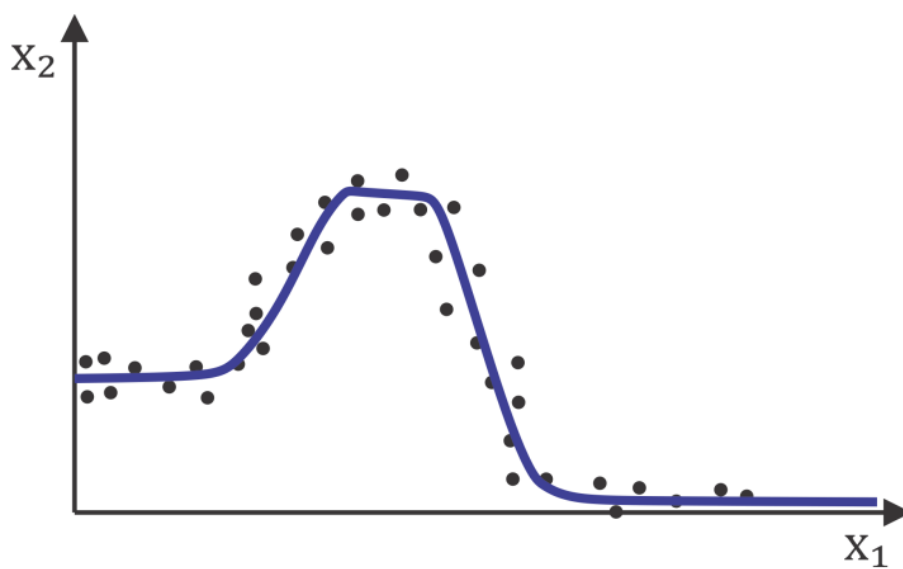


# INFORME -TP

## Sugeno



### **Integrantes del grupo:**

- Alan Gutiérrez
- Francisco Stimmler

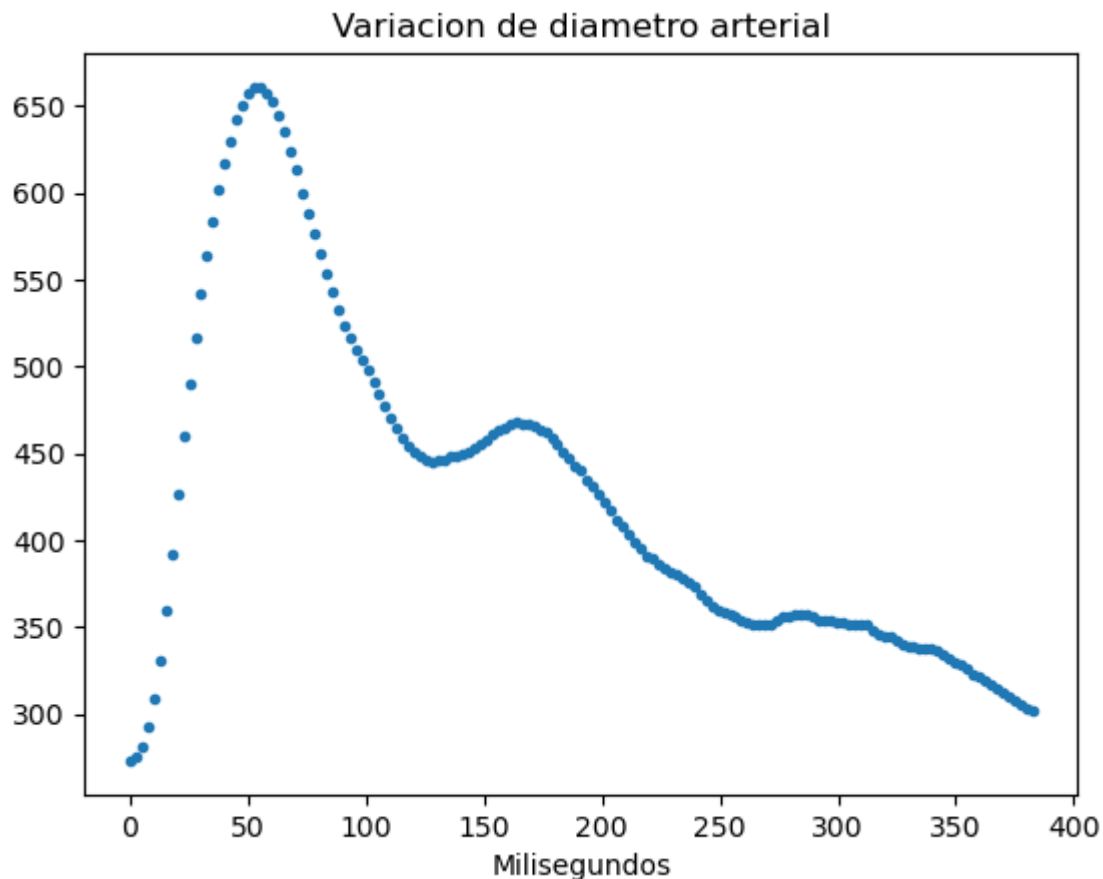
# Introducción:

Se nos dio 4 archivos de texto distinto con señales de Variación de Diámetro Arterial, que fueron muestreadas con una frecuencia de 400 Hz.

El sistema Sugeno, también conocido como modelo de Sugeno, fue propuesto por Takagi-Sugeno en la década de 1980, y es un modelo basado en reglas de tipo si-entonces, las cuales representan relaciones entrada-salida locales lineales de sistemas no lineales. La principal característica de este modelo es que se caracteriza por su capacidad para generar salidas no sólo en forma de conjuntos difusos, sino también en forma de funciones lineales o polinómicas de las variables de entrada.

Se nos pide crear varios modelos de Sugeno con distinta cantidad de reglas hasta encontrar el que creamos mejor.

Datos del archivo 3 graficados:



# Metodología:

1) FUZZIFICACIÓN: se calculan los valores de verdad de los antecedentes.

Lo primero que hicimos fue buscar la mejor manera de dividir en clúster los datos (k-means, Clustering sustractivo), una dificultad que tuvimos fue que con clustering sustractivo nos tomaba dos picos en un mismo cluster.

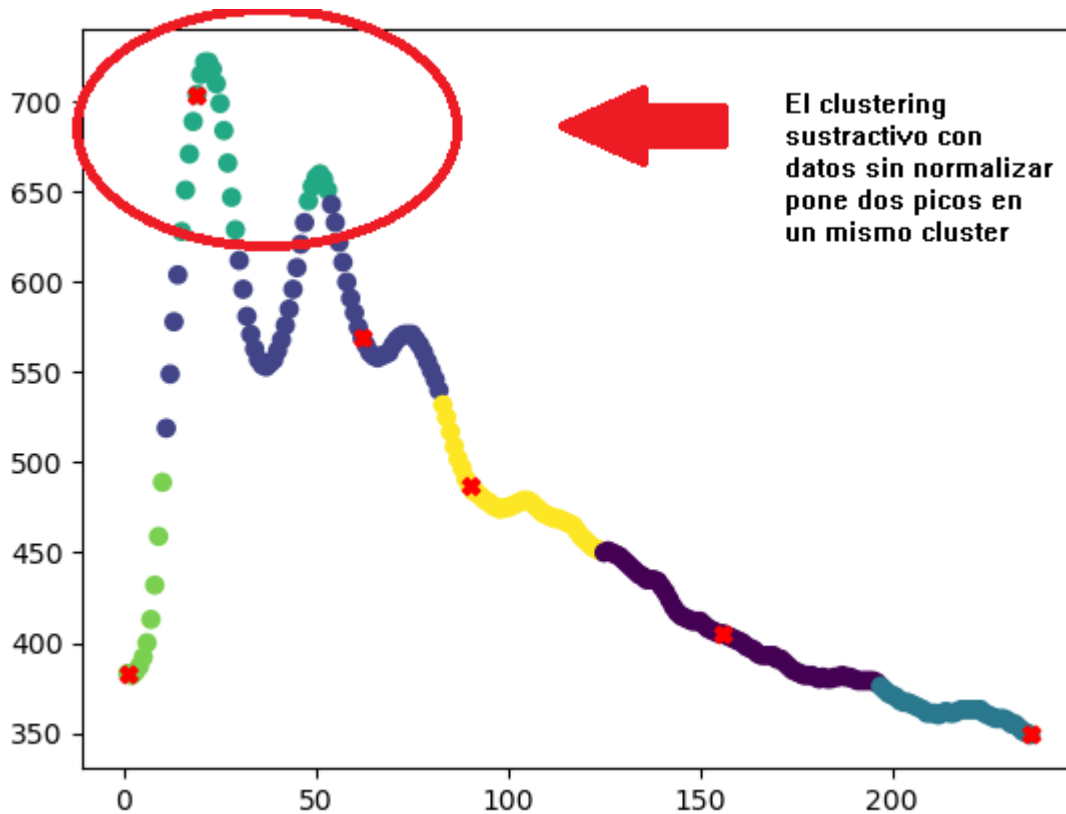


Figura 1. Clustering sustractivo con K=6

Decidimos entonces normalizar los datos y usar k-means además para poder probar con distintas cantidades de reglas.

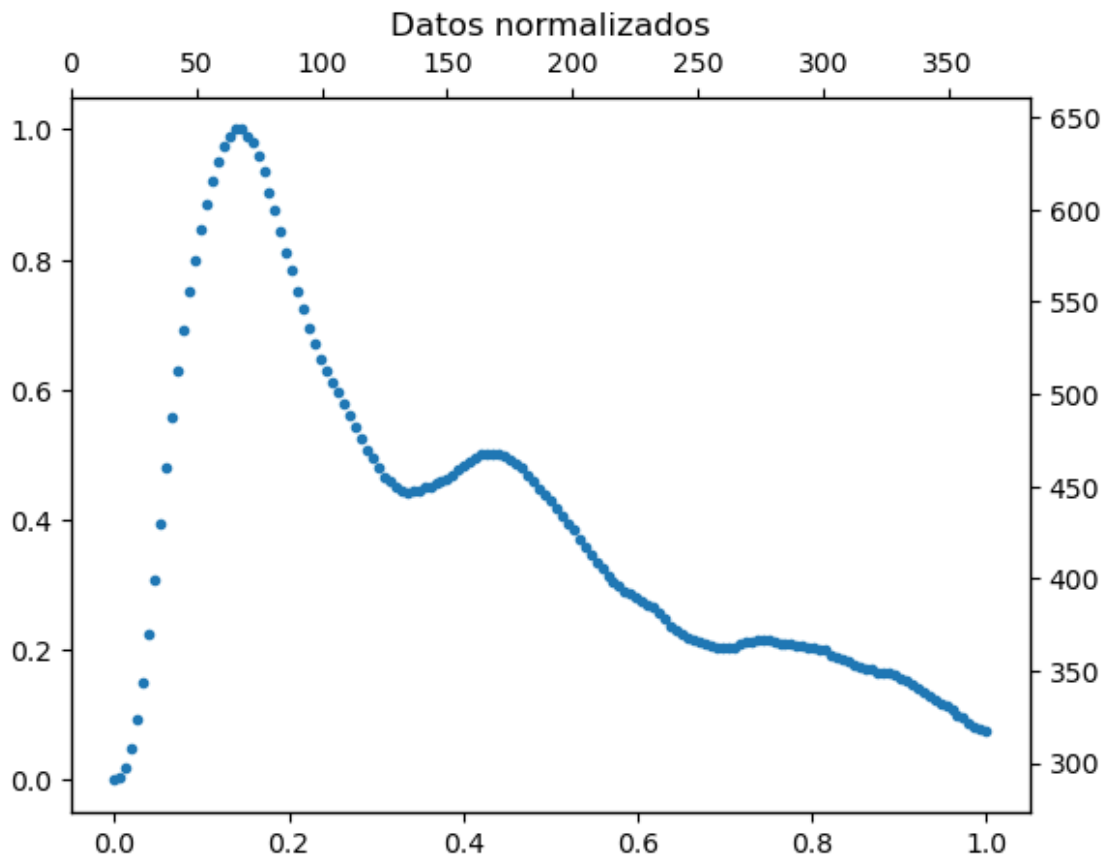


Figura 2. Datos normalizados

Datos coloreados con 7 clusters por k-means:

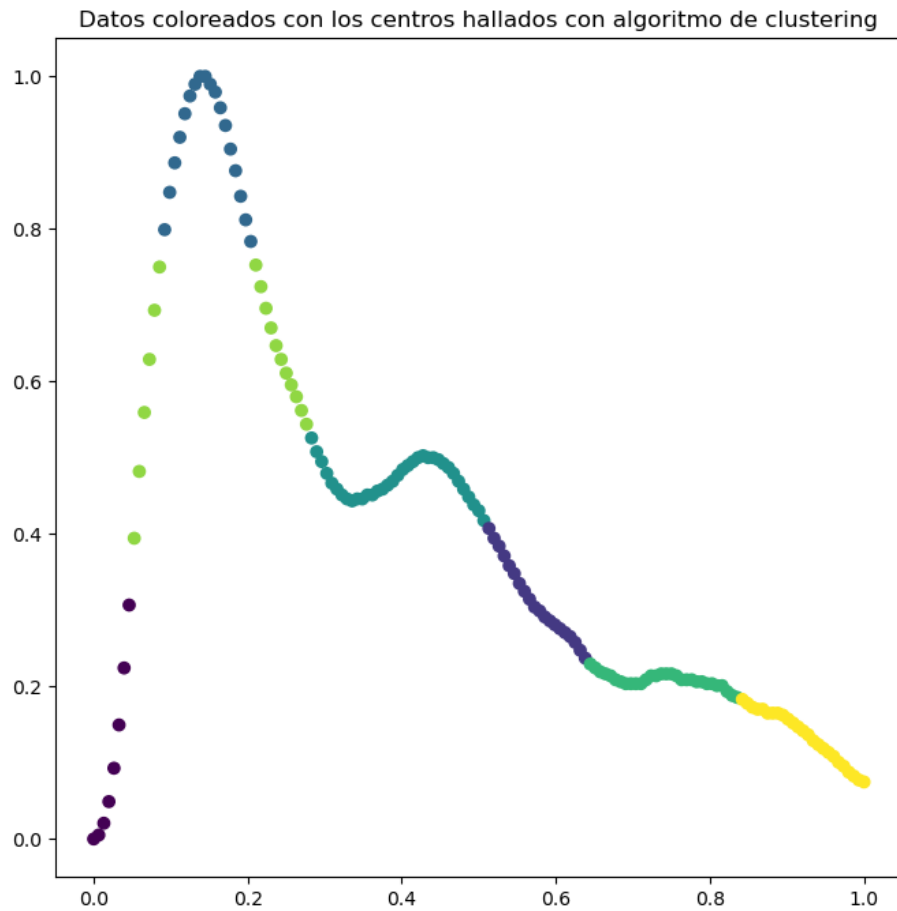


Figura 3. Datos con 7 clusters

Una vez encontrado los centros de clusters creamos las gaussianas de las funciones de pertenencia de cada regla (cantidad de clusters = cantidad de reglas).

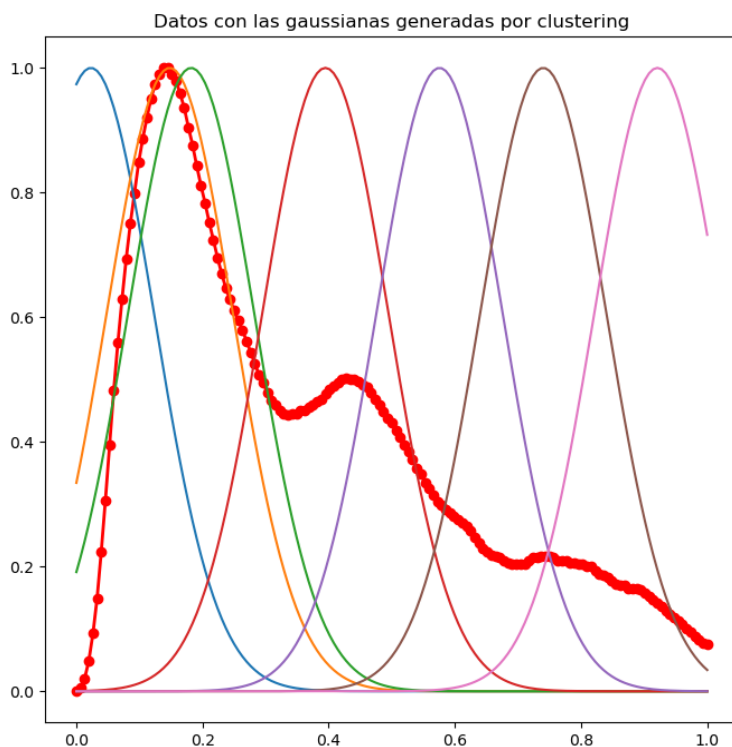


Figura 4. Gaussianas

2) INFERENCIA: se calculan los valores de las funciones de los consecuentes.

Una vez que tenemos definido los antecedentes, nos faltan los consecuentes, usamos el método de los mínimos cuadrados para calcular el valor de las funciones consecuentes:

**Método de los mínimos cuadrados (LS, *Least Squares*)** para minimizar:

$$\frac{1}{N} \sum_{i=1}^N error_i^2 = \frac{1}{N} \sum_{i=1}^N \left[ \frac{\mu_{1i}(a_1x_i + b_1) + \mu_{2i}(a_2x_i + b_2)}{\mu_{1i} + \mu_{2i}} - t_i \right]^2$$

3) AGREGACIÓN: se combinan los resultados de todas las reglas.

Luego combinamos los resultados de todas las reglas para obtener una curva final.

4) DEFUZZIFICACIÓN: se convierte un conjunto difuso en un valor de salida.

Resultado final, curva generada mediante sugeno.  
Algoritmo de cluster = K means. 7 reglas  
mse: 4.089680594370266e-05

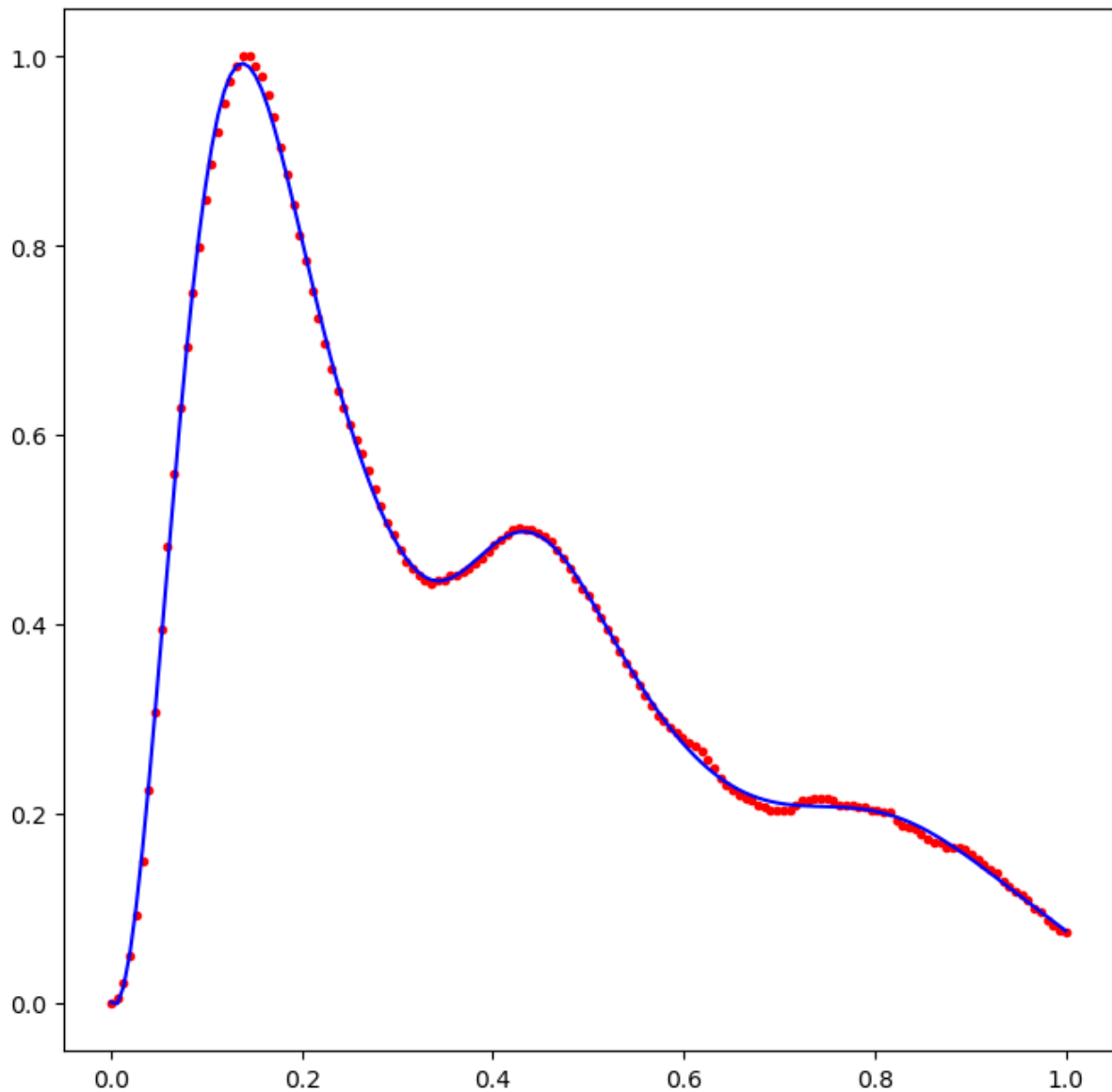
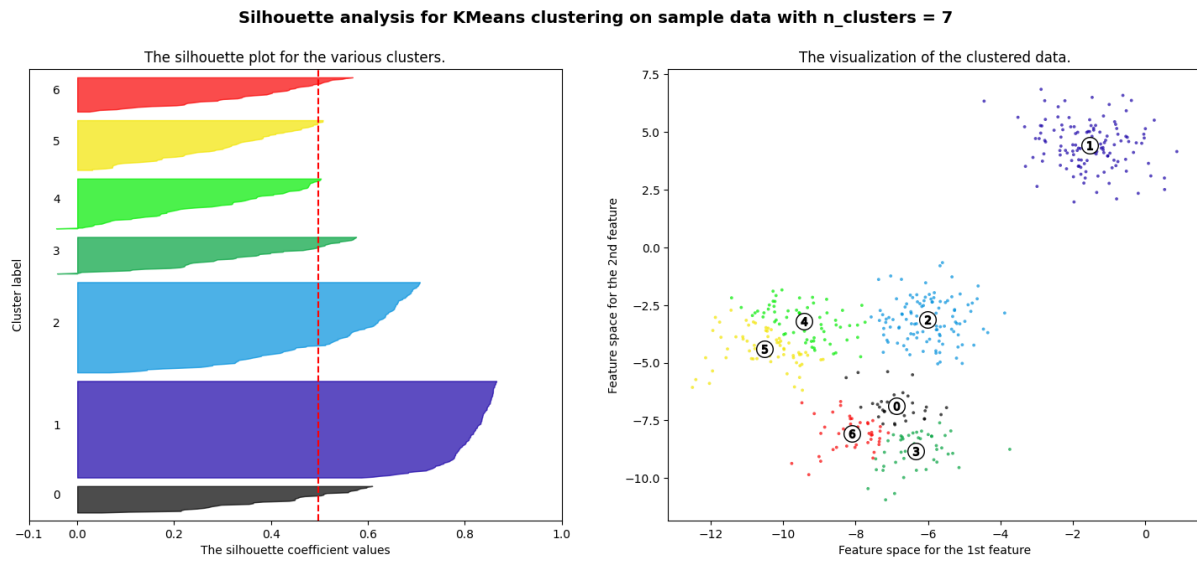


Figura 5. Aproximación por mínimos cuadrados

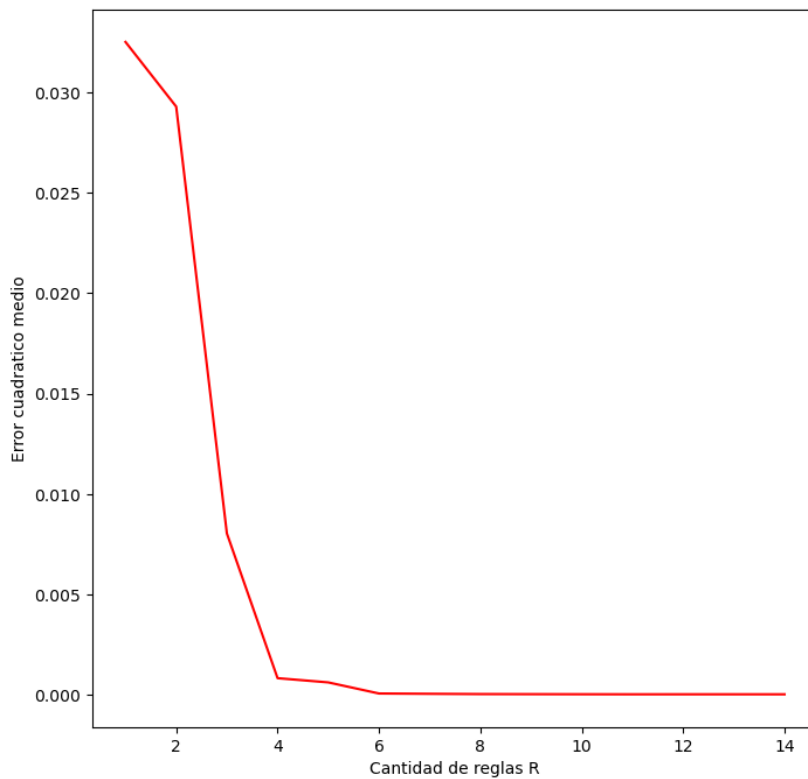
Ya que usamos K-Means decidimos probar distinta cantidad de clúster para obtener resultados distintos y comparar.

Comparamos el error cuadrático medio y además usamos Silhouette como otro criterio a comparar.



*Figura 6. Silhouette para K=7*

El error cuadrático medio de los distintos modelos con distintas cantidades de reglas.



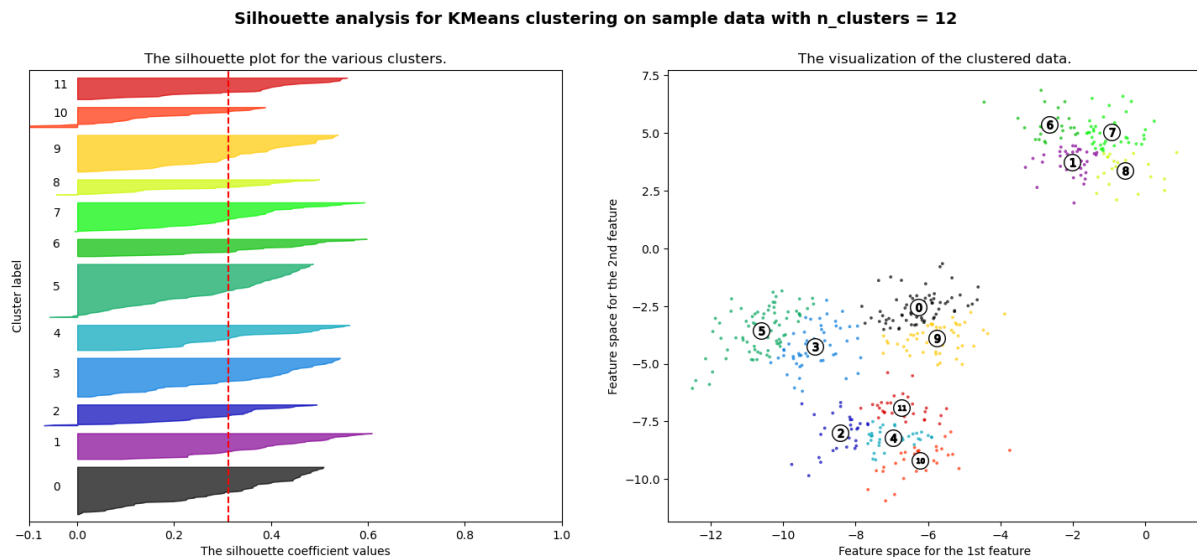
*Figura 7. Gráfico mse Vs R*

Se puede notar que con más de 4 reglas ya el cambio en el error es bajo y que no tiene sentido seguir.



# Resultados

Finalmente, se eligió usar un modelo de 11 reglas que contiene un error de apenas 0,0000160575949, y también porque en la representación gráfica de Silhouette, había muchos datos en clusters que tomaban valores negativos señalando un problema en los clusters.



*Figura 8. Silhouette para K=12*

Se puede ver, por ejemplo, que con  $K = 12$  el clúster 2 y 10 tienen varios datos con índices negativos, algo que se repite con más clústers.

Los resultados con el modelo elegido de 11 Reglas es el siguiente:

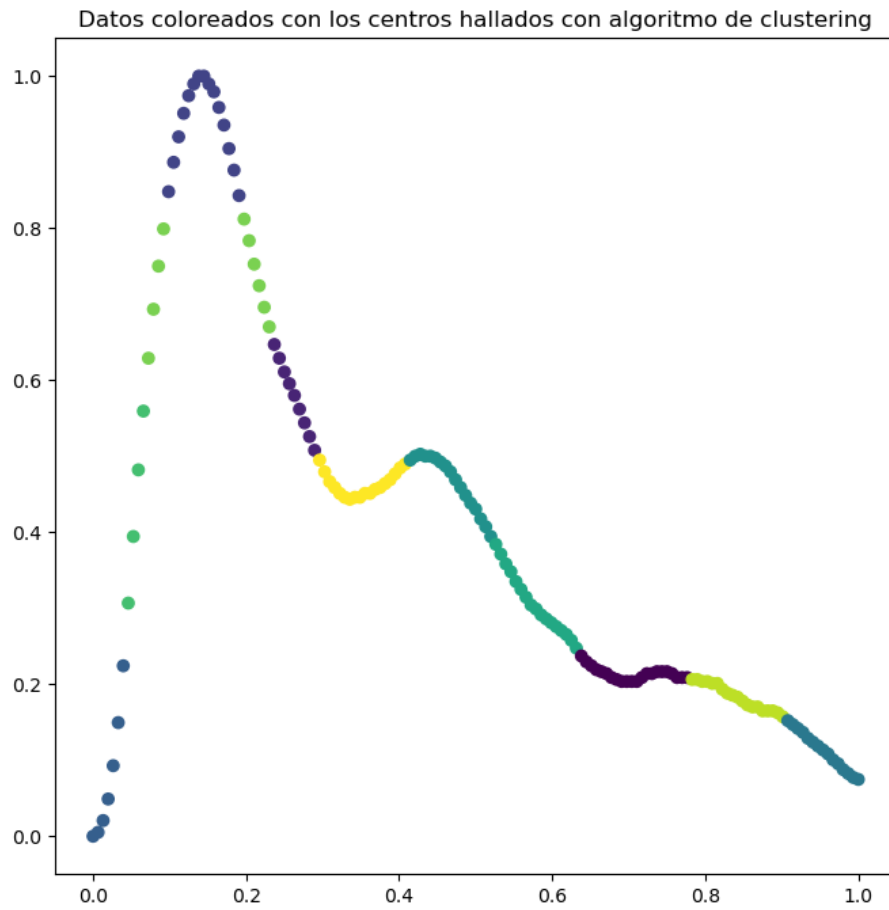


Figura 9. Clustering con  $K=11$

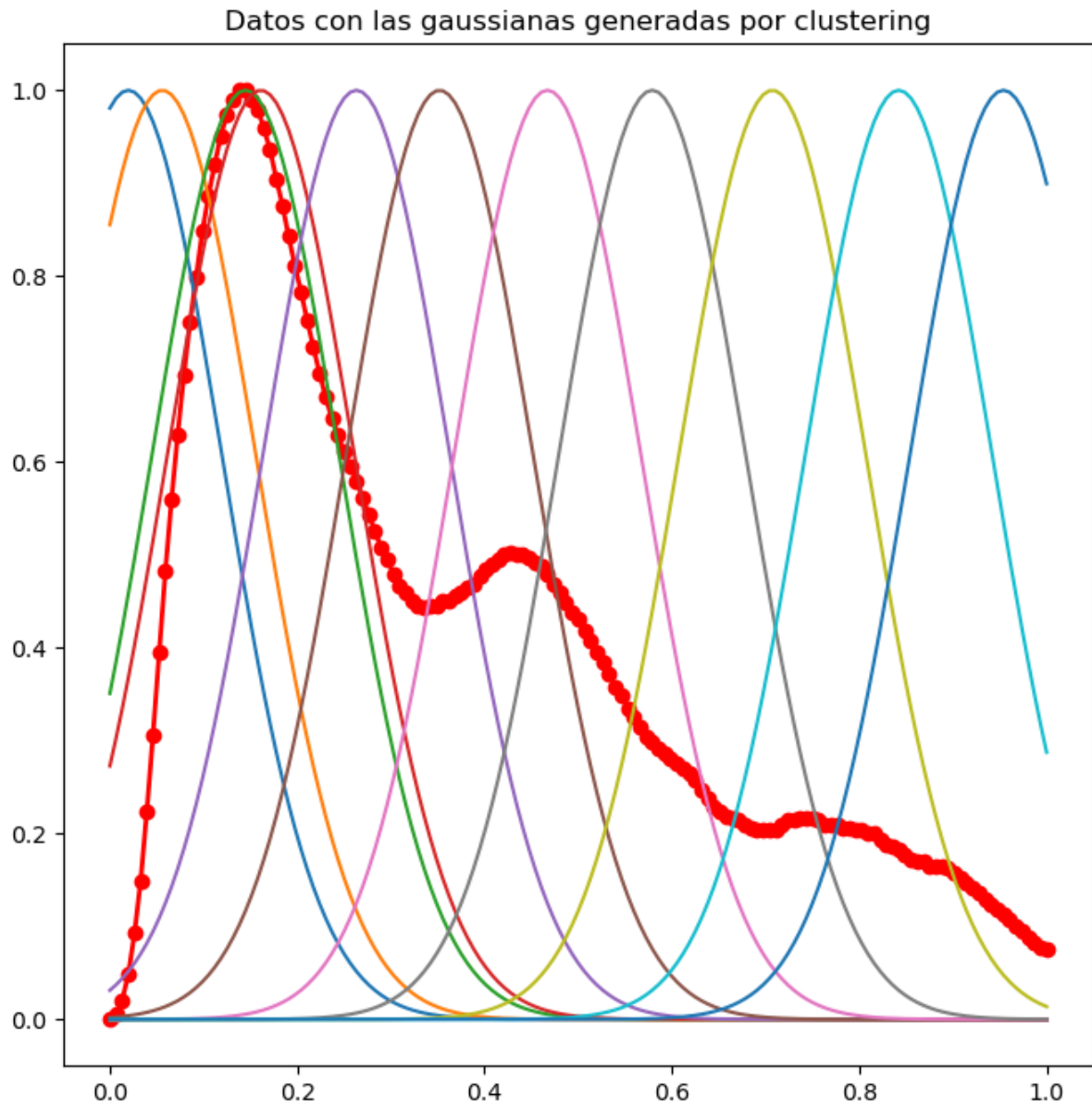


Figura 10. Gaussianas generadas para  $K=11$

Resultado final, curva generada mediante sugeno.  
Algoritmo de cluster = K means. 11 reglas  
mse: 1.605759493347939e-05

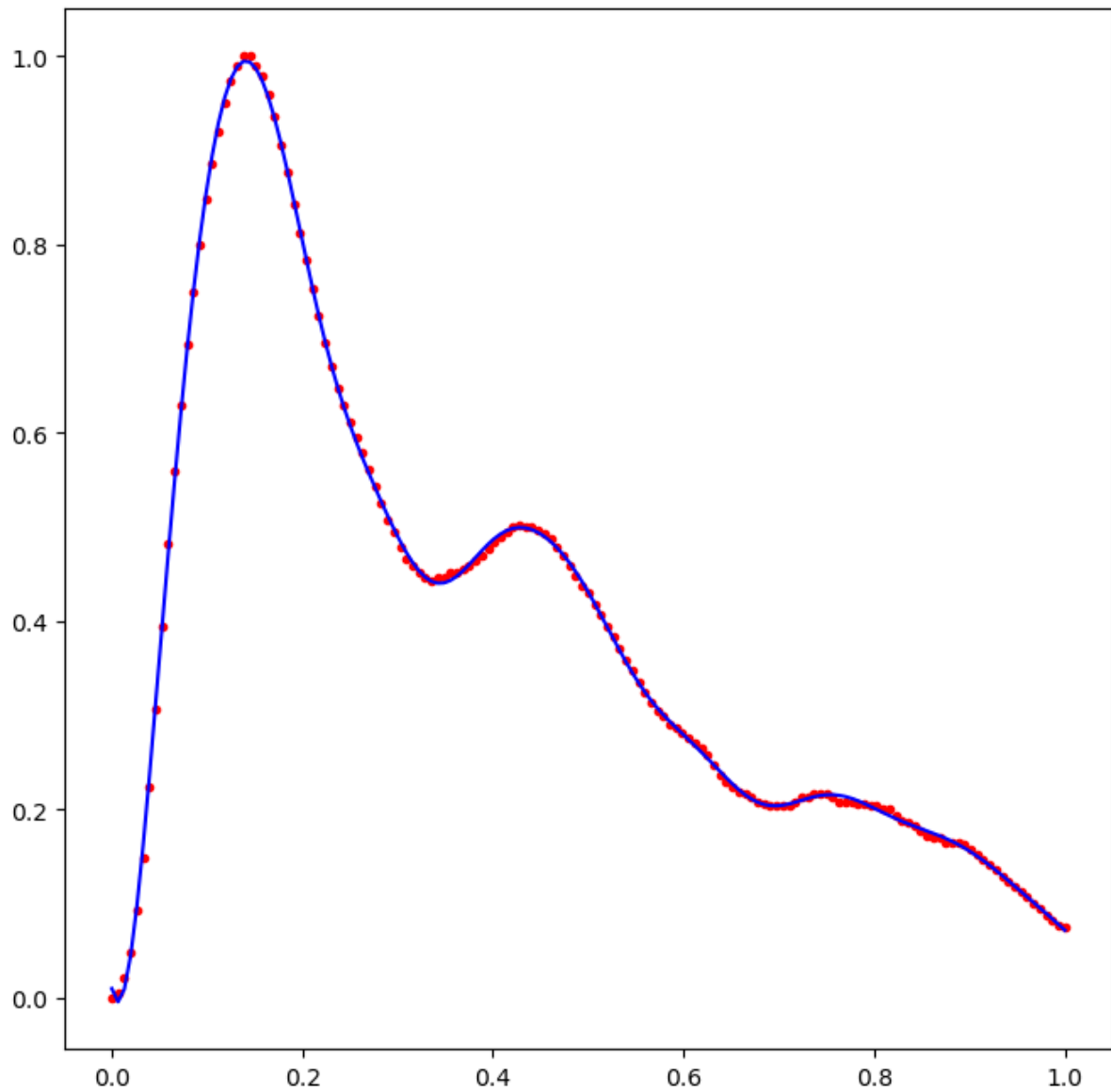


Figura 11. Curva generada

Finalmente, se nos pidió que hiciéramos un sobre muestreo de los datos con el modelo elegido, eligiendo valores dentro de la curva generada con un valor en el tiempo de entre medio de dos puntos.

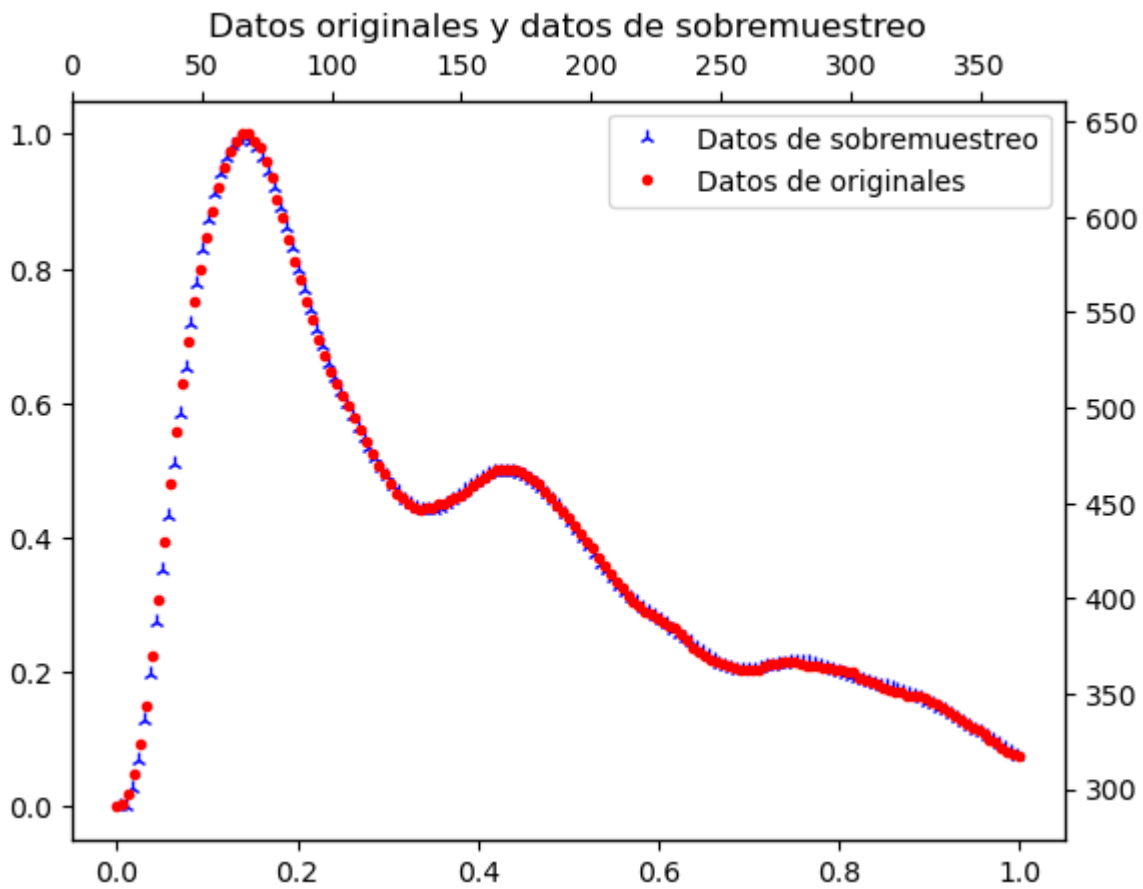


Figura 12. Sobremuestreo

Se puede observar que los datos de sobre muestreo parecen seguir los valores originales mostrando que no hay overfitting.

## Conclusiones

Se concluye que los datos dados, el mejor modelo es el de 11 reglas, también se remarca la importancia de normalizar los datos en casos como estos donde el rango de los ejes no coincide.