

**FORMULACIÓN PARA TRATAMIENTOS ADECUADOS A LA FIBROSIS QUÍSTICA UTILIZANDO
ÁRBOLES DE DECISIÓN**

**MIGUEL ÁNGEL ACEVEDO FRANCO
JIMMY ANDRÉS LEÓN MUÑOZ**

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERIAS
INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
PEREIRA
2015**

**FORMULACIÓN PARA TRATAMIENTOS ADECUADOS A LA FIBROSIS QUÍSTICA UTILIZANDO
ÁRBOLES DE DECISIÓN**

**MIGUEL ÁNGEL ACEVEDO FRANCO
JIMMY ANDRÉS LEÓN MUÑOZ**

PROYECTO DE GRADO

DIRECTOR

Msc. GUILLERMO ROBERTO SOLARTE MARTINEZ

**UNIVERSIDAD TECNOLÓGICA DE PEREIRA
FACULTAD DE INGENIERIAS
INGENIERÍA DE SISTEMAS Y COMPUTACIÓN
PEREIRA
2015**

Nota de aceptación

Firma del presidente de jurado

Firma del jurado

Firma del jurado

Pereira, 13 de marzo de 2015

TABLA DE CONTENIDO

INTRODUCCIÓN.....	9
1. PLANTEAMIENTO DEL PROBLEMA	10
1.1. ANTECEDENTES DEL PROBLEMA	10
1.2. FORMULACIÓN DEL PROBLEMA	11
1.3. DESCRIPCIÓN DEL PROBLEMA	11
2. JUSTIFICACIÓN	12
3. OBJETIVOS	13
3.1. OBJETIVO GENERAL.....	13
3.2. OBJETIVOS ESPECÍFICOS.....	13
4. MARCO DE REFERENCIAL.....	14
4.1. MARCO DE ANTECEDENTES	14
4.2. MARCO CONCEPTUAL	15
4.3. MARCO TEÓRICO	16
5. ESTRUCTURA DE LA UNIDAD DE ANÁLISIS, CRITERIOS DE VALIDEZ Y CONFIABILIDAD	22
6. DISEÑO METODOLÓGICO	23
6.1. HIPÓTESIS.....	23
6.2. TIPO DE INVESTIGACIÓN	23
6.3. POBLACIÓN	23
6.4. MUESTRA	23
6.5. VARIABLES.....	24
7. DESARROLLO DE LA INVESTIGACIÓN	25
7.1. RECOLECCIÓN DE DATOS DE SÍNTOMAS Y TRATAMIENTOS ADECUADOS	25
7.1.1. OBTENCION DE INFORMACION.....	25
7.1.2. ANÁLISIS DE SÍNTOMAS Y TRATAMIENTOS	27
7.1.3. AGRUPACION DE SINTOMAS EXISTENTES.....	29
7.2. MODELAMIENTO DEL ÁRBOL DE DECISIÓN CON LOS SÍNTOMAS Y TRATAMIENTOS	31
7.2.1. MINERÍA DE DATOS	31
7.2.2. MODELO PREDICTIVO	32
7.2.3. ÁRBOLES DE DECISIÓN.....	32
7.2.3.1. ALGORITMO ID3.....	32
7.2.3.1. ALGORITMO C4.5	33
7.2.4. GANANCIA DE INFORMACIÓN.....	34
7.2.5. GENERACION DEL ÁRBOL	34
7.2.5.1. TABLAS DE CONTINGENCIA	34

7.2.5.2. PROPORCIÓN DE LA GANACIA DE INFORMACIÓN	37
7.2.5.3. ENSAMBLE DEL ÁRBOL.....	39
7.3. DESARROLLO DEL PROTOTIPO DEL SIMULADOR PARA EL FORMULADOR DE TRATAMIENTOS PARA LA FIBROSIS QUISTICA	60
7.3.1. PSEUDOCODIGO DEL ALGORITMO C4.5	60
7.3.2. DESARROLLO DEL CÓDIGO	61
7.4. RESULTADO DE PRUEBAS CON EL PROTOTIPO DEL SIMULADOR PARA LA FIBROSIS QUISTICA	67
7.4.1. RESULTADOS POR TRATAMIENTO	67
7.4.2. GENERACION DE REGLAS DE CLASIFICACIÓN	87
8. CONCLUSIONES	89
9. IMPACTOS ESPERADOS.....	90
10. DIVULGACIONES	91
11. BIBLIOGRAFÍA	92
12. ANEXOS	96
ANEXO A	96
ANEXO B	115
ANEXO C	132
ANEXO D	134

LISTA DE TABLAS

Tabla 1. Sustrato de tabla con la información de los pacientes	27
Tabla 2. Medicamentos y síntomas asociados.	29
Tabla 3. Agrupación de tratamientos	29
Tabla 4. Administración de tratamientos	35
Tabla 5. Respuestas de Atributo “Peso” en el Tratamiento 1.	38
Tabla 6. Respuestas de Atributo “Peso” Primer Iteración.	39
Tabla 7. Atributo “Tos con expectoración de color” Primer Iteración.....	40
Tabla 8. Atributo “Dolor Abdominal” Primer Iteración.	40
Tabla 9. Atributo “Deposiciones anormales” Primer Iteración.	40
Tabla 10. Atributo “Pseudomona aeruginosa” Primer Iteración.	41
Tabla 11. Atributo “Pseudomona aeruginosa Multi resistente” Primer Iteración.....	41
Tabla 12. Atributo “Estafilococo aerus” Primer Iteración.	42
Tabla 13. Atributo “Asperguelis aerus” en el Primer Iteración.....	42
Tabla 14. Ganancia de información Primer Iteración.	45
Tabla 15. Datos de entrenamiento Peso NORMAL.....	46
Tabla 16. Datos de entrenamiento Peso ALTO.....	46
Tabla 17. Datos de entrenamiento Peso BAJO	46
Tabla 18. Atributo “Tos con expectoración de color” Segunda Iteración.....	47
Tabla 19. Atributo “Dolor Abdominal” Segunda Iteración.	47
Tabla 20. Atributo “Deposiciones anormales” Segunda Iteración.....	48
Tabla 21. Atributo “Pseudomona aeruginosa” Segunda Iteración.	48
Tabla 22. Atributo “Pseudomona aeruginosa Multi resistente” Segunda Iteración.....	49
Tabla 23. Atributo “Estafilococo aerus” Segunda Iteración.....	49
Tabla 24. Atributo “Asperguelis aerus” Segunda Iteración.	49
Tabla 25. Ganancia de información Segunda Iteración.	52
Tabla 26. Datos de entrenamiento Tos con expectoración SI	53
Tabla 27. Datos de entrenamiento Tos con expectoración NO	53
Tabla 28. Atributo “Dolor Abdominal” Tercera Iteración.	54
Tabla 29. Atributo “Deposiciones anormales” Tercera Iteración.	54
Tabla 30. Atributo “Pseudomona aeruginosa” Tercera Iteración.....	54
Tabla 31. Atributo “Pseudomona aeruginosa Multi resistente” Tercera Iteración.	55
Tabla 32. Atributo “Estafilococo aerus” Tercera Iteración.	55
Tabla 33. Atributo “Asperguelis aerus” Tercera Iteración.	56
Tabla 34. Ganancia de información Tercera Iteración.....	58
Tabla 35. Datos de entrenamiento Pseudomona SI	59
Tabla 36. Datos de entrenamiento Pseudomona NO.....	59

Tabla 37. Tabla de ganancia tratamiento 1 primera iteración.	68
Tabla 38. Tabla de ganancias tratamiento 1 segunda iteración	70
Tabla 39. Tabla de ganancias tratamiento 1 tercera iteración.	71
Tabla 40. Tabla de ganancias tratamiento 2 primera iteración	74
Tabla 41. Tabla de ganancias tratamiento 2 segunda iteración.	76
Tabla 42. Tabla de ganancias tratamiento 3 primer iteración.	78
Tabla 43. Tabla de ganancias tratamiento 4 primer iteración	81
Tabla 44. Tablas de ganancia tratamiento 4 segunda iteración.	82
Tabla 45. Tabla de ganancias tratamiento 5 primer iteración.	85
Tabla 46. Tabla de ganancias tratamientos 5 segunda iteración	86

LISTA DE FIGURAS

Figura 1. Ejemplo árbol Binario	19
Figura 2. Ejemplo árboles n-arios	20
Figura 3. Selección de atributo más adecuado "Peso"	45
Figura 4. Resultado de los cálculos del nodo raíz en el árbol	47
Figura 5. Selección de atributo más adecuado "Tos Con expectoración"	52
Figura 6. Resultado de los cálculos segunda iteración	53
Figura 7. Selección de atributo más adecuado "Pseudomona"	58
Figura 8. Resultado de los cálculos Tercera iteración.....	59
Figura 9. Árbol de decisión Tratamiento 1.	73
Figura 10. Árbol de decisión Tratamiento 2.	77
Figura 11. Árbol decisión del Tratamiento 3.	79
Figura 12. Árbol de decisión Tratamiento 4.	83
Figura 13. Árbol de decisión Tratamiento 5.	87

RESUMEN

La minería de datos como una herramienta científica, ha sido de gran importancia en muchas áreas del conocimiento, entre ellas la medicina. Ha tomado mucha fuerza en las últimas décadas gracias a la gran capacidad de manejo de datos importantes contenidos en una gran base de datos sin importar el tipo de investigación. Los pacientes con fibrosis quística debido a que su enfermedad tiene un número variado de síntomas y tratamientos aplicables; se buscan a partir de esta tecnología mejorar la calidad de vida de los pacientes que la padecen. Haciendo uso de las nuevas técnicas de la inteligencia artificial, se pueden resolver esta clase de problemas. Se busca que a partir de los árboles de decisión se logre dar un tratamiento adecuado para la problemática de los pacientes que sufren esta enfermedad. Se le dará explicación al uso de esta técnica aplicándola a esta enfermedad y se desarrollara un prototipo de la aplicación para que formule los tratamientos adecuados a los pacientes.

SUMMARY

Data mining as a scientific tool has been of big importance in many areas of knowledge, including medicine. It has taken a lot of strength in recent decades thanks to the big capacity of management of important data in a database regardless of the type of research. Patients with cystic fibrosis because their illness has a number of different symptoms and treatments applicable; seeks from this technology improve the quality of life of patients who suffer it. Using new techniques of artificial intelligence can solve this kind of problems. It is intended that from decision trees is achieved to provide adequate treatment for the problems of patients who are suffering this disease. Explanation will be given to the use of this technique by applying it to this disease and a prototype implementation to formulate appropriate treatments to patients.

INTRODUCCIÓN

La tecnología de hoy en día ha llegado hasta los puntos más remotos de nuestro planeta, y ha llegado a ser tan útil y necesaria en el área de la salud, ya que no nos podríamos imaginar un sistema de salud sin el soporte e interconexión de las redes informáticas. El algoritmo c4.5 que se mostrará en el desarrollo del proyecto, nos dará una idea para poder dar un mejor tratamiento a gran cantidad de enfermedades.

Los árboles de decisión, nos dan una idea más clara de cómo resolver un problema. Hoy en día es un problema que siempre ha surgido en la historia del hombre, es la capacidad de equivocarnos, algo que no es malo, ya que aprendemos de nuestros errores, pero en cuestiones de tratamientos de pacientes, un error puede costar vidas o simplemente, no ayudar al mejoramiento de una persona.

En este caso, aprovecharemos el poder de la tecnología para orientar al usuario en la aplicación de un tratamiento para la fibrosis quística. Este algoritmo ya fue empleado para enfermedades cardiovasculares y buscamos aplicarlo en esta oportunidad a una enfermedad respiratoria.

1. PLANTEAMIENTO DEL PROBLEMA

1.1 ANTECEDENTES DEL PROBLEMA

Actualmente no existe una herramienta o un algoritmo que apoye en la formulación de tratamientos de la fibrosis quística, sin embargo, existen trabajos similares para otras enfermedades como la del profesor Guillermo Solarte Martínez (2011) de la Universidad Tecnológica de Pereira, con su trabajo “Evaluar la utilidad de la metodología de redes bayesianas en la predicción y diagnóstico de enfermedades (cardiovasculares)”²³ el cual a través de métodos de minería de datos, realiza un diagnóstico analizando un conjunto de síntomas para determinar una enfermedad cardiovascular.

Los tratamientos existentes para el tratamiento de la fibrosis quística que son usados hoy en día, son la kinesiterapia respiratoria, broncodilatadores, tratamiento antibiótico de la exacerbación, corticosteroides, mucolíticos, inmunomoduladores, asistencia ventilatoria no invasiva, entre otros; los cuales dependen de los diferentes síntomas que presente un paciente, haciendo más complejo la formulación de un tratamiento adecuado ya que dependen del criterio del médico y de los cálculos que tenga hacer para la formulación del mismo.

Los trabajos realizados acerca de esta enfermedad han sido en relación al diagnóstico, como el “test del sudor”:

Es el examen fundamental para la comprobación diagnóstica. Su solicitud debe ser hecha frente a la sospecha clínica, después del primer mes de vida del paciente. La técnica estándar y confirmatoria es la de Gibson y Cooke, que consiste en recolección del sudor inducida por iontoforesis con pilocarpina, midiendo el cloro con cloridómetro digital. Una técnica alternativa, de buen rendimiento, considerada de screening, es la medición del cloruro de sodio a través de conductividad²¹.

Y las investigaciones realizadas como la de la egresada Zuray Fernanda Corredor (2005), de la Universidad del Valle, en su tesis “Frecuencia de la mutación AF508 en portadores y enfermos con fibrosis quística del Sur-Occidente colombiano”⁴¹, la cual hace un estudio sobre la variación de los síntomas en los pacientes de la Fibrosis Quística comprobando la dificultad para la formulación de tratamientos de esta enfermedad.

²³ GUILLERMO SOLARTE MARTINEZ, “Evaluar la utilidad de la metodología de redes bayesianas en la predicción y diagnóstico de enfermedades (cardiovasculares)” Trabajo de grado Ingeniero de sistemas y computación. Pereira: Universidad Tecnológica de Pereira. Facultad de ingenierías. Departamento de Ingeniería de Sistemas y computación.

²¹ Grupo Técnico, Programa de Fibrosis Quística Unidad de Salud Respiratoria Subsecretaría de Redes Asistenciales Ministerio de Salud, “Programa nacional de Fibrosis Quística orientaciones programáticas para diagnóstico y tratamientos 2012”. Internet: <http://respiratorio.minsal.cl/PDF/FIBROSIS/ADULTO/Guia_Clinica_FQ_2012.pdf>

⁴¹ ZURAY FERNANDEZ CORREDOR, “Frecuencia de la mutación AF508 en portadores y enfermos con fibrosis quística del Sur-Occidente colombiano”. (Biología con mención en genética). Cali: Universidad del Valle. Facultad de ciencias.

1.2 FORMULACIÓN DEL PROBLEMA

La formulación para los tratamientos de la fibrosis quística son imprecisos y no existen herramientas computacionales que apoyen a la formulación de los tratamientos.

1.3 DESCRIPCIÓN DEL PROBLEMA

La fibrosis quística es una enfermedad crónica que puede atacar a todos los órganos del cuerpo, pero se manifiesta más seguido en el sistema respiratorio y digestivo, haciendo que un paciente con fibrosis quística muestre diferentes síntomas; consiguiendo que su tratamiento sea algo muy complejo de dictaminar para un médico, provocando la imprecisión de este. Además se le suma, que no existen herramientas computacionales que apoyen al médico, para que formule los tratamientos adecuados dependiendo de la manifestación de los síntomas.

2. JUSTIFICACIÓN

Actualmente la fibrosis quística es la enfermedad más crítica para la raza caucásica³⁸, buscar una solución que disminuya la complejidad e imprecisión de la formulación los tratamientos, ayudaría a aliviar la carga de los médicos en el momento formularlos, además la calidad de vida de los pacientes mejoraría, ya que no tendrían que estar pasando por diferentes tratamientos sin resultados, por culpa de un dictamen errado y disminuyendo los costos gastados en tratamientos infructuosos. Por otra parte, el desarrollo de una herramienta así, impulsaría a otros investigadores para que busquen soluciones parecidas con otras enfermedades que aquejan a la humanidad; utilizando las diferentes técnicas computacionales.

³⁸ THE AMERICAN CONGRESS OF OBSTETRICIANS AND GYNECOLOGIST, "La fibrosis quística: pruebas de detección y diagnóstico prenatal". Internet: < http://www.acog.org/For_Patients/Search_Patient_Education_Pamphlets_-_Spanish/Files/La_fibrosis_quistica-_Pruebas_de_deteccion_y_diagnostico_prenatal >

3. OBJETIVOS GENERALES Y ESPECÍFICOS

3.1 Objetivo general:

Desarrollar un prototipo de un sistema experto, usando árboles de decisión, haciendo uso del algoritmo c4.5 y la minería de datos, para hacer la formulación de un tratamiento adecuado a la fibrosis quística.

3.2 Objetivos específicos:

- Identificar los síntomas y tratamientos de la fibrosis quística que son dependientes de la enfermedad.
- Modelar el árbol de decisión para la formulación de tratamientos de la fibrosis quística a partir los síntomas manifestados.
- Desarrollar el prototipo de la herramienta a partir del modelo de árbol de decisión para la formulación de tratamientos de la fibrosis quística.
- Realizar pruebas de la herramienta desarrollada para la formulación de tratamientos; usando la información obtenida de los pacientes que sufren esta enfermedad y evaluar la precisión de la herramienta con resultados mostrados.

4. MARCO REFERENCIAL

4.1 MARCO DE ANTECEDENTES

Con el desarrollo de nuevas tecnologías y el avance de la minería de datos, se pueden realizar grandes aplicaciones en los diferentes campos de investigación. Como lo propone el profesor Guillermo Solarte Martínez (2011) de la Universidad Tecnológica de Pereira, a través de su tesis “Evaluar la utilidad de la metodología de redes bayesianas en la predicción y diagnóstico de enfermedades (cardiovasculares)”²³ que demuestra que la utilidad de la minería de datos utilizando la técnica de redes bayesianas, además de evaluar la utilidad de la metodología bayesiana en la predicción y diagnóstico médico de enfermedades complejas (Cardiovasculares), a identificar conjuntos representativos de patrones que nos ayudan a la predicción o diagnóstico médico temprano. Otros trabajos relacionados a la utilización de los árboles de decisión lo realiza el ingeniero Reinel Aráis Montoya (2010) con su tesis “Detección temprana de fallas en la red de internet banda ancha aplicando minería de datos”³² en el cual expone como aplicar la técnica y desarrolla un software que se aplicará en la investigación de la causa de fallos de red o quejas de los usuarios de una red de banda ancha para acceso a Internet. Igualmente en este documento se realiza una descripción acerca de que es minería de datos, los medios de almacenamiento, métodos de clasificación, técnicas, herramientas y algoritmos más representativos de la minería de datos.

En relación con la problemática de la fibrosis quística, el cual es una enfermedad genética multisistémica, la egresada Zuray Fernanda Corredor (2005) de la Universidad del Valle con su tesis “Frecuencia de la mutación AF508 en portadores y enfermos con fibrosis quística del Sur-Occidente colombiano”⁴⁰ nos presenta la problemática que tiene los portadores de esta enfermedad en el sur occidente del país; y el especialista en biología molecular y biotecnología Rene Rodríguez Marín (2000) realizó la siguiente investigación “Determinación de incidencia de la mutación AF508 del gen CFTR, asociado a la fibrosis quística en la población del departamento de Risaralda”³³ exponiendo la problemática de los pacientes en la región de Risaralda.

²³ GUILLERMO SOLARTE MARTINEZ, “Evaluar la utilidad de la metodología de redes bayesianas en la predicción y diagnóstico de enfermedades (cardiovasculares)” Trabajo de grado Ingeniero de sistemas y computación. Pereira: Universidad Tecnológica de Pereira. Facultad de ingenierías. Departamento de Ingeniería de Sistemas y computación.

³² REINEL ARÁIS MONTOYA, “Detección temprana de fallas en la red de internet banda ancha aplicando minería de datos” Trabajo de grado Ingeniero de sistemas y computación. Pereira: Universidad Tecnológica de Pereira. Facultad de ingenierías. Departamento de Ingeniería de Sistemas y computación.

⁴⁰ ZURAY FERNANDEZ CORREDOR, “Frecuencia de la mutación AF508 en portadores y enfermos con fibrosis quística del Sur-Occidente colombiano”. Tesis (Biología con mención en genética). Cali: Universidad del Valle. Facultad de ciencias.

³³ RENE RODRÍGUEZ MARÍN, “Determinación de incidencia de la mutación AF508 del gen CFTR, asociado a la fibrosis quística en la población del departamento de Risaralda” Tesis especialización en biología molecular y biotecnología. Pereira: Universidad Tecnológica de Pereira. Facultad de Medicina.

4.2 MARCO CONCEPTUAL

Árboles de decisión³⁹: Técnica que permite analizar decisiones secuenciales basada en el uso de resultados y probabilidades asociadas.

Los árboles de decisión se pueden usar para generar sistemas expertos, búsquedas binarias y árboles de juegos

Minería de datos³¹: La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos.

Entropía⁵: La entropía también se puede considerar como la cantidad de información promedio que contienen los símbolos usados. Los símbolos con menor probabilidad son los que aportan mayor información; por ejemplo, si se considera como sistema de símbolos a las palabras en un texto, palabras frecuentes como "que", "el", "a" aportan poca información, mientras que palabras menos frecuentes como "corren", "niño", "perro" aportan más información. Si de un texto dado borramos un "que", seguramente no afectará a la comprensión y se sobreentenderá, no siendo así si borramos la palabra "niño" del mismo texto original. Cuando todos los símbolos son igualmente probables (distribución de probabilidad plana), todos aportan información relevante y la entropía es máxima.

El concepto de entropía es usado en termodinámica, mecánica estadística y teoría de la información. En todos los casos la entropía se concibe como una "medida del desorden" o la "peculiaridad de ciertas combinaciones". La entropía puede ser considerada como una medida de la incertidumbre y de la información necesaria para, en cualquier proceso, poder acotar, reducir o eliminar la incertidumbre. Resulta que el concepto de información y el de entropía están ampliamente relacionados entre sí, aunque se necesitaron años de desarrollo de la mecánica estadística y de la teoría de la información antes de que esto fuera percibido.

Sistema experto²⁴: Los Sistemas Expertos, rama de la Inteligencia Artificial, son sistemas informáticos que simulan el proceso de aprendizaje, de memorización, de razonamiento, de comunicación y de acción en consecuencia de un experto humano en cualquier rama de la ciencia.

Estas características le permiten almacenar datos y conocimiento, sacar conclusiones lógicas, tomar decisiones, aprender de la experiencia y los datos existentes, comunicarse con expertos humanos, explicar el porqué de las decisiones tomadas y realizar acciones como consecuencia de todo lo anterior.

³⁹ UHU, Árboles de decisión [en línea]

<<http://www.utm.mx/~jahdezp/archivos%20estructuras/DESICION.pdf>>

³⁰ MICROSOFT, Minería de datos [en línea]

<<http://msdn.microsoft.com/es-es/library/ms174949.aspx>>

⁵ Cuevas Agustín, Gonzalo, "Teoría de la información, codificación y lenguajes", Ed. SEPA (Sociedad para Estudios Pedagógicos Argentinos), Serie Informática 1986

²⁴ INFORMATICA INTEGRAL INTELIGENTE, Sistemas expertos [en línea]

<<http://www.informaticaintegral.net/sisexp.html>

Técnicamente un sistema experto, contiene una base de conocimientos que incluye la experiencia acumulada de expertos humanos y un conjunto de reglas para aplicar ésta base de conocimientos en una situación particular que se le indica al programa. Cada vez el sistema se mejora con adiciones a la base de conocimientos o al conjunto de reglas.

Fibrosis quística²⁹: Es una enfermedad hereditaria que provoca la acumulación de moco espeso y pegajoso en los pulmones, el tubo digestivo y otras áreas del cuerpo. Es uno de los tipos de enfermedad pulmonar crónica más común en niños y adultos jóvenes, y es un trastorno potencialmente mortal.

La fibrosis quística (FQ) es causada por un gen defectuoso que lleva al cuerpo a producir un líquido anormalmente espeso y pegajoso llamado moco. Este moco se acumula en las vías respiratorias de los pulmones y en el páncreas, el órgano que ayuda a descomponer y absorber los alimentos.

Esta acumulación de moco pegajoso ocasiona infecciones pulmonares potencialmente mortales y serios problemas digestivos. Esta enfermedad también puede afectar las glándulas sudoríparas y el aparato reproductor masculino.

4.3 MARCO TEÓRICO

Para comenzar con la teoría y dar solución al problema planteado en el proyecto debemos iniciar con la teoría de la decisión en cual se plantea lo siguiente:

Teoría de la decisión ⁴⁰

En la vida real, y tanto en el ámbito profesional como el personal, nos vemos enfrentados a multitud de situaciones en las que tenemos que decidir entre varias alternativas. La propia optimización no es más que una forma de tomar una decisión entre unas alternativas factibles.

Así, en su dimensión más básica, un proceso de toma de decisión puede entenderse como la elección de lo “mejor” entre lo “posible”. Ahora bien, según se defina qué es lo mejor y qué es lo posible nos enfrentaremos a distintas situaciones de decisión.

La optimización clásica tiene como característica general que lo mejor, el objetivo, es único y está claramente determinado (excepto en optimización multi objetivo) y que lo posible, las soluciones factibles, no vienen expresadas explícitamente sino en forma de restricciones y sin incertidumbre (excepto en optimización estocástica, que no es precisamente clásica). Pero, además de estos contextos de decisión de optimización clásica, existen otros que configuran lo que se suele denominar en términos amplios la teoría de la decisión. Tres grandes bloques son los que se suelen abordar en este análisis:

²⁹ MEDLINEPLUS, fibrosis quística [en línea]

<<http://www.nlm.nih.gov/medlineplus/spanish/ency/article/000107.htm>> [citado el 25 de octubre 2013]

⁴⁰ UNIVERSIDAD COMPLUTENSE DE MADRID, Teoría de la decisión [en línea]

<http://www.mat.ucm.es/~bvitoria/Archivos/a_dt_UCM.pdf> [citado el 29 de octubre 2013]

- a) La teoría de la decisión con incertidumbre o riesgo, en la que se analiza la toma de decisiones con aleatoriedad o incertidumbre en los resultados, de modo que las consecuencias de una decisión no están determinadas de antemano, sino que están sujetas al azar.
- b) La decisión multi criterio, en la que si bien dada una decisión sus consecuencias están perfectamente determinadas, lo que no está definido tan claramente es qué es lo mejor, existiendo varios objetivos en conflicto.
- c) La teoría de juegos, en la que las consecuencias de una decisión no dependen únicamente de la decisión adoptada, sino, también de la que elijan otros jugadores. En este contexto, los problemas de decisión con aleatoriedad del bloque anterior suelen ser denominados juegos frente a la naturaleza.

Entendiendo esta teoría podemos concluir para desarrollar una estrategia óptima cuando tomamos decisiones nos enfrentamos a varias alternativas de decisión y una incertidumbre o patrón de eventos futuros lleno de riesgos. Y la forma en la que atacaremos el problema para visualizar y organizar los cálculos que deben realizarse es con un árbol de decisiones.

Árbol de decisión ¹¹

El árbol de decisión es un gráfico que nos sirven como herramienta para la toma de decisiones en la empresa. Plantean el problema para que todas las opciones sean analizadas, y hace posible analizar las consecuencias de adoptar una u otra decisión. También nos permite cuantificar su coste y las probabilidades de ocurrencia de cada decisión.

Pueden aplicarse en muchas situaciones de la empresa a la hora de la toma de decisiones, como en inversión, reinversión, políticas de créditos y financiación a corto y largo plazo.

Al empresario en muchas ocasiones se le plantea la elección de una opción entre varias posibilidades, por lo que recurre a ésta herramienta. La elección de una alternativa supone el abandono de las demás opciones. Al tomar una decisión, el resultado de ésta vendrá determinado por un suceso incierto (estados de la naturaleza). Una vez producido ese estado de naturaleza, es posible elegir de nuevo entre distintas alternativas que dependen a su vez de nuevos estados de naturaleza.

Los elementos fundamentales en la elaboración de un árbol de decisión son:

- Puntos o nodos de decisión entre alternativas o estrategias; indica que una decisión necesita tomarse en ese punto de proceso.

¹¹ EMILIO SORIA, ANTONIO JOSÉ SERRANO Y JOSÉ DAVID MARTÍN "Arboles de decisión". Dpto. Ingeniería Electrónica, Internet:< http://ocw.uv.es/ciencias-de-la-salud/pruebas/1-2/1tema_6_ocw.pdf>

- Nudos aleatorios o de probabilidad: Ocurrencia de los posibles estados de la naturaleza
- Resultados esperados

Partimos del primer nudo siempre, que es la decisión inicial que debe de tomar el decisor, y a partir de ahí van saliendo las diferentes alternativas planteadas.

En el árbol de decisión existen dos clases de elementos:

- Arcos o Ramas: Punto de selección entre diferentes alternativas. Su representación gráfica es una *flecha* nos muestra los distintos caminos que se pueden emprender cuando tomamos una decisión (línea continua) o bien ocurre algún evento aleatorio (línea continua).

- Nudos o Vértices: existen dos tipos de vértices representados con notación diferente:

- a) Puntos de decisión que representan las opciones a adoptar. Se representan mediante un *cuadrado*.
- b) Sucesos inciertos, considera el entorno económico-financiero, y están representados por un *círculo*.

Una vez realizado el árbol de decisiones podemos tener una idea más clara del problema y construir un árbol binario para tratar de dar soluciones a nuestro problema planteado.

Árboles binarios ³¹

Un árbol es una estructura de datos con nodos enlazados en forma jerárquica y orientada. Es una estructura ordenada en que los hijos de un nodo generalmente tienen un valor menor que este y están ordenados de izquierda a derecha.

La raíz es el punto de entrada a la estructura. La raíz puede tener cero o más nodos descendientes desde ella.

El conjunto de estos nodos forman subárboles de la raíz. La raíz es el ancestro de estos subárboles.

Los nodos sin descendientes se llaman hojas. Los nodos internos son todos los nodos menos las hojas.

Una trayectoria del nodo n_i al nodo n_k , es una secuencia de nodos desde n_i hasta n_k , tal que n_i es el padre de n_{i+1} .

Existe un solo enlace (link) entre un padre y cada uno de sus hijos. El largo de una trayectoria es el número de enlaces en la trayectoria.

³¹ PROFESORES, arboles binarios [en línea]

<<http://profesores.elo.utfsm.cl/~tarredondo/info/datos-algoritmos/ELO-320%20Arboles%20binarios.pdf>>

Una trayectoria de k nodos tiene largo $k-1$. La altura de un nodo es el largo de la trayectoria más larga de ese nodo a una hoja.

La profundidad de un nodo es el largo de la trayectoria desde la raíz a ese nodo. La profundidad del árbol es la profundidad de la hoja más profunda.

Nodos a una misma profundidad están al mismo nivel.

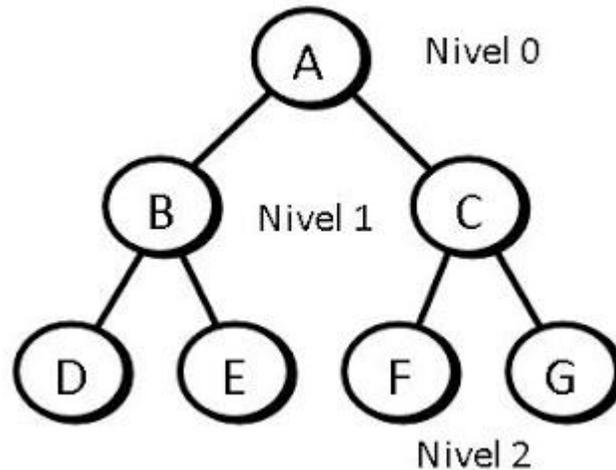


Figura 1. Ejemplo árbol Binario

Fuente: <http://geoacostac.blogspot.com/2007/08/arboles-concepto-y-definiciones-en.html>

Árboles n-arios

Los árboles n -arios son estructuras recursivas, en la cual cada nodo tiene un número cualquiera de nodos asociados. Estos nodos que forman árboles n -arios corresponden a la generalización de los árboles binarios. La diferencia entre estos radica en que los árboles n -arios puede manejar múltiples subárboles asociados a cada elemento, y no solamente dos, como es en caso de la estructura de árboles binarios.

Los conceptos que se manejan en árboles binarios como nodo padre, hijo, hermano, nivel, altura, peso, camino, etc. También son manejados en los árboles n -arios

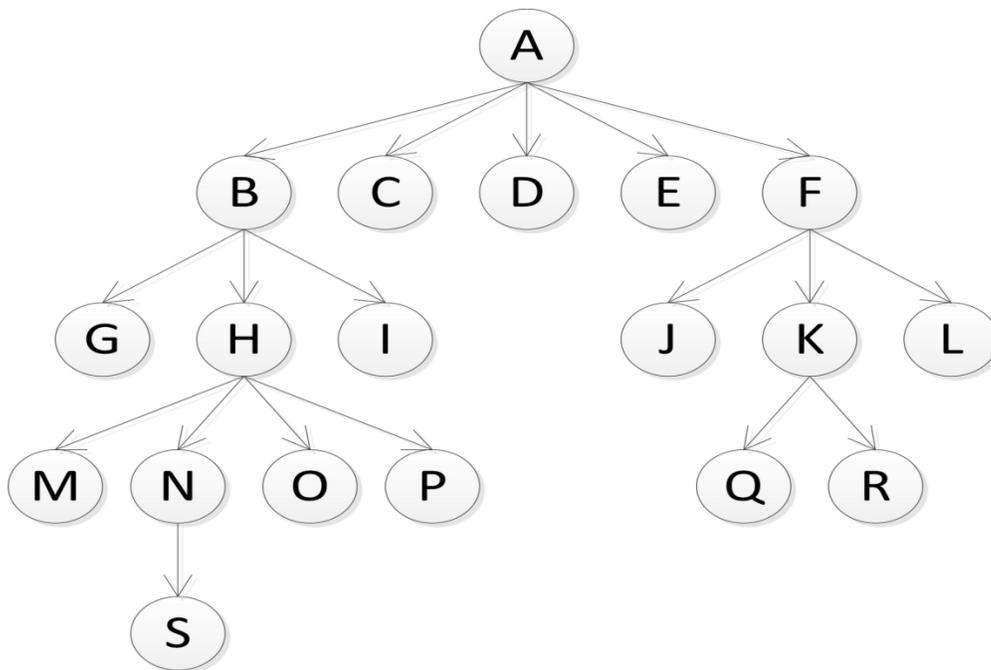


Figura 2. Ejemplo Árboles n-arios

Fuente: <http://zomwi.blogspot.com/2012/05/convertir-un-arbol-n-ario-binario.html>

Síntomas de la fibrosis quística ⁷

Existen signos indicativos que, a edad temprana, pueden inducir a pensar que nos encontramos frente a una afectación de fibrosis quística. Estos signos pueden ser:

- Problemas respiratorios frecuentes o crónicos (fácilmente confundidos con bronquitis, neumonías o patologías similares).
- Tos.
- Fiebres altas.
- Cansancio
- Dolor de vientre.
- Heces grasas y malolientes.
- Bajo peso y complexión débil.
- Casos de muertes tempranas o enfermedades pulmonares en los ascendentes familiares.

La fibrosis quística es una enfermedad multi sistémica, es decir, que se manifiesta en diferentes sistemas y aparatos del cuerpo humano.

⁷ DMEDICINA, Síntomas de la fibrosis quística

<<http://www.dmedicina.com/enfermedades/respiratorias/fibrosis-quistica>>

Los efectos más destacables son:

- Sinusitis.
- Pólipos nasales.
- Bronquitis.
- Bronquiolitis.
- Bronquiectasia.
- Tos.
- Malabsorción.
- Diabetes.
- Íleo mecánico o equivalente.
- Prolapso rectal.
- Pancreatitis.
- Cirrosis hepática.
- Colelitiasis.
- Acropaquia.

5. ESTRUCTURA DE LA UNIDAD DE ANÁLISIS, CRITERIOS DE VALIDEZ Y CONFIABILIDAD

- **UNIDAD DE ANÁLISIS:** Pacientes que sufren fibrosis quística ubicados en la ciudad de Pereira.
- **ESTRATEGIA DE MEDICIÓN:** Como el prototipo de la herramienta de árboles de decisión, será desarrollada para formular tratamientos a las personas que sufren fibrosis quística, se ingresarán un conjunto de síntomas el cual presente el paciente y dependiendo de estos, la herramienta determina que tratamientos son los más adecuados.
- **CRITERIOS DE VALIDEZ Y CONFIABILIDAD:** Los tratamientos obtenidos con la herramienta son de carácter cuantitativo, lo que nos permite comparar los resultados con casos clínicos similares que ya tenido éxito. A partir de esto la herramienta de formulación, mostrará su validez cuando se realicen pruebas con distintos casos clínicos y se haga un análisis comparativo de los datos obtenidos. Con este seguimiento se realiza una validez de contenido, ya que se estará midiendo que tan precisos son los tratamientos al momento de formularlos. La validez de constructo, ya que con esto se demostrarán las teorías que buscan formular tratamientos más precisos dependiendo de los síntomas del paciente.

El nivel de precisión de la herramienta será muy óptimo ya que los síntomas para analizar son de tipo cuantitativo, lo que permite obtener los mismos tratamientos para el mismo caso clínico, y gracias a la herramienta nos permite hacer cálculos más precisos a la hora de manejar cálculos matemáticos.

6. DISEÑO METODOLÓGICO

6.1 HIPÓTESIS

“Se puede disminuir, la imprecisión en la formulación de los tratamientos de la fibrosis quística; desarrollando una herramienta que utilice la técnica de árboles de decisión.”

6.2 TIPO DE INVESTIGACIÓN

El proyecto planteado, implementará tecnologías informáticas, la cual en este caso, usaremos la técnica de árboles de decisión para disminuir la imprecisión de los médicos en el momento de formular tratamientos, por lo que nuestro tipo de investigación es cuantitativa.

6.3 POBLACIÓN

La población al cual estará dirigido el proyecto, son las personas que padecen de fibrosis quística, en la ciudad de Pereira y según un estudio de la revista de neumología pediátrica en el año 2003 14.000 personas aproximadamente sufren este trastorno en Pereira⁸; estos datos se están corroborando con la facultad de medicina.

6.4 MUESTRA

Se utilizara el método de muestreo probabilístico “Muestreo aleatorio simple”¹⁰, como método para la selección de la muestra, utilizando la distribución normal estándar.

Para definir el tamaño de la muestra se utiliza la fórmula, teniendo en cuenta que nuestra población es finita.

$$n = \frac{N\sigma^2Z^2}{(N-1)e^2 + \sigma^2Z^2}$$

Dónde:

3158.75/113.617525

N = 14.000 (población de enfermos).

σ = 0.5 (varianza estándar).

⁸ DRA. CATALINA VÁSQUEZ, DR. RICARDO ARISTIZÁBAL, DR. WILSON DAZA, “Fibrosis quística en Colombia”. Internet: <<http://www.neumologia-pediatria.cl>>

¹⁰ EDGAR LEONEL GOMEZ NARCISO, “Estadística, Matemática y Computación”. Universidad rural de Guatemala maestría en investigación y proyectos. Internet: <<http://reyesestadistica.blogspot.com/2011/07/muestreo-simple-aleatorio.html>>

$Z = 0.95$ (nivel de confianza estándar).

$e = 0.09$ (error estándar).

n = tamaño de la muestra.

$$n = \frac{(14.000)(0,5)^2(0,95)^2}{(14.000-1)(0,09)^2+(0,5)^2(0,95)^2} = 27,80$$

Redondeando la cifra, la muestra total sería de 28 personas.

6.5 VARIABLES

Las variables son:

- Tiempo de ejecución: tiempo que se demora la herramienta en formular un tratamiento.
- Cantidad de tratamientos: según los síntomas la herramienta puede dictaminar uno o más tratamientos.
- Precisión de los tratamientos encontrados: según el caso clínico los tratamientos deberán ser efectivos en el paciente.

6.6 RECOLECCIÓN DE INFORMACIÓN

Para tomar los datos que usará la herramienta, se deberá usar el caso clínico realizado por el médico. En estos, se tomarán los síntomas que tiene el paciente para que la herramienta analice sus datos y formule un tratamiento.

Formato de caso clínico²

Casos Clínicos (Mínimo 10 máximo 20 páginas)

- Presentación caso
- Antecedentes personales, familiares, mórbidos, contextuales (escuela, trabajo, red de amigos, conocidos, colegas)
- Historia de la Consulta
- Motivo de Consulta
- Hipótesis DCA o de trabajo
- Foco, objetivos, lineamientos del trabajo
- Intervenciones realizadas
- Vínculo terapéutico y paciente del terapeuta
- Efectos e implicancias de las intervenciones
- Estado del proceso (alta, deserción, derivación, otro)
- Correlato teórico - clínico (teoría de la técnica)

²APALEX, "normas de presentación de caso clínico", Internet:

http://www.apalex.es/uploads/documentos/NORMAS_PRESENTACION_CASOS_CLINICOS_apalex.pdf

7. DESARROLLO DE LA INVESTIGACIÓN

7.1 CAPÍTULO 1: RECOLECCIÓN DE DATOS DE SINTOMAS Y TRATAMIENTOS ADECUADOS

7.1.1 Obtención de Información

Para la recolección de información sobre la fibrosis quística, al ser una enfermedad que compromete los sistemas pulmonar, digestivo entre otros; y el cual se manifiesta desde corta edad, no se puede obtener la información de los pacientes de forma corriente por lo especial de la enfermedad, y se requiere de un especialista que trate esta enfermedad en la ciudad de Pereira.

Este experto debía ser un neumólogo pediatra, el cual no posee actualmente la facultad de medicina de la Universidad Tecnológica de Pereira por lo tanto se hace contacto con una neumóloga externa recomendada por la facultad de medicina. La pediatra neumóloga Bertha Inés Agudelo Vega antes docente de la facultad de medicina, que ahora trabaja independiente y gracias a su gran trayectoria. Ella ha tratado con pacientes que padecen esta enfermedad desde hace mucho tiempo.

La doctora Bertha, en un gesto de buena voluntad y colaboración hacia el proyecto, ha proporcionado material de trabajo y suministrado las historias clínicas de los pacientes que padecen fibrosis quística. Estas historias clínicas antes de ser entregadas por la doctora Bertha son eliminadas los nombres y datos personales de los pacientes para no tener inconvenientes legales.

La información contenida en las historias clínicas es acerca de los síntomas que presentó el paciente, el examen físico que realiza el médico al paciente, el diagnóstico que el médico dictamina basado en los síntomas y el examen físico presentados, la fórmula con los medicamentos recetada por el doctor y si es necesario los exámenes y estudios que se debe realizar para una mejor valoración.

Se realiza una tabla con la información recogida en las historias clínicas generando los siguientes registros número del paciente, edad, peso, talla, IMC (índice de masa corporal), antecedentes de medicamentos, síntomas, examen físico, impresión diagnóstica, fórmula médica.

Número del paciente: como no se manejan nombres de pacientes se usa una identificación genérica (paciente 1).

Edad: la mayoría de pacientes tratados por la doctora Bertha son menores de edad entre los 4 y 18 años de edad.

Peso: peso del paciente.

Talla: estatura que tiene el paciente.

IMC: es una medida de asociación entre el peso y la talla del paciente para determinar si su peso es adecuado.

Antecedentes de medicamentos: son los medicamentos y tratamientos que se ha realizado o que actualmente se está realizando.

Síntomas: son las manifestaciones que tienen los pacientes el cual por ser fibrosis quística puede presentar diferentes en un solo paciente.

Examen físico: las alteraciones palpables que tiene el paciente.

Impresión diagnóstica: ya que es fibrosis quística solo se mira cual es el tipo de manifestación que está presentando el paciente.

Fórmula médica: son los diferentes medicamentos que se le formula al paciente.

N° paciente	Edad	Peso (kg)	Talla (ms)	IMC	Antecedentes de medicamentos	Síntomas	Examen físico	Fórmula medica
1	9	22	122	14,78	* dornasa alfa *lipasa adoside 25000 *pediasure *vitaminas a,d,e,k *iontoforesis	*dolor abdominal *deposiciones poco formadas *fibrosis hepáticas *estafilococo	* hipocratismos digital * lengua geográfica * otoscopia anormal * auscultación cardiopulmonar *hernia umbilical pequeña	* Acido ursodesoxicólico X 50 MG tabletas #60 tomar una al día por dos meses *Ranitidina x 150 MG tabletas #60 tomar una cada 12 horas
2	9	24,4	130	14,44	*clindamician *ranitidina *vitaminas a,d,e,k *ciprofloxacina	*tos seca *sobreinfección por pseudomonas aeruginosa y estafilococo	* Hipertrofia severa de los cornetes *cerumen impide ver tímpano izq.	*Trimetorim x 240MG suspensión frasco #4 tomar 14cc cada 12 horas por 21 días *ranitidina x 150 MG tabletas #270 tomar una cada día por 3 meses *tobramicina polvo para inhalación en capsula dura (TOBY PODHALER) x 28MG #224 para inhalar cada 12 horas por 28 días

3	9	25	133	14,13	* lipasa pancreática *pediasure *nebulizaciones tobramicina	*sin tos *deposiciones normales *sobreinfección pseudomona y estilococo controlada	* Hipertrofia severa de los cornetes	* sales de rehidratación oral x 45 Meq frasco #90 tomar un frasco al día por 3 meses
4	14	42,7	142	21,18	* cefatizidima *atreonam *amikacina *toma antibioticos para sinusitis *tres cirugias por sinusitis	*sinusitis severa *pseudomona	*a febril no tiene *rinoscopia con buena permeabilidad *otoscopia normal *auscultación pulmonar normal *abdomen bien	* aztreonam *rinofluimicil *tobramicina nebulizada
5	14	42,5	142	21,18	*mereonem *amikacina *cefepime *tobramician	*tos *expectoración de color verde *sinusitis refractaria al tratamiento	* hipertrofia severa de los cornetes *edema de la mucosa nasal *escurrimiento posterior	* aztreonam *klaritromicina x 500 MG tabletas #40 tomar una cada 12 horas *tobramicina nebulizada

Tabla 1. Sustrato de tabla con la información de los pacientes

Fuente: Autores

7.1.2 Análisis de síntomas y tratamientos

Gracias a la gran cantidad de información obtenida en las historias clínicas, se encuentra que muchos datos no son necesarios para realizar un diagnóstico y con la ayuda de la doctora Bertha se realiza una identificación para saber cuáles son estos datos. Una vez hecha la identificación de estos síntomas y tratamientos que no son necesarios para el diagnóstico se realiza una tabla de relación con los medicamentos y sus síntomas asociados.

Tratamientos	Síntomas asociados
ácido ursodexicolico	* dolor abdominal * deposiciones blandas o mal formadas
lipasa pancreática	* dolor abdominal * pérdida de apetito * bajo peso * prolapso rectal *diarrea *deposiciones blandas, muy duras, fétidas, mal formadas
Trimetroprim	*tos seca *expectoración de color *psudomona *estafilococo aerus *H. influenzae
Tobramicina	*tos seca *expectoración de color *psudomona multiresistente
Aztreonam	*tos seca *expectoración de color *psudomona
amoxicilina+acido clavulanico	*tos seca *expectoración de color *estafilococo aerus *H. influenzae
Levofloxacina	*tos seca *expectoración de color *psudomona multiresistente *estafilococo aerus multiresistente
Colastina	*tos seca *expectoración de color *estafilococo aerus multiresistente
Ciprofloxacina	*tos seca *espectoracion de color *psudomona multiresistente
Azitromicina	*tos seca *espectoracion de color *psudomona multiresistente *spergelis aerus *bronquilitis obliterante

dornasa alfa	*tos *espectoracion *bronquiactasis
Amikacina	*tos seca *expectoración de color *psudomona multiresistente
Cefepime	*tos seca *expectoración de color *psudomona multiresistente

Tabla 2. Medicamentos y síntomas asociados.
Fuente: Autores

7.1.3 Agrupación de tratamientos existentes

Se encontró con este análisis que existen diferentes tratamientos dependiendo del tipo de infección pulmonar o si el paciente tiene compromiso digestivo pero se logró agrupar los diferentes medicamentos en cinco tratamientos que se usaran para emitir la formulación adecuada.

Tratamiento preventivo	Pseudomona	P. multiresistente	Estafilococo+psudomona+aspergilus	Complicaciones digestivas
tratamiento 1	tratamiento 2	tratamiento 3	tratamineto 4	tratamineto 5
dornasa alfa lipasa pancreatica vitaminas azitromicina Terapia respiratoria	dornasa alfa trimetroprim aztreonam tobramicina solucion salina Terapia respiratoria	dornasa alfa tobramicina levofloxa amikacina azitromicina cefepime ciprofloxa colistina piperacilina-tazobactam Terapia respiratoria	dornasa alfa trimetroprim tobramicina amoxicilina+acido clavulanico azitromicina vancomicina Terapia respiratoria	lipasa pancreatica ácido ursodesoxicolico vitamina K Terapia respiratoria

Tabla 3. Agrupación de tratamientos
Fuente: Autores

Tratamiento 1:

La fibrosis quística es una enfermedad que produce una gran acumulación de moco espeso y pegajoso en los pulmones, el tubo digestivo, entre otras áreas del cuerpo; el cual genera complicaciones en estos sistemas. El objetivo de este tratamiento es reducir la cantidad de moco acumulado para evitar estas complicaciones, utilizando medicamentos como la dornasa alfa que disuelve las secreciones espesas y la lipasa pancreática, la cual es una enzima digestiva que ayuda a reducir la acumulación de secreciones en el páncreas. La terapia respiratoria es indispensable para los pacientes con fibrosis quística, por lo cual, va incluida para todos los tratamientos. El objetivo de esta terapia es mantener las vías respiratorias despejadas. La azitromicina elimina posibles infecciones las cuales pueden surgir por la acumulación de moco en las vías aéreas, por último, las vitaminas son indispensables para estos pacientes ya que estos casi siempre rozan niveles por debajo del índice de masa corporal normal.

Este tratamiento también es usado cuando el paciente está sufriendo de *bronquiectasias* o *pancreactasis*, la cual, es la inflamación en los bronquios o el páncreas por las constantes infecciones, la cual se busca aliviar un poco la afección del paciente ya que la mayoría de veces estos terminan siendo operados en este punto.

Tratamiento 2:

Uno de los grandes peligros de la formación de mucosa en la vía respiratoria, es que esta se convierte en un ambiente prospero para la procreación de bacterias aeróbicas como la *pseudomona aeuroginosa* la cual, infecta pulmones y vías urinarias o puede generar otras infecciones generalizadas en el organismo, además de neumonías si esta bacteria no es tratada rápidamente. En el peor de los caso puede causar la muerte. Los medicamentos para combatir esta bacteria son variados, como el trimetoprim, aztreonam, tobramicina, o la solución salina la cual, el medico puede recetar combinándolos o elegir solo uno según su criterio. El resto de medicamentos sugeridos para este tratamiento, cumplen la misma función del tratamiento preventivo pero con el fin de eliminar el medio de procreación de esta bacteria.

Tratamiento 3:

En algunas ocasiones, el tratamiento general para la *pseudomona aeuroginosa* puede ser inefectivo, ya que esta puede generar resistencia a los tratamientos comunes. Por lo que se formulan una serie de tratamientos más potentes el cual se pueden combinar o no para hacer más efectiva su tarea de erradicar esta infección dependiendo del criterio del médico. Estos medicamentos son la tobramicina, levofloxina, amikacina, azitromicina, cefepime, ciprofloxa, colistina, piperacilina, tazobactam.

Tratamiento 4:

En algunos casos, la mucosa puede albergar varias cepas de bacterias como el estafilococo aerus causante de tos con expectoración, mareos, neumonía, diarrea y vómito entre otros síntomas; o el *aspergillus aerus* el cual es un hongo causante de la sinusitis alérgica; además de la *pseudomona* que puede ser también resistente. Por lo que estos patógenos tienen diferentes estructuras

celulares, los medicamentos usados actualmente son los siguientes: trimetoprim, tobramicina, amoxicilina más ácido clavulánico, azitromicina, vancomicina.

Tratamiento 5:

Otro de los problemas más comunes de los pacientes con fibrosis quística, es la mucosa que se acumula en el sistema digestivo, provocando dolores abdominales, mal absorción de los nutrientes alimenticios, deposiciones mal formadas y fétidas. Y para eliminar estas secreciones y mejorar la absorción de los nutrientes los medicamentos usados son las enzimas como la lipasa pancreática y el ácido ursoxicólico.

Es normal que estos pacientes tengan un bajo índice de masa corporal, por lo que las vitaminas son muy importantes en tratamiento en especial la vitamina k.

7.2 CAPÍTULO 2: MODELAMIENTO DE ÁRBOL DE DECISIÓN CON LOS SÍNTOMAS Y TRATAMIENTOS

7.2.1 Minería de datos

La minería de datos es una tecnología que tenido su auge en los últimos años y ha cogido gran importancia, el cual permite el conjunto de diferentes áreas (estadística, inteligencia artificial, matemáticas, biología y medicina), además ayuda a identificar la información oculta significativa que se encuentra en los grandes volúmenes de datos⁴, cuyo objetivo específico es que dicha información encontrada sirva de base para la toma de decisiones de acuerdo al caso de estudio.

En esta información base encontrada, se pueden encontrar patrones. Entiéndase por patrón modelo predictivo, una segmentación de datos, un sumario de datos o un modelo de dependencias. Los patrones encontrados dependerán de la clase de minería ejecutada. Para un conjunto finito de datos, siempre existirá un número de posibles patrones extremadamente grande y de hecho mucho mayor que la cantidad de datos disponibles.

Con lo anterior los métodos exhaustivos o de búsqueda óptima se convierten en imprácticos y se requieran técnicas de búsqueda basadas en heurísticas.

Los métodos de minería de datos se dividen en dos modelos: modelos descriptivos y modelos predictivos.

Los modelos descriptivos buscan la segmentación u obtención de grupos/individuos aislados de datos usando técnicas tales como análisis de clúster para la segmentación de los clientes y detección de anomalías para evitar fraudes y abusos.

⁴ *Campbell, Mary. base IV Guía de Auto enseñanza. España. Editorial McGraw Hill Interamericana. 1990. pp110/111,121/122,16,169, 179-191/192.. (4 Mar 2009).*

Los modelos predictivos o de clasificación incluyen los métodos de máquinas de soporte vectorial, redes bayesianas, redes neuronales, y árboles de decisión.

Este trabajo se centra específicamente en la elaboración de la técnica de árboles de decisión para la formulación de tratamientos de la fibrosis quística usado en el modelo predictivo.

7.2.2 Modelo predictivo

El modelo predictivo se usa para responder preguntas sobre datos futuros variados tales como:

- ¿Esta persona le gusta lo que vendemos?
- ¿Esta persona contestará las cartas que le enviamos?
- ¿Esta persona entrará a nuestro sitio web?
- ¿Esta persona tendrá presión alta?
- ¿Esta persona sufrirá de algún dolor?

El enfoque de la minería de datos es el modelado predictivo, cuyo propósito es predecir uno o varios campos en los datos haciendo uso del resto de campos disponibles. Cuando la variable a predecir es categórica, el problema es llamado clasificación. Cuando la variable es continua, se habla de regresión. La clasificación ha sido un tema bastante estudiado. Los enfoques más sencillos incluyen las técnicas estándar de regresión lineal, regresión lineal generalizada y análisis discriminante.

7.2.3 Árboles de decisión

Los árboles de decisión son una de las técnicas de aprendizaje inductivo supervisado no paramétrico, se utiliza para la predicción y es empleado en el campo de inteligencia artificial, donde a partir de una base de datos se construyen diagramas de construcción lógica, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren en forma repetitiva para la solución de un problema.

Los árboles de decisión tienen dos algoritmos para su construcción y evaluación; ID3 (Induction Decision Tree) y C4.5 el cual es la extensión del ID3.

7.2.3.1 Algoritmo ID3

Una de las dificultades que se presenta al realizar el proceso de construcción de un árbol de decisión es escoger el atributo más apropiado. Este atributo debe ubicarse en la raíz del árbol para lo cual se debe realizar una prueba estadística a cada uno de los atributos que permita determinar qué tan acertado se están clasificando los ejemplos de entrenamiento. Una vez se obtiene el atributo más apropiado, se selecciona y se utiliza como nodo prueba en la raíz del árbol, luego para cada uno de los otros atributos se procede a generar un nuevo descendiente. Los datos de

entrenamiento son divididos y asignados al nodo descendiente adecuado, es decir, se organizan las ramas de acuerdo al valor que toma cada atributo. Este procedimiento se realiza recursivamente en cada nodo descendiente, utilizando los datos de entrenamiento correspondientes.

Para decidir qué atributo es el más apropiado a usar en cada nodo del árbol se utiliza una propiedad estadísticas llamada ganancia de información, que mide que tan bien clasifica ese atributo a los datos de entrenamiento. Así que elige el nodo del árbol que tenga mayor ganancia de información y luego expande sus ramas utilizando la misma metodología.

7.2.3.2 Algoritmo C4.5 ³⁵

El algoritmo C4.5 genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente. El árbol se construye mediante la estrategia de profundidad-primero.

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. Para cada atributo discreto, se considera una prueba con n resultados, siendo n el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos. En cada nodo, el sistema debe decidir cuál prueba escoge para dividir los datos.

Los tres tipos de pruebas posibles propuestas por el C4.5 son:

La prueba "estándar" para las variables discretas, con un resultado y una rama para cada valor posible de la variable.

Una prueba más compleja, basada en una variable discreta, en donde los valores posibles son asignados a un número variable de grupos con un resultado posible para cada grupo, en lugar de para cada valor.

Si una variable A tiene valores numéricos continuos, se realiza una prueba binaria con resultados $A \leq Z$ y $A > Z$, para lo cual debe de terminarse el valor límite Z .

Todas estas pruebas se evalúan de la misma manera, mirando el resultado de la proporción de ganancia, o alternativamente, el de la ganancia resultante de la división que producen. Ha sido útil agregar una restricción adicional: para cualquier división, al menos dos de los subconjuntos de C deben contener un número razonable de casos. Esta restricción, que evita las subdivisiones casi triviales, es tomada en cuenta solamente cuando el conjunto C es pequeño.

³⁵ Salvatore Ruggieri, "Efficient C4.5" Departamento de informática Universidad de Pisa
<idb.csie.ncku.edu.tw/tsengsm/COURSE/DM/Paper/ec45.pdf>

En este proyecto se usará el algoritmo C4.5, ya que en comparación con el ID3, este tiene la capacidad de realizar podas para eliminar ramas que no son necesarias y reemplazarlas con los nodos hoja, proporcionando resultados más eficientes.

Ventajas:

- Evitar sobreajuste de los datos.
- Determinar qué tan profundo debe crecer el árbol de decisión.
- Reducir errores en la poda.
- Condicionar la Post-Poda.
- Manejar atributos continuos.
- Escoger un rango de medida apropiado.
- Manejo de datos de entrenamiento con valores faltantes.
- Manejar atributos con diferentes valores.
- Mejorar la eficiencia computacional

7.2.4 Ganancia de información

Para determinar el atributo más apropiado a usar en cada nodo del árbol se usa una propiedad estadística llamada ganancia de información que mide que tan bien clasifica ese atributo a los datos de entrenamiento. El cual se define en la siguiente fórmula:

$$G(S, A) \equiv H(S) - H(S, A)_{(1)}$$

En la anterior expresión el primer término $H(S)$ corresponde a la entropía de S , el segundo término corresponde al valor esperado de la entropía después de que S ha sido particionado de acuerdo al atributo A .

Como podemos observar el segundo término de la fórmula de ganancia no es más que la sumatoria de entropías de cada subconjunto de S , ponderado por la fracción.

7.2.5 Generación del árbol

Aplicando los conceptos anteriores se aplicará a un conjunto de pacientes con fibrosis quística de acuerdo a un concepto médico se les fórmula uno de los 5 tipos de tratamiento posibles de acuerdo a la tos con expectoración, dolores abdominales, deposiciones anormales o tipo de infección pulmonar que tenga el paciente (pseudomona aeruginosa, estafilococo aerus, aspergelis aerus) (ver tabla 4).

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomonas	Pseudomonas multi resistente	Estafilococos aeruos	Aspergillus aeruos	Tratamiento 1	Tratamiento 2	Tratamiento 3	Tratamiento 4	Tratamiento 5
1	NORMAL	NO	SI	SI	NO	NO	SI	NO	NO	NO	NO	NO	SI
2	NORMAL	SI	NO	NO	SI	NO	SI	NO	NO	NO	NO	SI	NO
3	NORMAL	SI	NO	NO	SI	NO	NO	NO	NO	SI	NO	NO	NO
4	NORMAL	SI	SI	NO	NO	SI	NO	NO	NO	NO	SI	NO	SI
5	BAJO	SI	NO	NO	SI	NO	NO	NO	SI	SI	NO	NO	NO
6	BAJO	NO	NO	NO	SI	NO	SI	SI	NO	NO	NO	SI	NO
7	BAJO	SI	NO	NO	NO	NO	NO	NO	SI	NO	NO	NO	NO
8	BAJO	SI	SI	SI	NO	NO	NO	NO	SI	NO	NO	NO	SI
9	BAJO	SI	NO	NO	SI	NO	NO	NO	SI	SI	NO	NO	NO
10	ALTO	SI	NO	SI	NO	SI	NO	NO	SI	NO	SI	NO	SI
11	NORMAL	SI	NO	NO	NO	SI	NO	NO	NO	NO	SI	NO	NO
12	NORMAL	NO	SI	SI	NO	SI	NO	NO	NO	NO	SI	NO	SI
13	NORMAL	SI	NO	NO	NO	SI	SI	NO	NO	NO	SI	SI	NO
14	BAJO	SI	NO	SI	NO	NO	NO	NO	SI	NO	NO	NO	SI
15	BAJO	SI	NO	NO	SI	NO	NO	NO	NO	SI	NO	NO	NO

Tabla 4. Administración de tratamientos

Fuente: Autores

El conjunto de datos S de la tabla 4 que contiene valores positivos o negativos sobre una variable dicotómica para calcular la entropía de S relativa a su clasificación booleana se debe definir:

Pp: es la probabilidad de que las respuestas sean positivas según el conjunto S.

Pn: es la probabilidad de que las respuestas sean negativas según el conjunto S.

$$Pn = 1 - Pp \quad (2)$$

En este caso de estudio, se tiene cinco tratamientos posibles y para cada una se debe calcular su entropía. Se puede observar, que de 15 pacientes, los tratamientos 1 y 5 tienen 6 respuestas positivas y 9 respuestas negativas, el tratamiento 2 tiene 4 respuestas positivas y 12 respuestas negativas, y el tratamiento 3 tiene 5 respuestas positivas y 9 respuestas negativas, el tratamiento 4 tiene 3 respuestas positivas y 12 respuestas negativas.

Tratamientos 1 y 5:

$$Pp = \frac{6}{15} = 0,4$$

$$Pn = \frac{9}{15} = 0,6$$

Tratamientos 2:

$$Pp = \frac{4}{15} = 0,27$$

$$Pn = \frac{11}{15} = 0,73$$

Tratamiento 3:

$$Pp = \frac{5}{15} = 0,33$$

$$Pn = \frac{10}{15} = 0,66$$

Tratamiento 4:

$$Pp = \frac{3}{15} = 0,2$$

$$Pn = \frac{12}{15} = 0,8$$

Para calcular la entropía se define con base a las probabilidades anteriores, así:

$$H(S) = -Pp \log_2 Pp - Pn \log_2 Pn \quad (3)$$

Según la ecuación (3), la entropía del conjunto de los 15 datos respecto a las variables "Tratamientos 1, 2, 3, 4, 5" se calcula de la siguiente manera:

Entropía tratamiento 1 y 5:

$$H(S) = \sum_{i=1}^c -P_i \log_2 P_i = -\frac{6}{15} \log_2 \frac{6}{15} - \frac{9}{15} \log_2 \frac{9}{15} = 0,97$$

Entropía tratamiento 2:

$$H(S) = \sum_{i=1}^c -P_i \log_2 P_i = -\frac{4}{15} \log_2 \frac{4}{15} - \frac{11}{15} \log_2 \frac{11}{15} = 0,83$$

Entropía tratamiento 3:

$$H(S) = \sum_{i=1}^c -P_i \log_2 P_i = -\frac{5}{15} \log_2 \frac{5}{15} - \frac{10}{15} \log_2 \frac{10}{15} = 0,92$$

Entropía tratamiento 4:

$$H(S) = \sum_{i=1}^c -P_i \log_2 P_i = -\frac{3}{15} \log_2 \frac{3}{15} - \frac{12}{15} \log_2 \frac{12}{15} = 0,72$$

7.2.5.1 Tablas de contingencia

La entropía es una medida de desorden e impureza en un conjunto de datos. Para la clasificación de los datos se utiliza la medida ganancia de información explicada anteriormente el cual reduce la entropía obtenida al realizar la división de los datos en los subconjuntos de entrenamiento.

$$H(A, S) \equiv \sum_{V \in \text{Valores}(A)} \frac{|S_V|}{|S|} = H(S_V) \quad (4)$$

Donde S es el conjunto de muestras para clasificar (en este caso los pacientes con fibrosis quística de la tabla 4), A son los atributos (los síntomas) y S_V es un subconjunto de S (que puede tomar cierto valor dependiendo del atributo).

Para facilitar los cálculos anteriores se usan tablas de contingencia para el tratamiento 1, que se usarán como ejemplo, estas tablas son obtenidas a partir de la tabla 4. Para calcular la entropía del conjunto de datos, H(S) se procede a calcular la entropía de cada uno de los valores de A con la fórmula (4).

Tratamiento 1:

	BAJO	NORMAL	ALTO	TOTAL
SI	5	0	1	6
NO	2	7	0	9
TOTAL	7	7	1	15

Tabla 5. Respuestas de Atributo "Peso" en el Tratamiento 1.

Fuente: Autores

$$H(S_{P=BAJO}) = - \sum_{i=1}^C P_i \log_2 P_i = -\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} = 0,8631$$

$$H(S_{P=NORMAL}) = - \sum_{i=1}^C P_i \log_2 P_i = -\frac{0}{7} \log_2 \frac{0}{7} - \frac{7}{7} \log_2 \frac{7}{7} = 0$$

$$H(S_{P=ALTO}) = - \sum_{i=1}^C P_i \log_2 P_i = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

7.2.5.2 Proporción de la Ganancia de Información

Con los cálculos de las entropías por conjunto de cada atributo del tratamiento 1, se calculan las ganancias de información con la fórmula (1), se calcula la información de la división para cada atributo con la fórmula (5) este término es sensitivo a que tan amplia y uniformemente el atributo separa los datos²².

$$I_{div}(S, A) \equiv - \sum_{i=1}^C \left| \frac{S_i}{S} \right| \log_2 \left| \frac{S_i}{S} \right| \quad (5)$$

Por último se calcula proporción de la ganancia de la información el cual penaliza los atributos que tienen demasiados valores²² y favorece aquellos atributos que, en igualdad de ganancia separa los datos en menos clases.

$$PGan(S, A) \equiv \frac{G(S, A)}{I_{div}(S, A)} \quad (6)$$

²² Guillermo Solarte Arboles de decisiones en el diagnóstico de enfermedades cardiovasculares revista Scientia et Technica Año XVI, No 49, Diciembre de 2011. Universidad Tecnológica de Pereira. ISSN 0122-1701

²²IDEM

Tratamiento 1:

Peso

$$G(S, P) = 0,97 - \frac{7}{15} 0,8631 - \frac{7}{15} 0 - \frac{1}{15} 0 = 0,5672$$

$$Idiv(S, P) = -\frac{7}{15} \log_2 \frac{7}{15} - \frac{7}{15} \log_2 \frac{7}{15} - \frac{1}{15} \log_2 \frac{1}{15} = 1,2866$$

$$PGan(S, P) = \frac{0,5672}{1,2866} = 0,44$$

7.2.5.3 Ensamble del Árbol**Primera Iteración**

A partir de la tabla 4, se hacen los cálculos correspondientes para generar las tablas de contingencia, y obtener de ellas los valores de las proporciones de las ganancias de información y así poder deducir cual será el nodo raíz.

	BAJO	NORMAL	ALTO	TOTAL
SI	5	0	1	6
NO	2	7	0	9
TOTAL	7	7	1	15

Tabla 6. Respuestas de Atributo "Peso" Primer Iteración.

Fuente: Autores

$$H(S_{P=BAJO}) = -\sum_{i=1}^C P_i \log_2 P_i = -\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} = 0,8631$$

$$H(S_{P=NORMAL}) = -\sum_{i=1}^C P_i \log_2 P_i = -\frac{0}{7} \log_2 \frac{0}{7} - \frac{7}{7} \log_2 \frac{7}{7} = 0$$

$$H(S_{P=ALTO}) = -\sum_{i=1}^C P_i \log_2 P_i = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

	SI	NO	TOTAL
SI	6	0	6
NO	6	3	9
TOTAL	12	3	15

Tabla 7. Atributo "Tos con expectoración de color" Primer Iteración.

Fuente: Autores

$$H(S_{TE=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = - \frac{6}{12} \log_2 \frac{6}{12} - \frac{6}{12} \log_2 \frac{6}{12} = 1$$

$$H(S_{TE=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = - \frac{0}{3} \log_2 \frac{0}{3} - \frac{3}{3} \log_2 \frac{3}{3} = 0$$

	SI	NO	total
SI	1	5	6
NO	3	6	9
Total	4	11	15

Tabla 8. Atributo "Dolor Abdominal" Primer Iteración.

Fuente: Autores

$$H(S_{DA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = - \frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0,8112$$

$$H(S_{DA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = - \frac{5}{11} \log_2 \frac{5}{11} - \frac{6}{11} \log_2 \frac{6}{11} = 0,9940$$

	SI	NO	total
Si	3	3	6
No	2	7	9
Total	5	10	15

Tabla 9. Atributo "Deposiciones anormales" Primer Iteración.

Fuente: Autores

$$H(S_{EA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0,9709$$

$$H(S_{EA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{3}{10} \log_2 \frac{3}{10} - \frac{7}{10} \log_2 \frac{7}{10} = 0,8812$$

	SI	NO	total
Si	2	4	6
No	4	5	9
Total	6	9	15

Tabla 10. Atributo "Pseudomona aeruginosa" Primer Iteración.

Fuente: Autores

$$H(S_{PA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{2}{6} \log_2 \frac{2}{6} - \frac{4}{6} \log_2 \frac{4}{6} = 0,9182$$

$$H(S_{PA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{4}{9} \log_2 \frac{4}{9} - \frac{5}{9} \log_2 \frac{5}{9} = 0,9910$$

	SI	NO	total
Si	1	5	6
No	3	6	9
Total	4	11	15

Tabla 11. Atributo "Pseudomona aeruginosa Multi resistente" Primer Iteración.

Fuente: Autores

$$H(S_{PM=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0,8112$$

$$H(S_{PM=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{11} \log_2 \frac{5}{11} - \frac{6}{11} \log_2 \frac{6}{11} = 0,9940$$

	SI	NO	total
Si	0	6	6
No	4	5	9
Total	4	11	15

Tabla 12. Atributo "Estafilococo aerus" Primer Iteración.

Fuente: Autores

$$H(S_{E=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{4} \log_2 \frac{0}{4} - \frac{4}{4} \log_2 \frac{4}{4} = 0$$

$$H(S_{E=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{6}{11} \log_2 \frac{6}{11} - \frac{5}{11} \log_2 \frac{5}{11} = 0,9940$$

	SI	NO	total
Si	0	6	6
No	1	8	9
Total	1	14	15

Tabla 13. Atributo "Aspergelis aerus" en el Primer Iteración.

Fuente: Autores

$$H(S_{A=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$H(S_{A=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{6}{14} \log_2 \frac{6}{14} - \frac{8}{14} \log_2 \frac{8}{14} = 0,9852$$

A partir de los valores obtenidos, aplicamos los resultados a las funciones de ganancia de información, información de la división, y la proporción de la ganancia de la información.

Peso

$$G(S, P) = 0,97 - \frac{7}{15} 0,8631 - \frac{7}{15} 0 - \frac{1}{15} 0 = 0,5672$$

$$I_{div}(S, P) = -\frac{7}{15} \log_2 \frac{7}{15} - \frac{7}{15} \log_2 \frac{7}{15} - \frac{1}{15} \log_2 \frac{1}{15} = 1,2866$$

$$PGan(S, P) = \frac{0,5672}{1,2866} = 0,44$$

Tos con expectoración de color

$$G(S, TE) = 0,97 - \frac{12}{15}1 - \frac{3}{15}0 = 0,17$$

$$Idiv(S, TE) = -\frac{12}{15}\log_2 \frac{12}{15} - \frac{3}{15}\log_2 \frac{3}{15} = 0,7219$$

$$PGan(S, TE) = \frac{0,17}{0,7219} = 0,2354$$

Dolor abdominal

$$G(S, DA) = 0,97 - \frac{4}{15}0,8112 - \frac{11}{15}0,994 = 0,0247$$

$$Idiv(S, DA) = -\frac{4}{15}\log_2 \frac{4}{15} - \frac{11}{15}\log_2 \frac{11}{15} = 0,8366$$

$$PGan(S, DA) = \frac{0,0247}{0,8366} = 0,0295$$

Deposiciones anormales

$$G(S, EA) = 0,97 - \frac{5}{15}0,9709 - \frac{10}{15}0,8812 = 0,0589$$

$$Idiv(S, EA) = -\frac{5}{15}\log_2 \frac{5}{15} - \frac{10}{15}\log_2 \frac{10}{15} = 0,9182$$

$$PGan(S, EA) = \frac{0,0589}{0,9182} = 0,0641$$

Pseudomona aeruginosa

$$G(S, PA) = 0,97 - \frac{6}{15}0,9182 - \frac{9}{15}0,9910 = 0,0081$$

$$Idiv(S, PA) = -\frac{6}{15}\log_2 \frac{6}{15} - \frac{9}{15}\log_2 \frac{9}{15} = 0,97$$

$$PGan(S, PA) = \frac{0,0081}{0,97} = 0,0083$$

Pseudomona aeruginosa multiresistente

$$G(S, PM) = 0,97 - \frac{4}{15}0,8112 - \frac{11}{15}0,994 = 0,0636$$

$$Idiv(S, PM) = -\frac{4}{15} \log_2 \frac{4}{15} - \frac{11}{15} \log_2 \frac{11}{15} = 0,9183$$

$$PGan(S, PM) = \frac{0,0636}{0,9183} = 0,0693$$

Estafilococo aerus

$$G(S, E) = 0,97 - \frac{4}{15} \cdot 0 - \frac{11}{15} \cdot 0,994 = 0,241$$

$$Idiv(S, E) = -\frac{4}{15} \log_2 \frac{4}{15} - \frac{11}{15} \log_2 \frac{11}{15} = 0,8366$$

$$PGan(S, E) = \frac{0,241}{0,8366} = 0,2892$$

Aspergelis aerus

$$G(S, A) = 0,97 - \frac{1}{15} \cdot 0 - \frac{14}{15} \cdot 0,9852 = 0,05$$

$$Idiv(S, A) = -\frac{1}{15} \log_2 \frac{1}{15} - \frac{14}{15} \log_2 \frac{14}{15} = 0,2604$$

$$PGan(S, A) = \frac{0,05}{0,2604} = 0,1455$$

A continuación se puede ver en la tabla 14 los resultados obtenidos de cada atributo.

Atributos	Información de la división	Ganancia de información	Proporción de la Ganancia de la Información
Peso	1.2866	0,5672	0,44
Tos con expectoración de color	0,7219	0,17	0,2354
Dolor abdominal	0,8366	0,0247	0,0295
Deposiciones anormales	0,9182	0,0589	0,0641
Pseudomona	0,97	0,0081	0,0083
P. multi resistente	0,8366	0,0247	0,0295
Estafilococo aerus	0,8366	0,241	0,0288
Aspergelis aerus	0,2604	0,05	0,192

Tabla 14. Ganancia de información Primer Iteración.

Fuente: Autores

Como se observa en la tabla 14, el atributo que se debe seleccionar como nodo raíz es "Peso", ya que de acuerdo con la medida de la proporción de la ganancia de información (0,44), se puede ver que es el síntoma más adecuado para ser nodo inicial (raíz), desprendiendo de este tres ramas (Bajo, Medio, Alto).

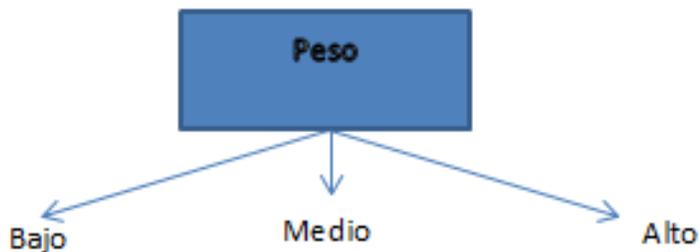


Figura 3. Selección de atributo más adecuado "Peso"

Fuente: Autores

Continuando con el paso a paso se debe aplicar la misma técnica en cada uno de los nuevos nodos creados, pero en cada uno solo se usa un subconjunto de los datos como se observa en las tablas 15, 16, 17.

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomonas	Pseudomonas multi resistente	Estafilococcus aerus	Aspergillus aerus	Tratamiento 1
1	NORMAL	NO	SI	SI	NO	NO	SI	NO	NO
2	NORMAL	SI	NO	NO	SI	NO	SI	NO	NO
3	NORMAL	SI	NO	NO	SI	NO	NO	NO	NO
4	NORMAL	SI	SI	NO	NO	SI	NO	NO	NO
11	NORMAL	SI	NO	NO	NO	SI	NO	NO	NO
12	NORMAL	NO	SI	SI	NO	SI	NO	NO	NO
13	NORMAL	SI	NO	NO	NO	SI	SI	NO	NO

Tabla 15. Datos de entrenamiento Peso NORMAL

Fuente: Autores

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomonas	Pseudomonas multi resistente	Estafilococcus aerus	Aspergillus aerus	Tratamiento 1
10	ALTO	SI	NO	SI	NO	SI	NO	NO	SI

Tabla 16. Datos de entrenamiento Peso ALTO

Fuente: Autores

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomonas	Pseudomonas multi resistente	Estafilococcus aerus	Aspergillus aerus	Tratamiento 1
5	BAJO	SI	NO	NO	SI	NO	NO	NO	SI
6	BAJO	NO	NO	NO	SI	NO	SI	SI	NO
7	BAJO	SI	NO	NO	NO	NO	NO	NO	SI
8	BAJO	SI	SI	SI	NO	NO	NO	NO	SI
9	BAJO	SI	NO	NO	SI	NO	NO	NO	SI
14	BAJO	SI	NO	SI	NO	NO	NO	NO	SI
15	BAJO	SI	NO	NO	SI	NO	NO	NO	NO

Tabla 17. Datos de entrenamiento Peso BAJO

Fuente: Autores

Como podemos apreciar en la tabla 15 y en la tabla 16, vemos que la columna "Tratamiento 1", tiene los mismos valores, por tanto, de allí se desprende un nodo hoja (definitivo) con el valor "NO" y "SI" en la columna de "Peso" con valores "NORMAL" y "ALTO" respectivamente.

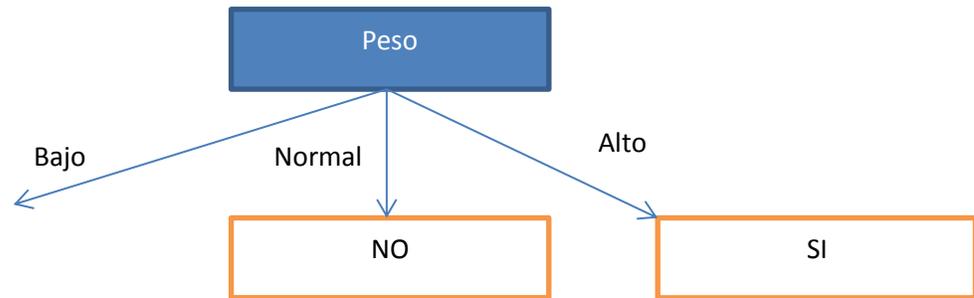


Figura 4. Resultado de los cálculos del nodo raíz en el árbol
Fuente: Autores

Ya que la tabla 16, no dio como resultado un nodo hoja. El algoritmo itera nuevamente, con la tabla 16 ya mencionada.

Segunda iteración:

	SI	NO	TOTAL
SI	5	0	5
NO	1	1	2
TOTAL	6	1	7

Tabla 18. Atributo "Tos con expectoración de color" Segunda Iteración.
Fuente: Autores

$$H(S_{TE=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0.65$$

$$H(S_{TE=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

	SI	NO	Total
SI	1	4	5
NO	0	2	2
Total	1	6	7

Tabla 19. Atributo "Dolor Abdominal" Segunda Iteración.
Fuente: Autores

$$H(S_{DA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

$$H(S_{DA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} = 0,9183$$

	SI	NO	Total
Si	2	3	5
No	0	2	2
Total	2	5	7

Tabla 20. Atributo "Deposiciones anormales" Segunda Iteración.
Fuente: Autores

$$H(S_{EA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} = 0$$

$$H(S_{EA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0,9710$$

	SI	NO	Total
Si	2	3	5
No	2	0	2
Total	4	3	7

Tabla 21. Atributo "Pseudomona aeruginosa" Segunda Iteración.
Fuente: Autores

$$H(S_{PA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 1$$

$$H(S_{PA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} = 0$$

	SI	NO	Total
Si	0	5	5
No	0	2	2
Total	0	7	7

Tabla 22. Atributo "Pseudomona aeruginosa Multi resistente" Segunda Iteración.

Fuente: Autores

$$H(S_{PM=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{0} \log_2 \frac{0}{0} - \frac{0}{0} \log_2 \frac{0}{0} = 0$$

$$H(S_{PM=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} = 0,8631$$

	SI	NO	total
Si	0	5	5
No	1	1	2
Total	1	6	7

Tabla 23. Atributo "Estafilococo aerus" Segunda Iteración.

Fuente: Autores

$$H(S_{E=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$H(S_{E=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0,65$$

	SI	NO	total
Si	0	5	5
No	1	1	2
Total	1	6	7

Tabla 24. Atributo "Aspergelis aerus" Segunda Iteración.

Fuente: Autores

$$H(S_{A=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$H(S_{A=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0,65$$

A continuación se realizar los cálculos de la entropía del tratamiento 1, la ganancia de información, la división de la información y la proporción de la ganancia de información para la segunda iteración.

Entropía tratamiento 1

$$H(S) = \sum_{i=1}^c -P_i \log_2 P_i = -\frac{5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} = 0,8631$$

Tos con expectoración de color

$$G(S, TE) = 0,8631 - \frac{6}{7} 0,65 - \frac{1}{7} 0 = 0,3060$$

$$I_{div}(S, TE) = -\frac{6}{7} \log_2 \frac{6}{7} - \frac{1}{7} \log_2 \frac{1}{7} = 0,5917$$

$$PGan(S, TE) = \frac{0,3060}{0,5917} = 0,5171$$

Dolor abdominal

$$G(S, DA) = 0,8631 - \frac{1}{7} 0 - \frac{6}{7} 0,9183 = 0,0760$$

$$I_{div}(S, DA) = -\frac{1}{7} \log_2 \frac{1}{7} - \frac{6}{7} \log_2 \frac{6}{7} = 0,5916$$

$$PGan(S, DA) = \frac{0,0760}{0,5916} = 0,1284$$

Deposiciones anormales

$$G(S, EA) = 0,8631 - \frac{2}{7} 0 - \frac{5}{7} 0,9710 = 0,1695$$

$$I_{div}(S, EA) = -\frac{2}{7} \log_2 \frac{2}{7} - \frac{5}{7} \log_2 \frac{5}{7} = 0,8631$$

$$PGan(S, EA) = \frac{0,0589}{0,9182} = 0,1964$$

Pseudomona aeruginosa

$$G(S, PA) = 0,8631 - \frac{4}{7}1 - \frac{3}{7}0 = 0,2917$$

$$Idiv(S, PA) = -\frac{4}{7}\log_2 \frac{4}{7} - \frac{3}{7}\log_2 \frac{3}{7} = 0,9852$$

$$PGan(S, PA) = \frac{0,2917}{0,9852} = 0,2961$$

Pseudomona aeruginosa multi resistente

$$G(S, PM) = 0,8631 - \frac{0}{7}0 - \frac{7}{7}0,8631 = 0,1069$$

$$Idiv(S, PM) = -\frac{0}{7}\log_2 \frac{0}{7} - \frac{7}{7}\log_2 \frac{7}{7} = 0$$

$$PGan(S, PM) = \frac{0,1069}{0} = 0$$

Estafilococo aerus

$$G(S, E) = 0,8631 - \frac{1}{7}0 - \frac{6}{7}0,65 = 0,3060$$

$$Idiv(S, E) = -\frac{1}{7}\log_2 \frac{1}{7} - \frac{6}{7}\log_2 \frac{6}{7} = 0,5917$$

$$PGan(S, E) = \frac{0,3060}{0,5917} = 0,5171$$

Aspergelis aerus

$$G(S, A) = 0,8631 - \frac{1}{7}0 - \frac{6}{7}0,65 = 0,3060$$

$$Idiv(S, A) = -\frac{1}{7}\log_2 \frac{1}{7} - \frac{6}{7}\log_2 \frac{6}{7} = 0,5917$$

$$PGan(S, A) = \frac{0,3060}{0,5917} = 0,5171$$

A continuación se puede ver en la tabla 25 los resultados obtenidos de cada atributo para la segunda iteración.

Atributos	Información de la división	Ganancia de información	Proporción de la Ganancia de la Información
Tos con expectoración de color	0,5917	0,306	0,5171
Dolor abdominal	0,5916	0,076	0,1284
Deposiciones anormales	0,8631	0,1695	0,1964
Pseudomona	0,9852	0,2917	0,2961
P. multi resistente	0	0,1069	0
Estafilococo aerus	0,5917	0,306	0,5171
Aspergelis aerus	0,5917	0,306	0,5171

Tabla 25. Ganancia de información Segunda Iteración.

Fuente: Autores

Como se observa en la tabla 25, el atributo que se debe seleccionar como siguiente nodo es "Tos con expectoración de color", ya que de acuerdo con la medida de la proporción de la ganancia de información (0,5171), es el valor más alto. Sin embargo este empata con el estafilococo aerus y el aspergillus aerus, en este caso se elige el primero ya que computacionalmente el algoritmo lo elige, y de este se desprende dos ramas (SI, NO).

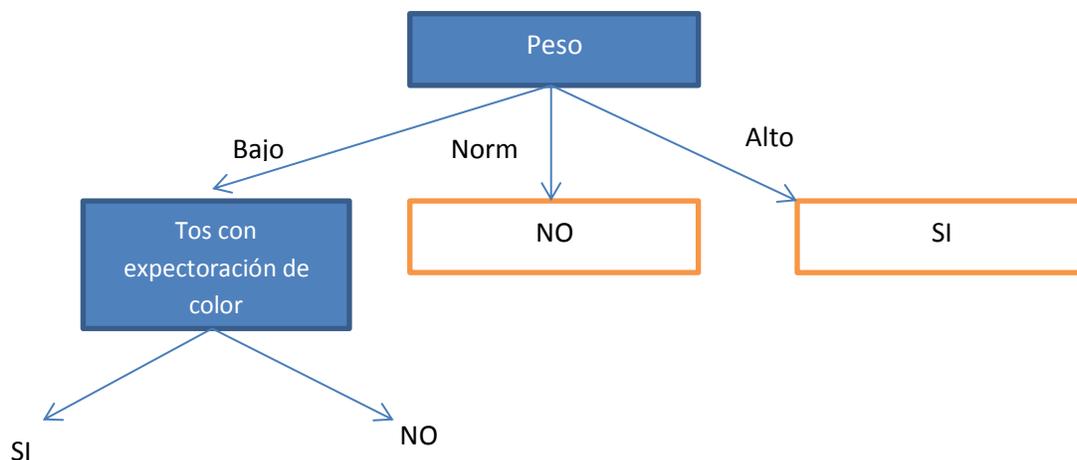


Figura 5. Selección de atributo más adecuado "Tos con expectoración"

Fuente: Autores

Continuando con el paso a paso se debe aplicar la misma técnica en cada uno de los nuevos nodos creados, pero en cada uno solo se usa un subconjunto de los datos como se observa en las tablas 26 y 27.

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomona	Pseudomona multi resistente	Estafilococo aerus	Aspergillus aerus	Tratamiento 1
5	BAJO	SI	NO	NO	SI	NO	NO	NO	SI
7	BAJO	SI	NO	NO	NO	NO	NO	NO	SI
8	BAJO	SI	SI	SI	NO	NO	NO	NO	SI
9	BAJO	SI	NO	NO	SI	NO	NO	NO	SI
14	BAJO	SI	NO	SI	NO	NO	NO	NO	SI
15	BAJO	SI	NO	NO	SI	NO	NO	NO	NO

Tabla 26. Datos de entrenamiento Tos con expectoración SI

Fuente: Autores

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomona	Pseudomona multi resistente	Estafilococo aerus	Aspergillus aerus	Tratamiento 1
6	BAJO	NO	NO	NO	SI	NO	SI	SI	NO

Tabla 27. Datos de entrenamiento Tos con expectoración NO

Fuente: Autores

Como podemos apreciar en la tabla 27, vemos que la columna "Tratamiento 1", tiene en su única fila un valor negativo por tanto, de allí se desprende un nodo hoja (definitivo) con el valor "NO".

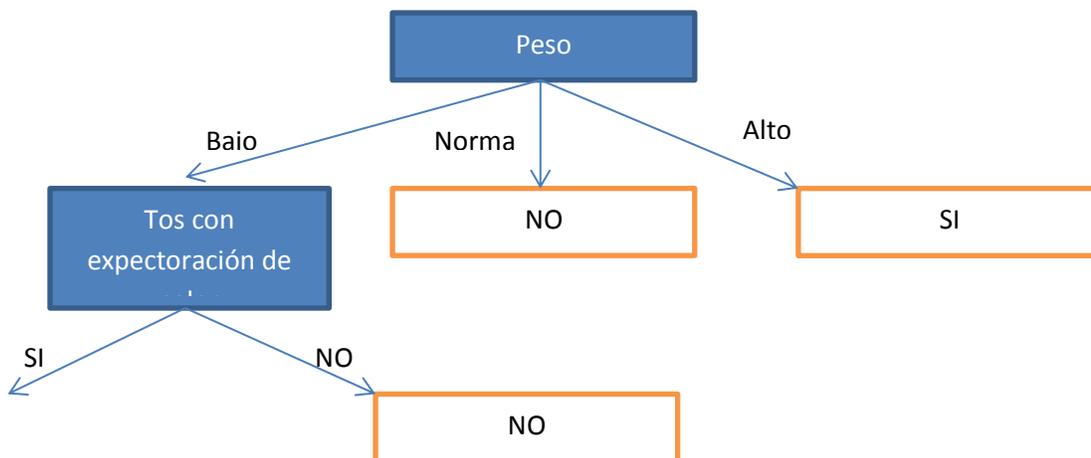


Figura 6. Resultado de los cálculos segunda iteración

Fuente: Autores

Tercera iteración

	SI	NO	Total
SI	1	4	5
NO	0	1	1
total	1	5	6

Tabla 28. Atributo "Dolor Abdominal" Tercera Iteración.
Fuente: Autores

$$H(S_{DA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5} = 0$$

$$H(S_{DA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0,7219$$

	SI	NO	Total
si	2	3	5
no	0	1	1
total	2	4	6

Tabla 29. Atributo "Deposiciones anormales" Tercera Iteración.
Fuente: Autores

$$H(S_{EA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0$$

$$H(S_{EA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0,8113$$

	SI	NO	Total
si	2	3	5
no	1	0	1
total	3	3	6

Tabla 30. Atributo "Pseudomona aeruginosa" Tercera Iteración.
Fuente: Autores

$$H(S_{PA=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0,9183$$

$$H(S_{PA=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{3}{3} \log_2 \frac{3}{3} - \frac{0}{3} \log_2 \frac{0}{3} = 0$$

	SI	NO	Total
si	0	5	5
no	0	1	1
total	0	6	6

Tabla 31. Atributo "Pseudomona aeruginosa Multi resistente" Tercera Iteración.

Fuente: Autores

$$H(S_{PM=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{0} \log_2 \frac{0}{0} - \frac{0}{0} \log_2 \frac{0}{0} = 0$$

$$H(S_{PM=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0,65$$

	SI	NO	total
si	0	5	5
no	0	1	1
total	0	6	6

Tabla 32. Atributo "Estafilococo aerus" Tercera Iteración.

Fuente: Autores

$$H(S_{E=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$H(S_{E=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0,65$$

	SI	NO	total
si	0	5	5
no	0	1	1
total	0	6	6

Tabla 33. Atributo "Aspergelis aerus" Tercera Iteración.

Fuente: Autores

$$H(S_{A=SI}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{0}{1} \log_2 \frac{0}{1} - \frac{1}{1} \log_2 \frac{1}{1} = 0$$

$$H(S_{A=NO}) = - \sum_{i=1}^c P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0,65$$

A continuación se realizar los cálculos de la entropía del tratamiento 1, la ganancia de información, la división de la información y la proporción de la ganancia de información para la tercera iteración.

Entropía tratamiento 1

$$H(S) = \sum_{i=1}^c -P_i \log_2 P_i = -\frac{5}{6} \log_2 \frac{5}{6} - \frac{1}{6} \log_2 \frac{1}{6} = 0,65$$

Dolor abdominal

$$G(S, DA) = 0,65 - \frac{1}{6} \cdot 0 - \frac{5}{6} \cdot 0,7219 = 0,0484$$

$$I_{div}(S, DA) = -\frac{1}{6} \log_2 \frac{1}{6} - \frac{5}{6} \log_2 \frac{5}{6} = 0,65$$

$$PGan(S, DA) = \frac{0,0484}{0,65} = 0,0745$$

Deposiciones anormales

$$G(S, EA) = 0,65 - \frac{2}{6} \cdot 0 - \frac{4}{6} \cdot 0,8113 = 0,1091$$

$$I_{div}(S, EA) = -\frac{2}{6} \log_2 \frac{2}{6} - \frac{4}{6} \log_2 \frac{4}{6} = 0,9182$$

$$PGan(S, EA) = \frac{0,1091}{0,9182} = 0,1188$$

Pseudomona aeruginosa

$$G(S, PA) = 0,65 - \frac{3}{6}0,9183 - \frac{3}{6}0 = 0,1909$$

$$Idiv(S, PA) = -\frac{3}{6}\log_2 \frac{3}{6} - \frac{3}{6}\log_2 \frac{3}{6} = 1$$

$$PGan(S, PA) = \frac{0,1909}{1} = 0,2961$$

Pseudomona aeruginosa multi resistente

$$G(S, PM) = 0,65 - \frac{0}{6}0 - \frac{6}{6}0,65 = 0,00002$$

$$Idiv(S, PM) = -\frac{0}{6}\log_2 \frac{0}{6} - \frac{6}{6}\log_2 \frac{6}{6} = 0$$

$$PGan(S, PM) = \frac{0,00002}{0} = 0$$

Estafilococo aerus

$$G(S, E) = 0,65 - \frac{0}{6}0 - \frac{6}{6}0,65 = 0,00002$$

$$Idiv(S, E) = -\frac{0}{6}\log_2 \frac{0}{6} - \frac{6}{6}\log_2 \frac{6}{6} = 0$$

$$PGan(S, E) = \frac{0,00002}{0} = 0$$

Aspergelis aerus

$$G(S, A) = 0,65 - \frac{0}{6}0 - \frac{6}{6}0,65 = 0,00002$$

$$Idiv(S, A) = -\frac{0}{6}\log_2 \frac{0}{6} - \frac{6}{6}\log_2 \frac{6}{6} = 0$$

$$PGan(S, A) = \frac{0,00002}{0} = 0$$

A continuación se puede ver en la tabla 34 los resultados obtenidos de cada atributo para la segunda iteración.

Atributos	Información de la división	Ganancia de información	Proporción de la Ganancia de la Información
Dolor abdominal	0,65	0,0484	0,0745
Deposiciones anormales	0,9182	0,1091	0,1188
Pseudomona	1	0,2917	0,2917
P. multi resistente	0	0	0
Estafilococo aerus	0	0	0
Aspergelis aerus	0	0	0

Tabla 34. Ganancia de información Tercera Iteración.

Fuente: Autores

Como se observa en la tabla 34, el atributo que se debe seleccionar como siguiente nodo es "Pseudomona", ya que de acuerdo con la medida de la proporción de la ganancia de información (0,2917), es el valor más alto, de este se desprende dos ramas (SI, NO).

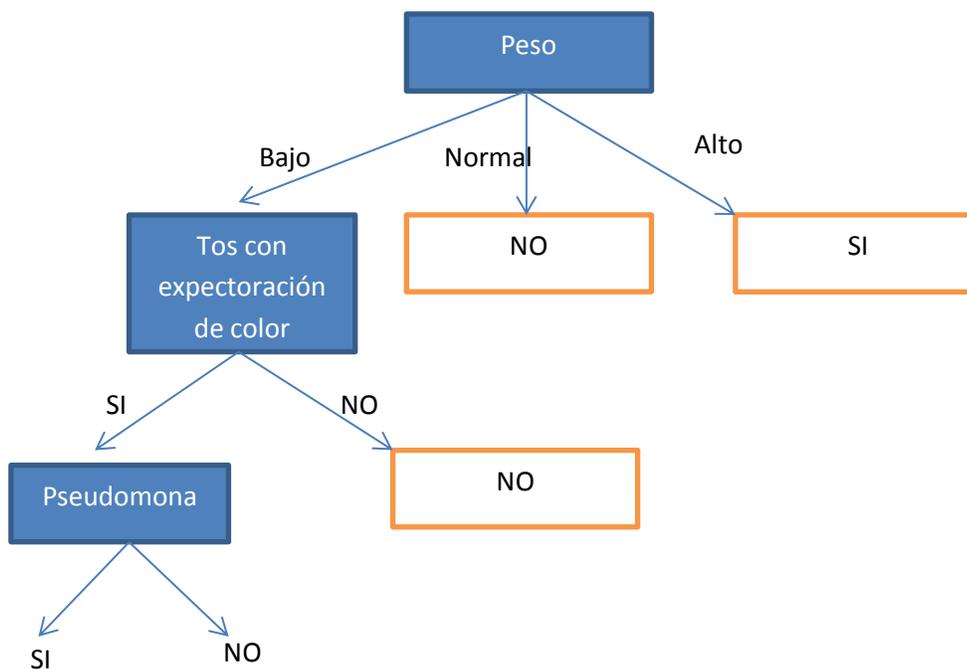


Figura 7. Selección de atributo más adecuado "Pseudomona"

Fuente: Autores

Continuando con el paso a paso se debe aplicar la misma técnica en cada uno de los nuevos nodos creados, pero en cada uno solo se usa un subconjunto de los datos como se observa en las tablas 35 y 36.

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomona	Pseudomona multi resistente	Estafilococo aerus	Aspergillus aerus	Tratamiento 1
5	BAJO	SI	NO	NO	SI	NO	NO	NO	SI
9	BAJO	SI	NO	NO	SI	NO	NO	NO	SI
15	BAJO	SI	NO	NO	SI	NO	NO	NO	NO

Tabla 35. Datos de entrenamiento Pseudomona SI

Fuente: Autores

Paciente	Peso	Tos con expectoración de color	Dolor abdominal	Deposiciones anormales	Pseudomona	Pseudomona multi resistente	Estafilococo aerus	Aspergillus aerus	Tratamiento 1
7	BAJO	SI	NO	NO	NO	NO	NO	NO	SI
8	BAJO	SI	SI	SI	NO	NO	NO	NO	SI
14	BAJO	SI	NO	SI	NO	NO	NO	NO	SI

Tabla 36. Datos de entrenamiento Pseudomona NO.

Fuente: Autores

Como podemos apreciar en la tabla 36, vemos que la columna "Tratamiento 1", tiene en todas sus filas un valor positivo por tanto, de allí se desprende un nodo hoja (definitivo) con el valor "SI". Según los resultados en la tabla 34, la ganancia de información y la proporción de la ganancia de información tienen el mismo valor (0,2917) esto quiere decir que no es necesario seguir expandiendo el árbol y dejar un nodo hoja con el valor "SI" ya que es el mayor peso en la tabla 35.

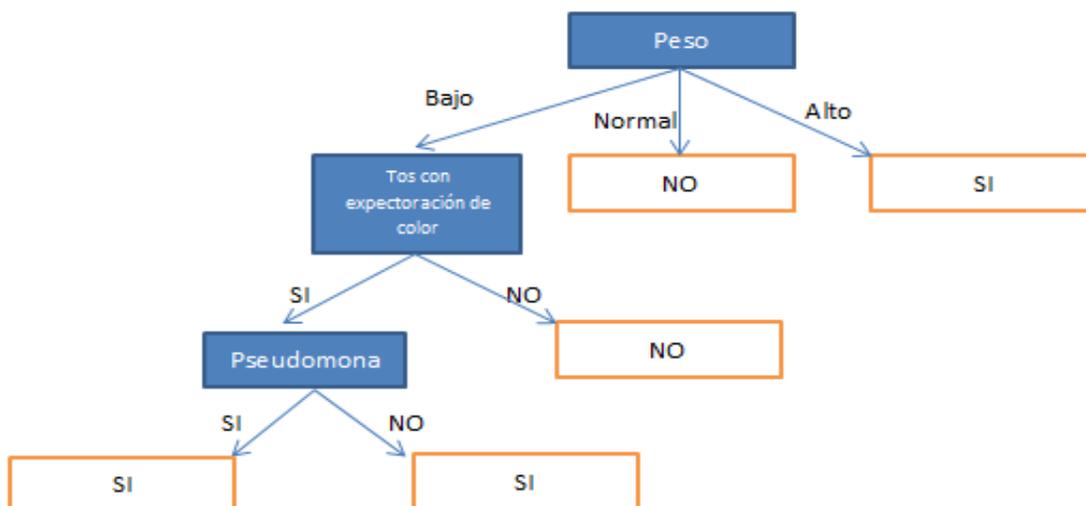


Figura 8. Resultado de los cálculos Tercera Iteración

Fuente: Autores

7.3 CAPÍTULO 3: DESARROLLO DEL PROTOTIPO DEL SIMULADOR PARA LA FORMULACIÓN DE TRATAMIENTOS PARA LA FIBROSIS QUISTICA

El lenguaje de programación elegido para el prototipo del “simulador de formulación de tratamientos para la fibrosis quística” es el lenguaje orientado a objetos JAVA el cual es multiplataforma, es decir, cualquier programa creado a través de Java podrá funcionar correctamente en cualquier sistema operativo. Otra de las ventajas principales de Java es que con este lenguaje es posible modelar cualquier elemento o aplicación. Además de que cuenta con un gran número de librerías y es un lenguaje que ha sido muy utilizado para crear simuladores como el proyecto de grado en la Universidad Francisco de Paula Santander (Cúcuta), realizado por Wilfred Uriel García - Cindy Yulieth Pabón Sánchez "Simuladores para Estudio de Estructuras de Datos"⁴².

7.3.1 Pseudocódigo del Algoritmo C4.5 ²⁵

A continuación se muestra el Pseudocódigo del algoritmo C4.5 en el cual se muestra como debería de funcionar el código del prototipo.

R: conjunto de datos no clasificadores
 C: atributo clasificador
 S: conjunto de entrenamiento

Inicio

Si el conjunto S es vacío,

Devolver un único nodo con valor falla

Si todos los registros de S tienen el mismo valor para el atributo clasificador,

Devolver un único nodo con dicho valor;

Si el conjunto R está vacío, entonces

Devolver un único nodo con el valor más frecuente del atributo clasificador en los registros de S [Nota: habrá errores, es decir, registros que no estarán bien clasificados en este caso];

Si el conjunto R no está vacío, entonces

D es el atributo con mayor proporción de ganancia (D, S) entre los atributos de R;

Sean $\{D_j \mid j=1, 2, \dots, m\}$ los atributos de D;

Sean $\{S_j \mid j=1, 2, \dots, m\}$ los subconjuntos de S correspondientes a los valores de D_j respectivamente;

⁴² Universidad Francisco de Paula Santander - Programa Ingeniería de Sistemas

Recurso Documental para Proyecto de Grado SEED "Simuladores para Estudio de Estructuras de Datos" Desarrolladores del Proyecto: Wilfred Uriel García - Cindy Yulieth Pabón Sánchez 2014.

²⁵ ING. Bruno Lopez Takeyas, Algoritmo C4.5 Nuevo Laredo, Tamaulipas, Noviembre del 2005.

Devolver un árbol con la raíz nombrada como D y con los arcos nombrados D_1, D_2, \dots, D_m que van respectivamente a los árboles C4.5 (R- {D}, C, S1), C4.5 (R- {D}, C, S2),..., C4.5 (R- {D}, C, Sm);

Fin

7.3.2 Desarrollo del código

A continuación se muestra el fragmento del código en java para calcular la entropía general del tratamiento.

```
private static double calcularEntropiaGeneral() {
    double positivos = 0.0;
    double negativos = 0.0;

    for (int i = 0; i < base_datos.length; i++) {
        if(base_datos[i][tratamiento].equals("SI")) {
            positivos++;
        }else if(base_datos[i][tratamiento].equals("NO")) {
            negativos++;
        }
    }

    double pos = (positivos/base_datos.length);
    double neg = (negativos/base_datos.length);

    return (-1 * pos) * (Math.log(pos)/Math.log(2)) - (neg) *
    Math.log(neg)/Math.log(2);
}
```

El siguiente fragmento de código calcula las tablas de ganancia y proporción de la ganancia de la información de cada uno de los síntomas.

```
private static LinkedList<Ganancia>
calcularTablaGanancias(String[][] dataBase) {
    int pos = 0;
    LinkedList<Ganancia> ganancias = new LinkedList<>();
    double entropiaGeneral = calcularEntropiaGeneral();

    while(pos < (base_datos[0].length - 5)) {

        ramas_temp = new LinkedList<>();
        obtenerListaRamas(pos);

        double[][] cantidades = new
double[ramas_temp.size()][2];
```

```

//Llenar la matriz de ceros
for (int i = 0; i < ramas_temp.size(); i++)
    for (int j = 0; j < 2; j++)
        cantidades[i][j] = 0;

for (int i = 0; i < ramas_temp.size(); i++) {
    for (int j = 0; j < dataBase.length; j++) {

if(dataBase[j][pos].equals(ramas_temp.get(i))){
        if(dataBase[j][tratamiento].equals("SI")){
            cantidades[i][0]++;
        }else
if(dataBase[j][tratamiento].equals("NO")){
            cantidades[i][1]++;
        }
    }
}

double[] entropias = new double [ramas_temp.size()];
double temp1, temp2, temp, cant;

for (int i = 0; i < ramas_temp.size(); i++) {
    cant = cantidades[i][0]+cantidades[i][1];

    temp1 = cantidades[i][0] / cant;
    temp2 = cantidades[i][1] / cant;
    temp = (-1 * temp1)*(Math.log(temp1)/Math.log(2))-
(temp2) * Math.log(temp2)/Math.log(2);

    if(Double.isNaN(temp))
        entropias[i] = 0.0;
    else
        entropias[i] = temp;
}

double gananciaSintoma = entropiaGeneral;
for (int i = 0; i < cantidades.length; i++) {
    gananciaSintoma -= (entropias[i] *
((cantidades[i][0]+cantidades[i][1])/dataBase.length));
}

double informacionDivision = 0.0;

for (int i = 0; i < cantidades.length; i++) {
    temp1 = (cantidades[i][0]+cantidades[i][1]) /
dataBase.length;
    temp = temp1*(Math.log(temp1)/Math.log(2));
    informacionDivision -= temp;
}

```

```

    }

    if (infoDetallada) {
        System.out.println("-----");
        System.out.println("DATOS ADICIONALES");
        System.out.println("");
        System.out.println("Entropía general: " +
entropiaGeneral);
        System.out.println("Síntoma: " + sintomas[pos]);
        System.out.println("");
        System.out.println("Tablas de cantidades");
        for (int i = 0; i < cantidades.length; i++) {
            System.out.print("-----\nRama: " +
ramas_temp.get(i) + "\nSI ( ");
            for (int j = 0; j < 2; j++) {
                System.out.print(cantidades[i][j] + " ");
            }
            System.out.println(") NO\n-----");
        }
        System.out.println("");

        for (int i = 0; i < ramas_temp.size(); i++)
            System.out.println("Entropía de " +
ramas_temp.get(i) + ":\t" + entropias[i]);
        System.out.println("");
        System.out.println("Ganancia de " + sintomas[pos]
+ ": "+gananciaSintoma);
        System.out.println("Información de la división: "
+ informacionDivision);
    }

    Ganancia nodoTemporal = new Ganancia(gananciaSintoma,
informacionDivision);
    ganancias.add(nodoTemporal);
    pos++;
}

imprimirTablaGanancias (ganancias);
return ganancias;
}

```

A continuación se muestra la función que crea una nueva base de datos a partir de la original dependiendo cual sea la condición del síntoma que se está tratando.

```

private static String[][] obtenerNuevaBaseDatos (String[][]
baseDatos, int col, int fil, int colOm, int i){
    String[][] db = new String[fil][col];

```

```

String[][] baseDatosOriginal = baseDatos;

int columna= 0;
int fila = 0;
for (int j = 0; j < baseDatosOriginal.length; j++){
    for (int k = 0; k < baseDatosOriginal[0].length; k++)
        if(k != colOm &&
baseDatosOriginal[j][colOm].equals(ramas_temp.get(i))) {
            db[fila][columna] = baseDatosOriginal[j][k];
            columna++;
        }

        columna = 0;

if(baseDatosOriginal[j][colOm].equals(ramas_temp.get(i)))
    fila++;
}

return db;
}

```

La siguiente función verifica si la proporción de la ganancia de información y la ganancia de información en el síntoma elegido tiene el mismo valor si es verdad la función devuelve un true son un false.

```

private boolean verificarCorte(LinkedList<Ganancia> lista, int
pos){
    if(lista.get(pos).getGanancia_de_informacion() -
lista.get(pos).getProporcion_de_ganancia() == 0)
        return true;
    return false;
}

```

La siguiente función busca entre la tabla de ganancia de información el síntoma con mayor valor de ganancia.

```

private int encontrarMejor(LinkedList<Ganancia> lista){
    int mayor;
    if(lista == null) return mayor = -1;
    mayor = 0;

    for (int i = 0; i < lista.size(); i++)
        if(lista.get(i).getGanancia_de_informacion() >
lista.get(mayor).getGanancia_de_informacion())
            mayor = i;

    return mayor;
}

```

La siguiente función busca la cantidad de veces que se está repitiendo un dato en una columna determinada.

```
private static int cantidadDatosRepetidos(int columna, String
dato){
    int cont = 0;
    for (int i = 0; i < base_datos.length; i++)
        if(base_datos[i][columna].equals(dato))
            cont++;

    return cont;
}
```

La siguiente función obtiene las posibles ramas que puede tener ese nodo.

```
private static void obtenerListaRamas(int sintoma){
    ramas_temp = new LinkedList<>();
    for(int i = 0; i < base_datos.length; i++)
        if(!ramas_temp.contains(base_datos[i][sintoma]))
            ramas_temp.add(base_datos[i][sintoma]);
}
```

El siguiente fragmento genera el árbol de decisión de forma recursiva usando las anteriores funciones.

```
public void generarArbol(){
    if(!esFinal()){
        int posicionMejorNodo;
        LinkedList<Ganancia> listaGanancias =
calcularTablaGanancias(base_datos);
        posicionMejorNodo = encontrarMejor(listaGanancias);

        contenido = sintomas[posicionMejorNodo];
        tipo = false;
        System.out.println("El mejor nodo es: " + contenido);
        System.out.println("-----
-----");

        obtenerListaRamas(posicionMejorNodo);
        imprimirRamas();

        for (int i = 0; i < ramas_temp.size(); i++) {
            if(!verificarCorte(listaGanancias,
posicionMejorNodo)){

                int filas =
cantidadDatosRepetidos(posicionMejorNodo,ramas_temp.get(i));
                int columnas = base_datos[0].length-1;
                int columnaOmitir = posicionMejorNodo;
```

```

        String[][] bdTemp =
obtenerNuevaBaseDatos(base_datos, columnas, filas, columnaOmitir, i);
        String[] sintomasTemp =
obtenerNuevosSintomas(columnaOmitir, sintomas);

        LinkedList<Nodo> originalHijos = hijos;
        String[][] originalDB = base_datos;
        String[] originalSintomas = sintomas;
        String rama = rama_padre;
        String nodo_p = nodo_padre;
        int originalTratamiento = tratamiento;

        Nodo temporal = new Nodo(bdTemp, sintomasTemp,
tratamiento - 1, infoDetallada, sintomas[posicionMejorNodo],
ramas_temp.get(i));
        temporal.generarArbol();
        hijos = originalHijos;
        rama_padre = rama;
        nodo_padre = nodo_p;
        base_datos = originalDB;
        sintomas = originalSintomas;
        tratamiento = originalTratamiento;
        hijos.add(temporal);
    }else{
        LinkedList<String> valores = new
LinkedList<>();

        for (int j = 0; j < base_datos.length; j++)

if(!valores.contains(base_datos[j][tratamiento]))

valores.add(base_datos[j][tratamiento]);

        int[] tam = new int[valores.size()];
        for (int j = 0; j < tam.length; j++) tam[j] =
0;

        for (int j = 0; j < base_datos.length; j++)
            for (int k = 0; k < valores.size(); k++)

if(valores.get(k).equals(base_datos[j][tratamiento]))
                tam[k]++;

        int mayor = 0;
        for (int j = 1; j < tam.length; j++) {
            if(tam[mayor] < tam[j])
                mayor = j;

```

```

        }

        tipo = true;

        contenido = valores.get(mayor);
        System.out.println("NODO FINAL: " +
contenido);

    }

    }else{
        tipo = true;
        contenido = base_datos[0][tratamiento];
        System.out.println("NODO FINAL: " + contenido);

    }

}

```

7.4 CAPÍTULO 4: RESULTADO DE PRUEBAS DEL PROTOTIPO DE FORMULACIÓN DE TRATAMIENTOS PARA LA FIBROSIS QUÍSTICA

Los resultados del prototipo desarrollado en Java para la formulación de tratamientos de la fibrosis quística son:

- Entropía.
- Ganancia de información.
- Proporción de la ganancia de información.
- Generación de la reglas de clasificación.

7.4.1 Resultados por tratamiento

Tratamiento 1

Primera iteración

Entropía general: 0.9709505944546686

Síntoma Peso

Entropía de ALTO: 0.0
Entropía de NORMAL: 0.0
Entropía de BAJO: 0.863120568566631

Síntoma Tos con expectoración de color

Entropía de SI: 1.0
Entropía de NO: 0.0

Síntoma Dolor abdominal

Entropía de NO: 0.9940302114769565

Entropía de SI: 0.8112781244591328

Síntoma Deposiciones anormales

Entropía de SI: 0.9709505944546687

Entropía de NO: 0.8812908992306927

Síntoma Pseudomona

Entropía de NO: 0.9910760598382221

Entropía de SI: 0.9182958340544896

Síntoma Pseudomona multi resistente

Entropía de SI: 0.7219280948873623

Entropía de NO: 1.0

Síntoma Estafilococo aerus

Entropía de NO: 0.9940302114769565

Entropía de SI: 0.0

Síntoma: Aspergelis aerus

Entropía de NO: 0.9852281360342516

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.5681609957902407	0.4415669800285151
Tos con expectoración de color	0.17095059445466854	0.23679725954056524
Dolor abdominal	0.025654272849131854	0.030663427637541305
Deposiciones anormales	0.05977313014931718	0.06509136591135878
Pseudomona	0.008986624929939457	0.009255491454729236
Pseudomona multi resistente	0.06364122949221451	0.06930362431377228
Estafilococo aerus	0.24199510603823393	0.2892461410339147
Aspergelis aerus	0.05140433415603374	0.1454732592614753

Tabla 37. Tabla de ganancia tratamiento 1 primera Iteración.

Fuente: Autores

Segunda Iteración

Entropía general: 0.863120568566631

Síntoma: Tos con expectoración de color

Entropía de SI: 0.6500224216483541

Entropía de NO: 0.0

Síntoma: Dolor abdominal

Entropía de NO: 0.9182958340544896

Entropía de SI: 0.0

Síntoma: Deposiciones anormales

Entropía de NO: 0.9709505944546687

Entropía de SI: 0.0

Síntoma: Pseudomona

Entropía de SI: 1.0

Entropía de NO: 0.0

Síntoma: Pseudomona multi resistente

Entropía de NO: 0.863120568566631

Entropía de SI: 0.0

Síntoma: Estafilococo aerus

Entropía de NO: 0.6500224216483541

Entropía de SI: 0.0

Síntoma: Aspergelis aerus

Entropía de NO: 0.6500224216483541

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Tos con expectoración de color	0.3059584928680418	0.517107603971795
Dolor abdominal	0.076009853662782	0.12846603124941064
Deposiciones anormales	0.1695844296704390	0.1964782625352851
Pseudomona	0.2916919971380596	0.29606543547586917
Pseudomona multi resistente	0	0
Estafilococo aerus	0.3059584928680418	0.517107603971795
Aspergelis aerus	0.3059584928680418	0.517107603971795

Tabla 38. Tabla de ganancias tratamiento 1 segunda iteración
Fuente: Autores

Tercer Iteración

Entropía general: 0.6500224216483541

Síntoma: Dolor abdominal

Entropía de NO: 0.7219280948873623

Entropía de SI: 0.0

Síntoma: Deposiciones anormales

Entropía de NO: 0.8112781244591328

Entropía de SI: 0.0

Síntoma: Pseudomona

Entropía de SI: 0.9182958340544896

Entropía de NO: 0.0

Síntoma: Pseudomona multi resistente

Entropía de NO: 0.6500224216483541

Entropía de SI: 0.0

Síntoma: Estafilococo aerus

Entropía de NO: 0.6500224216483541

Entropía de SI: 0.0

Síntoma: Aspergelis aerus

Entropía de NO: 0.6500224216483541

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Dolor abdominal	0.048415675908885514	0.07448308596203652
Deposiciones anormales	0.10917033867559889	0.1188836261987452
Pseudomona	0.19087450462110933	0.19087450462110933
Pseudomona multi resistente	0	0
Estafilococo aerus	0	0
Aspergelis aerus	0	0

Tabla 39. Tabla de ganancias tratamiento 1 tercera iteración.

Fuente: Autores

=====ÁRBOL GENERADO=====

Nodo padre: null

Rama padre: null

Nodo: Peso

Ramas:

* NORMAL

* ALTO

* BAJO

 Nodo padre: Peso

Rama padre: NORMAL

Nodo Final: NO

 Nodo padre: Peso

Rama padre: ALTO

Nodo Final: SI

 Nodo padre: Peso

Rama padre: BAJO

Nodo: Tos con expectoración de color

Ramas:

* SI

* NO

Nodo padre: Tos con expectoración de color
Rama padre: SI
Nodo: Pseudomona

Ramas:

* SI

* NO

Nodo padre: Pseudomona
Rama padre: SI

Nodo Final: SI

Nodo padre: Pseudomona
Rama padre: NO

Nodo Final: SI

Nodo padre: Tos con expectoración de color
Rama padre: NO

Nodo Final: NO

Según los datos obtenidos por el simulador el árbol generado para el tratamientos 1 es el siguiente.

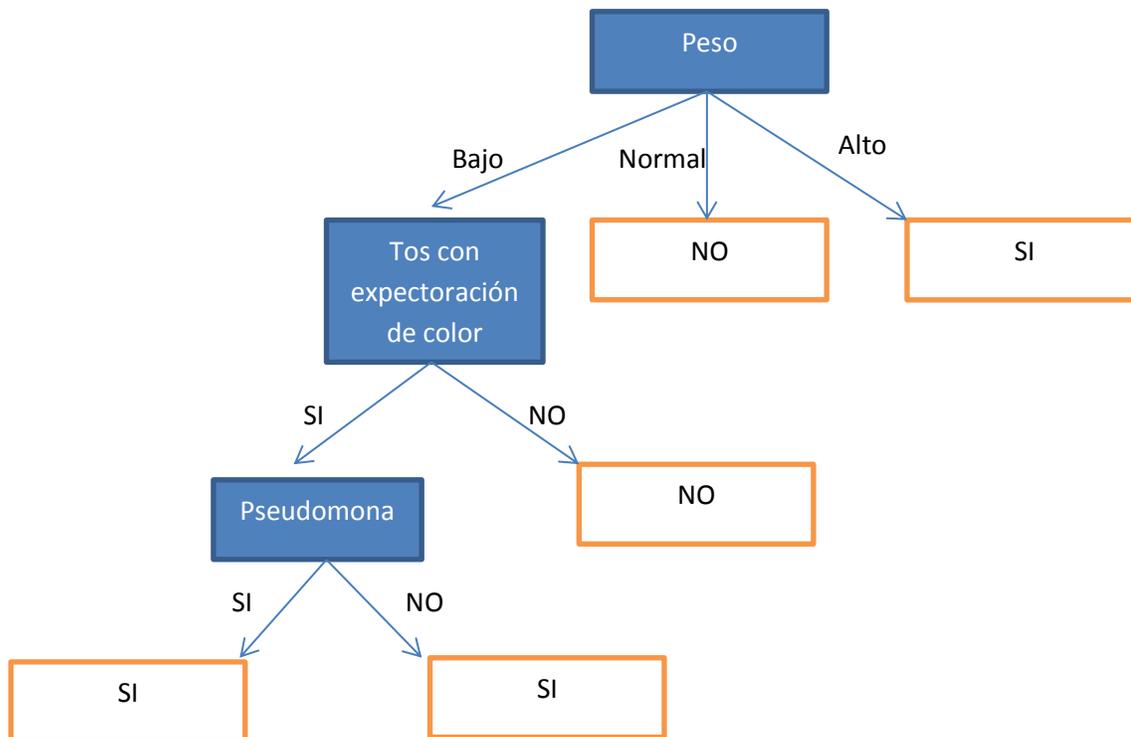


Figura 9. Árbol de decisión tratamiento 1.
Fuente: Autores

Tratamiento 2

Primer Iteración

Entropía general: 0.8366407419411673

Síntoma: Peso

Entropía de ALTO: 0.0
 Entropía de NORMAL: 0.5916727785823275
 Entropía de BAJO: 0.9852281360342516

Síntoma Tos con expectoración de color

Entropía de SI: 0.9182958340544896
 Entropía de NO: 0.0

Síntoma Dolor abdominal

Entropía de NO: 0.9456603046006402

Entropía de SI: 0.0

Síntoma Deposiciones anormales

Entropía de SI: 0.0

Entropía de NO: 0.9709505944546686

Síntoma Pseudomona

Entropía de NO: 0.0

Entropía de SI: 0.9182958340544896

Síntoma Pseudomona multi resistente

Entropía de SI: 0.0

Entropía de NO: 0.9709505944546686

Síntoma Estafilococo aerus

Entropía de NO: 0.9456603046006402

Entropía de SI: 0.0

Síntoma: Aspergelis aerus

Entropía de NO: 0.9852281360342516

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.10075364845343049	0.0783043619750009
Tos con expectoración de color	0.10200407469757566	0.14129395353908575
Dolor abdominal	0.14315651856736455	0.17110871057416344
Deposiciones anormales	0.18934034563805502	0.20618665425288207
Pseudomona	0.4693224083193715	0.48336384054944115
Pseudomona multi resistente	0.18934034563805502	0.20618665425288207
Estafilococo aerus	0.14315651856736455	0.17110871057416344
Aspergelis aerus	0.031061544612311698	0.08790356312626829

Tabla 40. Tabla de ganancias tratamiento 2 primera iteración

Fuente: Autores

Segunda Iteración

Entropía general: 0.9182958340544896

Síntoma: Peso

Entropía de ALTO: 0.0

Entropía de NORMAL: 1.0

Entropía de BAJO: 0.8112781244591328

Síntoma Tos con expectoración de color

Entropía de SI: 0.7219280948873623

Entropía de NO: 0.0

Síntoma Dolor abdominal

Entropía de NO: 0.9182958340544896

Entropía de SI: 0.0

Síntoma Deposiciones anormales

Entropía de SI: 0.0

Entropía de NO: 0.9182958340544896

Síntoma Pseudomona multi resistente

Entropía de SI: 0.0

Entropía de NO: 0.9182958340544896

Síntoma Estafilococo aerus

Entropía de SI: 0.0

Entropía de NO: 0.0

Síntoma: Aspergelis aerus

Entropía de NO: 0.7219280948873623

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.044110417748401076	0.04803508424256193
Tos con expectoración de color	0.31668908831502096	0.4871971762327021
Dolor abdominal	0	0
Deposiciones anormales	0	0
Pseudomona multi resistente	0	0
Estafilococo aerus	0.9182958340544896	1
Aspergelis aerus	0.31668908831502096	0.4871971762327021

Tabla 41. Tabla de ganancias tratamiento 2 segunda iteración.

Fuente: Autores

=====ÁRBOL GENERADO=====

Nodo padre: null
 Rama padre: null
 Nodo: Pseudomona

Ramas:

- * NO
- * SI

 Nodo padre: Pseudomona
 Rama padre: NO

Nodo Final: NO

 Nodo padre: Pseudomona
 Rama padre: SI
 Nodo: Estafilococo aerus

Ramas:

- * SI
- * NO

 Nodo padre: Estafilococo aerus
 Rama padre: SI

Nodo Final: NO

Nodo padre: Estafilococo aerus
 Rama padre: NO

Nodo Final: SI

Según los datos obtenidos por el simulador el árbol generado para el tratamiento 2 es el siguiente.

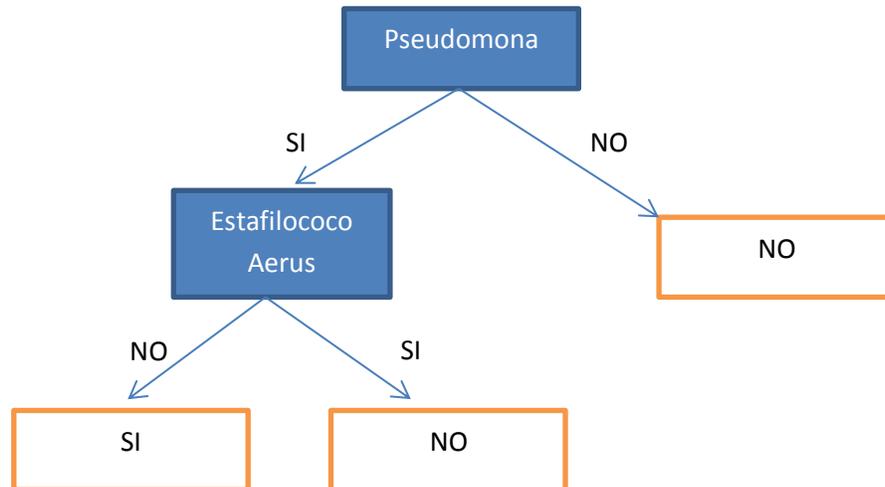


Figura 10. Árbol de decisión Tratamiento 2.

Fuente: Autores

Tratamiento 3

Primer Iteración

Entropía general: 0.9182958340544896

Síntoma: Peso

Entropía de ALTO: 0.0

Entropía de NORMAL: 0.9852281360342516

Entropía de BAJO: 0.0

Síntoma Tos con expectoración de color

Entropía de SI: 0.9182958340544896

Entropía de NO: 0.9182958340544896

Síntoma Dolor abdominal

Entropía de NO: 0.8453509366224365

Entropía de SI: 1.0

Síntoma Deposiciones anormales

Entropía de SI: 0.9709505944546686

Entropía de NO: 0.8812908992306927

Síntoma Pseudomona

Entropía de NO: 0.9910760598382222

Entropía de SI: 0.0

Síntoma Pseudomona multi resistente

Entropía de SI: 0.0

Entropía de NO: 0.0

Síntoma Estafilococo aerus

Entropía de NO: 0.9456603046006402

Entropía de SI: 0.8112781244591328

Síntoma: Aspergelis aerus

Entropía de NO: 0.9402859586706309

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.45852270390517214	0.35635759430530356
Tos con expectoración de color	-2.7755575615628914E-17	-3.844645444912271E-17
Dolor abdominal	0.03170514719803613	0.03789577247275108
Deposiciones anormales	0.007118369749138265	0.007751717349853377
Pseudomona	0.32365019815155627	0.3333333333333334
Pseudomona multi resistente	0.9182958340544896	1
Estafilococo aerus	0.008470777491584702	0.010124748971620566
Aspergelis aerus	0.040695605961900694	0.1151677681288189

Tabla 42. Tabla de ganancias tratamiento 3 primer iteración.

Fuente: Autores

=====ÁRBOL GENERADO=====

Nodo padre: null
 Rama padre: null
 Nodo: Pseudomona multi resistente

Ramas:
 * NO
 * SI

 Nodo padre: Pseudomona multi resistente
 Rama padre: NO

Nodo Final: NO

 Nodo padre: Pseudomona multi resistente
 Rama padre: SI

Nodo Final: SI

 Según los datos obtenidos por el simulador el árbol generado para el tratamiento 3 es el siguiente.

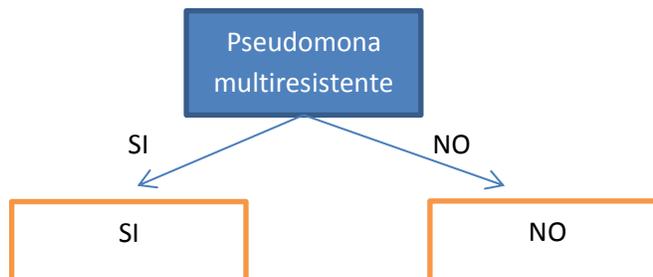


Figura 11. Árbol decisión del tratamiento 3.

Fuente: Autores

Tratamiento 4

Primer Iteración

Entropía general: 0.7219280948873623

Síntoma: Peso

Entropía de ALTO: 0.0
Entropía de NORMAL: 0.863120568566631
Entropía de BAJO: 0.5916727785823275

Síntoma Tos con expectoración de color

Entropía de SI: 0.6500224216483541
Entropía de NO: 0.9182958340544896

Síntoma Dolor abdominal

Entropía de NO: 0.8453509366224365
Entropía de SI: 0.0

Síntoma Deposiciones anormales

Entropía de SI: 0.0
Entropía de NO: 0.8812908992306927

Síntoma Pseudomona

En Entropía de NO: 0.5032583347756457
Entropía de SI: 0.9182958340544896

Síntoma Pseudomona multi resistente

Entropía de SI: 0.7219280948873623
Entropía de NO: 0.7219280948873623

Síntoma Estafilococo aerus

Entropía de NO: 0.0
Entropía de SI: 0.8112781244591328

Síntoma: Aspergelis aerus

Entropía de NO: 0.5916727785823275
Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.04302453288451502	0.033438080392211196
Tos con expectoración de color	0.01825099075778111	0.025280898315266
Dolor abdominal	0.10200407469757555	0.1219209985649354
Deposiciones anormales	0.13440082873356718	0.1463589659773978
Pseudomona	0.052654760400179024	0.05423011294385417
Pseudomona multi resistente	0	0
Estafilococo aerus	0.5055872616982602	0.6043062886528817
Aspergelis aerus	0.1697001682105233	0.4802481536259281

Tabla 43. Tabla de ganancias tratamiento 4 primer iteración

Fuente: Autores

Segunda Iteración

Entropía general: 0.8112781244591328

Síntoma: Peso

Entropía de ALTO: 0.0

Entropía de NORMAL: 0.9182958340544896

Entropía de BAJO: 0.0

Síntoma Tos con expectoración de color

Entropía de NO: 1.0

Entropía de SI: 0.0

Síntoma Dolor abdominal

Entropía de SI: 0.0

Entropía de NO: 0.0

Síntoma Deposiciones anormales

Entropía de SI: 0.0

Entropía de NO: 0.0

Síntoma Pseudomona

Entropía de NO: 1.0

Entropía de SI: 0.0

Síntoma Pseudomona multi-resistente

Entropía de NO: 0.9182958340544896

Entropía de SI: 0.0

Síntoma Estafilococo aerus

Entropía de NO: 0.0

Entropía de SI: 0.8112781244591328

Síntoma: Aspergelis aerus

Entropía de NO: 0.9182958340544896

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.12255624891826566	0.15106563978903303
Tos con expectoración de color	0.31127812445913283	0.31127812445913283
Dolor abdominal	0.8112781244591328	1
Deposiciones anormales	0.8112781244591328	1
Pseudomona	0.31127812445913283	0.31127812445913283
Pseudomona multi resistente	0.12255624891826566	0.15106563978903303
Aspergelis aerus	0.12255624891826566	0.15106563978903303

Tabla 44. Tablas de ganancia tratamiento 4 segunda iteración.

Fuente: Autores

=====ÁRBOL GENERADO=====

Nodo padre: null

Rama padre: null

Nodo: Estafilococo aerus

Ramas:

* SI

* NO

Nodo padre: Estafilococo aerus

Rama padre: SI

Nodo: Dolor abdominal

Ramas:

* SI

* NO

Nodo padre: Dolor abdominal
Rama padre: SI

Nodo Final: NO

Nodo padre: Dolor abdominal
Rama padre: NO

Nodo Final: SI

Nodo padre: Estafilococo aerus
Rama padre: NO

Nodo Final: NO

Según los datos obtenidos por el simulador el árbol generado para el tratamiento 4 es el siguiente.

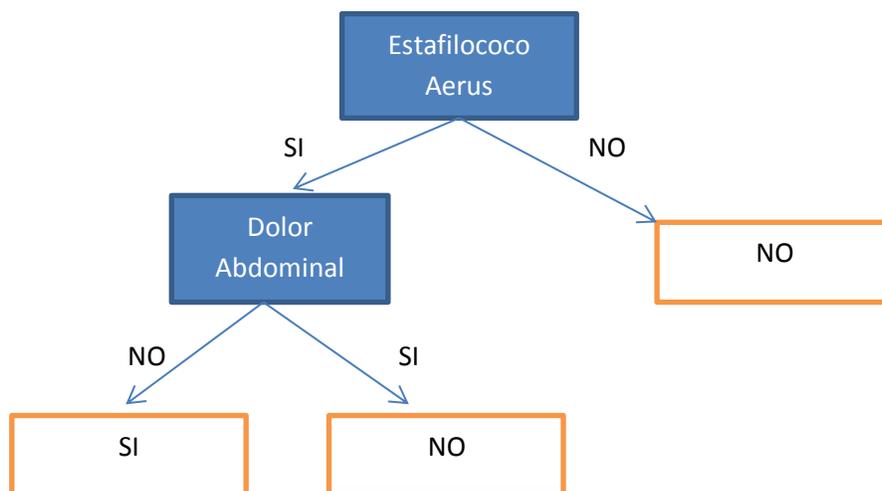


Figura 12. Árbol de decisión Tratamiento 4.
Fuente: Autores

Tratamiento 5

Primer Iteración

Entropía general: 0.9709505944546686

Síntoma: Peso

Entropía de ALTO: 0.0

Entropía de NORMAL: 0.9852281360342516

Entropía de BAJO: 0.863120568566631

Síntoma Tos con expectoración de color

Entropía de SI: 0.9182958340544896

Entropía de NO: 0.9182958340544896

Síntoma Dolor abdominal

Entropía de NO: 0.6840384356390417

Entropía de SI: 0.0

Síntoma Deposiciones anormales

Entropía de SI: 0.0

Entropía de NO: 0.4689955935892812

Síntoma Pseudomona

Entropía de NO: 0.9182958340544896

Entropía de SI: 0.0

Síntoma Pseudomona multi-resistente

Entropía de SI: 0.9709505944546687

Entropía de NO: 0.8812908992306927

Síntoma Estafilococo aerus

Entropía de NO: 0.9940302114769565

Entropía de SI: 0.8112781244591328

Síntoma: Aspergelis aerus

Entropía de NO: 0.9852281360342516

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.10838786564092334	0.08423757149367675
Tos con expectoración de color	0.05265476040017899	0.07293629486520312
Dolor abdominal	0.46932240831937144	0.5609604992824677
Deposiciones anormales	0.6582868653951478	0.7168570747932703
Pseudomona	0.4199730940219749	0.4325380677663126
Pseudomona multi resistente	0.05977313014931718	0.06509136591135878
Estafilococo aerus	0.025654272849131854	0.030663427637541305
Aspergelis aerus	0.05140433415603374	0.1454732592614753

Tabla 45. Tabla de ganancias tratamiento 5 primer iteración.

Fuente: Autores

Segunda Iteración

Entropía general: 0.4689955935892812

Síntoma: Peso

Entropía de ALTO: 0.0

Entropía de NORMAL: 0.7219280948873623

Entropía de BAJO: 0.0

Síntoma Tos con expectoración de color

Entropía de SI: 0.5032583347756457

Entropía de NO: 0.0

Síntoma Dolor abdominal

Entropía de NO: 0.0

Entropía de SI: 0.0

Síntoma Pseudomona

Entropía de SI: 0.0

Entropía de NO: 0.8112781244591328

Síntoma Pseudomona multi resistente

Entropía de NO: 0.0

Entropía de SI: 0.9182958340544896

Síntoma Estafilococo aerus

Entropía de SI: 0.0

Entropía de NO: 0.5916727785823275

Síntoma: Aspergelis aerus

Entropía de NO: 0.9852281360342516

Entropía de SI: 0.0

TABLA DE GANANCIAS		
Síntoma	Ganancia	Proporción
Peso	0.10803154614560007	0.10803154614560007
Tos con expectoración de color	0.016063092291200065	0.03424998552388784
Dolor abdominal	0.4689955935892812	1
Pseudomona	0.14448434380562808	0.14880710164946886
Pseudomona multi resistente	0.1935068433729344	0.21957204317195692
Estafilococo aerus	0.054824648581652036	0.06220948001336476
Aspergelis aerus	0.016063092291200065	0.03424998552388784

Tabla 46. Tabla de ganancias tratamientos 5 segunda iteración

Fuente: Autores

=====ÁRBOL GENERADO=====

Nodo padre: null

Rama padre: null

Nodo: Depositiones anormales

Ramas:

* SI

* NO

Nodo padre: Depositiones anormales

Rama padre: SI

Nodo Final: SI

Nodo padre: Depositiones anormales

Rama padre: NO

Nodo: Dolor abdominal

Ramas:

* NO

* SI

Nodo padre: Dolor abdominal

Rama padre: NO

Nodo Final: NO

Nodo padre: Dolor abdominal

Rama padre: SI

Nodo Final: SI

Según los datos obtenidos por el simulador el árbol generado para el tratamiento 5 es el siguiente

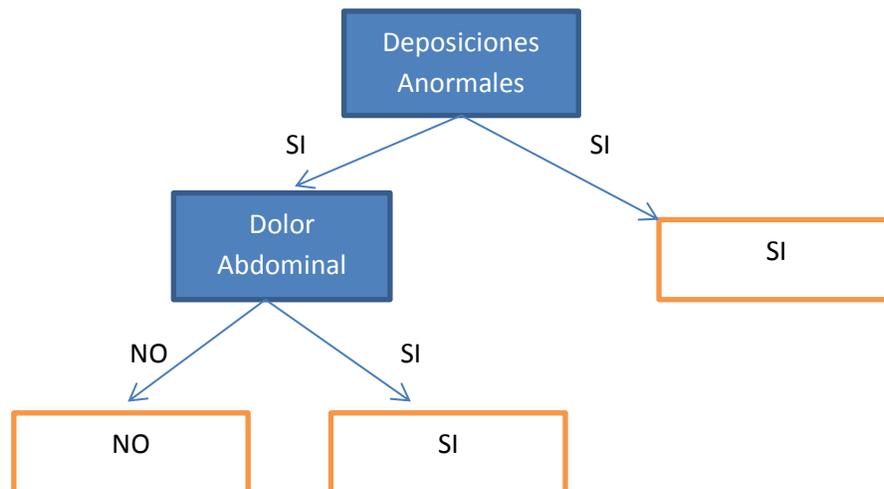


Figura 13. Árbol de decisión Tratamiento 5.

Fuente: Autores

7.4.2 Generación de reglas de clasificación

Según los gráficos 9, 10, 11, 12, 13 se puede observar que para aplicar a un paciente uno de los cinco tratamientos debemos tener en cuenta las siguientes reglas generadas.

Se puede administrar el tratamiento 1, Si:

- El paciente tiene bajo peso, tiene tos con expectoración de color y no tiene pseudomona.
- El paciente tiene bajo peso tiene tos con expectoración de color y tiene pseudomona.
- El paciente tiene un peso alto.

No se administra el tratamiento 1, Si:

- El paciente tiene un peso normal.
- El paciente tiene bajo peso pero no tiene tos con expectoración de color.

Se puede administrar el tratamiento 2, Si:

- El paciente tiene Pseudomona y no tiene estafilococo aerus.

No se administra el tratamiento 2, Si:

- El paciente no tiene pseudomona.
- El paciente tiene Pseudomona y tiene estafilococo aerus.

Se puede administrar el tratamiento 3, Si:

- El paciente tiene pseudomona multi resistente.

No se administra el tratamiento 3, Si:

- El paciente no tiene pseudomona multi resistente.

Se puede administrar el tratamiento 4, Si:

- El paciente tiene estafilococo aerus y tiene dolor abdominal.

No se administra el tratamiento 4, Si:

- El paciente tiene estafilococo aerus y tiene dolor abdominal.
- El paciente no tiene estafilococo aerus.

Se puede administrar el tratamiento 5, Si:

- El paciente tiene deposiciones anormales y tiene dolor abdominal.
- El paciente tiene deposiciones anormales.

No se administra el tratamiento 5, Si:

- El paciente tiene deposiciones anormales y no tiene dolor abdominal.

8. CONCLUSIONES

- A partir del análisis hecho con el especialista y los datos de los pacientes, se encontraron cinco tipos de tratamientos diferentes. Lo que provoca que se generen cinco árboles de decisión con diferentes combinaciones resultantes.
- La estructura de los arboles está dada por la cantidad de datos usados en la muestra. Nuestra muestra, fue limitada, ya que la accesibilidad a las historias clínicas de los pacientes con fibrosis quística no es muy sencilla y no hay facilidades de acceso a esta información por parte de los médicos, ya que tratan de mantener protegida la confidencialidad de sus pacientes.
- En la ejecución del prototipo existió un caso donde tres pacientes tenían los mismos síntomas; a dos se le aplicó el tratamiento y al otro no, por lo que se podría concluir que la probabilidad para aplicar ese tratamiento en ese caso sería del 66.66%. Lo que nos hace ver que el error humano existe, dándole al software la posibilidad de elegir como mejor opción el valor de mayor probabilidad.
- Se encontraron algunos errores en los cálculos manuales de los datos hechos por los autores, los cuales, fueron comparados con los resultados obtenidos por el prototipo de la herramienta desarrollada, y se re-verificaron estos datos y se comprobó que el error humano existe a la hora de formular tratamientos.
- Se demostró que la herramienta tiene cálculos mucho más precisos a los que podemos calcular usualmente los humanos. Dando así mayor precisión y una mejor estructura al árbol de decisión.

9. IMPACTOS ESPERADOS

Este proyecto, está desarrollado como un prototipo para la formulación de tratamientos para la fibrosis quística, usando arboles de decisión como metodología de desarrollo, por lo que se espera en el futuro se realicen más investigaciones de este tipo tratando de mejorar y profundizar en los resultados actuales usando otro tipo de metodologías como redes bayesianas, redes neuronales entre otras líneas de desarrollo de la inteligencia artificial y minería de datos.

Se deja una serie de códigos para la educación superior que pueden servir como base para desarrollos futuros para personas que quieran continuar con esta investigación.

Además se espera que este proyecto, despierte el interés en la comunidad de los profesionales de la salud, para que se animen y apoyen la investigación y desarrollo de nuevas tecnologías, que les puedan dar mejores soportes a la hora de formular tratamientos de diferentes tipos de patologías, en la cuales estas nuevas tecnologías ayudarán y facilitarán su trabajo. Mencionando también que el desarrollo de nuevas tecnologías puede ayudar a los pacientes que sufren alguna afección, tener una vida más llevadera o incluso salvar sus vidas al tener herramientas de mayor precisión evitando el error humano.

10. DIVUGACIONES

JIMMY ANDRÉS LEÓN MUÑOZ, MIGUEL ÁNGEL ACEVEDO FRANCO, “Formulación para tratamientos adecuados a la fibrosis quística utilizando árboles de decisión” Categoría B. Revista ciencia e ingeniería Neogranadina. EN EVALUACIÓN. Fecha de entrega: 14 de Abril, 2015

11. BIBLIOGRAFÍA

- [1] Accrue Software. An Introduction to OLAP Multidimensional Terminology and Technology. [online] Available: http://www.accrue.com/olap/wp_intro_olap.pdf
- [2] APALEX, “normas de presentación de caso clínico”, Internet: <http://www.apalex.es/uploads/documentos/NORMAS_PRESENTACION_CASOS_CLINICOS_apalex.pdf>
- [3] Breiman, L., Friedman, J., Olshen, R. & Stone, C. (1984), Classification and Regression Trees, Wadsworth. Belmont.
- [4] Campell, Mary. Base IV Guía de Auto enseñanza. España. Editorial McGraw Hill Interamericana. 1990. pp110/111,121/122,16,169, 179-191/192. (4 Mar 2009).
- [5] Cuevas Agustín, Gonzalo, "Teoría de la información, codificación y lenguajes", Ed. SEPA (Sociedad para Estudios Pedagógicos Argentinos), Serie Informática 1986
- [6] Davies J, Bush A. Infection in patients with cystic fibrosis and congenital immune deficiencies. Curr Opin Infect Dis 1997; 10:268-74
- [7] DMEDICINA, enfermedades respiratorias [en línea] <www.dmedicina.com/enfermedades/respiratorias/fibrosis-quistica> [citado en 25 de octubre 2013]
- [8] DRA. CATALINA VÁSQUEZ, DR. RICARDO ARISTIZÁBAL, DR. WILSON DAZA, “Fibrosis quitica en Colombia”. Internet: <<http://www.neumologia-pediatria.cl>>
- [9] DMAE, arboles binarios [en línea] <http://www.dmae.upct.es/~mcr Ruiz/Telem06/Teoria/arbol_decision.pdf> [citado en 25 de octubre 2013]
- [10] EDGAR LEONEL GOMEZ NARCISO, “Estadística, Matemática y Computación”. Universidad rural de Guatemala, maestría en investigación y proyectos. Internet: <<http://reyesestadistica.blogspot.com/2011/07/muestreo-simple-aleatorio.html>>
- [11] EMILIO SORIA, ANTONIO JOSÉ SERRANO Y JOSÉ DAVID MARTÍN “Arboles de decisión”. Dpto. Ingeniería Electrónica, Internet: <http://ocw.uv.es/ciencias-de-la-salud/pruebas/1-2/1tema_6_ocw.pdf>
- [12] FEDERACION ESPAÑOLA CONTRA LA FIBROSIS QUISTICA, Galenas, D.L: M ISBN: 101 preguntas de la fibrosis quística [Consulta: 26 octubre 2013] Disponible en: <http://www.fibrosisquistica.org/images/recursos/163.pdf>

- [13] FEDERACION ESPAÑOLA CONTRA LA FIBROSIS QUISTICA [en línea]
<www.fibrosisquistica.org> [citado en 25 de octubre 2013]
- [14] FEDERACION ESPAÑOLA CONTRA LA FIBROSIS QUISTICA, Galenas, D.L: ISBN: Los tres pilares del tratamiento en fibrosis quística [Consulta: 26 octubre 2013] Disponible en:
<http://www.fibrosisquistica.org/images/recursos/123.pdf>
- [15] FEDERACION ESPAÑOLA CONTRA LA FIBROSIS QUISTICA, Galenas, D.L: V, ISBN: Manual de Urgencias de la fibrosis quística [Consulta: 26 octubre 2013] Disponible en:
<http://www.fibrosisquistica.org/images/recursos/194.pdf>
- [16] FEDERACION ESPAÑOLA CONTRA LA FIBROSIS QUISTICA, libro blanco de atención de la fibrosis quística [Consulta: 26 octubre 2013] Disponible en:
<http://www.fibrosisquistica.org/images/recursos/31.pdf>
- [17] GALEON, análisis de decisiones [en línea]
<<http://alcemaestriarrhh.galeon.com/productos2219556.html>> [citado en 25 de octubre 2013]
- [18] Gispert SS. Identificado el gen de la fibrosis quística. RESUMED 1991; 4(1):3-6
- [19] Goglino D. Minería de datos. [Online] Available:
<http://www.infonews21.com/columnas/goglino/goglino.htm>
- [20] Griffin J. Transcutaneous electrical nerve stimulation, en Ramamurthy S y Rogers J. Decision making in Pain Management. Mosby-Year Book Inc. 194-195.
- [21] Grupo Técnico, Programa de Fibrosis Quística Unidad de Salud Respiratoria Subsecretaría de Redes Asistenciales Ministerio de Salud, “Programa nacional de Fibrosis Quística orientaciones programáticas para diagnóstico y tratamientos 2012”. Internet:
<http://respiratorio.minsal.cl/PDF/FIBROSIS/ADULTO/Guia_Clinica_FQ_2012.pdf>
- [22] Guillermo Solarte Arboles de decisiones en el diagnóstico de enfermedades cardiovasculares revista Scientia et Technica Año XVI, No 49, Diciembre de 2011. Universidad Tecnológica de Pereira. ISSN 0122-1701
- [23] Guillermo Solarte Martínez, “Evaluar la utilidad de la metodología de redes bayesianas en la predicción y diagnóstico de enfermedades (cardiovasculares)” Trabajo de grado Ingeniero de sistemas y computación. Pereira: Universidad Tecnológica de Pereira. Facultad de ingenierías. Departamento de Ingeniería de Sistemas y computación.
- [24] INFORMATICA INTEGRAL INTELIGENTE, Sistemas expertos [en línea]
<<http://www.informaticaintegral.net/sisexp.html>>
- [25] ING. Bruno Lopez Takeyas, Algoritmo C4.5 Nuevo Laredo, Tamaulipas, Noviembre del 2005.

- [26] INVESTIGACION DE OPERACIONES [en línea] <www.investigacion-operaciones.com/Curso_inv-Oper_carpeta/Clase24_II.pdf> [citado en 25 de octubre 2013]
- [27] J. C. Aguado Franco “Teoría de la decisión y los juegos” Mecaservi, Lucas de la Heras, S.L.L c/serrano 41Pta, 6 of 6, 28001 España Madrid España.
- [28] Larose, Daniel T. (2005). *Discovering Knowledge in Data an Introduction to Data Mining*. Hoboken, New Jersey. John Wiley & Sons, Inc Publication. 222p.
- [29] MEDLINEPLUS, fibrosis quística [en línea] <<http://www.nlm.nih.gov/medlineplus/spanish/ency/article/000107.htm>> [citado en 25 de octubre 2013]
- [30] MICROSOFT, Minería de datos [en línea] <<http://msdn.microsoft.com/es-es/library/ms174949.aspx>>
- [31] PROFESORES, arboles binarios [en línea] <<http://profesores.elo.utfsm.cl/~tarredondo/info/datos-algoritmos/ELO-320%20Arboles%20binarios.pdf>>
- [32] REINEL ARÁIS MONTOYA, “Detección temprana de fallas en la red de internet banda ancha aplicando minería de datos” Trabajo de grado Ingeniero de sistemas y computación. Pereira: Universidad Tecnológica de Pereira. Facultad de ingenierías. Departamento de Ingeniería de Sistemas y computación.
- [33] RENE RODRÍGUEZ MARÍN, “Determinación de incidencia de la mutación af508 del gen cftr, asociado a la fibrosis quística en la población del departamento de Risaralda” Tesis especialización en biología molecular y biotecnología. Pereira: Universidad Tecnológica de Pereira. Facultad de Medicina.
- [34] R. L-Beltra. “Bioinformática Simulación, vida artificial e inteligencia artificial”, ediciones Días de Santos, S.A Albazans., 2807 Madrid ISBN: 97-84-7978-181-1
- [35] Salvatore Ruggieri, “Efficient C4.5” Departamento de informática Universidad de Pisa <[idb.csie.ncku.edu.tw/tsengsm/COURSE/DM/Paper/ec45.pdf](http://db.csie.ncku.edu.tw/tsengsm/COURSE/DM/Paper/ec45.pdf)>
- [36] Schivab VE, Wold AE, Carson JL, Leigh MW, Cheng P, Gilligan PH, et al. Increased adherence of *Staphylococcus aureus* from cystic fibrosis lungs to airway epithelial cells. *Am Rev Respir Dis* 1993; 148:365-9.
- [37] Stenvang PS, Hoiby N, Shand GH, Fressler T. Antibody response to *Pseudomonas aeruginosa* antigens in cystic fibrosis. *Antibiot Chemoter* 1989; 82:130-53.
- [38] THE AMERICAN CONGRESS OF OBSTETRICIANS AND GYNECOLOGIST, “La fibrosis quística: pruebas de detección y diagnóstico prenatal”. Internet:

<http://www.acog.org/For_Patients/Search_Patient_Education_Pamphlets_-_Spanish/Files/La_fibrosis_quistica-_Pruebas_de_deteccion_y_diagnostico_prenatal>

[39] UHU, introducción a la teoría de decisiones [en línea]
<http://www.uhu.es/eyda.marin/apuntes/admon/tema6_II.pdf> [citado en 25 de octubre 2013]

[40] UNIVERSIDAD COMPLUTENSE DE MADRID, Teoría de la decisión [en línea]
<http://www.mat.ucm.es/~bvitoria/Archivos/a_dt_UCM.pdf> [citado el 29 de octubre 2013]

[41] ZURAY FERNANDEZ CORREDOR, "Frecuencia de la mutación AF508 en portadores y enfermos con fibrosis quística del Sur-Occidente colombiano". Tesis (Biología con mención en genética). Cali: Universidad del Valle. Facultad de ciencias.

[42] Universidad Francisco de Paula Santander - Programa Ingeniería de Sistemas
Recurso Documental para Proyecto de Grado SEED "Simuladores para Estudio de Estructuras de Datos" Desarrolladores del Proyecto: Wilfred Uriel García - Cindy Yulieth Pabón Sánchez
2014.

12. ANEXOS

ANEXO A

Historias clínicas entregadas por la especialista.

N° paciente	edad	peso(kg)	talla (ms)	IMC	antecedentes de medicamentos	sintomas	examen fisico	impresión diagnóstica	formula medica
1	9	22	122	14,78	* dornasa alfa * lipasas 25000 * pediasure * vitaminas a,d,e,k * iontoforosis	* dolor abdominal * deposiciones pocas formadas * fibrosis hepáticas * estafilococo	* hipocratismos digital * lengua geográfica * otoscopia anormal * auscultación cardiopulmonar * hernia umbilical pequeña	* Fibrosis quística con manifestaciones pulmonares * Fibrosis quística con manifestaciones intestinales	* Acido ursodesoxicólico X 50 MG tabletas #60 tomar una al día por dos meses * Ranitidina x 150 MG tabletas #60 tomar una cada 12 horas
2	9	24,4	130	14,44	* clindamicina * ranitidina * vitaminas a,d,e,k * ciprofloxacina	* tos seca * sobreinfección por pseudomonas aeruginosa y estafilococo	* hipertrofia severa de los cornetes * cerumen impide ver timpano izq	* Fibrosis quística con manifestaciones pulmonares * Fibrosis quística con manifestaciones intestinales	* Trimetoprim x 240MG suspensión frasco #4 tomar 14cc cada 12 horas por 21 días * ranitidina x 150 MG tabletas #270 tomar una cada día por 3 meses * tobramicina polvo

									para inhalacion en capsula dura (TOBY PODHALER) x 28MG #224 para inhalar cada 12 horas por 28 dias
3	9	25	133	14,13	* lipasa pancreatica *pediasure *nebulizaciones tobramicina	*sin tos *deposiciones normales *sobreinfección pseudomona y estilococo controlada	* hipertrofia severa de los cornetes	* Fibrosis quística con manifestaciones pulmonares * Fibrosis quística con manifestaciones intestinales	* sales de rehidratación oral x 45 Meq frasco #90 tomar un frasco al dia por 3 meses
4	14	42,7	142	21,18	* cefatizidima *atreonam *amikacina *toma antibioticos para sinusitis *tres cirugias por sinusitis	*sinusitis severa *pseudomona	*afebril no tiene *rinoscopia con buana permebilidad *otoscopia normal *auscultacion pulmonar normal *abdomen bien	*Fibrosis quística con otras manifestaciones	* aztreonam *rinofluimicil *tobramicina nebulizada
5	14	42,5	142	21,18	*mereon en *amikacina *cefepim	*tos *especto racion de color verde	* hipertrofia severa de los cornetes	*Fibrosis quística con otras manifestaciones	* aztreonam *klaritromicina x 500 MG

					e *tobrami cian	*sinusitis refractari a al tratamie nto	*edema de la mucosa nasal *escurrimi ento posterior	*Fibrosis quistica con manifest aciones intestinal es	tabletas #40 tomar una cada 12 horas *tobramici na nebulizada
6	14	42,5	142	21,18	* 4 cirugias por sinustis severa	*tos *especto racion de color verde *mareo a cualquier hora del dia *pseudo mona aerugino sa *aspergill us *estiloco auresu		*Fibrosis quistica con manifest aciones pulmona res *Fibrosis quistica con manifest aciones intestinal es	*amoxicili na *acido clavulanic o dosis de 100 MG por kilo de amoxicilin a
7	14	42,8	150	19,02	*dornasa alfa *tobrami cina *lipas pancreati ca *vitamin a z,k *ranitidi na	*vertigo y mareos *lesion vestibula r *tos *especto racion verde purulent a *obstruc cion nasal *dolores muscular es en piernas	*no tiene dificultad respiratori a *hipertrofi a severa de los cornetes *otoscopia normal *abdomen bien	*Fibrosis quistica con manifest aciones pulmona res *Fibrosis quistica con manifest aciones intestinal es	*lipasa pancratica *vitaminas a,d,e,k,z *calcio

8	14	42,7	151	18,76	*dornasa alfa *tobramicina *enzimas pancreaticas	*espectoracion verde *sordera en oido izq *hiperglicemia *hemoglobina glicosilada *sobreinfeccion *pancreatico endocrino	*hipertrofia severa de los cornetes *eritema timpano izq	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	
9	15	47	157	19,07	*dornasa alfa *tobramicina *levofloxacina *colistina *lipasa pancreatica	*psudomona multiresistente *dolor abdominal *epistaxis severa *tos *espectoracion verde	*hipertrofia severa de los cornetes *abdomen normal	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*dornasa alfa *tobramicina *levofloxacina *ensoy diabetes *lipasa pancreatica *acido ursodesoxicolico *ranitidina
10	15	51	161	19,68	*ceftazidime *ciproan *azteronam	*psudomona multiresistente *tos *espectoracion verde	*edema de la mucosa nasal *eritema de la mucosa nasal *otoscopia normal *auscultacion pulmonar normal	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*coloistin *dornasa alfa *lipasa pancreatica *tobramicina *salmeterol/fluticasona *acido ursodesoxicolico

							*abdomen bien		*ensoy diabeticos *vitamina k, e
11	15	54,7	162	20,84	*imipenon *colistina venosa *ciprofloxacina	*infeccion por pseudomonas resistente	*edema de la mucosa nasal *otoscopia normal *auscultacion pulmonar normal *abdomen bien	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*ciproflaxina x 500 MG tabletas # 120 tomar una y media tableta cada 8 horas por 20 dias *azitromicina x 500 MG tabletas #36 tomar una tableta 3 veces a la semana por 3 meses *colistina ampollas x 150 MG #30 nebulizar 1 al dia *dornasalfa x 2.5 MG en 2.5 ML ampollas #180 nebulizar una al dia por 6

									meses
12	15	54	165	19,83	*lipasa pancreatica *tobramicina	*tos *exacerbacion	*hipertrofia severa de los cornetes *otoscopia normal *auscultacion pulmonar entre los limites normales *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares	*deflazacort *salbutamol *paracodina *klaricid *lipasa pancreatica *dornasa alfa *vitaminas k,e *ensoy diabetes
13	15	54	165	19,83	*colistin venoso *amikacina	*tos con espectoracion verde *broquectasis *dolor de torax *pseudomona resistente	*no tiene signo de deficiencia respiratoria *tiene saturacion del 94% sin oxigeno *tiene secrecion purulenta por fosas nasales *auscultacion pulmonar con creptios basales	*Fibrosis quistica con manifestaciones pulmonares	*natrol ampollas #20 mezclar una ampolla con 30 cc de agua destilada y hacer lavado nasal 3 veces al dia *deflazacort x 30 MG tabletas #20 dar una y edia tabletas al dia por 7

							*abdomen bien	dias *agua destilada bolsas # 20 x 100 cc para lavado nasal *ogastro x30 MG capsulas #120 tomar una cada 12 horas por 2 meses *tobramici na x 300 Mg x 5 Ml ampollas #56 nebulizar una cada 12 horas *ensure x 9000 Gm lata polvo # 8 tomar 2 veces al dia *vitamina k ampollas x 2mg #12 tomar una ampolla semanal por 3 meses *hidraplus 45 frasco # 90 tomar un frasco al dia *dornasa alfa una vez al dia
--	--	--	--	--	--	--	------------------	--

14	15	34,5	152	14,93	*enzimas digestivas *vitaminas a,d,e,k *tobracima	*tos seca *espectoracion blanca *psudomona *deposiciones normales	*hipertrofia de los cornetes *engrosamiento de la membrana timpanoeder *asculacion pulmonar normal *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares	*tobromacina *enzimas *vitaminas *zinc
15	15	34,5	152	14,93		*tos intensa *espectoracion purulenta *sin apetito *fatiga	*hipertrofia de los cornetes *eritema en timpano izq *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares	*dornasa alfa *salmetrol /fluicasona *lipasa pancreatica *vitaminas *pediasure *terapia respiratoria
16	15					*psudomona *estafilococo *aspergilus		*Fibrosis quistica con manifestaciones pulmonares	*ciprox500 tabletas #60 dar una tapleta cad 8 horas por 3 semans *tobracima x 300 MG apolletas para nebulizar #56 nebulizar una al dia

								por 28 dias
17	15				*ciproflaxina	*psedomona resistente ciprofloxi na sensible cefepime *estafiloco	*Fibrosis quistica con otras manifestaciones *Fibrosis quistica con manifestaciones intestinales *Fibrosis quistica con manifestaciones pulmonares	*cefepime x1 gm ampollas aplicar 1.8 gm iv cada 12 horas por 1 dias *tobracima nebulizada una ampolla nebulizada cada 12 horas *terapia respiratoria 3 veces al dia completas *dornasa alfa ampollas x 2.5 mg nebulizar una al dia *lipasa pancreatic a x 25000 u una antes de cada comida *pediasure *centrium *vitamina e,z bec *cateter

18					*cepefime	*psedomona resistente ciprofloxi na sensible cefepime *estafilococo		*Fibrosis quística con manifestaciones pulmonares	*cepefime x 1gm apollas #84 aplicar 1.8 gm iv cada 8 horas por 14 días *gentamicina x160 mg ampollas aplicar una ampolla iv lenta en una hora al día
19	15	36,8	153	15,72	*terapia respiratoria *dornasa alfa *enzimas digestivas *vitaminas complementos nutricionales *tobramicina	*tos disminuida *espectoración mas clara		*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*dornasa alfa *pediasure *enzimas *vitaminas *terapia respiratoria
20	15	37	157	15,1		*deposiciones bladas y fetidas *tos *espectoración amarillenta *alcaligenes fecalis *hongos	*hipeertrofea de los cornetes *rinorea purulenta *auscultación pulmonar normal	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*trimetoprim *dornasa alfa *enzimas *terapia respiratoria

						no especifico		es	
21	15	36,2	155	15,7	*dornasa alfa *lipasa pancreatica *enzimas *vitaminas z bec *pediasure *terapia respiratia	*fiebre *disnea *tos *espectoracion verde	*edema de la mucosa nasal *no tiene escurrimiento posterior *auscucion pulmonar con crepitos en las bases *dolor en el hipocondriaco con volumen hepatico *infiltrado intrseccional en el lobulo superior derecho *bronquiactasis *saturacion 99%	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*acido clavulanic *medrol *dornasa alfa *enzimas *vitaminas
22					*tmp sulfa no mejoro	*tos *espectoracion verde *fatiga *infiltrado bronconeumonico		*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica	*ciproflaxina x 500 MG tabletas # 120 tomar una y media tableta cada 8

						*infiltrados hemitorax *psudomona *hongo por definir		con manifestaciones intestinales	horas por 20 dias *tobramicina x 300 mg ampollas para nebulizar #56 una cada 12 horas *terapia respiratoria una vez al dia por 2 semanas
23	16	39			*cefepime *amikacina *tobramicina *esteroides	*klebsiella pneumoniae *tos *espectogramas claros *sinusitis crónica *bronquiectasia	*saturación del 93% *leve taquicardia *edema de la mucosa nasal *hipertrofia de los cornetes *crepitos en el hemitorax izq *abdomen con hernia umbilical	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*dornasalfa *lipasa pancreática *terapia respiratoria domiciliar
24	16	40,5			*tobramicina *terapia respiratoria	*tos *resonancia en pecho *insomnio *fiebre *espectogramas verdes	*hipertrofia de los cornetes *escurrimiento posterior verde *otoscopia con congestión oído izq	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones	*dornasalfa *trimetoprima *vitaminas *ensure

							*auscultación pulmonar con crepitos en las bases *higado inflamado	aciones intestinales	
25	16	40	157	16,23	*terapia respiratoria *lipasa pancreática *vitaminas *tobramicina *deronun cilocde *moxifloxacin	*tos *especto racion amarillenta *bajo de peso *	*edema de la mucosa nasal *leve eritema de la membrana del timpano *tiene crepitos en el hemitorax der *abdomen blando *no tiene masas ni megalas	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*pulmocar e *azitromicina *terapi respiratoria *lipasa ncreatica *vitaminas z bec *capella
26	16	41,5	161	16,01	*dornasa alfa *vitaminas a,d,e *ensure *lipasa pancreática	*tos *especto racion	*crepitos en la base derecha *otoscopia con leve opacidad de mt *secrecion nasal purulenta *abdomen bien	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*terapia respiratoria *azitromicina *ensure *avamys seretide m *sulbutamol *dornasa alfa *lipasa pancreatic a

27	17	41,5	162	15,81	*dornasa alfa *terapia respiratoria	*mal apetito *tos *especto racion amarilla *fiebre *diarrea *estreñimiento	*muy delgado *hipertofia de los cornetes *otoscopia bien *auscucion pulmonar normal *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares	*dornasa alfa *lipasa pancreatica *azitromicina *vitaminas *terapia respiratoria *ensure
28	17	43	162	16,38		*fiebre *tos *especto racion purulenta	*ferebil *saturacion del 94% *hipertrofia de los cornetes *otoscopia normal *auscucion pulmonar con crepitos en la base *soplo tubarico en la base del hemitorax *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*meropenen *ciproflaxina
29	17					*psudomona *fiebre		*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones	*cipro *azitromicina *voriconazole *terapia respiratoria con tubo nebulizador

								intestinal es	
30	18	48,5	169	16,98	*lipasa pancreati ca *zinnat *cipro	*tos humeda *especto racion verde claro *hemopit isis *pseudo mona	*hipetrofia de los cornetes *edema nasal *eritema nasal *otoscopia bien *orofaring e congestiva *crepitos en las bases pulmonare s *abdomen bien	*Fibrosis quistica con manifest aciones pulmona res *Fibrosis quistica con manifest aciones intestinal es	*cipro *tobramici na *lipasa pncreatica *vitaminas *pediasure *dornasa alfa *terapia respiratori a
31	13	25	127	15,5	*panzitrat	*tos leve *especto racion blanca *deposici ones normales	*tiene silabancias sin dificulatad respiratora	*Fibrosis quistica con manifest aciones intestinal es	
32	14	26,8	128	16,36	*hormon as de crecimie nto *lipasa pancreati ca *vitamin as S,D,E y K *dornasa alfa	*tos	*leve hipertrofia de los cornetes *otoscopia normal *cardiopul monar bien *abdomen bien	*Fibrosis quistica con manifest aciones pulmona res	
33	14	26,8	128	16,36		* tiene gripa		*Fibrosis quistica con	

								manifestaciones intestinales	
34	14	27,8	129	16,71	*cipro *tobra *dornasa alfa *pediasure *vitaminas *prteinas	*tos seca *especto racion amarilla	*paldia sin signos de dificultad respiratoria *otoscopia anormal *cardiopulmonar bn	*Fibrosis quistica con manifestaciones pulmonares	*proteinx *sinthroid *vitamina e *centrium junior *z bec granulado *seretide *kanakion *zantac *hidraplus *creon *dornasa alfa *pediasure
35	14	28	132	16,07		*tos *especto racion amarillenta	*rinorea clara *otoscopia bien *cardiopulmonar con silibacias	*Fibrosis quistica con manifestaciones pulmonares	*dornasa alfa *tobramicina
36	14	29	134			*tos *especto racion amarillenta *come bien *deposiciones normales	*buena condicion afebril *eritema con edema en la mucosa nasal *cardiopulmonar normal	*Fibrosis quistica con manifestaciones intestinales	
37	15	33	138	17,49				*Fibrosis quistica con manifestaciones intestinales	

38	15	33	138	17,33				*Fibrosis quística con manifestaciones pulmonares	
39	10	43	143	21,1	*azteronam *tobra *dornasalfa *prevenar *zantac	*pansinusitis crónica *psudomona multiresistente *obstrucción en la nariz *deposiciones blandas	*buena condición afebril *dolor abdominal recurrente *sin signos de dificultad respiratoria *leve opacación en la membrana de timpano izq *orografía bien *cardiopulmonar bien	*Fibrosis quística con manifestaciones pulmonares *Fibrosis quística con manifestaciones intestinales	*lipasa pancreática
40	12	45	150	20	*colistina nebulizada *tobramicina *levofloxacina	*sinusitis severa *bronquiectasis *psudomona multiresistente	*buena condición afebril *sin problemas respiratorios *otoscopia normal *edema en l mucosa nasal *crepito en el hemitorax	*Fibrosis quística con manifestaciones intestinales	*tobramicina nebulizada *levofloxacina *azitromicina *lipasa pancreática *vitaminas

41	13	55	165	20,2	*doripneme *tobramicina *ciproflaxina *enzimas	*tos humeda *espectoracion verde *fiebre *pseudomona multiresistente	*buen estado afebril *sin deficiencias respiratorias *severa hipertrofia de los cornetes *obstruccion nasal intensa *otoscopia bien *orofaringe muy congestiva *auscultaciones pulmonar con roncus y movilizaciones de crepitos *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*colastina *cipro* vitaminas e *encimaz digestivas *lipasa pancreatica *dornasas alfa
42	4	10,3	75	18,31	*dornasas alfa *tobramicina	*pseudomona *fiebre *tos *rinorea *deposiciones blandas	*aceptable condicion afebril *aspiraciones prolongadas con roncos *escurremiento posterior *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*fisopred *clavulin

43	8	20	115	15,12	*dornasa alfa *lipasa pancreatica *vitaminas *terapia respiratoria	*sin tos *sin espectoracion *deposiciones normales	*buen estado afebril *sin problemas respiratorios *hipertrofia en los cornetes *otoscopia normal *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*dornasa alfa *lipasa pancreatica *vitaminas *terapia respiratoria
44	8	19,5	115	15	*dornasa alfa *lipasa pancreatica *vitaminas *terapia respiratoria	*espectro racion verde *deposiciones duras con sangrado	*buena condicion afebril *sin deficiencia respiratoria *otoscopia normal *cardiopulmonar normal *abdomen bien	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*dornasa alfa *lipasa pancreatica *vitaminas *terapia respiratoria *tirimetor pim sulfa *pediasure *ibuprofeno
45	9	20,7	117	15,12	*lipasa pancreatica *dornasa alfa *vitaminas *pediasure *terapia respiratoria		*edema de la mucosa nasal *cardiopulmonar normal *abdomen normal	*Fibrosis quistica con manifestaciones pulmonares *Fibrosis quistica con manifestaciones intestinales	*lipasa pancreatica *dornasa alfa *vitaminas *pediasure *terapia respiratoria

ANEXO B

Clase del prototipo "Algoritmo_c45.java".

```

package Algoritmo_c45;

import java.util.LinkedList;

import simulador_fibrosisquistica.Ganancia;

/**
 *
 * @author Jimmy Andrés León Muñoz
 * @author Miguel Ángel Acevedo Franco
 */

public class Nodo {

    public String contenido; //Valores que tendrán los nodos,
    pueden ser síntomas o resultados

    private static String tree = "\n\n=====ÁRBOL
    GENERADO=====";

    private static String rama_padre; //Valor de la rama de donde
    proviene el nodo

    private static String nodo_padre; //Valor del nodo padre

    private static String [][] base_datos;

    private static String [] sintomas;

    private static int tratamiento;

```

```

private static boolean infoDetallada;

private static LinkedList<String> ramas_temp;

private static boolean tipo; // TRUE: Nodo Final - FALSE: Nodo
Síntoma

public LinkedList<Nodo> hijos;

String np = "";

int flag = 0;

public Nodo(String[][] dataBase, String[] sintoma, int
treatment, boolean data, String nodo_p, String rama_p){

    base_datos = dataBase;

    tratamiento = treatment;

    infoDetallada = data;

    sintomas = sintoma;

    nodo_padre = nodo_p;

    rama_padre = rama_p;

    hijos = new LinkedList<>();

    imprimirParametrosIniciales();

}

public boolean esFinal(){

    int cont = 0;

    String valor = base_datos[0][tratamiento];

    for (int i = 1; i < base_datos.length; i++)

```

```

        if(!valor.equals(base_datos[i][tratamiento]))
            cont++;

    if(cont == 0) return true;
    return false;
}

public void printTree(){
    System.out.println(tree);
}

public void generarArbol(){
    if(!esFinal()){
        int posicionMejorNodo;

        LinkedList<Ganancia> listaGanancias =
calcularTablaGanancias(base_datos);

        posicionMejorNodo = encontrarMejor(listaGanancias);

        contenido = sintomas[posicionMejorNodo];
        tipo = false;

        System.out.println("El mejor nodo es: " + contenido);

        System.out.println("-----");
        "-----");

        obtenerListaRamas(posicionMejorNodo);
    }
}

```

```

imprimirRamas();

//-----

tree += "\nNodo padre: " + nodo_padre + "\n";
tree += "Rama padre: " + rama_padre + "\n";
tree += "Nodo: " + contenido + "\n\n";
tree += "Ramas:\n";
for (int i = 0; i < ramas_temp.size(); i++) {
    tree += " * " + ramas_temp.get(i) + "\n";
}

tree += "-----\n";
//-----

for (int i = 0; i < ramas_temp.size(); i++) {
    if(!verificarCorte(listaGanancias,
posicionMejorNodo)){

        int          filas          =
cantidadDatosRepetidos(posicionMejorNodo, ramas_temp.get(i));

        int columnas = base_datos[0].length-1;

        int columnaOmitir = posicionMejorNodo;

        String[][]          bdTemp          =
obtenerNuevaBaseDatos(base_datos, columnas, filas, columnaOmitir, i);

```

```

        String[]          sintomasTemp          =
obtenerNuevosSintomas(columnaOmitir, sintomas);

        LinkedList<Nodo> originalHijos = hijos;

        String[][] originalDB = base_datos;

        String[] originalSintomas = sintomas;

        String rama = rama_padre;

        String nodo_p = nodo_padre;

        int originalTratamiento = tratamiento;

        Nodo temporal = new Nodo(bdTemp, sintomasTemp,
tratamiento - 1, infoDetallada, sintomas[posicionMejorNodo],
ramas_temp.get(i));

        temporal.generarArbol();

        hijos = originalHijos;

        rama_padre = rama;

        nodo_padre = nodo_p;

        base_datos = originalDB;

        sintomas = originalSintomas;

        tratamiento = originalTratamiento;

        hijos.add(temporal);

    }else{

        LinkedList<String> valores = new
LinkedList<>();

```

```
        for (int j = 0; j < base_datos.length; j++)

if(!valores.contains(base_datos[j][tratamiento]))

valores.add(base_datos[j][tratamiento]);

        int[] tam = new int[valores.size()];

        for (int j = 0; j < tam.length; j++) tam[j] =

0;

        for (int j = 0; j < base_datos.length; j++)

            for (int k = 0; k < valores.size(); k++)

if(valores.get(k).equals(base_datos[j][tratamiento]))

                tam[k]++;

        int mayor = 0;

        for (int j = 1; j < tam.length; j++) {

            if(tam[mayor] < tam[j])

                mayor = j;

        }

        tipo = true;

        if(flag < ramas_temp.size()-1){

            np = contenido;
```



```

//-----
}
}

private static String[][] obtenerNuevaBaseDatos(String[][]
baseDatos, int col, int fil, int colOm, int i){

    String[][] db = new String[fil][col];

    String[][] baseDatosOriginal = baseDatos;

    int columna= 0;

    int fila = 0;

    for (int j = 0; j < baseDatosOriginal.length; j++){
        for (int k = 0; k < baseDatosOriginal[0].length; k++)
            if(k != colOm &&
baseDatosOriginal[j][colOm].equals(ramas_temp.get(i))){
                db[fila][columna] = baseDatosOriginal[j][k];
                columna++;
            }

        columna = 0;

    if(baseDatosOriginal[j][colOm].equals(ramas_temp.get(i)))
        fila++;
    }
}

```

```
        return db;
    }

    private static String[] obtenerNuevosSintomas(int colOm,
String[] sintomaOriginal){

        int temp=0;

        String[] original = sintomaOriginal;

        String[] sintoma = new String[original.length-1];

        for (int j = 0; j < original.length; j++)

            if(j != colOm){

                sintoma[temp] = original[j];

                temp++;

            }

        return sintoma;

    }

    private boolean verificarCorte(LinkedList<Ganancia> lista, int
pos){

        if(lista.get(pos).getGanancia_de_informacion() -
lista.get(pos).getProporcion_de_ganancia() == 0)

            return true;

        return false;

    }
}
```

```

private int encontrarMejor(LinkedList<Ganancia> lista){

    int mayor;

    if(lista == null) return mayor = -1;

    mayor = 0;

    for (int i = 0; i < lista.size(); i++)

        if(lista.get(i).getGanancia_de_informacion() >
lista.get(mayor).getGanancia_de_informacion())

            mayor = i;

    return mayor;
}

private static double calcularEntropiaGeneral(){

    double positivos = 0.0;

    double negativos = 0.0;

    for (int i = 0; i < base_datos.length; i++) {

        if(base_datos[i][tratamiento].equals("SI")){

            positivos++;

        }else if(base_datos[i][tratamiento].equals("NO")){

            negativos++;

        }

    }
}

```

```

    }

    double pos = (positivos/base_datos.length);
    double neg = (negativos/base_datos.length);

    return (-1 * pos)*(Math.log(pos)/Math.log(2)) - (neg) *
    Math.log(neg)/Math.log(2);
}

private static int cantidadDatosRepetidos(int columna, String
dato){

    int cont = 0;

    for (int i = 0; i < base_datos.length; i++)

        if(base_datos[i][columna].equals(dato))

            cont++;

    return cont;
}

private static void obtenerListaRamas(int sintoma){

    ramas_temp = new LinkedList<>();

    for(int i = 0; i< base_datos.length; i++)

        if(!ramas_temp.contains(base_datos[i][sintoma]))

            ramas_temp.add(base_datos[i][sintoma]);
}

```

```

private static LinkedList<Ganancia>
calcularTablaGanancias(String[][] dataBase){

    int pos = 0;

    LinkedList<Ganancia> ganancias = new LinkedList<>();

    double entropiaGeneral = calcularEntropiaGeneral();

    while(pos < (base_datos[0].length - 5)){

        ramas_temp = new LinkedList<>();

        obtenerListaRamas(pos);

        double[][] cantidades = new
double[ramas_temp.size()][2];

        //Llenar la matriz de ceros

        for (int i = 0; i < ramas_temp.size(); i++)
            for (int j = 0; j < 2; j++)
                cantidades[i][j] = 0;

        for (int i = 0; i < ramas_temp.size(); i++) {
            for (int j = 0; j < dataBase.length; j++) {

if(dataBase[j][pos].equals(ramas_temp.get(i))){

                if(dataBase[j][tratamiento].equals("SI")){

```

```

        cantidades[i][0]++;

        }else
if (dataBase[j][tratamiento].equals("NO")){

        cantidades[i][1]++;

        }

        }

    }

}

double[] entropias = new double [ramas_temp.size()];

double temp1, temp2, temp, cant;

for (int i = 0; i < ramas_temp.size(); i++) {

    cant = cantidades[i][0]+cantidades[i][1];

    temp1 = cantidades[i][0] / cant;

    temp2 = cantidades[i][1] / cant;

    temp = (-1 * temp1)*(Math.log(temp1)/Math.log(2))-
(temp2) * Math.log(temp2)/Math.log(2);

    if(Double.isNaN(temp))

        entropias[i] = 0.0;

    else

        entropias[i] = temp;

}

```

```

double gananciaSintoma = entropiaGeneral;

for (int i = 0; i < cantidades.length; i++) {
    gananciaSintoma -= (entropias[i] *
((cantidades[i][0]+cantidades[i][1])/dataBase.length));
}

double informacionDivision = 0.0;

for (int i = 0; i < cantidades.length; i++) {
    temp1 = (cantidades[i][0]+cantidades[i][1]) /
dataBase.length;
    temp = temp1*(Math.log(temp1)/Math.log(2));
    informacionDivision -= temp;
}

if(infoDetallada){
    System.out.println("-----
-----");
    System.out.println("DATOS ADICIONALES");
    System.out.println("");
    System.out.println("Entropía general: " +
entropiaGeneral);
    System.out.println("Síntoma: " + sintomas[pos]);
    System.out.println("");
    System.out.println("Tablas de cantidades");
}

```

```

        for (int i = 0; i < cantidades.length; i++) {
            System.out.print("-----\nRama: " +
ramas_temp.get(i) + "\nSI ( ");
            for (int j = 0; j < 2; j++) {
                System.out.print(cantidades[i][j] + " ");
            }
            System.out.println(" NO\n-----");
        }
        System.out.println("");

        for (int i = 0; i < ramas_temp.size(); i++)
            System.out.println("Entropía de " +
ramas_temp.get(i) + ":\t" + entropias[i]);

        System.out.println("");

        System.out.println("Ganancia de " + sintomas[pos]
+ ": "+gananciaSintoma);

        System.out.println("Información de la división: "
+ informacionDivision);
    }

    Ganancia nodoTemporal = new Ganancia(gananciaSintoma,
informacionDivision);

    ganancias.add(nodoTemporal);

    pos++;
}

```

```

        imprimirTablaGanancias(ganancias);

        return ganancias;
    }

    private static void imprimirBaseDatos(){

        System.out.println("-----BASE
DE DATOS-----");

        for (int i = 0; i < sintomas.length; i++)

            System.out.print(sintomas[i] + ", ");

        System.out.println("\n-----
-----");

        for (int j = 0; j < base_datos.length; j++) {

            for (int k = 0; k < base_datos[0].length; k++)

                System.out.print(base_datos[j][k] + " ");

            System.out.println("");

        }

        System.out.println("\n-----
-----");

    }

    public void imprimirParametrosIniciales(){

System.out.println("////////////////////////////////////
////////////////////////////////////");

        System.out.println("Nodo padre: " + nodo_padre);

        System.out.println("Rama padre: " + rama_padre);

```

```

        System.out.println("Valor tratamiento: " + tratamiento);
        System.out.println("");
        imprimirBaseDatos();
    }

    private static void imprimirRamas(){
        for (int i = 0; i < ramas_temp.size(); i++)
            System.out.println("Rama: " + ramas_temp.get(i));
    }

    private static void imprimirTablaGanancias(LinkedList<Ganancia> ganancias) {
        System.out.println("-----TABLA DE GANANCIAS-----");
        System.out.println("Ganancia\t\tProporción\t\tSíntoma");
        System.out.println("-----");

        for (int i = 0; i < ganancias.size(); i++)
            System.out.println(ganancias.get(i).getGanancia_de_informacion() +
                "\t" + ganancias.get(i).getProporcion_de_ganancia() +
                "\t" + sintomas[i]);

        System.out.println("-----");
    }

```

```
}
```

ANEXO C

Clase del prototipo “simulador_fibrosisquistica.java”

```
package simulador_fibrosisquistica;

/**
 *
 * @author Jimmy Andrés León Muñoz
 * @author Miguel Ángel Acevedo Franco
 */
public class Ganancia {

    private double ganancia_de_informacion;
    private double proporcion_de_ganancia;
    private double division_de_ganancia;

    public Ganancia (double ganancia, double division){
        ganancia_de_informacion = ganancia;
        division_de_ganancia = division;
        proporcion_de_ganancia = (ganancia / division);
        if(Double.isNaN(proporcion_de_ganancia))
            proporcion_de_ganancia = 0.0;
    }
}
```

```
public double getGanancia_de_informacion() {  
    return ganancia_de_informacion;  
}  
  
public void setGanancia_de_informacion(double  
ganancia_de_informacion) {  
    this.ganancia_de_informacion = ganancia_de_informacion;  
}  
  
public double getProporcion_de_ganancia() {  
    return proporcion_de_ganancia;  
}  
  
public void setProporcion_de_ganancia(double  
proporcion_de_ganancia) {  
    this.proporcion_de_ganancia = proporcion_de_ganancia;  
}  
  
public double getDivision_de_ganancia() {  
    return division_de_ganancia;  
}  
  
public void setDivision_de_ganancia(double  
division_de_ganancia) {
```

```
        this.division_de_ganancia = division_de_ganancia;
    }
}
```

ANEXO D

Clase del prototipo "Nodo.java"

```
import Algoritmo_c45.Nodo;

/**
 * @author Jimmy Andrés León Muñoz
 * @author Miguel Ángel Acevedo Franco
 * @version 1
 */
public class Main {

    private final static int tratamiento = 8; // 8 - 12

    private final static boolean infoDetallada = true; //Ver
información detallada

    private static final String [] sintomas =

        {"Peso",

        "Tos con expectoración de color",

        "Dolor abdominal",

        "Deposiciones anormales",

        "Pseudomona",
```

```

        "Pseudomona multi-resistente",
        "Estafilococo aerus",
        "Aspergelis auerus"
    };

    private static String [][] base_datos =
    {
        {"NORMAL", "NO", "SI", "SI", "NO", "NO", "SI", "NO",
        "NO", "NO", "NO", "NO", "SI"},
        {"NORMAL", "SI", "NO", "NO", "SI", "NO", "SI", "NO",
        "NO", "NO", "NO", "SI", "NO"},
        {"NORMAL", "SI", "NO", "NO", "SI", "NO", "NO", "NO",
        "NO", "SI", "NO", "NO", "NO"},
        {"NORMAL", "SI", "SI", "NO", "NO", "SI", "NO", "NO",
        "NO", "NO", "SI", "NO", "SI"},
        {"ALTO", "SI", "NO", "SI", "NO", "SI", "NO", "NO",
        "SI", "NO", "SI", "NO", "SI"},
        {"BAJO", "SI", "SI", "NO", "NO", "SI", "NO", "NO",
        "SI", "NO", "NO", "SI", "NO", "NO"},
        {"BAJO", "NO", "NO", "NO", "SI", "NO", "SI", "SI",
        "NO", "NO", "NO", "SI", "NO"},
        {"BAJO", "SI", "NO", "NO", "NO", "NO", "NO", "NO",
        "SI", "NO", "NO", "NO", "NO"},
        {"BAJO", "SI", "SI", "SI", "NO", "NO", "NO", "NO",
        "SI", "SI", "SI", "NO", "NO", "NO", "NO"},
        {"BAJO", "SI", "NO", "NO", "SI", "NO", "NO", "NO",
        "SI", "SI", "NO", "NO", "NO"},
        {"NORMAL", "SI", "NO", "NO", "NO", "SI", "NO", "NO",
        "NO", "NO", "SI", "NO", "NO"}
    }

```

```

        {"NORMAL",
"NO", "NO", "SI", "NO", "SI"},
        {"NORMAL",
"NO", "NO", "SI", "SI", "NO"},
        {"BAJO",
"SI", "NO", "NO", "NO", "SI"},
        {"BAJO",
"NO", "SI", "NO", "NO", "NO"}

};

```

```

public static void main(String[] args) {

    Nodo arbol = new Nodo(base_datos, sintomas, tratamiento,
infoDetallada, null, null);

    arbol.generarArbol();

    arbol.printTree();

}

}

```