



San Marcos

MIEMBRO DE LA RED
ILUMINO

SISTEMAS DE APOYO PARA LA TOMA DE DECISIONES



SISTEMAS DE APOYO PARA LA TOMA DE DECISIONES

ANÁLISIS MULTIDIMENSIONAL OLAP

OLAP es el acrónimo en inglés de procesamiento analítico en línea (*On-Line Analytical Processing*). Es una solución utilizada en el campo de la Inteligencia de Negocios (*Business Intelligence*), la cual consiste en consultas a estructuras multidimensionales,

LOS CUBOS OLAP SON ESTRUCTURAS MULTIDIMENSIONALES (CUBOS) QUE PERMITEN ANALIZAR BASES DE DATOS RELACIONALES DE GRAN VOLUMEN Y VARIEDAD CON UNA GRAN AGILIDAD Y RAPIDEZ, REDUCIENDO ENORMEMENTE EL TIEMPO Y LOS RECURSOS EMPLEADOS EN EL ANÁLISIS.

frecuentemente se denomina cubos OLAP y que le permite al usuario tener una visión más rápida e interactiva de los mismos. Se usa en informes de negocios de ventas, *marketing*, informes de dirección, minería de datos y áreas similares.

El análisis del hipercubo organiza la información según los parámetros que se

consulten, de manera tal que a partir de estructuras multidimensionales que contienen los datos resumidos de Sistemas Transaccionales, conocidos como OLTP (*Online Transactional Processing*) o de grandes bases, se obtendrá la información requerida.

Es muy utilizado en el área de *marketing*, ventas, informes, entre otros, especialmente, debido a que las respuestas a consultas complejas se obtienen muy rápidamente y además porque puede obtener los datos tanto de una fuente externa como de una base interna.

La razón de usar OLAP para las consultas es la velocidad de respuesta. Una base de datos relacional almacena entidades en tablas discretas si han sido normalizadas. Esta estructura es buena en un sistema OLTP pero para las complejas consultas multitabla es relativamente lenta.



Un modelo mejor para búsquedas, aunque peor desde el punto de vista operativo, es una base de datos multidimensional. La principal característica que potencia a OLAP es que es lo más rápido a la hora de hacer *selects*, en contraposición con OLTP que es la mejor opción para **INSERTS, UPDATES y DELETES**.

Etapas de OLAP:



Flujograma 1. Etapas OLAP

- **Recopilar:** es el cúmulo de información en un computador para calcularla según diferentes categorías. Por ejemplo, se recoge toda la información sobre las ventas, para calcular una previsión de las tendencias en ventas.
- **Profundizar:** permite a los usuarios navegar en la información accediendo a todo tipo de detalles. Por ejemplo, un usuario puede conocer las especificaciones de cada producto.
- **Especificar:** en esta fase se analiza la información en función de categorías determinadas y, se extrae una estrategia que permita alcanzar el objetivo fijado.
- **Evaluar:** se trata de evaluar esta estrategia bajo las diferentes opciones que ofrece la herramienta. Es decir, la estrategia anteriormente diseñada en ventas y sus efectos en el departamento de marketing o de producción.

La base de datos OLAP se personaliza y se configura según un esquema multidimensional, teniendo en cuenta complejas y actualizadas preocupaciones analíticas, en tiempo real.

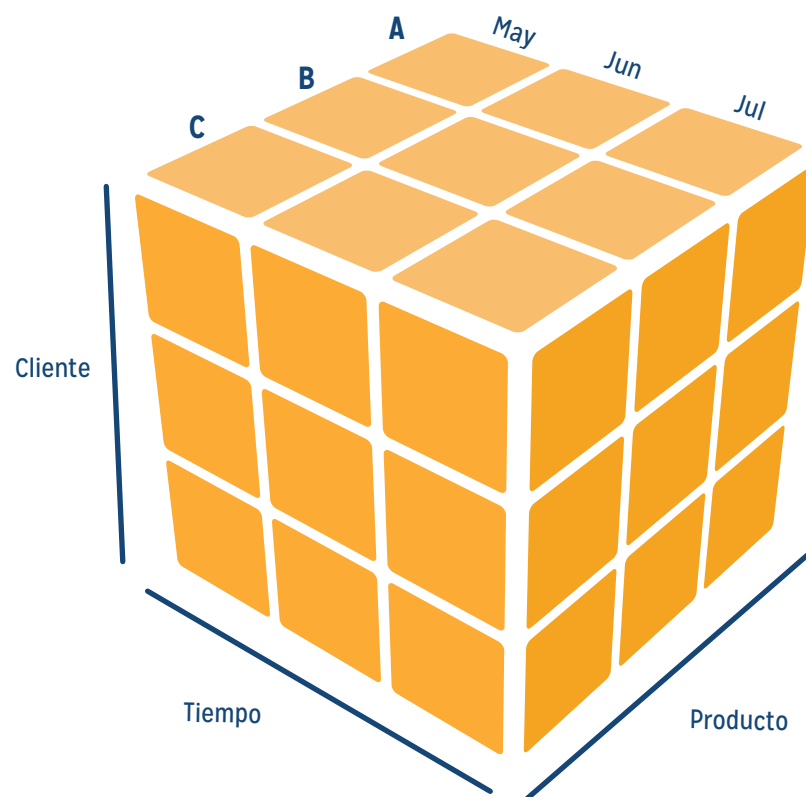


Figura 1. Ejemplo de un cubo OLAP

Existen algunas clasificaciones entre las implementaciones OLAP:

ROLAP es una arquitectura que accede a los datos almacenados en un *Datawarehouse* para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales. La principal ventaja de esa arquitectura es que permite el análisis de una enorme cantidad de datos.

La arquitectura ROLAP es capaz de usar datos precalculados si estos están disponibles, o de generar de manera dinámica los resultados desde los datos elementales si es preciso. Esta arquitectura accede directamente a los datos del *Datawarehouse* y soporta técnicas de optimización de accesos para acelerar las consultas.

MOLAP es una implementación OLAP que almacena los datos en una base de datos multidimensional. Para optimizar los tiempos de respuesta, el resumen de la información es usualmente calculado por adelantado. Estos valores precalculados o agregaciones son la base de las ganancias de desempeño de este sistema. Algunos sistemas utilizan técnicas de compresión de datos para disminuir el espacio de almacenamiento en disco debido a los valores precalculados.

HOLAP (*Hybrid OLAP*) almacena algunos datos en un motor relacional y otros en una base de datos multidimensional.

Un desarrollo un poco más reciente ha sido la solución OLAP híbrida (HOLAP), la cual combina las arquitecturas ROLAP y MOLAP para brindar una solución con las mejores características de ambas: desempeño superior y gran escalabilidad. Un tipo de HOLAP mantiene los registros de detalle (los volúmenes más grandes) en la base de datos relacional, mientras que mantiene las agregaciones en un almacén MOLAP separado.

DOLAP es un OLAP orientado a equipos de escritorio (*Desktop OLAP*). Trae toda la información que necesita analizar desde la base de datos relacional y la guarda en el escritorio. Desde ese momento, todas las consultas y análisis son hechas contra los datos guardados en el escritorio.



DATAMINING

Se define como *Datamining* a la extracción de información oculta y predecible de grandes bases de datos, es una poderosa tecnología nueva con gran potencial para ayudar a las compañías a concentrarse en la información más importante de sus Bases de Información (*Datawarehouse*).

Las herramientas de *Datamining* predicen futuras tendencias y comportamientos, permitiendo en los negocios tomar decisiones proactivas y conducidas por un conocimiento acabado de la información (*knowledge-driven*).

Las herramientas de *Datamining* pueden responder a preguntas de negocios que, tradicionalmente, consumen demasiado tiempo para poder ser resueltas y a los cuales los usuarios de esta información casi no están dispuestos a aceptar. Estas herramientas exploran las bases de datos en busca de patrones ocultos, encontrando información predecible que un experto no puede llegar a encontrar porque se encuentra fuera de sus expectativas.

Muchas compañías ya colectan y refinan cantidades masivas de datos. Las técnicas de *Datamining* pueden ser implementadas, rápidamente, en plataformas ya existentes de *software* y *hardware* para acrecentar el valor de las fuentes de información existentes y pueden ser integradas con nuevos productos y sistemas pues son traídas en línea (*on-line*).

LOS FUNDAMENTOS DEL DATAMINING

Las técnicas de *Datamining* son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. *Datamining* toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva.





Datamining ya se puede aplicar en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras:

- **Recolección masiva de datos**
- **Potentes computadoras con multiprocesadores**
- **Algoritmos de Datamining**

EL ALCANCE DE *DATAMINING*

Dadas bases de datos de suficiente tamaño y calidad, la tecnología de *Datamining* puede generar nuevas oportunidades de negocios al proveer las siguientes capacidades:

- **Predicción automatizada de tendencias y comportamientos**

Datamining automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que, tradicionalmente, requerían un intenso análisis manual, ahora pueden ser contestadas directa y rápidamente desde los datos, lo cual permite anunciar pronósticos de problemas financieros futuros y otras formas de incumplimiento, e identificar segmentos de población que probablemente respondan de forma similar a eventos dados.

LAS HERRAMIENTAS DE *DATAMINING* BARREN LAS BASES DE DATOS E IDENTIFICAN MODELOS PREVIAMENTE ESCONDIDOS EN UN SOLO PASO.

- **Descubrimiento automatizado de modelos previamente desconocidos**

Las herramientas de *Datamining* barren las bases de datos e identifican modelos previamente escondidos en un solo paso. Otros problemas de descubrimiento de modelos incluyen detectar transacciones fraudulentas de tarjetas de crédito e identificar datos anormales que pueden representar errores de tipeado en la carga de datos.

UNA ARQUITECTURA PARA DATAMINING

Para aplicar mejor estas técnicas avanzadas, deben estar totalmente integradas con el *Datawarehouse* así como con herramientas flexibles e interactivas para el análisis de negocios.

Actualmente, varias herramientas de *Datamining* operan fuera del *Warehouse*, requiriendo pasos extra para extraer, importar y analizar los datos. Además, cuando nuevos conceptos requieren implementación operacional, la integración con el *Warehouse* simplifica la aplicación de los resultados desde *Datamining*. El *Datawarehouse* analítico resultante puede ser aplicado para mejorar procesos de negocios en toda la organización, en áreas tales como manejo de campañas promocionales, detección de fraudes, lanzamiento de nuevos productos, entre otros.

Un *server* multidimensional OLAP permite que un modelo de negocios más sofisticado pueda ser aplicado cuando se navega por el *Datawarehouse*. Las estructuras multidimensionales permiten que el usuario analice los datos de acuerdo con la forma como quiera mirar el negocio resumido por línea de producto u otras perspectivas claves para su negocio. El *server* de *Datamining* debe estar integrado con el *Datawarehouse* y el *server* OLAP para insertar el análisis de negocios directamente en esta infraestructura.

DATAMINING VS. ESTADÍSTICA

Los conceptos de *Datamining* y estadística tienen el mismo objetivo: mejorar la toma de decisiones mediante un conocimiento del entorno. Este entorno lo facilitan los datos almacenados en la compañía, cuantitativos o cualitativos y mediante información de terceras empresas.





Datamining aventaja a la estadística en los siguientes aspectos:

Las técnicas estadísticas se centran, generalmente, en técnicas confirmatorias, mientras que las técnicas de *Datamining* son por lo general exploratorias. Sin embargo, cuando el objetivo es meramente exploratorio (para concretar un problema o definir cuáles son las variables más interesantes en un sistema de información) surge la

ALGORITMO: SE TRATA DE UNA SERIE DE INSTRUCCIONES O REGLAS ESTABLECIDAS QUE, POR MEDIO DE UNA SUCESIÓN DE PASOS, PERMITEN ARRIBAR A UN RESULTADO O SOLUCIÓN.

necesidad de delegar parte del conocimiento analítico de la empresa en técnicas de aprendizaje (inteligencia artificial), utilizando *Datamining*.

Utilizar técnicas de *Datamining* como árboles de decisión nos permitirá encontrar relaciones inéditas para luego concretar la investigación sobre las variables más interesantes.

Las técnicas de *Datamining* son menos restrictivas que las estadísticas. Una vez encontrado un punto de partida interesante puede suceder que los datos no satisfagan los requerimientos del análisis estadístico. Entonces, las variables deberán ser examinadas para determinar qué tratamiento permite adecuarlas al análisis, no siendo posible o conveniente en todos los casos.

Datamining o estadística son perspectivas que constituyen una sinergia y que no son excluyentes una de otra. En este sentido, la metodología de un proyecto de *Datamining* ha de contener referencias a la estadística en dos partes destacables del proceso: preparación de los datos (tratamiento de valores erróneos, valores omitidos,...) y aproximación a las variables de estudio, despliegue del proyecto y posible generación de hipótesis a refutar con una metodología y técnica estadística.

Así pues, *Datamining* y estadística son técnicas complementarias que permiten obtener conocimiento inédito en los almacenes de datos o dar respuestas a cuestiones concretas de negocio.

COMPORTAMIENTO DE LOS CLIENTES

La misión de la tecnología es ayudar y hacerlo, especialmente, en aquellas funciones que resultan tediosas y largas para la mente humana y una pérdida de tiempo absoluta para un profesional en la empresa.

El cliente es cada vez un consumidor más exigente y mejor informado, el cual busca rapidez, un servicio o un producto impecable, a buen precio y que, para ser fiel a sus proveedores y sus marcas, exige todos esos requisitos de una forma constante y creciente. En el desarrollo de la empresa, eso se traduce en una necesidad de mantenerse avanzando en el mercado intentando anticiparse lo más posible a los que serán los gustos y las necesidades de ese cliente.

Utilizar *Datamining* (minería de datos) permite descubrir comportamientos y predecir acciones de compra en la información de los clientes. Las más avanzadas técnicas *Datamining* utilizan las últimas técnicas de IA (inteligencia artificial) como las redes neuronales, que permiten encontrar gustos comunes o clientes similares en miles o millones de datos, estas herramientas lo que hacen es emular parte del proceso mental humano, de manera que se pueden construir arquitecturas capaces de reconocer patrones y aprender de ellos.

El sistema puede señalar alarmas cuando los clientes inician comportamientos de un abandono de la organización, para aprender de la experiencia y evitar la pérdida de otros similares. Pero poder predecir los que podrían marcharse en el futuro más cercano, eso no tiene precio, para evitar esa fuga se debe crear un plan de fidelización rápido, adecuado y contundente. Estas son solo algunas de las posibilidades de utilizar correctamente técnicas de *Datamining*, pero no las únicas. Antes de perder a un cliente, la inteligencia artificial pone al servicio de las empresas una serie completa de herramientas de decisión que permiten múltiples medidas para fidelizar a la clientela y mejorar la eficacia de su *marketing*.

La inteligencia artificial y el *Datamining* ponen al alcance de las empresas averiguar qué vender a sus clientes y cómo, para tenerlos satisfechos, además de apoyar decisiones estratégicas en datos científicos sobre el comportamiento de la clientela, las posibilidades de éxito se multiplican por cien.



BÚSQUEDA Y RECUPERACIÓN DE INFORMACIÓN EN LA WEB: ¿QUÉ HA PASADO Y QUÉ PODEMOS ESPERAR EN EL FUTURO?

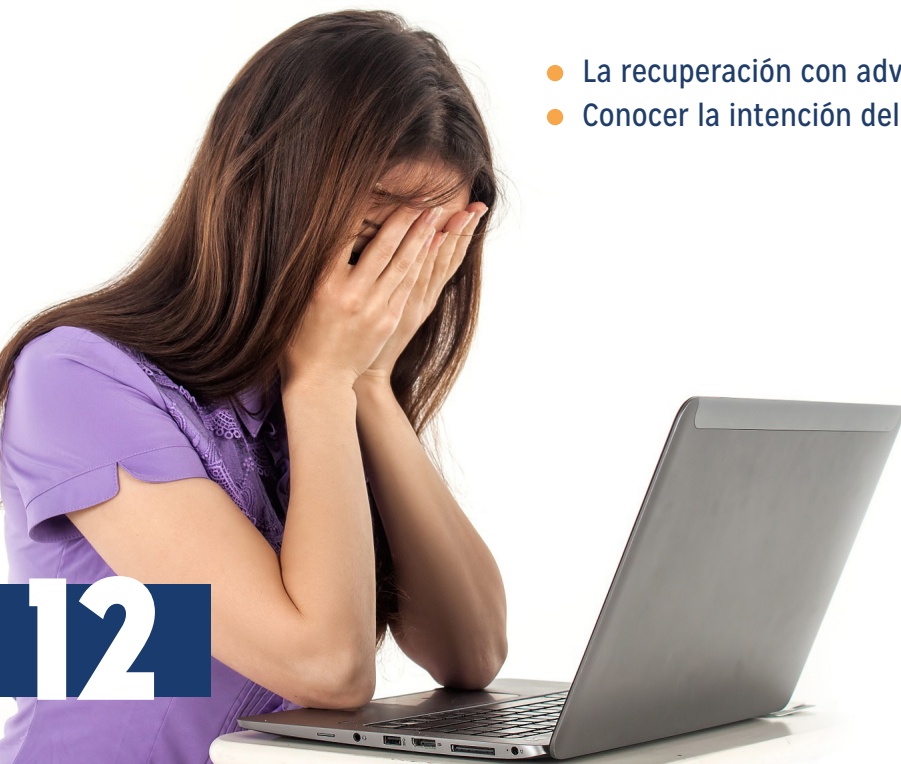
La recuperación de información ha experimentado más avances e innovaciones gracias a la web. El motivo es que hasta entonces, la recuperación de información (RI a partir de ahora) había estado confinada en entornos aislados y controlados a la vez.

Todo cambió a finales de los 90 con la idea de los fundadores de Google, Sergey Brin y Larry Page, de aplicar el análisis de enlaces para ordenar los resultados. La idea en síntesis era la siguiente: el filtrado de la información no tenía demasiado problema, el problema era la ordenación de los resultados. Había ya tanta información en la web en aquella época que encontrar miles o decenas de miles de páginas web sobre cualquier tema era fácil. El problema era el algoritmo de ordenación, basado en los cálculos estadísticos de la recuperación de información clásica, muy influenciados a su vez por las frecuencias de los términos presentes en cada documento, aunque luego esas frecuencias se ponderasen con frecuencias globales del conjunto de índice, entre otros.

Brin y Page dedujeron que si eran capaces de construir una especie de base de datos donde pudieran saber cuántos enlaces recibe cada página, entonces sería fácil determinar qué páginas eran las más valiosas para cada pregunta. Con un importante matiz posterior (no todos los enlaces valen lo mismo, sino que los enlaces de páginas a su vez muy enlazadas son más valiosos), esta forma de determinar la importancia de una web y después de influenciar su posición en la ordenación de la página de resultados recibió el nombre del ahora tan conocido, discutido e imitado PageRank.

La recuperación de información en la web tiene dos desafíos fundamentales:

- La recuperación con adversario
- Conocer la intención del usuario cuando lanza una búsqueda



LA RECUPERACIÓN CON ADVERSARIO

Esto se parece cada vez más a una carrera de armamentos donde a cada innovación de los buscadores para evitar el *spam*, sigue alguna técnica para burlar esa innovación. Por ejemplo, en su momento, cuando quedó clara la importancia de los enlaces de entrada en el posicionamiento de una página para expulsar del índice a las páginas de mala calidad, muchos autores sin escrúpulos se dedicaron a montar auténticas granjas de enlaces para autoenlazarse.

La clasificación de texto con adversario se puede definir como aquella tarea de clasificación de texto en la que existe un adversario cuyo objetivo es que el clasificador producido, ya sea manual o automáticamente con aprendizaje, deje de ser útil.

El elemento fundamental de la clasificación de texto con adversario es la presencia del adversario que pretende que el clasificador deje de ser útil. Las dos maneras más obvias para lograr que un clasificador deje de ser útil es lograr que cometa un número de errores sustancial, que pueden ser por defecto (por ejemplo, en el caso de un filtro de *spam*, que no detecte un número importante de mensajes de correo basura, es decir, que incurra en muchos falsos negativos) o por exceso (por ejemplo, que clasifique muchos mensajes legítimos como basura, es decir, que incurra en muchos falsos positivos).

CONOCER LA INTENCIÓN DEL USUARIO CUANDO LANZA UNA BÚSQUEDA

Dado que un buscador como Google es en realidad una interfaz unificada de acceso al repositorio de información multimedia más grande y diversa que jamás soñó la humanidad, resulta clave poder interpretar las intenciones del usuario. Pero esta intención solo puede determinarse por medios indirectos. No es pensable someter a cada usuario a un cuestionario previo antes de permitirle lanzar una búsqueda. Acertar en mayor medida en esta predicción (si el usuario en realidad busca noticias de actualidad, o vídeos, o sitios web, o un lugar donde realizar una transacción, o una definición, o una entrada de diccionario) puede marcar la diferencia entre que el usuario considere que la búsqueda ha sido acertada o más bien un fracaso.





Por el mismo motivo, en los últimos años la página de resultados también ha ido evolucionando hacia lo que se denomina la "búsqueda universal". Esto significa que actualmente, la mayoría de los motores de búsqueda combinan en la misma página de resultados diversas tipologías y morfologías de la información, generalmente imágenes, videos y noticias (además de páginas web).

INTRANET: RED INFORMÁTICA QUE UTILIZA LA TECNOLOGÍA DEL PROTOCOLO DE INTERNET PARA COMPARTIR INFORMACIÓN, SISTEMAS OPERATIVOS O SERVICIOS DE COMPUTACIÓN DENTRO DE UNA ORGANIZACIÓN.

Incluso el mismo concepto de buscador ha evolucionado, y junto a los buscadores tradicionales (o generalistas) se desarrollan nuevos tipos de buscadores. Especialmente interesantes son los casos de los metabuscadores y los buscadores especializados.

En los últimos años, se desarrolló el proyecto de la web semántica, el objetivo consiste en conseguir una web cuyos contenidos sean mejor interpretados por los ordenadores y, entre otras cosas, los buscadores

funcionen de forma "inteligente". Ahora bien, para ello sería necesario que la información esté marcada y descrita con metadatos siguiendo un formalismo muy bien determinado. Esto ya es bastante difícil de conseguir dada la naturaleza abierta y heterogénea, carente de todo control central que caracteriza a la web.

Para contar con una web como la descrita, sería necesario contar con procesos que permitan interpretar, comparar y fusionar los datos, los metadatos y las informaciones procedentes de diferentes recursos.

Por último, sería necesaria la existencia de herramientas capaces de relacionar y hacer conclusiones con estos contenidos para satisfacer las necesidades de información de los usuarios, que equivaldrían a los "buscadores de la web semántica", y que serían los denominados "agentes inteligentes de búsqueda".

Este proyecto de la web semántica ha motivado el desarrollo completo de diversos estándares (como XML, RDF, SKOS, OWL o Dublin Core, entre otros) que pueden contribuir, de manera decisiva, a la mejora de los sistemas de búsqueda de la web, especialmente si son aplicados en entornos más controlados como son los repositorios, las bibliotecas digitales o las intranets.

BIBLIOGRAFÍA

- Bages, B. (2014). Herramienta OLAP. Recuperado de <http://belindabages.com/herramienta-olap/>[Consulta 14 de enero de 2016].
- Carne, C. P. (2007). Datamining. Recuperado el de <http://www.monografias.com/trabajos/datamining/datamining.shtml/> [Consulta 11 de enero de 2016].
- Fontao, J. (2010). Datamining vs. Estadística. Recuperado de <http://crmamedida.blogspot.com/2010/05/data-mining-vs-estadistica.html/> [Consulta 11 de enero de 2016].
- Lluís Codina, R. P.-J. (2011). Búsqueda y recuperación de información en la web: qué ha pasado y qué podemos esperar en el futuro. Recuperado de <http://bid.ub.edu/27/pedraza2.htm/>[Consulta 11 de enero de 2016].
- OLAP, A. M. (2012). Teban. Recuperado de <http://analisismultidimensional6.blogspot.com/> [Consulta 11 de enero de 2016].
- Sinnexus. (2007-2016). Persistencia MOLAP, ROLAP, HOLAP. Recuperado de http://www.sinnexus.com/business_intelligence/olap_avanzado.aspx/ [Consulta 15 de enero de 2016].
- Viñals, J. J. (2011). Recuperado de http://www.factorhuma.org/index.php?option=com_content&view=article&id=5692%3A%C2%BFSe+puede+llegar+a+predecir+el+comportamiento+de+los+clientes%3F&catid=4%3Anoticias&Itemid=11&lang=es/ [Consulta 11 de enero de 2016].

