela de Posgrado Unidad de Posgrado de la Facultad de Ingeniería de Sistemas e Informática Programa de Maestría en Ciencias con mención en Tecnología de la Información https://orcid.org/0009-0001-2077-9600</p>
<p>Asesor: Ing. Dr. Miguel Angel Valles Coral</p>	<p>Dependencia local de soporte: Facultad de Ingeniería de Sistemas e Informática Escuela Profesional de Ingeniería de Sistemas e Informática Unidad o Laboratorio Ingeniería de Sistemas e Informática https://orcid.org/0000-0002-8806-2892</p>

Dedicatoria

Este logro lo dedico a mi mamá, Aylen Bartra Sanchez, quien me ha brindado un valioso apoyo en el inicio de esta maestría. Gracias a su aliento, puedo desarrollarme profesionalmente y adquirir las habilidades necesarias para ser más competitivo.

Agradecimientos

Agradezco a mi asesor, Miguel Angel Valles Coral, coordinador del Grupo de Investigación en Inteligencia Artificial de la Facultad de Ingeniería de Sistemas e Informática; gracias a su supervisión he logrado cumplir cada uno de los objetivos planteados en la presente investigación, y generar un aporte a la sociedad.

Asimismo, agradezco a los docentes de la Maestría en Ciencias con Mención en Tecnología de la Información, por compartir sus conocimientos y aplicar metodologías de enseñanza eficaces para nuestra formación como profesionales.

Índice general

Declaratoria de autenticidad	7
Ficha de identificación	7
Dedicatoria	8
Agradecimientos	9
Índice general.....	10
Índice de tablas	12
Índice de figuras.....	13
RESUMEN	14
ABSTRACT	15
CAPÍTULO I INTRODUCCIÓN A LA INVESTIGACIÓN	16
CAPÍTULO II MARCO TEÓRICO	20
2.1. Antecedentes de la investigación	20
2.2. Fundamentos teóricos	22
CAPÍTULO III MATERIALES Y MÉTODOS	33
3.1. Ámbito y condiciones de la investigación.....	33
3.1.1 Contexto de la investigación	33
3.1.2 Periodo de ejecución.....	33
3.1.3 Autorizaciones y permisos	33
3.1.4 Control ambiental y protocolos de bioseguridad	33
3.1.5 Aplicación de principios éticos internacionales	33
3.2. Sistema de variables	34
3.2.1 Variables principales	34
3.2.2 Variables secundarias.....	34
3.3 Procedimientos de la investigación	35
3.3.1 Objetivo específico 1	36
3.3.2 Objetivo específico 2	36
3.3.3 Objetivo específico 3	37
CAPÍTULO IV RESULTADOS Y DISCUSIÓN	38

CONCLUSIONES	50
RECOMENDACIONES.....	51
REFERENCIAS BIBLIOGRÁFICAS	52
ANEXOS	57

Índice de tablas

Tabla 1. Descripción de variables por objetivo general.....	34
Tabla 2. Estadísticos de los clústeres.....	47
Tabla 3. Distribución media de las características de inmigrantes irregulares según cluster	48
Tabla 4. Nivel de riesgo de inmigración irregular por cluster	49

Índice de figuras

Figura 1. Flujo de procedimientos.....	37
Figura 2. Distribución de los datos según el país de nacimiento	40
Figura 3. Matriz de correlación de los datos.....	41
Figura 4. Relación entre los parámetros (EPS y Min_pts) y el coeficiente de Silhouette alcanzado.....	44
Figura 5. Relación entre los parámetros (EPS y Min_pts) y el número de clusteres generados	44
Figura 6. Relación entre los parámetros (EPS y Min_pts) y el número de puntos ruido	45
Figura 7. Representación tridimensional de los clústeres generados.....	46

RESUMEN

Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

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. Sin embargo, la migración irregular plantea desafíos para los países receptores, ya que implica el ingreso de personas sin la documentación adecuada y puede generar inseguridades en términos de seguridad nacional y control de fronteras. En este contexto, este estudio tuvo como objetivo evaluar la aplicación de un algoritmo no supervisado para clasificar extranjeros según el riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones. Para ello, se utilizó el algoritmo DBSCAN (Agrupamiento Espacial Basado en la Densidad de Aplicaciones con Ruido). La metodología del estudio involucró la construcción de un dataset con datos de extranjeros reportados por la Superintendencia Nacional de Migraciones, la aplicación del algoritmo DBSCAN para clasificar a los extranjeros en diferentes clústeres según su nivel de riesgo de inmigración irregular, y la determinación del coeficiente de Silhouette como medida de la calidad de la clasificación. Los resultados mostraron que el algoritmo de aprendizaje no supervisado DBSCAN logró clasificar a los extranjeros en cuatro clústeres representando los niveles de riesgo de inmigración irregular: alto, medio alto, medio bajo y bajo. El coeficiente de Silhouette obtenido fue de 0,534, lo cual indica una clasificación significativa y consistente. En conclusión, este estudio demostró que la aplicación de DBSCAN como algoritmo no supervisado es una estrategia efectiva para la clasificación de extranjeros según el riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones. Estos resultados tienen implicaciones importantes para la toma de decisiones informadas por parte de los inspectores migratorios, contribuyendo a la reducción de la inmigración irregular y al mantenimiento del estatus migratorio adecuado en el país.

Palabras clave: calidad migratoria, control migratorio, migración, algoritmo no supervisado, inspectores migratorios

ABSTRACT

Application of an unsupervised algorithm to classify foreigners according to risk of irregular immigration in the National Superintendence of Migration.

Migration is a social phenomenon that affects the structure and distribution of the population and is motivated by the search for better opportunities and living conditions. However, irregular migration poses challenges for the receiving countries, since it involves the entry of people without proper documentation and can generate insecurities in terms of national security and border control. In this context, this study aimed to evaluate the application of an unsupervised algorithm to classify foreigners according to the risk of irregular immigration in the National Superintendence of Migration. For this purpose, the DBSCAN (Density-Based Spatial Clustering of Noisy Applications) algorithm was used. The methodology of the study involved the construction of a dataset with data on foreigners reported by the National Superintendence of Migration, the application of the DBSCAN algorithm to classify foreigners into different clusters according to their level of risk of irregular immigration, and the determination of the Silhouette coefficient as a measure of the quality of the classification. The results showed that the DBSCAN unsupervised learning algorithm was able to classify foreigners into four clusters representing the levels of irregular immigration risk: high, medium high, medium low and low. The Silhouette coefficient obtained was 0.534, indicating a significant and consistent classification. In conclusion, this study demonstrated that the application of DBSCAN as an unsupervised algorithm is an effective strategy for the classification of foreigners according to the risk of irregular immigration in the National Superintendency of Migration. These results have important implications for informed decision making by migration inspectors, contributing to the reduction of irregular immigration and the maintenance of appropriate immigration status in the country.

Keywords: migration quality, migration control, migration, unsupervised algorithm, migration inspectors.



CAPÍTULO I

INTRODUCCIÓN A LA INVESTIGACIÓN

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. 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 (Abel et al., 2019; Carling & Schewel, 2018; Gutiérrez Silva et al., 2020).

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 tuvo 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 (IOM's GMDAC, 2021).

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. 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. (Juste Ruiz, 2019; Lotero-Echeverri & Pérez-Rodríguez, 2019).

Los migrantes venezolanos son un claro ejemplo de migraciones forzadas debido a la crisis política, económica y humanitaria. Niño Arguelles et al. (2020), manifiestan que 7 de cada 10 venezolanos en Ecuador se dedican al comercio informal y venta ambulante, y gran parte tienen estatus migratorio irregular porque la obtención de documentos como visa, pasaporte o antecedentes penales, es complicado.

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. Para el caso del contexto nacional, 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 (INEI, 2018).

La situación de inmigrantes irregulares, no solo dirigidos a ciudadanos venezolanos, sino de diferentes extranjeros residentes en Perú, fue el problema identificado en el presente estudio, pues la permanencia en calidad migratoria irregular o ilegal puede afectar el estatus del Estado como se indica anteriormente. 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 (Azizi & Yektansani, 2020), 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. 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.) (Linares, 2018), reflejando sesgos en la asignación de permisos que, eventualmente, se conviertan en riesgos de inmigraciones irregulares (Ruiz Albornoz, 2020).

Las técnicas de minería de datos y el reconocimiento de patrones, aplicados mediante algoritmos supervisado y no supervisados, son ampliamente usados en este contexto (Chandy et al., 2017; Ren & Bloemraad, 2022; Tarasyev et al., 2018; Vazquez et al., 2020). Sin embargo, en la Superintendencia Nacional de Migraciones, se 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 (Molnar, 2019), y en base a ello, afronten la decisión de entregar una situación migratoria acertada.

Bajo este panorama, en esta investigación se formuló 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 de un algoritmo no supervisado para clasificar

extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones; y los objetivos específicos: i) Construir el datasets de clasificación de extranjeros según riesgo de inmigración irregular reportados por la Superintendencia Nacional de Migraciones; ii) Aplicar DBSCAN (Agrupamiento Espacial Basada en la Densidad de Aplicaciones con Ruido) como algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones; y iii) Determinar el coeficiente de Silhouette de la aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones. En cuanto a la hipótesis de la investigación, se planteó: 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.

El proyecto fue conveniente porque brinda clasificaciones de extranjeros según riesgo de inmigración irregular que solicitaron una calidad migratoria en la República del Perú a través de la Superintendencia Nacional de Migraciones, de tal manera que los inspectores migratorios posean y se basen en información verídica y estratégica respecto a las características de los solicitantes y así puedan tomar una decisión acertada. De este modo se contribuye a la reducción de inmigrantes irregulares que puedan perjudicar el estatus de la nación.

Por otra parte, el estudio implica una relevancia social en el contexto peruano, ya que contribuye a la solución de un problema que afecta las condiciones sociales y demográficas dentro de un territorio, pues a pesar de que los inmigrantes cuentan con el respaldo de los derechos humanos, estos tienen la responsabilidad de regirse a las normas o políticas de control y regulación de migraciones establecidas en Perú, por ende, resulta importante conocer cuáles son las clasificaciones de extranjeros según riesgo de inmigración irregular para contrarrestar la situación de inmigrantes ilegales, indocumentados, etc.

La investigación también posee justificación teórica-práctica. Poco se sabe de estudios locales o nacionales que abarquen la solución tecnológica planteada, siendo los algoritmos no supervisados técnicas en minería de datos y reconocimiento de patrones ampliamente usados en diferentes contextos. Es así que esta investigación genera una base de conocimientos para clasificar extranjeros según riesgo de inmigración irregular a través de DBSCAN que pueda replicarse en otros escenarios.

En un sentido metodológico, si bien se usa el método del algoritmo DBSCAN, también la investigación abrirá oportunidades para aplicar otros tipos de algoritmos no supervisados (K-means, puntos aleatorios, etc) cuyas finalidades sean discutir los resultados del

presente y ampliar posibilidades de soluciones mediante resultados de índices de validación de clustering más exactos. Además, se pueden proponer estudios con objetivos de reconocimiento de patrones respecto al comportamiento de los extranjeros irregular acentuados en Perú; y en base a dicha información, se mejoren las políticas de regulación de migraciones.

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes de la investigación

La literatura científica es escasa en cuanto a la aplicación de algoritmos no supervisados para clasificar extranjeros según riesgo de inmigración irregular, por lo que los antecedentes presentados tienen similitud en cuanto a la aplicación de algoritmos de aprendizaje, siendo necesarios y fundamentales para conocer las metodologías aplicadas y ampliar la discusión de los resultados.

Viloria et al. (2020), en su artículo “Algoritmos de aprendizaje no supervisado aplicados a problemas de agrupación” el objetivo de la investigación fue poder identificar las desviaciones sobre los casos del registro de eventos de los negocios, de acuerdo al requerimiento del mismo. El método empleado fue mediante la aplicación de métodos de aprendizaje no supervisado, principalmente el agrupamiento espacial basado en la densidad de aplicaciones con sonido (DBSCAN), para poder determinar las mejores alternativas de agrupamiento de las tazas. Los resultados revelaron la facilidad que brinda el agrupamiento espacial basado en la densidad de aplicaciones con sonido (DBSCAN) para lograr identificar los parámetros de agrupamientos más sencillos.

Sonde et al. (2020), menciona en su artículo “Detección y análisis de la expansión urbana mediante la clasificación no supervisada de datos de imágenes de alta resolución del área de Jawaharlal Nehru Port Trust en India” que el objetivo fue medir la expansión urbana del distrito Uran Taluka con el propósito de controlar el crecimiento desordenado. Se empleó algoritmos no supervisados para analizar las imágenes satelitales y las QuickBird. Los resultados demostraron que gracias a la interpretación y análisis de los algoritmos empleados especialmente K-means, se logró identificar que entre los años 2006 a 2014 el área urbana de la muestra del estudio se redujo en un 6,585%.

Hossain et al. (2021), en su artículo de investigación “Investigación de patrones de accidentes fatales y con lesiones de conductores adolescentes con algoritmos de aprendizaje no supervisados” proponen como objetivo lograr determinar y clasificar los patrones de accidentes fatales y con lesiones de conductores mediante la minería de reglas de asociación y el análisis de correspondencia conjunta (JCA). Los resultados evidenciaron como el algoritmo (JCA) no supervisado es herramienta útil para clasificar asociaciones significativas con la finalidad de descubrir las posibles causas y efectos de los accidentes.

Mansour et al. (2021,) en su investigación titulada “Modelo de codificador automático variacional basado en aprendizaje profundo no supervisado para el diagnóstico y la clasificación de COVID-19” proponen el objetivo del desarrollo de técnicas para la predicción de personas con COVID-19, a través del modelo no supervisado (UDL-VAE) basado en el aprendizaje profundo (DL), se empleó a demás técnicas de procesamiento en base al modelo adaptativo de Wiener para dar mayor precisión. Los resultados del diagnóstico evidenciaron la utilidad del modelo desarrollado (UDL-VAE) para clasificar a personas asintomáticas.

Paixão et al. (2022), en la investigación titulada “Intervenciones de actividad física no supervisada para personas con EPOC: una revisión sistemática y metaanálisis”, el objetivo fue lograr identificar y simplificar los efectos de las intervenciones AF no supervisadas en el tratamiento de pacientes con enfermedades pulmonares obstructivas crónicas (EPOC). Se realizó estudios cuasiexperimentales y ensayos al azar para comparar la atención habitual brindada a los pacientes con EPOC con la AF no supervisada, cuyos resultados principales fueron la mejora en la fatiga, fuerza muscular y función emocional mediante el análisis del algoritmo no supervisado AF.

Fidan & Yuksel (2022), en su investigación “Un estudio comparativo para determinar los niveles de riesgo de Covid-19 mediante métodos de aprendizaje automático no supervisados” propusieron como objetivo determinar las ciudades con mayor índice de contagios a través de técnicas de aprendizaje no supervisado, teniendo como índices el número de casos registrados y las medidas ambientales. Para lograrlo se emplearon distintos algoritmos de agrupamiento relacional. El resultado evidenció que el agrupamiento relacional gris es el método más exitoso para identificar las ciudades con mayor índice de contagios por COVID-19.

Un estudio propuesto por Martin et al. (2022), titulado “Análisis de los principales determinantes para ser inmigrante en Cuenca (Ecuador) a partir de un enfoque de agrupamiento borroso” analizó los determinantes para ser inmigrante en Cuenca. Para ello, aplicaron un método de agrupamiento difuso para analizar las diferencias observadas en los principales determinantes observados sobre una serie de rasgos según sus similitudes con tres perfiles obtenidos: (1) inmigrantes extremadamente exigentes; (2) inmigrantes extremadamente innecesarios; y (3) inmigrantes con exigencias intermedias. Los resultados demostraron que elementos como el acceso a internet y los beneficios para jubilados fueron muy valorados por algunos inmigrantes; así como los canales de información, los motivos para emigrar, la ubicación de la vivienda, el modo principal de transporte, los ingresos y la principal fuente de ingresos.

2.2. Fundamentos teóricos

Algoritmos de aprendizaje

Los algoritmos de aprendizaje son productivos para la extracción de información útil desde datos complejos. Su capacidad se basa en la detección de estructuras sutiles o ligeras de datos relacionados que son usados para pronosticar eventos. Estos algoritmos ahorran recursos humanos y máxima la precisión de los resultados. Se emplean para resolver problemas de clasificación, agrupación de imágenes, reconocimiento de voz, detección de valores atípicos y regresión (Brynjolfsson & Mitchell, 2017; Sohail & Arif, 2020); e implica una perspicacia matemática que incluyen sub campos del análisis funcional, optimización, estadística, topología y análisis numérico, que tienen el propósito de mapear datos discretos en una dimensión y espacios de características. Acá, los datos pueden ser agrupados en clases separados de mayor precisión (Sohail & Arif, 2020).

En el área de la Inteligencia Artificial (IA), la utilidad de los algoritmos de aprendizaje automático son componentes clave para soluciones de digitalización. Denominado *Machine Learning* (ML) en inglés, los algoritmos se procesan en una computadora para realizar tareas programadas y se reconoce que la máquina mejora a medida que gana experiencia en la ejecución de estas tareas; es entonces que toma decisiones propias y realiza predicciones o pronósticos basados en datos. Un ejemplo de ML es cuando un programa de computadora aprende a detectar/predecir cáncer a partir de los informes médicos de pacientes, el mismo que mejorará su rendimiento a medida que acumule experiencia de una población más amplia de pacientes, y su rendimiento se valorará por el recuento de predicciones correctas y detecciones de casos de cáncer validados por un oncólogo experimentado (Ray, 2019).

Ray (2019) menciona los siguientes campos de mayor aplicabilidad de ML:

- Robótica
- Asistentes personales virtuales
- Juegos de computadora
- Reconocimiento de patrones
- Procesamiento de lenguaje natural
- Minería de datos
- Predicción de tráfico
- Red de transporte en línea
- Recomendación de productos
- Predicción del mercado de acciones

- Diagnóstico médico
- Predicción de fraude en línea
- Asesoramiento agrícola
- Refinamiento de resultados de motores de búsqueda
- Chatbots o agentes conversacionales
- Filtrado de spam por correo electrónico
- Predicción de delitos a través de video sistema de vigilancia
- Servicios de redes sociales

Tipos de algoritmos de aprendizaje automático

Dependiendo de los problemas a identificar (clasificación, regresión y agrupamiento) y la disponibilidad de tipos y categorías de los datos de entrenamiento, es posible seleccionar entre las técnicas de aprendizaje supervisado, aprendizaje no supervisado, aprendizaje semisupervisado y aprendizaje de refuerzo, para aplicar el algoritmo de aprendizaje apropiado. A continuación, se describen los algoritmos más usados según (Ray, 2019).

- **Algoritmo de Descenso de Gradiente**

Método iterativo en el que el objetivo es minimizar una función de coste. Los coeficientes se calculan en cada iteración tomando el negativo de la derivada y reduciendo los coeficientes en cada paso por una tasa de aprendizaje multiplicada por la derivada para que puedan alcanzar los mínimos locales después de algunas iteraciones. Las iteraciones se detienen cuando converge el valor mínimo de la función de coste. Existen tres métodos: 1) Descenso de gradiente estocástico (SGD), tiene la ventaja de que las actualizaciones frecuentes dan como resultado una tasa detallada de mejora; 2) Descenso de gradiente por lotes (BGD), es algoritmo de mayor eficiencia computacional porque produce un gradiente de error y convergencia estable; y 3) Descenso de gradientes de mini lote (MBD), algoritmo que combina enfoques de SGD y BGD, se utiliza con frecuencias para entrenar redes neuronales.

- **Algoritmo de Regresión Lineal**

Está enfocado en aprendizaje supervisado y se utiliza para modelar variables continuas y hacer predicciones. En una regresión se tiene datos etiquetados y el valor de la variable de salida está determinado por los valores de la variable de entrada. Los algoritmos de regresión lineal intentan ajustar una línea recta (hiperplano recto) al conjunto de datos y es posible cuando la relación entre variables es lineal. Tiene como ventaja que es fácil de entender y evitar sobreajustes mediante regularización. Es una buena opción si se conoce la existencia de relación entre las covariables y la variable de respuesta de forma lineal.

No obstante, no es recomendado para aplicaciones prácticas, pues simplifica los problemas del mundo real.

- **Análisis de Regresión Multivariante**

La regresión lineal múltiple existe cuando hay una relación de mucho a uno, entre variables independientes (entrada/predictor) y una variable dependiente (salida/respuesta). Cuando se agregan variables de entrada no significa que la regresión será mejor ni ofrecerá mejores predicciones, a veces empeoran las cosas ya que resultan en un ajuste excesivo. La técnica de regresión multivariante brinda una idea profunda de la relación entre conjunto de variables independientes y dependientes. Esto se logra con regresión múltiple, técnicas de tabulación y correlacional parcial. Algunos de sus inconvenientes es que es compleja y requiere conocimientos en técnicas y modelados estadísticos, y la muestra debe ser alto si se desea obtener niveles de precisión confiables en el resultado de análisis.

- **Regresión Logística**

Los algoritmos de regresión logística son empleados para tratar problemas de clasificación. Estos brindan como resultado una decisión binomial basado en probabilidades de que si un evento ocurrirá o no (términos de 0 y 1) en función de los valores de las variables de entrada. Se puede presentar resultados multinomiales, por ejemplo, la predicción de tipos de comida preferidas, así como clasificaciones de productos. La regresión logística se ocupa de la predicción de la variable objetivo que es categórico, por el contrario, la regresión lineal predice valores de variables continuas. Entre sus ventajas destacan la simplicidad de la implementación, eficiencia computacional, eficiencia del entrenamiento y facilidad de regulación. Además, es pertinente señalar que no se ve afectada por un pequeño ruido en los datos y la multicolinealidad. Las desventajas que presenta son la incapacidad para solucionar problemas lineales y es propensa al sobreajuste.

- **Árbol de Decisión**

Enfoque de aprendizaje supervisado para solucionar problemas de clasificación y regresión a través de la división continua de datos en función de un determinado parámetro. El árbol de clasificación trabaja con variables de decisión tipo categórica (Si/No), y el árbol de regresión tipo continua. Las hojas respaldan las decisiones, mientras que los datos se dividen en los nodos. Este enfoque brinda facilidad de interpretación, facilidad de manejo de valores categóricos y cuantitativos, es capaz de llenar los valores faltantes en los atributos con valores que tengan mayor probabilidad de igualdad, y tiene alto rendimiento dado a la eficiencia del árbol algoritmo transversal. Sin embargo, presenta

desventajas en cuanto a inestabilidad, complicación para controlar el tamaño del árbol, propenso a errores de muestreo y solo se limita a soluciones locales.

- **Máquinas de Vectores de Soporte**

Son algoritmos que manejan problemas de clasificación y regresión. Para ello, es pertinente definir el hiperplano (límite de decisión), en el caso que exista un conjunto de objetos (lineales o no lineales) que pertenecen a diferentes clases. Estos algoritmos necesitan funciones matemáticas complejas llamadas núcleo para dividir objetos miembros de diferentes clases. Su propósito es clasificar correctamente los objetos en función de ejemplos en el conjunto de datos de entrenamiento. Las máquinas de vectores de soporte manejan datos semiestructurados y estructurados, y tienen menos probabilidad de sobreajuste. Asimismo, son capaces de escalar con datos de alta dimensión y no se obstruye en los óptimos locales. Entre sus desventajas se encuentra la disminución de rendimiento en conjuntos grandes de datos dado al tiempo incremental del entrenamiento, además no funciona correctamente cuando el conjunto de datos es ruidoso.

- **Aprendizaje Bayesiano**

Este algoritmo permite seleccionar una distribución de probabilidad previa y luego se actualiza para obtener una distribución posterior. Las redes bayesianas pueden manejar conjuntos de datos incompletos, evitando el ajuste excesivo de los datos, no habiendo la necesidad de eliminar las contradicciones de los datos. Presenta desventajas ya que la selección previa es difícil, la distribución posterior es influenciada a veces por la anterior y si estas son incorrectas las predicciones fallaran.

- **Naives Bayes**

Se considera un algoritmo simple basada en la probabilidad condicional. Dentro de este enfoque se presenta una tabla de probabilidad actualizada mediante datos de entrenamiento. Naives Bayes se fundamenta en que los datos tienen valores característicos donde es esencial buscar las probabilidades de clase para predecir. Las ventajas que presenta son la fácil implementación, buen rendimiento, funcionalidad con pocos datos, manejo de datos continuos y discretos, manejo de problemas de clasificación binaria y múltiples, realizar predicciones probabilísticas, y no es sensible a características irrelevantes.

- **Algoritmo K vecinos más próximos**

Es un algoritmo de clasificación. Emplea una base de datos que tiene puntos de datos agrupados en varias clases y el algoritmo busca clasificar el punto de datos de muestra

que se le ha otorgado como un problema de clasificación. Se denomina no paramétrico porque no asume distribución de datos subyacente. Son ventajas de este algoritmo ser una técnica simple implementado fácilmente, rentable, flexible, adecuado para clases multimodales, y la tasa de error es como máximo el doble del error de Bayes. No obstante, tiene desventajas porque su aprendizaje es lento y no hace generalización sobre los datos de entrenamiento.

- **Algoritmo K-Means**

El algoritmo de agrupamiento de medios K se usa con mayor frecuencia para resolver problemas de agrupamiento, y se considera una forma de aprendizaje no supervisado. Presenta beneficios al ser computacionalmente más eficiente que el agrupamiento jerárquico cuando existe variables grandes. La atracción de uso de este algoritmo se enmarca en su facilidad de implementación e interpretación de los resultados del agrupamiento. Son desventajas del K-Means la dificultad de predecir el valor de K, bajo rendimiento cuando los cúmulos son globulares y existe diferencias en los tamaños y la densidad de los grupos de datos de entrada, la suposición esférica es difícil de satisfacer ya que rompe la correlación entre las características, por último, es sensible a los valores atípicos.

Redes neuronales artificiales

Las redes neuronales artificiales son dispositivos computacionales paralelos que procesan grupos de elementos interconectados denominados neuronas. Se caracterizan por tipologías, reglas de entrenamiento y particularidades computacionales de sus elementos. Las redes neuronales se dividen en diferentes capas, la primera de ellas se denomina capa de entrada, y es la que recibe información del exterior, correspondiente a las variables independientes empleadas como entradas. La última capa es la capa de salida, donde sus neuronas dominan la salida de la red. Las capas que se encuentran entre la entrada y salida se llaman capas ocultas. Cada capa realiza cálculos independientes y pasan los resultados a otras capas. Las redes interconectadas entre sí en la misma capa o capas anteriores se denominan redes de retroalimentación (Niculescu, 2003).

De forma sencilla, las redes neuronales reproducen la forma en que el cerebro biológico organiza, almacena y procesa la información. Estas pueden aprender utilizando solo un tipo de aprendizaje: supervisado (la información de entrada y salida se usa para entrenamiento) y no supervisado (solo la información de entrada es procesada) (Niculescu, 2003). A medida que ha ido evolucionando, las redes neuronales artificiales se han integrado a métodos avanzados como la lógica difusa y el análisis de ondículas, con el fin

de optimizar la interpretación y modelados de los datos; así evitar la subjetividad del entrenamiento en el procesamiento de los algoritmos (Huang, 2009).

Ventajas de las redes neuronales artificiales

Niculescu (2003) señala las siguientes ventajas de las redes neuronales artificiales:

- Son adaptativas, toman datos y aprenden de ellas
- No imponen condiciones a las variables pronosticas
- Son esencialmente no lineales y capaces de aprender interacciones complejas
- Proporcionan respuestas a problemas sin soluciones algorítmicas
- Capacidad de extracción de información esencial
- Tolerantes a fallas al ser capaz de manejar datos ruidosos o incompletos
- No contaminan el resultado con una simplificación excesiva o ideas preconcebidas

Desventajas de las redes neuronales artificiales

Algunas de las desventajas de las redes neuronales artificiales son (Niculescu, 2003):

- Describen asociaciones y no causas
- La transparencia del modelo suele ser subjetiva para informáticos novatos

Algoritmos no supervisados

El aprendizaje de algoritmos no supervisados encuentra patrones en los datos sin importar que estos sean ruidosos o no estructurado. En el enfoque no supervisado, las máquinas reciben entradas, pero no tienen resultados objetivos supervisados. Aunque parece misterioso, la máquina es capaz de aprender a pesar que no recibe retroalimentación de su entorno. El objetivo del aprendizaje no supervisado es construir representaciones de los datos de entrada para tomar decisiones, predecir futuras entradas, comunicar con eficiencia las entradas a otras máquinas, etc. El agrupamiento y la reducción de la dimensionalidad son dos ejemplos clásicos del aprendizaje no supervisado (Ghahramani, 2004).

Por otro lado, los algoritmos no supervisados se han considerado herramientas útiles en la ciencia de datos por permitir la agrupación en clústeres. Este método se ha aplicado para encontrar la estructura en conglomerados de un conjunto de datos caracterizados por presentar similitud (dentro del conglomerado) y mayor disimilitud (distancia) (entre conglomerados). Estadísticamente, los métodos de agrupamientos aplicados en los algoritmos de aprendizaje no supervisado se dividen en modelos de probabilidad y

enfoques no paramétricos. El punto importante en estos métodos es definir la disimilitud entre un punto y un prototipo de conglomerado (Sinaga & Yang, 2020).

DBSCAN

DBSCAN (*Density-based spatial clustering of applications with noise*, por sus siglas en inglés) es un algoritmo de agrupamiento que se fundamenta en una noción de agrupamiento basada en la densidad. Se diseñó con el fin de descubrir grupos de forma arbitraria mientras es capaz de manejar el ruido o valores atípicos de modo efectivo. DBSCAN tiene como clave que, para cada objeto de datos de un grupo, la vecindad de un radio dado (Eps) tiene que tener al menos un número mínimo ($MinPts$) (Zhou et al., 2000).

DBSCAN selecciona un valor de k -dist global para la agrupación. Para las regiones más delgadas, el número de objetos contenidos en las vecindades de sus objetos centrales con un radio Eps igual a k -dist es k (el valor predeterminado de k en DBSCAN es 4). Sin embargo, para los otros clústeres, el número de objetos contenidos en la mayoría de las vecindades de sus objetos centrales del mismo radio es más que k . El algoritmo lleva a cabo una operación de consulta de región para cada objeto contenido en la vecindad del objeto central. Para un objeto central dado p en el plumero C , es concebible que las vecindades de los objetos contenidos en p se cruzarán entre sí (Zhou et al., 2000).

Pseudocódigo de DBSCAN (Han et al., 2012)

Entradas:

D : dataset que contiene los n objetos.

E : Parámetro del radio

$MinPts$: umbral para la densidad del vecindario

Salida:

Un conjunto de clústeres basados en la densidad

Procedimiento:

- (1) Marcar todos los objetos como "no visitados"
- (2) Repetir
- (3) Seleccionar aleatoriamente un objeto "no visitado" p
- (4) Marcar p como "visitado".
- (5) Si el E -vecindario de p tiene al menos $MinPts$ objetos
- (6) Crear un nuevo clúster C , y agregar p a C
- (7) Sea N el conjunto de objetos en la vecindad de p
- (8) Para cada punto p' en N
- (9) Si p' está marcado como "no visitado"
- (10) Marcar p' como visitado
- (11) Si el E -vecindario de p' tiene al menos $MinPts$ objetos
- (12) Agregar aquellos puntos en N
- (13) Si p' no está en ningún clúster, agregar p' a C .
- (14) Fin Para
- (15) Devolver C

(16) Sino Marcar *p* como ruido

(17) Hasta que no haya objetos marcados como “no visitados”

Migración

La migración es un fenómeno complejo que involucra muchos aspectos económicos, sociales y psicológicos de la vida de las personas (Cazzuffi & Fernández, 2018). De acuerdo con Pencea & Curteanu (2020), los flujos migratorios influyen, tanto cuantitativa como cualitativamente, en la seguridad demográfica, el desarrollo social y económico de todos los países. Estos flujos migratorios tienen un papel muy importante en el funcionamiento de la economía global y aparecieron en respuesta al desarrollo desigual de la economía nacional y las disparidades en el desarrollo económico de diferentes grupos de países.

Según la Organización Internacional para las Migraciones (IOM, 2019), los migrantes son personas y/o familias que se trasladan a otro país o región de forma permanente o temporal, con el fin de mejorar sus condiciones materiales o sociales, así como sus perspectivas y/o las de su familia. Sus causas pueden ser muy variadas e ir desde los motivos políticos o económicos, hasta abarcar algún tipo de catástrofe o tragedia natural, ocasionando comúnmente consecuencias negativas como la separación familiar, la pérdida de la identidad cultural, al igual que la desaparición parcial o total de comunidades (Fong & Shibuya, 2020). Raymer & Wiśniowski (2018) aclaran que, las migraciones son vistas desde dos puntos diferentes: emigración e inmigración; la emigración hace referencia al proceso de salida o abandono del lugar de origen hacia otro, y la inmigración ocurre cuando una persona o grupo de personas llegan a vivir a un país, estado o región.

Tipos de migración

Castelli (2018), alude que, dentro de la migración, se puede clasificar según el lugar del destino, según el tiempo y según su carácter, descritas a continuación:

Según su destino. Según Williams & Efendic (2019), existen dos tipos de migración según el lugar del destino:

- Migración interna. Movimientos migratorios que tienen como destino otra localidad, municipio o entidad del mismo país.
- Migración externa. Surge cuando se va de un país a otro.

Según tiempo. Chen et al. (2019), consideran dos tipos diferentes de migración según tiempo, que pueden ser empleados por diferentes individuos en el mismo período o en diferentes períodos:

- Temporal. Ocurre cuando el individuo está limitado por el tiempo, tienden a realizar viajes cortos y luego regresan al lugar de origen.
- Permanente. Ocurre cuando el individuo se muda de un lugar a otro y no tiene planes de regresar a su hogar original, tienden a recorrer distancias más largas sin intención de regresar.

Según carácter. Becker & Ferrara (2019), mencionan que las migraciones pueden ocurrir por las siguientes razones:

- Migración forzada. Cuando las condiciones de vida en el lugar de origen no permiten seguir viviendo en aquel lugar, por ejemplo: las guerras, conflictos políticos, desastres naturales, pobreza, falta de oportunidades, etc.
- Migración voluntaria. Cuando el migrante decide libremente irse a otro lugar, con el deseo de tener mejores oportunidades, buscar la felicidad y tranquilidad, por ejemplo: razones académicas, razones familiares, razones laborales, etc.

Consecuencias de la migración

De acuerdo a León Castillo (2015), los flujos migratorios producen una serie de consecuencias sociales, económicas y culturales, de manera positiva y negativamente tanto en el lugar de origen como en el lugar receptor de migrantes, donde se colocan de manifiesto oportunidades de crecimiento y desarrollo si el fenómeno migratorio es bien manejado; de lo contrario, surgirán problemas que atentan contra la estabilidad de la persona y el resto de la población, aunado a los desequilibrios económicos que se generan.

Para Gutiérrez Silva et al. (2020), el acto de migrar no sólo implica el cruce de una frontera, sino que también lleva a la instalación parcial o permanente de la persona en el país escogido para residir, donde tendrá que trabajar, consumir bienes y servicios, compartir costumbres, asistir a servicios de atención médica, solicitar seguridad física y posiblemente la nacionalización con el pasar del tiempo (si es el caso), entre otras atenciones que demandará.

Entre las consecuencias de la migración para el país de origen se describen a continuación:

- Impacto positivo: Disminución del conflicto social y político, puestos de trabajos disponibles, obtención de remesas y transferencia de tecnología, entre otros.
- Impacto negativo: Desequilibrio en el tamaño, distribución y composición de la población; pérdida de mano de obra calificada y/o joven; pérdida de ingresos fiscales, entre otros.

Entre las consecuencias de la migración para el país de destino se describen a continuación:

- Impacto positivo: Incremento de mano de obra calificada y/o joven, resurgimiento de sectores productivos, incremento de recaudación de impuestos; entre otros.
- Impacto negativo: Aumento de la competencia laboral; discriminación, xenofobia, explotación; problemas de integración y adaptación; entre otros.

Definición de términos básicos

Algoritmo: Un algoritmo es un “objeto A que satisface los postulados de secuenciación temporal, de estados abstractos y de exploración delimitada.” (p. 321) (Guerevich, 2000 citado en Mota, 2015).

Base de datos: Las bases de datos contienen datos de procedencias, características y tipologías diferentes. Pueden ser relacionales o no relacionales, además almacenan grandes cantidades de información.

Clúster: En la aplicación de algoritmos automáticos, un clúster representa el agrupamiento de datos típicos o atípicos, y es útil para detectar conglomerados y analizar tendencias de grupos relacionados.

Datos: Los datos identifican cuanti o cualitativamente eventos ocurridos. En informática, se asigna como dato a cualquier palabra, letra o cifra que ingresan a las computadoras, y son procesados o transformados para devolver información (Favaretto et al., 2019).

Eficiencia computacional: Es la forma de cuantificar a un algoritmo mediante el cálculo de datos procesados y los recursos usados para ejecutar los trabajos programados (Hernandez & Brown, 2020).

Inteligencia artificial: Sirve para simular comportamientos inteligentes de los humanos en computadoras, como el juicio, la toma de decisión y el razonamiento, siendo su objetivo replicar con gran precisión las actividades intelectuales humanas (Zhang & Lu, 2021).

Minería de datos: Es la práctica de examinar una gran cantidad de datos preexistentes para crear nueva información. Se centra en técnicas de agrupamiento, decisión, predicciones y redes neuronales (Sidow Osman, 2019).

Patrones: Los patrones de datos capturan diversas formas en que los datos se presentan y utilizan en los flujos de información. Su determinación se concibe a partir de aplicación de tecnologías y lenguajes de modelado específico (Russell et al., 2005).

Predicción de eventos: Se genera a partir de la minería de series temporales de datos. Tiene un papel importante en la toma de decisiones al visionar comportamientos o actividades eventuales que puedan ocurrir en la vida real (Molaei & Keyvanpour, 2015).

CAPÍTULO III MATERIALES Y MÉTODOS

3.1. Ámbito y condiciones de la investigación

3.1.1 Contexto de la investigación

La presente investigación se llevó a cabo en formato gabinete, ya que no se requirió realizar intervenciones en campo. Sin embargo, la información se obtuvo mediante la base de datos de la Superintendencia Nacional de Migraciones, previa autorización correspondiente.

3.1.2 Periodo de ejecución

El estudio fue desarrollado durante cuatro meses, de enero a abril de 2023.

3.1.3 Autorizaciones y permisos

Se realizó una solicitud de acceso a la información pública sobre los inmigrantes registrados por la Superintendencia Nacional de Migraciones, en cumplimiento de la Ley N° 27806 - Ley de Transparencia y Acceso a la Información Pública. Dicha solicitud fue presentada el 11 de marzo de 2022 y se recibió la respuesta correspondiente el 16 de marzo del mismo año (ver Anexo 3).

3.1.4 Control ambiental y protocolos de bioseguridad

No aplica.

3.1.5 Aplicación de principios éticos internacionales

En esta investigación, se aplicó el principio ético de beneficencia, el cual implica la responsabilidad de maximizar los beneficios y minimizar los riesgos para los participantes o sujetos de estudio. En el contexto de esta tesis, aplicar este principio implica garantizar que la aplicación del algoritmo no supervisado genere resultados que sean valiosos y útiles para la Superintendencia Nacional de Migraciones, especialmente en el proceso de toma de decisiones relacionadas con la otorgación de la calidad migratoria.

3.2. Sistema de variables

3.2.1 Variables principales

Variable independiente: Aplicación de un algoritmo no supervisado

Variable dependiente: Clasificación de extranjeros según riesgo de inmigración irregular

Tabla 1.

Descripción de variables por objetivo general

Objetivo general: Evaluar la aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones.			
Variable abstracta	Variable concreta	Medio de registro	Unidad de medida
VI: Aplicación de un algoritmo no supervisado	DBSCAN	DBSCAN	Número (Silhouette)
Variable abstracta	Variable concreta	Medio de registro	Unidad de medida
VD: Clasificación de extranjeros según riesgo de inmigración irregular	Índices internos:		
	Cohesión		
	Separación		
	Coefficiente de Silhouette		
	Correlación		
	Índices estadísticos:		
	Desviación estándar		
	Media		
	Varianza		
	Asimetría		
	Error estándar de asimetría	Ficha de registro	Numérico
	Curtosis		
	Rango		
	Mínimo		
	Máximo		
	Suma		
	Percentiles		
Mediana			
Moda			
Error estándar de la media			

3.2.2 Variables secundarias

No corresponde.

3.3 Procedimientos de la investigación

a) Tipo y nivel de investigación

La investigación fue de tipo básica, ya que se buscó aplicar teorías existentes para abordar un problema práctico y concreto relacionado con la clasificación de extranjeros según el riesgo de inmigración irregular mediante un algoritmo no supervisado. Se utilizó un enfoque cuantitativo, donde se procesaron datos recopilados de la Superintendencia Nacional de Migraciones y se aplicaron técnicas de análisis convencionales y de aprendizaje automático para obtener agrupaciones. El nivel de investigación correspondió al descriptivo, ya que se describió la caracterización de los datos de los extranjeros para establecer un comportamiento o estructura que los clasificara según el riesgo de inmigración.

b) Diseño de investigación

Se utilizó un diseño 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. No se ejerció control sobre las variables. En este sentido, el esquema que mejor se adecuó al diseño del estudio fue:



Donde M representaba el conjunto de datos procesados de los extranjeros y O_{VD} hacía referencia a la observación, análisis y descripción de los resultados obtenidos mediante el uso de DBSCAN para la clasificación de extranjeros según el riesgo de inmigración irregular.

c) 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 se aplicó un algoritmo no supervisado, se consideró que, a mayor cantidad de datos, los resultados de las agrupaciones o clústeres serían más confiables y precisos. 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 fue 31 625, empleando un muestreo no probabilístico por conveniencia.

3.3.1 Objetivo específico 1

Para generar un dataset consistente con el cual agrupar extranjeros según el riesgo de inmigración irregular, inicialmente, se accedió a los registros proporcionados por la Superintendencia Nacional de Migraciones, los cuales contienen información detallada sobre las solicitudes de cambio de calidad migratoria, así como otros datos demográficos y socioeconómicos relevantes. Estos registros fueron recopilados en un formato estructurado, lo que permitió su manipulación y análisis posterior. Luego, se procedió a realizar una exploración y limpieza de los datos, eliminando aquellos campos que presentaban poca o nula relevancia para el objetivo del estudio. También se realizaron ajustes y transformaciones necesarios para garantizar la integridad y coherencia de los datos, como la eliminación de registros duplicados o la imputación de valores faltantes en algunas variables clave.

Posteriormente, se llevó a cabo una selección de las variables más relevantes para la clasificación del riesgo de inmigración irregular. Esto implicó identificar aquellas características que podrían tener un impacto significativo en la determinación del riesgo, como la nacionalidad del individuo, la duración de su estancia en el país, los antecedentes migratorios previos, entre otros. Una vez realizada la selección de variables, se procedió a la transformación de los datos en un formato adecuado para su análisis. Esto incluyó la codificación de variables categóricas y la normalización de variables numéricas. Finalmente, se consolidaron todos los datos seleccionados y transformados en un dataset final en formato CSV (delimitado por comas), que fue utilizado como base para los análisis posteriores de agrupación de extranjeros según el riesgo de inmigración irregular.

3.3.2 Objetivo específico 2

A diferencia de otros algoritmos de agrupamiento, como K-means, DBSCAN no requiere especificar previamente el número de clusters a encontrar. En cambio, se basa en la densidad de los puntos en el espacio de datos para agruparlos. La aplicación de DBSCAN implica la configuración de dos parámetros clave: epsilon (eps) y min_pts. El parámetro epsilon define la distancia máxima entre dos puntos para que se consideren vecinos y se agrupen juntos. El parámetro min_pts especifica el número mínimo de puntos requeridos en una vecindad para que se forme un cluster.

Con el fin de seleccionar los parámetros eps y min_pts óptimos, se realizó un análisis exhaustivo para determinar dichos valores y que estos, permitieran obtener resultados significativos y representativos. Para ello, llevamos a cabo un proceso iterativo en el cual, utilizando diferentes combinaciones de valores para estos parámetros, se evaluaron los

resultados en función de las métricas de evaluación de clústeres, como el coeficiente de Silhouette, ruido y número de clústeres generados, seleccionando aquellos con los valores más óptimos en cada criterio. El coeficiente de Silhouette fue una métrica clave utilizada para medir la calidad de los clusters obtenidos. Este coeficiente se calcula para cada punto y proporciona una medida de qué tan bien se ajusta el punto a su propio cluster en comparación con los clusters vecinos más cercanos. Además de los resultados de clustering, también se calcularon los índices estadísticos para cada cluster individual y para todo el conjunto de datos. Los índices permitieron obtener una comprensión más profunda de las características y distribución de los datos en cada cluster, ayudando a enriquecer la interpretación de los resultados.

3.3.3 Objetivo específico 3

Se aplicó el coeficiente de Silhouette para evaluar la calidad de los clusters generados por el algoritmo DBSCAN en la clasificación de extranjeros según el riesgo de inmigración irregular. Dicho coeficiente proporcionó una medida cuantitativa de la calidad de los clusters obtenidos, que permitió evaluar y comparar diferentes configuraciones de parámetros. Ello facilitó la selección de los parámetros más adecuados para el algoritmo DBSCAN en este contexto específico, y proporcionó una base sólida para la toma de decisiones en la clasificación de extranjeros según el riesgo de inmigración irregular. Finalmente, tras obtener grupos bien constituidos, se procedió a etiquetar los clústeres generados con ayuda de los indicadores estadísticos de cada una de las agrupaciones obteniendo como resultado el modelo final.

La figura 1 muestra gráficamente el procedimiento descrito en los puntos anteriores.



Figura 1

Flujo de procedimientos

CAPÍTULO IV RESULTADOS Y DISCUSIÓN

4.1 Resultado específico 1: Construir el dataset de clasificación de extranjeros según riesgo de inmigración irregular reportados por la Superintendencia Nacional de Migraciones

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

Paso 1: Se cargó 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.

```
from google.colab import drive
drive.mount('/gdrive')
%cd /gdrive/
%cd MyDrive/Migraciones

# Cargamos el dataset desde el directorio en la nube
ruta_archivo = '/gdrive/MyDrive/Migraciones/dataset_0.csv'
df = pd.read_csv(ruta_archivo)
```

Paso 2: Se realizó un proceso de limpieza y preprocesamiento de los datos. Se eliminaron 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 se descartaron los campos relacionados con los servicios de internet, servicios básicos, tipo de vivienda y situación de salud.

```
# Eliminamos los campos de poca y nula relevancia
df = df.drop(columns='Ord')
df['Trabaja'].dropna(inplace=True)
df['Profesion'].dropna(inplace=True)
df.drop(columns='Fecha de ingreso', inplace=True)
df.drop(columns='Profesion', inplace=True)
df.drop(columns='Estado civil', inplace=True)
df.drop(columns='Embarazo', inplace=True)
df.drop(columns='Actividad económica', inplace=True)
df.drop(columns='Nivel Educativo', inplace=True)
```

```
df.drop(columns='Grado superior', inplace=True)
df.drop(columns='Discapacidad', inplace=True)
df.drop(columns='Disc Física', inplace=True)
df.drop(columns='Disc Sensorial', inplace=True)
df.drop(columns='Disc Cognitiva', inplace=True)
df.drop(columns='Disc Intelectual', inplace=True)
df.drop(columns='Servicio internet', inplace=True)
df.drop(columns='Servicios básicos', inplace=True)
df.drop(columns='Tipo de vivienda', inplace=True)
df.drop(columns='Situación de salud', inplace=True)
```

Paso 3: Se examinó la distribución de los datos en la columna "País de nacimiento" para comprender la procedencia de los migrantes. Se generó un gráfico de barras que mostraba la cantidad de migrantes por país de nacimiento (Figura 1).

```
# Analizamos las distribuciones de los datos dentro de la columna "País de nacimiento"
d_pais_nacimiento = df['País de nacimiento'].value_counts()
d_pais_nacimiento.sort_values(ascending=False)
plt.figure(figsize=(12, 12))
d_pais_nacimiento.plot.bar()
plt.title('Distribución: País de nacimiento')
plt.xlabel('País')
plt.ylabel('Cantidad')
plt.show()
```

Ver página siguiente.

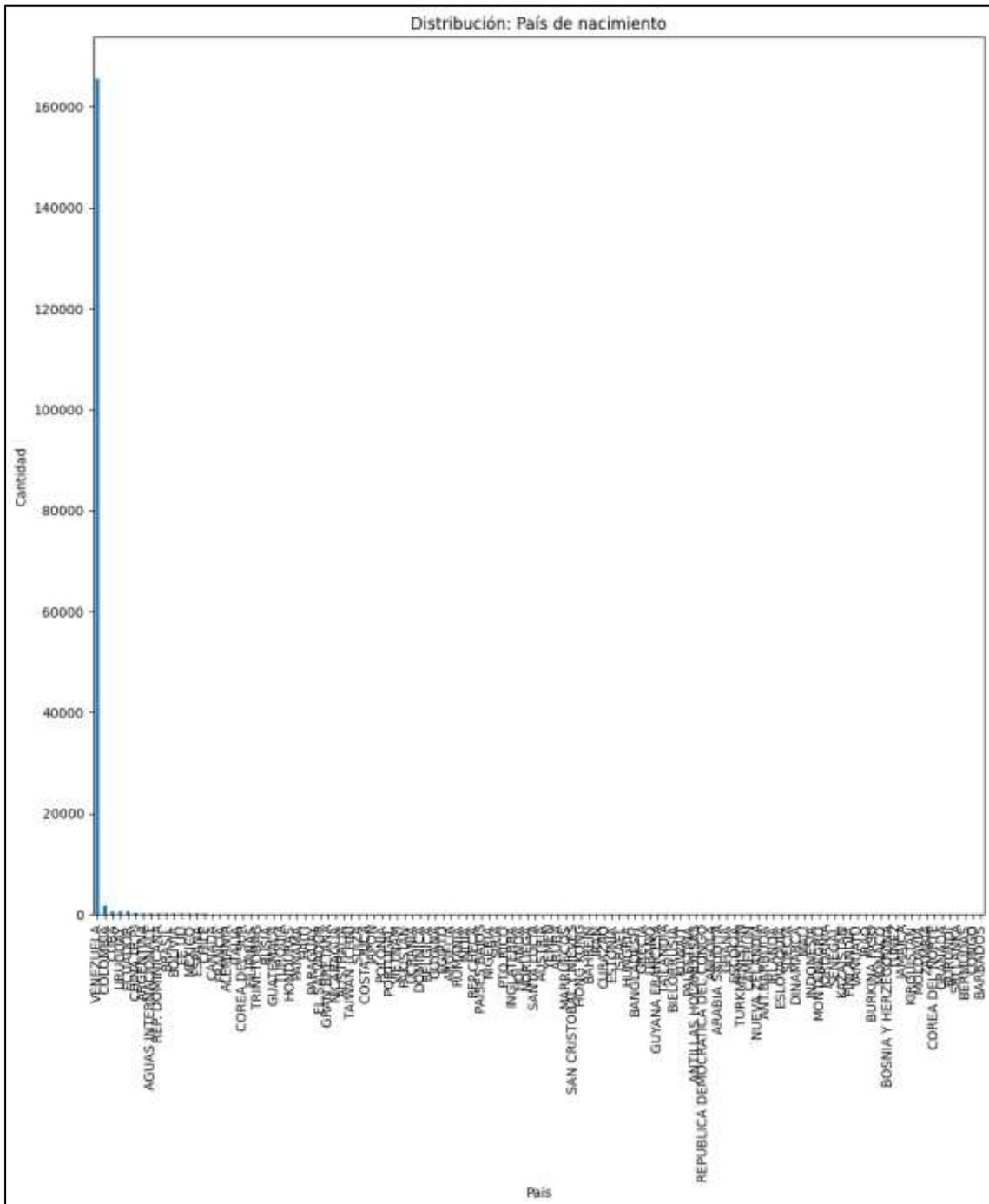


Figura 2

Distribución de los datos según el país de nacimiento

Paso 4: Para la construcción del dataset, se estableció como objetivo identificar el riesgo de inmigración irregular en cinco países de interés: Venezuela, Colombia y Cuba, Ecuador, Uruguay. Se filtraron los registros que no correspondían a la calidad migratoria de turista y se eliminó la columna "Calidad migratoria". Luego, se seleccionaron únicamente los

migrantes nacidos en los países mencionados utilizando la función `isin()` de pandas (Figura 2).

```

países_control = ['VENEZUELA', 'COLOMBIA', 'CUBA']
df = df.drop(df.loc[df['Calidad migratoria']!= 'TURISTA'].index)
df.drop(columns='Calidad migratoria', inplace=True)
df = df.drop(df.loc[~df['País de nacimiento'].isin(países_control)].index)

```

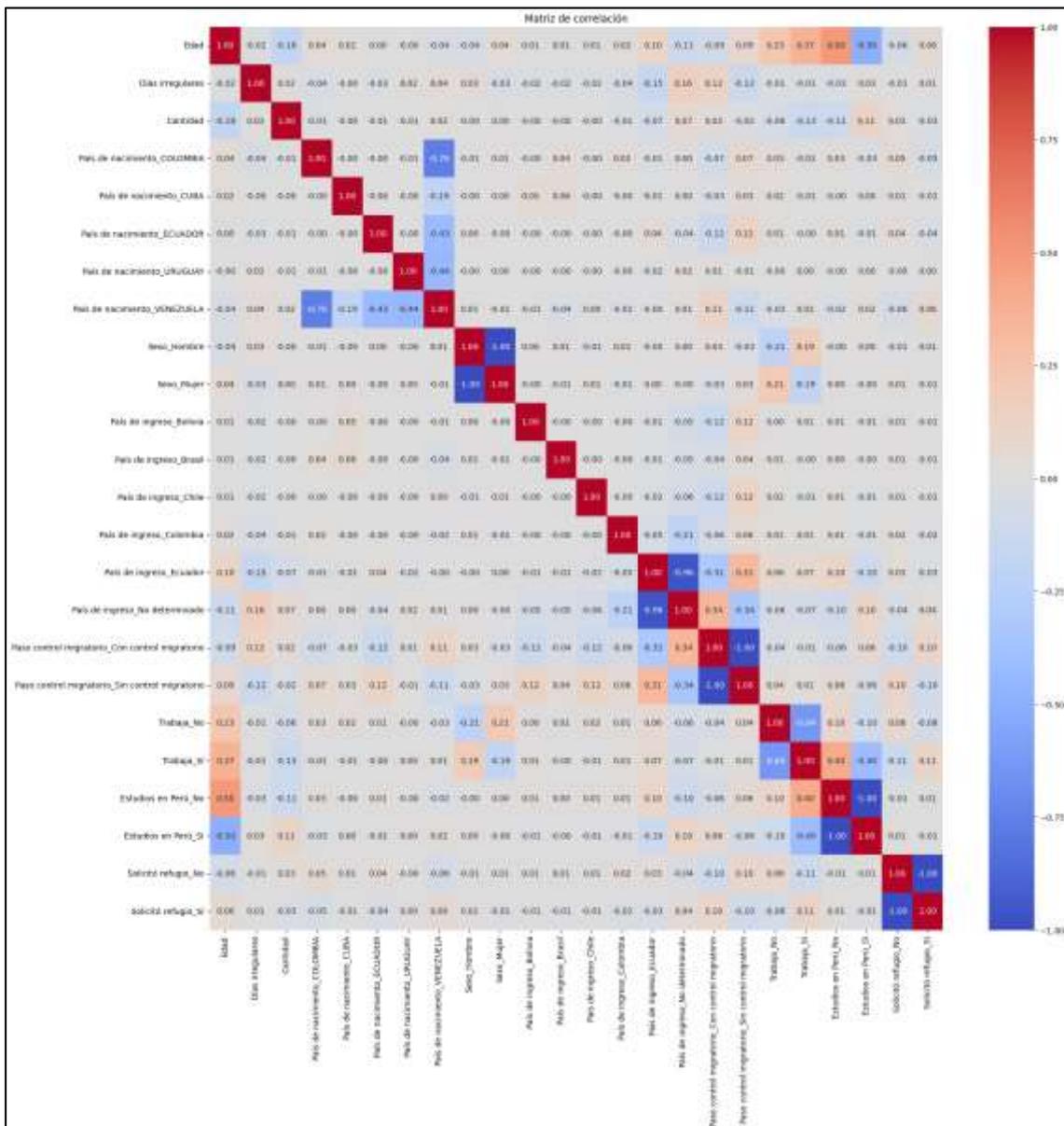


Figura 3
Matriz de correlación de los datos

Paso 5: Se convirtieron las variables categóricas en valores numéricos utilizando el enfoque de one-hot encoding mediante la función `pd.get_dummies()`. Esto permitió

representar las características categóricas como variables binarias, lo cual es necesario para utilizar algoritmos de clasificación.

```
df = pd.get_dummies(df)
df.head()
```

Paso 6: Con el objetivo de mejorar el rendimiento del modelo de clasificación y reducir la dimensionalidad del dataset, se aplicó una técnica de escalamiento de características utilizando `StandardScaler()` para estandarizar los valores de los atributos.

```
# Normalizamos los datos
dataset = StandardScaler().fit_transform(df)
df_normalized = pd.DataFrame(dataset, columns = df.columns)
matriz_correlacion = df_normalized.corr()
```

Paso 7: Se utilizó 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.

```
# Reducimos la dimensionalidad del dataset de 22 a 10 con ayuda del Analisis de Componentes
Principales (PCA)
pca = PCA(n_components = 10)
pca.fit(dataset);
dataset = pca.transform(dataset)
```

4.2 Resultado específico 2: Aplicar DBSCAN (Agrupamiento Espacial Basada en la Densidad de Aplicaciones con Ruido) como algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

Para desarrollar el segundo objetivo específico, se continuó con los pasos procedimentales de aplicación del algoritmo DBSCAN:

Paso 8: Con ayuda del dataset obtenido en los pasos previos, se llevó cabo la aplicación del algoritmo DBSCAN sobre dicho set de datos. Este proceso se realizó en dos etapas principales. En la primera etapa, se efectuó iteraciones utilizando diferentes valores de EPS (epsilon) y Min_pts (número mínimo de puntos) para explorar diferentes configuraciones de agrupamiento. Se calcularon métricas como el número de clusters, el coeficiente de Silhouette y la cantidad de puntos considerados como ruido para evaluar la calidad de cada configuración.

```

# Calcular valores EPS y Min_pts
valores_eps = np.arange(0.5, 3.6, 0.5)
min_pts = np.arange(95, 351, 10)
params_dbscan = list(product(valores_eps, min_pts))
nro_clusters = []
pts_sil = []
pts_ch = []
pts_db = []
_valores_eps = []
_pts_min = []
ruido = []
for p in params_dbscan:
    dbscan = DBSCAN(eps=p[0], min_samples=p[1]).fit(dataset)
    n_clusters_ = len(set(dbscan.labels_) - (-1 if -1 in dbscan.labels_ else 0))
    _valores_eps.append(p[0])
    _pts_min.append(p[1])
    nro_clusters.append(n_clusters_)
    ruido.append(list(dbscan.labels_).count(-1))
    _cluster = eliminar_ruido(dbscan.labels_, dataset)
    pts_sil.append(silhouette_score(_cluster[1], _cluster[0]))

eps_min = list(zip(nro_clusters, pts_sil, _valores_eps, _pts_min, ruido))
eps_min_df = pd.DataFrame(eps_min, columns=['Number of clusters', 'Silhouette score', 'EPS',
'Min pts', 'Noise'])

```

Paso 9: En la segunda etapa, se seleccionó 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 (0.533846492). Los parámetros estimados se muestran en las figuras 3, 4 y 5.

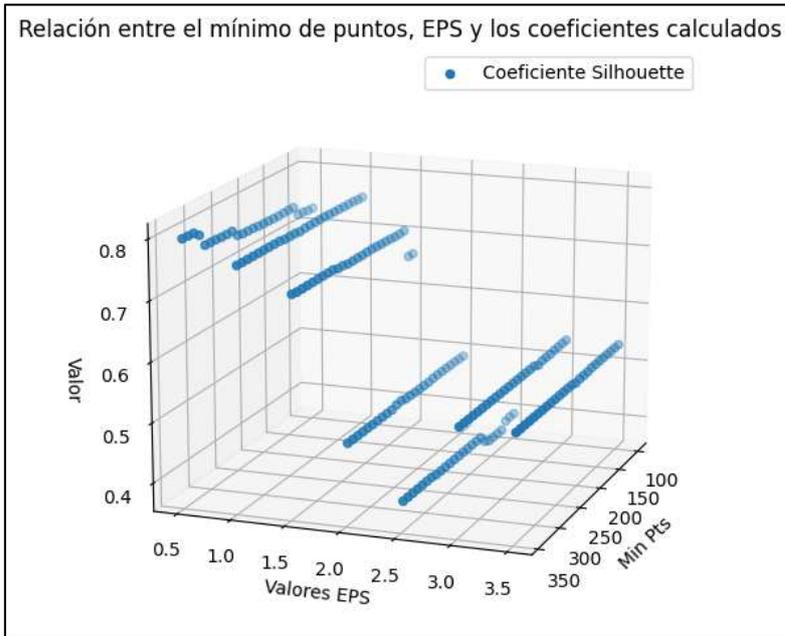


Figura 4

Relación entre los parámetros (EPS y Min_pts) y el coeficiente de Silhouette alcanzado

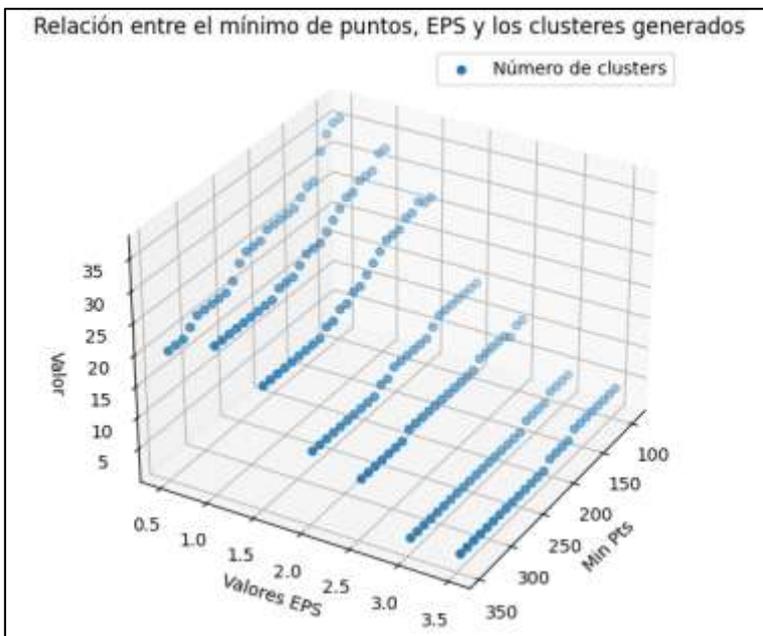


Figura 5

Relación entre los parámetros (EPS y Min_pts) y el número de clusteres generados

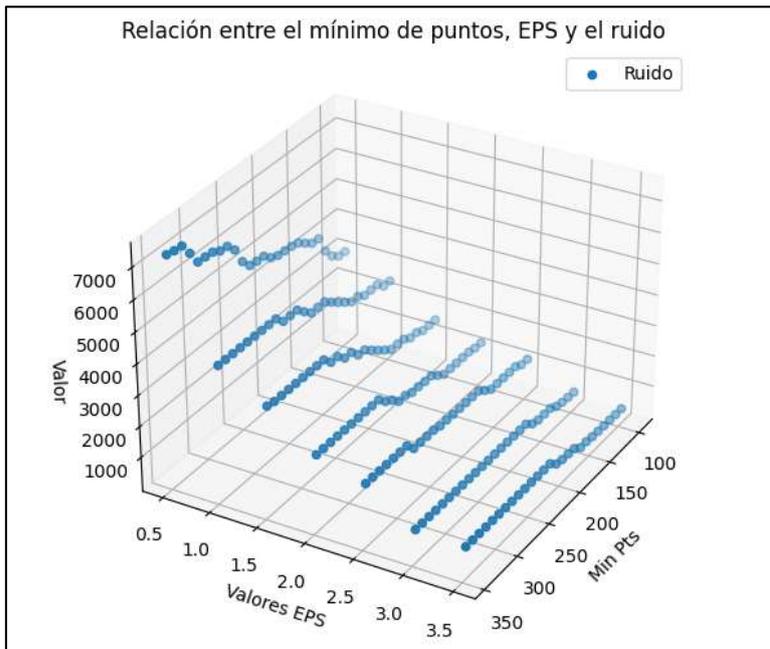


Figura 6

Relación entre los parámetros (EPS y Min_pts) y el número de puntos ruido

Paso 10: Tras seleccionar los hiperparámetros del modelo, se aplicó nuevamente el algoritmo DBSCAN utilizando estos valores. Los resultados obtenidos fueron guardados en un archivo Excel denominado 'resultados_DBSCAN.xlsx'. Este archivo contiene la información clasificada de los extranjeros, incluyendo sus atributos originales y las etiquetas de cluster asignadas por el algoritmo DBSCAN.

```
dbscan = DBSCAN(eps=3.5, min_samples=125)
dbscan.fit(dataset)
labels = dbscan.labels_
df['cluster'] = labels
df.to_excel('resultados_DBSCAN.xlsx', index=False)
```

Paso 11: Para una mejor visualización de los resultados, se aplicó el análisis de componentes principales (PCA) para reducir la dimensionalidad del conjunto de datos a tres componentes principales. Luego, se creó un gráfico en 3D que representaba los diferentes clusters identificados por el algoritmo DBSCAN (Figura 6).

```
pca = PCA(n_components = 3)
pca.fit(dataset);
dataset_reduced_pca = pca.transform(dataset)
```

```
dataset_reduced_df = pd.DataFrame(dataset_reduced_pca, columns=['Componente 1',  
'Componente 2', 'Componente 3'])  
dataset_reduced_df['cluster'] = labels
```

```
# Personalizamos el gráfico  
fig = plt.figure(figsize=(10, 10))  
ax = fig.add_subplot(111, projection='3d')  
clusters = dataset_reduced_df['cluster'].unique()  
for cluster in clusters:  
    cluster_points = dataset_reduced_df[dataset_reduced_df['cluster'] == cluster]  
    ax.scatter(cluster_points['Componente 1'], cluster_points['Componente 2'],  
cluster_points['Componente 3'], label=f'Cluster {cluster}')
```

```
ax.set_xlabel('Componente 1')  
ax.set_ylabel('Componente 2')  
ax.set_zlabel('Componente 3')  
ax.set_title('DBSCAN Clusters')  
ax.legend()  
plt.show()
```

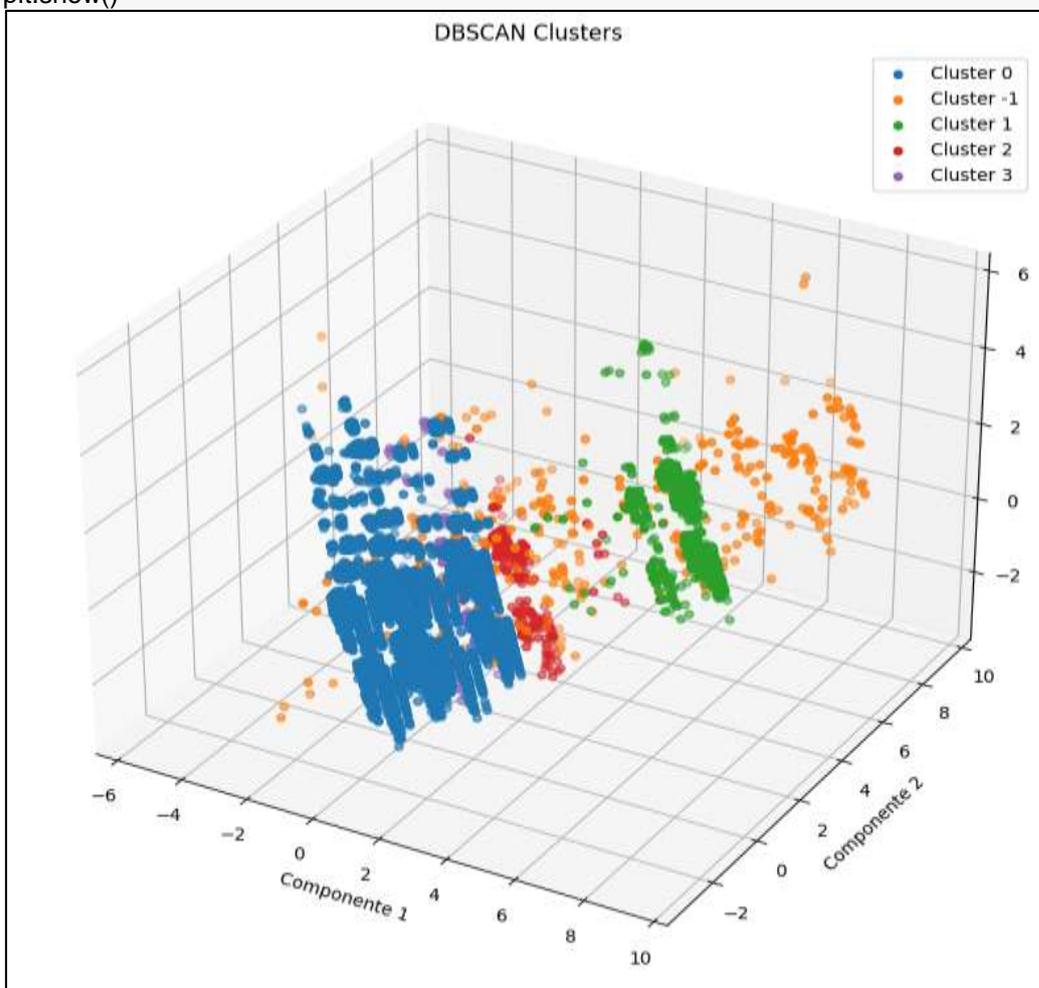


Figura 7

Representación tridimensional de los clústeres generados

El enfoque de selección de parámetros durante el presente estudio se alinea fuertemente al enfoque utilizado por Jin-Heng et al. (2022), en el que 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.

4.3 Resultado específico 3: Determinar el coeficiente de Silhouette de la aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

Después de haber explorado las diferentes combinaciones de parámetros de valor EPS (épsilon) y Min_pts (puntos mínimos), se calcularon múltiples métricas de evaluación para cada combinación de parámetros, entre ellas, el coeficiente de Silhouette. Este coeficiente es una medida de la cohesión intra-cluster y la separación inter-cluster de los puntos de datos. Un valor más cercano a 1 indica una clasificación más adecuada y bien definida, mientras que un valor cercano a -1 indica una clasificación deficiente. Cabe destacar la remoción del conjunto de datos considerados como ruido para evitar involucrar dichos datos dentro de la sumatoria final de cada cluster.

```
def eliminar_ruido(arr1, arr2):
    for i, e in reversed(list(enumerate(arr1))):
        if(e==-1):
            arr1 = np.delete(arr1, i, axis=0)
            arr2 = np.delete(arr2, i, axis=0)
    return [arr1, arr2]

_cluster = eliminar_ruido(dbscan.labels_, dataset)
silhouette_score(_cluster[1], _cluster[0])
```

Dados los parámetros seleccionados, se obtuvieron los siguientes coeficientes:

Tabla 2.

Estadísticos de los clústeres

Cluster	0	1	2	3
Silhouette	0.419657598	0.566275991	0.388101172	0.581932093
Desviación estándar	9130.394528	8997.536375	9170.22081	9146.51102
Media	15809.9114	15902.59277	15994.83178	16113.76547
Varianza	83364104.25	80955660.81	84092949.71	83658663.83
Asimetría	-0.00053951	0.001172007	0.012051062	-0.081774338
Error estándar de asimetría	-3.12309E-06	4.0035E-05	0.000823794	-0.004667108
Curtosis	-1.200544918	-1.165183548	-1.202240364	-1.243580606
Rango	31624	31582	31588	31374

Mínimo		0	15	18	104
Máximo		31624	31597	31606	31478
Suma		471799376	13628522	3422894	4946926
	25%	7906.25	8185	7477.25	8356
Percentiles	50%	15812.5	15750	16603	16383
	75%	23715.75	23710	23713.5	24168.5
Mediana		15812.5	15750	16603	16383
Moda		0	15	18	104
Error estándar de la media		52.8537226	307.3499877	626.8633894	522.0189565

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 cluster y buena separabilidad respecto a los elementos de los demás, y con ello aceptar la hipótesis planteada: 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 los de Anitha & Patil (2022) en el cual se alcanzó un coeficiente de Silhouette de 0.362159752.

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 cluster

Tabla 3.

Distribución media de las características de inmigrantes irregulares según cluster

Cluster	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
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

En función a la distribución media de los datos, es posible reconocer ciertos patrones en los datos que sugieren un mayor riesgo de inmigración irregular, tales como la cantidad de días irregulares y si paso por control migratorio, así como su relación entre los demás tipos de datos.

Tabla 4.

Nivel de riesgo de inmigración irregular por cluster

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

CONCLUSIONES

1. Se logró 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 como algoritmo no supervisado para la clasificación de extranjeros en riesgo de inmigración irregular es una estrategia efectiva.
2. Se construyó un dataset de clasificación de extranjeros según el riesgo de inmigración irregular, utilizando los reportes proporcionados por la Superintendencia Nacional de Migraciones. Se llevaron a cabo las adecuaciones y limpieza de los datos, eliminando campos de poca relevancia y nulos. Además, se realizó una exploración de los datos, identificando la distribución de los países de origen de los extranjeros.
3. Se aplicó el algoritmo no supervisado DBSCAN (Agrupamiento Espacial Basado en la Densidad de Aplicaciones con Ruido) para llevar a cabo la clasificación de los extranjeros según el riesgo de inmigración irregular. Para ello, se seleccionó parámetros óptimos, como el valor de epsilon (eps) y el número mínimo de puntos (min_pts), mediante la evaluación de diferentes combinaciones de estos valores y la posterior selección de los más adecuados, siendo 3.5 para el valor de eps y 125 para el valor del parámetro min_pts.
4. Se determinó el coeficiente de Silhouette como medida para evaluar la calidad de los clusters generados por el algoritmo DBSCAN. Mediante el cálculo de este coeficiente para cada punto de datos y su posterior promediado, se obtuvo el coeficiente de Silhouette global (0.534), que brinda una medida general de la calidad de los clusters generados.

RECOMENDACIONES

1. Se recomienda continuar recopilando y actualizando los datos de la Superintendencia Nacional de Migraciones para asegurar el mantenimiento del dataset utilizado en el estudio con el fin de no mermar la calidad de la clasificación ante los cambios y nuevas tendencias en la inmigración irregular.
2. Se sugiere evaluar otros algoritmos no supervisados y comparar sus resultados con el propósito de explorar diferentes enfoques y poder determinar el más adecuado para la clasificación de extranjeros según el riesgo de inmigración irregular.
3. Se aconseja realizar análisis adicionales para identificar las características más relevantes que contribuyan al riesgo de inmigración irregular, los cuales podrían incluir técnicas de selección de características o análisis de importancia de las variables. De esta manera, se proporcionaría información valiosa para comprender los factores que influyen en este riesgo y mejorar su detección temprana.
4. Finalmente, se recomienda utilizar de este tipo de modelos para desarrollar y fortalecer medidas preventivas y políticas migratorias enfocadas en la reducción del riesgo de inmigración irregular. Aprovechar la clasificación realizada para implementar controles más eficientes, mejorar la vigilancia en fronteras y desarrollar estrategias de intervención específicas para los grupos identificados como de mayor riesgo.

REFERENCIAS BIBLIOGRÁFICAS

- Abel, G. J., Brottrager, M., Crespo Cuaresma, J., & Muttarak, R. (2019). Climate, conflict and forced migration. *Global Environmental Change*, *54*, 239–249.
<https://doi.org/10.1016/j.gloenvcha.2018.12.003>
- Anitha, P., & Patil, M. M. (2022). RFM model for customer purchase behavior using K-Means algorithm. *Journal of King Saud University - Computer and Information Sciences*, *34*(5), 1785–1792. <https://doi.org/10.1016/J.JKSUCI.2019.12.011>
- Azizi, S., & Yektansani, K. (2020). Artificial Intelligence and Predicting Illegal Immigration to the USA. *International Migration*, *58*(5), 183–193.
<https://doi.org/10.1111/imig.12695>
- Becker, S. O., & Ferrara, A. (2019). Consequences of forced migration: A survey of recent findings. *Labour Economics*, *59*, 1–16. <https://doi.org/10.1016/j.labeco.2019.02.007>
- Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*(6370), 1530–1534.
<https://doi.org/10.1126/science.aap8062>
- Carling, J., & Schewel, K. (2018). Revisiting aspiration and ability in international migration. *Journal of Ethnic and Migration Studies*, *44*(6), 945–963.
<https://doi.org/10.1080/1369183X.2017.1384146>
- Castelli, F. (2018). Drivers of migration: why do people move? *Journal of Travel Medicine*, *25*(1). <https://doi.org/10.1093/jtm/tay040>
- Cazzuffi, C., & Fernández, J. (2018). *Rural youth and migration in Ecuador, Mexico and Peru*. Serie documento de trabajo N° 235. Programa Jóvenes Rurales, Territorios y Oportunidades: Una estrategia de diálogos de políticas.
- Chandy, R., Hassan, M., & Mukherji, P. (2017). Big Data for Good: Insights from Emerging Markets. *Journal of Product Innovation Management*, *34*(5), 703–713.
<https://doi.org/10.1111/jpim.12406>
- Chen, J. J., Kosec, K., & Mueller, V. (2019). Temporary and permanent migrant selection: Theory and evidence of ability-search cost dynamics. *Review of Development Economics*, *23*(4), 1477–1519. <https://doi.org/10.1111/rode.12617>

- Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. *Journal of Big Data*, 6(1), 12. <https://doi.org/10.1186/s40537-019-0177-4>
- Fidan, H., & Erkan Yuksel, M. (2022). A comparative study for determining Covid-19 risk levels by unsupervised machine learning methods. *Expert Systems with Applications*, 190, 116243. <https://doi.org/10.1016/j.eswa.2021.116243>
- Fong, E., & Shibuya, K. (2020). Migration Patterns in East and Southeast Asia: Causes and Consequences. *Annual Review of Sociology*, 46(1), 511–531. <https://doi.org/10.1146/annurev-soc-121919-054644>
- Ghahramani, Z. (2004). Unsupervised Learning. *Advanced Lectures on Machine Learning*, 72–112. https://doi.org/10.1007/978-3-540-28650-9_5
- Gutiérrez Silva, J. M., Romero Borré, J., Arias Montero, S. R., & Briones Mendoza, X. F. (2020). Migración: Contexto, impacto y desafío. Una reflexión teórica. *Revista de Ciencias Sociales*, 26(2), 299–313.
- Gutiérrez Silva, J., Romero Borré, J., Arias Montero, S., & Briones Mendoza, X. (2020). Migración: Contexto, impacto y desafío. Una reflexión teórica. *Revista de Ciencias Sociales*, 26(2). <https://doi.org/10.31876/rcs.v26i2.32443>
- Han, J., Kamber, M., & Pei, J. (2012). Introduction. In *Data Mining* (pp. 1–38). Elsevier. <https://doi.org/10.1016/B978-0-12-381479-1.00001-0>
- Hernandez, D., & Brown, T. B. (2020). *Measuring the Algorithmic Efficiency of Neural Networks*.
- Hossain, M. M., Sun, X., Mitran, E., & Rahman, M. A. (2021). Investigating fatal and injury crash patterns of teen drivers with unsupervised learning algorithms. *IATSS Research*, 45(4), 561–573. <https://doi.org/10.1016/j.iatssr.2021.07.002>
- Huang, Y. (2009). Advances in Artificial Neural Networks – Methodological Development and Application. *Algorithms*, 2(3), 973–1007. <https://doi.org/10.3390/alg02030973>
- INEI. (2018). Capítulo 3: Situación migratoria. In *Condiciones de Vida de la Población Venezolana que reside en Perú*. Instituto Nacional de Estadística e Informática.
- IOM. (2019). *Glossary on Migration*. International Organization for Migration.
- IOM's GMDAC. (2021). *Poblaciones de migrantes internacionales*. Estadísticas de La Inmigración y Emigración.

<https://www.migrationdataportal.org/es/themes/poblaciones-de-migrantes-internacionales>

- Jin-Heng, G., Jia-Xiang, L., Zhen-Chang, Z., & Han-Yu, L. (2022). CDBSCAN: Density clustering based on silhouette coefficient constraints. *2022 International Conference on Computer Engineering and Artificial Intelligence (ICCEAI)*, 600–605. <https://doi.org/10.1109/ICCEAI55464.2022.00128>
- Juste Ruiz, J. (2019). El Derecho internacional de las migraciones: entre la crisis y la renovación. *Anuario Español de Derecho Internacional*, 35, 535–551. <https://doi.org/10.15581/010.35.535-551>
- León Castillo, L. A. (2015). *Análisis Económico de la Población. Demografía*. Universidad de Alicante.
- Linares, M. D. (2018). La Dirección Nacional de Migraciones entre 2003 y 2015: acciones para el cambio de cultura organizacional. Na mirada “desde adentro.” *REMHU: Revista Interdisciplinaria Da Mobilidade Humana*, 26(53), 187–203. <https://doi.org/10.1590/1980-85852503880005312>
- Lotero-Echeverri, G., & Pérez-Rodríguez, M. A. (2019). Migraciones en la sociedad contemporánea: Correlación entre migración y desarrollo. *Retos*, 9(17), 145–159. <https://doi.org/10.17163/ret.n17.2019.09>
- Mansour, R. F., Escorcia-Gutierrez, J., Gamarra, M., Gupta, D., Castillo, O., & Kumar, S. (2021). Unsupervised Deep Learning based Variational Autoencoder Model for COVID-19 Diagnosis and Classification. *Pattern Recognition Letters*, 151, 267–274. <https://doi.org/10.1016/j.patrec.2021.08.018>
- Martin, J. C., Bustamante-Sánchez, N. S., & Indelicato, A. (2022). Analyzing the Main Determinants for Being an Immigrant in Cuenca (Ecuador) Based on a Fuzzy Clustering Approach. *Axioms*, 11(2), 74. <https://doi.org/10.3390/axioms11020074>
- Molaei, S. M., & Keyvanpour, M. R. (2015). An analytical review for event prediction system on time series. *2015 2nd International Conference on Pattern Recognition and Image Analysis (IPRIA)*, 1–6. <https://doi.org/10.1109/PRIA.2015.7161635>
- Molnar, Petra. (2019). New technologies in migration: human rights impacts. *Forced Migration Review*, 61, 7–9.
- Mota, S. (2015). ¿Qué es un algoritmo? Una respuesta desde la obra de Wittgenstein. *ENDOXA*, 36, 317–328. <https://doi.org/10.5944/endoxa.36.2015.14967>

- Niculescu, S. P. (2003). Artificial neural networks and genetic algorithms in QSAR. *Journal of Molecular Structure: THEOCHEM*, 622(1–2), 71–83.
[https://doi.org/10.1016/S0166-1280\(02\)00619-X](https://doi.org/10.1016/S0166-1280(02)00619-X)
- Niño Arguelles, Y. L., Álvarez Santana, C. L., & Giovanni Locatelli, F. (2020). Migración Venezolana, Aporofobia en Ecuador y Resiliencia de los Inmigrantes Venezolanos en Manta, Periodo 2020. *Revista San Gregorio*, 43, 82–108.
- Paixão, C., Rocha, V., Brooks, D., & Marques, A. (2022). Unsupervised physical activity interventions for people with COPD: A systematic review and meta-analysis. *Pulmonology*. <https://doi.org/10.1016/j.pulmoe.2022.01.007>
- Pencea, G. C., & Curteanu, A. B. (2020). Internal and External Migration. *Postmodern Openings*, 11(1), 57–70. <https://doi.org/10.18662/po/108>
- Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 35–39. <https://doi.org/10.1109/COMITCon.2019.8862451>
- Raymer, J., & Wiśniowski, A. (2018). Applying and testing a forecasting model for age and sex patterns of immigration and emigration. *Population Studies*, 72(3), 339–355. <https://doi.org/10.1080/00324728.2018.1469784>
- Ren, C., & Bloemraad, I. (2022). New Methods and the Study of Vulnerable Groups: Using Machine Learning to Identify Immigrant-Oriented Nonprofit Organizations. *Socius: Sociological Research for a Dynamic World*, 8, 237802312210769. <https://doi.org/10.1177/23780231221076992>
- Ruiz Albornoz, M. R. (2020). *Gobierno digital en el control migratorio del Puesto de Control Migratorio del Aeropuerto Internacional Jorge Chávez, Callao 2020 (Tesis de grado)*. Universidad César Vallejo.
- Russell, N., ter Hofstede, A. H. M., Edmond, D., & van der Aalst, W. M. P. (2005). Workflow Data Patterns: Identification, Representation and Tool Support. *International Conference on Conceptual Modeling*, 353–368. https://doi.org/10.1007/11568322_23
- Sidow Osman, A. (2019). Data Mining Techniques: Review. *International Journal Of Data Science Research*, 2(1).
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-Means Clustering Algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>

- Sohail, A., & Arif, F. (2020). Supervised and unsupervised algorithms for bioinformatics and data science. *Progress in Biophysics and Molecular Biology*, 151, 14–22. <https://doi.org/10.1016/j.pbiomolbio.2019.11.012>
- Sonde, P., Balamwar, S., & Ochawar, R. S. (2020). Urban sprawl detection and analysis using unsupervised classification of high resolution image data of Jawaharlal Nehru Port Trust area in India. *Remote Sensing Applications: Society and Environment*, 17, 100282. <https://doi.org/10.1016/j.rsase.2019.100282>
- Tarasyev, A. A., Agarkov, G. A., & Hosseini, S. I. (2018). *Application of machine learning methods to migrant behavior modeling*. 050004. <https://doi.org/10.1063/1.5079102>
- Vazquez, P., Garcia, J. C., Luna, M. J., & Vaca, C. (2020). Temporal topics in online news articles: Migration crisis in Venezuela. *2020 Seventh International Conference on EDemocracy & EGovernment (ICEDEG)*, 106–113. <https://doi.org/10.1109/ICEDEG48599.2020.9096804>
- Viloria, A., Lizardo Zelaya, N. A., & Varela, N. (2020). Unsupervised learning algorithms applied to grouping problems. *Procedia Computer Science*, 175, 677–682. <https://doi.org/10.1016/j.procs.2020.07.099>
- Williams, N., & Efendic, A. (2019). Internal displacement and external migration in a post-conflict economy: Perceptions of institutions among migrant entrepreneurs. *Journal of International Entrepreneurship*, 17(4), 558–585. <https://doi.org/10.1007/s10843-019-00244-5>
- Zhang, C., & Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects. *Journal of Industrial Information Integration*, 23, 100224. <https://doi.org/10.1016/j.jii.2021.100224>
- Zhou, A., Zhou, S., Cao, J., Fan, Y., & Hu, Y. (2000). Approaches for scaling DBSCAN algorithm to large spatial databases. *Journal of Computer Science and Technology*, 15(6), 509–526. <https://doi.org/10.1007/BF02948834>

ANEXOS

Anexo 1. Matriz de consistencia

<p>Problema general ¿En qué medida la aplicación de un algoritmo no supervisado permite clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones?</p>	<p>Objetivo general Evaluar la aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones.</p> <p>Objetivos específicos 1. Construir el datasets de clasificación de extranjeros según riesgo de inmigración irregular reportados por la Superintendencia Nacional de Migraciones. 2. Aplicar DBSCAN (Agrupamiento Espacial Basada en la Densidad de Aplicaciones con Ruido) como algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones. 3. Determinar el coeficiente de Silhouette de la aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones.</p>	<p>Hipótesis general 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.</p>		<p>Técnica DBSCAN (Agrupamiento espacial basada en la densidad de aplicaciones con ruido)</p> <p>Observación</p> <p>Instrumentos Ficha de registro de datos de métricas de validación del clustering.</p>											
<p>Diseño de investigación</p>	<p>Población y muestra</p>	<p>Variables y dimensiones</p>													
<p>Estudio de tipo básico, de enfoque cuantitativo y nivel descriptivo.</p> <p>El diseño de investigación fue no experimental.</p>	<p>La población se conformó por 100 000 datos de extranjeros registrados en la base de datos de la Superintendencia Nacional de Migraciones del Perú.</p> <p>La muestra se conformó por 31 625, mediante muestreo no probabilístico por conveniencia.</p>	<table border="1"> <thead> <tr> <th>Variables</th> <th>Dimensiones</th> <th></th> </tr> </thead> <tbody> <tr> <td>VI: Aplicación de un algoritmo no supervisado</td> <td>DBSCAN</td> <td></td> </tr> <tr> <td rowspan="2">VD: Clasificación de extranjeros según riesgo de inmigración irregular</td> <td>Índices internos</td> <td></td> </tr> <tr> <td>Índices externos</td> <td></td> </tr> </tbody> </table>	Variables	Dimensiones		VI: Aplicación de un algoritmo no supervisado	DBSCAN		VD: Clasificación de extranjeros según riesgo de inmigración irregular	Índices internos		Índices externos			
Variables	Dimensiones														
VI: Aplicación de un algoritmo no supervisado	DBSCAN														
VD: Clasificación de extranjeros según riesgo de inmigración irregular	Índices internos														
	Índices externos														

Anexo 3. Solicitud de acceso a información pública

SOLICITUD DE ACCESO A LA INFORMACIÓN PÚBLICA (Texto Único Ordenado de la Ley N° 27806, Ley de Transparencia y Acceso a la Información Pública, aprobado por Decreto Supremo N° 043-2003-PCM)		
FORMULARIO		N° DE REGISTRO
I. FUNCIONARIO RESPONSABLE DE ENTREGAR LA INFORMACIÓN		
II. DATOS DEL SOLICITANTE		
APELLIDOS Y NOMBRES/RAZON SOCIAL		DOCUMENTO DE IDENTIDAD
LAZO BARTRA ULISES		43025310
DOMICILIO		
Calle./Jr./Av./Psje.		N°/Dpto./Int.
Calle Comandante espinar 260.		Torre3/701
Urbanización		Distrito
Miramar		San Miguel
Provincia		Departamento
Lima		Lima
Teléfono		Correo electrónico
942916766		ulaba@hotmail.com
Asimismo, solicito que todo acto administrativo derivado del presente procedimiento, se me notifique en el correo electrónico (E-mail) consignado en el presente formulario. (Ley N° 27444, numeral 20.1.2 del artículo 20° y numeral 123.1 del artículo 123°)		
SI <input type="checkbox"/> NO <input type="checkbox"/>		
III. INFORMACIÓN SOLICITADA		
Actualmente soy estudiante universitario de la UNSM - Tarapoto, llevando una maestría en tecnologías de la información, en mi curso de tesis deseo construir un algoritmo para clasificar a los extranjeros según su riesgo de inmigración irregular. Por tal motivo solicito, información de extranjeros que hayan actualizado sus datos para solicitar regularización migratoria, según el siguiente detalle: Calidad migratoria, país y ciudad de nacimiento, edad, sexo, puesto fronterizo, fecha de ingreso, ingreso regular, exceso de días, tuvo control migratorio, nivel educativo, grado académico, labora actualmente en Perú, tiene educación completa, profesión, estado civil, religión, estudia en Perú, actividad económica, salario, tiene familia en Perú, tiene hijos, cantidad de hijos, solicito refugio, embarazada, tiene discapacidad, tipo de discapacidad, tiene enfermedad, tipo de vivienda, pago alquiler, tiene servicios básicos, tiene internet, tiene cuenta bancaria. Por favor espero su pronta respuesta, con el fin de generar un aporte a ustedes y lograr alcanzar mi desarrollo profesional.		
IV. DEPENDENCIA DE LA CUAL SEREQUIERE LA INFORMACIÓN		
REGISTRO MIGRATORIO		
V. FORMA DE ENTREGA DE LA INFORMACIÓN (Marcar con una X)		
Copia simple <input type="checkbox"/> CD <input type="checkbox"/> Correo electrónico <input checked="" type="checkbox"/>		
Fecha y hora de recepción		Firma
11/03/2022		
Observaciones		



"Decenio de la igualdad de oportunidades para mujeres y hombres"
"Año del Fortalecimiento de la Soberanía Nacional"

Breña, 16 de Marzo del 2022

CARTA N° 000116-2022-TP/MIGRACIONES

Sr(a).

LAZO BARTRA ULISES

Calle Comandante Espinar N° 260 Torre 3/701 urb. Miramar – San Miguel

Correo electrónico: ULABA@HOTMAIL.COM

Presente.-

Asunto: Respuesta a solicitud por de Acceso a la Información Pública

Referencia: Solicitud de Acceso a la Información Pública registrada con N° 2022031161288

Mediante la presente, tengo el agrado de dirigirme a usted, en atención al documento de la referencia, por medio del cual solicitó textualmente lo siguiente:

"(...) solicito, información de extranjeros que hayan actualizado sus datos para solicitar regularización migratoria, según el siguiente detalle:
Calidad migratoria, país y ciudad de nacimiento, edad, sexo, puesto fronterizo, fecha de ingreso, ingreso regular, exceso de días, tuvo control migratorio, nivel educativo, grado académico, labora actualmente en Perú, tiene educación completa, profesión, estado civil, religión, estudia en Perú, actividad económica, salario, tiene familia en Perú, tiene hijos, cantidad de hijos, solicito refugio, embarazada, tiene discapacidad, tipo de discapacidad, tiene enfermedad, tipo de vivienda, pago alquiler, tiene servicios básicos, tiene internet, tiene cuenta bancaria." sic.

Al respecto, la Oficina de Planeamiento y Presupuesto mediante Memorando N° 000854-2022-OPP/MIGRACIONES de fecha 15 de marzo de 2022, remitió la información solicitada cumpliendo con informar lo siguiente:

1. Las variables ciudad de nacimiento, salario, tiene familia en Perú, tiene hijos, cantidad de hijos, pago alquiler y tiene cuenta bancaria no han sido preguntadas a los ciudadanos extranjeros que se preinscribieron para el proceso de regularización migratoria, motivo por el cual no es posible remitir lo solicitado, de conformidad con lo establecido en el tercer párrafo del artículo 13¹ del T.U.O. de la Ley N° 27806, Ley de Transparencia y Acceso a la Información Pública.
2. Las variables puesto fronterizo, fecha de ingreso e ingreso regular son expresadas como País fronterizo (por el que ingresó) y Fecha de ingreso.
3. La variable religión no se puede proporcionar porque se encuentra dentro del ámbito de datos sensibles, según el marco de la Ley N° 29733, Ley de Protección de Datos Personales, en el artículo 2 y concordancia con el inciso 13.6 del artículo 13.
4. Las demás variables se encuentran agrupadas en el anexo "Preinscripción datos". (Se adjunta copia del Memorando antes citado)

En ese sentido, de conformidad con lo que establece la Ley N° 27806 – Ley de Transparencia y Acceso a la Información Pública, se da por atendida su solicitud.

¹ Artículo 13* - Denegatoria de acceso.

(...)

"La solicitud de información no implica la obligación de las entidades de la Administración Pública de crear o producir información con la que no cuente o no tenga obligación de contar al momento de efectuarse el pedido. En este caso, la entidad de la Administración Pública deberá comunicar por escrito que la denegatoria de la solicitud se debe a la inexistencia de datos en su poder respecto de la información solicitada" (...)

Superintendencia Nacional de Migraciones
Av. España N° 734 Breña, Lima – Perú – T (511) 200 – 1000
informes@migraciones.gob.pe / www.migraciones.gob.pe



Sin otro en particular, hago propicia la oportunidad para expresarle los sentimientos de mi especial consideración y deferente estima.

Atentamente,

ROSE MARY RAMIREZ ESCARATE
Responsable de Entregar la Información de Acceso Público
DOCUMENTO FIRMADO DIGITALMENTE

(RRE/jgs)

Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

por Ulises Lazo Bartra

Fecha de entrega: 22-sep-2023 01:59p.m. (UTC-0500)

Identificador de la entrega: 2173918035

Nombre del archivo: tesis_maestria_Ulises_Lazo_Bartara_1.docx (2.38M)

Total de palabras: 12602

Total de caracteres: 76401

Aplicación de un algoritmo no supervisado para clasificar extranjeros según riesgo de inmigración irregular en la Superintendencia Nacional de Migraciones

INFORME DE ORIGINALIDAD



FUENTES PRIMARIAS

1	repositorio.unsm.edu.pe Fuente de Internet	7%
2	www.ridaa.unicen.edu.ar Fuente de Internet	1%
3	repositorio.unan.edu.ni Fuente de Internet	1%
4	tesis.unsm.edu.pe Fuente de Internet	1%
5	hdl.handle.net Fuente de Internet	1%
6	Submitted to Pontificia Universidad Católica del Ecuador - PUCE Trabajo del estudiante	1%
7	Submitted to Universidad Tecnológica Centroamericana UNITEC Trabajo del estudiante	1%
8	Submitted to University of Southern California	