



Universidad Nacional José Faustino Sánchez Carrión

Facultad de Ciencias
Escuela Profesional de Estadística e Informática

Simulación de un módulo de sistema experto basado en casos como apoyo al diagnóstico médico

Tesis

Para optar el Título Profesional de Licenciado(a) en
Estadística e Informática

Autores

Nelson George Falcon Zuñiga
Assiry Sheila Penas Malpartida De Urbizagastegui

Asesora

Dra. Mirtha Sussan Trejo De Ríos

Huacho – Perú
2024



Reconocimiento - No Comercial – Sin Derivadas - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Reconocimiento: Debe otorgar el crédito correspondiente, proporcionar un enlace a la licencia e indicar si se realizaron cambios. Puede hacerlo de cualquier manera razonable, pero no de ninguna manera que sugiera que el licenciante lo respalda a usted o su uso. **No Comercial:** No puede utilizar el material con fines comerciales. **Sin Derivadas:** Si remezcla, transforma o construye sobre el material, no puede distribuir el material modificado. **Sin restricciones adicionales:** No puede aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros de hacer cualquier cosa que permita la licencia.



UNIVERSIDAD NACIONAL
JOSÉ FAUSTINO SÁNCHEZ CARRIÓN
LICENCIADA

(Resolución de Consejo Directivo N° 012-2020-SUNEDU/CD de fecha 27/01/2020)

FACULTAD DE CIENCIAS

ESCUELA PROFESIONAL DE ESTADÍSTICA E INFORMÁTICA

INFORMACIÓN

DATOS DEL AUTOR (ES):		
NOMBRES Y APELLIDOS	DNI	FECHA DE SUSTENTACIÓN
Assiry Sheila Penas Malpartida De Urbizagastegui	46145711	20/03/2024
Nelson George Falcon Zuñiga	44686465	20/03/2024
DATOS DEL ASESOR:		
NOMBRES Y APELLIDOS	DNI	CÓDIGO ORCID
Mirtha Sussan Trejo De Ríos	32812343	0000-0002-2755-9950
DATOS DE LOS MIEMROS DE JURADOS – PREGRADO/POSGRADO-MAESTRÍA-DOCTORADO:		
NOMBRES Y APELLIDOS	DNI	CÓDIGO ORCID
Miguel Ángel Aguilar Luna Victoria	17854491	0000-0003-1699-1913
Jorge Israel Santa Cruz Alvites	15736009	0000-0003-0956-5959
Julio César Valencia Bardales	15608608	0000-0002-7069-551X

Simulación de un módulo de sistema experto basado en casos como apoyo al diagnóstico médico

INFORME DE ORIGINALIDAD

18%

INDICE DE SIMILITUD

17%

FUENTES DE INTERNET

8%

PUBLICACIONES

10%

TRABAJOS DEL ESTUDIANTE

FUENTES PRIMARIAS

1	www.igi-global.com Fuente de Internet	2%
2	revistas.unjfsc.edu.pe Fuente de Internet	1%
3	www.coursehero.com Fuente de Internet	1%
4	uvadoc.uva.es Fuente de Internet	1%
5	Submitted to Universidad Nacional Jose Faustino Sanchez Carrion Trabajo del estudiante	1%
6	Submitted to Mansoura University Trabajo del estudiante	<1%
7	es.wikipedia.org Fuente de Internet	<1%
8	www.csc.liv.ac.uk Fuente de Internet	<1%

**Simulación de un módulo de sistema
experto basado en casos como apoyo al
diagnóstico médico**

DEDICATORIA:

Con especial cariño a nuestras familias que son el motor y motivo para avanzar en nuestro desarrollo profesional.

AGRADECIMIENTO

A Dios nuestro Creador, quien siempre ilumina la senda por la que recorreremos conjuntamente con nuestras familias y nos brinda enseñanzas aún en los momentos difíciles.

A nuestros compañeros de vida, por el apoyo constante al ser la fusión asertiva para nuestro diario caminar.

A nuestros profesores de la escuela profesional de Estadística e Informática, quienes nos alentaron a seguir adelante en los estudios a pesar de las dificultades en el trayecto de la vida universitaria.

ÍNDICE

ÍNDICE DE TABLAS	x
ÍNDICE DE FIGURAS	xi
RESUMEN	xii
ABSTRACT	xiii
INTRODUCCIÓN	xiv
CAPÍTULO I PLANTEAMIENTO DEL PROBLEMA	16
1.1. Descripción de la realidad problemática.....	16
1.2. Formulación del problema.....	18
1.2.1. Problema General	18
1.2.2. Problemas Específicos	18
1.3. Objetivos de la investigación.....	18
1.3.1. Objetivo General	18
1.3.2. Objetivos Específicos	18
1.4. Justificación de la investigación	19
1.5. Delimitaciones del estudio.....	20
CAPÍTULO II	22
MARCO TEÓRICO	22
2.1. Antecedentes de la investigación.....	22
2.2. Investigaciones internacionales	24
2.3. Antecedentes nacionales	25

2.3.1. Bases teóricas	26
2.3.2. Bases filosóficas	34
2.3.3. Definición de términos básicos.....	34
2.4. Hipótesis de investigación	36
2.4.3. Operacionalización de las variables.....	37
CAPÍTULO III.....	38
METODOLOGÍA	38
3.1. Diseño metodológico	38
3.2. Población y Muestra	38
3.3. Técnicas de recolección de datos.....	38
3.4. Técnicas para el procesamiento de la información Recolección de los datos ..	41
CAPÍTULO IV	46
RESULTADOS.....	46
4.1. Análisis de resultados	46
CAPÍTULO V	57
DISCUSIÓN.....	57
5.1 Discusión de resultados	57
CAPITULO VI	59
CONCLUSIONES Y RECOMENDACIONES	59
5.1. Conclusiones.....	59
5.2. Recomendaciones	60

REFERENCIAS	61
7.1. Fuentes documentales	61
7.2. Fuentes bibliográficas	62
7.3. Fuentes hemerográficas	64
7.4. Fuentes electrónicas	64

ÍNDICE DE TABLAS

Tabla 1. Características principales y diagnóstico de los pacientes en estudio	47
Tabla 2. Segmento de 9 casos con sus respectivas características.....	48
Tabla 3. Características generales de los datos en estudio.....	499
Tabla 4. Porcentaje de acierto entre el sistema experto y el médico	5454
Tabla 5. Concordancia entre el simulador y el diagnóstico médico	55
Tabla 6. Índice de Kappa de Cohen y significancia de la prueba	56

ÍNDICE DE FIGURAS

Figura 1. Distribución de las pacientes por diagnóstico de la diabetes	4848
Figura 2. Importación de los datos almacenados en fichero .csv a partir de Protégé	500
Figura 3. Clase paciente y sus respectivos slots (características)	5151
Figura 4. Ejemplo del slot diagnóstico (nombre, tipo de valor y valores permitidos)	5151
Figura 5. Medida de similitud para el slot glucosa	5252
Figura 6. Consulta ingresada y casos similares recuperados	5454

RESUMEN

La presente tesis denominada SIMULACIÓN DE UN MÓDULO DE SISTEMA EXPERTO BASADO EN CASOS COMO APOYO AL DIAGNÓSTICO MÉDICO, involucra aspectos generales sobre razonamiento basado en casos (RBC), con el objetivo general de simular un módulo de sistema de apoyo al diagnóstico médico, orientado como un método excelente para simular el raciocinio o manera de pensar del experto o médico especialista. La pregunta de investigación fue la siguiente: ¿Es posible simular un módulo de sistema experto basado en casos como apoyo al diagnóstico médico? El tipo de estudio fue de investigación tecnológica de tipo analítico no experimental; para la construcción del módulo, se tuvo en consideración la metodología de RBC la cual es parte de la inteligencia artificial y para probar la eficacia de este módulo, se tomó la base de datos de Pima Indians Diabetes Database, que considera datos de un grupo de pacientes con diferentes síntomas y características que fueron diagnosticados luego como diabéticos o sanos. Finalmente, el módulo permite escoger a una paciente en particular o ingresar ciertos síntomas y dar el diagnóstico respectivo. Como resultado, al realizar el análisis estadístico mediante el indicador Kappa de Cohen, se llegó a la conclusión que hubo concordancia entre el diagnóstico del módulo de sistema experto y el del médico.

Palabras clave: Sistema experto. Sistema de razonamiento basado en casos (SRBC). Ciclo. Similitud de casos. Base de casos. Diagnóstico. Enfermedad

ABSTRACT

This thesis called SIMULATION OF AN EXPERT SYSTEM MODULE BASED ON CASES AS SUPPORT FOR MEDICAL DIAGNOSIS, involves general aspects of reasoning based on cases, with the general objective of simulating a module of support system for medical diagnosis, oriented as a Excellent method to simulate the reasoning or way of thinking of the expert or medical specialist. The research question was the following: Is it possible to simulate an expert system module based on cases to support medical diagnosis? The type of study was of technological research of non-experimental analytical type; For the construction of the module, the case-based reasoning methodology was considered, which is part of artificial intelligence and to test the effectiveness of this module, the Pima Indians Diabetes Database database, which considers data from a group of patients with different symptoms and characteristics that were later diagnosed as diabetic or healthy. Finally, the module allows you to choose a particular patient or enter certain symptoms and give the respective diagnosis. As a result, when performing the statistical analysis using Cohen's Kappa indicator, it was concluded that there was agreement between the diagnosis of the expert system module and that of the doctor.

Theis Authors.

Keywords: Expert system. Case based reasoning system (SRBC). Cycle. Similarity of cases. Case Base Diagnosis. Disease

INTRODUCCIÓN

El aprendizaje humano es casi siempre por experiencias, aún en la enseñanza dichas experiencias se van almacenando en nuestra memoria para luego sacarlas a relucir ante cualquier problema que se suscite. Sin embargo, la situación que se nos presenta en muchas ocasiones no es exactamente la misma a la almacenada en memoria, entonces nos preguntamos, cómo hacemos útil esa experiencia, la respuesta es simple, pues por la similitud de características; y, eso es lo que sucede en el ámbito médico cuando los especialistas en salud realizan un diagnóstico.

Abdel-Badeeh y Hodhod (2002) indican que “el RBC puede proporcionar un diagnóstico apropiado para los síntomas, signos e investigaciones presentados por un paciente con el factor de certeza correspondiente pudiéndose utilizar como asistente de diagnóstico médico en la educación de los médicos de pregrado y posgrado” (pág. 304).

Por otro lado, (Juárez Herrero, 2007), propone en su tesis “...y desarrolla una arquitectura multimodal que utiliza razonamiento basado en casos, cuya base de conocimientos permite almacenar tanto patrones diagnósticos como casos temporales” (págs. 77-80).

Dicho esto, y tomando como referencia las experiencias humanas, es que en la tesis que presento he tomado este campo de la inteligencia artificial llamado razonamiento basado en casos en sus diferentes procesos como son la consideración del problema, la recuperación de casos similares, el rehúso del caso solucionado, la revisión del caso y por último el caso aprendido, todo esto utilizando a su vez softwares libres para comprobar la hipótesis general acerca de la simulación de un módulo de sistema experto orientado al diagnóstico de una enfermedad y probar finalmente si efectivamente da los mismos resultados que un experto en salud.

Para la consecución de los objetivos esta tesis tiene cinco capítulos; en el primero se da una descripción de la realidad problemática, presentando además el problema en estudio y sus respectivos objetivos. El segundo se corresponde con el marco teórico, donde se incluyen algunos antecedentes de investigación y se formulan las hipótesis respectivas correspondientes al problema de estudio. El tercer capítulo involucra la metodología, la descripción de los instrumentos a utilizar que para nuestro caso son los softwares de código abierto (open source) Protégé y MyCBR, así como las técnicas y el procesamiento de los datos e información con estos softwares. El cuarto capítulo brinda los resultados de la tesis, mostrando algunas pantallas de salida para una mejor comprensión del módulo creado mediante software y finalmente en el capítulo cinco se da una breve discusión sobre lo expuesto y las conclusiones y recomendaciones.

La implicancia de este estudio puede servir como un tutor para estudiantes de medicina o enfermería y que luego en un trabajo futuro se podría implementar en una plataforma Web, concordando con lo que explica Delgado (2002) o el Medical College of Winsconsin mencionados en los antecedentes de este trabajo.

CAPÍTULO I

PLANTEAMIENTO DEL PROBLEMA

1.1. Descripción de la realidad problemática

El diagnóstico de una enfermedad implica muchas veces el uso de herramientas como la anamnesis, la historia clínica, la exploración física y pruebas complementarias. Ambos grupos tienen un procedimiento de diagnóstico similar. Por lo general, no está directamente relacionado con el estado de la enfermedad, ya que cada condición subyacente o hallazgo del examen indicando el comienzo de la enfermedad. (Friedman, H., 2004)

Las exploraciones complementarias como pruebas de laboratorio, diagnóstico por imagen (radiografía, ecografía, etc.), endoscopías, biopsias, etc. aportan a reafirmar el diagnóstico el cual es bueno cuando ofrece resultados positivos en enfermos y negativos en pacientes sanos, con el menor rango de error posible.

Sin embargo, en muchas ocasiones no hay médicos especialistas o por otro lado no hay equipos o laboratorios suficientes, especialmente en las zonas alejadas de nuestro país como la sierra o la selva, donde el cuidado de la salud se ve comprometido cuando no se cuentan con suficientes recursos, no están presentes en los hospitales urbanos. En los hospitales rurales o regionales, donde hay poca competencia por la experiencia, los médicos deben poseer una amplia gama de habilidades que sus homólogos urbanos también puedan demostrar. Además, debido a la limitada oferta de recursos médicos, también hay una escasa disponibilidad de equipos y personal especializado. La responsabilidad de atender a una población diversa a menudo se delega a los médicos de los centros de salud rurales o alejados, a diferencia de los de los hospitales urbanos, donde tratan a un gran número de pacientes.

Además, en determinadas regiones rurales toda la asistencia sanitaria la proporciona un pequeño hospital o puesto médico donde "uno a uno" de los médicos y profesionales, así como de las enfermeras que trabajan allí, asumen muchas responsabilidades diferentes, incluso para aquellos sin especialidades. Además de manejar emergencias médicas, este personal de salud también maneja otros aspectos

médicos como la extensión comunitaria y la administración de instalaciones de atención médica.

En el ámbito de la atención médica a las comunidades indígenas, esta se brinda en establecimientos ubicados en áreas remotas, distantes de los centros urbanos. Muchos de estos lugares carecen de médicos, obstetras y enfermeras para proporcionar una atención integral, según informes defensoriales de 2015. Además, se observan indicadores preocupantes, como el 89,6% de las viviendas en Imaza que no tienen acceso a agua potable y el 94,6% sin servicios de eliminación de excretas conectados a la red pública. Estas carencias generan problemas de salud, especialmente infecciones estomacales y conjuntivitis, que afectan principalmente a los niños, exacerbando la situación de emergencia debido a la falta de especialistas. Similarmente sucede en los Distritos de Raymondi, comunidades de Manú, Machiguenga y otros (Informe Defensorial N° 134, 2017)

Así, el presente proyecto pretende crear un módulo utilizando softwares ya existentes relacionados a la inteligencia artificial, utilizando un sistema experto, en particular el de razonamiento basado en casos para apoyar al diagnóstico en la atención médica y que serviría sobre todo en los centros rurales o lugares alejados de las ciudades, donde la atención es incipiente, y la participación de un equipo médico es nula o en el mejor de los casos, hay un médico general, muchas veces inconforme, sin especialidad en particular y sin el equipamiento necesario para diagnosticar la enfermedad. En el peor de los casos se tiene a una auxiliar de enfermera o profesora como persona responsable del centro de salud o posta, a todo esto, se suma, la falta de saneamiento básico y los bajos niveles de educación. Como lo manifiesta una investigación de la Universidad Cayetano Heredia:

“... En zonas alejadas de la capital, una enfermedad o un accidente que podría tratarse con asistencia médica en la ciudad aún puede provocar la muerte. Aquellos que son jóvenes, nuevos profesionales e instituciones en las escuelas deben utilizar su facultad y capacidad para continuar preparándose.” (Nayda M., 2015).

Por otro lado, también este módulo servirá como apoyo en el desarrollo de la práctica médica en alumnos que siguen carreras de ciencias de la salud, como si fuera

un tutor brindando capacitación a los estudiantes para probar sus habilidades de diagnosticar tal o cual enfermedad.

1.2. Formulación del problema

1.2.1. Problema General

¿Es posible simular un módulo de sistema experto basado en casos como apoyo al diagnóstico médico?

1.2.2. Problemas Específicos

- ¿Es posible que el módulo de sistema experto basado en casos simulado pueda diagnosticar una enfermedad?
- ¿En qué medida el módulo de sistema experto basado en casos es tan efectivo en el diagnóstico de la enfermedad?

1.3. Objetivos de la investigación

1.3.1. Objetivo General

Simular un módulo de sistema experto basado en casos como apoyo al diagnóstico médico.

1.3.2. Objetivos Específicos

- Diagnosticar mediante el módulo de sistema experto basado en casos una enfermedad.
- Medir la efectividad del módulo de sistema experto basado en casos en el diagnóstico de la enfermedad.

1.4. Justificación de la investigación

Un sistema experto basado en casos es dinámico, fundamentándose en el razonamiento a través de la experiencia de casos. Encuentra numerosas aplicaciones en diversas áreas profesionales tales como el derecho, la medicina, la ingeniería, la administración, entre otros. Este enfoque ha demostrado ser exitoso al abordar problemas de clasificación de patrones que son comúnmente encontrados en los mencionados campos. (Aamodt y Plaza, 1994)

Dicho lo anterior, en lo que concierne a las profesiones de la salud (medicina, enfermería y otras), es de suma importancia medir el desempeño de estos profesionales, aun desde su pregrado; por lo tanto, un sistema experto basado en casos puede ser de gran utilidad, al poder capacitar a un estudiante de la salud mediante simulación. Por otro lado, en tiempo real, permitirá al practicante apoyarse en este simulador y a la vez que van acopiando práctica médica, también pueden ir almacenando los casos que tienen por consulta. En este contexto, contar con el respaldo de un sistema experto basado en casos sería sumamente beneficioso, ya que permitiría comparar los síntomas de un paciente con una base de casos previos, facilitando así un diagnóstico más preciso.

Es crucial brindar asistencia y apoyo al personal que sea profesional y pueda viajar desde pueblos o asentamientos rurales en la montaña o la selva, especialmente si trabajan en instalaciones hospitalarias pequeñas con acceso limitado a servicios hospitalarios. Esto se debe a que algunos especialistas no pueden diagnosticar una enfermedad o una emergencia médica en el momento en que es necesario.

En este contexto, el sistema experto basado en casos resulta atractivo en el campo de las ciencias de la salud, ya que un caso base preexistente contiene información almacenada, incluyendo los síntomas, el diagnóstico, el tratamiento y los resultados correspondientes a cada paciente. (Periklis, et al. 2011)

Entonces, el desarrollo de este proyecto de tesis se justifica, pues contribuirá en diferentes aspectos como apoyo en los ámbitos siguientes:

- **Enseñanza-aprendizaje:**

El sistema experto puede ser un recurso útil para los estudiantes de ciencias de la salud, ya que funciona principalmente como simbiosis con los síntomas

reales del paciente y los datos de su base de datos, ayudando a concluir un diagnóstico específico.

- **Aspecto social:**

Su utilización como sistema de apoyo por parte de proveedores de atención sanitaria especializados en regiones alejadas de las capitales redundará en beneficios sustanciales para los residentes locales. Al recibir un tratamiento médico que cumpla con los estándares necesarios para sus dolencias, estos residentes tendrán mayor confianza en los profesionales. Además, puede ayudar a implementar estrategias de atención médica proactivas para áreas remotas que se encuentran a varios kilómetros de un centro de atención médica totalmente equipado.

- **En la salud:**

Un respaldo para el médico, quien, en colaboración con sus conocimientos médicos, logrará vincular de manera más efectiva sus experiencias y proporcionar un diagnóstico más preciso. Además, podrá ampliar su base de datos para incluir a los pacientes a los que atiende, lo que ayudará a que su base de datos crezca.

- **En la economía:**

En numerosas zonas remotas, hay enfermedades comunes que podrían prevenirse fácilmente. La falta de diagnóstico oportuno y tratamiento adecuado puede convertirlas en epidemiológicas, generando un costo para el Estado. Es crucial que el ministerio de salud adopte un enfoque preventivo, ya que esto contribuiría significativamente al ahorro evitando gastos derivados de enfermedades no tratadas a tiempo.

1.5. Delimitaciones del estudio

El presente estudio incorpora técnicas relacionadas con la inteligencia artificial y que tienen su base en la minería de datos la cual tiene su base en análisis multivariado y por otro lado el razonamiento basado en casos, los cuales al integrarse forman un sistema experto y que se concentran en un software de fácil manejo.

En esta investigación se estudiará los síntomas relacionados a la enfermedad de la diabetes en el módulo de sistema experto basado en casos simulados que puedan diagnosticar esta enfermedad.

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes de la investigación

Al realizar cualquier diagnóstico, no necesariamente el médico suele utilizar el reconocimiento de patrones o características. Se trata de determinar a qué clase pertenece una muestra desconocida dentro de un conjunto de clases o hipótesis previamente definidas. El uso de estos métodos requiere un conjunto de características o medidas que puedan usarse para clasificar la muestra desconocida.

Las características del diagnóstico médico incluyen una serie de síntomas que fueron previamente evaluados en el laboratorio y luego manifestados en un individuo, dando lugar a un diagnóstico individualizado. Varias enfermedades se clasifican en distintas categorías. Es difícil determinar qué categoría de patrón de enfermedad corresponde a las características de un paciente en particular. Estas categorías pueden abarcar una variedad de enfermedades pancreáticas, con posibles inclusiones en los resultados de los análisis de sangre o en los síntomas de la diabetes. (Weiss S., 1974).

Weiss et al. (1978) En la década de 1970, la Universidad de Rutgers desarrolló CASNET (Causal Associational NETwork), un sistema experto para diagnosticar y tratar afecciones oculares, incluido el glaucoma y otras enfermedades visuales. El sistema experto utiliza una red de vínculo causal (CASNET) o modelo de red semántica para interpretar resultados específicos del paciente y describir los mecanismos fisiopatológicos y el curso clínico de las enfermedades tratadas y no tratadas a través

del proceso de inferencia. Al utilizar un modelo cualitativo del proceso de la enfermedad, proporcionar interpretaciones complejas, considerar el seguimiento de los pacientes, incluir opiniones alternativas de expertos sobre temas controvertidos y utilizar redes informáticas colaborativas para probar y actualizar su precisión, este sistema ha realizado contribuciones significativas.

(Bichindaritz & Sullivan M, 2002) elaboraron un módulo de sistema denominado sistema de atención Care-Partner. Se trata de un sistema de formación diseñado para estudiantes de medicina a través de Internet, brindándoles apoyo caso por caso que les ayude a practicar sus habilidades mostrando los resultados de la solución, a pesar del fracaso o éxito al seguir los procedimientos de diagnóstico.

Los expertos en ciencias de la salud consideran el uso del razonamiento interactivo basado en casos de sistemas expertos como una forma importante de resolver diversos problemas de diagnóstico secuencial. Además, destaca la introducción de una interfaz denominada "RBC Strategist" y su implementación en el campo del diagnóstico de fallos informatizado. (Maedche & Staab, 2001, págs. 72-79)

(Nilsson & Sollenborn, 2004) señala que se cuenta con una perspectiva integral del ámbito de la Razonamiento Basado en Casos (RBC) en el sector médico. Los sistemas de clasificación se dividen en cuatro categorías relacionadas con aplicaciones médicas: sistemas de diagnóstico, sistemas de clasificación, sistemas de enseñanza (por ejemplo, biomédicos) y sistemas de planificación. Además, se examinan los avances en la fabricación específica de razonamientos basados en casos para instalaciones médicas.

Dicho lo anterior, se observa que existe información bastante rescatable sobre este tipo de sistema experto; sin embargo, a nivel nacional es poca. A continuación, se

dan algunos antecedentes sobre los trabajos que involucran este tema a nivel internacional y nacional.

2.2. Investigaciones internacionales

(Delgado, M., 2009), cuenta con un sistema de razonamiento basado en casos clínicos en el campo de las ciencias médicas, en la plataforma web, Cuba.

Se han obtenido resultados de cierta relevancia en el desarrollo de software para el Razonamiento Basado en Casos (RBC), principalmente para aplicaciones de diagnóstico médico. Estos resultados resultaron valiosos no sólo desde un punto de vista práctico sino también desde un punto de vista teórico, al que la amplia información recopilada en la base de datos de casos contribuyó de manera importante. Su importancia educativa y clínica queda demostrada por su capacidad para proporcionar a los estudiantes de medicina y enfermería, así como a residentes y especialistas, una base de datos de casos que ayude a completar el diagnóstico y proporcione información integral relacionada con ese diagnóstico. Esto ayuda a mejorar la experiencia y la educación de los profesionales de la salud en diferentes campos.

En el Medical College of Winsconsin (2014), Estados Unidos de América.

Actualmente se está evaluando un sistema denominado PROTOISIS, basado en el Razonamiento Basado en Casos (RBC). Este sistema sirve como preparación clínica inicial para la toma de decisiones médicas. Se ha probado y utilizado con éxito en procedimientos seleccionados de diagnóstico por imágenes y tiene aplicaciones potenciales como soporte de radiología y tomografía computarizada. PROTOISIS ayuda a los profesionales a elegir los procedimientos adecuados para su uso en medicina.

El Grupo de Aplicaciones de Inteligencia Artificial (GAIA, 2011) de la Universidad Complutense desarrolló en España la plataforma JColibrí (software libre).

Permite el desarrollo rápido y sencillo de aplicaciones de razonamiento basado en casos en diversos campos. Más de una docena de universidades de todo el mundo lo utilizan como principal herramienta educativa. El impacto de esta iniciativa es evidente ya que JColibrí se ha descargado 5.000 veces en más de 70 países. Varios grupos de investigación internacionales contribuyen a su mejora y desarrollo continuo.

(Carrillo A. et al. 2011), realizó el trabajo titulado “Sistema Recomendador Orientado a la Educación Terapéutica del Paciente Diabético” desarrollado en la Universidad Complutense de Madrid, España.

Utilizando datos y mediciones analíticas del historial del paciente, el sistema ofrece recomendaciones y contenidos formativos con el propósito de ayudar al paciente a preservar su salud y mejorar su calidad de vida. Para llevar a cabo esta función, el sistema empleó la herramienta previamente mencionada, JColibrí.

2.3. Antecedentes nacionales

(Parisaca et al. 2003) En el Primer Congreso Nacional de Científicos Peruanos en la Universidad Cayetano Heredia, donde se trató el tema "Reducción de la base de datos de casos por conglomerado"

A pesar de no contar con trabajos sobre SRBC que se enfoquen únicamente en diagnosticar y tratar enfermedades, a investigadores peruanos se les solicitó información sobre su sistema. La propuesta sugiere un método para disminuir la base de casos en los sistemas RBC agrupando los casos utilizando tecnología de inteligencia artificial especializada y luego reduciéndola en función de las características del sistema.

(Cortez V. Augusto et al. 2010). SRBC es aplicable a sistemas de línea de productos de software.

Utilizando ejemplos, casos o datos conocidos, este estudio busca tomar decisiones sobre nuevos casos. En la minería de datos, podemos asignar una

clase a instancias arbitrarias simplemente viendo clases similares, como parte de nuestras actividades de clasificación. Asignaremos una nueva muestra al grupo con los individuos más similares en las tareas de agrupamiento.

(Sistema RBC en la Contraloría General de la República, 2009) en el ámbito de gestión.

Se creó e implementó un sistema de gestión del conocimiento en la Contraloría General de la República a través de la experiencia.

(Cruz, L., 2009). Modelo Híbrido para el Análisis de Riesgo Crediticio en Pymes de la Caja Municipal de Arequipa, en el ámbito financiero.

En este trabajo, presentan un enfoque híbrido que incorpora redes neuronales y razonamiento que distingue entre mayúsculas y minúsculas para evaluar las solicitudes de crédito MES. El objetivo es ayudar a las instituciones financieras a realizar análisis de crédito.

2.3.1. Bases teóricas

En la realización de cualquier diagnóstico, que no necesariamente sea médico, se suele emplear el reconocimiento de patrones o características. Es necesario identificar la clase dentro de un conjunto previamente establecido de hipótesis o clases para determinar la identidad del patrón desconocido. Estos métodos requieren un conjunto de atributos o medidas para categorizar la tendencia no identificada.

El diagnóstico medicinal implica identificar los síntomas del paciente o los resultados de las pruebas de laboratorio, que luego se utilizan para realizar diagnósticos. Las enfermedades se clasifican en grupos específicos. Una tarea difícil es distinguir entre los diferentes tipos de enfermedad a los que pertenece el patrón desconocido mediante la observación de las características de un paciente en particular. Varios tipos de enfermedades cardíacas se pueden clasificar en categorías y sus características pueden incluir pruebas de presión arterial o síntomas como dolor de cabeza. (Weis, 1974).

Desde la década de 1970 hasta 1978, la Universidad de Rutgers empleó el sistema CASNET de Weiss y Kulikowski, un sistema experto diseñado para ayudar en el diagnóstico y tratamiento de enfermedades oculares, incluido el glaucoma, así como otros trastornos visuales. Esto se logró a través de la investigación. El sistema experto utiliza una red de vínculo causal (CASNET) o modelo de red semántica para interpretar los resultados de los pacientes a través de un proceso de inferencia que describe el mecanismo fisiopatológico y el curso clínico de la enfermedad, tanto tratada como no tratada. El sistema genera descripciones complejas de modelos cualitativos de procesos patológicos, profundiza en el seguimiento detallado de los pacientes, incorpora opiniones alternativas de expertos sobre temas controvertidos y los envía para su prueba y actualización a través de una red informática colaborativa con testigos presenciales para estudiar el glaucoma.

Bichindaritz y Sullivan (2002) diseñaron un sistema denominado Care-Partner para brindar atención a los estudiantes de medicina a través de Internet. Este sistema proporciona tutorías, permitiendo a los estudiantes practicar casos y evaluar sus habilidades. Además, Care-Partner ofrece retroalimentación al estudiante indicando si ha fallado, cumplido parcialmente o cumplido completamente con todas las normas.

Según los expertos médicos, la intervención de un sistema experto que utiliza un razonamiento interactivo basado en casos es un método eficaz para abordar muchos problemas de diagnóstico secuenciales. También se destaca el uso de una interfaz denominada "RBC strategist", demostrando su aplicación en el campo de los diagnósticos fallidos a través del ordenador. (Maedche et.al., 2001).

Según Nilson y Sollenborn, el Razonamiento Basado en Casos (RBC) es bien conocido en el campo médico. Los sistemas de diagnóstico, clasificación, tutoría y planificación son los cuatro tipos principales de sistemas que abarcan aplicaciones médicas. Estos sistemas se pueden clasificar en categorías específicas. También proporcionan información sobre tendencias particulares de construcción de RBC para sistemas médicos. (Nilson et al., 2004).

Dicho lo anterior, se observa que existe información bastante rescatable sobre este tipo de sistema experto; sin embargo, a nivel nacional es poca. A continuación, se dan algunos antecedentes sobre los trabajos que involucran este tema a nivel internacional y nacional.

La temática principal de esta tesis tiene sus fundamentos en la informática, específicamente en el campo de la inteligencia artificial, que es la disciplina de la cual surgen distintos avances, incluyendo agentes inteligentes como el razonamiento basado en conocimiento y/o en casos. Esta estrategia busca acercar la tecnología informática a los usuarios, facilitando así diversos escenarios en los que trabajan profesionales de distintas áreas. A continuación, se presentan algunas bases teóricas que se emplearán en el desarrollo del proyecto.

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

Los SBC se basan en el conocimiento y existen dos tipos de conocimiento en cada área temática: información pública e información privada. El conocimiento público incluye definiciones, hechos y teorías publicadas; La experiencia suele consistir en más que eso. La mayoría de los especialistas humanos tienen acceso al conocimiento privado. Los sistemas basados en el conocimiento requieren una comprensión del conocimiento que se posee dentro de ese campo, y es crucial para exponer y reproducir este conocimiento en el SBC. (Alonso, et. al.,2012)

Los usuarios (expertos) deben definir sus conocimientos y el sistema los integra directamente en las aplicaciones existentes. Por este motivo conviven dos paradigmas de sistemas inteligentes:

- a) Sistemas de conocimiento que están estructurados cognitivamente y sus características incluyen:
 - Más que tener habilidades intelectuales generales, el conocimiento de la naturaleza y la naturaleza de los problemas es esencial para resolverlos.
 - La división entre la codificación de información en la máquina de base de conocimiento y las técnicas deductivas (motor de inferencia).
- b) Marcos expertos centrados en la funcionalidad o el comportamiento, siendo sus características:
 - Cuando un experto humano resuelve problemas en campos especializados, se examina su actividad para determinar si puede ser asistido o emulado.
 - Capacidad para mejorar gradualmente la experiencia y comunicarse con los usuarios, explicando sus razonamientos.

Según los paradigmas descritos anteriormente, el SBC se considera un sistema que utiliza conocimientos sobre diversos dominios para abordar problemas en esa área. Una persona que tenga experiencia en el dominio del problema recibiría una solución casi idéntica a la suya cuando se enfrente al mismo problema.

- **Tipos de sistemas basados en conocimiento**

Los métodos para obtener conocimiento y representarlo difieren según el tipo de conocimiento, como descriptivo, procedimental o metaconocimiento. Asimismo, existen varios paradigmas alternativos de resolución de problemas, como el paradigma lógico, la búsqueda heurística, el enfoque basado en el conocimiento, el enfoque conexionista y el enfoque orientado a la experiencia. El paradigma basado en la experiencia se compone de RBR, RBC y ontologías. Le daremos más atención a este paradigma, específicamente a RBC.

- **Base de conocimiento**

Incluye información relevante para la tarea. Codifica el conocimiento dentro de una base utilizando formalismos representacionales. La etapa de memoria a largo plazo es donde se reflejan la estructura cognitiva y los procesos humanos, y contiene los hechos y conocimientos sobre la experiencia que te brindaron.

- **Motor de inferencias**

Estrategias de gestión y uso del conocimiento. Al utilizar datos de entrada, noticias o conocimientos, el sistema puede razonar y producir resultados. Este motor controla la secuencia de pasos en el razonamiento, entrada y salida. Esta etapa puede interpretarse como una máquina que realiza funciones de razonamiento para resolver problemas.

- **Sistema experto**

Los sistemas expertos fueron introducidos por el Dr. Edward E. Feigenbaum de la Universidad de Stanford, quien enfatizó la importancia de tener experiencia en un campo particular en lugar de simplemente técnicas de programación y formalismos para resolver problemas informáticos. (Feigenbaum, 1977).

Finalmente se puede decir que la inteligencia artificial ve los sistemas expertos como aquellos que intentan emular los comportamientos de individuos con experiencia en un área particular de especialización.

- **Sistema experto basado en reglas**

El procedimiento permite la creación de un proceso deductivo a partir de un conjunto inicial de reglas, que puede utilizarse hasta que no haya más reglas a las que hacer referencia. El conocimiento se fundamenta en estas reglas.

- **Sistema experto basado en redes bayesianas**

Es una aplicación de inteligencia artificial que utiliza el marco teórico de las redes bayesianas para modelar y resolver problemas complejos. Las redes bayesianas son un tipo de modelo probabilístico que representa las relaciones de dependencia entre diferentes variables mediante un grafo acíclico dirigido, asimismo se puede decir que un sistema experto basado en redes bayesianas combina el conocimiento experto con la capacidad de modelar y manejar la incertidumbre, lo que lo hace útil para situaciones en las que se deben tomar decisiones bajo condiciones de incertidumbre y variabilidad.

- **Sistema experto basado en casos**

El sistema se basa en experiencias pasadas, poseídas por el sistema o el experto, y las conecta con posibles soluciones. Desarrolla estos hallazgos. Los tres métodos que se utilizan actualmente son los más similares a la forma en que piensan los humanos.

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

Un enfoque de inteligencia artificial consiste en utilizar experiencias similares para resolver problemas, que se almacenan en casos y se revisan en busca de similitudes con el caso específico. Luego, el sistema recupera la mejor similitud realizando una comparación simple de los datos mediante la consulta. Al utilizar un modelo de dominio con conocimiento general explícito, RBC trabaja reutilizando el mejor caso o casos coincidentes como parte de la solución de su problema. Luego se revisa la solución para que se ajuste al problema actual y la nueva solución del caso se puede almacenar en la base del caso para referencia futura. (Aamodt, 2004).

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

El razonamiento basado en casos es un enfoque de la inteligencia artificial que utiliza experiencias pasadas para resolver nuevos problemas. El ciclo del proceso de razonamiento basado en casos normalmente consta de los siguientes pasos:

- **Recuperación (Retrieval):**
 - **Recopilación de Casos:** Se recopilan y almacenan casos anteriores que contienen información sobre situaciones similares o problemas previos.
 - **Búsqueda y Selección:** Se busca en la base de casos para encontrar casos relevantes que sean similares al problema actual.
- **Reutilización (Reuse):**
 - **Adaptación de Casos:** Se identifican y seleccionan los aspectos relevantes de los casos recuperados que pueden aplicarse al problema actual.
 - **Aplicación de Soluciones Anteriores:** Se toma la solución o parte de la solución de casos anteriores y se aplica al problema actual.
- **Revisión (Revise):**
 - **Evaluación de la Solución:** La solución obtenida se evalúa en términos de su relevancia y eficacia para el problema actual.
 - **Ajuste de la Solución:** Si es necesario, se ajusta la solución obtenida para adaptarse mejor al problema específico en cuestión.
- **Retención (Retain):**
 - **Actualización de la Base de Casos:** Si la solución ha demostrado ser efectiva, se puede almacenar el caso actualizado en la base de casos para su uso futuro.
 - **Razonamiento (Reasoning): Aplicación de Conocimiento General:** Si no se encuentra una solución adecuada en la base de casos, se puede recurrir a la aplicación de reglas generales o conocimiento experto para razonar y resolver el problema.
- **Aprendizaje (Learning):**
 - **Actualización del Sistema:** Los casos nuevos y la experiencia obtenida durante el proceso de resolución se utilizan para mejorar y actualizar el sistema de razonamiento basado en casos.

- Aprendizaje Continuo: El sistema puede adaptarse y mejorar su rendimiento a lo largo del tiempo a medida que se enfrenta a nuevos problemas y casos.

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

La estructura de un sistema de Razonamiento Basado en Casos (RBC) puede variar según la implementación y los requisitos específicos de la aplicación. Sin embargo, se puede describir una estructura básica que a menudo incluye cuatro contenedores principales. Estos son:

- **Base de Casos (Case Base):**
 - Descripción: Contiene una colección de casos pasados, cada uno de los cuales incluye información sobre un problema específico y su solución asociada.
 - Propósito: Proporciona ejemplos anteriores que sirven como base de conocimiento para la resolución de problemas actuales.
- **Base de Conocimiento (Knowledge Base):**
 - Descripción: Contiene reglas, hechos y principios generales que son aplicables a una variedad de problemas dentro del dominio de aplicación.
 - Propósito: Ofrece conocimiento contextual y general que puede ser utilizado junto con la información específica de los casos para realizar razonamientos más amplios.
- **Base de Adaptación (Adaptation Base):**
 - Descripción: Almacena adaptaciones o transformaciones aplicadas a casos anteriores para ajustar soluciones existentes a nuevos problemas.
 - Propósito: Permite la reutilización de soluciones existentes con ajustes según la especificidad del problema actual.
- **Base de Experiencia (Experience Base):**
 - Descripción: Contiene información adicional relacionada con la ejecución y los resultados de casos anteriores, como datos sobre la eficacia de las soluciones aplicadas en situaciones específicas.

- Propósito: Proporciona información valiosa sobre la efectividad de las soluciones en diversas circunstancias y puede contribuir al aprendizaje continuo del sistema.

Esta estructura básica con cuatro contenedores refleja la idea fundamental del RBC, que es aprender de casos pasados para resolver problemas futuros. La base de casos proporciona ejemplos específicos, la base de conocimiento aporta reglas generales, la base de adaptación permite ajustar soluciones, y la base de experiencia ofrece información adicional sobre la ejecución y resultados de casos anteriores.

Es importante tener en cuenta que la implementación exacta puede variar según las necesidades específicas del sistema y el dominio de aplicación. Algunos sistemas pueden tener estructuras adicionales o características específicas según los requisitos particulares de la aplicación.

- **Validez de una prueba diagnóstica**

Indicando si la prueba es positiva o negativa para cada paciente, indicando si está sano o enfermo. Como resultado, si se encuentra allí y no en la enfermedad que se está estudiando, generalmente es positivo. La validez se puede determinar evaluando la precisión del indicador estadístico Kappa o utilizando los indicadores de sensibilidad y especificidad.

- **Características de una prueba diagnóstica**

Para Prieto realizar un test de diagnósticos es indispensable por ello se requiere las siguientes condiciones que principalmente son tres (Prieto, 2005):

- **Validez:** Se considera válido probar eficientemente los resultados obtenidos mediante esta prueba con otras pruebas más rigurosas y complejas.
- **Reproductividad,** es la capacidad de una prueba para producir los mismos resultados cuando se usa repetidamente en circunstancias similares. La capacidad reproductiva está determinada por la variabilidad biológica del evento observado, introducida por el propio observador y resultante de la propia prueba.
- **Seguridad:** La certeza de predecir la presencia o ausencia de una enfermedad en un paciente se llama seguridad.

- **Indicador estadístico Kappa**

Al evaluar un conjunto de sujetos u objetos, ofrece un indicador del nivel de acuerdo entre dos observadores, jueces y situaciones. Nuestro caso de estudio evaluaría el diagnóstico proporcionado por el módulo del sistema experto frente al diagnóstico auténtico realizado por simulación o los diagnósticos clínicos del médico. Este es nuestro objetivo. Este número varía de 0 a 1, donde 0 indica desacuerdo y 1 sugiere máximo acuerdo, lo que implica acuerdo compartido en ambos escenarios.

2.3.2. Bases filosóficas

Las bases filosóficas que se utilizarán para el desarrollo del proyecto se detallan a continuación:

- **I.A en la filosofía**

La intersección entre la inteligencia artificial y la filosofía es un campo en constante evolución a medida que las tecnologías avanzan y surgen nuevas preguntas éticas y conceptuales. Los filósofos contemporáneos están cada vez más involucrados en debates sobre el desarrollo y la implementación de la inteligencia artificial, explorando las implicaciones filosóficas de estas tecnologías emergentes.

Según él, la inteligencia artificial implica la creación de máquinas que pueden realizar tareas que serían imposibles de realizar para los humanos. Cada concepto sigue una larga trayectoria que no es ni recta ni uniforme, y su apariencia no se puede medir con precisión.

- **(Hawking, S., 2014)**

Basándose en la inteligencia artificial, Hawking dijo que "es seguro decir que la inteligencia artificial podría ser el fin de la humanidad". El científico advirtió en una entrevista con la BBC que el fin de la humanidad puede ser inminente debido a la codificación de la inteligencia artificial.

2.3.3. Definición de términos básicos

- **Diagnóstico médico**

La evaluación clínica del estado fisiológico del paciente es sólo indicativa de necesidades invasivas. El Profesional es el único clínico que puede brindar

orientación sobre el manejo del paciente o su diagnóstico. Los expertos clínicos son responsables de tratar a los pacientes y brindar tratamiento de emergencia según el caso de cada individuo. Para diagnosticar a un paciente es necesario considerar dos aspectos de la lógica, es decir, el análisis y la síntesis, lo que implica el uso de diversas herramientas como la anamnesis, la historia clínica, el examen físico y exámenes adicionales. (Altman, 1994)

- **Especificidad de una prueba**

La especificidad se refiere a la tasa de verdaderos negativos o la probabilidad de un resultado negativo en la prueba en ausencia de enfermedad. El diagnóstico de la enfermedad en individuos sanos es el mismo que el de los resultados falsos positivos. Es específico para personas sanas.

$$\text{Especificidad} = \frac{\text{Verdaderos Negativos}}{\text{Verdaderos Negativos} + \text{Falsos Positivos}}$$

- **Sensibilidad de una prueba**

La capacidad de nuestro estimador para presentar casos verdaderos enfermos como casos positivos, con una tasa de identificación adecuada, está indicada por la sensibilidad. Por lo tanto, esto significa que la prueba puede detectar enfermedades en personas enfermas.

$$\text{Sensibilidad} = \frac{\text{Verdaderos Positivos}}{\text{Verdaderos Positivos} + \text{Falsos Negativos}}$$

La sensibilidad es la fracción de verdaderos positivos (FVP) y la especificidad la fracción de verdaderos negativos (FVN).

- **Simulación**

La simulación se refiere al proceso de imitar el comportamiento de un sistema, proceso o fenómeno en un entorno controlado para comprender, analizar o prever su funcionamiento. Es una técnica utilizada en diversas disciplinas para estudiar el comportamiento de sistemas complejos o para entrenar a individuos en situaciones específicas. La simulación puede tener aplicaciones en campos como la ciencia, la ingeniería, la medicina, la educación, la economía y más. (Shanon, 1976).

- **Inteligencia artificial**

Los sistemas informáticos reproducen los procesos cognitivos que permiten el aprendizaje, el análisis y la toma de decisiones racionales basados en datos recopilados y comprendidos en organismos vivos y humanos.

- **Razonamiento basado en casos**

Un sistema experto se basa en la resolución de problemas basándose en las soluciones de problemas anteriores para tomar decisiones específicas de su dominio. (Aamodt et al., 2004)

2.4. Hipótesis de investigación

2.4.1. Hipótesis General

El módulo simulado de sistema experto basado en casos posibilita el apoyo al diagnóstico médico.

2.4.2. Hipótesis Específicas

- El módulo simulado de sistema experto basado en casos permite diagnosticar una enfermedad.
- El módulo de sistema experto basado en casos es de alta efectividad en el diagnóstico de la enfermedad.

2.4.3. Operacionalización de las variables

Variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores	Instrumentos a utilizar
Módulo de sistema experto	Alberga el conocimiento, realiza inferencias y proporciona respuestas o soluciones basadas en dicho conocimiento (Diccionario RAE, 2010)	Sistema de instrucciones interconectadas para determinar el tratamiento médico.	Antecedentes familiares Embarazos Nivel de glucosa a las 2 horas de alimentos Grosor de piel en tríceps Índice de masa corporal Nivel de insulina a las 2 horas de alimentos Presión diastólica Prueba bilirrubina Historial ginecológico Diagnóstico	Rango normal de características del estudio del paciente. (Guardar en la base de datos)	SOFTWARES PROTÉGÉ Y MYCBBR ESTADÍSTICO KAPPA DE COHEN

CAPÍTULO III

METODOLOGÍA

3.1. Diseño metodológico

Este proyecto tiene como objetivo crear avances tecnológicos que resulten en la producción de módulos de sistemas expertos utilizando herramientas de simulación, sin los costes asociados. Se espera que incluya módulos de diagnóstico que ayuden a los profesionales de la salud a diagnosticar determinadas enfermedades.

El nivel es descriptivo – analítico, no experimental, también con enfoque cuantitativo y se mide mediante un indicador que mide la precisión del diagnóstico proporcionado por el módulo del sistema experto y proporciona un diagnóstico imparcial.

3.2. Población y Muestra

Por ser un estudio de tipo tecnológico, no se ha seleccionado una muestra, sino más bien se ha tomado una base de datos concernientes a setecientos sesenta y ocho (768) pacientes diabéticos con sus respectivos síntomas desde la página DataHub y que luego se comparó con el análisis médico.

3.3. Técnicas de recolección de datos

Para recopilar información se utilizará el sitio web www.datahub ya que es una de las muchas plataformas que almacenan datos reales de calidad de diferentes tipos y al mismo tiempo comparten conocimientos. De esta base de datos se extraerá una muestra de conveniencia no probabilística de cincuenta (50) unidades de observación. Para ello se consideraron los siguientes enlaces web:

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

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

La tecnología DataHub sirve como contenedor de base de datos y herramienta de automatización. El archivo diabetes.csv se utilizará para seleccionar los datos sobre prevalencia y diagnóstico de diabetes de esta página. La base de casos consta de setecientos sesenta y ocho (768) casos que se obtuvieron de pacientes con los mismos rasgos mostradas en la tabla 1.

Tabla 1. Atributos o características de los pacientes en la base de casos

DIMENSIÓN	Nº Caract.	CARACTERÍSTICAS	Tipo de dato*	Rango normal	Unidad	Min-Prom-Máx
Demographics	1	Residence	C	{ Urban, Rural}	-	-
	2	Occupation	C	{Farmer, ...}	-	-
	3	Gender	C	{Male, Female}	-	-
	4	Age	N	-	year	29-48.117-74
	5	BMI	N	18.5 - 25	kg/m2	20-33.117-45
Diabetes Lab Tests	6	HbA1C	N	<=5	mmol/L	5-6.373-7.4
	7	2h PG	N	< 139	mg/dl	165-202.733-235
	8	FPG	N	< 99	mg/dl	96-129.633-156
Hematological Profile	9	Prothrombin INR	N	0 - 1	%	1-1.16-1.4
	10	Red cell count	N	4.2 - 5.4	10 ⁶ /cmm	3.8-5.194-5.88
	11	Hgb	N	Dic-16	g/dL	9.8-12.332-13.4
	12	Haematocrit (PCV)	N	37 - 47	vol%	31.1-35.215-36.8
	13	MCV	N	80 - 90	fl	26.8-71.908-76.4
	14	MCH	N	27 - 32	pg	3.3-25.47-29.4
	15	MCHC	N	30 - 37	%	1.8-35.465-41.7
	16	Platelet count	N	150 - 400	10 ³ /cmm	135-316.183-2000
	17	White cell count	N	4-Nov	10 ³ /cmm	6-8.055-9.2
	18	Basophils	N	0 - 1	%	0-1.013-5
	19	Lymphocytes	N	20 - 45	%	21.2-25.768-29
	20	Monocytes	N	2-Oct	%	1.7-2.942-4
	21	Eosinophils	N	1-Abr	%	1-1.897-3.4
Symptoms	22	Urination Frecuency	O	-	-	-
	23	Vision	O	-	-	-
	24	Thirst	O	-	-	-
	25	Hunger	O	-	-	-
	26	Fatigue	O	-	-	-
Kidney Function Lab Test	27	Serum Potassium	N	3.5 - 5.3	mEq/L	2.4-3.767-4.3
	28	Serum Urea	N	May-50	mg/dL	17-31.56-67
	29	Serum Uric acid	N	3.0 - 7.0	mg/dL	3-4.237-7.9
	30	Serum Creatinine	N	0.7 - 1.4	mg/dL	0.9-1.35-3.6
	31	Serum Sodium	N	135 - 150	mEq/L	134-137.833-158
Lipid Profile	32	LDL cholesterol	N	0 - 130	mg/dL	50-94.917-170
	33	Total cholesterol	N	0 - 200	mg/dL	158-209.367-275
	34	Triglycerides	N	60 - 160	mg/dL	78-144.767-189

Tumor Markers	35	HDL cholesterol	N	45 - 65	mg/dL	30-55.533-65
	36	FERRITIN	N	28 - 397	ng/mL	-
	37	AFP Serum	N	0.5 - 5.5	IU/ml	-
	38	CA-125	N	1.9 - 16.3	U/mL	-
Urine Analysis	39	Protein	O	-	-	-
	40	Blood	O	-	-	-
	41	Bilirubin	O	-	-	-
	42	Glucose	O	-	-	-
	43	Ketones	O	-	-	-
	44	Urobilogen	O	-	-	-
	45	Pus	O	-	-	-
	46	RBCs	O	-	-	-
	47	Crystals	O	-	-	-
Liver Function Tests	48	S. Albumin	N	3.5 – 5.0	g/dL	1.9-4.082-5.4
	49	Total Bilirubin	N	0.0 – 1.0	mg/dL	0.8-1.317-3
	50	Direct Bilirubin	N	0.0 – 0.3	mg/dL	0.3-0.533-1.6
	51	SGOT (AST)	N	0 – 40	U/L	35-54.567-165
	52	SGPT (ALT)	N	0 – 45	U/L	35-57.317-183
	53	Alk. Phosphatase	N	64 - 306	U/L	170-214.2-360
	54	γ GT	N	7 – 32	U/L	18-35.833-98
	55	Total Protein	N	6.0 – 8.7	g/dL	3.1-4.858-8.7
Females History	56	Amenorrhea	O	-	-	-
	57	Birth	O	-	-	-
	58	Dysmenorrhea	O	-	-	-
Diagnosis	59	Diabetes Diagnosis	C	-	-	-
	60	Nephropathy check	C	-	-	-
	61	Hypercholesteremia check	C	-	-	-
	62	Tumor Markers	C	-	-	-
	63	Liver problem	C	-	-	-
	64	Radiological Diagnosis	C	-	-	-

* Tipo de dato= {N=Numérico, C=Categorico, O=Ordinal}

Fuente: Base de datos coleccionada del Registro de salud electrónico (EHR) Mansura University y colectados por Pima Indians Diabetes Database (<https://www.openml.org/d/37>)

Los datos se guardarán en una hoja de cálculo para su transferencia y se evaluarán las notas y columnas de cada paciente que contienen síntomas y factores de riesgo. Los parámetros de evaluación para el diagnóstico de enfermedades han sido evaluados previamente:

Tabla 2. Niveles de los síntomas de una enfermedad

Parámetros de Evaluación	Niveles de la dolencia	Parámetros de Evaluación	Niveles de la dolencia
Síntoma 1	A	Riesgo 1	A
	B		B
	C		C
Síntoma 2	A	Riesgo 2	A
	B		B
	C		C
Síntoma 3	A	Riesgo 3	A
	B		B
	C		C
Síntoma 4	A	Riesgo 4	A
	B		B
	C		C
Síntoma 5	A	Riesgo 5	A
	B		B
	C		C
Síntoma 6	A	Riesgo 6	A
	B		B
	C		C

3.4. Técnicas para el procesamiento de la información

Recolección de los datos

Debido a la investigación tecnológica, los módulos de sistemas expertos basados en casos utilizarán principalmente software libre como:

- **MyCBR**

Es un acrónimo que significa "my Case-Based Reasoning". MyCBR es un sistema de software que se utiliza para desarrollar y gestionar sistemas basados en el razonamiento basado en casos (RBC). El razonamiento basado en casos es un enfoque en inteligencia artificial que implica resolver problemas al recuperar y reutilizar soluciones de casos anteriores.

El objetivo principal de myCBR es disminuir la cantidad de esfuerzo requerido en el diseño de CBR que requieren medidas de similitud intensivas en conocimiento. MyCbra proporciona interfaces gráficas de usuario fáciles de usar para modelar atributos específicos, similitudes y calidad de recuperación de pruebas. Este proyecto de código abierto contiene herramientas que pueden generar automáticamente representaciones de casos a partir de datos sin procesar.

- **Protégé**

El Centro de Investigación en Informática Biomédica de la Universidad de Stanford, en colaboración con la Universidad de Manchester, ha creado una herramienta de código abierto con una sólida comunidad académica y usuarios empresariales, que se utiliza para crear soluciones utilizando su experiencia en diversos campos, como la biomedicina, el comercio electrónico y el modelado organizacional. Esta herramienta es gratuita. Este programa ofrece una interfaz gráfica fácil de usar para crear ontologías y emplea clasificadores deductivos para verificar la coherencia del modelo y generar nueva información a través del análisis de una ontología. Es una aplicación basada en Java, que depende en gran medida de swing para su interfaz completa. Actualmente, más de 240.000 personas están registradas en el sitio web.

○ **Instrumento descriptivo para medir la confiabilidad del módulo CBR**

Se realizará un análisis descriptivo de casos de estudio simulados, incluidas muestras sanas y enfermas, para evaluar la confiabilidad de los resultados obtenidos del módulo del sistema experto basado en casos. La comparación entre el diagnóstico del médico y la simulación proporcionada por las historias clínicas y los casos de enfermedad se incluye en la Tabla 2.

Como resultado, se controlará el porcentaje de éxito entre el módulo experto y el médico (ver Tabla 2). Mediante la utilización de estadísticas, el estudio también tendrá como objetivo validar y validar científicamente la precisión del módulo RBC propuesto para hacer coincidir el diagnóstico clínico con el de los pacientes afectados. El índice estadístico Kappa se emplea para medir el nivel de acuerdo entre observadores o jueces y un grupo de sujetos u objetos.

Tabla 3. Comparación simulada de los diagnósticos

Caso Evaluado	Experto (médico)	Modelo RBC	Caso Evaluado	Experto (médico)	Modelo RBC
Caso_2001	Sano	Diabetes	Caso_2026	Diabetes	Diabetes
Caso_2002	Diabetes	Diabetes	Caso_2027	Diabetes	Diabetes
Caso_2003	Diabetes	Diabetes	Caso_2028	Diabetes	Diabetes
Caso_2004	Diabetes	Diabetes	Caso_2029	Diabetes	Diabetes
Caso_2005	Diabetes	Diabetes	Caso_2030	Diabetes	Diabetes
Caso_2006	Diabetes	Diabetes	Caso_2031	Diabetes	Diabetes
Caso_2007	Diabetes	Diabetes	Caso_2032	Diabetes	Diabetes
Caso_2008	Diabetes	Diabetes	Caso_2033	Diabetes	Diabetes
Caso_2009	Diabetes	Diabetes	Caso_2034	Diabetes	Diabetes
Caso_2010	Diabetes	Diabetes	Caso_2035	Diabetes	Diabetes
Caso_2011	Diabetes	Diabetes	Caso_2036	Diabetes	Diabetes
Caso_2012	Diabetes	Diabetes	Caso_2037	Diabetes	Diabetes
Caso_2013	Diabetes	Diabetes	Caso_2038	Diabetes	Diabetes
Caso_2014	Diabetes	Diabetes	Caso_2039	Diabetes	Diabetes
Caso_2015	Diabetes	Diabetes	Caso_2040	Diabetes	Diabetes
Caso_2016	Diabetes	Diabetes	Caso_2041	Diabetes	Diabetes
Caso_2017	Diabetes	Diabetes	Caso_2042	Diabetes	Diabetes
Caso_2018	Diabetes	Diabetes	Caso_2043	Diabetes	Diabetes
Caso_2019	Diabetes	Diabetes	Caso_2044	Diabetes	Diabetes
Caso_2020	Diabetes	Diabetes	Caso_2045	Diabetes	Diabetes
Caso_2021	Diabetes	Diabetes	Caso_2046	Diabetes	Sano
Caso_2022	Diabetes	Diabetes	Caso_2047	Diabetes	Diabetes
Caso_2023	Diabetes	Sano	Caso_2048	Diabetes	Diabetes
Caso_2024	Diabetes	Diabetes	Caso_2049	Diabetes	Diabetes
Caso_2025	Sano	Diabetes	Caso_2050	Diabetes	Diabetes

○ **Indicador estadístico para medir la confiabilidad del módulo CBR**

El índice Kappa de Cohen (k) es un método para medir el grado de acuerdo entre múltiples observadores, técnicas y otros sujetos sobre un fenómeno observado. Esto significa que para que el módulo del sistema experto tenga mayor confiabilidad científica, se debe estimar utilizando la siguiente ecuación [Cohen, 1960]:

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

Los valores P_0 y P_e son hallados partiendo de una tabla de contingencia de 2x2 en la que se tiene clasificada la concordancia observada y la concordancia por azar de dos métodos, mostrados en la tabla siguiente:

Tabla 4. Concordancia observada y esperada para el acuerdo entre el modelo RBC y el diagnóstico médico

MÉTODO 1	MÉTODO 2				Total Frec.
	Clasificación 1		Clasificación 2		
	Concord. observada	Concord. esperada	Concord. observada	Concord. esperada	
Clasificación 1	a	A	b	B	a + b
Clasificación 2	c	C	d	D	c + d
TOTAL	a + c	A + B	b + d	B + D	n

Proporción de concordancia observada global (P_0)

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

Proporción de concordancia esperada global (P_e)

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

El coeficiente Kappa se calcula utilizando las proporciones de concordancia observada (P_0), concordancia esperada por azar (P_e), y la concordancia máxima posible no debida al azar ($1 - P_e$). El numerador del coeficiente Kappa representa la diferencia entre la proporción de acuerdo observado y el acuerdo esperado, mientras que el denominador es la diferencia entre un acuerdo total y la proporción esperada por azar. En resumen, el índice Kappa corrige el acuerdo solo por azar, reflejando la proporción del acuerdo observado que supera la proporción esperada por azar. Si el valor es 1, indicaría una concordancia perfecta (100% de acuerdo) con cero proporción por azar. Cuando el valor es 0, hay un total desacuerdo, y la proporción esperada por azar se iguala a la proporción observada.

La evaluación del grado de acuerdo se da mediante el valor de k y clasificado según la escala mostrada en la tabla 4.

Tabla 5. Valoración del índice Kappa

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

Para la construcción del módulo de sistema experto propuesto, como se dijo anteriormente, se utilizaron los softwares MyCBR y PROTÉGÉ, el primero especialmente utilizado para comparar el caso en consulta a través de las medidas de similitud y recuperación del caso más adecuado y, el segundo para finalmente concluir en un diagnóstico.

4.1. Análisis de resultados

Es fundamental destacar que los datos fueron recopilados y extraídos del expediente de salud electrónico (EHR). Estos datos también están disponibles en la Pima Indians Diabetes Database (<https://www.openml.org/d/37>), que consta de setecientos sesenta y ocho (768) registros de pacientes femeninas con el idioma en inglés. La base de datos incluye diversas dimensiones relacionadas con las características del paciente, como datos demográficos, pruebas de laboratorio, perfil hematológico, síntomas, pruebas de laboratorio para la función renal, perfil lipídico, análisis de orina, prueba de bilirrubina, historial femenino en caso de ser mujer y, finalmente, los diagnósticos de diabetes (Diabetes Diagnosis) y del hígado (Liver Diagnosis).

Se utilizaron atributos para asignar valores numéricos, ordinales, nominales o categóricos al paciente estudiado como representación de estos casos. Se han eliminado todos los registros con datos faltantes.

Los factores que son más prevalentes entre todas las variables incluyen antecedentes, edad, frecuencia de embarazos (ebrz), nivel de glucosa, grosor del pliegue

cutáneo en la región del tríceps, índice de masa corporal e IMC, niveles de insulina y presión pulmonar. En la Tabla 8 se presenta un resumen de las características de los pacientes.

Tabla 1. Características principales y diagnóstico de los pacientes en estudio

Nº	Caract. categóricas	Coded Values	Original value
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	+++
10	Diabetes Diagnosis	0	Normal
		1	Prediabetic
		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 2.

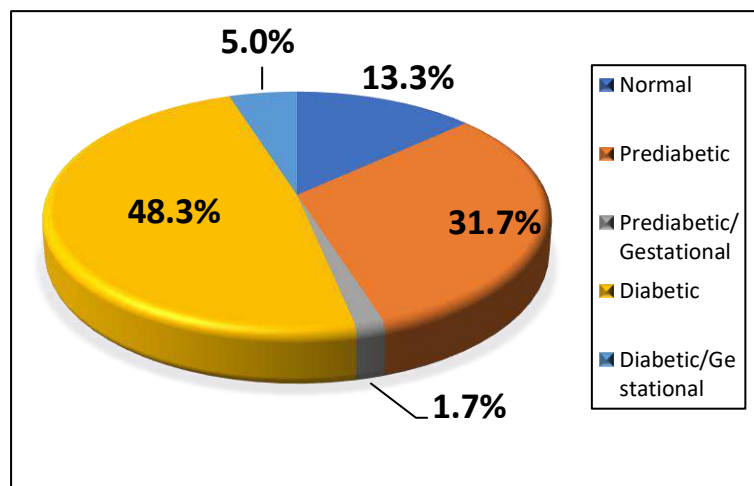
Tabla 2. Segmento de 9 casos con sus respectivas características

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

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 1. Distribución de las pacientes por diagnóstico de la diabetes



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

Tabla 3. Características generales de los datos en estudio

Caract.	Tipo de valor	Valores	Medidas en representación gráfica
Diagnóstico	Nominal	2 únicos valores	<p>Bar chart showing the distribution of diagnostic values. The x-axis has two categories: 'tested_negative' and 'tested_positive'. The 'tested_negative' bar is blue and has a value of 500. The 'tested_positive' bar is black and has a value of 268.</p>
Número de embarazos	Numérico	17 únicos valores	<p>Box plot for 'Número de embarazos'. The x-axis ranges from -5 to 20. The median is approximately 4. The interquartile range (IQR) is from approximately 1 to 7. Whiskers extend from 0 to 17.</p>
Glucosa	Numérico	136 únicos valores	<p>Box plot for 'Glucosa'. The x-axis ranges from -50 to 250. The median is approximately 110. The IQR is from approximately 80 to 150. Whiskers extend from 0 to 200.</p>
Presión diastólica	Numérico	47 únicos valores	<p>Box plot for 'Presión diastólica'. The x-axis ranges from -25 to 150. The median is approximately 65. The IQR is from approximately 50 to 85. Whiskers extend from 0 to 125.</p>
Grosor de la piel en el tríceps	Numérico	51 únicos valores	<p>Box plot for 'Grosor de la piel en el tríceps'. The x-axis ranges from -25 to 125. The median is approximately 25. The IQR is from approximately 10 to 40. Whiskers extend from 0 to 100.</p>
Insulina	Numérico	186 únicos valores	<p>Box plot for 'Insulina'. The x-axis ranges from -250 to 1000. The median is approximately 100. The IQR is from approximately 50 to 200. Whiskers extend from 0 to 800.</p>
Índice de masa corporal	Numérico	048 únicos valores	<p>Box plot for 'Índice de masa corporal'. The x-axis ranges from -20 to 80. The median is approximately 30. The IQR is from approximately 20 to 40. Whiskers extend from 0 to 65.</p>
Antecedentes.	Numérico	517 únicos valores	<p>Box plot for 'Antecedentes.'. The x-axis ranges from -0.5 to 2.5. The median is approximately 0.5. The IQR is from approximately 0.2 to 1.0. Whiskers extend from 0 to 2.0.</p>
Edad	Numérico	50 únicos valores	<p>Box plot for 'Edad'. The x-axis ranges from 0 to 100. The median is approximately 35. The IQR is from approximately 20 to 45. Whiskers extend from 20 to 85.</p>

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

Importación de los datos y creación de la clase paciente

Utilizando el software Protégé, los datos se importaron en formato CSV. desde la página anterior con una coma árabe como separador y se desarrolló la clase 'Paciente' junto con sus características. Las dos figuras siguientes ilustran los registros importados que son instancias en el software Protégé:

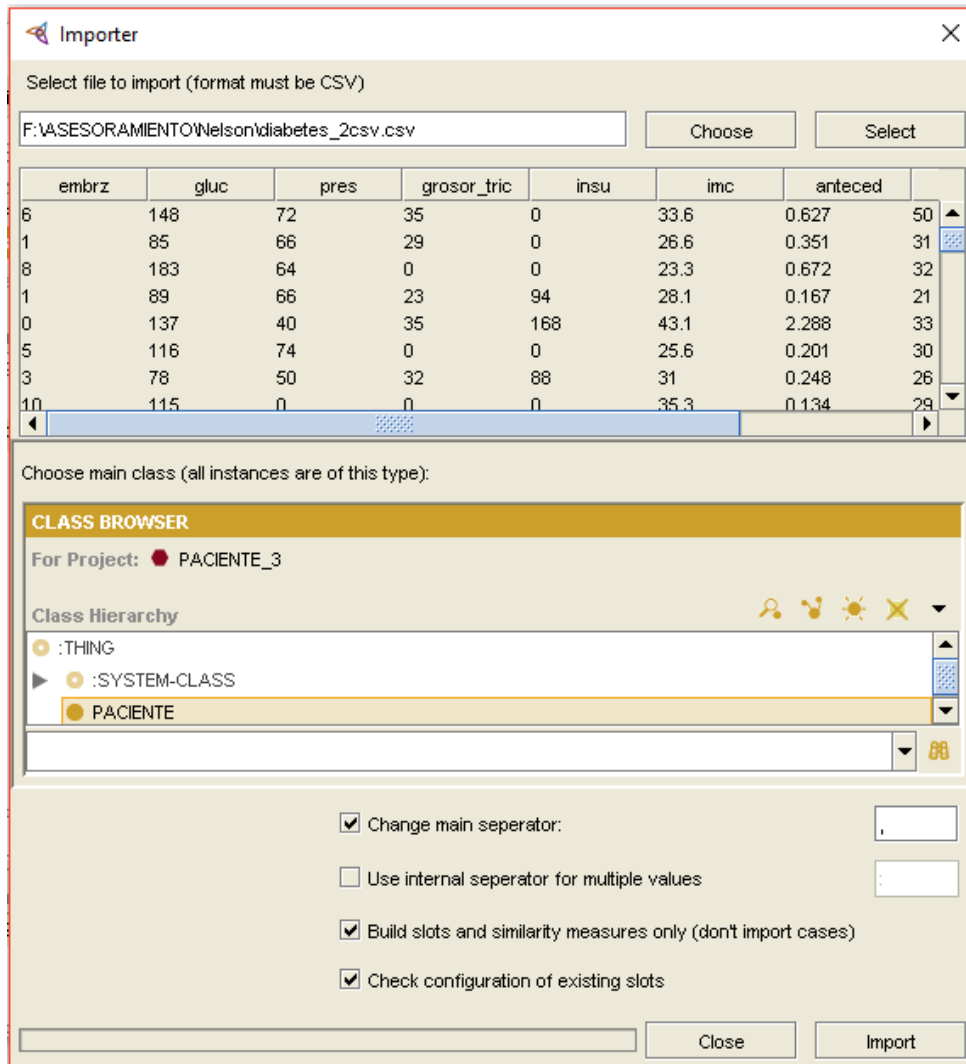


Figura 2. Importación de los datos almacenados en fichero .csv a partir de Protégé

Fuente: Elaboración propia partiendo del Software Protégé

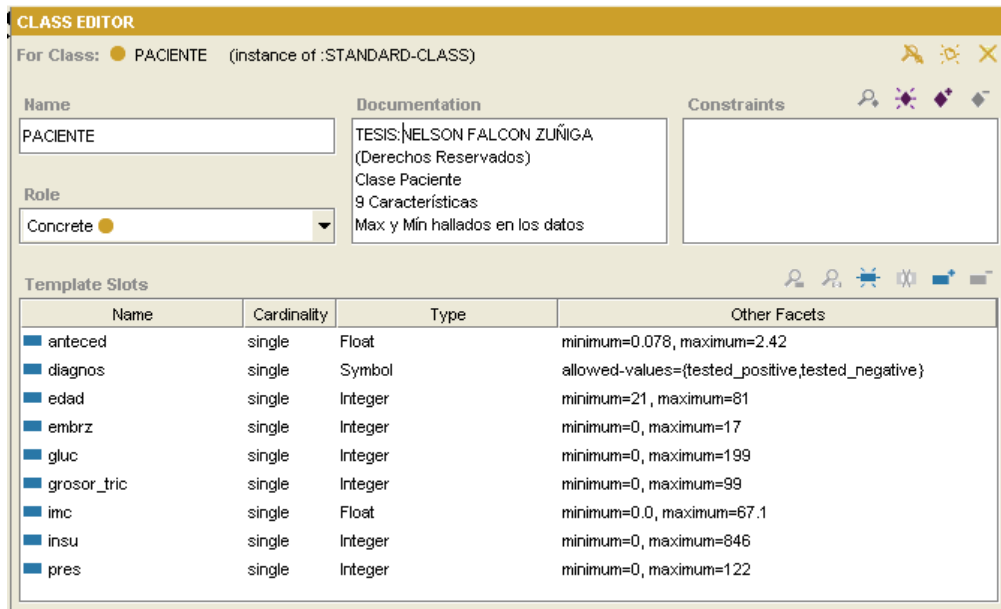


Figura 3. Clase paciente y sus respectivos slots (características)

Fuente: Elaboración propia partiendo del Software Protégé

El tipo de valor para la característica de diagnóstico es Símbolo y tiene dos valores: test_positive y test-negative, y cada ranura se asigna a una clase de paciente.

De manera similar a cómo se configuró para las otras funciones:

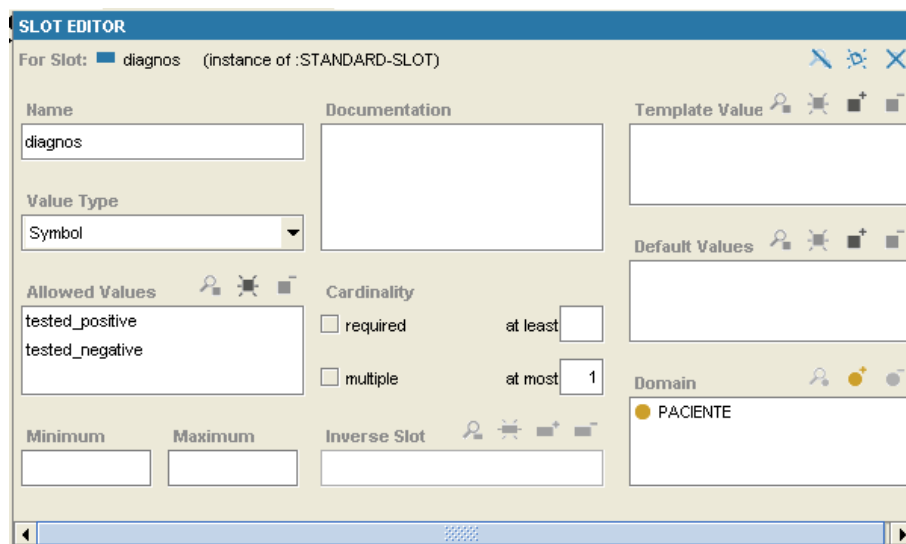


Figura 4. Ejemplo del slot diagnóstico (nombre, tipo de valor y valores permitidos)

Fuente: Elaboración propia partiendo del Software Protégé

El entorno del editor de software Protégé del complemento de software MyCBR se ha ajustado para utilizar medidas de similitud para las ranuras, como se describe a continuación:

- Se ha elegido el comportamiento de la función, que puede presentar simetría o asimetría.
- Se ha definido la función de distancia, teniendo la opción de ser una distancia simple o un cociente.
- La simetría o asimetría se ha configurado asignando valores al polinomio. En los espacios donde se va a ajustar la asimetría, el panel izquierdo calcula la similitud para valores de consulta mayores que los valores en un caso específico, mientras que el panel derecho lo hace para valores de consulta menores a un caso específico.
- En nuestro caso de estudio, todas las medidas de similitud han sido simétricas, como se muestra en el gráfico inferior del editor de medidas de similitud. Por ejemplo, para la glucosa, el gráfico presenta simetría, indicando que el valor más alto lo tiene un paciente con 199.

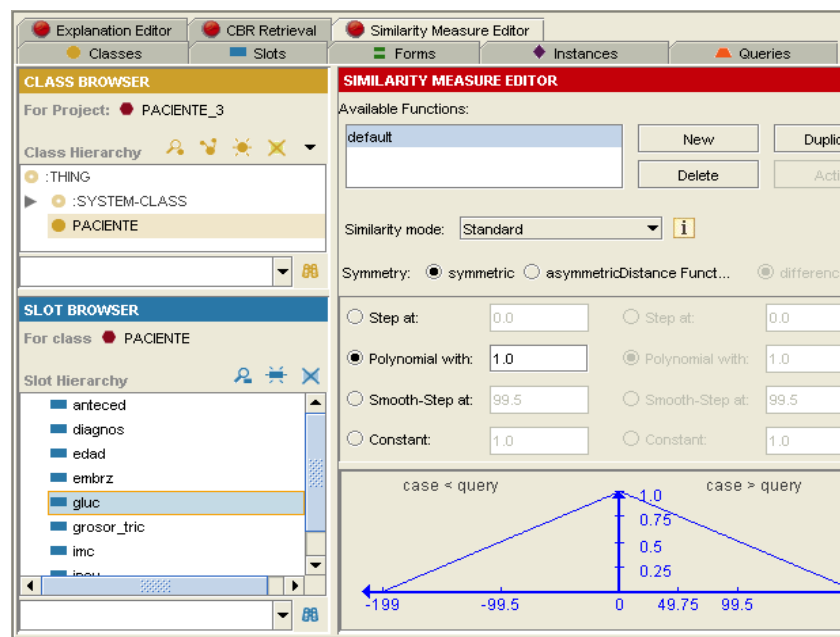


Figura 5. Medida de similitud para el slot glucosa

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 más 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 11.

Query	KB_862466_Class671	KB_862466_Class500	KB_862466_Class96
	1	2	3
	0.94	0.94	0.93
anteced	1.0	0.631	0.255
diagnos	_undefined_	tested_negative	tested_negative
edad	44	49	40
embrz	6	6	6
gluc	140	165	144
grosor_tric	24	26	27
imc	38.0	33.6	33.9
insu	188	168	228
pres	68	74	72

Figura 6. Consulta ingresada y casos similares recuperados

Fuente: Elaboración propia partiendo del Query del Software MyCBR

Validez del módulo de sistema experto

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

Tabla 4. Porcentaje de acierto entre el sistema experto y el médico

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

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. Para medir este grado de acuerdo se desarrolló la prueba de Hipótesis con el siguiente planteamiento:

H_0 : No existe concordancia entre el módulo de sistema experto propuesto y el diagnóstico médico

H_1 : Si existe concordancia entre el módulo de sistema experto propuesto y el diagnóstico médico

En la tabla 12 se dispone de los resultados observados y esperados.

Tabla 5. Concordancia entre el simulador y el diagnóstico médico

SIMULADOR	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

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 (sig) menor al 1%, lo cual indica que hay suficiente evidencia estadística que la concordancia entre módulo y es muy alta, como se muestra en la tabla 11.

Tabla 6. Índice de Kappa de Cohen y significancia de la prueba

	Valor	Significancia
Medida de acuerdo kappa	0,986	0,000

CAPÍTULO V

DISCUSIÓN

5.1 Discusión de resultados

El objetivo de un sistema experto basado en casos, es resolver un problema, buscando el caso más similar de entre los que encuentra en su almacenamiento de memoria, decidiendo aplicar la solución a la resolución del nuevo; sin embargo, en algunas ocasiones hay errores por lo que es muy importante entender bien las medidas de similitud existentes entre el caso nuevo y el de la base de casos.

En ese sentido, para desarrollar la tesis se siguió el modelo de Schank, a pesar que su metodología es de hace varios años, ésta se sigue utilizando, para mostrar de la manera más simple las capacidades de la metodología (Schank, 1986) y que ahora mediante el uso del software MyCBR, se hace bastante amigable el entendimiento del razonamiento basado en casos.

Esta metodología ya ha demostrado eficacia en la solución de problemas reales, por lo que mi estudio para aportar un valor agregado, ha importado la información de la base de datos correspondiente a pacientes del sexo femenino de Pima Indians Diabetes Database (<https://www.openml.org/d/37>), obteniendo resultados interesantes en lo que respecta a la similitud.

Asimismo, se ha llegado a entender que este módulo puede servir como un tutor para estudiantes de medicina o enfermería y que luego en un trabajo futuro se podría implementar en una plataforma Web, concordando con lo que explica Delgado (2009) o el Medical College of Winsconsin mencionados en los antecedentes de este trabajo.

Finalmente, se puede decir que el uso de los softwares Protégé y MyCBR se complementan muy bien y son fáciles de implementar las funciones de similitud obtenidas en MyCBR, esto porque de lejos el uso de la interface gráfica ayuda al investigador en el manejo del editor de similitud (Similarity Measure Editor)

CAPITULO VI

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Con los 768 casos, se hizo la evaluación del simulador, tomando varios casos de esta base de datos obteniéndose en todas las pruebas un promedio de similitud con el caso en consulta de cerca del 99%; el caso que se tomó como ejemplo es sólo uno tomado intencionalmente para corroborar si una paciente no diagnosticada con diabetes tendría alguna similitud, ya que el simulador estaba orientado a encontrar pacientes enfermos. En ese sentido, el módulo del sistema experto ha cumplido con una tarea realmente eficiente, dejando las siguientes conclusiones:

- La simulación del módulo de sistema experto presentado en esta tesis, constituye una herramienta importante para ayudar a encontrar un diagnóstico, pues a partir de los datos ingresados como consulta correspondiente a características especiales de los pacientes, es posible detectar la enfermedad, que para nuestro caso fue el diagnóstico de la diabetes. De esta manera se ha dado respuesta al Problema General, aceptando a su vez la Hipótesis correspondiente.
- En cuanto a la Hipótesis específica 1, se acepta ésta pues, el diagnóstico con el módulo propuesto es casi del 99%; es decir existe una concordancia o acercamiento entre el diagnóstico del simulador del módulo y el dado por el médico y que se encuentra en la base de datos.
- Para la Hipótesis específica 2, se ha recurrido al indicador Kappa de Cohen, para verificar estadísticamente la efectividad del diagnóstico, concluyendo que hay

suficiente evidencia estadística para afirmar que a un nivel de significancia del 1%, el módulo de sistema experto tiene una fuerza de concordancia muy buena pues es del 99% efectivo, siendo el rango del valor de este indicador de (0,81 a 1,00) considerado como muy bueno.

5.2. Recomendaciones

- Para nuestro estudio se ha trabajado con la función de similitud en modo simétrico porque así lo requirieron los datos; sería importante realizar un trabajo con el método asimétrico, para lo cual habría que hacer un ajuste en los valores polinomiales del software MyCBR para cada uno de los slots o características de los casos en estudio.
- Existe otro software para realizar este mismo de estudio denominado JColibri perteneciente al Grupo de Aplicaciones de Inteligencia Artificial (GAIA), por lo que sería importante realizar el mismo estudio con la misma base de datos y observar si hay concordancia con los resultados mostrados en esta tesis.
- Asimismo, existen no pocos algoritmos para medidas de similitud como Naive Bayes, Árboles de decisión, Redes neuronales, etc. que también podrían muy bien ser implementados en estos softwares, pero que requieren de mayor apoyo en la construcción del código y que se sugiere para aquellos que son especialmente programadores.

REFERENCIAS

7.1. Fuentes documentales

Agnar, A. (2004). *Knowledge-intensive case-based reasoning in creek. Advances in Case-Based Reasoning, 7th European Conference, ECCBR 2004*, 1-15.

Contraloría General de la República (2009). CGR en Coordinación con el Ministerio Federal de Cooperación Económica y Desarrollo-Alemania. Una Experiencia de Gestión del Conocimiento. ISBN 978-612-454-12-09.

Andritsos, P., Jurisica, I., & Glasgow, J. (2014). *Case-Based Reasoning for Biomedical Informatics and Medicine*. Auckland New Zealand: Springer, Berlin, Heidelberg.

Bichindaritz, I., & Sullivan M, K. (2002). Generating Practice Cases for Medical Training from a Knowledge-Based Decision-Support System. *ResearchGate*.

Feigenbaum, E. (1977). *The Art of Artificial Intelligence. International Joint Conference on Artificial Intelligence*, Cambridge, MA, 1014-1029.

Informe Defensorial N° 134 (2017). La salud de las comunidades nativas: Un reto para el Estado. *Defensoría del Pueblo*. Biblioteca Nacional del Perú N° 2008-05810. Portal. <http://www.defensoria.gob.pe>

Informes defensoriales (2015). La defensa del derecho de los pueblos indígenas amazónicos a una salud intercultural. *Defensoría del Pueblo*. Informe 169. Jefatura del Programa de Pueblos Indígenas de la Defensoría del Pueblo.

Kahn, C. (1990). Clinical Trial and Evaluation of a Prototype Case-Based System for. Milwaukee, , Wisconsin, USA. Obtenido de Section of Information and Decision Sciences.

M. H. Göker, R. J. Howlett, y J. E. Price, (2006). *Case-based reasoning for diagnosis applications*. The Knowledge Engineering Review, 20(03), 277–281.

Prieto, J. (2005). *Exploración clínica práctica*. Madrid, España: Elsevier.

Richter, M. (2005). The knowledge contained in similarity measures. *FB Informatik*.

7.2. Fuentes bibliográficas

Aamodt, A. (1994), *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. AI Communications. IOS Press, 7(1), 39-59.

Aamodt, A. and Enric. P. (1994). *Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches*. *Artificial Intelligence Communications* 7(1), 39-52

Altman, D. (1994). *Diagnostic tests. 1: Sensitivity and specificity*. Department of Public Health Sciences, St George's Hospital Medical School, London. SW17, ORE.

Carlos, M. (2002). Sistema experto de diagnóstico médico del síndrome de Guillian Barré. Lima, Lima, Perú.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. (D. Albarracín, Ed.) *Psychological Bulletin*, 70(4), 213-220.

Cortés-Reyes, É., Rubio-Romero, J. A., & Gaitán-Duarte, H. (2010). Métodos estadísticos de evaluación de la concordancia y la reproductibilidad de pruebas diagnósticas. 61(3), 247-255.

Cortez, A., Navarro, C. y Pariona, J. (2010). SRBC Aplicados a Sistemas de Líneas de Productos de Software. *Revista de Investigación de Sistemas e Informática*. UNMSM.

- Cruz, L. (2009). *Propuesta de Modelo Híbrido para Análisis de Riesgo Crediticio en Pymes: Caso Crédito MES Caja Municipal de Arequipa*. Universidad Nacional San Agustín de Arequipa.
- Delgado, M. (2009). *Definición del Modelo de Negocio y del Dominio utilizando Razonamiento Basado en Casos*. Centro de Estudios en Ingeniería de Sistemas, Cuba.
- Friedman, H. (2004). *Manual de diagnóstico médico*. Clinical professor of meidicne. Edit. Masson S.A. ISBN 84-458-1150-9. 5ª Edición. Barcelona, España
- Juárez, J. (2007). Una Aproximación Multimodal al Diagnóstico. *Inteligencia Artificial*. *Revista Iberoamericana - REDALYC*, 11(36), 77-80.
- Kolodner, J. (1993). *Case-Based Reasoning*, Morgan Kaufmann, San Mateo. ISBN-13: 978-0898596441. Georgia Institute of Technology. Edic. Schildt. Germany
- Kulikowski, C; (1978). Glaucoma Consultation by Computer, *Comput Biol Med*, 1978, 8, 25-40. Edic. Mc GrawHill
- Maedche, A. & Staab, S. (2001). *Ontology Learning for the semantic web*. IEEE Intelligent Systems, 16(2). Limua Willey. 5º Edic., 72–79
- Nilsson, M., & Sollenborn, M. (2004). Advancements and trends in medical case-based reasoning:. (V. Barr, & Z. Markov, Edits.) *Proceedings of the 17th International Florida Artificial Intelligence Research Society*, 178-183.
- Parisaca, A., Mamani, L., Mercedes, M. y Tapia, S. (2003). *Reducción de Bases de Casos Clusterizadas*. Universidad Nacional de San Agustín, Arequipa.
- Prieto, J. (2005). *Exploración clínica práctica*. Elsevier España, 26 Edic.

- Rodríguez, G., Berdún, L., Soria, Á., Amandi, A., & Campo, M. (2015). Análisis de Métricas de Similitud en Razonamiento. *Simposio Argentino de Inteligencia Artificial*, 57-64.
- Salem, A.-B. M., & Hodhod, R. (2002). Hybrid Expert System Supporting Diagnosis of Heart Diseases. En K. A. Publishers, *Intelligent Information Processing - IFIP 17th World Computer Congress* (pág. 304). Montréal Québec: Springer.
- Stahl, A., Thomas, R., & Berghofer, R. (2008). Rapid Prototyping of CBR Applications with the Open Source Tool myCBR. *German Research Center for Artificial Intelligence (DFKI)*, 1-15.
- Shanon, R. (1976). Systems simulation the art and science. *IEEE Transactions on Systems, Man and Cybernetics*. 6(10), 723-724
- Weis, S. (1974), Glaucoma Consultation by Computer, *Comput Biol Med*, 1978, 8, 25-40. Edic. Mc GrawHill
- Weiss, M. (1974). A system for model-based computer-aided diagnosis and therapy. Rutgers University. The State University of New Jersey, Ph. D.

7.3. Fuentes hemerográficas

- Hawking, S. (2014). Entrevista con BBC. Recuperado de https://www.bbc.com/mundo/ultimas_noticias/2014/12/141202_ultnot_hawking_inteligencia_artificial_riesgo_humanidad_egn

7.4. Fuentes electrónicas

- Alonso, J. & Gutiérrez, M. (2012). Introducción a los Sistemas Basados en Conocimiento. Recuperado de <http://www.cs.us.es/cursos/iic-2012/temas/tema-02-iic03.pdf>

- Alonso, J. & Gutiérrez, M. (2000-2001). Tema 10: Introducción al Aprendizaje Automático. *Razonamiento Automático*, (pág. 18). Sevilla. Obtenido de <https://www.cs.us.es/~jalonso/cursos/ra-00/temas/tema-10.pdf>
- Azán-Basallo, Y., Bravo-García, L., Rosales-Romero, W., Trujillo-Márquez, D., García-Romero, A., & Pimentel-Rivero, A. (2014). *Revista Cubana de Ciencias Informáticas*, 52-68. Recuperado el 30 de Marzo de 2013, de https://www.researchgate.net/publication/304673274_Solucion_basada_en_el_Razonamiento_Basado_en_Casos_para_el_apoyo_a_las_auditorias_informaticas_a_bases_de_datos
- Carrillo, A., Llovet, J. y Rodríguez, I. (2011). Sistema recomendador orientado a la educación terapéutica del paciente diabético. Recuperado de http://eprints.ucm.es/13097/1/Sistema_recomendador_orientado_a_la_educacion_terapeutica_del_paciente_diabetico.pdf
- Cortez Vasquez, A., Navarro Depaz, C., & Pariona Quispe, J. (2010). Sistema de razonamiento basado en casos aplicado a sistemas de línea de productos software. *Revista de Investigación de Sistemas e Informática*, 7(2), 43-48. Obtenido de <https://es.scribd.com/document/378330970/A-Sistemas-de-Lineas-de-Productos-Software>
- Delgado, M. (2002). Definición del modelo del negocio y del dominio utilizando Razonamiento Basado en Casos. *Fundación Dialnet*. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=2097256>

- Díaz-Agudo, B., Gonzalez-Calero, P. A., Recio-García, J., & Sánchez-Ruiz-Granados, A. (2007). Building CBR systems with jCOLIBRI. *ScienceDirect*, 69-75. Obtenido de <https://core.ac.uk/download/pdf/82148971.pdf>
- GAIA (2011). Juan Recio-García, Belén Díaz-Agudo, and Pedro A. González Calero. jcolibri 2 tutorial. Technical report, University Complutense of Madrid, September 16, 2008. Recuperado de <http://gaia.fdi.ucm.es/projects/jcolibri/jcolibri2/docs.html> ISBN 978-84-691-6204-0.
- Maedche, A., & Staab, S. (March de 2001). Ontology Learning for the Semantic Web. *ResearchGate*(16(2)), 72-79. Obtenido de https://www.researchgate.net/publication/220629103_Ontology_Learning_for_the_Semantic_Web
- Pariasca Vargas, A., Mamani Sánchez, L. P., Morveli Espinoza, M. M., & Tapia Tarifa, S. L. (2003). Reducción de Bases de Casos Clusterizadas. En S. P. Computación, *I Congreso Internacional de Científicos Peruanos* (pág. 263). Lima. Obtenido de <https://socios.spc.org.pe/ecuadros/papers/ICICP.pdf>
- Periklis, A., Andritsos, P., A., Jurisica, I. y Glasgow, J. (2011). Case-Based for Biomedical Informatic and Medicine. Machine Learning Methods for the Analysis, Modeling and Knowledge Discovery from Bioinformatics Data. 207-208. Recuperado de <http://www.cs.toronto.edu/~periklis/pubs/cbr14.pdf>
- Pueblo, D. d. (2015). *La Salud de las Comunidades*. Lima: 'Defensoría del Pueblo'. Obtenido de <http://www.defensoria.gob.pe>
- Sigillito, V. (6 de Abril de 2014). *OpenML*. Obtenido de <https://www.openml.org/d/37>

Torres, N. (2015). Conducción de la salud en las zonas rurales del Perú. *Ensayo Los jóvenes y la salud pública: retos y oportunidades*". Universidad Peruana Cayetano Heredia. Facultad en salud pública y administración. <http://www.upch.edu.pe/faspa/index.php/sobre-faspa/2015-04-14-16-46-55/noticias/item/43-ensayo-conduccion-de-la-salud-en-las-zonas-rurales-del-peru>

Weiss, S., Kulikowski, C., & Safir, A. (1978). Glaucoma consultation by computer. *ScienceDirect*, 25-40. Obtenido de <https://www.sciencedirect.com/science/article/pii/0010482578900112>