

**SEP**

SECRETARÍA DE  
EDUCACIÓN PÚBLICA



**TECNOLÓGICO NACIONAL DE MÉXICO**



**INSTITUTO TECNOLÓGICO DE CD. GUZMÁN**

**PROGRAMA DE MAESTRÍA EN CIENCIAS  
DE LA COMPUTACIÓN**

**TESIS**

**TEMA:**

**ANÁLISIS DE DATOS PARA LA TRAZABILIDAD DE LA MIEL**

**QUE PARA OBTENER EL GRADO DE:  
MAESTRO EN CIENCIAS DE LA COMPUTACIÓN**

**PRESENTA:**

**I.S.C. HÉCTOR RODRIGO SIERRA GUTIÉRREZ**

**DIRECTORAS:**

**DRA. ROSA MARÍA MICHEL NAVA  
M.C. CYNTHIA ALEJANDRA MARTÍNEZ PINTO  
DRA. FATIMA EZZAHRA HOUSNI**

**CD. GUZMÁN JALISCO, MÉXICO, AGOSTO DE 2018**

Cd. Guzmán, Jal. a 09/Agosto/2018

Oficio No. DEPI/58/18

ASUNTO : AUTORIZACIÓN DE IMPRESIÓN

**C. HÉCTOR RODRIGO SIERRA GUTIÉRREZ**  
**N.C. M16290036**

En cumplimiento con el documento normativo de las disposiciones para la operación de estudios de posgrado del Tecnológico Nacional de México y con base en la aprobación del Comité Tutorial comisionado para su revisión; la División de Estudios de Posgrado e Investigación le otorga la autorización de impresión de su trabajo de tesis intitulado:

**“ANÁLISIS DE DATOS PARA LA TRAZABILIDAD DE LA MIEL”**

dirigido por la **Dra. Rosa María Michel Nava**, desarrollado como requisito parcial para la obtención del grado de Maestro en Ciencias de la Computación, de acuerdo al plan de estudios MCOM-2011-05.

Sin otro asunto en particular, quedo de usted.

**ATENTAMENTE**



**DR. HUMBERTO BRACAMONTES DEL TORO**  
**JEFE DE DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN**



S.E.P. TecNM  
INSTITUTO TECNOLÓGICO  
DE CD. GUZMÁN  
DIVISIÓN DE ESTUDIOS  
DE POSGRADO E  
INVESTIGACIÓN

C.p. Archivo



## **Agradecimientos**

Agradezco a Dios por darme la oportunidad de concluir mis estudios de la Maestría en Ciencias de la Computación, por poner fortaleza, sabiduría e inteligencia en mí, para poder persistir en todos los momentos y nunca desistir.

Gracias al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico brindado en parte de mis estudios, que fueron vitales para la culminación de la maestría.

Dra. Rosa María Michel Nava, le agradezco el apoyo constante, los consejos, sus conocimientos, la motivación y el tiempo dedicado en sus asesorías para lograr terminar mis estudios de posgrado.

Agradezco a mis compañeros de la maestría, por hacer de esta etapa de mi vida, una etapa especial, divertida y sobre todo llena de buenos momentos.

Gracias a mi madre, la Sra. Elia, por sembrar en mí valores y dar lo mejor de ella para apoyar todos mis estudios previos; a mi hermana Gabriela por ser la meta a alcanzar y superar.

Pero, sobre todo, gracias a mi esposa Vane por sus paciencia, comprensión, amor y solidaridad con mis estudios, por el tiempo que me ha concedido, un tiempo que les he robado a la familia. A mi pequeña hija Regina Ailé, por ser el motor de mi vida, y quien me motiva por ser el mejor papá del universo y su mejor ejemplo a seguir. Sin su apoyo y motivación nunca habría terminado mis estudios de posgrado, por lo tanto, este trabajo es también el suyo.

## **Reconocimientos**

El presente proyecto de investigación, no se habría hecho posible sin el gran apoyo de mi asesora externa, la Dra. Fatima Ezzahra Housni, quien colaboró de manera directa en la recolección de información, que fueron los parámetros de entrada de la herramienta computacional generada, a la cual reitero el reconocimiento.

Así mismo, externo mi más grato reconocimiento a la M.C. María Eugenia Puga Nathal y la M.C. Cynthia Alejandra Martínez Pinto, quienes además de fungir como revisoras del documento, contribuyeron con ideas, alternativas y soluciones al problema de investigación, aportando de su tiempo, experiencias y conocimientos.

De igual manera, se reconoce el apoyo de las maestras M.C. María de Jesús Cárdenas Chávez y a la Dra. Gloria Estela Cárdenas Gómez, por su colaboración en el análisis del impacto económico y social, el cual forma parte de este documento de tesis.

Mi reconocimiento a la empresa Frutimielmex, la cual colaboró de manera directa en el proyecto de investigación, al proporcionar información muy relevante, que sirvió de base para el desarrollo de la herramienta computacional, con datos verídicos, actuales y de primera mano.

Y de manera particular, mi mayor reconocimiento a la Dra. Rosa María Michel Nava, directora de tesis del presente proyecto, quien además de cumplir al ciento por ciento su labor, aportó sus conocimientos, experiencia profesional, así como diferentes soluciones a cada uno de los problemas que se generaron durante el desarrollo del proyecto.

## **Resumen**

El presente trabajo de investigación está orientado al estado de Jalisco, en un área de trabajo de 15,552 kilómetros cuadrados y aborda la problemática dada a nivel mundial, la cual se refiere a la mortandad de las abejas.

Que debido a distintas situaciones ambientales y humanas, han perjudicado de manera sustancial a este grupo de fauna, ya que las recientes investigaciones han arrojado como resultado que “la fumigación con pesticidas y plaguicidas muy fuertes ha provocado la mortandad de las abejas” (Torres, 2017). Razón por la cual, con este proyecto se busca a través del uso de técnicas y métodos de análisis de datos, como Minería de Datos mediante la herramienta de árboles de decisión, predecir niveles de riesgo para las abejas, derivados de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, existente en las cercanía de los apiarios; obteniendo mediante una aplicación de software, el nivel de riesgo existente en la zona, así como un mapeo de los lugares que representen menor riesgo para las abejas, con la finalidad de contribuir a disminuir esta problemática que se está dando a nivel mundial y que ocasiona principalmente la baja de la producción, aumento en los costos productivos y por ende la reducción en las ventas de las empresas apícolas en el estado de Jalisco, impactando tanto económica, ecológica y socialmente.

Así mismo, se generó un software apícola que permite llevar el control de las actividades administrativas y de producción, encaminadas a buscar la certificación de la miel, a través de la trazabilidad de la misma.

Los resultados y publicaciones de este proyecto, brindan información sobre los niveles de riesgo categorizados como alto, medio o bajo, derivados de la predicción obtenida mediante el software; de tal manera que permitan determinar si es viable reubicar al apiario para reducir la mortandad de las abejas.

## **Índice general**

Índice de figuras .....	v
Índice de tablas .....	viii
Capítulo I. Introducción.....	1
Capítulo II. Fundamento teórico.....	12
Capítulo III. Marco Metodológico.....	32
Capítulo IV. Resultados.....	66
Capítulo V. Conclusiones .....	88
Fuentes de información .....	90
Glosario .....	94
Anexos.....	96

## Índice de figuras

<i>Figura 1.1</i> Zona de trabajo total del proyecto.....	2
<i>Figura 1.2</i> División de las áreas de toda la zona de trabajo.....	2
<i>Figura 2.1</i> Captura de pantalla de software APITECNIC obtenido de la página de internet <a href="http://magya.omixom.com/">http://magya.omixom.com/</a> .....	15
<i>Figura 2.2</i> Captura de pantalla de software TRACTUS obtenido de la página de internet <a href="http://tractus.es/miel.html">http://tractus.es/miel.html</a> .....	17
<i>Figura 2.3</i> Captura de pantalla de software APITECNIC obtenido de la página de internet <a href="http://tecnocustomer.powweb.com/apitecnic/index.php">http://tecnocustomer.powweb.com/apitecnic/index.php</a> .....	18
<i>Figura 2.4</i> Esquema general de algoritmo de árbol de decisión.....	25
<i>Figura 2.5</i> Ubicación de los Estados Unidos Mexicanos dentro del continente americano. ....	30
<i>Figura 2.6</i> Ubicación del estado de Jalisco dentro de la república mexicana.....	30
<i>Figura 2.7</i> Zona de trabajo establecida de 15,552 kilómetros cuadrados.....	31
<i>Figura 3.1</i> Flujo de proceso iterativo.....	38
<i>Figura 3.2</i> Flujo de proceso iterativo adaptado.....	38
<i>Figura 3.3</i> Diagrama Entidad-Relación del software apícola.....	41
<i>Figura 3.4</i> Diagrama de bloques del sistema apícola.....	43
<i>Figura 3.5</i> Organización de archivos Excel de las zonas de trabajo.....	48
<i>Figura 3.6</i> Extracto de información de una zona en archivo Excel.....	48
<i>Figura 3.7.</i> Mapa de identificación de vegetación.....	49
<i>Figura 3.8</i> Mapa de niveles de riesgo por zonas.....	51
<i>Figura 3.9</i> Concentrado de información por zonas y factores de riesgo.....	52
<i>Figura 3.10</i> Concentrado de información ponderado con niveles de riesgo.....	53
<i>Figura 3.11</i> Extracto de información de coordenadas y niveles de riesgo.....	53
<i>Figura 3.12</i> Estructura del archivo ARFF en Sublime Text.....	54
<i>Figura 3.13</i> Weka utilizando Bayes Net (Red Bayesiana).....	55
<i>Figura 3.14</i> Weka utilizando Naive Bayes (Bayes Genuino).....	55

<i>Figura 3.15</i> Weka utilizando AdaBoost.....	56
<i>Figura 3.16</i> Weka utilizando Decision Table (Tabla de decisiones) .....	56
<i>Figura 3.17</i> Weka utilizando Decision Tree (árboles de decisiones).....	57
<i>Figura 3.18</i> Modelo final del árbol de decisión .....	60
<i>Figura 3.19</i> Función para ubicar zona codificada en JavaScript .....	62
<i>Figura 3.20</i> Tabla porcentajeszonas.....	62
<i>Figura 3.21</i> Función para conectar con R .....	63
<i>Figura 3.22</i> Programa de predicción en entorno R .....	63
<i>Figura 3.23</i> Ejecución de código manualmente .....	64
<i>Figura 4.1</i> Comparativa de zonas de trabajo con porcentajes totales .....	68
<i>Figura 4.2</i> Comparativa de zonas de trabajo con porcentajes ponderados .....	69
<i>Figura 4.3</i> Niveles de riesgo por zona .....	70
<i>Figura 4.4</i> Zonas elegidas para el análisis de riesgo alto.....	71
<i>Figura 4.5</i> Gráfica de puntos zona 6 .....	72
<i>Figura 4.6</i> Gráfica de puntos zona 19 .....	73
<i>Figura 4.7</i> Gráfica de puntos zona 23 .....	74
<i>Figura 4.8</i> Punto de equilibrio en modo gráfico .....	77
<i>Figura 4.9</i> Inicio de sesión.....	78
<i>Figura 4.10</i> Menú principal.....	78
<i>Figura 4.11</i> Interfaz gestionar usuarios.....	79
<i>Figura 4.12</i> Interfaz códigos QR.....	80
<i>Figura 4.13</i> Interfaz respaldar/restaurar base de datos.....	80
<i>Figura 4.14</i> Interfaz principal del módulo predicciones .....	81
<i>Figura 4.15</i> Interfaz mostrar información, módulo predicciones .....	82
<i>Figura 4.16</i> Interfaz del proceso de generación de predicción, módulo predicciones .....	82
<i>Figura 4.17</i> Interfaz mostrando resultado de predicción de nivel de riesgo, módulo predicciones .....	83



<i>Figura 4.18</i> Interfaz principal del módulo de movilidad .....	84
<i>Figura 4.19</i> Interfaz desplegando información del apiario, del módulo de movilidad .....	85
<i>Figura 4.20</i> Interfaz de comparación de información de dos puntos, del módulo de movilidad .....	86
<i>Figura 4.21</i> Interfaz complemento de información, del módulo movilidad .....	86

## **Índice de tablas**

Tabla 3.1 Tabla de clasificación de niveles de riesgos para las abejas.....	50
Tabla 3.2 Resultados de comparativa entre algoritmos.....	58
Tabla 4.1 Resultado del análisis de riesgo de la zona 6.....	71
Tabla 4.2 Resultado del análisis de riesgo de la zona 19.....	72
Tabla 4.3 Resultado del análisis de riesgo de la zona 23.....	73
Tabla 4.4 Resultados con promedio ponderado.....	75
Tabla 4.5 Punto de equilibrio de micro productores de miel.....	76

# Capítulo I. Introducción

## 1.1 Descripción del trabajo de investigación

El presente proyecto impacta de forma general a los apicultores del estado de Jalisco, principalmente a los que cuentan con apiarios en la zona de trabajo establecida de 15,552 kilómetros cuadrados, que se encuentra bajo las coordenadas delimitadas por los puntos geográficos, indicados en grados decimales que se citan a continuación; punto 1: 20.570193°, -103.908013°, punto 2: 20.570193°, -102.986697°, punto 3: 19.052650°, -102.986697° y punto 4: 19.052650°, -103.908013°. Dichas zonas corresponden a los municipios de Acatlán de Juárez, Tala, Tlaquepaque, Tlajomulco de Zúñiga, Villa Corona, Zacoalco de Torres, Jocotepec, Ameca, San Martín Hidalgo, Cocula, Ixtlahuacán de los Membrillos, El Salto, Tonalá, Zapotlán del Rey, Juanacatlán, Zapotlanejo, Chapala, Ponciltán, Chiquilistlán, Tecolotlán, Atemajac de Brizuela, Techaluta de Montenegro, Teocuitatlán de Corona, Tuxcueca, Tonaya, Tapalpa, Sayula, Atoyac, Concepción de Buenos Aires, Tamazula de Gordiano, Gómez Farías, Valle de Juárez, Mazamitla, Manzanilla de la Paz, Tizapán el Alto, Tuxcacuesco, San Gabriel, Zapotlán El Grande, Tolimán, Zapotitlán de Vadillo, Zapotiltic, Tonila, Tuxpan, Tecalitlán, Santa María del Oro, Jilotlán de los Dolores y Pihuamo. Esta área de trabajo se dividió en 28 zonas de 24 por 24 kilómetros, con la finalidad de facilitar el análisis de la información y determinar la cantidad de hectáreas de cultivos de agricultura protegida, agricultura de riego (cultivo de aguacate) y agricultura de temporal, que fueron la base de la investigación de este proyecto. Cabe mencionar que la zona 25 no fue considerada, ya que incluye parte del territorio del estado de Colima, y el objeto de esta investigación está centrada solo al estado de Jalisco.

Con el uso del dispositivo GPS (Global Positioning System/Sistema de Posicionamiento Global) se obtuvieron las coordenadas de cada uno de los apiarios, y con la herramienta de representación cartográfica de ArcGIS, se obtuvieron imágenes satelitales provenientes del satélite LANDSAT, en las cuales se ubicaron las

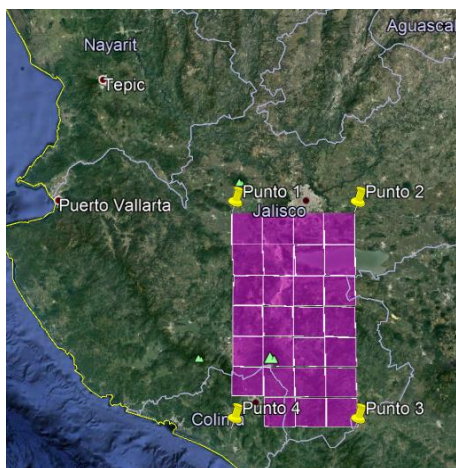
coordenadas previamente recabadas, además de la vegetación y cultivos antes mencionados, cuerpos de agua y zonas urbanas, existentes en cada una de las 27 zonas de trabajo establecidas de 576 kilómetros cuadrados.

En la figura 1.1 se muestra toda la zona de trabajo.



*Figura 1.1* Zona de trabajo total del proyecto.

En la figura 1.2 se muestra la división de las áreas en las que se dividió toda la zona de trabajo.



*Figura 1.2* División de las áreas de toda la zona de trabajo.

Posteriormente se procesó la información para obtener niveles de riesgo al posicionar el apiario en un nuevo punto dentro del área de trabajo de los 15,552 kilómetros cuadrados, teniendo así un panorama del nivel de riesgo que tendría el apiario en esa nueva ubicación, sin necesidad de moverlo físicamente.

Para el presente proyecto, se desarrollaron dos herramientas computacionales: la primera es un software apícola, que permite controlar los procesos administrativos y productivos de la empresa; y la segunda que está basada en el uso de minería de datos con la técnica de árboles de decisión, la cual a su vez está conformada por dos módulos.

Para el caso del software apícola, se diseñó una base de datos y un conjunto de interfaces, que permitieran al usuario llevar el control administrativo y de producción de la empresa Frutimielmex, del cual carecían, y que por tal motivo, era de suma importancia su desarrollo. El software a grandes rasgos está conformado por los módulos de: control de apiarios, ventas menudeo y barriles, rentas, control de abejas reinas, control de medicación y alimentación, reportes, usuarios y respaldo y recuperación de la base de datos.

Cabe mencionar que de manera específica se tuvo participación en el diseño del diagrama Entidad-Relación general del software, además de un seguimiento detallado del desarrollo de todo el sistema. Pero de manera particular, se desarrolló el diseño y la programación de los módulos de usuarios y de respaldo y recuperación de la base de datos.

La segunda herramienta computacional desarrollada, se divide en dos módulos.

En uno de ellos se obtiene como resultado el nivel de riesgo de un punto en específico (tomando las coordenadas geográficas en grados decimales: latitud y longitud), así como las características y número de hectáreas de los tipos de cultivos que rodean al punto y que fueron considerados como parte de la entrada de datos de entrenamiento del árbol de decisión. Posteriormente estas coordenadas son procesadas para determinar el riesgo y mostrar la información relacionada con este punto.

El segundo módulo tiene como finalidad permitir al usuario hacer el cambio de ubicación de determinado apiario (seleccionado de una lista). Posteriormente la información relativa a este apiario se muestra como referencia. Se elige un nuevo punto del cual se desea conocer la predicción del nivel de riesgo, para determinar si es conveniente mover al apiario a estas nuevas coordenadas. Una vez hecho el cambio de ubicación, se procede a actualizar la nueva información en la base de datos correspondiente.

La clasificación de los niveles de riesgo está basada en parámetros proporcionados por la empresa Frutimielmex, la cual brindó la información relacionada con los porcentajes de hectáreas considerados para ellos como riesgo alto, medio y bajo, de los cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, considerando la suma de porcentajes de cada tipo de factor en relación con el total de hectáreas que cuenta cada una de las zonas de trabajo. Para cada factor se le asignó una ponderación de riesgo, quedando de la siguiente manera:

- Cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles): 40%.
- Agricultura de riego (cultivo de aguacate): 40%.
- Agricultura de temporal de ciclo anual: 20%.

De tal manera que la clasificación quedó delimitada de la siguiente forma:

- Riesgo bajo: rango de 0.0% a 11.0%.
- Riesgo medio: rango de 11.1% a 27.0%.
- Riesgo alto: rango de 27.1% a 40.0%.

Para poder lograr dichos resultados, se usaron herramientas informáticas de minería y análisis de datos y lenguajes de programación, como R, programación web, MySQL,

Weka, PHP y JavaScript, las cuales se describen en el capítulo II Fundamento Teórico, numeral 2.2 marco teórico y 2.3 marco conceptual.

## **1.2 Definición del problema**

El cuidado de las abejas y la apicultura en general, es un tema de suma importancia actualmente a nivel mundial, por lo cual se requiere un alto grado de atención en esta actividad.

Las abejas, incluidas las abejas melíferas, abejorros y abejas solitarias, son el grupo prominente y económicamente más importante de polinizadores en todo el mundo. El 35% de la producción mundial de cultivos alimentarios depende de los polinizadores, lo que representa un valor anual de 153,000 millones de euros, (Blacquièrre, Smagghe, van Gestel & Mommaerts, 2012).

El investigador apícola de la Universidad de Guadalajara, Macías (2016), expuso que “el 70 por ciento de las frutas y hortalizas que existen en la tierra dependen de las abejas para su producción, si no existiera este insecto habría escasez de alimentos, porque si ellas no polinizan las flores, no hay frutas”.

Es ahí donde radica la importancia de la apicultura a nivel mundial, ya que sin ella existiría un descontrol en la cadena alimenticia, además de la falta de polinización de las flores, Macías (2016).

Los últimos estudios realizados en el estado de Jalisco han arrojado como resultado que la principal causa de muerte de las abejas es la vegetación contaminada con insecticidas, esto ha influido considerablemente en la productividad de la miel.

El presidente de la Asociación de Apicultores Unidos de Zapotlán El Grande, Hernández (2016), explicó textualmente “que por cambios climáticos, uso de pesticidas y otros factores, ha bajado la producción de miel en los últimos cinco años hasta un 40 por ciento, no solo en el municipio de Zapotlán El Grande, sino en todo México. Además mencionó que la apicultura se está transformando, ya que pasó de ser

el fuerte la producción de miel, hacia un mercado de servicio de polinización, por lo cual se buscó productores agrícolas conscientes del papel de la abeja en los procesos”.

Los neonicotinoides son actualmente la clase más utilizada de insecticidas en todo el mundo. Estos pesticidas son cada vez más frecuentes en entornos terrestres y acuáticos. Los neonicotinoides son absorbidos por las plantas y transportados a todos los órganos, incluidas las flores, contaminando así el polen y el néctar, así como cualquier fluido producido por la planta. Hay una creciente preocupación por el impacto de estos pesticidas sistémicos, no solo en organismos no objetivo, especialmente polinizadores como abejas melíferas y abejas silvestres, así como en otros invertebrados terrestres y acuáticos, sino también en vertebrados, incluidos los humanos. (Mitchell et al., 2017).

La crisis que está viviendo el sector por la muerte de las *colmenas* y por ende la baja en producción de miel hace que los precios de este producto, que el año pasado rondaba en 80 pesos el litro, este año se coloque en 120 pesos, Macías (2016).

Como una propuesta para combatir la problemática de la mortandad de las abejas, se desarrolla el presente proyecto que contribuye, mediante el análisis de datos con la identificación de las características de vegetación, zonas urbanas, cuerpos de agua y cultivos de aguacate, temporal y agricultura protegida, a generar una herramienta de predicción de niveles de riesgos en los diferentes puntos de la zona de trabajo, para así evaluar su riesgo de acuerdo a la cantidad de hectáreas de cada uno de los factores, causados por los productos químicos utilizados en los procesos de producción y post producción, que son una de las causas de muerte para las abejas, de tal manera que antes de mover físicamente el apiario a determinada zona, sea posible evaluar su riesgo a través de predicciones computacionales, con técnicas de minería de datos.

Así mismo, y de manera integral con la misma herramienta, se desarrolla el software apícola que permite controlar los procesos administrativos y productivos de la empresa, dada la carencia del sistema dentro de la misma, lo cual en su conjunto permitirán alcanzar la trazabilidad de la miel y por ende avanzar en la certificación de la misma.



### **1.3 Justificación**

La importancia de la investigación radica en aplicar técnicas de análisis de datos, a través de árboles de decisión, para la predicción de niveles de riesgo para las abejas, que contribuya a la identificación oportuna de los riesgos previo a realizar la reubicación física de los mismos, teniendo un panorama general de los riesgos como cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual.

Así mismo, con el software apícola, se contribuyó de manera precisa en lograr el control computarizado de los procesos administrativos y de producción que permiten alcanzar la trazabilidad de la miel, y por consecuencia, la certificación de la misma.

El impacto llegó más allá de la propia investigación, ya que actualmente existe una problemática a nivel mundial con los productores de miel y con las mismas abejas en sí, lo cual ha ocasionado un alto índice de mortandad en los apiarios, que además de producir pérdidas económicas a los productores, ha generado un descontrol ambiental debido a su importante labor dentro de la naturaleza.

Las abejas son esenciales para la agricultura, cumplen una función de ser indicadoras de la salud del medio ambiente y promueven la diversidad biológica, son, en pocas palabras, guardianes del ecosistema mundial, pero día a día enfrentan peligros que las acercan a la extinción, menciona el director general de la FAO (Food and Agriculture Organization of the United Nations / Organización de las Naciones Unidas para la Alimentación y la Agricultura) Graziano (2016).

Con esta investigación se logró ubicar de forma satelital cada uno de los apiarios de la empresa Frutimielmex de Zapotlán S.P.R. de R.L. de C.V. que se encuentran en el estado de Jalisco, además de generar la información del tipo de vegetación, cuerpos de agua, zonas de cultivo de aguacate, zonas de cultivos de agricultura protegida, entre otros, que se encuentran alrededor de los apiarios, y que han sido afectados de forma directa, para así establecer niveles de riesgo categorizados en alto, medio y bajo como se mencionó en el apartado de 1.1 “Descripción del trabajo de investigación”, que de

acuerdo a la cantidad de hectáreas de cultivo de agricultura protegida, cultivo de aguacate y cultivo de temporal establecidas en cada una de las zonas de trabajo, es su clasificación.

De igual manera, con la implementación del software apícola, se automatizaron los procesos de la empresa, en el ámbito administrativo y de producción, permitiendo así optimizar costos y tiempos, pero principalmente, simplificar el camino para lograr la trazabilidad de la miel.

La investigación fue viable, ya que se contó con los medios informáticos, humanos y de desarrollo tecnológico para aplicarlo en la generación del análisis de datos a través de minería de datos, mediante técnicas de árboles de decisión para la predicción de niveles de riesgo para las abejas, así como para el desarrollo del software apícola.

Además se tuvo la colaboración de la empresa Frutimielmex, dado que es prioridad para ellos y para los productores de miel del estado de Jalisco contar con una herramienta de predicción de niveles de riesgo.

## **1.4 Objetivos**

### **Objetivo general**

Aplicar análisis de datos para predecir niveles de riesgo en el ecosistema apícola, al cambiar de ubicación física derivados de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual.

### **Objetivos particulares o específicos**

Para que la empresa logre alcanzar el registro computacional de la trazabilidad de la miel, fue necesario desarrollar un sistema apícola que permita llevar el control de cada

uno de los procesos administrativos y de producción que se llevan a cabo en la empresa.

Por otro lado, para contribuir a la disminución de la mortandad de las abejas de la empresa, fue necesario desarrollar un módulo que permitiera hacer la detección de niveles de riesgo y posteriormente el registro de la movilidad de los apiarios. Para desarrollar este módulo fue necesario llevar a cabo lo siguiente:

- Recabar información de las ubicaciones de los apiarios en un área de trabajo de 15,552 kilómetros cuadrados en el estado de Jalisco.
- Identificar los tipos de vegetación, cultivos, zonas urbanas y cuerpos de agua que rodea a los apiarios.
- Identificar los factores de riesgo para la supervivencia de las abejas.
- Generar mapas que reflejen la ubicación, la vegetación y los factores de riesgo en torno a los apiarios.
- Investigar las características de las diferentes técnicas aplicables para el análisis y minería de datos.
- Definir la técnica más adecuada para obtener la predicción de las nuevas coordenadas de ubicación.
- Diseñar las diferentes interfaces de software para la aplicación del análisis de datos de las coordenadas.
- Desarrollar el software.
- Aplicar la técnica de árboles de decisión.
- Mostrar la información relacionada con la cantidad de hectáreas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, de la

zona donde se encuentra ubicado el apiario seleccionado para su posible reubicación.

- Obtener mediante el software, un mapeo general de los niveles de riesgos para las abejas, clasificándolos en nivel alto, medio y bajo.
- Mostar los nuevos índices de riesgo de la nueva ubicación de los apiarios, para su contrastación con las zonas anteriores.
- Procesar los resultados obtenidos.
- Contrastar los resultados entre las distintas ubicaciones.
- Difundir los resultados obtenidos.

## **1.5 Hipótesis**

Los niveles de riesgo en el ecosistema apícola, al cambiar de ubicación física derivados de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, se predicen aplicando análisis de datos.

## **1.6 Motivación**

La elección del presente trabajo de investigación radica principalmente en tomar el reto de aprender nuevas tecnologías y herramientas de minería de datos, de manera específica técnicas de predicción como árboles de decisión, lenguaje R, Weka, PHP, HTML, Java Script, entre otras, las cuales fueron aplicadas en la resolución de un problema real que aqueja a la rama apícola en general, y de manera específica dando atención a las necesidades de la empresa Frutimielmex de Zapotlán S.P.R. de R.L. de C.V. generando un panorama general de los niveles de riesgo de instalar o reubicar un apiario en determinado punto, mediante predicciones computacionales. Además del

desarrollo de una herramienta computacional que permite el control administrativo y de producción de la empresa, encaminado a la trazabilidad de la miel.

La problemática que actualmente existe en la rama apícola no sólo afecta en la parte productiva, post productiva y de comercialización de los productos y sub productos derivados de la miel, sino que va más allá de eso, ya que afecta directamente al medio ambiente, al formar parte las abejas del proceso tan importante de polinización de gran variedad de flora. Según información publicada por la FAO (2016), las abejas pueden tener un papel clave para mejorar la producción de unos 2,000 millones de pequeños campesinos en el mundo y asegurar los alimentos y la nutrición de la creciente población del planeta. La agencia de la ONU (Organización de las Naciones Unidas) citó un artículo publicado en la revista Science que sostiene que impulsar la producción agrícola aprovechando los procesos naturales es una de las vías sostenibles para aumentar el suministro de alimentos.

Dadas las razones anteriores y además, de manera precisa, la problemática local con la que cuenta la empresa Frutimielmex, la cual tiene apiarios en gran parte del estado de Jalisco en un área total de 15,552 kilómetros cuadrados, se buscó predecir a través del análisis de datos, mediante la minería de datos, haciendo uso de árboles de decisión, los niveles de riesgo al moverse sobre el área de trabajo establecida, para categorizarlos en riesgo alto, medio o bajo, de acuerdo a la cantidad de hectáreas de cultivos de agricultura protegida, cultivo de aguacate y agricultura de temporal de ciclo anual, con el apoyo de diferentes tecnologías de la información.

Así mismo, la elección del trabajo se basó en la afinidad de la forma de trabajo de la asesora y directora de tesis, la cual además maneja una línea de investigación afín a las características profesionales formadas en mi desarrollo académico a nivel licenciatura, aunado a la motivación de contar con un tema de investigación que me permita crecer profesionalmente y que además pueda tener una contribución a la posible resolución de un problema actual y real.

## Capítulo II. Fundamento teórico

### 2.1 Estado del arte

La apicultura en México tiene una gran importancia socioeconómica y ecológica, ya que es considerada como una de las principales actividades pecuarias generadora de ingresos. Según datos del Servicio de Información Agroalimentaria y Pesquera (SIAP, 2016) a nivel nacional en el 2016 se registró una producción de miel de 55,358 toneladas, generando un ingreso de \$2,278'810,000. En el estado de Jalisco se obtuvo una producción anual en el mismo 2016 de 1,438 toneladas.

De manera muy usual esta actividad se asocia únicamente con la producción de la miel, polen, jalea real, propóleos. Sin embargo, las abejas son fundamentales para un equilibrio del medio ambiente, ya que éstas al obtener el alimento de las flores fomentan en las plantas la capacidad de fecundarse. Muestra de ello es que en 2015, la Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) reportó que en México había 316 especies de plantas, de las cuales 286 se destinan para la alimentación y 80 como insumos para la industria, de los cuales el 80% depende de un polinizador para su producción.

Actualmente existe un problema muy grave con el alto índice de mortandad de las abejas, que debido a distintas situaciones ambientales y humanas, han perjudicado de manera sustancial a este grupo de fauna. El gran problema que tiene en jaque a los apicultores es el uso desmesurado e irresponsable de los plaguicidas en los campos agrícolas y su falta de regulación, según un estudio del Consejo Nacional de Ciencia y Tecnología (CONACYT) realizado en el año 2016. En estados como Chihuahua, Jalisco, Tamaulipas, San Luis Potosí, Michoacán y Yucatán, se reportaron la pérdida de 53 por ciento de las *colmenas*, (CONACYT, 2016).

Existen programas y grupos que promueven crear conciencia en la población, así como en los productores agrícolas y ganaderos, para frenar la mortandad de las mismas.

La problemática ha ocasionado principalmente la baja de la producción, aumento en los costos productivos y por ende la reducción significativa en las ventas de las empresas apícolas.

De acuerdo con Contreras (2015), investigadora del Centro Universitario del Sur, ubicado en Zapotlán El Grande, Jalisco, la mortandad de las abejas es un problema que se ha presentado desde hace más de una década, con mayor intensidad en ciudades de Europa y Asia, donde hay poblaciones de abejas que han disminuido hasta en un 85 por ciento. Así mismo, la Dra. Contreras, expone que “México, estaba de espectador, no obstante, en otoño de 2015 se comenzaron a conocer casos de mortalidad de abejas en algunas zonas del estado de entre 35% y hasta un 50%. Respecto al sur Jalisco, se estima que han muerto un 35% de *colmenas*, situación que es alarmante dado que Jalisco es el segundo estado productor de miel a nivel nacional y México es de los tres primeros productores que exportan a nivel mundial”.

En el estado de Jalisco, la académica explicó que las autoridades han estado dando diagnósticos a los apicultores quienes han tomado muestras de sus abejas para analizarlas, indicando que las abejas han muerto porque éstos no les dan un adecuado manejo. Reconoció que hasta ahora no se tienen identificada una sola causa para este problema, sino que la suma de factores, como el uso de agroquímicos, el cambio climático, enfermedades de las abejas y la generación de monocultivos.

En el ámbito del análisis de datos y de manera específica en el campo de la minería de datos, se han integrado de manera significativa el uso de estas tecnologías al campo agrícola y pecuario, mediante la inteligencia de negocios y las herramientas de análisis predictivo, lo cual genera factores decisivos en el desarrollo de las potencialidades competitivas de la empresa.

En la investigación realizada se encontró que se ha estado trabajando en aplicaciones relacionadas con el análisis de datos en el campo de la minería de datos y predicción de datos, en el rubro agrícola-pecuario, las cuales están orientadas al uso de los datos climáticos para realizar predicciones meteorológicas ajustadas según territorios y tipos de actividad, para la optimización de los costos de la cadena de suministro según las

condiciones del entorno, el arraigamiento de las medidas preventivas que permitan deducir las incidencias y los costos derivados de plagas, entre otros usos. Enseguida se enlistan algunos programas encontrados, así como sus características, enfocados a la incorporación de análisis y minería de datos con sus diferentes herramientas.

### **Programa de Red de Estaciones Meteorológicas**

Es una iniciativa originada en Argentina que se despliega por la provincia de Córdoba con el objetivo de brindar datos de utilidad para la toma de decisiones de los productores agropecuarios, pero también para toda la comunidad. A través de este programa, se intenta colaborar y facilitar la información que resulta de carácter estratégico para el trabajo en el campo, tanto en lo netamente productivo como en lo referido a aplicaciones periurbanas, entre otras utilidades. La información que brindan las estaciones se sube a un servidor en Internet, lo cual permite que los usuarios puedan acceder a los datos utilizando cualquier dispositivo que tenga conectividad, como un celular, laptop o computadora de escritorio. Una vez en el servidor, los datos son presentados en un software que permite visualizar datos actuales, graficar, generar reportes o realizar cálculos que sirven para la elaboración de predicciones meteorológicas, estudios climáticos ajustados a las variables fenológicas de los cultivos para determinar la oportunidad de la aplicación terrestre o aérea de fitosanitarios, entre otros usos. El acceso a los datos es instantáneo, gratuito y sin clave de acceso. En la figura 2.1 se aprecia una captura de pantalla del programa antes mencionado:



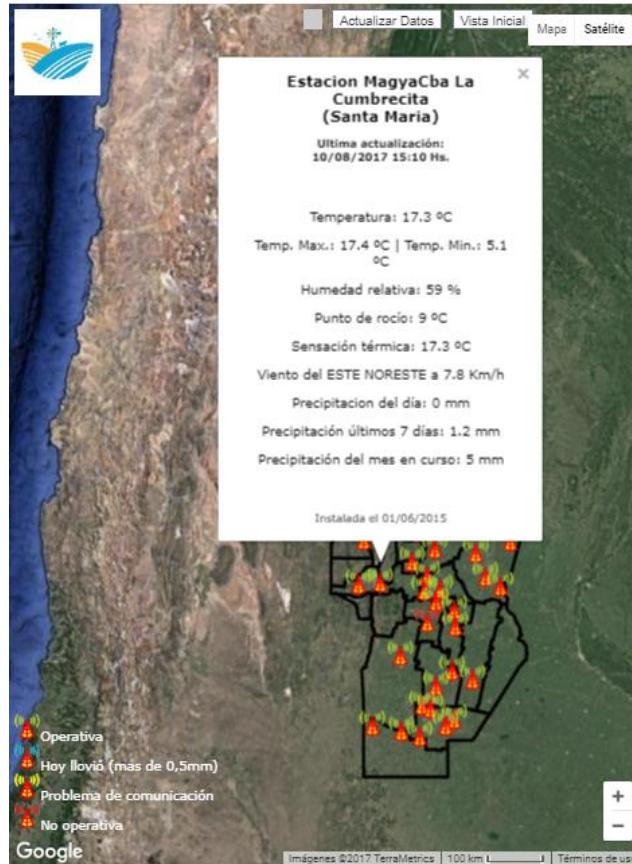


Figura 2.1 Captura de pantalla de software RED METEOROLÓGICA PROVINCIAL obtenido de la página de internet <http://magya.omixom.com/>

## ApisRAM

Es un modelo detallado de colonia de abejas en fase de desarrollo para EFSA (European Food Safety Authority / Autoridad Europea de Seguridad Alimentaria). Las oportunidades que ofrece este modelo y el enfoque utilizado para desarrollarlo son: integrar impactos de múltiples factores en la salud de las abejas; simular las interacciones entre los componentes de la abeja melífera y sus ambientes; y para predecir la dinámica futura del sistema en condiciones cambiadas de la gestión agrícola, apícola y el ambiente biológico y físico. En la medida en que estas oportunidades se pueden completar, están limitadas por aspectos de disponibilidad y gestión de datos. Los impactos de los factores estresantes o el manejo en las colonias

de abejas son difíciles de predecir. Se caracterizan por interacciones de cadena larga y retroalimentaciones no lineales.

Este sistema es capaz de integrar los impactos de múltiples estresores (químicos, biológicos y ambientales), simular interacciones entre componentes (entre las abejas y el ambiente) y predecir sistemas dinámicos complejos (cambios de temperatura y disponibilidad de recursos).

El éxito de ApisRAM está restringido por los datos, esto es, que sin ellos no se puede obtener una simulación real. Estas restricciones de los datos están dadas por: disponibilidad y acceso, recopilación, colaboración y gestión, análisis y comunicación.

En el ámbito concreto aplicado a la predicción de nivel de riesgos en el estado de Jalisco, hasta el momento no se encontró en la investigación aplicaciones o herramientas informáticas que implementen técnicas de minería de datos aplicados a dichos procesos.

En el mercado actual del software y aplicaciones informáticas, existen diversas herramientas que ayudan a mejorar el control de cada una de las fases de los procesos productivos y post productivos apícolas. Cabe mencionar que ninguno de ellos ofrece al usuario la opción de predicción de niveles de riesgo para las abejas. A continuación se describen algunos de ellos:

### **Tractus**

Es una aplicación informática que ayuda a cumplir con la normativa vigente en materia de gestión alimentaria. Consta de dos módulos; uno de producción y otro de comercialización, los cuales son personalizables y que se pueden contratar conjuntamente o por separado, según las necesidades de la empresa. El software está disponible para su compra en la página de internet <http://tractus.es/miel.html>. En la figura 2.2 se muestra una captura de pantalla de la aplicación informática Tractus.



Figura 2.2 Captura de pantalla de software TRACTUS obtenido de la página de internet <http://tractus.es/miel.html>

## Apitecnic

Es un software de control integral apícola, el cual contempla a grandes rasgos módulos de control de producción como lo son apiarios, alimentación, almacén, localización, entre otros. Está disponible para su compra en la página de internet <http://tecnocustomer.powweb.com/apitecnic/index.php>. En la figura 2.3 se ilustra una captura de pantalla del software denominado Apitecnic.



Figura 2.3 Captura de pantalla de software APITECNIC obtenido de la página de internet <http://tecnocustomer.powweb.com/apitecnic/index.php>

Después de realizada la investigación, no se ha encontrado en el ámbito del análisis de datos y herramientas de minería de datos, algo similar que permita predecir los niveles de riesgo para las abejas, por lo cual se dice que es una propuesta innovadora, originada de manera inicial para la empresa Frutimielmex de Zapotlán S.P.R. de R.L. de C.V. ubicada en el municipio de Zapotlán el Grande, Jalisco.

## 2.2 Marco teórico

La CONABIO reporta que hay 316 especies de plantas de las cuales 286 se destinan para la alimentación y 80 como insumos para la industria, de los cuales el 80% depende de un polinizador para su producción. Por su parte, investigadores de la UNAM identificaron 345 especies de plantas comestibles aprovechadas donde el 86% dependen de la polinización y estimaron el valor de la polinización en 43 mil millones de pesos.

Estos datos dan la certeza de la gran importancia que tienen las abejas en la conservación de la flora y la fauna, y por tal motivo es relevante implementar acciones que permitan dicha conservación.

Según la Coordinación General de Ganadería (CGG), en su publicación NOTIABJEA 2015-1, menciona que la especie apícola ha sufrido grandes pérdidas en los últimos 15 años, ya que la cantidad registrada de colonias ha estado disminuyendo a nivel mundial, y al cual se le ha denominado como “síndrome del colapso de las colonias”. La tasa de mortalidad de las abejas es de alrededor del 30% cada año desde 2007. La comunidad científica no ha encontrado a ciencia cierta las causas de muerte, pero las investigaciones mencionan una combinación de factores como la presencia de parásitos principalmente *Varroa*, nuevos agentes patógenos, el cambio climático y la exposición a pesticidas usados a gran escala en la agricultura.

Para el desarrollo del presente proyecto, fue indispensable el análisis y tratamiento de la información antes mencionada, mediante herramientas computacionales, que facilitaran su manejo. Para ello, fue necesaria la investigación y comparación de algunas herramientas de predicción que utiliza minería de datos como son, Bayes Net (Red Bayesiana), Naive Bayes (Bayes Genuino), AdaBoost, Decision Table (Tabla de decisiones) y Decision Tree (árboles de decisiones), las cuales se describen a continuación:

- Bayes Net (Red Bayesiana): según López (2012), las redes bayesianas son herramientas de modelado estadístico destinadas a representar un conjunto de incertidumbres relacionadas. Su estructura gráfica y su fundamento probabilístico las hace apropiadas para modelar sistemas multivariados orientados a la clasificación, el diagnóstico y la toma de decisiones. Las redes bayesianas son herramientas de modelado estadístico destinadas a representar un conjunto de incertidumbres relacionadas. Su estructura gráfica y su fundamento probabilístico las hace apropiadas para modelar sistemas multivariados orientados a la clasificación, el diagnóstico y la toma de decisiones.
- Naive Bayes (Bayes Genuino): es un clasificador probabilístico. Se basa en modelos de probabilidad que incorporan fuertes suposiciones de independencia. Los supuestos de independencia a menudo no tienen un impacto en la realidad.

Por lo tanto, se los considera ingenuos. Puede derivar modelos de probabilidad usando el teorema de Bayes (acreditado a Thomas Bayes). Dependiendo de la naturaleza del modelo de probabilidad, se puede entrenar el algoritmo de Naive Bayes en un entorno de aprendizaje supervisado. ("IBM Knowledge Center", 2018).

- AdaBoost: el algoritmo AdaBoost propone entrenar iterativamente una serie de clasificadores base, de tal modo que cada nuevo clasificador preste mayor atención a los datos clasificados erróneamente por los clasificadores anteriores, y combinarlos de tal modo que se obtenga un clasificador con elevadas prestaciones. Para ello, durante una serie de iteraciones entrena un clasificador que implementa una función asignándole un peso de salida, y lo añade al conjunto de modo que la salida global del sistema se obtenga como combinación lineal ponderada de todos los clasificadores base. ("Métodos de Boosting | Grupo de Gestión y Procesamiento de la Información", 2018).
- Decision Table (Tabla de decisiones): una tabla de decisiones es una entrada de lógica de reglas planificadas, en formato de tabla, que se compone de condiciones, representadas en las cabeceras de columna y fila, y acciones, representadas como puntos de intersección de los casos condicionales de la tabla. Las tablas de decisiones son especialmente idóneas para las reglas de negocio que tienen varias condiciones. Añadir otra condición es tan fácil como añadir otra fila o columna. Al igual que el conjunto de reglas if/then (si/entonces), la tabla de decisiones es controlada por la interacción de condiciones y acciones. La diferencia principal radica en que en una tabla de decisiones, la acción se decide a través de más de una condición, y se puede asociar más de una acción con cada conjunto condiciones. Si se cumplen las condiciones, se realiza la acción o acciones correspondientes. ("IBM Knowledge Center", 2018).

Y por último, Decision Tree (árbol de decisión), la cual fue la herramienta elegida para el desarrollo del proyecto, y que fue estudiada en todos sus sentidos.

Un árbol de decisión es una estructura de diagrama de flujo en la que el *nodo* interno representa la prueba de un atributo, cada rama representa el resultado de la prueba y cada *nodo* de *hoja* representa la etiqueta de clase. Un camino de raíz a *hoja* representa las reglas de clasificación. En análisis de decisión, un árbol de decisión y el diagrama de influencia estrechamente relacionado, se utiliza como una herramienta de apoyo a la decisión visual y analítica, para los valores esperados (Ran & Wang, 2015).

Como se menciona en "IBM Knowledge Center" (2018), hay varios algoritmos disponibles para realizar un análisis de segmentación y clasificación. Todos estos algoritmos son básicamente similares: examinan todos los campos del conjunto de datos para detectar el que proporciona la mejor clasificación o predicción dividiendo los datos en subgrupos. El proceso se aplica de forma recursiva, dividiendo los subgrupos en unidades cada vez más pequeñas hasta completar el árbol (según se definan determinados criterios de parada). Los campos objetivo y de entrada utilizados en la generación del árbol pueden ser continuos (rango numérico) o categóricos, dependiendo del algoritmo que se utilice. Si se usa un objetivo continuo, se genera un árbol de regresión; si se usa un objetivo categórico, se genera un árbol de clasificación.

Algunos ejemplos de algoritmos para generar árboles de decisión son:

- El algoritmo C4.5, genera un árbol de decisión a partir de los datos mediante particiones realizadas recursivamente. El árbol se construye mediante la estrategia de profundidad-primero (depth-first). El algoritmo considera todas las pruebas posibles que pueden dividir el conjunto de datos y selecciona la prueba que resulta en la mayor ganancia de información. Para cada atributo discreto, se considera una prueba con  $n$  resultados, siendo  $n$  el número de valores posibles que puede tomar el atributo. Para cada atributo continuo, se realiza una prueba binaria sobre cada uno de los valores que toma el atributo en los datos. En cada nodo, el sistema debe decidir cuál prueba escoge para dividir los datos.
- El *nodo* de árbol de clasificación y regresión (C&R), es un método de predicción y clasificación basado en árboles. este método utiliza la partición

reiterada para dividir los registros de entrenamiento en segmentos con valores de campo de salida similares. El nodo C&RT comienza por realizar un examen de los campos de entrada para buscar la mejor división, que se ha medido mediante la reducción del índice de impureza resultado de la división. La división define dos subgrupos, que se siguen dividiendo en otros dos subgrupos sucesivamente hasta que se activa un criterio de parada. Todas las divisiones son binarias (solamente se crean dos subgrupos).

- El *nodo* CHAID o detección automática de interacciones mediante chi-cuadrado (del inglés Chi-squared Automatic Interaction Detection), es un método de clasificación para generar árboles de decisión mediante estadísticos de chi-cuadrado para identificar divisiones óptimas. CHAID examina en primer lugar las tablas de tabulación cruzada entre los campos de entrada y los resultados para, a continuación, comprobar la significación mediante una comprobación de independencia de chi-cuadrado. Si varias de estas relaciones son estadísticamente importantes, CHAID seleccionará el campo de entrada de mayor relevancia (el valor P más pequeño). Si una entrada cuenta con más de dos categorías, se compararán estas categorías y se contraerán las que no presenten diferencias en los resultados. Para ello, se unirá el par de categorías que presenten menor diferencia, y así sucesivamente. Este proceso de fusión de categorías se detiene cuando todas las categorías restantes difieren entre sí en el nivel de comprobación especificado. En el caso de campos de entrada nominales, pueden fundirse todas las categorías. Sin embargo, en los conjuntos ordinales, únicamente podrán fundirse las categorías contiguas.
- El *nodo* QUEST, o árbol estadístico eficiente insesgado y rápido, es un método de clasificación binario para generar árboles de decisión. Una de las principales motivaciones para su desarrollo ha sido la reducción del tiempo de procesamiento necesario para los análisis de C&RT de gran tamaño con varias variables o varios casos. Un segundo objetivo de QUEST consiste en reducir la tendencia de los métodos de clasificación de árboles para favorecer a las entradas que permiten realizar más divisiones, es decir, los campos de entrada



continuos (rango numérico) o los correspondientes a varias categorías. QUEST utiliza una secuencia de reglas basada en comprobaciones de significación para evaluar los campos de entrada de un nodo. A efectos de selección, únicamente deberá realizar una sola comprobación en las distintas entradas de un nodo.

- El *nodo* Tree-AS, es similar al nodo CHAID existente; sin embargo, el nodo Tree-AS se ha designado para procesar grandes cantidades de datos (Big Data) para crear un solo árbol y mostrar el modelo resultante en el visor de salida que se ha añadido en SPSS Modeler versión 17. El nodo genera un árbol de decisiones utilizando estadísticas de chi-cuadrado (CHAID) para identificar divisiones opcionales. Este uso de CHAID puede generar árboles no binarios, lo que significa que algunas divisiones generarán más de dos ramas. Los campos de entrada y objetivo pueden ser continuos (rango numérico) o categóricos. CHAID exhaustivo es una modificación de CHAID que examina con mayor precisión todas las divisiones posibles, aunque necesita más tiempo para realizar los cálculos.
- El *nodo* árboles aleatorios, es similar al nodo C&RT existente; sin embargo, el nodo Árboles aleatorios se ha diseñado para procesar grandes cantidades de datos (Big Data) para crear un solo árbol y mostrar el modelo resultante en el visor de la salida que se ha añadido en SPSS Modeler versión 17. El nodo Árboles aleatorios genera un árbol de decisión que se utiliza para predecir o clasificar observaciones futuras. El método utiliza la partición reiterada para dividir los registros de entrenamiento en segmentos minimizando las impurezas en cada paso, donde un nodo se considera puro si el 100% de los casos del nodo corresponden a una categoría específica del campo objetivo. Los campos de entrada y objetivo pueden ser continuos (rango numérico) o categóricos (nominal, ordinal o marca). Todas las divisiones son binarias (sólo se crean dos subgrupos).
- El *nodo* C5.0, es el algoritmo elegido, y del cual se hará referencia a detalle en los siguientes párrafos.

De acuerdo a la literatura revisada y el tipo de problema de investigación realizada, se optó por utilizar el algoritmo C5.0, el cual divide la muestra en función del campo que ofrece la máxima ganancia de información. Las distintas sub muestras definidas por la primera división se vuelven a dividir, por lo general basándose en otro campo, y el proceso se repite hasta que resulta imposible dividir las sub muestras de nuevo. Por último se vuelven a examinar las divisiones del nivel inferior, y se eliminan o podan las que no contribuyen significativamente con el valor del modelo. Un árbol de decisión es una descripción sencilla de las divisiones que se han encontrado en el algoritmo. Los distintos *nodos* terminales (o "de *hoja*") describen un subconjunto de datos de entrenamiento, y cada uno de los casos incluidos en los datos de entrenamiento pertenece exactamente a un *nodo* terminal del árbol. En otras palabras, es posible realizar exactamente una predicción para cada registro de datos específico presente en un árbol de decisión.

Para entrenar un modelo C5.0, debe existir un campo categórico objetivo (por ejemplo, nominal u ordinal) y uno o más campos "Entrada" de cualquier tipo. Se ignorarán los campos establecidos en "Ambos" o "Ninguno". Los tipos de los campos utilizados en el modelo deben estar completamente instanciados. También se puede especificar un campo de ponderación. Los modelos C5.0 son bastante más robustos cuando aparecen problemas como datos perdidos y un número elevado de campos de entrada. Por lo general no precisan de largos tiempos de entrenamiento para calcular las estimaciones. Además, los modelos C5.0 suelen ser más fáciles de comprender que algunos tipos de modelos, ya que la interpretación de las reglas derivadas del modelo es muy directa. C5.0 también ofrece el eficaz método del aumento para obtener una mayor precisión en tareas de clasificación.

Según Herrera-Sements & Gago-Alonso (2015), existen varios algoritmos para obtener árboles de decisión, entre los cuales se encuentra el C5.0. Dicho algoritmo sigue el esquema general del algoritmo Árbol-Decisión que se muestra en la figura 2.4.

**Entrada:**  $D_t$  - conjunto de entrenamiento  
**Salida:**  $R$  - conjunto de reglas encontradas

```

1  $R \leftarrow \emptyset$ ;
2  $T \leftarrow \emptyset$ ; //conjunto de árboles
3  $t \leftarrow \text{Generar-Árbol-Decisión}(D_t)$ ;
4  $T \leftarrow \text{Podar-Árbol-Decisión}(t)$ ;
5  $R \leftarrow \text{Optimizar}(T, R)$ ;
6 return  $R$ ;

```

Figura 2.4 Esquema general de algoritmo de árbol de decisión.

En la siguiente sección se describen los conceptos, herramientas y lenguajes de programación relacionados con Árboles de Decisión, que fue la teoría principalmente utilizada en este proyecto.

## 2.3 Marco conceptual

Para el desarrollo del trabajo de investigación, fue de suma importancia adquirir conocimientos y datos referentes a las abejas, así como las tecnologías a utilizar para la generación del software. Por tal motivo, a continuación se abordan los temas que se tocaron para llevar a cabo y a buen término la presente investigación.

### Recomendaciones para la instalación de los apiarios

Según el Manual Básico Apícola publicado en 2015 por la Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA) a través de la CGG, recomienda los siguientes puntos para la instalación de los apiarios:

- No agrupar en un mismo apiario más de 30 *colmenas*.
- Ubicar el apiario cerca de donde exista abundancia de flores, ya que de ellas depende la producción de miel y polen. Las abejas dominan una zona de 2 a 3 kilómetros, sin embargo entre más cerca se encuentren, menor será el desgaste de energía de las abejas y tendrán un mayor rendimiento.

- La *colmena* debe estar orientada de manera que el sol dé a la *piquera* de la misma.
- Evitar lugares con mucha humedad.
- En zonas de calor, ubicarlas en zonas con sombra, pero no total.
- El terreno debe ser preferentemente plano; si es un cerro deberá estar al pie del mismo, para evitar desgaste innecesario de las abejas.
- Los apiarios deben estar situados a una distancia mínima de 3 kilómetros entre uno y otro.
- La separación entre las *colmenas* dentro del apiario debe ser de 1.5 metros entre una y otra.
- El apiario debe situarse como mínimo a 300 metros de distancia de viviendas, vías públicas y paso de animales.

### **Movilidad de los apiarios**

Consiste en mover la ubicación física de los mismos hacia otro lugar, de acuerdo a los factores de riesgo mencionados anteriormente, buscando aquellas zonas en las que existe un menor nivel de riesgo, que permita la sobrevivencia de las abejas en ese entorno.

### **Minería de datos**

Según Patel & Upadhyay (2012), es la extracción de información predictiva oculta en grandes bases de datos. Es una nueva y poderosa tecnología con un gran potencial para ayudar a las empresas a concentrarse en la información más importante en sus almacenes de datos. La minería de datos es en realidad parte del proceso de descubrimiento de conocimiento.

El proceso de descubrimiento de conocimiento en bases de datos, consta de algunos pasos que van desde colecciones de datos en bruto, hasta alguna forma de conocimiento nuevo.

La clasificación en la minería de datos encuentra un modelo para el atributo de clase en función de los valores de otros atributos. El clasificador genera descripciones significativas para cada clase que se utilizan para clasificar instancias del conjunto de datos dado.

Hay varios enfoques para la clasificación: red neuronal, SVM (Support Vector Machines/ Máquina de Soporte Vectorial), árbol de decisión, entre otros; siendo este último el utilizado.

### **Árboles de decisión**

La metodología de árbol de decisión es un método de minería de datos, comúnmente utilizado para establecer sistemas de clasificación basados en múltiples covariables o para desarrollar algoritmos de predicción para una variable objetivo. Este método clasifica una población en segmentos tipo rama, que construyen un árbol invertido con un *nodo* raíz, *nodos* internos y *nodos* hoja. El algoritmo no es paramétrico y puede tratar de manera eficiente conjuntos de datos grandes y complicados sin imponer una estructura paramétrica complicada. Cuando el tamaño de la muestra es lo suficientemente grande, los datos de estudio se pueden dividir en conjuntos de datos de entrenamiento y conjunto de datos de validación. Usando el conjunto de datos de entrenamiento para construir un modelo de árbol de decisión y un conjunto de datos de validación para decidir sobre el tamaño de árbol apropiado requerido para lograr el modelo final óptimo (Song, 2015).

### **R language**

Es un entorno de software libre para la informática estadística y los gráficos. Se compila y ejecuta en una amplia variedad de plataformas como UNIX, Windows y MacOS.

R language proporciona modelos estadísticos como los lineales y no lineales, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupación, entre otros, así como técnicas gráficas ("R: The R Project for Statistical Computing", 2018).

### **Weka**

Es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos se pueden aplicar directamente a un conjunto de datos o llamar desde su propio código Java. Weka contiene herramientas para el pre procesamiento de datos, clasificación, *regresión*, *clustering*, reglas de asociación y visualización. También es adecuado para desarrollar nuevos esquemas de aprendizaje automático. Es un software de código abierto emitido bajo la Licencia Pública General de GNU ("Weka 3 - Data Mining with Open Source Machine Learning Software in Java", 2018).

### **PHP (Hypertext Preprocessor)**

Es un lenguaje de código abierto, especialmente adecuado para el desarrollo web y que puede ser incrustado en HTML. PHP está enfocado principalmente a la programación de scripts del lado del servidor, por lo que se puede hacer cualquier cosa que pueda hacer otro programa CGI (Common Gateway Interface/Interfaz de Entrada Común), como recopilar datos de formularios, generar páginas con contenidos dinámicos, o enviar y recibir *cookies*. Existen principalmente tres campos principales donde se usan scripts de PHP: scripts del lado del servidor, scripts desde la línea de comandos y escribir aplicaciones de escritorio.

PHP puede emplearse en todos los sistemas operativos principales, incluyendo Linux, muchas variantes de Unix (incluyendo HP-UX, Solaris y OpenBSD), Microsoft Windows, Mac OS X, RISC OS, entre otros ("PHP: Manual de PHP - Manual", 2018).

## **JavaScript**

Es un lenguaje de programación que permite realizar actividades complejas en una página web. Permite crear contenido nuevo y dinámico, controlar archivos de multimedia, crear imágenes animadas y entre otras cosas más ("JavaScript", 2018).

## **MySQL**

Es un sistema de gestión de bases de datos relacional de código abierto basado en lenguaje de consulta estructurado SQL (Structured Query Language/Lenguaje de Consulta Estructurado), que se ejecuta en la mayoría de los sistemas operativos y que permite migrar los datos de un sistema a otro sin mayor dificultad ("MySQL | La base de datos de código abierto más popular | Oracle México", 2018).

## **2.4 Marco contextual**

El proyecto de investigación se llevó a cabo en la república Mexicana, en el estado de Jalisco, en un área establecida de trabajo de 15,552 kilómetros cuadrados.

La república mexicana, cuyo nombre oficial es Estados Unidos Mexicanos, se encuentra ubicado en el continente americano. Tiene una superficie de 1'984,375 km<sup>2</sup> y una población de 112.4 millones de habitantes (INEGI, 2010) distribuida en 31 estados federales y el Distrito Federal (Ciudad de México). A continuación se muestra en la figura 2.5 el mapa de la ubicación de México.



Figura 2.5 Ubicación de los Estados Unidos Mexicanos dentro del continente americano.

La macro localización del proyecto se encuentra en el estado de Jalisco. Su capital es Guadalajara. Está ubicado en la región oeste del país, limitando al norte con Nayarit, Durango, Zacatecas y Aguascalientes, al noreste con San Luis Potosí, al este con Guanajuato, al sur con Michoacán y Colima, y al oeste con el océano Pacífico. Tiene 7'844,830 habitantes (INEGI, 2015) y cuenta con una extensión territorial de 78,599 km<sup>2</sup>. A continuación se muestra en la figura 2.6 la ubicación del estado de Jalisco.



Figura 2.6 Ubicación del estado de Jalisco dentro de la república mexicana.



La micro localización del proyecto es como se había mencionado antes, en una zona de trabajo establecida de 15,552 kilómetros cuadrados que se encuentra delimitadas por los siguientes puntos de coordenadas geográficas en grados decimales que se citan a continuación; punto 1: 20.570193°, -103.908013°, punto 2: 20.570193°, -102.986697°, punto 3: 19.052650°, -102.986697° y punto 4: 19.052650°, -103.908013°.

En la figura 2.7 se muestra toda la zona de trabajo.



*Figura 2.7* Zona de trabajo establecida de 15,552 kilómetros cuadrados.

El foco de atención del proyecto está orientado hacia la ubicación de los apiarios de la empresa Frutimielmex de Zapotlán S.P.R. de R.L. de C.V. con sede en el municipio de Zapotlán El Grande, Jalisco, la cual ha colaborado de manera activa en la investigación, compartiendo información y datos técnicos.

## Capítulo III. Marco Metodológico

### 3.1 Tipo de Investigación

De acuerdo al tipo de problema que se resolvió con este trabajo, se llevó a cabo una investigación aplicada. Ya que mediante el uso de la tecnología computacional, se desarrolló un sistema de control apícola para que la empresa pudiera llevar a cabo sus registros de trazabilidad de la miel, y específicamente con el análisis y minería de datos, se abordó una problemática mundial y de gran importancia, como lo es el riesgo que viven las abejas con su ubicación en zonas rodeadas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual.

La población está conformada por los apiarios del estado de Jalisco, que son el foco de atención de la investigación. La muestra contemplada serán los apiarios de la empresa Frutimielmex, que se encuentra en dicho estado de la república Mexicana.

Con respecto a las técnicas y herramientas de estadísticas utilizadas en análisis y minería de datos, la que se utilizó en esta investigación es la conocida como árboles de decisión, ya que se tendrá una gran variedad de datos referidos a los factores de riesgos definidos que rodea a las diferentes coordenadas generadas por el software en cuestión, que fue necesario para analizar y evaluar los resultados que se arrojan siguiendo las distintas opciones, para poder predecir el nivel de riesgo en determinado punto, para así el usuario determine si decide reubicar los apiarios.

De acuerdo al tipo de problema que se logró resolver con el trabajo, se llevó a cabo una investigación aplicada. Para Murillo (2008), la investigación aplicada recibe el nombre de “investigación práctica o empírica”, que se caracteriza porque busca la aplicación o utilización de los conocimientos adquiridos, a la vez que se adquieren otros, después de implementar y sistematizar la práctica basada en investigación. El uso del conocimiento y los resultados de investigación que da como resultado una forma rigurosa, organizada y sistemática de conocer la realidad. Es decir, es aquella que se

realiza para la adquisición de nuevos conocimientos, dirigida hacia un objetivo o fin práctico, que responda a una demanda específica y determinada, y que en el campo que se está aplicando, que en este caso específico se busca a través de tecnología resolver el problema.

La investigación aplicada está fundamentada en los resultados de la investigación básica.

Carvajal (2002) refiere que los resultados de la investigación aplicada son considerados jurídicamente como derechos y en el marco nacional e internacional asumen la forma de propiedad intelectual, bajo sus diversas modalidades: patente, marca, modelo de réplica, dibujo industrial.

La población está conformada por los apiarios, que son el foco de atención de la investigación. Así mismo se hicieron pruebas mediante la movilidad de los apiarios, esto con base la analítica de la información recabada previamente. La muestra contemplada para la investigación fueron los apiarios del estado de Jalisco, ubicados en la zona delimitada en el capítulo I de este documento, pertenecientes a la empresa Frutimielmex.

El procedimiento a seguir para llevar a cabo la investigación a buen término se basó en lo siguiente:

- Investigación previa. En la que se recabó la información de los requerimientos que se tendrían en el sistema de control apícola, así como de las coordenadas de ubicación de los apiarios de la empresa caso de estudio; además de los tipos de vegetación alrededor de los apiarios, cuerpos de agua, zonas urbanas, cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, para identificar los factores de riesgo para las abejas, de tal manera que se puedan generar mapas que reflejen toda esta información.
- Análisis. Se investigaron las características de diferentes leguajes de programación que fueran aptos para el desarrollo del sistema de control apícola,

además de las diferentes técnicas existentes para análisis y minería de datos, para definir la que más se adecuó al proyecto: árboles de decisión.

- **Diseño.** Se diseñaron tanto las interfaces del software de control apícola en el que se lleva el registro de la trazabilidad de la miel, como del que obtiene la predicción de los niveles de riesgo para las abejas, basándose en las coordenadas y los factores definidos previamente.
- **Desarrollo.** Por una parte se implementó el software de control apícola que está conformado por diferentes módulos para el seguimiento de todos los procesos de trazabilidad de la miel, y por la otra se implementó el software donde se aplicó la técnica de árboles de decisión, para obtener, mostrar y contrastar los niveles de riesgo para la posible reubicación física de los apiarios.
- **Prueba.** Se puso a prueba todo el software desarrollado para procesar los resultados obtenidos y verificar su correcto funcionamiento.
- **Documentación.** Se elaboró un informe final del proyecto, el impacto económico y algunos artículos científicos en los que se difundan los resultados obtenidos.

### **3.2 Universo, población o unidades de análisis**

El universo de esta investigación está conformado por todos los apicultores del estado de Jalisco.

### **3.3 Criterios de inclusión/exclusión**

Los criterios de inclusión de la presente investigación, están basados en el hecho que los apiarios pertenezcan a la empresa Frutimielmex.

### **3.4 Muestreo**

La investigación utiliza un muestreo no probabilístico, ya que se orientó a los apiarios de la empresa Frutimielmex.

De acuerdo con Hernández, Baptista & Fernández (2010), en las muestras no probabilísticas, la elección de los elementos no depende de la probabilidad, sino de causas relacionadas con las características de la investigación o de quien hace la muestra. Aquí el procedimiento no es mecánico ni con base en fórmulas de probabilidad, sino que depende del proceso de toma de decisiones de un investigador o de un grupo de investigadores y, desde luego, las muestras seleccionadas obedecen a otros criterios de investigación.

### **3.5 Muestra**

La muestra de la investigación, son los apiarios ubicados dentro de la zona de trabajo de 15,552 kilómetros cuadrados en el estado de Jalisco, que se encuentra bajo las coordenadas delimitadas por los puntos de coordenadas geográficas grados decimales que se citan a continuación; punto 1: 20.570193°, -103.908013°, punto 2: 20.570193°, -102.986697°, punto 3: 19.052650°, -102.986697° y punto 4: 19.052650°, -103.908013°, pertenecientes a la empresa Frutimielmex.

### **3.6 Instrumentos**

Para la recolección de datos se requiere de instrumentos de medición, las cuales deben representar verdaderamente la(s) variable(s) de investigación, y que deben tener cumplir con validez, confiabilidad y objetividad (Hernández et al., 2010).

Para la presente investigación, se utilizaron diferentes instrumentos de recolección de datos, entre los cuales destaca la entrevista, que mediante la ayuda de un cuestionario básico de preguntas se procedió a entablar la conversación de forma personal y

presencial en las instalaciones de la empresa Frutimielmex, en donde existía una retroalimentación entre cada una de las preguntas, entre el entrevistador y el entrevistado.

Así mismo, el personal administrativo proporcionó formatos y datos referentes a la operación de la empresa, que fueron necesarios e indispensables para el desarrollo del proyecto. De igual manera, ofrecieron un recorrido por las instalaciones, dando la información de cada uno de los procesos productivos y administrativos.

### **3.7 Aparatos**

Para llevar a buen término la investigación, fueron requeridos aparatos electrónicos y físicos, los cuales son:

- Computadora portátil.
- GPS.
- Equipo de impresión.
- Medios de almacenamiento portátil como USB y discos duros.
- Dispositivos móviles.

### **3.8 Procedimientos**

En los siguientes párrafos, dentro de los apartados 3.8.1 y 3.8.2, se describe el procedimiento utilizado en el presente proyecto de investigación, explicando a detalle cada uno de los pasos que se llevaron a cabo para el desarrollo del software de control apícola, y el de movilidad y predicciones.

### 3.8.1 Software de control apícola

El software de control apícola fue desarrollado para plataforma Web, haciendo uso de base de datos MySQL, y con el apoyo de los lenguajes de programación JavaScript, HTML y PHP, así como las herramientas de diseño de diagramas DIA y StarUML. De igual manera, se generó un módulo para móvil, programado para el sistema operativo Android, el cual se conecta con la base de datos MySQL, y posteriormente se visualiza la información en la plataforma Web.

Para llevar a cabo el software apícola, dada la naturaleza del mismo, se tomó la decisión de trabajar con una metodología híbrida, integrando el marco de trabajo SCRUM con un flujo de proceso iterativo.

Una estructura general para la ingeniería de software define cinco actividades estructurales:

- Comunicación.
- Planeación.
- Modelado.
- Construcción.
- Despliegue.

El flujo de proceso es el cómo estas actividades están organizadas, y existen varios tipos de flujo, que describen la interacción de estas. En la figura 3.1 (Pressman, Campos Olguín, Enríquez Brito, Villegas Quezada & Ferro Castro, 2010) se observa el flujo de proceso iterativo, en el cual se repite una o algunas de las actividades antes de pasar a la siguiente.

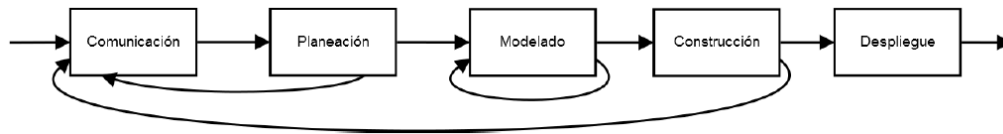


Figura 3.1 Flujo de proceso iterativo

SCRUM, por otra parte, es un marco de trabajo que fue creada por Ken Schwaber y Jeff Sutherland. Dicho marco se basa en la teoría del control de procesos empírica, es decir, que con este tipo de control se toman decisiones con base en el conocimiento que procede de la experiencia. Así mismo, SCRUM emplea un enfoque iterativo e incremental para reducir el riesgo de fallas.

Tomando en cuenta esta información, se llegó a la conclusión de tomar ambas formas de trabajo para ofrecer un resultado óptimo, quedando estructurada de la siguiente forma: se hicieron reuniones semanales para revisar los avances, así como establecer nuevas metas, y siguiendo el flujo de proceso iterativo, para facilitar la organización de las actividades, así como la integridad del módulo.

Para este proyecto, como se mencionó antes, se tomaron estas formas de trabajo para desarrollarlo. Sin embargo, se decidió remplazar la etapa de Construcción por las etapas de Codificación y Pruebas, por cuestión de organización, quedando entonces de como se muestra en la figura 3.2.

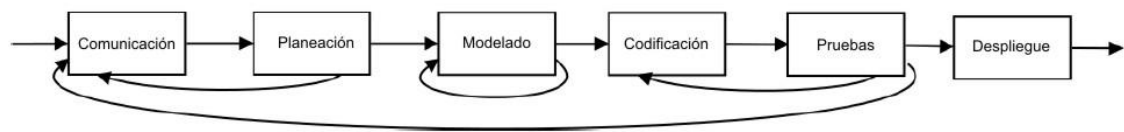


Figura 3.2 Flujo de proceso iterativo adaptado



A continuación, se explica y se detallan cada una de las etapas del flujo de proceso iterativo, así como las actividades realizadas en cada una de ellas.

### **3.8.1.1 Comunicación**

La actividad estructural de comunicación, consiste en hablar y colaborar con el cliente, entender los objetivos de los participantes en el proyecto y reunir requerimientos para poder definir las funciones y características del software.

Para tal fin, se requirió una amplia colaboración con la empresa Frutimielmex, ya que se necesitó una gran cantidad de la información para entender lo que la empresa requería. Hubo un total de cuatro reuniones: la primera fue para obtener los requerimientos generales, y las otras tres para recabar las especificaciones necesarias para desarrollar el software. Además, en una de las visitas ya antes mencionadas, se dio un recorrido por la empresa, para observar los procesos productivos y administrativos que se llevan a cabo dentro de la misma.

### **3.8.1.2 Planeación**

La planeación, dentro de las actividades estructurales, es donde se crea el plan del proyecto de software, y donde se plasman las tareas, riesgos y recursos, así como el producto y la programación del mismo.

Para la planeación de este módulo, así como las siguientes actividades estructurales, se organizaron reuniones semanales, en las cuales se establecieron las siguientes actividades a realizar:

- Definir los requerimientos del software, tomando en consideración todo lo que la empresa necesitaba para su software, así como sus limitantes en cuestión de infraestructura y la naturaleza propia de las tareas que debía realizar el software una vez terminado.
- Diseño de interfaces iniciales, las cuales sólo se trabajaron como un *Mockup*, para mostrar de forma gráfica como quedaría el software una vez terminado.

- Determinación de los tipos de lenguajes y manejadores a utilizar. Tomando en cuenta que el proyecto fue desarrollado por módulos, cada uno de ellos por varios alumnos residentes del Instituto Tecnológico de Ciudad Guzmán, fue necesario tomar decisiones para que el trabajo en equipo fuera el correcto, evitando así inconsistencias en cualquiera de las fases posteriores. Así mismo, tomar decisiones respecto a la forma de programar y las herramientas para realizar el desarrollo, con el fin de entregar un resultado óptimo.
- Determinar el alcance de cada módulo, ya que el proyecto está integrado por varios módulos y desarrollado por diferentes personas, y que en conjunto integran un mismo software, tanto en su parte Web como en su parte móvil, por lo cual fue de vital importancia marcar los límites de cada módulo.

### **3.8.1.3 Modelado**

El modelado, como parte de las actividades estructurales, es de suma importancia, ya que es en donde se crean modelos, a fin de plasmar de una manera fácil de entender, los requerimientos y el diseño del software. Y para hacerlo de manera eficiente, se utilizó UML (Unified Modeling Language/Lenguaje Unificado de Modelado). UML cumple con la función de informar de manera práctica a cualquier persona que esté involucrada en el diseño, codificación, pruebas y mantenimiento del software, independientemente de la metodología de desarrollo, todos los aspectos del software de una manera fácil de comprender.

Dentro del modelado, se desarrolla la parte lógica, en la cual se realiza la descripción del diseño de la base de datos, apoyándose en las estructuras de datos que un SGBD (Sistema Gestor de Bases de Datos) procesa. Para llevar a cabo esto, se requiere un diagrama Entidad-Relación que permita especificar la función de cada elemento que lo conforma, las relaciones existentes entre ellos, sus limitantes de mapeo, así como las características de los datos que se almacenarán en cada una de ellos. En la figura 3.3 se muestra el diagrama Entidad-Relación del sistema apícola.

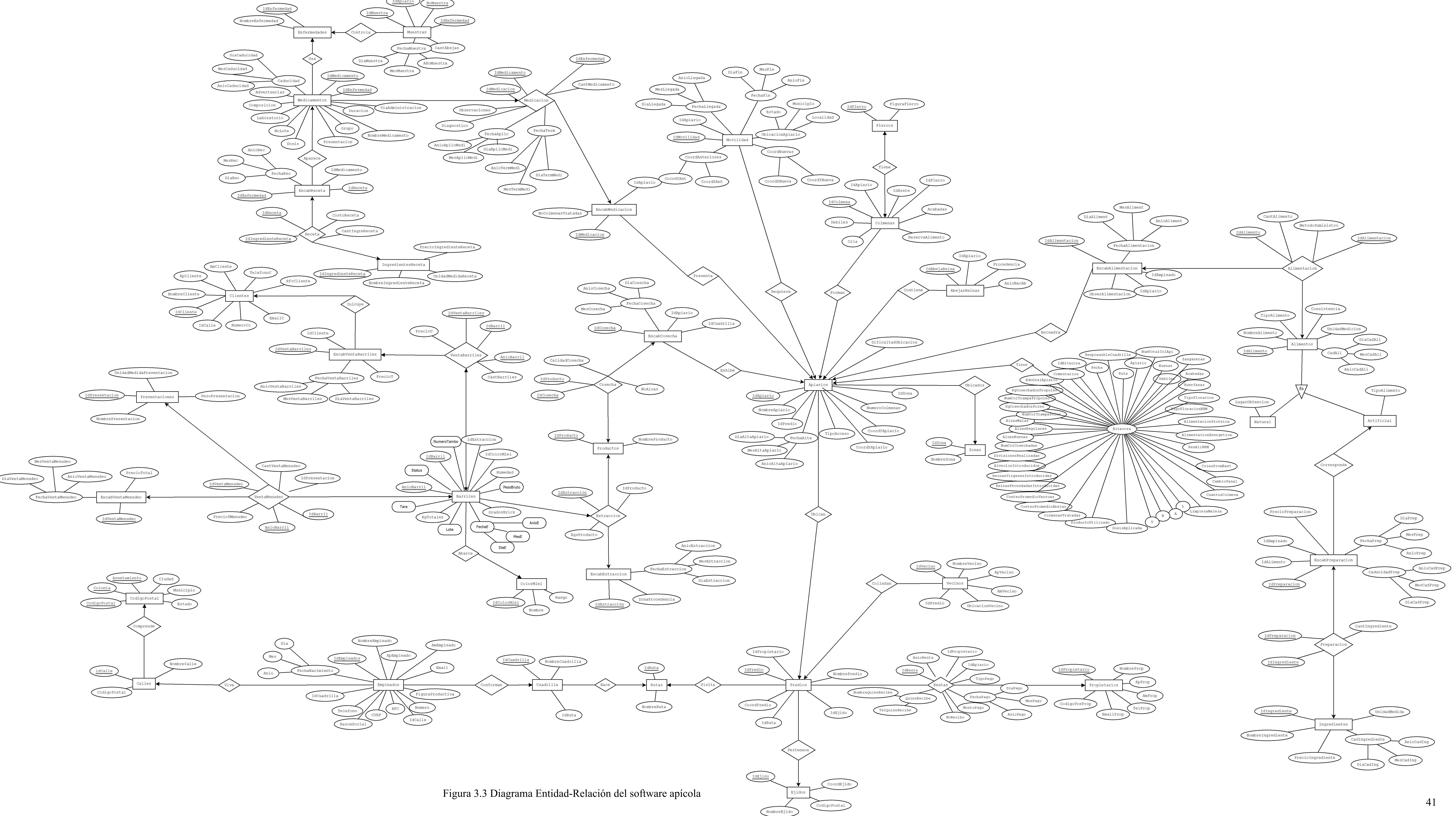


Figura 3.3 Diagrama Entidad-Relación del software apícola

Cabe mencionar que del total de módulos que integran el software apícola, se estuvo colaborando de forma directa en la programación en la parte de Usuarios, Respaldo y Recuperación de la base de datos, reportes de movilidad, así como en la consulta de las zonas haciendo uso de Google Maps.

#### **3.8.1.4 Codificación**

La codificación es donde se genera el producto, es decir, donde se realiza la programación de éste. Esencialmente, es llevar a código fuente, con el lenguaje o lenguajes que se hayan seleccionado, todo lo que fue modelado en la actividad anterior. Aunque parece que es la actividad que consume mayor tiempo durante el flujo de proceso, esto puede variar, ya que las actividades anteriores son la base del proyecto, y un error en la etapa anterior puede generar gran tiempo de codificación, pero si la etapa de modelado es realizada de manera correcta, la codificación puede realizarse en menor tiempo, siempre tomando en cuenta las características del sistema, que tan crítico es éste y el lenguaje de programación.

Para describir de manera gráfica este módulo, a continuación en la figura 3.4 se aprecia un diagrama de bloques del software, que permite identificar claramente cada parte del sistema.

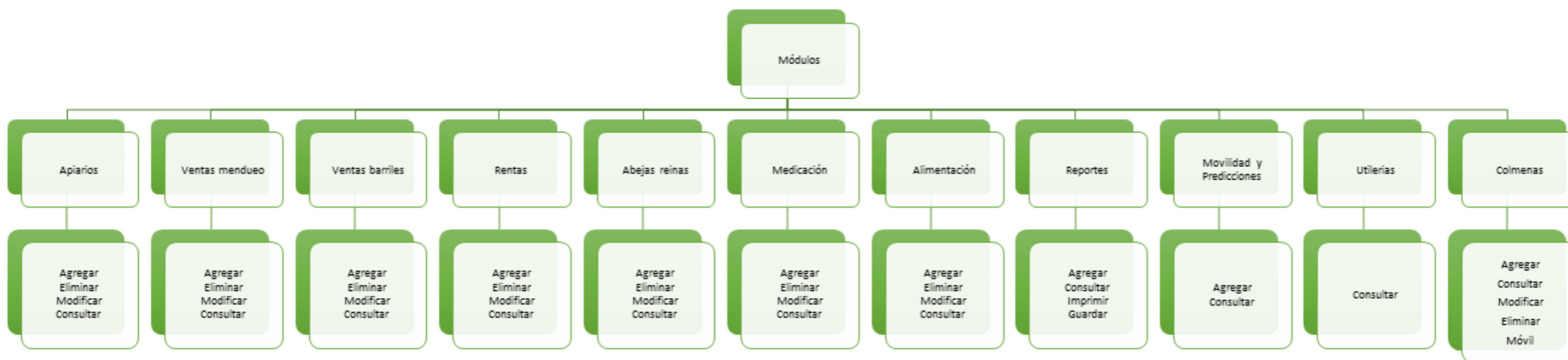


Figura 3.4 Diagrama de bloques del sistema apícola

Enseguida se describe el diagrama de bloques anterior, detallando cada uno de sus módulos.

- Módulo apiarios: permite llevar el control de los apiarios de la empresa, considerando aspectos de ubicación, número de apiarios y colmenas. Cuenta con las funciones de agregar, modificar, eliminar y consultar.
- Módulo ventas menudeo: permite llevar el control de las ventas de los productos comercializados a menudeo, tales como la miel, propóleos, polen, entre otros, gestionando los mismos a través de las funciones de agregar, modificar, eliminar y consultar.
- Módulo ventas barriles: permite controlar la venta de barriles de miel, gestionando los mismos mediante las funciones de agregar, modificar, eliminar y consultar.
- Módulo rentas: controla lo referente a la renta de predios donde se encuentran ubicados los apiarios, así como sus propietarios, formas y fechas de pago, así como el control de los datos de los empleados que laboran en la empresa y que realizan una parte de las operaciones. Permite agregar, modificar, eliminar y consultar.
- Módulo abejas reinas: en este módulo se lleva el control de las abejas reinas con las que cuenta la empresa, con información de año de nacimiento, procedencia y ubicación de las mismas en los apiarios. Permite agregar, modificar, eliminar y consultar.
- Módulo medicación: permite gestionar la medicación de cada uno de los apiarios con los que cuenta la empresa, así como las enfermedades detectadas y medicamentos aplicados.
- Módulo alimentación: permite gestionar la alimentación de las abejas, mediante el control de los alimentos, preparaciones y bitácora general de alimentación. Cuenta con las funciones de agregar, eliminar, modificar y consultar.

- Módulo reportes: este módulo permite la generación, consulta, impresión y almacenamiento de reportes de los diferentes módulos del sistema, tales como apiarios, alimentos, medicamentos, abejas reinas, ventas, entre otros.
- Módulo movilidad y predicciones: permite llevar a cabo la reubicación física de los apiarios. Así mismo, realizar predicciones de los niveles de riesgo de determinado punto en Google Maps.
- Módulo utilerías: en este módulo se puede realizar la gestión de usuarios del sistema, generación de códigos QR para los apiarios, así como el respaldo y restauración de la base de datos.

### **3.8.1.5 Pruebas**

Las pruebas de software son un elemento esencial en cualquier desarrollo, ya que revelan si el software funciona de manera correcta y detecta si tiene fallo o funciona de manera inadecuada conforme a diseño y/o codificación, para así poder corregirlos, esto con el fin de entregar un producto de calidad.

Una prueba es un conjunto de actividades que se pueden planear y realizarse de manera sistemática con el objetivo de verificar y validar lo diseñado y codificado.

Las pruebas que se realizaron fueron la llamada caja negra y la conocida como caja blanca. A continuación se describen cada una de ellas.

- Caja negra: también conocidas como pruebas de comportamiento, son las pruebas funcionales del software. Permite probar el software sin necesidad de conocer como está codificado, y se realizan aplicando un determinado grupo de datos y observando las salidas para determinar si dicha función se desempeña de forma correcta. Para esto se tomó en cuenta todo aquello que arrojaba el sistema, como datos que desplegaba y la forma en que los desplegaba. También se verificó que las búsquedas o reportes arrojaran la información realmente solicitada.

- Caja blanca: también llamada prueba de vidrio. Dicha prueba utiliza la estructura de control descrita como parte del diseño a nivel de componentes para derivar casos de prueba, es decir, son pruebas estructurales. Siguiendo una estructura lógica, y conociendo el código, se prueba destinadas a comprobar si, efectivamente, el código hace correctamente lo que el diseño marca. Se probó módulo por módulo, revisando los datos de entrada, su procesamiento, y que datos de salida se obtenían de dicho módulo.

Así mismo, se realizó una prueba de seguridad del sistema, la cual es vital dentro de un entorno Web, ya que garantiza la integridad de los datos que se manejan en ella, así como el del sistema en general.

Las pruebas se diseñan con el objetivo de detectar debilidades del lado del cliente, las comunicaciones de red que ocurren conforme los datos pasan del sitio web al servidor y viceversa.

Dentro de este sistema se revisó lo concerniente a:

- Inyecciones SQL.
- Encriptados.
- Autorizaciones (usuarios y contraseñas del sistema, así como a la base de datos).

#### **3.8.1.6 Despliegue**

En la actividad de despliegue, el software se entrega al cliente para su evaluación y posteriormente se retroalimenta con base a los resultados obtenidos.

Para el sistema apícola, y dado que en la empresa Frutimielmex no existía un software previo, fue implementado sin necesidad de hacerlo en paralelo con algún otro software.



### 3.8.2 Software de movilidad y predicciones

En lo correspondiente al segundo software realizado, el objetivo fue generar una herramienta computacional de movilidad y predicción, la cual tiene sus bases en la minería de datos, haciendo uso de la técnica de árboles de decisión. Mediante este software se pretende demostrar el cumplimiento de la hipótesis, así como el logro del objetivo general planteados en esta investigación. Dicha aplicación está integrada por dos módulos: en el primero se obtiene el nivel de riesgo de un punto específico de la zona de trabajo; y el segundo da la posibilidad al usuario de realizar la movilidad del apiario, es decir, cambiar de ubicación el apiario actual hacia una nueva ubicación, considerando los niveles de riesgo de la ubicación actual y la nueva.

Para alcanzar el objetivo principal de la investigación y de manera específica de este software, se llevaron a cabo los siguientes pasos:

- **Recabar información de las ubicaciones de los apiarios en un área de trabajo de 15,552 kilómetros cuadrados en el estado de Jalisco:** con el apoyo de la empresa Frutimielmex, de la Dra. Fatima Ezzahra Housni, quien funge como codirectora de este proyecto, así como de Iaritza Noriega Valdez que es su colaboradora, y con el uso de la herramienta de representación cartográfica de ArcGIS, se recabaron las ubicaciones geográficas de cada uno de los 196 apiarios con los que contaba la empresa en el momento del levantamiento de información, obteniendo de cada uno de ellos la latitud y longitud.

Con base en la obtención de las coordenadas de los apiarios, y para facilitar el manejo de la información, se establecieron 27 zonas de trabajo de 576 kilómetros cuadrados, obteniendo así un área total de 15,552 kilómetros cuadrados.

Dicha información se almacenó en archivos de Excel, generando así un total de 27 archivos, nombrados como TABLA\_1, con numeración consecutiva hasta la TABLA\_28, omitiendo la TABLA\_25 por la razón que dicha zona no

corresponde al estado de Jalisco. En la figura 3.5 se muestra la organización de los archivos.



Figura 3.5 Organización de archivos Excel de las zonas de trabajo

- **Identificar los tipos de vegetación, cultivos, zonas urbanas y cuerpos de agua que rodea a los apiarios:** una vez obtenidas las coordenadas geográficas de los apiarios y delimitada la zona total de trabajo total, a través de la herramienta ArcGIS, se detectaron la cantidad de hectáreas en cada una de las zonas de trabajo de cultivos de agricultura protegida, agricultura de riego (cultivo de aguacate), agricultura de temporal, vegetación, cuerpos de agua y zonas urbanas, almacenando la información en archivos de Excel, un archivo por cada una de las 27 zonas, para después migrarlo a la base de datos.

En la figura 3.6 se muestra un extracto del contenido de un archivo de Excel en el cual se aprecia la superficie del área analizada, clave interna de vegetación, el tipo de vegetación identificada, el tipo de formación (clasificación), la superficie dada en hectáreas y la superficie dada en miles de hectáreas.

	A	B	C	D	E	F	G
1	Shape_Leng	Shape_Area	CVE_VE	TIP_VEG	FORMACION	SUP_HA	SUP_MHA
47	19496.22953190000	1996453.128600000000	BQ/VSA	VEGETACIÓN SECUNDARIA ARBOREA DE BOSQUE DE ENCINO	LATIFOLIADAS	199.65	0.20
59	4695.63015663000	118209.303112000000	BQ/VSA	VEGETACIÓN SECUNDARIA ARBOREA DE BOSQUE DE ENCINO	LATIFOLIADAS	11.82	0.01
323	29233.75103880000	3977387.783110000000	ZU	ZONA URBANA	AREAS NO FORESTALES	397.74	0.40
324	24.92907148980	33.65886726980	ZU	ZONA URBANA	AREAS NO FORESTALES	0.00	0.00
333	2327514.48958000000	594573396.615000000000	ZU	ZONA URBANA	AREAS NO FORESTALES	59457.34	59.46
334	2600.61684182000	315196.870270000000	H2O	CUERPO DE AGUA	AREAS NO FORESTALES	31.52	0.03
335	3412.57862827000	485665.634308000000	H2O	CUERPO DE AGUA	AREAS NO FORESTALES	48.57	0.05
336	4543.40477788000	236565.591431000000	H2O	CUERPO DE AGUA	AREAS NO FORESTALES	25.66	0.03
337	1828.94237713000	125093.762283000000	TA	AGRICULTURA DE TEMPORAL DE CICLO ANUAL	AREAS NO FORESTALES	12.51	0.01
338	1505.09795344000	64512.516708200000	TA	AGRICULTURA DE TEMPORAL DE CICLO ANUAL	AREAS NO FORESTALES	6.45	0.01
359	2470.36786866000	108268.730392000000	TA	AGRICULTURA DE TEMPORAL DE CICLO ANUAL	AREAS NO FORESTALES	10.83	0.01
360	6701.39236670000	111887.521515000000	TA	AGRICULTURA DE TEMPORAL DE CICLO ANUAL	AREAS NO FORESTALES	11.19	0.01
361	2223.54507877000	90021.868743400000	TA	AGRICULTURA DE TEMPORAL DE CICLO ANUAL	AREAS NO FORESTALES	9.00	0.01
362	5347464.74628000000	2078213975.310000000000	TA	AGRICULTURA DE TEMPORAL DE CICLO ANUAL	AREAS NO FORESTALES	207821.40	207.82
363							
364							

Figura 3.6 Extracto de información de una zona en archivo Excel

Es importante mencionar que por cuestiones de confidencialidad con la empresa, no se muestran los datos contenidos en cada una de los archivos de la figura anterior.

Así mismo, se generó un mapa de la área de trabajo establecida, ubicando los tipos de vegetación, cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate), agricultura de temporal de ciclo anual, zonas urbanas y cuerpos de agua. En la figura 3.7 se aprecia dicho mapa.

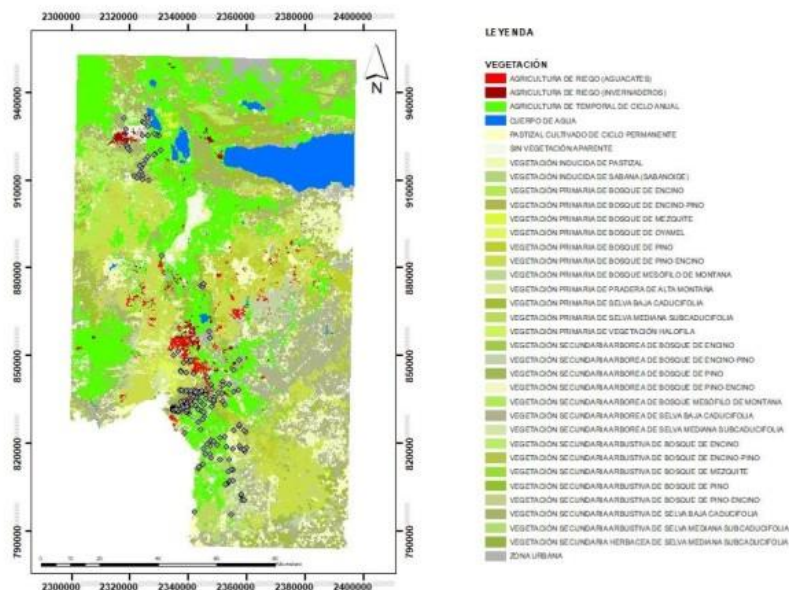


Figura 3.7. Mapa de identificación de vegetación

- Identificar los factores de riesgo para la supervivencia de las abejas:** derivado de la investigación en la literatura, artículos científicos y de la propia experiencia de la empresa Frutimielmex, se determinaron como factores de riesgo los cultivos de agricultura protegida, cultivos de agricultura de riego (aguacate) y cultivos de temporal, no siendo en sí el riesgo el propio cultivo, sino el uso desmesurado de agroquímicos, fertilizantes, pesticidas, insecticidas y herbicidas, utilizados durante los procesos productivos, causando la muerte de las abejas. De acuerdo con Mitchell et al. (2017), el uso de los neonicotinoides

va en aumento, y el impacto genera gran preocupación por las abejas, tanto melíferas como silvestres. La clasificación se catalogó como riesgo alto, medio o bajo, considerando la suma de porcentajes de cada factor en relación con el total de hectáreas de cada zona de trabajo. En la tabla 3.1 se muestra la clasificación de los niveles de riesgo, así como el rango de porcentaje para cada uno de los niveles, siendo el máximo porcentaje el 40%.

**Tabla 3.1**

*Clasificación de niveles de riesgos para las abejas.*

<b>Nivel de riesgo</b>	<b>Rango</b>
<b>BAJO</b>	0.0% - 11.0%
<b>MEDIO</b>	11.1% - 27.0%
<b>ALTO</b>	27.1% - 40.0%

- **Generar mapas que reflejen la ubicación, la vegetación y los factores de riesgo en torno a los apiarios:** una vez obtenidos los elementos mencionados en el párrafo anterior, se generaron mapas con la herramienta de Google Earth, donde se muestra la ubicación de cada uno de los 196 apiarios, vegetación, cultivos de agricultura protegida, temporal y de riego, zonas urbanas y cuerpos de agua en la área de trabajo total. La ubicación de los apiarios puede ser consultada a través de la misma herramienta computacional generada, mediante la base de datos y visualizada en Google Maps, el cual está integrado a la misma herramienta.

En la figura 3.8 se muestra un mapa del área total de trabajo, con divisiones de las zonas, en la cual se catalogan con nivel de riesgo alto, medio y bajo, dibujadas en color rojo, amarillo y verde respectivamente, tomando como consideración el nivel de riesgo general de la zona, y no de un punto específico dentro de la misma.

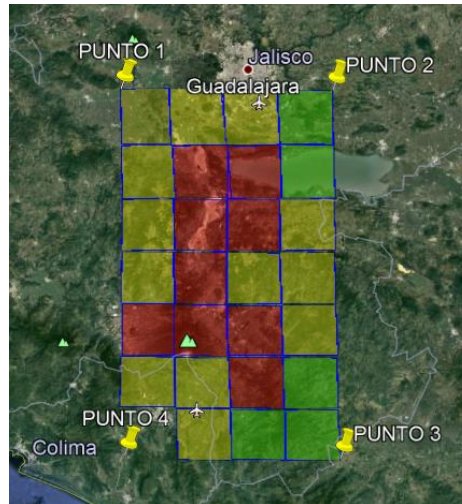


Figura 3.8 Mapa de niveles de riesgo por zonas

- **Investigar las características de las diferentes técnicas aplicables para el análisis y minería de datos:** durante esta etapa se investigó información referente a la minería de datos y las herramientas que utiliza. La minería de datos es una de las ramas del descubrimiento de información, que cuenta con una sub rama denominada aprendizaje supervisado, de la cual se deriva la clasificación, y la herramienta utilizada en el proyecto, que es árboles de decisión (Matsudo, 2001). Los árboles de decisión se implementaron a través del lenguaje de programación R, con el uso del algoritmo C5.0, el cual fue el más apto, tal como se describe en el numeral 2.2 marco teórico y 2.3 marco conceptual.
- **Definir la técnica más adecuada para obtener la predicción de las nuevas coordenadas de ubicación:** de acuerdo a la investigación realizada, y previa comparación en el apartado 2.2 marco teórico de las diferentes herramientas de predicción que utiliza minería de datos como son, Bayes Net (Red Bayesiana), Naive Bayes (Bayes Genuino), AdaBoost, Decision Table (Tabla de decisiones) y Decision Tree (árboles de decisiones); siendo esta última la que se eligió dado su alto porcentaje de instancias correctamente clasificadas. Dicho porcentaje

oscila entre el 87% y 88%, dadas las pruebas realizadas con el programa Weka, así como con el entorno de desarrollo de RStudio.

Para hacer las pruebas en el programa Weka, fue necesario realizar previamente lo siguiente:

- o Se llevó a cabo un análisis de la información de la cantidad de hectáreas con las que contaba cada tipo de vegetación identificada en cada una de las zonas, haciendo énfasis en los factores de riesgo (cultivos de aguacate, agricultura protegida y temporal) generando así un solo archivo en Excel con dicha información. En la figura 3.9 se aprecia dicho análisis.

	A	B	C	D	E	F	G	H	I	J
	ZONA	HAS AGUACATE	% AGUACATE	HAS INVERNADERO	% INVERNADERO	HAS TEMPORAL	% TEMPORAL	TOTAL HAS ZONA	% SUMA A+H+T	PROMEDIO
1	1	0.00	0.000%	0.00	0.000%	322,122.05	90.868%	354,493.75	90.868%	30.283%
2	2	0.00	0.000%	365,654.34	55.834%	139,296.68	21.270%	654,893.23	77.104%	25.701%
3	3	0.00	0.000%	0.00	0.000%	353,207.14	78.395%	450,547.54	78.395%	26.132%
4	4	0.00	0.000%	0.00	0.000%	208,644.76	64.113%	325,434.48	64.113%	21.371%
5	5	0.00	0.000%	121,894.38	28.223%	201,803.84	45.583%	431,623.42	74.813%	24.338%
6	6	1,462,619.76	33.540%	1,950,177.18	44.720%	854,445.38	19.533%	4,360,875.58	37.653%	32.618%
7	7	132,870.48	3.957%	2,681,469.56	79.849%	370,133.10	11.022%	3,356,160.73	94.628%	31.609%
8	8	0.00	0.000%	9,013.52	2.360%	215,363.40	56.396%	381,875.43	58.757%	19.586%
9	9	142,170.53	56.036%	0.00	0.000%	35,200.27	13.674%	253,714.08	63.910%	23.303%
10	10	605,424.90	72.409%	0.00	0.000%	124,055.03	14.740%	841,641.24	87.149%	23.050%
11	11	251,076.36	23.343%	487,539.32	45.338%	247,335.08	23.007%	1,075,340.22	91.687%	30.562%
12	12	31,497.10	35.812%	0.00	0.000%	27,205.37	10.648%	255,489.41	46.461%	15.487%
13	13	75,754.27	30.317%	3,236.01	1.320%	53,067.35	23.639%	249,872.50	55.276%	18.425%
14	14	1,681,021.92	28.644%	3,179,278.60	54.174%	860,900.66	14.663%	5,868,843.39	37.488%	32.496%
15	15	411,392.15	49.166%	9,207.06	1.100%	263,717.13	31.517%	836,749.55	81.783%	27.261%
16	16	177,474.95	40.363%	3,650.15	0.842%	13,467.51	3.108%	433,261.21	44.913%	14.971%
17	17	670,313.44	73.081%	43,207.81	4.711%	43,208.43	4.711%	917,220.18	82.503%	27.501%
18	18	510,118.77	56.111%	131,896.58	14.506%	133,854.74	14.723%	909,127.17	85.342%	28.447%
19	19	421,023.11	41.238%	164,835.37	16.107%	250,055.16	24.434%	1,023,393.57	81.759%	27.253%
20	20	19.75	0.006%	130,563.20	41.557%	940.26	0.236%	315,110.37	41.866%	13.953%
21	21	178,718.63	43.344%	0.00	0.000%	211.32	0.052%	406,632.23	43.996%	14.665%
22	22	89,168.50	37.543%	0.00	0.000%	42,501.85	17.895%	237,508.22	55.438%	18.473%
23	23	371,647.49	77.186%	0.00	0.000%	89,454.72	18.574%	481,619.31	95.740%	31.913%
24	24	0.00	0.000%	0.00	0.000%	43,207.81	53.809%	80,298.84	53.809%	17.336%
25	25	0.00	0.000%	0.00	0.000%	42,162.25	71.836%	58,632.37	71.836%	23.945%
26	26	1,208.25	0.955%	0.00	0.000%	42,767.38	35.223%	121,396.54	36.225%	12.078%
27	27	71,447.06	24.372%	0.00	0.000%	39.80	0.014%	293,152.38	24.366%	8.123%
28										
29										
30										

Figura 3.9 Concentrado de información por zonas y factores de riesgo

- o Una vez concentrada la información, se obtuvo el porcentaje previamente ponderado para cada uno de los factores de riesgo, y con esos datos, se asignó el nivel de riesgo a cada zona, catalogándola como alto, medio o bajo. En la figura 3.10 se muestran la información.

	A	L	M	N	O	P	Q	R	S	T
1	ZONA	HAS AG 40%	HAS INV 40%	HAS TMP 20%	SUMA HA	%40 AG	%40 INV	%20 TMP	% SUMA	RIESGO
2	1	0.0000	0.0000	64424.4100	64424.4100	0%	0%	18%	18.174%	MEDIO
3	2	0.0000	146261.9760	27859.3360	174121.3120	0%	22%	4%	26.588%	MEDIO
4	3	0.0000	0.0000	70641.4280	70641.4280	0%	0%	16%	15.679%	MEDIO
5	4	0.0000	0.0000	41728.9520	41728.9520	0%	0%	13%	12.823%	BAJO
6	5	0.0000	48753.9920	40236.1680	88990.1600	0%	11%	9%	20.608%	MEDIO
7	6	585047.9040	780070.8720	170889.0760	1536007.8520	13%	18%	4%	35.222%	ALTO
8	7	53148.1920	1072587.8240	74026.6200	1193762.6360	2%	32%	2%	35.727%	ALTO
9	8	0.0000	3605.4080	43072.6800	46678.0880	0%	1%	11%	12.223%	BAJO
10	9	56868.2120	0.0000	7040.0540	63908.2660	22%	0%	3%	25.183%	MEDIO
11	10	243763.3600	0.0000	24811.0060	268580.3660	23%	0%	3%	31.912%	ALTO
12	11	100430.5440	195015.9680	49467.0160	344913.5280	3%	18%	5%	32.075%	ALTO
13	12	36598.8400	0.0000	5441.0740	42033.9140	14%	0%	2%	16.455%	MEDIO
14	13	30301.7080	1319.2040	11813.4700	43434.3820	12%	1%	5%	17.383%	MEDIO
15	14	672408.7680	1271711.4400	1172180.1320	2116300.3400	11%	22%	3%	36.061%	ALTO
16	15	164556.8600	3682.8240	52743.4260	220983.1100	20%	0%	6%	26.410%	MEDIO
17	16	70989.9800	1460.0600	2693.5020	75143.5420	16%	0%	1%	17.344%	MEDIO
18	17	268125.3760	17283.1240	8641.6980	234050.1980	23%	2%	1%	32.059%	ALTO
19	18	204047.5080	52758.6320	26770.9480	283577.0880	22%	6%	3%	31.192%	ALTO
20	19	168729.2440	65934.1480	50011.0360	284674.4280	16%	6%	5%	27.817%	ALTO
21	20	7.9000	52385.2800	168.0520	52581.2320	0%	17%	0%	16.687%	MEDIO
22	21	71487.4520	0.0000	42.2640	71529.7160	18%	0%	0%	17.588%	MEDIO
23	22	35667.4000	0.0000	8500.3700	44167.7700	15%	0%	4%	18.596%	MEDIO
24	23	148658.9960	0.0000	17890.9440	166549.9400	31%	0%	4%	34.581%	ALTO
25	24	0.0000	0.0000	8641.5620	8641.5620	0%	0%	11%	10.762%	BAJO
26	26	0.0000	0.0000	8432.4500	8432.4500	0%	0%	14%	14.367%	MEDIO
27	27	483.3000	0.0000	8553.4760	9036.7760	0%	0%	7%	7.444%	BAJO
28	28	28578.8240	0.0000	7.9600	28586.7840	10%	0%	0%	9.751%	BAJO

Figura 3.10 Concentrado de información ponderado con niveles de riesgo

- o Al tener los niveles de riesgo para cada zona, posteriormente se generó otra tabla, donde se eligieron para cada zona, 5 puntos al azar dentro de la misma. Por lo tanto, se obtuvo una tabla con 135 registros. Es decir, cinco puntos para cada zona, y siendo un total de 27 zonas. En la figura 3.11 se encuentra un extracto de la información previamente descrita.

	A	B	C
1	CoordY	CoordX	Riesgo
2	20.46222811830	-103.79267618400	MEDIO
3	20.54262800000	-103.86470600000	MEDIO
4	20.54403100000	-103.72490300000	MEDIO
5	20.41080600000	-103.87069200000	MEDIO
6	20.40296700000	-103.70298300000	MEDIO
7	20.46047857910	-103.56262624300	MEDIO
8	20.54337200000	-103.65120300000	MEDIO
9	20.54182200000	-103.50251400000	MEDIO
10	20.41730600000	-103.63413100000	MEDIO
11	20.42549400000	-103.49898100000	MEDIO
12	20.45842493890	-103.33260242800	MEDIO
13	20.55566100000	-103.42160000000	MEDIO
14	20.56692200000	-103.28035000000	MEDIO
15	20.42401700000	-103.39949700000	MEDIO
16	20.43041900000	-103.26789400000	MEDIO
17	20.45606735480	-103.10260891500	BAJO
18	20.51260300000	-103.15745600000	BAJO
19	20.51565800000	-103.18455800000	BAJO
20	20.39513100000	-103.17163600000	BAJO
21	20.39105000000	-103.03006100000	BAJO
22	20.24540479380	-103.79436023900	MEDIO
23	20.30157200000	-103.86118900000	MEDIO
24	20.31213100000	-103.73355600000	MEDIO
25	20.17698600000	-103.85321700000	MEDIO
26	20.17641100000	-103.72144200000	MEDIO

Figura 3.11 Extracto de información de coordenadas y niveles de riesgo

- o Al contar con el archivo en el formato tal como se muestra en la figura anterior, se guarda el archivo en formato CSV (Comma Separated Values/Valores Separados por Comas), para su posterior uso. Se realizó una copia del mismo, y se convirtió a formato ARFF (Attribute-Relation File Format/Formato de Archivo de Relación de Atributo), con la ayuda del programa Sublime Text, dentro del cual se dio una estructura de cabeceras, tal como se muestra en la figura 3.12.

```
1 @RELATION 135datosriesgo
2
3 @ATTRIBUTE CoordX REAL
4 @ATTRIBUTE CoordY REAL
5 @ATTRIBUTE class {ALTO,MEDIO,BAJO}
6
7 @DATA
8 20.46222811830, -103.79267618400, MEDIO
9 20.54262800000, -103.86470600000, MEDIO
10 20.54403100000, -103.72490300000, MEDIO
11 20.41080600000, -103.87069200000, MEDIO
12 20.40296700000, -103.70298300000, MEDIO
13 20.46047857910, -103.56262624300, MEDIO
14 20.54337200000, -103.65120300000, MEDIO
15 20.54182200000, -103.50251400000, MEDIO
16 20.41730600000, -103.63413100000, MEDIO
17 20.42549400000, -103.49898100000, MEDIO
18 20.45842493890, -103.33260242800, MEDIO
19 20.55566100000, -103.42160000000, MEDIO
20 20.56692200000, -103.28035000000, MEDIO
21 20.42401700000, -103.39949700000, MEDIO
22 20.43041900000, -103.26789400000, MEDIO
23 20.45606735480, -103.10260891500, BAJO
```

Figura 3.12 Estructura del archivo ARFF en Sublime Text

- o Al finalizar la acción anterior, ya se cuenta con el archivo preparado para su manipulación con el programa Weka.



En las figura 3.13, 3.14, 3.15, 3.16 y 3.17 se aprecian capturas de pantalla de la interfaz del programa Weka, en el cual fueron probados los algoritmos con los mismos datos.

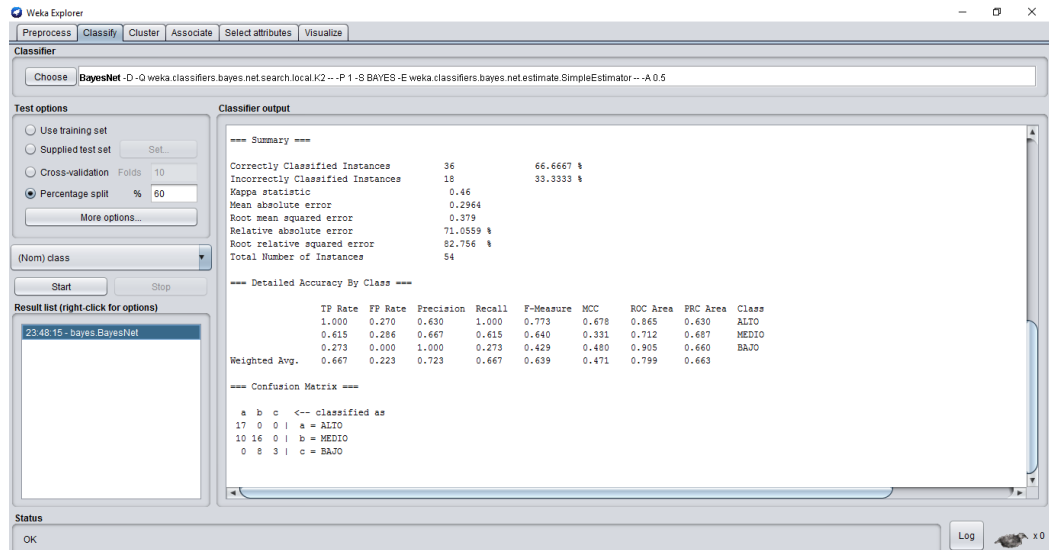


Figura 3.13 Weka utilizando Bayes Net (Red Bayesiana)

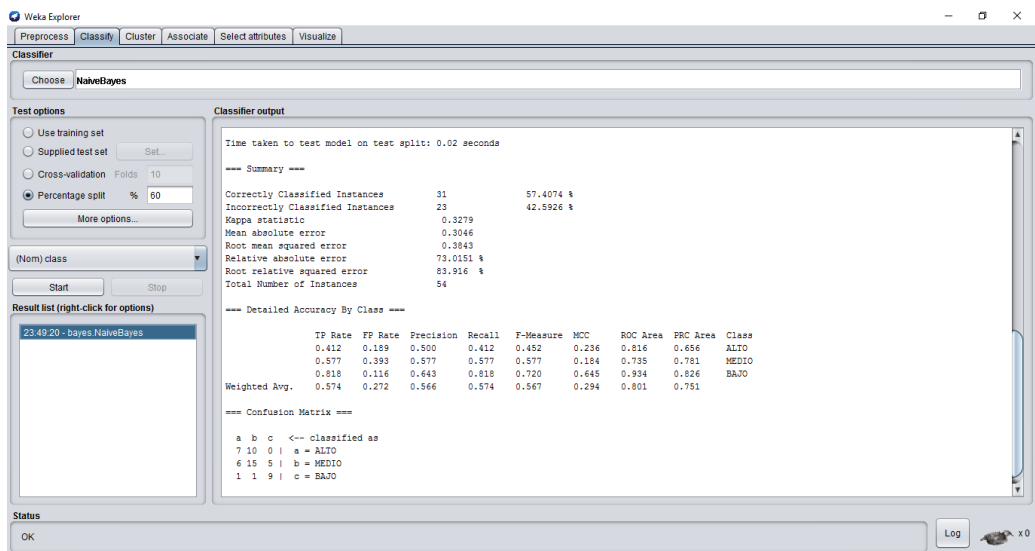


Figura 3.14 Weka utilizando Naive Bayes (Bayes Genuino)

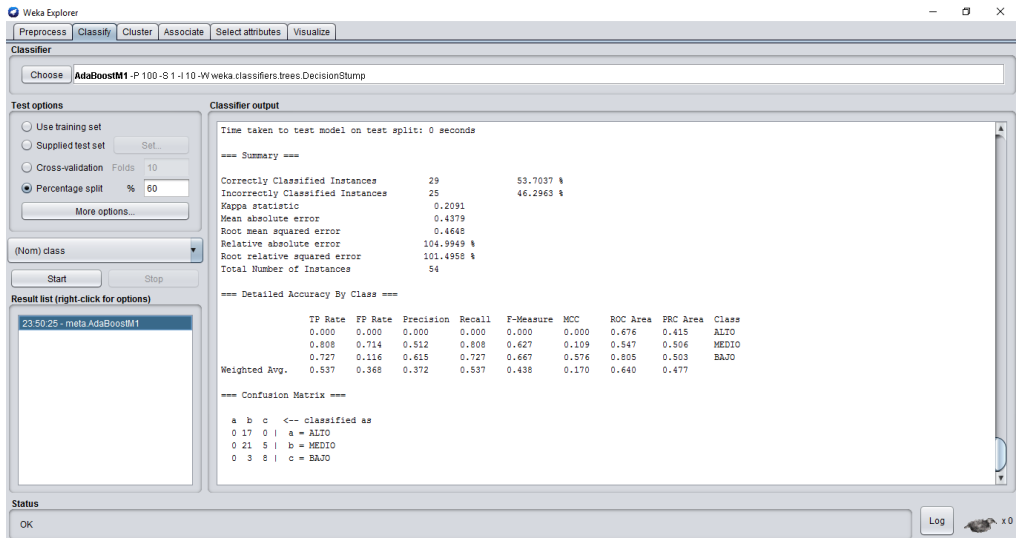


Figura 3.15 Weka utilizando AdaBoost

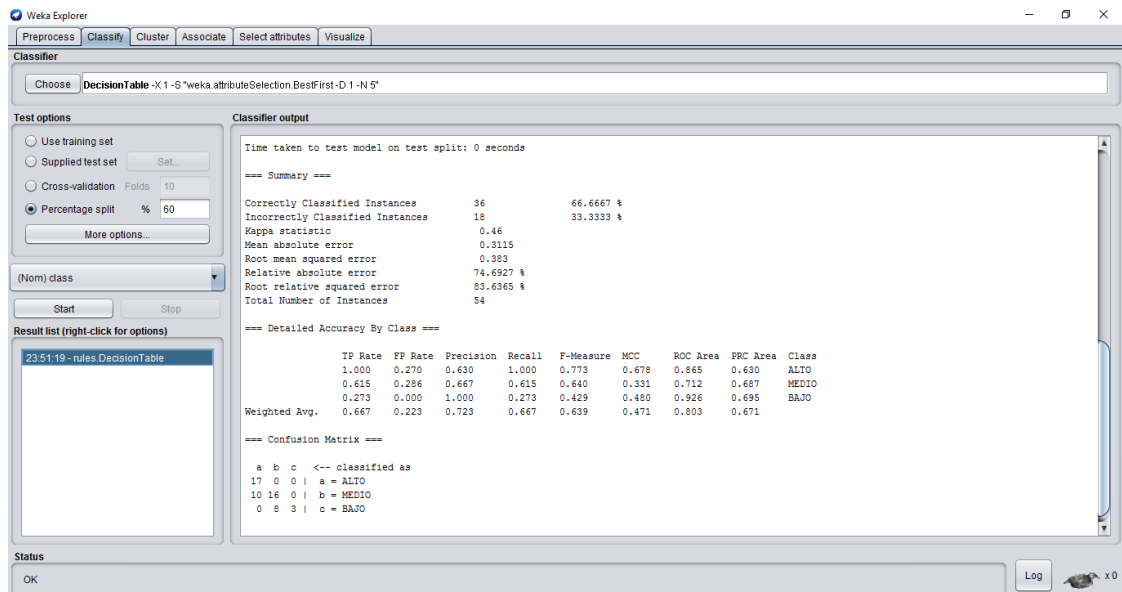


Figura 3.16 Weka utilizando Decision Table (Tabla de decisiones)

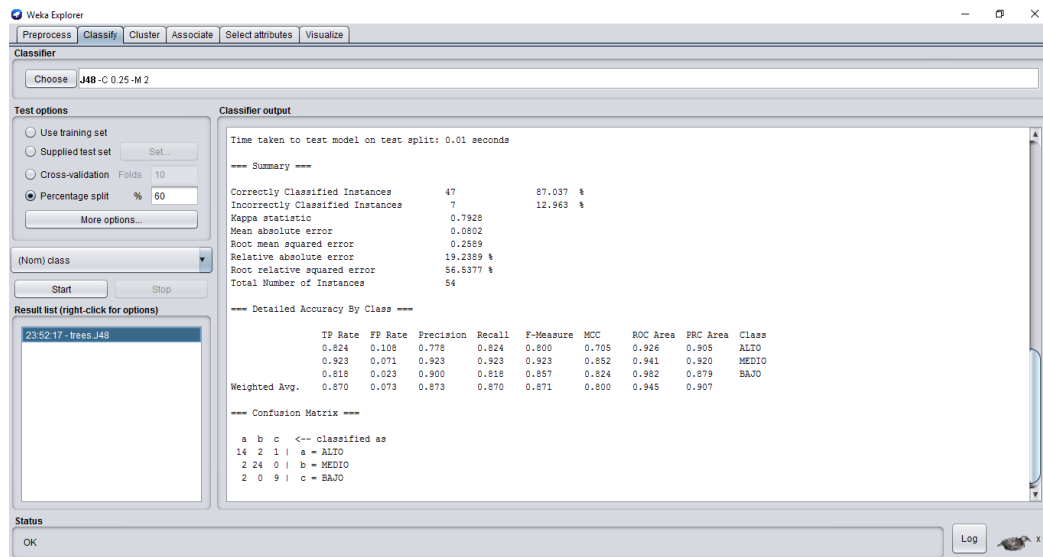


Figura 3.17 Weka utilizando Decision Tree (árboles de decisiones)

En la tabla 3.2 se muestran los resultados de la comparativa entre los cinco algoritmos utilizados con los mismos datos, considerando que para todos los casos se utilizó como parámetro un porcentaje de división del 60%. Donde el que obtuvo mayor efectividad, fue árboles de decisión.

**Tabla 3.2**

*Resultados de comparativa entre algoritmos*

<b>Algoritmo</b>	<b>Porcentaje de instancias correctamente clasificadas</b>	<b>Porcentaje de instancias incorrectamente clasificadas</b>
Bayes Net (Red Bayesiana)	66.6667 %	33.3333 %
Naive Bayes (Bayes Genuino)	57.4074 %	42.5926 %
AdaBoost	53.7037 %	46.2963 %
Decision Table (Tabla de decisiones)	66.6667 %	33.3333 %
Decision Tree (árboles de decisiones)	87.037 %	12.963 %

Una vez decidido el método de predicción, se procedió a realizar una investigación y comparativa de algunos los diferentes algoritmos que existen para generar árboles de decisión, entre los cuales se analizaron, C5.0, C4.5, árbol de clasificación y regresión (C&R), CHAID, QUEST y Tree-AS, tal como se describieron en el apartado 2.2 marco teórico. De los algoritmos comparados, el que mejor se adaptó al problema a resolver y al tipo de datos de entrada con los que se contaba, además de dar un mejor porcentaje de instancias correctamente clasificadas, fue el C5.0, el cual fue usado mediante librería pre

cargada en entorno de desarrollo de RStudio, por lo que no fue necesaria su programación, solo su implementación. Así mismo, para realizar la predicción, el mismo entorno de desarrollo, cuenta con la función denominada “predict()”, la cual posee internamente las instrucciones para realizar la predicción, con base en un modelo previamente creado dentro del mismo programa en R.

El modelo final del árbol de decisión, con los datos ingresados, y con el entrenamiento previo, quedó de la manera como se muestra en la figura 3.18.

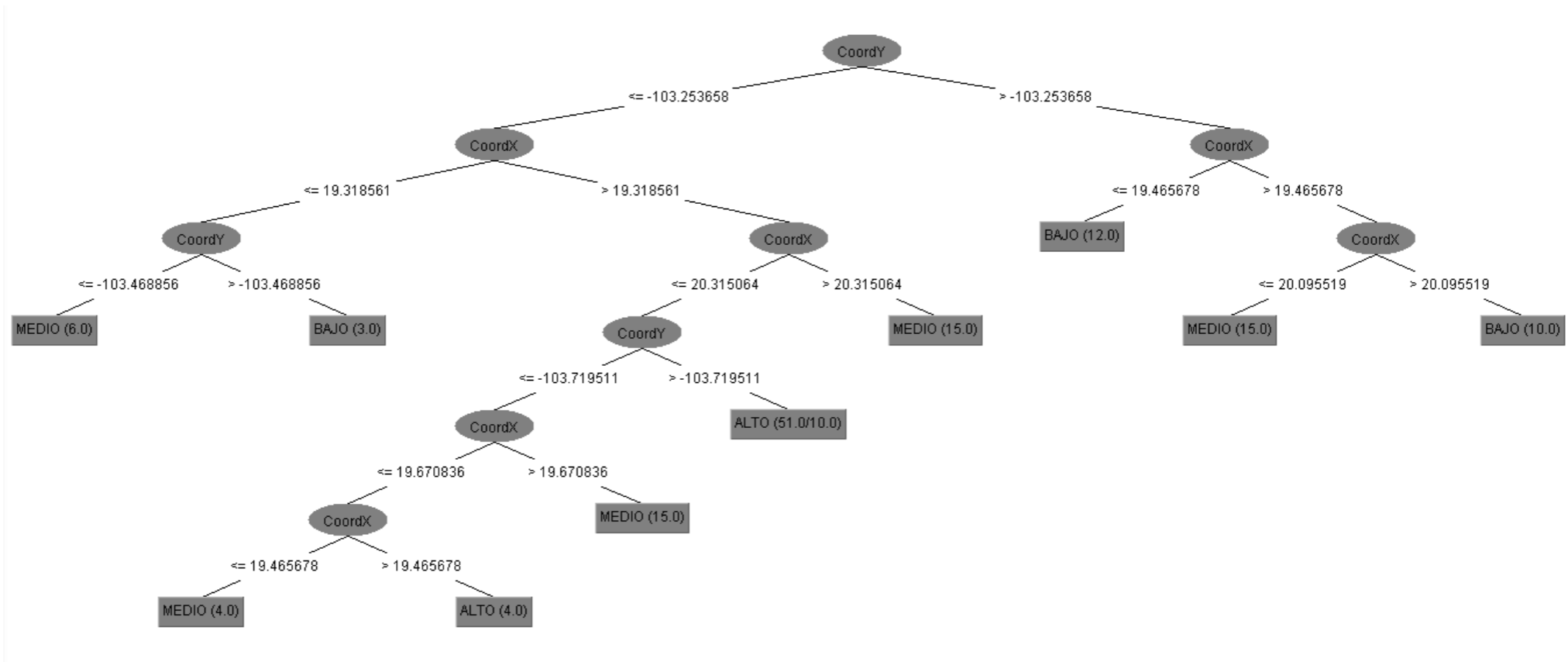


Figura 3.18 Modelo final del árbol de decisión

- **Diseño de interfaces y desarrollo del software para aplicación del análisis de datos, aplicando técnica de árboles de decisión:** una vez definido el método de predicción y el algoritmo para generarlo, se procedió a realizar el diseño de las interfaces y programación de las funciones que permiten interactuar al usuario con la aplicación. Dicha aplicación fue desarrollada para entorno web, utilizando HTML, PHP, JavaScript, Google Maps JavaScript API, así como conexión a base de datos MySQL. Las interfaces se pueden observar a mayor detalle en el apartado 4.5.2 Software de predicciones y movilidad.
- **Mostrar la información relacionada con la cantidad de hectáreas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, de la zona donde se encuentra ubicado el apiario seleccionado para su posible reubicación:** una vez ejecutada la aplicación, dentro de la interfaz, el usuario puede elegir cualquier punto dentro de la zona de trabajo, arrastrando el puntero del *mouse* y soltando en el lugar deseado, para obtener la información de la cantidad de hectáreas de los cultivos antes mencionados, así como los porcentajes de cada uno de ellos en relación con el total de hectáreas de la zona seleccionada. Esto se logra mediante la obtención de coordenadas de un punto específico a través del cursor de Google Maps, y mediante una función desarrollada en JavaScript, en la cual los parámetros de entrada son la latitud y la longitud, obteniendo de la función como retorno el número de zona elegida, tal como se muestra en la figura 3.19. Posteriormente, el número de zona se envía a través de una consulta a la base de datos a la tabla denominada porcentajes zonas, a través de PHP, donde se obtienen los porcentajes de cultivos de aguacates, cultivos de agricultura protegida y cultivos de temporal, como se aprecia en la figura 3.20, para su posterior visualización en un globo informativo dentro de Google Maps.

```

149 // Funcion para validar la ubicacion de la zona y mostrar el marcador con la información
150 function ubicarZona (cX, cY)
151 {
152     marker3.setMap(null);
153     //alert("valores funcion ubicarZona: "+cX+", "+cY);
154     // cY Latitud cX Longitud
155     // Se establecen los límites del polígono de trabajo, inferior, superior, izquierda y derecha
156     if((cX < -103.908013 || cX > -102.986697 || cY > 20.570193 || cY < 19.057682) || ((cX >= -11
157     {
158         alert("COORDENADAS FUERA DE ZONA DE TRABAJO");
159         num_zona=0;
160     }
161     else {
162         //verificación de las coordenadas para obtener la zona donde se encuentra
163         // Zona 1
164         if((cX >= -103.908013 && cX <= -103.677834) && (cY >= 20.35412 && cY <= 20.570193))
165         {
166             num_zona=1;
167         }
168         // Zona 2
169         if((cX >= -103.677834 && cX <= -103.447655) && (cY >= 20.35412 && cY <= 20.570193))

```

Figura 3.19 Función para ubicar zona codificada en JavaScript

+ Opciones		IdPorcZonas	NumZona	Porc_AR	Porc_AP	Porc_TMP	Porc_SUMA
<input type="checkbox"/>	Editar Copiar Borrar	1	1	0	0	0.91	0.91
<input type="checkbox"/>	Editar Copiar Borrar	2	2	0	0.56	0.21	0.77
<input type="checkbox"/>	Editar Copiar Borrar	3	3	0	0	0.78	0.78
<input type="checkbox"/>	Editar Copiar Borrar	4	4	0	0	0.64	0.64
<input type="checkbox"/>	Editar Copiar Borrar	5	5	0	0.28	0.47	0.75
<input type="checkbox"/>	Editar Copiar Borrar	6	6	0.34	0.45	0.2	0.98
<input type="checkbox"/>	Editar Copiar Borrar	7	7	0.04	0.8	0.11	0.95
<input type="checkbox"/>	Editar Copiar Borrar	8	8	0	0.02	0.56	0.59
<input type="checkbox"/>	Editar Copiar Borrar	9	9	0.56	0	0.14	0.7
<input type="checkbox"/>	Editar Copiar Borrar	10	10	0.72	0	0.15	0.87
<input type="checkbox"/>	Editar Copiar Borrar	11	11	0.23	0.45	0.23	0.92
<input type="checkbox"/>	Editar Copiar Borrar	12	12	0.36	0	0.11	0.46
<input type="checkbox"/>	Editar Copiar Borrar	13	13	0.3	0.01	0.24	0.55
<input type="checkbox"/>	Editar Copiar Borrar	14	14	0.29	0.54	0.15	0.97

Figura 3.20 Tabla porcentajes zonas

- Obtener mediante el software, un mapeo general de los niveles de riesgos para las abejas, clasificándolos en nivel alto, medio y bajo:** la aplicación es capaz de ofrecer al usuario los niveles de riesgo de un punto seleccionado de la zona de trabajo, esto mediante la obtención de las coordenadas geográficas decimales, siendo éstas el parámetro de entrada al árbol de decisión, el cual se ejecuta y arroja como resultado el nivel de riesgo del punto previamente elegido. A detalle, lo que sucede es lo siguiente: primero el usuario elige un punto en Google Maps, posicionando el cursor con forma de abeja. Al realizar esta acción, se obtienen las coordenadas compuestas por latitud y longitud, las



cuales son enviadas al programa R, a través de una función codificada en PHP, tal como se muestra en la figura 3.21.

```
2 <?php
3
4 $coordX = $_POST['coordX'];
5 $coordY = $_POST['coordY'];
6
7 $output=shell_exec("Rscript riskPrediction.R $coordX $coordY");
8
9 echo $output;
10
11 ?>
```

Figura 3.21 Función para conectar con R

Posteriormente el programa en R recibe los parámetros y realiza el procedimiento de entrenamiento, predicción y retorno de resultados, tal como se muestra en el extracto de código de la figura 3.22.

```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
riskPrediction.R
1 args <- commandArgs(TRUE)
2
3 coordenadaX <- args[1]
4 coordenadaY <- args[2]
5
6 # Carga el paquete específico del Arbol de clasificación C5.0
7 library(c50)
8
9 datos <- read.csv("c:/xampp/htdocs/www/php/135datosriesgo.csv", header = TRUE)
10
11
12 # Selección de una submuestra del 70% de los datos
13 set.seed(94)
14
15 tamaño.total <- nrow(datos)
16 tamaño.entreno <- round(tamaño.total*0.7)
17 datos.indices <- sample(1:tamaño.total, size=tamaño.entreno)
18 datos.entreno <- datos[datos.indices,]
19 datos.test <- datos[-datos.indices,]
20
21 # Acortamiento de nombres de ALTO, MEDIO Y BAJO
22 risk <- vector(length = dim(datos)[1])
23 risk[datos$riesgo=="ALTO"] <- "A"
24 risk[datos$riesgo=="MEDIO"] <- "M"
25 risk[datos$riesgo=="BAJO"] <- "B"
26 datos$riesgo <- factor(risk)
27
28 modelo <- C5.0(Riesgo ~ ., data = datos.entreno)
```

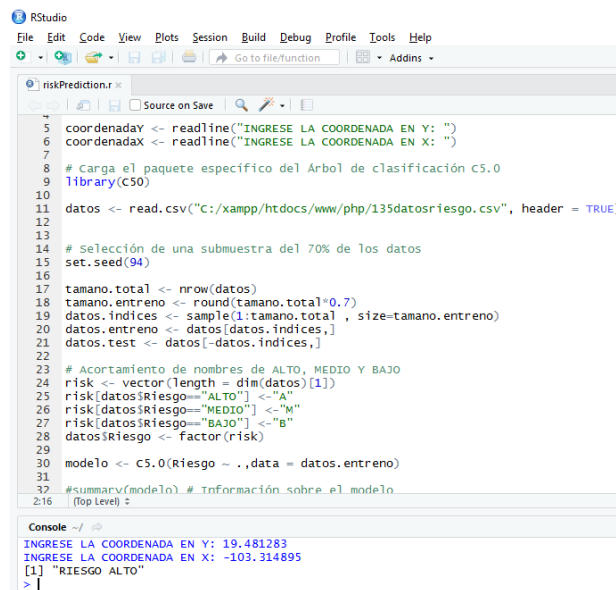
Figura 3.22 Programa de predicción en entorno R

- **Mostrar los nuevos índices de riesgo de la nueva ubicación de los apiarios, para su contrastación con las zonas anteriores:** al elegir un nuevo punto, se lleva a cabo una comparación del punto anterior respecto al nuevo, en relación a los niveles de riesgo, así como la información referente a la cantidad de hectáreas que se encuentran en la zona de cada uno de los cultivos que son factor de riesgo, además de los porcentajes de cada uno de ellos en relación con la zona. El procedimiento interno para la predicción, se lleva a cabo de la

misma manera como se describió en el punto anterior, y como se ilustra en la figura 3.21 y 3.22.

- **Procesar los resultados obtenidos:** al contar con la información actualizada de los niveles de riesgo de determinados puntos, se procesan los datos, obteniendo gráficas que permitan evaluar de manera numérica y visual los resultados, haciendo contraste en los niveles de riesgo de cada una de las 27 zonas de trabajo, considerando los factores de riesgo de agricultura protegida, agricultura de riego (cultivo de aguacate) y agricultura de temporal.

Para lograr procesar los resultados, de forma manual, se ejecutó el código en R, tal como se muestra en la figura 3.23, donde el programa solicitaba las coordenadas geográficas, las cuales eran obtenidas posicionando el cursor de Google Maps en determinado punto, y posteriormente introducidas en cada lectura de datos, obteniendo al final el nivel de riesgo. Al ejecutar el programa varias veces, en diferentes puntos, pero en la misma zona, se observaron algunas diferencias, tal como se muestra en el apartado 4.3 Resultados obtenidos.



```
5 coordinaday <- readline("INGRESE LA COORDENADA EN Y: ")
6 coordinadax <- readline("INGRESE LA COORDENADA EN X: ")
7
8 # Carga el paquete específico del Árbol de clasificación c5.0
9 library(c50)
10
11 datos <- read.csv("C:/xampp/htdocs/www/php/135datosriesgo.csv", header = TRUE)
12
13
14 # Selección de una submuestra del 70% de los datos
15 set.seed(94)
16
17 tamaño.total <- nrow(datos)
18 tamaño.entreno <- round(tamaño.total*0.7)
19 datos.indices <- sample(1:tamaño.total, size=tamaño.entreno)
20 datos.entreno <- datos[datos.indices,]
21 datos.test <- datos[-datos.indices,]
22
23 # Acortamiento de nombres de ALTO, MEDIO Y BAJO
24 risk <- vector(length = dim(datos)[1])
25 risk[datos$riesgo=="ALTO"] <- "A"
26 risk[datos$riesgo=="MEDIO"] <- "M"
27 risk[datos$riesgo=="BAJO"] <- "B"
28 datos$riesgo <- Factor(risk)
29
30 modelo <- C5.0(Riesgo ~ ., data = datos.entreno)
31
32 #summary(modelo) # Información sobre el modelo
33
34 Console ~/
35 INGRESE LA COORDENADA EN Y: 19.481283
36 INGRESE LA COORDENADA EN X: -103.314895
37 [1] "RIESGO ALTO"
38 > |
```

Figura 3.23 Ejecución de código manualmente

- **Difundir los resultados obtenidos:** los avances de investigación se dieron a conocer mediante ponencia en el IX Coloquio Internacional de Investigación para Estudiantes Universitarios (CIEU), en el municipio de Zapotlán El Grande, Jalisco. De igual manera, los resultados fueron difundidos en un artículo científico titulado “Predicción computacional para la reubicación física de apiarios en el estado de Jalisco”, además de participar en la ponencia denominada “Reubicación física de apiarios con ayuda de la computación” en el marco del V Día del Apicultor Jalisciense, con sede en Tonalá, Jalisco. Así mismo, los resultados son difundidos en la propia investigación, dentro del capítulo siguiente, en el cual se detallan las pruebas realizadas, recolección y procesamiento de datos, evaluación de los resultados, así como los impactos generados.

Para lograr la instalación y monitoreo de los módulos del software desarrollado para la empresa Frutimielmex, se llevó a cabo una estadía técnica en la misma, en un periodo comprendido del día 11 de Junio de 2018 al día 10 de Julio del 2018.

## Capítulo IV. Resultados

### 4.1 Pruebas realizadas

Las pruebas que se llevaron a cabo dentro de la investigación, radican principalmente en contrastar y expresar mediante gráficas los niveles de riesgo que tienen cada una de las zonas de trabajo, tanto de forma general considerando el total de hectáreas de la zona y su respectivo porcentaje en relación con las hectáreas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual; así como de forma ponderada, considerando los mismos factores, pero tomando como referencia los porcentajes asignados a cada factor en la investigación, los cuales son:

- Cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles): 40%.
- Agricultura de riego (cultivo de aguacate): 40%.
- Agricultura de temporal de ciclo anual: 20%.

De igual manera, se tomaron puntos al azar dentro de una zona de trabajo considerada como riesgo alto, y se probó la aplicación, siendo los parámetros de entrada para el árbol de decisión las coordenadas geográficas de dichos puntos, y obteniendo como resultado el nivel de riesgo para cada caso específico, observando que el nivel de riesgo de la zona se ve afectado por los niveles de riesgo de las zonas vecinas, es decir, el árbol de decisión basa su resultado al evaluar no sólo la zona de trabajo en cuestión, sino sus alrededores, siendo así una predicción más real.

### 4.2 Recolección y procesamiento de datos

La recolección de datos parte desde la obtención de los mismos con el dispositivo GPS, en la cual se levantaron coordenadas geográficas de los 196 apiarios de la empresa. El

dispositivo GPS guardó en su almacenamiento interno dichos datos, para posteriormente vaciarlos a archivo Excel, donde se relacionó el número de apiario con sus respectivas coordenadas. Posteriormente con la ayuda de la herramienta de representación cartográfica de ArcGIS, se obtuvieron imágenes satelitales provenientes del satélite LANDSAT, en la cual la Dra. Fatima Ezzahra Housni, identificó la vegetación, zonas urbanas, cuerpos de agua y cultivos de aguacate, temporal y agricultura protegida, con sus respectivas cantidades de hectáreas, categorizándolos por las zonas previamente definidas.

Dicha información, se almacenó en archivos de Excel, generando un archivo por cada una de las 27 zonas.

Una vez teniendo los archivos de Excel con la información, se procedió a analizarla y realizar un filtro, es decir, de cada uno de los 27 archivos, se extrajo únicamente la información de las hectáreas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, realizando la suma de cada uno de esos factores. Con base en estos factores, se obtuvo el porcentaje en relación al total de hectáreas, obteniendo así un valor porcentual para cada uno de los elementos. Posteriormente, se asignó una ponderación a cada uno de los factores considerados como riesgo, para obtener un nuevo porcentaje, y con base a ello, clasificar cada zona como riesgo alto, medio o bajo.

Enseguida se creó un archivo CSV (Comma Separated Values/ Valores Separados por Comas), en el cual se almacenó la coordenada geográfica central de cada una de las 27 zonas, así como 4 puntos más elegidos de forma aleatorio dentro de la misma zona, y su respectivo nivel de riesgo. El archivo CSV, que en total contenía un total de 135 filas, se utilizó como datos de entrada para el árbol de decisión, del cual se tomó el 70% de datos, como de entrenamiento.

El árbol de decisión fue programado en el lenguaje R, el cual ya cuenta con librerías y algoritmos pre cargados, que facilitan el trabajo. Dicho programa en R está conectado

con el lenguaje PHP, con el cual existe una conexión para el paso de parámetros de entrada para la predicción de los nuevos niveles de riesgo en un punto determinado.

En el siguiente apartado, se describen los resultados obtenidos una vez ejecutada la aplicación.

### 4.3 Resultados obtenidos

Derivado de las pruebas realizadas, se obtuvieron los siguientes resultados:

Se realizó una comparativa de cada una de las 27 zonas de trabajo de 576 kilómetros cuadrados, en la cual se expresa el porcentaje que existe de hectáreas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, en relación con el total de la zona. En la figura 4.1 se muestra la gráfica correspondiente, donde los datos del eje vertical presentan los porcentajes y los datos en el eje horizontal representan cada una de las 27 zonas de trabajo (la zona número 25 se omite porque no es objeto de análisis).

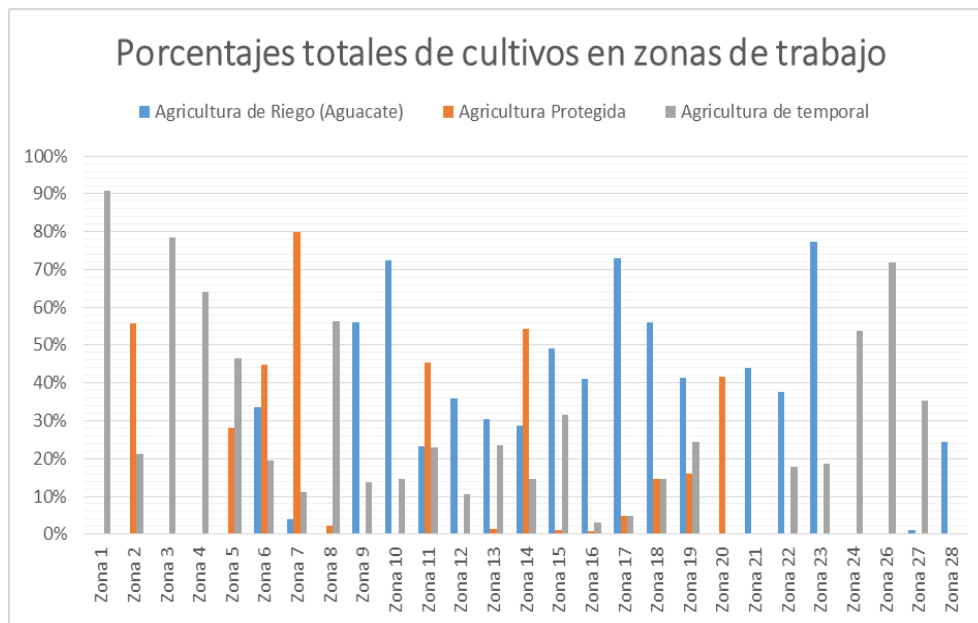


Figura 4.1 Comparativa de zonas de trabajo con porcentajes totales

De acuerdo a la ponderación dada a cada uno de los cultivos considerados como factores de riesgo descritos en el apartado 1.1 Descripción del trabajo de investigación del Capítulo I. Introducción, se generó la gráfica para visualizar de manera clara los diferentes porcentajes de cultivo. En la figura 4.2 se muestra la gráfica correspondiente, donde los datos del eje vertical representan el porcentaje ponderado que tiende a 40% y los datos en el eje horizontal cada una de las 27 zonas de trabajo.

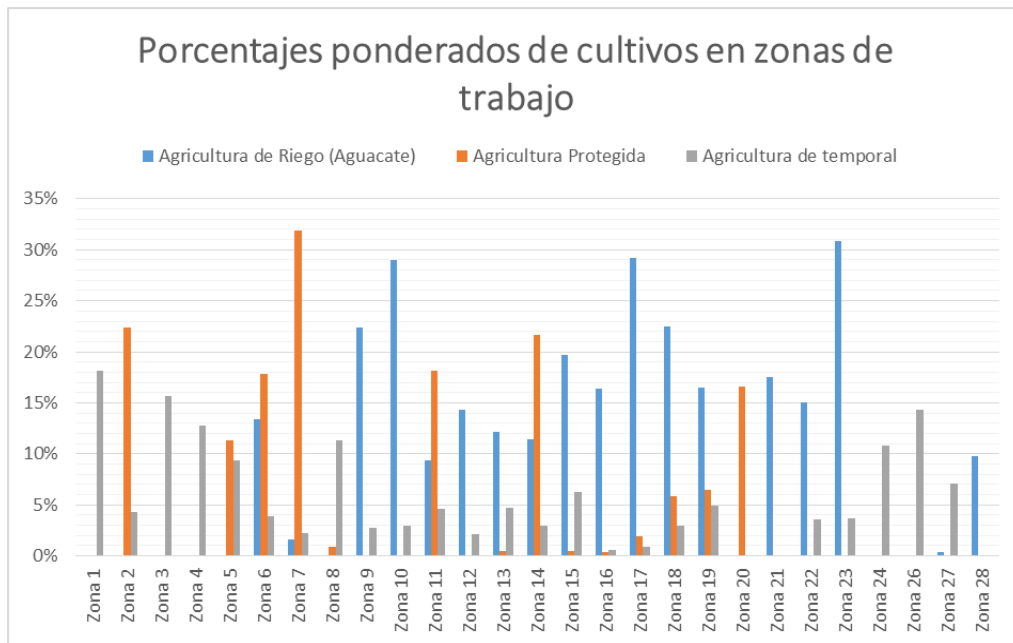


Figura 4.2 Comparativa de zonas de trabajo con porcentajes ponderados

Así mismo, se graficaron los niveles de riesgo identificados de forma general en cada una de las zonas, esto derivado únicamente del análisis de datos, tomando como referencia los porcentajes de la Tabla 3.1, del Apartado 3.8.2 Software de movilidad y predicciones, sin tomar en cuenta los resultados obtenidos del árbol de decisión. En la figura 4.3 se muestra la gráfica, donde los datos del eje vertical representan el nivel de riesgo de la zona, de tal manera que 1 es riesgo bajo, 2 riesgo medio y 3 riesgo alto; y los datos en el eje horizontal representan cada una de las 27 zonas de trabajo.

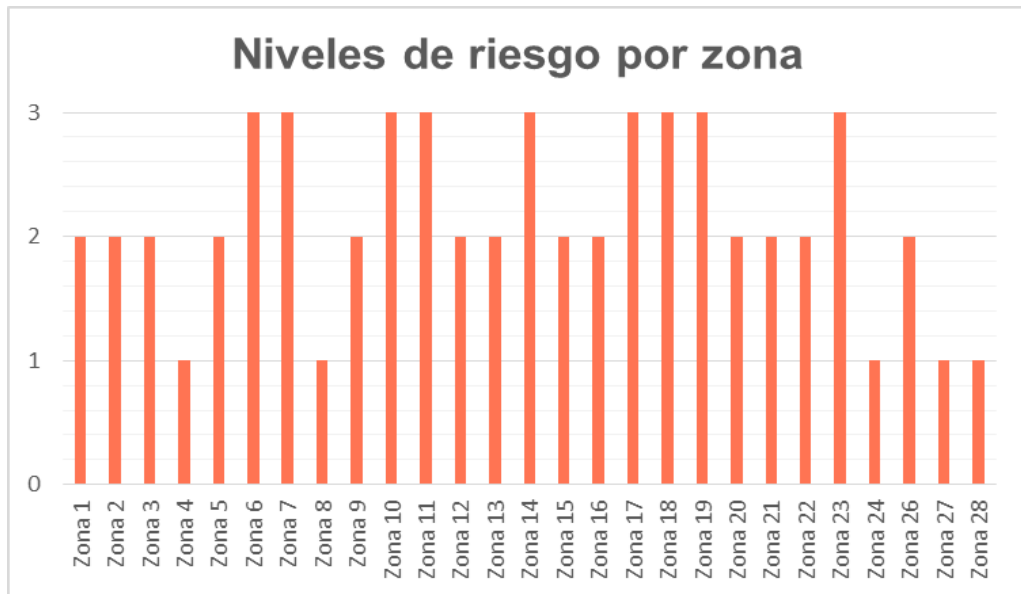


Figura 4.3 Niveles de riesgo por zona

De manera específica, y donde entró en juego el papel del árbol de decisión, fue mediante el análisis de los resultados de las predicciones, tomando como referencia 3 zonas de trabajo en donde el nivel de riesgo en general es considerado como Alto, y para los cuales se analizaron 6 puntos de coordenadas aleatorias ubicadas dentro de la zona, obteniendo resultados derivados de la predicción y que están ciertamente influenciados por el nivel de riesgo de las zonas vecinas, respecto a la zona analizada. Las zonas elegidas fueron la 6, la 19 y la 23; en la figura 4.4 se muestran dichas zonas.



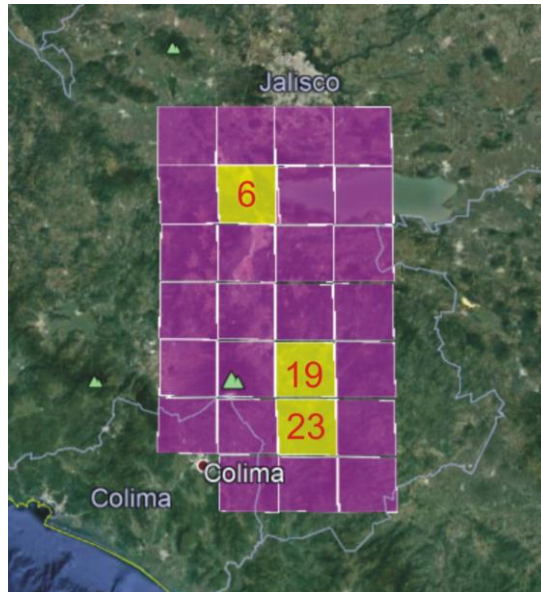


Figura 4.4 Zonas elegidas para el análisis de riesgo alto

Tomando como referencia la figura anterior, a continuación se muestran los datos obtenidos para cada una de las 3 zonas analizadas, donde por una parte se muestra la tabla con las coordenadas de cada uno de los 6 puntos, así como el nivel de riesgo obtenido de la predicción del árbol de decisión, como se aprecia en las tablas 4.1, 4.2 y 4.3. Y por otra parte se muestra la gráfica correspondiente a cada análisis, donde los datos del eje vertical representan el nivel de riesgo de la zona, donde 1 es riesgo bajo, 2 riesgo medio y 3 riesgo alto, tal como aparecen en las figuras 4.5, 4.6 y 4.7; y los datos en el eje horizontal representan cada una de los 6 puntos analizados.

**Tabla 4.1**

*Resultado del análisis de riesgo de la zona 6*

<b>ZONA 6 - RIESGO ALTO</b>			
<b>PUNTO</b>	<b>LATITUD</b>	<b>LONGITUD</b>	<b>RIESGO</b>
<b>1</b>	20.330276	-103.658143	MEDIO
<b>2</b>	20.152513	-103.647802	ALTO
<b>3</b>	20.15226	-103.547572	ALTO
<b>4</b>	20.243235	-103.658651	ALTO
<b>5</b>	20.337459	-103.460731	MEDIO
<b>6</b>	20.239428	-103.449909	ALTO

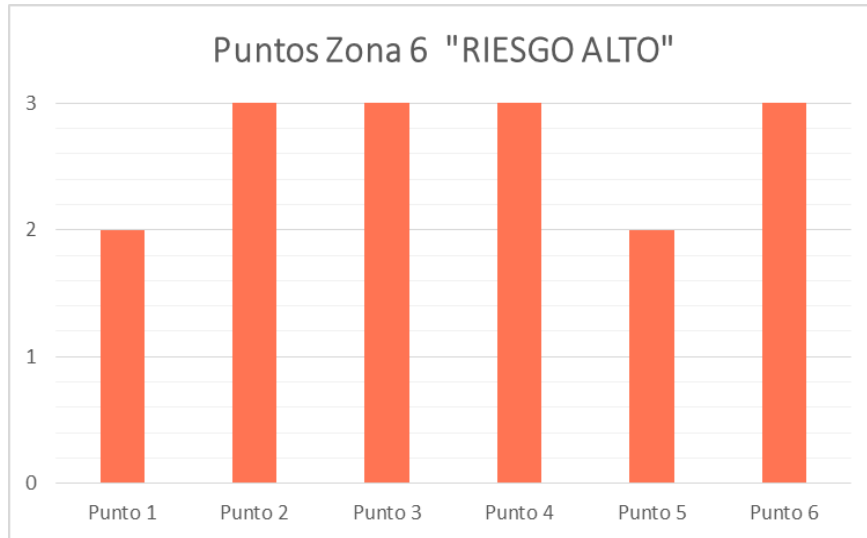


Figura 4.5 Gráfica de puntos zona 6

**Tabla 4.2**

*Resultado del análisis de riesgo de la zona 19*

<b>ZONA 19 - RIESGO ALTO</b>			
<b>PUNTO</b>	<b>LATITUD</b>	<b>LONGITUD</b>	<b>RIESGO</b>
<b>1</b>	19.566606	-103.332474	ALTO
<b>2</b>	19.689195	-103.429575	ALTO
<b>3</b>	19.517091	-103.410274	ALTO
<b>4</b>	19.507927	-103.315199	ALTO
<b>5</b>	19.677549	-103.231265	MEDIO
<b>6</b>	19.513452	-103.227922	MEDIO

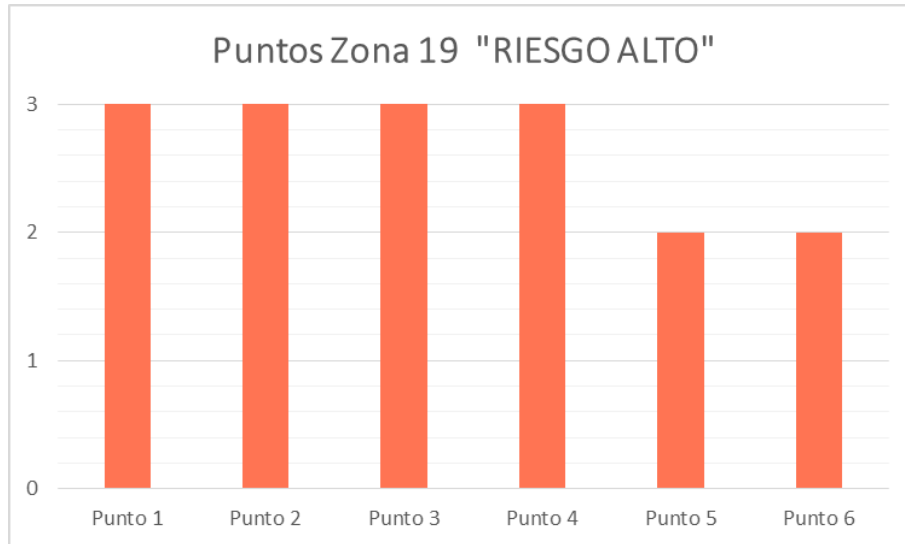
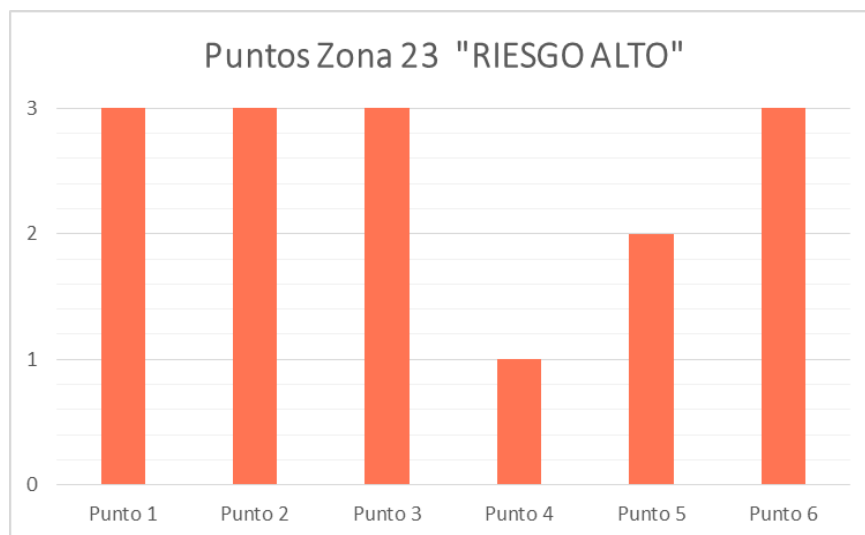


Figura 4.6 Gráfica de puntos zona 19

**Tabla 4.3**

*Resultado del análisis de riesgo de la zona 23*

<b>ZONA 23 - RIESGO ALTO</b>			
<b>PUNTO</b>	<b>LATITUD</b>	<b>LONGITUD</b>	<b>RIESGO</b>
<b>1</b>	19.477226	-103.421627	ALTO
<b>2</b>	19.303934	-103.408282	ALTO
<b>3</b>	19.296921	-103.320519	ALTO
<b>4</b>	19.383061	-103.224217	BAJO
<b>5</b>	19.475717	-103.226251	MEDIO
<b>6</b>	19.481283	-103.314895	ALTO



*Figura 4.7* Gráfica de puntos zona 23

#### **4.4 Evaluación o impacto económico y social**

El proyecto de investigación presenta una alternativa para predecir los niveles de riesgos para las abejas, derivados de algunos de los factores que son considerados de mayor importancia en el ámbito de la mortandad de las abejas, de acuerdo a la literatura y a la información proporcionada por la empresa Frutimielmex. La herramienta computacional buscará generar un impacto económico y social, en los apicultores que la utilicen, siendo la principal beneficiada, la empresa Frutimielmex.

Para la evaluación económica y social, se realizó un estudio de tendencia, donde los resultados indicaron que los ingresos van a la baja y los costos de producción en aumento, por lo que los empresarios se enfrentan al riesgo de pérdida monetaria, de la cual se percatan hasta el momento que la misma está presente. Por tal motivo, se recomienda realizar proyecciones financieras, que anteceden los riesgos económicos en los que se encuentran un gran número de familias mexicanas.

Para el presente estudio se aplicaron encuestas a los productores de miel del estado de Jalisco, en el mes de junio del 2018, en donde se logró una muestra de 76 resultados fidedignos, de los cuales se toman los datos para evaluar el impacto financiero actual,

con la obtención de datos fidedignos de primera mano. En los anexos del documento se presenta el formato de la encuesta aplicada (Anexo I), así como una muestra de 5 encuestas realizadas (Anexo II). De igual manera, es de suma importancia mencionar que la problemática mundial de la muerte de las abejas, se ve reflejada en el estado de Jalisco, y de manera específica, en la empresa Frutimielmex, la cual ha sido afectada de manera significativa por este fenómeno causado principalmente por los productos químicos utilizados en las labores agrícolas. Para reforzar la investigación y los datos mencionados, se aplicó de manera particular la misma encuesta a la empresa a la cual está orientada el proyecto, misma que se integra en el apartado de anexos (Anexo III).

El presidente de los apicultores del Estado de Jalisco, señor Salvador Hernández, informa que número total de colmenas son 280,000, distribuidas en 1,710 empresas registradas en el año 2018.

Se clasifican de la siguiente manera:

- MICRO: de 10 a 200 colmenas que equivalen a 1,250 empresas.
- PEQUEÑAS: de 201 a 800 colmenas que equivalen a 350 empresas.
- MEDIANAS: de 801 hasta 5,000 colmenas que equivalen a 90 empresas.
- GRANDES: más de 5,000 colmenas que equivalen a 20 empresas.

A continuación, en la tabla 4.4 se presentan los resultados con un promedio ponderado de datos básicos obtenidos de las encuestas aplicadas:

**Tabla 4.4**

*Resultados con promedio ponderado*

<b>DATOS ANUALES</b>	MICRO	PEQUEÑAS	MEDIANAS	GRANDES
Colmenas	10 a 200	201 a 800	801 a 5000	más de 5000
Producción por Colmena	21.72	22.73	Sin respuesta	Sin respuesta
Precio de Venta	94.84	99.25	Sin respuesta	Sin respuesta
Costo de produccion	63.94	61.15	Sin respuesta	Sin respuesta
Costos fijos	600	1,650	Sin respuesta	Sin respuesta

El análisis financiero del estudio, se focaliza al sector Micro, que es el de mayor número de empresas, tomando en consideración que cuenta con los datos de mayor riesgo financiero, se calcula con datos que se presentan en la tabla anterior, que se obtuvieron en la muestra, multiplicados por 1,250 representativas de este sector en el estado de Jalisco. En seguida, en la tabla 4.5, se presenta un análisis punto de equilibrio a nivel estatal, para mostrar información del riesgo financiero que generan el sector más vulnerable en el ámbito apícola.

**Tabla 4.5**

*Punto de equilibrio de micro productores de miel*

AÑO 2018	
P. VENTA por kilo de miel Promedio	\$94.84
Producción de miel al año	27150
INGRESOS por venta total anual	\$2,574,906.00
<b>COSTO DE PRODUCCION</b>	
Costo unitario de producción por Kilo de miel	\$63.94
Costos de producción total	\$1,735,971.00
Costos fijos (gasto gral. Anual)	\$750,000.00
<b>COSTO TOTAL ANUAL</b>	<b>\$2,485,971.00</b>
CALCULOS	
COSTO UNITARIO TOTAL (variable+fijo)	\$ 91.56
COSTOS FIJOS ANUALES	\$ 750,000.00
COSTOS VARIABLES	\$ 1,735,971.00
COSTO VARIABLE UNITARIO	\$ 63.94
MARGEN DE CONTRIB. MARGINAL	\$ 30.90
PUNTO DE EQUILIBRIO UNIDADES	24,272
PUNTO DE EQUILIBRIO EN DINERO	\$ 2,301,941.75

En la tabla anterior, se refleja el por qué la apicultura en Jalisco se encuentra con un alto nivel de riesgo, y que en cualquier momento puede caer en déficit, ya que para encontrarse en un punto de equilibrio y poder cubrir los costos fijos y variables se necesita producir 24,272 kilos, y actualmente se producen 27,150 kilos de miel, lo que representa un margen muy ajustado. Así mismo se presenta el punto de equilibrio en dinero a nivel estatal con los micro empresarios de miel, que en promedio el total de ellos deben ingresar \$2'301,941.75 y están ingresando \$2'574,906.00. Estos datos dan origen a asegurar que existe un gran riesgo para el sector mayoritario de los apicultores.

En la figura 4.8, se muestra una gráfica, donde al cruzar los ingresos con los costos totales, en promedio está muy cerca del límite de producción y ventas.

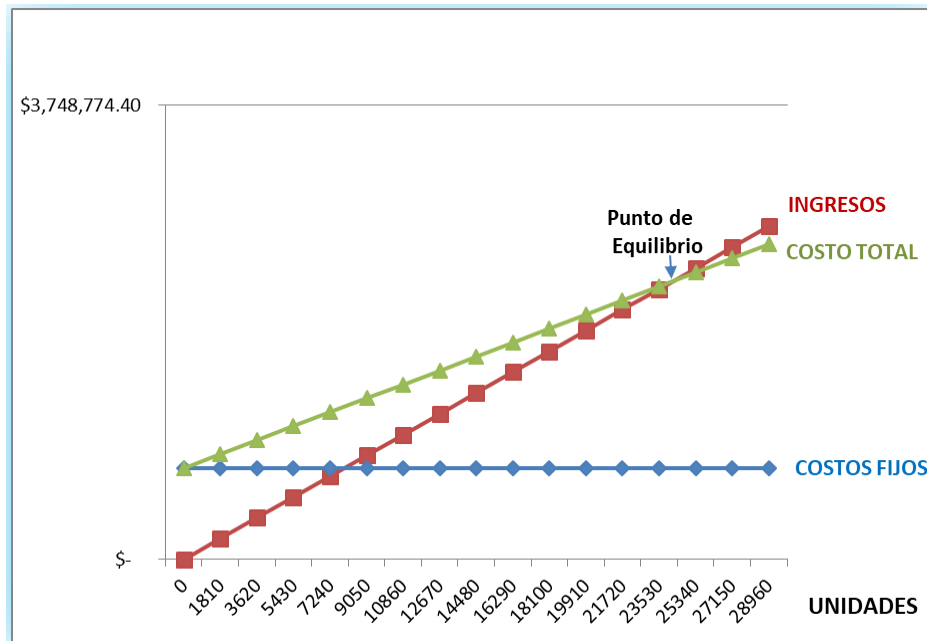


Figura 4.8 Punto de equilibrio en modo gráfico

## 4.5 El software

El software desarrollado en el presente proyecto está dividido en 2 grandes grupos. El primero es el software de control apícola y el segundo es el software de predicciones y movilidad. En los siguientes párrafos se hará referencia a cada uno de ellos.

### 4.5.1 Software de control apícola

El software apícola permite el control administrativo y productivo de la empresa Frutimielmex. Dicho software, fue realizado en conjunto con un grupo de residentes del Instituto Tecnológico de Ciudad Guzmán, los cuales diseñaron y programaron varios de los módulos que forman parte del sistema. Por tal motivo, la totalidad de capturas de pantalla del software no se muestran en el presente documento.

El sistema cuenta con una pantalla de inicio de sesión, para garantizar la seguridad de los usuarios, tal como se muestra en la figura 4.9.

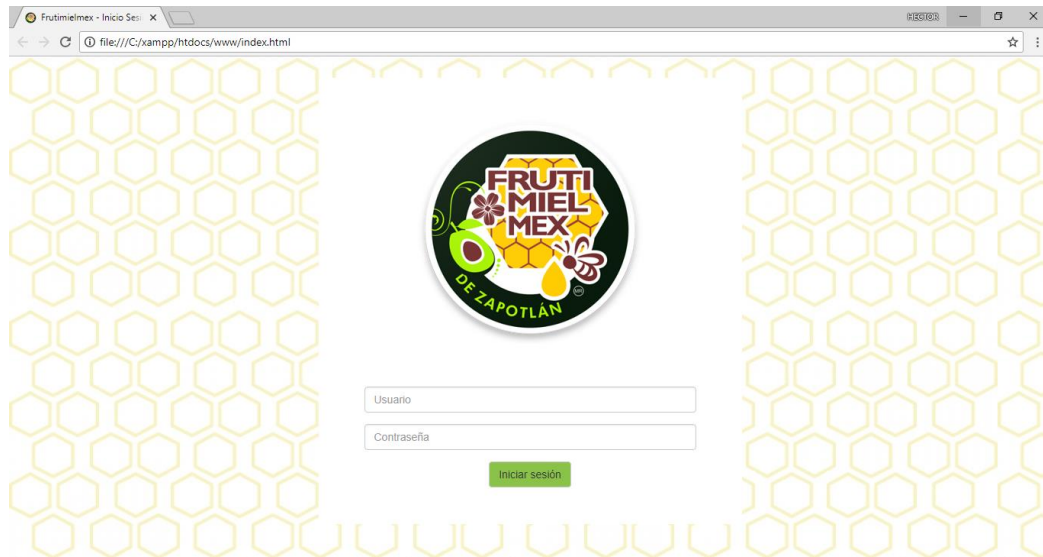


Figura 4.9 Inicio de sesión

Posteriormente, se ofrece el menú principal, donde se visualizan todas las funciones y módulos del sistema, como se aprecia en la figura 4.10. En el cual se puede observar módulos tales como alimentación, medicación, producción, ventas, rentas, apiarios, barriles, entre otros.

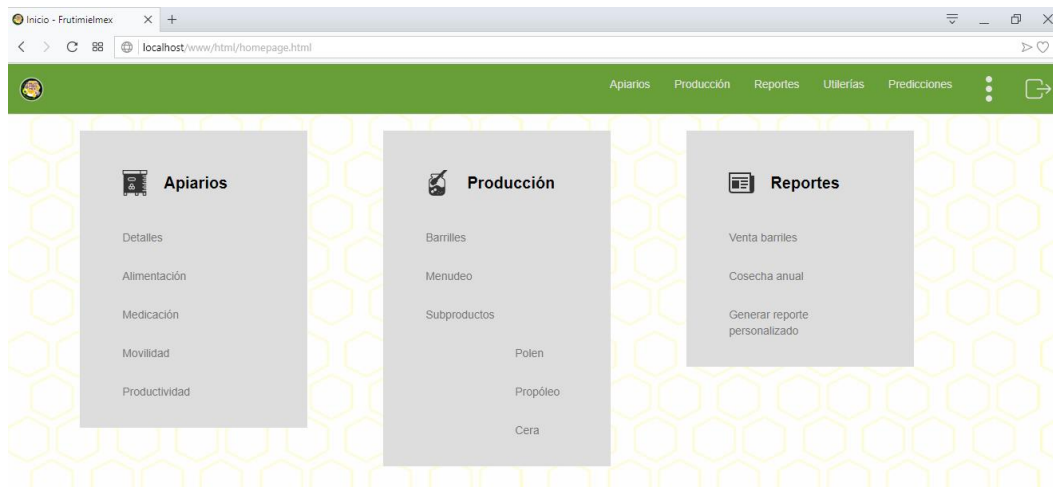


Figura 4.10 Menú principal



Dentro del menú de opciones, existen tres módulos que fueron completamente diseñados y codificados por el autor del presente documento, los cuales fueron: usuarios, código QR, respaldar y recuperar la base de datos.

El módulo usuarios contiene las funciones básicas de agregar, modificar y eliminar un usuario del sistema, con la variante de que el usuario puede ser general o empleado. Si es usuario general, puede ser administrador (control total del sistema) o usuario general (acceso a algunas funciones), pero si es usuario empleado, la diferencia es que su nivel de uso es muy limitado. La interfaz general de gestionar usuarios se muestra en la figura 4.11.

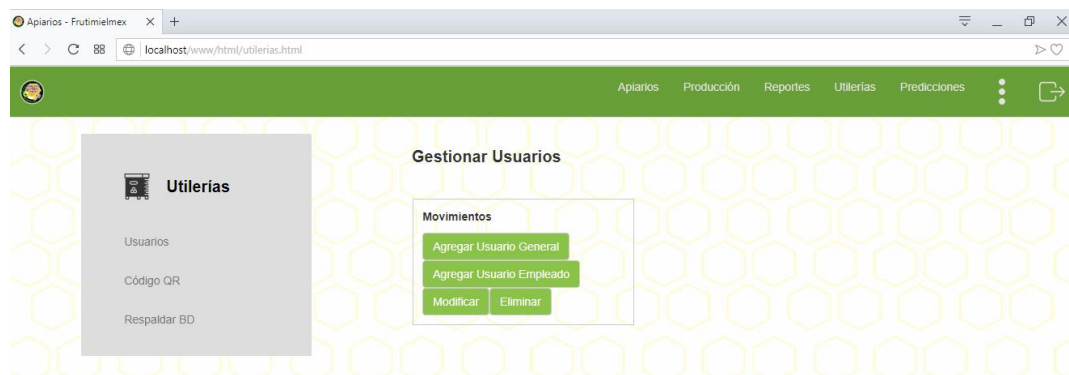


Figura 4.11 Interfaz gestionar usuarios

El módulo generación de códigos QR, consiste en realizar una búsqueda en la base de datos de determinado apiario, elegirlo y generarle una imagen en formato PNG (Portable Network Graphics/Gráfico de Red Portátil), la cual puede ser almacenada por el usuario para su posterior impresión, dado que dicha imagen es colocada a manera de etiqueta en cada uno de los apiarios para ser identificado y leído por un dispositivo móvil para generar las bitácoras diarias. La interfaz del módulo se muestra a continuación en la figura 4.12.

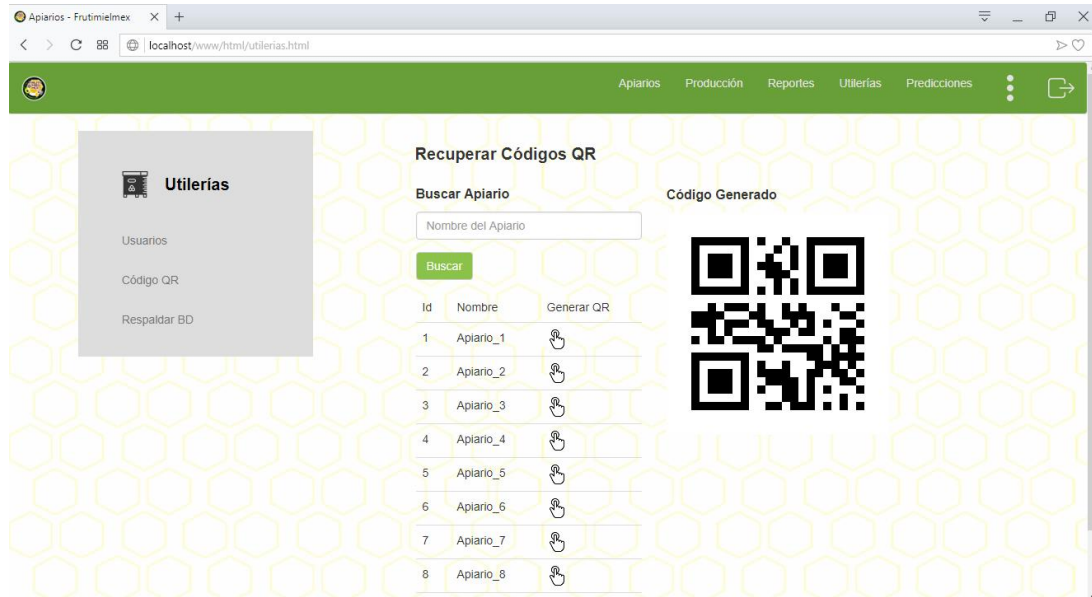


Figura 4.12 Interfaz códigos QR

El módulo para respaldar o restaurar la base de datos, permite realizar un respaldo de forma manual cuando el usuario así lo decida. Además, permite hacerlo a la inversa, es decir, a partir de un respaldo previamente guardado, restaurar la base de datos, con su estructura e información almacenada. En la figura 4.13 se aprecia dicho módulo.



Figura 4.13 Interfaz respaldar/restaurar base de datos

## 4.5.2 Software de predicciones y movilidad

La herramienta computacional de predicciones y movilidad está conformada por 2 módulos:

- Predicciones.
- Movilidad.

El primero de ellos, consiste a grandes rasgos, en mostrar mediante un mapa, la delimitación del área de trabajo, y en la cual el usuario mediante el posicionamiento del marcador en forma de abeja, elige un punto determinado. Al elegir algún punto, se mostrará su información relacionada, como lo es el número de la zona en que se encuentra, latitud y longitud, porcentajes de agricultura protegida, riego y temporal, en relación al total de la zona. Así mismo permitirá evaluar el riesgo en dicho punto, emitiendo el sistema como resultado riesgo alto, medio o bajo. En la figura 4.14, 4.15, 4.16 y 4.17, se muestran las interfaces de la herramienta en dicho módulo.

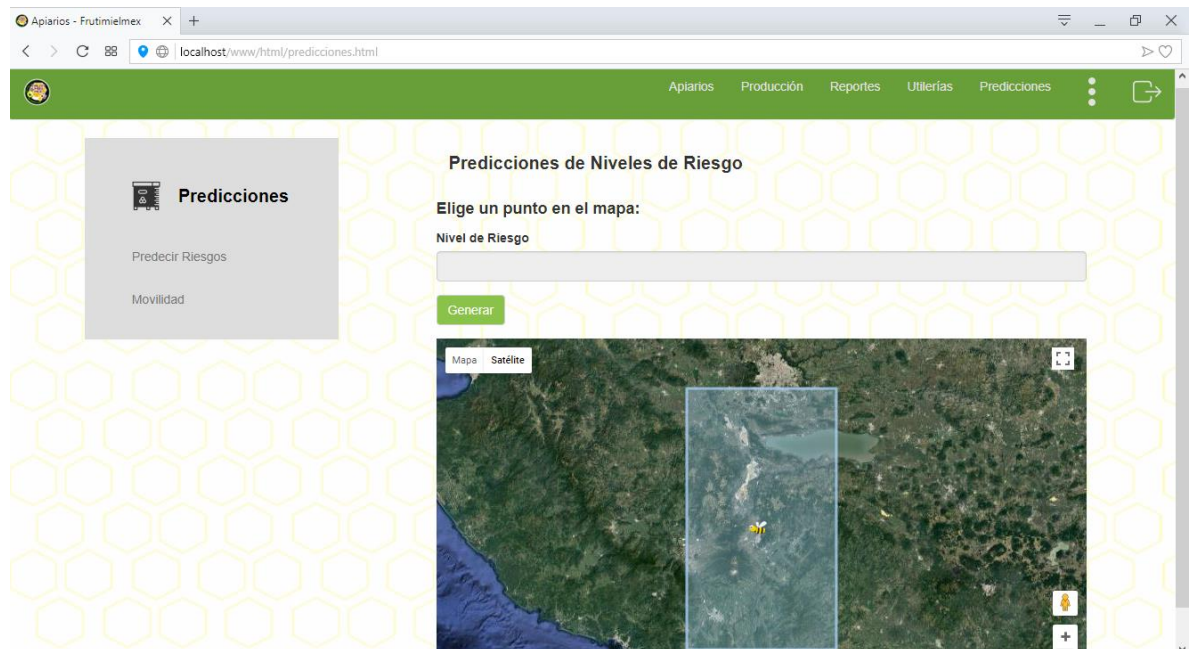


Figura 4.14 Interfaz principal del módulo predicciones

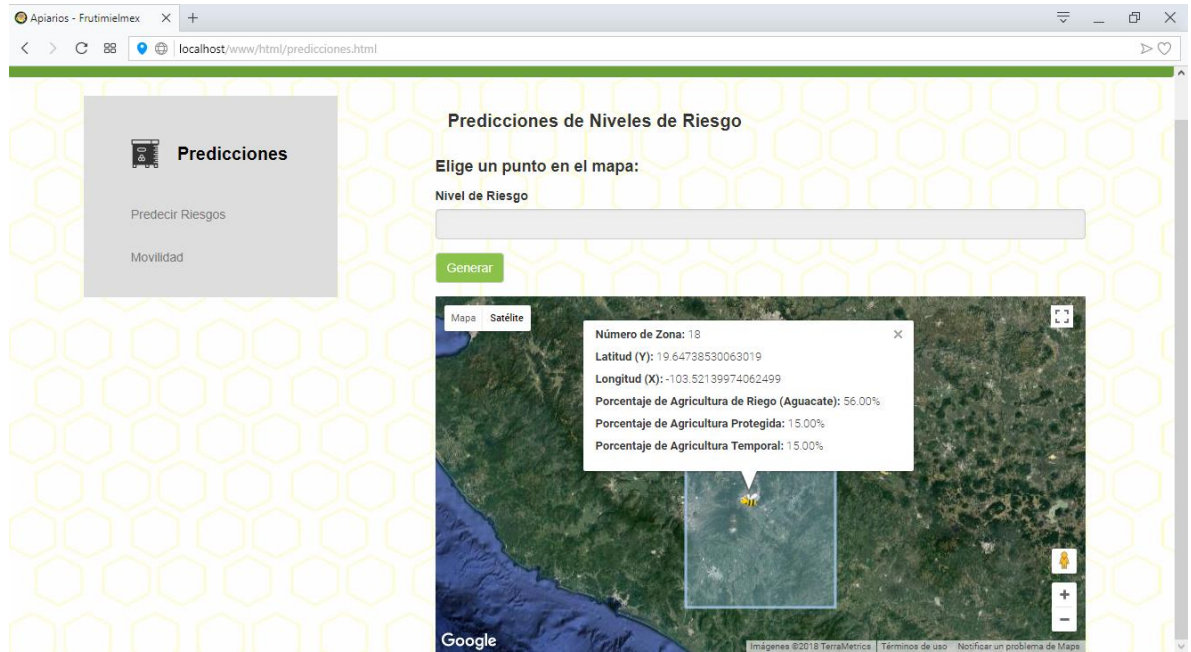


Figura 4.15 Interfaz mostrar información, módulo predicciones

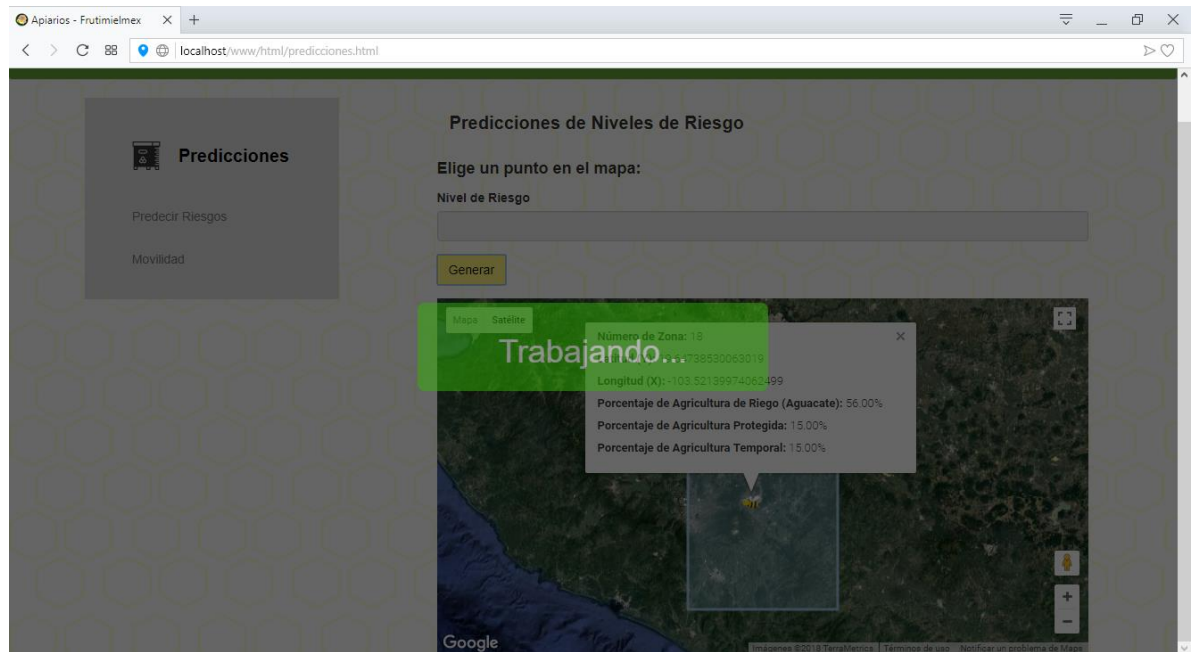


Figura 4.16 Interfaz del proceso de generación de predicción, módulo predicciones

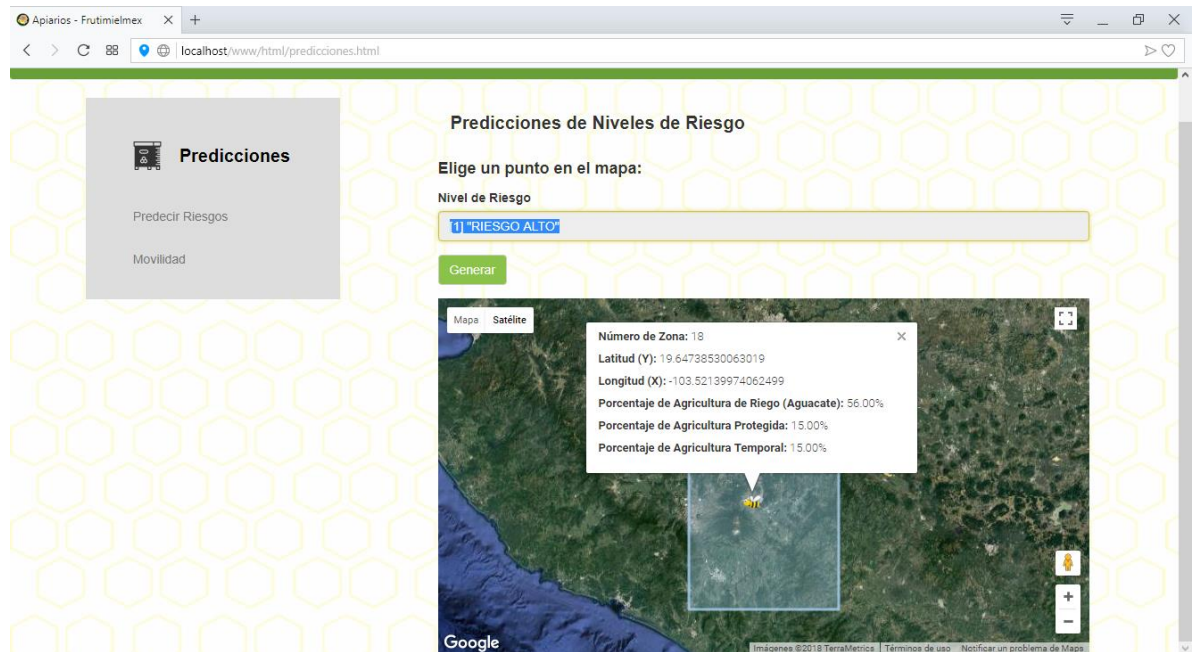


Figura 4.17 Interfaz mostrando resultado de predicción de nivel de riesgo, módulo predicciones

El segundo módulo, denominado movilidad, consiste básicamente en obtener la ubicación de determinado apiario, cargado en un listado obtenido de la base de datos del sistema, para posteriormente obtener la información relevante del mismo, así como la predicción del nivel de riesgo del apiario.

En la figura 4.18 se muestra la interfaz principal del módulo de movilidad, en el cual se despliegan todos los apiarios que se tienen registrados.

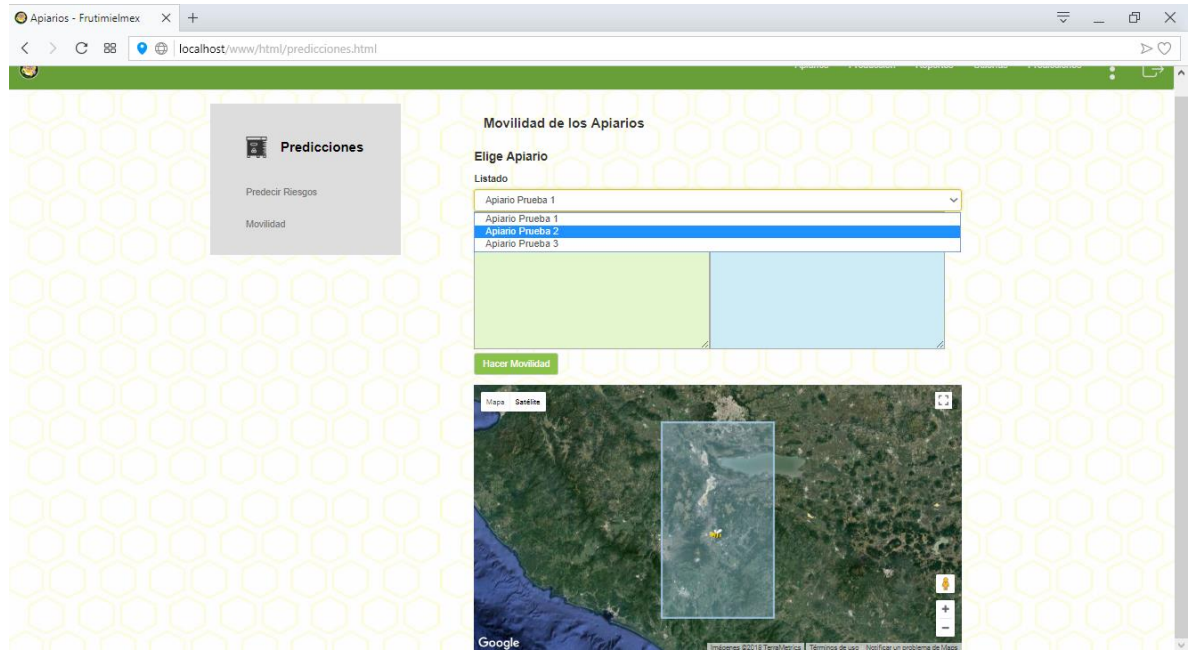


Figura 4.18 Interfaz principal del módulo de movilidad

En la figura 4.19, se ilustra la interfaz donde al elegir determinado apiario, la información de dicho apiario se carga en el recuadro del lado izquierdo en color verde, la cual contiene el número de la zona en que se encuentra, latitud y longitud, porcentajes de agricultura protegida, riego y temporal, en relación al total de la zona, así como la predicción del nivel de riesgo, que en ese caso se ilustra con el marcador en forma de abeja.

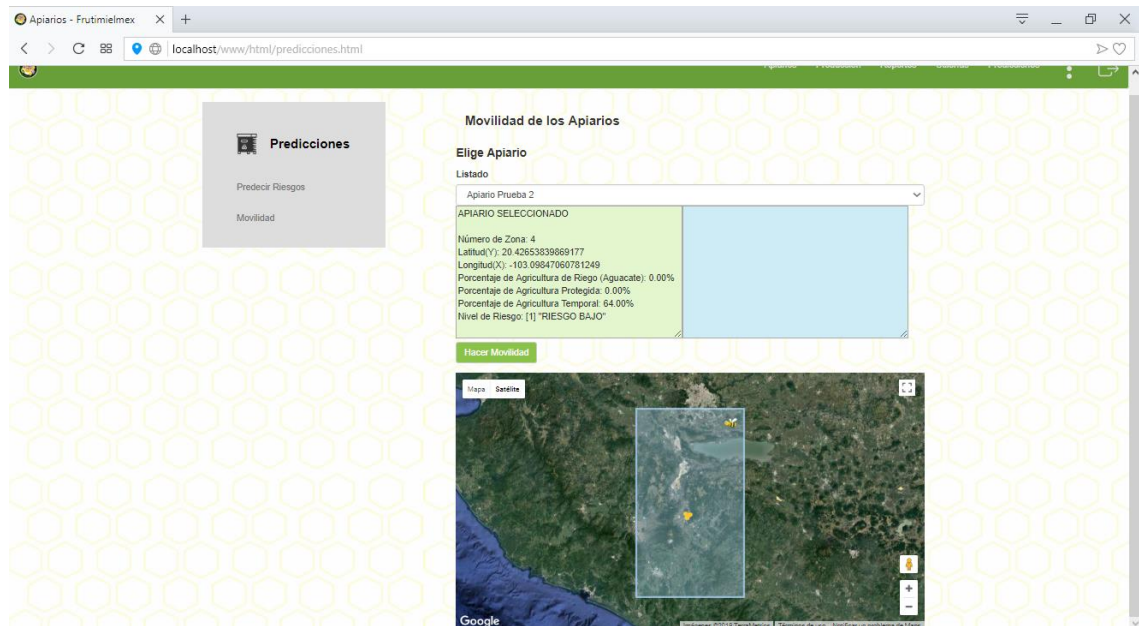


Figura 4.19 Interfaz desplegando información del apiario, del módulo de movilidad

Posteriormente, existe la posibilidad de buscar un nuevo punto para realizar la movilidad del apiario. Para ello se dispone de un marcador en forma de colmena, el cual al situarlo en cualquier punto del área de trabajo, se muestra la misma información del apiario seleccionado, pero con los datos de ese nuevo punto, para que así el usuario realice una comparación de ambos puntos, previo a realizar la movilidad, tal como se aprecia en la figura 4.20.

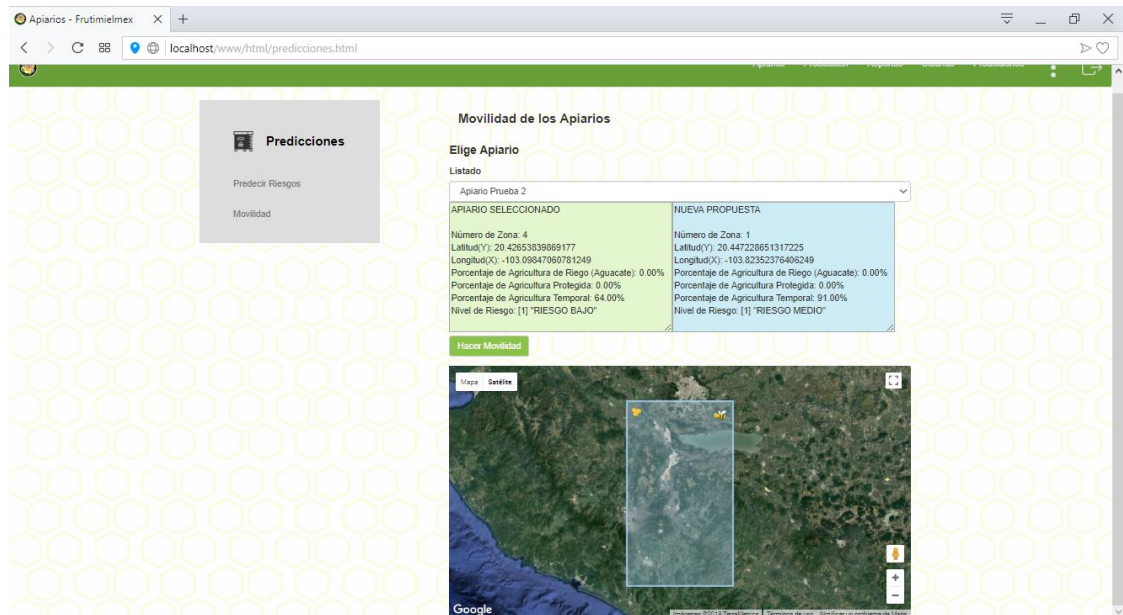


Figura 4.20 Interfaz de comparación de información de dos puntos, del módulo de movilidad

Cuando el usuario decide realizar la movilidad, presiona el botón “Hacer Movilidad”, y se despliega el complemento de información a llenar, para así, llevar a cabo de forma correcta la movilidad del apiario, tal como se muestra en la figura 4.21.

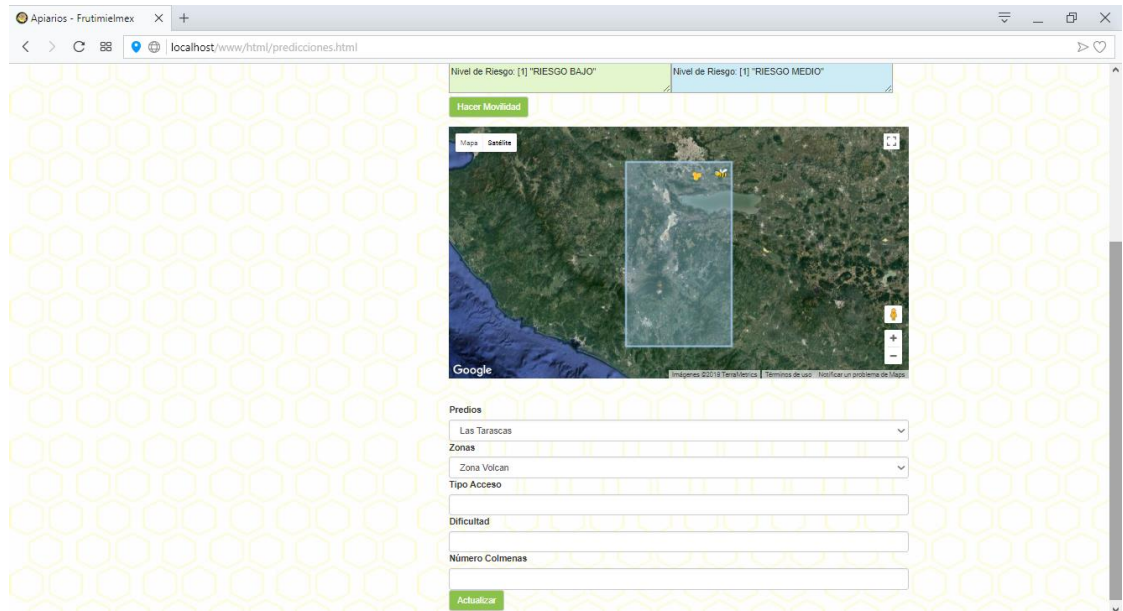


Figura 4.21 Interfaz complemento de información, del módulo movilidad



Para culminar el proceso de movilidad, el usuario presiona el botón “Actualizar”, e internamente, se realiza la actualización de la información en la base de datos, tanto en la tabla apiarios, histórico apiarios y en la tabla de movilidad.

## **Capítulo V. Conclusiones**

### **5.1 Conclusiones**

Con el trabajo de investigación se logró cumplir a detalle el objetivo general de la investigación, el cual es “Aplicar análisis de datos para predecir niveles de riesgo en el ecosistema apícola, al cambiar de ubicación física derivados de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual”, ya que se generó una herramienta computacional que permite evaluar los niveles de riesgo de un determinado punto, además de permitir al usuario la decisión de cambiar de ubicación física un apiario, para buscar un nivel de riesgo menor, en el cual se pretenda disminuir la mortandad de las abejas.

La investigación es innovadora y tiene una aportación importante, ya que previo a realizar la herramienta computacional, se llevó a cabo una revisión exhaustiva del estado del arte, no hallando hasta ese momento una herramienta que cumpliera con los objetivos de la aplicación desarrollada.

Queda abierta la posibilidad de implementar el mismo modelo predictivo, para llevar a cabo investigación en otros ambientes y entornos, y con otro tipo de factores, pudiendo así ser la herramienta pionera en su área, sirviendo de base para generar nuevos modelos y niveles de riesgo.

### **5.2 Recomendaciones**

El presente trabajo de investigación surge de la problemática a nivel mundial que aqueja al entorno apícola, y de manera específica a la contribución en la disminución del índice de mortandad de las abejas de la empresa Frutimielmex. Se recomienda el uso de la herramienta de predicción como tal, como una aplicación de apoyo en la toma de decisiones al cambiar de ubicación física de los apiarios, y queda a plena

consideración del usuario, la decisión final de realizar o no la acción. También es importante considerar que la herramienta computacional de predicción, está basada únicamente en evaluar los niveles de riesgo considerando los factores de la cantidad de hectáreas de cultivos de agricultura protegida (invernaderos, malla sombra y macro túneles), agricultura de riego (cultivo de aguacate) y agricultura de temporal de ciclo anual, y que dichos factores de consideraron de acuerdo a una exhaustiva revisión de la literatura, así como de la información recabada con la empresa Frutimielmex; y por lo tanto, debe tomarse en cuenta que existen otros factores de riesgo que también afectan a las abejas, pero que para este proyecto, no se tomaron en cuenta.

### **5.3 Trabajo futuro**

Dada la investigación realizada, tanto en la literatura, artículos científicos y la información proporcionada por la empresa Frutimielmex, se generó la herramienta de predicción, que de acuerdo a la ubicación de sus apiarios, se delimitó un área de trabajo de 15,552 kilómetros cuadrados en el estado de Jalisco. La herramienta computacional está preparada para que con el análisis de cualquier zona establecida de trabajo, y con la obtención de los factores de riesgo y su respectiva cantidad de hectáreas, se pueda realizar la predicción de los niveles de riesgo. El trabajo futuro está dado para analizar otras zonas, tanto en el mismo estado de Jalisco, como en otros estados de la república mexicana, ya que se cuenta con el sustento de investigación teórico y documental, para llevar a cabo otros análisis.

## Fuentes de información

- Beetight. (2016). *Beetight: Beekeeping software and online hive records*. Recuperado el día 8 de octubre de 2016, de: <https://www.beetight.com/>
- Blacquièrre, T., Smaghe, G., van Gestel, C., y Mommaerts, V. (2012). Neonicotinoids in bees: a review on concentrations, side-effects and risk assessment. *Ecotoxicology*, 21(4), 973-992. Doi: <http://dx.doi.org/10.1007/s10646-012-0863-x>
- Carvajal, L. (2002). *Metodología de la investigación*. Colombia: Faid.
- Consultoría Integral Apícola [Pitecnic]. (2016). *Gestión y trazabilidad apícola*. Recuperado el 3 Mayo 2018, de: <http://tecnocustomer.powweb.com/apitecnic/index.php>
- Díaz, A. (2016). Jalisco enmielado. *Mural*. Recuperado el día 3 Mayo 2018, de: <https://www.mural.com/aplicacioneslibre/articulo/default.aspx?id=783936&md5=1e3b31496020bb9de80f21226c11074e&ta=0dfdbac11765226904c16cb9ad1b2efe>
- El Informador. (2018). *Jalisco, tercer lugar nacional como productor de miel*. Recuperado el día 2 de mayo de 2018, de: <http://www.informador.com.mx/economia/2016/679301/6/jaliscotercer-lugar-nacional-como-productor-de-miel.htm>
- Franco-Arcega, A., Carrasco-Ochoa, J., Sánchez-Díaz, G., y Martínez-Trinidad, J. (2013). *Decision Tree based Classifiers for Large Datasets*. Recuperado de: <http://www.cys.cic.ipn.mx/ojs/index.php/CyS/article/view/1485/1495>
- Hernández, S. R., Baptista, L. P., y Fernández, C. C. (2010). *Metodología de la investigación*. pp. 170-195. México: McGraw-Hill Interamericana.
- Instituto Internacional de Ciencias de Datos. (2017). *Algunas consideraciones sobre la analítica de datos o "data analytics"*. Recuperado el día 29 de junio de 2017, de: <http://www.i2ds.org/algunas-consideraciones-sobre-la-analitica-de-datos-o-data-analytics/>

International Business Machines Corporation [IBM]. (2013). *Analítica de datos: un proyecto de generación de valor*. Recuperado el día 26 de septiembre de 2016, de: [http://www.ibm.com/midmarket/es/es/att/pdf/Analitica\\_de\\_datos\\_para\\_pymes.pdf](http://www.ibm.com/midmarket/es/es/att/pdf/Analitica_de_datos_para_pymes.pdf)

International Business Machines Corporation [IBM]. (2018). Knowledge Center. Recuperado el día 3 de mayo de 2018, de: [https://www.ibm.com/support/knowledgecenter/es/SS3RA7\\_sub/modeler\\_mainhelp\\_client\\_ddita/cle](https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/cle)

International Business Machines Corporation [IBM]. (2016). *¿Qué es Big Data?* Recuperado el día 27 de septiembre de 2016, de: <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/>

International Business Machines Corporation [IBM]. (2018). Knowledge Center. Recuperado el día 2 de mayo de 2018, de: [https://www.ibm.com/support/knowledgecenter/en/SSEPGG\\_9.7.0/com.ibm.im.overview.doc/c\\_naive\\_bayes\\_classification.html](https://www.ibm.com/support/knowledgecenter/en/SSEPGG_9.7.0/com.ibm.im.overview.doc/c_naive_bayes_classification.html)

International Business Machines Corporation [IBM]. (2018). Knowledge Center. Recuperado el día 2 de mayo de 2018, de: [https://www.ibm.com/support/knowledgecenter/es/SSFPJS\\_8.6.0/com.ibm.wbpm.admin.doc/topics/cbre\\_busiru\\_decisiontable.html](https://www.ibm.com/support/knowledgecenter/es/SSFPJS_8.6.0/com.ibm.wbpm.admin.doc/topics/cbre_busiru_decisiontable.html)

JavaScript. (2018). *Documentación web de MDN*. Recuperado el día 30 Abril 2018, de: <https://developer.mozilla.org/es/docs/Learn/JavaScript>

Matsudo, N. L. (2001). *Árboles de decisión, una técnica de Data Mining, desde una perspectiva informática y estadística* (Tesis Licenciatura). Recuperado de <https://www.dc.uba.ar/academica/tesis-de-licenciatura/2001/>

Mitchell, E., Mulhauser, B., Mulo, M., Mutabazi, A., Glauser, G., y Aebi, A. (2017). A worldwide survey of neonicotinoids in honey. *Science*, 358(6359), 109-111. Doi: <https://dx.doi.org/10.1126/science.aan3684>

- Métodos de Boosting | Grupo de Gestión y Procesamiento de la Información. (2018). Recuperado el día 15 de mayo de 2018, de: <http://g2pi.tsc.uc3m.es/es/Boosting-es>
- Oracle México. (2018). *MySQL. La base de datos de código abierto más popular*. Recuperado el día 3 de mayo de 2018, de: <https://www.oracle.com/mx/mysql/index.html>
- Patel, N., y Upadhyay, S. (2012). Study of Various Decision Tree Pruning Methods with their Empirical Comparison in WEKA. *International Journal Of Computer Applications*, 60(12), 20-25. Doi: <http://dx.doi.org/10.5120/9744-4304>
- Pressman, R., Campos Olguín, V., Enríquez Brito, J., Villegas Quezada, C., & Ferro Castro, B. (2010). *Ingeniería del software* (7ma ed.). México: The McGraw-Hill.
- PHP. (2018). *Manual de PHP*. Recuperado el día 30 Abril 2018, de: <http://php.net/manual/es/>
- Puga, J. (2012). Cómo construir y validar redes bayesianas con Netica. Recuperado el día 2 de mayo de 2018, de: <https://dialnet.unirioja.es/servlet/articulo?codigo=4107401>
- R Foundation. (2018). The R Project for Statistical Computing. Recuperado el 30 Abril 2018, de: <https://www.r-project.org/>
- Ran, C., y Wang, X. (2015). Building a Decision Tree Model for Campus Information Score Based on the Algorithm C5.0. *Applied Mechanics And Materials*, (719-720), 805-811. Doi: <http://dx.doi.org/10.4028/www.scientific.net/amm.719-720.805>
- Rzedowski, J. (1991). Diversidad y orígenes de la flora fanerogámica de México. *Acta Botánica Mexicana*, 14, 3-21. Recuperado de: <http://www.redalyc.org/pdf/574/57401402.pdf>
- Schmelkes, C. (1998). *Manual para la presentación de anteproyectos e informes de investigación*. Primera Edición. México: Oxford University Press México.
- Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación [SAGARPA/SENASICA]. (2015). *Manual de buenas prácticas pecuarias en la producción de miel*. 3a edición. Ciudad de México.

- Secretaría de Desarrollo Rural [SEDER]. (2018). *Jalisco: Gigante Agroalimentario, Edición 2014*. Recuperado el día 2 de mayo de 2018, de: <http://seder.jalisco.gob.mx/content/jalisco-gigante-agroalimentario>
- Song, Y. Y., y Lu, Y. (2015). Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatry*, 27. 130–135.
- TRACTUS. (2016). *Software de gestión apícola para la trazabilidad de la miel, polen y demás productos apícolas*. Recuperado del día 2 de octubre de 2016, de: <http://tractus.es/miel.html>
- Weka 3. (2018). *Data Mining with Open Source Machine Learning Software in Java Cs.waikato.ac.nz*. Recuperado el día 30 de abril de 2018, de: <https://www.cs.waikato.ac.nz/ml/weka/>

## **Glosario**

### **C**

Clustering: es una técnica de minería de datos dentro de la disciplina de Inteligencia Artificial, que identifica de forma automática agrupaciones o clústeres de elementos de acuerdo a una medida de similitud entre ellos.

Colmena: habitación proporcionada a las abejas por el hombre.

Cookies: es una pequeña información enviada por un sitio web y almacenado en el navegador del usuario, de manera que el sitio web puede consultar la actividad previa del usuario.

### **H**

Hoja: componente de un árbol de decisión, que representa un valor de la variable de destino dados los valores de las variables de entrada representados por el camino desde la raíz a la hoja.

### **M**

Mockup: es un boceto básico y de baja calidad del desarrollo de una página web o el diseño de una interfaz. Su finalidad es mostrar al cliente un diseño o boceto rápido y facilitar la comunicación entre cliente y desarrollador.

Mouse: es un dispositivo de entrada portátil que controla el cursor o el puntero de una pantalla de computadora, junto con la forma en que se mueve en una superficie plana.

### **N**

Nodo: es uno de los elementos de un árbol de decisión, y se representan con un cuadro o un círculo con líneas que están conectados hasta lograr el resultado deseado.

### **P**

Piquera: cualquier abertura de la colmena que permita el paso de las abejas.






## **R**

Regresión: retroceso en el transcurso de un proceso o acción.

## Anexos

### I. Formato de encuesta aplicada.

SECRETARÍA DE EDUCACIÓN PÚBLICA  
TECNOLOGICO NACIONAL DE MÉXICO  
INSTITUTO TECNOLÓGICO DE CIUDAD GUZMÁN

**ENCUESTA**

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F \_\_\_ M \_\_\_ Localidad: \_\_\_\_\_

**Preguntas:**

1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?  
\_\_\_\_\_

2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?  
( ) 10 a 200 ( ) 201 a 800 ( ) 801 a 5000 ( ) más de 5000

3.- ¿Cuál es su producción en Kilos por colmena actualmente? \_\_\_\_\_

4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?  
Año 2015 \_\_\_\_\_ 2016 \_\_\_\_\_ 2017 \_\_\_\_\_

5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ \_\_\_\_\_

6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?  
Año 2015 \_\_\_\_\_ 2016 \_\_\_\_\_ 2017 \_\_\_\_\_

7.- ¿Cuál es el costo de producción por colmena actualmente al año? \_\_\_\_\_

8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?  
Año 2015 \_\_\_\_\_ 2016 \_\_\_\_\_ 2017 \_\_\_\_\_

9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$ \_\_\_\_\_

10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI \_\_\_\_\_ NO \_\_\_\_\_

11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI \_\_\_\_\_ NO \_\_\_\_\_

12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? \_\_\_\_\_

13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?  
Enfermedades \_\_\_\_\_ Pesticidas \_\_\_\_\_ Cambio climático \_\_\_\_\_  
Otras \_\_\_\_\_

**Gracias por su cooperación!**

## II. Muestra de 5 encuestas aplicadas.



TECNOLÓGICO NACIONAL DE MÉXICO  
INSTITUTO TECNOLÓGICO DE CIUDAD GUZMÁN



### ENCUESTA

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F \_\_\_ M  Localidad: San Martín

**Preguntas:**

1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?

Apicultura

2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?

1 a 200 ( ) 201 a 800 ( ) 801 a 5000 ( ) más de 5000

3.- ¿Cuál es su producción en Kilos por colmena actualmente? 20

4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?

Año 2015 25 2016 25 2017 25

5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ 90

6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?

Año 2015 100 2016 95 2017 95

7.- ¿Cuál es el costo de producción por colmena actualmente al año? 70

8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?

Año 2015 60 2016 60 2017 70

9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$     

10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI \_\_\_ NO

11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI  NO \_\_\_

12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? 40%

13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?

Enfermedades  Pesticidas  Cambio climático \_\_\_

Otras \_\_\_

**Gracias por su cooperación!**

### ENCUESTA

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F \_\_\_ M  Localidad: Zapotlán

**Preguntas:**

1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?

Apicultor

2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?

10 a 200    ( ) 201 a 800    ( ) 801 a 5000    ( ) más de 5000

3.- ¿Cuál es su producción en Kilos por colmena actualmente? 20

4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?

Año    2015 15    2016 15    2017 15

5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ 80

6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?

Año    2015 90    2016 90    2017 85

7.- ¿Cuál es el costo de producción por colmena actualmente al año? 70

8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?

Año    2015 65    2016 65    2017 70

9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$ \_\_\_\_\_

10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI \_\_\_\_\_ NO

11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI  NO \_\_\_\_\_

12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? 15%

13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?

Enfermedades  Pesticidas  Cambio climático \_\_\_\_\_

Otras \_\_\_\_\_

**Gracias por su cooperación!**



**ENCUESTA**

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F  M  Localidad: Jamay

**Preguntas:**

- 1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?  
Apicultura
- 2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?  
( ) 10 a 200    () 201 a 800    ( ) 801 a 5000    ( ) más de 5000
- 3.- ¿Cuál es su producción en Kilos por colmena actualmente? 22
- 4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?  
Año    2015 28                      2016 23                      2017 20
- 5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ 90
- 6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?  
Año    2015 80                      2016 80                      2017 90
- 7.- ¿Cuál es el costo de producción por colmena actualmente al año? 70
- 8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?  
Año    2015 60                      2016 60                      2017 70
- 9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$ \_\_\_\_\_
- 10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI \_\_\_\_\_ NO
- 11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI  NO \_\_\_\_\_
- 12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? 20%
- 13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?  
Enfermedades \_\_\_\_\_ Pesticidas  Cambio climático   
Otras \_\_\_\_\_

**Gracias por su cooperación!**



**ENCUESTA**

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F  M  Localidad: Cdi

**Preguntas:**

1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?

APIGULTOR

2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?

( ) 10 a 200 (X) 201 a 800 ( ) 801 a 5000 ( ) más de 5000

3.- ¿Cuál es su producción en Kilos por colmena actualmente? 20

4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?

Año 2015 25 2016 22 2017 21

5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ 90

6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?

Año 2015 80 2016 85 2017 90

7.- ¿Cuál es el costo de producción por colmena actualmente al año? 70

8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?

Año 2015 60 2016 65 2017 65

9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$ \_\_\_\_\_

10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI  NO

11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI  NO

12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? 30%

13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?

Enfermedades \_\_\_\_\_ Pesticidas  Cambio climático \_\_\_\_\_

Otras \_\_\_\_\_

**Gracias por su cooperación!**



**ENCUESTA**

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F  M  Localidad: Atotonilco

**Preguntas:**

1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?

Productor

2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?

( ) 10 a 200    ( ) 201 a 800    ( ) 801 a 5000    ( ) más de 5000

3.- ¿Cuál es su producción en Kilos por colmena actualmente? 25

4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?

Año    2015 22    2016 23    2017 22

5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ 95

6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?

Año    2015 85    2016 95    2017 95

7.- ¿Cuál es el costo de producción por colmena actualmente al año? 75

8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?

Año    2015 70    2016 75    2017 75

9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$ \_\_\_\_\_

10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI \_\_\_\_\_ NO

11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI  NO \_\_\_\_\_

12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? 32%

13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?

Enfermedades  Pesticidas  Cambio climático \_\_\_\_\_

Otras \_\_\_\_\_

**Gracias por su cooperación!**

### III. Encuesta aplicada a la empresa Frutimielmex.



TECNOLÓGICO NACIONAL DE MÉXICO  
INSTITUTO TECNOLÓGICO DE CIUDAD GUZMÁN



#### ENCUESTA

**Objetivo:** Recabar datos para una investigación relacionada con el impacto económico en que se encuentran actualmente las empresas apícolas en Jalisco.

**Instrucciones:** Se le pide de la manera más atenta tenga a bien contestar con datos aproximados a la realidad las siguientes preguntas (la encuesta es anónima y personal).

**Datos generales:** Sexo: F  M  Localidad: Cd. Guzmán Jal.

**Preguntas:**

- 1.- ¿Cuál es su actividad actualmente en la industria de la apicultura?  
Administrador de empresa apícola
- 2.- Si trabaja con colmenas, ¿Cuántas son las que atiende actualmente?  
( ) 10 a 200 ( ) 201 a 800 ( ) 801 a 5000  más de 5000
- 3.- ¿Cuál es su producción en Kilos por colmena actualmente? 15
- 4.- ¿Cuál ha sido su producción en Kilos por colmena en promedio los últimos 3 años?  
Año 2015 20 2016 18 2017 15 *antes durante 50 años el promedio es de 20 kg por colmena*
- 5.- ¿Cuál es el precio de venta por kilo que maneja actualmente? \$ 90 *60 mayoreo menor*
- 6.- ¿Cuál ha sido el precio de venta en los últimos 3 años?  
Año 2015 <sup>May 50</sup> Men 80 2016 <sup>May 55</sup> Men 80 2017 <sup>May 60</sup> Men 90
- 7.- ¿Cuál es el costo de producción por colmena actualmente al año? 1000 pesos
- 8.- ¿Cuál ha sido el costo de producción por colmena los últimos 3 años?  
Año 2015 900 2016 800 2017 1000
- 9.- Además de los costos de producción ¿tiene costos fijos anuales en su empresa y a cuánto ascienden aproximadamente al año? \$ ya están incluidos
- 10.- ¿Su actividad como apicultor le permite tener una empresa con suficientes utilidades para sus necesidades básicas personales y familiares? SI \_\_\_\_\_ NO
- 11.- ¿Ha tenido problemas de mortandad de abejas en sus apiarios? SI  NO \_\_\_\_\_
- 12.- ¿Qué porcentaje de mortandad de abejas ha sufrido en los últimos años? 30%
- 13.- ¿Cuáles considera que han sido las causas principales que provocan esta mortandad?  
Enfermedades \_\_\_\_\_ Pesticidas  Cambio climático \_\_\_\_\_  
Otras \_\_\_\_\_

**Gracias por su cooperación!**