

Universidad de las Ciencias Informáticas

Facultad 5



Diagnóstico de Enfermedades de Transmisión Sexual mediante técnicas de Inteligencia Artificial.

**Trabajo de diploma para optar por el título de
Ingeniero en Ciencias Informáticas.**

Autores:

Yanet Bañobre Corpas.

Yulie Brossard González.

Tutor:

MSc.Roberto Millet Luaces

Ciudad de la Habana

Junio -2009

El trabajo del pensamiento se parece a la perforación de un pozo: el agua es turbia al principio, mas luego se clarifica.

DATOS DEL CONTACTO:

Tutor: Roberto Millet Luaces

Breve currícul:

- Graduado de Ingeniero Eléctrico en 1986, en la Universidad de Camagüey.
- Profesor Auxiliar
- MSc. en Ciencias Matemáticas.
- Imparte docencia en universidades desde 1987.

Ubicación: Universidad de las Ciencias Informáticas, Cuba.

E-mail: milletp@uci.cu

DEDICATORIA

A mi mamá por haberme dado siempre todo de sí, por su entrega sin límites y ese amor infinito que me profesa.

A mi papá por quererme tanto, por apoyarme en cada una de mis decisiones, por ser siempre mi lalo.

A mi hermano por darme la seguridad de que siempre estará ahí para mí, por ser mi amigo, por su cariño.

A mi tío Finfi por haber sido como un padre para mí durante toda mi vida, por su incondicionalidad infinita para conmigo.

A mi sobrinita Isabella por ser una fuente de inspiración para mí.
Yanet

A mi mamá por su sacrificio, amor incondicional y cariño que me ha dado durante toda mi vida, Gracias mami.

A mi papá por su apoyo y amor incondicional durante toda mi vida.

A mi hermana por sus consejos y cariño.

A mis sobrinos Adachelis y Leonardo por sus cariños y por ser unas de las cosas más importantes que he tenido en mi vida.

A mi novio Yeinier por su amor, comprensión y cariño durante estos 3 años de estudios.

A mi abuelita por su apoyo desde que era una niña.

Yulie

AGRADECIMIENTOS

A mi mamá por hacerme resurgir como ave Fénix de mis propias cenizas cuando me he vuelto polvo, por su confianza, por hacerme sentir fuerte, por su ternura ,por su grandeza, sin ti mami este sueño no hubiese sido posible, gracias por ser mi rayo de luz ,mi guía.

A mi papá por ser ese hombre que me ha expresado su amor no solo con palabras si no con hechos, por no haberme fallado nunca, por haber sabido mostrarme el camino a seguir en la vida ,por habernos puesto antes que a cualquier cosa, por tener un corazón tan grande.

A mi hermano por hacerme participe de su vida y de sus decisiones, por apoyarme y quererme, por sus consejos, por cuidarme, por hacerme saber las cosas tal como son, por todas las alegrías que me ha dado en la vida , estoy muy orgullosa de ti.

A mi segundo padre finfi, por siempre tenerme presente en cada momento, por preocuparse por mí, por toda su dedicación, las palabras no me alcanzan, pero espero un día poder agradecerte por todo.

A mis primsa Rossana y Yalena por haber estado siempre en las buenas y en las malas, por ser como mis hermanas.

A mi tía Carmen y mi tío Angel por preocuparse por mí, por su cariño.

A Kenia por su amistad, a Eliza, Joan, Yulie e Irene por haber estado siempre

A mi tutor Millet por todo su apoyo.

A todas aquellas personas que de una forma u otra han estado presentes en estos cinco años.

Yanet

A mi mamá y papá por ser lo más importante que he tenido en mi vida y por darme la suficiente fuerza de ser quien soy.

A mi tutor Roberto Millet por su apoyo, amistad y dedicación.

A mis tíos Narciso, Flabiano, Ibrain, Analidia, Caridad, Iria, Fefa y Ernis por sus preocupaciones y atenciones que han tenido conmigo.

A mis primos Gerlendys, Ernis y Yanisleidi por su compañía, atención, apoyo y cariño que han tenido conmigo.

A mis compañeras de aula Liliana, Arneris, Elizabeth, Geimi, Yaimaris, Danelys e Irene por su apoyo cuando lo necesitaba.

A mi amiga Kenia por sus consejos, apoyo y amistad durante estos 5 años.

A mi compañera de tesis y gran amiga Yanet por sus consejos siempre que formaba mis lloraderas y por su compañía durante estos 5 años en el mismo apto y cuarto.

A mis compañeros de aula Idelvis, Yordanys y Yusnier por estar siempre cuando los necesitaba.

A mi gran amigo Argelio por su compañía de asiento y apoyo en casi todos mis problemas, eres como un hermano para mí.

A mi amigo Joan por siempre oír mis reclamos y a acompañarme a comer.

A mis otros amigos por siempre estar presentes en los momentos buenos y malos.

Yulie.

RESUMEN

La Inteligencia Artificial (IA) en la actualidad está en constante desarrollo, pudiendo verse sus aplicaciones en diversos sectores. Dentro de la IA se han desarrollado de forma vertiginosa las técnicas que permiten adquirir conocimiento de conjuntos de datos.

El presente trabajo de diploma surge debido al creciente aumento de las Enfermedades de Transmisión Sexual (ETS) en la Universidad de las Ciencias Informáticas (UCI), en especial la Blenorragia y la Clamidia, que aunque no tan frecuente como la primera al presentar los mismos signos, en ocasiones se hace difícil saber de cuál de las dos se trata, representando un gasto para el ya mencionado centro, las pruebas que se realizan para diagnosticarlas. Por esta razón, se decidió crear un sistema en el que se incorpore el conocimiento adquirido a través de una técnica de IA, para el diagnóstico de las ETS Blenorragia y Clamidia.

Durante el desarrollo de este trabajo se abordaron los principales conceptos, técnicas y algoritmos relacionados con la obtención de conocimiento de conjuntos de datos. Primeramente se realizaron consultas a expertos para determinar cuales eran los síntomas más relevantes de ambas enfermedades y la relación existente entre ellos. Después se creó una Base de Hechos para la Blenorragia y otra para la Clamidia, las cuales se procesaron en el software WEKA mediante la técnica Árboles de Decisión (AD), usando para esto el algoritmo ID3. Una vez obtenido el conocimiento se utilizó el mismo para la creación de una aplicación basada en reglas, que permita diagnosticar si una persona está infectada por las ETS Blenorragia o Clamidia.

Palabras Claves: Inteligencia Artificial (IA), Enfermedades de Transmisión Sexual (ETS), Árboles de Decisión (AD), algoritmo ID3.

DEDICATORIA.....	I
AGRADECIMIENTOS.....	II
RESUMEN	IV
INTRODUCCIÓN:	1
1. FUNDAMENTACIÓN TEÓRICA.....	5
INTRODUCCIÓN:.....	5
1.1. INTRODUCCIÓN A LA INTELIGENCIA ARTIFICIAL.	5
1.1.1. <i>La Inteligencia Artificial.</i>	5
1.1.2. <i>Aprendizaje y Conocimiento.</i>	7
1.2. TIPOS DE APRENDIZAJE Y SUS TÉCNICAS.	9
1.2.1. <i>Aprendizaje Inductivo.</i>	9
1.2.1.1. <i>Aprendizaje Supervisado.</i>	9
1.2.1.1.1. <i>Vecino más Cercano.</i>	10
1.2.1.1.2. <i>Árboles de Decisión.</i>	11
1.2.1.1.2.1 <i>Algoritmo ID3.</i>	15
1.2.1.1.2.2. <i>Algoritmo C4.5.</i>	16
1.2.1.2. <i>Aprendizaje No Supervisado.</i>	18
1.2.1.2.1. <i>Agrupamiento.</i>	18
1.2.1.2.2. <i>Análisis de Dependencias.</i>	20
1.2.2. <i>Aprendizaje Deductivo.</i>	21
1.2.2.1. <i>Aprendizaje Basado en Explicación (EBL).</i>	21
1.2.3. <i>Aprendizaje por Refuerzo.</i>	22
1.2.3.1. <i>Elementos del Aprendizaje por Refuerzo.</i>	23
1.2.3.2. <i>Procesos de Decisión de Markov y Algoritmo Q.</i>	23
1.3. LAS ETS BLENORRAGIA Y CLAMIDIA EN LA UCI.	25
1.3.1. <i>Blenorragia.</i>	25
1.3.1.1. <i>Síntomas de la Blenorragia.</i>	26
1.3.2. <i>Clamidia.</i>	27
1.3.2.1. <i>Síntomas de la Clamidia.</i>	27
1.3.3. <i>Diagnóstico.</i>	28
1.4. INTELIGENCIA ARTIFICIAL Y DIAGNÓSTICO MÉDICO.....	29

1.4.1. Aplicaciones para el diagnóstico médico en el mundo.....	29
1.5. ELEMENTOS A UTILIZAR.....	30
1.5.1. Consulta a Expertos.....	31
1.5.1.1. Statgraphics.....	32
1.5.2. WEKA.....	32
1.5.3. Borland C++ Builder.....	33
2. OBTENCIÓN DEL CONOCIMIENTO.....	34
INTRODUCCIÓN:.....	34
2.1. MÉTODO DELPHI.....	34
2.2. ENCUESTA APLICADA A LOS EXPERTOS.....	35
2.2.1. Cuestionario de preguntas.....	35
2.2.2. Resultados obtenidos de la Encuesta.....	37
2.3. TRATAMIENTO ESTADÍSTICO.....	40
2.3.1. Gráficos Obtenidos.....	42
2.4. PROCESO DE DESCUBRIMIENTO DEL CONOCIMIENTO.....	43
2.4.1. Base de Hechos de la Blenorragia.....	44
2.4.2. Base de Hechos de la Clamidia.....	45
2.4.3. Procesamiento de datos.....	46
2.4.3.1. Preparación de los Datos.....	46
2.4.3.2. Análisis del tipo de conocimiento a obtener.....	47
2.3.4.3. Procesamiento de la Base de Hechos de la Blenorragia y resultados obtenidos.....	48
2.3.4.4. Procesamiento de la Base de Hechos de la Clamidia y resultados obtenidos.....	50
2.5. ÁRBOLES DE DECISIÓN.....	51
2.5.1. Árbol de decisión de la Blenorragia.....	51
2.5.2. Árbol de decisión de la Clamidia.....	52
2.6. PRUEBAS.....	53
2.6.1. Pruebas para la Blenorragia y resultados obtenidos.....	53
2.6.2. Pruebas para la Clamidia y resultados obtenidos.....	55
3. PROPUESTA DE SOLUCIÓN.....	57
INTRODUCCIÓN:.....	57
3.1. SISTEMA DE REGLAS.....	57

3.1.1. Reglas para la ETS Blenorragia	57
3.1.2. Reglas para la ETS la Clamidia.....	59
3.2. CARACTERÍSTICAS DE LA APLICACIÓN.....	60
3.3. INTERFACES DE LA APLICACIÓN.	60
3.4. PRUEBAS HECHAS A LA APLICACIÓN.	64
3.5. VENTAJAS DEL USO DE LA APLICACIÓN.	65
CONCLUSIONES.....	66
RECOMENDACIONES.....	67
ANEXOS.....	68
REFERENCIAS BIBLIOGRÁFICAS	72
BIBLIOGRAFÍA.....	73

Introducción:

La inteligencia artificial (IA) es una rama de las ciencias de la computación, que surgió a principios de 1956, permite la creación de programas para computadoras que estén preparados para semejar el comportamiento y la comprensión humana, siendo la creación de hardware y software que se comporten de forma inteligente. En los últimos tiempos el desarrollo de la IA la ha llevado a ser centro de atención debido a sus múltiples aplicaciones en muy diversos campos, pudiéndose obtener con ella conocimientos sobre casi cualquier tema.

En la IA hay un término acuñado como aprendizaje, cuyo objetivo es proveer técnicas que permitan obtener conocimiento o lo que también es conocido como patrones, los cuales una vez obtenidos sirven como base para la creación de sistemas.

Para valorar la importancia de las técnicas que implementa el aprendizaje, sólo se debe ver la amplia gama de aplicaciones que tienen, como por ejemplo: clasificación de secuencias de ADN, reconocimiento del habla y del lenguaje escrito, diagnóstico médico y otras.

Hoy día se ha ido extendiendo por el mundo la utilización de la IA y todas sus ramas en el área de la medicina, donde se resuelven sustanciales problemas gracias al desarrollo de esta ciencia, la relación entre la IA y la medicina es tan estrecha que las aplicaciones de este tipo son conocidas como Artificial Intelligence in Medicine (AIM) [1].

En Cuba también se han insertado técnicas de IA para el diagnóstico y tratamiento de algunas enfermedades como la hipertensión arterial, pero no se han encontrado referencias de que se utilicen en el diagnóstico de Enfermedades de Transmisión Sexual.

En la actualidad en la Universidad de las Ciencias Informáticas, las ETS representan un alto índice en comparación con otras enfermedades. En los últimos tres años las mismas han sufrido un considerable aumento, siendo la Blenorragia una de las más frecuentes. Por otra parte, hay otra ETS: la Clamidia, que presenta síntomas muy similares a la Blenorragia y que también está afectando, aunque en menor grado. En ocasiones se hace difícil decidir en presencia de cuál de las dos se está, debido a la similitud existente

entre ambas. Además muchos pacientes presentan síntomas de estas ETS pero les resulta embarazoso acudir al médico, mientras tanto la enfermedad continúa tomando fortaleza y propagándose a otras personas.

Por la problemática anteriormente expuesta el **problema a resolver** de este trabajo de diploma sería:

¿Cómo contribuir al diagnóstico de las ETS Blenorragia y Clamidia en la UCI?

Con el presente trabajo se pretende dar solución al problema existente con las ETS Blenorragia y Clamidia en la Universidad de las Ciencias Informáticas, creando una aplicación para su diagnóstico, cuya base sea el conocimiento obtenido a partir de alguna técnica de IA. Con dicha aplicación no se pretende sustituir a los médicos especialistas en este tema, sino mas bien ayudarlos a ellos y aquellas personas que presenten síntomas de estas ETS, pero que por diversas razones no acuden al médico.

Objetivo general.

Desarrollar una aplicación que contribuya a la realización del diagnóstico de las ETS Blenorragia y Clamidia.

Objeto de estudio.

El proceso de aplicación de técnicas de Inteligencia Artificial como métodos de obtención de conocimiento.

Campo de acción.

Técnicas de Inteligencia Artificial como métodos de obtención de conocimiento para el diagnóstico de las ETS Blenorragia y Clamidia en la UCI.

Objetivo específicos

1. Revisión de la bibliografía existente.
2. Consulta a expertos (Especialistas en el tema).

3. Estudio de conceptos y algoritmos relacionados con las técnicas de IA referentes a la obtención de conocimiento.
4. Estudio de elementos de estadística (asistente Statgraphics).
5. Estudio y aplicación del software WEKA.
6. Desarrollo de una aplicación para el diagnóstico de las ETS Blenorragia y Clamidia.

Para la realización de esta tesis se utilizarán los métodos científicos que se enuncian a continuación:

- Métodos Teóricos
 1. Método Histórico
 2. Analítico - Sintético
- Métodos Empíricos
 1. La Entrevista.
 2. La Encuesta.

De este modo en este trabajo de diploma se tratarán los contenidos de la siguiente manera: **Capítulo 1: Fundamentación Teórica**, en este capítulo se abordará lo concerniente al fundamento teórico de algunos de los tipos de aprendizaje que existen y las técnicas que utilizan los mismos para obtener conocimiento de conjuntos de datos. También se hará mención a los principales aspectos de las ETS Blenorragia y Clamidia y las herramientas y elementos que se utilizarán a lo largo del desarrollo de este trabajo.

Capítulo 2: Obtención del Conocimiento, en este capítulo se hará referencia a cómo se llevará a cabo el tratamiento de las encuestas aplicadas para determinar la relación existente entre las variables, mostrándose todo el proceso. Además, se expondrán las Bases de Hechos pertenecientes a las ETS Blenorragia y Clamidia, así como la forma en la que se procesarán los datos contenidos en las mismas, para poder obtener los Árboles de Decisión que contendrán el conocimiento. También se hará reseña a las pruebas aplicadas sobre el conocimiento se obtenga, así como los resultados que se logren con las mismas.

Capítulo 3: Propuesta de Solución, en este capítulo se presentarán los conjuntos de reglas pertenecientes a las ETS Blenorragia y Clamidia, las cuales constituirán la base de la aplicación. Se

mostrará la manera en que fue creado el sistema para el diagnóstico de las ETS mencionadas, así como las interfaces que lo conformarán, sus principales características y algunos resultados que se obtengan de la utilización del mismo.

1. Fundamentación Teórica.

Introducción:

La fundamentación teórica de las investigaciones permite conocer las opiniones que pueden tener diversos autores sobre un determinado tema a través de las búsquedas bibliográficas que se realizan. En este capítulo el marco teórico se centrará en los principales conceptos y técnicas mediante las cuales sea posible la obtención de conocimiento, así como en las principales características de las ETS Blenorragia y Clamidia.

1.1. Introducción a la Inteligencia Artificial.

1.1.1. La Inteligencia Artificial.

Cuando se habla de técnicas de obtención de conocimiento en la informática, se está haciendo referencia a técnicas de IA. Por tanto antes de mencionar dichas técnicas y los algoritmos que implementan las mismas, se hace necesario hacer una breve introducción acerca de la IA.

Cada día en muchos lugares del mundo, se realizan grandes esfuerzos para lograr que los ordenadores puedan simular de cierta forma la inteligencia humana, por eso al tratarse de IA, muchos lo asocian a súper computadoras inteligentes, que algún día puedan llegar a sobrepasar el conocimiento y entendimiento de los seres humanos. Sin embargo, aún no se sabe hasta donde podría llegar la inteligencia de las máquinas que el hombre sea capaz de crear. Mas lo que sí se hace evidente son los logros que se han obtenido hasta hoy en esta rama, los cuales han contribuido de forma positiva a la eliminación de obsoletas técnicas de trabajo y al desarrollo de las ciencias en general, dando nuevas perspectivas sobre cómo podría ser el mundo del mañana.

Puede resultar difícil definir el término **Inteligencia Artificial** de manera que sea aceptado por diversos expertos de esta área, entre las definiciones más difundidas están las siguientes:

- Hayes “La Inteligencia Artificial es la implementación de razonamientos inteligentes mediante técnicas propias de la computación.”[2]

- Marvin Minsky “La Inteligencia Artificial es una ciencia dedicada a la construcción de máquinas que implementan tareas propias de humanos dotados de Inteligencia.”[2].

Por tanto se puede decir que la IA es el conjunto de técnicas que tratan de conseguir que los ordenadores simulen la inteligencia humana, que les permitan “pensar” y resolver complejas tareas con la misma capacidad y análisis con que lo haría un ser racional.

La idea de algo parecido a la IA existe hace millones de años, desde que el hombre tomó conciencia de que tenía la habilidad de pensar y de que un ser inteligente era capaz de crear otro. Pero realmente la IA comenzó a desarrollarse a partir de 1943 cuando Warren McCulloch y Walter Pitts presentaron su modelo de neuronas artificiales que se considera el primer trabajo en este campo, aunque por aquel entonces aún no se había acuñado el término.

Un tiempo después a principios de los años 50 Alan Turing la introdujo en la comunidad científica con su artículo "Maquinaria Computacional e Inteligencia" , donde dió a entender su convicción de que las computadoras podrían imitar la inteligencia humana. También creó el llamado Test de Turing que permite considerar o decidir si una máquina puede pensar como un ser humano. Por todos los aportes que realizó Turing es conocido como el padre de la IA.

Un tiempo después en 1956 se desarrolla la Lógica Teórica, que es considerada por muchos el primer programa de IA. Un año después Jhon McCarthy junto a otras autoridades en el tema, sentarían las bases de la IA en una conferencia en los Estados Unidos.

Los años venideros traerían grandes avances, como el desarrollo del programa Eliza, considerado el primer sistema dentro del campo del Procesamiento del Lenguaje Natural, también el famoso Sistema Experto DENDRAL que estaba orientado a la determinación de la estructura química de un compuesto, aunque el más influyente sería el Mycin que era capaz de diagnosticar trastornos en la sangre. También se realizó la primera implementación del lenguaje PROLOG (Programming in Logical), por el francés Alain Colmerauer, este lenguaje para la programación en IA estaría basado en la Lógica del Cálculo de

Predicados y en el Principio de Resolución, y fue considerado por los japoneses como el lenguaje de las máquinas del futuro, dicho lenguaje hoy día todavía es ampliamente utilizado.

En la actualidad han ido surgiendo nuevas tendencias en las más diversas áreas y sus aplicaciones se han ido incrementando. Por día se pueden encontrar sistemas inteligentes lo mismo para jugar ajedrez, que para realizar complejas investigaciones médicas, en todos los casos lo que se utilizan son técnicas de IA, de ellas las más difundidas son:

- Lingüística computacional.
- Procesamiento del lenguaje Natural
- Minería de datos.
- Redes Neuronales Artificiales.
- Redes Bayesianas.
- Sistemas Expertos.
- Lógica Difusa.
- Ingeniería del conocimiento.
- Aprendizaje Automático.

No constituyendo las únicas que existen, pero sí las más usadas.

1.1.2. Aprendizaje y Conocimiento.

Al hablar de obtención de conocimiento, se está hablando de diversas técnicas de inducción del mismo, surgiendo entonces la pregunta ¿qué es el conocimiento?

Conocimiento: Es el entendimiento sobre un determinado tema, en el cual una persona pueda expresar su opinión.

En el caso de la IA este concepto podría variarse un poco, siendo el conocimiento conjunto de patrones que se obtienen de grupos de datos.

También existe otro concepto que está estrechamente relacionado con el anterior: **Aprendizaje**, que no es más que adquirir conocimiento sobre alguna cosa o tema.

Hoy día se implementan sistemas basados en conocimientos adquiridos sobre disímiles materias, utilizándose para ello técnicas de aprendizaje de las correspondientes a la IA. Con el transcurso del tiempo algunos sistemas han intentado eliminar toda necesidad de expertos en el análisis de datos, pero siempre se debe tratar de establecer un marco de cooperación entre el experto humano y la computadora. De cualquier forma la intuición humana no puede ser reemplazada en su totalidad, ya que son los diseñadores del sistema los que especifican la forma de representación de los datos, así como los métodos de manipulación de los mismos.

Los seres humanos realizan el aprendizaje de forma mecánica y casi sin darse cuenta al ir aprendiendo del entorno que lo rodea. Sin embargo, a las máquinas sí hay que enseñarles, es decir, hay que incluirle en su dominio aquello que se quiere que sepan.

Existen distintos tipos de aprendizaje a través de los cuales se puede obtener conocimiento:

- **Aprendizaje Inductivo:** Se pretende crear modelos de conceptos a partir de la generalización de conjuntos de ejemplos [3].
- **Aprendizaje Analítico o Deductivo:** Se aplica la deducción para obtener descripciones generales a partir de un ejemplo de concepto y su explicación [3].
- **Aprendizaje por Refuerzo:** El algoritmo aprende observando el mundo que lo rodea [3].

Antes de hacer referencia a cada una de estos tipos de aprendizaje hay dos conceptos que son necesarios introducir: **instancia**, que es la pieza de información mínima a partir de la cual el clasificador inducirá un modelo y **atributos** que son los valores predeterminados que caracterizan a las instancias.

Para el desarrollo de casi todos los algoritmos relacionados con el aprendizaje se utilizan las Bases de Hechos que son creadas a partir de conjunto de ejemplos, los cuales tienen que ver con el tema que se quiera desarrollar, por tanto una **Base de Hechos** es una estructura de datos que contiene información sobre una situación concreta [4].

1.2. Tipos de Aprendizaje y sus técnicas.

Existen diversos tipos de Aprendizaje y sus técnicas van desde las más complejas hasta las más sencillas. En este epígrafe se tratarán tres tipos de ellos, así como las técnicas que los mismos implementan, para más tarde decidir cual de ellas será la que se utilice para lograr obtener conocimiento a partir de un conjunto de datos.

1.2.1. Aprendizaje Inductivo.

El **Aprendizaje Inductivo** es uno de los más estudiados. Su principal objetivo es lograr encontrar características comunes en un grupo de instancias, esto está dentro del razonamiento inductivo y este se realiza a partir de un grupo limitado de ejemplos. El razonamiento inductivo permite obtener conclusiones generales a partir de información específica, esta información que se obtiene es conocimiento nuevo.

Desde un punto de vista se puede decir que el resultado de los métodos inductivos no es válido, ya que son de naturaleza heurística, se supone que con un número limitado de ejemplos se obtienen todas las características que se quieren aprender, pero en muchas ocasiones no ocurre así.

1.2.1.1. Aprendizaje Supervisado.

El aprendizaje **supervisado** es un tipo de Aprendizaje Inductivo y es aquel en el que por cada ejemplo se indica el concepto al que pertenece, es realizado por contraste entre conceptos, se distinguen cuales características son distintas entre los ejemplos de un concepto y el resto. El resultado es un concepto o conceptos que sean capaces de describir los ejemplos que se utilizaron. Es decir, dado un conjunto de ejemplos en forma de pares que serían de la forma (entrada, salida) se infieren las reglas que mejor modelen esta relación entre las entradas y las salidas.

Por lo tanto el aprendizaje supervisado se plantearía de la siguiente forma:

Dado un conjunto de datos con la estructura (x, z) , donde X es la entrada y Z la salida:

Datos $\{(X_1, Z_1); (X_2, Z_2); \dots; (X_n, Z_n)\}$

se debe encontrar una función o regla que se aproxime lo más posible a la salida Z a partir de las entradas X .

1.2.1.1.1. Vecino más Cercano.

El Vecino más cercano es una técnica de Aprendizaje Supervisado que se encuentra entre los métodos Lazy, que son aquellos que no construyen ningún modelo. En él se almacenan todas las instancias y después se clasifican las nuevas en función de las más cercanas a ella. La distancia entre dos instancias se calcula a partir del valor de sus atributos, generalmente se utiliza la distancia Euclídea para hallar la distancia entre la instancia a clasificar y los demás.

Distancia Euclídea: Es la distancia entre dos puntos ordinarios [5]. En este caso sería entre instancias que tienen N atributos; P ($p_1; p_2 \dots p_n$) y Q ($q_1; q_2 \dots q_n$) por lo que la fórmula quedaría de la siguiente forma:

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

Existen dos formas de aplicar este método, ellas son:

➤ **Método Básico :solo un vecino (1-NN)**

En este caso se buscaría la instancia almacenada más parecida a la que se quiere clasificar y se devolvería la salida de esa instancia como respuesta.

➤ **Mas de un vecino (K-NN)**

Se buscarían las K instancias almacenadas que más se parezcan a la instancia que se desea clasificar y como salida se daría el valor más frecuente, entre las K instancias más cercanas a ella.

A continuación se explica como queda de forma general el procedimiento para la clasificación de una nueva instancia:

1. Se mide la distancia entre la instancia a clasificar y todas las demás instancias de entrenamiento almacenadas en el espacio de los atributos con la distancia Euclídea.
2. Se eligen las K instancias más próximas.
3. Se asigna como clase la mayoritaria entre las K instancias.

Cuando se utiliza este método se tiene una capacidad de representación muy elevada, pero no se tiene ninguna legibilidad porque no se crea ningún modelo. El tiempo de cómputo On-Line es lento ya que se hace necesario calcular distancia a todos los ejemplos de entrenamiento, sin embargo el tiempo de cómputo Off-Line es muy rápido ya que sólo se necesita tiempo para guardar los ejemplos de entrenamiento. Ante instancias de entrenamiento ruidosas es más robusto en el caso que sea K-NN, mientras más aumenta K, más robusto se vuelve el método, si fuese 1-NN no se tendría robustez.

1.2.1.1.2. Árboles de Decisión.

Los Árboles de Decisión es otra técnica del Aprendizaje Supervisado, como su nombre lo indica son útiles cuando es necesario la toma de alguna decisión entre un conjunto de posibles opciones, con ellos se pretende modelar los datos de ejemplos mediante un árbol.

Los AD están formados por:

1. **Nodos:** Nombres o identificadores de los atributos.
2. **Ramas:** Posibles valores del atributo asociado al nodo.
3. **Hojas:** Conjunto ya clasificado de ejemplos y etiquetado con el nombre de una clase.
4. **Atributos:** Factores que influyen la clasificación o decisión.
5. **Clase:** Posibles valores de solución.
6. **Ejemplos:** Conjunto de combinaciones de atributos dados.

En los AD, los nodos intermedios son los atributos de entrada de los ejemplos presentados, las ramas representan valores de dichos atributos y los nodos finales son los valores de la clase.

Desde el punto de vista del aprendizaje, se trata de encontrar un árbol que clasifique correctamente un conjunto de ejemplos. De esta forma en un AD se pueden distinguir los siguientes elementos:

1. Espacio de hipótesis: Todos los posibles AD.
2. Método: Escalada (hill-climbing), empezando por el árbol vacío.
3. Heurística que guía la búsqueda: Ganancia de información.

El espacio de hipótesis de los AD es completo. Esto quiere decir que cualquier función finita que tome valores discretos puede ser representada como un AD. Además, en cada elección se considera una única hipótesis, a diferencia de otros algoritmos del aprendizaje, en los que se consideran simultáneamente todas las hipótesis consistentes con los ejemplos.

El algoritmo considera en cada paso gran cantidad de ejemplos, mientras que otros consideran los ejemplos uno a uno. Esta es la razón de la robustez frente al ruido que presentan los AD. Si los ejemplos fueran considerados uno a uno y alguno de ellos fuese erróneo, el efecto sobre la corrección de los resultados sería mayor que si el dato erróneo se considera simultáneamente a otro conjunto de datos en los que no hay ruido.

Por otra parte, el algoritmo no permite retroceso; una vez se ha elegido un atributo para un nodo, los ejemplos de este se clasifican según ese atributo, y más adelante no hay posibilidad de volver a este punto y considerar un atributo distinto, por lo que podemos obtener óptimos locales en lugar de óptimos globales.

Ventajas de los AD

1. Resume los ejemplos de partida, permitiendo la clasificación de nuevos casos, siempre y cuando no existan modificaciones sustanciales en las condiciones bajo las cuales se generaron los ejemplos que sirvieron para su construcción.
2. Facilita la interpretación de la decisión adoptada.
3. Proporciona un alto grado de comprensión del conocimiento utilizado en la toma de decisiones.
4. Explica el comportamiento respecto a una determinada tarea de decisión.
5. El resultado que se obtiene de la modelación del árbol puede expresarse como un conjunto de reglas de tipo **If- Else**.

Utilidad de los AD

Los AD se adaptan más fácilmente a ciertos tipos de problemas, básicamente son apropiados para los siguientes:

1. Los ejemplos pueden ser descritos como pares valor-atributo.
2. La función objetivo toma valores discretos.

3. Ante una posible existencia de ruido en el conjunto de entrenamiento.
4. Los valores de algunos atributos en los ejemplos del conjunto de entrenamiento pueden ser desconocidos.

Valores perdidos

Un **valor perdido (missing value)** en inglés, es el valor de un atributo que se desconoce para un ejemplo, en casos reales este tipo de valores suelen aparecer y presentan problemas en dos circunstancias principalmente:

- Al crear subárboles: La solución que se propone es ignorar los ejemplos y luego elegir el valor más frecuente, después se le asigna una fracción de ejemplo a cada rama, proporcional a la distribución de valores, se incluye la instancia en cada rama y se crea una rama con valor desconocido
- Al predecir: La solución que se propone es si existe una rama para valores desconocidos, usarla y tomar el valor más frecuente. Explorar todas las ramas y dar una probabilidad perteneciente a cada clase y por último dar la clase mayoritaria.

Tratamiento del Ruido

Ruido: Son los ejemplos etiquetados con una clase incorrecta [6].

El ruido casi siempre está presente cuando se utilizan conjuntos de datos reales y da como resultado árboles de gran tamaño, los nodos extras aparecen para demostrar los ejemplos que están mal etiquetados. A este efecto se le denomina sobre-especialización y hace que el árbol obtenga malos resultados prediciendo nuevos ejemplos.

Para eliminar o minimizar el ruido se utiliza la poda (prunning) del AD, lo que se hace es ir podando subárboles erróneos, de tal forma que se aumente el grado de generalización. Con la poda se enlazan pedazos de árboles parecidos para reducir el número de nodos y simplificar las reglas generalizándolas.

Poda (Prunning) de los AD

Existen dos tipos de poda la Post-Poda (Post-Prunning) que se realiza después de crear el árbol y la Pre-Poda (Pre-Prunning) que se realiza durante la creación del árbol. Pero la que más se usa comúnmente es

la Post –Poda, esto se debe a que en la Pre –Poda, es muy difícil saber cuando se debe detener el crecimiento del árbol.

En la Post-Poda se siguen los siguientes pasos:

1. Construir el árbol de decisión a partir del conjunto de entrenamiento.
2. Convertir el árbol en un conjunto equivalente de reglas.
3. Podar (generalizar) cada regla, eliminando aquellas que conlleven a un aumento estimado del rendimiento, sobre el conjunto de test o incluso sobre el conjunto de entrenamiento.
4. Ordenar el árbol según lo que se eliminó.

Sobreajuste

A medida que se añaden niveles al AD, las hipótesis se refinan tanto que describen muy bien los ejemplos utilizados en el aprendizaje, pero el error de clasificación puede aumentar al evaluar los ejemplos. Es decir, clasifica muy bien los datos de entrenamiento pero luego no sabe generalizar al conjunto de prueba. Esto se debe a que aprende hasta el ruido del conjunto de entrenamiento, adaptándose a las regularidades del mismo.

Este efecto por supuesto es indeseado y pueden existir varias causas posibles para que esto ocurra, las principales son:

- Exceso de ruido (nodos adicionales).
- Un conjunto de entrenamiento demasiado pequeño como para ser una muestra representativa de la verdadera función objetivo.

Las estrategias para evitar el sobreajuste en los datos se dividen en dos:

- Estrategias que frenan el crecimiento del árbol antes de que llegue a clasificar perfectamente los ejemplos del conjunto de entrenamiento (Pre-Poda).
- Estrategias que permiten que el árbol crezca completamente, y después realizan una poda (Post-Poda).

De alguna forma con estas técnicas se trata de compensar la falta de backtracking en el proceso de inducción.

1.2.1.1.2.1 Algoritmo ID3.

El ID3 es un algoritmo recursivo, que pertenece a la familia de los TDIDT (Top-Down Induction of Decision Trees) y que fue propuesto por Quinlan en 1986. Se utiliza para crear un AD que explique cada instancia de la secuencia de entrada de la manera más compacta posible. Genera el árbol a partir de ejemplos de partida e intentado encontrar la forma más sencilla para separar mejor los ejemplos.

Este algoritmo para elegir qué atributos y en qué orden aparecerán en el árbol utiliza una función de evaluación llamada **ganancia de información** que no es más que la reducción de la entropía en el conjunto de ejemplos [7] y nos sirve para la elección del mejor atributo en cada nodo, la **entropía** a su vez es la medida de incertidumbre que hay en un sistema [8], es decir ante una determinada situación la probabilidad de que ocurra cada uno de los posibles resultados. La ganancia de información puede verse como una heurística para elegir cuál es el mejor atributo.

La función de entropía más utilizada es la binaria, pero aquí se tratará de forma general, por tanto quedaría de la siguiente forma:

$E_t(S) = \sum_i (-t_i/T \log_2 (t_i/T))$ donde t_i representa la clase de los objetos y T el número total de ejemplos.

La ganancia de información después de conocida la entropía se plantearía de esta manera:

$G(S,A) = E_t(S) - \sum_i (t_i/T E_t(S_t))$ para todos los valores diferentes del atributo A , donde t_i es el número de ejemplos de S etiquetados con ese valor, T es el número total de ejemplos y $E_t(S_t)$ la entropía de los ejemplos etiquetados con ese valor.

El algoritmo ID3 examina todos los atributos y escoge el de máxima ganancia, forma la ramificación y usa el mismo proceso recursivamente para formar sub-árboles a partir de los nodos que se han generados.

El AD con ID3 se recorre desde la raíz y tanto en ella como en cada uno de los demás nodos se decide cuál rama tomar basándose en el valor de algún atributo del ejemplar que se esté clasificando, hasta llegar a un nodo terminal (hoja), que corresponde a la clase en que queda clasificado el ejemplar.

Como se han definido las funciones heurísticas en ID3, solo se pueden tratar atributos cualitativos, pero hay ocasiones en que los atributos tienen valores cuantitativos, por lo tanto la solución que se halló para esto fue particionar el intervalo de valores en dos o más intervalos, esto en la práctica sería discretizar el atributo en dos o más valores.

El principal inconveniente de este método es que favorece atributos que tengan muchos valores, los cuales necesariamente no son los más útiles, además puede traer conflictos en la base de conocimiento porque diferentes soluciones se pueden alcanzar con variables que tengan los mismos valores asociados. Con este algoritmo en algunas ocasiones se generan grandes árboles y esto trae como consecuencia que las reglas pueden no ser todo lo eficientes que se desea.

1.2.1.1.2.2. Algoritmo C4.5.

Este algoritmo es una mejora ID3, construye un AD a partir de los datos mediante particiones realizadas recursivamente.

El algoritmo permite trabajar con valores continuos para los atributos, separando los posibles resultados en 2 ramas $A_i \leq N$ y $A_i > N$. Además los árboles son menos frondosos, ya que cada hoja cubre una distribución de clases no una clase en particular. Genera el árbol de decisión inicial a partir de un conjunto de datos de entrenamiento. Se basa en la utilización del criterio de **proporción de ganancia (gain ratio)** que no es más que una medida basada en información que considera diferentes números (y diferentes probabilidades) de los resultados de las pruebas [9]. De esta manera se consigue evitar que las variables con mayor número de posibles valores salgan beneficiadas en la selección. Antes de cada partición de datos, el algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información o en la mayor proporción de ganancia de información.

Las estructuras que se usan en C4.5 son las mismas que en ID3, ya que el primero está basado en el segundo, por tanto la estructura principal de ambos métodos es la misma. Este algoritmo va construyendo el árbol de decisión mediante el algoritmo Divide y Vencerás y evalúa la información en cada caso utilizando los criterios de Entropía, Ganancia o proporción de ganancia, según sea el caso.

Este algoritmo puede trabajar con distintos tipos de atributos ellos son:

1. **Atributos de valores continuos:** Incorpora atributos con valores continuos, simplemente dividiendo estos valores en intervalos discretos, de forma que el atributo tendrá siempre valores comprendidos en uno de estos intervalos.
2. **Atributos con valores perdidos:** En ciertos casos existen atributos de los cuales se conoce su valor para algunos ejemplos, y para otros no. Por ejemplo una base de datos médica en la que no a todos los pacientes se les ha practicado un análisis de sangre. En estos casos lo más común es estimar el valor basándose en otros ejemplos de los que sí conocemos el valor. Normalmente se fija la atención en los demás ejemplos de ese mismo nodo. Así, al ejemplo de valor desconocido se le da el valor que más aparezca en los demás ejemplos.
3. **Atributos con pesos diferentes:** En algunas tareas de aprendizaje los atributos pueden tener costes asociados. Por ejemplo, en una aplicación médica para diagnosticar enfermedades se pueden tener atributos como temperatura, resultado de la biopsia, pulso, análisis de sangre, etc., que varían significativamente en su coste, monetario y relativo a molestias para el paciente.

El C4.5 tiene algunas mejoras respecto a su antecesor el ID3, ellas son:

- Evita sobreajuste de los datos.
- Determina qué tan profundo debe crecer el árbol de decisión.
- Reduce errores en la poda.
- Condiciona la Post-Poda.
- Maneja atributos continuos.
- Escoge un rango de medida apropiado.
- Manejo de datos de entrenamiento con valores faltantes.
- Maneja atributos con diferentes valores.
- Mejora la eficiencia computacional.

El C4.5 utiliza la variante de poda Post-Poda. Una vez generado el árbol completo plantea qué es lo que se debe podar, para mejorar el rendimiento y obtener un árbol más pequeño. Además el C4.5 convierte el árbol en un conjunto de reglas antes de podarlo, existen tres razones fundamentales para hacer esto:

- Ayuda a distinguir entre los diferentes contextos en los que se usa un nodo de decisión, debido a que cada camino de la raíz a una hoja se traduce en una regla distinta.
- Deja de existir la distinción entre nodos que están cerca de la raíz y los que están lejos. Así no hay problemas para reorganizar el árbol si se poda un nodo intermedio.
- Mejora la legibilidad. Las reglas suelen ser más fáciles de entender.

Una vez podados, las hojas de los AD generados por el C4.5 tendrán dos números asociados: N y E. N es la cantidad de casos de entrenamiento cubiertos por la hoja, y E es la cantidad de errores predichos si un conjunto de N nuevos casos fuera clasificados por el árbol. La suma de los errores predichos en las hojas, dividido entre el número de casos de entrenamiento, es un estimador inmediato del error de un árbol podado sobre nuevos casos.

1.2.1.2. Aprendizaje No Supervisado.

El aprendizaje **no supervisado** es aquel en el que no existe una clasificación para los ejemplos, no se sabe a qué concepto pertenecen. Lo que se hace es buscar la manera más adecuada de particionar los ejemplos, casi siempre buscando su estructura y en la mayoría de los casos se obtiene una partición en grupo. El aprendizaje es guiado por la similaridad/disimilaridad de los ejemplos, los ejemplos que más se parezcan estarán juntos en un grupo y separados de los que menos se parezcan que formarán otro grupo distinto.

Por tanto en el Aprendizaje No Supervisado lo que se hace es a partir de ejemplos elaborar una hipótesis que generalice los ejemplos que se observan.

1.2.1.2.1. Agrupamiento.

Con las técnicas de agrupamiento lo que se pretende es que el sistema detecte agrupamientos o estructuras intrínsecas en el conjunto de datos, logrando identificar grupos homogéneos entre los

ejemplos de entrenamiento. Lo que se hace es definir cada ejemplo según sus propiedades y crear grupos a los que se asocien ejemplos muy parecidos entre sí y distintos a los demás.

Mapas de Kohonen

Los **Mapas de Kohonen** es un algoritmo que presenta un aprendizaje No Supervisado Competitivo y generalmente están basados en redes neuronales [10]. La red en este algoritmo debe descubrir rasgos comunes, regularidades o correlaciones en los datos de entrada, e incorporarlos a su estructura interna de conexiones. Se dice, por tanto, que las neuronas deben auto-organizarse en función de los datos procedentes del exterior.

El resultado gráfico de este método es un mapa en 2D de categorías, en cada categoría estarán agrupados los ejemplos de entrenamiento de acuerdo a sus similitudes y cada categoría ocupará un espacio proporcional a su cantidad de componentes.

Un Mapa de Kohonen esta dividido en dos capas de neuronas: la capa de entrada que esta formada por N neuronas de acuerdo a la cantidad de entradas y la encargada de transmitir a la capa de salida la información procedente del exterior. La capa de salida es la encargada de procesar la información y formar los grupos de acuerdo a los rasgos.

Entre las neuronas de la capa de salida, puede decirse que existen conexiones laterales de excitación e inhibición implícitas, pues aunque no estén conectadas, cada una de estas neuronas va a tener cierta influencia sobre sus vecinas. Esto se consigue a través de un proceso de competición entre las neuronas y de la aplicación de una función denominada de vecindad, que produce la topología o estructura del mapa. Las topologías más frecuentes son la rectangular y la hexagonal. Las neuronas adyacentes pertenecen a una vecindad N_j de la neurona j . La topología y el número de neuronas permanecen fijos desde el principio.

Las conexiones entre las dos capas que forman la red son siempre hacia delante, es decir, la información se propaga desde la capa de entrada hacia la capa de salida. Cada neurona de entrada i está conectada con cada una de las neuronas de salida j mediante un peso w_{ji} . De esta forma, las neuronas de salida

tienen asociado un vector de pesos W_j llamado vector de referencia (codebook), debido a que constituye el vector prototipo de la categoría representada por la neurona de salida j .

El proceso de aprendizaje de este método se realiza de la siguiente forma:

1. Un ejemplo o vector X es seleccionado del conjunto de datos de entrenamiento, se halla la similitud entre él y los vectores de referencia usando la distancia Euclídea.
2. Una vez encontrado el vector más próximo, se actualizan el resto de los vectores de referencia y el vector más próximo y sus vecinos se mueven hacia el vector X .
3. Los pasos 1 y 2 se van repitiendo hasta que el entrenamiento termina, el número de veces que se realizará debe ser fijado con antelación, para poder calcular la tasa de convergencia de la función de vecindad y el grado de aprendizaje que se obtuvo. Cuando se termine el entrenamiento, el mapa debe de haberse organizado en sentido topológico, n vectores topológicamente próximos se aplican en n neuronas adyacentes o incluso en la misma neurona.

1.2.1.2.2. Análisis de Dependencias.

El objetivo de esta técnica de Aprendizaje No Supervisado es tratar de encontrar relaciones significativas entre los datos, para poder lograr, teniendo el valor de una instancia, predecir el valor de otra que tenga relación con la primera.

El Análisis de Dependencias se basa en el concepto de **transacción** que es la secuencia de análisis o sucesos relacionados entre sí. Si interviene el tiempo las secuencias se denomina series temporales.

Se utilizan dos técnicas para desarrollar este método, ellas son:

- **Análisis de Asociaciones:** Detecta elementos en una transacción, que implican la presencia de otros elementos, sin importar el orden de los sucesos en la transacción.

Aquí existen reglas llamadas de asociación en las que A, B, C, \dots implican Z , entonces se habla de **soporte** que es el por ciento de transacciones totales donde aparecen A, B, C y de **confianza** que es el por ciento de transacciones donde aparecen A, B, C y también Z .

- **Análisis de Secuencia:** Se hace lo mismo que en el Análisis de Asociaciones sólo que aquí sí es relevante el orden de los sucesos en las transacciones.

1.2.2. Aprendizaje Deductivo.

El **Aprendizaje Deductivo** es aquel que aprende de los conceptos conocidos, este es un proceso iterativo donde se añade lo aprendido a lo que ya se sabía. En este tipo de aprendizaje las premisas respaldan a las conclusiones, por lo que se dice que preserva la verdad y se obtienen resultados validos y sólidos. Este tipo de aprendizaje sigue el esquema del razonamiento deductivo, a partir de un conocimiento general y un conjunto de primicias se llega a las conclusiones deseadas, logrando elaborar nuevas hipótesis más eficientes.

1.2.2.1. Aprendizaje Basado en Explicación (EBL).

El **Aprendizaje Basado en Explicación** es un tipo de Aprendizaje Deductivo, posiblemente el más utilizado y conocido, en él se usan reglas o hechos que ya son conocidos para generar otros nuevos. Para aplicar métodos de EBL es necesario que se tenga conocimiento previo del dominio, esto tiene como principal dificultad que en la práctica exige un grado alto de conocimiento previo, pero como ventaja tiene que con un ejemplo basta para que se analice y se pueda generalizar correctamente.

En este tipo de aprendizaje se obtiene una nueva hipótesis que se deduce de la que ya existía, por lo que el conocimiento que se incorpora no es realmente nuevo, pero sí asegura que ese conocimiento que se incorporó sea correcto.

La forma en la que se aplica el EBL es bastante simple, se trata de un proceso de dos fases, la primera es la Explicación, teniendo un ejemplo de entrenamiento, se usa la **Teoría de Dominio** actual que no es más que el conocimiento previo sobre el problema que se trata y se construye una explicación que demuestre que el ejemplo de entrenamiento es verdadero del **Concepto Objetivo**, que es la hipótesis que ya se tiene y se desea mejorar. Después se comienza el segundo paso que es la Generalización, donde se consideran todos los caminos de demostración del ejemplo para identificar nuevo conocimiento útil, en este caso lo que se hace básicamente es un proceso de regresión, en el cual se sustituyen constantes por variables de tal forma que se conserve la validez lógica. Al terminarse este proceso de aplicación del EBL

se pretende obtener una nueva hipótesis deducida del conocimiento previo, que permita demostrar ese ejemplo y otros que se parezcan en menos pasos y por tanto de forma más eficiente.

Para saber si la nueva hipótesis que se dedujo es útil o no se utiliza el **Operador de Operatividad** que indica la validez de la nueva hipótesis, teniendo en cuenta cuan eficiente es, porque si es menos eficiente que la anterior no tendría sentido que se convirtiera en la nueva hipótesis.

1.2.3. Aprendizaje por Refuerzo.

El **Aprendizaje por Refuerzo** consiste en aprender a decidir ante una situación determinada, qué acción es la adecuada para lograr un objetivo. Este tipo de aprendizaje se usa cuando se está tratando el tema de los agentes inteligentes. Un agente tiene que realizar una tarea y para ello consta de un conjunto de acciones para resolverla, el agente debe escoger la acción o el conjunto de acciones que crea que mejor le posibilitan resolver dicha tarea.

Este aprendizaje está orientado a sistemas que sean capaces de interactuar con el entorno y a su vez, podrá dar una cuantificación numérica del éxito o fracaso del agente. Los sistemas de este tipo aprenden mediante prueba y error realizando una tarea determinada repetidamente, para adquirir experiencia y mejorar, su comportamiento.

Existen dos tipos de Aprendizaje por Refuerzo: el **pasivo** que es aquel en el que la política del agente ya está fijada y la tarea es aprender la utilidad de los estados; y el **activo** en el que el agente debe aprender además de la utilidad de los estados también qué hacer, el tomará información del entorno y la utilizará para determinar la siguiente acción a realizar.

En este aprendizaje se producen señales de esfuerzo que pueden ser inmediata o retardada, esto condiciona dos tipos de refuerzo:

- **Refuerzo inmediato:** En el que se obtiene una crítica para cada acción efectuada, después de su realización, y la información que aporta es local a cada acción tomada. Debe aprenderse una acción a realizar en cada situación para obtener un refuerzo positivo después de su ejecución

- **Refuerzo retardado:** Se tiene una estimación global del comportamiento para cada acción y se presenta cuando no se completan la secuencia de acciones para resolver el problema.

1.2.3.1. Elementos del Aprendizaje por Refuerzo.

- **Estados que describen el entorno:** posibles situaciones para el entorno, pueden ser discretos o continuos.
- **Acciones:** posibles acciones que el agente o el sistema puede realizar en un momento determinado.
- **Función de refuerzo:** indica al agente el resultado obtenido con una acción sobre un estado.
- **Objetivo:** Encontrar una vinculación entre los estados y las acciones que maximice la ganancia que se obtiene a través de los refuerzos, es decir, aprender a escoger la mejor acción a realizar cuando el sistema se encuentre en un estado dado.

1.2.3.2. Procesos de Decisión de Markov y Algoritmo Q.

El caso más sencillo de Aprendizaje por Refuerzo es cuando se tiene un conjunto finito de estados, estos estados son observables y el resultado de las acciones es determinista. En este caso se puede modelar el problema como un Proceso de decisión de Markov.

Un **Proceso de Decisión de Markov** se modela de la siguiente forma:

1. Un agente puede percibir un conjunto S estados posibles.
2. Existe un conjunto A de acciones que se pueden realizar.
3. En cada intervalo (discreto) de tiempo del agente observa el estado (S_t) y elige una acción a realizar (A_t).
4. El entorno responde con una recompensa de acuerdo al estado y la acción ($r_t=r(S_t, A_t)$), produciendo el estado sucesor ($S_{t+1}=f(S_t, A_t)$).
5. r y f son funciones de entorno que no tiene que ser conocidas por el agente

Cuando se utiliza este proceso el objetivo del agente será derivar una función Q : donde S implique A , es decir que le permita elegir la acción A_t dependiendo del estado S_t . Lo que se intenta lograr es que las acciones que se escojan produzcan la mayor recompensa acumulativa en el tiempo.

En este caso lo que se necesitaría es un algoritmo que pueda generar la función P_i que es la que determinará la mejor acción en cada estado, que es el objetivo de este aprendizaje. Para eso se utiliza el algoritmo Q.

Algoritmo Q

Es difícil aprender la función P_i directamente porque lo único observable es la secuencia de refuerzos que recibe el agente en el tiempo r ($S_i ; A_i$) para $i = 0; 1; 2$. Por tanto se crea una nueva función denominada Q.

Se define la función $Q(S,A)$ como la que calcula el máximo refuerzo acumulativo desde el estado S tomando A como primera acción, representa la máxima recompensa que se podría obtener. Conocido $Q(S,A)$ el sistema sabe qué acción tiene que tomar en cada estado, en todos los casos sería la que brinde un mayor máximo para $Q(S,A)$.

Forma de inferir la función Q:

1. Inicialmente, $Q(s,a)$ toma valores aleatorios (a cada estado y a cada acción posible en ese estado se les asigna un valor Q aleatorio).
2. Partiendo del estado inicial, se realizan acciones aleatoriamente hasta que se cumple una de estas condiciones:
 - El sistema alcanza el objetivo buscado.
 - El número de acciones realizadas alcanza un límite máximo.
3. Se actualizan los valores de Q en función del resultado obtenido:
 - Si se ha alcanzado el objetivo se recompensan las acciones.
 - Si se ha alcanzado el número máximo de acciones sin alcanzar el objetivo final, no hay recompensa.
4. Las etapas 2 y 3 se repiten hasta que Q converge (alcanza cierta precisión) o hasta que se alcanza un valor máximo de repeticiones.

Está demostrado que de esta forma es posible obtener Q con total precisión si se recorren todos los estados y todas las acciones un número infinito de veces. En la práctica no es necesario alcanzar una

precisión total, pero sí se requiere un número elevado de repeticiones, porque para poder lograr el aprendizaje, el agente tiene que entrenarse la mayor cantidad de veces posibles en el entorno para que la función Q pueda converger.

1.3. Las ETS Blenorragia y Clamidia en la UCI.

En la actualidad en el mundo hay dos ETS que se han vuelto muy comunes: la Blenorragia y la Clamidia. La UCI no queda exenta de estas infecciones y en los últimos tres cursos la Blenorragia ha sido una de las ETS más frecuentes en el centro, sumando cifras que podrían no parecer alarmantes en otros lugares, pero que aquí constituyen un verdadero peligro. La Clamidia por su parte, aunque no tan abundante como la Blenorragia, también constituye motivo de preocupación debido a que los casos también han experimentado un aumento considerable.

En el centro médico de la UCI, un alto grado de pacientes que se presentan a la consulta de Infecciones de Trasmisión Sexual con secreción genital está infectado por una de estas dos enfermedades, las cuales no difieren mucho, debido a que ambas presentan los mismos síntomas, tanto en hombres como en mujeres.

Las **ETS** son aquellas que se transmiten de una persona a otra por contacto sexual, ya sea vaginal, anal u oral. También son conocidas como ITS porque hay algunas que infectan a las personas sin causar ninguna enfermedad, solo síntomas desagradables.

Diagnóstico: Proceso de definir por qué alguien está enfermo. Los médicos reúnen todo tipo de datos, para poder decir qué es lo que está afectando a un determinado individuo.

1.3.1. Blenorragia.

La Blenorragia es una ETS curable. Es causada por un tipo de bacteria llamada *Neisseria gonorrhoeae* que puede crecer y multiplicarse fácilmente en lugares húmedos y tibios del aparato reproductivo, incluidos el cuello uterino, las trompas de Falopio y el útero en la mujer y la uretra en el hombre. También se puede desarrollar en la boca, el ano, los ojos y la garganta, pero la mayor incidencia está en las áreas genitales.

Esta enfermedad infecciosa es una de las más comunes y la puede adquirir cualquier persona que tenga una actividad sexual activa. Es más frecuente encontrarla en las grandes ciudades, áreas citadinas y poblaciones con niveles generales de educación bajos.

Los principales factores de riesgos son los que incrementan la posibilidad de contraer esta enfermedad y vienen dados, por tener múltiples parejas, tener relaciones sexuales con personas que tienen antecedentes de ETS, no usar preservativos a la hora del acto sexual o por tener historial de ETS.

La principal vía de transmisión de esta infección es la sexual, ya sean orales, anales o vaginales y no hay necesidad de que haya eyaculación para que se contagie. De madre a hijo también se puede contagiar, durante el parto, pasando el germen a los ojos del bebe, pero estos casos casi nunca se presentan.

1.3.1.1. Síntomas de la Blenorragia.

Los primeros síntomas de esta enfermedad generalmente son leves y aparecen entre los dos y cinco primeros días después de haberla contraído. En otras ocasiones no se presentan síntomas hasta que la infección no está muy avanzada e incluso pueden no presentarse, esto trae como consecuencia que la persona no sepa que está infectada y no busque tratamiento médico, aumentando el riesgo de complicaciones y la posibilidad de pasársela a otra persona.

El 80% de los hombres presentan síntomas los cuales comprenden:

1. Aumento de la frecuencia o urgencia urinaria.
2. Secreción del pene.
3. Dolor o ardor al orinar.
4. Testículos sensibles o inflamados.
5. Mal Olor.
6. Fiebre.

En el caso de las mujeres, casi nunca presentan síntomas a corto plazo y muchas veces suelen confundirse con otras infecciones. Pero cuando los presentan los más comunes son los siguientes:

1. Secreción vaginal.
2. Dolor o ardor al orinar.
3. Aumento de la frecuencia o urgencia urinaria.
4. Dolor intenso en la parte baja del abdomen.
5. Relaciones sexuales dolorosas.
6. Mal Olor.
7. Sangrado menstrual irregular.

1.3.2. Clamidia.

La clamidia es una ETS causada por la bacteria *Chlamydia trachomatis*. Esta enfermedad es una de las más comunes y afecta las áreas genitales en los hombres y las mujeres, pudiendo también contagiar el ano y la boca, pero la mayor incidencia está en los genitales.

La infección por clamidia puede ser transmitida durante las relaciones sexuales vaginales, orales o anales. También se puede transmitir de madre a hijo durante el parto vaginal. Toda persona sexualmente activa puede contraer la infección por clamidia. Entre mayor número de parejas sexuales tenga la persona, mayor es el riesgo de infección. Las adolescentes y las mujeres jóvenes que son sexualmente activas están expuestas a un mayor riesgo de infección porque el cuello uterino no se ha formado completamente y es más susceptible a infecciones.

1.3.2.1. Síntomas de la Clamidia

La clamidia es conocida como la enfermedad silenciosa porque en un gran número de los casos no presenta síntomas, cuando lo hace aparecen de una a tres semanas después de haber contraído la infección.

En los hombres uno de cada cuatro no presenta síntomas, pero cuando lo hacen los más frecuentes pueden ser:

1. Secreción del pene que puede ser blanca o acuosa.
2. Dolor o ardor al orinar.
3. Aumento de la frecuencia urinaria.
4. Mal Olor.
5. Dolor en los testículos.

En el caso de las mujeres el setenta por ciento de ellas no presentan síntomas, en los casos que lo hacen pueden ser:

1. Incremento de las secreciones vaginales.
2. Dolor o ardor al orinar.
3. Dolor en la parte baja del abdomen.
4. Dolor al tener relaciones sexuales.
5. Sangrado irregular.
6. Mal Olor.
7. Aumento de la frecuencia urinaria.

Los síntomas de la Clamidia tanto en hombres como en mujeres son muy similares a los de la Blenorragia, por esta razón es necesario aplicar pruebas para saber cuál de las dos infecciones es la que se padece.

1.3.3. Diagnóstico.

El proceso de realizar un diagnóstico de las ETS Blenorragia y Clamidia es exactamente igual para ambas. Lo primero que se hace cuando llega un paciente con posible infección de ETS, es preguntarle por los síntomas que presenta.

En la UCI se utiliza el exudado genital para detectar estas ETS. Si el paciente está infectado por Blenorragia entonces el exudado dará como resultado la presencia de la bacteria diplococo, de lo contrario dará negativo descartándose esta infección y automáticamente diagnosticándose Clamidia.

Sin embargo hay ocasiones en las que no hay reactivo para hacer el exudado. En el presente caso el paciente describe los síntomas y es examinado por el médico, siendo la secreción el síntoma que más información representa y en función de ella y de los otros síntomas se dicta un diagnóstico y se le informa al paciente si está infectado por Blenorragia o Clamidia.

Tanto la Blenorragia como la Clamidia se repiten sólo si la persona vuelve a ser contagiada por otra y pueden traer complicaciones de no ser tratadas a tiempo, tales como: inflamación pélvica que puede provocar abscesos internos y dolor pélvico crónico prolongado, infertilidad en la mujer y en el hombre epidemitis y con ella la esterilidad.

1.4. Inteligencia Artificial y diagnóstico médico.

Desde el surgimiento de la IA, una de las ramas que se ha beneficiado con su creación ha sido la medicina, a través de los diagnósticos médicos y la predicción de enfermedades.

Actualmente se crean sistemas que puedan clasificar el estado de un paciente en una o varias enfermedades. La generalidad es tener una Base de Hechos en la que se tengan todos los datos de ejemplo. La entrada será los datos de un paciente y la salida coincidirá con la enfermedad que tiene (si es que padece alguna), de acuerdo a los síntomas que se ingresaron. El sistema puede inferir qué enfermedad tiene el paciente extrayendo un conjunto de patrones generales que se le puedan aplicar a este caso en particular, a partir de los ejemplos de entrenamiento.

1.4.1. Aplicaciones para el diagnóstico médico en el mundo.

En el mundo existen sofisticados sistemas para el diagnóstico médico, creados a partir de algoritmos de aprendizaje que generen el conocimiento necesario para la realización del mismo, algunos son:

1. **MIDAS (Medical Diagnosis Assistant):** Es un sistema experto avanzado capaz de proporcionar un diagnóstico médico a partir de los informes radiológicos/patológicos del paciente, basado en extracción de información a partir de historias clínicas de pacientes diagnosticados anteriormente.

Específicamente, el sistema automatiza la asignación de códigos ICD-9-CM (International Classification of Diseases) a informes médicos, logrando unos buenos resultados de precisión.

2. **Sistema artificial útil para la detección de tumores:** Este es un sistema basado en Redes Neuronales, que entre otras utilidades tiene la de diagnosticar tumores mamarios. Se combina la visión por ordenador con el de las Redes Neuronales. Ha demostrado que se pueden crear sistemas en el tratamiento y diagnóstico de imágenes.
3. **Predicción de la supervivencia en cirróticos tratados con TIPS (Desvío portosistémico intrahepático transyugular):** Sistema útil para el diagnóstico basado en histología del hígado. Consta de 77 clases. En este sistema se incluye conocimiento acerca de la enfermedad tomando en cuenta datos anteriores y es posible predecir la supervivencia de enfermos con TIPS. La correcta selección de las variables que se utilizaron permitió reducir los gastos, agilizar el proceso de diagnóstico de la enfermedad y construir modelos con mayor poder clasificatorio.

1.5. Elementos a utilizar.

En el desarrollo de este trabajo se contará con dos Bases de Hechos, los datos contenidos en ellas serán del tipo síntomas y diagnóstico obtenido, tomando en cuenta esos síntomas, por tanto se utilizará una técnica de aprendizaje inductivo del tipo inductivo supervisado, para la obtención del conocimiento, siendo los datos de la forma (entrada, salida).

La técnica a utilizar será **Árboles de Decisión**, debido a que se considera la más adecuada, porque como su nombre lo indica, con ella se pueden tomar decisiones sobre un determinado tema. En este caso específico lo que se quiere es dar un diagnóstico de si un paciente está infectado o no por *Blenorragia* o *Clamidia*. En este trabajo el uso de los AD permitirá derivar el conocimiento que los mismos contienen como un conjunto de reglas, las cuales serán el conocimiento que se usará para dar el diagnóstico de las ETS *Blenorragia* y *Clamidia*.

En los casos como el presente, los AD son la técnica perfecta y una de las más utilizadas al crear sistemas que estén relacionados con la toma de decisiones acerca de enfermedades, esto se debe a que generan reglas que resumen todos los ejemplos de partida sobre las cuales se pueden montar los sistemas.

El algoritmo que se usará es el ID3, con el que se creará el AD para cada una de las ETS. Fue escogido porque genera árboles robustos y más pequeños que otros algoritmos. Además en este caso se puede utilizar sin ninguna dificultad debido a que los conjuntos de entrenamiento que se utilizarán están compuestos por datos nominales (que es el tipo de datos que soporta el ID3) y no sería necesaria su conversión, por lo que el algoritmo se puede aplicar sin riesgos de la información que genere no sea del todo correcta o no se resuman de forma eficiente los ejemplos de partida. Otro aspecto que podría afectar la calidad del conocimiento que se obtenga con el ID3 sería la existencia de valores perdidos, pero en este caso tampoco constituye un problema, debido a que en los datos que se emplearán como conjuntos de entrenamiento no existen este tipo de valores.

1.5.1. Consulta a Expertos.

Se aplicará la Consulta a Expertos, que es una técnica que consiste en la ejecución de una encuesta a distintos expertos para llegar a un consenso y validar si las variables que se van a utilizar son propicias para generar el conocimiento que se desea obtener. Con ella se logra tener una idea y extensión del tema que se está tratando y contribuye a dar un mayor rigor científico a las investigaciones.

Experto: Un experto es una persona que tiene habilidades y conocimientos para sacar conclusiones acerca de un determinado tema, basándose en experiencias pasadas y centrándose en los aspectos más importantes de un problema [11]. Por esta razón se dice que el éxito de un experto se debe a la experiencia que haya podido adquirir.

Lo que se hace es realizar una encuesta a expertos de la especialidad que se está tratando, para llegar a un acuerdo final sobre la veracidad de lo que se quiere validar. Se compararan las opiniones de unos con las de otros para validar la información técnica y lograr un mejor aprovechamiento de la información. Posteriormente se realiza un análisis estadístico para saber si se aceptó el criterio de los expertos.

En casos como el presente en el que se trata de diagnóstico de enfermedades la Consulta a Expertos se considera de vital importancia para guiar a los investigadores, esto viene dado porque los expertos en las enfermedades que se tratan son las personas más calificadas para orientar acerca de cuáles son los principales aspectos de las mismas, así como las causas que las originan.

En este trabajo se utilizará la técnica de Consulta a Expertos para determinar los síntomas de las ETS Blenorragia y Clamidia y para asegurar que el conocimiento que se obtenga este respaldado por la opinión de los especialistas, lográndose con ello una mayor solidez en la investigación.

Dentro de la Consulta a Expertos se aplicará el Método Delphi como parte del análisis multicriterial y se le dará tratamiento estadístico a los resultados obtenidos, ya que esto será lo que permita verificar que los resultados pueden ser utilizados para el desarrollo del presente trabajo. Se empleará como apoyo para el análisis estadístico el software Statgraphics, tratándose de reunir la mayor cantidad posible de elementos que puedan aportar datos significativos.

1.5.1.1. Statgraphics.

Este software es un asistente estadístico, el cual se utilizará para dar tratamiento a la Consulta a Expertos. Se escogió porque proporciona potentes procedimientos estadísticos, y permite representar los resultados obtenidos en tablas y gráficos, para ilustrarlos mejor y se logre así un mejor entendimiento. Además, admite la generación de informes y el procesamiento de datos. Destaca especialmente por sus capacidades para el desarrollo de experimentos, previsiones y simulaciones en función del comportamiento de los valores.

1.5.2. WEKA.

WEKA, acrónimo de Waikato Environment for Knowledge Analysis, es un entorno para experimentación de análisis de datos que permite aplicar, analizar y evaluar las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario.

La licencia de Weka es GPL, lo que significa que este programa es de libre distribución y difusión. Además, debido a que Weka está programado en Java, es independiente de la arquitectura, y funciona en cualquier plataforma sobre la que haya una máquina virtual Java disponible.

Está constituido por una serie de paquetes de código abierto con diferentes, técnicas de preprocesado, clasificación, agrupamiento, asociación, y visualización, así como facilidades para su aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados. Estos paquetes pueden ser utilizados en cualquier proyecto de análisis de datos, esto se debe a que contienen numerosas técnicas de análisis, que permiten realizar extracción de conocimiento de los conjuntos de datos iniciales. Al ser Weka Open Source permite incluir a los algoritmos implementados en estos paquetes, otros nuevos creados por los usuarios.

Con este software se realizará el procesamiento de los datos contenidos en las Bases de Hechos, lo que dará lugar a la creación de los AD, de los cuales se obtendrán un conjunto de reglas a partir de las que estará soportado el sistema. Se escogió este software porque ya trae algoritmos implementados, con los que se puede generar el AD, incluido el ID3, que es el que se empleará en el desarrollo de este trabajo. Además es multiplataforma (Java) y tiene cuatro interfaces de usuario distintas, permitiendo múltiples funcionalidades. Actualmente es uno de los Software más usado para la Minería de Datos en la comunidad Open Source.

1.5.3. Borland C++ Builder.

El software Borland C++ Builder, se utilizará para la creación de la aplicación. Se escogió debido a que con él se pueden programar desde los sistemas más sencillos, hasta los más complejos. La creación de interfaces gráficas para las aplicaciones en este software es sumamente sencillo, pudiéndose usarse conjuntamente con otros programas para crear potentes sistemas. Además el lenguaje C++ es el más conocido por los desarrolladores de este trabajo y este es precisamente el que soporta el Borland C++ Builder. Su uso representa una ventaja en cuanto a la economización del tiempo destinado para el desarrollo de este trabajo, pues no sería necesario aprender un nuevo lenguaje para poder realizarlo.

2. Obtención del conocimiento.

Introducción:

En este capítulo se hará referencia a cómo se llevó a cabo el tratamiento de las encuestas aplicadas para determinar la relación existente entre las variables, mostrándose todo el proceso y las conclusiones a las que se arribaron al culminarlo. Además se mostrarán las Bases de Hechos pertenecientes a las ETS Blenorragia y Clamidia, así como la forma en la que se procesarán esos datos, para poder obtener los AD que contendrán el conocimiento necesario para lograr el diagnóstico de estas ETS. También se realizarán pruebas a los AD obtenidos, y se presentarán los resultados que se lograron con las mismas.

2.1. Método Delphi.

El Método Delphi es una técnica que se utiliza para estructurar un proceso de comunicación en grupo. Se basa en la interrogación a un grupo de expertos sobre un tema específico ayudándose de cuestionarios sucesivos, con el fin de poner de manifiesto la convergencia de opiniones hasta que eventualmente se llegue a un consenso.

Durante la aplicación de este método ningún experto conoce la identidad de los otros expertos que componen el grupo de debate, por lo que se dice que se aplica en el anonimato, se hace de esta manera para evitar los efectos líderes. Actualmente una de las vías más comunes que se utiliza es el correo electrónico.

En el proceso de llevar a cabo el método se hace con una iteración y retroalimentación controlada, esto se realiza presentado varias veces el mismo cuestionario, lo que permite ir disminuyendo las convergencias de criterios, ya que se consigue que los expertos vayan conociendo los diferentes puntos de vistas y puedan ir modificando su opinión. La información que se presenta a los expertos no es sólo aquella que representa la opinión de la mayoría sino todas las opiniones y se indica el grado de acuerdo que se ha obtenido.

La calidad de los resultados que se obtienen depende en gran medida de la forma en que fueron elaborados los cuestionarios y de la correcta elección de los expertos, siempre se debe tratar que los

mismos guarden la mayor relación posible con el tema que se trata, porque con esto se garantiza una mayor seguridad y confianza en los resultados.

La aplicación de este método tiene varias etapas sucesivas:

- **1ra Etapa:** Se realiza la formulación del problema a tratar.
- **2da Etapa:** Selección de los expertos en función del problema que se va a tratar y atendiendo a los criterios de experiencia, responsabilidad y disponibilidad. También se elige el tamaño de la muestra de acuerdo los recursos, medios y tiempo disponible.
- **3ra Etapa:** Elaboración y lanzamiento de los cuestionarios. Los cuestionarios se deben elaborar de manera tal que faciliten la respuesta por parte de los encuestados.
- **4ta Etapa:** Explotación de los resultados. El objetivo de los cuestionarios sucesivos es disminuir la dispersión y llegar a un consenso.

La aplicación del método Delphi trae consigo que se tengan que realizar varias rondas de preguntas, pero en la práctica casi nunca sucede así, porque con solamente una ronda o dos es suficiente para que se llegue a un consenso.

Entre las principales ventajas del método Delphi están que permite obtener información desde distintos puntos de vista sobre temas muy amplios o específicos. Además por ser anónimo permite la participación de un gran número de personas sin que se arme el caos eliminando o aminorando los efectos negativos de las reuniones “cara a cara.”

2.2. Encuesta aplicada a los expertos.

La Consulta a Expertos en el caso de este trabajo se realizó a través de la aplicación de una encuesta a catorce especialistas en el campo de la medicina agrupados en las ramas de Médicos de Consultorio, Enfermeras de ETS y Especialistas en ETS. Las encuestas fueron aplicadas en el Centro Médico de la UCI debido a que es donde se centra el problema que se quiere resolver.

2.2.1. Cuestionario de preguntas.

1. ¿Cómo clasifica usted la presencia de la Blenorragia y de la Clamidia en la UCI?

2. ¿La Blenorragia y la Clamidia presentan síntomas muy similares, esto hace que se torne difícil definir cuándo se trata de una y cuándo de otra?
3. ¿Cuáles son los síntomas más frecuentes de ambas ETS?
4. ¿De un paciente presentar el síntoma secreción genital cuál o cuáles son las características que la misma puede presentar para cada una de estas ETS?
5. ¿Cómo se diagnostican en la UCI la Blenorragia y la Clamidia?
6. ¿Cuál o cuáles son las principales causas de infección por Blenorragia y cuáles por Clamidia en la UCI?
7. ¿Cuánto tiempo puede durar la infección por Blenorragia y cuánto la infección por Clamidia?
8. ¿Qué medicamentos se recetan para la infección por Blenorragia y cuáles para la infección por Clamidia?
9. ¿Cree usted que sería de utilidad la creación de una aplicación que le sirva a los médicos como ayuda para diagnosticar cuando un paciente está infectado por Blenorragia y cuándo esta infectado por Clamidia?
10. ¿Las variables que se dan a continuación tienen dependencia? .Responda según los siguientes criterios: Ninguna Relación, Muy Poca Relación, Poca Relación, Relación, Bastante Relación.
 - Secreción, Ardor al Orinar.
 - Secreción, Dolor al Orinar.
 - Secreción, Urgencia de Orinar.
 - Secreción, Mal Olor.
 - Secreción, Sangramiento.
 - Secreción, Sexo Doloroso.
 - Secreción, Dolor en el vientre.
 - Secreción, Dolor en los testículos.
 - Secreción, Sexo.
 - Ardor al Orinar, Mal Olor.
 - Ardor Orinar, Sexo Doloroso
 - Ardor al Orinar, Sangramiento.
 - Ardor al Orinar, Sexo.
 - Urgencia de Orinar, Mal Olor.

- Sexo Doloroso, Sangramiento.
- Sexo, Dolor en los testículos.
- Sexo, Dolor en el Vientre.
- Sexo, Sexo Doloroso.
- Sexo, Sangramiento.

2.2.2. Resultados obtenidos de la Encuesta.

De la encuesta aplicada a los especialistas se sacaron las dependencias existentes entre las variables. Debido a que las opiniones de los expertos pueden ser distintas en cuanto a la relación que existe entre cada una de las variables, se da como resultado final aquel que represente un mayor porcentaje de coincidencia en los criterios que se emitieron. A continuación se mostrarán los resultados obtenidos.

Variable 1	Variable 2	Relación
Secreción (S)	Ardor al Orinar(AO)	Bastante Relación
Secreción	Dolor al Orinar(DO)	Bastante Relación
Secreción	Urgencia de Orinar(UO)	Relación
Secreción	Mal Olor(MO)	Bastante Relación
Secreción	Sangramiento(SGO)	Relación
Secreción	Sexo Doloroso(SD)	Relación
Secreción	Dolor en el vientre(DV)	Relación
Secreción	Dolor en los testículos(DT)	Relación
Secreción	Sexo	Poca Relación
Ardor al Orinar	Mal Olor	Bastante Relación
Ardor al Orinar	Sangramiento	Bastante Relación
Ardor al Orinar	Sexo Doloroso	Relación
Ardor al Orinar	Sexo	Relación
Urgencia de Orinar	Mal Olor	Ninguna
Sexo Doloroso	Sangramiento	Muy Poca Relación
Sexo(SX)	Dolor en los testículos	Relación

Sexo	Sexo Doloroso	Bastante Relación
Sexo	Dolor en el vientre	Relación
Sexo	Sangramiento	Bastante Relación

Tabla 1. Resultados obtenidos de la relación de las variables en la consulta a expertos.

Se definen S, AO, DO, UO, MO, SGO, SD, DV, DT y SX como los vértices del grafo también llamado red y las aristas serán las líneas en un sentido que representan las relaciones que hay entre las variables. Las saetas de doble flecha representan la existencia de mucha relación entre los dos vértices que une, los cuales representan las variables, mientras que la saeta simple representa que existe relación.

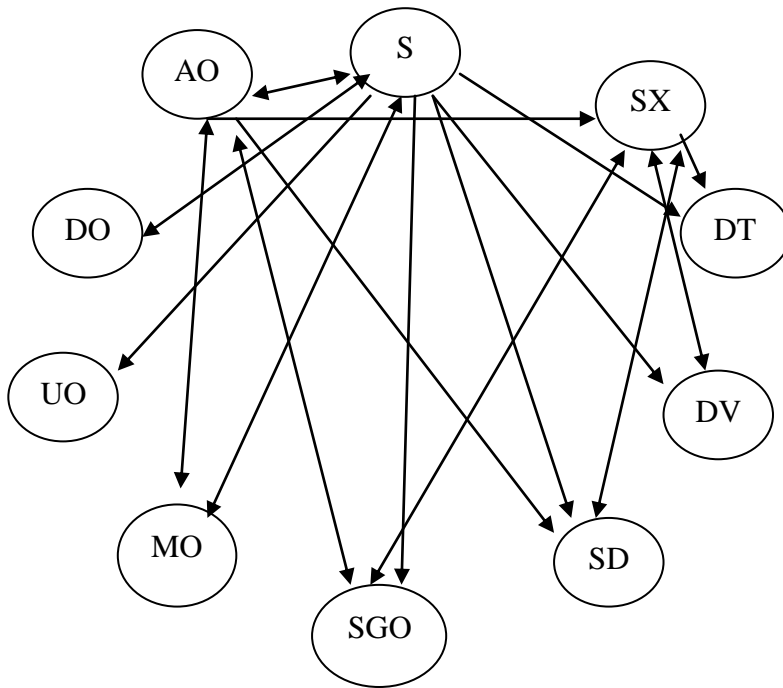


Fig. 1. Grafo de relaciones entre las variables.

Después de determinar las relaciones entre variables se llega a la conclusión de que las más dependientes son la S (Secreción) y el SX (Sexo), y que de la variable que más dependen otras es de la S (Secreción). Una vez halladas las dependencias se procede a construir la Matriz Booleana.

Matriz Booleana										
	S	AO	DO	UO	MO	SGO	SD	DV	DT	SX
S	0	1	1	1	1	1	1	1	1	0
AO	1	0	0	0	1	1	1	0	0	1
DO	1	0	0	0	0	0	0	0	0	0
UO	0	0	0	0	0	0	0	0	0	0
MO	1	1	0	0	0	0	0	0	0	0
SGO	0	1	0	0	0	0	0	0	0	1
SD	0	0	0	0	0	0	0	0	0	1
DV	0	0	0	0	0	0	0	0	0	1
DT	0	0	0	0	0	0	0	0	0	1
SX	0	0	0	0	0	1	1	1	1	0

Tabla 2. Matriz Booleana.

Leyenda

S- Secreción.

DV- Dolor en el vientre.

AO- Ardor al Orinar.

DO- Dolor al Orinar.

UO- Urgencia de Orinar.

MO- Mal Olor.

SX- Sexo.

SGO- Sangramiento

SD- Sexo Doloroso

DT- Dolor en los testículos.

1- Existe relación (Relación, Bastante Relación)

0- No existe relación (Ninguna, Muy Poca, Poca)

La tabla que se muestra a continuación representa los poderes de tipo A ó B y se crea a partir de los resultados obtenidos en la Matriz Booleana. La suma de los elementos de cada fila representan el número total de variables X, que necesitan directamente de la variable Y. Las variables que tengan el mismo valor, tendrán el mismo poder (peso). Al realizar este análisis se ordenan las variables respecto al valor del poder (Tipo A, Tipo B), lo que permite saber cuáles son las más importantes.

Tabla de Poder			
Variabes	Poder A	Poder B	Poder A + Poder B
S	8	3	11
AO	5	3	8
DO	1	1	2
UO	1	1	2
MO	2	2	4
SGO	2	3	5
SD	2	3	5
DV	1	2	3
DT	1	2	3
SX	4	5	9

Tabla 3. Tabla de Poder.

2.3. Tratamiento Estadístico.

Después de obtenidos los resultados de la encuesta, es necesario verificar que los mismos sean lo más confiables posible. Esto se logra aplicando procedimientos estadísticos, que demuestran cuán parecidos o no pueden ser los criterios de los expertos y si se debe aplicar una nueva ronda de encuesta.

Para validar el criterio de los expertos se encontró la media aritmética de los criterios de los mismos por cada elemento, determinando el coeficiente de variación por medio del cálculo de la desviación típica S .

Si $S > 1$ - Se rechaza el criterio de los expertos y se realiza una nueva ronda de preguntas.

Si $S < 1$ - Se acepta el criterio de los expertos.

Si $S = 1$ - Significa un acuerdo total entre los expertos.

El análisis estadístico para cada una de las combinaciones de variables se realizó mediante el paquete estadístico Statgraphics, quedando de la siguiente manera:

Variables	Cantidad	Promedio	Varianza	Desviación Típica	Mínimo	Máximo	Rango
S - AO	14	4.35714	0.862637	0.928783	2	5	3
S - DO	14	4.14286	0.90299	0.949262	2	5	3
S - UO	14	3.78571	1.1044	1.0509	1	5	4
S - MO	14	4.4287	0.571429	0.755929	3	5	2
S - SGO	14	3.71429	1.14286	1.06904	1	5	4
S - SD	14	4.0	0.769231	0.877058	2	5	3
S - DV	14	4.0	0.769231	0.877058	2	5	3
S - DT	14	3.64286	1.17033	1.08182	2	5	3
S - Sexo	14	2.64286	0.554955	0.744946	1	4	3
AO – MO	14	4.28571	0.681319	0.82542	3	5	2
AO – SGO	14	4.21429	0.950549	0.974961	2	5	3
AO – SD	14	3.78571	0.796703	0.892582	2	5	3
AO - Sexo	14	3.92857	0.686813	0.828742	2	5	3
UO - MO	14	1.92857	0.994505	0.997249	1	4	3
SD - SGO	14	1.92857	0.840659	0.916875	1	4	3
Sexo - DT	14	4.0	0.769231	0.877058	2	5	3
Sexo - SD	14	4.07143	0.994505	0.997249	2	5	3
Sexo - DV	14	4.0	0.769231	0.877058	2	5	3
Sexo - SGO	14	4.42857	0.571429	0.755929	3	5	2

Tabla 4. Resultados obtenidos del análisis estadístico.

En este caso se aceptó el criterio de los expertos en cuanto a las relaciones existentes entre las variables, que es lo que se quería validar con la aplicación del método Delphi. Esto ocurre debido a que al calcular la desviación típica para todos los pares de variables, la misma está por debajo de uno en casi todos los

pares, exceptuando tres de ellos, en los cuales está por encima, pero también se aceptan debido a que la diferencia por la que sobrepasan este valor no es significativa.

2.3.1. Gráficos Obtenidos.

Los gráficos que se presentan a continuación corresponden a la concordancia de los expertos al responder la pregunta: ¿Las variables que se dan a continuación tienen dependencia? .Responda según los siguientes criterios: Ninguna Relación, Muy Poca Relación, Poca Relación, Relación, Bastante Relación.

Se toma como muestra la relación entre el par de variables Secreción y Sexo, ya que las mismas son las que más valor tienen según los datos obtenidos en la tabla de poder.

Gráfico de Barras

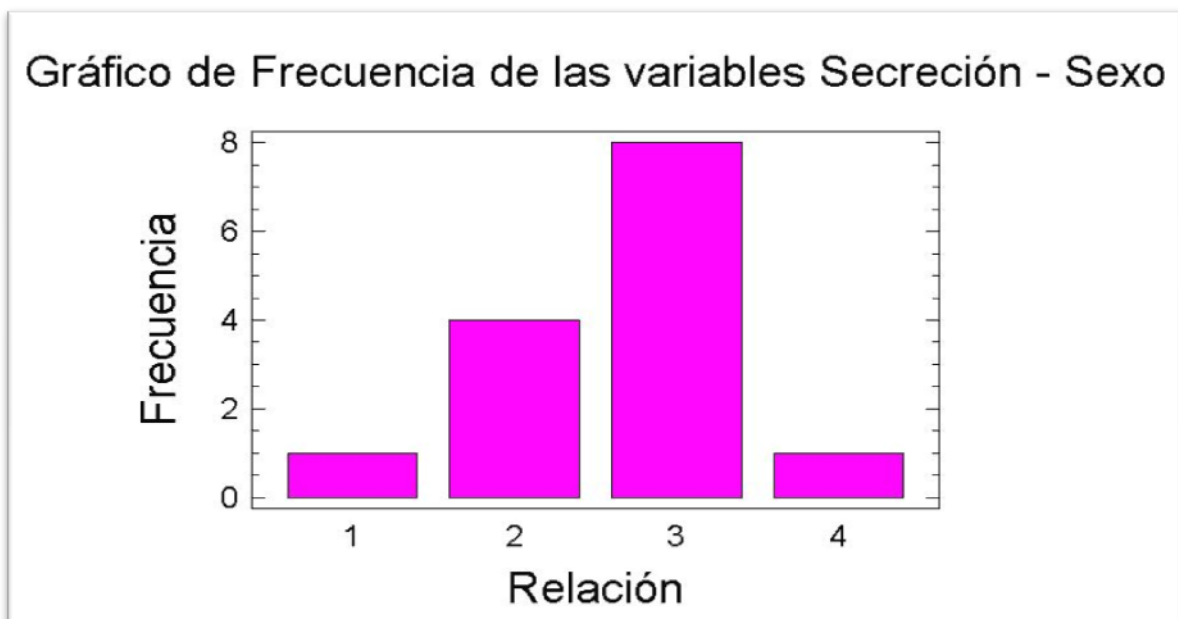


Fig. 2. Grafico de Barras de la relación de las variables Secreción-Sexo.

Gráfico de Pastel

Fig. 3. Grafico de pastel de la relación Secreción-Sexo.

Como se muestra en ambos gráficos, en el caso de la relación entre estas dos variables, ningún experto consideró que la misma fuese “Bastante” por tal razón no aparece dicho valor (5) en las gráficas.

2.4. Proceso de descubrimiento del Conocimiento.

En la actualidad la existencia de voluminosas Bases de Hechos conteniendo grandes cantidades de datos son una realidad en muchas organizaciones y esto trae consigo que frecuentemente se tomen decisiones en base a la intuición y experiencia del decisor, más que considerando la rica información almacenada.

En la realización de este trabajo se utilizarán dos Bases de Hechos pertenecientes a las enfermedades de Blenorragia y Clamidia para obtener el conocimiento que se necesita para la creación de una aplicación que ayude al diagnóstico de las mismas. Se decidió usar dos Bases de Hechos separadas ya que ambas enfermedades presentan los mismos síntomas y al intentarlas agrupar en una sola el conocimiento que se generaba no era del todo correcto debido a la intersección entre los atributos.

2.4.1. Base de Hechos de la Blenorragia.

La Base de Hechos de la Blenorragia consta de cuarenta y cinco instancias que serán utilizadas como ejemplos de entrenamiento para la obtención del conocimiento. Contiene once atributos, todos con valores nominales (cualitativos) y no posee valores perdidos.

A continuación se muestra una tabla de los atributos, con los posibles valores que pueden tomar, así como el orden en el que aparecen ordenados en la Base de Hechos.

Atributos	Valores
Sexo	F(Femenino),M(Masculino)
Secreción	Amarilla Pastosa, Verdosa, Blanca, No(en caso de no presentar ninguna)
Ardor al Orinar	S(Si),N(No)
Dolor al Orinar	S(Si),N(No)
Urgencia de Orinar	S(Si),N(No)
Dolor en el Vientre	S(Si),N(No)
Mal Olor	S(Si),N(No)
Sexo Doloroso	S(Si),N(No)
Sangramiento	S(Si),N(No)
Dolor en los Testículos	S(Si),N(No)
Enfermedad(Atributo Clase)	Blenorragia, No(en caso de no ser Blenorragia)

Tabla 5. Valores de la Base de Hechos de la Blenorragia.

En el caso del atributo “Secreción”, la de mayor predominio fue la Amarilla Pastosa con un 48.89 %, confirmándose con esto el criterio de los expertos acerca de que este tipo de secreción es el predominante en esta ETS.

La Base de Hechos fue creada en una plantilla Excel, donde se guardaron los valores de todos los campos para cada una de las instancias. [\(Ver Anexo 1\)](#).

2.4.2. Base de Hechos de la Clamidia.

La Base de Hechos de la Clamidia está conformada por 37 instancias, que al igual que las de la Blenorragia se utilizarán como datos de entrenamiento, tratándose en el momento de la recopilación, que las mismas fueran lo más diversas posibles. Tiene once atributos nominales (los cuales son los mismos que la Blenorragia pues los síntomas de estas dos ETS son iguales y ellos son los que conforman los atributos) y no incluye valores perdidos. Fue creada en una plantilla Excel ([Ver anexo 2](#)).

La tabla de atributos y los posibles valores que pueden tomar para la Clamidia quedaría de la siguiente manera:

Atributos	Valores
Sexo	F(Femenino),M(Masculino)
Secreción	Blanca, <i>Acuosa</i> , No(en caso de no presentar ninguna)
Ardor al Orinar	S(Si),N(No)
Dolor al Orinar	S(Si),N(No)
Urgencia de Orinar	S(Si),N(No)
Dolor en el Vientre	S(Si),N(No)
Mal Olor	S(Si),N(No)
Sexo Doloroso	S(Si),N(No)
Sangramiento	S(Si),N(No)
Dolor en los Testículos	S(Si),N(No)
Enfermedad(Atributo Clase)	<i>Clamidia</i> , No(en caso de no ser Clamidia)

Tabla 6. Valores de la Base de Hecho de la Clamidia.

La diferencia de esta Base de hechos con la de la Blenorragia, radica en que solo varían algunos valores del atributo Secreción donde en la Clamidia se incluye el valor *Acuosa* y se eliminan los valores Amarilla Pastosa y Verdosa. Además en el atributo clase enfermedad en este caso sería la ETS que se está tratando, permaneciendo el resto de los atributos con los mismos valores que en la Base de Hechos de la Blenorragia.

En el atributo “Secreción” el valor predominante fue la Acuosa con un 45.95% de ocurrencia, lo que resulta estar en concordancia con la opinión de los expertos que consideran este atributo el más importante y este valor el más significativo cuando se trata de esta ETS.

2.4.3. Procesamiento de datos.

El procesamiento de datos es el principal paso cuando se está tratando de obtener conocimiento de conjuntos de datos. En el caso de este trabajo el proceso consta de dos fases principalmente: la preparación de los datos a ser procesados y el procesamiento en sí. Una vez que las Bases de Hechos fueron creadas y revisadas se procedió al análisis de los datos utilizando el software WEKA.

2.4.3.1. Preparación de los Datos.

La primera acción que se realizó fue convertir las Bases de Hechos de la Blenorragia y de la Clamidia a un fichero extensión **.arff** (Attribute-Relation File Format) que es la que soporta el software a utilizar. Para esto se pasaron todos los datos de cada una de las Bases de Hechos a dos archivos Wordpad con nombres ETS (Blenorragia) y ETS (Clamidia) y se guardaron con la extensión especificada.

Como parte del proceso de preparación de los datos se eliminaron inconsistencias de forma manual tales como: errores a la hora de pasar valores de los atributos (que pueden traer como consecuencia la existencia de ruido) y la eliminación de instancias que se habían duplicado. Este paso se considera de vital importancia debido a que su correcta aplicación impacta de manera directa en el conocimiento a descubrir.

En ambas Bases de Hechos los atributos se declararon de tipo **symbolic** lo que significa que se especificarán los valores que pueden tomar entre llaves.

A continuación se muestra como quedaron los ficheros **.arff** para cada una de las ETS:

Blenorragia	Clamidia
-------------	----------

<pre> @relation ETS_Blenorragia.symbolic @attribute sexo {M,F} @attribute secrecion {amarillo_pastoso,verdosa,blanca,ho} @attribute ardor_orinar {S,N} @attribute dolor_orinar {S,N} @attribute urgencia_orinar {S,N} @attribute dolor_vientre {S,N} @attribute mal_olor {S,N} @attribute sexo_doloroso {S,N} @attribute sangramiento {S,N} @attribute dolor_testiculos{S,N} @attribute enfermedad {Blenorragia,N} @data M,verdosa,S,N,N,N,N,N,N,Blenorragia F,verdosa,N,N,N,N,N,N,N,N F,verdosa,N,N,S,N,N,N,N,N,Blenorragia M,verdosa,N,S,N,N,N,N,N,N,Blenorragia F,verdosa,S,N,N,N,N,N,N,N M,verdosa,N,N,N,N,N,N,N,N F,verdosa,N,N,N,N,S,N,N,N,Blenorragia F,verdosa,S,N,N,N,N,S,N,N,Blenorragia F,verdosa,S,N,N,N,N,N,S,N,Blenorragia M,amarillo_pastoso,S,N,N,N,N,N,N,N,Blenorragia M,amarillo_pastoso,N,N,N,N,N,N,N,S,Blenorragia M,amarillo_pastoso,N,S,N,N,N,N,N,N,Blenorragia </pre>	<pre> @relation ETS_Clamidia.symbolic @attribute sexo {M,F} @attribute secrecion {blanca,acuosa,no} @attribute ardor_orinar {S,N} @attribute dolor_orinar {S,N} @attribute urgencia_orinar {S,N} @attribute dolor_vientre {S,N} @attribute mal_olor {S,N} @attribute sexo_doloroso {S,N} @attribute sangramiento {S,N} @attribute dolor_testiculos{S,N} @attribute enfermedad {clamidia,No} @data F,blanca,S,N,N,N,N,N,N,No M,blanca,N,N,N,N,N,N,N,No M,blanca,S,N,N,N,N,N,N,N,clamidia M,blanca,N,S,N,N,N,N,N,N,clamidia M,blanca,N,S,N,N,N,N,N,N,No F,blanca,N,N,N,S,N,N,N,N,clamidia M,blanca,S,N,N,N,N,N,N,N,clamidia F,blanca,N,S,N,N,N,N,N,N,clamidia M,blanca,N,N,N,N,N,N,N,N,No M,blanca,S,N,N,N,N,N,N,N,clamidia F,blanca,N,N,N,N,N,N,N,S,clamidia M,blanca,N,N,N,N,N,N,N,N,No M,acuosa,S,N,N,N,N,N,N,N,clamidia </pre>
--	--

Tabla 7. Archivos .arff de la Blenorragia y Clamidia.

2.4.3.2. Análisis del tipo de conocimiento a obtener.

Antes de pasar al procesamiento de datos se decidió determinar el tipo de conocimiento que se desea obtener, la forma en que se obtendrá y para qué se utilizará, porque con esto se tendría una visión más clara de lo que realmente se desea. Para ello se plantearon una serie de preguntas:

1. **¿Cuál es el conjunto de datos que se consideran relevantes para obtener el conocimiento?**
 Se considera relevante para la obtención del conocimiento todos los datos contenidos en ambas Bases de Hechos, por lo que se decidió trabajar con ellas íntegras, sin excluir ningún atributo del procesamiento. Esto se debe a que todos los datos representan información vital a la hora de diagnosticar una posible infección por cualquiera de estas dos ETS.
2. **¿Cuál es el conocimiento que se desea obtener?**

Se desea obtener de cada Base de Hechos un conjunto de reglas que ayuden a tomar decisiones sobre si una persona está infectada o no por Blenorragia o Clamidia.

3. ¿Cuál es la mejor forma de obtener ese conocimiento?

Procesando los datos con la técnica árboles de decisión y el algoritmo ID3.

4. ¿Cómo se visualizará el conocimiento obtenido?

Para visualizar los patrones descubiertos se utilizará un AD por cada una de las Bases de Hechos.

Después de responder estas interrogantes se procedió a determinar cuál interfaz del software WEKA era la más propicia para el procesamiento.

El software posee cuatro interfaces:

1. **Simple CLI:** Ventana para ejecutar acciones a través de comandos.
2. **Experimenter:** Permite aplicar más de un algoritmo sobre uno o más conjuntos de datos y comparar estadísticamente los resultados.
3. **Explorer:** Es la opción que permite llevar a cabo la ejecución de los algoritmos de análisis implementados sobre los ficheros de entrada, una ejecución independiente por cada prueba.
4. **KnowledgeFlow:** Esta opción permite llevar a cabo las mismas acciones del Explorer, con una configuración totalmente gráfica

De estas interfaces se eligió la **Explorer** ya que entre los algoritmos que trae implementados están los de AD y el ID3 que es el que se desea utilizar.

2.3.4.3. Procesamiento de la Base de Hechos de la Blenorragia y resultados obtenidos.

A continuación de creados los archivos. *arrf* ETS (Blenorragia) y ETS (Clamidia) se procedió al procesamiento de los datos en el software.

El primer paso para el procesamiento de esta Base de Hechos fue cargar el fichero ETS (Blenorragia) en el software, donde se visualizan las características de los atributos (valores que pueden tomar, cantidad por valor, si hay o no valores perdidos y tipo de datos). A esta primera acción se le llama

2.5. Árboles de decisión.

El software WEKA permite con otros algoritmos como el C4.5 visualizar el AD, pero con el ID3 no ocurre así. Por esta razón a partir de los resultados obtenidos del procesamiento de datos de las Bases de Hechos de la Blenorragia y de la Clamidia se construyeron los AD de las mismas, para lograr con esto un mejor entendimiento del conocimiento que se alcanzó en el procesamiento.

2.5.1. Árbol de decisión de la Blenorragia.

El AD que se muestra a continuación fue obtenido basándose en los resultados del procesamiento de la Base de Hechos de la Blenorragia (Ver Fig. 4).

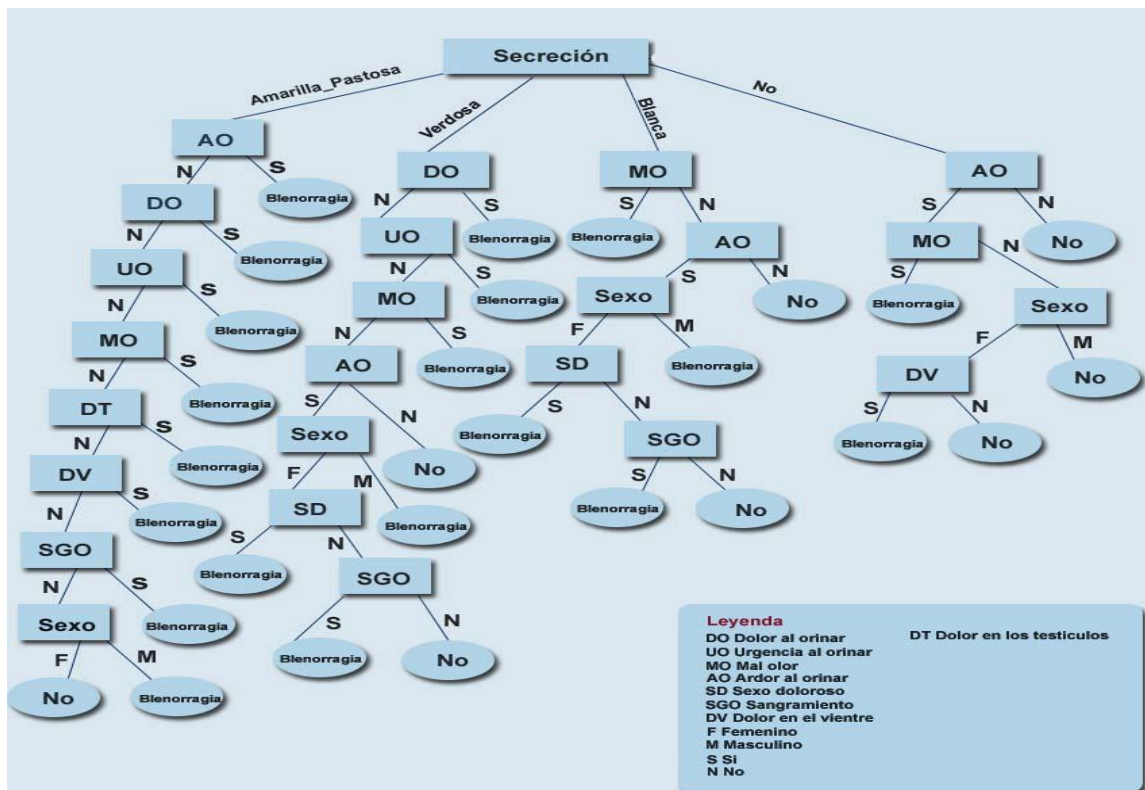


Fig. 6. Árbol de Decisión de la ETS Blenorragia.

El AD de la Blenorragia tiene nueve niveles, siendo su atributo raíz la secreción. En el se observan las dependencias que pueden existir entre los síntomas de esta enfermedad. Del mismo es posible derivar

veintiocho reglas de producción, representadas por el camino que se utiliza para llegar a cada hoja (clase).

2.5.2. Árbol de decisión de la Clamidia.

El AD basado en este resultado alcanzado en el procesamiento de la Base de Hechos de la Clamidia (Ver Fig. 5) quedaría de la siguiente manera:

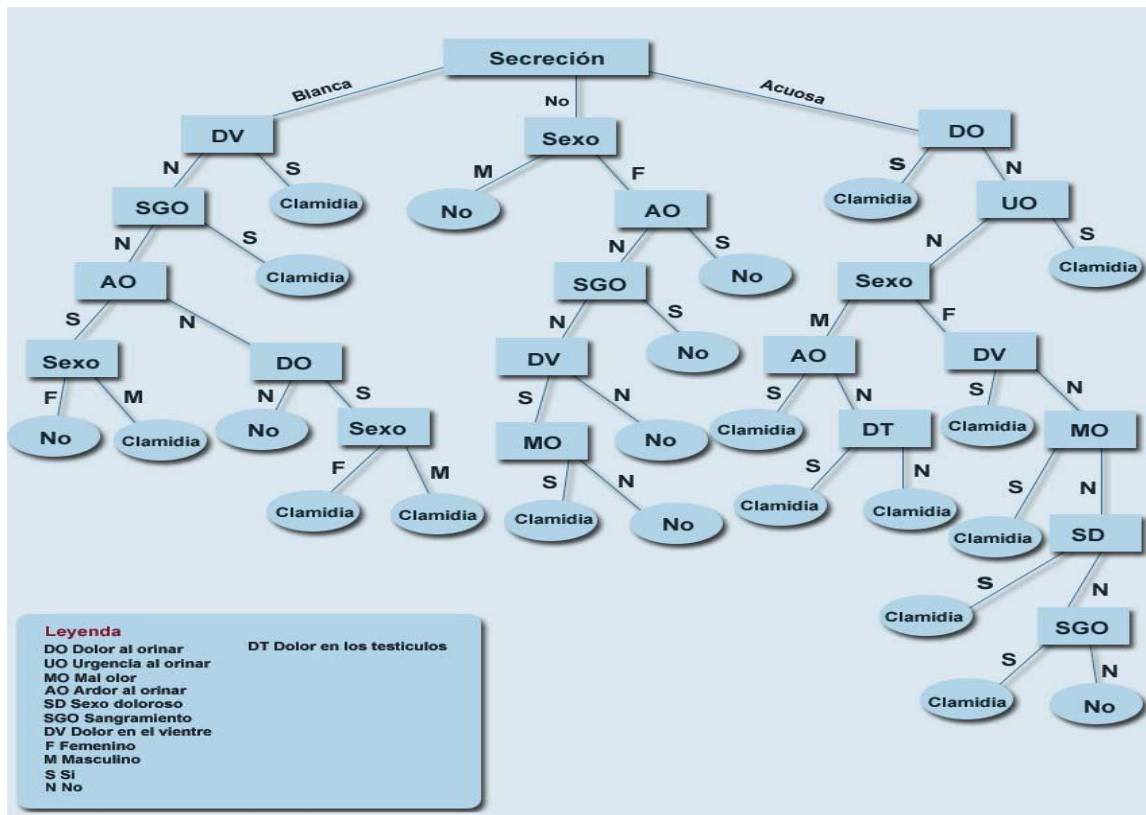


Fig. 7. Árbol de Decisión de la ETS Clamidia.

Como se observa este AD consta de ocho niveles y el atributo raíz al igual que en el de la Blenorragia también resulta ser la secreción. De él es posible derivar veintitrés reglas de producción.

2.6. Pruebas.

Una vez obtenidos los AD para las ETS Blenorragia y la Clamidia, se realizaron dos pruebas sobre cada uno de ellos. Con esto no se cambia en nada el resultado de los AD, ni la aplicación del algoritmo ID3 mediante el cual se obtuvieron, si no que permite saber qué por ciento de clasificación de instancias correctas e incorrectas se logran a partir de ellos.

Las dos pruebas que se le aplicaron a los AD fueron hechas en el propio software WEKA, el cual además de permitir realizar diversos tipos de prueba, visualiza los valores que se obtienen de las mismas.

Durante este proceso se utilizan en ambos casos un conjunto de entrenamiento que es aquel a partir del cual se obtiene el conocimiento y un conjunto de prueba que nos permite medir el alcance de ese conocimiento.

2.6.1. Pruebas para la Blenorragia y resultados obtenidos.

En una primera prueba se escogió la opción *Use Training Set* con la que los datos de entrada son los mismos para el conjunto de entrenamiento que de prueba, es decir, se utiliza la Base de Hechos completa para ambas operaciones.

Los resultados obtenidos de esta primera prueba se muestran a continuación:

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      44          97.7778 %
Incorrectly Classified Instances    1           2.2222 %
Kappa statistic                    0.9379
Mean absolute error                 0.0296
Root mean squared error             0.1217
Relative absolute error             7.9025 %
Root relative squared error         28.313 %
Total Number of Instances          45

```

Fig. 8. Resultado de aplicar pruebas a la Base de Hechos de la Blenorragia con la opción *Use Training Set*

Con esta prueba se obtuvo que de las 45 instancias que tiene la Base de Hechos, que cuarenta y cuatro fueron correctamente clasificadas, lo que da la seguridad de que se cubre una amplia cantidad de posibilidades con este árbol.

Para la segunda prueba se escogió la opción de Testeo (Prueba) **Supplied Test Set** que permite utilizar un conjunto de datos para entrenamiento y otro para prueba. Para el entrenamiento se usó el conjunto de todos los datos con el que se generó el árbol. Para el conjunto de prueba se aplicó un filtrado sobre la Base de Hechos utilizando el filtro de instancia **RemoveRange** que elimina todas las instancias en un rango que se le especifique. En este caso se dejaron las veinticuatro primeras instancias y se guardaron en un nuevo archivo **.arff** generado por el software llamado ETS (Blenorragia Testeo), el que fue cargado en el período de la prueba.

Seguidamente se observan los resultados alcanzados:

```

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      23      95.8333 %
Incorrectly Classified Instances     1       4.1667 %
Kappa statistic                     0.8636
Mean absolute error                  0.0278
Root mean squared error              0.1361
Relative absolute error              7.775 %
Root relative squared error          33.2863 %
Total Number of Instances           24

```

Fig. 9. Resultado de aplicar pruebas a la Base de Hechos de la Blenorragia con la opción *Supplied Test Set*

En el caso de esta prueba los resultados también demuestran ser favorables debido a que de 24 instancias que se utilizaron como prueba veintitrés fueron clasificados de forma correcta. Lo que quiere decir que el por ciento de clasificación de instancias incorrectas es muy bajo.

Por tanto se llegó a la conclusión de que el conocimiento que representa este árbol es fiable y puede ser utilizado como base de la aplicación a crear, porque las dos pruebas que se le aplicaron demostraron que

el por ciento de clasificación de instancia es elevado. Sirviendo esto como garantía de exactitud a la hora de realizar un diagnóstico de esta enfermedad.

2.6.2. Pruebas para la Clamidia y resultados obtenidos.

Al AD de la Clamidia se le aplicaron las mismas pruebas que al de la Blenorragia lográndose los resultados siguientes en la clasificación de las instancias con la opción **Use Training Set**.

```

=== Evaluation on training set ===
=== Summary ===

Correctly Classified Instances      35          94.5946 %
Incorrectly Classified Instances    2           5.4054 %
Kappa statistic                    0.8771
Mean absolute error                 0.0631
Root mean squared error             0.1776
Relative absolute error             13.767 %
Root relative squared error         37.1913 %
Total Number of Instances          37

```

Fig. 10. Resultado de aplicar pruebas a la Base de Hechos de la Clamidia con la opción **Use Training Set**.

Para la segunda prueba se aplicó el mismo filtro que para la Blenorragia (**RemoveRange**), sólo que esta vez se tomaron las 20 primeras instancias de la Base de Hecho de la Clamidia y se utilizaron como conjunto de prueba o testeo.

```

=== Evaluation on test set ===
=== Summary ===

Correctly Classified Instances      18          90 %
Incorrectly Classified Instances    2           10 %
Kappa statistic                    0.7826
Mean absolute error                 0.1
Root mean squared error             0.2297
Relative absolute error             21.1957 %
Root relative squared error         46.7307 %
Total Number of Instances          20

```

Fig. 11. Resultado de aplicar pruebas a la Base de Hechos de la Clamidia con la opción **Supplied Test Set**

Con ambas pruebas se obtuvieron resultados satisfactorios ya que el por ciento de clasificación de instancia incorrectas es de 2 en ambos casos, lo que también es un gran resultado y permite saber que el conocimiento que se almacena en el AD de la Clamidia, servirá para diagnosticar de forma correcta un elevado número de casos de esta enfermedad.

3. Propuesta de Solución.

Introducción:

En este capítulo se presentarán los conjuntos de reglas pertenecientes a las ETS Blenorragia y Clamidia, las cuales constituyen la base de la aplicación. Se mostrará la manera en que fue creado el sistema para el diagnóstico de estas ETS, así como las interfaces que lo conforman, sus principales características y algunos resultados obtenidos en la utilización del mismo.

3.1. Sistema de Reglas.

En el capítulo dos de la parte del procesamiento de datos, se obtuvieron dos AD, uno para la Blenorragia y otro para la Clamidia. Las pruebas aplicadas demostraron que ambos AD tienen un alto por ciento de clasificación de instancias. Por tanto la información almacenada en ellos se traducirá en reglas, pudiéndose afirmar que el conocimiento obtenido en el procesamiento se modelara como un sistema de reglas, las cuales constituyen la información esencial que se incluirá en la aplicación, y por tanto el conocimiento de el cual se servirá el sistema, para poder decidir si una persona está infectada o no por cualquiera de estas dos ETS (Blenorragia o Clamidia).

En los AD el camino hacia cada hoja representa una regla, pudiéndose extraer una amplia variedad de ellas de un mismo árbol, las mismas se pueden utilizar para tomar decisiones idénticas a si se empleara el árbol tal y como está. Por esta razón el conocimiento expresado en el árbol se mantiene intacto en el sistema de reglas, y la clasificación con ellas resulta ser bastante certera.

3.1.1. Reglas para la ETS Blenorragia.

El conjunto de reglas de la ETS Blenorragia fue obtenido a partir del AD de la misma ([Ver Fig. 6](#)) y suman un total de 28 reglas, agrupadas por el tipo de secreción: perteneciendo 9 de ellas a la secreción Amarilla Pastoso, 8 a la secreción Verdosa, 6 a la Blanca y 4 al grupo que no presenta ninguna secreción.

1. **SI** Secreción= Amarillo_Pastoso y AO=Si **ENTONCES Blenorragia**
2. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=Si **ENTONCES Blenorragia**
3. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=Si **ENTONCES Blenorragia**
4. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=No y MO=Si **ENTONCES Blenorragia**

5. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=No y MO=No y DT= Si **ENTONCES Blenorragia**
6. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=No y MO=No y DT= No y DV=Si **ENTONCES Blenorragia**
7. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=No y MO=No y DT= No y DV=No y SGO=Si **ENTONCES Blenorragia**
8. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=No y MO=No y DT= No y DV=No y SGO=No y Sexo=M **ENTONCES Blenorragia**
9. **SI** Secreción= Amarillo_Pastoso y AO=No y DO=No y UO=No y MO=No y DT= No y DV=No y SGO=No y Sexo=F **ENTONCES No**
10. **SI** Secreción= Verdosa y DO=Si **ENTONCES Blenorragia**
11. **SI** Secreción= Verdosa y DO=No y UO=Si **ENTONCES Blenorragia**
12. **SI** Secreción= Verdosa y DO=No y UO=No y MO=Si **ENTONCES Blenorragia**
13. **SI** Secreción= Verdosa y DO=No y UO=No y MO=No y AO=No **ENTONCES No**
14. **SI** Secreción= Verdosa y DO=No y UO=No y MO=No y AO=Si y Sexo=M **ENTONCES Blenorragia**
15. **SI** Secreción= Verdosa y DO=No y UO=No y MO=No y AO=Si y Sexo=F y SD=Si **ENTONCES Blenorragia**
16. **SI** Secreción= Verdosa y DO=No y UO=No y MO=No y AO=Si y Sexo=F y SD=No y SGO=Si **ENTONCES Blenorragia**
17. **SI** Secreción= Verdosa y DO=No y UO=No y MO=No y AO=Si y Sexo=F y SD=No y SGO=No **ENTONCES No**
18. **SI** Secreción=Blanca y MO=Si **ENTONCES Blenorragia**
19. **SI** Secreción=Blanca y MO=No y AO=No **ENTONCES No**
20. **SI** Secreción=Blanca y MO=No y AO=Si y Sexo=M **ENTONCES Blenorragia**
21. **SI** Secreción=Blanca y MO=No y AO=Si y Sexo=F y SD=Si **ENTONCES Blenorragia**
22. **SI** Secreción=Blanca y MO=No y AO=Si y Sexo=F y SD=No y SGO=Si **ENTONCES Blenorragia**
23. **SI** Secreción=Blanca y MO=No y AO=Si y Sexo=F y SD=No y SGO=No **ENTONCES No**
24. **SI** Secreción=No y AO=No **ENTONCES No**
25. **SI** Secreción=No y AO=Si y MO=Si **ENTONCES Blenorragia**
26. **SI** Secreción=No y AO=Si y MO=No y Sexo=M **ENTONCES No**

27. SI Secreción=No y AO=Si y MO=No y Sexo=F y DV=No **ENTONCES No**

28. SI Secreción=No y AO=Si y MO=No y Sexo=F y DV=Si **ENTONCES Blenorragia**

3.1.2. Reglas para la ETS la Clamidia.

Las reglas de la Clamidia fueron obtenidas a partir del AD de esta ETS (Ver Fig. 7). Al igual que las de la Blenorragia se dividieron por el tipo de secreción: 7 para la secreción Blanca, 6 para cuando no se presenta ninguna secreción y 10 para la secreción acuosa, la cual es la más característica de esta ETS.

1. **SI** Secreción=Blanca y DV=Si **ENTONCES Clamidia**
2. **SI** Secreción=Blanca y DV= No y SGO=Si **ENTONCES Clamidia**
3. **SI** Secreción=Blanca y DV= No y SGO=No y AO=Si y Sexo=M **ENTONCES Clamidia**
4. **SI** Secreción=Blanca y DV= No y SGO=No y AO=Si y Sexo=F **ENTONCES No**
5. **SI** Secreción=Blanca y DV= No y SGO=No y AO=No y DO=No **ENTONCES No**
6. **SI** Secreción=Blanca y DV= No y SGO=No y AO=No y DO=Si y Sexo=F **ENTONCES Clamidia**
7. **SI** Secreción=Blanca y DV= No y SGO=No y AO=No y DO=Si y Sexo=M **ENTONCES Clamidia**
8. **SI** Secreción=No y Sexo=M **ENTONCES No**
9. **SI** Secreción=No y Sexo=F y AO=Si **ENTONCES No**
10. **SI** Secreción=No y Sexo=F y AO=No y SGO=Si **ENTONCES No**
11. **SI** Secreción=No y Sexo=F y AO=No y SGO=No y DV=No **ENTONCES No**
12. **SI** Secreción=No y Sexo=F y AO=No y SGO=No y DV=Si y MO=Si **ENTONCES Clamidia**
13. **SI** Secreción=No y Sexo=F y AO=No y SGO=No y DV=Si y MO=No **ENTONCES No**
14. **SI** Secreción=Acuosa y DO=Si **ENTONCES Clamidia**
15. **SI** Secreción=Acuosa y DO=No y UO=Si **ENTONCES Clamidia**
16. **SI** Secreción=Acuosa y DO=No y UO=No y Sexo=M y AO=Si **ENTONCES Clamidia**
17. **SI** Secreción=Acuosa y DO=No y UO=No y Sexo=M y AO=No y DT=Si **ENTONCES Clamidia**
18. **SI** Secreción=Acuosa y DO=No y UO=No y Sexo=M y AO=No y DT=No **ENTONCES Clamidia**
19. **SI** Secreción=Acuosa y DO=No y UO=No y Sexo=F y DV=Si **ENTONCES Clamidia**
20. **SI** Secreción=Acuosa y DO=No y UO=No y Sexo=F y DV=No y MO=Si **ENTONCES Clamidia**
21. **SI** Secreción=Acuosa y DO=No y UO=No y Sexo=F y DV=No y MO=No y SD=Si **ENTONCES Clamidia**

22. SI Secreción=Acuosa y DO=No y UO=No y Sexo=F y DV=No y MO=No y SD=No y SGO=Si
ENTONCES Clamidia

23. SI Secreción=Acuosa y DO=No y UO=No y Sexo=F y DV=No y MO=No y SD=No y SGO=No
ENTONCES No

3.2. Características de la aplicación.

Para dar solución al problema que dio origen a este trabajo, se creó una aplicación cuyas principales características son las siguientes:

- ✓ Creada en el lenguaje C++.
- ✓ Programación orientada a objetos.
- ✓ Permite definir dada una entrada de datos si una persona está infectada por la ETS Blenorragia, Clamidia o por ninguna de las dos.
- ✓ Consta de una sola clase nombrada Diagnóstico, la misma contiene como atributos los correspondientes a las instancias de las Bases de Hecho: Sexo, Secreción, Ardor al Orinar, Dolor al Orinar, Urgencia de Orinar, Mal Olor, Dolor en el vientre, Dolor en los testículos, Sexo doloroso y Sangramiento.
- ✓ Se clasifica como un Sistema basado en reglas. Las reglas que lo soportan son las obtenidas de los AD de cada una de las ETS que se tratan.

3.3. Interfaces de la aplicación.

La aplicación consta de dos ventanas: la principal, que es en la que se lleva a cabo todo el proceso de entrada de datos y en la cual se observará el diagnóstico final obtenido y la ventana Error cuya función es mostrar mensajes de error al usuario cuando deje de llenar los campos Sexo y Secreción.

El funcionamiento de la aplicación se explicará a través de las interfaces que la conforman. Siendo la secuencia de imágenes que se muestra a continuación, la misma que se sigue para realizar un diagnóstico.

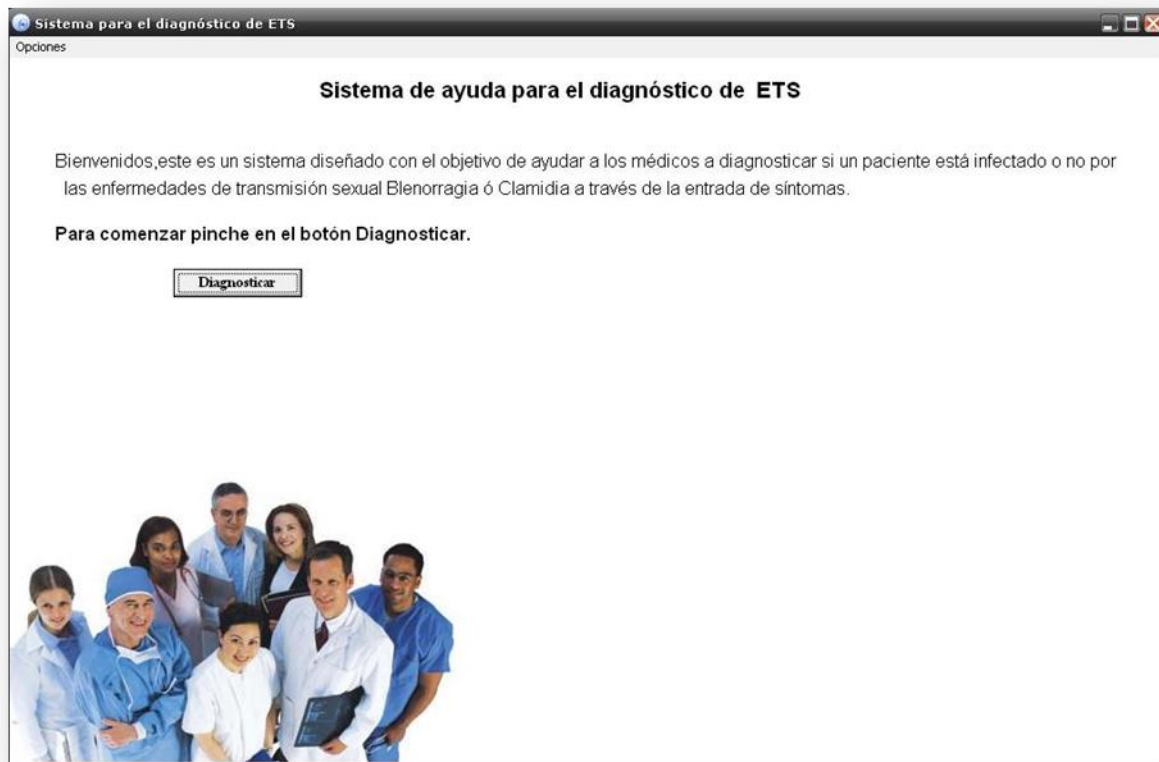


Fig. 12. Interfaz de Inicio.

La interfaz de inicio es lo primero que se le muestra al usuario al abrir la aplicación. En ella se da una breve explicación de la función del sistema y pinchando sobre el botón Diagnosticar se comienza el proceso de diagnóstico.



Fig.13. Interfaz Sexo.

En la interfaz Sexo se le solicita al usuario que entre el sexo que corresponda. La misma fue creada debido a que las interfaces donde se entran los síntomas son distintas para cada uno de los sexos.



Fig. 14. Interfaz Síntomas.

En la Interfaz Síntomas el usuario debe marcar los que padece. En la Fig. 14 se muestra la interfaz para un paciente masculino, ya que aparece el síntoma “Dolor en los testículos” que es propio de este sexo. De ser la interfaz del sexo femenino aparecerían los síntomas “Dolor en el vientre”, “Sangramiento” y “Sexo doloroso” acompañando a los demás.

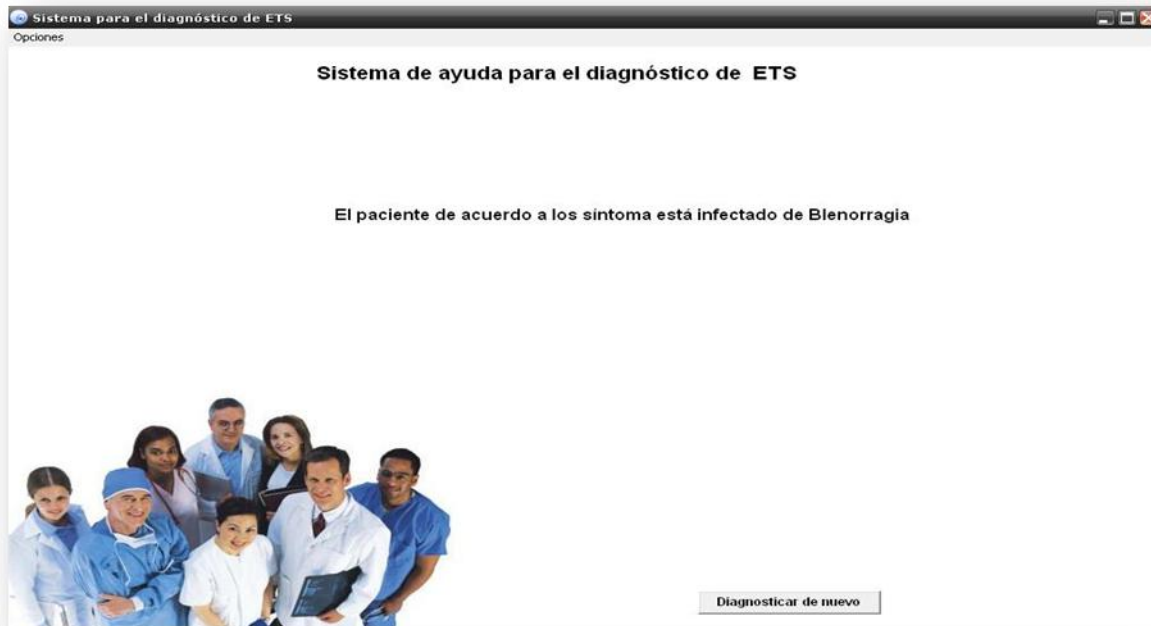


Fig. 15. Interfaz Diagnóstico.

En la interfaz Diagnóstico, se le muestra al usuario el diagnóstico que se obtuvo, de acuerdo a los síntomas que se le introdujeron a la aplicación. En el caso de la Fig. 15 se muestra un diagnóstico de Blenorragia, los otros tipos del mismo que se pudieran dar serían Clamidia, o que el paciente no está infectado ni de Blenorragia ni de Clamidia. Además brinda la opción de realizar un nuevo diagnóstico si se desea.



Fig. 16. Interfaz Error.

En la interfaz error se le muestra al usuario los campos que debe llenar para poder continuar avanzando. En ella se pueden señalar dos tipos de errores: el primero por no seleccionar el Sexo y el segundo por dejar vacío el campo Secreción.

3.4. Pruebas hechas a la aplicación.

Para verificar el correcto funcionamiento de la aplicación se le aplicaron algunas pruebas. Se utilizaron juegos de datos como entrada, conociéndose la salida que se espera con ellos. A continuación se muestran las pruebas aplicadas con tres juegos de datos distintos, uno para cada uno de las posibles salidas que puede brindar la aplicación.

Juego de Datos 1

Entrada: Sexo masculino, Secreción Verdosa, Dolor Testículos y Ardor al orinar.

Salida Esperada: El paciente según los síntomas esta infectado de Blenorragia.

Salida Obtenida: El paciente según los síntomas esta infectado de Blenorragia.

Juego de Datos 2

Entrada: Sexo femenino, Secreción Acuosa, Mal olor y Dolor en el vientre

Salida Esperada: El paciente según los síntomas esta infectado de Clamidia.

Salida Obtenida: El paciente según los síntomas esta infectado de Clamidia.

Juego de Datos 3

Entrada: Sexo femenino, Secreción Blanca, Ardor orinar.

Salida Esperada: El paciente según los síntomas no está infectado de Blenorragia ni Clamidia.

Salida Obtenida: El paciente según los síntomas no está infectado de Blenorragia ni Clamidia.

Las pruebas aplicadas con los distintos juegos de datos, se consideran satisfactorias. En los tres casos los resultados obtenidos estuvieron en relación con los esperados. Por esta razón se plantea que el sistema creado es lo bastante confiable en cuanto a los resultados que brinda, tomando siempre en cuenta que los mismos vienen dados por el conocimiento que él mismo tiene almacenado.

3.5. Ventajas del uso de la aplicación.

- ✓ Puede ser utilizada por personas que no tengan conocimiento sobre las ETS Blenorragia y Clamidia. La aplicación cuenta con un conjunto de imágenes que sirven como guía para el síntoma secreción, que posee varios tipos, siendo el más difícil de definir.
- ✓ Las interfaces que la conforman son fácilmente entendibles. En cada interfaz se informa lo que se debe hacer en ella.
- ✓ El diagnóstico es totalmente confidencial. Si la aplicación es utilizada por una persona que no sea un especialista, en el momento de dar el diagnóstico, solo ella sabrá el resultado del mismo. Por otra parte no se almacenan los resultados que se obtienen de los diagnósticos, por lo que no quedaría constancia una vez que se halla utilizado el sistema por una persona del diagnóstico que obtuvo.
- ✓ Se puede usar en paralelo con el exudado genital para dar un diagnóstico.

CONCLUSIONES

En el desarrollo de este trabajo de diploma se realizaron profundos estudios sobre las técnicas de Inteligencia Artificial a través de las cuales es posible obtener conocimiento. De todo este proceso se arribó a la conclusión de que la técnica de Árboles de Decisión constituye un medio idóneo como método de obtención de conocimiento para el diagnóstico de enfermedades y que el software WEKA es una potente herramienta de procesamiento de datos que permite la obtención de patrones de manera eficaz. Durante la realización de este trabajo se confeccionaron dos Bases de Hechos pertenecientes a las Enfermedades de Transmisión sexual Blenorragia y Clamidia, para su correcta creación se consultaron un grupo de especialistas en estas enfermedades. Se utilizó la técnica de Consulta a Expertos para conocer las relaciones existentes entre las variables que conformaban los Árboles de Decisión y para el análisis estadístico realizado a la misma se utilizó el software Statgraphics. La extracción del conocimiento se efectuó utilizando el software WEKA y los resultados obtenidos fueron sometidos a un proceso de prueba en el propio software. El conocimiento que se obtuvo se derivó en un conjunto de reglas, que sirvieron como soporte para la aplicación que se creó, por lo que la misma es considerada un sistema basado en reglas, al cual se le aplicaron pruebas con juegos de datos obteniéndose resultados favorables. Por todo lo planteado anteriormente se concluye que el objetivo general de este trabajo de diploma fue cumplido satisfactoriamente.

RECOMENDACIONES

- ✓ Ampliar la aplicación desarrollada en este trabajo a otros tipos de Enfermedades de Transmisión Sexual.
- ✓ Poner a disposición del centro médico de la Universidad de las Ciencias Informáticas y de la comunidad universitaria la aplicación creada en este trabajo de diploma para el diagnóstico de las Enfermedades de Transmisión Sexual Blenorragia y Clamidia.
- ✓ Extender la implementación de la aplicación para que permita el diagnóstico de las ETS Blenorragia y Clamidia utilizando Aprendizaje Automático.

ANEXOS

Anexo 1

Base de Hechos(Blenorragia) [Modo de compatibilidad] - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sexo	Secreción	AO	DO	UO	DV	MO	SD	SGO	DT	Enfermedad	
2	M	verdosa	S	N	N	N	N	N	N	N	Blenorragia	
3	F	verdosa	N	N	N	N	N	N	N	N	No	
4	F	verdosa	N	N	S	N	N	N	N	N	Blenorragia	
5	M	verdosa	N	S	N	N	N	N	N	N	Blenorragia	
6	F	verdosa	S	N	N	N	N	N	N	N	No	
7	M	verdosa	N	N	N	N	N	N	N	N	No	
8	F	verdosa	N	N	N	N	S	N	N	N	Blenorragia	
9	F	verdosa	S	N	N	N	N	S	N	N	Blenorragia	
10	F	verdosa	S	N	N	N	N	N	S	N	Blenorragia	
11	M	amarillo_pastoso	S	N	N	N	N	N	N	N	Blenorragia	
12	M	amarillo_pastoso	N	N	N	N	N	N	N	S	Blenorragia	
13	M	amarillo_pastoso	N	S	N	N	N	N	N	N	Blenorragia	
14	F	amarillo_pastoso	S	N	N	N	N	N	N	N	Blenorragia	
15	F	amarillo_pastoso	N	N	S	N	N	N	N	N	Blenorragia	
16	M	amarillo_pastoso	N	N	N	N	S	N	N	N	Blenorragia	
17	M	amarillo_pastoso	N	N	N	N	N	N	N	N	No	
18	M	amarillo_pastoso	N	N	N	N	N	N	N	S	Blenorragia	
19	M	amarillo_pastoso	S	S	N	N	N	N	N	N	Blenorragia	
20	F	amarillo_pastoso	N	N	N	N	S	N	N	N	Blenorragia	
21	M	amarillo_pastoso	N	S	N	N	N	N	N	N	Blenorragia	

Fig. 17. Base de Hechos en Excel de la Blenorragia.

Anexo 2

	A	B	C	D	E	F	G	H	I	J	K	L
1	Sexo	Secreción	AO	DO	UO	DV	MO	SD	SGO	DT	Enfermedad	
2	F	blanca	S	N	N	N	N	N	N	N	No	
3	M	blanca	N	N	N	N	N	N	N	N	No	
4	M	blanca	S	N	N	N	N	N	N	N	clamidia	
5	M	blanca	N	S	N	N	N	N	N	N	clamidia	
6	M	blanca	N	S	N	N	N	N	N	N	No	
7	F	blanca	N	N	N	S	N	N	N	N	clamidia	
8	M	blanca	S	N	N	N	N	N	N	N	clamidia	
9	F	blanca	N	S	N	N	N	N	N	N	clamidia	
10	M	blanca	N	N	N	N	N	N	N	N	No	
11	M	blanca	S	N	N	N	N	N	N	N	clamidia	
12	F	blanca	N	N	N	N	N	N	S	N	clamidia	
13	M	blanca	N	N	N	N	N	N	N	N	No	
14	M	acuosa	S	N	N	N	N	N	N	N	clamidia	
15	F	acuosa	N	N	N	N	N	N	N	N	No	
16	M	acuosa	N	N	N	N	N	N	N	N	clamidia	
17	F	acuosa	N	N	S	N	N	N	N	N	clamidia	
18	M	acuosa	N	S	N	N	N	N	N	N	clamidia	
19	F	acuosa	S	N	N	N	N	N	N	N	No	
20	M	acuosa	N	N	N	N	N	N	N	N	No	
21	F	acuosa	N	N	N	N	S	N	N	N	clamidia	
22	F	acuosa	N	N	N	N	N	S	N	N	clamidia	
23	F	acuosa	S	N	N	N	N	N	S	N	clamidia	
24	F	acuosa	N	N	N	S	N	N	N	N	clamidia	

Fig. 18. Base de Hechos en Excel de la Clamidia.

Anexo 3

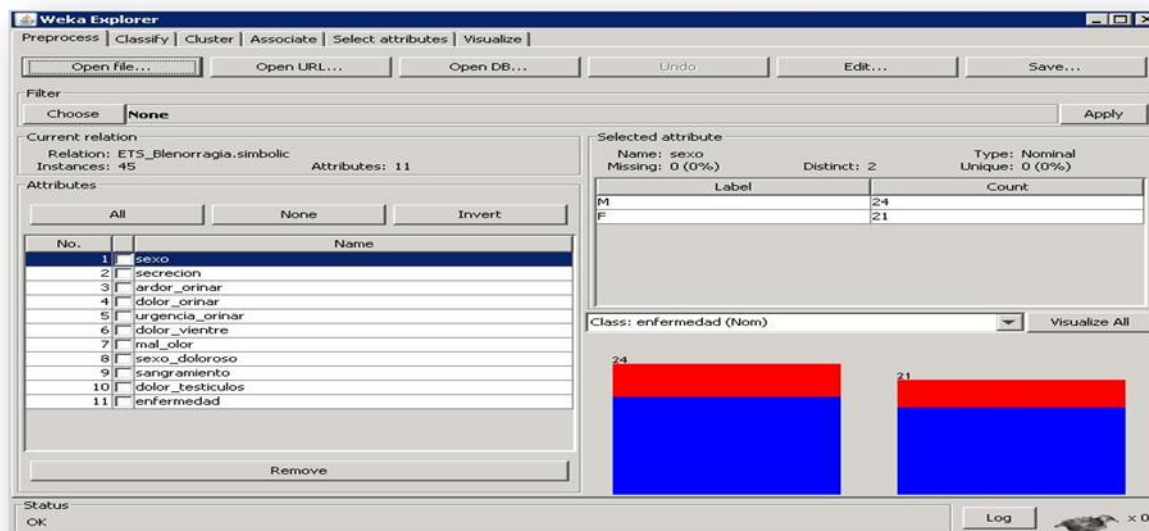


Fig. 18. Preprocesamiento en WEKA de la Base de Hechos de la Blenorragia.

Anexo 4

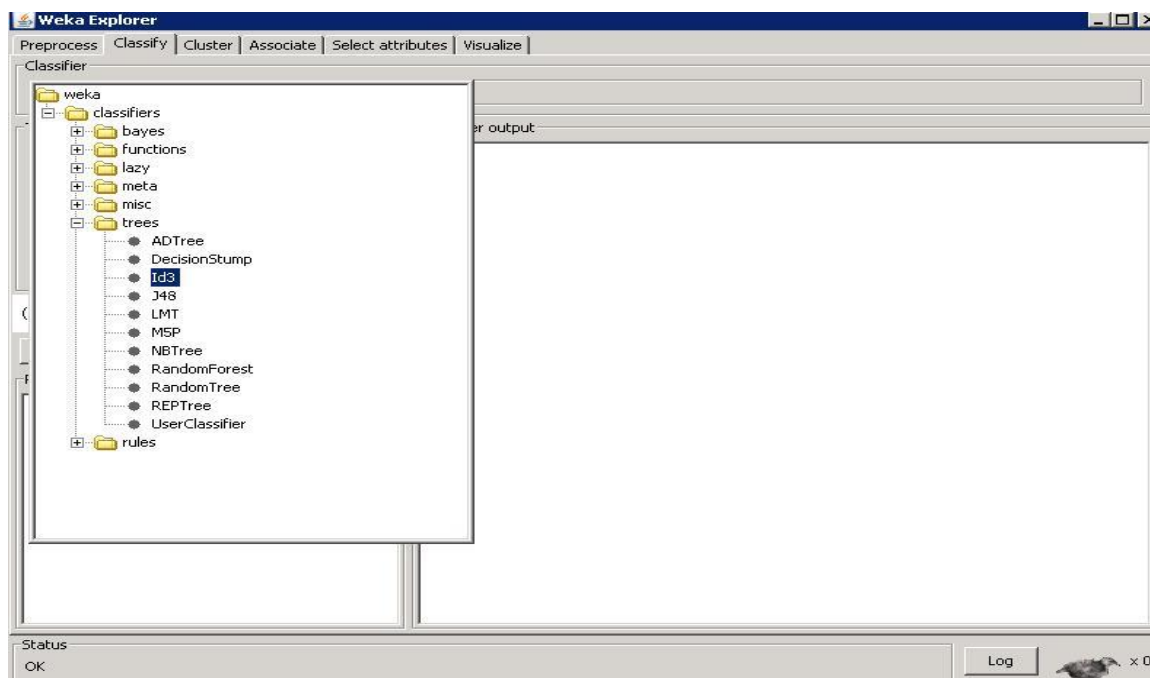


Fig. 19. Clasificación en WEKA.

Anexo 5

The screenshot shows the WEKA Explorer window with the 'Classify' tab selected. The 'Current relation' is 'ETS_Clamidia.symbolic' with 11 attributes and 37 instances. The 'Selected attribute' is 'sexo', which is a nominal attribute with 2 distinct values: 'M' (19 instances) and 'F' (18 instances). The interface includes a list of attributes, a filter section, and a visualization area with two stacked bar charts for 'M' and 'F'.

Label	Count
M	19
F	18

Fig. 20. Preprocesamiento en WEKA de la Base de Hechos de la Clamidia.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Capella, R. T. *Inteligencia Artificial en Medicina.Estado actual y perspectivas. Informática médica .*
- [2] *Sistemas Informativos Contables. Inteligencia Artificial. 2000.*
- [3] Béjar, Javier. *Inteligencia Artificial. Apuntes de Aprendizaje Automático. Curso 2005/2006.*
- [4] Guerra Hernández.Alejandro.*Aprendizaje.Clasificación.2004.*
- [5] Villena Román. Juan. *Inteligencia en Redes de Comunicaciones. Aprendizaje.*
- [6] *Guía para el uso de Árboles de Decisión. Alternativas de Uso de la Tierra para los Llanos Orientales de Colombia.*
- [7] López Takeyas .Bruno. *Algoritmo ID3.*
- [8] Béjar. Javier. *Árboles de Decisión. Aprendizaje Inductivo Supervisado. Curso 2006/2007.*
- [9] López Takeyas. Bruno. *Ingeniería en Sistemas Computacionales. Inteligencia Artificial. Algoritmo C4.5.*
- [10] Dürsteler. Juan. *Mapas de Kohonen.*
- [11] Aroche Reyes, Iirama.Suárez González, Dayanis. *La Teoría de la Incertidumbre aplicada a modelos de conocimiento. Habana .2008.*

BIBLIOGRAFÍA

Acuña, Edgar. *Clasificación usando Árboles de Decisión*.

Alonso González, Carlos J. *WEKA: Waikato Environment for Knowledge Analysis, Introducción Básica, Explorer*. Universidad de Valladolid.

Aroche Reyes, Iirama. Suárez González, Dayanis. *La Teoría de la Incertidumbre aplicada a modelos de conocimiento*. Habana .2008.

Barthelemy ,MsC .Jorge Haddad. *EL MÉTODO DELPHI Y LA GESTIÓN DE RECURSOS HUMANOS POR COMPETENCIA, UN NUEVO ENFOQUE*.

Béjar, Javier. *Aprendizaje Inductivo*. Curso 2007/2008.

Béjar, Javier. *Árboles de Decisión*. Curso 2006/2007.

Béjar, Javier. *Aprendizaje por Refuerzo*. Curso 2007/2008.

Cubero, J. L., Berzal, F., Herrera F. *Soft Computing y Sistemas Inteligentes, Fundamentos de Minería de Datos*. Universidad de Granada, España.

Dürsteler, Juan. *Mapas de Kohonen*.

Dúran ,Elena, Costaguta, Rossana. *Minería de datos para descubrir estilos de Aprendizaje*. Universidad Nacional de Santiago de Estero, Argentina.

Escribano Barreno, Julio. *Análisis de Datos Mediante Weka*.

López, Bruno. *Algoritmo ID3*.

Fernández, Fernando. *Taxonomía de los Algoritmos de Aprendizaje*.

García Jiménez , María, Álvarez Sierra, Aránzazu. *Análisis de datos en Weka-Pruebas de Selectividad*.
Universidad de Carlos III.

García Morate, Diego. *Manual de Weka*.

Gordon, T.J, Helmer ,Olaf, Dalkey, Norman. *Método Delphi*.

López Takeya, Bruno. *Ingeniería en Sistemas Computacionales, Algoritmo C4.5.2005*.

Pena Ribadas, Francisco José. *Aprendizaje Deductivo: Modelos de Razonamiento y Aprendizaje*.

Román, Julio. *Aprendizaje en Redes de Comunicaciones*.

Román, Julio, Sotelsek- Margalef, Anastasia. *MIDAS: An Information-Extraction Approach to Medical Text Classification*.

Santesteban Peña, Nayalys. *Modelación de una Base de Conocimientos. Aplicación de la Teoría de Errores*. 2007.

Witten, Ian H. Frank Eibe. *Weka Machine Learning Algorithms in Java*. *Universidad de Waikato Hamilton, New Zealand*.