



CAMPUS
DE EXCELENCIA
INTERNACIONAL



Graduado en Ingeniería Informática

Universidad Politécnica de Madrid

Escuela Técnica Superior de
Ingenieros Informáticos

TRABAJO FIN DE GRADO

Diseño e implementación de una herramienta de análisis de datos y automatización de procesos para mercados financieros

Autor: Antonio Fernández Pita

Director: Ernestina Menasalvas Ruiz

MADRID, JUNIO DE 2017

A mi familia, novia, amigos y
en especial a mi padre que
me animo a empezar este
grado en ingeniería informática.

Agradecimientos.

Mi especial agradecimiento a Ernestina Menasalva por la excepcional dirección y ayuda en el proyecto y todo ello a pesar de tener una agenda muy apretada.

Muchas gracias.

Abstract:

In this final degree project, a research about algorithmic trading as well as the learning of the data management is presented. A tool has been developed for that purpose and a data base has been deployed. Both are going to work together in the study of the financial markets.

To begin with, the planning of the architecture and structure for the data base has to be developed and the choice of what kind of data base fit better for the project has to be made. Furthermore, we needed to establish a connection between the financial markets data and our tool.

Secondly, the manage and the store of the information. The first assignment was the obtaining of the information. The creation of algorithms whereby takes the markets information in order to save them in our data base. In addition, the use of this information to model and calculate financial ratios to keep also in our data base.

Finally, the data science. With this amount of data is possible create algorithms which can find out patterns to save them in the data base. When the live data get into our tool, the implement create the pattern of this data, store it and compare with historical patterns. If there is a coincidence, the tool knows what happened in the past and we try to forecast the fluctuate.

To sum up, the object of this project is to study one of the many ways of the research of the financial market. The machine learning, the data science and automatic trading is for sure the present and the future of the investment.

Resumen:

En este proyecto de fin de grado, se realizará una investigación sobre el trading algorítmico, así como de la gestión de la información. Se desarrollará una herramienta para ese propósito y se desplegará una base de datos. Ambos trabajarán conjuntamente en el estudio de los mercados financieros.

En primer lugar, se describirá la planificación de la arquitectura y estructura de la base de datos, previo estudio del tipo de base de datos que mejor se ajusta al proyecto. Además, se establecerá una conexión entre los datos de los mercados financieros y nuestra herramienta.

En segundo lugar, la gestión y almacenamiento de la información. Se desarrollarán implementaciones que recopilan toda esta información de los mercados y la modelan para además calcular ratios financieros.

Por último, la minería de datos y el aprendizaje automático. Con esta cantidad de datos es posible crear algoritmos que puedan encontrar patrones para posteriormente almacenarlos. Cuando los datos en vivo entran en nuestra herramienta, el aplicativo crea el patrón de estos datos, lo almacena y compara con patrones históricos. Si existe una coincidencia, la herramienta consulta lo sucedido y tratamos de predecir la fluctuación.

En resumen, el objetivo de este proyecto es estudiar una de las muchas formas de investigación en el sector. El aprendizaje automático, la ciencia de los datos y el trading algorítmico es el presente y el futuro de la inversión.

Índice

1	Introducción	10
1.1	Introducción y motivación	10
1.2	Objetivos:	11
1.3	Estructura del trabajo	11
2	Estado del arte.....	12
2.1	Introducción	12
2.2	Empresas de servidores	12
2.3	Minería de datos	14
2.3.1	Generación de un modelo de datos	14
2.3.2	Tipos de problemas de data mining	16
2.3.3	Problemas descriptivos	16
2.3.4	Problemas Predictivos.....	17
2.4	Técnicas y enfoques para data mining	18
2.4.1	Enfoques	19
2.4.2	Operaciones de Data Mining	20
2.4.3	Técnicas de Data Mining	21
2.5	Minería de patrones temporales	23
2.5.1	Modelo punto de intervalo de tiempo	23
2.5.2	Modelo intervalo de tiempo	24
2.5.3	Transformadas.....	24
2.6	Operadores en intervalos de tiempo	25
2.6.2	Patrones de puntos temporales.....	27
2.6.3	Algoritmos de episodios de patrones	28
2.7	Base de Datos	28
2.7.1	MySQL	28
2.7.2	MongoDB	28
3	Establecimiento del problema.....	30
3.1	Introducción	30
4	Solución del Problema	31
4.1	Solución.....	31
4.1.1	Proveedor de datos	31
4.2	Especificación de requisitos	31
4.3	Diseño.....	33
4.3.1	Niveles de la arquitectura.....	34
4.3.2	Arquitectura	34
4.4	Implementación.....	35

4.4.1	Gestión de la información.....	35
4.4.2	Almacenamiento de datos.....	37
4.4.3	Análisis Técnico.....	39
4.4.4	Análisis Fundamental.....	41
4.4.5	Patrones.....	42
4.4.6	Estructura lógica de la herramienta.....	44
4.4.7	Framework.....	45
5	Valores.....	46
6	Discusión.....	48
7	Conclusión y líneas futuras.....	49
7.1	Conclusión.....	49
7.2	Líneas futuras.....	50
8	Bibliografía.....	51

1 INTRODUCCIÓN

1.1 INTRODUCCIÓN Y MOTIVACIÓN

Esta propuesta de trabajo de fin de grado se enmarca dentro del ámbito de los mercados financieros que como en muchos otros sectores productivos de la sociedad está inmerso en una evidente transformación tecnológica.

La cantidad de datos con los que se trabaja es muy elevada. Las nuevas tecnologías y paradigmas de uso de la información buscan mejorar la gestión de los mismos para así convertirlas en una herramienta eficiente y fundamental para la toma de decisiones y consecución de objetivos.

La minería de datos es un campo de la estadística y de las ciencias de computación que busca dar solución a esta problemática. Existen diferentes tipos de escenarios donde se suele aplicar esta práctica:

- **Pronóstico:** cálculo de las ventas y predicción de las cargas del servidor o del tiempo de inactividad del servidor.
- **Riesgo y probabilidad:** elección de los mejores clientes para la distribución de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- **Recomendaciones:** determinación de los productos que se pueden vender juntos y generación de recomendaciones.
- **Búsqueda de secuencias:** análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos.
- **Agrupación:** distribución de clientes o eventos en grupos de elementos relacionados, y análisis y predicción de afinidades.

En este proyecto vamos a desarrollar una herramienta que gracias a la minería de datos sea capaz de realizar un análisis y pronóstico de los mercados financieros. Vamos a gestionar una base de datos que en primera instancia recopile toda esta información, la almacene y realice con ella los estudios y análisis objeto de este proyecto. El análisis será dinámico ya que la herramienta obtendrá datos históricos y en directo cuando los mercados estén abiertos. Por tanto, las metas y completando lo escrito anteriormente son:

- Obtener los diferentes tipos de datos que existen en los mercados financieros. Esta información, se gestionará en directo y de forma automática.
- Almacenar y gestionar la información para que, de una manera eficiente, se pueda obtener una valía a esta cantidad de datos tan elevada.
- Desarrollar algoritmos que, gracias a esta información, compren o vendan acciones o índices financiero.

Para dar valía a la información obtenida, se calcularán patrones de comportamiento en los mercados para predecir en ciertos momentos puntuales (cuando se replique el patrón) hacia dónde va a fluctuar el mercado. El fin es validar si estos patrones obtenidos de datos pasados tienen tasa de acierto válida para una operativa de inversión en tiempo real.

1.2 OBJETIVOS:

- Establecer los requisitos funcionales.
- Estudio de los paradigmas de bases de datos actuales.
- Diseñar y desplegar la base de datos.
- Comunicar la base de datos con la fuente de datos de mercado.
- Diseño e implementación de la lógica de obtención y almacenamiento de datos.
- Estudio de fórmulas y ratios financieros que nos ayuden a modelar los datos obtenidos.
- Implementar operaciones de modelado los datos de mercado para almacenar resultados.
- Crear con todos los datos almacenados patrones de fluctuación de los mercados.
- Implementar lógica de comparación de patrones en directo con los patrones ya almacenados.
- Estudiar si a través de estos patrones se puede predecir la fluctuación del mercado.
- Validar herramienta.

1.3 ESTRUCTURA DEL TRABAJO

En el presente trabajo se ha estructurado la siguiente manera:

- El capítulo uno presenta la introducción y objetivos del trabajo.
- En el capítulo dos se muestra el estado del arte donde se presenta la situación actual del sector, empresas que ofrecen servicios para este tipo de herramientas, minería de datos y comparativa entre bases de datos SQL y NoSQL.
- En el capítulo tres observaremos los problemas a los que busca dar salida este trabajo de fin de grado.
- En el capítulo se muestra la solución al problema planteado en el capítulo anterior y una explicación de la implementación realizada.
- En el capítulo cinco se presentan las acciones, futuros y pares de divisas que se han utilizado en el análisis.
- En el capítulo seis se puede observar una discusión relacionada con el tema objeto del trabajo.
- Por último, en el capítulo siete se muestra la conclusión al proyecto realizado.

2 ESTADO DEL ARTE

2.1 INTRODUCCIÓN

Este trabajo presenta una solución basada en análisis de series temporales en mercados financieros. Por ello a continuación, revisaremos en el estado del arte la coyuntura actual de las soluciones relacionadas, técnicas de minería de datos y un análisis de los dos principales paradigmas de bases de datos en busca de cual se puede ajustar mejor a la operativa.

La negociación de alta frecuencia, también conocida en el ámbito financiero por su nombre en inglés high-frequency trading o por sus siglas HFT, es un tipo de negociación que se lleva a cabo en los mercados financieros utilizando intensamente herramientas tecnológicas sofisticadas para obtener información del mercado y en función de la misma intercambiar valores financieros tales como activos u opciones.

El análisis es cuantitativo ya que emplea algoritmos informáticos para analizar datos del mercado e implementar estrategias de negociación.

Cada posición de inversión se mantiene solo durante muy breves períodos de tiempo, para rápidamente ejecutar la posición y comprar o vender, según el caso, el activo de que se trate. En ocasiones esto se lleva a cabo miles, o incluso decenas de miles de veces al día. Los períodos de tiempo durante los que se mantienen las posiciones pueden ser de solo fracciones de segundo.

Suele llevarse a cabo por importantes salas de mercado, normalmente de fondos de inversión o bancos de inversión con carteras con gran volumen y muy diversificadas. Es un tipo de negociación muy sensible a la velocidad de procesamiento del mercado y al propio acceso al mercado.

Se estima que el 20%-50% de volumen negociado en los mercados provienen de algoritmos de high-frequency trading. Este ambiente de inversión genera opiniones enfrentadas.

2.2 EMPRESAS DE SERVIDORES

Actualmente, existen empresas que ofrecen servidores y soluciones tecnológicas enfocadas a la inversión automatizada. Ofertan servidores dedicados o virtuales de alto rendimiento que están ubicados en zonas muy cercanas a los principales gateways y servidores de los mercados financieros con el fin de minimizar el tiempo de latencia entre los mercados y la unidad de procesamiento. La velocidad es muy importante en el HFT.

Empresas como speedytraders poseen servidores ubicados en Chicago ciudad donde la latencia con CME (Chicago Mercantile Exchange), el mercado de futuros más importante, es de tan solo 1 ms. Del mismo modo, ofertan servidores ubicados en New-Jersey cerca de los data centers de NYSE (New York Stock Exchange), Nasdaq y BATS.

LARGE New-Jersey	XL New-Jersey	MEDIUM Chicago	LARGE Chicago
\$ 180 /month	\$ 220 /month	\$ 220 /month	\$ 260 /month
<ul style="list-style-type: none"> • Intel E3-1230v2 3.2 Ghz • 16 GB of RAM • 2 X 60 GB SSD disk • 1 Gb/s link • Windows Server 	<ul style="list-style-type: none"> • Xeon E3-1245v2 3.4Ghz • 32 GB Ram • 2 X 60 GB SSD disk • 1 Gb/s link • Windows Server 	<ul style="list-style-type: none"> • Intel i3-2120T 2.5 Ghz • 8 GB of RAM • 500GB Sata disk • 100 Mb/s Internet link • Windows Server 	<ul style="list-style-type: none"> • Xeon E3-1230 3.2 Ghz • 16 GB of RAM • 500 GB Sata disk • 100 Mb/s link • Windows Server

Oferta de la empresa Speedytraders

Empresas como Xenon no se quedan atrás y ofertan sus soluciones tecnológicas específicas para este campo de la inversión de alta frecuencia. Esta empresa posee una alta gama de productos que permiten customizar tu hardware a las necesidades de los requerimientos.

En primer lugar, el servidor. Las especificaciones de estos servidores son muy elevadas y son las siguientes:

1. Radon Duo RX1882 (20 cores):
 - Procesador Dual Intel E5.
 - 3.79GHz.
 - 10 núcleos.
 - NVMe SSD.
 - 8X8 DDR4-2400.
2. Radon Solo RX980 (8 cores):
 - Intel core i7-5960X.
 - 4.6GHz.
 - 8 núcleos.
 - 8X8 DDR4-2666.
 - 140W.
3. Radon Solo RX980 (10 cores):
 - Intel core i7-6950X.
 - 4.3GHz.
 - 8 núcleos.
 - 8X8 DDR4-2666.
 - 140W.
4. Radon Solo RX880 (4 cores):
 - Intel core i7-4790KX.
 - 4.8Ghz.
 - 4 núcleos.
 - 84W.

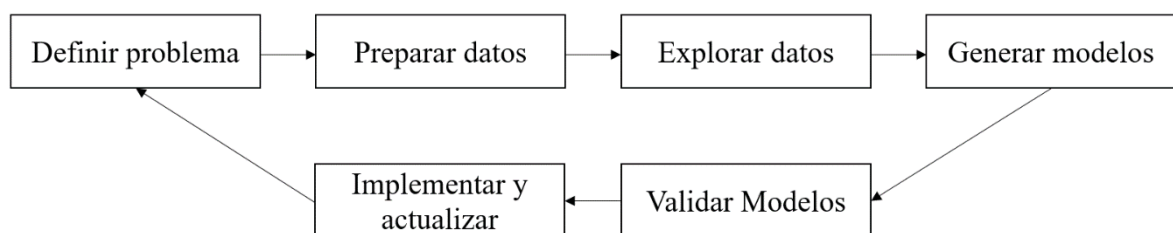
2.3 MINERÍA DE DATOS

La minería de datos es un paso esencial en el proceso global de extracción de conocimiento potencialmente útil a partir de un gran volumen de datos, proceso conocido como KDD (Knowledge Discovery in Databases). Consiste en la aplicación de técnicas de análisis de datos y algoritmos que dan como resultado una enumeración particular de estructuras sobre los datos.

2.3.1 Generación de un modelo de datos

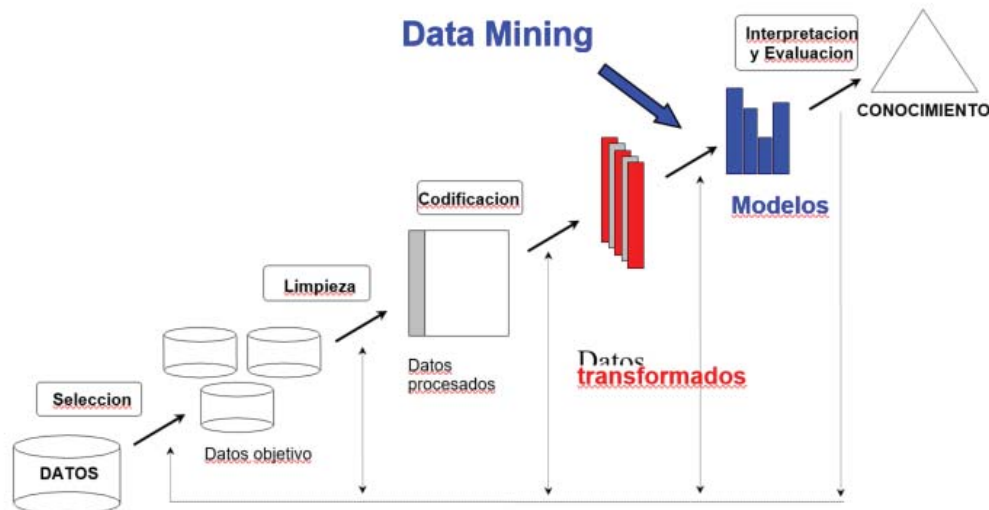
A la hora de realizar un proyecto de descubrimiento de conocimiento la planificación del mismo es esencial en el éxito del modelo:

- Definir el problema: el primer paso en la minería de datos consiste en definir el problema y considerar formas de usar los datos para proporcionar una respuesta.
- Preparar los datos: limpiar y consolidar la valía de la información para su posterior uso. Los datos como tal pueden estar almacenados u obteniéndose de una forma dispersa y sin coherencia. La limpieza de los datos es fundamental y puede implicar eliminar datos o interpolar datos que faltan.
- Explorar los datos: el tercer paso consiste en conocer los datos correctamente para así poder realizar un modelo de minería de datos con éxito.
- Generar los modelos: en este paso se crea la estructura y la lógica a seguir por el proceso.
- Validar los modelos: comprobar que los resultados obtenidos son correctos y de la misma manera todo el proceso.
- Implementar y actualizar es el último paso y consiste en afinar el modelo para que logre más satisfactoriamente su cometido.



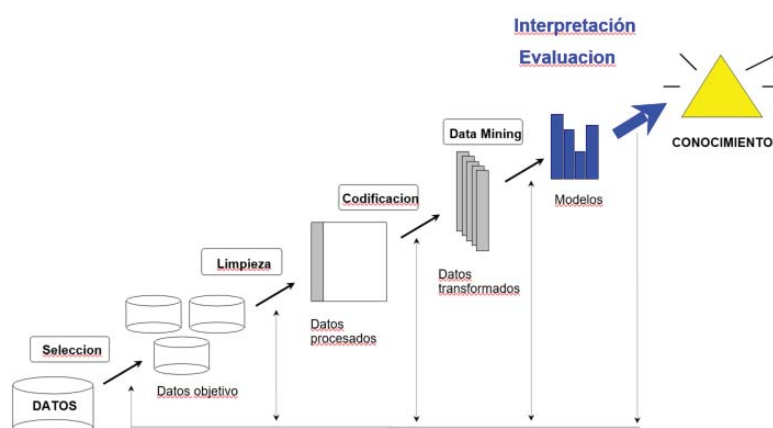
(*) Diagrama generación modelo de datos

La fase de *Data Mining* es la fase central del proceso de descubrimiento en la que se aplican los algoritmos de búsqueda de conocimiento a los datos previamente preprocesados. En realidad, este paso es inseparable del siguiente en la cadena que es el análisis de resultados. De hecho, a menudo el análisis de los resultados obtenidos provoca que se retroceda de nuevo a la fase de preparación de cara a obtención de más datos o más atributos.



Puesto que es en esta fase en la que se ejecutan los algoritmos de data mining, es fundamental en esta fase la actuación del analista de datos a menudo ayudados por los expertos de bases de datos. Para que el proceso salga correctamente es fundamental que el analista disponga del conjunto de datos preprocesado, los correspondientes metadatos y toda la información que de los datos se haya previamente extraído en las fases anteriores de análisis. Lo que ocurrirá en esta fase dependerá del tipo de meta a obtener. Esto es, no es lo mismo si el resultado final buscado es una caracterización de datos que si la meta perseguida es un modelo predictivo donde posiblemente el proceso será más largo y complicado.

El análisis de los resultados es uno de los pasos más importantes del proceso. Además, a pesar de los avances en interfaces gráficas de visualización de resultados, este paso sólo se puede realizar correctamente si trabajan en conjunción el analista de datos con expertos de la organización.



Para quien se acerca por primera vez a realizar un proceso de data mining el número de algoritmos existentes para resolver el mismo tipo de problema puede llevar a muchas confusiones, la situación viene empeorada por el hecho de que distintos investigadores utilizan distinta terminología para referirse al mismo término.

Si bien es difícil establecer una clasificación de los posibles tipos de problemas de data mining que nos podemos encontrar, es aún más difícil encontrar un procedimiento que establezca el mejor algoritmo aplicar para cada tipo de problema.

2.3.2 Tipos de problemas de data mining

Tal y como se ha comentado con anterioridad existen innumerables problemas de data mining con los que nos podemos encontrar:

- Predecir el nivel de morosidad de un cliente
- Saber quiénes son mis clientes
- Encontrar los productos que más frecuentemente se compran juntos
- Encontrar el perfil del comprador del producto A
- Detectar los clientes que están cometiendo acciones fraudulentas
- Encontrar el perfil del cliente que me abandonará el mes siguiente
- Encontrar los síntomas de enfermedades que más a menudo aparecen juntos
- Analizar los canales de televisión que a menudo se contratan juntas
- Calcular el valor potencial de un cliente
- ...

Es fácil comprobar ante esta lista que no pretende ser exhaustiva sino sólo mostrar algunos ejemplos de problemas data mining frecuentes como si bien la lista puede seguir creciendo, sin embargo, podemos establecer una primera clasificación de estos problemas en:

- Problemas Descriptivos
- Problemas Predictivos

Algunos autores denominan a los problemas descriptivos problemas de descubrimiento indirecto puesto que no hay una meta a predecir y a los problemas predictivos los denominan problemas de descubrimiento directo. No obstante, la primera clasificación es la más extendida y es la que se utilizará en este curso.

2.3.3 Problemas descriptivos

Entendemos en este contexto como un problema descriptivo aquellos cuya meta es simplemente encontrar una descripción de los datos de estudio. Pertenecen a este tipo de problemas el ejemplo de conocer cuáles son los clientes de una organización (características de los mismos), o el encontrar los productos que frecuentemente se compran juntos o síntomas de enfermedades que se presentan juntos.

La meta de todos estos problemas es una descripción del conjunto de datos origen. No obstante, analizando estos ejemplos más en detalle observamos que si bien ambos pretenden descubrir características del conjunto origen en el primer caso, (descripción de los clientes) lo que se pretende de alguna manera es agrupar a los clientes en grupos más o menos homogéneos y extraer las características de estos objetos. Sin embargo, en el segundo tipo de consultas

(productos que se compran juntos o síntomas de enfermedades que se presentan juntos, si bien el problema sigue siendo descriptivo, el tipo de descripción requerida es diferente pues lo que se pretende es encontrar asociaciones esta vez no entre los objetos origen sino entre los valores de atributos o propiedades de estos objetos. Esto provoca una división más detallada del problema descriptivo en:

- **Análisis de segmentación:** Se refiere a los problemas donde la meta es encontrar grupos homogéneos en la población de objetos origen. A estos problemas se los denomina también problemas de aprendizaje no supervisado. El típico ejemplo de segmentación es realizar una segmentación de los clientes
- **Análisis de asociaciones:** Hace referencia a los problemas en los que se persigue obtener relaciones entre los valores de atributos de una base de datos. El ejemplo más típico es el de análisis de la cesta de la compra.

2.3.4 Problemas Predictivos

Por otra parte, existen consultas de data mining cuya meta es obtener un modelo que en un futuro pueda ser aplicado para predecir comportamientos. Este tipo de problemas es el que denominamos problemas predictivos o en entornos de inteligencia Artificial se denominan problemas de aprendizaje supervisado.

Es importante destacar aquí que el proceso de data mining es siempre un proceso inductivo en que no se realiza ninguna predicción de los datos. Se llaman problemas predictivos porque si bien las técnicas aplicadas para la obtención del modelo son técnicas de inducción sobre los datos de origen, el resultado (modelo) será aplicado para predecir.

Son entonces problemas de tipo predictivo de entre los que analizábamos con anterioridad los siguientes:

- Encontrar el perfil del comprador del producto A
- Encontrar el perfil del cliente que me abandonará el mes siguiente
- Calcular el valor potencial de un cliente
- Probabilidad de que un cliente devuelva un préstamo
- ...

Obsérvese como en todos estos ejemplos la meta buscada es un modelo que en un futuro pueda ser utilizado para saber si un cliente comprará o no el producto A a la vista de sus datos, saber si el cliente en vistas de su comportamiento me dejará, poder saber la primer vez que un cliente se acerca a mi organización cuál será su valor potencial o poder saber ante un cliente que está solicitando un préstamo si este lo devolverá o no y con qué probabilidad.

No obstante, una vez más podemos analizar estos problemas con más detenimiento para observar que por ejemplo en el primer caso la variable a predecir es una variable categórica (si

compra o no un producto) sin embargo, en el caso del préstamo, la variable a predecir es la probabilidad de devolución del mismo que es una variable numérica.

Esta distinción en el tipo de variables que el modelo predecirá nos lleva a una distinción dentro de los problemas predictivos que es:

- Problemas de clasificación: Hacen referencia a los problemas en los que la variable a predecir tiene un número finito de valores, esto es la variable es categórica. Un ejemplo de este tipo de problemas sería encontrar un modelo que a la vista de un histórico de clientes clasificados como “buenos”, “regulares” y “malos”, establezca qué tipo de cliente es uno nuevo.
- Problemas de predicción de valores. Se refieren a los problemas en los que la variable a predecir es numérica. En este caso como ejemplo podríamos tener el caso de encontrar un modelo que me establezca la probabilidad de que un cliente que está pidiendo un préstamo lo devolverá o no

El hacer esta distinción es importante pues dependiendo del tipo de problema así será la técnica que se utilizará para solucionarlo.

2.4 TÉCNICAS Y ENFOQUES PARA DATA MINING

Las técnicas son implementaciones específicas de los algoritmos que se utilizan para llevar a cabo las operaciones de data mining.

No todos los algoritmos para resolver un determinado problema de data mining son iguales y cada uno de ellos tendrá una serie de ventajas e inconvenientes. La conveniencia de aplicar un determinado algoritmo depende no sólo del tipo de problema con el que nos estemos enfrentando, sino que depende en gran medida del tipo de los datos con los que se está tratando. En este sentido conviene analizar los distintos enfoques y algoritmos que existen en la literatura pues en la vida real nos encontramos que las herramientas comerciales ofrecen todo un abanico de posibles algoritmos y es el usuario final el que tiene que decidir cuál de ellos utilizar.

En la Tabla que se muestra a continuación puede observarse un primer resumen de las distintas operaciones que hemos analizado y de posibles técnicas de *Data Mining* que se podrían aplicar en cada uno de los casos [Cabena 97].

<i>Operaciones</i>	<i>Técnicas</i>
Modelos predictivos	- Clasificación - Predicción de valores
Segmentación de bases de datos	- Clustering demográfico - Clustering neuronal
	- Descubrimiento de asociaciones

Análisis de relaciones	<ul style="list-style-type: none"> - Descubrimiento de patrones secuenciales - Descubrimiento de patrones en series temporales
-------------------------------	--

(*) Tabla Operaciones y técnicas de Data Mining

Se va a analizar a continuación los distintos enfoques que se han presentado para resolver problemas de data mining para pasar con posterioridad a analizar técnicas provenientes de cada uno de estos enfoques

2.4.1 Enfoques

Hasta el momento han sido varios los enfoques que han surgido y que van dirigidos hacia la investigación en KDD. Entre los enfoques más significativos podríamos mencionar:

2.4.1.1 Enfoques estadísticos

Estos enfoques se dirigen hacia la construcción de modelos estadísticos para el análisis de datos.

A este tipo de enfoques pertenecen tanto la inferencia bayesiana, utilizada como herramienta para la extracción de conocimiento, como las redes bayesianas, empleadas para descubrir relaciones causales entre objetos.

2.4.1.2 Machine Learning

Este enfoque utiliza un modelo cognitivo para imitar el proceso de aprendizaje humano. Son numerosas las técnicas que pertenecen a este enfoque, podrían citarse, entre otras, el aprendizaje por ejemplos, el *clustering* conceptual y la inducción mediante árboles de decisión. Todas estas técnicas tienen en común la generalización de modelos cognitivos fáciles de entender por los humanos, sin embargo suponen que el conjunto de datos es pequeño, lo que implica una gran restricción al hablar de grandes volúmenes de datos.

2.4.1.3 Enfoques orientados a las bases de datos

En este tipo de enfoques se utilizan técnicas específicas de bases de datos para extraer reglas de las bases de datos relacionales. Dentro de este tipo de enfoques se podrían incluir técnicas como la generalización y el uso de jerarquías conceptuales para la extracción de reglas, así como el uso del algoritmo "A Priori", propuesto por Agrawal para la extracción de reglas de asociación de grandes bases de datos transaccionales.

2.4.1.4 Enfoques matemáticos

Estos enfoques se dirigen hacia la construcción de modelos matemáticos para la extracción de reglas, regularidades y patrones del modelo.

A este enfoque, entre otras, pertenece la teoría de los *Rough Sets*, utilizada para el descubrimiento de conocimiento en bases de datos como herramienta para encontrar

dependencias en los datos y para derivar tablas de decisión de las que se pueden obtener las reglas de decisión. Al mismo tiempo, los Rough Sets se han utilizado en el cálculo de reductos de una tabla de decisión, cálculo que consiste en la eliminación de los atributos o propiedades que sean dependientes o redundantes en dicha tabla.

2.4.1.5 Enfoques integrados

Integran varios enfoques con el fin de obtener mayor eficiencia. Así, por ejemplo, el sistema QUEST, desarrollado por IBM, integra técnicas de *data mining* con enfoques orientados a las bases de datos, el sistema DBMINER, propuesto por la universidad Simon Fraser de Canadá combina técnicas de *data mining* con tecnología de bases de datos, el sistema de KDD que está en desarrollo en el *Departamento de lenguajes y sistemas informáticos de la Facultad de Informática de Madrid*, RSDM, integra la teoría de los Rough Sets con las técnicas de bases de datos relacionales para extraer reglas de clasificación y caracterización.

Además de estos, existen otros enfoques dentro de los cuales pueden incluirse las técnicas de visualización y de representación del conocimiento.

A continuación, se describirán un poquito más en detalle primero las operaciones de *Data Mining* y luego las diferentes técnicas que se aplican para cada tipo de operación.

2.4.2 Operaciones de *Data Mining*

A continuación, se va a comentar cada una de las diferentes operaciones de *Data Mining* que aparecen en la Tabla.

- **Modelos predictivos**

Los modelos predictivos se basan en la experiencia del aprendizaje humano, donde se utiliza la observación para construir un modelo esencial de algo y luego lo utilizamos para clasificar algo similar. Por ejemplo observando varios perros, creamos un modelo genérico de perro que es un ser vivo de cuatro patas que ladra y luego las distintas razas de perro las asociamos con el modelo de perro que hemos creado previamente, sin necesidad de memorizar todos y cada uno de los tipos de perro que existen.

En los algoritmos de *Data Mining* se utilizan los datos para formar un modelo con las características esenciales de esos datos. Los modelos se crean en dos fases, una de entrenamiento y otra de prueba. Durante la fase de entrenamiento se utiliza un conjunto de datos históricos y en la fase de prueba se valida el modelo con datos reales. Al final estos modelos se reducen a reglas del tipo IF-THEN.

La clasificación se utiliza para determinar a qué clase pertenece un registro de la base de datos. La clase debe pertenecer a un conjunto finito y determinado previamente. Este tipo de aprendizaje es aprendizaje supervisado.

- **Segmentación de bases de datos**

El objetivo de segmentar una base de datos es particionar la base de datos en conjuntos de objetos con características similares, es decir conjuntos homogéneos. Dentro de un conjunto los elementos deben ser homogéneos, pero los elementos de diferentes segmentos deben ser diferentes. Para clasificar un elemento se utiliza una distancia que se define en función del algoritmo.

Este tipo de operación se utiliza para descubrir subpoblaciones dentro de una base de datos y poder crear perfiles que caractericen a esas subpoblaciones.

Este tipo de aprendizaje es aprendizaje no supervisado, puesto que en ningún momento se les indica a los algoritmos las clases a utilizar.

Las técnicas que se utilizan son clustering demográfico o clustering neuronal, que se diferencian en el tipo de datos que permiten como entrada, en la distancia que utilizan para calcular la pertenencia a una clase y en la forma de organizar los segmentos de salida.

- **Análisis de asociaciones**

Mediante esta técnica se trata de buscar relaciones entre diferentes registros de la base de datos. Estas relaciones suelen llamarse asociaciones. Un ejemplo de este tipo de operación es el descubrimiento de aquellos productos o servicios que los clientes suelen comprar o utilizar juntos. Como consecuencia, los resultados a menudo se utilizan para realizar venta cruzada o para realizar el seguimiento de los objetivos de ventas de una empresa o el movimiento de precios en el mercado de algún bien.

2.4.3 Técnicas de Data Mining

A continuación, se van a describir las técnicas de *Data Mining* que se pueden aplicar para resolver las operaciones de *Data Mining* descritas.

- **Modelos predictivos: clasificación**

En estos modelos se utiliza aprendizaje supervisado. Se suelen utilizar árboles de decisión, árboles de inducción y redes neuronales.

Estos modelos utilizan un conjunto de datos de entrenamiento para crear el modelo, que posteriormente se utiliza para clasificar individuos desconocidos. El problema de las redes neuronales es que únicamente aceptan en la entrada valores numéricos.

- **Modelos predictivos: predicción de valores**

Para la predicción de valores se utiliza la regresión lineal o la regresión no lineal. También se utiliza para la creación de estos modelos la técnica de funciones de base radial (RBF), que es una técnica más robusta que la de las regresiones. Se suelen utilizar redes neuronales RBF que son modelos supervisados de aprendizaje.

- **Segmentación de bases de datos: clustering demográfico**

Se compara cada registro de la base de datos con todos los segmentos creados por la función de creación de los segmentos. Se basa en el principio de votación (*condorset*) para medir la distancia del registro de entrada con los segmentos ya creados y asignar el registro de entrada al segmento correspondiente. El número de clusters se ajusta automáticamente. Esto se usa fundamentalmente con datos categóricos.

- **Segmentación de bases de datos: clustering neuronal**

Se suelen utilizar redes neuronales basadas en aprendizaje no supervisado, como por ejemplo mapas de Kohonen. Este tipo de técnica de *Data Mining* es apropiado cuando los datos de entrada son de tipo numérico, pero se pueden utilizar también con datos

categoricos si los transformamos previamente. La distancia Euclídea se utiliza para crear los clusters.

- **Análisis de relaciones: asociaciones**

El objetivo de esta técnica de *Data Mining* es encontrar elementos que implican la presencia de otros elementos dentro de una misma transacción. El resultado de esta técnica son reglas del tipo “if X then Y”. En las reglas a X se denomina cabeza de la regla e Y se denomina cuerpo.

Uno de los algoritmos de asociación más utilizados es Apriori [Agrawal 94]. Apriori se basa en contar las ocurrencias de todos las posibles combinaciones de elementos. Lo que hace es contar las ocurrencias de todos los elementos presentes en las transacciones de la base de datos y crea un vector donde cada elemento del vector lleva una cuenta de un elemento de la base de datos. Aquellas celdas del vector cuyo valor esté por debajo del nivel de soporte (umbral) se ignoran. Para los que no se han ignorado se forma una matriz para almacenar las ocurrencias de cada elemento con otros posibles elementos y otra vez se eliminan aquellas ocurrencias que no superan el umbral, y así sucesivamente hasta que ningún elemento supere el umbral. El problema de este algoritmo es que hay que leer los datos de la fuente de datos cada vez que creamos una nueva matriz.

Las reglas obtenidas tienen un soporte y una confianza que se definen de la forma que aparece en la Ecuación 2.1.

$$\text{Soporte} = \frac{\text{Número de transacciones que soportan la regla}}{\text{Número total de transacciones}}$$

$$\text{Confianza} = \frac{\text{Número de transacciones que soportan la regla}}{\text{Número de transacciones que soportan la cabeza de la regla}}$$

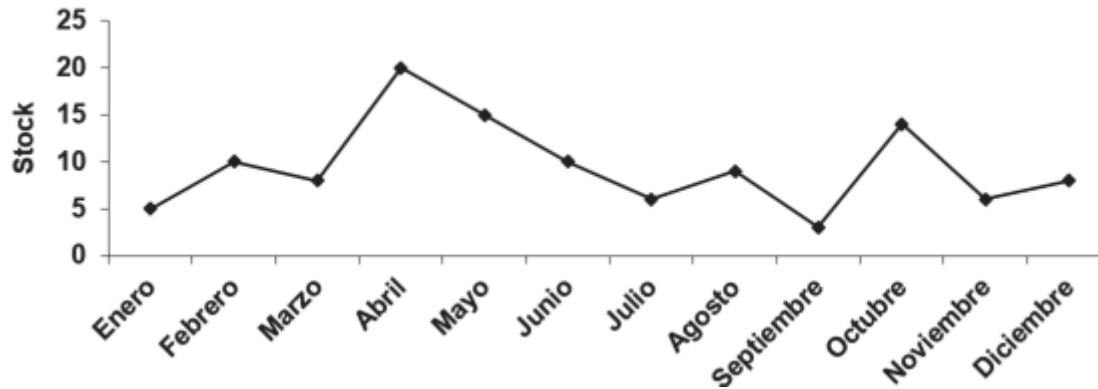
(*) Ecuación: Definición de soporte y confianza

- **Análisis de relaciones: patrones secuenciales**

Trata de descubrir patrones entre transacciones en las que un conjunto de elementos van seguidos de otro conjunto de elementos distanciados un periodo de tiempo determinado.

- **Análisis de relaciones: patrones en series temporales**

Con esta técnica se pretenden descubrir ocurrencias o secuencias similares a una dada en una base de datos que almacene información que represente una serie temporal, como puede ser la evolución de los precios de mercado o datos de telemetría provenientes de algún sensor, es decir, una serie temporal es un conjunto de valores de una variable en un periodo de tiempo. Si representamos la serie en una gráfica en el eje horizontal aparece el tiempo y el eje vertical se representan los valores de la variable como puede observarse en Figura 2.4.



(*) Figura 2.1: Ejemplo de serie temporal

Para poder considerar dos series o subseries como similares hay que ajustar la definición de igualdad para conseguir un significado de similaridad. Hay dos enfoques válidos para esta aproximación.

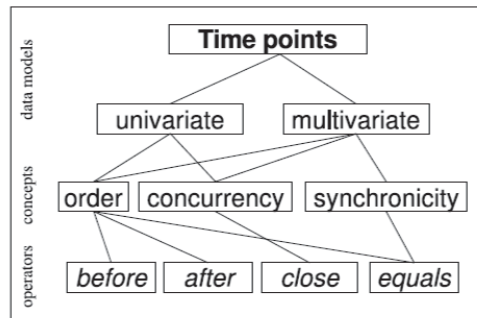
El primer enfoque es introducir un margen de error que es el máximo valor en el que pueden diferenciarse dos series o subseries para que se consideren iguales. Este enfoque asegura una tolerancia para la variabilidad de las series temporales sin perder los patrones que puedan considerarse similares.

El segundo enfoque es utilizar un intervalo llamado *mismatch gap*, que es el número de unidades consecutivas para las cuales los valores que no son similares se ignoran. De esta forma se asegura que pequeños picos o hundimientos en dos series no enmascaran un patrón similar. Una restricción de las series temporales es que los datos deben ser cuantitativos, es decir cantidades, y además deben depender del tiempo. La desventaja de esta técnica es que hay que tener mucho cuidado a la hora de definir los parámetros, puesto que si se especifica un margen de error pequeño, puede que no se descubran secuencias similares presentes en los datos, pero si el margen de error es grande, puede que se consideren como series similares aquellas que no lo son. Asimismo, si se le da un valor bajo al parámetro *mismatch gap* se pueden ocultar secuencias similares, pero si se le da un valor alto, se considerarían secuencias que no son similares como similares.

2.5 MINERÍA DE PATRONES TEMPORALES

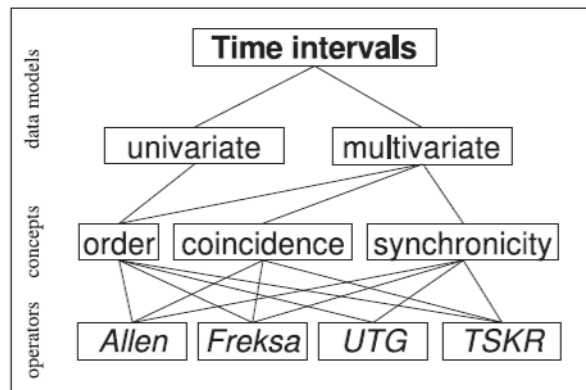
2.5.1 Modelo punto de intervalo de tiempo

Punto de intervalo de tiempo para un modelo de datos se refiere a las observaciones asociadas a momentos específicos. Puntos de la línea temporal concretos. Si se observan estos puntos de una forma paralela, se dice que el modelo es multivariante. Si el observador es único se denomina como univariante.



2.5.2 Modelo intervalo de tiempo

Los modelos de datos basados en intervalos de tiempo analizan los datos entre dos puntos de la línea temporal. Si se observan estos puntos de una forma paralela, se dice que el modelo es multivariante.

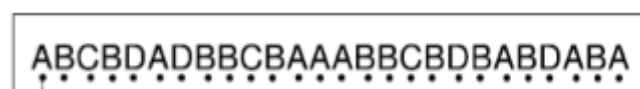


2.5.3 Transformadas

En las series de datos temporales los valores que se reciben son numéricos. Cada punto de tiempo tiene un valor y la progresión del tiempo crea una gráfica con los valores de la serie de datos temporales. Al analizar datos, se realiza una transformación de los datos con el fin de crear patrones de comportamiento. Lo fundamental es el comportamiento que se puede reflejar en transformadas ya que los valores numéricos son difícilmente repetible y replicable.

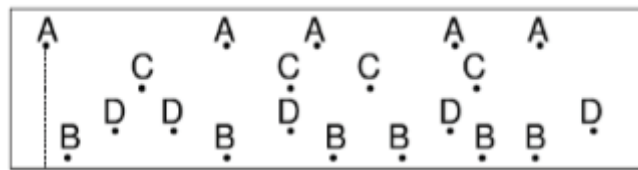
Podemos clasificar las transformadas de modelos de datos en:

- Series de tiempo simbólicas: se trata de una serie de tiempo con un valor nominal en el tiempo que sustituye al valor numérico.



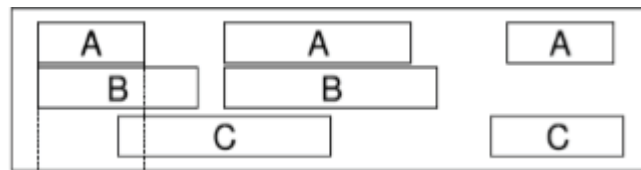
(*) Series de tiempo simbólicas

- Secuencias de tiempo simbólicas: se obtienen con valores nominales con posibles datos duplicados en un punto del tiempo.



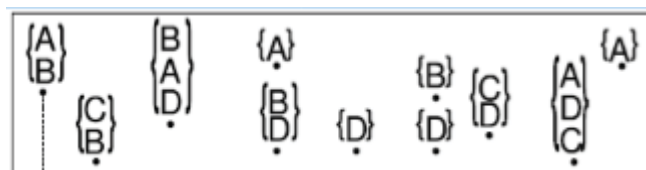
(*) Secuencias de tiempo simbólicas

- Secuencia de intervalos simbólicas: con valores nominales que se solapan una a otras.



(*) Secuencia de intervalos simbólicas

- Secuencias simbólicas de tiempo: secuencia de tiempo con conjuntos de valores nominales asignados a cada punto de tiempo. Esta es derivada de la secuencia de intervalos simbólicas y secuencias de tiempo simbólicas.




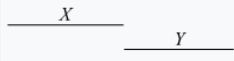

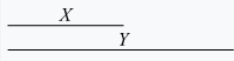
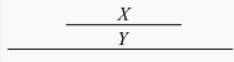
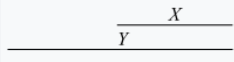
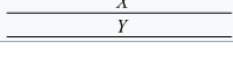
(*) Secuencias simbólicas de tiempo

2.6 OPERADORES EN INTERVALOS DE TIEMPO

2.6.1.1 Relación de intervalo de Allen

La relación de intervalo de Allen fue expuesta por James F. Allen en 1983 y se usa en algebra para el cálculo de razonamientos temporales. Esta relación enuncia diferentes bases en el análisis de series temporales:

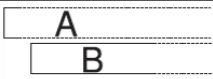
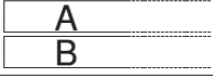
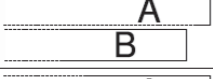
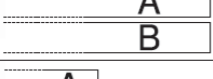
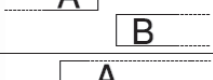
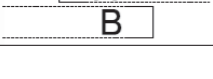
- Trece relaciones forman un algebra o una base de posible relación entre dos bases.

Relation	Illustration	Interpretation
$X < Y$ $Y > X$		X takes place before Y
$X m Y$ $Y mi X$		X meets Y (<i>i</i> stands for <i>inverse</i>)
$X o Y$ $Y oi X$		X overlaps with Y
$X s Y$ $Y si X$		X starts Y
$X d Y$ $Y di X$		X during Y
$X f Y$ $Y fi X$		X finishes Y
$X = Y$		X is equal to Y

- Dos intervalos cualquiera tiene exactamente una de estas relaciones expuestas anteriormente.
- Resuelve umbrales y extensiones difusas de algunos problemas.

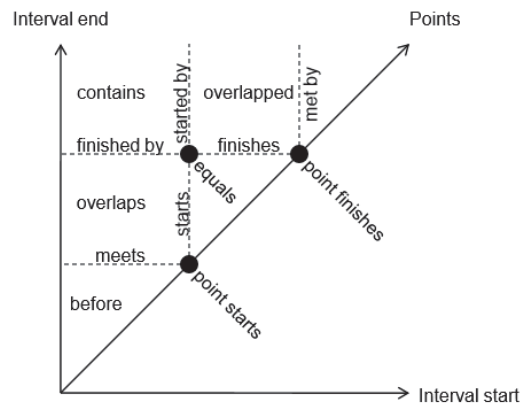
2.6.1.2 Relación de semi intervalos de Freksa

Enunciada por Christian Freksa en 1992. En ella se define como la relación entre los puntos inicio y final de dos intervalos son suficientes para identificar de manera única la relación. No usada ampliamente en la minería de datos.

older <i>younger</i>	
head to head	
survives <i>survived by</i>	
tail to tail	
precedes <i>succeeds</i>	
born before death <i>died after birth</i>	

2.6.1.3 Relación de intervalos/puntos de Reich

Enunciada en 1994 por Reich es una extensión de lo que ya enuncio Allen en 1983. En ella se proponen tres puntos más que dan lugar a cinco relaciones más. Además, propuso relaciones para diversificar el tiempo de razonar en mundos de futuros múltiples.



2.6.1.4 Relación de intervalos de Roddick

Enunciada por Roddick y Mooney en 2005 se trata del mismo modo de una extensión de la que ya enuncio Allen pero incluyendo relación de intervalos mediante puntos intermedios. Se diferencia principalmente en los nueve tipos de versiones de solapamientos que se enuncian este teorema y que da lugar a cuarenta y nueve relaciones en total. Este teorema se diseñó principalmente para agrupar con un orden local y arbitrario.

2.6.2 Patrones de puntos temporales

A continuación, se van a enumerar algoritmos de estudio de patrones para puntos temporales. En esta metodología de análisis se utilizan estructuras de datos especiales para representar los datos y derivar patrones frecuentes para buscar eficiencia de análisis.

Algoritmos de patrones secuenciales. Patrones que guardan una secuencia de símbolos. Se basan en el concepto de orden:

- AprioriAll [Agrawal/Srikant 1995].
- SPADE [Zaki 1998].
- PrefixSpan [Pei et al. 2001].

Algoritmos de patrones secuenciales cerrados:

- CloSpan [Yan et al. 2003].
- BIDE (BIDirectional Extension checking) [Wang/Han 2004].
- BIDEMargin [Fradkin/Moerchen 2010]

Otras variaciones:

- Regular expressions [Garofalakis et al. 1999]
- Multivariate data [Pinto et al. 2001]
- Allow mismatches [Kum et al. 2003][Zhu et al. 2007]
- Generators [Lo et al. 2008]

2.6.3 Algoritmos de episodios de patrones

Enunciado por Manila entre los años 1995 y 1997. Estos patrones buscan episodios que identifiquen patrones. En definitiva, ordena la relación existente entre símbolos. Este tipo de patrones expresan y trabajan con orden y de la misma manera con concurrencia de entrada de símbolos.

- Episodios en serie: orden entre símbolos o episodios.
- Episodios paralelos: símbolos observados de una manera paralela en una ventana temporal.
- Episodios: en principio orden arbitrario parcial de símbolos, pero a menudo combinación de episodios seriales y paralelos

Estos patrones tienen su razón de ser y son muy utilizados en el sector de las telecomunicaciones para el envío de mensajes de larga longitud.

2.7 BASE DE DATOS

A día de hoy, existen dos paradigmas en lo que arquitectura de base de datos se refiere. En primer lugar, las bases de datos SQL y su arquitectura entidad-relación y por otro lado las bases de datos NoSQL.

Para la realización de este proyecto, se ha realizado un estudio comparativo entre estas dos plataformas para averiguar cuál de ellas se ajusta mejor a las especificaciones del proyecto. Finalmente se ha optado por el desarrollo en MongoDB, arquitectura NoSQL.

2.7.1 MySQL

MySQL es la plataforma de base de datos más popular del mercado. El desarrollo, distribución y soporte corre cargo de la empresa Oracle Corporation. Entre las características principales de MySQL es la del almacenamiento de los datos en filas y estas filas en tablas. Además, usa un lenguaje estructurado de consultas como es SQL.

2.7.2 MongoDB

MongoDB por su parte, es una plataforma open-source desarrollada por MongoDB Inc. Su particularidad fundamental es que almacena los datos en ficheros JSON. Estos documentos se dividen en campos con los datos a almacenar. Pueden variar su estructura y por tanto no son estáticos. La información que guarda relación se almacena junta para un acceso más rápido mediante el lenguaje de consultas propio de MongoDB.

Bondades tecnológicas fundamentales de MongoDB:

- Plataforma que trabaja con esquemas dinámicos. Lo que significa que la política y definición de la base de datos permite variaciones a futuro, es decir, escalabilidad.
- Almacenamiento de estructuras complejas como arrays, listas u otras se realizan y gestionan de una manera menos compleja.
- Los ficheros que guardan relación se distribuyen dentro de la base de datos en diferentes colecciones.
- Alta escalabilidad.

- Alta disponibilidad.
- Out of the box replication*.
- Auto-sharing.

MySQL	MongoDB
Tabla	Colección
Fila	Documento
Columna	Campo
Joins	Documentos incrustados

**Comparativa entre MySQL y MongoDB*

	MySQL	MongoDB
Rich Data Model	No	Si
Dynamic Schema	No	Si
Typed Data	Si	Si
Data Locality	No	Si
Field Updates	Si	Si
Easy for Programmers	No	Si
Complex Transactions	Si	No
Auditing	Si	Si
Auto-Sharding	No	Si

**Comparativa de funcionalidades entre MySQL y MongoDB*

Entre las razones del uso de MongoDB se encuentra la velocidad, el alto manejo de diversos tipos de datos y la gestión más eficiente de escalar la arquitectura. Otro de los motivos fundamentales es la simplificación de la arquitectura. Al no trabajar con mapeado entidad relación, se elimina toda la complejidad que requiere la creación y planificación de una base de datos de estas características.

MongoDB es muy popular para aplicativos que necesitan bajas latencias y que además trabajan en tiempo real. Para un proyecto de estas características, que tiene como fin gestionar un almacén de datos, pero también datos en directo, MongoDB es la herramienta idónea.

3 ESTABLECIMIENTO DEL PROBLEMA

3.1 INTRODUCCIÓN

La cantidad de datos que se genera a día de hoy es elevada. Esta tendencia lejos de reducirse va a aumentar con los años. Datos ofrecidos por la OBS Business School estiman que en 2014 se transfirieron 1.570 Terabytes de información por minuto y que en 2020 el número de dispositivos conectados a internet será de 30 mil millones. A nivel técnico y científico la capacidad de gestionar y dar valía a esta información es todo un reto.

Este entorno está generando una inversión creciente en el sector del análisis de datos. El 73% de las organizaciones mundiales está invirtiendo o planificando grandes inversiones en este campo.

En el mundo de los mercados financieros se está produciendo este cambio de una manera acelerada y no existe una gran institución que no utilice herramientas tecnológicas en sus análisis, toma de decisiones, estudios o inversiones. Existen fondos de inversión que basan por completo su inversión en algoritmos. Para esta práctica es vital realizar un estudio de datos automatizado, exhaustivo y exitoso que sepa modelar esta información y generar rentabilidades económicas.

El primer reto o problema del desarrollo de una herramienta de estas características es la obtención de datos. Al no ser un agente o miembro del mercado se necesita un intermediario que te provea la información. La velocidad de recepción de información y procesamiento de toma de decisión es fundamental y lógicamente a nivel particular las latencias son más altas.

El segundo reto es obtener la información adecuada. La selección de los valores a estudiar junto a las variables que se van a obtener es importante. Por otra parte, el *timeframe* de la obtención de los datos es también fundamental.

Para que todo esto funcione es necesario el despliegue de una base de datos local que se adecue a los requisitos funcionales del aplicativo. El rendimiento y la velocidad de consultas de esta base de datos debe ser muy elevada.

La problemática principal es elaborar una estrategia de análisis de datos que sea capaz de acertar en las predicciones del mercado. Realizar un estudio de series temporales para modelar esta información y poder calcular patrones que posteriormente sean consultados para buscar semejanzas y así poder generar aciertos en la predicción.

4 SOLUCIÓN DEL PROBLEMA

4.1 SOLUCIÓN

A continuación, comienza la solución donde se expondrá elementos clave para la progresión del proyecto como son el proveedor de datos y la especificación de los requisitos funcionales del servicio.

4.1.1 Proveedor de datos

La obtención de los datos de mercado, se ha realizado mediante la conexión con un intermediario. Este bróker, es el proveedor de datos de mercado, así como que el canal de procesamiento de órdenes de compra o venta. El nombre de la empresa elegida es Interactive Brokers (IB).

Interactive Brokers es una empresa norte americana de servicios financieros. Se fundó en 1978 y es la empresa tecnológica del sector más antigua. Ofrece todo tipo de productos tales como acciones, opciones, mercado de divisas, futuros...

La particularidad diferencial en la elección de esta empresa es la API que Interactive Brokers pone a disposición de los desarrolladores y clientes para poder acceder a toda esta operativa. Esta interfaz permite desarrollo en los siguientes lenguajes de programación:

- Java
- Python;
- .NET (C#)
- C++
- ActiveX
- DDE

El lenguaje de programación elegido para el desarrollo es Java y como IDE eclipse. Para poder conectar nuestro proyecto en eclipse con los mercados financieros es necesario la Trading WorkStation (TWS) un programa desarrollado por Interactive Brokers que permite esta conexión.

4.2 ESPECIFICACIÓN DE REQUISITOS

Especificación de Requisitos				
Código	Nombre	Fecha	Grado de necesidad	
1	Conectar con los mercados	1-2-17	Alto.	
Descripción	La herramienta conectará con un proveedor de datos y gestionará la conexión.			
Entrada	Fuente	Salida	Destino	Restricciones
N/A	Proveedor de datos.	Confirmación de conexión levantada.	N/A	Conectar cuando los mercados a

				estudio estén abiertos.
Proceso	La herramienta a través de la API puesta a disposición por la empresa proveedora de datos gestionará la conexión con los mercados financieros conectando y desconectando cuando los mercados estén abiertos o cerrados. En caso de que falle la conexión gestionará la reconexión para la normal ejecución de la herramienta.			

Especificación de Requisitos				
Código	Nombre	Fecha	Grado de necesidad	
2	Obtener datos históricos.	1-2-17	Alto.	
Descripción	La herramienta conectará con un proveedor de datos y obtendrá los datos de mercado solicitados.			
Entrada	Fuente	Salida	Destino	Restricciones
Nombres de acciones, divisas o futuros.	Proveedor de datos.	Histórico de un mes en velas de cinco minutos.	Base de datos.	N/A
Proceso	La herramienta a través de la API puesta a disposición por la empresa proveedora de datos solicitará los datos históricos de las empresas deseadas. Estos datos se obtendrán y se almacenarán en la base de datos desplegada.			

Especificación de Requisitos				
Código	Nombre	Fecha	Grado de necesidad	
3	Obtener datos en directo.	5-2-17	Alto.	
Descripción	La herramienta conectará con un proveedor de datos y obtendrá los datos de mercado solicitados.			
Entrada	Fuente	Salida	Destino	Restricciones
Nombres de acciones, divisas o futuros.	Proveedor de datos.	Velas de cinco minutos.	Base de datos.	Solicitar información cuando los mercados se encuentren abiertos.
Proceso	La herramienta a través de la API puesta a disposición por la empresa proveedora de datos solicitará los datos de las empresas deseadas. Estos datos se obtendrán y se almacenarán en la base de datos desplegada. Los datos seguirán llegando cada cinco minutos mientras la herramienta siga en ejecución.			

Especificación de Requisitos				
Código	Nombre	Fecha	Grado de necesidad	
4	Calcular indicadores técnicos.	12-3-17	Medio.	
Descripción	La herramienta conectará con la base de datos y calculará datos estadísticos a partir de los datos de mercado obtenidos.			

Entrada	Fuente	Salida	Destino	Restricciones
Nombres de acciones, divisas o futuros.	Base de datos.	Indicadores técnicos.	Base de datos.	N/A
Proceso	La herramienta accederá a la base de datos y para cada empresa, divisa o futuro calculará sus datos técnicos. Los resultados se almacenarán en la base de datos. Este proceso ejecutara las operaciones con los datos históricos y posteriormente con datos en directo.			

Especificación de Requisitos				
Código	Nombre	Fecha	Grado de necesidad	
5	Obtener datos fundamentales.	5-2-17	Baja.	
Descripción	La herramienta conectará con un proveedor de datos y obtendrá los datos fundamentales solicitados.			
Entrada	Fuente	Salida	Destino	Restricciones
Nombres de acciones.	Proveedor de datos.	Datos fundamentales.	Base de datos.	No existen datos fundamentales para divisas o futuros.
Proceso	La herramienta a través de la API puesta a disposición por la empresa proveedora de datos solicitará los datos de las empresas deseadas. Estos datos se obtendrán y se almacenarán en la base de datos desplegada. Los datos se actualizarán mensualmente.			

Especificación de Requisitos				
Código	Nombre	Fecha	Grado de necesidad	
6	Calcular patrones.	26-4-17	Alto.	
Descripción	La herramienta conectará con la base de datos y calculará los patrones definidos a partir de los datos de mercado obtenidos.			
Entrada	Fuente	Salida	Destino	Restricciones
Nombres de acciones, divisas o futuros.	Base de datos.	Indicadores técnicos.	Base de datos.	N/A
Proceso	La herramienta accederá a la base de datos y para cada empresa, divisa o futuro calculará su patrón. Los resultados se almacenarán en la base de datos. Este proceso ejecutara las operaciones con los datos históricos y posteriormente con datos en directo.			

4.3 DISEÑO

En diseño explicaremos información sobre los niveles de la arquitectura, así como el flujo lógico de transferencia de los datos.

4.3.1 Niveles de la arquitectura

El aplicativo a desarrollar se trata de un servicio que a partir de unas entradas de datos es capaz de generar y calcular unos resultados explicados más adelante. Es por ello que podemos definir una estructura basada en niveles:

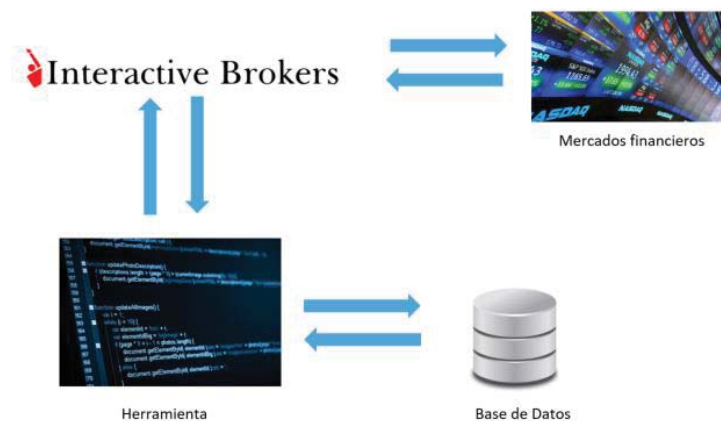


(*) Niveles de la arquitectura

- Nivel de servicio: donde se invoca a las funcionalidades de la herramienta. Se introducen los valores o empresas y comienza el análisis.
- Nivel de análisis: donde están implementados los algoritmos y operaciones a utilizar para el modelado de la información.
- Nivel de datos: almacena, gestiona y recoge nuevos datos de los mercados financieros.

4.3.2 Arquitectura

La arquitectura planteada es la siguiente:



Podemos designar como nivel de datos la base de datos y la empresa a la cual hacemos las consultas de datos. En la parte de análisis la herramienta desarrollada es la encargada de obtener esta información y modelarla según la solución planeada. La herramienta a través del nivel de servicio permite al usuario ejecutar la operativa y el análisis objeto de este trabajo de fin de grado.

4.4 IMPLEMENTACIÓN

En este apartado explicaremos la manera en la cual se ha implementado el servicio objetivo de este trabajo de fin de grado.

4.4.1 Gestión de la información

A continuación, las operaciones que hacen posible la consulta y obtención de la información así como la lógica en el modelado de los datos.

4.4.1.1 Históricos de mercado

Existen diferentes tipos de datos de mercado que se pueden consultar en IB. En primer lugar, los históricos de mercado. Mediante la operación `reqHistoricalData()`, IB devuelve los datos históricos de mercado desde una fecha introducida hasta el momento actual. Es importante almacenar y gestionar el histórico para el futuro análisis. Concretamente las variables que brinda IB son los siguientes:

Parameter	Type	Description
<code>reqId</code>	int	The ticker Id of the request to which this bar is responding.
<code>date</code>	String	The date-time stamp of the start of the bar. The format is determined by the <code>reqHistoricalData()</code> <code>formatDate</code> parameter.
<code>open</code>	double	The bar opening price.
<code>high</code>	double	The high price during the time covered by the bar.
<code>low</code>	double	The low price during the time covered by the bar.
<code>close</code>	double	The bar closing price.
<code>volume</code>	int	The volume during the time covered by the bar.
<code>count</code>	int	When TRADES historical data is returned, represents the number of trades that occurred during the time period the bar covers
<code>WAP</code>	double	The weighted average price during the time covered by the bar.
<code>hasGaps</code>	boolean	Whether or not there are gaps in the data.

Esta llamada trabaja con velas, es decir, agrupa en un `timeframe*` determinado (especificado en la llamada) la fluctuación del valor consultado. Datos a destacar y que son de relevancia para el estudio

1. Id.
2. Fecha del dato.
3. Valor en la apertura de la vela.
4. Valor en el cierre de la vela.
5. El precio máximo del valor en la vela.
6. El precio mínimo del valor en la vela.

7. El volumen negociado.
8. El contador de transacciones realizadas.

4.4.1.2 Datos en directo

Una vez obtenido el histórico, empezaremos a trabajar con datos en directo. Aquí existen dos tipos de llamadas diferentes. La primera de ellas es `reqRealTimeBars()` que trabaja con las mismas variables y velas que `reqHistoricalData`. Esta operación es recursiva y devuelve cada vez que se cumple el tiempo del frame los datos de la nueva vela.

En segundo lugar tenemos `reqMktData()`. Esta llamada tiene una visión más general y trabaja con el valor actual de la acción en el momento de la consulta. Es importante destacar, que el precio de la acción tiene dos valores. El primero es el de compra (Bid) y el segundo el de venta (Ask). Estos dos precios fluctúan y nunca son el mismo. La diferencia entre estos dos valores se conoce como Gap. Los datos que nos devolverá son:

1. Id de la consulta.
2. Precio de compra.
3. Precio de venta.
4. Último precio medio.
5. Precio máximo de la sesión.
6. Precio mínimo de la sesión.
7. Precio de cierre de la última sesión y por tanto precio de la apertura de la actual sesión.

4.4.1.3 Datos fundamentales

En los análisis y estudios bursátiles financieros se estudia también lo denominado como fundamentales. A partir de este análisis obtenemos ratios que son de vital importancia en el estudio de los mercados financieros. Estas ratios, se calculan partir de datos contables de la empresa, valores de mercado, coyuntura macroeconómica... IB con la operación `reqFundamentalData()` da acceso a esta información.

4.4.1.4 Datos ligados a las estrategias y al valor de cuenta

Este trabajo de fin de carrera tiene como objetivo el estudio del comportamiento de los mercados financieros y crear estrategias aprovechen este estudio realizado. Almacenaremos en la base de datos el histórico de todas las órdenes de compra venta realizadas. Datos importantes a almacenar:

- Id de la orden.
- Fecha de compra.
- Fecha de venta.
- Empresa o índice donde se realizó la orden.
- Dinero invertido.
- Ganancias o pérdidas de la operación.

- Estrategia que realizo la operación. Quizás sería más eficiente para la gestión separar la ordenes por estrategias en diferentes tablas para hacer un estudio más concreto del comportamiento de la estrategia.

Por último, los datos relacionados con el valor contable de la cuenta que nos permita realizar Money Management. Con la operación reqAccountUpdates(), IB nos devuelve la situación económica actual de la cuenta. Datos importantes:

- Cash Balance: dinero líquido disponible y que por lo tanto esta fuera del mercado.
- NetLiquidation: valor de la cuenta si se cierran todas las posiciones. Valor liquidativo de la cuenta.

4.4.2 Almacenamiento de datos

En este proyecto se han desplegado diferentes bases de datos con diferentes finalidades. Cada una de ellas trabaja con colecciones que se refieren a los datos en particular de una empresa o divisa a analizar. Las dependencias instaladas son las siguientes:

- MarketData: cada empresa, dentro del estudio realizado, tiene su propia colección y almacena los datos de mercado en *velas temporales** de 5 minutos. En la primera ejecución de la herramienta, se realiza una consulta de datos financieros históricos de hace un mes hasta la actualidad. Posteriormente, se concatena las velas del histórico con las velas obtenidas en directo. Por motivos técnicos, el proveedor de datos IB limita las consultas de históricos de velas de 5 minutos a un mes. Al tratarse de velas de la misma dimensión temporal se pueden concatenar. Los campos que vamos a almacenar de datos en directo como de datos históricos son:
 1. Time: fecha con formato “yyyymmdd hh:mm:ss” del dato de mercado.
 2. Open: precio al que abrió la vela.
 3. High: precio más alto de la vela.
 4. Low: precio más bajo de la vela.
 5. Close: precio al que cerro la vela.
 6. Volume: volumen monetario de la vela.
 7. Wap: precio promedio ponderado de la vela.
 8. Count: número de acciones negociadas en la vela.
- FundamentalData: es una base de datos que almacena los ratios fundamentales de las empresas a analizar (divisas y futuros no entran dentro de este análisis). Cada empresa tiene su colección. La llamada a este servicio es mucho más pasiva que la anterior ya que los ratios fundamentales varían con menos frecuencia. Los campos o ratios a almacenar son:
 1. NetProfitMargin: ganancias después de impuestos. Campo expresado en porcentaje.
 2. LowPrice: precio de la acción más bajo en su histórico de 12 meses.
 3. RevenueGrowtRate: tasa de crecimiento de ingresos.
 4. EPS: ingreso ajustado de los accionistas comunes.
 5. BookValuePerShareTangible: valor contable de los bienes tangibles de la empresa dividido por el número de acciones de la empresa.

6. PriceCashFlowPerShare: precio actual de la acción dividido por el dinero líquido por acción de la empresa.
7. GrossMargin: porcentaje de ingresos después de pagar todos los gastos directos de producción.
8. CashFlowPerShare: valor de *flujo efectivo** de doce meses dividido por el valor promedio de acciones en circulación.
9. CurrentRatio: proporción de *activos corrientes** dividido por el total de *pasivos corrientes**.
10. PriceBook: precio actual la actual dividido por el BookValuePerShareTangible.
11. Revenue: suma de todas las partidas de ingresos de la empresa.
12. MarketCap: número de acciones de la empresa multiplicado por el valor de la acción. Capitulación bursátil de la empresa.
13. PayoutRatio: porcentaje de ingresos primarios/básicos por acción. Excluidos dividendos.
14. OperatingMargin: ingresos después de todos los gastos operativos.
15. RevenueChange: porcentaje que calcula la variación anual de ventas.
16. ReturnAverageAssets: ganancias después de impuestos dividido por el número de acciones de la empresa
17. EPSNorm: ingreso normalizado de los accionistas comunes.
18. ReturnAverageEquity: ingresos disponibles para los accionistas comunes.
19. PER: precio actual de la acción dividido por los gastos por acción de la empresa.
20. EPSChange: cambio porcentual del EPS en un año.
21. ReturnInvestment: valor que divide las ganancias después de impuestos por el promedio del valor de la deuda a largo plazo.
22. RevenueChangePorc: cambio porcentual de las ganancias en un año.
23. Currency: divisa.
24. GrowthRateDividend: ratio del rendimiento de los dividendos de la empresa.
25. EPSChange1
26. CashPerShare: dinero líquido de la empresa dividido por el número de acciones.
27. Date: fecha de los datos
28. PreTaxMargin: ganancias antes de impuestos.
29. PerExclItem: precio de la acción dividido por los gastos de la empresa excluyendo elementos extraordinarios.
30. QuickRatio: dinero líquido más inversiones a corto plazo más cuentas por cobrar dividido por el *total de pasivos corrientes**.
31. EarningsBeforeTaxesNorm: ganancias antes de impuestos normalizados
32. RevenueShare: ingreso total de doce meses, dividido por el promedio de acciones para los doce meses siguientes
33. NetIncomeAvailable: acumulado de dólares ganados por los accionistas por dividendo en los últimos 12 meses.
34. EarningsBeforeTaxes: ganancias antes de impuestos.
35. EBITD: ganancias antes de inversiones, impuestos, depreciación y amortizaciones.
36. PriceTangibleBook: precio de los valores tangibles que posee la empresa.
37. NetIncomeAvailableNorm: NetIncomeAvailable normalizado.
38. EPSGrowthRate
39. LTTotalDebtEquity deuda a largo plazo dividido por el *patrimonio neto**.
40. TotalDebtEquity: deuda a corto plazo dividido por el *patrimonio neto**.
41. BookValuePerShare: valor contable de los libros de la empresa dividido por las acciones.

42. ClosingPrice: precio de la acción.
 43. HighPrice: precio de la acción más alto en su histórico de 12 meses.
- TechnicalData: es la base de datos que almacena lo que en el sector financiero se denomina como datos técnicos. El análisis técnico* hace uso de datos almacenados en MarketData y realiza en directo con ellos diferentes operaciones para cada indicador técnico. Al ser en directo, este servicio va a ir de la mano en la ejecución con la consulta de datos de mercado. Del mismo modo se desarrollará un proceso que calcule los datos técnicos del histórico almacenado. El resultado de estas operaciones es lo se almacena en la base de datos. Los indicadores técnicos que vamos a utilizar son:
 1. RSI.
 2. Media Móvil.
 3. Media Móvil Exponencial.
 4. MACD.
 5. ATR.
 6. Bears Power.
 7. Bulls Power.
 8. ROC.
 9. CCI.
 10. ADX.
 - Porfolio: base de datos que almacena la situación financiera de la cuenta desde la cual se está realizando la operativa. Los campos que vamos a cargar a la base de datos son:
 1. CashBalance: dinero líquido que tiene la cuenta.
 2. BuyingPower: dinero disponible para operar.
 3. EquityWithLoanValue: patrimonio con valor de prestamo
 4. InitMarginReq: margen inicial de la cuenta.
 5. MaintMarginReq: margen de mantenimiento de la cuenta.
 6. NetLiquidation: valor liquidativo de la cuenta.
 - Positions: base de datos que almacena las posiciones abiertas y cerradas.
 1. Contract: símbolo de la empresa.
 2. Position: número de acciones adquiridas.
 3. MarketPrice: precio del instrumento negociado.
 4. MarketValue: valor de mercado.
 5. AverageCost: coste de la acción más comisiones dividido por el número de acciones compradas.
 6. UnrealizedPNL: diferencia entre costes y valor de mercado de la posición.
 7. RealizedPNL: ganancias de las operaciones cerradas.
 8. Fecha: fecha de los datos.

4.4.3 Análisis Técnico

En el análisis técnico es un método de evaluación de los mercados financieros basado en el uso de la estadística asociada a la actividad de la fluctuación, usando datos tales como precio, volumen...

El análisis técnico no realiza mediciones en busca de un valor objetivo o intrínseco del valor en cuestión, sino que utiliza datos de gráficos para buscar patrones de comportamiento. Estos patrones son usados para tomas de decisiones en inversiones. En esta metodología hay diferentes formas de analizar. En primer lugar, los patrones de comportamiento comentados anteriormente y por otro lado los indicadores u osciladores. Los indicadores u osciladores se calculan operando con los datos de mercado.

4.4.3.1 Máximas del análisis técnico

Existen tres suposiciones en las que se basa el análisis técnico:

1. Los mercados descuentan absolutamente todo. Esta teoría se la denomina como hipótesis del mercado eficiente y expone que el precio de los mercados ya está cuadrados, tasados o ajustados a toda la casuística a la que le empresa o valor se enfrenta. Por tanto, todos los movimientos del mercado se basan únicamente en un producto de oferta y demanda de la acción en particular.
2. El precio se mueve sobre tendencias. Estas pueden ser de corto, medio o largo plazo y muchas veces se auto contienen unas a otras.
3. Las tendencias tienden a replicarse y a repetirse a lo largo del histórico.

4.4.3.2 Indicadores u osciladores utilizados

Los indicadores que se han calculado y almacenado en la base de datos, como hemos visto anteriormente, son los siguientes:

1. RSI: índice de fuerza relativa (Relative Strength Index). Indicador tipo oscilador que muestra la fuerza del precio mediante la comparación de movimientos individuales al alza o la baja de sucesivos precios de cierre.

$$RSI = 100 - 100 / \left(1 + \left(\frac{\text{Media Subidas}}{\text{Media Bajadas}} \right) \right)$$

2. Media Móvil: cálculo utilizado para analizar un conjunto de datos en modo de puntos para crear series de promedios.

$$\text{Media Móvil} = \frac{n + (n + 1) + \dots + (n + x)}{\text{Número de elementos}}$$

3. Media Móvil Exponencial: media móvil ponderada exponencialmente. Esta media se calcula usando la anterior media exponencial calculada.

$$\text{Multiplicador} = \frac{2}{(\text{Periodos de tiempo} + 1)}$$

$$EMA = \text{Media Móvil Exponencial calculada anteriormente}$$

$$\text{Media Móvil Exponencial} = (\text{Precio} - \text{EMA}) * \text{multiplicador} + \text{EMA}$$

4. Bears Power: determina el poder bajista del mercado. Para su cálculo se utiliza el precio más bajo de la vela y la media móvil exponencial que por tanto ha tenido que haber sido calculado anteriormente.

$$\text{Bears Power} = \text{Precio mínimo de la vela} - \text{Media Móvil Exponencial}$$

5. Bulls Power: determina el poder al alza de los mercados. Para su cálculo se utiliza el precio más alto de la vela y la media móvil exponencial que por tanto ha tenido que haber sido calculado anteriormente.

$$\text{Bulls Power} = \text{Precio máximo de la vela} - \text{Media Móvil Exponencial}$$

6. CCI (Commodity Chanel Index) es utilizado para identificar una nueva tendencia como para calcular condiciones extremas del mercado. Para el cálculo se utiliza los precios máximos y mínimos de la vela, como el precio de la vela, así como la media móvil.

$$\text{Constante} = 0.015$$

$$\text{Precio Típico} = \frac{\text{Precio máximo} + \text{Precio mínimo} + \text{Precio}}{3}$$

$$\text{CCI} = \frac{\text{Precio Típico} - \text{Media Movil}}{\text{Constante} - \text{Media Movil}}$$

4.4.4 Análisis Fundamental

El análisis fundamental es la técnica que busca determinar cuál es el valor objetivo o intrínseco de la acción. Para ello se centra no es el gráfico de la fluctuación de la acción, sino en datos contables, situación del negocio o perspectivas futuras del negocio.

Esta metodología de análisis es mucho más extensa en el número de variables que se estudian que el análisis técnico. En el análisis fundamental introduce al estudio no solo los datos relacionados con la empresa, sino que también datos de otras empresas, sector del negocio o coyuntura macroeconómica nacional e internacional. Por tanto, el estudio es mucho más complejo y minucioso.

El análisis fundamental se puede agrupar en dos categorías:

1. Análisis fundamental cualitativo.
2. Análisis fundamental cuantitativo.

4.4.4.1 Análisis fundamental cualitativo.

El análisis cualitativo se centra en asuntos como la naturaleza del negocio, posición de la compañía dentro de la industria, sus características físicas, geográficas y operativas, carácter de la directiva y, finalmente, las perspectivas de la compañía, de su industria y de los negocios en general. Estos asuntos son difícilmente plasmables en métricas y números como si es posible en el análisis fundamental cuantitativo. Esto se debe a que nos es información o datos tangibles.

La capacidad de realizar un estudio informático y por tanto automatizado de un análisis de estas características es mucho más complejo y no entra dentro del alcance del desarrollo realizado en este proyecto de fin de grado.

4.4.4.2 Análisis fundamental cuantitativo.

Este análisis es el que mejor se ajusta a las características de este proyecto de fin de carrera dentro del análisis fundamental. El análisis fundamental cuantitativo genera y utiliza métricas para él estudio. La exposición estadística de la empresa. A diferencia del análisis técnico, en esta elaboración de métricas se utilizan variables como pérdidas y ganancias, unidades producidas, capacidad de producción...

Las variables del análisis fundamental cuantitativo se etiquetan en diferentes subcategorías:

1. Capitalización: datos con el valor de la compañía.
2. Beneficios y dividendos.
3. Activos y pasivos.
4. Estadísticas operativas.

Este tipo de datos son con los que se han utilizado en el estudio objeto de este proyecto. En el punto “5.4 Datos almacenados”, se explica que variables se obtienen, así como el significado de cada una de las variables utilizadas.

4.4.5 Patrones

En la herramienta desarrollada se ha desplegado una colección de la base de datos exclusiva para patrones. Al contrario que en otras colecciones desplegadas, en esta se almacenan juntos todos los patrones de empresas, divisas y futuros del estudio. La razón es que entendemos que estos patrones se pueden replicar independientemente de la acción en cuestión y que además técnicamente nos permitirá una mayor velocidad de consulta de la información.

Cada patrón va a constar de una actividad de 20 minutos ya que va a generar el patrón a partir de 4 velas temporales de 5 minutos cada una. De la primera vela a la quinta, posteriormente de la segunda vela almacenada a la sexta y así sucesivamente.

A continuación, se muestra un ejemplo de un registro en la base de datos de un patrón:

```

_id: ObjectID('590b70a99c8cd25fa1d8183e')
Company: "GOOG"
▼ Pattern: Array
  ▼ 0: Array
    0: "BAJA"
    1: "SUBE"
    2: "BAJA"
    3: "BAJA"
    4: "SUBE"
    5: "BAJA"
    6: "SUBE"
  ▼ 1: Array
    0: "SUBE"
    1: "BAJA"
    2: "BAJA"
    3: "BAJA"
    4: "BAJA"
    5: "BAJA"
    6: "BAJA"
  ▼ 2: Array
    0: "BAJA"
    1: "SUBE"
    2: "SUBE"
    3: "SUBE"
    4: "BAJA"
    5: "SUBE"
    6: "IGUAL"
  ▼ 3: Array
    0: "SUBE"
    1: "BAJA"
    2: "BAJA"
    3: "BAJA"
    4: "SUBE"
    5: "BAJA"
    6: "SUBE"
Date: "20170405 15:50:00"
Profit: 0.02
Calculate: true

```

Los campos que se almacenan en el registro de un patrón son:

1. El primer campo “Company” se trata de las siglas de la empresa, cruce de divisa o futuro del patrón en cuestión.
2. El segundo campo es la matriz o el patrón propiamente dicho.
3. La fecha con la hora exacta en la que se registró dicho patrón.
4. El campo “Profit” que se trata del porcentaje que vario la fluctuación del valor 5 minutos después de suceder el patrón.
5. El ultimo campo registra si se ha calculado el campo “Profit” o no. Esta variable se utiliza fundamentalmente para los patrones que se generan en directo y no con el histórico. Cuando llegan los datos en directo y se genera el patrón se almacena sin el “Profit” ya que aún no ha sucedido. Cuando dentro de 5 minutos llega el siguiente dato de mercado, este se encarga de actualizar en base de datos este campo no sin antes calcular el porcentaje de variación del mercado.

4.4.5.1 Matriz del patrón

El campo que almacena el patrón se trata de una matriz de 4X7. Una de las bondades de MongoDB es que permite almacenar este tipo de estructuras de datos de una manera simple. En las filas de la matriz figuran las diferentes velas con su timeframe específico. En las columnas las variables asociadas a las velas:

1. Open: precio al que abrió la vela.
2. High: precio más alto de la vela.
3. Low: precio más bajo de la vela.
4. Close: precio al que cerró la vela.
5. Volume: volumen monetario de la vela.
6. Wap: precio promedio ponderado de la vela.
7. Count: número de acciones negociadas en la vela.

A través de una lógica que compara el valor de los campos de la vela anterior con la siguiente se genera una fila con valores que pueden tratarse de “SUBE”, “BAJA” o “IGUAL”. Estos valores resultado de la comparación es lo que se almacena en la base de datos.

Los valores numéricos se han abstraído de la lógica del patrón por diversas razones. La primera de ellas es que el valor exacto con decimales de la acción es difícilmente replicable y repetible. Lo que buscamos analizar es el comportamiento de subidas o bajadas, ósea un patrón propiamente dicho. La segunda de las razones es que los patrones puedan servir para todos los valores o acciones del estudio. Calcular patrones con números impediría realizar patrones genéricos para analizar diversos mercados y valores.

4.4.6 Estructura lógica de la herramienta

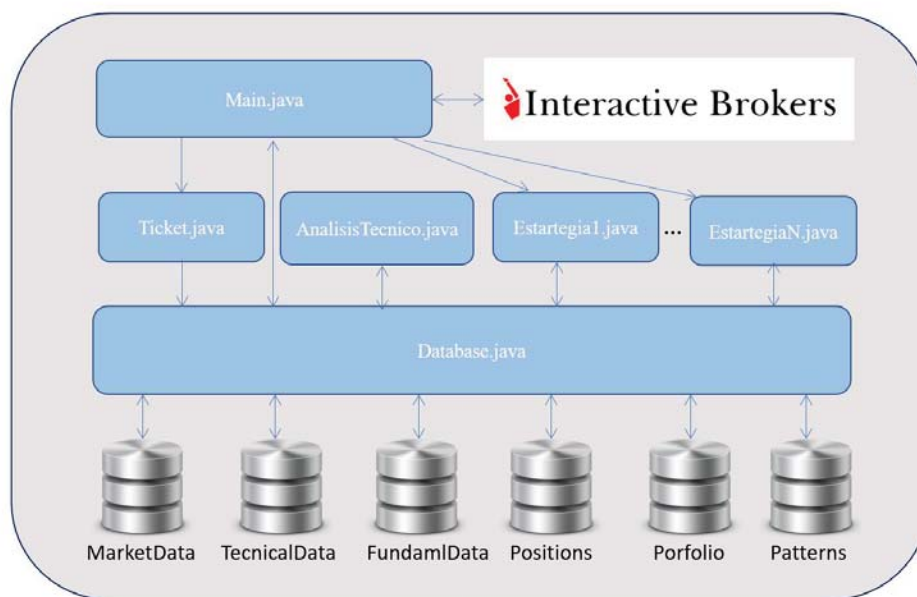
1. Main.java: como programa principal invoca y gestiona todas las operaciones de la herramienta.
2. Database.java: nos permite interactuar y comunicar nuestra herramienta con la base de datos desplegada. Esta clase contiene todas las operaciones que levantan las conexiones con las diferentes estancias de la base de datos, así como todas las operaciones de inserción o consulta de datos en nuestra base de datos. Del mismo modo también realiza las operaciones internas de la base de datos como por ejemplo el cálculo de los datos fundamentales o cálculo de patrones. Por último, también permite la desconexión con la base de datos.
3. AnalisisTecnico.java: contiene todas las operaciones que calculan los datos técnicos del mercado.
4. Ticket.java: es la clase que contiene el objeto modelo de datos a utilizar en el proyecto para los datos de mercado

```

public Ticket(String name,int id,String time,double open,double high,double low,
              double close,double volume,double wap,double count){
    this.name=name;
    this.id=id;
    this.time=time;
    this.open=open;
    this.high=high;
    this.low=low;
    this.close=close;
    this.volume=volume;
    this.wap=wap;
    this.count=count;
}

```

4.4.7 Framework



(*) Framework de la herramienta.

La clase “Main.java” es la que se encarga de gestionar y llamar a todos los servicios disponibles en la herramienta. Como podemos ver en el diagrama, los datos llegan directamente de Interactive Brokers donde la clase main crea el objeto “Ticket” para que finalmente la clase “Database” la almacene. Como podemos ver la clase “Database.java” es una capa que gestiona todas las comunicaciones con la base de datos.

En el framework podemos ver reflejadas las diferentes estancias de la base de datos que han sido explicadas anteriormente.







En las clases “estrategias.java” se quiere plasmar las operaciones que podrán aprovechar el análisis de minería de datos y de predicción de series temporales realizado.

5 VALORES

A continuación, se enumerarán los valores objeto de este estudio. El análisis se va a centrar en acciones de empresas tecnológicas, futuros y en Forex o mercado de divisas. Las empresas tecnológicas son de mucha popularidad a nivel mundial y por tanto gozan de una liquidez y un movimiento de mercado que es necesario para realizar un estudio de estas características. En cuanto a futuros, se puede decir que es uno de los mercados más líquidos que existe al igual que los mercados de divisas. Este último, goza de la particularidad que es un mercado continuo, no cierra nunca. Estudiar qué tipo de mercado se adapta mejor a esta operativa también es un objetivo de este trabajo de fin de grado.

- Las empresas seleccionadas son:

Google	Buscador, proveedor de servicios en internet y desarrollador del sistema operativo Android.	
Apple	Empresa que comercializa principalmente dispositivos móviles y ordenadores con su propio sistema operativo.	
Amazon	Empresa de e-commerce además de proveedor de soluciones tecnológicas para empresas y particulares.	
Ebay	Página de subastas y de comercio electrónico.	
AMD	Empresa de desarrolladora de componentes. Desarrolla procesadores, tarjetas gráficas...	
Nvidia	Empresa popular por el desarrollo de tarjetas gráficas.	
Facebook	Red social.	
Netflix	Empresa de contenido multimedia bajo demanda.	
Twitter	Red social.	

IBM	Empresa productora de ordenadores y de soluciones tecnológicas.	
Snapchat	Red social.	
Qualcomm	Empresa desarrolladora de procesadores para terminales móviles	
Microsoft	Desarrolladora del sistema operativo Windows.	
Cisco	Empresa desarrolladora de componentes fundamentalmente destinados a las telecomunicaciones.	
Tesla	Empresa pionera y líder en la construcción de coches eléctricos, energía verde además de apostar por la introducción de coches autónomos.	

- En el mercado de futuros vamos a introducir al estudio el S&P500 e-mini. El S&P500 o Standard & Poors' 500 es uno de los índices mas importantes de los Estados Unidos. Este índice aúna la fluctuación de 500 grandes empresas que cotizan en las bolsas del NYSE o NASDAQ. Concretamente vamos a estudiar el S&P500 e-mini que es un futuro sobre el S&P500 que se comercializa en el Stock Exchange de Chicago, el mercado de futuros más importante del mundo. Este futuro en concreto es el más comercializado del mundo, por tanto, es de un interés elevados para este estudio ya que en su fluctuación se concretan muchas transacciones.
- Forex o mercado de divisas. Cada valor enfrenta dos divisas. Este mercado es continuo no cierra y es descentralizado. Varios actores aúnan el comercio de divisas, pero ninguno en particular como si ocurre con las acciones, futuros, opciones... Las pares que se van a estudiar son.

EUR-USD	Euro frente al dólar
USD-CAN	Dólar frente al dólar canadiense
USD-CHF	Dólar frente al franco suizo
USD-HKD	Dólar frente al dólar de Hong Kong
USD-JPY	Dólar frente al yen japonés
USD-KRW	Dólar frente al Won surcoreano
USD-MXN	Dólar frente al peso mexicano
USD-RUB	Dólar frente al rublo ruso
USD-ZAR	Dólar frente el Rand sudafricano
AUD-USD	Dólar australiano frente dólar
GBP-USD	Libre inglesa frente dólar
NZD-USD	Dólar neozelandés frente al dólar.

6 DISCUSIÓN

La práctica de la inversión automatizada y de alta frecuencia es un tema que está en boca de todos en la actualidad. ¿Queremos unos mercados financieros operados únicamente por servidores?

En el lado positivo, la práctica del trading de alta frecuencia aporta a los mercados una mayor liquidez lo cual es importante destacar. Además, mejora la oferta de compra-venta de acciones ya que reduce el *spread* (diferencia entre valor de compra y venta).

En la cara opuesta de la moneda, la negativa de los operadores intradía que ven como es más complicado obtener retornos con estos agentes de mercado operando. Se opina, además, que estas prácticas pueden crear un ecosistema de mercado mucho más *volátil* y peligroso para la economía mundial. Este peligro se vio reflejado en el “flash crash” del 6 de mayo del 2010 a las 2:45 donde el índice Dow Jones Industrial registro un desplome de aproximadamente un 9% por un algoritmo que estaba mal programado.

Es importante recalcar que con estos agentes del mercado funcionando, el efecto ha sido contrario a lo que muchos analistas opinaban y predecían. La volatilidad de los mercados esta en mínimos históricos y es que los algoritmos trabajan con rentabilidad muy bajas, pero con una operativa de miles de operaciones por segundo esto produce un efecto “allanador” en la volatilidad de los mercados. A continuación, gráfico del VIX el índice más importante en la representación de la volatilidad en los mercados.



Gráfico VIX S&P500

Como ya se mencionó anteriormente, la transformación de este sector es ya una realidad. Es por eso que el perfil profesional del sector también ha evolucionado. Podemos hablar de tres perfiles necesarios:

1. Traders: personas especializadas en conocimientos financieros, contabilidad, mercados, bolsa y economía en general.
2. Quant Traders: personas del ámbito científico-técnico que sean capaces de crear los modelos matemáticos y estadísticos necesarios para esta práctica.
3. Expertos e ingenieros tecnológicos en sistemas, data science y big data que sean capaces de dar cabida tecnológica a toda esta operativa.

El estudio de los datos es ya una pieza fundamental no solo en el ámbito de estudio planteado para este trabajo de fin de grado sino en muchos otros sectores. Lejos de tratarse de máquinas operando automáticamente, se puede deslumbrar como realmente detrás de toda esta inteligencia y tecnologías hay personas trabajando y que el trabajo de los operadores de bolsa no va a desaparecer, sino que presumiblemente se va a transformar.

7 CONCLUSIÓN Y LÍNEAS FUTURAS

7.1 CONCLUSIÓN

A la hora de desarrollar una herramienta de análisis de mercados financieros, el tipo de análisis que mejor se ajusta es el técnico. Esta inversión se enfocada a periodos de corto medio plazo. Sin embargo, el análisis fundamental se enmarca dentro de una inversión que suele estar enfocada principalmente al largo plazo.

Como se ha visto anteriormente, el análisis técnico trabaja con indicadores o patrones, algo que desde el punto vista informático es mucho más factible de traducir a una lógica de programa. Dentro del ámbito de la inversión automática y de alta frecuencia, como su propio nombre indica, se realizan inversiones a muy corto plazo y con una velocidad de ejecución muy elevada. Por tanto, los ordenadores se ajustan perfectamente a estos dos requerimientos. Un análisis de tan corto plazo solo es posible con un análisis como el técnico.

A la hora de realizar un estudio basado en predicción de series temporales hemos estudiado como la traducción desde números a símbolos es mucho más factible. Este método de análisis perdería mucho sentido si todos los agentes del mercado utilizaran el mismo sistema de predicción. Para conseguir retornos otro tiene que sufrir pérdidas. Debido a esto un algoritmo de predicción nunca conseguirá ganar todas las operaciones que ejecute, pero mediante un proceso de optimización, se puede lograr que el número de aciertos sea mayor al número de errores en el pronóstico de la fluctuación del mercado. Del mismo, hay eventos que tienen un impacto muy alto en los mercados y que son impredecibles.

Por su parte, el análisis fundamental cuantitativo trabaja con métricas que también pueden ser utilizadas fácilmente por una herramienta de estas características. Sin embargo, los datos que se obtienen de este análisis son mucho más estáticos y no requieren de una velocidad de ejecución alta.

7.2 LÍNEAS FUTURAS

A pesar de haber conseguido alcanzar los objetos, este trabajo de fin de grado deja abiertas las siguientes líneas futuras:

- Extender el rango de análisis a mas valores, mercados, ventanas temporales y datos.
- Realizar pruebas para evaluar el comportamiento y rendimiento de las series temporales en un periodo de largo plazo.
- Optimizar el análisis para un mayor éxito en la predicción de la fluctuación del mercado.
- Comprobar el comportamiento del sistema siendo ejecutado en un servidor 24X7.

8 BIBLIOGRAFÍA

Amazon Web Services, Inc. (2017). *¿Qué es NoSQL? – Amazon Web Services (AWS)*. [online] Available at: <https://aws.amazon.com/es/nosql/> [Accessed 4 Jun. 2017].

Anon, (2017). *Temporal pattern mining in symbolic time point and time interval data*. [online] Available at: <https://www.siam.org/meetings/sdm11/moerchen.pdf> [Accessed 4 Jun. 2017].

Bbvacontuempresa.es. (2017). *¿Qué es el High Frequency Trading?*. [online] Available at: <http://www.bbvacontuempresa.es/a/high-frequency-trading-que-es> [Accessed 4 Jun. 2017].

Bolsawallstreet.com. (2017). *Medias Móviles Simples y Exponenciales*. [online] Available at: <http://bolsawallstreet.com/medias-moviles-simples-y-exponenciales/> [Accessed 4 Jun. 2017].

Bolsawallstreet.com. (2017). *Medias Móviles Simples y Exponenciales*. [online] Available at: <http://bolsawallstreet.com/medias-moviles-simples-y-exponenciales/> [Accessed 4 Jun. 2017].

Chu, B. (2017). *This is how 'the Hound of Houndslow' could have caused the flash crash - one of the scariest days in stock market history*. [online] The Independent. Available at: <http://www.independent.co.uk/news/business/news/flash-crash-explainer-how-the-hound-of-houndslow-could-have-caused-one-of-the-scariest-days-in-stock-10198805.html> [Accessed 4 Jun. 2017].

Economipedia. (2017). *Cómo se calcula el RSI - Ejemplo*. [online] Available at: <http://economipedia.com/definiciones/como-se-calcula-el-rsi-ejemplo.html> [Accessed 4 Jun. 2017].

ESTRATEGIAS DE TRADING. (2017). *Bandas de Bollinger: volatilidad con desviación estándar*. [online] Available at: <https://estrategiastrading.com/bandas-de-bollinger-volatilidad-con-desviacion-estandar/> [Accessed 4 Jun. 2017].

ESTRATEGIAS DE TRADING. (2017). *El indicador ATR (Average True Range) : medir la volatilidad*. [online] Available at: <https://estrategiastrading.com/indicador-atr/> [Accessed 4 Jun. 2017].

Fundssociety.com. (2017). *El papel del High-Frequency Trading en el mercado bursátil / Funds Society*. [online] Available at: <http://www.fundssociety.com/es/noticias/mercados/el-papel-del-high-frequency-trading-en-el-mercado-bursatil> [Accessed 4 Jun. 2017].

Google.es. (2017). *calcular fiboanci - Buscar con Google*. [online] Available at: https://www.google.es/search?q=calcular+fiboanci&ie=&oe=#q=como+calcular+fibonacci+finanza&*&spf=1 [Accessed 4 Jun. 2017].


Mercados, B. and tendencias, A. (2017). *ADX: el indicador técnico identificador de tendencias / Estrategias de Inversión*. [online] Estrategiasdeinversion.com. Available at: <http://www.estrategiasdeinversion.com/invertir-corto/analisis/20140211/adx-indicador-tecnico-identificador-tendencias> [Accessed 4 Jun. 2017].

Metatrader5.com. (2017). *Bulls Power - Osciladores - Uso de indicadores técnicos - Gráficos de cotizaciones, análisis técnico y fundamental - Guía de ayuda de MetaTrader 5*. [online] Available at: <https://www.metatrader5.com/es/terminal/help/indicators/oscillators/bulls> [Accessed 4 Jun. 2017].

Mongodb.github.io. (2017). *Quick Start*. [online] Available at: <http://mongodb.github.io/mongo-java-driver/3.4/driver/getting-started/quick-start/> [Accessed 4 Jun. 2017].

Técnicas de Trading. (2017). *Commodity Channel Index-CCI | Análisis Técnico | Técnicas de Trading*. [online] Available at: <http://www.tecnicasdetrading.com/2013/05/commodity-channel-index-cci.html> [Accessed 4 Jun. 2017].

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=Facultad de Informatica - UPM, C=ES
Fecha/Hora	Thu Jun 08 19:59:57 CEST 2017
Emisor del Certificado	EMAILADDRESS=camanager@fi.upm.es, CN=CA Facultad de Informatica, O=Facultad de Informatica - UPM, C=ES
Numero de Serie	630
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)