



GRADO EN ADMINISTRACIÓN Y DIRECCIÓN DE EMPRESAS

CURSO ACADÉMICO 2019-2020

TRABAJO DE FIN DE GRADO

**Ventajas del Business Intelligence para las empresas:
caso práctico aplicado al análisis del uso de un
servicio en streaming.**

**Business Intelligence advantages: practical example
applied to the analysis of the use of a streaming
service.**

AUTORA

Marta Floranes Elorza

DIRECTORA

Eliana Rocío Rocha Blanco

Fecha: Diciembre 2020

ÍNDICE

RESUMEN.....	4
SUMMARY	4
1. INTRODUCCIÓN	5
1.1. JUSTIFICACIÓN Y MOTIVACIÓN	5
1.2. OBJETIVOS	6
2. MARCO TEÓRICO	6
2.1. DEFINICIÓN DEL BUSINESS INTELLIGENCE	6
2.2. VENTAJAS DEL BUSINESS INTELLIGENCE.....	7
2.3. COMPONENTES DEL BI	8
2.4. ETAPAS DE DESARROLLO.....	9
2.5. ERRORES COMUNES EN LA IMPLANTACIÓN DE UN SISTEMA DE BI..	14
3. HERRAMIENTAS DE BI	15
4. FUNCIONES DE BI Y TOMA DE DECISIONES	20
5. BUSINESS INTELLIGENCE EN LA ACTUALIDAD.....	21
6. CASO PRÁCTICO	22
6.1. MINERÍA DE DATOS.....	23
6.1.1. Fases de la Minería de Datos	23
6.1.2. Técnicas de la Minería de Datos	24
6.2. DATOS DEL CASO PRÁCTICO	25
6.2.1. Pre-procesamiento de los datos	27
6.2.2. Aplicación de técnicas de Minería de Datos.....	27
6.3. APLICACIÓN DE TÉCNICAS DE ASOCIACIÓN	30
6.4. APLICACIÓN DE TÉCNICAS DE AGRUPAMIENTO O CLUSTERING.....	33
7. CONCLUSIONES	36
BIBLIOGRAFÍA.....	37

ÍNDICE DE FIGURAS Y TABLAS

Figura 2.1.....	9
Figura 2.2.....	10
Figura 2.3.....	11
Figura 2.4.....	13
Figura 3.1.....	17
Tabla 3.1.....	18
Figura 6.1.....	25
Tabla 6.1.....	26
Figura 6.2.....	28
Figura 6.3.....	28
Figura 6.4.....	29
Figura 6.5.....	30
Figura 6.6.....	31
Figura 6.7.....	32
Figura 6.8.....	33
Figura 6.9.....	34
Figura 6.10.....	35

RESUMEN

En vista de los grandes volúmenes de datos e información que generan y están a disposición de las organizaciones, el proceso de toma de decisiones de una manera efectiva se convierte en un desafío para los empresarios.

Así como un desafío, resultarán procesos clave la gestión y análisis de estos datos para conseguir los objetivos empresariales generando conocimiento útil. Esto se hará a través del uso del Business Intelligence, un sistema utilizado ampliamente ahora que, además, impulsará la eficacia e innovación de la organización.

Durante este trabajo, se revisará el concepto de *Business Intelligence*, además de la diferentes fases, componentes y herramientas que lo engloban. Por otro lado, se hace referencia a las ventajas que tiene su implementación para la empresa y los errores comunes en los que se puede incurrir.

Para finalizar, se habla de cómo ha ayudado el *Business Intelligence* en la situación actual de la pandemia del COVID-19 en el sector salud y se plantea un caso práctico estudiando la situación de una plataforma digital frecuente hoy en día. Este caso práctico se realiza a través de una aplicación de minería de datos denominada Weka, pretendiendo extraer conocimiento útil.

SUMMARY

In view of the huge volumes of data and information generated and that are available for the organizations, the effective decision-making process has become a big challenge for managers.

As a challenge, there will be a key process the management and analysis of data for achieving business goals generating useful knowledge. This will be done through using Business Intelligence, a system used widely nowadays that in addition, will boost efficiency and innovation across the organization.

Over this project, it will be what Business Intelligence is, also the different phases, elements and tools that embrace it. On the other hand, it will be also mentioned the benefits of its implementation and the common mistakes we can run into.

Finally, we talk about how Business Intelligence has helped in the situation of a frequent platform used nowadays and a practical case is analyzed studying the situation of a typical platform. This practical case is done through the use of Weka, which is a data mining application, with the aim of abstract useful knowledge.

1. INTRODUCCIÓN

En el mundo tan competitivo en el que nos encontramos actualmente, es de vital importancia para el éxito de las empresas que estas encuentren formas de diferenciarse de la competencia.

El *Business Intelligence* (a continuación, BI) se ha convertido en una técnica clave para alcanzar esta ventaja y cada vez es más importante para conseguir el éxito de la empresa en cualquiera que sea la industria en la que se desenvuelve.

Tradicionalmente, estas técnicas eran utilizadas exclusivamente por grandes organizaciones y multinacionales que tenían implantadas las tecnologías más novedosas ya que eran las que podían hacer frente al coste que supone, siendo muy común en el sector de la banca, el gran consumo y las telecomunicaciones.

Conforme ha ido pasando el tiempo, ha sido (y es) una técnica aplicada por la mayoría de empresas de todas las industrias debido a que es una gran manera de mejorar el servicio y los productos que se ofrecen a los ciudadanos, ya que agiliza y ayuda el proceso de toma de decisiones.

Actualmente, con el abaratamiento de las técnicas de computación, las licencias de uso comerciales y la existencia de muchas opciones de aplicaciones de código abierto, hace que cualquier empresa pueda implantar proyectos de BI a un coste asumible. Por el contrario, las empresas que no apuestan por las tecnologías, serán absorbidas por otras grandes corporaciones.

1.1. JUSTIFICACIÓN Y MOTIVACIÓN

El BI, que se analizará en profundidad a lo largo de este trabajo, es un pilar fundamental en el panorama empresarial siendo uno de los campos con mayor crecimiento. Antiguamente, los datos eran recopilados de manera manual, pero poco a poco fueron surgiendo nuevos métodos para agilizar y mejorar este proceso cuando la cantidad de datos e información aumentó tanto que supuso la incapacidad por parte de las personas de seguir haciéndolo manualmente. Hoy en día, con el creciente e importante progreso que ha experimentado la tecnología, se almacenan grandes cantidades de información en bases de datos. Con la implementación de estrategias de BI, que los que se encargan de la toma de decisiones en las empresas sean capaces de integrar diferentes fuentes de datos, predecir tendencias, mejorar el rendimiento y así identificar oportunidades de negocio. Se puede añadir que, en el siglo XXI, el BI es uno de los principales campos de batalla para las organizaciones, y todas las empresas están intentando hacer uso de ello para atraer y retener las mejores ventajas competitivas y anticiparse a sus competidores.

Esta importancia que tiene el BI en el ámbito empresarial ha sido una de las principales motivaciones que me ha llevado a desarrollar este proyecto. Además de la enorme amplitud del concepto, todo lo que le rodea y le compone y lo útil que es hoy en día para prácticamente cualquier sector que podamos imaginar. Conocer en profundidad los aspectos más relevantes y tener un conocimiento más amplio sobre la aplicación de estos sistemas de BI en las organizaciones me será de gran utilidad para el desarrollo de mi trayectoria profesional y crecimiento personal.

1.2. OBJETIVOS

Pese a la importancia que ha alcanzado el BI en la actualidad, continúa siendo un concepto desconocido debido a su amplitud y extensión para la sociedad en general. El objetivo principal de este trabajo es estudiarlo y aclararlo mediante la descripción de qué es y lo que lo compone. En primer lugar, se comienza con un marco teórico donde se expone una definición clara para que todo el mundo pueda comprender de qué se trata. También se detallan las fases y principales ventajas que una estrategia de BI puede proporcionar a las organizaciones, además de los errores comunes que pueden surgir a la hora de implementarlo.

Por otro lado, se esclarecen cuales son las herramientas más relevantes y las funciones de BI, así como cómo ayuda en la toma de decisiones.

Esta primera parte teórica finaliza con el análisis de las facilidades alcanzadas con los sistemas de BI en la actualidad, las ventajas que supone su implementación sobretodo en el sector salud para agilizar el proceso de toma de decisiones y recopilación de datos de manera eficiente durante la pandemia del COVID-19 y la importancia que ha tomado el elaborar correctamente estos sistemas para mantener la posición competitiva.

Tras esta primera parte, se sigue el trabajo con un caso práctico. Con la realización de este caso práctico, se pretende estudiar y evidenciar alguna de las ventajas que supone el uso de los sistemas de BI para las empresas. En este punto, se hace uso de una aplicación de minería de datos denominada Weka. Debido a la amplitud y complejidad de la minería de datos, también se ha explicado de manera detallada las fases y técnicas que lo componen para ayudar a su comprensión.

Podemos concluir que el objetivo no es otro que, tras el estudio del comportamiento de los usuarios, se saque conocimiento útil que suponga una ventaja y ayuda para la toma de decisiones de la empresa.

2. MARCO TEÓRICO

2.1. DEFINICIÓN DEL BUSINESS INTELLIGENCE

La evidencia sugiere que las universidades modernas han luchado, y los siguen haciendo, para educar a los profesionales en el tremendo impacto que el Business Intelligence puede tener en una organización si es usado correctamente. El Business Intelligence, (en adelante BI), es la ventajosa implementación de las tecnologías de soporte de decisiones para la empresa con el fin de hacer posible que los trabajadores del conocimiento como son ejecutivos, gerentes y analistas tomen mejores decisiones y más rápidas” (Davenport, 2006).

Las distintas empresas pueden definir el éxito de manera diferente, pero el éxito en BI generalmente será medido como la efectividad del análisis de datos a la hora de facilitar la toma de decisiones. (Hemsley, 2011).

El uso del BI puede estar directamente relacionado con el aumento de los beneficios, si la información no es utilizada para la maximización de los mismos, será inútil.

Así que, en resumen, se puede decir que el objetivo de BI es apoyar de forma razonable y continua a las organizaciones para mejorar su competitividad, facilitando la información necesaria en el momento adecuado para la toma de decisiones. (Villamizar Gutiérrez, 2010).

Antes del BI, la mayoría de las empresas analizaban sus operaciones empresariales utilizando aplicaciones de soporte de decisiones que consultan e informan directamente sobre los datos almacenados en las bases de datos de las transacciones empresariales. Existen varios problemas con este enfoque. Las 5 claves a tener en cuenta son (Murugesan y Karthikeyan, 2016):

- Los datos no suelen estar en una forma adecuada para informar
- Los datos pueden tener problemas de calidad
- El proceso de toma de decisiones degrada la ejecución de los procesos comerciales
- Los datos suelen estar dispersos en muchos sistemas diferentes
- Hay una falta generalizada de información histórica

2.2. VENTAJAS DEL BUSINESS INTELLIGENCE

El BI es una técnica que toma datos en bruto y los convierte en información relevante y práctica que las empresas pueden usar para fortalecer sus posiciones y decisiones comerciales. Las diferentes herramientas del BI analizan grandes volúmenes de datos basados normalmente en consultas escritas para obtener determinados tipos de información. Los resultados se formatean y como resultado obtenemos esta información en forma de resúmenes, gráficos, tablas, mapas...que serán útiles para analizarlos en cuanto a la toma de decisiones.

Las empresas quieren conocer todo lo posible a sus clientes y competidores y así actuar en consecuencia con sus gustos, preferencias, inquietudes, etc. Para ello, recopilan grandes cantidades de datos que manualmente sería imposible de analizar, aquí es donde entra en juego el papel del BI. A continuación, se verá algunos de los beneficios del uso del BI para la mejora de los procesos internos recopilando información basada en el flujo de trabajo:

- Reducción del tiempo necesario para tomar decisiones
- Optimizar los procesos internos ayudando así a los empleados a centrarse en el trabajo de mayor valor
- Reducción del tiempo de sacar un producto o servicio al mercado
- Aumento de la satisfacción del cliente mediante mejores eficiencias y un mejor servicio
- Liberar más tiempo para centrarlo en cosas como la calidad e iniciativas de retención de clientes
- Mejora general de la eficiencia operativa
- Agilización de procesos

2.3. COMPONENTES DEL BI

Como se ha ido mencionando a lo largo de este trabajo, el BI, entre otras cosas, permite a las empresas analizar y combinar datos procedentes de distintas fuentes y obtener una visión conjunta y simplificada con el fin de anticiparse a lo que sucederá y hacer la toma de decisiones más sencilla y efectiva.

Para conseguir que esto funcione, es importante entender que el BI se desarrolla a través de una herramienta tecnológica y un conocimiento profundo del *core business* de la empresa. (Solano, 2017).

Según Laudon y Laudon (2012), el BI está formado por tres componentes y cuatro procesos. Como podemos ver en la Figura 2.1, los tres componentes serían las fuentes de datos, el almacén de datos o *Data Warehouse*; y, por último, está el tablero de control o *Dashboard*. Las fuentes de datos hacen referencia al conjunto de datos extraídos por la organización que serán utilizados para alimentar el almacén de datos y con ello realizar las tareas de análisis y filtrado. El *Data Warehouse*, se define como una base de datos integral que consolida, estandariza y almacena datos actuales e históricos de potencial interés para aquellos que se encargan de la toma de decisiones dentro de la empresa. El tercer y último componente es el *Dashboard* que son pantallas interactivas que muestran información del negocio de forma visual y didáctica, normalmente usando umbrales de evaluación, elementos gráficos de fácil evaluación e indicadores clave de desempeño (KPIs).

También se puede observar en la Figura 2.1 los cuatro diferentes procesos comprendidos en el BI:

1. Proceso de extracción, transformación y carga de datos (*ETL Process*): este proceso se refiere al traspaso de la información de las fuentes de datos hacia el *Data Warehouse*. Su objetivo es el procesamiento sistemático de los datos de origen para que estén disponibles en el mejor formato posible para el uso previsto. (Galici et al., 2020).
2. Proceso analítico en línea (*OLAP Process*): capacidad de manipular y analizar grandes cantidades de datos desde diferentes perspectivas y además apoya la exploración interactiva de esos conjuntos de datos.
3. Proceso de reporte de datos (*Reporting*): este proceso consiste en la elaboración de informes y consultas a detalle sobre temas determinados de un área específico.
4. Proceso de minería de datos (*Data Mining*): la minería de datos proporciona información sobre datos que no se pueden obtener a través del proceso analítico en línea (*OLAP*) al encontrar patrones y relaciones ocultos en grandes bases de datos y sacar reglas de estos datos para predecir el comportamiento futuro. (Laudon y Laudon, 2014).

Figura 2.1. Componentes de un sistema de Business



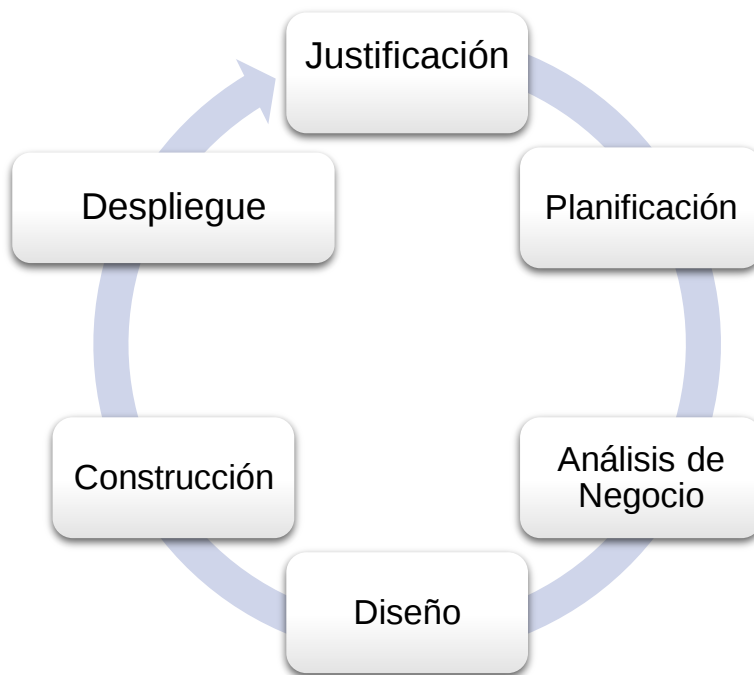
Fuente: Elaboración propia a partir de Solano (2017).

2.4. ETAPAS DE DESARROLLO

Casi todos los proyectos de BI se desarrollan en 6 etapas, que a su vez tienen otras subetapas. Como se puede ver en la Figura 2.2, las etapas que lo comprenden son: justificación, planificación, análisis de negocio, diseño, construcción y finalizando con la implementación (Moss y Atre, 2003). Pero, se ha de saber que aun que finalice con la implementación del proyecto, no se acaba ahí. Estos proyectos están en continuo control y supervisión para continuar mejorándolos y perfeccionándolos, basándose en el feedback de los diferentes usuarios que lo utilicen. Como se observa, se trata también de un proceso repetitivo.

1. Justificación: evaluar las necesidades de negocio que dan lugar a este tipo de proyectos.
2. Planificación: desarrollar planes tanto estratégicos como tácticos que exponen cómo el proyecto será realizado e implementado.
3. Análisis de Negocio: realizar un análisis detallado de los problemas y oportunidades del negocio para ganar un entendimiento sólido de los requerimientos comerciales para establecer una posible solución.
4. Diseño: proyectar un producto que resuelva nuestro problema o establezca una oportunidad de negocio.
5. Construcción: construir el producto, el cual nos debe devolver lo invertido con los beneficios que se alcancen en un marco de tiempo predeterminado, que normalmente será a medio o largo plazo.
6. Implementación: instalar o vender el producto final, después medir su efectividad para ver si cumple, no cumple o excede con lo previsto para la recuperación de lo invertido.

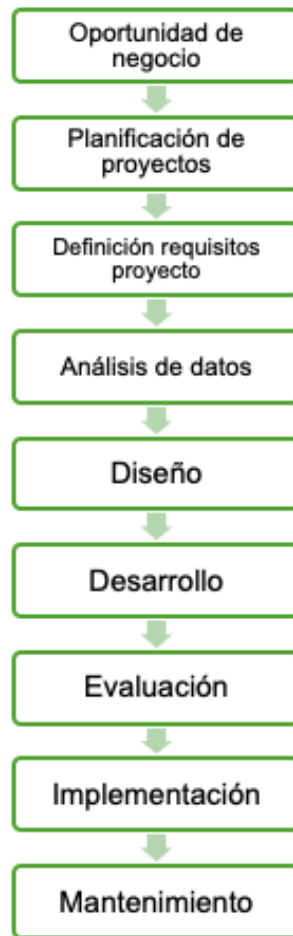
Figura 2.2. Etapas del desarrollo del BI.



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

En el pasado, los sistemas no se diseñaban o implementaban nunca con el fin de ser algo integrado. Todos los sistemas tenían un comienzo y un fin y, además, se diseñaban para solucionar un solo problema en un momento determinado. Esas prácticas podían servir por aquel entonces pero no sirven para proyectos de BI, ya que como se ha comentado anteriormente, se necesita la involucración de toda la organización y proporcionar un soporte a la toma de decisiones eficiente. Esto difiere mucho con lo que se pensaba antiguamente ya que no solo se creía que la involucración completa de la empresa en las diferentes actividades era innecesaria, sino que también se pensaba que entorpecía el proceso porque desaceleraban la implementación de los proyectos. Las metodologías en cascada convencionales, como la que se puede ver en la Figura 2.3, pueden ser suficientes para el desarrollo de sistemas no integrados ya que proporcionan orientación suficiente para la planificación, construcción y aplicación de sistemas autónomos. Sin embargo, estos métodos tradicionales no cubren la planificación estratégica, el análisis de negocios de toda la organización, o la evaluación de las nuevas tecnologías con cada proyecto; no aceptando tampoco el concepto de comunicación entre aplicaciones. (Moss y Atre, 2003). Como se observa en la Figura 2.3, estos métodos tradicionales suelen empezar con una necesidad, tienen diferentes procesos pasando por el análisis y desarrollo hasta llegar al mantenimiento. No es necesario llevar a cabo los pasos en secuencia, es más, la mayoría de organizaciones los lleva a cabo en paralelo (Villamizar Gutiérrez, 2010).

Figura 2.3. Despliegue convencional en cascada.



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

Etapa de Justificación

- Paso 1: Evaluación del caso de negocio: se define el problema o la oportunidad de negocio, y se propone una solución de BI.

Etapa de Planificación

- Paso 2: Evaluación de infraestructura de la empresa: como hemos concluido, los proyectos de BI son iniciativas que incumben a toda la empresa, esta debe crear una infraestructura para apoyarlas. La infraestructura de una organización tiene que tener dos componentes: infraestructura técnica y no técnica.
- Paso 3: Planificación de Proyectos: los proyectos de BI son extremadamente dinámicos. Los cambios en el personal, en el presupuesto, en la tecnología, etc.,

pueden afectar seriamente al éxito del proyecto. Por esto, la planificación ha de ser detallada y supervisada. (Villamizar Gutiérrez, 2010).

Etapa de Análisis de Negocio

- Paso 4: Definición de los requisitos del proyecto: quienes realizan este tipo de proyectos saben, o deben saber, que los requisitos van cambiando a medida que se va desarrollando. Por esto y por la necesidad de reducir el querer tener todo al instante hace que se convierta en algo difícil pero imprescindible de conseguir.
- Paso 5: Análisis de Datos: el mayor desafío de los proyectos de BI es conseguir buena calidad de los datos de origen, algo que va a ser crucial para el éxito de todo el proyecto. Esto va a dedicar un porcentaje muy alto en el tiempo invertido en el proyecto al completo.
- Paso 6: Prototipo de la aplicación: gracias a las nuevas herramientas y lenguajes de programación, los programadores pueden probar con relativa rapidez ideas o conceptos.
- Paso 7: Análisis del repositorio de metadatos (Data Warehouse): en nuestro proyecto tendremos diferentes tipos de metadatos (técnicos, de negocio...) que tienen que estar almacenados en un repositorio de metadatos lo que les hace tener que ser lógicos. Además, deberán de ser analizados.

Etapa de Diseño

- Paso 8: Diseño de la base de datos: en BI se pretende almacenar de forma general y detallada los datos del negocio en función de lo que se pretenda conseguir. Los esquemas de diseño de bases de datos deben coincidir con los requisitos de acceso a la información de la empresa. (Moss y Atre, 2003)
- Paso 9: Diseño del ETL: este paso es el más complicado de todo el proyecto de BI, por lo que requerirá mucho tiempo.
- Paso 10: Diseño del repositorio de metadatos (Data Warehouse): si estamos construyendo nosotros un repositorio de metadatos (y no lo compramos), se debe tomar la decisión de si este diseño lo basamos en entidad-relación u orientado a objetivos.

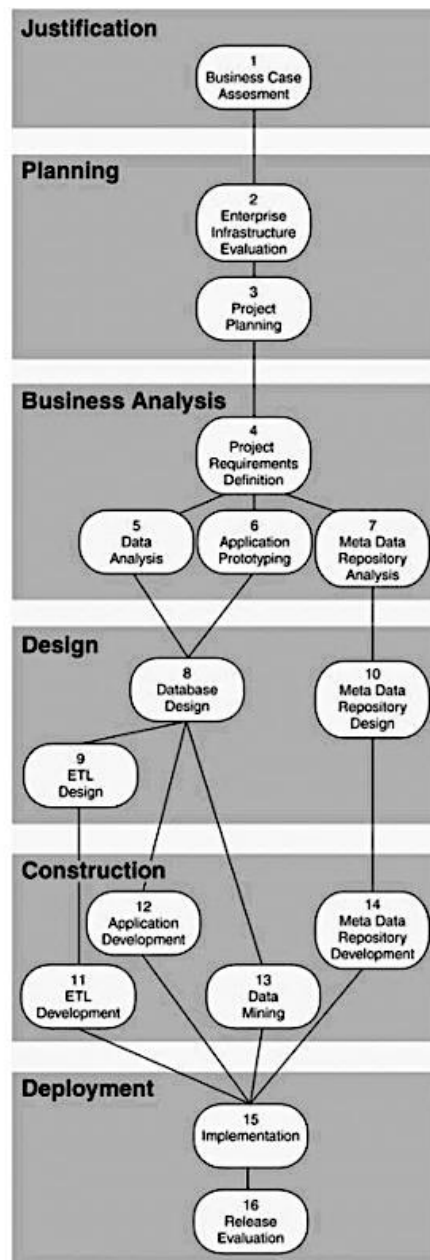
Etapa de Construcción

- Paso 11: Desarrollo del ETL: se requiere con frecuencia el pre-procesamiento de los datos y la creación de ampliaciones para complementar las capacidades de esta herramienta (Moss y Atre, 2003)
- Paso 12: Desarrollo de aplicaciones: puede ser más simple como la finalización de un prototipo operativo, o algo mas complejo como un esfuerzo de desarrollo involucrado con herramientas de acceso y análisis diferentes y más fuertes.
- Paso 13: Minería de Datos: hay datos ocultos que solo podemos descubrir con las herramientas que nos proporciona la minería de datos.
- Paso 14: Desarrollo del repositorio de metadatos: si se decide construir un repositorio de metadatos en lugar de comprarlo, este proceso será encomendado a un equipo de trabajo independiente (sub-proyecto)

Etapas de Despliegue

- Paso 15: Implementación: una vez que el equipo a probado a fondo todos los componentes del proyecto de BI, libera las bases de datos y aplicaciones que lo componen estando la información disponible.
- Paso 16: Evaluación del lanzamiento: a pesar de haber terminado el proyecto, este deberá ser examinado y supervisado preparándolo para siguientes versiones y eliminar o ajustar aquello que no estaba siendo útil.

Figura 2.4. Etapas y pasos de un sistema de BI.



Fuente: Moss y Atre (2003).

2.5. ERRORES COMUNES EN LA IMPLANTACIÓN DE UN SISTEMA DE BI

Las diferentes ventajas, así como los beneficios que se alcanzan cuando se implanta un sistema de BI, sólo se hacen efectivas, como es lógico, cuando este se implementa con éxito.

Según Medina La Plata (2012) existen 7 errores comunes a la hora de la implementación de Business Intelligence, que, por desgracia, las empresas siguen cometiendo.

1. Enfoque exclusivamente técnico: sería erróneo creer que el BI se trata únicamente de llevar a la herramienta analítica elegida los reportes que la empresa dispone a partir de las fuentes de datos transaccionales u hojas de cálculos. Con esto se estaría consiguiendo sólo una automatización del proceso y no agregando el valor que la organización demanda. (Medina La Plata, 2012). La esencia que debe guiar a la hora de implementar BI en una empresa debe ser un cambio funcional y generalizado de la cultura de la organización, basándose en el análisis de información. Es decir, se tiene que entender lo que la organización necesita para así aplicar los procesos necesarios que ayuden a la toma de decisiones sin adquirir tecnología que no vaya a ser utilizada, involucrando a todos los departamentos y no creyendo que simplemente se trata de algo tecnológico.
2. Mala selección del proveedor: es muy importante dejar en manos de un experto los proyectos de BI. Es habitual, debido a la creciente demanda de este área, que cada vez más proveedores se sumen a la realización de este tipo de tareas. Sin embargo, es muy necesario confirmar que tengan a su alcance equipos de trabajo preparados para estas tareas, así como del conocimiento necesario para conseguir los objetivos para los que se encomienden. Como se ha comentado en el error anterior, otro punto a tratar en relación a este aspecto es que algunos proveedores enfocan estos proyectos meramente como algo tecnológico, lo cual nos llevaría al mismo lugar que lo mencionado anteriormente. Medina La Plata (2012), recomienda establecer una etapa de selección de tecnología, que tenga en consideración las necesidades de información globales de las áreas de oportunidad y, a partir de ahí, establecer los aspectos que califiquen tanto las herramientas como los proveedores.
3. Mala calidad de los datos: un factor crítico de éxito en los proyectos de BI, es la calidad de los datos (Medina La Plata, 2012). Se ha de estar totalmente seguro de que la información resultante de los datos que se analizan es verídica, o por el contrario, sólo se conseguirá que los usuarios que reciben esta información no confíen en la solución y dejen de usarla. Hay que prestar especial atención aquí, ya que es algo complicado debido a los grandes volúmenes de datos que reciben las empresas.
4. Falta de planificación: como una de las etapas de desarrollo de BI, la planificación puede ser objeto de error común de las empresas. Muchas organizaciones se lanzan al primer proyecto, porque está demostrado que

aumenta los beneficios y es cada vez más común en empresas de todo tipo, sin planificar adecuadamente antes. Es importante determinar cuales son las áreas de oportunidad de la empresa donde sería efectivo la implementación de este tipo de iniciativas, así como estudiar cual sería la mejor técnica para cubrir las necesidades de dichas áreas. Son numerosas las preguntas a plantearse antes de iniciar un proyecto de este calibre.

5. Presupuesto inadecuado: una iniciativa de Business Intelligence supone una gran inversión ya que son varios aspectos a los que hay que tener en cuenta, como, por ejemplo: una infraestructura tecnológica, licencias de herramientas analíticas, licencia de base de datos y consultoría, además, a esto se le suma el compromiso de hacer crecer anualmente el numero de usuarios, a lo que se sumaria un mayor numero de licencias. No hay que menospreciar ninguno de estos aspectos ya que todos son relevantes y necesarios si se quiere que el proyecto de BI dé resultado y se obtengan beneficios que estén por encima de estos costes.
6. Mala selección de las herramientas: esto es uno de los errores más usuales que originan el deterioro de los proyectos. Hay que asegurarse que se identifican las áreas de oportunidad de la empresa y según esto buscar la alternativa tecnológica que mejor se adecue para satisfacer las necesidades de información y funcionalidades que se tengan en ese determinado momento. Es imprescindible también involucrar a los usuarios y no sólo al área tecnológica o informática.
7. No propiciar el cambio: la implementación de un sistema de BI debe comenzar con el convencimiento de su necesidad en la organización, además de estar dispuestos a dar los pasos necesarios para que sea un proyecto exitoso. Es importante comprender que lo que principalmente se quiere conseguir con este tipo de proyectos es un cambio en la organización basado en atender sus necesidades de gestión con un acertado uso de la tecnología y la información.

3. HERRAMIENTAS DE BI

Si hay algo peor que no tener información es tenerla y no saber qué hacer con ella. Por esto, surgen las diferentes herramientas del BI que han ido evolucionando a lo largo del tiempo. En el campo del Business Intelligence (BI), la tendencia se esta utilizando hacia la utilización de herramientas basadas en memoria y con capacidades de lógica asociativa de consultas, que además permitan dar soporte a la toma de decisiones, presentando la información de una manera sencilla e intuitiva permitiendo al usuario poseer el conocimiento que necesita oportunamente. (Ayala, J., et al., 2018)

Anteriormente y durante décadas, las tecnologías tradicionales de BI basadas en disco, han estado afectadas por tiempos de respuestas lentos. Al necesitar velocidad en el análisis de los datos y gracias al avance de la tecnología, fueron surgiendo las soluciones de BI basadas en memoria para reemplazar a las anteriores, que ofrecen mejores tiempos de respuesta.

Este tipo de herramientas basadas en memoria y en la utilización de lógica asociativa de consultas, les permiten volverse más ágiles y más flexibles, requisitos que serán buscados por los usuarios de BI, además de ser más sencillos de utilizar. Otra de las características que recoge este tipo de herramientas es que son más receptivos ante los cambios que puedan producirse.

Con el objetivo de revisar las herramientas líderes del mercado, se muestra a continuación una la clasificación para la evaluación de soluciones de BI que se denomina el “Cuadrante Mágico de Gartner”, en el cual se indican las herramientas de BI de mejor rendimiento. Para crear este cuadrante, el cual se observa en la Figura 3.1., Gartner ha evaluado más de 200 soluciones de BI diferentes disponibles en el mercado.

Aquellos que recibieron las puntuaciones mas altas en cuanto a su valor y habilidad pudieron pasar a formar parte de dicho cuadrante. Dichas puntuaciones, se basan en evaluaciones de los usuarios reales de las soluciones de BI, junto con una auditoria del rendimiento del producto clave realizada por expertos. (Tripathi, A., & Bagga, T, 2020).

Cada uno de los productos que forman parte de este cuadrante, ofrecen una utilidad y un equilibrio de competencia que hacen que destaquen. Tras la evaluación de Gartner y como se puede apreciar en la Figura 5, se puede concluir que los líderes entre todos los softwares son Microsoft, Tableau, Qlik y ThoughtSpot. Siendo este último una nueva incorporación a los demás que llevaban siendo lideres varios años; irrumpió en el cuadrante en 2017 posicionado en la zona de “Niche Players”, en 2018 en el cuadrante de “visionaries” y en 2019 ya irrumpió en el de los líderes uniéndose a los otros tres ya más veteranos. (Vicente Ferrer, J., 2019)

Figura 3.1. Cuadrante Mágico de Gartner.



Fuente: Gartner, Inc. (2019).

A continuación, se hará un breve análisis de las herramientas líderes de BI que están presentes en el “Cuadrante Mágico de Gartner”.

- Tableau: fue desarrollado con la visión de proporcionar a los usuarios empresariales un software que permitiera un análisis intuitivo. Además, se buscaba que fuese fácil de usar y que permitiese tener una visión de los datos. Como estrategia, pretende ofrecer una experiencia de inteligencia de negocios que permita a los usuarios interpretar mejor sus propios datos mediante la visualización y que para utilizarlo no sea necesaria gran cualificación ni entrenamiento. La interfaz de usuario intuitiva, la inteligencia incorporada y la utilización de la memoria principal centrada en optimizar el rendimiento contribuyen a la popularidad de esta solución. Así como la conectividad con varias fuentes de datos. . (Ayala, J., et al., 2018)
- Microsoft BI: Microsoft ofrece un conjunto de soluciones de BI en las que se puede identificar las que están relacionadas con la generación de las bases de

datos y otras soluciones relacionadas con los informes y visualizaciones que se pueden hacer sobre estos datos. . (Ayala, J., et al., 2018)

- QlikView: es una plataforma de BI moderna, creada para dotar a los usuarios de una aplicación eficiente para el análisis y descubrimiento de datos, capaz de impulsar el valor empresarial. Además, promueve el análisis de datos sin restricciones, ayudando a los usuarios a descubrir tendencias ocultas y hacer descubrimientos que impulsan a tomar decisiones. . (Ayala, J., et al., 2018)
- Thoughtspot: este sistema se basa en el uso de conjuntos de datos estructurados. También ha desarrollado una base de datos en memoria, esto permite que la información se almacene en la RAM para un procesamiento más rápido.

En la Tabla 3.1., se muestran las ventajas y desventajas principales de las 3 primeras herramientas líderes previamente mencionadas:

Tabla 3.1. Cualidades y debilidades de las herramientas líderes.

HERRAMIENTA	CUALIDADES	DEBILIDADES
Tableau	• Capacidad de análisis de datos con acceso a una variedad de fuentes de datos • Interfaz de usuario fácil de usar • Análisis visual con inteligencia integrada • No necesario grandes conocimientos de programación • Facilita la integración analítica de información	• Integración y modelización de daos tiene lugar en el cliente escritorio • Falta de soporte de navegación guiada • Riesgo de “explosión de informe”
Microsoft BI	• Ofrece productos con capacidades para informes con formato, informes ad hoc, análisis y cuadros de mando sin tener que incurrir en cargos de licencia adicionales	• Existen dependencias entre las versiones de las diferentes líneas de productos (SQL Server, Office...) • Esos productos no tienen repositorio

	• SQL Server como sistema de bases de datos • Excel llamado el BI de autoservicio y herramienta de análisis • Amplia red de socios comerciales, proporcionando a los clientes las competencias de Microsoft	central de metadatos • Clientes confusos por la existencia de herramientas superpuestas • Baja satisfacción del producto
QlikView	• Brinda capacidades de autoservicio (fácil uso) • Tecnología asociativa en memoria que permite a los usuarios analizar y procesar datos rápidamente • Preservar todas las asociaciones entre los campos que se analizan • No se precisa de gran tiempo para ver resultados • Exploración de datos instantáneamente y en tiempo real • Traduce grandes cantidades de datos complejos de múltiples fuentes • Procesamiento en memoria capaz de comprimir datos y optimizar la potencia • No se requiere almacén de datos	• Separación limitada de datos de la estructura propietaria • Problemas de consistencia de datos cuando se crean varios modelos • Dependencia de scripts a la hora de realizar cargas iniciales • Costo de mantenimiento alto

Fuente: Elaboración propia a partir de Ayala, J. et al. (2018).

4. FUNCIONES DE BI Y TOMA DE DECISIONES

El apoyo a la toma de decisiones, como un concepto tradicional en la gestión, ha tenido un papel vital en la competitividad o la supervivencia de las empresas y, en la actualidad, el BI, como una impresión moderna, tiene también varias contribuciones en el relación con este proceso.

Según Rouhani et al. (2016), el BI, como un termino general, constituye un conjunto de tecnologías como el *Data Warehouse*, *Data Mining*, *OLAP*, sistemas de apoyo a la toma de decisiones, balances de control...para mejorar el flujo de trabajo y el proceso de toma de decisiones. Además, se cree que esta tecnología es una de las áreas líderes en información tecnológica y es una de las prioridades para muchos ejecutivos.

Revisando el concepto de la toma de decisiones, Rouhani et al, (2016) quisieron hacer un estudio para ver si realmente las funciones de BI tenían algo que ver con los beneficios obtenidos en la toma de decisiones y a nivel organizacional. Por ello, para comprobarlo, decidieron separar los beneficios en dos niveles diferentes, incluyendo los beneficios de la toma de decisiones y los beneficios de la empresa. Entendemos los beneficios del apoyo a la toma de decisiones como os principales beneficios que son el resultado de este proceso de toma de decisiones. Por otro lado, los beneficios de la empresa son el resultado a largo plazo.

De todos los criterios que establecieron como funciones de BI, los agruparon en seis principales incluyendo: apoyo analítico e inteligente a la toma de decisiones, proporción de experimentos relacionados e integrados con la información ambiental, optimización y modelo recomendado, razonamiento, mejora de herramientas para la toma de decisiones; y la satisfacción de las partes interesadas.

Los beneficios del apoyo a la toma de decisiones los dividieron en tres: mejor conocimiento del proceso, reducción del tiempo de decisión y reducción del coste de decisión. Los tres criterios serán influidos por las funciones de BI.

Los beneficios de la empresa los separaron en decisión efectiva, ventaja competitiva y satisfacción de las partes interesadas, ya que son realmente los resultados de la toma de decisiones a un nivel organizativo.

La toma de decisiones estratégica, como una actividad fundamental, tiene un intenso impacto humano, financiero y organizativo en la empresa. (Rouhani et al., 2016).

En resumen, este estudio concluyó con el resultado de que las funciones anteriormente citadas tenían relación con los beneficios de la toma de decisiones y de la organización, ya que tenían un impacto sobre ellas. Resulta imprescindible detectar qué funciones son las que más importancia tienen para nuestra organización, así como tener un plan claro y bien estructurado para ver qué funciones tienen mayor impacto en nuestros beneficios (y por el contrario cuales no son inútiles). También, son herramientas que, si son utilizadas correctamente, con decisiones efectivas nos proporcionarán ventajas competitivas importantes, valores que mayor coeficiente de relación tenían.

Cabe destacar que realizar el proceso de la toma de decisiones de una manera efectiva, es una de las misiones más desafiantes a la que los directivos se enfrentan dentro de sus responsabilidades de gestión.

5. BUSINESS INTELLIGENCE EN LA ACTUALIDAD

Como ya se sabe, la pandemia del COVID-19 es uno de los virus más mortales que ha habido en el mundo.

Desde la propagación del virus, las empresas más esenciales y las que no son tan esenciales, no han tenido más opción que reducir su cuota de mercado, reducir personal, cerrar temporalmente y cruzar los brazos esperando ayudas del gobierno o arriesgarse al completo colapso de sus negocios. (A. O. Asare et al., 2020)

En España, el 80% de las empresas que han entrado en concurso en los seis meses siguientes a la propagación del virus, pertenecen a sectores directamente afectados o con una implicación destacable por el coronavirus, siendo las micro y pequeñas empresas las más afectadas. (Merca2, 2020).

Las diferentes empresas se han visto obligadas a implementar diferentes estrategias de BI para adaptarse rápidamente a los cambios, ya que cuánto más rápido puedan responder a la actual situación, más probabilidad tendrán de sobrevivir a corto y largo plazo. Algunas de estas herramientas son la gestión del riesgo, sistemas de inteligencia artificial y el análisis de datos que han servido para ayudar a tomar decisiones consistentes y así mejorar las operaciones comerciales en medio del COVID-19.

Entre otras cosas, el uso de herramientas de business intelligence ayuda a las empresas a ser adaptativas, flexibles e innovadoras para su supervivencia y seguir siendo competitivas durante estos tiempos desafiantes.

Considerablemente, los cambios actuales como resultado del COVID-19 hacen que sea necesario una revisión de las dimensiones de gestión del riesgo a través de, entre otras cosas, la integración de sistemas de inteligencia artificial y análisis de datos para ayudar a los directivos a tener diferentes percepciones entre lo que ocurre interna y externamente en la organización. El objetivo de tener diferentes percepciones y comparar lo que ocurre en el entorno de la empresa y lo que sucede dentro, es que se tomen las acciones necesarias, particularmente con flexibilidad y responsabilidad para mejorar el rendimiento empresarial.

Donde esta siendo muy útil la implementación de sistemas de BI o de inteligencia artificial es en el sector salud, ya que es uno de los sectores que entró en una de las mayores fases de cambio generado por la pandemia, además de ser un ámbito en el que no se deja de invertir e investigar.

Para la detección del COVID-19 se emplean dos tipos de pruebas estándar: la prueba viral que indica si el paciente está infectado y la prueba de anticuerpos que permite observar si el paciente ha pasado ya la enfermedad.

Aun que estas técnicas son refinadas, su sensibilidad en cuanto a las muestras es relativa. Donde surge el verdadero problema es a la hora de recolectar las pruebas, ya que se hace manualmente. A esto se le suma la cantidad de información que se requiere procesar, tales como datos relacionados con la rápida transmisión, etc. (Márquez, J., 2020).

Aquí entra en escena la aplicación de técnicas relacionadas con la inteligencia artificial o de BI, ya que ellas permiten realizar estudios pormenorizados en diferentes escalas estadísticas, imagenológicas y probabilísticas de información, que se condensan en los denominados sistemas de representación de datos (*dataset*). (Márquez, J., 2020).

Las herramientas de IA que han sido utilizadas son conocidas como, por un lado, Aprendizaje Automático (*Machine Learning*) que es una especializada en la que convergen en sistemas que trabajan con técnicas algorítmicas avanzadas para procesar diversos tipos de señales, optimizar modelos y procesar información masiva mediante el uso algorítmicos de regresión, clasificación y *clustering*. Por otro lado, esta el Aprendizaje Profundo (*Deep Learning*) para evaluar datos tipo imagen, video y audio empleando redes neuronales de memoria a corto y largo plazo, entre otras. Bajo este panorama, ha venido tomando fuerza el desarrollo de software de código abierto, donde la inteligencia colectiva es el engranaje principal para obtener un programa de altas prestaciones. (Márquez, J., 2020). El objetivo de esto es que los investigadores que trabajen con estas técnicas contribuyan a mejorar el sistema, perfeccionándolo hasta llegar a un producto cien por cien fiable y operativo.

El BI entra también en escena a través de uso de la analítica retrospectiva y descriptiva avanzada, ya que, a través de ellas se permite focalizar el estudio mediante diferentes indicadores y tendencias en el tiempo. Esto da lugar a predicciones futuras a través de las cuales se buscan correlaciones recurrentes por medio del cruce de información hasta llegar a alguna solución aceptable, que por lo general permitirá anticiparse a cualquier evento, pretendiendo estar mejor preparados a próximas pandemias o incluso poder prevenirlas.

En conclusión, lo que se puede rescatar de esta pandemia es que ha impulsado desarrollos tecnológicos en materia de inteligencia artificial y de negocios en diferentes áreas y sectores. Herramientas que se categorizan como fundamentales para ayudar, en este caso al sector salud, a controlar y gestionar el virus con cierto éxito.

6. CASO PRÁCTICO

Con el objetivo de poner en práctica lo expuesto, en este último punto se hará uso de estrategias de Minería de Datos para el análisis de la visualización de series de TV y películas de Netflix en el mundo dependiendo de diferentes factores, que llamaremos atributos, como puede ser el género, diferentes ratios de reseñas en plataformas de internet o el país de origen. Precisamente, se hará uso de la herramienta de minería de datos llamada Weka. La minería de datos, que se verá a continuación con mayor precisión, proporcionará a este estudio información sobre el *dataset* elegido ya que encuentra relaciones ocultas entre los atributos útiles para predecir comportamientos

futuros. Como se verá, la extensión y complejidad de la minería de datos es un hecho, y dentro de las diferentes técnicas existentes se hará uso de dos: la técnica de agrupación y la técnica de asociación o *clustering*.

6.1. MINERÍA DE DATOS

Sacar conocimiento de grandes cantidades de información ya es un reto. Por ello, surge la Minería de Datos o *Data Mining*, con el objetivo de buscarle sentido a la explosión de información que podemos almacenar.

Según Riquelme, J.C. et al. (2006), los datos en bruto raramente son beneficiosos de una manera directa. Su verdadero valor se basa en, por un lado, la habilidad de extraer información útil para ayudar a la toma de decisiones o para la exploración y, por otro lado, la comprensión de las fuentes de datos.

Anteriormente, el análisis de datos era un proceso manual, donde uno o más analistas familiarizados con los datos, con la ayuda de diferentes técnicas estadísticas, analizaban datos y generaban informes. Pero esto cambió, lógicamente, cuando el volumen de datos era tan grande que se escapaba de la capacidad humana para manipularlos manualmente.

Esto nos lleva a la necesidad de metodologías de análisis inteligente de datos, las cuales puedan descubrir conocimiento útil de estos. El termino KDD (*iniciales de Knowledge Discovery in Databases*) se refiere a todo el proceso de extracción del conocimiento a partir de una base de datos y marca un cambio de paradigma en el que lo importante es el conocimiento útil que se sea capaz de descubrir de los datos. (Riquelme, J.C. et al., 2006).

A menudo, los términos de Minería de Datos y KDD pueden ser confundidos por sinónimos, pero entenderemos la Minería de Datos como un paso fundamental dentro del proceso de extracción del conocimiento KDD.

6.1.1. Fases de la Minería de Datos

El proceso de KDD comentado anteriormente, consta de diferentes fases que analizaremos a continuación, pero cabe destacar una fase inicial que se trata de comprender el dominio de aplicación. Esto quiere decir que aquí se incluye los conocimientos previos relevantes y las metas de aplicación.

1. **Identificar los objetivos:** recogida de datos, evaluación de la calidad de los mismos y el uso de un análisis de exploración para familiarizarse con ellos.
2. **Preparación de los datos:** dentro de esta etapa hay diferentes fases como son la de limpieza, transformación, integración y reducción de datos.
 - Limpieza de datos: se deben eliminar el mayor numero de datos erróneos o inconsistentes e irrelevantes, detectar y corregir datos anómalos (*outliers*) y destapar valores desaparecidos (*data missing*).

- **Transformación:** en esta etapa se emplearán diferentes técnicas avanzadas como son la técnica de discretización y técnicas de aumento o reducción de la dimensión.
- 3. **Aplicación de técnicas de Minería de Datos:** es la fase fundamental del proceso y se llevará a cabo a partir del desarrollo de modelos predictivos y descriptivos y mediante el análisis de datos. Una vez se hayan recogido los diferentes datos que son de interés, un explorador establecerá el patrón a descubrir. En función de esto, elegiremos a técnica de minería de datos a emplear.
- 4. **Interpretación y análisis:** se deberán de interpretar o evaluar los diferentes patrones y comprarlos con los objetivos establecidos. Si fuese necesario, se volverá a las fases iniciales para una nueva iteración.
- 5. **Difusión y uso del conocimiento.** En esta fase se hace uso del conocimiento creado. Será importante que esté expresado de manera comprensible para que pueda llegar a muchos usuarios diferentes y ser difundido. Las herramientas utilizadas serán de visualización, presentación y transformación de datos.

6.1.2. Técnicas de la Minería de Datos

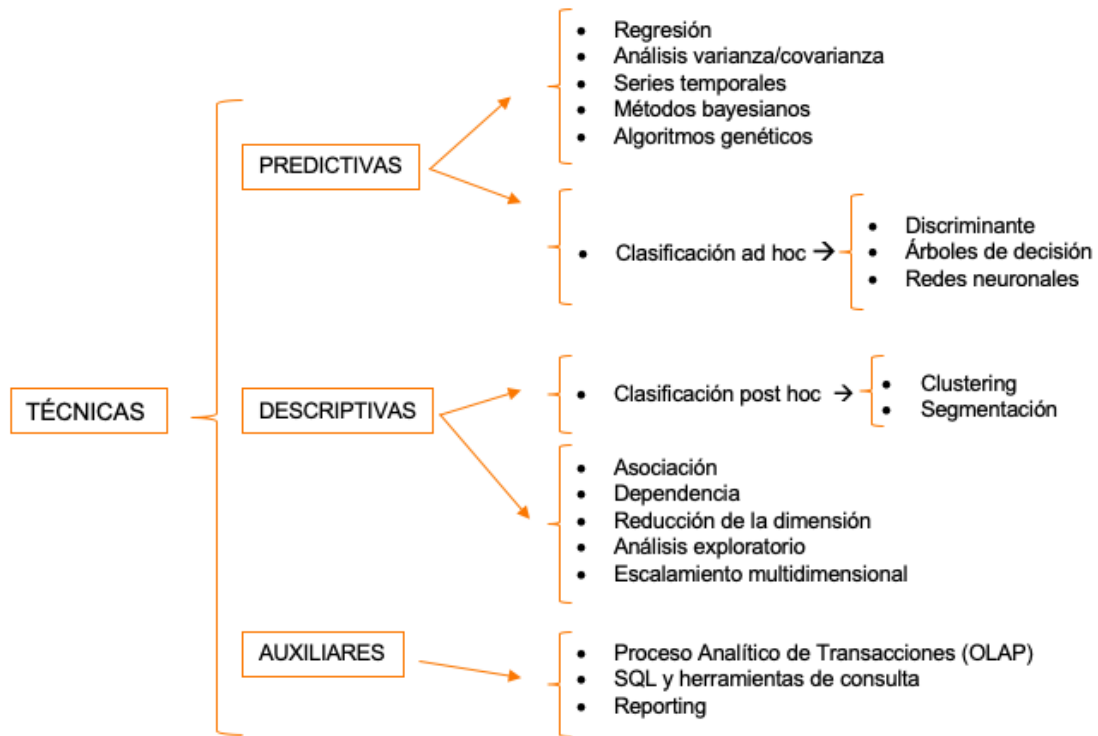
Una vez conocidas las diferentes etapas que componen la minería de datos, es importante analizar las distintas técnicas a aplicar en este proceso. Siendo tres fases, hay diferentes percepciones según Bayrak, T. (2015), y, las divide en técnicas descriptivas, prescriptivas y predictivas. Por otro lado, Pérez, C. y Santín, D. (2008), las dividen en técnicas predictivas, descriptivas y auxiliares.

Utilizando esta última percepción, las analizaremos a continuación:

- **Técnicas predictivas:** especifican el modelo para los datos en base a un conocimiento previo (Pérez, C. y Santín, D. , 2007). Además, este modelo debe ser contrastado antes de que sea aceptado como un modelo valedero. La aplicación de todo modelo debe superar diferentes fases, como son la identificación objetiva, estimación, diagnosis y predicción. A veces, el modelo se obtiene a partir de la mezcla del conocimiento obtenido antes y después de la aplicación de la minería de datos y, al igual que se explicaba antes, para darse como válido deberá ser comprobado.
- **Técnicas descriptivas:** el análisis descriptivo proporciona información significativa sobre el rendimiento empresarial y permite a los usuarios supervisar y gestionar mejor sus procesos empresariales haciendo uso de la minería de datos y el BI. (Bayrak, T., 2015). Al no asignar ningún papel predeterminado a las variables de estudio, no se supondrá la existencia de variables dependientes o independientes de los datos. Los modelos se crean de manera automática cuando se van reconociendo diferentes patrones.
- **Técnicas auxiliares:** son herramientas de apoyo más superficiales y con algunas limitaciones. Son métodos nuevos basados en técnicas estadísticas descriptivas, consultas e informes que, por lo general, suelen estar enfocados hacia la verificación.

A continuación, podemos ver en la Figura 6.1 una clasificación de las diferentes técnicas, haciendo referencia a las diferentes técnicas típicas que están dentro de las ya explicadas.

Figura 6.1. Técnicas de la Minería de Datos.



Fuente: Elaboración propia a partir de Pérez y Santín (2007).

Cabe destacar que, observando la Figura 6.1, se aprecia que las técnicas de clasificación pueden ser tanto técnicas predictivas (discriminante, árboles de decisión, redes neuronales) como descriptivas (*clustering*, segmentación). Las técnicas de clasificación predictivas se denominan técnicas de clasificación ad hoc ya que clasifican observaciones dentro de grupos que han sido definidos previamente. Las técnicas de clasificación descriptivas se denominan post hoc ya que realizan clasificaciones sin especificaciones previas.

6.2. DATOS DEL CASO PRÁCTICO

Para poder aplicar las técnicas de minería de datos, es necesario disponer de un conjunto de datos disponible para analizar y explotar. Se han extraído los datos de

Kaggle¹, plataforma online donde hay diferentes conjuntos de datos (*datasets*) proporcionados por una comunidad de científicos y profesionales de aprendizaje automático y que están disponibles para los usuarios con fines de investigación. Su fuente original es Shivam Mansal (2020).

En este trabajo, se ha utilizado una base de datos que consta de diferentes películas y series disponibles en la plataforma Netflix a partir del año 2019. El conjunto de datos se recopila de Flixable, que es un motor de búsqueda de Netflix de terceros. Esta *dataset* con 6056 registros y 11 atributos. En la Tabla 6.1. se describen los datos disponibles.

Tabla 6.1. Descripción de los atributos del caso práctico.

ATRIBUTO	DESCRIPCIÓN
Week	Variable que indica la semana en la que ha sido añadida
Show_type	Variable que indica el tipo de identificador (TV show, movie, documental)
Title	Variable que hace referencia al título de la serie o película
Ori_country	País de origen de la serie/película
Genre	Género de la serie/película
Releasedate	Fecha en la que fue añadida a Netflix
isNFori	Variable binaria sobre si es original de Netflix. Si lo es ("true"), si no ("false")
Imbdrating	Ratio de <i>Internet Movie Database</i> , una base de datos de reseñas de películas en internet
Rtrating	Ratio de Rotten Tomatoes, sitio web estadounidense de reseñas de cine y TV
Countrychart	País donde se ha visto la película/serie
Continent	Continente a donde pertenecen los que ven la película/serie

Fuente: Elaboración propia a través de Mansal (2020).

¹ Kaggle. Netflix Movies and TV Shows. <https://www.kaggle.com/shivamb/netflix-shows>

6.2.1. Pre-procesamiento de los datos

Como se ha comentado anteriormente, la Minería de Datos consta de diferentes fases. Una vez determinada la primera (identificación de los objetivos y recogida de datos) y analizados los diferentes atributos con los que cuenta la base de datos que disponemos, continuaremos con la siguiente: preparación o pre-procesamiento de los datos.

Como el dataset que se disponía era muy extenso se han seguido una serie de fases para adecuarlo a las necesidades pertinentes:

1. Primero se ha empezado con la integración de ficheros csv., se disponía de 80 ficheros, que se han integrado en un único fichero (con la opción de comandos `copy* .csv → nombrefichero.csv`).
2. Se ha convertido los porcentajes a número (se han quitado los caracteres de % para que fuese posible su lectura).
3. Se ha adaptado el atributo “fecha” para su adecuada lectura en la herramienta (incluyendo comillas al inicio y final de cada valor).
4. Para los valores *null* (desconocidos), se ha insertado el carácter “?”
5. Corrección de valores mal escritos (EURweek se ha cambiado por EUR con la opción de reemplazar).

6.2.2. Aplicación de técnicas de Minería de Datos.

Una vez seguidos los pasos anteriores, la base de datos estaba disponible para cargarla en Weka (pasando a la fase de utilización de herramientas de Minería de Datos) y la etapa de pre-procesamiento o preparación de los datos había finalizado.

Como se observa en la Figura 6.2, esto sería lo primero que se vería cuando abrimos la base de datos, teniendo ya información relevante.

La herramienta Weka muestra un conjunto de datos de 6056 registros y 11 atributos. Estos atributos se pueden analizar y obtener información estadística sobre ellos. Poniendo un ejemplo, en el atributo “imdbrating”, que es el ratio que hace referencia a una base de datos en línea que almacena información sobre películas, programas y series de televisión, etc. y recibe más de 100 millones de usuarios únicos al mes. La puntuación que dan en forma de reseñas va del 0% (valor mínimo) al 100% (valor máximo) se indica que el porcentaje máximo es del 94%, siendo la media del 34% y la desviación típica del 36%.

Esto en cuanto a una variable numérica, si la variable que se analiza es una nominal, como es el caso de “genre”, (género de la película, serie, documental...) se aprecia que hay 21 tipos diferentes de géneros y que el más común es la comedia seguido del drama.

En la Figura 6.3., aparecen las representaciones gráficas de los atributos mediante histogramas.

Figura 6.2. Carga del *dataset* en Weka.

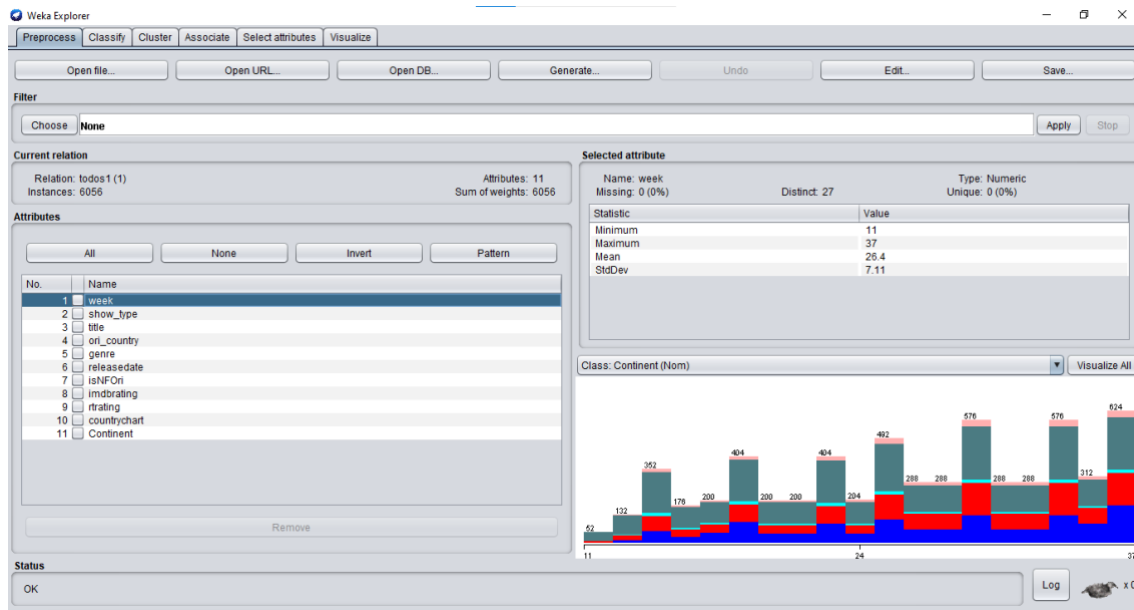
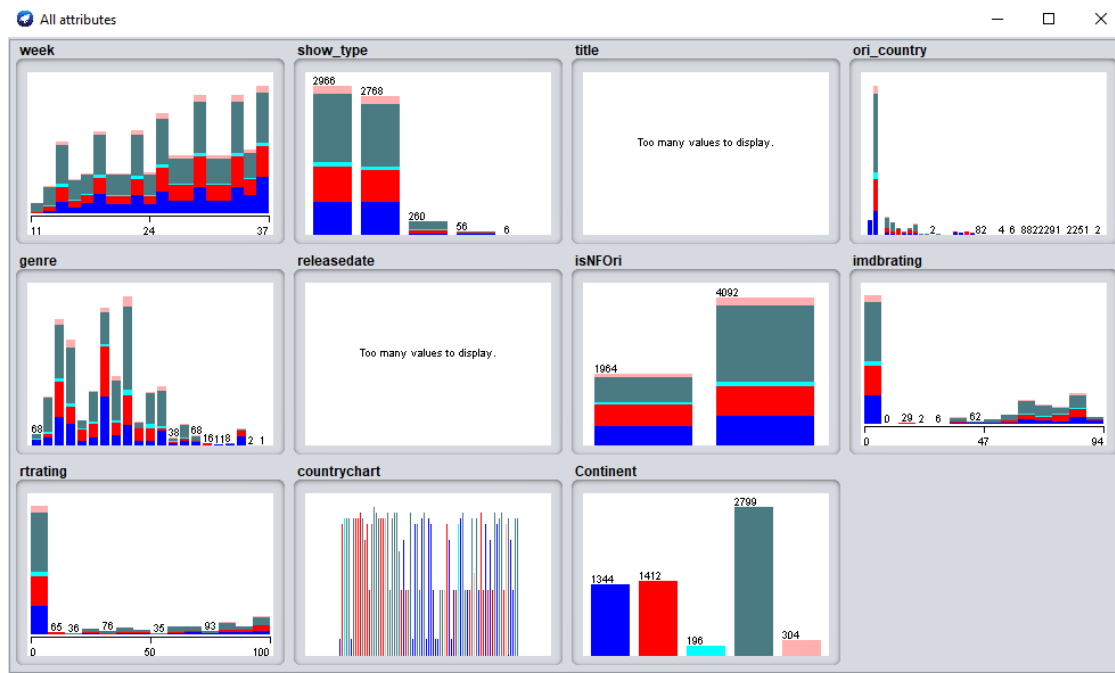
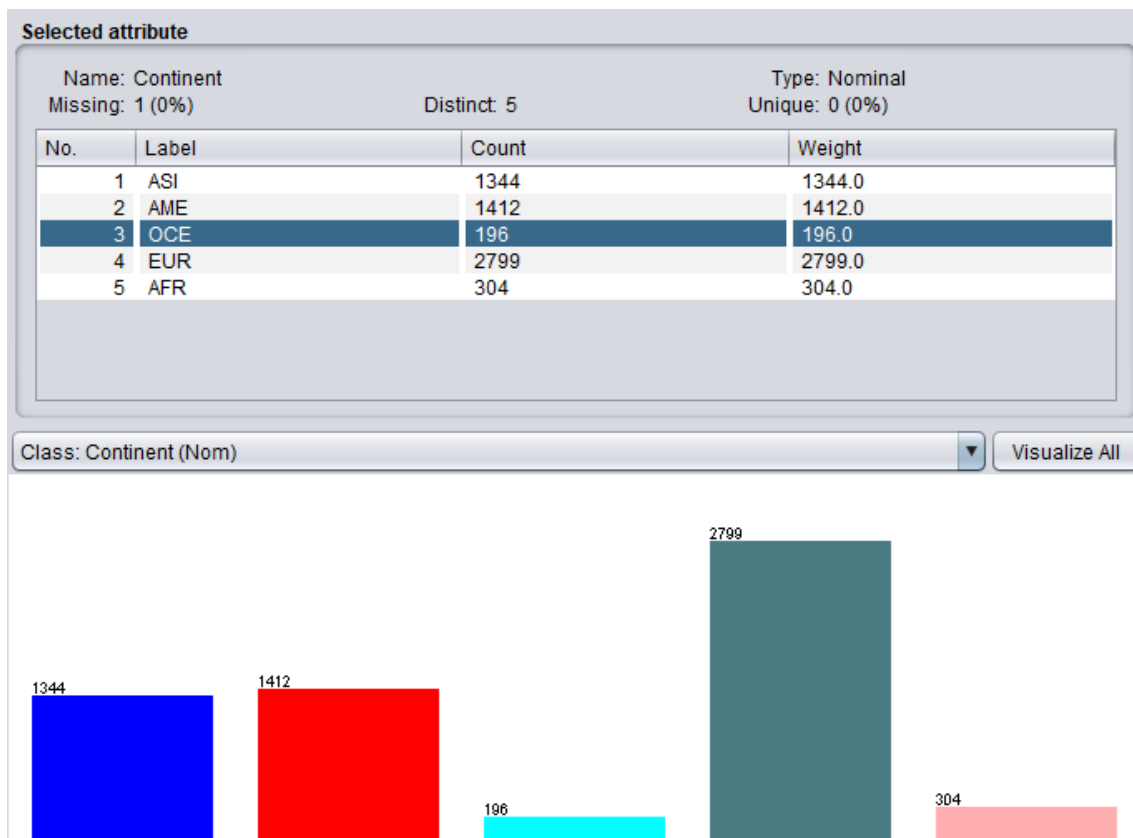


Figura 6.3. Representación de los 11 atributos mediante histogramas.



Poniendo otro ejemplo de una variable nominal, en este caso el atributo “continent”(continente en el que se ha visto el documento), se puede ver que hay cinco categorías diferentes (representando los cinco continentes). Además, desde el histograma que se observa en la Figura 6.4., se observa que el mayor uso de Netflix se hace en Europa (2799), mientras que el continente donde hay menor registro de uso es Oceanía (196).

Figura 6.4. Histograma del atributo “continent”.



A través de la una opción “*Edit*” (Figura 6.5), se puede ver la forma en la que los valores están estructurados y si existen datos erróneos o incorrectos para corregirlos si fuese necesario. En el caso de esta base de datos, se tiene valores desconocidos (*NULL*), que como se comentó previamente han tenido que ser editados para que Weka pudiese cargarlos.

Figura 6.5. Estructura del *dataset*.

No.	1: week	2: show_type	3: title	4: ori_country	5: genre	6: releasedate	7: isNFOri	8: imdbrating	9: rrating	10: countrychart	11: Continent
	Numeric	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Numeric	Numeric	Nominal	Nominal
1	37.0	Movie	Alive	South Korea	Horror	08/09/2020	False	0.62	0.86	ARE	ASI
2	37.0	TV Show	Away	USA	Scien...	04/09/2020	True	0.71	0.73	ARE	ASI
3	36.0	Movie	Skys...	USA	Action	04/07/2018	False	0.61	0.48	ARE	ASI
4	36.0	TV Show	Lucif...	USA	Sup...	25/01/2016	True	0.83	0.87	ARE	ASI
5	35.0	Movie	Proj...	USA	Action	14/08/2020	True	0.61	0.63	ARE	ASI
6	35.0	TV Show	Lucif...	USA	Sup...	25/01/2016	True	0.83	0.87	ARE	ASI
7	37.0	Movie	The ...	USA	Anim...	28/07/2017	False	0.14	0.08	ARG	AME
8	37.0	TV Show	Cob...	USA	Action	02/05/2018	False	0.88	0.94	ARG	AME
9	36.0	Movie	The ...	USA	Anim...	28/07/2017	False	0.14	0.08	ARG	AME
10	36.0	TV Show	Cob...	USA	Action	02/05/2018	False	0.88	0.94	ARG	AME
11	35.0	Movie	The ...	Argentina	Crime	20/08/2020	False	0.43		ARG	AME
12	35.0	TV Show	Lucif...	USA	Sup...	25/01/2016	True	0.83	0.87	ARG	AME
13	34.0	Movie	Proj...	USA	Action	14/08/2020	True	0.61	0.63	ARG	AME
14	34.0	TV Show	The ...			14/08/2020	False			ARG	AME
15	33.0	Movie	Won...	USA	Drama	13/11/2017	False	0.8	0.93	ARG	AME
16	33.0	TV Show	Lock...	Spain	Drama	20/04/2015	True	0.83	1.0	ARG	AME
17	32.0	Movie	Won...	USA	Drama	13/11/2017	False	0.8	0.93	ARG	AME
18	32.0	TV Show	Lock...	Spain	Drama	20/04/2015	True	0.83	1.0	ARG	AME
19	31.0	Movie	The ...	United Kin...	Rom...	24/07/2020	True	0.62	0.28	ARG	AME
20	31.0	TV Show	Dark...	Mexico	Drama	15/07/2020	True	0.75		ARG	AME
21	30.0	Movie	MILF	France	Com...	02/05/2018	False		0.14	ARG	AME
22	30.0	TV Show	Dark...	Mexico	Drama	15/07/2020	True	0.75		ARG	AME
23	29.0	Movie	The ...	USA	Sup...	10/07/2020	True	0.82	0.8	ARG	AME
24	29.0	TV Show	Dark...	Mexico	Drama	15/07/2020	True	0.75		ARG	AME
25	28.0	Movie	Ame...	USA	Action	14/09/2017	False	0.7	0.35	ARG	AME
26	28.0	TV Show	Dark	Germany	Mystery	01/12/2017	True	0.87	0.94	ARG	AME
27	27.0	Movie	Ame...	USA	Action	14/09/2017	False	0.7	0.35	ARG	AME
28	27.0	TV Show	Dark	Germany	Mystery	01/12/2017	True	0.87	0.94	ARG	AME
29	26.0	Movie	Lost...	France	Crime	19/06/2020	True	0.62		ARG	AME

6.3. APLICACIÓN DE TÉCNICAS DE ASOCIACIÓN

Una vez que se ha abierto el *dataset* disponible en la herramienta de Weka y se ha visto que nos puede proporcionar información relevante, se ha de utilizar alguna de las técnicas disponibles para extraer conocimiento útil que nos permita sacar conclusiones de los datos para la toma de decisiones.

Se comenzará por el uso de la Técnica de Asociación, que se evalúa en función del soporte y la confianza de los datos y se utiliza para encontrar hechos que ocurren en común dentro de un conjunto de datos. (Camana, R.G., 2016).

A través de la pestaña *Associate* que se puede ver en el panel superior de la aplicación Weka (cuarta pestaña), se pueden usar algoritmos de asociación. Es importante reseñar que este método sólo funciona con datos nominales, por ello, será necesario discretizar previamente todos los atributos numéricos. (García Morate, s.f). Además, para mejorar la eficiencia del algoritmo, elimina aquellos atributos que tengan sus valores desconocidos. (Molina, J.M y García, J., 2006).

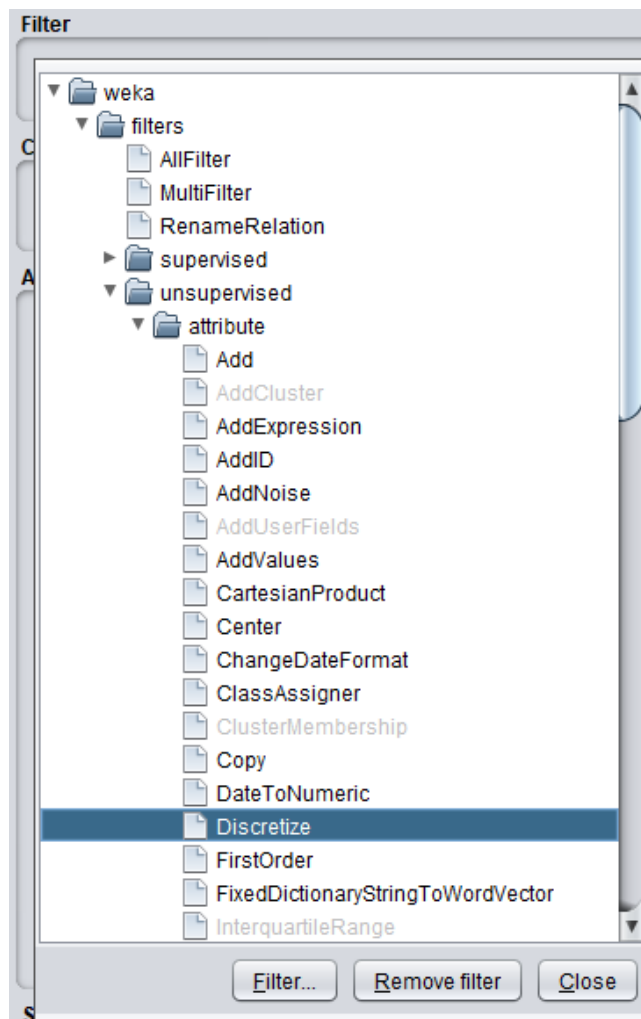
Dentro de esta pestaña, el algoritmo que se utilizará será el de A priori (una vez discretizados los atributos), algoritmo más destacado que tiene Weka y uno de los primeros desarrollados para la búsqueda de reglas de asociación. Las dos etapas que tiene este algoritmo son:

- Identificar todos los itemsets que ocurren con una frecuencia por encima de un límite.

- Identificar esos itemsets en reglas de asociación.

Para discretizar los valores que se ha comentado previamente, se tendrá que volver a la pestaña inicial de *Preprocess*. Una vez allí, se elegirá el filtro *unsupervised* (no supervisado) dando lugar a atributos discretizados en *bins*.

Figura 6.6. Aplicación de filtros para discretizar los atributos.



Una vez se ha realizado esto, volviendo a la pestaña de *Associate*, se puede aplicar ya el algoritmo de asociación *Apriori*.

Lo que se obtiene de este proceso son diferentes reglas útiles, cambiando el valor de "*minMetric*" (nivel de confianza) al 0.7 y se han pedido 10 reglas para evaluar:

- De las 1252 series vistas en Europa, 1158 son originales de Netflix.
- 1474 series producidas en E.E.U.U de las cuales 1318 son originales de Netflix.
- De las 2768 de las series que están dentro del *dataset*, 2450 son originales de Netflix.

- De 1964 documentos presentes en la base de datos y no originales de Netflix, 1634 son películas; y de éstas 912 han sido producidas en E.E.U.U.
- De los 1913 diferentes métodos de entretenimiento disponibles en el dataset de estudio producidos en Estados Unidos y vistos en Europa, 1536 son originales de Netflix.
- Existen 2799 archivos vistos en Europa de los cuales 2100 son originales de Netflix.
- De las 1252 diferentes series más vistas en Europa, 922 son procedentes de Estados Unidos.
- Y, por último, de 1332 películas originales de Netflix, el país de origen de 976 es E.E.U.U.

Figura 6.7. Aplicación del algoritmo de asociación *Apriori*.

```

Apriori
=====

Minimum support: 0.15 (908 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 13

Size of set of large itemsets L(2): 15

Size of set of large itemsets L(3): 6

Best rules found:

1. show_type=TV Show Continent=EUR 1252 ==> isNFOri=True 1158    <conf:(0.92)> lift:(1.37) lev:(0.05) [312] conv:(4.27)
2. show_type=TV Show ori_country=USA 1474 ==> isNFOri=True 1318    <conf:(0.89)> lift:(1.32) lev:(0.05) [322] conv:(3.04)
3. show_type=TV Show 2768 ==> isNFOri=True 2450    <conf:(0.89)> lift:(1.31) lev:(0.1) [579] conv:(2.81)
4. ori_country=USA isNFOri=False 1072 ==> show_type=Movie 912    <conf:(0.85)> lift:(1.74) lev:(0.06) [386] conv:(3.4)
5. isNFOri=False 1964 ==> show_type=Movie 1634    <conf:(0.83)> lift:(1.7) lev:(0.11) [672] conv:(3.03)
6. ori_country=USA Continent=EUR 1913 ==> isNFOri=True 1536    <conf:(0.8)> lift:(1.19) lev:(0.04) [243] conv:(1.64)
7. week='(-inf-20.5]' 1516 ==> isNFOri=True 1190    <conf:(0.78)> lift:(1.16) lev:(0.03) [165] conv:(1.5)
8. Continent=EUR 2799 ==> isNFOri=True 2100    <conf:(0.75)> lift:(1.11) lev:(0.03) [208] conv:(1.3)
9. show_type=TV Show Continent=EUR 1252 ==> ori_country=USA 922    <conf:(0.74)> lift:(1.23) lev:(0.03) [174] conv:(1.52)
10. show_type=Movie isNFOri=True 1332 ==> ori_country=USA 976    <conf:(0.73)> lift:(1.23) lev:(0.03) [180] conv:(1.5)

```

De esta información, es interesante que la mayoría de series disponibles no sólo en la base de datos sino también en la plataforma Netflix son originales. Esto se debe a la creciente apuesta de Netflix por sacar series, películas y documentales propios. También es de resaltar la gran producción estadounidense de películas y series que, además, la mayoría son vistas en Europa. Esto se debe a que Estados Unidos está a la cabeza de productores de contenido audiovisual, para ser más concretos, el segundo detrás de Japón con más de 5600 títulos disponibles en 2018.

Se puede destacar que el número cuando se menciona series es superior a cuando se menciona películas ya que la producción de series se ha triplicado casi desde 2010 y, por otra parte, la producción de películas ha disminuido considerablemente.

“Netflix tiene muy claro las preferencias audiovisuales de su audiencia. Los datos les dicen el tipo de programa que sería interesante que hicieron. Hacen el producto que su algoritmo y sus datos les dicen que su audiencia quiere”. (Bejerano, P.G., 2020).

6.4. APLICACIÓN DE TÉCNICAS DE AGRUPAMIENTO O CLUSTERING

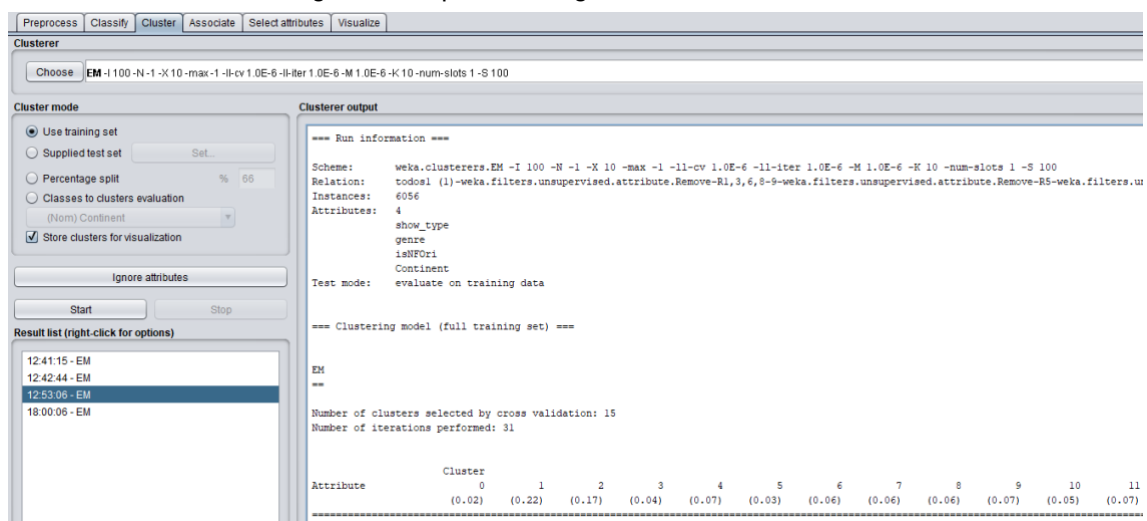
Otra de las técnicas de las que se va a hacer uso es la Técnica de Clustering o Agrupamiento. Esta técnica permite la identificación de tipologías o grupos donde los elementos guardan cierta similitud, es decir, donde los atributos son completa o prácticamente homogéneos. Es un método muy fácil de combinar con otro tipo de técnicas y dar lugar a sistemas híbridos debido a su naturaleza flexible.

Para ver cómo de homogéneos son los datos del estudio, se ha modificado la base de datos reduciendo el número de atributos.

Los atributos que se han querido analizar son 4 (“*show_type*”, “*genre*”, “*isNFOri*”, “*Continent*”).

Desde la pestaña de *Cluster*, se ha seleccionado el algoritmo a aplicar. En este caso ha sido el algoritmo EM (*Expectation Maximization*), el cual empieza adivinando los parámetros de las distribuciones y, a continuación, los utiliza para calcular las probabilidades de que cada parámetro pertenezca a un *cluster*. (Molina, J.M y García, J., 2006).

Figura 6.8. Aplicación algoritmo EM.



Al salir un número muy elevado de *clusters* (15), se comentarán aquellos con mayor probabilidad. Esto es el caso de los *clusters* 1,2 y 11.

Como se puede ver en la figura 6.9., el *cluster* 1 (22%) indica que la mayoría eran series de televisión relacionadas con el mundo de los súper héroes. A su vez, la mayoría eran series originales de Netflix y el continente donde más se veían era Europa.

Por otro lado, en el *cluster* 2 (17%) se puede apreciar que la mayoría eran películas de comedia, coincidiendo con el *cluster* 1 en que son originales de Netflix y el continente donde más vistas han sido es Europa.

Figura 6.9. Evaluación *clusters 1 y 2*.

Fantasy	1.0508	275.6437	1.183
Adventure	1.1183	1	1.2175
Reality-Show	1.0332	95.6324	1.0047
Thriller	2.0911	1.0021	22.3716
Animation	1.0271	1.0302	1.0009
Movie	1.0854	1	1.1799
War	1.0192	1	1.156
Unknown	1.1286	1.0005	1.215
Musical	1.0075	1	1.1121
Situation	1.0395	1	1.0555
[total]	125.9914	1327.936	1024.535
isNFORi			
False	15.4771	1.5396	194.4558
True	91.5143	1307.3964	811.0792
[total]	106.9914	1308.936	1005.535
Continent			
ASI	10.1455	188.2589	68.7469
AME	80.2681	32.5557	172.6027
OCE	2.9066	45.0057	34.3842
EUR	9.1306	931.9417	670.4216
AFR	7.5406	114.174	62.3796
[total]	109.9914	1311.936	1008.535

Attribute	Cluster		
	0 (0.02)	1 (0.22)	2 (0.17)
=====			
show_type			
Movie	102.2609	2.652	1000.0401
TV Show	4.6772	1306.2632	5.4388
Documentary TV	1.0377	1.0093	1.0441
Documentary	1.0118	1.0114	1.0014
Short	1.0038	1	1.0105
[total]	109.9914	1311.936	1008.535
genre			
Horror	1.0853	1.1255	1.2174
Science Fiction	2.809	106.4638	67.5979
Action	40.0772	81.0389	64.1701
Superhero	21.9998	336.0307	121.8541
Animated	1.294	2.1259	2.3142
Crime	14.9137	48.9005	19.1771
Drama	3.1475	190.0239	3.8758
Romance	14.5018	1.0019	158.5218
Comedy	12.4815	133.751	551.2998
Mystery	1.0342	36.1006	1.0043
Documentary	1.0468	12.0643	1.0065
Fantasy	1.0508	275.6437	1.183
Adventure	1.1183	1	1.2175
Reality-Show	1.0332	95.6324	1.0047
Thriller	2.0911	1.0021	22.3716

Por último, tenemos el *cluster* 11 (7%). En la Figura 6.10. se observa que la mayoría de películas fueron de romance. Esta vez no originales de Netflix y viéndose en Europa.

Se puede destacar que en los tres *clusters* el continente donde más visto ha sido era Europa. Esto se corrobora con que de los 10 países donde más se utiliza Netflix, 7 de ellos están en Europa.

Podemos concluir que, aunque las probabilidades y los números de *clusters* indiquen que la agrupación es difícil ya que los datos no son homogéneos, una cosa en común que tienen los *clusters* con más probabilidad (1 y 2) aparte de lo ya comentado anteriormente es que son originales de Netflix. Esto se debe a la gran apuesta de Netflix por el contenido original, siendo dependiente (para lo bueno y para lo malo) del mismo debido al gran auge de creación de diferentes plataformas (*Disney +*, *Prime Video...*). Tanto es así que en 2018 estrenaron 240 series, películas y documentales propios, cifra que aumentó hasta los 371 en 2019. (Millán, 2020).

Figura 6.10. Evaluación *cluster* 11.

9 (0.07)	10 (0.05)	11 (0.07)	12 (0.05)			
				1.1551	3.7264	36.3083
439.8102	15.4382	447.6535	1.025	5.2842	1.0458	16.983
6.1154	263.5288	1.3903	6.0971	1.0026	3.4321	1.0003
1.5801	4.2744	1.002	256.8945	4.0723	1.0476	12.1943
1.0036	1.0266	1.0019	56.6934	1.0002	1.0057	1
1.2201	1.9135	2.4238	1.0002	2.0404	1.036	4.1788
449.7293	286.1814	453.4715	321.7101	1.1043	1.0259	4.5113
				6.7849	54.5635	7.8894
2.4551	1.0449	29.756	1.0014	1.0475	1.0131	2.6549
2.7275	1.0559	38.5558	1.0209	1.1027	1.0073	1.3413
114.036	56.9743	15.0393	1.0201	465.7293	302.1814	469.4715
16.5768	1.0661	28.6483	1.0216			
81.5014	19.9698	33.933	1.0092	441.4902	279.5205	447.2784
22.5136	39.4146	15.3679	1.0221	5.239	3.661	3.1931
85.6691	4.1481	35.8679	1.0204	446.7293	283.1814	450.4715
40.754	1.0551	171.8004	1.003			
72.8977	106.5014	10.4394	1.0262	20.699	57.9519	10.4821
1.0002	1.012	1	1.017	376.0979	119.8648	6.1301
1.0038	1.0363	1.0019	317.4955	14.8884	4.4355	32.7463
1.1551	3.7264	36.3083	1.0174	27.9869	94.5637	376.8343
5.2842	1.0458	16.983	1	10.057	9.3656	27.2788
1.0026	3.4321	1.0003	1.0183	449.7293	286.1814	453.4715
4.0723	1.0476	12.1943	1.003			

7. CONCLUSIONES

A lo largo de este trabajo, mediante el análisis de lo que compone el concepto de BI, ha quedado reflejada la importancia que tiene para las organizaciones y la sociedad en general. Es importante elegir bien la estrategia y la implementación, ya que siendo efectuada de manera correcta será muy beneficiosa y útil para tratar las grandes cantidades de información y datos de los que disponen las empresas hoy en día. Es un hecho que la información forma parte del patrimonio de la empresa y será de vital importancia la velocidad en el análisis de los datos, así como la obtención de respuestas rápidas para hacer efectiva la toma de decisiones que será ayudada por los sistemas de BI. Tanto es así que permite obtener una ventaja competitiva y, por el contrario, quienes no se adecuen a los cambios tecnológicos actuales, estarán corriendo el riesgo de quedarse obsoletos, ya que, el implemento de estrategias de BI que hace unas décadas se veía como una opción, más bien hoy en día es una obligación. Cabe destacar que, como se ha visto en ejemplos durante el estudio del BI, es aplicable a todos los sectores e industrias.

Por otro lado, mediante la aplicación de herramientas de minería de datos a través de la aplicación Weka, se ha podido corroborar las facilidades y ayuda que puede suponer el uso del BI para las empresas. En el caso práctico en particular realizado en este trabajo, se han utilizado dos técnicas de las muchas disponibles en minería de datos. Estas dos técnicas son la técnica de agrupamiento o *clustering* y la técnica de asociación. Ambas técnicas han coincidido en que la mayoría de archivos disponibles en Netflix son originales, hecho que es beneficioso para la compañía y que tiene que seguir apostando por ello. Por otro lado, Europa es uno de los continentes que más uso hace de esta plataforma en streaming y el contenido estadounidense es el que más se visualiza.

Para finalizar, he de destacar que a pesar de la amplitud del dataset utilizado (6056 instancias y 11 atributos) se ha podido obtener información relevante y conocimiento útil.

BIBLIOGRAFÍA

- Agencias. (2020, octubre 23). El 80% de empresas que han entrado en concurso pertenece a sectores más afectados por Covid. Merca2.es. <https://www.merca2.es/empresas-covid-19-empresas/>
- Asare, A., Addo, P., Sarpong, E. and Kotei, D. (2020) COVID-19: Optimizing Business Performance through Agile Business Intelligence and Data Analytics. *Open Journal of Business and Management*, **8**, 2071-2080. doi: [10.4236/ojbm.2020.85126](https://doi.org/10.4236/ojbm.2020.85126).
- Ayala, J., Ortiz, J., Guevara, C., & Maya, E. (2018). Herramientas de Business Intelligence (BI) modernas, basadas en memoria y con lógica asociativa. *revistapuce*. <https://doi.org/10.26807/revpuce.v0i106.144>
- Bayrak, T. (2015). A Review of Business Analytics: A Business Enabler or Another Passing Fad. *Procedia - Social and Behavioral Sciences*, 195, 230-239. <https://doi.org/10.1016/j.sbspro.2015.06.354>
- Bejerano, P. G. (2020, octubre 18). Cinco años de Netflix: El impacto en la producción audiovisual española. EL PAÍS RETINA. https://retina.elpais.com/retina/2020/10/16/innovacion/1602843710_314589.html
- Business Intelligence: Cuadrante mágico de Gartner 2019. ThoughtSpot, un nuevo competidor entre los líderes.* (s. f.). Recuperado 11 de noviembre de 2020, de <https://www.inforges.es/post/business-intelligence-cuadrante-magico-de-gartner-2019-thoughtspot-un-nuevo-competidor-entre-los-lideres>
- Camana, R. G. (2016). Potenciales Aplicaciones de la Minería de Datos en Ecuador. *Revista Tecnológica - ESPOL*, 29(1). Recuperado a partir de <http://www.rte.espol.edu.ec/index.php/tecnologica/article/view/464>
- CESAR, P. L., & DANIEL, S. G. (2007). *Minería de datos. Técnicas y herramientas: técnicas y herramientas*. Editorial Paraninfo.

Estos son los países en los que más se ve Netflix. (s. f.). GQ España. Recuperado 26 de noviembre de 2020, de <https://www.revistagq.com/noticias/cultura/articulos/netflix-paises-en-los-que-mas-se-ve-series-mas-vistas/30205>

Forbes, I. (2019, octubre 23). *¿Por qué no debes confundir Big Data, Business Intelligence e Inteligencia artificial?* • Forbes México. Forbes México. <https://www.forbes.com.mx/por-que-no-debes-confundir-big-data-business-intelligence-e-inteligencia-artificial/>

Fraga, A. I. (2017, octubre 7). Big Data y Business Intelligence, ¿son lo mismo? ¿en qué se... *TICbeat*. <https://www.ticbeat.com/tecnologias/big-data-y-business-intelligence-son-lo-mismo-en-que-se-diferencian/>

Galici, R., Ordile, L., Marchesi, M., Pinna, A., & Tonelli, R. (2020). Applying the ETL Process to Blockchain Data. Prospect and Findings. *Information*, 11(4), 204. <https://doi.org/10.3390/info11040204>

García Morate, D. (s. f.-a). *MANUAL DE WEKA*. http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf

Group, I. D. M. (2020, marzo 25). *Cómo usar la inteligencia empresarial para mejorar los procesos | Predictive Analytics. Discover The New*. <https://discoverthenew.ituser.es/predictive-analytics/2020/03/como-usar-la-inteligencia-empresarial-para-mejorar-los-procesos>

Laudon, K. C., & Laudon, J. P. (2014). *Management Information Systems: Managing the Digital Firm* (Thirteenth Edition). Pearson.

Los 10 países con el mayor catálogo en Netflix. (2018, junio 27). Expansión. <https://expansion.mx/tecnologia/2018/06/27/los-10-paises-con-el-mayor-catalogo-para-ver-en-netflix>

M. Hemsley, L. (s. f.). *BUSINESS INTELLIGENCE TOOLS IN A DEVELOPMENTAL ENVIRONMENT: AN ACADEMIC MODULE*.

- Márquez, J. (2020). Inteligencia Artificial y Big Data como soluciones frente al COVID-19. *Revista de Bioética y Derecho*, 0(50), 315-331. <https://doi.org/10.1344/rbd2020.50.31643>
- Medina La Plata, E. H. (2012). *Business Intelligence: Errores comunes en su implementación*. <http://hdl.handle.net/10757/333710>
- Millán, V. (2020, julio 15). *Netflix y su avalancha de contenido original: ¿Es cada vez peor?* Hipertextual. <https://hipertextual.com/2020/07/netflix-contenido-original-peor>
- Molina López, J. M., & García Herrero, J. (2006). *TÉCNICAS DE ANÁLISIS DE DATOS: Aplicaciones prácticas utilizando Microsoft Excel y Weka*. http://matema.ujaen.es/jnavas/web_recursos/archivos/weka%20master%20recursos%20naturales/apuntesAD.pdf
- Moss, L. T., & Atre, S. (2003). *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision Support Applications*. Pearson.
- Murugesan, M., & Karthikeyan, K. (2016). *Business Intelligence Market Trends and Growth in Enterprise Business*.
- Netflix Movies and TV Shows*. (2020). Recuperado 20 de noviembre de 2020, de <https://kaggle.com/shivamb/netflix-shows>
- Nuevas Tendencias en Business Intelligence: Del Big Data al Social Intelligence. (2013). *www.stratebi.com*. http://www.stratebi.es/todobi/May13/Nuevas_Tendencias_BI.pdf
- Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). *Minería de Datos: Conceptos y Tendencias*. <https://idus.us.es/handle/11441/43290>
- Rodríguez Suárez, Yuniét y Díaz Amador, Anolandy y (2009), "Herramientas de Minería de Datos." *Revista Cubana de Ciencias Informáticas*, Vol. 3, núm.3-4, pp.73-80 [Consultado: 30 de Noviembre de 2020]. ISSN: 1994-1536. Disponible en : <https://www.redalyc.org/articulo.oa?id=3783/378343637009>

- Rouhani, S., Ashrafi, A., Zareravasan, A., & Afshari, S. (2016). The impact model of business intelligence on decision support and organizational benefits. *Journal of Enterprise Information Management*.
https://www.researchgate.net/publication/277003403_The_impact_of_business_intelligence_on_decision_support_and_organizational_benefits
- Solano, L. E. S. (2017). Business Intelligence: un balance para su implementación. *InnovaG*, 3, 27-36.
- Tapia, J. A. P., Palacios, M. M. T., Medina, E. H., & Crespo, J. D. O. (2020). Business Intelligence aplicado al sector Salud. *Revista Arbitrada Interdisciplinaria Koinonía*, 5(Extra 3), 622-650.
- ThoughtSpot promete una experiencia de BI similar a Google. (s. f.). *small-business-tracker.com*. <https://spa.small-business-tracker.com/thoughtspot-promises-google-like-bi-experience-736099>
- Tripathi, A., & Bagga, T. (2020). *Leading Business Intelligence (BI) Solutions and Market Trends* (SSRN Scholarly Paper ID 3568414). Social Science Research Network.
<https://papers.ssrn.com/abstract=3568414>
- Trujillo: «España es uno de los mejores actores en el campo de Business Intelligence». (2020, febrero 4). *Economía3*. <https://economia3.com/2020/02/04/247729-trujillo-lucentia-lab-espana-es-uno-de-los-mejores-actores-en-el-campo-de-bi/>
- Villamizar Gutiérrez, L. A. (2010). Cómo abordar un proyecto de Business Intelligence en una empresa u organización [EAFIT Medellín].
https://repository.eafit.edu.co/bitstream/handle/10784/411/LeonelAlfonso_VillamizarGutierrez_2010.pdf?sequence=1&isAllowed=y