



Universidad CENFOTEC

Maestría en Tecnología de Base de Datos
Proyecto de Investigación Aplicada 2

Tema:

Creación de un modelo de inteligencia de negocios en la atención de emergencias
del Benemérito Cuerpo de Bomberos de Costa Rica

Estudiante
Ceciliano Villalobos, Cristian

Enero 2017

Tabla de contenido

Resumen Ejecutivo.....	iv
Capítulo 1. Introducción.....	1
1.1. Generalidades.....	1
1.2. Antecedentes del Problema.....	1
1.3. Definición y Descripción del Problema.....	1
1.4. Justificación.....	2
1.5. Viabilidad.....	3
1.5.1. Punto de Vista Técnico.....	3
1.5.2. Punto de Vista Operativo.....	3
1.5.3. Punto de Vista Económico.....	3
1.6. Objetivos.....	4
1.6.1. Objetivo General.....	4
1.6.2. Objetivos Específicos.....	4
1.7. Alcances y limitaciones.....	4
1.7.1. Alcances.....	4
1.7.2. Limitaciones.....	5
1.8. Marco de Referencia Organizacional y Socioeconómico.....	5
1.8.1. Historia.....	5
1.8.2. Tipo de Negocio y Mercado Meta.....	10
1.8.3. Misión, Visión y Valores.....	10
1.9. Estado de la Cuestión.....	11
Capítulo 2. Marco Teórico.....	20
¿Qué es inteligencia de Negocios?.....	20
Modelo de Datos.....	21
Calidad de datos.....	22
Extracción de datos.....	23
Data Warehouse.....	24
Data Mart.....	24
Estructuras de Inteligencia de Negocios.....	25
Arquitectura de diseño para Data Warehouse.....	25
Bill Inmon.....	25
Ralph Kimball.....	26
Minería de Datos.....	27
OLAP.....	41
En Resumen.....	42
Capítulo 3. Marco Metodológico.....	44
3.1. Tipo de Investigación.....	44
3.2. Alcance de la Investigación.....	44
3.3. Enfoque.....	44
3.4. Diseño.....	45
3.5. Instrumento de recolección de datos.....	46
3.6. Técnicas de Análisis de la Información.....	46
Capítulo 4. Análisis de Diagnostico.....	47
Capítulo 5. Propuesta de Solución.....	48
Exploración y calidad de los datos.....	48
Creación de las bases de datos de BI.....	54
Construcción del stage área.....	55

Construcción del Data Warehouse	55
Base de datos garbage (basurero)	56
Extracción de datos	56
Creación del cubo	58
Implementación de Minería de datos	59
Visualizador de BI	62
OLAP	62
Minería de Datos	65
Capítulo 6. Conclusiones y Recomendaciones	68
Conclusiones	68
Recomendaciones	72
Capítulo 7 Reflexiones Finales	73
Capítulo 8 Trabajos a Futuro	74
Glosario	75
Referencias	79

Tabla de Figuras

<i>Figura 1: Paradigma: Bill Inmon</i>	25
<i>Figura 2: Paradigma: Ralph Kimball</i>	26
<i>Figura 3: Diagrama Inteligencia de Negocios</i>	43
<i>Figura 4: Diagrama Inteligencia de Negocios</i>	43
<i>Figura 5: Diagrama de Ishikawa: Análisis causa-efecto de los problemas</i>	46
<i>Figura 6: Total de Emergencias sin posición geográfica</i>	49
<i>Figura 7: Diagrama de Flujo Primera Parte</i>	50
<i>Figura 8: Diagrama de Flujo Segunda Parte</i>	51
<i>Figura 9: Diagrama de Flujo Tercera Parte</i>	51
<i>Figura 10: Diagrama de Flujo Cuarta Parte</i>	53
<i>Figura : Total de Emergencias sin posición geográfica</i>	54
<i>Figura : Modelo de la arquitectura de BI</i>	57
<i>Figura Modelo del Cubo de BI</i>	59
<i>Figura Grafico de comparación de algoritmos de minería de datos</i>	60
<i>Figura : Modelo de Visualización de Datos</i>	63
<i>Figura : Herramienta para desarrollar informes POWER BI Microsoft</i>	63
<i>Figura : Ejemplo de informe con POWER BI de Microsoft</i>	64
<i>Figura : Ejemplo de informe geográfico con POWER BI de Microsoft</i>	65
<i>Figura : Ejemplo de uso del algoritmo de minería de datos con Excel</i>	66
<i>Figura : Ejemplo de uso del algoritmo de minería de datos con Excel</i>	67

Resumen Ejecutivo

El BCBCR (Benemérito Cuerpo de Bomberos de Costa Rica) es una institución autónoma sujeta a la ley en materia de gobierno, la cual brinda servicio a la población costarricense en emergencias que requieran la asistencia de dicha organización para salvaguardar la vida de los ciudadanos.

BCBCR existe desde hace más de 150 años, iniciando operaciones en el año 1865 llevando protección a la ciudadanía, se encuentra distribuida en estaciones ubicadas alrededor de todo el país contando actualmente con 74 de estas estaciones.

BCBCR brinda servicio y protección en cualquier evento categorizado como emergencia en el que se requiera la intervención de los bomberos desde un incendio en estructuras hasta un recate de persona. El proceso de estos eventos queda registrado en sistemas institucionales los cuales dan la facilidad de dar un seguimiento a la emergencia y recopilar información para eventuales análisis a futuro.

El proyecto en sí, implementa un modelo de inteligencia de negocios el cual permite centralizar la información de bomberos en un único repositorio permitiendo aplicar métodos de análisis que ayudan a transformar el dato en información y la información en conocimiento para la toma de decisiones, el modelo aparte de análisis de información, también incorpora algoritmos de minería de datos que ayudan a montar escenarios de predicción para proyectar temas de uso de recursos

en el BCBCR o crear estrategias que ayuden ir un paso adelante para atender una eventual emergencia.

Capítulo 1. Introducción

1.1. Generalidades

Una de las funciones principales del BCBCR (Benemérito Cuerpo de Bomberos de Costa Rica), es ir de la mano con la calidad de su servicio, esto permite siempre estar alertas y atender lo más pronto posible cualquier emergencia que les sea asignada. Para estar en este constante proceso de mejora la institución registra cada evento donde participa por medio de un sistema, en el cual se le da seguimiento a la emergencia desde que inicia hasta que finaliza. Toda esta información es registrada en una base de datos transaccional.

1.2. Antecedentes del Problema

A raíz del estudio del estado de la cuestión no se encontraron investigaciones relacionadas a inteligencia de negocios aplicado a una institución como lo es el departamento de Bomberos a nivel de Latino América, existen herramientas de visualización de datos como lo es el departamento de bomberos de Quito en Ecuador pero sin una adecuada estructura de análisis de información. En nuestro BCBCR no se ha explorado proyectos de Inteligencia de Negocios, siempre ha sido un proceso manual y descentralizado.

1.3. Definición y Descripción del Problema

En el BCBCR actualmente el realizar un análisis de información cuya finalidad sea estudios estadísticos como, por ejemplo: comparaciones en periodos de tiempo o realizar estudios de probabilidades para conocer eventos a futuro de las emergencias atendidas en Costa Rica, se ha hecho una tarea manual sumamente

compleja, imprecisa y costosa en recurso y procesamiento. Ya que se debe ingresar a un sistema transaccional para sacar listados de reportes estáticos (estos listados se pueden obtener de varias pantallas, pero sus resultados no suelen ser los mismos). Además, son ejecutados por diferentes departamentos lo que conlleva a una consolidación entre los mismos para llegar a un único resultado.

1.4. Justificación

Para facilitar la toma de decisiones en el BCBCR es necesario analizar los datos, conocer y predecir el comportamiento de la institución. Las estrategias de inteligencia de negocios buscan apoyar a los jefes de batallón, director de operaciones o cualquier otro departamento con interés de analizar información, para identificar, medir, predecir y seguir estrategias que los lleven al cumplimiento de la visión de la organización.

Para lograr los objetivos y tener altos estándares de calidad en el servicio que se le brinda al país reaccionando ágilmente ante la atención de emergencia e información de interés nacional, es de gran ayuda contar con herramientas que permitan realizar estrategias basadas en la experiencia acumulada durante los años de servicio para ello se puede utilizar inteligencia de negocios. El cómo hacerlo es el objetivo de este proyecto, al crear un modelo de inteligencia de negocios que permita el análisis de información en la atención de emergencias.

1.5. Viabilidad

1.5.1. Punto de Vista Técnico

El BCBCR cuenta con licencias actuales de herramientas robustas para extracción de información, implementación de diseños de inteligencia de negocios y visualización de datos, además dispone de recurso suficiente para cubrir las necesidades a nivel de hardware y almacenamiento físico y lógico para la implementación de un proyecto de inteligencia de negocios.

1.5.2. Punto de Vista Operativo

A nivel operativo, se tendría un único centro de datos el cual vendría a facilitar en tiempos de respuesta realizar análisis de información precisas, que permitan conocer el estado actual de la organización en cuanto a la atención de emergencias relacionadas con el BCBCR.

1.5.3. Punto de Vista Económico

El recurso a nivel humano, utilizado para el análisis, desarrollo e implementación del proyecto no es necesario realizar una inversión monetaria dado que el mismo es un servicio que no se va a cobrar. El BCBCR ya cuenta con licencias y hardware para desarrollar el proyecto no es necesario realizar algún tipo de licitación para adquirir equipo de cómputo o licencias de herramientas de inteligencia de negocios. Aun así se cuenta con presupuesto para ser utilizado en caso de que sea necesario para el proyecto.

1.6. Objetivos

1.6.1. Objetivo General

Crear un modelo de inteligencia de negocios que permita un análisis estratégico de la información para la toma de decisiones en la atención de emergencias del Benemérito Cuerpo de Bomberos de Costa Rica.

1.6.2. Objetivos Específicos

1. Adquirir conocimiento a un nivel general de la organización.
2. Comprender el comportamiento de los datos de la organización.
3. Crear un sistema de almacenamiento de datos (Data Warehouse)
4. Implementar un algoritmo de minería de datos para la predicción de atención de emergencias en las estaciones.
5. Implementar una herramienta de procesamiento analítico de datos (OLAP)
6. Implementar un software de visualización de datos que se conecte al modelo de BI.
7. Elaborar informes con el resultado de los análisis de datos.
8. Clasificar la información de acuerdo a los criterios del negocio.
9. Comparar los resultados de análisis en diferentes periodos del tiempo.

1.7. Alcances y limitaciones

1.7.1. Alcances

- Creación de un Data Warehouse(Almacén de datos)
- Extracción de la Información.
- Proceso de calidad de datos.
- Creación del modelo de inteligencia de negocios.

- Manual de Usuario
 - Manual técnico
 - Capacitación de usuario final
 - Limitaciones
-
- No se dará consultorías sobre licitaciones para adquisición de equipo hardware.
 - La adquisición del equipo hardware para la ejecución del proyecto queda a responsabilidad del negocio.
 - No se dará consultorías sobre características para adquisición de software.
 - La adquisición de las licencias de software para la ejecución del proyecto queda a responsabilidad del negocio.
 - Capacitación técnica de la implementación.

1.8. Marco de Referencia Organizacional y Socioeconómico

1.8.1. Historia

Desde 1543 se tiene conocimiento de incendios en Costa Rica. En esa ocasión los aborígenes quemaron dos poblados como protesta por los vejámenes hechos a sus Caciques. A través de los años los incendios han seguido presentes en el país, muchos de estos provocados intencionalmente como los que se presentaron en Nicoya en 1681 y en Esparza en 1687, causados por los piratas. El estado de la cosas continuó y aún en el siglo XIX, todavía Costa Rica no contaba con un grupo organizado para combatir los incendios.

La historia registra que los sucesos ocurridos en San José, el 26 de enero de 1864, a raíz de un voraz incendio en la casa propiedad de don Francisco María Iglesias, indujeron a los costarricenses para formar un cuerpo de bomberos debidamente organizado.

Con la enorme preocupación que dejó entre los vecinos ese siniestro, el 15 de febrero de ese mismo año la Municipalidad de San José acordó iniciar gestiones, incluso recoger una contribución voluntaria entre los ciudadanos, para traer de los Estados Unidos una “bomba para incendios”, la cual llegó a la capital el 20 de junio de 1865. Simultáneamente, el Ayuntamiento de San José preparó y presentó al Poder Ejecutivo, el primer Reglamento Oficial del Cuerpo de Bomberos, el cual fue aprobado con fecha **27 de julio de 1865**, la cual marca el inicio en Costa Rica de las actividades de una organización de esa índole.

Se constituyó así el primer cuerpo de respuesta, a cargo de la Municipalidad de San José, cuya operación continuó en sus primeros años matizada con bastantes penalidades. Algunos de los primeros miembros de la Directiva de esa naciente organización fueron don Alfredo García, capitán; don Fernando Estreber, secretario y don Guillermo Nanne, primer teniente.

Por serias dificultades económicas de la Corporación Municipal, en 1914 los bomberos dejaron de pertenecer a ella y de funcionar como tal. Entonces pasaron a ser dependencia del gobierno y los miembros de la Policía de Orden y Seguridad asumieron la responsabilidad de operar la bomba para incendios, junto con la ayuda espontánea de ciudadanos que acudían en los momentos que cada siniestro ocurría. Sin embargo, debido a la impericia y falta de conocimientos sobre técnicas de extinción, los improvisados apagafuegos no brindaban un servicio adecuado en los momentos que más se les requería.

Posteriormente, como consecuencia de la proliferación en el país de agencias de seguros de compañías extranjeras, así como la falta de legislación apropiada en materia de seguros, se traspasaba la barrera de la moral y surgen los primeros indicios de lo que algunos llamaron “plaga de incendiarismo” que llegó a su clímax a principios del siglo XX. Con el ánimo de obtener ganancias fáciles, las propiedades se aseguraban contra incendio con doble seguro y luego las quemaban intencionalmente para cobrar la respectiva indemnización.

A pesar de los efectos perniciosos que significaba el incendiarismo, el Cuerpo de Bomberos seguía manejándose por medio del gobierno. En 1912 cuando las preocupaciones de la ciudadanía se acrecentaban, surgió la idea de modernizar el equipo de extinción existente, con dos bombas de vapor y adquirir una bomba “sistema automóvil”. De la misma forma, resurgió la necesidad de que la municipalidad capitalina se preocupara por el restablecimiento de los bomberos, por cuanto la policía no soportaba más el cumplimiento de su doble función de apagar incendios y mantener el orden público.

En junio de 1913 llegó a San José una bomba construida por la fábrica Knox, de Springfield, Massachusetts en Estados Unidos, primera de sistema automotor en el país. Ese mismo año se implantó un sistema de corte militar en la organización formal de los bomberos.

En 1924 se presentó un cambio radical en la concepción del seguro en Costa Rica que culminó con la promulgación de la Ley de Monopolio de Seguros en favor del estado, el 30 de octubre. Entre los factores que mediaron para su aprobación se encontraba el acabar definitivamente con el problema del incendiarismo.

Los gobernantes de la época observaron la necesidad de una mejor prevención, protección y combate de incendios y poco después de la creación del

Banco Nacional de Seguros, el 29 de mayo de 1925, por medio del Decreto Ejecutivo N°4 del entonces presidente de la República el Lic. Ricardo Jiménez Oreamuno, se dispuso que el Cuerpo de Bomberos pasara a ser una dependencia del citado Banco, hoy el Instituto Nacional de Seguros (INS) y que éste fuera el encargado de su administración y dotación.

Dicho decreto indicaba que:

“Artículo 1°.- El Cuerpo de Bomberos estará, desde el primero de junio próximo, bajo la inmediata dependencia del Banco Nacional de Seguros, con las dotaciones que marca el Presupuesto.

Artículo 2.- El Banco proveerá a dicho Cuerpo de los útiles necesarios, para su mayor eficiencia, corriendo de su cuenta los gastos, superiores al Presupuesto, que demande la reorganización y la compra de elementos.

Artículo 3.- Queda facultado el Banco para organizar Cuerpos de Bomberos voluntarios, en la capital y en la demás ciudades que juzgare conveniente.

Artículo 4°.- El Banco presentará a la aprobación de la Secretaría de Seguridad Pública el reglamento de esas organizaciones.

Artículo 5°.- El personal del Cuerpo o Cuerpos de Bomberos tendrá carácter militar y estará bajo la superior vigilancia de la Secretaría de Seguridad Pública.” (Citado por Instituto Nacional de Seguros, 1973, p.269).”

Los dos últimos artículos fueron eliminados mediante el Decreto Ejecutivo N° 6 del primero de setiembre de 1925 con el fin de que el Banco pudiera administrar a los bomberos sin ninguna interferencia.

De ahí en adelante se inició una nueva era para los bomberos cuya organización se transformó totalmente. El Banco Nacional de Seguros se preocupó

por dotarlos de equipos modernos, con un número de miembros acorde a las circunstancias; se comenzaron a construir edificios especialmente para Estaciones de Bomberos y se extendió el servicio por distintas partes del país. El Banco autorizó al Cuerpo de Bomberos para crear un servicio voluntario de bomberos, el primero de los cuales se organizó en 1927, constituyéndose formalmente el 25 de julio.

Al cumplir los cien años de su fundación en 1965, con veinte estaciones de bomberos operando en el país, el Cuerpo de Bomberos de Costa Rica recibió el benemeritazgo.

En agosto del 2010, la organización inicia un nuevo hito en la historia del Benemérito Cuerpo de Bomberos de Costa Rica, con la puesta en marcha para la aprobación de la Ley de Fortalecimiento Económico del Benemérito Cuerpo de Bomberos de Costa Rica, N° 8992, la cual entró en vigencia el 22 setiembre del 2011, mediante la publicación en el diario oficial La Gaceta. Con ello se logra la aprobación de un tributo solidario equivalente al 1.75% de la facturación mensual por consumo de electricidad (entre los 100 kw y los 1750 kw) como fuente complementaria de ingresos para la operación y el crecimiento sostenible del Cuerpo de Bomberos de Costa Rica.

En virtud de lo anterior, el año 2012 se convirtió en el inicio de una nueva etapa de expansión y mejoramiento de los servicios que brinda la organización. En la actualidad, Cuerpo de Bomberos ha logrado tener un total de 73 estaciones de bomberos distribuidas en todo el territorio, que brindan sus servicios de protección a la sociedad costarricense. (Bomberos de Costa Rica, s.f.)

1.8.2. Tipo de Negocio y Mercado Meta

Es un negocio sin fines de lucro y de bien social, que se dedica a salvaguardar la vida de toda la población costarricense y en determinados casos la de seres vivos del reino animal.

1.8.3. Misión, Visión y Valores

Misión

Brindar a la sociedad costarricense protección cuando la vida, los bienes y el medio ambiente se encuentran amenazados por incendios y situaciones de emergencia, basados en los más altos principios humanos y en la búsqueda permanente de la excelencia.

Visión

Ser una organización estatal de primera respuesta reconocida por sus altos estándares de calidad, eficacia y eficiencia, al atender las emergencias de su competencia y proveer servicios de prevención de incendios que integralmente, contribuyan al desarrollo del país, mediante la mejora de los índices de protección a la vida, la propiedad y el medio ambiente.

Valores

Los valores del Benemérito Cuerpo de Bomberos de Costa Rica son:

- Abnegación
Actitud voluntaria para ayudar a las personas sin esperar nada a cambio.
- Honor
Cualidad para comportarse apropiadamente ante el deber.
- Disciplina
Actitud para acatar las normas, protocolos, lineamientos y procedimientos que rigen las actividades de la organización.

- Trabajo en Equipo
Actitud de participación en forma conjunta de todos sus miembros en la ejecución de sus actividades.
- Solidaridad
Actitud de fraternidad para identificarse con las personas afectadas por condiciones adversas.
- Servicio
Disposición de respuesta con actitud de entrega, colaboración y espíritu de atención.
- Responsabilidad
Cumplir los deberes y competencias, realizando, de manera correcta, las actividades encomendadas.
- Preservación del Patrimonio
Actitud para legar el conocimiento, costumbres, tradiciones y salvaguardar los bienes de la organización para concienciar y preservar la historia.
- Honestidad
Actuar con apego a los principios y valores éticos.
- Equidad
Brindar servicios sin distinción de género, raza, condiciones sociales, de aseguramiento, económicas, entre otras.

1.9. Estado de la Cuestión

Ejecución de la Revisión

Ejecución de la selección en la fuente IEEE y Google Académico

Se detalle cómo se realizó la ejecución de la revisión en las fuentes de datos IEEE y Google Scholar, en cuanto a cómo se adaptó la cadena de búsqueda al motor en cuestión, los resultados obtenidos y el análisis de cada estudio primario.

Selección de estudios iniciales

Las búsquedas en la IEEE y Google Académico fueron realizadas mediante la opción de “búsqueda avanzada”, seleccionando las siguientes opciones:

- Búsqueda en: title, abstract y keywords.
- Sources: journals, book series y handbooks.
- Subject: Business intelligence, Inteligencia de negocios.
- Datas: 2008 -2015
- Idioma: Inglés y español.

Una vez configurada la búsqueda se pasó a realizar las consultas adaptando la cadena de búsquedas a la sintaxis propia de la fuente.

Books:(("Inteligencia de Negocios") OR ("Minería de Datos" AND Algoritmos)) AND OLAP) –Alumnos -Emocional

Tabla 1: Cadena de Búsqueda adaptada a la IEEE y Google Académico

(("Business intelligence") AND ("Data Mining" AND Algorithms) AND OLAP)

Tabla 2: Cadena de Búsqueda adaptada a la IEEE y Google Académico

La ejecución de la búsqueda en Google Académico nos arrojó 12 resultados y en IEE arrojó una cantidad muy elevada de resultados, por lo que se optó por tomar tres documentos con más relevancia para la investigación como criterio de inclusión. Luego de aplicar el criterio de exclusión se consideraron los siguientes documentos como estudios primarios:

	<p>On data integration and data mining for developing business intelligence</p> <p>Ping-Tsai Chung; Chung, S.H.</p> <p>Systems, Applications and Technology Conference (LISAT), 2013 IEEE Long Island, Pages 1-6</p>
	<p>An intelligent framework (O-SS-E) for data mining, knowledge discovery and business intelligence</p> <p>Rennolls, K.</p> <p>Database and Expert Systems Applications, 2005. Proceedings. Sixteenth International Workshop on, Pages 715-719</p>
	<p>Research on Extract, Transform and Load(ETL) in Land and Resources Star Schema Data Warehouse</p> <p>Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on Qin Hanlin; Jin Xianzhen; Zhang Xianrong, Pages 120-123</p>
	<p>La inteligencia del negocio: Etapas del proceso</p> <p>Vásquez Castrillón Jhon Bayron, Sucerquia Osorio Andrés, 2011 Páginas 1-5</p>
	<p>An Application of Business Intelligence Based on Patent in Data Integration and Analysis</p> <p>Dongsheng Zhai; Wenhui He</p> <p>Web Information Systems and Mining (WISM), 2010 International Conference on, 2010, Pages 288-292</p>
	<p>Análisis de los sistemas de ayuda a la toma de decisiones y su aplicación a los procesos empresariales</p> <p>Año 2010</p>

Páginas 1-71
Hernández Ruedas, Jordi
Inteligencia de negocios: Estado del arte
Vol. 1, núm. 44 (2010)
<i>Alveiro Alonso Rosado Gómez, Dewar Willmer Rico Bautista</i>

Tabla 3: Fuentes seleccionadas para el estudio del estado de la cuestión

Identificación	
Título:	On data integration and data mining for developing business intelligence
Publicación:	Systems, Applications and Technology Conference (LISAT), 2013 IEEE Long Island
Autores:	Ping-Tsai Chung; Chung, S.H.
Referencia	Inteligencia de Datos y Minería de Datos
Descripción	
Área:	Minería de Datos e Inteligencia de Negocios
Resumen	En este trabajo se discuten dos estudios de caso sobre la integración de datos y minería de datos. El primer caso es para los análisis de datos utilizando técnicas tradicionales de bases de datos relacionales , tales como bases de datos Oracle y la herramienta Cognos BI para integrar la minería y el sitio web de una empresa.
Aspectos a destacar	
	Tecnologías con Web Services,

Bunas herramientas y técnicas de minería de datos

Tabla 4: Ficha técnica de la fuente “On data integration and data mining for developing business intelligence”

Identificación	
Título:	An intelligent framework (O-SS-E) for data mining, knowledge discovery and business intelligence
Publicación:	Database and Expert Systems Applications, 2005. Proceedings. Sixteenth International Workshop on
Autores:	Rennolls, K
Referencia	
Descripción	
Área:	Minería de Datos
Resumen	Vincula las ontologías de datos y modelos que proporcionan dicha orientación, y una teoría del conocimiento que permita la evaluación de los conocimientos.
Aspectos a destacar	
	Sugerencias front-end y back, para el desarrollo de metodologías.

Tabla 5: Ficha técnica de la fuente “An intelligent framework (O-SS-E) for data mining, knowledge discovery and business intelligence”

Identificación	
Título:	Research on Extract, Transform and Load (ETL) in Land and

	Resources Star Schema Data Warehouse
Publicación:	Computational Intelligence and Design (ISCID), 2012 Fifth International Symposium on
Autores:	Qin Hanlin; Jin Xianzhen; Zhang Xianrong
Referencia	
Descripción	
Área:	Inteligencia de Negocios
Resumen	En este trabajo de investigación de las tecnologías clave de ETL, incluyendo los datos de extracción, los datos de transformación, los datos de carga gradual y transmisión de puntos de ruptura.
Aspectos a destacar	
	Métodos de extracción de datos ETL, transformación de datos, métodos de carga gradual.

Tabla 6: Ficha técnica de la fuente “Research on Extract, Transform and Load (ETL) in Land and Resources Star Schema Data Warehouse”

Identificación	
Título:	An Application of Business Intelligence Based on Patent in Data Integration and Analysis
Publicación:	Web Information Systems and Mining (WISM), 2010 International Conference on
Autores:	Dongsheng Zhai; Wenhui He
Referencia	
Descripción	

Área:	Inteligencia de Negocios
Resumen	Este documento sugiere la estructura de integración de datos y el análisis basado en la inteligencia de negocio (BI) para ayudar a las empresas a tomar las decisiones eficaces acerca de la estrategia y orientación del desarrollo tecnológico mediante la extracción de información.
Aspectos a destacar	Se centra en los procesos que conlleva el implementar un modelo de BI, desde la extracción de bases de datos heterogeneas, asi como la tranformación y limpieza de los mismos. Análisis de KPI y publicación con herramientas como Share Point

Tabla 7: Ficha técnica de la fuente “An Application of Business Intelligence Based on Patent in Data Integration and Analysis”

Identificación	
Título:	Análisis de los sistemas de ayuda a la toma de decisiones y su aplicación a los procesos empresariales
Publicación:	Web Information Systems and Mining (WISM), 2010 International Conference on
Autores:	Hernández Ruedas, Jordi
Referencia	
Descripción	
Área:	Ingeniería del Software

Resumen	El presente trabajo, aborda en primer lugar, una revisión de la literatura del proceso de toma de decisiones y de los sistemas de apoyo a la toma de decisiones en las organizaciones.
Aspectos a destacar	
Planteamiento de teorías, para la toma de decisiones	

Tabla 7: Ficha técnica de la fuente “Análisis de los sistemas de ayuda a la toma de decisiones y su aplicación a los procesos empresariales”

Identificación	
Título:	Inteligencia de negocios: Estado del arte
Publicación:	
Autores:	Vásquez Castrillón Jhon Bayron, Sucerquia Osorio Andrés
Referencia	
Descripción	
Área:	Inteligencia de Negocios
Resumen	La investigación comienza con la definición y aplicaciones de BI; además se muestran trabajos relevantes en algunas de las herramientas para hacer BI, como son Data Warehouse (Bodega de Datos), Olap (Cubos Procesamiento Analítico en Línea), Balance Scorecard (Cuadro de Mando) y Data Mining (Minería de Datos).
Aspectos a destacar	
Balance Scorecard	
Data Warehouse	

Olap
Data Mining

Tabla 8: Ficha técnica de la fuente “Inteligencia de negocios: Estado del arte”

Identificación	
Título:	El impacto de las herramientas de inteligencia de negocios en la toma de decisiones de los ejecutivos.
Publicación:	
Autores:	Calzada, Leticia y José Luis Abreu
Referencia	
Descripción	
Área:	Inteligencia de Negocios
Resumen	Se presenta en este trabajo de investigación una serie de conceptos, procedimientos y técnicas que forman parte de una metodología encaminada a crear sistemas de Inteligencia de Negocios. Con estas herramientas tecnológicas y con los elementos necesarios para planear, crear e implementar soluciones de Inteligencia de Negocios los ejecutivos de las empresas podrán tomar mejores decisiones
Aspectos a destacar	
Beneficios y Desventajas, procedimientos y técnicas	

Tabla 8: Ficha técnica de la fuente “Inteligencia de negocios: Estado del arte”

Capítulo 2. Marco Teórico

¿Qué es inteligencia de Negocios?

Inteligencia de negocios es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

Desde un punto de vista más práctico, y asociándolo directamente con las tecnologías de la información, podemos definir inteligencia de negocios como el conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa en implementaciones de análisis OLAP, dando así soporte a la toma de decisiones sobre el negocio. (Sinnexus, Sinnexus Business Intelligence Informática estratégica, s.f.)

La inteligencia de negocios actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto, crecimiento interno, administración de los recursos entre otros.

Así bien se establece que el objetivo principal de la inteligencia de negocios es ser un proceso de tratamiento e integración de los datos, para convertirlos en

información veraz utilizando herramientas de visualización, esto permite generar conocimiento a los ejecutivos de las empresas, administradores de empresas y otros usuarios finales a tomar decisiones de negocio.

Un dato se define como Información en bruto o desorganizada formada por: letras del alfabeto, números o símbolos. Que se refieren o representan, las condiciones, ideas o los objetos. Los datos son ilimitados y están presentes en todo el universo (Data, s.f.).

La información está constituida por un grupo de datos ya supervisados y ordenados, que sirven para construir un mensaje basado en un cierto fenómeno o ente. La información permite resolver problemas y tomar decisiones, ya que su aprovechamiento racional es la base del conocimiento. (Información, s.f.)

El conocimiento es un conjunto de información almacenada mediante la experiencia. En el sentido más amplio del término, se trata de la posesión de múltiples datos interrelacionados que al ser tomados por si solos, poseen un menor valor cualitativo. (Conocimiento, s.f.)

Modelo de Datos

Los datos se encuentran en su mayoría en la organización, normalmente estos están aislados y en diferentes fuentes y formatos, en modelos estructurados o simples archivos.

En la mayoría de casos los datos transaccionales por diferentes razones pierden su calidad, ya sea por no cumplir con reglas de negocio o errores producto de los sistemas que reciben estos datos. Por lo que antes de iniciar una implementación de BI dentro del modelo de datos debe existir una revisión de

calidad de datos y de ser necesario aplicar técnicas que ayuden a mejorar la calidad de los datos para obtener mejores resultados.

Calidad de datos

Según Jonathan G. Geiner define calidad de datos como la conformidad con los requisitos válidos del negocio y no necesariamente significa que debe tener cero defectos (Geiger), para definir calidad se debe:

- Determinar quién establecen los requisitos.
- Determinar cómo se establecen los requisitos.
- Determinar el grado de conformidad que se necesita.
- La calidad de datos en la inteligencia de negocios

En tecnologías de la información se habla sobre un concepto el cual define que si se tiene una entrada de datos con errores, se tendrá como salida datos con errores, a esto se le conoce como GIGO (Garbage in – Garbage Out). Por eso es que la importancia de la calidad en los datos en proyectos de BI (Business Intelligence o Inteligencia de Negocios) sobrepasa cualquier consideración que podamos hacer sobre otros elementos importantes en entornos de BI, su papel más que importante es fundamental, porque en los datos reside la clave de éxito de cualquier proyecto de esta índole. Se pueden utilizar técnicas de calidad de datos como la limpieza de datos que intenta resolver la problemática de la detección y corrección en los datos, con el fin de mejorar su calidad.

Extracción de datos

Estos datos deben ser extraídos, integrados y sincronizados en una sola base de datos permanentemente lo que se denomina como “Data Warehouse” o almacén de conocimiento organizacional. El proceso de creación y enriquecimiento del “Data Warehouse” se conoce como “Data warehousing”. Este proceso obtiene de las diferentes fuentes, a través de reglas de actualización, validación y seguridad; la información base del modelo. Puede realizarse de muchas formas y niveles de complejidad según el proyecto o fuente de datos. A ello se le conoce como ETL.

El ETL es un lenguaje para el entendimiento y sincronización de datos entre diferentes plataformas tecnológicas, el cual respeta los esquemas de seguridad de la fuente e interpreta las reglas de acceso necesarias. Se da mediante los procesos de:

- Extracción: obtención de información de las distintas fuentes tanto internas como externas.
- Transformación: filtrado, limpieza, depuración, homogeneización y agrupación de la información.
- Carga: organización y actualización de los datos y los metadatos en la base de datos.

El ETL alimenta los Data Mart según las reglas de vigencia y el diseño establecido por el equipo de consultores. Un subconjunto del Data Warehouse, que involucre datos y reglas de negocio específicos para cada área o proyecto de Inteligencia de Negocio se denomina “Data Mart”. Cada área tiene un “Data Mart” que constituye su base de datos particular. Sobre los Data Mart se realizan consultas y reportes resumidos y consolidados de acuerdo a las reglas de

integración, simplificando muchos procesos de detalle, consolidación y enriquecimiento de los datos en el origen. Sin embargo, el “Data Warehouse” o un Data Mart, en si no son el objetivo final de los proyectos de Inteligencia de Negocio, solamente constituye una herramienta de soporte a las aplicaciones en este sentido. Al “Data Warehouse” o Data Mart se le pueden aplicar técnicas de análisis de datos, como lo es la minería de datos o herramientas OLAP. (Sinnexus, Sinnexus Business Intelligence Informática estratégica, s.f.).

Data Warehouse

Palabra inglesa cuyo significado no es otro que almacén de datos, nació en la década de los 80 ante la necesidad de desarrollar un sistema de almacenamiento de datos que garantizase la fluidez, por lo tanto se define como un contenedor en el que se almacenan los datos procedentes de las distintas fuentes que puedan existir en una organización, quedando éstos integrados, depurados y ordenados en una única base de datos centralizada, donde los datos se almacenan durante el tiempo requerido, para cumplir con las necesidades del negocio.

Data Mart

Es un sistema orientado a la consulta, cuya distribución interna de los datos es clara y no hay dudas al respecto, estando éstos estructurados en modelos dimensionales de estrella o copo de nieve. Sin embargo, no sé puede decir lo mismo del Data Warehouse, para el que hay diferentes enfoques en cuanto a sus características y funciones.

Estructuras de Inteligencia de Negocios

Arquitectura de diseño para Data Warehouse

Cuando se desarrollan implementaciones para diseños de Data Warehouse, existen diferentes propuestas pero las que más se han usado y dado resultado son dos:

Bill Inmon

Él ve la necesidad de transferir la información de los diferentes OLTP de las organizaciones a un lugar centralizado donde los datos puedan ser utilizados para el análisis. En otras palabras, dentro de una empresa que tiene un Data Warehouse, los Data Marts obtienen su información a partir de este Data Warehouse.



Figura 1: Paradigma: Bill Inmon

Fuente de <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimENSIONAL/>

Ralph Kimball

Él establece que el Data Warehouse es un conglomerado de todos los Data Marts dentro de una empresa, siendo una copia de los datos transaccionales estructurados de una forma especial para el análisis, de acuerdo al Modelo Dimensional (no normalizado), el cual incluye las dimensiones de análisis y sus atributos, su organización jerárquica, así como los diferentes hechos de negocio que se quieren analizar.

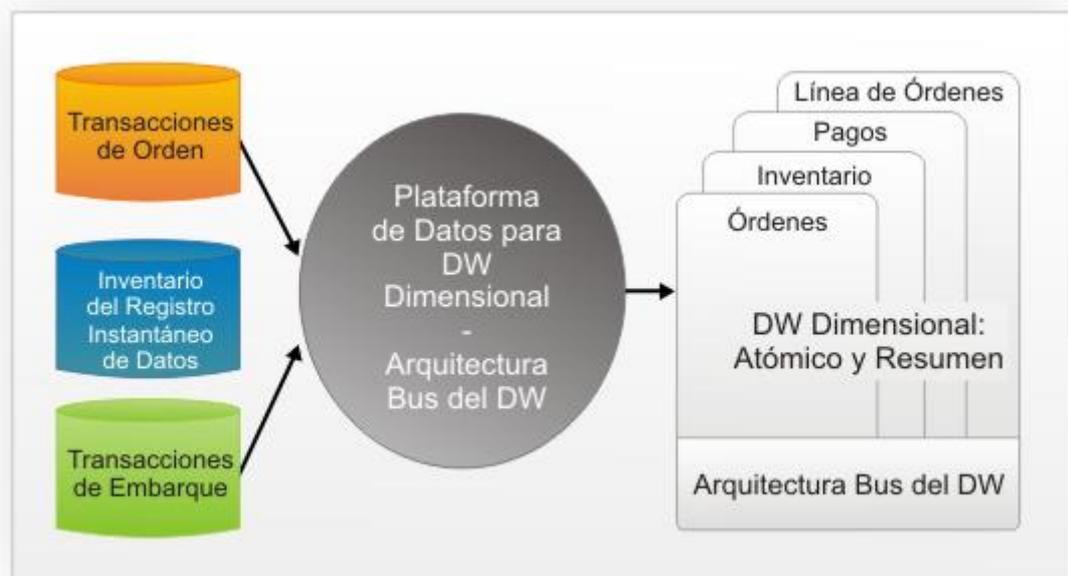


Figura 2: Paradigma: Ralph Kimball

Fuente de <https://churriwifi.wordpress.com/2010/04/19/15-2-ampliacion-conceptos-del-modelado-dimencional/>

Minería de Datos

El data mining (minería de datos), es el conjunto de técnicas y tecnologías que permiten explorar un almacén de datos de manera automática o semiautomática con el objetivo de encontrar patrones repetitivos, tendencias, predicciones o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Existen diferentes bases de algoritmos ya creados y adaptables para ser utilizados en un conjunto de información y poder realizar análisis de predicción en ellos, para efectos del proyecto se definen los algoritmos proporcionados por Microsoft en su sitio web (Microsoft, Microsoft, s.f.). Utilizando su servicio de Analysis Services que se explicaran a continuación:

Algoritmo de asociación de Microsoft

Es un algoritmo de asociación suministrado por Analysis Services, útil para escenarios donde tenga como fin recomendar un (producto, servicio, etc.) esto basado en los elementos que ya han adquirido o en los que tiene interés. Estos modelos de asociación se generan basándose en conjuntos de datos que contienen identificadores para casos individuales y para casos que contienen casos dentro de sí.

¿Cómo funciona el algoritmo?

El algoritmo recorre un conjunto de datos para hallar elementos que aparezcan juntos en un caso, seguido agrupa en conjuntos de elementos todos los elementos asociados que aparecen como mínimo en el número de casos especificado en el parámetro.

Datos Requeridos

Una columna llave: cada modelo debe contener una columna numérica o de texto que identifique cada registro

Una única columna de predicción: Solo puede tener una columna de predicción, normalmente se trata de la columna clave de la tabla, estos valores deben ser discretos o discretizados.

Por variable discreta se define en aceptar solamente valores que pertenecen a un conjunto de datos.

Columnas de entrada: Las columnas de entrada deben ser discretas. Los datos de entrada de un modelo de asociación suelen encontrarse en dos tablas. Es posible incluir estos datos en un modelo mediante el uso de una tabla anidada.

Rendimiento

El proceso de crear conjuntos de elementos y contar las correlaciones puede llevar mucho tiempo. Aunque el algoritmo “Reglas de asociación de Microsoft” utiliza técnicas de optimización para ahorrar espacio y agilizar procesamiento, es prudente tener en mente que pueden presentar problemas de rendimiento cuando:

- El conjunto de datos es grande y tiene muchos elementos individuales.
- El tamaño mínimo establecido para el conjunto de elementos es demasiado pequeño.

Algoritmo de árboles de decisión de Microsoft

Es un algoritmo de clasificación y regresión para modelos de predicción de atributos discretos y continuos.

Para los atributos discretos, el algoritmo hace predicciones basándose en las relaciones entre las columnas de entrada de un conjunto de datos. Utiliza los valores, conocidos como estados de estas columnas para predecir los estados de una columna que se designa como elemento de predicción. Específicamente, el algoritmo identifica las columnas de entrada que se correlacionan con la columna de predicción. En otras palabras el árbol de decisión realiza predicciones basándose en la tendencia hacia un resultado concreto.

Para los atributos continuos, el algoritmo usa la regresión lineal para determinar donde se divide el árbol de decisión.

¿Cómo funciona el algoritmo?

El algoritmo genera un modelo de minería de datos mediante la creación de una serie de divisiones en el árbol. Estas divisiones se representan como nodos. El algoritmo agrega un nodo al modelo cada vez que una columna de entrada tiene una correlación significativa con la columna de predicción, la forma en que el algoritmo determina una división varía en función de si predice una columna continua o una columna discreta.

Datos requeridos

Una columna llave: cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.

Una columna de predicción: Se requiere al menos una columna de predicción. Puede incluir varios atributos de predicción en un modelo y pueden ser de tipos

diferentes, numérico o discreto. Sin embargo, el incremento del número de atributos de predicción puede aumentar el tiempo de procesamiento.

Columnas de entrada: Se requieren columnas de entrada, que pueden ser discretas o continuas. Aumentar el número de atributos de entrada afecta al tiempo de procesamiento.

Rendimiento

Si se utiliza demasiados atributos de predicción o de entrada al diseñar un modelo de minería de datos, el modelo puede tardar mucho tiempo en procesarse o incluso quedarse sin memoria, para ello los algoritmos de Microsoft utilizan la selección de características para guiar la selección de atributos más útiles.

Algoritmo de clústeres de Microsoft

El algoritmo utiliza técnicas iterativas para agrupar los casos de un conjunto de datos dentro de clústeres que contienen características similares. Estas agrupaciones son útiles para la exploración de datos, la identificación de anomalías en los datos y la creación de predicciones.

El algoritmo de clústeres se diferencia de otros algoritmos de minería de datos, como los árboles que no tiene que designar una columna de predicción para generar un modelo de agrupación de clústeres. El algoritmo entrena el modelo de forma estricta a partir de las relaciones que existen en los datos y de los clústeres que identifica el algoritmo.

¿Cómo funciona el algoritmo?

El algoritmo de clústeres de Microsoft identifica primero las relaciones de un conjunto de datos y genera una serie de clústeres basándose en ellas. Un gráfico de dispersión es una forma útil de representar visualmente el modo en que el algoritmo agrupa los datos.

Después de definir los clústeres, el algoritmo calcula el grado de perfección con que los clústeres representan las agrupaciones de puntos e intenta volver a definir las agrupaciones para crear clústeres que representen mejor los datos. El algoritmo establece una iteración en este proceso hasta que ya no es posible mejorar los resultados.

Datos requeridos

Una columna llave: cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.

Columnas de entrada: Cada modelo debe tener al menos una columna de entrada que contenga los valores que se utilizan para generar los clústeres. Puede tener tantas columnas de entrada como desee, pero dependiendo del número de valores existentes en cada columna, la adición de columnas adicionales podría aumentar el tiempo necesario para entrenar el modelo.

Una columna de predicción opcional: El algoritmo no necesita una columna de predicción para generar el modelo, pero puede agregar una columna de predicción de casi cualquier tipo de datos. Los valores de la columna de predicción se pueden tratar como entradas del modelo de agrupación en clústeres, o se puede especificar que solo se utilicen para las predicciones.

Algoritmo de regresión lineal de Microsoft

El algoritmo de regresión lineal de Microsoft es una variación del algoritmo de árboles de decisión de Microsoft que ayuda a calcular una relación lineal entre una variable independiente y otra dependiente.

Cada punto de datos del diagrama tiene un error asociado con su distancia con respecto a la línea de regresión. Los coeficientes a y b de la ecuación de regresión ajustan el ángulo y la ubicación de la recta de regresión. Puede obtener la ecuación de regresión ajustando a y b hasta que la suma de los errores asociados a todos los puntos alcance su valor mínimo.

¿Cómo funciona el algoritmo?

El algoritmo de regresión lineal de Microsoft es una variación del algoritmo de árboles de decisión de Microsoft, eso sí con parámetros que restringen el comportamiento del algoritmo y requieren ciertos tipos de datos de entrada.

Datos requeridos

Una columna llave: cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.

Una columna de predicción: Se requiere al menos una columna de predicción. Se pueden incluir varios atributos de predicción en un modelo, pero deben ser tipos de datos numéricos continuos. No se puede utilizar un tipo de datos de fecha y hora como atributo de predicción aunque el almacenamiento nativo para los datos sea numérico.

Columnas de entrada: Deben contener datos numéricos continuos y se les debe asignar el tipo de datos adecuado.

Algoritmo de regresión logística

La regresión logística es una técnica estadística conocida que se usa para modelar los resultados binarios.

Existen varias implementaciones de regresión logística en la investigación estadística, que utilizan diferentes técnicas de aprendizaje. El algoritmo de regresión de Microsoft se ha implementado utilizando una variación del algoritmo de Red neuronal de Microsoft. Este algoritmo comparte muchas de las cualidades de las redes neuronales, pero es más fácil de entrenar.

Una de las ventajas de la regresión logística es que el algoritmo es muy flexible, puede tomar cualquier tipo de entrada y admite varias tareas analíticas diferentes:

- Usar datos demográficos para realizar predicciones sobre los resultados, como el riesgo de contraer una determinada enfermedad.
- Explorar y ponderar los factores que contribuyen a un resultado. Por ejemplo, buscar los factores que influyen en los clientes para volver a visitar un establecimiento.
- Clasificar los documentos, el correo electrónico u otros objetos que tengan muchos atributos.

¿Cómo funciona el algoritmo?

La implementación de Microsoft usa una red neuronal modificada para modelar las relaciones entre las entradas y los resultados. Se mide el efecto de cada entrada en el resultado y se ponderan las diversas entradas en el modelo acabado.

Datos requeridos

Una columna de una sola clave: cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.

Columnas de entrada: cada modelo debe tener al menos una columna de entrada que contenga los valores que se utilizan como factores en el análisis. Puede tener tantas columnas de entrada como desee, pero dependiendo del número de valores existentes en cada columna, la adición de columnas adicionales podría aumentar el tiempo necesario para entrenar el modelo.

Al menos una columna de predicción: el modelo debe contener al menos una columna de predicción de cualquier tipo de datos, incluidos datos numéricos continuos. Los valores de la columna de predicción también se pueden tratar como entradas del modelo, o se puede especificar que solo se utilicen para las predicciones. No se admiten tablas anidadas en las columnas de predicción, pero se pueden usar como entradas.

Algoritmo Bayes naive de Microsoft

El algoritmo Bayes naive de Microsoft es un algoritmo de clasificación basado en los teoremas de Bayes y que Microsoft SQL Server Analysis Services proporciona para el modelado de predicción. La palabra naïve (ingenuo en inglés)

del término Bayes naive proviene del hecho que el algoritmo utiliza técnicas Bayesianas pero no tiene en cuenta las dependencias que puedan existir.

Desde el punto de vista computacional, el algoritmo es menos complejo que otros algoritmos de Microsoft y, por tanto, resulta útil para generar rápidamente modelos de minería de datos que detectan las relaciones entre las columnas de entrada y las columnas de predicción. Puede utilizar este algoritmo para realizar la exploración inicial de los datos y, más adelante, aplicar los resultados para crear modelos de minería de datos adicionales con otros algoritmos más complejos y precisos desde el punto de vista computacional.

¿Cómo funciona el algoritmo?

El algoritmo Bayes naive de Microsoft calcula la probabilidad de cada estado de cada columna de entrada, dado cada posible estado de la columna de predicción.

Datos requeridos

Una columna de una sola clave: cada modelo debe contener una columna numérica o de texto que identifique cada registro de manera única. No están permitidas las claves compuestas.

Columnas de entrada: en un modelo Bayes naive, todas las columnas deben ser discretas o discretizadas. En un modelo Bayes naive, también es importante asegurarse de que los atributos de entrada sean independientes unos de otros. Esto es particularmente importante al utilizar el modelo para la predicción. El motivo es que, si utiliza dos columnas de datos que ya están estrechamente relacionadas, el efecto sería multiplicar la influencia de esas columnas, lo que puede ocultar otros factores que influyen en el resultado.

Al menos una columna de predicción: el atributo de predicción debe contener valores discretos o discretizados.

Algoritmo de red neuronal de Microsoft

En SQL Server Analysis Services, el algoritmo de red neuronal de Microsoft combina cada posible estado del atributo de entrada con cada posible estado del atributo de predicción, y usa los datos de entrenamiento para calcular las probabilidades. Posteriormente, puede usar estas probabilidades para la clasificación o la regresión, así como para predecir un resultado del atributo de predicción basándose en los atributos de entrada.

Los modelos de minería de datos construidos con el algoritmo de red neuronal de Microsoft pueden contener varias redes, en función del número de columnas que se utilizan para la entrada y la predicción, o solo para la predicción. El número de redes que contiene un único modelo de minería de datos depende del número de estados que contienen las columnas de entrada y las columnas de predicción que utiliza el modelo.

¿Cómo funciona el algoritmo?

El algoritmo de red neuronal de Microsoft crea una red formada por hasta tres niveles de neuronas. Estas capas son una capa de entrada, una capa oculta opcional y una capa de salida.

Nivel de entrada: las neuronas de entrada definen todos los valores de atributos de entrada para el modelo de minería de datos, así como sus probabilidades.

Nivel oculto: las neuronas ocultas reciben entradas de las neuronas de entrada y proporcionan salidas a las neuronas de salida. El nivel oculto es donde se asignan pesos a las distintas probabilidades de las entradas. Un peso describe la relevancia o importancia de una entrada determinada para la neurona oculta. Cuanto mayor sea el peso asignado a una entrada, más importante será el valor de dicha entrada. Los pesos pueden ser negativos, lo que significa que la entrada puede desactivar, en lugar de activar, un resultado concreto.

Nivel de salida: las neuronas de salida representan valores de atributo de predicción para el modelo de minería de datos.

Datos Requeridos

El modelo de red neuronal debe contener una columna de clave, una o más columnas de entrada y una o más columnas de predicción.

Los modelos de minería de datos que usan el algoritmo de red neuronal de Microsoft están muy influenciados por los valores que se especifican en los parámetros disponibles para el algoritmo. Los parámetros definen cómo se muestrean los datos, cómo se distribuyen o cómo se espera que estén distribuidos en cada columna, y cuándo se invoca la selección de características para limitar los valores usados en el modelo final.

Algoritmo de clústeres de secuencia de Microsoft

El algoritmo de clústeres de secuencia de Microsoft es un algoritmo de análisis de flujos que proporciona Microsoft SQL Server Analysis Services. Puede utilizar este algoritmo para explorar los datos que contienen eventos que pueden vincularse mediante rutas o secuencias. El algoritmo encuentra las secuencias más

comunes mediante la agrupación, o agrupación en clústeres, de las secuencias que son idénticas, es similar en muchas maneras al algoritmo de clústeres de Microsoft. Sin embargo, en lugar de encontrar clústeres de casos que contienen atributos similares, el algoritmo de clústeres de secuencia de Microsoft encuentra clústeres de casos que contienen rutas similares en una secuencia.

¿Cómo funciona?

El algoritmo de clústeres de secuencia de Microsoft es un algoritmo híbrido que combina técnicas de agrupación en clústeres con el análisis de cadenas de Markov para identificar los clústeres y sus secuencias. Una de las marcas distintivas del algoritmo de clústeres de secuencia de Microsoft es que utiliza los datos de las secuencias. Estos datos suelen representar una serie de eventos o transiciones entre los estados de un conjunto de datos. El algoritmo examina todas las probabilidades de transición y mide las diferencias, o las distancias, entre todas las posibles secuencias del conjunto de datos con el fin de determinar qué secuencias es mejor utilizar como entradas para la agrupación en clústeres.

Datos requeridos

Una columna de una sola llave: Un modelo de agrupación en clústeres de secuencia requiere una clave que identifique los registros.

Una columna de secuencia: para los datos de la secuencia, el modelo debe tener una tabla anidada que contenga una columna de identificador de secuencia. El identificador de secuencia puede ser cualquier tipo de datos ordenable. Por ejemplo, puede usar el identificador de una página web, un número entero o una cadena de texto, con tal de que la columna identifique los eventos en una secuencia. Solo se

admite un identificador de secuencia por cada secuencia y un tipo de secuencia en cada modelo.

Atributos opcionales no relacionados con la secuencia: el algoritmo admite la incorporación de otros atributos que no tengan que ver con las secuencias. Estos atributos pueden incluir las columnas anidadas.

Algoritmo de serie temporal de Microsoft

Proporciona los algoritmos de regresión que se optimizan para la previsión en el tiempo de valores continuos tales como las ventas de productos. Mientras que otros algoritmos de Microsoft, como por ejemplo los árboles de decisión, requieren columnas adicionales de nueva información como entrada para predecir una tendencia, los modelos de serie temporal no las necesitan. Un modelo de serie temporal puede predecir tendencias basadas únicamente en el conjunto de datos original utilizado para crear el modelo. Es posible también agregar nuevos datos al modelo al realizar una predicción e incorporar automáticamente los nuevos datos en el análisis de tendencias.

Una característica importante del algoritmo de serie temporal de Microsoft es su capacidad para llevar a cabo predicciones cruzadas. Si entrena el algoritmo con dos series independientes, pero relacionadas, puede utilizar el modelo generado para predecir el resultado de una serie basándose en el comportamiento de la otra.

¿Cómo funciona el algoritmo?

Al inicio el algoritmo de serie temporal de Microsoft utilizaba un único algoritmo llamado ARTXP este se optimizó para predicciones a corto plazo y por consiguiente predecía el siguiente valor probable en una serie. A partir de SQL

Server 2008 el algoritmo de serie temporal de Microsoft incorpora un segundo algoritmo, ARIMA. Este algoritmo está optimizado para la predicción a largo plazo. El algoritmo de serie temporal de Microsoft utiliza una mezcla de los dos algoritmos al analizar patrones y realizar predicciones. El algoritmo entrena dos modelos independientes sobre los mismos datos: uno de los modelos utiliza el algoritmo ARTXP y el otro modelo utiliza el algoritmo ARIMA, el algoritmo combina los resultados de los dos modelos para obtener la mejor predicción sobre un número variable de intervalos de tiempo. Dado que ARTXP obtiene mejores resultados en las predicciones a corto plazo, se le da mayor importancia al principio de una serie de predicciones. Sin embargo, a medida que los intervalos de tiempo que se están prediciendo se adentran en el futuro, se va dando más importancia a ARIMA

Datos requeridos

Una única columna llave de tiempo: Cada modelo debe contener una columna numérica o de fecha que se utilizará como serie de casos y que define los intervalos de tiempo que utilizará el modelo. El tipo de datos para la columna de clave temporal puede ser un tipo de datos datetime o bien numérico. Sin embargo, la columna debe contener valores continuos y éstos deben ser únicos para cada serie. La serie de casos para un modelo de serie temporal no pueden estar almacenada en dos columnas como por ejemplo una columna Año y una columna Mes.

Una columna predecible: Cada modelo debe contener por lo menos una columna predecible alrededor de la que el algoritmo generará el modelo de serie temporal. El tipo de datos de la columna predecible debe contener valores continuos. Por ejemplo, es posible predecir la manera en que los atributos numéricos tales como ingreso, ventas o temperatura, varían con el tiempo. Sin embargo, no es

posible utilizar como columna predecible una columna que contenga valores discretos tales como el estado de las compras o el nivel de educación.

Una columna de clave de serie opcional: Cada modelo puede tener una columna de clave adicional que contenga valores únicos que identifiquen a una serie. La columna de clave de serie opcional debe contener valores únicos.

OLAP

Es el acrónimo en inglés de procesamiento analítico en línea (Online Analytical Processing). Es una técnica cuyo objetivo es agilizar la consulta de grandes cantidades de datos. Para ello utiliza estructuras multidimensionales (Cubos OLAP) estas estructuras supera las limitaciones de las bases de datos relacionales y proporciona un análisis rápido de datos. Los cubos pueden mostrar y sumar grandes cantidades de datos, a la vez que proporcionan a los usuarios acceso mediante búsqueda a los puntos de datos. De este modo, los datos se pueden resumir o reorganizar según sea necesario, para procesar la variedad más amplia de preguntas pertinentes al área de interés de un usuario. (Microsoft, Microsoft System Center, s.f.)

Al usuario final se le presentan herramientas OLAP, están permiten organizar los datos en una estructura multidimensional (cubos OLAP), es decir con índices agrupadores y criterios de resumen producto de las relaciones entre los datos y respetando las reglas del modelo, todo esto sin que el usuario deba conocer la estructura interna del almacén de los datos.

En Resumen

Los sistemas y componentes de BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa típicamente que, en un Data Warehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado de datos. En este sentido, los procesos ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados.

En definitiva, una solución BI completa permite:

- Observar ¿qué está ocurriendo?
- Comprender ¿por qué ocurre?
- Predecir ¿qué ocurriría?
- Colaborar ¿qué debería hacer el equipo?
- Decidir ¿qué camino se debe seguir?

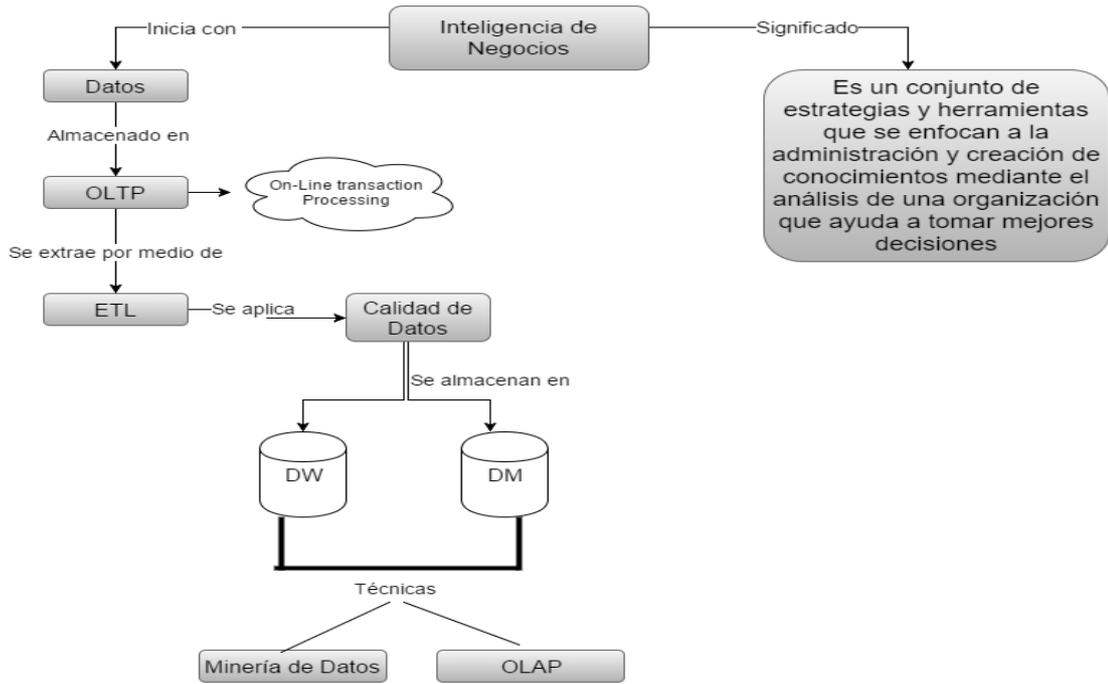


Figura 3: Diagrama Inteligencia de Negocios

Fuente: Elaboración Propia (2016)

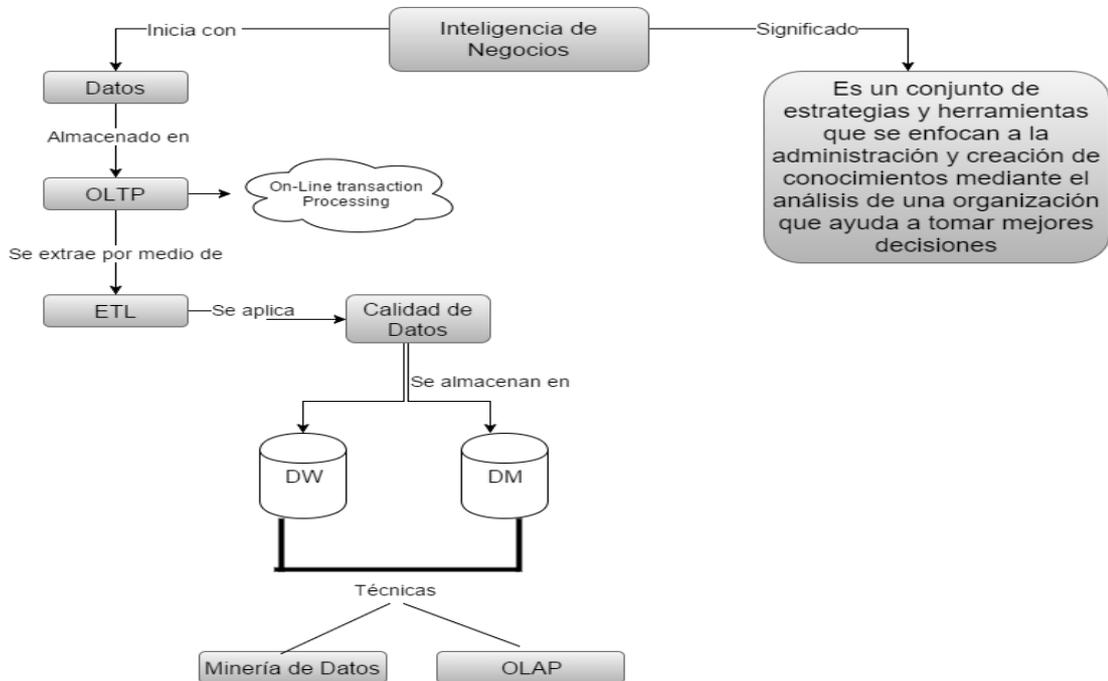


Figura 4: Diagrama Inteligencia de Negocios

Fuente: Elaboración Propia (2016)

Capítulo 3. Marco Metodológico

3.1. Tipo de Investigación

La investigación en curso es de tipo aplicada por que hace uso de los métodos del pasado, los conocimientos o teorías o de investigación básica para resolver un problema existente.

3.2. Alcance de la Investigación

La investigación tiene un alcance Explicativo, dado que el propósito final de la investigación es el responder las causas de relación entre variables (eventos, sucesos, comportamientos o fenómenos)

3.3. Enfoque

El enfoque de la investigación será mixto, cualitativo y cuantitativo. Cuantitativo porque consiste “en utilizar la recolección y el análisis de datos para contestar preguntas de investigación y probar hipótesis establecidas previamente, y confía en la medición numérica, el conteo y frecuentemente en el uso de la estadística para establecer con exactitud patrones de comportamiento en una población”.¹

¹ Roberto Hernández Sampieri, Carlos Fernández Collado, Pilar Baptista Lucio, Metodología de la Investigación, 4ta Edición, México D.F, 2006, 1998,1991, Pag. 5

Cualitativo porque “consiste en utilizar la recolección de datos sin medición numérica para descubrir o afinar preguntas de investigación en el proceso de interpretación”.²

3.4. Diseño

Etnometodológico

La Etnometodología fue propuesta por el sociólogo Harold Garfinkel al publicar "Studies in Ethnomethodology" (Estudios en Etnometodología) en el año 1967. Se trató de una respuesta crítica a la sociología estructural de su tiempo, ya que rechaza aquellos discursos sociales que calificaban la acción humana de automatista.

Aportes de la Etnometodología

La etnometodología es la que estudia los métodos o procedimientos con los que los integrantes de la sociedad dan sentido a la vida cotidiana o actúan en ella; desde la consideración de que el orden social está determinado por los continuos actos interpretativos de los sujetos Implicados.

Estudia los procedimientos por los que los miembros de otras disciplinas concretan su objeto inteligible de investigación. “Los estudios etnometodológicos analizan las actividades cotidianas como métodos que sus miembros usan para hacer que esas actividades sean racionalmente visibles y reportables para todos los efectos prácticos, es decir, “explicables”. (Harold , 1967)

² Roberto Hernández Sampieri, Carlos Fernández Collado, Pilar Baptista Lucio, Metodología de la Investigación, 4ta Edición, México D.F, 2006, 1998,1991, Pag. 8

Capítulo 4. Análisis de Diagnóstico

Actualmente el BCBCR, para realizar análisis de información o predicción sobre sus datos utiliza el sistema transaccional para extraer la información de las diferentes pantallas de donde lo pueda obtener, dicha información se exporta en documentos Excel para en un siguiente paso consolidar la información en un solo archivo, normalmente esta información es revisada contra un conteo de base de datos para garantizar la veracidad de dicha información. Cuando es validada se envía a los altos jerarcas del BCBCR, los cuales la analizan para hacer uso de ella en la toma de decisiones.

Para la predicción de eventos actualmente no se realiza ningún procedimiento manual o automático. En este caso para solventar este tipo de análisis acuden a la experiencia de funcionarios claves que se han desempeñado durante mucho tiempo en funcione de atención de emergencias.

Capítulo 5. Propuesta de Solución

El presente proyecto de tesis consistió en analizar, diseñar e implementar una solución de inteligencia de negocios , que permite realizar el análisis a nivel gerencial del total de atención de emergencias a nivel nacional del Benemérito Cuerpo de Bomberos de Costa Rica, la solución consiste en la construcción de un modelo de inteligencia de negocios que contiene extracción, calidad, transformación y almacenamiento de los datos, los cuales son presentados al usuario por medio de una herramienta con elementos gráficos, flexibles y de acceso eficiente, permitiendo con ello, lograr un análisis adecuado de los datos y distribuirlos por distintos filtros como fechas, ubicación geográfica, tipos de emergencias entre otros y dar facilidad a los usuarios para que interpreten mejor la información. De igual modo la solución permite la aplicación de minería de datos para realizar predicciones basadas en la información ya almacenada.

Para explicar el diseño de la solución se realizó una división por procesos, los cuales son explicados en pasos para una mayor comprensión de la solución desarrollada.

Exploración y calidad de los datos

Se inició con una exploración de los datos en el modelo de la base de datos del sistema transaccional que contiene toda la información referente a la atención de emergencias, esto para poder identificar las tablas de proceso y si la misma presenta problemas de calidad que tengan un fuerte impacto en los resultados finales de la solución. Para ello se realizan consultas en busca de campos clave con

valores nulos, vacíos o que presenten algún problema de formato en la tabla de proceso identificada como “Boleta_Incidente”, luego de este análisis se concluye que existe un grado alto de registros cuyo campo de zona geográfica constituido por provincia, cantón y distrito se encuentran con valores nulos en determinados años del total de información, para explicar mejor el problema se muestra el siguiente gráfico:

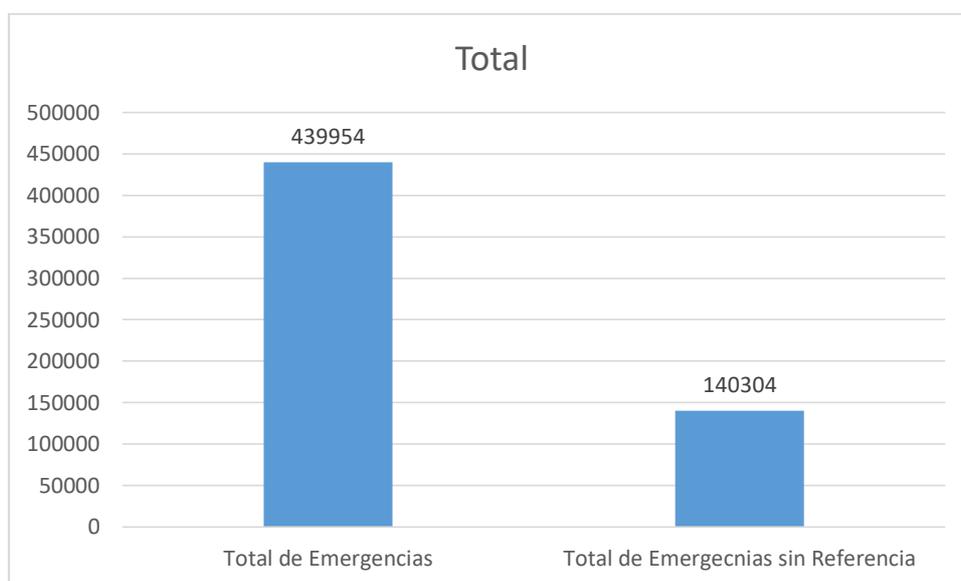


Figura 6: Total de Emergencias sin posición geográfica

Fuente: Elaboración Propia (2016)

Como se observa en la figura 4, existe mucha información que no está referenciada geográficamente 140304 para ser exactos de 439954 del total de registros. Lo cual representan un 32% aproximadamente del total, esto hace que para efectos de un proyecto de inteligencia de negocios el análisis de información y predicción de los datos se vea afectado al ser un dato sensible para la organización.

Para mitigar dicho incidente se desarrolló un servicio el cual consta de un método que obtiene todos los registros de la tabla “Boleta_incidente” objeto que se

encuentra en la base de datos SIGAE donde los campos provincia, cantón y distrito se encuentren vacíos e inicia el análisis de la siguiente forma:

Primer paso: Obtiene las provincias que existen en la base de datos y mediante un recorrido procesa provincia por provincia, utilizando la función contains() verifica si la provincia que se analiza en ese momento existe dentro del campo de texto “Direccion”, de ser así lo almacena dentro de una cadena de texto separando cada registro por un carácter especial para esta parte del algoritmo se utilizó “;”.

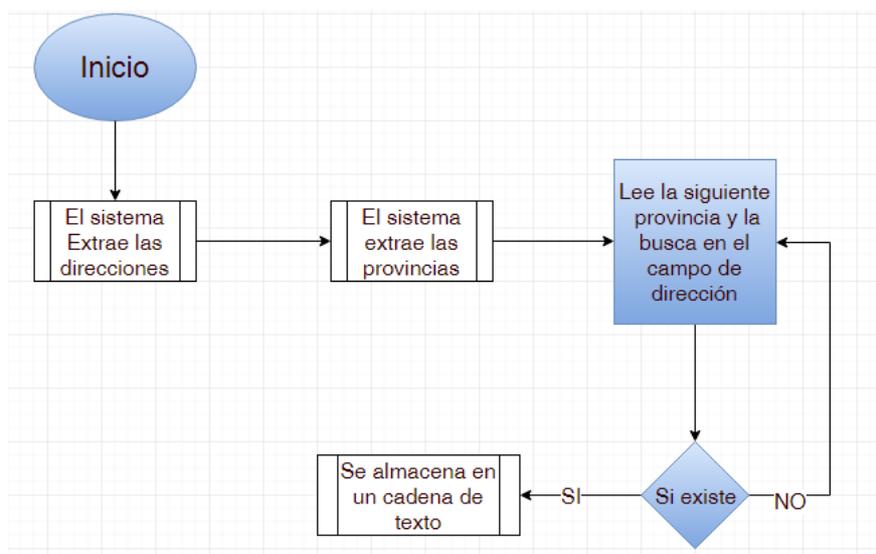


Figura 7: Diagrama de Flujo Primera Parte

Fuente: Elaboración Propia (2016)

Segundo paso: Obtiene los cantones que existen en la base de datos y mediante un recorrido procesa cantón por cantón, utilizando la función contains() verifica si el canton que se analiza en ese momento existe dentro del campo de texto “Direccion”, de ser así lo almacena dentro de una cadena de texto separando cada registro por un carácter especial para esta parte del algoritmo se usó “;”.

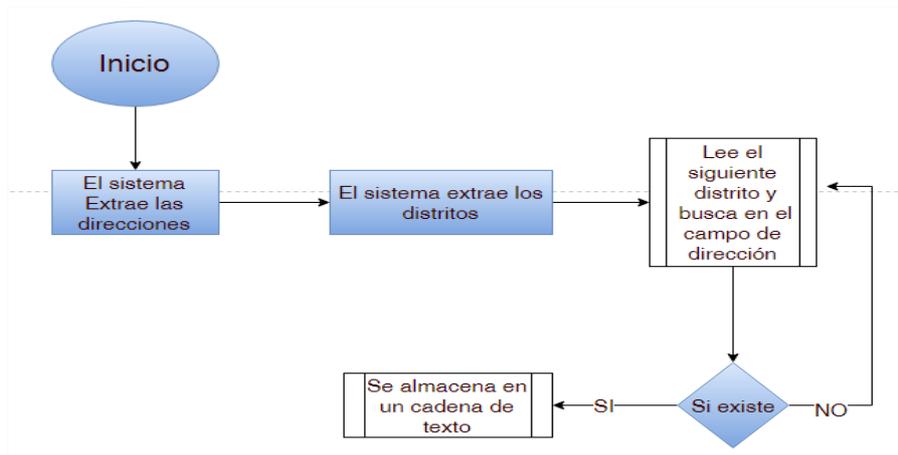


Figura 8: Diagrama de Flujo Segunda Parte

Fuente: Elaboración Propia (2016)

Tercer paso: Obtiene los distritos que existen en la base de datos y mediante un recorrido procesa distrito por distrito, utilizando la función contains() verifica si el distrito que se analiza en ese momento existe dentro del campo de texto “Direccion”, de ser así lo almacena dentro de una cadena de texto separando cada registro por un carácter especial para esta parte del algoritmo se usó “;”.

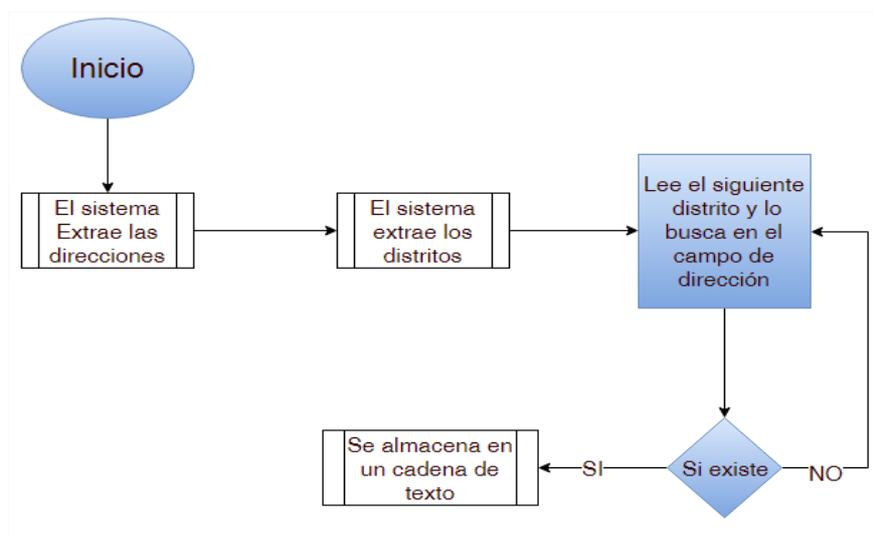


Figura 9: Diagrama de Flujo Tercera Parte

Fuente: Elaboración Propia (2016)

Cuarto paso: cada cadena de texto se convierte en un arreglo de string mediante la función `Split()`, seguido se inicia procesando el arreglo de provincias y cantones donde se hace un recorrido de cada provincia, y cada cantón para verificar la relación de ambos dentro de una posición geográfica. En otras palabras se toma la provincia y por cada cantón localizado se verifica si el mismo pertenece a esa provincia, de ser así se da por encontrada la relación se termina el análisis y sigue con los distritos. Los distritos se analizan de la misma forma, por cada cantón se analiza si el distrito existe en ese cantón para poder garantizar con exactitud el resultado final.

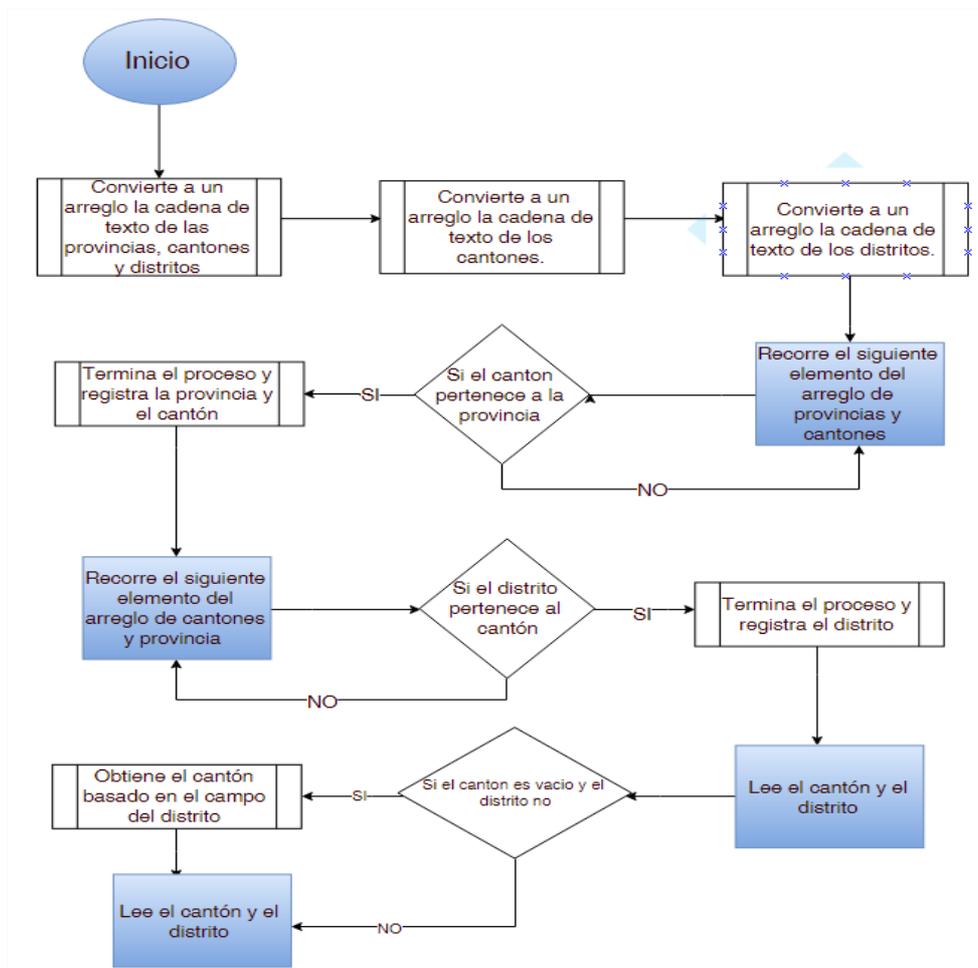


Figura 10: Diagrama de Flujo Cuarta Parte

Fuente: Elaboración Propia (2016)

Para una mejor comprensión se desarrolla el siguiente ejemplo:

Dirección: “SAN JOSE HATILLO COSTADO ESTE DEL RANCHO GUANACASTE, COLONIA 15 DE SETIEMBRE”

El algoritmo toma como palabras clave para provincia “SAN JOSE y GUANACASTE” y las concatena en una cadena de texto separado por un “;” en el caso de cantón no encuentra palabras claves y para distrito encuentra “HATILLO”, el siguiente paso es verificar si el distrito Hatillo existe en algún cantón de la provincia de Guanacaste, como no es así, ahora realiza la validación con San Jose, en donde

encuentra que la palabra Hatillo existe como distrito del cantón San José de la provincia San José, si para este mismo ejemplo la provincia que se analiza en primera instancia hubiera sido San José encuentra los datos de cantón y distrito y da por finalizada la búsqueda, ya no toma en cuenta el siguiente dato para provincia.

Estos resultados son mostrados en pantalla ante el negocio, para que establezca cuales emergencias son aceptables aplicar el cambio y poder asignar la provincia, cantón y distrito a los registros que antes no los tenían.

Por aplicado el proceso se obtiene una mejora sustancial en la calidad de los datos sin posición geográfica, encontrando dicha ubicación por medio del campo de dirección.

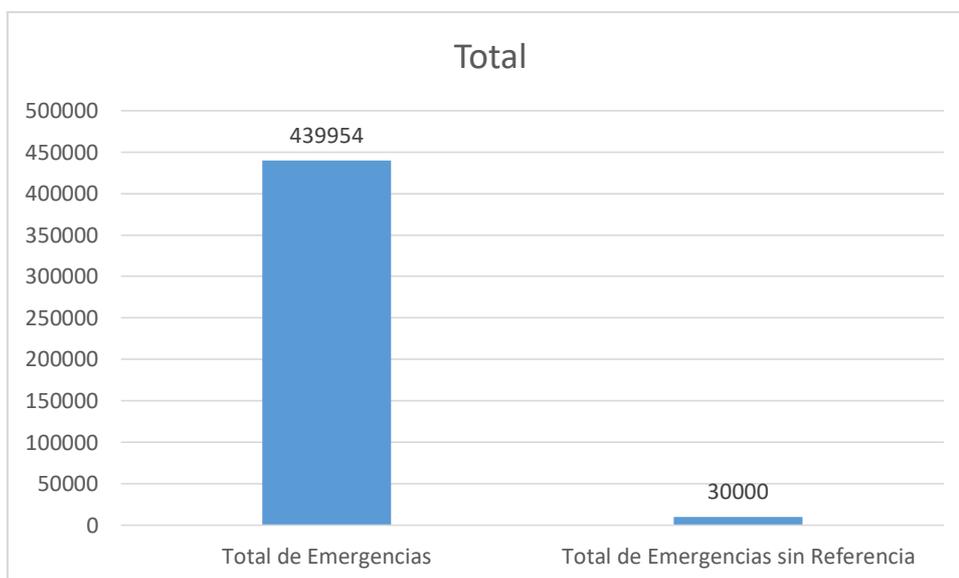


Figura 11: Total de Emergencias sin posición geográfica

Fuente: Elaboración Propia (2016)

Después de realizada la depuración de los datos para disminuir las emergencias sin una localización geográfica, se inicia con la construcción de la base de datos stage area y del Data Warehouse. Para ello se identificó las tablas del sistema transaccional que participan dentro del modelo de inteligencia de negocios, dado que es un sistema muy grande el cual además del registro de las emergencias procesa otras actividades propias de la organización es necesaria esta segmentación de tablas para no extraer información que no se va a utilizar dentro de un modelo de análisis.

Construcción del stage area

Esta base de datos es de uso temporal a nivel de almacenamiento de datos, está constituida solamente por las tablas que se identificaron para el modelo de análisis de BI y su principal función es la de almacenar los datos de producción temporalmente para ser transformados y transportados por medio de ETL a su próximo destino que es el Data Warehouse.

Construcción del Data Warehouse

El diseño del Data Warehouse es un modelo base de datos des normalizada bajo la arquitectura expuesta por Ralph Kimball en la cual se desarrolla un conjunto de Data Mart que juntos forman un concepto de Data Warehouse, este paradigma se escogió debido a que bomberos actualmente quiso conformar un modelo de BI para uno de sus procesos en específico, el cual es la atención de emergencias, el utilizar el paradigma de Kimball cumplió con este requerimiento de forma perfecta,

abarcando solamente lo que al negocio le interesa y dejando el modelo abierto para futuros desarrollos de nuevos Data Mart, para el diseño del esquema de las tablas de la base de datos se utilizó un modelo estrella dado que este ofrece un mayor rendimiento a la hora de procesar la información, lo cual significa que no existen tablas con un nivel jerárquico, solamente la relación que tiene cada dimensión con la tabla de hechos.

La tabla de hechos almacena la información de las emergencias atendidas por el BCBCR y las diversas tablas de dimensión corresponden a los distintos puntos de vista por los que se quiere analizar la tabla de hechos.

Base de datos garbage (ODS)

Es una base de datos que se creó con el fin de almacenar aquellos datos que no cumplen con reglas de negocio o no fue posible de corregir en procesos de limpieza antes de la extracción.

La información almacenada que corresponde a datos que no son aptos para analizar por medio de técnicas OLAP o de minería de datos se le creó un reporte estático el cual permite visualizar dicha información donde el negocio decide qué hacer con ella.

Extracción de datos

La extracción de datos se da por medio de ETL, y está desarrollada en dos procesos de extracción, en la primera etapa se extrae la información de la base de datos del sistema transaccional a una base de datos stage area. Con la diferencia que esta base de datos no contiene llaves primarias, relaciones foráneas, restricciones o cualquier otra validación de integridad, el quitar las reglas de integridad permite garantizar que la tarea de extracción cause el menor impacto posible en el sistema transaccional disminuyendo los tiempos de duración en extraer la información, este proceso se realiza de forma automática por medio de tareas de flujo de datos programada con integration services disponible en la solución de Microsoft SQL Server 2014. Por realizado este proceso, se inicia con la segunda parte, la carga y transformación de los datos del stage area al Data Warehouse en la que solamente se cargan los datos necesarios para el análisis de información referente a la atención de emergencias, este proceso se realiza por medio de Transac- SQL el cual se ejecuta de forma automática por medio de tareas de Integration Services.

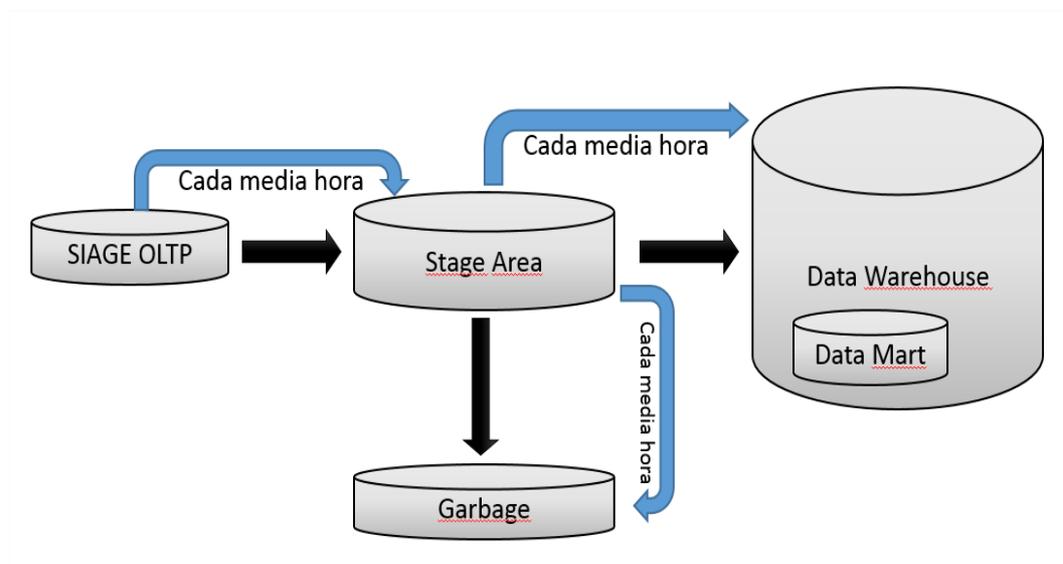


Figura 12: Modelo de la arquitectura de BI

Fuente: Elaboración Propia (2016)

Creación del cubo

El desarrollo del cubo se creó utilizando un proyecto de Analysis Services proporcionado dentro del paquete de Microsoft SQL Server 2014. Para ello se crea un origen de datos el cual se usa para conectar con el Data Warehouse, este origen de datos permite que se configuren las tablas a utilizar en la vista de datos, también se puede interpretar como una representación abstracta de las tablas del Data Warehouse y se procede a crear el cubo en el cual se define quien es la tabla de hechos y sus respectivas dimensiones para obtener un modelo estrella como el siguiente:

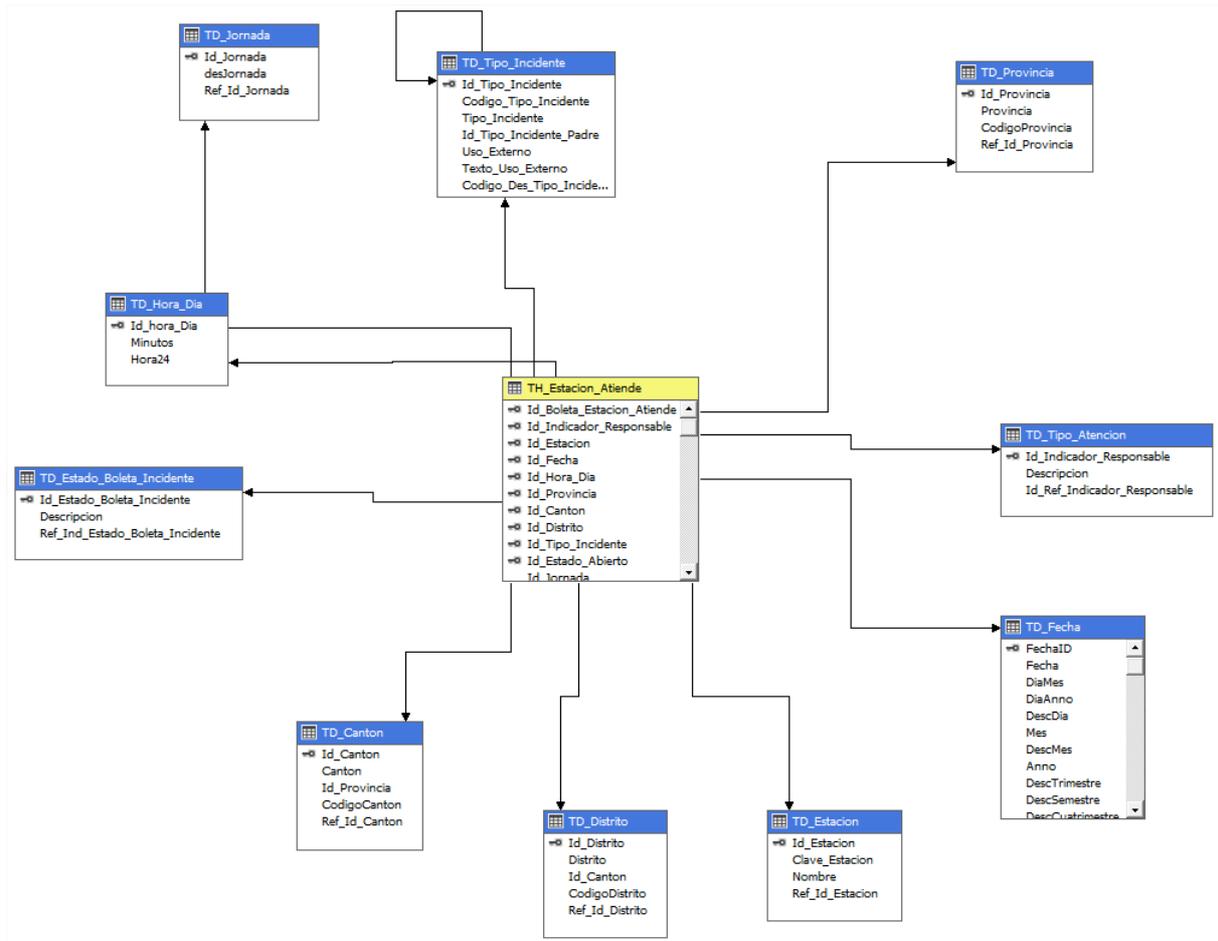


Figura 13 Modelo del Cubo de BI

Fuente: Elaboración Propia (2016)

Implementación de Minería de datos

Selección del algoritmo

Para poder determinar que algoritmo utilizar en el modelo de minería de datos, se estudiaron basado en sus características y rendimiento tres algoritmos en específico, el algoritmo de Árboles de Decisión de Microsoft, algoritmo de Bayes Naive de Microsoft y el algoritmo de Clústeres de Microsoft

Para ello luego de trabajar con los tres algoritmos y observar diferentes resultados, se utilizó una funcionalidad que permite analizar los tres algoritmos e

identificar cual es el más cercano a la realidad en términos de predicción para los datos que se estudian en cuanto a la atención de emergencias del BCBCR.

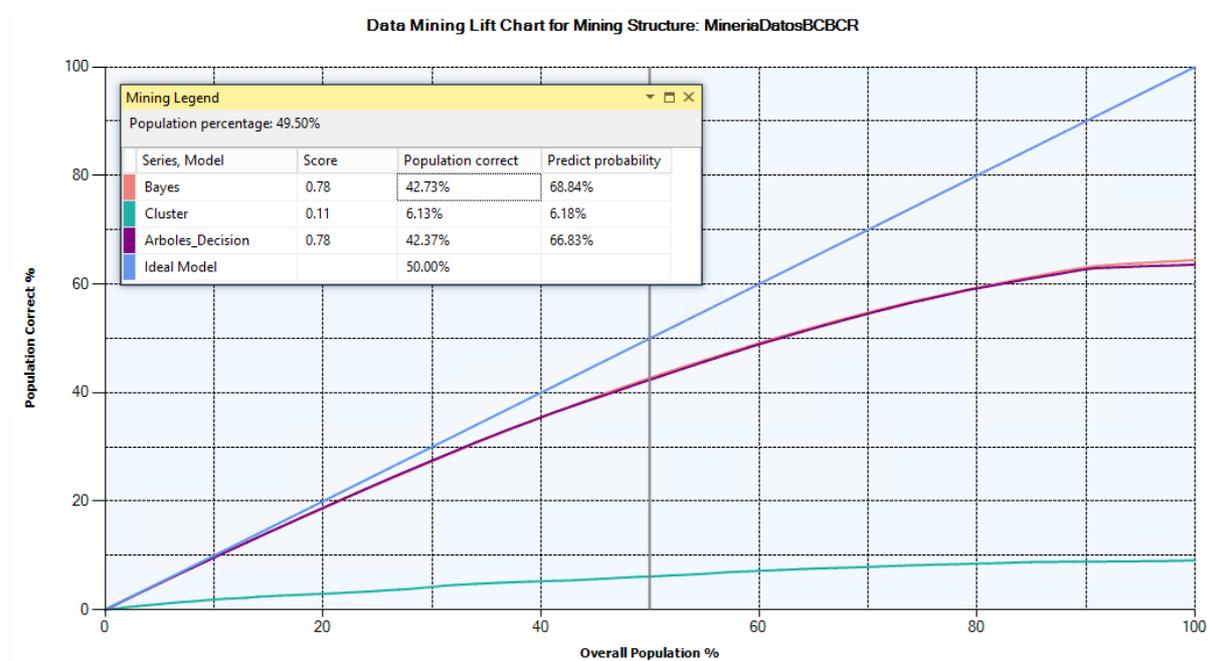


Figura 14 Grafico de comparación de algoritmos de minería de datos

Fuente: Elaboración Propia (2016)

En la figura se muestra un gráfico que ubica dos algoritmos con mismas probabilidades de precisión en el caso de árboles de decisión y Bayes, dejando por fuera a clústeres.

A partir del análisis anterior, se implementó el algoritmo de Bayes por las características que el mismo presenta para el análisis de información en la atención de emergencias. Para ello se llevó a cabo un proceso de preparación de datos los cuales son entrenados para que el algoritmo funcione de la mejor forma, esto se realizó creando dos vistas de base de datos, las vistas de base de datos son semejantes a una tabla de base de datos pero virtual, no tiene una estructura física.

La primera vista reúne los campos que contienen la información histórica referente a zonas geográficas, tipo de incidente, y estaciones involucradas en cada emergencia registrada en el Data Warehouse, la segunda vista contiene información que reúne emergencias por zona geográfica y tipo de incidente, pero sin la estación, el trabajo del algoritmo es predecir este campo para poder establecer prioridades de atención por estación de bomberos entre otros criterios para evaluar.

El algoritmo trabaja de la siguiente manera:

1. Se indica cual es la tabla que contiene los casos de estudio, en otras palabras, de donde la minería de datos va a tomar como base para poder predecir algo.
2. Se indica cual es la llave de la tabla, los campos de entrada y valor a predecir
3. Se define el tipo de contenido de cada campo de entrada o valor a predecir entre valor continuo o discreto. Continuo cuando tengo muchas posibilidades indefinidas a trabajar y discretas cuando estos valores son definidos.
4. Se establece el porcentaje de datos para prueba, se recomienda utilizar el 30% este valor ya se establece por defecto en la configuración del algoritmo a raíz de diferentes pruebas y experiencia en la que han dictado que un 30% es un valor alto para poder predecir eventos a futuro.
5. Se establece un nombre para el modelo.
6. Se visualiza la información.

Es importante mencionar que el diseño de la solución del algoritmo no está estrictamente ligado a un único algoritmo, si en un futuro las características de los datos cambian o el volumen de información aumenta, el modelo se adapta a la implementación de cualquiera de los otros algoritmos que la herramienta de analysis service de Microsoft ofrece, con solo agregar el algoritmo al modelo y realizar la publicación, queda habilitado para el usuario final e iniciar con el análisis respectivo de la información.

Visualizador de BI

OLAP

Cuando se desarrolla un modelo de inteligencia de negocios, la parte fundamental del proyecto por parte del interesado ósea el usuario final es la visualización y manipulación de la información, la cual le permite convertirla en conocimiento para la toma de decisiones de la institución. En pro de proporcionar una herramienta que se adapta a las necesidades actuales del BCBCR, la organización adquirió licencias de usuario para utilizar Power BI, herramienta de la marca Microsoft especialidad en soluciones de inteligencia de negocios.

Para ello se realizó la descarga de su versión escritorio que es gratuita desde la cual por medio de una conexión predefinida se conecta a los cubos de BI y queda a disposición del usuario para crear los respectivos reportes, mismos que pueden ser publicados en la nube para ser vistos desde una interfaz Web o en sus versiones móviles (Android, Apple, Windows Phone). Para utilizar en tiempo real los reportes en la web o el dispositivo móvil se creó un Gateway (puerta de enlace) virtual, el cual permite que el sistema web se comuniquen por medio de un servidor público, con el servidor de base de datos y así evitar una conexión directa al repositorio y poder cumplir con las políticas de la institución en normar de seguridad.

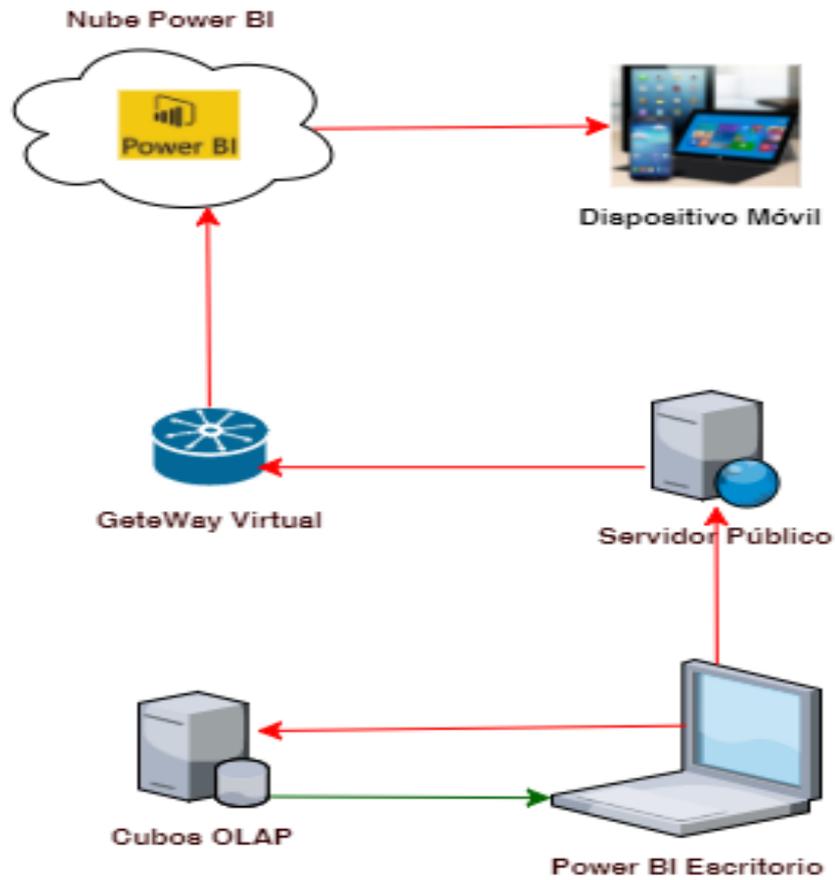


Figura 15: Modelo de Visualización de Datos
Fuente: Elaboración Propia (2016)

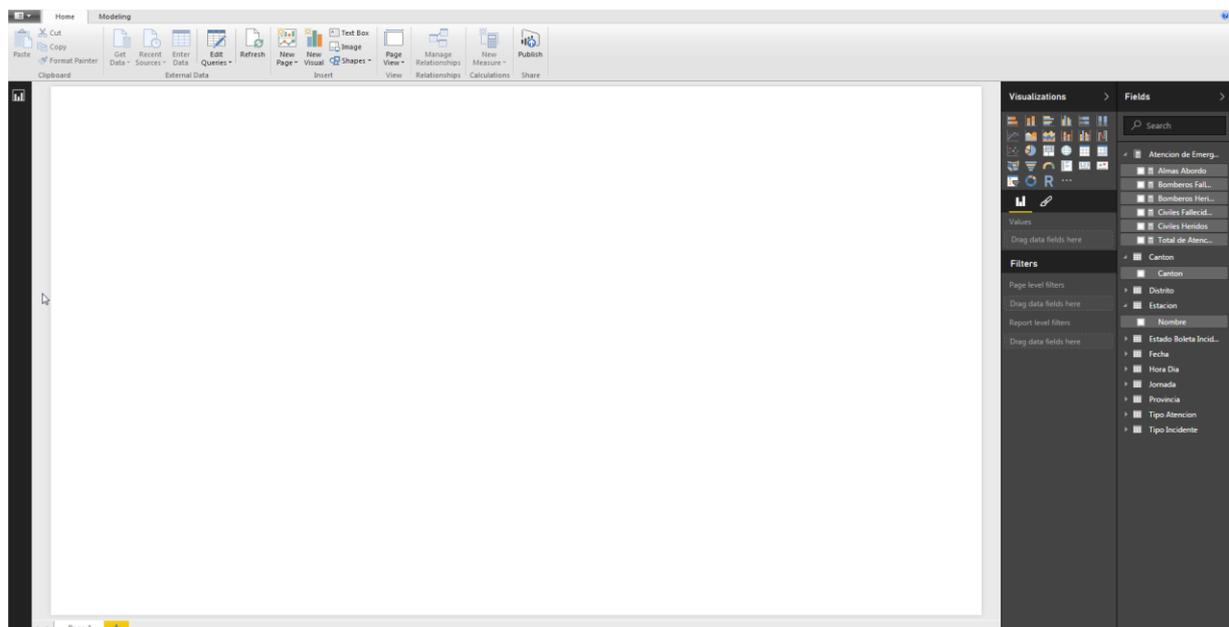


Figura 16: Herramienta para desarrollar informes POWER BI Microsoft.
Fuente: Elaboración Propia (2016)

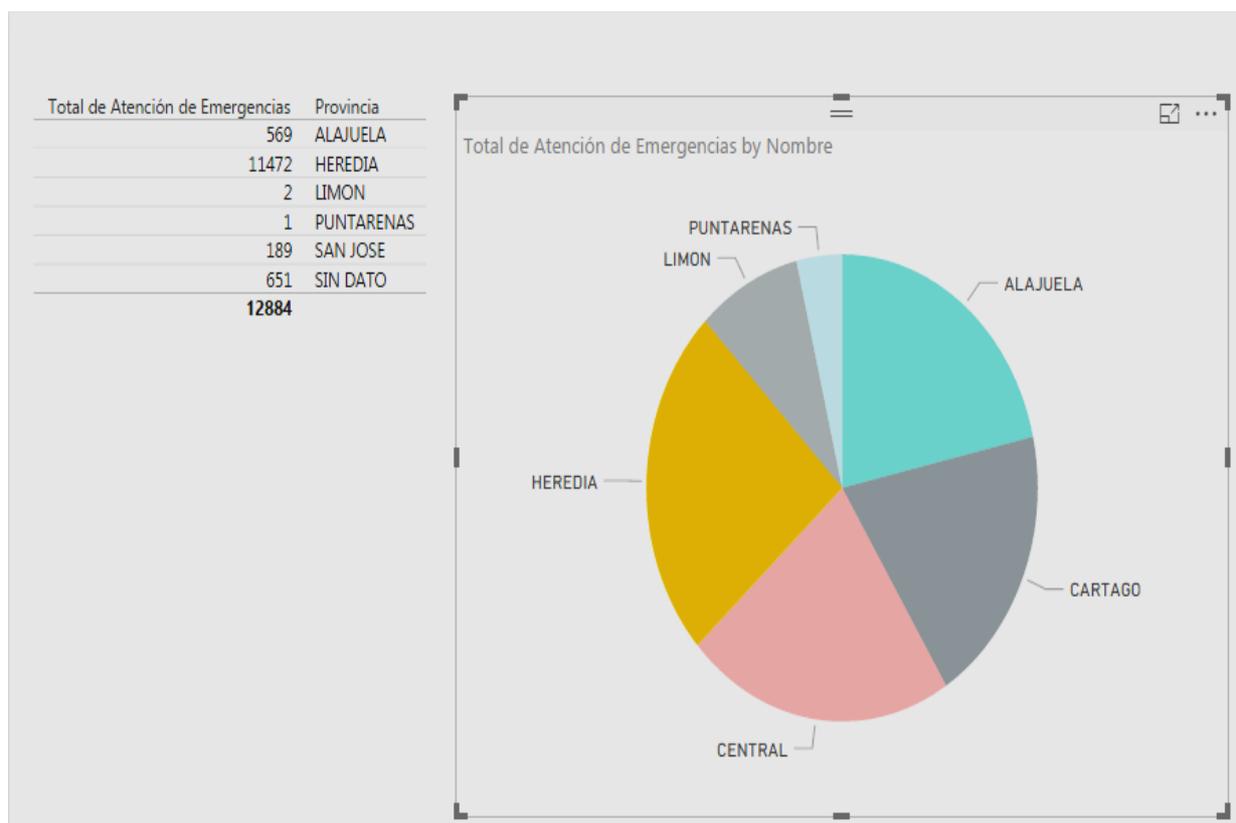


Figura 17: Ejemplo de informe con POWER BI de Microsoft.

Fuente: Elaboración Propia (2016)

En la figura 16, se muestra el resultado de las emergencias atendidas por determinadas estaciones en el país y en qué provincia fueron en auxilio de la emergencia.

Por ejemplo en el caso de Heredia como estación atendió 569 incidentes en la provincia de Alajuela tan y como se muestra al lado izquierdo de la figura 16.

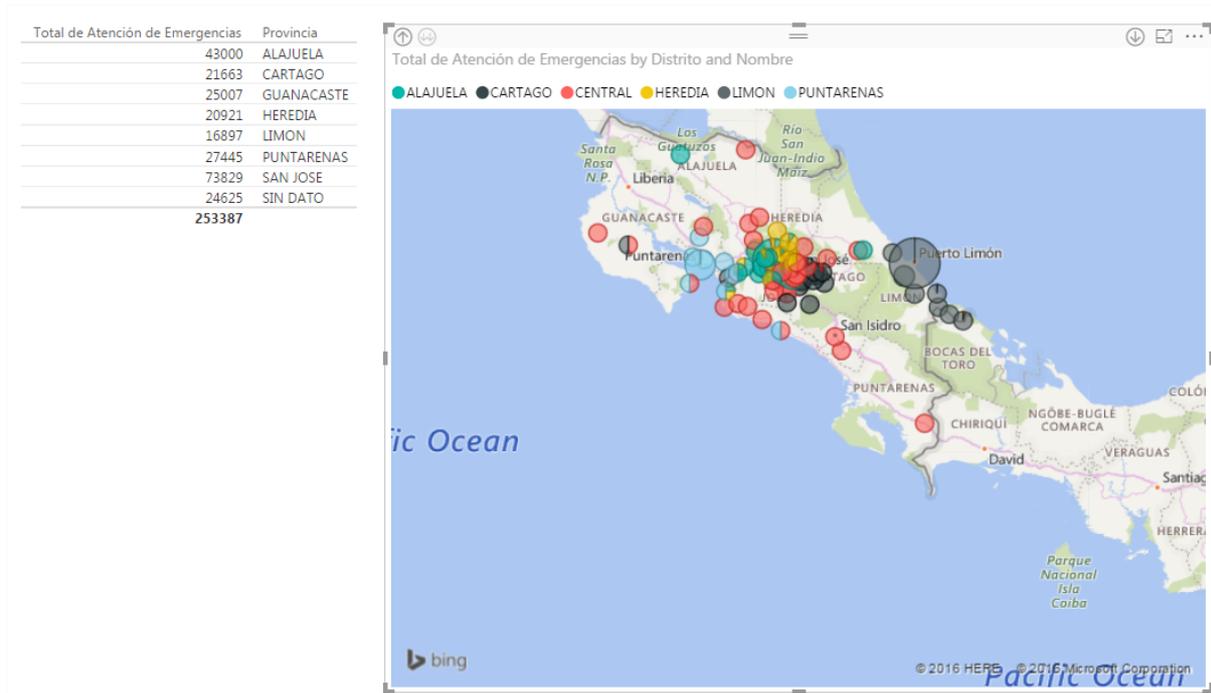


Figura 18: Ejemplo de informe geográfico con POWER BI de Microsoft.

Fuente: Elaboración Propia (2016)

En la figura 17 se observa un gráfico similar al anterior pero con la posibilidad de reflejar la información en un mapa de Costa Rica, permitiendo obtener panoramas aún más realista y acordes a zonas de afectación ante una eventual emergencia.

Minería de Datos

Para la visualización de minería de datos se utilizó un complemento que se incorpora dentro de Excel herramienta proporcionada dentro del paquete de Office, el complemento se llama Data Mining Client for Excel se puede descargar desde la página oficial de Microsoft. Esta herramienta permite interactuar con el modelo de minería de datos desarrollado con las herramientas de Microsoft, de una forma más natural y flexible para el usuario final. La escogencia del mismo se dio a partir de las pocas herramientas

existentes para hacer uso de la minería de datos y su flexibilidad de integración con Excel.

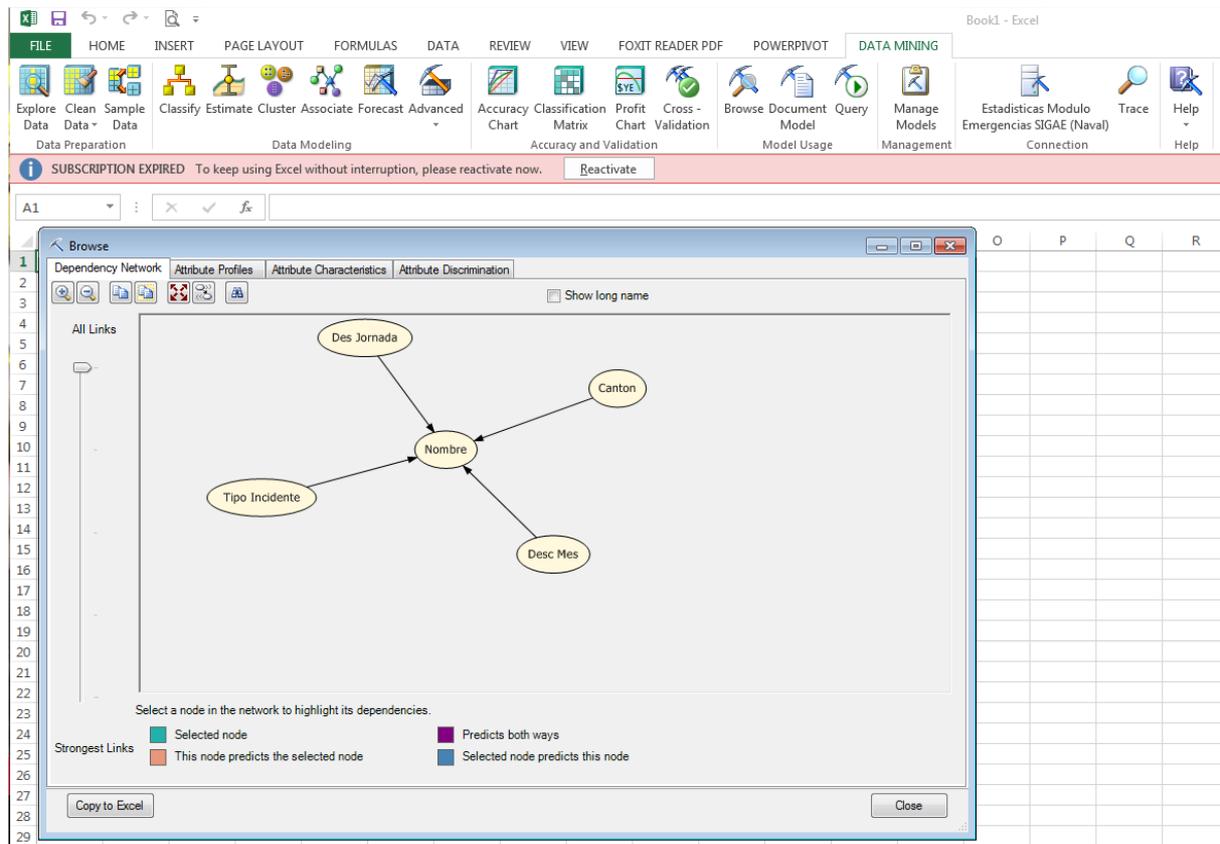


Figura 19: Ejemplo de uso del algoritmo de minería de datos con Excel.

Fuente: Elaboración Propia (2016)

En la figura 18 se observa como el complete que se incorpora a Excel interactúa con el modelo de minería de datos que se desarrolló utilizando SQL Server Microsoft. En la imagen el valor a predecir es la estación que atendería una futura eventualidad basado en cuatro criterios Jornada, Mes, Cantón, Tipo Incidente.

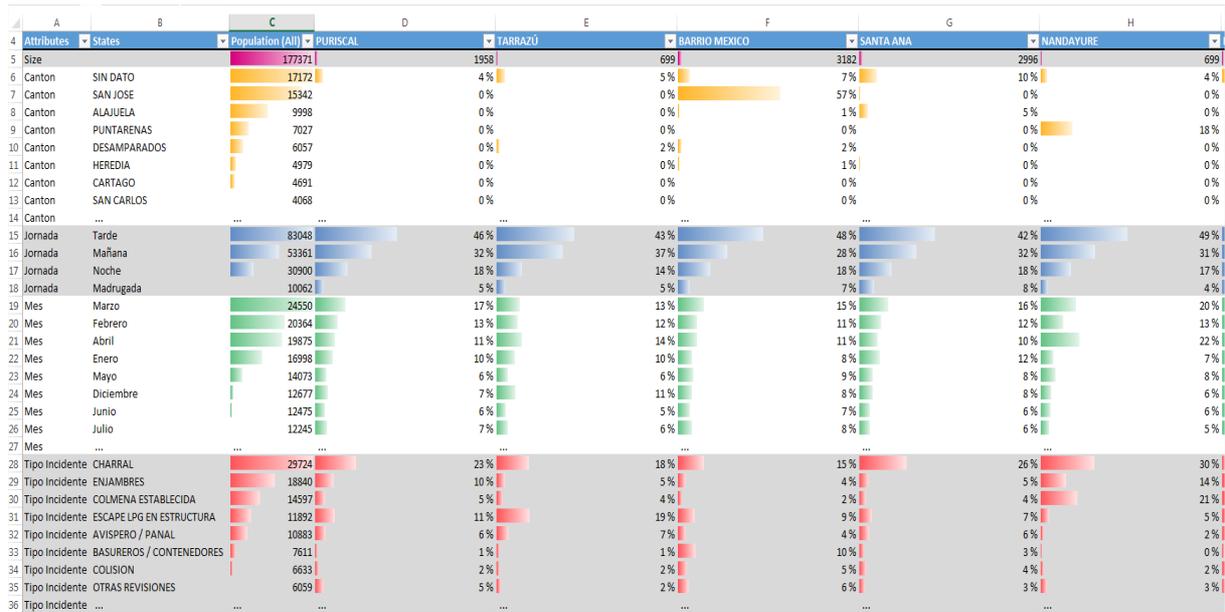


Figura 20: Ejemplo de uso del algoritmo de minería de datos con Excel.

Fuente: Elaboración Propia (2016)

En la figura 19 se observa los porcentajes de acierto que encontró para cada caso que se analizó, esta información se genera de forma automática y lista para la interpretación y análisis del usuario.

Capítulo 6. Conclusiones y Recomendaciones

Conclusiones

En la actualidad Bomberos de Costa Rica y su principal trabajo que es la atención de emergencias van siempre de la mano con la tecnología con el fin de ofrecer un servicio de calidad a la población costarricense. Para fortalecer la calidad del servicio se trabajó en la creación de un modelo estratégico de análisis de información Como lo es la inteligencia de negocios, la cual permitió generar conocimiento al negocio que se deriva de la correcta utilización de la información generada dentro y fuera de la institución. Con este conocimiento adquirido Bomberos de Costa Rica pudo hacer uso de sus recursos de forma más precisa y equilibrada para garantizar la mejor atención a la ciudadanía.

Para bomberos de Costa Rica es de vital importancia el poder analizar su información de diferentes formas y combinaciones, clasificando la información de acuerdo a los criterios en análisis, BI en sus siglas en inglés, por medio de un visualizador permitió tal flexibilidad para cumplir con la necesidad del negocio en ese apartado. Consiguiendo realizar análisis de información por categorías como posiciones geográficas, tipo de atención, estaciones entre otros, incluso realizar combinaciones entre ellos. Esto permitió que el BCBCR, establezca planes de mejora que refuercen aquellas estaciones donde exista una mayor incidencia de un tipo específico de emergencias.

El poder realizar comparativas en diferentes líneas del tiempo utilizando como criterio de análisis el tipo de emergencia atendida es un trabajo del día a día en el BCBCR, el cual antes de aplicar BI requería de un procedimiento complejo, exhaustivo, el cual tomaba muchos días y organización para llegar a una conclusión aproximada a la realidad. Con BI este trabajo se reduce a segundos, es exacto y se realiza con el esfuerzo de una sola persona. Permitted comparar el trabajo de atención de emergencias por parte de bomberos entre diferentes años consiguiendo así crear cuadros comparativos para reflejar el esfuerzo de trabajo entre un año y otro. Y dar a conocer a la ciudadanía esta información con el fin de crear conciencia para reducir los índices de atención.

La predicción de datos es un tema poco conocido en el país en general, este tipo de propuestas son muy interesantes y se puede utilizar en lugares donde tengan al menos cinco años de historia. Bomberos no es la excepción a esta regla y con la implementación del modelo de BI se incorporó minería de datos, herramienta aplicada para la predicción en cuanto a probabilidades de atención de emergencias por estación. Bomberos logro establecer planes de respuesta en cuanto a que estación debe atender una emergencia con relación a la zona donde se produce un incidente, acortando los tiempos de reacción por parte de los despachadores de unidades.

Antes de la implementación del modelo de inteligencia de negocios, el realizar análisis sobre comportamiento de los datos era una tarea imposible de hacer puesto que el proceso solo estaba dirigido a realizar estadísticas de forma manual, con este proyecto se abren muchas posibilidades en cuanto al tema de análisis, permitiendo realizar trazos y dar seguimiento a un patrón que permita estudiar dicho comportamiento obtenido resultados que permitan por ejemplo evidenciar porque un día se tiene altos valores de atención de incidentes por fuego y al siguiente menos.

La presentación de los datos obtenidos en el BCBCR no ha sido una tarea sencilla durante el tiempo en que se ha requerido presentar la misma a quien lo requiera, se han utilizado métodos engorrosos donde el informe suele pasar por varias personas que lo manipulan y sin un formato estandarizado, dando origen a varias revisiones para garantizar su autenticidad. Con el proyecto se implementa un visualizador Power BI de Microsoft, este software permitió que la tarea de crear informes sea flexible amigable y en únicos formatos de presentación, e incluso permitiendo compartirla con la persona que lo requiera con unos cuantos clic.

El crear un único almacén de datos donde se puede integrar la información proveniente de los sistemas de atención de emergencias que Bomberos de Costa Rica genera día con día se convirtió en una solución innovadora que ha permitido una mejora en la calidad del servicio que presta bomberos a la ciudadanía costarricense ya que con ella temas de operaciones estratégicas con la toma de decisiones y utilización de los recursos además de fortalecer puntos como la calidad de datos se han podido fortalecer al tener un único recurso de consulta, con ello se

eliminó problemas de duplicidad de información y de funciones en cuanto a quien consulta información y de donde se obtiene.

Para el BCBCR no era suficiente poder analizar y conocer sus datos solo con un almacén de datos donde se centraliza la información, para ello implemento y utilizo una de las formas más usadas para poder analizar su información, herramientas OLAP (o base de datos multidimensionales), este tipo de solución analizo la información por niveles de jerarquía permitiendo así por ejemplo obtener el número de emergencias atendidas por la estación de bomberos de Alajuela en la provincia de San Jose, entre los meses de marzo y abril de los años 2014 y 2015. Análisis que tiene un tiempo de respuesta muy alto y con agilidad ante grandes niveles de información.

Para BCBCR el poder tener un almacén de datos y herramientas de análisis fueron las bases para la construcción del modelo de inteligencia de negocios pero nada de eso es funcional para un bombero si la información no se puede visualizar en un entorno amigable, fluido y simplificado para el análisis de dichos datos, es aquí donde implementar Power BI se convirtió en la imagen de entrada para la solución de BI de cara al usuario, dicha herramienta se conectó a la base de datos multidimensional y expuso los datos de una forma simple para su uso con facilidades como la publicación online para que indiferentemente de la posición geográfica de la persona interesada en la información pueda consultarla por medio de una conexión de internet en su dispositivo móvil. Esto permitió ahorrar costes, tiempo y esfuerzos en la monitorización de procesos, la gestión del desempeño y el diseño/modificación de estrategias

Recomendaciones

Por haber finalizado la implementación del diseño del modelo de inteligencia de negocios en el BCBCR, se recomienda como proyecto a futuro implementar un proceso de calidad de datos que permita contener información precisa y acorde a la realidad, iniciando por el sistema transaccional el cual es la primera entrada de datos revaluando las reglas de negocio, formatos de campo y valores requeridos, con la información ya contenida se puede realizar escaneos con herramientas ya existentes de patrones en los datos que permitan evidenciar si existe redundancia de información o valores nulos que no deberían serlo. Realizando luego de estos pasos los cambios que sean necesarios y con ello aumente los grados de exactitud en el análisis y predicción de datos utilizando el modelo implementado en este proyecto.

Capítulo 7 Reflexiones Finales

Hoy en día las organizaciones manejan un flujo de información el cual era inimaginable apenas unas décadas atrás. Gracias a que nuestro mundo es ahora mucho más instrumentado, la capacidad de recolectar datos es más sencilla. Basta con observar la información tan valiosa que recolecta un supermercado por medio de las tarjetas de cliente frecuente o la forma en que los call center recopilan información para entender el porqué de las quejas y el abandono de sus clientes.

Es aquí donde entra la inteligencia de negocios, que se refiere directamente a la práctica y al conjunto de herramientas que pueden ayudar a las empresas a adquirir un mejor entendimiento de ellas mismas, logrando explotar su información con la intención de manipular dicho recurso de una forma más sencilla.

Las empresas tienden a considerar que este tipo de soluciones solo están para implementarse en grandes organizaciones porque tienen mucho que ganar y suficiente presupuesto para lograr el desarrollo, pero no es así, las medianas empresas se pueden beneficiar de igual manera sin necesariamente gastar mucho dinero y tiempo dado que cuentan con objetivos más focalizados y el costo de implementación suele ser mucho menor.

Inteligencia de negocios resulta ser sumamente importante para cualquier tipo de compañía. Estas herramientas aseguran que las decisiones tomadas sean siempre las mejores.

Capítulo 8 Trabajos a Futuro

En el Benemérito Cuerpo de Bomberos de Costa Rica, la información se ha convertido en una fuerte herramienta para aumentar la calidad del servicio que se brinda al pueblo costarricense, con ello el contar con este modelo de inteligencia de negocios ayuda a la tarea de analizar información para la toma de decisiones se ha convertido en un proyecto sólido el cual no termina aquí.

EL BCBCR tiene ya en su presupuesto, un porcentaje destinado en la conformación de un equipo de investigación que se especialice en el uso de esta herramienta creada para el análisis y predicción de información. Este equipo tendrá la tarea de utilizar de forma ágil dicho modelo y crear estrategias con el fin de utilizar el recurso humano y los activos de la organización de la mejor forma posible para disminuir los tiempos de atención en las emergencias que día con día se da en el país.

Glosario

A

Algoritmo: Conjunto ordenado de operaciones sistemáticas que permite hacer un cálculo y hallar la solución de un tipo de problemas.

Analysis Services: es un motor de datos analíticos en línea que se usa en la ayuda para la toma de decisiones y en análisis empresarial.

Análisis Cualitativo: está orientado a revelar cuáles son las características de alguna cosa. De este modo, lo cualitativo se centra en la calidad.

Análisis Cuantitativo: presenta información sobre una cierta cantidad.

C

Calidad de datos: se refiere a los procesos, técnicas, algoritmos y operaciones encaminados a mejorar la calidad de los datos existentes en empresas.

D

Data Mart: es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.

Data Mining o Minería de Datos: es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática,

con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto.

Dato: El dato es una representación simbólica de un atributo o variable cuantitativa.

Data Warehouse: En el contexto de la informática, un almacén de datos es una colección de datos orientada a un determinado ámbito, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad.

Desnormalizado: Es el proceso que procura optimizar el desempeño de una base de datos por medio de la agregación de datos redundantes.

E

Etnometodología: Es una corriente sociológica surgida en los años sesenta a través de los trabajos de Harold Garfinkel.

ETL: Abreviatura de Extract, Transform and Load, es el proceso que permite a las organizaciones mover datos desde múltiples fuentes, reformatearlos y limpiarlos, y cargarlos en otra base de datos, data mart, o data warehouse para analizar.

H

Hardware: Conjunto de elementos físicos o materiales que constituyen una computadora o un sistema informático.

G

Google Académico: es un buscador de Google enfocado en el mundo académico que se especializa en literatura científica-académica.

I

Información: Es un conjunto organizado de datos procesados, que constituyen un mensaje.

Inteligencia de Negocios o BI: es la habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

Integration Services: es una plataforma para compilar soluciones de transformaciones de datos e integración de datos de nivel empresarial.

IEEE: Abreviatura de Institute of Electrical and Electronics Engineers, es la mayor organización profesional técnica del mundo dedicada al avance de la tecnología en beneficio de la humanidad.

L

Licitaciones: procedimiento administrativo para la adquisición de suministros, realización de servicios o ejecución de obras que celebren los entes, organismos y entidades que forman parte del Sector Público.

O

OLAP: Es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). Es una solución utilizada en el campo de la llamada Inteligencia de negocios (o Business Intelligence) cuyo objetivo es agilizar la consulta de grandes cantidades de datos.

OLTP: Es la sigla en inglés de Procesamiento de Transacciones En Línea (OnLine Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación y procesamiento de transacciones (gestor transaccional).

S

Software: Conjunto de programas y rutinas que permiten a la computadora realizar determinadas tareas.

Stage Area: se puede traducir como área de pruebas o área de ensayo, es un área intermedia de almacenamiento de datos utilizada para el procesamiento de los mismos durante procesos de extracción, transformación y carga (ETL).

String: Es una cadena de caracteres, palabras o frases.

Función Split: Extrae subcadenas de una cadena original por medio de un delimitador.

Referencias

Bomberos de Costa Rica. (s.f.). Obtenido de <http://www.bomberos.go.cr/historia/>

Conocimiento. (s.f.). Obtenido de Definición.De: <http://definicion.de/conocimiento/>

Data. (s.f.). Obtenido de BusinessDictionary:
<http://www.businessdictionary.com/definition/data.html>

Geiger, J. G. (s.f.). *Data Quality Management, The Most Critical Initiative You Can Implement*. Obtenido de <http://www2.sas.com/proceedings/sugi29/098-29.pdf>

Harold , G. (1967). *Google books*. Obtenido de Google books:
https://books.google.co.cr/books?id=3vVUb0LyvrsC&pg=PR6&lpg=PR6&dq=Harold+Garfinkel+%22Studies+in+Ethnomethodology%22+editorial&source=bl&ots=l8CFnGrJoC&sig=oARvVP3wNAid382NdIK26UYHxyM&hl=es-419&sa=X&redir_esc=y#v=onepage&q=Harold%20Garfinkel%20%22Studies

Información. (s.f.). Obtenido de Definición.De: <http://definicion.de/informacion/>

Microsoft. (s.f.). *Microsoft*. Obtenido de [https://msdn.microsoft.com/es-es/library/ms175595\(v=sql.120\).aspx](https://msdn.microsoft.com/es-es/library/ms175595(v=sql.120).aspx)

Microsoft. (s.f.). *Microsoft System Center*. Obtenido de [https://technet.microsoft.com/es-es/library/hh916543\(v=sc.12\).aspx](https://technet.microsoft.com/es-es/library/hh916543(v=sc.12).aspx)

pdcahome. (s.f.). *Diagrama de Ishikawa: Análisis causa-efecto de los problemas.*

Obtenido de PDCA Home: <http://www.pdcahome.com/diagrama-de-ishikawa-2/>

Sinnexus. (s.f.). *Sinnexus Business Intelligence Informática estratégica.* Obtenido de

http://www.sinnexus.com/business_intelligence/datawarehouse.aspx

Sinnexus. (s.f.). *Sinnexus Business Intelligence Informática estratégica.* Obtenido de

http://www.sinnexus.com/business_intelligence/