



Implementación de regresión polinomial y clasificación K-Means para la predicción de precios en acciones

J. Antonio Rivera-García¹ and Miguel A. Lopez-Olvera¹

¹ ENES Juriquilla, Universidad Nacional Autónoma de México, Querétaro, Querétaro, México

I. INTRODUCCIÓN

1.1 Acciones y valores

En el mercado financiero, una acción (*share*) es un título emitido por una Sociedad Anónima o una Sociedad Comanditaria por Acciones que representa el valor de una de las fracciones iguales en que se divide su capital social. El dueño de las acciones o participaciones es conocido como accionista (*shareholder*, *stockholder*) de la sociedad. Una acción es una unidad indivisible de capital, que otorga derechos y obligaciones al propietario de esta, como la participación en las decisiones de la compañía o en los beneficios de la empresa, usualmente en función de la cantidad de capital que cada accionista ha invertido.

Se le denomina *stock* al conjunto de todas las acciones en las cuales está dividida una empresa. El dividendo es el capital obtenido producto de la propiedad de acciones. Para evaluar las acciones y, en términos generales, las empresas, se requieren métodos basados en la previsión de los beneficios y de los dividendos futuros. Las acciones pueden ser compradas o vendidas en las bolsas de valores (*Stock exchanges*).

1.2 Bolsa de valores

Una bolsa (*exchange*, *bourse*) es un mercado organizado en donde se puede comprar o vender valores, mercancía, divisas o derivados (futuros y opciones financieras).

La bolsa de valores (*Stock exchange*) es una organización pública o privada que brinda las facilidades necesarias para que sus miembros, atendiendo los mandatos de sus clientes, introduzcan órdenes y realicen negociaciones de compra y venta de valores, tales como acciones de sociedades o compañías anónimas, bonos públicos y privados, certificados, títulos de participación y una amplia variedad de instrumentos de inversión.

Las bolsas de valores fortalecen al mercado de capitales e impulsan el desarrollo económico y financiero en la mayoría de los países del mundo, donde existen en algunos casos

desde hace siglos, a partir de la creación de las primeras entidades de este tipo creadas en los primeros años del siglo XVII.

Los participantes de la bolsa son básicamente los demandantes de capital (empresas, organismos públicos o privados y otras entidades), los oferentes de capital (ahorradores, inversionistas) y los intermediarios

La mayoría de las bolsas son lugares en donde los interesados se reúnen para llevar a cabo las transacciones. Sin embargo, también existen bolsas conformadas por una red de computadoras, como el NASDAQ, en donde las operaciones son electrónicas.

Existen 16 bolsas con una capitalización de mercado superior a un billón de dólares (un trillón en inglés), que en 2015 representaron el 87 % de toda la capitalización mundial. Las principales bolsas pueden apreciarse en la tabla 1.

	Bolsa	Región	Cap.
1	New York Stock Exchange	EUA	22.923
2	NASDAQ	EUA	10.857
3	Japan Exchange Group	Japón	5.679
4	London Stock Exchange	Reino Unido	4.590
5	Shanghai Stock Exchange	China	4.026
6	Hong Kong Stock Exchange	Hong Kong	3.936
7	Euronext	Unión Europea	3.927
8	Toronto Stock Exchange	Canadá	3.256
9	Shenzhen Stock Exchange	China	2.504
10	Bombay Stock Exchange	India	2.056
11	National Stock Exchange	India	2.030
12	Deutsche Börse	Alemania	1.864
13	SIX Swiss Exchange	Suiza	1.523
14	Korea Exchange	Corea del Sur	1.463
15	Nasdaq Nordic Exchanges	Países Nórdicos	1.372
16	Australian Securities Exchange	Australia	1.328
24	Bolsa Mexicana de Valores	México	0.530

TABLA 1: CAPITALIZACIÓN EN BILLONES USD DE LAS PRINCIPALES BOLSAS DE VALORES EN 2019 (WORLD FEDERATION OF EXCHANGES)

1.3 Mercado de valores

Un mercado financiero es el conjunto de compradores y vendedores que comercian con valores y derivados con bajos costos de transacción (No confundir con bolsa, que se refiere de manera más precisa a las organizaciones en donde estas se efectúan y facilitan el comercio).

Los mercados financieros pueden clasificarse de diferentes maneras:

- *Stock markets* o mercado de valores (acciones).
- *Bond market* o mercado de bonos (títulos de deuda).
- *Commodity market* o mercado de materias primas.
- *Money market* o mercado monetario (activos financieros de corto plazo y bajo riesgo).
- *Derivatives market* o mercado de derivados (futuros, contratos u opciones financieras).
- *Futures market* o mercado de futuros.
- *Foreign exchange market* o mercado de divisas.
- *Cryptocurrency market* o mercado de criptomonedas.

El mercado de valores puede entonces definirse como el conjunto de personas que realizan el comercio con acciones o participaciones. Las acciones son clasificadas dependiendo de diversos criterios, como el país en donde la empresa está domiciliada. Por ejemplo, Nestlé o Novartis tienen su sede en Suiza, por lo que son consideradas parte del mercado bursátil suizo, aunque sus acciones puedan ser vendidas en otros países.

Comercializar en la bolsa de valores consiste la transferencia de acciones o garantías de un vendedor a un comprador, a cambio de dinero. Esto sugiere un acuerdo financiero entre ambas partes. Entre participantes del mercado de valores se encuentran pequeños inversores de acciones individuales hasta grandes inversores con personalidad física o jurídica, con posible sede en cualquier parte del mundo, y pueden incluir bancos, compañías de seguros, fondos de pensiones y de inversión. Las órdenes de compra y venta de estas entidades financieras pueden estar ejecutadas en su nombre o por un operador de bolsa.

En un intercambio, un comprador potencial ofrece a un determinado precio una acción, y un vendedor potencial pide a un precio determinado una acción. Cuando los precios de oferta y demanda coinciden, tiene lugar la venta.

Comprar o vender en el mercado implica aceptar los precios pedidos u ofrecidos para la acción deseada. El propósito de una bolsa de valores es facilitar el intercambio de títulos entre compradores y vendedores, creando así un mercado. Además, estos intercambios proporcionan información en tiempo real sobre el mercado e informa sobre la variación del precio de las acciones.

1.4 Índice bursátil

Un índice bursátil (*Stock market index*) es un índice que mide un mercado de valores o una parte de este. Esto resulta útil a los inversionistas para comparar los precios actuales de determinadas acciones con los precios anteriores para calcular el rendimiento de mercado. Los métodos de construcción de los índices son especificados, de manera que sea lo más transparente posible.

El índice estadounidense más antiguo existente hoy en día es el Dow Jones Industrial Average o simplemente conocido como Dow Jones. Este mismo fue creado por Charles Henry

Dow en conjunto con el Wall Street Journal para medir la actividad económica y financiera de los Estados Unidos de América a finales del siglo XIX. En un principio estaba compuesto por 12 compañías como General Electric, North American Company o American Tobacco Company. Actualmente está formado por 30 compañías.

Los índices bursátiles pueden clasificarse y segmentarse por el conjunto de acciones de cobertura del índice. La cobertura de un índice es el grupo subyacente de acciones, generalmente agrupadas junto con algún fundamento de su economía o la demanda de los inversores, que el índice está tratando de representar o rastrear. Un índice bursátil 'mundial' o 'global', como el MSCI World o el S&P Global 100, incluye acciones de todo el mundo y satisface la demanda de los inversores de un índice para acciones globales amplias. Los índices regionales que componen el índice MSCI World, como el MSCI Emerging Markets, incluyen acciones de países con un nivel similar de desarrollo económico, lo que satisface la demanda de los inversores de un índice para acciones de mercados emergentes que pueden enfrentar fundamentos económicos similares.

Los índices de cobertura global representan el desempeño del mercado de valores global. La serie FTSE Global Equity Index incluye más de 16.000 empresas.

Los índices de cobertura de países representan el desempeño del mercado de valores de una nación determinada y, por representación, reflejan el sentimiento de los inversores sobre el estado de su economía. **Los índices de mercado que se cotizan con mayor frecuencia son índices nacionales compuestos por acciones de grandes empresas que cotizan en las bolsas de valores más grandes de un país, como el índice S&P 500 en los Estados Unidos, el Nikkei 225 en Japón, el DAX en Alemania, el NIFTY 50 en India y el FTSE 100 en Reino Unido.**

Los índices de cobertura regional representan el desempeño del mercado de valores de una región geográfica determinada. Algunos ejemplos de estos índices son el índice FTSE Developed Europe y el índice FTSE Developed Asia Pacific.

Los índices de cobertura pueden basarse en el intercambio, como el NASDAQ-100 o grupos de intercambios, como el Euronext 100 u OMX Nordic 40. Otros rastrean el desempeño de sectores específicos del mercado. Algunos ejemplos incluyen el Wilshire US REIT Index, que rastrea más de 80 fideicomisos de inversión inmobiliaria y el NASDAQ Biotechnology Index, que consta de aproximadamente 200 empresas de la industria biotecnológica.

Otros índices pueden rastrear empresas de cierto tamaño, cierto tipo de administración o criterios más especializados, como los índices de base fundamental.

El S&P/BMV Índice de Precios y Cotizaciones (S&P/BMV IPC) es el principal índice bursátil de la Bolsa Mexicana de Valores, su muestra concentra las 35 principales empresas listadas en la Bolsa Mexicana de Valores. La recomposición del índice tiene una seria metodología que origina y expone la propia BMV, y que puede cambiar cada trimestre o año dependiendo de lo que la junta y dirección de la empresa BMV considere refleje mejor el valor de su muestra accionaria del mercado de valores en México.



1.5 Regresión lineal

En estadística, la regresión lineal o ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y y variables independientes X . Este modelo puede ser expresado como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_m X_m \quad (1)$$

En donde Y es la variable dependiente o variable de respuesta, mientras que X_1, X_2, \dots, X_m son las variables explicativas, independientes o regresoras.

Por su parte, los elementos $\beta_0, \beta_1, \beta_2, \dots, \beta_m$ son los parámetros del modelo, miden la influencia que las variables explicativas tienen sobre el regrediendo. El término β_0 es la intersección o término constante y las β_i ($i \geq 1$) son los parámetros respectivos a cada variable independiente.

Finalmente, m es el número de parámetros independientes a tener en cuenta en la regresión.

1.6 Algoritmos de agrupación en clústeres basados en distancia

Los representantes más conocidos de esta familia de métodos son los algoritmos **k-means** y **k-medoides**. Toman el parámetro de entrada k y dividen el espacio de datos en k grupos, de modo que la similitud entre los objetos de un grupo sea máxima. Además, minimizan la similitud entre objetos de diferentes grupos. El valor de similitud se calcula como la distancia desde el objeto hasta el centro del grupo. La principal diferencia entre estos métodos radica en la forma en que se define el centro del clúster. Con el algoritmo de **k-means**, la similitud es proporcional a la distancia al centro de masa del grupo. El centro de masa del grupo es el valor promedio de las coordenadas de los objetos del grupo en el espacio de datos.

El algoritmo de **k-means** se puede describir brevemente con los siguientes pasos:

1. Al principio, seleccionamos k objetos aleatorios
2. Definimos cada uno de ellos como un prototipo de grupo que representa el centro de masa del grupo.
3. Luego, los objetos restantes se adjuntan al grupo con mayor similitud.
4. Después de eso, se vuelve a calcular el centro de masa de cada grupo. Para cada partición obtenida, se calcula una función de evaluación particular, cuyos valores en cada paso forman una serie convergente
5. El proceso continúa hasta que la serie especificada converge a su valor límite.

En otras palabras, mover objetos de un grupo a otro termina cuando los grupos permanecen sin cambios. Minimizar la función de evaluación permite que los clústeres resultantes sean lo más compactos y separados posible. El método de **k-means** funciona bien cuando los clústeres son nubes compactas que están significativamente separadas entre sí. Es útil para procesar grandes cantidades de datos, pero no es aplicable para detectar grupos de formas no convexas o grupos con tamaños muy diferentes. Además, el método es susceptible

al ruido y a los puntos aislados, ya que incluso un pequeño número de tales puntos puede afectar significativamente el cálculo de la masa central del grupo.

II. PROBLEMA, OBJETIVOS E HIPÓTESIS

2.1 Planteamiento del problema

Muchas empresas que recurren a los mercados financieros son pequeñas empresas que, para continuar su desarrollo, requieren de capital de inversión. Por ejemplo, Orchid Biosciences pudo recaudar 48 millones de dólares en el 2000 al vender acciones[4].

La predicción del precio de las acciones es uno de las aplicaciones más importantes en Machine Learning aplicado a las finanzas y por supuesto, predecir el precio de una acción es el sueño de todos los inversores. Así mismo, son muchos los factores que influyen en la predicción, desde factores físicos hasta comportamientos psicológicos racionales o irracionales. Con una inmensidad de factores que pueden influir, el cambio en el precio de las acciones se vuelve muy volátil y por lo tanto, muy difícil de predecir.

La diversificación de una cartera de acciones es fundamental para mantener un equilibrio y es una técnica de inversión que consiste en componer la cartera con distintos tipos de activos con el objetivo principal de reducir los riesgos naturales de la inversión. Diversificar sería, de forma coloquial, no poner todos los huevos en una misma cesta para que, en caso de que se caiga, no se nos rompan todos.

Con el nacimiento y rápido crecimiento del análisis de datos y con ayuda del uso de la inteligencia artificial, diversos algoritmos de predicción de precios fundamentarían la toma de decisiones de inversión a inversionistas generales para la obtención de mayores rentabilidades con un menor grado de riesgo, además, se podría reducir aún más el riesgo de las inversiones si con el uso de algoritmos de clusterización se logra agrupar diversos tipos de acciones dependiendo de su volatilidad e ingresos anuales.

2.2 Objetivo general

Implementar los algoritmos de regresiones y clusterizado vistos en la materia de Inteligencia Artificial y encontrar sus aplicaciones para su uso en la vida diaria.

2.3 Objetivos particulares

Aplicar el algoritmo de regresión lineal utilizando polinomios de distintos grados con la finalidad de obtener una predicción sobre precio de las acciones, y así intentar reducir el riesgo al momento de realizar inversiones en el mercado de valores.

Usar el algoritmo de clusterizado conocido como K-Means para agrupar diferentes acciones de las 500 principales empresas que integran el Índice SP 500. para la diversificación de las carteras de inversión.

2.4 Hipótesis

Se logrará implementar el algoritmo de regresión lineal y se aplicará correctamente para la predicción del precio de las acciones de la empresa Tata Global Beverages, reduciendo

así el riesgo al momento de comprar acciones que pertenecen a la misma y que se encuentran en el mercado de valores.

Lograremos diversificar una cartera de inversión para las empresas que pertenecen al índice SP500 usando el algoritmo de K-Means y reduciremos el riesgo al momento de realizar inversiones en las 500 grandes empresas que poseen acciones que cotizan en las bolsas NYSE o NASDAQ

III. DESARROLLO E IMPLEMENTACIÓN

3.1 Regresiones

Tata Global Beverages es una de las diez principales compañías de alimentos y bebidas de India. La empresa cotiza en el *National Stock Exchange (NSE)* de dicho país.

A partir de un set de datos proporcionado por *Quandl*, a *NASDAQ platform* con información sobre el precio de las acciones de TGB durante un periodo de cinco años (08/10/2013 - 08/10/2018), se implementó un modelo de regresión lineal y polinomial para predecir los costos en el mercado de un conjunto de participaciones en función de otro grupo de días.

Date	Open	High	Low	Last	Close	Total	Trade	Quantity	Turnover (Lacs)	Day
2013-10-08	157.00	157.80	155.20	155.80	155.80	1720413.0	2688.94	1		
2013-10-09	155.70	158.20	154.15	155.30	155.55	2049580.0	3204.49	2		
2013-10-10	156.00	160.80	155.85	160.30	160.15	3124853.0	4978.80	3		
2013-10-11	161.15	163.45	159.00	159.80	160.05	1880046.0	3030.76	4		
2013-10-14	160.85	161.45	157.70	159.30	159.45	1281419.0	2039.09	5		
...
2018-10-01	234.55	234.60	221.05	230.30	230.90	1534749.0	3486.05	1231		
2018-10-03	230.00	237.50	225.75	226.45	227.60	1708590.0	3960.27	1232		
2018-10-04	223.50	227.80	216.15	217.25	218.20	1728786.0	3815.79	1233		
2018-10-05	217.00	218.60	205.90	210.25	209.20	3519515.0	7407.06	1234		
2018-10-08	208.00	222.25	206.85	216.00	215.15	4642146.0	10062.83	1235		

Fig. 1: Set de datos utilizado para la regresión.

Específicamente, para los datos en el eje X, se cambió el formato de fecha por días enumerados del 1 al 1235. No se cuentan con todos los días en el rango de cinco años debido a que las bolsas no suelen operar durante los fines de semana ni en días festivos.

La cotización de las acciones varía durante cada día, por lo que para el eje Y, se tomaron los precios de cierre. Es decir, los costos de estas al finalizar el día hábil.

En un comienzo se realizaron las regresiones de dos maneras: Primero se implementó el algoritmo de forma manual, elaborando los métodos de selección de datos de prueba, regresión lineal, predicción y cálculo del error cuadrático medio. Después se realizó un algoritmo similar, pero empleando las herramientas disponibles en la biblioteca de *SciKitLearn*.

Los resultados en el cálculo de la ordenada al origen, de los coeficientes y del error cuadrático medio en ambos códigos fueron exactamente los mismos, por lo tanto, por simplicidad, en este proyecto se empleó únicamente el algoritmo que utiliza la biblioteca *SciKitLearn*, el cual se encuentra anexo al presente documento.

Se realizó la regresión lineal, además de regresiones polinomiales de distintos grados. Al final, se trabajó con 9 modelos correspondientes a regresiones de orden 1 al 9, con la finalidad de encontrar el modelo que mejor se ajuste a los datos.

Además, se iteró 50 veces para cada modelo, con el objetivo de encontrar el que presente un menor error. Se muestran

también los valores de la ordenada al origen y de los coeficientes de dicha regresión, así como la gráfica de estos.

3.2 Clusterización K-Means

Primero, se realizaron algunos cálculos para la obtención de los retornos anuales y la volatilidad de cada una de las acciones integradas en el SP500, las fórmulas desarrolladas fueron las siguientes:

1. Retorno Anual

$$\sum_{i=1}^N \%ch_i = \bar{x}_{\%ch} N = \frac{\sum_{i=1}^N \%ch_i}{N} N \quad (2)$$

Donde N=252 (Días hábiles en el mercado de valores-1)

2. Volatilidad

$$\sqrt{\sum_{i=1}^N (\%ch_i - \bar{x}_{\%ch})^2} = \sigma_{\%ch} \sqrt{N} \quad (3)$$

$$\sigma_{\%ch} \sqrt{N} = \sqrt{\frac{\sum_{i=1}^N (\%ch_i - \bar{x}_{\%ch})^2}{N}} \sqrt{N} \quad (4)$$

Una vez obtenidos los valores para el retorno anual y la volatilidad, se graficó la curvatura de Elbow para conocer la cantidad óptima de clústeres para el set de datos. Seguido, se implementó el algoritmo de K-Means desde cero siguiendo los pasos descritos en la sección 1.6 de 'Algoritmos de agrupación en clústeres basados en distancias' se obtuvieron los centroides actualizados por medio del mismo algoritmo, iterando hasta llegar a un punto en el cual el valor absoluto de la resta de la distancia de la suma de los centroides viejos, menos los nuevos, fuera de cero. Así, se obtuvieron los clústeres y se graficaron para su posterior comparación.

Se hizo uso del paquete de Scikit-Learn para comparar la efectividad del algoritmo desarrollado y de igual manera se graficó y se obtuvieron los 5 centroides respectivos.

Los datos usados en este proyecto fueron extraídos de la web y se observa un resumen de los mismos en la Figura 2.

	Date	ABT	ABBV	...	ZBH	ZION	ZTS
0	2020-03-31	77.702164	72.322906	...	100.376968	25.804100	117.007439
1	2020-04-01	75.397980	69.693497	...	89.006607	24.415541	111.270905
2	2020-04-02	78.224052	71.316696	...	92.065186	24.483038	115.357056
3	2020-04-03	78.233894	69.646042	...	91.370049	23.798401	113.945297
4	2020-04-06	81.463692	71.886253	...	97.258812	25.968027	126.054657
...
248	2021-03-25	119.050003	103.879997	...	157.639999	54.740002	152.880005
249	2021-03-26	122.070000	105.980003	...	161.320007	55.849998	156.149994
250	2021-03-29	122.230003	106.730003	...	160.210007	53.889999	158.389999
251	2021-03-30	119.750000	106.790001	...	161.220001	55.910000	157.039993
252	2021-03-31	119.839996	108.220001	...	160.080002	54.959999	157.479996

Fig. 2: Datos extraídos uno a uno de Yahoo finanzas para realizar la clusterización con el algoritmo de K-Means

IV. RESULTADOS Y DISCUSIÓN

4.1 Regresiones

Las siguientes gráficas corresponden a los diferentes modelos aplicados para las regresiones:

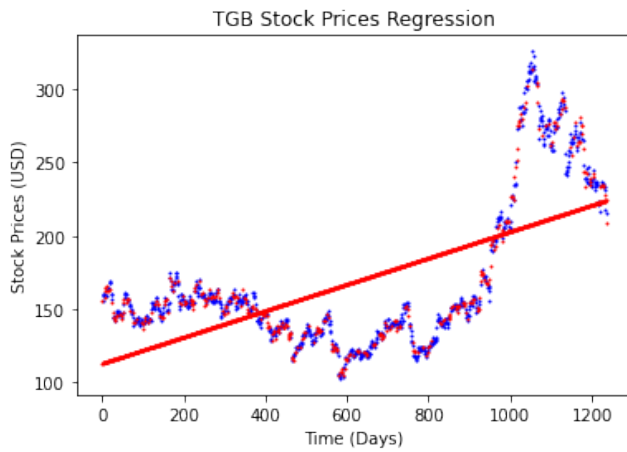


Fig. 3: Regresión de grado 1. Error = 1455.5656875390089

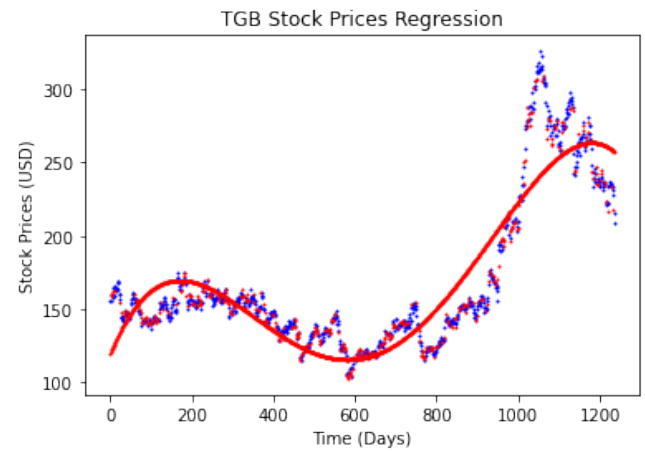


Fig. 6: Regresión de grado 4. Error = 378.25469605858785

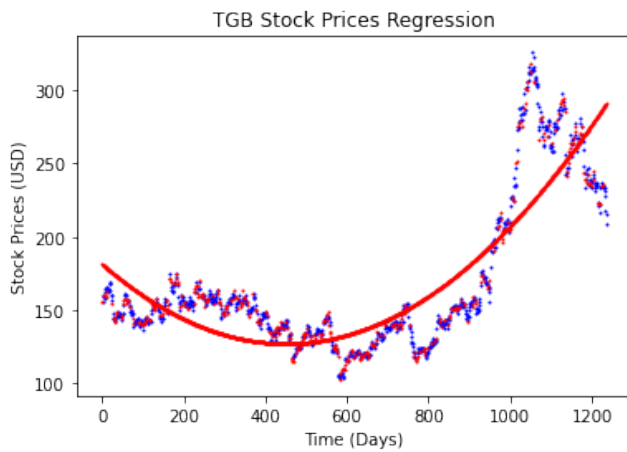


Fig. 4: Regresión de grado 2. Error = 627.8711419183087

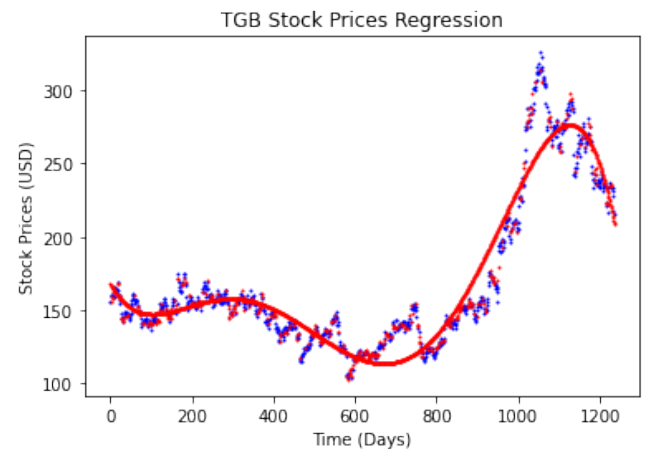


Fig. 7: Regresión de grado 5. Error = 198.60076338755556

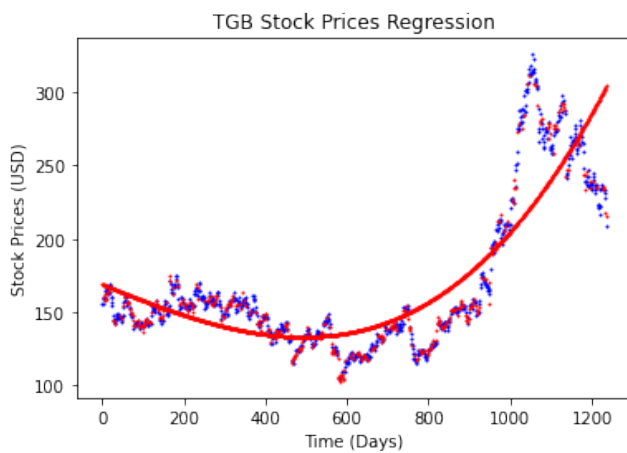


Fig. 5: Regresión de grado 3. Error = 581.7023054000116

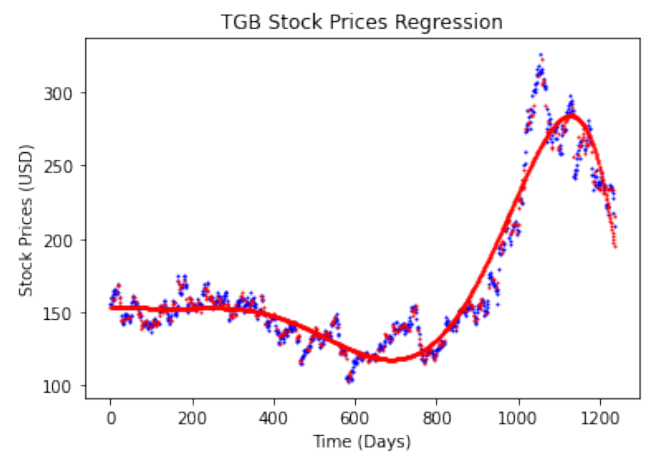


Fig. 8: Regresión de grado 6. Error = 175.63996162649022

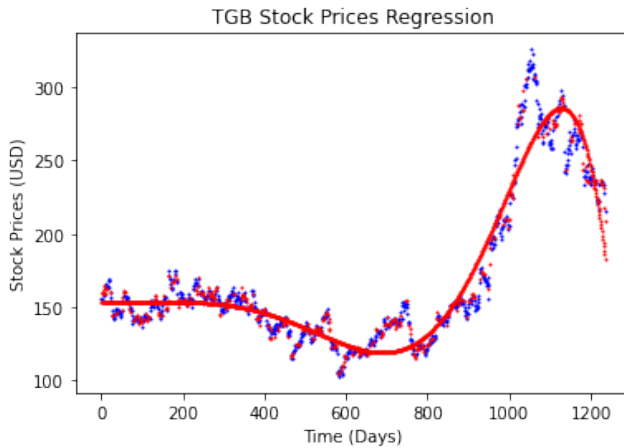


Fig. 9: Regresión de grado 7. Error = 169.7309276828444

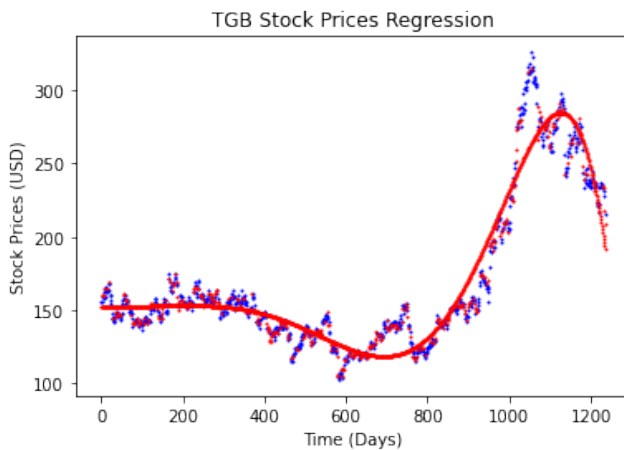


Fig. 10: Regresión de grado 8. Error = 172.89460110109536

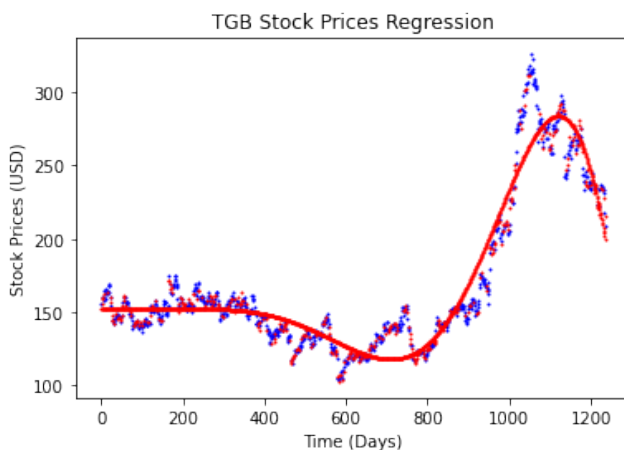


Fig. 11: Regresión de grado 9. Error = 174.5877391580006

Como puede apreciarse en las gráficas, y como podría esperarse, el modelo de la regresión lineal es el que presenta el error más grande, mientras que las regresiones cuadrática y cúbica reducen el error más de la mitad.

Por su parte, las regresiones de grado 6, 7, 8 y 9 presentan errores bastante similares. Incluso al observar las gráficas, los cambios entre cada una de estas son casi imperceptibles. A pesar de que, para no extender de más el presente documento, el modelo de mayor complejidad fue el de grado 9, parece que a partir de estos últimos, el error podría empezar a aumentar lentamente e incluso presentar *overfitting* en un grado no muy alejado y volver a mostrar errores incluso mayores a la regresión de primer grado en algún punto.

Aunque podría parecer que el modelo polinomial de grado 7 fue el más efectivo, sobra decir que debido a la naturaleza del algoritmo, los errores a obtener suelen oscilar dentro de un rango de posibles resultados, por lo que no podría decirse con certeza que, efectivamente, este sea el mejor modelo para predecir datos en cuestión. Cabe mencionar que el error obtenido en cada uno de los modelos fue resultado de una sola ejecución del código. Muy probablemente, si se realizaran más iteraciones con los modelos de 6, 7, 8 y 9, podrían alcanzarse errores aún más pequeños. De hecho, después de ejecutarse los modelos y de calcular los errores aquí presentes, se iteró más veces el modelo con polinomio de grado 9, y en un caso se obtuvo un error de 164. Resultaría un ejercicio interesante el poder conocer la campana de posibles errores utilizando métodos de probabilidad y estadística, pero eso va más allá del objetivo de este proyecto.

4.2 Clusterización K-Means

A continuación, en la Figura 12, se observa la curvatura de Elbow implementada para el set de datos utilizado en este proyecto, además de los resultados de la clusterización con el uso del paquete de Scikit-Learn.

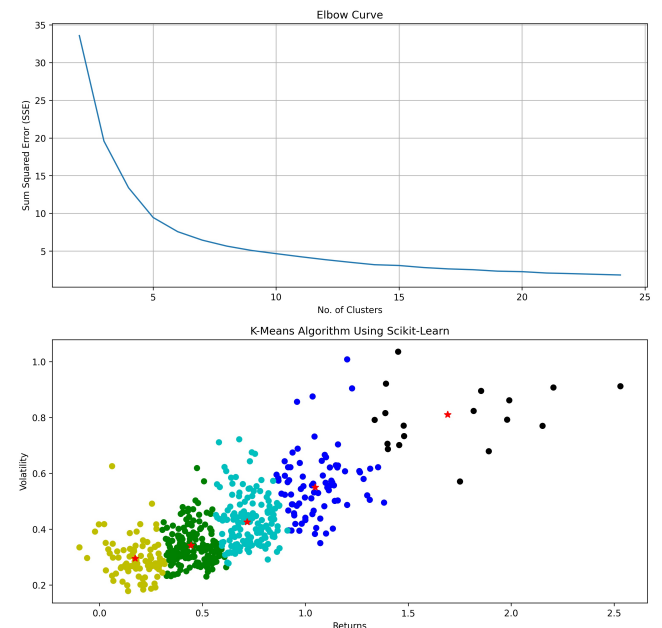


Fig. 12: Curvatura de Elbow que nos muestra un óptimo número de clústeres entre 4 y 6, se usó un total de 5 clústeres para este caso. Además, podemos ver el gráfico de los clústeres generados con el paquete de Scikit-Learn, usando el algoritmo de K-Means



Casi para terminar, el algoritmo de K-Means implementado desde cero se muestra en el gráfico de la Figura 13.

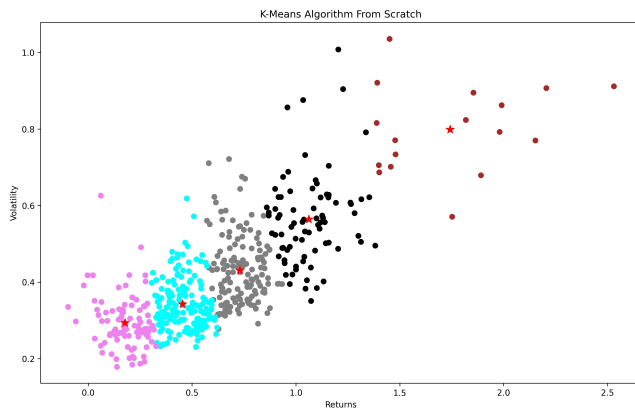


Fig. 13: Clusterización usando el algoritmo de K-Means implementado desde cero SIN el uso de Scikit-Learn

Por último, se presentan los centroides obtenidos con los algoritmos de K-Means, con y sin el uso de Scikit-Learn.

Centroids:	Cluster
[1.69231624 0.8096697]	1 0.177086 0.293660
[1.04856041 0.54953601]	2 0.730133 0.430273
[0.44481367 0.34124635]	3 1.741984 0.799054
[0.1739266 0.29452491]	4 0.453395 0.342337
[0.71815368 0.4252093]	5 1.061610 0.563941

Fig. 14: Centroides obtenidos con cada una de las implementaciones, los centroides de la izquierda corresponden a la implementación con Scikit-Learn

ciones, estos algoritmos tienen una gran variedad de aplicaciones en prácticamente cualquier área del conocimiento.

Si bien las regresiones lineales y los métodos de clasificación representan las herramientas más básicas en el amplio mundo de la inteligencia artificial, también son la base para seguir aprendiendo algoritmos cada vez más complejos y con aplicaciones increíblemente útiles, que podrían ayudarnos a solucionar bastantes problemas en el futuro, así como a tener una gran variedad de herramientas disponibles listas para ser empleadas en la vida profesional.

REFERENCIAS

- [1] Géron, Aurélien. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (1-276). Estados Unidos: O'Reilly.
- [2] Steven Jorge Pedrosa (11 de octubre, 2015). Diversificación de carteras. Economipedia.com
- [3] Temas y conceptos aprendidos durante la clase de Temas Selectos de Computación.
- [4] Yahoo. (2021). Yahoo Finanzas - Mercado de valores en vivo, cotizaciones, negocios y noticias financieras. <https://es-us.finanzas.yahoo.com/>
- [5] Kolodiazhnyi, K. (2020). Hands-On Machine Learning with C ++: Build, train, and deploy end-to-end machine learning and deep learning pipelines.

V. CONCLUSIONES

Se puede decir que con la implementación de ambos algoritmos se obtuvieron resultados favorables. Los códigos empleados funcionaron de manera esperada y permitieron cumplir el objetivo que se tenía en mente en un principio.

Por un lado, gracias a las regresiones polinomiales se pudo desarrollar una ecuación que logra aproximar bastante bien la fluctuación de las acciones de la empresa Tata Global Beverages. Como se mencionó previamente, incluso quedó una cierta incertidumbre sobre si se podría conseguir que los polinomios pudieran reducir el error aún más.

Por otro lado, se logró desarrollar un algoritmo capaz de diversificar una cartera de inversiones con el uso del algoritmo de K-Means, tanto con el paquete de Scikit-Learn, como sin el mismo. La eficiencia de ambos algoritmos para realizar el clusterizado es favorable, sin embargo, se realizó una medición del tiempo de ejecución de ambos y se encontró que el algoritmo con el uso de Scikit-Learn es 211.22 veces más rápido que el implementado sin el paquete.

La especulación financiera es una tarea bastante compleja que involucra una infinidad de posibles variables y la predicción de los precios futuros de las acciones no puede reducirse simplemente al uso de estos algoritmos, que funcionan basados en el comportamiento previo de las fluctuaciones.

Sin embargo, la implementación exitosa de ambos códigos, resulta en dos herramientas bastante útiles, y aunque no puedan utilizarse como un medio de predicción fiable de ac-