

**Análisis mundial del clima**

Carlos Castellanos Mateo

José Manuel Pinto Lozano

Arturo Aguirre Calvo

Índice

C

Conclusiones 8

I

Implementación 6

Introducción 3

M

Metodología 5

O

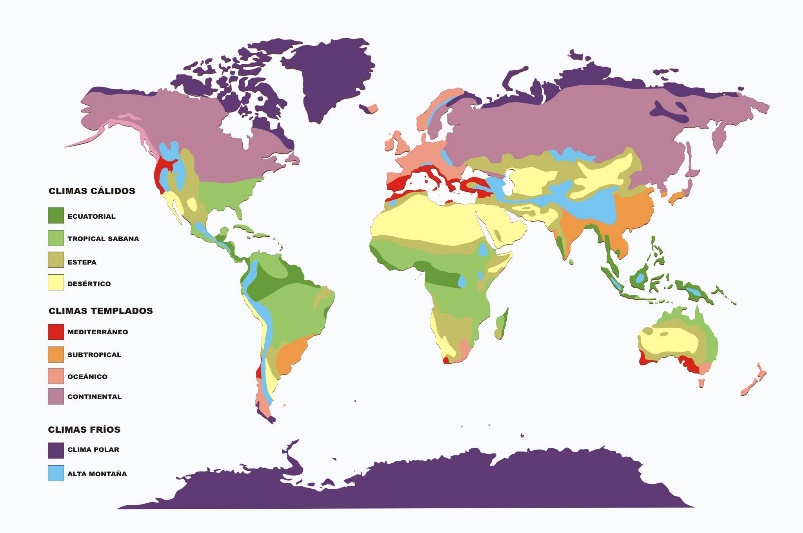
Objetivos 4

R

Resultados y evaluación 7

(índice automático, cuando terminemos el documento le damos a actualizar y calcula las páginas y ya lo ponemos en orden)

Introducción



Clima: conjunto de condiciones atmosféricas que caracterizan una región.

Existe un agrupamiento en tipos de clima conforme a ciertas características de la región como temperatura, lluvia, viento, etc. A su vez dentro de cada tipo se pueden diferenciar matices más específicos. Según la fuente y los parámetros, podemos encontrar distintas agrupaciones, algunas de ellas son:

Según la temperatura.

* Clima cálido
  + Ecuatorial
  + Tropical seco
  + Subtropical árido
  + Desértico
  + Semidesértico
* Clima templado
  + Subtropical húmedo
  + Mediterráneo
  + Oceánico
  + Continental
* Fríos
  + Polar
  + Montaña
  + Tundra

Según las precipitaciones.

* Clima Árido
* Clima Semiárido
* Clima Subhúmedo
* Clima húmedo
* Clima muy húmedo

Además, en determinadas situaciones se pueden crear microclimas, los cuales son climas con características distintas a las que están en la zona en que se encuentra. Podemos diferenciar los siguientes:

* Urbano
* Incendios
* Erupciones

En resumen, las condiciones meteorológicas no son algo exacto y aunque hay varias clasificaciones no se puede determinar con exactitud.

(…)

Objetivos

El objetivo principal de este proyecto es determinar el clima que posee un país en función a unos datos climáticos obtenidos a lo largo de los años. Dado que no se posee una clasificación de los países en los distintos climas, no se puede aplicar aprendizaje supervisado y por ello el objetivo es aplicar un algoritmo de *clustering* que permita determinar cuántos climas diferentes podrían estar representados en los *datasets* y tras esto analizar a que clima pertenece cada *clúster*. Tras esto se podría asignar una etiqueta a cada ejemplo que contenga el clima del *clúster* al que se le ha asignado de forma que ya podría aplicarse aprendizaje supervisado.

Además de esto también se pretende analizar las relaciones del clima entre los distintos países a lo largo de los años.

Metodología

Para la realización de este proyecto se utiliza una metodología SEMMA cuyas fases son detalladas a continuación:

1. **Sample**

Se obtienen datos de dos webs mediante los scripts *descargaClimateKnowledge.r* y *obtencionDatosTuTiempo.r,* los cuales almacenan los datos en archivos .csv para facilitar la lectura en las siguientes fases.

El script *descargaClimateKnowledge.r* obtiene los datos de temperatura y lluvias de diferentes países a partir de los siguientes links:

[**https://climateknowledgeportal.worldbank.org/api/data/get-download-data/historical/tas/1901-2016/**](https://climateknowledgeportal.worldbank.org/api/data/get-download-data/historical/tas/1901-2016/)

[**https://climateknowledgeportal.worldbank.org/api/data/get-download-data/historical/pr/1901-2016/**](https://climateknowledgeportal.worldbank.org/api/data/get-download-data/historical/pr/1901-2016/)

A estos enlaces se les concatena el nombre del país del que se quieren obtener los datos y utilizando el método *getURL(),* se obtiene un *string* que debe transformarse a dataframe. Una vez conseguidos ambos *dataframes* de cada país se unen mediante un inner join y se guarda el *dataframe* resultante en un archivo .csv

El script *obtencionDatosTuTiempo.r* obtiene los datos de la web:

[**https:/en.tutiempo.net/climate/**](https://en.tutiempo.net/climate/)

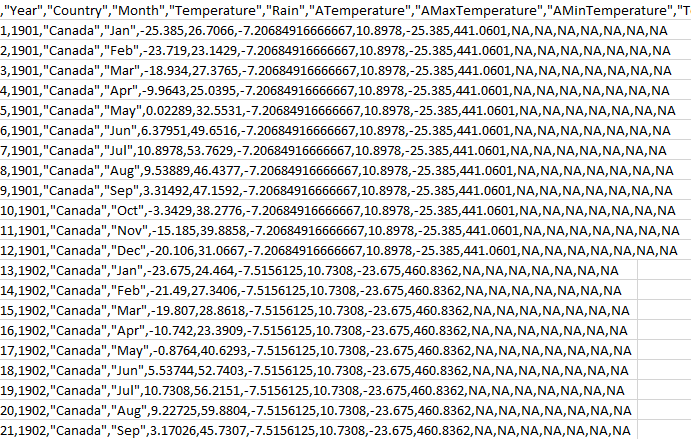
para ello se accede a cada continente y dentro de estos a cada país, y dentro de estos a cada ciudad, donde aparece una tabla con los datos. Esta tabla es cargada mediante el metodo *readHTMLtable(),* tras esto se transforma dicha tabla a *dataframe* y los valores ‘-’ son sustituidos por NA.

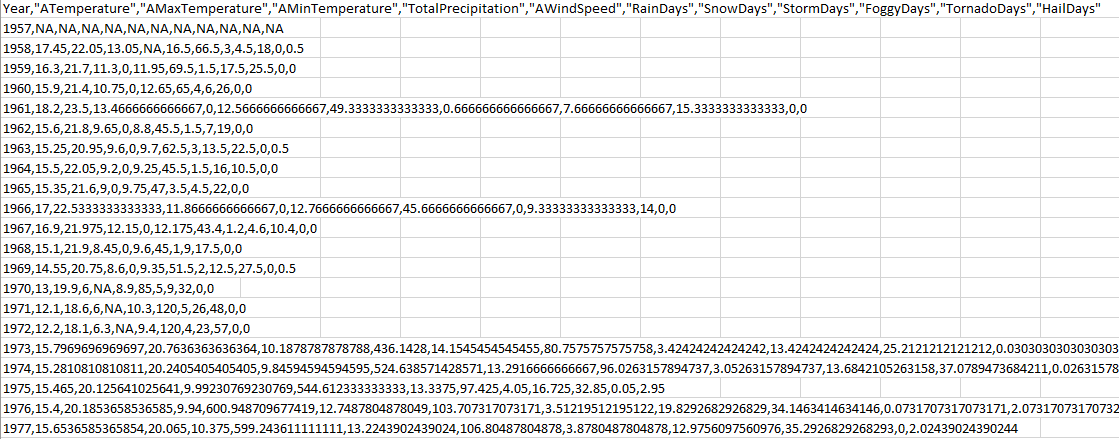
Por último para obtener datos agregados por países, se calculan las medias por año de todas las variables ya que anteriormente se tenía un ejemplo de cada año por ciudad. Tras esto se guardan los datos en un .csv.

1. **Explore**

En esta fase analizamos los datos para entender su estructura y detectar anomalías imprevistas. Además empezamos a pensar cómo abordarlos en el posterior procesamiento.

Por ejemplo este es el *dataset* que obtenemos de las temperaturas de Canadá de una fuente:



Analizamos otra fuente con información más variada de España: temperatura, precipitación, viento, etc.

En el análisis nos damos cuenta de que cuanto más atrás vamos en el tiempo menos datos hay, es decir, más huecos en blanco que no nos aportan información a la hora del procesamiento. Y cuanto más actual es la fecha, más datos y más precisos. Una vez comprendida la estructura de los *datasets* podemos empezar a pensar en el procesamiento de estos.

1. **Modify**

En la parte de modelización nos encargamos de modificar los datos para estén en una forma estandarizada y que el programa pueda leerlos sin problemas. Modificamos los campos en blanco y los ponemos como NA. Si hay algún día o año que no tiene ningún dato, lo descartamos directamente. De esta forma facilitamos el proceso de modelización.

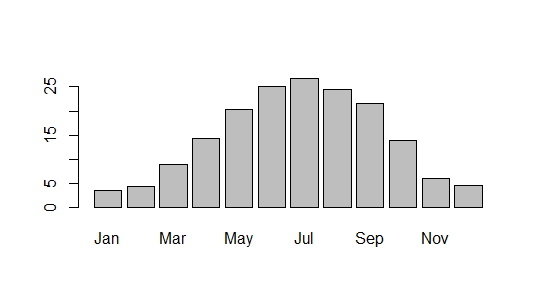
1. **Model**

Aplicamos los modelos para el procesamiento de los datos y nos quedamos con aquel que produce unos resultados más correctos.

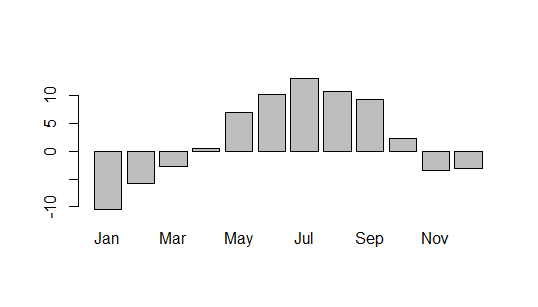
(…)

1. **Assess**

Analizamos los resultados del proceso de Minería de datos. Primero elaboramos unas gráficas para que la información pueda ser interpretada de forma más sencilla. Por ejemplo la temperatura de Afganistán de 2016:



De Noruega del mismo año:



Implementación

Resultados y evaluación

Conclusiones