

UNIVERSIDAD NACIONAL JOSÉ FAUSTINO SÁNCHEZ CARRIÓN
FACULTAD DE INGENIERÍA INDUSTRIAL, SISTEMAS E INFORMÁTICA
ESCUELA PROFESIONAL DE INGENIERÍA INDUSTRIAL



TESIS

**MODELO DE UN SISTEMA BASADO EN CASOS
ORIENTADOS AL DIAGNÓSTICO MÉDICO DE UNA
ENFERMEDAD**

PRESENTADO POR:

Bach. Carlos Saturnino Barreto Vite

ASESOR:

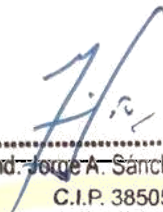
Mg. Jorge Antonio Sánchez Guzmán

Para Optar el Título Profesional de Ingeniero Informático

Huacho - Perú

2022

ASESOR Y MIEMBROS DEL JURADO



.....
Ing. Ind. Jorge A. Sánchez Guzmán

C.I.P. 38505

ASESOR



Ing. ALDO FELIPE LAOS BERNAL

PRESIDENTE



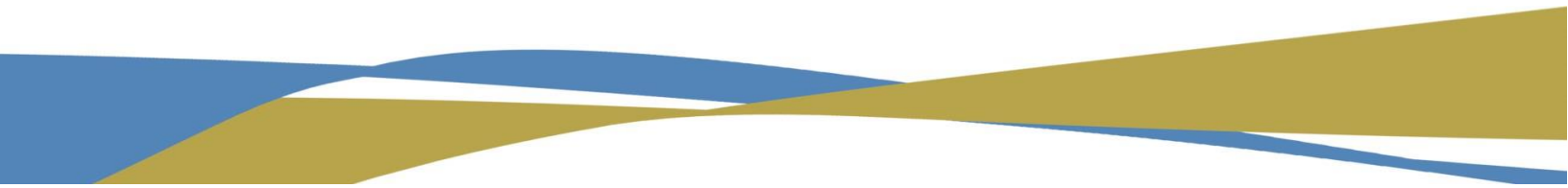
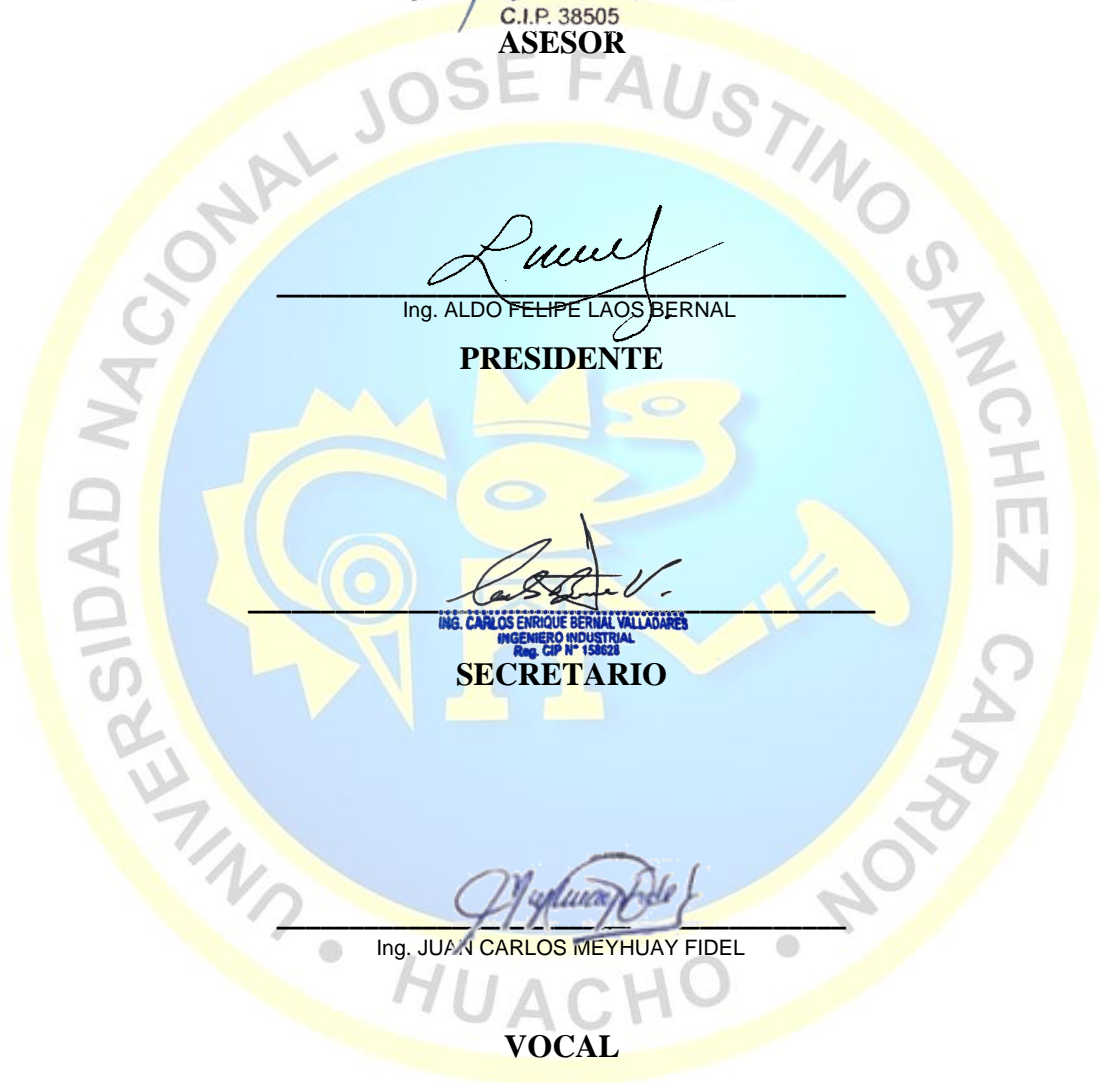
ING. CARLOS ENRIQUE BERNAL VALLADARES
INGENIERO INDUSTRIAL
Reg. CIP N° 158623

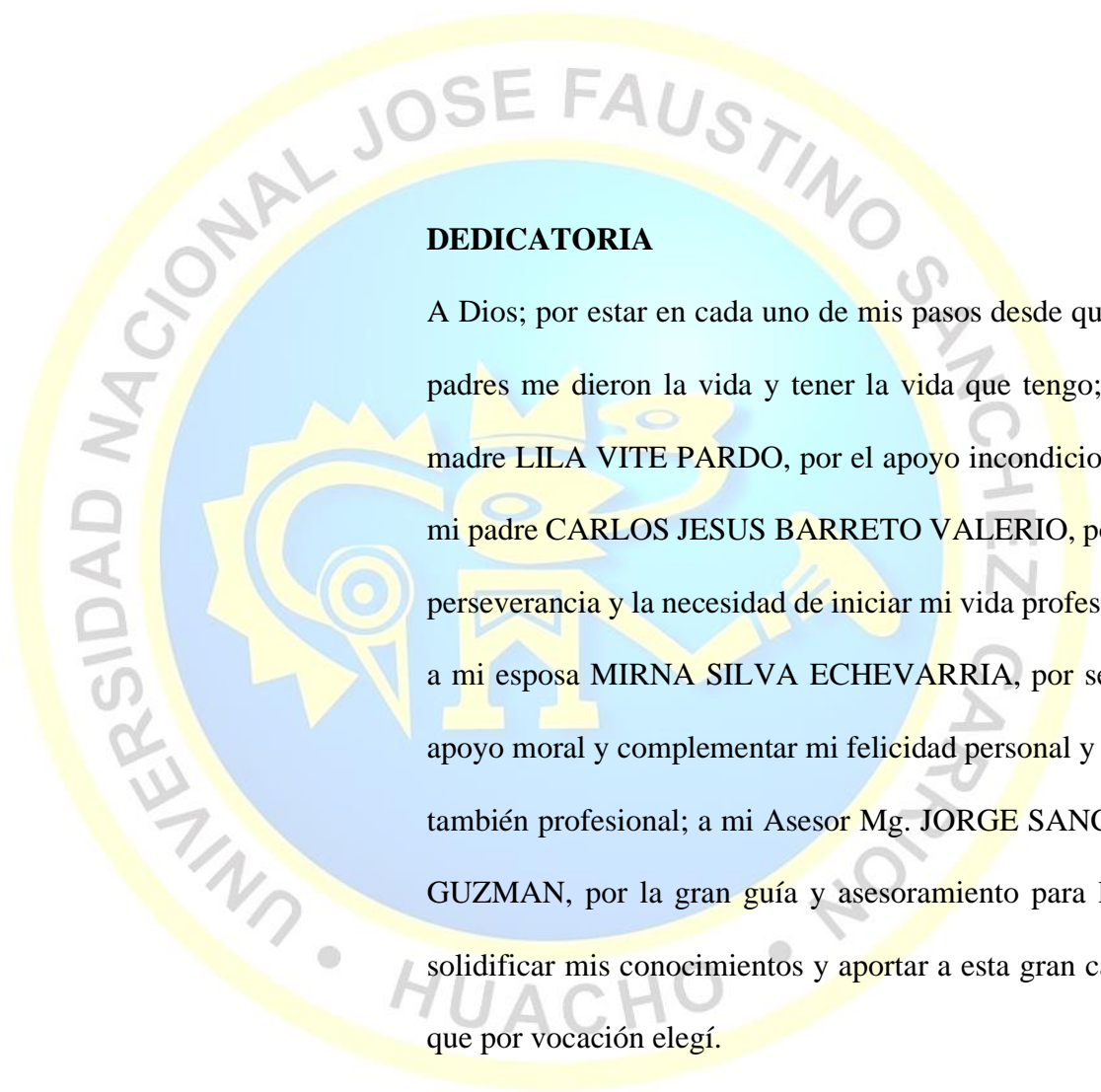
SECRETARIO



Ing. JUAN CARLOS MEYHUAY FIDEL

VOCAL





DEDICATORIA

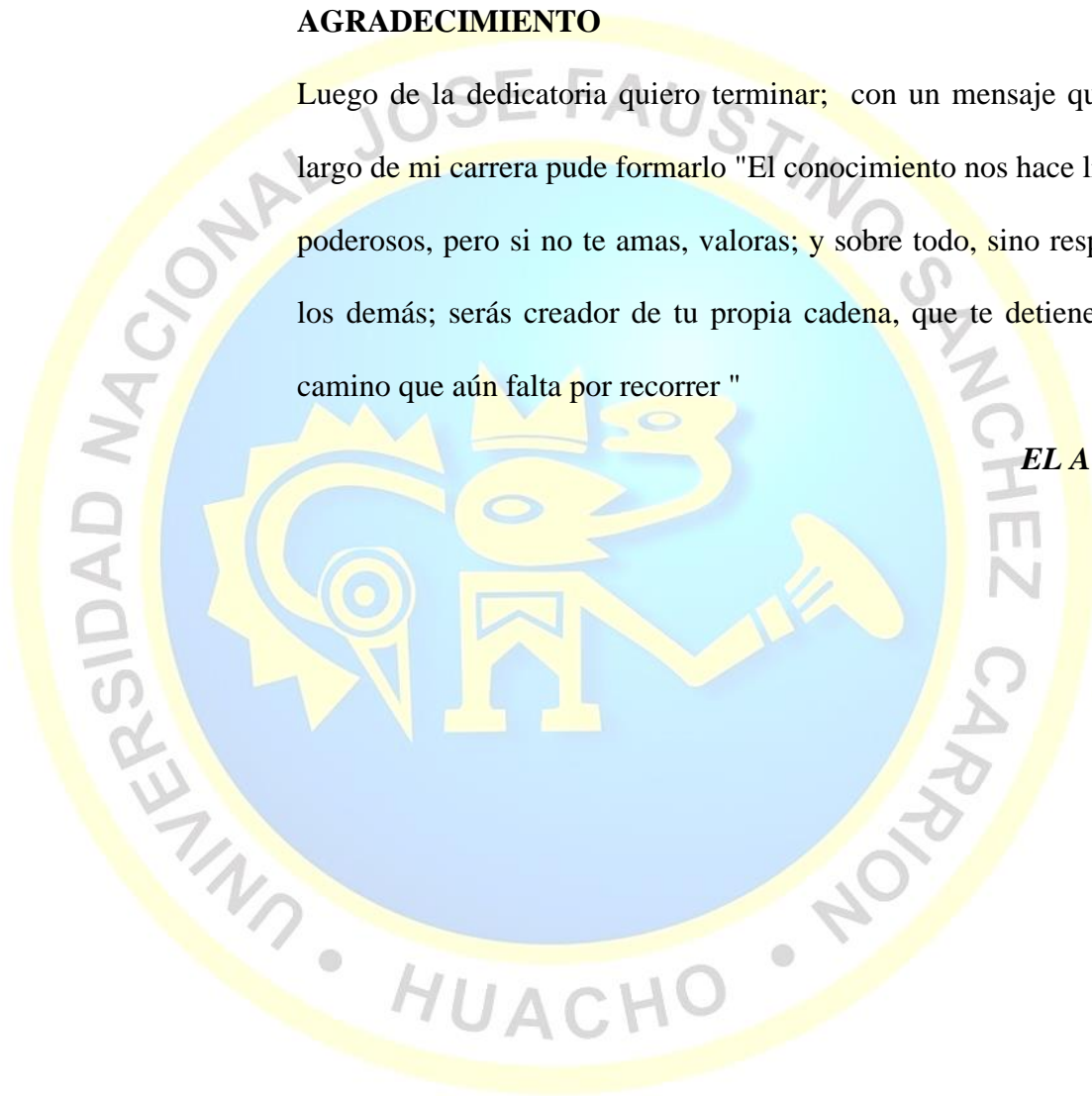
A Dios; por estar en cada uno de mis pasos desde que mis padres me dieron la vida y tener la vida que tengo; a mi madre LILA VITE PARDO, por el apoyo incondicional; a mi padre CARLOS JESUS BARRETO VALERIO, por esa perseverancia y la necesidad de iniciar mi vida profesional; a mi esposa MIRNA SILVA ECHEVARRIA, por ser ese apoyo moral y complementar mi felicidad personal y ahora también profesional; a mi Asesor Mg. JORGE SANCHEZ GUZMAN, por la gran guía y asesoramiento para lograr solidificar mis conocimientos y aportar a esta gran carrera que por vocación elegí.

EL AUTOR

AGRADECIMIENTO

Luego de la dedicatoria quiero terminar; con un mensaje que a lo largo de mi carrera pude formarlo "El conocimiento nos hace libres y poderosos, pero si no te amas, valoras; y sobre todo, sino respetas a los demás; serás creador de tu propia cadena, que te detiene en tu camino que aún falta por recorrer "

EL AUTOR



INDICE

ASESOR Y MIEMBROS DEL JURADO	ii
DEDICATORIA	iii
AGRADECIMIENTO	iv
INDICE	v
INDICE DE TABLAS	viii
INDICE DE GRÁFICOS	ix
RESUMEN	x
ABSTRACT	xi
INTRODUCCIÓN	xii
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	1
1.1. Descripción de la realidad problemática	1
1.2. Formulación del problema	6
1.2.1. Problema general.	6
1.2.2. Problemas específicos.....	6
1.3. Objetivos de la investigación	6
1.3.1. Objetivo general.....	6
1.3.2. Objetivos específicos.....	6
1.4. Justificación de la investigación.....	6
1.5. Delimitación del estudio.....	9
1.6. Viabilidad del estudio.....	10
CAPÍTULO II: MARCO TEÓRICO.....	12
2.1. Antecedentes de la investigación	12
2.1.1. Antecedentes internacionales	12
2.1.2. Antecedentes nacionales.....	16
2.2. Bases teóricas	20
2.3. Bases filosóficas.....	29
2.4. Definición de términos básicos	31

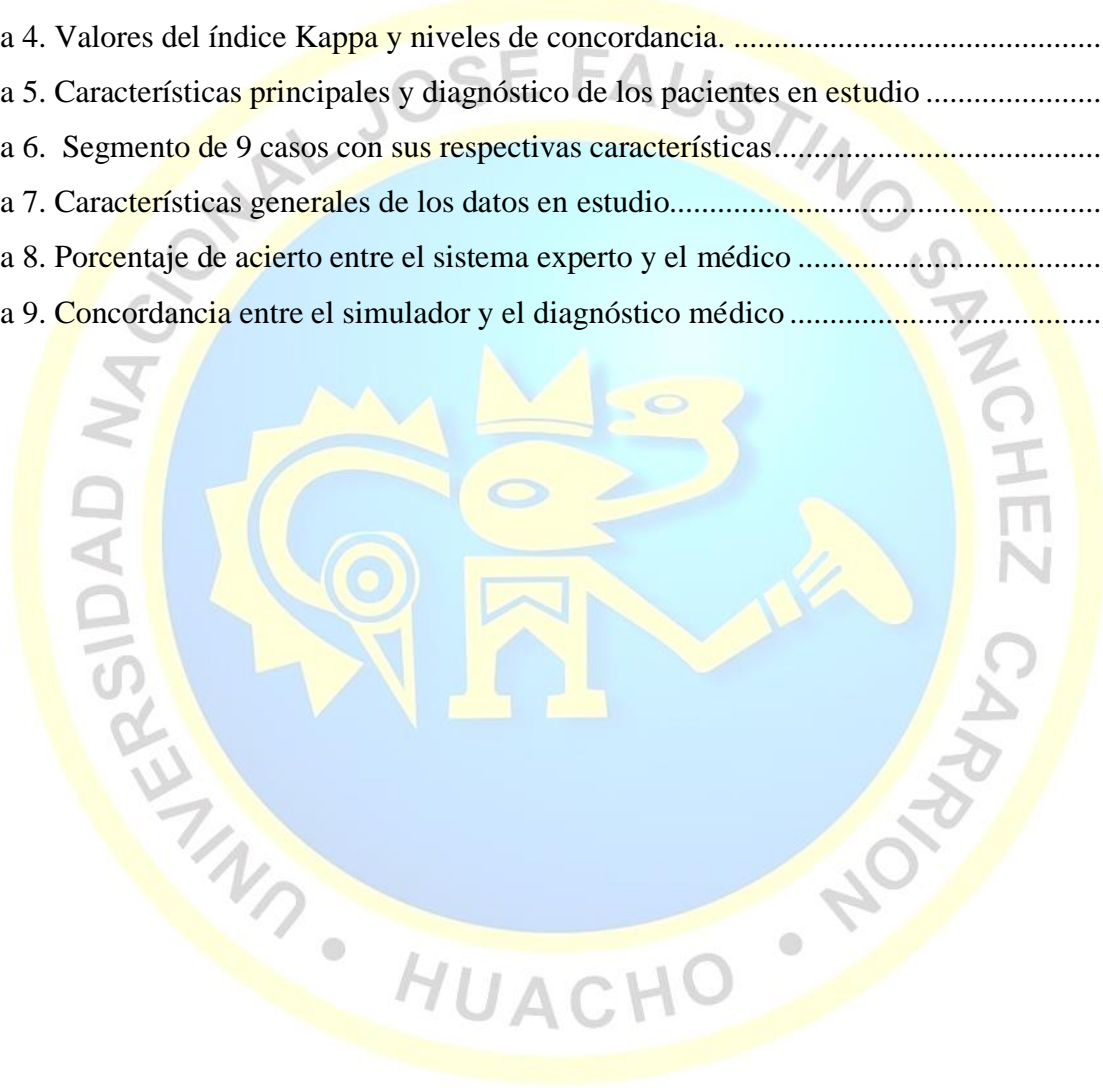
2.5.	Hipótesis de investigación.....	34
2.5.1	Hipótesis general	34
2.5.2	Hipótesis específicas	34
2.6.	Operacionalización de las variables	34
CAPÍTULO III: METODOLOGÍA.....		36
3.1.	Diseño metodológico.....	36
3.1.1.	Tipo.....	36
3.1.2.	Nivel y Enfoque	36
3.2.	Población y muestra	36
3.3.	Técnica de recolección de datos	36
3.4.	Técnicas para el procesamiento de la información	38
CAPÍTULO IV: RESULTADOS.....		42
4.1.	Análisis de resultados.....	42
4.1.1	Descripción del modelo propuesto	42
4.1.2	Descripción de los datos	44
4.1.3	Importación de los datos y creación de la clase paciente.....	48
4.1.4	Explicaciones que brinda el modelo dentro del software myCBR.....	52
4.1.5	Consulta de los casos	53
4.1.6	Concordancia entre el modelo expuesto y el diagnóstico médico	54
4.2.	Contrastación de las hipótesis	55
4.2.1.	Análisis para la contrastación de la Hipótesis General	55
4.2.2.	Análisis para la contrastación de las Hipótesis Específicas.....	55
CAPÍTULO V: DISCUSIÓN		58
5.1.	Discusión de resultados.....	58
CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES.....		60
6.1.	Conclusiones	60
6.2.	Recomendaciones.....	60

REFERENCIAS O FUENTES DE INFORMACIÓN	62
Fuentes documentales	62
Fuentes electrónicas	63
ANEXOS	64
Anexo 01. Matriz de consistencia	65
Anexo 02. Base de datos (Pima Indians).....	66



INDICE DE TABLAS

Tabla 1. Operacionalización de las variables.....	34
Tabla 2. Evaluación de los síntomas y clasificación del riesgo.....	37
Tabla 3. Comparación entre el experto médico y el sistema RBC	40
Tabla 4. Valores del índice Kappa y niveles de concordancia.	41
Tabla 5. Características principales y diagnóstico de los pacientes en estudio	45
Tabla 6. Segmento de 9 casos con sus respectivas características.....	46
Tabla 7. Características generales de los datos en estudio.....	47
Tabla 8. Porcentaje de acierto entre el sistema experto y el médico	54
Tabla 9. Concordancia entre el simulador y el diagnóstico médico	56



INDICE DE GRÁFICOS

Figura 1 Razonamiento basado en casos y el ciclo de consulta para un nuevo caso	27
Figura 2. Modelo para el sistema de diagnóstico y sus relaciones	43
Figura 3. Clasificación de los pacientes por tipo de diabetes	47
Figura 4. Importación de la base de datos desde el software Protégé.....	49
Figura 5. Clase paciente y sus respectivos slots (características)	50
Figura 6. Ejemplo de una característica del paciente.....	50
Figura 7. Similitud para la característica glucosa	52
Figura 8. Recuperación de casos similares a la consulta	54



RESUMEN

El estudio que conlleva esta tesis denominada “modelo de un sistema basado en casos orientados al diagnóstico médico de una enfermedad”, tiene como **objetivo** establecer un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad, que para nuestro caso fue la diabetes mellitus. **Metodología:** Investigación de tipo tecnológica, con nivel descriptivo analítico, con enfoque cualitativo y cuantitativo, cuantificando el nivel de concordancia entre los resultados del modelo y el diagnóstico médico. Se tomó la base de casos de Pima Indians Database sobre la enfermedad diabetes que contiene 769 casos coleccionados de pacientes con las características de síntomas y afecciones. La técnica empleada estuvo referida a el razonamiento basado en casos (RBC), en sus diferentes fases. **Resultados:** El modelo construido contiene el ciclo RBC orientado al diagnóstico médico, desarrollándose un aplicativo mediante los softwares integrados en Plugin Protégé y myCBR, los cuales permiten hacer la consulta de un caso médico y el software reporta que casos son los mas similares al consultado **Conclusión:** Es perfectamente factible la construcción de este modelo, llegando a tener una concordancia bastante alta con el diagnóstico médico, medido a través del índice Kappa de Cohen con un p-valor = $0,000 < 0,01$, el cual es altamente significativo..

Palabras claves: Razonamiento basado en casos, similitud, revisar, reutilizar, concordancia

ABSTRACT

The study that entails this thesis called "model of a system based on cases oriented to the medical diagnosis of a disease", has an **objective**: to establish a model of a system based on cases oriented to the medical diagnosis of a disease, which in our case was the Mellitus diabetes.

Methodology: Technological research, with an analytical descriptive level, with a qualitative and quantitative approach, quantifying the level of agreement between the results of the model and the medical diagnosis. The case base of the Pima Indians Database on the disease diabetes was taken, which contains 769 cases collected from patients with the characteristics of symptoms and conditions. The technique used was referred to case-based reasoning (RBC), in its different phases. **Results**: The built model contains the RBC cycle oriented to medical diagnosis, developing an application through the software integrated in Plugin Protégé and myCBR, which allow consulting a medical case and the software reports which cases are the most similar to the one consulted. **Conclusion**: The construction of this model is perfectly feasible, reaching a fairly high concordance with the medical diagnosis, measured through Cohen's Kappa index with a p-value = $0.000 < 0.01$, which is highly significant.

Keywords: Case-based reasoning, similarity, review, reuse, concordance

INTRODUCCIÓN

La presente investigación tiene como objetivo crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad, que para nuestro caso fue la diabetes mellitus.

En el desarrollo de esta investigación, se analizaron algunos estudios a nivel internacional y nacional, como antecedentes sobre el tema. Asimismo, la investigación está dividida en los siguientes capítulos:

En el capítulo I. Se presenta el planteamiento del problema, y la realidad problemática, así como el planteamiento de los objetivos, justificación, delimitación y viabilidad de estudio.

En el capítulo II. Se presenta el marco teórico, fundamental para entender el tema de investigación, el cual comprende los antecedentes, bases teóricas, definiciones conceptuales y las respectivas hipótesis planteadas.

En el capítulo III. En este capítulo se presenta la metodología del trabajo de investigación comprendido en el diseño metodológico, la base de datos de estudio, operacionalización de las variables y técnica de procesamiento de la información en los softwares Protégé y myCBR, donde se implementó el modelo

En el capítulo IV. Contiene los resultados sintetizados en la presentación de los tablas, figuras e interpretaciones y el proceso de contraste de las hipótesis.

En el capítulo V. Se muestra la discusión de la investigación, las conclusiones y recomendaciones finales.

CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la realidad problemática

Existen variadas aplicaciones de sistemas informáticos aplicados a ciencias de la salud o más específicamente a la medicina, esto permite que los nuevos métodos asistidos por computadora permitan la efectividad en cuanto a almacenamiento y gestión administrativa de bases de datos se refiere; como el manejo de historias clínicas, consultas en tiempo real, etc. Asimismo, por otro lado, existen las aplicaciones orientadas al apoyo de la investigación; como por ejemplo, en la biomedicina, en la robótica para prótesis, como apoyo al tratamiento, etc. y que mayormente se efectúa con un sistema experto (Hersh, 2002).

Así pues, en la actualidad el conocimiento sobre técnicas de inteligencia artificial (IA) se está expandiendo rápidamente y con variadas aplicaciones en las diferentes áreas, siendo una de ellas el área de la medicina, quien a su vez necesita aplicaciones para su mejor desarrollo como por ejemplo en lo que a diagnóstico de una enfermedad se refiere lo que hace deseable un sistema de diagnóstico asistido por computadora. Dicho sistema puede brindarle al médico una segunda opinión. Los recientes avances en Inteligencia Artificial ofrecen métodos y técnicas con el potencial de resolver tareas que antes eran difíciles de resolver con sistemas informáticos en dominios médicos. La investigación en todo el mundo se centra en las nuevas aplicaciones en el campo médico y, en particular, en el diagnóstico.

El uso de computadoras para ayudar a los médicos con el diagnóstico de cualquier enfermedad se ha intentado de diversas maneras. La inteligencia artificial (IA) y la teoría

del control se han aplicado a la detección y gestión de problemas, mientras que se han empleado varias formas de telemedicina para recopilar una mayor variedad de datos y / o aumentar la comunicación entre pacientes y médicos (Parker y otros, 2011)

El apoyo del diagnóstico es especialmente útil para aquellos médicos que se están iniciando y que no tienen experiencia o para aquellos que no son especialistas en tal o cual especialidad. Así pues, se presentan algunos casos al médico y como profesional desea conocer la respuesta de otros pacientes a un determinado tratamiento; sin embargo, no se cuenta con una base de casos, tampoco de tratamientos que hayan permitido registrar la práctica médica como diagnóstico y tratamiento de otros especialistas médicos, describiendo detalladamente a la enfermedad, en cuanto al tratamiento, los efectos, la persistencia de la enfermedad y otras características, que permita al médico decidir acertadamente en cuanto al diagnóstico y por ende al posterior tratamiento. Al analizar los casos similares, el especialista diagnostica de manera eficiente la enfermedad dando el adecuado tratamiento, a través de la indicación de los síntomas por parte del paciente; es decir se comparan estos con el conjunto de síntomas almacenados en la base de casos, contribuyendo así a obtener un diagnóstico mas eficiente.

Precisamente, el sistema de Razonamiento Basado en Casos (RBC), permite el almacenamiento como si fuera una base de datos, pero ahora es una base de casos. Cada paciente es un caso, con su respectivo diagnóstico y tratamiento con características propias del mismo, como síntomas, medicamentos, efectos del tratamiento, etc. y cada una de estas características podrían ser inclusive ponderadas y que sirven como razonamiento para un nuevo caso. Así pues, el RBC es adecuado en el dominio médico, especialmente por su modelo cognitivamente adecuado, la facilidad para integrar diferentes tipos de conocimiento y su representación de casos, que es posible obtener de

los registros de los pacientes. En particular, el diagnóstico de un paciente en el ámbito médico depende de la experiencia. Un médico recién egresado puede comenzar su práctica con alguna experiencia inicial (casos resueltos), luego tratar de utilizar esta experiencia pasada para resolver un nuevo problema y al mismo tiempo aumentar sus experiencias; es decir, su base de casos (Bichindaritz y Marling, 2006).

Es decir, en el campo médico, el RBC se modela para todas las tareas de diagnóstico, tratamiento, planificación y asistencia, se logra mediante la recopilación de casos y se implementa mediante la identificación de características significativas que construyen un caso, de la misma manera que los médicos clínicos se basan en experiencias pasadas para determinar uno nuevo. El RBC muestra el estado del paciente como método de aprendizaje en el área de IA y almacena los casos referidos como ejemplos de entrenamiento para luego acceder a ellos para resolver nuevos problemas o para obtener una predicción extrayendo aquellos casos que son similares o cercanos a los nuevos, y esto se logra acumulando datos para construir un modelo de comprensión de un nuevo problema almacenando las condiciones del nuevo problema, luego se establece una colección de reglas para una situación dada, tomando una o más acciones para resolverla.

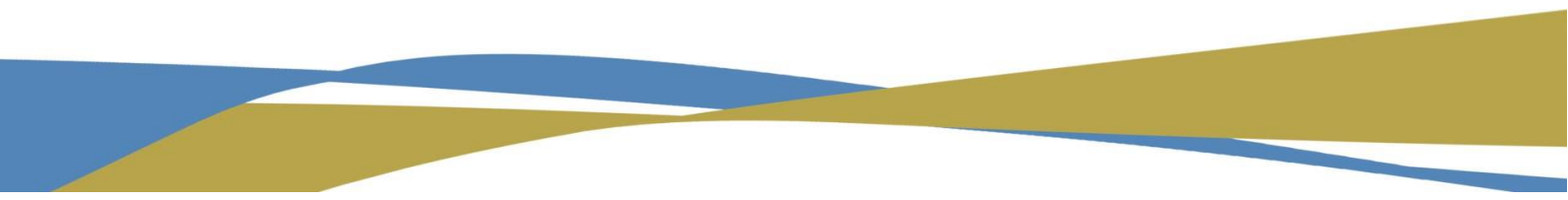
El sistema luego se enfoca en reconocer la similitud de nuevos problemas con los existentes, donde un caso es un registro de una situación pasada o un problema que ya ha sido resuelto, información que se registra considerando el dominio con el que se asociará el caso, especificaciones detalladas de la situación, descripción de todos los atributos y circunstancias del entorno, junto con la solución de ese caso con todos los hechos y los pasos que lo hacen una solución exitosa y un caso resuelto, y se puede representar en diferentes tipos y formatos, lo que hace que el RBC sea el método más flexible (Aamodt, 2004).

Este sistema podría ser especialmente útil en aquellas localidades alejadas de nuestro país, como por ejemplo en entornos rurales, donde mayormente no se encuentran hospitales sino mas bien las denominadas postas médicas mal implementadas y con poca atención por parte de los especialistas de salud, donde casi siempre va un serumista (profesional médico asignado al servicio de comunidades) quien se ve en la obligación de atender diferentes enfermedades aun no siendo un especialista y que bien podría apoyarse en este sistema experto. Así pues, en el entorno rural, son muchos y variados los casos que se le presentan al médico y éste desea conocer la similitud de síntomas de otros pacientes; sin embargo, en esas zonas el historial clínico es deficiente por lo que se desea tener un repositorio de casos donde se registre la práctica médica de médicos especialistas y su procedimiento para el respectivo tratamiento.

Por ejemplo, por ámbito geográfico, es en la Selva donde se encuentra mayor diferencia de 53,4 puntos porcentuales entre las personas con (83,5%) y sin (30,1%) discapacidad que padecen enfermedad o malestar crónico. En la Sierra, la brecha es la segunda más grande con una diferencia de 53,3 puntos porcentuales (INEI, 2019); aquí según estos índices podría muy bien ser adecuado el uso de este sistema experto, porque sirve la pregunta la discapacidad es de nacimiento o adquirida, pudiendo evitarse con un adecuado diagnóstico e inmediato tratamiento brindado pro este sistema experto como apoyo a los especialistas de salud en esos entornos.

Lo expresado entonces, el planteamiento del problema para esta investigación de tesis surge ante la necesidad de apoyar a los especialistas de salud quienes se encuentran en lugares bastante alejados de urbes o ciudades metrópolis, zonas donde la salud de los ciudadanos corre peligro, y que muy bien este sistema experto puede inclusive servir de

prevención al apoyar tempranamente en el diagnóstico del médico las dolencias mas simples que luego pueden llegar a ser mayores y convertirse en enfermedades crónicas.



1.2. Formulación del problema

1.2.1. Problema general.

¿Se puede crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad?

1.2.2. Problemas específicos.

1. ¿Se puede crear un entorno de almacenamiento en software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos?
2. ¿Existe concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico?

1.3. Objetivos de la investigación

1.3.1. Objetivo general.

Crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad.

1.3.2. Objetivos específicos.

1. Crear un entorno de almacenamiento en un software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos.
2. Determinar la concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico.

1.4. Justificación de la investigación

- **Conveniencia:**

La creación de un módulo simulado del sistema basado en casos para el diagnóstico de una enfermedad es conveniente, porque es una fuente de consulta del personal de salud para comparar el diagnóstico del médico o

personal de salud (que inclusive puede ser una enfermera) con la base de casos, aprovechando las experiencias no de uno sino de varios profesionales de salud respecto a una dolencia en particular y resolver el problema de diagnóstico. A parte, se puede decir que la actuación de un especialista en salud siempre tiene un nivel de aprendizaje, pero el razonamiento clínico se mantiene como un elemento clave en cualquiera de esas actuaciones; sin embargo, dar cátedra sobre estas habilidades cognitivas no es sencillo, entonces también este módulo puede servir como fuente de capacitación en alumnos que están llevando algunas de las carreras profesionales orientadas a ciencias de la salud, permitiéndoles cimentar sus conocimientos prácticos al comparar los propios con las experiencias de personal calificado y que están almacenadas como casos. Resumiendo, entonces en cuanto a conveniencia, el estudio se justifica por los siguientes motivos:

- En el aspecto de enseñanza-aprendizaje de la medicina, para los estudiantes de ciencias de la salud, pues se puede afirmar que este módulo se ajusta muy bien como consulta o apoyo en comparar los síntomas reales de un paciente con los almacenados en una base de casos, para luego concluir con tal o cual diagnóstico.
- En el aspecto social, porque si es utilizado como módulo de apoyo por un profesional de la salud destacado en zonas alejadas de las capitales de provincia, habrá un beneficio para aquellos pobladores de estos lugares, pues éstos sentirán la confianza necesaria en dichos profesionales, al recibir un diagnóstico certero y por tanto seguir con el tratamiento adecuado, inclusive podría servir como prevención para

estos pobladores alejados quienes cuya localidad de vivienda se encuentra a varios kilómetros de distancia de un centro de salud verdaderamente equipado.

- En el aspecto de la salud, porque este sistema experto servirá como soporte al médico quien conjuntamente con sus conocimientos médicos lograría relacionar mejor sus experiencias y dar un mejor diagnóstico, inclusive podrá incrementar la base de casos con los pacientes que se va recibiendo.

En el aspecto económico, porque en muchos lugares alejados, existen enfermedades comunes que fácilmente se pueden prevenir, y que al no ser diagnosticadas a tiempo y por lo tanto no tener el tratamiento adecuado, se vuelven endémicas, siendo un gasto para el Estado, cuyo ministerio de salud debe de tener también un carácter preventivo, pues esto disminuiría el gasto por una enfermedad que pudo haberse diagnosticado tempranamente y que luego se puede transformar en incurable,

- **Relevancia:**

El concretar este estudio es relevante dentro del ámbito social, porque será de ayuda tanto para la parte académica como se dijo anteriormente, así como a la sociedad al servir de apoyo al personal de salud en aquellas comunidades alejadas, permitiendo que la calidad de salud mejore, observando una mayor confianza por parte del poblador hacia este profesional, al reconocer que le van a brindar una atención acorde a un diagnóstico adecuado y por lo tanto un tratamiento adecuado. Es decir, con este módulo RBC, el apoyo no sólo es para

el diagnóstico, sino que puede servir como acción de salud preventiva en dichas comunidades.

- **Valor teórico**

El objetivo fundamental de la docencia en medicina es promover el desarrollo del razonamiento experto de sus estudiantes, pero muchas veces el mismo docente así sea experimentado, no es consciente del razonamiento que se usa para llegar a un diagnóstico certero. Con este módulo RBC, entonces puede brindar una mejor enseñanza al tener el apoyo correspondiente.

A parte, el concretar similitudes entre la decisión del médico y lo que arroja este módulo involucra teorías matemáticas que, aunque simples, pueden también servir a los estudiantes de ciencias, en especial a los de matemáticas o estadística, para mejorar la metodología del análisis de decisión a través de las funciones de similitud, como son las distancias euclidianas, el método del vecino mas cercano, similitud bayesiana, etc. permitiendo afinar mejor los resultados en una decisión.

1.5. Delimitación del estudio

- **Delimitación espacial**

Esta tesis por ser de tipo tecnológico y orientado a simular un módulo basado en casos, no tiene una circunscripción referente a un área espacial geográfica; si en cambio se trabajará con algunas bases de datos tomándolas como casos que pertenecen a páginas oficiales tanto internacionales como nacionales como por ejemplo, la base de datos de Pima Indians Diabetes Database (Sigillito,

2014) o la de plataforma de datos abiertos a nivel nacional (<https://www.datosabiertos.gob.pe/>).

- **Delimitación temporal**

Los datos recopilados de la base de datos indicada están comprendidos en el periodo entre enero 2016 y 2019

- **Delimitación de contenido**

Específicamente el contenido de estudio de esta tesis se centra en la utilización de una parte de la inteligencia artificial denominada razonamiento basado en casos, cuyo ciclo está dividido en cuatro procesos bien definidos como son: recordar la similitud de los casos, utilizar el conocimiento almacenado de un caso para la solución de uno en particular, luego examinar la propuesta de solución y finalmente retenerla para su utilización posterior.

1.6. Viabilidad del estudio

- **Viabilidad temática**

Por ser una investigación tecnológica orientada a la construcción de un simulador de un sistema experto orientado finalmente al diagnóstico de una enfermedad, utilizando el RBC, cuya metodología de razonamiento basado en casos, está ampliamente documentada desde hace dos décadas, por lo cual en lo que a teoría se refiere es totalmente factible.

De igual manera en lo referente a la utilización de softwares, tanto MyCBR como PROTÉGÉ que serán utilizados, éstos pertenecen a la categoría de open source y además de ser libres, existe suficiente información en fuentes primarias y secundarias sobre el tema en mención y relacionados con estos

softwares, por lo que su manipulación será bastante cómoda para el usuario final y de ejecución totalmente factible.

En lo que respecta a los recursos económicos para la solvencia de esta tesis, éstos serán íntegramente asumidos por el autor, pues no se requiere de un alto financiamiento más que el necesario en lo referente a impresión y medios de tecnología digital al alcance propio así como el apoyo de asesoría correspondiente, por lo que igualmente es totalmente viable.



CAPÍTULO II: MARCO TEÓRICO

2.1. Antecedentes de la investigación

2.1.1. Antecedentes internacionales

Paruchuri¹ y Granville (2020) con su investigación titulada: “Un sistema de razonamiento basado en casos para ayudar a los médicos en la toma de decisiones”. Perteneciente al Colegio de Medicina la Universidad del Estado de Florida. Objetivo: Desarrollar un sistema de razonamiento basado en casos (RBC) para abordar la organización eficaz y la recuperación de información vital del paciente para ayudar a los médicos a tomar decisiones. Material y Métodos: Investigación de simulación de un entorno de prueba RBC mediante un nuevo algoritmo como método para recuperación de nuevos casos y recomendar a los médicos. Resultados: El sistema considera un elemento como un tratamiento en si mismo y puede tratarse como un caso que involucra varios pasos o estados pero, en este sistema, se representa como un solo elemento, además se probó en un entorno simulado y los resultados demuestran que el sistema puede adaptarse a cambios como nuevos procedimientos médicos o tratamientos que tienen lugar en el campo médico.

Rumui y otros (2018), con su artículo de investigación sobre “CBR denominado Razonamiento basado en casos para el diagnóstico de enfermedades por accidente cerebrovascular”. Objetivo: Implementar el concepto RBC utilizando puntaje siriraj, índice denso y método de coeficiente de Jaccard para realizar el cálculo de similitud entre casos. Material y Métodos: Investigación tecnológica de corte transversal. Los casos que se utilizaron en este estudio son los datos de registros médicos de pacientes con accidente cerebrovascular obtenidos de DKT Dr. Soetarto, Hospital, Yogyakarta. Los casos totales

son 180 casos que consisten en 112 casos de accidente cerebrovascular isquémico (56 casos de trombótico y 56 casos embólicos), 56 casos de accidente cerebrovascular hemorrágico (28 casos subaracnoideos y 28 casos intracelulares) y 12 casos sin accidente cerebrovascular. La prueba se realiza utilizando el método de validación cruzada de k-veces con 180 casos divididos en 4 veces. Resultados: El estudio dio como resultado un sistema RBC para el diagnóstico de accidentes cerebrovasculares mediante la aplicación del método del coeficiente de jaccard calculando la similitud entre los casos y la puntuación de siriraj para distinguir los tipos de accidentes cerebrovasculares isquémicos y hemorrágicos, y también utilizando el índice denso como indexación para realizar la recuperación de casos de la base de datos. Los resultados de la prueba que utilizan el método de validación cruzada de k-veces indican que el sistema RBC sin indexación ofrece una mejor especificidad y precisión que el sistema RBC que utiliza indexación para la aplicación del valor umbral de similaridad 0,65, 0,7, 0,75 y 0,8. Finalmente, se concluye que la prueba de tiempo de recuperación con una base de datos de 135 casos indica que el sistema RBC que usa indexación requiere un tiempo más rápido en recuperar el caso que el sistema RBC sin indexación, donde el sistema RBC que usa indexación requiere 9.099 segundos de tiempo promedio total mientras que el sistema CBR sin indexación necesita 16,958 segundos de tiempo total de recuperación.

Addisu y otros (2020), en su investigación sobre un nuevo entorno de software para el diagnóstico de la malaria, perteneciente a la Universidad de Mekelle-Etiopía, denominada Case-Based Reasoning Framework for Malaria Diagnosis. Objetivo: Explorar la aplicabilidad del razonamiento basados en casos comparando diversos entornos con el presentado en esta investigación en el dominio de la salud para mejorar el problema de toma de decisiones en el diagnóstico de malaria en Tigray aumentando el

número de casos y seleccionando los atributos más relevantes para verificar el efecto en el marco RBC para el diagnóstico de esta enfermedad en todos los aspectos de evaluación.

Material y Métodos: Investigación de tipo tecnológica, no experimental y de corte transversal. La técnica utilizada fue el de medir el rendimiento del entorno, con análisis estadístico, pruebas de aceptación del usuario, pruebas de similitud de casos, y evaluación comparativa. En cuanto a los datos, el proceso de recopilación de los mismos implica el análisis de documentos de diferentes fuentes y entrevistas con expertos en el dominio de esta enfermedad en el Hospital Especializado de Abera, Kahasay y Ayder, así como entrevistas como medio de discusión detallada entre el investigador y los expertos, seleccionándolos en función de diferentes criterios, como la experiencia en el dominio, el rango académico y otros.

Resultados: Específicamente para el parámetro de evaluación de las pruebas de aceptación del usuario "¿Es el nuevo entorno CBRFMD aplicable para el diagnóstico de malaria?", 50 %, 37,5 % y 12,5 % de los expertos califican el marco diseñado como excelente, muy bueno y bueno, respectivamente. Esta calificación de los encuestados muestra que el marco podría considerarse como un remedio en las instituciones de salud de Tigray y podría mejorar el servicio de diagnóstico de la malaria.

Kouser y otros (2018), en su investigación sobre un sistema de predicción de enfermedades cardíacas utilizando redes neuronales artificiales, función de base radial y razonamiento basado en casos, perteneciente a Kingston Engineering College, en la Universidad de Tamilnadu. **Objetivo:** proponer este sistema como apoyo a los médicos no solo a diagnosticar y predecir la enfermedad cardíaca al lograr niveles de precisión, sino que también ayuda a recetar el medicamento con éxito de acuerdo con la enfermedad predicha. **Material y Métodos:** Investigación no experimental, de tipo tecnológica. Se utilizó dos metodologías ANN (redes neuronales artificiales) al probar los conjuntos de

datos, búsqueda de similitud de imágenes RBC (razonamiento basado en casos) al mapear las similitudes de las imágenes de pacientes antiguos almacenados en la base de datos para la predicción de enfermedades cardíacas. Los datos corresponden a 910 registros con 16 aspectos médicos, se tomaron de paciente cardiacos del Hospital de Cleveland. El conjunto de datos consta de 78 atributos sin procesar. Se almacenaron en la base de datos los registros médicos sobre la enfermedad, considerándolos como los casos en registros antiguos de pacientes con enfermedades del corazón,. Cada registro consta de (nombre del paciente, número del paciente, síntomas, tipo de enfermedad cardíaca, nombre del médico, nombre del hospital, recetas y, sobre todo, imágenes del corazón afectado, etc.). Resultados: Se obtuvo una precisión del 97 %, observando que el RBC se adapta utilizando los registros de pacientes antiguos y también sirvió para predecir el tipo de enfermedad cardíaca. Con esta salida de CBR que no solo tiene el tipo de enfermedad cardíaca sino también el medicamento recetado, se utiliza para conocer el medicamento comparándolo con el medicamento original y los medicamentos administrados por RBF (Función de base radial), obteniéndose un 98% de acierto.

Brnamini y otros (2018), en su investigación titulada Diagnóstico de diabetes por razonamiento basado en casos y lógica comparativa, perteneciente a la Universidad de Oran, Algeria. Objetivo: El objetivo de este trabajo es demostrar el valor de un sistema de inferencia guiado por minería de datos en el modelado razonamiento basado en casos (RBC). Se trata de la originalidad de la integración de diferentes técnicas derivadas de la inteligencia artificial (IA) en las etapas de RBC y su capacidad para utilizar conjuntamente la experiencia y el aprendizaje de los datos en un mismo contexto, especialmente que las herramientas de inteligencia artificial se conviertan en parte de la diabetes. cuidado de la salud. Material y Métodos: Investigación no experimental, de tipo

tecnológica y comparativa, concentrándose en la etapa de recuperación de casos por diferentes técnicas, de las cuales se compararán las literaturas sobre este tema, sobre todo se propone una metodología de un plan de seguimiento y tratamiento *Y* que toma sus valores del conjunto de planes de cuidado de la diabetes. para tomar la glucemia todos los días. Resultados: Se desarrolló una aplicación denominada FDTCBR, cuyo propósito es mejorar la precisión de la clasificación de la Diabetes, integrando un árbol de decisión difuso para mejorar el paso de recuperación de casos del proceso de RBC..

2.1.2. Antecedentes nacionales

Capuñay (2021) en su investigación de tesis para obtener el título de ingeniero de sistemas y computación de la Universidad Católica Santo Toribio de Mogrovejo de la Ciudad de Trujillo, titulada: “Desarrollo de un sistema experto para ayudar en la prevención de infarto agudo de miocardio en personas de 30 a 75 años”. Objetivo: Desarrollar un sistema experto capaz de evaluar distintos casos mediante reglas y así determinar el óptimo para calcular el porcentaje probable del padecimiento de infarto en una persona. Material y Métodos: Investigación de tipo tecnológica aplicada y descriptiva, con diseño de pos-prueba con un solo grupo. Se tomo una muestra de 382 individuos y se recolectó la información mediante cuestionario. Resultados: El sistema experto creado evaluó los distintos casos de pacientes mediante programación en PROLOG de reglas estructuradas, obteniendo como resultado la probabilidad de que una persona pueda tener infarto. Se concluye que la técnica óptima es la de Framingham la cual contiene determinados factores de influencia sobre el riesgo de una enfermedad partiendo del cálculo de probabilidad.

Medrano (2020), en su investigación de tesis para obtener el título de ingeniero de sistemas y computación de la Universidad Daniel Alcides Carrión de la Ciudad de Cerro

de Pasco, titulada: “Diseño e implementación de un sistema experto para el diagnóstico de desnutrición en niños menores de 2 Años en el área de pediatría del centro de salud Tupac Amaru – Chaupimarca”. Objetivo: Diseñar e implementar un Sistema Experto que permita mejorar el diagnóstico de la desnutrición en niños menores de 2 años en el Área de Pediatría del Centro de Salud Tupac Amaru - Chaupimarca. Material y Métodos: El tipo de investigación es: Descriptiva – Correlacional – Aplicada, se empleó el método hipotético deductivo o de contrastación de hipótesis, de diseño no experimental de corte transversal, porque la investigación se realiza en un solo tiempo. La población lo conforma el personal de planta y usuarios del Centro de Salud Tupac Amaru – Chaupimarca, es de 95 personas de los cuales se colectó una muestra de 48 a quienes se les administró un cuestionario. Resultados: La arquitectura del sistema experto nos ofrece una mayor escalabilidad, ya que el experto en conocimiento puede seguir agregando más casos de síntomas identificados dentro de la base de conocimiento y así tener mas precisión en futuros diagnósticos. Se evidencia que la implementación del sistema experto mejora el diagnóstico de desnutrición en niños menores de 2 años en el Área de Pediatría del Centro de Salud Tupac Amaru-Chaupimarca.

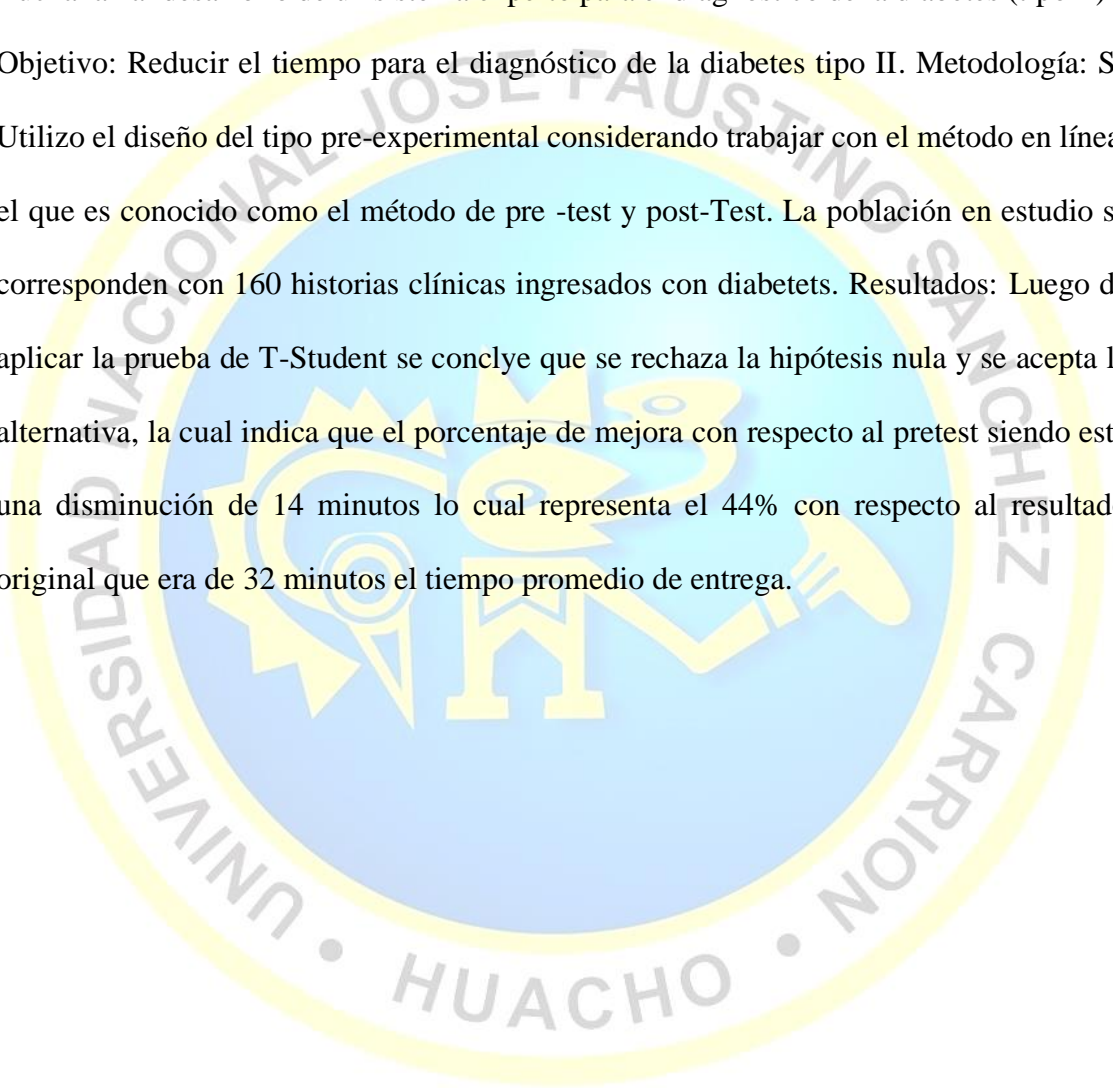
Chacaltana (2017), en su investigación de tesis para obtener el título de ingeniero de sistemas y cómputo de la Universidad Inca Garcilazo de la Vega, titulada: “Sistema Experto para el Diagnóstico de Enfermedades Respiratorias en el Hospital Central de la Policía Nacional del Perú Luis N. Sáenz”. Objetivo: Crear un sistema experto para el diagnóstico de enfermedades respiratorias, el cual será utilizado inicialmente en el Área de Neumología del Hospital Central PNP Luis N. Sáenz; permitiendo con ello atender de manera inmediata las inquietudes de los pacientes. Material y Métodos: Se creó un prototipo mediante programación orientada a objetos con desarrollo web y utilizando a

su vez la metodología CommonKADS. Resultados: Se logró determinar que el sistema experto tiene una gran influencia, puesto que hubo una mejora del proceso de atención de los pacientes, ya que, en general, hubo un incremento del nivel de satisfacción de los atendidos con respecto a éstas. La funcionalidad del sistema experto para el apoyo del diagnóstico de enfermedades respiratorias influye satisfactoriamente en los procesos de eficiencia y eficacia, ya que llega a cubrir y satisfacer los requerimientos funcionales de sus usuarios.

Aguilar (2017), en su investigación de tesis para obtener el grado de magister en computación e informática de la Universidad Nacional Mayor de San Marcos, titulada: “Sistema de razonamiento basado en casos, para la mejora de atención de salud en un centro rural”. Objetivo: Diseñar un modelo de sistema utilizando el razonamiento basado en casos (RBC) como apoyo al médico, para diagnosticar enfermedades más comunes en pobladores de un centro rural, con la finalidad de paliar en parte las necesidades básicas de salud en aquellos lugares. Material y Métodos: Investigación tecnológica, de corte transversal con enfoque cuantitativo y cualitativo, con aplicación a una muestra de 40 casos (pacientes) con enfermedades de paludismo y leishmaniasis. La metodología propuesta fue la del razonamiento basado en casos considerando cuatro etapas como son, etapa de recuperación, de adaptación, de revisión, de evaluación y finalmente de aprendizaje, estas dos últimas juntas, utilizándose los softwares JColibri donde se estableció el modelo y Protégé, donde se creó el nivel jerárquico y la ontología de dominio como son los síntomas y la interrelación entre ellos. Resultados: Resultados: Se construyó el sistema, observándose que los aciertos del modelo creado para este sistema es del 98% al compararse con el diagnóstico del médico evidenciado en la historia clínica, lo cual es bastante alto; para el diagnóstico de leishmaniasis el acierto es en un 100% (acierto total)

y para la malaria la concordancia con el diagnóstico médico fue de un 98% y para el diagnóstico de sano o sin estas enfermedades fue de un 80%.

Huarcaya (2021), en su investigación de tesis para obtener el título de ingeniero de sistemas de la Universidad César Vallejo, titulada: “Aplicación de la metodología Buchanan al desarrollo de un sistema experto para el diagnóstico de la diabetes (tipo II)”. Objetivo: Reducir el tiempo para el diagnóstico de la diabetes tipo II. Metodología: Se Utilizo el diseño del tipo pre-experimental considerando trabajar con el método en línea, el que es conocido como el método de pre -test y post-Test. La población en estudio se corresponden con 160 historias clínicas ingresados con diabetets. Resultados: Luego de aplicar la prueba de T-Student se concluye que se rechaza la hipótesis nula y se acepta la alternativa, la cual indica que el porcentaje de mejora con respecto al pretest siendo esta una disminución de 14 minutos lo cual representa el 44% con respecto al resultado original que era de 32 minutos el tiempo promedio de entrega.



2.2. Bases teóricas

La base teórica fundamental en el que se sustenta esta tesis, es el razonamiento basado en casos, la cual es una rama de los agentes inteligentes perteneciente al área de inteligencia artificial. Esta técnica permite a los usuarios finales realizar la toma de decisiones de una manera mas sencilla ante los diversos escenarios de cada una de las áreas del conocimiento, por tanto, es especialmente útil para los profesionales de cualquier especialidad que están a la busca de estas técnicas para desarrollar, aprender o tomar decisiones bien fundamentadas. Además, se tomará información sobre algunas otras bases teóricas que se utilizará para esta tesis y que son las siguientes:

- **Sistemas basados en conocimiento (SBC)**

Los SBC tienen su base, como su nombre lo indica, en el conocimiento de cualquier área de estudio o especialidad; el cual puede ser público o privado. El primero, incluye teorías, definiciones, hechos que han sido publicados; en cuanto al segundo se refiere a las competencias o habilidades de un experto y que es muy particular para u área específica. Las habilidades de un ser humano generalmente están relacionadas con el conocimiento privado. los SBC, precisamente representan “el conocimiento de los expertos en el dominio; especificar y reproducir tal conocimiento corresponde a la tarea principal en la construcción de sistemas basados en conocimiento”. (Jimenez & Gutiérrez , 2012)

Asimismo, el experto capta el conocimiento y lo almacena como caso en la base de casos existente dentro de un sistema, debido a eso es que existen dos modelos para los sistemas inteligentes:

a) Sistemas orientados al conocimiento con estructura cognoscitiva, cuyas características especiales son:

- Tienen la cualidad de solucionar determinados problemas, radica en un “saber”, o tener un conocimiento muy particular sobre un problema específico, más que en tener conocimientos generales.
- Permite la diferenciación del conocimiento interno de la máquina inteligente mediante su respectiva codificación (base de conocimientos) y los componentes o módulos de programas de deducción (motor de inferencias)

b) El sistema experto con orientación práctica, con características como:

- Trata de emular actividades de un experto o especialista al resolver problemas específicos dentro de un área muy particular.
- Tiene la capacidad para incrementar la experiencia.
- Tiene la capacidad para interactuar con los usuarios y explicar su razonamiento, para poder entender mejor la situación.

Considerando estos modelos o paradigmas, entonces que los SBC son sistemas que utilizan el conocimiento de un dominio en particular, con la finalidad de solucionar problemas del dominio específico. La solución entonces, de un problema mediante la máquina, sería la misma que la obtenida por una persona con experiencia cuando se enfrenta al mismo problema, siendo en este caso a persona un experto.

- **Tipos de sistemas basados en conocimiento**

La adquisición de conocimiento y su representación, según sea el tipo de conocimiento, se realiza por diferentes técnicas. También hay muchos modelos no excluyentes para la solución de problemas: por ejemplo, el modelo lógico, el modelo de búsqueda heurística, el basado en conocimiento, el conexionista, basado en experiencia, etc. Dentro del modelo basado en las experiencias, se encuentran el razonamiento basado en reglas (RBR), el razonamiento basado en casos (RBC) y las famosas ontologías, últimamente muy apreciadas y que inclusive hay softwares para ellas como el Protégé. Nosotros nos centraremos en este modelo RBC.

- **Base de conocimiento**

Es la que contiene las instrucciones de conocimientos para la realización de una tarea. Usa diagramas formales para la representación del conocimiento y los codifica dentro de un conjunto de normas. Esta etapa muestra la estructura del conocimiento o cognitiva y los procesos que un humano integra para dar una determinada solución, llamada memoria de largo plazo, donde se almacenan las experiencias (base de hechos) y el conocimiento (base de conocimientos) para el dominio en particular de esa experiencia.

- **Motor de inferencias**

El SBC puede controlar y aplicar los conocimientos al generalizar casos particulares, a esto se le llama inferencia, constituyendo el motor del sistema, mediante el cual se razona a partir del ingreso de los datos o conocimientos de entrada para luego dar un resultado o salida. Este motor gobierna el orden del

razonamiento, acepta ingresos y produce respuestas o soluciones. Esta fase se puede entender como un motor cuya función es la de razonar para resolver problemas.

- **Sistema experto**

El Doctor Edward E. Feigenbaum fue quien acuñó el término de sistema experto por primera vez, indicando que el poder de resolución de los problemas en la computadora u ordenador comienza en el conocimiento de un dominio específico, y que luego se traduce en código de programación que lo contiene (Feigenbaum, 1977).

En resumen, el sistema experto, es un sistema que emula el comportamiento de una persona experta en alguna temática (base de conocimiento); es decir, imita el accionar de un ser humano para lograr solucionar algún problema de un dominio específico.

- **Sistema experto basado en reglas**

Es un proceso que permitirá a partir de un conjunto de reglas iniciales desarrollar un proceso deductivo que concluirá al momento en que no quede ninguna otra regla por utilizar. Dichas reglas constituyen la base del conocimiento.

- **Sistema experto basado en redes bayesianas**

Este sistema basa su funcionamiento en las redes bayesianas. El motor de inferencia que utiliza para procesar las evidencias se basa en la teoría de probabilidades y más concretamente con el Teorema de Bayes.

- **Sistema experto basado en casos**

Este sistema basa su funcionamiento en experiencias anteriormente vividas, ya sea por el propio sistema o bien por la persona experta, y a partir de este conocimiento de vivencias realizar una asociación con estas experiencias para extraer una solución de esto. Como podemos ver de los tres métodos existentes, este es el que más se asemeja al modo de pensar que tenemos los seres humanos.

- **Razonamiento basado en casos (RBC)**

Es un sistema basado en conocimiento como parte de la inteligencia artificial para resolver problemas utilizando experiencias análogas para la solución respectiva, estas experiencias se almacenan como casos, haciendo que el sistema revise la el caso similar en la base de éstos, comparándolo con el caso específico, recuperando luego el que mejor se asemeja; al hacer la consulta, el caso es recuperado por comparación simple. El RBC funciona mediante un la especificación del dominio con un conocimiento explícito o muy particular, para luego escoger el mejor caso o casos de coincidencia y que serán nuevamente usados para formar parte de la solución del problema e incrementar la experiencia; la solución luego es revisada para ver si es adaptable al problema actual. El sistema retiene el nuevo caso como una solución para su uso posterior, pasando al almacén de la base de casos, creando un ciclo de razonamiento del conocimiento (Aamodt, 2004).

- **Ciclo del proceso del razonamiento basado en casos**

El RBC se divide en cuatro procesos, que son recordar o recuperar, reutilizar o reusar, revisar y retener. Esos procesos están claramente diferenciados.

- **Recordar** los casos iguales o similares al que se toma como caso base
- **Reutilizar** el conocimiento o experiencia que se tiene para resolver el problema caso base
- **Revisar** la solución propuesta utilizando los dos procesos anteriores
- **Retener** la nueva experiencia que puede ser útil para futuros casos que se proponen,

El ciclo RBC, se explica como un conocimiento contextualizado que representa la experiencia y que enseña una lección fundamental para lograr los objetivos del razonado. Los casos suelen presentarse en forma de problema discutido y solución disponible en detalle, la parte problemática del caso explica la situación que requiere ser resuelta, mientras que la segunda parte es la parte de solución que proporciona una respuesta adecuada a la situación actual. A continuación se describe con más detalle el ciclo.

- 1. Fase de recuperación del ciclo RBC:** La fase de recuperación de casos es el paso más importante en el proceso de RBC. Este paso se considera esencial porque el uso del sistema RBC de fase de recuperación calcula la similitud entre dos casos. La situación existente se concibe como un caso nuevo y se compara entre los casos almacenados sobre la base de la similitud.

El cálculo de similitud permite recuperar uno o más casos del repositorio de casos, donde los valores de similitud entre los casos van de 0 a 1, donde 0

refleja que no hay similitud en los casos almacenados y 1 muestra la coincidencia del 100% del nuevo caso con los casos almacenados. Se emplean múltiples algoritmos de recuperación en aplicaciones RBC. Las matrices de similitud son la base de la semejanza entre los casos que se almacenan en el repositorio de casos, siendo las técnicas de recuperación más utilizadas las que corresponden a los algoritmos del vecino más cercano, los árboles de decisión y sus derivados.

- 2. Fase de recuperación del ciclo RBC:** Esta es la fase de adaptación o reutilización, en ella se **reprocesa** el caso extraído de la base de conocimientos o del repositorio de casos y lo presenta como una solución recomendada para un caso de entrada actual. Los casos generalmente almacenados en el repositorio de casos deben modificarse para el nuevo caso. El procedimiento para ajustar la solución pasada para el caso de entrada actual se llama adaptación. Evalúa las variaciones generadas entre el caso de entrada y los casos almacenados que se recuperan del repositorio de casos y luego emplea diferentes técnicas para proponer la solución. Se necesitan expertos en el dominio para realizar la adaptación.
- 3. Fase de revisión del ciclo RBC:** Una vez que se recupera y adapta el caso, se procede a la etapa de revisión. En esta etapa la precisión y se ofrece como una solución validada para el nuevo caso problemático.
- 4. Fase de retención del ciclo RBC:** esta es la etapa final o última del ciclo CBR y es responsable de integrar nuevos casos en el repositorio de casos o en la base de conocimientos para su utilización futura.

En la figura 1 se observa el engarce del ciclo RBC

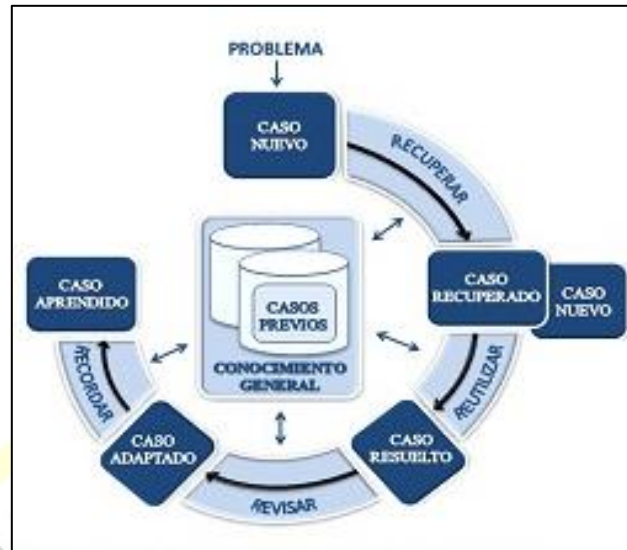


Figura 1 Razonamiento basado en casos y el ciclo de consulta para un nuevo caso

- **Contenedores del Conocimiento en RBC (Richter, 2005)**

Hay un modelo contenedor establecido y es conocido como el modelo de la estructura del sistema RBC y que precisamente tiene cuatro contenedores

- **El contenedor de la base de conocimiento** conteniendo los casos.
- **El contenedor de las medidas de similitud** la cual tiene los algoritmos de similitud.
- **La adaptación del conocimiento** mantiene la información para lograr que un caso antiguo se adapte a la base de casos, siempre comparándolo con el caso en cuestión.
- **El vocabulario** contenedor de los atributos de los casos

- **Validez de una prueba diagnóstica**

Mediante la clasificación de cada paciente como sano o enfermo, en función de que el resultado de la prueba sea positivo o negativo. De este modo, un resultado positivo se corresponde normalmente con la presencia de la

enfermedad en estudio y un resultado negativo con la ausencia de ésta. La validez se mide ya sea por los aciertos mediante el indicador estadístico Kappa o mediante los indicadores de sensibilidad y especificidad.

- **Características de una prueba diagnóstica**

Las condiciones que deben ser exigidas en un test diagnóstico son principalmente tres:

- **Validez**, es la frecuencia con la que los resultados obtenidos con este test pueden ser confirmados por otros más complejos y rigurosos.
- **Reproductividad**, Es la capacidad de un test de ofrecer los mismos resultados cuando se repite su aplicación en circunstancias similares. La reproductividad viene determinada por la variabilidad biológica del hecho observado, la introducida por el propio observador y la derivada del propio test.
- **Seguridad**, Es la certeza de que un test predecirá la presencia o ausencia de enfermedad en un paciente.

- **Indicador estadístico Kappa**

Proporciona una medida de grado acuerdo existente entre dos observadores, jueces o situaciones al evaluar una serie de sujetos u objetos. Para nuestro caso de estudio compararíamos el resultado del diagnóstico que dará el módulo del sistema experto y lo real brindado por la simulación o en todo caso los diagnósticos clínicos del médico. El valor de este indicador está entre 0 y 1 debiendo interpretarse a 0 como un acuerdo nulo (desacuerdo) y 1 un acuerdo máximo, entendiendo como acuerdo a concordar entre ambas situaciones.

2.3. Bases filosóficas

Desde la antigua Grecia, hace ya más de 2000 años se discutía sobre la ética y la moral; ahora, en nuestros tiempos, lejos de desvanecerse en el éter de la historia, se sigue con este debate y todo gracias a las nuevas tecnologías de información, una de ellas la inteligencia artificial (IA), de la cual dependemos cada vez más, pues está en todo y para todos alrededor nuestro, siendo dependientes a tal manera, que ya no importa para algunos inclusive “cultos”, el desaprovechar la intimidad familiar o de amistades; es decir, pareciera que la tecnología es buena pero los malos somos nosotros.

Y, por otro lado, el emular la IA con la inteligencia humana pareciera estar muy lejos todavía, porque no se comporta como el cerebro humano. Por ejemplo, hace unos pocos años, Google a través de su aplicación de “Photos”, cometió un error al estar mal direccionada, pues los usuarios se sorprendían al observar que dicha aplicación confundía los rostros con gorilas. El algoritmo para esa aplicación no tenía la capacidad para encontrar las diferencias y los técnicos de Google se pusieron a eliminar a los monos, gorilas y chimpancés, de las búsquedas para eliminar el error, pero ese no era el objetivo. Así, entonces todos los algoritmos basados en IA, que analizan diferentes actividades del ser humano, como control remoto del auto, automatización de artefactos, transacciones, transferencias, etc. pueden marcar una vida para siempre si se producen errores fatales.

Tal vez sea tiempo de revisar los principios que estableció Sócrates, que dice “el conocimiento es virtud; y el mayor de los males es la acción injusta; la razón es igual a la virtud”. Es evidente que la tecnología a través de computadoras, pueden realizar algunas actividades al igual que un ser humano, incluyendo acciones que se podría pensar, requieren intuición y discernimiento. Sin embargo, también es cierto, que existen algunas

actividades donde la inteligencia artificial no sobresale. Por otro lado algunos filósofos consideran que un robot con inteligencia artificial que haya pasado el Test de Turing no quiere decir que piense realmente como un humano, sino que es simplemente una analogía simulada a la acción de discernir humanamente (Russell & Norving, 2004).

Sin embargo, hay argumentos consistentes que se han postulado para el uso por ejemplo de las probabilidades dentro de la IA, como lo indican Cox (1946) y Carnap (1950). Imagínense ya desde hace más de 70 años el gran adelanto, ahora nos permite decir que las demostraciones prácticas muchas veces explican mejor las teorías. Por ejemplo, el éxito del razonamiento basado en la teoría de las probabilidades como base de los sistemas actuales demuestra que los incrédulos que decían nunca habrá IA están volviéndose crédulos.

En cuanto a la técnica de razonamiento basado en casos visto desde la perspectiva filosófica, es necesario comprender el por qué se utiliza, para lo cual tendremos que fijarnos el contexto en el que se va a desarrollar; así entonces, el RBC desde el punto de vista de la IA es un sistema que trata de emular a un ser humano en cuanto a una competencia y/o capacidad en particular, para solucionar un problema como lo haría un experto o reemplazar una función humana (Hammond, 1989)

Estos sistemas basan su funcionamiento en experiencias, ya sea del experto humano o por el propio módulo de sistema, y a partir de este reconocimiento de experiencias, realizar una asociación con las mismas para extraer una solución que resuelva el problema. Si hacemos una sinonimia con los niños, el razonamiento que utilizan está basado en sus experiencias con el entorno. Un ejemplo clásico relacionado con este tema, es cuando a un niño le sucede algo, por ejemplo, se quema por tocar un

recipiente caliente, entonces habrá adquirido una experiencia nueva. Si se le repite la misma situación, sabe que no debe tocar ese recipiente por la experiencia vivencial. Aun así puede ver a su madre coger ese recipiente con los paños de cocina, adquiriendo más experiencias o reconociendo patrones que debe de realizar para ese acto, reconociendo que si lo quiere repetir el acto, debe de primero usar los paños para no quemarse. Este es un ejemplo claro del funcionamiento del RBC. Se observa como el niño pequeño a través de anteriores vivencias, ya sea con errores o con aciertos, adquiere experiencias que utilizará cuando tenga que repetir la situación constituyéndose en su nueva base de razonamiento. Así pues funciona el RBC dentro de la IA teniendo a la vez adeptos como detractores y filosóficamente hablando diremos que la inteligencia artificial ya es aunque no lo deseen muchos, esencial en nuestra vida, aún algunos la utilicen para fines perversos.

2.4. Definición de términos básicos

- **Diagnóstico médico**

Es un juicio clínico sobre el estado psicofísico de un paciente, representa una manifestación en respuesta a una demanda para determinar tal estado, el único que puede indicar un proceso diagnóstico o manejo del paciente es el Profesional (Médico clínico que lleva a cabo la recepción del paciente y realiza tratamiento de emergencia según los casos clínicos del paciente). Para realizar el diagnóstico se debe tener en cuenta los dos aspectos de la lógica, es decir, el análisis y la síntesis, utilizando diversas herramientas como la anamnesis, la historia clínica, exploración física y exploraciones complementarias. (Altman, D., 1994)

- **Inteligencia artificial**

Se define como la reproducción mediante sistemas informáticos de los procesos mentales que en los seres vivos (y, en concreto, en el ser humano) dan lugar a las capacidades de aprendizaje, análisis y toma de decisiones racionales en base a los datos adquiridos y comprendidos (Aamodt, 2004).

- **Simulación**

Es el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias -dentro de los límites impuestos por un cierto criterio o un conjunto de ellos para el funcionamiento del sistema. La alternativa que nos presentan los entornos virtuales es la capacidad de recrear entornos físicos dentro de un ordenador o una simulación. A este tipo de mundos artificiales se les conoce actualmente como “Digital Twins”. Estos modelos, que no son más que réplicas virtuales exactas de entornos o sistemas físicos existentes en la vida real, han podido ser recreados mediante la utilización de motores físicos muy parecidos a los que se utilizan hoy en día en la industria de los videojuegos para la simulación de entornos virtuales (Napolitano, 2021)

- **Razonamiento basado en casos**

Es un sistema experto cuyo proceso es el de solucionar nuevos problemas basándose en las soluciones de problemas anteriores y que ayuda a tomar decisiones mientras se resuelven ciertos problemas concretos.

- **Caso**

Un caso es un evento, situación o fenómeno bajo un determinado contexto y que casi siempre hace referencia a una situación sui géneris o a una persona, diagnóstico, objeto, etc. Dentro de nuestro estudio, nos referimos a un paciente con sus determinadas dolencias para una enfermedad en particular. De allí que se debe de establecer el dominio de estudio de tal manera que, bajo una situación en particular, la solución proporcionada, por y para un caso sea mas precisa reflejando lo que realmente sucede en un contexto determinado (Juárez y Palma, 2015).

- **Índice de casos**

La indexación de los casos, garantizan su eficiente recuperación, considerando todos los aspectos relacionados al tiempo de recuperación y a la búsqueda misma, de tal manera que ningún caso relevante esté fuera del ciclo de búsqueda. Los índices se mantienen en memoria mediante una estructura de datos que puede ser evaluada y recorrida mediante un algoritmo de forma rápida. Normalmente, los índices están constituidos por característica o descriptores más importantes de los casos para evaluar sus diferencias entre unos casos de otros (Althoff 2011).

- **Especificidad de una prueba**

Es necesario indicar que tanto la especificidad como la sensibilidad son pruebas que permitirán observar que tanto es correcta la toma de una decisión de un buen diagnóstico. En cuanto a la especificidad, es la probabilidad de que un sujeto sano tenga un resultado negativo en la prueba. La especificidad es el

porcentaje de verdaderos negativos o la probabilidad de que la prueba sea negativa si la enfermedad no está presente. Los falsos positivos son sujetos sanos diagnosticados como enfermos. Con la especificidad lo que se detecta son los individuos sanos.

2.5. Hipótesis de investigación

2.5.1 Hipótesis general

Es factible crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad.

2.5.2 Hipótesis específicas

- Es factible crear un entorno de almacenamiento en un software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos.
- Existe concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico.

2.6. Operacionalización de las variables

Tabla 1. Operacionalización de las variables

Variable y su tipo	Definición conceptual	Definición operacional	Ingreso de datos para el RBC	Indicadores Carac. Norm	Instrumentos a utilizar
V.I. Sistema experto basado en casos para apoyo al médico	Programa que permite simular determinadas tareas como si fuera un experto (Diccionario RAE, 2010)	Conjunto de instrucciones que interactúan entre sí orientadas a realizar un diagnóstico médico.	Antecedentes familiares Embarazos Nivel de glucosa a las 2 horas de alimentos	Rangos de las características de estudio del paciente (Almacenados en la base de casos)	Softwares Protégé y Mýcbr Estadístico Kappa de Cohen
V.D.			Grosor de piel en triceps		

Diagnóstico
médico

Tipo de
variable:
Cualitativa

Índice de masa
corporal

Nivel de
insulina a las 2
horas de
alimentos

Presión
diastólica

Prueba
bilirrubina

Historial
ginecológico

Diagnóstico



CAPÍTULO III: METODOLOGÍA

3.1. Diseño metodológico

3.1.1. Tipo

Esta investigación de tesis está catalogado dentro del tipo de desarrollo tecnológico, orientado a obtener un producto final que será la obtención de un módulo de sistema experto partiendo de un modelo de razonamiento basado en caso, utilizando softwares libres; es decir, se espera finalmente contar con un módulo que diagnostique una enfermedad y que será de apoyo a los especialistas de salud.

3.1.2. Nivel y Enfoque

El nivel para esta investigación de tesis es el descriptivo analítico, no experimental, además de considerar el enfoque cualitativo y cuantitativo, pues se cuantificará mediante un indicador que tanto es el acierto del diagnóstico brindado por el módulo del sistema experto con un diagnóstico real.

3.2. Población y muestra

Por ser un estudio de tipo tecnológico no se consideró la toma de muestra, lo que si se pone a consideración es el modelo propuesto y que se verificó con bases de datos sobre diagnóstico médico.

3.3. Técnica de recolección de datos

Para la toma de la información se recurrió a la página www.datahub, la cual es una de las muchas que almacenan datos reales de calidad y de diversa índole, a la vez que comparten conocimiento. El link es el siguiente:

<https://datahub.io/machine-learning/diabetes>

<https://datahub.io/machine-learning/diabetes/r/1.html>

Estos datos también se puede encontrar en la base de datos de Pima Indians Diabetes

Database: <https://www.openml.org/d/37>

Como tecnología DataHub, además de ser un contenedor de bases de datos, es una herramienta para la gestión y automatización de los mismos, de esta página se escogerá los datos concernientes a prevalencia y diagnóstico de una enfermedad en particular; por ejemplo, la diabetes a través del archivo diabetes.csv. La base de casos contiene 769 casos coleccionados de pacientes con las características de síntomas y afecciones

En cuanto al traslado de los datos, éstos se almacenarán en una hoja de cálculo, donde cada registro será considerado para un paciente y cada columna con los síntomas y factores de riesgo, donde previamente se evaluarán los parámetros de evaluación en el diagnóstico de enfermedades de la siguiente manera:

Tabla 2. Evaluación de los síntomas y clasificación del riesgo

Parámetros de Evaluación	Niveles de la dolencia	Parámetros de Evaluación	Niveles de la dolencia
Síntoma 1	A B C	Riesgo 1	A B C
...
Síntoma n	A B C	Riesgo n	A B C

Fuente: Elaboración propia

3.4. Técnicas para el procesamiento de la información

Por ser una investigación tecnológica, los principales instrumentos a utilizar para la creación del módulo del sistema experto basado en casos, serán aquellos softwares libres que son especiales para este tipo de razonamiento, dentro de los cuales están:

- **MyCBR**

Es un aplicativo de software de código abierto o también denominado open source, fue construido por el Centro de Investigación Alemana DFKI para la Inteligencia Artificial (DFKI: Deutsches Forschungszentrum für künstliche intelligenz), se concatena muy bien con el software Protegé permitiendo la creación de prototipos de razonamiento basado en casos.

Este software myCBR tiene como orientación el apoyar en la construcción de aplicaciones RBC que requieren el uso de medidas de similitud entre lo existente y lo solicitado coordinando el conocimiento adquirido como experiencia, myCBR proporciona además, interfaces gráficas de usuario muy cómodas para modelar los tipos de atributos específicos, similitud y para evaluar la recuperación del caso resultante. Este software libre, aunque es antiguo se sigue utilizando por ser uno de los mejores incluyendo también herramientas para la generación automática y representativa de nuevos casos a partir de datos existentes sin procesar.

La idea subyacente del razonamiento basado en casos es la reutilización de casos anteriores para resolver problemas futuros (Aamodt y Plaza, 1994). Siguiendo este principio, también debemos capturar nuestra experiencia al crear sistemas RBC y proporcionarla para uso futuro. Esta es una de las razones para desarrollar la herramienta

de código abierto myCBR, nuestro objetivo es utilizar myCBR en proyectos y, por lo tanto, desarrollar funciones que se devuelven a la comunidad de investigadores.

myCBR como herramienta para la creación de sistemas RBC debe ser capaz de cubrir una gran variedad de tareas tales como apoyo a la toma de decisiones, diagnóstico, planificación, etc. Especialmente cuando se construyen aplicaciones basadas en el conocimiento, se deben considerar dos aspectos: durante la fase de definición, las bases de conocimiento tienen que ser creados y discutidos con expertos. Por lo tanto, una interfaz de usuario es crucial para hacer que este conocimiento sea transparente. Por otro lado, para ejecutar el sistema RBC, necesitamos un motor que haga uso de los conocimientos creados previamente. myCBR ofrece ambos: el primero se conoce como unidad de interface gráfica (myCBR GUI) y el último como myCBR SDK. La GUI hace uso del SDK y lo envuelve en una interfaz de interacción de usuario.

- **Protégé**

Es una herramienta open source, que cuenta con una sólida comunidad académica y de usuarios corporativos, sirve para la construcción de soluciones con base en el conocimiento de áreas tan diversas como la biología, medicina, negocio electrónico y modelado organizacional; fue desarrollado por el Centro de Investigación para la Informática Biomédica de la Universidad de Stanford, conjuntamente con la Universidad de Manchester. Protégé proporciona una interface bastante amigable al usuario para definir ontologías, incluyendo clasificadores que sirven para validar los modelos a crear y generalizar nueva información basada en ontologías. Este software se desarrolló ampliamente en Java. Actualmente existen más de 240 000 personas inscritas en su página

electrónica, muchos de los cuales son también desarrolladores del mismo software, por ser libre.

○ **Confiabilidad del modelo RBC**

Para dar una confiabilidad más científica al módulo de sistema experto, a parte del mostrado líneas arriba, se utilizará el estadístico Kappa de Cohen (Cohen, 1960), el cual mide la concordancia de dos o más situaciones sobre el mismo fenómeno en estudio, el cual queda expresado dicotómicamente, por ejemplo, al comparar la presencia o ausencia de una característica, que puede ser una enfermedad, un atributo, etc., la concordancia se expresa mediante del índice de Kappa (k), el cual se expresa por la siguiente fórmula

$$k = \frac{P_0 - P_e}{1 - P_e}$$

Las proporciones P_0 y P_e se hallan luego de construir una tabla de contingencia de 2x2 donde se encuentran los valores observados y los esperados por el azar y que se muestra en la tabla 3.

Tabla 3. Comparación entre el experto médico y el sistema RBC

EXPERTO	MODELO PROPUESTO (RBC)				Total Frec.
	Clasif. 1		Clasif. 2		
	Valores observados	Valores esperados	Valores observados	Valores esperados	
Clasif. 1	a	A	b	B	a + b
Clasif. 2	c	C	d	D	c + d
TOTAL	a + c	A + B	b + d	B + D	n

Proporción de observada global (P_0)

$$P_0 = \frac{(a + d)}{n}$$

Proporción esperada global (P_0)

$$P_e = \frac{(A + D)}{n}$$

Donde P_0 es la proporción observada de concordancia, P_e es la proporción esperada de concordancia por azar y $1 - P_e$, representa la máxima concordancia posible que no es debida al azar. El numerador del coeficiente Kappa es la diferencia de la proporción del acuerdo observado con el acuerdo esperado, en tanto que el denominador es la diferencia entre un acuerdo total y la proporción brindada por el azar. Así, el índice k corrige el acuerdo que solo nos daría el azar, en tanto es la proporción del acuerdo observado que excede la proporción por azar. Si el valor de k es igual a 1, se dice que hay una concordancia perfecta y por lo tanto, la proporción por el azar no existe, es decir es cero. Cuando el valor es 0, el desacuerdo es total, siendo que la proporción esperada por el azar es igual a la proporción observada.

El valor del estadístico k representa la fuerza de concordancia y está clasificado según los intervalos mostrados en la tabla 4.

Tabla 4. Valores del índice Kappa y niveles de concordancia.

Valor de k	Fuerza de concordancia
< 0,20	Pobre
0,21 - 0,40	Débil
0,41 - 0,60	Moderada
0,61 - 0,80	Buena
0,81 - 1,00	Muy buena

Fuente: Adaptado de Landis JR, Koch GG. Biometrics 1977; 33: 159-174

CAPÍTULO IV: RESULTADOS

4.1. Análisis de resultados

4.1.1 Descripción del modelo propuesto

El modelo propuesto orientado al diagnóstico médico, establece dos capas principales, la primera, concerniente a los componentes centrales que forman el conocimiento artificial y que está compuesto en gran parte por la base de casos y sus interrelaciones y la segunda referente al razonamiento para tomar una decisión a la que se denomina máquina de inferencia, la cual accederá a la base de conocimientos evaluando todos los casos de los pacientes de acuerdo a sus síntomas que previamente han sido ingresados por el que maneja el modelo mediante el aplicativo. En la figura 2 se muestra gráficamente el modelo, partiendo por la consulta del usuario a través de una interface amigable por computador. La consulta llega inmediatamente a una capa de base de casos que ha sido implementada por el experto, en este caso se considera las características del paciente y que casi siempre se evalúa partiendo de una historia clínica. Esta consulta pasa ahora a comparar el caso nuevo con el módulo de razonamiento donde se encuentra el conocimiento general y que corresponde a la capa de aplicación construida mediante el software myCBR, este módulo, es aquí donde se realiza todo el ciclo del razonamiento, es decir, se ingresa el caso nuevo, se recupera un caso o casos los mas semejantes posibles, luego se resuelve el caso, se lo adapta a la base de casos y el módulo se fortalece aprendiendo con el nuevo caso.

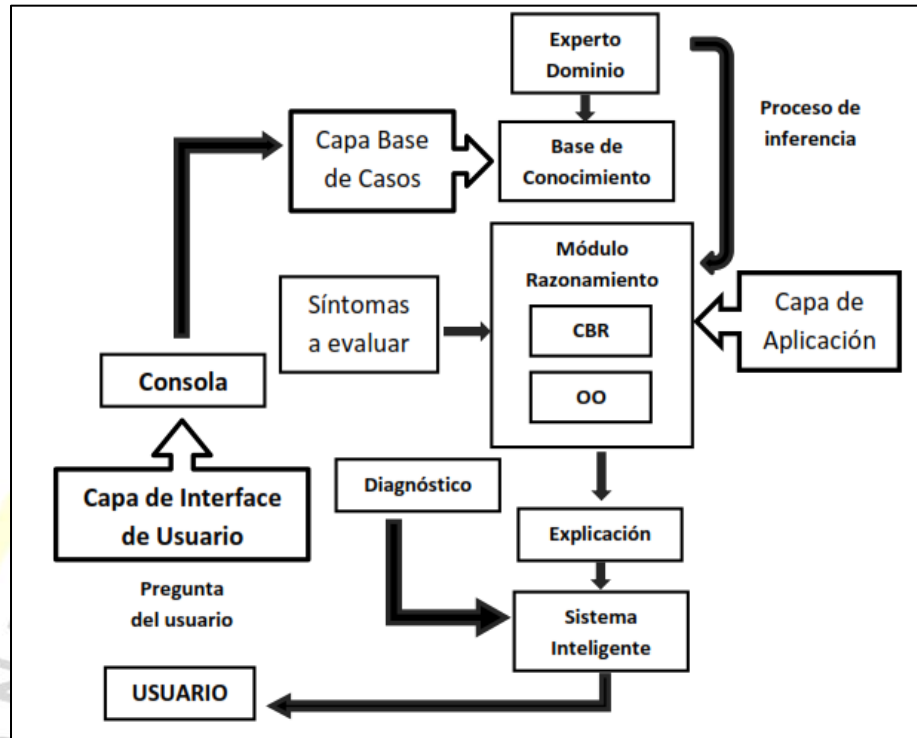


Figura 2. Modelo para el sistema de diagnóstico y sus relaciones

Lo anteriormente expuesto es parte del modelo y se detalla a continuación siguiendo la metodología del razonamiento basado en casos, y que permite realizar las siguientes tareas:

- 1) Se ingresa los síntomas a diagnosticar del paciente y que se considera como un caso.
- 2) El sistema mediante un algoritmo busca dentro del registro de memoria un caso con los síntomas iguales o muy parecidos a los del caso ingresado.
- 3) Luego de ubicar un caso, se orienta a la solución, que para nuestro caso es el diagnóstico. Inmediatamente el caso ingresado se anexa como un caso mas, enriqueciendo la base de casos.
- 4) Si no ubica un caso exactamente igual, se recupera un caso con bastante similitud al de entrada, para lo cual previamente se ha establecido una función de similitud en el modelo.

- 5) De acuerdo al paso anterior, el sistema puede modificar internamente algunas características al no cumplir con el caso de entrada, lo cual hace que el sistema se auto modifique, llegando a un caso de adaptación, pero indicando un porcentaje de similitud.
- 6) El sistema admite mediante el proceso de adaptación, generar una nueva solución enriquecida generando un caso nuevo, el cual es agregado en la memoria para nuevamente ser usado en un momento posterior.

El dominio para este modelo fue implementado previamente en el software Protégé, el cual da un conocimiento general, incluyendo las clases, conceptos, el vocabulario a utilizar y sobre todo las relaciones para representar dicho conocimiento orientado al diagnóstico médico. En cuanto al manejo de las similitudes, éstas serán implementadas en el software myCBR, donde se encuentra la estructura de cada uno de los casos y la implementación del esquema del RBC, es decir, la representación del caso, localización del mismo, recuperación, revisión, adaptación y posterior reutilización y que se detalló ampliamente en el acápite 3.4.

4.1.2 Descripción de los datos

Es necesario recalcar que los datos fueron colectados y recuperados del registro de salud electrónico (EHR), éstos también se encuentran en la base de datos de Pima Indians Diabetes Database (<https://www.openml.org/d/37>) que corresponden a 768 registros de pacientes del sexo femenino y se encuentran en idioma inglés, posee las características del paciente que a su vez incluyen varias dimensiones como son características demográficas, pruebas de laboratorio, perfil hematológico, síntomas, pruebas de laboratorio para la función Kidney, perfil lipídico, análisis de orina, prueba de bilirrubina, si es mujer, historial femenino y finalmente el diagnóstico de la diabetes (Diabetes Diagnosis), así como diagnóstico del hígado (Liver Diagnosis).

La representación de los casos, se hicieron mediante atributos, dando valores numéricos, ordinales, nominales o categóricos a las características del paciente en estudio. Aquellos registros que han tenido mas del 50% de los datos perdidos han sido removidos.

De todas estas variables se han tomado las mas preponderantes como son: antecedentes (anteced), edad, número de embarazos (embrz), nivel de glucosa (gluc), grosor de los pliegues de la piel del triceps (grosor_tric) , índice de masa corporal (imc), nivel de insulina (insu) y la presión diastólica (pres). En la tabla 8, se presentan las principales características de los pacientes

Tabla 5. Características principales y diagnóstico de los pacientes en estudio

Nº	Caract. categóricas	Valores Codif.	Valores Originales
1	Antecedentes	0	Sin antecedentes
		1	Con antecedentes
0	Embarazos	0	Normal
		1 a 0	+
		3 a 5	++
		> 5	++
3	Glucosa	0	Normal
		1	+
		0	++
4	Grosor triceps	0	Normal
		1	+
		0	++
		3	+++
5	IMC	0	Normal
		1	+
		0	++
		3	+++
6	Insulina	0	Normal
		1	+
		0	++
		3	+++
8	Presión diastólica	0	Normal
		1	+
		0	++
		3	+++

		0	Normal
		1	Prediabetic
10	Diabetes Diagnosis	0	Prediabetic/Gestational
		3	Diabetic
		4	Diabetic/Gestational

Fuente: Pima Indians Diabetes Database.
Recuperado de <https://www.openml.org/d/37>

Asimismo, las características para una muestra de nueve pacientes (casos), extraídos directamente de la base de casos se muestran en la siguiente tabla 8.

Tabla 6. Segmento de 9 casos con sus respectivas características

Gluc	Pres	Grosor _tric	Insu	Imc	Anteced	Edad	Diag
148	70	35	0	33.6	0.607	50	testeadado_positivo
85	66	09	0	06.6	0.351	31	testeadado_negativo
183	64	0	0	03.3	0.670	30	testeadado_positivo
89	66	03	94	08.1	0.167	01	testeadado_negativo
137	40	35	168	43.1	0.088	33	testeadado_positivo
116	74	0	0	05.6	0.001	30	testeadado_negativo
78	50	30	88	31.0	0.048	06	testeadado_positivo
115	0	0	0	35.3	0.134	09	testeadado_negativo
197	70	45	543	30.5	0.158	53	testeadado_positivo

Fuente: Construcción propia en base a los datos de Pima Indians Diabetes Database.

A continuación, se presentan algunas características de los datos en estudio.

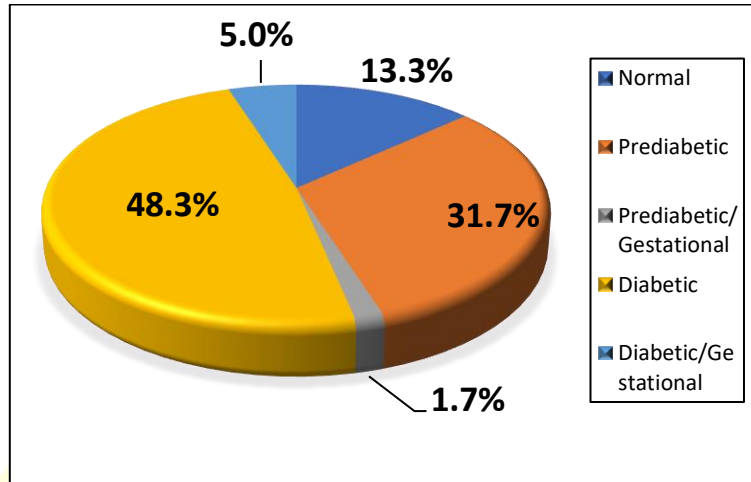
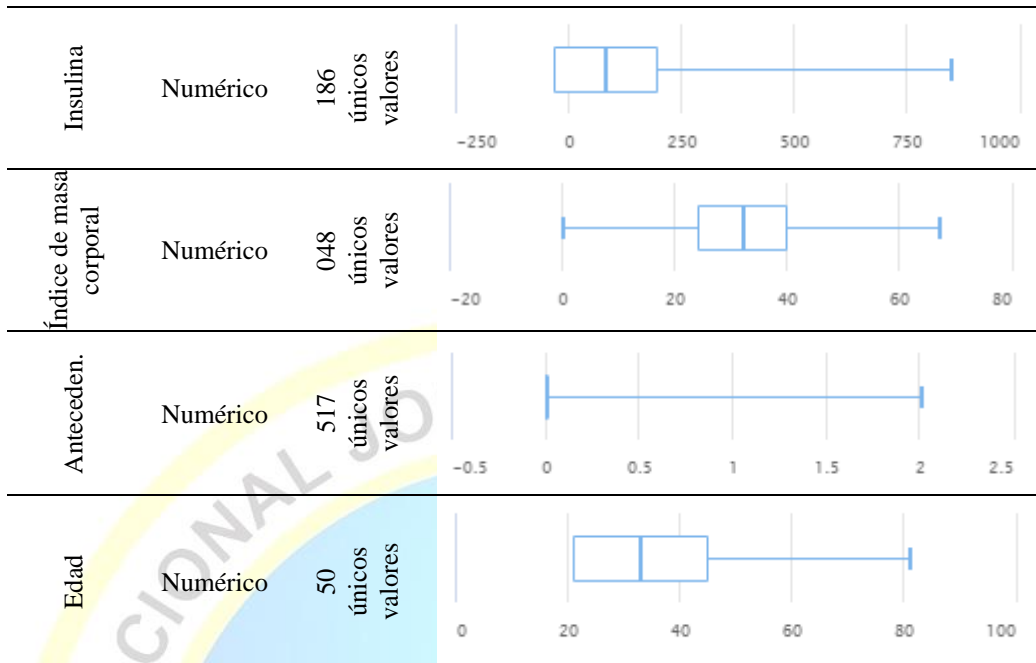


Figura 3. Clasificación de los pacientes por tipo de diabetes

Fuente: Datos colectados del registro de salud electrónico Pima Indians Diabetes Database.

Tabla 7. Características generales de los datos en estudio

Caract.	Tipo de valor	Valores	Medidas en representación gráfica
Diagnóstico	Nominal	2 únicos valores	
Número de embarazos	Numérico	17 únicos valores	
Glucosa	Numérico	136 únicos valores	
Presión diastólica	Numérico	47 únicos valores	
Grosor de la piel en el tríceps	Numérico	51 únicos valores	



Fuente: Pima Indians Diabetes Database. Recuperado de <https://www.openml.org/d/37>

4.1.3 Importación de los datos y creación de la clase paciente

Mediante el software Protégé se importó los datos desde la página mencionada anteriormente en formato .CSV cuyo separador fue la coma arábiga y a la vez se creó la clase “Paciente” con sus respectivas características. Cada uno de los registros importados el software Protégé los considera instancias, como se muestra en las siguientes dos figuras

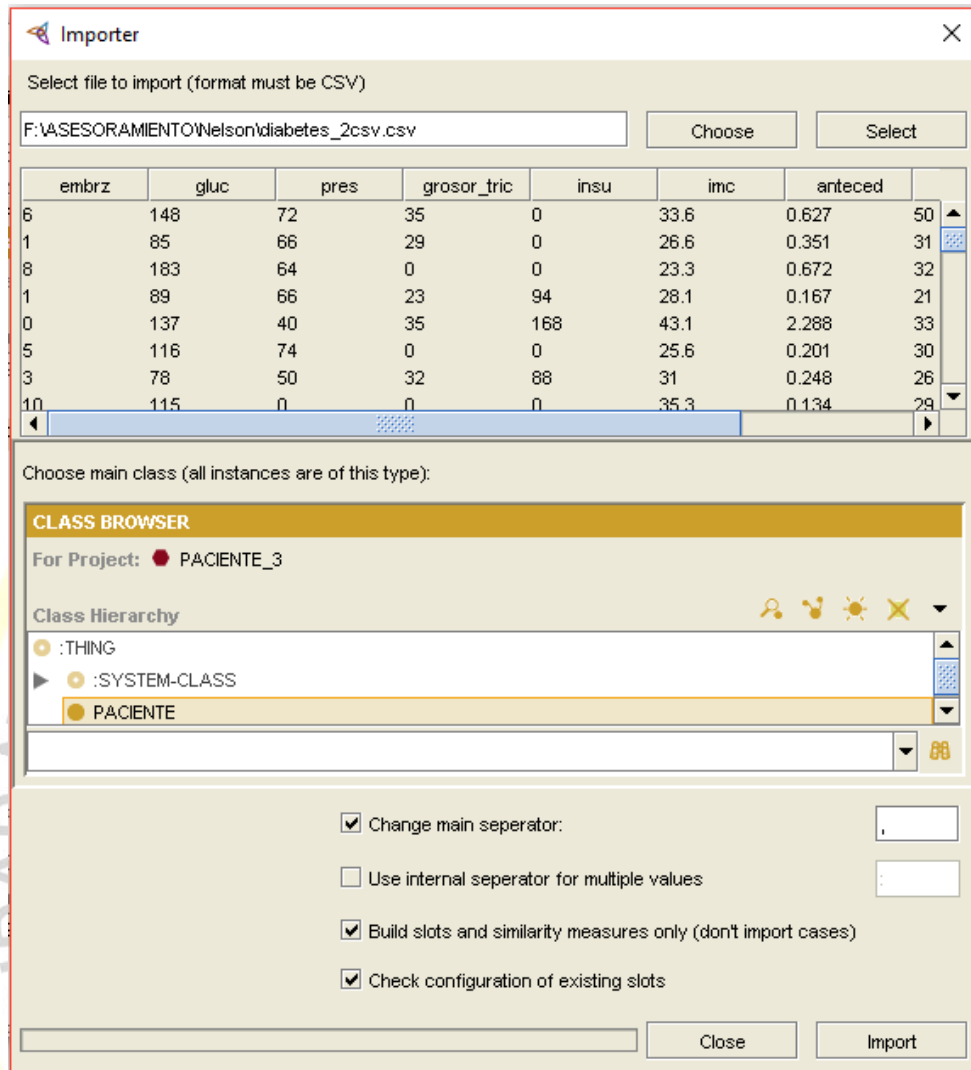


Figura 4. Importación de la base de datos desde el software Protégé
Fuente: Elaboración propia con Protégé

CLASS EDITOR
For Class: ● PACIENTE (instance of :STANDARD-CLASS)

Name: PACIENTE

Documentation: TESIS: NELSON FALCON ZUÑIGA (Derechos Reservados)
Clase Paciente
9 Características
Max y Min hallados en los datos

Constraints

Role: Concreto ●

Template Slots

Name	Cardinality	Type	Other Facets
anteced	single	Float	minimum=0.078, maximum=2.42
diagnos	single	Symbol	allowed-values={tested_positive, tested_negative}
edad	single	Integer	minimum=21, maximum=81
embrz	single	Integer	minimum=0, maximum=17
gluc	single	Integer	minimum=0, maximum=199
grosor_tric	single	Integer	minimum=0, maximum=99
imc	single	Float	minimum=0.0, maximum=67.1
insu	single	Integer	minimum=0, maximum=646
pres	single	Integer	minimum=0, maximum=122

Figura 5. Clase paciente y sus respectivos slots (características)
Fuente: Elaboración propia con Protégé

A su vez se ha tenido especial cuidado en construir los valores permitidos para cada característica o slot, por ejemplo, para la característica diagnóstico, el tipo valor es Symbol, con dos valores permitidos `tested_positivo` y `tested_negativo`, el dominio al que pertenece este slot es precisamente la clase Paciente. De igual manera se ha configurado para las demás características

SLOT EDITOR
For Slot: ■ diagnos (instance of :STANDARD-SLOT)

Name: diagnos

Documentation

Value Type: Symbol

Allowed Values: tested_positive, tested_negative

Cardinality: required at least
 multiple at most

Minimum: Maximum: Inverse Slot

Template Value

Default Values

Domain: ● PACIENTE

Figura 6. Ejemplo de una característica del paciente
Fuente: Elaboración propia con Protégé

Luego, para cada uno de los slots se ha configurado las medidas de similitud en el entorno del editor del software Protégé ya integrado con el plugin del software MyCBR, de la siguiente manera:

- Se ha seleccionado el comportamiento de la función el cual puede ser simétrico o asimétrico.
- Se ha especificado la función distancia, pudiendo ser distancia simple o cociente.
- Se ha configurado la simetría o asimetría dándole valores al polinomial. Aquellos slots en los que se va a configurar la asimetría, el panel de la izquierda calcula la similitud para valores de consulta mayores que los valores en un caso dado y el panel de la derecha para valores de consulta menores a un caso dado.
- Para nuestro caso de estudio, todas las medidas de similitud han sido simétricas y eso se observa en el gráfico inferior del editor de medidas de similitud. Por Ejemplo, en el caso de la glucosa, se observa el gráfico simétrico, donde el valor mayor lo tiene un paciente con 199

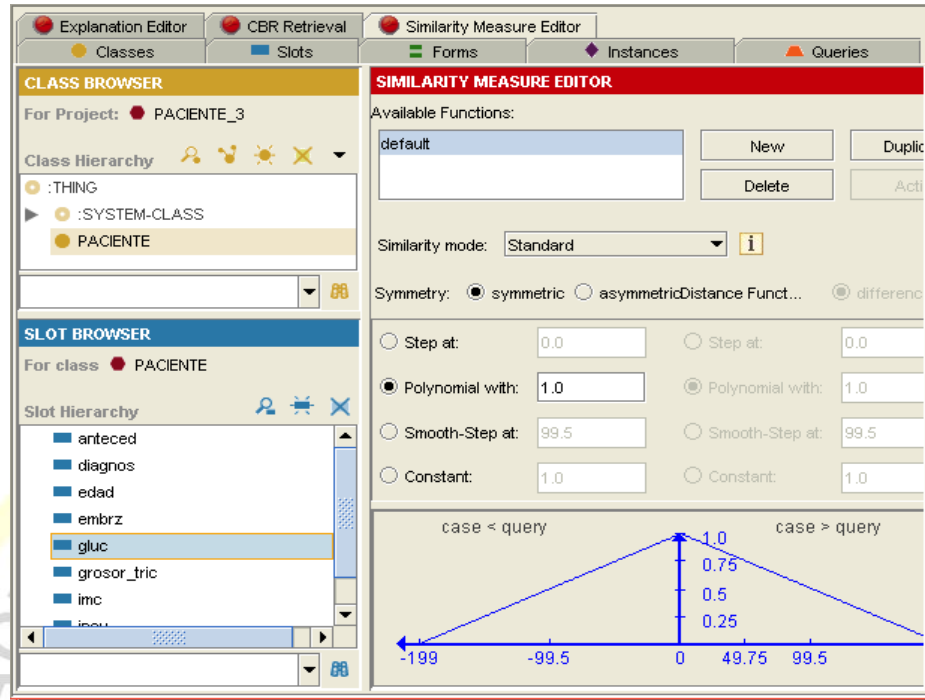


Figura 7. Similitud para la característica glucosa
Fuente: Elaboración propia con Protégé y myCBR

4.1.4 Explicaciones que brinda el modelo dentro del software myCBR

En el software myCBR, el modelo establecido para diagnóstico, ofrece tres diferentes explicaciones con los datos: las explicaciones conceptuales que son las que describen elementos del contenedor de conocimientos de vocabulario al usuario final mediante descripciones textuales o mediante enlaces a información adicional, que para nuestro caso son básicamente los síntomas de los pacientes. Luego muestra las explicaciones sobre el resultado de la recuperación en relación con una consulta particular al caso que ha ingresado el usuario y finalmente, las explicaciones anticipadas que ayudan en el modelado y el mantenimiento de la base de casos.

4.1.5 Consulta de los casos

Para probar este modelo se ha ingresado un caso nuevo, cuya descripción se ingresa al sistema como consulta y la salida nos presenta a aquellos pacientes ordenados por probabilidad de similitud al caso.

El caso en consulta tiene los siguientes parámetros, la mayoría de ellos normales:

- Antecedentes : 1
- Diagnóstico : indefinido
- Edad : 40 años
- Nro. de embarazos : 6
- Glucosa a 0 horas : 140
- Grosor triceps : 04
- IMC : 38
- Insulina a 0 horas : 188
- Presión diastólica : 68

Estos valores se ingresan en la columna consulta (Query), obviamente no ingresamos el diagnóstico (_undefined_). Luego de hacer una recuperación (Retrieve), aparecen los casos mas similares al consultado, que para nuestro ejemplo son el 671, el 500 y el 96, todos ellos testeados negativos con probabilidades 0,94, 0,94 y 0,93. Esto se muestra en la figura 8.

Query	KB_862466_Class671	KB_862466_Class500	KB_862466_Class96
	1	2	3
	0.94	0.94	0.93
anteced	1.0	0.631	0.839
diagnos	undefined_	tested_negative	tested_negative
edad	44	49	39
embrz	6	6	6
gluc	140	165	154
grosor_tric	24	26	32
imc	38.0	33.6	29.3
insu	188	168	193
pres	68	68	74

Figura 8. Recuperación de casos similares a la consulta
Fuente: Elaboración propia con MyCBR

4.1.6 Concordancia entre el modelo expuesto y el diagnóstico médico

Según los datos se tiene 500 pacientes testeados negativos y 069 testeados positivos. Después de haber ejecutado el simulador de sistema experto en el software myCBR, se muestra a continuación el porcentaje de aciertos entre el simulador y el diagnóstico médico encontrado en la base de datos.

Tabla 8. Porcentaje de acierto entre el sistema experto y el médico

Diagnóstico del simulador	Aciertos	% de aciertos
Tested_positive	(068 de 068)	100,0
Tested_negative	(495 de 500)	99,0
Total	(763 de 768)	99,3

Fuente: Elaboración propia

Se observa que hay bastante acercamiento entre el diagnóstico del simulador y el del médico (99,3%). Para darle una mayor validez a los resultados, se trabajó con el Índice Kappa de Cohen, el cual sirve para comprobar si la concordancia existente no solo se debe al azar.

4.2. Contrastación de las hipótesis

4.2.1. Análisis para la contrastación de la Hipótesis General

“Es factible crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad”.

Respecto a este planteamiento, se puede afirmar que es totalmente factible crear dicho modelo, al haber sido implementado en los softwares Protégé y myCBR, quedando establecido al aceptar las dos hipótesis específicas.

4.2.2. Análisis para la contrastación de las Hipótesis Específicas

H.1: “Es factible crear un entorno de almacenamiento en un software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos”.

Según 4.1.3. se importó los datos y se creó la clase paciente, con sus respectivos registros que el software Protégé los denomina instancias, considerando como se vio anteriormente algunas medidas con valores cualitativos o cuantitativos como son embarazo, glucosa, presión, etc. Luego de eso cada registro es considerado como un paciente y que fue importado de la data Pima Indians, formando la base de casos de pacientes con sus respectivas características. La consulta de un caso en 4.1.5, brinda las salidas respectivas para conocer con que caso o casos tiene mayor similitud el caso ingresado como consulta.

Con esto queda probada la hipótesis específica 1, permitiéndonos decir que es factible crear un entorno de almacenamiento en un software para los síntomas de una enfermedad que para nuestro caso de estudio los entornos han sido los softwares integrados Protégé y myCBR; en cuanto los síntomas de la enfermedad han sido los de la diabetes dentro del modelo del sistema basado en casos.

H.2: “Existe concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico”

Para probar esta hipótesis nos remitimos a 4.1.6, donde se observa bastante similitud entre los resultados que arroja el modelo y el diagnóstico médico dado en la base de datos; sin embargo, para probar estadísticamente la concordancia entre ambos resultados se utilizó la prueba Kappa de Cohen y que la desarrollamos mediante la siguiente prueba de hipótesis:

H_0 : No hay concordancia entre los resultados del modelo y el diagnóstico médico

H_1 : Si hay concordancia entre los resultados del módulo y el diagnóstico médico

Para probar esto disponemos en una tabla de doble entrada los resultados del modelo y del médico encontrando los valores observados y esperados para cada conjunto de diagnóstico

Tabla 9. Concordancia entre el simulador y el diagnóstico médico

MODELO	MÉDICO (sacado de los datos)				Total	
	tested_positive		tested_negative		Recuento	Recuento esperado
	Recuento	Recuento esperado	Recuento	Recuento esperado		
tested_positive	a: 068	A: 95,3	b: 5	B: 177,7	a + b 073	073,0
tested_negative	c: 0	C:170,7	d: 495	D: 300,3	c + d 495	495,0
Total	068	068,0	500	500,0	768	768,0

Fuente: Elaboración propia

Proporción de concordancia observada global (P_0)

$$P_0 = \frac{(a + d)}{n} = \frac{(068 + 495)}{768} = 0.993$$

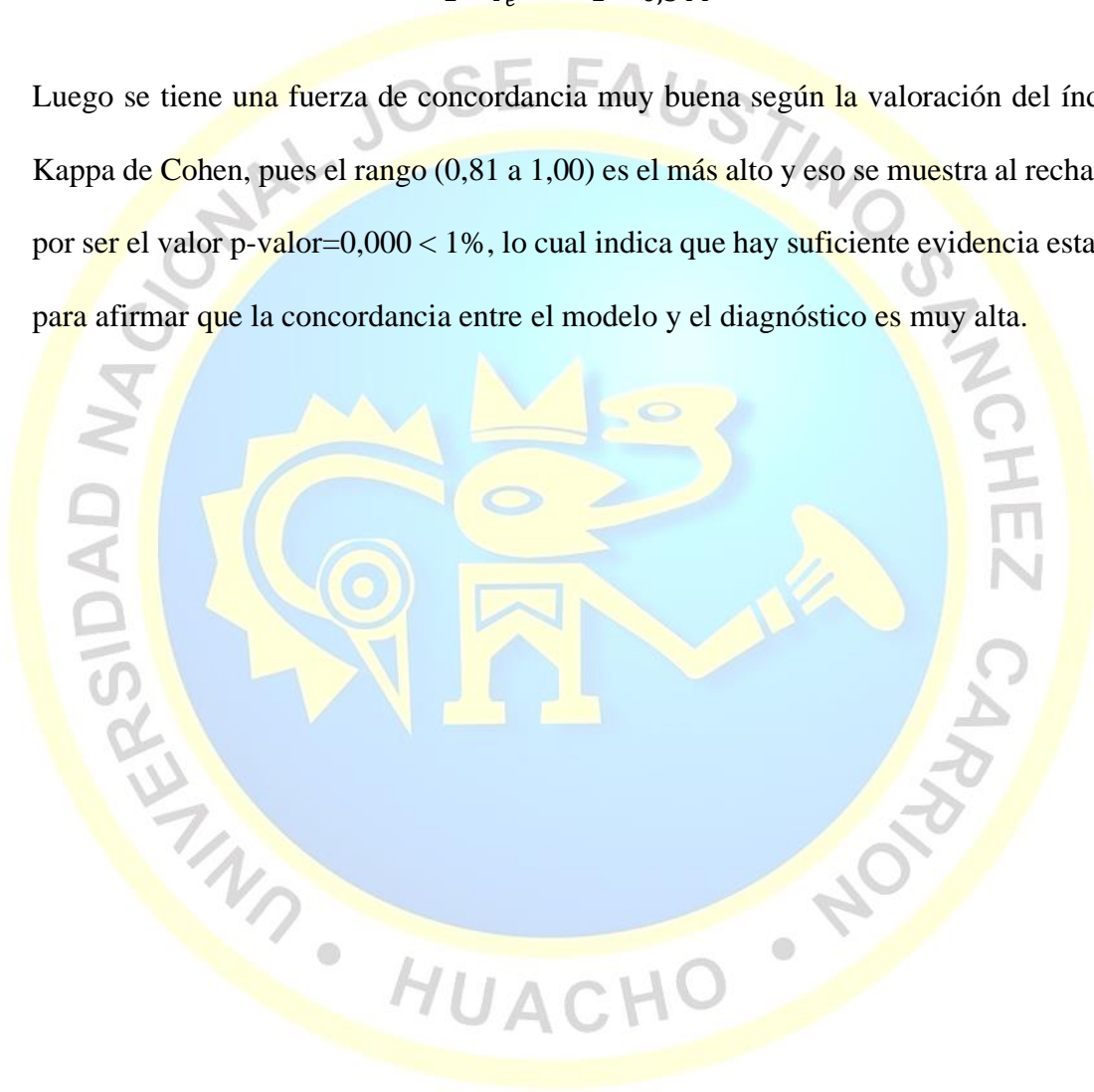
Proporción de concordancia esperada global (P_e)

$$P_e = \frac{(A + D)}{n} = \frac{(95,3 + 300,3)}{768} = 0,54$$

Según los cálculos para obtener el índice k se tiene:

$$k = \frac{P_0 - P_e}{1 - P_e} = \frac{0,993 - 0,544}{1 - 0,544} = 0,98$$

Luego se tiene una fuerza de concordancia muy buena según la valoración del índice de Kappa de Cohen, pues el rango (0,81 a 1,00) es el más alto y eso se muestra al rechazar H_0 por ser el valor p-valor=0,000 < 1%, lo cual indica que hay suficiente evidencia estadística para afirmar que la concordancia entre el modelo y el diagnóstico es muy alta.



CAPÍTULO V: DISCUSIÓN

5.1. Discusión de resultados

El objetivo de la tesis fue crear un modelo de un sistema basado en casos, orientado al diagnóstico médico de una enfermedad y que luego fue implementado en los softwares Protégé y myCBR con la finalidad de comprobar si efectivamente este tipo de razonamiento basado en casos, el cual es un área de la inteligencia artificial, puede ser de aplicación y ayuda en el campo del diagnóstico médico. La base de datos utilizada fue la de Pima Indians Diabetes Database para dicha enfermedad y que constó con 769 casos de pacientes con sus respectivos síntomas y que fue importada mediante archivo CSV al software Protégé (Ver Figuras 4 y 5).

Al respecto, se puede afirmar que el modelo creado cumple con las expectativas para el que fue creado, encontrando similitud con el diagnóstico médico, como se puede apreciar con índice estadístico Kappa de Cohen al encontrarse una concordancia de 0,986, la cual es considerada como alta, además de obtener un p-valor de 0,000, lo cual indica que existe suficiente evidencia estadística para afirmar que esa concordancia es altamente significativa (Ver 4.2.2) y que es bastante alta al comparar con Rumui y otros (2018), cuyos valores mas altos de similaridad fueron de 0,75 y 0,8.

Es necesario, aclarar que el modelo presentado en esta tesis está orientado al diagnóstico, previo almacenamiento de caos de una enfermedad en particular, pudiendo extenderse mediante un módulo mas, a que luego de diagnosticada la enfermedad, se pase al tratamiento, para escoger luego el mejor de ellos, dentro de también una base de casos,

como lo indica Paruchuri y Granville (2020); y, así se puede ir incrementando y enriqueciendo mediante módulos el modelo transformándose en un real sistema experto de inteligencia artificial con orientación medica.

Nuestro modelo implementado mediante estos dos softwares , puede servir entonces, tanto para la capacitación de personal de salud, incluyendo aquellos que no son médicos como enfermeras u otros, como también para los serumistas que van a lugares alejados de las capitales de provincias, como por ejemplo, en a sierra o selva, donde los servicios de salud, casi siempre son escasos y no hay especialistas en tal o cual enfermedad, coincidiendo en este caso con Shahina (2009) quien indica en su tesis que su sistema de diagnóstico de sensibilidad al estrés en psicofisiología puede utilizarse como experto para un médico con menos experiencia o como una segunda opción para un médico con experiencia en su proceso de toma de decisiones en el diagnóstico de estrés. O como lo indica Chacaltana (2017) cuyo sistema experto creado en su tesis y que luego de su aplicación, tiene una gran influencia, puesto que hubo una mejora del proceso de atención de los pacientes, debido a su funcionalidad para el apoyo del diagnóstico de enfermedades respiratorias, ya que llega a cubrir y satisfacer los requerimientos funcionales de sus usuarios

Como se dijo anteriormente, el modelo se implementó integrándose en dos softwares que son Protégé y myCBR, el primero para la construcción de las características propias del paciente, es decir su diagnóstico y el segundo donde se implementó el razonamiento basado en casos, coincidiendo en parte con Aguilar (2017) quien utilizó el software Protégé pero en vez de myCBR utilizó el software JColibrí, cuya diferencia está básicamente en el establecimiento de las interrelaciones, siendo el software myCBR mas amigable al usuario.

CAPÍTULO VI: CONCLUSIONES Y RECOMENDACIONES

6.1. Conclusiones

Luego de elaborado el modelo, de haberse implementado en los softwares mencionados y probado con la base de datos de Pima Indians Diabetes Database (Sigillito, 2014) de pacientes testeados como diabéticos y no diabéticos, se establece que el modelo cumple con las expectativas para las cuales fue creado y llegando a las siguientes conclusiones:

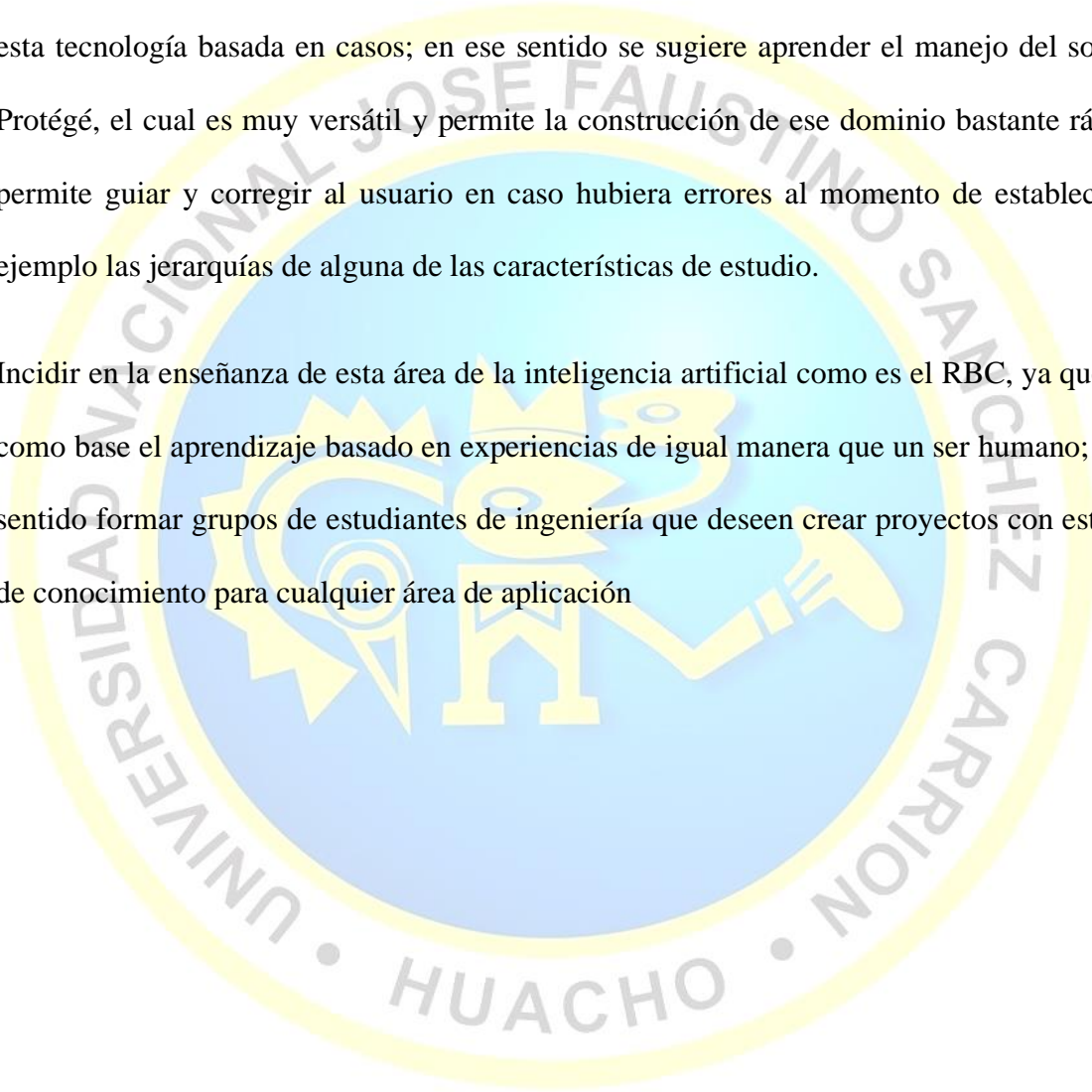
- a) Es factible crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad.
- b) Es factible crear un entorno de almacenamiento en un software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos.
- c) Existe concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico.

6.2. Recomendaciones

El estudio de la presente tesis se centró en la creación de un modelo RBC y su posterior implementación en los softwares ya mencionados; considerando que se puede establecer mejoras, el autor de esta tesis se permite dar las siguientes recomendaciones.

- a) Agregar un módulo mas al modelo, estableciendo no solo el diagnóstico, sino que una vez teniendo dicho diagnóstico se pueda prescribir el mejor tratamiento, también adquirido de una base de casos, de tal manera que este modelo se vea enriquecido y sea más completo, con la finalidad de apoyar al área de ciencias de la salud.

- b) Tomar bases de datos de los hospitales de nuestra ciudad e ir integrándolas en estos softwares que son gratuitos o freeware e inclusive son de código abierto, pudiendo mejorarlos de acuerdo al criterio del creador, con nuevas interfaces o capacidades de retorno de los resultados.
- c) Es especialmente útil establecer el dominio o alcance del estudio que se quiere desarrollar con esta tecnología basada en casos; en ese sentido se sugiere aprender el manejo del software Protégé, el cual es muy versátil y permite la construcción de ese dominio bastante rápido y permite guiar y corregir al usuario en caso hubiera errores al momento de establecer por ejemplo las jerarquías de alguna de las características de estudio.
- d) Incidir en la enseñanza de esta área de la inteligencia artificial como es el RBC, ya que tiene como base el aprendizaje basado en experiencias de igual manera que un ser humano; en ese sentido formar grupos de estudiantes de ingeniería que deseen crear proyectos con esta base de conocimiento para cualquier área de aplicación



REFERENCIAS O FUENTES DE INFORMACIÓN

Fuentes documentales

- Aamodt, A. (2004). Knowledge intensive case based reasoning in creek. Advances in case based reasoning. *7th European Conference, ECCBR 2004*, (págs. 1-152004).
- Althoff, K. D. (2011). Case-Based Reasoning. *Handbook of Software Engineering and Knowledge Engineering*.
- Bichindaritz, L., & Marling, C. (2006). Case-based reasoning in the health sciences: What's next? *In Artificial Intelligence in Medicine*, 36(2), 127-135.
- Capuñay, N. (2021). *Desarrollo de un sistema experto para ayudar en la prevención de infarto agudo al miocardio en personas de 30-75 años*. Chiclayo: Repositorio Universidad Católica Santo Toribio de Mogrovejo.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213-220.
- Feigenbaum, E. (1977). The Art of Artificial Intelligence. *International Joint Conference on Artificial Intelligence*, 1014-1029.
- Hammond, K. (1989). CHEF: a model of case-based planning-Fourth National Conference on Artificial Intelligence. Paris.
- Hersh, W. (2002). Medical Informatics: Improving healthcare through information. *JAMA: The journal of the American Medical Association*, 288(16), 1955-88.
- Huarcaya (2021). *Aplicación de la metodología Buchanan al Desarrollo de un sistema experto para el diagnóstico de la diabetes (tipo II)*. Tesis para obtener Título Profesional de: Ingeniero de Sistemas. UCV, Lima-Perú.
- INEI. (2019). *Perú: Caracterización de las condiciones de vida de la población con discapacidad, 2019*. Lima: INEI.
- Juarez, & Palma. (2015). Inteligencia Artificial. *Razonamiento Basado en Casos*, 74.
- Medrano, H. (2020). *Diseño e implementación de un sistema experto para el diagnóstico de desnutrición en niños menores de 2 Años en el área de pediatría del centro de salud Tupac Amaru-Chaupimarca*. Cerro de Pasco, Perú: Repositorio U. N. Daniel Alcides Carrión, Cerro de Pasco.
- Parker, R., Doyle, F., & Peppas, N. (2011). The intravenous route to blood glucose control. *IEEE. Engineering in Medicine and Biology*, 20(3), 65-73.
- Paruchuri, V., & Granville, B. (2020). A Case-Based Reasoning System for Aiding Physicians in Decision Making. *Intelligent Information Management*, 12, 63-74.
- Richter, M. (2005). The knowledge contained in similarity measures. *Talk at the First International Conference on Case-Based Reasoning, ICCBR'95*. Sesimbra.

- Rumui, N., Harjoko, A., & Musdholifah, A. (January de 2018). Case-Based Reasoning for Stroke Diseases Diagnosis. (*Indonesian Journal of Computing and Cybernetics Systems*, 12(1), 33~42. doi:DOI: 10.22146/ijccs.26331
- Russell, S., & Norving, P. (2004). *Inteligencia Artificial Un Enfoque moderno* (2da. ed.). (D. F. Aragón, Ed.) Salamanca, España: Pearson-Prentice Hall.
- Shahina, B. (2009). *A Case-Based Reasoning System For The Diagnosis Of Individual Sensitivity To Stress In Psychophysiology*. (V. S. Mälardalen University, Ed.) Västerås, Sweden.
- Wesley, M. (2009). *Problem Detection for situation assessment in case based reasoning for diabetes management*. Ohio, USA: School of Computer Science and the Russ College of Engineering and Technology .

Fuentes electrónicas

- Addisu, E.G., Boltena, A.S., y Amare, S. (2020). Case-Based Reasoning Framework for Malaria Diagnosis. *I.J. Information Technology and Computer Science*, 2020, 6, 31-48 Published Online December 2020 in MECS (<http://www.mecspress.org/>) DOI: 10.5815/ijitcs.2020.06.04
- Aguilar, M. (2017). *Sistema de razonamiento basado en casos, para la mejora de atención de salud en un centro rural*. Lima: UNMSM, Repositorio. Obtenido de <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/5690?show=full>
- Benamina, M y Atmani, B y Benbelkacem, S. *Diabetes Diagnosis by Case-Based Reasoning and Fuzzy Logic*, Laboratoire d'Informatique d'Oran (LIO), University of Oran 1 Ahmed Benbella (Algeria)
- Jimenez, A. y Gutiérrez, M. (2012). *Introducción a los Sistemas Basados en Conocimiento*. Obtenido de Paradigmas de la Inteligencia Artificial: <http://www.cs.us.es/cursos/iic-2012/temas/tema-02-iic03.pdf>
- Kouser R. y Thiagarajan, M. (2018). *Heart Disease Prediction System Using Artificial Neural Network, Radial Basis Function and Case Based Reasoning*. *Journal of Computational and Theoretical Nanoscience* Vol. 15, 2810–2817, 2018. <https://www.researchgate.net/publication/329838470>
- Napolitano, J. (18 de Febrero de 2021). *TechEdge*. Obtenido de Inteligencia Artificial, Machine Learning & Deep Learning, Rhinno, Reinforced Learning: <https://www.techedgegroup.com/es/blog/simulacion-entornos-inteligencia-artificial>
- Siguillito, V. (2014). *Kaggle*. Obtenido de Pima Indians Diabetes Database: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>



Anexo 01. Matriz de consistencia

Título: Modelo de un Sistema basado en casos orientado al diagnóstico médico de una enfermedad

PROBLEMA	OBJETIVOS	HIPÓTESIS	VAR.	MÉTODOS Y TÉCNICAS DE INVESTIGACIÓN
Problema General:	Objetivo General:	Hipótesis General:		
¿Se puede crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad?	Crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad	Es factible crear un modelo de un sistema basado en casos orientado al diagnóstico médico de una enfermedad	X1: Modelo basado en casos para apoyo al médico X2: Diagnóstico médico	Métodos: Estudio transversal Descriptivo Técnicas utilizadas <ul style="list-style-type: none"> • Estad. descriptivas • RBC. • Kappa de Cohen Softwares utilizados <ul style="list-style-type: none"> • Protégé • myCBR Recolección de datos: Base de datos de Pima Indians
Problemas Específicos:	Objetivos Específicos:	Hipótesis Específicas:		
1 ¿Se puede crear un entorno de almacenamiento en software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos?	1. Crear un entorno de almacenamiento en un software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos	1. Es factible crear un entorno de almacenamiento en un software para los síntomas de una enfermedad dentro del modelo del sistema basado en casos		
2 ¿Existe concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico?	2. Determinar la concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico.	2. Existe concordancia entre el diagnóstico que emite el modelo del sistema basado en casos y el diagnóstico médico		

Anexo 02. Parte de la base de datos (Pima Indians)

pres	skin	insu	mass	pedi	age	class
72	35	0	33.6	0.627	50	tested_positive
66	29	0	26.6	0.351	31	tested_negative
64	0	0	23.3	0.672	32	tested_positive
66	23	94	28.1	0.167	21	tested_negative
40	35	168	43.1	2.288	33	tested_positive
74	0	0	25.6	0.201	30	tested_negative
50	32	88	31	0.248	26	tested_positive
0	0	0	35.3	0.134	29	tested_negative
70	45	543	30.5	0.158	53	tested_positive
96	0	0	0	0.232	54	tested_positive
92	0	0	37.6	0.191	30	tested_negative
74	0	0	38	0.537	34	tested_positive
80	0	0	27.1	1.441	57	tested_negative
60	23	846	30.1	0.398	59	tested_positive
72	19	175	25.8	0.587	51	tested_positive
0	0	0	30	0.484	32	tested_positive
84	47	230	45.8	0.551	31	tested_positive
74	0	0	29.6	0.254	31	tested_positive
30	38	83	43.3	0.183	33	tested_negative
70	30	96	34.6	0.529	32	tested_positive
88	41	235	39.3	0.704	27	tested_negative
84	0	0	35.4	0.388	50	tested_negative
75	26	0	36	0.546	60	tested_negative