

elaboración de un portafolio de acciones conformado por las empresas más representativas del mercado. La ponderación de las acciones está en función de su variación con respecto al mercado. Sin embargo, tal indicador da un panorama general del comportamiento del mercado, sin tomar en cuenta que los diversos sectores de las compañías incluidas en el IPyC no se comportan de la misma manera. i.e. si un inversionista busca invertir en una acción de Santander, consultar los movimientos recientes del sector financiero es una práctica que tiene sentido. En contraste, el inversionista podría consultar los movimientos del sector energético, el cual tendrá poca o nula relación con el precio de la acción deseada. Desde este punto de vista, hay tres factores que se tomarán en cuenta:

- i.El desempeño propio de la emisora durante un periodo de tiempo determinado (diario, quincenalmente, etc.).
- ii.El movimiento del mercado en general (IPyC).
- iii.El comportamiento del sector al que pertenece la emisora.

Como se mencionó anteriormente, han sido publicadas diversas herramientas y metodologías para analizar el precio futuro de una acción. Por un lado, se tienen los métodos clásicos como el CAPM (Capital Asset Pricing) o el APT (Arbitrage Pricing Theory). Ambos utilizan metodologías de valuación de activos, a través de conjuntos eficientes y análisis de sensibilidad, calculan el rendimiento esperado de un activo (en este caso una acción). Posteriormente, se utiliza dicho rendimiento para evaluar los flujos futuros asociados a la acción [9]. El interés de esta investigación es estimar la probabilidad de alza o baja de una acción, condicionada a la variación de precios de acciones que cotizan en el sector financiero. Cabe destacar que BMV clasifica a las diferentes emisoras en los siguientes sectores:

- i. Energía
- ii. Materiales
- iii. Industrial
- iv. Servicios y bienes de consumo no básico
- v. Productos de consumo frecuente
- vi. Salud
- vii. Servicios Financieros
- viii. Tecnología de la información
- ix. Servicios de telecomunicaciones
- x. Servicios públicos

Dado que la principal fuente de financiamiento para una entidad bancaria es la captación de capital, además de que según cifras de el diario El Economista, para 2018, cerca del 68% de los mexicanos poseían una cuenta bancaria, es natural inferir que el sector financiero de la BMV es de los más importantes del país [15]. Por lo tanto, para fines ilustrativos de la investigación, se llevará a cabo el análisis sobre tal sector, particularmente para acciones de grupo Santander.

Esta investigación tiene como objetivo estimar la probabilidad de alza o baja del precio de acciones de Santander, condicionado a movimientos de precios de emisoras del mismo sector de manera conjunta. A través de la regla de Bayes aplicada en el algoritmo ingenuo de Bayes, se buscará la combinación de eventos que maximizan la probabilidad de alza o baja de una acción, dado el movimiento del mercado. Esta

probabilidad servirá de apoyo al público inversionista o académico que posea cierta posición sobre una acción, i.e. si se tiene una posición corta, se espera que la acción se deprecie, mientras que, con una posición larga, la expectativa es que la acción se aprecie. En otras palabras, los resultados de la investigación serán de apoyo para saber si vender o comprar, dado el movimiento del mercado.

El enfoque proporcionado en esta investigación utiliza una aplicación de un algoritmo bayesiano que pretende predecir si una acción subirá o bajará de precio, en términos de probabilidades condicionales, lo cual es un tratamiento apropiado, puesto que se trabaja con variables dicotómicas.

2. CONCEPTOS BÁSICOS

A lo largo de la investigación, se trabajará con una base de datos de precios de cotización diario de empresas del sector financiero. Particularmente, se estimará la probabilidad de alza o baja de una acción de Santander, condicionada a los cambios de otros bancos que cotizan en la BMV. Por lo tanto, se hará uso de términos financieros y probabilísticos [10].

Entre los diferentes algoritmos que existe en la minería de datos, los algoritmos probabilísticos resultan ser una herramienta sumamente potente ya que se basa en probabilidad y estadística. Puesto que el algoritmo ingenuo de Bayes utiliza como base la probabilidad condicional, se definirán conceptos como eventos independientes, regla de Bayes, probabilidad conjunta, así como los estimadores de las probabilidades condicionales (según el algoritmo).

Acción: Un título representativo de una parte alícuota del capital social de una sociedad anónima, siempre otorga a su titular derechos patrimoniales, y dependiendo del tipo de acción también otorga derechos corporativos. Las acciones son emitidas en masa o en serie y cuentan con títulos llamados cupones.

Mercado: Conjunto de oferentes y demandantes de activos financieros a través de instituciones e intermediarios financieros.

Precio de apertura: Precio de cotización con el cual inicia un día de emisión.

Precio de cierre: Precio de cotización con el cual finaliza un día de emisión.

Clasificación: Problema de identificar a cuál categoría (subpoblación) pertenece una nueva observación, basado en un conocimiento previo de los miembros de la población cuya clasificación es conocida

Probabilidad conjunta: Sean A, B eventos de un espacio muestral Ω , la probabilidad de que ocurran A, B , de manera simultánea, $P(B, A)$, está dada por:

$$P(B, A) = P(B \cap A).$$

Eventos independientes: Sean A, B eventos de un espacio muestral Ω , se dice que A, B son independientes si:

$$P(B \cap A) = P(B)P(A).$$

Probabilidad condicional: Sean A, B eventos de un espacio muestral Ω , la probabilidad de que ocurra B , dado que A ya ocurrió, está dado por:

$$P(B|A) = \frac{P(B \cap A)}{P(A)}.$$

Regla de Bayes: Sean A, B eventos de un espacio muestral Ω , se cumple que, por probabilidad condicional, se dice que la probabilidad de B dado A está dada por:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

Esta regla es producto de relacionar $P(B|A)$ y $P(A|B)$. En palabras llanas, se calcula la probabilidad de que las acciones de Santander suban, dado que las otras emisoras subieron o bajaron, i.e. Si las acciones de Santander y las de las otras emisoras pueden bajar o subir, la probabilidad de que Santander suba, dado el mercado está dado por la siguiente expresión:

$$P(S|X) = \frac{P(X|S)P(S)}{P(X)}$$

Donde S es el evento de que Santander suba, y X es el evento de que el mercado suba o baje. Los conceptos anteriores son definidos para subconjuntos de un espacio muestral, sin embargo, son heredados para variables aleatorias. Véase [10].

Algoritmo ingenuo de Bayes (también llamado Naïve Bayes): Considere un ejemplo x representado por k valores, el algoritmo de Naïve Bayes se basa en hallar la hipótesis más probable que describa ese ejemplo. Sean C_i las categorías de clasificación de una población, en este caso, las categorías son las siguientes:

Sean Y_t, Z_t los precios de apertura y de cierre de Santander, respectivamente para $t = 2, \dots, T$ (donde T es el número de precios diarios en la muestra) entonces:

$$C = \begin{cases} 0, & \text{si } Y_t < Z_{t-1} \\ 1, & \text{si } Y_t > Z_{t-1} \end{cases}$$

Sean $X_{j,t}$, variables predictoras, para $j \leq n$ donde n es el número de predictores, luego:

$$X_{j,t} = \begin{cases} 0, & \text{si Precio de cierre}_t < \text{Precio de apertura}_t \\ 1, & \text{si Precio de cierre}_t > \text{Precio de apertura}_t \end{cases}$$

En otras palabras, se toma en cuenta si las acciones de Santander amanecen cotizando por encima del cierre del día anterior, por otro lado se considera si los predictores (cotizaciones de empresas del mismo sector) tienen un cierre positivo el mismo día, es decir cierran con un precio mayor al de apertura.

Finalmente, se tiene que el estimador c^* :

$$c^* = \max_c P(C = c | X_1 = x_1, X_2 = x_2 \dots X_n = x_n) \\ = \max_c P(C = c) \prod_{i=1}^n P(X_i = x_i | C = c)$$

En el contexto de la investigación, el argumento máximo se interpreta como la combinación de eventos que maximiza la probabilidad de alza o baja de la acción. Para ampliar conceptos sobre algoritmo ingenuo de Bayes, véase [11]

Básicamente, se apelará al análisis técnico, pues se buscarán patrones de subida o bajada de precios, dado el movimiento del mercado, i.e. se utilizará el algoritmo de Bayes para determinar la combinación de eventos que maximiza la probabilidad de alza o baja de una acción. Dependiendo de la posición que posea el tenedor de la acción, se traducirá en las siguientes acciones: venta o compra, pues, dependiendo el caso, habrá un evento que maximiza la probabilidad de subida, y uno que maximiza el de bajada.

II. METODOLOGÍA Y DATOS

A continuación, se enlistarán los pasos que se llevaron a cabo para la estimación de probabilidades:

i. Obtención y limpieza de datos: Se descargaron precios de cotización diarios durante dos años de las acciones del sector financiero. Se considerarán los bancos BBVA, Banorte (BANORTE) y Banco del Bajío (BBAJIO). Se tomaron precios desde el 8 de junio de 2017 hasta el 30 de abril del 2019, obtenidos del portal financiero Yahoo Finance [14]. Dado que las emisoras no cotizan ciertos días, se discriminaron casos (limpieza) en los que las cuatro acciones no cotizaron de manera unánime. Dejando un total de 463 registros en la muestra. Los datos utilizados se ven resumidos en la gráfica de precios estandarizados (restando su media y dividiendo entre su desviación estándar) de la Figura 1 [12].

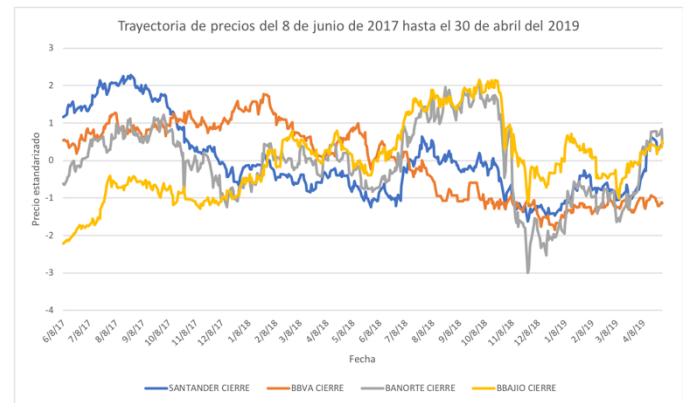


Figura 1. Precios estandarizados de cierre de algunas emisoras durante 2 años (tiempo en las abscisas y precio estandarizado en ordenadas). Fuente: Elaboración propia

ii. Cálculo de eventos: Pese a que se trabajará con variables dicotómicas, se harán las siguientes asignaciones bajo el criterio:

$$X_{i,j} = \begin{cases} 0, & \text{si Precio de cierre} < \text{Precio de apertura} \\ 1, & \text{si Precio de cierre} > \text{Precio de apertura} \end{cases} \\ \forall i = 1, \dots, n \\ \forall j = 1, \dots, k$$

donde k es el número de días analizados y n es el número de empresas analizadas.

iii. Cálculo de probabilidades: De acuerdo con el algoritmo ingenuo de Bayes, se calcularán las probabilidades de alza o baja de la acción como sigue: Sea

$$C = \begin{cases} 0, & \text{si la acción de Santander baja} \\ 1, & \text{si la acción de Santander sube} \end{cases}$$

Entonces, las probabilidades bayesianas están dadas por:

$$P(C = 1 | X_1 = x_1, X_2 = x_2 \dots X_n = x_n) = P(C = 1) \prod_{i=1}^n P(X_i = x_i | C = 1).$$

Análogamente, $C = 0$ si la acción baja.

iv. Clasificación: Una vez calculadas las probabilidades, se elige la combinación de eventos c^* , tal que:

$$c^* = \max_c P(C = 1 | X_1 = x_1, X_2 = x_2 \dots X_n = x_n)$$

v. Interpretación: Dado que c^* viene dado por una combinación de eventos conjuntos, que en el contexto de la investigación son: la probabilidad de que la acción de Santander suba de un día para otro, dado cierta combinación de eventos de emisoras del mismo sector (el día anterior), habrá

una combinación de eventos la cual, según la metodología, indica que la acción de Santander es más propensa a responder al alza. Para realizar el análisis y estimaciones del alza y baja del precio de una acción mediante el algoritmo de Bayes se utilizará el software libre R (The R Project for Statistical Computing), el cual es un conjunto integrado de librerías y un lenguaje de programación usado frecuentemente para estadística y econometría [13].

III. RESULTADOS

Utilizando el algoritmo ingenuo de Bayes definido en conceptos básicos, se diseñó un script en RStudio (Ver anexo 1) cuyos resultados son los siguientes:

i. Combinaciones de eventos posibles:

Puesto que se condicionó la probabilidad de que Santander subiera, los posibles eventos a los que responde la variación de la acción se muestran en el Cuadro 1:

Etiqueta del evento	BBVA	BANORTE	BBAJIO
E ₁	BAJA	BAJA	BAJA
E ₂	BAJA	BAJA	SUBE
E ₃	BAJA	SUBE	BAJA
E ₄	BAJA	SUBE	SUBE
E ₅	SUBE	BAJA	BAJA
E ₆	SUBE	BAJA	SUBE
E ₇	SUBE	SUBE	BAJA
E ₈	SUBE	SUBE	SUBE

FUENTE: ELABORACIÓN PROPIA

ii. Cálculo de probabilidades. La probabilidad de que una acción de Santander suba es:

$$P(C = 1) = \frac{\#de\ veces\ que\ sube\ Santander}{\#Total\ de\ Datos} = \frac{237}{462} = 0.5129$$

iii. Aplicación del algoritmo: Una vez calculada la probabilidad de que Santander suba en cualquier momento, se calculan las probabilidades condicionadas de acuerdo con el algoritmo ingenuo de Bayes que se muestran en el Cuadro 2. Se ilustra la metodología calculando la probabilidad de la etiqueta E₁:

$$\begin{aligned} P(C = 1|E_1) &= P(C = 1) * P(X_1 = 0|C = 1) \\ &\quad * P(X_2 = 0|C = 1) * P(X_3 = 0|C = 1) \\ &= 0.5129 * 0.5949 * 0.5021 * 0.5527 \\ &= 0.08470 \end{aligned}$$

el procedimiento es análogo para las demás etiquetas

Etiqueta del evento	Valor
P(C=1 E ₁)	0.0847
P(C=1 E ₂)	0.0685
P(C=1 E ₃)	0.084
P(C=1 E ₄)	0.068
P(C=1 E ₅)	0.0577
P(C=1 E ₆)	0.0467
P(C=1 E ₇)	0.0572
P(C=1 E ₈)	0.0463

FUENTE: ELABORACIÓN PROPIA

Ahora, aplicando la regla del algoritmo, donde los eventos E_k son especificados en el Cuadro 2.

$$\begin{aligned} c^* &= \max_c P(C = 1|E_k), \forall k \leq 8 \\ &= E_1 \end{aligned}$$

Es decir, el evento que maximiza la probabilidad de que la acción de Santander suba, dadas todas las combinaciones posibles del movimiento del mercado es:

BBVA: BAJA, BANORTE: BAJA, BBAJIO: BAJA

En otras palabras, de todos los movimientos que se pueden observar en el mercado, ésta es a la que Santander más responde hacia el alza. i.e. si se observa tal evento, teniendo una posición larga, los resultados indican que vender la acción es una estrategia inteligente.

Para validar los resultados obtenidos, se utiliza una lista de precios con las mismas emisoras, pero emitidos después de los precios utilizados para la estimación de parámetros mostrados en el Cuadro 2. Es decir, se valida con precios emitidos del 2 de mayo al 6 de septiembre de 2019 (estos precios no tuvieron interacción para el entrenamiento del modelo).

El procedimiento consiste en contar la proporción de ocurrencia de cada una de las combinaciones de eventos mostradas en el Cuadro 1 (cuando la acción de Santander subió). Estas proporciones que para efectos de la investigación, se interpretan como calidad de las predicciones del modelo.

BBVA	BANORTE	BBAJIO	Proporción
BAJA	BAJA	BAJA	22.50%
BAJA	SUBE	BAJA	20.00%
BAJA	SUBE	SUBE	15.00%
BAJA	BAJA	SUBE	7.50%
SUBE	BAJA	SUBE	12.50%
SUBE	BAJA	BAJA	5.00%
SUBE	SUBE	BAJA	7.50%
SUBE	SUBE	SUBE	10.00%

FUENTE: ELABORACIÓN PROPIA

La validación muestra que la etiqueta E₁ tuvo una proporción de ocurrencia del 22.50%, la cual es la más alta en el Cuadro 3, por lo que se podría inferir que el modelo predice correctamente los casos (como es de esperarse). Lo cual sugiere que el evento representado por la etiqueta E₁ sí es factor ante el alza de las acciones de Santander.

IV. CONCLUSIONES

Este trabajo proporciona evidencia de acercamiento al comportamiento del mercado financiero mediante herramientas de minería de datos. La herramienta propuesta en esta investigación pretende evaluar de manera objetiva probabilidad de alza de una acción condicionada a los posibles movimientos de empresas del mismo sector. Los resultados de la investigación reflejaron que para las acciones de Santander

reaccionan favorablemente ante ciertos eventos, por ejemplo, cuando BBVA: baja, BANORTE: baja, BBAJIO: baja. La metodología aquí planteada se puede extrapolar a cualquier empresa, y a otras emisoras del mismo sector. El método utilizado fue validado con la información proporcionada por la BMV, mostrando el gran potencial que proporciona la minería de datos como apoyo para estimar el alza o baja del precio de acciones. Este resultado permite apreciar la gran utilidad y potencial de herramientas de la inteligencia artificial en los mercados financieros.

Como se mostró en el Cuadro 3, las predicciones del modelo no son perfectas, ya que sólo se ha trabajado con precios históricos. Esto provoca que el modelo no cuente con los parámetros necesarios para realizar una predicción óptima. Pero puede verse que, para una mejor predicción, es necesario contar con una muestra más grande, así como mayor número de predictores.

La contribución principal de esta investigación ha sido proporcionar un enfoque de minería de datos con algoritmos probabilísticos únicamente con precios históricos. Otras investigaciones utilizan algoritmos como Máquinas de Soporte Vectorial, Redes Neuronales, Regresión, entre otras, por lo que este artículo enfatiza que pretende predecir a partir de probabilidades condicionales.

Por último, es importante destacar que, aunque los modelos sean buenos es importante contar con una estrategia de inversión que se ajuste al modelo propuesto; dejando abierta la posibilidad de un estudio más extenso y profundo en el que puedan contrastar diferentes algoritmos en diferentes sectores además de incluir algunas estrategias de inversión.

AGRADECIMIENTOS

El autor agradece a la Dra. María del Carmen Lozano Arizmendi por su apoyo a elaborar este documento, así como su guía para desarrollar la investigación.

REFERENCIAS

- [1] Díez de Castro L.T., Mercados financieros internacionales, Dykinson, España, 2007, p 9.
- [2] Lorenz, Edward. The Essence of Chaos. Washington. University of Washington Press, 1995.
- [3] Rao, A., et al. Survey: Stock Market Prediction Using Statistical, Computational Methodologies and Artificial Neural Networks. International, Research Journal of Engineering and Technology , 2015 (08).
- [4] De Arce, R. 20 años de modelos ARCH: una visión de conjunto de las distintas variantes de la familia. Estudios de economía aplicada, 2004, Vol. 22 (1)
- [5] Hernández, S. Pronóstico y volatilidad del IPC de la bolsa mexicana de valores. 2009, pp 25-36. <https://core.ac.uk/download/pdf/6436970.pdf>
- [6] Murphy, J. Análisis Técnico de los Mercados Financieros. Barcelona. España: Gestión 2007.
- [7] Fama, E., "The behavior of stock market prices". Journal of Business. 1965 (38): pp 34-105.
- [8] A. Pina. Análisis técnico de mercados financieros basado en técnicas de inteligencia artificial. Universidad de Murcia, 2014. Recuperado de <https://www.um.es/documents/118351/1884002/TFG-PINA+CANELLES.pdf/87e36c8c-5038-4a7d-bf27-e401f58b9548>, pp 9-11
- [9] Ross S.A., Westerfield R.W., Jaffe J.F. Finanzas Corporativas. McGraw Hill, 2012, pp 329-334

- [10] Port H., Introduction to Probability, Houghton Mifflin Company, 1971, pp 14-18.
- [11] P. Larrañaga, Iñaki Inza y Abdelalok Moujahid. Clasificadores Bayesianos. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad del País Vasco, 2019. Recuperado de https://www.researchgate.net/publication/265797118_Tema_12_Clasificadores_Bayesianos, pp 5-8.
- [12] Sánchez Arriaga L.A., Lista de precios de apertura y cierre de Santander, BBVA, Banco del bajío y Banorte. CLIDI. Archivo DOI. <https://doi.org/10.6084/m9.figshare.8312456>
- [13] Development Core Team, (2005). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponible en: <http://www.r-project.org>
- [14] Yahoo! (1997). R: Finance – Business Finance, Stock Market, Quotes, News. Disponible en: <https://finance.yahoo.com/>
- [15] El Universal (2018). Inclusión financiera en México se estanca; desde 2015 el nivel de bancarización es de 68%. Disponible en: <https://www.eleconomista.com.mx/sectorfinanciero/Inclusion-financiera-en-Mexico-se-estanca-desde-2015-el-nivel-de-bancarizacion-es-de-68-20181123-0033.html>

ANEXO 1

Código para el algoritmo en Rstudio

```
rm(list=ls())
library(readxl)
Datos=read_xlsx("/Users/luisarriaga/Dropbox/T
ESIS/dataset.xlsx")#Importación de datos
unos=c(rep(1,ncol(Datos)-2),rep(0,ncol(Datos)
-2))#Vector de unos y ceros

respuesta=1 #?La acción sube o baja?
priors=as.data.frame(rbind(numeric(ncol(Datos)
-1)))
colnames(priors)=colnames(Datos)[2:ncol(Datos)
]]
priors=t(colMeans(Datos[2:ncol(Datos)]))#Prob
abilidad de alza a priori
attach(Datos)
library(gtools)#Librería de combinaciones
combinaciones=permutations(2,ncol(Datos)-2,v=
unos,repeats.allowed = T)
probabilidades=combinaciones*0
c_estrella=c()

for(i in 1:nrow(probabilidades)){ #Cálculo de
probabilidades bayesianas
for(j in 1:ncol(probabilidades)){
probabilidades[i,j]=sum(SANTANDER==respuesta&
Datos[,2+j]==combinaciones[i,j])/sum(SANTANDE
R==respuesta)
}
c_estrella[i]=prod(probabilidades[i,])
}
c_=combinaciones[which.max(c_estrella),]
```