

Título del Proyecto de Investigación a que corresponde el Reporte Técnico:

Utilización de Técnicas de Inteligencia Artificial para brindar soporte cognitivo a cirujanos próximos a formarse en cirugía de cáncer de mama

TÍTULO DEL REPORTE TÉCNICO

Reporte técnico de estancia agosto-diciembre de 2020

Área del conocimiento al que pertenece

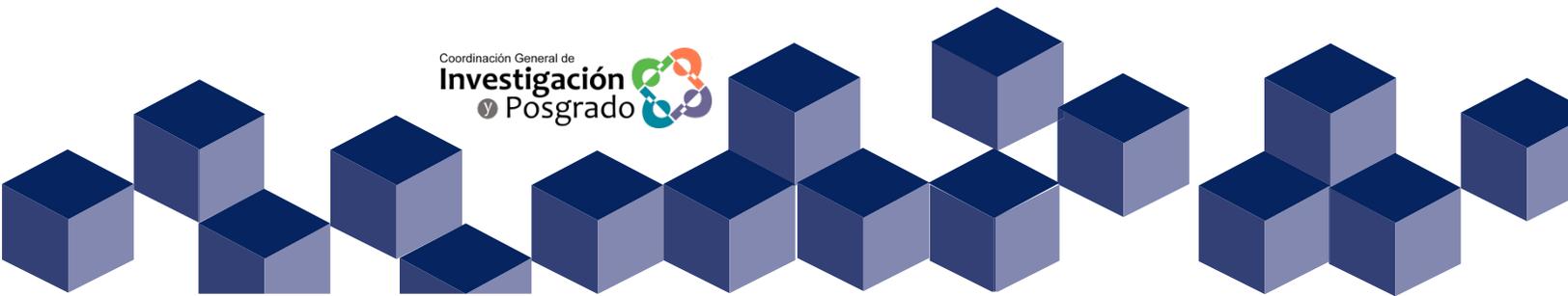
Aprendizaje Automático

Autores del reporte técnico: (Apellidos, Nombre(s))

Maestrante David Antonio Mota Medina 182962

Dr. Jorge Rodas-Osollo

Dr. Genaro Tamayo Perez



## TÍTULO DEL REPORTE TÉCNICO

Resumen del reporte técnico en español (mínimo 150 palabras):

Una clasificación mediante razonamiento basado en casos para la presentación de pacientes con posible cáncer de mama, con la ayuda de una base de datos de pacientes, es presentada en este documento. Una clasificación de pacientes permite organizar casos que tengan atributos similares de tal manera que, al presentarse un caso nuevo, este pueda ser comparado y clasificado con algunos de los casos previamente clasificados. Se presenta en el documento las técnicas de razonamiento basado en casos que son ampliamente utilizadas en inteligencia artificial y sistemas cognitivos, especialmente en aprendizaje automático, descubrimiento de conocimiento, minería de datos, etc. Los resultados esperados es mostrar una comparación de un paciente nuevo, con diferentes atributos (síntomas, edad, género), con los casos de pacientes tratados anteriormente, generando así una sugerencia para el oncólogo, que puede ser tomado como relevante para un diagnóstico más acertado, basado en datos de pacientes (casos).

Resumen del reporte técnico en inglés (mínimo 150 palabras):

A classification using case-based reasoning for possible diagnostic in patients with breast cancer, using a database of patients, is presented in this document. A classification allows us to organize cases that have features in common. With these, new cases can be compared and classified with older cases. In this document the techniques of case-based reasoning are presented. These techniques are widely used in areas like artificial intelligence, cognitive systems, specially in machine learning, data mining and knowledge discovery. The results expected is to show the case most similar case, from the database cases, to the one that is considered new. This will generate a suggestion to the oncologist that he can evaluate or contrast with his opinion and can be taken as important information to give a diagnostic because it is based on historical data of patients (cases).

Palabras clave (Al menos 3): Razonamiento basado en casos, Cáncer de Mama, Inteligencia Artificial

Usuarios potenciales (del proyecto de investigación):

Cirujanos oncológicos en formación profesional

Reconocimientos (agradecimientos a la institución que financió, estudiantes que colaboraron, instituciones que apoyaron a la realización del proyecto, etc.):

La estancia no fue apoyada por ningun recurso institucional.

1. INTRODUCCIÓN

2. PLANTEAMIENTO

- Antecedentes

- Marco teórico

3. METODOLOGÍA

4. RESULTADOS

5. CONCLUSIONES

REFERENCIAS (bibliografía)

ANEXOS

## 1. Introducción

En la actualidad, una de las principales causas de muerte en mujeres es el cáncer de mama. En el mundo representa alrededor del 16% de los casos de cáncer más comunes [1, <https://www.who.int/es/news-room/fact-sheets/detail/women-s-health>]. La detección temprana en el cáncer de mama disminuye su tasa de mortalidad en hasta un 20% según datos del Instituto Mexicano del Seguro Social.

### Estudios por imágenes

¿Para qué los médicos utilizan estudios por imágenes?

Los estudios por imágenes se usan por lo general para:

- Ver si hay cáncer en etapas iniciales (cuando es pequeño y no se ha propagado), y la persona no presenta síntomas. A esto se le puede llamar detección temprana o exámenes para detectar el cáncer.
- Buscar una masa o bulto (tumor) si una persona tiene síntomas. También pueden ayudar a descubrir si los síntomas son causados por un cáncer o por algún otro tipo de enfermedad.
- Pronosticar si es probable que un tumor sea canceroso. Esto puede ayudar a los médicos a decidir si es necesario realizar una biopsia (n una biopsia se extrae un pequeño fragmento de tejido y se examina al microscopio). Casi siempre se necesita una biopsia para saber con certeza que un cambio es cáncer.
- Mostrar dónde está el tumor, aunque esté muy adentro del cuerpo. Esto ayuda si es necesario realizar una biopsia.
- Ayudar a descubrir la etapa del cáncer (determinar si se extendió el cáncer y, de ser así, hasta qué punto se ha extendido).
- Planificar el tratamiento, como para mostrar a dónde se necesitan dirigir los rayos de la radioterapia.

- Mostrar si un tumor se ha reducido, ha quedado igual o ha crecido después del tratamiento. Esto puede darle al médico una idea de qué tan bien está funcionando el tratamiento.
- Averiguar si el cáncer ha regresado (recurrido) después del tratamiento.

Los estudios por imágenes son solo parte del diagnóstico y de tratamiento del cáncer que los médicos especialistas oncólogos, onco-ginecólogos y cirujanos oncológicos llevan a cabo para realizar un pre-diagnóstico. Un estudio completo del cáncer también incluye su historia clínica (una serie de preguntas sobre sus síntomas y factores de riesgo), un examen físico, análisis de sangre u otras pruebas de laboratorio.

Muchos médicos planean realizar radiografías u otros estudios por imágenes antes de iniciar el tratamiento. Estas imágenes se utilizan para el seguimiento de cambios que surgen durante el tratamiento. Estos se llaman estudios de referencia porque muestran cómo las cosas se veían al inicio. Se pueden comparar con imágenes posteriores a fin de ver los resultados del tratamiento con el paso del tiempo. De hecho, el pre-diagnóstico podría llegar a ser tan acertado como el resultado de la biopsia lo sea gracias a la experiencia que estos especialistas acumulan con la cantidad de tratamientos emitidos en su práctica profesional.

Es así que, los estudios por imágenes son de gran relevancia pues los usan para detectar el cáncer, saber cuán lejos se ha propagado y determinar cuál tratamiento podría ser más efectivo. En esencia un estudio por imágenes es una manera de saber que está sucediendo dentro del cuerpo de la paciente. Estos estudios envían formas de energía (como rayos X, ondas sonoras, partículas radiactivas o campos magnéticos) a través de su cuerpo. Los tejidos de su cuerpo cambian los patrones de energía para producir una imagen o fotografía. Estas imágenes muestran cómo los órganos internos se ven y funcionan para que los médicos puedan observar cambios que pudieran ser causados por enfermedades como el cáncer.

Los estudios por imágenes no son infalibles

Los estudios por imágenes pueden ser a menudo muy útiles, pero tienen límites. Por ejemplo, la mayor parte del tiempo, estos estudios por sí solos no pueden mostrar con certeza si un cambio es causado por el cáncer.

Los estudios por imágenes pueden encontrar grandes grupos de células cancerosas, sin embargo, ningún estudio por imágenes puede mostrar una sola célula cancerosa o incluso unas pocas. De hecho, se necesitan millones de células para hacer que un tumor sea lo suficientemente grande para que se vea en un estudio por imágenes. Por esta razón el tratamiento puede continuar incluso cuando las células cancerosas ya no pueden ser vistas en un estudio por imágenes. El objetivo es identificar cualquier célula sobreviviente de cáncer. Incluso una célula cancerosa puede crecer y con el tiempo, convertirse en un tumor que volverá a ser lo suficientemente grande como para causar problemas o se vea en un estudio por imágenes.

Por otro lado, a veces los estudios por imágenes pueden mostrar algo que se ve como cáncer, pero pruebas adicionales (tales como una biopsia) indican que no es cáncer. Incluso, un pre-diagnóstico puede complicarse debido a varios factores como lo son el contraste de la imagen, resolución, tamaño (rango de .1mm a .7mm) [2] y morfologías de tumores mimetizadas con músculos, vasos y tejidos de la glándula mamaria, pericia del médico... En este ámbito herramientas de la Inteligencia artificial (IA) pueden proveer al médico de conocimiento que ayuden a mejorar su práctica en cuanto a calidad, rapidez o ambas. Por mencionar algunas de las herramientas de la inteligencia artificial se encuentran: el aprendizaje automático, análisis inteligente de datos ambas utilizadas para clasificar y analizar información con el objetivo de generar conocimiento y que este pueda ser comunicado de forma adecuada para que el médico lo asimile y haga uso del mismo.

## **2. Planteamiento de la problemática**

En la actualidad se genera una gran cantidad de datos e información del área de la medicina, relativa a los pacientes, como pueden ser estudios por imágenes, expedientes clínicos... Por lo que, obtener información útil, como patrones o asociaciones entre

atributos, que caracterice a tales estudios y que permita mejorar el tratamiento de algún padecimiento cobra una gran importancia.

Tradicionalmente, se asume que el desempeño de un cirujano, con especialidad en cáncer de mama, se basa en su experiencia la cual tiene un componente extenso de conocimiento tácito. Dicho conocimiento lo utiliza cuando pretende definir un diagnóstico y en consecuencia determinar el tratamiento en particular para un paciente. A lo largo del tiempo de su práctica laboral el médico va acumulando una cantidad nada despreciable de casos que ha tratado, lo cual incrementa la dificultad de que éste pueda recordar la información relevante sobre tales pacientes.

Por otra parte, el conocimiento derivado de los casos suele estar solo a disposición del médico ya que este es tácito; o bien, no está explícito. Lo cual imposibilita el que un médico que se inicia en esta especialidad utilice dicho conocimiento y pueda reducir cualitativamente el tiempo de su curva de aprendizaje.

Otro factor importante que imposibilita la transferencia de conocimiento del médico experto hacia los menos experimentados es la cantidad de registros de pacientes, lo cuales crean una ruptura en la asimilación de conocimiento. Teniendo una cantidad elevada de registros se vuelve complicado recordar todos los registros y patrones que ayudaron a generar un diagnóstico por parte del médico experto. Dicho evento cuenta con conocimiento importante que se puede ser transferido a médicos menos experimentados. Por lo tanto, se debe generar un mecanismo de gestión de conocimiento en este dominio. Dicho mecanismo, permitirá que médicos que inician su práctica en tratamientos oncológicos puedan aprender de la experiencia acumulada al tiempo que van tratando nuevos pacientes con el conocimiento (diagnósticos) de registros de pacientes previamente atendidos. Adicionalmente, permite al médico experimentado poder contrastar nuevos diagnósticos con diagnósticos previamente realizados

Por lo que, es deseable implementar una herramienta del ámbito del cómputo cognitivo que: gestione casos conformados por la experiencia del médico cirujano con especialidad en cáncer de mama, identificar aquellos semejantes al caso que actualmente este por tratar,

comunicar de forma sencilla y concreta información relevante proveniente de tales casos en apoyo a una asimilación rápida por parte del médico en formación.

### **Antecedentes**

En la actualidad, una de las principales causas de muerte en mujeres es el cáncer de mama. En el mundo representa alrededor del 16% de los casos de cáncer más comunes [3]. Según datos del instituto mexicano del seguro social se ha reducido la tasa de mortalidad del cáncer de mama en un 20% sólo con la aplicación de campañas de prevención del cáncer de mama mediante mamografías. La detección temprana en el cáncer de mama también ayuda en la reducción de la tasa de mortalidad por este cáncer. Las imágenes generadas mediante las mamografías son parte de importante para la detección temprana.

Actualmente como se hace es a través de mastografías, la cual juega un papel muy importante en la detección del cáncer de mama ya que es uno de los primeros diagnósticos que se ofrecen al paciente. Actualmente se realiza el diagnóstico en base al conocimiento del doctor especialista y los que se pueda detectar de manera visual en la mastografía. Aquí es donde es importante que se utilicen herramientas del diseño asistido por computadora (CAD por sus siglas en inglés computer-aided design) que ayuden al especialista a generar un diagnóstico basado en casos anteriores.

Sin embargo, en ocasiones es complicado hacer un diagnóstico debido al poco contraste en las imágenes, las diferentes resoluciones de imagen, así como también los tamaños y formas de los tumores que pueden ser fácilmente confundidos con otros músculos, vasos y tejidos de la glándula mamaria. Los tumores pueden llegar a tener un tamaño en el rango de .1mm a .7mm [4], esto complica la identificación de estos.

Debido a que se encuentran disponible una cantidad importante de datos dentro del área de la medicina, como pueden ser imágenes, expedientes de pacientes, etc. Es importante que estos datos sean utilizados de tal manera que se conviertan en información útil. Esto puede ser desde encontrar patrones en los datos, clasificarlos, analizarlos, etc.

Mediante un marco computacional para resolver problemas de clasificación de fallos como lo es un Razonamiento Basado en Casos (CBR) se puede generar una herramienta que

permite que otros médicos que comienzan su incursión en el dominio, sin experiencia ni conocimientos más específicos, entender y visualizar el problema oncológico fundamental. También con esto puede orientar al oncólogo desde la visión no solo de la experiencia sino de experiencia explícita acumulada a lo largo del tiempo.

La Inteligencia artificial (IA) es una disciplina que está relacionada con las ciencias de la computación las cuales se relacionan con los cálculos y procedimientos matemáticos. Dentro de la inteligencia artificial se encuentra el cómputo cognitivo el cual es una aproximación innovadora para construir sistemas computacionales inteligentes. El análisis inteligente de datos se apoya del cómputo cognitivo en las técnicas y métodos para clasificar y analizar datos, esto con el objetivo de generar conocimiento. Parte del análisis inteligente de datos es la clasificación, la cual se enfoca en agrupar objetos con características similares. Por último, herramientas que son utilizadas en diferentes áreas, como por ejemplo los Rough Set uno de los temas que aborda para la clasificación, que en principio se encarga de ofrecer todos los atributos más importantes dentro de los datos para un posterior procesamiento de técnicas de minería de datos o aprendizaje automático.

Algunos de los trabajos están relacionados con la detección de irregularidades o anomalías en imágenes médicas. Estas imágenes médicas pueden ser mamografías, resonancias magnéticas, ultrasonidos, etc.

En [5] se propone un a ejecución digital de un modelo basado en el procesamiento de un histograma difuso y un algoritmo de agrupamiento c-mean difuso para la detección del cáncer de mama. Se realizan varias etapas como lo son el procesamiento de la imagen digital, seguido de un preprocesamiento de la imagen. Después e utiliza una segmentación para seguir con la extracción de características. Por último, se realiza la clasificación mediante reglas para poder detectar la presencia del cáncer en las imágenes digitales.

En [6] se presenta una conjunción ente rough sets y los conjuntos difusos a la par con técnicas de extracción de características mediante estadística. Presenta una clasificación de imágenes de cáncer de mama con solo dos salidas: cáncer y no cáncer. Propone una metodología que comienza con un preprocesamiento para aumentar el contraste de las imágenes y extraer la región de interés de la imagen. Seguido de la extracción de

características en la imagen. Después realiza una selección de características y generación de reglas mediante rough sets. Finalmente, las reglas son agregadas al clasificador para la discriminación de las diferentes regiones de interés. Para medir la similitud se utilizó una función de distancia basada en rough sets.

## **Marco Teórico**

### Dominios de estructura informal

Un dominio es un área específica donde se lleva a cabo una actividad. Esto se puede ver en la vida diaria desde unos dominios relativamente pequeños como en la cocina o la carpintería hasta las diferentes carreras universitarias como lo pueden ser la ingeniería de software, medicina, entre otras. Una característica importante que comparten estos dominios es que para resolver una problemática en estos dominios normalmente se requiere el apoyo de un especialista del dominio (ED). Estos especialistas basan su conocimiento en su experiencia adquirida a través de los años. En estos dominios es donde se utiliza el término de dominio de estructura informal [7] en donde el conocimiento es informal, incompleto, tácito, parcial y los datos e información puede o no tener una estructura definida. Es por esto por lo que los ED tienen un papel muy importante, ya que es necesario realizar la definición de los tipos de datos y sus relaciones, como también la formalización del conocimiento que posee el ED.

La principal diferencia entre los dominios de estructura informal a otros dominios está dada por la información con la que cuentan los especialistas del dominio, que se puede definir como el conocimiento tácito que la experiencia formó en el especialista. El conocimiento tácito se define como el conocimiento que se tiene sobre un dominio en específico pero que no se sabe cómo explicar dicho conocimiento, esto puede ser como la identificación de lesiones en una mamografía, donde no se puede explicar de manera explícita como se realiza esta actividad de identificación de lesiones. Esto trae como consecuencia un trabajo complejo a la hora de tratar de obtener y formalizar el conocimiento del ED. Al intentar transferir este conocimiento a personas ajenas al dominio lleva consigo una pérdida de información valiosa para llegar a comprender el dominio. Esto conlleva a que la curva de aprendizaje aumente, en donde se tiene que invertir más tiempo en aclaraciones y consultas.

Existe conocimiento tácito el cual nunca se va a poder explicar, por ello entonces, es importante que al trabajar con dominios de estructura informal se tenga al experto del dominio, ya que ellos valoraran las interpretaciones que se le estén dando a los datos.

#### Proceso sistemático KMoS-RE

KMoS-RE es un proceso sistemático de gestión de conocimiento para el análisis de requisitos que ese enfoca en obtener y estructurar el conocimiento de los dominios de estructura informal para posteriormente incorporarlo a los requerimientos de la solución de un problema, evitando ambigüedades o ausencias en los requisitos.

El proceso KMoS-RE© se apoya en tres ideas fundamentales [8]:

1. Utiliza un mecanismo para representar y comprender el dominio de aplicación
2. Incorpora el modelo de gestión de conocimiento SECI (Socialización, Externalización, Combinación e Internalización)
3. Gestiona el conocimiento tácito mediante la incorporación de métodos para identificar, capturar, indexar y hacer explícita la mayor cantidad posible de conocimiento tácito.

#### Razonamiento Basado en casos (RBC)

El razonamiento basado en casos aborda la resolución de problemas partiendo por problemas resueltos anteriormente apoyándose en la solución que se les dé a estos [9]. Este tipo de resolución de problemas tiene similitudes en como el ser humano resuelve problemas de la vida diaria, consciente o inconscientemente el ser humano recuerda eventos similares a la hora de abordar y querer solucionar un nuevo problema. El uso de la tecnología ha aumentado gradualmente la cantidad de información que se generan en los dominios, esto tiene como consecuencia que los especialistas del dominio se vean en la necesidad de apoyarse con sistemas que soporte la toma de decisiones en base a toda esta gran cantidad de información. Las capacidades de las herramientas de toma de decisiones ayudan a los ED a tomar mejores decisiones debido a los nuevos contextos que se generan en base a la información.

El proceso de resolver problemas y de aprender a base de experiencia, se puede observar frecuentemente en una buena cantidad de ámbitos, por ejemplo, la medicina, la aplicación de las leyes, problemas de diseño y estimaciones, y en gran medida son utilizados en situaciones dónde se toman decisiones; ya que es más sencillo comprender cómo poder resolver un problema a partir de la explicación de la solución que le dio en su momento a un problema resuelto.

### Representación de casos en un RBC

Para comenzar con la explicación de un RBC es preciso definir lo que es un caso, un caso es una experiencia para resolver un problema en concreto [11]. Esta incluye una representación de los casos de tal manera que se explique de qué manera se obtuvo la solución al problema planteado. La parte de la representación de los casos es una etapa precia al ciclo del RBC que debe ser bien aplicada ya que de esta dependerá la eficiencia de la base de casos que si forma parte del ciclo de RBC. Los problemas a los que se enfrenta la representación de los casos son:

- La definición de los atributos más importantes que describen los casos
- La definición de la estructura que describa el contenido de los casos
- Como se organizan los casos en la base de casos

Una vez entendido los problemas principales para la representación de los casos se tienen a revisar los deferentes métodos para la representación de los casos los cuales se dividen en dos tipos que son los tradicionales y los de conocimiento intensivo. Las representaciones tradicionales son las más conocidas y se conocen por su representación de conocimiento limitado, por no describir relaciones entre sus características y su efectividad para cuando se tienen pocos atributos para los casos. Esto tipo de representación funcionan bien para casos simple, para casos más complejos se utilizan los de conocimiento intensivo. Los de conocimiento intensivo son conocidos por su trabajo con ontologías (especificación formal de conceptualización compartida) y por las relaciones que tienen entre sus atributos para generar un mayor conocimiento en la base de casos. Estos métodos contienen reglas de tipo

de terminación y adaptación, donde la de terminación es para inferir en casos pasados mientras que las de adaptación son para como un caso pasado puede ser adaptado.

Los métodos de representación [11] tradicionales más comunes son:

- Representación como vector de características. En esta representación se utilizan un mismo tipo de atributos, que pueden ser numéricos o de atributo-valor. También no existen relaciones entre sus atributos y se categorizan los casos en base a sus atributos.
- Representación basada en marcos. En esta representación se utilizan ambos tipos de atributos numéricos y de atributo-valor. Estos en ocasiones pueden contener relaciones semánticas entre los atributos. En esta representación se puede utilizar el conocimiento del dominio donde se trabaja.
- Representación orientada a objetos. Solo se utilizan atributos de tipo atributo-valor. Se representan los casos como colecciones de objetos que están descritos a través de sus atributos.
- Representación de casos basados en predicados. Un predicado es la relación de los objetos que tiene condiciones de tipo SI y acciones de tipo ENTONCES. Donde los predicados que no tienen condicionales son hechos. Su dificultad radica en la articulación de los valores a los atributos de manera precisa, pero esto se puede resolver mediante herramientas de inteligencia artificial.

Los métodos de representación de conocimiento intensivo (semántico) son:

- Ontología como vocabulario del dominio en RBC. En este método se tienen dos opciones para el query (caso nuevo) donde se utiliza el mismo vocabulario que se encuentra en los casos, pero también se puede utilizar una ontología como vocabulario para una información más relevante y de obtención semántica.
- Ontología como base de casos y vocabulario del dominio. Este método utiliza una ontología no solo en el vocabulario si no también es incluida en la representación de los casos permitiendo una recuperación de los casos que es guiada por la adaptabilidad. Para esta representación el software de Jcoilibri es muy utilizado.

- Basado en OWL (ontology web language por sus siglas en ingles) y metodología de representación de casos en dominio médico. En este método se obtienen ontologías de gran peso donde la ontología está integrada con la ontología del dominio para una recuperación de casos semántica. Cuenta con conceptos estandarizados y conceptos específicos como ICD10. Las ventajas de la utilización de las ontologías en forma de casos son que tienen reusabilidad, se pueden compartir las ontologías y que pueden ser utilizadas en diferentes sistemas y ambientes.

### Ciclo de un RBC

Una vez definida la representación de los casos se puede pasar al ciclo del RBC que esta ejemplificado de manera gráfica en la figura 1. Este ciclo de manera general [9] se describe con los siguientes cuatro procesos:

- **RECUPERACIÓN** de los casos o caso más similar
- **REUTILIZAR** la información y conocimiento en el caso o casos para solucionar el problema
- **REVISAR** la solución propuesta
- **RETENER** las partes de esta experiencia que será útil para futuras soluciones a problemas

Además de los procesos existen otros conceptos importantes que hay que conocer [10] como lo es un nuevo caso (normalmente llamado query), la base de casos, caso más similar, la solución propuesta y por último confirmar solución. El nuevo caso o query es la información, sin solución o variable de salida, con la cual se hará el proceso de recuperar el caso más parecido a este, esto tiene el enfoque de buscar un caso similar que ya haya sido resuelto anteriormente y poder utilizar esa solución en este caso nuevo. Seguido se encuentra la base de casos que no es más que el contenedor de todos los casos construidos, cabe mencionar que estos casos ya contienen una solución o variable de salida. El caso más similar es el resultado del proceso de recuperar, aquí se realizan diferentes metodologías según la estructura y definición de los casos para obtener el caso más parecido de la base de casos en comparación con el nuevo caso. La solución propuesta es el resultado del proceso

de reusar el cual se puede dar por medio de copiar o integrar la solución de los casos o caso que fueron recuperados en el proceso de recuperación. Por último, en la confirmación de la solución es donde se le dice al sistema que el nuevo caso aporta gran conocimiento a la base de casos y se anexa a la misma. En este último paso es donde se puede decir que el sistema RBC está aprendiendo ya que adquiere conocimiento a través de la retroalimentación que le dan los casos nuevos.

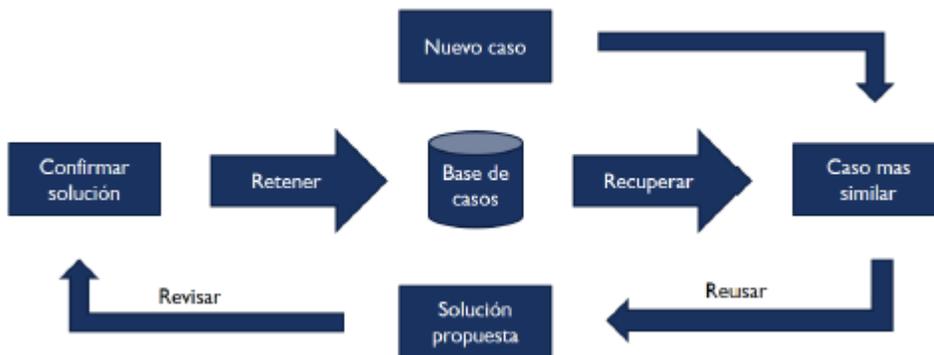


Figura 1 La figura muestra un ciclo de un CBR a partir de la aparición de un caso nuevo [10]

Con el objetivo de entender los procesos de CBR [12] más a fondo y ver que sucede en cada uno de ellos, a continuación, se describen cada uno de ellos:

- Procesos de recuperación. En estos procesos se utilizan las medidas de similitud entre la base de casos y el query. Estas medidas de similitud nos permiten saber que tan parecidos son los casos a través de sus atributos, donde estos pueden ser mediante valores numéricos o atributos-valor. Según los que se utilicen podrán obtenerse mejores resultados si junto al atributo se le asigna un valor de peso el cual influirá de manera directa a la hora de querer obtener los casos más parecidos al query.
- Procesos de reutilización. Aquí se utiliza la solución adaptada con la unión de las soluciones obtenidas en el proceso de recuperación o en su defecto utilizar directamente la solución del caso más parecido que se obtenga. Esto se da de esta manera ya que en el proceso de recuperación se obtiene una métrica de la cual se puede elegir la cantidad de casos que se quieran utilizar en este proceso.

- Proceso de revisar. Aquí es donde se deja a la retroalimentación al experto del dominio para saber si la solución propuesta o obtenida es la adecuada para el query. En este proceso se le da una evaluación a la solución y de ser necesario se repara la solución.
- Proceso de retención. En este proceso se decide qué información es la que se retiene de igual forma que el proceso anterior con ayuda del experto del dominio. En caso de que se considere que la información de query puede aportar para casos futuros se anexa la información de esta a la base de casos.

### Clusterización

Es una de las técnicas utilizadas en el aprendizaje no supervisado, donde no se tienen las etiquetas que pueden describir a los datos, que consiste en agrupar datos de manera automática basados en los valores de los datos, formando así grupos o clusters que tienen registros de datos que son similares entre los grupos. Es importante mencionar que al ser un tipo de aprendizaje no supervisado no existe una respuesta correcta, esto hace que la evaluación en los datos se vuelva subjetiva. Al final lo que se pretende con la clusterización es encontrar cual es el mejor agrupamiento de los datos. Las aplicaciones son variadas y van desde la agrupación de clientes en la parte administrativa hasta la agrupación de enfermedades en el ámbito médico.

El procedimiento de clusterización se muestra en la figura 2.

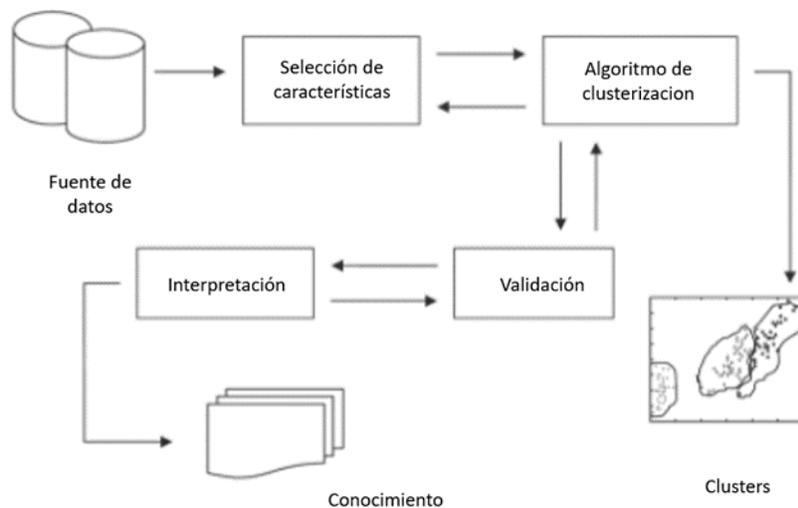


Figura 2 La figura muestra el proceso de clusterizacion [10]

Partiendo de los elementos en la figura 2 se tiene en un principio los datos que se van a clusterizar o agrupar, después se tienen las características que describen a esos datos. Dichas características deben ser revisadas de manera puntual para poder realizar una clusterizacion importante. Después sigue el algoritmo de clusterizacion donde se puede escoger dentro de una variedad de ellos donde el más común es el conocido como “K-means”. Durante el proceso de aplicación del algoritmo se pueden tener ayudas visuales sobre cómo se están agrupando los datos. En los últimos pasos se tiene la validación y la interpretación de los resultados. Estos pueden ser validados por alguna persona experta según el dominio de donde provengan los datos. Al final el objetivo es obtener un conocimiento que en este caso será representado en los grupos y en el proceso de que identifica a cada uno de los grupos.

#### Pruebas de significancia estadística

En términos generales, se dice que las pruebas de significancia estadística consisten en calcular la probabilidad de que las discrepancias observadas ocurran por el azar [18]. Si el valor que se obtiene de las pruebas es muy grande se dice que las diferencias o discrepancias no son significativas, y para el caso contrario donde el valor de la prueba es menor, en algunas pruebas es 0.05, se puede afirmar que las variaciones son significativas. Esto nos ayuda a realizar una validación de los datos que se tienen de alguno dominio.

El procedimiento general para la realización de una prueba de significancia estadística cuenta con los siguientes pasos:

- a) Plantear la hipótesis (nula) y la alternativa con base a los objetivos
- b) Calcular valores de la muestra
- c) Determinar la distribución de probabilidad adecuada al caso
- d) Escoger el nivel de significancia y calcular los grados de libertad si procede

- e) Buscar en la tabla correspondiente el valor de  $Z$  o  $t$  determinados por el nivel de significancia y los grados de libertad cuando corresponde
- f) Calcular con base a los valores de la muestra y en los de la población el valor de  $Z$  o  $t$  correspondientes al problema en particular
- g) Comparar los valores obtenidos en (f) con el obtenido en (e). Si este en términos absolutos es menor que aquel, se acepta la hipótesis nula. La hipótesis nula se rechaza cuando el valor calculado de (f), es, en términos absolutos, mayor que el encontrado en (e). En este caso se dice que la diferencia entre los valores comparados (promedios, proporciones, frecuencias, etc) son estadísticamente significativos
- h) Se obtiene la conclusión de acuerdo con el caso al que se le aplico la prueba.

El procedimiento puede llegar a ser un tanto largo y tedioso debido a las consultas en tablas y cálculo de valores, afortunadamente existen programas para computadora que realizan desde los cálculos hasta las consulta de los valores en tablas según los valores de significancia que introduce el usuario ahorrando así el tiempo de la realización de este procedimiento y expandiendo su aplicación para cuando se tiene una cantidad considerable de atributos de los cuales se quieren obtener pruebas de significancia estadística.

## **5. Metodología**

La metodología empleada durante la estacia para alcanzar el objetivo de dar inicio a la conformación de la case de casos es la que se señala para un RBC:

- RECUPERACIÓN
- REUTILIZAR
- REVISAR
- RETENER

Además de realizar los procesos de CBR:

- Procesos de recuperación.
- Procesos de reutilización.
- Proceso de revisar.
- Proceso de retención.

Se continua con la Clusterización cuyo procedimiento se indica en la figura 2: Se parte de los datos que se van a clusterizar o agrupar, después se tienen las características que describen a esos datos. Dichas características deben ser revisadas de manera puntual para poder realizar una clusterización importante. Después sigue el algoritmo de clusterización. En los últimos pasos se tiene la validación y la interpretación de los resultados.

Y, finalmente, las pruebas de significancia estadística que calculan la probabilidad de que las discrepancias observadas ocurran por el azar [18]. Esto nos ayuda a realizar una validación de los datos que se tienen de alguno dominio.

El procedimiento general para la realización de una prueba de significancia estadística cuenta con los siguientes pasos:

- a) Plantear la hipótesis (nula) y la alternativa con base a los objetivos
- b) Calcular valores de la muestra
- c) Determinar la distribución de probabilidad adecuada al caso
- d) Escoger el nivel de significancia y calcular los grados de libertad si procede
- e) Buscar en la tabla correspondiente el valor de  $Z$  o  $t$  determinados por el nivel de significancia y los grados de libertad cuando corresponde
- f) Calcular con base a los valores de la muestra y en los de la población el valor de  $Z$  o  $t$  correspondientes al problema en particular
- g) Comparar los valores obtenidos en (f) con el obtenido en (e). Si este en términos absolutos es menor que aquel, se acepta la hipótesis nula. La hipótesis nula se rechaza cuando el valor calculado de (f), es, en términos absolutos, mayor que el encontrado en (e).

En este caso se dice que la diferencia entre los valores comparados (promedios, proporciones, frecuencias, etc) son estadísticamente significativos

h) Se obtiene la conclusión de acuerdo con el caso al que se le aplico la prueba.

## **6.Resultados**

Se logro obtener un vector de atributos que permiten la discriminación de los datos de las pacientes de forma que facilite la implementación de los casos que conformaran la base de casos para el sistema de razonamiento basado en casos. Dicho vector fue validado por el especialista del área.

## **7.Conclusiones**

La representación de casos es un tema que exige tiempo de dedicación para realizar un arduo trabajo de preparación y gestión de los datos que serán utilizados para conformar los casos a utilizar por el RBC.

El apoyo del especialista, en especial para el ámbito médico, tiene un rol relevante para la determinación de que atributos serán tomados en cuenta para el sistema RBC. Sin la ayuda del mismo es muy probable que el RBC hubiera fracasado.

## **Referencias**

[1] A. Arokiyarny Delphina, M. Kamarasan and S. Sathiamoorthy, “Image Processing For Identification of Breast Cancer: A Literature Survey”, Asian Journal of Electrical Sciences, Volume 7 No.2, pp 28-37, July-December 2018

[2] Gaona, Enrique & Izel Osorio, Sofía & Arenas, Valeria & Salvador Bernal, Miguel & Tzitzitlini Labra, Samara & Molina-Frechero, Nelly & Franco Enríquez, Jesús & Rivera, Teodoro & Alvarez, Beatriz, “Efficiency Indicators of Mammography in the Detection of Breast Cancer Early Stages: Exploratory Study in Mexico” International Journal of Applied Science and Technology, Vol. 7, No. 4, December 2017

[3] A. Arokiyarny Delphina, M. Kamarasan and S. Sathiamoorthy, “Image Processing For Identification of Breast Cancer: A Literature Survey”, Asian Journal of Electrical Sciences, Volume 7 No.2, pp 28-37, July-December 2018

[4] Gaona, Enrique & Izel Osorio, Sofía & Arenas, Valeria & Salvador Bernal, Miguel & Tzitzitlini Labra, Samara & Molina-Frechero, Nelly & Franco Enríquez, Jesús & Rivera, Teodoro & Alvarez, Beatriz, “Efficiency Indicators of Mammography in the Detection of Breast Cancer Early Stages: Exploratory Study in Mexico” International Journal of Applied Science and Technology, Vol. 7, No. 4, December 2017

[5] Chiranjil Lal Chowdhary, D. P. Acharjya, “Breast Cancer Detection using Intuitionistic Fuzzy Histogram Hyperbolization and Possibilitic Fuzzy c-mean Clustering algorithms with texture feature based Classification on Mammography Images”, Proceedings of the International Conference on Advances in Information Communication Technology & Computing, August 2016

[6] AboulElla Hassanien, “Fuzzy rough sets hybrid scheme for breast cancer detection”, Image and Vision Computing, Volume 25, Issue 2, Pages 172-183, 2007

[7] Olmos-Sánchez, K., & Rodas-Osollo, J. “A Strategy of Requirements Engineering for Informally Structured Domains Advisor” Universidad Autónoma de Ciudad Juárez. International Journal of Combinatorial Optimization Problems and Informatics, 7(2), 49–56 2016

[8] Olmos-Sánchez, K., Rodas-Osollo, J., Fernández Martínez, L., & Morales Rocha, V. “Requirements engineering based on knowledge : a comparative case study of the KMoS-RE strategy and the DMS process”, (77), 88–94. 2015 <https://doi.org/10.17533/udea.redin.n77a11>

[9] Aamodt, A., & Plaza, E. “Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches”. AI Communications, 7(1), 39–59. 1994 <https://doi.org/10.3390/s120811154>

[10] Simon Shiu and Sankar K. Pal. “Foundations of Soft Case-Based Reasoning”. John Wiley & Sons, Inc., Hoboken, NJ, USA. 2004

[11] Shaker H. El-Sappagh and Mohammed Elmogy, “Case Based Reasoning: Case Representation Methodologies” International Journal of Advanced Computer Science and Applications (IJACSA), 6(11), 2015.

[12] El-Fakdi, A., Gamero, F., Meléndez, J., Auffret, V., Haigron, P.: EXiTCDS: “A framework for a workflow-based CBR for interventional clinical decision support systems and its application to TAVI”. Expert Syst. Appl. 41, 284–294 2014

[13] Google (2020). Google Forms [Online]. Available: <https://www.google.com/forms/about/>

[14] Mark Brocato (2020). Mockaroo [Online]. Available: <https://www.mockaroo.com>

[15] Guido van Rossum (2020). Python [Online]. Available: <https://www.python.org>

[16] Dataiku (2020). Dataiku Data Science Studio [Online]. Available: <https://www.dataiku.com/>

[17] Rui Xu, Don Wunsch: “Clustering”. Wiley-IEEE Press, 2008

[18] Ligia Moya de Madrigal: “Introducción Ala Estadística de la Salud”. Editorial Universidad de Costa Rica, 1986 Páginas 260-274

[19] SPSS Statistics (2020). IBM [Online]. Available: <https://www.ibm.com/products/spss-statistics>

## **Anexos**

### Anexo A

#### Léxico Oncología

1. Autoexploración. – forma en la que tocamos determinadas partes del cuerpo para identificar su condición.
2. Biopsia. - extracción de células o tejidos de nuestro cuerpo con distintos medios para identificar la enfermedad.
3. Cáncer. – reproducción anormal de uno o más tipos de células de nuestro cuerpo.
4. Células. – unidad anatómica y funcional de los seres vivos.

5. Diagnostico. – resultado de todos los estudios que se realizan para determinar si hay enfermedad.
6. Ganglios. – pequeñas estructuras que se encuentran a lo largo de nuestro cuerpo y nos ayudan a filtrar lo que este no necesita.
7. Ginecólogo. – médico especialista en la salud de la mujer.
8. Glándula mamaria. – órgano glandular ubicado en el pecho. Formado por tejido conjuntivo, graso y tejidos.
9. Mastectomía. – cirugía mediante la cual se remueve la glándula mamaria.
10. Mastografía. – Estudio radiológico de la glándula mamaria que se utiliza para ver si tiene anormalidades.
11. Metástasis. – Cuando el crecimiento de las células anormales ya sucede en varios órganos del cuerpo.
12. Oncólogo. – médico especialista en cáncer
13. Quimioterapia. – tratamiento utilizado para tratar el cáncer.
14. Radioterapia. – empleo de rayos para reducir el tamaño de los tumores
15. Remisión. – Cuando los estudios ya no muestran celulares anormales
16. Tumor. – crecimiento de muchas células anormales que se agrupan para formar una bola o masa.
17. Análisis de laboratorio. – procedimiento que evalúa una muestra de sangre, orina o sustancia del organismo para realizar un diagnóstico.
18. Benigno. – tumor no invade el tejido circundante ni se disemina a otras partes.
19. Diseminar. -separar cosas juntas por distintos lugares
20. Maligno. – tumor que es canceroso. Puede invadir tejido o diseminarse a otras partes.
21. Patólogo. -médico especialista en interpretar análisis de laboratorio y evaluar células, tejidos y órganos.
22. Predisposición. – la tendencia a tener una enfermedad que puede desencadenarse en determinadas condiciones.
23. Prueba por imágenes. - crear imágenes de partes internas del cuerpo, tejidos u órganos para realizar un diagnóstico.
24. Pronostico. – posibilidad de recuperación o predicción del resultado de la enfermedad.

## Anexo B

### Datos del proyecto

Variable	Tipo	Indicador	Rango	Reglas	Creado
Edad	Cuantitativa	Años de vida del	Entre 40 y 70	n/a	Python

		paciente			
<b>Densidad de la mama</b>	Cuantitativa	Porcentaje entre tejido y grasa	Entre 30 y 70	n/a	Mockaroo
<b>Edad de primera menstruación</b>	Cuantitativa	Edad en años de cuando la paciente comenzó a menstruar	Entre 10 y 15	n/a	Python
<b>Gestas</b>	Cuantitativa	Número embarazos	Entre 0 y 5	n/a	Mockaroo
<b>Paridad</b>	Cuantitativa	Número partos y/o cesáreas	Entre 0 y 5	<= Gestas	Mockaroo
<b>Menopáusica</b>	Cualitativa	Si se encuentra actualmente en la menopausia	<ul style="list-style-type: none"> <li>• True</li> <li>• False</li> </ul>	If <b>Edad</b> > 55 then True  Else False	Mockaroo
<b>Menopáusica quirúrgica</b>	Cualitativa	Si se realizó la extirpación del útero y los ovarios	<ul style="list-style-type: none"> <li>• True</li> <li>• False</li> </ul>	If menopáusic a is False Then false Else True or false	Mockaroo
<b>Peso</b>	Cuantitativa	Número, peso en kilogramos	Entre 40 y 100	n/a	Python
<b>Sobrepeso</b>	Cualitativa	Si la paciente tiene sobrepeso	<ul style="list-style-type: none"> <li>• True</li> <li>• False</li> </ul>	Distribución 70% true 30% false	Mockaroo
<b>Antecedentes de biopsia abierta</b>	Cualitativa	Si la paciente se le ha realizado una biopsia abierta anteriormente	Aleatorio entre: <ul style="list-style-type: none"> <li>• No</li> <li>• Si Izquierda</li> <li>• Si derecha</li> <li>• Si bilateral</li> </ul>	n/a	
<b>Antecedentes de biopsia por sacabocado</b>	Cualitativa	Si la paciente se le ha realizado biopsia por sacabocado anteriormente	Aleatorio entre: <ul style="list-style-type: none"> <li>• No</li> <li>• Si izquierda</li> <li>• Si derecha</li> <li>• Si bilateral.</li> </ul>	n/a	Mockaroo
<b>Antecedentes de carcinoma</b>	Cualitativa	Si la paciente ya ha presentado un cáncer anteriormente	Aleatorio entre <ul style="list-style-type: none"> <li>• No</li> <li>• Si izquierda</li> <li>• Si derecha.</li> </ul>	n/a	Mockaroo
<b>Antecedente de uso de hormonales</b>	Cualitativa	Si la paciente ha utilizado hormonales	<ul style="list-style-type: none"> <li>• True</li> <li>• False</li> </ul>	Distribución 80% true 20% false	Mockaroo
<b>Años de uso de hormonales Anticonceptivo</b>	Cuantitativa	Años que la paciente ha utilizado hormonales como anticonceptivo	Entre 0 y 10  0= nunca	IF <b>Antecedente de uso de hormonales</b> = True Then aleatorio (0,10)	Mockaroo

				Else 0	
<b>Años de uso de hormonales Terapia Hormonal de Reemplazo</b>	Cuantitativa	Años que la paciente ha utilizado hormonales como terapia hormonal de reemplazo	Entre 0 y 10  0= nunca	IF <b>Antecedente de uso de hormonales</b> = True Then aleatorio (0,10)  Else 0	Mockaroo
<b>Antecedentes familiares</b>	Cualitativa	Identificar si los familiares del paciente han presentado alguna de las enfermedades listadas	<ul style="list-style-type: none"> <li>• Desconocidos (Boolean)</li> <li>• Cáncer mamario (Boolean)</li> <li>• Diabetes mellitus (Boolean)</li> <li>• Presión alta (Boolean)</li> <li>• Cardiopatía (Boolean)</li> <li>• Oncológico primer grado (Boolean)</li> <li>• Oncológico no primer grado (Boolean)</li> </ul>	<b>Desconocidos</b> genera aleatoriamente (true or false)  El resto cada uno se rige por IF <b>Desconocidos</b> = true Then False  Else Aleatorio boolean	Mockaroo
<b>Mamografía anterior</b>	Cualitativa	Si la paciente tiene mamografías anteriores	<ul style="list-style-type: none"> <li>• True</li> <li>• False</li> </ul>	Distribución 28% true 72% false	Mockaroo
<b>Calidad de la imagen de mamografía</b>	Cuantitativa	Evaluar las condiciones en las que se encuentra la imagen de mamografía	Aleatorio entre: <ul style="list-style-type: none"> <li>• Muy buena</li> <li>• Buena</li> <li>• Regular</li> <li>• Baja</li> <li>• Muy baja</li> </ul>	n/a	Mockaroo
<b>Lesiones</b>	Cualitativa	Cuantificar lesiones potencialmente significativas en mamografía	Aleatorio entre: <ul style="list-style-type: none"> <li>• 1</li> <li>• 2</li> </ul>	n/a	Mockaroo
<b>Lesión principal</b>	Cualitativa	Identificar la lesión potencialmente más significativa.	Aleatorio entre: <ul style="list-style-type: none"> <li>• Izquierda</li> <li>• Derecha</li> <li>• Lesión principal</li> </ul>	Distribución 45% izq 45% der 10% calcificacion	

			son calcificaciones		
<b>Localización de lesión principal</b>	Cualitativa	Identificar la ubicación de la lesión principal	Aleatorio entre: <ul style="list-style-type: none"> <li>• Cse</li> <li>• Csi</li> <li>• Cie</li> <li>• Cii</li> <li>• Pezón y retroareolar</li> <li>• Axila</li> </ul>	Distribución: 50% en cse 15% csi 15% cie 8% cii 10% pezon 2% axila	Mockaroo
<b>Tamaño de la lesión principal</b>	Cuantitativa	Número, expresa el tamaño de la lesión	Entre 5 a 39 (mm)	Distribución normal	Python
<b>Forma de la lesión principal</b>	Cualitativa	Conocer la forma de la lesión	<ul style="list-style-type: none"> <li>• Ir a descripción de calcificaciones</li> <li>• Redonda</li> <li>• Ovalada</li> <li>• Lobulada</li> <li>• Irregular no espiculado (más común)</li> <li>• Irregular espiculado (más común)</li> </ul>	<b>IF Lesión principal =</b> Lesión principal son calcificaciones Then Ir a descripción de calcificaciones  Else Aleatorio entre el resto de los valores cualitativos	Mockaroo
<b>Densidad de la lesión principal</b>	Cualitativa	Conocer la densidad de la lesión	<ul style="list-style-type: none"> <li>• Ir a descripción de calcificaciones</li> <li>• = densidad del musculo</li> <li>• &lt; densidad del musculo (mas común)</li> <li>• &gt; densidad del musculo</li> </ul>	<b>IF Lesión principal =</b> Lesión principal son calcificaciones Then Ir a descripción de calcificaciones  Else Aleatorio entre el resto de los valores cualitativos	Mockaroo
<b>Márgenes de la lesión principal</b>	Cualitativa	Ver el margen de la forma de la lesión principal	<ul style="list-style-type: none"> <li>• Ir a descripción de calcificaciones</li> </ul>	<b>IF Lesión principal =</b> Lesión principal son	Mockaroo

			<ul style="list-style-type: none"> <li>• es</li> <li>• Circunscrito en al menos el 90% de la lesión</li> <li>• Obscurecido entre 50 y 90% de la lesión</li> <li>• Micro lobulado en más del 50% de la lesión</li> <li>• Mal definido en más del 50% de la lesión (más común)</li> <li>• Espiculado en más del 50% de la lesión (más común)</li> </ul>	<p>calcificaciones Then Ir a descripción de calcificaciones</p> <p>Else Aleatorio entre el resto de los valores cualitativos</p>	
<b>Calcificaciones asociadas a la lesión principal</b>	Cualitativa	Ver el tipo de calcificación asociada a la forma de la lesión principal	<ul style="list-style-type: none"> <li>• Ir a descripción de calcificaciones</li> <li>• Redondeadas</li> <li>• Centro lucido</li> <li>• Semilunas</li> <li>• Aisladas puntiformes</li> <li>• Burdos fragmentos de &gt;2 y &lt; 5mm</li> <li>• Burdos fragmentos grandes &gt; 5mm.</li> <li>• Pleomórficas área de al menos 5mm con al</li> </ul>	<p>IF <b>Lesión principal</b> = Lesión principal son calcificaciones Then Ir a descripción de calcificaciones</p> <p>Else Aleatorio entre el resto de los valores cualitativos</p>	Mockaroo

			<p>menos 5 microcalcificaciones (más común)</p> <ul style="list-style-type: none"> <li>• Vasculares</li> <li>• Cutáneas</li> <li>• Lineales (conductos) &gt; 5mm</li> <li>• Lineales (conductos) &lt; 5mm</li> <li>• No se aprecia calcificaciones</li> </ul>		
<b>Sucedáneos</b>	Cualitativa	Identificar propiedades de menor grado	<ul style="list-style-type: none"> <li>• No hay sucedáneos (Boolean)</li> <li>• Ganglio axilar de características benignas. (Boolean)</li> <li>• Ganglio axilar probablemente maligno (Boolean)</li> <li>• Ganglio axilar francamente maligno (Boolean)</li> <li>• Aglomerado de ganglios axilares malignos (Boolean)</li> <li>• Engrosamiento cutáneo asociado a la lesión principal (Boolean)</li> <li>• retracción cutánea asociada a la lesión principal</li> </ul>	<p><b>No hay sucedáneos genera</b> aleatoriamente (true or false)</p> <p>El resto cada uno se rige por</p> <p><b>IF No hay sucedáneos = true Then False</b></p> <p>Else Aleatorio boolean</p>	Mockaroo

			(Boolean) <ul style="list-style-type: none"> <li>• Presenta lesión asociada (Boolean)</li> </ul>		
<b>Lesión secundaria</b>	Cualitativa	Identificar la mama afectada	<ul style="list-style-type: none"> <li>• No hay lesión secundaria</li> <li>• Izquierda</li> <li>• Derecha</li> </ul>	IF <b>Lesiones</b> = 1 Then No hay lesión secundaria  Else Aleatorio entre el resto de las variables	Mockaroo
<b>Localización de lesión secundaria</b>	Cualitativa	Identificar la ubicación de la lesión principal	<ul style="list-style-type: none"> <li>• No hay lesión secundaria</li> <li>• Cse</li> <li>• Csi</li> <li>• Cie</li> <li>• Cii</li> <li>• Pezón y retroareolar</li> <li>• Axila</li> </ul>	IF <b>Lesiones</b> = 1 Then No hay lesión secundaria  Distribucion: 50% en cse 15% csi 15% cie 8% cii 10% pezon 2% axila	Mockaroo
<b>Tamaño de la lesión secundaria</b>	Cuantitativa	Medir el tamaño de la lesión	Entre 5 y 39 (mm) 0 no hay lesión secundaria	IF <b>Lesiones</b> = 1 Then 0  Else Distribucion normal	Mockaroo
<b>Forma de la lesión secundaria</b>	Cualitativa	Conocer la forma de la lesión	<ul style="list-style-type: none"> <li>• No hay lesión secundaria</li> <li>• Redonda</li> <li>• Ovalada</li> <li>• Lobulada</li> <li>• Irregular no espiculado (más común)</li> <li>• Irregular espiculado (más común)</li> </ul>	IF <b>Lesiones</b> = 1 Then No hay lesión secundaria  Else Aleatorio entre el resto de las variables	Mockaroo
<b>Densidad de la lesión secundaria</b>	Cualitativa	Conocer la densidad de la lesión	<ul style="list-style-type: none"> <li>• No hay lesión secundaria</li> <li>• = densidad del musculo</li> </ul>	IF <b>Lesiones</b> = 1 Then No hay lesión secundaria	Mockaroo

			<ul style="list-style-type: none"> <li>• &lt; densidad del musculo (mas común)</li> <li>• &gt; densidad del musculo</li> </ul>	<p>Else Aleatorio entre el resto de las variables</p>	
<b>Márgenes de la lesión secundaria</b>	Cualitativa	Ver el margen de la forma de la lesión principal	<ul style="list-style-type: none"> <li>• No hay lesión secundaria</li> <li>• Circunscrito en al menos el 90% de la lesión</li> <li>• Obscurecido entre 50 y 90% de la lesión</li> <li>• Micro lobulado en más del 50% de la lesión</li> <li>• Mal definido en más del 50% de la lesión (más común)</li> <li>• Espiculado en más del 50% de la lesión (más común)</li> </ul>	<p>IF Lesiones = 1 Then No hay lesión secundaria</p> <p>Else Aleatorio entre el resto de las variables</p>	Mockaroo
<b>Calcificaciones asociadas a la lesión secundaria</b>	Cualitativa	Ver el tipo de calcificación asociada a la forma de la lesión principal	<ul style="list-style-type: none"> <li>• No hay lesión secundaria</li> <li>• Redondeadas</li> <li>• Centro lucido</li> <li>• Semiluna</li> <li>• Aisladas puntiformes</li> <li>• Burdos fragmentos de &gt;2 y &lt; 5mm</li> <li>• Burdos fragmentos grandes &gt; 5mm.</li> </ul>	<p>IF Lesiones = 1 Then No hay lesión secundaria</p> <p>Else Aleatorio entre el resto de las variables</p> <p>Condicionar con "no se aprecian calcificaciones"</p>	Mockaroo

			<ul style="list-style-type: none"> <li>• Pleomórficas área de al menos 5mm con al menos 5 microcalcificaciones (más común)</li> <li>• Vasculares</li> <li>• Cutáneas</li> <li>• Lineales (conductos) &gt; 5mm</li> <li>• Lineales (conductos) &lt; 5mm</li> <li>• No se aprecia calcificaciones</li> </ul>		
<b>Calcificaciones como lesión principal</b>	Cualitativa	Ver el tipo de calcificación como lesión principal	<ul style="list-style-type: none"> <li>• Burdos fragmentos &gt; de 2 y &lt; 5mm</li> <li>• Burdos fragmentos grandes &gt; 5mm</li> <li>• Pleomórficas en un área de al menos 5mm con al menos 5 microcalcificaciones (mayoría)</li> <li>• Conductos lineales &gt; 5mm</li> <li>• Conductos lineales &lt; 5mm</li> <li>• No existen</li> </ul>	<b>IF Lesión Principal</b> = lesión principal son calcificaciones Then Aleatorio entre el resto de las variables (excepto "No existen")  Else No existen	Mockaroo
<b>Calcificación como lesión irrelevante</b>	Cualitativa	Ver calcificaciones como lesiones con poca relevancia	Aleatorio entre: <ul style="list-style-type: none"> <li>• Redondas (más común)</li> <li>• Centro lucido (más común)</li> </ul>	n/a	Mockaroo

			<ul style="list-style-type: none"> <li>• Semiluna (más común)</li> <li>• Aisladas puntiformes</li> <li>• Burdos fragmentos mayores de 2mm y menores que 5mm</li> <li>• Burdos fragmentos mayores que 5mm</li> <li>• Pleomórficas en un área de menos 5mm</li> <li>• Vasculares</li> <li>• Cutáneas</li> <li>• Conductos lineales mayores a 5mm</li> <li>• Conductos lineales menores a 5mm</li> <li>• No existen</li> </ul>		
<b>Distribución de la lesión principal</b>	Cualitativa	Ver la distribución de la lesión principal	<p>Aleatorio entre:</p> <ul style="list-style-type: none"> <li>• Solitaria</li> <li>• Agrupadas ocupan del 2 al 10% de la mama</li> <li>• Segmentarias entre 10 y 25% de distribución triangular</li> <li>• Regional o difusa entre el 25 y 40%</li> <li>• Generalizadas más del 40%</li> <li>• Aisladas menos del</li> </ul>	n/a	Mockaroo

			2%, <ul style="list-style-type: none"> <li>• Universales son bilaterales (en ambas mamas)</li> </ul>		
<b>Riesgo de lesión maligna (variable de salida para clasificar los casos)</b>	Cuantitativa	Probabilidad de cáncer de mama	Aleatorio entre: (escala liker) <ul style="list-style-type: none"> <li>• Definitivamente</li> <li>• Probable</li> <li>• Posible</li> <li>• Poco probable</li> <li>• Definitivamente no</li> </ul>	<b>IF lesión principal = Izquierda</b> Then 0  Else Aleatorio entre el resto de las variables	Mockaroo

