



UNIVERSIDAD AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

Análisis Multivariado en el Mercado

Accionario

Montoya Franco Luis Angel

Mtro.

José Gustavo Fuentes Cabrera.

14 de octubre de 2020

Índice

Introducción	4
Objetivos	7
Ciencia de datos aplicada en el mercado accionario	7
1. Modelación No Supervisada	10
1.1. Reducción de la dimensión	10
1.2. Clustering	11
2. Agrupamiento de las acciones del S&P500	13
2.1. El conjunto de datos y sus variables	13
2.2. Compresión de los datos	19
2.3. Análisis de Clustering	20
2.4. Empresas según su cluster	22
2.5. Análisis temporal de los clusters	23
3. Modelación Supervisada	25
3.1. Árboles de Decisión	25
3.2. Redes Neuronales	26
3.3. Máquinas Vector Soporte	27
3.4. K vecinos más cercanos	29
3.5. Gradiente Estocástico	29
3.6. Ensamblés	30
3.7. AdaBoost	30
3.8. RandomForest	31
3.9. XGBoost	32
4. Predicción de los precios usando regresión	34
4.1. Metodología	34
4.2. Resultados	36
Aplicación para los inversionistas	39

Conclusiones	42
Bibliografía	44

Resumen

El siguiente estudio tiene por objetivo analizar el precio de las acciones que conforman el índice mas influyente de la economía norteamericana **S&P 500**, con el fin de encontrar una estructura de tal forma que se proporcione una clasificación de acuerdo a sus características. Por otra parte se dispondrá de los antecedentes a fin de predecir el precio de las acciones mas influyentes de Norteamérica. El estudio se basa en los precios al cierre de las acciones que constituyen el índice mas importante de los Estados Unidos y el Mundo, el índice Standard & Poor's 500 el cual se sustenta en la capitalización bursátil de las 500 empresas mas grandes que poseen acciones en la Bolsa de Nueva York (NYSE) así como en NASDAQ, la razón de hacer uso de estos datos radica en su influencia sobre todas las economías.

Introducción

A lo largo de los años la administración del riesgo ha ido evolucionando de una manera acelerada debido a que para una empresa es vital mantener sus procesos seguros. En el caso del riesgo de mercado tanto para los empresarios como académicos no les ha resultado ser una tarea fácil poder modelar el movimiento de los precios accionarios en el tiempo. Ha

pasado más de un siglo desde que se propuso el primer modelo matemático para poder modelar el precio de las acciones como función del tiempo. Fue Bachelier en "*Théorie de la spéculation*" [Bachelier \(1900\)](#) donde se propuso la regla de Bachelier para modelar el precio de las acciones, la misma regla que Einstein utilizó para modelar el movimiento de partículas de las que no se era capaz de explicar su errático movimiento [Einstein \(1905\)](#), después se vió que ambos usaron el *movimiento browniano*¹. A primera vista podemos comparar un movimiento browniano contra una simulación de precios estos nos resultan muy similares, sin embargo si consideramos las diferencias de los procesos, es decir, lo que crece el precio de la acción notamos que el movimiento browniano se mantiene sin saltos importantes, sin embargo en el caso de los precios accionarios las diferencias tiene conglomerados de crecimientos des-proporcionales como en la Figura 1. Esto se debe principalmente a que se asume que hay normalidad en las diferencias de los precios, esto no siempre se cumple en la vida real (Comparación en [Bouchaud \(2008\)](#)).

¹Movimiento de las partículas en un fluido resultado de colisionar con moléculas en el mismo.

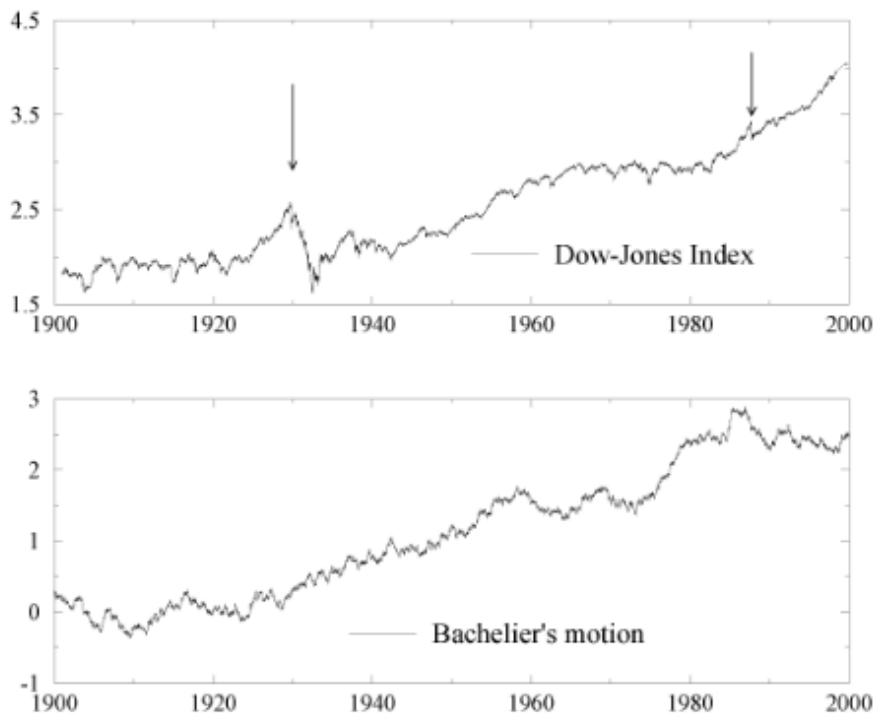


Figura 1: Comparación entre los precios reales y el movimiento browniano

En la actualidad es muy común que se asuma este modelo debido a que es muy sencillo llevarlo a la practica, pero se tienen errores grandes en las estimaciones y en consecuencia nuestras medidas de riesgo tienen errores, en particular el modelo para valuar opciones (call or put) propuesto por Black-Scholes² asume este hecho. Otros modelos mas complicados de llevar a la practica fueron desarrollados por Mandelbrot con base en los trabajos del matemático Paul Levy³. Con esta aportación las series de tiempo de las diferencias tiene saltos que varían en tamaño y no son constantes como en el caso del movimiento brownianoMandelbrot and Hudson (2006). Este nuevo tipo de proceso estocástico es llamado "*multifractal*", el termino fractal se refiere a que en cualquier segmento de tiempo tan pequeño como sea la distribución de los crecimientos tienen una proporción fija. Figura 2. Este modelo resulta ser muy realista. Sin embargo no es determinante para poder modelar en general los precios de las acciones pues tiene la desventaja de ser sensible ante perturbaciones de sus parámetros

² Modelo usado para determinar el precio justo para una opción de compra y venta

³ Levy propone una clase de procesos estocásticos más generales que el movimiento browniano

Sin embargo, gracias a estos modelos los administradores de riesgos tienen mayores herramientas para poder calificar inversiones o poder estructurar sus portafolios de inversión de manera que tengan un rendimiento deseado con el mínimo de riesgo. Aun teniendo modelos nuevos y que tienen un mejor comportamiento en casos de estrés, por lo general se usan modelos que asumen normalidad. Esto es debido a la facilidad con la que se pueden aplicar, pues desde cierto punto de vista asumir que nuestros crecimientos tienen una distribución normal es lo mismo a decir que para describir nuestros datos solo hace falta calcular la media y la varianza, por tanto ser capaces de describir toda la serie de tiempo. Esto sin embargo no resultará ser una limitación cuando hablamos del análisis multivariado pues nos será posible ajustar los parámetros necesarios para poder aplicar de manera correcta el modelo, esto es, tener la libertad de agregar variables que vayan siendo de importancia conforme se lleve el análisis.

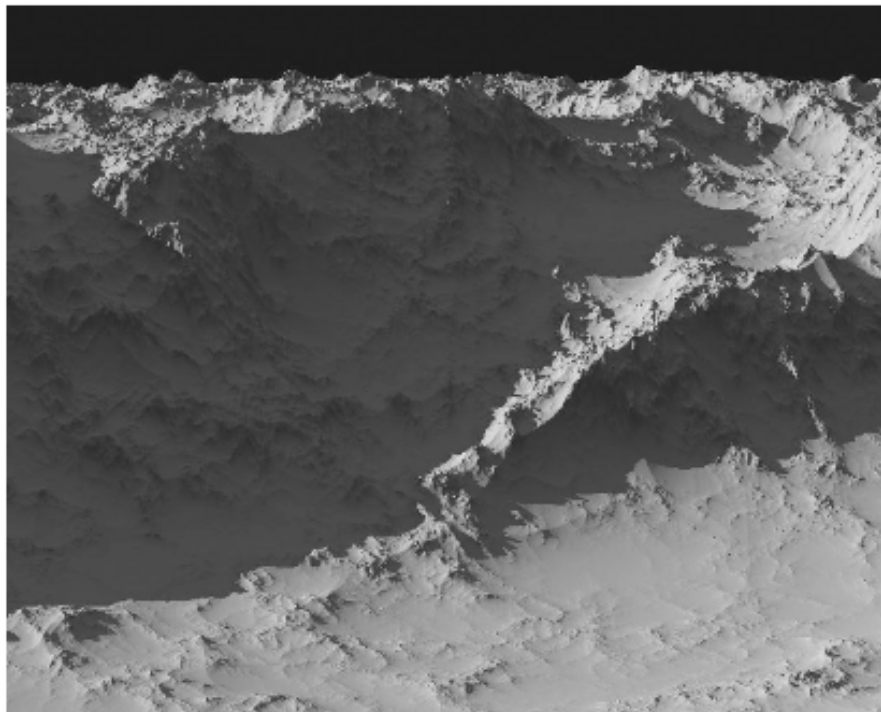


Figura 2: Representación gráfica de un modelo multifractal

Objetivos

- ✎ Etiquetar las acciones que conforman el índice mas relevante en el mercado accionario durante el año 2017, por medio del análisis no supervisado, considerando el valor esperado, volatilidad, ademas de dos estadísticos mas que nos dan una referencia de cual es el comportamiento de la información con respecto a los dos primeros.
- ✎ Conocer los atributos de cada grupo, ademas de estar al tanto de las empresas que conforman cada uno de los conjuntos.
- ✎ Predecir un día de cotización de las principales acciones tecnológicas de Norteamérica ademas de algunas otras que el día de hoy tienen gran presencia a nivel mundial.

Ciencia de datos aplicada en el mercado accionario

Actualmente el uso de Machine Learning se ha incrementado en el contexto de acciones, lo más actual es la incorporación de Procesamiento de Lenguaje Natural (PNL) , área de la inteligencia artificial encargada del estudio de la interacción entre el lenguaje humano y la computadora, entre sus principales componentes se encuentra el reconocimiento de voz y la minería de textos. La creación de reportes, prensa y artículos al respecto ha crecido en los últimos días, de esta forma [Xing et al. \(2018\)](#) llevó a cabo un estudio de PNL en los artículos que tuvieran que ver con acciones o pronósticos en ellas encontrando que desde 1998 se ha incrementado la publicación de trabajos relacionados Figura 3, para luego definir el Procesamiento de Lenguaje basado en Pronósticos Financieros (NLFF por sus siglas en inglés).

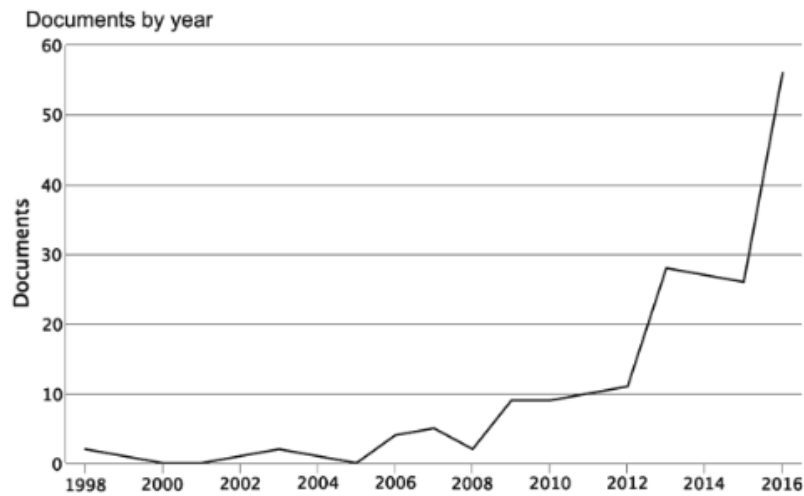


Figura 3: Trabajos publicados 1998-2016

Por otro lado, el mismo autor comenta que si sólo se hace el interés por predecir la dirección de la acción, más no la intensidad del movimiento de mercado, una Máquina Soporte Vectorial sería buena como una clasificador binario, también como regresión puede ser utilizada bajo ciertas modificaciones en los parámetros. Finalmente concluye con que NLFF es un área constituida diferentes disciplinas resumido en la Figura 4 y con el resultado que Análisis de sentimientos tiene una correlación grande con el retorno de mercado (*market returns*) pero con un poder predictivo muy débil a la hora de usarse para un pronóstico a corto plazo (semanalmente).

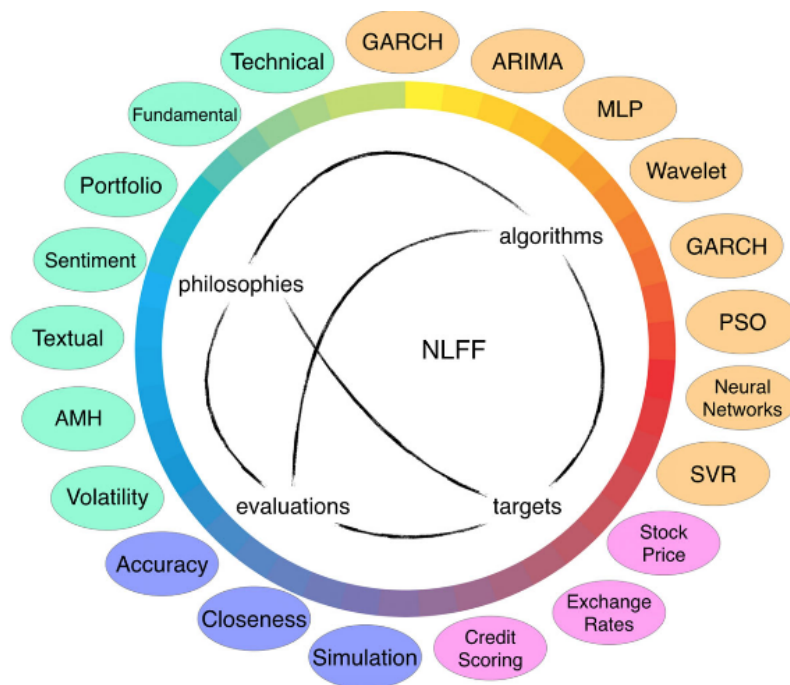


Figura 4: Diagrama de disciplinas usadas en NLFF

Sin embargo, tomemos en cuenta que el mercado accionario es un sistema que es afectado por diferentes factores además de complejas características como la volatilidad y la estacionalidad de esta manera métodos unilaterales de pronósticos (ARMA, GARCH, ARIMA) pueden ser tan confiables; algunos autores concuerdan que ente más robusto sea el algoritmo implica que pueda ser de más ayuda, aunque es cierto que no hay un algoritmo en el que concuerden tenga la mejor estimación, la mayoría de los que son escogidos se basan en la experiencia del investigador o en prueba y error. De esta manera algoritmos basado en procesos biológicos Göçken (2017) han tomado un papel importante por su construcción: aprender y adaptarse como seres biológicos, de los que destacan los evolutivos y basados en insectos (comportamiento colectivo) Binittha et al. (2012), además del uso de Extreme Machine Learning ó redes neuronales predictivas de una sola capa oculta han sido una buena técnica en el pronóstico del mercado accionario Ding et al. (2014); Huang et al. (2006), donde el papel que juega la computación biológica es para encontrar un conjunto de parámetros óptimos para usar algún algoritmo de EML.

1. Modelación No Supervisada

El análisis no supervisado tiene como objetivo inferir el comportamiento descriptivo de los datos, de donde destacan dos técnicas: Clustering y reducción de dimensión. La desventaja que tiene es que no se cuentan con métricas de qué tan bueno es el modelo pues recibe datos no etiquetados, esto reside más en el negocio o aplicación.

Clustering se basa en clasificar la información en un número finito de grupos que tengan sentido dentro del negocio para luego convertirse en un mercado al que se le puede hacer un marketing personalizado, esto se hace por medio de centroides, los cuales actúan como el punto de gravedad del clúster o grupo, a éste caerán todos aquellos que tengan una alta similitud con dicho centroide. Por otro lado, reducción de la dimensión es como compresión de los datos, lograr reducir la complejidad de la información preservando sus características más relevantes.

1.1. Reducción de la dimensión

Componentes Principales

Herramienta utilizada para la reducción de dimensiones con la finalidad de que esta reducción preserve la información necesaria para representar al mismo de gran escala. Su función es transformar un conjunto de variables posiblemente correlacionadas en un número reducido de variables no correlacionadas llamadas **componentes principales**.

Geométricamente es equivalente a una rotación del sistema coordenado original a uno ortogonal, llamados **ejes principales**.

Busca una combinación lineal de variables tal que la máxima varianza sea extraída de estas. Luego, remueve dicha varianza y busca una segunda combinación lineal la cual explica la máxima proporción de la varianza que sobró, esto de forma iterativa. Finalmente llega a un conjunto de variables ortogonales, es decir, no correlacionadas (Factores).

Los valores propios miden la cantidad de variación en la muestra total logrado por cada factor.

Un eigenvalor conceptualmente representa la cantidad de varianza alcanzada por cada factor. La proporción con respecto a estos es la proporción de los factores explicativos con res-

pecto a las variables originales. Si un factor tiene un valor propio pequeño entonces éste contribuye poco a la explicación de la varianza y puede ser descartado.

Una observación importante de esta técnica de compresión o reducción dimensional, más sin embargo, el resultado final, la reducción pueden no tener un sentido interpretable. Más aún, el principal objetivo de esta herramienta es seleccionar un subconjunto de variables de un espacio mayor (gran dimensión) basado en qué variables originales tenían la mayor correlación con el componente principal.

1.2. Clustering

K medias

Algoritmo de clustering que tiene como objetivo etiquetar datos de un determinado conjunto, el fin de este algoritmo es encontrar grupos en los datos, comparando las características de cada elemento con sus k vecinos más cercanos.

El algoritmo para agrupar por K-Means es el siguiente:

Definamos a Ω como el conjunto de n individuos a clasificar, todos con pesos iguales $\frac{1}{n}$.

- Inicialización.

Consiste en escoger al azar k objetos de Ω , los cuales tomaran el papel de nucleos iniciales, esto es:

$g_1, \dots, g_k \in \Omega$ además definiremos C_i como las clases de datos. Así

$C_i := \emptyset, \dots, C_k := \emptyset$

- Asignación

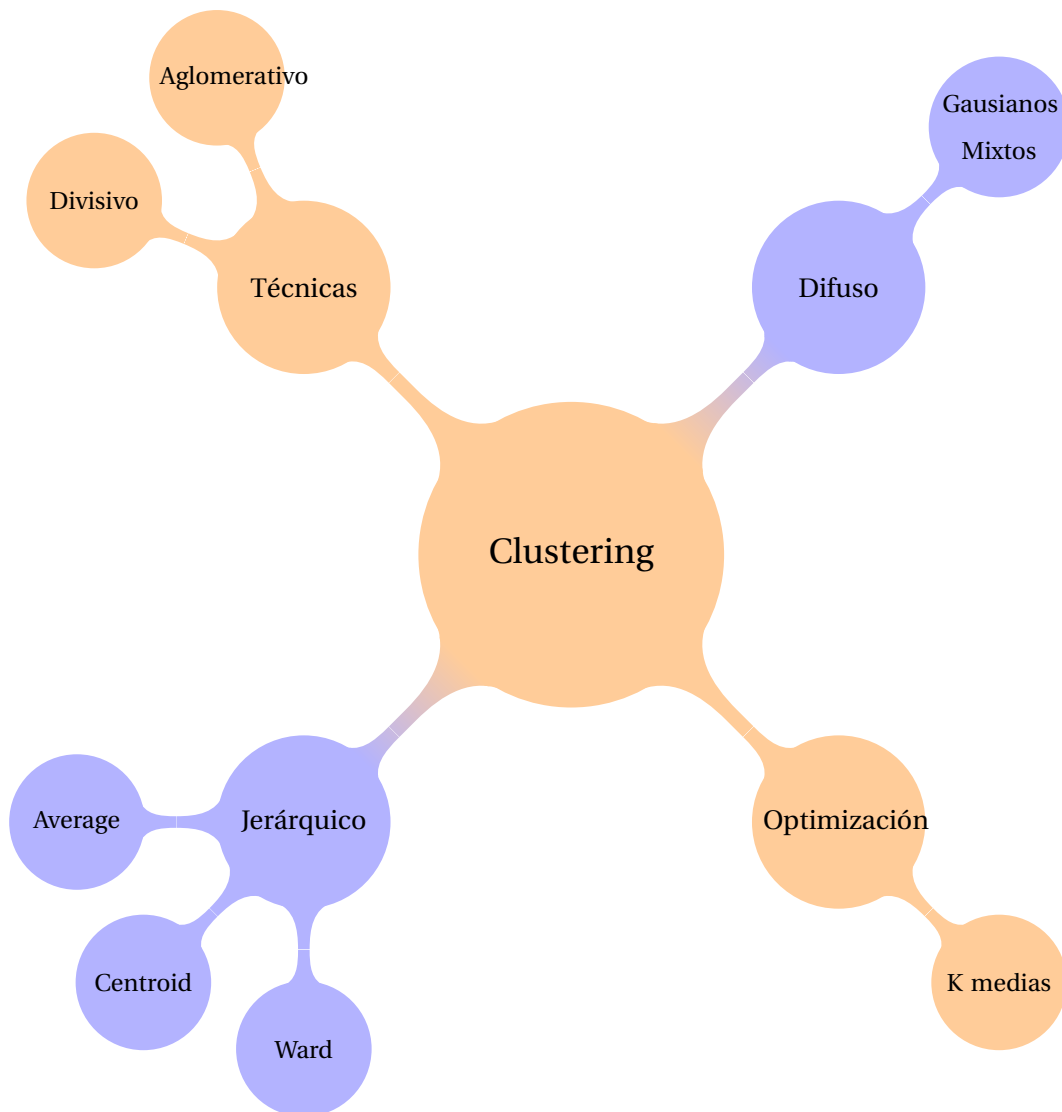
Asignar cada objeto a la clase del centro de gravedad mas cercano, es decir:

si $d(x_i, g_k^*) \leq d(x_i, g_k) \in k = 1, \dots, k$ entonces asignar x_i a la clase C_k^*

- Representación

Calcular los centro de gravedad. $W := \sum_{k=1} \sum_{i \in C_k} \|x_i - g_k\|^2$

Si la variación en el criterio W entre la iteración anterior y la presente, es menor a un umbral dado entonces la iteración se detiene y se forman los clusters finales.



2. Agrupamiento de las acciones del S&P500

El objetivo es clasificar a las 505 empresas mas representativas del mercado a partir del comportamiento de los precio de los últimos meses. Se clasificara mediante el método de clasificación K-Means el cual agrupa a partir de las distancias entre cada observación . El poder etiquetar a cada una de las acciones nos permitirá conocer las características de cada uno de los grupos ,conociendo así los puntos favorables y los menos favorables de cada uno.

2.1. El conjunto de datos y sus variables

Los datos utilizados en el modelo corresponden a datos diarios sobre el precio de cierre de las acciones que conforman el indice S P 500.El periodo para la obtención de datos fue considerado desde el 1 de Enero de 2017 hasta el 29 de Diciembre del mismo año, se eligió el precio al cierre ya que es el que mas se acerca al precio medio ponderado de entre las últimas 500 unidades contratadas, ademas de que es el precio con el cual se inician las transacciones del día siguiente.

Los precios entre acciones pueden diferir de acuerdo a la compañía de la cual se este hablando, por ejemplo, acciones del tamaño de Apple con alguna otra como Goofyear,la diferencia en precios es muy notoria, por lo que si se trabaja con los precios tendríamos valores extremos, lo cual no permitiría una buena clasificación. La forma en que este problema se puede solucionar es tomar la variación porcentual en lugar de los precios diarios.

Veamos el comportamiento de Apple una vez que hemos tomado la variación porcentual ademas de el histograma de la misma acción.

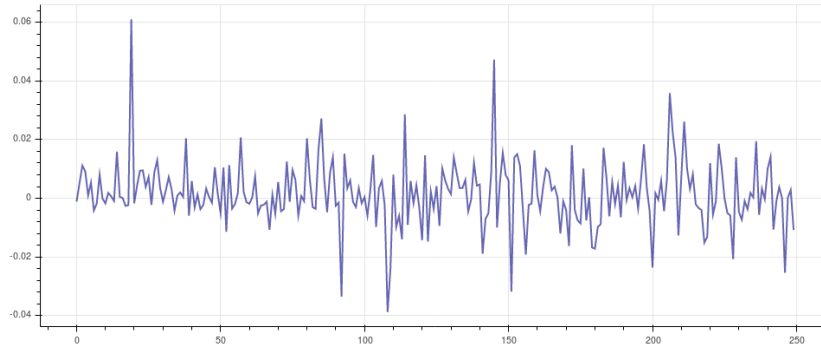


Figura 5: Variacion porcentual de Apple de los últimos meses

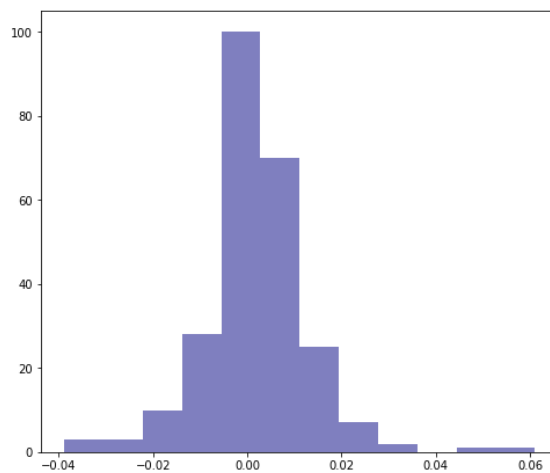


Figura 6: Histograma de rendimientos de Apple

El comportamiento de AAPL refleja la conducta de un cuantioso numero de las acciones que conforman nuestros datos.

Ahora que ya se ha definido los datos con los cuales se van a trabajar es debido definir las variables que determinaran el rumbo de la clasificación. A partir de los datos que poseemos es posible obtener distintos estadísticos de gran interés, en nuestro caso los de utilidad en el trabajo son:

- Media
- Desviación estándar
- Asimetría

- **Curtosis**

Tanto la Media como la Desviación Estándar son estadísticos básicos y conocidos, mientras que la asimetría y la curtosis son poco usados, a continuación proporcionaremos más información acerca de estos estadísticos.

- **Media**

Es lo que en promedio creció la acción diariamente durante el año que estamos estudiando, que este valor sea alto habla bien de la acción.

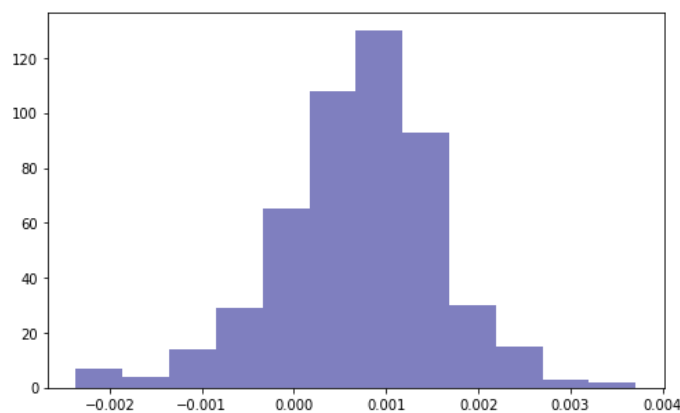


Figura 7: Histograma de la media de los datos

- **Desviación Estándar**

Indica en promedio que tanto se alejaron los datos de la media en promedio durante el año, un valor alto de este estadístico nos dice que se tiene mucha incertidumbre de cuál es el rendimiento de la acción. Mientras un valor pequeño nos indica que se tiene certeza de cómo se comportó la acción y que se mantuvo cerca de la media.

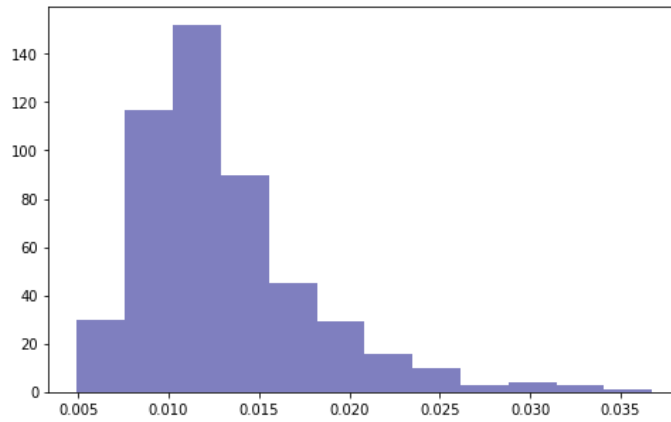


Figura 8: Histograma Desviación Estándar

■ Asimetría

Estadístico que caracteriza el grado de simetría de una distribución alrededor de su media. La forma en la que se calcula es la siguiente:

$$\alpha_2 = \frac{\frac{\sum (X - \bar{X})^3}{n}}{\left(\frac{\sum (X - \bar{X})^2}{n}\right)^{\frac{3}{2}}} = \frac{\mu_3}{\mu_2^{\frac{3}{2}}} \quad (1)$$

Nos indica la forma que tiene la gráfica de frecuencias de las ganancias diarias, este valor puede ser tanto positivo como negativo. Cuando es positivo significa que se tiene una concentración de datos a la izquierda de la media, por lo tanto que es más probable que haya valores por debajo de la media que por encima. Que sea negativo nos indica que hubo mayor concentración de datos a la derecha de la media, por que se tiene mayor probabilidad de que se tengan mejores rendimientos que los de la media

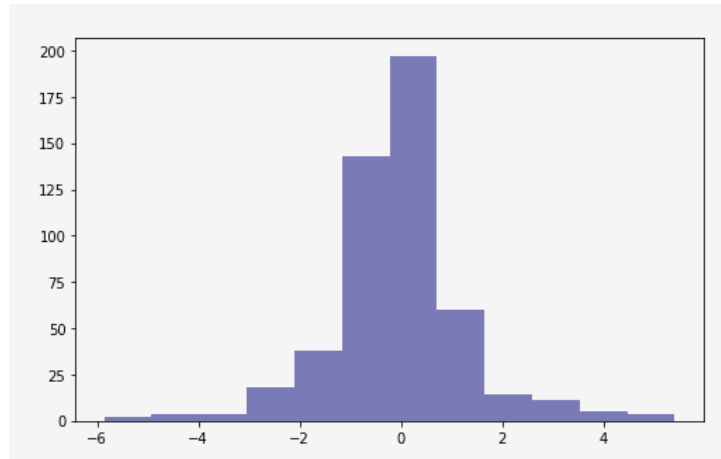


Figura 9: Histograma de la simetría de la variación porcentual

■ Kurtosis

Si la distribución es simétrica, lo siguiente a analizar es la forma de la distribución respecto a una Normal estándar. Mide la cantidad de valores extremos de nuestra distribución de ganancias diarias medias, que nuestra distribución tenga una kurtosis alta significa que se observaron muchos valores extremos, es decir que es una distribución de cola pesada, mientras que si se obtiene un valor negativo indica que en comparación con una normal, su pico central es mas bajo y mas ancho y sus colas son mas cortas y delgadas . El punto de referencia es la distribución normal, la cual posee una Kurtosis de 3.

$$\beta_2 = \frac{\frac{E(X-\mu)^4}{n}}{\frac{E((X-\mu)^2)^2}{n}} = \frac{\mu_4}{\sigma^2}$$

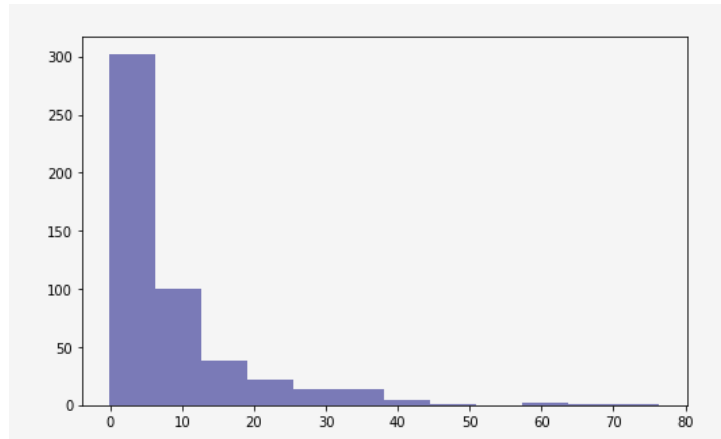


Figura 10: Histograma de la curtosis de la variación porcentual

Una vez que se han definido las variables y mencionado que representan de acuerdo a los datos utilizados es esencial conocer el grado en que las variables están correlacionadas, para analizar esto a continuación se presenta un mapa de calor el cual nos indica de acuerdo a una escala en nivel de correlación que hay entre las variables.

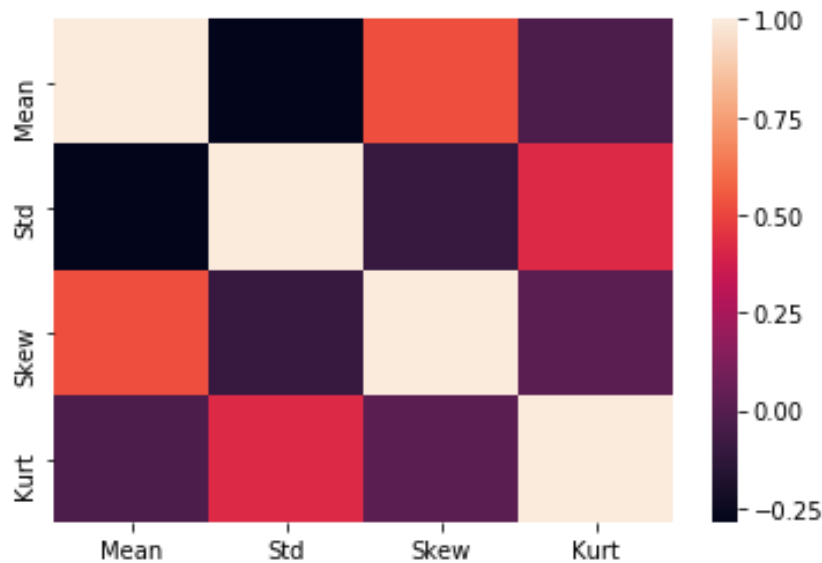


Figura 11: Correlacion de las variables

Las variables no presentan una correlación significativa, es decir, a valores altos de una no habrá un cambio significativo en la otra variable e igualmente con los valores bajos. En resumen se dispone de cuatro variables que están débilmente correlacionadas, ahora el inconveniente está en plasmar los datos ya que contamos con cuatro variables y al intentar

visualizar los datos seria casi imposible hacerlo por la dimensión, de manera que se hará uso de métodos estadísticos especializados para reducir dimensiones lo que facilitara visualizar los datos con los que contamos.

2.2. Compresión de los datos

Los métodos estadísticos mencionados anteriormente son Componentes Principales y Escalamiento Multidimensional, ambos métodos se clasifican como métodos de simplificación de dimensiones. El método de Componentes principales permite explicar de forma mas simple la estructura e interrelación de las variables originales. Por otra parte el Escalamiento Multidimensional tiene como objetivo modelar las proximidades entre los individuos de tal modo que pueda representarlos lo más exactamente posible en un espacio de baja dimensión, en este caso dos dimensiones.

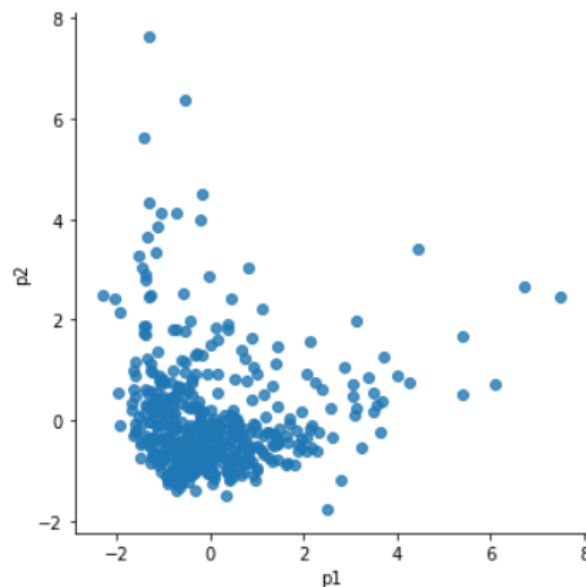


Figura 12: Gráfica con Componentes Principales

Ambos componentes principales representan al rededor del 75 por ciento de la varianza total, es decir, el gráfico que observamos nos presenta mas de la mitad de la información que se tiene, lo que indica que conserva gran parte de las características principales de los datos.

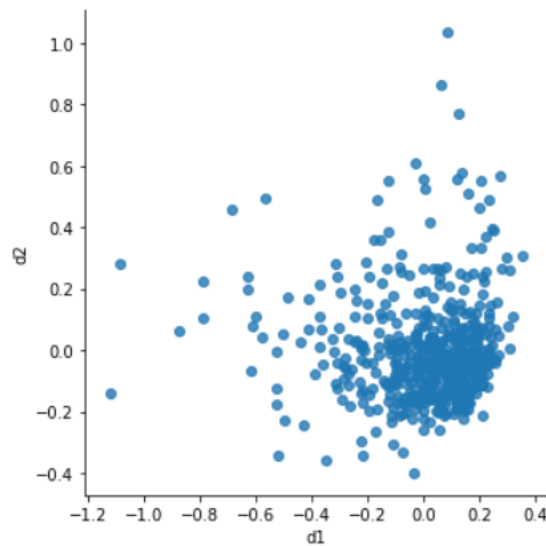


Figura 13: Gráfica con Escalamiento Multidimensional

2.3. Análisis de Clustering

Para aplicar el método de Clustering K-Means es necesario determinar el numero de grupos que deseamos tener, una mala elección de los mismos puede dar lugar a realizar agrupaciones de datos muy heterogéneos (pocos Clusters) o datos, que siendo muy similares unos a otros los agrupemos en Clusters diferentes (muchos Clusters).

Para resolver dicho problema existe el método del codo que nos ayudara a elegir un número apropiado de Clusters, dicho método utiliza los valores obtenidos de la inercia tras aplicar el K-means a un número diferente de Clusters, donde la inercia es la suma de las distancias al cuadrado de cada objeto del Cluster a su centroide. La siguiente gráfica corres-

ponde a los valores que tomo la inercia de acuerdo a los posibles clusters.

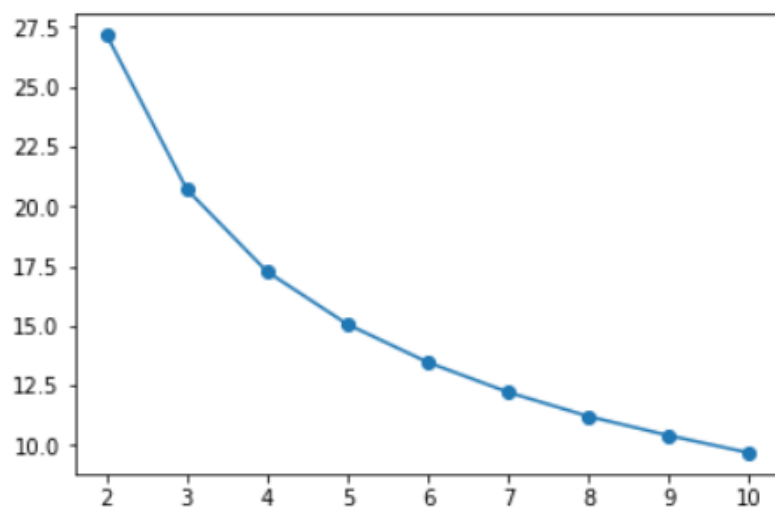


Figura 14: Inercia

En esta gráfica se aprecia un cambio brusco en la evolución de la inercia. El punto 4 es donde se observa ese cambio brusco el cual corresponde a el número óptimo (codo) de Clusters a seleccionar para nuestros datos.

Así, aplicando el método de clustering K-Means para cuatro clusters obtenemos las características de cada grupo de acuerdo a las variables con las cuales se trabajo.

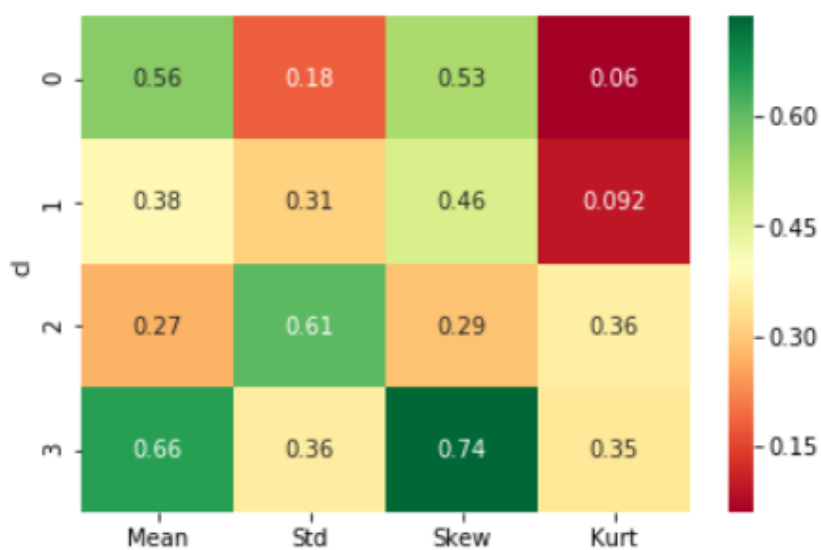


Figura 15: Características de los Clusters

Las características se interpretan de la siguiente manera:

- C_0 : Acciones con un rendimiento por encima del de una acción común y corriente que tiene un riesgo bajo en general y no tiene gran cantidad de casos extremos.
- C_1 : Acciones con rendimiento a la par de su riesgo, se puede decir que es el cluster menos extremo pues esta equilibrado en riesgo extremo, riesgo y rendimiento.
- C_2 : Acciones con rendimiento bajo y riesgo alto que son muy propensas a momentáneamente dar rendimientos por encima de la media, pero en largo plazo no. También es propensa a casos extremos.
- C_3 : Acciones con el rendimiento promedio más alto que de cualquier otro cluster con un riesgo razonable, por lo general obtendrás menor rendimiento (a corto plazo) que el que tuvo en promedio sin embargo aun así resulta ser alto.

2.4. Empresas según su cluster

En la Figura 16 son algunas empresas que conforman el correspondiente cluster.



Figura 16: Agrupamiento de empresas

Por otro lado, la siguiente imagen tenemos el comportamiento gráfico de los clusters por componentes principales Figura 18 y por multiescalamiento en la figura 17.

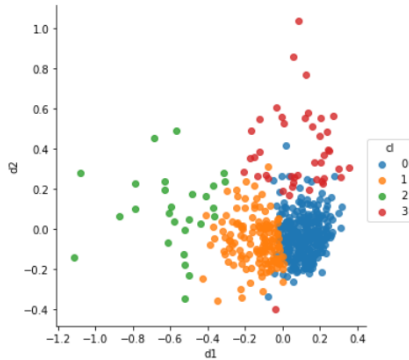


Figura 17: Clusters mediante MDS

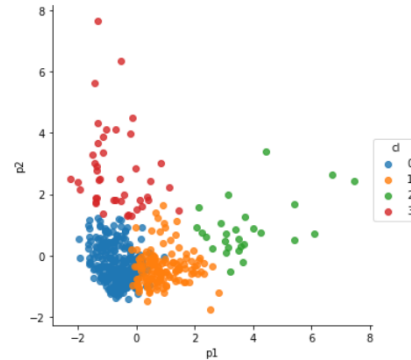


Figura 18: Clusters mediante CPA

2.5. Análisis temporal de los clusters

Se realizó un análisis de sensibilidad con respecto al tiempo pues cuando se calcularon estos solo estaba considerando el momento actual en el que encuentra la acción por lo que es muy fácil que los estadísticos de estas acciones cambien de manera muy rápida, lo que se hizo fue calcular los 4 estadísticos recorriendo a través de las fechas, teniendo de esta manera 4 estadísticos por cada día de cotización, de cada día se utilizaron los últimos 252 datos hasta la fecha de cotización.

Los resultados de este análisis fueron que efectivamente diariamente la acción cambia de posición en el espacio 4-dimensional, se observó que por lo general en cierto periodo de tiempo tienen saltos pequeños, que son los más comunes, y saltos grandes que por lo general en el periodo de tiempo que analizamos solo ocurrieron una o dos veces, este comportamiento se ejemplifica en las figuras 19 y 20 donde se muestra el comportamiento de las acciones de Amazon durante el año pasado, pasando de C_0 a C_3 .

Lo interesante de este análisis es poder saber cuanto tiempo falta para que una acción tenga un salto grande y cambie de cluster y a cual es más vulnerable que cambie, este tipo de preguntas pueden ser respondidas usando las predicciones hechas con la modelación supervisada.

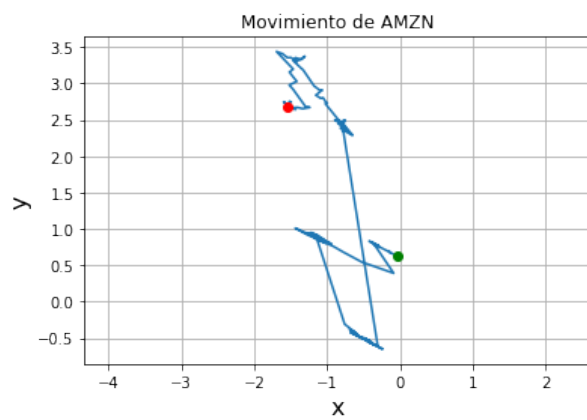


Figura 19: Acciones de Amazon en un año

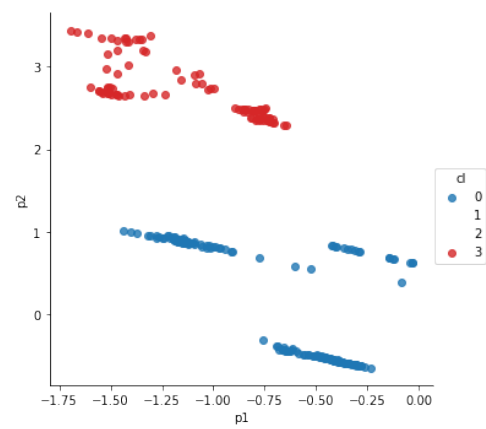


Figura 20: Clusters por los que paso la acción

3. Modelación Supervisada

En aprendizaje supervisado se basa en el descubrimiento de patrones a partir de ejemplos, la forma general consiste en lo siguiente: Es necesario contar con un numero de observaciones $x_1, x_2, x_3 \dots x_n$ las cuales están definidos a partir de un conjunto de atributos, es decir, propiedades $p_1, p_2, p_3 \dots p_m$ y para cada uno de ellos tenemos una clasificación observada $y_1, y_2, \dots y_n$, la tarea del aprendizaje supervisado es encontrar una función que nos permita inferir las clasificaciones de cada una de las observaciones para deducir la clasificación de nuevos datos a partir de sus propiedades. Obtener un mecanismo de este tipo seria de gran utilidad ya que se podría predecir comportamientos futuros a partir de sus propiedades. Existen distintos mecanismos para obtener una predicción fiable, los cuales se en listan a continuación.

3.1. Árboles de Decisión

Los arboles de decisión fundamentalmente se basa en proporcionar un conjunto de reglas que se aplican sobre las nuevas observaciones para determinar que clasificación es mas adecuada de acuerdo a sus propiedades.

El árbol de decisión están formados por una serie de nodos de decisión y nodos de respuesta. Un nodo de decisión se asocia a un atributo, tiene dos o mas ramas las cuales representan los valores que puede tomar el atributo asociado, por otra parte un nodo de repuesta se asocia a la clasificación que se pretende proporcionar, y devuelve la decisión del árbol con respecto a la entrada.

La generación de arboles de decisión consta de dos etapas: en la primer etapa se construye el árbol de decisión a partir de un conjunto de entrenamiento, mientras que en la segunda etapa cada objeto nuevo se clasifica por el árbol construido.

Los algoritmos de clasificación mas utilizados son ID3 y C4.5 .El algoritmo ID3 fue propuesto en el año 1979 por J. Ross Quinlan, tal algoritmo construye un árbol de arriba a abajo, de forma directa y se basa solo en los ejemplos iniciales proporcionados. Para ello se utiliza la ganancia de la información para seleccionar el atributo mas útil, es decir, decide la pregunta que mayor ganancia da en cada paso. El atributo mas informativo es el de menor entropia o de máxima ganancia de información . La entropia se define como una medida de incerti-

dumbre promedio, es la probabilidad de ocurrencia de cada uno de los eventos y se define como:

$$Entropia(S) = \sum -p_i \log_2 p_i \quad (2)$$

p_i es la proporción entre la probabilidad de que S pertenezca a la clase i . La ganancia de información es el complemento de la entropía. A mayor entropía menor información.

El algoritmo C4.5 es una extensión del algoritmo ID3 que permite manejar valores continuos y valores perdidos. Es computacionalmente rápido y las reglas de decisión son simples y legibles.

3.2. Redes Neuronales

Las redes neuronales es un modelo que está inspirado por la forma en la que funciona el cerebro humano. La unidad fundamental del cerebro humano es la neurona. La neurona recibe la información a través de las *dendritas*, la fuerza con la que cada impulso de cada neurona es variable, este impulso es transformado en el cuerpo de la neurona y la información viaja a la terminal sináptica de la neurona, y de esta manera es propagada la información a través de nuestro cerebro.

En nuestra red neuronal artificial nuestras neuronas reciben información por medio de x_1, x_2, \dots, x_n los cuales son multiplicados por ciertos pesos w_1, w_2, \dots, w_n para producir $z = \sum_{i=1}^n w_i x_i + b$, donde b es cierto sesgo, luego z es transformada con cierta función f y este valor es transmitido a las demás neuronas.

Para crear nuestra red neuronal es necesario un conjunto de neuronas que estén comunicadas de alguna manera, pero es fundamental que se tenga un sentido en el que se introduzcan datos por un lado y resulte un valor de nuestra regresión por el otro lado, en nuestro caso nos concentraremos en redes organizadas por capas, donde las neuronas no están conectadas entre ellas. Nuestras neuronas estarán conectadas mediante los pesos $W_{ij}^{(k)}$ donde k es el número de capa y i y j las neuronas que están conectando, este vector de pesos es denotado por θ .

El éxito que tenga nuestra red neural depende de un par de cosas, la función f y la forma en la que se determina θ , nuestro problema puede ser expresado en términos de matrices, la

matriz de $n \times m$ de los pesos de conexión entre cada neurona W , el vector de salida y entre cada capa y el vector de entrada x , entonces, $f(W^T x + b) = y$.

La forma en la que esta diseñada la red neuronal nos ayuda a que los parámetros se actualicen del vector de salida hacia el de entrada, se asigna cierta tasa de aprendizaje η y una función de pérdida $Loss$ entonces los pesos iniciales se van aproximando a los reales de la siguiente manera:

$$w_{ij} \leftarrow w_{ij} - \eta \left(\alpha \frac{\partial R(w_{ij})}{\partial w_{ij}} + \frac{\partial Loss}{\partial w_{ij}} \right)$$

En este sentido se pueden proponer distintas tasas de aprendizaje con las que pueda que se converja mas rápidamente a los valores de W .

Algunas neruonas muy comunes de usar son las siguientes [Buduma and Locascio \(2017\)](#):

- Sigmoide $f(z) = \frac{1}{1+e^{-z}}$
- Tangente hiperbólica $f(z) = \tanh(z)$
- ReLU $f(z) = \max(0, z)$
- Identidad $f(z) = z$

3.3. Máquinas Vector Soporte

La maquina vector soporte, es una técnica de modelación supervisada utilizada en problemas tanto de clasificación como de regresión. Consiste en construir un hiperplano(o un conjunto de hiperplanos) en una mayor dimensión, que incluso pudiese ser infinita, para separar patrones linealmente. Si consideramos la muestra de entrenamiento, como un conjunto de pares ordenados de la forma (x_i, d_i) , donde x_i representa al vector con los patrones de entrada y d_i la salida deseada.

Se asume que el patrón (clase/categoría) representado por el subconjunto $d_i = 1$ y que el patrón representado por el subconjunto $d_i = -1$ son “linealmente separables”, así, la ecuación que represente una superficie de decisión en la forma de un hiperplano que separe ambos patrones será:

$$w^T x + b = 0 \tag{3}$$

Donde x es un vector de entrada, w es un vector de pesos ajustables y b es un sesgo. Esto nos lleva a escribir que:

$$w^T x + b \geq 0 \quad (4)$$

para $d_i = 1$ y

$$w^T x + b < 0 \quad (5)$$

para $d_i = -1$

Dado un vector de pesos w y un sesgo b , la separación entre el hiperplano propuesto y el punto en los datos más cercano es llamada el margen de separación y se denota por la letra ρ . El objetivo de la máquina de soporte vectorial es encontrar el hiperplano particular tal que el margen de separación se maximice. Bajo esta condición, la superficie de decisión es llamada hiperplano óptimo. Si w^* y b^* son los valores óptimos del vector de pesos y el sesgo respectivamente, entonces, el hiperplano óptimo, representando una superficie lineal multidimensional de decisión en el espacio de entrada, está definida por:

$$w^* + b^* = 0 \quad (6)$$

Si definimos $g(x) = w^* x + b^*$ obtenemos una medida algebraica de la distancia desde x hasta el hiperplano óptimo, para ello, expresemos x como sigue:

$$x = x_p + r \frac{w^*}{\|w^*\|} \quad (7)$$

Donde x_p es la proyección ortogonal de x sobre el hiperplano óptimo y r es la distancia algebraica deseada. Donde r es positiva si x se encuentra en el lado positivo del hiperplano y negativa en caso de encontrarse del lado negativo. Dado que $g(x_p) = 0$, tenemos:

$$g(x) = w^T x + b^* = r \|w\| \rightarrow r = \frac{g(x)}{\|w^*\|} \quad (8)$$

El trabajo entonces, será encontrar los parámetros w^* y b^* para el hiperplano óptimo dado el conjunto de entrenamiento y dado que:

$$w^* x_i + b^* \geq 1 \quad (9)$$

para $d_i = 1$

$$w^* x_i + b^* \leq 1 \quad (10)$$

para $d_i = -1$.

Los puntos particulares (x_i, d_i) para los cuales las expresiones anteriores se convierten en igualdad son llamados vectores soporte y de ahí la derivación del nombre máquina de soporte vectorial. En términos conceptuales, tales vectores son los puntos que se encuentran más cerca de la superficie de decisión y son, en consecuencia, los más difíciles de clasificar.

3.4. K vecinos más cercanos

Este método consiste en suponer que la función de regresión es continua, pues considera que para valores cercanos con respecto a una métrica ρ definida en el espacio \mathbb{R}^n , los valores de la regresión serán de igual manera cercanos.

El algoritmo consiste en fijar una cantidad de vecinos más cercanos k y con respecto a esos k vecinos más cercanos se le asigna el valor $\hat{y} = \frac{1}{k} \sum_{i=1}^k y^{(i)}$, donde $y^{(i)}$ es el valor real de la regresión del i -ésimo vecino.

Se puede variar el valor k y la función de distancia ρ , algunas de las distancias más comunes son las siguientes:

- Euclidiana $\rho(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$
- Manhattan $\rho(x, y) = \sum_{i=1}^n |x_i - y_i|$
- Minkowsky $\rho(x, y) = [\sum_{i=1}^n (x_i - y_i)^p]^{1/p}$
- Chebyshev $\rho(x, y) = \max_i \{|x_i - y_i|\}$

3.5. Gradiente Estocástico

Es un enfoque simple pero muy eficiente. Este método define una función $E(W)$ que proporciona un error que comete la red en función del conjunto de pesos W . El objetivo será encontrar los pesos que corresponda al mínimo global de la función error, algunas veces basta con el mínimo local siempre y cuando sea lo suficientemente bueno.

Dado un conjunto de pesos $W(0)$ para el tiempo $t = 0$, se calcula la dirección de máxima variación del error. La dirección de máximo crecimiento de la función $E(W)$ en $W(0)$ está dado por el gradiente $\nabla E(W)$, se actualizan los pesos siguiendo el sentido contrario al indicado por $\nabla E(W)$, dicha dirección indica el sentido de máximo crecimiento, de modo que se produce un descenso hasta alcanzar un mínimo local.

$$W(t+1) = W(t) - \alpha \nabla E(W) \quad (11)$$

α indica el peso en cada iteración, si se toma un peso muy pequeño el proceso de entrenamiento resultaría muy lento, mientras que si se toma un peso muy grande se producen oscilaciones en torno al punto mínimo.

3.6. Ensamblés

En Machine Learning los ensambles son un conjunto de algoritmos usados para obtener un mejor modelo predictivo, su eficacia reside en la diversidad de algoritmos que lo constituyen. Por otro lado los algoritmos de ensamble con la cualidad Boosting tiene como objetivo minimizar el sesgo de predicción así como la varianza, dicho enfoque se basa en la hipótesis de que un conjunto de algoritmos de ML débiles (es decir, que tienen una estimación baja de predicción) pueden crear un algoritmo fuerte.

3.7. AdaBoost

Nombre acuñado de Adaptive Boost, es un algoritmo metaheurístico formulado por Freund y Schapire, es un algoritmo que por ser de la categoría de Boosting se constituye de una combinación lineal de algoritmos o modelos débiles h_t , (12), cuya estimación con respecto a los valores reales es baja, por su forma adaptativa hace que los modelos subsecuentes aprendan de las clasificaciones malas que hicieron los modelos anteriores.

$$f(x) = \sum_{t=1}^T a_t h_t(x) \quad (12)$$

3.8. RandomForest

Modelo de tipo ensamble donde el conjunto de modelos que lo constituyen son árboles de decisión, su ventaja reside en que minimiza el sobreajuste que comúnmente se encuentra como resultado de usar árboles de decisión individualmente, de esta forma RandomForest promedia todos estos árboles entrenándolos con distintas partes de la información destinada para entrenar el modelo con el objetivo de reducir la varianza.

La parte de entrenamiento se basa en la técnica de Tree Bagging donde dado un conjunto de entrenamiento X obtiene pequeñas muestras con remplazo a las que se le ajustan árboles un número B de iteraciones.

$$\hat{f} = \frac{1}{B} \sum_{b=0}^B f_b(x') \quad (13)$$

De (13) tenemos que dado una muestra X_b y un resultado de haberle aplicado un árbol clasificador o de regresión $f_b(X_b)$, iteradamente todos estos resultados se promedian. Más aún, x' representa aquellos datos que no se lograron observar.⁴

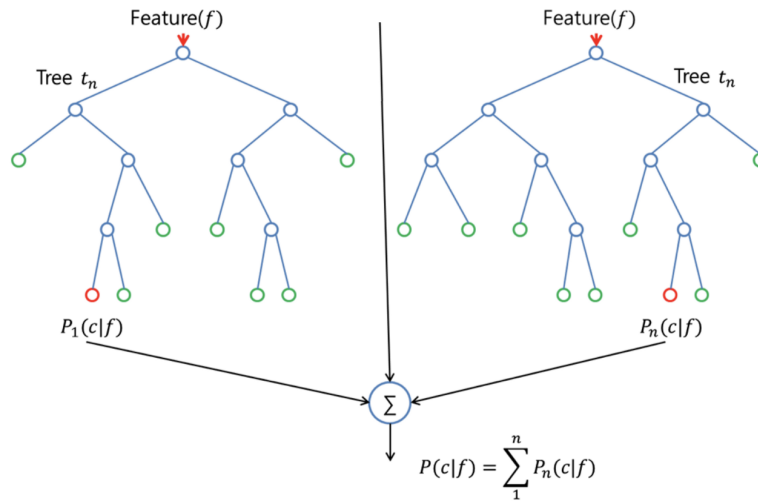


Figura 21: Ejemplo de un Bosque Aleatorio con sólo dos árboles

⁴Para más información visite [Towards Data Science](https://towardsdatascience.com/)

3.9. XGBoost

Nombre corto para Xtreme Gradient Boost algoritmo introducido por [Chen and Guestrin \(2016\)](#), es una mejora de los arboles de decisión, o más bien usa un conjunto de arboles aleatorios. El modelo asume que tenemos K arboles aleatorios entonces nuestra regresión de la siguiente manera

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i)$$

donde $f_k \in \mathcal{F}$ el conjunto de los arboles de decisión. Nuestra función objetivo para determinar las f_k 's es:

$$Obj = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k)$$

donde l mide que tan buena fue la aproximación Ω mide lo complejo que es el árbol de decisión, de esta manera al mismo tiempo que se optimiza que la regresión de mejores resultados también se busca que los modelos sean los más simples posibles, para así tener poco tiempo de computo.

La forma en la que se mide la complejidad de los arboles es:

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2$$

donde w_j es el peso de las hojas del arbol y T es el número de hojas, γ y λ son parámetros extra.

La forma en la que aproxima la regresión es con el entrenamiento aditivo, se empieza con una constante (en esta caso 0) y se van agregando las predicciones de los arboles.

$$\begin{aligned} y_i^{(0)} &= 0 \\ y_i^{(1)} &= y_i^{(0)} + f_1(x_i) \\ &\vdots \\ y_i^{(t)} &= \sum_{k=0}^t f_k(x_i) \end{aligned}$$

Pero la forma en la que se escogen las nuevas f_k 's que no es arbitraria, se escogen de manera que f_k minimice:

$$Obj^{(k)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(k-1)} + f_k(x_i)) + \Omega(f_k) + const$$

dada una función de pérdida l se requiere tener una aproximación de la función para poder realizar los cálculos en la computadora, para ello se requiere de la expansión de Taylor que esta dada de la siguiente manera:

$$Obj^{(k)} \approx \sum_{i=1}^n \left[l(y_i, \hat{y}_i^{(k-1)}) + g_i f_k(x_i) + \frac{1}{2} h_i f_k^2(x_i) \right] + \Omega(f_k) + const$$

donde $g_i = \partial_{\hat{y}_i^{(k-1)}} l(y_i, \hat{y}_i^{(k-1)})$ y $h_i = \partial_{\hat{y}_i^{(k-1)}}^2 l(y_i, \hat{y}_i^{(k-1)})$, de esta manera funciona el algoritmo, hasta llegar a tener K arboles.

4. Predicción de los precios usando regresión

Tener una herramienta que pueda pronosticar el precio de las acciones como hemos visto no es un tema nada nuevo, la pregunta más importante que se requiere responder acerca de una acción es ¿Cuanto valdrá esta acción el día de mañana?. Más haya de ofrecer un valor exacto de cuanto valdrá la acción se busca dar un poco de certidumbre a esta pregunta y poder tomar decisiones con nuestro pronóstico.

4.1. Metodología

Gracias a estos métodos de regresión multivariada es posible considerar una mayor cantidad de variables al momento de predecir el precio, en este sentido no tenemos ninguna limitación al querer agregar más información a nuestro modelo.

La forma en la que introducimos nuestras variables es muy similar al análisis de sensibilidad hecho con los clusters. Utilizamos el vector de precios diarios ajustados de las acciones de los últimos 500 días (Dos años aproximadamente) $\mathbf{P} = (P_1, \dots, P_{500})$, en este caso P_1 es el precio actual, con este vector se calcularon los primeros cuatro momentos con respecto a la media de su distribución en distintos subconjuntos del tiempo $T_s^i \subset \{1, 2, \dots, 500\} = T$, estos subconjuntos estan dados por $T_s^i = \{t | i \leq t \leq i + s, t \in T\}$, y los estadísticos por:

$$\hat{\mu}_{T_s^i} = \frac{1}{|T_s^i|} \sum_{t \in T_s^i} P_t \quad (14)$$

$$\hat{\sigma}_{T_s^i} = \sqrt{\frac{1}{|T_s^i|} \sum_{t \in T_s^i} (P_t - \hat{\mu}_{T_s^i})^2} \quad (15)$$

$$\hat{s}_{T_s^i} = \frac{\frac{1}{|T_s^i|} \sum_{t \in T_s^i} (P_t - \hat{\mu}_{T_s^i})^3}{\left[\frac{1}{|T_s^i| - 1} \sum_{t \in T_s^i} (P_t - \hat{\mu}_{T_s^i})^2 \right]^{3/2}} \quad (16)$$

$$\hat{\kappa}_{T_s^i} = \frac{\frac{1}{|T_s^i|} \sum_{t \in T_s^i} (P_t - \hat{\mu}_{T_s^i})^4}{\left(\frac{1}{|T_s^i|} \sum_{t \in T_s^i} (P_t - \hat{\mu}_{T_s^i})^2 \right)^2} - 3 \quad (17)$$

Tomando ventaja de que se puede agregar las variables que nos parezcan significativas utilizamos distintos periodos de tiempo para evaluar estos estadísticos, estos tiempos son $s = 2, 4, 7, 30, 60, 120, 250$, en el caso de la kurtosis κ solo se puede calcular cuando al menos se tienen 3 datos, por lo tanto tenemos un total de 27 variables que engloban tanto el compor-

tamiento de los últimos días como del pasado. Podemos reescribir nuestras variables como vectores de estos valores de s i.e. $\hat{\mu}^i = (\hat{\mu}_{T_2^i}, \hat{\mu}_{T_4^i}, \hat{\mu}_{T_7^i}, \hat{\mu}_{T_{30}^i}, \hat{\mu}_{T_{60}^i}, \hat{\mu}_{T_{120}^i}, \hat{\mu}_{T_{250}^i})$. De esta manera que nuestra tarea será estimar la función f tal que prediga el precio del día siguiente:

$$f(\hat{\mu}^i, \hat{\sigma}^i, \hat{s}^i, \hat{\kappa}^i) = P_{i+1}$$

Para lograr estimar f contamos con 250 ejemplos con los que entrenamos nuestros modelos, gracias a la API de Quandl ⁵ pudimos conseguir esta información de manera muy fácil con ayuda del lenguaje de programación python, para probar nuestros modelos se seleccionaron un conjunto pequeño de acciones, de las cuales heurísticamente pudiéramos preguntarnos ¿Cual será el valor de las acciones el día de mañana?, o que en general se tuviera un conocimiento previo acerca de como va la compañía de la que son las acciones, la lista de estas acciones se muestra en el cuadro 1

Ticker	Acción	Categoría
FB	Facebook	Tecnología
GOOGL	Google	Tecnología
AAPL	Apple	Tecnología
AMZN	Amazon	Tecnología
TSLA	Tesla	Automotriz
MSFT	Microsoft	Tecnología
AMD	Advanced Micro Devices	Tecnología
INTC	Intel Corporation	Tecnología
T	AT&T Inc	Telecomunicaciones
NFLX	Netflix Inc	Entretenimiento

Cuadro 1: Grupo de acciones seleccionadas

Cada una de estas acciones se intento ajustar alguno de los modelos que se explicaron en la sección anterior,. Con la ayuda de scikit-learn⁶ automatizamos el algoritmo para poder probar con las distintas variantes que tienen cada uno de los modelos con sus respectivos

⁵Quandl es una empresa que se dedica a vender información financiera quandl.com

⁶scikit-learn es una herramienta de machine learning en el lenguaje python, caracterizada por brindar los modelos de una manera muy sencilla y que el código pueda ser reutilizado <http://scikit-learn.org>

parámetros, para ello scikit-learn utiliza la métrica R^2 para poder comparar un modelo de otro.

La métrica R^2 (R cuadrada) que representa la cantidad de varianza que puede explicar nuestros valores que predcimos, el caso ideal sería cuando $R^2 = 1$, la forma en la que se calcula es la siguiente:

$$R^2 = 1 - \frac{\sum (y_i - f_i)^2}{\sum (y_i - \hat{y}_i)^2}$$

Donde y_i es el valor real y f_i son los valores que predijo nuestra regresión, de esta manera el algoritmo de scikit-learn diferencia entre modelos que tienen mayor valor predictivo.

4.2. Resultados

Con ayuda de scikit-learn y transformando nuestras variables escalándolas para ayudar a que nuestros algoritmos tengan mayor eficiencia, resumiendo el problema en una función del tipo $f : [0, 1]^{27} \rightarrow [0, 1]$. Calcular directamente la métrica R^2 no tiene tanto sentido pues se esta probando con los mismos datos que le dimos a los modelos para poder entrenar, por lo tanto es importante particionar nuestros datos para poder comprobar que nuestros modelos no sobre-ajusten, pues puede ocurrir que tenga un valor muy alto de R^2 pero con valores con los que nunca se entreno de valores fuera de la realidad.

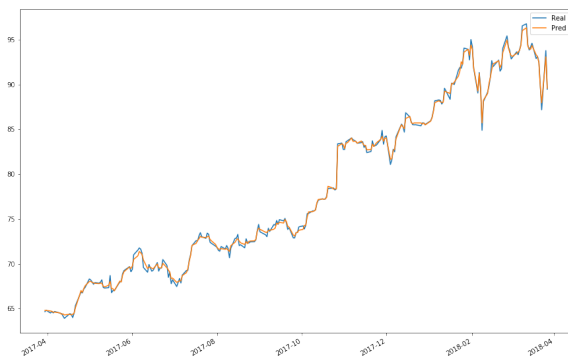
Para resolver este problema se entrena el modelo con el 70% de los datos tomados aleatoriamente y se calcula a R^2 tanto para los valores con los que se entreno como con el otro 30%. Para comparar a través de las 10 acciones que seleccionamos ajustamos el modelo a cada una de estas acciones y se promediaron los R^2 's tanto de entrenamiento (70%) como de validación (30%), los resultados se muestran en el cuadro 2.

Cuando la diferencia entre ambos R^2 's es demasiado alta, se puede decir que los modelos sobre-ajustaron, por lo general se tiene mejor resultado con el validate. XGBoost resulto el que mejor ajusto los datos y tiene un poder predictivo demasiado alto, que incluso sería muy complicado de mejorar agregando más información.

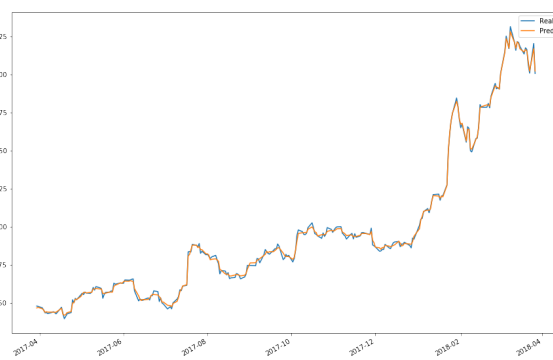
Otra manera de visualizar como resulto un modelo es graficar los valores reales con los valores que se predijeron, un ejemplo es la figura 23 donde están los valores reales de las acciones de Facebook comparados con los que nuestro modelo nos arrojo.

Modelo	Train R^2	Validate R^2	T. de computo
Red Neuronal	95.77 %	93.07 %	428 s
Maquina Vector Soporte	90.22 %	91.68 %	255 s
Arboles de decisi3n	93.01 %	89.19 %	2314 s
K vecinos	94.12 %	82.01 %	195 s
Gradiente Descendiente Estoc3stico	94.46 %	91.69 %	132 s
XGBoost	99.44 %	99.43 %	204 s
AdaBoost	97.16 %	93.59 %	199 s
RandomForest	98.87 %	92.27 %	191 s

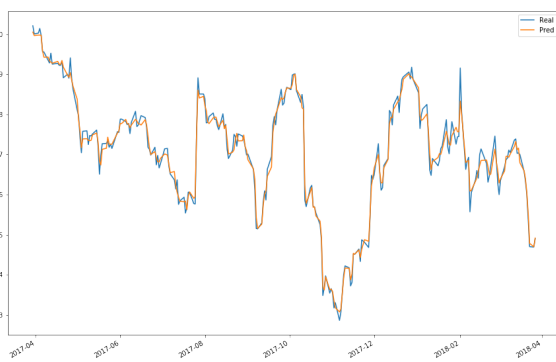
Cuadro 2: Tabla comparativa de resultados



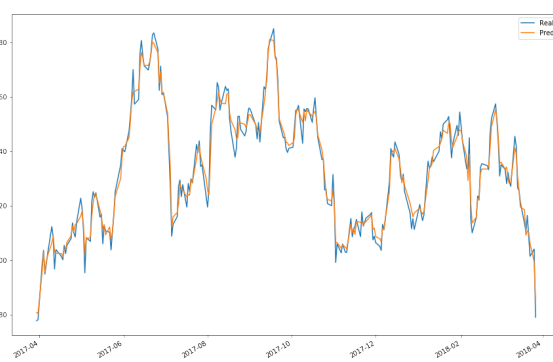
(a) Microsoft



(b) Netflix

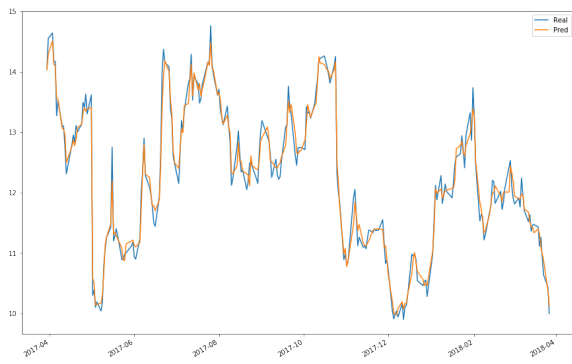


(c) AT&T

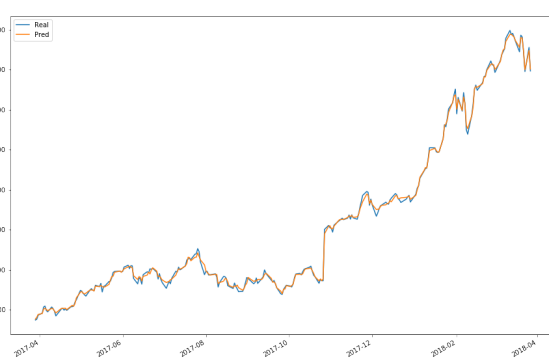


(d) Tesla

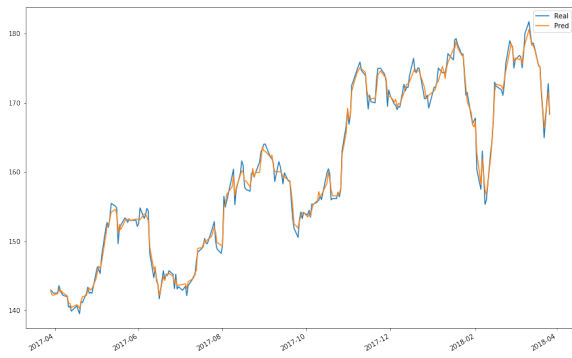
Figura 22: Gr3ficos de comparaci3n del modelo predictivo con el real



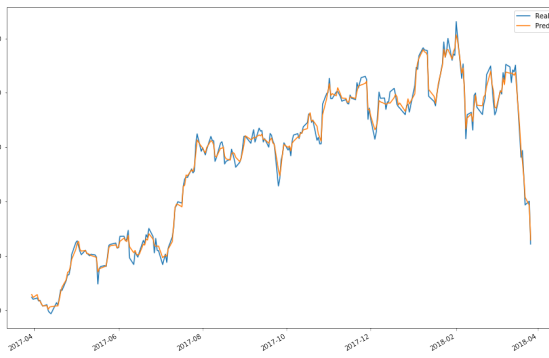
(a) AMD



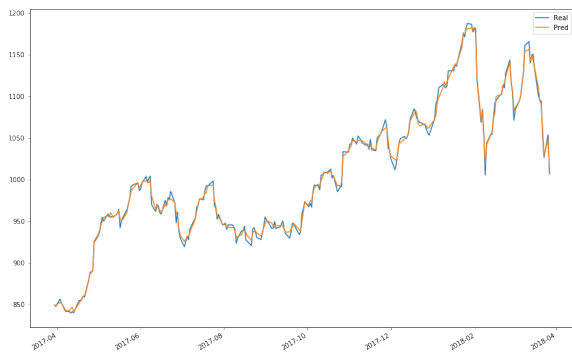
(b) Amazon



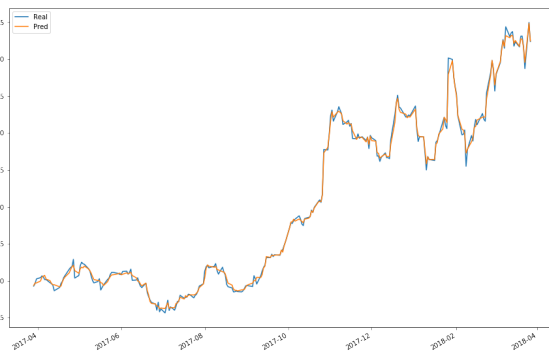
(c) Apple



(d) Facebook



(e) Google



(f) Intel Corp.

Figura 23: Gráficos de comparación del modelo predictivo con el real

Aplicación para los inversionistas

Cuando nos planteamos el objetivo final de poder predecir como se comportará el mercado accionario el día de mañana es poder tener certidumbre a la toma de decisiones, desde el punto de vista del inversionistas cada mañana antes de que abra el mercado accionario sabe que solo pueden ocurrir un par de cosa; que suba el precio de las acciones que posee o que bajen. Para el inversionista la información que existe acerca de este evento es difusa, tanto que le parece inverosímil tener una referencia clara de que ocurrirá por lo que su decisión sobre seguir poseyendo la acción no tiene un sustento claro.

Con nuestros modelos que hemos entrenado y probado se puede dar certidumbre a los inversionistas pues con la información conocida hasta el tiempo t fuimos capaces de dar un vistazo a lo que ocurrirá con P_{t+1} , el caso ideal para los inversionistas que tiene que tomar la decisión de hoy vender o seguir poseyendo es saber si $P_{t+1} > P_t$, lo que pueden hacer usando nuestro modelo es usar el valor que predecimos \hat{P}_{t+1} y poder evaluar la situación $\hat{P}_{t+1} > P_t$, comparar estos dos valores debe de ser mejor que lanzar una moneda y decidir si vender o no.

Ya hemos probado nuestros modelos usando la R^2 , sin embargo ese estadístico solo representa que tanto nuestro modelo entiende los datos, para poder probar que tan bueno es nuestro modelo de una manera realista debemos predecir los precios P_t usando estrictamente solo la información conocida hasta el tiempo $t - 1$. Para conocer el poder predictivo real se propone crear una función que nos cree modelos restringiendo la información y posteriormente comparar con lo que ocurrió en el mercado. Para esto usaremos el modelo obtenido con XGBoost pues según la métrica R^2 es el que mejor interpreto los datos. Usamos esta función por 30 días utilizando la serie de precios de AMZN (Amazon) y el resultado se muestra en la figura 24, podemos observar se tiene una mayor distancia entre las dos series que cuando usamos todos los datos para probar nuestro poder predictivo.

Utilizando nuestra idea de que lo más importante de nuestra predicción es si $\hat{P}_{t+1} > P_t$ nuestra predicción es mucho mejor que lanzar una moneda y usar el resultado como base de nuestra decisión.

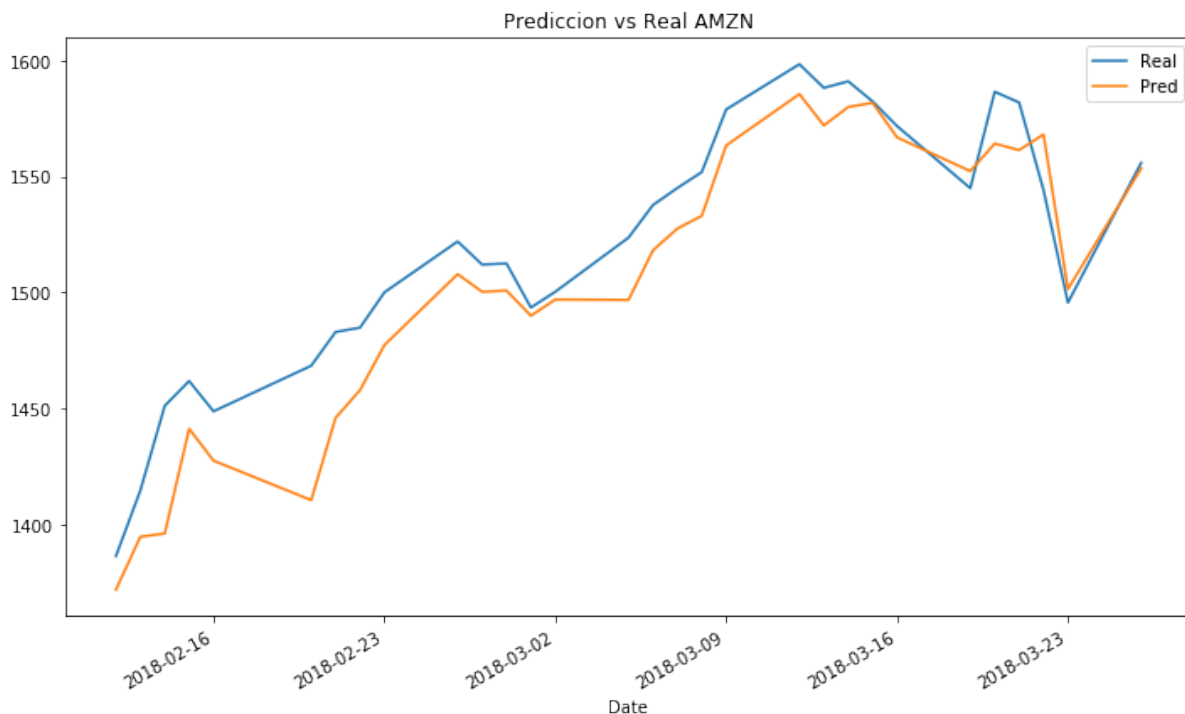


Figura 24: Comparación real de la predicción y el valor observado de la serie

Asumiendo que un cliente durante los 30 días hubiera tenido como criterio la predicción de nuestros modelos (en plural debido a que se entrena un nuevo modelo de manera diaria que solo sirve para predecir durante un día) hubiera mejora su oportunidad de obtener mejores rendimientos al final del periodo pues hubiera podido evitar los días en los que bajan los precios de las acciones que posee. En el caso de Amazon el resultado de haber seguido las recomendaciones se muestra en la figura 25. Esta gráfica muestra de una manera más clara los momentos en los que nuestro modelo cometió errores, sin embargo la serie al final del periodo supera la original debido a que tomo cierta ventajas los días que pronostico de manera correcta que la acción iba a bajar de precio, esos días la serie se mantiene constante y de esa manera evita la pérdida.

Desde un cierto de vista más critico los días que se vendió la acción y recuperamos nuestro dinero se pueden considerar como no producimos mas allá de que se haya evitado la pérdida pues no hubo ganancia, en ese caso se puede proponer que nuestro modelo busque en diferentes acciones alguna que se pronostique que va a aumentar su valor y se recomiende al usuario cambiar de las acciones de AMZN a las de la acción que encontró.

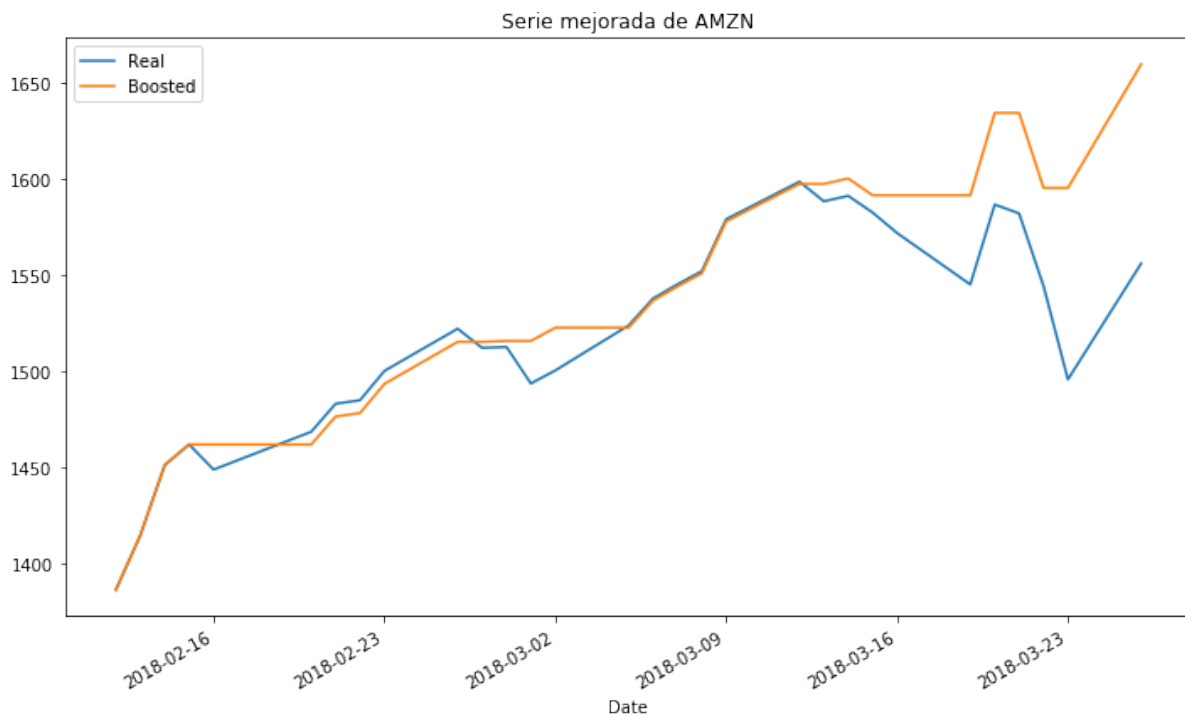


Figura 25: Serie de precios si el cliente hubiera seguido las recomendaciones durante 30 días

Más allá de que nuestro modelo se mejor que el azar prediciendo, se debe de tener en cuenta que no sería del todo correcto compararlo con los métodos que se utilizan para estructurar las inversiones hoy en día, o con la teoría de portafolios. Lo que podemos hacer es usar la metodología usual para estructurar portafolios de inversión y mejorar estos rendimientos con nuestro modelo, este esquema esta dedicado a los inversionistas agresivos que diariamente están queriendo sacar el mejor rendimiento posible.

Una propuesta para que los inversionistas puedan tener esta información diariamente es a través de una aplicación para dispositivos inteligentes donde puedan consultar tanto los clusters como las predicciones de sus acciones de interés. Esta información se actualizaría de manera diaria y les sería notificado los cambios existentes al cierre de los mercados.

Gracias a que las series de tiempo tienen un comportamiento fractalico se espera que cambiando la escala de tiempo se tengan resultados muy similares por lo que se podría esperar que se tenga un poder predictivo muy similar buscando el precio la siguiente semana o las próximas horas.

Conclusiones

En el mercado existen diversos tipos de inversionistas, pero todos tienen un objetivo en común y ese es generar rendimientos, nadie invierte su dinero buscando perderlo, lo que cambia, es que tan grandes quieren que sean esos rendimientos y el tiempo en obtenerlos. En ese sentido, el mercado es tan variado como los inversionistas, generar un modelo que englobe a todos sería prácticamente imposible, por lo que resulta más eficaz atacar el problema agrupando a los inversionistas o a las acciones.

De esta forma, ofrecemos ayudamos a los inversionistas a decidir sobre qué tipo de acciones invertir de acuerdo con sus características, esto gracias a la ayuda de las técnicas de “clustering” que nos permiten agrupar a las acciones, de acuerdo con su riesgo, rendimiento, tiempo en que tardan en generar esos rendimientos y entre otros factores.

Esto hace que los inversionistas se sientan más cómodos al saber que están invirtiendo en acciones de acuerdo con sus intereses y que para ello hay una teoría matemática detrás y no solo la experiencia de una persona. Aparte de que este agrupamiento permite a que las personas interesadas en invertir pero que no tienen un conocimiento amplio acerca de como se comportan todas las acciones a conocer el comportamiento de determinada acción, los clusters calculados pueden servir para identificar a grandes rasgos una acción de la que no se tiene conocimiento previo.

Además, la implementación de los modelos predictivos permite saber qué acciones podrían tener un mayor rendimiento, por lo que, el inversionista podría invertir en las acciones que se espera tengan mayor rendimiento, para así aumentar sus ganancias. Es una guía personalizada para los inversionistas en las que a su vez, gracias a la modelación supervisada les ayudamos a maximizar sus rendimientos.

Las técnicas de modelación multivariada nos abren una gran cantidad de oportunidades pues no tenemos limitaciones al momento de introducir factores que influyen en la predicción, sin embargo nuestro modelo predictivo sufre de ciertos problemas, el principal es que a pesar de que se considera la información reciente como del pasado, solo se considero el precio de las acciones en el pasado, se sabe que el precio de una acción depende no solo de como se ha comportado en el pasado si no también de factores externos que no pueden ser directamente explicados por nadie (en las noticias se intenta explicar porque fluctuó el

precio de una acción, pero en realidad es una especulación), fluctuaciones que solo ocurren eventualmente. Recordemos que en el mercado de acciones se tiende a especular demasiado, tanto sea la decisión de una máquina o de una persona, pues a fin de cuentas hay alguien pensando que el precio de la acción está baja (el que está comprando) y alguien más pensando que es muy alta (el que vende), mientras siga existiendo esta confrontación en la forma de pensar de los inversionistas, cualquier modelo tendrá esta ineficiencia.

Bibliografía

- Alexander, B. (1994). *Statistical Factor Analysis and Related Methods: Theory and Applications*. JOHN WILEY & SONS, INC.
- Bachelier, L. (1900). Théorie de la spéculation. *Annales scientifiques de l'École Normale Supérieure*, 17:21–86.
- Binitha, S., Sathya, S. S., et al. (2012). A survey of bio inspired optimization algorithms. *International Journal of Soft Computing and Engineering*, 2(2):137–151.
- Bouchaud, J.-P. (2008). Models of randomness and complexity, from turbulence to stock markets. *Leonardo*, 41(3):239–215.
- Buduma, N. and Locascio, N. (2017). *Fundamentals of deep learning: designing next-generation machine intelligence algorithms*. O'Reilly.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, pages 785–794, New York, NY, USA. ACM.
- Ding, S., Xu, X., and Nie, R. (2014). Extreme learning machine and its applications. *Neural Computing and Applications*, 25(3-4):549–556.
- Einstein, A. (1905). Über die von der molekularkinetischen theorie der wärme geforderte bewegung von in ruhenden flüssigkeiten suspendierten teilchen. *Annals of Physics*, 17:549.
- Fama, E. F. (1965). The behavior of stock-market prices. *The Journal of Business*, 38(1):34–105.
- Grimm, L. G. and Yarnold, P. R. (2000). *Reading and understanding MORE multivariate statistics*. American Psychological Association.
- Göçken, M. (2017). Hybridizing extreme learning machine and bio-inspired computing approaches for improved stock market forecasting. *2017 International Artificial Intelligence and Data Processing Symposium (IDAP), Artificial Intelligence and Data Processing Symposium (IDAP), 2017 International*, page 1.

- Hong, J.-Y. (2016). Study of stock index trend using tree-based ensemble classification.
- Huang, G.-B., Zhu, Q.-Y., and Siew, C.-K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3):489–501.
- Isogai, T. and Dam, H. C. (2017). Building classification trees on japanese stock groups partitioned by network clustering. 137(10):1387 – 1392.
- K. V. Mardia, J. T. Kent, J. M. B. (1979). *Multivariable Analysis*. Academic Press Limited.
- Mandelbrot, B. B. and Hudson, R. L. (2006). *The (mis)behavior of markets: a fractal view of risk, ruin, and reward*. Basic Books.
- Nair, B. B., Kumar, P. S., Sakthivel, N., and Vipin, U. (2017). Clustering stock price time series data to generate stock trading recommendations: An empirical study. *Expert Systems with Applications*, 70:20 – 36.
- Wu, S.-Q. (2017). A study of patent analysis for stock price prediction. *2017 4th International Conference on Information Science and Control Engineering (ICISCE), Information Science and Control Engineering (ICISCE), 2017 4th International Conference on, ICISCE*, page 115.
- Xing, F. Z., Cambria, E., and Welsch, R. E. (2018). Natural language based financial forecasting: a survey. *Artificial Intelligence Review*, 50(1):49–73.