



**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL  
FACULTAD DE INGENIERÍA EN ELECTRICIDAD Y COMPUTACIÓN**

**“SISTEMA PARA ANÁLISIS DE COMPORTAMIENTO DE  
CONSUMO DE LA POBLACIÓN DIABÉTICA ECUATORIANA  
APLICANDO TÉCNICAS DE MINERÍA DE DATOS”**

**PREVIO A LA OBTENCIÓN DEL TÍTULO DE:**

**INGENIERO EN COMPUTACION  
ESPECIALIZACION EN SISTEMAS TECNOLOGICOS  
INGENIERO EN COMPUTACION  
ESPECIALIZACION EN SISTEMAS TECNOLOGICOS  
INGENIERO EN COMPUTACION  
ESPECIALIZACION EN SISTEMAS DE INFORMACION**

Presentada por:

**Alfredo Salvador Cáceres Zambrano.**

**Mario Rubén Iturralde Orellana.**

**Eduardo David Perugachi Rojas.**

**Guayaquil - Ecuador**

**2006**

## **AGRADECIMIENTO**

Alfredo Cáceres

Agradezco a Dios, a nuestros padres, maestros y amigos que nos han iluminado y guiado nuestra vida estudiantil.

Mario Iturralde

A Dios, incondicional soporte y ayuda, a mis padres y hermanos, por su incansable preocupación y aliento y a Natalia por su apoyo en todo momento.

David Perugachi

A todos quienes con su cariño, inteligencia y paciencia supieron ser la mano amiga y la palabra exacta en los momentos difíciles. Gracias  
ONLYCONTROL.

## DEDICATORIAS

Alfredo Cáceres

A todas las personas, compañeros y seres queridos que siempre nos dieron fuerzas y motivaciones para cumplir nuestra meta, ser profesional.

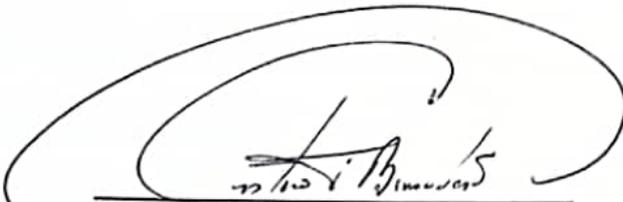
Mario Iturralde

A mis padres, quienes no desmayaron en su esfuerzo aliento y apoyo. Y a Natalia por estar presente siempre cuando más la necesito

David Perugachi

A Dios por ser luz, a mis padres por su sacrificio que me dio lo necesario para enfrentar la vida, a Susana por ser todo lo que necesito en mi vida.

TRIBUNAL DE GRADO



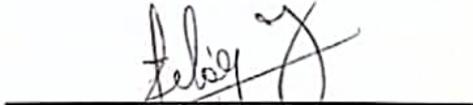
---

Ing. Gustavo Bermúdez Flores  
DECANO DE LA FIEC



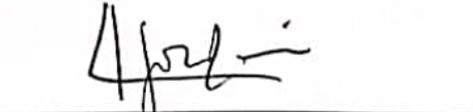
---

Ing. Fabricio Echeverría Briones  
DIRECTOR DE TÓPICO



---

Dr. Enrique Peláez Jarrín  
VOCAL PRINCIPAL



---

Ing. Carlos Jordán Villamar  
VOCAL PRINCIPAL

ESCUELA SUPERIOR POLITÉCNICA  
DEL CAYAMA  
FACULTAD DE INGENIERÍA  
BIBLIOTECA  
INV. No. \_\_\_\_\_

## DECLARACIÓN EXPRESA

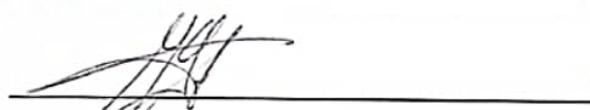
"La Responsabilidad del contenido de esta Tesis de Grado, nos corresponde exclusivamente; y el patrimonio intelectual de la misma a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL".

(Reglamento de Graduación de la ESPOL).



---

Alfredo Salvador Cáceres Zambrano



---

Mario Rubén Iturralde Orellana



---

Eduardo David Perugachi Rojas

## INDICE GENERAL

AGRADECIMIENTO.....	I
I	
DEDICATORIAS .....	III
TRIBUNAL DE GRADO .....	IV
DECLARACIÓN EXPRESA .....	V
INTRODUCCIÓN .....	1
CAPÍTULO 1 .....	6
1. ANTECEDENTES Y JUSTIFICACIÓN .....	6
1.1. OBJETIVOS.....	7
1.2. ANTECEDENTES .....	7
1.3. ENTENDIENDO LA DIABETES .....	8
1.3.1. DEFINICIÓN.....	8
1.3.2. FACTORES DE RIESGO .....	14
1.4. LÍNEAS DE PRODUCTOS PARA DIABÉTICOS .....	17
1.4.1. PROVEEDORES DEL MERCADO ECUATORIANO .....	21
1.4.2. COMPORTAMIENTO DEL MERCADO.....	22
1.5. JUSTIFICACIÓN .....	23
1.6. ALCANCES.....	24

CAPÍTULO 2.....	26
2. LA DIABETES Y EL CÍRCULO VIRTUOSO DE LA MINERÍA DE DATOS.....	26
2.1. ENFOQUE DE LA DIABETES CON RESPECTO A LA MINERÍA DE DATOS. ....	27
2.1.1. PERFILES DE DIABÉTICOS .....	28
2.1.2. COMPORTAMIENTO DE CONSUMO .....	28
2.2. TÉCNICAS DE MINERÍA DE DATOS Y MODELOS.....	29
2.2.1. CONCEPTOS Y DEFINICIONES DE LA MINERÍA DE DATOS.....	31
2.2.2. TIPOS DE MODELO DE LA MINERÍA DE DATOS.....	33
2.2.2.1. MODELO PREDICTIVO .....	38
2.2.2.2. MODELO DESCRIPTIVO .....	39
CAPÍTULO 3.....	41
3. MODELOS APLICADOS EN EL DESARROLLO (MINERÍA DE DATOS Y LENGUAJE DE MODELADO UNIFICADO).....	41
3.1. MODELOS DE MINERÍA DE DATOS APLICADOS.....	42
3.1.1. MODELO PREDICTIVO .....	49
3.1.2. MODELO PREDICTIVO .....	51

3.1.3. OPORTUNIDADES Y PROBLEMAS DE MODELOS APLICADOS .....	52
3.1.4. ANÁLISIS DE LOS DATOS Y SU ALCANCE (VISTA MINABLE).....	55
3.1.5. DATOS DE ENTRENAMIENTO DEL MODELO.....	55
3.1.6. DATOS DE PRUEBA DEL MODELO .....	56
3.2. MODELOS DE UML.....	57
3.2.1. MODELADO DE OBJETOS .....	59
3.2.2. MODELADO DINÁMICO .....	93
3.2.3. MODELADO FUNCIONAL .....	94
CAPÍTULO 4.....	96
4. DISEÑO E IMPLEMENTACIÓN DEL SISTEMA .....	96
4.1. VISIÓN GENERAL DEL DISEÑO .....	97
4.2. ELECCIÓN DE LAS HERRAMIENTAS DE DESARROLLO .....	102
4.2.1. CRITERIOS PARA LA ELECCIÓN DE LAS HERRAMIENTAS.....	102
4.2.2. JUSTIFICACIÓN PARA LA ELECCIÓN DE LAS HERRAMIENTAS SELECCIONADAS.....	104
4.3. RECOLECCIÓN E INTEGRACIÓN DE INFORMACIÓN PARA LA IMPLEMENTACIÓN DEL MODELO .....	105

4.4. TRANSFORMACIÓN DE LOS DATOS OBTENIDOS PARA REALIZAR EL MODELO .....	107
4.5. DISEÑO DEL MODELO RELACIONAL .....	108
4.6. DISEÑO DEL MODELO LÓGICO .....	112
4.7. DISEÑO MULTIDIMENSIONAL DE LA BASE DE DATOS DEL SISTEMA.....	113
4.8. IMPLEMENTACIÓN DEL MÓDULO PARA REGISTRO DE INFORMACIÓN .....	117
4.9. IMPLEMENTACIÓN DE MINERÍA DE DATOS Y ALGORITMOS DE INFERENCIA.....	123
4.9.1. IMPLEMENTACIÓN DEL MODULO PARA LA EJECUCIÓN DEL ANÁLISIS DE VENTAS .....	123
4.9.2. IMPLEMENTACIÓN DEL MODULO PARA ANÁLISIS DE PERFILES .....	131
4.10. RESULTADOS OBTENIDOS A PARTIR DE LAS TÉCNICAS DE MINERÍA DE DATOS APLICADAS .....	135
4.11. RESULTADOS DESEADOS VERSUS RESULTADOS OBTENIDOS.....	136
CAPÍTULO 5.....	137
5. FACTIBILIDAD Y COSTOS .....	137
5.1. PROPUESTA DEL VALOR COMERCIAL DEL SISTEMA.....	138

5.1.1. TCO COSTO TOTAL DE LA INVERSION.....	142
5.1.2. ROI RETORNO SOBRE LA INVERSION.....	145
5.2. EVALUACION DE COSTOS.....	148
CONCLUSIONES Y RECOMENDACIONES.....	158
BIBLIOGRAFÍA.....	160
ANEXOS.....	167
ANEXO 1.....	168
ANEXO 2.....	173
ANEXO 3.....	177
ANEXO 4.....	181
ANEXO 5.....	184

## ÍNDICE DE FIGURAS

Figura 3.1 Ciclo virtuoso de minería de datos	47
Figura 3.2 Diagrama de casos de uso	58
Figura 3.3 Diagrama de clases del sistema	71
Figura 3.4 Diagrama de interacción de objeto el administrador crea la encuesta exitosamente	72
Figura 3.5 Diagrama de interacción de objeto el administrador no puede crear la encuesta exitosamente	73
Figura 3.6 Diagrama de interacción de objeto el usuario envía sugerencia exitosamente	74
Figura 3.7 Diagrama de interacción de objeto el administrador no puede enviar las sugerencias exitosamente	75

Figura 3.8 Diagrama de interacción de objeto el usuario consulta información de diabetes exitosamente\_\_\_\_\_76

Figura 3.9 Diagrama de interacción de objeto el usuario no puede consultar información de diabetes exitosamente\_\_\_\_\_77

Figura 3.10 Diagrama de interacción de objeto el administrador genera análisis de reglas de asociación exitoso\_\_\_\_\_78

Figura 3.11 Diagrama de interacción de objeto el administrador no genera análisis de reglas de asociación exitoso\_\_\_\_\_79

Figura 3.12 Diagrama de interacción de objeto registro de usuario exitoso\_\_\_\_\_80

Figura 3.13 Diagrama de interacción de objeto registro de usuario no exitoso\_\_\_\_\_81

Figura 3.14 Diagrama de interacción de objeto el ingreso de usuario administrador exitoso\_\_\_\_\_82

Figura 3.15 Diagrama de interacción de objeto ingreso de usuario administrador no exitoso\_\_\_\_\_83

Figura 3.16 Diagrama de interacción de objeto el administrador procesa base de datos exitosamente_____	84
Figura 3.17 Diagrama de interacción de objeto el administrador no puede procesar base de datos_____	85
Figura 3.18 Diagrama de interacción de objeto el administrador genera análisis exitosamente_____	86
Figura 3.19 Diagrama de interacción de objeto el administrador no genera análisis exitosamente_____	87
Figura 3.20 Diagrama de interacción de objeto el administrador almacena análisis exitosamente_____	88
Figura 3.21 Diagrama de interacción de objeto el administrador no almacena análisis exitosamente_____	89
Figura 3.22 Diagrama de interacción de objeto el administrador consulta análisis exitosamente_____	90
Figura 3.23 Diagrama de interacción de objeto el administrador no consulta análisis exitosamente_____	91
Figura 3.24 Diagrama de estado: Naive Bayes_____	92

Figura 3.25 Diagrama de estado: Reglas de asociación	92
Figura 3.26 Diagrama de flujo: Naive Bayes	93
Figura 3.27 Diagrama de flujo: Reglas de asociación	94
Figura 4.1 Arquitectura de tres capas	97
Figura 4.2 Pantalla de inicio del sistema	100
Figura 4.3 Diagrama entidad-relación. Encuesta en línea	108
Figura 4.4 Diagrama entidad-relación. Transaccional de ventas	109
Figura 4.5 Base de Datos SACC	111
Figura 4.6 Diseño multidimensional. Encuesta en línea	112
Figura 4.7 Diseño multidimensional: Transaccional de ventas	114
Figura 4.8 Diseño multidimensional: Módulo análisis de perfiles.	115
Figura 4.9 Ventana de registro de usuarios	117
Figura 4.10 Opciones del perfil del administrador	118
Figura 4.11 SACC: Creación de pregunta	119

Figura 4.12 SACC: Eliminación de pregunta	120
Figura 4.13 SACC: Modificar pregunta	121
Figura 4.14 SACC: Encuesta usuario común	122
Figura 4.15 SACC: Selección de rango de fechas para el modelo	123
Figura 4.16 SACC: Selección de los parámetros de confianza y soporte	124
Figura 4.17 SACC: Resultados de modelo descriptivo	129
Figura 4.18 SACC: Procesar datos encuesta Naive Bayes	130
Figura 4.19 SACC: Definir variables para análisis	131
Figura 4.20 SACC: Definir variable objetivo y variables condición	132
Figura 4.21 SACC: Resultados de modelo predictivo	133

## ÍNDICE DE TABLAS

Tabla 1.1 Síntomas de la diabetes	10
Tabla 1.2 Laboratorios farmacéuticos de productos diabéticos Ecuador	21
Tabla 1.3 Diez principales causas de mortalidad año 2003	23
Tabla 2.1 Algoritmos para minería de datos predictivos	39
Tabla 2.2 Algoritmos para minería de datos descriptivos	40
Tabla 3.1 Cuadro Comparativo Reglas de Asociación y Clasificación	45
Tabla 4.1 Tabla comparativa de plataformas de programación	102
Tabla 4.2 Rango IMC	106
Tabla 4.3 Rango Edad	107
Tabla 5.1 Análisis FODA de SACC	140
Tabla 5.2 Modelo para calcular TCO	143
Tabla 5.3 Componentes para análisis TCO	144
Tabla 5.4 Costos fase de adquisición	152
Tabla 5.5 Costos fase de operación y desarrollo	153
Tabla 5.6 Costos fase de ampliaciones y cambios	153

Tabla 5.7 Ingresos totales anuales _____	154
Tabla 5.8 ROI total del Proyecto _____	154
Tabla 5.9 Flujo de Efectivo desde mes 0 hasta mes 9 _____	155
Tabla 5.10 Flujo de Efectivo desde mes 10 hasta mes 18 _____	156

# INTRODUCCIÓN

La minería de datos es el proceso de extracción de información y conocimiento significativo sobre un conjunto de datos no necesariamente heterogéneo y cuyo origen se da principalmente en los sistemas transaccionales. El objetivo principal es encontrar tendencias y correlaciones que permitan entender, descubrir y prever escenarios dentro del giro del negocio donde esta técnica fuese aplicada [1].

La Diabetes Mellitus es un grupo de trastornos metabólicos de carácter crónico caracterizados por un elemento común, la hiperglucemia, que contribuye al desarrollo de complicaciones macro vasculares, micro vasculares y neuropáticas, lo que la sitúa como una de las principales causas de mortalidad de las sociedades desarrolladas o en vías de desarrollo. Afecta a gran número de personas, con un aumento constantes de la prevalencia de la Diabetes Mellitus 1 y alarmante de la Diabetes Mellitus 2, esto último lo relaciona la Organización Mundial para la Salud con el crecimiento y envejecimiento de la población, el incremento de la obesidad, hábitos erróneos de la alimentación y modos de vida sedentarios, así mismo ocurre con la emergente Diabetes Mellitus 2 asociada a la obesidad en niños [2]. Todo esto lleva a que represente un problema personal y de salud pública de enormes proporciones

El objetivo del presente trabajo es aplicar minería de datos sobre una problemática de nivel social; para tener como resultado una herramienta que permita conocer los perfiles del consumidor de productos para diabéticos; así como también las tendencias que tiene una persona a sufrir esta enfermedad debido a su estilo de vida, factores de salud y características étnicas.

El Sistema para Análisis de Comportamiento de Consumo (SACC) tiene como propósito mostrar cuáles son las características demográficas de los diabéticos en el Ecuador; permitiendo describir cuáles son los patrones de consumo de productos especializados que ayudan al paciente al tratamiento de la diabetes y mejora de los síntomas, dando como resultado el ofrecer a los distribuidores de dichos productos potenciales oportunidades de mejora en la promoción y distribución de los mismos.

Además definirá la tendencia de una persona a desarrollar diabetes dependiendo de factores de comportamiento, consumo, hereditarios y de cuidado de su salud

En el capítulo uno se describe la enfermedad, su incidencia en Ecuador, revisión de productos destinados para el tratamiento de la misma y alcances generales del sistema

En el capítulo dos se expone de manera general el problema a resolver y se exponen las técnicas de minería disponibles para realizarlo.

En el capítulo tres se determinan cuales son los algoritmos más idóneos para la resolución del problema, se explica su funcionamiento y se muestra mediante lenguaje UML el diseño propuesto para la solución.

En el capítulo cuatro se determinan las herramientas para el desarrollo del sistema; además, se ilustran aspectos del diseño e implementación del sistema, tales como: el modelo relacional, modelo lógico, modelo multidimensional, módulos del sistema y resultados obtenidos de la ejecución de los algoritmos.

En el capítulo cinco se exponen las razones por las que se decidió diseñar e implementar esta herramienta de tipo informático. El sólo hecho de que agilice un proceso manual o que ofrezca una nueva manera de hacer las cosas a los usuarios no son razones suficientes para invertir en su desarrollo. Este capítulo demuestra las ventajas que

obtienen los usuarios de este sistema en comparación a los costos que implicó la realización del mismo.

# **CAPÍTULO 1**

## **1. ANTECEDENTES Y JUSTIFICACIÓN**

## **1.1. OBJETIVOS.**

Los objetivos para la implementación del SACC (Sistema para Análisis de Comportamiento de Consumo) son:

- Encontrar patrones de consumo y ventas de productos para pacientes diabéticos mediante reglas de asociación.
- Encontrar patrones de hábitos y perfiles de los pacientes diabéticos.
- Construir una interfaz para el usuario del sistema que sea intuitiva y fácil de entender.
- Demostrar que el proyecto de inversión es rentable y atractivo.

## **1.2. ANTECEDENTES**

Dentro del campo de las aplicaciones para el descubrimiento de conocimiento utilizadas en el Ecuador existen dos tendencias bien marcadas dentro del ámbito empresarial. La primera tendencia esta en las empresas que no utilizan una herramienta que les permita analizar los patrones de consumo de una población aplicada a cualquier producto, esta tarea generalmente debe ser realizada (o más bien simulada) por los departamentos de mercadeo mediante la utilización de

información histórica, análisis de variables estadísticas y casi siempre es representada mediante un archivo en formato Excel.

La segunda tendencia esta en el desarrollo de aplicaciones que se basan específicamente en la lógica del negocio de la empresa, de tal forma que se convierten en soluciones rígidas ya que no pueden ser utilizadas en otra empresa aunque ésta sea de un giro similar.

SACC está orientado al análisis de patrones de consumo de productos para diabéticos, analiza los patrones de venta de los productos, así como también el perfil de las personas que los consumen.

Sin embargo, utilizar SACC para otro tipo de análisis no resultaría muy complicado ya que se pueden parametrizar las fuentes de datos que alimentan al sistema.

### **1.3. ENTENDIENDO LA DIABETES**

#### **1.3.1. DEFINICIÓN**

La diabetes es una enfermedad en la que el organismo no produce insulina o no la utiliza adecuadamente. La insulina es una hormona necesaria para transformar el azúcar, el almidón y otros alimentos en la

energía que necesitamos para nuestra vida cotidiana. Aunque tanto los factores genéticos como medioambientales, tales como la obesidad y la falta de ejercicio, parecen desempeñar roles importantes, sin embargo la causa de la diabetes continúa siendo un misterio [3].

La diabetes puede provocar graves complicaciones y muerte prematura, pero las personas con diabetes pueden tomar medidas para controlar la enfermedad y disminuir el riesgo de sufrir complicaciones. La diabetes es una enfermedad crónica que no tiene cura.

En los Estados Unidos, hay 20.8 millones de personas, o el 7% de la población, que sufren de diabetes. Si bien ya se han diagnosticado con diabetes alrededor de 14,6 millones de personas, desafortunadamente 6,2 millones (o casi un tercio) no saben que padecen la enfermedad [2].

## **Síntomas**

Frecuentemente la diabetes pasa desapercibida debido a que sus síntomas aparentan ser inofensivos. Algunos de los síntomas de la diabetes incluyen:

<b>Síntomas comunes en la diabetes</b>	
<b>Diabetes tipo 1</b>	<b>Diabetes tipo 2<sup>1</sup></b>
-Ganas de orinar frecuentes	-Cualquiera de los síntomas de la diabetes tipo 1
-Sed inusual	-Infecciones frecuentes
-Hambre excesiva	-Visión borrosa
-Pérdida de peso inusual	-Heridas/moretos que tardan en curarse
-Fatiga extrema	-Cosquilleo/ adormecimiento en las manos o los pies
-Irritabilidad	-Infecciones recurrentes en la piel, las encías o la vejiga

**Tabla 1.1** Síntomas de la diabetes [4]

### **Métodos de detección de la enfermedad.**

Para determinar si un paciente tiene diabetes latente o diabetes, los profesionales de la salud realizan una prueba de glucosa en el plasma en ayunas (GPA) o una prueba oral de tolerancia a la glucosa (POTG) [5]. Con cualquiera de esas dos pruebas, puede diagnosticarse la diabetes latente o la diabetes. La American Diabetes Association recomienda la prueba de GPA porque es más económica, rápida y fácil de realizar.

Si en la prueba de GPA se detecta un nivel de glucosa en la sangre en ayunas entre 100 y 125 mg/dl, significa que la persona tiene una diabetes latente. Una persona con un nivel de glucosa en la sangre en ayunas de 126 mg/dl o superior padece diabetes [5].

<sup>1</sup> A menudo, las personas con diabetes tipo 2 no presentan síntomas

En el caso de la POTG, se mide el nivel de glucosa en la sangre de una persona que está en ayunas, y dos horas después de ingerir una bebida rica en glucosa. Si el nivel de glucosa en la sangre se encuentra entre 140 y 199 mg/dl pasadas las dos horas, significa que la persona tiene diabetes latente. Si el nivel de glucosa en la sangre después de las dos horas es de 200 mg/dl o superior, significa que la persona sufre de diabetes.

### **Tipos de Diabetes.**

Aproximadamente 230 millones de personas en el mundo sufren de diabetes en sus dos variedades: el "Tipo 1", grave y cuyos pacientes necesitan insulina a diario, y el "Tipo 2", que es una fase inicial de la enfermedad que se controla con pastillas [6].

### **Diabetes Tipo 1**

Anteriormente, se denominaba diabetes mellitus insulino dependiente (DMID) o diabetes juvenil a la diabetes tipo 1. La diabetes tipo 1 se desarrolla cuando el sistema inmunológico del cuerpo destruye las

células beta del páncreas, las únicas células del cuerpo que producen la hormona insulina que regula la concentración de glucosa en la sangre [7].

## **Diabetes Tipo 2**

La diabetes tipo 2 es la clase de diabetes más común. En este tipo de diabetes, el organismo no produce suficiente insulina o las células la ignoran. La insulina es necesaria para que el organismo pueda utilizar el azúcar [8]. El azúcar es el combustible esencial para las células del organismo y la insulina transporta el azúcar en la sangre hacia las células. Cuando la glucosa se acumula en la sangre en lugar de penetrar en las células, pueden presentarse dos problemas:

- En lo inmediato, las células pueden quedar privadas de energía.
- Con el paso del tiempo, los niveles altos de glucosa en la sangre pueden dañar los ojos, los riñones, los nervios o el corazón.

Descubrir que uno sufre de diabetes puede causar temor pero no hay que desesperarse. La diabetes tipo 2 es una enfermedad seria, pero las personas con diabetes pueden llevar una vida saludable y vivir por muchos años.

Si bien las personas de todas las edades y razas pueden padecer diabetes, algunos grupos corren más riesgos que otros de desarrollar diabetes tipo 2. La diabetes tipo 2 es más común entre los afroamericanos, los latinos, los nativos americanos y los asiáticos americanos/ de las islas Pacífico, así como los ancianos. Se calcula que aproximadamente entre el 90 y el 95% de los norteamericanos (17 millones) diagnosticados con diabetes padecen diabetes tipo 2 [9]. Este porcentaje se mantiene en el resto del mundo.

### **Diabetes gestacional**

La diabetes gestacional es una forma de intolerancia a la glucosa que se diagnostica a algunas mujeres durante el embarazo. La diabetes gestacional se manifiesta con mayor frecuencia en afroamericanos, hispanos/ latinos estadounidenses e indios americanos. También es más común en mujeres obesas y en mujeres con antecedentes familiares de diabetes. Durante el embarazo, la diabetes gestacional requiere de un tratamiento para normalizar los niveles de glucosa en la sangre de la madre, con el fin de evitar complicaciones en el bebé. Luego del embarazo, entre el 5% y el 10% de las mujeres que tuvieron diabetes gestacional desarrollan diabetes tipo 2. Las mujeres que han

tenido diabetes gestacional tienen una probabilidad de entre un 20% y un 50% de desarrollar diabetes en los 5-10 años siguientes. El 70% de las mujeres que tuvieron diabetes gestacional desarrollarán diabetes tipo 2 en algún momento de su vida [10].

Existen otros tipos de diabetes que se producen debido a trastornos genéticos específicos (como diabetes de adulto en la juventud), cirugías, medicamentos, desnutrición, infecciones y otras enfermedades. Esos tipos de diabetes representan entre el 1% y el 5% del total de los casos diagnosticados.

### **1.3.2. FACTORES DE RIESGO**

Todas las personas con diabetes sufren de hiperglucemia de vez en cuando. Si no se trata, puede ser un problema serio. La hiperglucemia constituye una de las principales causas de muchas de las complicaciones que sufren las personas con diabetes. Por ese motivo, es importante saber qué es la hiperglucemia, cuáles son los síntomas y cómo tratarla [11].

Hay muchas cosas que pueden causar hiperglucemia. Por ejemplo, si se tiene diabetes tipo 1, tal vez no se haya inyectado la cantidad

suficiente de insulina. Si se sufre de diabetes tipo 2, quizás el organismo sí cuente con la cantidad suficiente de insulina, pero no es tan eficaz como debería serlo. El problema quizás sea que se comió más de lo planeado o realizó menos actividad física de la programada. El estrés que provoca una dolencia como, por ejemplo, un resfrío o una gripe también podría ser la causa. Otras clases de estrés, tales como los conflictos familiares, los problemas en la escuela o los problemas de pareja, también podrían causar hiperglucemia.

Las señales y los síntomas incluyen: alto nivel de azúcar en la sangre, altos niveles de azúcar en la orina, aumento de la sed y necesidad frecuente de orinar.

El medir con frecuencia el nivel de azúcar en la sangre es una de las medidas preventivas que puede tomar para mantener la diabetes bajo control. Se debe consultar con un médico con qué asiduidad se debe controlar y cuáles deben ser los niveles de glucemia. Para evitar otros síntomas de la hiperglucemia es necesario que se controle y se trate los altos niveles de azúcar en la sangre sin perder tiempo.

Es importante tratar la hiperglucemia apenas se detecta. Si no se lo hace, se puede sufrir una afección denominada cetoacidosis (coma

diabético). La cetoacidosis se desarrolla cuando el organismo no cuenta con la suficiente cantidad de insulina. Sin la insulina, el organismo no puede utilizar la glucosa como combustible. Por ende, el organismo descompone las grasas (lípidos) para utilizarlas como energía.

Cuando el organismo descompone las grasas, aparecen productos residuales denominados cetonas. El organismo no tolera grandes cantidades de cetonas y trata de eliminarlos a través de la orina. Desafortunadamente, el organismo no puede deshacerse de todas las cetonas y, por lo tanto, se acumulan en la sangre. Como resultado, puede presentarse un cuadro de cetoacidosis.

La cetoacidosis puede provocar la muerte y requiere tratamiento inmediato. Los síntomas incluyen:

- Dificultad para respirar.
- Aliento con olor frutal.
- Náuseas y vómitos.
- Boca muy seca.

Por lo general, la actividad física ayuda a disminuir los niveles de glucemia. Sin embargo, si el nivel de azúcar se encuentra por encima

de 240 mg/dl, es necesario controlar la orina para detectar cetonas. En caso de presentar cetonas en la orina, no se debe realizar actividad física. Si tiene cetonas en la orina y realiza ejercicios, el nivel de azúcar en la sangre puede elevarse aún más.

#### **1.4. LÍNEAS DE PRODUCTOS PARA DIABÉTICOS**

Existen varias líneas de productos en el mercado ecuatoriano las cuales se clasifican en 3 grandes grupos:

##### **Medicamentos vía oral**

Con frecuencia, el primer tratamiento para la diabetes tipo 2 es la planificación de las comidas a fin de controlar el nivel de glucosa (azúcar) en la sangre, la pérdida de peso y la actividad física. En algunas ocasiones, esas medidas no alcanzan para reducir el nivel de glucosa en la sangre y acercarlo al nivel normal. El paso siguiente es administrar un medicamento para reducir el nivel de glucosa en la sangre. Existen dos tipos de medicamentos: los medicamentos orales (pastillas) y las inyecciones de insulina. Las pastillas para la diabetes no son insulina [12].

## **Cómo funcionan**

En personas con diabetes, el nivel de glucosa en la sangre es demasiado alto. Ese alto nivel se produce porque la glucosa permanece en la sangre en lugar de ingresar en las células, donde pertenece. Pero para que la glucosa pueda ingresar en una célula, la insulina debe estar presente y la célula debe tener “hambre” de glucosa.

Las personas con diabetes tipo 1 no producen insulina. Para ellas, las inyecciones de insulina son la única manera de mantener bajo el nivel de glucosa en la sangre. Las personas con diabetes tipo 2 suelen tener dos problemas: no producen suficiente insulina y las células de sus cuerpos no parecen incorporar glucosa con la avidez que debieran.

En la actualidad, todas las pastillas para la diabetes en venta pertenecen a cinco clases de drogas: sulfonilúreas, meglitinidas, biguanidas, tiazolidinedionas e inhibidores de las alfa-glucosidasas. Esas cinco clases de drogas funcionan de diferentes maneras para reducir el nivel de glucosa en la sangre [12].

## **Medicamentos inyectables**

Dentro del páncreas, las células beta producen la hormona insulina. Con cada comida, las células beta liberan insulina para ayudar al organismo a utilizar o almacenar la glucosa sanguínea que obtienen de los alimentos. En las personas que tienen diabetes tipo 1, el páncreas ya no fabrica más insulina. Las células beta han sido destruidas y la persona necesita inyectarse insulina para poder utilizar la glucosa de los alimentos [13].

Las personas con diabetes tipo 2 sí producen insulina, pero el organismo no responde adecuadamente a esa hormona. Algunas personas con diabetes tipo 2 necesitan tomar medicamentos para la diabetes o inyectarse insulina para ayudar a sus organismos a que utilicen la glucosa para obtener energía.

La insulina no puede administrarse en pastillas porque se destruiría durante la digestión, al igual que las proteínas de los alimentos. La insulina debe inyectarse en la grasa subcutánea para que penetre en la sangre.

### **Línea de consumo para diabéticos.**

Existen varios productos para la línea de consumo para diabéticos, entre los cuales resaltan las siguientes especialidades:

- Alergias.
- Barras de chocolate.
- Expectorantes.
- Gripes y Catarros.
- Sustitutivos de azúcar.
- Lociones y estimulantes para la circulación.
- Productos para la piel seca.
- Productos para los ojos.
- Vitaminas.

Existen muchas clases de insulina para distintas situaciones y estilos de vida y, en el Ecuador se distribuyen alrededor de 6 tipo de insulina en comparación de los mas de 20 tipos de insulina disponibles en los Estados Unidos [14], Esos tipos de insulina difieren en la forma en que están elaborados, la forma en que actúan dentro del organismo y el precio.

La insulina se fabrica en el laboratorio para que sea idéntica a la insulina humana, o bien es de origen animal (porcino). La disponibilidad de la insulina de origen animal en el futuro es incierta.

#### 1.4.1. PROVEEDORES DEL MERCADO ECUATORIANO

A continuación se detalla una lista de los laboratorios que comercializan productos para diabéticos.

<b>Laboratorios Farmacéuticos</b>
CHILE
COMERCIOSA
CORPORACION MAGMA
ECUAQUIMICA
ELI LILLY
GLAXOSMITHKLINE
GRUNENTHAL
JULPHARMA
MENARINI
MERCK
MK
NOVARTIS
OSMOPHARM
PFIZER
PHARMABRAND
ROCHE
ROEMMERS
SANOFI-AVENTIS
SERVIER

**Tabla 1.2** Laboratorios farmacéuticos de productos diabéticos Ecuador [15]

### **1.4.2. COMPORTAMIENTO DEL MERCADO**

Es necesario reconocer que no existe ningún estudio serio y público acerca del comportamiento del consumo de productos para la diabetes en el Ecuador, mas allá de los análisis que puedan realizar ciertas casas farmacéuticas con respecto a la comercialización de sus productos.

SACC tiene como objetivo ser una herramienta que permita llegar a conclusiones que permitan entender el comportamiento del consumo de productos para la diabetes, que está relacionado directamente con la evolución misma de la enfermedad, y así poder contar con índices tan específicos como los encontrados en estudios realizados en países tales como Estados Unidos.

En el anexo 1 podrá encontrar una tabla detallada de las ventas de productos para el tratamiento de la diabetes y su evolución con respecto al 2005 en el Ecuador.

## 1.5. JUSTIFICACIÓN

Según el censo nacional desarrollado en el 2003 por el INEC [16] la diabetes mellitus es la tercera causa de mortalidad en el Ecuador. Ver detalle en tabla a continuación.

N°	CAUSAS	NUMERO DE	%	TASA
ORDEN		MUERTES		
1°	Otras Enfermedades del Corazón	4830	9,00%	37,6%
2°	Enfermedades Cerebro Vasculares	2909	5,40%	22,7%
3°	Diabetes Mellitus	2434	4,50%	19,0%
4°	Neumonía	2428	4,50%	18,9%
5°	Enfermedades Isquémicas del Corazón	2296	4,30%	17,9%
6°	Enfermedades Hipertensivas	2195	4,10%	17,1%
7°	Accidentes de Transporte	2014	3,80%	15,7%
8°	Agresiones	1771	4,30%	13,8%
9°	Tumor Maligno del Estómago	1465	2,70%	11,4%
10°	Enfermedades del Hígado	141	2,60%	11,0%
11°	Síntomas, Signos y hallazgos anormales clínicos y de laboratorio, no clasificados en otra parte	6714	12,50%	
12°	Las demás causas de mortalidad	23055	43,10%	
	<b>TOTAL DEFUNCIONES</b>	<b>53,521</b>	100%	
	Población Estimada	12,842,578		

**Tabla 1.3** Diez principales causas de mortalidad año 2003

Siendo la diabetes una enfermedad que puede ser prevenida mediante la aplicación de correctos hábitos alimenticios y de vida, es necesario que la comunidad ecuatoriana cuente con una herramienta que le permita conocer acerca de la enfermedad.

Otra razón es que tomando en cuenta que el paciente que padece diabetes es una persona que debe consumir “toda su vida” productos para el tratamiento de la enfermedad, de tal forma que éstos lo ayuden a mitigar los efectos de la enfermedad y continuar con un estilo de vida casi normal, esto nos indica que es un mercado objetivo con un gran potencial para aplicar campañas de mercadeo.

## **1.6. ALCANCES**

SACC aplica dos algoritmos de minería de datos, que serán explicados detalladamente más adelante, que le permiten enfocarse en dos aspectos importantes del análisis de comportamiento de consumo. El primero tiene que ver con el perfil del consumidor, aquí se toman en cuenta variables de tipo demográfico tales como la edad, peso, ciudad en la que reside, entre otros; variables para el análisis del consumo y variables de tipo historial médico. El otro aspecto que se analizará desde SACC, es el comportamiento del flujo de ventas de una línea de productos para el tratamiento de la diabetes. La información que despliega SACC es valiosa tanto para casas farmacéuticas (puede ser aplicado al desarrollo correcto de una campaña de mercadeo), así como también para el desarrollo de estudios de tipo médico, ya que se analizan variables de ambos tipos.

SACC cuenta además con una herramienta que le permite diseñar cuestionarios en línea de forma dinámica.

## **CAPÍTULO 2**

### **2. LA DIABETES Y EL CÍRCULO VIRTUOSO DE LA MINERÍA DE DATOS**

## **2.1. ENFOQUE DE LA DIABETES CON RESPECTO A LA MINERÍA DE DATOS.**

Como hemos visto en el capítulo 1, la diabetes es una enfermedad crónica que afecta a un porcentaje considerable de la población mundial. Según datos de la Organización Mundial de la Salud existen aproximadamente 230 millones de personas que padecen la enfermedad actualmente y se estima que la cifra llegue a los 350 millones para el año 2025 [6].

Para tener una idea de la magnitud del problema, notemos que solo en Estados Unidos 18 millones de personas padecen diabetes, 13 millones tienen diagnóstico, pero 5.2 millones no saben que la padecen [17]. Esta dramática realidad adquiere mayor relevancia si tomamos en cuenta que los latinos tienen casi el doble de probabilidad de desarrollar diabetes en comparación con el resto de la población [11]. En el Ecuador es la tercera causa de muerte (ver tabla 1.3).

En este marco de referencia inevitablemente saltan preguntas que despiertan nuestra preocupación de cara a la enfermedad, preguntas como por ejemplo: ¿padezco diabetes?, ¿voy a sufrir diabetes en un futuro?

### **2.1.1. PERFILES DE DIABETICOS**

Es claro que existen factores de riesgo que aumentan la probabilidad de desarrollar la enfermedad, por lo que el verdadero problema es encontrar que factor o combinación de factores se dan sobre una población en particular y que me permitirán hacer una estimación de la probabilidad de padecer diabetes de cualquier nuevo individuo que pertenezca a esa población.

Es necesario analizar la población ya que, como hemos visto, el entorno juega un papel importante en el desarrollo de la enfermedad, sin embargo analizar toda una población pudiera convertirse en una tarea que consumirá mucho tiempo, esfuerzo y dinero, por lo que es necesario hallar un mecanismo de análisis de grandes cantidades de información minimizando todos estos recursos.

### **2.1.2. COMPORTAMIENTO DE CONSUMO**

Por otra parte, las compañías farmacéuticas tienen un especial interés en pacientes diabéticos y en general en pacientes con enfermedades crónicas, ya que un paciente de alguna enfermedad crónica tendrá que consumir medicamentos durante toda su vida lo que lo hace un cliente perfecto.

Esta premisa crea la necesidad de cualquier casa farmacéutica de satisfacer al paciente para captar su fidelidad a una marca o producto y dado que existen varias casas farmacéuticas en el mercado (tabla 1.2) y tomando en cuenta de que los precios de los productos farmacéuticos son controlados por el estado a través del Ministerio de Salud Pública, los clientes se inclinarán por el que les ofrezca mayor valor agregado.

Debido a esto se plantea el problema de descubrir los patrones de consumo de la población diabética y así establecer tendencias de compra basado en parámetros como por ejemplo: ciudad, rango de fechas y cliente que den una ventaja a las empresas a la hora de comercializar sus productos.

En ambos casos es necesario hacer el análisis de grandes volúmenes de información tratando de minimizar los recursos invertidos en dicho análisis, por lo que a continuación veremos como la minería de datos puede ser una herramienta útil en esta tarea en particular.

## **2.2. TÉCNICAS DE MINERÍA DE DATOS Y MODELOS.**

La tecnología disponible actualmente permite almacenar, transferir y gestionar información crítica dentro de una organización. Sin embargo,

si esta información no se transforma en conocimiento útil, carece de valor.

La minería de datos es el proceso de extracción de información significativa de cualquier conjunto de datos. Las técnicas de minería de datos revelarán tendencias y correlaciones ocultas que permitan una mejor comprensión de los datos, esto es, generar conocimiento útil. Por lo tanto, la minería de datos es un proceso de soporte al usuario.

Esta tecnología ayuda a las organizaciones a centrarse en la información más significativa contenida en sus bases de datos corporativas. Mediante esta información es posible predecir nuevas perspectivas y situaciones futuras con mayor precisión, lo que facilita la toma de decisiones y, empleada de forma apropiada, podría suponer una ventaja competitiva.

Los avances en la recolección de datos, el amplio uso de códigos de barras y la automatización de muchos negocios, tanto privados como gubernamentales, han generado una ingente cantidad de datos. El análisis de este volumen de datos ha generado una necesidad urgente de nuevas técnicas y herramientas que puedan asistir en la

transformación automática e inteligente de toda esta información en conocimiento útil.

### **2.2.1. CONCEPTOS Y DEFINICIONES DE LA MINERÍA DE DATOS**

Según el libro “Descubrimiento de conocimiento en bases de datos” (Piatetski-Shapiro G); la minería de datos es el conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar el conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con el objeto de predecir de forma automatizada tendencias y comportamientos y/o descubrir de forma automatizada modelos previamente desconocidos” [18].

Los avances en minería de datos y en el descubrimiento del conocimiento, han impulsado varios campos de investigación como la estadística basada en técnicas de lógica difusa, bases de datos universales, máquinas que aprenden e inteligencia artificial, todos ellos parte del creciente campo de minería de datos. De forma general la minería de datos tiene una serie de ventajas:

- Contribuye a facilitar la toma de decisiones tácticas y estratégicas proporcionando un procedimiento automatizado para identificar

información clave desde volúmenes de datos generados por procesos tradicionales y de e-Business.

- Permite dar prioridad a decisiones y acciones.
- Proporciona criterios que faciliten la toma de decisiones a los usuarios del negocio que mejor entienden el problema y el entorno.
- Habitualmente, genera modelos descriptivos.
- Permite que relaciones ocultas e identificadas a través del proceso de la minería de datos sean expresadas como reglas de negocio o modelos predictivos.
- Un sistema de minería de datos permite analizar factores de influencia en determinados procesos, predecir o estimar variables o comportamientos futuros, segmentar o agrupar ítems similares, además de obtener secuencias de eventos que provocan comportamientos específicos.

- La llegada de la minería de datos se considera como la última etapa de la introducción de métodos cuantitativos, científicos en el mundo del comercio, industria y negocios.

### **2.2.2. TIPOS DE MODELO DE LA MINERÍA DE DATOS**

La naturaleza de las diferentes técnicas de minería de datos pueden ser clasificadas en dos grupos: análisis de minería de datos de tipo descriptivo y análisis de minería de datos de tipo predictivo [1].

El análisis descriptivo se enfoca en presentar al usuario de manera intuitiva y parametrizable la información contenida en los diversos almacenes de información usando principalmente técnicas para el análisis multidimensional de las bases de datos. Las tareas que están clasificadas dentro del análisis de tipo descriptivo son: agrupamiento (clustering), las reglas de asociación, las reglas de asociación secuenciales y las correlaciones [1].

Por otra parte, el análisis de tipo predictivo se enfoca en generar modelos que descubren relaciones ocultas y complejas a partir de diversas operaciones. Las tareas asociadas al análisis de tipo predictivo son: las redes neuronales, árboles de decisión, modelos de regresión, entre otros [1]. A continuación se detallan cada una de estas tareas:

**La clasificación:**

Es quizá la tarea más utilizada, en ella cada instancia o registro de la base de datos, pertenece a una clase, la cual se indica mediante el valor de un atributo al que llamamos instancia de la clase. Este atributo puede tomar diferentes valores discretos cada uno de los cuales corresponde a una clase. El resto de atributos de la instancia se utilizan para predecir la clase. El objetivo es predecir a que clase pertenecen las nuevas instancias. Mas concretamente, el objetivo es maximizar la razón de precisión de la clasificación de las nuevas instancias, la cual se calcula como el cociente entre las predicciones correctas y el número total de predicciones [1].

Existen varias tareas de clasificación como: el aprendizaje de “rankings”, el aprendizaje de preferencias, el aprendizaje de estimadores de probabilidad, etc.

**La regresión.**

Es también una tarea predictiva que consiste en aprender una función real que asigna a cada instancia un valor real. Esta es la principal diferencia respecto a la clasificación; el valor a predecir es numérico. El

objetivo en este caso es minimizar el error (generalmente el error cuadrático medio) entre el valor predicho y el valor real [1].

Por ejemplo, si un empresario quiere conocer cual es el costo de un nuevo contrato basándose en los datos correspondientes a contratos anteriores. Para ello usa una fórmula de regresión lineal, ajustando con los datos pasados la función lineal y usándola para predecir el costo en el futuro [1].

### **El agrupamiento (clustering):**

Es una tarea descriptiva por excelencia y consiste en obtener grupos “naturales” a partir de los datos. Hablamos de grupos y no de clases, por que, a diferencia de la clasificación, en lugar de analizar datos etiquetados con una clase, los analiza para generar esta etiqueta. Los datos son agrupados basándose en el principio de maximizar la similitud entre los elementos de un grupo minimizando la similitud entre los distintos grupos. Es decir, se forman grupos tales que los objetos de un mismo grupo son muy similares entre sí y, al mismo tiempo, son muy diferentes a los objetos de otro grupo. Al agrupamiento también se le suele llamar segmentación, ya que parte o segmenta los datos en grupos que pueden ser o no disjuntos [1].

El agrupamiento es a menudo el primer paso en los análisis de Minería de Datos y sirve para apoyar la segmentación de clientes basándose en sus atributos demográficos. Por ejemplo, los hábitos de compra de múltiples segmentos de población podrían ser comparados para determinar a cuales segmentos enfocar las próximas campañas de mercadeo y publicidad [1].

### **La correlación**

Es una tarea descriptiva que se usa para examinar el grado de similitud de dos variables numéricas. Una fórmula estándar para medir la correlación lineal es el coeficiente de correlación  $r$ , el cual es un valor real comprendido entre -1 y 1. Si  $r$  es 1 o -1, se dice que las variables están perfectamente correlacionadas positiva o negativamente, mientras que si es 0 no hay correlación. Esto quiere decir que cuando  $r$  es positivo, las variables tienen un comportamiento similar (ambas crecen o decrecen al mismo tiempo) y cuando  $r$  es negativo si una variable crece, la otra decrece [1].

El análisis de correlaciones, sobre todo las negativas, puede ser muy útil para establecer reglas de ítems correlacionados, como por ejemplo:

Un inspector de incendios que desee obtener información útil para la prevención de incendios probablemente este interesado en conocer correlaciones negativas entre el empleo de distintos grosores de protección del material eléctrico y la frecuencia de ocurrencia de incendios.

### **Las reglas de asociación:**

Son también una tarea descriptiva, muy similar a las correlaciones, que tiene como objetivo identificar relaciones no explícitas entre atributos categóricos\*. Pueden ser de muchas formas, aunque la formulación más común es del estilo “si el atributo X toma el valor d, entonces el atributo Y toma el valor b”. Las reglas de asociación no implican una relación causa-efecto, es decir, puede no existir una causa para que los datos estén asociados. Este tipo de tarea se utiliza frecuentemente en el análisis de la cesta de la compra, para identificar productos que son frecuentemente comprados juntos, esta información puede utilizarse para: ajustar los inventarios, para la organización física del almacén o en campañas publicitarias. Las reglas se evalúan usando dos patrones: precisión y soporte o cobertura [1].

Un caso especial de reglas de asociación, que recibe el nombre de reglas de asociación secuenciales, se usa para determinar patrones secuenciales en los datos. Estos patrones se basan en secuencias temporales de acciones y difieren de las reglas de asociación en que las relaciones entre los datos se basan en el tiempo. Es decir, consiste en buscar asociaciones de la forma: "si sucede el evento X en el instante de tiempo t entonces sucederá el evento Y en el instante  $t+n$ ".

Por ejemplo: una tienda de ventas de electrodomésticos y equipos de audio analiza las ventas que ha efectuado usando análisis secuencial y descubre que el 30 por ciento de los clientes que compraron un televisor hace seis meses compraron un DVD en los siguientes dos meses.

#### **2.2.2.1. MODELO PREDICTIVO**

El modelo predictivo usa datos con resultados conocidos para desarrollar un modelo que puede ser usado para predecir un valor explícitamente. La meta de estos modelos es predecir el valor de una variable, desde valores conocidos de otras variables. Los modelos de predicción permiten hacer predicciones de clasificación y de regresión. En la clasificación las variables que se predicen son del tipo categóricas, mientras que en la regresión se predicen variables

cuantitativas [1]. A continuación se muestran los algoritmos que se encuentran dentro del modelo predictivo.

Redes Neuronales
Árboles de Decisión ID3, C4.5, C5.0
Árboles de decisión CART
Otros árboles de decisión
Regresión lineal y logarítmica
Regresión logística
<b>Naive Bayes (Aplicado en SACC)</b>
Vecinos más próximos
Algoritmos genéticos y evolutivos
Máquinas de vectores soporte
CN2 rules (cobertura)
Análisis discriminante multivariante

**Tabla 2.1** Algoritmos para minería de datos predictivos[11]

### 2.2.2.2. MODELO DESCRIPTIVO

El modelo descriptivo describe patrones dentro de los datos existentes los cuales pueden ser usados para guiar decisiones. Es lo opuesto a hacer predicciones explícitas. Los modelos descriptivos permiten describir la manera como los datos se distribuyen o agrupan en el espacio [1]. Algunos de estos modelos son: Modelos de distribución de probabilidad de los datos (estimación de densidad), Particionamiento de un espacio p-dimensional en grupos (análisis de cluster y segmentación),

A continuación se muestran los algoritmos que se encuentran dentro del modelo predictivo.

Redes Neuronales
Otros árboles de decisión
Redes de Kohonen
Regresión lineal y logarítmica
Regresión logística
Kmeans
<b>A priori (Aplicado en SACC)</b>
Vecinos más próximos
Análisis Factorial y de comp. princ.
Twostep, Cobweb
Algoritmos genéticos y evolutivos
Máquinas de vectores soporte
CN2 rules (cobertura)

**Table 2.2** Algoritmos para minería de datos descriptivos [41]

## **CAPÍTULO 3**

### **3. MODELOS APLICADOS EN EL DESARROLLO (MINERÍA DE DATOS Y LENGUAJE DE MODELADO UNIFICADO)**

### **3.1. MODELOS DE MINERÍA DE DATOS APLICADOS**

En esta sección se explica cual fue el ciclo que seguimos para el desarrollo de los procesos de descubrimiento de conocimiento a través de minería de datos y además se detallan los modelos utilizados por la herramienta para análisis de comportamiento de consumo a la que hemos denominado SACC.

Cabe mencionar que dentro del desarrollo del tópico de minería de datos se revisaron las siguientes técnicas: Kmean, KNN, reglas de asociación Apriori, escalamiento multidimensional, introducción a los algoritmos genéticos, análisis de probabilidades Naive Bayes, análisis discriminante multivariante y árboles de decisión.

Cada uno de estos métodos o técnicas fue desarrollado para resolver un problema específico, esto dio como resultado un mejor entendimiento de las ventajas y limitantes en la implementación de los mismos. Las técnicas de minería que no fueron cubiertas dentro del desarrollo del tópico, son mencionadas en el capítulo anterior, sin embargo nuestros criterios para la definición de los algoritmos a

implementar se basan principalmente en la experiencia adquirida debido a su implementación.

Teniendo en cuenta que las dos problemáticas que el sistema debe resolver son: el análisis de la venta de productos y el análisis de perfiles del consumidor<sup>2</sup>, a continuación se analizan las posibles técnicas de minería de datos más óptimas que nos ayudarían llevar a cabo dicha tarea.

El análisis de venta de productos se lo puede enfocar hacia dos tendencias: analizar características similares entre los grupos de clientes que adquieren los productos, o analizar correlaciones al momento de la venta de los productos.

Si se escoge la primera opción, se debería desarrollar una técnica que identifique grupos de clientes basados en los productos que adquieren. Para realizar este análisis se necesitarían algoritmos enfocados al agrupamiento que se encuentran descritos en la tabla 2.2.

En cambio, si se analizan las relaciones entre los productos al momento de su venta, se puede determinar qué combinaciones de productos son

---

<sup>2</sup> En el capítulo 2 sección 2.1 se analizaron las razones para enfocar el sistema en la solución de estos dos problemas.

las más requeridas o también cuáles son las menos frecuentes. Ésta información es valiosa para los distribuidores de este tipo de productos ya que pueden planificar estrategias de distribución.

Dado que el análisis del perfil de la persona, con o sin diabetes, es el segundo enfoque a analizar; el definir grupos de consumidores no es importante, ya que estos pueden identificarse con la implementación del segundo algoritmo.

Entre los métodos recomendados para el análisis de reglas de asociación, según tabla 2.2, se encuentran los siguientes: la regla de asociación con información A priori, la regresión logística, árboles de decisión, algoritmos genéticos y CN2 rules.

Según el libro Applied Logistic Regresión [19], la regresión logística tiene como objetivo primordial definir la probabilidad de aparición de un suceso, habitualmente dicotómico, la presencia o no de diversos factores y el valor o nivel de los mismos. Es ampliamente utilizada en datos de investigación clínica y epidemiología. Esta técnica no aporta con información relevante para el análisis de reglas de asociación de productos.

A continuación se detalla una tabla comparativa entre los algoritmos de reglas de asociación y reglas de clasificación:

REGLAS DE ASOCIACIÓN	REGLAS DE CLASIFICACIÓN
Diferentes combinaciones de atributos dependientes e independientes.	Predice un atributo (clase) a partir de otros.
Búsqueda completa (todas las reglas encontradas).	Búsqueda heurística (se encuentra un subconjunto de reglas).

**Tabla 3.1** Cuadro Comparativo Reglas de Asociación y Clasificación

Los árboles usan heurística de evaluación sobre un atributo, están basados en splitting, y normalmente realizan “sobreajuste” seguido de “podado” [1].

Las reglas de clasificación utilizan una heurística de evaluación de condición (basados en pares atributo-valor), consisten en la aplicación de técnicas de covering y utilizan principalmente criterios de parada (y a veces “sobreajuste” y “podado”) [1].

Las reglas de asociación se basan en medidas de confianza y soporte, consideran cualquier conjunto de atributos con cualquier otro conjunto de atributos [1].

El algoritmo de reglas de asociación con información Apriori, ha sido ampliamente utilizado para la resolución del problema de la canasta de compras. Es por esta razón que utilizamos este algoritmo para analizar la venta de productos para diabéticos.

Tomando en cuenta que para el análisis de los perfiles de las personas, con o sin diabetes, se ha realizado un levantamiento de información en base a la encuesta electrónica, el hallar patrones dentro de este almacén de datos se lo puede realizar mediante algoritmos que agrupen conjuntos de individuos o técnicas basadas en la probabilística.

Entre los algoritmos de regresión que se encuentran descritos en la Tabla 2.1, el algoritmo más simple de implementar, no por esto el menos preciso, es el algoritmo de Naive Bayes; además, el estudio de la población basado en la definición de condiciones permite determinar las probabilidades de eventos en una población, según consultado a especialista<sup>3</sup>.

Esta herramienta permite definir un escenario de condiciones y luego obtener como resultado un cuadro de probabilidades para una variable en particular.

---

<sup>3</sup> Dr. Gonzalo García Baranderan. Jefe del Centro de Diabetes del Hospital Clínica Kennedy Guayaquil. Teléfono Consultorio: 042395538

### El ciclo virtuoso de minería de datos en SACC

El proceso de minería de datos puede verse como un ciclo o círculo virtuoso, basado en la continua mejoría de un proceso de negocios, impulsado por el descubrimiento de nuevo conocimiento y por la toma de acciones correspondientes a ese conocimiento.

De manera general los estados del ciclo virtuoso de minería de datos están representados en la figura 3.1 [1]:



**Figura 3.1** Ciclo virtuoso de minería de datos

A continuación se detalla como se desarrollaron los estados del ciclo virtuoso de la minería al momento de implementar SACC

**Identificar un problema.**

Se identificaron las dos problemáticas más relevantes con respecto al análisis de una enfermedad, en este caso, la diabetes.

1. Análisis de los factores de riesgo e incidencia de la enfermedad (tomando en cuenta la población urbana del Ecuador).
2. El análisis de los patrones de consumo de los productos para el tratamiento de la enfermedad (tomando en cuenta la información de la base de datos transaccional de una empresa distribuidora de estos productos).

**Usar minería para transformar los datos en información.**

Mediante la ejecución de algoritmos de minería sobre los datos en bruto, se extrae información comprensible y útil para el usuario. SACC implementa los algoritmos Naive Bayes y reglas de asociación A Priori al momento de ejecutar este estado.

**Actuar basándonos en la información.**

Se debe proveer al usuario las herramientas para que pueda llegar a un juicio de valor basándose en la información presentada por el sistema. SACC permite almacenar los análisis realizados para que esta tarea sea más factible.

**Medir los resultados.**

Aquí se determina si los resultados obtenidos son valiosos para los usuarios y si éstos cubren sus expectativas. Esta suele considerarse la última fase del proceso pero no es así, puesto que los datos obtenidos en esta fase pueden realimentar nuevamente los modelos para mejorar las estimaciones realizadas.

**3.1.1. MODELO PREDICTIVO**

El modelo predictivo se enfoca en generar modelos que descubren relaciones ocultas y complejas a partir de diversas operaciones. Las

tareas asociadas al análisis de tipo predictivo son: las redes neuronales, árboles de decisión, modelos de regresión, entre otros [1].

Dado que uno de los enfoques que nuestra herramienta debe considerar es el análisis de los factores de riesgo e incidencia de la enfermedad, se seleccionó la utilización del algoritmo de Naive Bayes para estudiar y predecir el comportamiento de variables que se encuentran dentro de una encuesta. La encuesta cubre aspectos tales como: aspectos demográficos (edad, sexo, ciudad, entre otros), aspectos relacionados a la salud (incidencia de enfermedades cardiovasculares, tabaquismo, tipo de diabetes que lo afecta) y por último aspectos relacionados al consumo de productos para el tratamiento de la enfermedad.

La encuesta es la fuente para la captación de información que representa nuestro universo de datos, cada vez que un usuario realiza una nueva encuesta, obtendremos un universo de datos más amplio y como consecuencia un análisis de datos más preciso.

SACC implementa el modelo predictivo mediante la aplicación de algoritmo de Naive Bayes para calcular las propensiones sobre cualquiera de las variables que existen en el universo de datos. Por

ejemplo se puede calcular cuál es la probabilidad para la clase sexo (masculino, femenino) dadas las siguientes premisas: padece diabetes tipo 2, reside en Guayaquil, su edad está en el rango de adulto mayor y actualmente trabaja. Este análisis le puede indicar al usuario hacia qué grupo objetivo debería enfocar una campaña publicitaria para personas con el perfil descrito en la premisas o condiciones, si la aplicación del algoritmo Naive Bayes señala que las personas que cumplen éste perfil tienen una probabilidad más alta de que pertenezcan al sexo masculino, sería mas conveniente realizar la campaña publicitaria enfocada hacia este grupo.

### **3.1.2. MODELO DESCRIPTIVO**

El modelo descriptivo se enfoca en presentar al usuario de manera intuitiva y parametrizable la información contenida en los diversos almacenes de información usando principalmente técnicas para el análisis multidimensional de las bases de datos [1].

SACC implementa el análisis descriptivo de la venta de productos mediante la ejecución del algoritmo A Priori de reglas de asociación que encuentra las relaciones que existen entre los productos para diabéticos de una empresa farmacéutica al momento de su

comercialización. Por ejemplo se presenta como resultado de la ejecución del algoritmo que: los reactivos para la medición personal del nivel de glucosa están altamente relacionados con el consumo de medicamentos de tipo inyectables. Esto permitiría realizar una combinación de estos productos (a un menor precio), para una comercialización más efectiva.

### **3.1.3. OPORTUNIDADES Y PROBLEMAS DE MODELOS APLICADOS**

Mediante la implementación de los algoritmos en el desarrollo del tópico, hemos podido apreciar las siguientes ventajas y desventajas.

#### **Algoritmo a Priori Reglas de Asociación**

##### Oportunidades

- Las reglas de asociación están siempre definidas sobre atributos binarios.
- La generación de los atributos binarios para la aplicación de las reglas de asociación no es complicada inclusive en grandes bases de datos.

- El usuario puede determinar la confianza y cobertura de las reglas de asociación.

### Problemas

- El problema es que tal algoritmo eventualmente puede dar información que no es relevante.
- El algoritmo tiende a consumir altos niveles de recursos cuando no se define buenos parámetros de confianza y cobertura.

### **Algoritmo Naive Bayesiano**

#### Oportunidades

- La sencillez en su formulación, y por tanto la de los cálculos a realizar.
- El tiempo de cómputo necesario para su aprendizaje y validación es lineal en el número de variables.
- Reduce enormemente el coste computacional pues solo tiene en cuenta la distribución de la clase.

- Cada caso de entrenamiento cambia la probabilidad estimada de que una hipótesis sea correcta.
- El conocimiento previo puede ser utilizado para determinar la probabilidad de una hipótesis.
- Pueden dar predicciones probabilísticas.
- Pueden clasificar nuevas instancias combinando probabilísticamente distintas hipótesis.
- A pesar de su simplicidad es comparable con clasificadores sofisticados como: Las Redes neuronales y los Árboles de decisión.

#### Problemas

- Cuando la asunción de independencia condicional entre las variables del problema no se cumple, la pérdida de precisión clasificatoria de este modelo puede ser grande. Este supuesto será violado cuando entre las variables condicionantes existan redundancias, o bien, existan variables altamente correlacionadas unas con otras.
- El algoritmo no es apto para el manejo de variables aleatorias continuas.

#### **3.1.4. ANÁLISIS DE LOS DATOS Y SU ALCANCE (VISTA MINABLE)**

La vista minable es un contexto de base de datos relacionales, cuyos datos se encuentran en una tabla con todos los atributos relevantes para efectuar el proceso de minería de datos. Dichos atributos pueden ser numéricos (valores enteros, decimales, entre otros) y nominales (categóricos, enumerados o discretos).

El sistema, obtiene una vista minable de los datos tanto para efectuar el análisis del perfil del diabético, como para el análisis de las tendencias de consumo de la población diabética.

La naturaleza de los datos y su procesamiento serán analizados a profundidad en la sección 4.3.

#### **3.1.5. DATOS DE ENTRENAMIENTO DEL MODELO**

SACC obtiene los datos de entrenamiento a partir de la encuesta publicada en la pagina Web, cuyas respuestas son procesadas y almacenadas en la base de datos. La selección de las variables presentes en la encuesta esta basada en: Sistemas de encuestas

similares<sup>4</sup> y recomendaciones del especialista<sup>5</sup>. Estos datos de entrenamiento son esenciales para la aplicación del modelo Naive Bayes. A continuación se detalla la clasificación de los datos a partir de su naturaleza:

#### **Variables cuantitativas**

- Edad.
- IMC.
- Estatura.
- Peso.

#### **Variables cualitativas**

- Ciudad de residencia.
- Tipo de diabetes.
- Tipo de medicamento que consume para su tratamiento.
- Nivel de actividad física del diabético.
- Nivel de consumo de productos como aspirina.

### **3.1.6. DATOS DE PRUEBA DEL MODELO**

---

<sup>4</sup> Asociación Americana de Diabetes. [2]

<sup>5</sup> Dr. Gonzalo García Baranderan. Jefe del Centro de Diabetes del Hospital Clínica Kennedy Guayaquil. Teléfono Consultorio: 042395538

SACC define como datos de prueba a los nuevos análisis realizados en el sistema. Por ejemplo: cuando el usuario utiliza el sistema para ingresar su perfil basado en la encuesta, estos resultados son procesados con respecto a los datos de entrenamiento (tabla de hechos). Este proceso es iterativo, ya que los datos de prueba pasan a ser datos de entrenamiento para futuros análisis.

En el anexo 2 se exponen los datos de la encuesta realizada para la generación de los datos del modelo.

### **3.2. MODELOS EN UML**

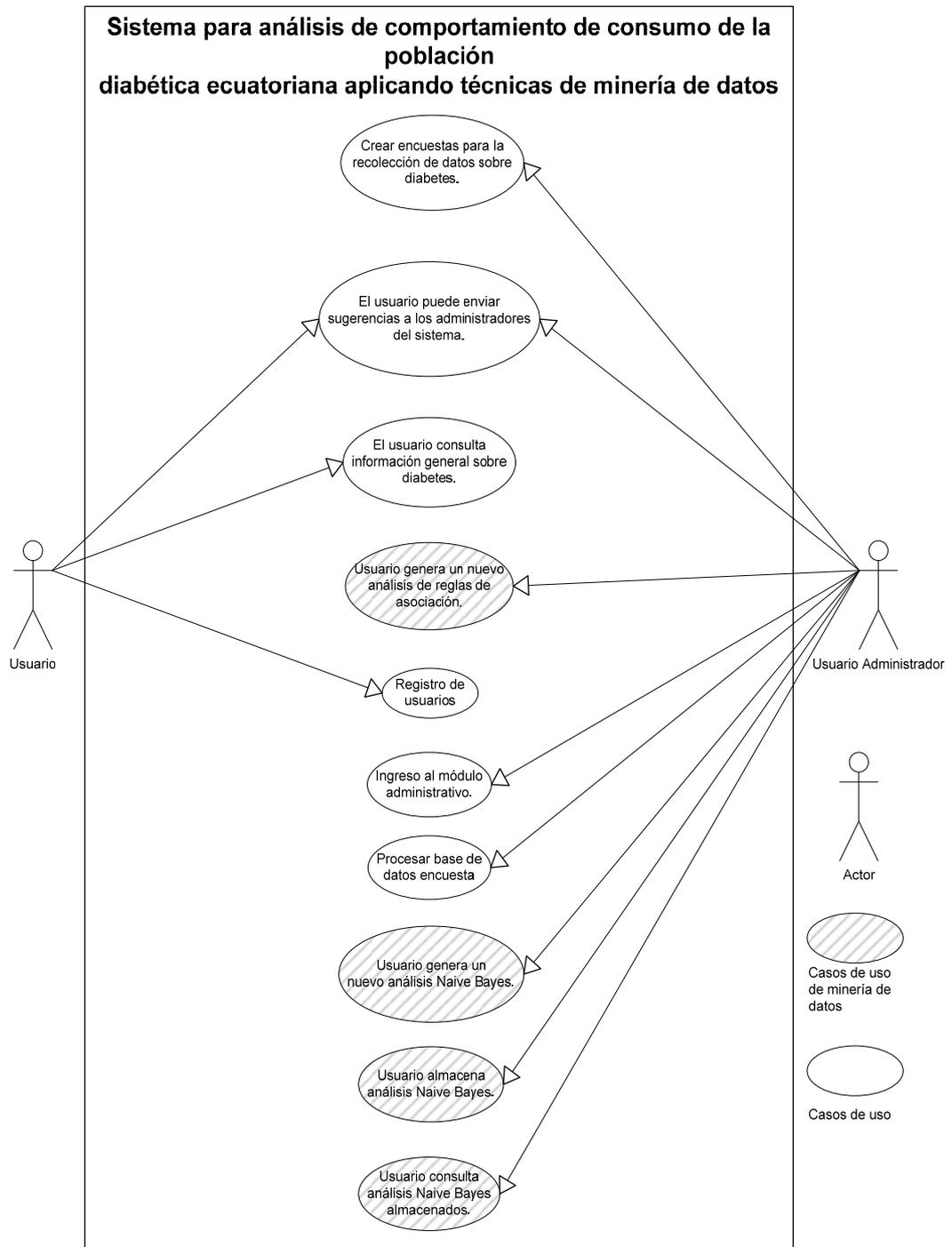
Ya descritos cuales son los modelos de minería utilizados en nuestra herramienta a continuación se detallan los modelos que definen el desarrollo de nuestra solución

El Lenguaje de Modelamiento Unificado (UML - Unified Modeling Language) es un lenguaje gráfico para visualizar, especificar y documentar cada una de las partes que comprende el desarrollo de software. UML entrega una forma de modelar cosas conceptuales como los procesos de negocio y funciones de sistema, además de cosas

concretas como lo son escribir clases en un lenguaje determinado, esquemas de base de datos y componentes de software reusables[20].

### 3.2.1. MODELADO DE OBJETOS

#### DIAGRAMA DE CASOS DE USO



## CASOS DE USO Y ESCENARIOS

<b>Caso de Uso 1:</b> Crear encuestas para la recolección de datos sobre diabetes.	
<b>Actor:</b> Usuario de tipo administrador en SACC.	
<b>Descripción:</b> Este caso de uso se refiere cuando el usuario administrador crea encuestas para analizar el consumo de productos, para el tratamiento de la diabetes y el perfil de las personas que los consumen.	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa su user y su password.	El sistema valida el user y password y da el ingreso al sistema activando las opciones que ese usuario tiene permitidas.
El usuario escoge la Opción de <i>Crear encuestas</i> .	El sistema presenta la ventana de crear encuesta.

### Escenarios

#### **Caso de uso 1: Crear encuestas para la recolección de datos sobre diabetes.**

##### **Escenario 1.1:**

**El administrador crea la encuesta exitosamente.**

##### **Supuestos:**

- El usuario tipo administrador escoge la opción crear encuesta del sistema.
- El usuario ingresa el nombre de la encuesta.
- El usuario selecciona el estado de la encuesta, los estados pueden ser: activa o inactiva.

##### **Salidas:**

- El sistema presenta el mensaje de encuesta creada.
- El sistema presenta la encuesta creada en la lista de encuestas almacenadas.

##### **Escenario 1.2:**

**El administrador no puede crear la encuesta exitosamente.**

##### **Supuestos:**

- El usuario tipo administrador escoge la opción crear encuesta del sistema.
- El usuario ingresa el nombre de la encuesta.

- El usuario no selecciona el estado de la encuesta, los cuales pueden ser: activa o inactiva.

**Salidas:**

- El sistema presenta el mensaje de error de encuesta no creada.

<b>Caso de Uso 2:</b> El usuario puede enviar sugerencias a los administradores del sistema.	
<b>Actor:</b> Usuario.	
<b>Descripción:</b> Este caso de uso se refiere cuando un usuario desea enviar una sugerencia a los administradores del sistema.	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa al sistema y selecciona la opción contáctenos.	El sistema le presenta la ventana de envío de correo al usuario del sistema.

## Escenarios

**Caso de uso 2:** El usuario puede enviar sugerencias a los administradores del sistema.

**Escenario 2.1:**

**El usuario envía sugerencia exitosamente.**

**Supuestos:**

- El usuario escoge la opción contáctenos del sistema.
- El usuario ingresa su dirección de correo electrónico.
- El usuario redacta las sugerencias que desea enviar.

**Salidas:**

- El sistema presenta el mensaje de mensaje enviado.

**Escenario 2.2:**

**El administrador no puede enviar las sugerencias exitosamente.**

**Supuestos:**

- El usuario escoge la opción contáctenos del sistema.
- El usuario ingresa su dirección de correo electrónico errónea.
- El usuario redacta las sugerencias que desea enviar.

**Salidas:**

- El sistema presenta el mensaje de error de mensaje no enviado.

<b>Caso de Uso 3:</b> El usuario consulta información general sobre diabetes.	
<b>Actor:</b> Usuario.	
<b>Descripción:</b> Este caso de uso le permite al usuario consultar información relacionada con la diabetes.	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa al sistema y selecciona la opción diabetes.	El sistema presenta la página que contiene información relacionada con la diabetes.

### Escenarios

**Caso de uso 3: El usuario consulta información general sobre diabetes.**

#### Escenario 3.1:

**El usuario consulta información de diabetes exitosamente.**

##### **Supuestos:**

- El usuario escoge la opción diabetes del sistema.
- Sitio Web disponible (en línea).

##### **Salidas:**

- El sistema presenta la página Web con la información de diabetes.

#### Escenario 3.2:

**El administrador no puede consultar información de diabetes exitosamente.**

##### **Supuestos:**

- El usuario escoge la opción diabetes del sistema.
- Sitio Web no disponible (no esta en línea).

##### **Salidas:**

- El sistema no presenta la página Web con la información de diabetes.
- El sistema presenta el mensaje de error de página Web con la información de diabetes no disponible.

<b>Caso de Uso 4:</b> Usuario genera un nuevo análisis de reglas de asociación.	
<b>Actor:</b> Usuario de tipo administrador en SACC.	
<b>Descripción:</b> Este caso de uso permite al usuario administrador extraer patrones de consumo de productos para el tratamiento de la diabetes.	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa al sistema y selecciona la opción diabetes.	El sistema presenta la página que contiene información relacionada con la diabetes.

**Caso de Uso 4: Usuario genera un nuevo análisis de reglas de asociación.**

**Escenario 4.1:**

**El administrador genera análisis de reglas de asociación exitoso.**

**Supuestos:**

- El usuario administrador decide escoger la opción reglas de asociación.
- El sistema muestra las transacciones de consumo de productos de diabetes.
- El usuario administrador realiza un filtro previo de la información que quiere analizar; dicho filtrado se lo puede realizar por: cliente, ciudad y fecha.
- El usuario administrador ingresa los valores de soporte y confianza.

**Salidas:**

- El sistema presenta las reglas de asociación generadas que cumplen con los parámetros de soporte y confianza.
- El sistema presenta la opción de consulta de nivel de inventario de los productos que intervienen en cada regla de asociación generada.
- El sistema presenta el mensaje de reglas de asociación generadas.

**Escenario 4.2:****El administrador no genera análisis de regla de asociación exitoso.****Supuestos:**

- El usuario tipo administrador decide escoger la opción reglas de asociación.
- El sistema muestra las transacciones de consumo de productos de diabetes.
- El usuario administrador realiza un filtro previo de la información que quiere analizar; dicho filtrado se lo puede realizar por: cliente, ciudad y fecha.
- El usuario ingresa valores de soporte y confianza que hace que el algoritmo diverja.

**Salidas:**

- El sistema no genera ninguna regla de asociación.
- El sistema presenta el mensaje de error reglas de asociación no generadas.

<b>Caso de Uso 5:</b> Registro de usuarios.	
<b>Actor:</b> Usuario.	
<b>Descripción:</b> El usuario se registra en el sitio Web.	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa al sistema y selecciona la opción de registro de usuario.	El sistema presenta la página con el formulario el registro.

**Caso de Uso 5: Registro de usuarios.****Escenario 5.1:****Registro de usuario exitoso.****Supuestos:**

- El usuario decide registrarse en el sitio Web.
- El sistema le presenta la página con el formulario de registro.
- El usuario llena el formulario.

**Salidas:**

- El sistema presenta el mensaje de usuario registrado satisfactoriamente.

**Escenario 5.2:****Registro de usuario no exitoso.****Supuestos:**

- El usuario decide registrarse en el sitio Web.
- El sistema le presenta la página con el formulario de registro.
- El usuario llena el formulario.
- El sistema pierde la conexión con la base de datos para realizar el registro.

**Salidas:**

- El sistema presenta el mensaje de error de Usuario no registrado error en la conexión de base de datos.

<b>Caso de Uso 6:</b> Ingreso al módulo administrativo.	
<b>Actor:</b> Usuario administrador.	
<b>Descripción:</b> El usuario administrador ingresa a las opciones del perfil administrador ya sea para realizar consultas o modificaciones al sitio Web.	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario administrador ingresa al sistema y selecciona la opción de modulo administrativo.	El sistema presenta la página la pagina Web de login.

**Caso de Uso 6: Ingreso al módulo administrativo del sistema.****Escenario 6.1:****Ingreso de usuario administrador exitoso.****Supuestos:**

- El usuario administrador selecciona la opción de modulo administrativo.
- El sistema le presenta la página de login.
- El usuario ingresa su usuario y contraseña correctamente.

**Salidas:**

- El sistema presenta el mensaje bienvenido al módulo administrativo.
- El sistema presenta la página Web del módulo administrativo del sitio.

**Escenario 6.2:****Ingreso de administrador no exitoso.****Supuestos:**

- El usuario administrador selecciona la opción de modulo administrativo.
- El sistema le presenta la página de login.
- El usuario ingresa su usuario y contraseña de manera incorrecta.

**Salidas:**

- El sistema presenta el mensaje de error de que el usuario administrativo debe ingresar usuario y contraseña válidos.

<b>Caso de Uso 7:</b> Procesar base de datos encuesta.	
<b>Actor:</b> Usuario de tipo administrador en SACC	
<b>Descripción:</b> Este caso de uso se refiere cuando el usuario procesa la base de datos de encuesta y actualiza la tabla de hechos para el análisis Naive Bayes	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa su user y su password.	El sistema valida el user y password y da el ingreso al sistema activando las opciones que ese usuario tiene permitidas.
El usuario escoge la Opción de <i>Procesar Base de Datos</i>	El sistema despliega la ventana de <i>Procesar base de datos</i> .
El usuario selecciona la opción Procesar	El sistema procede a recalcular la tabla de hechos

**Escenarios****Caso de uso 7: Procesar base de datos encuesta****Escenario 7.1:****El administrador procesa base de datos exitosamente.****Supuestos:**

- El usuario tipo administrador decide escoger la opción Procesar Base de Datos.
- La base de datos de encuesta contiene información.
- La base de datos de encuesta se encuentra en línea y operativa.

**Salidas:**

- El sistema procesa la base de datos de encuesta y actualiza la tabla de hechos.
- El sistema muestra un mensaje de proceso exitoso.

**Escenario 7.2:****El administrador no puede procesar base de datos exitosamente.**

**Supuestos:**

- El usuario tipo administrador decide escoger la opción Procesar Base de Datos.
- La base de datos de encuesta no contiene información.
- La base de datos de encuesta no se encuentra en línea y tampoco operativa.

**Salidas:**

- El sistema trata de procesar la base de datos de encuesta, pero un error no permite actualizar la tabla de hechos.
- El sistema muestra un mensaje de proceso fallido y muestra el error.

<b>Caso de Uso 8:</b> Usuario genera un nuevo análisis Naive Bayes.	
<b>Actor:</b> Usuario de tipo administrador en SACC	
<b>Descripción:</b> Este caso de uso se refiere cuando el usuario desea realizar un nuevo análisis Naive Bayes sobre los datos de la tabla de hechos	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ingresa su user y su password.	El sistema valida el user y password y da el ingreso al sistema activando las opciones que ese usuario tiene permitidas.
El usuario escoge la Opción de <i>Nuevo Análisis</i> .	El sistema despliega la ventana de <i>Análisis de Variables</i> .
El usuario consulta el estado de las variables y escoge la Opción de Seleccionar Variables.	El sistema despliega la ventana de Selección de Variables.
El usuario selecciona las variables por tres métodos: Selección personalizada, Selección de variables para análisis médico, Selección de variables para análisis de mercado. El usuario selecciona la opción ejecutar.	El sistema despliega la ventana para ejecución del algoritmo.
El usuario selecciona la variable a analizar y define los valores para el resto de variables que son de condición. El usuario procesa la información.	El sistema muestra los resultados y permite que el usuario cambie el valor de las variables.

## Escenarios

### Caso de uso 8: Usuario genera un nuevo análisis Naive Bayes

#### Escenario 8.1:

**El administrador genera análisis exitosamente.**

##### **Supuestos:**

- El usuario tipo administrador decide escoger la opción *Nuevo Análisis*.
- El sistema consulta tabla de hechos y muestra las variables disponibles para análisis (base de datos en línea y operativa).
- El usuario selecciona variables para análisis (opción personalizada, opción análisis de mercado, opción análisis médico).
- El usuario selecciona al menos tres variables para realizar el análisis.
- El usuario define cual es la variable objetivo.
- El usuario define cuáles son los valores para las variables condición.
- El usuario procesa la información.

##### **Salidas:**

- El sistema consulta la tabla de hechos y ejecuta el algoritmo.
- El sistema muestra los resultados de forma tabulada y grafica.

#### Escenario 8.2:

**El administrador no genera análisis exitosamente.**

##### **Supuestos:**

- El usuario tipo administrador decide escoger la opción *Nuevo Análisis*.
- El sistema consulta tabla de hechos y muestra las variables disponibles para análisis (base de datos no está en línea y tampoco operativa).
- El usuario selecciona variables para análisis (opción personalizada, opción análisis de mercado, opción análisis médico).
- El usuario no selecciona al menos tres variables para realizar el análisis.
- El usuario no define cual es la variable objetivo.
- El usuario no define cuáles son los valores para las variables condición.
- El usuario no procesa la información.

##### **Salidas:**

- El sistema no puede consultar la tabla de hechos y no ejecuta el algoritmo.
- El sistema muestra el error que no permite al usuario continuar con el análisis.

<b>Caso de Uso 9:</b> Usuario almacena análisis Naive Bayes.	
<b>Actor:</b> Usuario de tipo administrador en SACC	
<b>Descripción:</b> Este caso de uso se refiere cuando el usuario almacena un análisis Naive Bayes para futuras consultas	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario ha procesado las variables, ha llegado a un resultado destacable y escoge la opción <i>Guardar Análisis</i>	El sistema despliega la ventana Guardar Análisis.
El usuario ingresa el nombre con el que desea reconocer posteriormente el análisis.	El sistema almacena: cuáles fueron las variables que el usuario utilizó, cuáles fueron los valores para estas variables, y cuál fue el resultado del análisis; utilizando el nombre que el usuario ha definido.

## Escenarios

### Caso de uso 9: Usuario almacena análisis Naive Bayes

#### Escenario 9.1:

**El administrador almacena análisis exitosamente.**

#### **Supuestos:**

- El usuario tipo administrador decide escoger la opción *Guardar Análisis*.
- El sistema muestra la ventana para que el usuario ingrese un nombre descriptivo del análisis.
- El usuario ingresa nombre descriptivo
- El usuario pulsa el botón Guardar.

#### **Salidas:**

- El sistema consulta la tabla de análisis y almacena el análisis (variables y resultados) junto al nombre descriptivo
- El sistema muestra mensaje de proceso exitoso.

**Escenario 9.2:****El administrador no almacena análisis exitosamente.****Supuestos:**

- El usuario tipo administrador decide escoger la opción *Guardar Análisis*.
- El sistema muestra la ventana para que el usuario ingrese un nombre descriptivo del análisis.
- El usuario no ingresa nombre descriptivo
- El usuario no pulsa el botón Guardar.

**Salidas:**

- El sistema no permite almacenar el análisis.

<b>Caso de Uso 10:</b> Usuario consulta análisis Naive Bayes almacenados.	
<b>Actor:</b> Usuario de tipo administrador en SACC	
<b>Descripción:</b> Este caso de uso se refiere cuando el usuario consulta por uno de los análisis almacenados	
<b>Secuencia típica de Eventos:</b>	
<b>Actor</b>	<b>Sistema</b>
El usuario desea consultar un análisis almacenado y selecciona la opción <i>Análisis Almacenados</i> .	El sistema despliega la ventana <i>Análisis Almacenados</i> .
El usuario consulta desde una tabla todos los análisis almacenados y selecciona la opción <i>Mostrar Condiciones</i> para un análisis determinado.	El sistema muestra las variables que se utilizaron en dicho análisis y cuales fueron los valores asignados a las variables.
El usuario consulta desde una tabla todos los análisis almacenados y selecciona la opción <i>Mostrar Resultados</i> para un análisis determinado.	El sistema muestra los resultados para la variable objetivo de dicho análisis. Muestra valores y gráfico descriptivo.

## Escenarios

### Caso de uso 10: Usuario consulta análisis Naive Bayes almacenado.

#### Escenario 10.1:

##### **El administrador consulta análisis exitosamente.**

##### **Supuestos:**

- El usuario tipo administrador decide escoger la opción *Análisis Almacenados*.
- La base de datos de análisis está en línea y operativa
- El sistema muestra una tabla que describe los análisis almacenados y cual fue la variable analizada para cada uno.
- El usuario pulsa el botón Mostrar Condiciones.
- El usuario pulsa el botón Mostrar Resultados.

##### **Salidas:**

- El sistema consulta la tabla de análisis y muestra el análisis (variable, resultado y gráfico).
- El sistema muestra mensaje de proceso exitoso.

#### Escenario 10.2:

##### **El administrador no consulta análisis exitosamente.**

##### **Supuestos:**

- El usuario tipo administrador decide escoger la opción *Análisis Almacenados*.
- La base de datos de análisis no está en línea y tampoco operativa
- El sistema no muestra una tabla que describe los análisis almacenados y cual fue la variable analizada para cada uno.
- El usuario no pulsa el botón Mostrar Condiciones.
- El usuario no pulsa el botón Mostrar Resultados.

##### **Salidas:**

- El sistema no puede consultar la tabla de análisis y tampoco muestra el análisis (variable, resultado y gráfico).
- El sistema muestra mensaje que indica cuál fue el error.

## DIAGRAMA DE CLASES DEL SISTEMA

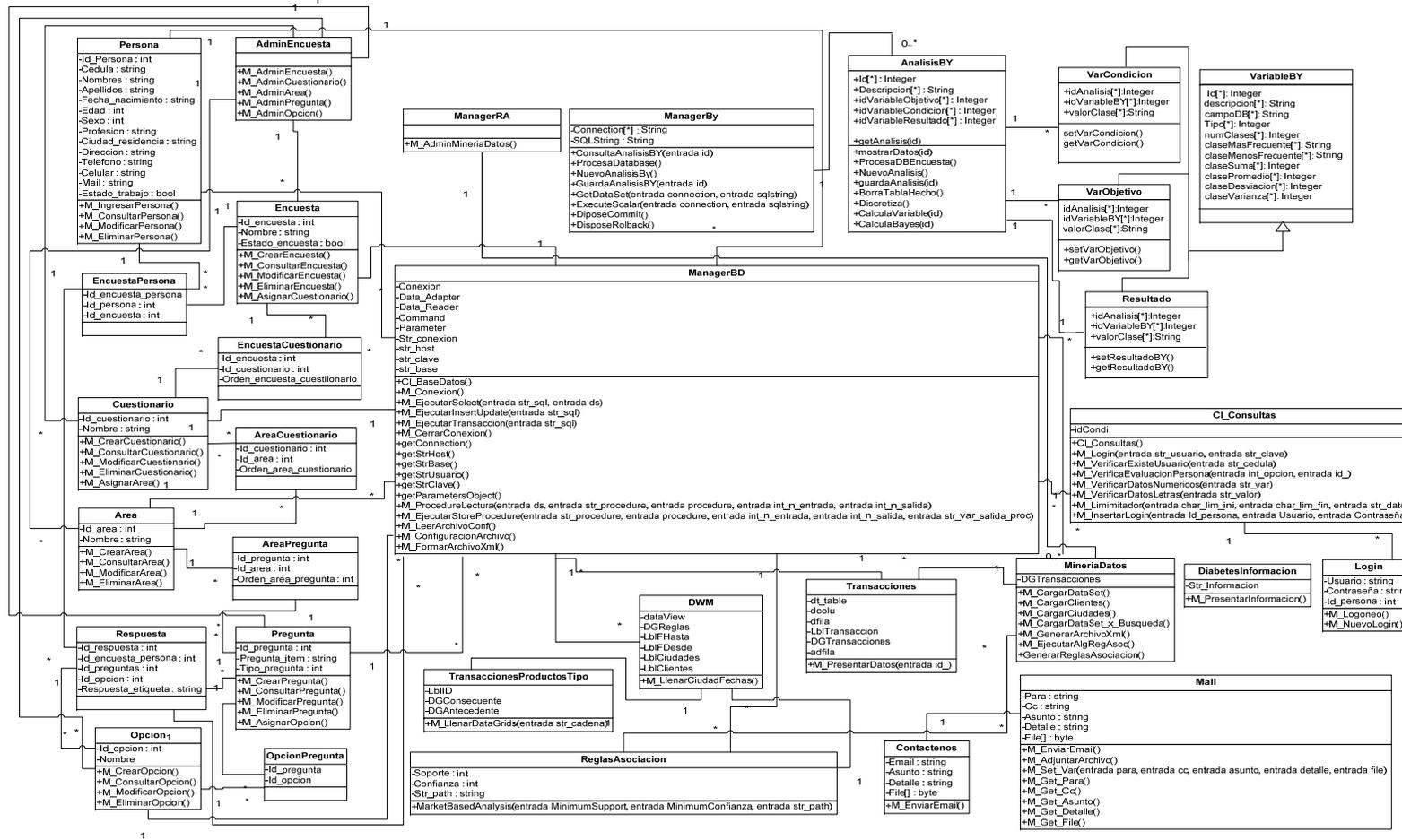
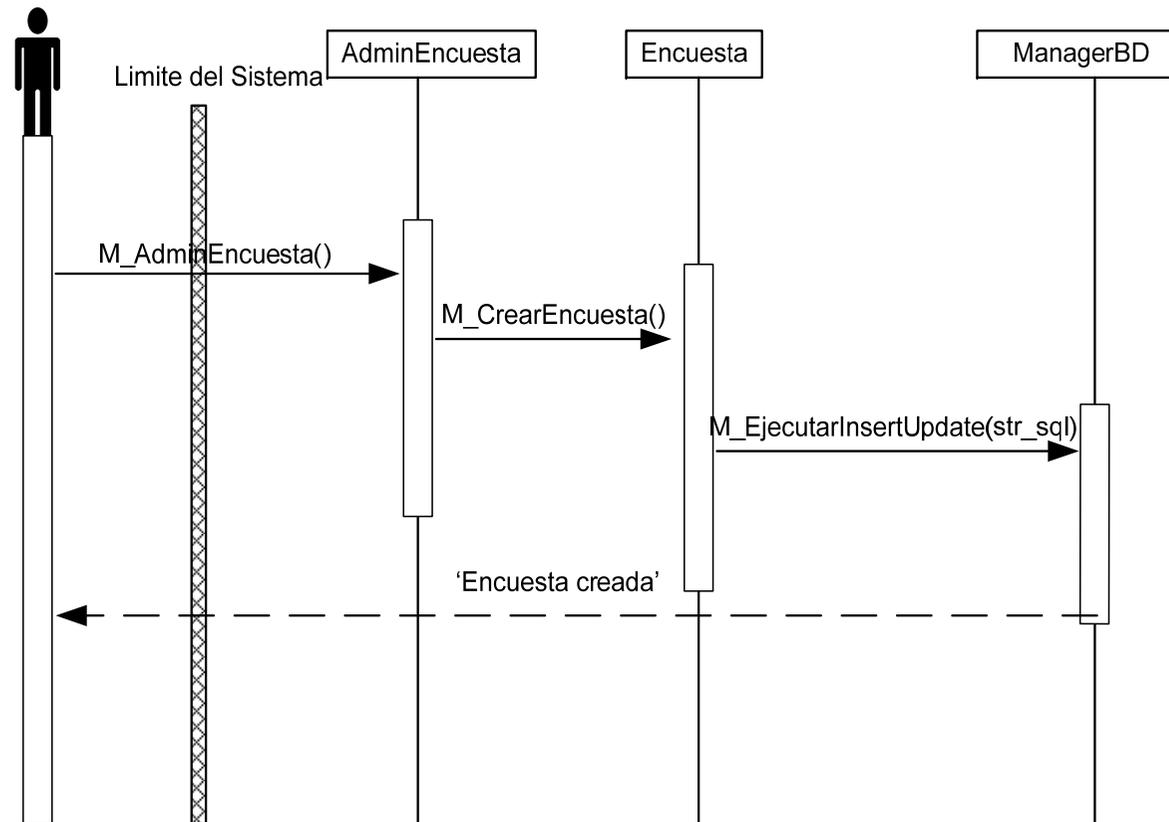


Figura 3.3 Diagrama de clases del sistema.

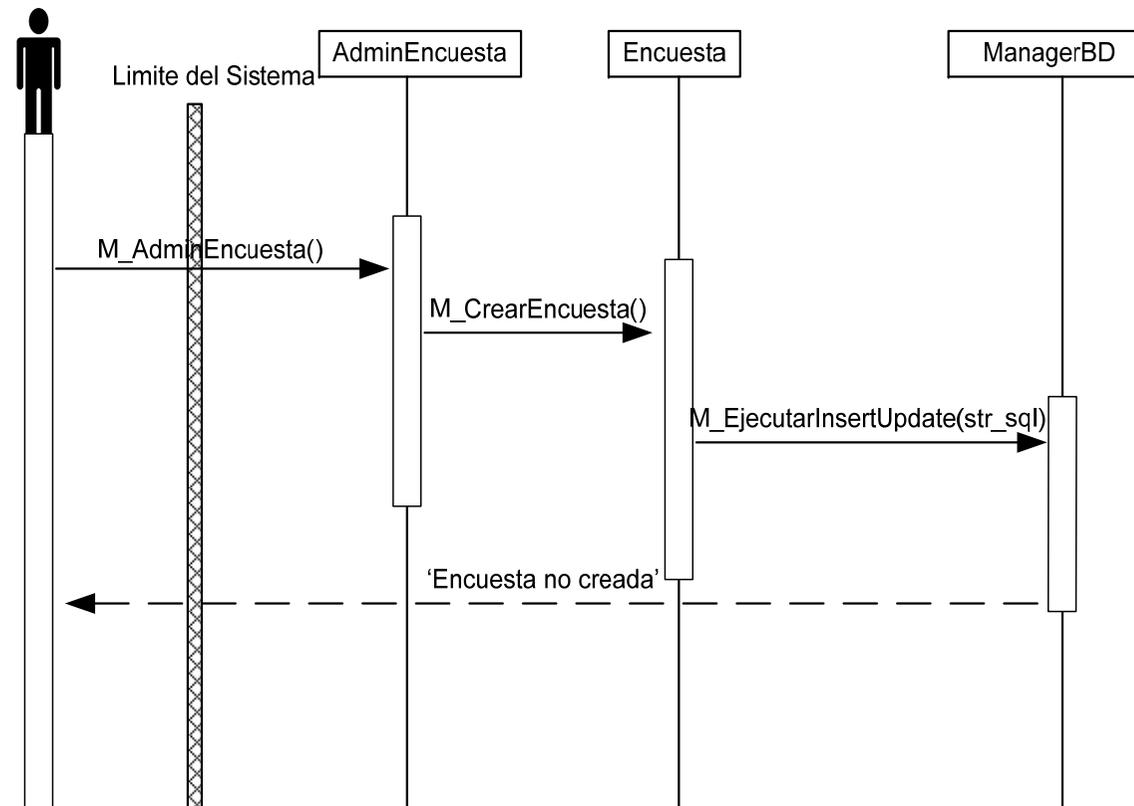
## DIAGRAMA DE INTERACCIÓN DE OBJETOS

1.1 El administrador crea la encuesta exitosamente.



**Figura 3.4** Diagrama de interacción de objeto el administrador crea la encuesta exitosamente.

### 1.2 El administrador no puede crear la encuesta exitosamente.



**Figura 3.5** Diagrama de interacción de objeto el administrador no puede crear la encuesta exitosamente.

## 2.1 El usuario envía sugerencia exitosamente.

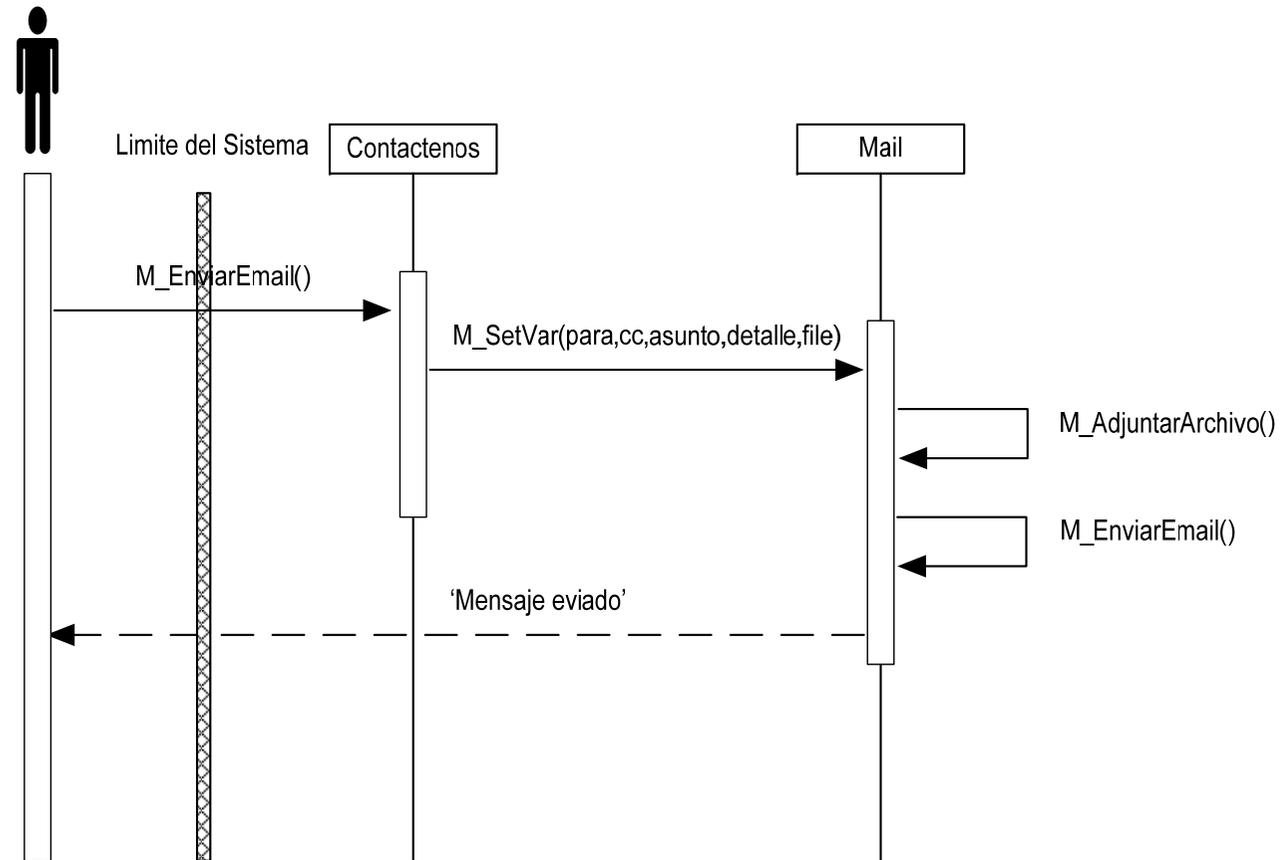
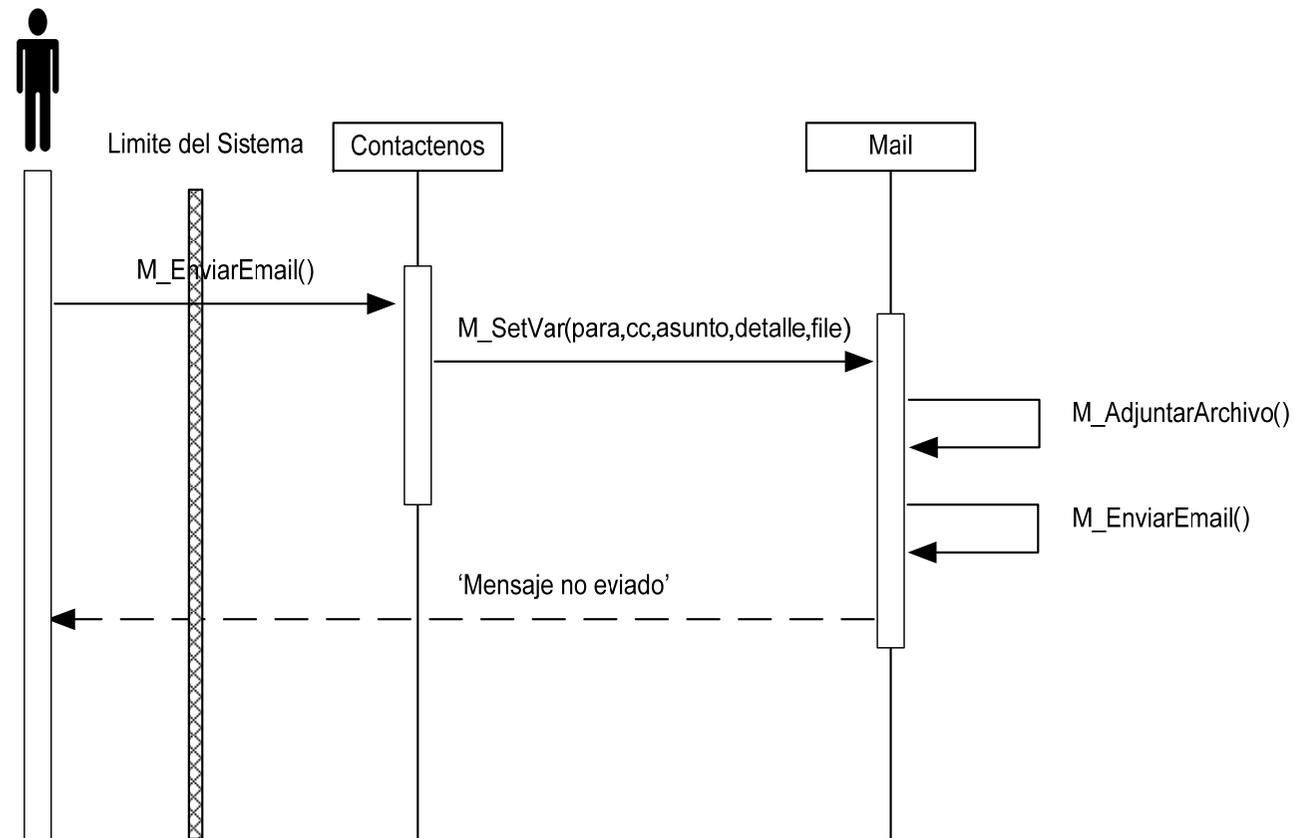


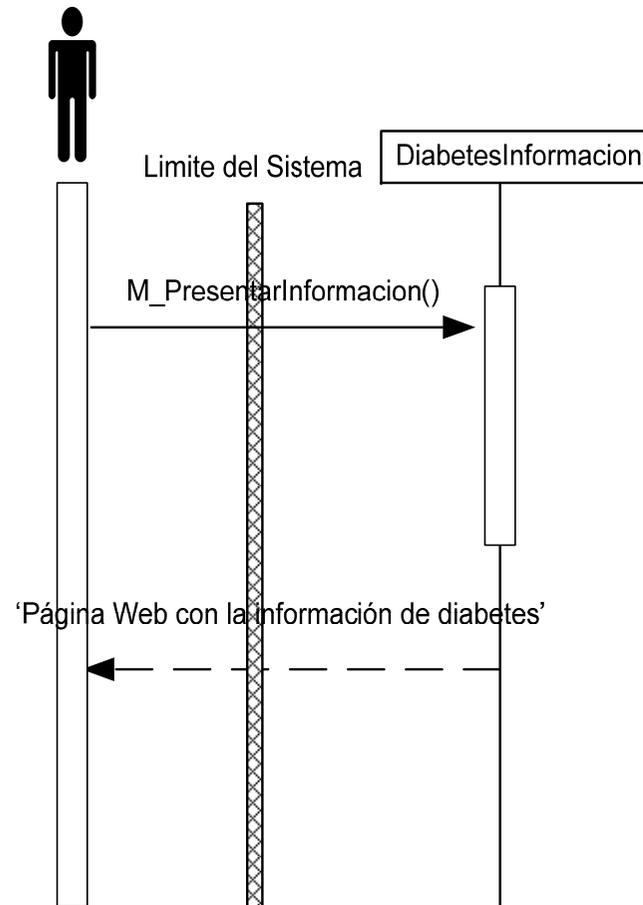
Figura 3.6 Diagrama de interacción de objeto el usuario envía sugerencia exitosamente.

## 2.2 El administrador no puede enviar las sugerencias exitosamente.



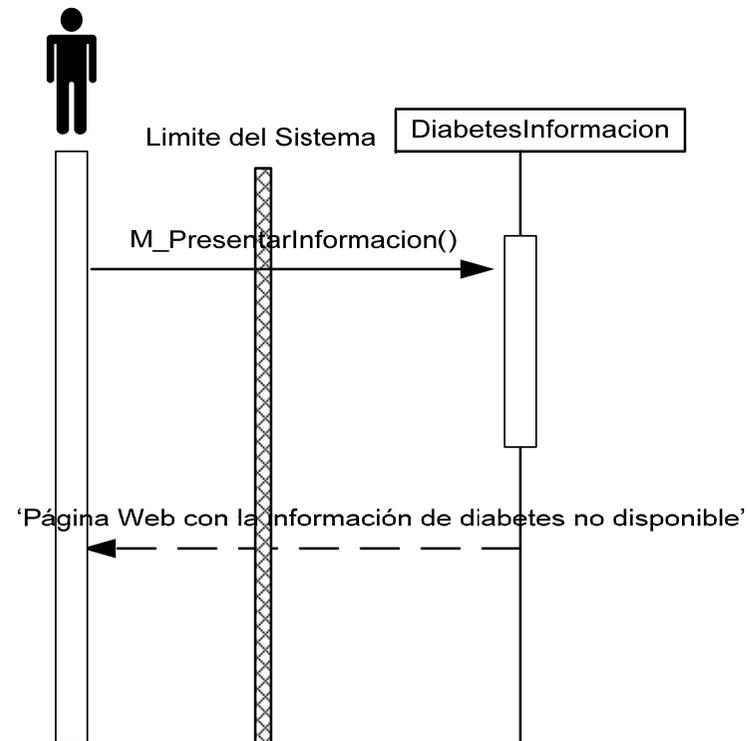
**Figura 3.7** Diagrama de interacción de objeto el administrador no puede enviar las sugerencias exitosamente.

3.1 El usuario consulta información de diabetes exitosamente.



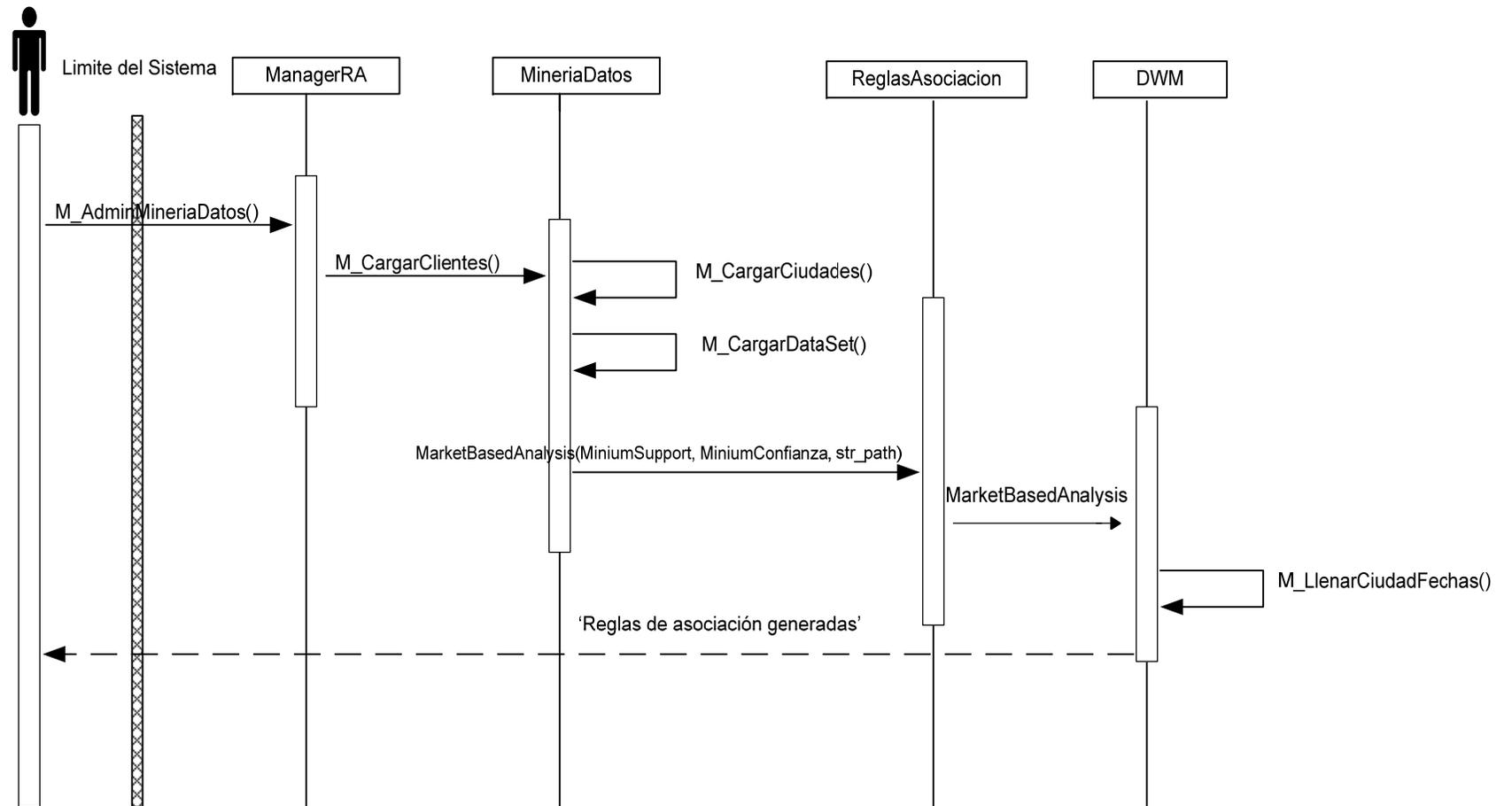
**Figura 3.8** Diagrama de interacción de objeto el usuario consulta información de diabetes exitosamente.

3.2 El usuario no puede consultar información de diabetes exitosamente.



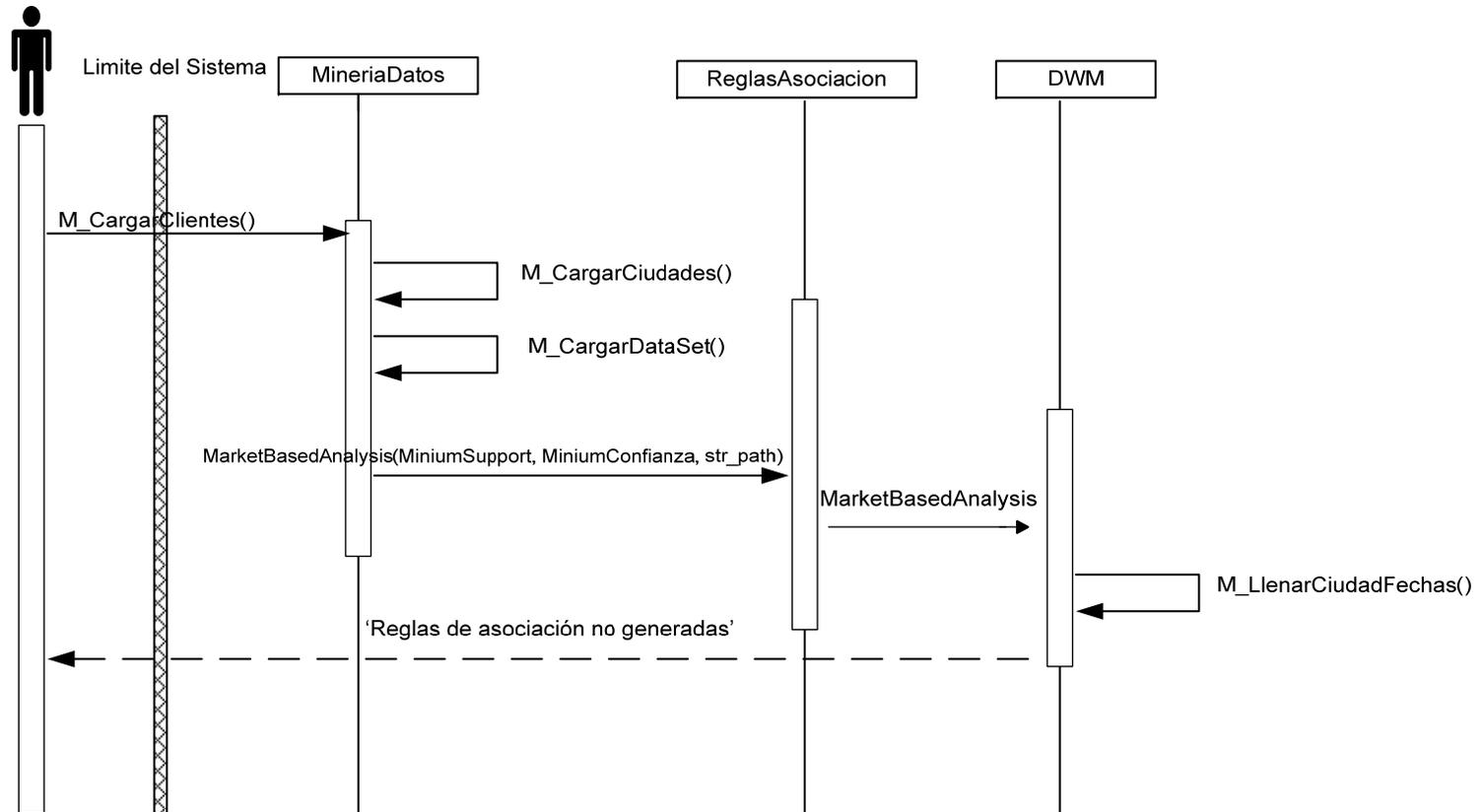
**Figura 3.9** Diagrama de interacción de objeto el usuario no puede consultar información de diabetes exitosamente.

#### 4.1 El administrador genera análisis de regla de asociación exitoso.



**Figura 3.10** Diagrama de interacción de objeto el administrador genera análisis de reglas de asociación exitoso.

#### 4.2 El administrador no genera análisis de reglas de asociación exitoso.



**Figura 3.11** Diagrama de interacción de objeto el administrador no genera análisis de reglas de asociación exitoso.

## 5.1 Registro de usuario exitoso.

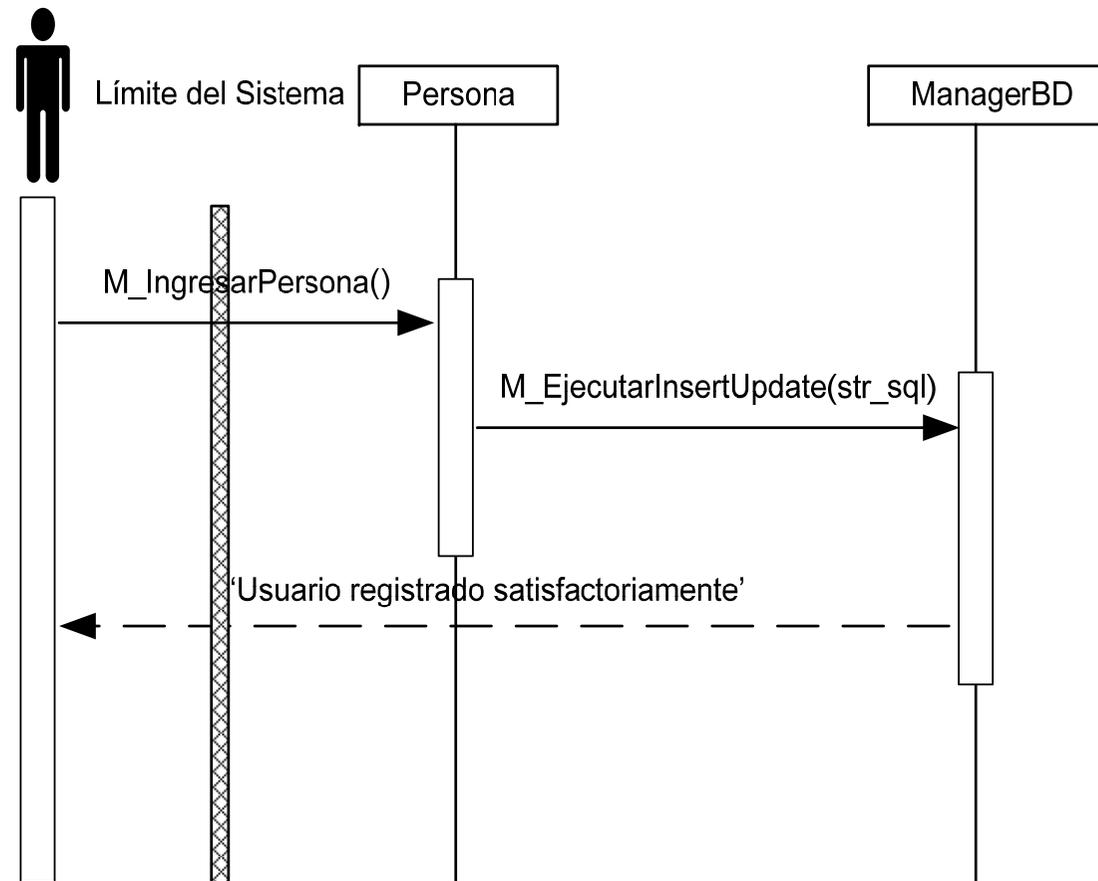
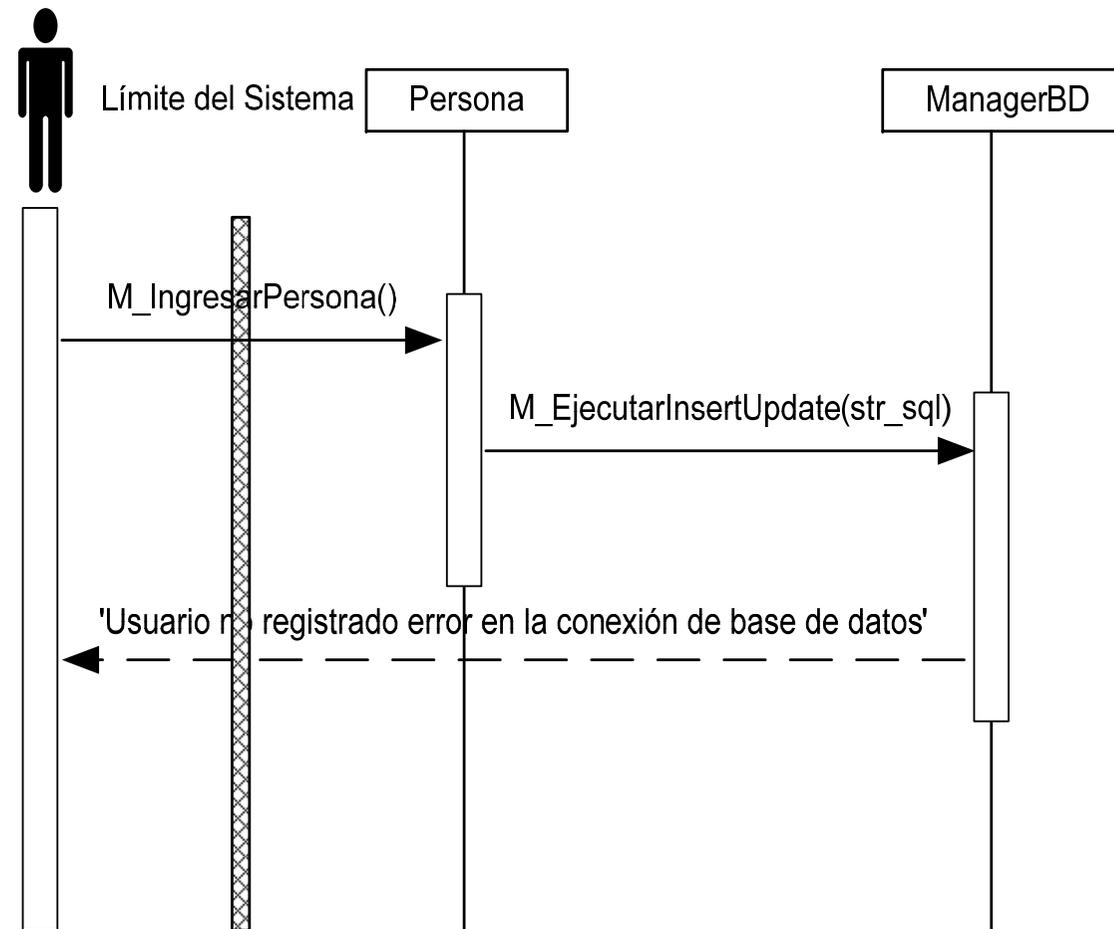


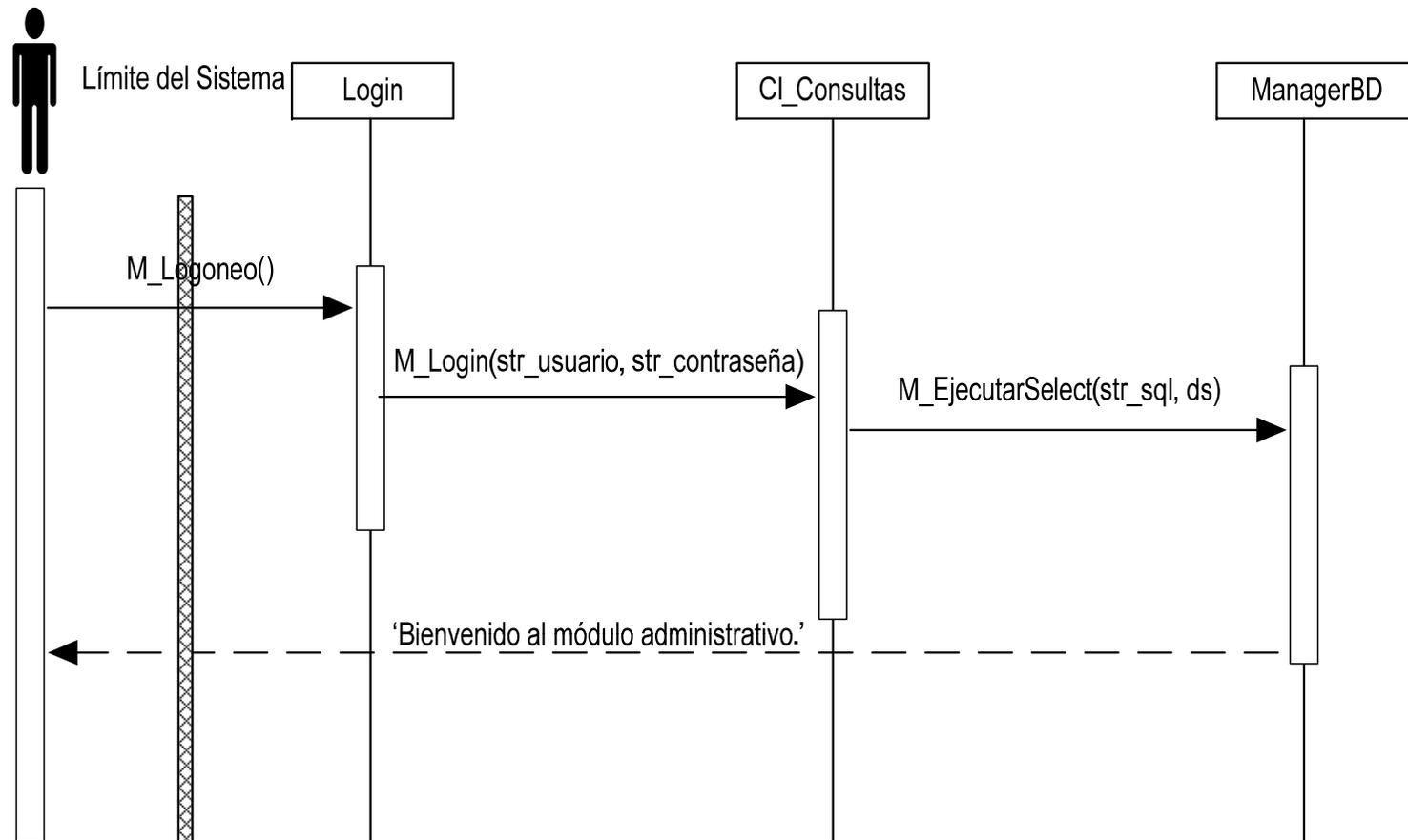
Figura 3.12 Diagrama de interacción de objeto registro de usuario exitoso.

## 5.2 Registro de usuario no exitoso.



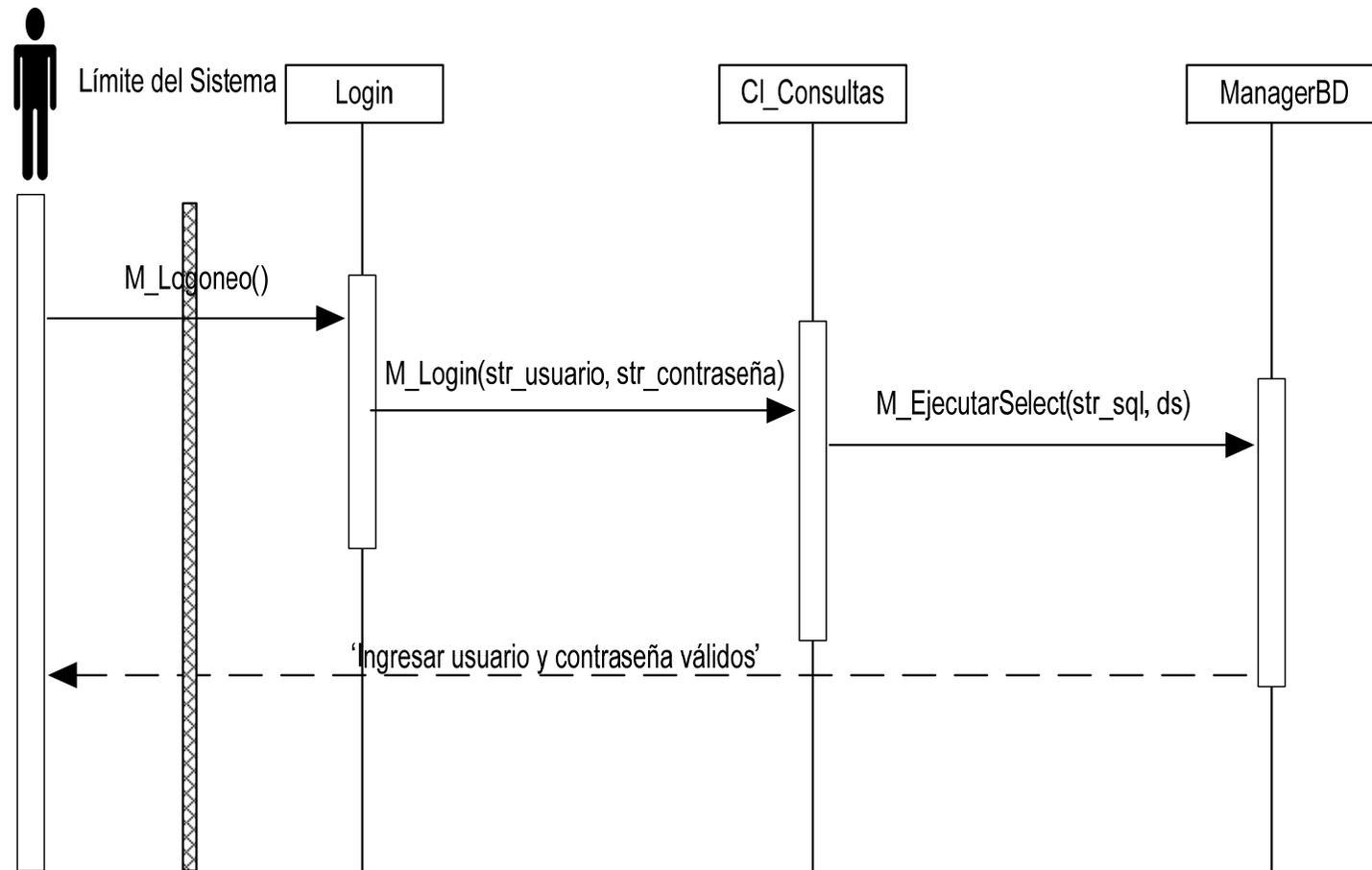
**Figura 3.13** Diagrama de interacción de objeto registro de usuario no exitoso.

## 6.1 Ingreso de usuario administrador exitoso.



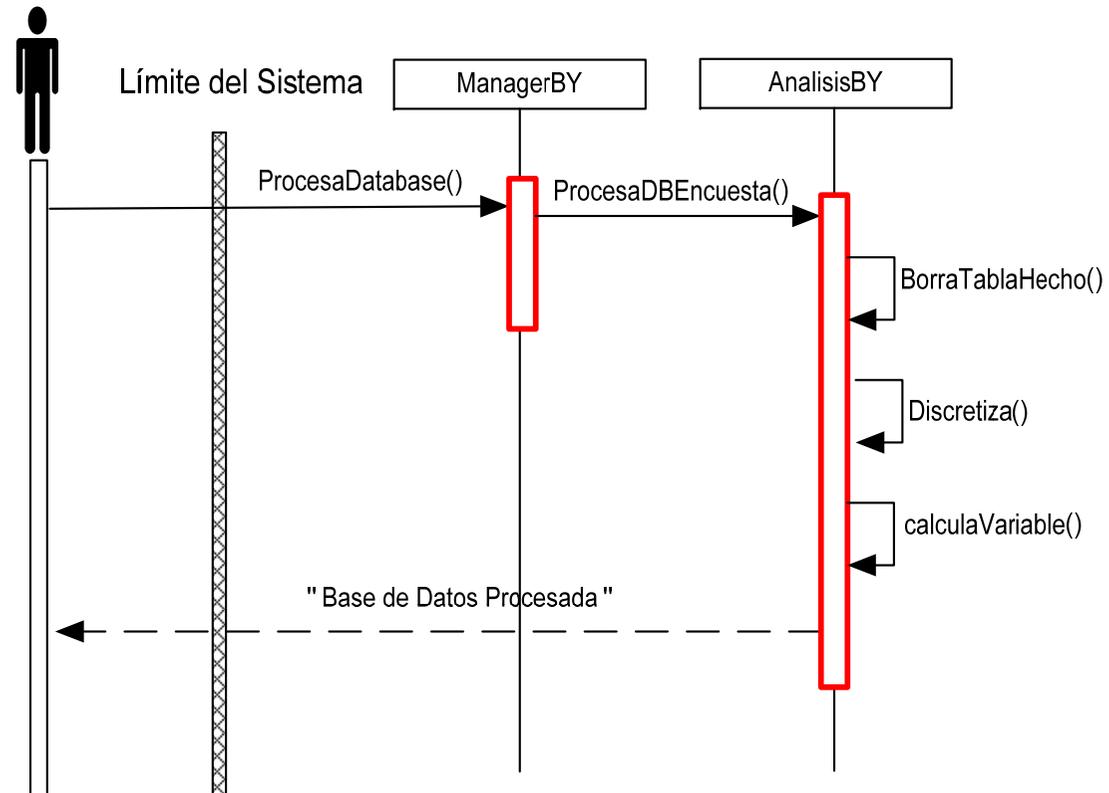
**Figura 3.14** Diagrama de interacción de objeto el ingreso de usuario administrador exitoso.

## 6.2 Ingreso de usuario administrador no exitoso.



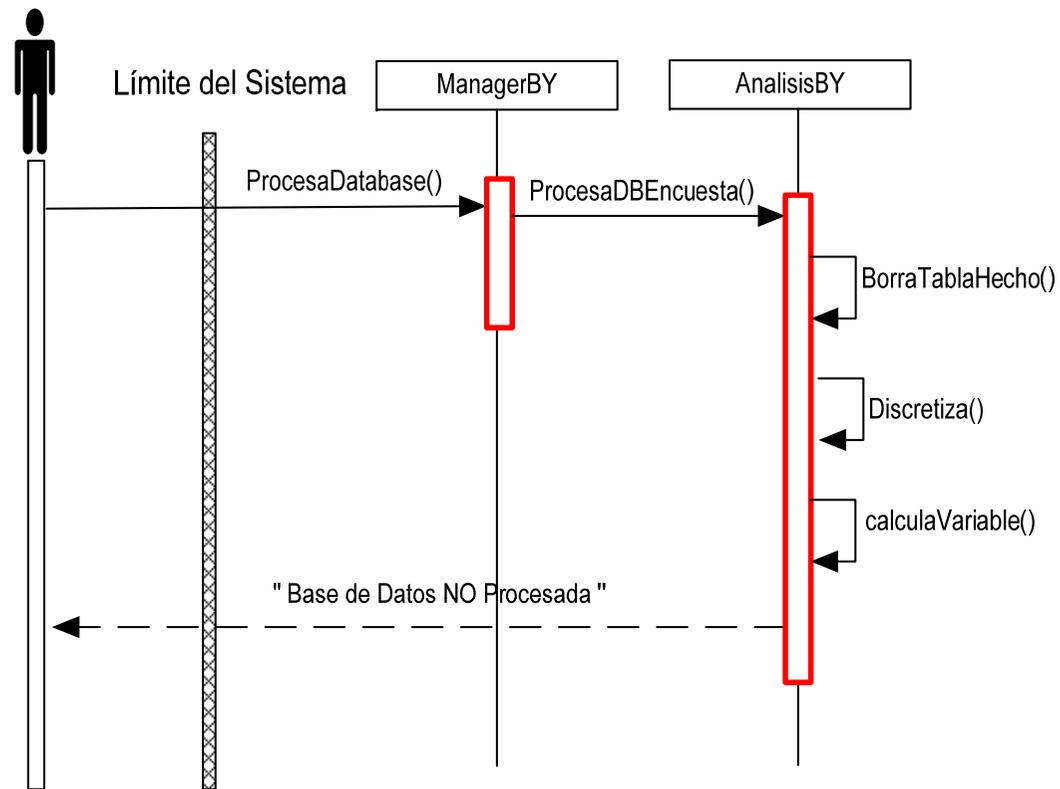
**Figura 3.15** Diagrama de interacción de objeto ingreso de usuario administrador no exitoso.

## 7.1 El administrador procesa base de datos exitosamente.



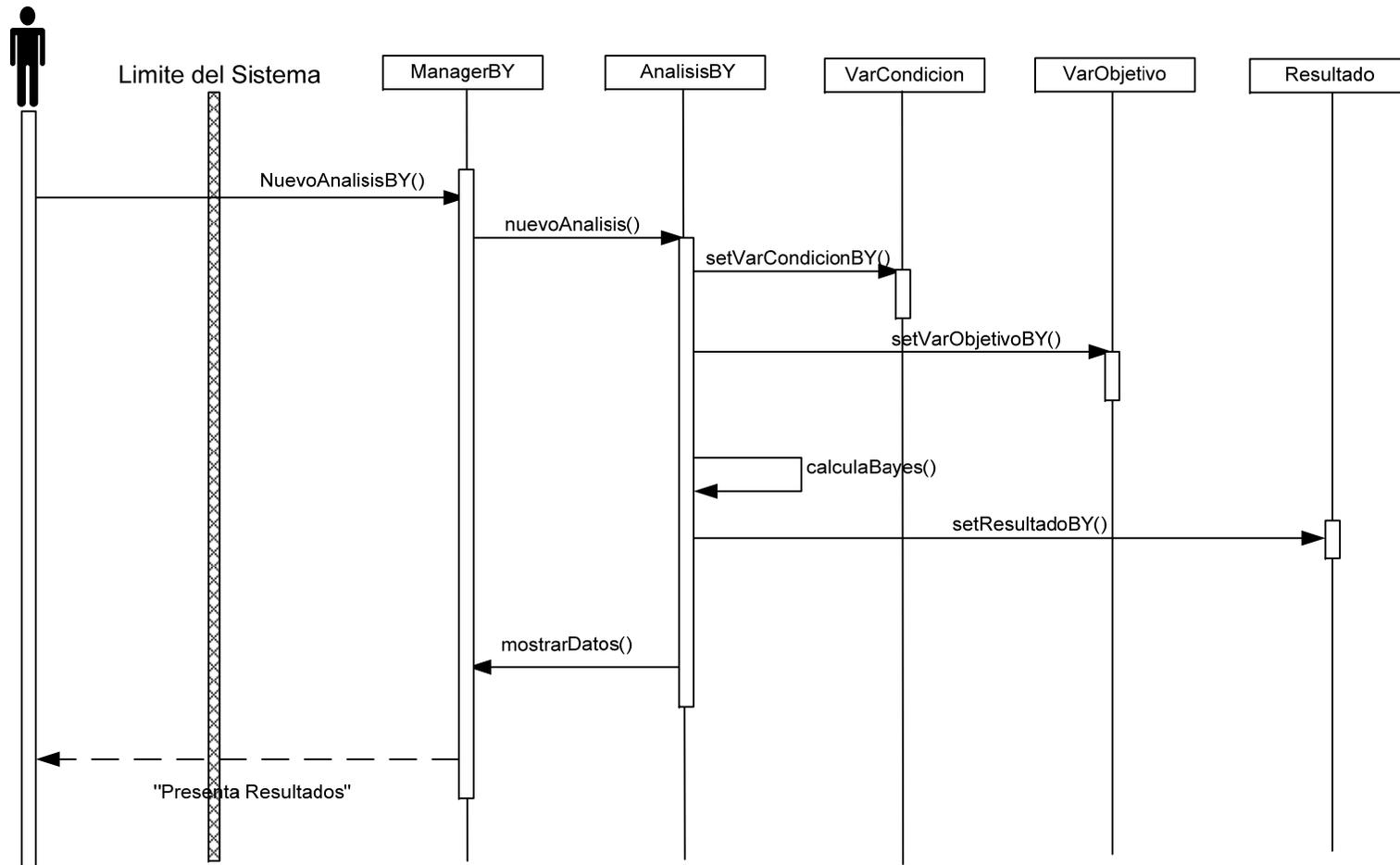
**Figura 3.16** Diagrama de interacción de objeto el administrador procesa base de datos exitosamente.

## 7.2 El administrador no puede procesar base de datos exitosamente.



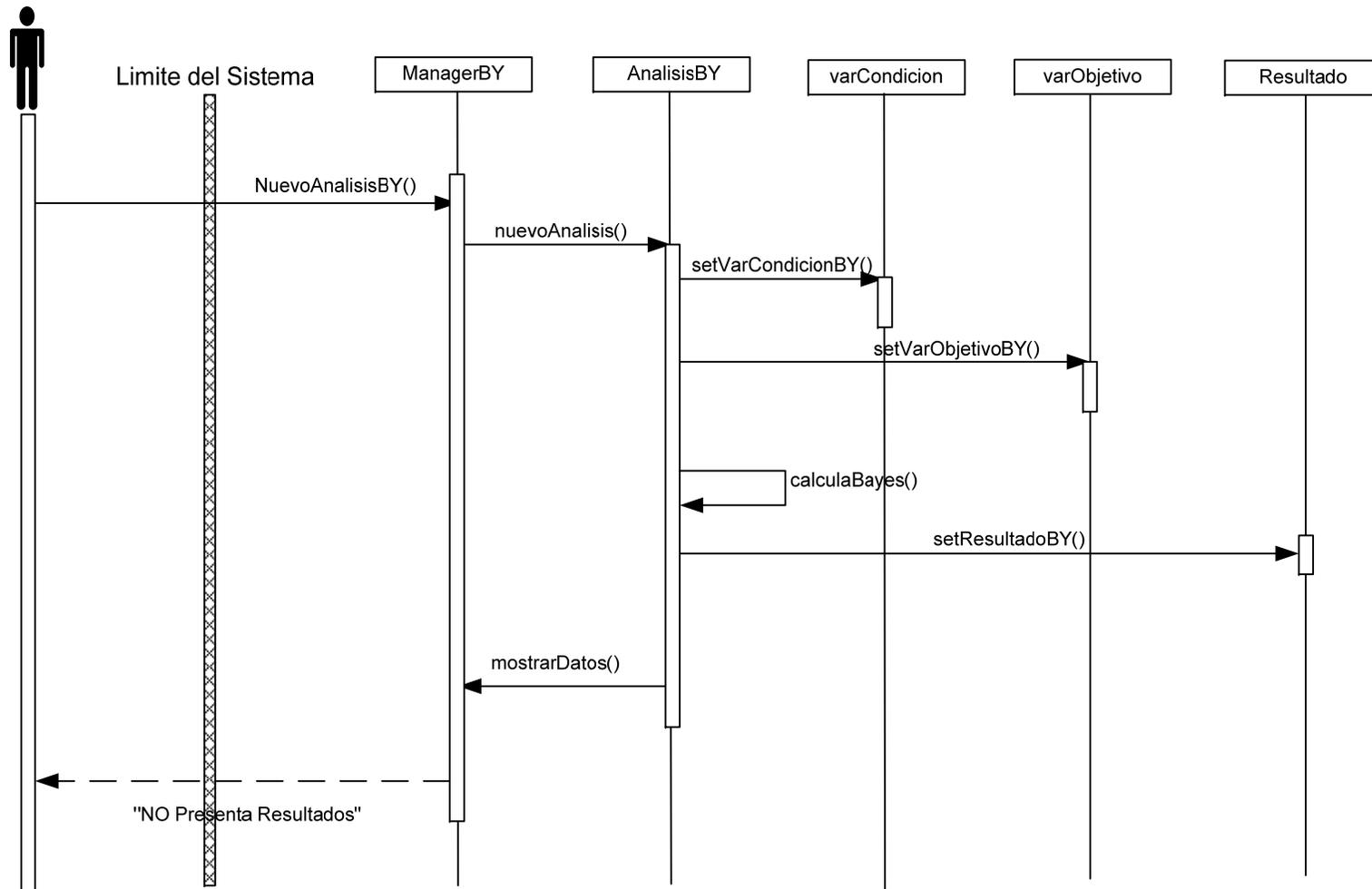
**Figura 3.17** Diagrama de interacción de objeto el administrador no puede procesar base de datos.

## 8.1 El administrador genera análisis exitosamente.



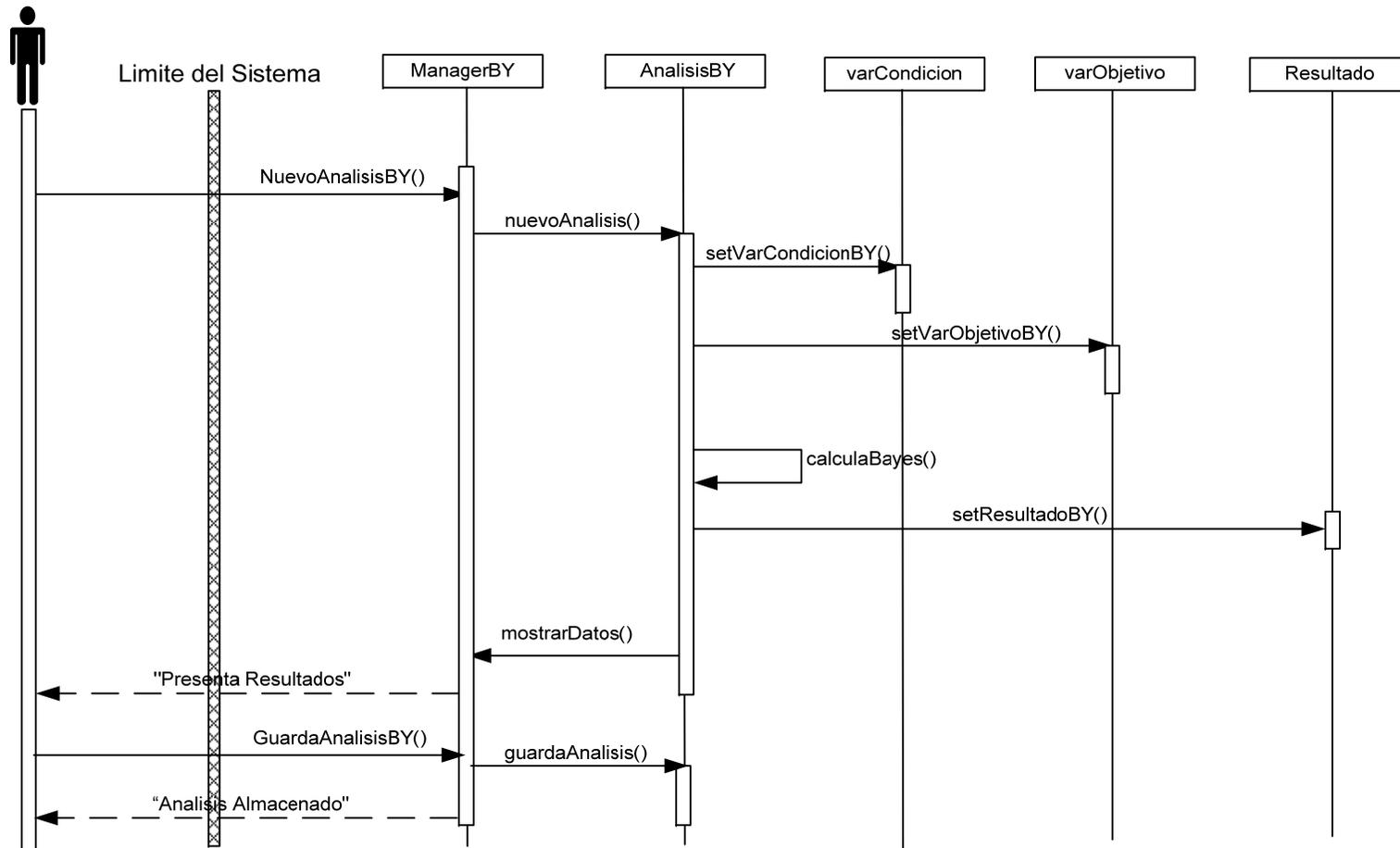
**Figura 3.18** Diagrama de interacción de objeto el administrador genera análisis exitosamente.

## 8.2 El administrador no genera análisis exitosamente.



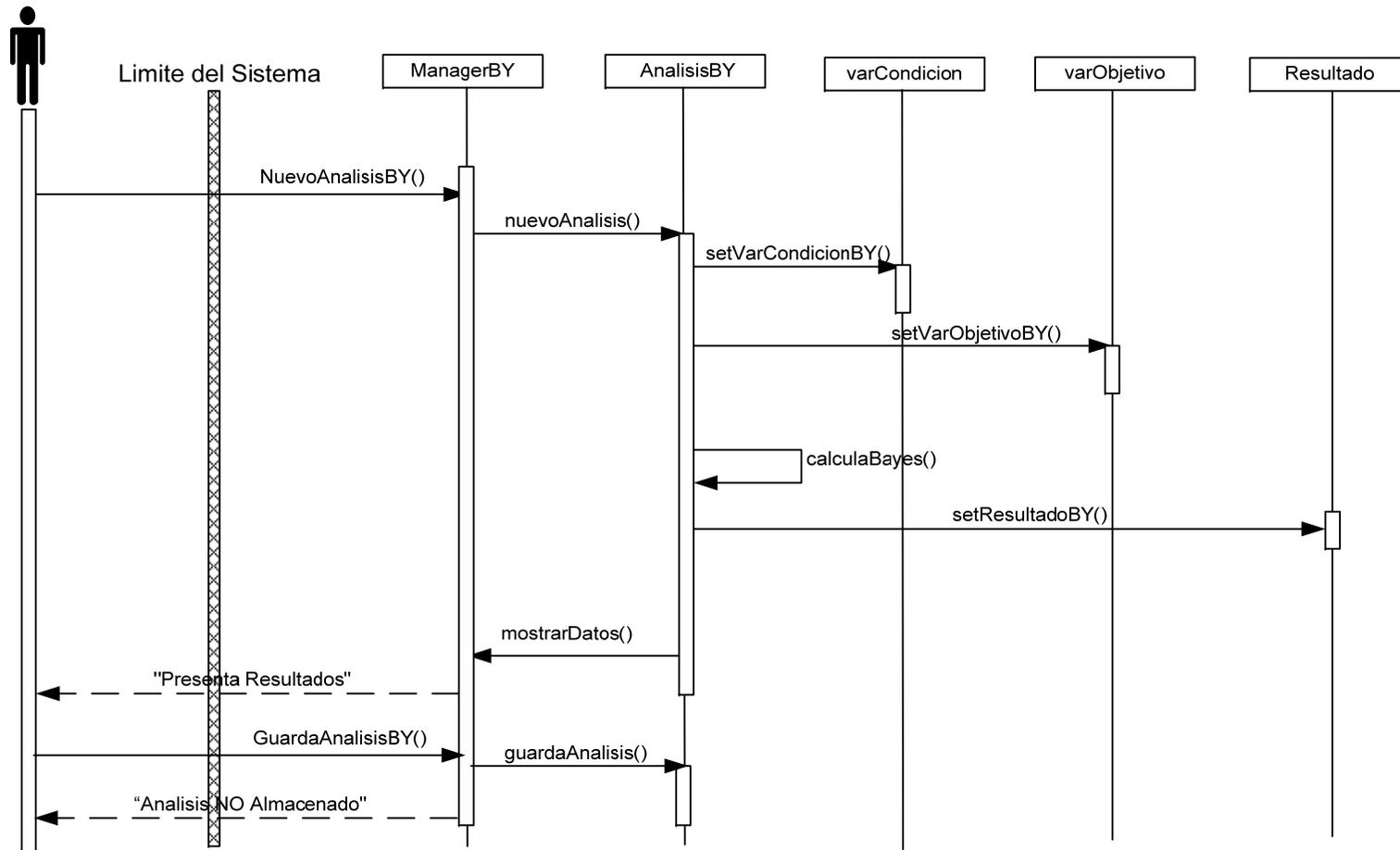
**Figura 3.19** Diagrama de interacción de objeto el administrador no genera análisis exitosamente.

## 9.1 El administrador almacena análisis exitosamente.



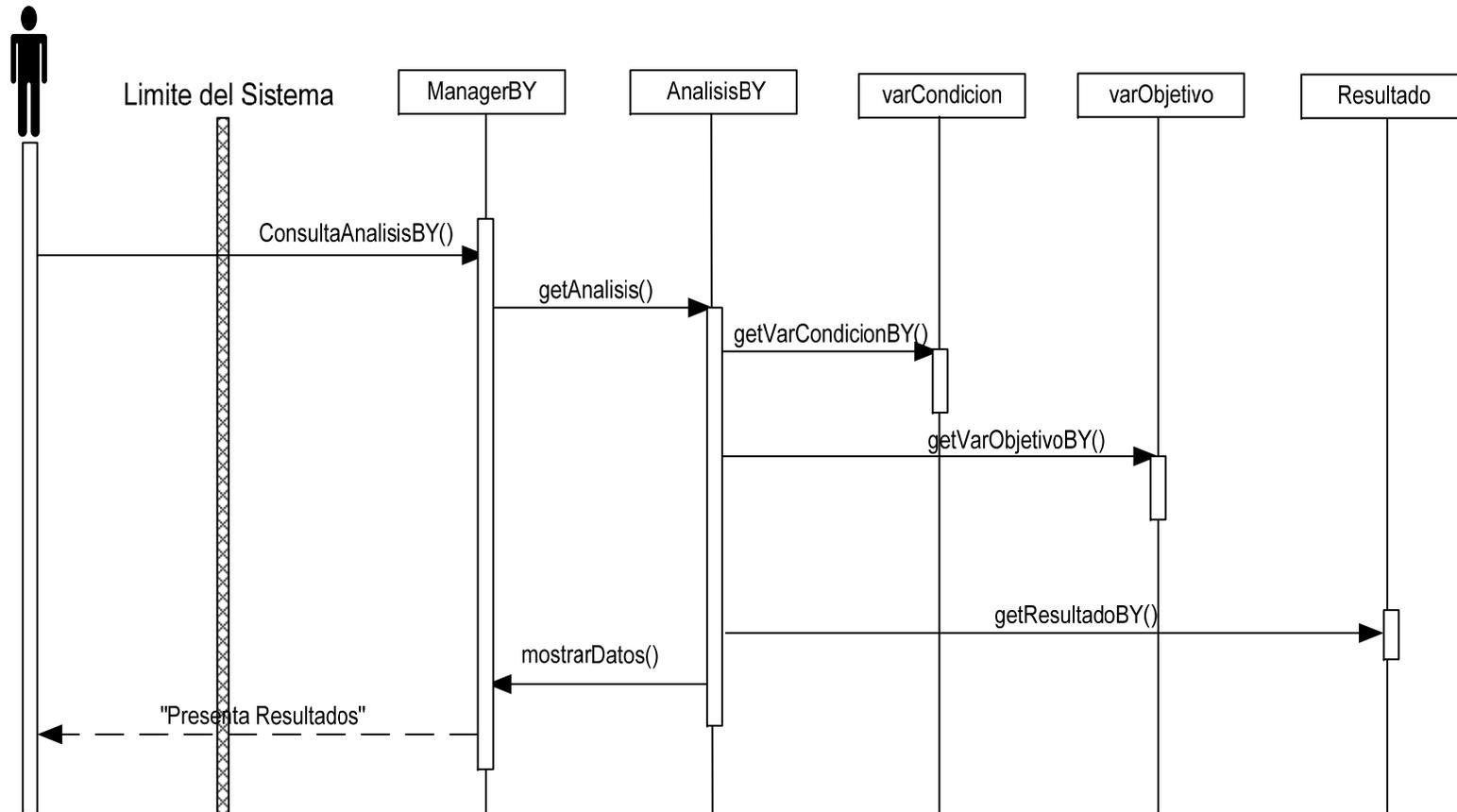
**Figura 3.20** Diagrama de interacción de objeto el administrador almacena análisis exitosamente.

## 9.2 El administrador no almacena análisis exitosamente.



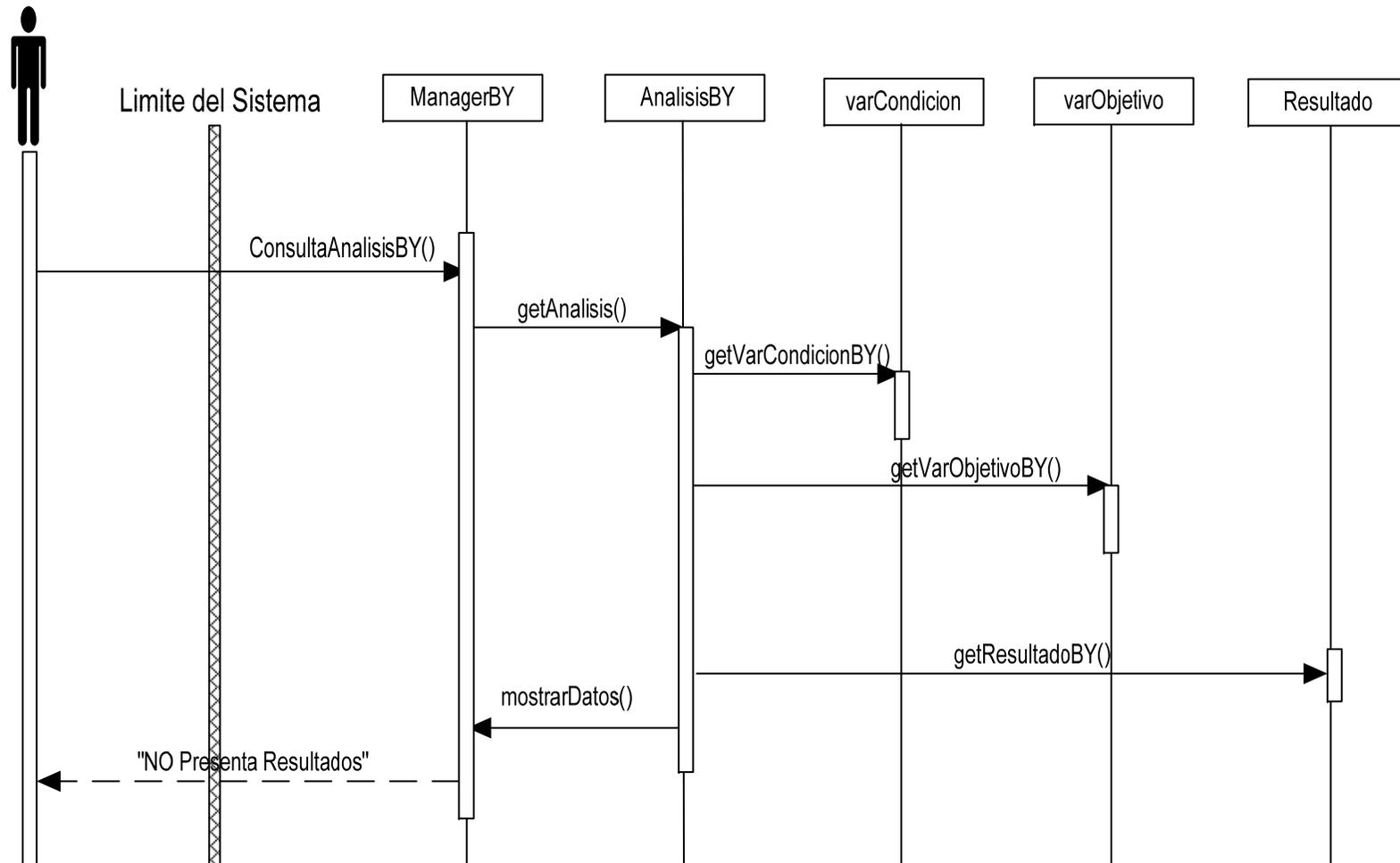
**Figura 3.21** Diagrama de interacción de objeto el administrador no almacena análisis exitosamente.

## 10.1 El administrador consulta análisis exitosamente.



**Figura 3.22** Diagrama de interacción de objeto el administrador consulta análisis exitosamente.

## 10.2 El administrador no consulta análisis exitosamente.



**Figura 3.23** Diagrama de interacción de objeto el administrador no consulta análisis exitosamente.

### 3.2.2. MODELADO DINÁMICO

#### Diagrama de estado. Modelo Naive Bayes

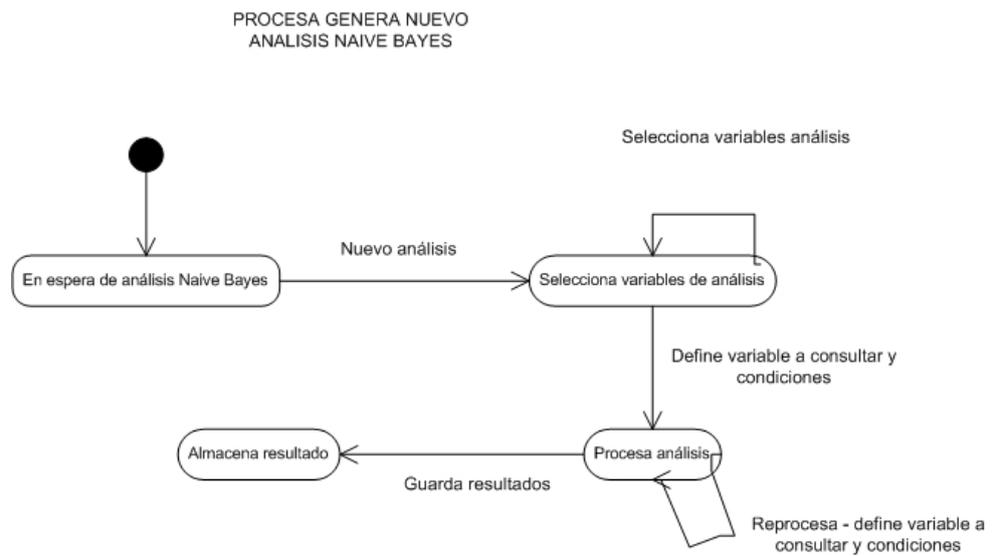


Figura 3.24 Diagrama de estado: Naive Bayes.

#### Diagrama de estado. Modelo reglas de asociación

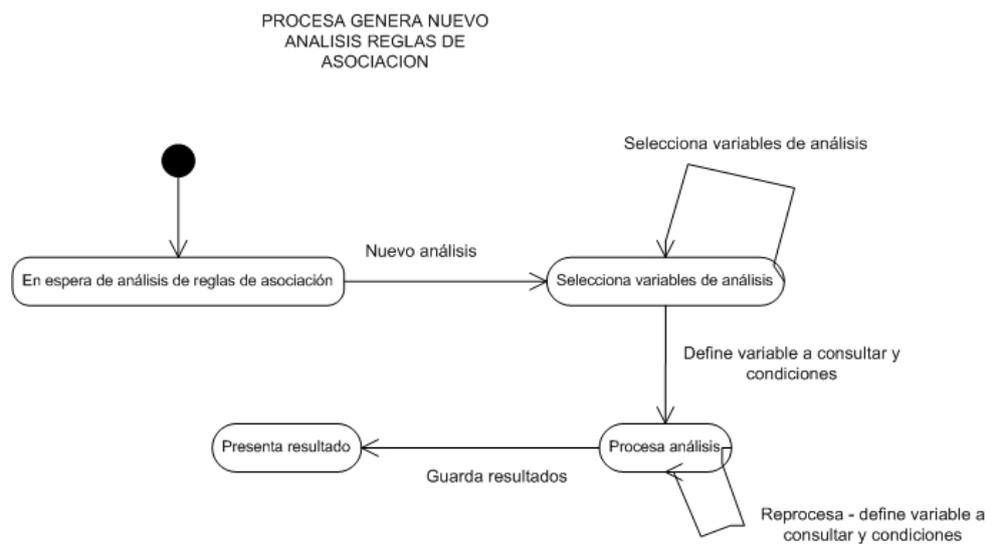


Figura 3.25 Diagrama de estado: Reglas de asociación.

### 3.2.3. MODELADO FUNCIONAL

#### Diagrama de flujo de datos modelo Naive Bayes

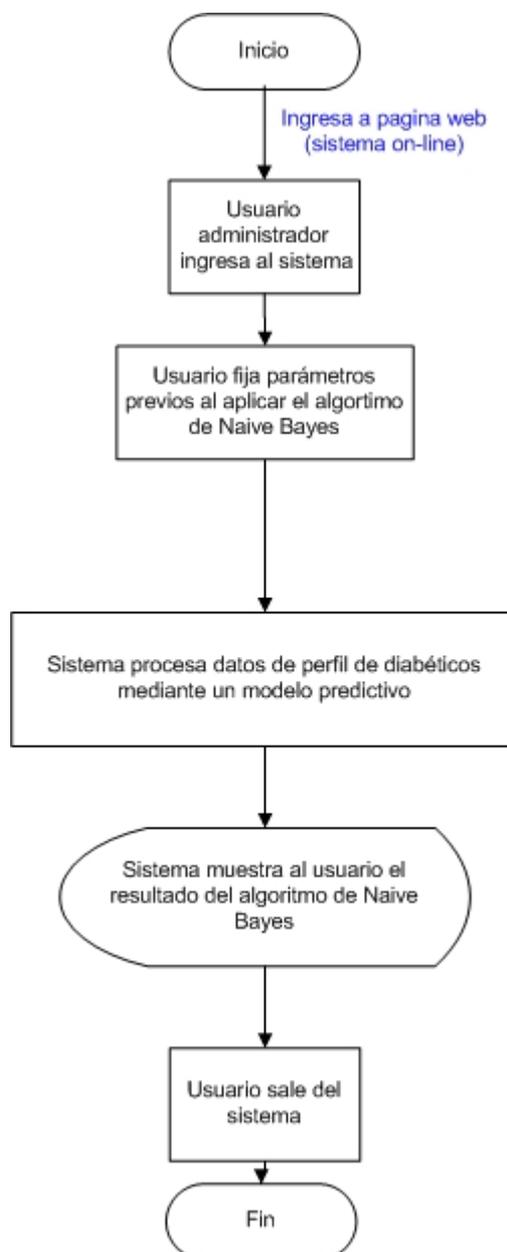
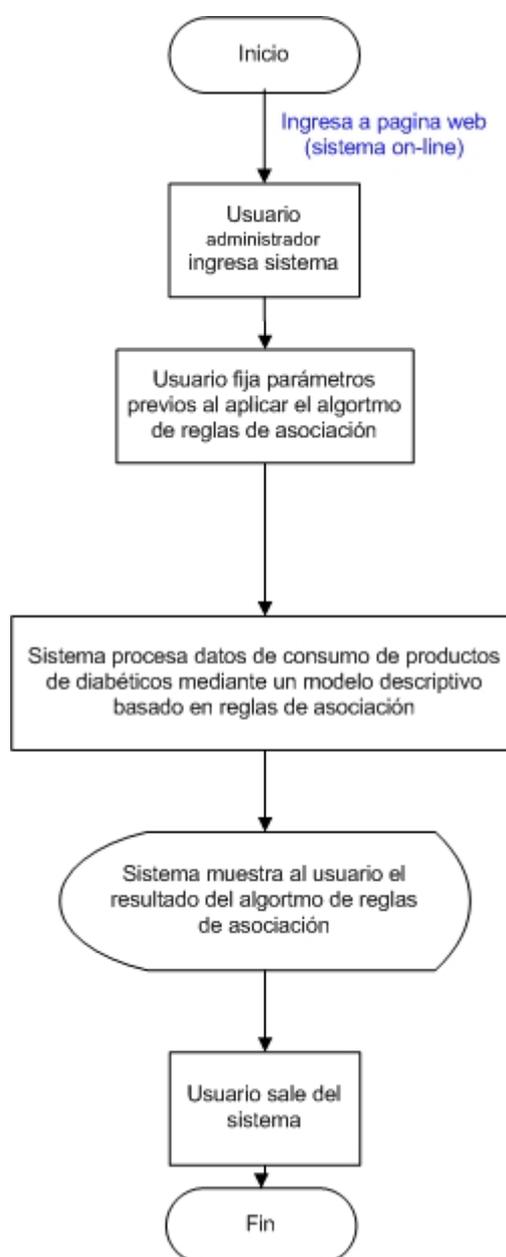


Figura 3.26 Diagrama de flujo: Naive Bayes.

### Diagrama de flujo de datos modelo reglas de asociación



**Figura 3.27** Diagrama de flujo: Reglas de asociación.

## **CAPÍTULO 4**

### **4. DISEÑO E IMPLEMENTACIÓN DEL SISTEMA**

#### **4.1. VISION GENERAL DEL DISEÑO**

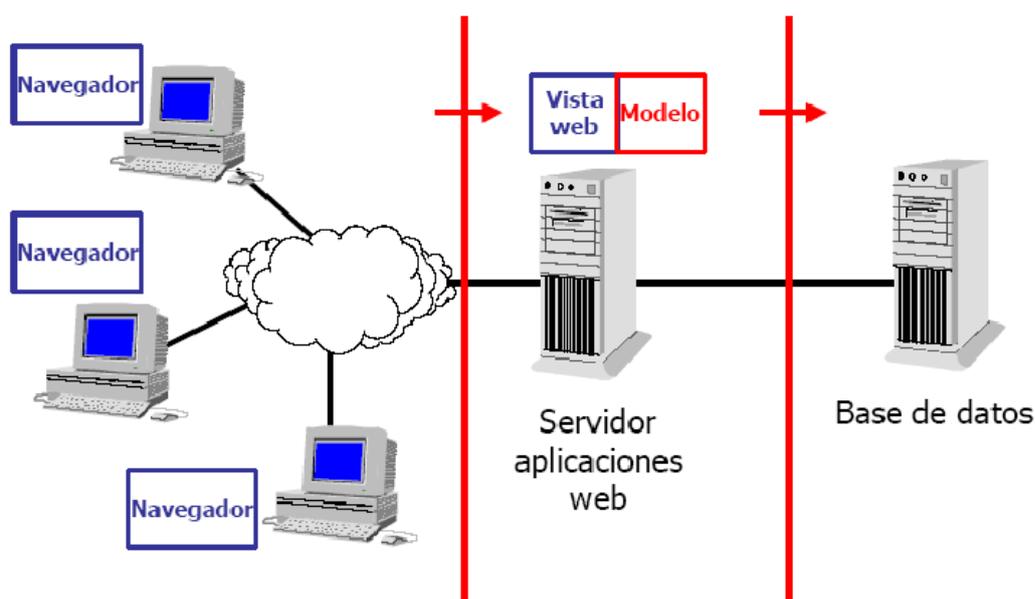
Las aplicaciones han pasado por un proceso evolutivo enorme. Desde su inicio con las aplicaciones monolíticas donde en una aplicación todo estaba ligado o mezclado por decirlo de alguna manera. Es decir, la interfaces de usuario, la lógica de cómo funcionaba la empresa y el manejo de la información almacenada y recuperada estaban juntas.

Luego la industria ha implementado un nuevo modelo de aplicaciones, las aplicaciones distribuidas cliente/servidor, que se convirtió en el estándar por un tiempo. Pero con la llegada de las aplicaciones Web se hacía necesario un nuevo estándar para las operaciones de los sistemas, y es por esto que se ha propuesto el modelo de las aplicaciones en n-capas.

Este modelo por lo general esta basado en un esquema de tres partes: acceso, lógica de negocios e interfaces de usuario. “Aunque es posible continuar sub dividiendo este modelo en sub capas para una mayor flexibilidad en la distribución del equipo de desarrollo y soporte durante la etapa de mantenimiento”[21].

### Arquitectura de aplicaciones con tres capas

En la figura 4.1 podemos apreciar un esquema que explica la implementación de una arquitectura de tres capas.



**Figura 4.1** Arquitectura de tres capas.

Arquitectura de aplicaciones en n-capas utilizando arquitectura DNA se ha convertido en el estándar para el software empresarial propuesto por Microsoft [22]. Se caracteriza por la descomposición de las aplicaciones

- Proporciona una escalabilidad, capacidad de administración y utilización de recursos mejorados.

- Cada capa es un grupo de componentes que realiza una función específica.
- Se puede actualizar una capa sin recompilar otras capas.

Por regla general, la capa de la presentación es una interfaz gráfica que muestra los datos a los usuarios. La capa de la lógica de negocios es responsable de procesar los datos recuperados y enviarlos a la capa de presentación. La capa de datos almacena los datos de la aplicación en un almacén persistente, tal como una base de datos relacional o archivos XML. Se pueden alojar todas las capas en el mismo servidor, pero también es posible alojar cada capa en varios servidores.

En resumen, las razones por las que la implementación de nuestra solución SACC utiliza arquitectura de 3 capas son las siguientes:

- Alta escalabilidad: El sistema SACC define en la capa de lógica del negocio, la implementación de los algoritmos de minería de datos, lo que permite agregar nuevos algoritmos de manera eficiente.
- Reutilización: Permite utilizar la capa de lógica de negocio en otras interfaces a futuro.

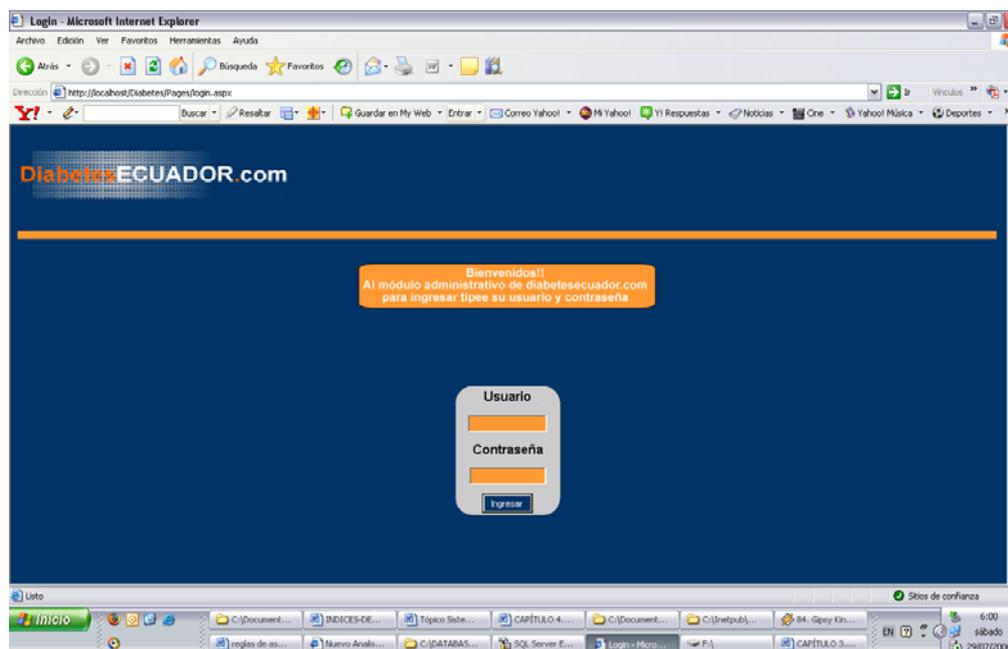
Sin embargo, debemos anotar que una de las desventajas de utilizar esta arquitectura, es que vuelve el diseño del sistema una tarea más compleja.

Aterrizando todos estos conceptos en nuestro proyecto se definieron las siguientes capas:

### **Capa de Aplicación (Browser cliente)**

Es el conjunto de páginas visibles para el cliente, y sobre las cuales interactúa directamente para obtener los resultados esperados. Esta interfase sirve también para recolectar los datos que el usuario ingresa al momento de registrarse y llenar la encuesta en línea que esta disponible para todos los usuarios.

Para la capa de Aplicación o presentación se utilizo la plataforma .NET y para el desarrollo se utilizó la tecnología aspx. En la figura 4.2 podemos apreciar la interfase Web que el usuario percibe al momento de ingresar al sistema.



**Figura 4.2** Pantalla de inicio al sistema.

### **Capa de lógica del Negocio (Web Server)**

La capa de lógica del negocio es la encargada de atender peticiones del usuario y procesar los datos (validarlos y gestionar su presentación y/o almacenamiento).

Como lenguaje de programación para implementar los objetos, clases y estructuras del sistema en conjunto con las páginas, se utilizó ASP.NET con C# y VB Net.

### **Capa de Almacenamiento (Data base Server)**

En él se guarda todo el repositorio donde se almacenan los datos, luego de que estos han sido recolectados, validados y procesados y sobre el cuál se aplicará las técnicas de minería de datos para extraer conocimiento. En esta capa se utilizó SQL Server como motor de base de datos.

## **4.2. ELECCIÓN DE LAS HERRAMIENTAS DE DESARROLLO**

En esta sección se muestra cuál fue el criterio utilizado para la selección de la herramienta de programación que permitió la implementación de nuestra solución.

### **4.2.1. CRITERIOS PARA LA ELECCIÓN DE LAS HERRAMIENTAS**

Para el desarrollo de este proyecto es necesario definir una herramienta que permita el desarrollo de aplicaciones Web y que al mismo tiempo tenga un comportamiento dinámico. Es por esto que se optó por la tecnología Active Server Pages (aspx) de la plataforma de desarrollo .NET utilizando como lenguaje de programación C# y VB Net.

A continuación se detalla una tabla comparativa entre las tecnologías más utilizadas para el desarrollo de aplicaciones Web. Los valores que se describen en la tabla corresponden a las habilidades del grupo de desarrollo de la aplicación.

	<b>Java Server Pages</b>	<b>PHP</b>	<b>Visual Basic.NET y C#.NET</b>
<b>Descripción</b>	Orientada a objetos	Orientada a objetos y estructural	Orientada a objetos
<b>Plataforma sistema operativo</b>	Toda Plataforma	Toda Plataforma	Microsoft Windows
<b>Integración con base de datos</b>	Ninguna Nativa. Mediante paquetes.	Nativa My-sql, Postgress	Nativa Sql-Server
<b>Aprendizaje</b>	Requiere mayor tiempo de entrenamiento	Requiere mayor tiempo de entrenamiento	Intuitivo, natural, documentado.
<b>Desempeño</b>	Bueno	Bueno	Bueno

**Tabla 4.1** Tabla comparativa de plataformas de programación.

La plataforma escogida fue Visual Basic .Net y C# . Net debido a calidad visual, desempeño funcional y potencia de procesamiento del lado del servidor. Además de que nuestro equipo de desarrolladores tiene más experiencia en la ejecución de proyectos utilizando esta herramienta.

EL motor de base de datos escogido es el SQL Server 2000 para aprovechar la máximo la compatibilidad de la plataforma Microsoft en conjunto con el servidor para publicación Internet Information Server.

## **4.2.2. JUSTIFICACIÓN PARA LA ELECCIÓN DE LAS HERRAMIENTAS SELECCIONADAS**

### **ASP.NET**

Las siglas ASP corresponden a las palabras Active Server Pages (Páginas Activas en el Servidor), es un completo marco de trabajo de construcción de aplicaciones Web. Una de las principales características de este modelo es la flexibilidad de escoger su lenguaje de programación.

Una de las características más importantes de las páginas ASP es la posibilidad de conectar con diferentes tipos de bases de datos, para extraer, agregar o eliminar datos de ellas y generar páginas con esos datos. Se generan en forma dinámica dependiendo de las sentencias que se establezcan, para así obtener los resultados del proceso realizado.

Puede conectarse a motores de bases de datos como: SQL, Access, Oracle, y cualquiera que soporte conexión ODBC<sup>6</sup> el cual proporciona un conjunto consistente de funciones para acceder a su base de datos.

---

<sup>6</sup> Open Data Base Connection (Conexión abierta a base de datos). Es una forma de conexión a veces de datos, independientemente del lenguaje o programa que estemos utilizando. Cada fabricante provee su propia librería (DLL) con las características de conexión a las bases de datos [23].

Sus funcionalidades permiten un acceso más sencillo y con la facilidad de poder emplear diferentes lenguajes de codificación, pero sin perder potencia de diseño ya que los principales: C#, Visual Basic.NET y Jscript.NET son orientados a objetos. Además que soportan: herencias, polimorfismos y encapsulación.

Su diseño e implementación es limpio y presenta un ambiente sofisticado de depuración de errores.

#### **4.3. RECOLECCIÓN E INTEGRACIÓN DE INFORMACIÓN PARA LA IMPLEMENTACIÓN DEL MODELO**

SACC utiliza dos fuentes de información para la aplicación de los algoritmos de minería. Datos obtenidos mediante las encuestas realizadas en la página Web e información obtenida mediante la base de datos transaccional del distribuidor de productos para diabéticos.

SACC se enfoca en el análisis de tres tipos de variables: variables de tipo demográfico, variables para medir el comportamiento de consumo y variables de perfil médico.

Las variables de tipo demográfico son: sexo, nombre, edad, ciudad, actividad, nivel de estudios. Las variables para el análisis de consumo son: consumo de cigarrillo, consumo de alcohol, consumo de productos para diabéticos, consumo de aspirina, consumo de medicina para diabetes, tipo de medicinas, consumo de reactivos. Las variables para definir el perfil médico son: antecedentes cardiovasculares, padece diabetes, edad en la que se diagnosticó diabetes, si tiene familiares que sufren diabetes, actividad física, consumo de alcohol, consumo de cigarrillo, consumo de aspirina.

Definitivamente se podían agregar más criterios de información sobre el perfil de las personas con respecto a la diabetes, sin embargo los utilizados nos permiten realizar un buen análisis orientado al consumo, así como también análisis orientado al perfil médico, según consulta realizada a especialista endocrinólogo <sup>7</sup>.

En la etapa de recolección y depuración de la información se realizaron procesos de discretización de las variables de tipo continua, tales como: edad, IMC (índice de masa corporal). Dentro de este proceso también se incluye la binarización de las transacciones de ventas para la generación de las reglas de asociación.

---

<sup>7</sup> Dr. Gonzalo García Baranderan. Jefe del Centro de Diabetes del Hospital Clínica Kennedy Guayaquil. Teléfono Consultorio: 042395538

#### 4.4. TRANSFORMACIÓN DE LOS DATOS OBTENIDOS PARA REALIZAR EL MODELO

Dentro de esta fase se generó también una nueva variable que es resultado de relacionar la estatura de la persona con su peso. Esta variable es conocida como índice IMC (índice de masa corporal)<sup>8</sup>, que es resultado de aplicar la siguiente fórmula.

$$\text{IMC} = \text{peso (Kg)} / \text{estatura}^2 \text{ (mt)}$$

IMC
BAJO PESO [0-18,5]
NORMAL [18,5 – 25]
SOBRE PESO [25 – 30]
OBESIDAD [ > 30 ]

**Tabla 4.3** Rango IMC

A continuación se presentan las tablas con las variables que fueron discretizadas:

---

<sup>8</sup> El índice de masa corporal (IMC, en sus siglas inglesas BMI) es un número (adimensional) que pretende determinar, a partir de la estatura y el peso, el rango más saludable de peso que puede tener una persona. Se lleva calculando desde principios de 1980. El IMC es igual al Valor Numérico (VN) resultante de la división del peso en kilogramos y el cuadrado de la estatura en metros [24].

Edad Discretizada
NINO[0-12]
ADOLESCENTE [12-18]
ADULTO JOVEN [18-35]
ADULTO MAYOR [35-65]
TERCERDA EDAD [MAYOR DE 65]

**Tabla 4.3** Rango Edad

En el anexo 3 se encuentra una tabla que detalla el resultado del proceso de discretización, paso necesario, antes de la aplicación del algoritmo Naive Bayes.

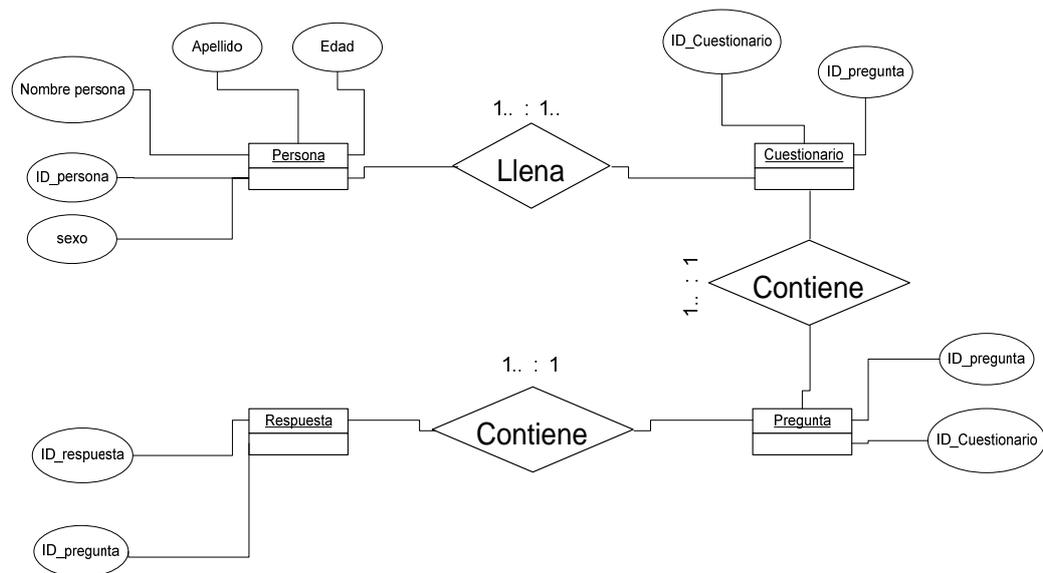
#### **4.5. DISEÑO DEL MODELO RELACIONAL**

En esta sección se describen los diferentes diagramas del modelo relacional de la base de datos, que son producto del análisis y diseño realizado en capítulos anteriores.

##### **Diagrama entidad relación**

El diagrama de entidad-relación (también conocido como DER, o diagrama E-R) es un modelo de red que describe con un alto nivel de abstracción la distribución de datos almacenados en un sistema [32].

Diagrama entidad relación para el sistema, tomando como datos la información recopilada a través de la encuesta en línea:



**Figura 4.3** Diagrama entidad-relación: Encuesta en línea.

### Explicación del diagrama

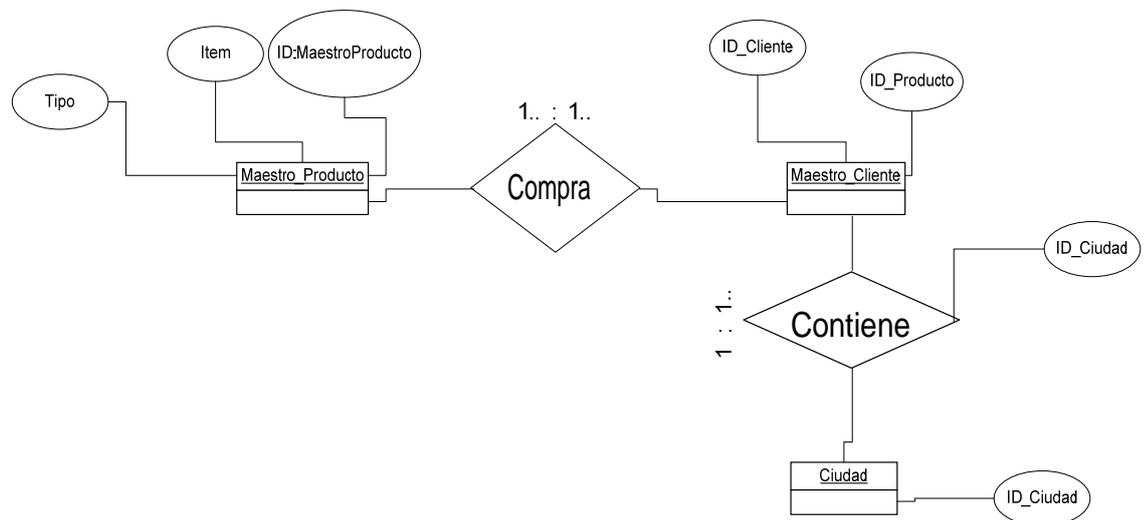
En el diagrama E-R se puede apreciar las entidades con sus respectivos atributos y la manera como dichas entidades se relacionan.

Por ejemplo: la relación entre persona y cuestionario (1.. : 1..) establece una relación de muchos a muchos entre si, es decir, una persona puede llenar varios cuestionarios y un cuestionario puede ser llenado por muchas personas. Esta relación de muchos a muchos nos

sugiere que muy probablemente nazca una tabla de hecho que resuelva dicha relación en el diseño multidimensional.

Por otro lado, si miramos la relación entre cuestionario y pregunta (1.. : 1), vemos que es una relación de muchos a uno, lo que nos indica que 1 cuestionario tiene muchas preguntas y una pregunta estar en 1 cuestionario.

A continuación se detalla el diagrama entidad relación del modelo descriptivo tomando como datos la información extraída de la base de datos transaccional de ventas.



**Figura 4.4** Diagrama entidad-relación. Transaccional de ventas.

### **Explicación del diagrama**

Podemos observar que la relación entre las entidades Maestro\_Cliente y Maestro\_Producto es de muchos a muchos ya que un cliente puede comprar muchos productos y un producto puede ser comprado por varios clientes. En efecto se deberá crear una tabla de hecho que relacione el evento de la venta de productos que involucra el detalle de los productos que compró el cliente y en que fecha se realizó la transacción.

Todos estos detalles se deben contemplar en el diseño multidimensional de la base de datos.

### 4.6. DISEÑO DEL MODELO LÓGICO

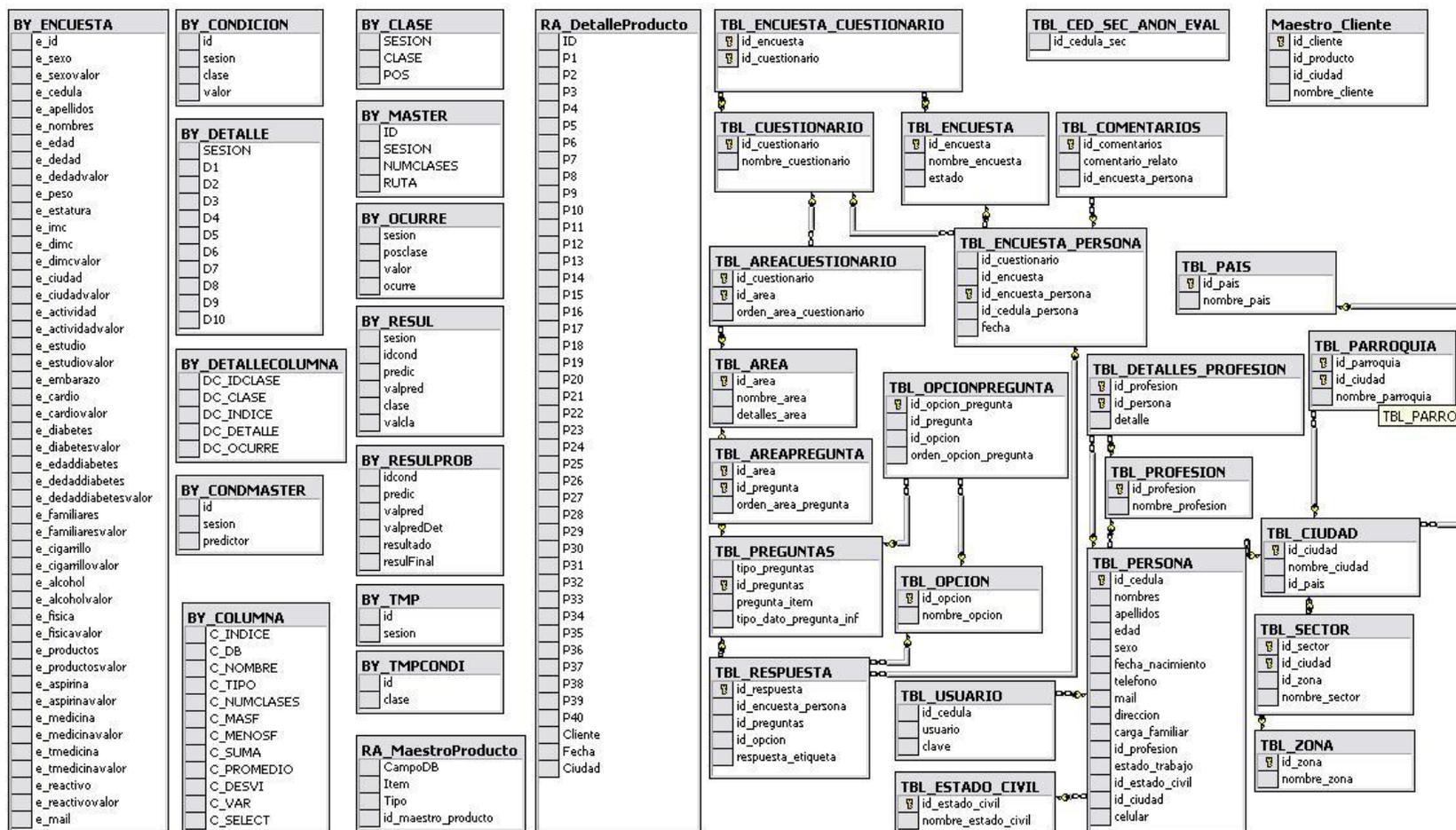
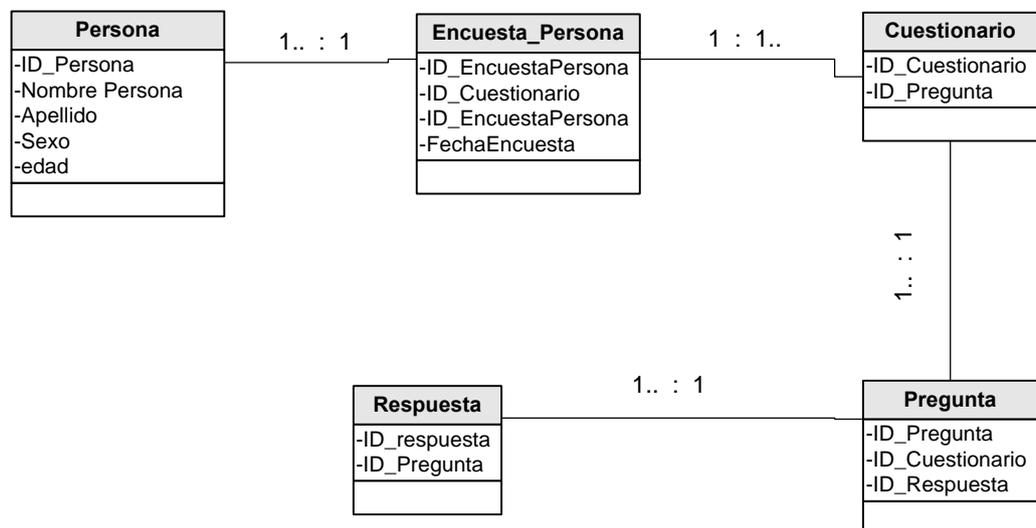


Figura 4.5. Base de datos SACC.

## 4.7. DISEÑO MULTIDIMENSIONAL DE LA BASE DE DATOS DEL SISTEMA

### Encuesta en línea

En la figura 4.6 podemos observar como la relación de muchos a muchos se transforma en una tabla de hecho en donde se registran los datos de la encuesta realizada, es en la tabla Encuesta\_Persona donde se relaciona a la persona con la encuesta y en que fecha la realizó. Esto nos da persistencia en los datos ya que permite recuperar una encuesta realizada en una fecha específica.



**Figura 4.6** Diseño multidimensional: Encuesta en línea.

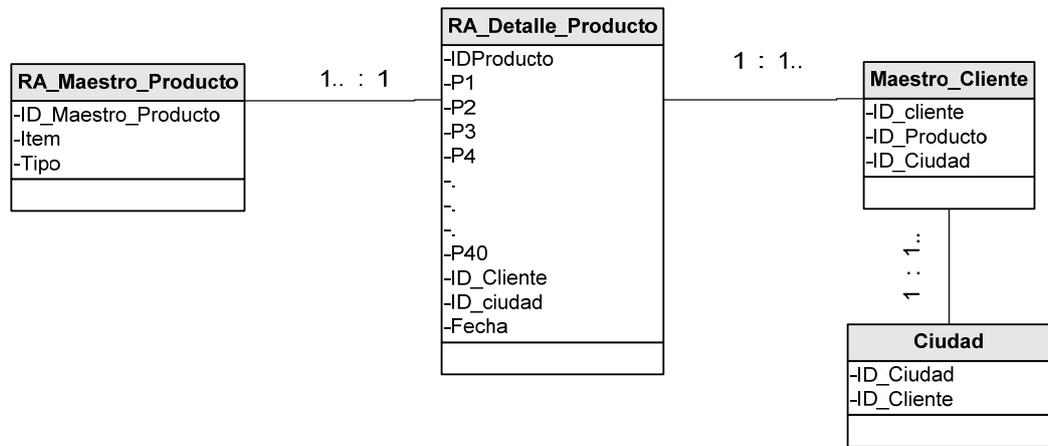
La persistencia de los objetos designa la capacidad de un objeto trascender en el espacio/tiempo salvando su estado en algún dispositivo físico justo antes de ser destruido, mediante algún mecanismo y en un determinado formato [25].

Además se observa como los atributos de las entidades son los atributos de las tablas y se identifican las claves primarias que son los identificadores únicos para cada instancia.

Las relaciones de las demás entidades se mantienen en las tablas.

### **Datos del transaccional de ventas**

En la figura 4.7 se observa como la relación de muchos a muchos entre las entidades Maestro\_Producto y Maestro\_Cliente da origen a una tabla de hecho llamada RA\_detalle\_Producto, en donde se guarda la información referente a que producto llevo cada cliente y en que fecha se realizó la transacción.

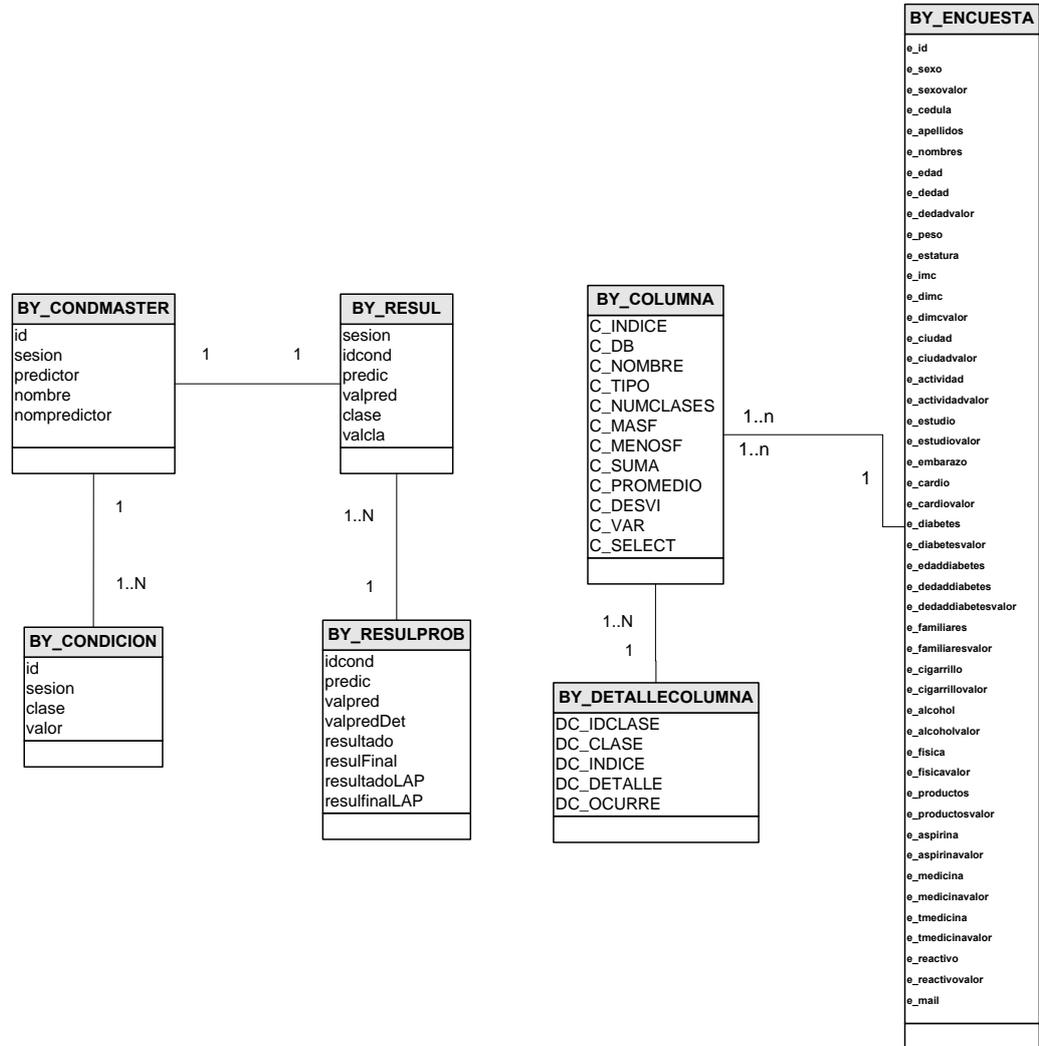


**Figura 4.7** Diseño multidimensional: Transaccional de ventas.

Es en esta tabla donde el SACC extrae el conocimiento basado en el algoritmo de reglas de asociación. Las relaciones de las demás entidades se mantienen en las tablas del modelo multidimensional.

### Datos del módulo análisis de perfiles

En la figura 4.8 se observa la tabla que contiene de manera discretizada los valores de respuesta a la encuesta BY\_ENCUESTA. Esta tabla es el resultado del proceso que recoge las respuestas de los usuarios a la encuesta y que posteriormente son ordenadas, depuradas y corregidas. En las tablas BY\_CONDMASTER, BY\_CONDICION, BY\_RESUL y BY\_RESULPROB se almacena la información de resultado de los análisis realizados sobre los datos de la encuesta.



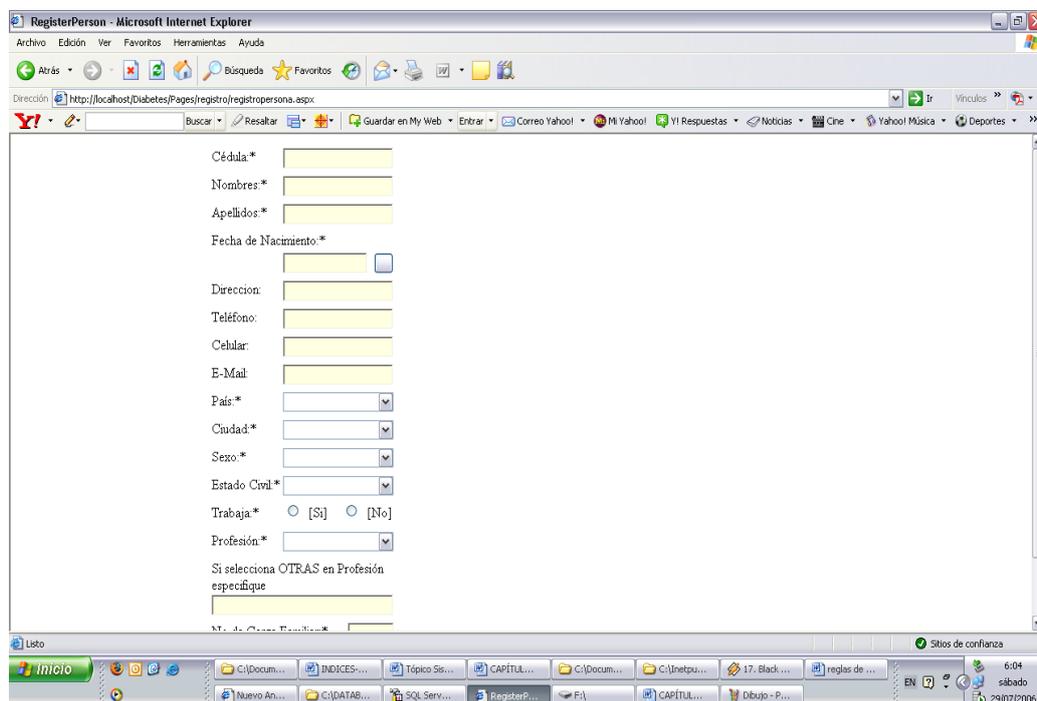
**Figura 4.8** Diseño multidimensional: Módulo análisis de perfiles.

#### **4.8. IMPLEMENTACIÓN DEL MÓDULO PARA REGISTRO DE INFORMACIÓN**

Como ya se ha mencionado, uno de los orígenes de información es la recolectada por el sitio Web: [www.diabetesecuador.com](http://www.diabetesecuador.com). Este sitio, desarrollado específicamente para nuestro proyecto, contiene 3 componentes principales que describiremos a continuación:

##### **Módulo de registro de usuarios**

En este módulo los usuarios deberán llenar un formulario de registro con información personal básica. Esta información se guarda en nuestra base de datos y enriquece la población sobre la que se realizará la tarea de minería de datos. Para el diseño se tomaron en cuenta portales similares y se personalizaron con preguntas relacionadas con la diabetes. En la figura 4.9 se muestra la apariencia del cuestionario en línea.



The image shows a screenshot of a Microsoft Internet Explorer browser window displaying a web form for user registration. The browser's address bar shows the URL: `http://localhost/Diabetes/Pages/registro/registropersona.aspx`. The form contains the following fields and options:

- Cédula\* (Text input)
- Nombres\* (Text input)
- Apellidos\* (Text input)
- Fecha de Nacimiento\* (Text input with a calendar icon)
- Dirección (Text input)
- Teléfono (Text input)
- Celular (Text input)
- E-Mail (Text input)
- País\* (Dropdown menu)
- Ciudad\* (Dropdown menu)
- Sexo\* (Dropdown menu)
- Estado Civil\* (Dropdown menu)
- Trabaja\* (Radio buttons for [Si] and [No])
- Profesión\* (Dropdown menu)
- Si selecciona OTRAS en Profesión especifique (Text input)
- M. de C. de Profesión\* (Text input)

The Windows taskbar at the bottom shows the system tray with the date and time: sábado 29/07/2006 6:04. Several application windows are open, including 'Nuevo An...', 'C:\DATAB...', 'SQL Serv...', 'RegisterP...', and 'CAPITUL...'.

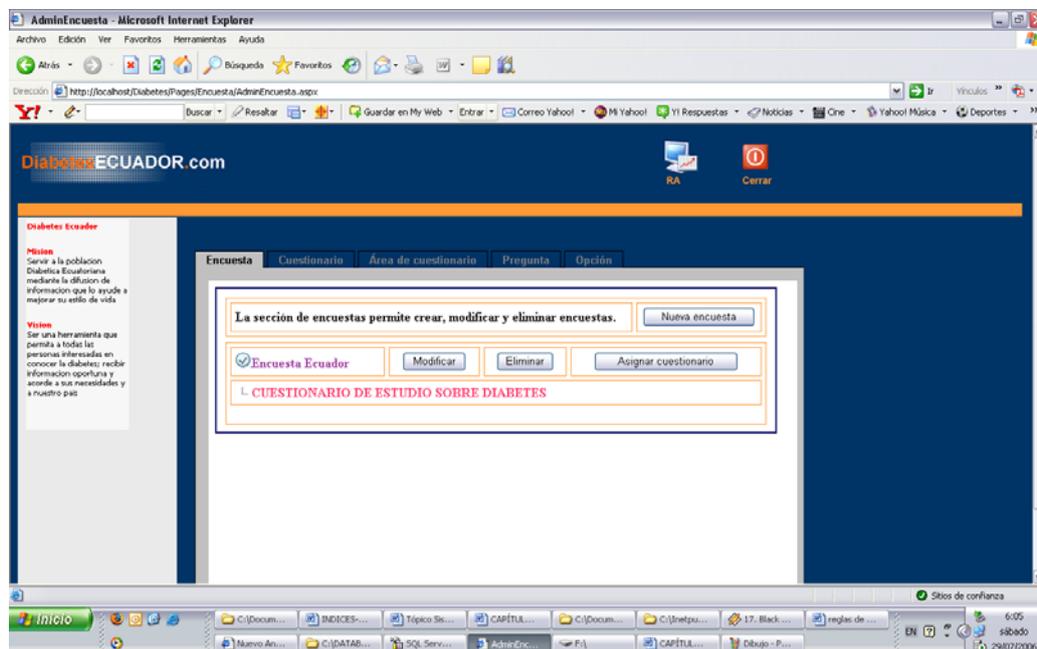
**Figura 4.9** Ventana de registro de usuarios.

## Módulo de cuestionarios

En este módulo se diferencian 2 perfiles de usuarios. El perfil administrador y el perfil usuario común. Cada uno tendrá opciones distintas que se detallan a continuación.

### Perfil usuario administrador

Posterior a la autenticación el usuario puede ingresar al módulo de administración de encuestas.



**Figura 4.10** Opciones del perfil administrador.

Como se muestra en la figura superior, desde éste perfil se pueden realizar modificaciones al cuestionario, sea en la estructura de las preguntas como en su contenido. Bajo éste perfil están las tareas de creación, modificación y eliminación de preguntas.

La creación de preguntas se refiere a la capacidad del sitio a crear nuevas preguntas dinámicamente e incorporarlas al cuestionario, el administrador puede modelar tanto la pregunta como las posibles respuestas. Ver figura siguiente.

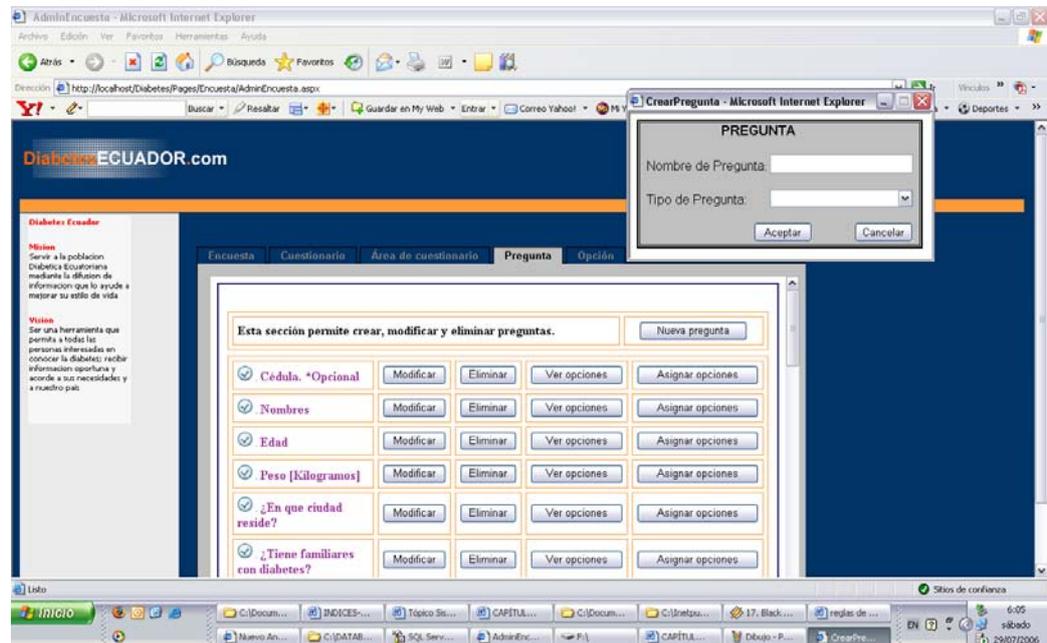
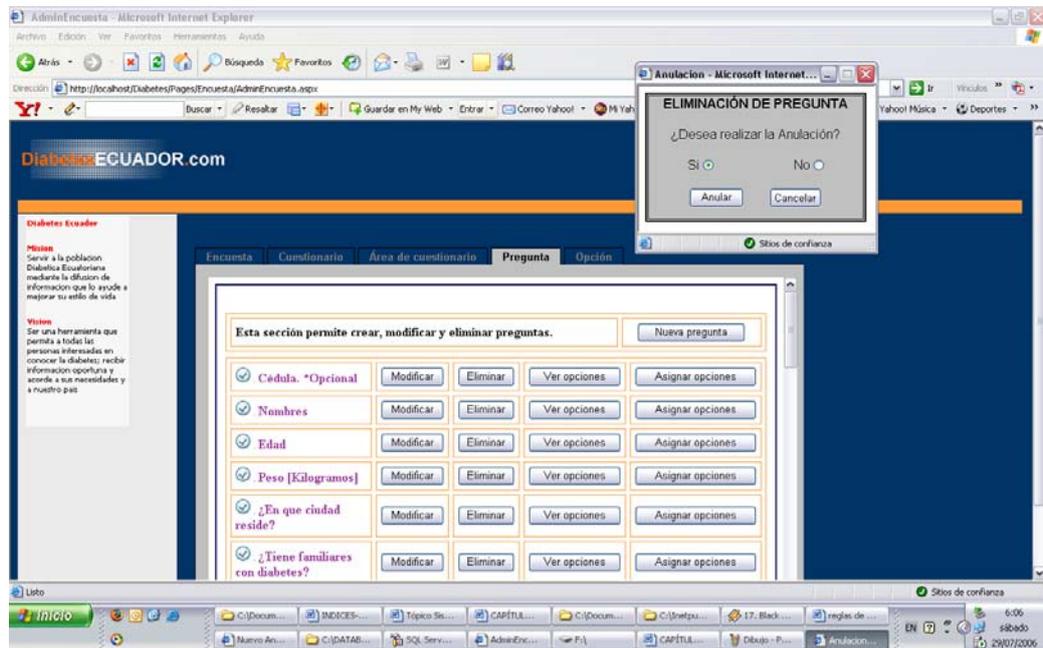


Figura 4.11 SACC: Creación de pregunta.

La eliminación de las preguntas se refiere a que el usuario con privilegios de administrador puede eliminar cualquier pregunta del cuestionario.



**Figura 4.12** SACC: Eliminación pregunta.

La modificación se refiere a que el usuario administrador puede modificar la estructura de la pregunta así como su contenido y la forma como se presentan las posibles respuestas, por ejemplo: una pregunta cuyas opciones múltiples de respuesta se las presente como un radio box, se las puede mostrar como una lista o como un check box. Esto ofrece mucha flexibilidad a la hora de diseñar una encuesta que sea atractiva para el usuario.

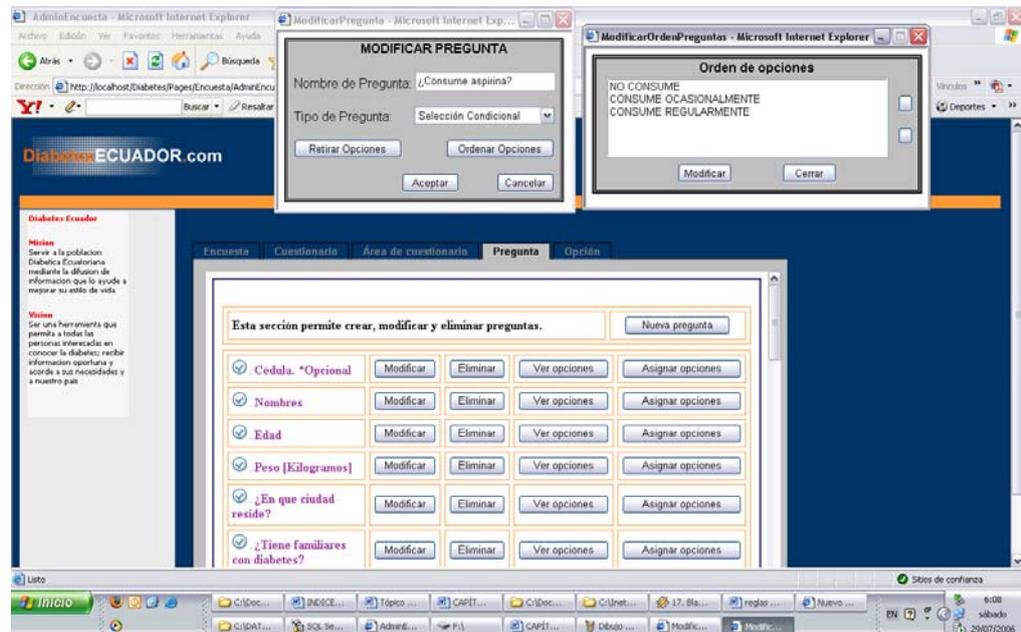


Figura 4.13 SACC: Modificar pregunta.

## Perfil usuario común

Este perfil se carga cuando un usuario ingresa al sistema y coloca su usuario correctamente. El usuario tendrá la opción de contestar un cuestionario y evaluar la probabilidad que ocurra un evento basado en parámetros preestablecidos.

The screenshot shows a web browser window titled 'PreguntaPersona - Microsoft Internet Explorer'. The address bar shows a local URL: 'http://localhost/Diabetes/Pages/Encuesta/PreguntaPersona.aspx?id\_encuesta=1&id\_cuestionario=1'. The page content is titled 'PERFIL DIABETICO' and includes a red warning message: 'Esta área presenta preguntas sobre su historial médico relacionado con la diabetes'. Below this, there are several questions with corresponding input fields:

- ¿Cuántos embarazos ha tenido? \*Ingrese 0 si nunca ha estado embarazada (text input)
- ¿Tiene familiares con problemas cardiovasculares? (dropdown menu)
- ¿Padece diabetes? (dropdown menu)
- Si usted padece diabetes ¿A qué edad se la diagnosticaron? (text input)
- ¿Tiene familiares con diabetes? (dropdown menu)
- ¿Consumes cigarrillo? (dropdown menu)
- Consumes bebidas alcohólicas (dropdown menu)
- Hábitos de actividad física (dropdown menu)

At the bottom of the form, there is a section titled 'HABITOS DE CONSUMO PARA TRATAMIENTO DE DIABETES.'.

**Figura 4.14** SACC: Encuesta usuario común.

## 4.9. IMPLEMENTACIÓN DE MINERÍA DE DATOS Y ALGORITMOS DE INFERENCIA

En esta sección se describen detalladamente y con ejemplos, cómo fueron ejecutados los algoritmos que analizan el comportamiento de las ventas de productos para diabéticos (reglas de asociación A priori) y el análisis de perfiles de personas diabéticas (Naive Bayes).

### 4.9.1. IMPLEMENTACIÓN DEL MODULO PARA LA EJECUCIÓN DEL ANÁLISIS DE VENTAS

Cuando se desea realizar la búsqueda en grandes bases de datos como sucede en minería de datos, se debe tratar de minimizar la cantidad de tiempo que se emplea en acceder las mismas, por cuanto estas operaciones de acceso a disco son por lo general las más lentas del proceso. SACC resuelve este problema ofreciendo al usuario la oportunidad de filtrar previamente la selección de los datos sobre los cuales se aplicara el modelo. Es así que el usuario puede escoger realizar el análisis entre un rango de fechas o por cliente o por ciudad.

The screenshot shows the SACC web application interface. The main content area is titled 'REGLAS DE ASOCIACION'. It features a search interface with the following elements:

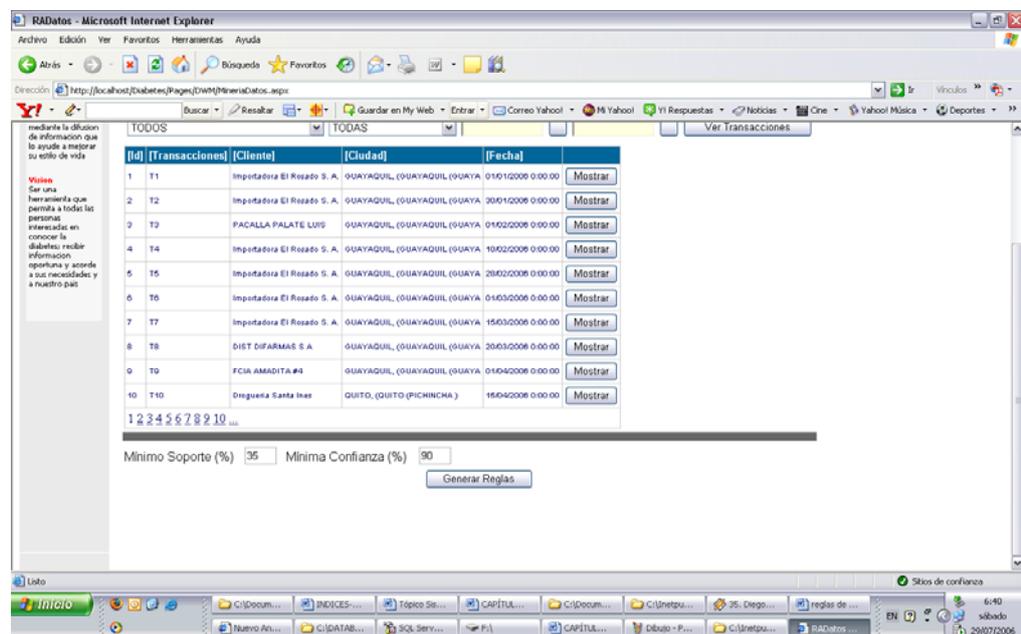
- Filters:** 'Cliente' (set to 'TODOS'), 'Ciudad' (set to 'TODAS'), 'Desde' (empty), and 'Hasta' (empty). A 'Ver Transacciones' button is located to the right of the 'Hasta' field.
- Table:** A table with columns: [Id], [Transacciones], [Cliente], [Ciudad], and [Fecha]. It lists 10 transactions.
- Calendar:** A calendar widget is open, showing the date selection process for the 'Desde' field. The date '29-Jul-2006' is highlighted.
- Buttons:** 'Mostrar' buttons are present next to the 'Fecha' column for each transaction row.

[Id]	[Transacciones]	[Cliente]	[Ciudad]	[Fecha]
1	T1	Impedansa El Resado S. A.	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	01/01/2006 0:00
2	T2	Impedansa El Resado S. A.	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	30/01/2006 0:00
3	T3	PACALLA PALATE LUIS	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	01/02/2006 0:00
4	T4	Impedansa El Resado S. A.	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	10/02/2006 0:00
5	T5	Impedansa El Resado S. A.	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	28/02/2006 0:00
6	T6	Impedansa El Resado S. A.	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	01/03/2006 0:00:00
7	T7	Impedansa El Resado S. A.	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	15/03/2006 0:00:00
8	T8	DIST DIFARMAS S A	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	20/03/2006 0:00:00
9	T9	FCA AMADITA #4	QUAYAGUIL, (QUAYAGUIL) (QUAYA)	01/04/2006 0:00:00
10	T10	Dispensa Santa Ines	QUITO, (QUITO) (PICHINCHA)	15/04/2006 0:00:00

Figura 4.15 SACC: Selección de rango de fechas para el modelo.

Un algoritmo como APriori satisface estos requerimientos, el mismo consta de dos pasos:

1. Encontrar todos los conjuntos frecuentes de ítems, es decir todos aquellos subconjuntos que cumplen con la condición  $s \geq s_{min}$ .
2. Usar los conjuntos frecuentes de ítems obtenidos para generar reglas que cumplan con la condición  $c \geq c_{min}$ .



**Figura 4.16** SACC: Selección de los parámetros de confianza y soporte.

La idea del algoritmo se basa en que si un conjunto de ítems cumple con la condición de soporte mínimo, entonces todo subconjunto de éste también la cumplirá, en realidad se usa el contra recíproco, cuando se obtiene un conjunto de ítems se chequea si todos los subconjuntos de éste, medidos en la iteración anterior, cumplían con la condición de

mínimo soporte, si se encuentra alguno que no la cumple se puede concluir, "a priori", que dicho conjunto no la cumplirá y por tanto no es necesario medirlo, esto evita mucha medición innecesaria, recuérdese que se quiere minimizar el tiempo total de acceso a la base de datos.

Para representar los conjuntos de ítems en cada uno de los pasos empleamos árboles, donde la profundidad de los mismos depende de la iteración que se esté realizando.

Al inicio se genera un árbol cuya raíz es una hoja, en la cual se representan todos los ítems mediante un arreglo de contadores. Realizar la medición entonces significa recorrer secuencialmente la base de datos incrementando en cada transacción los contadores correspondientes a los ítems presentes en ella, al final cada contador tendrá la cantidad total de veces que aparece este ítem en la base de datos.

A continuación se realiza un proceso de filtrado que significa eliminar todos aquellos ítems que no cumplen con la condición de soporte mínimo.

El siguiente paso es la generación de los candidatos que serán medidos a continuación, para esto se mezclan los ítems obteniendo todas las posibles combinaciones de dos, de esta forma se generará un nuevo árbol de profundidad 1 que tendrá en su raíz un nodo interior con una entrada para cada primer ítem que apuntará a una hoja donde están representados todos los conjuntos que contienen al mismo.

Este proceso se repite iterativamente hasta que se genere un árbol que no tiene ningún candidato, con la particularidad de que a partir del siguiente paso antes de realizar la medición se realiza un proceso de purga en el conjunto de candidatos, eliminando todos aquellos en que uno de sus subconjuntos no cumple con la condición de soporte mínimo.

Por ejemplo, para un caso de solo 5 ítems, se han generado en el conjunto de candidatos C los subconjuntos:

$$C = \{ \{1(\text{aciclovir}), 2(\text{ampicilina})\}, \{1(\text{aciclovir}), 4(\text{amoxicilina})\}, \{1(\text{aciclovir}), 5(\text{ranitidina})\}, \{2(\text{ampicilina}), 4(\text{amoxicilina})\}, \{2(\text{ampicilina}), 5(\text{ranitidina})\}, \{4(\text{amoxicilina}), 5(\text{ranitidina})\} \}$$

, el árbol queda:

1er nivel	1(aciclo)	2(amp)	3	4(amox)
2do nivel	2(amp),3,4(amox),5(rani)	3,4(amox),5(rani)	5(rani)	

Como se ve en el ítem 3 no existe hoja lo que significa que este fue filtrado en el paso anterior, en nuestra implementación no se hace distinción entre los ítems a nivel de la hoja, cuando se ha añadido el primero se añaden todos los demás aún cuando algunos de ellos nunca alcanzarán el soporte mínimo ( $\{1,3\},\{2,3\}$ ), como se verá más adelante este es la modificación principal al algoritmo a fin de garantizar máxima eficiencia.

En realidad en este árbol se representan los conjuntos:

$\{1,\{2,3,4,5\}\}, \{2,\{3,4,5\}\}, \{4,\{5\}\}$

A continuación se hace el proceso de medición, donde se toma cada transacción y se baja por el árbol hasta llegar a las hojas incrementándose los contadores correspondientes a los ítems restantes.

Por ejemplo sea:

$T = \{1,2,3,5\}$

Se baja por el 1 y se incrementan los contadores 2, 3 y 5, luego se baja por el 2 y se incrementan los contadores 3 y 5.

Pudiera pensarse que al incrementar los ítems 3, correspondientes a  $\{1,3\}$  y  $\{2,3\}$  se está realizando una operación innecesaria, sin embargo demostraremos que por el contrario, esto permite ganar en rapidez.

Cuando se va a bajar en el primer nivel es necesario hacer comparaciones para saber si está o no presente el hijo correspondiente al ítem analizado, en los nodos interiores esto es necesario y nosotros lo hacemos igual por cuanto se usa un arreglo de apuntadores y sólo tienen hijos aquellos caminos que fueron generados. Al llegar a la hoja si siguiéramos el algoritmo original sería necesario hacer lo mismo para cada ítem, por ejemplo si existen los ítems 2, 3 y 5, en el primer caso del ejemplo anterior, si la respuesta es positiva se incrementaría el contador, en otro caso no; luego al bajar por el 1 necesitaríamos hacer tres comparaciones, de ellas dos darían respuesta positiva (2 y 5) y la otra no (3), haciéndose por tanto dos incrementos. En el caso nuestro implementamos en la hoja un arreglo de contadores en lugar de un arreglo de apuntadores, por tanto solo hacemos los incrementos y así evitamos las comparaciones, en este ejemplo haríamos solo tres incrementos y ninguna comparación en la hoja, lo cual se hace en menos tiempo.

Nuestra implementación permite por tanto ahorrar todas las comparaciones en las hojas a cambio de un mayor gasto de memoria, por ejemplo: para una transacción de tamaño  $n$  en un árbol de nivel 2, al bajar por el primer ítem haríamos en la hoja  $n-1$  incrementos, el algoritmo haría en todos los casos  $n-1$  comparaciones además haría en el peor caso  $n-1$  incrementos, en el caso promedio  $(n-1)/2$  y solo en el mejor caso no haría ninguno, para todas las otras hojas sería similar. Aún en el peor caso, nuestra variante es superior pues se puede demostrar que en la mayoría de las máquinas el tiempo de ejecución de una operación de incremento es inferior al de una de comparación.

VINCES, (VINCES (LOS RIOS) YARUQUI (QUITO (PICHINCHA) ZARUMA, (ZARUMA (EL ORO))

Desde: ..... Hasta: .....

ID	Regla de Asociación	Confianza	
93	SI (COMPLEJO B MK TABLETAS, DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (DICLOXACILINA MK <math>\leq 500\text{MG}</math>)	97,2972972972973%	Mostrar Tipo
377	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (DICLOXACILINA MK <math>\leq 500\text{MG}</math>)	96,202531645696%	Mostrar Tipo
320	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, DICLOXACILINA MK <math>\leq 500\text{MG}</math> FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS) ENTONCES (COMPLEJO B MK TABLETAS)	95,8762336597938%	Mostrar Tipo
593	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, DICLOXACILINA MK <math>\leq 500\text{MG}</math>) ENTONCES (COMPLEJO B MK TABLETAS)	95,2%	Mostrar Tipo
336	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, DICLOXACILINA MK <math>\leq 500\text{MG}</math>, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (COMPLEJO B MK TABLETAS)	95%	Mostrar Tipo
378	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, DICLOXACILINA MK <math>\leq 500\text{MG}</math>, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS)	95%	Mostrar Tipo
91	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, DICLOXACILINA MK <math>\leq 500\text{MG}</math>, FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (COMPLEJO B MK TABLETAS)	94,7368421052632%	Mostrar Tipo
94	SI (COMPLEJO B MK TABLETAS, DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, DICLOXACILINA MK <math>\leq 500\text{MG}</math>, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS)	94,7368421052632%	Mostrar Tipo
571	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, FLUOXETINA MK <math>\leq 20\text{MG}</math> CAPSULAS) ENTONCES (COMPLEJO B MK TABLETAS)	94,3396226415094%	Mostrar Tipo
474	SI (DICLOFENACO MK <math>\leq 50\text{MG}</math> TABLETA, FUROSEMIDA MK <math>\neq 40\text{MG}</math> TABLETAS) ENTONCES (COMPLEJO B MK TABLETAS)	94,1860465116279%	Mostrar Tipo

Figura 4.17 SACC: Resultados de modelo descriptivo.

Estos resultados ayudan significativamente a la hora de ver que producto se está vendiendo más a que cliente. De esta manera se puede elaborar una estrategia de marketing o de negocios más efectivos y personalizados, o por ejemplo: se pueden armar combos dependiendo del volumen de las ventas por sector.

#### 4.9.2. IMPLEMENTACIÓN DEL MODULO PARA ANÁLISIS DE PERFILES

Antes de la ejecución del algoritmo se debe preprocesar la información para discretizar los valores continuos, posteriormente se debe contabilizar la ocurrencia de cada una de las clases de cada variable.

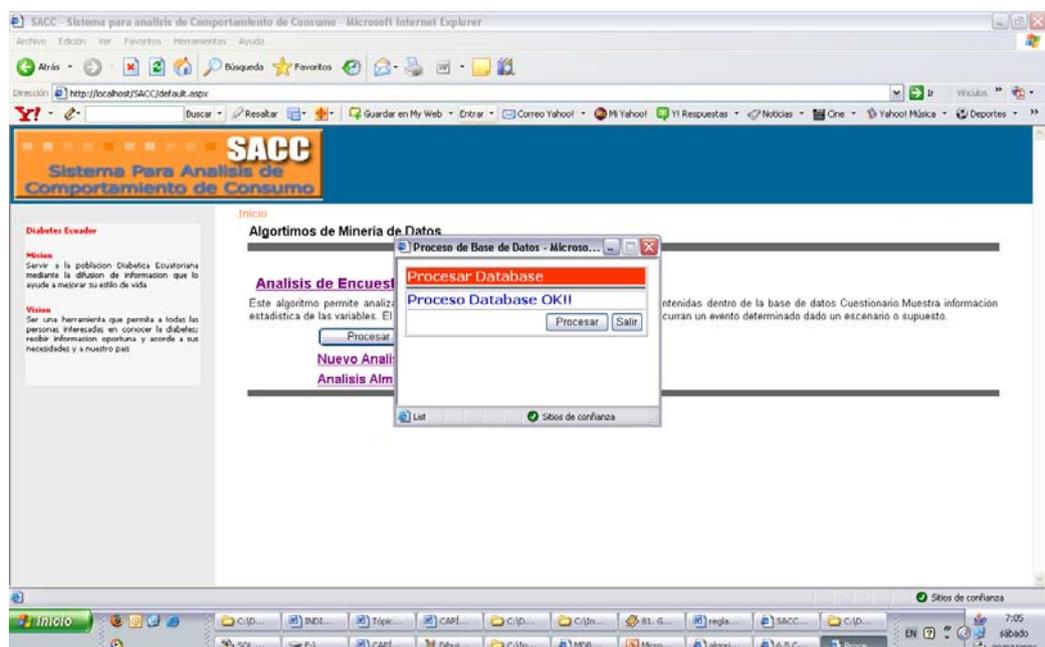
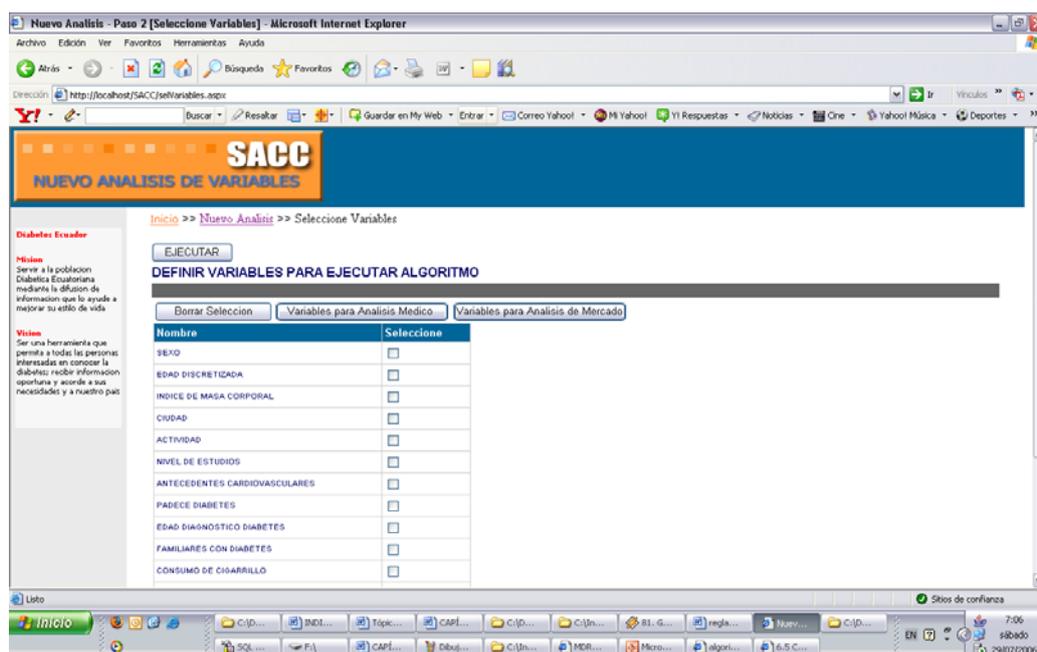


Figura 4.18 SACC: Procesar datos encuesta Naive Bayes.

Por ejemplo si se habla de la variable sexo, debemos contabilizar cuál es el número de ocurrencias para cada una de sus clases, que en este caso son sólo dos (masculino y femenino). Este proceso se repite para todas las variables.

Luego el usuario define cuales son las variables a intervenir en el proceso de análisis, de tal forma que puede trabajar con todo el universo de variables o hasta un mínimo de tres variables (para que el resultado sea significativo).

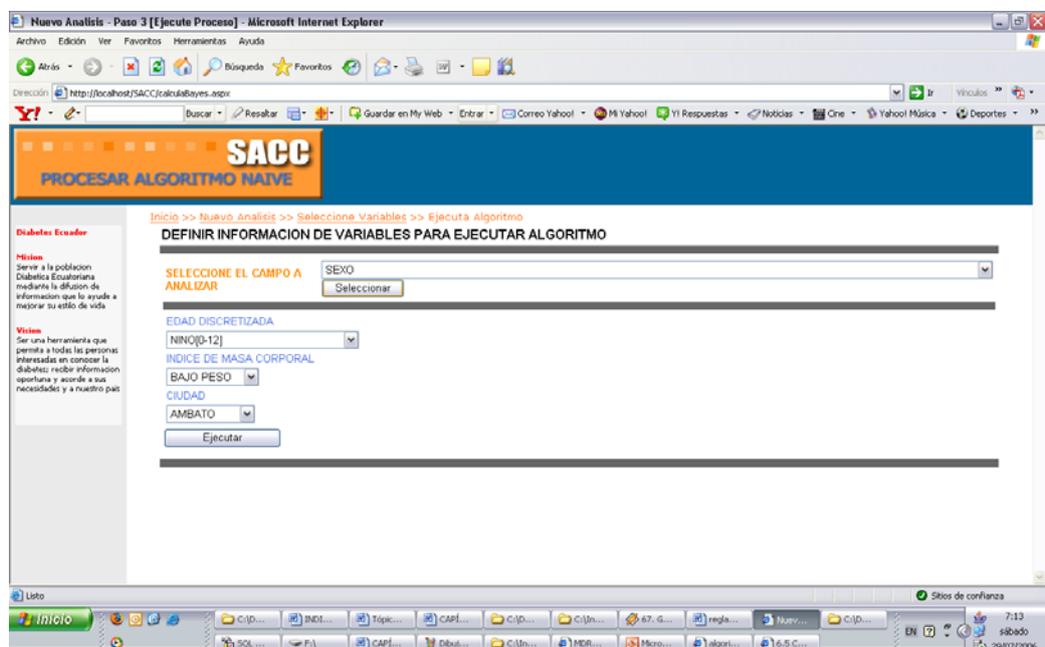


**Figura 4.19** SACC: Definir variables para análisis.

En el momento que el usuario selecciona las variables se genera una nueva sesión que automáticamente almacena esta información en el

modelo relacional, lo que permite que el usuario pueda utilizar cualquiera de las variables escogidas en la siguiente pantalla. Además el usuario puede seleccionar opciones predefinidas dependiendo del análisis que desea realizar; por ejemplo análisis de tipo médico o análisis de mercado.

Luego el usuario debe definir la variable objetivo a analizar y posteriormente cuales son las condiciones de las demás variables en la condición. Aquí se realiza un conteo de cada una de las clases de la variable que concuerda con el valor escogido para la variable a analizar junto a cada uno de los valores de clase de las variables de condición.



**Figura 4.20** SACC: Definir variable objetivo y variables condición.

Como debe recorrer todas las variables, si encontrase una que no tuviera ninguna ocurrencia, automáticamente detendría el análisis y la probabilidad sería 0. Luego de que tiene la probabilidad para cada clase de la variable a analizar, realiza una operación de normalización que devuelve las probabilidades para cada una de las clases. El sistema muestra todas las probabilidades y queda a discreción del usuario utilizar esta información.

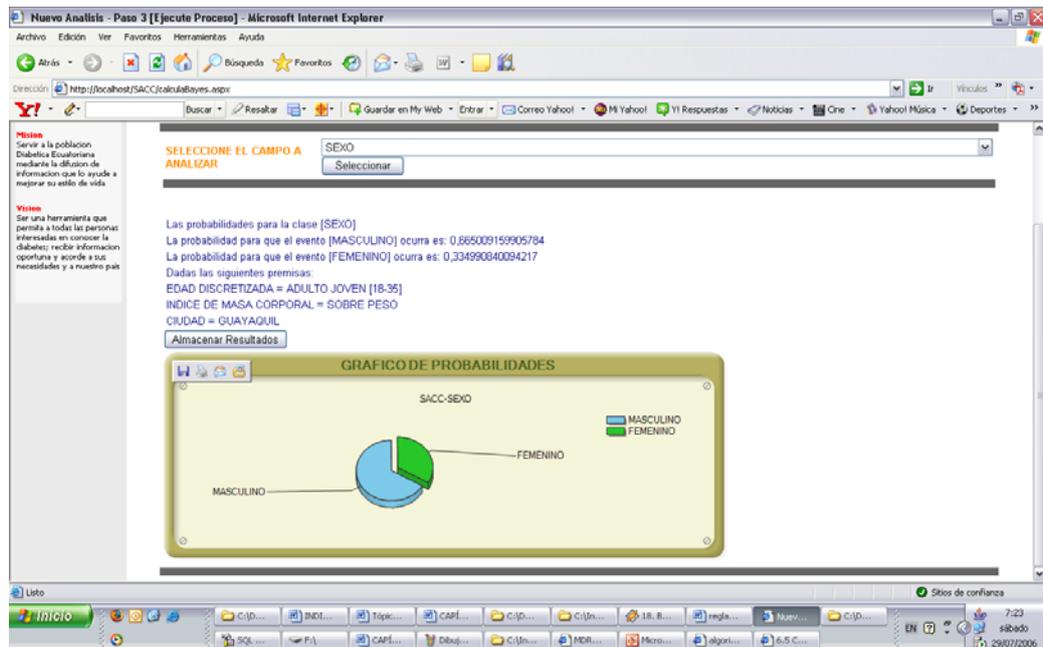


Figura 4.21 SAAC: Resultado del modelo predictivo.

Inclusive el usuario puede almacenar el resultado para futuras revisiones

#### **4.10. RESULTADOS OBTENIDOS A PARTIR DE LAS TÉCNICAS DE MINERÍA DE DATOS APLICADAS**

Los modelos de minería aplicados se enfocaron en resolver tanto el aspecto comercial, como el aspecto descriptivo de la enfermedad. A continuación se muestran alguno de los resultados que fueron obtenidos gracias a SACC.

##### **Modelo Predictivo – Naive Bayes**

Entre los resultados que devuelve la ejecución del algoritmo, los más relevantes son los siguientes:

- Existe una mayor incidencia de la enfermedad en los hombres que en las mujeres.
- Las mujeres en un mayor porcentaje nunca se han realizado un examen para determinar si tienen diabetes.
- Los adultos jóvenes son el grupo más numeroso que no realiza actividad física semanalmente y además sufren sobrepeso.
- El grupo de mujeres es más cuidadoso de comprar tiras para medición de diabetes que los hombres.

### **Modelo Descriptivo – Reglas de Asociación.**

Entre los resultados que devuelve la ejecución del algoritmo, los más relevantes son los siguientes:

- El consumo de medicamentos orales está fuertemente relacionado con la compra de consumibles para medición de niveles de glucosa.
- El consumo de insulinas inyectables no está fuertemente relacionado con el consumo de vitaminas.
- Los analgésicos y las vitaminas están relacionados.

#### **4.11. RESULTADOS DESEADOS VERSUS RESULTADOS OBTENIDOS**

Debido a que actualmente el proyecto se encuentra en fase inicial de publicación y difusión, la base de datos de encuesta solo tiene datos representativos de las ciudades principales y no de todos los grupos de personas. Sin embargo, a medida que más personas contesten el cuestionario en línea, la información tiende a ser más exacta.

## **CAPÍTULO 5**

### **5. FACTIBILIDAD Y COSTOS**

## 5.1. PROPUESTA DEL VALOR COMERCIAL DEL SISTEMA.

SACC es una herramienta que ayuda a resolver las interrogantes más comunes que surgen en todo tipo de empresa o institución que comercializa bienes o servicio. La metodología presentada en esta tesis se basa en el esquema implementado por la empresa multinacional Added Value™, definido como las 5 W (Who?, What?, When?, Where?, Why?); que traducidos al español representan ¿Quién es mi cliente?, ¿Qué productos esta comprando mi cliente?, ¿Cuándo mi cliente compra mis productos?, ¿Dónde mis clientes compran mis productos? y finalmente ¿Por qué compra mis productos?<sup>9</sup>

Descubrir y entender todos estos aspectos definitivamente le da a la empresa o institución una herramienta eficaz para alcanzar una de las metas que toda institución desea alcanzar que son: aumentar su participación en el mercado o simplemente vender más.

Inicialmente SACC está orientado a analizar los comportamientos de consumo de la población diabética ecuatoriana, mediante el uso de encuestas electrónicas que van alimentando una base de datos con preferencias y también mediante la relación de información de venta de

---

<sup>9</sup> Ver anexo 4 [26].

productos de la línea para diabéticos. Es necesario recalcar que el sistema maneja un perfil que puede ser cambiado a otro ámbito simplemente generando la información acorde al nuevo perfil deseado para que ésta sea procesada en el sistema.

Las razones que orientaron a escoger a la enfermedad Diabetes para aplicar un análisis de mercado e incidencia en nuestro país fueron:

- La diabetes es una enfermedad con una tasa de crecimiento constante en el Ecuador y en toda América,
- Una técnica efectiva para la prevención de la Diabetes Mellitus Tipo 2 es el correcto desenvolvimiento en hábitos alimenticios y cuidados de la salud, lo cual se puede alcanzar mediante acceso a la información correcta de tal forma las personas puedan incorporar estos consejos a su rutina diaria.
- No existe un directorio con información de médicos especialistas en el país.
- El uso y la difusión de Internet como canal de comunicación hacia los consumidores es más común en la actualidad.

- Los diabéticos son una población que depende del consumo regular de sus medicinas; por lo que se considera un mercado con alta lealtad.

Dadas las razones podemos identificar cuales serán nuestros grupos objetivo para la utilización de nuestro software y también realizaremos un análisis FODA de nuestro proyecto

### **Grupos Objetivos**

Nuestra herramienta permite enlazar las características de consumo de las personas que sufren diabetes con la información de comercialización de productos especializados para el tratamiento de la enfermedad.

Empresas que se dediquen a la comercialización de productos para diabéticos son nuestro primer grupo objetivo, debido a que las mismas pueden realizar un análisis de tendencias basado en información real del mercado, así como también incorporar información de su banco de datos de ventas.

El segundo grupo objetivo es la población diabética ecuatoriana y público en general, la cual podrá realizar consultas sobre temas de interés, consultar directorios de médicos especialistas, realizar una encuesta que mide la probabilidad de que sufra la enfermedad, entre otros.

### **Análisis FODA.**

El Sistema para Análisis de Comportamiento de Consumo (SACC) como todo proyecto cuenta con desafíos y amenazas en el mercado (Factores externos), así como también fortalezas e inconsistencias (factores internos) que se han analizado para la ejecución del proyecto.

<p><b>Fortalezas</b></p> <ul style="list-style-type: none"> <li>• Alto conocimiento de las herramientas que se usan, sistemas, algoritmos que aseguran calidad en los procesos y procedimientos.</li> <li>• Respaldo de empresas de sistemas que pueden avalar la tecnología usada.</li> </ul>	<p><b>Oportunidades</b></p> <ul style="list-style-type: none"> <li>• Creciente apogeo de personas con alto interés por los temas de salud.</li> <li>• Alianzas estratégicas con empresas farmacéuticas y/o de consumo interesadas en el mercado objetivo.</li> <li>• Pioneros en el mercado de software estadístico para investigación de consumo.</li> </ul>
<p><b>Debilidades</b></p> <ul style="list-style-type: none"> <li>• Poca estructura técnica y de personal para obtención de datos.</li> </ul>	<p><b>Amenazas</b></p> <ul style="list-style-type: none"> <li>• Fácil entrada de competidores (creadores de software)</li> <li>• Poca disposición de pagar el precio del software y del levantamiento de información por falta de conocimiento de los alcances del resultado.</li> </ul>

**Tabla 5.1** Análisis FODA de SACC

### 5.1.1. TCO COSTO TOTAL DE LA INVERSION

Una definición según Gartner Group<sup>10</sup>:"Total Cost of Ownership (Costo Total de Propiedad) es un modelo que ayuda a las empresas a analizar todos los costos y beneficios relacionados con la adquisición, desarrollo y uso de componentes de Tecnologías de Información (IT) a lo largo de su ciclo de vida".

El Total Cost of Ownership (TCO) ha sido permanentemente una ayuda significativa en el mundo de las Tecnologías de Información desde 1987, cuando Hill Kirwin de Gartner Group Inc., desarrolló el modelo para los PC o equipos de escritorio. Desde esta época el modelo se ha extendido a las redes de área local (LAN), sistemas cliente/servidor, computación distribuida, telecomunicaciones, centros de procesamiento de datos, y recientemente para sistemas portátiles (handheld).

El TCO es un modelo que ayuda tanto a los consumidores como a los ejecutivos de empresas evaluar los costos directos e indirectos que están relacionados con la compra de cualquier activo de capital, tal como el software y los computadores, pero no limitado sólo a ellos.

---

<sup>10</sup> Gartner Group, firma de investigación y consultoría que asesora a más de 10,000 clientes en el aprovechamiento de su tecnología para el logro de sus objetivos [27].

La valorización del TCO idealmente ofrece un costo que, además de considerar el precio de compra, incluye los costos relacionados con el uso y manutención de los equipos y sistemas.

Esto incluye los costos de implementación, capacitación de los usuarios, costos asociados con las fallas o períodos fuera de servicio (planeados o no planeados), incidentes de pérdida de rendimiento (repuestas más lentas), costos por incumplimientos (pérdida de reputación y pérdidas por la recuperación de la falla), espacio, energía, desarrollo, control de calidad, y muchos más. Por eso a veces el TCO es llamado Costo Total de Operación.

Hoy el TCO es usado para dar soporte a decisiones de compra y de planificación para un rango cada vez mayor activos que generan costos significativos de operación y manutención a lo largo de sus años de vida útil. El TCO es usado para apoyar la toma de decisiones que involucran equipos computacionales, sistemas de información, vehículos, equipos de laboratorio, equipos médicos, equipos para manufactura, maquinaria, entre otros.

A continuación se encuentra una tabla que define cuáles son los recursos que deben ser tomados en cuenta para un análisis en cada una de las etapas del ciclo de vida del sistema.

		CICLO DE VIDA DEL SISTEMA		
		ADQUISICION	OPERACION	AMPLIACIONES Y CAMBIOS
RECURSOS	HARDWARE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	SOFTWARE	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	PERSONAL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	REDES Y COMUNICACIONES	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
	INSTALACIONES	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**Tabla 5.2** Modelo para calcular TCO

Un buen análisis TCO permite descubrir los costos “ocultos” o los costos no obvios que podrían quedar fuera del proceso de compra o del proceso de planificación. El análisis comienza con el diseño de modelo adecuado, tal que cobra completamente los costos relacionados con el sistema a evaluar. La tabla 5.3 muestra un modelo básico para evaluar una compra de un sistema de información, que funciona bastante bien para muchas situaciones. Como se puede ver el modelo incluye los gastos de la adquisición, de la operación y los gastos de mejoras, en otras palabras se incluye la variable de la operación.

La siguiente tabla muestra componentes de costo considerados en los modelos de Gartner Group.

Costos Directos		Costos Indirectos	
Hardware		Costos de Usuarios	
	Inversiones en activos fijos y cuotas de leasing		Costos no presupuestados por auto soporte de los usuarios, capacitación informal de los usuarios
	Puede incluir las estaciones de trabajo, redes teléfonos		
Software		Downtime	
	Inversiones en activos fijos y cuotas de leasing		Perdida de productividad por tiempos planificados o no planificados de downtime
Administración de Sistemas		Costos de Oportunidad	
	Gastos por personal propio y pagos por servicios externos para la administración de la infraestructura IT		Otros beneficios no realizados como resultado de esta inversión
Soporte			
	Gastos por personal propio y pagos por servicios externos para soporte a los usuarios		
Desarrollo			
	Gastos de mano de obra para el diseño, testing, documentación y manutención		
Comunicaciones			
	Gastos anuales por arriendo de líneas, servicio de acceso remoto, Web, Wan		

**Tabla 5.3** Componentes para análisis TCO [28]

### 5.1.2. ROI RETORNO SOBRE LA INVERSION

Según Gartner Group. el retorno sobre la inversión (ROI) es un estimado del beneficio (el "retorno") sobre el dinero gastado (la

“inversión”) en una alternativa en particular, y consiste en determinar los beneficios, calcular los costes y resumir los resultados.

Fórmula: [26]

$$\text{ROI} = (\text{Beneficios/Costes}) \times 100$$

El retorno sobre la inversión ROI es el indicador que actualmente usan las empresas para medir el éxito de los proyectos y se calcula de forma directa. El cálculo real del ROI es muy simple, se divide los ingresos del proyecto para los costos, para obtener la relación. Los costos incluyen el análisis inicial, la adquisición del servicio de una empresa externa, el mantenimiento, la elaboración de informes, la administración y los costos de oficina. Aunque no se podrán cuantificar, siempre es conveniente señalar en los reportes los beneficios intangibles.

La necesidad de velar por el rendimiento de las inversiones hace que los esfuerzos del área de sistemas de las empresas se centren en la selección de proveedores que se enfoquen en resultados, que tengan experiencia y que den pautas concretas para conocer como calcular el ROI en forma práctica, precisa y eficiente

## **Beneficios al calcular el ROI**

Al aplicar el cálculo de ROI en proyectos que están orientados a la publicación de información en la Web tendremos que diferenciar entre los beneficios que nos trae publicar una página Web y los factores que producen rentabilidad. En los dos casos, tanto los beneficios como los Retornos Directos son palpables, pero mientras los primeros no son cuantificables los segundos si.

Según estudio realizado por el Dr. William Trochim [29], profesor de la universidad de Cornell, cuando un proyecto para la publicación de información y contenido Web está bien conceptualizado, siguiendo los parámetros de uso y acceso, estructura basada en objetivos, y un correcto posicionamiento en buscadores, se puede conseguir como beneficios típicos los siguientes:

- Nuevas oportunidades de negocio.
- Operación las 24 Horas los 365 días del año.
- Penetración en mercados nacionales.
- Penetración en mercados internacionales.
- Reforzar imagen corporativa.

- Branding.<sup>11</sup>
- Permanencia en el mercado.

Cuando los proyectos son conceptualizados buscando objetivos precisos, determinando desde el inicio cuales son los resultados que buscamos podemos definir los índices que justifican nuestra inversión.

El cálculo del ROI y TCO para nuestro proyecto esta detallado en los capítulos a continuación.

## **5.2. EVALUACIÓN DE COSTOS**

### **Análisis TCO y ROI para [www.diabetesecuador.com](http://www.diabetesecuador.com)**

Nuestro proyecto [www.diabetesecuador.com](http://www.diabetesecuador.com) debe tener en cuenta las siguientes consideraciones para su correcto análisis de costos.

### **Consideraciones Generales para Costos**

---

<sup>11</sup> Concepto Branding. Hace referencia a una serie de estrategias alrededor de la construcción de una marca. En Internet, branding va más allá de estrategias publicitarias. [30].

### **Fase de adquisición**

- No se asumen los costos de implementar una empresa desarrolladora de software, los costos están basados en la implementación del software específico.
- Se asume la adquisición de un equipo para desarrollo, ya que el arrendamiento o leasing por el periodo de desarrollo equivaldría al costo del equipo.
- La fase de análisis y desarrollo tiene como tiempo estimado de ejecución 3 meses.
- Se deben utilizar componentes que ayuden a la visualización de resultados y administración de contenido Web, dado que la naturaleza del proyecto es diferente al desarrollo de estos componentes es una mejor opción comprarlos e integrarlos.
- Proveedores de alojamiento Web, tales como BlazerNetwork [31], proporcionan planes acordes al tipo de proyecto a implementar, a continuación características básicas:
  - 800 Megabytes de espacio en Disco
  - 20 Gygabytes de espacio en Database
  - Interfaz para manejo de Base de Datos SQL- Server
  - Soporte para Microsoft ASP.net 1.1 y 2.0
  - Costo mensual: \$10.95

**Fase de desarrollo**

- La fase de despliegue y soporte al sistema tuvo como tiempo de ejecución 9 meses.
- Fueron necesarias alrededor de 1000 horas hombre para el desarrollo de todos los módulos y componentes necesarios. Se tomó en cuenta todas las fases del desarrollo de software. Se dividió el trabajo para tres integrantes con una carga horario de 20 horas semanales, con tiempo estimado para la ejecución en 3 meses.
- Se debe tener en cuenta el pago de servicios básicos para el desarrollo del proyecto.

**Fase de ampliaciones y cambios**

- Se debe tener en cuenta el pago de mano de obra para la persona encargada de actualizaciones de contenido y auspiciantes.

**Consideraciones generales para ingresos**

Se plantea el manejo de perfiles de auspiciantes para [www.diabetesecuador.com](http://www.diabetesecuador.com). Entre los perfiles se encuentran los siguientes:

### **Plan Auspiciante Oficial**

La categoría máxima de auspiciante. Tiene derecho a diseño de una página Web completa que puede mostrar información de sus productos e información general de empresa, presencia de su marca en la página principal de [www.diabetesecuador.com](http://www.diabetesecuador.com). El usuario puede ir a la página Web del auspiciante. El usuario puede utilizar SACC para sus análisis de mercado. Costo de suscripción: \$750 anuales

### **Plan Auspiciante Premium**

Tiene derecho a diseño de una página Web completa que puede mostrar información de sus productos e información general de empresa, presencia de su marca en la página principal de [www.diabetesecuador.com](http://www.diabetesecuador.com). El usuario puede ir a la página Web del auspiciante.

Costo de suscripción: \$250 anuales

### **Plan Auspiciante Gold**

Logotipo del auspiciante, presencia de su marca en páginas secundarias de [www.diabetesecuador.com](http://www.diabetesecuador.com). El usuario puede ir a la página Web del auspiciante.

Costo de suscripción: \$100 anuales

### **Plan Auspiciante Doctor**

Plan destinado a profesionales médicos relacionados al tratamiento de la diabetes. Publicación de información para contacto en la zona de doctores. El usuario puede utilizar SACC para sus análisis de médicos.

Costo de suscripción: \$25 anuales

Como marco de referencia, para el cálculo de las tarifas de los auspiciantes, nos hemos basado en precios de mercado. En el anexo 5 se encuentran propuesta de portales ecuatorianos semejantes.

## Tablas de Valores.

### COSTOS

		ADQUISICIÓN	
		CONCEPTO	COSTO
HARDWARE			
		Computador para Desarrollo	\$ 550,00
SOFTWARE			
		Licencia de desarrollo Aplicaciones	\$ 500,00
		Licencia de Base de Datos- Sql ServerMSDE	\$ 0,00
		Licencia Publicador Web	\$ 400,00
		Componentes Visualización	\$ 100,00
		Licencia Sistema Operativo Win Xp Pro	\$ 150,00
PERSONAL			
		Ninguno	\$ 0,00
REDES Y COMUNICACIONES			
		Pago Inicia de Hosting x 1 año	\$ 20,00
		Pago de Dominio - Inicio Contrato	\$ 10,00
INSTALACIONES			
		Ninguno	\$ 0,00
		Total	\$ 1.730,00

**Tabla 5.4** Costos fase de adquisición

		OPERACION / DESARROLLO	
		CONCEPTO	COSTO
HARDWARE			
		Ninguno	\$ 0,00
SOFTWARE			
		Ninguno	\$ 0,00
PERSONAL			
		Sueldo Programadores 3 meses x \$200	
		3 p x 3 meses x \$200	\$ 1.800,00
		Sueldo documentador 1p X 1mes x \$100	\$ 100,00
REDES Y COMUNICACIONES			
		Hosting 12 meses x \$ 15	\$ 180,00
INSTALACIONES			
		Pago de servicio básicos x 3 meses	\$ 60,00

Alquiler x 3 meses	\$ 150,00
Total	\$ 2.290,00

**Tabla 5.5** Costos fase de operación y desarrollo

		AMPLIACIONES Y CAMBIOS	
		CONCEPTO	COSTO
HARDWARE			
	Ninguno		\$ 0,00
SOFTWARE			
	Ninguno		\$ 0,00
PERSONAL			
	Sueldo Programador Mantenimiento WEB		
	1 p x 12 meses x \$50		\$ 600,00
REDES Y COMUNICACIONES			
	Ninguno		\$ 0,00
INSTALACIONES			
	Pago de Servicio básicos x 9 meses		
	(Agua, Luz, teléfono) = \$10 x 9 meses		\$ 90,00
	Alquiler x 9 meses		\$ 1050,00
	Total		\$ 1.740,00

**Tabla 5.6** Costos fase de ampliaciones y cambios

## INGRESOS

		TABLA DE INGRESOS ANUALES	
		Concepto	Valor
Publicidad tipo 1			
	Auspiciante Oficial		
	Página Web completa		
	Logotipo página principal		
	Link a Web de auspiciante		
	Uso de SAAC		
	\$750 x contrato anual x 5 subscriptores		\$ 3.750,00
Publicidad tipo 2			
	Auspiciante Premium		
	Página Web completa		
	Logotipo página principal		
	Link a Web de auspiciante		
	\$250 x contrato anual x 5 subscriptores		\$ 1.250,00
Publicidad tipo 3			
	Auspiciante Gold		
	logotipo en paginas secundarias		
	LINK A WEB DEL AUSPICIANTE		
	\$100 x CONTRATO ANUAL x 10 subscriptores		\$ 1.000,00
PUBLICIDAD TIPO 4			
	ZONA DE DOCTORES		
	DATOS PERSONALES Y DE CONTACTO		
	USO DE SACC		
	\$25 x CONTRATO ANUAL x 100 subscriptores		\$ 2.500,00
		Total	\$ 8.500,00

**Tabla 5.7 Ingresos totales anuales**

ROI TOTAL DEL PROYECTO	
Total beneficios	\$ 8.500,00
Total costos	\$ 5.760,00
Costo/beneficio	1,647286822
ROI	164,7%

**Tabla 5.8 ROI total del Proyecto**

Flujo de Efectivo										
Detalle	Mes 0	Mes 1	Mes 2	Mes 3	Mes 4	Mes 5	Mes 6	Mes 7	Mes 8	Mes 9
<b>INGRESOS</b>										
Ingresos por Ventas										
Auspiciante Oficial	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 312,50	\$ 312,50	\$ 312,50	\$ 312,50	\$ 312,50	\$ 312,50
Auspiciante Premium	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 104,17	\$ 104,17	\$ 104,17	\$ 104,17	\$ 104,17	\$ 104,17
Auspiciante Gold	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 83,33	\$ 83,33	\$ 83,33	\$ 83,33	\$ 83,33	\$ 83,33
Auspiciante Doctores	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 208,33	\$ 208,33	\$ 208,33	\$ 208,33	\$ 208,33	\$ 208,33
<b>Total Ingresos</b>	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 708,33	\$ 708,33	\$ 708,33	\$ 708,33	\$ 708,33	\$ 708,33
<b>Ingresos Acumulados</b>	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 708,33	\$ 1.416,66	\$ 2.124,99	\$ 2.833,32	\$ 3.541,65	\$ 4.249,98
<b>EGRESOS</b>										
Hardware	\$ 550,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
Software	\$ 1.150,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
Personal	\$ 0,00	\$ 600,00	\$ 600,00	\$ 700,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00
Redes y Comunicaciones	\$ 30,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00
Instalaciones	\$ 0,00	\$ 70,00	\$ 70,00	\$ 70,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00
<b>Total Egresos</b>	\$ 1.730,00	\$ 685,00	\$ 685,00	\$ 785,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00
<b>Egresos Acumulados</b>	\$ 1.730,00	\$ 2.415,00	\$ 3.100,00	\$ 3.885,00	\$ 4.010,00	\$ 4.135,00	\$ 4.260,00	\$ 4.385,00	\$ 4.510,00	\$ 4.635,00
<b>FLUJO DE EFECTIVO</b>	<b>\$ -1.730,00</b>	<b>\$ -2.415,00</b>	<b>\$ -3.100,00</b>	<b>\$ -3.885,00</b>	<b>\$ -3.301,67</b>	<b>\$ -2.718,34</b>	<b>\$ -2.135,01</b>	<b>\$ -1.551,68</b>	<b>\$ -968,35</b>	<b>\$ -385,02</b>

**Tabla 5.9** Flujo de Efectivo desde mes 0 hasta mes 9

Flujo de Efectivo								
Mes 10	Mes 11	Mes 12	Mes 13	Mes 14	Mes 15	Mes 16	Mes 17	Mes 18
\$ 312,50	\$ 312,50	\$ 312,50	\$ 312,50	\$ 312,50	\$ 312,50	\$ 0,00	\$ 0,00	\$ 0,00
\$ 104,17	\$ 104,17	\$ 104,17	\$ 104,17	\$ 104,17	\$ 104,17	\$ 0,00	\$ 0,00	\$ 0,00
\$ 83,33	\$ 83,33	\$ 83,33	\$ 83,33	\$ 83,33	\$ 83,33	\$ 0,00	\$ 0,00	\$ 0,00
\$ 208,33	\$ 208,33	\$ 208,33	\$ 208,33	\$ 208,33	\$ 208,33	\$ 0,00	\$ 0,00	\$ 0,00
\$ 708,33	\$ 708,33	\$ 708,33	\$ 708,33	\$ 708,33	\$ 708,33	\$ 0,00	\$ 0,00	\$ 0,00
\$ 4.958,31	\$ 5.666,64	\$ 6.374,97	\$ 7.083,30	\$ 7.791,63	\$ 8.499,96	\$ 8.499,96	\$ 8.499,96	\$ 8.499,96
\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00	\$ 0,00
\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00	\$ 50,00
\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00	\$ 15,00
\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00	\$ 60,00
\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00	\$ 125,00
\$ 4.760,00	\$ 4.885,00	\$ 5.010,00	\$ 5.135,00	\$ 5.260,00	\$ 5.385,00	\$ 5.510,00	\$ 5.635,00	\$ 5.760,00
<b>\$ 198,31</b>	<b>\$ 781,64</b>	<b>\$ 1.364,97</b>	<b>\$ 1.948,30</b>	<b>\$ 2.531,63</b>	<b>\$ 3.114,96</b>	<b>\$ 2.989,96</b>	<b>\$ 2.864,96</b>	<b>\$ 2.739,96</b>

**Tabla 5.10** Flujo de Efectivo desde mes 10 hasta mes 18

## **CONCLUSIONES Y RECOMENDACIONES**

## CONCLUSIONES

- La encuesta en línea mostró ser una herramienta eficaz y efectiva ya que permitió la recolección de grandes volúmenes de información de una manera sencilla al tiempo que se alimentaba directamente la base de datos. Gracias a esta forma de extracción de datos se ahorró tiempo y costos, en los cuales se hubiera tenido que incurrir en una tabulación manual.
- En las pruebas realizadas se pudo observar que el algoritmo de reglas de asociación es un descriptor muy preciso y útil a la hora de extraer conocimiento de un conjunto de datos que están relacionados entre si, ya que analiza todas las combinaciones de productos y descarta las que no cumplen con los parámetros mínimos de confianza y soporte. Con esto solo las más significativas son presentadas al usuario lo que ayudará a las empresas a ofertar acertadamente lo que el consumidor realmente quiere.
- El análisis financiero demuestra que el proyecto es rentable ya que la inversión que se genera es recuperada en tan solo un año; lo que lo hace atractivo para potenciales empresarios, inversionistas o distribuidores.

- Fue un acierto la selección del algoritmo Naive Bayes para el modelo predictivo ya que ofrece gran flexibilidad en la selección de las variables

## RECOMENDACIONES

- Se recomienda la utilización de una población extensa para que el algoritmo Naive Bayes sea preciso, ya que con cada individuo se entrena y se perfecciona.
- Los parámetros de cobertura y confianza son vitales a la hora de aplicar las reglas de asociación ya que estos inciden en la convergencia y exactitud del método. Se recomienda tener una confianza alta, así se garantiza que la regla sea efectiva. Hay que tener presente que la cobertura siempre es menor que la confianza.
- Dado la flexibilidad del sistema se puede adaptar para cualquier tipo de datos. Se recomienda incursionar en otros mercados como por ejemplo: autoservicios, heladerías, farmacias, boutiques y en general cualquier empresa con distintas líneas de productos.

## BIBLIOGRAFÍA

### Referencias de libros

- [1]** Libro: Introducción a la Minería de Datos. Autores: José Hernández Orallo, M<sup>a</sup> José Ramírez Quintana, César Ferri Ramírez. Editorial: Prentice Hall.
- [14]** Libro: Vademécum Ecuador. Autores: Editorial: .
- [15]** Libro electrónico: IMSPlus a Junio 2006 de empresas calificadas en el Ministerio de salud publica. Autores: Empresa IMS.
- [18]** Libro: Knowledge Discovery in Databases. Autores: Piatetski-Shapiro G.; Frawley W.J Editorial: AAAI/MIT Press. (1991).
- [19]** Libro: Applied Logistic Regression. Autores: David W. Hosmer, Stanley Lemeshow, Ed. John Wile. (1989).

### Referencias de Internet

- [2]** Entidad: Asociación Americana de Diabetes. Artículo: All about diabetes. URL: <http://www.diabetes.org/about-diabetes.jsp>
- [3]** Entidad: Las Enfermedades. Artículo: Diabetes, Definición, tipos y síntomas. URL: <http://www.lasenfermedades.com/21.html>
- [4]** Entidad: Asociación Americana de Diabetes. Artículo: Síntomas de la diabetes. URL: <http://www.diabetes.org/espanol/todo-sobre-la-diabetes/sintomas.jsp>
- [5]** Entidad: Asociación Americana de Diabetes. Artículo: Todo acerca de la diabetes. URL: <http://www.diabetes.org/espanol/todo-sobre-la-diabetes.jsp>
- [6]** Entidad: Odontodos. Artículo: Por el avance de la diabetes piden a la ONU una rápida intervención. URL: <http://www.odontodos.net/notas-interes/diabetes.htm>
- [7]** Entidad: Asociación Americana de Diabetes. Artículo: Información sobre la diabetes. URL: <http://www.diabetes.org/uedocuments/diabetes.pdf>

**[8]** Entidad: Asociación Americana de Diabetes. Artículo: Diabetes tipo 2. URL: <http://www.diabetes.org/espanol/todo-sobre-la-diabetes/diabetes-tipo-2.jsp>

**[9]** Entidad: Asociación Americana del Corazón. Artículo: Diabetes. URL: <http://www.americanheart.org/presenter.jhtml?identifier=3018643>

**[10]** Entidad: Asociación Americana de Diabetes. Artículo: Diabetes gestacional. URL: <http://www.diabetes.org/espanol/todo-sobre-la-diabetes/gestacional.jsp>

**[11]** Entidad: Asociación Americana del Corazón. Artículo: Diabetes. URL: <http://www.americanheart.org/presenter.jhtml?identifier=3018643>

**[12]** Entidad: Asociación Americana de Diabetes. Artículo: Medicamentos via Oral. URL: <http://www.diabetes.org/espanol/diabetes-tipo-2/medicamentos-via-oral.jsp>

**[13]** Entidad: Asociación Americana de Diabetes. Artículo: Insulina. URL: <http://www.diabetes.org/espanol/diabetes-tipo-2/insulina.jsp>

**[16]** Entidad: El Instituto Nacional de Estadística y Censos – INEC, es un órgano del Sistema Estadístico Nacional, con sede en

Quito. Es una entidad de derecho público, con personería jurídica, adscrita al Ministerio de Economía y Finanzas del Ecuador. Artículo: Diez principales causas de mortalidad año 2003. URL: <http://www.inec.gov.ec>

**[17]** Entidad: MCG Health System. Artículo: Estadísticas diabetes URL:

<http://www.mcghealth.org/Greystone/sadult/diabetes/stats.html>

**[20]** Entidad: Universidad de Chile . Artículo: Tutorial de UML URL: <http://www.dcc.uchile.cl/~psalinas/uml/introduccion.html>.

**[21]** Entidad: El Guille. Artículo: Aplicaciones distribuidas tres capas URL:

[http://www.elguille.info/colabora/NET2005/Sagara\\_AplicacionesDistribuidas3Capas.htm](http://www.elguille.info/colabora/NET2005/Sagara_AplicacionesDistribuidas3Capas.htm)

**[22]** Entidad: Expropia. Artículo: Introduction to Microsoft DNA. Understanding the 3 layers of an application. URL: [http://www.extropia.com/tutorials/dna/three\\_layers.html](http://www.extropia.com/tutorials/dna/three_layers.html)

**[23]** Entidad: Glosarium. Artículo: Diccionario Informático URL: Referencia: <http://www.glosarium.com/list/14/1,O,,xhtml>

**[24]** Entidad: Wikipedia. Artículo: Índice de masa corporal URL:

[http://es.wikipedia.org/wiki/%C3%8Dndice\\_de\\_masa\\_corporal](http://es.wikipedia.org/wiki/%C3%8Dndice_de_masa_corporal)

**[25]** Entidad: Mentores. Artículo: Programación orientada a objetos (POO) Parte 1 URL: <http://www.mentores.net/Default.aspx?tabid=104&site=67&parentid=26&type=art>

**[26]** Entidad: Added Value. Artículo: Discovering the Feeling: Applying Neuroscience to Marketing. URL: <http://www.added-value.com/AV/Default.aspx?tabid=98>

**[27]** Entidad: Gartner Group. Artículo: Discovering the Feeling: Decision Tools for Cost Management. URL: [www.gartner.com/4\\_decision\\_tools/measurement/decision\\_tools/tco/tco.html](http://www.gartner.com/4_decision_tools/measurement/decision_tools/tco/tco.html)

**[28]** Entidad: Solution Matrix. Artículo: Total Cost of Ownership Analysis (TCO). URL: <http://www.solutionmatrix.com/total-cost-of-ownership.html>

**[29]** Entidad: Cornell University. Artículo: Forget ROI: Measure "Return on Expectations". URL: [http://www.conceptsystems.com/papers/whitepapers/forget\\_roi/forget\\_roi.cfm](http://www.conceptsystems.com/papers/whitepapers/forget_roi/forget_roi.cfm)

**[30]** Entidad: Publidirecta. Artículo: Diccionario de Marketing URL: [http://www.publidirecta.com/dicc/diccionario\\_de\\_marketing\\_b.php](http://www.publidirecta.com/dicc/diccionario_de_marketing_b.php)

**[31]** Entidad: BlazerNetwork. Artículo: Windows Hosting

URL: <http://www.blazernetwork.com/WindowsHosting/index.html>

**[32]** Entidad: Instituto Tecnológico de La Paz. Artículo:

Diagramas de Entidad–Relación URL:

<http://www.itlp.edu.mx/publica/tutoriales/analisis/42.htm>

## **ANEXOS**

# **ANEXO 1**

DESCRIPCION DE PRODUCTOS	LABORATORIO	PARTICIPACION		EVOLUCION		PRECIO PROM
		2.006 MILES DE UNIDADES	2.006 US-DOLARES (miles)	2.006 MILES DE UNIDADES	2.006 US-DOLARES (miles)	2.006 US-DOLARES
EUGLUCON ROC	ROCHE	18,94	\$ 6,97	-9,81	\$ -8,26	\$ 9,28
EUGLUCON TABL 5 MG x 100		100	\$ 100,00	-9,81	\$ -8,26	\$ 9,28
GLUCOFAGE MCK	MERCK	13,28	\$ 13,28	22,49	\$ 24,26	\$ 14,09
GLUCOFAGE TABL E.C. 850 MG x 60		60,96	\$ 62,04	6,12	\$ 5,87	\$ 15,40
GLUCOFAGE TABL E.C. 500 MG x 60		31,63	\$ 27,77	37,26	\$ 36,49	\$ 13,28
GLUCOFAGE TABL E.C. 1000 MG x 30		6,81	\$ 9,22	493,56	\$ 502,89	\$ 10,24
GLUCOFAGE TABL XR 500 MG x 30		0,59	\$ 0,97	999	\$ 999,00	\$ 12,40
GLICEM OSM	OSMOPHARM	9,4	\$ 1,64	-8,44	\$ -8,66	\$ 4,41
GLICEM TABL 5 MG x 100		100	\$ 100,00	-8,44	\$ -8,66	\$ 4,41
BI-EUGLUCON M ROC	ROCHE	18	\$ 16,14	19,18	\$ 23,31	\$ 9,66
BI-EUGLUCON M TABL x 30		40,41	\$ 36,16	0,86	\$ -0,10	\$ 6,07
BI-EUGLUCON M TABL x 60		20,27	\$ 19,04	-9,35	\$ -7,62	\$ 12,73
BI-EUGLUCON M TABL M5 x 60		39,32	\$ 44,80	83,04	\$ 84,44	\$ 15,45
AMARYL S.A	SANOFI-AVENTIS	4,98	\$ 10,88	6,6	\$ 6,55	\$ 8,27
AMARYL TAB.RECU.RAN 2 MG x 15		48,92	\$ 35,42	7,63	\$ 7,87	\$ 5,98
AMARYL TAB.RECU.RAN 4 MG x 15		51,08	\$ 64,58	5,63	\$ 5,84	\$ 10,45
DAONIL S.A	SANOFI-AVENTIS	3,68	\$ 1,18	-9,97	\$ -12,27	\$ 2,26
DAONIL TABL 5 MG x 28		100	\$ 100,00	-9,97	\$ -12,27	\$ 2,26
GLUCOVANCE MCK	MERCK	11,36	\$ 10,30	41,71	\$ 48,39	\$ 6,86
GLUCOVANCE TABL RECUBI. 2.50 MG x 30 (/500)		44,21	\$ 40,35	19,51	\$ 20,61	\$ 6,26
GLUCOVANCE TABL RECUBI. 250 MG x 30 (/1.2)		15,89	\$ 11,79	11,21	\$ 11,24	\$ 5,08
GLUCOVANCE TABL RECUBI. 5 MG x 30 (/500)		39,9	\$ 47,87	106,89	\$ 105,07	\$ 8,23
DIAMICRON MR SVR	SERVIER	3,14	\$ 4,38	0,46	\$ -0,19	\$ 10,55
DIAMICRON MR TABL 30 MG x 30		100	\$ 100,00	0,46	\$ -0,19	\$ 10,55
LIBRAGLUCIL GRT	GRUNENTHAL	1,47	\$ 0,30	-33,52	\$ -36,29	\$ 1,04
LIBRAGLUCIL TABL RANUR. 5 MG x 20		100	\$ 100,00	-33,52	\$ -36,29	\$ 1,04
DIAMICRON SVR	SERVIER	1,23	\$ 1,56	-15,79	\$ -15,97	\$ 6,41

DIAMICRON TABL 80 MG x 20		100	\$ 100,00	-15,79	\$ -15,97	\$ 6,41
AVANDIA GSK	GLAXOSMITHKLINE	1,7	\$ 7,69	57,11	\$ 56,92	\$ 15,93
AVANDIA GRAG. 4 MG x 14		79,81	\$ 71,79	66,02	\$ 68,61	\$ 14,33
AVANDIA GRAG. 8 MG x 14		20,19	\$ 28,21	29,61	\$ 33,38	\$ 22,25
LIBRAFORMIN GRT	GRUNENTHAL	0,1	\$ 0,07	-84,31	\$ -84,16	\$ 3,52
LIBRAFORMIN TABL RECUBI. 500 MG x 20		100	\$ 100,00	-84,31	\$ -84,16	\$ 3,52
DIABINESE PFZ	PFIZER	0,74	\$ 0,32	-8,27	\$ 7,22	\$ 5,49
DIABINESE TABL 250 MG x 50		100	\$ 100,00	-8,18	\$ 7,32	\$ 5,49
DIABINESE TABL 250 MG x 250		0	\$ -	-100	\$ -100,00	\$ -
STARLIX NVR	NOVARTIS	0,32	\$ 0,53	-24,04	\$ -24,74	\$ 9,96
STARLIX TABL RECUBI. 120 MG x 24		100	\$ 100,00	-24,04	\$ -24,74	\$ 9,96
DIAFORMINA RMM	ROEMMERS	1,64	\$ 1,25	33,2	\$ 35,38	\$ 3,83
DIAFORMINA TABL RETARD 850 MG x 20		59,99	\$ 63,86	28,06	\$ 30,89	\$ 4,08
DIAFORMINA TABL RECUBI. 500 MG x 20		40,01	\$ 36,13	41,73	\$ 44,10	\$ 3,46
METFORAL MEN	MENARINI	1,05	\$ 0,87	-0,88	\$ 0,12	\$ 6,28
METFORAL TABL RECUBI. 850 MG x 30		100	\$ 100,00	-0,88	\$ 0,12	\$ 6,28
GLUCOBAY BAY	BAYER	0,58	\$ 0,82	32,89	\$ 27,68	\$ 10,72
GLUCOBAY TABL 50 MG x 30		61,42	\$ 48,90	49,81	\$ 48,12	\$ 8,53
GLUCOBAY TABL 100 MG x 30		38,58	\$ 51,10	12,63	\$ 12,79	\$ 14,20
GLIBENCLAMIDA MCK	MERCK	0,37	\$ 0,08	-30,98	\$ -30,45	\$ 1,61
GLIBENCLAMIDA TABL 5 MG x 30		100	\$ 100,00	-30,98	\$ -30,45	\$ 1,61
GLIBENCLAMIDA EQM	ECUAQUIMICA	1,3	\$ 0,16	67,79	\$ 52,01	\$ 0,91
GLIBENCLAMIDA TABL 5 MG x 30		100	\$ 100,00	67,79	\$ 52,01	\$ 0,91
GLIBENCLAMIDA GRT	GRUNENTHAL	0	\$ -	-99,9	\$ -99,91	\$ 1,00
GLIBENCLAMIDA TABL 5 MG x 20		100	\$ 100,00	-99,9	\$ -99,91	\$ 1,00
ACTOS ABT	ABBOTT	0,2	\$ 0,65	123,91	\$ 82,56	\$ 12,17
ACTOS TABL 15 MG x 15		100	\$ 100,00	123,91	\$ 82,56	\$ 12,17
NOVOLIN N HM N-N	ROCHE	0	\$ 0,09	-97,54	\$ -97,50	\$ 13,86
NOVOLIN N HM V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	-97,54	\$ -97,50	\$ 13,86
HUMULIN N LLY	ELI LILLY	0,07	\$ 3,87	69,46	\$ 57,34	\$ 14,51
HUMULIN N V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	69,46	\$ 57,34	\$ 14,51
INSUMAN N S.A	SANOFI-AVENTIS	0,01	\$ 0,22	-52,4	\$ -41,87	\$ 10,53

INSUMAN N CARTUCHO 100 IU 3 ML x 5 (/ML)		0	\$ -	-100	\$ -100,00	\$ -
INSUMAN N V.SC 100 IU 5 ML x 1 (/ML)		55,68	\$ 65,24	0,59	\$ -5,16	\$ 7,97
INSUMAN N OPTISET 100 IU x 5 (/3ML)		44,32	\$ 34,76	-69,65	\$ -63,22	\$ 26,67
GLIMEPIRIDA MK M.K	MK	0,07	\$ 0,09	-45,56	\$ -44,66	\$ 4,82
GLIMEPIRIDA MK TABL 4 MG x 15		64,16	\$ 77,12	-34,22	\$ -37,18	\$ 5,80
GLIMEPIRIDA MK TABL 2 MG x 15		35,84	\$ 22,88	-58,4	\$ -60,50	\$ 3,08
HUMULIN 30/70 LLY	ELI LILLY	0,03	\$ 1,91	102,43	\$ 80,29	\$ 14,26
HUMULIN 30/70 V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	102,43	\$ 80,29	\$ 14,26
NOVOLIN R HM N-N	ROCHE	0	\$ 0,04	-95,81	\$ -95,80	\$ 13,92
NOVOLIN R HM V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	-95,81	\$ -95,80	\$ 13,92
DIABESULF BI-	BIOINDUSTRIA	0,63	\$ 0,09	0,61	\$ 3,30	\$ 3,44
DIABESULF TABL 5 MG x 100		100	\$ 100,00	0,61	\$ 3,30	\$ 3,44
HUMULIN R LLY	ELI LILLY	0,02	\$ 0,99	56,78	\$ 47,89	\$ 15,11
HUMULIN R V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	56,78	\$ 47,89	\$ 15,11
INSUMAN R S.A	SANOFI-AVENTIS	0	\$ 0,09	-52,55	\$ -48,35	\$ 11,75
INSUMAN R CARTUCHO 100 IU 3 ML x 5 (/ML)		0	\$ -	-100	\$ -100,00	\$ -
INSUMAN R V.SC 100 IU 5 ML x 1 (/ML)		50,42	\$ 62,26	-22,78	\$ -20,26	\$ 8,76
INSUMAN R OPTISET 100 IU 3 ML x 5 (/ML)		49,58	\$ 37,75	-58,1	\$ -57,64	\$ 27,00
GLURENOR MEN		0	\$ -	0	\$ -	\$ -
GLURENOR TABL 30 MG x 40		0	\$ -	0	\$ -	\$ -
NOVOLIN 70/30 HM N-N	ROCHE	0	\$ 0,05	-97,18	\$ -97,18	\$ 13,66
NOVOLIN 70/30 HM V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	-97,18	\$ -97,18	\$ 13,66
AVANDAMET GSK	GLAXOSMITHKLINE	0,33	\$ 1,12	-64,04	\$ -62,55	\$ 11,96
AVANDAMET TABL RECUBI. 4 MG x 14 (/500)		62,17	\$ 75,45	-63,03	\$ -62,03	\$ 14,52
AVANDAMET TABL RECUBI. 2 MG x 14 (/500)		37,83	\$ 24,55	-65,57	\$ -64,08	\$ 7,76
LANTUS S.A	SANOFI-AVENTIS	0,03	\$ 3,91	34,4	\$ 36,19	\$ 36,43
LANTUS V.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	34,4	\$ 36,19	\$ 36,43
GLYZIDE JLP	JULPHARMA	1,12	\$ 0,25	-23,26	\$ -22,88	\$ 3,41
GLYZIDE TABL 80 MG x 60		100	\$ 100,00	-23,26	\$ -22,88	\$ 3,41
HIPOGLUCIN CM-	CHILE	0,64	\$ 0,56	-3,04	\$ -4,61	\$ 13,31
HIPOGLUCIN TABL 850 MG x 60		100	\$ 100,00	-3,04	\$ -4,61	\$ 13,31
INSULATARD HM ROC	ROCHE	0,06	\$ 1,81	631,97	\$ 571,82	\$ 14,45

INSULATARD HM V.SC 100 IU 10 ML x 1 (/ML)		46,21	\$ 72,61	684,44	\$ 655,53	\$ 12,93
INSULATARD HM PENFILL SC 100 IU 3 ML x 5 (/ML)		53,79	\$ 27,39	592,2	\$ 419,30	\$ 20,95
MIXTARD 30 HM ROC	ROCHE	0,03	\$ 0,81	519,2	\$ 414,15	\$ 15,95
MIXTARD 30 HM V.SC 100 IU 10 ML x 1 (/ML)		35,53	\$ 62,72	414,93	\$ 396,67	\$ 13,64
MIXTARD 30 HM PENFILL SC 100 IU 3 ML x 5 (/ML)		64,47	\$ 37,28	597,01	\$ 446,49	\$ 22,34
DIAFORMINA PLUS RMM	ROEMMERS	2	\$ 1,50	138,36	\$ 143,03	\$ 5,68
DIAFORMINA PLUS TABL 5 MG x 30 (/500)		56	\$ 60,09	154,87	\$ 153,64	\$ 6,09
DIAFORMINA PLUS TABL 2.50 MG x 30 (/500)		44	\$ 39,91	120,21	\$ 128,62	\$ 5,15
GLUCOCID PBD	PHARMABRAND	0,37	\$ 0,23	370,92	\$ 279,62	\$ 4,77
GLUCOCID TABL 850 MG x 30		100	\$ 100,00	370,92	\$ 279,62	\$ 4,77
BI-GLICEM OSM	OSMOPHARM	0,93	\$ 0,59	306,78	\$ 347,99	\$ 4,82
BI-GLICEM TABL 1.25 MG x 30 (/250)		16,15	\$ 13,85	54,46	\$ 54,71	\$ 4,13
BI-GLICEM TABL 2.50 MG x 30 (/500)		34,66	\$ 32,84	145,37	\$ 145,55	\$ 4,57
BI-GLICEM TABL 5 MG x 30 (/500)		49,19	\$ 53,32	999	\$ 999,00	\$ 5,23
GLIMEPAX S+W		0,14	\$ 0,22	286,02	\$ 364,54	\$ 6,16
GLIMEPAX TABL 4 MG x 15		55,37	\$ 69,29	326,93	\$ 373,84	\$ 7,71
GLIMEPAX TABL 2 MG x 15		44,63	\$ 30,71	245	\$ 344,76	\$ 4,24
NOVOLIN 70/30 ROC	ROCHE	0,01	\$ 0,64	999	\$ 999,00	\$ 12,96
NOVOLIN 70/30 A.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	999	\$ 999,00	\$ 12,96
NOVOLIN N ROC	ROCHE	0,02	\$ 0,95	999	\$ 999,00	\$ 12,78
NOVOLIN N A.SC 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	999	\$ 999,00	\$ 12,78
INSULINA HUMAN.RAP PIS	COMERCIOSA	0	\$ 0,12	999	\$ 999,00	\$ 12,76
INSULINA HUMAN.RAP AMP. 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	999	\$ 999,00	\$ 12,76
INSULINA N ISOFANA PIS		0	\$ 0,11	999	\$ 999,00	\$ 10,65
INSULINA N ISOFANA AMP. 100 IU 10 ML x 1 (/ML)		100	\$ 100,00	999	\$ 999,00	\$ 10,65
GLIBENCLAMIDA MAM	CORP. MAGMA	0,02	\$ -	999	\$ 999,00	\$ 0,90
GLIBENCLAMIDA TABL 5 MG x 30		100	\$ 100,00	999	\$ 999,00	\$ 0,90
NOVOLIN R ROC	ROCHE	0	\$ 0,16	999	\$ 999,00	\$ 10,62
NOVOLIN R A.SC 1000 IU 10 ML x 1		100	\$ 100,00	999	\$ 999,00	\$ 10,62
ACTRAPID ROC	ROCHE	0,02	\$ 0,51	193,7	\$ 135,28	\$ 15,09
ACTRAPID PENFILL SC 100 IU 3 ML x 5 (/ML)		60,8	\$ 35,19	273,51	\$ 199,11	\$ 22,43
ACTRAPID V.SC HM 100 IU 10 ML x 1 (/ML)		39,2	\$ 64,81	120,59	\$ 110,85	\$ 12,81

## **ANEXO 2**

## ENCUESTA

## INFORMACION GENERAL

Sexo	Masculino	<input type="checkbox"/>	Femenino	<input type="checkbox"/>
Nro. Cedula	<input type="text"/>			
Apellidos	<input type="text"/>			
Nombres	<input type="text"/>			
Correo Electronico	<input type="text"/>			
Edad	<input type="text"/>			
Peso (Kg)	<input type="text"/>			
Estatura (cm)	<input type="text"/>			
¿En que ciudad reside?	<input type="text"/>			
Actividad	Ama de casa	<input type="checkbox"/>	Estudiante	<input type="checkbox"/>
	Jubilado/Pensionado	<input type="checkbox"/>	Desocupado	<input type="checkbox"/>
	Trabaja	<input type="checkbox"/>	Otros	<input type="checkbox"/>
Nivel de Estudios	Sin estudios	<input type="checkbox"/>	Primario	<input type="checkbox"/>
	Secundario	<input type="checkbox"/>	Universitario	<input type="checkbox"/>
	Masterado/Doctorado	<input type="checkbox"/>		

## PERFIL DIABETICO

¿Cuantos embarazos ha tenido? Ingrese cero si nunca ha estado embarazada

¿Tiene Familiares con problemas cardiovasculares?	Si	<input type="checkbox"/>	No	<input type="checkbox"/>
¿Padece Diabetes?	Nunca me he realizado un examen de diabetes	<input type="checkbox"/>	Padezco Diabetes Tipo 1	<input type="checkbox"/>
	Padezco Diabetes Tipo 2	<input type="checkbox"/>	Padezco de Diabetes pero no se de que tipo	<input type="checkbox"/>
	Me diagnosticaron que no tengo Diabetes	<input type="checkbox"/>		

Si Usted Padece Diabetes, ¿a que edad se la Diagnosticaron?		<input type="text"/>
¿Tiene Familiares con Diabetes?	Si	<input type="text"/> No <input type="text"/>
Consumo Cigarrillos	Nunca he fumado	<input type="text"/> He Fumado Ocasionalmente <input type="text"/>
	Fuma actualmente y con regularidad	<input type="text"/>
Consumo Bebidas Alcoholicas	Nunca he consumido	<input type="text"/> Consumo en Ocasiones Especiales <input type="text"/>
	Consumo Regularmente	<input type="text"/>
Habitos de actividad Física	Poco o nada de actividad fisica semanalmente	<input type="text"/> Realiza ejercicio al menos 2 dias a la semana <input type="text"/>
	Realiza Ejercicio mas de 2 dias a la semana	<input type="text"/> Realiza Ejercicio todos los dias de la semana <input type="text"/>
<b>Hábitos de Consumo Para Tratamiento de la Diabetes</b>		
¿Consume Productos Especiales para diabéticos?		<input type="text"/>
	Consume Ocasionalmente	<input type="text"/> Consume Frecuentemente <input type="text"/>
	No consumo, no estoy interesado en consumir	<input type="text"/> No consumo, si estoy interesado en consumir <input type="text"/>
¿Consume aspirina?	Nunca la he consumido	<input type="text"/> Consume ocasionalmente <input type="text"/>
	Consume frecuentemente	<input type="text"/>
¿Consumo Medicinas para su tratamiento?	Si	<input type="text"/> No <input type="text"/>
¿Qué tipo de medicina consume? (Sólo conteste si es diabetico)		
	Ninguna	<input type="text"/> Inyectable <input type="text"/>
	Oral	<input type="text"/> Suministros <input type="text"/>

Compra habitualmente reactivos para la medición de niveles de glucosa? (Sólo conteste si es diabetico)

Si, compro regularmente las tiras de medición

No, no tengo el dispositivo

No, tengo el dispositivo pero no lo uso regularmente

○

## **ANEXO 3**

O

TABLA DE VALORES DE VARIABLES DE ENCUESTA			
Campo	Clase	Valor	Detalle
e_sexo	SEXO	1	MASCULINO
	SEXO	2	FEMENINO
e_dedad	EDAD DISCRETIZADA	0	NINO[0-12]
	EDAD DISCRETIZADA	1	ADOLESCENTE [12-18]
	EDAD DISCRETIZADA	2	ADULTO JOVEN [18-35]
	EDAD DISCRETIZADA	3	ADULTO MAYOR [35-65]
	EDAD DISCRETIZADA	4	TERCERDA EDAD [MAYOR DE 65]
e_dimc	INDICE DE MASA CORPORAL	0	BAJO PESO
	INDICE DE MASA CORPORAL	1	NORMAL
	INDICE DE MASA CORPORAL	2	SOBRE PESO
	INDICE DE MASA CORPORAL	3	OBESIDAD
e_ciudad	CIUDAD	1	AMBATO
	CIUDAD	2	AZOGUES
	CIUDAD	3	BABAHOYO
	CIUDAD	4	CUENCA
	CIUDAD	5	DURÁN
	CIUDAD	6	GUARANDA
	CIUDAD	7	GUAYAQUIL
	CIUDAD	8	IBARRA
	CIUDAD	9	QUITO
	CIUDAD	10	PORTOVIEJO
	CIUDAD	11	LATACUNGA
	CIUDAD	12	LOJA
	CIUDAD	13	MACHALA
	CIUDAD	14	MANTA
e_actividad	ACTIVIDAD	1	AMA DE CASA
	ACTIVIDAD	2	ESTUDIANTE
	ACTIVIDAD	3	JUBILADO/PENSIONADO
	ACTIVIDAD	4	DESOCUPADO
	ACTIVIDAD	5	TRABAJA
	ACTIVIDAD	6	OTROS
e_estudio	NIVEL DE ESTUDIOS	1	SIN ESTUDIOS
	NIVEL DE ESTUDIOS	2	PRIMARIO
	NIVEL DE ESTUDIOS	3	SECUNDARIO
	NIVEL DE ESTUDIOS	4	UNIVERSITARIO
	NIVEL DE ESTUDIOS	5	MASTERADO/DOCTORADO

e_cardio	Ant. CARDIOVASCULARES	1	NO
	Ant. CARDIOVASCULARES	2	SI
e_diabetes	PADECE DIABETES	1	NUNCA ME HE REALIZADO UN EXAMEN DE DIABETES
	PADECE DIABETES	2	PADEZCO DIABETES TIPO 1
	PADECE DIABETES	3	PADEZCO DIABETES TIPO 2
	PADECE DIABETES	4	PADEZCO DE DIABETES PERO NO SE DE QUE TIPO
	PADECE DIABETES	5	ME DIAGNOSTICARON QUE NO TENGO DIABETES
e_dedadidiabetes	EDAD DIAGNOSTICO DIABETES	0	NUNCA
	EDAD DIAGNOSTICO DIABETES	1	NINO[0-12]
	EDAD DIAGNOSTICO DIABETES	2	ADOLESCENTE [12-18]
	EDAD DIAGNOSTICO DIABETES	3	ADULTO JOVEN [18-35]
	EDAD DIAGNOSTICO DIABETES	4	ADULTO MAYOR [35-65]
	EDAD DIAGNOSTICO DIABETES	5	TERCERDA EDAD [MAYOR DE 65]
e_familiares	FAMILIARES CON DIABETES	1	NO
	FAMILIARES CON DIABETES	2	SI
e_cigarrillo	CONSUMO DE CIGARRILLO	1	NO FUMA
	CONSUMO DE CIGARRILLO	2	FUMA OCASIONALMENTE
	CONSUMO DE CIGARRILLO	3	FUMA REGULARMENTE
e_alcohol	CONSUMO DE ALCOHOL	1	NO CONSUME
	CONSUMO DE ALCOHOL	2	CONSUME OCASIONALMENTE
	CONSUMO DE ALCOHOL	3	CONSUME REGULARMENTE
e_fisica	ACTIVIDAD FISICA	1	POCO O NADA DE ACTIVIDAD FISICA SEMANALMENTE
	ACTIVIDAD FISICA	2	REALIZA EJERCICIO AL MENOS 2 DIAS A LA SEMANA
	ACTIVIDAD FISICA	3	REALIZA EJERCICIO MAS DE 2 DIAS A LA SEMANA
	ACTIVIDAD FISICA	4	REALIZA EJERCICIO TODOS LOS DIAS DE LA SEMANA
e_productos	CONSUME PRODUCTOS DIABETICOS	1	CONSUME OCASIONALMENTE
	CONSUME PRODUCTOS DIABETICOS	2	CONSUME FRECUENTEMENTE
	CONSUME PRODUCTOS DIABETICOS	3	NO CONSUMO, NO ESTOY INTERESADO EN CONSUMIR

	CONSUME PRODUCTOS DIABETICOS	4	NO CONSUMO, SI ESTOY INTERESADO EN CONSUMIR
e_aspirina	CONSUME ASPIRINA	1	NO CONSUME
	CONSUME ASPIRINA	2	CONSUME OCASIONALMENTE
	CONSUME ASPIRINA	3	CONSUME REGULARMENTE
e_medicina	CONSUME MEDICINA DIABETES	1	NO
	CONSUME MEDICINA DIABETES	2	SI
e_tmedicina	TIPO MEDICINA DIABETES	1	NINGUNA
	TIPO MEDICINA DIABETES	2	ORAL
	TIPO MEDICINA DIABETES	3	INYECTABLE
	TIPO MEDICINA DIABETES	4	INSUMOS
e_reactivo	COMPRA REACTIVOS	1	NO, NO TENGO EL DISPOSITIVO
	COMPRA REACTIVOS	2	NO, TENGO EL DISPOSITIVO PERO NO LO USO REGULARMENTE
	COMPRA REACTIVOS	3	SI, COMPRO REGULARMENTE LAS TIRAS DE MEDICIÓN

○

## **ANEXO 4**

## COMPORTAMIENTO DEL CONSUMIDOR

El comportamiento del consumidor ha sido investigado durante el inicio de la vida comercial, es decir, desde que existió el primer trueque de artículos, el expendedor buscó el interés del potencial comprador; así a través de los años múltiples metodologías se han implementado para el conocimiento de gustos, hábitos y actitudes de los consumidores, que permitan crear planes de venta y de comunicación, entiéndase publicidad, dirigidos y eficaces.

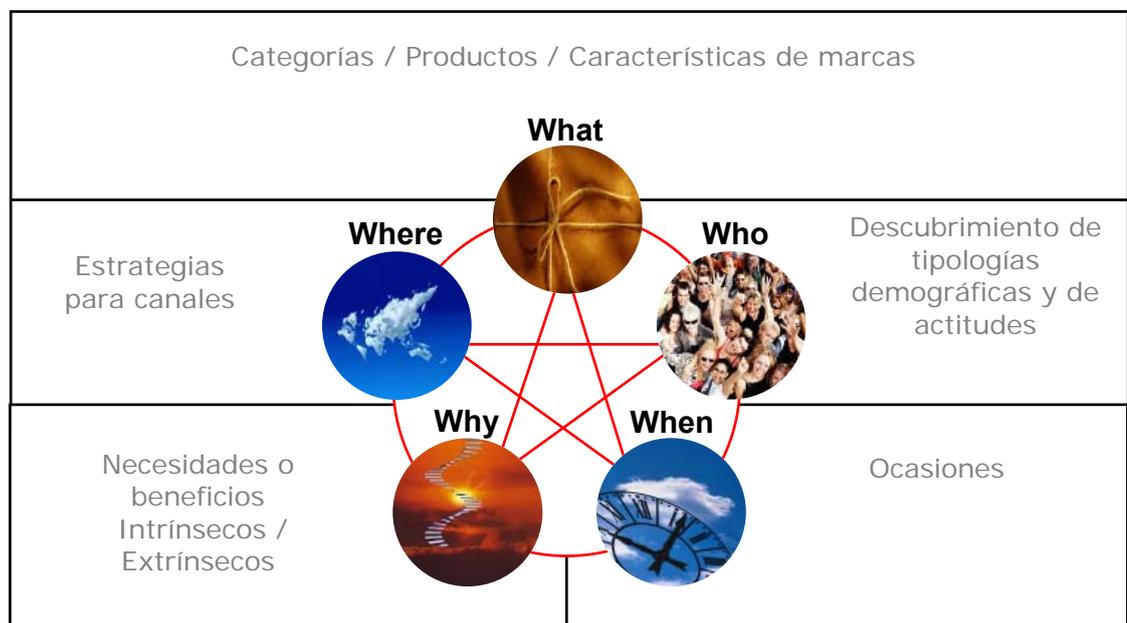
La metodología presentada en esta tesis se ancla en el estudio de una empresa multinacional denominada Added Value™ que usa más variables de las comunes para entender el comportamiento y actitudes hacia un producto o servicio.

Esta empresa multinacional está presente en 19 países y maneja marcas a nivel mundial tanto en investigación como en comunicación, entre las principales: Levi's, Porsche, Colgate – Palmolive, entre otras.

La metodología Market Mapping que es usada en esta tesis es sólo referencial y no constituye un estudio completo de hábitos. Tradicionalmente se clasifica a los consumidores por perfiles socioeconómicos y demográficos, pero esta metodología Market Mapping™ busca conocer más sobre el entorno y se hace 5 cuestionamientos, denominados las 5W's por su significado en inglés:

- Who? Quién consume? Perfiles Pen portraits: Edad, NSE, actitudes, etc.
- When? Cuándo? Ocasiones de consumo.
- Where? Dónde? Tipologías y canales de venta.

- What? Qué? Productos, empaques, contenido, marcas.
- Why? Por qué? Estados de necesidad.



## **ANEXO 5**

# EL FINANCIERO

PRIMER PERIÓDICO ECONÓMICO DEL ECUADOR

Tarifas On line (mensuales)		
Pagina principal	Exclusivo	Compartido
Banner superior (468 x 60 pix)	440	250
Mini banner inferior simple (90 x 60 pix)	125	n/a
Mini banner inferior Doble (180 x 60 pix)	225	n/a
Paginas interiores	Exclusivo	Compartido
Banner superior paginas economicas * (468 x 60 pix)	315	180
Banner interior otras paginas* (468 x 60 pix)	250	140
* Paginas Economicas: Dinero y Divisas, Valores y Productos, Reportes (Bancos, Financieras, Cooperativas y Seguros)		
** Otras Paginas: Produccion y Negocios, Marketing e Internet, Seminarios y Eventos		

## El Financiero On line

Este es un medio de ACTUALIZACION DIARIA mediante el cual ofrecemos a nuestros anunciantes información financiera y administrativa de interés, lo que nos convierte en un medio interactivo de consulta permanente.

En nuestra web usted puede anunciar su marca bajo las siguientes formas:

- Auspiciando una página o sección determinada
- Incluyendo un banner o viñeta en nuestra página principal

Servicio gratuitos que le ofrecemos para anuncios on line:

- Elaboracion del arte o banner
- Links hacia su página web

### Mayor informacion

#### En Guayaquil:

Shirley Mantilla, Gerente Nacional de Ventas  
[shirley.mantilla@gmail.com](mailto:shirley.mantilla@gmail.com) ;  
[shiman@telconet.net](mailto:shiman@telconet.net) Telfs.: 593-04-611000 /  
 593-04-09-9519437 /  
 09-4425112 / 09-6008612

#### En Quito:

Paulina Ramos, Gerente Regional de Ventas  
[finanmil@uio.satnet.net](mailto:finanmil@uio.satnet.net)  
 593-02-229007 / 008 / 009 / 593-02-09-  
 9163343