

## TRABAJO FIN DE MÁSTER

<https://othesoluciones.github.io/TFM/>

Director:  
**Antonio Sarasa Cabezuelo**

30 de Septiembre de 2016

# SISTEMA DE GENERACIÓN DE ALERTAS DE BROTES ALÉRGICOS POR GRAMÍNEAS

Autores:  
**Sara Otero de Navascués Tomé**  
**César Hernández Martín**

MASTER EN BD & BA

# Objetivo

Diseño, desarrollo e implementación de una Aplicación Web dirigida al estudio y predicción de brotes alérgicos producidos por el polen de gramíneas.

El sistema consistirá en una prueba de concepto para una única zona geográfica, la Comunidad de Madrid, implementándose de forma que se pueda ser extensible a cualquier zona y tipo de polen.



# Introducción

¿Qué son las gramíneas y por qué las hemos elegido para nuestro estudio?

Son una familia de plantas del orden de las monocotiledóneas, a la que pertenecen un gran número de especies usados tanto para el consumo humano (caña de azúcar, trigo, arroz, maíz...) como para el animal .

Es una de las familias más numerosas, consta de casi 700 géneros y de unas 12.000 especies. Y se calcula que cubren un 20% de la superficie del mundo.

Se considera al polen de las gramíneas como la principal causa de alergia en Europa.

# Configuración del proyecto

Herramientas utilizadas:

Control de Versiones



Repositorio:

<https://github.com/othsoluciones/TFM>

**Lenguaje de programación**



**Framework**



**PaaS (Platform as a Service)**



**Base de Datos**

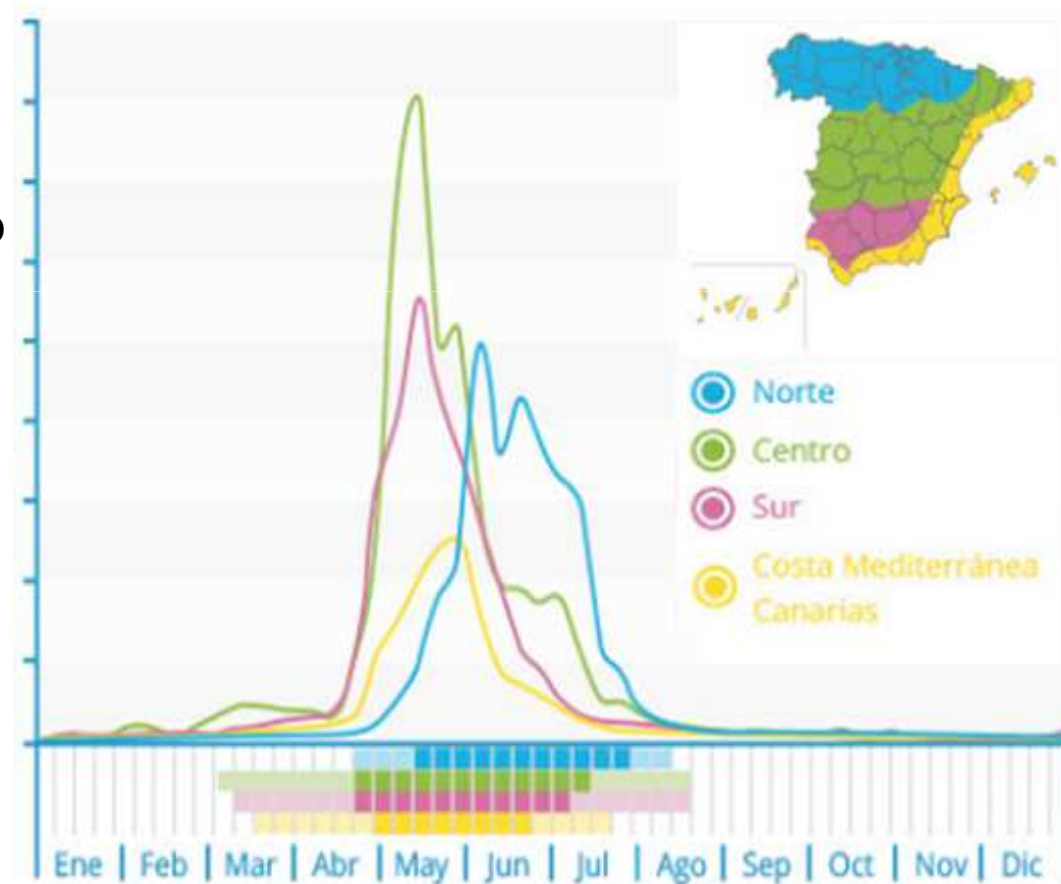


# Modelo

## Calendario Polínico:

Hemos partido del calendario polínico como base para realizar la predicción.

El nivel de polen es un dato estacional



Hemos establecido tres niveles de alerta, asignando a cada uno de ellos un valor numérico .

Y clasificando de forma inicial cada uno de los meses en uno de estos niveles.

El valor inicial se irá incrementando o reduciendo dependiendo del resto de factores incluidos en el algoritmo

MES	NIVEL
Enero	Bajo
Febrero	Bajo
Marzo	Medio
Abril	Alto
Mayo	Alto
Junio	Alto
Julio	Medio
Agosto	Bajo
Septiembre	Bajo
Octubre	Bajo
Noviembre	Bajo
Diciembre	Bajo

## Nivel Medio Diario de Polen:

La *SEAIC (Sociedad Española de Alergología e Inmunología Clínica)* nos proporciona un valor en función del número de  $\text{gr/m}^3$  de polen que hay en el aire para la semana anterior.

Gramos/ $\text{m}^3$ de polen en el aire		
Desde	Hasta	Valor
0	200	0
200,01	1000	1
1000,01	$\infty$	2

De esta forma vamos a ir obteniendo un valor numérico que finalmente traduciremos en el nivel de alerta calculado por nuestro sistema.



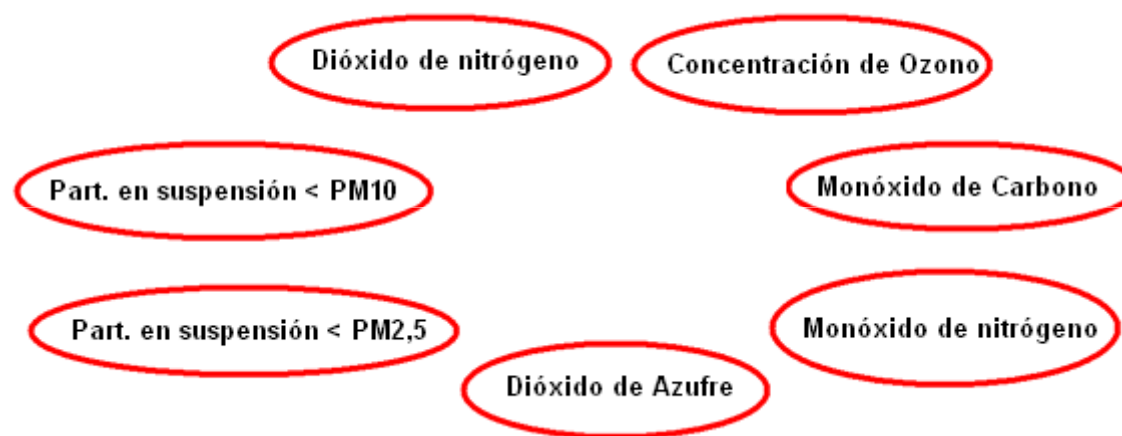
## Calidad del aire:

Varios estudios han demostrado que la polución, sobre todo la producida por las partículas emitidas por los motores de los coches diésel, modifica la naturaleza de los pólenes volviéndolos más agresivos.



Además la polución afecta a las vías respiratorias irritando las mucosas de nariz, faringe y pulmones, lo que **agrava** la sintomatología propia de la **alergia**. Esto explica que haya más alergias en las ciudades que en el campo, aunque los niveles de polen sean inferiores.

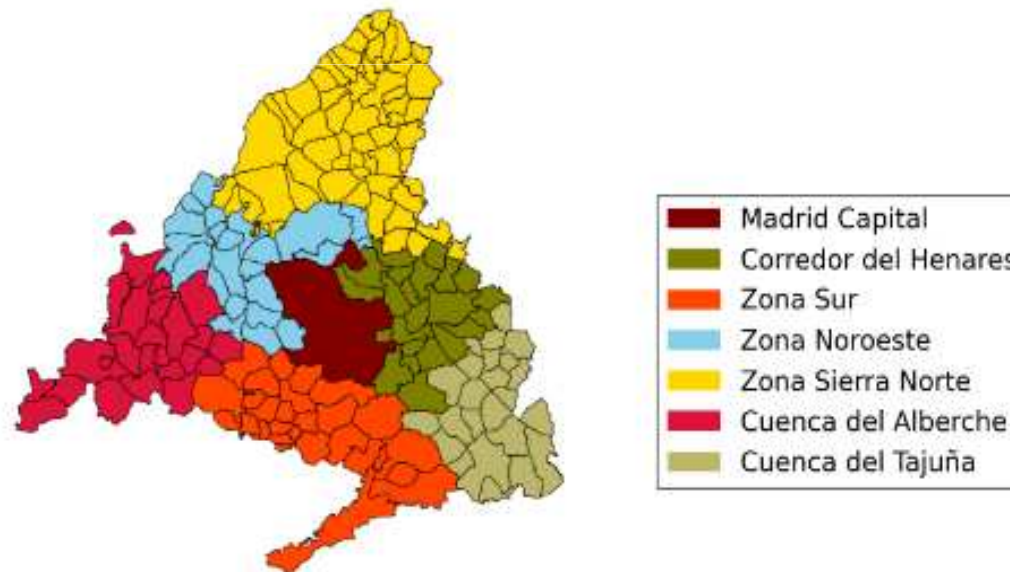
## Contaminantes incluidos en el algoritmo:



Los valores límite que se han tenido en cuenta para estos contaminantes han sido los establecidos por la legislación vigente.

## Captadores de calidad del aire:

No todos los municipios disponen de captadores que recojan los datos de calidad del aire. Por lo que se ha dividido la Comunidad de Madrid en 7 zonas.



## Datos meteorológicos:

El clima también influyen en las alergias al polen, en los días lluviosos y los días nublados y sin viento los síntomas disminuyen, ya que la lluvia limpia la atmósfera y la ausencia de viento evita el desplazamiento del polen.



Por el contrario, el tiempo cálido, seco y con viento favorece una mayor distribución del polen y, en consecuencia, un aumento de los síntomas.

Recuperamos las previsiones de los siguientes datos meteorológicos, proporcionados por el **AEMET**.



Hemos tenido en cuenta los valores para los próximos tres días en cada uno de los distintos municipios de la Comunidad de Madrid.

## Cálculo del nivel de alerta:

Con todos estos datos ejecutamos el algoritmo que nos devuelve un valor numérico que al redondearlo le asignamos un nivel de alerta.

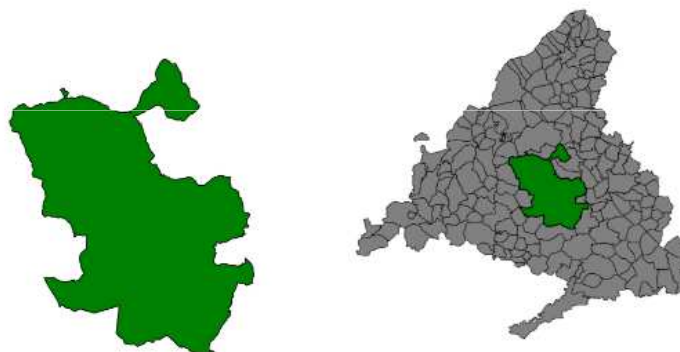
### Relación entre valor y nivel

VALOR CALCULADO	NIVEL OBTENIDO
<1	BAJO
>1 y <2	MEDIO
>2	ALTO

Todos estos datos se actualizan a primera hora de la mañana mediante un proceso batch que también ejecuta el modelo, generando la predicción inicial que se muestra en la Web.

Madrid

Para hoy, día 29-09-2016, el nivel de alerta es: Bajo



Pero a lo largo del día esta puede variar dependiendo de los reportes que hagan los usuarios .

Cada vez que un usuario realiza un reporte se vuelve a calcular el nivel de alerta para ese día. Esto permite que el servicio se ajuste a la realidad del momento.

# Búsqueda de noticias

Se ha creado un motor de búsqueda para recuperar noticias relacionadas con el polen en Madrid durante este año.

Este motor se ejecuta en el batch a primera hora de la mañana, por lo que cada día se actualizan las noticias que se muestran en la Web.

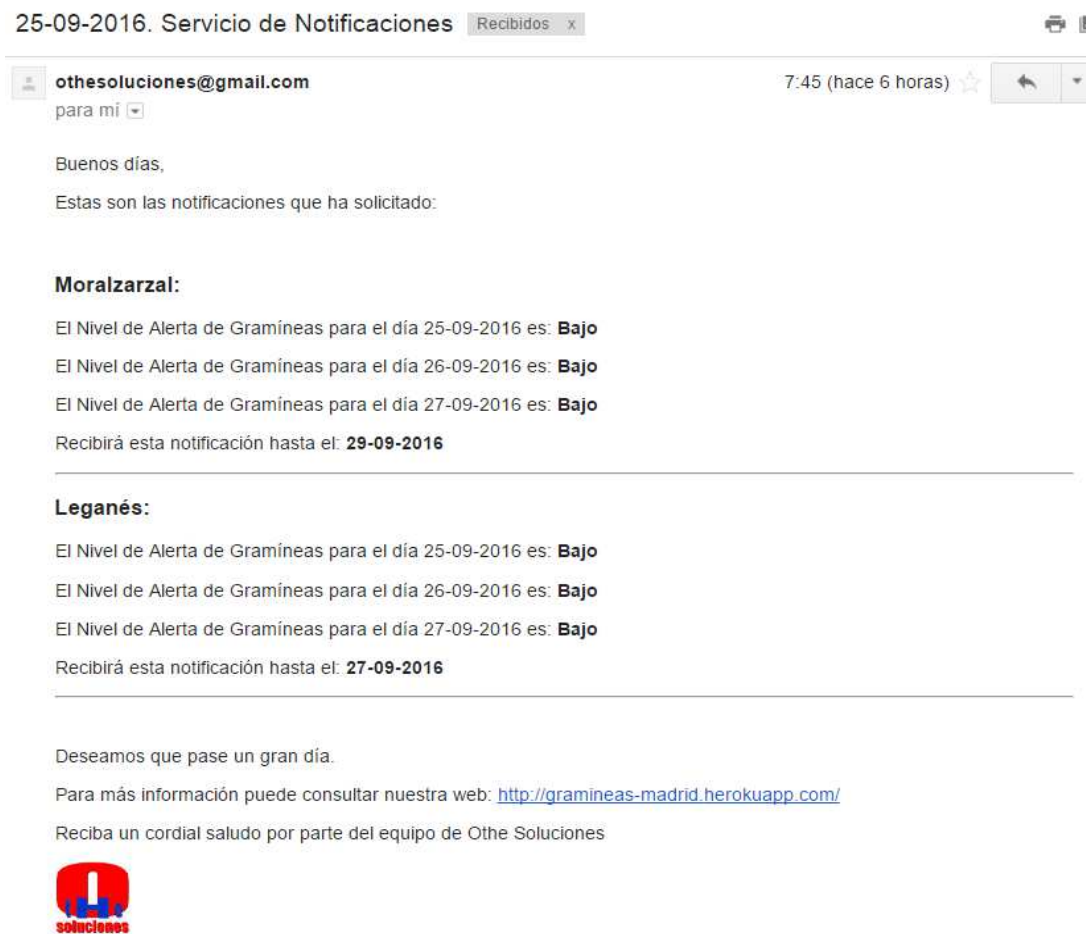
Las noticias se recuperan de los principales periódicos de la zona.



# Aplicación Web

<http://gramineas-madrid.herokuapp.com>

## Ejemplo de email recibido el 25-09-2016:



# Trabajos futuros

**Seguridad**

**Aplicación para dispositivos móviles**

**Geografía**

**Otros pólenes**

**Procesamiento de Lenguaje Natural**

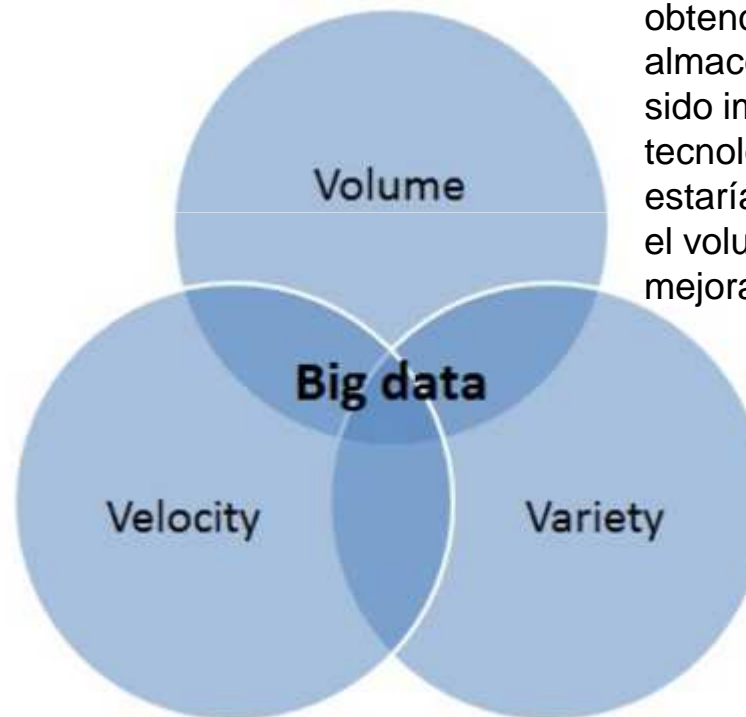
# Conclusión

Problemas encontrados:

- No se ha podido probar un ciclo anual completo.
- Problemas a la hora de recoger los datos. Red PalinoCAM.

## Las tres V's que definen el Big Data:

Los reportes de alertas actualizan la predicción en tiempo real.



Volumen no muy grande debido a las restricciones de mLab. Pero, las técnicas utilizadas para la obtención, el tratamiento o el almacenamiento de los datos han sido implementadas usando tecnología Big Data. Por lo que estaría preparado para incrementar el volumen de datos al aplicarle las mejoras comentadas.

Variedad en los datos, tanto por su origen y forma como por su formato

***Les recordamos que esta presentación, así como toda la información relativa a este proyecto, se encuentra disponible en la siguiente página web***

***<https://othesoluciones.github.io/TFM/>***

***Muchas gracias por su atención.***

***Reciban un cordial saludo.***



CIFF Trustees:



César Hernández Martín  
Sara Otero de Navascués Tomé