

# INTELIGENCIA DE NEGOCIOS APLICADA: CASO REGIÓN HUETAR NORTE DE COSTA RICA



**Marlen Treviño**

Departamento de Computación  
Instituto Tecnológico de Costa Rica, ITCR, Sede San Carlos  
+506 2401-3143  
[mtrevino@itcr.ac.cr](mailto:mtrevino@itcr.ac.cr)



**Lorena Valerio**

Departamento de Computación  
Instituto Tecnológico de Costa Rica, ITCR, Sede San Carlos  
+506 2401-3114  
[lvalerio@itcr.ac.cr](mailto:lvalerio@itcr.ac.cr)



**Óscar López**

Departamento de Computación  
Instituto Tecnológico de Costa Rica, ITCR, Sede San Carlos  
+506 2401-3117  
[olopez@itcr.ac.cr](mailto:olopez@itcr.ac.cr)

## Resumen

*La inteligencia de negocios es una estrategia para la administración del conocimiento mediante el análisis de los datos de la empresa. Las soluciones de BI deben ser accesibles para una amplia gama de usuarios, por lo que deben aportar al mercado una combinación atractiva de características, funciones y servicios. Entre los componentes más importantes se destaca los datawarehouse, las herramientas de consultas y la minería de datos. En la Región Huetar Norte de Costa Rica, el Sistema de Información Regional (SIR-ZEE) se ha dado a la tarea de integrar las técnicas de BI, en busca de poder monitorear las actividades de la Región para que la información pueda verse en línea y a partir de indicadores sea posible tomar decisiones. Por ello, en el presente artículo se muestra el impacto y beneficio de BI mediante una serie de ejemplos que puedan ser utilizados en potenciar el desarrollo económico local.*

### **Palabras clave**

*Inteligencia de negocios, pymes, desarrollo económico local, software libre.*

## 1. INTRODUCCIÓN

Vivimos en una época en que la información es la clave para obtener una ventaja competitiva. Aunado a este aspecto, se encuentra el ritmo acelerado que presentan los negocios, por ello en la década de 1990, surgieron una serie de herramientas automáticas con un alto grado de confiabilidad para apoyar la toma de decisiones. A este conjunto de herramientas, se les denominó Sistemas de Inteligencia de Negocios (BI, por sus siglas en inglés), las que están enfocadas en la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa [1].

La raíz conceptual de la Inteligencia de Negocios puede relacionarse con el libro “El arte de la guerra” de Sun Tzu. En este libro, se indica que: “para poder ser exitoso en la guerra, se debe conocer completamente las fortalezas y debilidades de uno mismo, así como las del enemigo. El no saber alguna de ellas podría significar la derrota” [2].

Tomando como base la premisa anterior, se diseñaron una serie de componentes para las soluciones de BI. Sin embargo, por el gran número de componentes que existen para dichas plataformas, se debe tener en consideración que estos sistemas al menos deben contar con los siguientes aspectos: (1) Multidimensionalidad, (2) Minería de datos, (3) Alertas, (4) Datawarehouse, (5) Sistemas de extracción, transformación y carga de datos (ETL, por sus siglas en inglés) y (6) Herramientas de consultas [3].

Por otro lado, entre los principales beneficios que ofrece BI se encuentra el facilitar el proceso de transformación de los datos en información, la información en conocimiento y el conocimiento en acción en el momento preciso. De este modo, surge el concepto de “La Cadena del Valor de la Inteligencia”, que está formado por cuatro componentes esenciales: (I) Datos, (II) Información, (III) Conocimiento e (IV) Inteligencia [4].

Con el fin de mostrar el impacto y beneficio de las tecnologías de BI, en el presente artículo se muestran ejemplos de uso de los principales componentes de la plataforma PENTAHO, con información correspondiente a la Región Huetar Norte de Costa Rica. De este modo, se pretende que los resultados proporcionados por dichos componentes sirvan como insumo en las diferentes fases de la cadena de valor de la inteligencia.

El resto del artículo se distribuye en cuatro secciones adicionales. La Sección 2 presenta la definición de los principales componentes que conforman una solución de BI. La Sección 3 describe el problema, el objetivo del estudio y el diseño metodológico de la investigación. En la Sección 4 se presentan los resultados del estudio. Finalmente, las conclusiones y trabajos futuros en la Sección 5 cierran el artículo.

## **2. INTELIGENCIA DE NEGOCIOS Y SUS COMPONENTES**

El mercado tradicional de BI se ha centrado principalmente en los analistas o usuarios con poder, que sólo representan un pequeño subgrupo en cualquier organización tradicional. Sin embargo, el objetivo de toda estrategia de BI a lo largo y ancho de cualquier organización es proporcionar la información correcta a la gente correcta y en el tiempo preciso para soportar los procesos de toma de decisiones en todos los niveles de la organización.

Para abordar este reto, las soluciones de BI deben ser más accesibles para una más amplia gama de usuarios, mediante una mayor visualización de datos, mejor interactividad, adición de acceso a contenidos y funcionalidad de texto presentada en una interfaz de búsqueda conocida que soporte la colaboración y el flujo de trabajo o análisis dirigido.

Las soluciones actuales de BI aporta al mercado combinaciones atractivas de características, funciones y servicios que proveen a las organizaciones cliente las herramientas necesarias para lanzar o continuar con sus esfuerzos hacia un buen desempeño administrativo. Por lo que, en esta sección se describen los principales componentes que debe poseer un sistema de BI.

### **2.1. Multidimensionalidad**

El objetivo del análisis multidimensional es ganar comprensión en el conocimiento contenido en las bases de datos. Su principal ventaja es que facilitan los análisis complejos y la visualización de los datos para procesos de toma de decisiones, reduciendo la confusión y disminuyendo las interpretaciones erróneas. Además, la velocidad de las operaciones es varias veces superior y más consistente de lo que es posible en otras estructuras de bases de datos [3].

### **2.2. Datawarehouse**

Se llama datawarehouse al almacén de datos que reúne la información histórica generada por todos los distintos departamentos de una organización, orientada a consultas complejas y de alto rendimiento. Un datawarehouse tiene una orientación corporativa que pretende conseguir que cualquier departamento pueda acceder a la información de cualquiera de los otros mediante un único medio, así como obligar a que los mismos términos tengan el mismo significado para todos. [3].

### **2.3. Minería de datos**

La minería de datos (Data Mining, en inglés) es una metodología que posibilita a los trabajadores del conocimiento descubrir tendencias futuras y patrones ocultos y detectar con anterioridad insospechadas correlaciones en la información. Es una vía muy importante de aprovechamiento de la información almacenada en el Datawarehouse [3].

### **2.4. Alertas**

Ante un cambio o situación de negocios inesperada o que se salga de los parámetros establecidos, una alerta indica dicha anomalía y permite tomar decisiones rápidas garantizando el normal funcionamiento del negocio y evitando pérdidas de tiempo y dinero [3].

### **2.5. ETL**

Es un componente que permite *Extraer, Transformar y Cargar* (ETL, por sus siglas en inglés) mediante la integración de datos cuando se tienen distintas fuentes (es decir, diferentes manejadores de bases de datos) y llevar información de las bases de datos operacionales a las bases de datos multidimensionales [3].

### **2.6. Herramientas de consultas**

Las soluciones de inteligencia de negocios proporcionan varias clases de herramientas, que permiten aprovechar al máximo los datos. Estas

herramientas permiten a los usuarios *visualizar reportes previamente elaborados o definir sus propias consultas a la base de datos*, es decir tener la posibilidad de elegir los campos, filtros, agrupaciones deseadas para la obtención de un reporte perfectamente adecuado a la necesidad de información. Es decir, los usuarios no necesitan conocer los aspectos técnicos de la base de datos [3].

### **3. CASO DE ESTUDIO**

En la Región Huetar Norte de Costa Rica se ha desarrollado la estrategia denominada Zona Económica Especial (ZEE), la que representa el esfuerzo por construir una visión de desarrollo orientada a estimular la inversión y generar empleo de calidad. Esa estrategia engloba al Sistema de Información Regional (SIR-ZEE, [www.sirzee.itcr.ac.cr](http://www.sirzee.itcr.ac.cr)), que proporciona información oportuna y relevante para impulsar las capacidades productivas y fortalecer las capacidades de gobernabilidad.

Debido al éxito de la experiencia SIR-ZEE, se ha planteado transferir dicha experiencia a otras zonas del país dentro del Programa de Regionalización Interuniversitaria del Consejo Nacional de Rectores de Costa Rica (CONARE). Específicamente, se está desarrollando el Sistema de Información Regional del Pacífico Central (SIR-PC) y el Sistema de Información Regional del Sur (SIR-SUR), mediante la vinculación interuniversitaria del Instituto Tecnológico de Costa Rica (ITCR), la Universidad Nacional (UNA), la Universidad Estatal a Distancia (UNED) y la Universidad de Costa Rica (UCR).

En el panorama actual de la Región, la toma de decisiones no se sustenta en datos históricos. Fundamentalmente se debe a que hay pocas herramientas de BI para pymes, las empresas no tienen recursos financieros, ni las tecnologías adecuadas; y carecen de información interna útil, confiable y oportuna. Los aspectos mencionados anteriormente provocan una baja fidelización de clientes, mala relación con proveedores, un bajo nivel de operación, pérdida de oportunidades de negocios y un bajo nivel de competitividad (ver Figura 1).

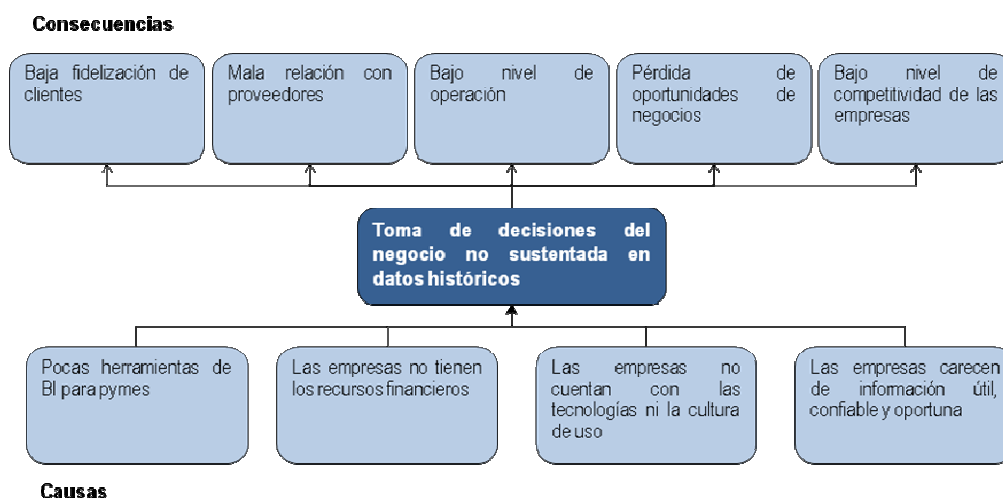
Por los motivos anteriormente expuestos, en el artículo “La inteligencia de negocios en el desarrollo económico local” [9], se efectuó un análisis para dar solución a dicho problema y se obtuvo como resultado que las no pymes cuenten con herramientas especializadas de inteligencia de negocios. Por lo que, se exploró la funcionalidad de la herramienta PENTAHO, trabajando específicamente con las estructuras multidimensionales de proceso analítico en línea (OLAP, por sus siglas en inglés).

Al trabajar con dicha plataforma se concluyó que la misma proporciona las características necesarias para facilitar la toma de decisiones a la pymes de la Región, por lo que el SIR-ZEE próximamente pondrá a disposición de sus usuarios la plataforma PENTAHO. Ésta consiste en una Suite Completa de BI que incluye todos los principales componentes requeridos para implementar soluciones de este tipo [7], tales como: (a) Informes, (b) Tableros de control, (c) Procesamiento analítico, (d) Procesos de extracción, transformación y carga, (e) Integración de datos, (f) Suscripciones, (g) Minería de datos, y (h) Alertas.

Tomando como punto de partida los aspectos anteriores, en este artículo se pretende aplicar los principales componentes de BI a la información almacenada en la base de datos SIR-ZEE.

### 3.1. Objetivo del estudio

El objetivo de este trabajo es aplicar los componentes de BI de la plataforma PENTAHO en la base de datos del SIR-ZEE, con el fin de explotar el funcionamiento de dicha plataforma en el caso específico de la Región Huetar Norte y de este modo proporcionar a los responsables de tomar decisiones un medio en el que puedan apoyarse.



*Figura 1. Análisis del problema de la toma de decisiones del negocio en la Región Huetar Norte.*

### 3.2. Diseño metodológico

El desarrollo de un proyecto de BI es un proceso iterativo, ya que una vez que está implementado se podrá mejorar y ampliar continuamente basado en la retroalimentación de las personas del negocio. El producto evoluciona y madura. Las etapas que conforman este proceso son:

- Etapa 1: Justificación - Evaluar la necesidad (oportunidad o problema) del negocio que dará origen al nuevo proyecto de BI.
- Etapa 2: Planificación - Desarrollar los planes del proyecto de BI.
- Etapa 3: Análisis del negocio - Analizar la necesidad del negocio para identificar los requerimientos de la solución (producto).
- Etapa 4: Diseño - Conceptualizar el producto que resolverá el problema del negocio o habilitará la oportunidad.
- Etapa 5: Construcción - Desarrollar el producto que proporcionará un retorno de la inversión en un tiempo preestablecido.
- Etapa 6: Implantación - Implantar el producto finalizado y evaluar si cumple con el retorno esperado.

Cada etapa requiere la ejecución de una serie de pasos, en la etapa de justificación se requiere evaluar el problema principal que se está abordando (ver la sección Caso de estudio). En la etapa de planificación se estudia la arquitectura de la plataforma (en este caso PENTAHO) y se efectúa un análisis de las fuentes

de datos (es decir, las bases de datos del SIR-ZEE). Seguidamente, en la etapa de análisis del negocio, se definen los principales requerimientos de las fuentes de datos, se selecciona la herramienta que se utilizará para desarrollar el prototipo, se diseñan los reportes, consultas y el prototipo. Además, se realiza una demostración. Para la etapa de diseño se desarrollan todas las consultas y vistas, se elaboran y prueban todas las funciones de ETL, así como las estructuras multidimensionales. En la etapa de construcción se desarrolla la aplicación e interpretan los resultados. Por último, en la etapa de implantación se deben instalar todos los componentes de la aplicación, cargar la base de datos de producción y efectuar una evaluación del producto.

#### **4. RESULTADOS Y DISCUSIÓN**

Se trabajó con tres de los componentes que proporciona PENTAHO, y se diseñaron una serie de ejemplos para poder aplicar los conocimientos sobre BI. El primer grupo de ejemplos consistió en desarrollar estructuras OLAP. El segundo grupo fue para la generación de reportes. Y, el tercer y último ejemplo fue para aplicar técnicas de minería de datos. En las secciones siguientes se analizarán los principales resultados que se obtuvieron con estos ejemplos, en los que se utilizó la base de datos del SIR-ZEE, la que está conformada actualmente por aproximadamente 210 tablas.

##### **4.1. Datawarehouse: Pentaho Cube Designer y Mondrian**

En la plataforma PENTAHO se incluyen las herramientas PENTAHO Cube Designer and OLAP Mondrian, que en conjunto permiten crear y publicar cubos de información para el análisis multidimensional.

En primera instancia, se elaboró un cubo OLAP para interactuar con la información correspondiente al Censo Agropecuario [8], que fue elaborado por el Ministerio de Agricultura y Ganadería (MAG) de Costa Rica y tiene como propósito fundamental obtener información actualizada, confiable y oportuna, sobre las estimaciones de producción de las principales actividades productivas de interés económico de la Región, de manera que esté a la disposición de los técnicos, productores, procesadores, comercializadores, exportadores y público en general [8].

Mediante dicha estructura, en la figura 2, se visualiza la información correspondiente al cultivo de piña en el distrito de Aguas Zarcas. Además, a nivel macro se aprecia la oferta regional de productos agrícolas por cantón, distrito y tipo de cultivo.

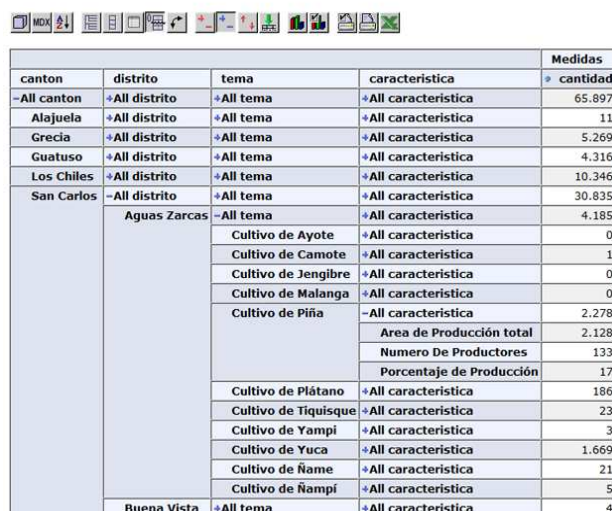
Para complementar la información que proporciona un cubo OLAP, se observa en la figura 3, un mapa digital de información geográfica (tomado del sitio web del SIR-ZEE) en el que se visualiza la distribución del cultivo de piña en la Región. Según la intensidad de los tonos cafés hay una mayor proporción de siembra de piña en el distrito, las zonas en donde se cultiva con mayor proporción la piña son Pital y La Virgen.

El segundo ejemplo desarrollado, muestra la pirámide poblacional de la región por provincia, cantón, distrito y rangos de edad. La figura 4, proporciona los

datos del distrito de Aguas Zarcas, en donde la mayor cantidad de personas está concentrada entre los rangos de 10 a 39 años.

Las herramientas Pentaho Cube Designer y Mondrian, cuentan con una gran variedad de opciones para modificar la consulta. Por ejemplo, se pueden agregar o quitar campos, generar gráficos y exportar datos a hojas electrónicas. Además, ofrecen flexibilidad en el análisis de la información, pues permiten efectuar diferentes cruces de variables. Y, combinando los resultados con sistemas de información geográficos (SIG), brindan al usuario una mejor forma de visualizar los datos.

Censo agropecuario



| canton      | distrito      | tema                 | caracteristica           | Medidas<br>+ cantidad |
|-------------|---------------|----------------------|--------------------------|-----------------------|
| -All canton | +All distrito | +All tema            | +All caracteristica      | 65.897                |
| Alajuela    | +All distrito | +All tema            | +All caracteristica      | 11                    |
| Grecia      | +All distrito | +All tema            | +All caracteristica      | 5.269                 |
| Guatuso     | +All distrito | +All tema            | +All caracteristica      | 4.316                 |
| Los Chiles  | +All distrito | +All tema            | +All caracteristica      | 10.346                |
| San Carlos  | Aguas Zarcas  | -All tema            | +All caracteristica      | 4.185                 |
|             |               | Cultivo de Ayote     | +All caracteristica      | 0                     |
|             |               | Cultivo de Camote    | +All caracteristica      | 1                     |
|             |               | Cultivo de Jengibre  | +All caracteristica      | 0                     |
|             |               | Cultivo de Malanga   | +All caracteristica      | 0                     |
|             |               | Cultivo de Piña      | -All caracteristica      | 2.278                 |
|             |               |                      | Area de Producción total | 2.128                 |
|             |               |                      | Numero De Productores    | 133                   |
|             |               |                      | Porcentaje de Producción | 17                    |
|             |               | Cultivo de Plátano   | +All caracteristica      | 186                   |
|             |               | Cultivo de Tiquisque | +All caracteristica      | 23                    |
|             |               | Cultivo de Yampi     | +All caracteristica      | 3                     |
|             |               | Cultivo de Yuca      | +All caracteristica      | 1.669                 |
|             |               | Cultivo de Ñame      | +All caracteristica      | 21                    |
|             |               | Cultivo de Ñampi     | +All caracteristica      | 5                     |
|             |               | Buena Vista          | +All tema                | +All caracteristica   |

Figura 2. Cubo OLAP del censo agropecuario en la Región Huetar Norte

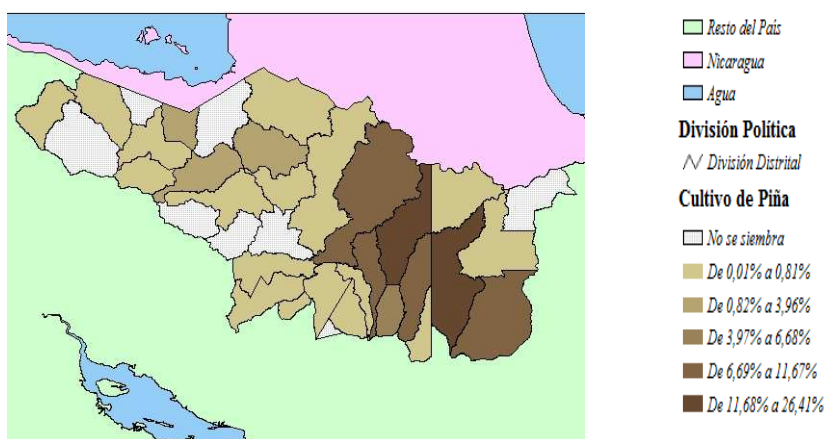


Figura 3. Distribución del cultivo de piña en la Región Huetar Norte

### Pirámide poblacional

| provincia               | canton               | distrito      | característica       | Medidas |
|-------------------------|----------------------|---------------|----------------------|---------|
| -All provincia          | +All canton          | +All distrito | +All característica  | 259.900 |
| Alajuela                | -All canton          | +All distrito | +All característica  | 214.465 |
|                         | Alajuela             | +All distrito | +All característica  | 2.493   |
|                         | Grecia               | +All distrito | +All característica  | 7.744   |
|                         | Guatuso              | +All distrito | +All característica  | 13.045  |
|                         | Los Chiles           | +All distrito | +All característica  | 19.732  |
|                         | San Carlos           | -All distrito | +All característica  | 127.140 |
|                         |                      | Aguas Zarcas  | -All característica  | 13.651  |
|                         |                      |               | Personas 1-4 Años    | 1.269   |
|                         |                      |               | Personas 10- 19 Años | 3.157   |
|                         |                      |               | Personas 20- 29 Años | 2.225   |
|                         |                      |               | Personas 30- 39 Años | 2.010   |
|                         |                      |               | Personas 40- 49 Años | 1.242   |
|                         |                      |               | Personas 5-9 Años    | 1.814   |
|                         |                      |               | Personas 50- 59 Años | 760     |
|                         |                      |               | Personas 60- 64 Años | 226     |
|                         | Personas 65- 69 Años |               | 222                  |         |
|                         | Personas 70- 74 Años | 135           |                      |         |
| Personas 75 Años Ó Más  | 234                  |               |                      |         |
| Personas Menos de 1 Año | 357                  |               |                      |         |
| Buena Vista             | +All característica  | 352           |                      |         |

**Figura 4.** Cubo OLAP de la pirámide poblacional de la Región Huetar Norte

#### 4.2. Herramienta de consultas: Pentaho Report Designer

La mayoría de las organizaciones utilizan reportes para registrar y visualizar análisis y resultados. Como consecuencia, los reportes son considerados una necesidad principal en BI. La unidad de reportes de Pentaho permite acceder, dar formato y distribuir fácilmente la información. De igual forma, provee acceso a fuentes de datos relacionales, OLAP o basadas en XML, así como ofrece varios formatos de salida como PDF, HTML, Excel o hasta texto plano. También permite llevar esta información a los usuarios finales vía web, e-mail, portales corporativos o aplicaciones propias.

Para ser congruentes con los dos ejemplos desarrollados en la sección anterior (el primero sobre censo agropecuario y el segundo sobre la pirámide poblacional), en este apartado se efectuaron reportes que complementen los cubos OLAP. En la figura 5, se presenta el área cultivada por distrito y tipo de cultivo en la Región, específicamente para Pital y Pocosol pertenecientes al cantón de San Carlos. En el caso de Pital, la mayor proporción cultivada es de yuca y en Pocosol es de Tiquisque. Otro ejemplo se presenta con los datos de la pirámide poblacional por cantón y distrito, ver Figuras 6 y 7. En la figura 6, se muestra un cuadro resumen con los diferentes rangos de edad para los distritos de Ciudad Quesada y La Tigra. En ambos casos, la mayor proporción de personas se ubica en el rango de 10 a 19 años. Por otro lado, en la figura 7, se observa un gráfico comparando el cantón de Upala y Los Chiles.

### Área de cultivo por distrito

| Distrito: Pital      |          |
|----------------------|----------|
| Tipo de cultivo      | Cantidad |
| Cultivo de Ayoite    | 0        |
| Cultivo de Camote    | 0        |
| Cultivo de Jengibre  | 0.3      |
| Cultivo de Malanga   | 0.4      |
| Cultivo de Ñame      | 39       |
| Cultivo de Ñameí     | 12       |
| Cultivo de Piña      | 399      |
| Cultivo de Plátano   | 6        |
| Cultivo de Tiquisque | 27.3     |
| Cultivo de Yamsí     | 1        |
| Cultivo de Yuca      | 1371     |

| Distrito: Pocosol    |          |
|----------------------|----------|
| Tipo de cultivo      | Cantidad |
| Cultivo de Ayoite    | 0        |
| Cultivo de Camote    | 0        |
| Cultivo de Jengibre  | 14.2     |
| Cultivo de Malanga   | 0        |
| Cultivo de Ñame      | 31       |
| Cultivo de Ñameí     | 4        |
| Cultivo de Piña      | 6        |
| Cultivo de Plátano   | 3        |
| Cultivo de Tiquisque | 81.3     |
| Cultivo de Yamsí     | 4.5      |

Figura 5. Reporte sobre Censo agropecuario Región Huetar Norte

### Pirámide poblacional por distrito

| Distrito: Quesada       |          |
|-------------------------|----------|
| Rango edad              | Cantidad |
| Personas Menos de 1 Año | 740      |
| Personas 1-4 Años       | 2863     |
| Personas 5-9 Años       | 4090     |
| Personas 10-19 Años     | 8244     |
| Personas 20-29 Años     | 5851     |
| Personas 30-39 Años     | 5721     |
| Personas 40-49 Años     | 4009     |
| Personas 50-59 Años     | 2237     |
| Personas 60-64 Años     | 776      |
| Personas 65-69 Años     | 594      |
| Personas 70-74 Años     | 476      |
| Personas 75 Años Ó Más  | 742      |

| Distrito: Tigra         |          |
|-------------------------|----------|
| Rango edad              | Cantidad |
| Personas Menos de 1 Año | 145      |
| Personas 1-4 Años       | 559      |
| Personas 5-9 Años       | 683      |
| Personas 10-19 Años     | 1226     |
| Personas 20-29 Años     | 880      |
| Personas 30-39 Años     | 792      |
| Personas 40-49 Años     | 488      |
| Personas 50-59 Años     | 301      |
| Personas 60-64 Años     | 156      |

Figura 6. Reporte de la pirámide poblacional de la Región Huetar Norte

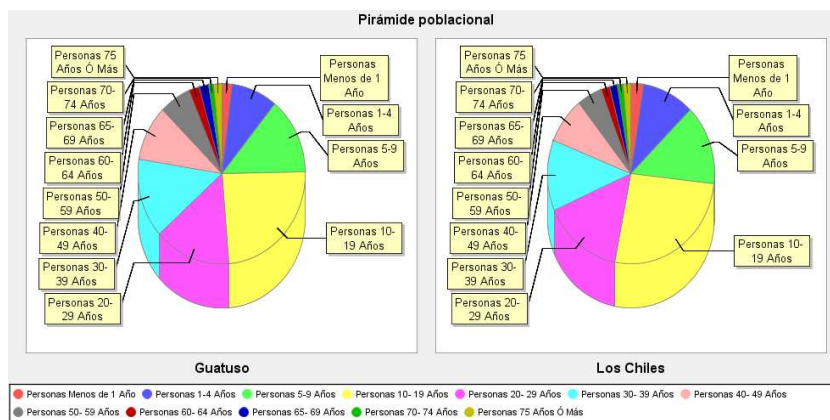


Figura 7. Gráfico de la pirámide poblacional de la Región Huetar Norte

### 4.3. Minería de datos: Weka

El último componente estudiado se denomina Weka, es un software para aprendizaje automático y minería de datos. Contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades.

Para explorar las funcionalidades de esta herramienta, se utilizará la técnica de clusters (con los algoritmos SimpleKmeans y FarthestFirst) para clasificar los 35 distritos de la Región. Para ello, se tomó como base las siguientes características sociodemográficas: (1) Índice de desarrollo social, (2) Tasa de desempleo abierta, (3) Tasa de defunción, (4) Tasa de fecundidad y (5) Densidad de población. Esta técnica consiste en crear un conjunto de elementos, los cuales tengan alguna característica común. Es importante que para aplicar este tipo de algoritmo, se asumió que los datos tienen un comportamiento normal.

En la Figura 8, se puede visualizar el resultado de aplicar el algoritmo SimpleKmeans al conjunto de datos. Este algoritmo busca formar clusters (grupos), los que son representados por K objetos y cada uno de estos K objetos es el valor medio de los objetos que pertenecen a dicho grupo. Como se puede apreciar se formaron tres grupos, el primer grupo contiene catorce elementos, el segundo grupo doce elementos y el tercer grupo nueve elementos.

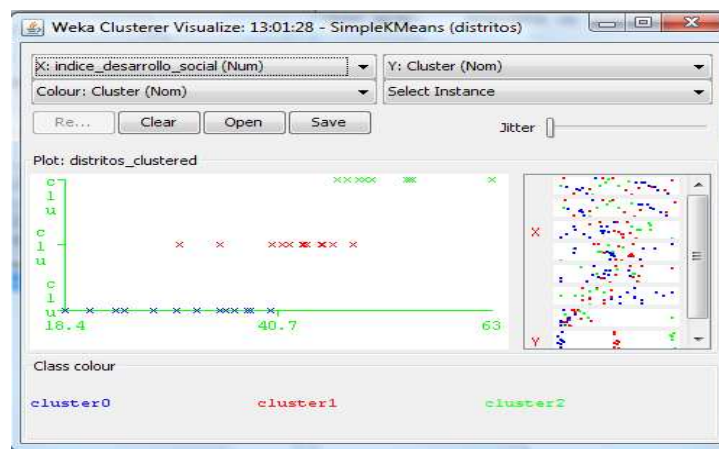
Cluster centroids:

| Attribute                | Full Data<br>(35) | Cluster#  |           |          |
|--------------------------|-------------------|-----------|-----------|----------|
|                          |                   | 0<br>(14) | 1<br>(12) | 2<br>(9) |
| indice_desarrollo_social | 40.2486           | 30.9857   | 42.0167   | 52.3     |
| tasa_deseempleo_abierta  | 5.92              | 6.65      | 6.3       | 4.2778   |
| Tasa_Defuncion           | 2.6086            | 1.7286    | 3.625     | 2.6222   |
| Tasa_Fecundidad          | 2.5771            | 2.9357    | 2.4167    | 2.2333   |
| Densidad_poblacion       | 33.9177           | 15.3927   | 32.9962   | 63.9631  |

Clustered Instances

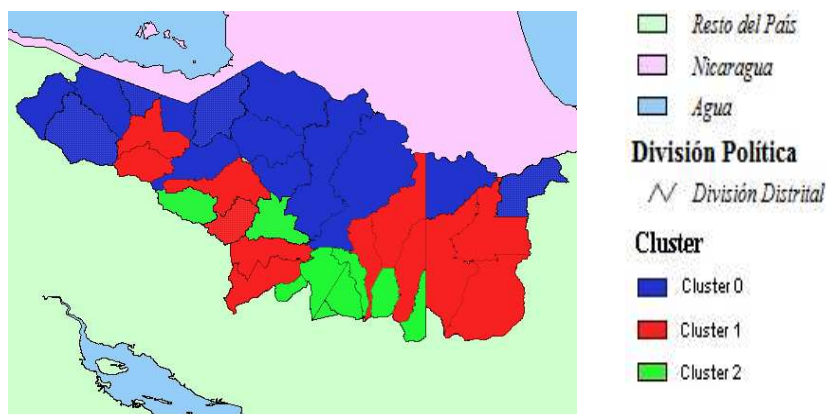
|   |           |
|---|-----------|
| 0 | 14 ( 40%) |
| 1 | 12 ( 34%) |
| 2 | 9 ( 26%)  |

**Figura 8.** Clusters de distritos con el algoritmo SimpleKmeans



**Figura 9.** Visualización del clusters de distritos con el algoritmo SimpleKmeans

Igualmente, en la Figura 9, se visualiza gráficamente como se encuentran agrupados los distritos. Al mismo tiempo, en la Figura 10, se presenta un mapa de la Región con la distribución de los clusters, evaluando la variable índice de desarrollo social. Con estos datos, se obtiene que los distritos con más bajo índice de desarrollo social se ubican en la zona fronteriza con Nicaragua.



**Figura 10.** Mapa del clusters de distritos con el algoritmo SimpleKmeans

```

=== Model and evaluation on training set ===

FarthestFirst
=====

Cluster centroids:

Cluster 0
 34.7 10.0 2.5 2.8 12.242
Cluster 1
 54.8 4.0 4.2 1.9 255.168
Cluster 2
 63.0 0.0 4.5 2.3 13.5

Clustered Instances

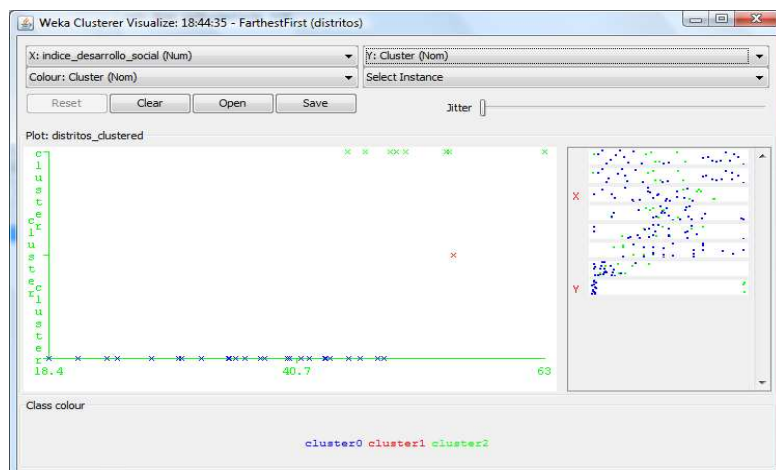
0      26 ( 74%)
1       1 (  3%)
2       8 ( 23%)

```

**Figura 11.** Clusters de distritos con el algoritmo FarthestFirst

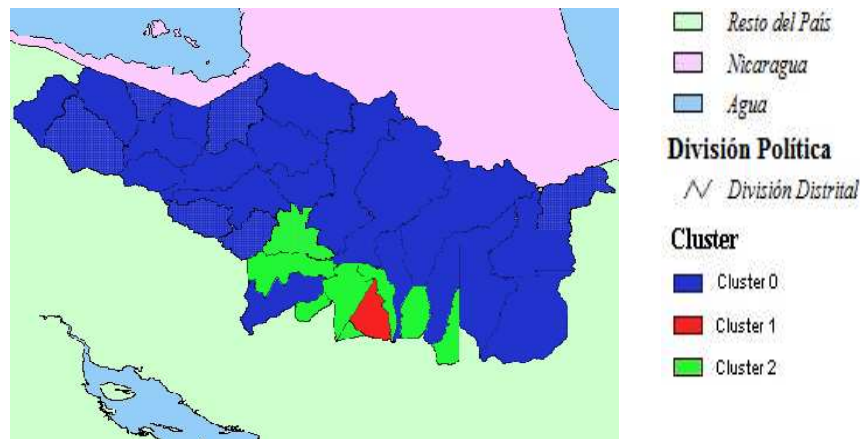
Por otra parte, al aplicar el algoritmo FarthestFirst, que es usado como técnica de agrupamiento para encontrar grupos homogéneos y de esta forma, detectar los estilos de aprendizaje predominantes buscando el vecino más cercano, se obtuvo el resultado que se muestra en la Figura 11. Igual al caso anterior se crearon tres clusters, el primer grupo con veintiséis elementos, el segundo grupo con un elemento y el tercer grupo con ocho elementos. Para cada clusters se muestra la media de los datos que forman ese conjunto.

Asimismo, en la Figura 12, se visualiza gráficamente como se encuentran agrupados los distritos. Y, en la Figura 13, se presenta un mapa de la Región con la distribución de los clusters, evaluando la variable índice de desarrollo social. Con este algoritmo, los resultados obtenidos están más dispersos, por ejemplo el caso del distrito de Ciudad Quesada que se encuentra en un solo cluster.



**Figura 12.** Visualización del clusters de distritos con el algoritmo FarthestFirst

Aunque, los resultados obtenidos en ambos casos son totalmente diferentes, ello no inválida la utilidad de los algoritmos. Pues, se debe tener claro que su aplicación está asociada a la respuesta que se quiere obtener. Sin embargo, en el caso del algoritmo FarthestFirst se obtuvieron resultados más cercanos a la realidad de la región, en donde el cantón de Ciudad Quesada sobresale del resto de los cantones.



**Figura 13.** Mapa del clusters de distritos con el algoritmo FarthestFirst

## 5. CONCLUSIONES Y TRABAJOS FUTUROS

En este artículo se han aplicado tres componentes de BI en el contexto de la Región Huetar Norte de Costa Rica, con apoyo de la herramienta denominada PENTAHO. Además, a pequeña escala se incursionó en trabajar algunos resultados de BI con el sistema de información geográfico que proporcionar el SIR-ZEE.

Uno de los problemas que se presentaron en la etapa de planificación, fue la forma en que cierta información estaba almacena en la base de datos. Pues, en algunos casos los datos se encontraban para toda la Región y esto imposibilitó el poder efectuar un análisis más detallado de variables importantes. Por ello, se debe tener en consideración que para aprovechar la información de una mejor manera es necesario que las bases de datos tengan una buena estructura.

Es importante destacar la factibilidad de aplicar los diferentes modelos de minería de datos al tratamiento de información relacionada con poblaciones. Además, el hecho de aplicar la técnica de minería de datos denominada cluster para agrupar los distritos de la Región utilizando características sociodemográficas, permite identificar los patrones de comportamiento. De este modo, este tipo de técnicas proporciona información sumamente valiosa para que las organizaciones puedan apoyarse en el proceso de toma de decisiones. Sin embargo, se debe tener en cuenta el algoritmo que se seleccione, ya que como se observó en la sección Resultados y Discusión los resultados varían dependiendo del algoritmo.

La idea fundamental de incorporar las técnicas de BI al SIR-ZEE, está enfocada en monitorear las actividades de la Región. Con el fin de tener información en tiempo real de lo que está ocurriendo con los diferentes tipos de datos y la información que está pasando por las diferentes instituciones se convierta en información accionable. Es decir, que pueda verse en línea lo que está ocurriendo en la Región y que a partir de indicadores sea posible analizar la información para tomar decisiones directamente desde las herramientas de análisis. En otras palabras cerrar el ciclo de la información llevándola a acciones específicas que generen una toma de decisiones adecuada y certera. Esto puede realizarse debido a que en la base de datos del SIR-ZEE se tiene una gran

variedad de información y por ello el área de aplicación de BI se torna muy extensa.

Entre las tareas pendientes que nos arroja este artículo se pueden enumerar (a) Investigar sobre nuevas técnicas de minería de datos, (b) Mejorar los datos almacenados en la base de datos del SIR-ZEE, (c) Trasladar el conocimiento adquirido a otras instituciones (por ejemplo a municipalidades mediante cursos y asesorías), (d) Combinar las técnicas de BI con los Sistemas de Información Geográfica (SIG) para aprovechar la gama de servicios del SIR-ZEE y ayudar al usuario final a comprender mejor los resultados. Por último, (e) Desarrollar un portal de información que integre los sistemas SIR-ZEE y en segunda instancia cuando estén operando, también podrá contar con la información del SIR-SUR y el SIR-PC. Además, que tenga conexión con otras bases de datos como del Estado de la Nación, Registro Nacional de Propiedad, entre otros.

## 6. REFERENCIAS

- [1] C. García. Gestión del desempeño corporativo. La evolución del Business Intelligence. Consultado el 3 de marzo del 2008. Disponible en: <http://www.sg.com.mx/content/view/499/>
- [2] S.n. Boletín de Asesoría Gerencial. La Inteligencia de Negocios. Consultado el 6 de febrero del 2009. Disponible en: [http://www.pwc.com/ve/spa/pdf/aseger\\_200810.pdf](http://www.pwc.com/ve/spa/pdf/aseger_200810.pdf)
- [3] S.n. El manejo inteligente de la información. Consultado el 6 de febrero del 2009. Disponible en: [http://www.igerencia.com/inteligencia\\_de\\_negocios/](http://www.igerencia.com/inteligencia_de_negocios/)
- [4] S.n. Business Intelligence y la Cadena de Value. Consultado el 6 de febrero del 2009. Disponible en: <http://www.agensystech.com/articlesitems/bivc>
- [5] LLombart, O. BI: Inteligencia aplicada al negocio. Consultado el 8 de febrero del 2009. Disponible en: [http://www.eldiarioexterior.com/conocimiento/docs/BI\\_Inteligencia\\_aplicada\\_al\\_negocio.pdf](http://www.eldiarioexterior.com/conocimiento/docs/BI_Inteligencia_aplicada_al_negocio.pdf)
- [6] S.n. Tutorial y ejemplos de Pentaho Business Intelligence. Consultado el 17 de febrero del 2009. Disponible en: <http://pentaho.almacen-datos.com/tutorial.html>
- [7] S.n. Pentaho Commercial Open Source Business Intelligence. Consultado el 17 de febrero del 2009. Disponible en: <http://www.pentaho.com/>
- [8] MAG. Informe Censo Agrícola 2005. Consultado el 22 de febrero del 2009. Disponible en: <http://www.sirzee.itcr.ac.cr/modules/Biblioteca/Usuario/archivos/Informe%20Final%20Censo%202005.doc>
- [9] Treviño, M. Valerio, A. López, O. La Inteligencia de Negocios en el Desarrollo Económico Local. COMPDES. León, Nicaragua. 2008.