

Capítulo 5

Resultados experimentales

Un conjunto de experimentos fueron ejecutados usando series de tiempo financieras descargadas de Yahoo! Finance. Se aplicaron dos versiones del modelo predictivo para analizar y predecir series de tiempo llevando a cabo una comparación de estos métodos predictivos.

La predicción es aplicada a series de tiempo dentro del sector financiero, específicamente datos de compañías de la bolsa de valores de Estados Unidos. Un archivo de texto que contiene el precio de la acción diaria por cada empresa ha sido descargada de la página web Yahoo! Finance donde el precio indicado es el precio del cierre de cada día. La longitud de la serie (número de días de negociación) es variable y depende de la historia de la compañía.

5.1 Resultados de predicción basada en representación esparsiva con diccionarios entrenados

El primer conjunto de experimentos incluye los resultados de predicción del modelo propuesto basado en diccionarios entrenados. Se ha probado un conjunto de empresas con diferentes parámetros del modelo. El modelo fue probado con toda la actividad de la empresa. La longitud de los diccionarios es variada para cada empresa desde 450 hasta 4500 átomos, y el número promedio de átomos de diferentes simulaciones de empresas es de 1850 átomos. La dimensión del átomo y un vector característica es de 50, y el tamaño del conjunto de señales de entrenamiento para entrenar un diccionario es 100 y el factor de esparsividad es 3.

El objetivo de un predictor artificial es hacer las predicciones más precisas que el resultado de una curva de ganancia de retorno acumulado sea más alta que la curva que proporciona el mercado. Se considera que una unidad monetaria es la entrada en el comienzo del análisis a predecir.

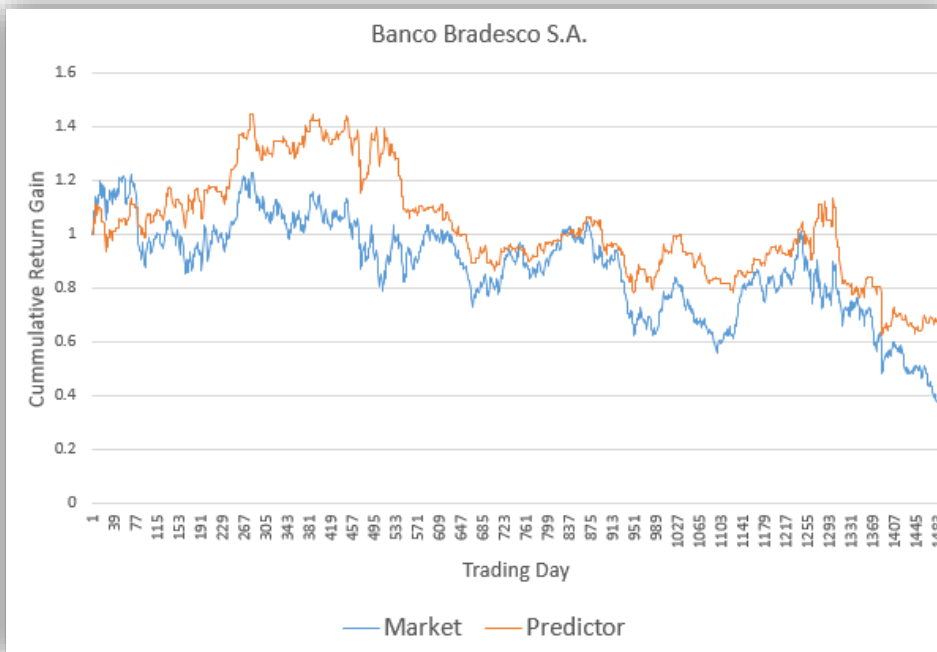
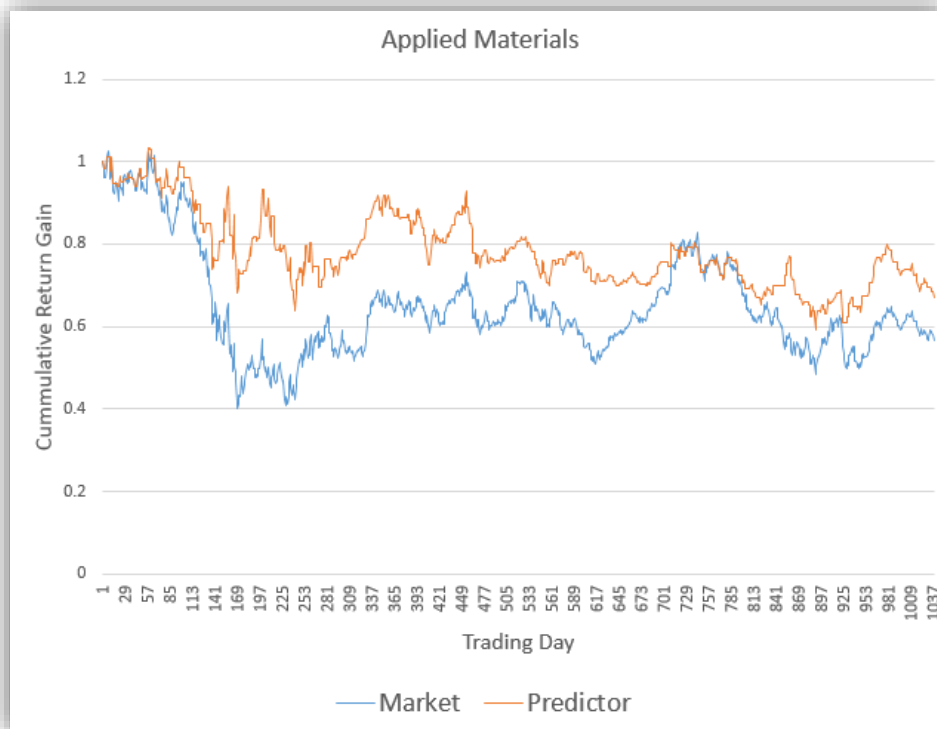
La figura 4 muestra algunos ejemplos de la óptima predicción conseguida por el modelo cuando se aplica para analizar las acciones de diferentes empresas. Se puede observar que la

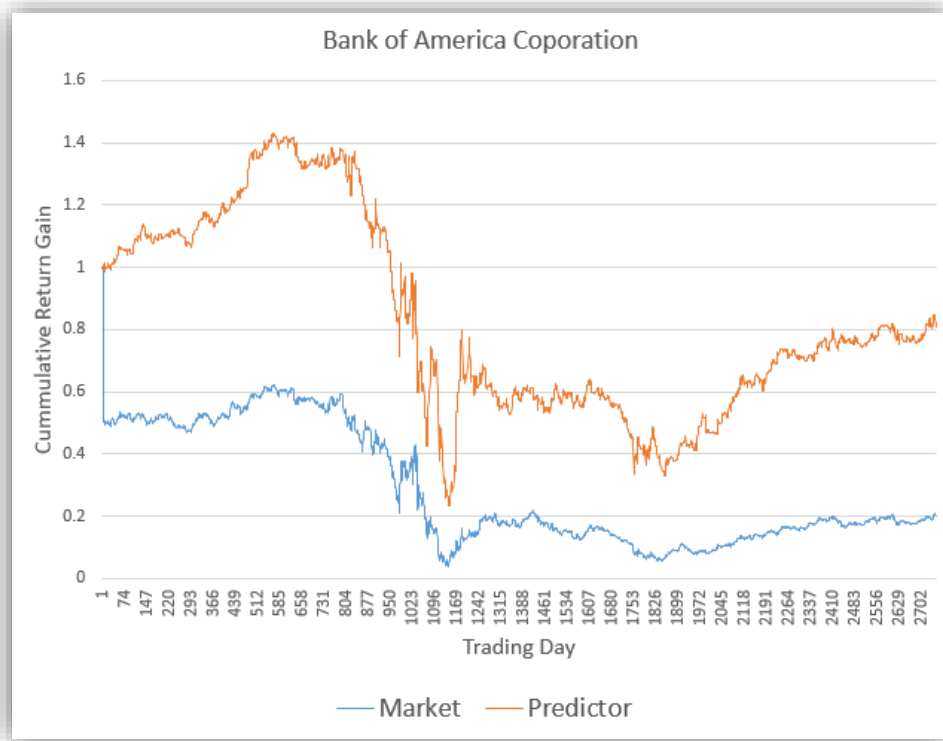
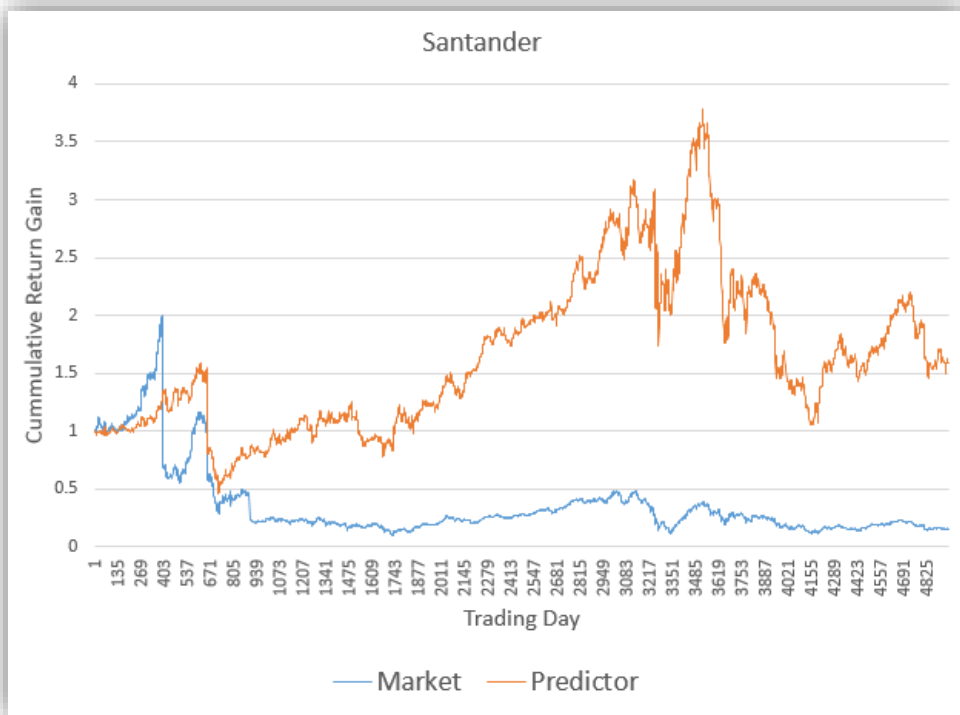
ganancia del retorno acumulado del sistema de predicción (naranja) por lo general supera al mercado (azul).

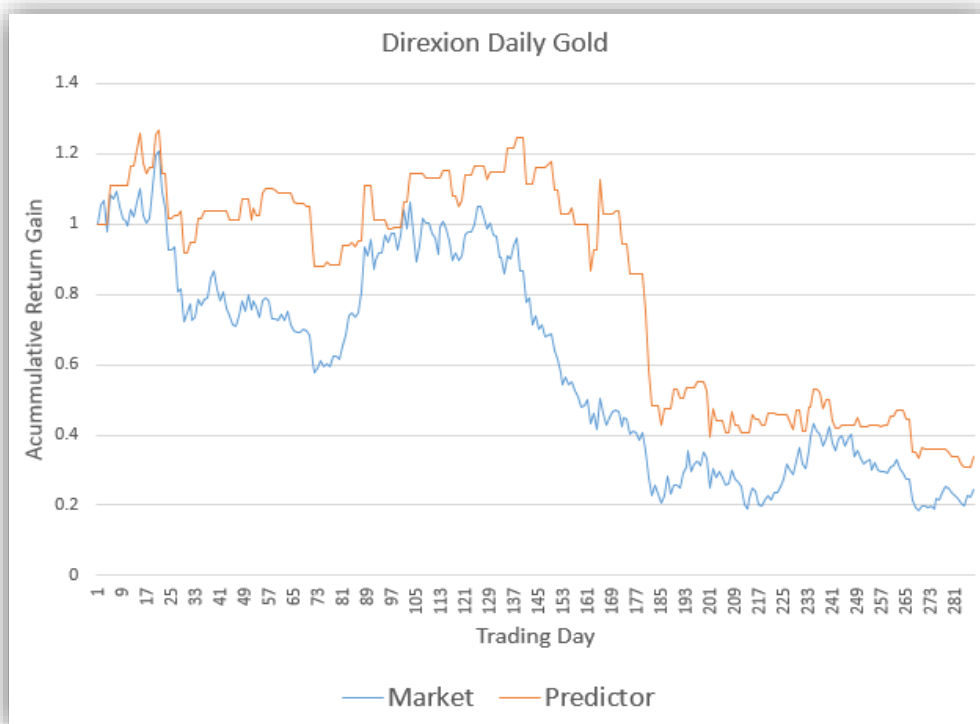
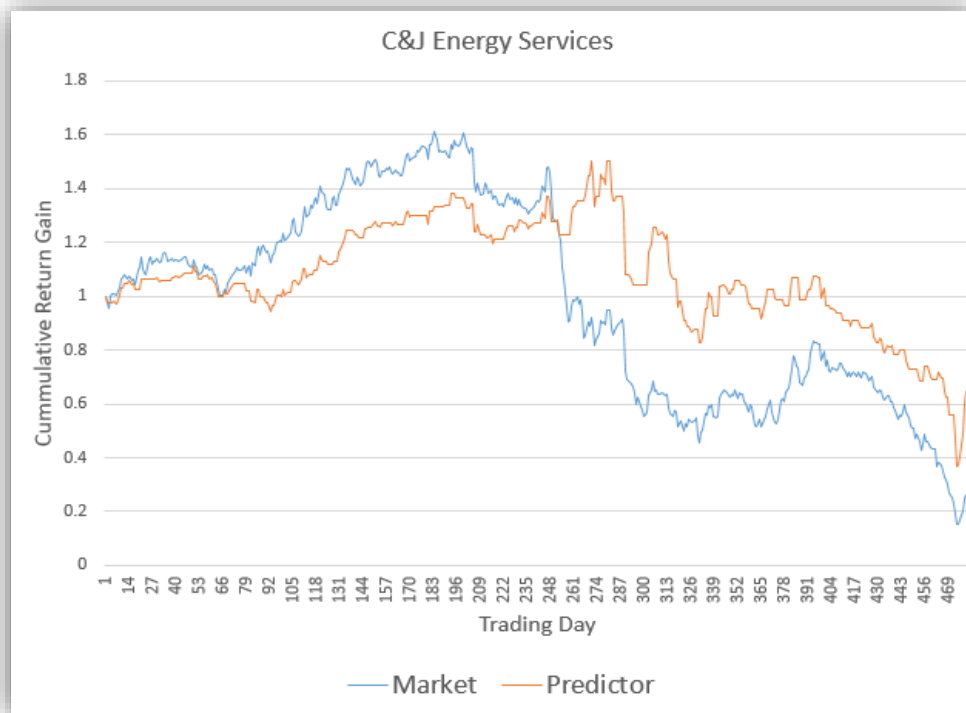
La ganancia de retorno acumulado es una serie geométrica $G = \prod_{i=1}^{current\ trading\ day} (1 + R_i)$.

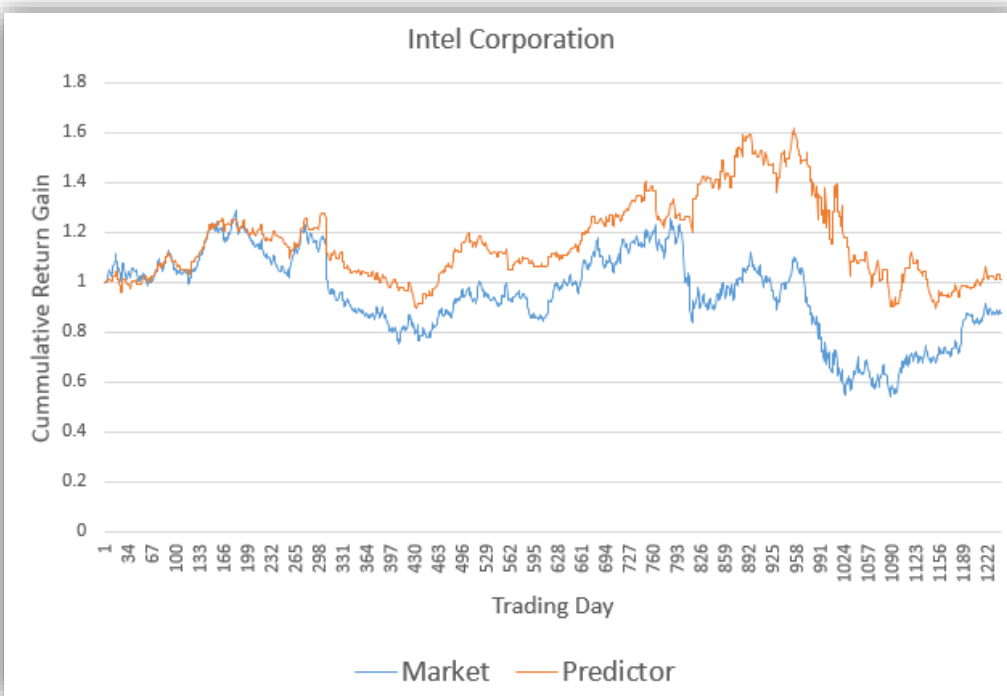
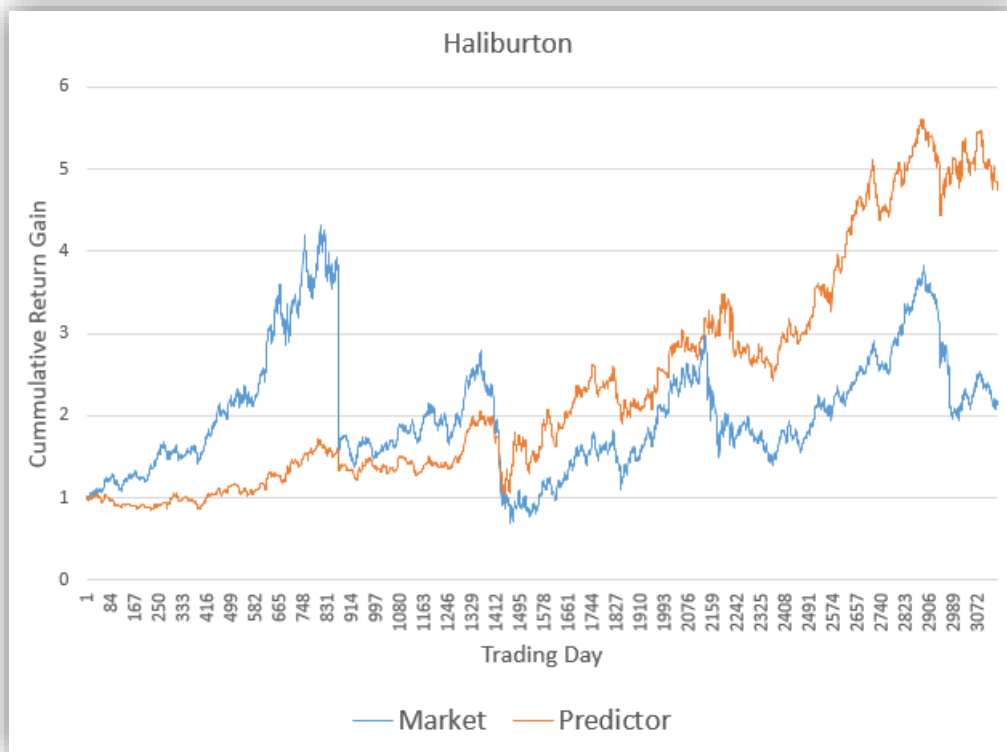
Una vez que esta serie alcanza un valor más grande que uno, que podría aumentar de forma exponencial, siempre y cuando si se multiplica por factores mayores que uno, $(1 + R_i) > 1$.

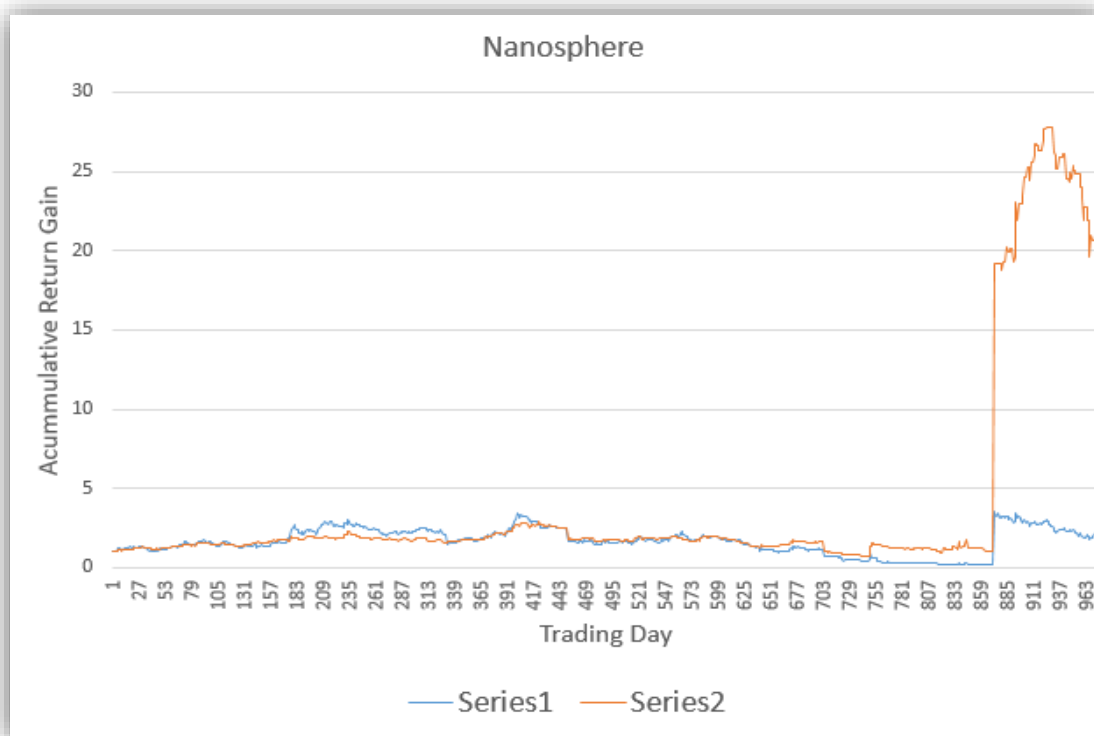
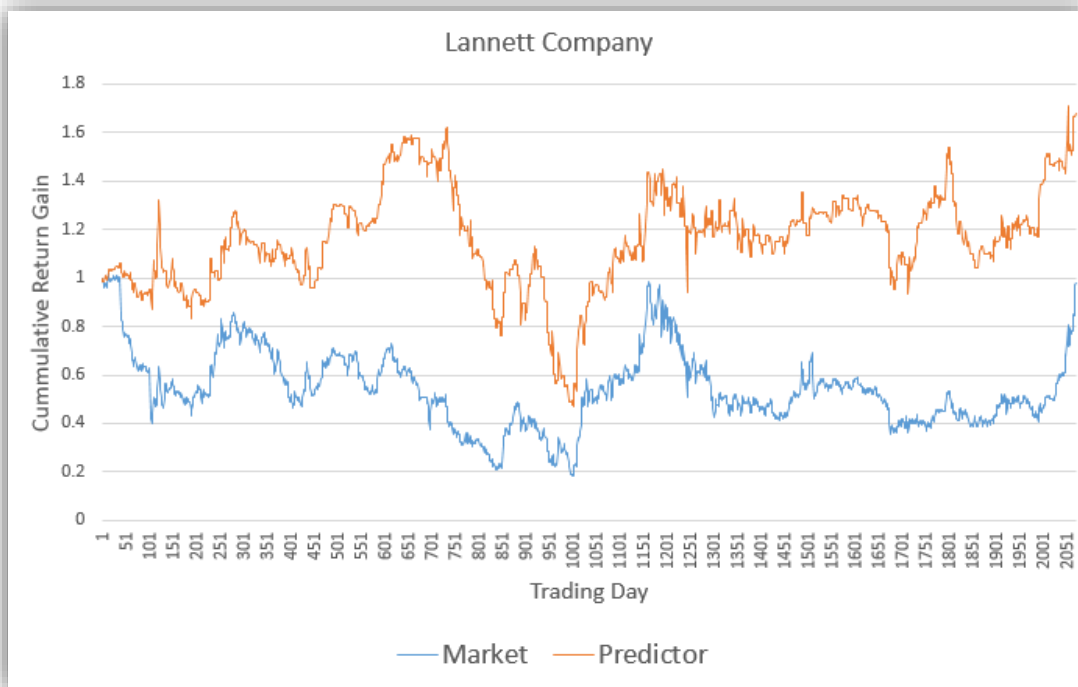
Cuando la serie geométrica está en un valor más bajo que uno es difícil para que se inicie un rápido aumento y se observa como una señal constante. Esta es la razón del porque hay algunos casos donde la ganancia del predictor crece exponencialmente mientras la ganancia del predictor se mantiene estable o incluso decrece tal y como se muestra en las curvas de Santander Bank iniciando con un entrenamiento del día 784 y Nokia iniciando con un entrenamiento del día 2821. Estos son casos donde el predictor artificial logra un buen desempeño en términos de vencer al mercado.

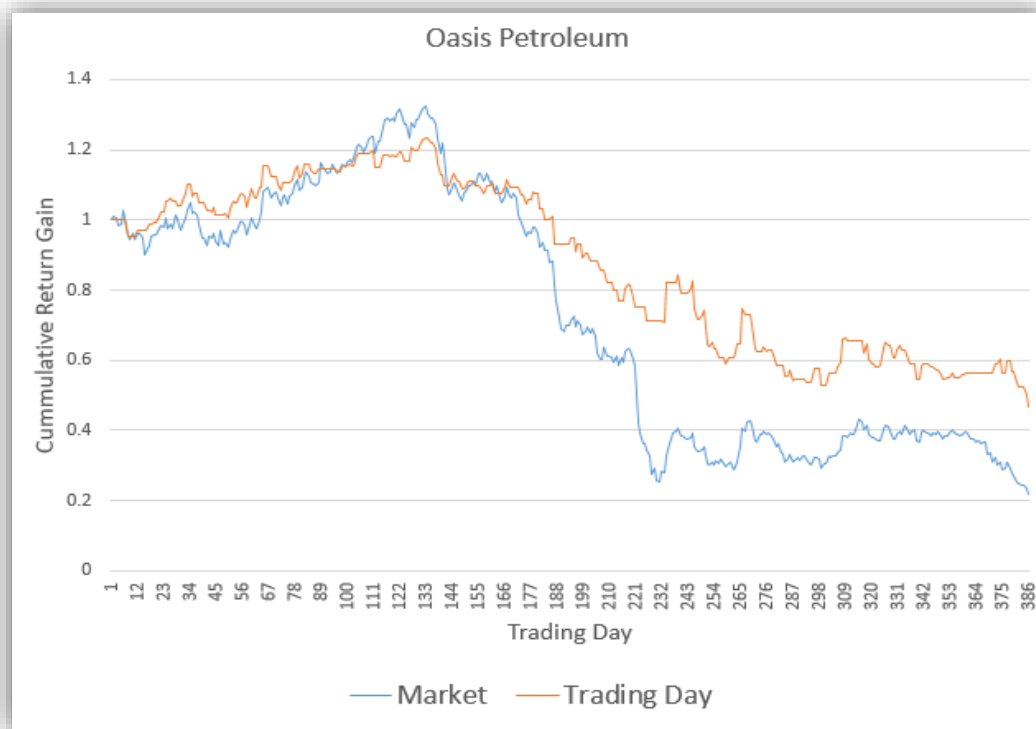
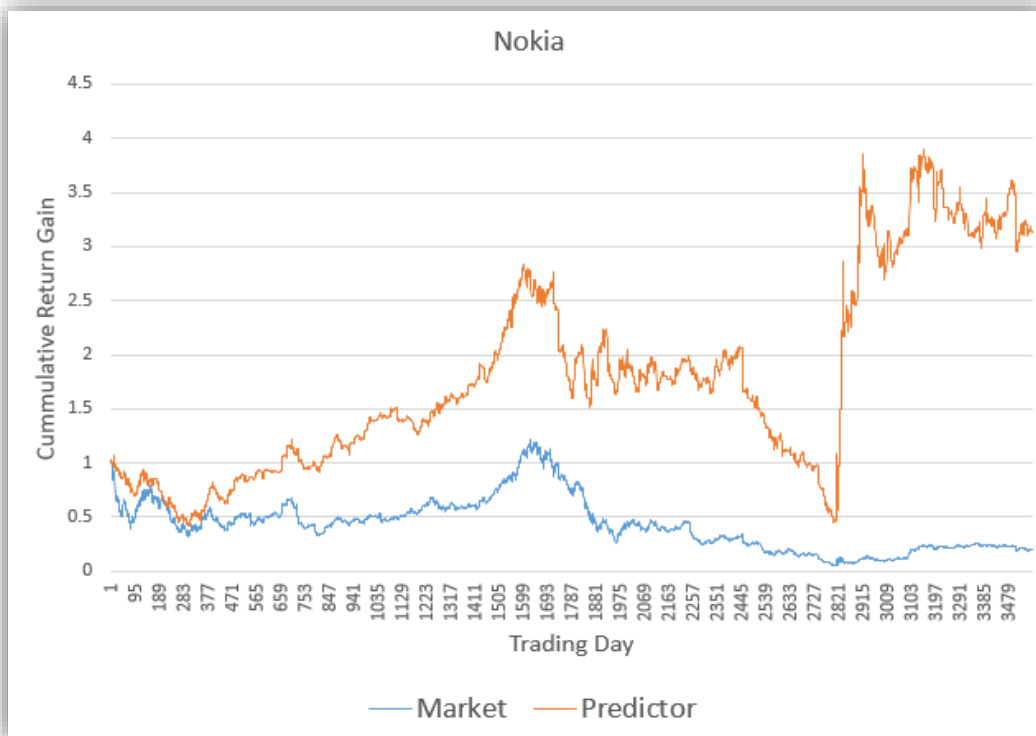


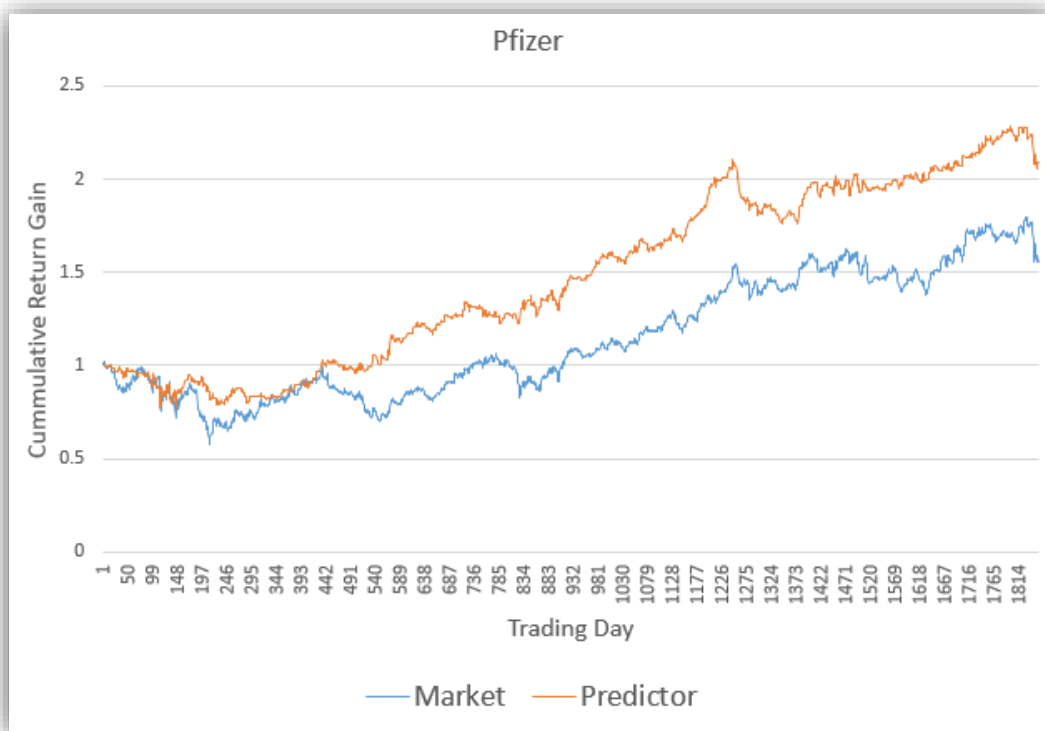
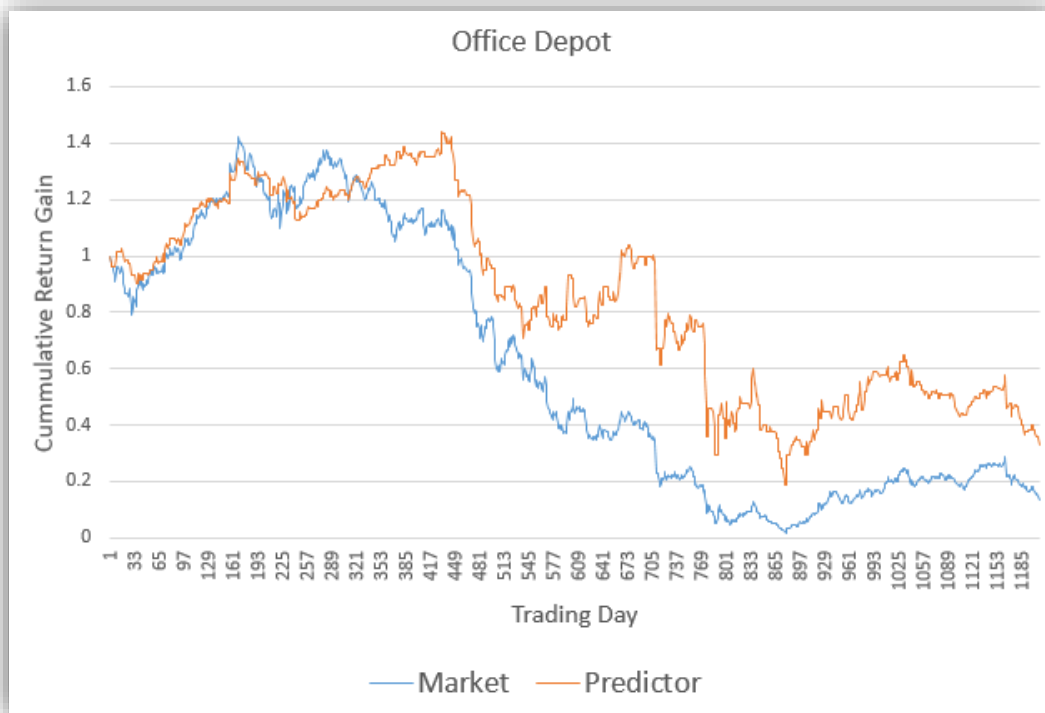


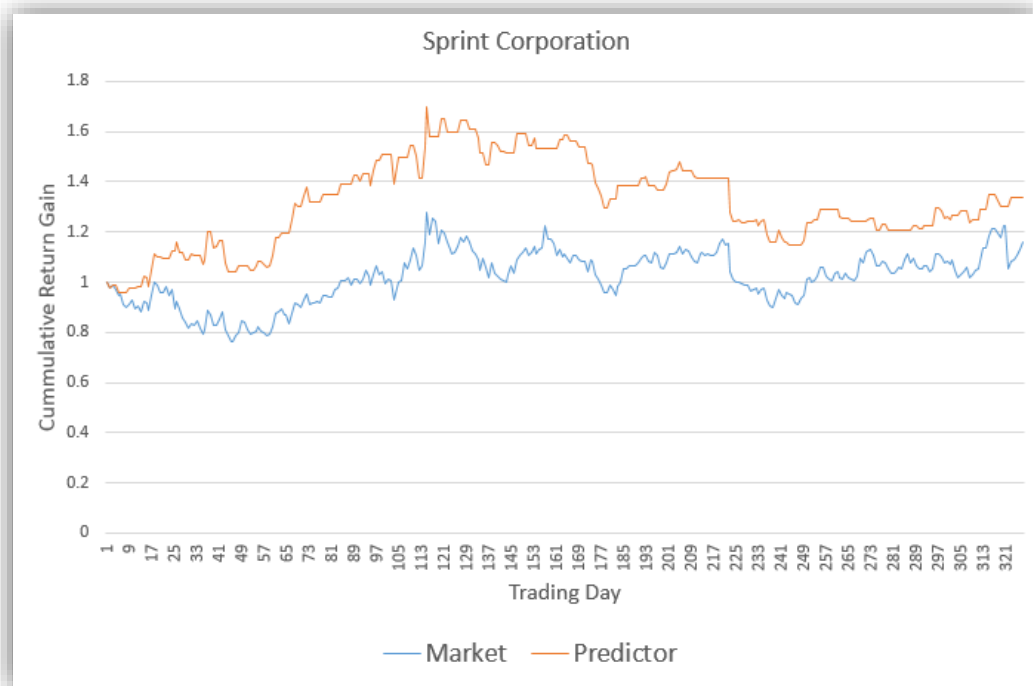
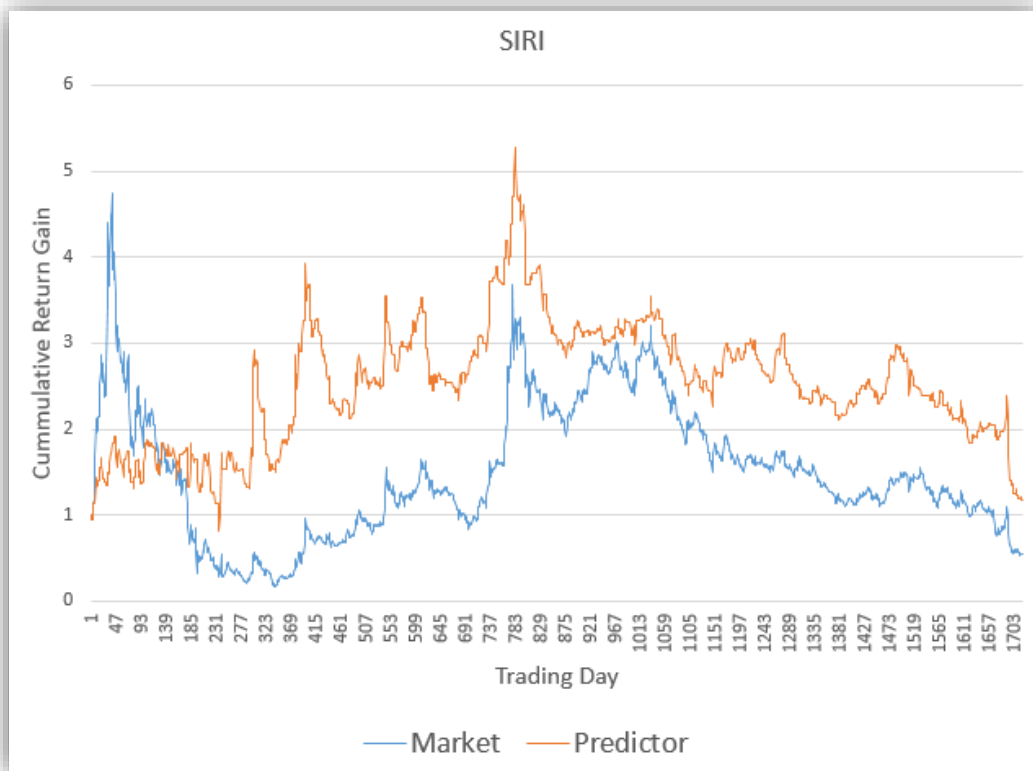


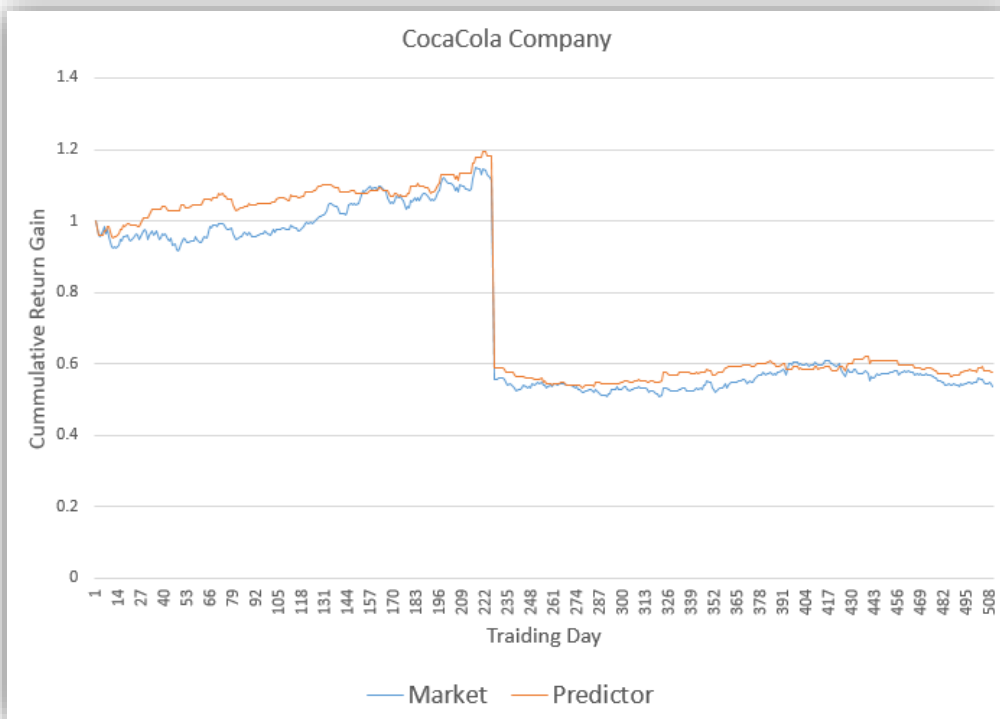
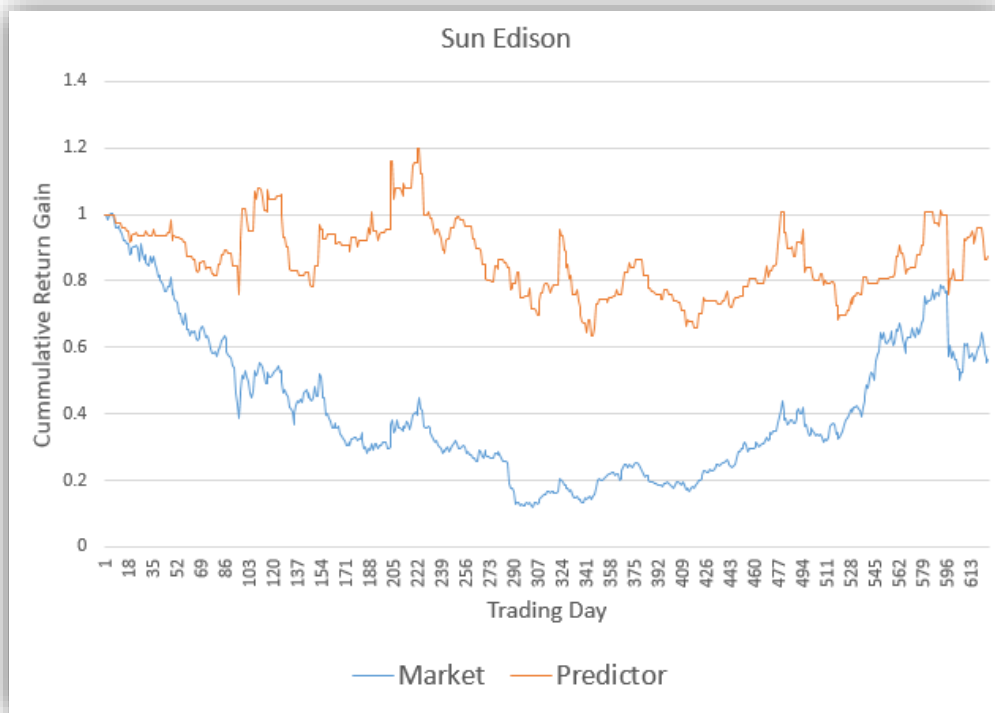












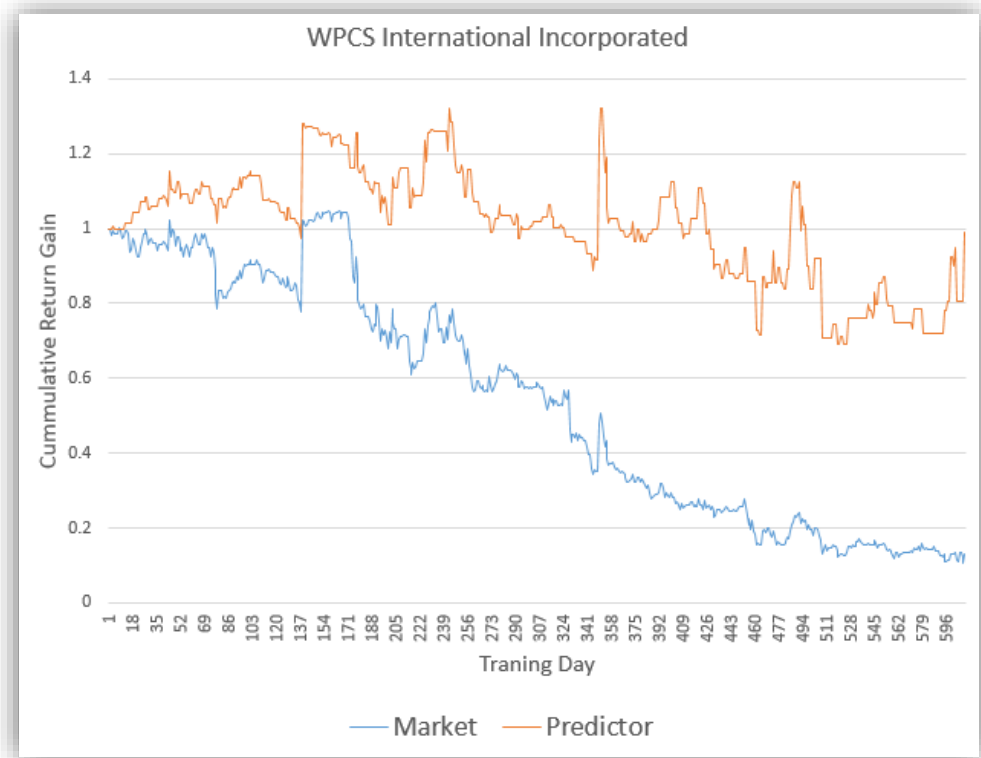
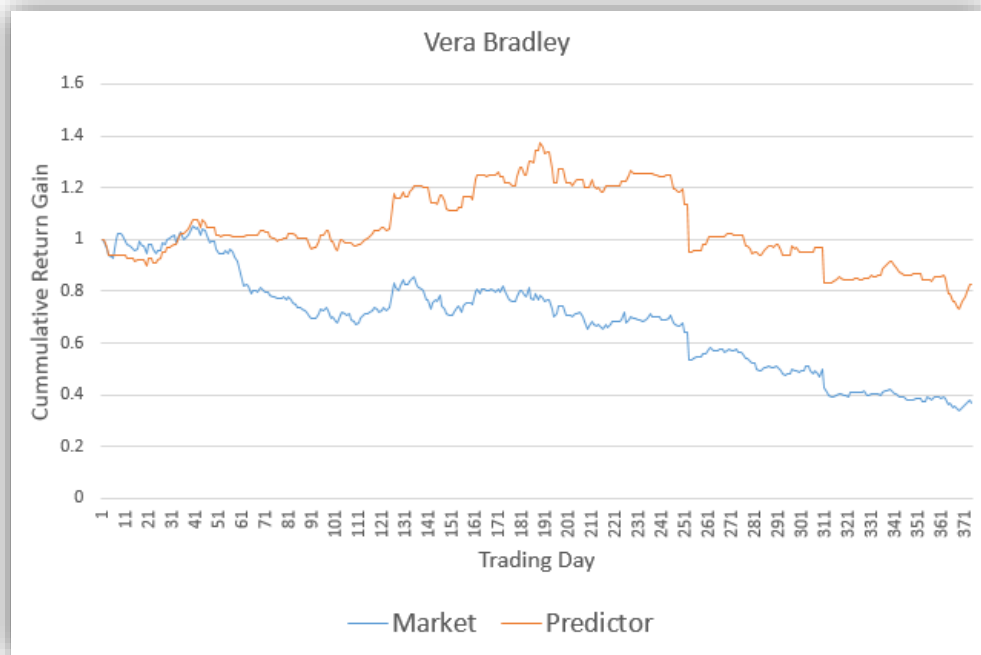


Figura 4 Ganancia de retorno acumulado del modelo predictivo basado en diccionarios entrenados vs la curva generada por el mercado.

La tabla 1 muestra un análisis cuantitativo del modelo predictivo basado en diccionarios entrenados. Una métrica eficaz es la tasa de predicción exitosa (RSP) que es definida como el ratio de esos días donde el predictor vence al mercado sobre el número total de días bajo la predicción. Otras métricas son generadas calculando la diferencia entre dos señales, la ganancia de retorno del predictor x_p (señal anaranjada) menos la ganancia de retorno del mercado x_m (señal azul). Métricas asociadas al histograma de la señal diferencial $x_{diff} = x_p - x_m$ son la diferencia promedio entre la ganancia del predictor y la ganancia del mercado $\overline{x_{diff}}$ y el valor máximo de la señal diferencial $\max_{trading\ day} x_{diff}(trading\ day)$. El rendimiento de la métrica final es el número de átomos por diccionario. Para el caso de diccionarios entrenados el número de átomos necesita ser grande para que el predictor pueda vencer al mercado mientras que los diccionarios no entrenados se caracterizan por tener diccionarios con un número de átomos más pequeño.

Hay dos casos donde el éxito de la tasa de predicción es baja, Halliburton Company con RSP=0.547627 y Nanosphere con RSP=0.620513; sin embargo Nanosphere es caracterizado también por los valores más altos para los parámetros de máxima diferencia (25.16187) y la diferencia media (2.322425). Los mejores valores se muestran en negritas.

Tabla 1 Análisis cuantitativo del modelo predictivo basado en diccionarios entrenados

	Con entrenamiento				Number of atoms
	number of trading days	Successful Prediction Rate	Average Difference	Maximum Difference	
Applied Materials	1037	0.933462	0.134551	0.411098	1800
Banco Bradesco	1493	0.931011	0.147604	0.529507	1700
Banco Santander	4949	0.925844	1.279362	3.406631	1900
Bank of America Corporation	2752	0.998547	0.54079	0.884993	4500
C&J Energy Services	481	0.482328	0.093648	0.607159	450
Direxion Daily Gold	287	0.989547	0.220402	0.623338	700
Haliburton Company	3139	0.547627	0.334887	3.128864	1800
Intel Corporation	1240	0.929839	0.209619	0.719994	2000
Lannett Company	2070	0.999034	0.62041	1.149219	1900
Nanosphere	975	0.620513	2.322425	25.16187	900
Nokia	3570	0.99944	1.351609	3.730181	1700
Oasis Petroleum	386	0.831606	0.14297	0.494709	800
Office Depot	1207	0.882353	0.237358	0.648902	3800
Pfizer	1852	0.967063	0.329787	0.638791	1800
SIRI	1718	0.927241	1.024333	2.965039	1700
Sprint Corporation	327	0.990826	0.295772	0.521129	1600
Sun Edison	625	0.992	0.449024	0.794063	1850
The CocaCola Company	510	0.907843	0.037209	0.110866	2500
Vera Bradley	374	0.919786	0.356279	0.588402	750
WPCS International Incorporated	609	0.995074	0.466195	0.910921	1700

5.2 Resultados de predicción basada en representación esparsiva con diccionarios sin entrenamiento.

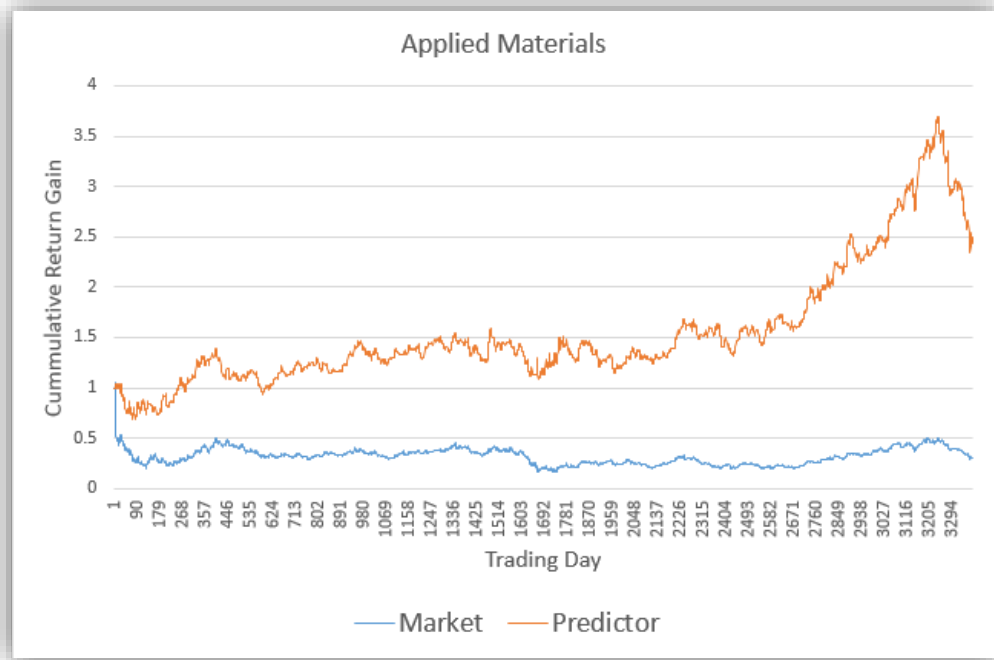
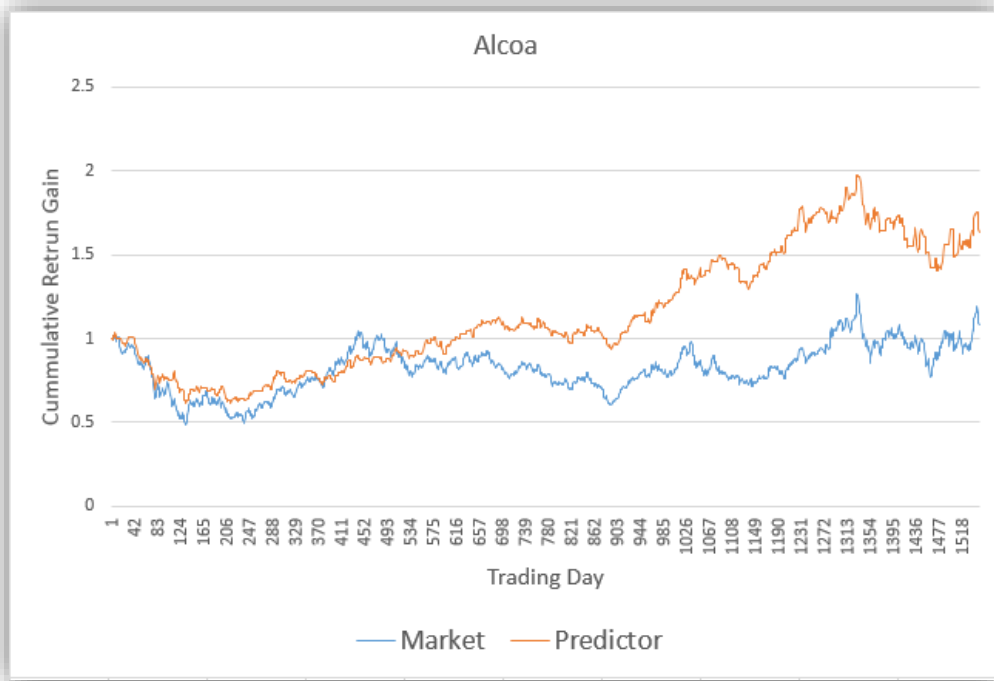
La figura 5 muestra algunos resultados de experimentos que son realizados con 21 compañías para aplicar el modelo predictivo basado en diccionarios sin entrenamiento. En este conjunto de experimentos los pasos 3 y 4 descritos en la sección 4 son ignorados.

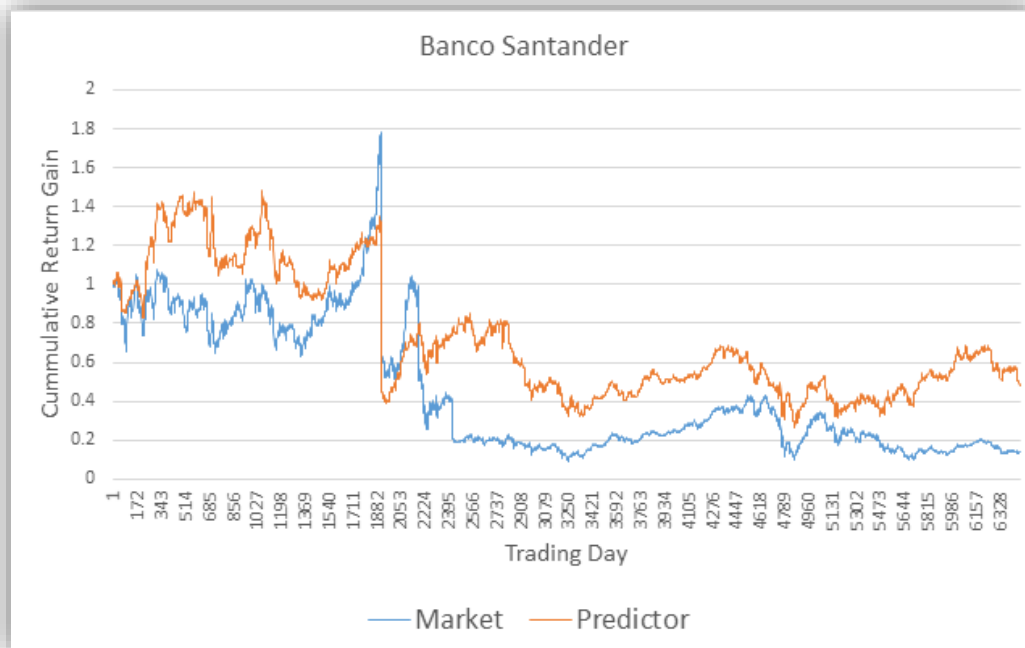
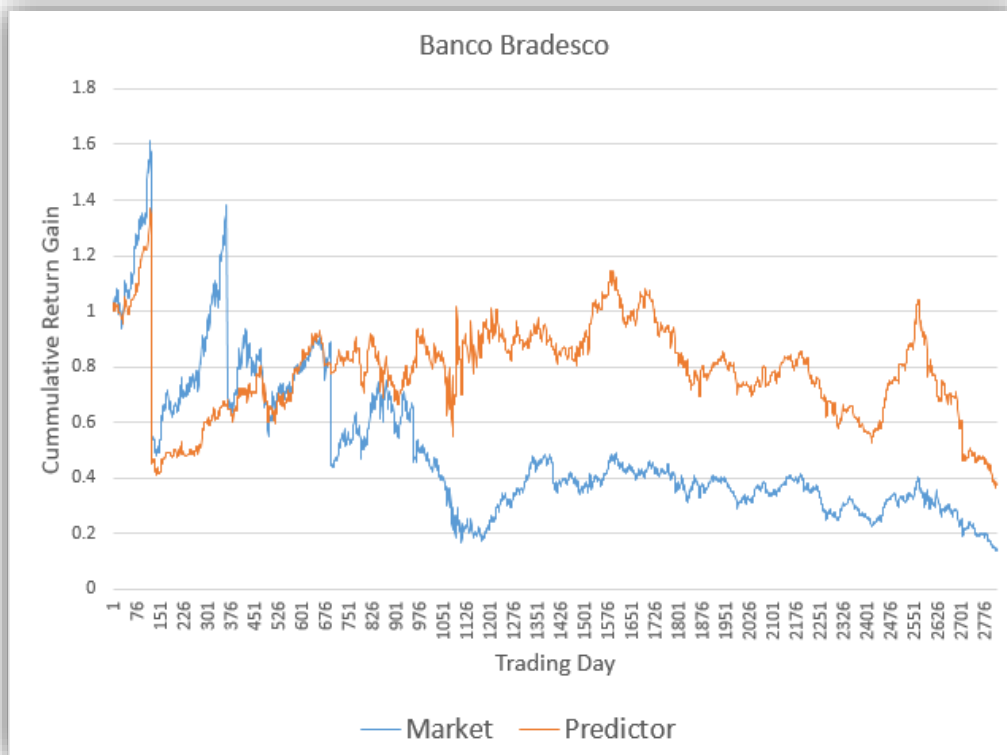
- El modelo se aplica a los datos de entrenamiento con diferentes valores para los parámetros del modelo:
- El tamaño del diccionario se ajusta a un número distinto de átomos (300, 350, 380, 400, 450).
- La dimensionalidad de cada átomo es 50.
- No hay ningún conjunto de señales para adaptar un diccionario.

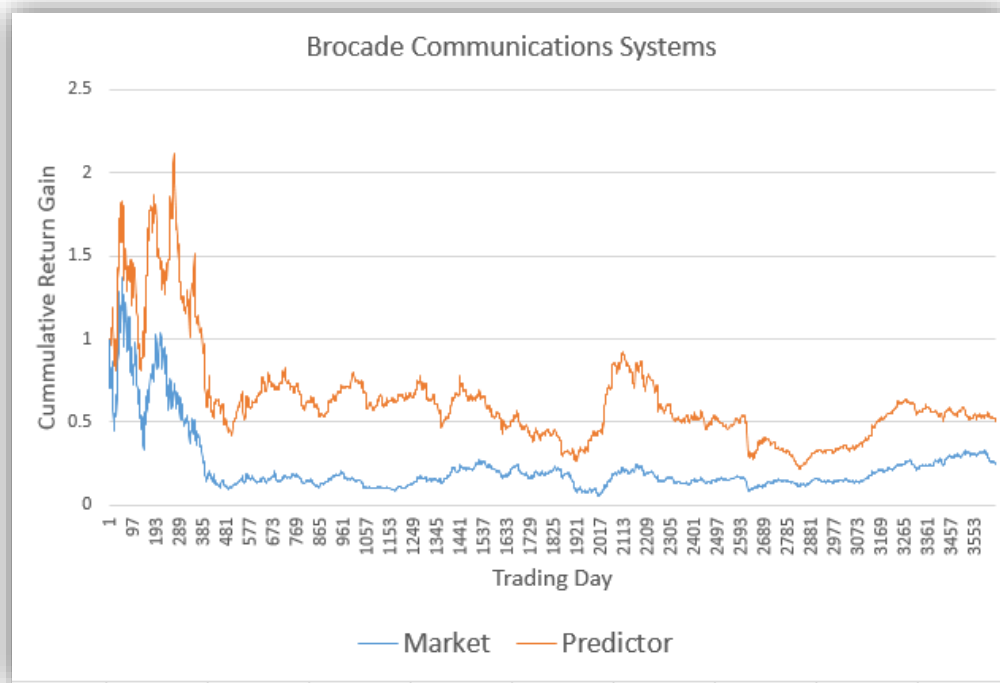
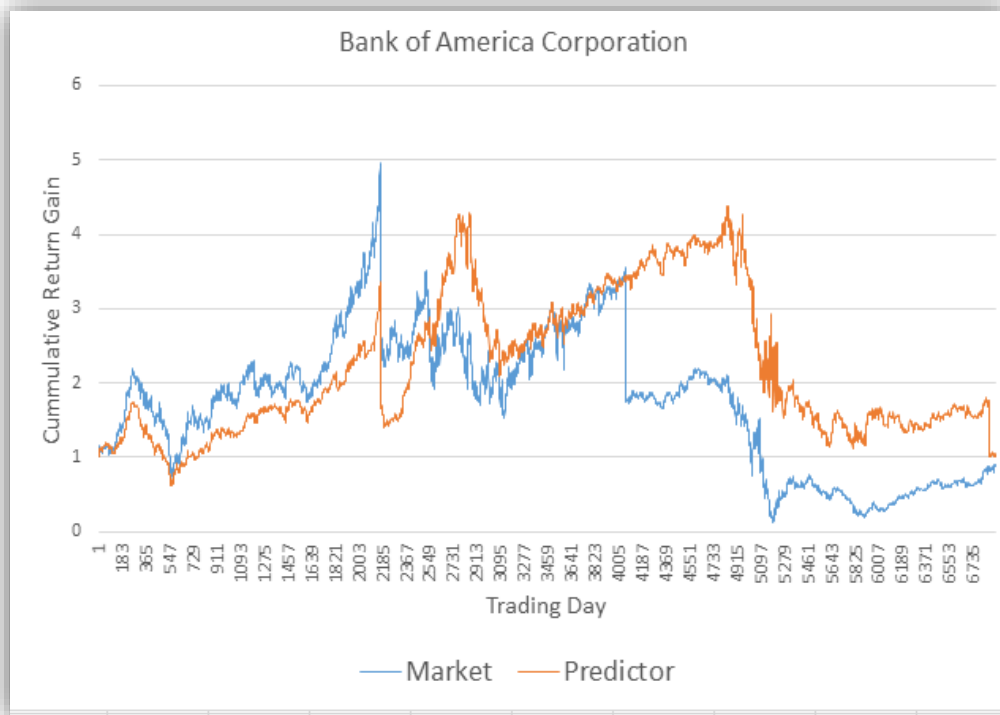
- El factor esparsivo es 3.

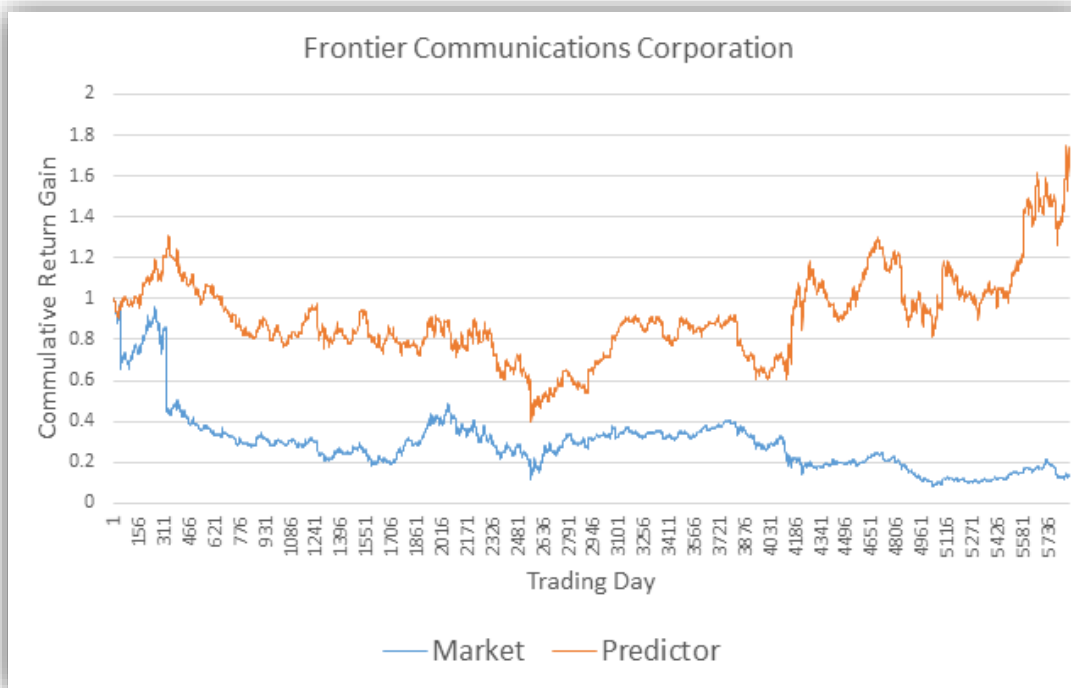
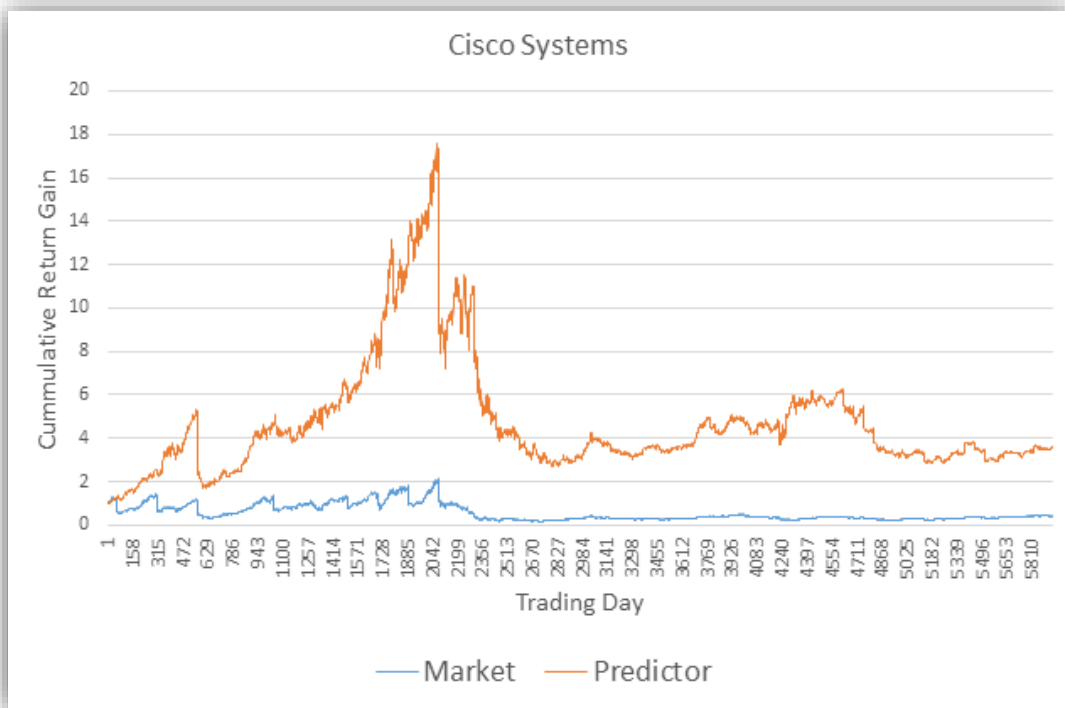
Para este conjunto de experimentos se observa que el modelo logra un mejor rendimiento que la del modelo basado en diccionarios con entrenamiento. Este modelo también se caracteriza por un tiempo de cálculo que es reducido considerablemente ya que no hay etapa de entrenamiento y el hecho de que el número de átomos es menor que la de los diccionarios entrenados.

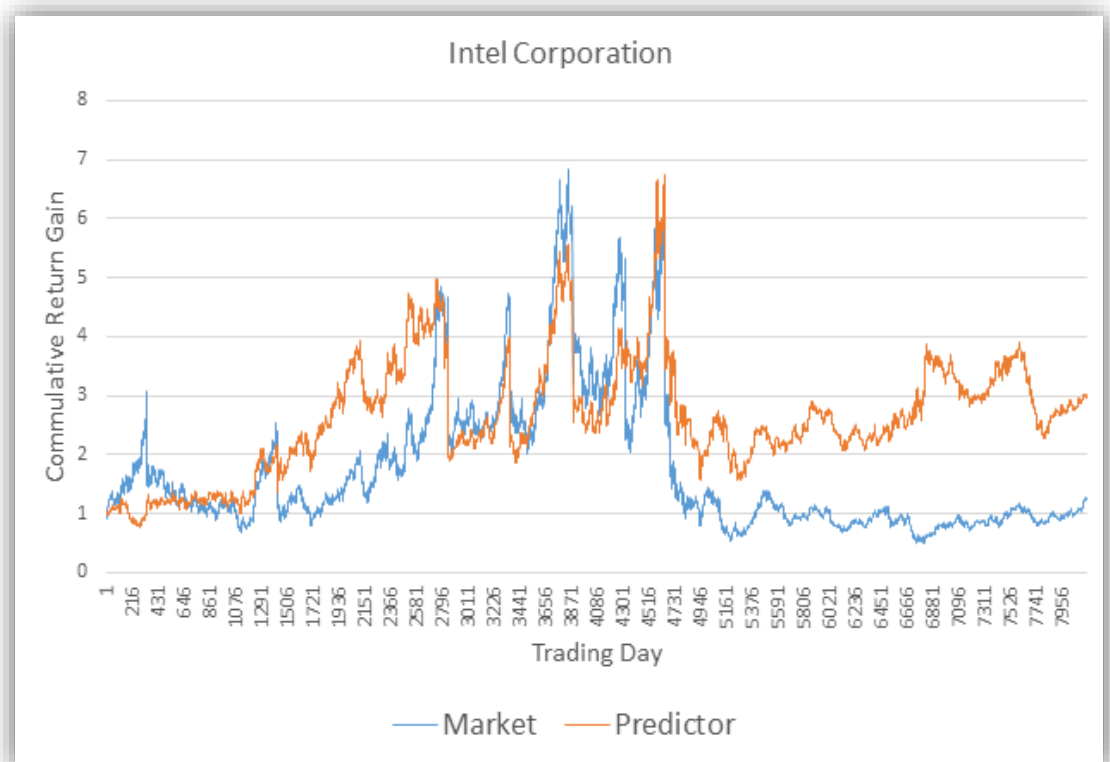
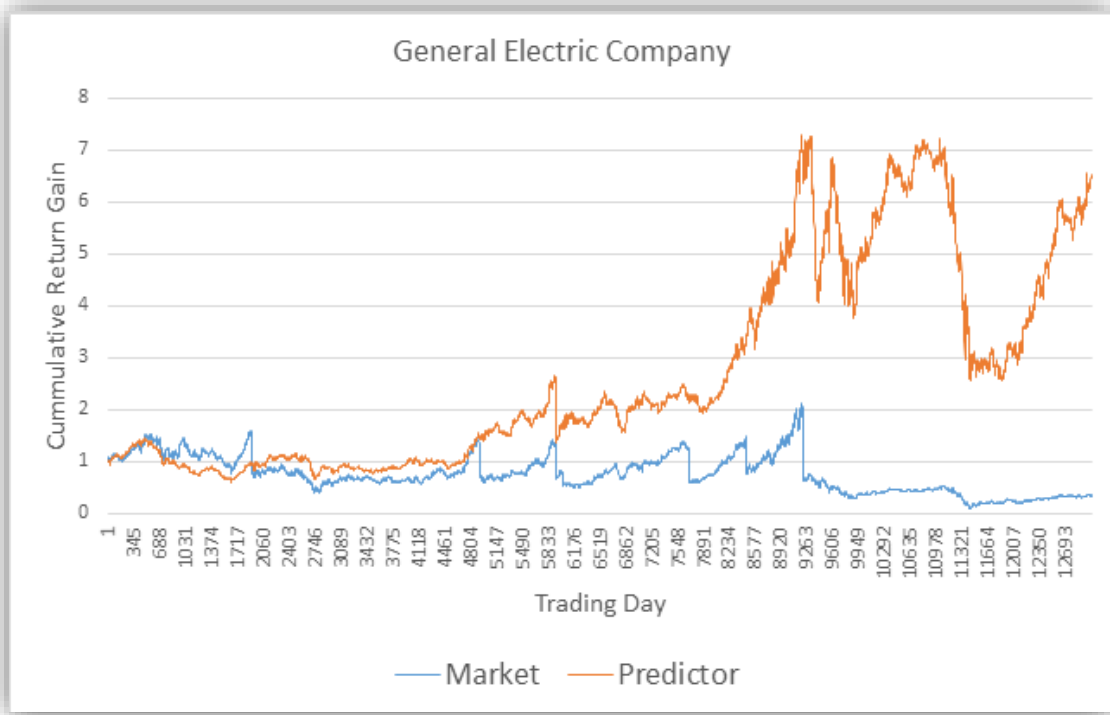
Hay algunos casos donde la serie geométrica corresponde al predictor alcanzando un valor mayor que uno e incrementando exponencialmente mientras que la serie del mercado se mantiene en valores menores a uno y por lo tanto se comporta como una señal constante. Algunos escenarios caracterizados por esta tendencia son las de Applied Materials a partir del día de negociación 2,671, Frontier Communications a partir del día 4239, General Electric a partir del día de negociación 8234 y Cisco Compañy entre el día 940 y 2192.

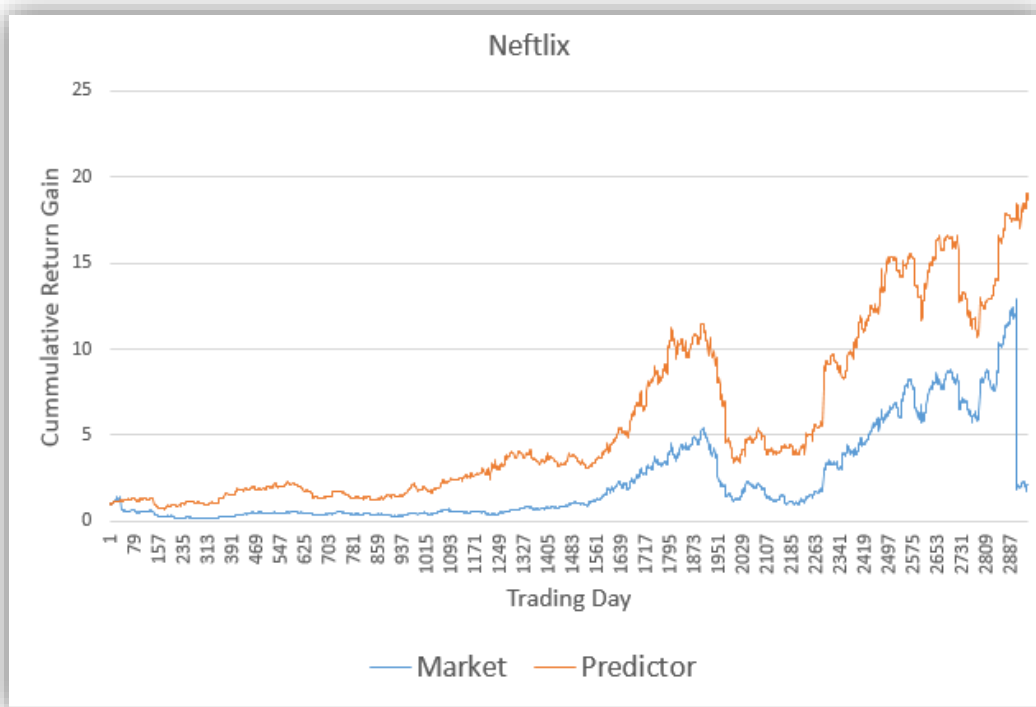
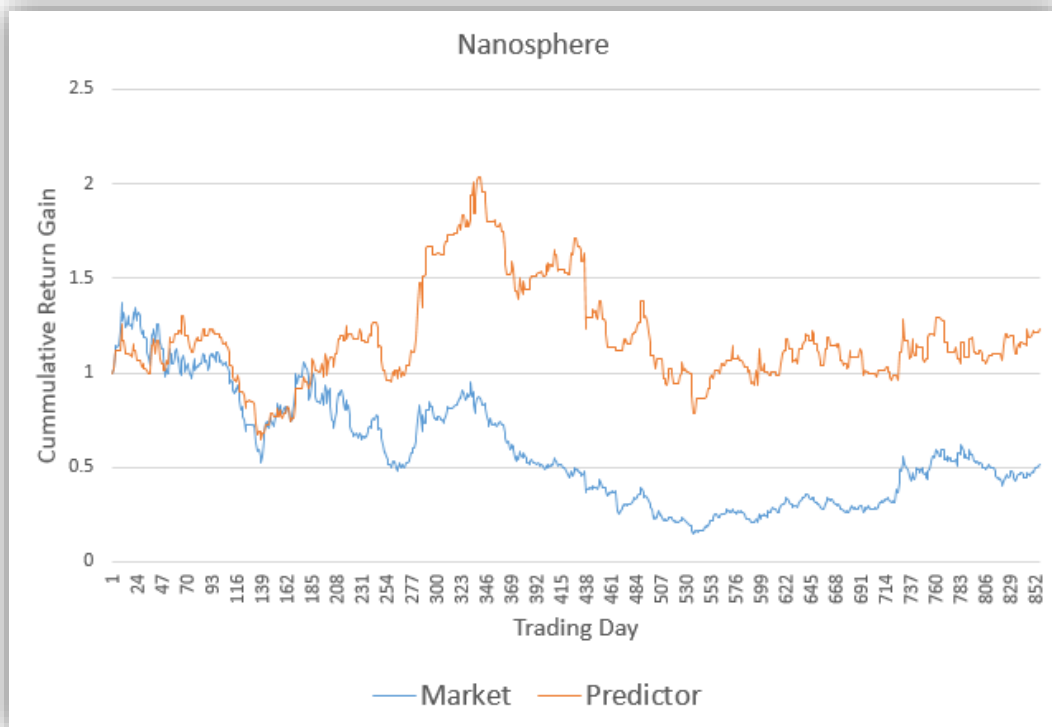


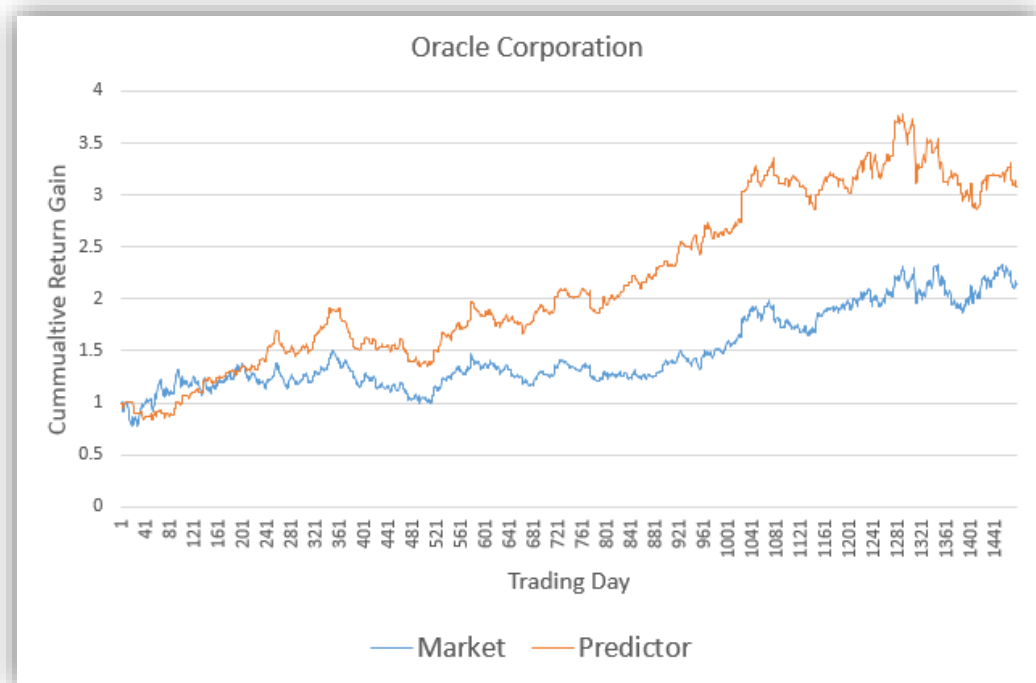
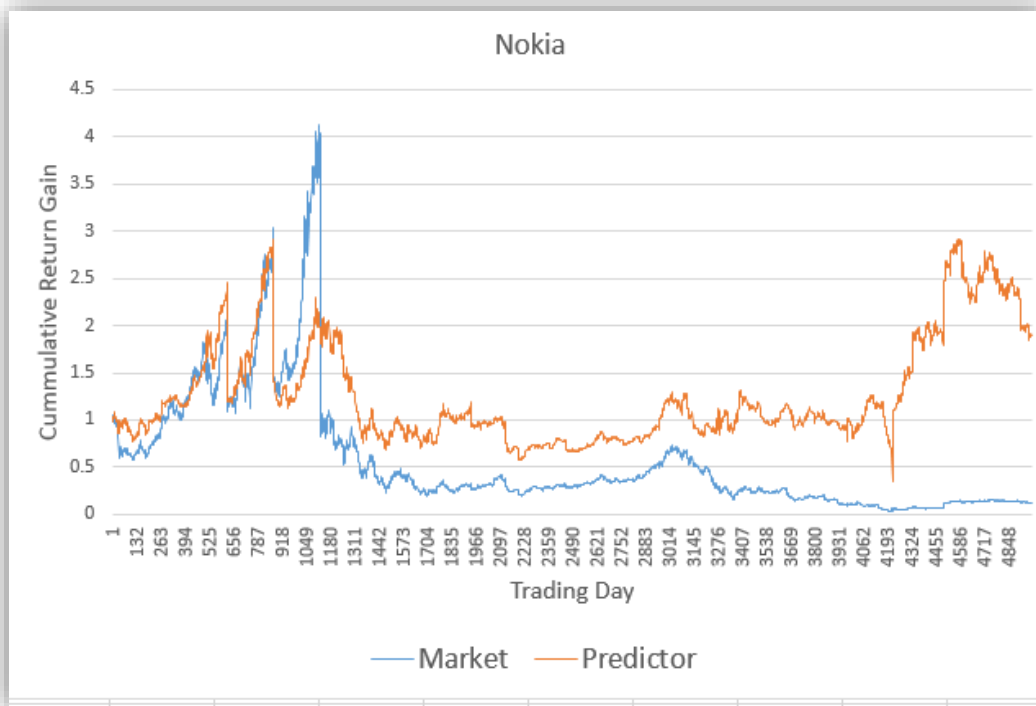


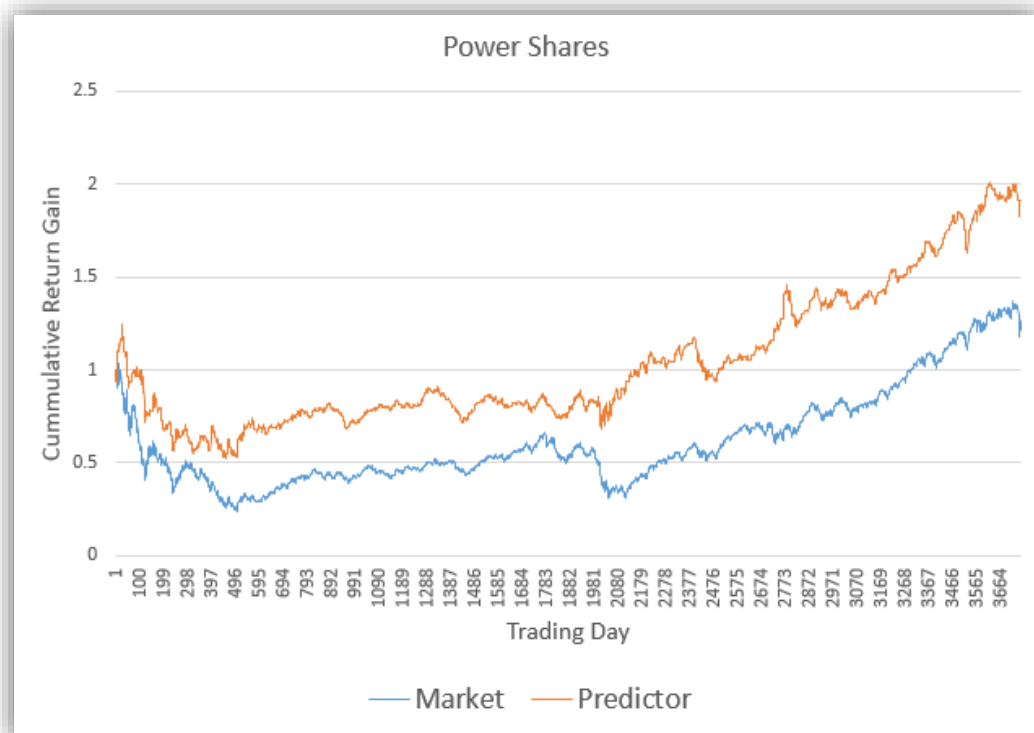
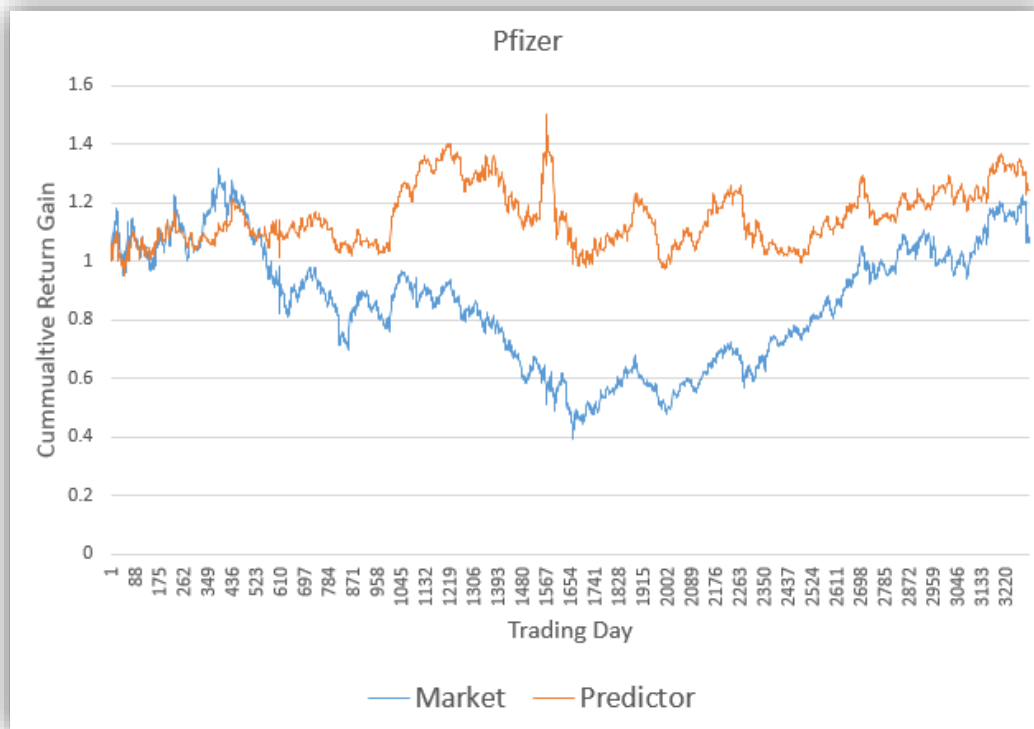


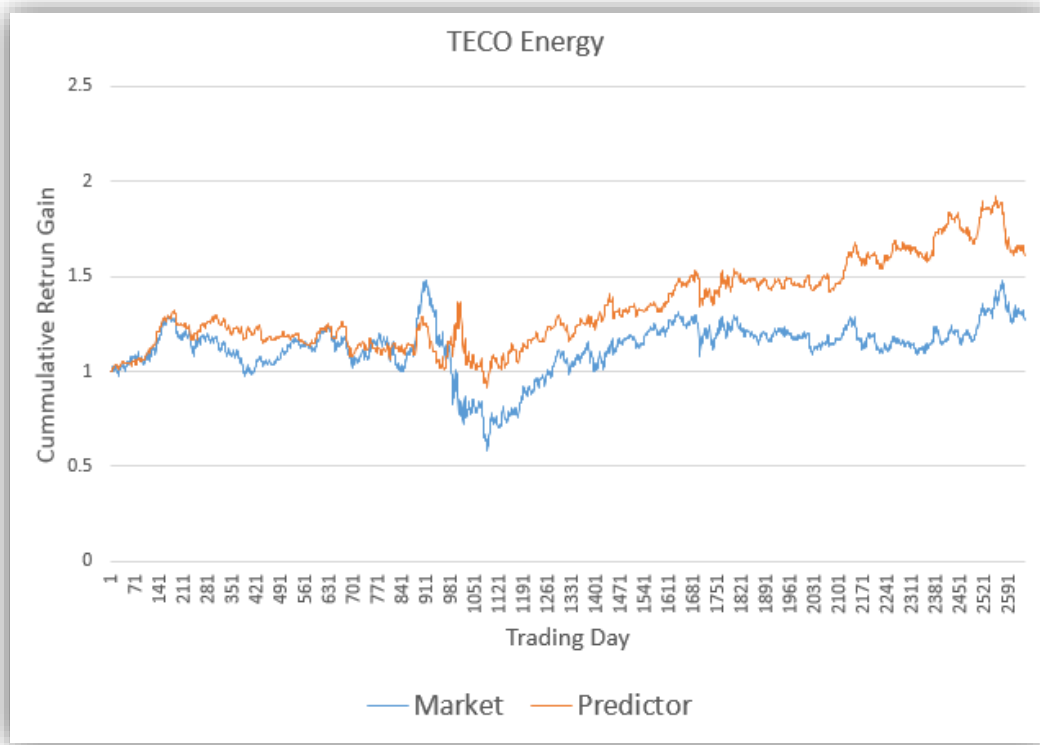
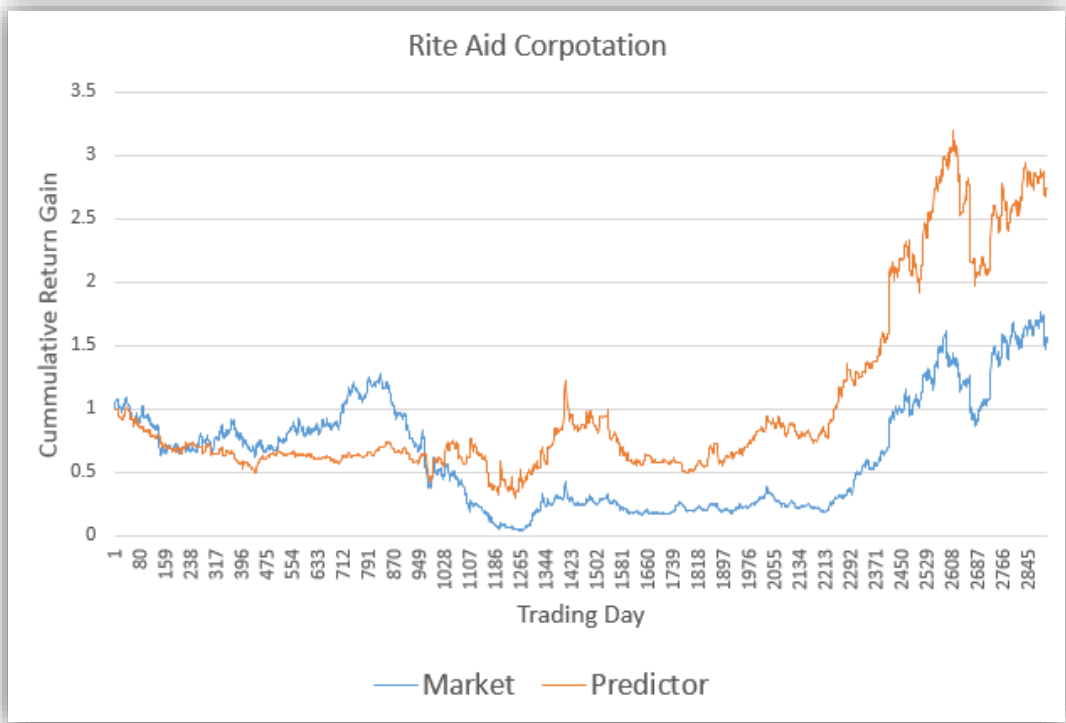


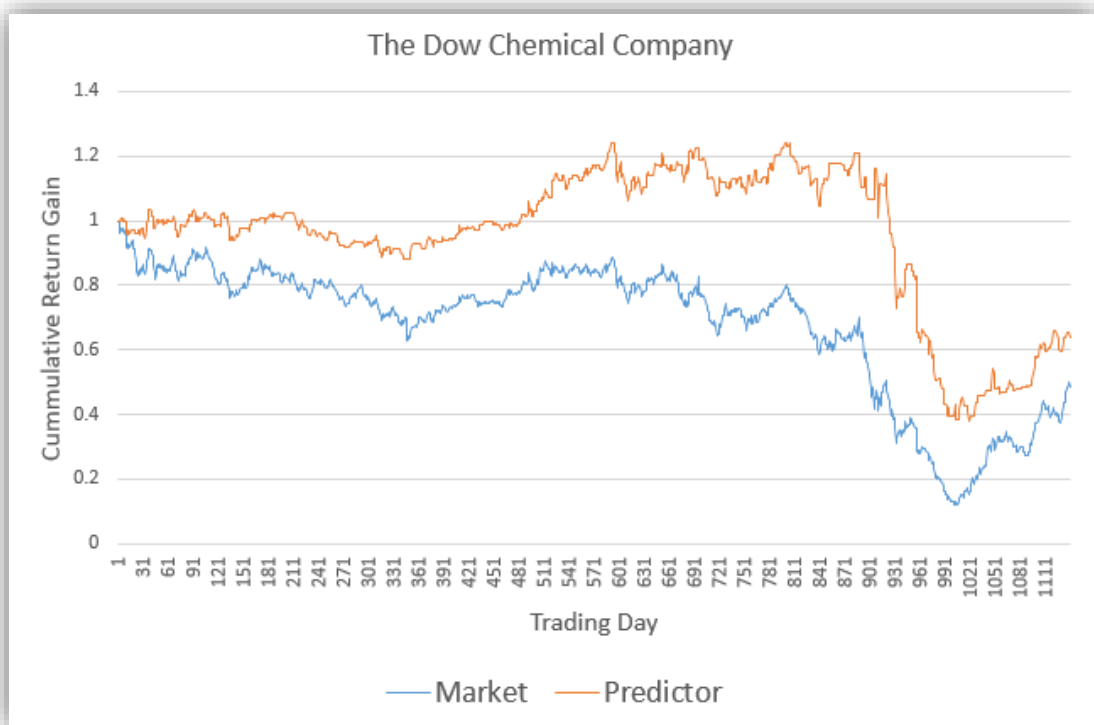
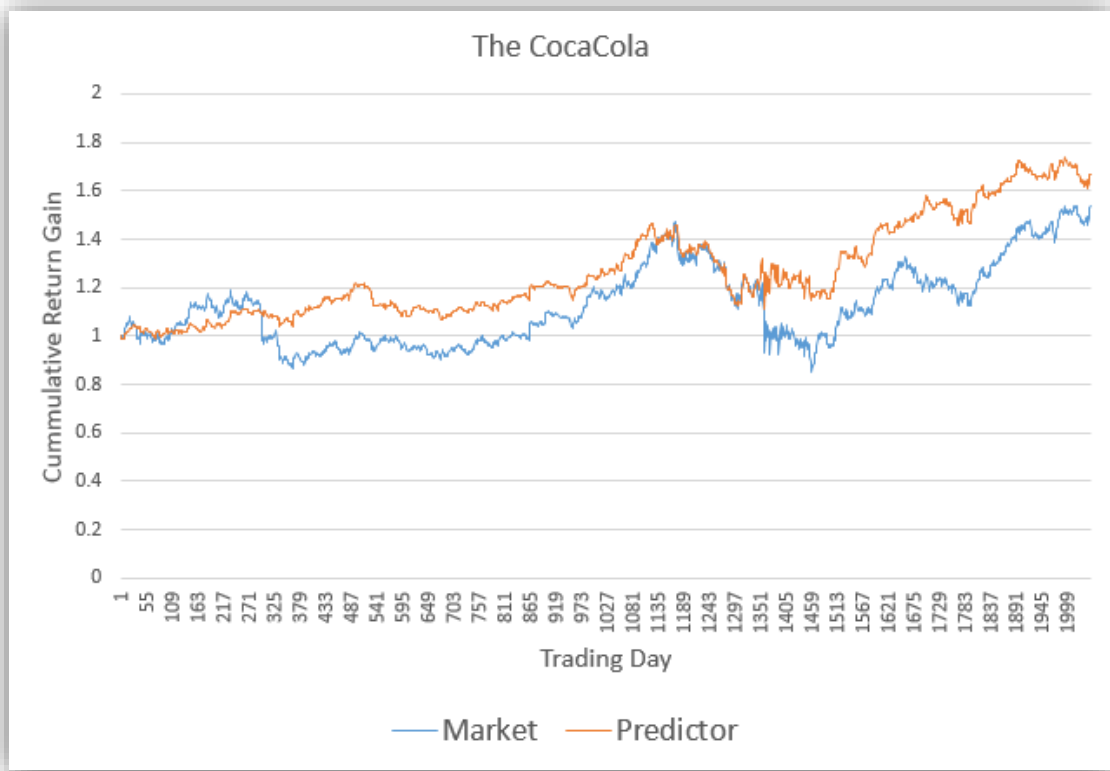












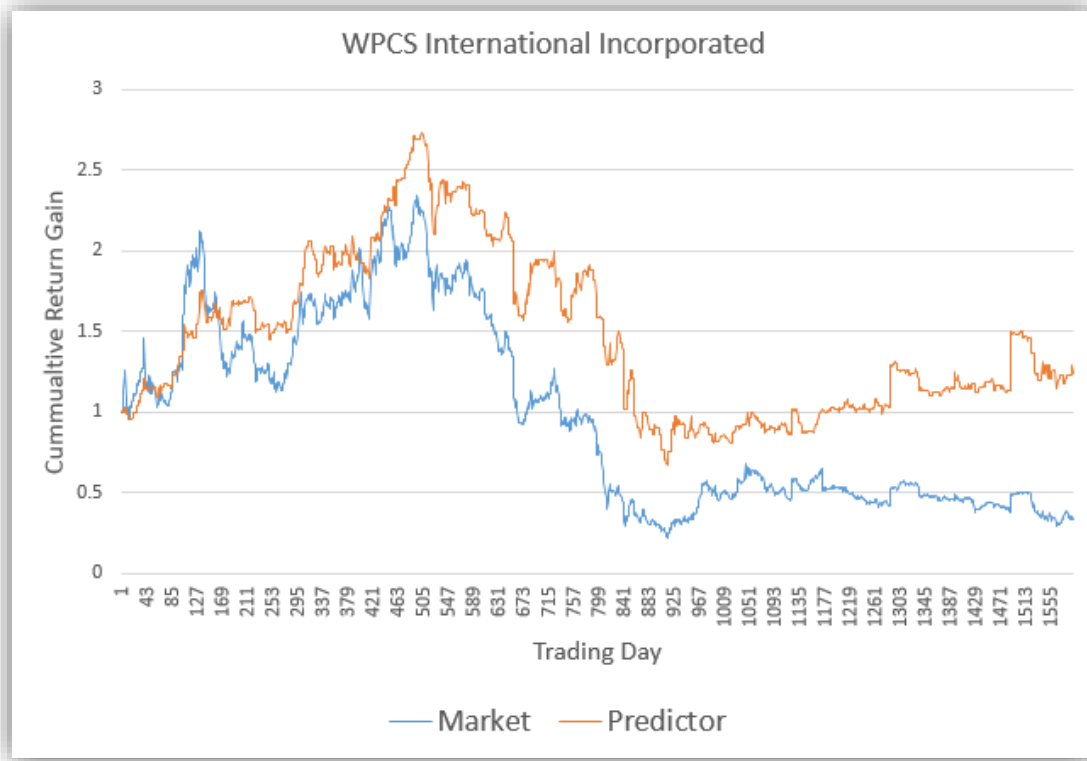


Figura 5 Ganancia de retorno acumulado del modelo predictivo basado en diccionarios sin entrenamiento vs la curva generada por el mercado

El análisis cuantitativo del modelo predictivo basado en diccionarios sin entrenamiento es mostrado en la tabla 2. Esta tabla presenta la misma métrica usada para analizar el potencia predictivo de los diccionarios con entrenamiento, el éxito de predicción, la diferencia promedio entre la ganancia del predictor y la ganancia del mercado, valor máximo y número de átomos por diccionario.

Los mejores resultados fueron obtenidos por Power Shares en términos de la tasa de éxito de predicción (0.99944), CISCO Systems en términos de diferencia promedio (4.143708), y Netflix en términos de diferencia máxima (17.22953).

Tabla 2 Análisis cuantitativo de modelo predictivo basado en diccionarios sin entrenamiento

Sin Entrenamiento					
	number of trading days	Successful Prediction Rate	Average Difference	Maximum Difference	Number of atoms
Alcoa Inc	1548	0.910207	0.322051	0.848233	350
Applied Materials	3377	0.998519	1.235327	3.205048	300
Banco Bradesco S.A	2813	0.79808	0.290351	0.777002	380
Banco Santander	6469	0.941104	0.273208	0.635865	380
Bank of America Corporation	6902	0.616778	0.463229	2.64655	450
Brocade Communications Systems	3637	0.999725	0.402858	1.423769	350
Cisco Systems	5935	0.992249	4.143708	15.53699	350
Frontier Communications Corporat	5864	0.997442	0.615141	1.607215	450
Intel Corporation	8158	0.743197	0.887895	3.251783	400
Nanosphere	855	0.912281	0.594047	1.23293	350
Netflix	2942	0.988443	3.363954	17.22953	300
Nokia	4970	0.914286	0.693434	2.775752	300
Oracle Corporation	1477	0.918754	0.694525	1.543832	450
Pfizer	3302	0.894912	0.293304	0.923765	350
PowerSharesQQQ	3748	0.999733	0.414269	0.757353	300
Rite Aid Corporation	2902	0.696072	0.384714	1.765183	300
TECO Energy	2638	0.920015	0.21106	0.638477	450
The CocaCola Company	2047	0.85393	0.14208	0.37172	400
The Dow Chemical Company	1140	0.998246	0.278879	0.714436	300
The General Electric Company	13034	0.88	1.9405494	6.774487	350
WPCS International Incorporated	1592	0.926508	0.477183	1.004763	350

5.3 Comparación de modelo predictivo basado en diccionarios con entrenamiento vs. modelo basado en diccionarios sin entrenamiento.

De acuerdo con algunos resultados que se muestran en la figura 4 y figura 5, así como en la tabla 1 y tabla 2, los diccionarios no entrenados superan a los diccionarios con entrenamiento para el caso de la predicción de series de tiempo financieras en términos de la ganancia de retorno acumulado.

Una ventaja adicional de los diccionarios no entrenados es la simplicidad y el menor tiempo de cálculo ya que no se requiere la extracción de los conjuntos de entrenamiento y la etapa

de entrenamiento del diccionario. Debido al menor tiempo de cálculo, los diccionarios no entrenados fueron utilizados para predecir los valores de retorno sobre los periodos de días de negociación mucho más largos, además para el caso de diccionarios sin entrenamiento, el número de átomos es considerablemente menor que el número de átomos en los entrenados.

Una motivación para usar diccionarios sin entrenamiento es el uso directo de las características de los átomos que son valores de ganancia de retorno sin cambios introducidos por una etapa de entrenamiento. Todos los átomos de un diccionario en espacio n $\mathbf{d}_i \in \mathbb{R}^n$, específicamente en la hiper-esfera unitaria $\|\mathbf{d}_i\|_2 = 1$. Otra característica de los diccionarios sin entrenamiento es que todos sus átomos están en el ortante positivo \mathbb{R}_+^n del espacio n , ya que todas las entradas de los átomos son valores de ganancia de retorno que no son negativos $\mathbf{d}_i \in \mathbb{R}_+^n$. De otra manera si los átomos son entrenados entonces los átomos no estarán en el ortante positivo.

Cualquier señal $\mathbf{x} \in \mathbb{R}_+^n$ es esparsivamente reconstruida sobre un diccionario y también es localizado en el ortante positivo desde que sus componentes son valores de ganancia de retorno. En conclusión los átomos en los diccionarios sin entrenamiento se encuentran más cerca de cualquier señal para ser esparsivamente reconstruida que con los átomos de diccionarios con entrenamiento. Esta es la razón por que el error $\|\mathbf{x} - \mathbf{D} \boldsymbol{\alpha}\|_2$ introducido por la reconstrucción esparsiva de una señal en el ortante positivo es más pequeño para el caso de diccionarios sin entrenamiento.