

# **DISEÑO E IMPLEMENTACION DE UN SISTEMA PARA PREDICCIÓN DE ENFERMEDADES TROPICALES TRANSMITIDAS POR VECTORES ARTRÓPODOS EN LA CIUDAD DE GUAYAQUIL**

## **AUTORES**

Sara Martha Castillo Fiallos <sup>1</sup>, Mario Eugenio Clavijo Rosero <sup>2</sup>, Juan Alvarado <sup>3</sup>.

## **RESUMEN**

Con el principal objetivo de obtener el título de Licenciados en Sistemas de Información (LSI), los estudiantes Sara Castillo y Mario Clavijo, han desarrollado un sistema que permite Predecir las condiciones y casos por las cuales se presentan las enfermedades tropicales transmitidas por Vectores artrópodos, haciendo uso de la Minería de Datos; logrando así finalizar una etapa más en su vida profesional.

Este proyecto está orientado en la atención brindada por el Hospital de Infectología de Guayaquil, uno de los más importantes centros de salud, considerando cada uno de los casos presentados.

With the main objective of obtaining the title of Graduates in Systems of Information (LSI), the students Sara Castillo and Mario Clavijo, they have developed a system that permit to prediction the conditions and cases by which the tropical diseases transmitted by artrópodos Vectors appear, making use of the Data Mining; being able this way to conclude a stage more in their professional life.

This project is oriented in the attention offered by the Hospital of Infectología in Guayaquil, one of the most important centers of health, considering each one of the presented cases.

## **INTRODUCCION**

El siguiente documento describe un modelo de minería de datos diseñado para Centros de Salud, que se dedican al tratamiento, cura y prevención de enfermedades transmitidas por vectores artrópodos.

El objetivo principal de este modelo, es determinar y predecir los casos y condiciones por las cuales sucedió la enfermedad.

La fuente de datos utilizada en este proyecto, proviene del: Hospital de Infectología, SNEM (Servicio Nacional de erradicación de la Malaria), INAMHI (Instituto Nacional de Metereología e Hidrología) e INEC (Instituto Nacional de Estadísticas y Censos); para uso educativo y no comercial.

---

<sup>1</sup> Egresada de Licenciada en Sistema de Información FIEC-ESPOL.

<sup>2</sup> Analista de Sistemas PROTCOM-ESPOL, Egresado de Licenciado en Sistema de Información FIEC-ESPOL.

<sup>3</sup> Director del Tópico. Magíster en Administración Pública con Mención en Liderazgo ESPOL, Ingeniero en Electricidad Especialización Computación ESPOL, Profesor de ESPOL.

## **CONTENIDO**

### **CAPITULO 1. DESCRIPCIÓN DEL NEGOCIO**

El Hospital de Infectología, fue creado como respuesta al incremento de enfermedades infecto-contagiosas en el país, fundado en el edificio del Asilo de Ancianos "Hospicio Corazón de Jesús" en la dirección del Dr. Isaías Medina (1946-1951) y bautizado con el nombre del Ilustre Guayaquileño *Dr. José Daniel Rodríguez Maridueña*, el 7 de agosto de 1975.

Está ubicado en las calles Julián Coronel 900 y José Mascote, parroquia Tarqui del Cantón Guayaquil, Catalogado como una *Unidad Hospitalaria Referencial*.

#### **CAPITULO 1.1. DEFINICION DEL PROBLEMA ACTUAL**

Actualmente, esta Unidad Hospitalaria no cuenta con una Sistema Integrado, así como ninguno de sus departamentos esta enlazado con un sistema. El Departamento de Estadística no posee un Sistema automatizado que le permita llevar el Registro y Control de las Enfermedades que se originan en una estación invernal.

Debido al crecimiento en número de pacientes, de enfermedades tropicales y de la detección de nuevas enfermedades infectocontagiosas, tiene la necesidad de crear un sistema el cual le permita llevar un control automatizado de sus pacientes.

Los registros son llevados en una hoja de cálculo llamada EXCEL, en donde el usuario lleva un registro manual de los casos atendidos por paciente y fecha en la que ingresó; agrupando los casos por enfermedad.

#### **CAPITULO 1.2. PROCESO GENERAL DEL NEGOCIO**

El Proceso de atención al paciente, comprende diferentes departamentos detallados en los siguientes pasos:

1. Información, dirige al paciente a obtener ayuda sobre que Departamento dirigirse.
2. Estadística, verifica que el paciente no posea Historial para proceder a crearle uno, de caso contrario se lo envía directamente a Caja.
3. Caja, recepta el pago del ticket de consulta para que el pacienta proceda a dirigirse a Admisión.
4. Admisión, luego de tomarle sus signos vitales, peso y talla; verifica si el paciente está grave para que sea atendido por Emergencia, caso contrario se lo envía a Consulta Externa.
5. Emergencia y Consulta Externa, procede a examinar al paciente si el mismo presenta signos y síntomas probables de contagio, se envía a Laboratorio para realizar las pruebas confirmatorias, caso contrario, procede a dar la medicación respectiva.
6. Laboratorio, procede a tomar la muestra de sangre para confirmar.
7. Con los resultados de Laboratorio se actualiza el Historial del Paciente y se lleva un registro manual de los casos atendidos.

## CAPITULO 2. DESCRIPCIÓN DE LOS DATOS

Los datos recopilados para el desarrollo de nuestro proyecto son de 4 años (2003-2006) para los meses de enero, febrero, marzo, abril y mayo de los casos atendidos: Dengue Clásico y Dengue Hemorrágico. Contamos además con datos de los niveles pluviométricos, temperatura, áreas de fumigación e índices socio-económicos por sectores censales del Censo realizado en el año 2001.

Los Índices censales son un total de 84, clasificados de la siguiente manera:

- Tipo de Vivienda
  - Tipo de techo
  - Tipo de Pared
  - Tipo de Piso
  - Tipo estructura
  - Formas de Instalación de Tubería para Agua
  - Formas de abastecimiento del Agua
  - Formas de desechar los excrementos
  - Uso de Luz eléctrica
  - Uso de Televisión
  - Formas de desechar la basura
  - Nivel de Educación
  - Grupos familiares
- } *Estructura de la vivienda*  
 } *Servicios Básicos*  
 } *Niveles de estudio*  
 } *Situación familiar*

### CAPITULO 2.1. MODELO ENTIDAD RELACIÓN DEL SISTEMA

El modelo relacional de las tablas creadas para el sistema es:

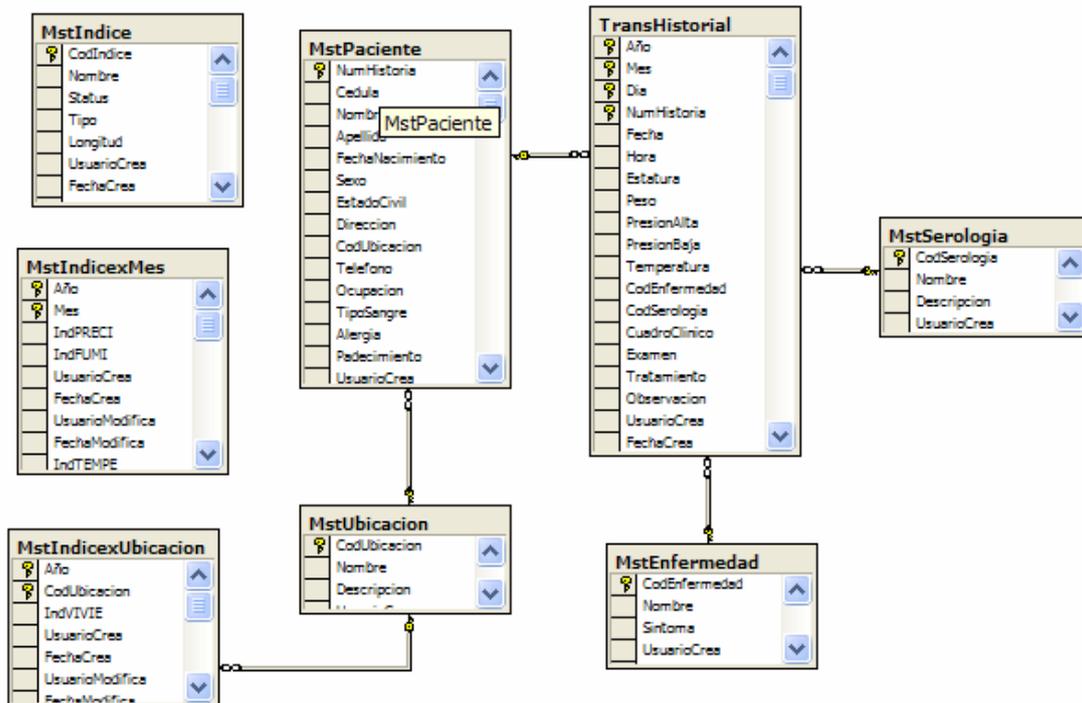


Figura <sup>1</sup>

<sup>1</sup> Figura del modelo entidad relación del sistema desarrollado.

## **CAPITULO 2.2. PROCESO DE MINERIA DE DATOS**

La Minería de datos es un proceso semi-automatizado para analizar grandes bases de datos, con el propósito de encontrar patrones que sean válidos, nuevos, útiles e interpretables. Tiene por objetivo extraer conocimiento a partir de los datos. Los datos sin procesar son inútiles y es necesario usar técnicas para extraer automáticamente información.

Minería de Datos es una manera de aprender del pasado, para hacer mejores decisiones en el futuro. Las mejores prácticas son diseñadas para conocer cosas que no son verdad o conocer cosas que son verdad pero no son útiles.

### **CAPITULO 2.2.1 METODOLOGIA APLICADA**

La metodología de minería de datos tiene 11 pasos que son:

1. Identificar el problema, trasladando el problema de negocio en un problema de minería de datos.
2. Seleccionar los datos apropiados.
3. Conocer los datos.
4. Crear el set de datos para crear el modelo.
5. Arreglar problemas con los datos.
6. Transformar los datos.
7. Construir el modelo, involucra hacer el algoritmo.
8. Evaluar el modelo, ver la utilidad del modelo.
9. Implementar el modelo en un sistema comercial.
10. Evaluar los resultados del modelo.
11. Empezar otra vez.

Los pasos tienen un orden, pero no es necesario completarlos de manera consecutiva, estos nos permiten regresar al paso anterior para obtener los resultados esperados y el modelo ideal, que mejor se ajuste a nuestra necesidad.

### **CAPITULO 2.2.2 DESCRIPCIÓN DEL PROCESO REALIZADO**

Entre los datos necesarios para cumplir los objetivos de este proyecto, se consideró la información en EXCEL proporcionada por el Hospital para su respectiva evaluación. Estos datos son almacenados en la base de datos.

Los Índices que se usarán para la Predicción del modelo, están dados por los Sectores Censales determinados en el Censo realizado en el 2001. Estos datos fueron facilitados en un archivo de EXCELL y nos permitirán evaluar mejor los datos e identificar claramente las condiciones de contagio de la Enfermedad. Existen 375 zonas censales divididas en varios sectores para la ciudad de Guayaquil.

Con estos Índices, las direcciones de los Pacientes corresponden a una Zona y Sector Censal. Se aplicó un preprocesamiento para usar sólo aquellos que añadan criterio y permitan evaluar.

Los pasos realizados en el procesamiento de los datos para poder aplicar el Modelo de Predicción, fueron:

1.- Agrupar cada una de las direcciones de los Pacientes, a la Zona y sector Censal correspondiente, haciendo uso de un mapa Censal obtenido en el INEC. Este paso se lo hizo de forma manual.

2.- Crear un procedimiento almacenado en la base de datos para seleccionar los datos de los Pacientes por Enfermedad y clasificarlos como de **Contagio**. Para los sectores censales en donde la enfermedad no se dio, se clasificó como de **No Contagio**. Para tener un balance en los datos, se incrementó el número de registros de Contagio, repitiendo los mismos para igualar el número de registros de Contagio y No Contagio. Luego se adicionaron los valores de los Índices seleccionados de acuerdo al sector censal correspondiente.

3.- Los datos generados por el procedimiento almacenado son registrados en una Tabla que tendrá toda la información necesaria para el Modelo de Predicción

### CAPITULO 2.2.3 MODELO ENTIDAD RELACIÓN APLICADO AL MODELO DE MINERÍA

El modelo relacional de las tablas creadas para el proceso de minería es el siguiente:

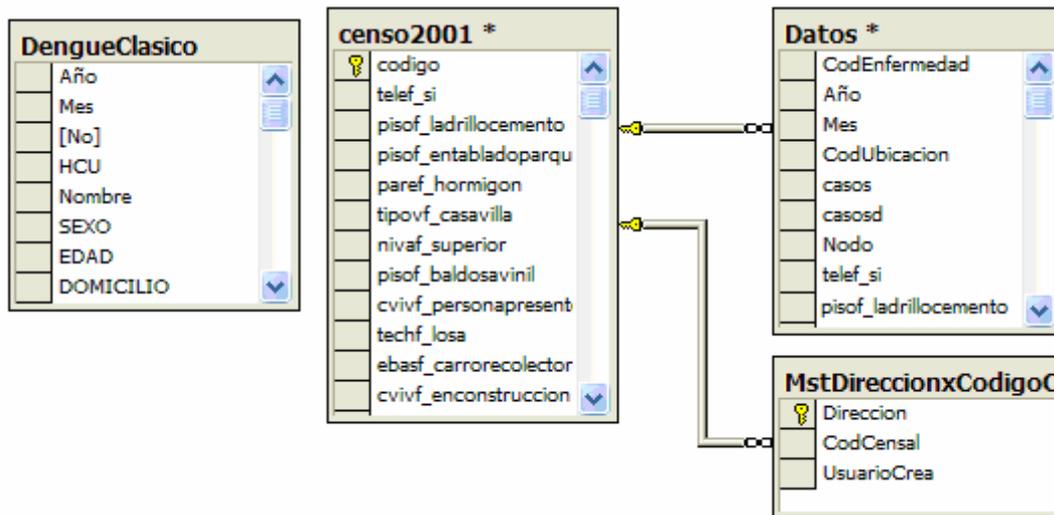


Figura <sup>2</sup>

### CAPITULO 3. PREPROCESAMIENTO DE LOS DATOS

La información recibida en archivos de Excel fue importada a la base de datos para su respectivo análisis, verificación y uso.

<sup>2</sup> Figura del modelo entidad relación del sistema desarrollado.

Para la información registrada como Historial de los Pacientes por Enfermedad, se realizó una verificación de datos nulos, blancos y eliminación de registros duplicados. El dato relevante es la Dirección del Paciente y todos los registros la tenían especificada. Para hacer uso de estos índices censales, estas direcciones fueron ubicadas a una Zona y sector censal correspondiente; haciendo uso de un mapa Censal obtenido en el INEC.

Para los Índices Censales se crearon nuevos valores, con el objetivo de poder representar en porcentaje el peso que tiene sobre el total de la categoría a la que pertenece. De esta forma, al momento de evaluar el algoritmo de Árbol de decisión el índice pueda interpretarse.

Para optimizar el proceso de Minería se evaluó los índices porcentuales, haciendo uso de un programa llamado WEKA; con el objetivo de usar los más representativos y relevantes. Mediante WEKA se hacen pruebas de Hipótesis, para establecer la significancia de las columnas (Índices) con la Variable de Respuesta.

Los índices seleccionados son 15, los cuales indicamos a continuación:

<b>INDICES SELECCIONADOS</b>	<b>DESCRIPCIÓN</b>
★ telef_si	Tiene teléfono
★ pisof_ladrillocimiento	Piso Ladrillo Cemento
★ pisof_entablado parquet	Piso de Tabla
★ peref_hormigon	Pared Hormigón
★ tipovf_casavilla	Vivienda Casa o Villa
★ nivaf_superior	Nivel estudio superior
★ pisof_baldosavinil	Piso baldosa o vinil
★ cvivf_personapresente	Persona presente en censo
★ techf_losa	Techo de loza
★ ebasf_carrecolector	Servicio de Carro Recolector de Basura
★ cvivf_enconstruccion	Vivienda en Construcción
★ eaguf_pozoseptico	Pozo Séptico para desecho de excremento
★ pisof_tierra	Piso de Tierra
★ peref_canarevestida	Pared de caña revestida
★ tipovf_departamento	Vivienda Departamento

#### **CAPITULO 4. SELECCIÓN Y EVALUACIÓN DEL MODELO**

El modelo de Minería de Datos se llama: “Árbol de Decisión”, donde obtendremos una serie de reglas en forma jerárquica que deducen el resultado de lo que se quiere predecir.

La Predicción que se necesita realizar determinará las condiciones en que se puedan dar casos de Contagios y No Contagio por Enfermedad. Estas condiciones son presentadas en forma de un Árbol, indicando el número de registros que contiene un segmento.

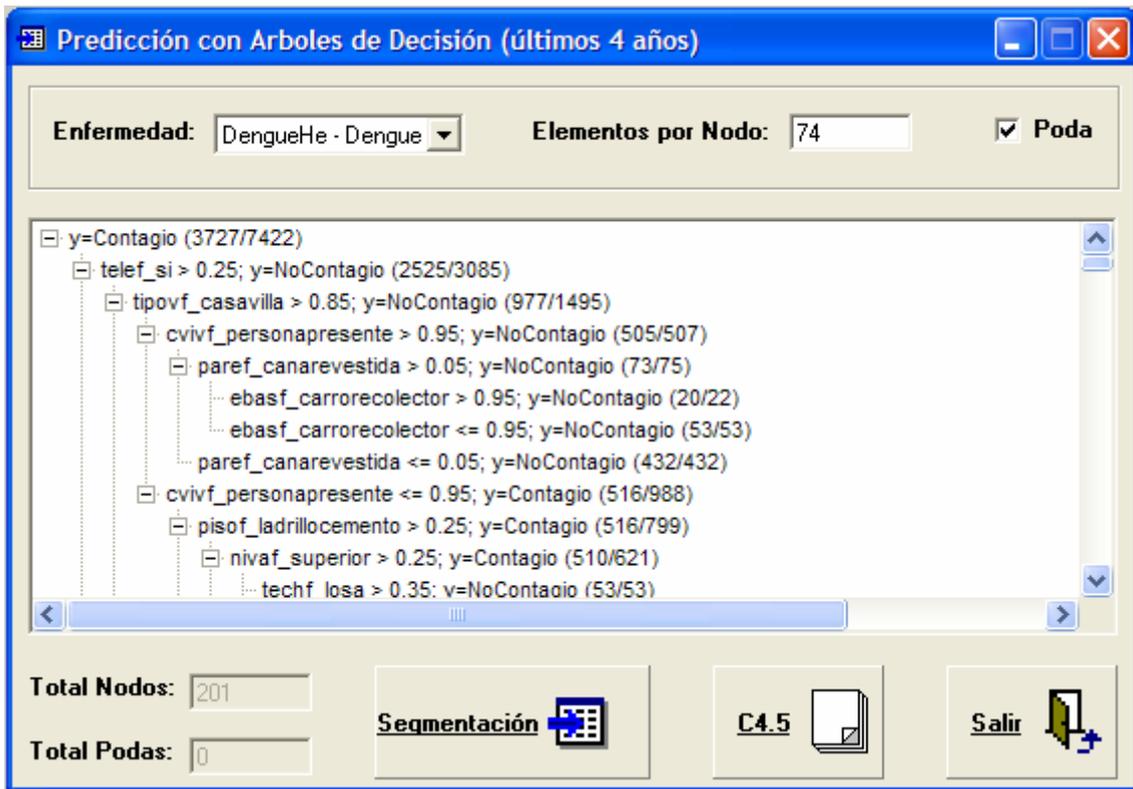


Figura <sup>3</sup>

#### CAPITULO 4.1. ANALISIS DE LOS SEGMENTOS

Se llama segmento a una hoja del árbol de decisión; es decir, desde la raíz hasta el final de su ramificación.

Una vez creado el Árbol de decisión, se realiza el proceso de Segmentación, el cual permite evaluar los segmentos encontrados para determinar el Índice Élite. Este Índice élite nos indica el grado de significación del Segmento; el valor de este índice supera en mucho a la clase minoritaria existente.

Los Nodos con el Índice élite de mas valor, resultan en respuesta a lo que se está evaluando.

Esto lo podemos verificar y evaluar en la siguiente figura:

<sup>3</sup> Figura que muestra el árbol de decisión,, generado en el sistema.

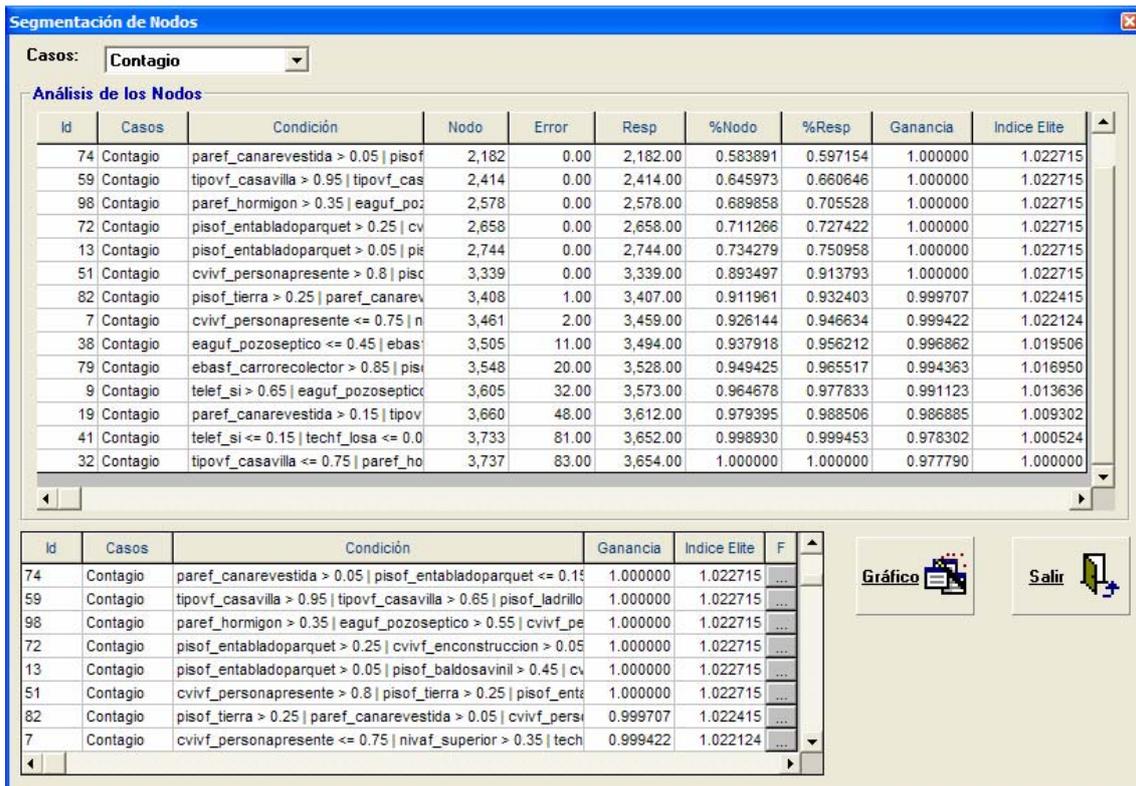


Figura 4

Una vez realizada la Segmentación, se muestra los segmentos con mayor índice élite para determinar las variables o condiciones de forma simplificada, en rangos mínimos y máximos de ocurrencia. Estas variables son el resultado de lo que se está queriendo obtener mediante el proceso de minería.

## CAPITULO 4.2. RESULTADOS ENCONTRADOS

Como resultado tenemos los segmentos con mayor Índice Élite, expresados en variables simplificadas con rangos mínimos y máximos, por casos.

Con esta información, resulta fácil determinar que el mayor grado de incidencia para estas enfermedades tropicales transmitidas por vectores artrópodos; se debe en primer lugar, a la estructura de la Vivienda y en segundo lugar, los Servicios Básicos que disponen los Pacientes.

Cabe recalcar que de acuerdo a su Dirección o Domicilio, se asociaron los índices censales, considerados para esta evaluación.

En lugares con viviendas en construcción, sin habitar, con pozo séptico (lugar para desecho de excrementos), piso de tierra, paredes de caña revestida y con un sistema de recolección de basura esporádico; existe una gran incidencia de Contagio.

<sup>4</sup> Figura que muestra los segmentos con sus índices Elites, generado en el sistema.

Esto también se debe a que los mosquitos existen en lugares con olores fuertes que resultan de un pozo séptico y la basura. Son lugares de reproducción al igual que donde existen aguas estancadas.

La estructura de la vivienda puede ayudarnos a identificar la situación económica del lugar, es posible que en dichos sitios no hagan uso de insecticidas, repelentes, toldos ó mallas metálicas en puertas y ventanas, para evitar la picadura de los mosquitos.

De acuerdo a los datos del SNEM (Servicio Nacional de erradicación de la Malaria), Guayaquil es fumigada en su totalidad por áreas, existe una planificación para llegar a cada una de estas áreas.

Con esta información podemos concluir que muchos son los factores que ayudan a la reproducción del mosquito del Dengue y Paludismo, otras deben ser las medidas de prevención y campañas de capacitación para llegar a los diferentes lugares con las características anteriormente mencionadas.

## **CAPITULO 5. IMPLEMENTACIÓN DEL SISTEMA**

Se implementará un sistema informático automatizado que permitirá realizar el ingreso de los datos y el procesamiento de los mismos.

Entre las opciones de Mantenimiento y Proceso tenemos:

### ***Mantenimiento***

- Paciente
- Historial
- Ubicación
- Enfermedad
- Serología
- Índice
- Índice por Mes
- Índice por Ubicación
- Dirección por Código Censal

### ***Proceso***

- Verificar Datos
- Actualizar Datos
- Predicción con Árbol de Decisión

Las opciones de Mantenimiento permitirán: Ingresar, Modificar, Eliminar, Consultar de forma específica y general. Cada una de estas pantallas evalúa la información que se ingrese, aceptando solo información valida y necesaria. El sistema verifica los datos obligatorios de ingresar, al momento de grabar, y sólo si son ingresados le permitirá grabar los datos.

Las opciones de Proceso ejecutarán la acción especificada en cada uno de los botones que tenga la pantalla.

## **CONCLUSIÓN**

Se ha proyectado que el Hospital Docente de Infectología "Dr. José Daniel Rodríguez Maridueña" sea a futuro un Hospital de Especialidades de enfermedades infectocontagiosas y tropicales, donde se brinde al paciente una atención oportuna con la calidad que se merece, dado por un equipo básico integral de Salud, solidarizándose con los usuarios y aplicando nuevas estrategias de gestión a través de la estructura funcional por procesos.

Frente a los avances tecnológicos, se ve la oportunidad de cubrir la necesidad de tomar decisiones para erradicación de la malaria. El presente trabajo ha sido desarrollado con el propósito de suplir y utilizar información relevante y necesaria para la determinación de los factores que influyen en el contagio de enfermedades transmitidas por vectores artrópodos; permitiendo analizar las diferentes condiciones que conllevan a que suceda el contagio y poder actuar con prevención.

## **RECOMENDACIÓN**

Nuestro producto es una ayuda para evaluar las condiciones por las que se presenta la enfermedad y sirve como herramienta para crear nuevos medios de prevención.

Consideramos que es importante realizar campañas para la prevención de estas enfermedades tropicales que pueden ocasionar hasta la muerte. Llevando un control y las precauciones del caso, podremos ayudar al bienestar de la familia, recordando siempre que un País sano es un país con esperanza.

## **REFERENCIAS**

### **a) Libros con edición.**

Data Mining Techniques for Marketing, Sales, and Customer Relationship Management - Segunda Edición (Michael J.A. Berry & Gordon S. Linoff).

Data Mining Practical Machine Learning Tools and Techniques - Segunda Edición (Ian H. Witten & Eibe Frank).

### **b) Referencias de Internet**

<http://www.inflexa.com/jsp/template.jsp?pag=mineria-datos.htm&mnu=mnu-mineria.htm>

<http://monografía/MINERIA\MineriaDatosBressan.htm>

<http://www.itba.edu.ar/capis/webcapis/RGMITBA/articulosrgm/R-ITBA-26-datamining.pdf>

<http://www.it.uc3m.es/jvillena/irc/practicas/04-05/16mem.pdf#search=%22como%20usar%20weka%22>