

Daily Ecopetrol Stock Performance Estimate by Using Estimation of Distribution Algorithms (EDAs)

Evolutionary Computation

Alejandro Peña Palacio, Camilo Palacio Orozco
Grupo de Investigación en Ingeniería del Software y
Modelamiento Computacional (GISMOC)
Escuela de Ingeniería de Antioquia - EIA
Envigado, Colombia
pjapena@eia.edu.co, capalacio6@hotmail.com

Sara Ramírez Arenas, Christian Lochmüller
Grupo de Investigación en Gerencia, Productividad y
Calidad (GPC)
Escuela de Ingeniería de Antioquia - EIA
Envigado, Colombia
sarisr23@hotmail.com, pfchloch81@eia.edu.co

Abstract — In this paper a model, which is based on the principles of evolutionary computation, will be developed and analyzed for financial forecasting and time series prediction. The model refers to returns, prices and transaction volumes of stocks that are listed on the Colombian Stock Exchange (BVC), and particularly to time series that represent the returns of the stocks of Ecopetrol. The proposed model uses the estimation of distribution algorithms (EDA's) for the short-term projection of the returns of such stocks. The EDAs, unlike other optimization techniques that are based on traditional evolution algorithms, provide greater statistical robustness in shaping the population of individuals and with respect to the number of generations that are required to solve the problem. For this purpose the EDA incorporates three dynamic populations that explore the solution space of the problem by using a mechanism which propagates the differences that exist among the best individuals. Additionally, it includes a strategy to preserve the diversity of the population. The proposed model allowed short-term projections for the future values of the returns of Ecopetrol stocks, departing from a set of parameters and variables that make up the stock price. This way the behavior of a specific asset can be described over time and the complexity which is associated with the application of traditional techniques, that allow the modeling of financial time series, can be eliminated.

Keywords- *Time Series, Estimation of Distribution Algorithms (EDA's), Evolutionary Computation, Evolutionary Mechanism, Ecopetrol*

I. INTRODUCCION.

Los numerosos periodos de inestabilidad que han experimentado los mercados financieros en los últimos años, a causa de la crisis financiera mundial desatada en el año 2008, ha hecho que las entidades encargadas de velar por la estabilidad de los mercados a nivel internacional, muestren gran interés por el modelamiento y la simulación de la dinámica de los mismos, centrados principalmente en el mercado de valores [22], [8]. Por su parte, los rendimientos de las acciones en los mercados de valores, han sido considerados por mucho tiempo como

aleatorios, pues incorporan información sobre la volatilidad de las acciones [2], lo que es determinante en el comportamiento de los mercados. Es por esto, que muchos modelos utilizados para describir este comportamiento, involucren parámetros y variables de tipo estocástico [17]. Asimismo, los actores del mercado financiero, han mostrado gran interés en los últimos años, por el uso de herramientas sofisticadas para la toma de decisiones, esto con el fin de maximizar sus ganancias, aumentar sus rendimientos en el mercado, y reducir sus pérdidas asociadas al riesgo inherente de sus activos [8].

Para la solución de este problema, se han desarrollado una serie de modelos que buscan determinar la dinámica de series temporales financieras, como los modelos EGARCH, GARCH, ARMA, ARIMA entre otros [16], y los cuales para su utilización, requieren de un alto conocimiento tanto estadístico como matemático, dada su complejidad. Es por esto que la incorporación de herramientas cada vez más complejas para la toma de decisiones en el ámbito financiero, ha motivado a diferentes autores a la construcción de modelos financieros como el presentado por Trejos et al. [23], el cual permite la predicción a corto y a mediano plazo, del precio promedio ponderado de la Acción de Cementos Argos (acción inscrita en la BVC). Este modelo, fue desarrollado mediante la utilización de la metodología de *Box and Jenkins*. De forma similar, en el Departamento de Economía Aplicada de la Universidad de Sevilla, se llevó a cabo un estudio para la proyección del comportamiento del Ibex 35 en España, para horizontes de pronóstico de 2, 5, 10 y 20 pasos hacia adelante, mediante la utilización de redes neuronales del tipo *multilayer feedforward* [15]. Los resultados obtenidos por los autores, mostraron las pocas variaciones que el modelo experimento frente a la estimación del Ibex35 para horizontes de predicción cortos, eso sí, enfatizando en la conveniencia de la utilización de métodos de proyección no paramétricos. Más adelante, Fernández [8] presentó en su artículo “EGARCH: un modelo asimétrico para estimar la volatilidad de series financieras”, la compilación de una serie de elementos que muestran las ventajas de los

modelos EGARCH, frente a otros modelos utilizados para la estimación y pronóstico de series de tiempo financieras. En este artículo, se puede observar el efecto que las noticias económicas generan sobre la volatilidad en una serie de tiempo [8].

Desde el punto de vista de la Inteligencia Computacional, los años 80 marcaron un gran avance en esta área del conocimiento, lo que fue fundamental para el desarrollo de modelos que explicaran la dinámica de los mercados financieros, empleando para ello técnicas promisorias basadas en sistemas por adaptación y aprendizaje. Es por esto que Mahfoud & Mani [13], desarrollaron un modelo que emplea los algoritmos genéticos integrados a las redes neuronales artificiales, para analizar el desempeño de una acción en un horizonte de proyección de 12 semanas. Los resultados arrojados por el modelo, mostraron el buen comportamiento de los algoritmos por evolución, en la solución de problemas que requieren la experiencia de un experto, tal y como sucede en los Sistemas Clasificadores Genéticos [5], [4].

De esta manera, se han desarrollado una serie de modelos para el pronóstico y la predicción de series temporales, mediante la utilización de algoritmos estadísticos por evolución (EDA's) [9], [10], o la utilización de modelos borrosos para la identificación del riesgo operacional en las actividades del negocio que involucra una entidad financiera [20], o en la asignación de puntajes crediticios para clientes beneficiados por un crédito de consumo [21].

Es por esto que en este artículo, se desarrolla y analiza un modelo basado en los principios de la computación evolutiva, para la proyección de los rendimientos a corto plazo, de una de las acciones colombianas más exitosas inscritas en la Bolsa de Valores de Colombia (BVC – <http://www.bvc.com.co/pps/tibco/portalbvc>), como es la acción Ecopetrol [6], y que dada su alta volatilidad, es necesario conocer su comportamiento. Es por esto que el modelo propuesto, lleva a cabo un proceso de identificación de las variables más relevantes que conforman el precio de la acción, permitiendo así identificar las dinámicas a las cuales se ve sometida en el mercado de Valores Colombiano. El modelo propuesto, incorpora además un Algoritmo de Estimación de la Distribución EDA's, el cual a diferencia de los algoritmos por evolución tradicional, no precisa de operadores genéticos, por lo que la eficiencia en la solución de un problema, radica en el tamaño de la población, y en el número de generaciones requerido para alcanzar una solución óptima.

De acuerdo con los principios de los EDA's, la estructura de solución o individuo, posee una serie de genes que permiten relacionar los valores de las variables que conforman el precio de la acción, con la estimación de sus rendimientos, y en donde la función de aptitud, está definida en términos del error obtenido por el modelo frente a la estimación de los valores que conforman la serie de tiempo de los rendimientos de la acción. Para la transformación de las posibles soluciones o individuos, el Algoritmo EDA propuesto, incorpora tres poblaciones dinámicas que exploran el espacio de solución del problema, mediante un mecanismo propagador de las diferencias entre los mejores individuos, y el cual posee una estrategia orientada a preservar la diversidad de las poblaciones por medio de distribuciones de probabilidad

normales. Los resultados obtenidos por el modelo propuesto, permitieron proyectar en el corto plazo, valores futuros de la serie temporal de los rendimientos de la acción de Ecopetrol, a partir de una serie de variables temporales asociadas, las cuales aportaron información relevante sobre la posible volatilidad, y el comportamiento de la acción. Finalmente, se presentan una serie de conclusiones y recomendaciones, que permiten identificar una serie de tendencias de desarrollo en esta área del conocimiento, y que permitirán extender este tipo de modelos, a otros comportamientos observados para diferentes acciones, tanto en el mercado colombiano, como en el mercado internacional, dando una mayor confiabilidad a la toma de decisiones en los mercados financieros.

II. METODOLOGÍA.

A. Experimentación.

Para el análisis y la validación del modelo, se llevó a cabo la construcción de una base de datos, la cual muestra la evolución diaria de los rendimientos de la acción Ecopetrol, para un período comprendido entre el 29 de Noviembre de 2007, y el 30 de Junio de 2011. El comportamiento de esta acción, fue obtenida a partir de la información histórica que posee la acción en la Bolsa de Valores de Colombia (BVC). Esta serie de tiempo, está compuesta por un total de 875 registros, los cuales determinaron igualmente, el comportamiento de cada uno de los parámetros y variables que conforman el valor del rendimiento de la acción, tal y como se puede observar en la Tabla 1 [6].

TABLE I. VARIABLES DE ENTRADA AL MODELO POR EVOLUCIÓN PROPUESTO.

RUIDO (R)	WTI	UTILIDAD NETA (UN)	VARIACIÓN % (VP)
PRECIO (P)	IGBC	VOLUMEN (T)	

Dónde:

IGBC: Representa el Índice General de la Bolsa de Colombia IBVC [\$].

WTI: Representa el índice de referencia para el barril de petróleo producido por Ecopetrol de acuerdo con la West Texas Intermediate dado [u\$].

RUIDO: Representan la señal aleatoria temporal, que permite incorporar el comportamiento de parámetros y variables no consideradas por el modelo.

Para el análisis y la validación del modelo, se llevó a cabo una etapa inicial de aprendizaje, en la cual se tomaron quinientos (500) registros sobre la serie de la acción, mientras que los datos restantes (375), fueron utilizados en una segunda etapa para la validación del modelo. Para el análisis del comportamiento del modelo, tanto en su etapa de entrenamiento, como en su etapa de validación, fue necesaria la utilización del modelo propuesto por Park (2007), el cual evalúa el comportamiento estadístico de un modelo frente a los datos de referencia, o datos de aprendizaje.

Finalmente, se llevó a cabo un análisis del comportamiento del modelo frente al proceso de estimación de los rendimientos de la acción Ecopetrol, en términos del número de individuos (NI) que componen la población inicial de individuos [4], y en términos de los principios que rigen el comportamiento de un Algoritmo

de Estimación de la Distribución (EDA's), frente a la solución del problema [12].

B. MODELO EVOLUTIVO PROPUESTO.

Uno de los elementos más importantes en el desarrollo actual de la computación evolutiva (CE), lo constituyen los algoritmos de estimación de la distribución (EDA's), los cuales incorporan mecanismos por evolución para la solución de un problema, basados principalmente en las características de las posibles soluciones o *individuos*, y que conforman una población [7], [11], [12].

1) Estructura de Solución o Individuo.

De acuerdo con los principios de la computación evolutiva, la estructura de solución o individuo, estará compuesta por dos *genotipos* de la siguiente manera:

- *Genotipo 1*: Está compuesto por una serie de genes que permiten subdividir la información de las variables de entrada. El tamaño de este genotipo, se define como: $NG1 = NE \cdot NPE$. Donde, $NGE1$: Indica el número de genes que conforma el *genotipo 1*. NE : Número de variables de entrada al modelo NPE : Número de valores de subdivisión de las entradas.
- *Genotipo 2*: Está compuesto por una serie de valores, que permiten agrupar la información por características, de acuerdo con la división de la información de entrada. Este número de genes se denota: $NG2 = NPE$.

De esta manera, la estructura de solución o individuo, se denota y define:

$G_{1,1}$	$G_{NE,NPE}$	$G_{2,1}$...	$G_{2,NPE}$
-----------	------	--------------	-----------	-----	-------------

Figure 1. Estructura de Solución o Individuo.

De acuerdo con lo anterior, la longitud del individuo o estructura de solución, poseerá un total de genes:

$$NGT = NG1 + NG2 \quad (1)$$

2) Función de Aptitud.

De acuerdo con la estructura de solución o *individuo*, la función de aptitud que califica la calidad de cada individuo frente a la solución del problema, se denota y define [3]:

$$FA = \frac{1}{\left\| \sum_{k=1}^{nd} (yd_k - yr_k)^2 \right\|} \quad (2)$$

Dónde:

yd_k : Indica los datos de referencia de la serie de tiempo que representa los rendimientos de la acción Ecopetrol (\$), mientras que yr_k , indica los rendimientos estimados por el modelo propuesto para la acción. nd : Indica el número de datos que conforman la serie de tiempo. De igual manera, el valor de yr_k se define de la siguiente manera, en términos de la estructura de solución o individuo.

$$yr_k = G_{2,j} \cdot G_{1,j,i} \cdot X_{i,k} \quad (3)$$

Dónde:

$X_{i,k}$: Representa el vector que contiene las variables que conforman el precio de la acción, para un instante de tiempo k , así como sus retardos temporales. i : Indica el número de entradas al modelo, j : indica el número de genes de subdivisión de la información.

3) Mecanismo por Evolución.

Para la transformación del conjunto de las posibles soluciones, el mecanismo por evolución propuesto, se basa en un Algoritmo de Estimación de la Distribución (EDA's) del tipo MAGO (MultiDynamics Global Optimization) [9], el cual es un mecanismo que inicia con la generación una *población de posibles soluciones* distribuidas de forma aleatoria sobre el espacio de solución del problema, y la cual es subdividida en tres poblaciones o grupos.

El primer grupo de individuos, está compuesto por una elite de individuos seleccionados en términos de la función de aptitud, la cual es transformada mediante un proceso de mutación llamado de *Dinámica Emergente*, y en donde la generación de nuevos individuos se da de la siguiente manera:

$$G_T^{(g)} = G_i^{(g)} + F^{(g)} \cdot (G_B^{(g)} - G_M^{(g)}) \quad (4)$$

Dónde, $G_B^{(g)}$: Indica los valores de los genes del mejor individuo de la población sobre la generación (g). $G_M^{(g)}$: Indica los valores de los genes de un individuo aleatorio. Para poder incorporar la información sobre las relaciones entre los genes, se obtiene la siguiente matriz de correlación genética, la cual se denota y define de la siguiente manera:

$$F^{(g)} = \frac{S^{(g)}}{\|S^{(g)}\|} \quad (5)$$

Donde $F^{(g)}$, representa la matriz de correlación entre las variables de la estructura de solución, y en dónde, $S^{(g)}$: representa la matriz de covarianza sobre la generación g . Es de anotar que este subgrupo, tiene el objetivo de hacer más rápida la convergencia.

Un segundo grupo, o de *Dinámica Másiva* (G_2), lleva a cabo la generación de nuevos individuos por muestreo aleatorio, sobre una distribución *hiper-rectangular* definida en el intervalo $[LB^{(g)}, UB^{(g)}]$ de la siguiente manera:

$$LB^{(g)} = G_m^{(g)} - \sqrt{\text{diag}(S^{(g)})}, \quad UB^{(g)} = G_m^{(g)} + \sqrt{\text{diag}(S^{(g)})} \quad (6)$$

Dónde, $G_m^{(g)}$: Representa la media de cada uno de los genes que componen la población de individuos, sobre la generación g . Este subgrupo, busca posibles soluciones en una vecindad cercana a la media poblacional. Por su parte, esta población permite una evolución más acotada.

Un tercer grupo (G_3), es sometido a un mecanismo de *Dinámica Accidental*, en donde los individuos son generados mediante muestras tomadas de una distribución uniforme en todo el espacio de solución del problema, de manera similar a como se hace en la generación de la población inicial. Esta población tiene dos funciones básicas, una, mantener la diversidad de la población, y dos, continuar con la exploración de todo el espacio de búsqueda.

Este mecanismo, lleva el proceso por evolución, hasta que se cumpla un criterio de parada dado, el cual se

vincula directamente con el error en la estimación de los valores de rendimiento para cada uno de los datos que conforma una serie de tiempo. Es de anotar que la función de aptitud de los nuevos individuos, es comparada con la función de aptitud de los padres en cada población, y en donde el mejor individuo entre padres e hijos, se mantiene en cada población. Una vez finalizado el proceso por evolución, se obtiene un individuo que contiene la información genética que mejor describe la estimación de los valores de la serie de tiempo de la acción Ecopetrol. Este mecanismo por evolución, si bien no garantiza una solución global, si garantiza una solución óptima al problema de estimación de los rendimientos de la acción.

III. ANÁLISIS DE RESULTADOS.

Para llevar a cabo el análisis y la validación del modelo propuesto, se hizo en primera instancia un análisis de la serie de tiempo que contiene los valores del precio de cierre de la acción Ecopetrol, como se muestra en la Figura 2.

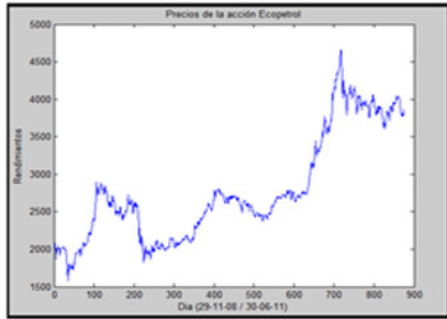


Figure 2. Serie de Precios de la Acción de Ecopetrol.

Para este análisis, se procedió con la estimación del diagrama de autocorrelación para cada una de las variables del modelo que conforman el rendimiento de la acción, el cual arrojó como resultado la correlación entre los datos que conforman la serie, tal y como se puede observar en la Figura 3.

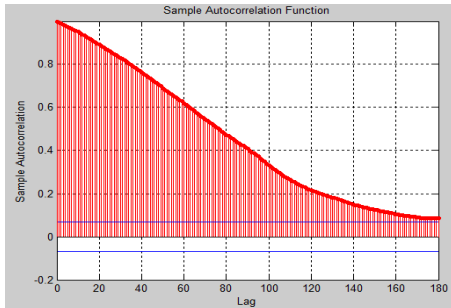


Figure 3. Diagrama de Autocorrelación Parcial para la Serie Temporal de precios de cierre de la acción Ecopetrol.

De acuerdo con este diagrama, se puede observar que la serie de precios de la acción, presenta tendencia de crecimiento en los rendimientos, lo que está de acuerdo con el precio internacional del Barril de Petróleo WTI para el período de tiempo que comprende la serie. Este diagrama de autocorrelación, fue estimado para las demás variables que conforman el precio de los rendimientos de la acción, lo que arrojó como resultado los siguientes retardos temporales (Tabla II).

TABLE II. RETARDOS TEMPORALES POR CADA UNA DE LAS VARIABLES QUE CONFORMAN EL PRECIO DE LA ACCIÓN.

Variable	Retardos	Variable	Retardos
Ruido	1	Variación %	2
Precio WTI	3	Precio	7
Utilidad Neta	2	IGBC	4
Volumen	2	Horizonte	1
		Na	21

De acuerdo con lo anterior, la estructura del modelo de pronóstico, se denota y define de forma general de la siguiente manera:

$$y_k = \sum_{i=0}^{na} a_i y_{k-i} + e(n) \quad (7)$$

Donde na , indica el número de retardos considerados por el modelo autorregresivo para cada una de las variables que conforman el rendimiento de la acción. Sin embargo, este número de retardos puede variar, dependiendo de los coeficientes obtenidos luego del proceso por evolución [1]. Es de anotar, que a_i se puede representar en términos de la estructura de solución o *individuo* de la siguiente manera:

$$a_i = G_{2,j} \cdot G_{1,j,1} \cdot X_{1,k} \quad (8)$$

Para el entrenamiento del modelo propuesto, se construyó una base de datos 868*21 registros, teniendo en cuenta para ello los retardos identificados anteriormente en cada una de las variables que conforman el precio de la acción. De esta manera, el aprendizaje del modelo se llevó a cabo sobre los 500 primeros registros, mientras que el proceso de validación se hizo para la serie de registros restantes (368). Los resultados obtenidos por el modelo para la fase de entrenamiento, se muestran en la Figura 4.

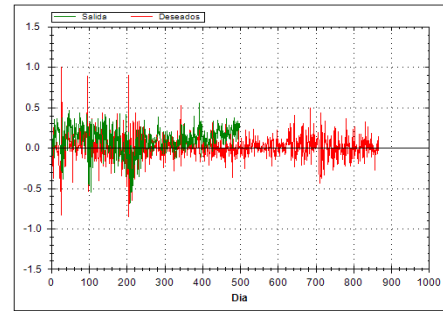


Figure 4. Resultados obtenidos luego del proceso por evolución para la serie de la acción Ecopetrol

De igual manera, en la Figura 5 se muestran los resultados obtenidos por el modelo durante la fase de validación, para un total de 368 registros.

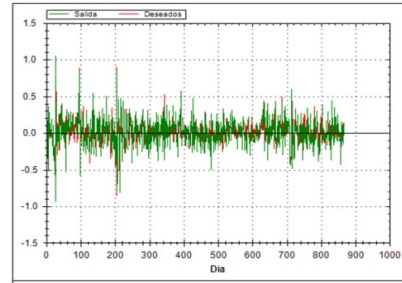


Figure 5. Validación del individuo frente a la totalidad de los registros.

Para la validación del modelo en sus dos etapas, se utilizó el modelo borroso propuesto por Park [19], el cual evalúa el desempeño de un modelo frente a los datos, integrando para ello ocho métricas estadísticas como: Fractional Bias (FB), Normalized Mean Square Error (NMSE), Geometric Bias Mean (MG), Geometric Bias Variance (VG), Within a Factor of Two (FAC2), Index of Agreement (IOA), Unpaired Accuracy of Peak (UAPC) and Mean Relative Error (MRE). De acuerdo con este modelo, cada una de las métricas es descrita cualitativamente en términos de los valores alcanzados por cada una de ellas de la siguiente manera: Good (G), OverFair (OF), Fair(F), UnderFair (UF), Poor(P). Para obtener el valor de desempeño, Park propone una serie de valores que son aditivos de acuerdo con las cualidades tomadas por cada métrica así: Good 7-10 (average 8.5), Fair 4-7 (average 5.5), OverFair (average 6), UnderFair (average 5) and Poor 1-4 (average 2.5), hasta lograr un valor máximo de 68 puntos, los cuales pueden ser obtenidos en porcentaje teniendo en cuenta el puntaje máximo. De acuerdo con lo anterior, los resultados arrojados por el modelo luego de la fase de entrenamiento y validación, se muestran en la Tabla III [18].

De acuerdo con la Tabla III, se puede observar de manera general, que el modelo presentó un mejor desempeño en la etapa de entrenamiento, que frente a la etapa de validación, hecho que es evidente en modelos basados en algoritmos por adaptación y aprendizaje, los cuales pueden tener problemas de generalidad, a pesar del score logrado. De esta manera igualmente, podemos observar que el modelo tuvo la tendencia a la sobreestimación de los datos, tal y como se presenta en la negatividad del índice FB, y en el valor tomado por el índice MG, el cual es cercano a la unidad. Esto demuestra que la media de los datos, se encuentra levemente por encima de los datos de referencia.

TABLE III. DESMPÑO DEL MODELO FRENTE A LOS DATOS DE REFERENCIA

MÉTRICA	Etapa de Aprendizaje		Etapa de Validación	
	VALOR	CUALIDAD	VALOR	CUALIDAD
FB	-0.03888683	G	-0.0461942	G
NMSE	0.04050601	G	0.04677059	G
MG	0.952555	G	0.84433884	G
VG	1.00004971	G	1.02963259	G
FAC2	0.94728637	G	0.84272685	G
IOA	0.80744553	G	0.78078305	G
UAPC2	0.012	G	-0.015	G
MRE	0.045528102	G	0.06835749	G
Score		68		68

Los resultados, mostraron una mayor dispersión en los datos de validación, tal y como se puede observar en los índices VG y NMSE. Con respecto al índice de predicción FAC2, los valores estuvieron por encima del 75%, lo que muestra el buen comportamiento del modelo, esto aunado al valor alcanzado por el IOA, el cual logró que la serie de tiempo resultante del aprendizaje, mantenga la forma de la serie de tiempo de referencia. Con respecto a la flexibilidad del modelo frente a cambios inesperados en los valores de la serie de rendimientos, el índice UAPC2 mostró valores más allá del 75%, lo que corrobora la poca discrepancia del modelo frente a los datos de aprendizaje, hecho que se evidencia aún más, en los valores alcanzados

por el índice MRE, los cuales estuvieron por debajo del 30%.

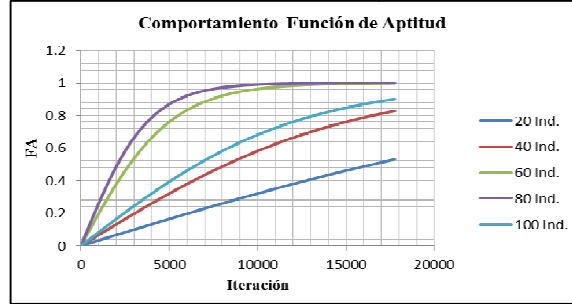


Figure 6. Comportamiento de la Función de Aptitud para una población de 80 individuos.

Para evaluar el comportamiento del mecanismo por evolución, el modelo inicia con la generación de una población inicial de individuos (NI), la cual se mantiene constante durante todo el proceso de estimación. Para tal efecto, se hicieron una serie de pruebas sobre la cantidad de individuos, iniciando la población con 10 individuos, y llegando gradualmente, a un total de 100 individuos. Para cada tamaño de población, se llevaron a cabo 20000 iteraciones, en donde para cada iteración, se identificó el valor promedio de la función de aptitud de la población, tal y como se muestra en la Figura 6. De acuerdo con lo anterior, se puede observar que las poblaciones que mejor comportamiento presentaron frente a la función de aptitud normalizada, fueron las poblaciones que tuvieron 80 individuos, y 60 individuos respectivamente, a diferencia de otras poblaciones, las cuales mostraron el crecimiento del valor de la función de aptitud algo más moderada.

IV. CONCLUSIONES.

El modelo evolutivo propuesto, permitió llevar a cabo el pronóstico y la predicción de la serie temporal de rendimientos de la acción Ecopetrol, mediante la incorporación de técnicas por evolución basadas en Algoritmos de Estimación de la Distribución (EDA's), esto gracias a la relación genética existente entre los genes que conforman la estructura de solución o *individuo*. La función de aptitud por su parte, permitió establecer la relación entre las variables temporales que conforman el precio de la acción, y sus valores de referencia de mercado.

La inclusión de una variable de ruido en el modelo propuesto, la cual es modelada como ruido blanco, permitió evaluar la robustez del modelo frente a la estimación de los valores que conforman la serie de rendimiento de la acción Ecopetrol. Esto se puede observar más ampliamente, en los valores obtenidos por el modelo frente al índice FAC2 e IOA, los cuales muestran que el modelo presenta estabilidad frente a cambios inesperados en las variables que conforman el precio de la acción.

La correcta definición del mecanismo por evolución, fue fundamental para el desempeño del modelo, ya que se logró un ahorro computacional por efecto de la ausencia de operadores genéticos complejos, tal y como ocurre en la evolución diferencial, propia de los algoritmos evolutivos. Sin embargo, y debido a la eficiencia en la estimación por iteración, el modelo incremento el número

de iteraciones llegando hasta 20000, y logrando mejorar ostensiblemente la población de individuos que da solución al problema.

Para mejorar el desempeño del modelo frente al pronóstico y la predicción de la serie de rendimientos de la acción Ecopetrol, es importante el diseño de funciones de aptitud mucho más completas, que involucren de una u otra forma, las métricas definidas por Park (2007) para evaluar el desempeño de un modelo frente a los datos.

El algoritmo de estimación de la distribución implementado (EDA's), se comportó mucho mejor que los algoritmos por evolución tradicional, debido principalmente a la estrategia de modificación de soluciones utilizada, en donde el factor reproductivo entre los individuos, fue determinante para mejorar la eficiencia en el proceso de estimación.

Uno de los elementos fundamentales para mejorar la eficiencia de los mecanismos por evolución basados en los Algoritmos de Estimación de la Distribución (EDA's), está centrado en el desarrollo de nuevos operadores que permitan promover la actividad reproductiva de los individuos dentro de la población, y que de este modo, permitan explorar de una forma más óptima y promisoría, el espacio de solución del problema.

REFERENCES

- [1] Arranz, J. M., & Zamora, M. M. (2011). *Master en Administración Electrónica de Empresas*. Recuperado el 28 de 10 de 2011, de Variables Ficticias en el Modelo de Regresión: <http://ciberconta.unizar.es/leccion/ficticia/variables%20ficticias.PDF>
- [2] Bahi, C. A. (2007). *Modelos de Medición de la Volatilidad en los Mercados de Valores: Aplicación al Mercado Bursátil Argentino*. Buenos Aires.
- [3] Chará Ordoñez, D. E. (2010). *Software de Pronóstico para la Concentración del Material Particulado de Valle de Aburrá*. Medellín: Escuela de Ingeniería de Antioquia.
- [4] Coello, C. (2007). *Introducción a la Computación Evolutiva*. México D.F.: Centro de Investigaciones Avanzadas CINVESTAV.
- [5] Drake, A. E., & Marks, R. E. (1998). *Genetic Algorithms in Economics and Finance: Forecasting Stock Market Prices and Foreign Exchange*. Sydney, Australia.
- [6] Ecopetrol. (13 de Febrero de 2011). *Ecopetrol*. Recuperado el 13 de Febrero de 2011, de <http://www.ecopetrol.com.co/contenido.aspx?catID=258&conID=39930>
- [7] Etxeberria, R., & Larrañaga, P. (2008). *Global Optimization Using Bayesian Networks*. Donostia.
- [8] Fernandez Castaño, H. (2009). Egarch: Un Modelo Asimétrico para Estimar la Volatilidad de Series Financieras. *Revista Ingenierías Universidad de Medellín*, 49-60.
- [9] Hernández Riveros, J. A., & Ospina Arango, J. D. (2009). *Estimación de Parámetros en Modelos de Volatilidad Estocástica Usando Algoritmos Evolutivos*. Medellín: Universidad Nacional de Colombia, Sede Medellín.
- [10] Hernández, J., & Opsina, J. (2010). A multi dynamics algorithm for global optimization. *Mathematical and Computer Modelling (XX)*, XX., doi:10.1016/j.mcm.2010.03.024.
- [11] Larrañaga, P., & Lozano, P. (2002). *Estimation of Distribution Algorithms*. Massachussets: Kluwer.
- [12] Lozano, J., Larrañaga, P., & Bengoetxea, E. (2006). Towards a New Evolutionary Computation: Advances on Estimation of Distribution Algorithms . Springer Verlag.
- [13] Mahfoud, S., & Mani, G. (1996). Financial Forecasting Using Genetic Algorithms. *Applied Artificial Intelligence*, 543-565.
- [14] Ok-Hyun, P., & Min-Gwang, S. (2007). Selection of an appropriate model to predict plume dispersion in coastal areas. *Atmospheric Environment*, 6095-6101.
- [15] Olmeda, E., Velasco, F., & Valderas, J. M. (2007). Caracterización no Lineal y Predicción no Paramétrica en el IBEX35. *Estudios de Economía Aplicada*, 815-842.
- [16] Ospina D'Aleman, F., & Giraldo Sánchez, D. (2009). Aplicación de los modelo GARCH a la estimación del Var de acciones colombianas. *Revista Soluciones de Postgrado EIA (3)* , 11-24.
- [17] Otero, L. A., & Durán, P. (2008). Diseño de un modelo de análisis financiero dinámico (DFA) aplicado al seguro de automóvil español. *Revista Europea de Dirección y Economía de la Empresa*, 63-84.
- [18] Palacio Orozco, J. C., & Lochmüller, C. (2011). *Modelo Neuroborroso para la Evaluación del Riesgo en el Otorgamiento de Créditos de Consumo*. Medellín: Universidad Nacional - Sede Medellín.
- [19] Park , O.-H., & Seok , M.-G. (2007). Selection of appropriate model to predict plume dispersion in coastal areas . *Atmospheric Environment (41)*, 6095-6101, doi: 10.1016/j.atmosenv.2007.04.010.
- [20] Patiño, A., Palacio, C., Peña P , A., Lochmuller , C., Murillo, J., & Pérez, M. (2011). Neurofuzzy model risk assessment in the placement of consumer credit. *VIII Coloquio Internacional de Estadística*. Medellín, Colombia: Universidad Nacional de Colombia.
- [21] Peña P., A., Lochmuller, C., Pérez, M., & Murillo, J. (2011). Modelo cualitativo para la asignación de créditos de consumo y ordinario. *Revista de Ingenierías - Universidad de Medellín 10(19)*.
- [22] Spronk, J., & Trinidad Segovia, J. E. (2005). Más de Medio Siglo en Busca de una Teoría Sobre los Mercados de Capitales. *Estudios de Economía Aplicada*, 29-44.
- [23] Trejos Carpintero, A., Nieto Salazar, S., & Carvajal Olaya, P. (2003). Modelo de Predicción de la Acción Ordinaria Cementos Argos. *Scientia et Technica*, 67-72.