

Trabajo final integrador

Predicción de tendencias de mercado
sobre commodities utilizando algoritmos de
Inteligencia Artificial

Autor: Alejandro Mildiner

Tutor: Gustavo Arjones

Instituto Tecnológico de Buenos Aires

4 de mayo de 2020

Índice

1. Introducción	1
2. Definición del problema	2
3. Investigación	3
3.1. Extracción/Obtención de datos	4
3.2. Análisis estadístico de los datos	4
3.3. Predicción a corto plazo	11
3.3.1. Modelo Lineal	11
3.3.2. Modelo Logístico	11
3.3.3. Modelo Random Forest	13
3.4. Predicción a largo plazo	14
3.4.1. Prophet simple	15
3.4.2. Prophet lineal	15
3.4.3. Prophet logístico	16
3.4.4. Prophet con conocimiento del dominio	17
3.4.5. Métricas para medir performance	18
4. Conclusiones	23
5. Anexo	25

1. Introducción

La evolución de las nuevas tecnologías de comunicación e internet generó un impacto en todas las industrias. La industria financiera no fue la excepción. A continuación, se muestran algunas citas de los principales medios del país y referentes del mundo.

“Los bancos digitales no detienen su crecimiento. Incluso en años recesivos (y con una fuerte devaluación incluida) como este 2018. Según pudo confirmar Infotechnology.com con la Cámara Argentina de Fintech -presidido por Juan Pablo Bruzzo-, la industria registró una envidiable tasa de crecimiento anual de 110 %.”[1]

“... El empresario [Matías Friedberg] agregó que en el 2018 había en el país 133 empresas Fintech y que este año el número creció 70 % hasta las 226 compañías.”[2]

“El sector Fintech se encuentra creciendo a pasos agigantados. Atendiendo a datos del Observatorio Sectorial DBK de INFORMA, el negocio de las empresas Fintech creció más de un 61.3 % a lo largo de 2018, con unos 430.5 millones de euros en el total de las operaciones realizadas”. [3]

“En 2018, PayPal, una compañía Fintech especializada en pagos, incremento su base de cuantas un 17 % desde 2017, alcanzando 267 millones de cuentas activas, con un record de 13.8 millones nuevas cuentas en el cuarto cuatrimestre del año. Debido a un numero creciente de negocios están optando por sistemas de pagos digitales, la demanda de soluciones Fintech esta creciendo e impulsando el crecimiento del mercado”. [4]

De las citas anteriores podemos observar que la industria financiera está siendo impactada en dos frentes: a) La aparición de un gran número de Fintechs tomando lugares sólo reservados a los grandes operadores y b) la cantidad y diversidad de nuevos inversores, cada uno de ellos con perfiles y necesidades diferentes.

La democratización de la industria financiera generó una explosión en el mercado, llevando a un círculo virtuoso, en el que las nuevas tecnologías abren la posibilidad a la generación de nuevas Fintechs que a su vez se acercan a la población buscando captar nuevos inversores. Con cada nuevo inversor y con cada nueva Fintech la necesidad de brindar un mejor asesoramiento se vuelve vital para las empresas, lo que a su vez potencia el desarrollo de nuevas tecnologías. Esta misma democratización permitió a pequeños inversores tomar control de sus operaciones bursátiles, comprando y vendiendo de acuerdo con su propio criterio y sensibilidad al riesgo.

Esta nueva forma de operar afianzo la necesidad tanto del pequeño in-

versor como de las Fintechs y los grandes operadores de bolsa, de querer reducir su exposición al riesgo a la hora de tomar decisiones sobre cómo operar para maximizar sus ganancias. Esta aversión al riesgo y necesidad de asesoramiento en tiempo real puso de manifiesto que el servicio de un buen asesoramiento pasó a tener un valorpreciado por su potencial para aumentar las ganancias.

Los grandes operadores cuentan con costosos, complejos y privados modelos de predicción de precios y tendencias. Si bien, un pequeño inversor puede asesorarse con grandes agentes de bolsa como JPMorgan Chase, Morgan Stanley, Goldman Sachs o diferentes fondos de inversión, sus servicios no son económicos ni están libres de errores. Exponiendo al pequeño inversor no sólo a perder el dinero invertido sino a pagar el costo de un asesoramiento que no rindió lo esperado.

En este trabajo, se propone usar el valor histórico de la acción de Apple para encontrar patrones en el comportamiento de la tendencia del precio y así poder predecirlo con un margen de error acotado y de esta manera asistir al operador de bolsa.

Este trabajo muestra dos tipos de predicción una a corto plazo (valor de la acción al día siguiente) y otra a largo plazo que es el valor de la acción a 365 días.

2. Definición del problema

Todos los días surgen nuevas plataformas y formas para invertir en el mercado de valores, a su vez cada día hay más gente invirtiendo directamente y no a través de un fondo o banco de inversión. Esta nueva masa de operadores suma aún más complejidad e imprevisibilidad al ya dinámico mundo financiero.

Al iniciar la rueda bursátil, todos estos operadores se encuentran con un confuso entorno donde los precios de las acciones cambian constantemente. Esto genera en todos los inversores bursátiles la siguiente pregunta: ¿Cómo mejorar el rendimiento del portfolio? Esta problemática es la que se intenta resolver en esta investigación.

3. Investigación

En este apartado se detalla el proceso utilizado en este trabajo de investigación. El mismo cuenta con una etapa de extracción de datos, luego estos datos son analizados y descriptos para tener conocimiento del dataset con el que se trabaja, siguiendo los métodos descriptos en el artículo “Exploratory Data Analysis: the Best way to Start a Data Science Project” [5].

Una vez analizados los datos se generan los primeros modelos predictivos. Estos modelos ponen en evidencia la necesidad de nuevos datos, los cuales deben ser extraídos, almacenados y agregados a los anteriormente obtenidos. Siguiendo el modelo predicción a escala orientado al negocio que intenta incluir al analista en el ciclo de la predicción a escala.[6]

Los valores de una acción pueden variar por un sin fin de motivos endógenos y exógenos a la compañía estudiada. Entre los endógenos encontramos los siguientes: nuevos lanzamientos y/o tecnologías desarrolladas, resultado de ventas, contratos globales, etc. Entre los exógenos podemos encontrar: valores de otros commodities, crisis económicas mundiales y pandemias entre otras catástrofes. Esta investigación incluye solamente variables endógenas como ser los valores de la empresa en la bolsa y las fechas de lanzamiento de nuevos productos

La investigación propone un modelo de predicción a largo plazo para que los operadores de bolsa puedan armar sus estrategias financieras y otro modelo a corto plazo para poder realizar rápidas correcciones.

La predicción a largo plazo busca responder a la pregunta ¿Qué valor tendrá la acción dentro de un año? Esta pregunta incluye otras incógnitas que deben ser analizadas y respondidas para poder responder la pregunta original. Entre las interrogantes intermedias encontramos: ¿Cuál es la tendencia del valor de la acción? ¿Hay periodicidad en la misma? Por nombrar algunas de las preguntas que aparecen.

Es por esto que al analizar la predicción a largo plazo se explican y detallan las virtudes del algoritmo de predicción “Prophet” [6] desarrollado por la empresa Facebook. El mismo es público y de uso libre.

La predicción a corto a plazo busca responder a la pregunta ¿El valor de la acción será mayor mañana? Esta pregunta puede ser replanteada de la

siguiente manera. Considerando el pasado y dado el valor actual de la acción, quiero clasificar la acción del día de mañana en dos grupos: uno donde su valor sube y otro grupo donde su valor no sube. Dado que solo hay dos grupos y que ambos son mutuamente excluyentes, se pueden utilizar los algoritmos de Machine Learning de clasificación binaria descritos en el libro “Introducción al aprendizaje estadístico” [7].

3.1. Extracción/Obtención de datos

Se utilizó un set de datos libre que se encuentra publicado en el sitio [kaggle.com](https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs). El mismo se puede descargar con el siguiente comando o desde el sitio.

```
kaggle datasets download -d
  borismarjanovic/price-volume-data-for-all-us-stocks-etfs
  -p ../data/raw
```

Con los datos descargados, se continuó con el proceso de exploración y análisis de los mismos. Buscando patrones y transformaciones que permitan la predicción de sus comportamientos.

3.2. Análisis estadístico de los datos

El dataset con el que se va a realizar la investigación cuenta con 7 columnas y 8364 filas. A continuación, se describe el contenido de cada columna:

ID: representa el ID de la observación.

Date: representa la fecha del registro.

Open: es el valor con el que la acción abrió ese día.

Close: es el valor con el que la acción cerró ese día.

High: representa el máximo valor de la acción ese día.

Low: representa el mínimo valor de la acción ese día.

Volume: representa la cantidad de acciones vendidas medidas por la cantidad de dinero.

ID	Date	Open	High	Low	Close	Volume
1	10/09/1984	0.42388	0.42516	0.41366	0.42134	18022532.0
2	11/09/1984	0.42516	0.43668	0.42516	0.42902	42498199.0
3	12/09/1984	0.42902	0.43157	0.41618	0.41618	37125801.0
4	13/09/1984	0.43927	0.44052	0.43927	0.43927	57822062.0
5	14/09/1984	0.44052	0.45589	0.44052	0.44566	68847968.0

Tabla 1: *Topo del conjunto de observaciones utilizado en la investigación.*

La Tabla 1 muestra los primeros 5 registros del dataset a modo de ejemplo.

La figura 1 muestra la evolución del valor de Open de la acción de Apple desde sus orígenes. En la misma se observa que hasta el 2005 el valor de la acción se mantiene bastante estable y que a partir de ese año el valor de la acción empieza a mostrar cierta volatilidad con una clara tendencia alcista. La figura 2 muestra con mayor detalle la evolución del valor de la acción a partir de Enero 2005.

Apple stocks price evolution

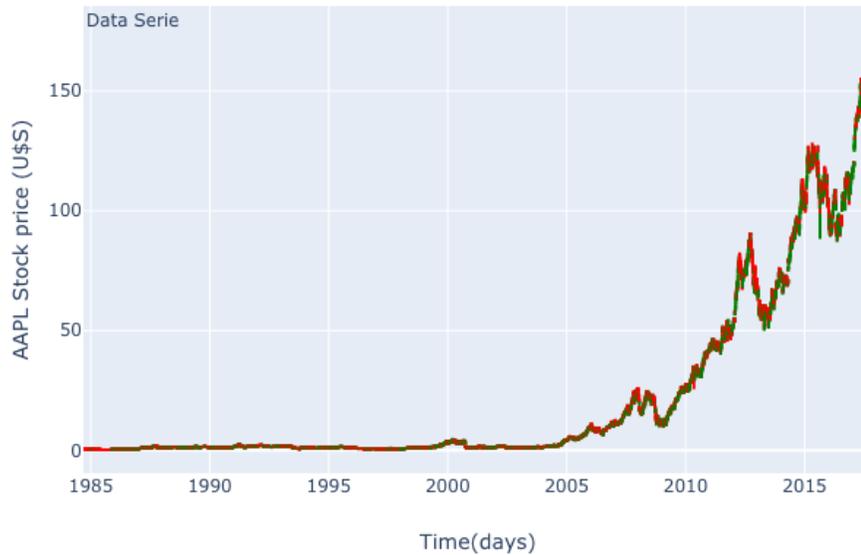


Figura 1: *Evolución del valor de la acción de Apple en el tiempo desde 1984.*

Apple stocks price evolution



Figura 2: *Evolución del valor de la acción de Apple en el tiempo desde 2005.*

Con esta observación para el modelo a largo plazo se decidió partir el dataset en dos partes tomando los valores desde Enero 2005 hasta Noviembre 2016 para entrenar los algoritmos y los datos restantes se utilizan para evaluar el modelo. La figura 3 muestra como queda particionado el dataset.

Con el objetivo de predecir el valor futuro de la acción, se analizó el dataset en búsqueda de algún patrón o correlación entre los datos. Para ello se usaron datos históricos y transformaciones de los mismos.

Las figuras 4 y 5 comparan la evolución del precio de la acción para una fecha y el día siguiente. En las mismas también se muestra la diferencia entre ambos, con la curva de color rojo. Las líneas horizontales de color púrpura son cotas de la diferencia a $\pm 2\%$ del valor del día anterior. La figura 4 compara el valor de la acción al abrir el mercado desde Enero 2005 mientras que la figura 5 compara el valor desde Noviembre de 2016.

Las figuras 6 y 7 muestran el volumen de acciones vendido por día. Mientras que la figura 6 reporta la evolución del volumen de acciones operado desde el 2005, la figura 7 lo describe desde Noviembre 2016.

Apple stocks price evolution



Figura 3: Corte en el dataset de entrenamiento y prueba.

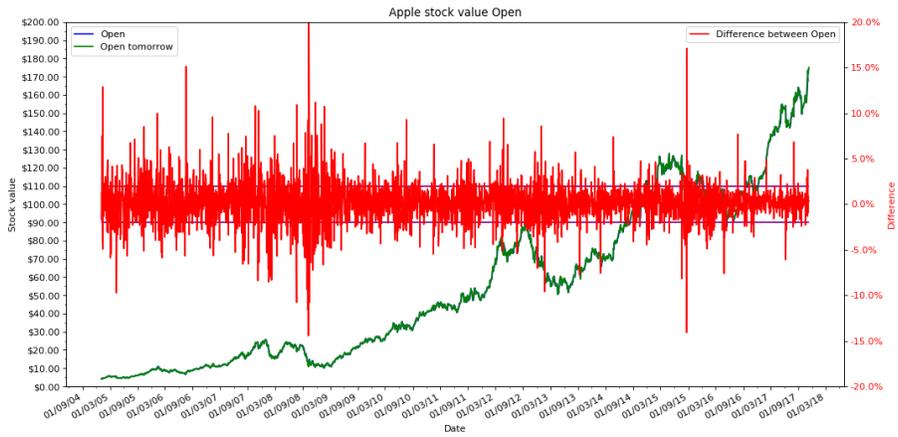


Figura 4: Comparación del valor de Open entre un día y el siguiente desde Enero 2005.

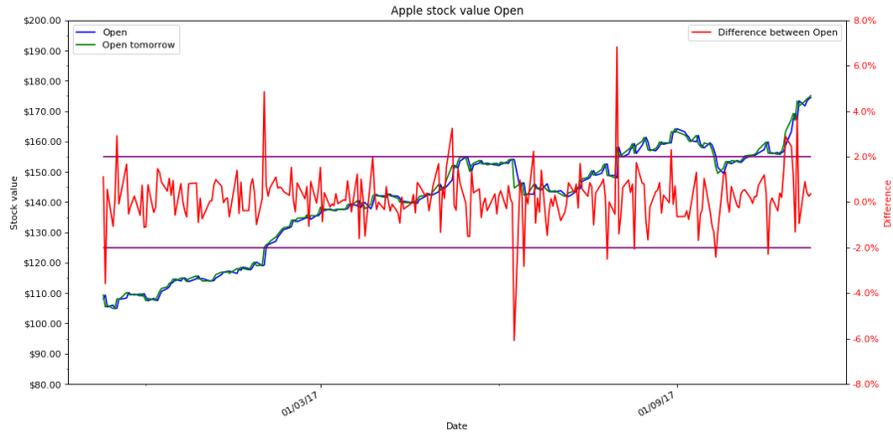


Figura 5: Comparación del Open entre un día y el siguiente desde Noviembre 2016.

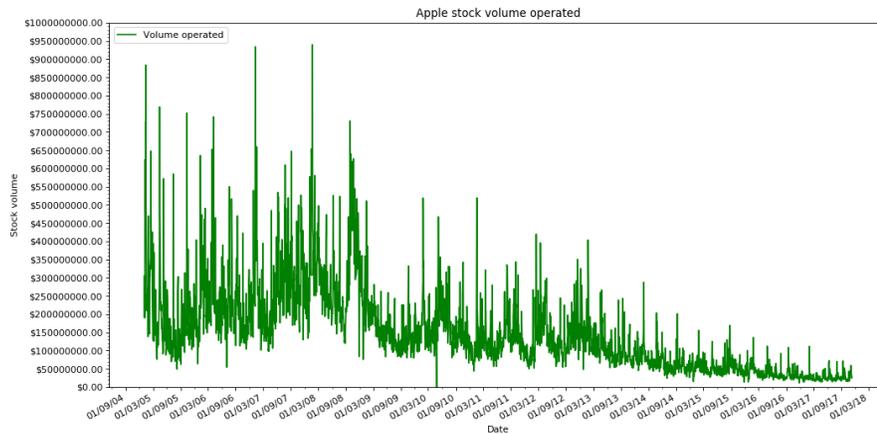


Figura 6: Evolución del volumen de acciones transaccionado desde Enero 2005.

Las figuras 8 y 9 muestran nubes de puntos donde se relaciona el valor de Open con el volumen para una determinada fecha. En esas figuras se pueden observar puntos rojos y puntos verdes. Los puntos verdes representan combinaciones de volumen y valores de Open que generaron un incremento en el valor del día siguiente. Mientras que los puntos rojos generaron una disminución en el valor del día siguiente. La figura 8 muestra la nube de

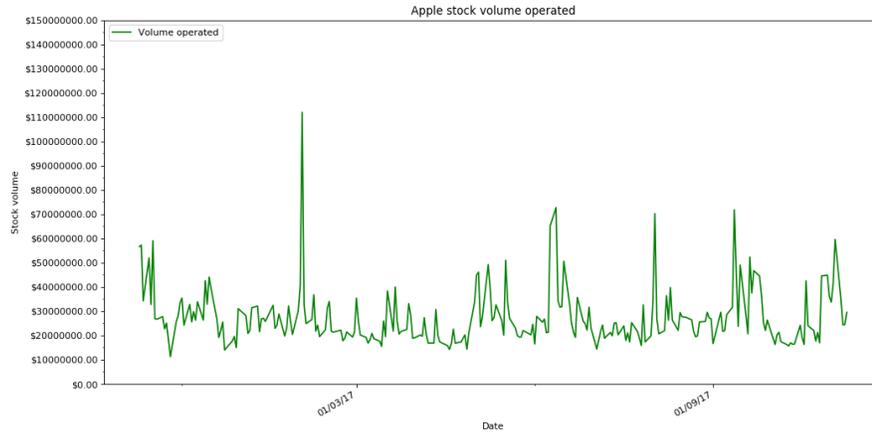


Figura 7: *Evolución del volumen de acciones transaccionado desde Noviembre 2016.*

puntos desde 2005 mientras que la figura 9 describe la nube de puntos desde 2016.

Del análisis de estas figuras parece que no hay ninguna correlación visible entre los datos.

Relation between Volume and Close Value log scale since 2005

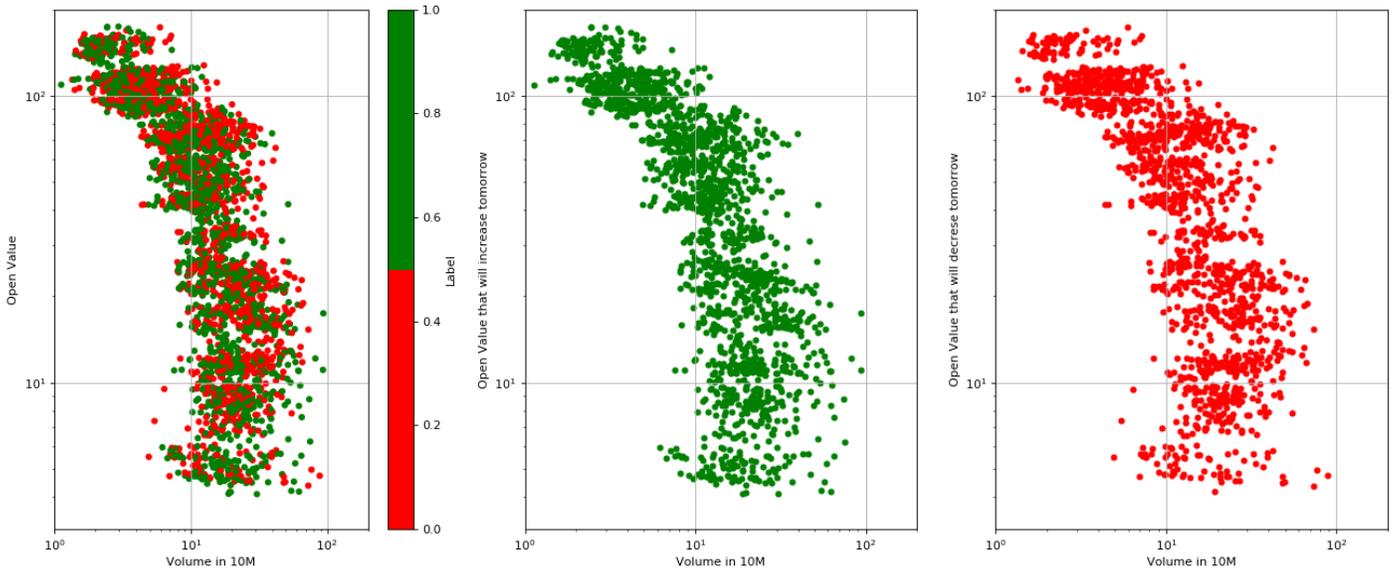


Figura 8: *Relación entre volumen y precio desde Enero 2005.*

Relation between Volume and Close Value log scale since 2016

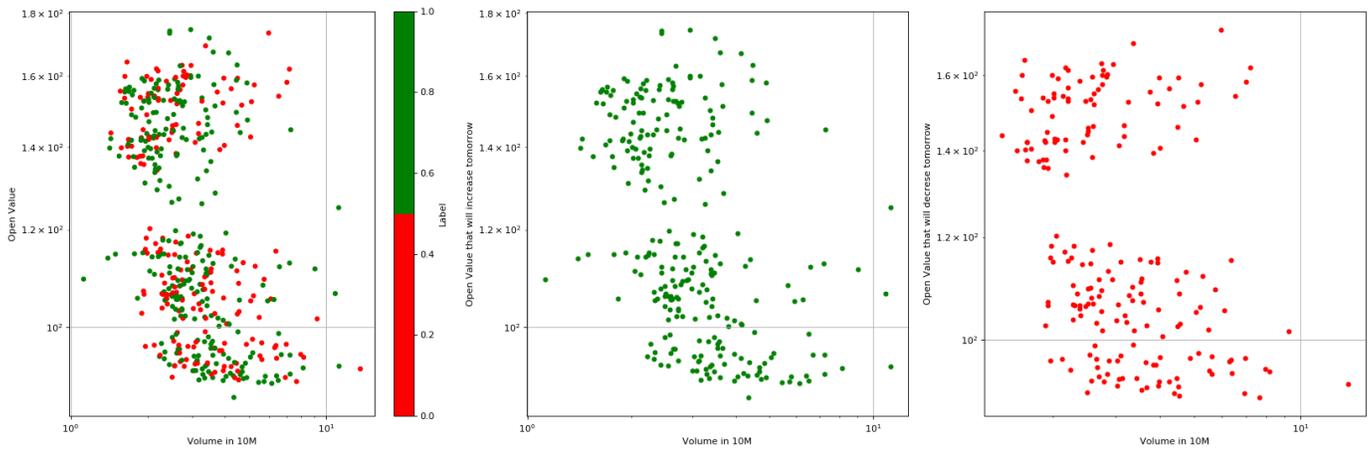


Figura 9: *Relación entre volumen y precio desde Noviembre 2016.*

3.3. Predicción a corto plazo

Como se mencionó en el apartado anterior, la predicción a corto a plazo busca responder a la pregunta: Considerando que quiero clasificar la acción según su valor suba o no el día de mañana. ¿A qué grupo pertenece la acción el día de hoy?

Dado que solo hay dos grupos y que ambos grupos son mutuamente excluyentes, se pueden utilizar los algoritmos de Machine Learning de clasificación binaria. Para la predicción a corto plazo se comparan varios modelos los cuales son descritos y analizados.

Con este fin se creó una nueva columna al dataset llamada 'Label' que representa la variable a ser estimada. Esta variable toma el valor positivo si al día siguiente la acción sube y toma el valor 0 si la acción se mantiene o baja su valor.

Para las predicciones a corto plazo, el dataset fue particionado de manera aleatoria generando una partición de training y de testing que estén balanceadas y de esa manera no sesgar el entrenamiento de los distintos algoritmos.

3.3.1. Modelo Lineal

Es el modelo más simple de clasificación y se lo usa para medir si el esfuerzo y el costo de implementación de otros modelos tiene sentido.

Dado que en el dataset de entrenamiento se obtuvo que el 50.9% de los casos muestran que la acción bajaría su valor, este es el grupo más frecuente y por ende el que el algoritmo intentará predecir.

Al evaluar el modelo obtenemos la matriz de confusión que muestra la figura 10 de la cual se obtiene que la exactitud del modelo es del 49.26%.

La figura 11 compara la predicción realizada por el algoritmo con el comportamiento real del mercado.

3.3.2. Modelo Logístico

La regresión logística es un modelo sencillo y muy utilizado. Este modelo, a diferencia del anterior, predice tanto casos positivos como negativos. Otra de sus virtudes es que posee parámetros de performance que pueden ser modificados para mejorar la precisión del algoritmo dependiendo de los datos que está prediciendo.

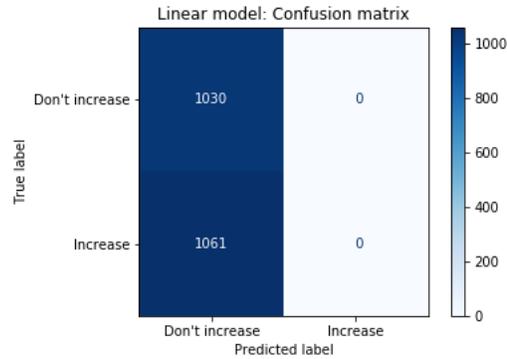


Figura 10: *Matriz de confusión describiendo los resultados del modelo lineal.*

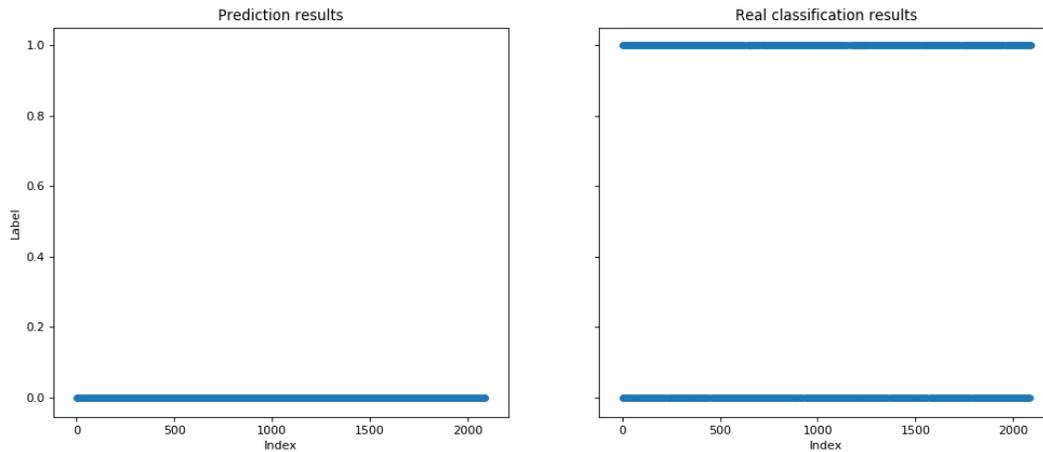


Figura 11: *Comparación de resultados entre el modelo lineal y el de la realidad.*

En la etapa de entrenamiento como en la de evaluación se utilizó el mismo set de datos que con el modelo anterior.

Durante el entrenamiento se utilizó el algoritmo de grid search para encontrar los parámetros que permiten al algoritmo realizar una mejor predicción. También se utilizó la técnica de estandarización de variables para que el algoritmo mejore su performance.

Los resultados obtenidos de combinar estas técnicas y algoritmos se observan en la matriz de confusión de la figura 12. De la misma obtenemos que la exactitud del mismo es del 86.18% y la precisión también es del 86%. Esta es una mejora muy significativa sobre el resultado anterior. La figura 13 compara la predicción realizada por el algoritmo con el comportamiento

real del mercado.

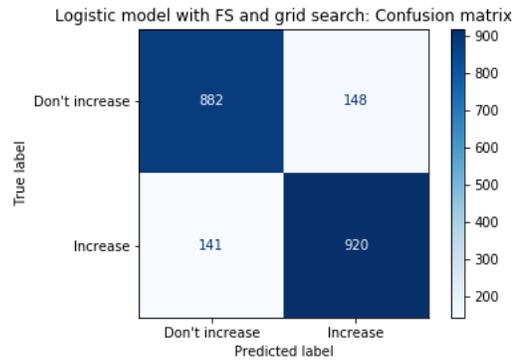


Figura 12: Matriz de confusión describiendo los resultados del modelo logístico.

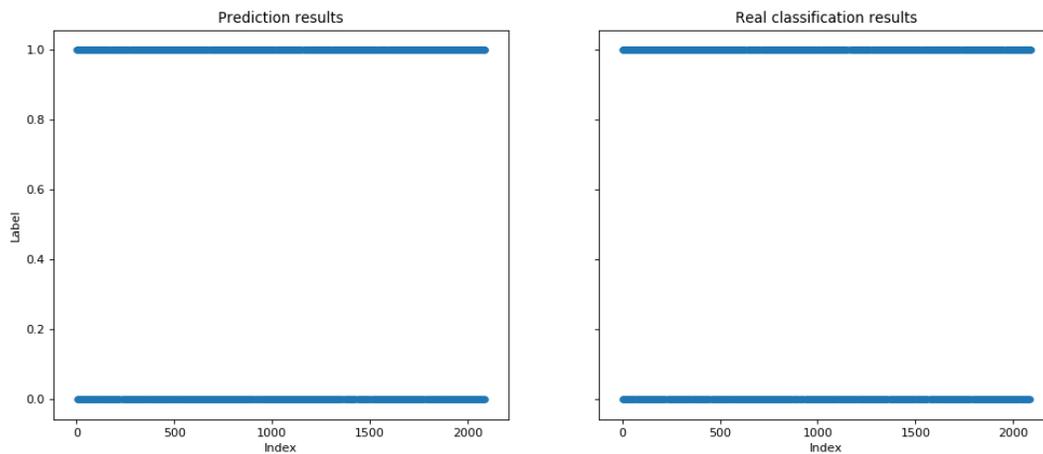


Figura 13: Comparación de resultados entre el modelo logístico y la realidad.

3.3.3. Modelo Random Forest

La técnica de random forest es un modelo de árboles, que también predice tanto casos positivos como negativos y también posee parámetros de performance que pueden ser modificados.

En la etapa de entrenamiento como de evaluación se utilizó el mismo set de datos que con el modelo anterior. También se utilizó el algoritmo de grid search para encontrar los parámetros de mayor rendimiento.

Los resultados obtenidos de combinar estos algoritmos se observan en la matriz de confusión de la figura 14. De la misma obtenemos que la exactitud es del 86.3 % y la precisión es del 86.9 %. La figura 15 compara la predicción realizada por el algoritmo con el comportamiento real de mercado.

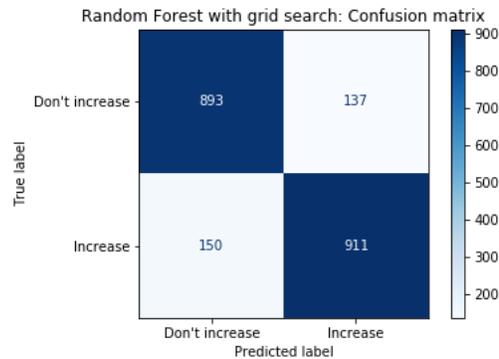


Figura 14: *Matriz de confusión describiendo los resultados del modelo random forest.*

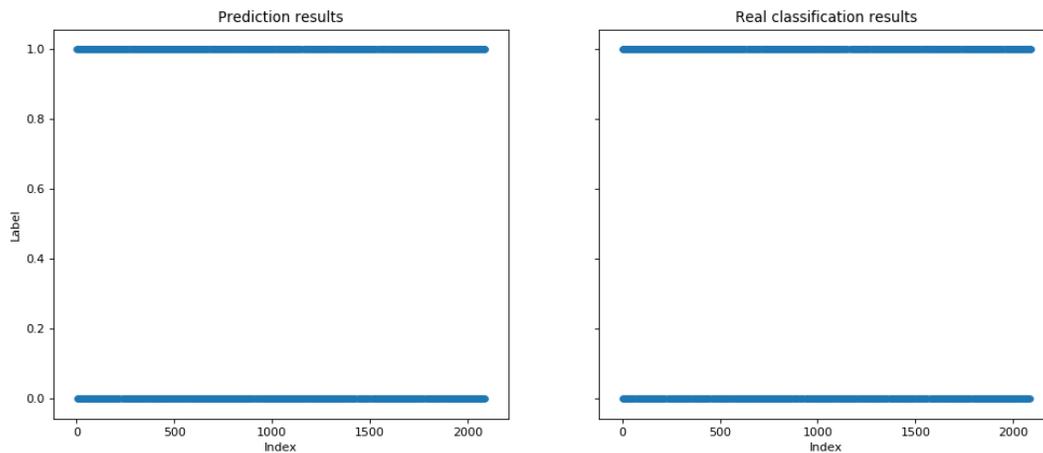


Figura 15: *Comparación de resultados entre el modelo random forest y la realidad.*

3.4. Predicción a largo plazo

El objetivo de esta predicción es que el agente pueda tener una noción con cierto margen de error de cuanto va a valer el valor de una acción en el futuro, basado solamente en los valores pasados.

Con esta predicción se intenta responder a la pregunta ¿Cuánto valdrá la acción dentro de un año? Para poder responder eso, primero debemos responder o al menos preguntarnos una serie de interrogantes: ¿La curva del valor muestra alguna tendencia que podamos proyectar a futuro? ¿Cuáles son los valores previos a la fecha a estimar? ¿Cuál es el error esperado en la estimación?

Dado que para esta estimación necesitamos estimar una variable continua (precio) en el tiempo, se utiliza un algoritmo del tipo serie de tiempo. El algoritmo utilizado es el Facebook “Prophet”. El mismo fue recientemente publicado y su objetivo principal es poder determinar el momento de los nuevos posteos de sus usuarios, así como la cantidad de hits que reciben.

Para poder responder a las preguntas antes mencionadas se fueron realizando diferentes experimentos con el mismo algoritmo. El dataset utilizado como base consta de dos columnas una con la fecha y la otra con el valor de “Open” de la acción en ese día. El periodo analizado es desde Enero 2005 hasta Noviembre de 2017.

El experimento consiste en predecir el valor y tendencia de la acción a un año en el futuro. Para el caso del experimento es hasta el Noviembre 2018.

Al final del apartado se analizan y comparan los distintos experimentos realizados.

3.4.1. Prophet simple

Es el modelo más sencillo de este algoritmo y se lo usa para comparar el esfuerzo y el costo de implementación de otras configuraciones del algoritmo.

La Figura 16 manifiesta el comportamiento del algoritmo y la predicción que genera. En el gráfico de la derecha se muestra cómo el algoritmo “siguió” la curva real, también se observa el intervalo de confianza en el cual el algoritmo estima que el valor real debería estar. El gráfico de la izquierda señala la tendencia hallada, así como los puntos de inflexión de la misma.

En el caso ideal se esperaría que, durante la etapa de entrenamiento, el valor real se encuentre en el medio del intervalo de confianza, para todos los valores de la etapa de entrenamiento. Más aun, se esperaría que la curva de la tendencia se superponga con la curva de los valores reales.

3.4.2. Prophet lineal

El algoritmo posee un conjunto de parámetros que pueden ser modificados para mejorar su rendimiento. En este apartado se muestra la configuración algoritmo en su modo lineal..

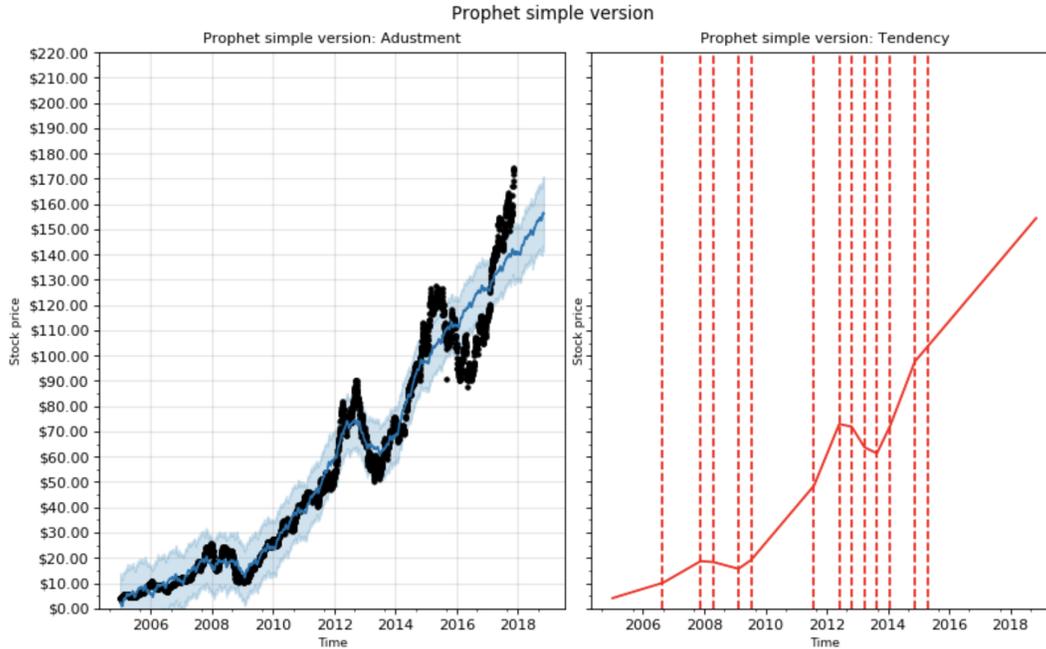


Figura 16: Comportamiento del algoritmo Prophet modo simple.

La figura 17 describe el mejor caso del modo lineal. En el lado izquierdo de la figura se observa como “siguió” a los valores reales y en la derecha se observa la tendencia y los puntos de inflexión que encontrados.

Se observa una gran mejora en la capacidad del algoritmo de seguir la curva real. También se puede entender que el intervalo de confianza de los primeros seis meses de estimación son muy pequeños mientras que en los últimos seis meses de estimación el intervalo crece ampliamente. Esta observación tiene mucho sentido dado que una predicción a mayor distancia en el tiempo acarrea una mayor incertidumbre.

3.4.3. Prophet logístico

Para este experimento se utilizó el modo de predicción logística del algoritmo.

La figura 18 muestra el mismo análisis realizado en el apartado anterior, pero ajustando los parámetros del algoritmo para usar su modo logístico. En el lado izquierdo de la figura se observa como “siguió” a los valores reales y en la figura de la derecha señala la tendencia y los puntos de inflexión que encontrados.

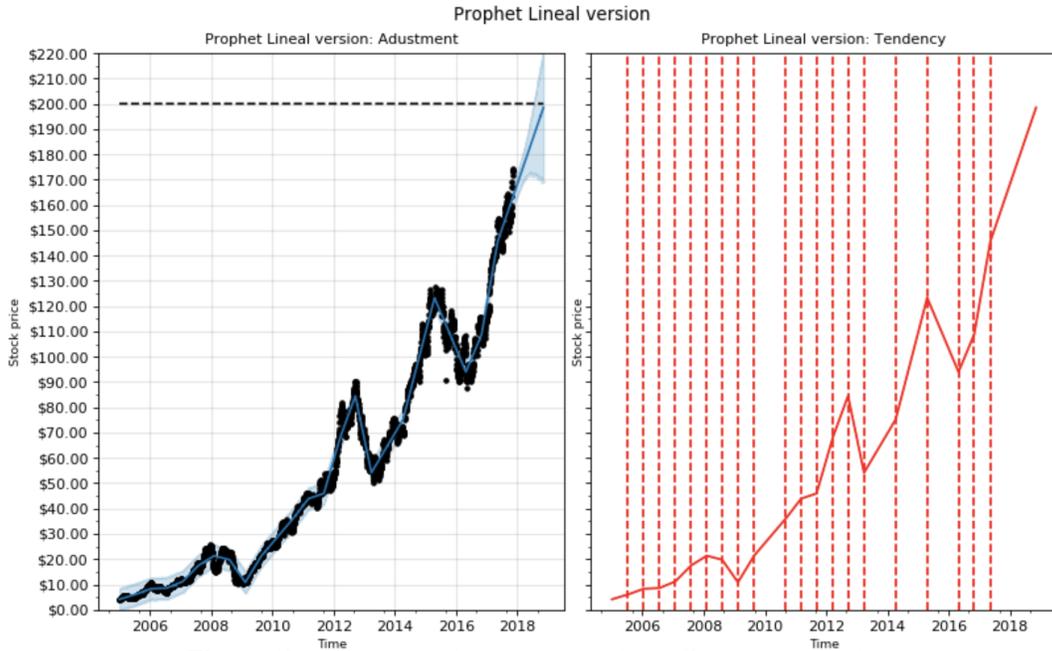


Figura 17: Comportamiento del algoritmo Prophet modo lineal

Se puede concluir que el algoritmo se ajusta bastante bien a la curva real y que el intervalo de confianza en la predicción futura se mantiene muy pequeño comparado con el caso lineal.

3.4.4. Prophet con conocimiento del dominio

Una de las virtudes del algoritmo es que permite incluir conocimiento del dominio para poder mejorar la predicción.

De esta forma, podemos analizar cómo ciertos eventos que Apple organiza afectan el valor de sus acciones. En particular, se tomaron los eventos Apple ‘Key-notes’, en los que la empresa anuncia sus nuevos lanzamientos al mercado. Apple organiza estos acontecimientos regularmente, una o dos veces por año dependiendo el caso.

Para esto, se extrajo del sitio oficial de Apple las fechas de los eventos y se incorporaron estos nuevos datos al análisis. Los mismos fueron incluidos en el dataset de entrenamiento y en el de prueba también dado que se conoce que días se realizará el evento con anticipación.

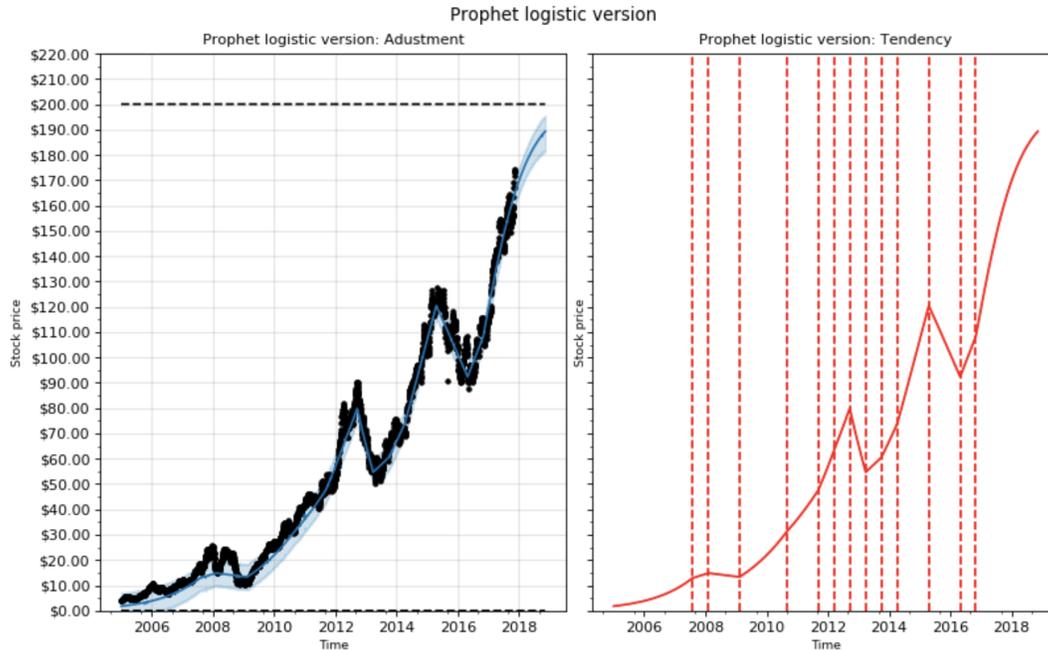


Figura 18: *Comportamiento del algoritmo Prophet modo logístico*

Primero se analizó con estos eventos como días feriados y luego como regresores en el algoritmo. Para este análisis se repitieron las experiencias anteriores para ver cómo el conocimiento del dominio impacta en la estimación. La figura 19 muestra una comparación del resultado de utilizar los eventos ‘key notes’ como días feriados o vacaciones sobre los 3 modos del algoritmo. Mientras que la figura 20 describe el resultado usando los eventos como regresores.

De este análisis queda en evidencia que la comparación visual no es suficiente para poder determinar que configuración es mejor.

En el siguiente apartado se analiza la técnica de “Cross Validation” para poder determinar el margen de error de cada configuración del algoritmo.

3.4.5. Métricas para medir performance

En este apartado se analizan las métricas obtenidas para poder comparar los resultados de las distintas configuraciones del algoritmo. Se utilizó la técnica de Cross Validation para entrenar y testear cada configuración del algoritmo. El experimento consistió en proveer al algoritmo de cross validation

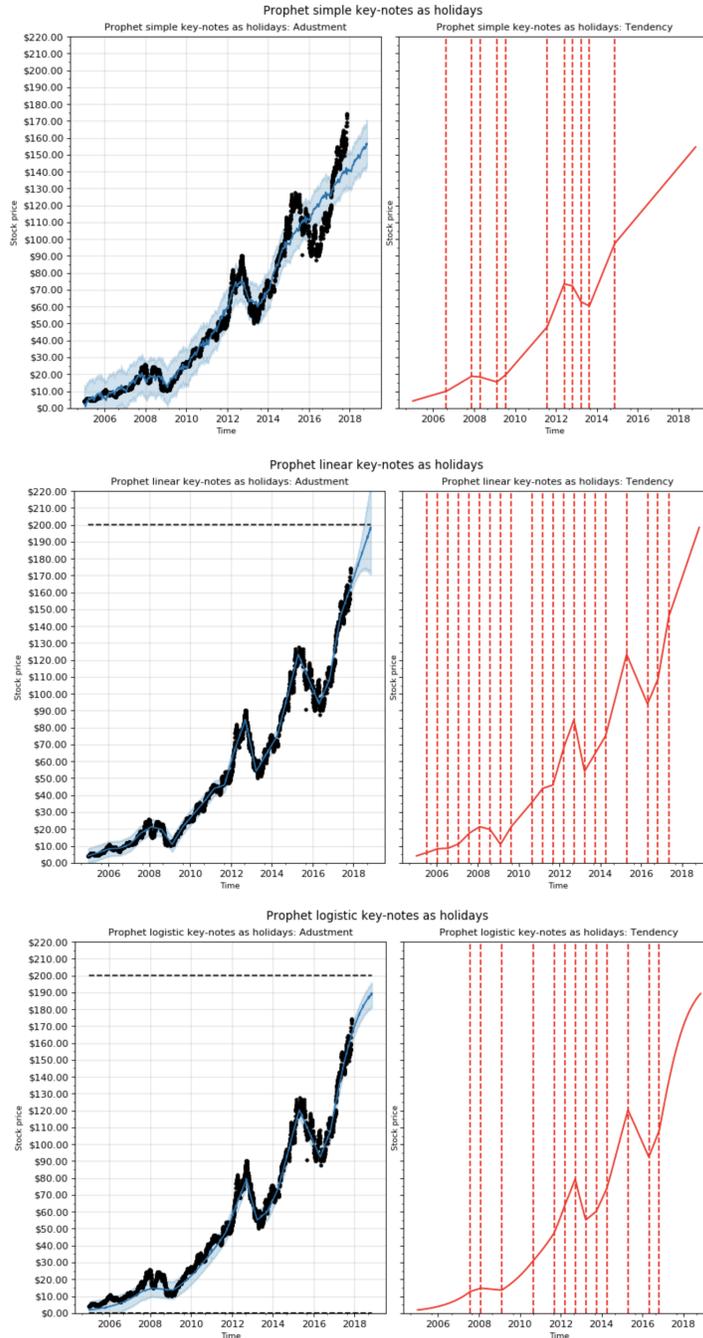


Figura 19: Comparación de resultados usando los eventos 'Key-Notes' como feriados

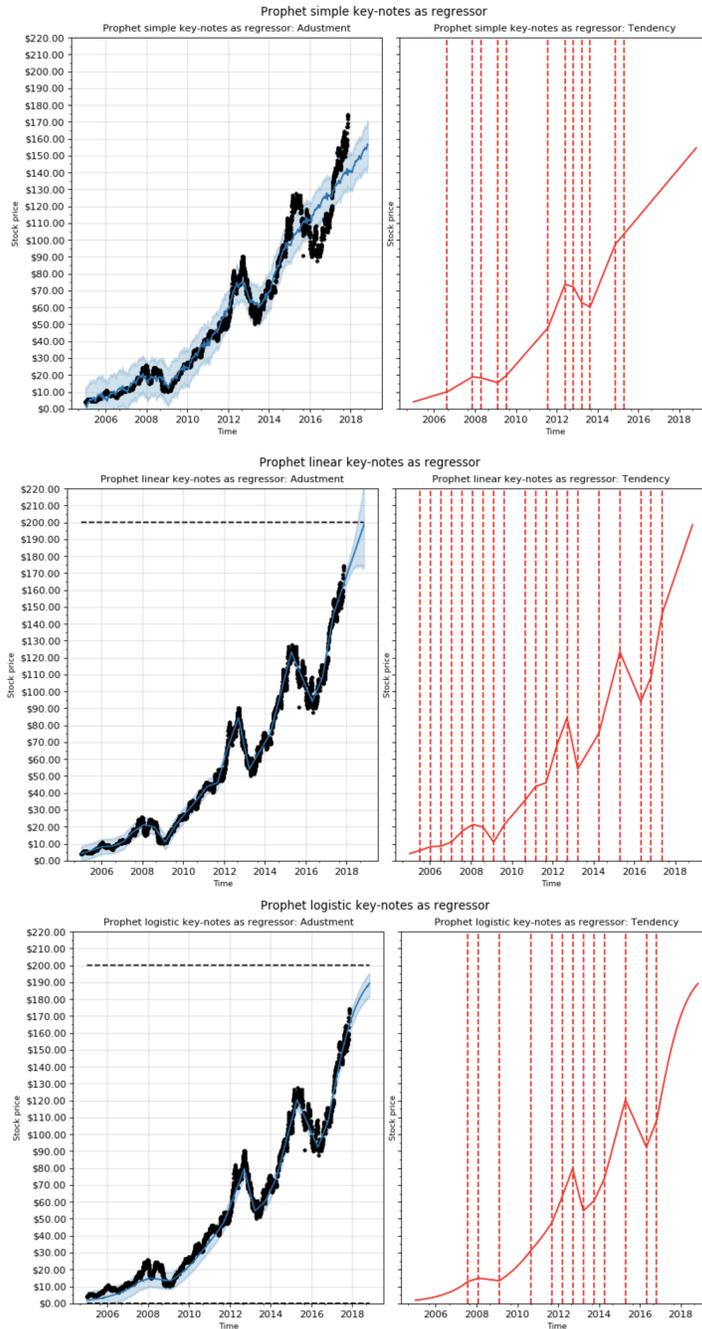


Figura 20: Comparación de resultados usando los eventos 'Key-Notes' como regresores

con cada configuración del predictor, una cantidad de días de entrenamiento y otra cantidad de prueba. De esta manera se optó por un periodo de entrenamiento desde Enero 2005 hasta Noviembre 2016 y el periodo de prueba es de un año y comprende desde Noviembre 2016 hasta Noviembre 2017.

La métrica utilizada para evaluar la virtud de cada configuración del algoritmo es el error porcentual absoluto medio, en adelante abreviado como MAPE por sus siglas en inglés. Las figuras 21 y 22 indican a modo de resumen una comparación de las métricas obtenidas por cada configuración descrita en los apartados anteriores. Estas figuras muestran la evolución del error día por día. Para obtener un único valor con el cual comparar la performance del algoritmo, se hizo un promedio de los errores de todos los días estimados. La tabla 2 describe el error promedio del algoritmo para cada una de las configuraciones mostradas en los apartados anteriores.

De estos resultados no se observan mejoras significativas en las distintas formas de utilizar el conocimiento de negocio. Se puede concluir que la incidencia del mismo en la estimación a largo plazo del valor de la acción es irrelevante comparada con la performance del algoritmo por sí mismo.

Mode	MAPE	
	Key-Note as Holidays	Key-Note as regressor
Simple	0.207511	0.207827
Lineal	0.056519	0.056348
Logístico	0.085881	0.086216

Tabla 2: Promedio de los errores de la estimación a 1 año por configuración del algoritmo.

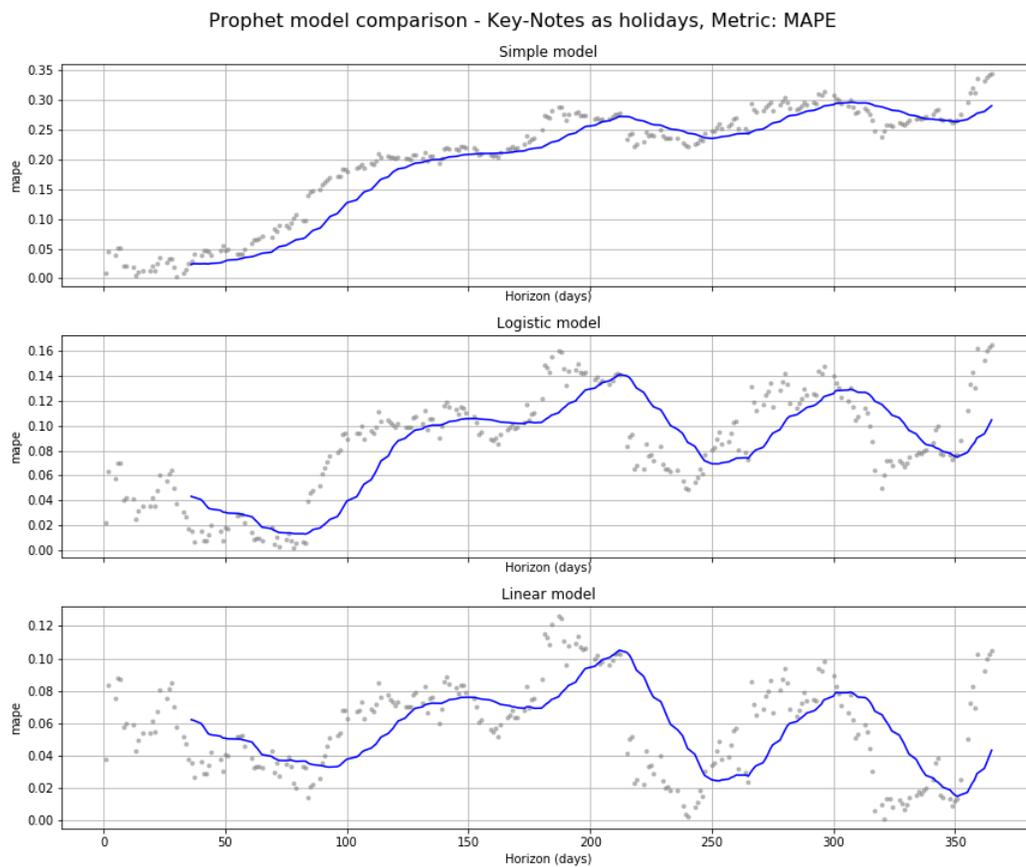


Figura 21: Comparación del error medio porcentual con los 'Key-Notes' como feriados

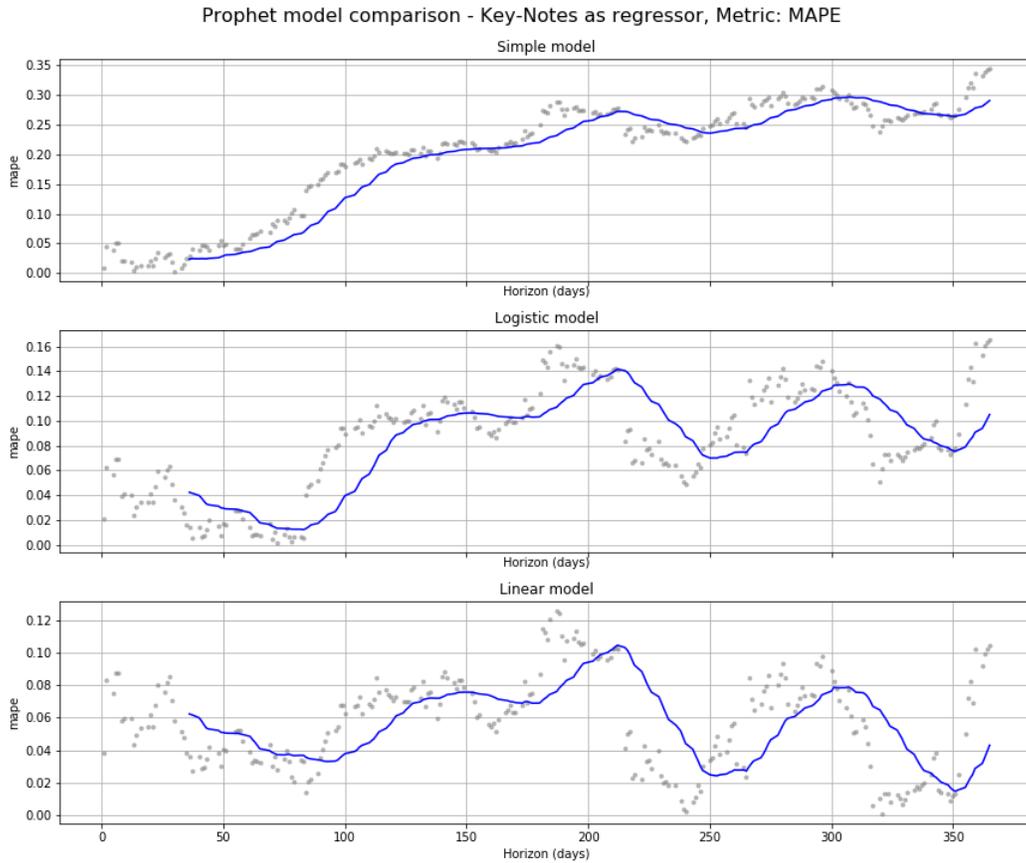


Figura 22: Comparación del error medio porcentual con los 'Key-Notes' como regresor

4. Conclusiones

Esta investigación inició con el propósito de proveer una herramienta de estimación al analista financiero u operador bursátil. El objetivo de la herramienta es brindar información para la toma de decisiones y que el operador bursátil pueda manejar su portfolio y crear estrategias que maximicen su ganancia. En este sentido, se obtuvieron resultados en las tres áreas que se investigaron: a) análisis estadístico/exploratorio de los datos; b) predicciones a corto plazo y c) predicciones a largo plazo.

La variación del precio de la acción mantiene una clara tendencia alcista si bien la misma parece estar desacelerándose. En el periodo comprendido

entre Noviembre 2016 y Noviembre 2017, las variaciones oscilan entre un -2% y un +2% de su valor anterior, con muy pocos casos que superan esas barreras. Al mismo tiempo, se observa que el volumen operado está disminuyendo.

Ambos comportamientos están relacionados: a medida que el valor de la acción se vuelve más estable, las ganancias obtenidas por su compraventa se reducen, lo que desalienta a muchos inversores ávidos de ganancias.

No se encuentra una clara relación entre el volumen operado, el valor de Open y la variación del precio de la acción.

Al crear el modelo de predicción a corto plazo, vemos que el modelo lineal sugiere que la acción mantiene o baja su valor mas veces que las que sube. Este modelo obtiene una exactitud del 49.3%. Este resultado no es bueno, no sólo por su baja performance sino que la predicción no es confiable por que siempre predice lo mismo.

Los modelos logístico y random forest ofrecen una significativa mejora por sobre el modelo lineal a costa de un mayor tiempo de procesamiento y análisis de optimizaciones. Con estos modelos mejoró tanto la exactitud como la precisión, lo que significa que ambos predicen tanto los casos en que la acción sube como los que no.

Al evaluar los modelos a corto plazo vemos que la predicción va mejorando, hasta alcanzar el 86% de casos acertados. Podemos concluir que el modelo de random forest o regresión logística son buenos modelos que pueden ser utilizados por los agentes de bolsa para las estimaciones a corto plazo.

Al crear el modelo de predicción a largo plazo, el mismo fue evolucionando hasta obtener un modelo que predice el valor de la acción con error porcentual promedio del 5.6% en un periodo de 1 año en el futuro. También se observa que intervalo de confianza en el que el algoritmo predice, crece para estimaciones mayores a los 6 meses.

El análisis de la predicción a largo plazo sugiere otro resultado muy significativo: el valor de la acción no se ve influenciado por conocimiento del negocio o al menos no por el conocimiento estudiado en esta investigación.

Futuras investigaciones podrían incluir acciones de otras compañías que presenten rendimientos mixtos, es decir que no presenten una clara tendencia, de esta forma se buscaría un algoritmo más genérico para la recomendación. También se sugiere futuras investigaciones sobre distintas compañías para ver si el conocimiento del negocio influye en la predicción del valor de la misma.

5. Anexo

El código desarrollado para la investigación es de carácter publico y se encuentra en el siguiente repositorio.

<https://github.com/MildoCentaur/stockforecasting>

Referencias

- [1] Carlos Altea *FINTECH ARGENTINA: creció 100 % en el año y va por más*. infotechnology.com, Buenos Aires, 2018.
- [2] *Creció 70 % la cantidad de empresas “Fintech” en los últimos doce meses*. Telam, Buenos Aires, 2019.
- [3] Azahara Benito Carrillo *Se dispara el crecimiento de las empresas Fintech en España*. viafirma.com, Julio 2019.
- [4] *Fintech Global Market Report 2020*. thebusinessresearchcompany.com, Enero 2020.
- [5] InData Labs *Exploratory Data Analysis: the Best way to Start a Data Science Project*. InData Labs, Junio 2017
- [6] Taylor SJ, Letham B. *Forecasting at scale*. PeerJ Preprints 5:e3190v2, 2017
- [7] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani *An Introduction to Statistical Learning* Springer, 2015
- [8] Dr. Alexander Elder *Trading for a living*. John Wiley & Sons, Wiley Finance Editions, New York, 1993.
- [9] John c Hull. *Options, Futures and other Derivatives*. Pearson, New Jersey, 2006
- [10] *Price Forecasting: Applying machine learning approaches to electricity, flights, hotels, real estate and stock pricing* Altex Soft Inc, Bussines and Finance newsletter
- [11] Vinay Arun, *An LSTM model to predict commodity prices* Medium Corporation, Data science newsletter

- [12] Robert C. Merton *Theory of Rational Option Pricing* The RAND Corporation, 1973
https://www.researchgate.net/publication/24048188_Rational_theory_of_option_pricing