

UNIVERSIDAD AUTÓNOMA DE MADRID  
ESCUELA POLITÉCNICA SUPERIOR



Doble Grado en Ingeniería Informática y  
Matemáticas

TRABAJO FIN DE GRADO

**ESTUDIO DE LA EVOLUCIÓN DE  
LOS DERECHOS LGBT CON  
APRENDIZAJE AUTOMÁTICO**

Autor: Ares Aguilar Sotos  
Tutor: David Domínguez Carreta

Enero 2017



# ESTUDIO DE LA EVOLUCIÓN DE LOS DERECHOS LGBT CON APRENDIZAJE AUTOMÁTICO

Autor: Ares Aguilar Sotos  
Tutor: David Domínguez Carreta

Dpto. de Ingeniería Informática  
Escuela Politécnica Superior  
Universidad Autónoma de Madrid  
Enero 2017



## Resumen

En las últimas décadas, Europa ha sido testigo de la consecución de los logros más importantes en materia de derechos LGBT. Sin embargo, la lucha por la igualdad se remonta siglos atrás, y en la mayor parte de los casos los países muestran una progresión similar en el reconocimiento de estos derechos.

En la era del big data, son numerosas las herramientas y algoritmos de aprendizaje automático que permiten analizar un número elevado de datos de gran complejidad. Concretamente, este trabajo aplica estas técnicas para la predicción de series temporales.

Este proyecto tiene tres objetivos fundamentales: recoger la información disponible en materia de derechos LGBT y procesarla para obtener series temporales, desarrollando mientras tanto una escala que permita medir el estado legal de las personas LGBT en un país determinado; analizar dicha información y establecer predicciones para el futuro con distintos algoritmos de aprendizaje automático y, finalmente, permitir la visualización de dichos datos mediante una interfaz gráfica con mapas interactivos.

Adicionalmente se han empleado algoritmos de clusterización para detectar los países pioneros en cada región geográfica y se ha aplicado el método de los escenarios para evaluar los retos a los que se podría enfrentar nuestra sociedad en el futuro.

Los resultados obtenidos muestran una tendencia al alza en la legislación a favor de la igualdad de las personas LGBT en Europa y América, mientras que en las regiones que actualmente castigan las relaciones homosexuales no se predicen mejoras en el futuro.

## Palabras Clave

Análisis de escenarios, Aprendizaje automático, Big data, Ciencia de datos, Comunidades, Derechos Humanos, Feminismo, Homosexualidad, LGBT, Modelización, Optimización, Predicción, Regresión, Series temporales, Sociología

## **Abstract**

Over the last few decades Europe has witnessed the achievement of the most important milestones in terms of LGBT rights. However, the fight for equal rights started centuries ago, and most countries go through a more or less standard sequence of steps legally recognising homosexuality.

In the age of big data, there are numerous tools and machine learning algorithms to analyze a huge number of highly complex data. In particular, this work applies these techniques to time series prediction.

This project has three main objectives: to gather up the available information in terms of LGBT rights and process it to obtain time series, while developing an index to measure the legal state of LGBT people in a particular country; to analyze such data establishing predictions for the future using different machine learning algorithms and -eventually- to display those data through a graphical user interface with interactive maps.

Besides, clustering algorithms have been used in order to detect the pioneer countries in every geographical region and scenario analysis has been applied with the purpose of evaluating the different challenges that our society may face in the future.

The results obtained show an increasing trend in the legal recognition of LGBT equality in Europe and America, whereas no improvement is predicted in those regions that currently punish homosexual orientation.

## **Key words**

Big data, Communities, Data science, Feminism, Homosexuality, Human rights, LGBT, Machine learning, Modelization, Optimization, Prediction, Regression, Scenario analysis, Sociology, Time series

# Agradecimientos

Los derechos solo se ganan por aquellos  
que hacen oír su voz

---

Harvey Milk

*A mi madre y a mi padre, que me enseñaron a levantar la voz contra las injusticias.*

*A David y a Mar, por confiar en mi trabajo y permitirme desarrollar mis ideas.*

*A Dani, conociéndole me conozco.*

*Y a todas las personas del colectivo LGBT que han dado su vida por el resto.  
Gracias a ellas este trabajo ha sido posible.*



# Índice general

<b>Índice de Figuras</b>	<b>IX</b>
<b>Índice de Tablas</b>	<b>x</b>
<b>1. Introducción</b>	<b>1</b>
1.1. Motivación . . . . .	1
1.2. Definición del proyecto . . . . .	1
1.2.1. Objetivos y alcance . . . . .	1
1.2.2. Metodología . . . . .	3
1.3. Tecnología y herramientas empleadas . . . . .	3
1.4. Estructura de la memoria . . . . .	4
<b>2. Estado del arte</b>	<b>5</b>
2.1. Introducción. Las personas LGBT en la actualidad . . . . .	5
2.2. Medidores de la igualdad LGBT . . . . .	6
2.3. Técnicas para el estudio de series temporales . . . . .	8
2.4. Predecir el futuro. El análisis de escenarios . . . . .	8
2.5. Problemática y retos futuros . . . . .	9
<b>3. Bases de datos y algoritmos empleados</b>	<b>11</b>
3.1. Tratamiento de datos . . . . .	11
3.1.1. Base de datos empleada . . . . .	11
3.1.2. Cuantificación de los derechos de los homosexuales . . . . .	11
3.2. Análisis de los algoritmos . . . . .	12
3.2.1. Algoritmos de regresión . . . . .	13
3.2.2. Algoritmos de clusterización . . . . .	17
<b>4. Estudio de la igualdad de derechos de los homosexuales</b>	<b>19</b>
4.1. Estudio de la serie temporal . . . . .	19
4.1.1. Perspectiva histórica . . . . .	20
4.1.2. Predicción . . . . .	23

4.2. Estudio de escenarios . . . . .	25
4.3. Estudio de pioneros y seguidores . . . . .	27
<b>5. Conclusiones y trabajo futuro</b>	<b>31</b>
5.1. Conclusiones . . . . .	31
5.1.1. Predicción de series temporales con aprendizaje automático . . . . .	31
5.1.2. Los derechos de los homosexuales en las próximas décadas . . . . .	32
5.2. Trabajo futuro . . . . .	32
<b>Glosario de acrónimos</b>	<b>35</b>
<b>Bibliografía</b>	<b>36</b>
<b>A. Base de datos empleada</b>	<b>41</b>
<b>B. Análisis de la aplicación por subsistemas</b>	<b>43</b>

# Índice de Figuras

3.1. Porcentaje de error de k-NN . . . . .	14
3.2. Porcentaje de error en selvas aleatorias . . . . .	15
3.3. Porcentaje de error en redes neuronales . . . . .	16
4.1. GILRHO en el año 1960 . . . . .	20
4.2. GILRHO en el año 1980 . . . . .	21
4.3. GILRHO en el año 2000 . . . . .	21
4.4. GILRHO en el año 2010 . . . . .	22
4.5. GILRHO en el año 2016 . . . . .	23
4.6. Predicción del año 2025 . . . . .	24
4.7. Predicción del año 2045 . . . . .	24
4.8. Escenario optimista del año 2025 . . . . .	26
4.9. Escenario pesimista del año 2025 . . . . .	26
4.10. Dendograma de la clusterización del GILRHO . . . . .	27
4.11. Clustering de 2 agrupaciones . . . . .	28
4.12. Clustering de 5 agrupaciones . . . . .	28
A.1. Muestra de la base de datos de Waal09 . . . . .	42
A.2. Muestra de la base de datos de Waal14 . . . . .	42
B.1. Diagrama por componentes de la aplicación . . . . .	44
B.2. Aplicación de visualización de los resultados . . . . .	44



# Índice de Tablas

4.1. Países pioneros en las leyes estudiadas por el GILRHO . . . . .	27
--	----



# 1

## Introducción

### 1.1. Motivación

---

La lucha de las personas homosexuales por su reconocimiento en la sociedad se remonta a las primeras décadas del siglo XX, en el contexto del proceso judicial a Oscar Wilde y el surgimiento de las primeras organizaciones homosexuales en Alemania. Este primer movimiento por la despenalización de la homosexualidad fue rápidamente apagado con la llegada de Hitler al poder en 1933 [1].

Tras la Segunda Guerra Mundial surgió en Europa Occidental y Estados Unidos el movimiento homófilo, que pretendía conseguir la respetabilidad de las personas homosexuales con la difusión de estudios científicos sobre la homosexualidad. Sin embargo, los disturbios de Stonewall, el 28 de junio de 1969, producidos en respuesta a la persecución policial de la comunidad LGBT, supusieron un punto de inflexión en la historia del movimiento.

Es en este contexto que comienza nuestro estudio del movimiento LGBT, que conseguirá la despenalización de la homosexualidad en la mayor parte de Europa y América. La siguiente reivindicación de los colectivos a principios del siglo XXI será el reconocimiento de las uniones civiles y el matrimonio de las personas del mismo sexo, así como luchar por la despenalización en el resto del mundo. A día de hoy, cuando 73 Estados criminalizan la actividad sexual entre personas del mismo sexo [2], y se asesina a personas homosexuales por el mero hecho de serlo, estas luchas siguen muy vigentes.

En la actualidad, el aumento de la calidad y el número de publicaciones que estudian el estado de las personas LGBT a través del mundo nos permite tener una visión global, atendiendo a los contextos geográficos e históricos, de la situación legal a la que se enfrentan.

### 1.2. Definición del proyecto

---

#### 1.2.1. Objetivos y alcance

Este trabajo tiene dos objetivos fundamentales: en primer lugar, estudiar la efectividad de varios algoritmos de aprendizaje automático a la hora de predecir la evolución de series temporales y compararlos entre sí; en segundo lugar, difundir la historia de la lucha por los

derechos del colectivo LGBT de forma didáctica a través de mapas interactivos, así como estudiar y predecir las tendencias mundiales en la igualdad de derechos. Para alcanzar estos objetivos, se han establecido cuatro hitos fundamentales.

El primer hito es la obtención de los datos legales necesarios para el estudio. Para ello, se deberán investigar las bases de datos existentes y, si fuera necesario, generar nuevas bases de datos que representen la evolución temporal. Además, se deberán estudiar desde un punto de vista estadístico para asegurar su calidad y realizar un primer análisis.

Por otra parte, será necesario desarrollar el código de los modelos que se utilizarán para predecir. Para ello, se compararán los resultados de distintos clasificadores y se optimizarán sus parámetros, para finalmente seleccionar los mejores en nuestro estudio. Una vez definidos los modelos a utilizar, se estudiarán distintas técnicas de agrupamiento y distintas configuraciones de clasificadores para minimizar el error de la predicción.

En tercer lugar se desarrollará una interfaz gráfica que permita visualizar la evolución temporal de los derechos de las personas LGBT y las predicciones generadas para el futuro y comparar los datos generados por cada uno de los algoritmos. Asimismo, se contempla incorporar a la interfaz información adicional de la situación legal en cada país.

En cuanto al alcance, se ha limitado a la situación legal de las personas homosexuales, dejando fuera de este estudio otras identidades sexuales dentro del colectivo LGBT expuestas a una discriminación incluso mayor.

Como veremos en el estudio, existe actualmente una gran limitación a la hora de obtener datos legales que reflejen el estado de las personas LGBT a lo largo del mundo, ya que es un movimiento relativamente reciente. Este es el motivo por el que este trabajo se centra en los varones homosexuales, siendo este un grupo privilegiado dentro de su situación de discriminación. No obstante, los datos legales en cuanto a los hombres homosexuales son perfectamente válidos para los bisexuales.

Por otra parte, las mujeres están ya discriminadas por el simple hecho de serlo en todo el mundo. Una de las vertientes de esta discriminación es la invisibilización de su sexualidad [3], que hace que en aquellos países donde la homosexualidad masculina está castigada, la femenina se permita. Esto no se debe a una aceptación social del lesbianismo, sino que responde, como ya se ha dicho, a la discriminación de la mujer; ya que ni siquiera se considera que la misma tenga una sexualidad separada del placer del hombre.

Por este hecho, este estudio no refleja la situación particular de las mujeres en los países que criminalizan la homosexualidad (en los países que la protegen se considera la homosexualidad sin importar el sexo), pero en la práctica su situación es, en este aspecto, la misma.

Finalmente, no nos podemos olvidar del colectivo transexual, fundamental en la defensa de los derechos LGBT, y uno de los más discriminados en todo el mundo. La situación de las personas transexuales es radicalmente distinta a la de las homosexuales: algunos países, como India, criminalizan la homosexualidad pero permiten la transexualidad; mientras que otros, como España, tienen leyes que protegen a las personas homosexuales pero siguen considerando la transexualidad una enfermedad.

Lo anterior deja claro que el campo de estudio está en continua evolución, por lo que la aplicación que se desarrolle deberá ser fácilmente adaptable a todas las particularidades anteriormente descritas. Por otra parte, el objetivo pedagógico obliga a que la aplicación sea intuitiva y los resultados se visualicen de forma clara y precisa.

### 1.2.2. Metodología

Tras analizar el estado del arte y estudiar los datos existentes relativos a los derechos de los homosexuales, quedó claro que se debería emplear una metodología ágil que facilitase la adaptación del proyecto a los problemas y necesidades conforme surgiesen.

Para la recolección de datos se comenzó con un estudio de los informes más importantes mundialmente de los derechos de los gays. De estos informes se extrajeron los factores más relevantes a la hora de calificar la situación de las personas LGBT en un país. Mediante técnicas de procesamiento de textos, estos indicadores se combinaron en la base de datos que emplearemos. Esta fase tuvo que ser desarrollada en dos ciclos, ya que una vez finalizado el primero se publicó el informe Homofobia de Estado de 2016 [2], que añadimos a los datos que teníamos.

Una vez generadas las bases de datos necesarias comenzó el desarrollo de los clasificadores. Este desarrollo se dividió en un estudio inicial mediante Weka de los clasificadores más adecuados para el trabajo, la optimización de los parámetros más relevantes en cada uno de ellos, y, una vez finalizado lo anterior, el desarrollo de técnicas para la combinación de dichos clasificadores.

Con los primeros resultados generados se contempló la posibilidad de estudiar el impacto geográfico en los derechos de las personas LGBT. Para ello se definieron dos nuevos hitos: la incorporación de algoritmos de clustering al estudio; y, a raíz de esto, la clasificación de los países en pioneros y seguidores.

Cuando ya se habían obtenido los resultados, se produjo la mayor masacre de personas LGBT del siglo en Orlando. Esta noticia supuso un verdadero shock y me llevó a plantearme la validez del proyecto. Así, se introdujo un nuevo hito: la incorporación a los métodos de aprendizaje automático del método de los escenarios [4], ampliamente utilizado en el campo económico, que permitiese realizar distintos análisis variando el grado de optimismo para obtener predicciones pesimistas y optimistas además de las ya generadas.

Finalmente se procedió a desarrollar la aplicación web que permitiese visualizar los datos del proyecto. A su vez, este desarrollo se subdividió en dos hitos fundamentales: la visualización de datos geotemporales en primer lugar y la aplicación completa en segundo, que incluirá dicho visualizador además de información estadística sobre las bases de datos.

## 1.3. Tecnología y herramientas empleadas

El objetivo didáctico está presente en todo el proyecto. Para facilitar su difusión y estudio, desde el primer momento estuvo claro que las tecnologías a emplear deberían ser gratuitas y, dentro de lo posible, libres. Además, el proyecto combina diversas tecnologías de análisis y representación de datos, pero también varios lenguajes de programación.

Por esto, en primer lugar se implementó un entorno de desarrollo que soportase todos ellos. Tras un estudio de las herramientas disponibles, se decidió utilizar Emacs [5], con una configuración personalizada para cada uno de los lenguajes utilizados. En mayor o menor medida, se han empleado los siguientes:

- R: como objetivo adicional del trabajo se ha decidido aprender R para, en primer lugar, generar los estadísticos necesarios de los datos y, en segundo lugar, unir todas las partes del proyecto en una única aplicación gráfica.
- AWK: el procesamiento inicial de los informes y la manipulación de los ficheros de datos se ha realizado, por comodidad, con AWK.

- Java: la parte del proyecto encargada de clasificar y optimizar parámetros está desarrollada en Java. Aunque la mayor parte del código fue utilizada sin modificaciones, en algunos casos se han incluido alteraciones en el código para adaptarlo mejor al problema concreto.

Además, se ha tratado de utilizar todos los recursos disponibles de forma libre en la red. Concretamente, se han utilizado las siguientes bibliotecas y utilidades:

- Weka: plataforma de software para aprendizaje automático y minería de datos de donde se han obtenido la mayor parte de los modelos.
- DataMaps: biblioteca de representación de datos geográficos en javascript que se ha modificado para admitir representación temporal.
- Shiny (paquete de R): marco de trabajo para la implementación de aplicaciones web en R. Se encargará de conectar todas las partes de la aplicación.

## **1.4. Estructura de la memoria**

---

Este documento se divide en cinco secciones. En esta primera, se presentan la motivación y los objetivos del proyecto y se define el alcance del mismo; a la vez que se definen la metodología y las herramientas empleadas.

La segunda sección recoge el análisis del estado del arte y repasa los estudios actuales sobre la legislación LGBT, así como los intentos que ha habido hasta la fecha de crear un índice que los cuantifique. Además, se discuten las técnicas actuales de predicción de series temporales, así como el método de los escenarios, y se plantean los problemas y retos del futuro.

Los algoritmos y los datos empleados en el estudio se describen detalladamente en la sección tercera. Por una parte, se presenta el método empleado para cuantificar los derechos de los homosexuales. Por otra, se describen teóricamente los algoritmos que se empleara, así como los detalles de su implementación.

En cuarto lugar se presenta el estudio de la igualdad de derechos de los homosexuales haciendo uso de los algoritmos anteriores. Se estudia no solo la serie temporal generada, sino también los países pioneros y seguidores, y se aplica el método de los escenarios para obtener una visión más rica del problema.

Por último, se recogen las conclusiones del trabajo. Estas conclusiones van en dos direcciones: la efectividad de los algoritmos de aprendizaje automático en la predicción de series temporales, y la evolución esperada para las próximas décadas en los derechos de los homosexuales. Además, se incluyen en anexos muestras de las bases de datos empleadas y el análisis por subsistemas de la aplicación desarrollada.

# 2

## Estado del arte

### 2.1. Introducción. Las personas LGBT en la actualidad

---

Según la última edición de *Homofobia de Estado* [2], 73 Estados criminalizan la actividad sexual entre dos personas del mismo sexo. Al mismo tiempo, el número de países donde está permitida la unión civil en cualquier grado ha ascendido a 46. Por otra parte, en los últimos años han proliferado las conocidas como "leyes de propaganda", que censuran la libertad de expresión de las personas LGBT y sus prácticas sexuales.

Numerosas son las voces que se han alzado contra esta situación. La ILGA (*International Lesbian and Gay Association*) publica anualmente su informe *Homofobia de Estado* en un intento de visibilizar las problemáticas LGBT mundiales y concienciar a los gobiernos para poner fin a estas situaciones de discriminación.

Desde el punto de vista legal, las acciones que pide la comunidad LGBT se engloban en tres bloques: en primer lugar, la descriminalización de las prácticas sexuales entre personas del mismo sexo; en segundo lugar, el establecimiento de leyes que prohíban explícitamente la discriminación en todos los aspectos por razón de condición sexual; y por último, el reconocimiento de las parejas de personas del mismo sexo con los mismos derechos que el matrimonio heterosexual.

En esta línea, ILGA divide el mundo en tres bloques de países: los que criminalizan la homosexualidad (**criminalización**), los que protegen a la comunidad mediante leyes explícitas (**protección**), y aquellos en los que se da la igualdad real en términos legales (**reconocimiento**). Cada uno de estos tres bloques se divide según el grado en el que se encuentre cada país.

Esta visión presenta un inconveniente: es frecuente que en países donde las relaciones entre personas del mismo sexo están permitidas, la homofobia social impida a las personas LGBT vivir su vida en libertad, por lo que el mapa de situación sugiere un resultado mucho más optimista del que se presenta en la actualidad.

Por esto, otros estudios como *The Global Divide on Homosexuality* [6] enfocan el problema ya no desde el punto de vista legal sino del de la opinión pública. Este estudio muestra en toda Europa, Estados Unidos y la mayor parte de América Latina, una tendencia al alza de la aceptación de la homosexualidad. Sin embargo, también refleja la variación de la opinión durante la crisis económica: el auge de la extrema derecha en países como Francia tiene su reflejo en la disminución de 9 puntos porcentuales en la aceptación de la homosexualidad. También

se ve claramente el efecto de las leyes anti-propaganda en Rusia, donde la aceptación de la homosexualidad ha disminuído 6 puntos respecto a 2002.

Además, se pone de manifiesto la preocupante situación en la que se encuentran las personas LGBT en África y, en general, en los países musulmanes. En África subsahariana, al menos 9 de cada 10 personas consideran que la homosexualidad no debería ser permitida, un porcentaje similar al que encontramos en la mayoría de países musulmanes.

Por otra parte, el efecto positivo sobre la opinión pública de la aprobación de leyes en favor de la comunidad LGBT es también considerable. Tras la aprobación de las leyes contra la discriminación y del matrimonio entre personas del mismo sexo, la opinión pública argentina favorable a la homosexualidad ha aumentado en 8 puntos porcentuales. Este efecto también ocurre en otros países de América Latina, como México o Chile, donde ya son dos tercios de la población quienes tienen una opinión favorable de la homosexualidad [6].

Esto señala un punto importante que el presente trabajo pretende abordar: la influencia de los países vecinos en las dinámicas sociales. En Europa por ejemplo, la igualdad de derechos para las personas LGBT ha ido en notable aumento desde la aprobación del matrimonio homosexual en los Países Bajos en 2001. Argentina, por su parte, actúa como país pionero en derechos LGBT en América Latina. Desde su adopción del matrimonio homosexual en 2010, Brasil, Uruguay y Colombia han adoptado medidas similares. En otros países de la región, con la opinión pública desfavorable, como Ecuador o Costa Rica, también se están adoptando medidas en la misma dirección.

En definitiva, en los últimos años hemos visto en occidente un interés creciente por la situación de las personas LGBT en todo el mundo. El número y la calidad de los informes ha aumentado notablemente, y las organizaciones más importantes en la materia están realizando un trabajo constante en aunar esfuerzos en la elaboración de estudios y otros materiales que visibilicen la problemática actual. Estos informes ayudan a ONGs, activistas LGBT, agencias gubernamentales y de la ONU a obtener una visión global de la problemática LGBT mundial. Es difícil exagerar la importancia de estas herramientas por cuanto nos permiten medir el progreso (o retroceso) en el transcurso de los años, al tiempo que exponen la arbitrariedad de las leyes persecutorias sobre la base de la orientación sexual de las personas.

## 2.2. Medidores de la igualdad LGBT

---

La lucha de las personas LGBT por la igualdad de derechos ha sido protagonista de la historia reciente de la humanidad. No obstante, hasta el siglo XXI, con la explosión del interés en el *Big Data*, no se ha empezado a trabajar en la elaboración de bases de datos que recojan de manera sistemática la situación de estas personas en todo el mundo.

En el año 2006, ILGA publicó por primera vez su informe *Homofobia de Estado*, donde se reoge información de la situación legal de las personas LGBT en los distintos países. Este informe muestra la situación actual de la legislación en materia LGBT cada año, y categoriza a los países según si criminalizan la homosexualidad, si prohíben la discriminación por sexualidad en distintos ámbitos y si permiten el matrimonio y la adopción de parejas del mismo sexo.

Sin embargo, *Homofobia de Estado* presenta una instantánea de los derechos de las personas LGBT cada año, centrándose más en poner de manifiesto los problemas actuales que en evaluar la tendencia que presenta cada país estudiado. No obstante, es el informe más importante a nivel mundial en esta materia, y el estudio de los materiales gráficos que lo acompañan permite obtener una visión global de la problemática actual.

El primer intento de crear un índice global de la inclusión de las personas LGBT llegó en el año 2009, de la mano de un estudio de Kees Waaldijk titulado *Legal recognition of homosexual orientation in the countries of the world* [7]. Este primer estudio sistematiza los medidores para la igualdad LGBT en el mundo, y los divide en 4 bloques de 2 medidores cada uno:

1. Discriminación: Estudia si las relaciones homosexuales son legales y con el mismo límite de edad que las heterosexuales.
2. Anti-discriminación: Se incluye si el país tiene medidas de protección de los homosexuales en materia laboral y en el resto de servicios (educación, vivienda o salud).
3. Reconocimiento de parejas: Se contempla si existe reconocimiento legal de cohabitación entre personas del mismo sexo y si existe la unión civil.
4. Reconocimiento familiar: Este bloque estudia si existe la posibilidad de matrimonio y adopción por parte de las parejas del mismo sexo.

Este estudio no se limita a plasmar la situación del momento en los distintos países, sino que recoge por primera vez información de los años en los que ha habido cambios en cada una de las cuestiones. Así, permite visualizar una tendencia global de mayor aceptación del colectivo LGBT.

No obstante, el propio estudio admite que no se trata de un trabajo completo. En primer lugar, porque trata exclusivamente la situación de la homosexualidad, dejando de lado por ejemplo la problemática transgénero. En segundo lugar, los datos legales disponibles eran muy limitados en algunos países, por lo que abundan los registros incompletos, ambigüedades y errores.

Sin embargo, el trabajo anterior fue continuado en un nuevo informe, *The Relationship between LGBT Inclusion and Economic Development: An Analysis of Emerging Economies* [8]. Este estudio trata de demostrar la tesis de que la exclusión social de las personas lesbianas, gays, bisexuales y transgénero tiene una repercusión negativa en el desarrollo económico de un país. Para ello construye dos índices: el GILRHO (*Global Index on Legal Recognition of Homosexual Orientation*), para medir la situación de la homosexualidad; y el TRI (*Transgender Rights Index*), para atender a la problemática de las personas transgénero.

El GILRHO supone la primera escala numérica para comparar la situación de las personas homosexuales, y se limita a darle un punto a cada medidor de los descritos en [8]. Por tanto, la máxima puntuación que puede obtener un país es 8; si permite los matrimonios homosexuales, las adopciones, y tiene leyes en contra de la discriminación.

Waaldijk estudia en [8] el GILRHO para el periodo comprendido entre los años 1980 y 2014 en un total de 39 países. Las conclusiones a las que llega muestran que cada derecho estudiado por el GILRHO supone un aumento de 1.400\$ en el PIB per cápita de un país, así como un aumento considerable del Índice de Desarrollo Humano (HDI).

Finalmente, el Programa de las Naciones Unidas para el Desarrollo y el Alto Comisionado de Naciones Unidas para los Derechos Humanos han anunciado en 2016 la creación de un índice de inclusión LGBT [2]. Para la elaboración de este índice, que girará en torno a cuatro puntos básicos (bienestar económico, participación cívica y política, seguridad personal y violencia, y salud), están colaborando representantes de la ONU, activistas LGBT y expertos en reunión de datos.

Se pretende que este índice tenga una naturaleza mundial y evolucione a lo largo del tiempo para fomentar la inclusión de las personas LGBT, garantizando que nadie se quede atrás.

## 2.3. Técnicas para el estudio de series temporales

---

A lo largo de la historia, filósofos y matemáticos han tratado de modelar estadísticamente procesos naturales observados a lo largo del tiempo. Los desarrollos teóricos en el análisis de series temporales comenzaron con los procesos estocásticos, aunque la primera aplicación con datos reales de modelos autorregresivos es la de los trabajos de G. U Yule [9] y G. Walker [10] a principios del siglo XX.

En 1970 G. E. P. Box y G. M. Jenkins publican *Time Series Analysis* [11], donde aparece el primer procedimiento completo para cada serie individual: especificación, estimación, diagnóstico y predicción. Los hoy llamados modelos de Box-Jenkins son los más utilizados, y han dado lugar a un gran número de desarrollos como las generalizaciones no lineales, principalmente ARCH (AutoRegressive Conditional Heteroscedasticity) y GARCH (Generalized ARCH). Ambos modelos permiten parametrización y predicción de varianza no constante, y han sido especialmente útiles en series temporales financieras [12]. De hecho, su invención, junto con el lanzamiento del modelo de corrección de error, dieron en 2003 el Premio en Ciencias Económicas en memoria de Alfred Nobel a C. W. J Granger y R. F. Engle.

En las últimas décadas, uno de los modelos más utilizados para la predicción de series temporales son las redes neuronales artificiales (ANN), ya que presentan una complejidad mucho menor a los métodos estadísticos anteriores. Además, estudios como los de Werbos [13] señalan que las ANNs predicen mejor que los métodos estadísticos clásicos y los modelos de Box-Jenkins. Más adelante, se han utilizado otros modelos de aprendizaje automático como árboles de decisión, máquinas de soporte vectorial (SVM) [14], y regresión de vecinos próximos [15].

Precisamente una de las líneas de investigación actuales es la combinación de distintos métodos para producir un único resultado, lo que disminuye notablemente el error final [16]. Estas técnicas se han aplicado en meteorología [17], en medicina [18] y, por supuesto, en economía [19].

## 2.4. Predecir el futuro. El análisis de escenarios

---

El campo de la futurología trata de imaginar escenarios futuros posibles, denominados futuros, y en ocasiones de determinar su probabilidad, con el fin último de planificar las acciones necesarias para evitar o acelerar su ocurrencia.

Esta disciplina, aunque siempre ha estado presente en la historia de la humanidad, se desarrolló a partir de 1946 con la creación de la RAND Corporation, que colaboraba con el ejército estadounidense en la planificación de largo alcance, viendo la tendencia sistemática, el desarrollo de escenarios, y la visión de futuro. Más adelante, las empresas privadas comenzaron a aplicar estas técnicas para la gestión de inversiones, predicciones del mercado, análisis de riesgos y diseño de estrategias. Se considera que los primeros trabajos clave en este campo son *The art of conjecture*, de Bertrand de Jouvenel, e *Inventing the future* de Dennis Gabor, ambos de la década de 1960.

El método de los escenarios, introducido por Herman Kahn en sus trabajos con las fuerzas armadas estadounidenses, no intenta predecir el futuro basándose en los datos del presente, evitando así incorporar las preferencias sociales de una época en su predicción. De esta forma, analiza distintos escenarios futuros posibles, permitiendo así una toma de decisiones más consciente y robusta.

Según lo describe Peter Schwartz, cofundador de la Global Business Network (GBN), el método de los escenarios "(...) es una herramienta para una mejor toma de decisiones (...) Empresas y gobiernos emplean esta herramienta porque les ayuda a tomar mejores decisiones estratégicas"[20].

Aunque en la actualidad el método de los escenarios se emplea ampliamente en la empresa privada y los gobiernos, no existen publicaciones que lo combinen con las predicciones obtenidas mediante inteligencia artificial. No obstante, dado que el campo que nos ocupa tiene un altísimo grado de variabilidad debido a la intervención humana, el uso de este método no solo es adecuado sino necesario para comprender la variabilidad de los futuros que nos esperan.

## **2.5. Problemática y retos futuros**

---

El primer reto al que se enfrentó este trabajo es a la escasa disponibilidad de datos sobre el colectivo LGBT. Esta dificultad se mantiene hoy, ya que la legislación al respecto, en la mayoría de los países, no es fácilmente accesible. Por esto, queda fuera de este trabajo el análisis de las condiciones de vida de las personas transgénero. Uno de los retos en el futuro es precisamente este, incluir medidores para todos los colectivos englobados como LGBT.

Predecir el futuro siempre conlleva un gran riesgo. Conocemos el impacto que tienen las medidas legales positivas o negativas para el colectivo LGBT en la opinión pública, pero no podemos predecir el auge de políticas que las lleven a cabo. Aunque Europa se caracteriza por una consolidada protección al colectivo homosexual, en los últimos años estamos siendo testigos del auge de nuevos fascismos [21]. En Madrid, por ejemplo, este año se han registrado el mayor número de agresiones a personas homosexuales desde que se tiene registro [22].

Por otro lado, nos encontramos con el problema de la opinión pública: existen países sin regulación explícita contraria a las prácticas homosexuales donde estas personas son asesinadas por su condición sexual. Resulta evidente la importancia de la opinión pública en la situación del colectivo LGBT, pero hasta los últimos años no se ha empezado a recoger esta opinión en encuestas, lo que impide incluir estos datos en nuestro análisis.

La ILGA, en su último informe [2], se hace eco de este problema. De hecho, este informe incluye algunos resultados de la encuesta *Global Attitudes Survey on LGBTI people*, hecha en colaboración entre Riwi Corp. y Logo. Se espera que con el avance del tiempo, estos informes de la opinión pública sean mayores y más frecuentes y puedan ser utilizados para mejorar la visión global y las predicciones sobre la situación LGBT mundial.



# 3

## Bases de datos y algoritmos empleados

### 3.1. Tratamiento de datos

---

#### 3.1.1. Base de datos empleada

Para elaborar la base de datos que emplearán los clasificadores, se han considerado variables de tipo temporal, geográfico y legal. A continuación se describe detalladamente cada una de ellas, según su ámbito:

- Temporal: Cada registro de la base de datos recoge el año de la muestra.
- Geográfico: Para incluir información geográfica en cada una de las muestras, se ha decidido incluir la región geográfica del país, según el estándar UN M.49 de la ONU [23]. Además, para representar la proximidad se ha incluido también la subregión de cada país, así como la media del GILRHO de los países de su región en el año anterior.
- Legal: Del trabajo inicial de Wladijk [7] y los informes anuales de la ILGA [2] se han extraído las 8 leyes que contemplamos para construir el GILRHO, y se han procesado para obtener la clase en una escala numérica del 0 al 8. Los detalles de estas bases de datos se encuentran en el anexo A de este trabajo.

#### 3.1.2. Cuantificación de los derechos de los homosexuales

La primera tarea a la que se enfrenta este trabajo es la de establecer una medida a los derechos de los homosexuales. Dado que se trata del único intento a nivel mundial de cuantificarlos, se ha elegido el GILRHO, descrito por Kees Wladijk en [7] y actualizado en [8]. El GILRHO supone una medida en una escala de 0 a 8 puntos, dividida en intervalos de 0,5, del grado de cumplimiento de 8 medidas legales que han sido y son fruto de batallas del colectivo LGBT.

Una de las hipótesis de este trabajo es que la mayoría de los países, en distintas épocas y a distinto ritmo, siguen una secuencia de pasos similar a la hora de reconocer legalmente la homosexualidad [7]. Por tanto, las 8 leyes observadas no tienen un orden arbitrario, sino que reflejan precisamente cada uno de esos pasos. Así, se le otorga 1 punto a un país si cumple una

determinada ley, y 0'5 puntos en el caso de que la cumpla tan sólo en algunas partes del territorio del país.

Las leyes que se consideradan para medir el GILRHO son las siguientes, agrupadas en 4 grandes bloques:

1. Descriminalización
  - a) Descriminalización de las relaciones sexuales entre personas del mismo sexo.
  - b) Igualdad de la edad legal de consentimiento entre relaciones sexuales homosexuales y heterosexuales.
2. Leyes anti-discriminación
  - a) Prohibición explícita de la discriminación en el empleo por motivos de orientación sexual.
  - b) Prohibición de la discriminación en relación a bienes y servicios (incluyendo vivienda, asistencia sanitaria y educación) por motivos de orientación sexual.
3. Reconocimiento de parejas homosexuales
  - a) Reconocimiento legal de la cohabitación no registrada entre personas del mismo sexo.
  - b) Reconocimiento legal de la unión civil o cohabitación registrada entre personas del mismo sexo, con los mismos derechos que las uniones heterosexuales.
4. Reconocimiento familiar
  - a) Reconocimiento legal del matrimonio homosexual.
  - b) Permision de la adopción por parte de personas y/o parejas homosexuales.

Cabe recordar que algunos países no siguen por completo el orden de estas medidas a la hora de aplicarlas, como Gran Bretaña, que introdujo en 1988 una ley en contra de la homosexualidad. Sin embargo, la mayoría sí que lo siguen, como Francia, Dinamarca o Finlandia. Por esto, el GILRHO representa, para un año dado, el estado en que se encuentra un determinado país en la línea temporal del reconocimiento de la homosexualidad.

## 3.2. Análisis de los algoritmos

---

Para la predicción del GILRHO de la serie temporal, se han empleado distintos algoritmos de regresión, considerando así el indicador como una medida continua, aunque en la realidad no lo sea.

Las implementaciones de los algoritmos se han tomado, cuando ha sido posible, de la plataforma Weka (*Waikato Environment for Knowledge Analysis*, [24]), y para acceder a las mismas desde R se ha empleado la biblioteca RWeka [25].

Además, se ha implementado un algoritmo para representar el método de los escenarios y forzar predicciones pesimistas u optimistas, que se utilizará para estudiar variaciones en el comportamiento de los algoritmos.

Finalmente, también se han empleado algoritmos de clusterización para estudiar las líneas temporales de los países hasta la actualidad, para ver su comportamiento en términos de pioneros y seguidores.

Para entrenar los algoritmos, así como para optimizar sus parámetros, se ha seguido en todo momento un esquema de validación cruzada [26] de cinco particiones.

En esta sección se presentan, además de un análisis teórico de cada algoritmo, detalles sobre su implementación, así como la metodología seguida para optimizar sus parámetros.

### 3.2.1. Algoritmos de regresión

Los algoritmos que se describen en esta sección pertenecen a los modelos autoregresivos de media móvil (ARMA [11]) usados ampliamente en el campo de la meteorología.

Dada una serie temporal  $X_t$ , estos modelos predicen los valores de dicha serie en dos partes: en primer lugar una parte autoregresiva que se encarga de predecir los valores futuros en función de los observados; y en segundo, la parte de media móvil, que modela el error obtenido como una combinación lineal de los errores pasados de la serie.

En nuestro caso, la media móvil se ha implementado añadiendo sendos atributos a cada ejemplo de la muestra para representar el GILRHO medio en la subregión en el año anterior. Por tanto, se ha añadido un paso extra a todos los algoritmos para actualizar estos valores antes de predecir.

### Vecinos Próximos

#### 1. Descripción

El algoritmo vecinos próximos, o k-nn (K nearest neighbors [27]) estima la probabilidad a posteriori de que un elemento  $x$  tenga un determinado valor a partir de los datos aportados por el conjunto de entrenamiento.

Originalmente, este es un algoritmo de clasificación, pero se ha implementado la regresión calculando la media entre los k vecinos más próximos en lugar de simplemente asignar la clase del más cercano.

#### 2. Implementación

Para implementar vecinos próximos, se ha utilizado la clase IBk [28] de Weka, que permite modificar la función de distancia para obtener los k vecinos más próximos.

En nuestro caso, se ha utilizado una búsqueda por fuerza bruta, y se ha implementado una variación sobre la distancia euclídea para dar más importancia a los países con evoluciones de la serie temporal más similares, en lugar de considerar únicamente a los más cercanos espacialmente.

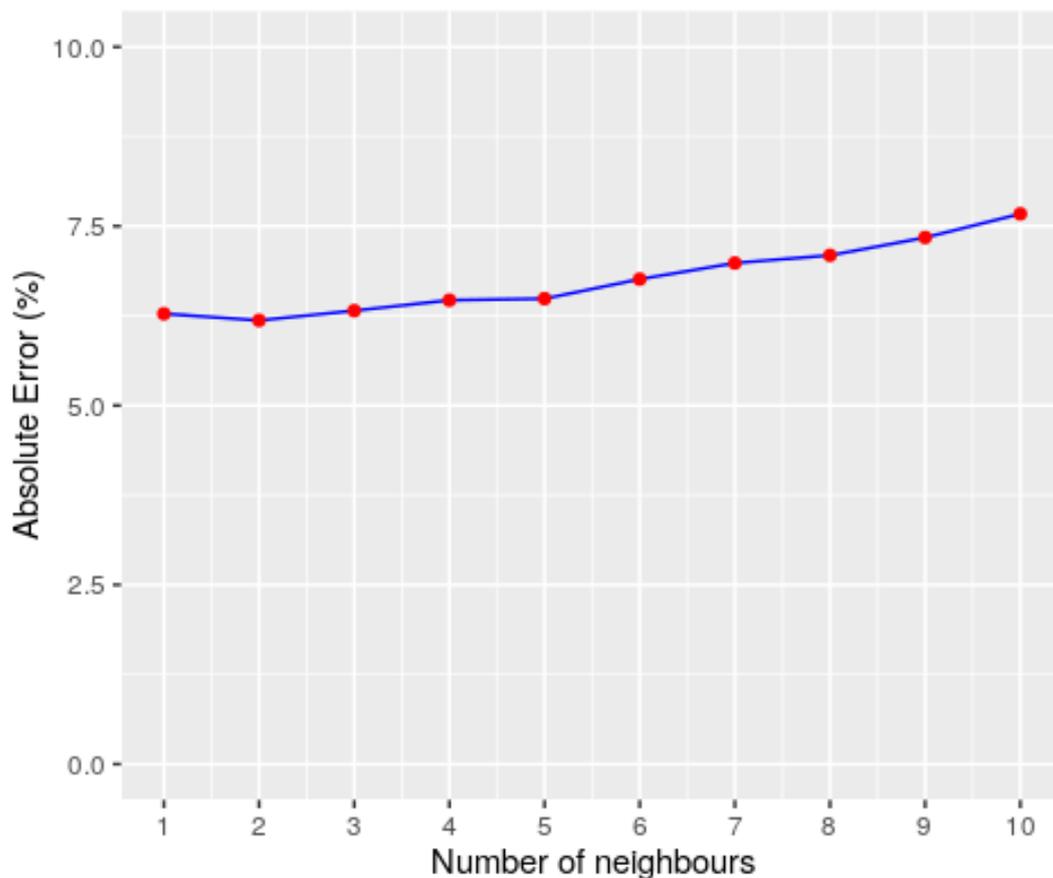
#### 3. Parametrización

En el caso de vecinos próximos, existe tan sólo un parámetro que optimizar (sin tener en cuenta la función de distancia a emplear, que ya se ha definido en la sección anterior) y es el número  $k$  de vecinos.

Para parametrizarlo, se ha entrenado el algoritmo con los valores de  $k$  variando entre 1 y 10 y se ha medido el error porcentual siguiendo un esquema de validación cruzada en 5 particiones.

En la figura 3.1 se aprecia claramente un mínimo en el error con  $k = 2$ . Sin embargo, para evitar un sobreajuste de los datos de entrenamiento, se ha escogido un valor de tres vecinos.

Figura 3.1: Porcentaje de error de k-NN



## Selvas Aleatorias

### 1. Descripción

Las selvas aleatorias (o random forest) son un método de empaquetado o bagging en el que varios árboles de decisión se combinan para obtener una única predicción. Este algoritmo fue desarrollado por Leo Breiman y Adele Cutler [29], y hoy en día es uno de los más utilizados por su eficiencia y certeza en las predicciones.

Durante la fase de entrenamiento, cada árbol se entrena independientemente del resto con la misma distribución de vectores aleatorios, pero con una selección de atributos iniciales diferente. Para predecir una variable numérica, se promedian las predicciones de cada uno de los árboles entrenados, obteniendo así un único valor.

### 2. Implementación

La implementación que se ha utilizado para las selvas aleatorias es el método de Weka RandomForest, entrenando todos los árboles con el 100 % del conjunto de entrenamiento y con profundidad ilimitada.

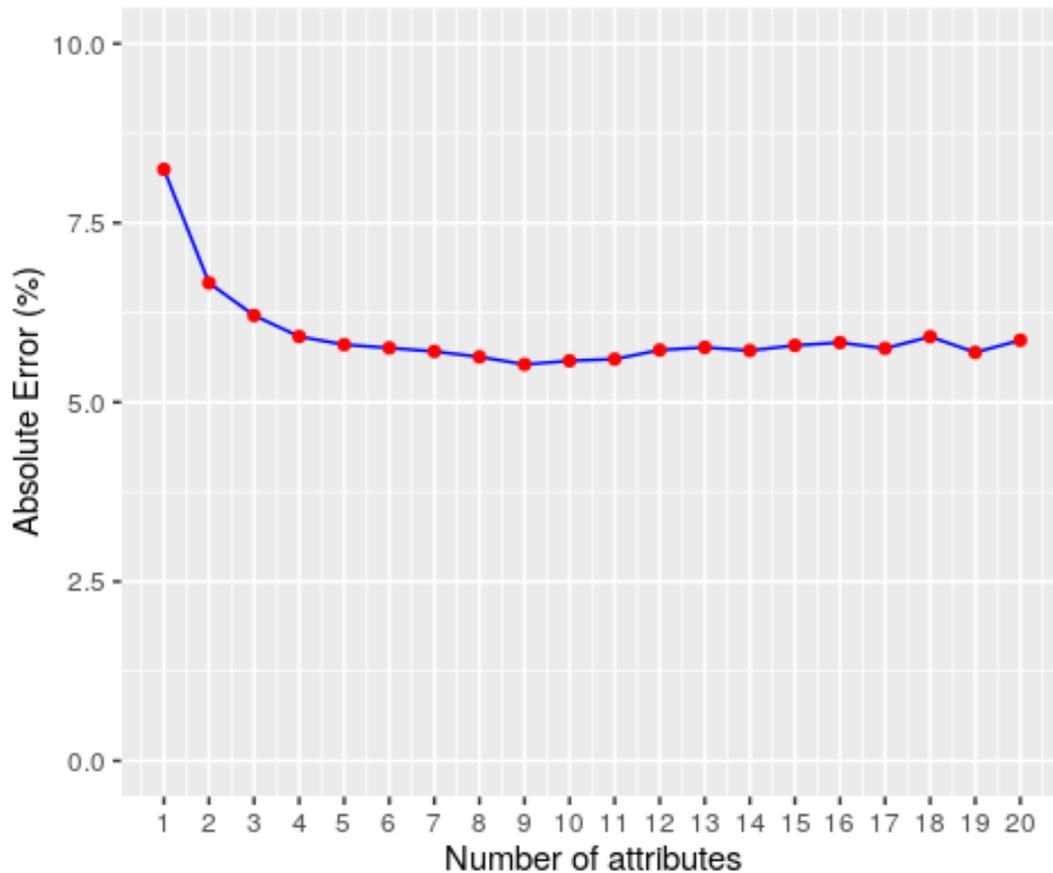
### 3. Parametrización

Una de las ventajas de las selvas aleatorias es que, salvo contadas excepciones [30], no sobreajustan los datos. Por tanto, el número óptimo de árboles a utilizar será siempre el mayor posible con la capacidad de cómputo disponible.

Sin embargo, el número de atributos que emplea cada árbol es un parámetro determinante en la certeza de las predicciones [31] y debe ser optimizado. Dado que el número óptimo

de atributos a emplear suele estar cerca de la raíz cuadrada del total, se han entrenado selvas aleatorias empleando de 1 a 20 atributos para minimizar el error.

Figura 3.2: Porcentaje de error en selvas aleatorias



Como muestra la figura 3.2, se alcanza un mínimo al utilizar un total de 9 atributos; por tanto, este será el valor que se ha dado al parámetro en la predicción.

## Redes Neuronales con Retropropagación

### 1. Descripción

Las redes de neuronas artificiales (o ANN, *Artificial Neural Networks*), introducidas en los trabajos de Warren McCulloch y Walter Pitts [32], son un método de aprendizaje automático basado en las interconexiones formadas por las neuronas en el cerebro. Una red neuronal consiste en unidades mínimas de computación o neuronas distribuidas en distintas capas, que pueden estar activas o inactivas en un momento dado, y que están conectadas entre sí.

Los parámetros que definen una red neuronal son de tres tipos: el grafo de conexión entre las neuronas, la función de aprendizaje que actualiza los pesos de dichas neuronas y la función de activación, encargada de convertir las entradas de la neurona en un valor de salida.

Las redes neuronales son, a día de hoy, uno de los algoritmos de aprendizaje automático más utilizados en una inmensa cantidad de campos que van desde la medicina [33] hasta la bolsa [34].

## 2. Implementación

Para la implementación de la red neuronal, se ha optado por la siguiente configuración de parámetros: en primer lugar, se ha optado por emplear el perceptrón [35] como modelo de neurona; y un grafo de tres capas, de entrada, salida y una capa oculta.

En cuanto a la función de aprendizaje, se ha decidido emplear, como es común en series temporales, retropropagación (del inglés *backpropagation*) [36]. Este algoritmo emplea dos fases: en la primera se produce la salida para un vector de entrada y en la segunda se calcula el error y se propaga desde la capa de salida hacia atrás al resto de neuronas, que actualizan sus pesos en función de los errores cometidos.

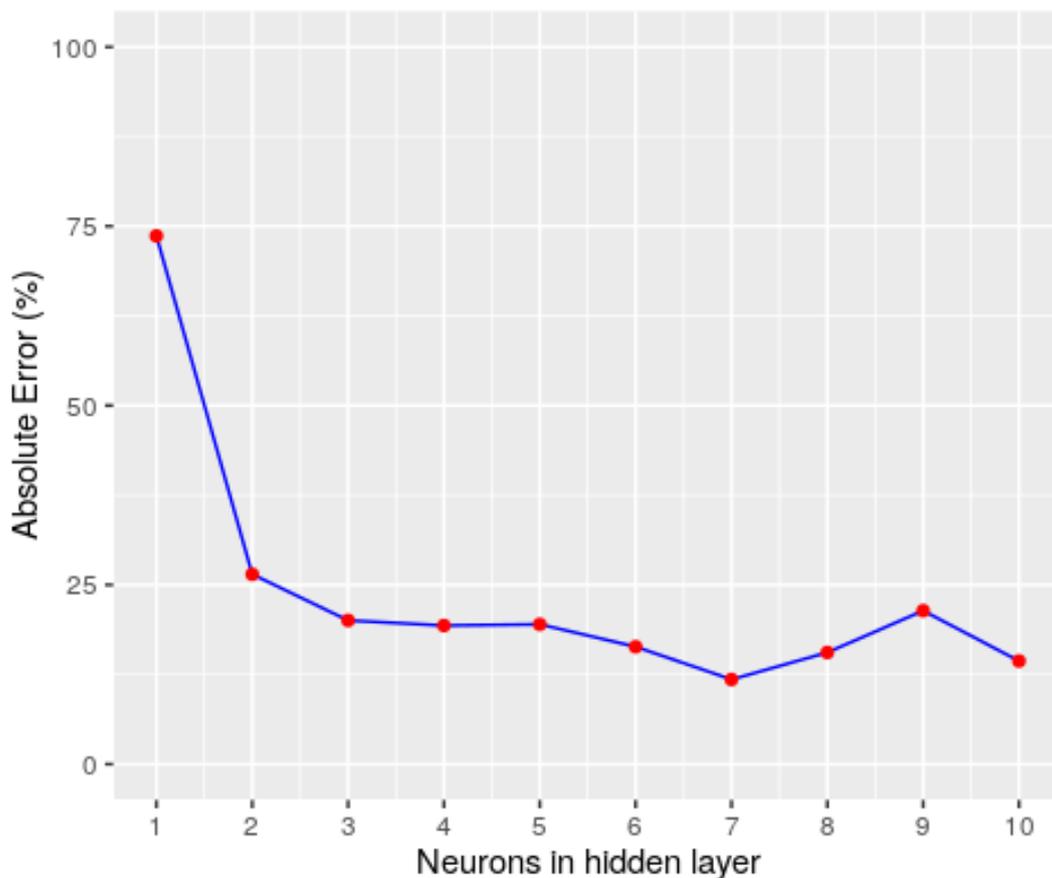
Por último, se ha elegido como función de activación la sigmoideal  $y(x) = \tanh(x)$ .

Para ello se ha utilizado la clase MultilayerPerceptron de Weka, que permite el uso de una interfaz gráfica para el diseño de la red.

## 3. Parametrización

Como ya se ha destacado, la mayor parte de los parámetros de la red se han decidido en la fase de implementación. Sin embargo, el número de neuronas de la capa oculta queda por determinar. Tradicionalmente se suelen escoger valores de aproximadamente la raíz cuadrada del número de atributos, por lo que se ha entrenado la red con valores de entre 1 y 10 para las neuronas de la capa oculta, midiendo el error cometido en cada caso.

Figura 3.3: Porcentaje de error en redes neuronales



En la figura 3.3 se puede observar un mínimo con 7 neuronas. Sin embargo, este mínimo corresponde al conjunto de entrenamiento y conlleva un claro sobreajuste. Por tanto, finalmente se ha decidido utilizar 6 neuronas en la capa intermedia, ya que el error se estabiliza a partir de este número.

## Método de los escenarios

Para implementar el método de los escenarios en el contexto de los algoritmos de predicción empleados, se ha decidido hacer uso de los valores discretos del GILRHO y continuos en las predicciones.

De esta forma, al terminar cada predicción se ajustan los valores del GILRHO según el escenario escogido. Para el normal, únicamente se comprueba que los valores no exceden el máximo de ocho puntos, y en caso de que así sea se reducen a este valor. En el escenario optimista, el GILRHO se redondea únicamente al alza; y el pesimista se define de forma análoga, redondeando el valor tan sólo en el caso de que sea a la baja.

### 3.2.2. Algoritmos de clusterización

Los algoritmos de clusterización son una de las técnicas más comunes de análisis de datos dentro del aprendizaje automático. En este trabajo se ha utilizado para estudiar las similitudes en la evolución de los derechos de los homosexuales entre los distintos países del mundo.

Concretamente, se ha empleado el clustering jerárquico, que permite generar dendogramas que reflejen las relaciones a distintos niveles de las series temporales estudiadas. Para ello, se ha empleado la base de datos del GILRHO hasta el año 2016, que es el último con datos reales.

Además, entre los algoritmos de clustering es común el uso de una distancia para medir estas similitudes; pero para el caso de las series temporales, una distancia simple como la euclídea o la distancia de Manhattan no se ajusta a la realidad: dos países pueden tener tendencias inversas (ir hacia la criminalización o ir hacia la protección de los homosexuales) pero su distancia euclídea podría ser menor que la de dos países con igual tendencia y mayor diferencia de puntos en la escala GILRHO.

Por esto, se ha optado por emplear *dynamic time warping* (DTW) como métrica, un algoritmo empleado ampliamente en el reconocimiento del habla y del hablante, ya que permite detectar similitudes entre series temporales incluso cuando una de ellas se presente acelerada o desacelerada frente a la otra.

### Dynamic time warping

*Dynamic time warping* (DTW) es un algoritmo empleado en el análisis de series temporales para medir la similitud entre dos secuencias temporales que pueden variar en velocidad.

Matemáticamente, el DTW no es una distancia, ya que no cumple la desigualdad triangular [37]; sino que calcula el emparejamiento óptimo entre dos series temporales. Las series se distorsionan de manera no lineal en la escala temporal para obtener la similitud entre ellas con independencia de distorsiones temporales no lineales.

El algoritmo DTW entre dos series  $s$  y  $t$  es el siguiente, donde  $d(x, y) = |x - y|$ :

```
int DTWDistance(s: array [1..n], t: array [1..m]) {
  DTW := array [0..n, 0..m]

  for i := 1 to n
    DTW[i, 0] := infinity
  for i := 1 to m
    DTW[0, i] := infinity
  DTW[0, 0] := 0
```

```

for i := 1 to n
  for j := 1 to m
    cost := d(s[i], t[j])
    DTW[i, j] := cost + minimum(DTW[i-1, j ],    // insertion
                                DTW[i , j-1],    // deletion
                                DTW[i-1, j-1])    // match

return DTW[n, m]
}

```

Para la implementación en R del algoritmo DTW se ha empleado el paquete de R `dtw`, que se integra en el método `dist` para facilitar su uso en algoritmos de clustering.

### Clustering jerárquico (UPGMA)

Para el clustering jerárquico se ha decidido emplear el método *Unweighted Pair Group Method with Arithmetic Mean* (UPGMA). Se trata de un método aglomerativo, ya que comienza con cada uno de los datos en su propio cluster y progresivamente los va emparejando según se avanza en el nivel jerárquico.

Se ha optado por este método aglomerativo dada su relativa rapidez computacional comparado con los métodos divisivos, ya que su complejidad es de  $O(n^2 \log(n))$  [38] frente a  $O(2^n)$ , propia de los segundos.

El método UPGMA, atribuido a Sokal y Michener [39], construye un dendograma a partir de una matriz de similitud que, en nuestro caso, se ha construido con el algoritmo DTW.

Al comienzo del algoritmo, cada dato representa su propio cluster. En cada iteración consecutiva del algoritmo, se unen los dos clústers más cercanos en uno nuevo. Para computar la distancia entre dos clústers  $A$  y  $B$ , con cardinalidades  $|A|$  y  $|B|$  respectivamente, se toma la media aritmética entre las distancias  $d(x, y) = DTW(x, y)$  de las muestras  $x \in A$  e  $y \in B$ :

$$D(A, B) = \frac{1}{|A| \cdot |B|} \sum_{x \in A} \sum_{y \in B} d(x, y)$$

Para la implementación en R del clustering jerárquico se ha utilizado el paquete `hclust`; que, además de UPGMA, ofrece implementaciones para los algoritmos WPGMA, WPGMC y UPGMC. Además, este paquete devuelve los datos para su representación en forma de dendograma a través de las herramientas estándares de visualización de R.

# 4

## Estudio de la igualdad de derechos de los homosexuales

Esta sección recoge el análisis de la igualdad de derechos de los homosexuales desde tres perspectivas distintas: en primer lugar, se estudia la serie temporal de la evolución de los derechos en el mundo, ofreciendo una perspectiva histórica así como un análisis de las predicciones obtenidas a través de los algoritmos descritos en la sección anterior.

En segundo lugar se aplica el método de los escenarios a los algoritmos de predicción, para ofrecer una visión más rica acerca de las posibilidades que depara el futuro a los derechos de los homosexuales en el mundo.

Finalmente, se estudian los resultados de la clusterización de las series temporales para distintos números de agrupaciones, obteniendo así una nueva perspectiva en el análisis de dichas series, esta vez en términos de países pioneros en la defensa de los derechos de los homosexuales y países seguidores. Además, los resultados de la clusterización aportan una visión clara de las zonas del mundo más problemáticas en este aspecto.

Para facilitar la visualización interactiva de todos estos resultados, se ha desarrollado una aplicación web en Shiny que los represente en mapas mediante la biblioteca leaflet de R (basada a su vez en la biblioteca de JavaScript leaflet.js). En el anexo B se puede ver el análisis por subsistemas de esta aplicación.

### 4.1. Estudio de la serie temporal

---

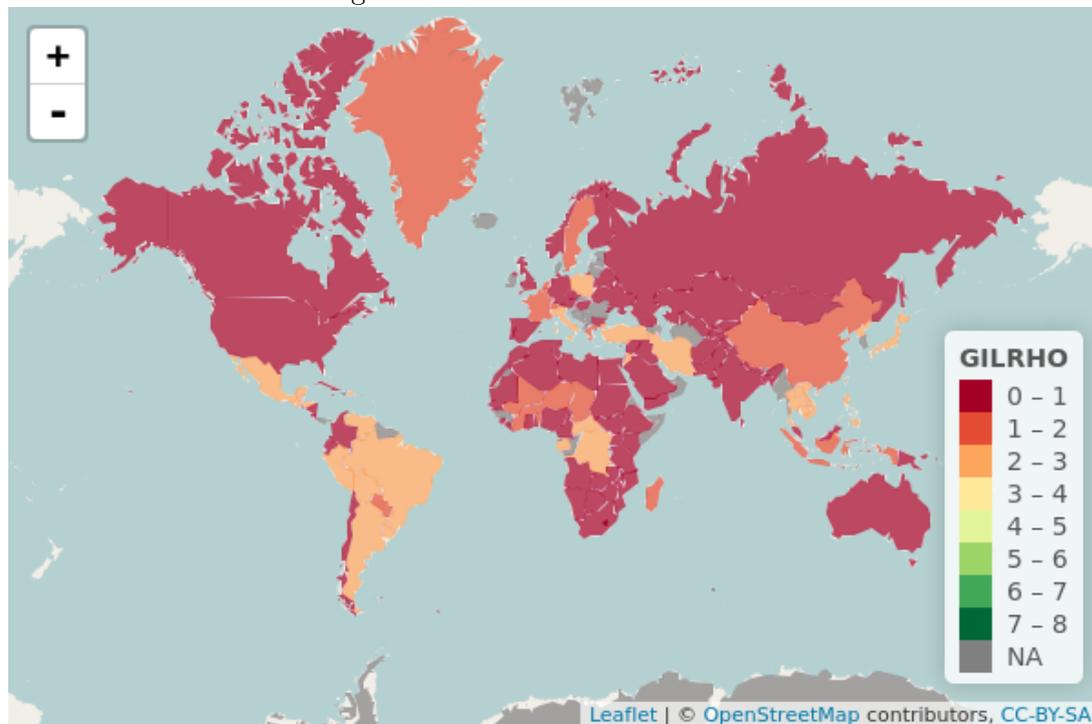
La lucha de los homosexuales para obtener reconocimiento y derechos iguales al resto de personas de la sociedad se remonta a los primeros años del siglo XX. Antes de este siglo, los únicos países en decriminalizar la homosexualidad (sin tener en cuenta los que nunca la han criminalizado en su código penal) fueron Francia (1791), Holanda (1811), Indonesia (1811), Brasil (1830), Guatemala (1871), México (1871) y Argentina (1886). Sin embargo, este estudio comienza, debido a la base de datos empleada, en el año 1960, en pleno auge de la decriminalización.

En primer lugar se estudia la evolución del GILRHO en el mundo desde 1960 hasta el año 2016 mediante la aplicación web desarrollada en Shiny; para continuar analizando las predicciones obtenidas mediante los distintos algoritmos empleados.

### 4.1.1. Perspectiva histórica

La situación de las personas homosexuales en el año 1960 refleja el inicio de la lucha por su descriminalización. Como se puede ver en la figura 4.1, la mayor parte del mundo estaba caracterizada por la criminalización de las relaciones homosexuales.

Figura 4.1: GILRHO en el año 1960



Algunos países africanos, como la República del Congo o Mali, tienen valores de GILRHO superiores a la media, pero como se verá más adelante, la no criminalización no es en estos casos sinónimo de aceptación sino de invisibilización, y en la realidad las relaciones homosexuales solo se dan en la clandestinidad.

Por otra parte, en el mapa se reflejan los primeros esfuerzos de los homosexuales por vivir libremente en la sociedad, con la descriminalización de la homosexualidad en Dinamarca (1933), Uruguay (1934), Suiza (1942), Suecia (1944), Grecia (1951) y Tailandia (1956).

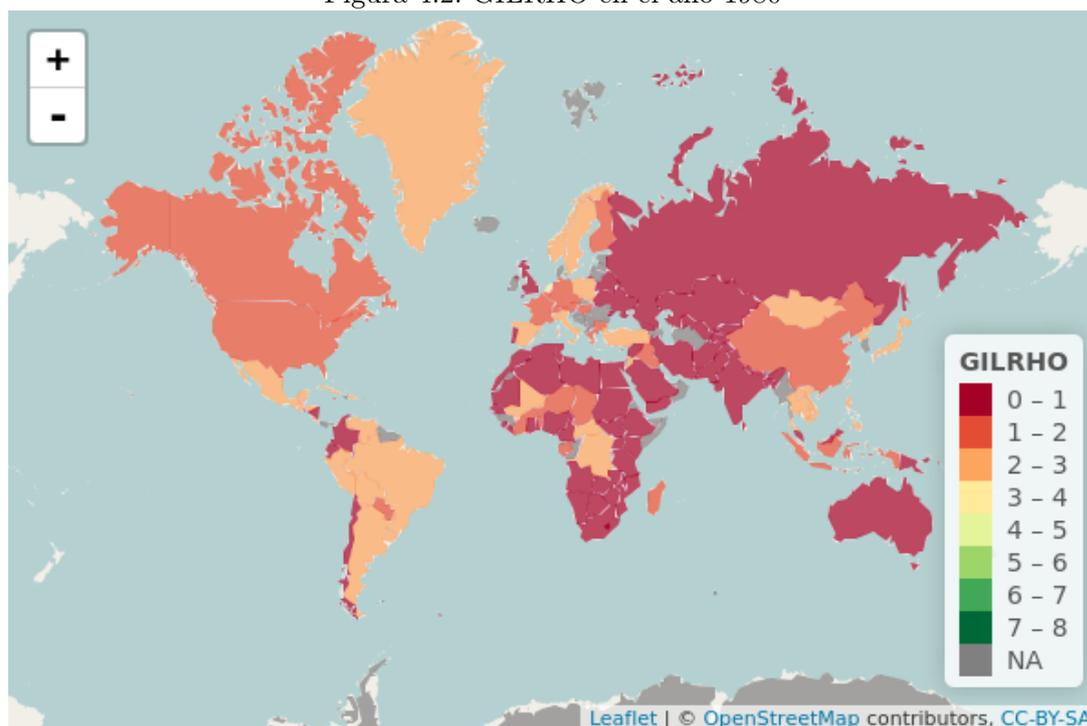
La década de los 60 está marcada por la aparición de las primeras revistas con temática homosexual y las manifestaciones públicas favorables a su descriminalización, así como el surgimiento de las primeras asociaciones de gays y lesbianas. Sin embargo, serán pocos los países que se sumen a esta tendencia creciente, por lo que el colectivo LGBT comienza a organizarse y a protestar sistemáticamente por el cese de su persecución.

Estas manifestaciones culminan en el año 1969 con los disturbios de Stonewall, que originarán la siguiente ola de descriminalizaciones en todo el mundo en la década siguiente, comenzando con Canadá en este mismo año y terminando en España en el año 1979.

En la figura 4.2 se pone de manifiesto esta situación. Países como Australia (1976), Cuba (1979), Costa Rica (1971) o Noruega (1972) alcanzan su primer punto de GILRHO; mientras que otros, como Groenlandia o los Países Bajos comienzan a situar el límite legal de consentimiento en la misma edad que las relaciones heterosexuales.

A partir de la década de los 80 se empiezan a establecer medidas legales prohibiendo la discriminación por motivos de orientación sexual en el trabajo y servicios básicos como la educación

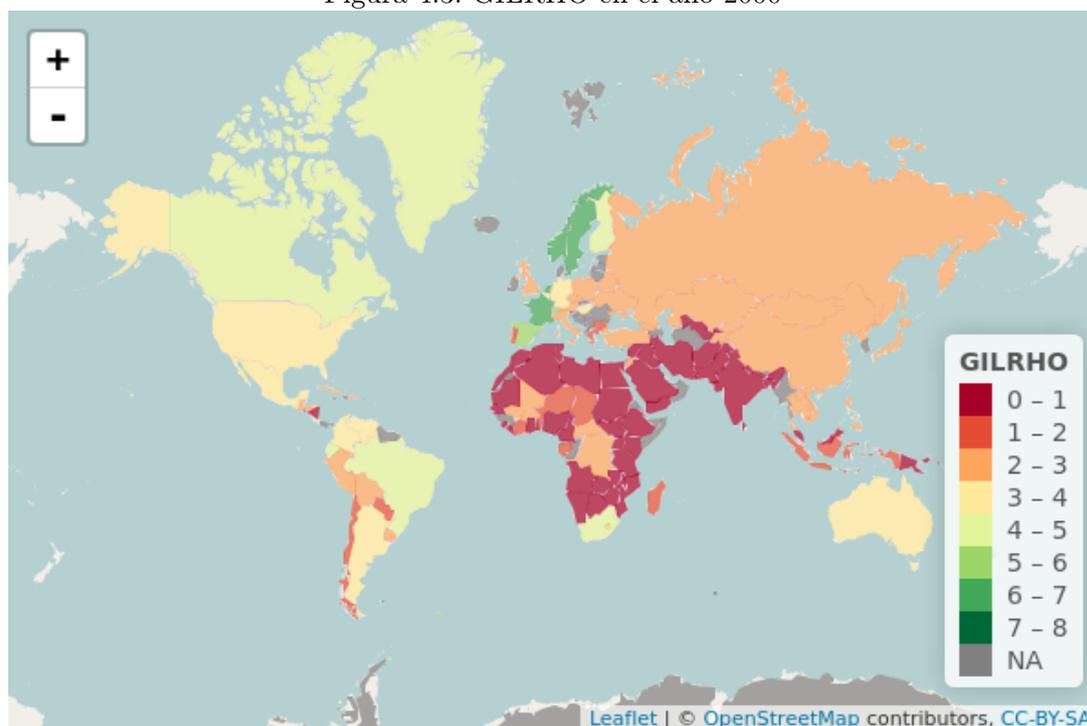
Figura 4.2: GILRHO en el año 1980



o la sanidad. Noruega fue el primer país en hacerlo, en el año 1981, y Francia le siguió años más tarde.

Sin embargo, habrá que esperar hasta los años 90 para que estas leyes se extiendan al resto del mundo. Algunos ejemplos son Sudáfrica (1994), Canadá (1995) o Ecuador (1998).

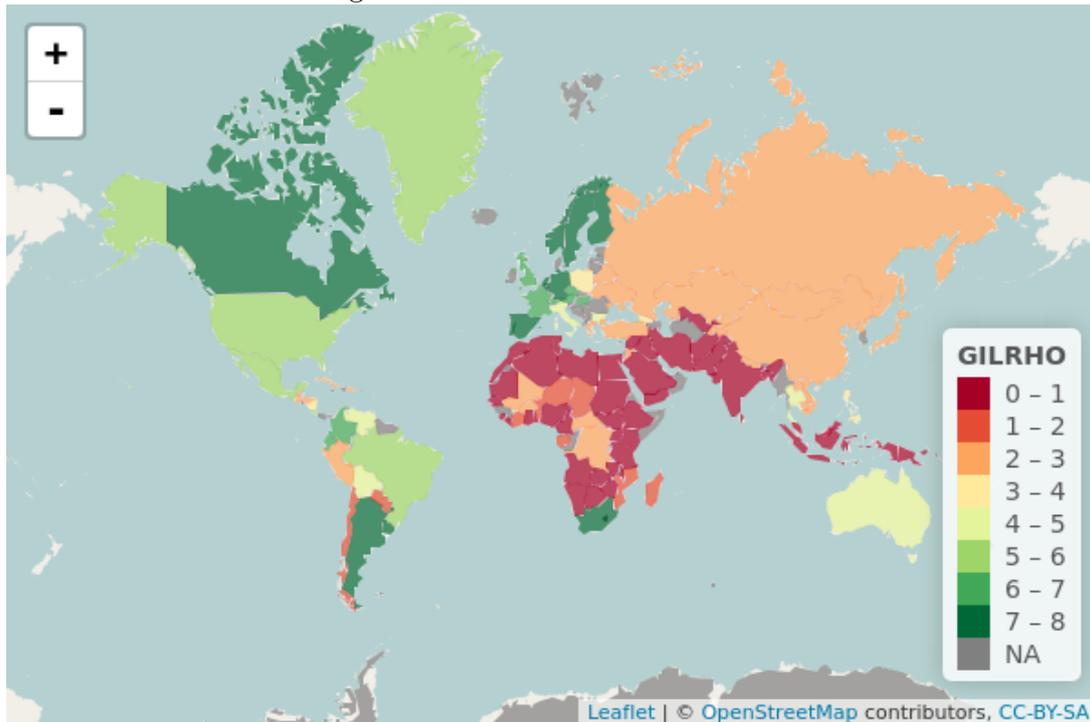
Figura 4.3: GILRHO en el año 2000



Como se observa en la figura 4.3, en el año 2000 las relaciones homosexuales ya no están castigadas por la ley, y tienen la misma edad de consentimiento que las heterosexuales, a excepción de la mayor parte de África y Oriente Medio.

Es en el siglo XXI cuando se empieza a luchar por la unión civil, con derechos iguales al matrimonio heterosexual a excepción de la adopción. Los únicos países que la permitían hasta la fecha eran Dinamarca (1989), Noruega (1993), Suecia (1995), Islandia (1996), los Países Bajos (1998) y Francia (1999). Los Países Bajos serán pioneros en el mundo al legalizar en 2001 el matrimonio entre personas del mismo sexo.

Figura 4.4: GILRHO en el año 2010

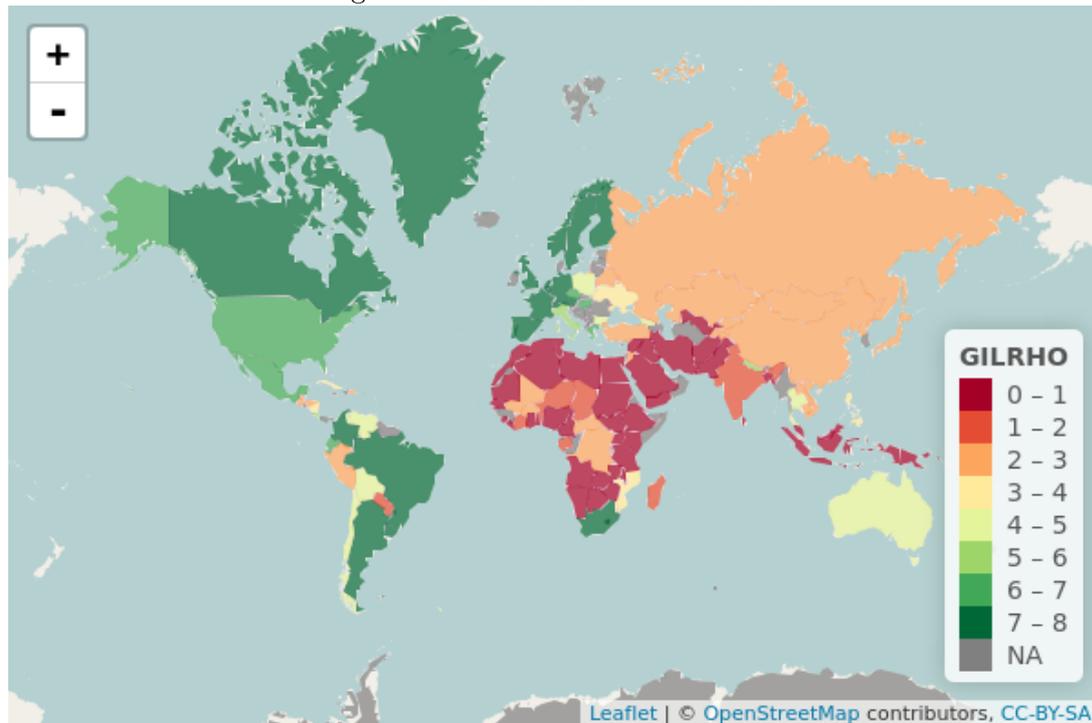


Durante toda la primera década del siglo XXI la unión civil se extiende por el mundo occidental, y los países más avanzados la sustituyen por el matrimonio. Esto da pie a la que será la última gran lucha del colectivo LGBT que se estudia en este trabajo, la de la igualdad de derechos en la adopción. En la figura 4.4 se ve como la situación en el año 2010 ha mejorado notablemente en la mayor parte de Europa del Sur así como en toda América, donde destacan como pioneros Canadá y Argentina.

Para terminar nuestro repaso de la historia, en la figura 4.5 se muestra la situación actual de los derechos de las personas homosexuales en el mundo. Es fácil ver que en la mayor parte de Europa y América del Norte ya se aplican las leyes estudiadas por el GILRHO. También es así en una gran parte de América del Sur, donde se siguen haciendo grandes progresos en la actualidad, el más reciente la legalización del matrimonio homosexual en Colombia (2016).

Sin embargo, la mayor parte de la región africana no presenta signos de modificar su legislación en favor del colectivo LGBT. Lo mismo ocurre en Rusia, donde, a pesar de permitirse la homosexualidad, en 2013 se aprobó la ley contra la propaganda homosexual, que condena la difusión de cualquier información positiva de la homosexualidad con multas y penas de cárcel.

Figura 4.5: GILRHO en el año 2016



#### 4.1.2. Predicción

La figura 4.5 es el punto de partida para la predicción que harán los distintos algoritmos estudiados. En esta sección se describen los distintos enfoques que da cada uno de ellos a la posible evolución futura de los derechos de las personas homosexuales en el mundo.

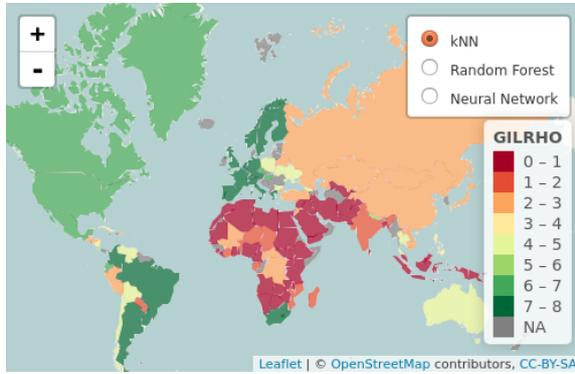
En primer lugar, las figuras 4.6 muestran la predicción de cada uno de los algoritmos para el año 2025. Lo primero que destaca es la similitud entre las tres predicciones, ya que todas suponen un futuro bastante similar a la situación en 2016. Esto tiene sentido, ya que la mayor parte de los países que han mostrado una historia de reconocimiento de los derechos de los homosexuales ya han alcanzado a día de hoy la máxima puntuación que otorga el GILRHO.

Por otra parte, aquellos países que a día de hoy siguen suponiendo una situación precaria para el colectivo LGBT no han dado muestras en su pasado de avanzar positivamente en este sentido. Es por esto que en la predicción del año 2025 no hay demasiadas variaciones entre algoritmos. Sin embargo, sí que se puede apreciar que vecinos próximos es el algoritmo más conservador, no solo evitando predecir mejoras sino incluso prediciendo una ligera regresión en algunas zonas de América.

La selva aleatoria resulta el algoritmo más optimista de los estudiados, prediciendo una clara mejoría en la situación de toda América, sobre todo en los países del Cono Sur, fruto de los avances de los últimos años. Por su parte, la red neuronal es la que más varía su predicción con respecto a la época actual. Por un lado, predice mejoras en América y en toda Europa del Este; pero, por el otro, la situación de toda África y Oriente Medio empeora con respecto a la actual. La red neuronal, por tanto, polariza su predicción del GILRHO, dividiendo el mundo en dos regiones, según si siguen el camino de igualar los derechos de los homosexuales al resto de la población o no.

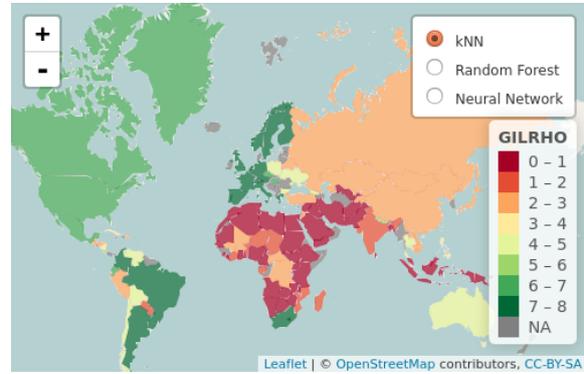
Para terminar esta sección, la figura 4.7 presenta las predicciones de los tres algoritmos para el año 2045. Evidentemente, la capacidad de predicción de los algoritmos decae progresivamente con el tiempo, pero esta predicción evidencia las características ya destacadas de cada uno

Figura 4.6: Predicción del año 2025

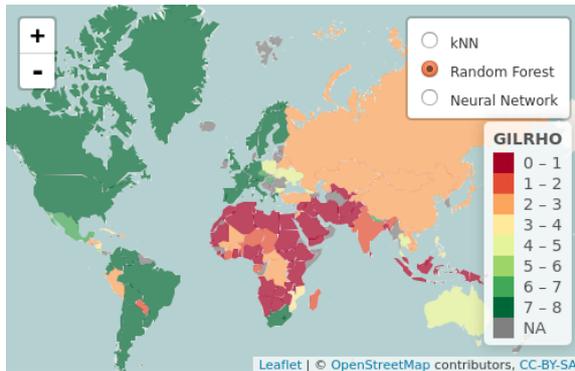


(a) Vecinos próximos

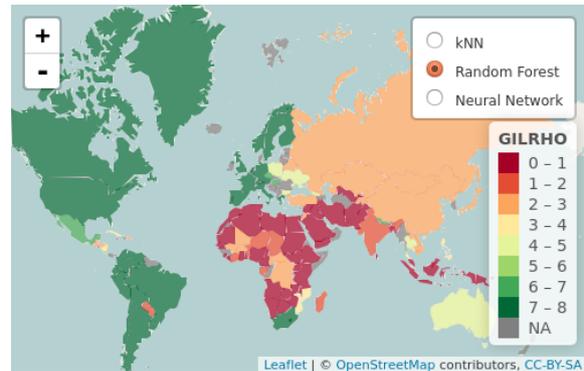
Figura 4.7: Predicción del año 2045



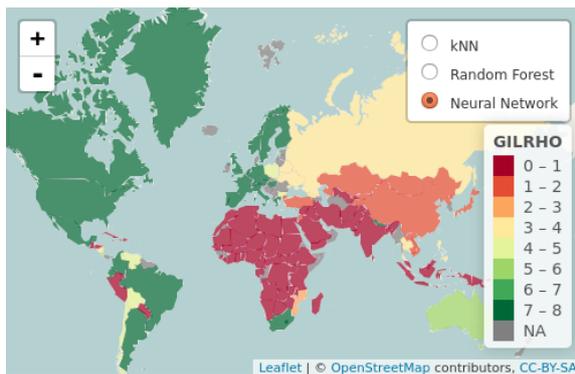
(a) Vecinos próximos



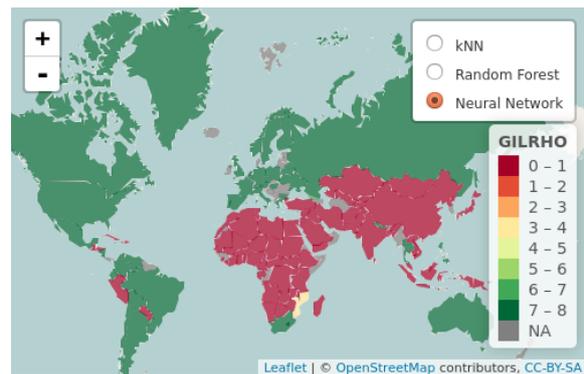
(b) Selva aleatoria



(b) Selva aleatoria



(c) Red neuronal



(c) Red neuronal

de ellos: vecinos próximos presenta un escenario muy similar al que se ha visto en 2025, la selva aleatoria sigue prediciendo mejoras en América Latina y la red neuronal continúa en su polarización, llegando a una situación en la que Europa, América, Groenlandia y Australia alcanzan la máxima puntuación del GILRHO, mientras que el resto del mundo mantiene o empeora su situación hasta llegar a la mínima puntuación.

## 4.2. Estudio de escenarios

---

Para el estudio de escenarios se han definido, además de la predicción normal analizada en la sección anterior, dos escenarios adicionales: uno optimista y otro pesimista.

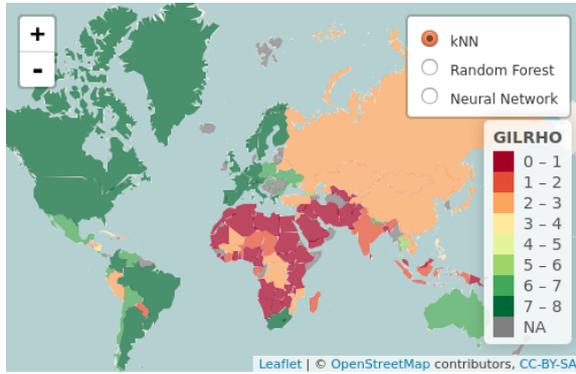
El escenario optimista otorga una probabilidad de 0.5 a la posibilidad de alcanzar en el año consecutivo el punto siguiente, a los países con un punto de GILRHO consolidado. No se tiene en cuenta para ello la situación geográfica de la muestra en cuestión, para evitar introducir información que está representada ya en la base de datos.

La figura 4.8 muestra las predicciones optimistas de cada uno de los algoritmos para el año 2025, con el fin de compararlas con las analizadas en 4.6. El mapa resultante tiene dos características diferenciales con respecto al escenario normal. Por un lado, los países con tendencias al alza en la igualdad de derechos, como los pertenecientes a Europa del Este, progresan en su línea temporal más rápidamente y hacia puntos de GILRHO mayores. Por el otro, aquellos países que no progresan en la igualdad se mantienen en su misma posición, en lugar de mostrar una tendencia regresiva. No obstante, es destacable que ninguno de los países estudiados sufre un cambio de tendencia al aplicar el escenario optimista: la línea temporal presente en estos países es tan regresiva que ni siquiera modificaciones al alza en un año concreto supondrían un cambio radical en dicha tendencia.

En cuanto al escenario pesimista, presentado en la figura 4.9, sí que se observan cambios de tendencia significativos en varios países estudiados. En los países americanos, sobre todo en los del sur, la obtención de derechos por las personas homosexuales es muy reciente, por lo que una pequeña variación anual en alguno de los derechos estudiados podría suponer el inicio de una nueva tendencia regresiva. El mismo comportamiento se pone de manifiesto en Europa del Este, África y Oriente Medio, donde todos los algoritmos coinciden en otorgar la puntuación más baja posible al aplicar el escenario pesimista.

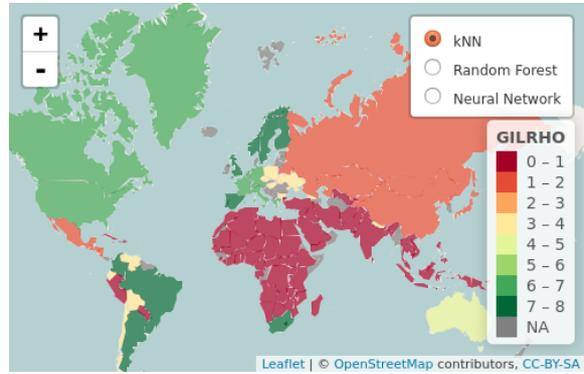
El caso de Rusia, quizá el más significativo debido a la reciente introducción de la ley anti-propaganda, deja claro que en pocos años, si insiste en su tendencia regresiva, podría situarse en muy pocos años junto a los países más atrasados del mundo en materia de derechos LGBT.

Figura 4.8: Escenario optimista del año 2025

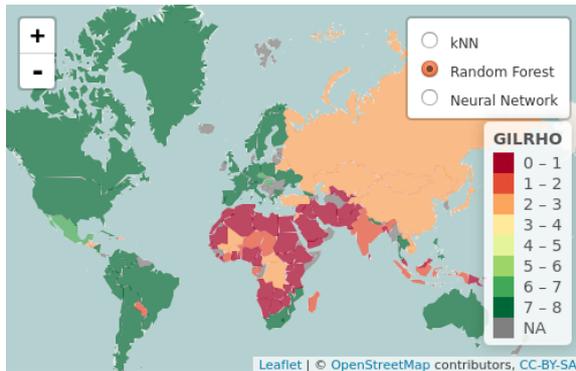


(a) Vecinos próximos

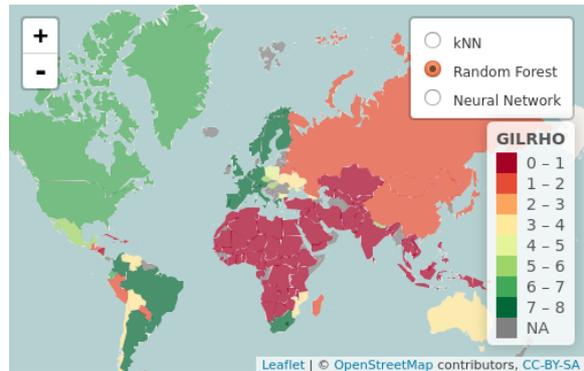
Figura 4.9: Escenario pesimista del año 2025



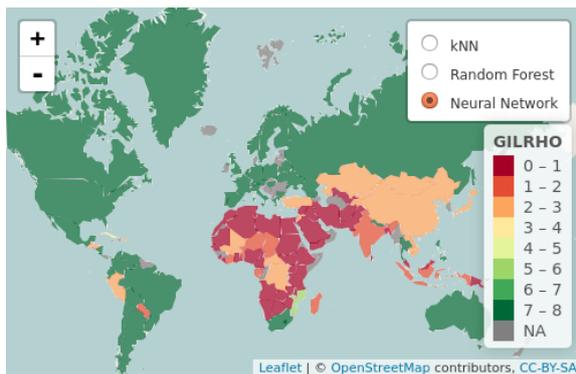
(a) Vecinos próximos



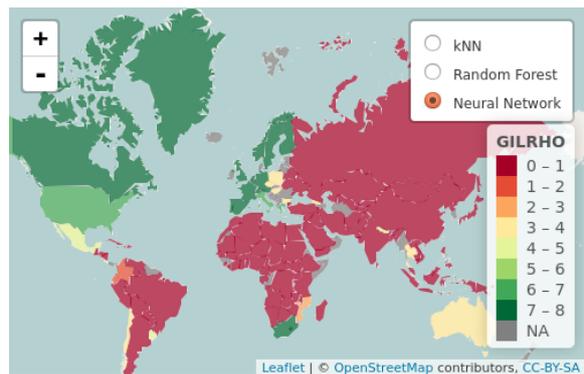
(b) Selva aleatoria



(b) Selva aleatoria



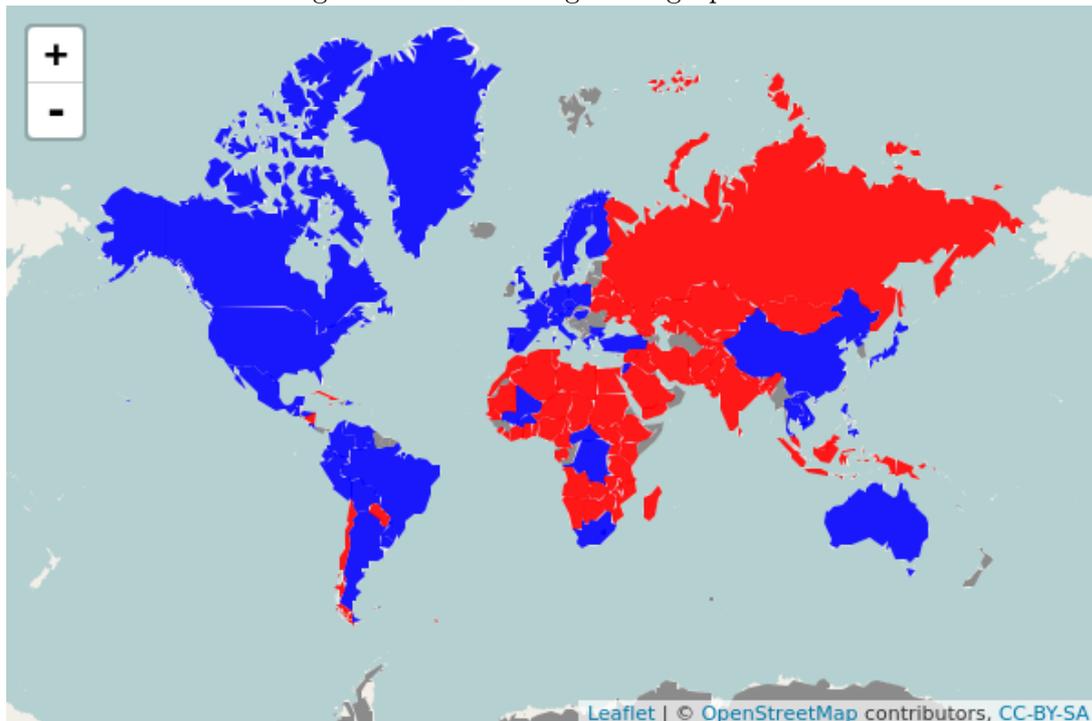
(c) Red neuronal



(c) Red neuronal

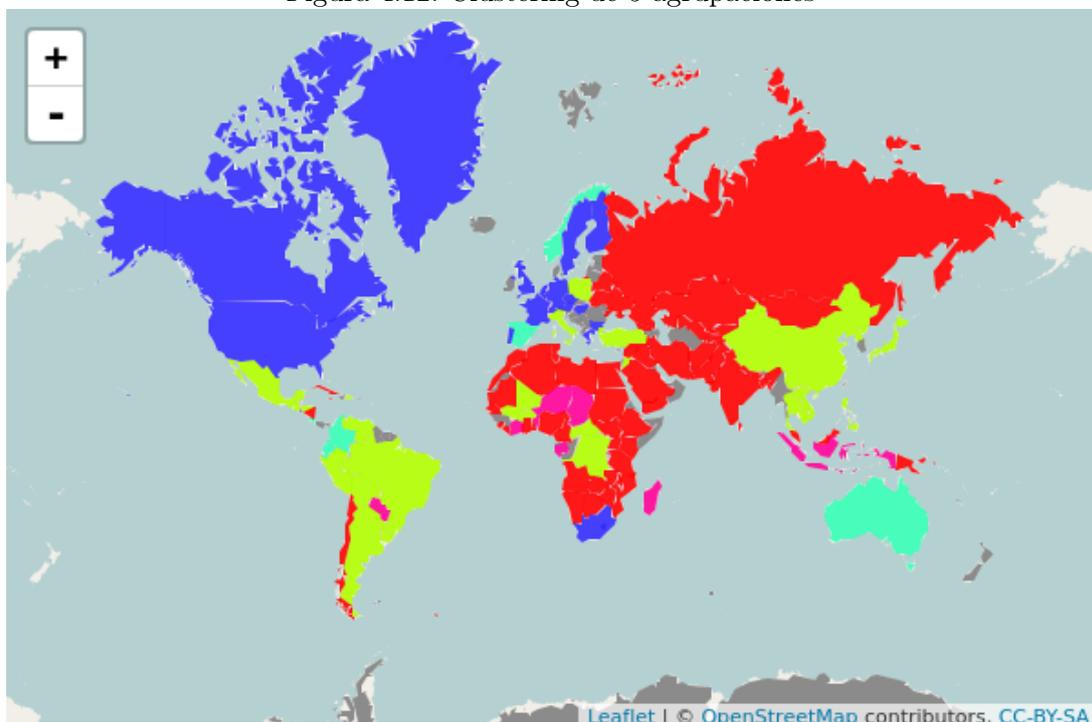


Figura 4.11: Clustering de 2 agrupaciones



El análisis de países pioneros y seguidores requiere que se aumente el número de agrupaciones para dividir la región azul en distintos clústeres. En la figura 4.12 se ha seleccionado un total de 5 de ellas.

Figura 4.12: Clustering de 5 agrupaciones



Para identificar el grupo de países pioneros se ha empleado la tabla 4.1. Se puede comprobar que todos los países identificados históricamente como pioneros en la lucha de los derechos LGBT

aparecen en el mapa dentro de la misma agrupación, en color azul (oscuro y claro). Dentro de Europa, prácticamente la totalidad de los países están en este grupo pionero. Italia supone una excepción, aun no siendo su situación preocupante queda claro que es el país europeo más rezagado en la aplicación de las leyes estudiadas en el GILRHO.

Dentro de este grupo de pioneros mundiales se pueden distinguir, marcados en color celeste, los países que son pioneros dentro de su región: España y Noruega en Europa, Colombia en América del Sur y Australia que, pese a no destacar mundialmente por su puntuación en el GILRHO, es el país más avanzado en materia LGBT de Oceanía.

En cuanto a África, esta nueva división aporta poca información nueva a la ya estudiada. Cabe destacar que Sudáfrica se posiciona en el grupo de pioneros mundiales en derechos LGBT, pero ningún país de su entorno parece adoptar medidas similares.

Finalmente, se puede identificar a Colombia como el país pionero en América Latina en lo que respecta a la igualdad del colectivo LGBT. No obstante, prácticamente del resto de países latinoamericanos aparece en el grupo de seguidores (de color verde en el mapa), lo que da una perspectiva muy optimista en cuanto al futuro de la región.



# 5

## Conclusiones y trabajo futuro

### 5.1. Conclusiones

---

#### 5.1.1. Predicción de series temporales con aprendizaje automático

Los algoritmos de aprendizaje automático empleados actualmente ya han demostrado estar a la altura de las técnicas clásicas de predicción de series temporales, como ARMA y ARIMA, a la hora de predecir la evolución de series en campos tan variados como la meteorología, la política, la estadística y la economía.

Este trabajo ahonda en las posibilidades que ofrecen tres de ellos: vecinos próximos, selvas aleatorias y redes neuronales. Los tres algoritmos se revelan no solo como sustitutos de las técnicas clásicas, sino como mejoras, ya que permiten trabajar con grandes cantidades de datos ofreciendo una eficiencia mucho mayor, y eliminando el estudio previo y descomposición de la serie temporal, necesarios en los modelos clásicos.

Por otra parte, los algoritmos de clusterización permiten un análisis de las series temporales que no está presente en las técnicas clásicas. Mediante el cálculo de la pseudodistancia DTW, en combinación con técnicas de clusterización propias del aprendizaje automático, las series temporales pueden compararse agrupándolas según su similitud, para facilitar el estudio cualitativo global en problemas con series similares pero distintas características geotemporales.

Con el auge del *big data*, estos algoritmos se han implementado en librerías para prácticamente la totalidad de los lenguajes actuales de programación orientados a estadística. Esto facilita notablemente el estudio de series temporales, ya que la recogida y análisis de datos, la programación de los algoritmos y la visualización final de las predicciones puede llevarse a cabo sin cambiar de entorno de programación.

Finalmente, este Trabajo de Fin de Grado demuestra que es posible realizar este tipo de estudios utilizando enteramente software de código abierto, y cada vez más, sin dejar de utilizar software libre. Este hecho supone que cualquier persona con acceso a un ordenador e internet puede llevar a cabo análisis de series temporales, estudiar el código subyacente, mejorarlo y compartirlo de nuevo con la comunidad.

### **5.1.2. Los derechos de los homosexuales en las próximas décadas**

La historia de la lucha de las personas homosexuales por tener iguales derechos al resto de la sociedad es relativamente reciente, comenzando a principios del siglo XX y llegando, en nuestro siglo, a conseguir la mayor parte de los objetivos legales que se había planteado en occidente.

Sin embargo, esta rápida expansión tiene, a su vez, consecuencias negativas: los derechos conseguidos no están asentados en la sociedad, pues una gran cantidad de países tiene una opinión mayoritaria incluso más retrógrada que sus estructuras jurídicas. En España, el país con mejor opinión de la homosexualidad en todo el mundo [6], el año 2016 fue en el que se produjo el mayor número de agresiones homófobas desde que se tiene registro.

De hecho, una posición política contraria a los derechos de los homosexuales supone, a la par, una regresión en la sociedad en cuanto a la aceptación de estas personas. Sobran los ejemplos en el último siglo en los que gobiernos, generalmente autoritarios, han atacado al colectivo LGBT y directamente a las personas que lo conforman o defienden.

La Alemania Nazi, por ejemplo, se caracterizó por la represión de los homosexuales y una vuelta a su clandestinidad. En España, pionera al abolir las leyes contra la sodomía en 1822, la dictadura franquista supuso una regresión brutal, política y de opinión pública, simbolizada con la muerte, en 1936, de Federico García Lorca, uno de los poetas más grandes de nuestra historia, por su condición sexual (y también, claro está, por sus ideas de izquierdas).

Queda fuera de los objetivos de este trabajo predecir el ascenso de fuerzas políticas homófobas o fascistas que supondrían, previsiblemente, la vuelta a la criminalización de la homosexualidad. Pero este no es el único reto al que se enfrenta el colectivo LGBT en nuestra época.

Este Trabajo de Fin de Grado deja claro, a través de las predicciones, del método de los escenarios y de la clusterización de los países estudiados, en qué regiones es más problemática la situación de la homosexualidad actualmente y en las próximas décadas. Ninguno de los 73 Estados que actualmente criminalizan la homosexualidad, proyecta predicciones de cambio de esta situación en el futuro. Es evidente que resulta necesario, para el progreso de estas regiones en materia LGBT, que se formen grupos locales de trabajo como los que originaron las primeras protestas en el mundo y que las personalidades más destacadas en dichos países tengan un posicionamiento claro en la materia.

Finalmente, este trabajo también deja espacio para la esperanza. Los últimos años han supuesto el auge de la consecución de derechos por parte de las personas homosexuales en la mayor parte de América Latina, una población tradicionalmente conservadora en temática LGBT pero con una mayor aceptación a medida que estas leyes entran en vigor. Por tanto, queda claro que hay aún mucho trabajo por hacer, pero que este trabajo, sin duda necesario, tiene resultados concretos que mejoran la vida de millones de personas que, en todo el mundo, solo buscan vivirla en igualdad.

## **5.2. Trabajo futuro**

---

La naturaleza cambiante de la vida política, así como la gran cantidad de estimadores susceptibles de ser estudiados, hace que el presente sea un trabajo con numerosos retos en el futuro.

En primer lugar, los algoritmos de aprendizaje automático son numerosos en la actualidad, y constantemente son mejorados y adaptados a retos concretos. Por esto, este trabajo plantea la posibilidad de estudiar los datos con nuevos algoritmos para evaluar su efectividad con respecto a los analizados.

Otro de los aspectos que no se han estudiado, pero que resultan de interés en el campo del aprendizaje automático aplicado a series temporales, es realizar una predicción de las series en el pasado, observando el comportamiento de los algoritmos con datos anteriores a los que se han utilizado en su entrenamiento.

Por otra parte, la base de datos empleada es susceptible de mejorar en dos aspectos fundamentales. Por un lado el aspecto temporal, tanto con datos pasados para incluir los inicios de la lucha LGBT como con datos futuros conforme estos sean recopilados. En el aspecto cualitativo también hay diversas mejoras a realizar: se ha hecho hincapié durante todo este Trabajo de Fin de Grado en que únicamente se han estudiado datos relativos a la homosexualidad, dejando de lado otros colectivos pertenecientes a la comunidad LGBT, como las personas transexuales. La dificultad de recopilar datos legales en este sentido los ha excluido de este trabajo, pero no cabe duda de que es uno de los aspectos más importantes a incluir en un futuro. También hay que destacar que los objetivos del colectivo LGBT van variando conforme se van consiguiendo. Por ello, la base de datos deberá ser actualizada para reflejar las nuevas luchas que se van llevando a cabo en todo el mundo, particularmente en aquellos países que ya obtienen la puntuación más alta en el GILRHO.

Finalmente, se está trabajando en la publicación de los resultados obtenidos en dos aspectos, el académico y el didáctico. En el primero, se está estudiando la publicación de artículos comparando los resultados obtenidos para el GILRHO con los del Trabajo de Fin de Grado de Jesús de los Nietos Valle, *Análisis evolutivo de la igualdad de género y su impacto en los informes de sostenibilidad GRI*.

En el aspecto didáctico, se está buscando colaborar con entidades reconocidas por su lucha en materia de derechos LGBT para difundir, gracias al desarrollo de los mapas del GILRHO, los logros conseguidos y las dificultades a las que se enfrentan las personas homosexuales en una gran parte del mundo.



## Glosario de acrónimos

- **ANN:** Red neuronal artificial (*Artificial Neural Network*)
- **ARCH:** Modelos autorregresivos con heterocedasticidad condicional (*Autoregressive conditional heteroskedasticity*)
- **ARIMA:** Modelos autorregresivos integrados de media móvil (*Autoregressive integrated moving average*)
- **ARMA:** Modelos autorregresivos de media móvil (*Autoregressive moving average*)
- **DTW:** Distorsión temporal dinámica (*Dynamic time warping*)
- **GARCH:** Modelos autorregresivos con heterocedasticidad condicional generalizados (*Generalized autoregressive conditional heteroscedasticity*)
- **GILRHO:** Índice legal del reconocimiento de la orientación homosexual (*Global Index on Legal Recognition of Homosexual Orientation*)
- **GVN:** Global Business Network
- **HDI:** Índice de desarrollo humano (*Human Development Index*)
- **ILGA:** *International Lesbian and Gay Association*
- **k-NN:** k vecinos próximos (*k nearest neighbours*)
- **LGBT:** Lesbianas, gays, bisexuales y personas transgénero
- **ONU:** Organización de las Naciones Unidas
- **SVM:** Máquina de soporte vectorial (*Support vector machine*)
- **UPGMA:** Método de agrupación sin pesos con media aritmética (*Unweighted Pair Group Method with Arithmetic Mean*)
- **Weka:** Entorno para análisis del conocimiento de la Universidad de Waikato (*Waikato Environment for Knowledge Analysis*)



# Bibliografía

- [1] Fundación Triángulo. Exterminio bajo el nazismo. *Orientaciones. Revista de Homosexualidades*, (5), 1 2003.
- [2] Aengus Carroll. *State-Sponsored Homophobia 2016: A world survey of sexual orientation laws: criminalisation, protection and recognition*. International Lesbian, Gay, Bisexual, Trans and Intersex Association, 11 edition, 5 2016.
- [3] María Luisa Femenias, M. Calderón, and R. Osborne. Mujer, sexo y poder: Aspectos del debate feminista en torno a la sexualidad. *Reis*, (51):226, 1990.
- [4] Antonio Martelli. From scenario building to scenario planning: Intuitive logic and trend impact analysis. *Models of Scenario Building and Planning*, pages 124–156, 2014.
- [5] GNU. Gnu emacs - gnu project. <https://www.gnu.org/software/emacs/>, 2013. (Accessed on 01/13/2017).
- [6] PewResearch. The global divide on homosexuality. *PewResearchCenter*, 6 2013.
- [7] Kees Waaldijk. Legal recognition of homosexual orientation in the countries of the world. a chronological overview with footnotes. Technical report, The Williams Institute (at UCLA), 2009.
- [8] M.V. Lee Badgett, S. Nezhad, K. Waaldijk, and Y. van der Meulen Rodgers. The relationship between lgbt inclusion and economic development: An analysis of emerging economies. Technical report, The Williams Institute, 11 2014.
- [9] G. Udny Yule. On a method of investigating periodicities in disturbed series, with special reference to wolfer’s sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 226:267–298, 1927.
- [10] Gilbert Walker. On periodicity in series of related terms. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 131(818):518–532, 1931.
- [11] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. Time series analysis. Jun 2008.
- [12] L. Esch, R. Kieffer, T. Lopez, C. Berbé, P. Damel, M. Debay, and J. F. Hannosset. *Time Series Models: ARCH-GARCH and EGARCH*, pages 373–373. John Wiley & Sons Ltd, 2005.
- [13] Paul Werbos. Beyond regression: New tools for prediction and analysis in the behavioral sciences. 1974.
- [14] K. R. Muller, A. J. Smola, G. Ratsch, B. Scholkopf, J. Kohlmorgen, and V. Vapnik. Using support vector machines for time series prediction. *Advances in kernel methods-support vector learning*, MIT Press, Cambridge, MA, pages 243–254, 1999.

- [15] J. Friedman, T. Hastie, and R. Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin, 2001.
- [16] Thomas G. Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000.
- [17] Imran Maqsood and Ajith Abraham. Weather analysis using ensemble of connectionist learning paradigms. *Applied Soft Computing*, 7(3):995–1004, 2007.
- [18] H. Moon, H. Ahn, R. L. Kodell, S. Baek, C. Lin, and J. J. Chen. Ensemble methods for classification of patients for personalized medicine with high-dimensional data. *Artificial Intelligence in Medicine*, 41(3):197–207, Nov 2007.
- [19] L. Yu, S. Wang, and K. K. Lai. Forecasting crude oil price with an emd-based neural network ensemble learning paradigm. *Energy Economics*, 30(5):2623–2635, Sep 2008.
- [20] James A Ogilvy. *Creating better futures: Scenario planning as a tool for a better tomorrow*. Oxford University Press, 2002.
- [21] Cas Mudde. *Populist radical right parties in Europe*, volume 22 of 8. Cambridge University Press, 2007.
- [22] Observatorio Madrileño contra la homofobia transfobia y bifobia. Informe sobre delitos de odio por lgtbfobia en la comunidad de madrid. Technical report, Arcópoli, 6 2016.
- [23] Department of Economic and Statistics Division Social Affairs. *Standard Country or Area Codes for Statistical Use*. Number ST/ESA/STAT/SER.M/49/REV.4 in Statistical papers. Series M no.49, rev. 4. United Nations. Statistical Division, 4 edition, 4 1999.
- [24] Stephen R Garner et al. Weka: The waikato environment for knowledge analysis. In *Proceedings of the New Zealand computer science research students conference*, pages 57–64. Citeseer, 1995.
- [25] K. Hornik, C. Buchta, and A. Zeileis. Open-source machine learning: R meets weka. *Computational Statistics*, 24(2):225–232, 2008.
- [26] W. Paul Vogt. Cross validation. *Dictionary of Statistics & Methodology*, 2005.
- [27] B. W. Silverman and M. C. Jones. E. fix and j.l. hodges (1951): An important contribution to nonparametric discriminant analysis and density estimation: Commentary on fix and hodges (1951). *International Statistical Review / Revue Internationale de Statistique*, 57(3):233, Dec 1989.
- [28] D. Aha and D. Kibler. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.
- [29] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.
- [30] Mark R Segal. Machine learning benchmarks and random forest regression. *Center for Bioinformatics & Molecular Biostatistics*, 2004.
- [31] Huang B.F.F. and Boutros P. C. The parameter sensitivity of random forests. *BMC Bioinformatics*, 17(1), Sep 2016.
- [32] Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4):115–133, Dec 1943.

- [33] Sara Moein. Artificial neural network for medical diagnosis. *Medical Diagnosis Using Artificial Neural Networks*, pages 85–94, 2014.
- [34] Suraphan Thawornwong and David Enke. Forecasting stock returns with artificial neural networks. *Neural Networks in Business Forecasting*, pages 47–79, 2003.
- [35] F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408, 1958.
- [36] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, Oct 1986.
- [37] Dmitri Fomin, Sergey Genkin, and Ilia V. Itenberg. The triangle inequality. *Mathematical World*, pages 51–56, Jul 1996.
- [38] Lior Rokach and Oded Maimon. Clustering methods. *Data Mining and Knowledge Discovery Handbook*, pages 321–352, 2005.
- [39] R.R. Sokal, C.D. Michener, and University of Kansas. *A Statistical Method for Evaluating Systematic Relationships*. University of Kansas science bulletin. University of Kansas, 1958.





## Base de datos empleada

Los datos empleados para la predicción y clusterización se han obtenido a partir de dos informes de Kees Waaldijk, *Legal recognition of homosexual orientation in the countries of the world* (en adelante, Waal09) y *The Relationship between LGBT Inclusion and Economic Development: An Analysis of Emerging Economies* (en adelante, Waal14).

La figura A.1 muestra un ejemplo de los datos en Waal09 para la situación de las personas homosexuales en los 15 estados miembros de la Unión Europea más antiguos. Esta tabla contiene, para cada país, el año en el que se alcanzó cada uno de los puntos que estudia el GILRHO.

Estas tablas contienen datos erróneos, como el propio Waaldijk admite, dudosos y en algunos casos directamente no aparecen. De todas formas, se han procesado con el lenguaje AWK para extraerlos a ficheros separados por comas que puedan ser tratados con R.

Una vez obtenidos los datos principales sobre los que se basará el estudio, se ha tratado de igual manera, también con AWK, la información de Waal14. En este segundo informe ya aparece la puntuación GILRHO en lugar del año en que se consiguen ciertos derechos. Así, se han cruzado las puntuaciones resultantes de tratar los datos de Waal09 con las que se dan en Waal14, y cuando ha habido incoherencias entre ambas se ha utilizado la puntuación del segundo, dado que es más reciente. La figura A.2 muestra el formato de los datos de Waal14 según aparecen en el informe.

Finalmente, para aumentar el alcance de este trabajo, se ha incorporado manualmente la información legal disponible en el informe de 2016 de la ILGA, *State-sponsored homophobia, A world survey of sexual orientation laws: criminalisation, protection and recognition*, para incluir más países de los contemplados en los dos informes de Waaldijk.



# B

## Análisis de la aplicación por subsistemas

La aplicación desarrollada se ha programado con tres objetivos fundamentales:

1. Automatizar la extracción de los datos de los informes empleados.
2. Generalizar el entrenamiento y predicción de modelos de aprendizaje automático.
3. Visualizar los datos y los resultados a través de tablas y mapas interactivos.

Para ello, se ha dividido en tres subsistemas, atendiendo cada uno a un objetivo concreto. Los subsistemas en cuestión son los siguientes:

1. Subsistema de datos: Encargado de la extracción de los datos legales de los informes que se emplearán, así como del procesado de los mismos para generar una base de datos con los atributos que se emplearán en este trabajo. Además se encarga de almacenar esta base de datos y proveer a los modelos de los objetos que cada uno de ellos necesita.
2. Subsistema de análisis: Su tarea principal es el entrenamiento de los modelos y la predicción o clusterización de los datos. Para ello se comunica con el subsistema de datos y con las clases de la aplicación Weka, abstrayendo todo el proceso para facilitar la tarea al analista, permitiendo la selección de distintos parámetros y generando gráficas con el error cometido.
3. Subsistema de visualización: Muestra los registros de la base de datos en forma de mapas, para las de predicción y clusterización, y tablas, para dar una visión clara de la evolución de los derechos en el mundo.

La figura B.1 muestra el diagrama de componentes, dividido en subsistemas, con el flujo de los datos dentro de la aplicación.

En la figura B.2 se puede ver una captura de pantalla del módulo de visualización, donde se muestra el mapa interactivo para las predicciones del GILRHO con cada uno de los algoritmos.

Figura B.1: Diagrama por componentes de la aplicación

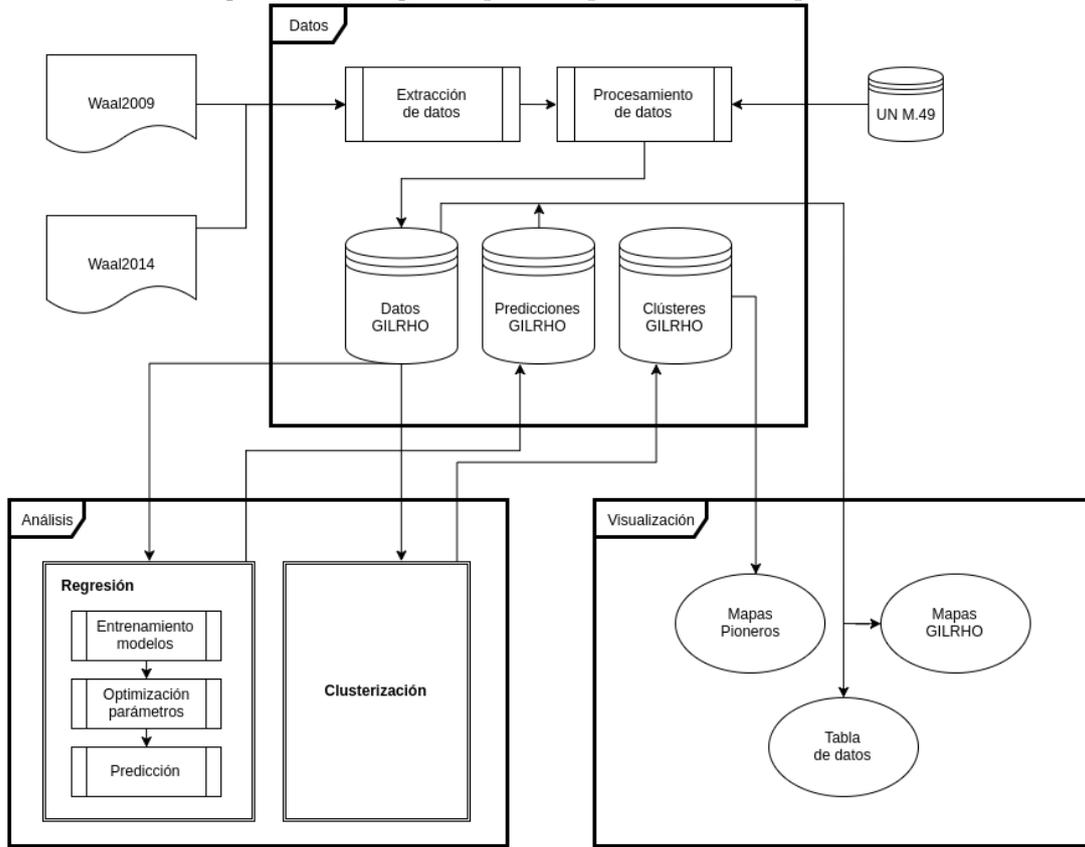


Figura B.2: Aplicación de visualización de los resultados

