



UNIVERSIDAD DE CHILE

FACULTAD DE ECONOMÍA Y NEGOCIOS

ESCUELA DE ECONOMÍA Y ADMINISTRACIÓN

PREDICCIÓN DE LA VARIACIÓN DEL IPSA: COMPARACIÓN DE TRES MODELOS

Seminario para optar al título de
Ingeniero Comercial, Mención Administración

Participantes: Patricio Arroyo Campos

Claudio Ormazábal Caris

Profesor guía: Jorge Gregoire Cerda

Santiago – Primavera 2013

Resumen

En el presente estudio se evaluaron tres modelos para predecir el signo de la variación del IPSA, los cuales fueron regresión lineal, Support Vector Machines (SVM) y redes neuronales utilizando el software de data mining RapidMiner. Las variables explicativas utilizadas fueron los rezagos semanales de los retornos del IPSA, del Dow Jones y tipo de cambio USDCLP. Con estos datos se obtuvo que los modelos no lineales poseen un rendimiento levemente superior a la regresión lineal, obteniendo un 76,6% versus un 76% respectivamente. Adicionalmente se probaron distintas proporciones muestrales de entrenamiento y de prueba de los datos, obteniendo en todas estas, porcentajes de predicción sobre el 70% y con pequeñas variaciones entre ellas. Finalmente se realizó el test de Pesaran y Timmermann (1992) entregando evidencia de que todos los modelos pueden predecir en algún grado la dirección de signo del IPSA.

Tabla de contenido

1. Introducción.....	5
2. Revisión de Literatura	7
3. Metodología y Datos	11
3.1 Regresión lineal	13
3.2 Redes neuronales	13
3.3 Support Vector Machines.....	15
4. Resultados	18
4.1 Resultados de predicción de tendencia	18
4.2 Medidas de robustez.....	22
4.3 Resultados test DA	23
5. Conclusión.....	25
Bibliografía	27
Anexo	30
Anexo 1: Descripción estadística de los retornos de los datos utilizados:	30
Anexo 2: Test de acierto de direccionalidad (DA)	31
Anexo 3: tablas de resultados entregados por RapidMiner.....	33

Tabla de ilustraciones y cuadros

Figura 1: Esquema de red neuronal sencilla	14
Figura 2: Resultado de la predicción de la tendencia bajo la metodología de regresión lineal	18
Figura 3: Resultado de la predicción de la tendencia bajo la metodología de redes neuronales.....	18
Figura 4: Resultado de la predicción de la tendencia bajo la metodología de SVM	¡Error! Marcador no definido.8
Figura 5: Resultado gráfico bajo la metodología de regresión lineal	¡Error! Marcador no definido.9
Figura 6: Resultado gráfico bajo la metodología de redes neuronales.....	20
Figura 7: Resultado gráfico bajo la metodología de SVM	20
Cuadro 1: Resultados muestra 70-30.....	21
Cuadro 2: Resultados muestra 20-80.....	22
Cuadro 3: Resultados muestra 30-70.....	¡Error! Marcador no definido.
Cuadro 4: Resultados muestra 50-50.....	¡Error! Marcador no definido.2
Cuadro 5: Resultados test DA muestra 70-30	¡Error! Marcador no definido.
Cuadro 6: Resultados test DA muestra 20-80	¡Error! Marcador no definido.
Cuadro 7: Resultados test DA muestra 30-70	¡Error! Marcador no definido.
Cuadro 8: Resultados test DA muestra 50-50	¡Error! Marcador no definido.

1. Introducción

Este estudio analiza la capacidad de tres modelos para predecir el signo de las variaciones semanales del índice de precios selectivo de acciones (IPSA), perteneciente al mercado bursátil chileno, los modelos a testar son de regresión lineal, Support Vector Machines (SVM) y de redes neuronales, cambiando la proporción de datos para trabajar en los modelos y probarlos. Nos centramos en el IPSA ya que es uno de los índices más importante de nuestro país, el cual incluye las 40 acciones con mayor presencia bursátil, siendo esta lista elaborada por la Bolsa de Comercio de Santiago y revisada anualmente.

Analizamos la variación del signo de este índice ya que un estudio de este tipo puede arrojar mejores resultados que aquellos basados en la proyección de la variable observada y actualmente la predicción del movimiento de retornos de mercado es un tema que cada vez adquiere mayor importancia en el ámbito financiero. En este contexto, la contribución de este estudio radica en el análisis de distintos modelos a fin de medir su capacidad de predicción de movimiento del IPSA, sobre la base del número de veces en que la dirección pronosticada es la correcta. Para llevar a cabo este estudio se utilizó el programa RapidMiner, software de Data Mining, con esto se logró la simplificación del proceso de predicción de signo debido a los variados algoritmos que incorpora en su aplicación, contando también con módulos que facilitan en gran medida el trabajo, ya que posee una interfaz bastante intuitiva.

Este trabajo constituye un complemento a estudios desarrollados con anterioridad, los cuales ahondan en el tema de la predicción de signo de variación del IPSA, intentando encontrar el modelo más idóneo sobre el tema ya señalado. Para esto utilizaremos un rango más extenso y más reciente de datos, a diferencia de los estudios ya existentes.

Dentro de la principal motivación para desarrollar este estudio contamos con el hecho de que si bien el mercado accionario chileno no es tan profundo como el de

los mercados desarrollados, cada vez se encuentra ganando más terreno, esto se puede ejemplificar con el hecho de que el mes de septiembre de 2013 se lanzó el primer ETF que replica al IPSA, primer instrumento de este tipo sobre el índice, demostrándonos además que el mercado chileno se está ampliando, debido a que este instrumento permitirá extender las posibilidades de estrategias para los inversionistas que busquen exposición en el mercado chileno.

El trabajo se organiza de la siguiente manera. En el capítulo 2 presentaremos una revisión de la literatura relacionada a este estudio. En el capítulo 3 se describen los datos utilizados y como se desarrolla la metodología. El capítulo 4 presenta los principales resultados y medidas de robustez, por último en el capítulo 5 se presentan las conclusiones.

2. Revisión de Literatura

La hipótesis de mercados eficientes (Fama, 1970) plantea que el mercado refleja completa y correctamente toda la información relevante para la determinación de los precios de los activos, ya que si vemos que el surgimiento de nueva información es de carácter aleatorio, los cambios que se ven reflejados en los activos, también lo serían y seguirían un camino aleatorio. Esto significaría, además el hecho de que la teoría de caminata aleatoria implicaría que al usar series de tiempo pasados de variaciones en el precio de las acciones, no permitiría obtener una rentabilidad mayor a la obtenida mediante una estrategia de buy and hold. Jensen y Benington (1970) apoyan este punto, los autores responden al trabajo de Levy (1967) usando dos técnicas de trading que utiliza este autor y encuentran resultados totalmente distintos a él. Al aplicar estas técnicas sobre otros títulos y en distintos períodos de tiempo, no encuentran evidencia que éstas generen un retorno mayor a una estrategia pasiva.

De acuerdo a lo anteriormente planteado, un comportamiento random walk de los precios de las series de tiempo invalidaría la premisa de tratar de predecir los retornos de los activos, ante eso surgen muchos estudios que han presentado evidencia significativa respecto a la predictibilidad de los precios de títulos de mercado, lo que confirmaría que estos no siguen un camino aleatorio, por lo que sí podríamos predecir movimientos del retorno de los activos, por esto se ha tratado de desarrollar modelos no lineales para mejorar el ajuste y predicción de modelos bursátiles y de acciones, uno de ellos es el obtenido por Lo y Mackinlay (1988) los cuales utilizando datos semanales de mercados desarrollados tales como Estados Unidos, Japón y Europa Occidental, encuentran una correlación serial positiva entre los retornos semanales, es decir que los precios accionarios no siguen un camino aleatorio y los retornos son predecibles en algún grado. Jegadeesh (1990) estudia la predictibilidad de los de los retornos mensuales sobre activos individuales y sus resultados evidencian una correlación serial negativa de primer

orden significativamente alta para rezagos de dos meses y una correlación serial positiva para rezagos mayores.

Además algunos investigadores han encontrado que los mercados se comportan de una manera no lineal (Bosarge, 1993) lo cual nos motiva a que dos de nuestros tres modelos de predicción permitan este comportamiento, en especial el de redes neuronales, ya que modelos de este tipo han presentado un buen desempeño en modelamientos no lineales (Widrow, Rumelhart y Lehr, 1994). De acuerdo a Martín del Brío y Sanz (1997), las redes neuronales artificiales “son sistemas de procesamiento que copian esquemáticamente la estructura neuronal del cerebro para tratar de reproducir sus capacidades”, definiendo además que existen tres características que permiten un mejor rendimiento de las redes neuronales sobre otros modelos, las cuales son su capacidad de tener sistemas paralelos (cuentan con una gran cantidad de neuronas o procesadores elementales, PEs, cada uno de los cuales trabaja paralelamente con una pequeña parte de un problema mayor), distribuidos (cuentan con muchos PEs a través de los cuales distribuyen su memoria) y adaptativos (tienen la capacidad de adaptarse al entorno modificando sus pesos y sinapsis de manera de encontrar una solución aceptable al problema) .

Entendemos además que la predicción de la dirección del movimiento del IPSA es relevante para generar estrategias de transacción que sean efectivas y que puedan entregar una mejor estrategia que tratar de obtener una proyección lo más exacta posible del precio de cierre semanal de la variable a tratar según lo encontrado por Leung, Daouk y Chen (2000). Estos autores realizan su estudio comparando la capacidad predictiva de los modelos de clasificación (*linear discriminant analysis, logit model, probit model y probabilistic neural network*) con los de estimación de nivel (*adaptive exponential smoothing, vector autoregression with Kalman filter, multivariate transfer function y multilayered feedforward neural network*) concluyendo que los primeros se desempeñan mejor a la hora de utilizarlos para predecir el signo o la dirección de retornos de los índices bursátiles estudiados. Lo anteriormente planteado sobre centrar la

atención en la predicción es muy relevante para los traders ya que apunta a que estos deben centrar sus esfuerzos en predecir con precisión la dirección de los movimientos en vez de minimizar la desviación estándar de las estimaciones de los valores observados, ya que estas estrategias pueden generar beneficios más altos (Wu y Zhang, 1997), esto también se encuentra apoyado por lo encontrado por O'Connor, Remus y Griggs (1997) quienes afirman que la utilidad será mayor al proyectar la dirección del cambio en el nivel de precios más que el nivel de precios en sí.

Hodgson y Nicholls (1991) en la línea de lo anteriormente planteado concluyen que el grado de predictibilidad de los retornos es generalmente considerado económicamente no significativo y proponen centrar el análisis en evaluar la significancia económica de predecir la dirección de cambio de los precios de activos lo cual sigue apoyando las estrategias basadas en esta predicción.

En el mercado chileno, se cuenta con el estudio desarrollado por Gregoire (1985), en su trabajo sobre el ajuste de los precios accionarios a la información en el mercado chileno, donde utiliza la metodología de Fama (1970), concluye que no se encuentra evidencia que las fluctuaciones de precios siguieran un random walk, pero no se puede rechazar la eficiencia débil de mercado al contrastar filtros lineales con una estrategia buy and hold. El comportamiento de los rendimientos parece compatible con un modelo de submartingala, agregando además que el mercado accionario chileno se caracteriza por un ajuste lento a la nueva información, en comparación a mercados desarrollados, esto debido a que se encuentra afectado principalmente a una baja liquidez del mercado y la frecuencia baja de transacciones, constituyendo factores que contribuyen a la existencia de patrones no lineales en los retornos, en línea con lo evidenciado años después por Urrutia (1994), el cual examinando la teoría del random walk para cuatro países latinoamericanos, Argentina, Brasil, Chile y México, encuentra evidencia que esta teoría sólo debería ser rechazada en Chile.

Continuando con los análisis desarrollados en nuestro país destaca el trabajo realizado por Parisi (2002) en donde se evalúa la capacidad de varios modelos de

redes neuronales para predecir el signo de variación semanal del IPSA, encontrando que el mejor modelo es una Red Ward Recursiva, con una capacidad predictiva extramuestral de 72%. Parisi (2004) realiza otro estudio también sobre el IPSA, pero ahora analizando si es mejor utilizar una red neuronal para predecir el signo de dirección semanal del IPSA o un algoritmo genético, encontrando que la Red Ward tendría mayor capacidad que el modelo de algoritmos genéticos y que el modelo naive para predecir el signo de las variaciones del índice chileno. Además para entregarle mayor fortaleza a sus resultados, realizó una evaluación de los modelos en 60 series generadas por un proceso *blok-bootstrap* y fueron considerados los costos de transacción.

3. Metodología y Datos

La muestra utilizada es de 527 datos, valores de cierre semanal del IPSA, el Dow Jones y el tipo de cambio USDCLP para el periodo comprendido entre el 24 de enero del 2003 hasta el 25 de enero del 2013¹, adicionalmente con estos valores de cierre se calcularon los retornos semanales como la diferencia porcentual entre el precio de cierre de la semana t y la semana $t+1$, luego se calcularon 3 rezagos en el pasado de cada una de las variables y retornos previamente obtenidos, es decir, se tienen los valores de $t-1$ que es donde se realiza la predicción y para poder obtener esta predicción se utilizan los valores de los retornos de las variables en los 3 periodos previos $t-2$, $t-1$ y t .

Se escogieron estas variables debido a que han sido utilizadas en estudios anteriores (Parisi 2004) acerca de la temática y porque creemos en su fuerte correlación con el desempeño de las empresas que están presentes en el índice bajo estudio.

La metodología a utilizar es a través del uso del programa de Data Mining RapidMiner5. Para comprender más sobre el software se procederá a realizar una breve introducción de él y sus utilidades.

RapidMiner provee softwares, soluciones y servicios en el campo de análisis predictivo, data mining y text mining. Pudiendo automática e inteligentemente analizar datos, incluyendo bases de datos y texto, en una gran escala (RapidMiner, 2013).

Este software ha sido desarrollado desde el 2001 y es una de las soluciones más utilizadas en el mundo para el análisis de datos. Adicionalmente, RapidMiner (ex Rapid-I) es actualmente el único proveedor de código abierto en ofrecer un rango completo de las herramientas analíticas de negocio. Estas comienzan con los datos de almacenamiento y ETL y finalizan con la creación de reportes basados en la web y cuadros de mando. A diferencia de las soluciones tradicionales los

¹ Ver Anexo 1: Descripción estadística de los retornos de los datos utilizados.

productos de RapidMiner siguen un procedimiento único, donde cada transformación, cada visualización y cada análisis siempre se modelan como un proceso con la misma herramienta. (Rapid-I, 2012)

Algunas características principales de RapidMiner son:

- Acceso a más de 40 tipos de archivos y manejo de datos, filtros de filas, columnas y valores perdidos, además de unión de bases de datos.
- Exportación e importación de modelos de RapidMiner, R y Weka.
- Operador para aplicar los modelos a los conjuntos de datos (Scoring)
- Sampling: Estratificado, balanceado, bootstrapping, etc.
- Data partitioning, Transformaciones (Normalizaciones y estandarizaciones, entre otras), binning, Reemplazo de datos, Ponderación y selección, generación de atributos y estadísticas descriptivas.
- Gráficos y visualizaciones.
- Similitudes y medidas de distancia.
- Modelos de clustering, de análisis de cesto de supermercado, árboles de decisiones, reglas de inducción, modelos bayesianos, regresiones, Redes Neuronales, Support Vector Machines, entre otros.
- Evaluación de modelos.
- Scripting, puede utilizar modelos escritos en R.
- Procesos de control, optimización automática, entre otras.

Adicionalmente, desde el software se pueden descargar variados complementos y extensiones que permiten trabajar distintos tipos de datos tales como análisis de series de tiempo, text mining, entre otros complementos. (Rapid-I, 2012)

Dentro del amplio rango de características de este software, para este trabajo se utilizaron 3 métodos distintos de predicción para poder conocer el valor futuro del IPSA con las variables antes mencionadas, que son en primer lugar una regresión lineal, seguida de un modelo SVM y finalmente el uso de redes neuronales. Para poder llevar a cabo estos modelos, se calcularon los retornos semanales utilizando los valores del IPSA conseguidos desde la plataforma Bloomberg.

A continuación se presenta una breve descripción de los modelos utilizados:

3.1 Regresión lineal

La regresión lineal es una técnica estadística utilizada para estudiar la relación entre variables. La regresión lineal puede utilizarse para explorar y cuantificar la relación entre una variable dependiente (Y), y una o más variables independientes o predictores () así como para desarrollar una ecuación lineal con fines predictivos (Valles Martínez et al, 2012). Este último es la utilidad que se le dio en el presente estudio. Las ecuaciones formadas por esta técnica son del tipo:

Con esto se tiene que para cada factor se debe tener un ponderador de manera que contribuyan a la pendiente de la recta. Este tipo de técnica solo captura los efectos lineales de las variables independientes sobre la dependiente.

En las regresiones de nuestro estudio las variables independientes serían cada uno de los rezagos del IPSA, Dow Jones y Tipo de cambio, y la variable dependiente sería el valor del retorno del IPSA para la semana según lo antes descrito.

3.2 Redes neuronales

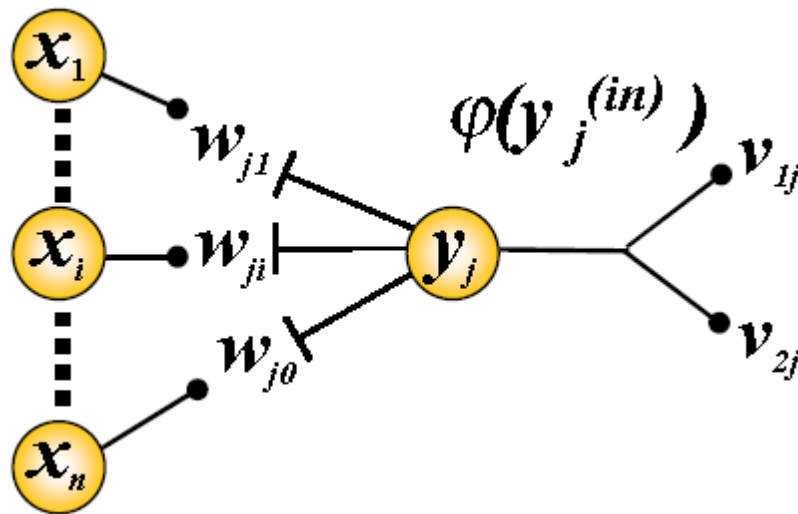
Una neurored es un procesador de información, de distribución altamente paralela, constituido por muchas unidades sencillas de procesamiento llamadas neuronas. Las neuroredes se caracterizan principalmente por:

- Tener una inclinación natural a adquirir el conocimiento a través de la experiencia, el cual es almacenado, al igual que el cerebro, en el peso relativo de las conexiones interneuronales.
- Tienen una altísima plasticidad y gran adaptabilidad, son capaces de cambiar dinámicamente con el medio.
- Posee un alto nivel de tolerancia a fallas.

- Tienen un comportamiento altamente no lineal, lo que les permite procesar información procedente de otros fenómenos no lineales.

Un esquema de representación matemático sencillo de este tipo de técnicas es el siguiente:

Figura 1: Esquema de red neuronal sencilla



Fuente: Saavedra, 2000.

En él se puede observar que la neurona de interés es que para el estudio representa el valor del retorno del IPSA en . Las n neuronas están enviando señales de entradas, que son valores numéricos de “algo”, lo que en nuestro estudio podrían ser los retornos pasados del IPSA, Dow Jones y tipo de cambio en distintas combinaciones. Los valores de representan los pesos sinápticos de las dendritas de , donde el primer índice de la neurona denota a la neurona hacia donde se dirige la información, el segundo índice denota de qué neurona procede la información. Lo que hace cada peso sináptico es simplemente multiplicar a su entrada correspondiente y define la importancia relativa de cada entrada (Saavedra, 2000), esta entrada varía dependiendo de las ecuaciones formadas por la red neuronal, porque este método para realizar la predicción utiliza múltiples ecuaciones lineales para así poder capturar efectos no lineales.

3.3 Support Vector Machines

El modelo de SVM determina el hiperplano óptimo que separa el conjunto de datos². Para este propósito, “linealmente separable” requiere encontrar el par (w, b) tal que clasifique correctamente los vectores de ejemplos en dos clases, es decir, para un espacio de hipótesis dado por un conjunto de funciones se impone la siguiente restricción:

Los hiperplanos que satisfacen se conocen como hiperplano canónicos. El objetivo de SVM es encontrar entre todos los hiperplanos canónicos que clasifican correctamente los datos, aquel con menor norma, lo que es equivalente a encontrar el hiperplano separador para el cual la distancia entre dos envolturas convexas, medida a lo largo de una línea perpendicular del hiperplano, es maximizada. Esta distancia se conoce como margen.

El problema de maximización del margen es:

—

Sujeto a:

A partir de esta formulación se construye el dual mediante la técnica de los multiplicadores de Lagrange. La formulación dual permitirá construir funciones de clasificación no lineales, lo que usualmente lleva a un mayor poder predictivo, esta se representa como:

—

² De ser posible separarlo, de ser conjuntos no linealmente separables se proyecta el conjunto de datos a un espacio de mayor dimensión, donde se construye un hiperplano separador de máximo margen.

Sujeto a:

Donde α_i representa los multiplicadores de Lagrange asociados a las restricciones. Los multiplicadores que cumplen con la condición de $\alpha_i > 0$ son llamados Support Vectors, debido a que son los únicos que participan en la construcción del hiperplano de clasificación. Se tiene además que $\alpha_i \geq 0$ y además que $\alpha_i = 0$ para cada Support Vector. La función de decisión puede escribirse como (Weber, 2012):

Ahora si incorporamos esta definición del modelo SVM a nuestro trabajo es equivalente a decir que para predecir el valor del IPSA, al igual que con la red neuronal se utilizaran distintas ecuaciones las que en este modelo utilizan unos multiplicadores de lagrange denominados Support Vectors para ponderarlas y que tratan de capturar los efectos lineales y no lineales de la variable dependiente (Retorno del IPSA) con respecto a las variables independientes que serían cada uno de los rezagos de los tres indicadores económicos seleccionados además de los rezagos del propio IPSA.

Una vez introducidos los modelos a utilizar, se procederá a caracterizar en mayor detalle los datos utilizados. Como se mencionó anteriormente los datos seleccionados corresponden al IPSA, Dow Jones y tipo de cambio USDCLP del periodo de estudio, a estas variables se les calculó tres rezagos para ver cómo afecta el pasado en su tendencia actual. Con esta transformación se perdieron 3 observaciones más 1 observación que se pierde por el uso de retornos en vez de precios de cierre, quedando con un total de 523 datos, de este total de datos se separó la muestra en 2 sub-muestras una de entrenamiento y una de prueba, separadas en aproximadamente un 70-30% de los datos, quedando la muestra de

entrenamiento con un total de 368 observaciones y la muestra de prueba con 155 datos.

El objetivo de este trabajo apunta más allá de obtener el valor exacto de retorno del IPSA sino más bien obtener el signo de la tendencia de este, para ello se utilizó un horizonte de 1 periodo en el tiempo, es decir deseamos conocer el la tendencia que tendrá el índice IPSA en una semana en el futuro. Dada las características del software utilizado la medición de la precisión de la tendencia es posible obtenerla mediante un operador que mide el Performance del modelo con esto se tiene una medida de fácil comparación entre las tres metodologías utilizadas.

Adicionalmente debido a que en la literatura empírica previa en Chile no existe un consenso, ni una clara explicación del tamaño o proporción de la muestra de entrenamiento y prueba, se ha decidido como medida de robustez de los resultados utilizar distintas proporciones de estas muestras, que son 20%-80%, 30%-70% y 50%-50% porcentajes para la muestra de entrenamiento y prueba respectivamente, los resultados de estas pruebas se pueden encontrar en la sección 4.2 Medidas de robustez.

Por otra parte, para examinar la precisión direccional se utilizará como medida adicional el test de Pesaran y Timmermann (1992), medida que ha sido utilizada en literatura previa como medida de desempeño para los modelos de redes neuronales (Parisi (2002), (Estrella A., 1998), (Qi, 2001)). En el anexo 2 se puede encontrar en mayor detalle las características de este Test.

4. Resultados

4.1 Resultados de predicción de tendencia

Cabe destacar que en este estudio, a pesar de que los modelos entregan valores de los retornos predichos, se testearán los porcentajes de aciertos de la dirección de la variación de ellos. Los resultados de la predicción de la tendencia obtenidos por las tres metodologías antes mencionadas se pueden observar en las imágenes a continuación:

Figura 2: Resultado de la predicción de la tendencia bajo la metodología de regresión lineal

```
prediction_trend_accuracy  
prediction_trend_accuracy: 0.760
```

Figura 3: Resultado de la predicción de la tendencia bajo la metodología de redes neuronales

```
prediction_trend_accuracy  
prediction_trend_accuracy: 0.766
```

Figura 4: Resultado de la predicción de la tendencia bajo la metodología de SVM

```
prediction_trend_accuracy  
prediction_trend_accuracy: 0.766
```

En ellas se puede observar en primer lugar el resultado de la predicción la tendencia en el modelo de regresión lineal con un 76,0% de efectividad, mientras que en el modelo bajo la metodología de SVM y redes neuronales se obtuvo un 76,6% de precisión en ambos casos, estos resultados son considerados aceptables debido a que se logró predecir correctamente en más de tres cuartas partes de la muestra el signo de tendencia de una forma eficaz.

Si analizamos los resultados de forma gráfica para los modelos obtenemos las siguientes imágenes:

Figura 5: Resultado gráfico bajo la metodología de regresión lineal

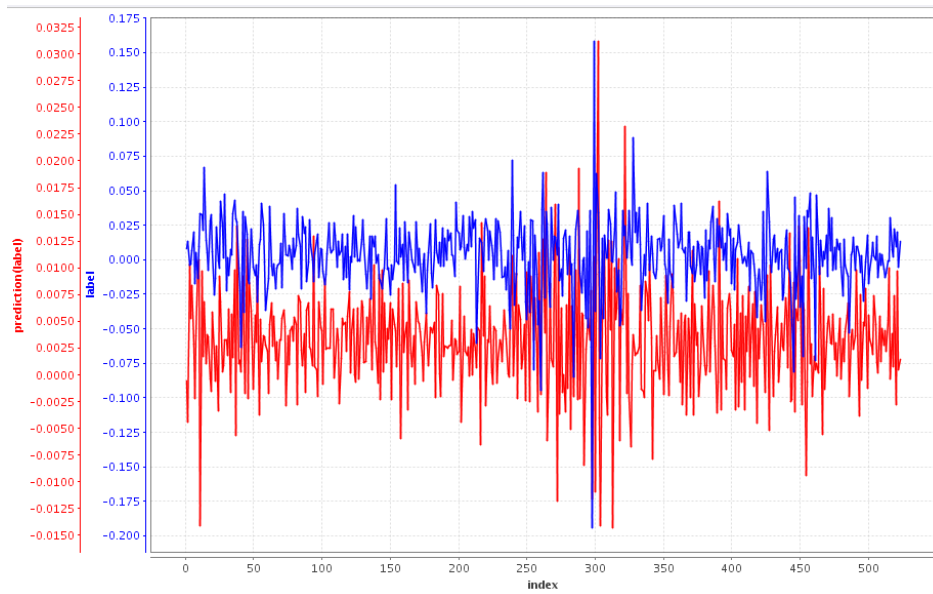


Figura 6: Resultado gráfico bajo la metodología de redes neuronales

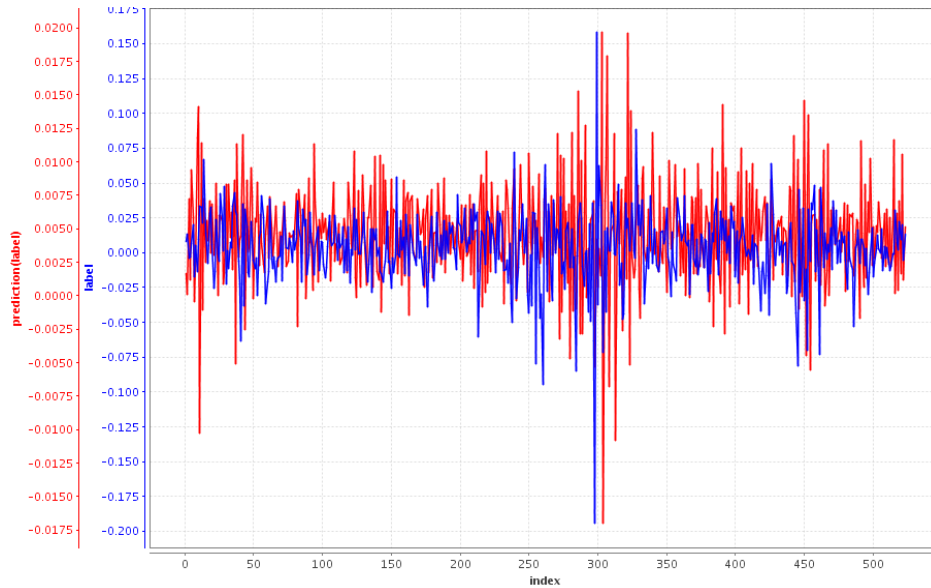
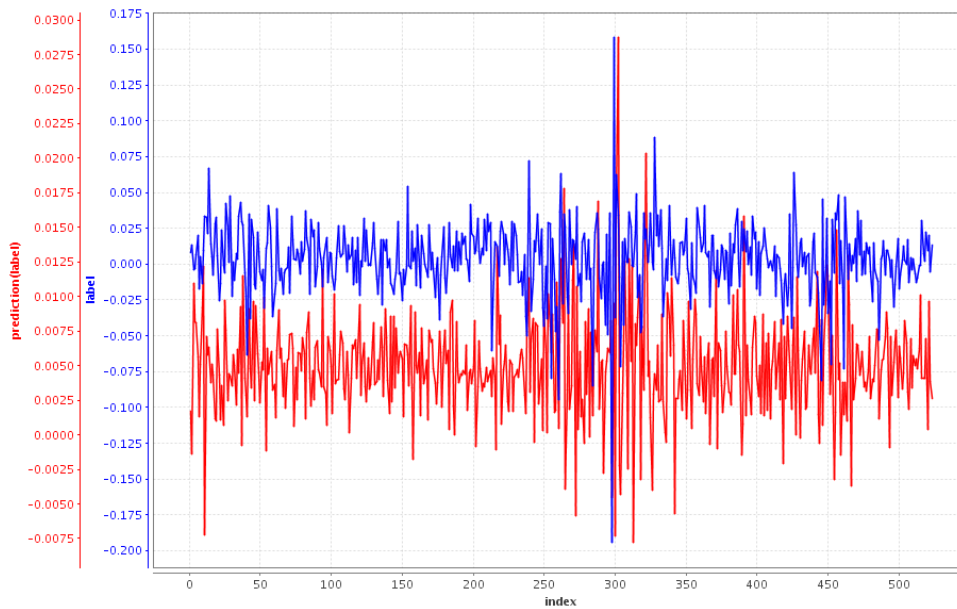


Figura 7: Resultado gráfico bajo la metodología de SVM



Con los gráficos anteriores se puede ver que los modelos siguen de manera similar al índice pero no se puede apreciar a primera vista cuál tiene un desempeño superior. Para esto, más adelante se comparan los modelos según sus distintas medidas de error en la tabla resumen de los modelos.

Un hecho que vale la pena recalcar es que para la regresión lineal el programa utiliza métodos estadísticos para determinar las variables estadísticamente significativas a los niveles convencionales tales como el retorno del Dow Jones con dos rezagos (p-value 0,001), el retorno del Dow Jones con un periodo de rezago (p-value 0,003), el retorno del tipo de cambio con dos rezagos (p-value 0,058) y finalmente el retorno del índice IPSA con dos rezagos (p-value 0,087). Como se aprecia la mayor cantidad de variables relevantes para la predicción del IPSA pertenecen a un horizonte temporal de dos semanas previas, con esto se podría plantear la hipótesis que el mercado toma sus decisiones incorporando información del pasado que ya comprueba su real tendencia más allá de incorporar nueva información con sobrerreacciones emocionales en ella, pudiendo estar acorde a la teoría del análisis técnico de que los inversionistas buscan puntos que les confirman la tendencia en base a la información pasada para tomar una posición más segura (Murphy, John J. (2000)). Esta es solo una hipótesis que abre camino a ser comprobada en estudios posteriores.

A continuación se presenta un cuadro resumen con la información de los modelos estimados de manera de hacer una comparación más fácil de ellos:

Cuadro 1: Resultados muestra 70-30

70-30			
Modelo	Predicción de tendencia	Error medio	Error cuadrático medio
Regresión lineal	0,760	0	0,063
Red Neuronal	0,766	-0,002	0,058
SVM	0,766	-0,001	0,057

En el cuadro anterior se puede observar que los 3 modelos presentan similares niveles de predictividad y a la hora de analizar sus errores estos tampoco presentan variaciones significativas, dado esto se puede plantear la hipótesis de que en el mercado chileno el efecto no lineal en las acciones no es tan significativo

debido a que una regresión lineal posee niveles de acierto elevados y similares a los de una red neuronal o modelo SVM.

4.2 Medidas de robustez

Como medida de robustez de los resultados se calculó la predicción anteriormente realizada utilizando distintas proporciones de división de la muestra, que son 20%-80%, 30%-70% y 50%-50%, siendo el primer dígito el porcentaje de muestra para entrenar el modelo y el segundo valor el porcentaje usado para testear el modelo, obteniendo los siguientes resultados³:

Cuadro 2: Resultados muestra 20-80

20-80			
Modelo	Predicción de tendencia	Error medio	Error cuadrático medio
Regresión lineal	0,748	0	0,063
Red Neuronal	0,739	0,004	0,076
SVM	0,722	-0,002	0,057

Cuadro 3: Resultados muestra 30-70

30-70			
Modelo	Predicción de tendencia	Error medio	Error cuadrático medio
Regresión lineal	0,730	0	0,063
Red Neuronal	0,725	0,002	0,069
SVM	0,749	-0,002	0,057

Cuadro 4: Resultados muestra 50-50

50-50			
Modelo	Predicción de tendencia	Error medio	Error cuadrático medio
Regresión lineal	0,738	0	0,063
Red Neuronal	0,731	-0,001	0,06
SVM	0,758	-0,002	0,057

³ En anexo 3 se puede encontrar los resultados entregados por RapidMiner.

Como se observa en los resultados recientemente presentados los valores obtenidos en las distintas proporciones de muestra no distan significativamente unos de otros por lo que los resultados se podrían considerar robustos, además cabe destacar, que en todas las simulaciones realizadas se obtuvo un porcentaje de acierto superior al 70% en el signo de variación del IPSA con un horizonte de una semana, manteniéndose incluso en la regresión lineal lo que reafirma la hipótesis de que los factores no lineales no poseen un peso relativo significativo dentro de este mercado, al considerar las variables en estudio. Entregando evidencia de que en el mercado chileno hay cabida para el análisis técnico, y que a través del análisis de los precios pasados y la inclusión de algunas variables fundamentales, se pueden obtener resultados interesantes de acierto en las predicciones de la variación de los precios de este índice.

4.3 Resultados test DA

A cada uno de los modelos se les aplicó el test DA de Pesaran y Timmermann (1992), con el objetivo de medir la significancia estadística de la capacidad predictiva de los modelos, estos autores plantean la hipótesis nula de que μ_1 e μ_2 se encuentran independientemente distribuidos. El test presenta una distribución normal estándar. Los resultados se presentan en los siguientes cuadros para las distintas proporciones muestrales estudiadas:

Cuadro 5: Resultados test DA muestra 70-30

70-30		
Modelo	Estadístico DA	H0
Regresión lineal	-5,494	Se Rechaza
Red Neuronal	-3,049	Se Rechaza
SVM	-2,551	Se Rechaza

Cuadro 6: Resultados test DA muestra 20-80

20-80		
Modelo	Estadístico DA	H0
Regresión lineal	-5,494	Se Rechaza
Red Neuronal	-12,476	Se Rechaza
SVM	-2,637	Se Rechaza

Cuadro 7: Resultados test DA muestra 30-70

30-70		
Modelo	Estadístico DA	H0
Regresión lineal	-5,494	Se Rechaza
Red Neuronal	-7,066	Se Rechaza
SVM	-2,637	Se Rechaza

Cuadro 8: Resultados test DA muestra 50-50

50-50		
Modelo	Estadístico DA	H0
Regresión lineal	-5,494	Se Rechaza
Red Neuronal	-3,294	Se Rechaza
SVM	-2,148	Se Rechaza

Los resultados muestran que para los 3 modelos, en las distintas proporciones muestrales bajo estudio se rechaza la hipótesis nula, lo que permite concluir que es posible predecir el cambio direccional del signo semanal del IPISA, en todos los escenarios estudiados, con esto se puede confirmar que no existen diferencias significativas en los resultados al modificar las proporciones utilizadas para entrenamiento y prueba del modelo por lo menos bajo el uso de RapidMiner.

5. Conclusión

De los diversos modelos que hemos utilizado en este estudio, la red neuronal y el modelo SMV fueron los que tuvieron un rendimiento levemente superior en la predicción del signo de las variaciones semanales del IPSA, en relación a la evaluación de la capacidad de la prueba del signo, alcanzando una capacidad predictiva de un 76,6%, luego tenemos el nivel de predicción de la regresión lineal, el cual es menor, ya que alcanzó un 76% de aciertos en el signo.

Para entregarle más robustez a nuestros resultados y dado que la literatura en Chile no ha hecho referencia en relación a la cantidad de observaciones que deben ser utilizadas para entrenar un modelo y luego probarlo, decidimos realizar cuatro medidas más para los tres modelos con distintos porcentajes de datos utilizados, encontrando que la relación 70-30 (70% de los datos para entrenar el modelo y 30% para probarlo) es la que entrega la mayor capacidad de predicción para todas las metodologías que hemos analizado, pero su eficacia es solo levemente superior, debido a que en todas las proporciones muestrales se obtuvo un rendimiento mayor al 70%. Adicionalmente los resultados del test DA nos entregaron que en todos los casos de estudio se rechaza la hipótesis nula, con esto se concluye que todos los modelos son capaces de predecir el cambio direccional de la rentabilidad del IPSA.

Debemos destacar también que se pudo observar que en general, las variables de dos rezagos de anterioridad afectan de manera significativa la predicción de la tendencia del IPSA actual, siendo aún más relevante que la información de la semana más reciente, observando que en promedio los inversionistas debiesen esperar a confirmar las informaciones y reacciones del mercado antes de tomar sus posiciones. Lo anterior ha mostrado que el signo del retorno accionario de uno de los principales índices bursátiles del país es predecible en algún grado, ratificando estudios empíricos anteriores en Chile que se mencionaron en el capítulo de revisión de literatura.

Finalmente una ampliación de este estudio podría ser medir el impacto económico de estas predicciones, de esta manera el uso de estas herramientas de inteligencia artificial puede ser útil para los operadores de mesa de trading o analistas técnicos, ya que quedó demostrada su buena capacidad de predicción, lo cual puede ser un buen complemento para sus estrategias a la hora de seleccionar sus portafolios de inversión.

Bibliografía

- Bosarge W.E. (1993). Adaptive processes to exploit the nonlinear structure of financial markets. In R.R. Trippi & E. Turban (eds.), *Neural Networks in finance and investing*. New York, Irwin, 371-402.
- Estrella A., y. M. (1998). Predicting US recessions: Financial variables as leading indicators. *The Review of Economics and Statistics* , 45-61.
- Fama, Eugene (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance* 25, 383-417.
- Gregoire, J. (1985). El ajuste de los Precios Accionarios a la Información: Resultados Empíricos. *Paradigmas en Administración (Santiago de Chile)* 7, pp. 121–131.
- Hodgson A., Nicholls D (1991). The impact of index futures markets on Australian share market volatility. *Journal of Business Finance and Accounting* 18, 267-280.
- Jegadeesh, Narasimhan (1990). Evidence of predictable behavior of security returns. *The Journal of Finance*, Vol. 45(3), 881-898 (July).
- Jensen, M., Bennington, G. (1970). "Random walks and technical theories: Some additional evidence". *Journal of Finance*, 25, 469-482.
- Leung Mark, Daouk Hazem, Chen An-Sing (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* (16) 2, 173-190.
- Levy, R. (1967a). Random Walks: Reality or Myth. *Financial Analyst Journal*. November-December, 1967a).
- Levy, R. (1967b). Relative Strength as a Criterion for Investment Selection. *Journal of Finance*. XXII December, 1967b), 595-610.

Lo, A., MacKinley A. (1988). Stock Market Price do not follow random walk: Evidence from a simple specification test. *Review of Financial Studies* 1, 41-66.

Martín del Brío, B., Sanz, A. (1997). *Redes neuronales y sistemas borrosos: Introducción, teórica y práctica*. Primera Edición. Ra-ma. 387 pp.

Murphy, John J. (2000). *Análisis técnico de los mercados financieros*. New York Institute of Finance, Cap 4: Conceptos básicos de tendencia.

O'Connor, M., Remus, W. y Griggs, K. (1997). Going up-going down: How good are people at forecasting trends and changes in trends?. *Journal of Forecasting* 16, 165-176.

Parisi, Antonino. (2002). Evaluación de modelos de redes neuronales de predicción del signo de variación del IPSA. *Estudios de Administración*, vol. 9, N°1, 67-103. Chile.

Parisi A. (2004). Modelos de algoritmos genéticos y redes neuronales en la predicción del signo de variación del IPSA. *Estudios de Administración*, CL 11(1), 55-79.

Parisi A., Parisi F, Díaz D. (2006). Modelos de algoritmos genéticos y redes neuronales en la predicción de índices bursátiles asiáticos. *Cuadernos de economía, Latin American Journal of Economics*. 128, 251-284.

Pesaran, M. y Timmermann, A. (1992). A simple nonparametric test of predictive performance. *Journal of Business and Economic Statics* 10, pp. 461-465.

Qi, M. (2001). Predicting US recessions with leading indicators via neural network models. *International Journal of Forecasting*, 383-401.

Rapid-I. (2012). *Fact Sheet: RapidMiner and RapidAnalytics*. Dortmund: Rapid-I GmbH.

RapidMiner. (2013). *RapidMiner*. Recuperado el 17 de 12 de 2013, de Sección About us: <http://rapidminer.com/about-us/>.

Saavedra, C. y. (2000). *Redes Neuronales Artificiales*. Concepción, Chile: Departamento de Física, Universidad de Concepción.

Urrutia, J. (1994). Time series properties of four Latin American equity markets: Argentina, Brazil, Chile and Mexico. *Estudios de Administración* 2, 1-8.

Valles Martínez, M. S., Arribas, A., Díaz, E., Álvarez, M. R., & D'Ancona, F. B. (16 de Julio de 2012). SPSS 10. GUÍA PARA EL ANÁLISIS DE DATOS. *Capítulo 18 Análisis de regresión lineal: El Procedimiento de regresión lineal*. Madrid, Madrid, España: Universidad Complutense de Madrid.

Weber, S. M. (2012). Modelo de selección de atributos para Support Vector Machines. *Revista de ingeniería y sistemas*, 49-70.

Widrow B., Rumelhart D.E., Lehr M.A. (1994). Neural Networks: Applications in industry, business and science *Communications of the ACM*, 37, 3 (March), 93-105.

Wu Y., Zhang H. (1997) Forward premiums as unbiased predictors of future currency depreciation: A non-parametric analysis. *Journal of International Money and Finance* 16, 609-623.

Anexo

Anexo 1: Descripción estadística de los retornos de los datos utilizados:

1.1 Descripción Estadística de los retornos del IPSA

Descripción Estadística de los retornos del IPSA

Fecha	Nº de Obs.	Desviación estándar	Media	Mediana	Mínimo	Máximo	Skewness	Kurtosis
24/01/2003 - 25/01/2013	523	0,025788686	0,00323616	0,00394603	-0,19424609	0,15798391	-0,8344616	9,45405126

1.2 Descripción Estadística de los retornos del Dow Jones

Descripción Estadística de los retornos del Dow Jones

Fecha	Nº de Obs.	Desviación estándar	Media	Mediana	Mínimo	Máximo	Skewness	Kurtosis
24/01/2003 - 25/01/2013	523	0,024102066	0,00132022	0,00257158	-0,18151293	0,11290914	-0,55738363	7,80982648

1.3 Descripción Estadística de los retornos del USDCLP

Descripción Estadística de los retornos del USDCLP

Fecha	Nº de Obs.	Desviación estándar	Media	Mediana	Mínimo	Máximo	Skewness	Kurtosis
24/01/2003 - 25/01/2013	523	0,017031608	-0,00070548	-0,00116838	-0,06197389	0,11804996	1,3405778	7,95812786

Anexo 2: Test de acierto de direccionalidad (DA)

El test de acierto de direccionalidad fue desarrollado por Pesaran y Timmermann (1992), el cual compara el signo de la proyección \hat{y}_i , con el del valor observado y_i , para cada i -ésima observación del conjunto ($i=1,2,\dots$), donde el signo indica la dirección en que se moverá el mercado accionario: al alza si es positivo, o a la baja, si es negativo. Si los signos coinciden aumenta la efectividad del modelo que se está utilizando, por otro lado, en caso de no existir coincidencia, aumenta el error de predicción del modelo.

El test DA de Pesaran y Timmermann se construye de la siguiente manera:

Primero se define un ratio de éxito (SR):

Donde “ m ” representa el total de datos del conjunto e $I(\cdot)$ es una función que toma el valor 1 cuando su argumento es cierto y 0 en otro caso. Además:

Y

El ratio de éxito en el caso de independencia de \hat{y}_i e y_i , SRI está dado por la siguiente expresión:

Cuya varianza es:

Por otra parte, la varianza de SR queda definida como:

Por último el test DA de Pesaran y Timmermann (1992) queda definido como:

-

Como dicho test sigue una distribución normal estándar, los valores de rechazo de la hipótesis nula serán el -1,96 y el 1,96, considerando un nivel de confianza del 95%, dicho de otra manera, si el valor DA se encuentra entre los valores de rechazo, no se rechaza la hipótesis nula de que las variaciones observaciones están independientemente distribuidas de las variaciones proyectadas.

Anexo 3: tablas de resultados entregados por RapidMiner

3.1 Resultados muestra 70-30%

Resultados regresión lineal

Table View
 Text View
 Annotations

Attribute	Coefficient	Std. Error	Std. Coefficient	Tolerance	t-Stat	p-Value	Code
Retorno DJI-2	-0.182	0.053	-3.804	0.980	-3.407	0.001	****
(Intercept)	0.003	0.001	?	?	3.015	0.003	***
Retorno DJI-1	0.129	0.056	2.735	0.999	2.299	0.025	**
Retorno Ipsa-2	0.095	0.053	0.771	0.934	1.802	0.087	*
Retorno tc-2	-0.118	0.060	3.098	0.967	-1.969	0.058	*
Retorno Ipsa-3	-0.087	0.053	-0.707	0.973	-1.650	0.123	
Retorno Ipsa-1	-0.050	0.052	-0.403	0.940	-0.948	0.352	
Retorno DJI-3	-0.026	0.055	-0.551	0.957	-0.475	0.639	
Retorno tc-1	0.011	0.056	-0.294	1.000	0.198	0.845	

Resultados Red Neuronal

ImprovedNeuralNet		
Hidden 1 ===== Node 1 (Sigmoid) ----- Retorno Ipsa-3: -0.106 Retorno Ipsa-2: -0.029 Retorno Ipsa-1: 0.022 Retorno DJI-3: -0.063 Retorno DJI-2: -0.141 Retorno DJI-1: 0.057 Retorno tc-3: 0.042 Retorno tc-2: 0.006 Retorno tc-1: -0.001 Bias: -0.154 Node 2 (Sigmoid) ----- Retorno Ipsa-3: 0.013 Retorno Ipsa-2: -0.047 Retorno Ipsa-1: 0.022 Retorno DJI-3: -0.084	Retorno DJI-2: -0.088 Retorno DJI-1: -0.003 Retorno tc-3: 0.042 Retorno tc-2: 0.033 Retorno tc-1: -0.009 Bias: -0.136 Node 3 (Sigmoid) ----- Retorno Ipsa-3: -0.081 Retorno Ipsa-2: -0.016 Retorno Ipsa-1: -0.059 Retorno DJI-3: -0.058 Retorno DJI-2: -0.144 Retorno DJI-1: 0.092 Retorno tc-3: 0.047 Retorno tc-2: 0.003 Retorno tc-1: 0.010 Bias: -0.134 Node 4 (Sigmoid) ----- Retorno Ipsa-3: 0.019 Retorno Ipsa-2: -0.044	Retorno Ipsa-1: 0.037 Retorno DJI-3: 0.018 Retorno DJI-2: 0.066 Retorno DJI-1: -0.042 Retorno tc-3: 0.101 Retorno tc-2: 0.053 Retorno tc-1: -0.013 Bias: -0.122 Node 5 (Sigmoid) ----- Retorno Ipsa-3: -0.226 Retorno Ipsa-2: 0.158 Retorno Ipsa-1: -0.071 Retorno DJI-3: -0.159 Retorno DJI-2: -0.382 Retorno DJI-1: 0.303 Retorno tc-3: 0.068 Retorno tc-2: -0.228 Retorno tc-1: 0.042 Bias: -0.130 Node 6 (Sigmoid) ----- Retorno Ipsa-3: -0.222 Retorno Ipsa-2: 0.137

```
Retorno Ipsa-1: -0.043
Retorno DJI-3: -0.139
Retorno DJI-2: -0.342
Retorno DJI-1: 0.264
Retorno tc-3: 0.030
Retorno tc-2: -0.179
Retorno tc-1: 0.007
Bias: -0.160
```

```
Output
=====
```

```
Regression (Linear)
```

```
-----
Node 1: 0.078
Node 2: -0.039
Node 3: 0.101
Node 4: -0.174
Node 5: 0.540
Node 6: 0.465
Threshold: -0.310
```

Resultados modelos SVM

Kernel Model

```
Total number of Support Vectors: 523
Bias (offset): 0.005
```

```
w[Retorno Ipsa-3] = 0.000
w[Retorno Ipsa-2] = 0.002
w[Retorno Ipsa-1] = -0.002
w[Retorno DJI-3] = 0.000
w[Retorno DJI-2] = -0.003
w[Retorno DJI-1] = 0.004
w[Retorno tc-3] = -0.000
w[Retorno tc-2] = -0.000
w[Retorno tc-1] = 0.000
```

3.2 Resultados muestra 20-80%

Resultados regresión lineal

prediction_trend_accuracy

prediction_trend_accuracy: 0.748

Table View Text View Annotations

Attribute	Coefficient	Std. Error	Std. Coeffici...	Tolerance	t-Stat	p-Value	Code
Retorno DJI-2	-0.182	0.053	-3.804	0.980	-3.407	0.001	****
(Intercept)	0.003	0.001	?	?	3.015	0.003	***
Retorno DJI-1	0.129	0.056	2.735	0.999	2.299	0.025	**
Retorno Ipsa-2	0.095	0.053	0.771	0.934	1.802	0.087	*
Retorno tc-2	-0.118	0.060	3.097	0.967	-1.969	0.058	*
Retorno Ipsa-3	-0.087	0.053	-0.706	0.973	-1.649	0.123	
Retorno Ipsa-1	-0.050	0.052	-0.403	0.939	-0.948	0.352	
Retorno DJI-3	-0.026	0.055	-0.549	0.957	-0.475	0.640	
Retorno tc-1	0.011	0.056	-0.294	1.000	0.199	0.844	

Resultados Red Neuronal

prediction_trend_accuracy

prediction_trend_accuracy: 0.739

ImprovedNeuralNet

Hidden 1

=====

Node 1 (Sigmoid)

Retorno Ipsa-3: -0.252
 Retorno Ipsa-2: 0.121
 Retorno Ipsa-1: -0.101
 Retorno DJI-3: -0.157
 Retorno DJI-2: -0.360
 Retorno DJI-1: 0.287
 Retorno tc-3: 0.021
 Retorno tc-2: -0.153
 Retorno tc-1: 0.106
 Bias: -0.122

Node 2 (Sigmoid)

Retorno Ipsa-3: 0.048
 Retorno Ipsa-2: -0.127
 Retorno Ipsa-1: -0.001
 Retorno DJI-3: -0.020

Retorno DJI-2: 0.126
 Retorno DJI-1: -0.173
 Retorno tc-3: 0.039
 Retorno tc-2: 0.129
 Retorno tc-1: -0.002
 Bias: -0.103

Node 3 (Sigmoid)

Retorno Ipsa-3: -0.081
 Retorno Ipsa-2: 0.038
 Retorno Ipsa-1: -0.054
 Retorno DJI-3: -0.005
 Retorno DJI-2: -0.039
 Retorno DJI-1: -0.025
 Retorno tc-3: 0.052
 Retorno tc-2: 0.013
 Retorno tc-1: 0.026
 Bias: -0.161

Node 4 (Sigmoid)

Retorno Ipsa-3: -0.055
 Retorno Ipsa-2: -0.003
 Retorno Ipsa-1: -0.020
 Retorno DJI-3: -0.030

Retorno DJI-2: -0.110
 Retorno DJI-1: -0.016
 Retorno tc-3: 0.050
 Retorno tc-2: -0.029
 Retorno tc-1: 0.020
 Bias: -0.143

Node 5 (Sigmoid)

Retorno Ipsa-3: -0.062
 Retorno Ipsa-2: 0.051
 Retorno Ipsa-1: -0.080
 Retorno DJI-3: -0.075
 Retorno DJI-2: -0.146
 Retorno DJI-1: 0.131
 Retorno tc-3: -0.009
 Retorno tc-2: -0.070
 Retorno tc-1: 0.049
 Bias: -0.131

Node 6 (Sigmoid)

Retorno Ipsa-3: -0.284
 Retorno Ipsa-2: 0.138
 Retorno Ipsa-1: -0.037
 Retorno DJI-3: -0.101

```
Retorno DJI-2: -0.353
Retorno DJI-1: 0.272
Retorno tc-3: -0.003
Retorno tc-2: -0.111
Retorno tc-1: 0.049
Bias: -0.110
```

```
Output
=====
```

```
Regression (Linear)
```

```
-----
```

```
Node 1: 0.490
Node 2: -0.321
Node 3: -0.018
Node 4: 0.014
Node 5: 0.155
Node 6: 0.473
Threshold: -0.265
```

Resultados modelos SVM

prediction_trend_accuracy

```
prediction_trend_accuracy: 0.722
```

Kernel Model

```
Total number of Support Vectors: 523
Bias (offset): 0.005
```

```
w[Retorno Ipsa-3] = 0.000
w[Retorno Ipsa-2] = 0.001
w[Retorno Ipsa-1] = -0.002
w[Retorno DJI-3] = 0.000
w[Retorno DJI-2] = -0.003
w[Retorno DJI-1] = 0.003
w[Retorno tc-3] = -0.001
w[Retorno tc-2] = -0.000
w[Retorno tc-1] = -0.000
```

3.3 Resultados muestra 30-70%

Resultados regresión lineal

prediction_trend_accuracy

prediction_trend_accuracy: 0.730

Table View Text View Annotations

Attribute	Coefficient	Std. Error	Std. Coeffici...	Tolerance	t-Stat	p-Value	Code ▼
Retorno DJI-2	-0.182	0.053	-3.804	0.980	-3.407	0.001	****
(Intercept)	0.003	0.001	?	?	3.015	0.003	***
Retorno DJI-1	0.129	0.056	2.735	0.999	2.299	0.025	**
Retorno Ipsa-2	0.095	0.053	0.771	0.934	1.802	0.087	*
Retorno tc-2	-0.118	0.060	3.098	0.967	-1.969	0.058	*
Retorno Ipsa-3	-0.087	0.053	-0.707	0.973	-1.650	0.123	
Retorno Ipsa-1	-0.050	0.052	-0.403	0.939	-0.948	0.352	
Retorno DJI-3	-0.026	0.055	-0.550	0.957	-0.475	0.639	
Retorno tc-1	0.011	0.056	-0.294	1.000	0.199	0.844	

Resultados Red Neuronal

prediction_trend_accuracy

prediction_trend_accuracy: 0.725

ImprovedNeuralNet	Retorno DJI-2: -0.281	Retorno DJI-2: -0.422
Hidden 1 ===== Node 1 (Sigmoid) ----- Retorno Ipsa-3: -0.055 Retorno Ipsa-2: -0.000 Retorno Ipsa-1: -0.050 Retorno DJI-3: -0.030 Retorno DJI-2: -0.197 Retorno DJI-1: 0.108 Retorno tc-3: 0.006 Retorno tc-2: 0.022 Retorno tc-1: 0.040 Bias: -0.120 Node 2 (Sigmoid) ----- Retorno Ipsa-3: -0.128 Retorno Ipsa-2: 0.071 Retorno Ipsa-1: -0.047 Retorno DJI-3: -0.077	Retorno DJI-1: 0.196 Retorno tc-3: -0.005 Retorno tc-2: -0.137 Retorno tc-1: 0.057 Bias: -0.086 Node 3 (Sigmoid) ----- Retorno Ipsa-3: -0.164 Retorno Ipsa-2: 0.046 Retorno Ipsa-1: -0.032 Retorno DJI-3: -0.102 Retorno DJI-2: -0.236 Retorno DJI-1: 0.116 Retorno tc-3: 0.010 Retorno tc-2: -0.088 Retorno tc-1: 0.078 Bias: -0.112 Node 4 (Sigmoid) ----- Retorno Ipsa-3: -0.245 Retorno Ipsa-2: 0.249 Retorno Ipsa-1: -0.133 Retorno DJI-3: -0.139	Retorno DJI-1: 0.301 Retorno tc-3: 0.023 Retorno tc-2: -0.226 Retorno tc-1: 0.056 Bias: -0.141 Node 5 (Sigmoid) ----- Retorno Ipsa-3: -0.058 Retorno Ipsa-2: -0.019 Retorno Ipsa-1: -0.026 Retorno DJI-3: -0.018 Retorno DJI-2: -0.080 Retorno DJI-1: 0.034 Retorno tc-3: 0.025 Retorno tc-2: 0.070 Retorno tc-1: 0.046 Bias: -0.110 Node 6 (Sigmoid) ----- Retorno Ipsa-3: 0.088 Retorno Ipsa-2: -0.147 Retorno Ipsa-1: 0.023 Retorno DJI-3: 0.030

```
Retorno DJI-2: 0.173
Retorno DJI-1: -0.231
Retorno tc-3: 0.062
Retorno tc-2: 0.143
Retorno tc-1: 0.035
Bias: -0.137
```

```
Output
=====
```

```
Regression (Linear)
```

```
-----
```

```
Node 1: 0.122
Node 2: 0.316
Node 3: 0.234
Node 4: 0.595
Node 5: -0.014
Node 6: -0.408
Threshold: -0.283
```

Resultados modelos SVM

prediction_trend_accuracy

```
prediction_trend_accuracy: 0.749
```

Kernel Model

```
Total number of Support Vectors: 523
```

```
Bias (offset): 0.005
```

```
w[Retorno Ipsa-3] = 0.000
w[Retorno Ipsa-2] = 0.001
w[Retorno Ipsa-1] = -0.002
w[Retorno DJI-3] = 0.000
w[Retorno DJI-2] = -0.003
w[Retorno DJI-1] = 0.003
w[Retorno tc-3] = -0.000
w[Retorno tc-2] = -0.000
w[Retorno tc-1] = -0.000
```

3.4 Resultados muestra 50-50%

Resultados regresión lineal

prediction_trend_accuracy

prediction_trend_accuracy: 0.738

Table View Text View Annotations

Attribute	Coefficient	Std. Error	Std. Coeffi...	Tolerance	t-Stat	p-Value	Code ▾
Retorno DJI-2	-0.182	0.053	-3.804	0.980	-3.407	0.001	****
(Intercept)	0.003	0.001	?	?	3.015	0.003	***
Retorno DJI-1	0.129	0.056	2.735	0.999	2.299	0.025	**
Retorno Ipsa-2	0.095	0.053	0.771	0.934	1.802	0.087	*
Retorno tc-2	-0.118	0.060	3.098	0.967	-1.969	0.058	*
Retorno Ipsa-3	-0.087	0.053	-0.707	0.973	-1.650	0.123	
Retorno Ipsa-1	-0.050	0.052	-0.403	0.939	-0.948	0.352	
Retorno DJI-3	-0.026	0.055	-0.550	0.957	-0.475	0.640	
Retorno tc-1	0.011	0.056	-0.294	1.000	0.199	0.845	

Resultados Red Neuronal

prediction_trend_accuracy

prediction_trend_accuracy: 0.731

ImprovedNeuralNet		
Hidden 1 ===== Node 1 (Sigmoid) ----- Retorno Ipsa-3: -0.225 Retorno Ipsa-2: 0.111 Retorno Ipsa-1: -0.027 Retorno DJI-3: -0.083 Retorno DJI-2: -0.352 Retorno DJI-1: 0.140 Retorno tc-3: 0.038 Retorno tc-2: -0.105 Retorno tc-1: 0.033 Bias: -0.116 Node 2 (Sigmoid) ----- Retorno Ipsa-3: 0.015 Retorno Ipsa-2: -0.018 Retorno Ipsa-1: 0.016 Retorno DJI-3: -0.027	Retorno DJI-2: -0.020 Retorno DJI-1: -0.078 Retorno tc-3: 0.110 Retorno tc-2: 0.059 Retorno tc-1: 0.027 Bias: -0.124 Node 3 (Sigmoid) ----- Retorno Ipsa-3: -0.077 Retorno Ipsa-2: 0.059 Retorno Ipsa-1: 0.000 Retorno DJI-3: -0.093 Retorno DJI-2: -0.126 Retorno DJI-1: -0.008 Retorno tc-3: 0.102 Retorno tc-2: -0.013 Retorno tc-1: 0.037 Bias: -0.084 Node 4 (Sigmoid) ----- Retorno Ipsa-3: -0.322 Retorno Ipsa-2: 0.179 Retorno Ipsa-1: -0.010 Retorno DJI-3: -0.147	Retorno DJI-2: -0.388 Retorno DJI-1: 0.271 Retorno tc-3: 0.025 Retorno tc-2: -0.164 Retorno tc-1: 0.056 Bias: -0.137 Node 5 (Sigmoid) ----- Retorno Ipsa-3: 0.023 Retorno Ipsa-2: -0.018 Retorno Ipsa-1: -0.012 Retorno DJI-3: -0.026 Retorno DJI-2: 0.004 Retorno DJI-1: -0.065 Retorno tc-3: 0.074 Retorno tc-2: 0.021 Retorno tc-1: 0.002 Bias: -0.163 Node 6 (Sigmoid) ----- Retorno Ipsa-3: -0.052 Retorno Ipsa-2: 0.015 Retorno Ipsa-1: -0.057 Retorno DJI-3: -0.091

```
Retorno DJI-2: -0.108
Retorno DJI-1: 0.022
Retorno tc-3: 0.065
Retorno tc-2: -0.009
Retorno tc-1: 0.011
Bias: -0.135
```

```
Output
=====
```

```
Regression (Linear)
```

```
-----
Node 1: 0.389
Node 2: -0.108
Node 3: 0.076
Node 4: 0.560
Node 5: -0.111
Node 6: 0.054
Threshold: -0.262
```

Resultados modelos SVM

prediction_trend_accuracy

prediction_trend_accuracy: 0.758

Kernel Model

Total number of Support Vectors: 523
Bias (offset): 0.005

```
w[Retorno Ipsa-3] = 0.000
w[Retorno Ipsa-2] = 0.001
w[Retorno Ipsa-1] = -0.002
w[Retorno DJI-3] = 0.000
w[Retorno DJI-2] = -0.003
w[Retorno DJI-1] = 0.003
w[Retorno tc-3] = -0.001
w[Retorno tc-2] = -0.000
w[Retorno tc-1] = -0.000
```