



FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

TRABAJO DE FIN DE GRADO

GRADO EN ECONOMÍA

CURSO ACADÉMICO 2018-2019

**BUSINESS INTELLIGENCE PARA LA TOMA DE DECISIONES
ESTRÁTEGICAS: UN CASO DE APLICACIÓN DE MINERÍA DE DATOS
DENTRO DEL SECTOR BANCARIO**

**BUSINESS INTELLIGENCE FOR STRATEGIC DECISION-MAKING: A CASE
OF APPLICATION OF DATA MINING WITHIN THE BANKING SECTOR**

AUTOR: Vicente López Güemes

Directora: Eliana Rocío Rocha Blanco

Junio 2019

RESUMEN

En un contexto como el actual, donde las organizaciones generan enormes cantidades de datos, aprender a gestionar, procesar y analizar estos datos resulta de vital importancia para toda aquella organización que pretenda ser competitiva. Lograr extraer información que después pueda convertirse en conocimiento, proporciona una mejor toma de decisiones a las organizaciones, lo que sin duda las ayudará en la consecución de sus objetivos. Las organizaciones hacen uso del *Business Intelligence* para lograr extraer ese conocimiento que las permita ser más competitivas en sus respectivos sectores. A lo largo de este proyecto, se analiza este concepto de *Business Intelligence* intentando indagar en los diferentes componentes, técnicas y metodologías que lo rodean. Con ello, se pretende poner de manifiesto la importancia que el Business Intelligence tiene para las organizaciones en el entorno de gran competencia empresarial en el que nos encontramos.

Por otro lado, se aborda un caso práctico donde el *Business Intelligence* puede resultar beneficioso para una organización. Mediante la aplicación de minería de datos y haciendo uso de la herramienta Weka, se pretende extraer un conocimiento que resulte útil para la toma de decisiones de una entidad bancaria.

SUMMARY

In a context like the current one, where organizations generate huge amounts of data, learning how to manage, process and analyze this data is of vital importance for any organization that wants to be competitive. Extracting information that can later become knowledge provides better decision making to organizations, which will undoubtedly help them to achieve their objectives. Organizations use Business Intelligence to extract that knowledge that allows them to be more competitive in their respective sectors. Throughout this project, this concept of Business Intelligence is analyzed trying to investigate the different components, techniques and methodologies that surround it. With this, it is intended to highlight the importance that Business Intelligence has for organizations in the environment of great business competition in which we find ourselves.

In addition, a practical case is analyzed where Business Intelligence can be beneficial for an organization. Through the application of data mining and using the Weka tool, we pretend to extract knowledge that is useful for the decision making of a bank.

ÍNDICE DE CONTENIDOS

RESUMEN.....	2
SUMMARY	2
1. INTRODUCCIÓN.....	4
1.1. JUSTIFICACIÓN Y MOTIVACIÓN.....	4
1.2. OBJETIVOS	5
2. MARCO TEÓRICO	6
2.1. DEFINICIÓN DE BUSINESS INTELLIGENCE	6
2.2. COMPONENTES DEL BUSINESS INTELLIGENCE	6
2.3. BENEFICIOS DERIVADOS DEL BUSINESS INTELLIGENCE.....	9
2.4. DISEÑO DE UNA ESTRATEGIA DE BUSINESS INTELLIGENCE.....	9
2.5. APLICACIONES DEL BUSINESS INTELLIGENCE.....	13
3. TÉCNICAS Y HERRAMIENTAS DE BUSINESS INTELLIGENCE	15
4. APLICACIÓN CASO PRÁCTICO.....	16
4.1. CONCEPTO, FASES Y TÉCNICAS DE MINERÍA DE DATOS.....	16
4.1.1. TÉCNICAS DE MINERÍA DE DATOS.....	17
4.1.2. APLICACIONES DE LA MINERÍA DE DATOS EN LAS ORGANIZACIONES	19
4.2. FUENTE DE DATOS DEL CASO PRÁCTICO.....	20
4.2. OBJETIVOS E HIPÓTESIS DEL ANÁLISIS	21
4.3. PREPARACIÓN DE LOS DATOS	21
4.4. APLICACIÓN DE ALGORITMOS DE CLASIFICACIÓN	23
4.5. APLICACIÓN DE TÉCNICAS DE CLUSTERING O AGRUPAMIENTO	25
4.6. APLICACIÓN DE TÉCNICAS DE REGLAS DE ASOCIACIÓN.....	28
5. CONCLUSIONES	30
6. BIBLIOGRAFÍA.....	32

1. INTRODUCCIÓN

En la actualidad, las organizaciones generan enormes cantidades de datos en el transcurso de sus actividades cotidianas. Todas las actividades de las compañías (compras, ventas, envíos, gestión de terceros, datos de clientes etc.) son gestionadas y almacenadas en grandes bases de datos. Estos datos almacenados contienen información de gran utilidad para las empresas. Sin embargo, las compañías deben ser capaces de gestionar, procesar y analizar las inmensas cantidades de datos de las que disponen para lograr extraer esa información valiosa que les ayude en su toma de decisiones empresariales. Precisamente este hecho, *“la conversión de los datos operativos de las empresas en información que dé soporte a la toma de decisiones, es lo que se conoce como inteligencia de negocio o Business Intelligence”* (Curto, 2010). En un contexto como el actual, donde las empresas destinan muchos recursos con el fin de mejorar su competitividad en el mercado, es de vital importancia para ellas aprender a gestionar la información de la que disponen y lograr extraer el conocimiento que se encuentra tras los datos almacenados para con ello lograr una toma de decisiones empresariales mejor, más rápida y con menor incertidumbre. Es por ello, por lo que la inteligencia de negocio o *Business Intelligence* se ha convertido en un área de estudio de gran interés para profesionales e investigadores tanto del campo de las ciencias económicas y empresariales como del sector informático y tecnológico. Como se verá a lo largo de este trabajo, son numerosos los beneficios que una estrategia de *Business Intelligence* correctamente diseñada puede proporcionar a una organización. Todas las herramientas, técnicas y metodologías que engloba el *Business Intelligence*, facilitarán el proceso de gestión de los datos y la extracción de conocimiento que permita mejorar el proceso de toma de decisiones. Hoy en día, cualquiera de las empresas punteras a nivel mundial, independientemente del sector en el que desempeñen su actividad, utiliza el *Business Intelligence* como un pilar básico dentro de su funcionamiento empresarial. Por lo tanto, todas aquellas compañías que pretendan prosperar deberán aprender a elaborar correctamente una estrategia de BI que les permita ser competitivas.

1.1. JUSTIFICACIÓN Y MOTIVACIÓN

El concepto de *Business Intelligence*, en adelante BI, que será analizado en profundidad en el presente trabajo, ha tomado gran importancia en los últimos tiempos en el panorama empresarial. Décadas atrás, las empresas tuvieron la necesidad de disponer de soportes informáticos que les permitiesen recopilar todos los datos referentes a su actividad empresarial de manera más eficiente, dejando atrás los métodos tradicionales de almacenado mediante archivos en papel y libros contables. Hoy en día, una vez logrado ese objetivo, las compañías disponen de sistemas de información que dan apoyo a todos sus procesos básicos (sistemas de información para la gestión). Sin embargo, actualmente su necesidad va más allá y se encuentra enfocada hacia otro tipo de prestaciones por parte de los sistemas de información. Los esfuerzos de las empresas se centran en lograr obtener un sistema que, además de almacenar y gestionar la información, sirva de apoyo en sus decisiones empresariales (sistemas de

información para la toma de decisiones). Es en esta cuestión donde el BI cobra real importancia. (Pérez Marqués, 2015).

En un contexto de globalización como el actual, donde la competencia entre empresas es máxima, tener la capacidad de disponer de la información en tiempo real y de manera precisa es un factor diferencial en la toma de decisiones. Aquellas empresas que logren una mejor canalización y administración de la información podrán llegar a tomar decisiones, que en algunos casos pueden llegar a ser cruciales para el devenir de la compañía, con mayor celeridad y anticipación que sus competidores. Esta importancia que el BI puede llegar a tener en el ámbito empresarial ha sido la principal motivación que me ha llevado a desarrollar el presente proyecto. Llegar a conocer en mayor profundidad algunos de los conceptos del BI desarrollados en diversas asignaturas a lo largo del Grado y llegar a tener un conocimiento más amplio sobre la aplicación del BI en las organizaciones, ya que seguramente pueda llegar a ser fundamental para el desarrollo de mi trayectoria profesional.

1.2. OBJETIVOS

A pesar de la crucial importancia que el BI puede llegar a tener para las empresas u organizaciones, sigue siendo un concepto muy abstracto para el público en general. El objetivo principal que se persigue en este trabajo es clarificar este concepto de BI mediante una descripción de los principales componentes y técnicas que lo rodean. En primer lugar, se intenta dar una definición formal y clarificadora del concepto de BI, con el objetivo de que cualquier posible lector tenga una noción básica. Se analizarán todos los elementos que componen el BI, la función que desempeña cada uno de estos elementos y el modo en que se integran entre sí para lograr un correcto funcionamiento en conjunto. Además, se detallarán las principales ventajas que una estrategia de BI puede proporcionar a las organizaciones. Por otro lado, se abordará el proceso de diseño de una estrategia de BI y las distintas fases en su implementación. Posteriormente se pretende realizar un análisis de las herramientas de BI más relevantes y algunas de sus aplicaciones prácticas más utilizadas por las organizaciones.

Tras esta primera parte teórica y como objetivo más específico dentro del trabajo, se procederá a la realización de un caso práctico que permita reflejar algunas de las ventajas y beneficios de la aplicación del BI en las compañías. En concreto, se utilizará la minería de datos para la extracción de conocimiento dentro del sector bancario. Para ello, se analizarán la minería de datos y las diferentes fases en su aplicación, así como las diferentes técnicas que la rodean. El objetivo es poder aplicarlas posteriormente en un caso práctico y lograr extraer conocimiento útil para la organización. Mediante la utilización de la herramienta Weka¹, se aplicarán técnicas propias de minería de datos sobre una base de datos basada en la campaña de marketing directo de una entidad bancaria portuguesa. El objetivo no es otro que la extracción de conocimiento que sirva de apoyo en la toma de decisiones de la entidad y la permita obtener mejores resultados.

¹ Herramienta open source para la minería de datos <https://www.cs.waikato.ac.nz/ml/weka/>

2. MARCO TEÓRICO

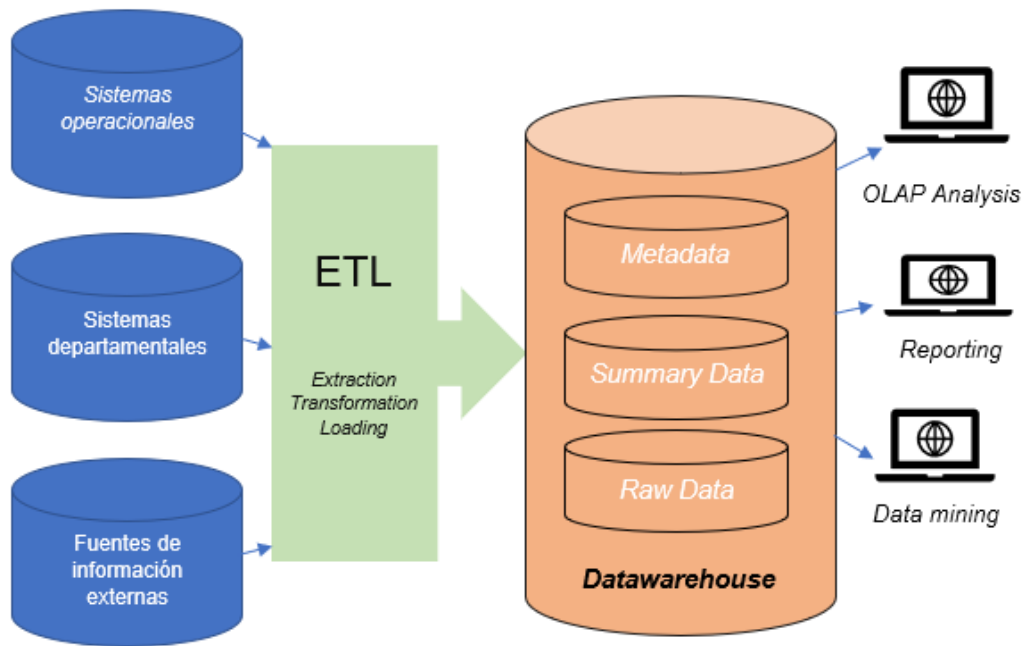
2.1. DEFINICIÓN DE BUSINESS INTELLIGENCE

Cabe destacar que la definición de BI ha ido evolucionando con el paso del tiempo al igual que la tecnología. En el año 1958, Hans Peter Luhn propuso una primera definición para el término de BI *“La habilidad de aprender las relaciones de hechos presentados de forma que guíen las acciones hacia una meta deseada”*. Posteriormente, en 1989 Howard Dresden realizó una definición más formal del concepto y lo expresó como *“Conceptos y métodos para mejorar las decisiones de negocio mediante el uso de sistemas de soporte basados en hechos”*. En la actualidad, son diversos los autores que han realizado una definición formal del concepto de BI, pero cabe destacar la dada por Josep Curto Díaz en su libro *Introducción al Business Intelligence: “Conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización”* (Curto, 2010). Esta definición formal del concepto de BI engloba perfectamente todo lo que representa el BI y sobre lo que se profundizará a lo largo del proyecto. En definitiva, el BI no hace referencia a ninguna aplicación, técnica o método de análisis por separado, sino a una combinación de todas ellas. Por tanto, es crucial entender el BI como un conjunto de elementos interconectados entre sí que ayudan a gestionar, procesar y extraer información de los datos para así lograr un conocimiento sobre el que basar la toma de decisiones empresariales.

2.2. COMPONENTES DEL BUSINESS INTELLIGENCE

Antes de que se comente todos los beneficios y ventajas que un sistema de BI proporciona a las empresas, se debe analizar cuáles son los elementos que lo componen. Cabe señalar que una correcta estrategia de BI requiere de una serie de análisis previos (relacionados con el alcance del proyecto, objetivos, costes etc.), los cuales se explicarán a lo largo del apartado 2.4, y que facilitarán la toma de decisiones a la hora del diseño de los componentes del sistema de BI. Para Cano (2007), se pueden distinguir cuatro partes o elementos que componen un sistema de BI, cada una ellas con una función propia y específica dentro del sistema. Como refleja la figura 2.1, los componentes de un sistema de BI son:

Figura 2.1. Componentes de un sistema de Business Intelligence



Fuente: Elaboración propia a partir de Cano (2007).

- **Fuentes de información.** Este será el lugar de donde se obtendrá toda la información que luego se trasladará al *Datawarehouse*. Las fuentes de información a las que se accede pueden ser de origen interno o externo (Cano, 2007).
 - De origen interno de la organización: sistemas de información departamentales (presupuestos, hojas de cálculo, previsiones etc.) y sistemas operacionales y transaccionales como los ERP, CRM, SCM, etc.
 - Fuentes de información externa: redes sociales, fuentes de datos gubernamentales o estudios de mercado encargados a terceros.
- **Procesos ETL (extracción, transformación y carga).** Mediante este proceso se obtendrán datos de las distintas fuentes de información y se trasladarán al *Datawarehouse*. Este proceso será clave dentro del proyecto de BI y será necesario destinarle gran parte del tiempo total destinado al proyecto. Dentro del proceso ETL se pueden diferenciar a su vez cinco subprocesos (Cano, 2007).
 - Extracción: se basa en la recuperación de los datos a partir de las fuentes de información. Se consigue obtener un conjunto de datos en bruto listos para ser procesados que temporalmente se alojarán en almacenes de datos intermedios (*Data staging*).
 - Limpieza: se procede a comprobar la calidad de los datos eliminando duplicados y corrigiendo posibles errores y valores vacíos. Se obtiene un conjunto de datos de mayor calidad a los obtenidos tras el proceso de extracción

- Transformación: los datos de alta calidad obtenidos tras la limpieza se estructuran y procesan en los distintos modelos de análisis. Se obtienen datos consistentes, homogéneos y útiles.
- Integración: se cargan los datos en el *Datawarehouse* y se integran en los distintos modelos y áreas de negocio que se hayan definido. Debe comprobarse que los datos cargados sean compatibles con el formato y definición dada al *Datawarehouse*.
- Actualización: Este proceso permitirá actualizar con cierta periodicidad los datos del *Datawarehouse*.
- **Datawarehouse o almacén de datos.** Se trata de una base de datos compuesta por todos los datos agregados desde las distintas fuentes de información a través de los procesos ETL. De esta manera los usuarios dispondrán de información consistente, histórica y preparada para la realización de análisis y una posterior toma de decisiones. El *Datawarehouse* o almacén de datos representa el sistema de información central de todo proyecto o sistema de BI. Es posible que la construcción del *Datawarehouse* requiera ciertos plazos de tiempo y eso genere ciertos inconvenientes a las compañías, por ese motivo se diseñaron los Data Mart. Se trata de bases de datos generalmente de menor tamaño que los *Datawarehouse* y que almacenan la información de ciertas áreas concretas dentro de la organización. Los Data Mart podrán ser dependientes o independientes en función de si obtienen la información directamente de las fuentes de información de la empresa o si la obtienen desde el *Datawarehouse* corporativo. Como explica Cano (2007), existen dos estrategias principales a la hora de construir un *Datawarehouse*:
 - La primera de ellas defendida por Inmon (2005), propone construir un *Datawarehouse* corporativo común para los distintos departamentos de la organización. A partir de él, se van diseñando distintos modelos de análisis para los diferentes departamentos de la organización. Cada departamento construiría su propio *Data Mart* dependiente de los datos del *Datawarehouse* corporativo. Esta forma de diseño del *Datawarehouse* es lo que comúnmente se conoce como método *Top-Down*.
 - En segundo lugar, se encuentra la propuesta por Kimball (2002), también conocida como metodología Bottom-up. Este método sugiere diseñar distintos *Data Marts* en función de las distintas necesidades de la organización, no ve necesario la construcción de un *Datawarehouse* corporativo. De esta manera el *Datawarehouse* estaría conformado por la unión de los Data Mart de cada uno de los departamentos o áreas de la empresa.

Si bien es cierto que ambas estrategias son válidas si se desarrollan correctamente, la más común es la segunda. Se comienza construyendo un *Data Mart* único al que luego se adhieren otros *Data Mart* que abarcar distintas áreas de negocio. Este método requiere menores inversiones que el primero y es sencillo de implementar, aunque es cierto que también puede provocar que permanezcan los silos de información en la compañía (Cano, 2007).

- **Herramientas de Business Intelligence.** Existen diversas tecnologías que permiten el análisis de la información depositada en los *Datawarehouse*. Algunas de las más destacadas serían: herramientas OLAP (On-line Analytical Processing), EIS (Executive Information System), sistemas de informes o consultas avanzadas y herramientas de minería de datos. Estas diferentes

herramientas de BI se describirán en mayor detalle en el apartado 3.1. del trabajo.

2.3. BENEFICIOS DERIVADOS DEL BUSINESS INTELLIGENCE

Un entorno de BI que apoye la toma de decisiones empresariales proporciona numerosos beneficios tanto tangibles como intangibles. Curto (2010) destaca las principales ventajas que una estrategia de BI puede proporcionar a las organizaciones.

- Creación de un flujo importante de información. Los datos son transformados en información y esta a su vez genera un conocimiento que permite una toma de decisiones mejoradas, clave para la obtención de mejores resultados.
- Elaboración y manejo de indicadores claves de rendimiento (KPI, *Key Performance Indicator*) e indicadores claves de metas (KGI, *Key Goal Indicator*) que resultan fundamentales para las empresas.
- Permite tener una visión de todos los datos de la organización mucho más amplia y de mayor calidad.
- Elimina los silos de información de las organizaciones.
- Permite diferenciar la información de la que dispone la empresa, separando la valiosa de la irrelevante. Además, permite analizar cantidades mucho mayores de información, que con métodos tradicionales sería imposible de procesar.
- Proporciona un acceso mucho más rápido a la información y una mayor celeridad en la toma de decisiones.
- Facilita una mayor coordinación entre el departamento de tecnologías de la información (TI) y el resto de la organización.

2.4. DISEÑO DE UNA ESTRATEGIA DE BUSINESS INTELLIGENCE

En algunas ocasiones, la creación de un entorno de BI puede conllevar elevados costes para las compañías. Es por ello por lo que toda empresa necesita una estrategia de BI y una justificación comercial que le permita analizar el balance entre los costes incurridos y los beneficios que se obtienen. En muchas ocasiones, los proyectos relacionados con BI terminan en abandono o fracaso y esto es debido en gran parte a una planificación inadecuada. Los proyectos de BI presentan seis etapas en su desarrollo (Moss y Atre, 2003). En la Figura 2.2 se pueden observar las distintas etapas que deben seguirse en el desarrollo de un proyecto de BI.

Figura 2.2. Etapas en el desarrollo de un proyecto de BI.



Fuente: Elaboración propia a partir de Moss y Atre (2003).

Dentro de cada una de estas etapas se pueden encontrar a su vez subetapas, las cuales será necesario realizar para lograr un desarrollo óptimo del proyecto de BI. Un análisis profundo de cada una de estas etapas requeriría de gran detalle y extensión. Sin embargo, un análisis detallado y por minorizado de estas etapas no es el objetivo del presente Trabajo de Fin de Grado y por lo tanto se realizará un análisis tan solo general de estas etapas con el objetivo de entender cuál es el proceso en el desarrollo de una estrategia de BI. A continuación, se analizarán las etapas en el desarrollo de un proyecto de BI:

1) **Justificación.** En primer lugar, se debe encontrar una justificación para el proyecto. Se procede a un análisis del negocio donde se debe detectar un problema u oportunidad de negocio e intentar proponer una solución de BI. Se deberá detectar alguna necesidad de información empresarial que no pueda satisfacerse con los métodos ya existentes. Dicha necesidad de negocio deberá tener algún tipo de vinculación financiera para la compañía (sobrecoste, pérdida de ingresos etc.) Una vez se detecta el problema comercial y se han detectado las carencias del sistema actual, se establecerán los objetivos de la aplicación de BI. La aplicación de BI que se proponga debe tener un coste justificado y se deben definir claramente los beneficios que aportara para la resolución del problema o el aprovechamiento de la oportunidad de negocio. Además, deberá realizarse un análisis de riesgos en términos de tecnología, integración, inversión financiera etc. (Mos y Atre, 2003).

2) **Planificación.** El siguiente paso consiste en planificar la implementación de esta solución de BI. Se debe tener en cuenta que las aplicaciones de BI requieren de una infraestructura empresarial para su correcto funcionamiento, por lo tanto, lo primero que se debe realizar es una evaluación de la infraestructura ya existente con el fin de detectar las posibles carencias que presente. Es que posible que existan componentes de la infraestructura empresarial que aún no existan en la empresa y que sin embargo sea necesario para el desarrollo de la aplicación de BI. En este caso, dichos componentes deberán desarrollarse como parte del proyecto de BI. Una vez se ha analizado rigurosamente la infraestructura empresarial necesaria para nuestro proyecto

se procederá a la planificación de este. Se deberán establecer unas metas u objetivos, el alcance del proyecto, requisitos del proyecto, analizar la situación en la que se encuentran los archivos y datos de origen, evaluación de los riesgos y restricciones, análisis de los costes en los que se deberá incurrir, establecer mecanismos de gestión de problemas y mecanismos de gestión de cambios. Algunas de estas cuestiones deberán analizarse muy genéricamente ya que se profundizará en ellas en otras etapas posteriores del proyecto (Moss y Atre, 2003).

3) **Análisis de negocio.** El siguiente paso en el diseño de un proyecto de BI es el análisis de negocio. Para Moss y Atre (2003) esta etapa presenta cuatro subapartados que resultan cruciales para el éxito de esta fase de análisis de negocio.

- En primer lugar, deben establecerse los **requisitos del proyecto**. Inicialmente los requisitos serán básicos e irán evolucionando hacia una mayor complejidad a medida que avance el proyecto. Algunos ejemplos de los requisitos que se establecerán en esta etapa son: la clase de información que necesitan los usuarios de nuestra empresa y que clase de consultas e informes realizarán.
- Una vez se establezcan los requisitos, se procederá a un **análisis de los datos**. Esta fase dependerá mucho de los de la calidad de los datos ya existentes. En muchas ocasiones la calidad de los datos de origen es uno de los mayores problemas que se encuentran en el desarrollo de proyectos de BI. En muchos casos se encuentra que durante años se venían realizando multitud de malos hábitos en la gestión de los datos y corregirlos requerirá de buena parte del tiempo destinado al proyecto. En otras ocasiones, se requiere del uso de datos externos los cuales no presentan el mismo formato y estructura que los internos. Por tanto, será necesario identificar estas diferencias y corregirlas.
- Con el objetivo de poder testar las diferentes ideas o conceptos que van surgiendo, se realizarán **prototipos**. Esto permitirá ver el potencial del proyecto de BI y con ello ajustar mejor los requisitos y expectativas del proyecto.
- Además, será necesario contar con un **repositorio de metadatos**, ya sea mediante la construcción de uno propio o comprándolo a un proveedor. *“Un repositorio de metadatos es una base de datos. Pero a diferencia de las bases de datos ordinarias, un repositorio de metadatos no está diseñado para almacenar datos comerciales para una aplicación empresarial. En su lugar, está diseñado para almacenar información contextual sobre los datos comerciales”* (Moss y Atre, 2003). Los metadatos son de gran importancia para un entorno de apoyo a decisiones de BI ya que ayudan a transformar los datos comerciales en información. Los metadatos permiten que la interpretación de los datos comerciales sea la correcta (Moss y Atre, 2003).

4) **Etapas de diseño.** Como ha ocurrido con la etapa anterior, dentro de esta etapa de diseño se pueden diferenciar varias subetapas según Moss y Atre (2003).

- **Diseño de la base de datos**, donde se incorporarán todos los datos comerciales de la empresa debidamente procesados. El esquema de diseño de la base de datos deberá coincidir con los requisitos de acceso a la información por parte de la compañía o empresa que hará uso de ella. *“Las bases de datos de destino de BI están diseñadas para una recuperación de datos simplificada y de alto rendimiento, no para la eficiencia del almacenamiento y mantenimiento de datos*

(que son consideraciones de diseño importantes para las bases de datos operacionales)” (Moss y Atre, 2003).

- **Extraer/Transformar/Carga (Diseño de Procesos ETL).** Este proceso o etapa seguramente sea la de mayor dificultad dentro del proyecto de BI. Las fuentes de datos originales de las aplicaciones de BI provienen de plataformas diferentes. El objetivo del proceso ETL es lograr combinar todos estos datos de las diferentes plataformas en un formato estándar, para su posterior depósito en la base de datos o *Datawarehouse* (Moss y Atre, 2003).
- **Diseño del repositorio de metadatos.** Los metadatos tienen gran importancia en cualquier entorno de BI ya que son necesarios entre otras cosas para el funcionamiento de los sistemas gestores de bases de datos (SGBD) y las demás herramientas del entorno (indican a las herramientas ETL que datos debe transformar o a las herramientas OLAP como agregar y resumir datos etc. Para (Moss y Atre ,2003) existen tres formas de diseñar un repositorio de metadatos:
 - Un repositorio centralizado de metadatos. Esta sería la solución más común y la más sencilla de implementar.
 - Repositorio de metadatos descentralizado. Esta solución almacena los metadatos en múltiples bases de datos en múltiples ubicaciones.
 - Solución distribuida de metadatos habilitados para XML. Si bien esta solución puede ser la más interesante también es la más compleja de instalar. En esta solución los metadatos permanecerían en su ubicación original que serían los diferentes diccionarios de herramientas.

5) Etapa de construcción.

- **Desarrollo del proceso ETL.** En todo proceso ETL deberán reproducirse tres tipos totales de reconciliación: conteos de dominio, conteos de registro y conteos de calidad. Con ello se conseguirá que los datos de las bases de datos con destino BI coincidan en gran medida con los datos de los sistemas de origen, puesto que esta es una de las quejas más comunes que reciben las aplicaciones de BI. La herramienta ETL estará sometida a un gran número de pruebas y testeos similares a los recibidos por los sistemas operativos. El proceso ETL será uno de los pilares básicos de la estrategia de BI y por lo tanto debe asegurarse de que la herramienta ETL cumple con todos los requisitos y propósitos exigibles (Moss y Atre, 2003).
- **Desarrollo de aplicaciones.** Tras el desarrollo de los diferentes prototipos y su correspondiente testeo para confirmar que cumplen con los requisitos funcionales, se estará en disposición de comenzar el desarrollo de la aplicación de acceso y análisis. Para Moss y Atre (2003) las actividades de desarrollo de aplicaciones que deberían seguirse serían las siguientes:
 - Determinar los requisitos finales del proyecto. Se procederá a una revisión del prototipo analizando los problemas que aparecieron durante el testeo y evaluando los cambios que se necesiten.
 - Diseñar los programas de aplicación. Se deberán diseñar los componentes de acceso y análisis de la aplicación de BI, los informes, las consultas, interfaz de usuario y la función de ayuda en línea.

- Construir y testear los programas de aplicación. Mediante la utilización de unos datos de prueba se comprobará si los programas y scripts se compilan sin error y realizan todas sus funciones correctamente. Se compararán los resultados esperados con los obtenidos en las pruebas y se analizarán posibles incidencias.
- Proporcionar acceso a datos y capacitación de análisis. Deberán identificarse las necesidades de capacitación del personal y usuarios que harán uso de la aplicación de BI. Una vez detectas estas necesidades, se programan sesiones de capacitación y se asegurara su efectividad.
- **Minería de datos (Data mining).** En muchas ocasiones, las organizaciones no sacan todo el provecho a su entorno de apoyo a decisiones de BI. Es habitual ver compañías que limitan sus aplicaciones de BI a la realización de informes o al reemplazo de los ya existentes. Mientras tanto acumulan enormes cantidades de datos sin sacarles ningún provecho. A través de herramientas de minería de datos las compañías pueden ser capaces de extraer información muy valiosa que se encuentra tras estos datos almacenados. Es posible generar modelos de análisis para encontrar patrones y pautas de comportamiento tras los datos, y posteriormente utilizar esta información para la obtención de ventajas competitivas (Moss y Atre, 2003).
- **Desarrollo del repositorio de metadatos.** En caso de que se haya decidido construir un repositorio de metadatos propio, se habrá encargado el desarrollo de este a un equipo de trabajo por separado. Por lo tanto, el desarrollo del repositorio de metadatos se convertirá en un subproyecto importante dentro del proyecto principal de BI.

6) **Etapas de Implementación.** Una vez que la aplicación de BI está correctamente construida y testeada, lo siguiente será su implementación en el entorno de producción. Las bases de datos y las aplicaciones desarrolladas se pondrán en marcha. Todo el mecanismo de soporte también comienza a funcionar, esto incluye la mesa de ayuda, monitorear el rendimiento, el ajuste y mantenimiento de las bases de datos y programar los trabajos por lotes de ETL. Por último, se realizará una evaluación de la versión de la aplicación que se ha puesto en funcionamiento. Aprender de los plazos requeridos, los costes excesivos, las disputas experimentadas y todos los ajustes realizados durante el proceso serán clave para las futuros proyecto de BI que se pongan en marcha. Toda herramienta, técnica o procedimiento que no resulte útil deberá volverse a evaluar y ajustar (Moss y Atre, 2003).

2.5. APLICACIONES DEL BUSINESS INTELLIGENCE

Después de analizar el proceso de diseño de una estrategia de BI, resulta interesante conocer algunos de los casos prácticos o situaciones en las que se suele utilizar el BI con mayor frecuencia. Méndez del Río (2006) aborda en su artículo Más Allá del Business Intelligence algunas de los casos prácticos donde el BI resulta de mayor utilidad. Como se verá, el BI resulta útil en prácticamente todas las áreas de la compañía

y su aplicación resulta beneficiosa para múltiples fines. Algunas de las aplicaciones más frecuentes del BI son:

- **Análisis del valor del cliente:** El BI permite obtener una visión mucho más global de los clientes. Conocer los datos históricos del cliente, su situación actual y el valor que potencialmente podría aportar en un futuro permite a las compañías valorar más claramente el valor real de los clientes.
- Otra área donde el BI resulta muy útil es en la **fidelización de los clientes**. El BI ayudará a predecir posibles escenarios futuros y disponer de un mayor conocimiento de los clientes, factores que sin duda facilitarán la valiosa tarea de fidelizar clientes.
- **Optimizar las campañas de marketing.** Mediante la utilización del BI se pueden automatizar los procesos de planificación y puesta en funcionamiento de las campañas de marketing. Facilita la recepción de información referente a la respuesta de los clientes a estas campañas, es decir, permite captar mejor el feedback del cliente.
- **Estimación de la demanda.** Las herramientas de BI facilitan a las compañías el análisis de los hechos pasados y la predicción de los futuros. De esta manera, las compañías son capaces de adaptarse y predecir la demanda de sus clientes, unos clientes que presentan nuevas y crecientes necesidades.
- **Mejor optimización de la cadena de suministros.** La mejora en la gestión de la cadena de suministros por parte de las compañías conducirá casi con total seguridad a la obtención de unos mejores resultados financieros. Algunas aplicaciones de BI permiten identificar y analizar todos los costes asociados a cada elemento o actividad de la cadena de suministros.
- **Gestión de costes basados en actividades.** Existen un tipo de soluciones de BI que se conocen como aplicaciones para la gestión de costes basados en actividades (ABM). Este tipo de aplicaciones de BI permite a los responsables de las compañías un mejor análisis de los costes y márgenes involucrados en las actividades de sus compañías.
- Las soluciones de BI también permiten una mejor **gestión del riesgo** por parte de las compañías. Las diversas soluciones de BI permiten la simulación de escenarios y la percepción de riesgos por parte de las compañías. Esto permite a las compañías asignar distintos niveles de riesgo a los diferentes clientes, grupos de clientes o productos. Este tipo de soluciones de BI de gestión del riesgo pasan a ser críticas en algunos sectores como puede ser el sector bancario y financiero.
- **Gestión financiera.** Las herramientas de BI se han vuelto claves en el campo de la gestión financiera. Un entorno de apoyo a decisiones permite a los responsables de los departamentos financieros tener una mayor precisión y efectividad en sus planificaciones y presupuestos. Además, facilita la tarea de alinear las operativas diarias con los objetivos a medio y largo plazo de la compañía y disponer de una mayor certeza en la predicción de cambios en el mercado y la anticipación de escenarios.

3. TÉCNICAS Y HERRAMIENTAS DE BUSINESS INTELLIGENCE

Existen diversas herramientas de BI que facilitan la toma de decisiones, el autor Cano (2007) considera que las más destacables son las siguientes:

- **EIS (*Executive Information Systems*)**. Se trata de un sistema de información y un conjunto de herramientas asociadas las cuales permiten a los usuarios acceder a la información de estado y las actividades de gestión. Esta herramienta se especializa en el análisis del estado diario de la organización a través de indicadores clave. De esta forma los usuarios pueden enterarse con mayor celeridad sobre los cambios que se produzcan. La mayor parte de la información facilitada por esta herramienta es de carácter numérico (nivel de stock, volumen de ventas etc.) y se representa gráficamente facilitando un panorama completo sobre el estado de los indicadores de negocio.
- **Herramientas OLAP (*Online Analytical Processing*)**. Se trata de herramientas mucho más genéricas que las anteriores. Permiten realizar agregaciones y combinaciones de datos mucho más complejas que facilitan realizar análisis más estratégicos. Permite a los usuarios realizar análisis con diferentes niveles de agregación y múltiples dimensiones. Un ejemplo podría ser analizar el nivel de ventas de un producto en función de los clientes, la región geográfica o un determinado periodo de tiempo. Es lo que se conoce como análisis multidimensionales, ya que permite realizar el análisis desde diferentes perspectivas o dimensiones.
- **Sistemas de informes o consultas avanzadas**. Por lo general, se basan en sistemas u objetos relacionales y los resultados suelen presentarse de forma tabular. Utilizan operadores clásicos como proyección, selección, concatenación y agrupamiento entre otros.
- **Minería de datos**. Son capaces de extraer tendencias, patrones, regularidades, descubrir relaciones que permiten una mejor descripción y comprensión de los datos. Esto permite realizar predicciones sobre comportamientos o sucesos futuros. Algunos de los usos más comunes de la minería de datos son la segmentación, sendas de consumo, clasificación, previsiones, venta cruzada, optimización etc.
- **Herramientas para la planificación y modelización**. Permite a los usuarios la realización de planes de negocio y simulaciones a partir de la información de BI. Pueden ser utilizadas para la realización de presupuestos, previsiones o planificaciones. Trasmite a los Dashboards y los Scorecards los objetivos y los umbrales de las métricas.
- **Herramientas de Dashboard y Scorecard**. Estas herramientas facilitan a los usuarios una visualización rápida de la información crítica de la compañía a través de representaciones gráficas y posibilitando una visualización detallada si fuese necesario. Los Dashboards muestran indicadores de áreas de negocio sin necesidad de que exista una relación entre ellos. Es decir, puede que los indicadores solo correspondan a parte de la organización. Por el contrario, los Scorecards son más estratégicos ya que establecen relaciones entre los indicadores y generalmente hacen referencia a la organización al completo.

4. APLICACIÓN CASO PRÁCTICO

Con el objetivo de poner en práctica parte de lo expuesto en los puntos anteriores, en este apartado se hará uso de estrategias de Minería de Datos para el análisis de datos comerciales. En concreto, se hará uso de Weka, herramienta de minería de datos, para el análisis de datos del sector bancario. Con ello, a partir de datos comerciales, se pretende extraer nuevo conocimiento que permita una mejor toma de decisiones dentro del sector bancario.

4.1. CONCEPTO, FASES Y TÉCNICAS DE MINERÍA DE DATOS

La existencia de enormes volúmenes de información y el uso de diversas herramientas informáticas han transformado por completo el análisis de datos. Hoy en día, el análisis de datos se encuentra enfocado hacia una serie de técnicas especializadas englobadas bajo el nombre de minería de datos o *Data Mining*. La minería de datos puede definirse en un primer momento como “*un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos*” (Pérez, 2007). Por tanto, el objetivo principal de las técnicas de minería de datos sería el descubrimiento del conocimiento que se encuentra tras la información almacenada de forma ordenada en bases de datos. Mediante el análisis de los datos las técnicas de minería buscan descubrir todo tipo de patrones, tendencias o perfiles que de otra manera quedarían ocultos tras la información almacenada. En el apartado 4.1.1 se abordarán las diferentes técnicas de minería de datos que se pueden llevar a cabo para la extracción de este conocimiento. Como afirma Pérez (2007), es importante entender la minería de datos como una fase fundamental dentro del proceso de extracción del conocimiento KDD (*Knowledge Discovery in Databases*). Este proceso KDD consta de varias fases o etapas las cuales se analizarán a continuación.

- En primer lugar, se lleva a cabo una fase de análisis previo para comprender las necesidades que presenta la organización y concretar el problema que presenta. En esta fase se establecen los objetivos e hipótesis a los que se pretende responder a través de la minería de datos.
- Posteriormente vendría lo que se conoce como fase de **pre-procesamiento de datos**. Dentro de esta fase de preparación de los datos se encuentra las siguientes fases:
 - **Fase de selección:** Esta primera fase resulta fundamental ya que condicionará la capacidad o no de extraer conocimiento válido y útil en las fases posteriores. En esta fase será necesario la integración de las diferentes fuentes de datos existentes, la identificación y selección de las variables o atributos más relevantes de los datos. Para la realización de estas tareas será necesario aplicar técnicas de muestreo que resulten adecuadas.
 - **Fase de limpieza:** Durante esta etapa se pretende detectar y corregir la presencia de valores atípicos (*outliers*), imputar valores faltantes (*data missing*) y se eliminan o corrigen los datos incorrectos.
 - **Transformación:** Se realiza mediante técnicas de reducción o aumento de la dimensión, técnicas de discretización y numeración o técnicas de escalado simple o multidimensional.

- Tras la fase de preparación de los datos, vendría la fase aplicación de técnicas propias de minería de datos. Esta fase se desarrolla a partir de modelos predictivos y descriptivos y mediante el análisis de los datos. Se decidirá qué tipo de conocimiento se quiere descubrir y en función de esto se escogerá la técnica de minería de datos a utilizar.
- Posteriormente se realizará una fase de evaluación. Deberán evaluarse los patrones “minados” y compararlos con los objetivos establecidos al principio del proceso. Si fuese necesarios se volvería a las fases anteriores para buscar una mejor iteración. Se volverían a analizar las fases que componen la preparación de los datos y detectar sus efectos sobre los datos obtenidos tras el algoritmo de minería de datos.
- Por último, estaría la fase de difusión y uso del conocimiento. En esta fase se haría uso del conocimiento obtenido y se intentaría expresar de una manera comprensible para poder distribuirlo en la organización. Durante esta fase se utilizan herramientas de visualización, presentación y transformación de datos. Los conocimientos obtenidos durante el proceso ayudarán a realizar predicciones sobre sucesos futuros y permitirán una toma de decisiones mejorada.

4.1.1. TÉCNICAS DE MINERÍA DE DATOS

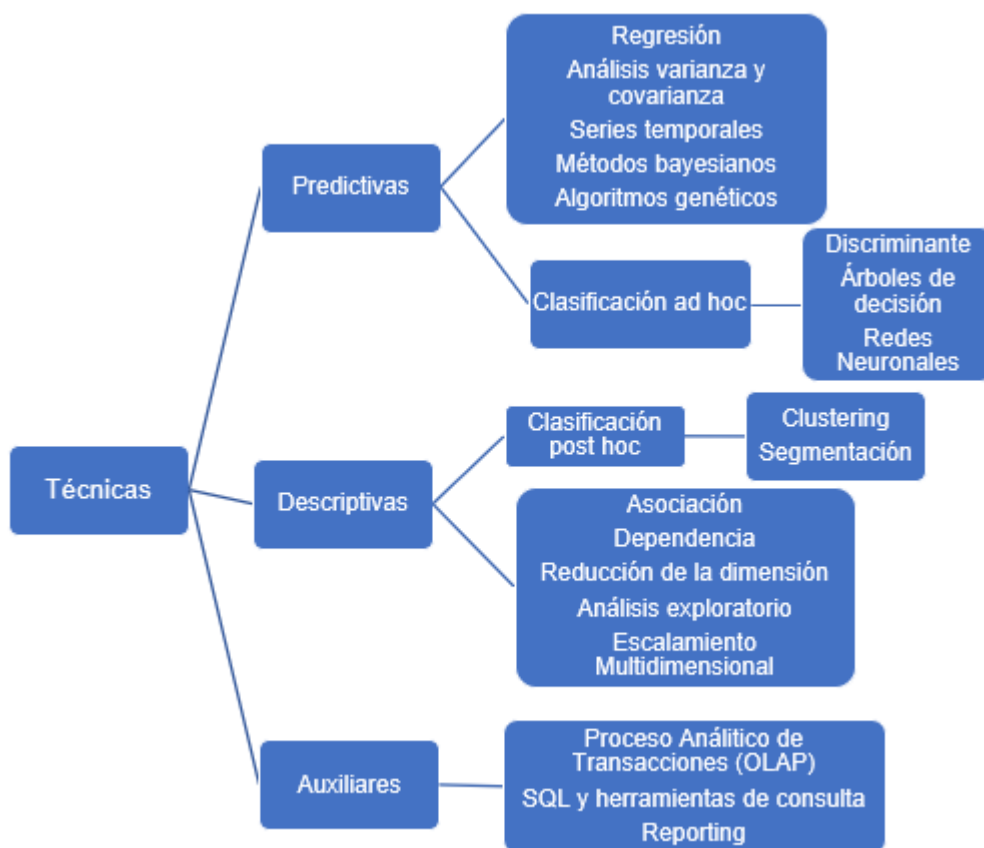
Una vez se han explicado las diferentes fases y elementos que componen la minería de datos, se deben analizar las diferentes técnicas que se pueden aplicar en el proceso anteriormente descrito. Para el autor Cesar Pérez (2007) las técnicas de minería de datos deben clasificarse inicialmente entre técnicas predictivas, descriptivas y auxiliares.

- **Técnicas predictivas.** *“Especifican el modelo para los datos en base a un conocimiento teórico previo. El modelo supuesto para los datos debe contrastarse después del proceso de minería de datos antes de aceptarlo como válido”* (Pérez, 2007). La aplicación de los modelos deberá superar las fases de identificación objetiva, estimación, diagnosis y predicción. En algunas ocasiones, el modelo se obtendrá de la mezcla del conocimiento obtenido antes y después del *Data Mining* y también deberá revisarse antes de aceptarlo como válido. Dentro de este grupo de técnicas se pueden incluir todos los tipos de regresión, análisis de varianza y covarianza, series temporales, análisis discriminante, arboles de decisión, redes neuronales, algoritmos genéticos y técnicas bayesianas.
- **Técnicas descriptivas.** En este tipo de técnicas no se asigna ningún papel predeterminado a las variables. No se supondrá la existencia de variables dependientes o independientes ni la existencia de ningún modelo previo para los datos. Los modelos se irán creando automáticamente a medida que se vayan reconociendo patrones. Dentro de esta clasificación de técnicas encontraremos técnicas de *clustering* y segmentación, técnicas de asociación y dependencia, técnicas de análisis exploratorio de los datos y técnicas de reducción de la dimensión y escalamiento multidimensional.
- **Técnicas auxiliares.** Se trata de herramientas de apoyo mucho más superficiales y con ciertas limitaciones. Consiste en nuevos métodos basados en

las técnicas estadísticas descriptivas y las consultas e informes, todos ellos enfocados hacia la verificación.

A continuación, se observa un esquema sobre la clasificación de los diferentes tipos de técnicas. Cabe destacar que existen técnicas de clasificación que pertenecen al grupo de las técnicas predictivas (árboles de decisión, redes neuronales y discriminante) y otras que pertenecen al grupo de técnicas descriptivas (*clustering* y segmentación). Las técnicas de clasificación descriptivas se denominan técnicas de clasificación *post hoc* mientras que las técnicas de clasificación predictivas se denominan técnicas de clasificación *ad hoc*.

Figura 4.1. Esquema de las diferentes técnicas de minería de datos.



Fuente: Elaboración propia a partir de Pérez (2007)

4.1.2. APLICACIONES DE LA MINERÍA DE DATOS EN LAS ORGANIZACIONES

Como se destacó anteriormente, la minería de datos permite a las organizaciones extraer conocimiento oculto tras los datos almacenados y con ello poder anticiparse a eventos futuros, descubrir nuevas oportunidades de negocio o mejorar la competitividad de la organización. En la actualidad, existen muchas áreas e industrias en las que la aplicación de minería de datos puede resultar beneficiosa. Por tanto, destacando que el campo de aplicación de la minería de datos es bastante amplio y multidisciplinar, el objetivo que se persigue en este apartado es recoger algunas de las aplicaciones más típicas de la minería de datos dentro del sector comercial y financiero. En la tabla 4.1, se recogen algunas de las aplicaciones más comunes que la minería de datos tiene en las industrias según SAS Institute.

Tabla 4.1. Aplicaciones de la minería de datos

Pregunta de negocio	Aplicación	¿Qué se predice?
¿Hacia que perfiles dirigir las ofertas de productos y servicios?	Perfiles y segmentación	Comportamientos y necesidades de clientes por segmentos
¿Qué producto/servicio recomendar?	Venta cruzada y venta directa	Posibles compras de los clientes.
¿Cómo aumentar los clientes valiosos y retenerlos?	Adquisición y retención	Preferencias de los clientes y patrones de compra
¿Cómo dirigir mejor las ofertas?	Gestión de campañas	El éxito en las comunicaciones con los clientes
¿En qué clientes invertir y cómo atraerlos mejor?	Rentabilidad y valor de vida	Factores de valor futuro (margen y retención)
¿Cómo evaluar y controlar el riesgo dentro del portafolio de consumo actual o nuevo?	Calificación de crédito (banca)	Solvencia de grupos de clientes y nuevos clientes
¿Cómo incrementar las ventas mediante ventas cruzadas/directas, programas de lealtad y promociones?	Sistemas de recomendación (<i>retail</i> en línea)	Productos que potencialmente se compren la siguiente vez
¿Cómo reducir las pérdidas por fraudes y disminuir los positivos falsos?	Gestión de fraudes y seguridad informática (seguros, bancos, gobierno)	Casos de fraude desconocidos y riesgos futuros
¿Cómo reducir las interrupciones operativas y los costos de mantenimiento?	Mantenimiento de activos (servicios, manufactura, petróleo y gas)	Factores reales de fallo de activos o equipo.

Fuente: SAS Institute Inc (2015)

4.2. FUENTE DE DATOS DEL CASO PRÁCTICO

En primer lugar, lo primero de lo que se debe disponer es de conjunto de datos sobre la que poder aplicar la herramienta de minería de datos. En este caso se ha seleccionado datos referentes a la campaña de marketing directo de una entidad bancaria portuguesa. Los datos han sido extraídos del repositorio *UCI Machine Learning Repository*² y cuya fuente original es Moro, Cortez y Rita (2014). El objetivo de la campaña de marketing era conseguir que los clientes contratasen un depósito a plazo. Se cuenta con 4521 registros y 17 atributos o variables predictorias. A continuación, se hace una breve descripción de los diferentes atributos de los que se dispone para cada registro. Cabe destacar que los datos utilizados es un extracto significativo (10%) de la base de datos original la cual cuenta con 45212 registros.

Tabla 4.2. Descripción de atributos

Atributo	Descripción
Age	Edad del cliente
Job	Variable categórica sobre la ocupación de los clientes ("admin", "unknown", "unemployed", "management", "housemaid", "entrepreneur", "student", "blue-collar", "self-employed", "retired", "technician", "services")
Marital	Variable categórica sobre el estado civil ("married", "divorced", "single")
Education	Variable categórica sobre el nivel de estudios de los clientes ('desconocido', 'secundario', 'primario', 'terciario')
Default	Variable binaria sobre si el estado del crédito es default ("yes", "no")
Balance	Saldo medio anual en euros
Housing	Variable categórica que indica si el cliente tiene préstamo para vivienda
Loan	Variable categórica que indica si el cliente tiene un préstamo personal
Contact	Variable categórica que indica si la llama fue mediante teléfono móvil o fijo
Day	Día del mes del último contacto con el cliente
Month	Mes del último contacto
Duration	Duración en segundos del último contacto
Campaign	Número de contactos realizados con el cliente durante esta campaña
Pdays	Número de días desde que el cliente fue contactado por última vez en campañas anteriores. (-1 indica que el cliente no fue contactado con anterioridad)
Previous	Número de contactos realizados previamente a la campaña para este cliente
Poutcome	Resultado de la campaña anterior de marketing ("fracaso", "desconocido", "otro", "éxito")
Y	Variable binaria que indica si el cliente contrata el depósito a plazo ("yes", "no")

Fuente: Elaboración propia a partir de Moro, Cortez y Rita (2014).

² UCI Machine Learning Repository. Bank Marketing <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

4.2. OBJETIVOS E HIPÓTESIS DEL ANÁLISIS

Como se ha explicado con anterioridad, en este caso práctico se pretende la extracción de conocimiento mediante la aplicación de minería de datos, para con ello, lograr mejorar la toma de decisiones de una entidad bancaria. Pero antes de todo se deben plantear los objetivos o metas que se pretenden alcanzar con este proceso. El objetivo que se plantea es conocer las características más comunes que presentan los clientes que contrataron el depósito a plazo fijo. De manera que, conociendo algunos de estos atributos se podría llegar a predecir con una elevada probabilidad de acierto si el cliente contratará o no el servicio. Esto permitirá a la entidad enfocar más eficientemente sus campañas de marketing hacia determinados sectores de la población y lograr un mayor éxito en sus campañas.

4.3. PREPARACIÓN DE LOS DATOS

Una vez que se ha realizado el análisis previo y se han logrado concretar las necesidades que presentan la organización y los objetivos que se pretenden alcanzar, se debe realizar la preparación de los datos para la posterior aplicación de técnicas de minería de datos. Como se destacó anteriormente, esta fase resulta fundamental para poder obtener unos resultados válidos y útiles. Durante esta fase de preparación de los datos se analizan todos los registros para comprobar que no exista presencia de valores erróneos. Por otro lado, se hace un análisis de los atributos que componen la base de datos con el objetivo de eliminar o modificar aquellos que no resulten relevantes para el análisis. También existe la posibilidad de discretizar atributos numéricos transformándolos en atributos simbólicos. Esto resulta necesario en algunas ocasiones ya que ciertas herramientas requieren de atributos simbólicos para su funcionamiento.

Como se muestra en la figura 4.2 al cargar la base de datos en Weka ya se puede observar información relevante. La herramienta Weka muestra un conjunto de datos de 4521 registros y 17 atributos. Se pueden analizar los diferentes atributos y Weka muestra información estadística sobre ellos. Por ejemplo, en el caso del atributo “age” nos indica que la edad media de los clientes era de 41 años, siendo la edad mínima y máxima 19 y 87 años respectivamente. Además, nos indica si el atributo es nominal o numérico y nos proporciona el porcentaje de registros que no tienen información sobre el atributo mediante el indicador “Missing”. Por otro lado, como se ve en la figura 4.3, proporciona un histograma para cada uno de los atributos que permite ver cómo se distribuyen los datos.

Figura 4.2. Carga de la base de datos en Weka

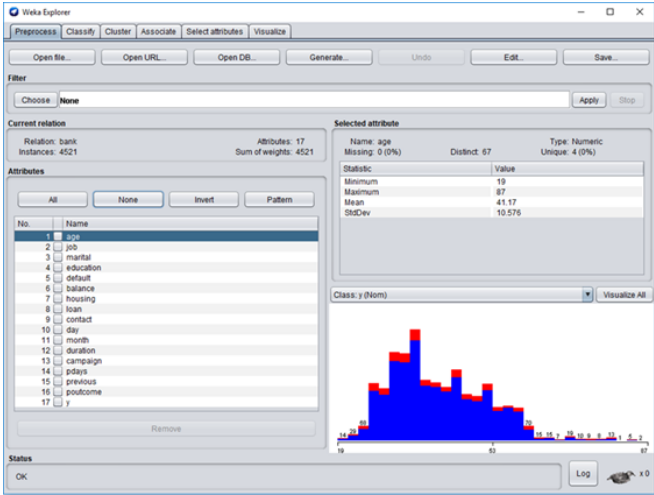
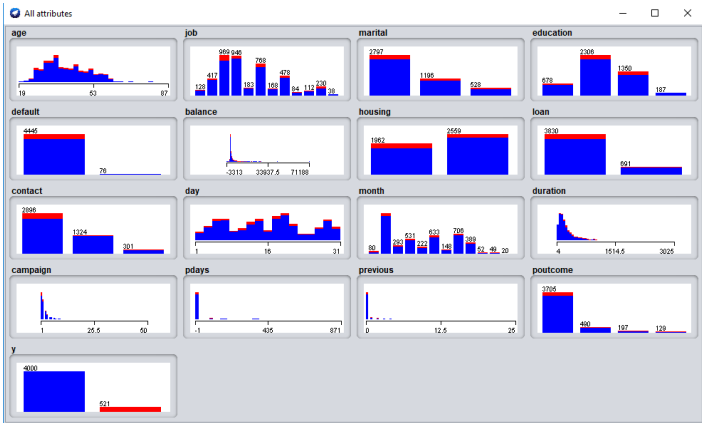


Figura 4.3. Histogramas de los atributos.



Volviendo al ejemplo de la edad, mediante el histograma se puede observar que la mayor parte de los clientes se encuentran comprendidos entre los 24 y 53 años. En el caso de los atributos nominales, Weka muestra la cantidad de registros que pertenecen a cada categoría del atributo. Dentro del atributo marital, puede verse que 2797 clientes estaban casados, 1196 solteros y 528 estaban divorciados o viudos.

Figura 4.4. Variable categórica sobre el estado civil

Selected attribute			
Name: marital		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 3	
No.	Label	Count	Weight
1	married	2797	2797.0
2	single	1196	1196.0
3	divorced	528	528.0

Mediante la opción “Edit” se puede observar la forma en la que están estructurados los datos y se puede comprobar la presencia de valores erróneos o incorrectos para corregirlos. En la figura 4.5 se observa la estructura de la base de datos que se está utilizando.

Figura 4.5. Estructura de la base de datos

No	1: age	2: job	3: marital	4: education	5: default	6: balance	7: housing	8: loan	9: contact	10: day	11: month	12: duration	13: campaign	14: pdays	15: previous	16: poutcome	17: y
	Numeric	Nominal	Nominal	Nominal	Nominal	Numeric	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Numeric	Numeric	Numeric	Nominal	Nominal
1	30.0	une...	married	primary	no	1787.0	no	no	cellular	19.0	oct	79.0	1.0	-1.0	0.0	unknown	no
2	33.0	serv...	married	secondary	no	4789.0	yes	yes	cellular	11.0	may	220.0	1.0	339.0	4.0	failure	no
3	35.0	man...	single	tertiary	no	1350.0	yes	no	cellular	16.0	apr	185.0	1.0	330.0	1.0	failure	no
4	30.0	man...	single	tertiary	no	1476.0	yes	yes	unkno...	3.0	jun	199.0	4.0	-1.0	0.0	unknown	no
5	59.0	blue...	married	secondary	no	0.0	yes	no	unkno...	5.0	may	226.0	1.0	-1.0	0.0	unknown	no
6	35.0	man...	single	tertiary	no	747.0	no	no	cellular	23.0	feb	141.0	2.0	176.0	3.0	failure	no
7	36.0	self...	married	tertiary	no	307.0	yes	no	cellular	14.0	may	341.0	1.0	330.0	2.0	other	no
8	39.0	tech...	married	secondary	no	147.0	yes	no	cellular	6.0	may	151.0	2.0	-1.0	0.0	unknown	no
9	41.0	ent...	married	tertiary	no	221.0	yes	no	unkno...	14.0	may	57.0	2.0	-1.0	0.0	unknown	no
10	43.0	serv...	married	primary	no	-88.0	yes	yes	cellular	17.0	apr	313.0	1.0	147.0	2.0	failure	no
11	39.0	serv...	married	secondary	no	9374.0	yes	no	unkno...	20.0	may	273.0	1.0	-1.0	0.0	unknown	no
12	43.0	adm...	married	secondary	no	254.0	yes	no	cellular	17.0	apr	113.0	2.0	-1.0	0.0	unknown	no
13	36.0	tech...	married	tertiary	no	1109.0	no	no	cellular	13.0	aug	328.0	2.0	-1.0	0.0	unknown	no
14	20.0	stud...	single	secondary	no	502.0	no	no	cellular	30.0	apr	251.0	1.0	-1.0	0.0	unknown	yes
15	31.0	blue...	married	secondary	no	360.0	yes	yes	cellular	29.0	jan	89.0	1.0	241.0	1.0	failure	no
16	40.0	man...	married	tertiary	no	194.0	no	yes	cellular	29.0	aug	189.0	2.0	-1.0	0.0	unknown	no
17	56.0	tech...	married	secondary	no	4073.0	no	no	cellular	27.0	aug	239.0	5.0	-1.0	0.0	unknown	no
18	37.0	adm...	single	tertiary	no	2317.0	yes	no	cellular	20.0	apr	114.0	1.0	152.0	2.0	failure	no
19	25.0	blue...	single	primary	no	-221.0	yes	no	unkno...	23.0	may	250.0	1.0	-1.0	0.0	unknown	no
20	31.0	serv...	married	secondary	no	132.0	no	no	cellular	7.0	jul	148.0	1.0	152.0	1.0	other	no
21	38.0	man...	divorced	unknown	no	0.0	yes	no	cellular	18.0	nov	96.0	2.0	-1.0	0.0	unknown	no
22	42.0	man...	divorced	tertiary	no	16.0	no	no	cellular	19.0	nov	140.0	3.0	-1.0	0.0	unknown	no
23	44.0	serv...	single	secondary	no	106.0	no	no	unkno...	12.0	jun	109.0	2.0	-1.0	0.0	unknown	no
24	44.0	ent...	married	secondary	no	93.0	no	no	cellular	7.0	jul	125.0	2.0	-1.0	0.0	unknown	no
25	26.0	hou...	married	tertiary	no	543.0	no	no	cellular	30.0	jan	169.0	3.0	-1.0	0.0	unknown	no
26	41.0	man...	married	tertiary	no	5883.0	no	no	cellular	20.0	nov	182.0	2.0	-1.0	0.0	unknown	no
27	55.0	blue...	married	primary	no	827.0	yes	no	unkno...	5.0	may	247.0	1.0	-1.0	0.0	unknown	no
28	67.0	retired	married	unknown	no	696.0	no	no	teleph...	17.0	aug	119.0	1.0	105.0	2.0	failure	no
29	56.0	self...	married	secondary	no	784.0	no	yes	cellular	30.0	jul	149.0	2.0	-1.0	0.0	unknown	no
30	59.0	adm...	married	secondary	no	485.0	no	no	cellular	24.0	aug	74.0	1.0	-1.0	0.0	unknown	no

4.4. APLICACIÓN DE ALGORITMOS DE CLASIFICACIÓN

Con el objetivo de predecir si el cliente contrata o no el depósito a plazo fijo se utilizarán técnicas de clasificación supervisada. Dentro de las técnicas de clasificación están los árboles de decisión y se aplicará el algoritmo J48 que permite predecir la categoría de una variable objetivo a partir de una serie de atributos. Dentro de los modos de evaluación se utiliza la opción *Use training set* que permite evaluar el clasificador sobre el mismo conjunto sobre el que se construye el modelo predictivo para determinar el error, que en este caso se denomina "error de resustitución" (Molina y García, 2006).

Figura 4.6. Clasificación mediante algoritmo J48

Classifier
Choose: J48-C 0.25-M 2

Test options
☒ Use training set
☐ Supplied test set
☐ Cross-validation Folds: 10
☐ Percentage split %: 66
 More options...

Classifier output

```

| | | | | job = unknown: no (0.0)
| | | | | duration > 924: yes (18.0/1.0)
| | | | | poutcome = failure
| | | | | duration <= 834: yes (2.0)
| | | | | duration > 834: no (3.0)
| | | | | poutcome = other: yes (3.0/1.0)
| | | | | poutcome = success: no (1.0)

Number of Leaves :    104
Size of the tree :    146

Time taken to build model: 0.1 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

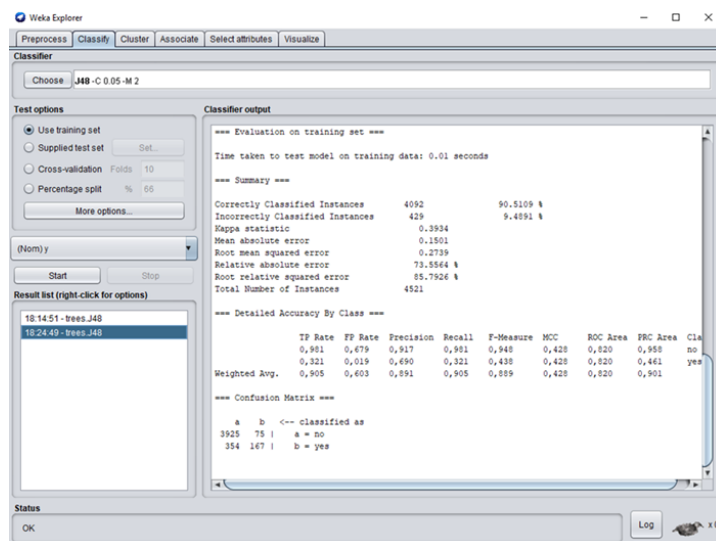
=== Summary ===

Correctly Classified Instances      4204      92.9893 %
Incorrectly Classified Instances    317       7.0117 %
Kappa statistic                    0.6103
Mean absolute error                 0.1146
Root mean squared error             0.2394
Relative absolute error             56.1578 %
Root relative squared error         74.9626 %
Total Number of Instances          4521

```

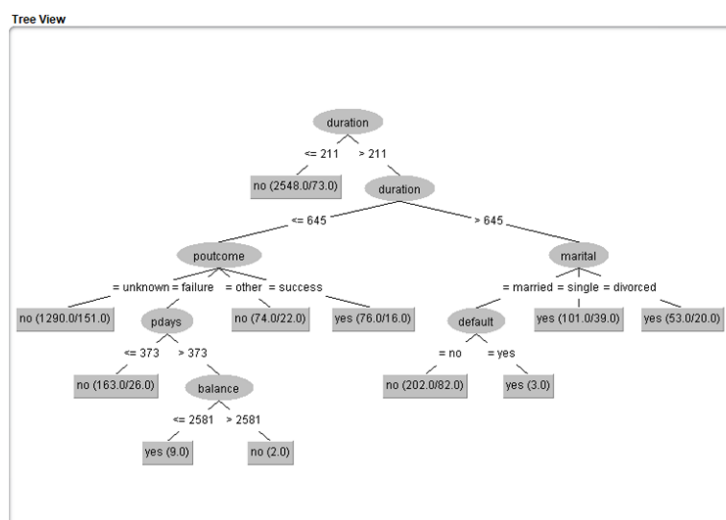
Como puede verse en la figura 4.6, el algoritmo clasifica correctamente el 92.98% de las instancias. Sin embargo, el árbol de decisión que se obtiene consta de 104 hojas y un tamaño de 146 lo cual no permite interpretarlo fácilmente. Para lograr un árbol de decisión que sí permita el análisis se disminuye el coeficiente de confianza pasando de un nivel de confianza del 25%, valor que trae por defecto Weka, a un 5% de confianza obteniéndose los resultados de la figura 4.7.

Figura 4.7. Clasificación mediante algoritmo J48 con un 5% de confianza.



Se puede ver que ahora el 90,51% de las instancias se han clasificado correctamente. Además, analizando la matriz de confusión, se observa que de las personas que no contrataron el depósito a plazo, 3925 fueron clasificadas correctamente mientras que 75 fueron clasificadas erróneamente lo que supone un 98,1% de acierto. Sin embargo, únicamente el 32,1% de las personas que si contrataron el depósito a plazo fueron correctamente clasificadas. Por otra parte, al disminuir porcentaje de confianza, se obtiene un árbol de decisión más fácil de analizar como puede verse en la figura 4.8.

Figura 4.8. Árbol de decisión para un nivel de confianza del 5%.



A partir de este árbol de decisión se pueden observar varias cosas relevantes:

- Cuando la duración de la llamada fue de 211 segundos o menos, ningún cliente contrató el depósito. Un total de 2621 clientes no contrataron el depósito, de los cuales 2548 fueron clasificados correctamente.
- Cuando la duración de la llamada fue de entre 211 y 645 segundos, si además el resultado de la anterior campaña de marketing fue “desconocido” u “otro”, entonces el cliente no contrató el depósito. Sin embargo, cuando el resultado de la anterior campaña de marketing había sido de éxito, entonces el cliente si contrató el depósito. Cuando el resultado de la anterior campaña de marketing fue de “fracaso”, el número de días desde el último contacto en campañas anteriores era mayor a 373 y el saldo de la cuenta del cliente era inferior a 2581 euros, entonces el cliente si contrató el depósito.
- En el caso de las llamadas con una duración mayor a 645 segundos, si el cliente estaba soltero, divorciado o viudo si contrató el depósito. En el caso de que el cliente estuviese casado, únicamente cuando el estado del préstamo era de incumplimiento contrató el depósito a plazo.

4.5. APLICACIÓN DE TÉCNICAS DE CLUSTERING O AGRUPAMIENTO

Otra de las técnicas que se pueden aplicar es la utilización de algoritmos de agrupamiento (clustering), permitiendo al usuario agrupar instancias que posean características similares. Para ello, desde la pestaña de *Cluster* se selecciona el algoritmo a utilizar y se fija el método de evaluación de los resultados. En este caso práctico, se ha obtenido una submuestra en la que se encuentran únicamente los clientes que sí contrataron el depósito a plazo, un total de 521 clientes. Esto permitirá obtener clusters que ayuden a identificar patrones o caracterizar a los clientes que contrataron el depósito a plazo fijo. Aplicando el algoritmo de clasificación EM, implementado en Weka, en dicha submuestra se obtienen los resultados de la figura 4.9 y 4.10. *“El algoritmo EM busca el número de “cluster” más probable dados los datos. La base de este tipo de agrupamiento se encuentra en un modelo estadístico llamado mezcla de distribuciones, donde cada distribución representa la probabilidad de que un objeto tenga un conjunto particular de pares atributo-valor”* (Pérez, Puldón y Espín, 2010).

Figura 4.9. Utilización del algoritmo de agrupamiento EM

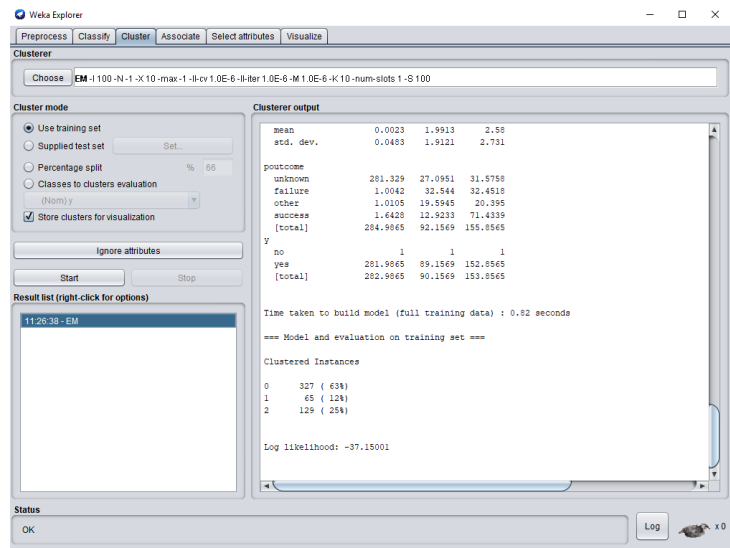


Figura 4.10. Clusters obtenidos utilizando el algoritmo EM

Attribute	Cluster		
	0 (0.54)	1 (0.17)	2 (0.29)
=====			
age			
mean	39.5162	39.2103	49.9012
std. dev.	10.2416	9.156	16.3943
job			
unemployed	8.7305	2.8233	4.4461
services	21.6217	8.8506	10.5277
management	78.9033	22.7235	32.3732
blue-collar	44.9167	20.697	6.3863
self-employed	16.5955	2.0876	4.3169
technician	47.7145	15.0262	23.2593
entrepreneur	10.6203	4.3041	3.0756
admin.	28.3852	14.7907	17.8241
student	11.9933	2.0933	7.9134
housemaid	7.8145	2.1556	7.0299
retired	13.0139	2.6116	41.3745
unknown	2.6772	1.9935	5.3293
[total]	292.9865	100.1569	163.8565
marital			
married	138.6444	44.755	96.6006
single	102.4709	32.1727	35.3563
divorced	42.8712	14.2292	22.8995
[total]	283.9865	91.1569	154.8565
education			
primary	30.7611	9.729	26.5099
secondary	132.5544	50.0663	65.3792
tertiary	112.2646	30.2741	53.4612
unknown	9.4064	2.0875	10.5061
[total]	284.9865	92.1569	155.8565
default			
no	274.2121	87.9342	152.8536
yes	8.7744	2.2227	1.0029
[total]	282.9865	90.1569	153.8565
balance			
mean	1285.2144	912.4729	2485.3739
std. dev.	1732.7267	1036.03	3611.5057
housing			
no	162.1377	19.5681	122.2942
yes	120.8488	70.5888	31.5624
[total]	282.9865	90.1569	153.8565

loan			
no	256.4706	74.0131	150.5163
yes	26.516	16.1439	3.3402
[total]	282.9865	90.1569	153.8565
contact			
cellular	220.3373	71.1838	127.4789
unknown	48.2562	14.1726	1.5711
telephone	15.393	5.8005	25.8065
[total]	283.9865	91.1569	154.8565
day			
mean	15.7317	16.5474	15.0066
std. dev.	8.2852	7.9363	8.2313
month			
oct	15.709	4.7264	19.5646
may	54.3172	32.842	8.8408
apr	31.6141	11.0652	16.3207
jun	35.9513	9.2787	12.77
feb	21.319	3.6156	16.0654
aug	49.8396	11.341	20.8194
jan	6.1138	1.0462	11.8399
jul	37.8569	13.8245	12.3186
nov	20.6391	5.0201	16.3408
sep	4.8254	2.2832	12.8914
mar	12.7015	1.5811	9.7174
dec	2.0995	3.533	6.3675
[total]	292.9865	100.1569	163.8565
duration			
mean	590.0494	782.8561	350.1257
std. dev.	360.146	541.4067	192.7983
campaign			
mean	2.1781	3.6149	1.6482
std. dev.	1.4243	3.9078	0.9572
pdays			
mean	-0.992	178.938	133.449
std. dev.	0.4809	162.1405	124.2564
previous			
mean	0.0023	1.9913	2.58
std. dev.	0.0483	1.9121	2.731
poutcome			
unknown	281.329	27.0951	31.5758
failure	1.0042	32.544	32.4518
other	1.0105	19.5945	20.395
success	1.6428	12.9233	71.4339
[total]	284.9865	92.1569	155.8565
y			
no	1	1	1
yes	281.9865	89.1569	152.8565
[total]	282.9865	90.1569	153.8565

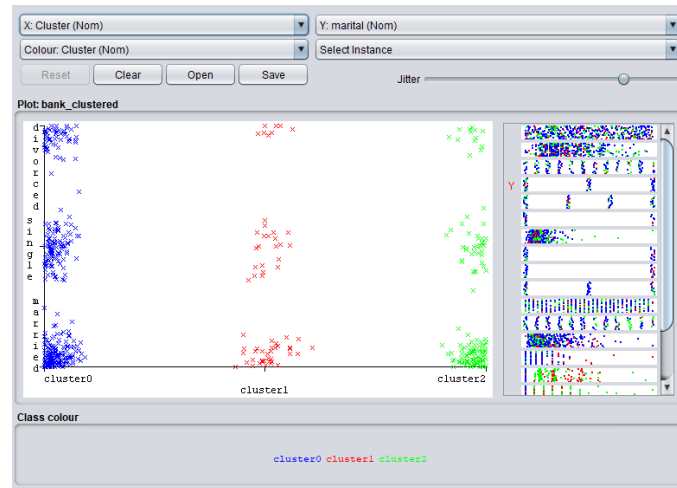
Puede verse como Weka clasifica los registros en tres *clusters*. Analizando cada uno de los clusters se puede observar información muy relevante:

- **Clusters 0.** La media de edad de los clientes de este cluster era de 39,5 años. La mayoría de los clientes de este cluster eran directivos, estaban casados y tenían un nivel de estudios secundario. El estado del crédito de estos clientes no era de incumplimiento y disponían de un saldo medio de 1285 euros. Por otro lado, la mayoría de estos clientes no disponía de créditos personales ni de vivienda. A la mayor parte de los clientes se les contactó a través de teléfono móvil durante el mes de mayo, entre los días 15 y 16 de dicho mes, y la duración media de las llamadas fue de 590 segundos. Por último, a estos clientes se les contactó 2,17 veces de media durante la campaña, no se les había contactado previamente a esta campaña y el resultado de la anterior campaña de marketing era desconocido para la mayoría de ellos.
- **Cluster 1.** La media de edad para los clientes de este cluster era de 39,2 años. La mayoría de los clientes de este cluster eran directivos, estaban casados y tenían un nivel de estudios secundario. El estado del crédito de estos clientes no era de incumplimiento y disponían de un saldo medio de 912 euros. En su mayoría, estos clientes disponían de un crédito para vivienda, pero no tenían ningún préstamo personal. A la mayor parte de estos clientes se les contactó a través de teléfono móvil durante el mes de mayo, entre los días 16 y 17 de dicho mes, siendo la duración media de las llamadas de 782 segundos. Por último, estos clientes fueron contactados de media 3,61 veces durante esta campaña y se les había contactado 1,99 veces en campañas anteriores. Habían transcurrido 179 días desde que los clientes habían sido contactados en campañas anteriores y se desconocía el resultado de la anterior campaña de marketing.
- **Cluster 2.** La media de edad para los clientes de este *cluster* era de 49,9 años. La mayoría de los clientes de este *cluster* eran jubilados, estaban casados y tenían un nivel de estudios secundario. El estado del crédito de estos clientes no era de incumplimiento y disponían de un saldo medio de 2485 euros. No disponían de créditos personales ni de vivienda. A la mayor parte de estos clientes se les contactó a través de teléfono móvil durante el mes de agosto, durante el día 15 de dicho mes, siendo la duración media de las llamadas de 350 segundos. Este grupo de clientes fue contactado 1,64 veces de media durante la campaña y se les había contactado 2,58 veces en campañas anteriores. Habían transcurrido 133 días desde el último contacto en campañas anteriores y el resultado de la anterior campaña de marketing fue de éxito para la mayoría de los clientes.

Se puede destacar que en los tres *clusters*, la mayoría de los clientes estaban casados y tenían un nivel de estudios secundario. Además, en los tres *clusters* el estado del crédito no era de incumplimiento y se les había contactado a través de teléfono móvil. Por otro lado, a pesar de que el mes del último contacto no fue el mismo para todos los *clusters*, este contacto tuvo lugar a mediados de mes en todos los casos. También hay que destacar, que los *clusters* 0 y 1 tienen varias características más en común. En ambos *clusters*, la mayoría de los clientes eran directivos y tenían una media de edad prácticamente similar. Además, a la mayoría de los clientes de estos dos *clusters* se les contactó durante el mes de mayo y el resultado de la anterior campaña de marketing era desconocido en ambos casos.

Weka también nos permite realizar un análisis visual de los *clusters* generados, como puede verse en la figura 4.11. En este caso, vemos representados los tres *clusters* y el atributo “marital” referente al estado civil en el eje de ordenadas. Puede verse que el cluster más poblado es el 0 y que en todos los clusters existe un mayor número de clientes casados.

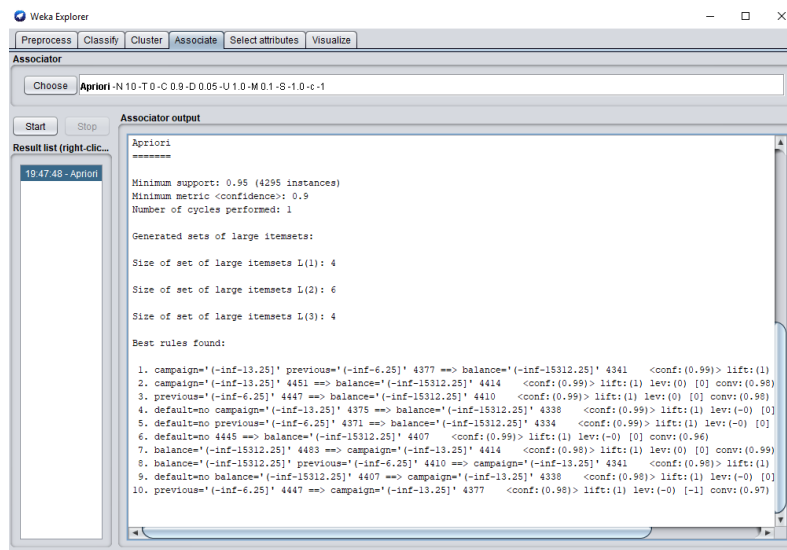
Figura 4.11. Visualización gráfica de los *clusters*.



4.6. APLICACIÓN DE TÉCNICAS DE REGLAS DE ASOCIACIÓN

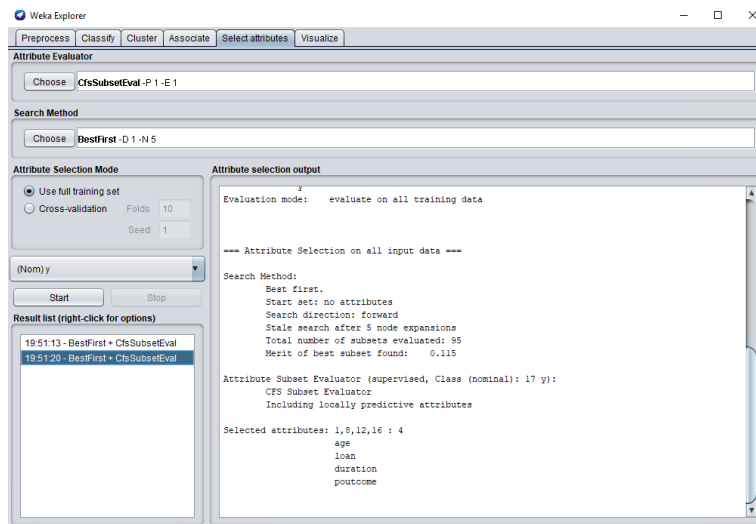
A través de la pestaña *Associate* que proporciona Weka, pueden utilizarse algoritmos de asociación. Estos algoritmos realizan una búsqueda automática de reglas que relacionen conjuntos de atributos entre sí. El algoritmo de asociación más destacado que presenta Weka, es el algoritmo *Apriori*. Este algoritmo únicamente es capaz de buscar reglas entre atributos simbólicos, por lo tanto, será necesario discretizar todos los atributos numéricos (Molina y García, 2006).

Para discretizar los atributos numéricos, desde la pestaña *Preprocess* se elige el filtro no supervisado *Discretize*, discretizando los atributos en cuatro *bins*. Una vez se han discretizado los atributos para convertirlos en simbólicos, se puede proceder a aplicar el algoritmo de asociación *Apriori*. Dicho algoritmo se ha aplicado exigiendo un nivel mínimo de confianza del 90%, obteniéndose los resultados de la figura 4.12.

Figura 4.12. Aplicación del algoritmo de asociación *Apriori*

Sin embargo, las reglas que se obtienen carecen de información útil para el análisis. Con el objetivo de obtener reglas que sean de más utilidad, se filtran aquellos atributos más importantes. Mediante la pestaña de Weka *Select Attributes* pueden verse cuales son los atributos más significativos que ayudan a explicar la variable objetivo, en este caso la variable “y” referente a la contratación o no del depósito a plazo. Se ejecuta esta selección de atributos con el método de evaluación y el método de búsqueda en sus valores por defecto. Se obtiene que los atributos que mejor explican la variable “y” son *age*, *loan*, *duration* y *poutcome*.

Figura 4.13. Selección de atributos.



A continuación, se procede a ejecutar de nuevo el algoritmo de asociación Apriori, pero esta vez utilizando únicamente los atributos obtenidos en la figura 4.13 y el atributo “y”. Se obtienen las reglas de asociación de la figura 4.14, donde destacan las siguientes reglas.

Figura 4.14. Reglas de asociación obtenidas mediante el algoritmo A priori.

```
1. loan=no poutcome=unknown y=no 2806 ==> duration='(-inf-759.25]' 2738 <conf:(0.98)> lift:(1.03) lev:(0.02) [74] conv:(2.07)
2. poutcome=unknown y=no 3368 ==> duration='(-inf-759.25]' 3285 <conf:(0.98)> lift:(1.03) lev:(0.02) [88] conv:(2.04)
3. y=no 4000 ==> duration='(-inf-759.25]' 3898 <conf:(0.97)> lift:(1.03) lev:(0.02) [101] conv:(1.98)
4. loan=no y=no 3352 ==> duration='(-inf-759.25]' 3266 <conf:(0.97)> lift:(1.03) lev:(0.02) [84] conv:(1.96)
5. poutcome=unknown 3705 ==> duration='(-inf-759.25]' 3517 <conf:(0.95)> lift:(1) lev:(0) [0] conv:(1)
6. loan=no poutcome=unknown 3113 ==> duration='(-inf-759.25]' 2954 <conf:(0.95)> lift:(1) lev:(-0) [0] conv:(0.99)
7. loan=no 3830 ==> duration='(-inf-759.25]' 3632 <conf:(0.95)> lift:(1) lev:(-0) [-3] conv:(0.98)
8. duration='(-inf-759.25]' poutcome=unknown 3517 ==> y=no 3285 <conf:(0.93)> lift:(1.06) lev:(0.04) [173] conv:(1.74)
9. loan=no duration='(-inf-759.25]' poutcome=unknown 2954 ==> y=no 2738 <conf:(0.93)> lift:(1.05) lev:(0.03) [124] conv:(1.57)
10. poutcome=unknown 3705 ==> y=no 3368 <conf:(0.91)> lift:(1.03) lev:(0.02) [89] conv:(1.26)
```

- Hubo 4000 clientes que no contrataron el depósito, la duración de la llamada fue inferior a 759,25 segundos para 3898 de ellos con un nivel de confianza del 97%.
- Se desconoce el resultado de la anterior campaña de marketing para 3705 clientes, para 3517 de estos clientes la duración de llamada fue inferior a 759,25 segundos con un nivel de confianza del 95%.
- Existen 3705 clientes para los que se desconoce el resultado de la anterior campaña de marketing, 3368 de esos clientes no contrataron el depósito a plazo con un nivel de confianza del 91%.

5. CONCLUSIONES

A lo largo del trabajo y mediante el análisis de los diferentes componentes, técnicas y metodologías que rodean al BI, se ha puesto de manifiesto la importancia que el BI tiene para las organizaciones. Como se ha visto, una correcta estrategia de BI resulta muy beneficiosa y proporciona ventajas significativas. La utilización del BI resulta fundamental para que las organizaciones consigan tratar de manera eficiente las inmensas cantidades de información que producen. A través de la aplicación del BI dentro de las organizaciones, éstas son capaces de procesar los datos, extraer información y lograr obtener un conocimiento que les permita una toma de decisiones mejorada. Por tanto, el BI supone una vía fundamental a través de la cual las organizaciones pueden ser capaces de obtener ventajas competitivas y lograr mejorar su funcionamiento. En los últimos tiempos, el BI se ha convertido en una oportunidad para las organizaciones, pero también en una obligación. En un entorno de competencia como el actual, no hacer uso del BI para mejorar el funcionamiento de la entidad, es un lujo que ninguna organización puede permitirse. Además, debe destacarse que el campo de aplicación del BI es muy amplio. No se circunscribe a ciertos sectores o industrias, sino que puede ser utilizado para múltiples fines y propósitos. Por ello, toda organización debería tener la capacidad de hacer uso del BI en su beneficio, ya que esto le ayudará a lograr sus objetivos.

Por otro lado, mediante la aplicación de minería de datos a través de la herramienta Weka, se ha podido comprobar algunas de las utilidades y ventajas que el BI puede proporcionar a las organizaciones. En este caso práctico, mediante la utilización de técnicas de clasificación, agrupamiento y reglas de asociación, se ha logrado extraer un conocimiento muy útil para la toma de decisiones de una entidad bancaria. Se ha logrado extraer conocimiento que la entidad bancaria puede utilizar para mejorar sus campañas de marketing. Gracias a la utilización de técnicas de clasificación, se ha logrado obtener información relevante en cuanto a la duración de las llamadas. Se han

analizado cuáles son los atributos que caracterizan a los clientes que sí contrataron el depósito a plazo para los distintos intervalos de duración de las llamadas. Por otro lado, mediante técnicas de agrupamiento se logró caracterizar a los clientes que sí contrataron el depósito a plazo. Se observó que la mayoría de estos clientes estaban casados y poseían un nivel de estudios secundario. Además, se trataba de clientes en su mayoría con la ocupación de directivos o ya jubilados. Esto resulta de gran utilidad para que la entidad bancaria conozca cuáles eran las características de los individuos que contrataron el depósito a plazo y pueda obtener prototipos de potenciales clientes para la entidad. Por último, mediante reglas de asociación, no se han logrado obtener reglas lo suficientemente relevantes para el análisis.

Cabe destacar, que la utilización de una submuestra de 4521 registros supone cierta limitación en el análisis, ya que sólo representa entorno al 10% de los registros de la muestra original. La utilización de la muestra original con todos los registros, seguro ayudaría en la obtención de unos resultados más representativos y definitorios. Pese a ello, conocer algunos de los rasgos que caracterizan a los clientes que contratan el depósito a plazo, sin duda ayudará a la entidad a perfilar sus campañas de marketing hacia estos sectores de la población. Esto le permitirá obtener mejores resultados en sus campañas de marketing futuras y centrar sus recursos hacia aquellos clientes más susceptibles de contratar el depósito a plazo que oferta la entidad bancaria.

6. BIBLIOGRAFÍA

- BALAGUERÓ, T. (2018). *Big Data: Componentes del Business Intelligence*. Deusto Formación
Disponible en: <https://www.deustoformacion.com/blog/gestion-empresas/big-data-componentes-business-intelligence>
- CANO GINER, J.L. (2007). *Business Intelligence: Competir con Información*. Fundación Cultural Banesto, Madrid.
- CURTO DÍAZ, J. (2010). *Introducción al Business Intelligence*. Editorial UOC, 2010.
- ELSALAMONY, H. A., & ELSAYAD, A. M. (2013). *Bank direct marketing based on neural network and C5. 0 Models*. *International Journal of Engineering and Advanced Technology (IJEAT)*, 2(6).
- ESCOBAR, C. (2014). *Componentes de Business Intelligence. Inteligencia de Negocios*.
Disponible en: <https://es.slideshare.net/elemento3000/componentes-de-business-intelligence>
- INMON, W.H. (2005). *Building the Data Warehouse*, Fourth Edition. John Wiley & Sons Inc. ISBN-13: 978-0-7645-9944-6.
- KIMBALL, R. Y ROSS, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*. Second Edition. John Wiley & Sons Inc. ISBN 0-471-20024-7
- MÉNDEZ DEL RIO, L. (2006). *Más allá del Business Intelligence*. Revista Partida Doble, núm. 181, pág. 48-55.
- MORO, S., CORTEZ, P. Y RITA, P. (2014). *A Data-Driven Approach to Predict the Success of Bank Telemarketing*. *Decision Support Systems*. Elsevier, 62:22-31.
- MOSS, L. T. Y ATRE, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision- Support Applications*. Addison-Wesley Professional.
- MOLINA LÓPEZ, J.M. Y GARCÍA HERRERO, J. (2006). *Técnicas de Análisis de Datos. Aplicaciones prácticas utilizando Microsoft Excel y Weka*.
Disponible en: http://matema.ujaen.es/inavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf
- MURILLO JUNCO, M.J. Y CÁCERES CASTELLANOS, G. (2013). *Business Intelligence y la toma de decisiones financieras: una aproximación teórica*.

Revista Logos, Ciencia & Tecnología, vol. 5, núm. 1, julio-diciembre, pág. 119-138.

Disponible en: <https://www.redalyc.org/html/5177/517751547010/>

- NAVAS MORENO, F. (2016). *Introducción a la minería de datos con Weka: Aplicación a un problema económico*. Facultad de Ciencias Sociales y Jurídicas. Universidad de Jaén.
Disponible en : <http://tauja.ujaen.es/bitstream/10953.1/6984/1/TFG%20-%20navas%20moreno%2C%20%20Francisco.pdf>
- PÉREZ LÓPEZ, C. (2007). *Minería de Datos. Técnicas y Herramientas*. Thomson, Madrid. ISBN: 978-84-9732-492-2.
- PÉREZ JIMÉNEZ, S., JAIME PULDÓN, J. Y ESPÍN ANDRADE, R.A. (2010). *Aplicación clustering en la Gestión de Procesos de Negocio*.
Disponible en: <http://ccia.cujae.edu.cu/index.php/siia/siia2010/paper/download/932/79>
- PÉREZ MARQUÉS, M. (2015). *Business Intelligence. Técnicas, herramientas y aplicaciones*. San Fernando de Henares (Madrid) RC Libros.
- SAS Institute Inc. (2015). *La Minería de Datos de la A a la Z: Cómo Descubrir Conocimientos y Crear Mejores Oportunidades*.
Disponible en: https://www.sas.com/content/dam/SAS/es_mx/doc/assets/data-minig.pdf
- UCI MACHINE LEARNING REPOSITORY. *Bank Marketing Data Set*
Disponible en: <http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>