

CONVOCATORIA PROTOTIPOS ORIENTADOS AL MERCADO EN TCUE CURSO 2018-2019

MEMORIA FINAL DEL PROYECTO (Extensión máxima 30 páginas) (debe contener, al menos, los siguientes apartados)

“Esta actuación se encuadra en el Plan TCUE 2018-2020, y ha sido seleccionada en el marco de un programa operativo cofinanciado por el Fondo Europeo de Desarrollo Regional (FEDER) y la Junta de Castilla y León”

1. DATOS GENERALES

Beneficiario/a :	Nombre: Álvaro	Apellidos: Bartolomé del Canto
Tutor/a Académico	Nombre: Pablo	Apellidos: Chamoso Santos
	Nombre: Fernando	Apellidos: De La Prieta Pintado
	Nombre: Elena	Apellidos: Hernández Nieves
Otros colaboradores:		
Título del proyecto: Sistema de Recomendación de Inversión Bursátil en base a Predicciones de Mercado y Análisis de Opinión de Twitter		
Fecha de inicio: 01/01/2019		Fecha de finalización: 01/09/2019

2. INTRODUCCIÓN

El proyecto consiste en la investigación y desarrollo de un sistema de recomendación que combine el cálculo de factores técnicos propios del análisis de mercado tradicional con la aplicación de algoritmos de Inteligencia Artificial (IA) de cara a determinar el comportamiento futuro del mercado para así poder identificar señales de compra/venta. Adicionalmente, para apoyar dicha predicción basada meramente en los valores de dicha acción de apertura y cierre, se propone la creación de un modelo de lenguaje en español, para el pre-procesamiento de la opinión de cara a la creación de un modelo de análisis de sentimientos, que combinado con un algoritmo para identificar tendencias en el mercado, sea capaz de determinar la correlación entre el comportamiento del mercado y la opinión volcada por los usuarios en redes sociales, en este caso Twitter.

Esto implica que se requerirá la creación de una plataforma de cara a acercar al usuario los resultados del estudio e investigación realizado. Esto así supone una innovación tanto tecnológica como conceptual por la combinación de técnicas tradicionales para el estudio del comportamiento del mercado con algoritmos de *Machine Learning* y la creación de modelos de *Natural Language Processing*. Actualmente, hay algunos estudios que hablan de la posibilidad de combinar distintos parámetros ajenos al mercado, pero que puedan tener cierto impacto sobre el comportamiento de este, por lo que se propone el estudio e implementación de un estudio centrado en el impacto de la opinión extraída de Twitter para con el comportamiento del mercado. Así este proyecto acercará a los usuarios el estudio realizado, permitiéndoles analizar el comportamiento futuro del mercado por medio de la combinación de distintas técnicas de predicción y recomendación.

A nivel técnico cabe destacar: la creación de **investpy**, un paquete de Python para la recuperación de datos en tiempo real de Investing.com de acciones, fondos y ETFs de todo el mundo; la creación de **trendet** que es un paquete de Python para la identificación de tendencias alcistas o bajistas del mercado en series temporales a partir de los valores de cierre y aplicable a cualquier conjunto de datos de una serie temporal; la creación de **twipper** que es un *wrapper* de Twitter para Python, que permite recuperar tweets tanto en tiempo real (*streaming*) como a partir de un histórico (*batch*) y que integra funciones tanto de la API Free de Twitter como de la API Premium; la creación de un **modelo de lenguaje en español** dado que las principales bibliotecas de NLP son para inglés; la investigación y creación de un modelo de **análisis de sentimientos en español** que sea capaz de clasificar un tweet en un sentimiento: negativo, neutral o positivo; y el estudio de la combinación de las distintas características mencionadas previamente junto con la integración de las mismas y la creación de una plataforma que englobe todas ellas.

De este modo, se presentará la plataforma completa que integra todos y cada uno de los módulos mencionados previamente, pero que en caso de ser necesarios se podrán utilizar de manera aislada dado que todos los paquetes de Python creados son de código abierto, con lo que, a pesar de ser un sistema global completo, las funcionalidades individuales se podrán utilizar con otro propósito en caso de ser necesitado.

La plataforma así ofrece al usuario un análisis del mercado totalmente objetivo, es decir, muestra al usuario información objetiva, dado que son valores de mercado, con total transparencia; y la información procesada y los algoritmos utilizados se le muestran para que así tenga en todo momento una visión de lo que internamente está sucediendo en la plataforma.

A la hora de llevar a cabo el proyecto en lo que a gestión se refiere se ha seguido un desarrollo y gestión del software basado en metodologías ágiles, en este caso, *SCRUM* y *Kanban*; de modo que el proyecto resultante se ajustase a las necesidades definidas inicialmente, pudiendo estar sujeto a cambios, pero dando lugar a un proyecto consistente, robusto, escalable y modular.

3. OBJETIVOS

Tal y como se ha mencionado previamente, dado que el principal foco de investigación se basa en determinar la viabilidad de la combinación de técnicas de predicción del comportamiento de mercado en base a técnicas tradicionales como el análisis técnico financiero con técnicas actuales basadas en Inteligencia Artificial; el objetivo principal es llevar a cabo la **combinación del análisis técnico con los algoritmos de Machine Learning con el fin de identificar señales de compra/venta** para poder realizar una recomendación.

Adicionalmente, a parte del análisis de los valores de cierre, apertura, máximo y mínimo del mercado, se propone la adición de otros valores que puedan tener influencia sobre el mercado, por tanto, surge un caso de estudio que gira en torno a los datos extraídos de Twitter mediante técnicas de Procesamiento del Lenguaje Natural. A raíz de esto, surge un nuevo objetivo que es el **determinar el impacto de la opinión volcada en Twitter de una acción por parte de los usuarios**.

De todos modos, dado que el análisis y la predicción previamente realizados son de acciones de empresas del mercado continuo español, el modelo de NLP a realizar ha de ser en español, y dado a la carencia de fuentes tanto de cara al análisis como el pre-procesamiento de texto plano en español, surge un nuevo objetivo que es la **creación de un modelo de procesamiento de lenguaje natural en español**. También, a la hora de analizar el sentimiento de cada uno de los tweets, se propone la creación de una plataforma para la clasificación supervisada debido a la ausencia de datos clasificados en sentimientos: negativo, neutral y positivo, en español; con lo que se propone la creación de dicha plataforma, que, tras clasificar cierto número de tweets, dará lugar a la **creación de un modelo de análisis de sentimientos en español**.

Por tanto, los objetivos definidos en el proyecto van desde el análisis de los datos históricos de una acción, como la combinación del análisis de los mismos para con el desarrollo e investigación de factores ajenos al mercado que puedan tener impacto sobre el mismo.

4. INTERÉS DEL PROYECTO: MOTIVACIÓN Y ASPECTOS INNOVADORES DEL PROYECTO

1. Motivación

La motivación principal del proyecto es intentar determinar el impacto de la opinión volcada sobre las redes sociales, en concreto sobre Twitter, en el comportamiento de las acciones del mercado continuo español. De este modo, se busca combinar técnicas tradicionales con técnicas basadas en Inteligencia Artificial, como lo son el *Machine Learning* y el *Natural Language Processing*. Por tanto, la motivación del proyecto no solo es predecir el comportamiento futuro del mercado mediante la identificación de señales de compra/venta, sino el de determinar el impacto o no de la opinión volcada en Twitter por parte de los usuarios a una empresa, viendo cómo y si afecta positiva o negativamente al valor de mercado de la misma.

2. Aspectos Innovadores

Por tanto, en la misma línea que el apartado de Motivación, los principales aspectos innovadores se centran en la combinación de técnicas de Inteligencia Artificial con técnicas de Análisis Técnico financiero usadas en el análisis tradicional, con el fin de predecir el comportamiento futuro del mercado. Adicionalmente, para apoyar dicha combinación de técnicas, se añade la adición de la minería de opinión de Twitter junto con el análisis de sentimientos de la misma.

Con lo que el principal aspecto innovador gira en torno a la combinación del cálculo de los factores técnicos junto con los algoritmos de regresión utilizados para predecir el comportamiento futuro del mercado y, en consecuencia, identificar señales de compra/venta de acciones en función de dicha predicción. Además, la creación de un modelo de lenguaje natural en español supone un avance dado que los modelos de lenguaje desarrollados y de uso libre actualmente tan solo son en inglés, con lo que la creación del mismo conlleva la realización de todas las pruebas pertinentes de cara a crear un modelo no solo de lenguaje natural, sino de análisis de sentimientos.

5. ESPECIFICACIONES TÉCNICAS Y FASES DE DESARROLLO

1. Especificaciones Técnicas

A nivel técnico, introducir que se ha utilizado Python como lenguaje de programación principal, ya que es un lenguaje muy utilizado para el análisis de datos e Inteligencia Artificial, en la versión 3.7.4. Adicionalmente, para la plataforma se ha utilizado Django que es un framework de desarrollo web para Python basado en un patrón MVT (*Model-View-Template*) que, por tanto, combina Python con HTML, CSS y JavaScript. Como IDE se ha utilizado PyCharm Professional con una licencia de estudiante de la USAL; y también se ha utilizado Jupyter Notebooks, que es un IDE web interactivo que permite desarrollar código en Python de manera dinámica.

Ahora, se procederá a la explicación desglosada de cada una de las funcionalidades implementadas y las técnicas utilizadas para la conclusión de las mismas:

- **investpy**: es un paquete de Python para la extracción de datos históricos de Investing.com de los principales productos financieros, como Acciones, Fondos o ETFs entre otros. De este modo, investpy se basa en el uso de técnicas de *Web Scraping* sobre los datos proporcionados por Investing.com, donde a través de Python dichos datos son recuperados, y un proceso de *Web Scraping* es realizado. Una vez los datos son recuperados estos se formatean y se retornan bien en un *pandas.DataFrame* o en un objeto *JSON*. Destacar que las funcionalidades desarrolladas por investpy se han incluido en un paquete de Python homónimo, de modo que sea extensible y más usable, a la par que flexible en lo que a implementación de nuevas características se refiere. Para el desarrollo de las funciones de extracción de datos se han utilizado las bibliotecas de Python: *requests* y *lxml*, para la recuperación del HTML y el envío de las peticiones *GET* y *POST*, y para la extracción de datos de la respuesta, respectivamente. La documentación sobre su uso puede encontrarse en <https://investpy.readthedocs.io/> o bien un informe detallado sobre su uso en <https://github.com/alvarob96/tcue/raw/master/docs/investpy.pdf>.
- **Análisis de Algoritmos de ML**: tras extraer los datos históricos de las acciones del mercado continuo español se ha procedido a la investigación e implementación de los distintos algoritmos de regresión de Machine Learning. La investigación implica comprender el funcionamiento de los algoritmos con el fin de determinar qué algoritmo o algoritmos se ajustan mejor a las necesidades de la problemática planteada. De este modo, tras determinar los algoritmos a utilizar, se procederá a la realización de la validación cruzada de los hiper-parámetros de estos, ya que se requiere ajustar los parámetros para cada uno de los algoritmos y para cada una de las acciones, ya que distintas acciones no se comportan igual a lo largo del tiempo, con lo que los algoritmos que pueden funcionar muy bien para unas acciones, pueden ser los peores para otras. La documentación técnica sobre lo mencionado previamente se encuentra en el informe <https://github.com/alvarob96/tcue/raw/master/docs/investpy-ml.pdf>.

- **Plataforma de Visualización:** se ha creado una plataforma para la visualización de los datos extraídos de **investpy** en base al análisis realizado con el fin de acercar al usuario dichos datos y hacer transparente al mismo el estudio y análisis de los mismos. Adicionalmente, la plataforma de visualización realizada en *Django* siguiendo el patrón *MVT*, ofrece dos alternativas, bien la visualización de los datos de la acción española introducida o bien la visualización y recomendación. En la recomendación, a parte de la visualización, se incluyen los algoritmos de ML previamente estudiados con los mejores hiper-parámetros de cara a la creación del modelo de regresión para predecir el comportamiento futuro del mercado, donde los algoritmos utilizados se ordenarán en función de la precisión de los mismos. Para montar la plataforma, consultar la documentación en https://github.com/alvarob96/tcue/tree/master/tcue_platform.
- **Plataforma de Clasificación Supervisada de Tweets en Sentimientos:** se ha creado una plataforma adicional para apoyar la clasificación supervisada de tweets en sentimientos con el fin de hacerlo más automático, de modo que la clasificación de los mismos sea más fácil y rápida. Los datos recuperados a través de **twipper** se insertan en una base de datos *noSQL*, *MongoDB* en este caso. La plataforma así ofrece tweets extraídos de *MongoDB* en *streaming* y botones para la clasificación de los mismos en sentimiento: negativo, neutral y positivo. Para montar la plataforma, consultar la documentación en https://github.com/alvarob96/tcue/tree/master/tcue_sentiment.
- **trendet:** es un paquete de Python creado para la identificación de tendencias tanto alcistas como bajistas en base a los datos históricos de una acción recuperados por **investpy**, con un algoritmo de creación propia. Este paquete se centra en la identificación de las tendencias de datos históricos de mercado, pero del mismo modo puede ser utilizado para cualquier matriz de datos numéricos en base a una serie temporal, aunque el comportamiento esperado no es tan bueno. La documentación sobre su uso puede encontrarse en <https://trendet.readthedocs.io> o bien un informe detallado sobre su uso en <https://github.com/alvarob96/tcue/raw/master/docs/trendet.pdf>.
- **twipper:** es un paquete de Python creado para cubrir y encapsular las llamadas (*endpoints*) de la API de Twitter, tanto de la versión *Premium* como de la versión *Free*. De este modo, busca cubrir una necesidad debido a las carencias de los paquetes actuales, en los que no se combinan las llamadas en un único paquete, junto con este motivo, el hecho de evitar dependencias en paquetes de terceros, se crea **twipper**. Se basa en la creación de una clase para la autenticación con Twitter, a partir de la cual se realizarán las llamadas a los endpoints desde las distintas funciones, mediante la biblioteca de Python, *requests*. Por tanto, básicamente consiste en encapsular las llamadas a los *endpoints* de Twitter a través de Python, permitiendo la modificación de los campos de la petición mediante parámetros de Python. La documentación sobre su uso puede encontrarse en <https://twipper.readthedocs.io> o bien un informe detallado sobre su uso en <https://github.com/alvarob96/tcue/raw/master/docs/twipper.pdf>.
- **Modelo de Lenguaje y Análisis de Sentimientos en español:** se ha desarrollado también para la parte de *NLP* un modelo de lenguaje en español para la tokenización de

los tweets, es decir, la conversión de cada uno de los tweets en listas de palabras o *tokens* válidas, lo cual implica eliminar tanto las *stopwords*, siendo estas palabras no útiles a nivel de análisis del texto con *NLP*, como los caracteres que no sean palabras. Además, se ha creado un modelo de clasificación para el análisis de sentimientos en español en base a los tweets previamente clasificados de forma supervisada. El modelo se ha creado utilizando los datos extraídos de la plataforma y la biblioteca de *sklearn* para la implementación y uso de los algoritmos, tanto de *Feature Extraction* como de *Classifiers*. De este modo, el mejor modelo en base a una validación cruzada y a pruebas supervisadas se ha determinado que es la combinación de *TfidfVectorizer* junto con *MultinomialNB*, llegando a tener un 70% de precisión el modelo mejorado, aún con un alto porcentaje de mejora. La documentación técnica sobre lo mencionado previamente se encuentra en el informe <https://github.com/alvarob96/tcue/raw/master/docs/twipper-sentiment.pdf>.

- **Análisis del impacto del sentimiento de Twitter para con el comportamiento/tendencias del mercado:** siendo esta una fase de investigación, como resultado de la combinación de los paquetes creados previamente, es decir, **trendet** y **twipper**, se propone la combinación de los mismos con el fin de determinar el impacto del análisis del sentimiento de los tweets escritos sobre una acción, en la franja temporal en la que se ha identificado una tendencia, y comprobar si se cumple que para una tendencia alcista el análisis de sentimiento en positivo y, por el contrario, para una tendencia bajista el sentimiento es negativo, lo cual ayudará a determinar si la opinión volcada en Twitter tiene o no algún impacto sobre el comportamiento del mercado. La documentación técnica sobre lo mencionado previamente se encuentra en el informe <https://github.com/alvarob96/tcue/raw/master/docs/trendet-twipper.pdf>.

Por último, mencionar que los paquetes de Python se han alojado en PyPI (*Python Package Index*), que es un hosting de paquetes de Python instalables a través del comando en terminal: *pip*. Además, la documentación de estos se ha generado con *sphinx* que es una biblioteca de Python para generar la documentación técnica en Python de una forma fácil y rápida en formato HTML; la cual ha sido *hosteada* en *Read the docs* que es una plataforma para el hosting de documentación en línea. Todas las herramientas utilizadas mencionadas previamente están perfectamente integradas con *GitHub* en cada uno de los repositorios creados,

2. Fases de Desarrollo

Para el desarrollo del proyecto se ha seguido una planificación basada en metodologías ágiles, en este caso se han utilizado *SCRUM* y *Kanban*. *SCRUM* se ha utilizado gracias a su flexibilidad, facilidad de uso y escalabilidad, que, dado que el equipo desarrollador está formado únicamente por una persona, el resto de las metodologías orientadas a equipos se han descartado. El hecho de que se haya utilizado *SCRUM* por la flexibilidad implica que, aunque inicialmente las tareas definidas en el proyecto hayan sido unas, estas están sujetas a cambios y a modificaciones en

caso de ser requeridas para el desarrollo y conclusión del proyecto. Adicionalmente, se utiliza *Kanban* con el fin de gestionar las tareas, agrupándolas en columnas: *To do*, *Doing* y *Done*, de modo que se agrupan las tareas en función de su estado de finalización; aunque adicionalmente también se pueden etiquetar en categorías. Para realizar la planificación y las estimaciones pertinentes para la gestión y organización del proyecto con *SCRUM* se ha utilizado *Microsoft Excel*; mientras que para la gestión de las tareas del proyecto con una estructura y organización de *Kanban* se han utilizado *GitHub Projects* y *Trello*.

La planificación de tareas propuesta contempla como fecha de inicio del proyecto el 1 de enero de 2019 y como finalización el 1 de septiembre de 2019, tal y como se desglosan en la siguiente figura; las cuales serán desglosadas y explicadas brevemente más adelante.

#	Tarea	M1	M2	M3	M4	M5	M6	M7	M8
1	Estudio del problema y desarrollo de la metodología								
2	Análisis de los requisitos y definición de objetivos								
3	Desarrollo del paquete de extracción de datos (<i>investpy</i>)								
4	Estudio de algoritmos de regresión de Machine Learning								
5	Implementación de la plataforma e integración con <i>investpy</i>								
6	Desarrollo del paquete de detección de tendencias (<i>trendet</i>)								
7	Desarrollo del wrapper de Twitter (<i>twipper</i>)								
8	Estudio y análisis del impacto de la opinión de Twitter sobre las acciones								
9	Corrección de errores y preparación de la versión beta de la plataforma								
10	Documentación								

Tareas:

- 1- Estudio del problema y desarrollo de la metodología: es la fase inicial en la que se estudia el problema en base a la propuesta realizada, y se comienza el desarrollo de la metodología a seguir.
- 2- Análisis de los requisitos y definición de objetivos: esta fase consiste en el análisis de los requisitos propuestos inicialmente y, en consecuencia, en la identificación de una serie de objetivos en base a los mismos.
- 3- Desarrollo del paquete de extracción de datos (***investpy***): en esta fase, dado que no existía ninguna fuente de datos disponible o, en este caso, ningún paquete de Python que supliese las necesidades del proyecto, se creará un paquete de Python que extraiga los datos de *Investing.com* de las acciones españolas y que los formatee para su uso integrado con Python.
- 4- Estudio de algoritmos de regresión de Machine Learning: en esta fase, según se comience a tener datos, bien extraídos por *investpy* o bien datos de ejemplo, se procederá a la prueba e implementación de algoritmos de ML de regresión para la predicción de los futuros valores de cierre a partir de los valores de apertura de mercado. En esta fase, se determinarán los algoritmos a utilizar y se realizará una validación cruzada de los hiper-

parámetros a utilizar por cada uno de ellos, con el fin de determinar la mejor combinación (la que dé una mayor precisión).

- 5- Implementación de la plataforma e integración: en esta fase se llevará a cabo el desarrollo de la plataforma web y su integración tanto con investpy como con los algoritmos de ML previamente probados con el fin de visualizar tanto el análisis de los datos, como la predicción.
- 6- Desarrollo del paquete de detección de tendencias (**trendet**): en esta fase se creará un paquete de Python para, de nuevo, suplir una necesidad en el proyecto a la hora de determinar las tendencias de una serie temporal en base a los datos históricos de una acción de forma automática.
- 7- Desarrollo del wrapper de Twitter (**twipper**): en esta fase se creará un paquete de Python para, de nuevo, suplir una necesidad en el proyecto en lo que a extracción de datos de Twitter a través de su API se refiere, dado que se requiere la combinación de las llamadas a la API Free y API Premium en un mismo paquete. Además, dado que las APIs están sujetas a cambios, con el fin de hacer del proyecto uno flexible y escalable, se ha determinado que la creación de dicho paquete beneficiará al proyecto en caso de cambios en la API de Twitter.
- 8- Estudio y análisis del impacto de la opinión de Twitter sobre las acciones: en esta fase, se combinarán los paquetes previamente creados, *trendet* y *twipper*, de modo que en base a las tendencias alcistas y bajistas identificadas se analizará el sentimiento de forma que se buscará la correlación entre el sentimiento para con la tendencia.
- 9- Corrección de errores y preparación de la versión beta de la plataforma: en esta fase se corregirán los errores identificados en la plataforma en base a la experiencia de usuario y a las pruebas de esta con usuarios reales; mientras que los errores en los paquetes de Python bien se corregirán por la identificación de los mismos por el desarrollador, o bien en base a los identificados por los usuarios del mismo a través de los *issues* de *GitHub*.
- 10- Documentación: esta fase tendrá lugar a lo largo de todo el proyecto a partir de la fase inicial, debido a la alta importancia de documentar tanto los avances realizados como el código desarrollado.

Por tanto, siguiendo dicha organización de las tareas, se ha concluido el proyecto en septiembre de 2019, sobre los plazos fijados e incluso aprovechando *SCRUM* para la adición de características no contempladas inicialmente, pero que son beneficiosas para la plataforma, con el fin de poder presentar una versión inicial totalmente funcional.

6. ANÁLISIS BÁSICO sobre las posibilidades de explotación del prototipo obtenido (mercado, creación de una empresa, protección del resultado...)

Las posibilidades de explotación de este proyecto son infinitas, también en parte no solo a la idea central del proyecto, la cual dado que aún se encuentra en una fase de investigación podría ampliarse enormemente; sino también gracias al hecho de haber seguido un desarrollo del proyecto modular, de forma que surgen líneas de desarrollo y explotación auxiliares. De este modo, para cada uno de los paquetes de Python podría dar lugar a la explotación de cada uno de los mismos bien en conjunto o bien individualmente con el fin de extender su funcionalidad, así como los modelos de NLP desarrollados para el lenguaje español.

Por tanto, el desarrollo modular facilita la ampliación de las posibilidades de explotación del proyecto, pudiendo así desarrollarse cada módulo individualmente, o darle uso en otro proyecto. El proyecto por tanto podrá tener una explotación total con el fin de crear una empresa *spin-off* en base al proyecto, vender el proyecto a una empresa o empresas interesadas, o bien una explotación particular en caso de querer explotar cualquiera de los módulos creados, pudiendo así ser integrados para con otros proyectos y/o empresas en caso de que estas necesiten o vean en ellas una vía de desarrollo futuro de cara a obtener rentabilidad.

También cabría tener en cuenta la posibilidad no solo de mejora de la plataforma sino la adaptación de esta a otros mercados, dando así lugar a la cobertura de un mayor rango de productos bancarios como los fondos o las ETFs, y la ampliación a todos los países posibles, no solo a los productos financieros de España. Y, la mejora esperada, referente a la mejora del sistema de predicción con casos de éxito y la ampliación del rango de predicción a medio-largo plazo, no solo a corto plazo.

Debido a lo mencionado anteriormente, por tanto, se podría también dar lugar a la creación de una empresa que gestione y siga con el desarrollo de la plataforma presentada, cuyo objetivo sea hacerla escalable y mejorarla de cara a proporcionar soporte a otras empresas. El hecho de dar soporte a otras empresas puede ser beneficioso de cara a la necesidad de estas a la resolución de necesidades particulares a partir del amplio espectro de soluciones ofertadas en el proyecto, que como se menciona previamente, cabe la posibilidad de que el interés de una empresa no sea para con todo el proyecto sino para con alguno de los módulos que lo integran.

7. MANUAL DE USUARIO

En esta sección se pretende abarcar todos los aspectos referentes a la plataforma web desarrollada para con la cual interactuará el usuario, de este modo en esta sección se presentará tanto la plataforma como el uso e interacción del usuario con la misma, dado que los detalles técnicos ya han sido desglosados previamente.

La plataforma se ha desarrollado en base a las directrices de diseño web marcadas por *Jakob Nielsen* en el libro “*Designing Web Usability*”, centradas en la simpleza del diseño y la usabilidad, por lo que se ha desarrollado una plataforma con *Django* de cara a visualizar los resultados de estudio y presentarlos a alto nivel (nivel de usuario).

1. Presentación de la Plataforma

La plataforma **Django** propuesta consta de una vista principal en la que el usuario podrá introducir el nombre de una acción de las pertenecientes al mercado continuo español listadas en Investing.com, dado que esta es la fuente de datos principal de la que extrae los datos **investpy**. En dicha vista principal se le ofrece una breve descripción de la plataforma al usuario en la que se reflejan los aspectos técnicos de la plataforma siendo estos en análisis de datos, la aplicación de funciones matemáticas propias del análisis técnico financiero y el uso de algoritmos de ML para determinar el comportamiento futuro del mercado y así ofrecer una recomendación al usuario.



Plataforma de Visualización y Recomendación de Inversión Bursátil en Acciones del Mercado Continuo Español

Visualización

Visualización y Recomendación

Álvaro Bartolomé del Canto, 2018-2019 @ alvarob96 en GitHub

De este modo, cuando el usuario introduzca el nombre de una de las acciones del mercado continuo español de las indexadas en Investing.com, la plataforma procederá a la búsqueda de los datos a través de **investpy** y a preparar su visualización y análisis en caso de seleccionar la opción *Visualización*, que en caso de seleccionar *Visualización & Recomendación* también realizará la predicción a partir de los factores técnicos y los algoritmos de ML, que tendrán como resultado una recomendación al usuario en función de las señales de compra/venta identificadas.

Cabe destacar que la plataforma tiene control de errores, de modo que en caso de que el nombre de la acción introducida no sea válido, se mostrará un panel de error en el que se describirá el error junto con el código de este y una posible solución al mismo.

Error 400 Petición Errónea

¡Algo ha ido mal!

Por favor, asegúrese de que el nombre de la acción introducida es correcto. Puede comprobar los nombres de las acciones disponibles en: [Investing.com - Acciones de España](#).

Volver

Álvaro Bartolomé del Canto, 2018-2019 @ alvarob96 en GitHub

2. Visualización de Datos

En caso de haber seleccionado la opción *Visualización*, la plataforma procederá a la recuperación de los datos en tiempo real a través de **investpy**, que incluye tanto el perfil de empresa de la acción introducida, así como los datos históricos de los últimos 5 años de apertura, cierre, más alto, más bajo y volumen de cada día.

El perfil de empresa se muestra de modo que acerque al usuario la actividad y descripción de la empresa que tiene la acción introducida, pudiendo tener así una visión general de dicha empresa, dado que el perfil de empresa es un extracto de texto redactado por la propia empresa en la que se recoge su actividad general.

Banco Bilbao Vizcaya Argentaria S.A. - BBVA

BBVA es un grupo financiero global fundado en 1857 con una visión centrada en el cliente. BBVA disfruta de una sólida posición de liderazgo en el mercado español, es la mayor institución financiera de México y cuenta con franquicias líderes en América del Sur y en la región "Sunbelt" de EE. UU. Además, cuenta con una presencia relevante en Turquía y China (a través de inversiones estratégicas en Garanti Bank y CITIC), y opera en una amplia red de oficinas en todo el mundo. Su negocio diversificado está enfocado a mercados de alto crecimiento y concibe la tecnología como una ventaja competitiva clave. BBVA es uno de los primeros bancos de la eurozona por rentabilidad de los recursos propios (ROE) y eficiencia. La responsabilidad corporativa es inherente a su modelo de negocio, impulsa la inclusión y la educación financieras y apoya la investigación científica y la cultura. BBVA opera con la máxima integridad, visión a largo plazo y mejores prácticas, y está presente en los principales índices de sostenibilidad. BBVA emplea a más de 110.000 personas en más de 30 países, tiene más de 50 millones de clientes y más de un millón de accionistas.

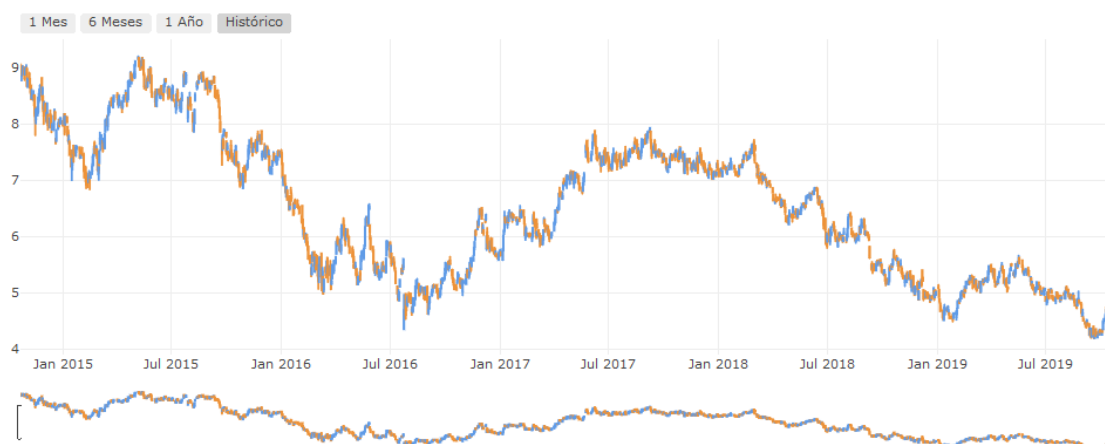
[Información Adicional >](#)

Los datos históricos de los últimos 5 años de dicha acción se representarán de varias formas con el fin de ayudar al usuario a comprender los datos y sus valores a lo largo del tiempo, como parte del análisis básico presentado. Las representaciones básicas son una gráfica de la serie temporal recuperada con la cual se puede interactuar y que muestra todos los datos extraídos de Investing.com a través de investpy; y, también, un gráfico de velas, que es una representación de datos muy utilizada en las series temporales de los valores de mercado diarios, en los que se refleja el comportamiento general de dicha acción en el mercado en ese día, ya que cada vela abarca tanto los valores de cierre y apertura, como los valores máximo y mínimo.

Serie Temporal de los Precios de Cierre de BBVA



Gráfico de Velas de BBVA



Además, los datos se presentan también en una tabla con las columnas: Fecha, Apertura, Máximo, Mínimo, Cierre/Último, Volumen y Divisa; siendo estos respectivamente, la fecha de dichos valores de mercado (siendo los días hábiles en los que el mercado está abierto en España de lunes a viernes excepto festivos), los valores de apertura, máximo, mínimo y cierre de ese día; el volumen, que es la cantidad de títulos negociados; y, por último, la divisa en la que los valores del mercado están especificados.

Datos Históricos de BBVA

Fecha	Apertura	Máximo	Mínimo	Cierre/Último	Volumen	Divisa
19/09/2019	4.722	4.785	4.694	4.785	0	EUR
18/09/2019	4.72	4.743	4.646	4.692	16880000	EUR
17/09/2019	4.771	4.782	4.669	4.697	18510000	EUR
16/09/2019	4.825	4.846	4.758	4.795	16079999	EUR
13/09/2019	4.77	4.872	4.758	4.865	28150000	EUR
12/09/2019	4.678	4.775	4.554	4.729	30150000	EUR
11/09/2019	4.74	4.772	4.653	4.676	21110000	EUR
10/09/2019	4.6	4.702	4.542	4.696	26240000	EUR
09/09/2019	4.508	4.602	4.501	4.602	16020000	EUR
06/09/2019	4.483	4.56	4.46	4.49	15900000	EUR

De este modo concluye la parte de visualización y análisis inicial de los datos de dicha acción, dado que esta vista tan solo ofrece esas características, que serán completadas con la predicción y recomendación posterior mediante la selección de la opción *Visualización & Recomendación*.

3. Sistema de Recomendación

En caso de seleccionar la opción *Visualización & Recomendación*, se realizará tanto la extracción y visualización de los datos como la predicción mediante el cálculo de los factores técnicos y la aplicación de los algoritmos de ML de cara a la identificación de señales de compra/venta para poder realizar en base a estas una recomendación al usuario.

Sistema de Recomendación de BBVA

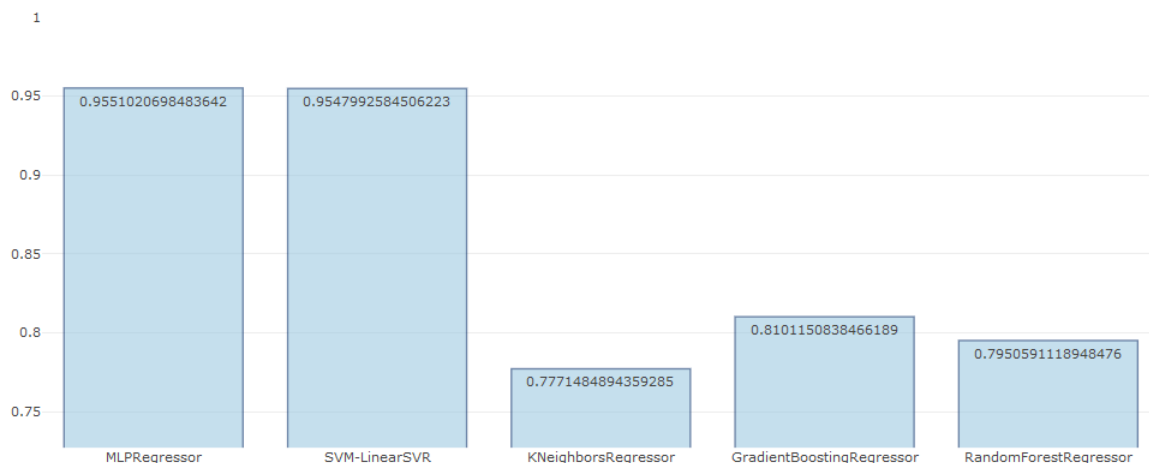
A continuación se presenta el sistema de recomendación basado en la identificación de señales de compra/venta como resultado de la combinación del análisis técnico financiero (técnica tradicional para determinar el comportamiento futuro del mercado) con algoritmos basados en Inteligencia Artificial. Esto implica por tanto la combinación de sistemas tradicionales de predicción con sistemas modernos, con el fin de determinar la correlación o no entre estos a la hora de identificar señales de compra/venta en base a los valores históricos de una acción (en este caso, una del mercado continuo español).

Los algoritmos de ML utilizados han sido los de regresión, en los que a partir de una variable de entrada (predictor) se predice una variable objetivo (target), analizando así la relación entre ambos. En este caso, el sistema implementado, a partir de valores históricos de apertura del mercado, busca predecir los valores de cierre.

Los factores técnicos son estrategias propias del análisis técnico que se utilizan para predecir futuros movimientos u oscilaciones del precio de una acción, analizando los datos históricos y actuales de dicha acción. Adicionalmente, el análisis técnico también emplea otras medidas como las medias móviles (SMA & EMA) con el fin de determinar la dirección que tomará el mercado en el futuro en base a la tendencia actual. Otro de los posibles usos de las medias móviles se centra en la posibilidad de observar datos suavizados en vez de centrarse en los valores diarios.

Adicionalmente, tras la sección de *Visualización* en esta vista se presentará el sistema de recomendación creado para cada una de las distintas acciones de las empresas que forman el mercado continuo español dado que los hiper-parámetros utilizados en los algoritmos se han elegido en función de cuales daban un mejor resultado en la predicción de cada una de las acciones.

Precisión de los Algoritmos de Regresión



El mejor algoritmo en base a su precisión comparando la predicción con los valores reales es: "MLPRegressor"

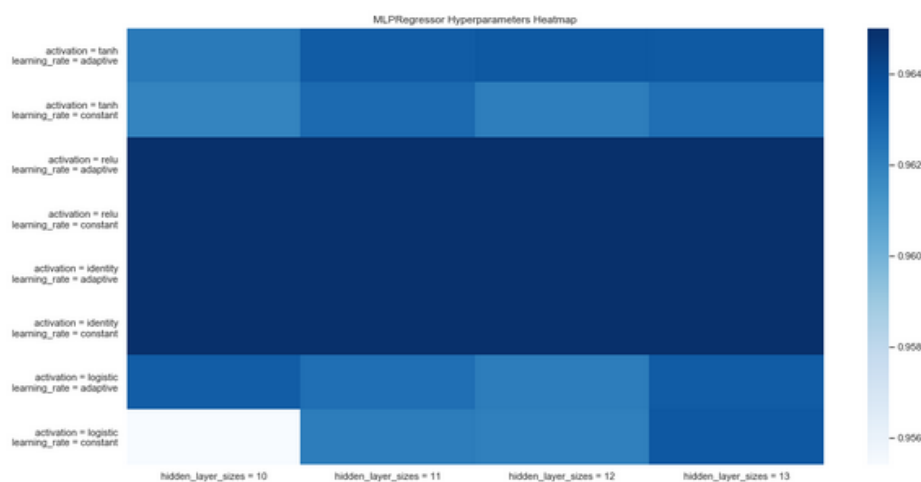
Mapas de Calor de Selección de Híper-Parámetros

Por tanto, se presentará el resultado en cuanto a precisión obtenido por cada uno de los algoritmos aplicados internamente por la plataforma, de modo que el usuario pueda ver cual de los algoritmos aplicados ha funcionado mejor.

Mapas de Calor de Selección de Híper-Parámetros

MLPRegressor

A multilayer perceptron (MLP) is a class of feedforward artificial neural network. A MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable.



LinearSVR

In machine learning, support-vector machines (SVMs, also support-vector networks) are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). In SVR we try to fit the error within a certain threshold. Similar to SVR with parameter kernel='linear' in this case.

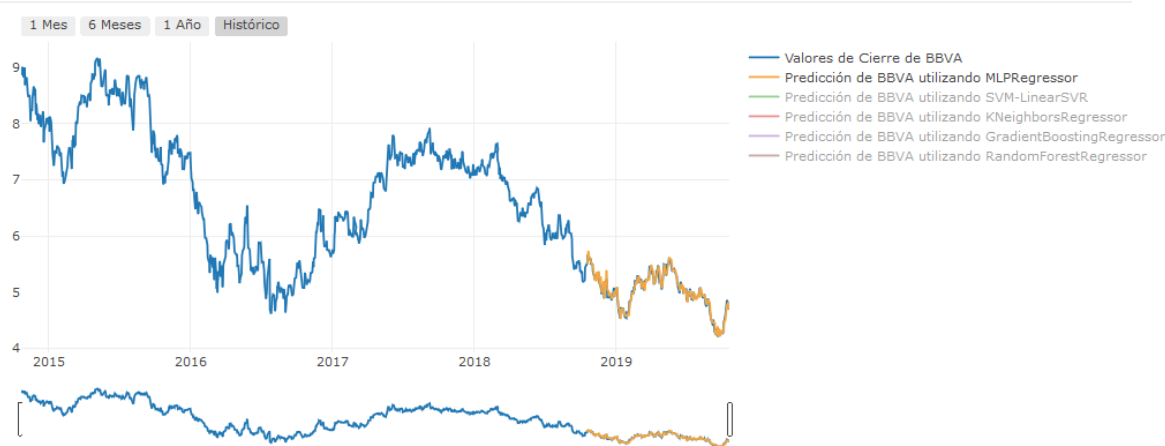


También se ofrece al usuario la posibilidad de observar una breve descripción de los distintos algoritmos utilizados junto con un mapa de calor que representa el resultado de la validación cruzada realizada para cada acción con el fin de determinar los híper-parámetros a utilizar.

De este modo, se presenta la visualización del resultado de la predicción superpuesto a los datos históricos de entrada, para que el usuario pueda visualizar la precisión del algoritmo en el conjunto

de *test* a partir del conjunto de *train*, junto con la recomendación extraída en función de los datos resultantes de aplicar el algoritmo para los valores de cierre de ese día hábil a partir de su valor de apertura. Además, la biblioteca de visualización utilizada, *Plotly.js*, permite seleccionar los valores de la serie temporal a mostrar, así como interactuar con ellos.

Predicción de BBVA utilizando Algoritmos de Regresión



Dado que el valor de la predicción para hoy es **4.72334** y el valor de apertura fue **4.722**.

La tendencia de hoy es **ALCISTA**, por lo que la recomendación en base a los algoritmos de regresión es **VENDER** dado que el valor actual de la acción es **4.785**.

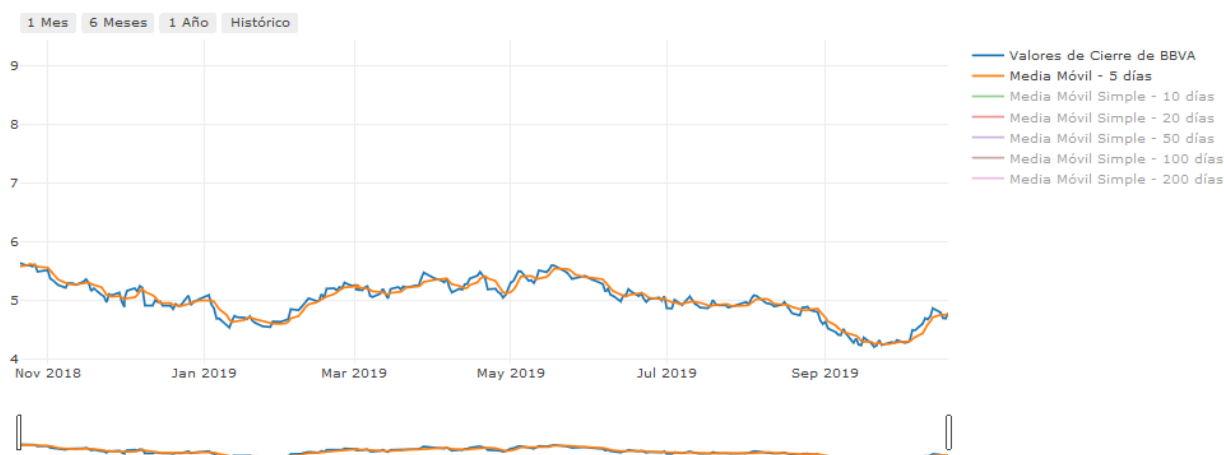
A parte de los resultados de la predicción con los algoritmos de ML, la plataforma también muestra los resultados obtenidos del análisis técnico financiero realizado que calcula tanto las medias móviles (*SMA* y *EMA*) como los factores técnicos (*Momentum Indicators*), junto con la recomendación en función del valor de cierre calculado para ese día en función del valor de apertura de este.

Así las medias móviles representan lo siguiente:

- *Simple Moving Average (SMA)*: es un indicador técnico que indica si el valor de una acción va a cambiar de tendencia, lo cual da lugar a la identificación de señales de compra/venta en función del comportamiento del mercado, tomando así distintas ventanas temporales.

• Medias Móviles Simples (SMA)

Las Medias Móviles (MA) son una importante herramienta analítica con el fin de identificar las tendencias actuales del mercado en función de sus precios y, también, determinar el potencial de cambio de una tendencia ya establecida (bien sea alcista o bajista). Una Media Móvil Simple suaviza la volatilidad sobre el precio de una acción, haciendo así **más fácil observar la certeza de una tendencia**.



Ventanas Temporales	SMA	Señal de SMA
5 days	4.7668	COMPRAR
10 days	4.7027	COMPRAR
20 days	4.5055	COMPRAR
50 days	4.5526	COMPRAR
100 days	4.78066	COMPRAR
200 days	4.96086	VENDER

Si la Media Móvil Simple va hacia arriba, implica que el precio está aumentando, por lo que se identifica una señal de **VENTA**.

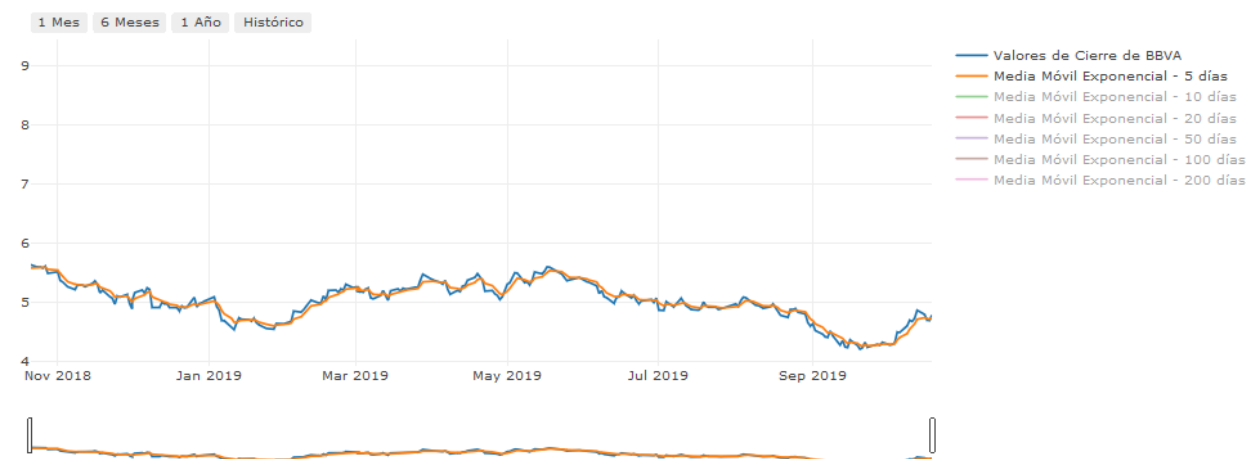
Si la Media Móvil Simple va hacia abajo, implica que el precio está disminuyendo, por lo que se identifica una señal de **COMPRA**.

En este caso, se han identificado 5 señales de COMPRA y 1 señales de VENTA, con lo que la recomendación es un movimiento de **COMPRA FUERTE**.

- **Exponential Moving Averages (EMA)**: similar a SMA, pero con la diferencia de que EMA data un mayor peso al comportamiento reciente del mercado, ponderando más los cruces y divergencias del valor de EMA respecto del valor real del mercado.

• Medias Móviles Exponenciales (EMA)

Las Medias Móviles Exponenciales (EMA) son un tipo de Media Móvil que asigna un mayor peso e importancia a los valores más recientes en lo que a cambios en el mercado se refiere. La principal diferencia, por tanto, entre las SMA y las EMA reside en que EMA da mayor peso a las observaciones recientes mientras que SMA asigna un peso igual a todas las observaciones.



Ventanas Temporales	EMA	Señal de EMA
5 days	4.7668	COMPRAR
10 days	4.67818	COMPRAR
20 days	4.58804	COMPRAR
50 days	4.59977	COMPRAR
100 days	4.74115	COMPRAR
200 days	4.97385	VENDER

Las Media Móviles Exponenciales (EMA) son mejores para los mercados bursátiles, dado que cuando el mercado está en una sólida y consistente tendencia alcista, la línea de la EMA también mostrará una tendencia alcista; y, viceversa para las tendencias bajistas.

En este caso, se han identificado 5 señales de COMPRA y 1 señales de VENTA, con lo que la recomendación es un movimiento de **COMPRA FUERTE**.

También se realiza el cálculo de los siguientes factores técnicos:

- *Relative Strength Index (RSI)*: es un indicador que mide el impacto de los cambios de precio recientes con el fin de identificar señales de compra/venta en base al histórico de una acción.
- *Stochastic Oscillator (STOCH)*: es un indicador que compara un precio de cierre de una acción dentro de un rango de los precios de cierre que ha tenido a lo largo del tiempo.
- *Ultimate Oscillator (ULTOSC)*: es un indicador que mide el *momentum* de precio de una acción a lo largo de una serie temporal. Las señales de compra/venta se generan calculando las divergencias entre la media ponderada de tres franjas temporales distintas con los valores de cierre reales.

- *Williams %R (WILLR)*: es un indicador cuyo valor oscila entre 0 y -100 y mide los niveles de sobre-compra o sobre-venta. Este indicador busca encontrar puntos de entrada o salida del mercado.

• Technical Factors

Factor Técnico	Resultado	Señal
RSI - 14 días	62.9368	COMPRAR
STOCH - 9,6 días	79.3436	COMPRA FUERTE
ULTOSC - 7,14,28 días	60.454	COMPRAR
WILLR - 14 días	-13.80952	SOBRECOMPRA

Finalmente, el sistema ha establecido una recomendación en base a la identificación de señales de compra/venta a partir de los datos históricos de los últimos 5 años de una acción, de modo que se pueda contrastar las señales resultantes del análisis técnico financiero respecto de las identificadas por los algoritmos de ML.

Fecha de presentación: 20/09/2019

Lugar de presentación: Salamanca

Firmado:



Nombre, Apellidos y DNI: Álvaro Bartolomé del Canto, 70959831R

La presente MEMORIA FINAL tiene que acompañarse de los siguientes archivos en formato electrónico:

- **Power Point de presentación del prototipo**
- **Video demostrativo**
- **Ficha resumen de los resultados obtenidos**
- **Código de aplicación (*)**
- **Ejecutable (*)**

(*): Archivos requeridos únicamente en proyectos relacionados con TIC's.

La memoria final, previamente revisada por el/la tutor/a académica, deberá ser remitida en formato electrónico (pdf) junto con los archivos anteriormente mencionados a la dirección: tcue@usal.es teniendo como fecha límite de entrega el 20 de septiembre de 2019. La entrega de la documentación final requerida condicionará el pago final de la ayuda concedida.

Los datos facilitados servirán para ofrecerle nuestra mejor atención en el desarrollo de la actividad con fines de gestión, estadísticos y de control. La Fundación General de la Universidad de Salamanca introducirá sus datos en un fichero del que es responsable esta entidad y se compromete, de acuerdo con la normativa vigente en materia de protección de datos, a su deber de guardarlos, y a adoptar las medidas necesarias para evitar su alteración, pérdida, tratamiento o acceso no autorizado. Se permite a los interesados ejercer en todo momento sus derechos de acceso, rectificación, cancelación y oposición dirigiéndose a la Fundación General de la Universidad de Salamanca, sin que medie contraprestación alguna.
