

Determinantes económicos, sociales y demográficos de  
la migración en Colombia: Un enfoque probabilístico  
y machine learning (2023-2025)

Alan Steven Ríos Munar  
Juan Esteban Samaniego Poveda  
Sebastian Casas Poloche

Universidad Externado de Colombia  
IA con aplicaciones en economía  
Lina María Castro

## Objetivo

Se quiere determinar qué factores económicos, demográficos y sociales son causantes de la migración interna del país y cuál es la probabilidad de migración de los ciudadanos colombianos, tanto entre departamentos como municipios. Para poder identificar cuales son las regiones que tienen mayor concentración en cuanto a los flujos de entrada y de salida, cuál es la causa del desplazamiento y la probabilidad de una persona migrar a otra región.

## Descripción

Desarrollar un modelo que permita caracterizar las principales variables que producen migración en el territorio nacional y la probabilidad de que una persona migre a otra región del país, mediante el análisis y comparación de diferentes modelos probabilísticos. Además se quiere usar un KNN para agrupar de manera fácil y óptima las personas que compartan características para una identificación de forma grupal. Por otra parte, se hará uso del desarrollo de mapas de calor para visualizar el flujo de migrantes en el territorio a nivel departamental y a nivel municipal. Clasificando los departamentos y municipios con mayor movilidad de personas y con mayor probabilidad de migración.

## Desafíos

Unir en una sola base de datos la GEIH (gran encuesta integrada de hogares) realizada por el DANE de los años 2023, 2024, 2025, para poder tener la mayor cantidad de datos y lograr una mayor tasa de efectividad en la predicción de la probabilidad de migración y sus causas. Teniendo en cuenta que la información dentro de cada encuesta está dada por meses y que cada una de ellas representa un año específico.

## Entrega de valor

Este proyecto aporta una innovación metodológica y aplicada al estudio de la migración interna en Colombia al integrar modelos probabilísticos, uso de machine learning y análisis espacial con base en la unificación de la GEIH del año 2023, 2024, 2025. A diferencia de los estudios de causalidad, la investigación estima la probabilidad individual de migración interna, identifica perfiles migratorios mediante algoritmos de clasificación como KNN, compara distintos modelos probabilísticos para determinar cuál ofrece mayor capacidad predictiva e incorpora análisis territorial a través de mapas de calor y flujos interregionales.

## Stakeholders

- Ministerio de Trabajo
- Planeación Nacional

- Gobernaciones
- Alcaldías
- Programas de desarrollo regional
- Planeación urbana

#### Técnicas que se utilizaran

- Análisis exploratorio de datos
- Clustering
- Machine learning
- Modelos probabilísticos
- Análisis espacial
- Integración y transformación de datos

#### Fuentes de datos

Gran Encuesta Integrada de hogares (GEIH) 2023-2025

Banco de la República

- Indicadores regionales de actividad económica.
- Desempleo regional.
- Informes económicos departamentales.

Departamento Administrativo Nacional de Estadística

- PIB departamental.
- Tasa de desempleo por ciudad.
- Índice de pobreza monetaria.
- Índice de pobreza multidimensional.

#### Variables

Demográficas:

- Edad
- Sexo
- Tamaño del hogar
- Estado civil

Socioeconómicas:

- Ingresos
- Condición laboral

- Tipo de ocupación
- Formalidad/informalidad
- Nivel educativo

Territoriales:

- Cambio de municipio de residencia.
- Municipio de residencia anterior

Migración:

- Migración reciente (últimos 12 meses)
- Motivo de migración

Violencia:

- Desplazamiento por violencia asociada al conflicto armado.
- Desplazamiento por violencia no asociada al conflicto armado.
- Motivo de migración por inseguridad o amenazas.