

Universidad de Lima

Facultad de Ingeniería y Arquitectura

Carrera de Ingeniería de Sistemas



PREDICCIÓN DE PRECIOS DE ACCIONES DE BOLSA DE VALORES UTILIZANDO SUPPORT VECTOR REGRESSION

Trabajo de Investigación para optar el título profesional de Ingeniero de Sistemas

Víctor Andrés Edgard Cáceres Chian

Código 20122365

Asesor

Juan Manuel Gutiérrez Cárdenas

Lima – Perú

Abril 2018



**PREDICCIÓN DE PRECIOS DE ACCIONES DE
LA BOLSA DE VALORES UTILIZANDO
SUPPORT VECTOR REGRESSION**

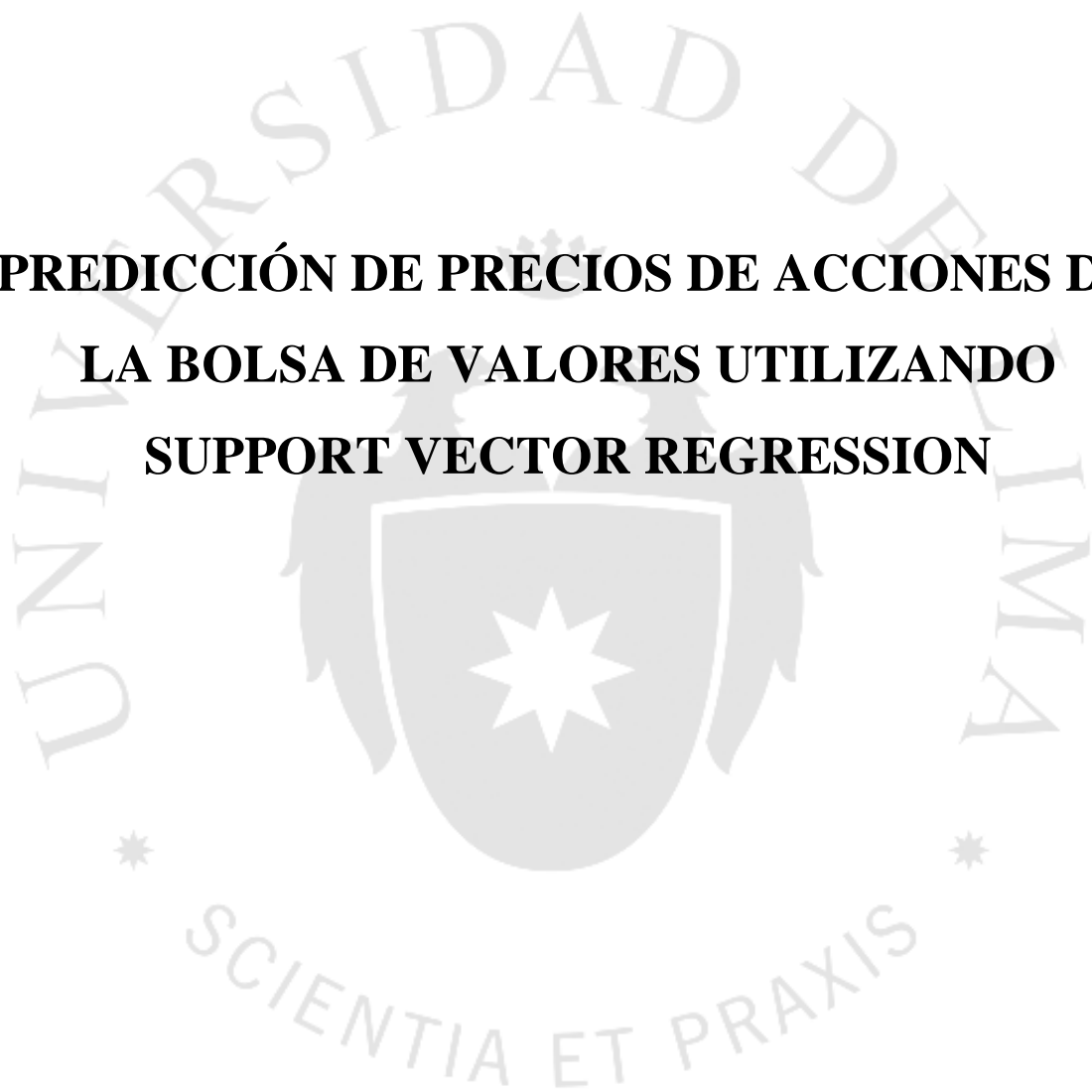
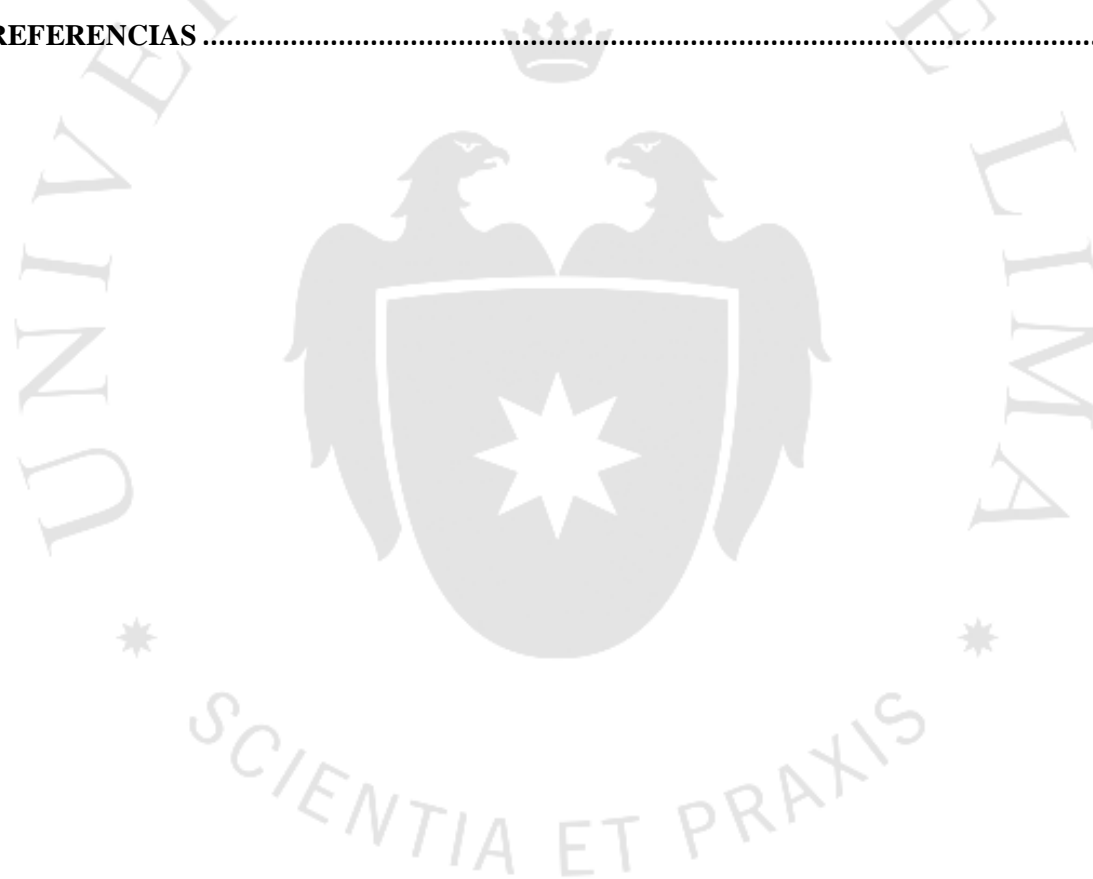


TABLA DE CONTENIDO

Contenido

RESUMEN	1
ABSTRACT	2
INTRODUCCIÓN	3
DESCRIPTORES TEMÁTICOS	4
CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA	5
1.1. Formulación del problema.....	5
1.2. Objetivo de la investigación	6
1.2.1. Objetivo general.....	6
1.2.2. Objetivos específicos	6
1.3. Justificación.....	6
1.4. Aportes	7
CAPÍTULO II: ESTADO DEL ARTE	8
CAPÍTULO III: MARCO TEÓRICO	17
3.1. Mercados de Valores	17
3.2. Acciones Ordinarias	19
3.3. Análisis Fundamental	21
3.4. Análisis Técnico	25
3.5. Standard and Poor Dow Jones Indexes	29
3.6. Minería de Datos	30
3.7. Support Vector Machine.....	31
3.8. Kernels	34
3.9. Support Vector Regression.....	36
CAPÍTULO IV: PROPUESTA DE SOLUCIÓN	39
4.1. Métodos de investigación	39
4.2. Alcance.....	40
4.3. Riesgos	40
4.4. Entregables.....	40

CAPÍTULO V: PLAN DE TRABAJO	41
CAPÍTULO VI: PRUEBAS Y RESULTADOS	47
6.1. Análisis Técnico	47
6.2. Análisis Fundamental	50
6.3. Análisis Técnico y Fundamental	53
6.4. Subgrupo de Análisis Técnico y Fundamental	56
6.5. Comparación de Resultados	59
6.6. Comparación con Support Vector Machine	64
CONCLUSIONES	67
Recomendaciones	69
REFERENCIAS	70



ÍNDICE DE TABLAS

Tabla 5. 1 Indicadores de Acciones	42
Tabla 6. 1 Análisis Técnico: Indicadores.....	47
Tabla 6. 2 Análisis Técnico: Parámetros de Modelo	48
Tabla 6. 3 Análisis Técnico: Resultados.....	48
Tabla 6. 4 Análisis Fundamental: Indicadores.....	50
Tabla 6. 5 Análisis Fundamental: Parámetros de Modelo.....	51
Tabla 6. 6 Análisis Fundamental: Resultados.....	51
Tabla 6. 7 Análisis Técnico y Fundamental: Indicadores.....	53
Tabla 6. 8 Análisis Técnico y Fundamental: Parámetros de Modelo	54
Tabla 6. 9 Análisis Técnico y Fundamental: Resultados	54
Tabla 6. 10 Subgrupo Análisis Técnico y Fundamental: Indicadores	56
Tabla 6. 11 Subgrupo Análisis Técnico y Fundamental: Parámetros de Modelo.....	57
Tabla 6. 12 Subgrupo Análisis Técnico y Fundamental: Resultados	57
Tabla 6. 13 Comparación de Resultados	59
Tabla 6. 14 Comparación de Rendimientos.....	61
Tabla 6. 15 Comparación de Rendimientos (Negativos).....	63
Tabla 6. 16 Support Vector Machine: Resultados	64
Tabla 6. 17 Support Vector Machine: Resultados (2).....	64
Tabla 6. 18 Support Vector Machine: Rendimientos	65



ÍNDICE DE FIGURAS

Figura 2. 1 Indicadores de Entrada	11
Figura 2. 2 Indicadores de Entrada	12
Figura 2. 3 Indicadores de Entrada	12
Figura 2. 4 Indicadores de Entrada	13
Figura 3. 1 Ciclo Financiero	17
Figura 3. 2 Rendimiento de Cartera de Acciones	25
Figura 3. 3 Promedios Móviles.....	27
Figura 3. 4 Promedios Móviles y Momentum	28
Figura 3. 5 Sectores de Empresas de S&P.....	30
Figura 3. 6 Datos separados por Múltiples Planos.....	32
Figura 3. 7 Hiperplano Óptimo de Separación	32
Figura 5. 1 Flujograma de Trabajo	41
Figura 6. 1 Análisis Técnico: Real vs Pronóstico. Datos Normalizados	49
Figura 6. 2 Análisis Técnico: Real vs Pronóstico. Datos Desnormalizados.....	49
Figura 6. 3 Análisis Fundamental: Real vs Pronóstico. Datos Normalizados	52
Figura 6. 4 Análisis Fundamental: Real vs Pronóstico. Datos Desnormalizados	52
Figura 6. 5 Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Normalizados.....	55
.....	55
Figura 6. 6 Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Desnormalizados.	55
.....	55
Figura 6. 7 Subgrupo Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Normalizados	58
Figura 6. 8 Subgrupo Análisis Técnico y Fundamental. Real vs Pronóstico. Datos Desnormalizados.....	58

ÍNDICE DE ANEXOS

Anexo 1: Comandos SVMLight	80
----------------------------------	----



RESUMEN

En esta tesis se investiga la aplicación de Support Vector Machine y Support Vector Regression en la predicción de precios de acciones. Cuatro modelos de cada algoritmo se construyen basados en el análisis técnico y análisis fundamental, ambos utilizados en finanzas para estudiar el movimiento de los precios de acciones. Los modelos son luego evaluados por el rendimiento que obtienen las acciones, uno de los principales criterios de decisión para inversiones. Los resultados favorecen un modelo combinado de Support Vector Regression basados en el análisis técnico y análisis fundamental como forma de decisión frente a una variedad de acciones.



ABSTRACT

In this paper, Support Vector Machine and Support Vector Regression application to forecasting of stock prices is investigated. Four models of each algorithm are built based on technical analysis and fundamental analysis, both used in finance to study stock price movement. The forecast models are then evaluated by their resulting actual return, one of the main decision criteria for investments. Results favor a combined Support Vector Regression model of technical and fundamental analysis as way to decide upon a variety of stocks.



INTRODUCCIÓN

Las tecnologías que se han desarrollado por las últimas décadas han influenciado muchos aspectos de la vida actual, cambiando el enfoque y estilo de algunas de ellas y creando, en otras, oportunidades antes no posibles.

El Internet ha tenido un gran rol en la apertura del mundo de las inversiones a una mayor cantidad de personas. Mayor número de canales de información y facilidades para la compra y venta de instrumentos financieros han sido resultado del desarrollo de las comunicaciones. Hace 10 años, alrededor del 50% de estadounidenses ya era partícipe del mercado financiero, teniendo a su nombre acciones comunes de diversas empresas (Gitman, Joehnk y Aguado, 2005, p. 3).

Dentro de los instrumentos financieros, se encuentran las acciones comunes, teniendo entre sus características principales la flexibilidad y una alta rentabilidad a comparación de otros instrumentos como los bonos o acciones preferentes. Sin embargo, la característica negativa más resaltante es el riesgo, pudiendo ser entendido como la variabilidad de un resultado respecto a su valor esperado (Welch, 2014, p. 123). Esto puede resultar en rentabilidades menores a las esperadas o incluso negativas, perdiendo parte de las inversiones iniciales.

Por tanto, los precios de las acciones han sido los datos financieros más analizados por de las últimas décadas, siendo muchas técnicas empleadas con diferentes grados de éxito (Granger, 1992). Support Vector Machine ha sido una de ellas, mostrando progresos en pronósticos de datos financieros.

El proyecto consiste en implementar un algoritmo de predicción de precios de acciones basado en Support Vector Machine, que permita a inversionistas individuales y administradores de fondos mejorar la rentabilidad de sus inversiones, puesto que la predicción precisa de los precios de los instrumentos financieros es esencial para tomar mejores decisiones de inversiones con mínimo riesgo (Das y Padhy, 2012)

DESCRIPTORES TEMÁTICOS

- Mercado de Valores
- Acciones Comunes
- Análisis Técnico
- Análisis Fundamental
- Standard & Poor
- Minería de Datos
- Support Vector Machine
- Support Vector Regression
- Kernels



CAPÍTULO I: PLANTEAMIENTO DEL PROBLEMA

1.1. Formulación 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 (Gitman et al., 2008, p. 28).

Uno de los roles de las instituciones financieras es de canalizar los ahorros individuales a la inversión privada, necesaria para el desarrollo de las empresas. Otra es disminuir el riesgo, financiando proyectos entre gran cantidad de personas mediante las acciones en lugar de que una sola persona arriesgue su capital (Berk, DeMarzo y Harford, 2009, p.18).

Por otro lado, los compradores utilizan las acciones como formas de inversión con el objetivo de obtener una rentabilidad a futuro superior a otras alternativas como depósitos o bonos. Si bien las acciones no se caracterizan por los dividendos pagados (siendo menores que otros instrumentos financieros como los bonos), son una de las mejores opciones para ganancias de capital. Por ejemplo, las acciones promediaron desde el año 1950 hasta el 2000 una rentabilidad anual de 12%, mientras que los bonos tuvieron una rentabilidad anual de 6% (Gitman et al., 2008, p. 150-151).

Sin embargo, tienen la desventaja significativa del riesgo. Las acciones se encuentran sujetas a varios tipos de riesgo, incluyendo de negocio, financieros, mercados y eventos externos, que pueden afectar negativamente a los beneficios de las acciones y la rentabilidad esperada (Gitman et al., 2008, p. 155).

La tendencia de incremento del capital de las acciones junto con el riesgo que traen, ponen como centro de atención a la compra y venta de acciones. Los precios de las acciones son el elemento esencial en la compra y venta de las mismas, asistiendo a los inversionistas en las decisiones de compraventa en el mercado (Gitman et al., 2008, p. 159).

La rentabilidad de las acciones proviene mayormente por las ganancias de capital, que son las diferencias de precios entre la compra y venta de la acción. Como no se conoce esta variación en el precio a futuro al momento de comprar, solo se puede determinar la rentabilidad de una acción al momento de venderla. El riesgo consiste en que esta rentabilidad sea menor o incluso mayor que la esperada, perdiendo oportunidades si es que no se optó por estas u otras acciones.

Por tanto, un algoritmo de predicción de precios asistiría a tomar mejores decisiones de inversión a inversionistas individuales, administradores de fondos y demás participantes para obtener una mayor rentabilidad con menor riesgo.

1.2. Objetivo de la investigación

1.2.1. Objetivo general

Implementar y evaluar la factibilidad de un modelo de Support Vector Machine para la predicción de precios de acciones a mediano plazo (un trimestre) de la bolsa de valores de S&P.

1.2.2. Objetivos específicos

Los objetivos específicos de esta tesis permitirán alcanzar el objetivo general:

- a) Determinar variables de análisis técnico y análisis fundamental a utilizar
- b) Desarrollar modelos utilizando distintas variables y compararlos *
- c) Comparar modelos de Support Vector Machine y Support Vector Regression

1.3. Justificación

El objetivo de la investigación es evaluar un modelo que permita una mejor predicción de precios de acciones a mediano plazo (un trimestre) para asistir en las decisiones de compraventa por parte de los inversionistas en el mercado de valores.

Un modelo de predicción de precios asistirá a inversionistas, administradores de fondos comunes y demás participantes de la bolsa de valores a mejorar sus inversiones,

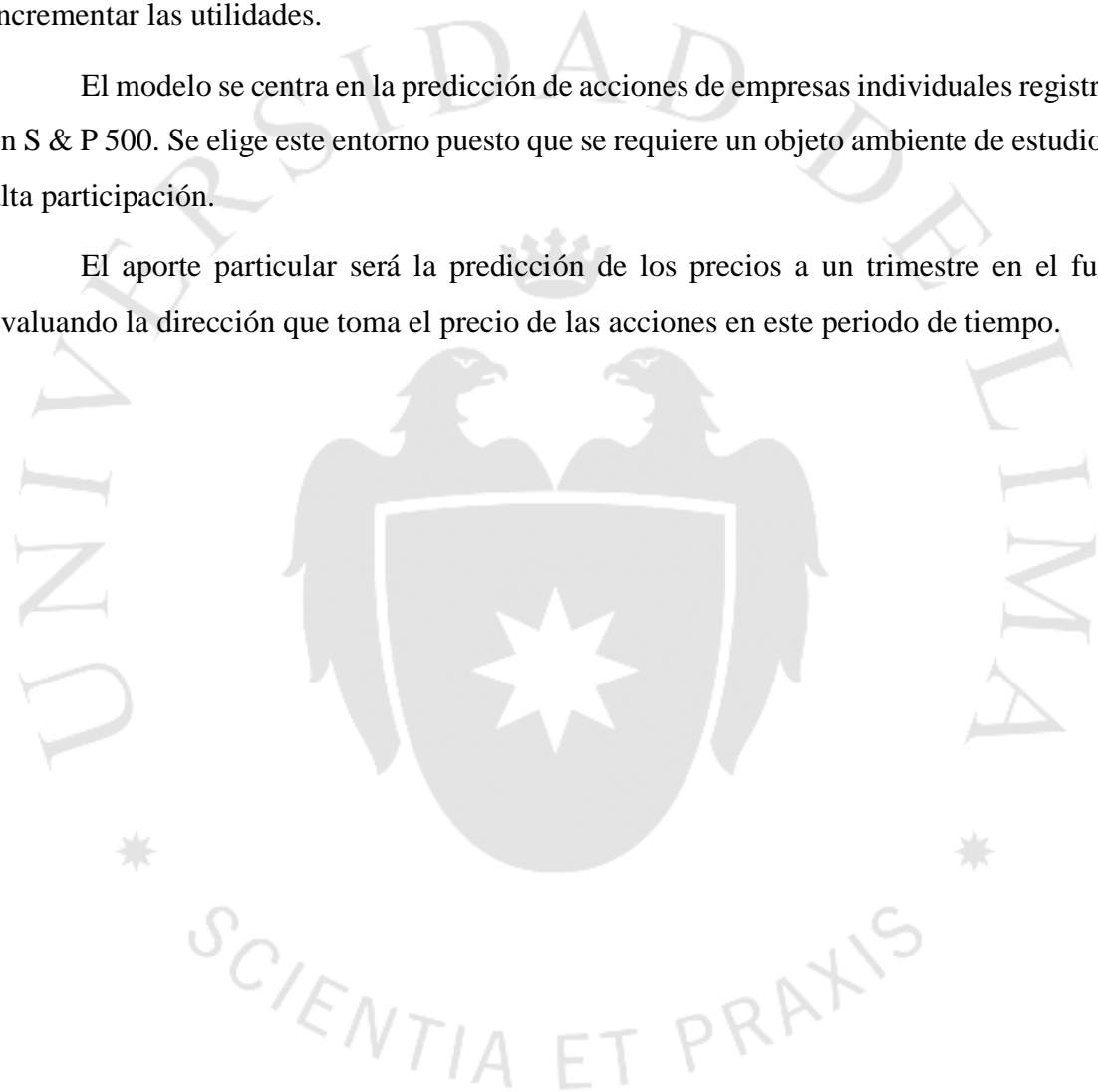
reduciendo las posibilidades de utilidades negativas, mejorando sustancialmente las inversiones tomadas e incrementando utilidades como resultado final.

1.4. Aportes

El aporte este modelo será la predicción de precios de acciones que sirva como soporte a las decisiones de inversión para reducir los riesgos que tiene toda inversión y a largo plazo incrementar las utilidades.

El modelo se centra en la predicción de acciones de empresas individuales registradas en S & P 500. Se elige este entorno puesto que se requiere un objeto ambiente de estudio con alta participación.

El aporte particular será la predicción de los precios a un trimestre en el futuro, evaluando la dirección que toma el precio de las acciones en este periodo de tiempo.



CAPÍTULO II: ESTADO DEL ARTE

La tecnología se ha desarrollado a gran velocidad en los últimos años y ha tenido un impacto significativo en todos los ámbitos, creando nuevas oportunidades en todos ellos y brindando nuevas herramientas para abordar problemas existentes con distintos enfoques con la esperanza de encontrar nuevas soluciones que tengan mayor éxito.

Hace 34 años se fundó el International Institute of Forecasting, una organización sin fines de lucro que busca generar, distribuir y utilizar el conocimiento de predicciones para, entre otros, mejorar la calidad y la utilidad de las mismas (International Institute of Forecasters IIF, 2016)

Uno de los aspectos que cubre este instituto es la predicción de series de tiempo y, desde su fundación, se ha observado un creciente interés en la predicción de las series de tiempo financieras por razones académicas y prácticas (Dai, Shao y Lu, 2013) representando casi un tercio de las publicaciones del instituto (De Gooijer y Hyndman, 2006).

Por la parte académica, estas series son consideradas una de las aplicaciones de mayor desafío de predicciones de series de tiempo (Tay y Cao, 2001) y, en la parte práctica, representan uno de gran dificultad que afrontan las organizaciones financieras, empresas e inversionistas privados (Dai et al., 2013).

La dificultad de las series de tiempo financieras radica en que el mercado es un sistema complejo, evolutivo y dinámico, siendo esta área caracterizada por gran cantidad de datos, ruido, no-estacionalidad, de naturaleza no estructurada, de alto grado de incertidumbre y con dependencias escondidas (Huang, Nakamori y Wang, 2005). Otro punto que contribuye a la dificultad es que muchos factores interactúan en el mercado financiero, que incluyen eventos políticos, condiciones económicas y hasta las expectativas de los mismos participantes. Sin embargo, los movimientos no son aleatorios, sino que se comportan de una forma dinámica no lineal (Choudhry y Garg, 2008).

En el artículo de Cao (2002), se explica que las series de tiempo tienen dos problemas principales: el ruido y la no estacionaridad. El ruido de las series se refiere a la falta de

información disponible para determinar la dependencia entre los precios pasados y futuros pudiendo derivar en un problema de sobreajuste o subajuste del modelo, ocasionando un pobre desempeño con nuevos datos. La no estacionaridad se refiere a las relaciones cambiantes entre las variables de entrada y de salida a lo largo del tiempo, variando el grado de sus dependencias. Tay y Cao (2001) le agregan una característica a las series financieras, que son deterministamente caóticas. Esto significa que a largo plazo son consistentes pero a corto plazo difíciles de predecir.

Por tanto, muchas técnicas de inteligencia artificial y de machine learning han sido empleadas en los últimos años para predecir los precios de las acciones comunes en el mercado de valores. La técnica más popular ha sido la de redes neuronales (Choudhry y Garg, 2008; Dai et al., 2013).

Comparando redes neuronales con modelos estadísticos tradicionales, éstas son modelos impulsados por los datos, siendo menos susceptibles a los problemas de especificación de parámetros. Son también flexibles, adecuándose a sistemas dinámicos mediante un proceso de retención de patrones de datos (Tay y Cao, 2001). Poseen la capacidad de poder capturar las características no lineales de las series financieras, identificando dichas relaciones no lineales entre las entradas y las salidas (García, Jalal, Garzón y López, 2013).

Estos modelos fueron inspirados en las redes neuronales biológicas, intentando replicar el comportamiento de las células neuronales. Son un conjunto interconectado de nodos y enlaces entrenados para resolver problemas de clasificación (Tan, Steinbach y Kumar, 2005, p. 247).

Sin embargo, en dos artículos (Das y Padhy, 2012; Tay y Cao, 2001) se exponen las debilidades del modelo de redes neuronales. Se explican tres problemas que tiene, siendo el primero la presencia de un gran número de parámetros en las redes neuronales. El segundo problema es de sobreajuste derivado de una gran capacidad de capturar información resultando en buenos resultados con datos de entrenamiento pero de menor grado con nuevos datos. El tercer y último problema identificado es la dificultad de obtener una solución estable, puesto que no es posible encontrar una solución única global óptima para la serie de tiempo analizada.

Vladimir Vapnik desarrolló otra técnica de clasificación llamada Support Vector Machine (Das y Padhy, 2012). Está basada en el principio de minimización de riesgo estructural para encontrar una función resistente al problema de sobreajuste de las redes neuronales y tenga buen rendimiento con nuevos datos. Tiene también la característica clave de tener una sola solución única y global (Huang et al., 2005). El modelo como resultado ha tenido mejoras considerables en muchas aplicaciones predictivas (Das y Padhy, 2012).

Support Vector Machine es un tipo de clasificadores de máximo margen. Se busca encontrar un plano óptimo que separe a dos clases de datos, maximizando la distancia del plano encontrado con los datos de entrenamiento. El plano es llamado hiperplano de separación óptimo y los datos de entrenamiento más cercanos son denominados vectores de soporte. Los nuevos datos se pueden entonces clasificar dependiendo de su posición respecto al hiperplano hallado (Choudhry y Garg, 2008).

Las variables de entrada utilizadas por distintos artículos difieren entre sí. El artículo de Tay y Cao (2001) por ejemplo utiliza el precio de cierre como datos de entrada y cinco indicadores adicionales: porcentaje de diferencia de precios relativos agrupados en periodos de 5 días (RDP-5, RDP-10, RDP-15 y RDP-20) y un precio de cierre transformado (EMA 15), mostrados en la figura 2.1. El artículo de Das y Padhy (2012) añade a los anteriores 6 tipos de datos de entrada: precio de apertura, precio máximo, precio mínimo, volumen y valor intercambiado.

Figura 2. 1

Indicadores de Entrada

Table 2
Input and output variables^a

Indicator	Calculation
Input variables	
EMA15	$P(i) - \overline{EMA_{15}(i)}$
RDP-5	$(p(i) - p(i - 5))/p(i - 5) * 100$
RDP-10	$(p(i) - p(i - 10))/p(i - 10) * 100$
RDP-15	$(p(i) - p(i - 15))/p(i - 15) * 100$
RDP-20	$(p(i) - p(i - 20))/p(i - 20) * 100$
Output variable	
RDP+5	$\frac{\overline{p(i + 5)} - \overline{p(i)}}{\overline{p(i) = EMA_3(i)}} * 100$

^aEMA_n(i) is the n-day exponential moving average of the ith day; and p(i) is the closing price of the ith day.

Fuente: Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. Omega, 29(4), 309-317.

Por su parte, Choudhry y Garg (2008) basan su selección de variables en el análisis técnico, recogiendo 35 indicadores técnicos de la figura 2.2 que utilizados por expertos financieros. Algunos de éstos indicadores son impulso o momento (diferencia de precios en 5 días), ratio de cambio, disparidad en 5 y 10 días (en porcentajes). Dai et al. (2013) por su parte utilizan 11 indicadores técnicos como variables de entrada para su modelo mostrados en la figura 2.3. Entre ellos se encuentran, el precio de apertura, promedio móvil, indicadores de fuerza relativa y línea psicológica del mercado (número de días de subida de los últimos 5 días).

SCIENTIA ET PRAXIS

Figura 2. 2

Indicadores de Entrada

Feature Name	Formula
Momentum	$(C(i)/C(i-N))* 100$
Williams %R	$(HH(n)-C(t))/ (HH(n)-LL(n))* 100$
Rate of Change (ROC)	$(C(t) - C(t-n))/C(t-n)$
5 Day Disparity	$(C(t)/MA(5))*100$
10 Day Disparity	$(C(t)/MA(10))*100$
Stochastic %K	$(C(t) - L(t))/(H(t) - L(t))$
Price Volume Trend (PVT)	$((C(t) - C(t-1)) / C(t-1)) * V$

Fuente: Choudhry, R., & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39(3), 315-318.

Figura 2. 3

Indicadores de Entrada

Variables	Description
X1	Today's opening index value (Opening)
X2	Previous day's highest index value (Highest)
X3	Previous day's lowest index value (Lowest)
X4	Previous day's closing index value (Closing)
X5	5-day moving average (MA5): $MA5 = \frac{\sum_{i=1}^5 \text{Closing index on day } i}{5}$
X6	10-day moving average (MA10): $MA10 = \frac{\sum_{i=1}^{10} \text{Closing index on day } i}{10}$
X7	5-day relative strength indicator (RSI5): $RSI5 = 100 - \frac{100}{1+RS}$, where $RS = \frac{\text{Average of the past 5 days' up closes}}{\text{Average of the past 5 days' down closes}}$
X8	10-day relative strength indicator (RSI10): $RSI10 = 100 - \frac{100}{1+RS}$, where $RS = \frac{\text{Average of the past 10 days' up closes}}{\text{Average of the past 10 days' down closes}}$
X9	5-day psychological line (PSY5): $PSY5 = \frac{\text{Number of upward days over the past 5 days}}{5} \times 100$
X10	5-day BIAS (BIAS5): $BIAS5 = \frac{\text{Closing index today} - MA5}{MA5}$
X11	5-day price rate of change (ROC5): $ROC5 = 100 \times \frac{(\text{Closing index today} - \text{closing index 5 days ago})}{\text{Closing index 5 days ago}}$

Fuente: Dai, W., Shao, Y. E., & Lu, C. J. (2013). Incorporating feature selection method into support vector regression for stock index forecasting. *Neural Computing and Applications*, 23(6), 1551-1561.

El artículo de Wang, Hou, Wang y Shen (2016) utiliza 20 indicadores técnicos para su predicción, mostrados en figura 2.4. Entre ellos el ratio de cambio de cantidad y de valor, promedios móviles de diferentes cantidades y también una línea psicológica del mercado.

Figura 2. 4

Indicadores de Entrada

Number	Indicator	Explanations
1	Opening	Opening price of the current day
2	Closing	Closing price of the current day
3	Highest	Highest price of the current day
4	Lowest	Lowest price of the current day
5	Volume	Trading volume of the current day
6	Amount	Transaction amount of the current day
7	Change ratio	Change ratio of the current day
8	Change amount	Change amount of the current day
9	MA5	Moving average for 5 days
10	MA10	Moving average for 10 days
11	MA20	Moving average for 20 days
12	MACD	Moving average convergence-divergence
13	DIF	Difference
14	KDJ.K	Stochastic Oscillator
15	KDJ.D	Stochastic Oscillator
16	PSYMA6	Psychological line for 6 days
17	RSI6	Relative strength index for 6 days
18	RSI12	Relative strength index for 12 days
19	BIAS6	Bias ratio for 6 days
20	BIAS24	Bias ratio for 24 days

Fuente: Wang, J., Hou, R., Wang, C., & Shen, L. (2016). Improved v-Support vector regression model based on variable selection and brain storm optimization for stock price forecasting. *Applied Soft Computing*, 49, 164-178.

Otro estudio como el de Huang et al. (2005) incluyen también variables ajenas a los precios de las acciones pero que influyen del mercado. Dentro de estas variables se encuentran las tasas de interés, el índice del consumidor, el producto bruto interno y la tasa de cambio de divisas.

La gran cantidad de variables que se pueden tomar en cuenta para la predicción de precios de acciones comunes lleva a que sea importante un análisis o filtro previo de las variables antes de introducirlas al modelo.

Una desventaja de Support Vector Machine que subrayan Dai et al. (2013) es la deficiencia de poder capturar la importancia relativa de las variables independientes sobre las dependientes cuando muchas variables independientes son consideradas.

Dai et al. (2013) por tanto propusieron un modelo de MARS (multivariate adaptive regression splines) para la clasificación de variables de entrada antes de utilizar el modelo de predicción basado en Support Vector Machine, que tenía mejores resultados comparado a utilizar el mismo modelo sin clasificación de variables.

Análogamente, Guajardo, Weber y Miranda (2006) propusieron una metodología genérica para la predicción de series de tiempo que involucra un algoritmo de agrupación de variables buscando un grupo óptimo para mejores resultados. Si bien su metodología no es específicamente para series de tiempo financieras, puede ser utilizada al ser de tipo genérica.

Wang et al. (2016) usan el PCA o Principal Component Analysis para determinar las variables más influyen en la variancia de los resultados. De esta forma pueden reducir la cantidad de dimensiones utilizadas en el modelo propuesto.

De la misma forma, Choudhry y Garg (2008) utilizan un algoritmo genético para determinar las variables óptimas a utilizar antes de ser empleadas por el modelo de predicciones de Support Vector Machine. De los 35 indicadores, utilizan este algoritmo para determinar la combinación que tenga los mejores resultados utilizando Support Vector Machine

Los experimentos realizados por previas investigaciones se concentran mayormente en el pronóstico de los índices del mercado de valores en lugar de acciones comunes de empresas individuales. Los mercados objetos de estudio son distintos.

Das y Padhy (2012) por ejemplo predice el comportamiento del índice de NSE (National Stock Exchange) de la India. Huang et al. (2005) utilizan el índice de Nikkei 225 que corresponde a las 225 acciones intercambiadas en el mercado de valores de Tokyo (Tokyo Stock Exchange). Dai et al. (2013) también utilizan este índice en conjunto con otros dos, el de TAIEX de Taiwan (Taiwan Stock Exchange Capitalization Weighed Stock Index) y el de HSI de China (Hang Seng Index). Tay y Cao (2001) realizan su experimento sobre el índice de Standard & Poor 500 (CME-SP) y del gobierno francés (MATIF-CAC40) en conjunto con algunos series de tiempo de bonos del gobierno de Estados Unidos y de Alemania.

Por otra parte, Choudhry y Garg (2008) realizan su experimento prediciendo los precios de acciones individuales de tres empresas de tecnología de la India.

De igual forma, hay distintos métodos de medir los experimentos de pronóstico de precios, puesto que se puede optar por predecir el precio de las acciones cómo un número mediante Support Vector Regression, regresión derivada de Support Vector Machine, o se puede escoger predecir la dirección del movimiento de los precios, si es que estos suben o bajan respecto al actual precio.

Das y Padhy (2012) optan por predecir los precios utilizando 3 indicadores estadísticos para medir la precisión de las predicciones. Estos tres indicadores son el NMSE (Normalized Mean Squared Error), MAE (Mean Absolute Error) y DS (Directional Symmetry). Los dos primeros son la medida de la desviación de los valores predichos respecto a los valores actuales. DS es el porcentaje de aciertos en dirección de la predicción relativo los valores reales. Tay y Cao (2001) además de las tres medidas utilizan WDS (Weighed Directional Symmetry) que mide la magnitud del error en dirección y número, penalizando incorrectas predicciones.

Dai et al. (2013) utilizan otras medidas estadísticas, siendo 6 indicadores los empleados, que son distintos a los anteriores. Utilizan RMSE (Root Mean Squared Error), MAD (Mean Absolute Difference) y MAPE (Mean Absolute Percentage Error) para medir la desviación entre los valores de las predicciones y los valores actuales. Otro indicador, DA (Directional Accuracy), mide la exactitud de la dirección del movimiento del precio mientras que los últimos dos, CP (Correct Up Trend) y CD (Correct Down Trend), miden en cambio la tendencia que sigue la serie de tiempo.

Otros optan por predecir la dirección del movimiento mediante Support Vector Machine, puesto que sostienen que los inversionistas sólo requieren de la dirección del precio para tomar la decisión de vender o comprar acciones (Choudhry y Garg, 2008) y que la compraventa guiada por una predicción de dirección con mayor precisión es más rentable que una guiada por predicciones de precios con menor precisión (Huang, Nakamori y Wang, 2005).

Ambos artículos utilizan el “Hit ratio” para medir la precisión de las predicciones en términos de porcentajes. Esto se refiere al porcentaje de veces que el modelo de Support

Vector Machine estuvo acertado y tuvo éxito en predecir la dirección del movimiento de los precios de las acciones.

Todos los artículos mencionados previamente implementan Support Vector Machine como técnica de predicción para los experimentos y para tener una referencia repiten el experimento con otros modelos para tener un punto de comparación. Algunos comparan el rendimiento de Support Vector Machine con Redes Neuronales (Guajardo et al., 2006; Dai et al., 2013; Tay y Cao, 2001; Das y Padhy, 2012; Huang et al., 2005).

Otros estudios también comparan el modelo de Support Vector Machine con modelos estadísticos previos como ARMAX (Guajardo et al., 2006), MARS (Dai et al., 2013) y modelos lineales como LDA (Linear Discriminant Analysis) y QDA (Quadratic Discriminant Analysis) (Huang et al., 2005).

Los experimentos de los artículos muestran que Support Vector Machine tiene significativamente mejores resultados con respecto a los otros modelos utilizados en los mismos escenarios. Por tanto, sugieren a otros investigadores escoger el modelo de Support Vector Machine, mejorando la selección de variables Choudhry y Garg (2008) y refinando el modelo con mayor detenimiento para tener mejores resultados. (Tay y Cao, 2001).

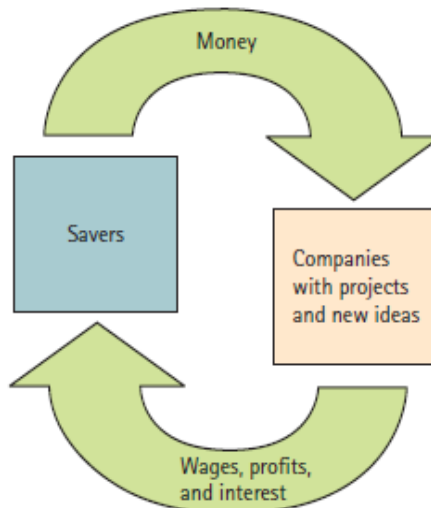
CAPÍTULO III: MARCO TEÓRICO

3.1. Mercados de Valores

Hoy en día todas las personas tienen contacto con instituciones financieras con frecuencia. Estas instituciones son entidades que dan servicios financieros como recibir depósitos, realizar préstamos o manejar inversiones (Berk, DeMarzo y Harford, 2009, p.17).

Las instituciones financieras participan de un ciclo financiero ilustrado en la figura 3.1. Las personas guardan e invierten su dinero, y este dinero, por medio de préstamos y acciones, se dirige a empresas que lo utilizan para crecer mediante nuevos proyectos o productos y luego retorna a los ahorros de los inversionistas.

Figura 3. 1
Ciclo Financiero



Fuente: Berk, DeMarzo & Harford. (2009). *Fundamentals of Corporate Finance*. Prentice Hall; 2nd edition.

Muchos tipos de instituciones financieras escogen invertir en bonos y acciones como métodos de hacer crecer el dinero que reciben de las personas contribuyentes. Fondos mutuos, fondos de pensión y aseguradoras son algunos de ellos que buscan la máxima rentabilidad posible (Berk et al., 2009, p.18).

Los mercados de valores son intermediarios entre las instituciones financieras y las empresas. Son foros donde se realizan transacciones financieras entre emisores y compradores de títulos (Gitman y Joehnk, 2009, p. 33).

Se clasifican en dos tipos: mercados de dinero o mercados de capital. En los mercados de dinero son aquellos donde se adquieren títulos a corto plazo, con vencimientos menores a un año, mientras que en los mercados de capital se compran y venden títulos a largo plazo, con vencimientos mayores al año, como acciones y bonos. Por su parte, los mercados de capital se clasifican en mercados primarios o secundarios (Gitman y Joehnk, 2009, p. 33).

Un mercado primario se refiere a cuando las empresas venden nuevos títulos a los inversionistas y son ellas quienes se benefician directamente de la venta (Berk et al., 2009, p.14). Una salida a la bolsa por primera vez por parte de la empresa se llama IPO (Initial Public Offering). Generalmente son pequeñas empresas con expectativas de crecimiento acelerado y que requieren de mayor capital para su desarrollo (Gitman y Joehnk, 2009, p. 33).

Una vez que se han vendido los títulos por primera vez, éstos continúan siendo intercambiados en los mercados secundarios por los inversionistas sin beneficiar o involucrar de alguna forma la empresa que originalmente emitió los títulos (Berk et al., 2009, p.14).

Los mercados secundarios son donde se lleva a cabo la mayor parte de las transacciones que realizan los inversionistas. Se subdividen en dos tipos en referencia a la manera de negociar los títulos: Mercados de corredores y mercados de dealers. Los mercados de corredores son las bolsas de valores, dónde los compradores y vendedores de títulos se reúnen físicamente a intercambiarlos. La mayor bolsa de valores es la NYSE (New York Stock Exchange) (Gitman y Joehnk, 2009, p. 37).

En el mercado de dealers, por contraste, el vendedor y comprador no establecen contacto directamente, sino que las compras y ventas se ejecutan por agentes que crean mercados para comprar o vender los títulos. En cada transacción se encuentra involucrado el dealer, como comprador o como vendedor (Gitman y Joehnk, 2009, p. 38). El mayor mercado de dealers es NASDAQ (National Association of Securities Dealers Automated Quotation), siendo un sistema automatizado de cotización de títulos (Gitman y Joehnk, 2009, p. 41).

3.2. Acciones Ordinarias

Uno de los títulos comercializados son las acciones ordinarias. Éstas representan una participación en una corporación siendo una fracción de la propiedad de la empresa. Las acciones ordinarias son el instrumento más conocido de inversión luego de las viviendas y los instrumentos de corto plazo (Gitman y Joehnk, 2009, p. 8).

El rendimiento de las acciones proviene de dos fuentes, dividendos y ganancias de capital. Los dividendos son los pagos que realiza la corporación a los dueños de las acciones a partir de las ganancias de la empresa. Las ganancias de capital, en cambio, provienen de la venta de acciones compradas a un precio mayor que al que fueron adquiridas (Gitman y Joehnk, 2009, p. 8).

El rendimiento es “el nivel de beneficio producto de una inversión” (Gitman y Joehnk, 2009, p. 127). Es una variable de decisión para inversionistas, que permite comparar las ganancias de diferentes opciones de inversión.

El Holding Period Return o rendimiento periodo de tenencia es una medida utilizada para valuar una inversión durante un tiempo específico, un año o menos. Se calcula de la siguiente forma según Gitman y Joehnk (2009, p.17)

$$\text{Rendimiento} = \frac{\text{Ingresos corrientes} + \text{Ganancias de Capital}}{\text{Valor inicial de la inversión}} \quad (1)$$

Esto ofrece un método de comparación, que señala el retorno o eficiencia por cada unidad de dinero invertido. De esta forma, se pueden comparar inversiones de distintos tipos y diferentes tamaños.

No obstante, la propiedad que tienen los accionistas es en verdad una propiedad residual, lo que significa que la empresa debe atender a todas sus otras obligaciones antes de que los dueños de las acciones comunes reciban cualquier tipo de pago (Alexander, 2010, p .241).

Una de las principales ventajas de una corporación es la responsabilidad limitada. Cuando una sociedad anónima no cumple con sus obligaciones, no se puede obligar a los accionistas a pagarlas. Esto resulta en que, debido al incumplimiento, el valor de las acciones de tal empresa sea prácticamente insignificante. Como resultado, los accionistas pierden su inversión pero no más que ésta (Berk et al., 2009, p. 241).

La ventaja principal de poseer acciones son las oportunidades de rendimiento que ofrecen. A comparación con otros instrumentos financieros, tienen mayor rendimiento. Los bonos corporativos desde 1950 al 2005 tuvieron un rendimiento promedio de 6%, sólo alrededor de la mitad del rendimiento de las acciones. Si bien hay algunos años en que los rendimientos de los bonos superan a los de las acciones, por lo general ocurre lo contrario. Las acciones también son una protección ante la inflación, puesto que generalmente tienen rendimientos superiores a la tasa inflacionaria (Gitman y Joehnk, 2009, p. 229).

Otros beneficios de las acciones es que son de fácil compra y venta con moderados costos de transacción, siendo la información necesaria del mercado ampliamente difundida. El costo de una acción individual también se encuentra al alcance de la mayoría de inversionistas individuales, tan sólo de 50 o 60 dólares. Los bonos por lo general requieren inversiones superiores a los 1000 dólares (Gitman y Joehnk, 2009, p. 229).

Sin embargo, también poseen desventajas. Una de ellas es el sacrificio de ingresos corrientes. Muchos otros instrumentos pagan niveles más altos de ingresos corrientes con mayor certeza. Otra desventaja y quizás la de mayor importancia es el riesgo. Las acciones se encuentran expuestas a muchos tipos de riesgo que pueden influir negativamente en sus ganancias, dividendos, apreciación y como resultado en la tasa de rendimiento del inversionista. Esta desventaja a su vez resulta en otra desventaja: la difícil valoración de una acción. El futuro incierto de una acción hace que sea difícil seleccionar consistentemente las mejores acciones (Gitman y Joehnk, 2009, p. 229).

Los valores de una acción se pueden describir de diferentes formas, los siguientes términos son definidos por Gitman y Joehnk (p. 236) son:

1. Valor Nominal: Se refiere al valor declarado, utilizado casi exclusivamente para valores contables. No es utilizado para inversionistas y en la actualidad se emiten acciones con muy bajo valor nominal o incluso sin éste.
2. Valor en Libros: Representa el valor total del patrimonio de los accionistas en una empresa. Se puede obtener el valor en libro por acción si se divide el monto entre la cantidad de acciones existentes. Por lo general, casi todas las acciones tienen precios de mercado muy por arriba de los valores en libro.
3. Valor de Mercado: Es el precio de mercado vigente de una acción. Si se multiplica por el número de acciones en el mercado se obtiene la capitalización de mercado de la empresa.
4. Valor de inversión: Es el valor que los accionistas dan a la acción. Representa el precio máximo al que estarían dispuestos a pagar para obtener el título.

3.3. Análisis Fundamental

El análisis fundamental se basa en estudiar la situación financiera de una empresa para comprender mejor el precio de sus acciones, puesto que proviene del supuesto que el precio de la acción es influenciado directamente por el desempeño de la empresa a la que pertenece. Si la empresa tiene un buen desempeño, entonces las acciones subirán de precio y, de lo contrario, bajarán (Gitman y Joehnk, 2009, p. 283).

El análisis fundamental utiliza los estados financieros de la empresa para estudiar las razones financieras, que son indicadores para evaluar dichos estados financieros. Tiene el motivo de comprender el pasado de la empresa para tratar de entender y pronosticar su futuro con cierto grado de confianza. Hay 5 grupos de razones financieras (Gitman y Joehnk, 2009, p. 288-289).

1. Liquidez: Se relaciona con la capacidad de la empresa para cumplir con las obligaciones que tiene a corto plazo.
 - a. Razón Corriente:

$$\text{Razón Corriente} = \frac{\text{Activos Corrientes}}{\text{Pasivos Corrientes}} \quad (2)$$

Una medida superior a la unidad refleja que la empresa tiene un buen nivel de activos líquidos.

b. Capital de Trabajo Neto:

$$\text{Capital de Trabajo Neto} = \text{Activos Corrientes} - \text{Pasivos Corrientes} \quad (3)$$

Es preferible tener cifras altas o crecientes.

2. Actividad: Comparan las ventas de la empresa para medir la eficiencia en cuanto al uso de sus activos.

a. Rotación de Cuentas por Cobrar:

$$\text{Rotación de CxC} = \frac{\text{Ventas Anuales}}{\text{Cuentas por Cobrar}} \quad (4)$$

Cuanto mayor sea la cifra de rotación más favorable es.

b. Rotación de Inventarios:

$$\text{Rotación de Inventarios} = \frac{\text{Ventas Anuales}}{\text{Inventarios}} \quad (5)$$

Cuanto más alta sea la cifra, menor es el tiempo que un artículo permanece en inventario y mejor el rendimiento de la empresa sobre sus fondos de inventario.

c. Rotación de Activos totales:

$$\star \text{ Rotación de Activos Totales} = \frac{\text{Ventas Anuales}}{\text{Total de Activos}} \quad \star (6)$$

Una cifra alta sugiere que los recursos se encuentran bien administrados.

3. Apalancamiento: Analiza la cantidad de deuda utilizada para respaldar los recursos financieros.

a. Relación Deuda-Capital:

$$\text{Relación Deuda Capital} = \frac{\text{Deuda a Largo Plazo}}{\text{Capital de Accionistas}} \quad (7)$$

Es preferible que sea baja, puesto que sugiere menor exposición a riesgo financiero

b. Cobertura de Interés:

$$\text{Cobertura de Interés} = \frac{\text{Utilidad antes de intereses e impuestos}}{\text{Gasto por intereses}} \quad (8)$$

Se buscan medidas altas, superiores a ocho o nueve unidades como resultado.

4. Medidas de Rendimiento: La medida relativa del éxito de la empresa, comparando utilidades con sus ventas, activos o capital propio.

a. Margen Utilidad Neta:

$$\text{Margen Utilidad Neta} = \frac{\text{Utilidad neta después de impuestos}}{\text{Ventas Totales}} \quad (9)$$

Indica la tasa de utilidad de la empresa, a mayor la cifra mejor la situación.

b. Rendimiento sobre los Activos:

$$\text{Rendimiento sobre Activos} = \frac{\text{Utilidad neta después de impuestos}}{\text{Activos Totales}} \quad (10)$$

Mayor la cifra, más rentable es la empresa.

c. Rendimiento sobre el Capital:

$$\text{Rendimiento sobre Capital} = \frac{\text{Utilidad neta después de impuestos}}{\text{Capital de Accionistas}} \quad (11)$$

Vigilado de cerca por inversionistas; mayor la cifra, mejor la situación.

5. Mercado: Indican qué parte de las utilidades, dividendos y capital propio se asigna a cada acción.

a. Dividendos por Acción:

$$\text{Dividendos por Acción} = \frac{\text{Dividendos anuales pagados}}{\text{Número de acciones ordinarias}} \quad (12)$$

Son los pagos que reciben los dueños de las acciones.

b. Razón de Pago:

$$\text{Razón de Pago} = \frac{\text{Dividendos por acción}}{\text{Utilidades por Acción}} \quad (13)$$

Las razones de pago altas pueden ser preocupantes.

Alexander (2010, p .257) utiliza 3 medidas para juzgar el crecimiento de una acción comparada con su valor, los valores se muestran en la figura 3.2. La primera es la razón valor en libros contra el valor de mercado y la segunda es la razón utilidades contra el precio.

1. Razón Valor en libros – Valor de Mercado (BV/MV – book-value-to-market-value)

Se calcula utilizando el valor de libros de las acciones de la empresa utilizando el último balance general publicado. Luego se determina la capitalización de la empresa usando el precio de mercado más reciente multiplicado por la cantidad de acciones en el mercado. Se divide la primera cifra entre la segunda cifra para encontrar el valor de BV/MV. El autor cita a un estudio realizado por Fama y French encontrando que entre mayor el tamaño de BV/MV mejores eran los rendimientos de las acciones.

$$\text{Razón Libros Mercado} = \frac{\text{Valor en Libros}}{\text{Valor de Mercado}} \quad (14)$$

2. Razón de Utilidades – Precio (E/P, earnings-to-price). Se calcula utilizando las utilidades después de impuestos de la empresa proveniente del estado de resultados más reciente y dividiéndola entre el número de acciones en circulación. Luego se obtiene el precio de mercado de una acción ordinaria de la empresa. Por último se divide la primera cifra entre la segunda para encontrar el valor de E/P. Se cita nuevamente a Fama y French que encuentran que a mayor la tasa de E/P, mayor son las tasas de rendimiento de las acciones.

$$\text{Razón Utilidades Precio} = \frac{\text{Utilidades}}{\text{Precio de Acción}} \quad (15)$$

3. Tamaño: Es la capitalización de mercado de la empresa. Fama y French también realizaron un experimento por tamaño y encontraron una relación inversa entre el tamaño y el rendimiento promedio. Se puede aludir a un efecto de “empresa pequeña”.

$$\text{Tamaño} = \text{Capitalización de Empresa} \quad (16)$$

Figura 3. 2

Rendimiento de Cartera de Acciones

Base de cartera	Número de cartera ^a											
	1	2	3	4	5	6	7	8	9	10	11	12
(a) Razón BV/MV												
Rendimiento	.30	.67	.87	.97	1.04	1.17	1.30	1.44	1.50	1.59	1.92	1.83
BV/MV	.11	.22	.34	.47	.60	.73	.87	1.03	1.23	1.52	1.93	2.77
(b) Razón E/P ^b												
Rendimiento	1.04	.93	.94	1.03	1.18	1.22	1.33	1.42	1.46	1.57	1.74	1.72
E/P	.01	.03	.05	.06	.08	.09	.11	.12	.14	.16	.20	.28
(c) Tamaño												
Rendimiento	1.64	1.16	1.29	1.24	1.25	1.29	1.17	1.07	1.10	.95	.88	.90
ln(MV) ^c	1.98	3.18	3.63	4.10	4.50	4.89	5.30	5.73	6.24	6.82	7.39	8.44

^aCarteras formadas con base en clasificaciones de medida financiera indicada para acciones de la más pequeña (1) a la más grande (12); el rendimiento indica, en porcentaje, el rendimiento mensual promedio.

^bLa cartera de acciones que tuvo rendimientos negativos tuvo un rendimiento mensual promedio de 1.46.

^cln(MV) indica el valor logarítmico del valor de mercado de las acciones promedio de la cartera (expresado en millones de dólares).

Fuente: Eugene F. Fama y Kenneth R. French, “The Cross-Section of Expected Stock Returns”, Journal of Finance 47, núm. 2 (junio de 1992), tablas 2 y 4 como se citó en Alexander, G. J. (2010). *Fundamentals of investments/ fundamentos de inversiones*. Place of publication not identified: Prentice Hall.

3.4. Análisis Técnico

Por otra parte el análisis técnico es el estudio de la información interna que existe en la bolsa de valores. Mientras que el análisis fundamental mira hacia adelante a largo plazo, el análisis técnico trata de predecir los movimientos a corto plazo y busca los momentos precisos para comprar o vender acciones (Alexander, 2010, p .286).

Los analistas técnicos sostienen que el análisis fundamental pueda que identifique acciones que se encuentran subvaluadas en el mercado, pero que esto se dé, no indica que las acciones no permanecerán mucho tiempo sin reflejar el valor que verdaderamente le corresponda (Sáez y Doblado, 2013, p .31).

También asumen que toda la información que influye de alguna forma en el mercado, sea esta de índole económica, política, sociológica, etc, se ve de alguna forma reflejada en el comportamiento de los precios de las acciones y por tanto que los precios ya incluyen toda la información existente (Sáez y Doblado, 2013, p .32).

El análisis técnico estudia los movimientos de las acciones desde los precios de los mismos, analizando la intensidad y dirección de sus fuerzas, tomando como referencia los patrones identificados de precio (Sáez y Doblado, 2013, p .32).

La existencia de las tendencias es abiertamente reconocida por todos los enfoques del mercado. Una tendencia alcista es definida por el crecimiento de los precios en una serie de mínimos y máximos. Lo máximos solo confirman la tendencia o el agotamiento temporal, mientras que los mínimos son los que indican los posibles cambios de las tendencias. En una tendencia bajista ocurre lo contrario (Sáez y Doblado, 2013, p .35).

Los promedios móviles son útiles para identificar tendencias en los precios y para identificar cambios de tendencias lo más temprano posible. El promedio móvil simple es el más utilizado. Por ejemplo, para obtener un promedio móvil de 5 días, se suman los últimos 5 precios de cierre y se divide entre 5. Cada nuevo día se añade el nuevo precio de cierre y se descarta el más antiguo.

Se toma el ejemplo de Credit Suisse (s.f.), una empresa proveniente de Suiza que ofrece servicios financieros, que utiliza 3 promedios móviles. De corto plazo de 21 días, de mediano plazo de 55 días y de largo plazo de 144 días, mostrado en la figura 3.3.

Figura 3. 3

Promedios Móviles



Fuente: Credit Suisse (s.f.) *Technical Analysis - Explained*

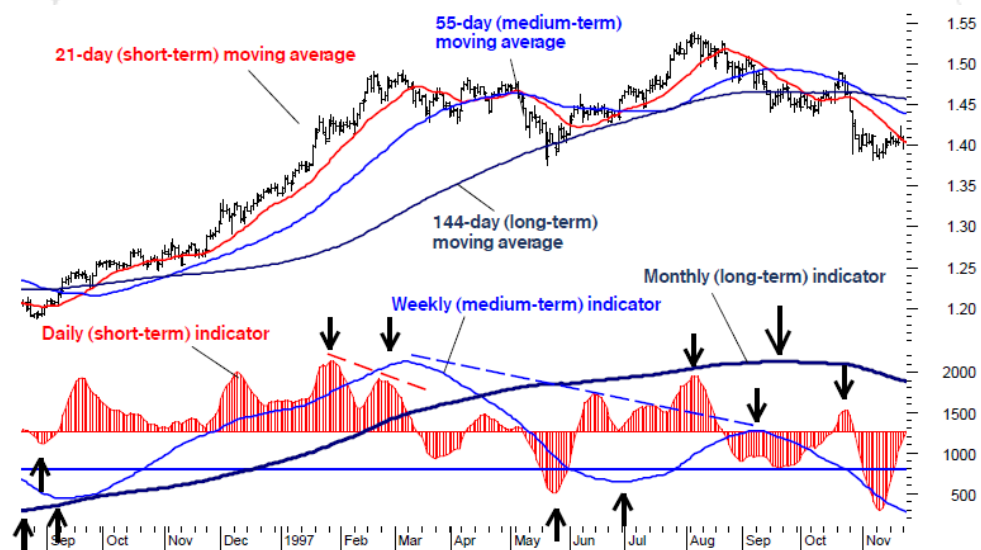
1. Corto plazo: Una señal de compra se da cuando el precio se incrementa sobre la línea roja del promedio móvil de 21 días. Una señal de venta es en la dirección opuesta.
2. Mediano plazo: Una señal de compra se da cuando el precio se incrementa sobre la línea azul del promedio móvil de 55 días y se confirma cuando la línea roja del promedio móvil de 21 días también lo supera. Una señal de venta es en la dirección opuesta.
3. Largo plazo: Una señal de compra se da cuando el precio se incrementa sobre la línea negra del promedio móvil de 144 días y se confirma cuando la línea azul del promedio móvil de 55 días también lo supera. Una señal de venta es en la dirección opuesta.

El mejor desempeño es cuando la línea roja del promedio móvil de 21 días se encuentra sobre la línea azul del promedio móvil de 55 días que se encuentra sobre la línea negra del promedio móvil de 144 días.

Otro indicador utilizado por la empresa es el momentum o ímpetu. En lugar de calcular la suma, se calcula la diferencia entre los precios. Por ejemplo para 5 días, se calcula la diferencia del precio de cierre y el precio de cierre de 5 días atrás.

Este indicador muestra la aceleración o desaceleración de la variación de precios. Se utilizan 3 indicadores nuevamente, mostrados en la figura 3.4: a corto plazo de 10 días, a mediano plazo de 10 semanas y a largo plazo de 10 meses. Los inversionistas deben comprar acciones cuando el momentum se encuentra en lo más bajo y cuando el precio sobrepasa el promedio móvil. Deben vender, en cambio, cuando el momentum empieza a bajar y el precio cae debajo del promedio móvil.

Figura 3. 4
Promedio Móviles y Momentum



Fuente: Credit Suisse (s.f.) *Technical Analysis - Explained*

La combinación de los 6 indicadores muestra el camino más probable que seguirán los precios de las acciones. El escenario más positivo es cuando el precio se encuentra sobre el promedio móvil de 10 días, que se encuentra sobre el promedio móvil de 55 días que a su vez se encuentra sobre el promedio móvil de 155 días y al mismo tiempo todos los indicadores de momentum incrementan. El escenario más negativo es cuando ocurre lo opuesto.

3.5. Standard and Poor Dow Jones Indexes

S&P Indexes y Dow Jones Indexes se unieron en el año 2012 para formar el proveedor más grande del mundo de índices de mercados financieros. Posee muchos índices de mercado entre ellos S&P 500 que fue lanzando en 1957. (S&P Dow Jones Indices S&P, 2017)

S&P 500 es considerado como el mejor visualizador de las capitalizaciones de mercado de empresas en Estados Unidos. El índice captura 500 compañías líderes de las bolsas de NYSE o NASDAQ y alrededor de 80% de las capitalizaciones de mercado. Es la base de muchos instrumentos financieros. (S&P Dow Jones Indices S&P, 2017)

El índice de mercado se calcula cada día hábil utilizando el valor de cierre de cada acción del índice en la siguiente ecuación (Gitman y Joehnk, 2009):

$$\text{Índice S\&P} = \frac{\text{Valor de cierre de acción 1} + \text{Valor de cierre de acción 2} + \dots + \text{Valor de cierre de última acción}}{\text{Valor de cierre de acción 1 en periodo base} + \text{Valor de cierre de acción 2 en periodo base} + \dots + \text{Valor de cierre de última acción en periodo base}} * 10 \quad (17)$$

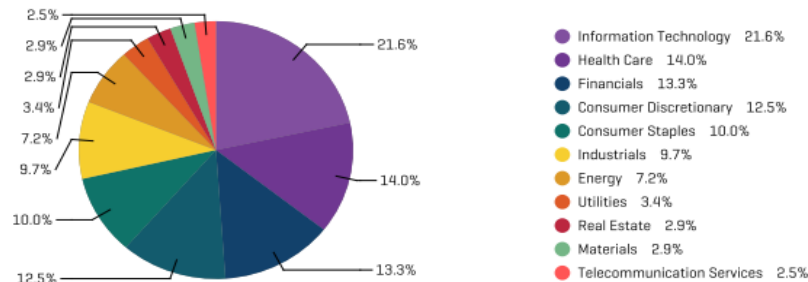
El periodo base utilizado es de 1941-1943. Un índice de 1258.60 como el 12 de julio del 2006 significa que los valores de las acciones que componen el índice aumentaron 125.860 veces su valor con relación al periodo 1941-1943

La figura 3.6 muestra la composición de las 500 empresas consideradas en el índice S&P 500 según el sector en el que participan.

Figura 3. 5

Sectores de Empresas de S&P

Sector* Breakdown



*Based on GICS® sectors

The weightings for each sector of the index are rounded to the nearest tenth of a percent; therefore, the aggregate weights for the index may not equal 100%.

Fuente: S&P Dow Jones Indices. Consultado el 10 de Abril de 2017 de: <http://us.spindices.com/indices/equity/sp-500>

Se utiliza S&P debido a que el análisis técnico requiere un mercado de alta participación, porque el precio es un reflejo de la voluntad de los participantes. Un mercado con pocos participantes es volátil, puesto que movimientos de socios mayoritarios de empresas tienen gran influencia en los precios. En este tipo de mercados se puede pasar de rentabilidades de bajas a altas y viceversa con solo algunos títulos vendidos (Sáez y Doblado, 2013, p.107).

3.6. Minería de Datos

La minería de datos es un proceso que consiste en descubrir información útil dentro de grandes cantidades de datos. Estas técnicas sirven para encontrar patrones en los datos que de otra forma pasen desapercibidos. Dan también la posibilidad de poder pronosticar el resultado de futuras observaciones basado en los datos anteriores (Tan et al., 2005, p.2).

Para Han, Kamber y Pei (2013), la minería de datos es un proceso más largo que debería llamarse en realidad descubrimiento de conocimiento de los datos. Parte del proceso lo compone los siguientes pasos:

1. Limpieza de datos: para eliminar el ruido y los outsiders
2. Integración de datos: combinando múltiples fuentes
3. Selección de datos: Obtener datos relevante de la fuente
4. Transformación de datos: Se transforma mediante operación de agregación
5. Minería de datos: Aplicación de métodos inteligentes para obtener patrones
6. Evaluación de patrones: Identificar patrones de interés en los datos
7. Presentación de Conocimiento: Se presentan los descubrimientos resultados del proceso.

Los cuatro primeros pasos son parte del preprocesamiento de datos. Usualmente el término de minería de datos se utiliza para denominar todo este proceso. Entonces, en una vista más amplia, la minería de datos es el proceso de descubrir patrones de interés y conocimiento de grandes cantidades de datos (Han et al., 2013, p. 8).

3.7. Support Vector Machine

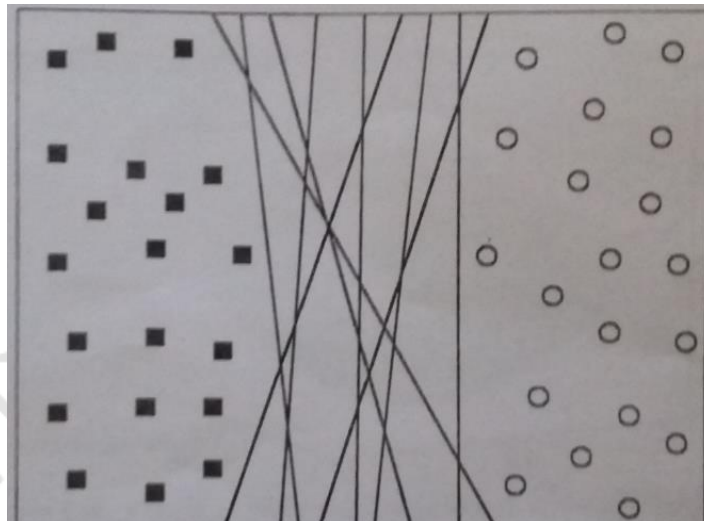
Support Vector Machine es una técnica de clasificación alternativa a las redes neuronales. Su actual versión fue publicada por Vladimir Vapnik y Corinna Cortes (1995) como nuevo algoritmo de machine learning para problemas de clasificación binaria.

El objetivo es obtener un plano bidimensional que separe a dos clases de datos y posteriormente utilizarlo para clasificar nuevos datos basada en su posición respecto al plano.

Dos problemas surgen del enfoque anterior: El primero es que no todos los planos generalizan óptimamente, es decir, no todos los planos que separan los datos de entrenamiento podrán separar correctamente los datos de prueba y datos reales, como lo muestra la figura 3.7. El segundo problema es separar datos con gran número de dimensiones (Cortes y Vapnik, 1995).

Figura 3. 6

Datos separados por Múltiples Planos

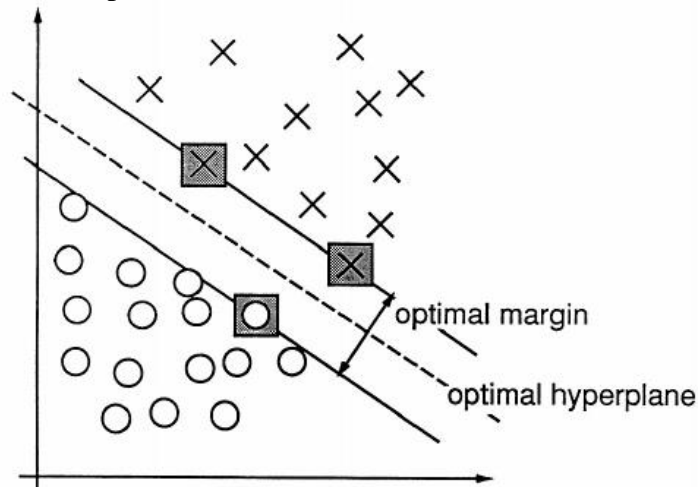


Fuente: Tan, P.-N., Steinbach, M., & Kumar, V. (2015). *Introduction to data mining*. Dorling Kindersley: Pearson.

La solución que dieron Cortes y Vapnik (1995) es encontrar un hiperplano óptimo que maximice la distancia entre los vectores de las dos clases, como lo muestra la figura 3.8. Para construir estos planos solo se requiere de una pequeña cantidad de datos de entrenamiento, llamada vectores de soporte.

Figura 3. 7

Hiperplano Óptimo de Separación



Fuente: Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297..

El plano se encuentra definido por:

$$w \cdot x + b = 0 \quad (18)$$

w y b son parámetros del modelo y x corresponde al conjunto de atributos de un punto de datos.

La ecuación de decisión entonces se encuentra dada por:

$$w \cdot x_i + b \quad (19)$$

Si el resultado es positivo los datos serán de un tipo y si el resultado es negativo será del otro tipo.

Esto a su vez resulta en que los vectores de soporte (los vectores ubicados en el margen óptimo) se encuentran definidos por:

$$y_i(w \cdot x_i + b) - 1 = 0 \quad (20)$$

y_i es +1 para valores positivos y -1 para valores negativos.

El valor del margen se encuentra definido por la ecuación:

$$\frac{2}{\|w\|} \quad (21)$$

Como el objetivo es maximizar el margen, entonces minimizar la siguiente ecuación da el mismo resultado:

$$\frac{\|w\|^2}{2} \quad (22)$$

Aplicando multiplicadores de Lagrange se obtiene la equivalencia de

$$w = \sum_i \alpha_i y_i x_i \quad (22)$$

Y también se obtiene la función a minimizar:

$$L_D = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \quad (23)$$

Los α son los multiplicadores de Lagrange.

Esta ecuación muestra que la maximización del margen depende de los productos punto de pares de ejemplos de datos.

Reemplazando la equivalencia en la ecuación de decisión se obtiene:

$$\sum_i \alpha_i y_i u_i \cdot x_i + b = 0 \quad (24)$$

Esta ecuación también muestra que la clasificación depende solo de los productos punto de pares de ejemplos de datos (Tan et al., 2005, p.264).

3.8. Kernels

Para casos en que los datos no sean linealmente separables, es decir, que los datos puestos en un plano no puedan ser separados por una línea en las dimensiones originales, se aplican transformaciones de tipo $\phi(x_i)$ que permiten trabajarlos en otra perspectiva y hacen posible una separación lineal.

Como la maximización de las ecuaciones tan solo depende de los productos punto de los vectores entonces se tiene que una transformación de tipo

$$K(x_i, y_i) = \phi(x_i) \cdot \phi(y_i) \quad (25)$$

La función se denomina función Kernel, y para resolverlo solo se requiere la transformación de los productos puntos y no de la ecuación completa

Según Andrew Ng (s.f), para distinguir entre los dos grupos de variables antes y luego de las transformaciones, se denominan atributos (attributes) a las variables de entrada originales mientras que a las variables transformadas se denominan características (features).

La función que transforma a los atributos en características se denomina Kernel o ϕ . Entonces se aplica la función a todos los atributos de un vector. Como el algoritmo de minimización depende de solamente del producto punto de dos vectores $\langle x, y \rangle$, entonces se reemplaza esos productos punto por los productos punto de los vectores transformados $\langle \phi(x), \phi(z) \rangle$

Si bien la función $\phi(x)$ es computacionalmente costosa de calcular puesto que $\phi(x)$ es un vector con gran cantidad de dimensiones, el producto punto $\langle \phi(x), \phi(z) \rangle$ es

computacionalmente poco costoso de calcular. Support Vector Machine entonces calcula los productos puntos sin tener que calcular los vectores directamente. Se resuelve entonces el segundo problema planteado por Vapnik sobre la complejidad del cálculo en mayores dimensiones.

Las características de las variables se determinan utilizando la función del kernel que mide la “cercanía” de un punto de datos con respecto a todos los demás puntos de datos. Las “distancias” entonces son arregladas en un vector de características que luego es utilizado como variables en el algoritmo de Support Vector Machine en lugar de los atributos iniciales.

Como resultado se tiene soluciones a problemas que originalmente no eran linealmente separables.

Existen varios tipos de kernels siendo los más populares el lineal y el exponencial.

Kernel Lineal o Polinomial:

$$K(x_i, y_i) = (x \cdot y + 1)^p \quad (26)$$

Kernel Radial o Exponencial

$$K(x_i, y_i) = e^{-\frac{\|x-y\|^2}{2\sigma^2}} \quad (27)$$

Kernel Sigmoidal

$$K(x_i, y_i) = \tanh(kx_i \cdot y_i - \delta) \quad (28)$$

Cada uno de los kernels resulta en una transformación diferente y no existe realmente una regla para determinar cuál de ellos produce el SVM o SVR de mejor precisión. Asimismo, comenta que en la práctica, la utilización de un tipo u otro tipo de kernel no tiene un gran impacto en los resultados de las predicciones del modelo

Hsu, Chang, y Lin (2003) comparan el kernel radial con el kernel lineal. El kernel radial o RBF kernel transforma los datos a un espacio de mayor cantidad de dimensiones, por lo que puede manejar casos en que las variables y los datos no se encuentran linealmente relacionadas. El RBF kernel también tiene menor cantidad de parámetros por lo que resulta en una primera elección por su mayor simplicidad. La última razón es que el resultado del

kernel va de 0 a 1 por lo que tiene menor cantidad de dificultades en el momento del cálculo, puesto que el kernel lineal puede tener valores hasta el infinito.

El kernel lineal, en cambio se utiliza cuando se tiene una mucha mayor cantidad de variables de entradas que de datos de entrenamiento o de prueba. Un segundo caso es cuando tanto la cantidad de datos y de variables son de gran magnitud. Un último caso especial, porque se recomienda el uso de kernels no lineales, pero también se puede utilizar el kernel lineal con un parámetro especial de grado 2 es cuando el número de datos es mucho mayor que el número de variables de entrada

Aplicando Kernels a la ecuación de SVM, se obtiene la siguiente ecuación (Tan et al., 2005, p. 274):

$$f(x) = \text{signo} \left(\sum_i \alpha_i y_i K(u_i, x_i) + b \right) \quad (30)$$

3.9. Support Vector Regression

Es una técnica basada en Support Vector Machine pero que en lugar de ser de tipo clasificación binaria, es utilizada para la regresión. Ha tenido exitosos rendimientos en este campo y en series de tiempo (Smola y Schölkopf, 2004).

En regresión de tipo ε - SV, la meta es encontrar una función de los Support Vectores (SV) teniendo como máximo una desviación ε con respecto a los verdaderos valores y al mismo tiempo es tan plana como posible, en el sentido de menor sobreajuste. En otras palabras, se aceptan errores de predicción pero no más que ε . Esto es una función de pérdida basada en ε . La siguiente ecuación determina los dos casos posibles, en que el resultado de la regresión es menor o mayor al valor real y_i :

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \quad (31)$$

Se introducen también variables de holgura para cumplir con casos que se tengan desviaciones mayores a ε . Por lo que se tiene como ecuación:

$$\min \sum_{i=1}^n \varepsilon_i + \sum_{i=1}^n \varepsilon'_i$$

$$\text{sujeto a } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \varepsilon_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \varepsilon'_i \\ \varepsilon_i, \varepsilon'_i \geq 0 \end{cases} \quad (32)$$

Se modifica la función para obtener una similar a SVM y poder aplicar los mismos conceptos utilizados en casos no linealmente separables. Se agrega $\frac{1}{2} \|w\|^2$ para obtener una función plana y evitar los sobreajustes a los datos.

$$\text{minimizar } \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^n \varepsilon_i + \sum_{i=1}^n \varepsilon'_i \right)$$

$$\begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \varepsilon_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \varepsilon'_i \\ \varepsilon_i, \varepsilon'_i \geq 0 \end{cases} \quad (33)$$

Luego de aplicar los multiplicadores de Lagrange y realizar algunas transformaciones en las ecuaciones, se deriva a una ecuación de optimización:

$$\max \begin{cases} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)(x_i, x_j) \\ -\varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \end{cases} \quad (34)$$

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*), \alpha_i, \alpha_i^* \in [0, C]$$

El parámetro b se obtiene con las siguientes ecuaciones mediante las condiciones de Kuhn-Tucker que sostiene que en la solución las variables duales y las condiciones deben desaparecer:

Se debe entender que $\alpha_i \alpha_i^* = 0$ y que ambos son 0 para datos dónde el error es menor que ε . Entonces b es finalmente una combinación lineal de algunos ejemplos de los datos multiplicado por los coeficientes de Lagrange.

Por tanto en todas las ecuaciones, por ser combinaciones lineales, se puede utilizar los Kernels de la siguiente forma:

$$K(X, X') = \varphi(X)^T \varphi(X') \quad (35)$$

Se tiene entonces como resultado la ecuación de optimización de Support Vector Regression con un Kernel aplicado (Smola y Schölkopf, 2004):

$$\max \begin{cases} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \\ -\varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \end{cases} \quad (36)$$

$$\sum_{i=1}^n (\alpha_i - \alpha_i^*), \alpha_i, \alpha_i^* \in [0, C]$$

CAPÍTULO IV: PROPUESTA DE SOLUCIÓN

4.1. Métodos de investigación

Para validar lo propuesto en este trabajo de investigación, se llevará a cabo un experimento de pronóstico de precios de acciones utilizando modelos basados en el algoritmo de Support Vector Regression y de Support Vector Machine utilizando las variables determinadas en el marco teórico.

Se obtendrá datos de algunas empresas pertenecientes a Standard & Poor 500 en relación a los estados financieros y también información de las cotizaciones de sus acciones.

Posteriormente, se utilizarán datos para, mediante transformaciones, obtener los indicadores que se utilizarán luego en los modelos basados en Support Vector Regression y Support Vector Machine.

Se procederá después a una fase de optimización de los modelos obtenidos, eligiendo los parámetros para poder obtener mejores predicciones.

Para evaluar los resultados de los modelos, se utilizarán indicadores estadísticos como MAE (Mean Average Error / Promedio de Error), MAPE (Mean Average Percentage Error / Promedio Porcentual de Error), MSE (Mean Squared Error / Promedio de Error Cuadrado) y DA (Directional Accuracy / Precisión de Dirección), que han sido usados por varios autores para evaluar similares casos de pronóstico de precios de acciones. (Wang et al., 2016; Dai et al., 2013).

También se utilizarán gráficos de comparación para ver la variación entre los precios pronosticados y los precios reales.

Se realizará también comparaciones entre los propios modelos de predicción obtenidos.

Finalmente, se mostrarán las conclusiones del trabajo de investigación y recomendaciones para futuros investigadores en este campo.

4.2. Alcance

- Se realizará un experimento con algunas empresas del mercado de Standard and Poor
- Se trabajará con acciones individuales.
- Se utilizará Support Vector Regression y Support Vector Machine para realizar el pronóstico de precios

4.3. Riesgos

El riesgo principal del proyecto es no poder encontrar un modelo adecuado para el pronóstico satisfactorio de los precios de acciones.

De no obtener predicciones que cumplan con las expectativas se debe evaluar las diferentes opciones que se tienen:

- Utilización de distintos kernels
- Utilización de diferentes parámetros del algoritmo
- Utilización de un sub-grupo de variables.

4.4. Entregables

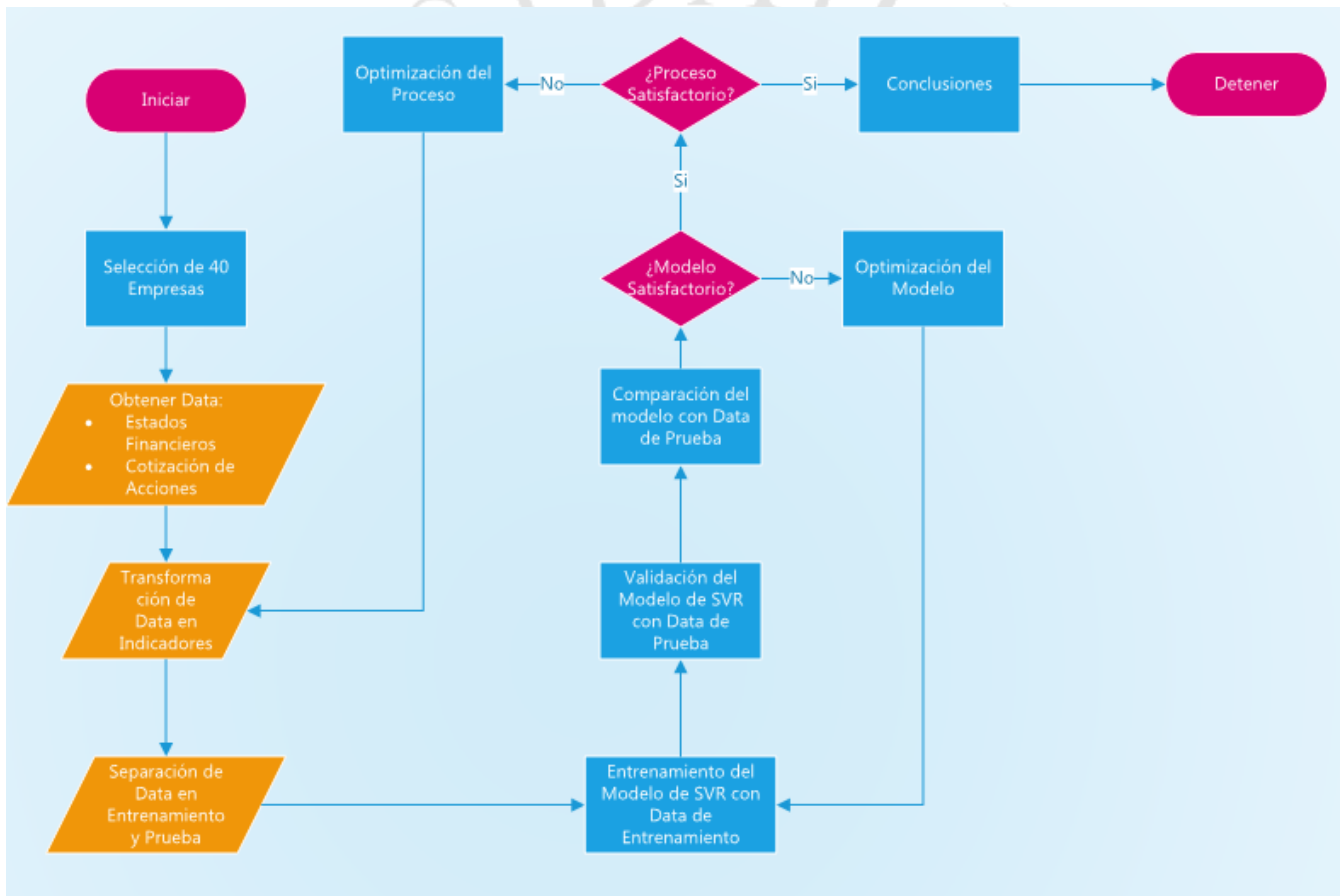
- Datos de acciones de empresas utilizada
- Algoritmo de Support Vector Regression evaluado.
- Comparación de entre Support Vector Regression y Support Vector Machine

CAPÍTULO V: PLAN DE TRABAJO

En este capítulo se presenta la metodología de desarrollo utilizado en la investigación.

Figura 5. 1

Flujograma de Trabajo



Fuente: Elaboración Propia.

Para el desarrollo de la investigación se utilizó datos financieros correspondientes a empresas y sus acciones. Se obtuvo datos de 40 empresas catalogadas en el índice de Standard & Poor 500, de las bolsas de NASDAQ o NYSE. Los datos correspondieron a los estados financieros y cotizaciones de acciones de los años 2011 a 2015. La fuente de la información fue Economatic del Laboratorio de Capitales de la Universidad de Lima.

Se realizó una transformación de cada dato para obtener los indicadores que fueron explicados en el marco teórico dentro de la investigación, correspondiente al análisis fundamental y análisis técnico. Las fechas de las cotizaciones de las acciones fueron escogidas de acuerdo a la fecha de publicación del estado financiero correspondiente. Es decir, cada dato corresponde a un estado financiero entre los años contemplados, obteniéndose los indicadores del mismo, así como los indicadores de las cotizaciones de sus acciones que corresponden a esa fecha, agregándole como dato a pronosticar el precio de la acción en la fecha del siguiente estado financiero.

Los indicadores utilizados y sus componentes para cada dato son los siguientes:

Tabla 5. 1

Indicadores de Acciones

Precio de Acción	
Precio de Acción	<i>Precio de cierre del día</i>
Indicadores de Análisis Técnico	
Promedio Móvil de 21 Días	<i>Promedio 21 últimos días</i>
Promedio Móvil de 55 Días	<i>Promedio 55 últimos días</i>
Promedio Móvil de 144 Días	<i>Promedio 300 últimos días</i>
Momentum de 10 Días	<i>Precio de Acción – Precio de Acción 10 días antes</i>
Momentum de 70 Días	<i>Precio de Acción – Precio de Acción 70 días antes</i>
Momentum de 300 Días	<i>Precio de Acción – Precio de Acción 300 días antes</i>
Indicadores de Análisis Fundamental	
Razón Corriente	$\frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$
Capital de Trabajo	<i>Activo Corriente – Pasivo Corriente</i>
Rotación de Cuentas por Cobrar	$\frac{\text{Ventas}}{\text{Cuentas por cobrar}}$
Rotación de Inventarios	$\frac{\text{Ventas}}{\text{Inventarios}}$
Rotación de Activos Totales	$\frac{\text{Ventas}}{\text{Total de Activos}}$

Relación Deuda Capital	$\frac{\text{Deuda a Largo Plazo}}{\text{Capital Accionistas}}$
Cobertura de Interés	$\frac{\text{Utilidad Antes de Impuestos}}{\text{Gastos por Interés}}$
Margen de Utilidad Neta	$\frac{\text{Utilidad Neta}}{\text{Ventas}}$
Rendimiento sobre los Activos	$\frac{\text{Utilidad Neta}}{\text{Activos Totales}}$
Rendimiento sobre el Capital	$\frac{\text{Utilidad Neta}}{\text{Capital de accionistas}}$
Dividendos por acción	$\frac{\text{Dividendos pagados}}{\text{Número de Acciones Ordinarias}}$
Razón de pago	$\frac{\text{Dividendos por Acción}}{\text{Utilidades por Acción}}$
Razón Valor en libros – Valor de Mercado	$\frac{\text{Valor de acciones en Libros}}{\text{Capitalización de Empresa}}$
Razón de Utilidades	$\frac{\text{Utilidad Neta}}{\text{Número de Acciones}}$
Tamaño	$\text{Capitalización de Empresa}$

Fuente: Elaboración Propia.

Las acciones son de gran variedad de precios y los indicadores de diferentes escalas. Por tanto, se lleva a cabo una normalización que intenta dar a todos los atributos una misma importancia independiente de su tamaño original. (Han, Kamber y Pei, 2013, p. 113) La normalización consiste en transformar los datos a un rango de menor tamaño mediante el método de min-max, mediante la siguiente ecuación, donde v es el dato real y v' es el dato normalizado hacia una escala de 1 a 10:

$$v' = \frac{v - \min}{\max - \min} (10 - 1) + 1 \quad (37)$$

Luego, se arman los datos en un formato especial para ser utilizado por el algoritmo de Support Vector Regression, donde el primer valor es el valor a pronosticar y el resto son los atributos que corresponden al dato. El formato es el siguiente:

$\langle \text{Valor a Pronosticar} \rangle$ 1: $\langle \text{Atributo 1} \rangle$ 2: $\langle \text{Atributo 2} \rangle$... n: $\langle \text{Atributo n} \rangle$

Mediante un programa de elaboración propia, se divide los datos en datos de Entrenamiento y datos de Prueba, con un porcentaje de 60% a 40%, 60% siendo datos de entrenamiento del modelo y el resto para validación del mismo.

El algoritmo que sigue el programa es el siguiente:

- Entrada: Lista 'DatosEntrenamiento': una lista que contiene los 800 datos recolectados.
- Salidas son dos listas: DatosEntrenamiento y DatosPrueba, con 60% y 40% de los datos respectivamente.

Algoritmo 1

Algoritmo de Separación de Datos

```
1   Datos(DatosEntrenamiento)
2   C ← Cantidad(DatosEntrenamiento)
3   T ← C*0.6
4   Para i desde 0 hasta T-1
5       F← Aleatorio (0 a Cantidad(DatosEntrenamiento))
6       DatosPrueba ← DatosEntrenamiento(F)
7       Remove(DatosEntrenamiento(F))
8   {Fin Ciclo}
```

Fuente: Elaboración Propia.

Luego de obtener ambas listas, se procede a entrenar el modelo de predicción. El algoritmo de Support Vector Regression es una técnica de regresión derivada de Support Vector Machine que tiene el objetivo de hallar un plano óptimo de separación entre datos usado para posterior predicción de nuevos datos. Los kernels son utilizados para poder trabajar con datos que no es linealmente separable

La Ecuación del Support Vector Regression, explicada en el Marco Teórico, es:

$$\begin{aligned}
 & \text{Max} \begin{cases} -\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)K(x_i, x_j) \\ -\varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + \sum_{i=1}^n y_i(\alpha_i - \alpha_i^*) \end{cases} \\
 & \text{sujeto a } \sum_{i=1}^n (\alpha_i - \alpha_i^*), \alpha_i, \alpha_i^* \in [0, C]
 \end{aligned} \tag{38}$$

Para el algoritmo de Support Vector Regression y Support Vector Machine, se utilizará la librería de SVMLight, recomendada por Tan et al. (2005), calificada como una de las más robustas para usos avanzados de las SVM.

Se utilizan datos de entrenamiento con el algoritmo para obtener un modelo de predicción. Luego se utiliza este modelo con datos de prueba para obtener predicciones de precios y se procede a comparar estos pronósticos con los resultados reales.

Para evaluar los resultados del algoritmo, se utilizaron indicadores estadísticos como MAE (Mean Average Error / Promedio de Error), MAPE (Mean Average Percentage Error / Promedio Porcentual de Error), MSE (Mean Squared Error / Promedio de Error Cuadrado) y DA (Directional Accuracy / Precisión de Dirección) que han sido usados por varios autores para evaluar similares casos de pronóstico de precios de acciones.

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{39}$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \frac{|\hat{y}_i - y_i|}{y_i} \tag{40}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \tag{41}$$

$$DA = \frac{100}{n} \sum_{i=1}^n d_i, \text{ donde} \quad (42)$$

$$d_i \begin{cases} 1 & (\hat{y}_i - x_i)(y_i - x_i) \geq 0 \\ 0 & (\hat{y}_i - x_i)(y_i - x_i) < 0 \end{cases}, \text{ donde } x \text{ es el Precio Inicial}$$

También se utilizarán gráficos de comparación para ver la variación de los precios pronosticados y sus verdaderos valores.

Luego se procede a una iteración de entrenamiento del modelo con diferentes parámetros para buscar uno de mejores resultados.

Finalmente, se realiza una comparación de la medida del movimiento de los precios (DA) con modelos basados en Support Vector Machine, explicada en el Marco Teórico, para ilustrar las diferencias, siendo su ecuación con kernel aplicado:

$$f(x) = \text{sign} \left(\sum_i \alpha_i y_i K(u_i, x_i) + b \right) \quad (43)$$

CAPÍTULO VI: PRUEBAS Y RESULTADOS

En este capítulo se describen los resultados obtenidos por el modelo de predicción de precios.

Se desarrollaron cuatro modelos de Support Vector Regression basados en distintos grupos de indicadores que fueron expuestos en el capítulo anterior para realizar predicciones.

Se utilizó los mismos datos de Entrenamiento y datos de Prueba para entrenar a los 4 modelos y posteriormente realizar las pruebas. Los datos varían en los indicadores que contiene, de acuerdo al modelo.

6.1. Análisis Técnico

Para empezar, se utilizó el algoritmo de Support Vector Regression utilizando los indicadores del análisis técnico adicionado al Precio de Cierre de la Acción.

Los indicadores utilizados para este modelo se muestran en la tabla 6.1:

Tabla 6. 1

Análisis Técnico: Indicadores

Precio de Acción	
Precio de Acción	<i>Precio de cierre del día</i>
Indicadores de Análisis Técnico	
Promedio Móvil de 21 Días	<i>Promedio 21 últimos días</i>
Promedio Móvil de 55 Días	<i>Promedio 55 últimos días</i>
Promedio Móvil de 144 Días	<i>Promedio 300 últimos días</i>
Momentum de 10 Días	<i>Precio de Acción – Precio de Acción 10 días antes</i>
Momentum de 70 Días	<i>Precio de Acción – Precio de Acción 70 días antes</i>
Momentum de 300 Días	<i>Precio de Acción – Precio de Acción 300 días antes</i>

Fuente: Elaboración Propia.

Los parámetros del mejor algoritmo encontrado, entrenado con datos de Entrenamiento, se muestran en la tabla 6.2:

Tabla 6. 2

Análisis Técnico: Parámetros de Modelo

Kernel Utilizado	Kernel Polinomial
Grado del Kernel	1
Parámetro ϵ	0.01
Parámetro C	0.75

Fuente: Elaboración Propia.

Los resultados del modelo con los datos de Prueba se muestran en la tabla 6.3:

Tabla 6. 3

Análisis Técnico: Resultados

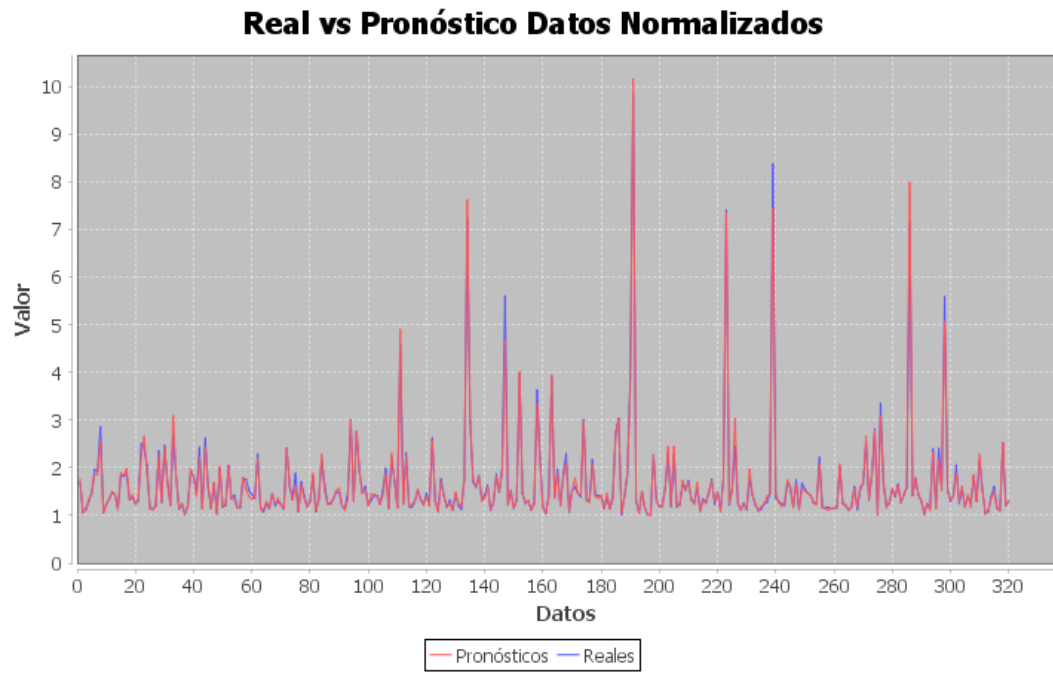
MAE (Datos Normalizados)	0.061110426
MAPE (Datos Normalizados)	2.923603%
MSE (Datos Normalizados)	0.016496316
Dirección	62.1875%
MAE	5.2711134
MAPE	9.798934%
MSE	122.732956

Fuente: Elaboración Propia.

Los siguientes gráficos, figuras 6.1 y 6.2, comparan los datos real con los pronósticos obtenidos.

Figura 6. 1

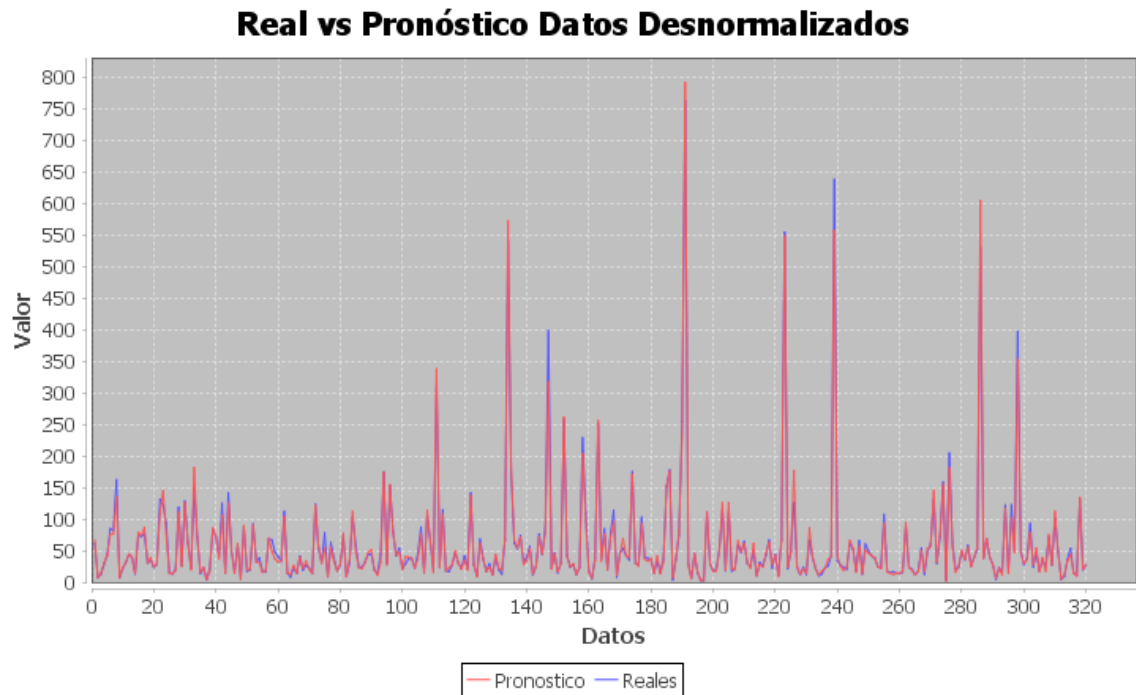
Análisis Técnico: Real vs Pronóstico. Datos Normalizados



Fuente: Elaboración Propia.

Figura 6. 2

Análisis Técnico: Real vs Pronóstico. Datos Desnormalizados



Fuente: Elaboración Propia.

6.2. Análisis Fundamental

El segundo modelo de Support Vector Regression se desarrolló utilizando los indicadores del análisis fundamental adicionado al Precio de Cierre de la Acción.

Los indicadores utilizados para este modelo se muestran en la tabla 6.4:

Tabla 6. 4

Análisis Fundamental: Indicadores

Precio de Acción	
Precio de Acción	<i>Precio de cierre del día</i>
Indicadores de Análisis Fundamental	
Razón Corriente	$\frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$
Capital de Trabajo	<i>Activo Corriente – Pasivo Corriente</i>
Rotación de Cuentas por Cobrar	$\frac{\text{Ventas}}{\text{Cuentas por cobrar}}$
Rotación de Inventarios	$\frac{\text{Ventas}}{\text{Inventarios}}$
Rotación de Activos Totales	$\frac{\text{Ventas}}{\text{Total de Activos}}$
Relación Deuda Capital	$\frac{\text{Deuda a Largo Plazo}}{\text{Capital Accionistas}}$
Cobertura de Interés	$\frac{\text{Utilidad Antes de Impuestos}}{\text{Gastos por Interés}}$
Margen de Utilidad Neta	$\frac{\text{Utilidad Neta}}{\text{Ventas}}$
Rendimiento sobre los Activos	$\frac{\text{Utilidad Neta}}{\text{Activos Totales}}$
Rendimiento sobre el Capital	$\frac{\text{Utilidad Neta}}{\text{Capital de accionistas}}$
Dividendos por acción	$\frac{\text{Dividendos pagados}}{\text{Número de Acciones Ordinarias}}$
Razón de pago	$\frac{\text{Dividendos por Acción}}{\text{Utilidades por Acción}}$
Razón Valor en libros – Valor de Mercado	$\frac{\text{Valor de acciones en Libros}}{\text{Capitalización de Empresa}}$
Razón de Utilidades	$\frac{\text{Utilidad Neta}}{\text{Número de Acciones}}$
Tamaño	<i>Capitalización de Empresa</i>

Fuente: Elaboración Propia.

Los parámetros del mejor algoritmo encontrado, entrenado con datos de Entrenamiento, se muestran en la tabla 6.5:

Tabla 6. 5

Análisis Fundamental: Parámetros de Modelo

Kernel Utilizado	Kernel Polinomial
Grado del Kernel	1
Parámetro ϵ	0.01
Parámetro C	0.75

Fuente: Elaboración Propia.

Los resultados del modelo con datos de Prueba se muestran en la tabla 6.6:

Tabla 6. 6

Análisis Fundamental: Resultados

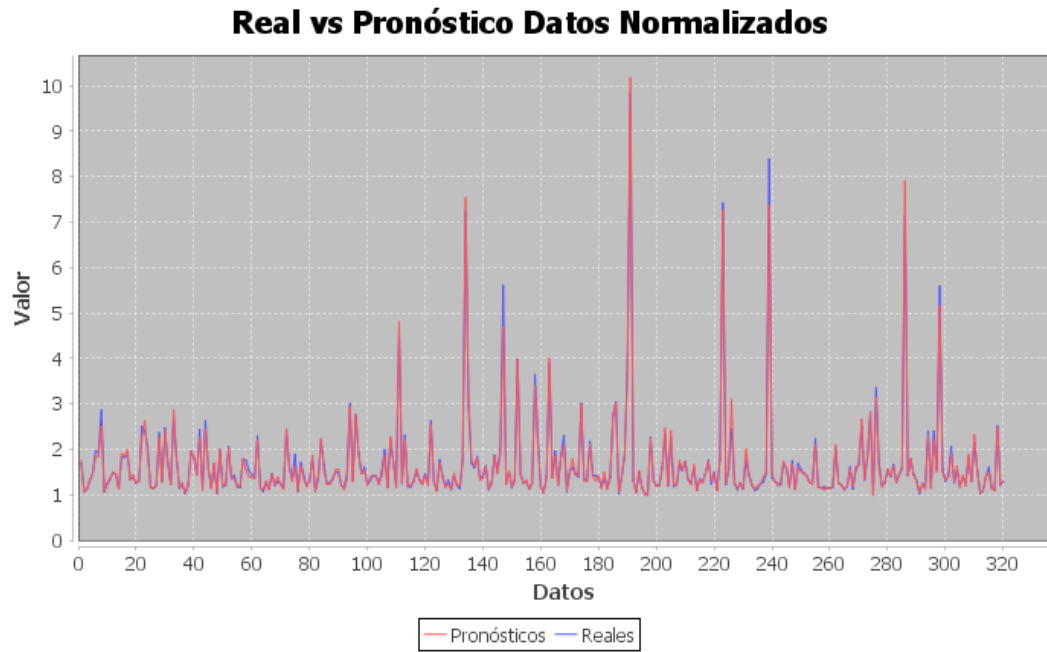
MAE (Datos Normalizados)	0.060776137
MAPE (Datos Normalizados)	2.9640436%
MSE (Datos Normalizados)	0.01566691
Dirección	61.562504%
MAE	5.24228
MAPE	10.507793%
MSE	116.56218

Fuente: Elaboración Propia.

Los siguientes gráficos, figuras 6.3 y 6.4, comparan datos real con los pronósticos obtenidos.

Figura 6. 3

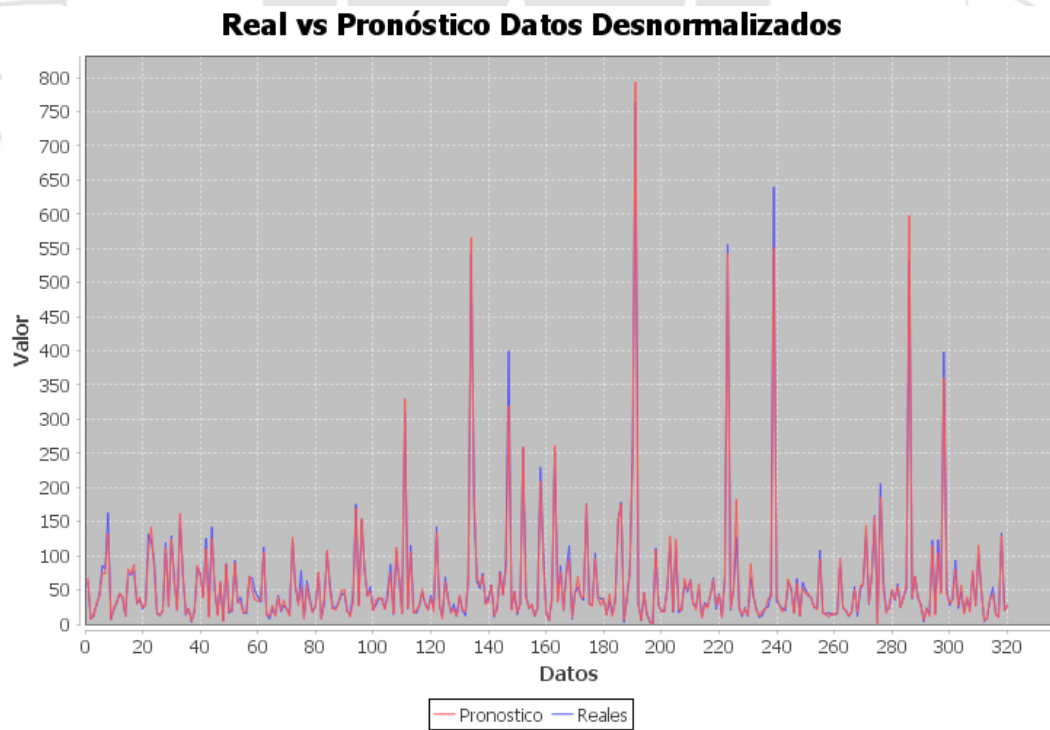
Análisis Fundamental: Real vs Pronóstico. Datos Normalizados



Fuente: Elaboración Propia.

Figura 6. 4

Análisis Fundamental: Real vs Pronóstico. Datos Desnormalizados



Fuente: Elaboración Propia.

6.3. Análisis Técnico y Fundamental

Luego utilizó el algoritmo de Support Vector Regression utilizando todos los indicadores investigados, del análisis técnico como del análisis fundamental adicionado al Precio de Cierre de la Acción.

Para este modelo, se utilizaron todos los indicadores expuestos en el capítulo anterior, se muestran en la tabla 6.7:

Tabla 6. 7

Análisis Técnico y Fundamental: Indicadores

Precio de Acción	
Precio de Acción	<i>Precio de cierre del día</i>
Indicadores de Análisis Técnico	
Promedio Móvil de 21 Días	<i>Promedio 21 últimos días</i>
Promedio Móvil de 55 Días	<i>Promedio 55 últimos días</i>
Promedio Móvil de 144 Días	<i>Promedio 300 últimos días</i>
Momentum de 10 Días	<i>Precio de Acción – Precio de Acción 10 días antes</i>
Momentum de 70 Días	<i>Precio de Acción – Precio de Acción 70 días antes</i>
Momentum de 300 Días	<i>Precio de Acción – Precio de Acción 300 días antes</i>
Indicadores de Análisis Fundamental	
Razón Corriente	$\frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$
Capital de Trabajo	$\text{Activo Corriente} - \text{Pasivo Corriente}$
Rotación de Cuentas por Cobrar	$\frac{\text{Ventas}}{\text{Cuentas por cobrar}}$
Rotación de Inventarios	$\frac{\text{Ventas}}{\text{Inventarios}}$
Rotación de Activos Totales	$\frac{\text{Ventas}}{\text{Total de Activos}}$
Relación Deuda Capital	$\frac{\text{Deuda a Largo Plazo}}{\text{Capital Accionistas}}$
Cobertura de Interés	$\frac{\text{Utilidad Antes de Impuestos}}{\text{Gastos por Interés}}$
Margen de Utilidad Neta	$\frac{\text{Utilidad Neta}}{\text{Ventas}}$
Rendimiento sobre los Activos	$\frac{\text{Utilidad Neta}}{\text{Activos Totales}}$

Rendimiento sobre el Capital	$\frac{\text{Utilidad Neta}}{\text{Capital de accionistas}}$
Dividendos por acción	$\frac{\text{Dividendos pagados}}{\text{Número de Acciones Ordinarias}}$
Razón de pago	$\frac{\text{Dividendos por Acción}}{\text{Utilidades por Acción}}$
Razón Valor en libros – Valor de Mercado	$\frac{\text{Valor de acciones en Libros}}{\text{Capitalización de Empresa}}$
Razón de Utilidades	$\frac{\text{Utilidad Neta}}{\text{Número de Acciones}}$
Tamaño	$\text{Capitalización de Empresa}$

Fuente: Elaboración Propia.

Los parámetros del mejor algoritmo encontrado, entrenado con datos de Entrenamiento, se muestran en la tabla 6.8:

Tabla 6. 8

Análisis Técnico y Fundamental: Parámetros de Modelo

Kernel Utilizado	Kernel Polinomial
Grado del Kernel	1
Parámetro ϵ	0.01
Parámetro C	0.75

Fuente: Elaboración Propia.

Los resultados del modelo con datos de Prueba se muestran en la tabla 6.9:

Tabla 6. 9

Análisis Técnico y Fundamental: Resultados

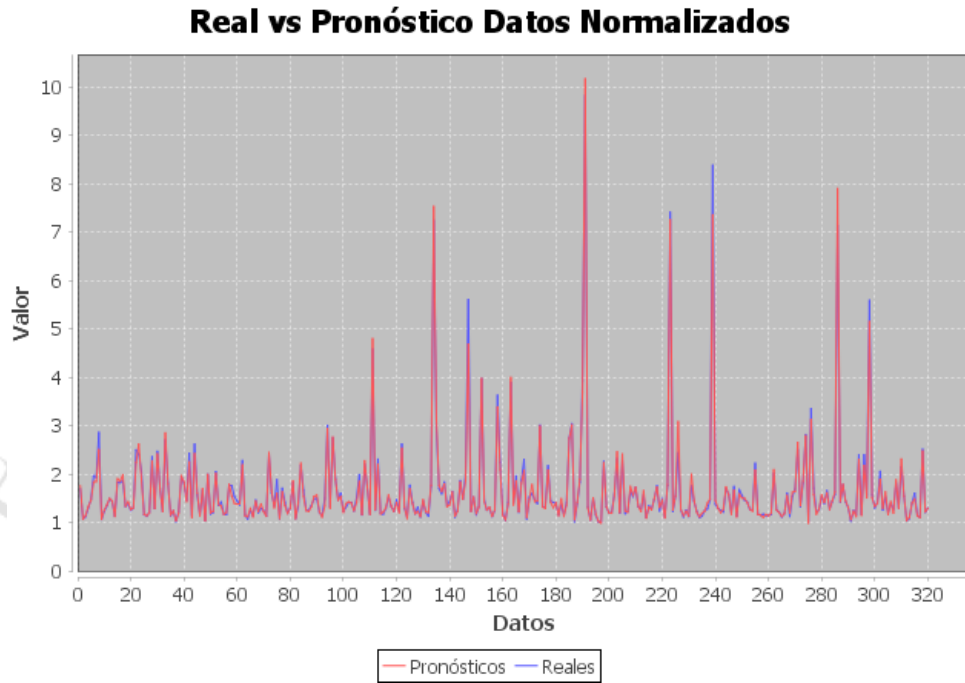
MAE (Datos Normalizados)	0.06191671
MAPE (Datos Normalizados)	2.9872935%
MSE (Datos Normalizados)	0.016454436
Dirección	59.375%
MAE	5.340661
MAPE	10.377501%
MSE	122.42137

Fuente: Elaboración Propia.

Los siguientes gráficos, figuras 6.5 y 6.6, comparan datos real con los pronósticos obtenidos.

Figura 6. 5

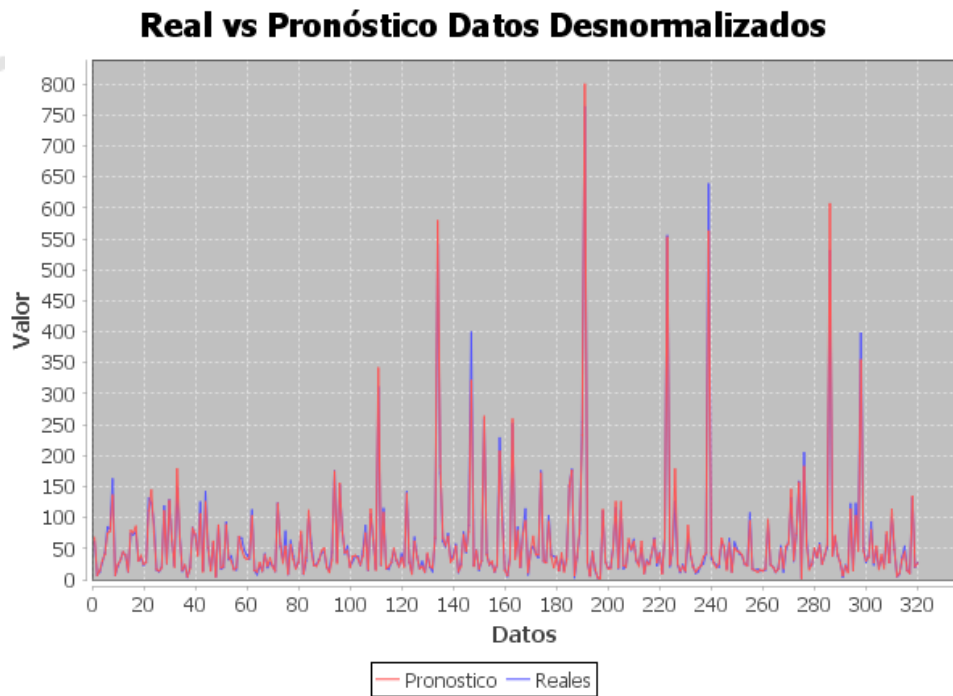
Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Normalizados



Fuente: Elaboración Propia.

Figura 6. 6

Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Desnormalizados



Fuente: Elaboración Propia.

6.4. Subgrupo de Análisis Técnico y Fundamental

Por último, se decidió usar un subgrupo de variables a criterio tanto del análisis técnico y del análisis fundamental.

Los siguientes indicadores se muestran en la tabla 6.10:

Tabla 6. 10

Subgrupo Análisis Técnico y Fundamental: Indicadores

Indicadores de Análisis Técnico	
Promedio Móvil de 21 Días	<i>Promedio 21 últimos días</i>
Promedio Móvil de 55 Días	<i>Promedio 55 últimos días</i>
Promedio Móvil de 144 Días	<i>Promedio 300 últimos días</i>
Momentum de 10 Días	<i>Precio de Acción – Precio de Acción 10 días antes</i>
Momentum de 70 Días	<i>Precio de Acción – Precio de Acción 70 días antes</i>
Momentum de 300 Días	<i>Precio de Acción – Precio de Acción 300 días antes</i>
Indicadores de Análisis Fundamental	
Razón Corriente	$\frac{\text{Activo Corriente}}{\text{Pasivo Corriente}}$
Cobertura de Interés	$\frac{\text{Utilidad Antes de Impuestos}}{\text{Gastos por Interés}}$
Margen de Utilidad Neta	$\frac{\text{Utilidad Neta}}{\text{Ventas}}$
Rendimiento sobre los Activos	$\frac{\text{Utilidad Neta}}{\text{Activos Totales}}$
Rendimiento sobre el Capital	$\frac{\text{Utilidad Neta}}{\text{Capital de accionistas}}$
Razón Valor en libros – Valor de Mercado	$\frac{\text{Valor de acciones en Libros}}{\text{Capitalización de Empresa}}$
Razón de Utilidades	$\frac{\text{Utilidad Neta}}{\text{Número de Acciones}}$

Fuente: Elaboración Propia.

Los parámetros del mejor algoritmo encontrado, entrenado con datos de Entrenamiento, se muestran en la tabla 6.11:

Tabla 6. 11

Subgrupo Análisis Técnico y Fundamental: Parámetros de Modelo

Kernel Utilizado	Kernel Polinomial
Grado del Kernel	1
Parámetro ϵ	0.01
Parámetro C	0.75

Fuente: Elaboración Propia.

Los resultados del modelo con datos de Prueba se muestran en la tabla 6.12:

Tabla 6. 12

Subgrupo Análisis Técnico y Fundamental: Resultados

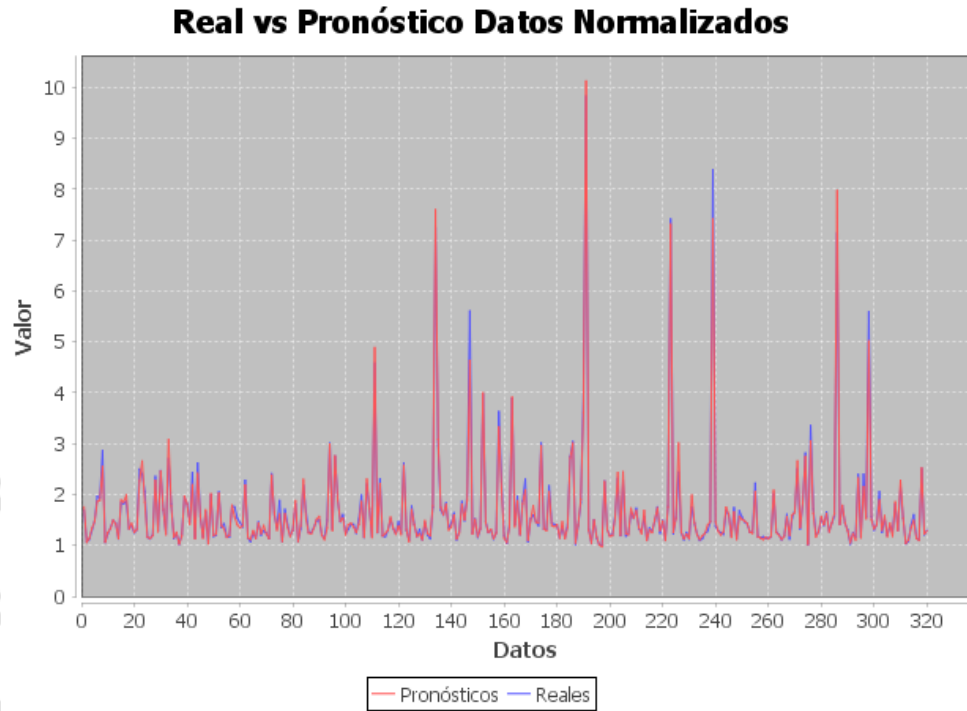
MAE (Datos Normalizados)	0.06168791
MAPE (Datos Normalizados)	2.9583266%
MSE (Datos Normalizados)	0.016885426
Dirección	59.375%
MAE	5.320922
MAPE	10.197345%
MSE	125.62804

Fuente: Elaboración Propia.

Los siguientes gráficos, figura 6.7 y 6.8, comparan los datos reales con los pronósticos obtenidos

Figura 6. 7

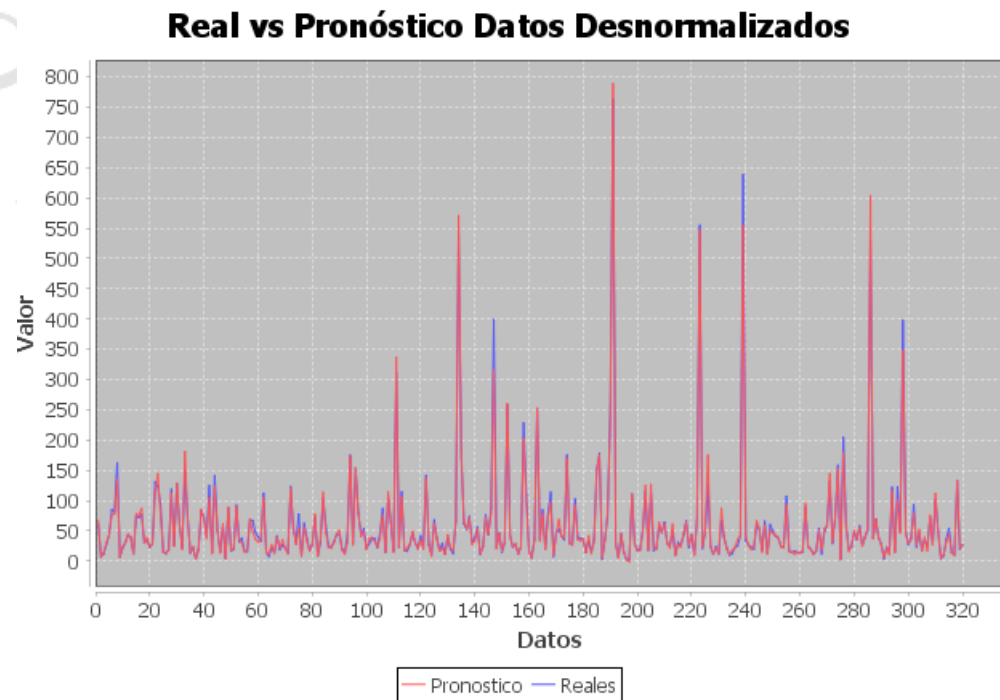
Subgrupo Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Desnormalizados



Fuente: Elaboración Propia.

Figura 6. 8

Subgrupo Análisis Técnico y Fundamental: Real vs Pronóstico. Datos Desnormalizados



Fuente: Elaboración Propia.

6.5. Comparación de Resultados

La tabla 6.13 resume los resultados de los 4 modelos.

Tabla 6. 13

Comparación de Resultados

Indicador	Análisis Técnico	Análisis Fundamental	Todos los Indicadores	Subgrupo de Indicadores
MAE (Datos Normalizados)	0.061110426	<u>0.060776137</u>	<u>0.06191671</u>	0.06168791
MAPE (Datos Normalizados)	<u>2.923603%</u>	2.9640436%	<u>2.9872935%</u>	2.9583266%
MSE (Datos Normalizados)	0.016496316	<u>0.01566691</u>	0.016454436	<u>0.016885426</u>
DA	<u>62.1875%</u>	61.562504%	<u>59.375%</u>	<u>59.375%</u>
MAE	5.271114	<u>5.24228</u>	<u>5.340661</u>	5.320922
MAPE	<u>9.798934%</u>	<u>10.507796%</u>	10.377501%	10.197345%
MSE	122.732956	<u>116.56218</u>	122.42137	<u>125.62804</u>

Fuente: Elaboración Propia.

La tabla 6.13 muestra que los cuatro modelos tienen resultados similares en indicadores estadísticos utilizados para medir los pronósticos de cada uno comparados con datos real.

Los datos que son subrayados con una línea indican que el modelo tuvo la mayor performance para ese indicador comparado con los otros modelos. Los datos que son subrayados con dos líneas, indican que el modelo tuvo la menor performance comparado con los tres modelos restantes. Es importante recalcar que la diferencia en todos los casos entre el mayor y el menor es del 8% relativo al mayor, por lo que los modelos se pueden considerar de similar performance.

Las mejores medidas lo tienen los modelos de análisis técnico y análisis fundamental cuando se encuentran separados uno del otro, teniendo una performance disminuida cuando los indicadores son combinados, sea en el modelo que incluye todos los indicadores o el subgrupo de indicadores.

Algunos motivos de obtener estos resultados pueden ser la redundancia de indicadores entre ambos tipos de análisis financiero, es decir que los precios parte del análisis técnico ya incluyen indirectamente la información financiera que contienen los indicadores del análisis fundamental.

Otro motivo puede ser la correlación que existe entre los indicadores de ambas ramas de análisis, si el precio sube y tiene buenos indicadores de tendencias de parte del análisis técnico, puede que ser refleje también en los estados financieros de la empresa, origen del análisis fundamental.

Sin embargo, si bien los dos primeros modelos de ramas de análisis independientes se presentan como mejores alternativas, los cuatro modelos muestran resultados aceptables no muy distintos entre ellos. Por lo que otro análisis de resultados desde un nuevo punto de vista puede que establezca mejores diferencias entre los modelos. Por tanto, se analizarán los resultados de los modelos desde el punto de vista: el rendimiento de acciones

La tabla 6.14 muestra escenarios de rendimientos de acciones, bajo el supuesto de elegir las acciones cuya predicción es mayor que cierto porcentaje, teniendo como resultados la cantidad de acciones y el rendimiento del conjunto elegido.

Rendimiento de Prueba: 320 Acciones - 3,994753%

Tabla 6. 14

Comparación de Rendimientos

Indicador	Análisis Técnico		Análisis Fundamental		Todos los Indicadores		Subgrupo de Indicadores	
	#	%	#	%	#	%	#	%
1.05	<u>78</u>	6,931	96	<u>5,017</u>	<u>117</u>	7,174	93	<u>7,481</u>
1.06	<u>53</u>	<u>4,938</u>	71	7,751	87	<u>8,277</u>	64	6,863
1.07	<u>33</u>	<u>4,210</u>	52	7,263	66	5,363	47	<u>8,557</u>
1.08	<u>24</u>	7,929	37	<u>7,025</u>	49	8,445	36	<u>11,318</u>
1.09	<u>15</u>	<u>14,942</u>	29	<u>7,401</u>	35	10,301	29	12,246
1.10	<u>11</u>	<u>18,652</u>	22	<u>6,400</u>	30	9,911	18	16,869

Fuente: Elaboración Propia.

Nuevamente, los datos que son subrayados con una línea indican que el modelo tuvo la mayor performance para esa categoría comparado con los otros modelos. Los datos que son subrayados con dos líneas, indican que el modelo tuvo la menor performance comparado con los tres modelos restantes.

Este escenario muestra diferencias de mejor forma entre los modelos propuestos. Se encuentra diferencias especialmente al considerar como objetivos rendimientos mayores.

Los modelos con mejores resultados en este escenario, uno de los más importantes puesto que el objetivo de las acciones es conseguir mayores rendimientos, son el análisis técnico y el subgrupo de indicadores.

El conjunto de las 320 acciones de datos de entrenamiento tuvo un rendimiento promedio de 3,994753%. Los modelos de predicción expuestos en esta tabla permiten escoger las acciones que tuvieron mejor rendimiento mediante las predicciones realizadas.

El modelo del análisis fundamental es el modelo que tiene los resultados menos óptimos. De igual forma tiene mejores resultados que el promedio de rendimiento del conjunto de acciones escogidas. El análisis fundamental por la teoría busca acciones cuyas empresas haya mostrado buenos resultados en el Análisis Financiero, y esto se muestra reflejado hasta cierto punto en las acciones escogidas. Quizás a mayor plazo de tiempo tendría este modelo mejores resultados.

El modelo del análisis técnico por otro lado, por buscar tendencias, tiene mejores desempeños a un horizonte de tiempo de un trimestre como es este caso. Por tanto, muestra mejores resultados que el Análisis Financiero en cuestión del potencial de obtener altos rendimientos escogiendo ciertas acciones determinadas por el modelo. Su debilidad sería en el número de acciones escogidas, puesto que una mayor cantidad de acciones da una mejor diversificación de riesgo.

El modelo que combina los Indicadores del análisis fundamental y el análisis técnico tiene mejores resultados en cuestión a rendimiento que el análisis fundamental y mejores resultados en cuestión a cantidad de acciones que el análisis técnico.

El modelo de subgrupo de Indicadores tiene superiores resultados que el modelo previo, puesto que de cierta forma optimiza qué indicadores son utilizados del este modelo, teniendo mejores resultados en rendimiento y cantidad de acciones. En cierto aspecto es superior al modelo de análisis técnico puesto que considera mayor cantidad acciones en sus resultados obteniendo rendimientos comparables.

Esto último indica la importancia de un método de selección de variables que determine un subconjunto de las mismas para ser utilizado por el algoritmo de Support Vector Regression.

Como adicional, la tabla 6.15 muestra escenarios de rendimientos de acciones, bajo el supuesto de elegir las acciones cuya predicción es menor que cierto porcentaje, teniendo como resultados la cantidad de acciones y el rendimiento del conjunto elegido.

Tabla 6. 15

Comparación de Rendimientos (Negativos)

Indicador	Análisis Técnico		Análisis Fundamental		Todos los Indicadores		Subgrupo de Indicadores	
	#	% (Rend.)	#	% (Rend.)	#	% (Rend.)	#	% (Rend.)
1.00	38	2.8366	<u>36</u>	<u>0.3702</u>	43	6.0766	<u>61</u>	3.5182
0.99	<u>13</u>	<u>3.8957</u>	21	<u>-1.5274</u>	26	2.8361	<u>40</u>	2.7473
0.98	<u>6</u>	-1.0859	17	<u>-4.1695</u>	16	0.6308	<u>25</u>	<u>5.3328</u>
0.97	<u>2</u>	2.9847	13	<u>-4.8693</u>	11	<u>5.6948</u>	<u>15</u>	-1.4348
0.96	<u>1</u>	4.0462	<u>12</u>	<u>-4.7330</u>	10	<u>6.5466</u>	<u>12</u>	0.8259
0.95	<u>1</u>	4.0462	<u>13</u>	<u>-3.2702</u>	8	<u>8.0343</u>	10	0.4463

Fuente: Elaboración Propia.

La importancia de la predicción de rendimientos negativos se debe a dos razones. La primera es poder determinar qué acciones no se deberían conservar puesto que llevarían a pérdidas mientras que la segunda es por la existencia de las opciones de venta. Estas son un tipo de inversión que consiste en un contrato entre dos partes en donde una adquiere un derecho de poder vender cierta cantidad de acciones de determinada compañía a un determinado precio en un determinado periodo de tiempo a cambio de un pago inicial (Alexander, 2010). Esto permite obtener ganancias en el futuro mediante predicciones de rendimientos negativos de acciones.

Los resultados de esta tabla muestran que los modelos en conjunto no tienen la misma performance para predecir rendimientos positivos que rendimientos negativos. El modelo que mejor se adapta es el de análisis fundamental mientras que el segundo mejor adaptado sería el de subgrupo de Indicadores.

Sin embargo, se muestra también que, debido a que el análisis fundamental tiene buen rendimiento este escenario, existe un grupo de indicadores dentro todos los considerados que podría predecir con mejor precisión rendimientos negativos de acciones.

Se debería considerar que un modelo de rendimientos negativos puede ser diferente en estructura e indicadores que uno para rendimientos positivos.

6.6. Comparación con Support Vector Machine

Se realizó también un experimento utilizando los modelos detallados anteriormente, es decir, tanto los indicadores y los parámetros de cada modelos (en el caso que sean aplicables) para construir nuevos modelos. Estos nuevos modelos son de clasificación basados en Support Vector Machine, que busca pronosticar la dirección del cambio de precio que tiene la acción.

Los resultados preliminares de porcentaje de acierto se muestran en la tabla 6.16.

Tabla 6. 16

Support Vector Machine: Resultados

	Análisis Técnico	Análisis Fundamental	Todos los Indicadores	Subgrupo de Indicadores
SVM	65.31%	64.69%	64.69%	65.31%

Fuente: Elaboración Propia

Analizando los resultados a mayor detalle, se encuentra un problema en estos modelos de SVM. Casi todas las predicciones son de tipo positivo, asignando a cada caso un valor mayor a 0. Esto quiere decir que el modelo predice que la acción sube de precio en prácticamente todos los casos.

Regresando a la teoría, los valores entre -1 y +1 son valores ambiguos, en el que el modelo no se encuentra seguro de su clasificación, puesto que caen dentro del margen mínimo. Si se analizan los resultados en que el modelo asigna valores mayores a +1 o menores a -1, se muestran en la tabla 6.17:

Tabla 6. 17

Support Vector Machine: Resultados (2)

	Análisis Técnico		Análisis Fundamental		Todos los Indicadores		Subgrupo de Indicadores	
	#	%	#	%	#	%	#	%
SVM	19	57.89%	136	65.44%	117	67.52%	24	83.33%

Fuente: Elaboración Propia.

El modelo de análisis técnico en Support Vector Machine no tiene buenos resultados, pues tiene poca cantidad de acciones y al mismo tiempo un bajo porcentaje de aciertos.

Tanto el modelo de análisis fundamental y el modelo que contiene todos los indicadores tienen una mayor cantidad de acciones y a su vez un buen porcentaje de aciertos.

El modelo de Subgrupo de Indicadores tiene una menor cantidad de acciones pero un muy buen porcentaje de aciertos frente a los demás modelos.

Para poder observarlo de otra perspectiva, se medirá el escenario de rendimientos en el caso de todas las acciones con pronóstico +1, es decir, en el que el modelo está seguro de que subirán de precio.

La tabla 6.18 muestra los resultados:

Tabla 6. 18

Support Vector Machine: Rendimientos

	Análisis Técnico		Análisis Fundamental		Todos los Indicadores		Subgrupo de Indicadores	
	#	% (Rend.)	#	% (Rend.)	#	% (Rend.)	#	% (Rend.)
SVM	19	3.7920	133	3.6232	114	3.9082	24	10.5757

Fuente: Elaboración Propia.

Existe una clara diferencia en los rendimientos de las acciones seleccionadas por los modelos, donde el modelo de Subgrupo de Indicadores tiene los mejores resultados en cuestión de obtener rendimientos de acciones.

Se refuerza la idea de que una selección de indicadores en el ayudaría a obtener mejores resultados, y se resalta la posibilidad de pueden existir formas de predecir rendimientos de acciones.

Comparando ambos modelos se observa que los modelos basados en Support Vector Regression se ajustan mejor que los modelos de Support Vector Machine para predicción de acciones.

La razón financiera de un mejor ajuste de Support Vector Regression es que el objetivo no es solo buscar acciones que tengan rendimientos positivos. En los mercados financieros, al ser las acciones instrumentos de inversión, se espera que todas tengan rendimientos positivos y que estos sean superior a los rendimientos positivos de otros instrumentos de inversión.

Support Vector Machine, al clasificar las acciones, pronosticando la dirección de precios, no distingue si este incremento es superior o inferior al del mercado. Por tanto, no se puede distinguir qué acciones escogidas por el modelo tendrán un rendimiento superior al promedio del mercado. Esto resulta en que no se sabe qué acciones se deben elegir de todas las recomendadas, puesto que pueden mostrar un incremento pero no alineado a las expectativas.

Support Vector Regression, al pronosticar precios, permite buscar acciones que tengan rendimientos superiores al promedio del mercado, que es el objetivo de todo inversionista que incurra en la bolsa de valores.

Por tanto, Support Vector Regression es un modelo que mejor se adecúa a los objetivos de los participantes de este mercado que un modelo de Support Vector Machine.

CONCLUSIONES

En esta investigación se ha realizado un experimento de predicción de precios mediante modelos basados en Support Vector Machine y Support Vector Regression. Se han implementado varios modelos utilizando indicadores del análisis técnico y análisis fundamental, ambas metodologías utilizadas para el análisis financiero de acciones comunes en la bolsa de valores.

Los resultados de los modelos de Support Vector Regression son evaluados mediante indicadores estadísticos como DA, MAE, MAPE y MSE. Los cuatro modelos elegidos, uno de análisis técnico, otro de análisis fundamental, uno combinado de las dos metodologías y finalmente un subgrupo de indicadores de último grupo, tienen resultados similares entre sí. Tienen como resultados promedio una dirección correcta de 60%, un error promedio de 5 dólares, un error porcentual de 10% y un error promedio cuadrado de 120 dólares.

Los resultados de los modelos de Support Vector Machine son evaluados por la cantidad de veces que se precisó correctamente la dirección del precio futuro con respecto al inicial. El modelo que mejor performance tuvo fue el del subgrupo de indicadores con 83% de dirección, pero tan solo clasificando 24 acciones de las 320 de forma segura. Los modelos de análisis fundamental y el modelo de todos los indicadores clasificaron a 133 y 114 acciones con un éxito de alrededor de 66%.

Al no tener claras diferencias entre los modelos, se opta por realizar un análisis de otro punto de vista, evaluando el rendimiento de las acciones escogidas por el modelo. El modelo más destacado para rendimientos positivos es el modelo de subgrupo de indicadores tanto para los modelos basados en Support Vector Regression y en Support Vector Machine con rendimientos de hasta 18,652% y de 10,5757% respectivamente, en un contexto donde el conjunto es de 3,9947%

La diferencia entre los modelos es el contexto del mercado, al ser las acciones un instrumento financiero de inversión, entonces las expectativas son que las rentabilidades sean positivas. Por tanto, para obtener rendimientos superiores al mercado, no es suficiente poder

predecir la dirección de las acciones, dado por los modelos de Support Vector Machine, sino se requiere predecir qué tanto subirán los precios de estas acciones, dado por los modelos de Support Vector Regression.

Otra diferencia es poder predecir rendimientos negativos de acciones, que en este caso no es posible por los modelos de Support Vector Machine evaluados, pero si por los modelos de Support Vector Regression desarrollados. El modelo de análisis fundamental predice con buenos resultados los rendimientos negativos de las acciones.

Por tanto, y por los resultados expuestos en la investigación, se determina que, para el contexto del problema, un modelo de Support Vector Regression se adecúa más a un modelo de Support Vector Machine. Asimismo, los modelos Support Vector Regression propuestos y discutidos en la investigación permiten identificar acciones que tendrían un rendimiento superior al promedio del mercado, objetivo de los inversionistas de la bolsa de valores.



Recomendaciones

- La automatización de transformación de datos de estados financieros y Cotizaciones de acciones es una ayuda para futuras investigaciones y experimentos en acciones.
 - Automatización de búsqueda de parámetros de los modelos.
 - Como se indicó en la investigación, la determinación de indicadores evitando factores como la autocorrelación puede mejorar los resultados de los modelos para que se puedan adecuar mejor al problema. Una investigación de preprocesamiento de todos los indicadores nombrados a lo largo de la investigación para utilizar en un modelo de Support Vector Regression es el siguiente paso a esta investigación.
 - La realización de experimentos en otra época puede ser una prueba para validar el modelo. La investigación se realizó en una época en conjunto de crecimiento positivo por lo que una investigación utilizando datos de una época de crecimiento negativo sería recomendable.
 - Se recomienda una investigación sobre la correlación de las acciones escogidas puesto que, para conformar un portafolio en el mercado, se requieren acciones que estén negativamente correlacionadas para protección frente al riesgo.
 - Una investigación sobre el riesgo de acciones complementaría esta investigación para una exitosa formación de un portafolio de acciones.



SCIENTIA ET PRAXIS

REFERENCIAS

- Alexander, G. J. (2010). *Fundamentals of investments/ fundamentos de inversiones*. Place of publication not identified: Prentice Hall.
- Berk, DeMarzo & Harford. (2009). *Fundamentals of Corporate Finance*. Prentice Hall; 2nd edition.
- Cao, Lijuan. (2001). Support vector machines experts for time series forecasting. *Neurocomputing* 51: 321-339 [http://dx.doi.org/10.1016/S0925-2312\(02\)00577-5](http://dx.doi.org/10.1016/S0925-2312(02)00577-5)
- Choudhry, R., & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39(3), 315-318.
- Clive W.J. Granger. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting* 8(1):3-13 doi:10.1016/0169-2070(92)90003-R
- Choudhry, R., & Garg, K. (2008). A hybrid machine learning system for stock market forecasting. *World Academy of Science, Engineering and Technology*, 39(3), 315-318.
- Credit Suisse (s.f.) Technical Analysis – Explained.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.
- Dai, W., Shao, Y. E., & Lu, C. J. (2013). Incorporating feature selection method into support vector regression for stock index forecasting. *Neural Computing and Applications*, 23(6), 1551-1561.
- Das, S. P., & Padhy, S. (2012). Support vector machines for prediction of futures prices in Indian stock market. *International Journal of Computer Applications*, 41(3).
- De Gooijer JG, Hyndman RJ (2006) 25 years of time series forecasting. *Int J Forecast* 22(3):443–473
- García, M. C., Jalal, A. M., Garzón, L. A., & López, J. M. (2013). Métodos para predecir índices bursátiles. *Ecos de Economía*, 17(37), 51-82.
- Gitman, L. J., & Joehnk, M. D. (2009). *Fundamentos de inversión*. México: Pearson Educación.
- Granger, C. W. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting*, 8(1), 3-13.

- Guajardo, J., Weber, R., & Miranda, J. (2006). A forecasting methodology using support vector regression and dynamic feature selection. *Journal of Information & Knowledge Management*, 5(04), 329-335.
- Han, J., Kamber, M., & Pei, J. (2013). *Data mining: Concepts and techniques* (Third ed.). Burlington, MA: Elsevier.
- Huang, W., Y. Nakamori, & S. Wang. (2005) Forecasting Stock Market Movement Direction with Support Vector Machine. *Computers & Operations Research*
- Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). *A practical guide to support vector classification*
- Institute of Financial Forecasters. (s.f.) Recuperado de <https://forecasters.org/about/>
- S&P Dow Jones Indices (s.f.) Recuperado de <http://us.spindices.com/our-company/>
- S&P Dow Jones Indices (s.f.) Recuperado de <http://us.spindices.com/indices/equity/sp-500>
- Ng, A. (s.f) *Support Vector Machines*.
- Sáez, . C. A., & Doblado, C. (2013). *Mercados financieros: Todos por una causa*.
- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.
- Tan, P.-N., Steinbach, M., & Kumar, V. (2015). *Introduction to data mining*. Dorling Kindersley: Pearson.
- Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *omega*, 29(4), 309-317.
- Wang, J., Hou, R., Wang, C., & Shen, L. (2016). Improved v-Support vector regression model based on variable selection and brain storm optimization for stock price forecasting. *Applied Soft Computing*, 49, 164-178.
- Welch, I. (2011). *Corporate finance*. Ivo Welch.



ANEXOS

Anexo 1: Comandos SVMLight

Para el uso de la librería se utilizaba los siguientes comandos via CMD

svm_learn [options] example_file model_file

- “*example_file*” contenía el set de datos de entrenamiento
- “*model_file*” representaba el nombre del archivo donde se guardaría el Modelo generado de SVM o SVR

Las opciones utilizadas eran las siguientes:

-z {c,r}	Seleccionando C para Clasificación (SVM) y R para regresión (SVR)
-w {}	Parámetro para la regresión ϵ
-c {}	Costo de los errores
-t 1	Seleccionando 1 para el Kernel Polinomial
-d {}	Grado de Kernel Polinomial

svm_classify [options] example_file model_file output_file

- “*example_file*” contenía el set de datos de prueba
- “*model_file*” representaba el nombre del archivo donde se había guardado el Modelo generado de SVM o SVR que se utilizaría para las predicciones
- “*output_file*” es el nombre dado al archivo que almacenaría las predicciones del set de datos de prueba

Para el caso de predicciones, no se utilizaba ninguna opción especial.