UNIVERSIDAD DE SANTIAGO DE CHILE
FACULTAD DE INGENIERÍA
DEPARTAMENTO DE INGENIERÍA INFORMÁTICA

Laboratorio 2 -Agrupamiento K-medias

Integrantes: Sofía Castro
Gonzalo Cuevas

Curso: Análisis de Datos

Sección A-1

Profesor: Max Chacón Pacheco

7 de Septiembre de 2022

Tabla de contenidos

Índice de cuadros

Índice de figuras

1. Introducción

El agrupamiento (clustering) de datos es una técnica no supervisada de la minería de datos que permite encontrar similitudes entre variables clasificándolas en grupos homogéneos, y por otra parte, la heterogeneidad entre los distintos grupos sea elevada, por lo tanto, Larrañaga et al. (2005) comenta que su objetivo es minimizar la variabilidad dentro de los grupos y maximizar la variabilidad entre los distintos grupo . Esta técnica se divide en dos categorías, jerárquico y no jerárquico, lo cuales también se sub-dividen en otras clases según Fidan Kaya Gülagiz (2017). La elaboración de este laboratorio se centrara en agrupamiento no jerárquico, en especifico el método K-Means, el cual es conocido por su simple implementación y eficiencia en ejecución.

La motivación de ejercitar el método K-Means en la base datos, consiste en poder obtener una vista y comprensión de los tipos de clases (grupos) que pueden definir la información pertinente.

El informe esta compuesto por una leve descripción de los métodos implementados en la sección del marco teórico, permitiendo al lector entender los pasos llevados a cabo en los resultados, tales como, la preparación de los datos, conocido como el pre-procesamiento, para luego formular el problema e implementar el agrupamiento, obteniendo los resultados necesarios a fin de analizar sus valores y concluir con respecto a ellos y las dificultades presentadas durante el progreso del laboratorio.

1.1. Objetivos

1. Extraer el conocimiento del problema asignado, mediante el uso del software R, utilizando el algoritmo de clustering.

2. Realizar el análisis respectivo de los resultados obtenidos.

3. Comparar los resultados con lo expuesto en la literatura externa.

4. Analizar por grupo e identificar aquellas características relevantes e inferir conocimiento respecto a ello.

2. Marco Teórico

La implementación de este laboratorio depende de nuestro conocimiento previo y el obtenido en cátedra, por lo que a continuación se explican los contenidos básicos e indispensables para el desarrollo.

2.1. Clustering (agrupamiento)

Técnica de minería de datos que divide un set de datos en diferentes categorías al calcular la similitud entre las variables. Hay un gran grado de similaridad dentro de cada cluster clasificado por el algoritmo y la similitud entre clusters posee un valor bajo, significando mayor distancia entre ellos Cui et al. (2020).

Como se menciono en la introducción, esta técnica se divide en dos categorías. El algoritmo agrupamiento jerárquico (aglomerativo), obtiene como resultado en su gráfico un árbol, llamado dendrograma y este se consigue al implementar los siguientes pasos según Michael et al. (2000): Calcular la similitud entre pares de datos, mediante una matriz de distancia, donde la posición Xij da la distancia entre los datos Xi y Xj, unir los datos o grupos (clusters) más cercanos, actualizar la matriz de distancia, para reflejar las medidas pertinentes entre el nuevo cluster y los datos restantes, para finalmente, repetir pasos 2 y 3 hasta obtener un cluster restante.

Por otro lado, el método no jerárquico (particional) se basa en especificar inicialmente el numero de grupos (clusters) e iterativamente asignar objetos que converjan en grupos. Los métodos heurísticos populares son K-Means y K-Mediods, el primer algoritmo se explica en la sub-siguiente sección mientras que a continuación se explaya sobre el segundo algoritmo mencionado. La estrategia de K-Mediods es que cada cluster sea representado por un sujeto que se encuentre cerca del centro del cluster Madhulatha (2012). Este algoritmo escoge de manera aleatoria k sujetos como centros de un cluster, por consiguiente, comienza una iteración evaluando el intercambio entre un sujeto central i y no central j, escogiendo el que produce un mejor grupo. La función objetivo usada es la suma de las distancias entre cada sujeto y el centro mas cercano. En cada etapa (iteración), se debe seleccionar el sujeto como centro que minimice la función objetivo Reynolds et al. (2004).

2.2. Métodos de segmentación

Se distinguen dos métodos que permiten determinar el número de grupos:

Método del codo: Trabaja mediante el calculo de la la diferencia al cuadrado para los k valores solicitados, con esto, a medida que los valores de k incrementan, el grado de distorsión promedio disminuye, siendo así un mejor numero de grupos aquel que logro un grado de distorsión mayor Syakur et al. (2018).

Método de la silueta: Trabaja calculando la distancia que posee cada elemento con cada grupo, haciendo que este tome valores entre -1, 1. Donde los valores mas altos obtenidos implican una mejor asignación de grupos Řezanková (2018). Este método puede ser utilizado para validar si un elemento se encuentra en el grupo correcto o no.

2.3. K-means

Método propuesto J. MacQueen en 1967, en sus palabras, ïnformalmente, el procedimiento k-means consiste simplemente en comenzar con k grupos donde cada uno consta de un solo punto aleatorio, a partir de ahí, se agregan nuevos punto al grupo cuya media hacia un punto nuevo es la mas cercana. Después de que un punto es incorporado a un grupo, la media de ese grupo se ajusta para tener en cuenta el nuevo punto. Así, en cada etapa, las k-means son, de hecho, las medias de los grupos que representan"MacQueen et al. (1967).

Los k primeros datos son llamados centroides iniciales, es decir, conglomerados con un único elemento, de esta manera, cada uno de los objetos se va asignando al conglomerado con centroide más próximo, con la característica de que al efectuar cada asignación se recalculan las coordenadas del nuevo centroide Larrañaga et al. (2005).

2.4. Distancia de Euclidiana

Es la distancia entre dos puntos en dos, tres o n dimensiones. Cuanto menor sea la distancia euclidiana entre los datos, mayor será la similitud entre ellos.Cui et al. (2020) Definida para n dimensiones, la distancia euclidiana corresponde a la siguiente formula:

$$d = \sqrt{\sum (x_{i1} - x_{i2})^2}, i = 1, 2, 3...n \tag{1}$$

3. Resultados

La base de datos a utilizar, corresponde a la base de datos *allhypo*, creada por Ross Quinlan, perteneciente al instituto Garavan de Sydney, Australia y obtenida por **?**, la cual cuenta con los registros asociados a la enfermedad de la tiroide, y en concreto al hipotiroidismo, contando con un total de 29 variables de las cuales 23 son variables categóricas booleanas correspondientes a características que presentaban los pacientes, mientras que las otras 6 son variables numéricas que describen el nivel de las hormonas tiroideas en cada individuo. En este apartado se presentan las tecnicas implementadas en los datos para poder concluir con el objetivo inicial, visualizar el agrupamiento de los sujetos y descubrir nuevas relaciones por grupo.

3.1. Pre-procesamiento

Cada columna de la base de datos fue estudiada cuidadosamente para identificar cuales realmente aportaban información relevante para el análisis y cuales no, como es el caso de aquellas que indican si se ha medido el nivel de una hormona, en referencia a las columnas TT4 measured, TBG measured, T3 measured, FTI measured y TSH measured, del mismo modo la columa TBG se encuentra compuesta solo por valores NA, por lo que también fue eliminada.

Sin embargo, la base de datos contaba con una gran cantidad de valores NA (no answer), lo cual dificulta la obtención de un estudio correcto sobre la misma, por lo que se se decidió aplicar como criterio la eliminación de datos ausentes para aquellas columnas donde la proporción de NA sea inferior al 5 % Dong and Peng (2013). mientras que en las columnas con una proporción superior al 5 %, se ha utilizado como técnica la imputación única mediante el remplazo de estos valores con la mediana de dicha columna, pues posteriormente la información sera dividida en grupos causando que la incidencia de esta métrica no afecte significativamente el resultado Jadhav et al. (2019). Con esto en mente, para cada columna fue calculada la proporción de Na existente, obteniendo como resultado que esta problemática de valores Na solo se encontraba presente en las 6 variables numéricas y no en las categóricas. Donde en algunos casos, los valores de Na eran superiores al 10 % de las observaciones como

lo fue en las variables TSH y FTI e incluso, llegando a valores preocupantes como el caso de la hormona T3, la cual llegaba a un 20 % de valores Na, los cuales posteriormente serian reemplazados.

Como las variables categóricas existentes solo contenían un valor booleano, se procedió a estudiar cada columna numérica de la base de datos, con el fin de determinar aquellos valores que no se encontraran dentro de los rangos normales de dichas variables. Según describe The National Academy of Clinical Biochemistry los rangos normales para las hormonas asociadas a la tiroides se encuentran en la tabla (nombre tabla) Spencer (2003).

Hormona	**Rango**	**Medida**
TSH	0.4-4.5	nmol/L
TT4	58-160	nmol/L
T4U	0.7–1.2	ug/dL
FTI	77.22 - 141.57	nmol/L
T3	1.2 -2.7	nmol/L

Cuadro 1: Rango de valores normales para las hormonas tiroideas.

Ahora bien, luego de eliminar o cambiar por algún parámetro los valores atípicos asociados a las hormonas, se descubrió que una cantidad significante de información desaprovechada, pues la mayoría de los casos en que las personas estaba vinculadas a algún tipo de hipotiroidismo eran efectivamente aquellos que presentaban una o mas hormona con valores atípicos, siendo así los objetivos de estudio de mayor interés y el motivo por el cual se decidió no restringir estas variables. Por otro lado, si fue restringida la columna asociada a la edad, pues esta presentaba un valor extremadamente atípico al ser incluso cuatro veces mayor a cien años.

Dado que el algoritmo K-Means es para variables continuas y no cualitativas, las variables categóricas fueron convertidas en variables dicotómicas, esto quiere decir que por cada nivel de una variable categórica, fue creada una nueva columna con su nombre y sus atributos equivalentes a 0 o 1 si el sujeto se atribuye o no la categoría, un ejemplo explicativo se puede visualizar en el cuadro 2.

Sujeto	Animal
1	Dog
2	Cat
3	Frog

Sujeto	Cat	Dog	Frog
1	0	1	0
2	1	0	0
3	0	0	1

Cuadro 2: Conversión de variable categórica.

A pesar de modificar las variables categóricas como numéricas, la distancia euclidiana es considerada sensible a los cambios en las diferencias entre los datos, por lo que si existe una distancia mayor entre dos sujetos, la relevancia seria determinada por el número mayor, además, el algoritmo K-Means es incapaz de manejar datos ruidosos y valores desconocidos. Las técnicas de pre-procesamiento, normalización de datos en este caso, son a menudo aplicado a los conjuntos de datos para hacerlos más limpios, consistente y libre de ruido, estandarizando los atributos de la base de datos, retribuyendo una significancia equitativa a cada dato, para que los sujetos redundantes y ruidosos puedan ser eliminados, elevando la precisión de los resultados Virmani et al. (2015). Por lo tanto, se opto por realizar la normalización de las variables mediante la función *scale*, pues la escala de mediciones entre las variables difería en exceso, en especial luego de que se decidió contemplar los datos atípicos. Con esto, se logro mejorar la eficiencia y eficacia del algoritmo, permitiendo una mayor representación de la base de datos en los clusters. Scale utiliza la ecuación 2 para normalizar los datos de forma predeterminada, donde x representa el valor real del sujeto, $\overline{x}$ el promedio de los datos y sd la desviación estándar de los valores.

$$xScaled = \frac{x - \overline{x}}{sd} \tag{2}$$

3.2. Obtención del Clúster

Luego del pre-procesamiento de los datos y la elección de las variables a considerar, conforme a la importancia de cada una, se procede a evaluar el numero de clusters adecuado según la entrada dada, por medio del método de la silueta y del codo explicados en el capitulo anterior. Las entradas corresponden a 5 grupos distintos de pruebas, que contienen:

1. Datos-0: Todas las variables de la base de datos.
2. Datos-1: Solo variables numéricas (hormonas), tales como, TT4, T3, TSH, FTI y T4U.
3. Datos-2: Las variables de Datos-1 y las cuatro clases negative, primary, secondary y compensated.
4. Datos-3: Los valores de Datos-2, la columna female, male y edad.
5. Datos-4: Este set incluye la hormona TSH y las clases negative, primary, secondary y compensated.

Los valores de K obtenidos de acuerdo a cada método se pueden visualizar en el cuadro 3.

Grupo	**Método Codo**	**Método Silueta**
Datos-0	2 ó 6	12
Datos-1	2	4
Datos-2	3	3
Datos-3	4	6
Datos-4	4	6

Cuadro 3: Valores de K conforme al método utilizado.

Por consiguiente, cada dato de entrada fue ingresado al algoritmo de K-means junto a el numero de grupos recomendado, obteniendo los resultados dispuestos en las figuras 1, 2, 3 y 4.

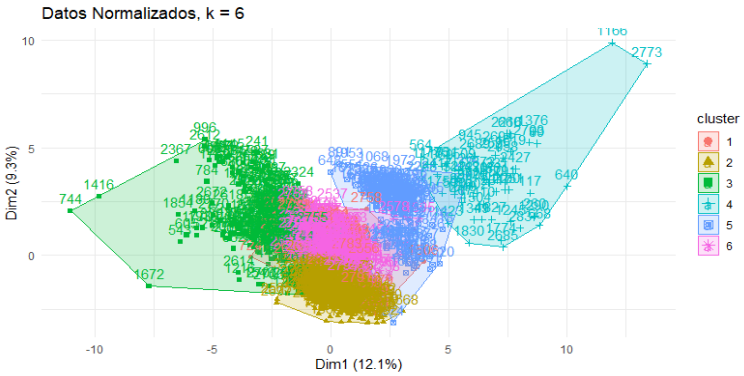

Figura 1: Datos-0, k = 6

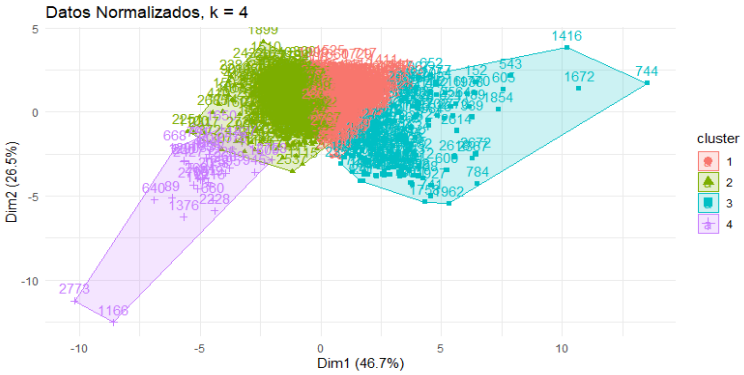

Figura 2: Datos-1, k = 4

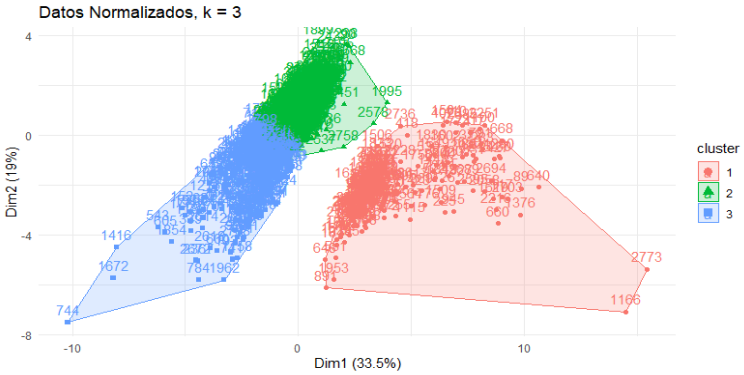

Figura 3: Datos-2, k = 3

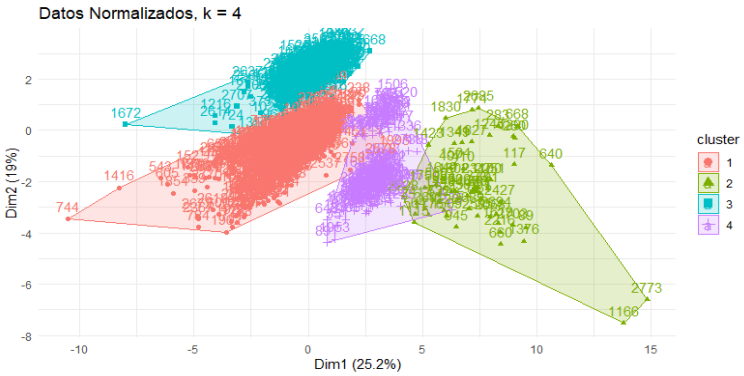

Figura 4: Datos-3, k = 4

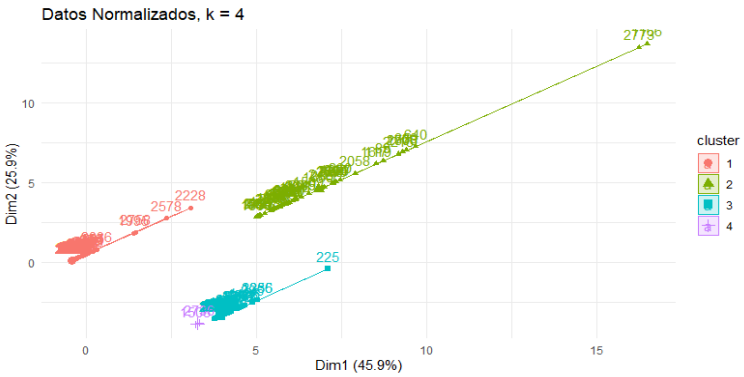

Figura 5: Datos-4, k = 4

Sin embargo, los resultados expuestos solo corresponden a lo que se consideró como la mejor agrupación de los datos, pues para la mayoría de las entradas, al aumentar la cantidad de grupos, estos visualmente eran solapados unos por otros, causando que no se tuviera una clasificación concreta para la información. Esto puede ser visible en la entrada de los datos-0, descrita por la figura 1, donde existen grupos completamente solapados por otros causando que no exista distinción entre ellos. Con esto, a través de la experiencia, se concluyo que una mejor representación para los grupos pertenecientes a la entrada de datos-0 se obtiene cuando el valor de k es igual a 5, es decir, cuando no toma uno de los recomendados por los método, lo cual se puede ver en la figura 6, donde si existe distinción de cada uno de los grupos formados.

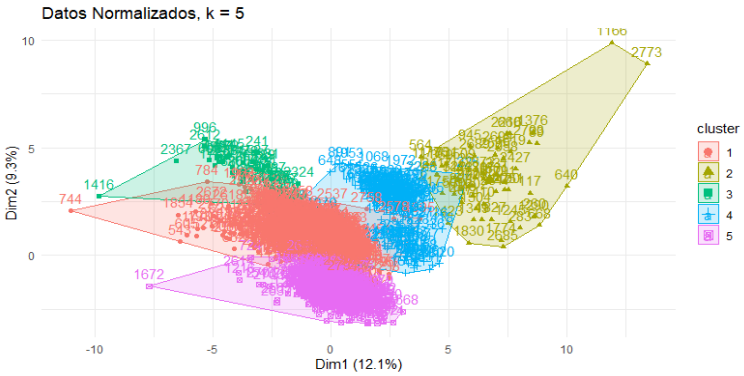

Figura 6: Datos-0, k = 5

3.3. Análisis de los Cluster

Si bien las figuras obtenidas para la mayoría de las entradas muestran grupos definidos, solo serán analizadas las figuras 4 y 6 correspondientes a los datos-3 y datos-0 pues ambos trabajan con el mayor numero de variables en conjunto y además, uno contiene una mayor representatividad en los ejes de sus dimensiones en comparación del otro, como es el caso de la información expuesta en la figura 6.

Para la figura 6 que contiene la información asociada a los datos-0 , se distinguen claramente los cinco grupos. Al analizar cuales son los individuos involucrados en cada cluster, se permite comprender la clasificación del cluster 1, pues este se encuentra compuesto solamente por mujeres, las cuales no tienen hipotiroidismo, es decir, presentan un resultado negativo, el cluster 2 involucra a todas las personas con hipotiroidismo primario, así mismo, el cluster 3 se encuentra contiene todas las mujeres embarazadas que no tienen hipotiroidismo, mientras que el cluster 4 contempla a todas las personas con hipotiroidismo secundario y compensado, por ultimo el cluster 5 representa a los hombres sin hipotiroidismo. De esta manera, a grandes rasgos, la dimensión 1 de la figura se encuentra asociada al sexo del individuo, mientras que la dimensión 2 representa el estado de el, es decir, si presenta la enfermedad o no. Donde los cuadrantes I y II están asociados a los individuos con sexo femenino mientras los cuadrantes III y IV a los individuos con sexo masculino, también los cuadrantes I y IV corresponden a las personas que padecen alguna clase de hipotiroidismo y los otros dos cuadrantes a aquellas que no.

Considerando los resultados mencionados, al asociar el análisis obtenido de los gráficos con la información entregada por la NACB respecto a las hormonas tiroideas, se esperaría que las personas con niveles altos de TSH padezcan alguna clase de hipotiroidismo, al igual que quienes contengan niveles bajos de la hormona T4 o niveles bajos de FTI Spencer (2003).

De los resultados obtenidos, este supuesto se cumple, ya que los niveles de TSH aumentan de izquierda a derecha, mientras que los niveles de T4 y FTI aumentan de derecha a izquierda, pues como se describió anteriormente, en el lado izquierdo de la figura 6 se encuentran las personas que no padecen hipotiroidismo, mientras que en el lado derecho están quienes si lo padecen. Asimismo, aquellos valores que se encuentran muy alejados del grupo que se les asigno, corresponden a alguno de los valores máximos que presentan estas variables numéricas, por lo que aquellas personas que se encuentran en los cuadrantes II y III y a su vez son muy distantes de sus grupos asignados, posiblemente padezcan hipertiroidismo al no encontrarse cercanos a los valores normales.

Las mujeres embarazadas son una de las poblaciones mas propensas a desordenes hormonales, según la NACB, estas tienden a aumentar sus niveles de TT4 durante el periodo de gestación, siendo esta una de las causales de abortos o el desarrollo de bocio Spencer (2003). Al analizar la figura 6 es posible notar que el grupo de las mujeres embarazadas con un resultado negativo se encuentra ligeramente mas a la izquierda que el centro de la conglomeración de mujeres no embarazadas con resultado negativo. Esto nos permite aseverar que sus niveles de TT4 pueden ser ligeramente inferiores en comparación del otro grupo y por ende, pueden ser mas propensas a desarrollar hipertiroidismo en lugar de hipotiroidismo

Si bien los resultados son acorde a lo estipulado, la representatividad que posee cada dimensión obtenida para los datos-0 es considerada baja, al poseer un porcentaje inferior al 20 % para cada dimensión apreciable en los ejes de la figura 6. Esto puede deberse a la cantidad de variables a analizar, pues se debe intentar agrupar cada uno contemplando todos estos valores, lo que indudablemente dificulta al proceso de asignación.

Es por ello que se trabajo con distintas entradas de datos, de las cuales, se decidió elegir la entrada vinculada a los datos-3, pues esta contiene una menor cantidad de variables, pero a su vez, contempla las necesarias para poder realizar un análisis similar con la condición

de que aumenta la representatividad de las dimensiones, apreciable en los ejes de la figura 4.

En este caso (figura 4), el cluster 1 representa a todas las mujeres con un resultado negativo de hipotiroidismo, mientras que el cluster 2 posee tanto hombres como mujeres con hipotiroidismo primario, por otra parte, el cluster 3 concentra a todos los hombres con un diagnostico negativo en hipotiroidismo y finalmente el cluster 4 contiene hombres y mujeres que padecen de hipotiroidismo compensado o secundario. Por lo tanto, a diferencia del resultado con los datos-0 en este no se pueden diferenciar las mujeres embarazadas de los distintos clusters, siendo esta la unica diferencia, pues lo ejes se distribuyen de la misma forma, donde el nivel de TSH disminuye de derecha a izquierda y las hormonas T3 y TT4 aumentan a la derecha.

Lo llamativo de los resultados es el fenómeno que ocurre con hipotiroidismo compensado y secundario, pues ambos convergen en el mismo cluster, mientras que el diagnostico de hipotiroidismo primario se separa completamente, al igual que las personas que no padecen de hipotiroidismo. El hipotiroidismo secundario se debe a que la glándula pituitaria no libera un nivel necesario de TSH, mientras que en el hipotiroidismo compensado la misma glándula libera en exceso niveles TSH. Por lo tanto, permite confirmar lo estipulado por Benvenga et al. (2018) donde afirma que ÇH incluye hipotiroidismo causado por alteración de la secreción de TSH (hipotiroidismo secundario)"donde CH alude a hipotiroidismo central también conocido como clínico o compensado, por lo tanto, se puede deducir que este es causado o proviene del hipotiroidismo secundario, atribuyendo aceptación a la unión de estas dos clases en el cluster 4.

Cabe destacar que los otros datos, tanto datos-1, datos-2 y datos-3 no tuvieron un análisis debido a la redundancia de estos, ya que siempre recaen en la separación de un diagnostico negativo y positivo de hipotiroidismo, al igual que con los niveles de la hormona TSH, siendo este el motivo del análisis de dos gráficos, pues ambos brindan mejor interpretación y representación de la base de datos.

4. Conclusión

A partir de la investigación, fue posible comprender visualmente como se comportan las variables asociadas a los registros de los individuos involucrados con el hipotiroidismo, logrando generar grupos definidos que contemplen la información dispuesta y que permiten analizar y comparar lo obtenido con otros resultados asociados al padecimiento descrito.

Respecto al procedimiento realizado, si bien fue posible alcanzar los objetivos estipulados, el camino hacia este no estuvo exento de problemas, en especial asociadas a la escala con la que se trabajo y la determinación del numero de grupos. Esto porque en un principio, se considero trabajar con los datos sin normalizar, lo que conllevo a que los grupos no se definieran correctamente y existiera un nivel de solapamiento entre grupos elevado, así mismo, incluso luego de la normalización, aplicando los criterios descritos para encontrar un valor óptimo de grupos, no era posible encontrar un valor especifico que lograra representar la información correctamente y que a su vez permitiese lograr un análisis efectivo de los datos, por lo que fueron realizadas varias pruebas agregando o quitando datos y variables o modificando los valores de estas hasta poder alcanzar los resultados descritos en este informe. Siendo también una de las principales complicaciones de esta base de datos el excesivo numero de valores NA existentes, pues si estos no eran contemplados o su valor se reemplazaba por alguno que no fuera considerado dentro de esta instancia, la clasificación en grupos resultante no era apta para inferir correctamente como estos se comportaban.

Sin embargo, aunque el proceso fue engorroso se puede afirmar el cumplimiento de los objetivos propuestos inicialmente, pues se pudo implementar las técnicas que conllevan al algoritmo de clustering, al igual que el análisis de cada grupo y resultado importante obtenido. También la motivación estipulada es fructosa, puesto que hemos descubierto un nuevo conocimiento en relación a la base de datos, donde el hipotiroidismo compensado y secundario poseen una relación fuerte, debido a las alteraciones que afectan a la glándula pituitaria.

Finalmente, se espera en futuras evaluaciones lograr inferir y modelar otras características que no fueron contempladas en esta instancia y que son conocidas por la comunidad científica, por lo que el objetivo se concentrara en hacer un descubrimiento entre las varia-

bles que pueda ayudar a las personas que padecen de hipotiroidismo o tienen secuelas de esta enfermedad.

Bibliografía

Benvenga, S., Klose, M., Vita, R., and Feldt-Rasmussend, U. (2018). Less known aspects of central hypothyroidism: Part 1 – acquired etiologies. *Journal of clinical translational endocrinology*, 14:25–33.

Cui, M. et al. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1(1):5–8.

Dong, Y. and Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus*, 2(1):1–17.

Fidan Kaya Gülagiz, S. S. (2017). Comparison of hierarchical and non-hierarchical clustering algorithms. In *Clustering*, volume 9, pages 6–14. International Journal of Computer Engineering and Information Technology, Dubai.

Jadhav, A., Pramod, D., and Ramanathan, K. (2019). Comparison of performance of data imputation methods for numeric dataset. *Applied Artificial Intelligence*, 33(10):913–933.

Larrañaga, P. et al. (2005). Clustering. In *Clustering*, pages 1–8. Departamento de Ciencias de la Computación e Inteligencia Artificial, Universidad del País Vasco,País Vasco.

MacQueen, J. et al. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA.

Madhulatha, T. S. (2012). An overview on clustering methods. *CoRR*, abs/1205.1117.

Michael, S., George, K., and Vipin, K. (2000). A comparison of document clustering techniques. *Department of Computer Science and Egineering, University of Minnesota*.

Reynolds, A. P., Richards, G., and Rayward-Smith, V. J. (2004). The application of k-medoids and pam to the clustering of rules. In Yang, Z. R., Yin, H., and Everson, R. M., editors, *Intelligent Data Engineering and Automated Learning – IDEAL 2004*, pages 173–178, Berlin, Heidelberg. Springer Berlin Heidelberg.

Řezanková, H. (2018). Different approaches to the silhouette coefficient calculation in cluster evaluation. In *21st International Scientific Conference AMSE Applications of Mathematics and Statistics in Economics*, pages 1–10.

Spencer, C. A. (2003). Thyroid testing for the new millenium. *Thyroid*, 13(1):2–2.

Syakur, M., Khotimah, B., Rochman, E., and Satoto, B. D. (2018). Integration k-means clustering method and elbow method for identification of the best customer profile cluster. In *IOP conference series: materials science and engineering*, volume 336, page 012017. IOP Publishing.

Virmani, D., Shweta, T., and Malhotra, G. (2015). Normalization based K means clustering algorithm. *CoRR*, abs/1503.00900.

Anexos