

Integración económica de la población venezolana en el Perú 2022

Cynthia Zhou Chun Miao Zhou Lin

Facultad de Ingeniería
Universidad del Pacífico
Lima, Perú
cz.zhoul@alum.up.edu.pe

Fabrizio Hernan Montalvo Pascal

Facultad de Ingeniería
Universidad del Pacífico
Lima, Perú
fh.montalvop@alum.up.edu.pe

Paolo Cesar Salazar Patricio

Facultad de Ingeniería
Universidad del Pacífico
Lima, Perú
pc.salazarp@alum.up.edu.pe

Abstract—La migración venezolana hacia Perú ha sido uno de los movimientos más grandes en América del Sur en la última década. Este trabajo busca identificar y clasificar los perfiles socioeconómicos de los migrantes venezolanos utilizando *data mining* para comprender mejor su realidad y su integración en la economía peruana. Los resultados ayudarán a optimizar políticas públicas dirigidas a esta población vulnerable.

Index Terms—migración venezolana, Perú, integración económica, *data mining*, mercado laboral.

I. INTRODUCCIÓN

La crisis económica y política en Venezuela ha provocado una de las mayores migraciones en la historia reciente de América Latina. Desde 2015, millones de venezolanos han abandonado su país debido a la hiperinflación, el desabastecimiento de alimentos y productos básicos, y la falta de oportunidades laborales, buscando refugio en países vecinos [1]. Según estimaciones del ACNUR (Alto Comisionado de las Naciones Unidas para los Refugiados, 2024), más de 7.7 millones de venezolanos han migrado [2], de los cuales más de 1.5 millones se encuentran en Perú, convirtiéndolo en uno de los principales países receptores [3]. Esta migración masiva ha generado retos tanto para los migrantes como para los países anfitriones. En Perú, la mayoría de los venezolanos se ha integrado al sector informal de la economía, donde enfrentan condiciones laborales precarias, xenofobia y barreras para acceder a trabajos formales debido a la dificultad de validar sus títulos académicos y profesionales [4]. A pesar de que existen políticas migratorias peruanas que buscan proteger los derechos humanos de los migrantes, en la práctica, estas suelen ser insuficientes, lo que perpetúa la vulnerabilidad de esta población y limita su contribución al desarrollo económico [5]. En este contexto, el presente trabajo tiene como objetivo identificar y clasificar los perfiles socioeconómicos de los migrantes venezolanos residentes en Perú mediante técnicas de minería de datos (*data mining*) aplicadas a los datos de la Encuesta a la Población Venezolana Residente en el Perú (ENPOVE 2022). Este enfoque permitirá generar una comprensión más profunda de la realidad de los migrantes y sus necesidades, contribuyendo al diseño de políticas públicas más efectivas que promuevan su integración en la sociedad y el mercado laboral peruanos.

II. OBJETIVOS

A. Objetivo general

- Identificar y clasificar los diferentes perfiles socioeconómicos de los migrantes venezolanos residentes en Perú utilizando técnicas de *data mining* para comprender su realidad e impulsar su integración en la economía y el mercado laboral del Perú.

B. Objetivos específicos

- Encontrar una base de datos con las características socioeconómicas de la población venezolana.
- Realizar un preprocesamiento efectivo para obtener una base de datos limpia, ordenada y preparada para aplicar cualquier técnica de minería de datos.
- Aplicar técnicas de minería de datos, como análisis de *clústeres*, para segmentar a la población migrante en grupos homogéneos basados en sus características socioeconómicas.
- Identificar patrones y tendencias significativas dentro de los rasgos socioeconómicos que revelen las necesidades, desafíos y oportunidades específicas de los perfiles de migrantes identificados.

III. ESTADO DEL ARTE

La aplicación de técnicas de *data mining* para identificar y clasificar perfiles socioeconómicos en poblaciones vulnerables ha sido ampliamente estudiada en diversos contextos migratorios y de pobreza. Estas investigaciones proporcionan un marco conceptual y metodológico valioso para analizar la situación de los migrantes venezolanos en el Perú. En Colombia, un estudio realizado por la Organización Internacional del Trabajo (OIT), empleó el análisis de *clústeres* para segmentar a los migrantes venezolanos según variables como el nivel educativo, la situación laboral y el tipo de vivienda. Este estudio identificó cuatro grupos principales: (a) profesionales con dificultades de homologación, (b) trabajadores informales, (c) personas en extrema vulnerabilidad y (d) emprendedores [6]. En el informe resaltaron que los perfiles más vulnerables enfrentan barreras significativas en el acceso a servicios básicos, y la segmentación permitió sugerir políticas específicas, como la simplificación de trámites

para validar títulos académicos [6]. Por otro lado, en Turquía, un estudio respaldado por la Organización Internacional para las Migraciones (OIM) aplicó el análisis discriminante y k-means para segmentar a los refugiados sirios según su nivel de acceso a empleo, vivienda y educación [7]. Se identificaron tres grupos principales: (a) refugiados con acceso limitado o nulo a empleo formal y educación, concentrados mayormente en áreas rurales, (b) aquellos en regiones urbanas con empleo informal y acceso parcial a servicios educativos, y (c) un grupo minoritario en ciudades principales con mejores condiciones laborales y educativas [7]. Las variables clave incluyeron nivel educativo, tipo de empleo, acceso a vivienda adecuada y la región de asentamiento. Los resultados destacaron que factores como la falta de redes sociales y la barrera del idioma eran determinantes críticos en los grupos más vulnerables, lo que permitió una mejor asignación de recursos humanitarios y educativos hacia las zonas rurales y los programas de integración laboral y académica [7]. Finalmente, en Brasil, investigaciones del Banco Mundial y el ACNUR aplicaron análisis de datos para evaluar la inclusión económica y social de los migrantes venezolanos [8]. Los resultados indicaron que los adultos venezolanos tienen un 64% menos de probabilidades de estar empleados en el sector formal, y los niños un 53% menos de probabilidades de asistir a la escuela, en comparación con la población local [8]. Estos hallazgos resaltan la necesidad de políticas públicas que aborden las barreras de integración y promuevan la inclusión socioeconómica de los migrantes.

IV. DISEÑO DEL EXPERIMENTO

A. Descripción de la base de datos

Los datos empleados provienen del Instituto Nacional de Estadística e Informática (INEI), obtenidos a través de la "Encuesta Dirigida a la Población Venezolana que Reside en el País (ENPOVE)".

Fuente oficial: <https://proyectos.inei.gob.pe/microdatos/>.

Ciudades incluidas: Tumbes, Piura, Chiclayo, Trujillo, Chimbote, Ica, Arequipa, Lima Metropolitana y Callao.

Archivos:

- **Conjunto de datos 1: Salud.** Este dataset recoge información sobre el estado de salud de los encuestados, acceso a servicios médicos y otros aspectos relacionados con la atención médica, como afiliación a sistemas de salud, enfermedades crónicas, discapacidades y datos sobre COVID-19.

Dimensión del dataset: 11,621 filas x 116 columnas.

- **Conjunto de datos 2: Características de la vivienda y del hogar.** Incluye detalles sobre el tipo de vivienda, materiales de construcción, acceso a servicios básicos, y bienes presentes en el hogar.

Dimensión del dataset: 3,769 filas x 49 columnas.

- **Conjunto de datos 3: Características de los residentes del hogar.** Proporciona información demográfica y socioeconómica de los residentes, incluyendo parentesco, edad, estado civil, número de celular y condición migratoria.

Dimensión del dataset: 12,487 filas x 42 columnas.

- **Conjunto de datos 4: Características de los empleados.** Contiene información sobre la situación laboral de los encuestados, como ocupación, tipo de empleo, sector económico, horario, condiciones laborales y remuneración.

Dimensión del dataset: 10,242 filas x 128 columnas.

- **Conjunto de datos 5: Educación.** Contiene información sobre el nivel de educación de los inmigrantes venezolanos.

Dimensión del dataset: 10,739 filas x 55 columnas.

B. Metodología

En esta sección, se explicarán los pasos que se siguieron en la metodología. Esta es una versión modificada de la metodología KDD (Knowledge Discovery Dictionary) adaptada al presente proyecto.



Fig. 1. Metodología del estudio.

- **Entendimiento del problema:** Definir claramente los objetivos y preguntas a responder con la investigación. Recopilar información de la literatura reciente sobre la problemática.
- **Entendimiento de los datos:** Exploración de bases de datos provenientes de la Encuesta de Población Venezolana e Instituto Nacional de Estadística e Informática. Selección de variables relevantes para el tema a tratar como salud, empleos e infraestructura de los hogares.
- **Integración de bases de datos:** Integración de los datos en una única base utilizando los atributos previamente analizados como el número de casa, conglomerado, grupo, etc.
- **Preparación de los datos:** Limpieza y preprocesamiento de los datos con técnicas de programación para garantizar la calidad para el análisis.
 - Se verificó la existencia de valores nulos en las diversas columnas. Todos los registros vacíos y espacios fueron reemplazados por valores nulos.
 - Se eliminaron los registros que tenían valores nulos pero representan menos del 5% de los registros totales de las columnas.
 - Se eliminaron las columnas con más del 70% de registros nulos.
 - Se imputaron con la moda los registros nulos que representan más del 5% de los registros totales de las columnas.
 - Se eliminaron columnas de identificación y/o columnas con información irrelevante. Por ejemplo, la pregunta "En este hogar vive algún venezolano?" tenía siempre la misma respuesta: "Sí".
 - Se realizó un proceso de conversión de variables categóricas a numéricas. Para las variables con un

orden se aplicó el LabelEncoder, mientras que para las variables sin un orden específico se aplicó One-HotEncoding.

- Se eliminaron registros y columnas que representaban un valor NaN luego de la conversión.

Finalmente, la base de datos preprocesada tuvo 10,212 registros y 49 columnas.

- **Análisis exploratorio de los datos:** Elaborar gráficos sobre la base de datos final para visualizar tendencias y/o relaciones entre las variables de estudio y obtener insights de ello.
- **Selección de técnicas de Data Mining:** En base a los análisis realizados, seleccionar las técnicas de minería de datos más adecuadas para poder cumplir los objetivos propuestos. En el presente estudio, se escogieron las técnicas de Análisis de Componentes Principales (PCA), Non-Negative Matrix Factorization (NMF), y K-means.
- **Experimentación:** Desarrollar, implementar y evaluar el funcionamiento de las técnicas seleccionadas con métricas de desempeño.
- **Análisis y conclusiones:** Analizar los resultados obtenidos, elaborar conclusiones y plantear recomendaciones dirigidas a la resolución del problema.

V. ANÁLISIS EXPLORATORIO DE DATOS

En primer lugar, se analizó la distribución de estados civiles de la población venezolana (ver Figura 2).

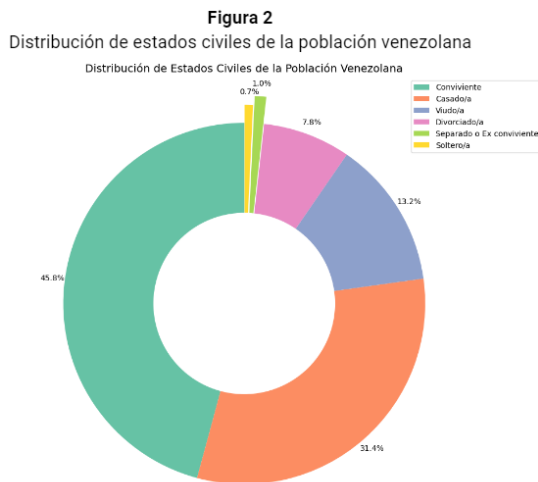


Fig. 2. Distribución de estados civiles de la población venezolana

Se observaron las siguientes características:

- **Conviviente:** La porción más grande, con un 45.8%, representa a las personas que viven en una unión consensual pero no están legalmente casadas.
- **Casado/a:** El 31.4% de la población está legalmente casada.
- **Soltero/a:** El 7.8% de la población se clasifica como soltera.
- **Viudo/a:** El 13.2% de las personas son viudas.

- **Divorciado/a:** El 7.8% de la población está divorciada.
- **Separado o Ex conviviente:** Este grupo constituye el 1.0%.
- **Otra categoría pequeña:** El 0.7% corresponde a un grupo no especificado.

El gráfico muestra que una parte significativa de la población está en uniones informales o casadas, seguidas por personas viudas y solteras. Esto ofrece una visión general de las tendencias de estado civil y relaciones de los venezolanos en el Perú.

Posteriormente, mediante una matriz de correlación entre las variables, se identificó que las horas de trabajo pueden estar relacionadas con los días de la semana (ver Figura 3).

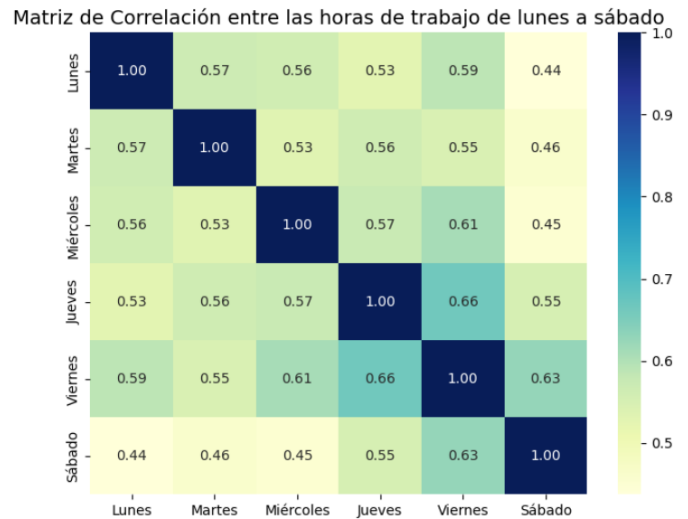


Fig. 3. Matriz de correlación entre horas de trabajo y días de la semana

En la Figura 3 se observa que los lunes tienen una correlación de 0.57 con los martes, lo que indica que si una persona trabaja más horas el lunes, es probable que también trabaje más horas el martes. Sin embargo, los sábados presentan las correlaciones más bajas con el resto de días de la semana, lo que podría indicar que las personas trabajan menos o de forma más variable los sábados.



Fig. 4. Ciudad de destino de los venezolanos

Por último, se elaboró un gráfico de barras para identificar los destinos de los migrantes venezolanos dentro del Perú (ver Figura 4). Lima es el principal destino, con más de 2000 personas, seguido por Trujillo, con aproximadamente 500 personas. Esto sugiere que Lima es el principal atractivo para los migrantes venezolanos.

VI. EXPERIMENTACIÓN Y RESULTADOS

A. Análisis de Componentes Principales (PCA)

Para realizar una tarea de agrupamiento efectiva, se realizó un Análisis de Componentes Principales (PCA) para reducir la dimensionalidad de los datos. Previamente, se realizó la estandarización de las variables numéricas.



Fig. 5. Gráfico de la varianza explicada por el número de componentes principales.

Al realizar y analizar la gráfica de varianza explicada según número de componentes, se eligieron 5 componentes principales debido a que estos explicaban más del 80% de la varianza.

Los 5 principales componentes que estaban correlacionados con algunas de las variables de la base de datos son:

- **Componente 1:** Nivel de estudios en Venezuela.
- **Componente 2:** Número de habitantes en el hogar.
- **Componente 3:** Estrato social.
- **Componente 4:** Persona que se considera de etnicidad morena.
- **Componente 5:** Persona que se considera de etnicidad blanca.

B. Non-Negative Matrix Factorization (NMF)

Posteriormente, se empleó otra técnica de reducción de dimensionalidad para comparar los resultados obtenidos en la primera técnica. Previamente se realizó la transformación MinMax a las variables numéricas.

Al realizar y analizar la gráfica de varianza explicada según número de componentes, se eligieron 5 componentes principales debido a que estos explicaban más del 80% de la varianza. Los 5 principales componentes que estaban correlacionados con algunas de las variables de la base de datos son:

- **Componente 1:** Nivel de estudios en Venezuela.
- **Componente 2:** Persona que se considera de etnicidad morena.

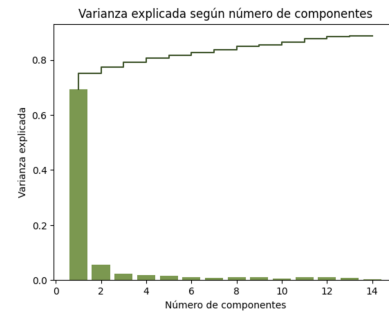


Fig. 6. Gráfico de la varianza explicada por el número de componentes principales.

- **Componente 3:** Persona con último grado de estudios solo en Venezuela.
- **Componente 4:** Estrato social.
- **Componente 5:** ¿Tenía trabajo antes de partir de Venezuela?

C. K-means

Luego se realizó un análisis de KMEANS tanto con los componentes obtenidos con PCA como los obtenidos con NMF.

1) *K-means con PCA:* Se optimizó el número de clústeres mediante el método del codo.

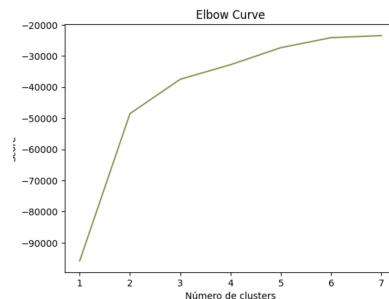


Fig. 7. Método del codo para determinar el número óptimo de clústeres con PCA.

Luego del análisis, se eligió generar 3 clústeres

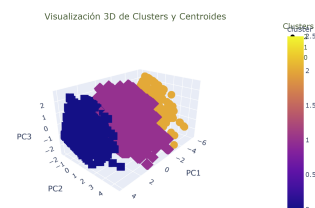


Fig. 8. Gráfico de clústeres encontrados con PCA.

Los clústeres azules y púrpuras tienen una distribución más compacta, indicando menor dispersión de los datos. Por otro lado, el clúster amarillo muestra mayor variabilidad, lo que podría reflejar mayor heterogeneidad dentro de este grupo.

2) *K-means con NMF*: Se optimizó el número de clústeres mediante el método del codo.

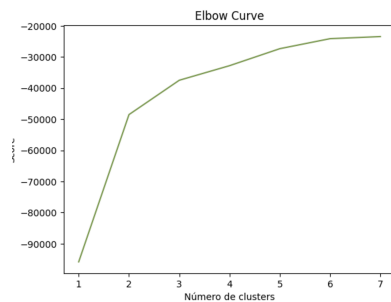


Fig. 9. Método del codo con NMF.

Luego del análisis, se eligió generar 3 clústeres

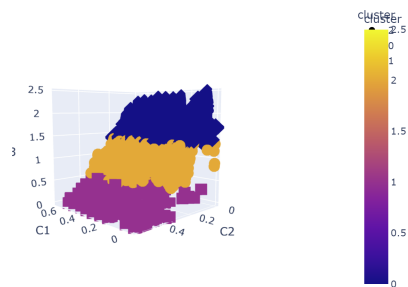


Fig. 10. Gráfico de clústeres encontrados con NMF.

En este caso, también se identificaron tres clústeres principales (azul, amarillo y púrpura). Las separaciones entre clústeres son más claras que en el gráfico obtenido con PCA, aunque hay zonas de solapamiento entre el clúster amarillo y el púrpura.

D. DBSCAN

Se aplicó el algoritmo DBSCAN tanto con PCA como con NMF.

1) *DBSCAN con PCA*: Se optimizaron los parámetros *eps* y *min_samples* mediante *gridsearch*.

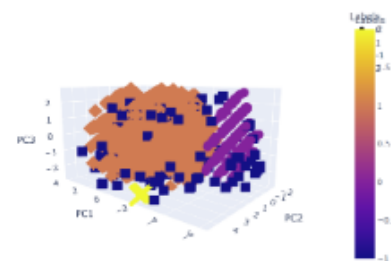


Fig. 11. Gráfico de clústeres encontrados con DBSCAN (PCA).

2) *DBSCAN con NMF*: Se optimizaron los parámetros *eps* y *min_samples* mediante *gridsearch*.

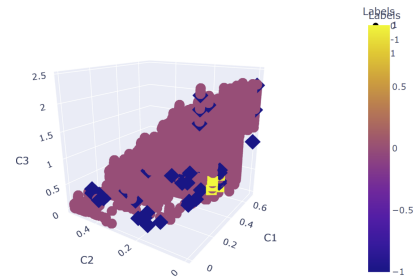


Fig. 12. Gráfico de clústeres encontrados con DBSCAN (NMF).

REFERENCES

- [1] ACNUR: "Situación de Venezuela", 2024. [Online]. Available: <https://www.acnur.org/emergencias/situacion-de-venezuela>
- [2] El Peruano: "En Perú reside un millón y medio de venezolanos y solo 589 podrán votar", 2024. [Online]. Available: <https://www.elperuano.pe/noticia/248440-en-peru-reside-un-millon-y-medio-de-venezolanos-y-solo-589-podran-votar#:~:text=21%2F07%2F2024%20En%20Per%C3%BA,900.000%20tienen%20edad%20para%20votar>
- [3] Eguren, F. & Koechlin, J., "El éxodo venezolano: entre el exilio y la emigración", 2018.
- [4] Loayza, C., "Investigación en Ciencias Sociales: Reflexiones sobre la migración", 2020.
- [5] Blouin, C., "Complejidades y contradicciones en la política migratoria", 2021.
- [6] Crespo, R., "Discursos sobre la migración venezolana en Perú", 2022.
- [7] Galarza, F. & Torres, J., "Integración humana de migrantes venezolanos en el Perú", 2019.
- [8] Lovón, R., "Explotación laboral y vulnerabilidad de migrantes venezolanos", 2021.
- [9] INEI. "Encuesta Dirigida a la Población Venezolana que Reside en el País - ENPOVE". INEI. Accedido el 21 de septiembre de 2024. [Online]. Available: <https://proyectos.inei.gob.pe/microdatos/>
- [10] Riquelme, J. C., Ruiz, R., & Gilbert, K., "Minería de datos: Conceptos y tendencias", Inteligencia Artificial, 2006. [Online]. Available: <https://www.redalyc.org/pdf/925/92502902.pdf>
- [11] Martínez, B. B., "Minería de datos: Cómo hallar una aguja en un pajar", Ingenierías, 2001. [Online]. Available: <https://www.cs.buap.mx/~bbeltran/NotasMD.pdf>
- [12] Bravo, C. A. & Soto, D. L., "Midiendo la integración de los migrantes usando experimentos: Un análisis de la integración venezolana en el Perú", Trabajo de Suficiencia Profesional, Facultad de Economía y Finanzas, Lima, Perú, marzo 2024.
- [13] Condori, M., Reyna, G. A., Villavicencio, A. C., Párraga, C. L., & Vilcapoma, D., "Éxodo venezolano, inserción laboral y discriminación social en la ciudad de Huancayo, Perú", Revista Espacios, vol. 41, no. 27, art. 8, pp. 72-83, 2020.

VII. ANEXOS

Github, <https://github.com/paolosalazarp/InmigracionVenezolana>