Logotipo, nombre de la empresa

Descripción generada automáticamente

MÁSTER EN DATA SCIENCE Y BUSINESS ANALYTICS PRESENCIAL

El fenómeno de la inmigración en España: causas, orígenes y tendencias

TFM elaborado por:

Cristian de Andrade Correia

Erick Ernesto Hernández Lara

Tutor/a de TFM:

Miguel Martín

Madrid, 30 de octubre de 2024

# Índice

[Índice 2](#_Toc1944136957)

[Resumen 3](#_Toc178280302)

[Introducción 4](#_Toc496217161)

[Antecedentes 7](#_Toc1399428220)

[Objetivo del proyecto 8](#_Toc354199448)

[Material y métodos 9](#_Toc325344134)

[Resultados 13](#_Toc2124761362)

[Adquisición de fuentes de datos 13](#_Toc1228042720)

[Descripción de fuentes de datos 13](#_Toc1868360534)

[Análisis Exploratorio de Datos o EDA 13](#_Toc744081032)

[Preprocesado de datos 19](#_Toc535968376)

[Selección de variables 19](#_Toc1593234350)

[Partición de los datos en conjuntos train y test 19](#_Toc567517263)

[Aplicación de modelos de aprendizaje automático 19](#_Toc590846205)

[Evaluación y validación de modelos de aprendizaje automático 20](#_Toc1170663162)

[Comparativa de resultados 20](#_Toc341495043)

[Conclusiones 20](#_Toc1472635573)

[Referencias bibliográficas 21](#_Toc591251290)

[Anexos 22](#_Toc284330677)

# Resumen

# Introducción

Según el Informe de Migraciones de 2022 de la Organización Internacional de Migraciones (OIM)1, de la Organización de las Naciones Unidas (ONU), la cantidad de migrantes internacionales ha incrementado de 153 millones en 1990 a 281 millones en 2020, acercándose a duplicar dicho valor, y que es, a su vez, más del triple de lo estimado en 1970.

De estos 281 millones de migrantes en el 2020, Europa y Asia acogieron alrededor de 87 y 86 millones respectivamente, que representa el 61% de la población mundial de migrantes, seguidos de Estados Unidos de Norte América con 21%.

Ya en el 2016 el Banco Europeo de Inversiones (BEI) mencionaba, en su análisis de desafíos y oportunidades de la migración en Europa2, que la Unión Europea (UE) fue el mayor aportador durante la escalada de inmigrantes y aplicantes de asilo durante los años 2015 y 2016, con una asignación de 10,1 billones de euros en ambos años, indicando que una deficiente integración de los migrantes podría traducirse en una reducción del Producto interno bruto (PBI) per cápita en los años subsiguientes. Sin embargo, mantuvo la visión que, a largo plazo, las migraciones internacionales tendrían un efecto positivo en Europa y resaltaba la necesidad de modificar y mejorar los marcos legislativos en torno a las políticas de integración y optimización del flujo de migrantes.

En consecuencia, el éxito recae en el sinergismo de los diversos procesos que regulan el ritmo migratorio y la integración de los inmigrantes, que a su vez están asociados a la capacidad económica, judicial, logística y de defensa de los países receptores para dar respuesta a las exigencias que plantea el continuo aumento de los movimientos migratorios internacionales.

Por otra parte, desde el punto de vista del individuo, también tenemos numerosos factores que impactan en su decisión por embarcarse en una travesía migratoria. La Organización Internacional para las Migraciones expone un diagrama (Figura 1) que hace un excelente sumario de la relación entre variables “macro”, “meso” y “micro” que afectan la decisión de las personas en migrar, las cuales involucran variables como la seguridad, libertad, empleo, precios, servicios básicos y salud (macro), actuando en conjunto con las condiciones, posibilidades y ventajas personales (micro y meso).

A diagram of a decision

Description automatically generated

**Figura 1**. Factores que propician la migración internacional.

(International Organization for Migration)

Así, podemos observar que el estudio del proceso migratorio es un reto complejo debido a la diversidad de factores que lo afectan y el dinamismo de nuestro actual entorno globalizado. Esto se evidencia también en la creciente implementación de políticas en relación con la movilidad humana promovida por la ONU a partir de 2010 que incluye áreas como el cambio climático, desastres y desarrollo sustentable3. Si, además, consideramos eventos de origen bélico, como el actual conflicto entre Ucrania y Rusia, o Israel y el grupo terrorista HAMAS, tendremos una visión general de la complejidad del fenómeno de las migraciones internacionales.

Indudablemente, el mayor esfuerzo y responsabilidad recae sobre los países con mayor recepción de dichos inmigrantes internacionales. De acuerdo con la OIM, tres países de la UE formaban parte del top 10 de países destino para migrantes internacionales a nivel global en el 2020: Alemania, Francia y España, siendo España el país que ocupaba el décimo lugar a nivel global y el tercero de la UE1, que, además, alcanzó en el 2022 un saldo migratorio de 727.005 personas, el máximo nivel en 10 años4.

Aunque el fenómeno de la migración también atiende a decisiones personas que pueden estudiarse desde un punto de vista psicosocial, en este proyecto mantendremos nuestro enfoque en variables macro (por encima del individuo).

A razón de lo expuesto, este proyecto busca aprovechar estas tecnologías para estudiar los datos oficiales de inmigrantes en España y desarrollar un modelo predictivo como herramienta de estrategia y planificación que permita estimar el número de inmigrantes al territorio español. Y, aunque el fenómeno de la migración también atiende a decisiones personas que pueden estudiarse desde un punto de vista psicosocial, en este proyecto mantendremos nuestro enfoque en variables macro (por encima del individuo).

# Antecedentes

Entre la diversidad de aplicaciones en los avances de la inteligencia artificial “clásica” o generativa, tenemos que los desarrollos en Aprendizaje Automático (AA), uno de los subcampos de la Inteligencia Artificial (IA), han permitido abordar problemáticas como la expuesta mediante algoritmos que permitan, por ejemplo, generar predicciones de inmigrantes y grupos específicos5, o incluso prever flujos migratorios relacionados con asilo mediante alertas usando datos de eventos y tendencias de Google6.

Exponer lo que hicieron, como y lo que obtuvieron

Relacionar los hallazgos de los distintos

Buscar un par mas de antecedentes (pueden ser también estrategias mas antiguas que no fuesen de ML, probablemente con metodos estadisticos)

# Objetivo del proyecto

* Predecir el número de inmigrantes en España mediante modelo de aprendizaje automático para prever flujos migratorios internacionales hacia el país.

# Material y métodos

* **Base de datos**: Los datos recopilados y usados en este proyecto son bases de datos estructuradas/relacionales. Siendo un estudio temporal de inmigración, las claves centrales que relacionan los distintos ficheros empleados se basan en: año y nombre del país/código y sexo; las relaciones varían según los datos que aporta cada fichero.
* **Repositorio**: Se creó un repositorio de GitHub para los ficheros, notebooks, *exports* resultantes y el escrito final, trabajando progresivamente en el proyecto mediante ramas a nivel local, haciendo *commits* de trabajoy revisiones conjuntas.
* **Infraestructura computaciona**l: Se trabajo en local mediante equipos personales y software de uso libre y/o estándares: VS Code, GitBash, MS Excel, MS Word y MS PowerPoint (diagramas). Adicionalmente, las librerías principales que se usaron en python durante el proyecto fueron: pandas, numpy, matplot-lib, altair, scikit-learn (*machine learning* y redes neuronales) y tensorflow (redes neuronales).

Debido a la variedad de conjuntos de datos usados y la necesidad de limpieza/preprocesamiento que cada uno presentaba, se decidió dividir el proyecto en 4 fases (Figura X) que proporcionaran una secuencia ordenada de limpieza, análisis y prueba de algoritmos de *machine learning*:

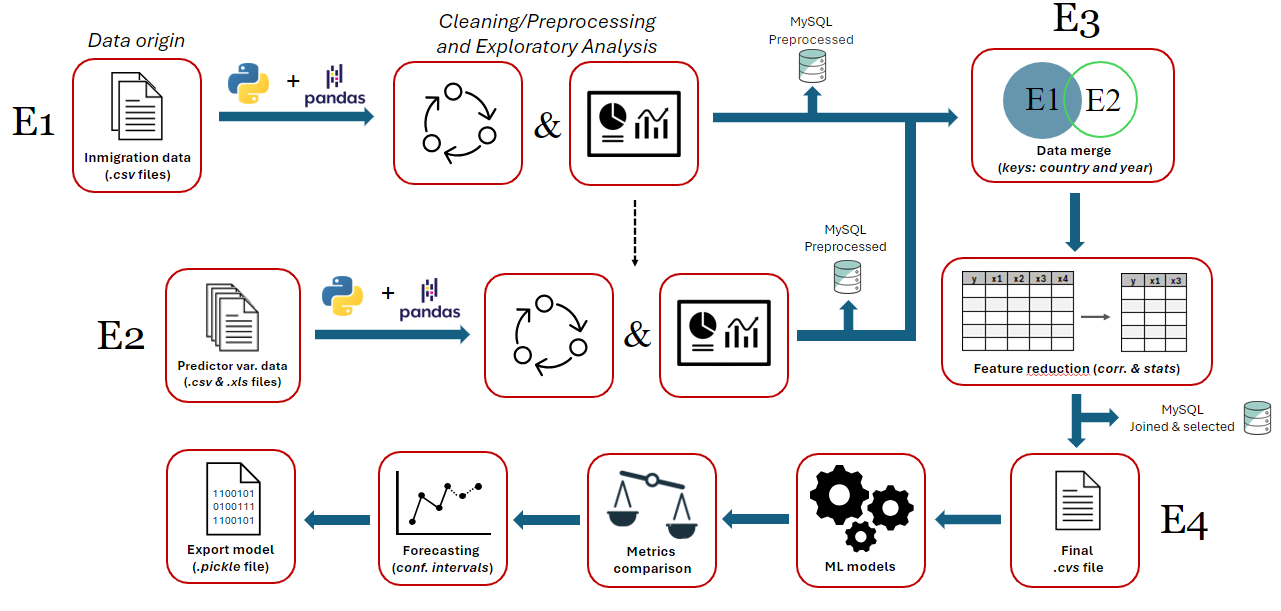
* **Etapa 1:** Centrada en los datos de inmigración de España.
* **Etapa 2**: Enfocada en las variables predictoras a relacionar con los datos de inmigración que alimentaran el modelo predictivo.
* **Etapa 3**: Unificación de las variables para análisis de selección.
* **Etapa 4**: *Machine learning*.

La Figura Y muestra un esquema más detallados de las distintas etapas.

... mencionar los *notebooks*

En primer lugar,

A grandes rasgos, el proyecto busca encontrar desde un punto de vista estadístico, la relación que guarda el fenómeno de inmigración (representado por el número de personas que migran a España) con variables del tipo macroeconómico, social y político. Exploramos una causalidad mediante la aplicación de estudio estadístico (exploración de datos, preprocesamiento, reducción dimensional y de variables y entrenamiento de modelos de *machine learning*) con la intención de encontrar una relación veraz, descriptiva y tangible sobre este fenómeno y su trasfondo.



# Resultados

### Adquisición de fuentes de datos

(pendiente:)

La adquisición de datos originales supuso el planteamiento del objetivo en distintas fases o partes que mutuamente se complementaran. El fenómeno de la migración depende de múltiples factores macroeconómicos, sociales, políticos y también de factores personales, psicológicos o laborales de cada persona sujeto. Los principales factores que identificamos para elaborar nuestros *datasets* originales se dividían en las siguientes categorías:

1. Distribución poblacional por residentes permanentes en España, de origen foráneo
2. Países de habla hispana e inglesa
3. Países de habla hispana oficial
4. Países y regiones del mundo (división política)
5. Índices de desarrollo económico (tales como PIB, estabilidad política, estado de derecho, etc.)
6. Conflictos armados en el mundo (actuales y pasados)
7. Cantidad de turistas anuales en España
8. Índices sobre libertad de expresión o respeto a derechos humanos
9. Datos sobre el padrón de residentes en España
10. Régimen político de países alrededor del mundo
11. Índices de libertad democrática

Las fuentes de datos contemplan estudios o pronunciamientos de organismos internacionales, ONG y gobiernos.

### Descripción de fuentes de datos

Nuestros datos comprenden de distintos tipos de *datasets* que miden fenómenos que poco tienen que ver entre sí, si los miramos desde un punto de vista pragmático. Contamos con variables que cuentan población en número de residentes, valores booleanos para representar ausencia o presencia de alguna característica como un índice y también valores estandarizados de 0 a 1 para representar el grado de solidez de un indicador.

En el Anexo, presentamos una descripción detallada de todas las variables que usamos en nuestros *datasets*, misma que fue sujeta a un preprocesamiento, análisis exploratorio y reducción dimensional previo al entrenamiento de modelos estadísticos.

### Análisis Exploratorio de Datos o EDA

El análisis exploratorio fue dividido en varios *notebooks* con la finalidad de evitar una confusión al introducir muchas variables en uno solo.

A manera de exposición, exponemos los gráficos más relevantes para comprender el tipo de análisis que ejecutamos, manteniendo el análisis completo en cada *notebook*.

A graph with a line and a blue line

Description automatically generated

Figura 1, total de inmigrantes en España de 2008 a 2022, elaboración propia

A graph of a number of people

Description automatically generated

Figura 2, distribución de la población inmigrante en España por sexo y grupo de edad, elaboración propia

Entre 2008-2013 hay una caída de número de inmigrantes y a partir del 2014 aumenta progresivamente hasta alcanzar un pico de 750,480 inmigrantes en 2019, para luego disminuir en el 2020 y 2021 (período de pandemia y restricciones sanitarias relacionadas al COVID 19) y mostrar un gran salto hasta ≈1.25 millones de inmigrantes en el 2022 (post-pandemia), superando al 2019 en aproximadamente 500,000 inmigrantes.

Observando los inmigrantes que llegaron a España durante estos 15 años según la edad, vemos que predomina la inmigración de personas jóvenes de entre 25-34 años, seguidos de jóvenes de entre 15-24. A partir de ese punto, es notable la disminución progresiva del número de inmigrantes con el incremento de la edad.

En cuanto al sexo, vemos que la cantidad de mujeres y hombres es similar en todos los grupos, especialmente en los grupos de mayor presencia (15-24 y 25-34 años). Únicamente los grupos de entre 0-14 y 55-64 años muestran una mayor diferencia, el primero hacia los hombres y el segundo hacia las mujeres.

A screenshot of a computer

Description automatically generated

Figura 3, distribución poblacional de inmigrantes en España por cada año de 2008 a 2022, elaboración propia

La distribución por año parece indicar que a medida que hay menor inmigración, la dispersión de los datos es menor y los valores atípicos menos extremos en relación con el rango intercuartílico (caja), mientras que cuando hay aumenta la inmigración en el tiempo, la dispersión también aumenta.

Además, vemos que la mayoría de los datos se concentran en un rango medio/bajo de número de inmigrantes (observar rango intercuartílico entre 2008-2021 que se mantiene alrededor o por debajo de 10,000) y sólo algunas nacionalidades presentan valores "atípicamente" superiores al resto. Y son precisamente estas nacionalidades con grandes valores atípicos los que son de nuestro interés de estudio y modelado, pues en ellas se condensan la mayor masa de inmigrantes hacia España. Más aún, es interesante observar cómo algunas nacionalidades mantienen un número atípicamente alto de inmigración a lo largo de todo el período 2008-2022, como son españoles, colombianos, rumanos y marroquíes. A graph of blue rectangular objects

Description automatically generated with medium confidence

Figura 4, porcentaje por total de inmigrantes en España por país de origen, elaboración propia

Previamente, ya habíamos notado la cantidad de inmigrantes españoles, colombianos, marroquíes y rumanos, y ahora lo observamos también en su distribución porcentual en base al total.

Luego de los top 4, parece variación porcentual menor entre las subsiguientes nacionalidades.

A graph of blue rectangular bars with white text

Description automatically generated

Figura 5, desempeño por indicadores de libertad para Argentina de 2008 a 2022 en promedio, elaboración propia

Para los índices de libertad, obtuvimos gráficas por cada país para conocer el desempeño por cada uno de los indicadores. En el caso de Argentina, observamos que las libertades civiles alcanzan un grado de plenitud (donde 1 significa certeza y 0 ausencia). En general, los países seleccionados obtienen un promedio de desempeño alrededor del 0.5 de todos los indicadores.

A graph with a line

Description automatically generated

Figura 6, desempeño global en indicador de igualdad en salubridad de 2008 a 2022, elaboración propia

Este indicador es particularmente interesante dado el contexto de la pandemia del COVID. El indicador responde a la accesibilidad de los ciudadanos a servicios de salud de alta calidad, para permitirles ejercer sus derechos políticos en la adultez. En franco declive desde 2008, la situación se ve un poco acelerada para 2019, tocando fondo en 2020 y teniendo un rebote. Aunque se aprecia un esfuerzo global por mejorar los servicios de salud en los países, nuevamente se pierde el avance para 2022.

A graph of a graph of a graph

Description automatically generated with medium confidence

Para efectos del estudio, podemos afirmar que los migrantes en España provienen de un trasfondo diverso, muchos marcados por carencias palpables y otros con un nivel de vida alto. Las gráficas anteriores muestran el ejemplo concreto para Pakistán mismo que tiene altibajos de un indicador a otro.

### Preprocesado de datos

### Selección de variables

### Partición de los datos en conjuntos train y test

### Aplicación de modelos de aprendizaje automático

### Evaluación y validación de modelos de aprendizaje automático

### Comparativa de resultados

# Conclusiones

# Referencias bibliográficas

1. International Organization for Migration. *Factores que propician la migración internacional*. Año 2011. Recuperado de <https://emm.iom.int/es/handbooks/contexto-global-de-la-migracion-internacional/factores-que-propician-la-migracion> el 19 de septiembre 2024.
2. Organización Internacional de Migraciones (OIM). *Informe sobre las Migraciones en el Mundo 2022*. Geneva. Disponible en: <https://publications.iom.int/books/informe-sobre-las-migraciones-en-el-mundo-2022>
3. Banco Europeo de Inversiones (BEI). *Migración y las naciones europeas. Retos, oportunidades y el papel del BEI*. Marzo 2016. Disponible en: <https://publications.iom.int/books/informe-sobre-las-migraciones-en-el-mundo-2022>
4. CLIMB Database: Human Mobility in the Context of Disasters, Climate Change and Environmental Degradation Database. Red de las Naciones Unidas. Disponible en: <https://migrationnetwork.un.org/climb#accordion>
5. Instituto Nacional de Estadística (INE). *Estadística de Migraciones y Cambios de Residencia (EMCR). Año 2022*. Nota de Prensa. Disponible en: <https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736177098&menu=ultiDatos&idp=1254735573002>
6. AYDEMİR Belgin, AYDIN Hakan, ÇETİNKAYA Ali y POLAT Doğan Şafak. *Predicting the Income Groups and Number of Immigrants by Using Machine Learning* *(ML).* International Journal of Multidisciplinary Studies and Innovative Technologies. 2022. 6 (2), 162-168.
7. CARAMMIA Marcello, MARIA IACUS Stefano y WILKIN Teddy*. Forecasting asylum‑related migration flows with machine learning and data at scale*. Nature, Scientific Reports. 2022. 12:1457.

# Anexos