

Agradecimientos

El trabajo desarrollado durante los estudios de todo el programa en general y en la elaboración de esta tesis en particular no hubiera sido posible sin la ayuda de una gran cantidad de personas maravillosas. Les estoy profundamente agradecido a todos. Desde mis colegas en el Campus León del Tec de Monterrey hasta todos mis compañeros de estudios en general y en particular a los del área de finanzas. Muchas gracias por todo su apoyo y soporte durante esta etapa de mi vida. Agradezco en particular a Andrés Sierra por las horas de camino que resultaron más gratas con su peculiar estilo de conversación.

Agradezco a los directivos del Campus León por permitirme desarrollarme en este espacio del conocimiento. Obviamente nada es posible sin el apoyo económico y el estímulo profesional.

Por supuesto a mi asesor el Dr. Francisco Javier Cuevas de la Rosa quien me permitió conocer e intentar aprender una serie de particularidades de la ciencia que me han hecho fascinarme con el deseo de aprender más y proponer más tópicos relacionados a éste en el futuro. Así mismo hago mención del apoyo del Centro de Investigaciones en Óptica por brindarme un espacio de reflexión y experimentación.

Mención especial merece el Ing. Jorge Humberto Abarca Winkler quien logró que mis balbuceos computacionales, se convirtieran en programas sólidos y ejecutables.

Pero el trabajo que he desarrollado durante toda mi vida ha sido siempre soportado, apoyado, comprendido, aconsejado, ilustrado y amado por Verónica. Ella es el eje de todo lo que hago, sin ella nada hubiera sido posible. Su calma cuando las cosas no van bien, su organización cuando el caos parece reinar y su energía cuando el desaliento llega, constituye la luz necesaria en medio de la oscuridad.

Nena, muchas gracias por todos estos años de apoyo y soporte.

Y gracias también a César, Chío y Richie. De verdad no tienen una idea de cómo iluminan mi vida, como la hacen mejor y como los he extrañado en cada viaje corto o largo que he hecho durante estos años.

Comparación de tres modelos de algoritmos genéticos, un algoritmo de conteo y un algoritmo voraz a la información de 10 años de los rendimientos de 40 emisoras de la Bolsa Mexicana de Valores

Resumen

Por: J. Eddie César Villegas Zermeño
Instituto Tecnológico y de Estudios Superiores de Monterrey
Campus Ciudad de México
Febrero de 2005

Existen muchas formas de predecir el comportamiento de los mercados financieros de manera experimental, desde los modelos clásicos de pronósticos como lo son los modelos econométricos, las series de tiempo, las relaciones de causalidad y las metodologías de Box-Jenkins; hasta los modelos que aplican la heurística y la volatilidad estocástica y que hacen uso de una gran cantidad de información para intentar una mayor precisión en el pronóstico de los rendimientos de un mercado en particular.

A pesar de esto las aplicaciones prácticas siguen utilizando fundamentalmente promedios móviles-28 para obtener un pronóstico de los rendimientos y tomar sus decisiones de inversión.

La presente disertación propone el uso de las técnicas de la computación suave y la minería de datos para obtener modelos más precisos del rendimiento de las acciones en los mercados financieros internacionales.

De partida se toma la información de 10 años de la Bolsa Mexicana de Valores tomando una muestra de 40 emisoras y con esta muestra se desarrollan portafolios de 10 emisoras que posteriormente se comparan unos con otros hasta encontrar aquel que mayor rendimiento hubiera tenido a lo largo del horizonte temporal bajo estudio.

Se toman diferentes fases de tiempo, a saber: diario, semanal y mensual para determinar cuál medición de tiempo da mayores rendimientos. Sin embargo se ignora la parte práctica de las comisiones.

Posteriormente se aplica una minería de datos a un algoritmo de conteo que efectúa hasta $2^{123,389,622,640}$ combinaciones para encontrar aquella que forma el portafolio de inversión que mayor rendimiento da en el periodo.

El resultado de este algoritmo se compara con los encontrados para las mismas fases de tiempo con un algoritmo genético con tres diferentes medidas de

aptitud a saber: el rendimiento global, la distribución normal y el movimiento geométrico browniano.

Finalmente se hace uso de un algoritmo voraz para encontrar otros portafolios para compara con los obtenido previamente.

Los resultados muestran que los portafolios formados con al algoritmo genético que usa como medida de aptitud la distribución normal generan mayores rendimientos anualizados.

TABLA DE CONTENIDOS

RESUMEN	ii
LISTA DE FIGURAS	vii
LISTA DE ECUACIONES	vii
LISTA DE TABLAS	vii
CAPÍTULO 1: INTRODUCCIÓN.....	1
1.1 Contenido de la tesis.....	2
CAPÍTULO 2: MODELO DE HARRY MARKOWITZ SOBRE PORTAFOLIOS EFICIENTES.....	5
2.1 Hipótesis del modelo de Markowitz.....	5
2.2 Rendimiento esperado de un portafolio.....	5
2.3 Efectos de la diversificación: reducción del riesgo de un portafolio	6
2.4 Frontera eficiente para valores múltiples.....	8
2.5 Preferencias de los inversionistas	9
CAPÍTULO 3: MINADO DE DATOS	11
3.1 Minado de datos y aprendizaje inteligente de las máquinas	12
3.1.1 Aprendizaje inteligente de las máquinas	14
3.1.2 Minado de datos.....	16
3.1.3 Decisiones que involucran juicios	16
3.1.4 Aprendizaje inteligente de las máquinas y estadística ...	17
3.1.5 El minado de datos y las finanzas	18
3.1.6 Enfoques del minado de datos.....	21
CAPÍTULO 4: ANÁLISIS DE CONGLOMERADOS	24
4.1 Medidas de proximidad.....	26
4.2 Algoritmos de agrupamiento	28
4.3 Evaluación de los grupos	30
CAPÍTULO 5: ALGORITMOS GENÉTICOS	32
5.1 Un vistazo histórico a las teorías de la evolución de las especies y de computación evolutiva	33
5.1.1 El origen de las ideas.....	33
5.1.2 El lamarckismo	33
5.1.3 La teoría del germoplasma	34
5.1.4 Russell y Darwin.....	34
5.1.5 La teoría de la combinación.....	35
5.1.6 Las leyes de la herencia de Mendel	35
5.1.7 La teoría de la pangénesis.....	35
5.1.8 La teoría de la mutación.....	36
5.1.9 La teoría cromosómica de la herencia.....	36
5.1.10 Neodarwinismo.....	36
5.1.11 Inspiración biológica	36

5.1.12 Alexander Fraser	37
5.1.13 EVOP	37
5.1.14 Evolución de programas de Friedberg	37
5.1.15 Friedman y la robótica evolutiva	37
5.1.16 Vida artificial	38
5.1.17 La optimización evolutiva de Bremerman	38
5.1.18 La programación evolutiva.....	38
5.1.19 Las estrategias evolutivas.....	39
5.1.20 Los algoritmos genéticos	39
5.1.21 Ecosistemas artificiales.....	39
5.1.22 Programación genética.....	40
5.1.23 Dinámica Evolutiva	40
5.2 Conceptos básicos de algoritmos genéticos	40
5.2.1 El modelo básico de los algoritmos genéticos.....	41
5.2.2 Nomenclatura de la biología utilizada para los algoritmos genéticos	45
5.2.3 Pasos del proceso que realizan los algoritmos genéticos.....	48
5.3 Algoritmos genéticos y finanzas	53
5.3.1 Pronósticos de rendimientos.....	53
5.3.2 Optimización de portafolios	54
5.3.3 Descubrimiento de las reglas de intercambio	56
5.3.4 Optimización de las reglas de intercambio	57
CAPÍTULO 6: ESPECIFICACIONES.....	59
6.1 Selección de la población.....	59
6.2 Selección del tamaño y de las componentes de la muestra	61
6.2.1 Selección del tamaño de la muestra	61
6.2.2 Selección de las componentes de la muestra	63
6.3 Determinación del tamaño del portafolio.....	68
6.4 Escenarios posibles para el análisis de la información.....	68
CAPÍTULO 7: SELECCIÓN DEL PORTAFOLIO QUE MEJOR PREDICE EL COMPORTAMIENTO DE LA BOLSA MEXICANA DE VALORES USANDO UN ALGORITMO DE CONTEO	70
7.1 Diseño del algoritmo de conteo	70
7.2 ¿Qué hace el algoritmo de conteo?.....	71
7.3 Selección del mejor portafolio	73
7.2.1 Escenario diario.....	73
7.2.2 Escenario semanal.....	73
7.2.3 Escenario mensual.....	74
CAPÍTULO 8: SELECCIÓN DEL PORTAFOLIO QUE MEJOR PREDICE EL COMPORTAMIENTO DE LA BOLSA MEXICANA DE VALORES USANDO UN ALGORITMO GENÉTICO CON TRES DIFERENTES MEDIDAS DE APTITUD... 75	75
8.1 El algoritmo.....	75
8.1.1 Resultados del algoritmo genético usando como función de aptitud el rendimiento promedio	78

8.1.2 Resultados del algoritmo genético usando como función de aptitud la distribución normal.....	79
8.1.3 Resultados del algoritmo genético usando como función de aptitud el movimiento geométrico browniano	80
CAPÍTULO 9: COMPARACIÓN DE RESULTADOS, CONCLUSIONES E INVESTIGACIÓN POSTERIOR.....	82
9.1 Comparación de resultados.....	82
9.2 Algoritmo voraz	86
9.3 Conclusiones	87
9.4 Investigación posterior	89
ANEXO 1: CORRIDAS COMPLETAS DE LOS ALGORITMOS GENÉTICOS	91
ANEXO 2: EL MÉTODO DE LA RULETA	94
ANEXO 3: PANTALLAS DE APLICACIÓN ECONOMATICA.....	96
ANEXO 4: LOS ALGORITMOS	105
Diseño del algoritmo de conteo	105
El algoritmo genético	112
REFERENCIAS.....	119

LISTA DE FIGURAS

Figura 2.1 Combinaciones riesgo-rendimiento para un portafolio de inversión	7
Figura 2.2 Frontera eficiente para valores múltiples.....	8
Figura 2.3 Preferencias de los inversionistas	10
Figura 5.1 Representación del algoritmo genético estándar.....	41
Figura 5.2 Reproducción de un gen dominante contra un recesivo	42
Figura 5.3 Representación de la no aplicación de la ley del sorteo independiente	43
Figura 6.1 Diversificación del riesgo	63

LISTA DE ECUACIONES

Ecuación 2.1 Rendimiento promedio de un portafolio	5
Ecuación 2.2 Riesgo de un portafolio	6
Ecuación 2.3 Correlación de un portafolio.....	7
Ecuación 4.1 Correlación producto-momento de Pearson.....	27
Ecuación 4.2 Distancia Euclidiana entre dos puntos.....	27
Ecuación 6.1 Rendimiento de un portafolio.....	60
Ecuación 6.2 Volatilidad de un portafolio	60
Ecuación 6.3 Beta de un portafolio	60

LISTA DE TABLAS

Tabla 6.1 Diversificación del riesgo	62
Tabla 6.2 Muestra de acciones seleccionadas.....	67

LISTA DE FIGURAS

Figura 2.1 Combinaciones riesgo-rendimiento para un portafolio de inversión	7
Figura 2.2 Frontera eficiente para valores múltiples.....	8
Figura 2.3 Preferencias de los inversionistas	10
Figura 5.1 Representación del algoritmo genético estándar.....	41
Figura 5.2 Reproducción de un gen dominante contra un recesivo	42
Figura 5.3 Representación de la no aplicación de la ley del sorteo independiente	43
Figura 6.1 Diversificación del riesgo	63

LISTA DE ECUACIONES

Ecuación 2.1 Rendimiento promedio de un portafolio	5
Ecuación 2.2 Riesgo de un portafolio	6
Ecuación 2.3 Correlación de un portafolio.....	7
Ecuación 4.1 Correlación producto-momento de Pearson.....	27
Ecuación 4.2 Distancia Euclidiana entre dos puntos.....	27
Ecuación 6.1 Rendimiento de un portafolio.....	60
Ecuación 6.2 Volatilidad de un portafolio	60
Ecuación 6.3 Beta de un portafolio	60

LISTA DE TABLAS

Tabla 6.1 Diversificación del riesgo	62
Tabla 6.2 Muestra de acciones seleccionadas.....	67

LISTA DE FIGURAS

Figura 2.1 Combinaciones riesgo-rendimiento para un portafolio de inversión	7
Figura 2.2 Frontera eficiente para valores múltiples.....	8
Figura 2.3 Preferencias de los inversionistas	10
Figura 5.1 Representación del algoritmo genético estándar	41
Figura 5.2 Reproducción de un gen dominante contra un recesivo	42
Figura 5.3 Representación de la no aplicación de la ley del sorteo independiente	43
Figura 6.1 Diversificación del riesgo	63

LISTA DE ECUACIONES

Ecuación 2.1 Rendimiento promedio de un portafolio	5
Ecuación 2.2 Riesgo de un portafolio	6
Ecuación 2.3 Correlación de un portafolio	7
Ecuación 4.1 Correlación producto-momento de Pearson	27
Ecuación 4.2 Distancia Euclidiana entre dos puntos	27
Ecuación 6.1 Rendimiento de un portafolio	60
Ecuación 6.2 Volatilidad de un portafolio	60
Ecuación 6.3 Beta de un portafolio	60

LISTA DE TABLAS

Tabla 6.1 Diversificación del riesgo	62
Tabla 6.2 Muestra de acciones seleccionadas.....	67



Hacemos constar que en la Ciudad de México, el día 25 de febrero de 2005, el alumno:

J. Eddie César Villegas Zermeño

sustentó el examen oral en defensa de la Tesis titulada:

Comparación de Tres Modelos de Algoritmos Genéticos, un Algoritmo de Conteo y un Algoritmo Voraz a la Información de 10 Años de los Rendimientos de 40 Emisoras de la Bolsa Mexicana de Valores

Presentada como requisito final para la obtención del Grado de:

Doctor en Administración

Ante la evidencia presentada en el trabajo de tesis y en este examen, El Comité Examinador, presidido por el Dr. Macario Schettino Yáñez, ha tomado la siguiente resolución:

— **APROBADO** —

Dr. Francisco Javier Cuevas de la Rosa
Director de Tesis

Dr. Miguel Torres Cisneros
Lector

Dr. David Rivera Caballero
Lector

Dr. César Augusto Coutiño Gómez
Lector

Dr. Macario Schettino Yáñez
Director del Programa Doctoral

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY

CAMPUS CIUDAD DE MÉXICO



COMPARACIÓN DE TRES MODELO DE ALGORITMOS
GENÉTICOS, UN ALGROTIMO DE CONTEO Y UN
ALGORITMO VORAZ A LA INFORMACIÓN DE 10 AÑOS
DE LOS RENDIMIENTOS DE 40 EMISORAS DE LA
BOLSA MEXICANA DE VALORES

DOCTORADO EN ADMINISTRACIÓN

TESIS PRESENTADA POR

J. EDDIE CÉSAR VILLEGAS ZERMEÑO

FEBRERO 2005

Capítulo 1

INTRODUCCIÓN

Posiblemente una de las grandes ambiciones del hombre a través de la historia ha sido la de tener la capacidad de predecir el futuro y por lo tanto tener el control de las cosas en los eventos por venir.

A los magos, a los brujos y a los chamanes les hemos conferido la potestad de prever y por lo tanto de poder cambiar el curso de las cosas. Quién no quisiera tener esa bola de cristal mágica y a través de ella poder echar un vistazo al futuro social, político, deportivo y financiero.

Las ciencias y en particular la matemática tienen su mago en la figura de la estadística; a través de ella se pretende predecir el comportamiento y los resultados de eventos inciertos pero que tienen algún significado para algún tópico científico en particular.

Por supuesto la estadística tradicional no se compromete del todo y sólo se atreve a decir que con un $(1-\alpha)100\%$ de confianza ocurrirá tal o cual resultado.

Y sin embargo parecería ser que la posibilidad de predecir el futuro en su aspecto básico y fundamental sigue siendo una apuesta. Así que cuando se toma la decisión de formar un portafolio de inversión, a pesar de los análisis y estudios que se puedan hacer al respecto, esto sigue siendo una apuesta.

Por supuesto, como en las carreras de caballos, siempre se trata de encontrar el mejor caballo, aquel que tenga la mayor fortaleza y *aptitud* para alcanzar en primer lugar la meta; esto se logra analizando el historial y las fortalezas de los animales en conjunto con el medio ambiente que está predominado y de esta manera determinar cuáles son las bestias que dadas estas condiciones tendrían mayor probabilidad de éxito.

De igual manera esto es lo que se pretende en la formación de un portafolio de inversión, encontrar la "bestia" que en las condiciones predominantes sea la que mayor *aptitud* tendrá para resultar vencedora, donde en este caso no significa llegar primero a la meta, sino obtener el mayor rendimiento posible entre el conjunto de portafolios participantes.

El problema se vuelve más complejo porque a diferencia de la carrera de caballos, en este caso los portafolios no son entes individuales, sino que son conjuntos provenientes de diversos lugares (firmas) y que cada una en lo individual tienen cierto comportamiento, pero que en conjunto pueden ver acelerado o atemperado su rendimiento, según la naturaleza de su origen y su comportamiento individual.

Adicionalmente a esto debemos recordar el principio de la racionalidad limitada de Herbert Simon¹ que afirma que la racionalidad de los actores económicos está limitada por su capacidad restringida para recibir, tratar, almacenar y buscar información.

Debido a la racionalidad limitada, la especificación de la mejor decisión en general y del mejor portafolio para nuestro caso, es prácticamente imposible o muy costosa y tardada.

Podemos afirmar entonces que es más fácil para un individuo identificar al caballo que obtendrá el primer lugar en la siguiente carrera en el hipódromo, que identificar aquel portafolio que obtendría el mayor rendimiento en la bolsa².

Por lo tanto es necesario que los sujetos contemos con alguna clase de herramienta de apoyo para que de una manera mucho más rápida y expedita, tomando en consideración todos los factores que rodean el medio ambiente, sea posible identificar los elementos que formen la mejor apuesta al momento de elegir un portafolio de inversión.

1.1 Contenido de la tesis

El presente trabajo hace uso de dos herramientas computacionales para la elección del mejor portafolio de inversión tomando como medida de aptitud el resultado del rendimiento de las acciones en la Bolsa Mexicana de Valores: la minería de datos (*data mining*) y los Algoritmos genéticos; estas herramientas actuarán como nuestros guías en esta búsqueda de predecir el futuro y tener éxito en el intento.

Lo que pretende este estudio es hacer uso de la técnica de la computación evolutiva conocida como Algoritmos genéticos (AG) para encontrar portafolios que pronostiquen el rendimiento de las acciones de la Bolsa Mexicana de Valores y comparar los portafolios encontrados mediante esta técnica contra otros encontrados por un proceso de minería de datos o conteo y de esta forma determinar cuál de las dos metodologías predice mejor el comportamiento del mercado.

En el capítulo 2 se presentan los aspectos fundamentales de la teoría de portafolios de inversión de Harry Markowitz sobre portafolios eficientes.

En el capítulo 3 se presentan las definiciones básicas del minado de datos, la relación entre esta técnica y el aprendizaje inteligente de las máquinas, algunos casos de las decisiones que involucran juicios, la relación entre el aprendizaje

¹ SIMON, Herbert A.; *Naturaleza y Límites de la Razón Humana*, Fondo de Cultura Económica, 1989

² Esto es debido al tamaño del espacio de búsqueda. En el caso de los caballos hay N posibles soluciones, donde N, obviamente es el número de caballos. En el caso del portafolio de inversión existirán aproximadamente $(N-1)(N-2)\dots(N-M)$ posibles soluciones, donde N es el total de acciones y M es el tamaño del portafolio de inversión.

inteligente de las máquinas y la estadística, y una serie de casos de aplicaciones del minado de datos a las finanzas junto con los parámetros que usa el minado de datos cuando de finanzas se trata, finalmente presentamos los enfoques del minado de datos.

En el capítulo 4 presentamos las bases teóricas de uno de los enfoques del minado de datos que es el análisis de conglomerados y que en nuestro caso fue el que se adecuó a los datos con los que teníamos que trabajar, es decir, de la gran cantidad de información que teníamos había que reducirla a un número suficientemente grande para que fuera significativa, pero suficientemente pequeño para que fuera manejable.

Ahí presentamos los fundamentos básicos del análisis de conglomerados que son las medidas de proximidad y los Algoritmos de agrupamiento así como la evaluación de los grupos.

En el capítulo 5 definimos Algoritmos genéticos y utilizamos el trabajo del Dr. Carlos Coello para mostrar la evolución histórica de los Algoritmos genéticos, que sin esta descripción histórica no es posible comprender los conceptos, el alcance y la importancia de los Algoritmos genéticos hoy en día.

Se presentan también como la biología da la base de la nomenclatura usada en la computación evolutiva y se muestran los pasos para realizar un Algoritmo genético. Finalmente se presenta una serie de literatura revisada sobre aplicaciones de Algoritmos genéticos en el campo de las finanzas.

El capítulo 6 muestra las especificaciones de esta tesis, es decir, selección de la población, determinación del tamaño de la muestra, de los componentes de la misma y del tamaño del portafolio.

También se describen los escenarios para el análisis de la información.

El capítulo 7 muestra como se obtiene el mejor portafolio con la información dada utilizando un Algoritmo de conteo, se presenta el diseño del Algoritmo, el programa computacional diseñado para que este se ejecutara y la selección de los mejores portafolios usando este Algoritmo para cada uno de los escenarios previamente definidos.

El capítulo 8 presenta la obtención de los portafolios usando algoritmos genéticos con tres diferentes medidas de aptitud, se presenta la estructura de los algoritmos y los pasos y corridas hasta encontrar los mejores para cada uno de los escenarios definidos.

Se presentan también los resultados obtenidos mediante un algoritmo de los conocidos como "codiciosos" o Greedy.

Finalmente en el capítulo 9 presentamos de manera muy concisa la comparación de resultados, las conclusiones y algunas sugerencias sobre investigación posterior.

Capítulo 2

MODELO DE HARRY MARKOWITZ SOBRE PORTAFOLIOS EFICIENTES³

Harry Markowitz buscó recoger en forma explícita en su modelo, los rasgos fundamentales de lo que podríamos considerar como conducta racional del inversionista, consistente en buscar aquella composición del portafolio que haga máximo el rendimiento para un determinado nivel de riesgo, o bien, mínimo el riesgo para un rendimiento dado.

2.1 Hipótesis del modelo de Markowitz:

El modelo de Markowitz parte de las siguientes hipótesis:

1. El rendimiento de cualquier título o portafolio, es una variable aleatoria cuya distribución de probabilidad para el periodo de referencia es conocido por el inversionista. Se acepta como medida del rendimiento de la inversión la media o la esperanza matemática de dicha variable aleatoria.
2. Se acepta como medida del riesgo la dispersión, medida como la variancia o la desviación estándar de la variable aleatoria que describe el rendimiento, ya sea de un valor individual o de un portafolio.
3. El inversionista elegirá aquellos portafolios con mayor rendimiento y mayor riesgo.

El inversionista se encuentra presionado por dos fuerzas en sentido opuesto: por un lado el deseo de obtener ganancias y por el otro la aversión al riesgo. La selección de una determinada "ganancia-riesgo", dependerá de la mayor o menor aversión al riesgo del inversionista.

2.2 Rendimiento esperado de un portafolio

El rendimiento esperado, como señalamos anteriormente, puede encontrarse mediante una medida estadística: media o esperanza matemática.

El rendimiento medio de un portafolio o cartera es igual a un promedio ponderado de los rendimientos esperados para los valores que comprenden el portafolio. Entonces:

$$R_p = \omega_1 R_1 + \omega_2 R_2 + \dots + \omega_{n-1} R_{n-1} + \omega_n R_n$$

$$R_p = \sum_{i=1}^n \omega_i R_i$$

Ecuación 2.1: rendimiento promedio de un portafolio

³ Material tomado de Collatti, M.B. *Teoría de Carteras*. Programa de formación 2002. Bolsa de Comercio de Rosario. 2002. Rosario Argentina, basado en Van Horne, J. C. *Administración Financiera* 10a edición. Prentice may Hispanoamericano.

donde:

R_p : rendimiento del portafolio de inversión

R_i : rendimiento del activo i

W_i : porcentaje del fondo invertido en el activo i

2.3 Efectos de la diversificación: reducción del riesgo de un portafolio

Markowitz centró su atención en la diversificación de portafolios y mostró cómo un inversionista puede reducir el riesgo de un portafolio eligiendo valores cuyas oscilaciones no sean paralelas, es decir, valores que tengan poca relación, de manera que unos aumenten su valor, mientras otros experimenten bajas en sus precios.

Esto puede deberse a sensibilidades opuestas ante determinados factores macroeconómicos.

La intención es que en algunos momentos en la caída en el valor de una de las acciones quede compensada por la subida en el precio de la otra. Por lo tanto, que exista una oportunidad de reducir el riesgo por medio de la diversificación.

El riesgo de un portafolio no es la suma de los riesgos de los valores que lo componen, sino que existe otra variable vinculada al riesgo total y es la *covariancia* de los rendimientos.

Esta es una medida del grado al que se espera van a variar juntas, en lugar de independientes una de la otra.

Entonces la desviación estándar o riesgo de un portafolio es:

$$\sigma_p = \sqrt{\sum_{i=1}^n \sum_{j=1}^n \omega_i \omega_j \sigma_{(R_i, R_j)}}$$

Ecuación 2.2: riesgo de un portafolio

donde:

n : total de valores en el portafolio

W_i y W_j : proporción total de fondos invertidos en i y j .

$\sigma_{(R_i, R_j)}$: covariancia de los rendimientos posibles para i y j .

Las dos sumatorias significan que podemos considerar las covariancias para todas las combinaciones posibles en pares de valores en el portafolio.

Las covariancias son útiles en términos del cálculo de los coeficientes de correlación entre los valores del portafolio, este coeficiente toma valores de -1 hasta 1 y en donde el valor extremo de -1 significa que se tiene una perfecta correlación negativa, mientras que el valor de 1 refleja una

perfecta correlación positiva; el valor de 0 indicaría la ausencia total de correlación.

Este coeficiente es el cociente de la covariancia entre el producto de las variancias individuales:

$$\rho = \frac{\sigma_{(R_1, R_2)}}{\sigma_{R_1} \sigma_{R_2}}$$

Ecuación 2.3: correlación de un portafolio

La gráfica siguiente muestra las combinaciones riesgo-rendimiento para diferentes coeficientes de correlación, desde cuando toma el valor de 1 y hasta que toma el valor de -1.

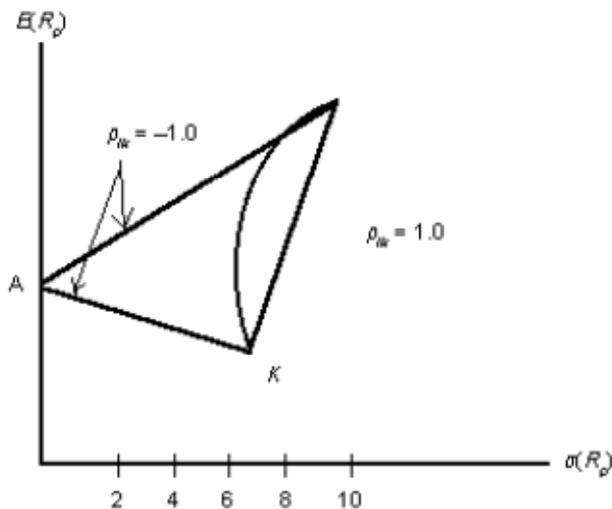


Figura 2.1: Combinaciones riesgo-rendimiento para un portafolio de inversión

La línea IK muestra las combinaciones posibles de riesgo-rendimiento cuando el coeficiente de correlación es igual a 1.

La curva IK muestra las combinaciones posibles de riesgo-rendimiento cuando existe una correlación entre 1 y -1.

Las líneas AK y AI muestran las combinaciones posibles de riesgo y rendimiento cuando la correlación es -1.

La curva IK se denomina conjunto de oportunidades y representa aquellas combinaciones de activos disponibles en el mercado. Siempre tendrá una forma similar, aunque se trabaje con portafolios que tengan k activos.

Podemos apreciar el efecto de la diversificación al comparar la línea curva con la línea recta.

La línea recta describirá el conjunto de oportunidades si predomina una perfecta correlación (es decir, si el coeficiente de correlación fuera igual a uno).

Al disminuir la correlación entre los valores, la línea curva se distancia de la línea recta.

Observamos cómo la línea curva domina a la recta, ya que sobre la primera, para un determinado nivel de riesgo se obtienen mayores rendimientos y para un determinado nivel de rendimientos, el riesgo es menor.

Con un coeficiente de correlación muy pequeño, es evidente un efecto considerable de la diversificación por la distancia entre las dos líneas.

Aunque parezca imposible, podemos reducir la desviación estándar respecto de la esperada con una inversión del 100% en un valor, si invertimos en un valor de mayor riesgo.

Este resultado contra intuitivo se debe al efecto de la diversificación, donde los rendimientos inesperados de un valor quedan balanceados por movimientos opuestos de los rendimientos del otro valor.

2.4 Frontera eficiente para valores múltiples

En la práctica, no es común que exista la limitación de invertir en sólo dos valores sino que se forman portafolios de múltiples valores que pueden adquirirse en el mercado. A continuación graficamos la frontera eficiente para valores múltiples:

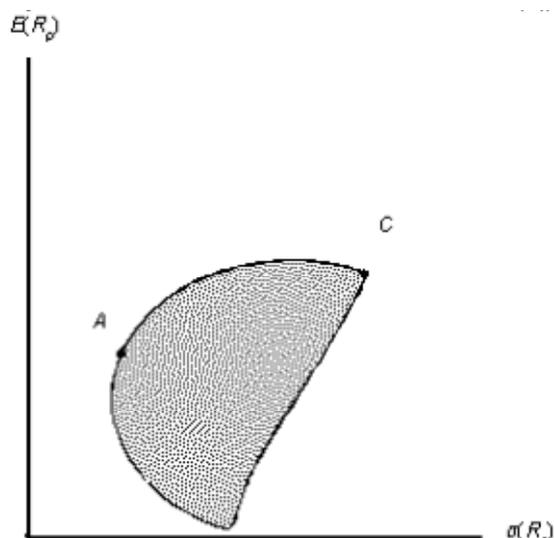


Figura 2.2: frontera eficiente para valores múltiples

A la línea sólida se le conoce como el conjunto de eficiencia, y el punto A es el comienzo de este conjunto eficiente ya que es la del portafolio de mínima variancia.

Los puntos sobre el interior representan combinaciones de riesgo y rendimiento ofrecidas por diferentes títulos individuales, mientras que la línea sólida representa los portafolios finales que se pueden crear provenientes de los activos individuales disponibles en el mercado.

Combinando estos títulos en diferentes proporciones se puede obtener una amplia gama de posibilidades de riesgos y rendimientos esperados.

Si desea aumentar el rendimiento esperada y reducir la desviación estándar, estará interesado únicamente en aquellos portafolios que se encuentren sobre la curva que va desde A hasta C.

Harry Markowitz los llamó *Portafolios Eficientes*.

A partir de aquí, la elección del portafolio dependerá del grado de aversión al riesgo del inversionista.

2.5 Preferencias de los inversionistas

Si bien es verdad que los inversionistas buscan rendimientos esperados altos y menores riesgos, no podemos precisar que portafolio preferirá un determinado inversionista, ya que esto depende de la actitud frente al riesgo del mismo.

Si es un inversionista racional elegirá una que se encuentre sobre el portafolio eficiente descrita anteriormente.

No obstante todavía quedan muchos portafolios factibles para recoger.

Un inversionista atrevido o amante del riesgo, quizá este dispuesto a correr altos riesgos para obtener mayores rendimientos.

Otro, conservador, preferirá arriesgar menos, sacrificando rendimientos futuros.

La preferencia de los inversionistas están representadas a través de curvas construidas en forma tal que cada curva individual representa diferentes combinaciones de riesgo/rendimiento que sean igualmente atractivas para un inversionista. Por lo que esta clase de curvas se conocen como *curvas de indiferencia*.

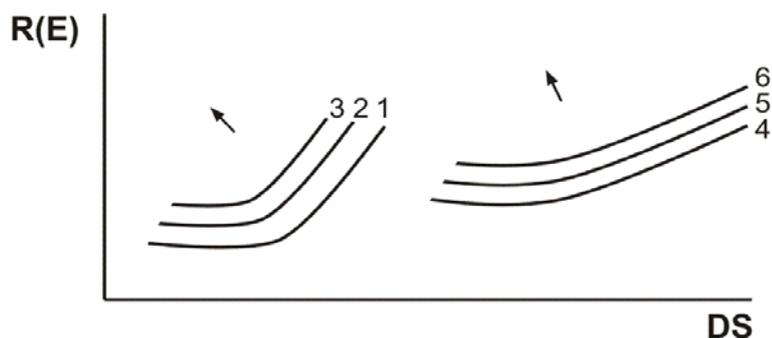


Figura 2.3: preferencias de los inversionistas

Las curvas 1, 2 y 3 pertenecen a un inversionista conservador. El inversionista menos conservador posee curvas de indiferencia con menor pendiente como las curvas 4, 5 y 6, ya que estará dispuesto a afrontar riesgos mayores para incrementar sus rendimientos.

El inversionista conservador, encontraría igualmente atractivo todos los puntos sobre la curva 2, pero seguramente preferirá estar en cualquier punto de la curva 2 y no estar sobre la curva 1.

En términos de la gráfica, el inversionista conservador preferirá encontrarse sobre la curva más alta que fuera obtenible, lo mismo es cierto para el inversionista atrevido.

Sin embargo, el trabajo del inversionista no culmina eligiendo la curva más alta, sino que deberá atenerse a las posibilidades disponibles en el mercado.

Quizás no sea posible para estos inversionistas alcanzar las curvas más altas.

Si podemos determinar un grupo de preferencias que estén implícitas mediante las curvas de utilidad y se conoce la información sobre las oportunidades de inversión que están disponibles para los inversionistas, resulta entonces posible determinar qué oportunidades de inversión seleccionarán realmente los distintos inversionistas.

Capítulo 3

MINADO DE DATOS

Witten⁴ (2000) menciona que la fertilización in Vitro del ser humano implica el recoger varios huevos de los ovarios de una mujer, que, después de la fertilización con la esperma del donante, producen varios embriones.

Algunos de éstos se seleccionan y se transfieren al útero de la mujer.

El problema es seleccionar los "mejores" embriones para utilizar los que tengan más probabilidades de sobrevivir.

La selección se basa en alrededor de sesenta características registradas del embrión en cuanto su morfología, folículo y la muestra de la esperma.

El número de características es lo suficientemente grande, lo que hace difícil que un embriólogo puede determinarlas todas simultáneamente y correlacionar los datos históricos con el resultado crucial de si ese embrión resultó o no en un niño vivo.

En un proyecto en Inglaterra, el aprendizaje inteligente de las máquinas⁵ se está investigando como técnica para hacer la selección, usando como datos del entrenamiento sus expedientes el historial de embriones y de sus resultados.

Otro ejemplo que presenta Witten⁶ es el de que cada año, lecheros en Nueva Zelanda tiene que tomar una difícil decisión: qué vacas deben conservar en su manada y cuáles a vender a un matadero.

Típicamente un quinto de las vacas en una manada de vacas se entresaca cada año cerca del final de la estación de ordeño mientras que disminuyen las reservas de la alimentación. Cada historia de la crianza de la vaca y de la producción de leche influye en esta decisión.

Otros factores incluyen edad (una vaca se está acercando al final de su vida productiva en ocho años), problemas de salud, mal historia de fertilidad, rasgos indeseables de temperamento (patear cercas, saltarlas), y no estar cargada con becerro para la estación siguiente.

⁴ WITTEN, I. (2000). *Data mining. Practical Machine Learning Tools and Techniques wit Java Implementation*. Morgan-Kaufman. San Francisco. CA. USA.

⁵ En este trabajo denominamos *Aprendizaje inteligente de las máquinas* al concepto conocido en inglés como *Machine Learning*. El adjetivo inteligente lo incluimos para hacer énfasis en el hecho de que la máquina (computadora) de algún modo aprende los procesos de raciocinio y toma de decisiones y no sólo un aprendizaje por repetición.

⁶ *Ibíd.*

Cerca de setecientas cualidades para cada millón de vacas se han registrado en los últimos los años.

Se están investigando técnicas de aprendizaje inteligente de máquinas como manera de comprobar qué factores son considerados por los *granjeros* para no automatizar la decisión, sino propagar sus habilidades y experiencia a otras.

Vida y muerte. Familia y negocio. El aprendizaje inteligente de las máquinas es una burbujeante nueva tecnología para el conocimiento del minado de datos, una tecnología que mucha gente está comenzando a tomar seriamente.

3.1 Minado de datos y aprendizaje inteligente de las máquinas

Nos agobian los datos. La cantidad de datos en el mundo, en nuestras vidas, parece seguir sin parar y continuar incrementándose sin ninguna señal de su fin. Las computadoras personales omnipresentes hacen demasiado fácil el guardar cosas que previamente nosotros habríamos desechado.

La electrónica registra nuestras decisiones, nuestras opciones en el supermercado, nuestros hábitos financieros, nuestras venidas e idas.

Luchamos por nuestro lugar en el mundo, cada lucha es un registro en un banco de datos. El World Wide Web nos agobia con información; entretanto, cada opción que nosotros hacemos se graba.

Y todas estas son simplemente opciones personales: ellos tienen colegas innumerables en el mundo de comercio e industria. Nosotros testificaremos a la brecha creciente entre la generación de datos y nuestra comprensión de ellos.

Así como la cantidad de datos generada aumenta inexorablemente, la proporción de la gente que los “entiende” disminuye de manera alarmante.

Quedando escondida en todos estos datos se encuentra información potencialmente útil que raramente se hace explícita o se le toma ventaja alguna ventaja.

No es nada nuevo la búsqueda de patrones en los datos.

Las personas han estado buscando tendencias desde que la vida humana empezó.

Los cazadores buscan patrones de conducta en la migración animal, los granjeros buscan modelos en el crecimiento de la cosecha, los políticos buscan tendencias en la opinión del votante, los amantes buscan patrones en las respuestas de su pareja.

El trabajo de un científico (como un bebé) es darle sentido a los datos, para

descubrir las tendencias que gobiernan los trabajos mundiales físicos y encapsularlos en teorías que pueden usarse para predecir lo que pasará en nuevas situaciones.

El trabajo del empresario es identificar oportunidades, es decir, modelos en la conducta que puedan convertirse en un negocio provechoso para después explotarlo.

En el minado de datos, los datos son almacenados electrónicamente y la búsqueda es automatizada o al menos tratada por una computadora. Esto no es particularmente nuevo.

Economistas, estadísticos, expertos en pronósticos e ingenieros en comunicación han trabajado desde hace tiempo con la idea de que los patrones en los datos pueden ser identificados de manera automática, además de ser validados para ser usados en una predicción.

Lo que es nuevo es el incremento en las oportunidades para encontrar dichas tendencias en los datos.

El innegable crecimiento en las bases de datos en los últimos años, pone al minado de datos como una nueva tecnología para los negocios.

Se ha estimado que el aumento de los datos almacenados en las bases de datos del mundo se duplica cada veinte horas.

Conforme el mundo crece en complejidad, y nos agobia con los datos que genera, el minado de datos se convierte en nuestra única esperanza para traer a relieve las tendencias que esconden.

Datos inteligentemente analizados son un recurso valioso. Pueden desembocar desde nuevos descubrimientos y nuevos enfoques hasta ventajas competitivas.

El minado de datos sirve para la resolución de problemas, al analizar datos ya capturados en bases de datos.

En el mundo competitivo de hoy, es decir en las empresas ya sean comerciales o de servicios, los datos son la materia prima que encienden el crecimiento de las empresas, claro siempre y cuando esos datos puedan ser minados.

El minado de datos se define como el proceso de descubrir tendencias en los datos.

El proceso debe de ser automático o más usualmente semi-automático. Las tendencias descubiertas deben de ser significativas para que otorguen una ventaja, usualmente del tipo económico.

Los datos se presentan invariablemente en cantidades sustanciosas.

Pero ¿cómo se expresan estas tendencias? Los patrones útiles nos permiten hacer predicciones no triviales en nuevos datos.

Existen dos extremos para expresar un patrón o tendencia: como una caja negra cuyo contenido es efectivamente incomprensible y una caja transparente cuya construcción revela la estructura del patrón.

Suponemos que las dos pueden resultar en muy buenas predicciones.

La diferencia recae en que si los patrones que han sido minados se representan en términos de una estructura, pueden ser examinados, razonados, y utilizados para informar en decisiones futuras.

Aquellos son llamados patrones estructurales, porque capturan la estructura de la decisión de una manera explícita.

En otras palabras, ayudan a explicar algo sobre los datos.

3.1.1 Aprendizaje inteligente de las máquinas

¿Qué es el aprendizaje? ¿Qué es el aprendizaje inteligente de las máquinas?

Estas son preguntas filosóficas, pero de cualquier forma vale la pena invertir unos cuantos momentos para clarificar estos conceptos fundamentales, para observar lo capciosos que pueden tornarse, antes de ver el aprendizaje inteligente de las máquinas de una manera práctica.

Nuestro diccionario define "aprender"⁷ como el adquirir conocimientos mediante el estudio, ejercicio, experiencia o ser enseñados; el enterarse por información o por observación; el guardarse en la memoria; el ser informado de; el recibir instrucciones.

Estos significados tienen algunos defectos cuando se trata de las computadoras.

Para los primeros dos, es virtualmente imposible el probar lo que el aprendizaje ha logrado o no.

¿Cómo saber si una máquina ha "aprendido de"?

Probablemente ni siquiera es lógico hacernos esta pregunta; si lo fuera no estaríamos probando su habilidad de aprender, sino más bien su habilidad de responder preguntas.

⁷ <http://www.diccionarios.com>

Lo mismo ocurre con el aprendizaje por observación, sería una respuesta del tipo filosófico la que habría de surgir de ésta premisa.

En lo que respecta a los últimos dos significados, es muy fácil denotar lo que significan en términos humanos, el “guardarse en la memoria” y el “recibir instrucción”, pero para aplicarlo al aprendizaje inteligente de las máquinas podría resultar un poco difícil de implementarlo.

Realmente podemos recibir instrucción, sin obtener beneficios.

Podemos comernos a la memoria o ser informados, sin poder aplicar el nuevo conocimiento a situaciones nuevas.

Anteriormente definimos de manera operativa al minado de datos como el proceso de descubrir tendencias o patrones, de manera automática o semi-automática, en grandes cantidades de datos y que estos patrones deberían de ser útiles.

Una definición operativa puede, así mismo, ser formulada para el aprendizaje.

Él cómo las cosas o seres aprenden cuando cambian su comportamiento en una manera que los hará desempeñarse mejor en el futuro.

Esto une al aprendizaje con el desempeño, en lugar del conocimiento. Usted puede probar al aprendizaje al observar el comportamiento y comparándolo con el comportamiento pasado.

Este es un tipo de definición más objetivo, que al parecer también es más satisfactorio.

Pero persiste un problema. El aprendizaje es un concepto abstracto. Muchas cosas cambian su comportamiento en formas que hacen tener un mejor desempeño en el futuro, aun así no querríamos decir que realmente aprendieron.

Un buen ejemplo es una sandalia cómoda. ¿Ha aprendido la forma de tu pie?

Ciertamente ha cambiado su comportamiento para desempeñarse mejor como sandalia que otra cosa, sin embargo no querríamos llamarle aprendizaje.

En el lenguaje de diario, usualmente utilizamos la palabra “entrenamiento” para denotar un aprendizaje del tipo práctico.

Entrenamos a los animales e inclusive a las plantas, aunque sería estirar un poco el sentido de la palabra hablar de entrenar a las sandalias, que

son objetos inertes.

Pero aprender es diferente, el aprender implica propósito. Algo que aprende lo debe de hacer de manera intencional.

Es por ello que no querríamos decir que una viña no aprendió a crecer en un viñedo, querríamos mejor decir que fue entrenado.

Aprender sin propósito es meramente entrenar. O, más al punto, en el aprendizaje el propósito es el aprendiz, mientras que en el entrenamiento el propósito es el maestro.

Al hacer un examen más cercano la segunda definición de aprendizaje, de manera operacional y orientada al desempeño, también trae sus propios problemas cuando la tratamos de aplicar a las computadoras.

Para decidir si algo realmente ha aprendido, necesitamos ver si tuvo la intención de hacerlo, si hubo algún motivo o propósito incluido.

Lo hace difícil aplicar a las máquinas, porque es poco claro si se pueden comportar con algún propósito propio.

Es por ello que las discusiones filosóficas de lo que realmente se quiere decir con "aprendizaje" y con "intención" y "propósito" están plagadas de dificultades.

3.1.2 Minado de Datos

Estamos interesados en técnicas para encontrar y describir patrones estructurales en los datos, como una herramienta para ayudarnos a explicar los datos y hacer predicciones de ellos.

Muchas de las técnicas de aprendizaje buscan descripciones estructurales de lo que se aprende, descripciones que pueden resultar bastante complejas y que típicamente son expresadas como un conjunto de reglas como las descritas porque pueden ser comprendidos por la gente, estas descripciones sirven para explicar lo que ha sido aprendido, para explicar las bases de las nuevas predicciones.

La gente frecuentemente utiliza al minado de datos para obtener conocimiento, no solo para predecir.

Ganar conocimiento de los datos ciertamente suena como una buena idea que uno puede hacer.

3.1.3 Decisiones que involucran juicios

Cuando usted pide un préstamo, debe de llenar un cuestionario

respondiendo información personal y financiera relevante.

La información se utiliza por la compañía acreedora como la base para su decisión para determinar si hace o no la transacción.

Las decisiones típicamente se hacen en dos etapas.

Primero, se utilizan métodos estadísticos para determinar casos claro de "aceptar" o "rechazar".

Los casos que no son tan claros necesitan un juicio humano.

Por ejemplo, una compañía que presta dinero utiliza el procedimiento de decisiones estadísticas para calcular el parámetro numérico basado en la información proporcionada en el cuestionario.

Los solicitantes son aceptados si el parámetro excede un valor preestablecido y rechazadas si caen por debajo.

Esto generalmente cubre el 90% de los casos, el 10% restante son decididos por los oficiales de la institución crediticia.

Al examinar datos históricos para determinar si el solicitante realmente pago los prestamos adquiridos, no sirvió de nada con los que apenas cubrían los requisitos ya que la mitad de los solicitantes que fueron acreditados con un préstamo cayeron en morosidad.

Mientras sería tentador simplemente negar el crédito a los clientes que apenas califican, los profesionales de la industria de crédito señalaron que estos clientes son los de mayor condición crónica volátil y se constituyen como los mejores clientes, claro sería excelente dicen determinar de manera confiable si serán capaces de pagar el principal en el futuro.

También se necesita una conciliación de intereses en donde el contador no quiere cartera mala y el ejecutivo de ventas no quiere renunciar a los negocios.

3.1.4 Aprendizaje inteligente de las máquinas y la estadística

¿Cuál es la diferencia entre aprendizaje inteligente de las máquinas y la estadística?

Algunos cínicos, explotando el interés comercial en ésta área, igualan estos conceptos.

En verdad, no debe buscar una línea divisora entre la estadística y el minado de datos, porque realmente existe una continua y multidimensional con las técnicas del análisis de datos.

Algunas derivan de las habilidades enseñadas con los cursos de estadística estándar, y otros están más asociadas con el tipo de aprendizaje inteligente de las máquinas que se ha originado conforme la ciencia de las computadoras ha avanzado.

Históricamente, los dos lados han gozado de tradiciones diferentes. Si se nos forzará a hacer énfasis en una diferencia, señalaríamos que la estadística se enfoca más en la prueba de hipótesis, y el aprendizaje inteligente de las máquinas se enfoca más en formular procesos de generalización en la búsqueda de posibles hipótesis.

Pero esta es una simplificación a groso modo: la estadística se compromete con más aspectos aparte de la prueba de hipótesis y muchas técnicas de aprendizaje inteligente de las máquinas no incluyen la búsqueda de todas.

En el pasado, esquemas muy similares fueron desarrollados en paralelo al aprendizaje inteligente de las máquinas y estadística. Uno es la inducción al árbol de decisión.

Cuatro estadísticos Breiman⁸ et al. (1984) publicaron un libro llamado "Clasificación y Árboles de Regresión" en la mitad de los 80's, mientras que en los 70's y principio de los 80's investigadores prominentes del aprendizaje inteligente de las máquinas, por ejemplo J. Ross Quinlan⁹ (1993), estaba desarrollando sistemas para inferir árboles de clasificación dados algunos datos.

3.1.5 El minado de datos y las finanzas

Las técnicas del minado de datos se aplican en las finanzas para descubrir patrones y tendencias escondidas en las bases de datos financieras.

La verdadera discusión radica en cómo separar tendencias y patrones reales de los imaginarios.

Como ya hemos dicho el minado de datos proviene de dos fuentes principales: la administración de las bases de datos y las tecnologías del aprendizaje inteligente de las máquinas.

El objetivo de este aprendizaje inteligente de las máquinas es construir programas de computadora que automáticamente mejoren con la experiencia.

⁸ BREIMAN, L., FRIEDMAN, J., STONE, C.J., OLSHEN, R.A. (1984) *Classification and Regression Trees*. Kluwer Academic Publishers. Boston, MA. USA.

⁹ ROOS QUINLAN, J. (1993). *C4.5: Programs for Machine learning*. Morgan-Kaufman. San Mateo, CA. USA.

El pronóstico financiero ha sido ampliamente estudiado como un caso de predicción de análisis de series de tiempo.

La dificultad asociada a esto se debe fundamentalmente a problemas de no-linealidad, no-estacionaridad y de distribución de los datos.

Esto por lo tanto conduce a que sea necesaria la aplicación de técnicas adaptivas de pronósticos.

Parámetros

Los parámetros utilizados por el minado de datos en las finanzas son:

1. Tipo de datos: hay dos grupos principales de datos, atributos o relaciones.
2. Conjunto de datos: existen también dos grandes grupos, la serie de tiempo o bien el uso de todas las variables que pueden influir en dicha serie de tiempo.
3. Algoritmo matemático: este sería el método o modelo a usar de una variedad de modelos estadísticos, redes neuronales y metodologías lógicas. Sin embargo la combinación de diferentes modelos muchas veces resulta en un mejor desempeño que el que proporcionan modelos individuales.
4. Relación funcional: los métodos del minado de datos suponen la existencia de una relación funcional para ser modelada.

Las redes neuronales son el método más común en los pronósticos de los mercados financieros.

Los datos financieros son frecuentemente representados como series de tiempo de una variedad de atributos tales como el precio de las acciones y los índices.

La expectativa de descubrir reglas dinámicas en las finanzas se basa en una idea que se toma prestada de la física, en donde se establece que el comportamiento de una molécula no es predecible, pero el comportamiento de un conjunto de moléculas (en forma de gas por ejemplo) si lo es.

De esta forma no es predecible el comportamiento de un operador individual en el mercado, pero si es posible que exista una regla que gobierne el comportamiento del mercado global en su conjunto.

Lo anterior parece ir en contra de la *Teoría de la eficiencia del mercado* que establece que es prácticamente imposible fijar un modelo de pronósticos a largo plazo usando la información histórica del mercado accionario, pues si el mercado presenta cierta clase de regularidad entonces alguien obtendría ventaja de dicha regularidad y esta desaparecería; en otras palabras la evolución de los precios de cada variable económica es un *random walk*.

Sin embargo esta teoría no excluye que ciertas condiciones condicionales de corto plazo escondidas puedan existir.

Estas regularidades no trabajan para siempre y de hecho deberían ser corregidas frecuentemente.

Drake¹⁰ (1997) ha mostrado que los datos financieros no son aleatorios y que la hipótesis del mercado eficiente es meramente un subconjunto de una mayor hipótesis del mercado caótico.

El minado de datos crea herramientas que pueden ser utilizadas para descubrir patrones y tendencias condicionales muy sutiles de una amplia variedad de datos financieros.

Las aplicaciones del minado de datos a las finanzas tiene una gran variedad de formas y enfoques, por ejemplo Weber¹¹ (2002) afirma que la contabilidad gubernamental está cambiando para aprovechar los adelantos tecnológicos.

Las oficinas, alguna vez lleno de contadores de nivel medio que hacían las entradas de contabilidad de fondos, está viendo el advenimiento de sistemas financieros que rastrear y controlan los fondos automáticamente.

Las auditorias hoy día, no se basan exclusivamente en la revisión de estados financieros, sino en el estudio de sistemas de bases de datos y del flujo de información de las empresas.

Por su parte Saarevirta¹² (1998) afirma que la era de información actual requiere que las compañías utilicen todos los recursos a su disposición para competir eficazmente y hacer crecer las ganancias de los accionistas.

Un recurso clave que puede usarse más eficazmente son los datos que las organizaciones capturan como un derivado de operar sus negocios.

Un proceso que las compañías pueden usar para transformar sus datos operativos en información de soporte para la toma de decisiones es la minería de los datos

¹⁰ DRAKE, K., KIM, Y. (1997). "Abductive Information Modeling Applied to Financial Time Series Forecasting. Finance & Technology Publishing.

¹¹ WEBER, C.M. (2002) "Getting it right the first time: Accounting, auditing, financial systems and the federal government". The Journal of Government Financial Management. Alexandria, VA.USA.

¹² SAARENVIRTA, G. (1998) "Data mining to improve profitability". The Management Accounting Magazine. Hamilton, N.J. USA.

Vojinovic¹³ et al. (2001) describen una de las relativamente nuevas técnicas de minería de datos que puede usarse para pronosticar las series de tiempo del proceso del tipo de cambio.

Los resultados son significativamente mejores que la predicción de un modelo tradicional lineal autorregresivo.

McQueen¹⁴ et al. (1997) hace una comparación de rendimientos de 1946 a 1995 en un portafolio que toma el Dow-10 contra otro portafolio formado con el Dow-30 y muestran que el Dow-10 obtiene rendimientos significativamente superiores al Dow-30.

Wu¹⁵ (2001) presenta la aplicación del minado de datos al análisis de las relaciones entre las divisas de diferentes países y verifica como afectan ciertos eventos de una divisa al valor relativo de otra divisa.

Lawera¹⁶ (1999) presenta una serie de modelos que capturan las variancias no estacionarias y dependientes en los rendimientos de los futuros de gas natural. Se exploran modelos univariados y multivariados basados en las metodologías ARIMA y Markov.

3.1.6 Enfoques del minado de datos

Podemos hablar de dos grandes enfoques en el minado de datos:

1. Aprendizaje supervisado.
2. Modelos de agrupamiento.

Zighed¹⁷ (1996) presenta ambos modelos con las siguientes definiciones y diagramas:

- $W=\{w\}$, una muestra conocida como la muestra de entrenamiento tomada de una población . Cada individuo w en W es conocido como ejemplo de entrenamiento.
- $X(w)$, el estado de las n variables conocidas como atributos para cada ejemplo de entrenamiento w .
- $Y(w)$, la función meta que asigna el valor meta para cada ejemplo de entrenamiento w . Los valores de $Y(w)$ son llamados clases si ellos

¹³ VOJINOVIC, Z., KECMAN, V., SEIDEL, R. (2001). "A data mining approach to financial time series modeling and forecasting". International Journal of Intelligent Systems in Accounting, Finance and Management. Chinchester. NZ.

¹⁴ McQUEEN, G., SHIELDS, K., THORLEY, S.R. (1997). "Does the "Dow-10 investment strategy" Beat the Dow statistically and economically?" Financial Analysts Journal. Charlottesville, NC. USA.

¹⁵ WU, Q. (2001). "Data mining and Knowledge discovery in financial research. Empirical investigation under currency". McGill University. Montreal CA.

¹⁶ LAWERA, M. (1999). "Future prices: Data Mining and Modeling Approaches". Rice University. Houston, TX. USA.

¹⁷ ZIGHED, DA. (1996). Sipina-W, <http://eric.univ-lyon2.fr/~ricco/sipina.html>. Universitié Lumière, Lyon. FR.

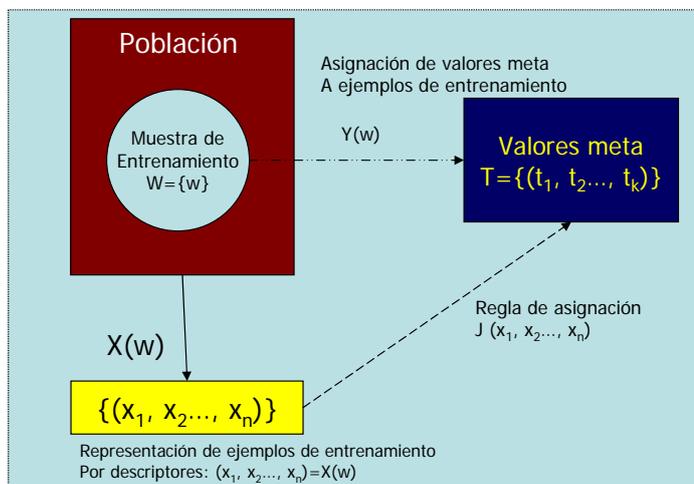
representan una clasificación de ejemplos de entrenamiento.

- La intención es encontrar una regla o modelo J que sea capaz de predecir el valor meta de la función meta $Y(w)$.

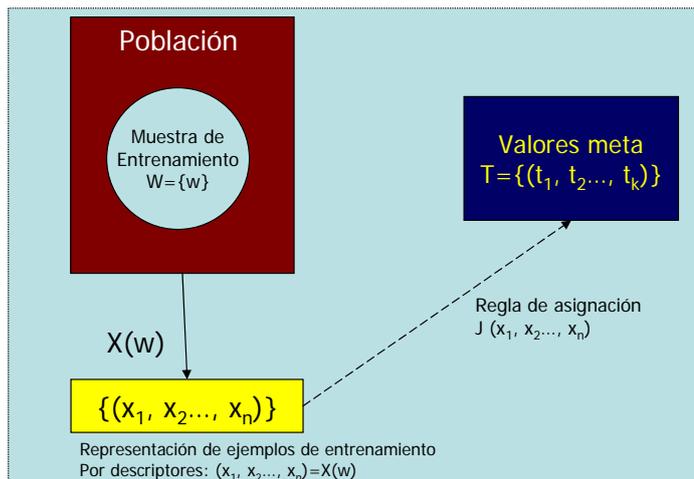
En el caso del modelo supervisado la selección de los valores meta a los ejemplos de entrenamiento es totalmente definida por el investigador, mientras que en el caso del cluster esto no es así pues se deja que se asignen grupos mediante alguna o algunas técnicas del análisis de conglomerados.

Con tantas metas puede resultar un poco confusa la explicación, así que un diagrama ayudaría mucho a clarificar estos aspectos:

Aprendizaje supervisado



Modelos de agrupamiento



A estos modelos genéricos es posible aplicar una serie de técnicas que originalmente han sido tratadas en otras áreas de la estadística o la informática, pero que en este caso se adaptan a los modelos del data mining.

Entre las técnicas del minado de datos se encuentran:

- **Redes neuronales artificiales:** modelos predecibles no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica.
- **Árboles de decisión:** estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Árboles de Clasificación y Regresión (CART: Classification And Regression Tree) y Detección de Interacción Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection)
- **Regla de inducción:** la extracción de reglas if-then de datos basados en significado estadístico.
- **Algoritmos genéticos:** técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución. Para este trabajo esta técnica en particular es tan importante que será desarrollada de manera profunda en el capítulo 4.
- **Análisis de conglomerados:** una técnica que hace grupos tomando individuos buscando los más parecidos o los menos diferentes. En particular esta técnica también es muy importante que será desarrollada de manera profunda en el capítulo 4.

Para este trabajo las técnicas que usaremos será el análisis de conglomerados y los Algoritmos genéticos. Los sustentos conceptuales de los presentamos en los capítulos siguientes.

Capítulo 4

ANÁLISIS DE CONGLOMERADOS¹⁸

No obstante que el viejo adagio afirma que los opuestos se atraen, en la práctica parece ser más evidente que los semejantes son aquellos que se atraen.

Las aves de plumajes similares y muchos otros objetos animados e inanimados que tienen características similares tienden a agruparse¹⁹.

Se pueden formar grupos de objetos similares usando uno de los procedimientos más viejos de la minería de datos conocido como análisis de conglomerados.

En biología, el análisis de conglomerados es utilizado para clasificar animales y plantas. Esto es conocido como taxonomía numérica.

En medicina, el análisis de conglomerados es utilizado para identificar enfermedades y sus diferentes etapas.

En mercadotecnia, el análisis de conglomerados se utiliza para identificar personas con similares hábitos de consumo; examinando estas características se es capaz de desarrollar estrategias de mercados más eficientes.

Como en cualquier otra metodología estadística, una serie de preguntas deben de ser contestadas antes de embarcarse en realizar dicho análisis.

- ¿Qué variables servirán de base para formar los grupos?
- ¿Cómo será medida la distancia entre los casos?
- ¿Qué criterio será utilizado para combinar los casos y formar los clusters?

La selección de las variables para incluir en el análisis siempre es crucial.

Si hay variables importantes que son excluidas, resultados pobres o engañosos pueden ser encontrados.

En el análisis de conglomerados la elección inicial de las variables determina las características que pueden ser utilizadas para identificar los subgrupos.

Los conceptos de distancia y similitud son básicos en muchas técnicas estadísticas. La distancia es la medida de que tan lejos están dos objetos; la similitud es la medida de que tan cerca están estos dos objetos.

Las medidas de distancia son pequeñas y las de similitud son grandes para objetos parecidos.

¹⁸ La técnica es conocida también como *Análisis de conglomerados* o en su denominación en inglés *Cluster Análisis*.

¹⁹ NORUSIS, M.J. (1994) *SPSS Professional Statistics 6.1*. SPSS. Chicago, Ill. USA.

En el análisis de conglomerados estos conceptos son particularmente importantes dado que los casos son agrupados en términos de la "cercanía".

De una manera resumida diríamos que el análisis de conglomerados es el nombre genérico de una amplia variedad de procedimientos relacionados con la identificación de grupos dentro de los datos.

Mediante la organización de datos multivariantes dentro de tales subgrupos, el agrupamiento puede ayudar al investigador a descubrir las características de cualquier estructura o patrón presente.

Los grupos se encuentran basados en alguna medida de proximidad o distancian entre las entidades de agrupamiento e involucran cientos de Algoritmos diferentes.

El análisis de conglomerados tiene numerosas ventajas sobre otros métodos de clasificación.

En primer lugar, está diseñado para clasificar y proporcionar resultados en forma de grupos.

El resultado de un análisis de conglomerados es típicamente un pequeño número de grupos mutuamente excluyentes²⁰ y la información acerca de las entidades de agrupamiento cae dentro de cada grupo.

En segundo lugar, el análisis de conglomerados es capaz de tratar con una amplia variedad de información, incluyendo series de tiempo, análisis cross-sectional y datos longitudinales.

Inclusive datos categóricos pueden ser utilizados junto con datos de intervalo.

Siempre y cuando una matriz de proximidad pueda calcularse cualquier tipo de dato puede ser utilizado.

Finalmente, el análisis de conglomerados no hace muchos supuestos. Mientras que cada método o Algoritmo pueda ser construido bajo ciertas premisas es usualmente posible encontrar un método que es aplicable para el método para el problema.

Sin embargo, el análisis de conglomerados no está libre de los problemas.

La dificultad principal la representa el gran número de formas de desarrollarlo, que deja al investigador con la tarea de seleccionar el método apropiado.

²⁰ Una excepción es el agrupamiento difuso (Fuzzy clustering) que proporciona la probabilidad de que una entidad esté en un grupo en particular.

Esto puede producir un cierto grado de sesgo, dado que el investigador podría tener una idea preconcebida de los resultados y decidir cambiar los métodos hasta que encuentre uno que apoye los resultados preconcebidos.

Un segundo problema resulta del rango de disciplinas académicas que usan el análisis de conglomerados, que van desde la arqueología hasta la zoología.

Esto lleva denominaciones inconsistentes y a resultados contradictorios. Sólo en los últimos años los investigadores de varias disciplinas han celebrado reuniones para discutir el análisis del conglomerados.

Un último problema es el resultado del más rápido desarrollo de nuevas técnicas de conglomerados que de la validación que estas tienen.

Se proponen muchos Algoritmos de conglomerados pero nunca se prueban, y por consiguiente dichas técnicas sólo limitan al uso del investigador que la desarrolla y en su caso la aplica.

El análisis de conglomerados involucra tres pasos:

1. Determinar la medida de proximidad.
2. Seleccionar el Algoritmo de agrupamiento.
3. Evaluar los resultados de los grupos.

4.1 Medidas de proximidad

Existe una amplia variedad de métodos para calcular medidas de proximidad.

Por ejemplo, la correlación de Pearson y la distancia Euclidiana.

En la investigación de la finanzas, la hasta la fecha más popular medida de proximidad es el coeficiente de correlación.

King²¹ (1996), Meyers²² (1973), Farrell²³ (1974), Livingston²⁴ (1977) y Arnott²⁵ (1980) usan el coeficiente producto-momento de correlación de Pearson de los rendimientos las compañías en el tiempo.

²¹ KING, B. (1996). "Market and Industry Factors in Stock Price Behavior". Journal of Business.

²² MEYERS, S. (1973). "A Re-Examination of Market and Industry Factors in Stock Price Behavior". Journal of Finance.

²³ FARREL, J. (1974). "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings". Journal of Business.

²⁴ LIVINGSTON, M. (1977). "Industry Movements of Common Stocks". Journal of Finance.

²⁵ ARNOT, R. (1980). "Cluster Analysis and Stock Price Co movement". Financial Analyst Journal, Nov-Dec.

Este coeficiente es definido como:

$$\rho_{y_1, y_2} = \frac{\text{cov}(y_1, y_2)}{\sqrt{\sigma_{y_1} \sigma_{y_2}}}$$

Ecuación 4.1: correlación producto-momento de Pearson

Se supone que las compañías con un gran coeficiente de la correlación positivo son más similares que las compañías con un coeficiente de correlación pequeño o negativo.

En el contexto de la optimización de un portafolio de Markowitz, este coeficiente de correlación tiene un significado fundamental.

En este marco, los inversionistas usan los promedios, variancias, y covariancias del rendimiento de las acciones para calcular la ponderación óptima en un portafolio de varios activos.

De esta manera lo más importante no es el nivel de rendimientos, sino la correlación de los rendimientos.

Usar la correlación de Pearson como una medida de proximidad debe producir grupos de compañías cuyos rendimientos varían juntos a través del tiempo.

Tal vez la más popular medida de distancia para el análisis de conglomerados es la distancia Euclidiana entre dos puntos x_1 e x_2 :

$$d(x_1, x_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$$

Ecuación 4.2: distancia euclidiana entre dos puntos x_1 y x_3

La distancia Euclidiana cae dentro de la categoría de las medidas de distancia llamadas métricas.

Las métricas son representaciones de similitud o distancia que pueden ser representadas en un espacio de coordenadas que satisfacen lo siguiente:

1. $d(x_1, x_2) \geq 0$; positividad
2. $d(x_1, x_1) = 0$; igualdad de los idénticos
3. $d(x_1, x_2) = d(x_2, x_1)$; simetría
4. $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$; desigualdad del triángulo

La distancia Euclidiana fue usada inicialmente por Elton y Gruber²⁶ (1971) y ha sido usada recientemente por Abraham, Goetzmann, y Wachter²⁷ (1994), Brown y Goetzmann²⁸ (1997) y Brown, Goetzmann, y Grinblatt²⁹ (1997).

Los argumentos contra el uso de cualquier medida de similitud o distancia de que no satisfacen los cuatro criterios un una métrica se presenta en Jardine y Sibson³⁰ (1971) y Clifford y Stephenson³¹ (1975).

Debido a que la distancia Euclidiana es la dominante en la literatura actual y que satisface los requerimientos de una métrica, se usará como la medida de proximidad en este estudio.

4.2 Algoritmos de agrupamiento

El gran número métodos de agrupamiento puede subdividirse en cuatro grupos: los datos pueden ser agrupados ya sea usando Algoritmos heurísticos o Algoritmos de optimización (Kaufmann y Rousseeuw³² (1990)).

Adicionalmente, los métodos pueden describirse como jerárquico o de división (Everitt³³ (1993)).

Los Algoritmos heurísticos son esencialmente descripciones matemáticas del agrupamiento basado en diagramas.

Con mayor frecuencia, los Algoritmos heurísticos también son jerárquico aglomerativos, empezando con cada objeto como su propio grupo y uniendo todos los grupos hasta que sólo quede un grupo final.

El proceso de agrupamiento puede desplegarse en un diagrama llamado dendograma que se parece las raíces de un árbol.

Sin embargo, este proceso de la fusión puede ser problemático, dado que es imposible deshacer los grupos una vez formados.

²⁶ ELTON, E.J., GRUBER, M.J. (1970). "Homogeneous Groups and the Testing of Economic Hypotheses". *Journal of Financial and Quantitative Analysis*.

²⁷ ABRAHAM, J.M., GOETZMANN, W.N., WATCHER, S.M. (1994). "Homogeneous Grouping of Metropolitan Housing Markets". *Journal of Housing Economics*.

²⁸ BBROWN, S.J., GOETZMANN, W.N. (1997). "Mutual Funds Style". *Journal of Financial Economics*, 43.

²⁹ BBROWN, S.J., GOETZMANN, W.N., GRINBLATT, M. (1997) "Positive Portfolio Factors", Working Paper.

³⁰ JARDINE, N., SIBSON, R. (1971). *Mathematical Taxonomy*. John Wiley & Sons. London. UK.

³¹ CLIFFORD, H., STEPHENSON, W. (1975). *An Introduction to Numerical Taxonomy*. Academic Press. New York, USA.

³² KAUFMAN, L., ROUSSEEUW, P.J. (1987). "Clustering by means of medoids in Dodge Y. *Statistical Data Analysis Based on the L1-norm*. Amsterdam.

³³ EVERITT, B. (1993) *Cluster Analysis*. 3rd Edition. Edward Arnold. London.

Los métodos jerárquicos divisorios empiezan con un grupo que es separado en grupos sucesivos.

Al igual que los métodos jerárquicos aglomerativos no se pueden deshacer los grupos una vez formados.

En contraste con los métodos heurísticos, los Algoritmos de optimización no están basados en argumentos gráficos, sino que buscan minimizar algún tipo de función de pérdida.

Quizás la más popular rutina de optimización es el método de Ward, un algoritmo jerárquico aglomerativo que busca minimizar el incremento de la suma de los cuadrados o la variancia dentro de los grupos que se van formando.

La mayoría de los Algoritmos de optimización no son jerárquicos, pero por división: los miembros de un número preestablecido k de grupos son cambiados para poder minimizar alguna función.

Por mucho, el método de división más popular, el método de los k -medias, minimiza las distancias cuadradas a las medias de los grupos preestablecidas que son llamadas semillas.

Al usar el método de las k -medias, estas semillas pueden escogerse al azar o pueden encontrarse usando otro Algoritmo de agrupamiento.

Existen criterios matemáticos que pueden potencialmente ayudar en la selección del Algoritmo de agrupamiento.

Jardine y Sibson³⁴ (1971) muestran que el Algoritmo de eslabonamiento simple es el único método que satisface un conjunto de condiciones matemáticas. Sin embargo, Kaufman y Rousseeuw³⁵ (1990) argumenta que "la respetabilidad matemática de un método depende de ha establecido el criterio".

Tomando de base a Rousseeuw³⁶ (1985), ilustran que un diferente conjunto de condiciones conduce a que el método de eslabonamiento promedio es el único Algoritmo admisible sin que estos autores deseen establecer que este es "el único método aglomerativo razonable".

En general, los métodos de agrupamiento caen en uno de tres grupos: métodos de eslabonamiento, métodos de sumas de cuadrados o variancia, y métodos de centroide.

³⁴ Ibid 30

³⁵ Ibid 30

³⁶ ROUSSEEUW, P.J., (1985). "Some Thoughts on Agglomerative Cluster Analysis" CM ISI, Amsterdam.

Todos están basados ya sea en una matriz de distancias o de similitud entre las parejas de casos.

Los métodos difieren en cómo estiman las distancias entre los grupos en las sucesivas etapas del método.

Dado que la fusión de grupos a cada paso depende de las medidas de distancia, diferentes medidas de distancia pueden producir diferentes grupos de soluciones para el mismo Algoritmo de agrupamiento.

Hay una gran variedad de Algoritmos de agrupamiento, como ya hemos dicho hay casi tantos como autores sobre el tema haya; sin embargo haciendo una lista muy sencilla y comúnmente aceptada, esta sería la lista de Algoritmos de agrupamiento que de manera general es son los más utilizados:

- **Eslabonamiento simple** (el vecino más cercano): las combinaciones se hacen agrupando los elementos con la menor distancia al nuevo grupo en formación.
- **Eslabonamiento completo** (el vecino más lejano): las combinaciones se hacen agrupando los elementos con la mayor distancia al nuevo grupo en formación.
- **Eslabonamiento promedio** (UPGMA³⁷): se toma el promedio del nuevo elemento al nuevo grupo en formación.
- **Eslabonamiento promedio dentro de los grupos**: es un UPGMA que no sólo agrupa por los más parecidos sino que separa por los menos parecidos.
- **Método de Ward**: busca hacer grupos de igual magnitud y ponderación entre toda la población de inicio.
- **Método del centroide**: calcula la distancia entre dos grupos como la distancia a la media de todos los elementos de los grupos.

Para este trabajo hemos seleccionada el método del eslabonamiento promedio dentro de los grupos, pues toma semejanzas y diferencias que en este trabajo son tan importantes las unas como las otras.

4.3 Evaluación de los grupos

Everitt³⁸ (1993) describe una práctica recurrente de evaluar los agrupamientos de la manera siguiente:

³⁷ unweighted pair-group method using arithmetic averages

³⁸ Ibid 31

"La interpretación de resultados de un Algoritmo de agrupamiento es a menudo determinado por la intuición y la visión personal. Si al investigador le hacen sentido el número de grupos producidos, entonces el análisis de agrupamientos frecuentemente se juzga como un éxito".

No es necesario decir que este método puede producir decisiones arbitrarias y hacer difícil de comparar los resultados de los grupos.

Afortunadamente existen métodos objetivos para examinar los resultados del agrupamiento.

La validación de los grupos se usa para evaluar la calidad de los grupos y las reglas de alto se usan para determinar el número de grupos óptimos. Muchos de estos métodos recaen en el criterio conocido como "criterio de la variancia" que puede expresarse en términos de la suma de los cuadrados dentro del grupo SSW y la suma de los cuadrados entre grupos SSB , o la proporción de la suma de los cuadrados entre el grupos y dentro del grupo SSB/SSW .

Capítulo 5

ALGORITMOS GENÉTICOS

Los Algoritmos genéticos (GA´s) son métodos de adaptación que se pueden utilizar para solucionar problemas de búsqueda y optimización.

Se basan en los procesos genéticos de los organismos biológicos.

Por generaciones las poblaciones en la naturaleza se han desarrollado según los principios de la selección natural y de la supervivencia del más apto.

Imitando este proceso, los Algoritmos genéticos pueden evolucionar las soluciones a los problemas del mundo real, siempre y cuando estos problemas hayan sido codificados convenientemente.

Los principios básicos de los GA´s fueron desarrollados por Holland³⁹ en 1975.

Los GA´s trabajan con una población de "individuos" cada uno de los cuales representa una solución posible a un problema dado.

A cada individuo le es asignada una "medida de aptitud" de acuerdo a que tan buena solución es para el problema que se este resolviendo.

Los individuos con una alta aptitud tienen la oportunidad de reproducirse mediante una "cruza" con otros individuos de la población.

Esto produce nuevos individuos o "crías" que comparten algunas características tomadas de los "padres".

Los individuos menos aptos de la población tienen menos probabilidades de ser seleccionados para la reproducción y por lo tanto se extinguen.

Una población completamente nueva de posibles soluciones es producida por la selección de los mejores individuos de la actual generación y que son apareados para producir una nueva generación.

Esta nueva generación contiene una alta proporción de las características que poseían los buenos miembros de la generación anterior.

De esta manera, a través de las generaciones, las buenas características son propagadas a través de la población.

Favoreciendo el apareo de los individuos más aptos, las áreas más promisorias del espacio de búsqueda son exploradas.

³⁹ HOLLAND, John H. (1992). *Adaptation in Natural and Artificial Systems. An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press. Boston. MA.

Si el GA ha sido bien diseñado, la población converge a una solución óptima del problema.

5.1 Un vistazo histórico a las teorías de la evolución de las especies y de computación evolutiva⁴⁰

El origen de las ideas desarrolladas por la computación evolutiva se encuentra en las teorías biológicas de la evolución de las especies, donde la premisa fundamental es que los más fuertes sobrevivirán a las amenazas del medio ambiente y que dada esta supervivencia se generarán nuevas y mejores especies, con mayor resistencia y capacidad de adaptación.

Contrario a lo que muchos creen, las ideas que popularizó Charles Darwin en 1858 no se originaron con él, sino que estuvieron presentes en las mentes de una serie de científicos y pensadores en general que no se sentían satisfechos con la idea de que había un Dios creador de todas las especies del planeta y de que las especies estaban jerarquizadas por Dios de tal manera que el hombre ocupaba un rango superior al lado del creador.

5.1.1 El origen de las ideas

Georges Louis Leclerc (Conde de Buffon) fue tal vez el primero en especular, unos 100 años antes que Darwin, en su *Historie Naturelle* (una impresionante enciclopedia de entre 36 y 44 tomos, que describía todo lo que se sabía en aquel entonces sobre la naturaleza), que las especies se originaron entre sí.

Leclerc no sólo notó las similitudes entre el hombre y los simios, sino que incluso habla sobre un posible ancestro común entre estas dos especies.

Leclerc creía en los cambios orgánicos, pero no describió un mecanismo coherente que fuera responsable de efectuarlo, sino que especuló que era el ambiente el que influía directamente sobre los organismos.

5.1.2 Lamarckismo

Propuesta por el zoólogo francés Jean Baptiste Pierre Antoine de Monet (Caballero de Lamarck) a partir de 1801, enfatizaba la importancia de la naturaleza en el cambio de las especies, ya que afirmaba que los organismos no eran alterados en forma pasiva por su ambiente, sino que más bien un cambio en su ambiente produce cambios en las necesidades de los organismos, lo que hace que en consecuencia, estos cambien su comportamiento.

⁴⁰ Fragmentos tomados de COELLO, Carlos A. (2001). *Introducción a la computación evolutiva*, CINVESTAV-IPN, cap. 2; pp. 39-64. Usados con autorización del autor de acuerdo al e-mail del 27 de mayo de 2004.

Estos cambios en el comportamiento producen mutaciones como crecimiento de órganos o pérdida de miembros y lo ejemplificaba con las jirafas y el crecimiento de su cuello.

Aunque su teoría fue más bien tomada a burla y hoy en día el concepto de Lamarckismo es usado de manera peyorativa, sus ideas conducen a los mismos resultados de Darwin, es decir, que las especies sufren cambios adaptativos debido a la influencia del ambiente.

5.1.3 La teoría del germoplasma

August Weismann formuló en el siglo XIX una teoría denominada del *germoplasma*, según la cual el cuerpo se divide en células germinales (o germoplasma) que puede transmitir información hereditaria y en células somáticas (o somatoplasma) que no pueden hacerlo⁴¹.

Sus ideas se oponían al Lamarckismo y decidió probarlo cortando las cola a una serie de ratas y mostrando que las generaciones posteriores de dichos animales no eran afectados en el tamaño promedio de su cola por este hecho.

De alguna forma condujo al re-descubrimiento del trabajo de Mendel sobre las leyes de la herencia.

Hoy en día el germoplasma ha evolucionado al genotipo⁴² y el somatoplasma al fenotipo⁴³.

5.1.4 Russell y Darwin

Alfred Russell Wallace se interesó en el origen de las especies a mediados del siglo XIX e intuyó de manera autónoma la teoría de la selección natural sin saber que Darwin y se le había adelantado e irónicamente le escribió a este para pedirle que le ayudara a publicar sus ideas.

El resultado fue la presentación de un trabajo conjunto. Curiosamente las más grandes influencias de Russell y Darwin estuvieron fuertemente influidos por los trabajos del economista Thomas Robert Malthus y del geólogo Charles Lyell.

Malthus afirmaba que mientras las poblaciones se reproducen de manera geométrica, los recursos naturales sólo se reproducen de manera aritmética por lo que de cuando en cuando ocurren crisis para estabilizar las poblaciones.

⁴¹ WEISMANN, August, (1893). *The Germ Plasm: A Theory of Heredity*. Scott, London, UK. Citado por Coello.

⁴² El genotipo se usa para denotar la composición genética de un individuo.

⁴³ Fenotipo son los rasgos específicos de un individuo.

Lyell por su parte afirmaba que la forma de las montañas se debía a causas naturales y no divinas.

La influencia de la "NO DIVINIDAD" era un aspecto muy importante que debía enfrentar Darwin y de algún modo el hecho de que otro hiciera mención a esto lo llevo a convencerse de publicar *El origen de las especies* cuyo tiraje de 1250 ejemplares se agotó en un solo día.

Darwin se percató de que una especie que no sufriera cambios se volvería incompatible con su ambiente y que las características son heredadas de padres a hijos y que de generación ocurrían cambios cuya principal motivación era hacer a los nuevos individuos más aptos para sobrevivir.

5.1.5 La teoría de la combinación

Establece que las características hereditarias de los padres se mezclan o combinan de alguna forma en sus hijos.

La mayor debilidad de esta teoría es que no explica la ocurrencia de cambios repentinos.

5.1.6 Las leyes de la herencia de Mendel

El monje austriaco Johann Gregor Mendel sentía que su vida monástica era muy aburrida⁴⁴ y por eso realizó una serie de experimentos con chícharos durante una gran parte de su vida, estudiando las características básicas de esta planta.

Mediante un cuidadoso análisis de las diferentes características manifestadas por las plantas de los chícharos, Mendel descubrió tres leyes básicas que gobernaban el paso de una característica de un miembro de una especie a otro, dichas leyes las nombro como Ley de *segregación*, de *recombinación* y de *sorteo independiente*.

Aunque Mendel descubrió las leyes de la herencia, realmente no llegó a entender el mecanismo detrás de ella y los genes así como el mecanismo de transmisión de los mismos de generación en generación no fue descubierto sino hasta varios años después. Profundizaremos en este tema más adelante.

5.1.7 La teoría de la pangénesis

Teoría esbozada por Darwin que esgrime que cada órgano del cuerpo produce pequeñas partículas hereditarias llamadas *gémulas* o *pangenes* que son transportadas a través de la sangre y se recolectaban en los

⁴⁴ No existe evidencia de que esto en realidad ocurriera, es sólo una interpretación de quien esto escribo sobre la vida monstica.

gametos durante su formación con lo que las características de los padres se heredan a los hijos a través de la sangre.

5.1.8 La teoría de la mutación

Hugo DeVries creyó haber descubierto una nueva especie de planta al encontrar una flor roja entre una gran cantidad de flores amarillas.

Según DeVries esto se debía a una mutación abrupta de flores amarillas.

De Vries decía que los cambios eran abruptos y no paulatinos.

5.1.9 La teoría cromosómica de la herencia

En 1903 Walter Sutton leyó los trabajos de Mendel y DeVries y determino correctamente sin necesidad de experimentos que los cromosomas en el núcleo de las células es el lugar donde se almacenaban las características hereditarias.

Thomas Hunt Morgan confirmaría experimentalmente la hipótesis de Sutton.

5.1.10 Neo-darwinismo

Establece que la historia de la vasta mayoría de la vida en nuestro planeta puede ser explicada a través de un puñado de procesos estadísticos que actúan sobre y dentro de las poblaciones y especies: la reproducción, la mutación, la competencia y la selección.

La reproducción es la propiedad obvia de las formas de vida en el planeta ya que sin ella no sería posible la reproducción.

En cualquier sistema que se reproduce a sí mismo continuamente y que está en constante equilibrio, la mutación está garantizada.

5.1.11 Inspiración biológica

La evolución natural fue vista como un proceso de aprendizaje desde los años 30. W.D.

Cannon plantea que el proceso evolutivo es algo similar al aprendizaje por ensayo y error que suele manifestarse en los humanos.

Por otro lado el celebre matemático inglés Alan Mathison Turing reconoció también una conexión "obvia" entre la evolución y el aprendizaje de máquina en su artículo titulado *Computing Machinery and Intelligence* y que se considera el clásico de la inteligencia artificial.

5.1.12 Alexander Fraser

A fines de los años 50 y principios de los 60s el biólogo Alexander S. Fraser publicó una serie de trabajos sobre la evolución de los sistemas biológicos en una computadora digital, siendo la inspiración para lo que se convertiría más tarde en Algoritmos genéticos.

El trabajo de Fraser incluye entre otras cosas el uso de una reproducción binaria, de un operador de cruce probabilística de una población de padres que generaban una nueva población de hijos tras recombinarse y el empleo de un mecanismo de selección.

Su trabajo de más de 10 años en este tema se resumen en un libro titulado *Computer Models in Genetics*.

5.1.13 EVOP

George E.P. Box, un experto en estadística inglés, propuso un enfoque evolutivo para la optimización de la producción industrial.

Su técnica denominada EVOP (*Evolutionary Operation*) consistía en efectuar pequeños cambios a un conjunto de parámetros de producción, monitoreando ciertos datos estadísticos de los procesos para guiar la búsqueda.

Llegó a establecer claramente la analogía entre los cambios y las mutaciones que ocurren en la naturaleza, e hizo ver que el proceso de ajuste de parámetros que efectuaba con técnicas estadísticas era similar al proceso de selección natural.

5.1.14 La evolución de programas de Friedberg

R.M. Friedberg es considerado como uno de los primeros investigadores en intentar evolucionar programas de computadora.

Aunque Friedberg no usa la palabra evolución, resulta claro que ese es el enfoque que adoptó en su artículo original y en los posteriores, de hecho en un artículo posterior sus co-autores modelan la evolución como un proceso de optimización⁴⁵.

5.1.15 Friedman y la robótica evolutiva

George J. Friedman fue tal vez el primero en proponer una aplicación de técnicas evolutivas a la robótica; propuso evolucionar una serie de circuitos de control similares a los que hoy conocemos como redes

⁴⁵ DUNHAM, B., FRIDSHAL, D., NORTH, J.H. (1963). *Design by natural selection*. *Synthese*, 15:254-259, Citado por Coello.

neuronales, usando lo que el denominaba “retroalimentación selectiva”, en un proceso análogo a la selección natural.

También especuló que la simulación del proceso de reproducción sexual (cruza) y el de mutación nos conducirían al diseño de “máquinas pensantes”.

5.1.16 Vida artificial

Nils Aall Barricelli desarrolló simuladores de un sistema evolutivo en una computadora.

Utilizó la computadora IAS⁴⁶ y sus experimentos siguieron los lineamientos de la disciplina que después fue conocida en los 80s como “Vida artificial”. Sus conclusiones fueron que los genes deben satisfacer:

- Una cierta capacidad de reproducción.
- Una cierta capacidad para cambiar a formas alternas.
- Una necesidad de simbiosis con otros genes u organismos.

5.1.17 La optimización evolutiva de Bremerman

Hans Joachim Bremermann vio la evolución como un proceso de optimización y utilizó una técnica evolutiva para problemas de optimización con restricciones lineales.

Fue de los primeros en usar el concepto de “población” en la simulación de procesos evolutivos. Advirtió la importancia de la coevolución.

5.1.18 La programación evolutiva

Fogel y otros concibieron el uso de la evolución simulada en la solución de problemas.

Su técnica denominada “programación evolutiva” consistía en hacer evolucionar autómatas de estados finitos los cuales eran expuestos a una serie de símbolos de entrada y se esperaba que eventualmente fueran capaces de predecir las secuencias futuras de símbolos que recibirán.

La programación evolutiva se aplicó originalmente a problemas de predicción, control automático, identificación de sistemas y teoría de juegos.

⁴⁶ La computadora se desarrolló en el *Institute for Advanced Studies* por el legendario matemático John Von Neumann.

5.1.19 Las estrategias evolutivas

Peter Bienert, Ingo Rechenberg y Hans Paul Schwefel desarrollaron un método de ajustes discretos aleatorios inspirados en el mecanismo de mutación que ocurre en la naturaleza.

A esta técnica la denominaron "estrategia evolutiva". Bienert construyó un robot que podía efectuar de manera automática el proceso de optimización usando este método.

Al mismo tiempo Schwefel se dio a la tarea de implementar esta técnica en una computadora.

5.1.20 Los Algoritmos genéticos

En los años 60 John H. Holland se interesó en estudiar los procesos lógicos involucrados en la adaptación.

Inspirado por los estudios realizados en aquella época con autónomas celulares y redes neuronales.

Holland se percató de que el uso de reglas simples podría generar comportamientos flexibles, y visualizó la posibilidad de estudiar la evolución de comportamientos en un sistema complejo.

Aunque concebido originalmente en el contexto del aprendizaje de máquina, el Algoritmo genético se ha utilizado mucho en optimización, siendo una técnica sumamente popular en la actualidad. Este tema es profundizado más adelante.

5.1.21 Ecosistemas artificiales

Michael Conrad y Howard H. Pattee se cuentan entre los primeros en simular un ecosistema artificial jerárquico en el que un conjunto de organismos unicelulares estaban sujetos a una estricta ley de conservación de la materia que les inducía a competir para sobrevivir.

Los organismos simulados fueron capaces de simular cooperación mutua y de llevar a cabo estrategias biológicas tales como la recombinación genética y la modificación de la expresión de su genoma.

En esta implementación los organismos realmente consistían de subrutinas genéticas, por lo que el fenotipo de cada individuo se determinaba mediante la forma en que estas rutinas eran usadas por los organismos.

5.1.22 Programación genética

Hicklin y Fujiki usaron expresiones S en LISP para representar programas cuyo objetivo era resolver problemas de teoría de juegos.

Hicklin discutió la combinación de segmentos de programas mediante el copiado de subárboles de un individuo a otro.

Cramer y Koza propusieron de manera independiente el uso de una representación de árbol en la que se implementó un operador de cruce para intercambiar subárboles entre los diferentes programas de una población generada al azar.

Koza logro automatizar la función de aptitud cosa que Cramer no pudo y tuvo que hacerlo de manera interactiva.

5.1.23 Dinámica evolutiva

Thomas S. Ray desarrolló a principios de los 90 un simulador muy original en el que se evolucionaban programas en lenguaje ensamblador, los cuales competían por ciclos en el CPU de una computadora, a la vez que intentaban reproducirse (es decir copiarse a sí mismos) en la memoria de dicha computadora.

En este simulador denominado *Tierra*, se partía de un programa único con la capacidad de auto-replicarse al que se denominada "ancestro".

En base a este programa se generaban "criaturas" nuevas, las cuales a su vez se podían dividir para dar nacimiento a más criaturas.

5.2 Conceptos básicos de los Algoritmos genéticos

Los Algoritmos genéticos son Algoritmos de búsqueda basados en el mecanismo de la selección natural y la genética.

Combinan la supervivencia del más apto entre un grupo de strings binarios (es decir, listas de 1 y 0) con un intercambio estructurado de información aleatorio para formar un Algoritmo de búsqueda que contiene algo del instinto innovador que los seres humanos tenemos al investigar.

En cada generación un nuevo conjunto de criaturas artificiales (strings binarios) es creado usando bits y partes de aptitud de las criaturas anteriores; una parte de los nuevos strings es probada para el mejoramiento de la medición.

Aunque son aleatorios, los Algoritmos genéticos no son simples *random walks*, sino que utilizan de manera eficiente la información histórica para especular con

nuevos puntos de búsqueda con una expectativa de mejorar el desempeño. El Algoritmo genético estándar se puede representar de esta manera⁴⁷:

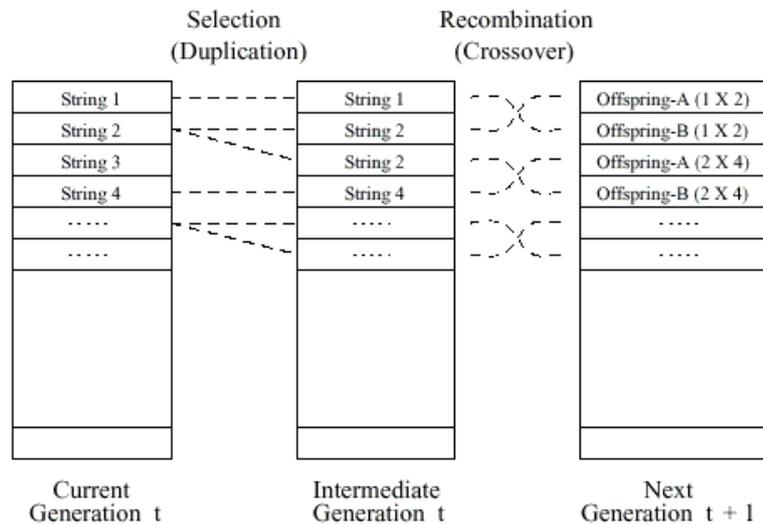


Figura 5.1: representación del algoritmo genético estándar

5.2.1 El modelo biológico de los Algoritmos genéticos

El ADN está conformado por cuatro bases nitrogenadas, las cuáles son Adenina, Guanina, Citosina y Timina. La Adenina y Timina se encuentran en una proporción de 28% y la Guanina y Citosina en una de 19%, permitiéndonos conocer que la Adenina hace pareja con la Timina y la Citosina con la Guanina.

El ADN en sí es una cadena de bases nitrogenadas unidas a otra cadena de bases nitrogenadas formando una doble hélice, pero lo que a nosotros nos interesa saber es que una porción de la molécula del DNA es igual a un gen, y el código genético está conformado por tripletes.

Cada miembro de un par de genes es llamado alelo y veintitrés de estos pares (llamados cromosomas) son los contenedores de toda la información genética del individuo.

Los miembros de cada par de cromosomas son idénticos en tamaño y forma llamándose cromosomas homólogos, todos menos el último par, el par 23 que determina el sexo del individuo.

⁴⁷ Tomado de Buseti, Franco. *Genetic Algorithms: an overview*. <http://www.geocities.com/francorbusetti/gaweb.pdf>

Este par es de cromosomas heterólogos debido a su naturaleza, y estas células sexuales son conocidas como "gametos". XY es para los machos y XX es para las hembras.

Las características físicas que se dan por el paquete genético de instrucciones se llaman fenotipo mientras que el paquete de instrucciones en sí se llama genotipo.

Hay genes que son "dominantes" mientras que otros son "recesivos", y esto se ha visto desde que Mendel instituyó sus leyes.

Según las mismas, la reproducción de un gen dominante contra uno recesivo se puede ilustrar de la siguiente manera:

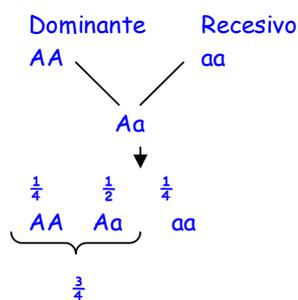


Figura 5.2: reproducción de un gen dominante contra un recesivo

Algunos ejemplos clásicos de esto son el albinismo (donde el gen recesivo es la pigmentación blanca de la piel), el Síndrome de Down, el Síndrome de Turner y el daltonismo.

Podemos concluir que el gen dominante es el "normal" mientras que el recesivo es el "defectuoso".

Como ya habíamos mencionado Mendel describió las siguientes leyes:

Ley de Segregación

- Cada individuo inicia la vida como un óvulo fecundado que contiene dos representantes de cada tipo de gen; Un miembro de cada par se deriva de un padre y el otro del otro padre⁴⁸.
- Un adulto transmite sólo un miembro de cada par de genes en cada gameto que él o ella produce.

⁴⁸ En el vocabulario utilizado en la computación evolutiva algunos autores hablan de padre y madre, y otros hacen referencia a dos padres. En este trabajo usaremos de manera indistinta ambos conceptos.

Ley de Recombinación

Cuando un organismo madura y produce gametos, los genes pueden entrar en los gametos en diferentes asociaciones sin importar el padre del que provengan.

Así cuando los genes derivados de un padre no necesariamente van unidos sino pueden agregarse y formar cualquier combinación posible.

Ley de Sorteo Independiente

Cuando dos o más características se heredaban independientemente una de la otra, la generación producida era el resultado de una distribución al azar de los factores hereditarios en los gametos y de las recombinaciones subsecuentes de éstos durante la fecundación.

Sin embargo se descubrió que a veces hay características que están ligadas entre sí.

En general se hace alusión a unas flores con cierta forma y ciertos filamentos los cuáles estaban relacionados entre sí, y para estos casos la ley de Sorteo Independiente de Mendel no es aplicable. Esto se representaría de esta manera:

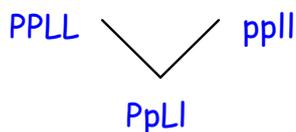


Figura 5.3: representación de la no aplicación de la ley del sorteo independiente

A veces (una probabilidad baja) este ligamento se rompe (el que une a ambas características mencionadas) y por lo tanto esto permite que haya un cruzamiento independiente, esto es conocido como "Entrecruzamiento".

El entrecruzamiento es de una gran ventaja biológica ya que sobre esta materia prima actúa la selección natural para moldear el cambio evolutivo.

Además existen otros fenómenos incontrolables aleatorios conocidos como "mutaciones" que causan cambios en un organismo, que pueden ser relativamente menores o ser graves.

En general las mutaciones tienden a ser maléficas más que benéficas, es como tratar de cambiar una palabra al azar en este texto por otra seleccionada de la misma forma y que el párrafo mejore o al menos

mantenga su coherencia original, obviamente sabemos que la probabilidad de que esto ocurra es demasiado baja.

Las mutaciones pueden ocurrir al azar o cuando se expone a una fuente de "radiación"⁴⁹ dependiente de su intensidad.

Un individuo está formado por su material genético heredado de sus padres así como del fenotipo que este paquete genera, y las circunstancias que le rodean.

Es decir, la actividad que desempeña, la intensidad con que la realiza, la tasa de producción, si está activo sexualmente, si está vivo, si está en celo (En caso de las hembras), y la edad en la que se encuentra.

El fenotipo de un individuo está conformado por el sexo que tiene, si es o no hermafrodita, su edad de inicio de reproducción, su edad de fin de reproducción, su edad límite de vida, su color, atractivo, carisma, inteligencia, resistencia y su tasa de obtención y uso de energía.

En la vida de los individuos pueden acontecer muchas cosas fuera de nuestro control que afecten su fenotipo y su status, lo cuál cambia al individuo, estas se toman en cuenta como "Sucesos".

En resumen podemos decir que cada individuo que habita el planeta posee ciertas características que lo identifican, estas características externas constituyen lo que se denomina el fenotipo del individuo y son dadas por herencia; así pues el fenotipo es el resultado de la interacción del medio ambiente en que se desarrolla un individuo y la herencia que este recibe de sus ancestros.

El fenotipo de cada individuo está determinado por las proteínas que produce y esto a su vez está definido en la información genética de cada una de las células.

La información está contenida en los cromosomas del individuo. Un cromosoma es una larga cadena de ADN, que como ya habíamos mencionado, está formada por cuatro distintos compuestos más simples llamados bases o nucleótidos: adenina (A), guanina (G), citosina (C) y timina (T). A las subcadenas de tres nucleótidos se les conoce como *codones* y al conjunto de nucleótidos que codifican una proteína completa se les llama *genes*. El valor que posee un gen determinado se denomina *alelo*.

Las células que poseen dos juegos de cromosomas se denominan *diploides*; los genes deben estar en pares homólogos.

⁴⁹ Se asigna dicho nombre porque se supone que en las mutaciones de los seres humanos el factor que mayor influencia tiene es la radiación solar.

Si ambos tienen el mismo alelo se dice que son *homocigos*, si no, son *heterocigos*, y en este último caso sólo uno de los alelos se manifestará en el fenotipo (éste se denomina *dominante*, su homólogo que no se manifiesta, se llama *recesivo*).

Toda la información genética de un individuo se llama *genoma* y el conjunto de genes contenidos en el genoma *genotipo*.

Hay unas células especiales llamadas *gametos* que intervienen en la reproducción. Los gametos no se reproducen por *mitosis* como el resto de las células, el proceso de división se llama en este caso *meiosis*.

En la mitosis las células producidas son *diploides*, mientras que en la meiosis el resultado son *haploides*, sólo tienen un juego de cromosomas.

Partiendo de una sola célula diploide el proceso meiótico es como sigue:

1. Se duplica el número de cromosomas en la célula.
2. Se cruza un juego de cromosomas del padre con uno de la madre, formándose dos juegos de cromosomas híbridos.
3. Se divide la célula dos veces y al final del proceso quedan cuatro células haploides.

En el paso dos del proceso de meiosis se mezclan las características del padre y la madre.

Para el cruzamiento de dos cromosomas se forman entre ellos puntos de ruptura y unión de las cadenas de ADN.

Estos puntos llamados *quiasmas* cortan el cromosoma en segmentos llamados *cromátidas* complementarias de dos distintos cromosomas.

Al final cada cromosoma que participó en la cruce queda constituido por segmentos que ya poseía y por otros que eran de su análogo.

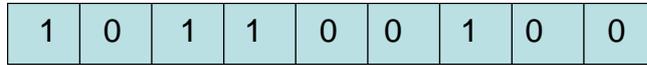
5.2.2 Nomenclatura de la biología usada por los Algoritmos genéticos

En general todos los autores se refieren a los conceptos de los Algoritmos genéticos en términos de sus correspondientes de los usados en la biología y ciencias naturales, algunos autores maneja una distinción entre los términos que llaman naturales y artificiales⁵⁰; refiriéndose por naturales precisamente a los de la biología y por artificiales a los de la computación.

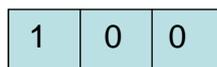
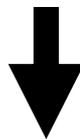
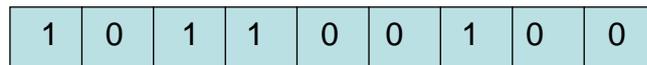
⁵⁰ Por ejemplo el clásico de Goldberg hace esta distinción: GOLDBERG, David, E. (1989). *Genetic Algorithms in search, optimization and machine learning*. Addison-Wesley-Longman. Boston, MA.

Estas son las definiciones de los conceptos que se usaran en los Algoritmos Genéticos aquí presentados y las equivalencias usadas por Goldberg entre paréntesis en *itálicas*:

- **Cromosoma** (*string*): estructura de datos que contiene una cadena de parámetros de diseño o genes.



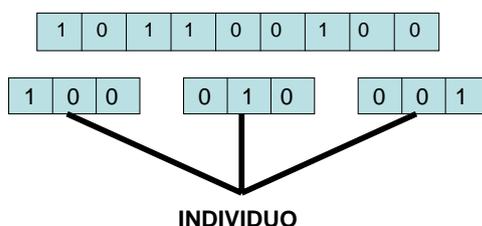
- **Gen o Gene** (*feature, character, detector*): subsección de un cromosoma que usualmente codifica el valor de un solo parámetro.



- **Genotipo** (*structure*): es la codificación de los parámetros que representa una solución del problema a resolverse.
- **Fenotipo** (*parameter set, alternative solution, decoded structure*): es la decodificación del cromosoma, es decir, es el valor obtenido al pasar de la representación binaria a la función objetivo.



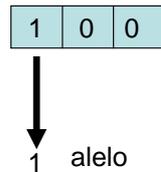
- **Individuo**: es un solo miembro de la población de soluciones potenciales a un problema. Cada individuo contiene un cromosoma (o de manera más general un genoma) que representa una solución posible al problema a resolverse.



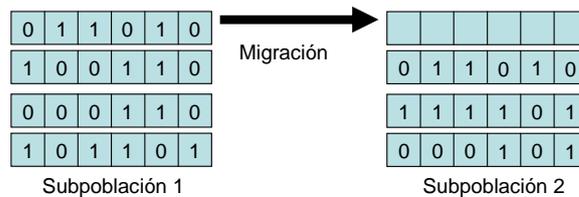
- **Aptitud**: es el valor que se asigna a cada individuo y que indica que tan bueno es éste con respecto a los demás para la solución de un problema. Es decir, es una función que asigna un

resultado a cada individuo que pase por el proceso definido en dicha función.

- **Paisaje de aptitud** (*fitness landscape*): es la hipersuperficie que se obtiene al aplicar la función de aptitud a cada punto del espacio de búsqueda.
- **Alelo** (*feature value*): es cada valor posible que puede adquirir una cierta posición genética.

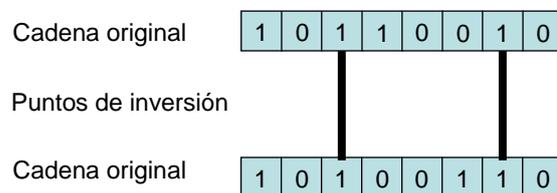


- **Generación:** es una iteración de la medida de aptitud y la creación de una nueva población por medio de operadores de reproducción.
- **Subpoblación:** es grupo que surge de la división de la población. En Genaro, sólo pueden cruzarse entre sí los individuos que provienen de la misma población.
- **Migración:** es la transferencia de los genes de un individuo de una subpoblación a otra. Si la cruce sólo se permite entre individuos de la misma subpoblación decimos que tenemos una especiación.



- **Población panmítica:** es aquella en la que cualquier individuo puede reproducirse con otro con una probabilidad que depende sólo de su aptitud.
- **Epístasis** (*nonlinearity*): es la interacción entre los diferentes genes de un cromosoma. Es la medida en que la contribución de aptitud de uno de los genes depende de los valores de otros genes.
- **Bloque constructor:** es un grupo pequeño y compacto de genes que han coevolucionado de tal forma que su introducción en cualquier cromosoma tiene una alta probabilidad de incrementar la aptitud de dicho cromosoma.

- **Decepción:** es la condición en que la combinación de buenos bloques constructores llevan a una reducción de la aptitud, en vez de un incremento.
- **Operador de reproducción:** es todo aquel mecanismo que influencia la forma en que se pasa la información genética de padres a hijos. Los operadores de reproducción caen en tres amplias categorías a saber: cruza, mutación y reordenamiento.
- **Cruza:** forma un nuevo cromosoma combinando partes de cada uno de sus cromosomas padre.
- **Mutación:** es un operador que forma un nuevo cromosoma a través de alteraciones (usualmente pequeñas) de los valores de los genes de un solo cromosoma padre.
- **Reordenamiento:** es aquel que cambia el orden de los genes de un cromosoma, con la esperanza de juntar los genes que se encuentren relacionados, facilitando así la producción de bloques constructores.
- **Inversión:** tipo de operador de reordenamiento en el que se invierte el orden de todos los genes comprendidos entre dos puntos seleccionados al azar.



- **Elitismo:** es el mecanismo utilizado en algunos Algoritmos evolutivos para asegurar que los cromosomas de los miembros más aptos de una población se pasen a la siguiente generación sin ser alterados por ningún operador genético. El elitismo asegura que la aptitud máxima nunca se perderá de una generación a la siguiente.

5.2.3 Pasos del proceso que realizan los Algoritmos genéticos

Dependiendo de las preferencias de los autores los términos de algunos de los pasos en el desarrollo de un Algoritmo genético suele tener varios nombres, sin embargo el significado esencial es el mismo; así mientras Kuri⁵¹ y Gen⁵² hablan de *codificación*; ese mismo concepto es llamado

⁵¹ KURI, Angel. (1999). *A Comprehensive Approach to Genetic Algorithms in Optimization and Learning. Theory and Applications. Volume 1. Foundations.* Instituto Politécnico Nacional. México.

⁵² GEN, Mitsuo, CHENG, Runwei. (2000) *Genetic Algorithms and Engineering Optimization.* Wiley-Interscience. New York.

por Michalewicz⁵³ y Martí⁵⁴ como *representación*, y, por otro lado el paso de evaluación es mencionado como *evaluación* por Kuri⁵⁵, por *medida de evaluación* por Martí⁵⁶ y como *función de evaluación* por Michalewicz⁵⁷, Gen⁵⁸, Holland⁵⁹ y Goldberg⁶⁰.

Así pues nosotros hablaremos que para poder representar un problema como Algoritmo genético es necesario realizar los siguientes pasos:

- Representación en forma de cromosoma (codificación).
- Población inicial.
- Selección.
- Evaluación.
- Cruza.
- Mutación.

Representación en forma de cromosoma (codificación): la solución hipotética al problema que se esté encarando se representa en forma de una cadena de caracteres binarios, la siguiente es una representación típica de la codificación⁶¹:

Cromosoma 1	1101100100110110
Cromosoma 2	1101111000011110

Población inicial: en general suele ser generada aleatoriamente, sobre todo en los casos en los que desconocemos todo acerca de las características de la solución de nuestro problema, en esos casos dejamos que la computadora seleccione N candidatos de solución.

Recientemente se están utilizando métodos heurísticos para generar soluciones de buena calidad, sin embargo hay que tener cuidado con estas pues se podría caer a convergencias prematuras. Las siguientes representaciones son todas de Kuri⁶²:

⁵³ MICHALEWICZ, Zbigniew. (1992). *Genetic Algorithms + Data Structures = Evolution Programs*. Springer. Berlin.

⁵⁴ <http://www.uv.es/~rmarti/genet.html>

⁵⁵ *Ibíd.*

⁵⁶ *Ibíd.*

⁵⁷ *Ibíd.*

⁵⁸ *Ibíd.*

⁵⁹ *Ibid* 36

⁶⁰ *Ibid* 56

⁶¹ Tomada de <http://cs.felk.cvut.cz/~xobitko/ga/operators.html>

⁶² *Ibíd.* 61

Número	Individuo
1	0111
2	0100
3	0101
4	0001
5	0000
6	1100
7	1000
8	1001
9	1000
10	1010

Selección: el proceso de selección de un Algoritmo genético es fundamental en la exactitud de los resultados que presentará dicho AG una vez procesado. Dado que el proceso de selección es probabilístico, aun los individuos menos aptos tienen cierta oportunidad de sobrevivir. Existen tres grandes grupos para clasificar las técnicas de selección de un AG, a saber:

- **Selección proporcional:** conjunto de esquemas de selección en los cuales se eligen individuos de acuerdo a su contribución de aptitud con respecto al total de la población. Las técnicas de selección son:
 - La ruleta.
 - Sobrante estocástica.
 - Universal estocástica.
 - Muestreo determinístico.
 - Escalamiento sigma.
 - Jerarquías.
 - Selección de Boltzmann.

En particular para este estudio seleccionaremos las técnicas de la ruleta y la selección de Boltzmann que serán discutidas con más detalle en el capítulo 5.

- **Selección mediante torneo:** métodos que requieren dos pasos a través de toda la población en cada generación: 1) calcular la aptitud media y, 2) Calcular el valor esperado de cada individuo. La idea básica del método es seleccionar con base en comparaciones directas de los individuos. Se habla de dos versiones, la probabilística y la determinística.
- **Selección de estado uniforme:** sólo los individuos menos aptos son reemplazados en cada generación y es muy útil cuando los miembros de una población resuelven colectivamente un problema.

Evaluación: a los individuos seleccionados se les evalúa la *aptitud* para determinar si estos candidatos a soluciones son “buenos” o no lo son.

Es en este punto es donde se compara la medida de aptitud que se definió en el problema y que de hecho ya no es una codificación sino un número real. Por ejemplo:

Número	Individuo	aptitud
1	0111	95.00
2	0100	29.00
3	0101	47.00
4	0001	-1.00
5	0000	-3.00
6	1100	285.00
7	1000	125.00
8	1001	159.00

Cruza: es un proceso complejo que ocurre entre parejas de cromosomas. Estos cromosomas se alinean, luego se fraccionan en ciertas partes y posteriormente intercambian fragmentos entre sí. La representación básica de una cruce es la siguiente:



Las técnicas de cruce básica son:

- **Cruza de un punto:** se selecciona aleatoriamente un punto en el genoma y después se intercambian los genes de dos individuos previamente seleccionados.

Antes de la cruce	
1 0 0	0 1 0 1 0 1 1 1 0
1 0 1	0 1 1 1 0 0 1 0 1
Después de la cruce	
1 0 0	0 1 1 1 0 0 1 0 1
1 0 1	0 1 0 1 0 1 1 1 0

- **Cruza de dos puntos:** el genoma es tratado como un anillo, esto es, se supone que los extremos de un anillo genético están anexados unos a otros:

Antes de la cruce		
1 0 0	0 1 0 1 0 1	1 1 0
1 0 1	0 1 1 1 0 0	1 0 1
Después de la cruce		
1 0 0	0 1 1 1 0 0	1 0 1
1 0 1	0 1 1 1 0 0	1 0 1

- **Cruza uniforme:** los genes a ser cruzados son seleccionados aleatoriamente de tal manera que el número de genes intercambiados son equivalentes en una razón de $\frac{1}{2}$:

Antes de la cruce					
Después de la cruce					

Mutación: es la modificación intencional de algunos de los elementos del código genético de ciertos individuos con el fin de explorar posibles soluciones más aptas.

La mutación se considera como un operador secundario en los Algoritmos genéticos canónicos, es decir, su uso es menos frecuente que el de la cruce.

El papel que juega la mutación en el proceso evolutivo, así como su comparación con la cruce, sigue siendo tema frecuente de investigación y debate en la comunidad de computación evolutiva.

Entre otras, las siguientes son las técnicas de mutación que se han propuesto en la literatura especializada.

- Mutación para permutaciones:
 - Mutación por inserción.
 - Mutación por desplazamiento.
 - Mutación por intercambio recíproco.
 - Mutación heurística.
- Mutación para programación genética.
- Mutación para representación real:
 - Mutación no uniforme.

- Mutación de límite.
- Mutación uniforme.
- Mutación basada en parámetros.

5.3 Algoritmos genéticos y finanzas

Los Algoritmos Genéticos son un enfoque válido para muchos problemas prácticos en finanzas, los cuales pueden ser complejos y por lo tanto requerir del uso de técnicas eficientes y robustas de optimización.

Algunas aplicaciones de los Algoritmos genéticos a problemas complejos de los mercados financieros incluyen: Pronósticos de rendimientos, optimización de portafolios, identificar reglas de transacciones y optimización de transacciones.

5.3.1 Pronósticos de rendimientos

En 1990 Packard⁶³ muestra como los Algoritmos genéticos pueden ser utilizados para pronosticar datos complejos.

Un ejemplo es una aplicación de Levitt que desarrolla una nueva técnica de aprendizaje de máquinas a la que denominó *Aprendizaje local basado en la genética* que es una combinación sinérgica entre Algoritmos genéticos y técnicas de predicción local.

Se ha mostrado que este método produce rendimientos estadísticamente significativos cuando se aplica al mercado cambiario.

Los Algoritmos genéticos también han sido utilizados en administración de inversiones para pronosticar rendimientos de diferentes clases de activos con el objetivo de hacer una *TAA (tactical asset allocation)*. Leinweber y Arnott⁶⁴ (1995) muestran que aplicando un Algoritmo genético a sus modelos TAA pueden mejorar substancialmente el rendimiento de sus fondos domésticos.

Posteriormente, Mahfoud, Mani y Reigel⁶⁵ (1997) muestran como los Algoritmos genéticos pueden ser utilizados para predecir los rendimientos relativos para acciones individuales que es un aspecto muy utilizado en la localización estratégica de activos.

⁶³ PACKARD, N.H. (1990). "A Genetic Learning Algorithm for the Analysis of Complex Data. *Complex Systems*", 45(5), 543-572.

⁶⁴ LEINWEBER, D. ARNOTT, R. (1995). "Quantitative and Computational Innovation in Investment Management" .*Journal of Portfolio Management*, 21(2), 8-15.

⁶⁵ MAHFOUD, S., MANI, G. REIGEL, S. (1997). "Nonlinear Versus Linear Techniques for Selecting Individual Stocks". Citado en WEIGEND, A.S., ABUMOSTAFA, Y. Y REFENES, A-P. *Decision Technologies for Financial Engineering*, World Scientific, Singapore, pp 65-75.

Ellos encontraron que el Algoritmo genético tenía una ventaja significativa sobre los modelos lineales al identificar que acciones tendrían un desempeño superior e inferior en el mercado.

5.3.2 Optimización de portafolios

Otra importante función en la administración de inversiones es determinar la ponderación adecuada que se debe dar a los valores individuales en un portafolio. Este proceso es conocido como optimización o construcción de un portafolio.

Loraschi y Tattamanzi⁶⁶ (1996) usan un Algoritmo genético para encontrar la ponderación adecuada de los valores de un portafolio mediante la minimización de los factores de riesgo.

Ellos encontraron que este enfoque es útil cuando se encuentra con un gran espacio de soluciones caracterizado por tener varios óptimos locales.

Dada la reciente popularidad de los índices bursátiles, la importancia de la réplica de índices basada en un enfoque muestral se ha convertido en un aspecto importante para muchos administradores de fondos de pasivos; este problema es especialmente relevante para los administradores que intentan hacer coincidir una índices con una amplia base de valores.

Eddelbiittel⁶⁷ (1996) encuentra que un Algoritmo genético proporciona un enfoque computacionalmente eficiente al problema del seguimiento de los índices del mercado usando sólo un subconjunto de acciones que lograr comprimir el índice.

Esta aplicación empírica considera el índice del mercado alemán DAX; pero puede ser fácilmente extendido a mercados con índices más grandes y complejos.

Del Valle⁶⁸ (2002) muestra la aplicación de los Algoritmos Genéticos en un modelo para ayuda a la toma de decisiones en portafolios de inversión.

El modelo considera la selección de portafolios acorde a las preferencias del usuario, en cuanto a riesgo y rendimiento, utilizando como medidas

⁶⁶ LORASCHI, A., TETTAMANZI, A. (1996). *An Evolutionary Algorithm for Portfolio Selection Within a Downside Risk Framework*. in Dunis, C. (ed.). *Forecasting Financial Market*. John Wiley and Sons, Chichester, pp 275-285.

⁶⁷ EDDELBIITTEL, D. (1996). "A Hybrid Genetic Algorithm for Passive Management". Segunda Conferencia de Computación en Economía y Finanzas. Society of Computacional Economics. Ginebra.

⁶⁸ DEL VALLE, David. (2002). *Modelo para Toma de Decisión en Portafolios de Inversión con Acciones Usando Algoritmos Genéticos*. Tesis Doctoral. Tecnológico de Monterrey, CCM. Cd. De México.

de aptitud cada uno de los elementos en lo individual y el cociente de ambos.

Los resultados muestran que los Algoritmos genéticos pueden ser otra opción para la ayuda en la selección de portafolios de inversión.

También muestra que combinado con un modelo de programación lineal mejora sensiblemente sus resultados.

Parisi, Parisi y Cornejo⁶⁹ (2003) utilizan valores de cierre semanales, correspondientes al período entre el 07 de Abril de 1998 y el 14 de Abril de 2003, analizan la eficiencia de los modelos multivariantes dinámicos, construidos a partir de *Algoritmos genéticos recursivos*, para predecir el signo de las variaciones semanales de los índices bursátiles IPC, TSE, Nasdaq y DJI.

Los resultados fueron comparados con los de un modelo AR(1) y de un modelo multivariable construido en forma aleatoria.

Los mejores modelos producidos por el *Algoritmo genético* arrojaron un porcentaje de predicción de signo (PPS) de un 59%, 60%, 59% y 59%, para los índices IPC, Nasdaq, TSE y DJI, respectivamente.

La capacidad predictiva resultó significativa en cada uno de los índices. Al analizar el PPS de los modelos AR(1), se encontró que estos fueron menores, resultando significativos únicamente en el caso del Nasdaq.

Los modelos multivariantes dinámicos construidos en forma aleatoria registraron el PPS más bajo (excepto en el índice TSE), siendo significativo para el Nasdaq sólo al considerar un nivel de significancia del 10%.

Además, los modelos construidos por el *Algoritmo genético* generaron el mayor rendimiento acumulado, excepto en el caso del Nasdaq, donde la rentabilidad más alta fue registrada por el modelo AR(1).

Se observó que, independientemente de la significancia de la capacidad predictiva de los modelos de proyección, estos superaron en rentabilidad a la estrategia de inversión pasiva en todos los índices analizados (a excepción del TSE), por lo que prácticamente siempre fue mejor gestionar la cartera indexada en función de alguno de ellos.

⁶⁹PARISI, A., PARISI, F., CORNEJO, E. (2003). "Algoritmos Genéticos: Predicción de Índices Bursátiles de América del Norte". Facultad de Ciencias Económicas y Administrativas. Departamento de Administración. Universidad de Chile.

5.3.3 Descubrimiento de la reglas de intercambio

Los Algoritmos genéticos también se han usado para descubrir las reglas de intercambio rentables. Bauer⁷⁰ (1994) usa un Algoritmo genético para desarrollar las reglas de las oportunidades del mercado para las acciones norteamericanas y los mercados de bonos.

Estas reglas son formuladas usando datos macroeconómicos mensuales para descubrir relaciones entre los mercados financieros y la economía. Los resultados obtenidos son consistentes con la habilidad de las oportunidades de mercado y ganancias pequeñas.

Sin embargo la importancia estadística de estas ganancias no se investiga. En un estudio posterior Bauer⁷¹ (1995) aplica esta metodología al mercado cambiario donde una vez más se encuentran resultados prometedores.

Allen y Karjalainen⁷² (1994) usan la técnica de la programación genética para descubrir reglas técnicas de intercambio rentables usando datos tomados del índice accionario S&P500.

La programación genética, desarrollada por Koza⁷³, es similar a un Algoritmo genético pero es menos restrictivo dado que no está atada a la longitud de la representación binaria de las soluciones.

Se encontró que la rentabilidad de estas reglas se puede generalizar a un período de prueba fuera de la muestra y es estadísticamente significativa.

Sin embargo, recientemente, Allen y Karjalainen⁷⁴ (1999) usando una serie de precios de acciones mucho mayor y modificando el diseño del experimento, descubrieron que hay una pequeña, tal vez ninguna, evidencia de rendimientos en exceso atribuibles a las reglas de intercambio.

Debido a los relativamente bajos costos de transacción en el mercado cambiario, posiblemente hay mayores oportunidades para el

⁷⁰BAUER, R.J. Jr. (1994), *Genetic Algorithms and Investment Strategies*, Wiley Finance Editions, John Wiley and Sons, New York.

⁷¹BAUER, R.J. Jr. (1995), "Genetic Algorithms and the Management of Exchange Rate Risk," in BIETHAHN, J., NISSEN, V. (eds.) *Evolutionary Algorithms in Management Applications*, Springer Verlag, Berlin, Heidelberg, New York, pp 253-263.

⁷²ALLEN, F., KARJALAINEN, R. (1994), "Using Genetic Algorithms to Find Technical Trading Rules," Working Paper, The Wharton School, University of Pennsylvania.

⁷³KOZA, J.R. (1992), *Genetic Programming: On the Programming of Computers By the Means of Natural Selection*, MIT Press, Cambridge, MA.

⁷⁴ALLEN, F., KARJALAINEN, R. (1999), "Using Genetic Algorithms to Find Technical Trading Rules" , The Wharton School. University of Pennsylvania.

descubrimiento de reglas de intercambio rentables usando tecnología computacional avanzada.

Las divisas redituables son encontradas por medio de la simulación por Colin⁷⁵ (1996) y Neely, Weller y Dittmar⁷⁶ (1997) aplicando un enfoque de programación genético a los numerosos tipos de cambio. Neely et al encontraron que las reglas de intercambio descubiertas por ellos generan excesos de retornos estadísticamente significativos.

Por otro lado, Colin no toma en cuenta ningún costo de transacción y no considera la significancia estadística de sus resultados.

5.3.4 Optimización de las reglas de intercambio

La aplicación de las reglas de intercambio basada en indicadores técnicos o fundamentales del mercado financiero requiere de la selección de valores apropiados de los parámetros.

En la práctica los intermediarios normalmente escogen estos parámetros de una manera subjetiva principalmente basados en la intuición y la experiencia.

También, numerosos estudios que examinan la rentabilidad en los mercados financieros han ignorado el aspecto de la optimización de los parámetros o han usado valores de los parámetros determinados a posteriori.

Esta práctica puede llevar a un sesgo de los datos véase Lo y MacKinley⁷⁷ (1990).

Un enfoque más objetivo y válido al problema de la selección de los parámetros involucra el uso de datos históricos. Para conducir una evaluación válida del desempeño de las reglas de intercambio libre de sesgos necesario escoger los valores de los parámetros a priori.

Esto puede lograrse usando un período tiempo previo a la muestra para determinar el valor a priori óptimo de los parámetros.

Hay un par de estudios que han aplicado los Algoritmos genéticos al problema técnico de la optimización de los parámetros en las reglas de

⁷⁵ COLIN, A. (1996), "A Genetic Programming-Based Approach to Generation of Foreign Exchange Trading Models, Conferencia de "Commerce, Complexity and Evolution, University of New South Wales". Citado por Perreira Ibid 28.

⁷⁶ NEELY, C.J., WELLER, P., DITTMAR, R. (1997), "Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach," *Journal of Financial Quantitative Analysis*, 32 (4), 405-426.

⁷⁷ LO, A.W., MACKINLEY, A.G. (1990), "Data Snooping Biases in Tests of Financial Asset Pricing Models," *The Review of Financial Studies*, 3, 431-467.

intercambio. Klimasauskas⁷⁸ (1995) desarrolla un indicador múltiple de las oportunidades del mercado y usa un Algoritmo genético para perfeccionar los parámetros del modelo. Pictet, Dacorogna, Dave, Chopard, Schirru y Tomassini⁷⁹ (1996) introducen la idea de optimización robusta en las técnicas de las reglas de intercambio usando los Algoritmos genéticos con las acciones recolectadas.

La optimización robusta se preocupa por encontrar valores de los parámetros con que no son necesariamente consistentes con el óptimo global pero que se encuentra en las regiones altas y llanas de la superficie de ganancia en el espacio de los parámetros.

⁷⁸ KLIMASAUSKAS, C. (1995), "Developing a Multiple-Indicator Market Timing System: Theory, Practice and Pitfalls," in Lederman, J. and Klein, R.A. (eds.), *Virtual Trading*, Irwin Publishing, New York, pp 127-166.

⁷⁹ PICTET, OX., DACOROGNA, M.M., DAVE, R.D., CHOPARD, B., SCHIRRU, R., TOMASSINI, M. (1996), "Genetic Algorithms with Collective Sharing for Robust Optimization in Financial Applications," Olsen and Associates, Zurich, Suiza.

Capítulo 6

Especificaciones

Las especificaciones que debemos establecer para desarrollar esta investigación son:

1. ¿Cuántas y cuáles acciones van a constituir nuestra población?
2. ¿Cuántas y cuáles acciones van a constituir la muestra?
3. ¿De cuántas acciones se van a constituir los portafolios?
4. ¿Cuáles son los posibles escenarios para el análisis de la información?

El segundo punto de los citados anteriormente es particularmente importante porque es ahí donde aplicamos la técnica del minado de datos que hemos seleccionado para utilizar en esta investigación, esto es, el análisis de agrupamientos.

Como es posible inferir, a partir de la selección de la población buscaremos la muestra de acciones a las que aplicaremos el programa de búsqueda enumerativa y los Algoritmos genéticos, es decir, tenemos que reducir de N (tamaño de la población) a n (tamaño de la muestra) acciones.

En la sección 6.2 damos las motivaciones para usar el análisis de agrupamientos, los pasos que se siguieron y los resultados producidos.

6.1 Selección de la población

La definición que probablemente sea la más exacta acerca del arte de pronosticar es aquella que dice que "pronosticar es sacar conclusiones sobre el futuro, en base a la información obtenida hasta el presente"⁸⁰; esta definición deja muy en claro que para poder sacar una conclusión sobre el futuro es una condición sine qua non tener información histórica de lo que en el pasado ha ocurrido sobre el fenómeno o problema en particular que se va a estudiar.

La información histórica de la BMV se puede conseguir fundamentalmente a través de la misma BMV, Invertia Plus®, Terra® y Economática®. Nos decidimos por esta última fundamentalmente por lo completo y exacto de

⁸⁰ BURES, Esperanza (1986). *Métodos de pronósticos aplicados a la administración*. ITESM. Monterrey.

la información que maneja, por su amigabilidad en el uso y por su compatibilidad con Microsoft Office® al poder trasladar la información de sus bases de datos a Excel® con un simple procedimiento de copiar-pegar.

Para determinar a la población decidimos incluir nueve variables que son:

Tipo de activo: de entre todas las posibilidades posibles para nuestro estudio sólo seleccionamos las *acciones*.

Estatus: el estudio pretende generar una herramienta de utilidad para el futuro, por lo que del grupo de acciones sólo nos interesaron aquellas con estatus de *activo*.

Precio de cierre: es el último valor de la acción en el día.

Rendimiento (R): es el beneficio dividido por la cantidad invertida. Formalmente, el rendimiento R, está dado por la fórmula:

$$R = \frac{P_i - P_{i-1}}{P_{i-1}}$$

Ecuación 6.1: rendimiento de un portafolio

Volatilidad (σ): es la medida de la incertidumbre acerca de los rendimientos que proporcionará una acción. Los valores típicos de la volatilidad de una acción están entre 20% a 40% por año. Esto se obtiene simplemente con la raíz cuadrada de la variancia esperada. Formalmente σ esta dada por:

$$\sigma_i = \sqrt{\sum_{j=1}^n p_i [R_{ij} - E(R_i)]^2}$$

Ecuación 6.2: volatilidad de un portafolio

Beta (β): mide la contribución de una acción individual a la variancia del portafolio del mercado como una fracción de la variación total del portafolio del mercado.

Esto se obtiene del cociente de la covariancia entre el rendimiento individual y el rendimiento del mercado dividido por la volatilidad del mercado elevado al cuadrado. Formalmente la β esta dada por:

$$\beta = \frac{COV(R_i, R_M)}{\sigma_M^2}$$

Ecuación 6.3: beta de un portafolio

Donde:

$$Cov(R_i, R_M) = \sum_{j=1}^n p_i [R_{ij} - E(R_i)][R_{Mj} - E(R_M)] \quad \sigma_M^2 = \sum_{j=1}^n p_i [R_{Mj} - E(R_M)]^2$$

Adicionalmente se agregaron las variables: fecha de última operación, fecha de inicio de la serie y volumen en pesos del último día de operaciones.

Una vez calculado esto para cada acción, se procedió a hacer un filtrado final para desechar todas aquellas acciones de las que no fue posible calcular algunas de las variables a considerar para hacer el análisis de agrupamientos, quedando después de esto el número final de 58 acciones para trabajar con ellas.

6.2 Selección del tamaño y de las componentes de la muestra

Una posible decisión era simplemente usar la base de datos entera de las 58 acciones que habían resultado del proceso de filtrado que es explicado posteriormente.

El problema resultaba de los cálculos que se tendría que hacer para el diseño de los portafolios.

Si, por ejemplo, se decidiera que los portafolios fueran de 10 acciones, entonces resultaba que para cada día de información de los 10 años de datos se tendría una total de 52,179´482,350 combinaciones posibles.

Si además a esto le sumáramos que en el escenario de días la base de datos tiene 2504 días, entonces para ese escenario se tendrían que buscar la combinación óptima de entre un total de 130´657,423´800,000 posibles.

El número resulta muy grande y por supuesto justifica la utilización tanto del Algoritmo de conteo y de los Algoritmos genéticos, sin embargo habría que considerar en términos del tiempo de uso computacional que tan buena es una diversificación en 58 acciones.

La cuestión es verificar si vale la pena hacer un uso tan extenso de información o bien si está puede reducirse de tal manera que aun teniendo un gran número de cálculos, se puede encontrar una muestra representativa.

6.2.1 Selección del tamaño de la muestra

La selección del tamaño de la muestra se obtuvo del trabajo de Statman⁸¹ que derivó de Elton et al.⁸² que un portafolio

⁸¹ STATMAN, M. (1987) *How many stocks make a diversified portfolio?* Journal of Financial and Quantitative Analysis 22 sep-1987. 353-364.

⁸² Elton, E.J. & Gruber, M.J. *Risk Reduction and portfolio size: An analytic solution* Journal of Business 50 oct-1977

diversificado muestra que una reducción significativa en el riesgo y un aumento en el rendimiento sólo hasta que tiene un tamaño de 40 componentes, a partir de ahí el agregar más miembros a la población de acciones reduce el riesgo y aumenta el rendimiento de manera marginal, por lo tanto no hay diferencia significativa entre hacer un estudio con 58 acciones o con 40.

Pero en términos de cálculos las combinaciones de 40 de 10 son solamente de 847 660,528 que al multiplicarlas por 2504 nos da un total de 2 122,541 962,000 combinaciones, un número suficientemente grande para que las conclusiones y las aplicaciones se justifiquen, pero que reduce los tiempos de cómputo en aproximadamente 15 semanas, sin pérdida de generalidad en los resultados.

En esencia el principio fundamental desarrollado por Statman es que *la diversificación reduce el riesgo, pero sólo hasta cierto punto.*

Las siguientes tabla y gráfica las presentan Corrado y Jordan⁸³ y las citan precisamente de Statman⁸⁴

Número de acciones en el portafolio	Riesgo promedio anual del rendimiento del portafolio	Razón del riesgo del portafolio al riesgo de un activo individual
1	49.24%	1.00
2	37.36%	0.76
4	29.69%	0.60
6	26.64%	0.54
8	24.98%	0.51
10	23.93%	0.49
20	21.68%	0.44
30	20.87%	0.42
40	20.46%	0.42
50	20.20%	0.41
100	19.69%	0.40
200	19.42%	0.39
300	19.34%	0.39
400	19.29%	0.39
500	19.27%	0.39
1000	19.21%	0.39

Tabla 6.1: diversificación del riesgo

Como puede observarse de la tabla y de la gráfica, no hay un mejoramiento significativo al incrementar el número de

⁸³ Corrado, C.J. & Jordan B.D. *Fundamentals of Investments: valuation and management* Irwin McGraw-Hill 2000 Boston

⁸⁴ Ibid

acciones más allá de 40. Debido a esta evidencia decidimos tomar una muestra de 40 acciones, que para determinar cuáles iban a ser utilizamos el análisis de agrupamientos con las etapas que especificamos en la siguiente sección.

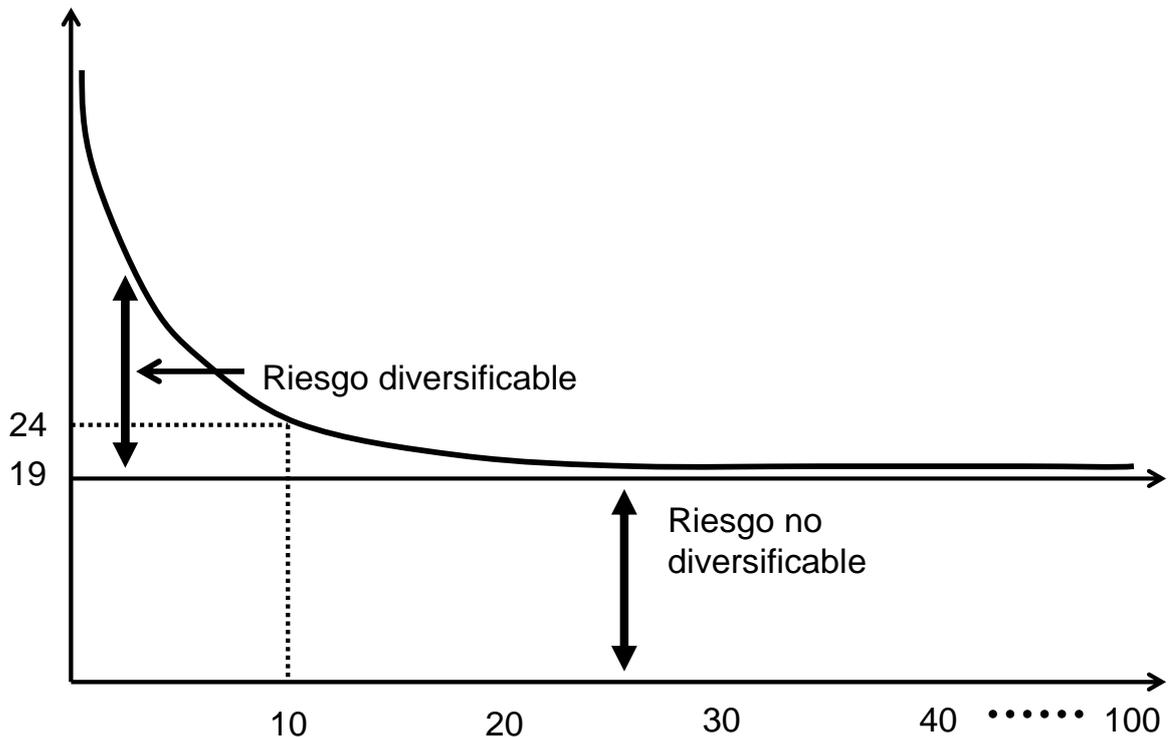


Figura 6.1: diversificación del riesgo

6.2.2 Selección de los componentes de la muestra

Para determinar esto es que precisamente se llevó a cabo el análisis de conglomerados.

De nuestro total de 58 debíamos elegir 40. Este es un proceso de selección que hicimos con las definiciones del análisis de conglomerados previamente presentadas.

Para hacer la agrupación utilizamos el paquete estadístico SPSS® en su versión 10.0 para Windows.

Una vez determinados los grupos había que seleccionar un elemento de cada uno de ellos para que fuera aquel que iba a formar parte de la muestra que nos permitiría pronosticar el mejor

portafolio usando tanto una técnica enumerativa como un Algoritmo genético.

La decisión fue simple, del conjunto de acciones de cada grupo seleccionaríamos aquella que tuviera el mayor rendimiento promedio en la base de datos de Economatica.

A continuación presentamos los 40 grupos con sus respectivas acciones y el rendimiento promedio de cada una de ellas almacenada en la base de datos.

GRUPO 1	
ACCION	RENDIMIENTO
Accel S.A. B	0.0484%
Ekco	0.0222%
Savia A	0.0190%
Maseca GI B	0.0076%

GRUPO 2	
ACCION	RENDIMIENTO
Cemex S.A. CPO	0.0927%
Alfa S.A. A	0.0894%

GRUPO 3	
ACCION	RENDIMIENTO
Apasco S.A.	0.1021%
Penoles Industrias	0.1018%
Bimbo Gpo A	0.0983%
Kimberly Clark Me	0.0960%

GRUPO 4	
ACCION	RENDIMIENTO
Ara Consorcio	0.1021%

GRUPO 5	
ACCION	RENDIMIENTO
Fragua Corporativ	0.0556%
Bachoco Industria	0.0131%

GRUPO 6	
ACCION	RENDIMIENTO
Telefs de Mex A	0.1110%
GInd Saltillo	0.1092%
Posadas Gpo L	0.1066%
Cementos Chihua	0.0937%

GRUPO 7	
ACCION	RENDIMIENTO
Cid Mega Resort	-0.1600%

GRUPO 8	
ACCIÓN	RENDIMIENTO
Wal Mart de Mexi	0.0888%
Comercial Mexica	0.0351%

GRUPO 9	
ACCIÓN	RENDIMIENTO
Continental Grupa	0.1520%
Desc Soc Fom In	0.0924%

GRUPO 10	
ACCIÓN	RENDIMIENTO
Corp Interam de B	0.1595%

GRUPO 11	
ACCIÓN	RENDIMIENTO
Dataflux B	-0.0621%

GRUPO 12	
ACCIÓN	RENDIMIENTO
Hylsamex B	0.0945%
Herdez S.A.	0.0924%
Desc Soc Fom In	0.0753%
Gigante Gpo	0.0497%

GRUPO 13	
ACCIÓN	RENDIMIENTO
Elektra Gpo	0.0758%

GRUPO 14	
ACCIÓN	RENDIMIENTO
Fomento Econ Me	0.1527%
Wal Mart de Mexi	0.0931%

GRUPO 15	
ACCIÓN	RENDIMIENTO
GCarso A1	0.0558%

GRUPO 16	
ACCIÓN	RENDIMIENTO
Geo Corporacion	0.0478%

GRUPO 17	
ACCIÓN	RENDIMIENTO
GFBanorte O	10.8000%

GRUPO 18	
ACCIÓN	RENDIMIENTO
GFBVA Bancom	0.0910%

GRUPO 19	
ACCION	RENDIMIENTO
39:KOF Coca-Col	0.1441%
24:GFInbursa O	0.0874%

GRUPO 20	
ACCION	RENDIMIENTO
GMaseca (Gruma	-0.0058%

GRUPO 21	
ACCION	RENDIMIENTO
GMex Desarrollo	-0.3887%

GRUPO 22	
ACCION	RENDIMIENTO
GMexico B	3.7800%

GRUPO 23	
ACCION	RENDIMIENTO
GModelo C	0.1051%

GRUPO 24	
ACCION	RENDIMIENTO
GNacional Provin	0.4818%

GRUPO 25	
ACCION	RENDIMIENTO
Hogar Consorcio	-0.1351%

GRUPO 26	
ACCION	RENDIMIENTO
ICA Soc Controla	-0.0078%

GRUPO 27	
ACCION	RENDIMIENTO
Imsa Gpo UBC	0.0292%

GRUPO 28	
ACCION	RENDIMIENTO
Iusacell Gpo	-0.0964%

GRUPO 29	
ACCION	RENDIMIENTO
Lamosa Gpo B	0.0172%

GRUPO 30	
ACCION	RENDIMIENTO
Posadas Gpo A	0.0613%

GRUPO 31	
ACCION	RENDIMIENTO
San Luis Corp CF	-0.0173%

GRUPO 32	
ACCION	RENDIMIENTO
Simec Grupo B	0.0270%

GRUPO 33	
ACCION	RENDIMIENTO
Soriana Organiza	0.1299%

GRUPO 34	
ACCION	RENDIMIENTO
Tekchem S.A. A	-0.0022%

GRUPO 35	
ACCION	RENDIMIENTO
Telecom Carso G	0.1034%

GRUPO 36	
ACCION	RENDIMIENTO
Telefs de Mex L	0.1024%

GRUPO 37	
ACCION	RENDIMIENTO
Televisa Gpo CP	0.0783%

GRUPO 38	
ACCION	RENDIMIENTO
TV Azteca CPO	-0.0312%

GRUPO 39	
ACCION	RENDIMIENTO
Valle Jugos del B	0.0668%

GRUPO 40	
ACCION	RENDIMIENTO
Vitro A	0.0291%

De tal manera que después de encontrados los 40 grupos, las acciones seleccionadas son:

Accel S.A. B	Cemex S.A. CPO	Apasco S.A.	Ara Consorcio	Fragua Corporativo B
Telefs de Mex A	Cid Mega Resort	Wal Mart de Mexico C	Continental Grupo	Corp Interam de Ent
Dataflux B	Hylsamex B	Elektra Gpo	Fomento Econ Mex UBD	GCarso A1
Geo Corporacion B	GFBanorte O	GFBBVA Bancomer B	KOF Coca-Cola L	GMaseca (Gruma) B
GMex Desarrollo B	GMexico B	GModelo C	GNacional Provincia	Hogar Consorcio B
ICA Soc Controlad	Imsa Gpo UBC	Iusacell Gpo	Lamosa Gpo B	Posadas Gpo A
San Luis Corp CPO	Simec Grupo B	Soriana Organizacio	Tekchem S.A. A	Telecom Carso Globa
Telefs de Mex L	Televisa Gpo CPO	TV Azteca CPO	Valle Jugos del B	Vitro A

Tabla 6.2: muestra de acciones seleccionadas

Como se puede observar, en esta lista no están aquellas acciones que durante el periodo dieron el mayor rendimiento; sino las que una vez dentro de un grupo eran las que tenían el mayor rendimiento.

También notamos que hubo muchos grupos que finalmente quedaron con una sola acción.

Esto lo consideramos bueno porque la interpretación superficial que le damos es que sólo se agruparon las acciones con un comportamiento tan similar que haber seleccionado una u otra era irrelevante para los portafolios resultantes.

Aun cuando hemos seleccionado acciones con un rendimiento negativo, si conservan su individualidad y por lo tanto entran en el conjunto de aquellas que es interesante estudiar.

6.3 Determinación del tamaño del portafolio

Como se hace notar en la sección 6.2, el tamaño del portafolio es de diez⁸⁵ acciones. Weston y Copeland⁸⁶ hacen referencia a un estudio empírico realizado por Wagner y Lau⁸⁷ donde demuestran los efectos de las diversificaciones en portafolios mayores a 10 acciones.

Los resultados que encuentran es que a medida que aumenta el número de valores en un portafolio, disminuye el riesgo de los rendimientos del portafolio, con reducciones que son más o menos pequeñas después de que se incluyen unos 10 valores en el portafolio.

En este sentido tanto la experiencia práctica usada por los profesionales, así como la investigación desarrollada desde inicios de los años setentas, sugieren que la formación de portafolios de 10 acciones es una buena base de partida para la simulación de la información financiera.

6.4 Escenarios posibles para el análisis de la información

Definimos tres escenarios posibles para la comparación de los rendimientos de los portafolios:

- Información de rendimientos diarios.

⁸⁵ La Lic. María del Carmen Salinas Fernández, subdirectora de promoción de Ixe Grupo Financiero, señala que un portafolio pequeño es aquel que tiene al menos 8 acciones y que uno ideal sería de 12 para estar bien diversificados. En ese sentido decidimos tomar 10 como el número intermedio.

⁸⁶ Weston, J.F. Copeland, T.E. *Managerial Finance* 11th Edition. 1999 Boston.

⁸⁷ Wagner, W.H. Lau, S.C. *The effect of diversification on risk* Financial Analyst Journal, nov-dec 1971. pp. 48-53

- Información de rendimientos semanales.
- Información de rendimientos mensuales.

En el primer caso, información diaria, la información es muy abundante y completa.

Sin embargo esta misma cantidad de información provoca que las variaciones extraordinarias sean magnificadas ya que un rendimiento extraordinariamente alto o bajo, se refleja en el resultado global sin haber permitido identificar si dicha variación es permanente o se debió simplemente a un cambio coyuntural de un día particular.

En el segundo caso, información semanal, hay una reducción importante en el número de datos, sin embargo de alguna manera la información puede ser más precisa ya que el rendimiento promedio refleja mejor los cambios que hubieran ocurrido al pasar de un periodo a otro.

De hecho, los administradores de fondos de inversión tienden a recomponer sus portafolios en función de pronósticos semanales.

En el tercer caso, información mensual, los datos se reducen aun más, la información es más manejable, es posible identificar más fácilmente valores extraordinarios; sin embargo en muchas ocasiones un periodo de un mes se considera un lapso demasiado grande para la toma de decisiones bursátiles.

Cada escenario podría tener sus ventajas y desventajas; nosotros tomamos la decisión de usar estos tres escenarios para hacer comparaciones entre los modelos conteos y de Algoritmos genéticos para cada uno de los tres lapsos de tiempo.

Capítulo 7

Selección del portafolio que mejor predice el comportamiento de la Bolsa Mexicana de Valores usando un Algoritmo de conteo

Una vez seleccionada la muestra procedimos a encontrar un portafolio de diez acciones que pronostique el rendimiento del mercado mexicano utilizando un Algoritmo de conteo.

7.1 Diseño del Algoritmo de conteo

El objetivo principal de este Algoritmo es que sea capaz de interpretar una Base de Datos de Access, normalizarla y mostrar el número de veces que combinaciones de diez acciones cumplen con el requisito de sobrepasar el promedio de todas las acciones en un período de tiempo dado con el promedio del rendimiento de las diez acciones que constituyen la combinación.

Una opción era realizar todo el trabajo de manera exhaustiva, es decir, realizar una a una todas las posibles combinaciones sin algún criterio de selección predefinido, pero después de haber simulado con una base de datos de prueba con seis acciones, llegamos a la conclusión que sería más sencillo y práctico poder distribuir el trabajo de las combinaciones a lo largo de su recorrido por las filas (los períodos de tiempo) para poder contar con la opción del paralelismo y finalmente unir todos los resultados para poder realizar las observaciones necesarias.

A fin de poder cumplir las metas propuestas para la aplicación se desarrollaron diferentes Algoritmos para realizar las siguientes acciones:

1. Sacar las combinaciones apropiada y efectivamente
2. Normalizar la Base de Datos para realizar las operaciones necesarias
3. Crear y hacer uso de tablas auxiliares que permitieran separar el trabajo en partes
4. Borrar la Base de Datos y las tablas que se usarían para recomenzar el procedimiento

La base del proyecto es el Algoritmo que permitirá sacar las combinaciones para después ir las probando.

Este Algoritmo debía tener la capacidad de poder ser interrumpido y reiniciar en caso de ocurrir una falla o simplemente para posponer los cálculos.

Para cumplir con estos requisitos era necesario que el Algoritmo no fuese recursivo sino determinístico.

El Algoritmo resultante presentado en pseudo-código, así como los procedimientos desarrollados se presenta en el anexo correspondiente.

7.2 ¿Qué hace el Algoritmo de conteo?

Una vez presentada la forma en que trabaja este Algoritmo explicamos exactamente qué es lo que hace.

A través de cada una de las filas de fecha obtienen en primer lugar el rendimiento promedio del mercado para, después comienza a formar los portafolios de 10 acciones por cada una de las filas, si el rendimiento de dicho portafolio es mayor que el rendimiento del mercado lo almacena, en caso contrario lo desecha.

Una vez que cubre toda la fila compara los rendimientos de todos los portafolios con un rendimiento superior al del mercado y de entre estos selecciona a aquel que haya obtenido el mayor rendimiento y lo almacena.

Esto lo hace para cada una de las filas y al terminar con todas las filas presenta de manera ordenada los portafolios que más veces se repitieron en el transcurso del periodo de tiempo señalado.

Entonces tomamos a aquel que más se repite como el portafolio representativo de esta metodología para compararlo contra los encontrados por un procedimiento de Algoritmos genéticos.

Los valores resultantes están representados con el valor del código de la acción definido en el SPSS y cuyos valores definidos para nuestra muestra son:

cod_emi	Acción
1	Accel S.A. B
7	Apasco S.A.
8	Ara Consorcio
15	Cemex S.A. CPO
16	Cid Mega Resort
19	Continental Grupo
20	Corp Interam de Ent B
22	Dataflux B
26	Elektra Gpo
27	Fomento Econ Mex UBD
28	Fragua Corporativo B
29	GCarso A1
32	Geo Corporacion B
33	GFBanorte O
34	GFBBVA Bancomer B
39	GMex Desarrollo B
40	GMexico B
41	GModelo C
42	GNacional Provincia
46	Hogar Consorcio B
47	Hylsamex B
48	ICA Soc Controlad
49	Imsa Gpo UBC
50	Iusacell Gpo
52	KOF Coca-Cola L
53	Lamosa Gpo B
54	Maseca GI B
59	Posadas Gpo A
61	San Luis Corp CPO
63	Simec Grupo B
64	Soriana Organizacio B
65	Tekchem S.A. A
66	Telecom Carso Globa A1
67	Telefs de Mex A
68	Telefs de Mex L
69	Televisa Gpo CPO
70	TV Azteca CPO
72	Valle Jugos del B
73	Vitro A
74	Wal Mart de Mexico C

7.3 Selección del mejor portafolio

7.3.1 Escenario diario

Para el escenario diario, el Algoritmo arrojó como mejor resultado el portafolio

cod_emi	acción
7	Apasco S.A.
15	Cemex S.A. CPO
29	GCarso A1
34	GFBBVA Bancomer B
48	ICA Soc Controlad
54	Maseca GI B
64	Soriana Organizacio B
69	Televisa Gpo CPO
73	Vitro A
74	Wal Mart de Mexico C

7.3.2 Escenario semanal

Para el escenario semanal, el Algoritmo arrojó como mejor resultado el portafolio

cod_emi	acción
7	Apasco S.A.
15	Cemex S.A. CPO
19	Continental Grupo
27	Fomento Econ Mex UBD
29	GCarso A1
34	GFBBVA Bancomer B
48	ICA Soc Controlad
64	Soriana Organizacio B
67	Telefs de Mex A
74	Wal Mart de Mexico C

7.3.3 Escenario mensual

Para el escenario semanal, el Algoritmo arrojó como mejor resultado el portafolio

cod_emi	emisora
7	Apasco S.A.
15	Cemex S.A. CPO
19	Continental Grupo
27	Fomento Econ Mex UBD
48	ICA Soc Controlad
54	Maseca GI B
64	Soriana Organizacio B
67	Telefs de Mex A
69	Televisa Gpo CPO
74	Wal Mart de Mexico C

Capítulo 8

Selección del portafolio que mejor predice el comportamiento de la Bolsa Mexicana de Valores usando un Algoritmo genético con tres diferentes medidas de aptitud

8.1 El Algoritmo. El objetivo principal de este proyecto es desarrollar un Algoritmo genético que permita predecir el comportamiento del mercado dando tres medidas de aptitud: el promedio histórico observado de cada una de las acciones, la distribución normal tomando como media y desviación estándar los valores calculados a partir de la información de la muestra y, de manera experimental, una medida de aptitud que siguiera un comportamiento del tipo Movimiento Geométrico Browniano con las definiciones de los parámetros que explicaremos en la sección particular dedicada a este modelo.

El uso de las tres medidas de aptitud tiene justificaciones independientes.

El uso del promedio es de manera general la medida tradicional que permite seleccionar el mejor portafolio de inversión.

Esto es, un inversionista preferirá siempre, en ausencia de mayor información, aquel portafolio cuyo rendimiento sea mayor.

El uso de la distribución normal es porque le agrega a la conformación del portafolio la medida del riesgo.

Recordemos que la medida de la utilidad del portafolio es la media, mientras que la medida del riesgo es la desviación estándar, así que esta medida de aptitud incluiría tanto el rendimiento como el riesgo.

Un inversionista con actitud de adversidad al riesgo siempre preferirá un portafolio con menor riesgo; en oposición con un inversionista del tipo living risk que preferirá un portafolio de mayor rendimiento.

La utilización del movimiento geométrico browniano como medida de aptitud se basa en el hecho de que el desarrollo de los modelos de ablución de opciones y de formación de portafolios eficientes y libres de riesgo se basan en este modelo; así los desarrollos del Lema de Îto que son utilizados para la prueba de la Black & Scholes tienen su sustento teórico en estos modelos, por lo que resulta interesante identificar que pasa en el mercado mexicano con los modelos de Brown.

Como ya hemos sugerido previamente, la validez de los Algoritmos genéticos se basa en el teorema presentado por Holland⁸⁸ en 1992, el cuál es conocido como

⁸⁸ Ibid 36

"Teorema de los Esquemas" y dice: "los esquemas pequeños de bajo orden arriba del promedio reciben un incremento exponencial de representantes en las siguientes generaciones de un Algoritmo genético".

Lo que puede explicarse sencillamente de la siguiente manera:

Los esquemas se construyen usando ceros, unos y comodines, los cuáles se representan con asteriscos.

Los comodines bien pueden tomar el valor de uno o cero, permitiéndonos así representar con un esquema varias cadenas. Por ejemplo, el esquema $1^{**}10$ representa a las cadenas: **100**10, **101**10, **110**10, **111**10. El "orden del esquema" está definido por el número de elementos que no son comodines en el esquema dado, en el ejemplo anterior podemos decir que el esquema $1^{**}10$ es un esquema de **tercer** orden.

También se define a la "longitud del esquema" como el número de elementos entre la primer posición fija, es decir, que no es comodín y la última posición fija.

Pero, ¿qué significa esto? Bueno, dado que un Algoritmo genético opera con una población en cada iteración se espera que el método converja de modo que al final de su proceso la población sea muy semejante, y en el caso de iteraciones infinitas, que la población sea reducida solamente a un solo individuo o a varias copias del mismo según sea la implementación del Algoritmo.

Para probar esta convergencia se ha desarrollado la teoría conocida como "el Teorema de Esquemas", el cual estudia el caso de las cadenas binarias y se basa principalmente en la consideración de que una cadena es un representante de una clase de equivalencia, un esquema, reinterpretando la búsqueda entre esquemas en lugar de entre cadenas. De este modo podemos concluir en lo conocido como paralelismo intrínseco que dice "en una población de n cadenas se están procesando implícitamente $O(n^2)$ esquemas⁸⁹".

A partir de ese resultado el teorema de esquemas prueba que la población converge a unos esquemas que cada vez son más y más parecidos y en límite a una única cadena.

Los usos de los Algoritmos genéticos bien pueden dividirse en cinco áreas bien definidas, las cuales son: búsqueda, clasificación, aprendizaje de máquina, predicción y optimización.

En este trabajo el área que se cubrirá será la de la predicción ya que dada una base de datos con rendimientos financieros, además de ciertas restricciones, correremos el Algoritmo durante unas generaciones para predecir el conjunto de portafolios que mayor rendimiento tendrá en el futuro.

⁸⁹ O a lo más $n \cdot 2^n$ esquemas para ser exactos.

Particularmente haremos una comparación del uso de este Algoritmo genético contra el método de conteo del capítulo anterior.

Una vez asentadas las bases de este teorema, veremos la composición de un Algoritmo genético tradicional.

Ya hemos mencionado que los Algoritmos genéticos son el Darwinismo aplicado en las ciencias computacionales.

En palabras simples se genera aleatoriamente un conjunto de soluciones iniciales, es decir una población, y luego ésta va evolucionando poco a poco utilizando tres operadores fundamentales: selección, cruzamiento y mutación hasta que en un número de iteraciones preestablecido por alguna tolerancia de error se obtiene una solución aceptable al problema.

Esta sencilla idea nos permite deducir que no es necesario poseer una cantidad extremadamente grande de información acerca del problema objeto de estudio⁹⁰ y que con los métodos tradicionales sería una condición imprescindible.

El Algoritmo genético básico estará conformado por las siguientes partes:

1. **Inicio:** Generar una población al azar o, como en nuestro caso, tomar una población dada de n cromosomas que podrían ser posibles soluciones al problema presentado.
2. **Aptitud:** Se evalúa la aptitud de cada cromosoma en la población.
3. **Nueva Población:** Se crea una nueva población siguiendo los pasos que se detallarán a continuación hasta que se tiene una nueva población:
 - a. **Selección:** Se seleccionan dos cromosomas padres de una población de acuerdo a su aptitud, mientras mejor aptitud tengan, mayor es la probabilidad de ser seleccionados.
 - b. **Cruza:** Dada una probabilidad de cruce se intercambia la información genética de los dos padres para formar un par de hijos, en caso que la probabilidad no permita el cruce, los hijos serán una copia de los padres.
 - c. **Mutación:** Dada una probabilidad de mutación se altera al azar un alelo del cromosoma.
 - d. **Aceptación:** Los hijos toman su lugar en la nueva población.
4. **Reemplazo:** Usar la nueva población generada para la próxima iteración del Algoritmo.

⁹⁰ Por supuesto es necesario tener un conocimiento básico a priori del problema y su medio ambiente para definir la función de aptitud adecuada.

5. **Prueba:** Si la condición de paro es satisfecha entonces se para la corrida y se arroja la mejor solución de la población actual.
6. **Ciclo:** Se realiza todo lo anterior comenzando desde el paso número 2.

Una vez que ya hemos conocido el procedimiento que nos permitirá desarrollar nuestro Algoritmo genético, definiremos las bases de nuestro diseño para así continuar con la programación y la justificación de la misma.

El desarrollo del algoritmo y sus especificaciones y pasos se muestran en el anexo correspondiente.

8.1.1 Resultados del Algoritmo genético usando como función de aptitud el rendimiento promedio

La función de aptitud usada en este AG fue el promedio histórico de todas las acciones en sus respectivos escenarios de tiempo.

Una de las principales decisiones al correr un Algoritmo genético es cuántas veces deberá ejecutarse dicho Algoritmo hasta determinar los resultados óptimos.

Por sugerencia personal de Coello hicimos 10 corridas y seleccionamos aquel portafolio que más veces se repitió para cada uno de los escenarios a saber: diario, semanal y mensual.

Para el escenario diario el portafolio seleccionado esta conformado por:

cod_emi	emisora
15	Cemex S.A. CPO
29	GCarso A1
34	GFBBVA Bancomer B
48	ICA Soc Controlad
50	Iusacell Gpo
54	Maseca GI B
64	Soriana Organizacio B
67	Telefs de Mex A
69	Televisa Gpo CPO
74	Wal Mart de Mexico C

Para el escenario semanal, el portafolio seleccionado estaría conformado por:

cod_emi	emisora
15	Cemex S.A. CPO
34	GFBBVA Bancomer B
42	GNacional Provincia
50	Iusacell Gpo
52	KOF Coca-Cola L
54	Maseca GI B
63	Simec Grupo B
68	Telefs de Mex L
73	Vitro A
74	Wal Mart de Mexico C

Para el escenario mensual, el portafolio seleccionado estaría conformado por:

cod_emi	emisora
7	Apasco S.A.
19	Continental Grupo
42	GNacional Provincia
50	Iusacell Gpo
52	KOF Coca-Cola L
54	Maseca GI B
63	Simec Grupo B
64	Soriana Organizacio B
67	Telefs de Mex A
69	Televisa Gpo CPO

8.1.2 Resultados del Algoritmo genético usando como función de aptitud la distribución normal

Al igual que en el caso anterior se corrió 10 veces y para el escenario diario, el portafolio seleccionado esta conformado por:

cod_emi	emisora
15	Cemex S.A. CPO
19	Continental Grupo
29	GCarso A1
33	GFBanorte O
50	Iusacell Gpo
63	Simec Grupo B
67	Telefs de Mex A
68	Telefs de Mex L
69	Televisa Gpo CPO
74	Wal Mart de Mexico C

Para el escenario semanal, el portafolio seleccionado estaría formado por:

cod_emi	emisora
7	Apasco S.A.
27	Fomento Econ Mex UBD
29	GCarso A1
42	GNacional Provincia
50	Iusacell Gpo
52	KOF Coca-Cola L
54	Maseca GI B
68	Telefs de Mex L
73	Vitro A
74	Wal Mart de Mexico C

Para el escenario mensual, el portafolio seleccionado estaría formado por:

cod_emi	emisora
7	Apasco S.A.
15	Cemex S.A. CPO
33	GFBanorte O
42	GNacional Provincia
48	ICA Soc Controlad
50	Iusacell Gpo
63	Simec Grupo B
64	Soriana Organizacio B
69	Televisa Gpo CPO
73	Vitro A

8.1.3 Resultados del Algoritmo genético usando como función de aptitud el movimiento geométrico browniano

Al igual que en el caso anterior se corrió 10 veces.

Para el escenario diario, el portafolio seleccionado esta formado por:

cod_emi	emisora
7	Apasco S.A.
15	Cemex S.A. CPO
29	GCarso A1
34	GFBBVA Bancomer B
50	Iusacell Gpo
52	KOF Coca-Cola L
67	Telefs de Mex A
68	Telefs de Mex L
69	Televisa Gpo CPO
74	Wal Mart de Mexico C

Para el escenario semanal, el portafolio seleccionado estaría conformado por:

cod_emi	emisora
19	Continental Grupo
29	GCarso A1
34	GFBBVA Bancomer B
42	GNacional Provincia
48	ICA Soc Controlad
50	Iusacell Gpo
52	KOF Coca-Cola L
54	Maseca GI B
63	Simec Grupo B
73	Vitro A

Para el escenario mensual, el portafolio seleccionado estaría conformado por:

cod_emi	emisora
1	Accel S.A. B
7	Apasco S.A.
33	GFBanorte O
34	GFBBVA Bancomer B
48	ICA Soc Controlad
52	KOF Coca-Cola L
64	Soriana Organizacio B
67	Telefs de Mex A
68	Telefs de Mex L
73	Vitro A

Capítulo 9

Comparación de resultados, conclusiones e investigación posterior

9.1 Comparación de resultados

Como ya hemos visto en los capítulos precedentes, hemos encontrado cuatro portafolios que se utilizarán en tres escenarios diferentes, para intentar determinar cuál es aquel que proporcionará el mayor rendimiento en un futuro, es decir, tenemos un total de 12 portafolios.

Para determinar el rendimiento futuro, y dado que el horizonte de tiempo se tomó hasta el 31 de diciembre de 2002, encontraremos el rendimiento obtenido por cada uno de ellos durante 2003 en cada uno de sus escenarios de tiempo y luego anualizaremos este rendimiento para determinar cual es aquel que mayor utilidad nos hubiera dado durante el año 2003.

Denotaremos los portafolios de la siguiente manera:

P_{CD} : portafolio por conteo diario.

P_{CS} : portafolio por conteo semanal.

P_{CM} : portafolio por conteo mensual.

P_{GRD} : portafolio genético, aptitud rendimiento promedio diario.

P_{GRS} : portafolio genético, aptitud rendimiento promedio semanal.

P_{GRM} : portafolio genético, aptitud rendimiento promedio mensual.

P_{GND} : portafolio genético, aptitud distribución normal diario.

P_{GNS} : portafolio genético, aptitud distribución normal semanal.

P_{GNM} : portafolio genético, aptitud distribución normal mensual.

P_{GBD} : portafolio genético, aptitud movimiento browniano diario.

P_{GBS} : portafolio genético, aptitud movimiento browniano semanal.

P_{GBM} : portafolio genético, aptitud movimiento browniano mensual.

De los capítulos 7 y 8 tenemos que para el escenario diario nuestros portafolios son:

$P_{CD} = \{ \text{Apasco S.A.}, \text{Cemex S.A. CPO}, \text{GCarso A1}, \text{GFBBVA Bancomer B}, \text{ICA Soc Controlad}, \text{Maseca GI B}, \text{Soriana Organizacio B}, \text{Televisa Gpo CPO}, \text{Vitro A}, \text{Wal Mart de Mexico C} \}$

$P_{GRD} = \{ \text{Cemex S.A. Cpo}, \text{GCarso A1}, \text{GFBBVA Bancomer B}, \text{ICA Soc Controlad}, \text{Iusacell Gpo}, \text{Maseca GI B}, \text{Soriana Organizacio B}, \text{Telefs de Mex A}, \text{Televisa Gpo Cpo}, \text{Wal Mart de Mexico C} \}$

$P_{GND} = \{ \text{Cemex S.A. Cpo}, \text{Continental Grupo}, \text{GFBanorte O}, \text{Iusacell Gpo}, \text{Simec Grupo B}, \text{Telefs de Mex A}, \text{Telefs de Mex L}, \text{Televisa Gpo Cpo}, \text{Wal Mart de Mexico C} \}$

$P_{GBD} = \{ \text{Apasco S.A.}, \text{Cemex S.A. CPO}, \text{GCarso A1}, \text{GFBBVA Bancomer B}, \text{Iusacell Gpo}, \text{KOF Coca Cola L}, \text{Telefs de Mex A}, \text{Telefs de Mex L}, \text{Televisa Gpo CPO}, \text{Wal Mart de Mexico C} \}$

Para el escenario semanal nuestros portafolios son:

$P_{CS} = \{ \text{Apasco S.A.}, \text{Cemex S.A. CPO}, \text{Continental Grupo}, \text{Fomento Econ Mex UBD}, \text{GCarso A1}, \text{GFBBVA Bancomer B}, \text{ICA Soc Controlad}, \text{Soriana Organizacio B}, \text{Telefs de Mex A}, \text{Wal Mart de Mexico C} \}$

$P_{GRS} = \{ \text{Cemex S.A. CPO}, \text{GFBBVA Bancomer B}, \text{GNacional Provincia}, \text{Iusacell Gpo}, \text{KOF Coca Cola L}, \text{Maseca GI B}, \text{Simec Grupo B}, \text{Telefs de Mex L}, \text{Vitro A}, \text{Wal Mart de Mexico C} \}$

$P_{GNS} = \{ \text{Apasco S.A.}, \text{Fomento Econ Mex UBD}, \text{GCarso A1}, \text{GNacional Provincia}, \text{Iusacell Gpo}, \text{KOF Coca Cola L}, \text{Maseca GI B}, \text{Telefs de Mex A}, \text{Vitro A}, \text{Wal Mart de Mexico C} \}$

$P_{GBS} = \{ \text{Continental Grupo}, \text{GCarso A1}, \text{GFBBVA Bancomer B}, \text{GNacional Provincia}, \text{ICA Soc Controlad}, \text{Iusacell Gpo}, \text{KOF Coca Cola L}, \text{Maseca GI B}, \text{Simec Grupo B}, \text{Vitro A} \}$

Y para el escenario mensual nuestros portafolios son:

$P_{CM} = \{ \text{Apasco S.A.}, \text{Cemex S.A. CPO}, \text{Continental Grupo}, \text{Fomento Econ Mex UBD}, \text{ICA Soc Controlad}, \text{Maseca GI B}, \text{Soriana Organizacio B}, \text{Telefs de Mex A}, \text{Televisa Gpo CPO}, \text{Wal Mart de Mexico C} \}$

$P_{GRM} = \{ \text{Apasco S.A.}, \text{Continental Grupo}, \text{GNacional Provincia}, \text{Iusacell Gpo}, \text{KOF Coca Cola L}, \text{Maseca GI B}, \text{Simec Grupo B}, \text{Soriana Organizacio B}, \text{Telefs de Mex A}, \text{Televisa Gpo CPO} \}$

$P_{GNM} = \{ \text{Apasco S.A.}, \text{Cemex S.A. Cpo}, \text{GFBanorte O}, \text{GNacional Provincia}, \text{ICA Soc Controlad}, \text{Iusacell Gpo}, \text{Simec Grupo B}, \text{Soriana Organizacio B}, \text{Televisa Gpo CPO}, \text{Vitro A} \}$

$P_{GBM} = \{ \text{Accel S.A. B, Apasco S.A., GFBanorte O, GFBBVA Bancomer B, ICA Soc Controlad, KOF Coca Cola L, Soriana Organozacio B, Telefs de Mex A, Telefs de Mex L, Vitro A} \}$

Los rendimientos promedio del año 2003 para cada uno de los portafolios son los siguientes:

Escenario diario:

$$P_{CD} = 0.1692 \%$$

$$P_{GRD} = 0.1821 \%$$

$$P_{GND} = 0.2276 \%$$

$$P_{GBD} = 0.1581 \%$$

Escenario semanal:

$$P_{CS} = 0.6786\%$$

$$P_{GRS} = 0.8320 \%$$

$$P_{GNS} = 1.2092 \%$$

$$P_{GBS} = 0.9783\%$$

Escenario mensual:

$$P_{CM} = 4.0290\%$$

$$P_{GRM} = 5.1227. \%$$

$$P_{GNM} = 6.2515\%$$

$$P_{GBM} = 3.7367. \%$$

Por lo tanto, si seleccionamos un portafolio para cada uno de los escenarios tendríamos que el mejor para el horizonte diario sería:

$P_{GND} = \{ \text{Cemex S.A. Cpo, Continental Grupo, GFBanorte O, Iusacell Gpo, Simec Grupo B, Telefs de Mex A, Telefs de Mex L, Televisa Gpo Cpo, Wal Mart de Mexico C} \}$.

Para el horizonte semanal sería:

$P_{GNS} = \{\text{Apasco S.A., Fomento Econ Mex UBD, GCarso A1, GNacional Provincia, Iusacell Gpo, KOF Coca Cola L, Maseca GI B, Telef de Mex A, Vitro A, Wal Mart de Mexico C}\}$

Para el horizonte mensual sería:

$P_{GNM} = \{\text{Apasco S.A., Cemex S.A. Cpo, GFBanorte O, GNacional Provincia, ICA Soc Controlad, Iusacell Gpo, Simec Grupo B, Soriana Organizacio B, Televisa Gpo CPO, Vitro A }\}$.

Los tres portafolios seleccionados como los mejores se obtuvieron mediante el Algoritmo genético con función de aptitud basada en la distribución normal, lo cual nos indica que con esta información los modelos desarrollados usando Algoritmos genéticos con esta clase de función de aptitud son mejores para formar portafolios de inversión que los modelos basados únicamente en el conteo.

Comparemos ahora los 12 portafolios utilizando⁹¹ el rendimiento anualizado para cada uno de ellos:

$$P_{CD} = 60.91\%$$

$$P_{GRD} = 65.56\%$$

$$P_{GND} = 81.94\%$$

$$P_{GBD} = 56.92\%$$

$$P_{CS} = 35.30\%$$

$$P_{GRS} = 43.42\%$$

$$P_{GNS} = 62.88\%$$

$$P_{GBS} = 50.87\%$$

$$P_{CM} = 48.34\%$$

$$P_{GRM} = 61.47\%$$

$$P_{GNM} = 75.025\%$$

$$P_{GBM} = 44.84\%$$

⁹¹ Utilizamos una base de 360 días para la anualización. En el caso de las semanas el factor es de 360/7 y para el caso de los meses es de 360/30.

Ordenando de mayor a menor los portafolios, tendríamos la siguiente lista:

P _{GND}	81.94%
P _{GNM}	75.03%
P _{GRD}	65.56%
P _{GNS}	62.88%
P _{GRM}	61.47%
P _{CD}	60.91%
P _{GBD}	56.92%
P _{GBS}	50.87%
P _{CM}	48.34%
P _{GBM}	44.84%
P _{GRS}	43.42%
P _{CS}	35.30%

Observamos que los dos primeros resultados son los generados por el Algoritmo genético normalizado los periodos diario y mensual.

Los primeros cinco provienen de portafolios obtenido por algún Algoritmo genético, y sólo hasta la sexta posición aparece el de conteo diario, en la novena posición aparecerá el de conteo mensual y al final el de conteo semanal.

Los Algoritmos genéticos normalizados aparecen en la primera, segunda y cuarta posición permitiéndonos afirmar que son los que encuentran las combinaciones que son capaces de obtener portafolios con mayores rendimientos.

Excepto el modelo browniano, en todos los casos se encontró un orden que presenta que el mayor rendimiento se encuentra con un horizonte de tiempo diario, le sigue el mensual y al final es el semanal; en el caso del browniano se intercambian los mensuales con los semanales en la segunda y tercera posición pero se respeta el diario como el mejor.

9.2 Algoritmo voraz⁹²

Los algoritmos voraces trabajan en fases. En cada fase se toma una decisión aparentemente buena, sin considerar las consecuencias futuras. Generalmente, esto significa que se ha seleccionado como solución un óptimo local. Esta estrategia de "tome lo que pueda ahora" es el origen del nombre para esta clase de algoritmos. Cuando el algoritmo termina, tenemos la esperanza de que el óptimo local sea igual al óptimo global. Si esto pasa el algoritmo es correcto, pero si no entonces se ha llegado a una solución subóptima.

⁹² <http://www.cs.man.ac.uk/~graham/cs2022/greedy/>

En nuestro caso desarrollamos un algoritmo voraz que sólo buscaba una solución óptima después de 10 iteraciones siempre y cuando esta pareciera la correcta. Los resultados obtenidos por el algoritmo voraz fueron:

Mejor solución diaria:

VD={Televisa, Wallmart C, Vitro, GCarso A1, GFBBVA Bancomer, Soriana, Maseca, Cemex, Apasco, Telefs L}

VS={ICA, Fomento Económico, Cemex, Apasco, GCarso A1, Continental, GFBBVA Bancomer, Soriana, Telefs A, WallMart C}

VM={Vitro, Accel, GFBBVA Bancomer, Telefs L, GCarso A1, KOF Coca Cola, GNacional, SIMEC, GF Banorte, Iusacell}

Los resultados obtenidos por este algoritmo voraz fueron:

VD=0.08%

Es decir un 28.8% anualizado.

VS=0.68%

Es decir un 35.36% anualizado

VM=5.25%

Es decir un 63% anualizado.

De estos tres resultados, sólo el mensual llegaría al cuarto lugar general en rendimiento,

Mientras que los otros dos darían de los peores resultados.

9.3 Conclusiones

Todo trabajo que pretenda incluir la fusión de dos ciencias para la consecución de sus resultados corre el riesgo de quedar mal con alguna, si no es que con ambas ciencias que pretenda incluir en dicho proyecto.

Esto se vuelve todavía más complejo y dramático si se pretende fusionar a las finanzas con la computación en un intento de formar portafolios de inversión tomando de base el modelo de Markowitz.

Por un lado, los expertos en computación probablemente quedarán insatisfechos con las metodologías, técnicas, procesos y métodos de selección y aun con el software utilizado. Siempre se hablará de una opción más rápida, más económica o más elegante. Esto se agrava si quien está elaborado la investigación es, por decir lo menos, un profano en la ciencia de la computación.

Por el otro lado los financieros podrán cuestionar el rigor de los métodos de las finanzas empíricas, la econometría financiera o el cálculo estocástico utilizados en el corazón del programa. ¿Qué tal que no aplico de manera correcta? ¿Qué pasa con la confiabilidad de los algoritmos con respecto a la precisión financiera? ¿Se podrá modelar la actitud, percepción o sensaciones de un experto financiero? ¿Valdrá la pena hacer millones y millones de cálculos para encontrar resultados que un experto hubiera encontrado por sí mismo?

Ante estas interrogantes surge la tentación de cuestionar si valdrá la pena desarrollar modelos que incluyan millones de combinaciones y cientos de horas hombre y máquina.

Podríamos comenzar negando la validez y concluir por el absurdo que si ha valido la pena el esfuerzo.

Sin embargo, la conclusión y la respuesta es más sencilla, y esa proviene de los resultados. Los portafolios encontrados generan rendimientos muy por encima del mercado en el año de 2003. Con la información de diez años se encontraron conjuntos de activos que dieron rendimientos muy por encima de lo normal.

Aún incluyendo las comisiones, los rendimientos siguen siendo lo bastante atractivos para que valga la pena tomar el riesgo de a partir de un nuevo punto t_0 ejecutar una vez más los algoritmos aquí presentados y apostar por los resultados obtenidos en un futuro inmediato.

La aportación más importante de este trabajo lo constituye la construcción del algoritmo en sí. Es totalmente original y diseñado ex profeso para el mercado mexicano. A pesar de que en la literatura (como ya se ha señalado) hay varias aplicaciones a este respecto, en general se utilizan algoritmos diseñados con anterioridad ya sea para otros mercados o para otras aplicaciones.

Con todas las reservas que cualquier medio de pronóstico, predicción, simulación o inferencia pueda tener; hemos desarrollado una herramienta que ha arrojado resultados que a la luz de compararlos contra lo que realmente sucedió, nos permiten decir que hacen un buen trabajo de predicción.

Los métodos actualmente utilizados por la mayoría de las casas de bolsa hacen uso fundamentalmente de métodos tradicionales de pronósticos como promedios móviles⁹³ con una longitud de variación de 30 días y estos mecanismos obtienen muy buenos resultados siempre y cuando se siga la metodología de entrada y salida que determina el modelo. Nosotros hemos diseñado y proponemos un modelo más versátil pues utiliza el aprendizaje inteligente de las máquinas más que la intervención del ser humano.

⁹³ Por ejemplo Ixe Grupo Financiero usa una extraordinaria herramienta de pronósticos econométricos llamado el emulador, que proporciona muy buenos resultados en el mediano plazo.

Esto permitiría seleccionar un portafolio de entrada al inicio de un periodo $t=t_0$ y después obtener las utilidades en el periodo $t=t_f$ sin la necesidad de estar haciendo compras y ventas de acciones donde se perdería el monto de las comisiones de compra y venta tantas veces como se hicieran operaciones.

Por supuesto el resultado puede no ser el óptimo global, pero si tiene una alta probabilidad de ser un óptimo local y al final del día obtener rendimientos sobre la inversión inicial, si bien puede no ser la mayor utilidad, si parece obtener una muy buena utilidad.

Por los resultados obtenidos podemos concluir que con esta información los Algoritmos genéticos generan combinaciones que forman portafolios más productivos en comparación que los generados por métodos de conteo, que dan muy buenos resultados comparados contra el rendimiento promedio del mercado y que por su velocidad de compilación en comparación con los métodos de conteo y sobre todo de los enumerativos bien vale la pena probar algunos de ellos.

Que en una escala de mayor a menor, podemos considerar a los portafolios del escenario diario como más productivos, seguidos de los portafolios mensuales y hasta el final los semanales; contradiciendo esto a la creencia afirmación de que los mejores rendimientos son los obtenidos por periodos semanales.

9.4 Investigación posterior

Como todo modelo éste es perfectamente perfectible, sobre todo a la luz de la gran cantidad de métodos de selección que tienen los Algoritmos genéticos. Una línea de trabajo es utilizar modelos de selección tales como el recocido simulado, modelos de Vasconcelos, de Nietzsche o la Ecuación de Woltzmann, para manejar patrones de temperatura más consistentes.

Otra área se da en la metodología de la selección de la población y muestra así como en el número de acciones, hardware más poderoso permitiría hacer mejor trabajo; y finalmente el trabajo de adecuar estos modelos a tiempo real, con la información del minuto de la BMV.

Por supuesto la consistencia y robustez del modelo sólo la daría el tiempo y una serie de repeticiones y simulaciones con diferentes espectros de tiempo, así que un área interesante sería la de probar otros horizontes de tiempo, otras longitudes de datos y usar alguna otra técnica de agrupación, como pudiera parecer el agrupamiento natural por sectores.

Dos cosas si se deberían incluir en una investigación posterior.

La primera de ellas será la más simple. No es posible formar portafolios de inversión sin considerar un aspecto duro de la vida real: las comisiones. A pesar de que para el modelado simple siempre consideramos de base un mercado libre

de ellas, la realidad es que esto no es así, por lo que habría que incluirlas como parte de un trabajo posterior.

El otro asunto no es trivial. Consiste en la aplicación dura de la metodología de Markowitz e incluir una ponderación diferente para cada uno de los componentes de los portafolios, esta será una tarea más retadora ya que ataca al corazón mismo de la construcción del algoritmo. Algunos estudios ya lo esbozan y consideramos que con algo de esfuerzo, dedicación y entrega se podrá considerar este aspecto que aquí se ha pasado por alto.

ANEXO 1

CORRIDAS COMPLETAS DE LOS ALGORITMOS GENÉTICOS

CORRIDAS AG BROWNIANOS DIARIO

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	1, 7, 27, 34, 50, 52, 54, 67, 69, 73	49
2	1, 7, 15, 33, 42, 48, 50, 54, 67, 69	106
3	29, 33, 34, 42, 48, 50, 63, 64, 73, 74	48
4	1, 27, 29, 42, 50, 52, 54, 63, 64, 74	56
5	27, 29, 33, 42, 48, 50, 52, 63, 67, 73	78
6	7, 19, 27, 34, 42, 48, 50, 67, 68, 73	111
7	7, 27, 29, 33, 34, 50, 52, 54, 67, 69	64
8	7, 15, 29, 34, 50, 52, 67, 68, 69, 74	149
9	1, 29, 33, 34, 50, 54, 63, 68, 73, 74	56
10	1, 15, 19, 27, 29, 48, 50, 69, 73, 74	104

CORRIDAS AG NORMALIZADO DIARIOS

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	29, 34, 50, 52, 63, 67, 68, 69, 73, 74	5
2	19, 27, 29, 34, 42, 50, 52, 63, 67, 69	4
3	15, 19, 29, 33, 50, 63, 67, 68, 69, 74	11
4	7, 27, 29, 33, 42, 50, 52, 67, 69, 73	9
5	19, 29, 34, 42, 50, 52, 54, 64, 67, 68	3
6	1, 7, 15, 19, 27, 34, 42, 48, 50, 73	4
7	7, 27, 29, 33, 50, 54, 63, 64, 67, 68	7
8	7, 19, 27, 48, 50, 52, 54, 63, 73, 74	11
9	15, 19, 33, 42, 48, 50, 54, 63, 68, 73	3
10	1, 7, 15, 19, 29, 34, 50, 67, 68, 74	3

CORRIDAS AG RENDIMIENTO DIARIO

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	1, 15, 27, 33, 34, 50, 64, 68, 69, 73	54
2	15, 29, 34, 48, 50, 54, 64, 67, 68, 74	120
3	7, 15, 19, 27, 42, 48, 50, 52, 54, 68	108
4	1, 15, 33, 50, 52, 54, 63, 64, 67, 68	106
5	15, 29, 34, 42, 50, 52, 54, 63, 69, 73	88
6	15, 27, 29, 33, 50, 63, 67, 68, 69, 73	101
7	33, 34, 42, 48, 50, 54, 64, 68, 69, 74	106
8	1, 7, 27, 29, 33, 34, 48, 50, 63, 69	100
9	7, 29, 34, 48, 50, 54, 63, 69, 73, 74	79
10	1, 7, 33, 42, 50, 63, 67, 68, 69, 74	76

CORRIDAS AG BROWNIANO SEMANAL

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	19, 29, 34, 42, 48, 50, 52, 54, 63, 73	99
2	1, 27, 33, 34, 42, 50, 54, 67, 68, 73	64
3	1, 19, 27, 50, 54, 64, 67, 68, 69, 74	75
4	1, 7, 19, 42, 48, 50, 54, 64, 68, 73	99
5	1, 7, 27, 29, 33, 50, 52, 64, 69, 74	77
6	7, 34, 42, 48, 50, 54, 63, 64, 67, 73	51
7	1, 7, 19, 50, 54, 63, 64, 69, 73, 74	98
8	27, 29, 34, 42, 48, 50, 64, 67, 69, 74	55
9	1, 19, 27, 29, 42, 48, 50, 54, 64, 73	72
10	19, 33, 34, 48, 50, 52, 54, 63, 64, 67	53

CORRIDAS AG NORMALIZADO SEMANAL

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	1, 7, 19, 29, 33, 42, 50, 63, 67, 73	9
2	7, 15, 27, 33, 42, 50, 63, 68, 73, 74	7
3	15, 27, 33, 50, 52, 54, 63, 67, 69, 74	3
4	7, 27, 29, 42, 50, 52, 54, 68, 73, 74	16
5	7, 27, 33, 34, 50, 52, 67, 69, 73, 74	11
6	1, 7, 19, 27, 29, 33, 50, 52, 63, 68	8
7	7, 19, 27, 29, 34, 42, 50, 54, 67, 73	9
8	7, 19, 27, 29, 34, 42, 50, 54, 63, 68	4
9	7, 29, 33, 34, 42, 50, 54, 64, 69, 73	3
10	15, 27, 34, 48, 50, 54, 67, 68, 74, 73	7

CORRIDAS AG RENDIMIENTO SEMANAL

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	33, 34, 42, 48, 50, 63, 64, 67, 68, 69	37
2	7, 29, 34, 42, 50, 52, 54, 64, 67, 69	77
3	1, 15, 19, 27, 33, 34, 48, 50, 63, 73	58
4	1, 7, 15, 29, 33, 42, 50, 54, 64, 67	59
5	7, 19, 27, 34, 48, 50, 54, 63, 64, 74	29
6	15, 34, 42, 50, 52, 54, 63, 68, 73, 74	170
7	29, 33, 48, 50, 54, 64, 67, 68, 69, 74	51
8	1, 19, 33, 34, 48, 50, 52, 54, 69, 73	71
9	27, 29, 33, 34, 48, 50, 63, 64, 68, 69	46
10	7, 19, 33, 34, 48, 50, 63, 67, 68, 73	125

CORRIDAS AG BROWNIANOS MENSUAL

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	1, 7, 33, 34, 48, 52, 64, 67, 68, 73	73
2	27, 29, 33, 34, 42, 48, 50, 54, 67, 68	47
3	29, 33, 42, 50, 52, 54, 63, 68, 69, 73	73
4	27, 33, 34, 50, 52, 63, 64, 68, 69, 73	9
5	7, 27, 33, 34, 42, 50, 63, 67, 73, 74	19
6	7, 19, 27, 33, 48, 67, 68, 68, 73, 74	31
7	1, 27, 29, 33, 42, 48, 50, 52, 68, 73	26
8	15, 19, 27, 34, 48, 52, 54, 67, 68, 73	5
9	19, 29, 33, 34, 48, 50, 52, 63, 69, 74	27
10	7, 19, 42, 50, 52, 54, 63, 64, 67, 69	43

CORRIDAS AG NORMALIZADO MENSUAL

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	15, 27, 29, 33, 42, 48, 50, 52, 54, 67	4
2	7, 15, 19, 29, 33, 34, 50, 52, 63, 73	2
3	1, 7, 15, 19, 29, 33, 34, 50, 67, 69	8
4	27, 33, 34, 50, 52, 63, 64, 68, 69, 73	9
5	7, 27, 33, 34, 42, 50, 63, 67, 73, 74	19
6	1, 7, 15, 19, 33, 48, 50, 54, 63, 68	6
7	1, 7, 15, 27, 29, 33, 42, 50, 63, 68	4
8	15, 19, 27, 34, 48, 52, 54, 67, 68, 73	5
9	15, 34, 48, 50, 52, 54, 63, 67, 69, 73	12
10	7, 15, 33, 42, 48, 50, 63, 64, 69, 73	29

CORRIDAS AG RENDIMIENTO MENSUAL

CORRIDA	MEJOR PORTAFOLIO	FRECUENCIA
1	1, 7, 15, 33, 34, 42, 50, 54, 63, 68	41
2	15, 19, 27, 33, 34, 50, 54, 63, 64, 73	28
3	7, 19, 27, 34, 48, 50, 52, 54, 67, 68	37
4	19, 29, 42, 48, 50, 52, 63, 64, 68, 69	25
5	1, 19, 42, 48, 50, 54, 64, 69, 73, 74	29
6	7, 19, 27, 33, 48, 67, 68, 69, 73, 74	31
7	1, 27, 29, 33, 42, 48, 50, 52, 67, 74	26
8	1, 15, 27, 29, 33, 42, 48, 50, 67, 74	42
9	19, 29, 33, 34, 48, 50, 52, 63, 69, 74	27
10	7, 19, 42, 50, 52, 54, 63, 64, 67, 69	43

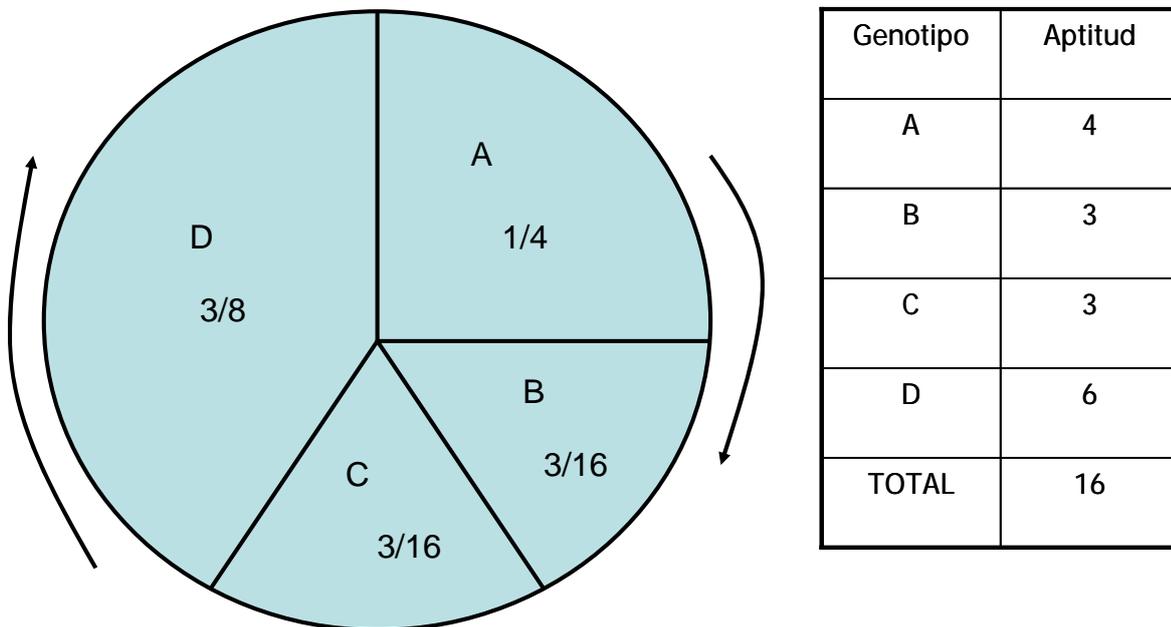
ANEXO 2

EL MÉTODO DE LA RULETA⁹⁴

El Algoritmo genético simple incluye en uno de sus pasos la selección de dos individuos de la población actual con una probabilidad proporcional a su aptitud, este tipo de selección proporcional es llamada también el método de la ruleta (*roulette wheel selection*) por lo siguiente: supóngase que se suman las aptitudes de todos los individuos completos de la población y esta suma es considerada el 100% de una circunferencia. Luego, cada individuo se le asigna el trozo que le corresponde de esta según la aportación que haga a la suma de las aptitudes. Es decir, si la aptitud de un individuo es x_i entonces le corresponde un segmento de circunferencia dado por la siguiente regla de tres:

$$s = 2\pi \frac{x_i}{\sum_j x_j}$$

¿Qué ocurrirá entonces si se considera ésta como una ruleta y se coloca una lengüeta que roce el borde de ella?



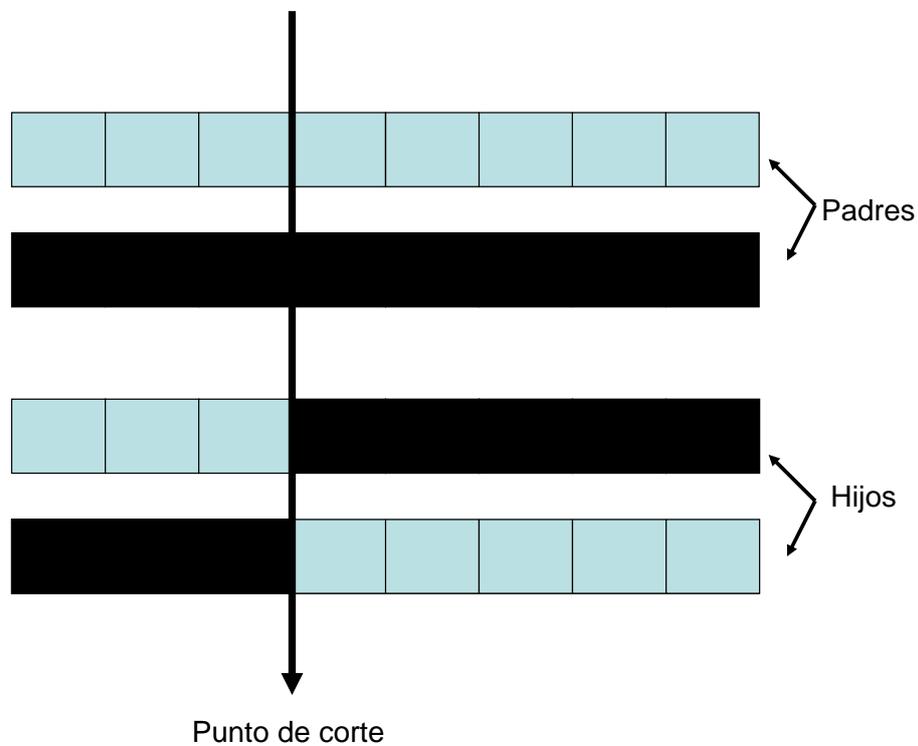
La probabilidad de que dicha lengüeta quede en el arco correspondiente al individuo de calificación x_i , cuando la rueda se detenga tras realizar algunos giros, es:

⁹⁴ Tomado de Kuri, A. y Galaviz, J. *Algoritmos Genéticos* IPN, UNAM, FCE

$$p(i) = \frac{x_i}{\sum_j x_j}$$

la que es proporcional a la calificación (x_i) del individuo.

Aun queda por aclarar como se mezclan los códigos de dos individuos para formar dos híbridos. En general hay muchas maneras de hacerlo. Una de las más populares es el cruzamiento de un punto (*1-point crossover*). En este tipo de cruzamiento, dados dos individuos se elige aleatoriamente un punto de corte entre dos bits cualesquiera del cromosoma. Esto define segmentos izquierdos y derechos de cada genotipo. Se procede entonces a intercambiar los segmentos izquierdos (o los derechos indistintamente) de cada individuo. De esto resultan dos híbridos:



ANEXO 4

LOS ALGORITMOS

Diseño del algoritmo de conteo

El objetivo principal de este Algoritmo es que sea capaz de interpretar una Base de Datos de Access, normalizarla y mostrar el número de veces que combinaciones de diez acciones cumplen con el requisito de sobrepasar el promedio de todas las acciones en un período de tiempo dado con el promedio del rendimiento de las diez acciones que constituyen la combinación.

Una opción era realizar todo el trabajo de manera exhaustiva, es decir, realizar una a una todas las posibles combinaciones sin algún criterio de selección predefinido, pero después de haber simulado con una base de datos de prueba con seis acciones, llegamos a la conclusión que sería más sencillo y práctico poder distribuir el trabajo de las combinaciones a lo largo de su recorrido por las filas (los períodos de tiempo) para poder contar con la opción del paralelismo y finalmente unir todos los resultados para poder realizar las observaciones necesarias.

A fin de poder cumplir las metas propuestas para la aplicación se desarrollaron diferentes Algoritmos para realizar las siguientes acciones:

1. Sacar las combinaciones apropiada y efectivamente
2. Normalizar la Base de Datos para realizar las operaciones necesarias
3. Crear y hacer uso de tablas auxiliares que permitieran separar el trabajo en partes
4. Borrar la Base de Datos y las tablas que se usarían para recomenzar el procedimiento

La base del proyecto es el Algoritmo que permitirá sacar las combinaciones para después ir las probando. Este Algoritmo debía tener la capacidad de poder ser interrumpido y reiniciar en caso de ocurrir una falla o simplemente para posponer los cálculos. Para cumplir con estos requisitos era necesario que el Algoritmo no fuese recursivo sino determinístico.

El Algoritmo resultante, presentado en pseudo-código, es el siguiente:

Parámetros:

- **FilaInicioDada**: Fila cargada de la Base de Datos de donde iniciará a correr el Algoritmo
- **InicioPrimeraAcción**: Número de Inicio de la Primera Acción donde iniciará
- **InicioSegundaAcción**: Número de Inicio de la Segunda Acción donde iniciará
- **InicioTerceraAcción**: Número de Inicio de la Tercera Acción donde iniciará
- **InicioCuartaAcción**: Número de Inicio de la Cuarta Acción donde iniciará
- **FilaFinDada**: Número de Fila donde el Algoritmo parará

- **FinPrimeraAcción:** Número de la Primera Acción donde el Algoritmo parará
 - **FinSegundaAcción:** Número de la Segunda Acción donde el Algoritmo parará
 - **FinTerceraAcción:** Número de la Tercera Acción donde el Algoritmo parará
 - **FinCuartaAcción:** Número de la Cuarta Acción donde el Algoritmo parará
 - **Bandera:** Bandera que indicará si cargar la Fila de Inicio, y los inicios de los Acciones
1. Desde Fila = FilalInicioDada Hasta NúmeroDeFilas:
 2. Promedio = ElPromedioDeLaFila
 3. Desde i = InicioPrimeraAcción Hasta NúmeroDeAcciones - 5
 4. Si Bandera = Falso Entonces InicioSegundaAcción = i + 1
 5. Desde j = InicioSegundaAcción Hasta NúmeroDeAcciones - 4
 6. Si Bandera = Falso Entonces InicioTerceraAcción = j + 1
 7. Desde k = InicioTerceraAcción Hasta NúmeroDeAcciones - 3
 8. Si Bandera = Falso Entonces InicioCuartaAcción = k + 1
 9. Desde l = InicioCuartaAcción Hasta NúmeroDeAcciones - 2
 10. Si (Fila = FilaFinDada y i = FinPrimeraAcción y j = FinSegundaAcción y k = FinTerceraAcción y l > FinCuartaAcción) ó (Fila = FilaFinDada y i = FinPrimeraAcción y j = FinSegundaAcción y k > FinTerceraAcción) ó (Fila = FilaFinDada y i = FinPrimeraAcción y j > FinSegundaAcción) ó (Fila = FilaFinDada y i > FinPrimeraAcción) ó (Fila > FilaFinDada) Entonces
 11. Finalizar
 12. Fin Si
 13. Total = 0
 14. Si Hay Rendimiento Reportado En i, j, k y l Entonces
 15. Total = PromedioDe(i, j, k y l)
 16. Si Total > Promedio Entonces
 17. Si La Combinación de i, j, k y l No Existe En aux Entonces
 18. Agregar La Combinación A Combinaciones Con 1
 19. SiNo
 20. Sumar 1 A Las Veces De La Combinación
 21. Fin Si
 22. Fin Si
 23. SiNo
 24. Si La Combinación de i, j, k y l No Existe En aux Entonces
 25. Agregar La Combinación A Combinaciones Con 0
 26. Fin Si
 27. Fin Si
 28. Actualizar iFila, i, j, k y l En aux
 29. Regresar El Control A La Aplicación
 30. Si BanderaDeContinuar = Falso Entonces Finalizar
 31. Fin Desde
 32. Bandera = Falso
 33. Fin Desde
 34. Fin Desde
 35. Fin Desde
 36. InicioPrimeraAcción = 1
 37. InicioSegundaAcción = 2
 38. InicioTerceraAcción = 3
 39. InicioCuartaAcción = 4
 40. Actualizar iFila, i, j, k y l En aux
 41. Fin Desde

La filosofía de este Algoritmo está basada en un experimento conteo realizado a mano con un conjunto de seis números y que sirvió de base para probar la base de datos a la que se hizo mención antes. Este experimento se reproduce a continuación para entender el concepto del

Algoritmo, el conjunto de números será del 1 al 6; sabemos que en este caso el resultado es 15 combinaciones diferentes⁹⁷:

```

1, 2, 3, 4, 5, 6
(1) (2, 3, 4, 5, 6)
    (1, 2) (3, 4, 5, 6)
        (1, 2, 3) (4, 5, 6)
            (1, 2, 3, 4) (5, 6)
                (1, 2, 3, 5) (6)
                (1, 2, 3, 6) ()
            (1, 2, 4) (5, 6)
                (1, 2, 4, 5) (6)
                (1, 2, 4, 6) ()
            (1, 2, 5) (6)
        (1, 3) (4, 5, 6)
            (1, 3, 4) (5, 6)
                (1, 3, 4, 5) (6)
                (1, 3, 4, 6) ()
            (1, 3, 5) (6)
        (1, 4) (5, 6)
    (2) (3, 4, 5, 6)
        (2, 3) (4, 5, 6)
            (2, 3, 4) (5, 6)
                (2, 3, 4, 5) (6)
                (2, 3, 4, 6) ()
            (2, 3, 5) (6)
        (2, 4) (5, 6)
    (3) (4, 5, 6)

```

Este experimento puede mostrarnos varias reglas que sólo podremos apreciar si prestamos atención con detenimiento. Lo primero que se puede observar es la profundidad máxima a la que se puede llegar. Es decir que se necesitan cuatro ciclos. La segunda cosa que podemos observar es que cada vez se van reduciendo más las posibilidades, es decir que se resta un elemento por cada vez que se entra a un nuevo ciclo y también en cada iteración del mismo ciclo. Esta segunda observación también nos permite ver que para el **ciclo de profundidad 1** (El primer ciclo) se requerirán de **i hasta n - 3** iteraciones (de 1 a 6 - 3 = 3), para el **ciclo de profundidad 2** (El segundo ciclo) se requerirán de **i + 1 hasta n - 2** iteraciones (de 2 a 6 - 2 = 4), para el **ciclo de profundidad 3** (El tercer ciclo) se requerirán de **i + 2 hasta n - 1** iteraciones (de 3 a 6 - 1 = 5) y para el **ciclo de profundidad 4** (El cuarto ciclo) se requerirán de **i + 3 hasta n - 0** iteraciones (de 4 a 6 - 0 = 6). Finalmente podemos observar que el ciclo de nivel i inicia en el número final del nivel i - 1 (Ejemplo: (1, 3) hará que inicie en (4, 5, 6), mientras que (1, 2) hará que inicie en (3, 4, 5, 6).)

Siguiendo estas observaciones simplemente se adaptaron en un Algoritmo sencillo que se probó en un programa de Visual Basic para comprobar si las combinaciones resultantes eran correctas. Tras haber hecho numerosas pruebas se concluyó que el Algoritmo trabajaba satisfactoriamente obteniendo los resultados esperados. Así que se le

⁹⁷ El experimento completo para 40 acciones nos da como resultado 847 660,528 combinaciones diferentes.

agregaron otras capacidades, como la de permitir darle los parámetros de inicio al Algoritmo y parámetros de fin (La condición larga simplemente checa si los parámetros de fin se han cumplido para así finalizar la ejecución del Algoritmo) y una Bandera para saber si cargará el estado dado o no. También se verificó dónde realizar las actualizaciones de la Base de Datos (ya que de ahí se sacarán los parámetros de inicio en caso de haber sido interrumpido) y se probó el Algoritmo en varias ocasiones, se depuró y se adaptó hasta llegar a la versión actual, la cuál dio resultados satisfactorios de acuerdo a lo esperado.

A continuación se mostrará la composición de la Base de Datos **consolid.mdb**, la cual contiene originalmente una sola tabla: **consolid**. Sin embargo, se le adicionan otras tablas para poder trabajar apropiadamente:

- **aux**: Tabla que contiene la Fila en la que se trabaja así como el número de índice de las cuatro acciones actuales. En el primer registro cuenta con el número actual, mientras que en el segundo con los índices donde debiera terminar.
- **Combinaciones**: Tabla que contiene todas las combinaciones posibles así como las veces que ha cumplido con el requisito establecido anteriormente y las fechas (períodos de tiempo) en que el requisito ha sido satisfecho.
- **CombinaFecha**: Tabla diseñada para contener el id de la Combinación y de la Fecha pensado para disminuir el espacio ocupado por la tabla de combinaciones.
- **Consolidados**: Tabla que contiene los datos de la tabla **consolid** una vez que fueron normalizados. Cada campo cuenta con la fecha (período de tiempo) y el comportamiento de cada una de las acciones en ese período de tiempo (es decir, su rendimiento).
- **Fechas**: Tabla que contiene todos los períodos de tiempo a tratar y su id para ser usado en la tabla **CombinaFecha**.
- **Promedios**: Tabla que contiene el promedio de todos los rendimientos de cada período de tiempo de la tabla **Consolidados**.

La composición de las tablas a detalle, donde se muestran los campos de cada tabla y debajo de ellos el tipo de dato que almacenan, se muestra a continuación:

Aux					
Id	iFila	i	J	k	l
Integer	Long	long	long	long	long

Combinaciones		
Nombre	Veces	Fecha
Text	Long	Text

CombinaFecha	
IdCombinacion	IdFecha
Long	Long

Consolidados					
Fecha	1	2	...	57	58
Date/Time	Text	Text	...	Text	Text

Fechas	
Id	Fecha
Long	Date/Time

Promedios	
Promedio	
Text	

Consolid		
FECHA	RENDIMIE	COD_EMI
Date/Time	Double	Double

La aplicación cuenta con más Algoritmos, que lo único que hacen es el manejo de Bases de Datos:

- Cargar la tabla consolid en la tabla de la interfaz de usuario a manera que se pueda apreciar por fecha el rendimiento de cada acción, es decir que traduce consolid a Consolidados en la forma principal. Una vez hecho esto, se ejecuta una sentencia SQL que guarda la Tabla Consolidados dentro de la Base de Datos **consolid.mdb**.
 - Borrar las tablas de la Base de Datos **consolid.mdb** que fueron seleccionadas en un diálogo que se muestra al presionar el botón "*Borrar*"
 - Obtener los índices de la Fila, la Primera Acción, la Segunda Acción, la Tercera Acción, la Cuarta Acción y hasta la décima Acción, tanto de inicio como de fin. Estos son obtenidos mediante un diálogo que se muestra al presionar el botón "*Opciones*". El diálogo en cuestión muestra unas barras que permiten deslizarse hasta alcanzar el número que se desea obtener (Es decir que dan un rango de números entre los cuales escoger). Finalmente, este diálogo cuenta con código que no permite tener valores inválidos en los índices, ya que el índice de la Segunda Acción no puede ser menor que el de la Primera, el índice de la Tercera no puede ser menor que el de la Segunda y el índice de la Cuarta no puede ser menor que el de la Tercera.
 - Desplegar el resultado de las combinaciones (con fecha si se desea) del número dado de acciones (Por ejemplo, los primeros 20)

Sin embargo, estos Algoritmos no son de gran interés para el objetivo principal. Por lo tanto no se explicarán a detalle. Todo lo explicado anteriormente es simplemente un extra que ayuda al manejo del Algoritmo que se mostró antes. De cualquier forma, se muestran a continuación las interfaces que maneja el programa. Cabe señalar que la

aplicación cuenta con una ventana que muestra el avance de los procesos largos y permite enviar a la barra de estado un icono representativo al minimizar la aplicación. Este icono muestra el avance brevemente si se deja el cursor unos breves momentos sobre él.

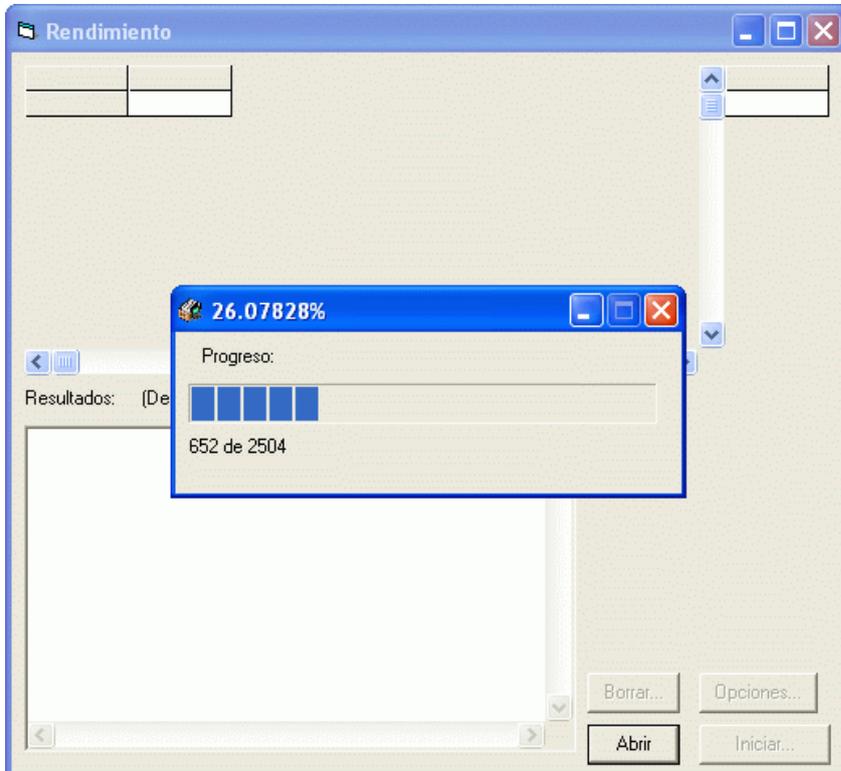


Imagen 17: El proceso de cargado de la Tabla consolid desde la Base de Datos consolid.mdb

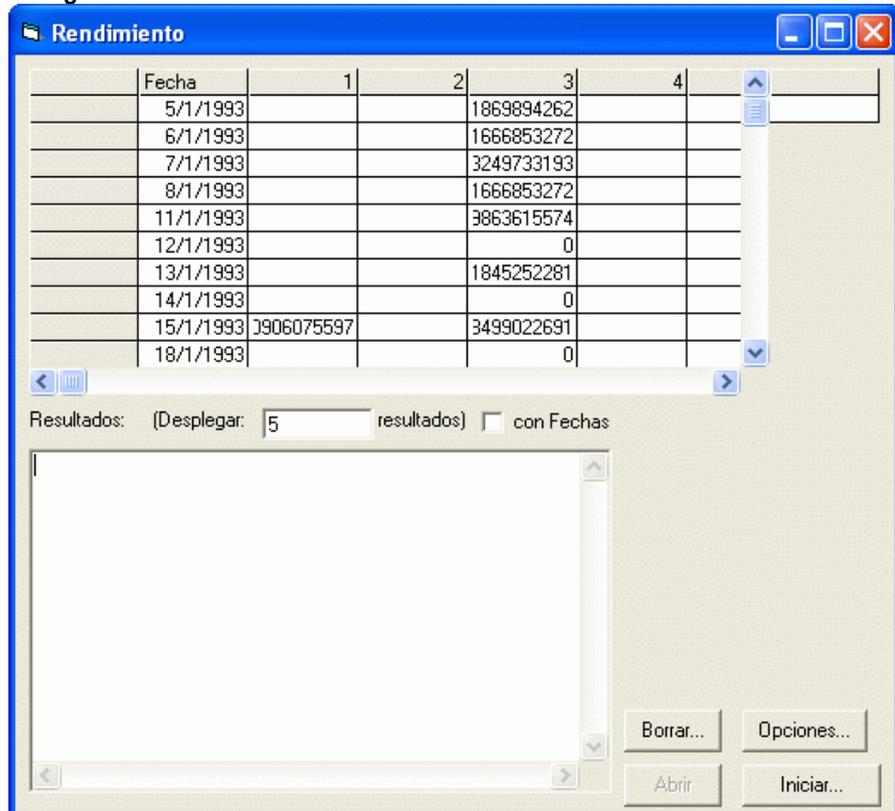


Imagen 18: La tabla de Consolidados, es decir, la tabla de consolid tras haber sido normalizada. Se pueden apreciar las opciones de Borrar, Opciones e Iniciar, así como la de desplegar cierto número de resultados con las fechas correspondientes.

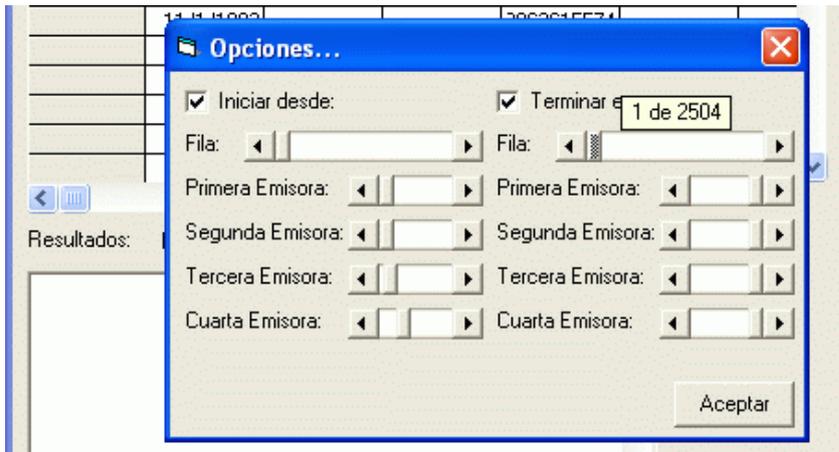


Imagen 19: Se muestra el diálogo de opciones que permite dar los parámetros de inicio al Algoritmo: Donde iniciará el Algoritmo (Fila y Acciones) y dónde terminará el Algoritmo (Fila y Acciones)

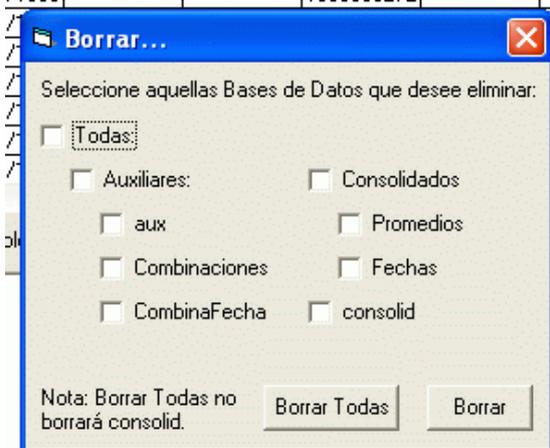


Imagen 20: Se muestra el diálogo de borrar que permite seleccionar las tablas a borrar de la Base de Datos para así evitar tener que entrar a la Base de Datos y hacer los movimientos desde el interior. También tuvo su función al hacer depuraciones y correcciones al Algoritmo y al comportamiento general de la aplicación.

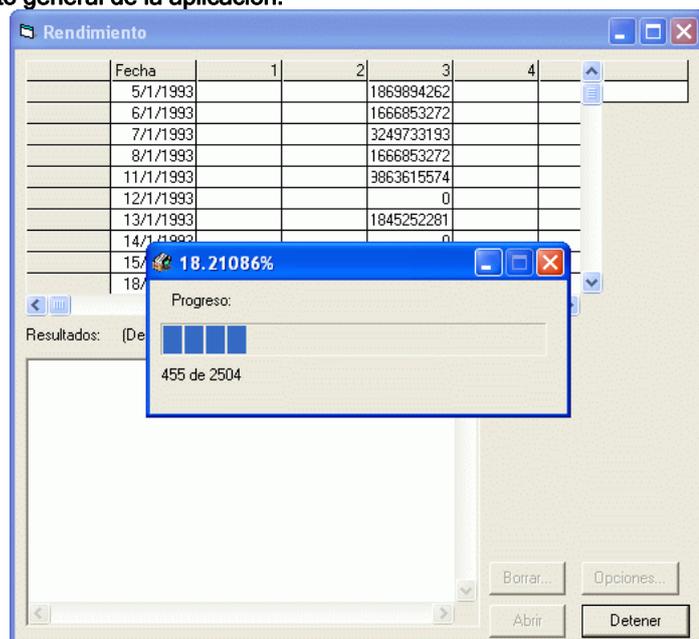


Imagen 21: Se muestra el progreso del cálculo de los promedios o el cargado de los promedios en la tabla a la derecha de Consolidados. Posteriormente se iniciará el proceso de cálculo de las combinaciones y se mostrará en el cuadro de texto de la forma.

El algoritmo genético

El Algoritmo genético básico estará conformado por las siguientes partes:

1. **Inicio:** Generar una población al azar o, como en nuestro caso, tomar una población dada de n cromosomas que podrían ser posibles soluciones al problema presentado
2. **Aptitud:** Se evalúa la aptitud de cada cromosoma en la población
3. **Nueva Población:** Se crea una nueva población siguiendo los pasos que se detallarán a continuación hasta que se tiene una nueva población:
 - a. **Selección:** Se seleccionan dos cromosomas padres de una población de acuerdo a su aptitud, mientras mejor aptitud tengan, mayor es la probabilidad de ser seleccionados.
 - b. **Cruza:** Dada una probabilidad de cruce se intercambia la información genética de los dos padres para formar un par de hijos, en caso que la probabilidad no permita el cruce, los hijos serán una copia de los padres.
 - c. **Mutación:** Dada una probabilidad de mutación se altera al azar un alelo del cromosoma.
 - d. **Aceptación:** Los hijos toman su lugar en la nueva población
4. **Reemplazo:** Usar la nueva población generada para la próxima iteración del Algoritmo
5. **Prueba:** Si la condición de paro es satisfecha entonces se para la corrida y se arroja la mejor solución de la población actual.
6. **Ciclo:** Se realiza todo lo anterior comenzando desde el paso número 2.

Una vez que ya hemos conocido el procedimiento que nos permitirá desarrollar nuestro Algoritmo genético, definiremos las bases de nuestro diseño para así continuar con la programación y la justificación de la misma.

Lo primero a tener en cuenta en el diseño del modelo para el Algoritmo genético es contar con un **Ambiente Computacional**, el cuál será un recipiente de la **Población** y otros datos misceláneos que permitirán la corrida de la *Selección Natural*, la *Reproducción* (Cruza) y la *Muerte*. La **Población** también debe ser considerada, la cual podría ser simplemente un arreglo de los **Portafolios** (*Pobladores*, *Cromosomas*) y finalmente los **Portafolios** que usaremos, que en sí son los *pobladores* o bien *cromosomas*.

La mayoría de Algoritmos genéticos utilizan la notación biológica⁹⁸ explicada previamente para su implementación y su programación. Esto es gracias al hecho de que existen al menos dos escuelas de codificación⁹⁹: la **binaria**, que fue iniciada por Holland en su trabajo, y la de **permutaciones**. La primera es un tanto purista y sostiene que el **genotipo** tiene que ser codificado por una

⁹⁸ Genoma, cromosoma, genotipo, fenotipo, etc.

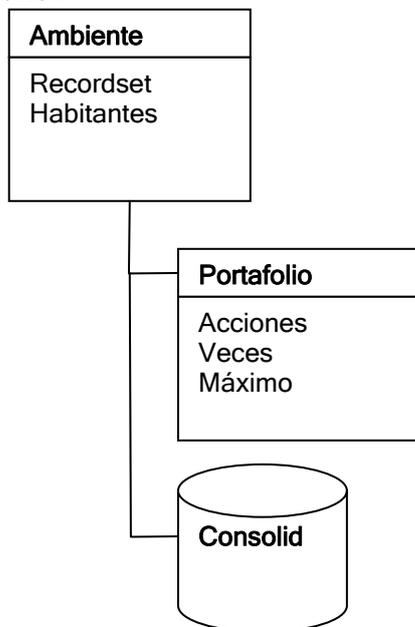
⁹⁹ Existen más escuelas además de las mencionadas aquí como son entre otras la real y la de árbol.

cadena binaria mientras que la segunda sostiene que se debe adaptar la codificación a la necesidad de los problemas¹⁰⁰.

Debido a las características del problema motivo de esta tesis se utilizó una codificación de cromosomas usando permutaciones de enteros.

Debido a que nuestro problema tiene una restricción (Los portafolios deben ser diferentes entre sí) usaremos la segunda técnica de codificación, la de permutaciones. Por ello llamaremos a los pobladores por el nombre de **Portafolio** ("Portafolio Financiero") y a su **genotipo** como "*Acciones*". Otro detalle a considerar es que tenemos todas las "soluciones" posibles dentro de una base de datos, es decir que podemos conocer las características de un **Portafolio** (su **fenotipo**) dado un **genotipo**, por lo que no es necesario almacenarlo en el **Portafolio**, ya que puede ser accedido desde la Base de Datos que tenemos.

Después de estas aclaraciones, observemos entonces cómo quedaría nuestro diseño:



Ambiente: El ambiente en el que trabajaremos, contiene a los *Habitantes* (**Portafolios**) y el *Recordset* (Base de Datos, **Consolid**) a utilizar que contiene todos los datos de los **Portafolios**. El ambiente contará con los siguientes procedimientos / funciones:

- a) **Crear Ambiente:** Este procedimiento recibirá la Base de Datos a usar, los portafolios a utilizar (el número de los portafolios. Por ejemplo: 1, 2, 3,... 58), y un valor booleano (Verdadero o Falso) que determinará si se llenará la población de una forma uniforme o mediante el uso de un random.
- b) **Llenar Tradicional:** Este procedimiento llenará la población con habitantes uniformes. Es decir que dado el portafolios del ejemplo anterior generará pobladores de la siguiente manera: (1, 2, 3, 4,..., 10), (11, 12, 13, 14,..., 20), (21, 22, 23,24,...,30),..., (31, 32, 33,..., ¿?). Los pobladores restantes, en caso de que falten, los obtendrá haciendo un random selectivo, es decir que si en el caso que vemos hará un random del 1 hasta el 40 para evitar repetir valores y de esta manera cumplir con la condicionante de no tener portafolios con valores repetidos.
- c) **Llenar Random:** Este procedimiento llenará la población haciendo uso del random selectivo mencionado anteriormente. Es decir, comenzará con toda la población, del 1 al 40 y luego quitará de la población el

¹⁰⁰ Las bases del teorema de los esquemas está fundamentado en el hecho de que la codificación es binaria utilizada en el Algoritmo genético simple de Holland.

- número que seleccionó, para hacer el random selectivo a esta nueva "población de números" para así llenar todos los habitantes.
- d) **Selección Natural:** Esta función recibirá como parámetro el número de veces que iterará el Algoritmo genético y para cada iteración llamará al procedimiento de **Reproducción**, al procedimiento de **Muerte** y se moverá al siguiente registro de la Base de Datos. Finalmente arrojará el resultado de aplicar la función **El Popular**.
 - e) **Reproducción:** Este procedimiento realiza una repetición de buscar dos padres mediante el método de selección de la **Ruleta**, el más usado para los Algoritmos genéticos. Se estableció una probabilidad de cruce de 0.65 entonces ocurrirá una cruce y los hijos de los padres se agregarán a la población (**Habitantes**). Estas repeticiones ocurren $n/2$ veces, donde n es el número de pobladores.
 - f) **Muerte:** Este procedimiento manda a **ordenar** a los habitantes por sus aptitudes de menor a mayor y se deshace de los primeros m habitantes para así mantener el número de la población constante. Por ejemplo, el número de pobladores originales es de 40, ocurre el procedimiento de Reproducción y el nuevo número de pobladores es de 64, entonces muerte ordena a los 64 pobladores y se deshará de los primeros 24 habitantes para así mantener el número de pobladores en 40.
 - g) **El Popular:** Esta función se mueve por todos los habitantes y regresa aquel que tenga las mayores veces de haber nacido en las poblaciones históricas. Ya que este poblador, en teoría, es al que tiende a evolucionar el Algoritmo cuando el límite es infinito.
 - h) **Ruleta**¹⁰¹: Existen varios métodos de de selección, como el Torneo, el de Rangos y el de elitismo, pero debido a que este método es el más popular será el que utilizaremos en esta implementación. Este método actúa como una ruleta trucada ya que se hace una sumatoria de todas las aptitudes de todos los habitantes y entonces toma un poblador random y desde ese punto comienza a moverse por los demás pobladores; por ello es conveniente usar un arreglo de pobladores, que en nuestro caso es **Habitantes**, sumando la aptitud de cada uno de ellos hasta que se llegue a la sumatoria de todas las aptitudes o se sobrepase de esa cantidad. Esto da una gran ventaja para escoger aquellos pobladores con una mejor Aptitud debido a que su probabilidad de ser seleccionados aumenta. La función arroja el número en el arreglo de pobladores en el que se detuvo la ruleta.

Algoritmo genético con función de aptitud promedio

Portafolio: El cromosoma que habita dentro del **Ambiente**. Contiene las diez **Acciones** que trabajan como su **genotipo**. El **fenotipo** del cromosoma es la codificación del **genotipo**, en este caso es el desglose verbal de las acciones y la función **Promedio** actúa para el cálculo de la *Aptitud* del *cromosoma*. Además tiene procedimientos y funciones para representarlo como un *cromosoma*, es decir que se tiene una

¹⁰¹ Una explicación completa y detallada de la técnica de la ruleta se puede ver en los anexos.

función de *cruza* y un procedimiento de *mutar*.

Procedimientos/Funciones:

- a) **nacer**: Procedimiento que recibe un arreglo que contiene las acciones y el número máximo que una acción puede tener. Usa estos datos para llenar los propios y así inicializar el objeto. Como el **Portafolio** tiene las acciones por separado como variables independientes usa **set Acciones** para así pasar del arreglo dado a las cuatro variables.
- b) **Set Acciones**: Procedimiento que recibe un arreglo que contiene las acciones y lo pasa a las diez variables (E1, E2, E3, E4, E5, E6, E7, E8, E9, E10) del **Portafolio**. Esto lo realiza tras llamar a la función **ordenar** que regresará un arreglo tras haber ordenado las acciones mediante sus números. Por ejemplo, si se recibe (5, 6, 3, 4, 2, 7, 1, 9, 10, 8) se ordena a (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) y se igualan las acciones E1 = 1, E2 = 2, E3 = 3, E4 = 4, E5 = 5, E6 = 6, E7 = 7, E8 = 8, E9 = 9, E10 = 10.
- c) **Get Acciones**: Función que convierte las diez variables en un arreglo y lo arroja de regreso.
- d) **Cruzar**: Función que recibe a otro **Portafolio** para así realizar la cruce entre ambos y regresa un arreglo con los hijos de ambos **Portafolios**. Esta función es un tanto complicada debido a la restricción de que las *Acciones* no pueden repetirse. Lo primero que hace es **Obtener Todas las Acciones Diferentes** entre los dos **Portafolios**. Por ejemplo, si llega (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) y (3, 4, 5, 6, 7, 8, 9, 10, 11, 12) entonces se obtendrá (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12). Luego se obtiene al azar un punto de cruce (Donde se realizará el corte y el intercambio de material genético) y se checa si las Acciones diferentes son 20, en caso de serlo realiza una cruce normal, y en caso de no ser así realiza una cruce especial. La cruce normal simplemente toma la primera parte del primer padre (cortando a ambos padres por el punto de cruce) y la segunda parte del segundo padre y forma a un hijo, luego toma la primera parte del segundo padre y la segunda parte del primer padre y forma al segundo hijo. Por ejemplo: (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) y (11, 12, 13, 14, 15, 16, 17, 18, 19, 20) en el punto de corte 2 daría (1, 2, 3, 4, 5 | 16, 17, 18, 19, 20) y (6, 7, 8, 9, 10 | 11, 12, 13, 14, 15). En caso de ser una cruce especial, se observan dos casos, el caso en que todas las acciones son iguales y el que sólo algunas lo son. En el primer caso simplemente se llena a ambos hijos con las acciones del primer padre y son arrojados de vuelta. En el segundo caso se usa la función **Obtener Repetidos** que obtendrá las acciones que se repiten y **Obtener Diferentes**, que obtendrá sólo las acciones que no se repiten. Entonces se pasará el número de repetidos necesario para llenar el punto de corte (o todos en el caso de que no se llene la condición) y luego se seleccionarán las partes del segundo o primer padre según sea el caso y se regresa a estos hijos formados. Obviamente, cada hijo es inicializado llamando a su función **nacer**

(En todos los casos) y luego se llama al procedimiento de **mutar**, que decidirá si mutar o no y en donde mutar.

- e) **Mutar:** Procedimiento que obtiene una probabilidad al azar, y si es menor o igual que 0.05 entonces seleccionará una localidad al azar de un arreglo de acciones (que se pide con **Get Acciones**) y entonces realizará un random selectivo, ya que formará un arreglo con todas las acciones desde el 1 hasta el **Máximo** que contiene el Portafolios (Para eso es el **Máximo**). Y luego se removerán todas aquellas acciones que contiene el portafolio y se seleccionará una acción al azar de todas las demás que quedarán, reemplazando el valor de la localidad que se sacó random también. Por ejemplo se mutará (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) en la posición 2. Entonces se generará (1, 2,..., 40), se eliminarán 1, 2, 3, 4, 5, 6, 7, 8, 9, 10y quedará (11,...,40) de donde se seleccionará al azar un número, digamos el 24 y se colocará en lugar de lo que tenía la posición 2: (1, 24, 3, 4, 5, 6, 7, 8, 9, 10). Posteriormente se usará el procedimiento **Set Acciones** para así asignar (y ordenar) las nuevas acciones del Portafolio.
- f) **Obtener Todas las Acciones Diferentes:** Esta función recibe un par de portafolios, obtiene sus acciones y entonces saca todas las diferentes acciones que tienen entre los dos. Por ejemplo: (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) y (10, 11, 12, 13, 14, 15, 16, 17, 18, 19) regresará un arreglo con (1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19).
- g) **Obtener Repetidos:** Esta función recibe un par de portafolios, obtiene sus acciones y entonces saca todas aquellas que se repiten entre ellas. Por ejemplo: (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) y (2, 4, 5, 6, 11, 12, 13, 14, 17) regresará un arreglo con (2, 4).
- h) **Obtener Diferentes:** Esta función recibe un par de portafolios, obtiene sus acciones y entonces saca todas aquellas que no se repiten entre ellas. Por ejemplo: (1, 2, 3, 4, 5, 6, 7, 8, 9, 10) y (2, 4, 5, 6, 11, 12, 13, 14, 17) regresará un arreglo con (1, 3, 7, 8, 9, 10, 11, 12, 13, 14, 17).
- i) **Ordenar:** Esta función recibe un arreglo de números y los ordena de mayor a menor.
- j) **Promedio:** Esta función recibe una Base de Datos (un recordset) y saca el promedio de las acciones del **Portafolio** (Usando el registro actual) arrojándolo como su resultado.
- k) **Aumentar Veces:** Este procedimiento puede recibir el incremento a sumar a las veces (obviamente el default es 1) y lo suma al atributo de *Veces* que tiene el **Portafolio**. Este procedimiento se manda a llamar una vez que se realiza la **Reproducción** en el **Ambiente** y el **Portafolio** a agregar a la población ya existe. En el caso contrario simplemente se agrega el **Portafolio** sin llamar a este procedimiento.

Consolid: Base de Datos que contiene toda la información de los **Portafolios** que se almacenó periódicamente. De aquí puede sacarse el

Promedio de todo un período o bien el Promedio de ciertas Acciones. Esta Base de Datos se obtiene mediante una consulta SQL a un archivo localizado en el folder relativo "*\BD\Consolida.mdb*". Es decir que si el proyecto está localizado en "*C:\Proyectos*" entonces la Base de Datos se buscará en "*C:\Proyectos\BD\Consolida.mdb*".

Una vez descrito el comportamiento de las clases identificadas podemos encontrar que necesitamos de una función que ordene a los Portafolios (Los **Habitantes**) dado una Base de Datos en cierto registro y regrese la colección ordenada, una función que regrese un número entre un rango dado y otra que pueda distinguir si un elemento existe dentro de una colección o no (esto se usa debido a que Habitantes es una colección que usa un *Id* para saber si el **Portafolio** ya existe o no y poder aumentar sus veces o simplemente agregarlo.). Estos procedimientos y funciones se almacenarán en un módulo común para su uso. También podemos darnos cuenta que se necesita hacer una **consulta SQL**, y debido a que ya se tiene un módulo que resuelve esta cuestión, mismo que se usa en el proyecto pasado y que se agrega a este proyecto también.

Finalmente, se diseñó una interfaz gráfica que permita el mayor aprovechamiento del proyecto, por lo cual se decidió incluir una opción para saber si se llenará al azar la población o no, una opción que permitirá usar todos los registros de la base de datos o ciertas iteraciones dadas por el usuario, una lista para desplegar el mejor elemento, un lugar donde mostrar cuántas veces apareció ese elemento y otro para desplegar su último valor, un botón que mande todos estos valores a la clase **Ambiente** y una barra de progreso que permitirá ver el progreso del Algoritmo genético para así evitar pensar que el proyecto se ha trabado o congelado tras varias iteraciones.

Por lo mismo, se muestra la interfaz descrita con anterioridad después del párrafo actual. Cabe mencionar que esta interfaz se encarga de crear una instancia (un objeto) de la clase **Ambiente**, y que será en base a ese objeto en el que se trabajará. También crea el recordset y lo llena con el resultado de la consulta SQL utilizando la Base de Datos que se ha mencionado anteriormente. Obviamente da como parámetro del arreglo de portafolios un arreglo que va desde el 1 hasta el 40, lo cual era parte del problema en sí, esto bien podría cambiarse fácilmente.

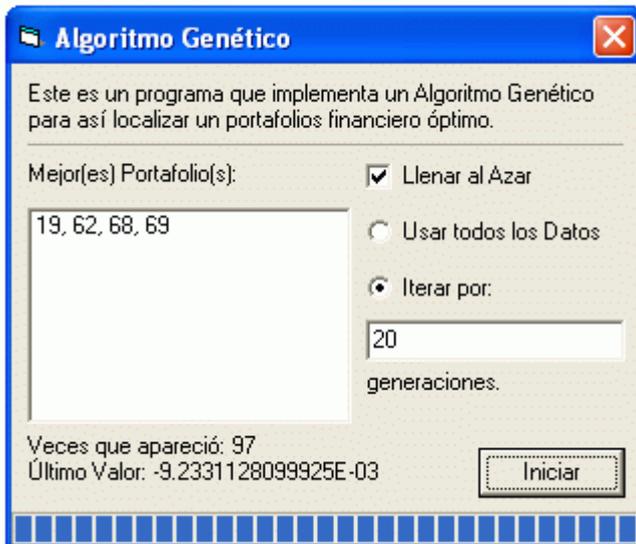


Imagen 25: Interfaz del programa del Algoritmo genético

Finalmente queda explicar el por qué se utiliza un porcentaje de 65% para la cruce y 5% para la mutación. En realidad el porcentaje de cruce, el porcentaje de mutación y el tamaño de la población son parámetros que se dan al Algoritmo genético. La probabilidad de cruce va a decir que tan a menudo se realizará la cruce, si no hay cruce se puede sacar una copia de los padres y usar esa. En nuestro caso simplemente no hacemos nada al respecto ya que la población siempre es la misma y nos deshacemos de los débiles. Esto es conocido como "elitismo", es decir que aparte de la ruleta estamos usando en cierta manera el mencionado elitismo. Si la probabilidad de cruce es del 100% entonces la nueva población es formada sólo por los hijos. Si la probabilidad de cruce es de 0% entonces toda la población es formada por los "padres". La cruce se hace esperando que los nuevos cromosomas sean mejores. En algunos experimentos se muestra que lo mejor es mantener este porcentaje desde 60% hasta 100%. La probabilidad de mutación decide que tan a menudo se mutarán los cromosomas. Si no hay mutación los hijos no cuentan con cambio alguno, si hay mutación una o más partes del cromosoma son cambiadas. Si la probabilidad de mutación es del 100% todo cromosoma mutará, si es de 0% entonces ningún cromosoma cambiará. La mutación generalmente evita que el Algoritmo genético caiga en máximos o mínimos locales. La mutación no debería ocurrir muy a menudo ya que el Algoritmo genético puede degenerar en una búsqueda al azar. La recomendación es que se mantenga un porcentaje entre 1 y 5%.

ANEXO 3

PANTALLAS DE APLICACIÓN ECONOMÁTICA

En la imagen 1 podemos observar la pantalla de bienvenida de Economática⁹⁵.

El sistema fue desarrollado en Brasil y actualmente tiene información entre otros países de Brasil, Chile, Argentina, Perú, Colombia, Venezuela, Estados Unidos y México, desde 1980 y hasta la fecha.

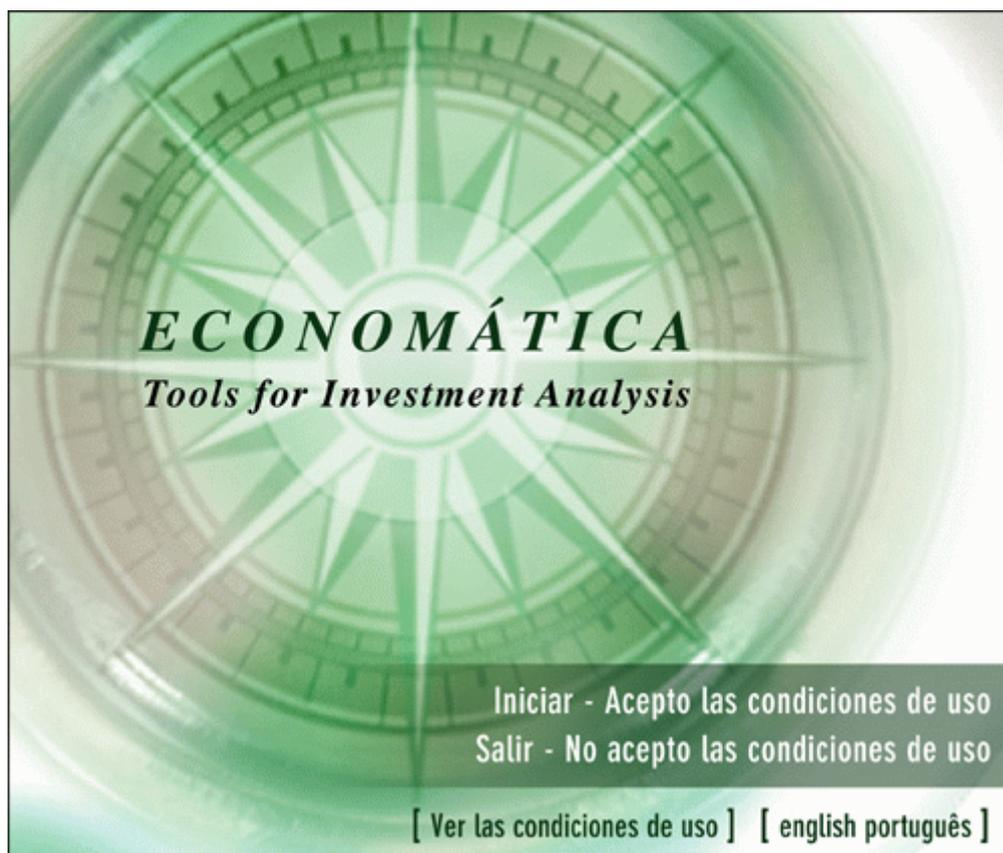


Imagen 1: Pantalla de bienvenida de Economática

Pasos para la obtención de las acciones a ser incluidas en la investigación

Para este estudio en particular sólo interesa la información de México por lo que se tomó sólo la base de datos de la BMV, en la imagen 2, que hemos partido en la parte inicial y en la parte final

⁹⁵ El sistema Economática es una herramienta para el análisis de inversiones en acciones utilizada por analistas de las más grandes instituciones financieras del mundo. Fundada en 1986, Economática está hoy presente en varios países. Su experiencia local es garantía de calidad incomparable de la base de datos y su software permite extrema agilidad en el manejo de las informaciones.

de la tabla mostrada por Economática, podemos observar que la información relativa a la BMV de inicio contiene un total de 968 registros, dichos registros corresponden a las diferentes emisiones de toda clase de instrumentos financieros que se han colocado en la BMV a través de la historia.

		Crear columna		Crear columna
1	3M Co.		933	SIRV SCOTIA9 B
2	A C Mexicana B1		934	SIRV SCOTINT B
3	Accel S.A. B		935	SIRV ST&ER-A B
4	Acer Latinoamerica		936	SIRV ST&ER-D B
5	Agro Ind Exportador A		937	SIRV ST&ER-I B
6	Alcoa Inc		938	SIRV ST&ER-M B
7	Alfa S.A. A		939	SIRV STERDOW B
8	Almacenadora Acce		940	SIRV VACRECI B
9	Almacenes Coppel SA 2		941	SIRV VAINDEX B
10	ALSEA		942	SIRV VALMX20 B
11	Altos Hornos de Mex		943	SIRV VALMX22 B
12	Altria Group Inc		944	SIRV VALMX24 B
13	America Movil A		945	SIRV VALMX26 B
14	America Movil AADR		946	SIRV VALOR4C B
15	America Movil L		947	SIRV VALUEV5 B
16	America Movil LADR		948	SIRV VALUEV6 B
17	America Telecom A1		949	SIRV VECTCR B
18	American Express Co		950	SIRV VECTIND B
19	Apasco S.A.		951	SIRV VECTPA B
20	Ara Consorcio		952	SIRV VECTPRO B
21	Arabela		953	SIRV ZCAP1 B
22	Arca Embotelladora		954	SIRV ZCAP2 B
23	Argos Embotelladora B		955	SIRV ZCAP4 B
24	Aristos Consorcio A		956	SIRV ZEVEERST B
25	Aristos Consorcio B		957	T-BOND 10 Yrs Yield
26	ASureste ADR		958	T-BOND 30 Yrs Yield
27	ASureste B		959	Tasas de Interes MX 182
28	AT&T Corp		960	Tasas de Interes MX 28
29	Autlan Cia. Minera B		961	Tasas de Interes MX 360
30	Bachoco Industrias ADR		962	Tasas de Interes MX 91
31	Bachoco Industrias UBL		963	Tasas de Interes MX TI28
32	Bafar Grupo B		964	Tasas de Interes MX TI91
33	Banamex Accival GF O		965	Tipo Cambio MX Euro
34	Bimbo Gpo A		966	Tipo Cambio MX FIX
35	Boeing Corp		967	Tipo Cambio MX SPOT
36	Bret Embotelladora		968	Unidades Inversion

Imagen 2: Parte inicial y final de la base de datos de Economática

Esta cantidad contiene toda la información de instrumentos colocados en la BMV desde 1980 y hasta la fecha, por lo que procedimos a efectuar un proceso de filtrado. En particular el tipo de activo que a nosotros nos interesa es el de las acciones, por lo

que utilizando la ventana de títulos de Economatica, sobre esta misma base de datos creamos una nueva columna que muestre el tipo de activo y en la base de datos resultante, insertamos un filtro para que sólo tome como tipo de activo *acciones*. Así, de una base de datos de 968 registros, se convirtió en una base de datos de tan sólo 206. Los pasos de este proceso se muestran en las imágenes 3, 4, 5, 6 y 7.

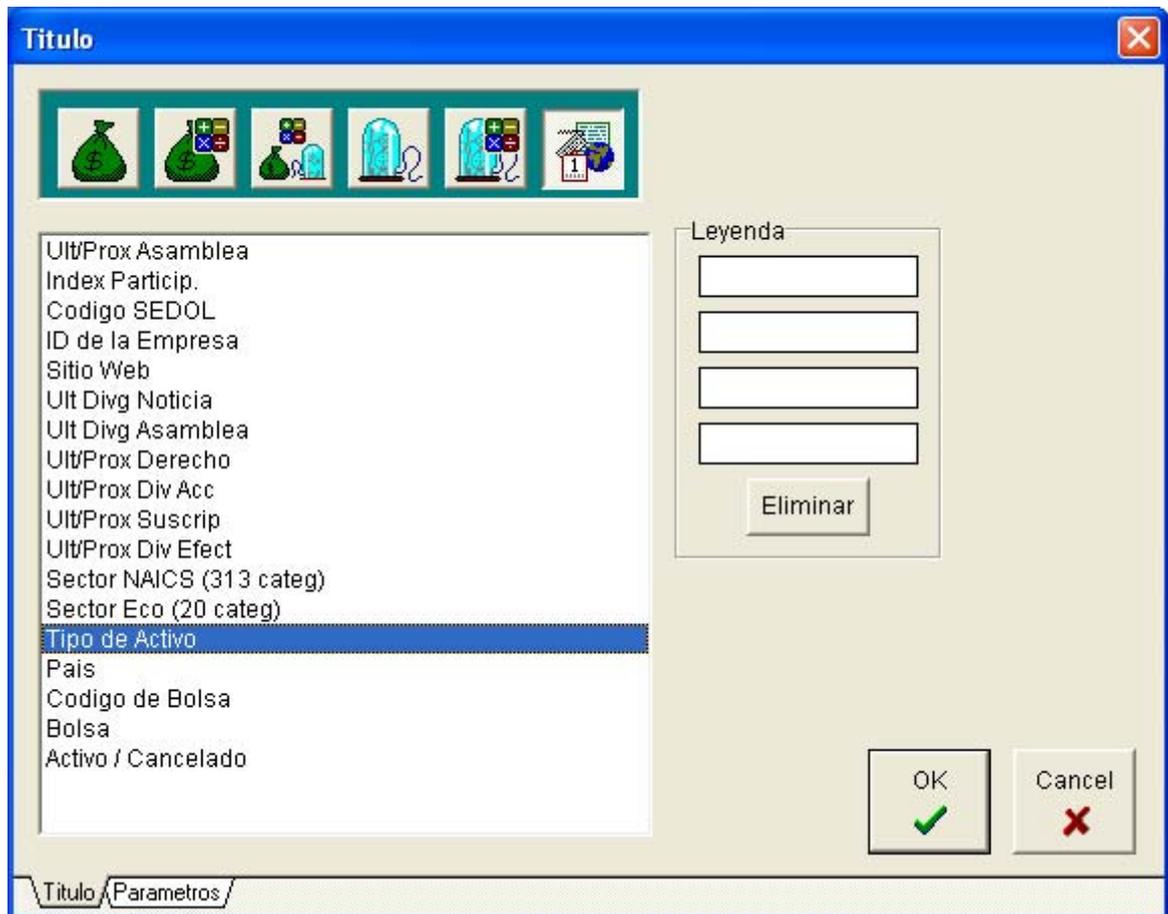


Imagen 3: Selección del tipo de activo en Economatica

1	3M Co.	ADR	949	SIRV VECTCR B	Cuota de Fonc
2	A C Mexicana B1	Accion	950	SIRV VECTIND B	Cuota de Fonc
3	Accel S.A. B	Accion	951	SIRV VECTPA B	Cuota de Fonc
4	Acer Latinoamerica	Accion	952	SIRV VECTPRO B	Cuota de Fonc
5	Agro Ind Exportador A	Accion	953	SIRV ZCAP1 B	Cuota de Fonc
6	Alcoa Inc	ADR	954	SIRV ZCAP2 B	Cuota de Fonc
7	Alfa S.A. A	Accion	955	SIRV ZCAP4 B	Cuota de Fonc
8	Almacenadora Acce	Accion	956	SIRV ZEVERST B	Cuota de Fonc
9	Almacenes Coppel SA 2	Accion	957	T-BOND 10 Yrs Yield	Renta Fija
10	ALSEA	Accion	958	T-BOND 30 Yrs Yield	Renta Fija
11	Altos Hornos de Mex	Accion	959	Tasas de Interes MX 182	Renta Fija
12	Altria Group Inc	ADR	960	Tasas de Interes MX 28	Renta Fija
13	America Movil A	Accion	961	Tasas de Interes MX 360	Renta Fija
14	America Movil AADR	ADR	962	Tasas de Interes MX 91	Renta Fija
15	America Movil L	Accion	963	Tasas de Interes MX TI28	Renta Fija
16	America Movil LADR	ADR	964	Tasas de Interes MX TI91	Renta Fija
17	America Telecom A1	Accion	965	Tipo Cambio MX Euro	Moneda
18	American Express Co	ADR	966	Tipo Cambio MX FIX	Moneda
19	Apasco S.A.	Accion	967	Tipo Cambio MX SPOT	Moneda
20	Ara Consorcio	Accion	968	Unidades Inversion	Renta Fija

Imagen 4: Parte inicial y final de la base de datos de Económica con tipo de activo

Imagen 5: Pantalla de filtrado de Económica

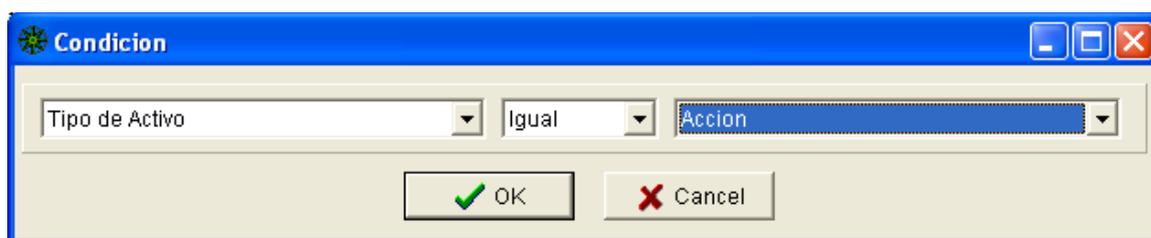


Imagen 6: condición de filtrado tipo de activo=acción

Aun con esta importante reducción en el número de registros, debemos tomar en consideración que muchos de ellos corresponden a acciones que ya no están cotizando y su estatus es de inactivas, por lo que procedimos a crear una nueva columna sobre el estatus de la acción, ya sea este activo o inactivo, de esta manera se logró una nueva reducción al pasar de 206 a 156 acciones. El proceso es similar al mostrado para el filtrado del tipo de acción *activo*, por lo que sólo presentamos en la imagen 8 la tabla final resultante de este procedimiento.

1	A C Mexicana B1	Accion	187	Tekchem S.A. A	Accion
2	Accel S.A. B	Accion	188	Telecom Carso Globa A1	Accion
3	Acer Latinoamerica	Accion	189	Telefs de Mex A	Accion
4	Agro Ind Exportador A	Accion	190	Telefs de Mex L	Accion
5	Alfa S.A. A	Accion	191	Televisa Gpo A	Accion
6	Almacenadora Acce	Accion	192	Televisa Gpo CPO	Accion
7	Almacenes Coppel SA 2	Accion	193	Tenaris S.A. de C.V	Accion
8	ALSEA	Accion	194	Tenedora US B-1	Accion
9	Altos Hornos de Mex	Accion	195	TMM GRUPO A	Accion
10	America Movil A	Accion	196	Tribasa Grupo	Accion
11	America Movil L	Accion	197	Tubos de Acero Mex	Accion
12	America Telecom A1	Accion	198	TV Azteca CPO	Accion
13	Apasco S.A.	Accion	199	Unefon A	Accion
14	Ara Consorcio	Accion	200	Union de Capitales B	Accion
15	Arabela	Accion	201	US Commercial B-1	Accion
16	Arca Embotelladora	Accion	202	Valle Jugos del B	Accion
17	Argos Embotelladora B	Accion	203	VALUEGF O	Accion
18	Aristos Consorcio A	Accion	204	Vitro A	Accion
19	Aristos Consorcio B	Accion	205	Wal Mart de Mexico C	Accion
20	ASureste B	Accion	206	Wal Mart de Mexico V	Accion

Imagen 7: Parte inicial y final de la base de datos de Economía con tipo de activo=acción

No siendo esta una disminución particularmente importante, si es trascendental el estatus, pues si pretendemos hacer un pronóstico de lo que va a ocurrir a futuro una de las grandes fuentes de ruido

sería el usar información que ya no fuera relevante como en este caso lo podría ser trabajar con acciones que ya están canceladas.

En un caso no fue posible evitar esto, ya que las acciones fueron canceladas cuando el trabajo ya estaba en un proceso muy avanzado y en función de la consistencia de la construcción del modelo no fue omitida.⁹⁶

Otra consideración que en particular nos pareció importante es que a la fecha de la selección de la muestra de acciones para efectuar los análisis, éstas hubieran cotizado, así que procedimos a efectuar un nuevo filtro de los registros tomando ahora el precio de cierre como parámetro. La imagen 9 muestra el resultado que se obtuvo con este nuevo filtro. Como se puede observar, se pasó de 158 a sólo 72 acciones.

Con 72 acciones bien podría haberse efectuado ya el análisis de agrupamientos para encontrar las acciones a las que se iba a trabajar tanto con el modelo de conteo como con el algoritmo genético.

Sin embargo con esta información la formación de grupos se hubiera basado exclusivamente en una variable que era la de precio de cierre. Esto que no es malo del todo, si llevaría a la formación de grupos que fueran parecidos sólo por el precio de cierre, lo cual para la formación del portafolio que obtenga el mayor rendimiento sería demasiado aleatorio. Por lo tanto decidimos incluir seis variables adicionales a la de precio de cierre para que el agrupamiento fuera más significativo.

Estas variables son: rendimiento, beta, volatilidad, volumen de operaciones en pesos, número de operaciones y tiempo de operar en días. Los conceptos de volumen de operaciones en pesos y número de operaciones son las que se observaron en la fecha del 7 de enero de 2004 y el dato está tomado de Economatica como se observa en la imagen 10. El tiempo de operar en días se obtuvo de contar los días desde el inicio de operaciones de la firma (como también se observa en la imagen 10) hasta el 7 de enero de 2004.

En cuanto a los conceptos de rendimiento, beta y volatilidad estos fueron calculados para cada una de las acciones usando el Microsoft Excel para el rendimiento y el mismo software de Economatica para el resto. Estos tres conceptos siempre son muy importantes cuando se habla de portafolios de inversión porque

⁹⁶ GFBBVA Bancomer.

representan las variables que se tratan de optimizar; ya sea maximizando el rendimiento o reduciendo la volatilidad, y en el

caso de la β porque representa el grado de afectación en la variación total del mercado, debido a la variación de un activo individual. Sus definiciones y el modo de calcularlos se presenta a continuación.

1	Accel S.A. B	Accion	activo
2	Agro Ind Exportador A	Accion	activo
3	Alfa S.A. A	Accion	activo
4	Almacenes Coppel SA 2	Accion	activo
5	ALSEA	Accion	activo
6	Altos Hornos de Mex	Accion	activo
7	America Movil A	Accion	activo
8	America Movil L	Accion	activo
9	America Telecom A1	Accion	activo
10	Apasco S.A.	Accion	activo
11	Ara Consorcio	Accion	activo
12	Arca Embotelladora	Accion	activo
13	Aristos Consorcio A	Accion	activo
14	Aristos Consorcio B	Accion	activo
15	ASureste B	Accion	activo
16	Autlan Cia. Minera B	Accion	activo
17	Bachoco Industrias UBL	Accion	activo
18	Bafar Grupo B	Accion	activo
19	Bimbo Gpo A	Accion	activo
20	Cablevision CPO	Accion	activo
139	Savia A	Accion	activo
140	Simec Grupo B	Accion	activo
141	Soriana Organizacio B	Accion	activo
142	Synkro Industrias A	Accion	activo
143	Synkro Industrias C	Accion	activo
144	Tekchem S.A. A	Accion	activo
145	Telecom Carso Globa A1	Accion	activo
146	Telefs de Mex A	Accion	activo
147	Telefs de Mex L	Accion	activo
148	Televisa Gpo A	Accion	activo
149	Televisa Gpo CPO	Accion	activo
150	TMM GRUPO A	Accion	activo
151	TV Azteca CPO	Accion	activo
152	Unefon A	Accion	activo
153	US Commercial B-1	Accion	activo
154	Valle Jugos del B	Accion	activo
155	VALUEGF O	Accion	activo
156	Vitro A	Accion	activo
157	Wal Mart de Mexico C	Accion	activo
158	Wal Mart de Mexico V	Accion	activo

Imagen 8: Parte inicial y final de la base de datos de Económica con activo/cancelado=activo

1	Accel S.A. B	Accion	activo	2.02
2	Agro Ind Exportador A	Accion	activo	0.30
3	Alfa S.A. A	Accion	activo	37.35
4	America Movil A	Accion	activo	16.34
5	America Movil L	Accion	activo	16.42
6	America Telecom A1	Accion	activo	15.51
7	Apasco S.A.	Accion	activo	92.19
8	Ara Consorcio	Accion	activo	28.76
9	Arca Embotelladora	Accion	activo	23.08
10	ASureste B	Accion	activo	19.40
11	Bachoco Industrias UBL	Accion	activo	19.50
12	Bimbo Gpo A	Accion	activo	22.29
13	Cablevision CPO	Accion	activo	12.20
14	Cementos Chihuahua	Accion	activo	14.00
15	Cemex S.A. CPO	Accion	activo	59.45
16	Cid Mega Resort	Accion	activo	3.29
17	Cintra S.A. A	Accion	activo	1.60
18	Comercial Mexicana UBC	Accion	activo	12.08
19	Continental Grupo	Accion	activo	20.50
20	Corp Interam de Ent B	Accion	activo	21.50
53	Nadro S.A. B	Accion	activo	4.90
54	Penoles Industrias	Accion	activo	50.00
55	Posadas Gpo A	Accion	activo	5.55
56	Posadas Gpo L	Accion	activo	5.55
57	San Luis Corp CPO	Accion	activo	3.00
58	Sare B	Accion	activo	6.15
59	Savia A	Accion	activo	1.68
60	Simec Grupo B	Accion	activo	29.80
61	Soriana Organizacio B	Accion	activo	26.67
62	Tekchem S.A. A	Accion	activo	0.404
63	Telecom Carso Globa A1	Accion	activo	15.57
64	Telefs de Mex A	Accion	activo	18.51
65	Telefs de Mex L	Accion	activo	18.51
66	Televisa Gpo CPO	Accion	activo	23.18
67	TV Azteca CPO	Accion	activo	6.18
68	US Commercial B-1	Accion	activo	4.90
69	Valle Jugos del B	Accion	activo	14.75
70	Vitro A	Accion	activo	13.20
71	Wal Mart de Mexico C	Accion	activo	31.13
72	Wal Mart de Mexico V	Accion	activo	33.65

Imagen 9: Parte inicial y final de la base de datos de Económica con precio de cierre no nulo

		Tipo de Activo	Activo / Cancelado	Cierre 07Ene04 ajust p/var cap en moneda orig	Beta Ene04 60 meses en moneda orig	Volatility 07Ene04 1 meses en moneda orig	Fecha de la Ult Operacion	Fecha del Inicio de la Serie	Volumen \$ 07Ene04 en moneda orig en miles
1	Accel S.A. B	Accion	activo	2.02	0.4	0.2	07/01/2004	30/01/1992	314
2	Alfa S.A. A	Accion	activo	37.35	1.1	0.3	07/01/2004	02/12/1991	80,890
3	Apasco S.A.	Accion	activo	92.19	0.6	0.1	07/01/2004	02/12/1991	16,752
4	Ara Consorcio	Accion	activo	28.76	0.8	0.2	07/01/2004	26/09/1996	37,031
5	Bachoco Industrias UBL	Accion	activo	19.50	0.3	0.1	07/01/2004	19/09/1997	5,206
6	Bimbo Gpo A	Accion	activo	22.29	0.4	0.2	07/01/2004	02/12/1991	8,019
7	Cementos Chihuahua	Accion	activo	14.00	0.6	0.3	07/01/2004	27/02/1992	3,708
8	Cemex S.A. CPO	Accion	activo	59.45	1.0	0.2	07/01/2004	03/04/1992	104,670
9	Cid Mega Resort	Accion	activo	3.29	-0.1	0.2	07/01/2004	29/09/1997	74
10	Comercial Mexicana UBC	Accion	activo	12.08	1.1	0.2	07/01/2004	02/12/1991	36,383
11	Continental Grupo	Accion	activo	20.50	0.8	0.4	07/01/2004	02/12/1991	21,199
12	Corp Interam de Ent B	Accion	activo	21.50	1.1	0.1	07/01/2004	19/12/1995	13,088
13	Dataflux B	Accion	activo	0.82	2.1	0.3	07/01/2004	21/07/1997	270
14	Desc Soc Fom Ind B	Accion	activo	4.25	0.7	0.5	07/01/2004	02/12/1991	6,953
15	Desc Soc Fom Ind C	Accion	activo	3.60	0.6	0.5	07/01/2004	08/02/1995	560
16	Ekco	Accion	activo	1.16	0.4	1.0	07/01/2004	10/12/1991	6
17	Elektra Gpo	Accion	activo	63.81	1.9	0.3	07/01/2004	26/09/1994	11,295
18	Fomento Econ Mex UBD	Accion	activo	42.77	1.0	0.2	07/01/2004	02/12/1991	186,301
19	Fragua Corporativo B	Accion	activo	22.50	0.4	0.0	07/01/2004	06/10/1997	380
20	GCarso A1	Accion	activo	41.75	1.1	0.2	07/01/2004	02/12/1991	24,644
21	Geo Corporacion B	Accion	activo	59.15	1.4	0.2	07/01/2004	13/09/1994	12,573
22	GFBanorte O	Accion	activo	39.89	1.0	0.3	07/01/2004	03/01/1994	45,767
23	GFBBVA Bancomer B	Accion	activo	9.96	1.1	0.2	07/01/2004	17/03/1992	215,463
24	GFInbursa O	Accion	activo	12.77	1.0	0.3	07/01/2004	04/01/1994	54,147
25	Gigante Gpo	Accion	activo	6.20	0.5	0.2	07/01/2004	29/11/1993	136
26	GInd Saltillo	Accion	activo	18.61	0.6	0.1	07/01/2004	10/12/1991	432
27	GMaseca (Gruma) B	Accion	activo	16.20	0.2	0.2	07/01/2004	29/04/1994	1,372
28	GMex Desarrollo B	Accion	activo	3.10	1.4	0.8	07/01/2004	08/02/1995	13
29	GMexico B	Accion	activo	30.25	0.8	0.4	07/01/2004	11/08/1994	57,383
30	GModelo C	Accion	activo	27.69	0.4	0.2	07/01/2004	01/07/1994	628,550
31	GNacional Provincia	Accion	activo	17.00	0.3	0.2	07/01/2004	17/12/1993	595
32	Herdez S.A.	Accion	activo	3.85	0.3	0.3	07/01/2004	13/09/1993	19
33	Hogar Consorcio B	Accion	activo	3.73	0.6	0.5	07/01/2004	23/05/1997	42
34	Hylsamex B	Accion	activo	6.70	0.3	0.2	07/01/2004	02/01/1995	493
35	ICA Soc Controlad	Accion	activo	2.98	1.2	0.4	07/01/2004	09/04/1992	46,713
39	KOF Coca-Cola L	Accion	activo	26.50	0.9	0.4	07/01/2004	29/11/1993	5,863
40	Lamosa Gpo B	Accion	activo	10.00	-0.0	0.4	07/01/2004	15/07/1998	330
41	Maseca GI B	Accion	activo	5.05	0.2	0.2	07/01/2004	02/01/1992	719
42	Penoles Industrias	Accion	activo	50.00	0.4	0.7	07/01/2004	03/12/1991	15,069
43	Posadas Gpo A	Accion	activo	5.55	0.4	1.4	07/01/2004	08/02/1995	46
44	Posadas Gpo L	Accion	activo	5.55	0.6	0.5	07/01/2004	21/12/1992	55
45	San Luis Corp CPO	Accion	activo	3.00	0.8	0.8	07/01/2004	10/03/1995	4
46	Savia A	Accion	activo	1.68	0.3	0.5	07/01/2004	03/12/1991	255
47	Simec Grupo B	Accion	activo	29.80	1.2	7.1	07/01/2004	02/12/1993	256
48	Soriana Organizacion B	Accion	activo	26.67	1.1	0.3	07/01/2004	03/12/1991	19,781
49	Tekchem S.A. A	Accion	activo	0.404	0.8	0.4	07/01/2004	11/10/1996	67
50	Telecom Carso Globa A1	Accion	activo	15.57	1.2	0.2	07/01/2004	26/07/1996	21,079
51	Telefs de Mex A	Accion	activo	18.51	0.9	0.1	07/01/2004	02/12/1991	353
52	Telefs de Mex L	Accion	activo	18.51	0.9	0.1	07/01/2004	02/12/1991	376,860
53	Televisa Gpo CPO	Accion	activo	23.18	1.6	0.2	07/01/2004	10/12/1991	122,178
54	TV Azteca CPO	Accion	activo	6.18	1.9	0.4	07/01/2004	15/08/1997	61,560
55	Valle Jugos del B	Accion	activo	14.75	-0.1	0.3	07/01/2004	26/01/1995	179
56	Vitro A	Accion	activo	13.20	0.6	0.4	07/01/2004	02/12/1991	27,412
57	Wal Mart de Mexico C	Accion	activo	31.13	1.0	0.2	07/01/2004	02/12/1991	82,048
58	Wal Mart de Mexico V	Accion	activo	33.65	1.0	0.2	07/01/2004	02/12/1991	170,277

Imagen 10: Emisoras resultante de todo el proceso de selección

REFERENCIAS

ABRAHAM, J.M., GOETZMANN, W.N., WATCHER, S.M. (1994). "Homogeneous Grouping of Metropolitan Housing Markets". *Journal of Housing Economics*.

ALEXANDER, S.S. (1961). Price movements in speculative markets: trends or random walks?, *Industrial Management Review* 2, pp. 7-26.

ALLEN, F. ; KARJALAINEN, R. (1999). Using genetic algorithms to find technical trading rules. *Journal of Financial Economics*, 51, 245-271.

ARNOT, R. (1980). "Cluster Analysis and Stock Price Co movement". *Financial Analyst Journal*, Nov-Dec.

BACK, T. (1993). *Optimal mutation rates in genetic search*, In, S. Forrest (ed.): *Proc. Fifth Int. Conf on Genetic Algorithms*", Morgan Kaufmann, San Mateo, CA, pp. 2-9.

BAUER, RICHARD J. Jr. (1994). *Genetic Algorithms and Investment Strategies*, John Wiley ; Sons, Inc.

BAUER, R.J. Jr. (1995), "Genetic Algorithms and the Management of Exchange Rate Risk," in BIETHAHN, J., NISSEN, V. (eds.) *Evolutionary Algorithms in Management Applications*, Springer Verlag, Berlin, Heidelberg, New York, pp 253-263.

BANZHAF, W.; NORDIN, P.; KELLER, R.E.; FRANCONI, F.D. (1998). Genetic Programming. An Introduction on the Automatic Evolution of Computer Programs and Its Applications, *Morgan Kaufmann Publisher, California*

BESSEMBINDER, H. ; CHAN, K (1995). The profitability of technical trading rules in the Asian stock markets, *Pacific-Basic Finance Journal* 3, pp. 257-284

BLUME, L., EASLEY D. ; O'HARA M. (1994). Market statistics and technical analysis: The role of volume. *Journal of Finance*, 49, 153-82.

BREIMAN, L., FRIEDMAN, J., STONE, C.J., OLSHEN, R.A. (1984) *Classification and Regression Trees*. Kluwer Academic Publishers. Boston, MA. USA.

BROCK, W.A.; LAKONISHHOCK, J.; LEBARON, B. (1992). Simple technical trading rules and the stochastic of stock returns, *Journal of Finance* 47, pp. 1731-1764

BROWN, S.J., GOETZMANN, W.N. (1997). "Mutual Funds Style". *Journal of Financial Economics*, 43.

BROWN, S.J., GOETZMANN, W.N., GRINBLATT, M. (1997) "Positive Portfolio Factors", Working Paper.

BROWN, S.; GOETZMANN, W.; KUMAR, A. (1998). The Dow Theory: William Peter Hamilton's Track Record Reconsidered, Working paper, *Stern school of Business*, New York University.

BURES, E. (1986). *Métodos de pronósticos aplicados a la administración*. ITESM. Monterrey.

COELLO, CARLOS A. (2001). *Introducción a la computación evolutiva*, CINVESTAV-IPN, cap. 2; pp. 39-64.

CORRADO, C.J.; LEE, S.H. (1992). Filter rule test of the economic significance of serial dependence in daily stock returns. *Journal of Financial Research* 15 Vol. 4, pp. 369-387

CHOPRA, N., LAKONISHOK J.; RITTER J. R. (1992). Measuring abnormal returns: Do stocks overreact? *Journal of Financial Economics* 31, 235-268.

CLIFFORD, H., STEPHENSON, W. (1975). *An Introduction to Numerical Taxonomy*. Academic Press. New York, USA.

COLIN, A. (1996), "A Genetic Programming-Based Approach to Generation of Foreign Exchange Trading Models, Conferencia de "Commerce, Complexity and Evolution, University of New South Wales".

CONRAD J.; KAUL G. (1988). Time-variation in expected returns. *Journal of Business* 61, 409-425.

CONRAD J.; KAUL G. (1989). Mean reversion in short-horizon expected returns. *Review of Financial Studies* 2, 225-240.

CUEVAS, F.J., SERVIN, M., RODRIGUEZ-VERA, R., SOSSA-AZUELA, J.H. A parametric method applied to phase recovery from a fringe pattern base don a genetic algorithm. *Optic Comunications*. 2002.

CUEVAS, F.J., SERVIN, M., RODRIGUEZ-VERA, R., SOSSA-AZUELA, J.H. Soft Computing for phase detection from fringe patterns. Kerala India, 2003.

DAVIS, LAWRENCE. (1991). *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, New York.

DAVIS, LAWRENCE. (1994). *Genetic Algorithms and Financial Applications*, in: Deboeck (ed., 1994) pp. 133-147.

- DEBONDT, W. F. M.; THALER R. (1985). Does the stock market overreact? *Journal of Finance* 40, 793-805.
- DRAKE, K., KIM, Y. (1997). "Abductive Information Modeling Applied to Financial Time Series Forecasting. Finance & Technology Publishing.
- DUNHAM, B., FRIDSHAL, D., NORTH, J.H. (1963). *Design by natural selection. Synthese*, 15:254-259
- EDDELBIITTEL, D. (1996). "A Hybrid Genetic Algorithm for Passive Management". Segunda Conferencia de Computación en Economía y Finanzas. Society of Computational Economics. Ginebra.
- ELTON, E.J., GRUBER, M.J. (1970). "Homogeneous Groups and the Testing of Economic Hypotheses". *Journal of Financial and Quantitative Analysis*.
- ESTRELLA, A.; MISHKIN, F. S. (1998). "Predicting US recessions: Financial variables as leading indicators". *The Review of Economics and Statistics* 80(1), p.p. 45-61.
- EVERITT, B. (1993) *Cluster Analysis*. 3rd Edition. Edward Arnold. London.
- FAMA, E.F. (1965). The behavior of Stock-Market Prices, *Journal of Business* 38, pp. 34-105.
- FAMA, E.F. (1970). Efficient capital markets: A review of theory and empirical work". *Journal of Finance* 25, p.p. 383-417.
- FAMA, E.F.; BLUME, M.E. (1966). Filter rules and stock market trading, *Journal of Bussiness* 39, pp. 226-241
- FAMA, E.; FRENCH K. R. (1988). Permanent and temporary components of stock prices. *Journal of Political Economy* 98, 247-273.
- FARREL, J. (1974). "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings". *Journal of Business*.
- FELDMAN, K.; TRELEAVEN, P. (1994). Intelligent systems in finance. *Appl. Mathematical Finance* 1, 195-207, Chapman ; Hall, London, UK.
- FERNANDEZ, F.; SOSVILLA, S.; ANDRADA J. (1999). Technical analysis in the Madrid stock exchange, *Documento de trabajo de 99-05 de la Fundación de Estudios de Economía Aplicada*.
- FOGEL, I.J.; OWENS, A.J.; WALSH, M.J. (1966). Artificial intelligence through simulated evolution, *Wiley*, New York

FRALEY, C., RAFERTY, A.J. (2000). "How many clusters? Which clustering method? Answers via model-based Cluster Analysis". Department of Statistics. University of Washington.

GEN, MITSUO, CHENG, RUNWEI. (2000) *Genetic Algorithms and Engineering Optimization*. Wiley-Interscience. New York.

GOLDBERG, D.E. (1989). Genetic Algorithms in Search, Optimization, and Machine Learning, *Addison Wesley*, Reading, Mass

GOLDBERG, D.E.; DEB, K. (1991). *A comparative analysis of selection schemes used in genetic algorithms*, In, G. Rawlins (ed.), *Foundations of Genetic Algorithms*, Morgan Kaufmann, San Mateo, CA, pp. 69-93.

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.

HOLLAND, JOHN H. (1962). Outline for a logical theory of adaptive systems, *Journal of the Association for Computing Machinery* 3, pp. 297-314

HOLLAND, JOHN H. (1975). *Adaptation in Natural and Artificial Systems*, The University of Michigan Press. Ann Arbor.

HOLLAND, JOHN H. (1980). Adaptive algorithms for discovering and using general patterns in growing knowledge-bases, *International Journal of Policy Analysis an Information System* 4, pp.217-240

HOLLAND, JOHN H. (1992). Genetic Algorithms, *Scientific American*, July, pp. 44-50.

HUDSON, R.; DEMPSEY, M.; KEASEY, K. (1996). A note on the weak form of efficiency of capital markets: The application of simple technical trading rules to the U.K. stock markets-1935-1994, *Journal of Banking and Finance* 20 Vol. 6, pp.11211132

JARDINE, N., SIBSON, R. (1971).*Mathematical Taxonomy*. John Wiley & Sons. London. UK.

KAUFMAN, L., ROUSSEEUW, P.J. (1987). "Clustering by means of medoids in Dodge Y. *Statistical Data Analysis Based on the L1-norm*. Amsterdam.

KIM, K. ; HAN, I. (2000). Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. *Expert Systems with Applications*, vol 19, N°2 Agosto, p.p. 125-132.

KING, B. (1996). "Market and Industry Factors in Stock Price Behavior". *Journal of Business*.

KINGDON, J. ; FELDMAN, K. (1995). Genetic algorithms and applications to finance. *Applied Mathematical Finance*, vol 2, N°2, Junio, p.p. 89-116.

KLIMASAUSKAS, C. (1995), "Developing a Multiple-Indicator Market Timing System: Theory, Practice and Pitfalls," in Lederman, J. and Klein, R.A. (eds.), *Virtual Trading*, Irwin Publishing, New York, pp 127-166.

KOZA, J.R. (1992). Genetic Programming: On the Programming of Computers By the Means of Natural Selection, MIT Pres, Cambridge, Mass

KURI, ANGEL. (1999). *A Comprehensive Approach to Genetic Algorithms in Optimization and Learning. Theory and Applications. Volume 1. Foundations*. Instituto Politécnico Nacional. México.

LAWERA, M. (1999). "Future prices: Data Mining and Modeling Approaches". Rice University. Houston, TX. USA.

LEINWEBER, D. ARNOTT, R. (1995). "Quantitative and Computational Innovation in Investment Management" .*Journal of Portfolio Management*, 21(2), 8-15.

LEUNG MARK T., DAOUK HAZEM, CHEN AN-SING (2000). Forecasting stock indices: a comparison of classification and level estimation models. *International Journal of Forecasting* (16)2, pp. 173-190.

LIVINGSTON, M. (1977). "Industry Movements of Common Stocks". *Journal of Finance*.

LO, A. W.; MACKINLEY A. C. (1988). Stock market price do not follow random walk: Evidence from a simple specification test. *Review of Financial Studies* 1, 41-66.

LO, A.W.; MACKINLEY (1990). Data snooping biases in test of financial asset pricing models, *The Review of Financial Studies* 3, pp. 431-467

LORASCHI, A., TETTAMANZI, A. (1996). *An Evolutionary Algorithm for Portfolio Selection Within a Downside Risk Framework*. in Dunis, C. (ed.). *Forecasting Financial Market*. John Wiley and Sons, Chichester, pp 275-285.

MAHFOUD, S., MANI, G. REIGEL, S. (1997). "Nonlinear Versus Linear Techniques for Selecting Individual Stocks". Citado en WEIGEND, A.S.,

ABUMOSTAFA, Y. Y REFENES, A-P. *Decision Technologies for Financial Engineering*, World Scientific, Singapore, pp 65-75.

MARSHALL, D. J.(1999), *Industrial Applications of Genetic Algorithms. Data Mining using genetic algorithms*. Edited by Charles Karr and L.Michael, CRC Press.

McQUEEN, G., SHIELDS, K., THORLEY, S.R. (1997). "Does the "Dow-10 investment strategy" Beat the Dow statistically and economically?" *Financial Analysts Journal*. Charlottesville, NC. USA.

MENDENHALL, W., SCHEAFFER, R.L., OTT, L. (1986). *Elementary survey sampling*. 3rd Edition. PWS Publishers. USA.

MEYERS, S. (1973). "A Re-Examination of Market and Industry Factors in Stock Price Behavior". *Journal of Finance*.

MICHALEWICZ, Z. (1992). *Genetic Algorithms + Data Structures = Evolutionary Programs*, Springer-Verlag, Berlin.

MIIHLENBEIN, H. (1992). *How genetic algorithms really work: mutation and hill-climbing*, In, R. Mdnner and B. Manderick (eds): *Parallel Problem Solving from Nature 2*, Elsevier Science, Amsterdam, pp. 15-26.

MILLS, T. (1996). Technical analysis and the London Stock Exchange: Testing trading rules using the FT30, *Department of Economics, Loughborough of Technology*, Economic Research Paper No. 96/1

NAVARRO, LUIS H. (2002). Estimación de la Volatilidad de la Tasa de Intercambio del Nuevo Sol versus el Dólar americano. *Núcleo de Estadística Computacional (NEC)*, Reporte Técnico.

NAVARRO, LUIS H. (2003) ¿Algoritmos Genéticos? *Núcleo de Estadística Computacional (NEC)*

NEELY, C.J.; WELER, P. ; DITTMAR, R. (1997). Is technical analysis in the foreign exchange market profitable A genetic programming aproach, *Journal of Financial Quantitative Analysis 32 Vol. 4*, pp. 405-426

NELSON, D.B. (1990). Stationarity and Persistence in the GARCH(1,1) model. *Econometric Theory*, 6:318-334.

NORUSIS, M.J. (1994) *SPSS Professional Statistics 6.1*. SPSS. Chicago, Ill. USA.

PACKARD, N.H. (1990). "A Genetic Learning Algorithm for the Analysis of Complex Data. *Complex Systems*", 45(5), 543-572.

PARISI, A., PARISI, F., CORNEJO, E. (2003). "Algoritmos Genéticos: Predicción de Índices Bursátiles de América del Norte". Facultad de Ciencias Económicas y Administrativas. Departamento de Administración. Universidad de Chile.

PARISI, A., PARISI, F.; GUERRERO, J. L. (2003). "Modelos Predictivos de Redes Neuronales en índices Bursátiles", *El Trimestre Económico*, México.

PEREIRA, ROBERT. (1996). Selecting parameters for technical trading rules using genetic algorithms, *Journal of Applied Finance and Investment*, Volume 1, No. 3, July/August pp. 27-34.

PEREIRA ROBERT. (1999). Forecasting ability but no profit ability: An empirical evaluation of genetic algorithm-optimised technical trading rules, Discussion Papers, School of Business, La Trobe University

PESARAN, M.H. ; TIMMERMANN A. (1992). "A simple nonparametric test of predictive performance". *Journal of Business and Economic Statistics* 10, pp. 461-465.

PICTET, OX., DACOROGNA, M.M., DAVE, R.D., CHOPARD, B., SCHIRRU, R., TOMASSINI, M. (1996), "Genetic Algorithms with Collective Sharing for Robust Optimization in Financial Applications," , Olsen and Associates, Zurich, Suiza.

POTERBA, J. M. ; SUMMERS L. H. (1988). Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics* 22, 27-59.

QI, M. (2001). "Predicting US recessions with leading indicators via neural network models". *International Journal of Forecasting* 17, p.p. 383-401.

ROOS QUINLAN, J. (1993). *C4.5: Programs for Machine learning*. Morgan-Kaufman. San Mateo, CA. USA.

ROUSSEEUW, P.J., (1985). "Some Thoughts on Agglomerative Cluster Analysis" CM ISI, Amsterdam.

SAARENVIRTA, G. (1998) "Data mining to improve profitability". The Management Accounting Magazine. Hamilton, N.J. USA.

SCHWEFEL, H.P. (1981). Numerical optimization of computer models, *Wiley*, New York

SIMON, HERBERT A. (1989). *Naturaleza y Límites de la Razón Humana*, Fondo

de Cultura Económica.

SUGAR, C., JAMES, G.M., (2003). "Finding the number of clusters in a data set : an information theoretic approach". Marshall School of Business. University of Southern California.

SUSUKI, Y.M., CUEVAS, F.J., S OSSA-AZUELA, J.H. CARPIO, J.M. Acomodo de patrones de materiales finitos usando Algoritmos genéticos. 2002.

TAYLOR, M.P. AND ALLEN, H. (1992). The use of technical analysis in the foreign exchange market, *Journal of Money and Finance* 11, pp. 304-314

VALLE DEL, DAVID (2002). *Modelo para Toma de Decisión en Portafolios de Inversión con Acciones Usando Algoritmos Genéticos*. Tesis Doctoral. Tecnológico de Monterrey, CCM. Cd. De México.

VOJINOVIC, Z., KECMAN, V., SEIDEL, R. (2001). "A data mining approach to financial time series modeling and forecasting". International Journal of Intelligent Systems in Accounting, Finance and Management. Chinchester. NZ.

WEBER, C.M. (2002) "Getting it right the first time: Accounting, auditing, financial systems and the federal government". The Journal of Government Financial Management. Alexandria, VA.USA.

WEISMANN, August, (1893). *The Germ Plasm: A Theory of Heredity*. Scott, London, UK.

WITTEN, I. (2000). *Data mining. Practical Machine Learning Tools and Techniques wit Java Implementation*. Morgan-Kaufman. San Francisco. CA. USA.

WU, Q. (2001). "Data mining and Knowledge discovery in financial research. Empirical investigation under currency". McGill University. Montreal CA.

ZIGHED, DA. (1996). Sipina-W, <http://eric.univ-lyon2.fr/~ricco/sipina.html>. Universitié Lumière, Lyon. FR.