

# **Salud y Minería de Datos: Análisis de conglomerados con el algoritmo PAM para la mejora de la gestión de turnos médicos en un centro de salud**

**Ing. Juan Miguel Moine**  
**Ing. Javier Matías Landini**

*Grupo de Investigación en Minería de Datos*  
*U.T.N. Facultad Regional Rosario*

## **Resumen**

*PAM (Partitioning Around Medoids) es un algoritmo de Minería de Datos utilizado para el análisis de conglomerados, es decir, la identificación de grupos o clusters en un conjunto de datos. Aplicamos este algoritmo sobre la base de datos de turnos cancelados en una especialidad médica con alta demanda de atenciones, oftalmología, en un centro de salud público. En el análisis pudimos determinar la existencia de tres grupos de turnos cancelados, cuya descripción y caracterización nos posibilita mejorar la gestión de atenciones médicas programadas en esta especialidad.*

## **Palabras clave**

Minería de Datos, PAM, k-medioides, cluster analysis, análisis de conglomerados, gestión de turnos médicos, salud pública.

## **Introducción**

En la actualidad, en los centros de salud existen ciertas especialidades médicas con agenda de turnos saturada, especialmente en instituciones del ámbito público. Cuando un paciente necesita ser atendido en alguna de estas especialidades, no logra conseguir un turno (atención programada) para una fecha a corto plazo, debiendo esperar una gran cantidad de días, incluso semanas, para poder asistir a una consulta médica.

Esta situación se genera, fundamentalmente, debido al reducido horario de atención de algunos médicos y a la gran demanda de atenciones médicas especializadas. El problema crece cuando tenemos una alta tasa de ausentismo, ya sea porque el paciente no asiste a la consulta, o bien porque cancela su turno sin anticipación, el mismo día de la atención. Empíricamente se sabe que estos turnos generalmente quedan sin poder aprovecharse por otros pacientes.

El caso en estudio es el de un centro de salud municipal ubicado en la provincia de Santa Fe, Argentina. En el mismo, se brinda atención para más de 15 especialidades médicas. Los turnos para los especialistas son gestionados mediante un sistema informático. La base de datos de este sistema constituye la fuente de información primaria de nuestro análisis.

Debido a la gran demanda de atenciones y a los pocos recursos disponibles, fundamentalmente en algunas especialidades, es necesario llevar a cabo en el centro de salud una política de administración de turnos que permita aprovechar al máximo el horario de atención del médico. Para atacar esta problemática, podemos trabajar desde distintos enfoques.

Los directivos de las organizaciones frecuentemente toman decisiones basándose en su experiencia, utilizando un enfoque subjetivo. Sin embargo, en la actualidad, se ha demostrado que es una mejor alternativa tomar las decisiones basándose en cifras, hechos y datos.

La *Minería de Datos (Data Mining)*, es el proceso de extraer conocimiento útil, comprensible y novedoso de grandes volúmenes de datos <sup>[1]</sup>. Su principal objetivo es encontrar información oculta o implícita, que no es posible obtener mediante métodos estadísticos convencionales.

Dentro de la Minería de Datos, las técnicas de *Agrupamiento (Clustering)*, también llamadas *análisis de conglomerados*, tienen como objetivo agrupar elementos de un conjunto de datos en grupos homogéneos en función de las similitudes o similitudes entre ellos. En esta técnica, se dividen las observaciones en grupos (clusters), donde aquellos elementos ubicados en el mismo grupo, se consideran “similares” <sup>[2]</sup>.

Para poder aplicar las técnicas de agrupamiento, debemos disponer de un conjunto de datos de  $n$  observaciones evaluadas en  $p$  variables. La elección del algoritmo de agrupamiento a utilizar, dependerá de la naturaleza de las variables en estudio (si son cualitativas o cuantitativas), el tamaño del set de datos, y si nos interesa obtener una estructura aglomerativa (métodos de particionamiento) o una estructura jerárquica (métodos jerárquicos).

Los métodos de particionamiento convencionales (partitioning methods), trabajan sólo sobre conjuntos de datos caracterizados por un tipo de variable, como en el caso de  $k$ -medias<sup>1</sup> (k-means) que trabaja principalmente con variables cuantitativas. En aplicaciones a problemas sociales, estos algoritmos resultan poco prácticos ya que los individuos o elementos en estudio son caracterizados por variables numéricas y cualitativas.

*PAM (Partitioning Around Medoids)* es un algoritmo de particionamiento basado en la búsqueda de  $k$  elementos representativos (llamados medioides o centroides) alrededor de los cuales se concentran los demás individuos, dividiendo de esta forma el conjunto de observaciones en  $k$  grupos. Este algoritmo fue publicado por Kaufman y Rousseeuw en 1990. Constituye una técnica factible de aplicar cuando el conjunto de datos está formado por variables mixtas.

### Objetivos

En este trabajo, presentamos una alternativa para mejorar la gestión de turnos médicos con Minería de Datos, específicamente, aplicando el algoritmo de agrupamiento PAM a un conjunto de turnos otorgados en una especialidad, oftalmología, que posee gran demanda de atenciones y agenda de turnos saturada.

Analizamos particularmente el caso de los turnos cancelados, es decir, aquellos turnos en los cuales el paciente avisa que no asistirá. La antelación con la que se realiza la

---

<sup>1</sup> *K-medias (k-means)* es un algoritmo de agrupamiento no jerárquico, que trabaja especialmente sobre un conjunto de datos formado por variables cuantitativas. El algoritmo divide el conjunto de observaciones en  $k$  grupos, minimizando la distancia euclídea de cada individuo al centro de su grupo.

cancelación es muy importante, ya que no tiene el mismo impacto en la agenda del médico el caso de un paciente que cancela su turno el mismo día de la atención, que otro que cancela su turno con anticipación. Probablemente en el primer caso, el turno quede sin otorgarse, impidiéndole la oportunidad de una atención más temprana a otros pacientes.

Mediante la aplicación del algoritmo P.A.M, evaluamos la existencia de una estructura de conglomerados (grupos) entre los turnos cancelados en oftalmología, analizando las características que identifican a cada uno.

### El Algoritmo PAM

PAM, Partitioning Around Medoids, es una implementación de k-medioides, algoritmo que divide un conjunto de datos en  $k$  grupos, buscando  $k$  observaciones representativas (llamadas medioides) alrededor de las cuales se agrupan el resto de los elementos.

El valor de  $k$ , cantidad de grupos buscados, es un valor establecido a priori, y debe ser suministrado como entrada al algoritmo.

El algoritmo PAM no se aplica directamente sobre los datos originales, sino sobre la matriz de disimilaridades<sup>2</sup> o distancias:

- **X**: matriz de datos de orden  $n \times p$ , ( $n$ : cantidad de observaciones,  $p$ : cantidad de variables)
- **D**: matriz de disimilaridades, de orden  $n \times n$ . Cada valor,  $d_{ij}$ , representa la disimilaridad o distancia entre dos individuos  $(i,j)$ .

Para obtener la matriz de disimilaridades, se calcula para cada par de individuos  $(i,j)$  el coeficiente de disimilaridad entre los mismos:

$$d_{ij} = \frac{\sum_{h=1}^p w_{ijh} d_{ijh}}{\sum_{h=1}^p w_{ijh}} \quad (1)$$

Donde:

- $d_{ijh}$  es la disimilaridad entre los elementos  $(i,j)$  en la  $h$ -ésima variable. El cálculo de este valor depende de la naturaleza de la variable  $h$ .
  - Si es cualitativa,  $d_{ijh}=0$  si  $x_{ih}=x_{jh}$ , y  $d_{ijh}=1$  en caso contrario.
  - Si es numérica,  $d_{ijh}$  se calcula como el cociente entre el valor absoluto de  $x_{ih}-x_{jh}$  sobre el rango de la variable  $h$ .
- $w_{ijh}$  es el peso de la variable  $h$ :
  - $w_{ijh}=0$ , si  $x_{ih}$  o  $x_{jh}$  son datos ausentes
  - $w_{ijh}=1$  en caso contrario

<sup>2</sup> La matriz de disimilaridades (D) de un conjunto de datos compuesto por  $n$  observaciones, es una matriz de orden  $n \times n$ , donde cada una de sus entradas  $d_{ij}$ , es una medida de la distancia o disimilaridad entre las observaciones  $(i,j)$ .

El coeficiente de disimilaridad global  $d_{ij}$  planteado en (1), es una leve variación del coeficiente de similaridad de Gower<sup>3</sup>.

Sobre la matriz de disimilaridades (D), se busca un subconjunto de elementos  $\{m_1, \dots, m_k\}$  perteneciente al conjunto de datos, tal que minimicen la función objetivo:

$$\sum_{i=1}^n \min_{t=1, \dots, k} d(i, m_t) \quad (2)$$

Donde  $d(i, m_t)$  es la distancia del elemento  $i$  a su medioide más próximo  $m_t$ .

Los  $k$  objetos representativos o medioides, deben minimizar esta función objetivo (2), la cual está formada por la suma de las disimilaridades de cada objeto  $i$  a su medioide más cercano.

Las iteraciones del algoritmo PAM<sup>[3]</sup> podemos dividir las en dos etapas:

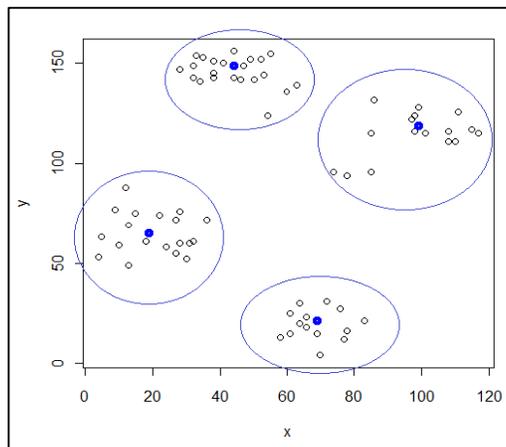
1. *Etapa BUILD*

Selecciona  $k$  objetos como los medioides iniciales. Generalmente se seleccionan como medioides iniciales, aquellos individuos con la menor distancia al resto de los elementos.

2. *Etapa SWAP*

Si la función objetivo se puede reducir intercambiando un medioide con un elemento que no lo es, se efectúa el cambio. Esta acción continúa hasta que el valor de la función objetivo no puede mejorarse más.

Como resultado del algoritmo, se obtiene finalmente una división del conjunto de observaciones en  $k$  grupos, siendo los medioides los elementos representativos de cada uno.



**Figura 1.** Ejemplo de aplicación del algoritmo PAM para un valor de  $k=4$ . Los medioides son las observaciones que mejor representan a los individuos de su grupo.

<sup>3</sup> Gower (1971) propuso un coeficiente de similaridad global entre dos individuos de una muestra, que permite trabajar con variables mixtas (cualitativas y cuantitativas)<sup>[4]</sup>.

Análisis de silueta: Evaluando la robustez del agrupamiento

Para estimar la robustez de la estructura de conglomerados resultante, llevamos a cabo un análisis de silueta (silhouette).

En este análisis, para cada objeto  $i$ , se calcula un valor de silueta <sup>[5]</sup> que se define como:

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (3)$$

Donde:

- $b(i)$  es la distancia promedio del individuo  $i$ , con respecto a los individuos del grupo más cercano. Si el objeto  $i$  fue agrupado en el cluster A, consideremos un cluster C (distinto de A), y definimos  $d(i, C)$  como la disimilaridad media de  $i$  con los objetos de C. Después de calcular  $d(i, C)$  para todos los clusters  $C \neq A$ , seleccionamos aquel que tiene el mínimo valor:

$$b(i) = \min_{C \neq A} d(i, C) \quad (4)$$

El cluster B donde se alcanza este mínimo, es decir  $d(i, B) = b(i)$ , se denomina grupo vecino del objeto  $i$ .

- $a(i)$  es la distancia promedio del individuo  $i$ , con respecto al resto de los individuos de su grupo:

$$a(i) := \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (5)$$

El valor de  $s(i)$ , mide qué tan buena es la asignación del elemento  $i$  a su cluster. Estará comprendido en el rango  $[-1, 1]$  y puede ser interpretado de la siguiente forma:

- $s(i) \approx 1$ , el objeto está bien clasificado (en el grupo A)
- $s(i) \approx 0$ , el objeto cae en el medio de dos grupos (entre A y B)
- $s(i) \approx -1$ , el objeto está mal clasificado (el objeto  $i$  está más cerca de B que de A)

La calidad del modelo se puede medir con el *coeficiente de silueta promedio*, que se define como el promedio de todos los valores de silueta  $s(i)$  de los individuos en el conjunto de datos. Los valores de este coeficiente pueden ser interpretados empíricamente según la tabla 1.

| Valores de Silueta promedio | Robustez de la aglomeración                                 |
|-----------------------------|---|
| Entre 0.71 – 1.00           | Se encontró una estructura fuerte y robusta.                |
| Entre 0.51 – 0.70           | Se encontró una buena estructura.                           |
| Entre 0.26 - 0.50           | La estructura no es buena, deberían probarse otros métodos. |
| < 0.25                      | No se encontró una estructura.                              |

Tabla 1: Interpretación empírica del coeficiente de silueta promedio

### Elementos del trabajo y metodología

Realizamos el estudio para el período comprendido entre 01/01/2008 y 31/12/2009. Utilizamos como fuente primaria de información la base de datos del sistema informático de turnos, donde se registran los movimientos asociados a los mismos con excepción de la asistencia del paciente a la atención, debido a políticas de la organización.

En este período, sólo un médico ha efectuado atenciones en oftalmología, con una frecuencia de dos días por semana. Se registraron 158 días de atención. Los horarios en la totalidad de los turnos otorgados (cancelados y no cancelados) se distribuyen uniformemente entre las 12:30 hs. y 18:00 hs. Entre los mismos encontramos 137 turnos cancelados.

En el Centro de Salud se considera que un paciente cancela su turno sin anticipación, si notifica que no asistirá el mismo día de la atención. De los turnos cancelados, 76 se cancelaron sin anticipación (aproximadamente un promedio de 1 turno por semana).

Las *unidades de análisis*, quedaron constituidas por los 137 turnos cancelados en la especialidad, para el período mencionado. Estas observaciones fueron evaluadas en las siguientes ocho variables:

| Variable                          | Tipo        | Valores posibles   | Descripción   |
|-----------------------------------|-------------|--|---|
| Horario del turno <sup>4</sup>    | Cualitativa | Primera tarde<br>[12:30 hs. , 15:00 hs.]<br>Segunda tarde<br>[15:00 hs, 18:00] | La categoría “primera tarde” corresponde a un horario generalmente “no laborable” en la ciudad donde reside el centro. La categoría “segunda tarde” corresponde a un horario laboral. |
| Edad del paciente                 | Numérica    | [0,100]  | Edad del paciente al momento de la atención.  |
| Sexo del paciente                 | Cualitativa | M/F  | Sexo del paciente.  |
| Primera vez                       | Cualitativa | Si/No  | Indica si es la primera vez que el paciente se atiende en la especialidad.  |
| Días antelación solicitud         | Numérica    | [0,30]   | Cantidad de días entre la solicitud del turno y la fecha de atención.   |
| Recepcionista que otorgó el turno | Cualitativa | [A,B,C,D]  | Variable que indica qué recepcionista otorgó el turno.  |
| Cancelado con anticipación        | Cualitativa | Si/No  | Variable que indica si el turno fue cancelado con anticipación (al menos un día antes).   |
| Volvió a otorgarse                | Cualitativa | Si /No   | Variable que indica si el turno cancelado pudo volver a otorgarse a otro paciente.  |

**Tabla 2:** Variables contempladas en el estudio

<sup>4</sup> Los comercios de la ciudad donde reside el centro de salud, trabajan con horario de atención discontinuo. Es decir, que permanecen cerrados entre el mediodía (aproximadamente 12:30 hs) y las primeras horas de la tarde (aproximadamente las 15 hs).

La distribución de cada variable se visualiza en la figura 2:

```
> summary(Datos)
usuarioOtorgo diasAntelacSol canceloConAnticipacion sexo volvioAotorgarse edad primeraVez horario
A:45 Min. : 0.00 false:80 F:82 false:82 Min. : 1.00 false:60 primeraTarde:43
B:27 1st Qu.: 7.00 true :57 M:55 true :55 1st Qu.:15.00 true :77 segundaTarde:94
C:51 Median :12.00 Median :29.00
D:14 Mean :12.91 Mean :31.32
3rd Qu.:19.00 3rd Qu.:45.00
Max. :30.00 Max. :84.00
```

**Figura 2:** análisis univariante del conjunto de datos en estudio con el software R

Analizando los valores de la figura 2 podemos ver que:

- Los usuarios A y C son los que mayor cantidad de turnos han otorgado.
- El tiempo promedio de antelación en la solicitud de los turnos es de 12 días.
- La mayor parte de los turnos fueron cancelados sin anticipación.
- Predominan pacientes de sexo femenino.
- La mayoría de los turnos cancelados no volvió a otorgarse.
- La edad promedio de los pacientes es de 30 años.
- La mayor cantidad de turnos se ha otorgado a pacientes nuevos en la especialidad.
- La mayor parte de los turnos cancelados corresponden al horario “segunda tarde”.  
Es importante destacar que, previamente, se determinó que esta variable se distribuye uniformemente en la totalidad de turnos otorgados.

Mediante el algoritmo de agrupamiento PAM buscamos una estructura de conglomerados en este conjunto de datos, caracterizado por variables cualitativas y cuantitativas.

Los datos fueron analizados con R <sup>[6]</sup>, un software libre y gratuito que nos proporciona un entorno completo de programación para análisis estadístico y gráfico, con interesantes funciones para minería de datos. Cuenta con un paquete (módulo) especial para análisis de conglomerados, llamado *cluster*. En este paquete encontramos un conjunto de comandos que nos permiten utilizar el algoritmo PAM.

La estructura final de los datos que se proporcionó como entrada al algoritmo, llamada también *vista minable*, consistió en una matriz de 137 observaciones ( $n=137$ ), y un conjunto de  $p$  variables, seleccionadas entre las presentadas en la tabla 2.

La variable “Volvió a otorgarse”, la cual se encuentra estrechamente vinculada a la variable “Canceló con anticipación”, se ha omitido para la construcción del modelo con el objetivo de utilizarla para un análisis posterior de los grupos obtenidos.

Se evaluaron distintas combinaciones de las variables en estudio, considerando primero todas las variables para el análisis ( $p=7$ ), y posteriormente distintos subconjuntos de las mismas.

En cada experimentación:

- Se tomó un subconjunto de variables de entrada ( $p \leq 7$ ) y se corrió el algoritmo en R.

- Analizamos los valores de silueta promedio para distintos valores de  $k$  (cantidad de grupos).
- Para valores de  $k$  con una buena silueta promedio ( $>0.5$ ), se examinaron las características de los grupos formados por medio de un análisis estadístico de las variables en cada cluster, buscando una estructura posible de interpretar en el dominio del problema.

### Resultados

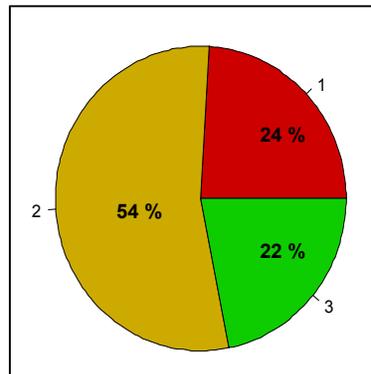
Luego de diversas experimentaciones, al suministrar como entrada al algoritmo las variables “Días antelación solicitud”, “Canceló con anticipación”, “Edad” y “Horario”, hemos detectado una buena estructura de conglomerados, con potencial interpretación en el dominio, para un valor de  $k=3$  (es decir, 3 grupos).

Los medioides encontrados por el algoritmo en los tres grupos, son descriptos en la tabla 3:

| N° Grupo | Observación medioide | Variable                  |                          |      |              |
|----------|----------------------|---------------------------|--------------------------|------|--------------|
|          |                      | Días antelación solicitud | Canceló con anticipación | Edad | Horario      |
| 1        | 8                    | 14                        | true                     | 10   | primeraTarde |
| 2        | 132                  | 9                         | false                    | 33   | segundaTarde |
| 3        | 65                   | 21                        | true                     | 31   | segundaTarde |

*Tabla 3: Medioides de los 3 grupos obtenidos*

Podemos ver en la figura 3, que el 24% de los elementos (33 turnos) fueron agrupados en el grupo 1, el 54% (74 turnos) en el grupo 2, y el 22% (30 turnos) en el grupo 3.



*Figura 3: cantidad de elementos agrupados en cada cluster.*

Con la intención de poder describir las características de los grupos formados, realizamos un análisis estadístico evaluando la distribución de las variables en cada cluster (figura 4).

```

> summary(objetosCluster1)
diasAntelacSolic canceloConAnticipacion edad horario
Min. : 2.00 false: 6 Min. : 1.00 primeraTarde:33
1st Qu.: 8.00 true :27 1st Qu.: 6.00 segundaTarde: 0
Median :14.00 Median :10.00
Mean :12.79 Mean :18.24
3rd Qu.:16.00 3rd Qu.:22.00
Max. :29.00 Max. :73.00

> summary(objetosCluster2)
diasAntelacSolic canceloConAnticipacion edad horario
Min. : 0.00 false:74 Min. : 5.00 primeraTarde:10
1st Qu.: 7.00 true : 0 1st Qu.:22.00 segundaTarde:64
Median : 9.50 Median :33.00
Mean :11.23 Mean :36.85
3rd Qu.:18.75 3rd Qu.:47.00
Max. :22.00 Max. :84.00

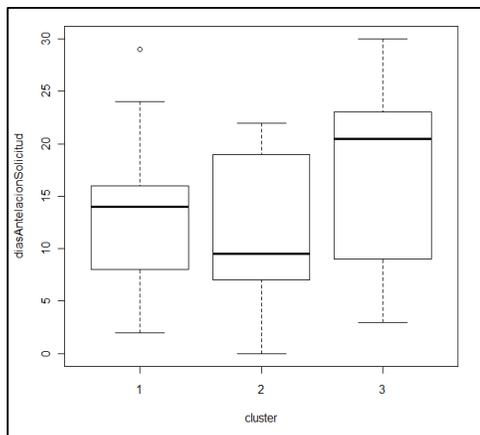
> summary(objetosCluster3)
diasAntelacSolic canceloConAnticipacion edad horario
Min. : 3.00 false: 0 Min. : 3.00 primeraTarde: 0
1st Qu.: 9.75 true :30 1st Qu.:18.50 segundaTarde:30
Median :20.50 Median :30.00
Mean :17.20 Mean :32.07
3rd Qu.:22.75 3rd Qu.:40.75
Max. :30.00 Max. :84.00

```

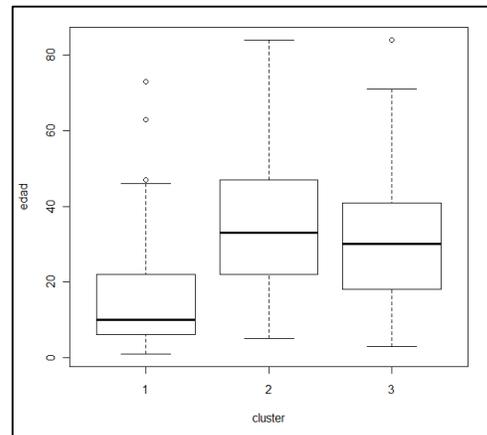
**Figura 4:** salida del software R al analizar la distribución de las variables en cada cluster.

Es interesante analizar cómo las variables adoptan distinta distribución según el grupo donde nos ubiquemos:

- **Variable “días antelación solicitud”:** la distribución de la variable (figura 5) indica que los turnos cancelados asignados al grupo 2 (cuya mediana es cercana a 9 días) fueron solicitados, en su mayoría, con menor antelación que en los grupos 1 y 3 (cuyas medianas son 14 y 21 días respectivamente).
- **Variable “edad del paciente”:** podemos ver en la figura 6, que los turnos asignados al cluster 1 corresponden en su mayoría a pacientes jóvenes y niños (el 75% es menor de 22 años). Caso contrario es el de los grupos 2 y 3, que corresponden a pacientes adultos (75% mayores a 22 años en el grupo 2 y 75% mayores a 19 años en el grupo 3).

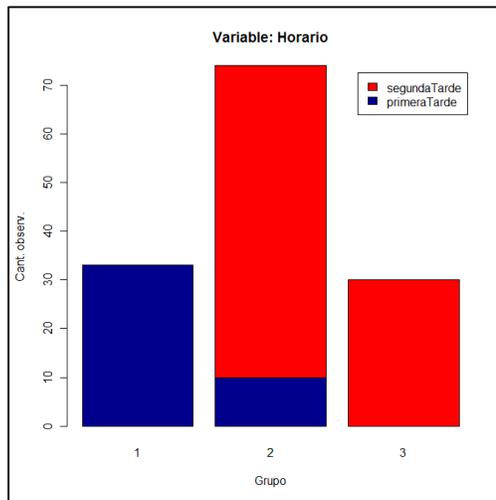


**Figura 5:** distribución de la variable “días antelación solicitud” en los grupos obtenidos.

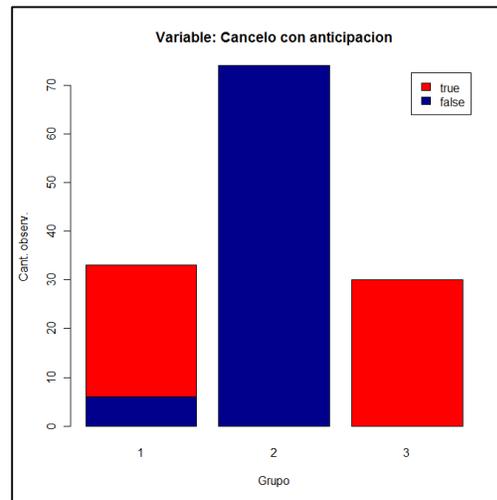


**Figura 6:** distribución de la variable “edad” en cada grupo.

- **Variable “Horario”:** se observa que la totalidad de los turnos asignados al grupo 1 corresponden a la primera tarde (entre las 12:30 hs. y las 15 hs., correspondiente a un horario “no laborable” en la ciudad donde reside el centro). La mayoría de los turnos cancelados asignados a los grupos 2 y 3 pertenecen a la segunda tarde (entre las 15 hs. y 18 hs., horario laborable en la ciudad).
- **Variable “Cancelo con anticipación”:** podemos ver que los turnos cancelados sin anticipación se concentran en el grupo 2, que resulta ser el de mayor tamaño. Los grupos 1 y 3 corresponden a turnos cancelados con anticipación.

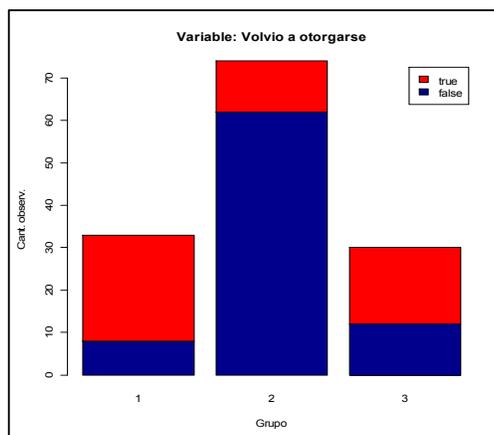


**Figura 7:** distribución de la variable “horario” en cada cluster.



**Figura 8:** distribución de la variable “canceló con anticipación” en cada cluster.

- **Variable “volvió a otorgarse”:** Posteriormente, analizamos en los grupos formados la variable “volvió a otorgarse”, la cual no fue considerada para armar el modelo. Podemos confirmar que los turnos que se cancelan sin anticipación (cluster 2) en su mayoría no vuelven a otorgarse.

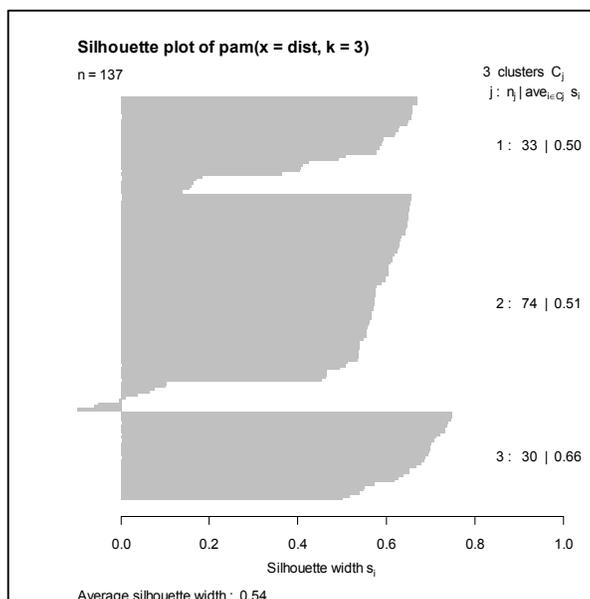


**Figura 9:** distribución de la variable “volvió a otorgarse” en los distintos grupos.

### Análisis de silueta

Al efectuar el análisis de silueta en la estructura encontrada por el algoritmo, obtenemos un *coeficiente de silueta promedio* de 0.54, valor que según la tabla 1 corresponde a un buen modelo.

El software R, nos proporciona un gráfico resumen (figura 10), donde podemos ver que la mayor parte de los individuos fueron asignados correctamente a su grupo (valores de silueta  $s(i) > 0$ ).



**Figura 10:** gráfico de siluetas para la estructura de conglomerados encontrada.

### **Discusión**

Podemos concluir que hemos encontrado una buena estructura de conglomerados formada por 3 grupos, posible de interpretar en el dominio del problema, al aplicar el algoritmo sobre las variables “Días antelación solicitud”, “Canceló con anticipación”, “Edad” y “Horario”.

Describimos los grupos encontrados de la siguiente manera:

- *El grupo 1*, corresponde a turnos cancelados de pacientes niños y jóvenes. Estos pacientes, al ser menores de edad, deben concurrir a la atención médica con sus padres, quienes solicitan el turno con antelación y lo cancelan con anticipación.
- *El grupo 2*, está formado por turnos que fueron cancelados sin anticipación. Corresponden a pacientes en su mayoría adultos que solicitaron el turno con menor antelación que en el resto de los grupos. Son turnos otorgados para la segunda mitad de la tarde (entre las 15 hs y las 18 hs), situación que podría estar relacionada a inconvenientes laborales que provocan la cancelación sin anticipación.

Este grupo es crítico, ya que al analizar la variable “volvió a otorgarse” en el mismo (figura 9), vemos que la mayor parte de estos turnos cancelados no pudieron ser aprovechados por otros pacientes.

- El grupo 3, corresponde a turnos cancelados de pacientes adultos que solicitan su turno con antelación, y a su vez lo cancelan con anticipación. Podemos ver que corresponden a la segunda parte de la tarde.

Del análisis anterior, podemos ver que los turnos que se cancelaron en oftalmología corresponden generalmente a la segunda parte de la tarde, y a pacientes adultos en edad laboral (grupos 2 y 3). Entre los mismos, tenemos aquellos pacientes que son “más responsables” que solicitan su turno con mayor antelación y lo cancelan con anticipación (grupo 3), y aquellos “menos responsables” que realizan la solicitud con menor antelación y cancelan el turno el mismo día de la atención (grupo 2).

La descripción realizada de cada cluster, se corresponde con las características de los medioides (tabla 3), confirmando que los mismos son representativos de los grupos a los que corresponden.

Es interesante observar que mediante la estructura de conglomerados, obtuvimos una distribución distinta de las variables en estudio respecto a su distribución en el conjunto de datos originales.

El centro de salud podría de esta forma mejorar la gestión de atenciones programadas en esta especialidad, identificando aquellos turnos que cumplen con las características del grupo 2, y confirmando mediante un contacto con el paciente, su asistencia al mismo.

Se deja pendiente en este trabajo la experimentación con otras técnicas de conglomerados, a fin de determinar su coincidencia con los grupos aquí obtenidos.

---

#### **Agradecimientos**

- Programa de Becas Bicentenario de Investigación y Posgrado, Universidad Tecnológica Nacional.
- A los directivos del Centro de Salud en el que fue realizado el estudio.
- A la Est. Mónica Grasso†, quien con entusiasmo y dedicación, nos guió en nuestros primeros pasos en el campo de la investigación.

#### **Referencias**

- [1] Peña Daniel, 2002, “Análisis de datos multivariantes”, Editorial Mc Graw Hill. España, ISBN 84-481-3610-1, Cap. 8.
- [2] Hernández Orallo y otros, 2004, “Introducción a la Minería de Datos”, Editorial Pearson Prentice Hall. España. ISBN 84-205-4091-9. Cap. 16.
- [3] Anja Struyf y otros, 1997, “Clustering in an Object-Oriented Environment”, Department of Mathematics and Computer Science, U.I.A.
- [4] J. Demey, 2008, “Diversidad genética en bancos de germoplasma: un enfoque biplot”, Cap1, Universidad de Salamanca, Departamento de estadística.
- [5] Peter Rousseeuw, 1986, “Silhouettes: a graphical aid to the interpretation and validation of cluster analysis”.
- [6] R-project, “The R Project for Statistical Computing”, <http://www.r-project.org>.

#### **Datos de Contacto**

Ing. Juan Miguel Moine, U.T.N. Facultad Regional Rosario, [juanmiguelmoine@gmail.com](mailto:juanmiguelmoine@gmail.com)  
Ing. Javier Matías Landini, U.T.N. Facultad Regional Rosario, [javierlandini@gmail.com](mailto:javierlandini@gmail.com)