

Evidencia 2

*El Aprendizaje No Supervisado como Herramienta de Predicción de Enfermedades en el
Área de la Salud*

Proyecto de Aprendizaje No Supervisado

Modelación del Aprendizaje con Inteligencia Artificial

TC2034.101

Gpo. 101

Raúl Correa Ocañas	A01722401
Sebastián Miramontes Soto	A01285296
Ericka Sofia Rodriguez Sanchez	A01571463

Docente:

Dr. Santiago Enrique Conant Pablos

Monterrey, Nuevo León, 10 de marzo de 2024

1. Introducción.....3

2. Marco Teórico.....	3
2.1 Aprendizaje No Supervisado en Salud.....	3
2.2 Descripción del Dataset.....	4
2.3 Preprocesamiento.....	5
2.4 Análisis Exploratorio con Gráficos.....	5
3. Metodología.....	7
3.1 Esqueleto del Código.....	7
3.2 Modelo 1: K-Means.....	9
3.3 Modelo 2: DBSCAN.....	10
4. Resultados.....	11
4.1 Comparativa de Métodos.....	11
4.3 Análisis, ventajas, desventajas, posibles mejoras.....	11
5. Conclusiones.....	12
5.1 Conclusión - Sebastián Miramontes Soto.....	12
5.2 Conclusión - Raúl Correa Ocañas.....	13
5.3 Conclusión - Ericka Sofia Rodriguez Sanchez.....	13
6. Referencias.....	13

1. Introducción

La humanidad siempre se encuentra en constante evolución, y con los avances tecnológicos y científicos se busca alcanzar una vida más cómoda, segura y extensa para el ser humano. La medicina y el área de salud son ramas que han logrado avances significativos con el paso del tiempo, encontrando cada vez mejores maneras de diagnosticar y curar enfermedades. Por otro lado, la tecnología y los conocimientos han crecido a gran medida, de una manera en la que se están combinando ramas como la estadística y la informática, para analizar datos de forma que las computadoras pueden aprender de ellos y realizar modelos efectivos capaces de clasificar y predecir etiquetas. Es entonces cuando aparece este reto de combinar ambas disciplinas para lograr en conjunto un mayor avance, y se deben de aprovechar los recursos en su totalidad para obtener resultados de calidad.

En el siguiente proyecto se llevará a cabo una investigación, mediante la cuál a través de una base de datos médica, se buscará encontrar los mejores modelos para clasificar enfermedades de acuerdo a síntomas presentados. Todo esto utilizando técnicas de aprendizaje no supervisado, para entrenar modelos, analizar su desempeño, y en base a ello tomar los conjuntos más adecuados para diagnosticar enfermedades.

2. Marco Teórico

2.1 Aprendizaje No Supervisado en Salud

Actualmente, en la medicina de precisión, el aprendizaje no supervisado está siendo utilizado para llegar a una mejor comprensión de los mecanismos de las enfermedades. Es gracias a esta información que se puede encontrar cuáles son las mejores rutas de tratamiento que se pueden tomar para estos casos. (Spencer, 2020)

Además, se ha mostrado como el aprendizaje no supervisado ha sido aplicado mediante la clusterización para agrupar y realizar visualizaciones de datasets de células. Un

ejemplo puntual es un análisis que se realizó al conjunto de datos *Breast Cancer Winsconsin Data Set*. El documento presenta los resultados de la aplicación de métodos como agrupamiento jerárquico, t-SNE (incrustación de vecinos estocásticos distribuidos en t), k-means, y análisis de componentes principales. (Rueda, 2019)

En cuanto a la segmentación de poblaciones, es de utilidad para separar a los pacientes en grupos que comparten características clínicas o genéticas. De esta manera se pueden descubrir patrones o relaciones que sean útiles para brindar una atención personalizada, identificar riesgos o comparar como ha sido la respuesta a tratamientos específicos a pacientes que presentan características similares. Adicionalmente, pueden ser utilizados para detectar anomalías en datos de salud o características genómicas para la detección temprana de enfermedades o condiciones.

2.2 Descripción del Dataset

Previo a comenzar con el proceso de presentar una solución de aprendizaje automatizado supervisado en un contexto médico, es crucial plantear qué clase de predictores son relevantes para el diagnóstico de enfermedades. Bien que saber únicamente los síntomas de un paciente no siempre es suficiente para realizar un diagnóstico confiable y correcto, proponer un modelo que no requiera considerar todos los síntomas y que logre clasificar adecuadamente una enfermedad puede reducir el número de preguntas hechas para realizar un diagnóstico. La base de datos *Disease Prediction Data*, tomada de [Kaggle](#), nos muestra un mapeo de cómo se distribuyen los síntomas en 42 enfermedades, con lo que se pueden aplicar conocimientos previamente adquiridos en el campo de la salud y ciencias.

Los conjuntos de datos fueron seleccionados con la intención de identificar y diagnosticar enfermedades a través de 132 síntomas en donde se incluían algunos muy generales como dolor de estómago, articulaciones o músculos, vómito, tos y fatiga, y algunas

más específicas. Los datos se inicializan cargando los datasets en forma de 2 diferentes conjuntos diferentes, una de *training* y otra de *testing*, los cuales previamente habían sido aplicados a un problema de clasificación. Se hace una unión entre las bases de datos para trabajar con un solo data frame *data*. El código fue implementado sobre el dataset de *data* para encontrar los mejores parámetros, el cual después se utiliza este mismo para evaluar entre los mejores modelos. En esta ocasión no se realiza un corte de *training* y *testing*, ya que los problemas de aprendizaje no supervisado no pueden ser evaluados de la misma manera que los métodos de aprendizaje supervisado.

2.3 Preprocesamiento

Para el preprocesamiento de los datos se inició con la eliminación de los datos que no aportan tanta información a la toma de decisiones en los modelos. Una variable identificada y descartada inmediatamente fue “Unnamed: 133”, debido a que la totalidad de esta columna resultó estar completa de NaNs. Al tener alrededor de 130 variables restantes, es evidente la posibilidad de que no todas aportan el mismo grado de información, por lo cual pueden ser redundantes. Pero al tener que implementar un método de agrupación se decidió utilizar el resto de las variables para que se pudieran observar todas los escenarios posibles, y que el algoritmo encontrará la mejor forma de dividirlos para que se obtengan los resultados totales. Sin tener más conocimiento en cuanto a las relaciones entre variables del set de datos, el realizar un análisis exploratorio de datos puede proporcionar un mejor panorama de las variables.

2.4 Análisis Exploratorio con Gráficos

Para el análisis exploratorio de los datos, se creó una tabla mostrando métricas estadísticas de la base de datos. Al notar que todas las variables predictoras tienen un rango

de $[0,1]$ y hacen referencia a algún síntoma, se infiere que su comportamiento es de estilo booleano. Esto indica que la base de datos ha sido codificada y transformada al estilo *One-Hot Encoding*.

Adicionalmente, se realizaron distintas gráficas para comprender mejor la naturaleza de los datos. Entre las generadas, se decidió mantener las más relevantes. La primera de éstas fue la gráfica de distribución de enfermedades, en la cuál se observan 42 distintas, todas relacionadas con ciertos de los síntomas utilizados como variables predictoras más adelante, se puede observar que existe una varianza muy cercana a cero, la cuál se calcula con la ecuación $Var[X] = p(1 - p)$.

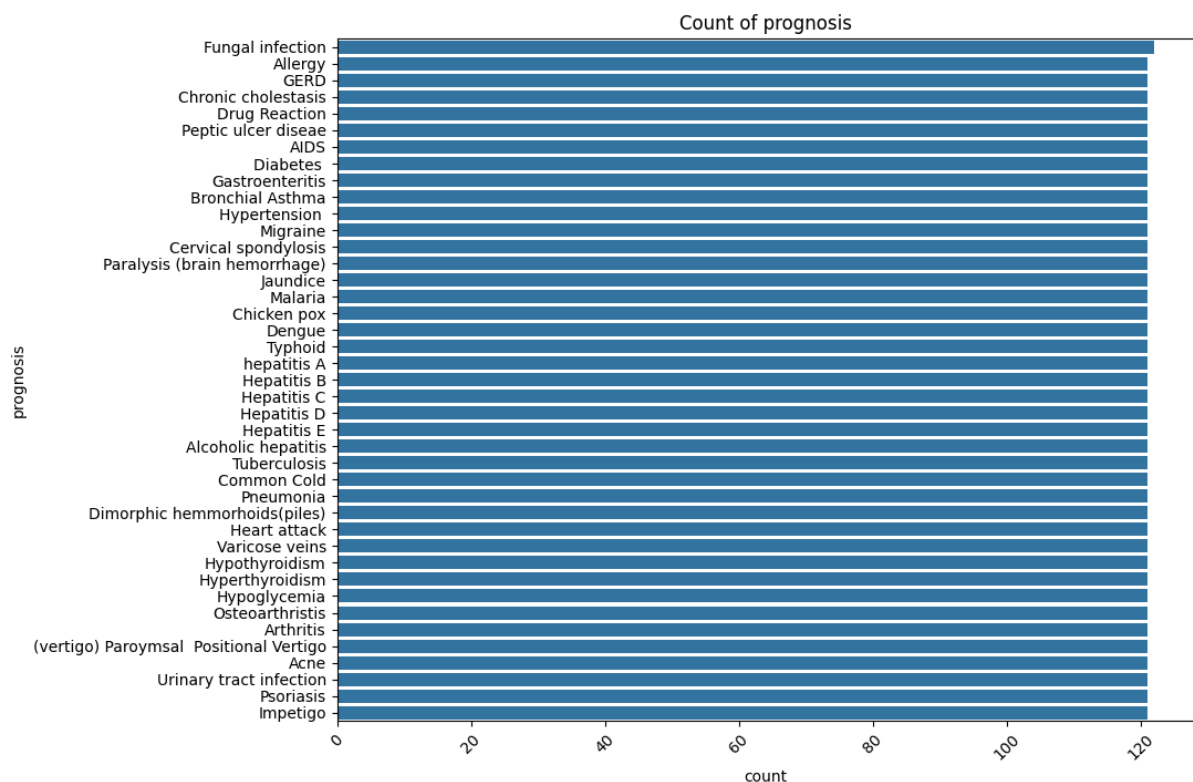


Imagen 1. Distribución de enfermedades.

Si se observa a detalle el gráfico, se puede observar que la única variable que tiene un número de casos un poco más alto que el resto, pero al final todos tienen poco más de 120, por lo cuál nos son bastante útiles para el trabajo.

El segundo gráfico que se analizó fue un mapa de calor en cuanto a la correlación entre síntomas. De forma visual, se pudo observar qué tan fuerte era la relación entre las variables. Las zonas rojas indican una relación directa fuerte, las blancas eran neutrales y las azules indican una relación inversamente fuerte. Se observó que en la zona del centro se encontraban cuadros rojos más grandes, mientras que otros se podían observar en zonas aleatorias. Además, se pudo notar que la mayoría del mapa tendía a ser neutral, descartando la diagonal que comparaba una variable consigo misma.

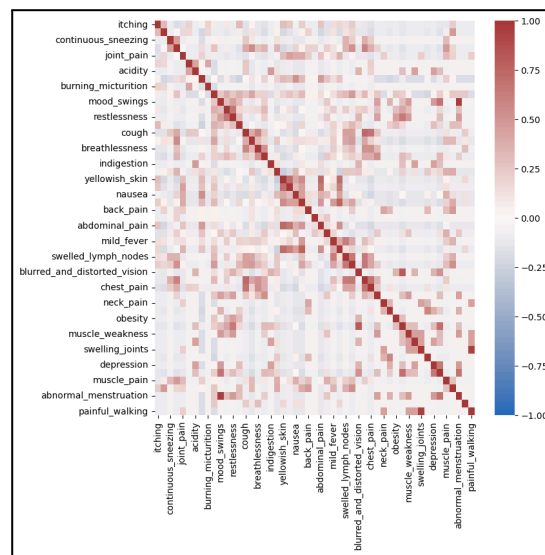


Imagen 2. Mapa de Calor de Correlación entre Síntomas

3. Metodología

3.1 Esqueleto del Código

Para la elaboración de este código se decidió utilizar un orden lógico similar al de la evidencia anterior. La primera etapa que se elaboró fue la importación de librerías, en este caso se ingresaron al código 13 librerías distintas, donde las más relevantes fueron “Pandas” para la lectura de las bases de datos, “Seaborn” y “Matplotlib” cuyas funciones son las creaciones de distintos gráficos para el análisis visual de las variables y los resultados, y la

librería “SkLearn” la cuál fue utilizada para la implementación de los códigos de los métodos de KMeans y DBSCAN.

La segunda etapa del código fue el preprocesamiento de los datos, el cuál consistió en la lectura de las bases de datos a utilizar, principalmente en esta etapa se eliminaron las columnas que no son relevantes para la evidencia. Lo primero que se realizó fue la eliminación de las variables NaNs de la tabla de valores, utilizando la función “data.dropna(axis=1)”, con esto también determinamos con la columna Unnamed: 133 era completamente irrelevante ya que la totalidad de los datos que contenía eran de este tipo de valores. Después de esto se eliminó la columna de Prognosis, esto debido a que por su tipo de respuesta puede llegar cierto sesgo al momento de realizar los métodos por clustering.

Para continuar con el código se realizó la fase de Análisis de Exploratorio de Datos para lo que se realizaron 2 gráficos y una tabla, con la única función de seguir analizando los datos y observar qué tan útiles eran. Lo primero que se hizo en está sección fue utilizar la función describe para obtener como se distribuyen los datos entre sí, esto ayudó a verificar si todos los datos contaban con la misma cantidad de respuestas, lo cuál era verdadero y tenían 4962 cada una, por otro lado también se pudo notar que que en todos los casos la media era menor a 0.2, y al tener solo variables binarias, se puede deducir que en más del 80% de los casos se obtuvieron respuestas negativas. Después de esto se realizaron el diagrama de Distribución de enfermedades y el de Mapa de Calor de Correlación entre Síntomas previamente mencionados, cuya función fue encontrar la distribución y correlación entre la variables respectivamente.

El cuarto paso realizado para el código fue la modelación del Método de K means Clustering, en donde se inicializan valores como best_ari, best_silhouette, best_n_components, best_eps , best_min_samples y best_clusters, que serán utilizados para que mediante un proceso de ciclos se encuentre la mejor solución posible para el reto,

buscando que no baje de 0.65 el silhouette y que la cantidad de clusters no supere la cantidad de enfermedades que se tienen en la base de datos, todo esto se aplica al conjunto de datos con las características predictoras transformadas. Después de esto, en base al resultado anterior se generaba una gráfica de dispersión para identificar a cada uno de los puntos divididos en los clusters generados.

El siguiente paso debe de ser bastante similar al cuarto, con la única diferencia de que ahora se debe de buscar encontrar la mejor solución utilizando el método DBSCAN, se mantienen los mismos valores y la misma estructura, y al final se genera la misma gráfica pero en esta ocasión sobre los resultados obtenidos en este segundo modelo.

El sexto y último paso fue el de evaluación/comparación de resultados en donde se utilizaron los parámetros generados en los dos pasos anteriores, pero ahora evaluados sobre los conjuntos de prueba, para así verificar el funcionamiento de los modelos en una base más grande, y ya una vez realizado esto, se comparan los resultados obtenidos para las dos métricas y así decidir cuál modelo fue más efectivo de los dos.

3.2 Modelo 1: K-Means

K-Means es un algoritmo que divide en k clusters un conjunto de datos. Estos grupos se forman de manera que los puntos en un mismo cluster tengan mayores similitudes entre ellos que con puntos de los otros clusters. (Ramírez, 2023) El proceso que sigue este método es el siguiente: primero se escoge el número k de clusters en el que se dividirán los datos; después se escogen k puntos; para saber en qué cluster estará cada punto, se calcula la distancia entre el punto y el centroide y se asigna al más cercano; posteriormente se recalculan los centroides; y así se repiten los dos últimos pasos hasta que los centroides ya no tengan un cambio significativo o se alcancen ciertas iteraciones. Una desventaja de este método es que se tiene que elegir k a priori, sin embargo, para esto existen maneras de

encontrar valores óptimos como el método del codo y técnicas de validación cruzada. Es importante considerar que este método es sensible a la elección inicial de centroides, no se puede verificar la respuesta pues no hay una correcta, es sensible a outliers y no es adecuado cuando los clusters son de distintos tamaños o densidades. Una de las razones por las que se eligió este método es por sus implementaciones en el área de salud, principalmente se ha podido utilizar para la agrupación de pacientes con cierta característica en común, como lo pueden ser los tumores o alguna enfermedad de gravedad (Hernandez, s.f). También se ha tenido uso en procesos científicos más complejos, como lo pueden ser la división de tejidos celulares, para aumentar su segmentación y obtener resultados para uso científico (Rueda, 2019).

3.3 Modelo 2: DBSCAN

El modelo Agrupamiento espacial basado en densidad de aplicaciones con ruido o por sus siglas en inglés DBSCAN, es un algoritmo no supervisado sencillo que define los clústeres mediante su densidad, donde se consideran conceptos claves como la distancia euclidiana entre los puntos (DataScientest, 2023). Este método depende principalmente de 2 parámetros importantes, el primero es el ϵ que especifica qué tan cerca deben de estar los puntos del cluster al que pertenecen, y el segundo son los puntos mínimos que son aquellos que le dan cierta densidad a los clusters, también una de las ventajas es que en este método se pueden utilizar clusters de cualquier forma, aparte de que es bueno manejando clusters de alta densidad, y no necesariamente circulares, por lo que esto podría ser útil para el caso debido a la gran diversidad de valores y datos (Gonzalez, 2021). Este método fue elegido principalmente por su practicidad, versatilidad y eficiencia en casos similares al aplicado, y por lo mismo se asumió que podría ser aplicable para casos específicos en ciencias o salud, cosa que no es para nada errónea, ya que este método en combinación con

otros logra agrupar a los pacientes con síntomas similares en un conjunto de grupos iniciales, uso que se le puede dar para permitir que el personal médico de algunas instituciones tome las medidas adecuadas en función del estado de los pacientes actuales (Al-Shammari et al., 2019).

4. Resultados

Antes de realizar el análisis y comparativa de resultados es importante explicar las métricas utilizadas para evaluar los métodos. En esta ocasión se utilizaron las métricas de ari y silhouette score, ambas ayudan a verificar qué tan fuerte y preciso es el modelo desarrollado. El silhouette lo que mide es la relación entre un objeto y su cluster y se indica que tiene que tener arriba del 70% para que se considere un modelo fuerte y arriba de 50% para que se considere decente, mientras que en el caso del ari lo que verifica es la relación entre clusters, con la diferencia en que en este caso un porcentaje del 60% para arriba ya es considerado bueno. Para ambos casos el 100% es considerado excelente pero es muy difícil encontrar un caso en que suceda, y más en casos de muchas variables.

4.1 Comparativa de Métodos

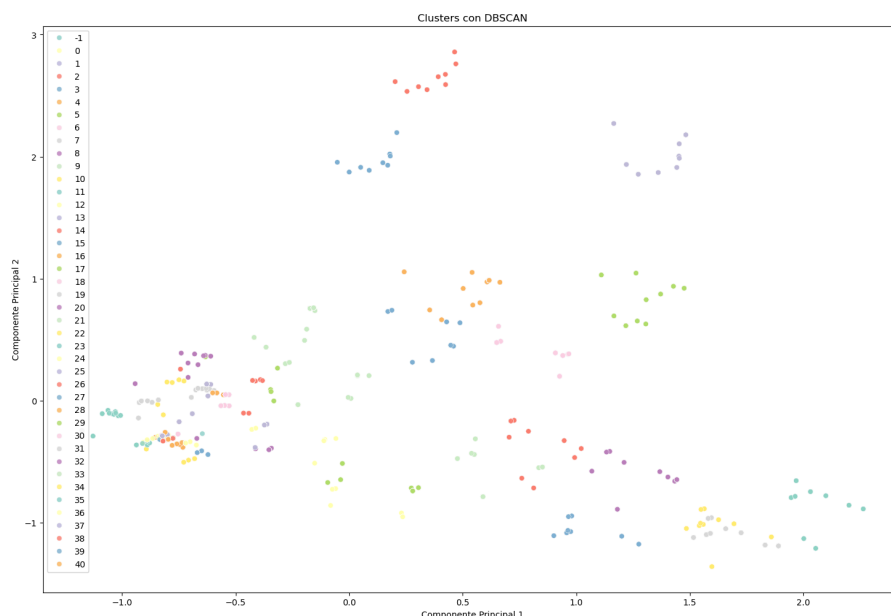
Teniendo ambos modelos listos, se hace una comparación de resultados con los mejores hiperparametros. Notando que al tener entre 25-35 componentes principales, se representa entre el 85% y 95% de la varianza de los datos. Por lo tanto, la búsqueda de hiperparámetros para los modelos se limita a este intervalo.

Para el caso de K Means, se hizo una búsqueda de cuadrícula entre diferentes números de componentes principales (obtenidos al realizar un PCA) y el número de clusters. Esta búsqueda resultó en la combinación de 35 componentes principales y 39 clusters. En el caso del modelo DBSCAN, se identificó que el mejor modelo utiliza 34 componentes principales, un parámetro de epsilon de 1.0 y un número mínimo de muestra de 5.

Creando una instancia de K Means con los parámetros encontrados y otra instancia de DBSCAN con sus parámetros correspondientes, se calcula el ARI y coeficiente de Silueta para cada modelo.

Modelo	Adjusted Random Index	Coef. de Silueta
DBSCAN	0.9997917520202194	0.8281497781045049
K Means	0.9519193444905224	0.8045194012516956

Con los resultados obtenidos, es posible identificar que tanto DBSCAN como K-Means tuvieron la capacidad de tener fuertes indicadores de clustering. DBSCAN se acercó mucho a tener una métrica muy cercana a un ARI perfecto, por lo que se implica que el modelo tuvo la capacidad de identificar adecuadamente maneras de hacer clusters muy similares a la enfermedad correspondiente. Aunque bien se esperaba buscar un método de clustering que brindara más información que las etiquetas que corresponden a estos síntomas, esto indica que el modelo pudo identificar la forma de cómo agrupar los síntomas para corresponder a las etiquetas reales. Adicionalmente, la métrica del coeficiente de silueta también es alta. Si bien tener un coeficiente mayor a 0.7 indica una fuerte indicación de clustering adecuado, tener una métrica incluso mayor es indicador de que el clustering realizado probablemente sea adecuado.



Clusters Formados usando DBSCAN y dos componentes principales

Con los resultados obtenidos, es posible identificar que K-Means también demostró un rendimiento notable en términos de clustering. Aunque su Adjusted Rand Index y Silhouette Coefficient son ligeramente menores a los de DBSCAN, aún muestran una capacidad significativa para agrupar los datos. Con un Adjusted Rand Index de 0.952 y un Silhouette Coefficient de 0.805, el algoritmo de K-Means logró identificar patrones relevantes en los datos y asignar adecuadamente las observaciones a los clústeres correspondientes. Esta capacidad de agrupación efectiva sugiere que el modelo pudo capturar la estructura de los datos y proporcionar agrupaciones útiles en el contexto de las enfermedades estudiadas. Además, el coeficiente de silueta, aunque ligeramente menor que el de DBSCAN, aún indica un clustering adecuado, con un valor superior a 0.7, lo que respalda la efectividad del enfoque de K-Means para este conjunto de datos.



Clusters Formados usando K Means y dos componentes principales

Es importante mencionar que el método de DBSCAN pudo realizar su búsqueda de hiper parámetros en una menor cantidad de tiempo, terminando su ejecución en 15 minutos. Por otro lado, la búsqueda de hiper parámetros de K Means fue de 25 minutos. Esto nos indica que para este conjunto de datos en específico, no solo tuvo mejores resultados, sino que también es más rápido implementar un modelo de DBSCAN para encontrar clusters de enfermedades. Por este motivo, se decide que este método es mejor para la problemática a resolver.

4.3 Análisis, ventajas, desventajas, posibles mejoras

Cuando se habla de programar un problema de modelación no supervisada, es muy importante que se consideren los tipos de datos, que se quiere resolver, los métodos a utilizar, la cantidad de datos, entre otros aspectos muy importantes. En términos generales del código se comprobó que ambos métodos son fáciles de implementar y tienen la capacidad de categorizar grandes cifras de datos, sobre todo numéricos. Aunque también existen ciertas desventajas que tienen en común, por ejemplo se necesita hacer un ajuste para que funcionen

de la mejor manera posible, y que los resultados tengan una relación fuerte, por otra parte también dependen mucho de los parámetros que se utilicen.

En términos específicos de su aplicación a la medicina, principalmente se puede decir que ambos son muy buenas fuentes de Agrupamiento de datos médicos, como lo fueron los síntomas en esta ocasión y la segmentación de pacientes, también ambos se pueden ver implementados con otros usos médicos, que aunque no es algo que se vea en este caso en particular se debe de tomar en cuenta. Una de las desventajas que se saben acerca de su uso, es que tienen grandes dificultades con datos desequilibrados o desnivelados, que aunque en este caso no sucedió así, muchas de las bases de datos médicas tienen estos errores de dejar bastantes cifras no determinadas.

Ahora hablando del modelo elaborado en esta ocasión, podemos decir que aunque se realizar varias cosas bien, también se puede considerar que existen posibles mejoras, a pesar de que nuestros resultados fueron buenos, se pudo haber experimentado con otros modelos para comparar de manera empírica y no solamente de acuerdo a la investigación, que fue el principal factor para elegir los métodos de K mean y DBSCAN.

5. Conclusiones

5.1 Conclusión - Sebastián Miramontes Soto

Durante la elaboración de esta evidencia pudimos observar la implementación de métodos no supervisados en un caso relacionado con la salud, consideramos que este método es más aplicable para un análisis visual que los utilizados en instancias anteriores, principalmente podemos hacer comparaciones mediante gráficas observando los puntos obtenidos mediante las bases de datos más los puntos obtenidos mediante predicciones, cosa que era un poco más difícil antes de entender, también podemos tener cierta facilidad para encontrar puntos distantes al resto o fuera de la tendencia que se muestra, y personalmente considero que para los casos de agrupación o clasificación, los métodos de clustering pueden ser considerados

los mejores o de los mejores para tratar estos temas. Considero que los resultados fueron diferentes a los esperados, ya que yo personalmente pensaba que el método visto en clases iba a ser completamente superior al investigado, pero la realidad es que no fue así en los porcentajes de efectividad, por otra parte, se podría considerar que si fue mejor en la creación de clusters, ya que para uno era mucho más difícil generarlo, de hecho en varias iteraciones se generaban más de 100 clusters por lo cuál tuvimos que hacer bastantes modificaciones en la forma de realizar el código, y con esto obtuvimos resultados mucho más acertados que utilizamos para el análisis final.

5.2 Conclusión - Raúl Correa Ocañas

Este proyecto me ayudó a ver una perspectiva más del aprendizaje automático. Previo a esta clase, nunca había trabajado con modelos de aprendizaje no supervisados, y al principio estuve algo perdido en como es la metodología de evaluación en los modelos. Así mismo, el aprender de tutoriales vistos en clase me brindaron el conocimiento de pipelines, en el que pude aprender sobre la automatización de un proceso dada una base de datos. Así mismo, me percaté de las diferencias en la forma de buscar cómo optimizar los hiper parámetros de un modelo de aprendizaje no supervisado. Aprender sobre las métricas utilizadas para evaluar si un número de clusters es apropiado para el problema que se busca responder también aportó mucho a mi conocimiento, ya que aprendí sobre la métrica del coeficiente de silueta y el adjusted random index. Me encantaría poder encontrar otras aplicaciones de aprendizaje no supervisado, en el que no tenga una referencia de etiquetas para determinar la efectividad de los modelos.

5.3 Conclusión - Ericka Sofia Rodriguez Sanchez

Durante la realización de este proyecto pude comprender mejor la diferencia entre los métodos de aprendizaje supervisado y no supervisado, así como las ventajas y desventajas de cada uno. En el proceso resolvimos diferentes problemas, uno de ellos fue el funcionamiento del método DBSCAN, ya que el modelo y la representación gráfica no se construían de acuerdo a los mejores parámetros que se habían calculado en una primera instancia, sin embargo después se pudo realizar y obtuvimos resultados bastante buenos y cercanos a los esperados. Esta clase me ayudó a visualizar de una mejor manera cómo la inteligencia artificial se puede utilizar para realizar clasificaciones y predicciones prácticamente de cualquier tema mientras se tenga una base de datos mediante la construcción de modelos capaces de aprender, y me dió una perspectiva diferente hacia la solución de problemas que suelen ser cotidianos y que se pueden solucionar por este medio.

6. Referencias

- Al-Shammari, A., Zhou, R., Naseriparsa, M., & Liu, C. (2019). An effective density-based clustering and dynamic maintenance framework for evolving medical data streams. *International Journal Of Medical Informatics*, 126, 176-186.
<https://doi.org/10.1016/j.ijmedinf.2019.03.016>
- DataScientest. (2023, 30 octubre). *Machine Learning & Clustering: el algoritmo DBSCAN*. Formación En Ciencia de Datos | DataScientest.com.
<https://datascientest.com/es/machine-learning-clustering-dbscan>
- González, L. (2021, 8 septiembre). *DBSCAN teoría*. Aprende IA.
<https://aprendeia.com/dbscan-teoria/>
- Hernandez, J. (s. f.). *Clustering basado en el algoritmo K-Means para la identificación de grupos de pacientes quirúrgicos*. Artículo de Internet. Recuperado 16 de marzo de

2024, de

https://www.udi.edu.co/congreso/historial/congreso_2016/ponencias/Ingenieria_de_sistemas/JAVIER_HERNANDEZ_CACERES.pdf

Ramírez, L. (2023, January 5). Algoritmo k-means: ¿Qué es y cómo funciona? Thinking for Innovation.

<https://www.iebschool.com/blog/algoritmo-k-means-que-es-y-como-functiona-big-data/>

Rueda, J. F. V. (2019, 11 noviembre). Modelos no supervisados en salud: clusterizando células - healthdataminer.com. healthdataminer.com.

<https://healthdataminer.com/analitica-en-accion/modelos-no-supervisados-en-salud-clusterizando-celulas/>

Spencer, R. (2020, 31 marzo). *10 Aplicaciones del aprendizaje automático en la atención médica*. Lanner.

<https://www.lanner-america.com/es/blog-es/10-aplicaciones-del-aprendizaje-automatico-en-la-atencion-medica/#:~:text=El%20aprendizaje%20no%20supervisado%20tambi%C3%A9n,de%20tratamiento%20para%20estas%20enfermedades.>

TEMA 2. INTRODUCCIÓN a LA INTELIGENCIA ARTIFICIAL. CONCEPTOS

FUNDAMENTALES y APLICACIONES EN SALUD. (s. f.).

<https://www.salusplay.com/apuntes/apuntes-de-salud-digital/tema-2-introduccion-a-la-inteligencia-artificial-conceptos-fundamentales-y-aplicaciones-en-salud>