UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA

INGENIERÍA DE SISTEMAS – OPTIMACIÓN FINANCIERA

MACHINE LEARNING EN EL PRONÓSTICO DEL ÍNDICE DE PRECIOS Y COTIZACIONES DE LA BOLSA MEXICANA DE VALORES.

TESIS

QUE PARA OPTAR POR EL GRADO DE:

MAESTRO EN INGENIERÍA

PRESENTA:

ING. GERARDO PALACIOS MORALES

TUTOR:

M. EN I. JORGE RODRÍGUEZ RUBIO

PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA

CIUDAD UNIVERSITARIA CD. MX. JUNIO 2019

JURADO ASIGNADO:

Presidente:	Dr. Ortiz Calisto Edgar
Secretario:	Dr. Martínez Miranda Elio Agustín
Vocal:	M. en I. Rodríguez Rubio Jorge
1 er. Suplente:	Dra. Sosa Castro Magnolia Miriam
2 d o. Suplente:	M. en I. Malfavón Ruiz Yonahandy

Lugar o lugares donde se realizó la tesis:

Universidad Nacional Autónoma de México
Posgrado de Ingeniería
Ciudad de México, México. Junio 2019

TUTOR DE TESIS:

MI. JORGE RODRÍGUEZ RUBIO

FIRMA

Agradecimientos

A la Universidad Nacional Autónoma de México por la oportunidad que me ha brindado para participar en el programa de maestría.

Al CONACYT por el apoyo económico que me otorgó durante el programa de maestría.

A mis padres, Laura y René y a mi hermano René, quienes me han brindado un apoyo incondicional a lo largo de mi vida y siempre me han motivado en los momentos difíciles para salir adelante. No me alcanzan las palabras para expresar el orgullo y lo agradecido que me siento por tener una familia tan asombrosa.

A Gloria y Renata por su motivación e inspiración en todos los aspectos de mi vida.

A mis amigos, Alejandra, Ivette, Joab, Alan y Julio por todos esos momentos divertidos, por impulsarme a concluir este trabajo y por su motivación en los momentos de estrés y frustración.

A mi tutor, Jorge Rodríguez Rubio por sus enseñanzas, y consejos.

A los integrantes del jurado, M. en I. Yonahandy Malfavón, Dra. Miriam Sosa, Dr. Edgar Ortiz y Dr. Elio Martínez por sus enseñanzas a lo largo de la maestría y sus comentarios y observaciones en la revisión de esta tesis.

Tabla de contenido

Lista de figuras

Lista de tablas

Resumen

Este proyecto de investigación emplea diferentes algoritmos de clasificación de *Aprendizaje Automático* (Machine Learning) para pronosticar el movimiento del precio de cierre diario del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (IPC). Los resultados son evaluados y comparados para determinar si el uso de estas herramientas es rentable para su aplicación en la toma de decisiones en estrategias efectivas de trading.

Se desarrollan siete modelos empleando algoritmos de Machine Learning (*Maquinas de Soporte Vectorial, K-Nearest Neighbor, Naive Bayes, Arboles de decisión, Bosques Aleatorios, Regresión Logística y Redes Neuronales Artificiales*) con el fin de predecir si el precio de una acción se incrementara o no al cierre del día y evaluar estas herramientas como apoyo en la toma de decisiones en operaciones intradiarias de compra y venta. Como datos de entrada (variables predictoras) para los modelos se emplearon 3 bases con configuraciones diferentes, formadas por variables macroeconómicas, e indicadores financieros globales y mexicanos, así como indicadores de análisis técnico, propios del IPC. Estos datos se consideran como factores que pueden incidir en los movimientos del índice.

Esta tesis se desarrolla en seis capítulos, los capítulos 1 y 2 se enfocan en el marco teórico donde se proporciona un resumen sobre la temática desarrollada en esta investigación.

El capítulo 3 aborda la metodología de trabajo empleada para la elaboración de los pronósticos, un detalle de las variables de entrada seleccionadas como variables predictoras, así como los parámetros y criterios empleados en la aplicación de los algoritmos.

En el capítulo 4 se revisan los resultados obtenidos a través de los métodos de evaluación de desempeño para modelos de clasificación.

El capítulo 5 muestra la simulación (backtesting) de una estrategia de inversión empleando las señales de compra y venta generadas por los modelos de pronósticos y aplicado al IPC en un periodo anual.

En el capítulo 6 se exponen las conclusiones al proyecto de investigación y se proporcionan sugerencias en áreas de oportunidad detectadas para futuros trabajos.

Glosario de términos

Algoritmo. Conjunto de pasos establecido para resolver un problema o lograr un fin especifico.

Algoritmo de aprendizaje. Un algoritmo de aprendizaje, o un algoritmo de aprendizaje automático, es cualquier algoritmo que puede producir un modelo mediante el análisis de un conjunto de datos.

Algoritmos de caja negra. Es aquel en el que el usuario no puede ver la forma interna de funcionamiento del algoritmo.

Big Data. El Data Warehousing Institute (2011) define a Big Data como la aplicación de técnicas avanzadas de análisis para conjuntos de grandes volúmenes de datos. La consultora IDC (2012) lo define como: “Una nueva generación de tecnologías y arquitecturas diseñadas para extraer el valor económico de grandes volúmenes de una amplia variedad de datos, al permitir a alta velocidad, la captura, descubrimiento y/o análisis”

Business Intelligence (Inteligencia de Negocios). Es una estrategia empleada para mejorar el rendimiento, la efectividad y la competitividad de los negocios a través de una óptima organización y análisis de los datos. Estas herramientas basan la toma de decisiones en la lectura de los datos.

Característica. Una característica es un atributo de un ejemplo, generalmente una parte de un vector de características. Puede ser numérico o categórico.

Clase. Una clase es un grupo al que puede pertenecer un ejemplo. Un ejemplo etiquetado consiste en un vector de características y una referencia a la clase a la que pertenece. Una clase específica asociada a un vector de características se llama etiqueta. Por ejemplo, en un modelo de clasificación binario que detecta spam, hay dos clases, "spam" y "no spam". En un modelo de clasificación de múltiples clases que identifica especies de plantas, las clases serían "árboles", "flores", "hongos", etc.

Cluster. Es un grupo de ejemplos (generalmente, este grupo es más pequeño que el conjunto de datos completo) que tiene cierta semejanza según una métrica de similitud.

Clustering. La agrupación en clústeres es un problema de asignar ejemplos a uno o más clústeres.

Conjunto de entrenamiento. Es un subconjunto del conjunto de datos total, utilizado por el algoritmo de aprendizaje para crear un modelo.

Dataset. Un conjunto de datos, es decir, una colección de ejemplos.

Data Science (Ciencia de Datos). Es una disciplina enfocada al tratamiento y análisis de grandes volúmenes de datos por medio de herramientas como SQL y NoSQL para la gestión de estos.

Deep Learning. El Deep Learning es una técnica de aprendizaje automático que enseña a los ordenadores a hacer lo que resulta natural para las personas: aprender mediante ejemplos. Con el Deep Learning, un modelo informático aprende a realizar tareas de clasificación directamente a partir de imágenes, texto o sonido. Los modelos de Deep Learning pueden obtener una precisión de vanguardia que, en ocasiones, supera el rendimiento humano. Los modelos se entrenan mediante un amplio conjunto de datos etiquetados y arquitecturas de redes neuronales que contienen muchas capas.

Ejemplo. Un ejemplo (también llamado instancia) es un miembro de un conjunto de datos. Típicamente, un ejemplo es un vector de características. Cada característica representa alguna propiedad específica del ejemplo. Por ejemplo, si el conjunto de datos contiene ejemplos de estrellas, entonces una característica puede representar el porcentaje de hidrógeno en la estrella, otra característica podría representar el diámetro de la estrella, una tercera característica podría representar alguna propiedad del campo magnético de la estrella, etc. Todos los ejemplos representados como vectores tienen vectores de la misma dimensionalidad y cada dimensión representa la misma característica.

Ejemplo etiquetado. Un ejemplo etiquetado es un par que contiene el vector de características y una etiqueta. Una etiqueta es típicamente la cantidad que el modelo intenta predecir.

Entrenamiento. Es el proceso de construir un modelo mediante la aplicación de un algoritmo de aprendizaje automático a los datos de entrenamiento.

Hiperplano. Es un límite que separa un espacio en dos subespacios. Por ejemplo, una línea es un hiperplano en dos dimensiones y un plano es un hiperplano en tres dimensiones. En el aprendizaje automático, un hiperplano suele ser un límite que separa un espacio de alta dimensión. Por ejemplo, el algoritmo de Máquina de soporte vectorial utiliza hiperplanos para separar las clases positivas de las clases negativas, a menudo en un espacio de muy alta dimensión.

Modelo. Un modelo, también conocido como modelo estadístico, es el resultado de un algoritmo de aprendizaje automático aplicado a los datos de entrenamiento. El modelo es a menudo una fórmula matemática parametrizada, donde los parámetros se aprenden mediante el algoritmo de aprendizaje automático. Dado un ejemplo de entrada, un modelo puede producir la etiqueta de clasificación o el valor de regresión directamente, o puede producir una probabilidad para cada valor posible (etiqueta).

Normalización. La normalización es el proceso de convertir un rango real de valores en un rango estándar de valores, típicamente en el intervalo [-1, +1] o [0, 1]

Selección de características. La selección de características es un proceso de eliminación de las características del conjunto de datos que parecen irrelevantes para el modelado.

Sesgo (bias). El sesgo es un error de suposiciones erróneas en el algoritmo de aprendizaje. Un alto sesgo puede hacer que un algoritmo pierda las relaciones relevantes entre las características y las salidas de destino (fallas).

Trading. Negociaciones de compra y venta de instrumentos financieros que se dan en los mercados.

Umbral de clasificación. El umbral de clasificación es un número real que define un criterio que se aplica a la puntuación predicha de un modelo para separar la clase positiva de la clase negativa. Por ejemplo, un umbral de clasificación se usa al asignar los resultados de la regresión logística a la clasificación binaria: considere un modelo de regresión logística

que determine la probabilidad de que un mensaje de correo electrónico dado sea spam. Si el umbral de clasificación es 0.9, los valores de regresión logística por encima de 0.9 se clasifican como spam y los por debajo de 0.9 se clasifican como no spam.

Variable binaria. Una variable binaria es una variable (una característica o el objetivo) que puede tomar valores ya sea "Sí" o "No" (Verdadero o Falso, uno o cero, etc.).

Para la construcción de este glosario se emplearon como fuentes: Daza (2016), Joyanes (2013), Semanti.ca, y Arimetrics.com

Introducción

Introducción y propósitos de la investigación

Históricamente la predicción de los mercados accionarios ha sido una de las tareas más desafiantes y complejas debido a la aleatoriedad e incertidumbre que los rige, sin embargo, un pronóstico funcional del precio de las acciones puede ser de gran utilidad para apoyar las tomas de decisiones que realizan los inversionistas en la elaboración de estrategias de trading.

Un pronóstico en el mercado de valores es un proceso complejo ya que se trata de un mercado dinámico y no lineal, lo que lo hace complicado, no paramétrico y en si, de naturaleza caótica. El mercado accionario genera datos no estacionarios y en cualquier momento puede presentar tendencias cíclicas, caminatas aleatorias o incluso una combinación de estas.

Además, el mercado bursátil se ve afectado por eventos políticos, condiciones económicas generales, políticas empresariales, las expectativas de los inversionistas, las elecciones de inversionistas institucionales, la psicología de los inversores y al estar en un entorno globalizado, el movimiento de otras bolsas de valores.

En el campo de las inversiones en la bolsa de valores, es evidente que para obtener ganancias se requiere invertir en la acción a un precio dado y venderla cuando haya alcanzado un precio superior. Por tanto, la clave del éxito en esta forma de inversión es predecir con certeza que una acción subirá de precio en un periodo de tiempo razonable para venderla y generar utilidades razonables.

En la búsqueda por predecir los precios de las acciones se asume que las ocurrencias futuras están basadas en parte en información del presente y del pasado. Sin embargo, las series de tiempo financieras se encuentran entre las más cambiantes con señales difíciles de pronosticar.

El riesgo es un elemento propio de las inversiones en instrumentos de renta variable, pero también hay grandes oportunidades de conservar el poder adquisitivo del capital, y aún de hacer excelentes utilidades, y es mediante el análisis como podemos saber que el precio

actual de una acción es atractivo con alta probabilidad de aumentar y de esta manera controlar el riesgo y mejorar las posibilidades de tener ganancias. Esto no es fácil de alcanzar, y no existe ningún método infalible que nos permita lograrlo en cada ocasión. Sin embargo, puede lograrse con disciplina y análisis constante.

El análisis técnico y el análisis fundamental son dos de los métodos mas utilizados por los inversionistas para predecir el comportamiento del precio de un instrumento financiero antes de considerar invertir, sin embargo, hoy en día la ciencia de los datos se ha vuelto tan accesibles que básicamente cualquier persona cuenta con la capacidad de emplear diferentes técnicas y modelos computacionales que les permitan identificar mas y mejores oportunidades, así como tomar mejores decisiones concernientes a sus negocios. Técnicas como Big Data, Minería de Datos y Machine Learning han tenido un desarrollo exponencial ya que en general tienen el objetivo de darle un uso al gran volumen de datos generados día a día por las empresas y por los mercados. La aplicación de estas técnicas computacionales nos brinda la posibilidad de encontrar información significativa en grandes volúmenes de datos, de manera rápida y con un uso eficiente de recursos.

Las técnicas de computación se aplican ampliamente a los problemas del mercado de valores, brindando herramientas útiles para pronosticar entornos ruidosos como los mercados de valores, capturando así, su comportamiento no lineal. El uso de sistemas inteligentes como redes neuronales, sistemas difusos y algoritmos genéticos para fines de predicción en el campo de las finanzas tiene interesantes avances. Últimamente, las redes neuronales artificiales (ANN) y las máquinas de vectores de soporte (SVM) se han aplicado con éxito para resolver los problemas de predicción de series de tiempo financieras, incluida la predicción del mercado bursátil financiero [Boyacioglu y Avci, 2010].

Es gracias a este desarrollo de las tecnologías y la inteligencia artificial que surgen métodos modernos como el Aprendizaje Automático, conocido como Machine Learning, un campo de la ciencia de la computación que consiste en el estudio y aplicación de algoritmos computacionales que poseen la habilidad de encontrar patrones, generalizar y aprender sin ser explícitamente programados. Estas técnicas poseen la habilidad de procesar grandes volúmenes de datos para construir modelos que se pueden emplear, por ejemplo, en la búsqueda de patrones que nos permitan obtener pronósticos precisos.

A través del uso de estos algoritmos se han desarrollado numerosas investigaciones basadas en la predicción del movimiento de divisas, acciones, índices accionarios, la mayoría, aplicadas en mercados desarrollados. Sin embargo, existe un limitado número de investigaciones enfocadas en los mercados emergentes, por lo que, a diferencia de investigaciones previas, la importancia de este trabajo radica en su aplicación a un mercado emergente como es el caso de México[1], además de que existe un número reducido de investigaciones que emplean modelos de clasificación para pronósticos.

En cuanto a la predicción del precio de las acciones, recientemente se han desarrollado diferentes investigaciones aplicando técnicas de Machine Learning, sin embargo, existen un número tan grande de algoritmos, combinaciones de parámetros y técnicas que se pueden aplicar a un sin fin de mercados accionarios, que el comportamiento de estos modelos puede resultar de interés en un mercado tan particular como el mercado accionario mexicano.

México es un país que cuenta con características propias, por lo que es probable que el precio de los activos no pueda reflejar completamente toda la información disponible al público, existiendo así la posibilidad de que un pronóstico de los movimientos futuros en los precios de las acciones pueda llevarnos a una ventaja comercial. Como se sabe el comercio en los mercados emergentes es menos eficiente a causa de su menor liquidez, menor competencia comercial, menor transparencia y equidad, debido en mayor parte a la incertidumbre política y económica. Lo antes señalado se transfieren al mercado comercial y contribuye a una mayor ineficiencia de los mercados financieros.

Como hipótesis para este proyecto de investigación se considera que la metodología de Machine Learning aplicada al análisis de datos financieros y económicos permitirá obtener pronósticos eficientes sobre el precio de cierre diario del IPC, lo cual permitiría la posibilidad de que inversionistas y operadores tomen decisiones eficientes en estrategias de compra y venta de activos. Se espera que el resultado generado por la aplicación de la metodología señalada nos proporcione información veraz, oportuna y concisa, que permita obtener

[1] México se encuentra clasificado como país emergente por diversas organizaciones calificadoras como Standard & Poors, Dow Jones y Morgan Stanley entre otras, basado en criterios como desarrollo económico e ingreso per capita.

resultados más precisos, eficientes y competitivos en las operaciones realizadas en el mercado bursátil mexicano.

Con base en lo señalado el objetivo de esta investigación es la aplicación y evaluación de modelos modernos de análisis de datos considerando variables macroeconómicas, índices financieros e indicadores de análisis técnico de influencia en términos del mercado bursátil mexicano. Algoritmos de clasificación de Machine Learning se emplean para pronosticar los movimientos diarios en el precio del Índice de Precios y Cotizaciones de la Bolsa Mexicana de Valores (IPC) con el fin de evaluar la funcionalidad de estos modelos de pronósticos aplicados a mercados emergentes, como es el caso de México. Los resultados son evaluados y comparados para determinar si el uso de estos algoritmos es útil para su aplicación en la toma de decisiones en estrategias de trading. Para este proyecto de investigación se consideró un periodo de estudio con datos correspondientes a 10 años (2 de enero de 2009 al 31 de diciembre de 2018) y una configuración con datos correspondientes a un periodo de 6 años y 4 meses aproximadamente (30 de agosto de 2012 al 31 de diciembre de 2018).

Los modelos desarrollados tienen la meta de pronosticar si el precio de una acción se incrementara o no, al cierre del día. En el caso de obtener resultados positivos contaríamos con un indicador que junto con otras herramientas empleadas actualmente por los inversores nos permita tomar mejores decisiones en operaciones intradiarias de compra y venta.

Para el caso de estudio de esta tesis consideramos el índice de precios y cotizaciones (IPC) para la aplicación de los modelos en su pronóstico. Una predicción precisa del movimiento del IPC puede no solo proporcionar un valor de referencia para que los inversionistas realicen una estrategia efectiva, sino también para que entes reguladores supervisen el mercado.

Como evaluación se propone la simulación (backtesting) de una estrategia de compraventa sencilla para actuar sobre el IPC basada en los pronósticos. Los modelos predicen el cambio entre la apertura y los valores de cierre del día de negociación. Si se pronostica que el mercado suba por cualquier cantidad, esto indica una compra. Por otro lado, si se prevé que caiga, esto señala una venta.

Antecedentes

El pronóstico del precio de las acciones ha generado un gran número de investigaciones aplicando diferentes modelos matemáticos, financieros y económicos, tal es el caso de Vera y Rosado (2010) quienes aplican la simulación Montecarlo para pronosticar el precio de acciones en la Bolsa Mexicana de Valores empleando como datos de entrada factores macroeconómicos, corporativos y de mercado, obteniendo como resultado de la simulación, la capacidad de predecir la tendencia general en los precios de los activos, sin embargo, pronosticando unos precios muy lejanos a los precios reales de las acciones.

En los últimos años las redes neuronales artificiales se han usado para pronosticar diferentes variables en un gran número de disciplinas, como lo es el pronóstico del clima, la velocidad del viento y la demanda en cajeros automáticos y un número importante de estudios ha demostrado que estas son herramientas útiles para el modelado y predicción de series de tiempo.

En el ámbito financiero investigaciones como la de Asil Oztekin (2015) estudian y comparan la eficacia de modelos analíticos modernos, realizando un pronóstico del BIST 100, el principal indicador de la bolsa de valores de Turquía a través de 3 modelos diferentes: redes neuronales artificiales (ANNs), maquinas de vectores de soporte (SVM) y Sistemas Adaptativos de Inferencia Neuro-Difusa (ANFIS). Estos modelos son alimentados con variables económicas y financieras, obteniendo como resultado que las maquinas de vectores de soporte sobrepasan al resto de los modelos con un 72% de probabilidades de acertar.

En su investigación, Kim Kyoung-jae (2003) pronostica el indicador de la bolsa de Valores de Korea empleando las maquinas de soporte de vectores utilizando indicadores técnicos como inputs y obteniendo resultados satisfactorios en la predicción de la dirección futura del índice.

Nicolás Sánchez Anzola (2015) emplea las maquinas de soporte de vectores y redes neuronales artificiales en la predicción del movimiento intradía del dólar y el peso colombiano (USD/COP), utilizando como datos de entrada indicadores de análisis técnico,

obteniendo un porcentaje de éxito de 57.8% para las redes neuronales y un 55.6% para las maquinas de soporte de vectores.

Magnus Olden (2016) aplica técnicas de Machine Learning para pronosticar el movimiento diario de 22 acciones en la bolsa de valores noruega usando 4 años de información, obteniendo como resultado que los mejores algoritmos superan el índice benchmark de Oslo (OBX).
Estudios realizados por Toro [Toro et al. 2006] comparan las redes neuronales y neuro-difusas para pronosticar precios en la bolsa de valores de Colombia, donde concluye que las redes neuronales presentan un mejor desempeño que las redes neuro-difusas, así como un menor tiempo de cálculo.

Behl [Behl et al., 2017] emplea una variedad de modelos de Machine Learning como redes neuronales y arboles de decisión para pronosticar el S&P 500 a través de indicadores económicos y de análisis técnico obteniendo una exactitud entre 70% y 73%.

Huertas (2015) aplica modelos predictivos y Machine Learning en el mercado Forex obteniendo un 60% de precisión al pronosticar los movimientos sobre la cotización EUR/USD.

Capítulo 1 Mercados Financieros: estructura, funcionamiento y análisis

En este capítulo se presenta un resumen con las bases del sistema y los mercados financieros relacionadas a esta tesis haciendo un énfasis en el mercado de capitales, donde se aplica esta investigación.

1.1 El Sistema Financiero

A lo largo del tiempo el mercado bursátil se ha encargado de canalizar el flujo de recursos financieros de la sociedad, fungiendo como una importante opción para los inversionistas que tratan de proteger e incrementar sus recursos y como una fuente de financiamiento para los proyectos de empresas y gobiernos; esta actividad es fundamental para el crecimiento económico de un país ya que permite la generación de empleos, flujo de recursos, infraestructura y posteriormente generar desarrollo económico.

Las empresas que requieren recursos (dinero) para financiar su operación o proyectos de expansión, pueden obtenerlo a través del mercado bursátil, mediante la emisión de valores (acciones, obligaciones, papel comercial, etc.) que son puestos a disposición de los inversionistas (colocados) e intercambiados (comprados y vendidos) en la Bolsa Mexicana, en un mercado transparente de libre competencia y con igualdad de oportunidades para todos sus participantes [Bolsa Mexicana de Valores, 2017].

Los rendimientos ofrecidos por el mercado accionario brindan la posibilidad de obtener una alta rentabilidad[2] aunque a un mayor riesgo, es por esto que, es de suma importancia tomar las decisiones adecuadas al momento de comprar y vender acciones.

Las acciones son la principal forma en que las corporaciones recaudan su capital social y emplean estos recursos para invertirlos en actividades económicas como lo son financiar su operación o proyectos de expansión. El financiamiento bursátil es considerado una alternativa eficaz para que las empresas cubran sus necesidades de capital. Por otra parte,

[2] La Bolsa mexicana de valores cerro 2017 con un rendimiento anual de 8.13%, mientras que en 2018 tuvo un rendimiento negativo de -15.6%.

los inversionistas acuden a las bolsas de valores buscando opciones para proteger y acrecentar su ahorro financiero, sin embargo, debido a la falta de educación financiera en México actualmente pocas personas participan en actividades de inversión dentro de los mercados financieros, mientras que en países desarrollados es una práctica común para los ahorradores[3], es por esto que la creación de nuevas herramientas que complementen a las ya existentes puede traer importantes benéficos, atrayendo a un número mayor de inversores que participen activamente en la bolsa de valores, generando un mayor flujo de recursos económicos hacia las empresas y en general una cultura financiera [Vera y Rosado, 2010].

En cualquier economía, el sistema financiero se constituye por un conjunto de mercados, instituciones y mecanismos legales, cuyo objetivo principal es canalizar de manera eficiente el ahorro generado por unidades económicas con superávit hacia aquellas con déficit [Díaz, 2004].

"El sistema financiero desempeña un papel central en el funcionamiento y desarrollo de la economía. Está integrado principalmente por diferentes intermediarios y mercados financieros, a través de los cuales una variedad de instrumentos moviliza el ahorro hacia sus usos más productivos. Procura la asignación eficiente de recursos entre ahorradores y demandantes de crédito. Un sistema financiero sano requiere, entre otros, de intermediarios eficaces y solventes, de mercados eficientes y completos, y de un marco legal que establezca claramente los derechos y obligaciones de las partes involucradas" [Banco de México, 2018].

El sistema financiero mexicano es el conjunto de personas y organizaciones, tanto públicas como privadas, que captan, administran, regulan y dirigen los recursos financieros que se negocian entre los diversos agentes económicos del país, dentro del marco de la legislación correspondiente. El conjunto de entidades que conforman este sistema se divide en cuatro grupos:

a) Instituciones reguladoras.

---

[3] El estudio "Acciones para democratizar el acceso al mercado bursátil" del 2015 elaborado por el Instituto Mexicano para la Competitividad (IMCO) concluyó que en México únicamente 2.9% de la población invierte en un fondo de inversión y apenas 0.20% participa en el mercado accionario. En contraste, poco más de la mitad de los estadounidenses (52%) invierten hoy en día en el mercado de valores, de acuerdo con datos de la firma de análisis y consultoría en Estados Unidos Gallup.
Fuente: https://www.eleconomista.com.mx/mercados/El-mercado-bursatil-necesita-mas-inversionistas-20171115-0126.html

b) Instituciones financieras o intermediarios financieros.
c) Personas y organizaciones que realizan operaciones con los intermediarios financieros.
d) Organizaciones secundarias, como son las asociaciones de bancos o de aseguradoras.

Entre el amplio grupo de instituciones que conforman el sistema financiero mexicano destacan, de manera preponderante, la banca y la bolsa. El sistema financiero y el mercado de valores con sus mecanismos, instrumentos e instituciones constituyen la puerta de entrada del dinero nacional y externo. Estas instituciones hacen llegar el dinero a las entidades que no lo tienen y que lo requieren para financiar sus necesidades productivas, recomponer sus estructuras, expandirse, crecer, etc. [Guerrero, 2005].

El mercado de valores esta constituido por la Bolsa Mexicana de Valores, las casas de bolsa y las operadoras de sociedades de inversión, todas ellas organizaciones supervisadas por la comisión Nacional Bancaria y de Valores [Díaz y Aguilera, 2005].

1.2 Los Mercados Financieros

Un mercado financiero es un lugar físico o virtual donde concurren oferentes y demandantes de recursos monetarios (dinero), donde se comercian o intercambian activos financieros [Fabozzi et al., 1996].

Court (2010) define a un mercado financiero como un mecanismo que reúne a vendedores y compradores de instrumentos financieros, el cual facilita las transacciones a través de sus sistemas.

Los activos que se negocian en el mercado financiero son activos intangibles cuyo valor se basa en el derecho a obtener una cantidad monetaria futura, “básicamente son documentos legales que representan una inversión o un derecho económico para quien esta entregando el dinero y son un mecanismo de financiación para quien lo esta emitiendo” [Villegas, 2010].

En un mercado financiero los compradores y vendedores de los activos financieros pueden clasificarse en [Mascareñas, 2012]:

a) **El emisor o prestatario**: Es la institución (el Estado, una empresa, etc.) que se compromete a realizar pagos en el futuro a cambio de vender/emitir ahora mismo un activo financiero a cambio de dinero. Es lo que, en términos económicos, se denomina una unidad de gasto con déficit.
b) **El inversor o prestamista**: Es el propietario del activo financiero que, a cambio de entregar dinero al emisor, obtiene el derecho a recibir una cantidad monetaria futura de este. Es lo que, en términos económicos, se denomina una unidad de gasto con superávit. Por supuesto, los inversores también pueden revender los activos financieros que previamente habían adquirido.

1.2.1 Estructura de los mercados financieros

Para ofertar públicamente los valores, la empresa acude a una casa de bolsa, que ofrece los valores a la venta en el llamado mercado primario de la bolsa de valores. El precio de una emisión en el mercado primario es fijo y lo determina la empresa asesorada por la casa de bolsa, antes de salir a la oferta pública. De esta manera, los emisores (la empresa) reciben los recursos resultantes de la venta de los valores que adquieren los inversionistas en el mercado primario. Después que las acciones han sido colocadas entre los inversionistas, éstas pueden ser vendidas y compradas en la bolsa de valores en el llamado mercado secundario, una vez más, a través de una casa de bolsa. En el mercado secundario, el precio de las acciones que se intercambian lo deciden en forma individual los compradores y los vendedores, mediante las posturas respectivas [Díaz, 2004].

Otra forma de distinguir entre mercados es sobre la base del vencimiento de los valores negociados en cada mercado. El mercado de dinero es un mercado financiero donde solo se negocian instrumentos de deuda a corto plazo (por lo general, aquellos que tienen vencimiento menor a un año); El mercado de capital es el mercado donde se negocian instrumentos de deuda a largo plazo (por lo general, aquellos que tienen vencimiento de un año o más).

Los valores a corto plazo tienen menores fluctuaciones de precios que los valores a largo plazo, lo cual los vuelve inversiones mas seguras. Algunos ejemplos de los instrumentos negociados en el mercado de dinero son los Certificados de la tesorería de EUA

(instrumentos emitidos para financiar al gobierno federal), Certificados de depósitos bancarios negociables, Papel comercial, Fondos federales (FED) y CETES en el caso de México [Mishkin, 2014]. Estos instrumentos, también llamados Instrumentos de renta fija generan unos flujos de dinero conocidos a lo largo del tiempo, lo que permite calcular la rentabilidad aproximada de la inversión. Se los denomina “bonos u obligaciones”, y los flujos que producen a lo largo del tiempo son llamados “cupones”. El hecho de que se llamen “instrumentos de renta fija” no indica que la rentabilidad sea siempre positiva y fija; de hecho, puede darse que una inversión de este tipo provoque rentabilidades diferentes de las esperadas, o incluso pérdidas [Court, 2010].

Los instrumentos del mercado de capital tienen fluctuaciones de precios mucho más amplias que los instrumentos del mercado de dinero y se consideran inversiones bastantes riesgosas. Estos instrumentos, también conocidos como Instrumentos de renta variable son aquellos en los que la rentabilidad de la inversión se origina principalmente a partir de un incremento en el precio futuro, el cual no es conocido, lo que genera una incertidumbre sobre las rentabilidades futuras. Así también, dependiendo de la utilidad del ejercicio, estos instrumentos pagan periódicamente un monto de dinero, conocido como “dividendos” [Mishkin, 2014] [Court, 2010].

Algunos de los activos financieros mas importantes que se negocian en los mercados son [Villegas, 2010]:

- **Acciones.** Se trata de títulos representativos del valor de una de las fracciones iguales en que se divide el capital social de una empresa. Las acciones de muchas compañías se transan en las bolsas de valores, que son mercados financieros especializados en este y otros tipos de activos financieros.

- **ETF´s.** Exchange Traded Fund, es un fondo que toma posiciones sobre un índice o sobre una canasta de acciones, es decir, el ETF va a comprar los activos que conforman el índice, en la misma proporción en la que el índice las tiene consideradas. Este fondo se puede negociar como si fuera una acción individual.

- **Contratos de futuros y opciones.** Estos activos pertenecen a los llamados derivados financieros, que son productos cuyo valor se basa en el precio de otro

activo. El activo del que depende se llama activo subyacente. Los subyacentes que se usan pueden ser de muchas clases: acciones, índices bursátiles, títulos de renta fija, tasas de interés o materias primas. Se comprende entonces que la característica fundamental de los derivados financieros es que su valor cambia en respuesta a los cambios de precio del activo subyacente. Los derivados financieros pueden negociarse en las bolsas de valores o también en mercados no organizados (en inglés se conocen como OTC, por Over The Counter). Un contrato de futuros es un acuerdo o contrato que obliga a las partes contratantes a comprar o vender un número determinado de unidades del activo subyacente en una fecha futura determinada. El punto clave es que el precio se define de antemano, fungiendo como una medida de cobertura ante la incertidumbre que se pueda presenta en el mercado. En las opciones el contratante tiene la posibilidad, pero no la obligación, de comprar o vender un número determinado de unidades del activo subyacente en una fecha futura determinada.

- **Monedas.** Aunque las monedas (o divisas) no son un activo financiero propiamente dicho, pues una empresa, una corporación, un banco o un individuo no pueden crear o emitir una moneda, sí pueden ser negociadas en los mercados financieros. El término “Forex” o también FX es la abreviación de “Foreign Exchange Market”, siendo su homónimo español mercado de divisas. Este mercado se basa en el intercambio de una divisa por otra, siendo con diferencia el mercado más grande del mundo con una media de negociación diaria de cuatro trillones de dólares.

Todo activo financiero viene caracterizado por tres variables: la rentabilidad, el riesgo y la liquidez [Mascareñas,2012]:

El **rendimiento** se refiere al beneficio que se espera genere al inversor, el activo financiero que ha adquirido, a lo largo de un periodo de tiempo concreto con relación al coste de su inversión.
El **riesgo** es la posibilidad de que el activo financiero al final genere un rendimiento diferente (superior o inferior) al prometido.
La **liquidez** implica la posibilidad de convertir en dinero líquido un activo como el que sea innecesario rebajar su precio para poder venderlo (sin perdida de valor). Los activos financieros que cotizan en mercados financieros organizados son mas líquidos que los que no cotizan.

1.3 Índices accionarios / Índice de Precios y Cotizaciones (IPC)

Los índices han surgido gracias a la necesidad de medir el desempeño de los mercados financieros, se emplean como un apoyo para conocer su comportamiento y estabilidad e incluso medir la salud de la economía de un país, un segmento de mercado o alguna clase de activos.

En general los índices tratan de medir el mercado mediante un sistema de puntos, que subirán o bajaran en la medida en que suban o bajen los precios de los activos que lo compone. Los índices modernos asignan un puntaje diferente a cada uno de los activos que lo componen, según el tamaño y la importancia de la empresa que los representa [Villegas, 2010].

Estos índices son un registro estadístico, que trata de reflejar los rendimientos de las acciones que lo integran. Las acciones que lo componen presentan rasgos parecidos, como un volumen de operación similar, pertenecer a un mismo sector industrial o a un mercado geográfico, nivel de capitalización, etc.

El uso de los índices nos permite analizar en una manera resumida el comportamiento general de los precios en una cifra que sea de fácil lectura, análisis y comprensión en diferentes periodos históricos.

En el mercado mundial cada bolsa de valores tiene uno o varios índices que la miden, algunos de estos son: el MERVAL de Argentina, BOVESPA en Brasil, IBEX 35 de España, NIKKEI en Japón, DAX 30 en Alemania, entre otros. En los Estados unidos, donde hay muchos mercados y miles de activos para operar existen muchos índices que tratan de monitorizar la economía. Los más populares son el Dow-Jones, con 30 acciones; el NASDAQ, que incluye 100 acciones; y el Standard and Poors, que se compone de 500 empresas. Por otra parte, debido al tamaño de los mercados mundiales, se han desarrollado índices que solo incluyen acciones de empresas de determinados sectores de la economía, como el tecnológico, el industrial, el de la construcción, etc. [Villegas, 2010].

“Los Índices de la Bolsa Mexicana de Valores, dependiendo de su enfoque y especialidad, son indicadores que buscan reflejar el comportamiento del mercado accionario mexicano en su conjunto, o bien de diferentes grupos de empresas con alguna característica en

común. El Índice De Precios Y Cotizaciones (IPC), con base octubre de 1978, tiene como principal objetivo, constituirse como un indicador representativo del Mercado Mexicano para servir como referencia y subyacente de productos financieros" [Bolsa Mexicana de Valores, 2018].

El índice de precios y cotizaciones es el principal indicador bursátil de la Bolsa Mexicana de Valores. Expresa el rendimiento del mercado accionario, en función de las variaciones de los precios y volúmenes negociados de una muestra de las 35 acciones más representativas cotizadas en la BMV. La muestra de acciones contenida en este índice se selecciona cada dos meses de acuerdo con su nivel de capitalización y de bursatilidad [Díaz y Aguilera, 2005].

El nivel de bursatilidad es la facilidad para comprar o vender una acción en un momento determinado, y toma en cuenta variables como el número de operaciones, el importe negociado, los días operados, y la relación entre el monto operado y el monto suscrito. El IPC reúne las acciones que más actividad registran en todos los sentidos en la Bolsa Mexicana de Valores, con el fin de condensar en una cifra el comportamiento general de todo el mercado.

El nivel de capitalización indica el valor que da el mercado a una empresa que cotiza en bolsa, y se calcula multiplicando el precio que tienen sus acciones por la cantidad de ellas que han sido emitidas [Kuspit, 2018].

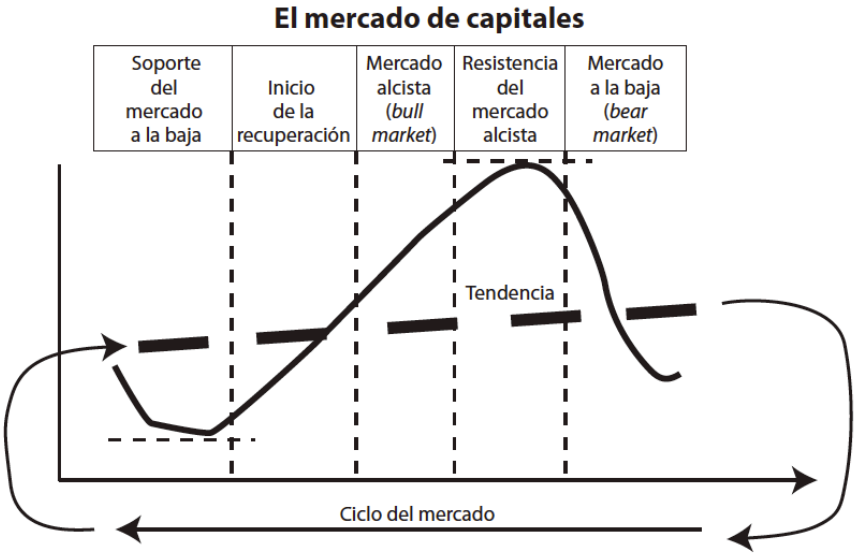
Por lo tanto, en términos generales, son las acciones que tienen mayor liquidez y que se negocian con mayor facilidad las que se integran en el IPC. El índice se modifica cuando las condiciones de bursatilidad de un título se reducen de manera significativa, lo que provoca su cambio por otro que tenga mayor operatividad en el mercado.

El IPC es un índice ponderado. Esto quiere decir que el comportamiento de las acciones de las empresas más grandes tiene mayor impacto en el índice que el de las acciones de las más pequeñas.

El mercado de valores presenta un comportamiento cíclico. En él se suceden periodos alcistas y periodos a la baja. El comportamiento del IPC refleja la tendencia del mercado,

marca cuándo se encuentra al alza (bull market) y cuándo se encuentra a la baja (bear market) como se muestra en la siguiente figura [Díaz y Aguilera, 2005].

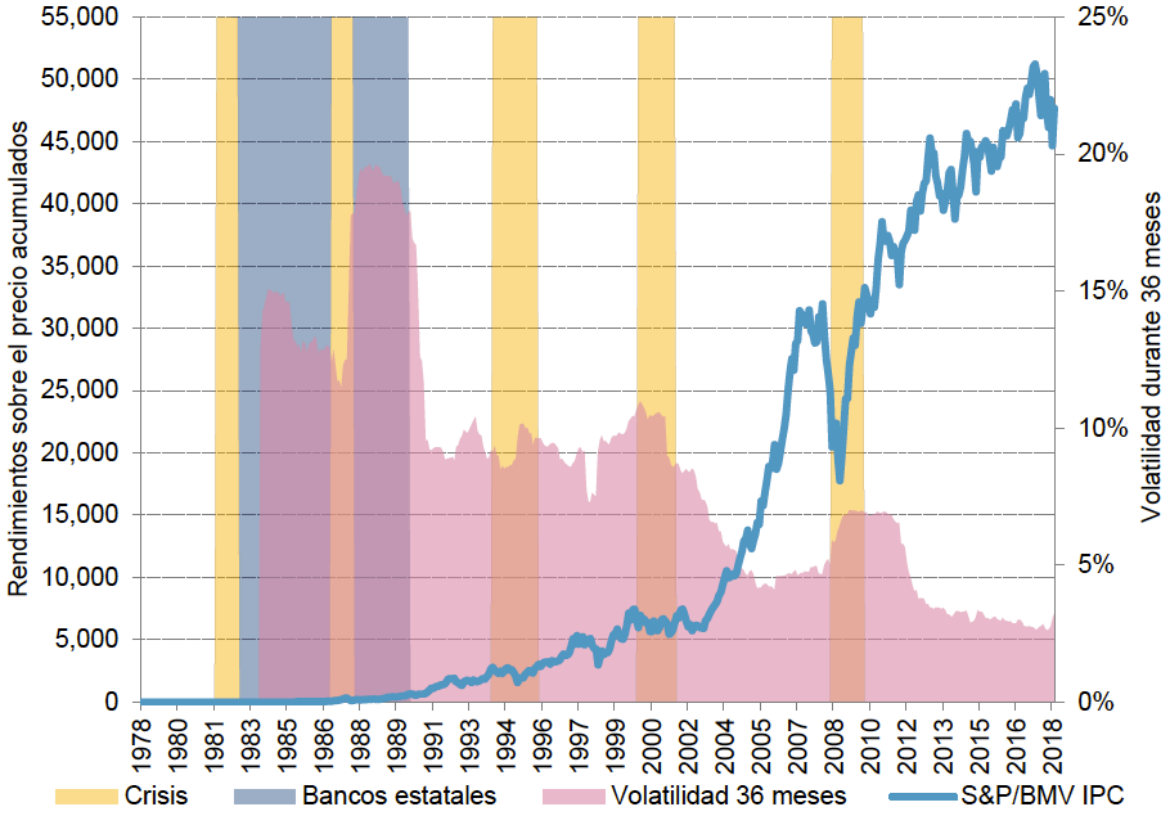
Figura 1.1 El comportamiento cíclico del mercado de valores.



Fuente: Díaz y Aguilera, 2005

Desde su creación en 1978 el IPC ha experimentado algunas reducciones significativas en su historia. Entre abril de 1979 y mayo de 1982, la baja máxima fue de -70.30%, de la cual le llevó 53 meses recuperarse hasta septiembre de 1983. La segunda mayor reducción de -69.24% ocurrió en un período más corto, entre septiembre de 1987 y diciembre de 1987. Al índice le llevó 21 meses recuperarse hasta junio de 1989. Ambos períodos coincidieron con tiempos de crisis local. La tercera mayor reducción de -44.48% fue durante la crisis financiera mundial, entre mayo de 2008 y febrero de 2009. El índice se recuperó en diciembre de 2009 [Sánchez, 2018].

Figura 1.2 Períodos de crisis – Volatilidad y rendimientos acumulados del IPC.



Fuente: S&P Dow Jones Indices LLC, 2019.

La siguiente gráfica muestra el precio de cierre diario del IPC para un periodo de 10 años, en los cuales a presentado rendimientos anuales mixtos.

Figura 1.3 Rendimiento histórico IPC.

Fuente: Elaboración propia con información de S&P Dow Jones Indices.

La tabla 1.1 presenta los perfiles de riesgo y rendimiento del IPC en los últimos 10 años. En este periodo el mejor rendimiento se obtuvo en el año 2009 con 43.52%, mientras que el peor rendimiento se dio el año 2018 con -15.63%.

Tabla 1.1 Rendimientos y riesgos del IPC.

RENDIMIENTO

NIVEL	RENDIMIENTOS			RENDIMIENTOS ANUALIZADOS			
	1 MES	3 MESES	YTD	1 AÑO	3 AÑOS	5 AÑOS	10 AÑOS
RENDIMIENTO TOTAL							
59,976.42	5.65%	0.78%	5.65%	-10.76%	2.41%	3.36%	10.39%
RENDIMIENTO SOBRE EL PRECIO							
43,987.94	5.64%	0.10%	5.64%	-12.82%	0.27%	1.48%	8.44%

RENDIMIENTO AÑO CALENDARIO

2018	2017	2016	2015	2014	2013	2012	2011	2010	2009
RENDIMIENTO TOTAL									
-13.62%	10.49%	8.15%	1.46%	1.99%	-0.02%	19.71%	-2.16%	21.61%	46.17%
RENDIMIENTO SOBRE EL PRECIO									
-15.63%	8.13%	6.20%	-0.39%	0.98%	-2.24%	17.88%	-3.82%	20.02%	43.52%

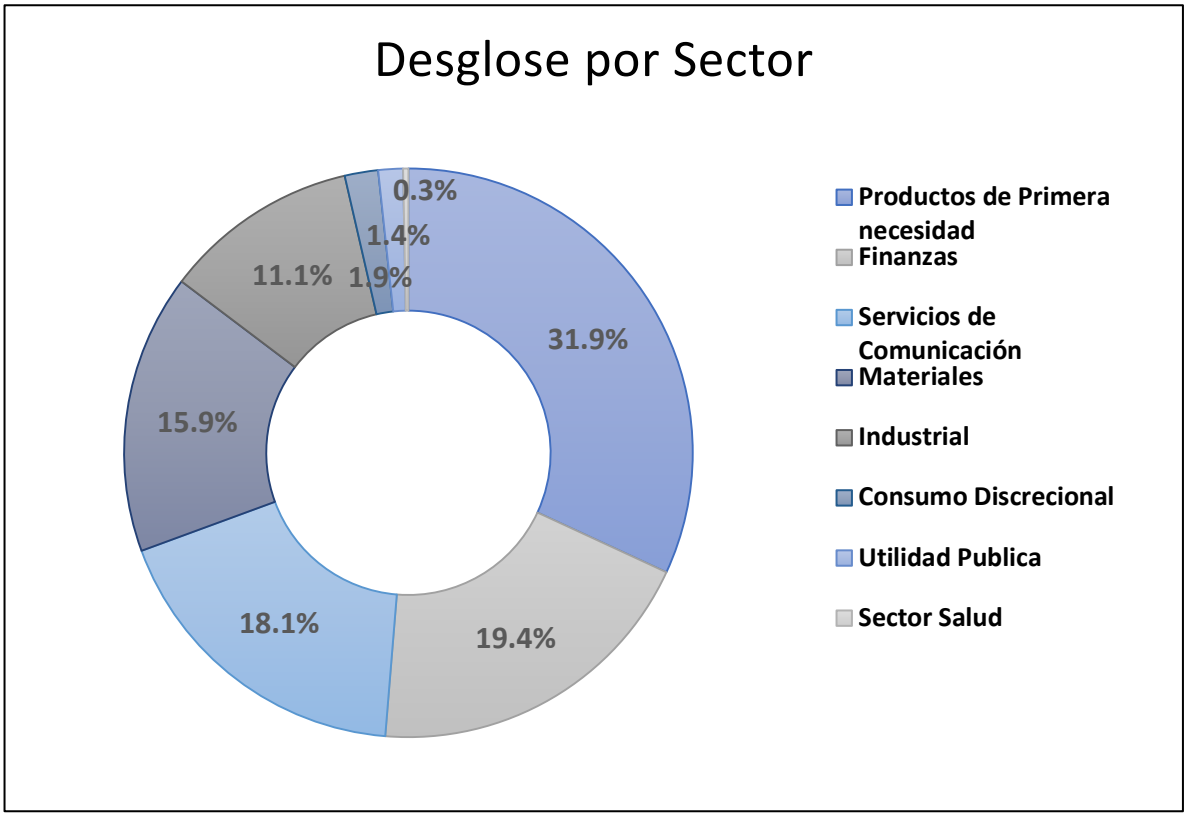
RIESGO

RIESGO ANUALIZADO			RENDIMIENTOS ANUALIZADOS AJUSTADOS POR RIESGO		
3 AÑOS	5AÑOS	10 AÑOS	3 AÑOS	5 AÑOS	10 AÑOS
DESVIACIÓN ESTÁNDAR					
13.39%	12.35%	14.04%	0.18	0.27	0.74

Fuente: S&P Dow Jones Indices LLC, 2019.

El siguiente grafico muestra la composición sectorial del IPC creada con el fin de describir la economía mexicana de manera general. Ocho sectores están representados en el IPC, considerando su última estructura. El sector de Productos de Primera Necesidad tiene la mayor ponderación en el índice con nueve compañías que representan 31.9% del indicador. El segundo sector de mayor tamaño es Finanzas con siete compañías que representan 19.4% del índice; seguido por Servicios de Comunicación con un 18.0% y tres empresas. El sector de Materiales representa 15.9% con seis empresas, mientras que el sector Industrial tiene seis compañías, pero un peso de 11.1%. Consumo Discrecional aporta 1.9% con dos compañías, Servicios de Utilidad Pública representa 1.4% con una sola empresa y, finalmente, el sector de Salud aporta el 0.3% restante con solo una compañía.

Figura 1.4 Composición del IPC por Sector



Fuente: Elaboración propia con información de S&P Dow Jones Indices.

Tabla 1.2 Principales componentes del IPC.

Los 10 componentes principales por ponderación

COMPONENTE	TICKER	SECTOR
América Móvil SAB de CV L	AMXL	Servicios de Comunicación
Fomento Económico Mexicano S.A.B. de C.V.	FEMSA UBD	Productos de Primera necesidad
Grupo financiero Banorte O	GFNORTE O	Finanzas
Wal-Mart de México SAB de CV	WALMEX*	Productos de Primera necesidad
Grupo México SAB de CV B	GMEXICO B	Materiales
Cemex SA CPO	CEMEX CPO	Materiales
Grupo Televisa SAB CPO	TLEVISA CPO	Servicios de Comunicación
Alfa SA A	ALFA A	Industrial
Grupo Bimbo S.A.B.	BIMBO A	Productos de Primera necesidad
Grupo Aeroportuario del Sureste SAB de CV B	ASUR B	Industrial

Fuente: S&P Dow Jones Indices LLC, 2019.

1.4 Pronósticos y el precio de las acciones

Las decisiones, difícilmente, se toman sin contar con algún tipo de pronóstico. Diversos autores definen los pronósticos como el estudio de datos históricos para descubrir sus patrones y tendencias fundamentales. Este conocimiento se utiliza para proyectar los datos a periodos futuros como pronósticos [Hanke, 1996].

En general los métodos para obtener un pronóstico se basan en técnicas subjetivas como la intuición, juicio personal, experiencia o la opinión de un experto, por otra parte, encontramos las técnicas de pronóstico cuantitativas, basadas en estadística, como los análisis de series de tiempo y de regresión. Una serie de tiempo estadística es un conjunto de valores numéricos de una variable aleatoria a lo largo de un horizonte de tiempo y pueden ser de los siguientes tipos: modelos de nivel constante, de nivel constante con efectos estacionales y con tendencia lineal. En el análisis de regresión, la variable que se va a pronosticar (dependiente) se expresa como una función matemática de otras variables (independientes). Este tipo de metodologías opera bien cuando el fenómeno en estudio presenta una tendencia de ajuste suave, es decir, cuando los datos en un intervalo no presentan cambios bruscos de pendiente [Toro, et.al, 2005].

La creciente complejidad en el mundo de los negocios ha generado una necesidad en las organizaciones por asegurar, a medida de lo posible, el futuro. Con el desarrollo de técnicas modernas traídas con el progreso de las tecnologías, el proceso de los pronósticos ha recibido mayor atención durante los años recientes, adquiriendo así una importante posición en la administración de las organizaciones.

Las facilidades tecnológicas permiten que, en la actualidad, cualquier administrador posea la capacidad de utilizar técnicas de análisis de datos complejas para fines de pronóstico, y una comprensión de dichas técnicas es esencial hoy en día para cualquier compañía.

Los sistemas inteligentes han surgido como una combinación de las técnicas cualitativas y cuantitativas, lo que nos permite evaluar variables con tendencias fluctuantes; en esta situación, las demás metodologías no son tan efectivas, si se compara el error medio cuadrático entre ellas.

A pesar del uso de técnicas modernas en los pronósticos, es raro que estos coincidan al pie de la letra con el futuro, es por esto por lo que se continúan desarrollando nuevas estrategias enfocadas de manera particular en los errores. Los errores son inevitables en cualquier procedimiento de pronóstico, sin embargo, el éxito de estos radica en intentar que sean tan pequeños como sea posible. [Hanke, 1996].

En la búsqueda por tener mejores resultados en los pronósticos surgen los modelos predictivos, estos reúnen gestión, tecnologías de la información, estadística y modelado, orientándose al trabajo con grandes volúmenes de datos y con el objetivo de realizar mejores predicciones basadas en un intento de relacionar un conjunto de variables con otro, extraer patrones de comportamiento, predecir tendencias e identificar riesgos y oportunidades. Las aplicaciones de los modelos predictivos han permitido que en la actualidad contemos con técnicas como la minería de datos y el aprendizaje automático.

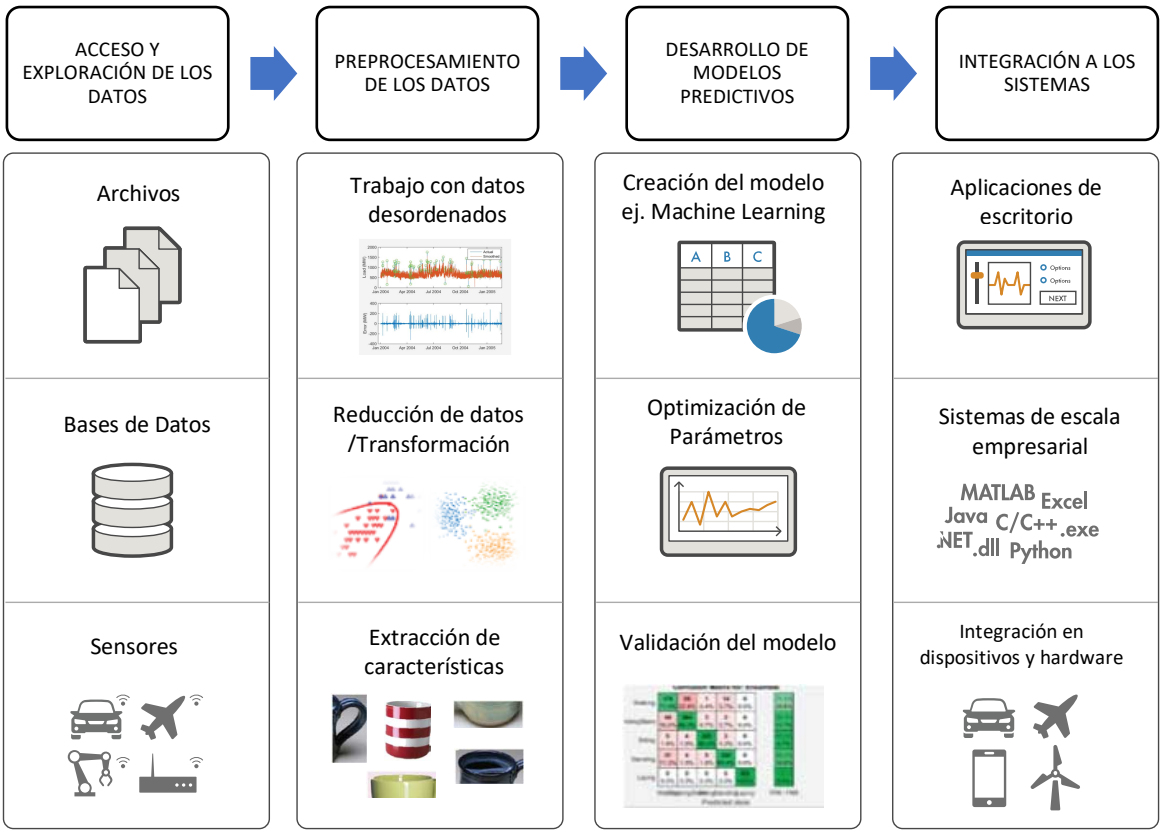
El análisis predictivo emplea datos históricos para predecir eventos futuros. Normalmente, los datos históricos se utilizan para crear un modelo matemático que capture las tendencias importantes. Este modelo predictivo se usa entonces con los datos actuales para predecir lo que pasará a continuación, o bien para sugerir acciones que llevar a cabo con el fin de obtener resultados óptimos.

El modelado predictivo utiliza métodos matemáticos y de cálculo para predecir un evento o un resultado. Estos modelos pronostican un resultado en algún estado o tiempo futuros en función de los cambios en las entradas del modelo. Mediante un proceso iterativo, se desarrolla el modelo mediante un conjunto de datos de entrenamiento y después se prueba y se valida para determinar su precisión con el fin de realizar predicciones. Normalmente, el flujo de trabajo para una aplicación de análisis predictivo incluye los siguientes pasos básicos [Matlab, 2019].

1. Importar datos de varias fuentes, tales como archivos web, bases de datos y hojas de cálculo.
2. Limpiar los datos mediante la eliminación de los valores atípicos y la combinación de las fuentes de datos.
3. Desarrollar un modelo predictivo preciso basado en los datos agregados mediante estadísticas, herramientas de ajuste de curvas o aprendizaje automático.

4. Integrar el modelo en un sistema de predicción de la carga en un entorno de producción.

Figura 1.5 Flujo de trabajo de análisis predictivo.



Fuente: Matlab, 2019

El modelado predictivo en el trading es un proceso en el que predecimos la probabilidad de un resultado utilizando un conjunto de variables predictoras. Se pueden construir modelos predictivos para diferentes activos como acciones, futuros, monedas, commodities, etc. Por ejemplo, podemos construir un modelo para predecir el cambio de precio al día siguiente para una acción, o un modelo para predecir los tipos de cambio de una moneda extranjera. El resultado que obtenemos de los modelos predictivos puede aprovecharse para tomar las decisiones de inversión correctas y para crear carteras rentables.

1.4.1 El precio de las acciones

Los mercados financieros tienen un comportamiento que, aunque puede parecer caótico (errático) al observador casual, se rigen por una serie de principios fundamentales que los hacen sujetos de estudio cuidadoso, medibles, analizables y de alguna manera predecibles.

En cualquier caso, si aceptamos el comportamiento caótico de los mercados financieros, podemos utilizar herramientas estadísticas para estudiar y predecir el comportamiento de los mercados, al menos en términos de probabilidades. Los mercados bursátiles no son entonces impredecibles, aleatorios ni imposibles de estudiar en modelos matemáticos. Por el contrario, todos los activos, en especial los muy líquidos y no manipulables, son sujetos de estudio ideales en la estadística moderna [Villegas, 2010].

Existen un gran número de variables que pueden afectar el precio de las acciones, David Cutler en su investigación “What moves stock prices” determina que las noticias macroeconómicas no pueden explicar mas que una tercera parte de su varianza además de que eventos políticos o noticias de diversa índole siempre afectan a los precios.

Sin embargo, no es solo una noticia, es la reacción a la noticia que también juega un papel importante en el precio de la acción. La reacción a las noticias depende de diferentes factores, así como de los propios mercados. En tiempos de alta volatilidad y ambigüedad (altas incertidumbres sobre la volatilidad), el mercado reacciona más fuertemente a las malas noticias que a las buenas. Además, el mercado puede volverse demasiado optimista o pesimista a veces [Pachanekar, 2018].

Los tipos de cambio, las variaciones en los precios de los commodities, los volúmenes de negociación de acciones e incluso noticias particulares sobre las compañías son algunas de las tantas fuerzas que afectan diariamente el precio que puede tener una acción.

Por lo tanto, parte del trabajo de un inversionista es evaluar cada nueva información y juzgar la posible reacción del mercado teniendo en cuenta el entorno empresarial, así como los factores macroeconómicos que podrían resultar en un ajuste en precio de las acciones de una empresa.

En términos simples, el mercado estima el valor presente de la empresa teniendo en cuenta sus ganancias futuras previstas. Sin embargo, una de las características importantes de un mercado es que este se corrige periódicamente.

1.4.2 Hipótesis de los mercados eficientes

Muchos especialistas han adoptado la hipótesis de los mercados eficientes creada por Eugene Fama en 1970. De acuerdo con esta hipótesis, un mercado es eficiente cuando los inversores negocian de forma racional utilizando información disponible, gratuita y al alcance de todos. Los precios actuales del mercado reflejan toda la información disponible sin dejar lugar a la especulación y establece que los cambios en los precios son independientes al pasado siguiendo una caminata aleatoria, por lo que la información histórica no tiene valor para pronosticar cambios en los precios.

De acuerdo con esta hipótesis, los ajustes en los precios son impredecibles y pronosticarlos es una tarea imposible de realizar. Cualquier ajuste en los precios representa la reacción inmediata a nuevos eventos o a nuevos e inesperados cambios en cifras de oferta y demanda. Si hubiera una oportunidad de beneficio esperado, los inversores la explotarían de manera inmediata, de manera que esto llevaría el precio de vuelta a un nivel donde ya no es rentable [Abu-Mostafa y Atiya, 1996].

La hipótesis de mercado eficiente establece que el precio de la acción refleja toda la información disponible, pero existe una brecha de tiempo antes de que el mercado se corrija a sí mismo debido a la nueva información que se procesa. Aquí se encuentra el concepto que los daytraders llaman, “cronometrar el mercado”. Saber cuándo entrar y salir es importante para un trader, y aquí las noticias juegan un papel vital [Pachanekar, 2018].

A pesar de que ha existido un gran debate acerca de la hipótesis de mercados eficientes, es difícil tanto probarla como refutarla. La experiencia en los mercados nos dice que estos, son de cierta manera predecibles. La existencia de tendencias en los precios en los mercados financieros y la correlación presente entre eventos fundamentales y cifras económicas que afectan al mercado, son algunas de tantas evidencias que existen en contra de la hipótesis de los mercados eficientes [Abu-Mostafa y Atiya, 1996].

Si bien, no sabemos como se comportará el mercado ante determinadas situaciones, existe una relación causa-efecto implícita en el mercado, por lo que este, puede reaccionar ante ciertos “estímulos”, es decir, este proceso no es completamente aleatorio y existe un proceso que explica los datos que observamos.

Aunque existen datos que pueden explicar el comportamiento del mercado y datos del mismo mercado que reflejan patrones que se repiten históricamente, es imposible identificar el proceso que afecta el cambio en el precio de un activo en su totalidad, sin embargo, identificando esas regularidades y relaciones se ha logrado desarrollar diferentes técnicas que nos permiten construir buenas aproximaciones.

1.4.3 Análisis técnico y fundamental

Históricamente los análisis mas empleados por los especialistas para considerar invertir en una acción han sido el análisis fundamental y el análisis técnico.

En el análisis fundamental los inversionistas se enfocan en el valor intrínseco de las acciones, el desempeño de la industria y la economía, el clima político, etc. Se basa en el estudio de las cuentas de las empresas, analizando su activos y pasivos, estudiando el comportamiento de sus beneficios, y comparando su estructura financiera con las de otras empresas del sector. Contempla que las noticias y eventos de factor económico influencian la oferta y la demanda de una empresa.

El análisis técnico se enfoca en el movimiento que tiene el precio de una acción, construyendo hipótesis a partir de precios y volúmenes históricos, calculando indicadores, osciladores y gráficos que ayudan a identificar patrones y tendencias que sugieren como se comportara el precio de un activo en el futuro. Estos gráficos buscan reflejar la interacción que existe entre la oferta y la demanda de un activo y considera que los patrones recientes en el precio pueden determinar una dirección futura del mismo. El análisis técnico busca correlaciones entre ciertos patrones y las posteriores direcciones que tomó el mercado, algunos expertos mencionan que el “precio de un activo tiene memoria” ya que es muy habitual que cuando llegue a un cierto nivel se comportará como ya lo ha hecho antes en ese nivel.

Capítulo 2 Ciencia de Datos y Machine Learning

Solo en los últimos dos años se generó el 90 por ciento de los datos en el mundo y actualmente creamos 2.5 quintillones de bytes de datos cada día, pero ese ritmo solo se está acelerando con el crecimiento del Internet de las Cosas [Forbes, 2018].

De acuerdo con estudios realizados, durante 2017 más de 15.2 millones de mensajes de texto fueron enviados cada minuto. 103,447,520 correos electrónicos no deseados se envían cada minuto, 3.6 millones de búsquedas en Google, Spotify agregó 13 nuevas canciones, tuiteamos 456,000 veces, se realizaron 45,787 viajes en Uber cada minuto, publicamos 46,740 fotos de Instagram y publicamos 600 nuevas ediciones de página en Wikipedia cada minuto [Domo.com, 2017].

Este rápido desarrollo e integración de datos provee a los científicos, ingenieros y en general, gente de negocios con un vasto recurso que puede ser analizado para hacer descubrimientos científicos, optimizar sistemas industriales y descubrir valiosos patrones financieros.

Para emprender estos grandes proyectos de análisis de datos los investigadores y los profesionales han adoptado algoritmos establecidos a partir de estadísticas, minería de datos, aprendizaje automático (Machine Learning) y Big Data, vitales para la cantidad de datos nuevos que las compañías e instituciones generan día a día, creando un ambiente en el cual, cualquier organización tiene una mayor facilidad y una amplia gama de opciones para obtener y analizar información valiosa acerca de sus operaciones y su entorno.

La acumulación de grandes volúmenes de datos en las organizaciones ha traído consigo una nueva forma de hacer negocios, permitiendo que el cliente que antes era anónimo se hiciera visible e identificable, de esta manera las organizaciones se han enfocado en desarrollar modelos de negocio orientados al cumplimiento de las necesidades particulares de sus clientes, adelantándose a los requerimientos futuros de sus clientes [Palma, 2009].

En el caso particular de las bolsas de valores actualmente contamos con el desarrollo de herramientas que nos permite acceder de manera rápida y sencilla a información bursátil, analizar el comportamiento del precio de las acciones, paginas web y software que nos

permiten visualizar gráficos basados en análisis técnico, todo esto brindándonos un importante soporte para la toma de decisiones.

El Algorithmic Trading ha tenido un desarrollo importante y se ha convertido en una de las herramientas mas utilizadas por las corporaciones para realizar compra y venta de instrumentos financieros de manera automática con base en parámetros definidos previamente.

Algorithmic trading es un sistema computarizado responsable de ejecutar ordenes de compra y venta de un activo, sustituyendo el trabajo manual de un trader. Este programa computacional sigue reglas predeterminadas para determinar como se debería de ejecutar cada orden. Basado en estas reglas, puede, por ejemplo, dividir porciones o pequeñas ordenes para enviar al mercado, dando seguimiento a las condiciones del mercado y posibles eventos [Johnson, 2010].

En la última década la prestigiosa empresa de Wall Street Goldam Sachs cambió su estructura, donde antes contaban con 600 traders comprando y vendiendo acciones ahora tan solo emplean a 200 ingenieros en computación y tan solo 2 traders apoyando sus estrategias. De acuerdo con la firma de investigación inglesa “Coalition” Hoy en día, casi el 45 por ciento de los ingresos generados con el comercio de acciones proviene de transacciones electrónicas.

2.1 Minería de datos

Actualmente la complejidad en las variables que afectan a los mercados y la volatilidad son un tema que emerge de forma recurrente creando esto, la necesidad de aplicar e integrar modelos cada vez más sofisticados que faciliten el desempeño de las organizaciones públicas y privadas.

La ciencia de extraer información útil de grandes conjuntos o bases de datos se conoce como minería de datos y uno de sus principales objetivos es el de explorar y descubrir hechos significativos en la historia de una organización que permitan vislumbrar, explicar y pronosticar el comportamiento de sus variables. De esta manera la minería de datos hace posible estimar cómo será el comportamiento de las empresas en un futuro cercano. Palma

(2009) afirma que: “Data Mining es *anticipar*” y lo define como “la explotación de bases de datos corporativas que registran características personales, familiares, socioeconómicas, conductas de compra y conductas de pago”,

Zaki y Meira (2014) definen a la minería de datos como el proceso de descubrir patrones internos y de interés, así como modelos descriptivos, entendibles y predictivos de una fuente de datos a gran escala

De esta manera podemos entender a la minería de datos como la extracción de datos previamente desconocidos y con un uso potencial a través tecnologías las cuales nos permiten buscar regularidades o patrones en estos datos. Estos patrones nos pueden ayudar a generar predicciones sobre datos futuros donde el descubrimiento de una nueva oportunidad puede ser la base de la construcción de una nueva ventaja competitiva.

Las técnicas y métodos desarrollados a través de la minería de datos buscan analizar de manera combinada la información generada de manera interna por las organizaciones con la información obtenida a través de fuentes internas como las diferentes variables económicas que podemos encontrar en el mercado. Los datos se revisan en busca de patrones y luego se aplican los criterios para determinar las relaciones más frecuentes e importantes. El objetivo de la minería de datos es derivar un valor de negocio a partir de estos patrones y relaciones nunca vistas en grandes conjuntos de datos.

Las técnicas de minería de datos en conjunto con la inteligencia artificial se utilizan para construir modelos de aprendizaje automático (Machine Learning).

El aprendizaje automático y la minería de datos utilizan los mismos algoritmos clave para descubrir patrones en los datos. Sin embargo, su proceso, y en consecuencia la utilidad, difieren. A diferencia de la minería de datos, en el aprendizaje automático, la máquina debe aprender automáticamente los parámetros de los modelos a partir de los datos. El aprendizaje automático utiliza algoritmos de autoaprendizaje para mejorar su rendimiento en una tarea con experiencia a lo largo del tiempo. Se puede utilizar para revelar información y proporcionar comentarios casi en tiempo real [Brooks et al., 2017].

El aprendizaje automático y la minería de datos son áreas de investigación de ciencias de la computación cuyo rápido desarrollo se debe a los avances en la investigación de análisis

de datos, el crecimiento en la industria de bases de datos y las necesidades del mercado resultante de métodos que son capaces de extraer valiosos conocimientos de grandes almacenes de datos [Fürnkranz et al., 2012].

El éxito en las técnicas de minería de datos ha generado la creación de metodologías o procesos que fungen como estándares en la industria, los cuales buscan definir una serie de pasos secuenciales que pretenden guiar la implementación de aplicaciones, no solo para proyectos de minería de datos, si no que también aplicables a modelos de Machine Learning y Big Data. En la actualidad las metodologías más populares son SEMMA y CRISP-DM.

La metodología SEMMA (Sample, Explore, Modify, Model, Assess) considera un proceso de 5 etapas: Extracción de una muestra representativa de los datos, exploración de los datos buscando tendencias y anomalías, modificación de los datos, creando, seleccionando y transformando las variables de acuerdo con el modelo que se aplicara, modelado de los datos permitiendo que el software busque automáticamente una combinación de datos que genere el resultado deseado de manera confiable y por último la evaluación de los datos para determinar el desempeño de la aplicación.

La metodología CRISP-DM (Cross Industry Standard Process for Data Mining) consiste en un ciclo de 6 etapas, al ser esta la metodología seleccionada para realizar este proyecto de investigación se revisará a fondo en el siguiente apartado.

2.1.1 Metodología CRISP-DM

El manejo de la información requiere de una metodología adecuada para el análisis de datos, para ello utilizaremos CRISP-DM (Cross Industry Standard Process for Data Mining) por ser considerado el método más empleado para proyectos de este tipo.

La metodología CRISP fue diseñada para brindar una guía a través de un modelo genérico que pudiera acoplarse a las necesidades de cualquier proyecto de minería de datos y análisis de datos.

Esta metodología consta de 6 fases (Molero,2008):

Figura 2.1 Fases de la metodología CRISP-DM.

Fuente: Elaboración propia

1. **Comprensión del negocio.** En esta fase inicial se debe comprender con claridad los objetivos y requerimientos del proyecto, con la finalidad de elaborar una buena planeación en el desarrollo. Esta fase aglutina las tareas de comprensión de los objetivos y requisitos del proyecto desde una perspectiva empresarial o institucional, con el fin de convertirlos en objetivos técnicos y en un plan de proyecto. Sin lograr comprender dichos objetivos, ningún algoritmo por muy sofisticado que sea permitirá obtener resultados fiables. Esta fase incluye tareas como Determinar los objetivos, Evaluar la situación, y la elaboración de un plan del proyecto. [Gallardo, 2009].

2. **Comprensión de los datos.** Se establece el contacto directo con el problema. Las actividades por realizar son: la recolección inicial de datos, la identificación de la calidad de los datos y el establecimiento de posibles relaciones más evidentes que permitan obtener las primeras hipótesis. Las principales tareas para desarrollar en esta etapa son: Recolección de datos iniciales, Descripción de los datos, Exploración de los datos y verificación de la calidad de los datos.

3. **Preparación de datos.** Aquí se realiza la selección de datos a los que se va a aplicar la técnica de modelado (variables y muestras), la limpieza de los datos, la generación de variables adicionales, la integración de diferentes conjuntos de datos y cambios de formato. Esta fase laboriosa, está directamente relacionada con la fase de modelado, puesto que, en función de la técnica a utilizar, los datos necesitan ser procesados en diferentes formas. Incluye tareas como: Selección de datos, limpieza de los datos, estructuración de los datos e integración y formateo de estos.

4. **Modelado.** Aquí se seleccionan las técnicas apropiadas para el desarrollo del proyecto. La técnica por emplearse en esta fase debe ser seleccionada en función a los siguientes criterios: ser apropiada al problema, disponer de datos adecuados,

cumplir los requerimientos del problema, y el conocimiento de la técnica. Las tareas en esta fase son: Seleccionar la técnica de modelado, generar el plan de prueba, construir el modelo y evaluar el modelo.

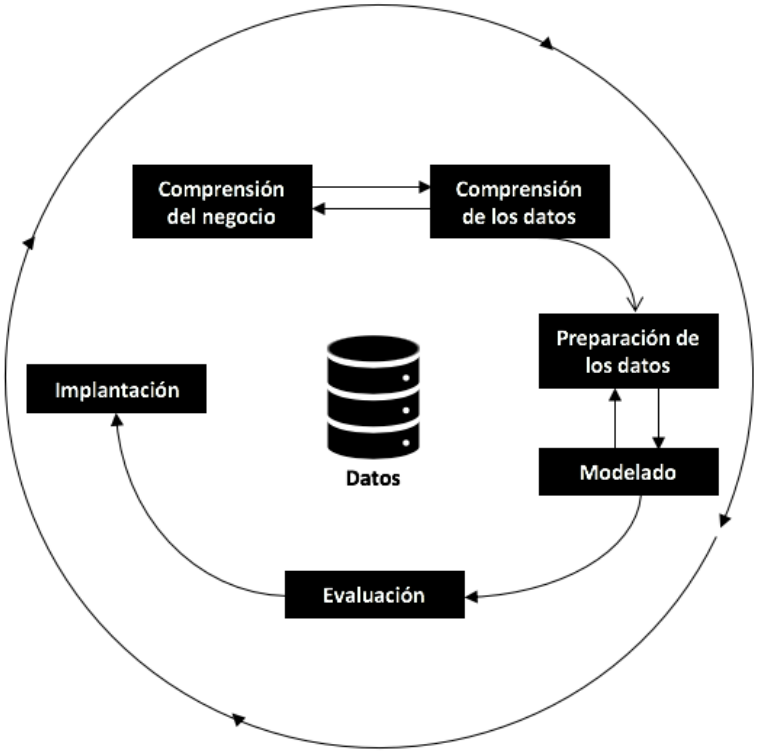
5. **Evaluación.** En esta fase se evalúa el modelo, no desde el punto de vista de los datos, sino del cumplimiento a los requerimientos iniciales. Antes de proceder a su implantación para su uso habitual, se debe revisar todo el proceso teniendo en cuenta los resultados obtenidos, e identificando posibles errores que llevan a repetir algún proceso anterior. Sus tareas consisten en evaluar los resultados, revisión del proceso y determinar los próximos pasos a seguir.

6. **Implementación.** Si el modelo generado es valido, desde el punto de vista de cumplimiento a los requerimientos iniciales, se procede a su implementación y explotación. Normalmente los proyectos de minería de datos no terminan en la implementación del modelo, sino que se deben documentar y presentar los resultados de manera comprensible para alcanzar un mejor entendimiento del conocimiento. En esta etapa se elabora un plan de implementación y se realiza una monitorización y mantenimiento.

La secuencia de las fases no es estricta, puesto que estas pueden interactuar entre sí durante el desarrollo del proyecto. De esta manera, la siguiente fase en la secuencia, a menudo depende de los resultados asociados con la fase precedente.

El siguiente diagrama representa las fases de la metodología CRISP-DM como un ciclo, en el que las flechas indican las dependencias significativas entre las diferentes fases.

Figura 2.2 Ciclo de la metodología CRISP-DM



Fuente: Chapman et al. 2000

2.2 Machine Learning (Aprendizaje Automático)

Machine Learning conocido como “Aprendizaje automático” en español, es una disciplina que nace a partir de la inteligencia artificial y del constante esfuerzo por tratar de construir programas de computadora que mejoren automáticamente con la experiencia, en otras palabras, construir maquinas que adaptan y modifican sus acciones o predicciones de tal manera que se vuelven mas precisas. Este “aprendizaje” le brinda a la computadora la capacidad de mejorar su desempeño, de manera similar al aprendizaje en los seres humanos.

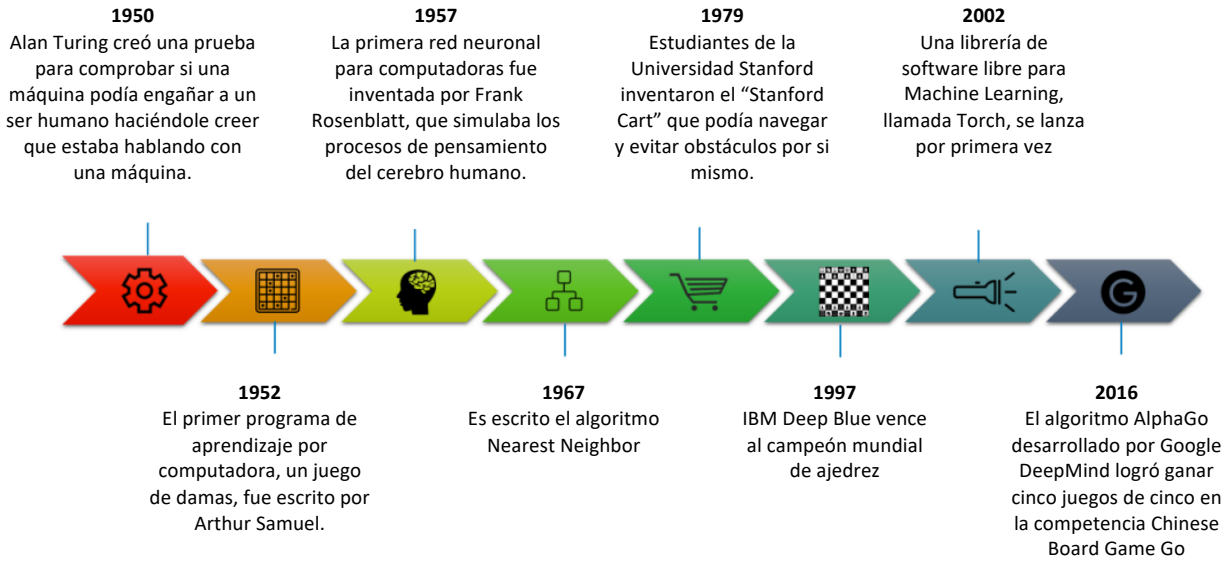
El termino Machine Learning fue creado por Arthur Samuel (1959), quien lo definió como una rama de la inteligencia artificial, cuyo campo de estudio da a las computadoras la habilidad de aprender algo sobre lo que no han sido explícitamente programadas.

Machine Learning permite que los equipos aprendan a partir de los datos y las experiencias y actúen sin haber sido programados de forma específica. Los usuarios pueden crear aplicaciones de inteligencia artificial que detecten, procesen y actúen de forma inteligente según la información, lo que aumenta la capacidad humana, la velocidad y la eficacia, y ayuda a las organizaciones a llegar más lejos [Microsoft, 2018].

Cuando en la informática tradicional escribimos un programa o un código para algún propósito específico utilizamos un conocimiento a priori para las reglas de código, es decir, estamos escribiendo un conjunto definido de instrucciones que la máquina seguirá. Mientras que, en el aprendizaje automático, los algoritmos no están diseñados para un solo dominio del problema, ingresamos un conjunto de datos a través del cual la máquina se ajusta a las muestras, identificando y analizando los patrones en el conjunto de datos y aprenderá a tomar decisiones de forma autónoma en función de sus observaciones y aprendizajes y, por extensión, crea sus propias reglas autodidactas.

El siguiente diagrama cubre algunos eventos críticos en el desarrollo de Machine Learning a lo largo de la historia:

Figura 2.3 Eventos históricos – Machine Learning



Fuente: Tahsildar, 2018

En el campo de la ciencia de los datos Machine Learning se enfoca en el diseño de algoritmos que pueden aprender sobre datos pasados para generar pronósticos. Al igual que en la minería de datos, ambos sistemas procesan grandes volúmenes de datos para construir modelos que puedan por ejemplo buscar patrones a través de los datos con que son alimentados, pero a diferencia de la minería de datos, Machine Learning no es solo un problema de base de datos, ya que este forma parte de la inteligencia artificial con base en que un sistema que está continuamente cambiando de entorno ha de poseer la habilidad de aprender. Su principal objetivo es convertir datos en conocimientos a través de ejemplos y experiencia.

Formalmente hablando, trata de crear programas y rutinas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos o experiencia pasada. Partimos de un modelo definido por unos parámetros y mediante el "aprendizaje" (la observación y análisis de ejemplos y experiencia pasada) optimizamos los parámetros del modelo, para luego usarlos en la predicción [Huertas, 2015].

"Aprender" en este contexto quiere decir identificar patrones complejos en millones de datos. La máquina que realmente aprende es un algoritmo que revisa los datos y es capaz de predecir comportamientos futuros. "Automáticamente", también en este contexto, implica que estos sistemas se mejoran de forma autónoma con el tiempo, sin intervención humana. Los datos históricos, debidamente organizados y tratados en bloque, generan una base de datos que se puede explotar para predecir futuros comportamientos, favorecer aquellos que mejoran los objetivos de negocio y evitar aquellos que son perjudiciales. Esa cantidad descomunal de datos son imposibles de analizar por una persona para sacar conclusiones y menos todavía para hacer predicciones. Los algoritmos en cambio sí pueden detectar patrones de comportamiento contando con las variables que le proporcionamos y descubrir cuáles son las que han llevado, a cierto comportamiento [González, 2014].

El proceso de Machine Learning es iterativo pues consiste en tener una base de datos que se coloca como entrada a un algoritmo. Los datos que suministramos a la maquina se dividen en dos partes: datos de entrenamiento y datos de prueba. El uso de algoritmos es el medio por el cual la computadora analiza estos datos, aprende de los patrones y características de los datos de entrenamiento y se entrena para tomar decisiones como identificar, clasificar o predecir nuevos datos. Para verificar con qué precisión la máquina puede tomar estas decisiones, las predicciones se aplican a los datos de prueba, los

resultados obtenidos son la medida de aprendizaje de la maquina, guiando la toma de decisiones.

Cierto es que no podemos identificar el proceso completo que define los datos, pero creemos poder construir una buena aproximación a éste: confiamos en poder identificar esas regularidades y relaciones. Ésta es la base del aprendizaje automático. La identificación de este comportamiento reiterativo nos ayudara a entender el proceso y a hacer predicciones [Huertas, 2015].

2.2.1 Disciplinas Relacionadas al ML

La precisión en modelos de Machine Learning se puede incrementar de manera sustancial al incorporar disciplinas como Big Data.

Big data es cualquier tipo de fuente de datos que tenga al menos una de las cuatro características compartidas, denominadas cuatro “V´s” [Hurwitz, 2018]:

- Volúmenes de datos extremadamente grandes
- La capacidad de mover esos datos a una alta velocidad de velocidad.
- Una variedad cada vez mayor de fuentes de datos
- Veracidad para que las fuentes de datos realmente representen la verdad.

Sin suficientes datos, estaríamos intentando tomar decisiones basados en pequeños subconjuntos de datos que podrían llevar a una mala interpretación de una tendencia o perder un patrón que recién comienza a emerger.

Con Big Data, ahora es posible virtualizar los datos para que puedan almacenarse de la manera más eficiente y rentable, ya sea en forma de premisas o en la nube. Además, las mejoras en la velocidad y confiabilidad de la red han eliminado otras limitaciones físicas de poder administrar grandes cantidades de datos a la velocidad aceptable. Adicional a esto el impacto de los cambios en el precio y la sofisticación de la memoria de la computadora, y con todas estas transiciones tecnológicas, ahora es posible imaginar cómo las empresas pueden aprovechar los datos de una forma que hubiera sido inconcebible hace solo cinco años [Hurwitz, 2018].

La estadística y la minería de datos son disciplinas que también cuentan con un rol importante en el desarrollo del Machine Learning principalmente en la comprensión de datos, describiendo las características de un conjunto de datos y encontrando relaciones y patrones en ellos para construir un modelo.

La estadística es la ciencia del análisis de los datos. Las estadísticas clásicas o convencionales son de naturaleza inferencial, lo que significa que se utilizan para llegar a conclusiones sobre los datos (varios parámetros). El modelado estadístico se enfoca principalmente en hacer inferencias y entender las características de las variables. Los modelos de aprendizaje automático aprovechan los algoritmos estadísticos y los aplican para predecir los análisis. En un modelo estadístico, una hipótesis es una forma comprobable de confirmar la validez del algoritmo específico. [Hurwitz, 2018].

La minería de datos, señalada en el apartado 2.1, se basa en los principios de la estadística y tiene el objetivo de explicar y comprender los datos. Las herramientas de minería de datos están destinadas a apoyar el proceso de toma de decisiones, mostrando patrones que pueden ser utilizados por los humanos. En contraste, Machine Learning automatiza el proceso de identificación de patrones que se utilizan para hacer predicciones.

2.2.2 Aplicaciones de Machine Learning

Gracias al aumento en Big data y a su incorporación en diversos paquetes y librerías aplicadas a lenguajes de programación como Python, R, C++, Matlab, etc. El número de aplicaciones y estudios se ha incrementado de manera sustancial. Su uso se implementa en una tarea tan simple como reconocer la escritura humana o tan complejo como los autos que conducen por sí mismos.

Algunos ejemplos de la variedad de campos en que se aplica Machine Learning son:

- Redes sociales:
    - Análisis de sentimientos
    - Filtrado de spam

- Transporte
    - Monitoreo
    - Control de trafico aéreo
    - Estimación de tiempos
- Servicios financieros
    - Administración de portafolios
    - Detección de fraudes
    - Calificación crediticia
- Cuidado de la salud
    - Diagnostico de enfermedades
    - Cirugía Robótica
- Comercio electrónico
    - Atención al cliente
    - Recomendación de productos
    - Publicidad
- Asistentes virtuales
    - Procesamiento de lenguaje
    - Reconocimiento facial
    - Procesamiento y clasificación de imágenes
    - Dictado a texto
    - Reconocimiento de voz
    - Detección de objetos

En el ámbito financiero los algoritmos de Machine Learning comienzan a ser utilizados por empresas para diferentes propósitos:

- Analizar el comportamiento histórico del mercado utilizando grandes conjuntos de datos.
- Determinar las entradas óptimas (predictores) para una estrategia.
- Determinación del conjunto óptimo de parámetros de estrategia.
- Haciendo predicciones comerciales, calificación crediticia, etc.

El mercado financiero genera un gran volumen de datos. Por lo tanto, puede resultar difícil para un analista estudiar los datos para encontrar un patrón y posteriormente diseñar una

estrategia que funcione. Los algoritmos de Machine Learning se pueden utilizar para analizar los datos y crear estrategias que nos ayuden a predecir los precios de los activos con una precisión significativa.

2.2.3 Machine Learning en el análisis de Series de tiempo

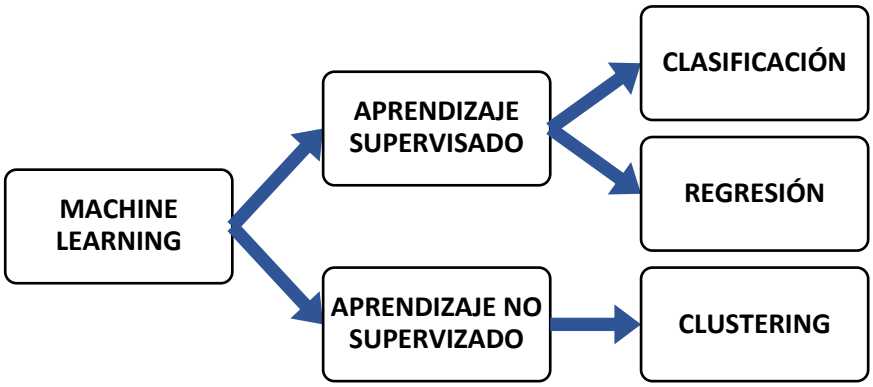
Recientemente Machine Learning se ha convertido en una aplicación popular para el pronóstico de series financieras, donde un gran número de investigaciones han obtenido una alta exactitud en la predicción de cambios en precios o tendencias en diferentes mercados financieros brindando la posibilidad de obtener ganancias basados en estos pronósticos.

Las series temporales son observaciones ordenadas en secuencias, generalmente ordenadas en el tiempo. El análisis de series de tiempo difiere de otros análisis, porque las observaciones dependen del tiempo. Esto requiere algunas consideraciones al realizar el aprendizaje automático, ya que no podemos aleatorizar el orden de las entradas ni afirmar de cuántas observaciones anteriores depende una observación actual. Los estudios han demostrado que los algoritmos de aprendizaje automático son adecuados para las series de tiempo de pronóstico y que el aprendizaje automático se puede utilizar completamente para uno y dos pasos de predicción [Bontempi, 2013] [Nesreen, 2010] [Olden, 2016].

2.3 Clasificación de los algoritmos de Machine Learning

Los modelos empleados en Machine Learning se clasifican en dos técnicas: Aprendizaje supervisado que entrena un modelo con datos de entrada y salida conocidos para que pueda predecir salidas futuras y el aprendizaje no supervisado, que encuentra patrones ocultos o estructuras intrínsecas en los datos de entrada.

Figura 2.4 Clasificación de Machine Learning



Fuente: Elaboración propia

2.3.1 Aprendizaje supervisado

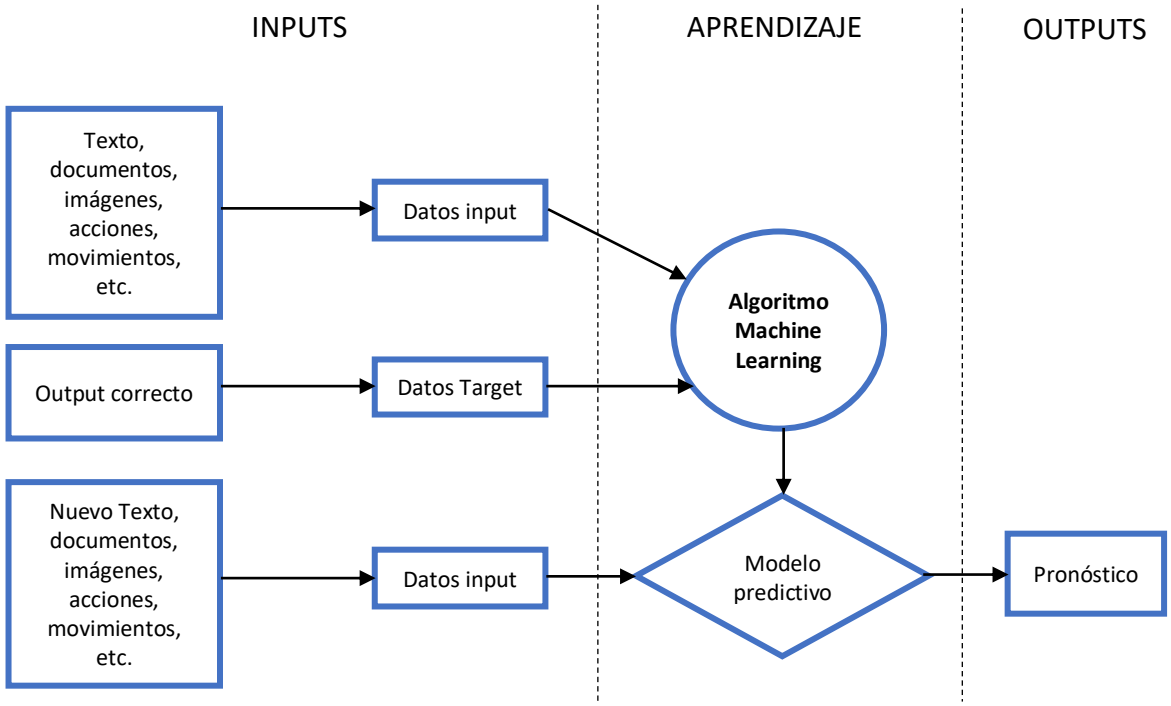
En este tipo de algoritmo, el conjunto de datos con los que se lleva a cabo el entrenamiento consiste en datos etiquetados, es decir, que alimentamos el modelo tanto con los parámetros de entrada como la salida requerida. Emplea una división de datos que consiste en inputs (X) y outputs (Y), siendo el output o *target* la respuesta que el algoritmo debe de producir a través de los datos de entrada.

En este tipo de aprendizaje se desarrollan modelos que definan correctamente la relación entre input y output, es decir, que asocien correctamente cada dato de entrada con su respectiva salida. El objetivo es predecir el valor de la variable respuesta basándose en numerosas variables de entrada, de forma que la variable a predecir "supervisa" el proceso de aprendizaje. Formalmente, se buscan los parámetros, θ, del modelo, *g(.)*, que describa la relación entre input y output [Huertas, 2015]:

$$Y = g\,(\,X \mid \theta\,)$$

De forma esquemática, se suelen representar como:

Figura 2.5 Aprendizaje supervisado



Fuente: Olden, 2016

A través del aprendizaje supervisado encontramos que los tipos de problemas se dividen en dos subgrupos: Clasificación para predecir categorías discretas y Regresión, para predecir valores continuos.

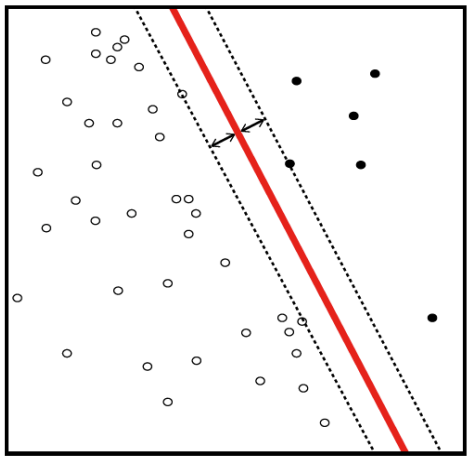
2.3.1.1 Clasificación

Es la predicción de valores discretos, donde la salida del problema se divide en clases y se busca asociar cada input con su respectiva clase. El problema consiste en tomar un nuevo input u observación y decidir en cual, de un número determinado de clases, pertenece. Se desempeña basado en ejemplos de entrenamiento de cada una de las clases. Algunos ejemplos de problemas de clasificación son: clasificar si un correo nuevo es spam o no (output) en función de su contenido (input); decidir si un paciente tiene una enfermedad (output) según sus síntomas e historial clínico (input); clasificar el riesgo de hacer un crédito a un cliente (output) en función de su capacidad financiera (input) [Huertas,2015].

Para el caso de estudio de esta investigación se emplearon algoritmos de clasificación binaria, también conocidos como “Two-Class Algorithms” (Algoritmos de dos clases) los cuales simplemente son algoritmos con dos clases previamente definidas, por ejemplo, verdadero y falso.

La siguiente figura muestra un ejemplo de modelo de clasificación en donde un separador lineal divide las dos clases existentes:

Figura 2.6 Modelo de Clasificación



Fuente: Olden, 2016

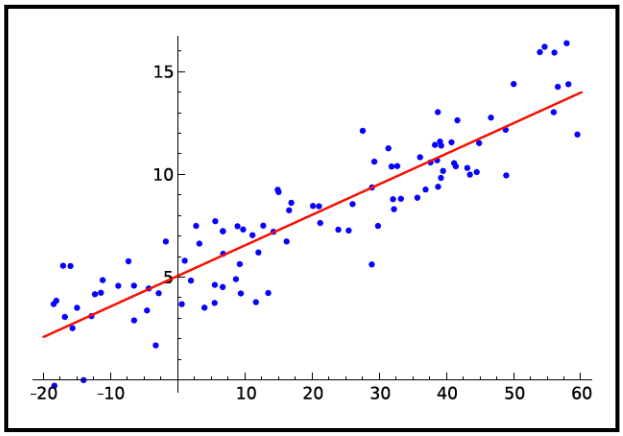
En general, la tarea de un clasificador es la de separar el espacio de variables en regiones de decisión (clases). Los límites entre diferentes regiones se conocen como fronteras de decisión. El calculo de estas fronteras de decisión es el objetivo de los modelos de clasificación [Huertas, 2015].

2.3.1.2 Regresión

Predice valores continuos, donde la salida se representa mediante un valor numérico calculado a partir de la entrada de datos. Los algoritmos de regresión funcionan estimando la relación existente entre las variables de entrada y la variable de salida, siendo esta relación como un cambio en los datos de entrada afectan los datos de salida.

Los problemas de regresión pueden ser la predicción del precio de una casa (output) según sus atributos (input); o la creación de coches autónomos, donde la salida sería el ángulo de giro del volante en cada instante y la entrada los datos tomados por distintos sensores dispuestos por el coche.

Figura 2.7 Modelo de regresión



Fuente: Olden, 2016

2.3.2 Aprendizaje no supervisado

A diferencia de los algoritmos de aprendizaje supervisado, donde empleamos datos etiquetados para la fase de entrenamiento, los datos no estarán etiquetados para los algoritmos de aprendizaje no supervisados. La agrupación de datos en un grupo específico se realizará sobre la base de las similitudes entre las variables.

En el aprendizaje no supervisado no existe la figura del “supervisor” y solo se dispone de los datos de entrada. El objetivo es encontrar regularidades y patrones en los datos que nos permitan asociar unos con otros. Suponemos que existe una estructura en el espacio de datos de entrada donde ciertas conductas son más frecuentes que otras, y nos interesamos en observar lo que generalmente ocurre y lo que no (lo que en estadística se conoce como estimación de la densidad) [Huertas, 2015].

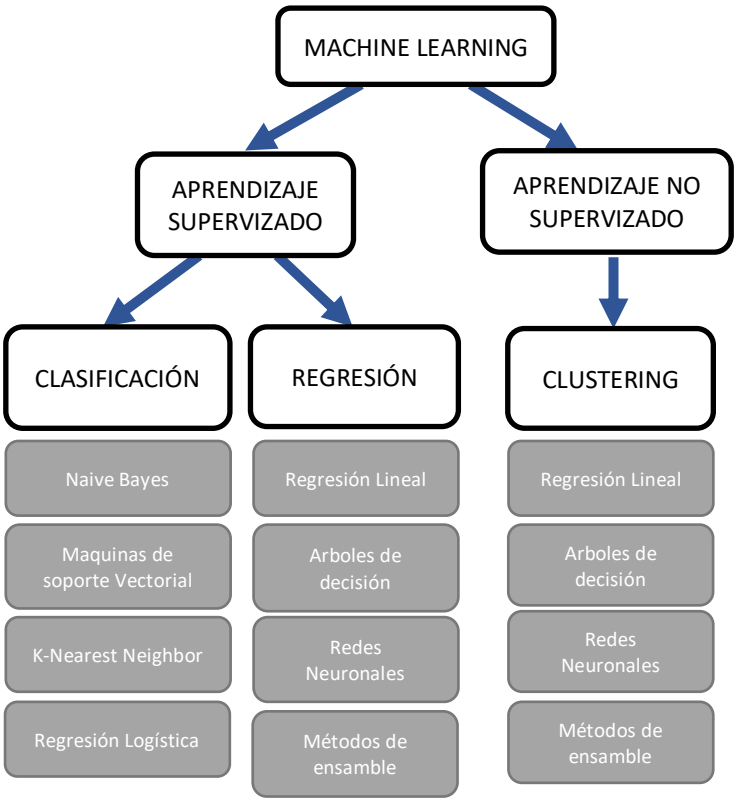
Un método muy utilizado en el aprendizaje no supervisado es el *clustering*, una agrupación de datos que es útil para, por ejemplo, compañías interesadas en establecer conjuntos de consumidores con el fin de ofrecer un producto que interese a la mayoría.

2.4 Algoritmos de Machine Learning

Existe una gran variedad de algoritmos de Machine Learning, por lo que la elección del algoritmo más adecuado para un problema puede ser complicado. El *Teorema del No Free Lunch* declara que no hay un algoritmo de Machine Learning óptimo para todos los problemas, y que algunos se ajustan mejor que otros para ciertos casos, esto quiere decir, que, por cada par de algoritmos de aprendizaje, hay tantos problemas en el que el primer algoritmo es mejor que el segundo, como problemas en el que el segundo es mejor que el primero. La principal causa de esto es que los diferentes algoritmos manejan problemas como el ruido y el overfitting de diferentes maneras. No hay manera certera de saber la manera en que un algoritmo se comportara frente a cada problema, por lo que tenemos que probar una variedad de diferentes algoritmos con el fin de encontrar cual se ajusta en la resolución de nuestro modelo [Wolpert, 2001] [Olden, 2016].

Existen docenas de algoritmos de aprendizaje automático supervisados y no supervisados, y cada uno ofrece un enfoque distinto del aprendizaje, por lo que la elección del algoritmo adecuado puede parecer abrumadora. Para encontrar el algoritmo correcto se utiliza en parte la técnica de ensayo y error ya que no podemos saber si un algoritmo funcionará sin probarlo. Pero la elección del algoritmo también depende del tamaño y el tipo de los datos con los que se trabaja, la información que se desea obtener de los datos y cómo se empleará dicha información.

Figura 2.8 Técnicas y algoritmos de aprendizaje automático

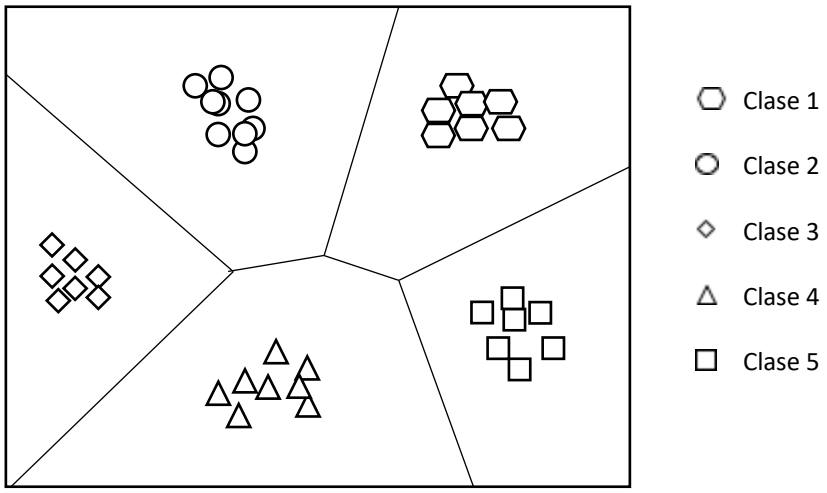


Fuente: Matlab, 2019

Para este proyecto se eligieron siete de los algoritmos de clasificación mas populares y empleados en diferentes artículos e investigaciones de pronósticos en series financieras.

En general, en un problema de clasificación con dos o más clases, un algoritmo clasificador divide el espacio de las características o variables en volúmenes llamados regiones o límites de decisión. Un límite de decisión es una hipersuperficie que divide el espacio vectorial subyacente en dos o más regiones, una para cada clase. La forma en que funciona el clasificador depende de qué tan bien el límite de decisión separe los ejemplos de diferentes clases entre sí. En la siguiente figura, las líneas que separan los ejemplos de cada clase entre sí son límites de decisión.

Figura 2.9 Ejemplo de un limite de decisión.



Fuente: Princeton.edu

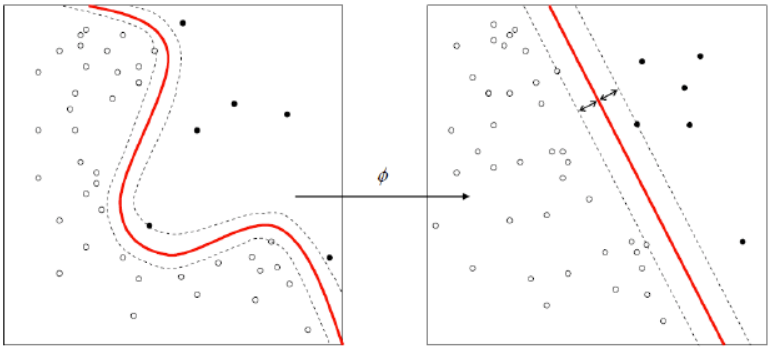
La forma en que funciona un clasificador depende de qué tan cerca se parezcan los patrones de entrada a clasificar. En el ejemplo de la figura 2.9, la correspondencia es muy estrecha y se puede anticipar un rendimiento excelente. Sin embargo, las cosas no siempre son tan buenas en la práctica, y uno debe entender las limitaciones de los clasificadores simples.

A continuación, se revisa los aspectos teóricos de los modelos de clasificación seleccionados para llevar a cabo este análisis.

2.4.1 Maquinas de soporte Vectorial (Support Vector Machine)

Las máquinas de vectores de soporte (SVM por sus siglas en ingles), son un esquema de clasificación que, cuando se construye el modelo, utiliza una función matemática para aumentar las dimensiones de las muestras hasta que pueda separar linealmente las clases en el conjunto de pruebas. La función matemática que aumenta las dimensiones se conoce como función de kernel; la función del kernel transforma los datos de tal manera que hay una mayor posibilidad de separar las clases. La siguiente figura muestra cómo un kernel ϕ transforma los datos en clases separables. Los SVM pueden usar funciones de kernel tanto lineales como no lineales [Olden, 2016].

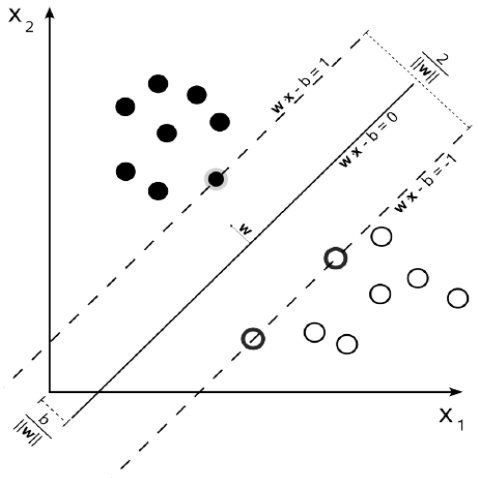
Figura 2.10 Kernel de una Maquina de soporte vectorial



Fuente: Olden, 2016

Cuando la SVM ha alcanzado un estado en el que puede separar linealmente las clases, intenta encontrar la separación óptima. La separación óptima es cuando el separador está igualmente lejos de la muestra más cercana en cada clase, como se muestra en la figura 2.11. Cuando el SVM ha construido su modelo, puede predecir sobre nuevos datos realizando la misma transformación del núcleo en los nuevos datos y luego observar a qué clase debería pertenecer. Los SVM son conocidos por funcionar bien en conjuntos de muestras de tamaño razonable, pero tienen un peor rendimiento en conjuntos grandes [Marsland, 2014].

Figura 2.11 Modelo de una separación optima.



Fuente: Olden, 2016

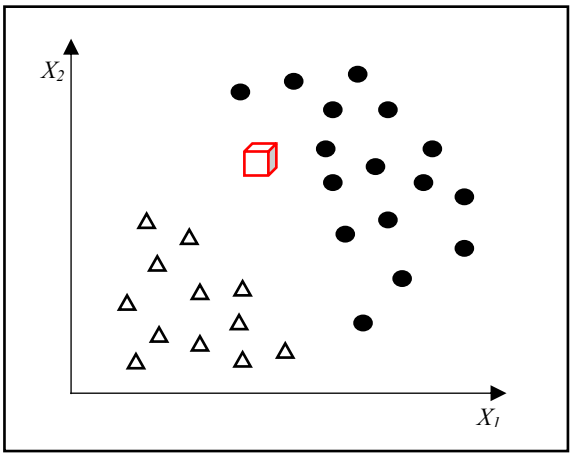
2.4.2 K-Nearest Neighbors (KNN)

Este algoritmo se utiliza para clasificar un conjunto de datos en grupos o clases específicas según las similitudes entre los datos.

El algoritmo de los *k-vecinos más próximos*, o KNN, es uno de los algoritmos mas simples utilizados en Machine Learning para problemas de regresión y clasificación. Este algoritmo almacena todos los casos disponibles y los clasifica en función de una medida de similitud, por ejemplo, la función de distancia. Un nuevo caso se clasifica en la clase predominante de su vecindario, es decir, en función de la "mayoría de votos" de sus vecinos los datos se asignan a la clase que tiene los vecinos más cercanos. A medida que aumenta el número de vecinos más cercanos (*k*) la precisión puede aumentar.

La siguiente gráfica muestra el ejemplo de un modelo KNN para el caso de dos variables predictoras (X1 y X2). El nuevo dato (cubo) se clasificaría como circulo debido a que es la clase predominante de su "vecindario".

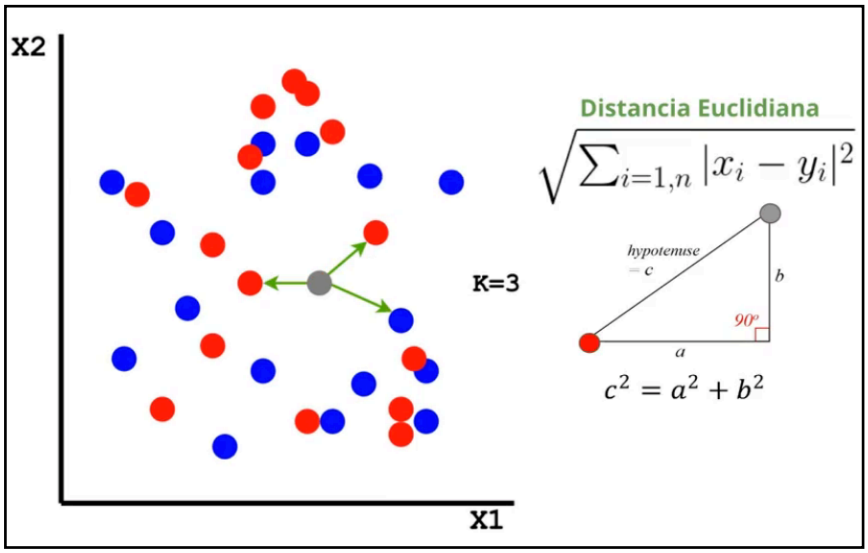
Figura 2.12 Ejemplo de clasificación con K-Nearest Neighbor



Fuente: Elaboración propia

Para agrupar los vecinos, se establece una métrica en el conjunto de puntos de entrenamiento, la cual puede ser la distancia Euclidiana, la distancia Manhattan (o rectangular) o la distancia de Minkowski. En el caso de variables categóricas, se suele usar la distancia de Hamming [Huertas, 2015].

Figura 2.13 Clasificación en K-Nearest Neighbor con distancia Euclidiana



Fuente: Garrido, 2018

Por tanto, el proceso de aprendizaje consiste en: una vez seleccionada una métrica adecuada para el problema, se agrupan los datos de entrenamiento en *K "vecindarios"*, de tal forma que esta agrupación sea óptima. Cada dato nuevo se clasifica en función de la clase a la que pertenece la mayoría de las variables de su vecindario [Huertas, 2015].

Este algoritmo presenta algunos problemas potenciales en su aplicación:

- Usa una cantidad muy grande de memoria para almacenar los datos.
- Usa demasiado poder computacional al realizar las predicciones.
- Tiene problemas cuando emplea bases de datos con demasiada información.

2.4.3 Arboles de decisión

Los árboles de decisión son un método de Machine Learning que se usa en problemas de clasificación y regresión, también conocido como CART.

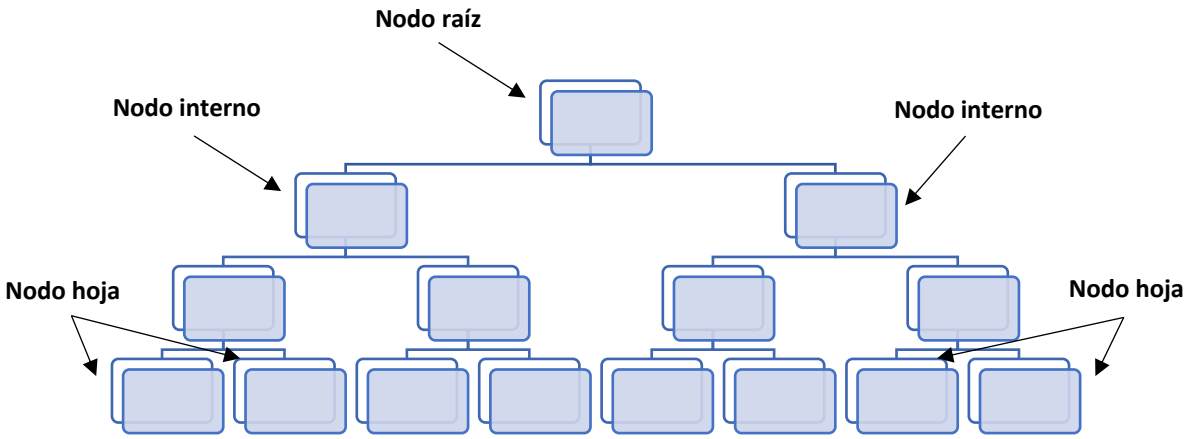
Son estructuras similares a los diagramas de flujo, en el que cada nodo contiene una prueba aplicada a una característica, cada rama representa la salida de dicha prueba y cada nodo final contiene una etiqueta de clase.

Cuando los árboles de decisión se utilizan como un algoritmo de clasificación, los árboles se crean dividiendo el conjunto de datos en subconjuntos según los resultados con respecto a una característica y crean los subárboles correspondientes. Esto se realiza hasta que cada subárbol derivado conduce a un nodo final [Olden, 2016].

La estructura de los árboles de decisión se compone de la siguiente manera:

- El **nodo raíz**, está en la parte superior y no tiene rutas entrantes.
- Los **nodos internos** o los nodos de prueba están en el centro y pueden estar en diferentes niveles o subespacios, tienen vías de entrada y salida.
- Los **nodos de hoja** o los **nodos de decisión** están en la parte inferior, tienen rutas entrantes, pero no rutas salientes y aquí podemos encontrar los resultados.

Figura 2.14 Estructura de los arboles de decisión



Fuente: Elaboración propia

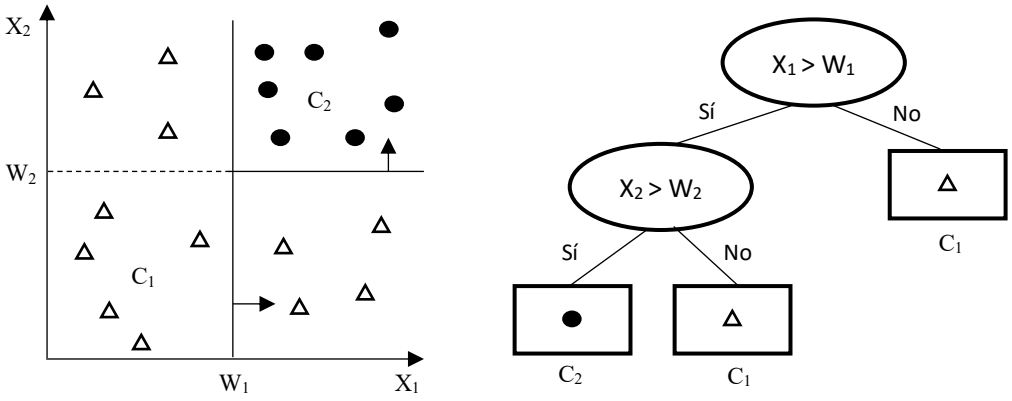
El proceso de construcción del árbol es su proceso de aprendizaje. El modelo analiza los datos y extrae de ellos las reglas lógicas que le permiten clasificar los datos. Se trata de un proceso recursivo, que consta de los siguientes pasos [Huertas, 2015]:

1. Se analizan las distintas particiones del espacio y se escogen aquellas que generen mejores regiones de decisión.

2. Se optimiza la separación.
3. Se repite el paso 1 con los nodos hijos que sean de decisión.

El ejemplo mostrado en la figura 2.15 muestra dos variables (X_1 y X_2) y dos clases (C_1 y C_2). A la izquierda, el espacio de variables con cada ejemplo. A la derecha, el árbol generado, donde los nodos ovalados son nodos de decisión y los rectangulares son nodos *hoja.*

Figura 2.15 Ejemplo de clasificación con arboles de decisión



Fuente: Huertas, 2015

Un árbol de decisión divide el espacio de la característica en regiones separadas y cada observación que cae en una región específica recibe una predicción de la media de esa región.

2.4.4 Bosques aleatorios

Un bosque aleatorio o *random forest* es una técnica de conjunto capaz de realizar tanto tareas de regresión como de clasificación con el uso de múltiples árboles de decisión y una técnica llamada Bootstrap Aggregation, comúnmente conocida como bagging. El bagging, en el método de bosque aleatorio, implica el entrenamiento de cada árbol de decisión en una muestra de datos diferente donde se realiza un muestreo con reemplazo. La idea básica

detrás de esto es combinar múltiples árboles de decisión para determinar el resultado final en lugar de confiar en árboles de decisión individuales [Hewa, 2018].

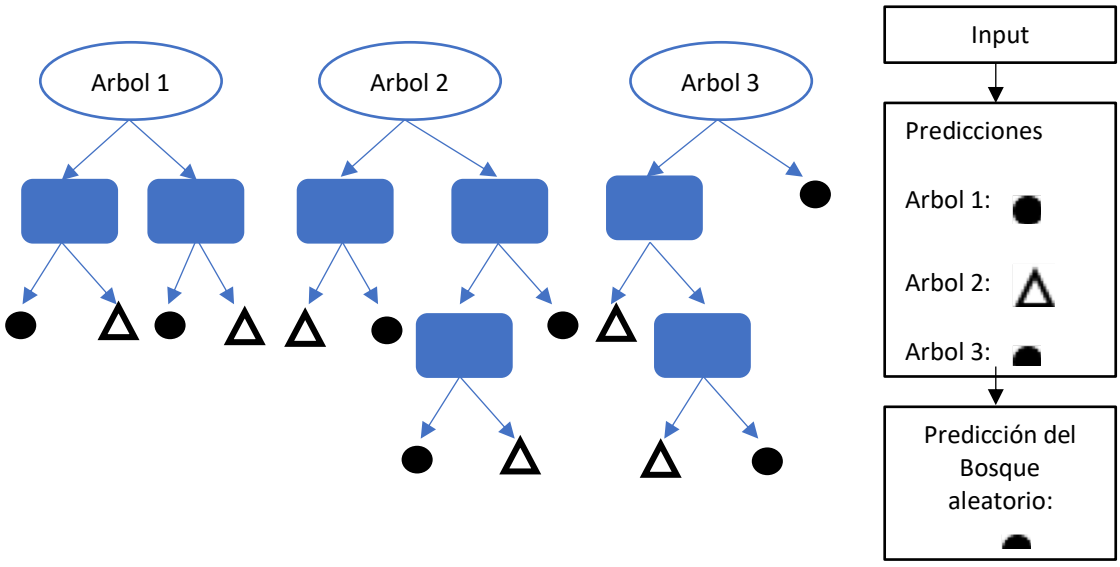
En general, el bosque aleatorio construye múltiples arboles de decisión y los fusiona para obtener una predicción mas precisa y estable.

El algoritmo agrega aleatoriedad adicional al modelo, mientras crecen los árboles. En lugar de buscar la característica más importante al dividir un nodo, busca la mejor característica entre un subconjunto aleatorio de características. Esto da como resultado una amplia diversidad que generalmente resulta en un mejor modelo. Por lo tanto, en el bosque aleatorio, el algoritmo para dividir un nodo solo tiene en cuenta un subconjunto aleatorio de las características. Incluso podemos hacer que los árboles sean más aleatorios, mediante el uso adicional de umbrales aleatorios para cada función en lugar de buscar los mejores umbrales posibles (como lo hace un árbol de decisión normal) [Donges, 2018].

Este algoritmo funciona en cuatro pasos:

- Selecciona muestras aleatorias de un conjunto de datos determinado.
- Construye un árbol de decisión para cada muestra y obtiene un resultado de predicción de cada árbol de decisión.
- Realiza una votación por cada resultado pronosticado.
- Selecciona el resultado de la predicción con más votos como la predicción final.

Figura 2.16 Funcionamiento del Bosque Aleatorio



Fuente: Polamuri, 2017

Finalmente, para clasificar un objeto en función de sus atributos, cada árbol proporciona una clasificación que se dice que "vota" por esa clase. El bosque elige la clasificación con el mayor número de votos. En el clasificador de bosque aleatorio, cuanto mayor sea el número de árboles en el bosque se obtienen resultados de mayor precisión.

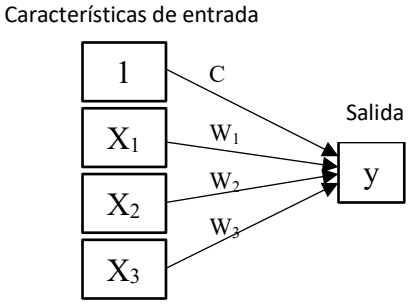
2.4.5 Regresión Logística

La regresión logística, a pesar de su nombre, no es un algoritmo de regresión sino un clasificador binario. La diferencia entre regresión lineal y logística es que la regresión logística se usa con variables dependientes categóricas (por ejemplo: Sí/No, Verdadero/Falso, 0/1, -1/1), a diferencia de las variables de valor continuo utilizadas en la regresión lineal.

La regresión logística ayuda a determinar la probabilidad de que una determinada variable esté en un determinado grupo. El modelo estima la probabilidad de un resultado binario basado en algunas características. Esto lo realiza midiendo la relación entre la variable dependiente categórica y una o más variables independientes mediante la estimación de probabilidades utilizando una función logística / sigmoide.

La regresión logística es un poco similar a la regresión lineal o podemos decirlo como un modelo lineal generalizado. En la regresión lineal, predecimos una salida con valores reales "y" basada en una suma ponderada de variables de entrada.

Figura 2.17 Regresión lineal

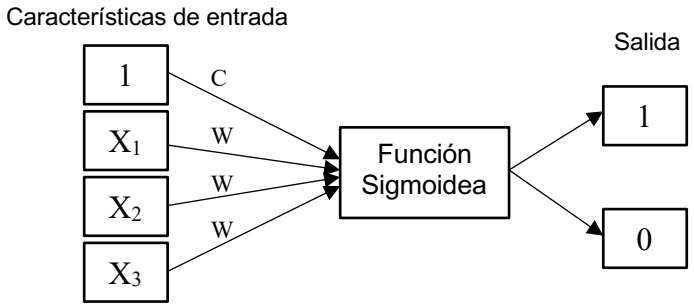


Fuente: Singh, 2018

El objetivo de la regresión lineal es estimar los valores de los coeficientes del modelo c, w1, w2, w3… .wn y ajustar los datos de entrenamiento con un mínimo error cuadrado y predecir la salida *y*.

La regresión logística hace lo mismo, pero con una adición. El modelo de regresión logística calcula una suma ponderada de las variables de entrada similares a la regresión lineal, pero ejecuta el resultado a través de una función especial no lineal, la función logística o la función sigmoidea para producir la salida *y*. Aquí, la salida es binaria o en la forma de 0/1 o -1/1.

Figura 2.18 Regresión Logística



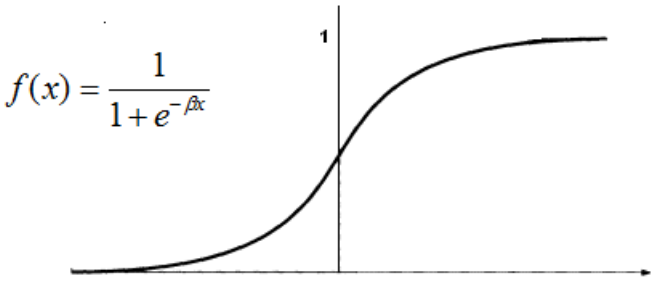
Fuente: Singh, 2018

La función sigmoidea / logística está dada por la siguiente ecuación:

$$y = \frac{1}{1+e^{-\beta x}} \qquad (2.1)$$

Como se puede ver en la figura 2.19, es una curva en forma de S que se acerca a 1 cuando el valor de la variable de entrada aumenta por encima de 0 y se acerca a 0 cuando la variable de entrada disminuye por debajo de 0. La salida de la función sigmoidea es 0.5 cuando la variable de entrada es 0:

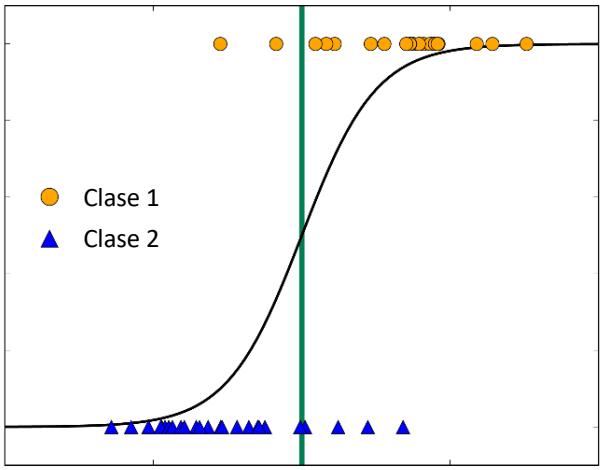
Figura 2.19 Función sigmoidea



Fuente: Singh, 2018

Por lo tanto, si el resultado es más de 0.5, podemos clasificar el resultado como 1 (o positivo) y si es menor que 0.5, podemos clasificarlo como 0 (o negativo) [Singh, 2018].

Figura 2.20 Ejemplo de regresión logística



Fuente: Microsoft Azure ML, 2017

Para el caso de estudio de este proyecto que consiste en predecir el movimiento del precio de las acciones. Si el precio de cierre de mañana es más alto que el precio de cierre de hoy, entonces compraremos las acciones (1), de lo contrario, las venderemos (-1). Si la producción es 0.7, entonces podemos decir que hay un 70% de probabilidad de que el precio de cierre de mañana sea más alto que el precio de cierre de hoy y lo clasifiquemos como 1.

2.4.6 Naive Bayes

Basado en el Teorema de Bayes que nos dice con qué frecuencia sucede A, dado que sucede B (P (A | B)), cuando sabemos con qué frecuencia ocurre B dado que ocurre A (P (B | A)), y qué tan probable es que A y B estén por su cuenta [Patel, 2017].

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \qquad (2.2)$$

Donde:

P(A|B) es "Probabilidad de A dado B", la probabilidad de A dado que B sucede.
P(A) es la probabilidad de A.
P(B|A) es "Probabilidad de B dado A", la probabilidad de B dado que A sucede.
P(B) es la probabilidad de B.

El clasificador Naive Bayes calcula las probabilidades para cada factor. Luego selecciona el resultado con mayor probabilidad.

Este clasificador asume que las características son independientes. De ahí la palabra ingenuo (Naive). Este algoritmo es utilizado para:

- Predicción en tiempo real
- Clasificación de texto / Filtrado de Spam
- Sistema de recomendaciones

El teorema de Bayes se puede convertir en un problema de clasificación adaptando la fórmula de la siguiente manera:

$$P(Clase_C|\{x_i\}) = \frac{P(\{x_i\}|Clase_C)\,P(Clase_C)}{P(\{x_i\})} \qquad (2.3)$$

$\{x_i\} = \{x_1\}, \{x_2\}, \ldots, \{x_n\}$ Variables independientes de una observación.
$P(Clase_C|\{x_i\})$ = Probabilidad de que dicha observación pertenezca a la clase C.
$P(\{x_i\}|Clase_C)$ = Ocurrencia de la clase C en el conjunto de observaciones.
$P(\{x_i\})$ = Ocurrencia de la observación x_i en el conjunto de observaciones.

Donde tenemos la probabilidad de que dada una observación X_i con todas las variables que tenemos pertenezca a una clase ($Clase_C$) determinada.

Para poder computar de forma estadísticamente significativa la ocurrencia de dicha observación en la $Clase_C$ necesitaríamos tener suficientes datos para cada combinación de todas las variables en el dataset, como lo demuestra la siguiente ecuación:

$$P(\{x_i\}|Clase_C) = P((x_1, x_2, \dots, x_n)|Clase_C) \qquad (2.4)$$

Con base en que lo anterior requiere un universo entero de datos que difícilmente se encuentran a nuestra disposición, este algoritmo asume de manera "ingenua" que las variables X_i son independientes las unas de las otras entonces se obtiene la siguiente ecuación:

$$P(\{x_i\}|Clase_C) = P\big((x_1, x_2, \dots, x_n)\big|Clase_C\big) \approx P(x_1|Clase_C) * P(x_2|Clase_C) * \dots * P(x_n|Clase_C) \qquad (2.5)$$

El clasificador Naive Bayes funciona aprendiendo las probabilidades de ver cada una de las variables de las observaciones de los datos de entrenamiento para cada clase, y usando el teorema de Bayes para predecir P ($Clase_C$ | $Observación_i$) [Garrido, 2018].

Naive Bayes ofrece tres alternativas para la formación de modelos [Patel, 2017]:

- Gaussiano: se usa en la clasificación y supone que las características siguen una distribución normal.
- Multinomial: Se usa para conteos discretos. Por ejemplo, un problema de clasificación de texto.
- Bernoulli: el modelo binomial es útil si sus vectores de características son binarios (es decir, ceros y unos). Una aplicación sería la clasificación de texto con el modelo de "bolsa de palabras" donde los 1 y 0 son "la palabra aparece en el documento" y "la palabra no aparece en el documento" respectivamente.

2.4.7 Redes Neuronales Artificiales

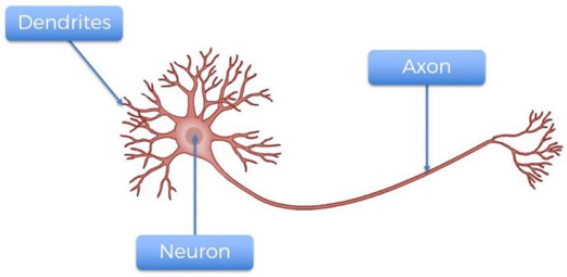
Deep Learning es un subcampo de Machine Learning que estudia el uso de las redes neuronales artificiales. Estos modelos están basados en el funcionamiento del cerebro humano. Al igual que el cerebro humano está formado por células nerviosas o neuronas que procesan la información mediante el envío y la recepción de señales, la red neuronal artificial en el Deep Learning consiste en capas de "neuronas" que se comunican entre sí y procesan información.

El "profundo" en Deep Learning señala la cantidad de capas dentro de la red; mientras mas grande sea el número de capas, más profunda es la red.

A diferencia de los modelos tradicionales de Machine Learning que carecen de un mecanismo para identificar errores y en los cuales el programador debe intervenir para ajustar el modelo y tomar decisiones más precisas, los modelos de Deep Learning pueden identificar la decisión inexacta y corregir el modelo por sí solos, sin intervención humana [Tahsildar, 2018].

Hay tres componentes en una neurona, las dendritas, el axón y el cuerpo principal de la neurona. Las dendritas son los receptores de la señal y el axón es el transmisor. Solo, una neurona no es de mucha utilidad, pero cuando está conectada a otras neuronas es capaz de realizar varios cálculos complicados.

Figura 2.21 Descripción de una célula nerviosa



Fuente: Singh, 2018

Una neurona de computadora se construye de manera similar, hay entradas a la neurona y la neurona emite una señal de salida después de algunos cálculos. La capa de entrada se parece a las dendritas de la neurona y la señal de salida es el axón.

Una red neuronal artificial es un modelo lógico matemático compuesto por estas neuronas o nodos. Cada nodo realiza una operación sencilla con una entrada seguida de una capa oculta para conectarse a través de una función de transferencia, y volverse a conectar a la capa oculta con la capa de salida que se enviara con otra función de transferencia en secuencias. Este procesamiento en paralelo tiene ventajas en el análisis de datos, ya que el procesar la información de esta manera no solo permite aprender de los errores, sino que permite aprender de ejemplos, y reconocer patrones en los datos [García y Morales, 2013].

Estructura de una red neuronal

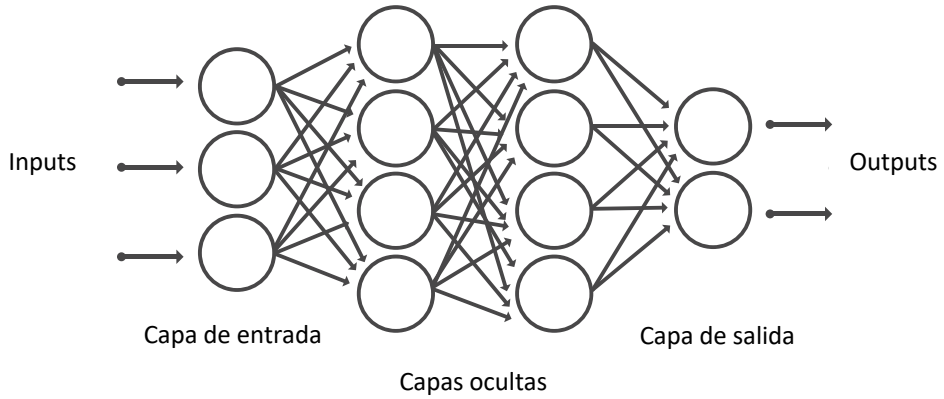
Cada red neuronal artificial consta de tres tipos de capas, que son [Tahsildar, 2018]:

- Capa de entrada: esta es la primera capa en una red neuronal artificial y proporciona los parámetros de entrada necesarios para procesar la información. Simplemente pasa estos parámetros a las capas adicionales sin ningún cálculo. Los parámetros con los que alimentamos la capa de entrada nos ayudarán a llegar a un valor de salida o hacer una predicción.

- Capas ocultas: estas capas en la red neuronal realizan los cálculos necesarios en las entradas recibidas de las capas anteriores y pasan el resultado a la siguiente capa. Es crucial decidir la cantidad de capas y la cantidad de neuronas en cada capa para aumentar la eficiencia del modelo. A mayor número de capas ocultas, más profunda es la red.

- Capa de salida: esta capa en la red neuronal profunda nos da la salida final después de recibir los resultados de las capas anteriores.

Figura 2.22 Redes neuronales organizadas en capas



Fuente: Matlab, 2019

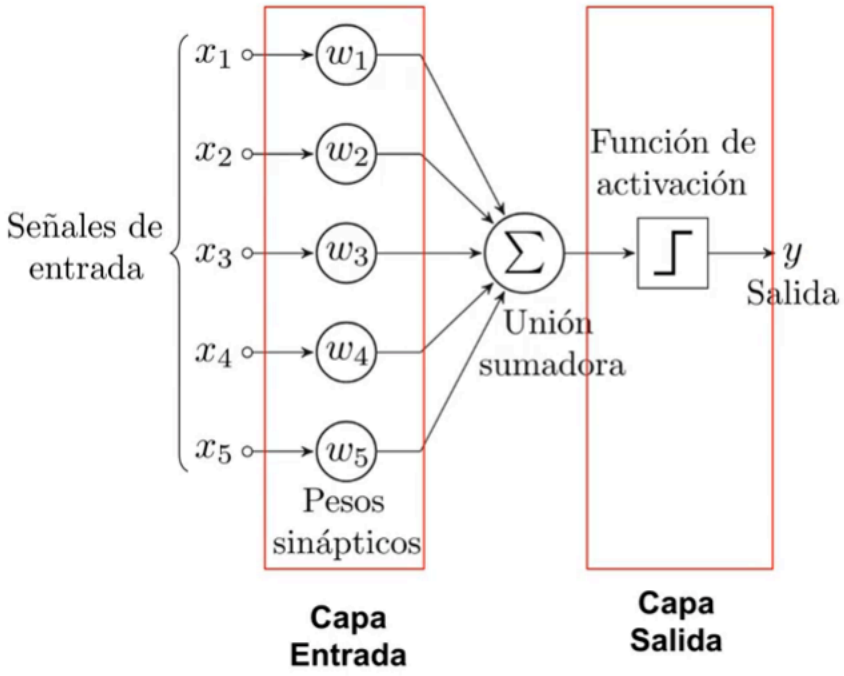
Perceptrón Multicapa

El perceptrón multicapa es una de las redes neuronales artificiales mas comunes, consta de una capa de entrada (inputs) y una capa de salida (outputs). Como se muestra en la figura 2.23 si recibimos un input (variables X1 a X5) el perceptrón asigna 5 pesos (W1 a W5), uno para cada variable.

Los pesos, como su nombre lo indica, se utilizan para agregar cierta ponderación a una determinada característica. Algunas características pueden ser más importantes que otras para obtener el resultado deseado [Tahsildar, 2018].

Posteriormente el perceptrón realiza una sumatoria del producto escalar de cada variable con su peso, después cuenta con una función de activación que modifica esa unión y lo convierte en una salida (output).

Figura 2.23 Perceptrón



Fuente: Garrido, 2018

En un caso simple de clasificación binaria, un perceptrón tendrá la siguiente función de activación:

$$f(X) = \begin{cases} 1 \text{ si } W * X + b > 0 \\ 0 \text{ en caso contrario} \end{cases}$$

Donde X es el input (las variables independientes X_1, $X_2 \ldots X_n$), W son los pesos de la neurona (W_1, $W_2 \ldots W_n$) y b es el sesgo.

El perceptrón es un algoritmo similar a las maquinas de soporte vectorial, ya que busca medir la distancia de un vector al hiperplano de decisión. El perceptrón es un clasificador lineal al hacer simplemente un producto vectorial y crear una frontera de decisión.

Aprendizaje del Perceptrón

Como se mencionó, los únicos parámetros que tiene el perceptrón son los pesos de la sinapsis. El objetivo de este algoritmo es encontrar aquellos pesos óptimos.

Si tenemos un dataset con S observaciones, D = $\{ (X_1,d_1),\ldots,(X_s,d_s) \}$

X_j las variables independientes de la observación j, y $X_{j,i}$ la variable i de la observación j,

$X_{j,0}$ = 1 el término de sesgo de cada observación

D_j el valor de la variabe objetivo para la observación j (1 o 0)

Siendo $W_i(t)$ el peso i de la sinapsis en el paso t,

γ el error máximo que queremos cometer, definido como:

$$\gamma = \frac{1}{s} \sum_{j=1}^{s} |d_j - y_j(t)| \qquad (2.6)$$

El error es la desviación media de los puntos, es decir, el output (objetivo) menos la predicción.

Los pasos en los que aprende el perceptrón son (Garrido, 2018):

1. Inicializa con los pesos de manera aleatoria, por ejemplo, a 0 o a un número aleatorio.

2. Para cada observación j en el dataset D:

   a. Calcula el output de la función de activación:

   $$y_j(t) = f[W(t) * X_j]$$
   $$= f[W_0(t)X_{j,0} + W_1(t)X_{j,1} + W_2(t)X_{j,2} + \ldots + W_n(t)X_{j,n}] \qquad (2.7)$$

   b. Actualiza los pesos sinápticos, para cada variable i de la observación j, calcula los nuevos pesos en función del error que se produjo:

   $$W_i(t+1) = W_i(t) + (d_j - y_i(t)) X_{j,i} \qquad (2.8)$$

3. Evalúa si ha llegado a cumplir el criterio de convergencia (error máximo). Si el error total es menos que el deseado para, si no, continua.

El perceptrón clásico no puede crear fronteras de decisión no lineales, por ello en las redes neuronales existen una variedad de funciones de activación mas avanzadas, todas ellas haciendo transformaciones no lineales.

La función de activación en la mayoría de los casos es una función continua no lineal, entre valores de su imagen o contra dominio entre valores de 0 y 1, o en otros casos entre -1 y 1. La función de activación se elige de acuerdo con las necesidades especificas: la función mas popular es la función sigmoidea o logística, como se muestra en la ecuación.

$$f(x) = \frac{1}{1 + e^{-\beta x}} \qquad (2.9)$$

Otra función muy utilizada es la de tipo tangente hiperbólica (TanH):

$$f(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad (2.10)$$

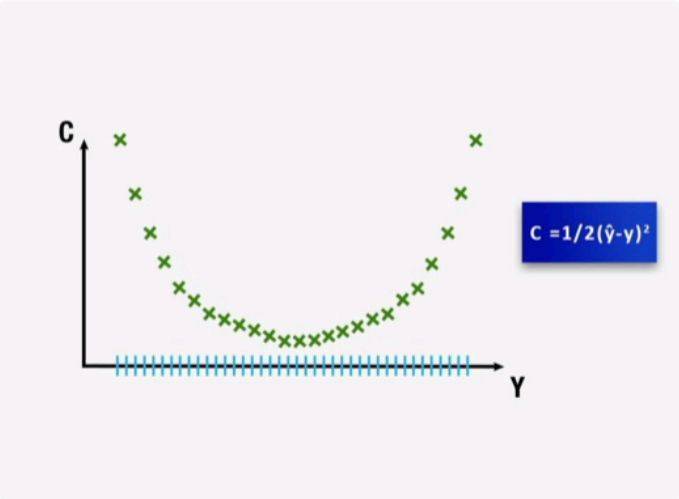
Estas funciones son las más utilizadas debido a su fácil diferenciación entre grupos, es decir funciones clasificadoras que permiten clasificar de manera adecuada [García y Morales, 2013].

Descenso de gradiente

Es un método de optimización que se emplea para minimizar o maximizar una función. En las redes neuronales artificiales generalmente se utiliza para minimizar una función de pérdidas, función que define el error de nuestro modelo.

Una forma de hacerlo es usando calculo. Supongamos que tomamos 1000 valores para los pesos y evaluamos la función de costo para estos valores. Cuando trazamos la gráfica de la función de costo, llegaremos a una gráfica como se muestra a continuación. El mejor valor para los pesos sería la función correspondiente a los mínimos de este gráfico.

Figura 2.24 Descenso de gradiente

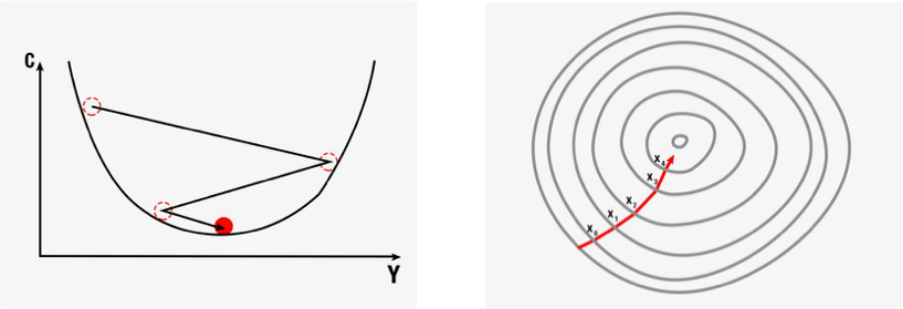


Fuente: Singh, 2018

Este enfoque podría ser exitoso para una red neuronal que involucra un peso único que necesita ser optimizado. Sin embargo, a medida que aumenta el número de pesos que se ajustan y aumenta el número de capas ocultas, el número de cálculos necesarios aumentará drásticamente. El tiempo que requerirá entrenar a un modelo de este tipo será extremadamente grande incluso en la supercomputadora más rápida del mundo. Por esta razón, es esencial desarrollar una metodología mejor y más rápida para calcular los pesos de la red neuronal. Este método es el Descenso de gradiente [Singh, 2018].

El descenso de gradiente implica analizar la pendiente de la curva de la función. En función de la pendiente, ajustamos los pesos para minimizar la función de pérdida en pasos en lugar de calcular los valores para todas las combinaciones posibles. La visualización del descenso de gradiente se muestra en las siguientes figuras. La primera gráfica es un solo valor de pesos y, por lo tanto, es bidimensional. Se puede ver que la bola roja se mueve en un patrón de zig-zag para llegar al mínimo de la función de costo. En el segundo diagrama, tenemos que ajustar dos pesos para minimizar la función. Por lo tanto, podemos visualizarlo como un contorno, como se muestra en la figura, donde nos estamos moviendo en la dirección de la pendiente más empinada, para alcanzar los mínimos en la duración más corta. Con este enfoque, no tenemos que hacer muchos cálculos y, como resultado, los cálculos no toman mucho tiempo, lo que hace que el entrenamiento del modelo sea una tarea factible [Singh, 2018].

Figura 2.25 Descenso de gradiente bidimensional y multidimensional

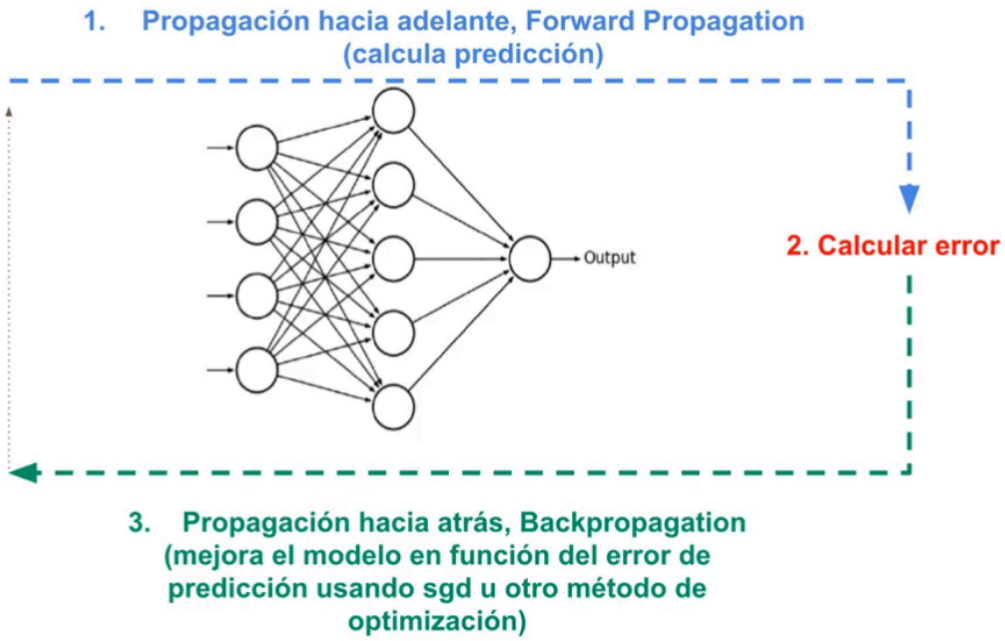


Fuente: Singh, 2018

Propagación hacia adelante y hacia atrás

El primer paso en el proceso de aprendizaje de una red neuronal es la propagación hacia adelante, donde en base al input vamos a generar una predicción, después se calcula el error de dicha predicción, finalmente se realiza una propagación hacia atrás, en la que se busca mejorar el modelo en función del error, usando el método de descenso de gradiente.

Figura 2.26 Propagación hacia adelante y hacia atrás



Fuente: Garrido, 2018

Aplicaciones del Deep Learning y las redes neuronales artificiales

Las aplicaciones del Deep Learning y las redes neuronales artificiales comprenden un amplio uso en diferentes campos como:

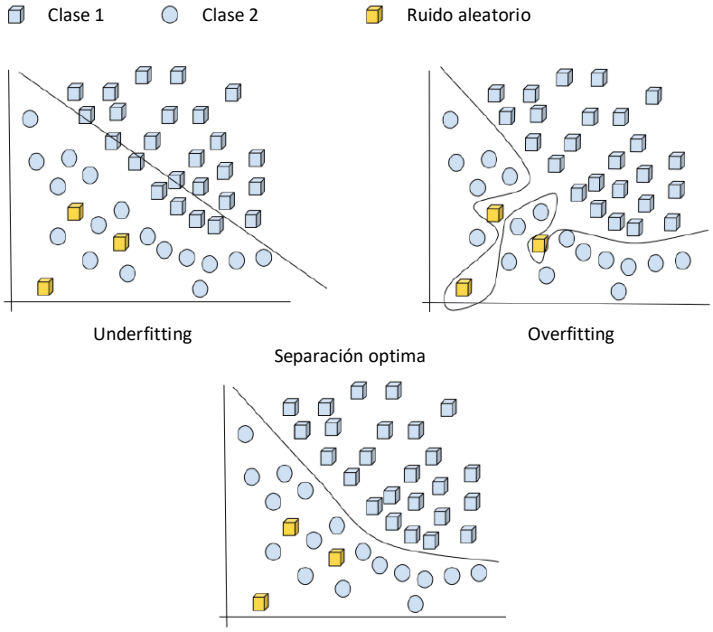
- Reconocimiento de imágenes
- Procesamiento de lenguaje natural
- Administración de portafolio y predicción de precios
- Robótica y Automóviles autónomos
- Reconocimiento de voz
- Descubrimiento de fármacos y diagnostico de enfermedades

2.5 Over-fitting / Under-fitting

Uno de los problemas más importantes en la aplicación del Machine Learning es el Over-fitting y el under-fitting. Problemas existentes a causa del “ruido”. Ruido se refiere a la información irrelevante o al azar que se encuentra presente en un conjunto de datos.

Los algoritmos intentan construir un modelo para que un conjunto de entradas pueda proporcionar la salida deseada, sin embargo, un modelo debe contener todo lo que es necesario para el modelado, y nada más. En términos de Aprendizaje automático, esto significa que el conjunto de entradas debe contener exactamente lo que se necesita para una buena predicción. La siguiente figura intenta mostrar los problemas de Over-fitting y under-fitting. Cuando el modelo enfatiza demasiado en tener un bajo error, el modelo crea un límite de decisión que es demasiado complicado e incluye el ruido. Cuando el modelo permite un error demasiado grande, no puede dividir las clases correctamente. Estos problemas pueden ser difíciles de manejar. Y, desafortunadamente, casi todos los conjuntos de datos del mundo real contienen algún tipo de ruido. En estos casos lo que sucede es que el algoritmo intenta predecir el ruido y crea un modelo que requiere demasiados inputs o que es demasiado complejo [Olden, 2016].

Figura 2.27 Ejemplos de overfitting, underfitting y separación optima



Fuente: Olden, 2016

Underfitting se refiere a un modelo que no modelará los datos de entrenamiento ni generalizará los datos nuevos. La mala adaptación ocurre porque el modelo no puede capturar la relación entre el conjunto de datos de entrada y la variable objetivo (target). Este problema causa que el modelo no pronostique los objetivos en los conjuntos de datos de entrenamiento con mucha precisión.

En el Overfitting el modelo se adapta de manera excesiva cuando se desempeña en el conjunto de datos de entrenamiento, pero no funciona de manera adecuada en los datos nuevos. Este ajuste excesivo ocurre cuando un modelo aprende los detalles y el ruido en los datos de entrenamiento en la medida en que impacta negativamente el rendimiento del modelo en datos nuevos.

Algunos métodos empleados para evitar o reducir problemas de Over y Underfitting son:

- Emplear una mayor cantidad de datos en la fase de entrenamiento del algoritmo.
- Técnicas de regularización
- Técnicas de validación cruzada que nos permiten evaluar un modelo varias veces con diferentes configuraciones en los datos. Este proyecto de investigación usa un método de validación llamado K-Fold Cross-Validation.

Capítulo 3 Diseño e implementación de los modelos

Como meta para este proyecto tenemos: Evaluar una selección de algoritmos de clasificación de Machine Learning para predecir el movimiento diario del precio del IPC y comparar su desempeño, así como evaluar su uso en estrategias de compra y venta de acciones en la BMV.

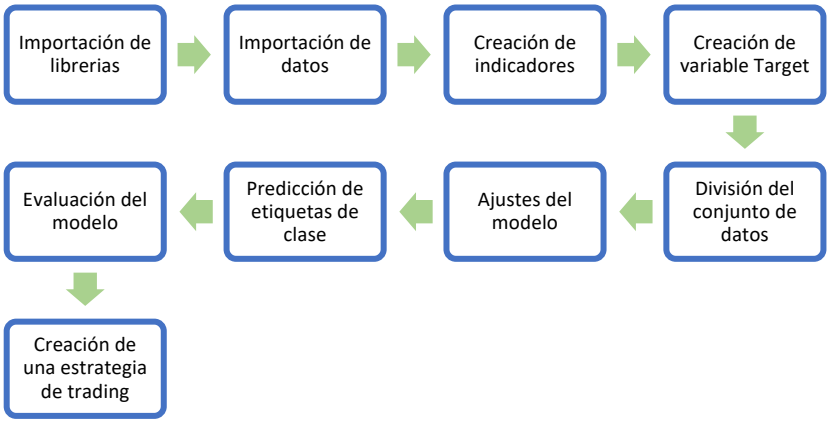
En primer lugar, el pronóstico se considera como un problema de predicción de dos clases, tomando como base si el precio subiría por encima de un umbral establecido o por el contrario si este bajaría en un periodo diario. De esta manera convertimos el problema en una pregunta de verdadero/falso para los algoritmos de Machine Learning.

Los algoritmos de Machine Learning fueron entrenados empleando diferentes configuraciones de datos (*DataSets*) como entrada, con el fin de evaluar su desempeño, así como destacar cual de estos se ajusta mas para el pronóstico de series de tiempo financieras.

En total se emplearon 7 algoritmos diferentes de Machine Learning y su desempeño se evaluó a través de medidas como: Exactitud (Accuracy), Precisión, Recall y F-Score empleadas comúnmente en modelos de clasificación.

El siguiente diagrama resume la metodología empleada en la implementación de los modelos de Machine Learning.

Figura 3.1 Metodología de implementación de los modelos



Fuente: Elaboración propia

3.1 Compilación de los datos

Empleando la metodología CRISP-DM las primeras tres fases se encargan de la selección, preparación y depuración de los datos a estudiar en un periodo determinado de tiempo, para el caso de este trabajo comprende la siguiente información de entrada que contiene los valores diarios de apertura y cierre según sea el caso.

El primer dataset (*dataset 1*) esta basado en indicadores financieros y macroeconómicos tanto globales como propios del mercado mexicano. El segundo dataset (*dataset 2*) solo consta de variables correspondientes al mercado mexicano. Finalmente, el tercer dataset (*dataset 3*) fue creado con base en indicadores de análisis técnicos propios del IPC.

Para realizar el pronóstico con los diferentes algoritmos, cada una de estas bases de datos se probaron con diferentes configuraciones de tiempo con base en la información disponible al momento de su elaboración.

La información incluida en el *dataset 1* y en el *dataset 3* abarcan el siguiente periodo:

- 10 años - 2 de enero de 2009 al 31 de diciembre de 2018

El *dataset 2* consiste en datos del siguiente periodo:

- 6 años y 4 meses aprox. - 30 de agosto de 2012 al 31 de diciembre de 2018

A continuación, se detallan los componentes de cada dataset seleccionados como variables de entrada para cada uno de los modelos.

Los datos fueron descargados de paginas de dominio publico como *yahoo finance, Investing.com, Banco de México, Bloomberg y Standard & Poors.*

Por otra parte, se creó una *variable target u objetivo* la cual es empleada como la variable que los algoritmos de clasificación de Machine Learning van a predecir. En este caso la *variable target* considera, si el precio de cierre del IPC cerrara al alza o a la baja comparado contra el precio de apertura del mismo día.

3.1.1 DataSet 1: Indicadores Financieros y macroeconómicos globales

El desempeño de las acciones de un país es a menudo inseparable de su crecimiento económico. Al igual que muchos índices en todo el mundo, el IPC se utilizó inicialmente como un indicador económico. Si bien el IPC no busca seguir todo el mercado, ya que hay muchas más empresas públicas en el mercado local, instrumentos de deuda, commodities, bienes raíces, etc., la evolución del mercado de valores tiene una estrecha relación con variables económicas [Sánchez, 2018].

El primer *dataset* consistió en una base conformada por indicadores financieros y Macroeconómicos del mundo y de México, que pueden afectar directamente el comportamiento en el precio de una acción mexicana, en este caso el IPC.

Con base en lo anterior los datos que conforman este *dataset* consisten en el precio de cierre diario de los siguientes 20 indicadores:

- IPC ($Cierre_{t-1}$ vs $Apertura_{t0}$)
- IPC Cierre
- Nasdaq Composite
- DJI – Dow Jones Industrial Average
- S&P 500 – Standard & Poors 500
- USD/MXN
- EUR/MXN
- TIIE 28
- S&P/BMV Government CETES Bond Index
- S&P/BMV Corporate Bond Index
- EEM – iShares MSCI Emerging Markets Index
- S&P Emerging Plus
- S&P Developed
- S&P Latin America 40
- XAU/USD
- Bloomberg Commodity Index
- Petróleo Mezcla Mexicana
- S&P China SX20

- S&P Brazil 15
- VIX - Volatility Index

La variable **IPC ($Cierre_{t-1}$ vs $Apertura_{t0}$)** esta conformada por el cambio porcentual del IPC entre el precio de cierre en el día t_{-1} y el precio de apertura de la mañana siguiente en t_0.

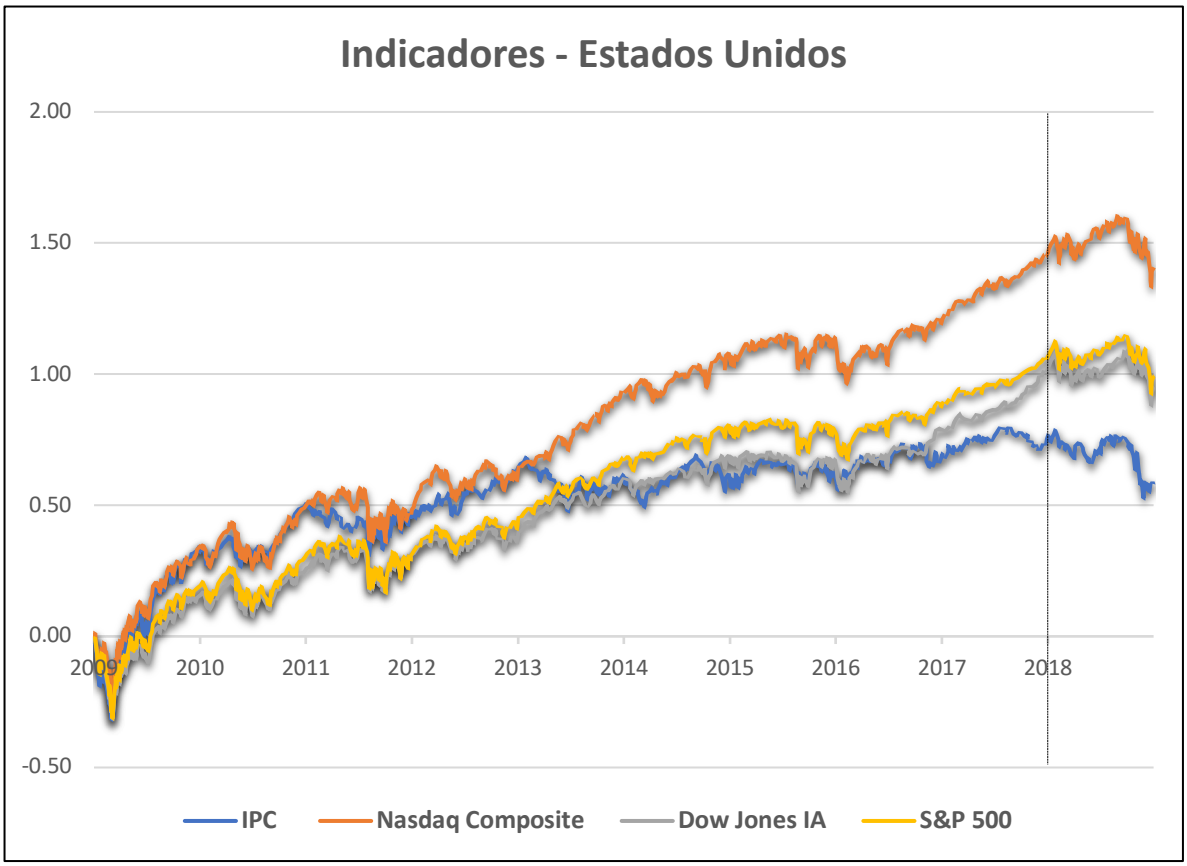
Las variables **USD/MXN** y **EUR/MXN** están compuestas por el cambio porcentual en el valor del peso mexicano cambiado por un dólar y un euro al cierre del día t_{-2} y en el día t_{-1}.

El resto de las variables se conforman por el cambio porcentual entre el valor de cierre en el día t_{-2} y el valor de cierre en t_{-1}.

Como se menciona anteriormente, las variables fueron seleccionadas con base en una posible relación que pudiese afectar directamente el precio del IPC.

Las siguientes gráficas muestran el comportamiento de las variables en el periodo seleccionado para realizar la investigación. La línea vertical representa la separación de los datos, entre los datos que se emplearon para la fase de entrenamiento y la fase de prueba.

Figura 3.2 Rendimientos del IPC vs Indicadores de Estados Unidos



Fuente: Elaboración propia

La economía mexicana se encuentra estrechamente relacionada con la economía de Estados Unidos, esto debido en parte a la economía globalizada. Como podemos observar en la gráfica el IPC mantiene una estrecha relación con los tres índices. Si bien en el periodo de 10 años en general podemos observar una tendencia positiva al alza, en ese lapso también podemos notar la caída de todos los índices tras la crisis global del 2008, así como una importante caída a finales del 2018 causada en parte por las políticas económicas de Donald Trump.

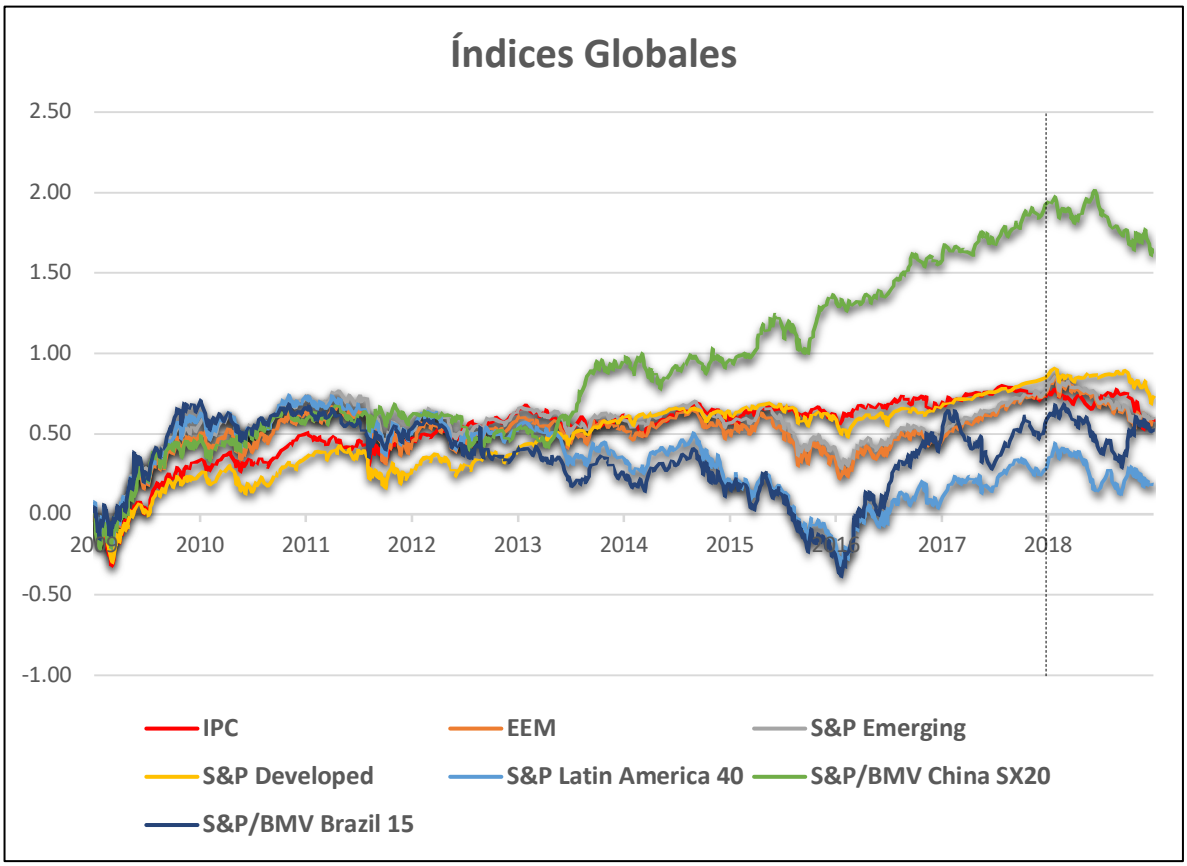
Se seleccionaron los tres mas importantes índices del mercado financiero de Estados Unidos:

El Nasdaq Composite es un índice bursátil que incluye los valores de las empresas presentes en el mercado Nasdaq, encargado de canalizar las empresas de alta tecnología, informática, telecomunicaciones y biotecnología entre otras.

El Promedio Industrial Dow Jones refleja el comportamiento de las acciones de las 30 compañías mas importantes de Estados Unidos listadas en la bolsa de valores de Nueva York (NYSE).

El índice Standard & Poor's 500 (S&P 500) es uno de los índices mas importantes de Estados Unidos. Esta formado por 500 empresas que cotizan en las bolsas NYSE y NASDAQ y se considera el índice mas representativo de la situación del mercado de Estados Unidos.

Figura 3.3 Rendimientos IPC vs Indicadores globales



Fuente: Elaboración propia

Todos estos índices fueron seleccionados con el fin de presentar variables globales que pueden afectar el desarrollo del IPC. En la gráfica podemos notar que los índices tuvieron movimientos similares hasta el año 2014, cuando el índice Chino comenzó un repunte importante, por otra parte, el índice Brasileño comienza una tendencia a la baja, propiciada por la crisis política y económica de ese país. El índice S&P Latin America 40 presenta una caída importante junto con el mercado brasileño causado por la influencia y composición de empresas brasileñas en este indicador.

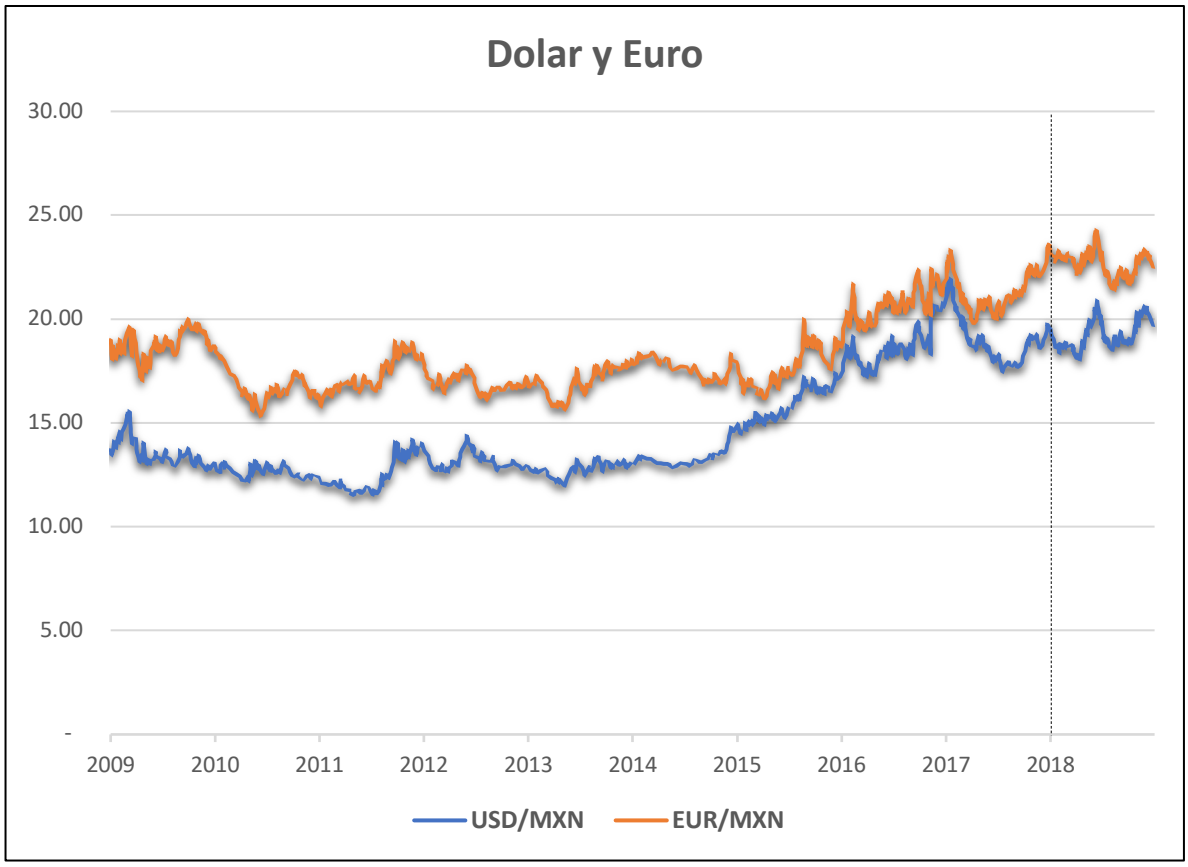
El ETF iShares MSCI Emerging Markets (EEM) busca rastrear los resultados de inversión de un índice compuesto por acciones de mercados emergentes de capitalización grande y mediana. Está diseñado para medir el desempeño del mercado de acciones en los mercados emergentes globales. El Índice consiste en valores de los siguientes 24 países de mercados emergentes: Brasil, Chile, China, Colombia, República Checa, Egipto, Grecia, Hungría, India, Indonesia, Malasia, México, Pakistán, Perú, Filipinas, Polonia, Qatar, Rusia, Sudáfrica, Corea del Sur, Taiwán, Tailandia, Turquía y los Emiratos Árabes Unidos y una parte significativa del Índice está representada por valores de compañías en las industrias o sectores financieros y de tecnología de la información.

S&P Dow Jones Indices crea una serie de índices que buscan cubrir ciertas particularidades, tal es el caso del índice S&P Emerging que captura compañías domiciliadas en mercados emergentes, el S&P Developed es un "benchmark" completo que incluye acciones de 25 mercados desarrollados. El S&P Latin America 40 incluye las 40 compañías líderes que capturan aproximadamente el 70% de la capitalización de mercado total de la región. Los componentes son elegidos entre los cinco mercados latinoamericanos principales: Brasil, Chile, Colombia, México y Perú.

El índice S&P/BMV China SX20 busca medir el rendimiento de los ADR y ADS chinos de mayor tamaño y liquidez, listados en la Bolsa de Valores de Nueva York (NYSE) y en el NASDAQ. Su objetivo es proporcionar un índice representativo, pero al mismo tiempo fácilmente replicable, que abarque el mercado chino a través de los certificados de depósitos Americanos ADR y ADS.

El índice S&P/BMV Brazil 15 busca medir el rendimiento de los 15 ADR brasileños de mayor tamaño y liquidez listados en la Bolsa de Valores de Nueva York (NYSE) y el NASDAQ.

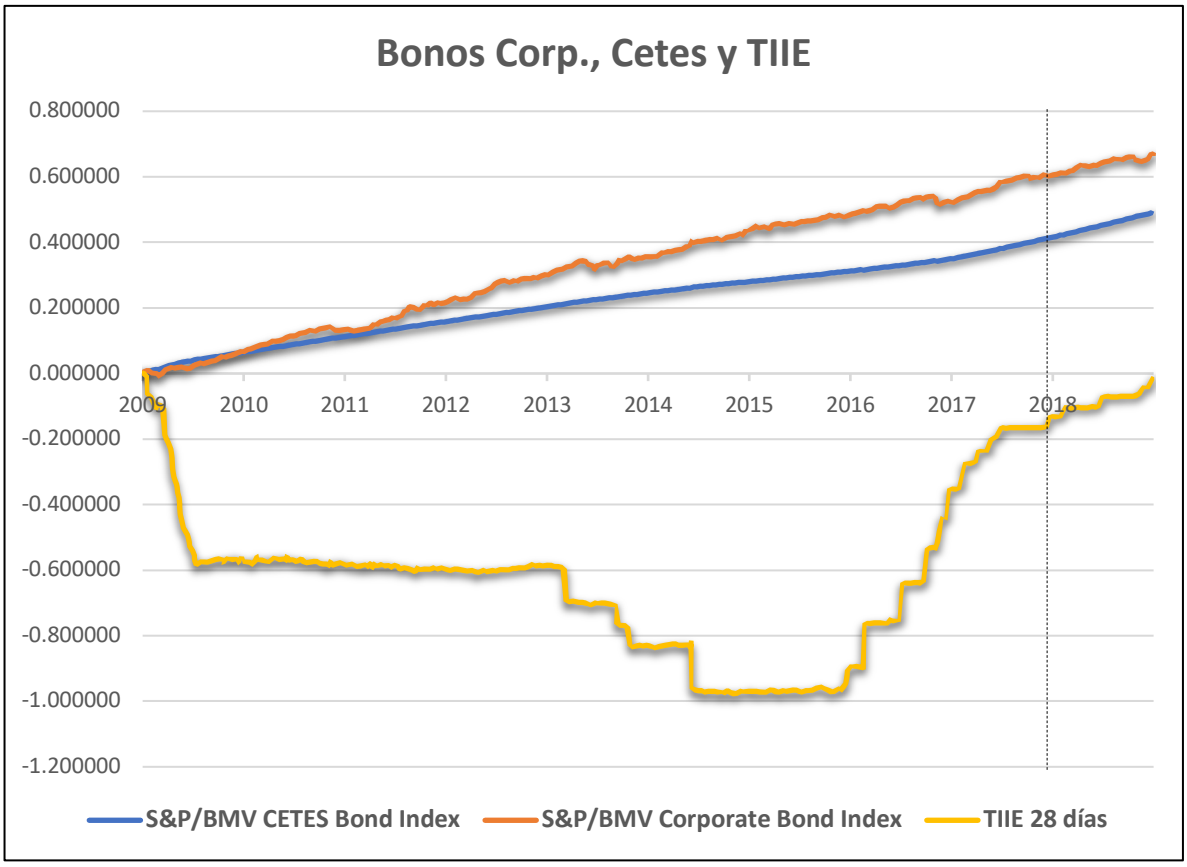
Figura 3.4 Precios Peso-Dólar y Peso-Euro



Fuente: Elaboración propia

La economía mexicana esta ligada a la economía norteamericana ya que es su principal socio comercial, y las operaciones se efectúan en dólares. De manera similar sucede un efecto parecido con el euro, aunque en menor proporción ya que las operaciones comerciales son reducidas. Así observamos la tendencia del dólar y el euro desde el 2009 hasta el 2018 en la gráfica. Donde se observa que el dólar rebasó la barrera de 15 pesos en 2009 con motivo de la crisis financiera y económica que azotó a todo el mundo. A partir de ese momento se incrementó la volatilidad en el tipo de cambio, producto de la propia recesión y el proceso de recuperación. Posteriormente el tipo de cambio permaneció estable. Sin embargo, desde principios del 2015 se ha observado una fuerte depreciación por factores como el alza de las tasas de interés de la Fed, la baja en el precio del petróleo, la preocupación por la debilidad china y el bajo crecimiento del país. El tipo de cambio peso-euro mantiene una simetría con el tipo de cambio peso-dólar.

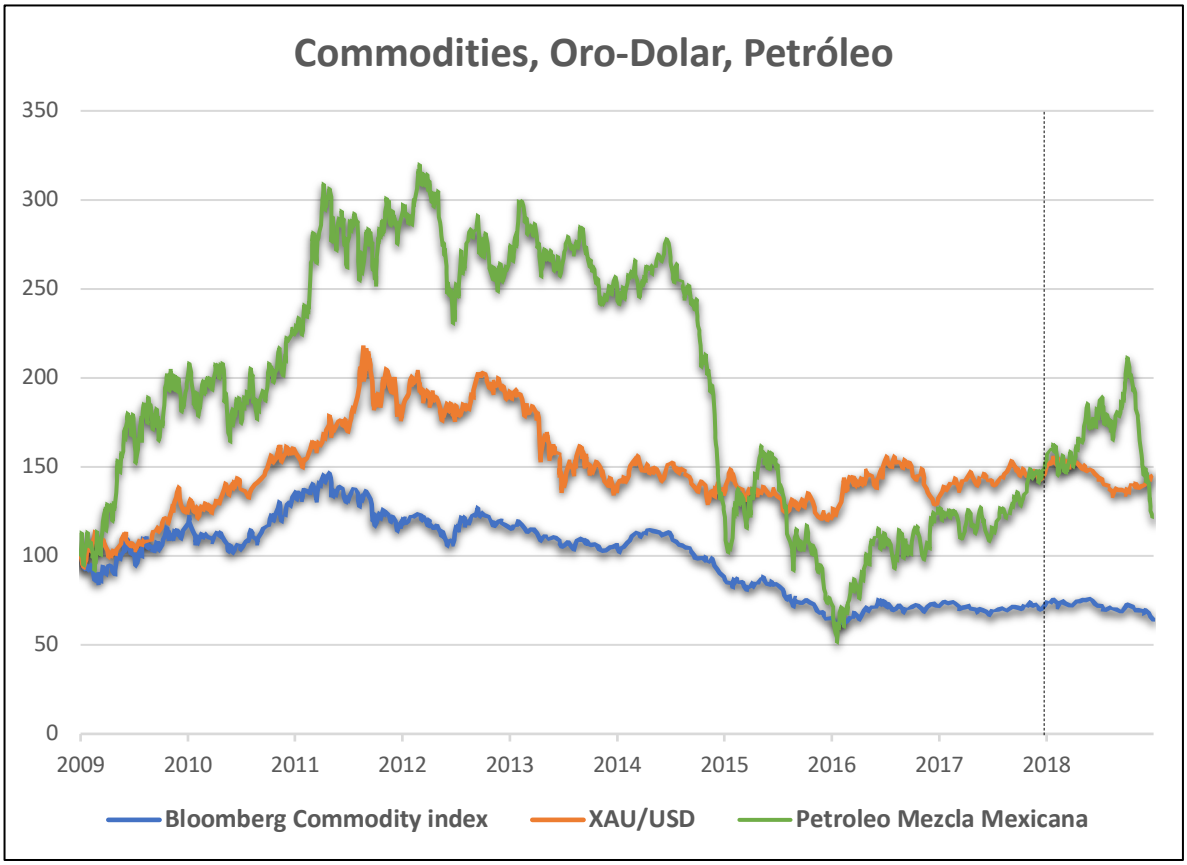
Figura 3.5 Rendimientos Cetes, Bonos Corporativos y TIIE



Fuente: Elaboración propia

La gráfica representa la participación del sector publico a través de los Certificados de la Tesorería de la Federación (**Cetes**) instrumentos que emite el gobierno para financiarse, ofreciendo a cambio una tasa de interés cobrable al final de un plazo establecido y el sector privado (**bonos corporativos**) en el mercado financiero, por otra parte, se observa el comportamiento de la Tasa de interés interbancaria de equilibrio (**TIIE**), la cual es una tasa representativa de las operaciones de crédito en la banca, generada por el Banco de México la cual se usa como referencia para establecer algunas tasas comerciales a nivel bancario.

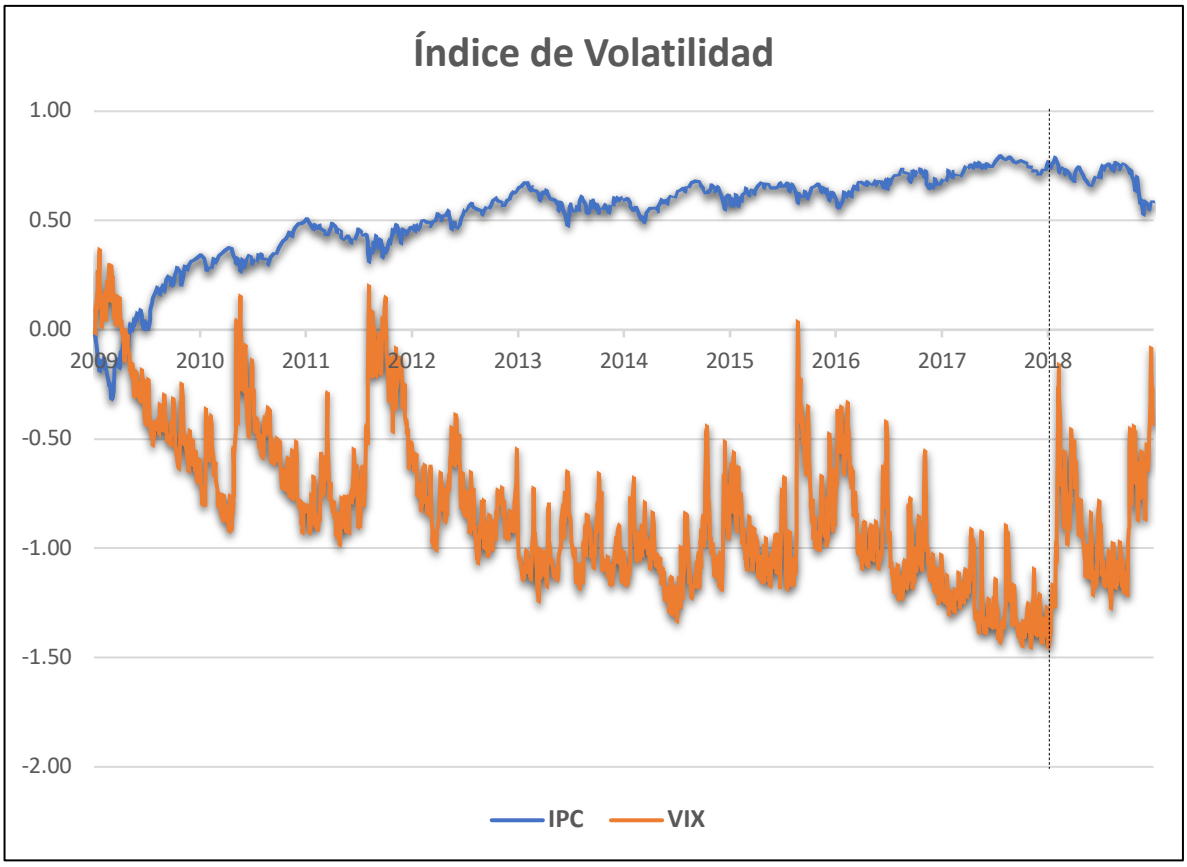
Figura 3.6 Comparativo base 100 Commodities, Oro-dólar y Petróleo Mezcla Mexicana



Fuente: Elaboración propia

La relación oro-dólar y el precio del petróleo influyen en el IPC ya que economías como la mexicana dependiente de los ingresos del petróleo y divisas provenientes de sus exportaciones principalmente de commodities, los cuales influyen en la trayectoria de las empresas que cotizan en el mercado bursátil.

Figura 3.7 Rendimientos IPC vs Índice de Volatilidad (VIX)



Fuente: Elaboración propia

El índice de volatilidad VIX conocido como el indicador del miedo se emplea para reflejar el miedo y pesimismo de los mercados financieros, se encuentra correlacionado al S&P 500 índice que tiene gran repercusión en el resto de las bolsas mundiales. Este índice es de gran importancia ya que nos dice el sentimiento del mercado sobre las bolsas mundiales. Cuando tenemos un índice VIX muy bajo, significa que la volatilidad es muy baja y por tanto no hay miedo en el mercado, lo que hace que las bolsas sigan subiendo. Las grandes caídas en las bolsas vienen cuando un VIX está muy bajo y comienza a subir. Por el contrario, un VIX muy alto significa que la volatilidad es muy alta y eso se traduce en nerviosismo en el mercado.

3.1.2 DataSet 2: Indicadores Financieros y macroeconómicos mexicanos

El segundo dataset, consiste exclusivamente en información del mercado mexicano para evaluar el comportamiento de los algoritmos y se conforma por 20 índices creados por la Bolsa Mexicana de Valores en conjunto con Standard and Poors.

Los índices de la Bolsa Mexicana de Valores y Standard and Poors, según su especialidad y enfoque, son indicadores que pretenden reflejar el comportamiento del mercado bursátil mexicano como un todo, o también agrupando a algunos emisores que comparten cierta característica [BMV, 2018].

Los indicadores seleccionados se clasifican de la siguiente manera:

Índices Sectoriales Analíticos de S&P/BMV

Los Índices Sectoriales Analíticos de S&P/BMV buscan medir el rendimiento de los sectores económicos dentro del mercado bursátil mexicano, de acuerdo con el sistema de clasificación industrial creado por la Bolsa Mexicana de Valores. Los componentes de estos índices son ponderados por capitalización de mercado [BMV, 2018].

- S&P/BMV Sector Materiales
- S&P/BMV Sector Industrial
- S&P/BMV Sector Productos de Consumo Frecuente
- S&P/BMV Sector Servicio de Consumo no Básico
- S&P/BMV Sector Salud
- S&P/BMV Sector Servicios Financieros
- S&P/BMV Sector Servicios de Comunicación

Índices de actividad económica

Los Índices de Actividad Económica de la Bolsa Mexicana de Valores y S&P Dow Jones son indicadores que reflejan el comportamiento de los diferentes Sectores (Primario, Secundario y Terciario) de la Bolsa Mexicana de Valores que conforman la Actividad Económica del país al incluir los más líquidos. Las empresas de cada actividad en sus listas constituyentes, en función de las variaciones de precios de una lista constitutiva equilibrada, ponderada y representativa [BMV, 2018].

Sector primario: Actividades en las que se extraen y explotan los recursos naturales, tal como se obtienen de la naturaleza, ya sea para alimentos o para la generación de materias primas (consumo o comercialización).

- Agricultura
- Explotación forestal
- Cría de ganado, ganadería
- Minería
- Pescar

Sector secundario: Se refiere a la industria en general, donde el uso de maquinaria y procesos automatizados para transformar las materias primas obtenidas del sector primario (incluidas fábricas, manufactura, talleres, laboratorios, etc.) es predominante. Los productos más complejos se desarrollan a partir de materias primas.

- Construcción
- Electricidad, gas y agua.
- Industria manufacturera

Sector terciario: No hay producción de bienes materiales como tal, aquí los productos fabricados en el sector secundario (incluidas las comunicaciones y el transporte) se venden o explotan (servicios)

Producen comodidad y satisfacción de las necesidades humanas a través de la prestación de servicios.

- Comercio
- Servicios
- Transporte

Con estos índices, la Bolsa de Valores de México y los Índices de S&P ofrecen opciones para seguir el comportamiento del mercado de valores al agrupar en las diferentes listas de constituyentes todas las actividades económicas que participan en él [BMV, 2018].

Los índices de actividad económica son los siguientes:

- S&P/BMV Extractiva (Minería y Agricultura)
- S&P/BMV Transforma (Industria manufacturera, electricidad, gas y agua)
- S&P/BMV Construye (Construcción)
- S&P/BMV Comercio (Casas Comerciales y Distribuidores)
- S&P/BMV Enlace (Infraestructura y Transportes)
- S&P/BMV Servicios (Servicios Financieros)
- S&P/BMV Servicios Comerciales (Comercio y Prestación de Servicios)

Índices fundamentales

Se compone de dos índices:

- S&P/BMV Bursa Óptimo
- S&P/BMV Índice de Calidad, Valor y Crecimiento

El índice S&P/BMV Bursa Óptimo busca medir el rendimiento de las acciones de mayor tamaño y liquidez listadas en la Bolsa Mexicana de Valores, mediante un esquema de ponderación que toma en cuenta la capitalización de mercado ajustada al capital flotante, la liquidez y variables fundamentales como los ingresos netos, ganancias operativas, margen de ganancia, relación precio-ganancias (P/E) y relación deuda neta-patrimonio [S&P, 2018].

El S&P/BMV Índice de Calidad, Valor y Crecimiento busca medir el rendimiento de acciones que forman parte del S&P/BMV IPC y que presentan la combinación más elevada de calidad, valor y crecimiento [S&P, 2018].

De manera adicional este modelo se alimenta con los siguientes índices representativos de los bonos emitidos por el gobierno y por el sector privado:

- S&P/BMV Government CETES Bond Index
- S&P/BMV Corporate Bond Index

El S&P/BMV Government CETES Bond Index busca medir el rendimiento de Certificados de la Tesorería (CETES) de México. Los componentes del índice deben ser títulos emitidos por el gobierno mexicano con plazos de vencimiento inferiores a un año y denominados en pesos mexicanos [S&P, 2018].

El S&P/BMV Corporate Bond Index, es diseñado para medir el rendimiento de los títulos emitidos por corporativos mexicanos, con vencimiento mayor a 31 días y denominados en pesos mexicanos [S&P, 2018].

3.1.3 DataSet 3: Indicadores de análisis técnico

Los indicadores generados a través del análisis técnico son empleados de manera común por gestores e inversionistas en la creación de estrategias de trading.

Esta base consiste en indicadores de análisis técnico, obtenidos a través de datos diarios del IPC, transformados por medio de fórmulas matemáticas establecidas a partir de los datos de apertura, máximo, mínimo, cierre y volumen.

La siguiente tabla resume los indicadores seleccionados como input para los modelos, así como las fórmulas empleadas en su calculo:

Tabla 3.1 Indicadores técnicos y fórmulas

Indicador Técnico	Fórmula
SMA (Media Móvil Simple)	$\frac{C_t + C_{t-1} + \ \ldots + C_{t-n}}{n}$
RSI (Relative Strength Index)	$100 - \frac{100}{1+\left(\sum_{i=0}^{n-1}\frac{Up_{t-i}}{n}\right) / \left(\sum_{i=0}^{n-1}\frac{Down_{t-i}}{n}\right)}$
Williams %R	$\frac{H_n - C_t}{H_n - L_n} * 100$
SAR	$SAR_{n+1} = SAR_n + \ \propto (EP - SAR_n)$
ADX (Average Directional Index)	$ADX = Average\ DX14$ $DX = 100 * \left(\frac{DI\ 14\ Diff}{DI\ 14\ Sum}\right)$
MACD (Moving Average convergence/Divergence)	$MACD(n)_{t-1} + \frac{2}{n} + 1 \ * (DIFF_t - MACD(n)_{t-1})$
Stochastic Oscilator %K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} * 100$
Stochastic Oscilator %D	$\frac{\sum_{i=0}^{n-1} K_{t-i}\%}{n}$
Momentum	$C_t - C_{t-n}$
Weighted Moving Average	$\frac{(n) * C_t + (n-1) * C_{t-1} + \cdots + C_n}{n + (n-1) + \cdots + 1)}$

C_t= Precio de cierre

L_t=Precio mínimo en t

H_t=Precio máximo en t

Up_t=Cambio de precio hacia el alza en t

$DOWN_t$=cambio en el precio hacia abajo en tiempo t

LL_t y HH_t= el mínimo más bajo y el máximo más alto en los últimos t días.

Indicador Técnico	Fórmula
ROC (Rate of Change)	$\frac{C_t - C_{t-n}}{C_{t-n}} * 100$
CCI (Commodity Channel Index)	$\frac{M_t - SM_t}{0.015 - D_t}$
Bollinger Bands	$TP = \frac{High+Low+Close}{3}$ $MidBand = SMA * TP$
AD (Accumulation/Distribution Line)	$CLV = \frac{(Close - Low) - (High - Close)}{(High - low)}$ $\text{AD} = AD_{-1} + CLV * Volume$
OBV (On Balance Volume)	$OBV = OBV_{-1}$+ Volume $OBV = OBV_{-1} -$ Volume $OBV = OBV_{-1}$
TR (True Range)	$TrueHigh = Highest\ of\ High_0\ or\ Close_{-1}$ $TrueLow = Lowest\ of\ Low_0\ or\ Close_{-1}$ $TR = TrueHigh - TrueLow$
Average Price	$\frac{Open + High + Low + Close}{4}$
Weighted Close	$\frac{High + Low + Close * 2}{4}$
EMA (Exponential Moving Average)	$K = \frac{2}{n+1}$ $EMA = K * input + (1 - K) * EMA_{-1}$

∝= Factor de aceleración

EP = Precio Extremo

TP=Typical Price

CLV = Close Location Value

Fuente: Elaboración propia con información de Anzola (2015), Python Ta-Lib y FMLabs.com

3.1.4 Variable de salida (Target)

Esta base se considera como la variable de salida (output) y es el dato o clase que los algoritmos buscaran predecir.

La variable se incorpora al modelo con el fin de “entrenar” a cada uno de los algoritmos en la fase de aprendizaje/entrenamiento.

Debido a que los modelos buscan predecir el cambio diario entre el precio de apertura y el precio de cierre, se creo una señal (binaria) que nos indica si el cambio porcentual entre el precio de apertura y el precio de cierre del día tiene un aumento o un decremento:

$$\text{Target} \begin{cases} 1, \text{cambio porcentual} > 0.5\% \\ -1, \text{cambio porcentual a la baja} \end{cases}$$

Si la variable indica un “1” quiere decir que el precio de cierre es mayor que el precio de apertura del día, esto podemos interpretarlo como una señal de “compra”, mientras que en caso de tener un “-1” se considera como una señal de “venta” ya que el precio de cierre es menor al precio de apertura de ese día.

Se aplico un umbral de 0.5% al cambio porcentual del día, que nos indica cuanto debe subir el precio de la acción para considerarlo como una señal positiva, con el fin de crear una estrategia de trading que nos permita maximizar las ganancias y cubrir posibles gastos de comisión por la compra/venta del instrumento a través de una casa de bolsa.

3.2 Normalización

Dado que valores con magnitudes diferentes pueden generar problemas en la ejecución de los algoritmos se aplica una normalización de los datos de entrada, en este caso, normalizar significa comprimir los valores de la variable para que estos se encuentren en un rango definido.

Para el caso del Dataset 1 y 2 se realizó un preprocesamiento de los datos con el objetivo de obtener valores relativos, es decir cambios porcentuales calculados a través de la siguiente fórmula y aplicadas a cada una de las variables mencionadas:

$$[(P_t - P_{t-1}) / P_{t-1}] \times 100 \qquad (3.1)$$

Al convertir estas variables de entrada y salida en tasas, se obtienen variables de entrada normalizadas.

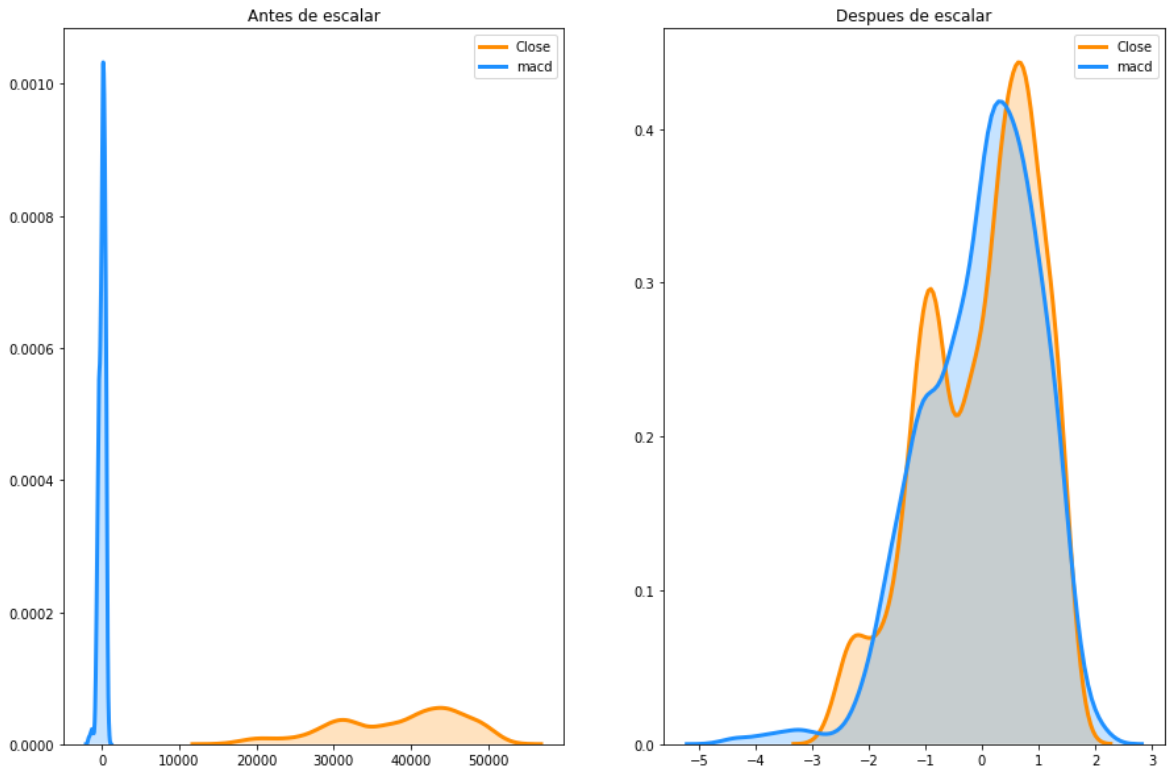
Para el Dataset 3 se aplicó un *Escalado Estándar o Estandarización*, esta técnica hace que los datos se reajusten para que tengan las propiedades de una distribución normal estándar convirtiendo la media de los datos a cero y su varianza a 1, permitiéndonos de esta manera normalizar los datos entre unos límites definidos (el máximo y el mínimo de la variable) empleando la siguiente fórmula:

$$X_{Normalizada} = \frac{X - X_{media}}{X_{desv.est.}} \qquad (3.2)$$

Esto nos ayuda a disminuir los sesgos (bias) al entrenar el modelo. Si no se realiza este paso corremos el riesgo de que los algoritmos den un mayor peso a aquellas características que tienen un valor promedio mas alto que otras.

La siguiente gráfica muestra una comparativa entre la distribución de los datos del precio de cierre diario del IPC y el indicador MACD antes y después del escalado.

Figura 3.8 Comparativo entre datos antes y después de escalar



Fuente: Elaboración propia

En la siguiente tabla, de igual manera, podemos observar el cambio que se lleva a cabo con el preprocesamiento de los datos.

Tabla 3.2 Comparativo entre datos antes y después de escalar

	Antes de escalar	Después de escalar
Media	38,552.78	-4.29
Desviación estándar	8,068.06	1.00
Valor Máximo	51,713.37	1.63
Valor Mínimo	16,891.02	-2.68

Fuente: Elaboración propia

3.3 Filtro Hodrick Prescott

Como caso de estudio adicional se realizó la aplicación de un filtro de suavizamiento al dataset 1 y al dataset 2 con el fin comparar el desempeño de los algoritmos cuando son probados con datos que contienen menos “ruido”.

El Filtro Hodrick-Precott (Filtro HP) es una herramienta matemática utilizada principalmente en macroeconomía para realizar la descomposición de series de tiempo, lo que implica separar una serie de tiempo en dos componentes; componentes de ciclo y componentes de tendencia.

El Filtro HP crea una representación no lineal suavizada de una serie de tiempo que es menos sensible a las fluctuaciones de corto plazo que a las fluctuaciones de largo plazo. El filtro de Hodrick-Prescott asume que cualquier serie de tiempo **X** determinada puede dividirse en un componente de tendencia $\mathbf{T_t}$ y un componente cíclico $\mathbf{C_t}$ y expresarse por la suma $\mathbf{X_t = T_t + C_t}$. El componente cíclico se puede obtener restando T de X dando $\mathbf{C_t = X_t - T_t}$. El componente cíclico y el componente de tendencia se pueden aislar resolviendo el siguiente problema de minimización [Larsen,2010]:

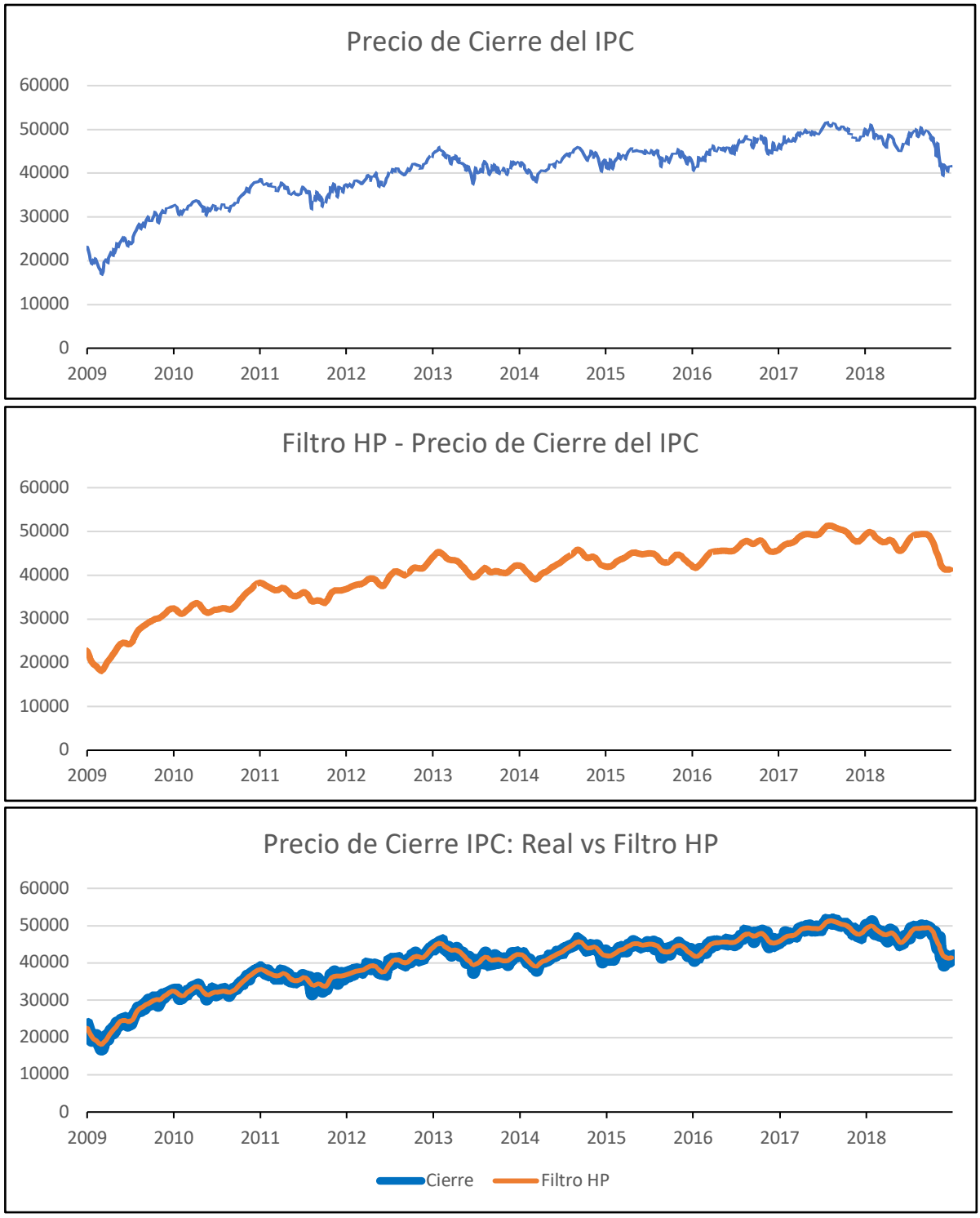
$$min_\tau \sum_{t=1}^{T} (x_t - |\tau_t)^2 + \lambda \sum_{t=2}^{T} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 x_t \in X \qquad (3.3)$$

Donde el primer término de la ecuación es la suma de la desviación al cuadrado de $\mathbf{C_t}$ y el segundo término es la primera diferencia del componente de tendencia. Al resolver el problema de minimización, el primer término penaliza los valores grandes de $\mathbf{C_t}$ **,** mientras que el segundo término penaliza la falta de suavidad en $\mathbf{T_t}$. La compensación entre los dos términos es controlada por el parámetro λ. En consecuencia, los valores más altos de λ penalizan las variaciones en la primera diferencia del componente de tendencia que causa una línea de tendencia más suave que es menos sensible a las fluctuaciones a corto plazo que a las fluctuaciones a largo plazo (esencialmente controla el grado de suavizado en las fluctuaciones a corto plazo). A medida que λ se aproxima a 0, el componente de tendencia se aproxima a la serie de tiempo original, y cuando λ se acerca a infinito $\mathbf{T_t}$ se aproxima a una tendencia lineal [Larsen, 2010] [Guerrero,2011].

Para esta investigación se utiliza el valor típico **λ = 1600** para el suavizado de los datos.

Las siguientes gráficas comparan el precio de cierre del IPC con los datos suavizados una vez que se aplico el Filtro HP. Este filtro se aplico a todas las variables de entrada y posteriormente se realizo la normalización de los datos indicada en la sección 3.2.

Figura 3.9 Datos reales vs filtro HP



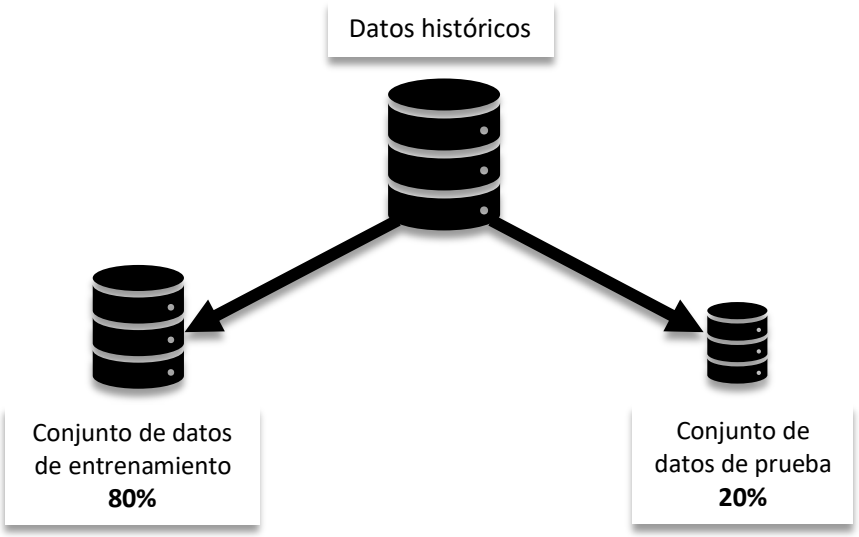
Fuente: Elaboración propia

3.4 División de datos

Aprender los parámetros de una función de predicción y probarla con los mismos datos es un error metodológico: un modelo que simplemente repetiría las etiquetas de las muestras que acaba de ver tendría una puntuación perfecta pero no podría predecir nada útil en datos aún no vistos. Para evitar el problema anterior, es una práctica común cuando se realiza un experimento de Machine Learning separar los datos disponibles en conjuntos de entrenamiento y de prueba [scikit-learn.org, 2018].

Para las primeras pruebas se selecciono el 80% de los datos para utilizarlos en la fase de entrenamiento de los algoritmos y el 20% de los datos restantes se reservó para la fase de pruebas.

Figura 3.10 División de los datos



Fuente: Elaboración propia

3.5 Validación cruzada de K iteraciones (K-Fold Cross-Validation)

Es una técnica que nos permite hacer un uso más eficiente de los datos que tenemos con el fin de minimizar el sesgo asociado con el muestreo aleatorio de los datos de entrenamiento y prueba en la comparación de la precisión predictiva de dos o más métodos [Desai et al., 1996]. A través de este método podemos disminuir los problemas asociados

al over y under-fitting revisados en la sección 2.5 y mantener la habilidad de generalización de los algoritmos de Machine Learning.

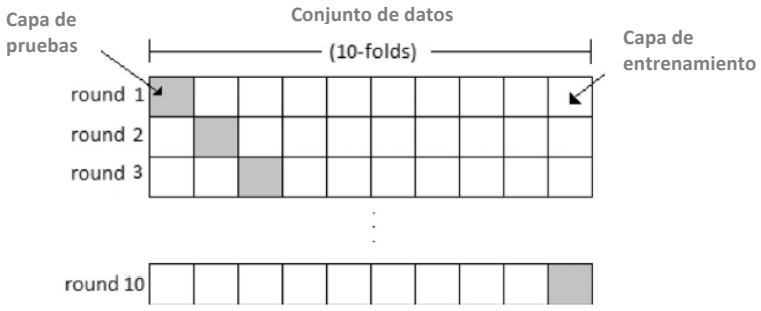
En la validación cruzada k-fold, el conjunto de datos completo (D) se divide aleatoriamente en k subconjuntos mutuamente exclusivos (bloques: D_1, D_2, ..., D_K) de aproximadamente el mismo tamaño. El modelo de clasificación es entrenado y probado k veces. Cada vez ($t \in \{1, 2, ..., k\}$), se entrena en todos menos un pliegue (D_t) que se reserva para la prueba. El desempeño de la validación cruzada k-fold se calcula simplemente como el promedio de las medidas de rendimiento individuales *k*:

$$CV = \frac{1}{k} \sum_{i=1}^{k} PM_i \qquad (3.4)$$

Donde *CV* significa validación cruzada, *k* es el número de bloques utilizados y *PM* es la medida de desempeño para cada bloque [Olson y Denle, 2008].

Para estimar el rendimiento de los modelos en esta investigación se utilizó una validación cruzada de 10 bloques o capas. Los estudios empíricos mostraron que 10 parece ser un número óptimo de pliegues que optimiza el tiempo que se tarda en completar la prueba y minimiza el sesgo y la varianza asociados con el proceso de validación [Kohavi, 1995]. En la validación cruzada de 10 bloques, todo el conjunto de datos se divide en 10 subconjuntos. Cada bloque se usa una vez para probar el rendimiento del modelo de predicción que se genera a partir de los datos combinados de los nueve bloques restantes, lo que lleva a 10 estimaciones de rendimiento independientes como se ve en la figura [Oztekin et al., 2016].

Figura 3.11 Representación gráfica de la validación cruzada de 10 bloques



Fuente: Oztekin et al., 2016

3.6 Implementación de los modelos de clasificación

Una vez que creamos un conjunto de datos históricos con entradas relevantes y dividimos los datos en conjuntos de entrenamiento y de prueba utilizamos los datos de entrenamiento para entrenar nuestros algoritmos y hacer un pronóstico sobre el precio futuro del IPC.

El proceso de aprendizaje se basa en deducir una función que asocie los datos con su respectiva clase (subida o bajada del precio) a partir de la observación de ejemplos, llamados datos de entrenamiento. Estos ejemplos son, como ya mencionamos inputs y outputs, correctamente asociados que el algoritmo analiza y de donde obtiene conclusiones, es decir detecta patrones y relaciones no visibles a simple vista. Por tanto, queremos crear una función capaz de predecir la clase correspondiente a cualquier objeto de entrada después de haber visualizado una serie de ejemplos. El proceso de adaptación de las normas y pautas aprendidas por el algoritmo a nuevos ejemplos (desconocidos hasta el momento) se conoce como generalización [Huertas, 2015].

Una vez que el algoritmo se aplica a los datos de entrenamiento, lo usamos en los datos de prueba y analizamos su rendimiento comparando la clase pronosticada con la variable *target* real.

Para la implementación de los algoritmos se empleó el lenguaje de programación Python en conjunto con la librería Scikit-learn especializada en módulos de análisis de datos y Machine Learning. A través de este entorno obtenemos el beneficio de reducir los tiempos en la implementación de los algoritmos, así como la facilidad de ajustar los parámetros de estos según se requiera.

3.7 Medidas de desempeño

Algunas de las medidas de desempeño mas populares en los modelos de clasificación son la Exactitud, Precisión, Recall y el F-Score. Estas medidas se calculan con base en el número de verdaderos/falsos positivos y negativos. Para efectos de esta investigación un verdadero positivo se da cuando el modelo pronostica correctamente que el precio de cierre del IPC aumentara (por encima del umbral definido) con respecto al precio de apertura.

Falso positivo se da cuando el modelo pronostica de manera incorrecta que el precio aumentara. Por otra parte, un verdadero negativo significa que el modelo pronostica correctamente que el precio no aumentara y un falso negativo significa que pronosticó incorrectamente que el precio no aumentaría.

3.7.1 Exactitud

Es un criterio de desempeño práctico que muestra el éxito direccional del modelo propuesto. Es la relación de las predicciones correctas de movimientos al alza o a la baja.

A través de esta medida calculamos la exactitud del modelo, ya sea la fracción o el conteo de las predicciones correctas. Si todo el conjunto de etiquetas predichas para una muestra coincide estrictamente con el verdadero conjunto de etiquetas, entonces la precisión del subconjunto es 1.0; De lo contrario es 0.0.

Si $y\hat{}$ es el valor pronosticado de la muestra i y y es el valor real correspondiente, entonces la fracción de predicciones correctas sobre el número de muestras $n_{muestras}$ se define como [scikit-learn, 2018]:

$$Exactitud(y, y\hat{}) = \frac{1}{n_{muestras}} \sum_{i=0}^{n_{muestras}-1} 1(y\hat{}_i = y_i) \qquad (3.5)$$

3.7.2 Reporte de clasificación

Es un informe de texto que muestra las principales métricas de clasificación. Miden la capacidad del modelo para reconocer los valores positivos y negativos.

El reporte de clasificación cuenta con las siguientes métricas [scikit-learn.org, 2018]:

- Precisión. La precisión es la relación tp/(tp+fp) donde tp es el número de verdaderos positivos y fp el número de falsos positivos. La precisión es intuitivamente la capacidad del clasificador para no etiquetar como positiva una muestra que es negativa.

- La recuperación (recall) es la relación tp/(tp+fn) donde tp es el número de verdaderos positivos y fn el número de falsos negativos. La recuperación es intuitivamente la capacidad del clasificador para encontrar todas las muestras positivas.

- La F-score se puede interpretar como una media armónica ponderada de la precisión y el recall, donde una puntuación F-beta alcanza su mejor valor en 1 y la peor puntuación en 0.

  La puntuación F-Score otorga pesos a la precisión y al recall por un factor de beta. Beta = 1 significa que la precisión y el recall son igualmente importantes.

En este contexto, podemos definir las nociones de precisión, recall y F-Score [scikit-learn, 2018]:

$$Precision = \frac{tp}{tp + fp} \tag{3.6}$$

$$Recall = \frac{tp}{tp + fn} \tag{3.7}$$

$$F\beta = (1 + \beta^2)\frac{precision * recall}{\beta^2\, precision + recall} \tag{3.8}$$

3.7.3 Matriz de confusión

Evalúa la precisión de la clasificación al calcular la matriz de confusión con cada fila correspondiente a la clase verdadera.

Una matriz de confusión es un resumen de los resultados del modelo de clasificación mostrados en una representación matricial. Las filas representan las clases reales y las columnas representan las clases predichas.

En una tarea de clasificación binaria, los términos ‘’positivo’’ y ‘’negativo’’ se refieren a la predicción del clasificador, y los términos ‘’verdadero’’ y ‘’falso’’ se refieren a si esa predicción corresponde a la observación real. Dadas estas definiciones, podemos formular la siguiente tabla:

Figura 3.12 Matriz de confusión

		Clase pronosticada
Clase actual (observación)	**TP (verdadero positivo)** Resultado correcto	**FP (falso positivo).** Resultado inesperado
	FN (falso negativo) Resultado faltante	**TN (verdadero negativo)** Correcta ausencia de resultado

Fuente: Elaboración propia

Capítulo 4 Resultados

En esta sección se revisan los resultados obtenidos para las diferentes configuraciones. La sección 4.1 contempla la aplicación de los modelos con el dataset 1 y el dataset 2 sin haber sido suavizados (datos brutos) así como el dataset 3. La sección 4.2 presenta los resultados cuando el dataset 1 y el dataset 2 fueron procesados con el filtro de suavizamiento Hodrick-Prescott. Esto nos permite comparar el desempeño de los modelos cuando son alimentados con datos que contienen mayor o menor ruido.

4.1 Resultados sin la aplicación del filtro HP

4.1.1 Exactitud de los modelos

En esta sección se muestra la exactitud conseguida por cada uno de los modelos con la división de datos de 80% para la fase de entrenamiento y 20% para la fase de pruebas en cada una de las tres diferentes configuraciones de datos. El siguiente cuadro muestra el número de registros empleado por los modelos para cada una de las fases:

Tabla 4.1 Número de registros por base de datos

Base de datos	Entrenamiento	Pruebas
DataSet 1	2,193	243
DataSet 2	1,389	154
DataSet 3	2,250	250

Fuente: Elaboración propia

Los resultados conseguidos por los modelos se encuentran en un rango de entre 48.22% obtenido por la maquina de soporte vectorial aplicado en el *dataset 2* (el peor) y 54.69% conseguido por los arboles de decisión en el *dataset 2* (el mejor) para la fase de pruebas, siendo la maquina de soporte vectorial el algoritmo con el menor porcentaje de éxito y los arboles de decisión el algoritmo con el mejor desempeño.

En estos resultados la exactitud en el pronóstico de la fase de pruebas nos proporciona los resultados reales de los modelos, al ser en esta fase en la que los algoritmos aplican los resultados obtenidos durante su fase de entrenamiento.

Tabla 4.2 Resultados de Exactitud

Dataset 1 (Var. Mundiales-Mexicanas)

	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
Entrenamiento:	72.18%	60.93%	59.50%	59.29%	54.62%	53.08%	98.10%
Pruebas:	48.57%	48.77%	50.00%	49.59%	52.46%	52.46%	52.87%

Dataset 2 (Var. Mexicanas)

	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
Entrenamiento:	62.32%	60.29%	58.91%	59.40%	56.81%	52.11%	100.00%
Pruebas:	48.22%	48.87%	54.69%	48.87%	51.13%	49.51%	52.75%

Dataset 3 (Var. Análisis técnico)

	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
Entrenamiento:	58.15%	62.85%	55.15%	56.20%	54.30%	52.00%	92.20%
Pruebas:	51.80%	51.20%	48.80%	51.60%	50.40%	49.00%	50.60%

Fuente: Elaboración propia

Como podemos observar en estos resultados, las diferentes configuraciones de datos no representan una diferencia significativa que nos ayude a obtener una mejora en el éxito de los pronósticos. Los modelos aplicados en el *dataset 1* promedian una exactitud de 50.67%, en el *dataset 2* un 50.58% y finalmente en el *dataset 3* promedian un 50.49%. Como se mencionó anteriormente tanto el peor como el mejor resultado obtenido por los modelos para todas las configuraciones se dio en el *dataset 2*. En el *dataset 1* la mayor exactitud la tiene el modelo de la Red Neuronal Artificial (52.87%) y la menor exactitud la obtuvo la maquina de soporte vectorial (48.57%). Para el dataset 3 el modelo con los mejores resultados fue la maquina de soporte vectorial (51.80%) y el de menor desempeño fueron los arboles de decisión (48.80%).

4.1.2 Reportes de Clasificación

En esta sección se presenta el reporte de clasificación y las matrices de confusión elaborado para cada uno de los modelos con la división de datos 80/20.

La siguiente tabla muestra el reporte de clasificación en el cual podemos medir la precisión, Recall y F1-Score así como la matriz de confusión para cada uno de los algoritmos probados.

Como se menciono en la sección 3.7.2 en la clasificación binaria, algunas de las medidas de rendimiento más comunes son Precisión, Recall y F-Score. Estas son medidas estadísticas que se calculan a partir del número de verdaderos / falsos positivos y negativos.

Para el problema en esta tesis, un verdadero positivo sería cuando el modelo predice correctamente que el precio aumentaría por encima del umbral establecido al final del día. Falso positivo se da cuando el modelo predice incorrectamente que el precio del índice aumentaría. Por otro lado, un verdadero negativo significa que predijo correctamente que el precio no aumentaría, y un falso negativo significa que predijo incorrectamente que el precio no aumentaría.

Si alguien compra acciones basándose en el pronóstico de un modelo binario, probablemente sería prudente utilizar un algoritmo que rara vez predice un falso positivo, ya que un falso positivo causaría que compre una acción y pierda dinero. Por esa razón, se puede argumentar que la precisión es la mejor medida, ya que es una medida de la relación entre verdaderos positivos y falsos positivos o en otras palabras, la proporción de cuántas veces el modelo obtendría una ganancia en comparación con cuantas veces el modelo crearía una pérdida, lo que significaría que minimizaría el riesgo de perder dinero. Debido a esto, el modelo con la mayor precisión puede ser el mejor si uno quiere obtener una ganancia. Esto sería especialmente cierto en períodos bajistas cuando no se espera que el mercado aumente.

Recall mide la proporción de la frecuencia con la que el algoritmo predice correctamente que una acción subirá frente al número de veces que falla. Cuanto mayor sea el recall, más a menudo el algoritmo identificará el aumento y, por lo tanto, obtendrá ganancias. Si bien

el uso de este indicador pudiera resultar riesgoso a menudo las estrategias de mayor riesgo pueden ser las más lucrativas.

El F-Score considera tanto la precisión como el recall y, por lo tanto, es más informativo que cualquiera de ellos por sí mismos. Se calcula utilizando la media armónica de los dos valores. Por lo tanto, es probable que el algoritmo con el F-Score más alto tenga un buen desempeño en períodos bajistas, alcistas y en general, en cualquier punto intermedio.

En este caso de estudio, en general, los modelos presentan desempeños neutros ubicándose cerca del 50 % para las 3 evaluaciones. En el dataset 1 los modelos con la mejor precisión y recall son las redes neuronales artificiales y Naive Bayes con 53% para ambas pruebas, en el F-Score las redes neuronales obtienen nuevamente un 53 % en promedio. En el dataset 2 los arboles de decisión obtienen los mejores resultados promedio con 56% para las pruebas de precisión y recall y 55% en la prueba F-score. Por último en el dataset 3 las maquinas de soporte vectorial y los bosques aleatorios consiguen un 52% promedio en la precisión y recall, en la F-score las maquinas de soporte vectorial nuevamente son el mejor algoritmo con 52%

Con respecto a las matrices de confusión obtenidas, para el dataset 1 la matriz de confusión que presenta los mejores resultados es la red neuronal artificial, acertando 111 señales de compra, sin embargo, clasificando erróneamente 134, con respecto a las señales de venta los arboles de decisión aciertan 204 veces y fallan solo en 39 señales.

Para el dataset 2 el algoritmo Naive Bayes acierta en pronosticar 87 señales de compra, fallando 76. En las señales de los bosques aleatorios son el algoritmo con mejor desempeño pronosticando correctamente 128 señales y tan solo fallando en 18 señales.

En el dataset 3 el algoritmo Naive Bayes y los algoritmos de decisión tiene el mejor pronóstico en cuanto a señales de venta obteniendo 230 y 226 señales correctas respectivamente. El mejor algoritmo para detectar las señales de venta fueron las maquinas de soporte vectorial pronosticando 135 señales de manera correcta.

Tabla 4.3 Reporte de clasificación

Dataset 1 (Var. Mundiales-Mexicanas)

Máquinas de soporte vectorial

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.48	0.31	0.38	77	168
1	0.49	0.66	0.56	83	160
Avg	0.49	0.49	0.47		

K-Nearest Neighbor

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.49	0.43	0.46	106	139
1	0.49	0.54	0.51	111	132
Avg	0.49	0.49	0.49		

Árboles de Decisión

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.51	0.16	0.25	40	205
1	0.50	0.84	0.63	39	204
Avg	0.51	0.50	0.44		

Bosques Aleatorios

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.50	0.29	0.37	71	174
1	0.50	0.70	0.58	72	171
Avg	0.50	0.50	0.48		

Regresión Logística

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.54	0.34	0.42	84	161
1	0.52	0.71	0.60	71	172
Avg	0.53	0.53	0.51		

Naïve Bayes

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.54	0.36	0.43	88	157
1	0.52	0.69	0.59	75	168
Avg	0.53	0.53	0.51		

Red Neuronal MLP

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.54	0.45	0.49	111	134
1	0.52	0.60	0.56	96	147
Avg	0.53	0.53	0.53		

Dataset 2 (Var. Mexicanas)

Máquinas de soporte vectorial

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.52	0.28	0.36	45	118
1	0.47	0.71	0.57	42	104
Avg	0.50	0.50	0.47		

K-Nearest Neighbor

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.52	0.40	0.46	66	97
1	0.47	0.58	0.52	61	85
Avg	0.50	0.49	0.49		

Árboles de Decisión

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.59	0.47	0.52	76	87
1	0.52	0.64	0.57	53	93
Avg	0.56	0.56	0.55		

Bosques Aleatorios

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.56	0.14	0.23	23	140
1	0.48	0.88	0.62	18	128
Avg	0.52	0.51	0.43		

Regresión Logística

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.55	0.41	0.47	67	96
1	0.49	0.62	0.55	55	91
Avg	0.52	0.52	0.51		

Naïve Bayes

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.52	0.53	0.53	87	76
1	0.46	0.45	0.46	80	66
Avg	0.49	0.49	0.50		

Red Neuronal MLP

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.56	0.50	0.53	82	81
1	0.50	0.55	0.53	65	81
Avg	0.53	0.53	0.53		

Dataset 3 (Var. Análisis técnico)

Máquinas de soporte vectorial

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.52	0.49	0.51	124	127
1	0.52	0.54	0.53	114	135
Avg	0.52	0.52	0.52		

K-Nearest Neighbor

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.51	0.56	0.53	140	111
1	0.51	0.47	0.49	133	116
Avg	0.51	0.52	0.51		

Árboles de Decisión

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.49	0.90	0.64	226	25
1	0.42	0.07	0.12	231	18
Avg	0.46	0.49	0.38		

Bosques Aleatorios

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.51	0.72	0.60	180	71
1	0.52	0.31	0.39	171	78
Avg	0.52	0.52	0.50		

Regresión Logística

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.51	0.49	0.50	124	127
1	0.50	0.51	0.51	121	128
Avg	0.51	0.50	0.51		

Naïve Bayes

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.50	0.92	0.64	230	21
1	0.42	0.06	0.11	234	15
Avg	0.46	0.49	0.38		

Red Neuronal MLP

Clase	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.51	0.56	0.53	141	110
1	0.50	0.45	0.48	137	112
Avg	0.51	0.51	0.51		

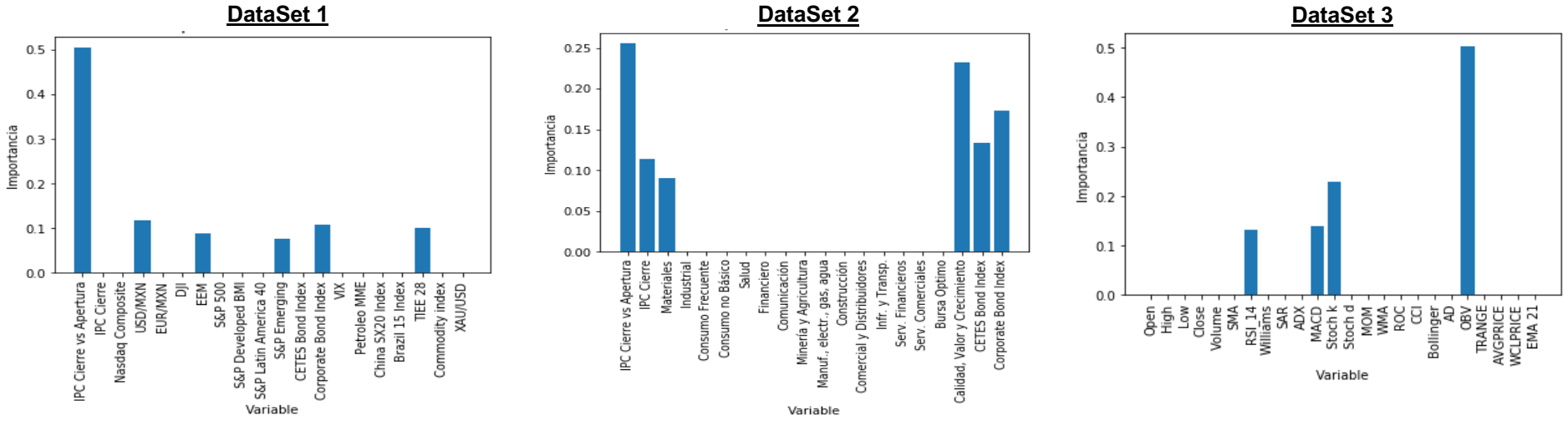
4.1.3 Importancia de las variables

Un inconveniente importante en la mayoría de los algoritmos de Machine Learning es el problema de la caja negra, sin embargo, los algoritmos de Arboles de decisión y Bosques aleatorios a diferencia del resto nos permiten medir la importancia de cada variable de entrada en la predicción, así como observar los parámetros que los modelos establecieron para cada uno de los arboles generados.

Las siguientes gráficas creadas por el modelo muestran la importancia relativa o la contribución de cada característica en la predicción. Calcula automáticamente la puntuación de relevancia de cada función en la fase de entrenamiento. Luego, reduce la relevancia para que la suma de todas las puntuaciones sea 1.

Árboles de decisión

Figura 4.1 Importancia de las variables – Arboles de decisión

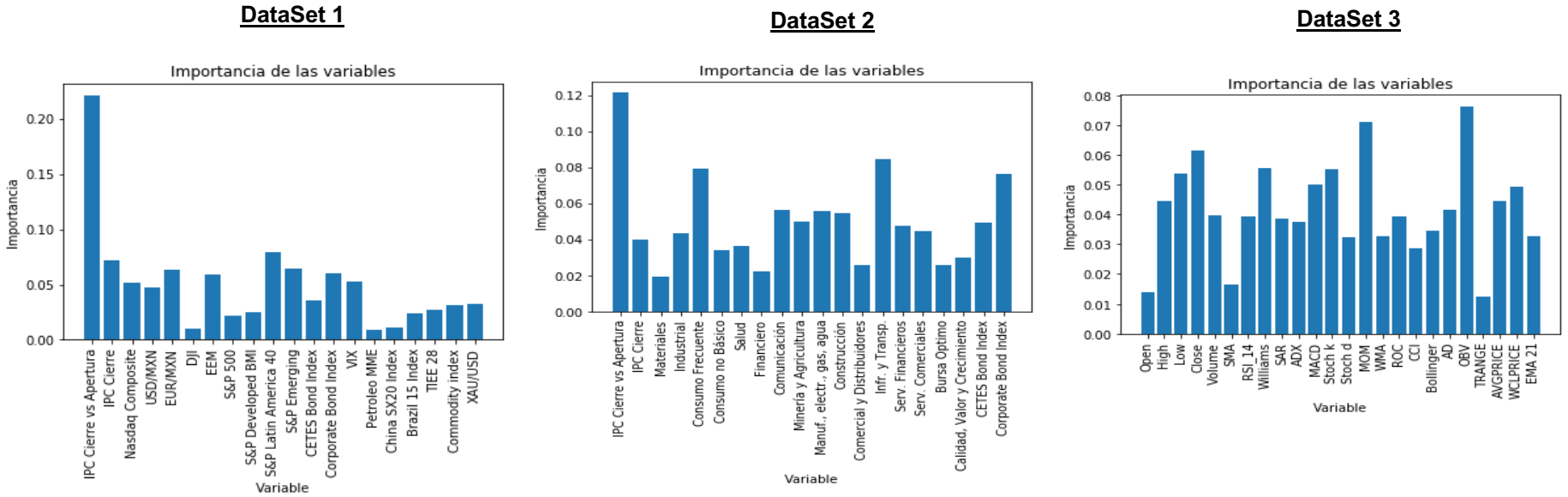


Fuente: Elaboración propia

Podemos observar que cuando el algoritmo fue entrenado con el dataset 1 y el dataset 2 da la mayor importancia a la diferencia porcentual que existe entre el precio de cierre del día anterior y el precio de apertura del día actual. Posteriormente con los datos del dataset 1 otorgo una mayor importancia al tipo de cambio Peso-Dólar, al índice de bonos corporativos, la TIIE 28 días y finalmente a los 2 índices que siguen el comportamiento de los mercados emergentes (EEM y S&P Emerging markets). Cuando el modelo fue entrenado con el dataset 2 otorgó una gran importancia al índice "Calidad, valor y crecimiento", seguido de los bonos corporativos, CETES así como el precio de cierre del IPC y el índice de materiales. Para el dataset 3 la mayor importancia corresponde al indicador OBV (On Balance Volume) que relaciona el volumen de negociación con los cambios de precio.

Bosques aleatorios

Figura 4.2 Importancia de las variables – Bosques aleatorios



Fuente: Elaboración propia

Para el caso de los Bosques aleatorios la mayor importancia nuevamente corresponde a la diferencia porcentual que existe entre el precio de cierre del día anterior y el precio de apertura del día actual para el dataset 1 y el dataset 2. De igual manera cuando el modelo fue entrenado con los datos del dataset 3 la mayor importancia corresponde al indicador técnico OBV (On Balance Volume) seguido del indicador MOM (Momentum) que compara el precio actual con el precio anterior de una cantidad en períodos anteriores, en este caso del IPC.

4.1.4 K-Fold Cross-Validation

La siguiente tabla muestra los resultados obtenidos a través de los modelos de K-Fold Cross Validation para cada una de las 10 capas.

Tabla 4.4 Resultados K-fold cross validation

Dataset 1 (Var. Mundiales-Mexicanas)

Fold No.	Máquinas Soporte Sectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
1	52.4	51.6	57.7	59.0	49.5	53.6	54.9
2	53.7	50.0	57.8	57.8	54.9	51.6	54.9
3	53.7	48.4	50.4	50.8	48.8	50.4	50.4
4	47.1	52.0	59.0	55.3	55.7	48.8	50.0
5	48.0	50.0	52.9	50.8	50.8	49.2	54.9
6	50.0	52.5	52.9	50.8	49.2	52.0	57.0
7	56.0	54.3	54.3	56.8	53.1	54.3	51.9
8	44.9	48.1	53.5	55.1	50.6	46.9	49.8
9	49.8	51.9	48.6	46.1	50.2	52.3	51.4
10	46.9	46.1	49.8	53.1	53.9	51.9	50.2
Mean	50.2	50.4	53.6	53.5	51.6	51.1	52.5
Std. Dev	3.4	2.3	3.4	3.7	2.3	2.1	2.4

Dataset 2 (Var. Mexicanas)

Fold No.	Máquinas de soporte vectorial	K-Nearest Neighbor	Árboles de Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
1	55.4	49.6	48.3	52.9	54.8	54.8	43.2
2	46.5	46.5	45.8	47.1	51.6	43.9	53.5
3	53.5	51.6	51.6	50.3	49.7	47.1	47.1
4	55.8	49.4	50.0	55.2	51.3	52.6	54.5
5	56.5	51.3	53.9	55.8	59.7	46.8	55.2
6	46.8	46.8	47.4	51.3	51.9	52.6	47.4
7	51.9	51.9	48.7	50.6	48.1	50.0	53.9
8	52.6	42.9	50.0	44.2	46.1	44.8	53.2
9	45.5	42.2	57.1	50.6	49.4	55.2	50.6
10	46.8	44.8	48.7	45.5	48.1	48.7	51.3
Mean	51.1	47.6	50.1	50.3	51.0	49.6	51.0
Std. Dev	4.1	3.4	3.1	3.6	3.7	3.8	3.7

Dataset 3 (Var. Análisis técnico)

Fold No.	Máquinas de soporte vectorial	K-Nearest Neighbor	Árboles de Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
1	57.9	54.4	57.1	57.5	55.2	57.5	53.2
2	58.8	50.8	58.4	58.0	55.6	59.6	46.4
3	47.2	49.2	48.4	49.2	52.0	49.6	52.8
4	46.0	53.6	44.8	54.0	44.4	48.4	53.6
5	49.2	48.8	53.6	52.4	52.8	50.8	54.8
6	52.4	50.4	52.4	52.0	52.8	49.6	52.4
7	49.6	49.2	49.6	50.0	50.8	52.8	48.8
8	51.6	50.4	46.0	46.8	52.0	44.0	51.2
9	54.8	54.4	49.6	54.0	50.8	49.6	46.0
10	48.8	46.8	48.0	50.0	52.4	48.0	48.4
Mean	51.6	50.8	50.8	52.4	51.8	51.0	50.7
Std. Dev	4.1	2.4	4.2	3.4	2.9	4.4	2.9

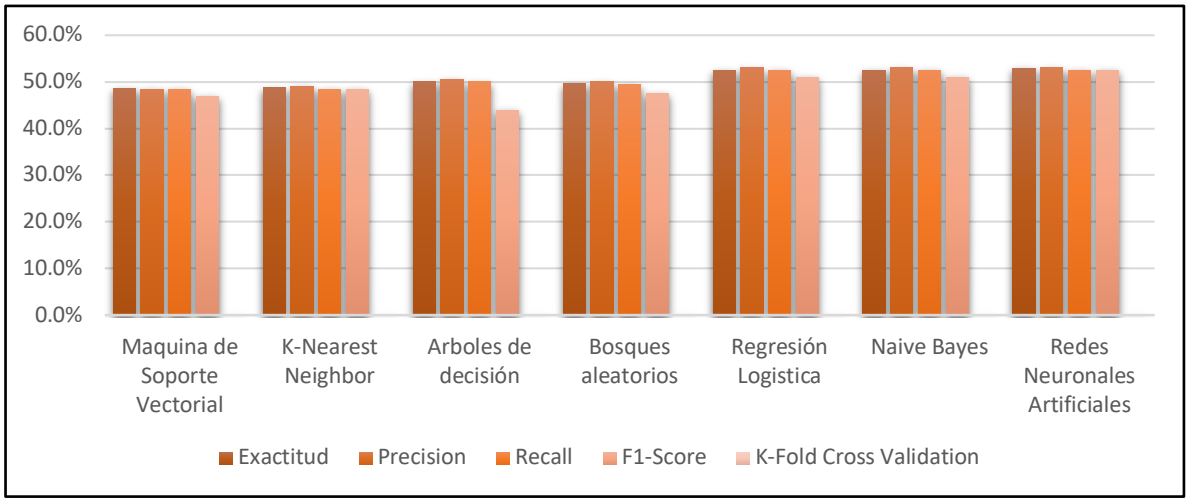
Fuente: Elaboración propia

A través de este método podemos notar que los modelos mejoran ligeramente sus resultados, aunque sin llegar a otorgar un aumento significativo en el porcentaje de éxito de los pronósticos. Los algoritmos de arboles de decisión y bosques aleatorios presentan los mejores resultados en cuanto a exactitud con un 53.6% y 53.5% respectivamente entrenados con el dataset 1. El peor desempeño lo obtiene el algoritmo K-Nearest Neighbor entrenado en el dataset 2 con un 47.6% de exactitud.

Las siguientes gráficas resumen y comparan los principales resultados obtenidos por los algoritmos en cada uno de los métodos de evaluación aplicados.

DataSet 1

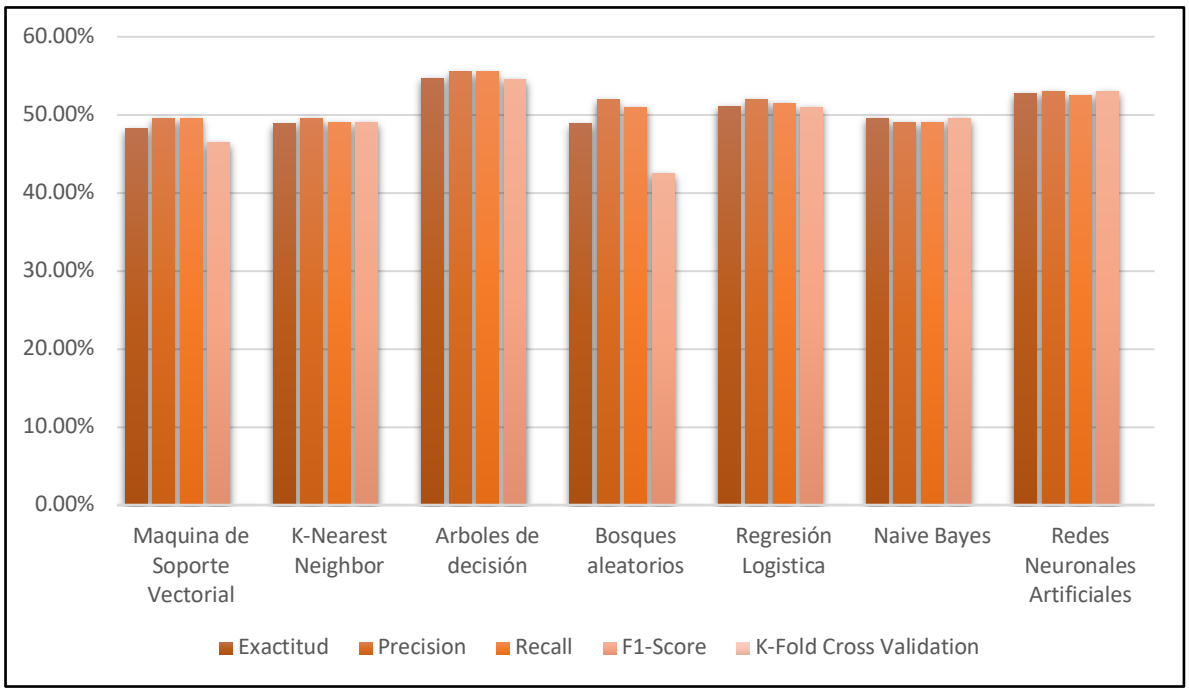
Figura 4.3 Comparativo de resultados - Dataset 1



Fuente: Elaboración propia

DataSet 2

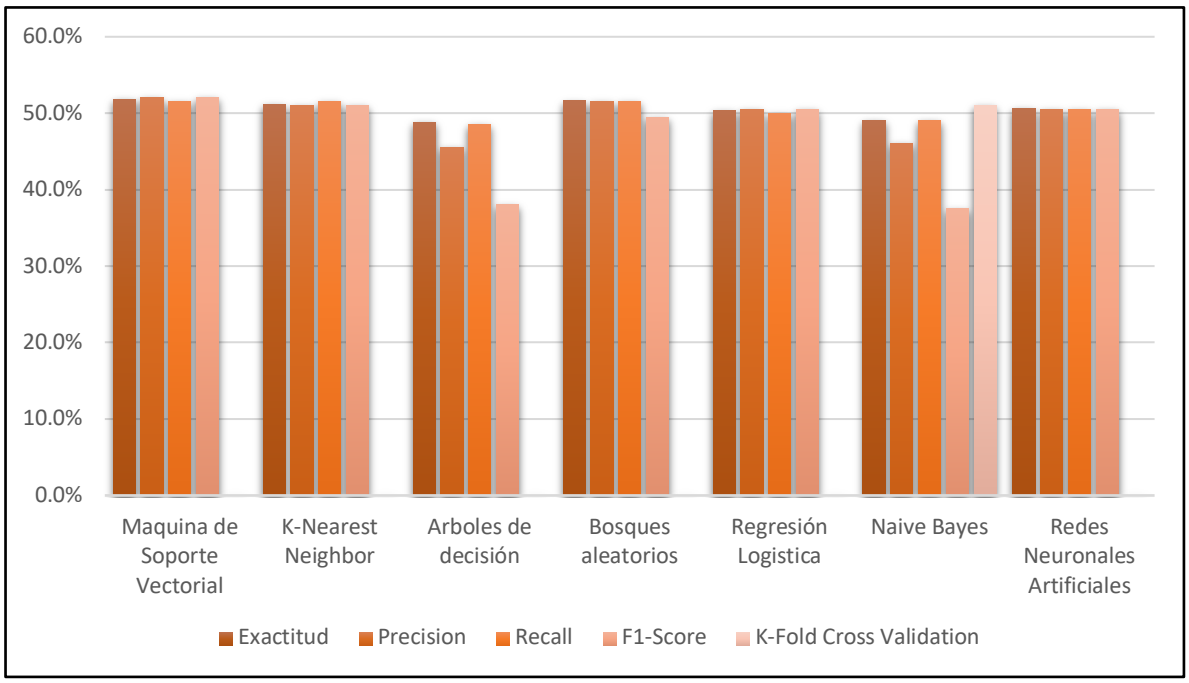
Figura 4.4 Comparativo de resultados - Dataset 2



Fuente: Elaboración propia

DataSet 3

Figura 4.5 Comparativo de resultados - Dataset 3



Fuente: Elaboración propia

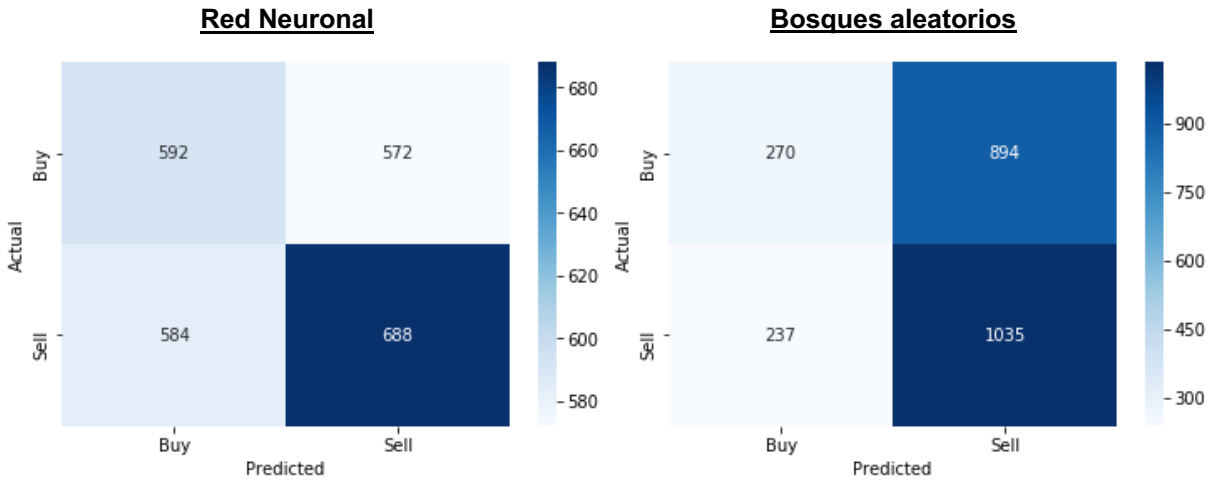
Como se menciona anteriormente todos los algoritmos se mantienen cercanos al 50% en cada una de las evaluaciones realizadas. Los diferentes conjuntos de datos en que estos fueron entrenados y probados no representaron diferencias significativas o mejoras considerables en su desempeño. Por otra parte, el tamaño en el periodo de tiempo que abarcan cada una de las bases tampoco influyo en obtener mejores o peores resultados.

4.1.5 Matrices de confusión

A continuación, se muestran solo las mejores matrices de confusión obtenidas por los algoritmos en cada una de las tres configuraciones de datos empleando la validación cruzada de 10 capas.

DataSet 1

Figura 4.6 Matrices de confusión – Dataset 1

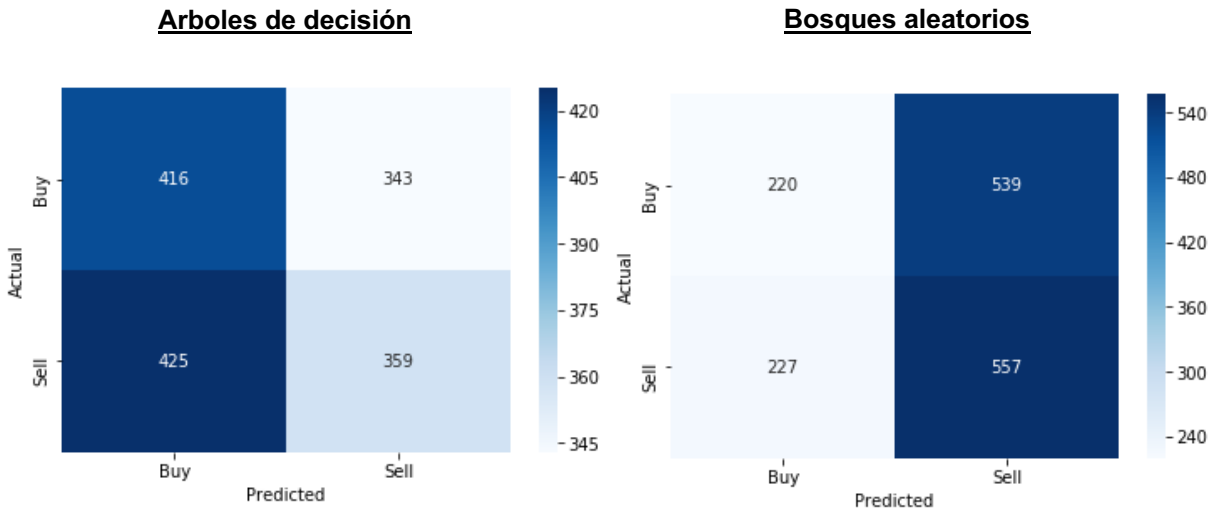


Fuente: Elaboración propia

Para las pruebas realizadas en el DataSet 1 las redes neuronales presentaron los mejores resultados en la predicción de las señales de compra al conseguir pronosticar 592 registros de manera correcta. Por otra parte, los Bosques aleatorios se desempeñaron de mejor manera que el resto de los algoritmos al pronosticar las señales de venta con 1,035 registros correctos y tan solo 237 señales incorrectas.

DataSet 2

Figura 4.7 Matrices de confusión – Dataset 2

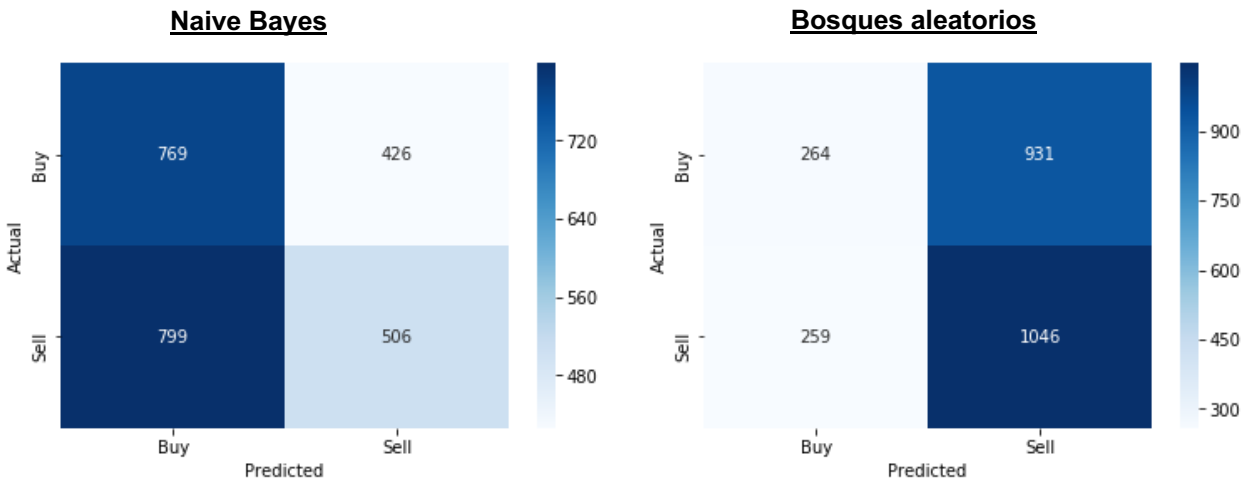


Fuente: Elaboración propia

En el DataSet 2 el modelo de los Arboles de decisión arrojo los mejores resultados en la predicción al pronosticar 416 señales de compra de manera correcta y 343 de manera incorrecta. Con respecto a las señales de venta los Bosques aleatorios nuevamente son el mejor en pronosticar las señales de venta con 557 registros correctos y 227 incorrectos.

DataSet 3

Figura 4.8 Matrices de confusión – Dataset 3



Fuente: Elaboración propia

Finalmente, para el DataSet 3 el modelo de Naive Bayes tuvo el mejor desempeño al pronosticar 769 señales de compra de manera correcta contra 426 señales de manera incorrecta. En cuanto a las señales de venta una vez mas los Bosques aleatorios destacan al pronosticar 1,046 señales de manera correcta y solo 259 de manera incorrecta.

De estos resultados podemos destacar que los modelos se desempeñan mejor al pronosticar un movimiento a la baja en el precio del IPC, esto quiere decir que la mayoría de los modelos se desempeñaría mejor durante un mercado bajista.

4.2 Resultados con la aplicación del filtro Hodrick Prescott

4.2.1 Exactitud de los modelos

Como podemos observar en la siguiente tabla, la aplicación de un filtro de suavizamiento nos permitió obtener mejores resultados una vez que se elimino parte del ruido presente en las series de tiempo financieras. Los resultados generados cuando los modelos se aplicaron al dataset 1 varían entre un 51.84% y un 56.56% correspondiente al algoritmo de Regresión Logística.

Por otra parte, los modelos entrenados con el dataset 2 presentan los mejores resultados de la investigación al tener una exactitud de entre 54.69% y 58.25% conseguida por los algoritmos de Regresión Logística y Bosques aleatorios.

Tabla 4.5 Resultados de Exactitud

Dataset 1 (Var. Mundiales-Mexicanas)

	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
Entrenamiento:	57.75%	60.63%	60.42%	58.06%	57.85%	56.47%	64.17%
Prueba:	52.25%	50.82%	53.48%	56.15%	56.56%	53.89%	51.84%

Dataset 2 (Var. Mexicanas)

	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
Entrenamiento:	57.13%	60.21%	59.48%	58.43%	57.86%	56.40%	63.13%
Prueba:	57.93%	54.69%	57.93%	58.25%	58.25%	57.28%	55.34%

Fuente: Elaboración propia

4.2.2 Reporte de clasificación

Tabla 4.6 Reporte de clasificación

Dataset 1 (Var. Mundiales-Mexicanas)

Máquinas de soporte vectorial

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.56	0.23	0.33	57	188
1	0.51	0.81	0.63	45	198
Avg	0.54	0.52	0.48		

K-Nearest Neighbor

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.51	0.41	0.46	101	144
1	0.51	0.60	0.55	96	147
Avg	0.51	0.51	0.51		

Árboles de Decisión

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.53	0.60	0.57	148	97
1	0.54	0.47	0.50	130	113
Avg	0.54	0.54	0.54		

Bosques Aleatorios

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.61	0.36	0.45	89	156
1	0.54	0.76	0.63	58	185
Avg	0.58	0.56	0.54		

Regresión Logística

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.57	0.53	0.55	130	115
1	0.56	0.60	0.58	97	146
Avg	0.57	0.57	0.57		

Dataset 2 (Var. Mexicanas)

Máquinas de soporte vectorial

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.61	0.58	0.59	95	68
1	0.55	0.58	0.56	62	84
Avg	0.58	0.58	0.58		

K-Nearest Neighbor

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.56	0.67	0.61	109	54
1	0.53	0.41	0.46	86	60
Avg	0.55	0.54	0.54		

Árboles de Decisión

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.57	0.80	0.67	131	32
1	0.60	0.33	0.42	98	48
Avg	0.59	0.57	0.55		

Bosques Aleatorios

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.60	0.63	0.61	103	60
1	0.56	0.53	0.54	69	77
Avg	0.58	0.58	0.58		

Regresión Logística

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.59	0.68	0.63	111	52
1	0.57	0.47	0.52	77	69
Avg	0.58	0.58	0.58		

Naïve Bayes

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.55	0.42	0.48	103	142
1	0.53	0.66	0.59	83	160
Avg	0.54	0.54	0.54		

Naïve Bayes

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.59	0.61	0.60	100	63
1	0.55	0.53	0.54	69	77
Avg	0.57	0.57	0.57		

Red Neuronal MLP

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.52	0.60	0.56	148	97
1	0.52	0.43	0.47	138	105
Avg	0.52	0.52	0.52		

Red Neuronal MLP

Class	Precision	Recall	F1-Score	Confusion Matrix:	
-1	0.59	0.53	0.55	86	77
1	0.52	0.58	0.55	61	85
Avg	0.56	0.56	0.55		

Fuente: Elaboración propia

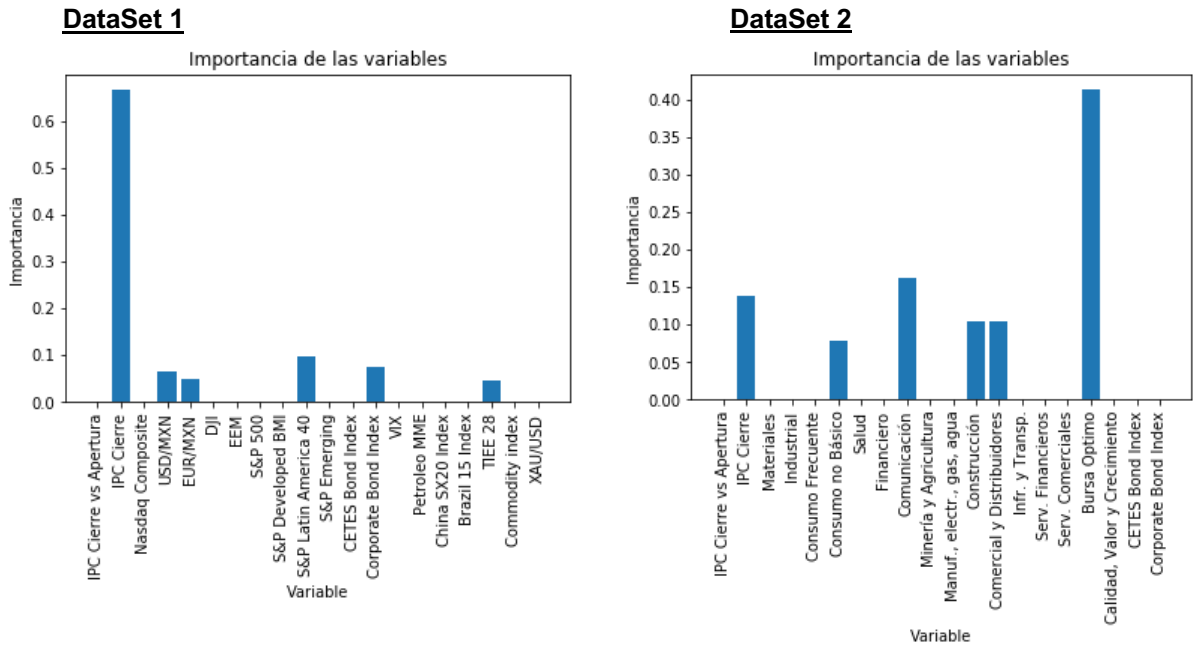
En los reportes de clasificación nuevamente notamos una mejora en los resultados obtenidos. Si bien, los algoritmos presentan desempeños similares, podemos destacar los resultados de las maquinas de soporte vectorial aplicadas en el dataset 2 (58%), los Arboles de decisión en el dataset 2 (59%) y los bosques aleatorios y regresión logística en ambas bases de datos (57/58%).

4.2.3 Importancia de las variables

Árboles de Decisión

Con respecto a la importancia que el modelo da a cada una de las variables observamos que para el dataset 1 el precio de cierre del IPC de los días anteriores representa la mayor proporción. Para el caso del dataset 2 el modelo otorgó una mayor importancia en la generación de su pronóstico al índice Bursa Óptimo que mide el rendimiento de las acciones de tomando en cuenta variables fundamentales como los ingresos netos, ganancias operativas, margen de ganancia, etc.

Figura 4.9 Importancia de las variables – Arboles de decisión

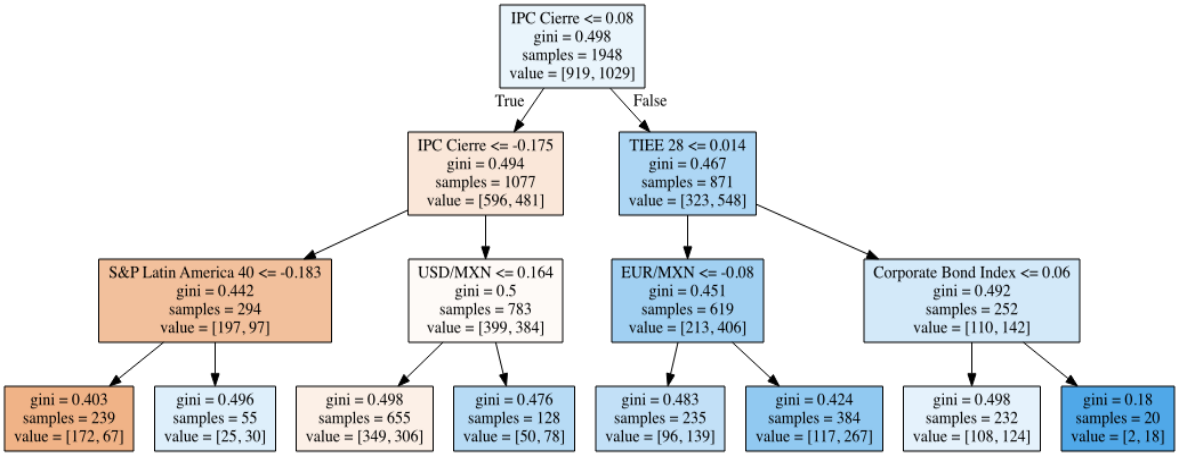


Fuente: Elaboración propia

Otra de las características de los algoritmos de arboles de decisión y bosques aleatorias es la posibilidad de visualizar los arboles generados por los modelos. A continuación, se muestran los arboles generados para cada uno de los datasets.

DataSet 1

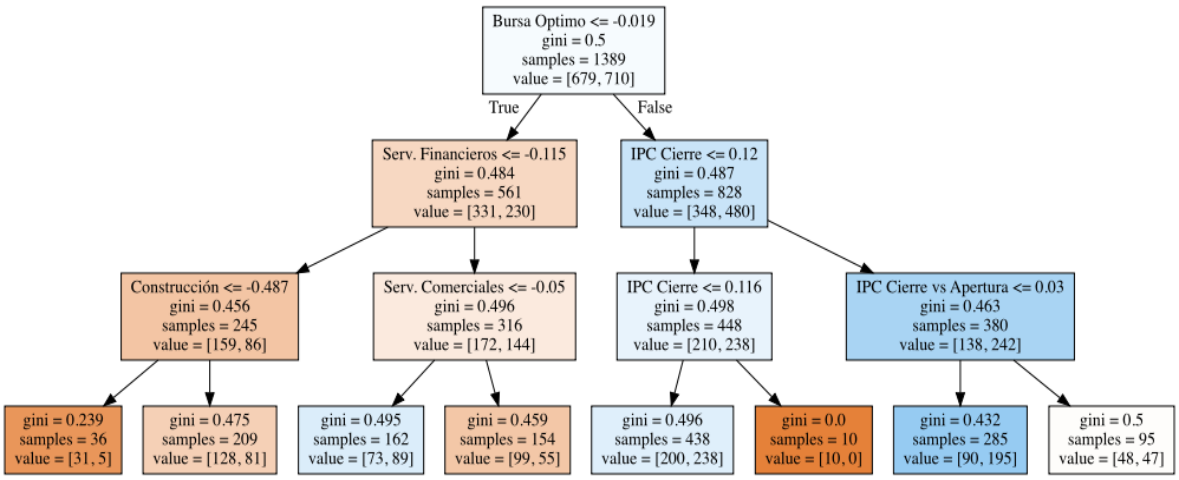
Figura 4.10 Árbol de decisión de dataset 1



Fuente: Elaboración propia

DataSet 2

Figura 4.11 Árbol de decisión de dataset 2



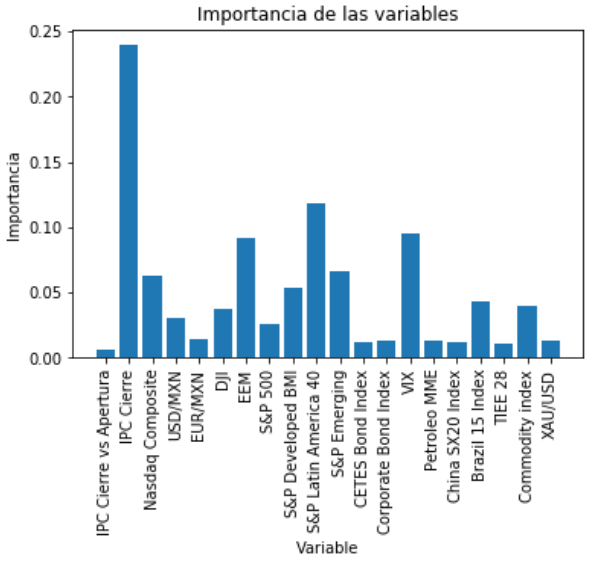
Fuente: Elaboración propia

Bosques Aleatorios

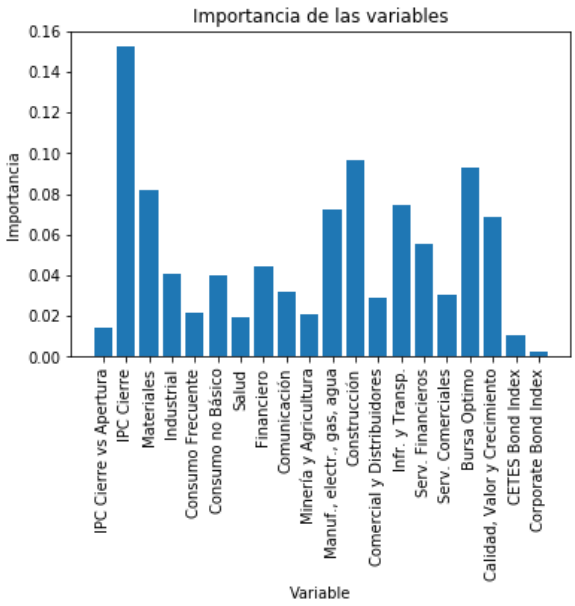
Los bosques aleatorios obtuvieron una mayor contribución para la creación de sus pronósticos del precio de cierre del IPC de los días anteriores tanto para el dataset 1 como para el dataset 2.

Figura 4.12 Importancia de las variables – Arboles de decisión

DataSet 1



DataSet 2

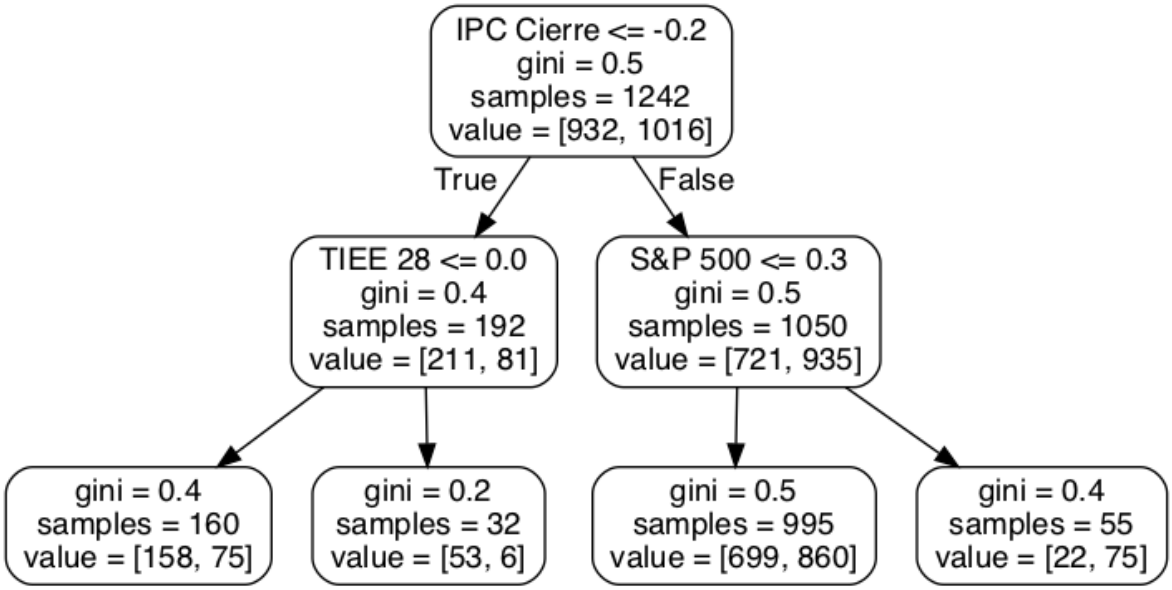


Fuente: Elaboración propia

Con respecto a los parámetros empleados en los algoritmos de bosques aleatorios se crearon 100 arboles de decisión para cada uno de los modelos, a continuación, se muestran algunos ejemplos de los arboles creados por los algoritmos para cada una de las bases de datos con las que fueron probados.

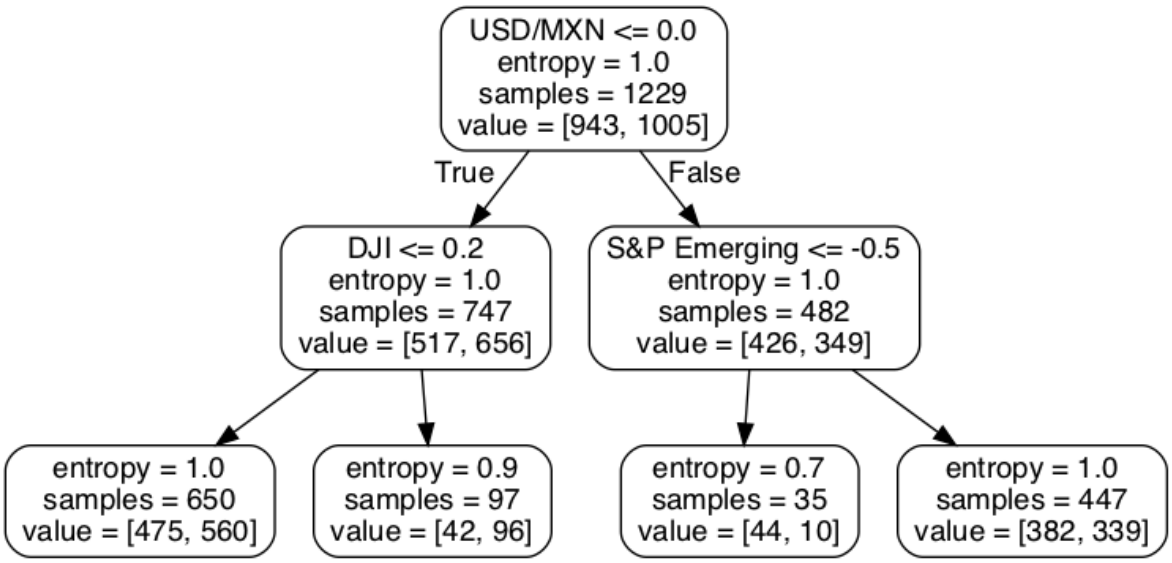
DataSet 1

Figura 4.13 Ejemplo 1. Árbol de decisión en dataset 1



Fuente: Elaboración propia

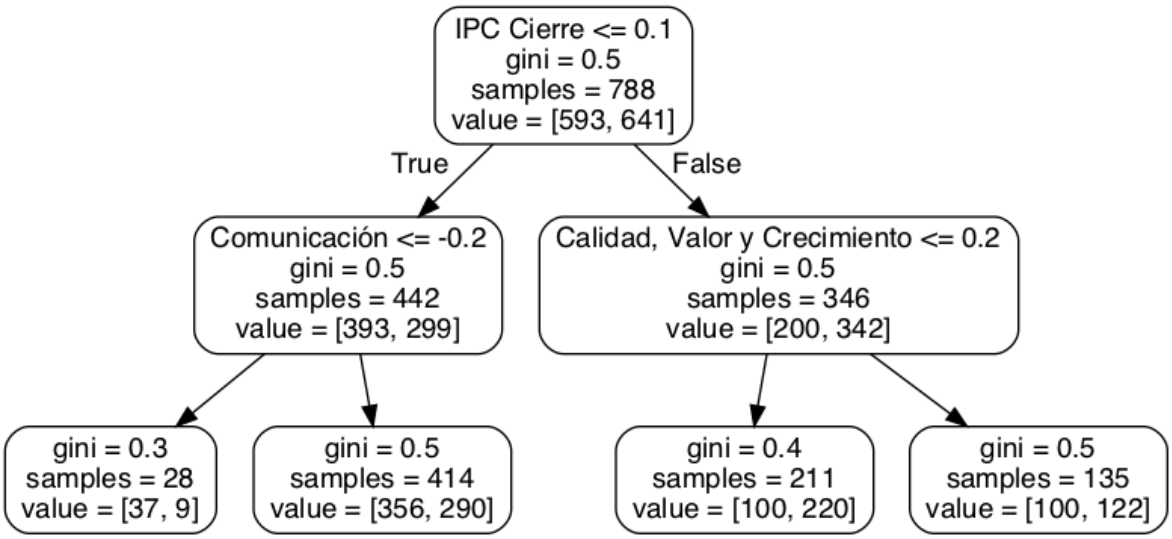
Figura 4.14 Ejemplo 2. Árbol de decisión en dataset 1



Fuente: Elaboración propia

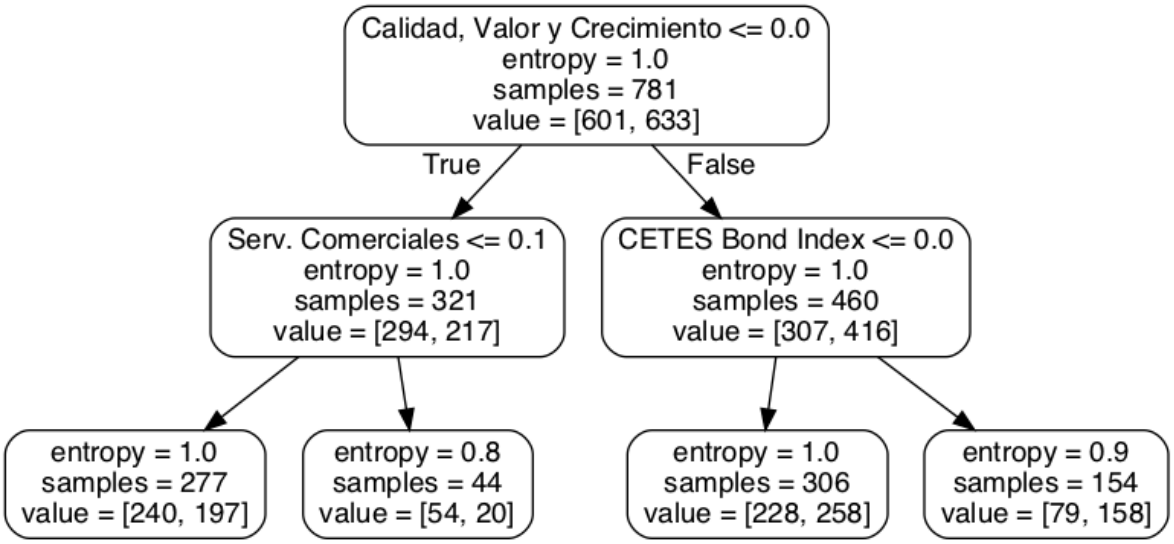
DataSet 2

Figura 4.15 Ejemplo 1. Árbol de decisión en dataset 2



Fuente: Elaboración propia

Figura 4.16 Ejemplo 2. Árbol de decisión en dataset 2



Fuente: Elaboración propia

4.2.4 K-Fold Cross Validation

Tabla 4.7 Resultados K-fold cross validation

Dataset 1 (Var. Mundiales-Mexicanas)

Fold No.	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles de Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
1	61.88	58.19	60.24	59.42	60.65	60.24	65.16
2	61.9	57.8	57.0	61.9	58.2	63.9	57.0
3	52.9	50.0	52.9	54.1	55.7	49.2	54.1
4	58.6	54.5	57.0	60.2	60.2	58.6	52.9
5	54.5	52.0	56.6	54.1	57.8	51.6	56.6
6	54.1	54.1	58.6	54.1	56.1	54.5	54.5
7	58.0	56.0	56.0	56.0	58.8	57.2	49.8
8	56.4	52.3	47.7	54.7	57.2	54.7	52.7
9	53.9	46.5	55.6	53.1	58.4	52.7	53.1
10	53.9	51.4	52.7	58.8	53.9	55.1	52.7
Mean	56.6	53.3	55.4	56.6	57.7	55.8	55
Std. Dev	3.2	3.4	3.3	2.9	2.0	4.1	3.9

Dataset 2 (Var. Mexicanas)

Fold No.	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles de Decisión	Bosques Aleatorios	Regresión Logística	Naïve Bayes	Red Neuronal MLP
1	60.64	56.13	60.0	61.29	60.0	56.12	56.77
2	60.0	56.1	57.4	60.6	61.9	59.4	60.0
3	55.5	54.2	60.0	56.8	56.1	56.8	48.4
4	55.2	53.9	56.5	57.1	53.2	57.8	51.9
5	56.5	48.1	58.4	56.5	59.1	58.4	50.6
6	55.8	56.5	56.5	55.8	55.8	54.5	57.1
7	58.4	57.1	58.4	58.4	58.4	58.4	63.6
8	53.9	48.1	51.9	53.9	55.8	54.5	54.5
9	58.4	52.6	57.1	59.1	57.1	58.4	54.5
10	59.1	57.1	57.8	59.1	58.4	58.4	58.4
Mean	57.4	54.0	57.4	57.9	57.6	57.3	56
Std. Dev	2.2	3.3	2.2	2.2	2.4	1.6	4.3

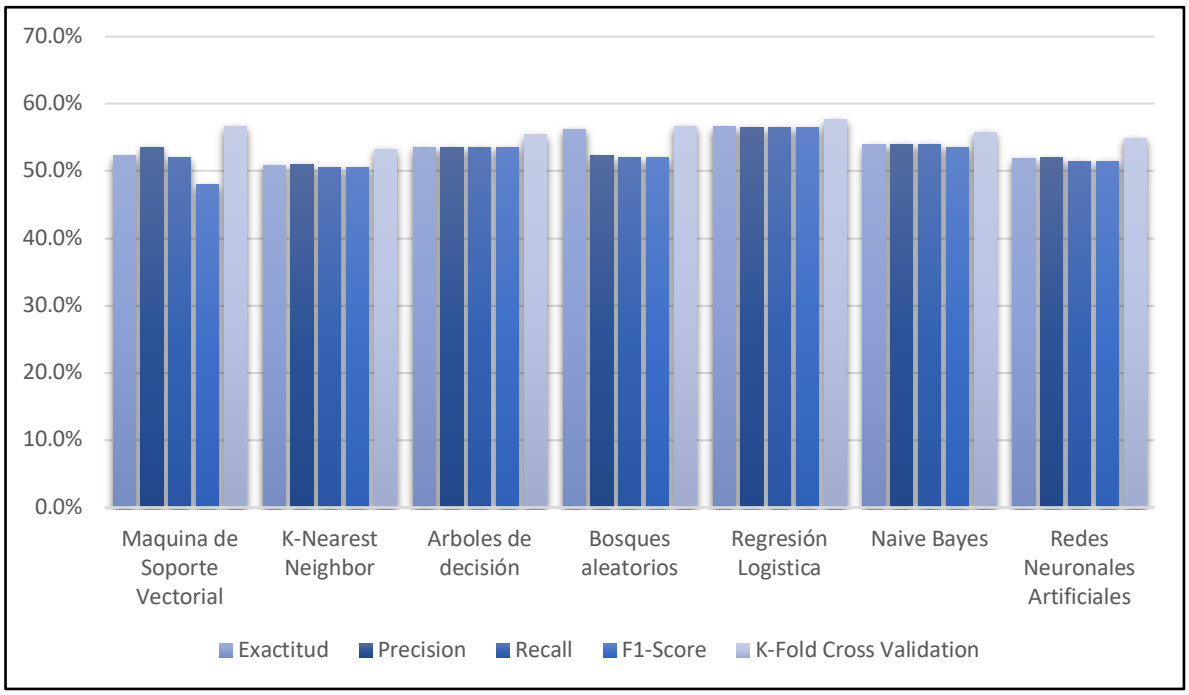
Fuente: Elaboración propia

Los resultados generados por la validación cruzada no generaron un aumento significativo en la exactitud de los modelos en comparación con la división de datos 80/20. Cuando los algoritmos fueron probados con el dataset 1 la Regresión logística obtuvo los mejores resultados (57.7%), mientras que con el dataset 2 los bosques aleatorios tienen el mejor desempeño (57.9%).

Las siguientes gráficas resumen los resultados obtenidos por los algoritmos en cada uno de los métodos de evaluación aplicados.

DataSet 1

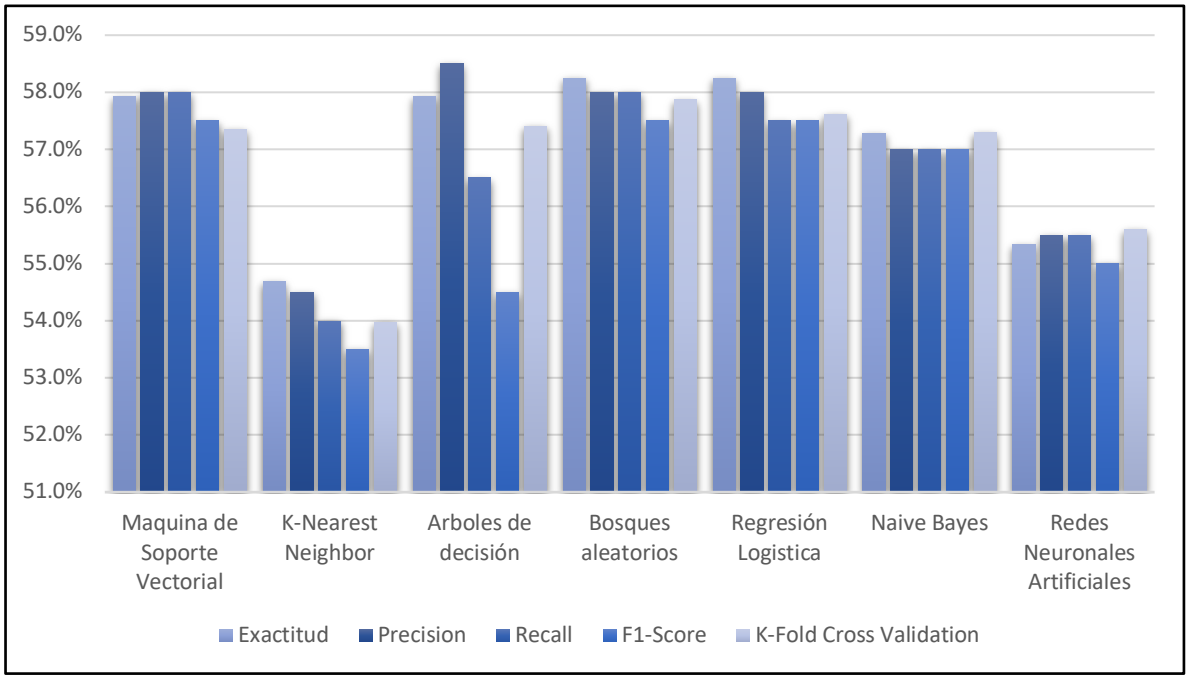
Figura 4.17 Comparativo de resultados - Dataset 1



Fuente: Elaboración propia

DataSet 2

Figura 4.18 Comparativo de resultados - Dataset 2



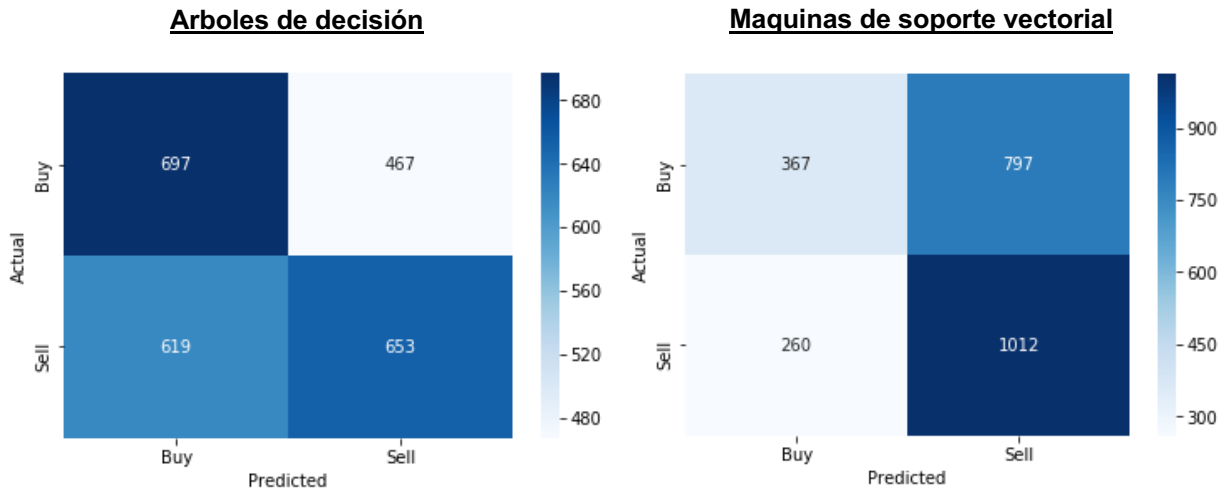
Fuente: Elaboración propia

En estas gráficas nuevamente notamos la mejora que tuvo el desempeño de los algoritmos al trabajar con datos suavizados. A diferencia de las pruebas realizadas sin la aplicación del filtro HP en su mayoría los algoritmos se acercan al 60% en las evaluaciones aplicadas.

4.2.5 Matrices de Confusión

DataSet 1

Figura 4.19 Matrices de confusión – Dataset 1

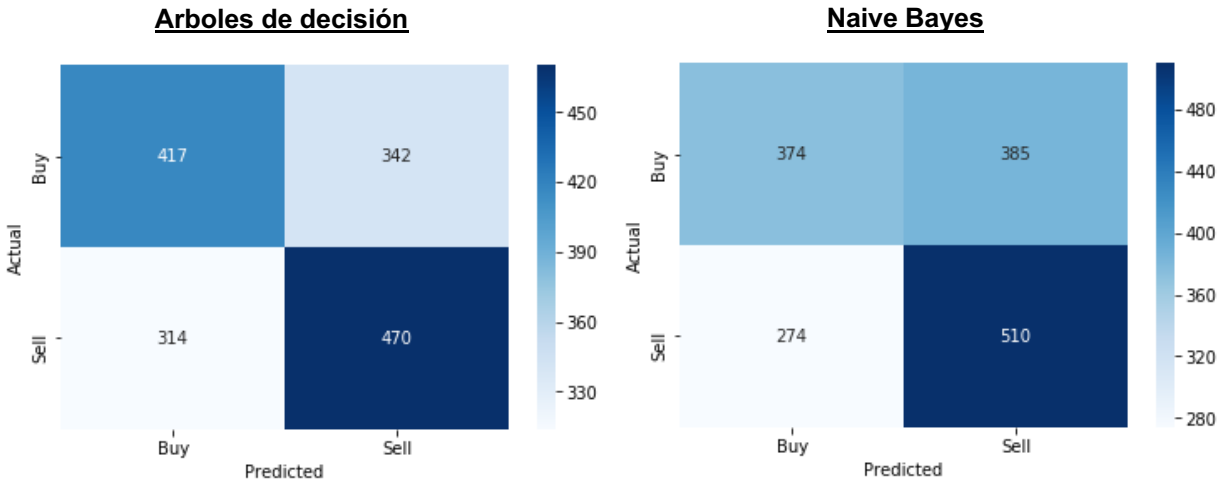


Fuente: Elaboración propia

En las pruebas realizadas en el DataSet 1 los arboles de decisión presentaron los mejores resultados en la predicción de las señales de compra al conseguir pronosticar 697 registros de manera correcta. Por otra parte, las maquinas de soporte vectorial se desempeñaron de mejor manera que el resto de los algoritmos al pronosticar las señales de venta con 1,012 registros correctos.

DataSet 2

Figura 4.20 Matrices de confusión – Dataset 2



Fuente: Elaboración propia

Para el caso del DataSet 2 el modelo de los Arboles de decisión arrojo los mejores resultados en la predicción al pronosticar 417 señales de compra de manera correcta y 342 de manera incorrecta. Con respecto a las señales de venta el algoritmo Naive Bayes es el mejor en pronosticar las señales de venta con 510 registros correctos y 274 incorrectos.

Capítulo 5 Backtesting – Simulación de una estrategia de inversión

Backtesting es una metodología utilizada para verificar y diagnosticar la eficiencia y los resultados de un modelo de trading. Esta metodología intenta determinar si las conclusiones del modelo son acertadas al realizar una “evaluación hacia atrás”, al aplicar los resultados de un modelo en periodos anteriores y compararlo con datos históricos.

En un backtesting se simula de forma exacta las reglas de entrada y salida impuestas por un modelo con el fin de probar una estrategia de trading antes de emplearla.

Con el propósito de evaluar de forma práctica el pronóstico generado por los algoritmos de Machine Learning a continuación se simuló una estrategia de inversión.

Dado que la precisión de esta metodología depende de las predicciones correctas de todos los movimientos del índice (hacia arriba o hacia abajo), realizar una evaluación a través de una estrategia de trading es una medida importante ya que las decisiones de compra / venta se basan en las predicciones de que el índice aumentará o disminuirá.

Aquí se propone como meta obtener una estrategia de compra/venta sencilla para actuar con base en los pronósticos. Los modelos predicen el cambio entre la apertura y los valores de cierre del día de negociación. Si se pronostica que el mercado aumentara por encima del umbral, esto indica una “compra”. Por otro lado, si se prevé que caiga, esto señala una “venta”.

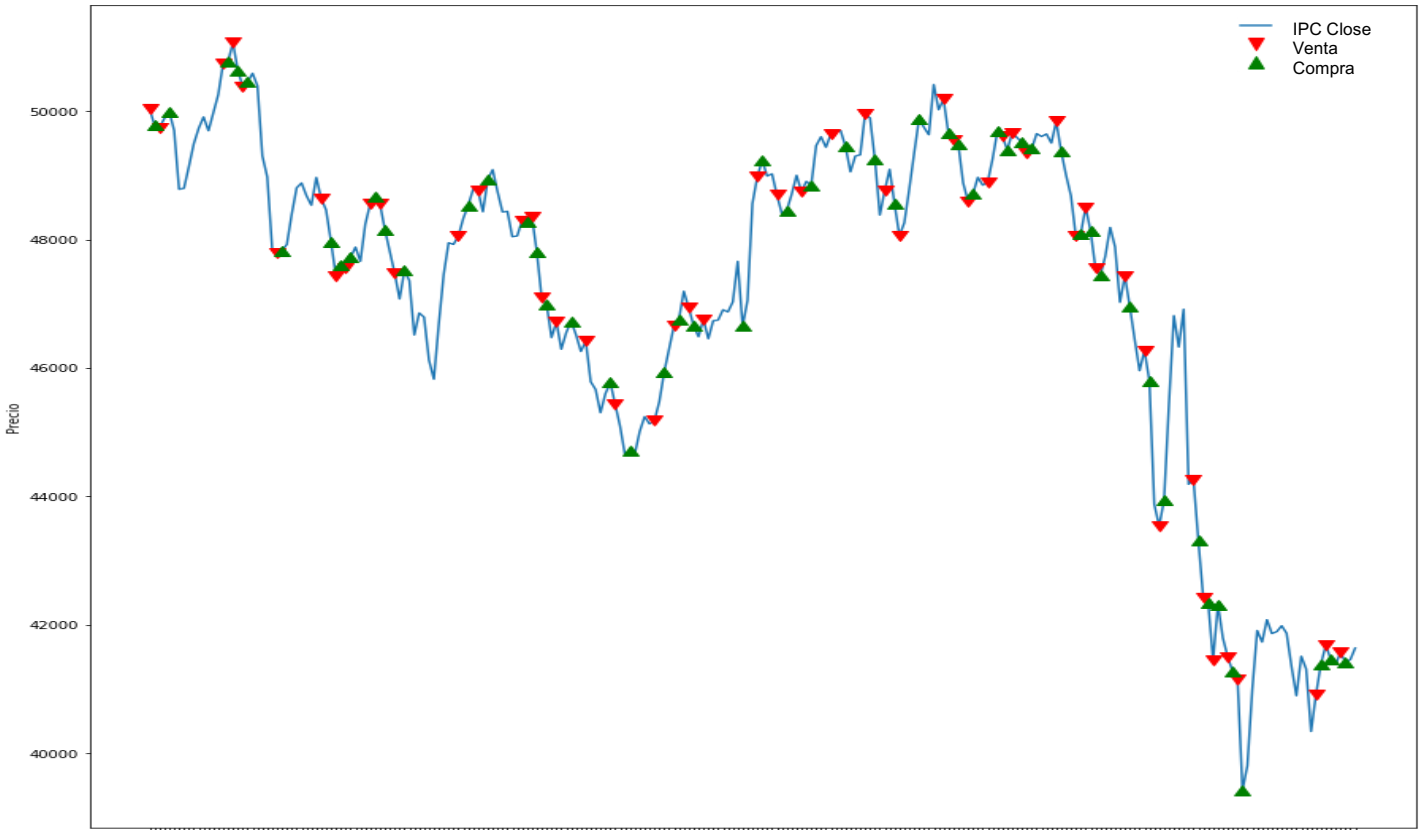
Todas las operaciones se realizan al comienzo del día de negociación al precio de apertura del IPC. El primer día que se recibe una señal de compra del modelo, se realiza una inversión en el IPC utilizando el monto total de capital. Las señales de compra en los días siguientes se tratan como instrucciones de espera hasta que se recibe una señal de venta. Todas las unidades de inversión que se mantienen actualmente se venden al precio actual después de una señal de venta, el capital se convierte nuevamente en efectivo. Las señales de venta de días posteriores se tratan nuevamente como instrucciones no comerciales, hasta que se recibe otra señal de compra y se repite el proceso.

5.1 Señales de compra venta

Con fines prácticos se creó una serie de gráficas para cada uno de los modelos, estas gráficas proporcionan una manera practica de visualizar las señales de compra y venta arrojadas por cada uno de los algoritmos el triangulo rojo indica una señal de venta, mientras el triangulo verde indica una señal de compra del instrumento.

La gráfica muestra un ejemplo de las señales, en este caso por la maquina de soporte vectorial entrenada con el DataSet 1 de indicadores globales.

Figura 5.1 Señales de Compra/Venta, algoritmo Maquina de Soporte Vectorial



Fuente: Elaboración propia

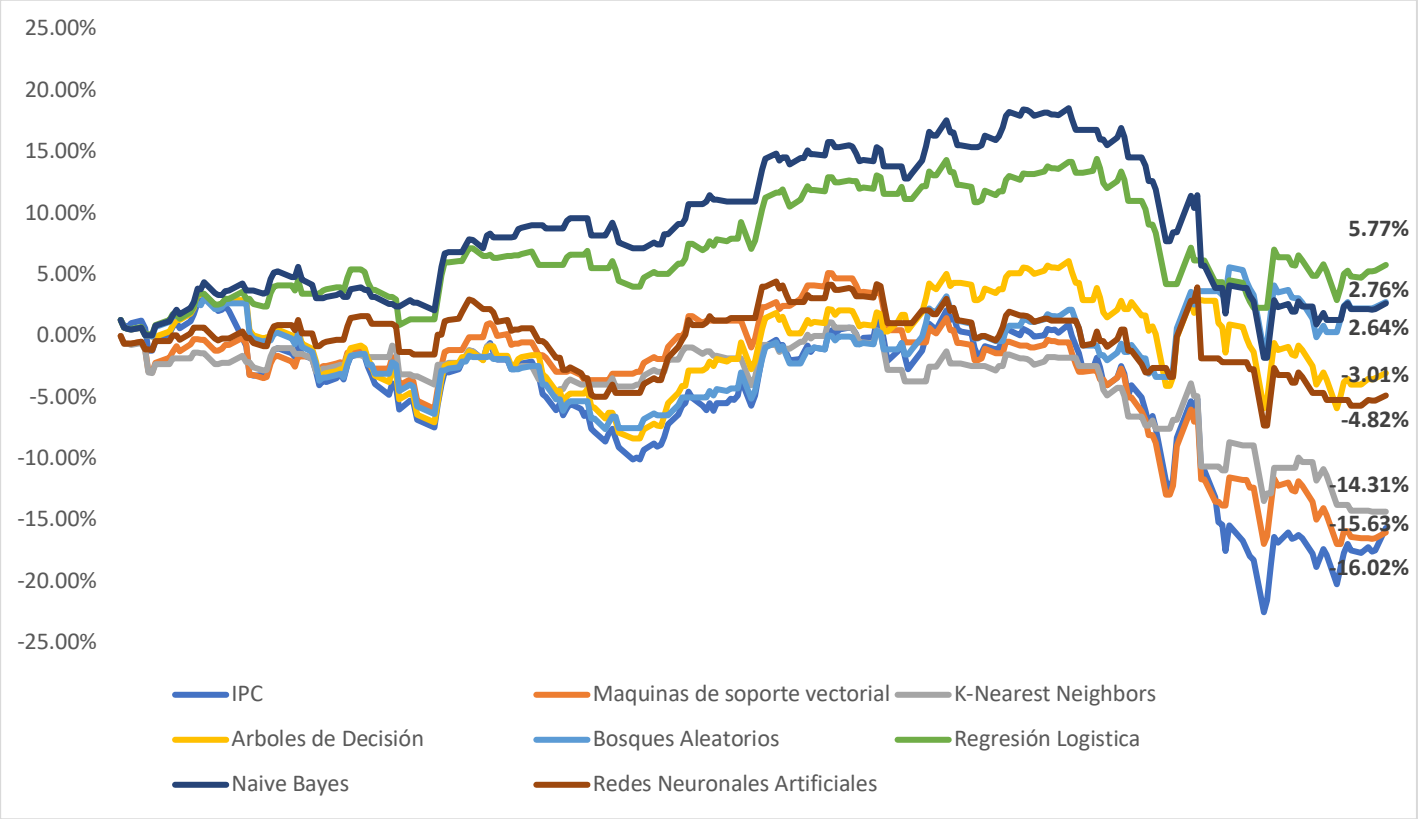
5.2 Rendimientos

Las siguientes gráficas contrastan el desempeño de los rendimientos conseguidos por medio de una estrategia basada en las señales pronosticadas por cada modelo. En este caso los rendimientos de cada modelo se comparan con los rendimientos obtenidos en una estrategia pasiva en la cual se genera una compra al inicio del periodo (2 de enero de 2018) y una venta al final de este (31 de diciembre del 2018).

Esta estrategia pasiva en el IPC generó un rendimiento de -15.63%, por lo que este es el parámetro principal con el que mediremos el desempeño de todos los modelos.

DataSet 1

Figura 5.2 Estrategia de inversión – DataSet 1



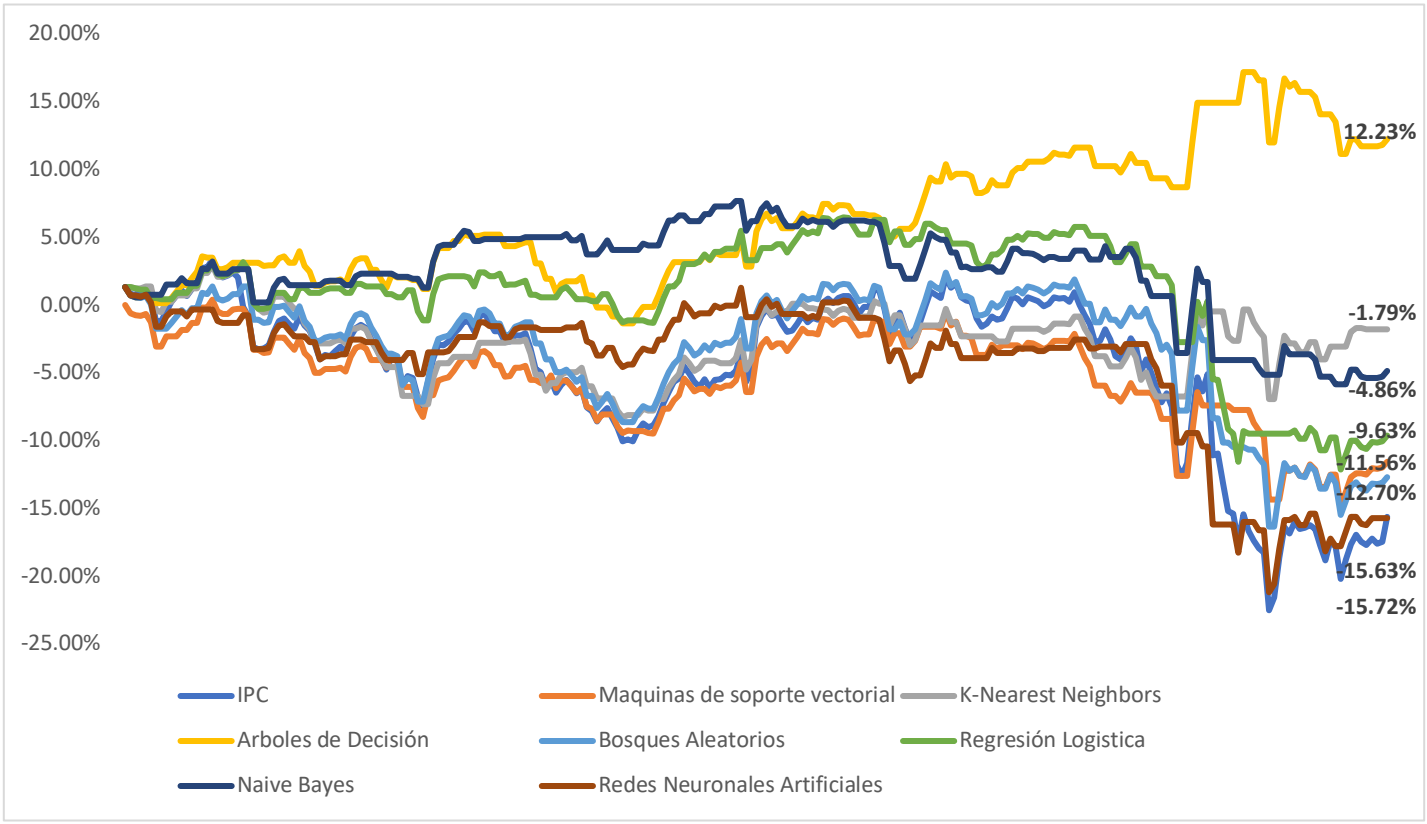
Fuente: Elaboración propia

Como resultado de realizar una estrategia de compra y venta con base en las señales pronosticadas por cada uno de los modelos obtenemos que la regresión logística consigue el mejor rendimiento (5.77%), le siguen los modelos de bosques aleatorios y Naive Bayes

con 2.76% y 2.64% respectivamente. El peor rendimiento se obtiene con la aplicación de las maquinas de soporte vectorial. Si bien el resto de los modelos presentan rendimientos negativos al igual que el IPC estos rendimientos son mejores que los rendimientos conseguidos a través de una estrategia pasiva en el índice.

DataSet 2

Figura 5.3 Estrategia de inversión – DataSet 2

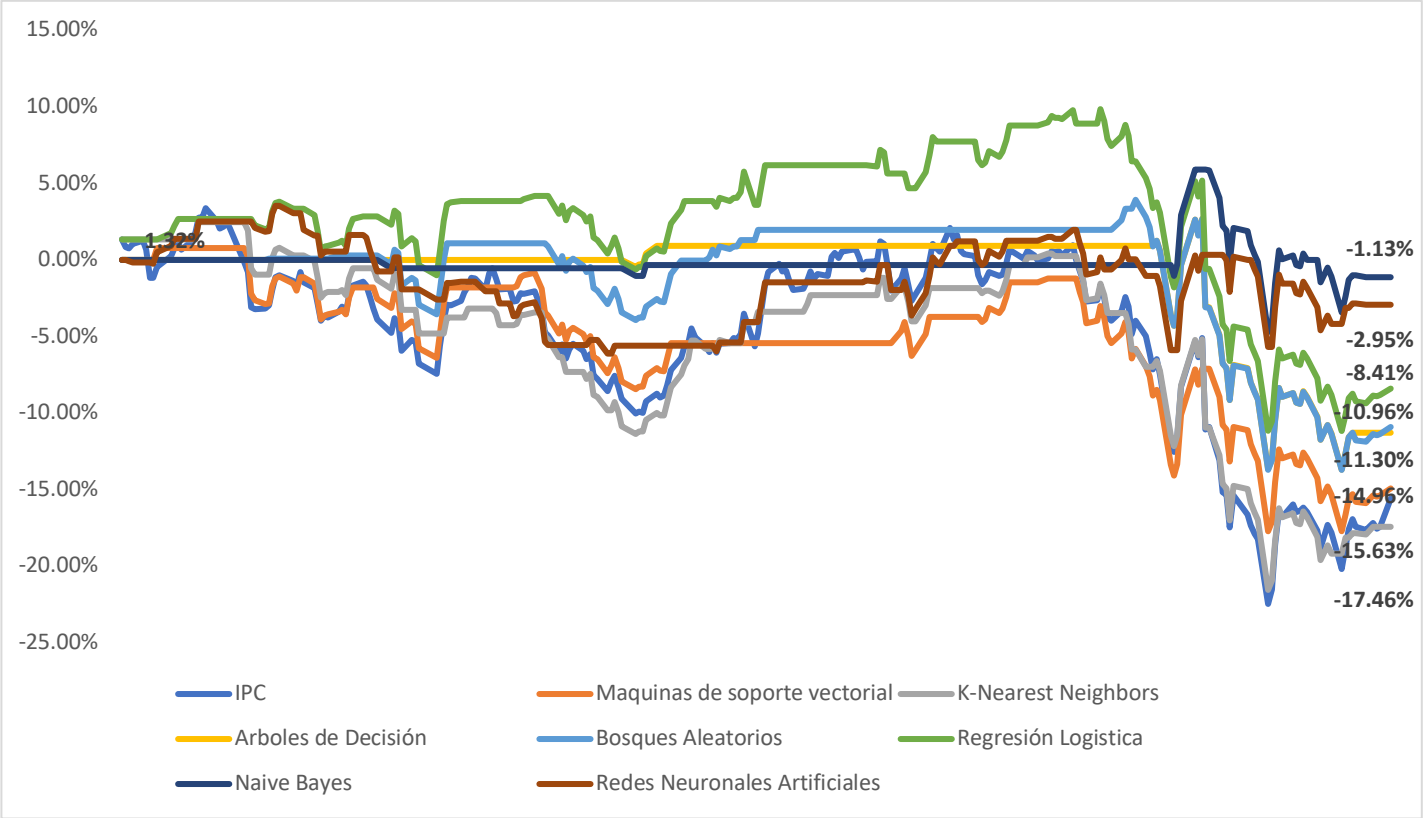


Fuente: Elaboración propia

Los modelos probados en el DataSet 2 compuesto por indicadores del mercado mexicano generaron rendimientos negativos con excepción de los Arboles de decisión los cuales obtuvieron un rendimiento sobresaliente del 12.23%. Las redes neuronales artificiales consiguieron el peor rendimiento (-15.72%). A pesar de que el resto de los modelos obtuvo rendimientos negativos en su mayoría estos resultados aún superan al rendimiento conseguido en la estrategia pasiva (-15.63%).

DataSet 3

Figura 5.4 Estrategia de inversión – DataSet 3



Fuente: Elaboración propia

Finalmente, la aplicación de los modelos en el DataSet 3 conformado por indicadores de análisis técnicos del IPC proporciona los resultados más bajos, ya que ninguno de estos consiguió rendimientos positivos, sin embargo, todos los modelos con excepción de K-Nearest Neighbor fueron superiores al -15.63% de la estrategia pasiva del IPC.

La siguiente tabla muestra un concentrado con los rendimientos conseguidos por cada estrategia.

Tabla 5.1 Resumen de rendimientos

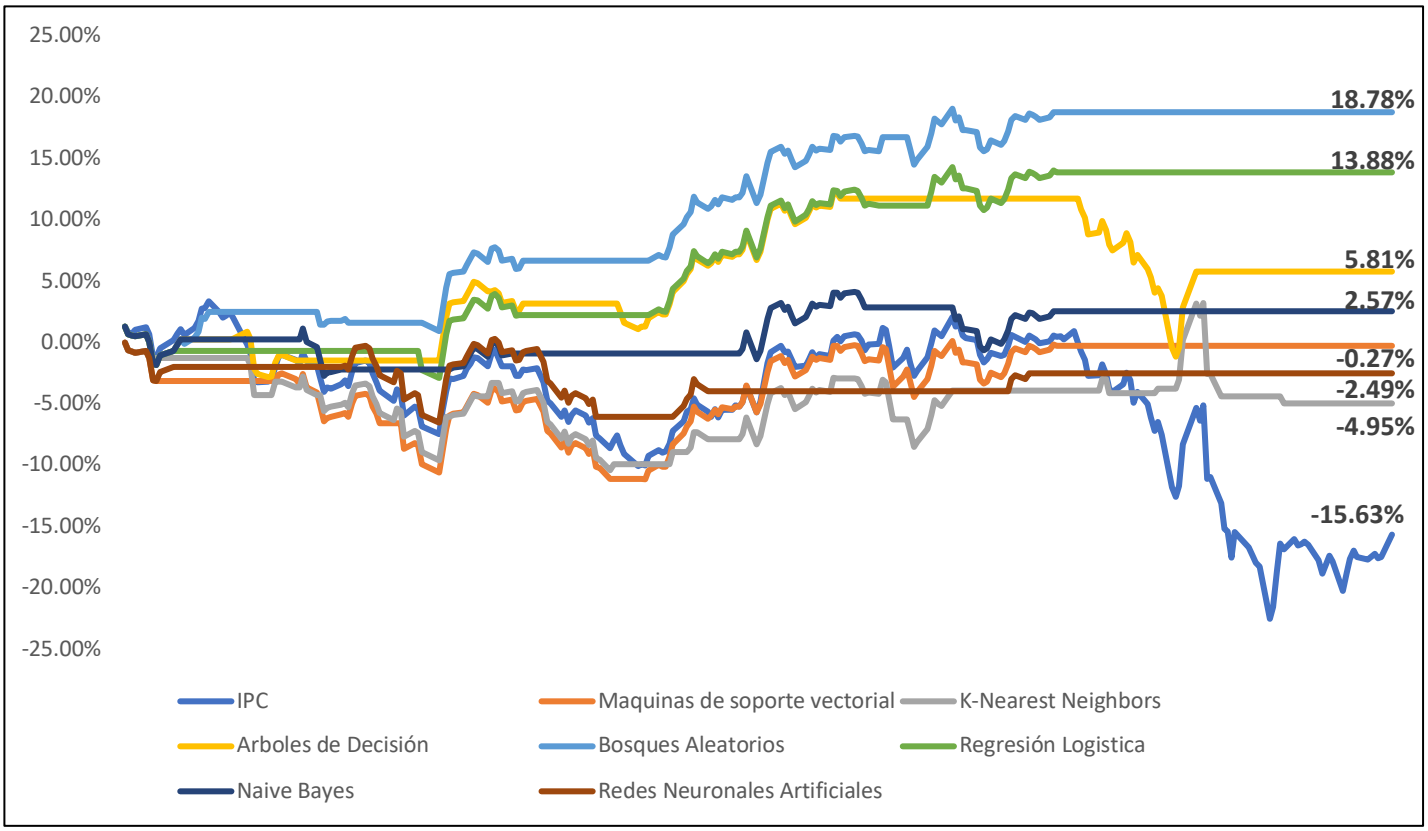
Base de Datos	IPC	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naive Bayes	Red Neuronal
Dataset 1	-15.63%	-16.02%	-14.31%	-3.01%	2.76%	5.77%	2.64%	-4.82%
Dataset 2	-15.63%	-11.56%	-1.79%	12.23%	-12.70%	-9.63%	-4.86%	-15.72%
Dataset 3	-15.63%	-14.96%	-17.46%	-11.30%	-10.96%	-8.41%	-1.13%	-2.95%

Fuente: Elaboración propia

Las siguientes gráficas muestran los resultados obtenidos cuando se aplicó el filtro de suavizamiento Hodrick-Prescott a los conjuntos de datos previo a su aplicación en los algoritmos.

DataSet 1 con Filtro HP

Figura 5.5 Estrategia de inversión con filtro HP – DataSet 1

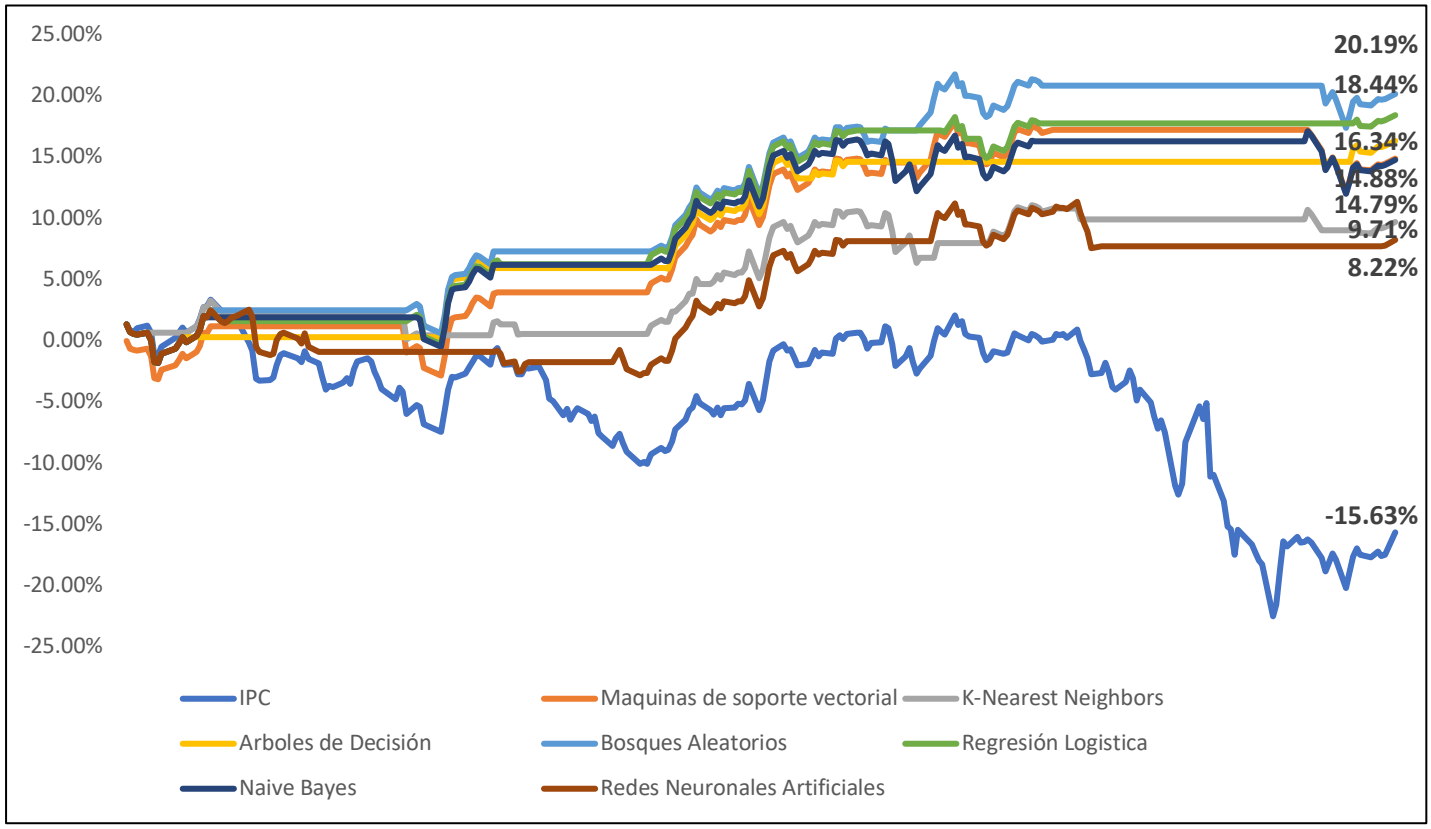


Fuente: Elaboración propia

En esta gráfica podemos observar que el desempeño de los algoritmos mejora considerablemente cuando fue entrenado con datos suavizados. Los bosques aleatorios, regresión logística y arboles de decisión obtienen los mejores resultados. Por otra parte las maquinas de soporte vectorial, redes neuronales artificiales y K-Nearest Neighbor son los únicos en obtener resultados negativos.

DataSet 2 con Filtro HP

Figura 5.6 Estrategia de inversión con filtro HP – DataSet 2



Fuente: Elaboración propia

Finalmente, en la última gráfica, donde los modelos fueron entrenados con el dataset 2 compuestos por indicadores del mercado de valores mexicano y suavizados con el filtro Hodrick-Prescott obtuvieron los mejores resultados de todas las pruebas realizadas. Una vez mas los bosques aleatorios, regresión logística y arboles de decisión tuvieron los mejores resultados.

De manera adicional en estas dos últimas gráficas podemos notar que los modelos funcionaron de manera sobresaliente al predecir la fuerte caída que se dio en el mercado de valores mexicano a partir de octubre del 2018. Entre octubre y diciembre los modelos mantuvieron señales pasivas y de esta manera evitaron importantes caídas en sus rendimientos.

La siguiente tabla resume los rendimientos obtenidos en ambas pruebas.

Tabla 5.2 Resumen de rendimientos con filtro HP

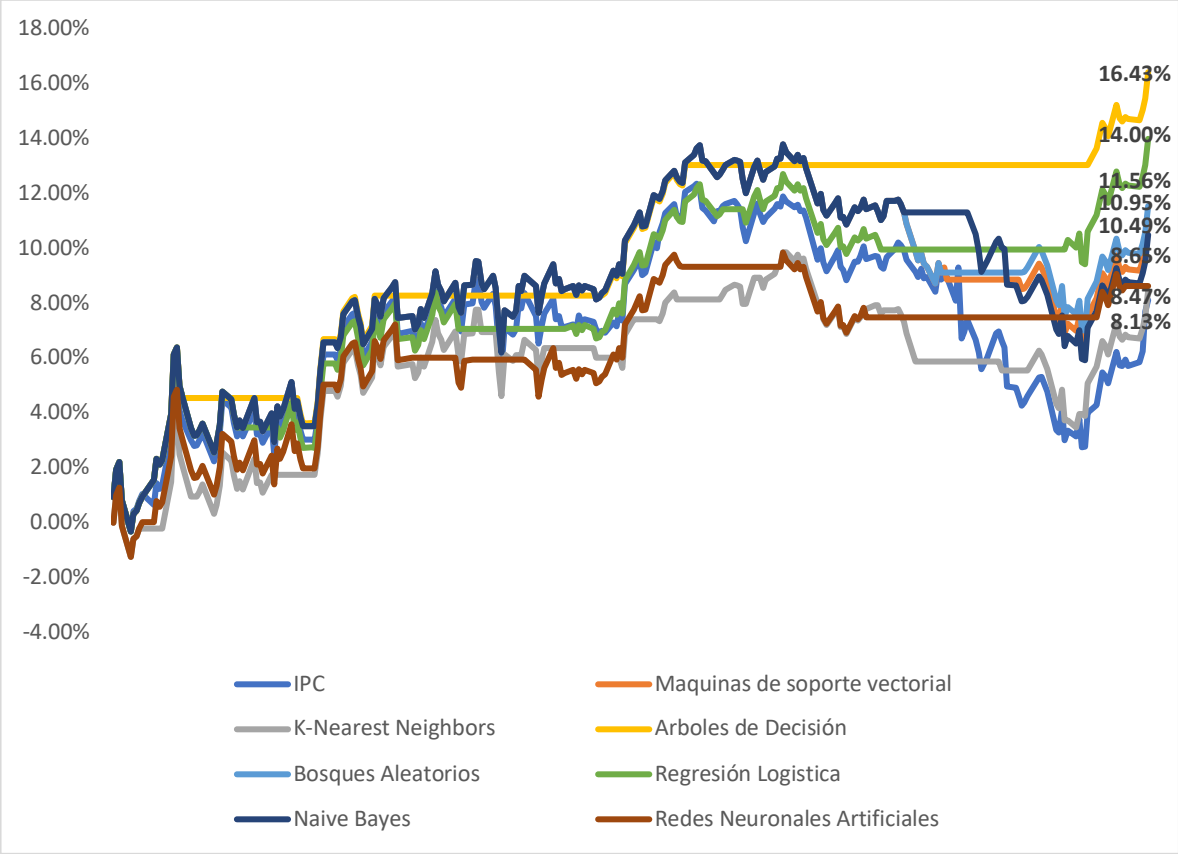
Base de Datos	IPC	Máquinas Soporte Vectorial	K-Nearest Neighbor	Árboles Decisión	Bosques Aleatorios	Regresión Logística	Naive Bayes	Red Neuronal
Dataset 1	-15.63%	-0.27%	-4.95%	5.81%	18.78%	13.88%	2.57%	-2.49%
Dataset 2	-15.63%	14.88%	9.71%	16.34%	20.19%	18.44%	14.79%	8.22%

Fuente: Elaboración propia

Como prueba adicional se realiza una evaluación de estos modelos aplicados al año 2017. En 2017 una estrategia pasiva de compra y venta en el IPC hubiera otorgado un rendimiento de 8.13 %. Las siguientes gráficas muestran los rendimientos obtenidos aplicando las señales creadas por los algoritmos.

DataSet 1 con Filtro HP aplicado en 2017

Figura 5.7 Estrategia de inversión con filtro HP – DataSet 1, 2017

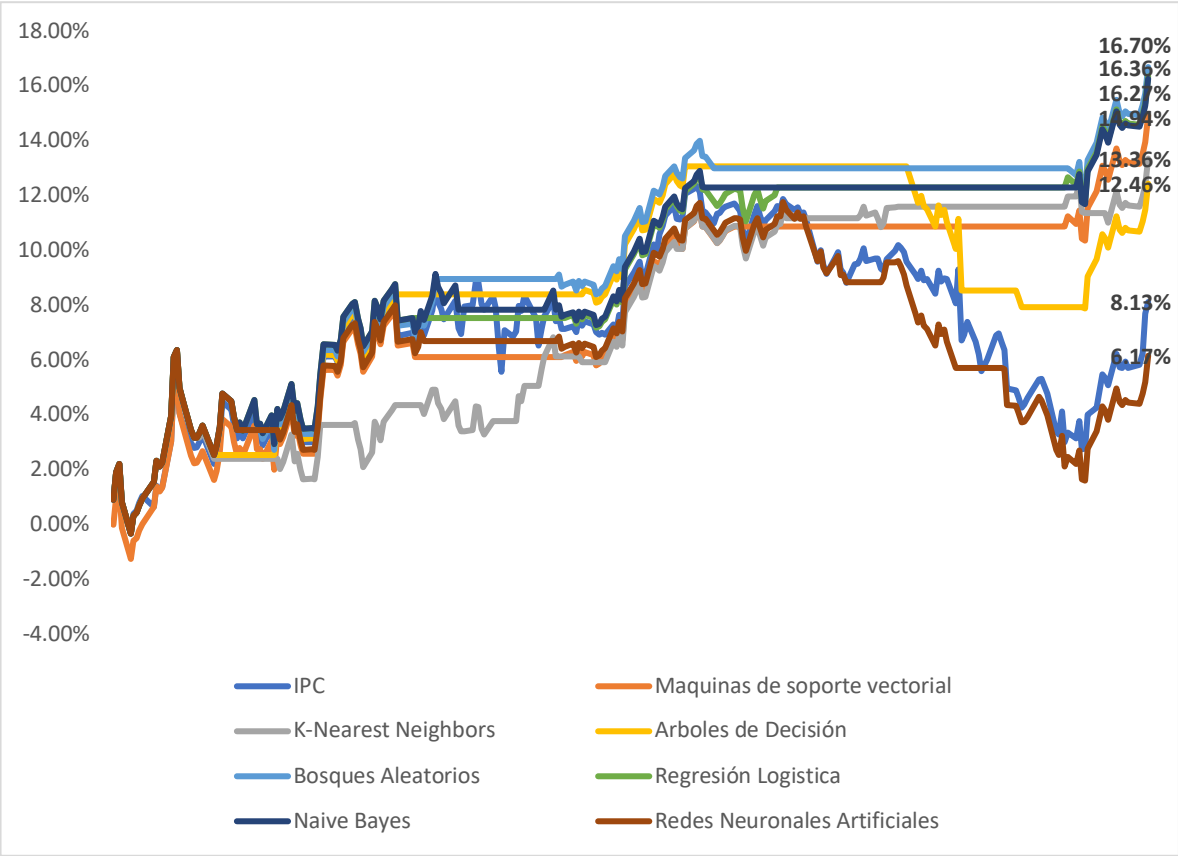


Fuente: Elaboración propia

Empleando el dataset 1 los arboles de decisión consiguen los mejores resultados (16.43%), seguidos de la regresión logística (14%). En esta prueba todos los algoritmos superan la estrategia pasiva de compra y venta del IPC.

DataSet 2 con Filtro HP aplicado en 2017

Figura 5.8 Estrategia de inversión con filtro HP – DataSet 2, 2017



Fuente: Elaboración propia

Con datos del dataset 2 todos los algoritmos con excepción de las redes neuronales artificiales superaron el rendimiento obtenido por el IPC. Los mejores resultados los obtiene el algoritmo de bosques aleatorios con 16.70%, regresión logística con 16.36 y Naive Bayes con 16.27%.

Tabla 5.3 Resumen de rendimientos con filtro HP, 2017

Base de Datos	IPC	Máquinas soporte vectorial	K-Vecinos más cercanos	Árboles de Decisión	Bosques Aleatorios	Regresión Logística	Naive Bayes	Red Neuronal MLP
Dataset 1	8.13%	10.95%	8.47%	16.43%	11.56%	14.00%	10.49%	8.65%
Dataset 2	8.13%	14.94%	13.36%	12.46%	16.70%	16.36%	16.27%	6.17%

Fuente: Elaboración propia

Capítulo 6 Conclusiones

Esta investigación tuvo como objetivo evaluar y determinar si es probable o no emplear algoritmos de clasificación de Machine Learning en la predicción de los movimientos del precio del IPC con la finalidad de crear una herramienta que funcione como apoyo en estrategias de compra y venta de acciones.

Como bien sabemos predecir el precio de las acciones es muy difícil. Una serie de investigaciones históricas y modernas insisten en que el mercado de valores se rige por una caminata aleatoria. Algunos dicen que es imposible hacer una predicción significativa de las acciones, y que la mejor estrategia de inversión es comprar acciones al azar. Otros investigadores afirman que, aunque los cambios de precios pueden no ser estrictamente aleatorios, la interdependencia es tan leve que es imposible obtener un beneficio. Es por eso por lo que obtener una conclusión definitiva con respecto a pronosticar el mercado bursátil parece casi imposible. Simplemente hay demasiados factores desconocidos que pueden influir en el mercado.

Por un lado, los resultados arrojados muestran que hay algoritmos de Machine Learning que aplicados en una estrategia de trading pueden generar buenos rendimientos y que pueden optimizarse en una medida mucho mayor, lo que probablemente aumentaría el beneficio producido. Esto demuestra que hay una posibilidad en la predicción de acciones con Machine Learning. Sin embargo, uno de los problemas subyacentes con la predicción del mercado de valores empleando estas técnicas es saber si el período de prueba tomado es representativo para cualquier período de tiempo futuro.

Las diferentes configuraciones en cuanto a periodo y cantidad de datos en las que fueron aplicados los modelos no representaron diferencias significativas en sus resultados. De igual manera el tipo de indicadores empleados en cada unas de las tres bases creadas no proporcionaron ninguna evidencia de que el rendimiento de los algoritmos mejorara.

Los mejores resultados se consiguieron una vez que los datos fueron preprocesados con el Filtro Hodrick-Prescot, lo que permitió disminuir el ruido presente en las series de tiempo financieras y obtener datos suavizados. En estas pruebas destacaron los algoritmos de Arboles de decisión, Bosques aleatorios y Regresión logística, al obtener los mejores

resultados en las pruebas de exactitud y clasificación, de igual manera obtuvieron los mejores rendimientos en el caso de simulación. Por otra parte los algoritmos K-Nearest Neighbor y las Maquinas de soporte vectorial presentaron los peores resultados en la mayoría de las pruebas realizadas.

Como pudimos observar en la simulación de inversión, la aplicación de las señales generadas por los modelos en su mayoría supera los rendimientos conseguidos en una estrategia pasiva de inversión. En esta simulación también podemos observar que los modelos se desempeñan mejor en periodos de mercado bajista, como lo fue el mercado mexicano a finales del año 2018 y en el que los modelos pronosticaron de manera correcta esta caída, permitiendo minimizar las perdidas en el rendimiento.

La exactitud de los modelos se ve reducida (50-57% en promedio) si lo comparamos con modelos desarrollados en investigaciones similares aplicadas a mercados como Estados Unidos o Europa, donde obtienen probabilidades de acertar cercanas al 70%. Esto se puede atribuir a que los modelos se desempeñan mejor con información derivada de economías estables (países desarrollados), donde existe una volatilidad menor en los mercados, por el contrario, países emergentes como México cuentan con un mercado volátil y son mas sensibles a la incertidumbre económica y política presente en los mismos. Un caso de estudio interesante seria la aplicación de estos modelos en economías desarrolladas de Europa o Estados Unidos empleando información creada a partir de la llegada de Donald Trump a la presidencia de Estados Unidos, donde sus políticas han generado constantes movimientos económicos y financieros a nivel global.

Artículos recientes con las investigaciones de Singh D. (2018) donde aplica redes neuronales artificiales de clasificación, Pisa M. (2018) con Arboles de Decisión, Singh V. (2018) con K Nearest Neighbor y Regresión Logística y Tahsildar S. (2018) empleando Bosques Aleatorios, todas aplicadas al pronostico del S&P 500 brindan resultados similares a este proyecto, donde obtienen una precisión entre el 52% y 55%, pero de igual manera generando rendimientos mayores a los de una estrategia pasiva.

Finalmente los resultados de este trabajo nos muestran que las estrategias de inversión empleando las señales generadas por los algoritmos de Machine Learning superan al índice de referencia (IPC). Sin embargo, los porcentajes de exactitud resultantes de los modelos nos impiden llegar a una conclusión indiscutible sobre el uso de estas estrategias de

inversión aplicadas al mercado financiero mexicano. El uso real de estos modelos requeriría de la creación de un portafolio de inversión que nos permita controlar el riesgo así como la incorporación de diferentes herramientas de análisis técnico y fundamental que permitan llevar a cabo una mejor estrategia en la compra y venta de las acciones.

Por otra parte podemos recalcar las ventajas que la metodología y técnicas de Machine Learning brindaron durante su aplicación en este proyecto de investigación, ya que permiten un manejo eficiente de recursos al permitir emplear grandes volúmenes de datos, tener tiempos reducidos de procesamiento, así como la facilidad para automatizar el trabajo e incluso integrarlo en diferentes plataformas que permiten un acceso a cualquier usuario sin la necesidad de tener conocimientos avanzados de programación o estadística.

Como áreas de oportunidad para trabajos futuros se propone la elaboración de una aplicación web que permita realizar el análisis de datos y la generación de los pronósticos diarios de una manera mas sencilla y accesible, actualmente plataformas como Azure Machine Learning de Microsoft permite integrar sus modelos con Microsoft Excel, lo que permite aumentar el alcance de estas técnicas en un ambiente en el que no se requieren grandes conocimientos de programación.

Otra propuesta consiste en realizar los experimentos empleando diferentes configuraciones y tipos de datos económicos y financieros. En casos donde se busque pronosticar el precio de las acciones de una empresa en particular buscaríamos adaptar las variables de entrada con variables que pudieran afectar directamente el precio de esta, un ejemplo sería incorporar variables económicas de turismo si estamos tratando de pronosticar el precio de una empresa de ese ramo, o incorporar precios de minerales en caso de aplicar los modelos en compañías mineras.

Referencias

- Abu-Mostafa, Y., Atiya, A. (1996). Introduction to financial forecasting. Applied Intelligence, (6), 205–213.

- Barry J. (2010). Algorithmic Trading & DMA. Londres: 4Myeloma Press

- BMV (Bolsa Mexicana de Valores). www.bmv.com.mx

- Bontempi G., Ben S., Le Borgne Y. (2013). Machine Learning strategies for time series forecasting. Universidad de Bruselas, Bélgica.

- Brooks R., Dahlke K. (2017). ¿Artificial Intelligence vs. Machine Learning vs. Data Mining 101 – What's the Big Difference? Recuperado de https://www.guavus.com/artificial-intelligence-vs-Machine-Learning-vs-data-mining-101-whats-big-difference/

- Daza A. (2016). Data Mining, Minería de datos. Perú: Editorial Macro.

- Dhakshayani N. (2016). What's the relationship between Machine Learning and data mining? Quora. Recuperado de https://www.quora.com/What-is-the-difference-between-data-mining-artificial-intelligence-and-Machine-Learning

- Diaz A., Aguilera V. (2005). Introducción al mercado bursátil. México: McGraw-Hill

- Dongles N. (2018). The Random Forest Algorithm. Towards Data Science. Recuperado de https://towardsdatascience.com/the-random-forest-algorithm-d457d499ffcd/

- Fabozzi, F., Modigliani F., Ferri M. (1996). Mercados e instituciones financieras. México: Prentice Hall.

- Fernandez F., Gonzalez C., Sosvilla S. (2000). On the profitability of technical trading rules based on artificial neural networks: Evidence from the Madrid stock market. Economics Letters, (69), 89–94.

- Gallardo J. (2009). Metodología para la Definición de Requisitos en Proyectos de Data Mining (Tesis de grado) Universidad Politécnica de Madrid, España.

- García O., Morales A. (2014). Empresas exitosas y no exitosas que cotizan en la BMV del sector comercial. Revista Estocástica, Finanzas y Riesgo, (4), 33-60.

- Garrido M. (2018). Machine Learning y Data Science con Python. Udemy.

- González A. (2014). ¿Qué es Machine Learning? Clever Data, Recuperado de https://cleverdata.io/que-es-Machine-Learning-big-data/

- Guerrero J. (2005). El mercado de valores en México y las opciones de inversión para extranjeros. Mercados y negocios, (11), 37-44.

- Guerrero V. (2011). Medición de la tendencia y el ciclo de una serie de tiempo económica, desde una perspectiva estadística. Realidad, datos y espacio. Revista internacional de estadística y geografía, INEGI, (2), 50-73.

- Hand D., Mannila H., Smyth P. (2001). Principles of Data Mining. The MIT Press. Inglaterra.

- Hanke J., Reitsch A. (2009). Pronósticos en los negocios. México: Prentice Hall.

- Huerta A. (2015). Modelos predictivos para el mercado Forex (Tesis de maestria). Universidad de Murcia, España.

- Investing.com. https://mx.investing.com/

- Johannes F., Dragan G., Nada L. (2012). Foundations of Rule Learning. Alemania: Springer-Verlag.

- Kara Y., Boyacioglu M., Kaan Ö. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector Machines: The sample of the Istanbul Stock Exchange. Expert Systems with Applications, (38), 5311-5319.

- Kyoung-jae K. (2003). Financial time series forecasting using support vector machines.

- Krishni H. (2018). A beginners guide to random forest regression. Medium. Recuperado de https://medium.com/datadriveninvestor/random-forest-regression-9871bc9a25eb/

- Larsen J. (2010). Predicting Stock Prices Using Technical Analysis and Machine Learning (Tesis de grado). Norwegian University of Science and Technology.

- López D. (2017). Escuela de Traders: Las 101 preguntas que cambiarán tu trading para siempre. España.

- Marsland S. (2014). Machine Learning: an algorithmic perspective. E.U.A.: CRC press.

- Mascareñas J., (2012). Introducción a los mercados financieros. Universidad Complutense de Madrid, España.

- Matlab. (2019). Análisis predictivo. Recuperado de https://la.mathworks.com/discovery/predictive-analytics.html

- Mishkin S. (2014). Moneda, banca y mercados financieros. México: Pearson Educación.

- Mohammed Z., Wagner M. (2014). Data Mining and analysis, fundamental concepts and algorithms. Primera edición. Cambridge University Press. E.U.A.

- Nesreen K. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. Universidad del Cairo, Egipto.

- Olden M., (2016). Predicting Stocks with Machine Learning, Stacked Classifiers and other Learners applied to the Oslo Stock Exchange (Tesis de grado). Universidad de Oslo, Noruega.

- Pachanekar R. (2018). Market Fundamentals: What Impacts the Share Price? Quant-Insti. Recuperado de https://www.quantinsti.com/blog/market-fundamentals-shareprice?utm_source=newsletter&utm_medium=email&utm_campaign=newsletter_dec2018

- Pachanekar R. (2019). Top 10 Machine Learning Algorithms For Beginners. Quant-Insti. Recuperado de https://www.quantinsti.com/blog/top-10-Machine-Learning-algorithms-beginners/

- Palma C., Palma W., Pérez R. (2009). Data Mining: El arte de anticipar. Chile: RIL editors.

- Patel S. (2017). Supervised Learning and Naive Bayes Classification. Medium. Recuperado de https://medium.com/Machine-Learning-101/chapter-1-supervised-Learning-and-naive-bayes-classification-part-1-theory-8b9e361897d5

- Pisa M. (2018). Decision Tree For Trading Using Python. Quant-Insti. Recuperado de https:// https://blog.quantinsti.com/decision-tree/

- Sánchez M. (2018). “El S&P/BMV IPC cumple 40 años”. S&P Dow Jones Indices LLC. México.

- Scikit Learn. https://scikit-learn.org/

- Semanti.Ca. (2018). Glosary of Machine Learning terms. Recuperado de https://semanti.ca/blog/?glossary-of-Machine-Learning-terms

- Singh V. (2018). Machine Learning Logistic Regression in Python: From Theory To Trading. Quant-Insti. Recuperado de https://www.quantinsti.com/blog/Machine-Learning-logistic-regression-python

- Singh D. (2018). Working Of Neural Networks For Stock Price Prediction. Quant-Insti. Recuperado de https://www.quantinsti.com/blog/working-neural-networks-stock-price-prediction

- Singh D. (2018). Training Neural Networks For Stock Price Prediction. Quant-Insti. Recuperado de https://www.quantinsti.com/blog/training-neural-networks-for-stock-price-prediction/

- S&P Dow Jones Indices. https://spindices.com/

- Tahsildar S. (2018). Introduction To Deep Learning And Neural Network. Quant-Insti. Recuperado de https://www.quantinsti.com/blog/introduction-deep-Learning-neural-network/

- Tahsildar S. (2018). Random forest Algorithm in Trading using Python. Quant-Insti. Recuperado de https://blog.quantinsti.com/random-forest-algorithm-in-python/

- Tan T., Quek C. (2005). Brain-inspired genetic complementary Learning for stock market prediction. IEEE Congress on Evolutionary Computation, 2653–2660.

- Villegas J. (2010). Ingresarios: 5 Pasos para aprender a Invertir en Bolsa. Colombia: Penguin Random House.

- Vera J., Rosado Y. (2010). Predicción del precio de acciones en la Bolsa Mexicana de Valores utilizando la “Simulación Monte Carlo”. Mercados y negocios, (11), 89-108.

- Wolpert D. (2001). The Supervised Learning No-Free-Lunch Theorems. NASA Ames Research Center. E.U.A.

- Yahoo Finance. https://finance.yahoo.com/