

INSTITUTO TECNOLÓGICO Y DE ESTUDIOS
SUPERIORES DE MONTERREY

CAMPUS CIUDAD DE MÉXICO



TOMA DE DECISIONES BAJO UN AMBIENTE DE
INCERTIDUMBRE: UN ANÁLISIS COMPARATIVO

DOCTORADO EN ADMINISTRACIÓN

TESIS PRESENTADA POR
JORGE ALBERTO RAMÍREZ LANDA

ASESOR
DR. RALF EDER LANGE

AGO 2007



**TECNOLÓGICO
DE MONTERREY**

Hacemos constar que en la Ciudad de México, el día 29 de agosto de 2007 el alumno:

Jorge Alberto Ramírez Landa

sustentó el examen oral en defensa de la Tesis titulada:

Toma de decisiones bajo un ambiente de incertidumbre: un análisis comparativo

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. Guillermo Rodríguez Abitia, ha tomado la siguiente resolución:

APROBADO POR UNANIMIDAD

Dr. Ralf Eder Lange
Director de Tesis

Dr. Guillermo Rodríguez Abitia
Lector

Dr. Raúl Antonio Trejo Ramírez
Lector

Índice general

Índice de figuras	5
Resumen	10
Capítulo 1. Introducción	12
1. Problema a investigar	18
2. Definición de términos	28
3. Problemas y limitaciones	29
4. El proceso de la investigación	30
Capítulo 2. Marco teórico	32
1. Pronósticos	33
2. Inteligencia Artificial	49
Capítulo 3. Procedimiento de investigación	58
1. Planteamiento del problema	58
2. Propuesta	60
3. Procedimientos	60
Capítulo 4. Análisis de resultados finales	66
1. Resultados de los casos de prueba	66
2. Casos de estudio	85
Capítulo 5. Conclusiones	94

ÍNDICE GENERAL	4
1. Resumen	94
2. Conclusiones	96
3. Recomendaciones y trabajo futuro	98
El software Magister (manual de usuario)	102
4. Ejecución de Magister	102
5. El menú ARCHIVO	104
6. El menú PROYECTO	104
7. El menú OPCIONES	104
8. Ejemplo	107
Bibliografía	110

Índice de figuras

1. Diagrama de dispersión aparentemente sin relación entre las variables dependiente e independiente.	20
2. Diagrama de dispersión lineal.	20
3. Diagrama de dispersión no lineal (parabólico).	21
4. Regresión lineal sobre una serie de datos	22
5. Regresión no lineal.	24
6. Aproximación lineal sobre un conjunto de datos cualquiera.	27
7. Regresión parabólica (no lineal) sobre el conjunto de datos de la figura 6.	27
8. Regresión simbólica sobre el conjunto de datos de la figura 6 y 7.	28
1. Serie de datos seudo caótica generada por la ecuación 1.	62
2. Serie de datos seudo caótica generada por la ecuación 2.	62
3. Serie de datos seudo caótica generada por la ecuación 3.	63
4. Serie de datos seudo caótica generada por la ecuación 4	63
5. Serie de datos seudo caótica generada por la ecuación 5.	64
6. Índice NASDAQ	64
7. Índice IPC mexicano	65

ÍNDICE DE FIGURAS

	6
1. Diagrama de dispersión con tendencia lineal.	66
2. Diagrama de dispersión con tendencia no lineal.	67
3. Diagrama de dispersión estacional.	67
4. Diagrama de dispersión estacional con tendencia lineal.	68
5. Diagrama de dispersión sin tendencia.	68
6. Comparativa entre los datos reales y los pronosticados por el método de promedios móviles.	69
7. Pronósticos utilizando el método de promedios móviles ponderados.	70
8. Comparativa utilizando la técnica de suavizamiento exponencial.	70
9. Regresión lineal.	71
10. Pronóstico obtenido mediante la técnica de programación genética en un ambiente lineal.	71
11. Análisis de pronósticos mediante promedios móviles en un ambiente no lineal.	72
12. Comparativa de valor real contra el pronóstico obtenido por el método de promedios móviles ponderados.	73
13. Pronóstico mediante suavizamiento exponencial en un ambiente no lineal.	73
14. Análisis de regresión lineal sobre un diagrama de dispersión no lineal.	74

15. Resultado obtenido mediante programación genética para realizar pronósticos sobre una serie de datos no lineales.	74
16. Comparativa entre los datos reales y los pronosticados por el método de promedios móviles en una serie de datos estacionales.	75
17. Pronósticos utilizando el método de promedios móviles ponderados en un ambiente estacional.	75
18. Comparativa utilizando la técnica de suavizamiento exponencial en datos estacionales.	76
19. Regresión lineal en un ambiente estacional.	76
20. Pronóstico obtenido por la programación genética en datos con tendencia estacional.	77
21. Análisis de pronósticos mediante promedios móviles en un ambiente estacionales con tendencia lineal ascendente.	77
22. Comparativa de valor real contra el pronóstico obtenido por el método de promedios móviles ponderados.	78
23. Pronóstico mediante suavizamiento exponencial en un ambiente estacional ascendente.	78
24. Análisis de regresión lineal sobre un diagrama de dispersión estacional con tendencia lineal ascendente	79
25. Resultado obtenido mediante programación genética para realizar pronósticos sobre una serie de datos estacionales con tendencia lineal.	79
26. Análisis de pronósticos mediante promedios móviles en un ambiente no lineal.	80

27. Comparativa de valor real contra el pronóstico obtenido por el método de promedios móviles ponderados.	81
28. Pronóstico mediante suavizamiento exponencial en un ambiente sin tendencia.	81
29. Análisis de regresión lineal sobre un diagrama de dispersión sin tendencia alguna.	82
30. Resultado obtenido mediante programación genética para realizar pronósticos sobre una serie de datos sin tendencia.	82
31. Resultados del primer caso experimental	83
32. Promedio de los porcentajes de error de los diferentes casos de prueba.	84
33. Resultados sobre la prueba de normalidad aplicados a los diferentes escenarios.	85
34. Pronóstico del año 2006 utilizando la serie de 1991 al 2005.	86
35. Pronóstico del IPC entrenando el método propuesto con un año de anterioridad.	87
36. Pronóstico del IPC entrenando el método propuesto con 2 años de anterioridad.	88
37. Pronóstico del IPC entrenando el método propuesto con 3 años de anterioridad.	88
38. Pronóstico del IPC entrenando el método propuesto con 4 años de anterioridad.	89
39. Pronóstico del IPC entrenando el método propuesto con 5 años de anterioridad.	90

ÍNDICE DE FIGURAS

9

40. Pronóstico del índice NASDAQ.	90
41. Pronóstico del NASDAQ entrenando el método propuesto con 1 año de anterioridad.	91
42. Pronóstico del NASDAQ entrenando el método propuesto con 2 años de anterioridad.	92
43. Pronóstico del NASDAQ entrenando el método propuesto con 3 años de anterioridad.	92
44. Pronóstico del NASDAQ entrenando el método propuesto con 4 años de anterioridad.	93
45. Pronóstico del NASDAQ entrenando el método propuesto con 5 años de anterioridad.	93
1. Interface gráfica del software Magister.	103
2. Controles para modificar el comportamiento del algoritmo genético en el submenú "Algoritmo Genético".	105
3. Controles para modificar las propiedades del árbol de expresión en el submenú "Árbol de expresión".	106
4. Controles para seleccionar el conjunto de operaciones a utilizar del submenú "Set de Operaciones".	106
5. Controles para seleccionar el tipo de codificación a utilizar del submenú "Método".	107

Resumen

Las organizaciones modernas trabajan en ambientes turbulentos, los cuales ponen a prueba su crecimiento y sobrevivencia. En la administración de una organización es de suma importancia el tomar las decisiones correctas para lograr el objetivo de la empresa. Para tomar las mejores decisiones es necesario evaluar los posibles escenarios futuros, es decir, elaborar buenos pronósticos, principalmente de la demanda y de las ventas. De este pronóstico depende la planeación de la producción, de las compras, de los requerimientos de mano de obra, de los recursos financieros, etc., por lo que, sin lugar a duda, se convierte en información muy valiosa y determinante para alcanzar las metas de cualquier organización.

La elaboración del pronóstico puede ser simple o compleja, dependiendo de distintos factores. Muchas veces la demanda (o lo que se vaya a pronosticar) es afectada por factores cíclicos y estacionales, y otras veces influyen elementos de muy difícil previsión. En cualquier caso la experiencia y el conocimiento del mercado son vitales, pero a veces esto no es suficiente para la elaboración de un buen pronóstico. El optimismo o pesimismo del planeador podrían afectar sus proyecciones, por lo que resulta de gran utilidad el empleo de técnicas que permiten

medir la tendencia, la estacionalidad, entre otras características de la demanda a lo largo del tiempo. El uso de estas técnicas junto con la experiencia y grado de conocimiento del planeador, le pueden llevar a lograr mejores pronósticos, con los cuales podrá realizar una correcta planeación de las actividades productivas.

Para cualquier área de una empresa es fundamental conocer estos métodos: poder seleccionar un método de pronóstico apropiado, emplearlo y obtener conclusiones de él. En este sentido la tecnología ha contribuido suministrando herramientas muy valiosas a la hora de procesar la información.

En este trabajo, se investiga y se propone un método nuevo para realizar pronósticos bajo elementos de difícil previsión, bajo ambientes de alta incertidumbre. Para validar dicho método, se realiza un análisis comparativo entre los métodos tradicionales y la técnica propuesta.

CAPÍTULO 1

Introducción

Los sistemas de información se han vuelto indispensables para crear compañías competitivas, dirigir corporaciones globales y proporcionar productos y servicios útiles a los clientes. Los tipos de sistemas que se construyen hoy son muy importantes para el desempeño general de la organización, sobre todo en la economía actual altamente globalizada y basada en información. Los sistemas de información están impulsando tanto las operaciones diarias como la estrategia de las organizaciones. La tecnología de información permite a los administradores realizar planificación de recursos de empresa, lo que aumenta la precisión de los planes, de los pronósticos y del monitoreo de los principales procesos del negocio.

Un sistema de información estratégico modifica las metas, las operaciones, los productos, los servicios o las relaciones con el entorno de las organizaciones, a fin de ayudar a éstas a conseguir una ventaja sobre sus competidores. Hoy en día, los sistemas de información pueden elevar de forma tan drástica la productividad y la eficiencia de una compañía, que los negocios ven a la información como un arma contra sus competidores y como un recurso estratégico.

Los sistemas de información pueden servir para apoyar la estrategia en los niveles de negocio, de compañía y de industria. En el nivel de negocio, los sistemas de información pueden servir para ayudar a las compañías a convertirse en el productor de bajo costo, a diferenciar sus productos o a atender nuevos mercados. Los sistemas de información también pueden servir para retener a clientes y proveedores. En el nivel de compañía, los sistemas de información pueden servir para lograr nuevas eficiencias. En el nivel de industria, los sistemas pueden promover la ventaja competitiva, facilitando la cooperación con otras compañías de la industria, creando consorcios o comunidades para compartir información, intercambiando transacciones o coordinando actividades.

Una clase de sistema de información son los sistemas de apoyo a las decisiones (DSS) que son sistemas interactivos controlados por el usuario que combinan datos, modelos y herramientas de análisis avanzados, y software fácil de usar, en un solo sistema potente que puede apoyar la toma de decisiones. Los DSS permiten a los usuarios extraer información útil que antes estaba enterrada en grandes cantidades de datos.

Los sistemas de apoyo a la toma de decisiones se están volviendo una de las principales herramientas para que un administrador pueda realizar su tarea con mayor eficiencia. Las grandes empresas y negocios están haciendo uso de estos sistemas para tomar decisiones estratégicas

de tal forma que hace a las empresas más competitivas y a los negocios más rentables. Los sistemas de información para el apoyo a la toma de decisiones son los más complicados de desarrollar ya que su finalidad es encontrar información relevante para los administradores. Para el descubrimiento de información, estos sistemas hacen uso de técnicas computacionales de la inteligencia artificial como la minería de datos, la cual hace uso de sistemas expertos, razonamiento basado en casos y redes neuronales.

La toma de decisiones es el proceso mediante el cual se realiza una elección entre las alternativas o formas para resolver diferentes situaciones de la vida, éstas se pueden presentar en diferentes contextos: a nivel laboral, familiar, sentimental, es decir, en todo momento se toman decisiones, la diferencia entre cada una de estas es el proceso o la forma en la cual se llega a ellas [Huber(1986)]. La toma de decisiones consiste, básicamente, en elegir una alternativa entre las disponibles.

Las decisiones se pueden clasificar según la frecuencia con la que se presentan [Huber(1986)]:

- *Decisiones programadas o estructuradas*: Son aquellas que se toman frecuentemente, es decir son repetitivas y se convierte en una rutina tomarlas; como el tipo de problemas que resuelve se presentan con cierta regularidad, ya se tiene un método bien establecido de solución y por lo tanto ya se conocen los pasos para abordar este tipo de problemas. En estas decisiones la

persona que toma la decisión no tiene la necesidad de diseñar ninguna solución, sino que simplemente se rige a la que se ha seguido anteriormente.

- *Decisiones no programadas o no estructuradas*: Son decisiones que se toman en problemas o situaciones que se presentan con poca frecuencia, o aquellas que necesitan de un modelo o proceso específico de solución, por ejemplo: El lanzamiento de un nuevo producto al mercado; en este tipo de decisiones es necesario seguir un modelo de toma de decisión para generar una solución específica para este problema en común.

Las situaciones, ambientes o contextos en los cuales se toman las decisiones, se pueden clasificar según el conocimiento y control que se tenga sobre las variables que intervienen o influyen en el problema, ya que la decisión final o la solución que se tome va a estar condicionada por dichas variables [Joseph E. McCann(1984)].

- *Ambiente de certidumbre (certeza)*. Se tiene conocimiento total sobre el problema, las alternativas de solución que se planteen van a causar siempre resultados conocidos e invariables. Al tomar la decisión sólo se debe pensar en la alternativa que genere mayor beneficio.
- *Ambiente de riesgo*. La información con la que se cuenta para solucionar el problema es incompleta, es decir, se conoce el problema, se conocen las posibles soluciones, pero no se conoce con certeza los resultados que pueden arrojar. En este tipo de decisiones, las posibles alternativas de solución tienen cierta

probabilidad conocida de generar un resultado. En estos casos se pueden usar modelos matemáticos o también el decisor puede hacer uso de la probabilidad objetiva o subjetiva para estimar el posible resultado. La probabilidad objetiva es la posibilidad de que ocurra un resultado basándose en hechos concretos, puede ser cifras de años anteriores o estudios realizados para este fin. En la probabilidad subjetiva se determina el resultado basándose en opiniones y juicios personales.

- *Ambiente de incertidumbre.* Se posee información deficiente para tomar la decisión, no se tienen ningún control sobre la situación, no se conoce como puede variar o la interacción de las variables del problema, se pueden plantear diferentes alternativas de solución pero no se le puede asignar probabilidad a los resultados que arrojen.

Con base en lo anterior hay dos clases de incertidumbre [**Joseph E. McCann(1984)**]:

- *Estructurada:* No se sabe que puede pasar entre diferentes alternativas, pero sí se conoce que puede ocurrir entre varias posibilidades.
- *No estructurada:* No se sabe que puede ocurrir ni las probabilidades para las posibles soluciones, es decir no se tienen ni idea de que pueda pasar.

El ambiente al que se enfrenta una organización puede conceptualizarse de acuerdo a su estabilidad. Emery y Trist [**F.E. Emery(1965)**] fueron los primeros investigadores sobre este tema. Ellos proponen el concepto de turbulencia y su contrario, los ambientes placenteros. Turbulencia

es una medida del cambio que ocurre en los factores o componentes del ambiente sobre el cual trabaja una organización. Por un lado se tienen los ambientes placenteros, donde los cambios son estáticos (no existen cambios) y por el otro se tienen los ambientes turbulentos, donde los cambios en los factores están en constante flujo. La cantidad de turbulencia en el ambiente está relacionada con el grado de incertidumbre al que se enfrenta una organización. Conforme el ambiente se hace más turbulento, los factores se hacen cada vez más difíciles de predecir, por lo tanto, se hacen cada vez más inciertos; los valores de las variables que son de importancia se mueven de forma errática.

El uso principal de los pronósticos está relacionado con la toma de decisiones. Algunas son "pequeñas" decisiones, por ejemplo las relacionadas al mantenimiento de un adecuado nivel de inventario. Algunas son "grandes" decisiones, como las relacionadas con inversiones. *En todo caso los pronósticos para ser útiles deben estar relacionados con la toma de decisiones [Radford(1978)].* Para que un pronóstico sea útil se debe presentar con oportunidad.

El uso de modelos matemáticos han ido en incremento para interpretar y predecir las dinámicas y controles en la toma de decisiones gerenciales. Dichas aplicaciones incluyen pronóstico de ventas, predicciones del impacto y efecto de campañas publicitarias, estrategias para

proteger desabastecimiento de inventarios y para determinar estrategias óptimas de inversión de portafolios.

La mayoría de las decisiones gerenciales están basadas en pronósticos [James W. Dean(1996)]. Cada decisión se hace efectiva en algún punto en el futuro, por lo tanto deberían estar basadas en pronósticos de las condiciones futuras.

1. Problema a investigar

La presente investigación tiene como finalidad proponer una nueva forma de realizar pronósticos en ambientes de alta incertidumbre.

Usualmente se pide al administrador que pronostique el comportamiento de un evento basado en las experiencias pasadas. La mayoría de las aplicaciones tienen una o muchas variables que influyen en el análisis del posible evento o variable de salida. Al evento de salida que comúnmente se desea pronosticar se le llama variable dependiente y a las variables que influyen en la variable de salida se les conoce como variables independientes o variables de entrada. El análisis de regresión estudia la relación entre las variables dependientes e independientes. Generalmente se procede escogiendo un modelo al cual se tratan de acomodar las variables independientes. Si este modelo concuerda lo suficiente con los datos como para ser considerado útil, éste puede ser utilizado para pronosticar el valor de la variable dependiente de acuerdo a valores de las variables independientes.

El proceso de regresión es especificado por un conjunto de variables de entrada independientes x y una variable de salida dependiente y . El objetivo es aproximar y utilizando x y un conjunto de coeficientes w tal que [Keijzer(2002)],

$$y = f(x, w) + e$$

Donde e representa un error. A cada una de las variables explicativas recibe el nombre de variable independiente o exógena (x). Y a cada una de las variables explicadas recibe el nombre de variable dependiente o endógena (y). La función matemática que cuantifica la relación existente entre ambos tipos de variables se denomina función de regresión (f). Se llama diagrama de dispersión a la representación de los valores observados del evento. Es conveniente realizar un diagrama de dispersión ya que la forma del diagrama ofrece una idea inicial del tipo de relación existente entre x y y . En la figura 1 se observa un diagrama de dispersión en el cual no se observa una relación clara entre x y y . En la figura 2 se muestra una relación lineal entre las variable exógena y endógena, mientras que en la figura 3 se observa una relación parabólica (no lineal) entre la variable dependiente e independiente.

Los análisis de regresión se clasifican de acuerdo al número de variables, al tipo de función o al contexto del problema. La clasificación se muestra a continuación [P. Raymond(2000)]:

1. *En cuanto al número de variables dependientes (y).*

Regresión univariante: sólo una variable

Regresión multivariante: más de una variable

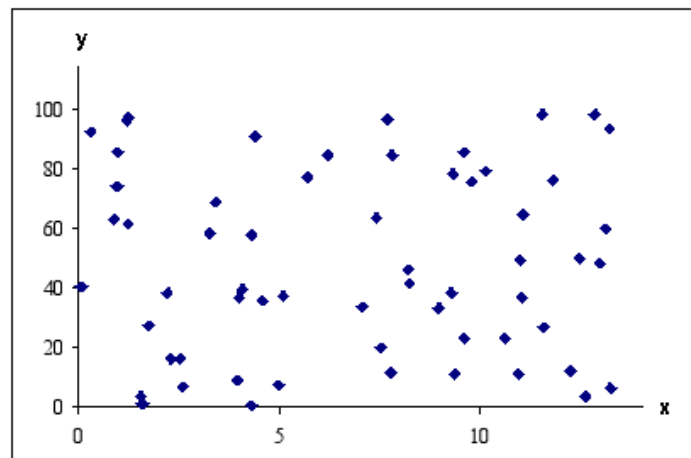


FIGURA 1. Diagrama de dispersión aparentemente sin relación entre las variables dependiente e independiente.

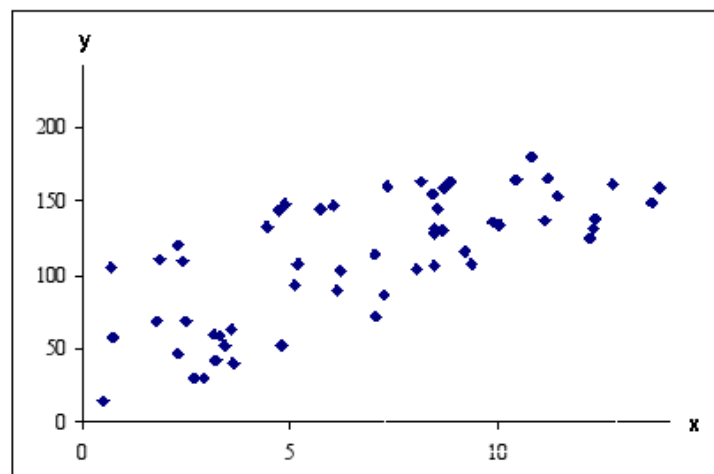


FIGURA 2. Diagrama de dispersión lineal.

2. *En cuanto al número de variables independientes (x).*

Regresión simple: sólo una variable

Regresión múltiple: más de una variable

3. *En cuanto a la forma matemática de la función de regresión (f):*

Regresión lineal: si dicha función es lineal

Regresión no lineal: si dicha función es no lineal

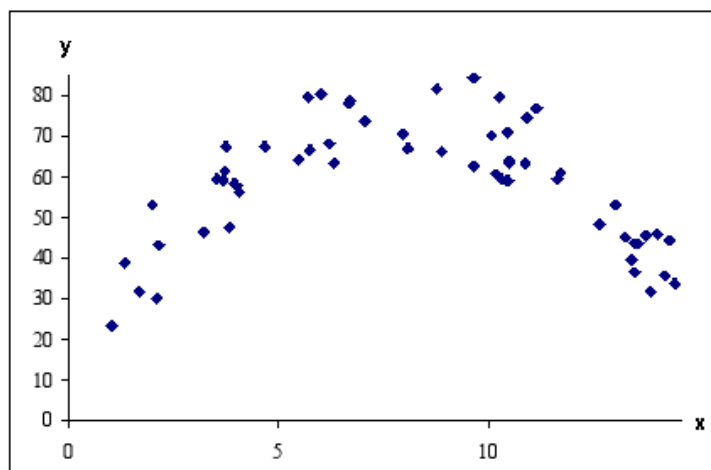


FIGURA 3. Diagrama de dispersión no lineal (parabólico).

4. *En cuanto al contexto temporal del problema*

Regresión estática

Regresión dinámica

En este trabajo se analizarán diferentes técnicas para realizar pronósticos con regresiones univariantes simples.

Debido a su simplicidad analítica, la forma funcional $f(x, w)$ que más se utiliza en la práctica es la relación lineal. Esta regresión [P. Raymond(2000)] es la forma más simple de análisis de regresión. En esta regresión, $f(x, w)$ debe ser de la forma:

$$f(x, w) = w_0 + w_1x_1 + \dots w_nx_n$$

Cuando sólo existe una variable independiente, esto se reduce a una línea recta, es decir, cuando

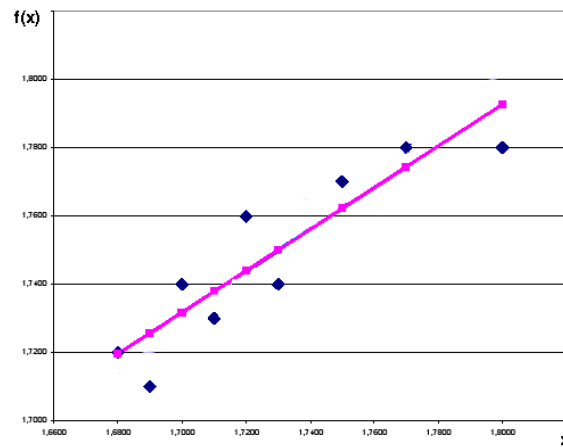


FIGURA 4. Regresión lineal sobre una serie de datos

$$f(x, w) = w_0 + w_1x$$

Donde w_1 (la pendiente) y w_0 (la intersección con el eje y) son constantes llamadas coeficientes de regresión lineal. Uno de los métodos más utilizados para determinar la línea es mediante mínimos cuadrados. A este tipo de regresión lineal se le conoce como regresión simple. En la figura 4 se observa un ejemplo de una serie de valores y la línea recta que mejor se aproxima a estos valores.

En el caso más general de la regresión múltiple, existen dos o más variables independientes.

$$f(x, w) = w_0 + w_1x_1 + \dots w_nx_n$$

La estimación de los coeficientes de una regresión múltiple es un cálculo bastante complicado y laborioso, por lo que se requiere del empleo de programas de computación especializados.

La regresión lineal no siempre da buenos resultados, porque a veces la relación entre x y y no es lineal sino que exhibe algún grado de curvatura. La estimación directa de los parámetros de funciones no lineales es un proceso bastante complicado. No obstante, a veces se pueden aplicar las técnicas de regresión lineal por medio de transformaciones de las variables originales.

Por ejemplo, una función no-lineal que tiene muchas aplicaciones es la función exponencial

$$y = Ax^w$$

donde A y w son constantes desconocidas. Si se aplican logaritmos, esta función también puede ser expresada como:

$$\log(y) = \log(A) + w\log(x)$$

Consideremos ahora la siguiente regresión lineal:

$$\log(y) = w_0 + w_1\log(x)$$

En esta regresión (denominada regresión doble-log), en lugar de calcular la regresión de y contra x , se calcula la regresión del logaritmo

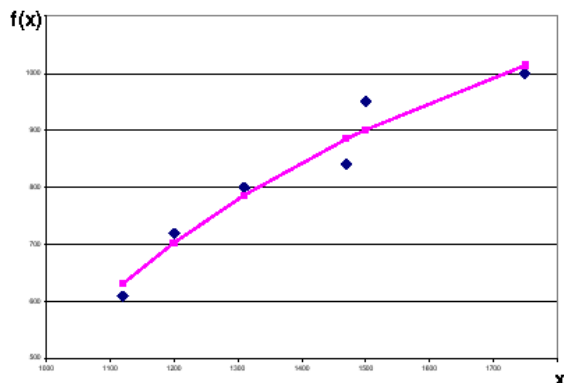


FIGURA 5. Regresión no lineal.

de y contra el logaritmo de x .

Otra forma de aproximar y es utilizando un modelo polinomial [P. Raymond(2000)]. Es decir, que los datos concuerden con un polinomio, donde el polinomio puede ser de cualquier grado. Si el polinomio es de primer grado, se habla de un modelo de línea recta, si es de segundo grado el modelo es una parábola, y así sucesivamente.

Para caso de una parábola, el modelo es,

$$f(x, w) = w_0 + w_1x + w_2x^2$$

Donde w_0 , w_1 , y w_2 son los coeficientes de regresión polinomial. Y en este caso también se suele utilizar el método de mínimos cuadrados. Un ejemplo de una curva parabólica que se aproxima a una serie de valores, se observa en la figura 5.

La regresión simbólica es el proceso mediante el cual se obtiene la función $f(x, w)$ representada por otras funciones básicas. Es decir, al contrario de las técnicas anteriores, la programación genética [Koza(1992)] aplicada a hacer una análisis de regresión, no utiliza una función previamente especificada, sino que hace uso de funciones básicas. Las cuales pueden ser combinadas para crear la función $f(x, w)$.

Por ejemplo, dado un conjunto de funciones básicas h_1, \dots, h_a , que toman un argumento y un conjunto de funciones g_1, \dots, g_b , que toman dos argumentos, la función $f(x, w)$ generada por programación genética puede tomar una gran variedad de formas. Las funciones h y g son generalmente funciones aritméticas como suma, resta, multiplicación y división, pero se pueden incluir funciones trigonométricas, lógicas, exponenciales, logarítmicas o seccionadas.

Un ejemplo de función podría ser:

$$f(x, w) = h_1(g_2(g_1(x_3, w_1), h_2(x_1)))$$

Si se sustituye cada uno de los símbolos h y g por las respectivas funciones básicas, este ejemplo puede llevar a una función f como:

$$f(x, w) = \text{sqr}t(*(+ (x_3, w_1), \text{exp}(x_1)))$$

Que puede ser traducida a la expresión

$$f(x, w) = \sqrt{((x_3 + w_1) * e^{x_1})}$$

Sin embargo, la programación genética no es el único método para realizar regresiones simbólicas. Algunas otras técnicas son: Multi Expressions Programming (MEP) [M. Oltean(2003)], Linear Genetic Programming (LGP) [Brameier(2004)], Gramatical Evolution (GE) [M. O'Neill(2001)], Gene Expression Programming (GEP) [Ferreira(2001)] y Evolving mathematical expressions (MEXE) [Grosan(2004)].

La ventaja que tiene la regresión simbólica sobre un análisis de regresión es que la forma de la función $f(x, w)$ tiene más riqueza en cuanto a su expresión, es decir, la función $f(x, w)$ está representada por funciones trigonométricas, lógicas, exponenciales entre otras, y no sólo como un polinomio (que es el caso de la regresión lineal y polinomial) *y no se debe tener un conocimiento previo de los datos para que la regresión simbólica obtenga la función.* Por otra parte, en la mayoría de las aplicaciones reales, no se tiene una relación lineal, aunque aparentemente existe una relación lineal como en la figura 4, en realidad no lo es. Un mejor ajuste se hace con un análisis de regresión no lineal, pero se tiene el inconveniente que es más complicado. La regresión simbólica permite hacer un ajuste de curva más exacto sin recurrir a la regresión no lineal. En la figura 6 se tiene una regresión lineal sobre un conjunto de datos (diagrama de dispersión). Para los mismos datos, en la figura 7 se observa una regresión no lineal (parabólica). En la figura 8 se muestra una regresión simbólica sobre el mismo diagrama de dispersión, en donde el error se minimiza lo más posible.

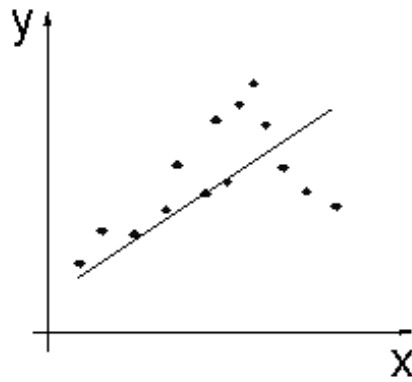


FIGURA 6. Aproximación lineal sobre un conjunto de datos cualquiera.

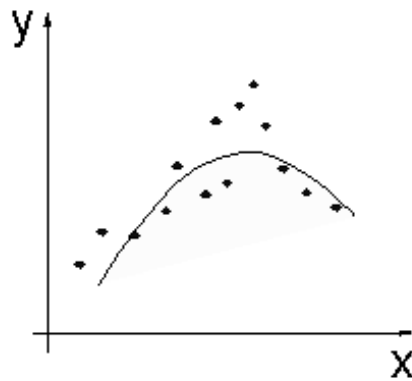


FIGURA 7. Regresión parabólica (no lineal) sobre el conjunto de datos de la figura 6.

En el presente trabajo se hace un análisis comparativo de las técnicas tradicionales más utilizadas para realizar pronósticos en contra del método no tradicional usando programación genética, para estudiar y mostrar los contextos en que cada una de las técnicas es más aplicable o recomendable.

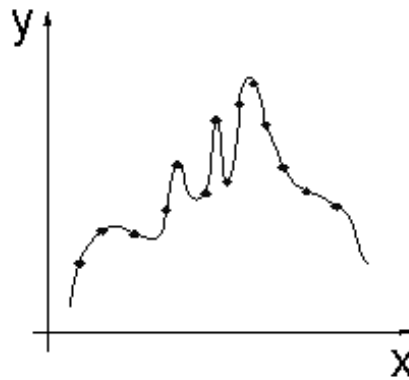


FIGURA 8. Regresión simbólica sobre el conjunto de datos de la figura 6 y 7.

Esta investigación da respuesta a preguntas como: ¿en qué casos (tipos de ambientes) se recomienda utilizar los métodos tradicionales?, en ambientes de alta incertidumbre ¿es posible obtener un pronóstico confiable mediante la programación genética?

2. Definición de términos

En la presente investigación se utilizará la siguiente terminología y variables:

- *Diagrama de dispersión.* Es el conjunto de datos que representan un ambiente, escenario, caso o contexto al cual es aplicable el pronóstico. Por ejemplo: ventas, acciones, futuros, compras, entre otros. El diagrama de dispersión se mide con respecto a un periodo de tiempo, por lo que también se le conoce como serie de tiempo. En esta investigación, otros sinónimos utilizados son ambiente, escenario y caso.

- *Ambiente de incertidumbre.* Es un diagrama de dispersión completamente aleatorio o caótico. En otras palabras, un ambiente que no tiene patrón alguno o tendencia alguna.
- *Periodo de tiempo.* Es una cantidad de tiempo fija, en la cual se hace una observación del fenómeno a pronosticar. Por ejemplo: cada segundo, cada día, cada bimestre o cada año.
- *Variable independiente.* Es la variable sobre la cual se realizará el experimento, en este caso es un escenario en el cual se realizará el pronóstico. Este escenario se manipulará intencionalmente para generar diferentes casos con determinadas características. Por ejemplo: un ambiente cuyo diagrama de dispersión tenga un tendencia lineal o uno sin tendencia.
- *Variable dependiente.* Es la variable que se desea medir. En este caso es el pronóstico para el siguiente periodo de tiempo del diagrama de dispersión.

3. Problemas y limitaciones

Dado que existen una infinidad de posibles ambientes (diagramas de dispersión), este trabajo se limitará a investigar sobre los siguiente casos:

- Se estudiaran los ambientes univariantes, es decir ambientes con una sola variable dependiente, como ventas, acciones o compras.

- También se trabajarán exclusivamente en ambientes simples, que únicamente consistan de una sola variable independiente. En este caso es el tiempo.
- Así mismo, se estudian ambientes lineales y no lineales. El ambiente de mayor interés, es un caso aleatorio, caótico, o de incertidumbre (un ambiente no lineal).

4. El proceso de la investigación

En el capítulo 2 se presenta el estado del arte en lo respectivo a la realización de pronósticos en la administración, mediante el análisis de series de tiempo. Se mencionan los diferentes enfoques para hacer pronósticos y se hace una breve explicación sobre las técnicas clásicas para realizar pronósticos. Dado que es fácil confundir los términos programación genética y algoritmos genéticos, se hace una breve explicación sobre qué es la programación genética y cuál es su diferencia con los algoritmos genéticos. Se explica por qué la programación genética puede ser una solución para realizar pronósticos y por qué los algoritmos no son una técnica aceptable.

En el capítulo 3 se describe el procedimiento realizado para la investigación. Así como la propuesta que se hace en esta investigación, además de cómo se diseñó el experimento para hacer la comprobación de la misma. Se describe cómo se seleccionó y analizó la muestra del experimento.

En el capítulo 4 se presentan el análisis de los resultados finales obtenidos por el experimento diseñado.

Finalmente, en el capítulo 5 se presentan las conclusiones finales obtenidas durante el desarrollo del presente trabajo.

Para la prueba de los diferentes casos, se realizó un software prototipo basado en programación genética. El uso de este software se presenta en el apéndice A.

CAPÍTULO 2

Marco teórico

La presente investigación tiene como finalidad proponer una nueva forma de realizar pronósticos en ambientes de alta incertidumbre. Para ello se realizará un estudio comparativo de las técnicas tradicionales en contra de la técnica propuesta, mostrando los contextos en que cada una es más recomendable. En esta nueva técnica se utilizará la programación genética que es un método utilizado en la inteligencia artificial para generar de forma automática programas computacionales. La programación genética se propone como el motor de búsqueda de ecuaciones para realizar el modelado de la información dentro del ambiente de alta incertidumbre.

Para entender mejor el problema a resolver en este trabajo, se hace una revisión sobre aquellas técnicas más utilizadas para realizar pronósticos. Así como la base teórica sobre la que se realiza la nueva técnica, que es la programación genética, la cual tiene sus orígenes en los algoritmos genéticos de la inteligencia artificial.

1. Pronósticos

Las herramientas para las decisiones tecnológicas tales como los modelos matemáticos han sido aplicados a una amplia gama de situaciones en la toma de decisiones dentro de diversas áreas de la gerencia. En la toma consciente de decisiones bajo incertidumbre, siempre realizamos pronósticos o predicciones. Podríamos pensar que no estamos pronosticando, pero nuestras opciones estarán dirigidas por la anticipación de resultados de nuestras acciones o inacciones.

El uso de modelos matemáticos han ido en incremento para interpretar y predecir las dinámicas y controles en la toma de decisiones gerenciales. Dichas aplicaciones incluyen pronóstico de ventas, predicciones del impacto y efecto de campañas publicitarias, estrategias para proteger desabastecimiento de inventarios y para determinar estrategias óptimas de inversión de portafolios.

La mayoría de las decisiones gerenciales están basadas en pronósticos. Cada decisión se hace efectiva en algún punto en el futuro, por lo tanto deberían estar basadas en pronósticos de las condiciones futuras.

Los pronósticos son necesarios en todas las áreas de una organización, y los mismos no deberían ser generadas por un grupo aislado de analistas. Tampoco se puede considerar ningún pronóstico como "terminado", los pronósticos son continuamente necesarios, y así como pasa el tiempo, el impacto de los pronósticos sobre el desempeño real

de acontecimientos debe ser medido. Los pronósticos originales son actualizados y las decisiones son modificadas, y así sucesivamente.

1.1. Enfoques de pronósticos. Hay dos enfoques generales al pronosticar, uno es el análisis cuantitativo y el otro el enfoque cualitativo [**Bures(1985)**]. Los pronósticos cuantitativos utilizan una variedad de modelos matemáticos que se apoyan en datos históricos o en variables causales para pronosticar la(s) variable(s) de interés. Los pronósticos cualitativos o subjetivos incorporan aquellos factores como la intuición, las emociones, las experiencias personales y el sistema de valores de quien toma la decisión para llegar al pronóstico.

1.1.1. Métodos cualitativos. Estos métodos reciben también el nombre de tecnológicos, porque históricamente se usaron primero para pronosticar cambios tecnológicos [**John Hanke(2002)**].

Las técnicas cualitativas se usan cuando los datos son escasos, por ejemplo cuando se introduce un producto nuevo al mercado. Estas técnicas usan el criterio de la persona y ciertas relaciones para transformar información cualitativa en estimados cuantitativos. Los métodos de uso más frecuente son: método delphi, estudios de mercados, consenso de un panel, pronóstico visionario y analogía histórica.

La posición central en estos métodos no la tienen los datos pasados, sino la experiencia de las personas. Frecuentemente se usa la experiencia y buen juicio de varios expertos. Hay 2 tipos de métodos [John Hanke(2002)]:

- *Exploratorios o Prospectivos*. Se parte de las experiencias pasadas y presentes para proyectar al futuro, sopesando las diferentes posibilidades.
- *Normativos o Deductivos*. En éstos se procede al revés, se parte de las metas u objetivos a lograr en el futuro y se analiza qué se necesita para lograrlos y eso nos da pauta para pensar cuando sucederán los eventos previstos para el futuro.

Un método tecnológico muy usado es el método Delphi; éste trata de obtener un consenso confiable entre diversos expertos para usarlo como base para pronosticar.

Al reunir a varios expertos para que emitan sus opiniones hay factores psicológicos que afectan al consenso. Algunos expertos pueden tener mayor claridad en sus exposiciones, ser más persuasivos, o mejores polemistas que otros, sin que tengan mayor razón. Por esto el método Delphi funciona evitando que los expertos se reúnan, toda la comunicación se hace mediante un coordinador.

Delphi necesita un grupo de expertos que estén dispuestos a contestar una serie de preguntas, y exponer sus razones, respecto a algún

desarrollo tecnológico, por ejemplo. El método Delphi funciona por rondas. Para ver como es esto, pongamos un ejemplo. Se hizo un estudio Delphi para saber cuáles eran los inventos y avances tecnológicos que se iban a dar en los 20 años siguientes. El estudio tuvo 4 rondas.

Primera ronda. Mediante una carta se le pidió a los miembros del panel una lista de los inventos y avances científicos que fueran a la vez útiles y factibles en los próximos 20 años. El coordinador, después de recibir las listas, hizo una con los 50 avances más mencionados.

Segunda ronda. La lista de 50 fue enviada a los expertos para que los acomodaran temporalmente (en 4 periodos de 5 años). Tomando como base el tiempo en que creyeran que había una probabilidad de 50 % de que se realice el avance.

Tercera ronda. El coordinador envió a los expertos dos listas:

- la lista de los avances en los que hubo consenso.
- la lista donde no hubo consenso indicándoles las medianas de los tiempos para los avances en que no hubo acuerdo.

Se les pidió que reconsideraran sus opiniones respecto a los avances en que no hubo consenso. A los expertos que difirieron mucho de los demás se les hizo notar. (Muchos de ellos cambiaron sus estimaciones).

Cuarta ronda. Se repitió la tercera ronda para "cerrar" más las opiniones de los expertos. El coordinador, elaboró un informe final; en este informe se obtuvo no sólo una lista de los avances que el panel de expertos consideró como alcanzables sino una estimación de los tiempos en que se van a alcanzar.

El método Delphi tiene las ventajas siguientes:

- queda documentado no sólo el resultado sino el proceso que se siguió.
- los expertos interactúan en forma anónima.
- se evitan divagaciones.

Las dificultades son:

- el coordinador debe permanecer "neutral" respecto a la discusión.
- puede haber dificultad en captar la atención de los expertos.
- gracias a la tecnología es posible acelerar la lentitud que va de la mano del correo.
- muchas veces las opiniones "delatan" al experto, dificultando el anonimato.

1.1.2. Métodos cuantitativos. Los métodos cuantitativos se basan en datos históricos [**John Hanke(2002)**]. Esta información pasada se encuentra en forma numérica. Las fuentes usuales son los registros de

la propia empresa o información oficial de diverso origen: gobierno, asociaciones de empresarios o profesionistas, organismos internacionales.

Se debe tener cuidado, sobre todo cuando la información proviene de la propia empresa (aunque en la proveniente de otras fuentes también hay que cuidarse), que haya sido cuantificada de manera uniforme. Para información sobre costos, por ejemplo, hay que asegurarse que los costos incluyan los mismos conceptos en todos los años que vamos a utilizar; de no ser así es preciso tratar previamente los datos.

Para aplicar los métodos cuantitativos es preciso convencernos, razonablemente, de que se cumple la llamada Hipótesis de Continuidad. Este supuesto es que los factores externos en los que se dieron los datos históricos no cambiarán en el futuro para el que estamos pronosticando. Estos factores son, en forma destacada:

- Economía en general.
- Competencia en el mercado (oferta).
- Estado del mercado (demanda).
- Estado tecnológico del producto ("ciclo de vida del producto").

Esta continuidad del ambiente nunca se da en forma perfecta, sino en forma gradual. Se requiere buen juicio para suponer que las violaciones a la continuidad no van a afectar a los resultados de la aplicación del método de pronóstico.

Los métodos cuantitativos son de dos tipos según la información en que se basan:

- *Métodos de Series de Tiempo*. Se usa información de la misma variable que se va a pronosticar.
- *Métodos Causales*. Se utiliza información de la variable que se va a pronosticar y de otras variables que influyen en ella o que están relacionadas con ella y cuyo pronóstico sea más simple.

1.2. Pronósticos de series de tiempo. Una serie de tiempo se basa en una secuencia de datos puntuales separados a intervalos iguales (semanas, meses, trimestres, etc.) [Barry Render(2004)]. Los datos de series de tiempo para pronósticos implican que los valores futuros se predicen solamente a partir de los valores pasados, y que se pueden ignorar otras variables, sin importar qué tan potencialmente valiosas sean.

1.2.1. Patrones de una serie de tiempo. Cuando se tienen datos para hacer un pronóstico, la herramienta más útil es el diagrama de dispersión. La gráfica que queremos es la de los datos contra el tiempo. En el eje horizontal ponemos los tiempos y en el sentido vertical señalamos el punto cuya altura corresponda a la magnitud de la observación que tengamos para cada tiempo. Por regla general, los datos se encuentran equiespaciados en el tiempo. Las diferentes formas que toma el arreglo de los datos en la gráfica nos indican como debemos proceder en el pronóstico.

Las características que, de manera primordial, buscamos en la gráfica son las regularidades que permitan la proyección del comportamiento observado en el pasado hacia el futuro. Los patrones regulares que nos son útiles son de varios tipos [**John Hanke(2002)**].

- *Datos horizontales.* Se presentan como un valor constante (recta horizontal) alrededor del cual los datos oscilan de forma irregular. Es el patrón de datos mas simple, la mejor manera de pronosticar en una situación como ésta es estimar la altura de la línea horizontal y usar ese valor como pronóstico.
- *Datos con tendencia.* Se presentan como una línea lisa (una recta o una curva suave) que sube o baja monóticamente y los datos oscilan erráticamente alrededor de ella. La manera de pronosticar que se ocurre primero, en este caso, es la de calcular una ecuación para la línea y usar ese valor para pronóstico.
- *Datos estacionales.* Muchas series de datos presentan este tipo de comportamiento repetitivo. Los datos se muestran como desviaciones erráticas alrededor de una senoide. El origen del nombre estacional son, precisamente las estaciones del año. Mucha de la actividad humana y muchos fenómenos naturales varían de acuerdo a las estaciones. Por extensión, en muchas actividades se presenta una oscilación semanal o mensual similar a la de las estaciones del año. Por ejemplo, no es raro observar que en algunos días de la semana se incrementa el ausentismo laboral. Tenemos otro ejemplo en la cantidad de transacciones que se realizan en las oficinas bancarias, estas

presentan dos "picos" mensuales, al principio/fin y al medio. Cuando se estudia una serie con esta característica, es deseable incorporarla al pronóstico.

- *Datos cíclicos.* Este se refiere a curvaturas de largo período asociadas con grandes ciclos económicos. El pronóstico en estas condiciones es mucho más complicado ya que la forma de estos ciclos no es simple y la teoría económica no se encuentra suficientemente desarrollada como para permitir una cuantificación confiable de ellos. Claro que si observamos tal patrón en los datos, es conveniente incorporarlo al pronóstico aún cuando sea de una manera imperfecta.

Dentro de los métodos de análisis de series de tiempo existen dos tipos generales:

- Aquellos métodos que tratan de encontrar el patrón total de los datos para proyectarlo al futuro (Promedios móviles, suavización exponencial y box Jenkinx)
- Aquellos métodos que consisten en separar la serie de tiempo en sus componentes, identificar el patrón de cada componente, para después hacer el pronóstico con la integración de los patrones (descomposición de series de tiempo)

1.2.2. Método de promedios móviles. El método de promedios móviles usa un número de valores de datos históricos reales para generar un pronóstico. Los promedios móviles son útiles si podemos suponer que

la demanda del mercado permanecerá relativamente estable en el tiempo. Un promedio móvil de 4 meses se encuentra simplemente sumando la demanda de los últimos 4 meses y dividiéndola entre cuatro. Al concluir cada mes, los datos del mes más reciente se agregan a la suma de los 3 meses anteriores y se elimina el dato del mes más antiguo. Esta práctica tiende a suavizar las irregularidades del corto plazo en las series de datos.

Matemáticamente, el promedio móvil simple (que sirve como estimación de la demanda del siguiente periodo) se expresa como

$$P_{t+1} = \frac{(X_t + X_{t-1} + X_{t-2} + \dots + X_{t-n+1})}{n}$$

en donde

P_{t+1} = Pronóstico para el siguiente periodo

X_t = Valor de la serie en el periodo t

n = Número de observaciones incluidos en el promedio móvil

Los promedios móviles son efectivos para suavizar las fluctuaciones repentinas en el patrón de la demanda, con el fin de obtener estimaciones estables. Sin embargo, los promedios móviles presentan tres problemas:

1. Aumentar el tamaño de n (el número de periodos promediados) si bien permite suavizar mejor las fluctuaciones, también resta sensibilidad al método ante los cambios reales en los datos.

2. Los promedios móviles no reflejan muy bien las tendencias. Puesto que son promedios, siempre se quedarán en niveles pasados, no predicen los cambios hacia niveles más altos ni más bajos. Es decir, retrasan los valores reales.
3. Los promedios móviles requieren amplios registros de datos históricos.

1.2.3. Método de suavizamiento exponencial. El suavizamiento exponencial es un sofisticado método de pronóstico de promedios móviles ponderados cuya aplicación sigue siendo muy sencilla. Implica mantener muy pocos registros de datos históricos. La fórmula básica para el suavizamiento exponencial se expresa como sigue:

Nuevo pronóstico = pronóstico del periodo anterior + α (demanda real en mes anterior - pronóstico del periodo anterior)

donde α es la ponderación, o constante de suavizado, elegida por quien pronostica, que tiene un valor entre 0 y 1.

$$F_t = F_{t-1} + \alpha(A_{t-1} - F_{t-1})$$

donde

F_t = nuevo pronóstico

F_{t-1} = pronóstico anterior

α = constante de suavizado (o ponderación) ($0 \leq \alpha \leq 1$)

A_{t-1} = demanda real en el periodo anterior

La constante de suavizado, α , se encuentra generalmente en un intervalo de 0.05 a 0.50 para aplicaciones comerciales. Puede cambiarse para dar más peso a los datos recientes (si α es alta) o más peso a los datos anteriores (si α es baja). Cuando α llega al extremo de 1.0, entonces en el pronóstico se vuelve un modelo intuitivo.

La ventaja de este método con relación a promedios móviles es que requiere mucho menor espacio en memoria cuando se usa una computadora para los cálculos y la desventaja es que la única forma para calcular el valor de α que produce el menor error de pronóstico es por prueba y error.

Hay muchas variaciones de los métodos de suavización exponencial: algunos son más versátiles que otros, requieren más tiempo de cómputo o son más complejos computacionalmente. Una versión de suavización exponencial (el método de Winters) puede manejar una serie de tiempo que presente variación estacional, cosa que no pueden hacer los métodos de promedios móviles.

1.2.4. Método de Box-Jenkins. El método de suavización exponencial es un caso especial de la técnica de Box-Jenkins. En este método a la serie de tiempo se le ajusta un modelo matemático que es óptimo en el sentido de que asigna errores más pequeños que cualquier otro método a los datos históricos. El tipo de modelo que va a usarse debe

identificarse y posteriormente calcular sus parámetros.

Esta es una de las técnicas más exactas que se usan actualmente, pero también es una de las más complejas matemáticamente y que consumen mucho tiempo de computadora. Esto significa que es una técnica cara, tanto por el tiempo de computadora requerido, como por la necesidad de personal altamente especializado.

1.2.5. Descomposición de series de tiempo. Este método consiste en romper una serie de tiempo en sus componentes. Cada componente se analiza por separado en un intento de aislar el patrón de comportamiento y su influencia resultante sobre la variable que se va a pronosticar.

Debido a que a menudo se encuentra que cada componente se comporta en forma diferente a través del tiempo, en muchos casos el hacer análisis separados de cada uno de esos componentes aumenta la exactitud del pronóstico.

La técnica general para descomponer series de tiempo es:

1. Determinar el componente estacional y eliminar su efecto de los datos originales (desestacionalizar la serie).
2. Calcular la tendencia y eliminar su efecto de los datos desestacionalizados.

3. Analizar el componente cíclico.

1.3. Pronósticos en modelos causales. Los modelos causales asumen que la variable que se va a pronosticar guarda una relación causa-efecto con una o más variables independientes. El propósito de los modelos causales es descubrir la forma de la relación entre las variables y usarla para predecir valores futuros de la variable dependiente.

Los modelos de más uso son los modelos de regresión, modelos econométricos, encuestas de intención de compra y anticipación, modelo de insumo-producto y análisis del ciclo de vida.

1.3.1. Modelos de regresión. En estos modelos se trata de encontrar la relación que existe entre una variable dependiente (generalmente ventas) y una o más variables independientes [Maddala(1996)].

1.3.2. Modelos econométricos. Un modelo econométrico es un sistema de ecuaciones de regresión interdependientes que describe algún sector de actividades econométricas, ventas o utilidades [Greene(1998)]. Los parámetros de las ecuaciones de regresión generalmente se estiman simultáneamente.

Como regla general estos modelos son relativamente caros de desarrollar y de mantener. Sin embargo, debido al sistema de ecuaciones inherentes en estos modelos, pueden expresar mejor que las ecuaciones

de regresión cualquier contingencia, y por lo tanto son mucho más exactos para predecir punto de cambio.

1.3.3. Encuestas de intenciones de compra y anticipaciones. Estas encuestas que se hacen al público en general determinan:

- Las intenciones de compra de ciertos productos.
- Derivan un índice que mide el sentimiento general sobre el consumo presente y futuro y estimen como afectarán estos sentimientos o los hábitos de compra.

Este enfoque para hacer pronósticos es más útil que otras técnicas para seguir el desarrollo de la demanda y para señalar puntos de peligro. El problema principal al usuario es que puede señalar en forma incorrecta los puntos de cambio.

1.4. Medición del error en el pronóstico. La precisión general de cualquier modelo de pronóstico - promedios móviles, suavizamiento exponencial o cualquier otro - se determina comparando los valores pronosticados con los valores reales u observados [**Barry Render(2004)**]. Si F_t denota el pronóstico en el periodo t , y A_t denota la demanda real del periodo t , el error de pronóstico (o desviación) se define como

Error del pronóstico = demanda real - valor pronosticado

$$\text{Error del pronóstico} = A_t - F_t$$

Existen varias medidas de uso común en la práctica para calcular el error global del pronóstico. Estas medidas sirven para comparar distintos modelos de pronóstico, así como para vigilar los pronósticos y asegurar su buen desempeño. Las medidas más comunes son desviación absoluta media, DAM (mean absolute deviation, MAD) y error cuadrático medio, ECM (mean squared error, MSE).

1.4.1. Desviación absoluta media. La primera medición del error global del pronóstico para un modelo es la desviación absoluta media (DAM). Su valor se calcula sumando los valores absolutos de los errores individuales del pronóstico y dividiendo entre el número de periodos de datos (n):

$$DAM = \frac{\sum |real - pronostico|}{n}$$

1.4.2. Error cuadrático medio. El error cuadrático medio es una segunda forma de medir el error global del pronóstico. El ECM es el promedio de los cuadrados de las diferencias entre los valores pronosticados y observados. Su fórmula es

$$ECM = \frac{\sum (real - pronostico)^2}{n}$$

Una desventaja de emplear ECM es que tiende a acentuar las desviaciones importantes debido al término al cuadrado. Por ejemplo, si el error de pronóstico para el periodo 1 es dos veces más grande que el error para el periodo 2, entonces el error al cuadrado en el periodo 1 es cuatro veces más grande que el del periodo 2. Por lo tanto, el empleo del ECM como medición del error de pronóstico usualmente indica

que se prefiere tener varias desviaciones pequeñas en lugar de una sola desviación grande.

2. Inteligencia Artificial

Las aplicaciones de inteligencia artificial son especialmente útiles para los sistemas de información porque se concentran en ayudar a encontrar información relevante dentro de la empresa y así ayudar a los administradores a tomar decisiones.

En una conferencia en Dartmouth College en 1956, John McCarthy propuso el uso del término inteligencia artificial para describir a las computadoras con la capacidad de imitar o duplicar las funciones del cerebro humano.

El comportamiento inteligente abarca varias características entre las que se incluyen las capacidades para: aprender de la experiencia y aplicar estos conocimientos a nuevas experiencias; manejar situaciones complejas y solucionar problemas para los que quizá se carezca de información; determinar la información importante en una situación determinada; pensar en una forma lógica y racional y proporcionar una respuesta rápida y correcta, y comprender imágenes visuales y procesar símbolos.

La computadora es más eficiente que los seres humanos en la transferencia de información, en la realización de una serie de cálculos con

rapidez y exactitud y hacer cálculos complejos, pero el ser humano es más eficiente que una computadora en todos los demás atributos de inteligencia. Algunas técnicas utilizadas en la inteligencia artificial son los sistemas expertos, el razonamiento basado en casos, la lógica difusa y los algoritmos genéticos.

Los sistemas expertos son programas de computadora que hacen uso intensivo de conocimientos y resuelven problemas que hasta ahora requerían saber y experiencia humanos. Los sistemas capturan un dominio limitado del conocimiento humano, mediante el uso de reglas o marcos. Los sistemas expertos son útiles sobre todo en problemas de clasificación o diagnóstico.

El razonamiento basado en casos representa el conocimiento de la organización como una base de datos de casos que se puede ampliar y refinar continuamente. Cuando el usuario se topa con un caso nuevo, el sistema busca casos similares, encuentra el que más se le parezca y aplica las soluciones de ese caso al caso nuevo. Este último se guarda, junto con las soluciones en la base de datos de casos.

Las redes neuronales consisten en hardware y software que intentan imitar los procesos de razonamiento del cerebro humano. Las redes neuronales son notables por su capacidad para aprender sin programación y para reconocer patrones que las personas no pueden describir fácilmente. Estas redes se están usando en la ciencia, la medicina y los

negocios, primordialmente para distinguir patrones en grandes cantidades de datos.

La lógica difusa es una tecnología de software que expresa, con cierta imprecisión cuidadosamente definida, la lógica que está más cercana a la forma en que la gente realmente piensa. Se ha empleado lógica difusa para controlar dispositivos físicos, y se está comenzando a usar en aplicaciones limitadas de toma de decisiones.

Los algoritmos genéticos desarrollan soluciones a problemas específicos, utilizando procesos con base genética, como aptitud, cruce y mutación para crear soluciones. Estos algoritmos se comienzan a aplicar a problemas que implican optimización, diseño de productos y monitoreo de sistemas industriales.

La programación genética es un tipo de algoritmo genético pero que sirve para generar programas computacionales para resolver un problema sin tener que decirle a la computadora lo que tiene que hacer. Una de las posibles aplicaciones de la programación genética es para hacer regresiones simbólicas como lo propone Koza en [Koza(1992)] y en aplicaciones financieras como en [Chidambaran(2003)] y [D. Rivero(2005)].

2.1. Algoritmos genéticos y programación genética. Los algoritmos genéticos son procedimientos computacionales de búsqueda utilizados para resolver de forma adecuada problemas bien definidos.

Estos procedimientos fueron desarrollados en los algoritmos genéticos por Holland (1975) [**Holland(1975)**] y posteriormente en la programación genética desarrollada por Koza (1992) [**Koza(1992)**]. Las principales características de ambos algoritmos incluyen: (i) una forma de representar las posibles soluciones a un problema y (ii) una forma de medir la "calidad" de la solución por un candidato a resolver el problema. Las dos clases de algoritmos producen "generaciones" de posibles candidatos que puedan resolver el problema. Estas generaciones se producen de forma sucesiva utilizando procedimientos que simulan los procesos de reproducción y recombinación genética. Cada generación es sometida a un proceso de "selección natural", en el cual, sólo las mejores soluciones son escogidas para reproducirse.

Para entender los principios involucrados en la programación genética es necesario entender la operación de un algoritmo genético simple.

2.2. Algoritmo Genético. Los algoritmos genéticos simulan el proceso de evolución descrito por Charles Darwin en *On the origin of species by means of natural selection* [**Darwin(2002)**]. Un algoritmo genético es una serie de instrucciones computacionales que transforman un conjunto (población) de individuos (genes), cada uno asociado a un valor de aptitud, en una nueva población (siguiente generación). Para ello se vale de operaciones descritas por Darwin sobre reproducción y sobre vivencia del más apto (cruza y mutación).

El algoritmo genético resuelve problemas al ir buscando en un espacio de posibles soluciones para el problema en particular. Cada una de estas posibles soluciones deben ser representadas de una forma tal que el algoritmo genético lo pueda procesar para crear una nueva generación y obtener la mejor solución [Zbigniew(1999)], [Spears(2000)].

Los algoritmos genéticos requieren que las soluciones potenciales a un problema específico sean representados como una cadena de caracteres de un tamaño fijo. Si se considera un problema en el que las soluciones candidatas son representadas en una cadena caracteres de forma binaria, es decir, formada por unos y ceros. Y dicha cadena tiene una longitud de cinco dígitos. Una posible solución puede estar representada como (01010). Asociada a esta cadena de dígitos binarios estaría una medida de aptitud que cuantifica qué tan bien resuelve el problema. En otras palabras se necesita un función de aptitud que evalúe todas las posibles soluciones al problema (representadas por cadenas de dígitos binarios) y las clasifique por orden de calidad de la solución. Después se introduce el operador de cruza. Dadas dos cadenas, se escoge de forma aleatoria un punto de cruza y la primera parte de una de las cadenas es combinada con la segunda parte de la otra. Por ejemplo, dadas las dos cadenas (00101) y (11010) y un punto de cruza entre los elementos dos y tres, la nueva cadena (00010) es generada. Las partes restantes de las cadenas originales son desechadas.

El algoritmo empieza generando aleatoriamente la población inicial de cadenas binarias y la evaluación de aptitud de cada cadena al aplicar la función de aptitud. Después, el programa produce una nueva (segunda) generación de posibles soluciones al seleccionar pares de cadenas al azar de la población inicial y aplicando el operador de cruza para crear nuevas cadenas. La probabilidad de seleccionar una cadena de caracteres de la población, es directamente proporcional a su aptitud, es decir, qué tan bien resuelve el problema. Este proceso se repite para producir de forma sucesiva las generaciones de cadenas, procurando dejar la población del mismo tamaño. El procedimiento "evoluciona" nuevas generaciones de soluciones potenciales al problema.

Antes de ejecutar un algoritmo genético es necesario realizar los siguientes pasos:

1. Determinar la representación para cada individuo.
2. Determinar la función de aptitud.
3. Determinar los parámetros y variables que controlan el algoritmo
4. Determinar la condición de terminación

La representación del individuo especifica la transformación que aparee un punto en el espacio de búsqueda de soluciones del problema a una cadena de caracteres y la transformación inversa que aparee cada cadena de caracteres a un punto en el espacio de búsqueda. Escoger una representación que facilite la solución a un problema, requiere

conocimiento del problema en sí.

La función de aptitud se encarga de asignar un valor numérico a cada individuo de la población, de acuerdo a que tan cercano es el individuo a la solución del problema. Dicho valor de aptitud es importante, ya que es el que permite seleccionar individuos que sobrevivirán y se reproducirán para la siguiente generación.

Los principales parámetros a determinar para ejecutar el algoritmo genético son: el tamaño de la población (M), el máximo número de generaciones (G), la probabilidad de cruce (p_c) y la probabilidad de mutación (p_m).

Cada ejecución de un algoritmo genético necesita de una condición de terminación, la cual está dada cuando se alcanza el máximo número de generaciones (G) o se satisface alguna condición específica al problema.

Una vez entendido el funcionamiento de un algoritmo genético, se presenta en la siguiente sección los conceptos principales sobre programación genética.

2.3. Programación genética. La programación genética fue concebida por John Koza a principios de los noventa. Los algoritmos genéticos requieren que las soluciones potenciales deben ser codificadas

como cadenas de caracteres. La aportación de Koza fue que en lugar de utilizar cadenas de caracteres usa cadenas acomodadas de forma jerárquica que pueden ser pensadas como árboles de decisión o programas computacionales. Sin embargo, la estructura básica de un programa genético es la misma que para un algoritmo genético. En otras palabras, la programación genética es una extensión de los algoritmos genéticos, en donde la población son programas de computadoras .

La programación genética permite dar una respuesta al problema de sobre como se puede hacer para que una computadora aprenda a resolver problemas sin ser explícitamente programada para resolver el problema [**W. Banzhaf(2001)**].

De acuerdo a Koza, para formar un programa computacional es necesario definir un conjunto de funciones (nodos internos del árbol) $F = f_1, f_2, \dots, f_N$ y un conjunto de variables terminales (nodos terminales u hojas del árbol) $T = t_1, t_2, \dots, t_n$.

Las funciones, también llamadas operadores, dentro del conjunto de funciones pueden incluir:

- Operaciones aritméticas (+, -, *, /, etc.)
- Funciones matemáticas (seno, coseno, exponencial, logaritmo, etc.)
- Operaciones lógicas (disyunción, conjunción , negación, etc.)
- Operadores condicionales (si-entonces)
- Funciones de iteración (hacer-hasta, mientras, etc.)

- Funciones recursivas

Las variables terminales pueden ser de dos tipos: constantes y variables, éstas últimas representan las entradas del problema específico.

CAPÍTULO 3

Procedimiento de investigación

1. Planteamiento del problema

Dentro de la variedad de temas que abarca la Econometría, destacan las técnicas y modelos que utilizan información fechada (series temporales), con las que se obtienen predicciones y simulaciones que sirven de soporte al proceso de toma de decisiones económicas y empresariales.

Una serie temporal, o serie cronológica, es una sucesión de valores observados de una variable referidos a momentos o a períodos de tiempo diferentes, generalmente regulares. La característica distintiva de una serie temporal, en contraposición a las observaciones de corte transversal, es que los datos aparecen ordenados cronológicamente.

Hay series temporales que, por su propia naturaleza, resultan prácticamente impredecibles: son series "aleatorias". Un ejemplo típico es el de la serie de números premiados en la Lotería Nacional. En el extremo opuesto, existen series temporales cuyo comportamiento es tan regular que cabe realizar con ellas pronósticos muy precisos. Por ejemplo, el horario de las mareas, la posición de los astros, etc. La mayoría de las series temporales económicas no pertenecen ni a un caso ni al otro, sino que contienen en su comportamiento una componente regular y otra

irregular.

El análisis de series temporales puede perseguir los dos fines siguientes:

- La descripción de una serie temporal
- El pronóstico de la evolución futura de una serie temporal

Para llevar a cabo tales objetivos se adopta el denominado enfoque univariante, que hace uso de los datos históricos de la variable (serie temporal) para elaborar un modelo que describa aceptablemente el comportamiento de la misma en el pasado. Tal modelo descriptivo puede usarse para realizar pronósticos de los valores futuros de la variable. Se pueden utilizar bajo este enfoque dos tipos de modelos:

- *Clásico*. En donde la componente una pauta de comportamiento sistemática, de la misma naturaleza que los errores de observación, es decir, si un hecho aislado incide en una variable, es incidencia se supone errática y sin ninguna repercusión en los cambios futuros
- *Moderna*. También llama estocástica, y en donde se toma en cuenta la componente irregular

En la actualidad se debido a los cambios en las economías, cada vez más globalizadas, son ambientes más irregulares.

2. Propuesta

La programación genética es una técnica de computación que permite realizar de forma eficiente pronósticos sobre una serie de datos que representen un ambiente de incertidumbre.

En otras palabras, para realizar pronósticos que ayuden a la toma de decisiones en las organizaciones cuyo ambientes no tengan alguna tendencia, la programación genética puede ayudar a hacer pronósticos de forma eficiente.

3. Procedimientos

Con la finalidad de comprobar la propuesta planteada, es necesario hacer una comparación entre las técnicas convencionales modernas para realizar pronósticos y las técnica propuesta basada en programación genética.

Las técnicas convencionales a usar en este trabajo de investigación, son las presentadas en el capítulo de marco teórico. Estos métodos son: promedio móvil, promedio móvil ponderado, suavizamiento exponencial y regresión lineal.

Estas técnicas se usaran para obtener el pronóstico sobre diferentes series de datos, los cuales representan diferentes tipos de ambientes.

Se medirá el error cuadrático medio del pronóstico obtenido para cada técnica y se comparará para el método propuesto.

3.1. Selección de la muestra. Para realizar el estudio comparativo se generarán diferentes casos de prueba que representen diferentes ambientes. Estos escenarios serán con tendencia lineal (figura 2), tendencia exponencial o parabólica (no lineal - figura 3), tendencia estacionaria, tendencia estacionaria con linealidad ascendente y sin tendencia (aleatorio -figura 1).

Después se explorará la utilidad del método propuesto con dos casos reales, cuyo comportamiento sea de difícil previsión.

3.2. Recolección de los datos. Los escenarios anteriores se obtendrán a partir de las siguientes series pseudo caóticas,

ecuación: 1; $r_t = 4r_{t-1}(1 - r_{t-1})$; figura 1.

ecuación: 2; $r_t = 1 + 0.3r_{t-2} - 1.4r_{t-1}^2$; figura 2.

ecuación: 3: $r_t = \begin{cases} 2r_{t-1}, & 0 \leq r_{t-1} < 0.5 \\ 2(1 - r_{t-1}), & 0.5 \leq r_{t-1} \leq 1 \end{cases}$; figura 3.

ecuación: 4; $r_t = 4r_{t-1}^3 - 3r_{t-1}$; figura 4.

ecuación: 5: $r_t = 8r_{t-1}^4 - 8r_{t-1}^2 + 1$; figura 5.

Cada una de estas series será montada en una línea recta con pendiente positiva para generar un ambiente con tendencia lineal ascendente. Para obtener el ambiente no lineal, a las series se le sumará una línea parabólica . El ambiente estacional se obtiene al montar la serie

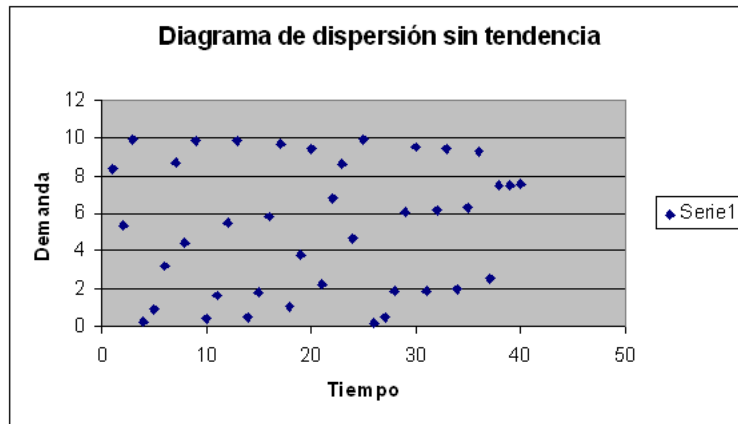


FIGURA 1. Serie de datos seudo caótica generada por la ecuación 1.

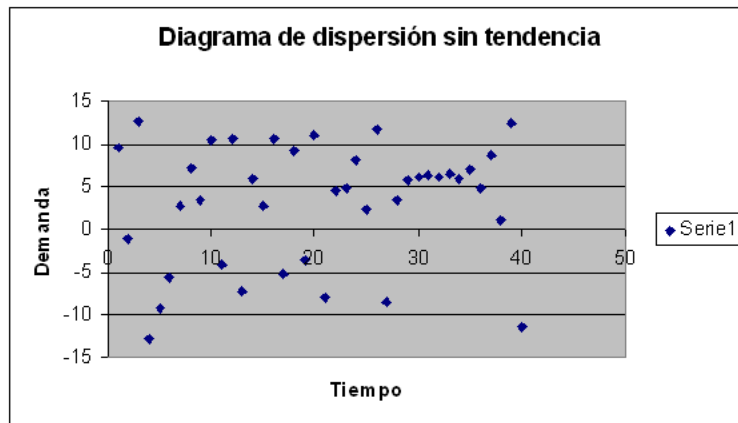


FIGURA 2. Serie de datos seudo caótica generada por la ecuación 2.

seudo caótica en una señal cosenoidal y finalmente para obtener el ambiente estacional ascendente se sumará una línea recta más una señal cosenoidal a la serie caótica.

Finalmente, se comprobarán estos resultados en dos problemas de la vida real como casos de estudio. El primer caso de estudio es el índice NASDAQ 6, el cual tiene como característica principal ser un ambiente

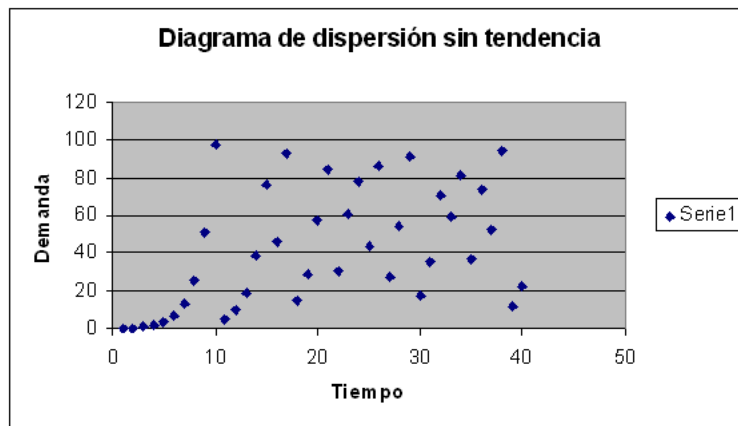


FIGURA 3. Serie de datos seudo caótica generada por la ecuación 3.

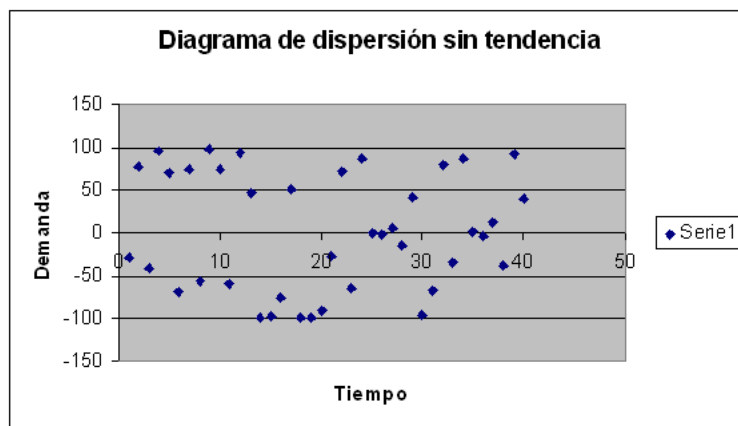


FIGURA 4. Serie de datos seudo caótica generada por la ecuación 4

de alta irregularidad, dado los cambios muy rápidos que existen en la tecnología. El segundo caso es el índice IPC mexicano 7.

3.3. Análisis de los datos. Los datos generados para el experimento y el caso de estudios serán analizados mediante una técnica cuantitativa. De todos los casos generados para comprobar la hipótesis, se medirá el error cuadrático medio, el pronóstico para el siguiente

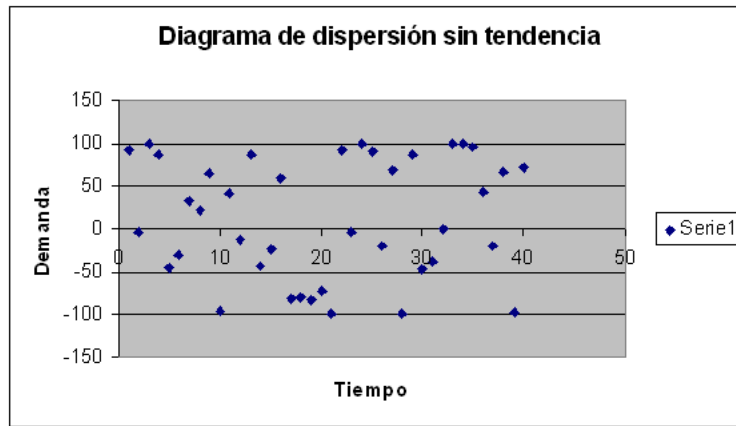


FIGURA 5. Serie de datos pseudo caótica generada por la ecuación 5.

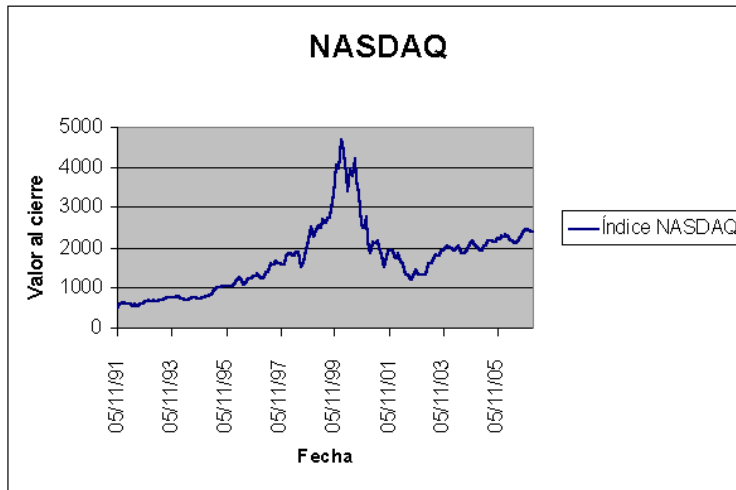


FIGURA 6. Índice NASDAQ

periodo de tiempo futuro y el porcentaje de error entre el pronóstico y el valor real. Se realizará una prueba de estadística descriptiva [Roberto Hernandez Sampieri(2005)] que implica realizar medidas de tendencia central que en este caso particular se obtendrá la media para el porcentaje de error.

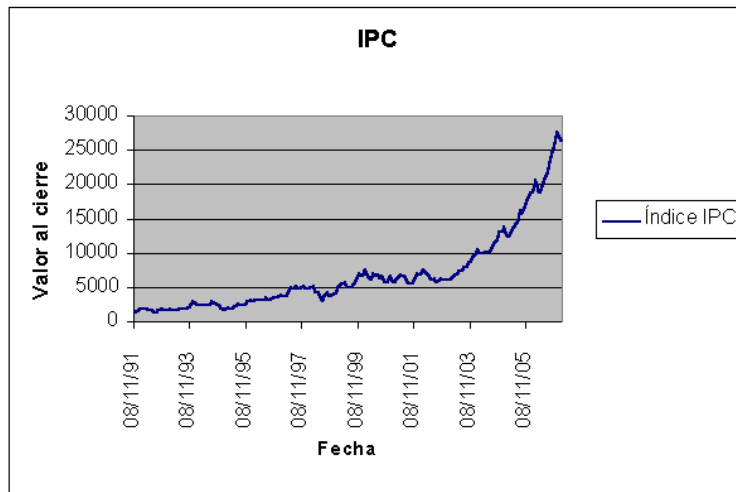


FIGURA 7. Índice IPC mexicano

Se compararán el porcentaje de error entre los diferentes casos de estudio y se aceptará o rechazará la propuesta.

Finalmente se probará la técnica en los casos reales de estudio y los métodos tradicionales, se obtendrá el error cuadrático medio y se compararán los resultados. Las diferentes técnicas serán utilizadas con el 75 % de los datos para realizar el pronóstico del 25 % restante de los datos reales. Esto se hace con la finalidad de poder comprobar y comparar el pronóstico realizado por las diferentes técnicas con los datos que se obtuvieron en la vida real.

CAPÍTULO 4

Análisis de resultados finales

En el presente capítulo se analizan los resultados obtenidos sobre los diferentes casos de prueba del experimento diseñado. Así mismo, se analizan los tres diferentes casos de estudio reales mencionados en el capítulo Procedimiento de investigación.

1. Resultados de los casos de prueba

Los casos de prueba sobre los que se realizaron los análisis fueron una serie de datos con diagrama de dispersión lineal (figura 1), no lineal (figura 2), estacional (figura 3), estacional con tendencia lineal ascendente (figura 4) y sin ninguna tendencia (figura 5), esta última representa un ambiente alto en incertidumbre.

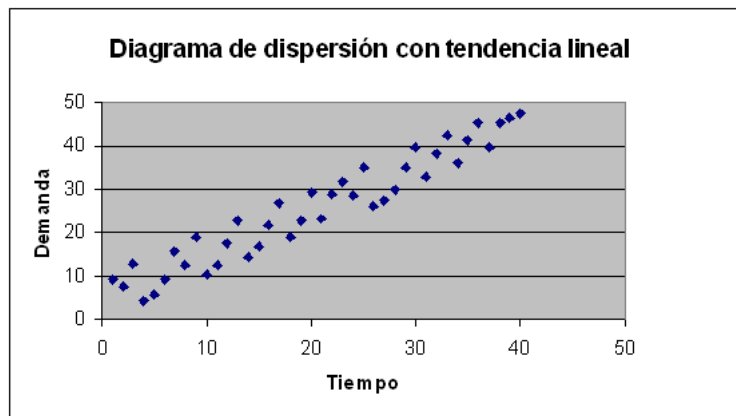


FIGURA 1. Diagrama de dispersión con tendencia lineal.

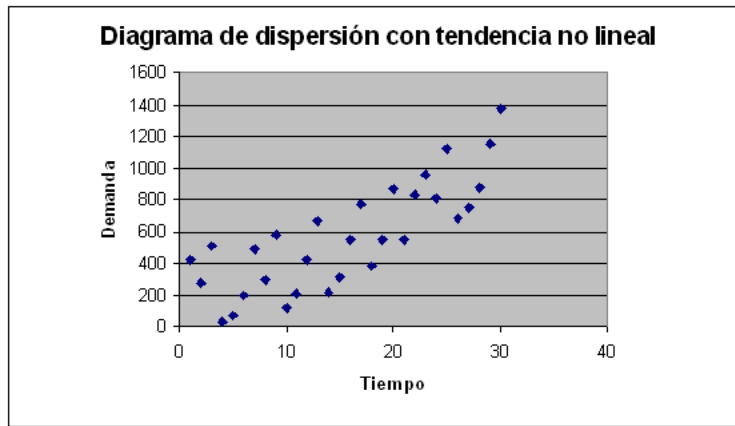


FIGURA 2. Diagrama de dispersión con tendencia no lineal.

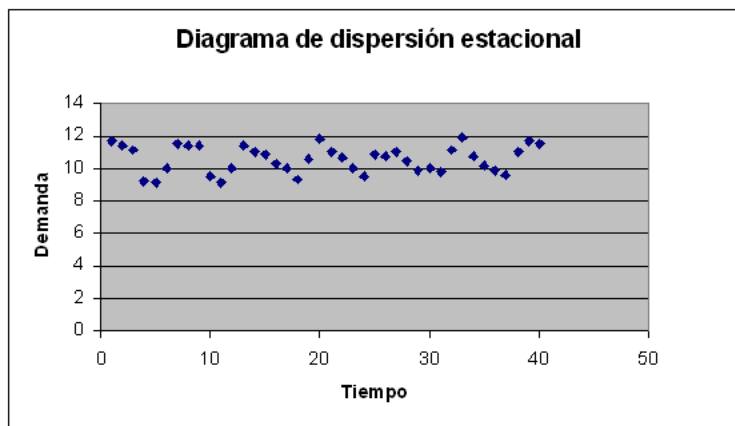


FIGURA 3. Diagrama de dispersión estacional.

A estos diagramas de dispersión se les aplicó los métodos de pronósticos de promedios móviles, promedios móviles ponderados, suavizamiento exponencial, regresión lineal y el método propuesto utilizando programación genética. Se realizó una medición del error utilizando el error cuadrático medio (ECM) y una comparación del siguiente valor de la serie con el valor pronosticado por la técnica tradicional.

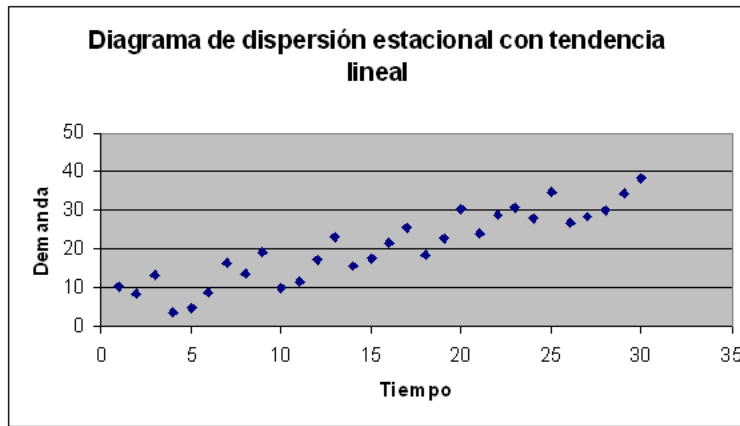


FIGURA 4. Diagrama de dispersión estacional con tendencia lineal.

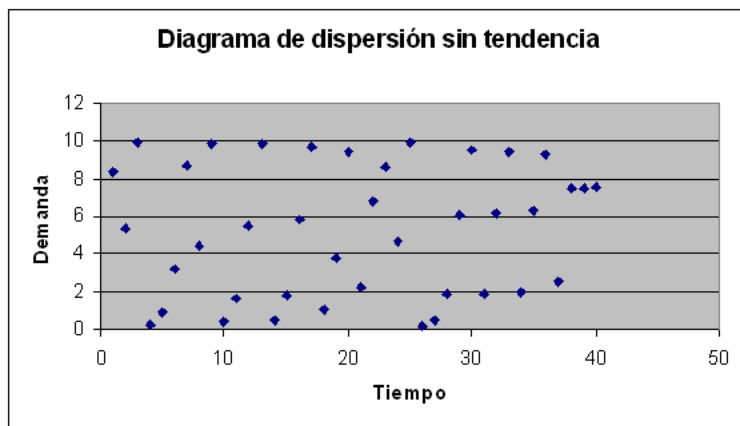


FIGURA 5. Diagrama de dispersión sin tendencia.

Los resultados obtenidos pueden proporcionar evidencias sobre que técnica tiene un mejor desempeño, es decir, tiene un menor error cuadrático medio. Así como cuál técnica pronostica mejor el siguiente valor futuro.

1.1. Resultados para el primer caso de prueba. En cada caso se presentan cinco gráficas, la primera representa la comparación

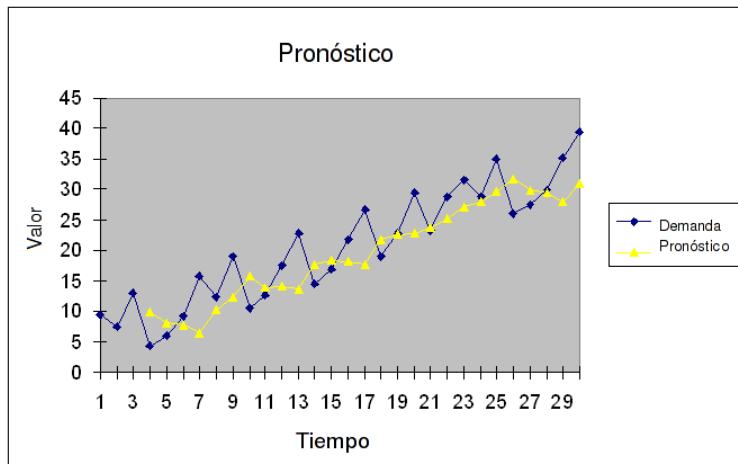


FIGURA 6. Comparativa entre los datos reales y los pronosticados por el método de promedios móviles.

entre los valores reales y el valor pronosticado por el método de promedios móviles, la segunda muestra los valores pronosticados utilizando la técnica de promedios móviles ponderados, en la tercera gráfica se observa la comparación utilizando el método de suavizamiento exponencial, la técnica de regresión lineal es la utilizada para obtener la cuarta gráfica y finalmente la quinta gráfica muestra la comparación utilizando la técnica de programación genética.

1.1.1. Ambiente lineal. En este caso se utilizó la serie de datos de la figura 1.

- *Promedios móviles.* La gráfica de la figura 6 muestra el resultado de aplicar la técnica de promedios móviles. En este caso se obtuvo un error cuadrático medio de 25.51 unidades. El valor para el siguiente periodo es de 34.83 unidades.

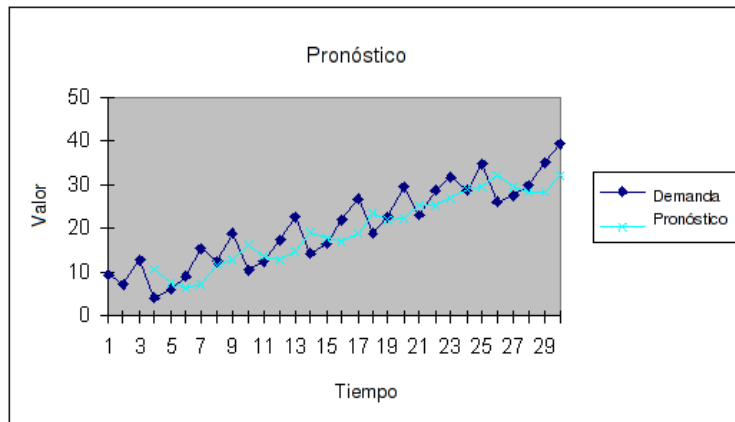


FIGURA 7. Pronósticos utilizando el método de promedios móviles ponderados.

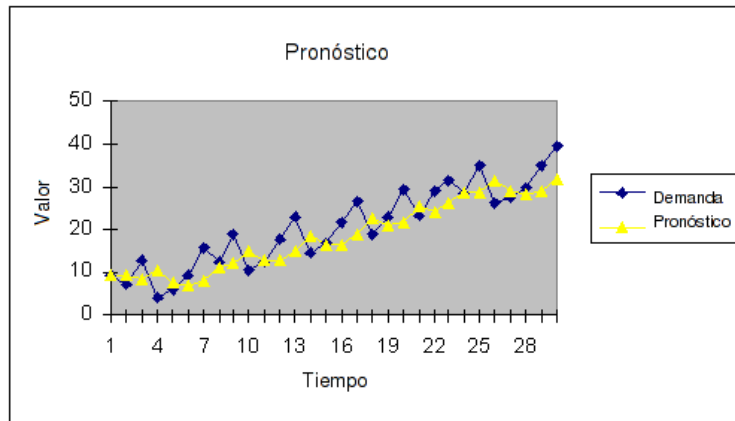


FIGURA 8. Comparativa utilizando la técnica de suavizamiento exponencial.

- *Promedios móviles ponderados.* En este caso el error cuadrático medio fue de 24.48 y el valor para el siguiente periodo es de 36.44 unidades. Figura 7.
- *Suavizamiento exponencial.* En valor del error en este caso fue de 23.67 y 35.26 como pronóstico para el siguiente periodo. Figura 8.

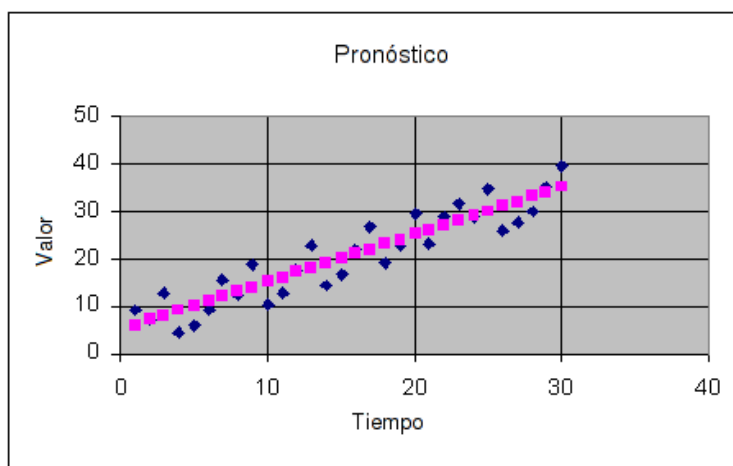


FIGURA 9. Regresión lineal.

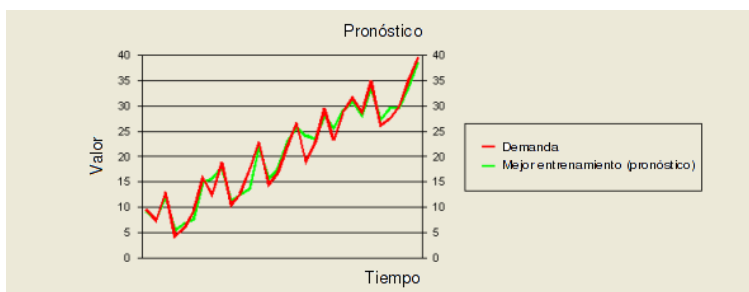


FIGURA 10. Pronóstico obtenido mediante la técnica de programación genética en un ambiente lineal.

- *Regresión Lineal.* En la figura 9 se observan los valores reales y la línea recta que mejor representa a estos datos. El error cuadrático medio fue de 2.94 unidades. Utilizando esta ecuación se obtiene un pronóstico para el siguiente periodo de 36 unidades.
- *Programación genética.* En la figura 10 se muestra el resultado obtenido por la ecuación encontrada por Magister, con un error del 2.77 y una pronóstico de 35.90 para el siguiente periodo.

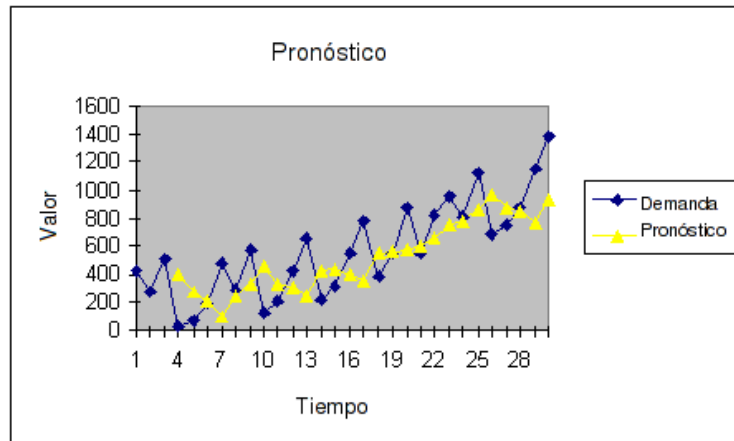


FIGURA 11. Análisis de pronósticos mediante promedios móviles en un ambiente no lineal.

1.1.2. *Ambiente no lineal.* La serie de datos de la figura 2 se utilizó para realizar los siguientes análisis:

- *Promedios móviles.* El error cuadrático medio fue 61368.66 unidades y el pronóstico para el siguiente periodo es 1133.36 unidades. Figura 11.
- *Promedios móviles ponderados.* En la figura 12 se observa el análisis de pronósticos mediante promedios móviles ponderados en un ambiente no lineal, con un valor de 60655.89 como error cuadrático medio y 1216.18 pronóstico para el siguiente periodo.
- *Suavizamiento exponencial.* Comparativa del valor real contra el pronóstico obtenido por el método de suavizamiento exponencial. Error cuadrático medio igual a 55262.18 y pronóstico para el siguiente periodo de 1154.81. Figura 13.

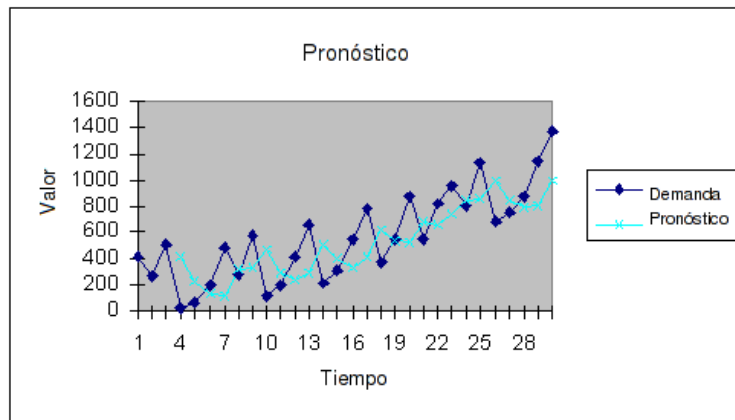


FIGURA 12. Comparativa de valor real contra el pronóstico obtenido por el método de promedios móviles ponderados.

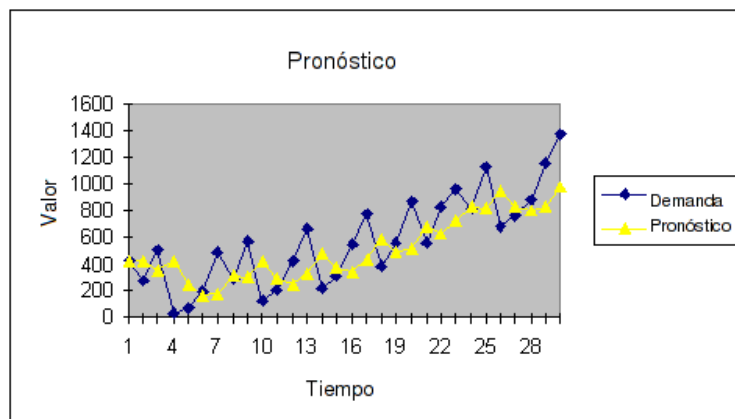


FIGURA 13. Pronóstico mediante suavizamiento exponencial en un ambiente no lineal.

- *Regresión Lineal*. Con un valor de error cuadrático medio de 39171.93 y mediante la línea estimada se encontró un valor de pronóstico para el siguiente periodo de 1045.79 . Figura 14.
- *Programación genética*. Para este ambiente el pronóstico de la figura 15 obtuvo un error cuadrático medio de 14,891.33 y un valor pronosticado para el siguiente periodo de 1,298.42 .

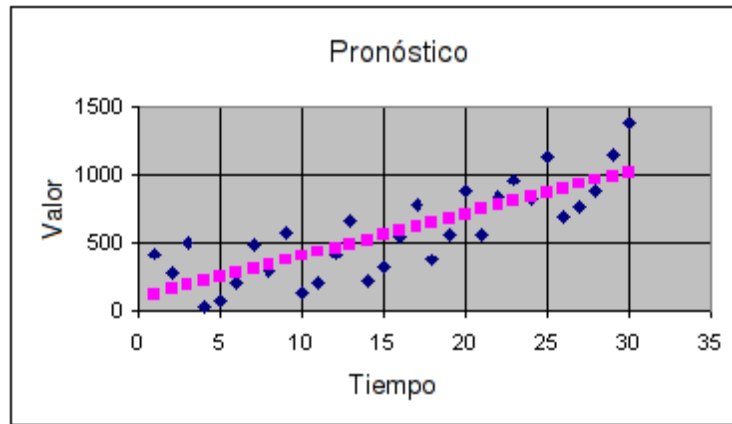


FIGURA 14. Análisis de regresión lineal sobre un diagrama de dispersión no lineal.

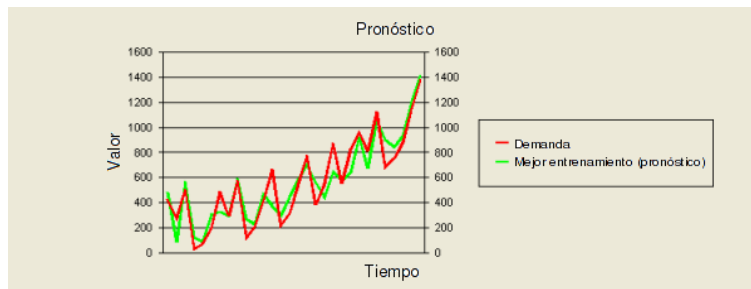


FIGURA 15. Resultado obtenido mediante programación genética para realizar pronósticos sobre una serie de datos no lineales.

1.1.3. *Ambiente estacional.* En este caso se utilizó la serie de datos de la figura 3.

- *Promedios móviles.* La gráfica de la figura 16 muestra el resultado de aplicar la técnica de promedios móviles. En este caso se obtuvo un error cuadrático medio de 1.29 unidades. El valor para el siguiente periodo es de 10.12 unidades.
- *Promedios móviles ponderados.* En este caso el error cuadrático medio fue de 1.06 y el valor para el siguiente periodo es de 10.4 unidades. Figura 7.

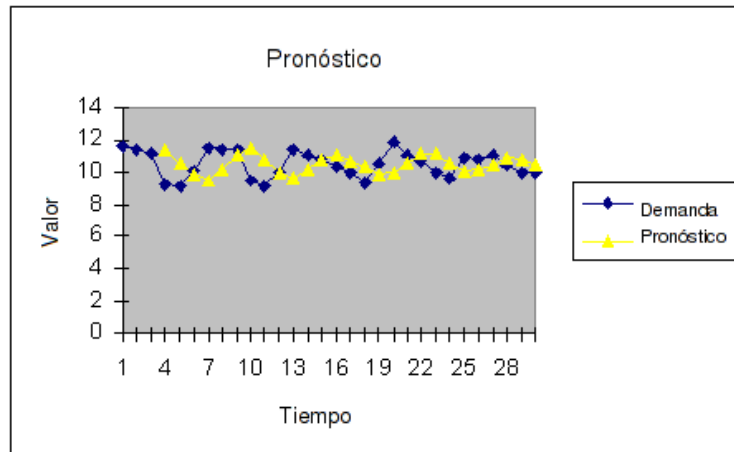


FIGURA 16. Comparativa entre los datos reales y los pronosticados por el método de promedios móviles en una serie de datos estacionales.

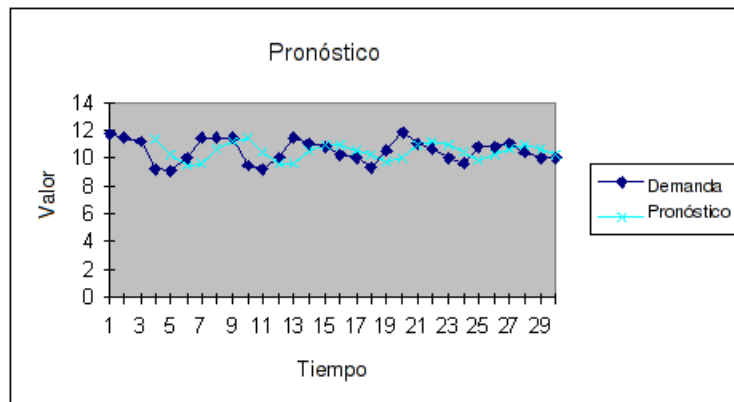


FIGURA 17. Pronósticos utilizando el método de promedios móviles ponderados en un ambiente estacional.

- *Suavizamiento exponencial*. En valor del error en este caso fue de 0.86 y 10.6 como pronóstico para el siguiente periodo. Figura 18.
- *Regresión Lineal*. En la figura 19 se observan los valores reales y la línea recta que mejor representa a estos datos. El error cuadrático medio fue de 0.63 unidades. Utilizando esta

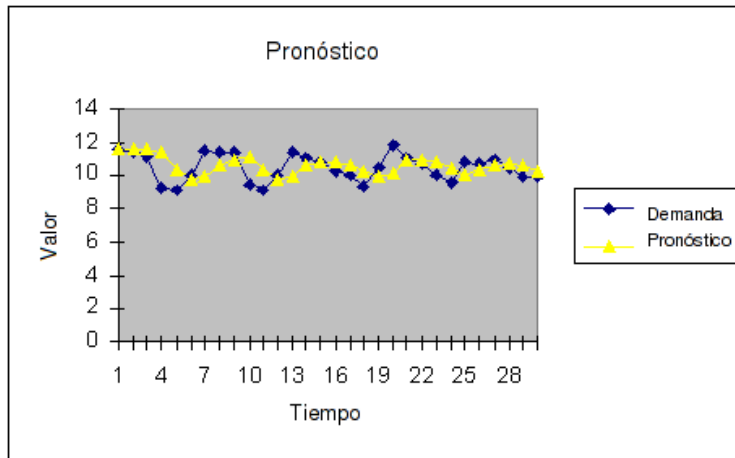


FIGURA 18. Comparativa utilizando la técnica de suavizamiento exponencial en datos estacionales.

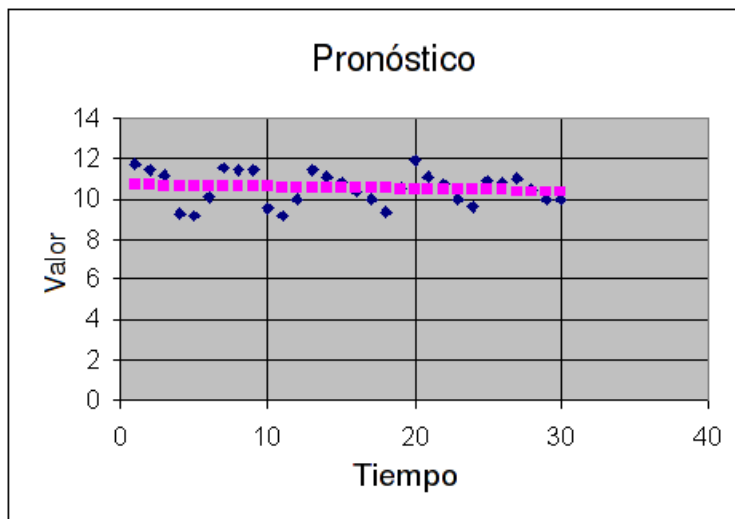


FIGURA 19. Regresión lineal en un ambiente estacional.

ecuación se obtiene un pronóstico para el siguiente periodo de 10.04 unidades.

- *Programación genética.* En este caso, el método propuesto obtuvo un pronóstico pr el siguiente periodo de 10.09 y un error cuadrático medio de 0.06 . Figura 20.

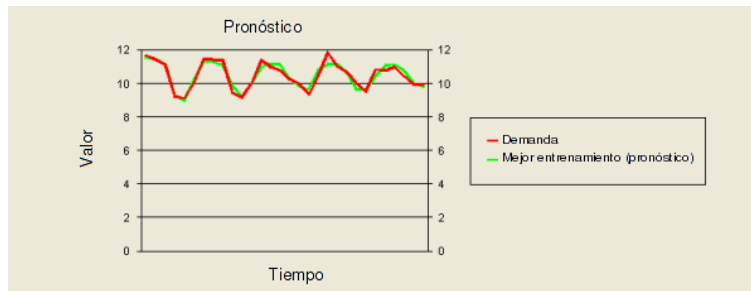


FIGURA 20. Pronóstico obtenido por la programación genética en datos con tendencia estacional.

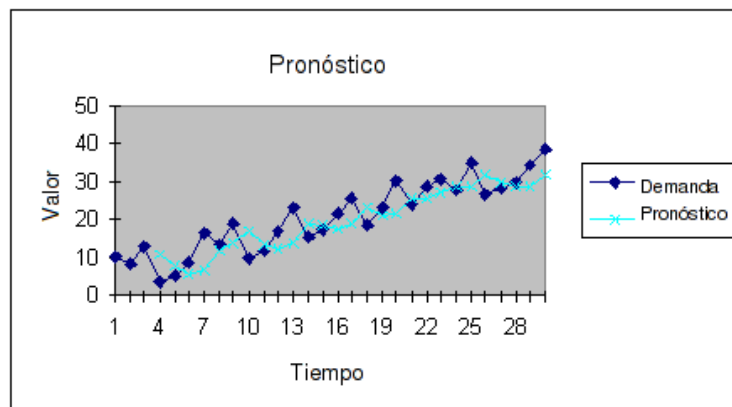


FIGURA 21. Análisis de pronósticos mediante promedios móviles en un ambiente estacionales con tendencia lineal ascendente.

1.1.4. *Ambiente con tendencia estacional y lineal.* La serie de datos de la figura 4 se utilizó para realizar los siguientes análisis:

- *Promedios móviles.* El error cuadrático medio fue 26.21 unidades y el pronóstico para el siguiente periodo es 34.37 unidades. Figura 21.
- *Promedios móviles ponderados.* En la figura 22 se observa el análisis de pronósticos mediante promedios móviles ponderados en un ambiente estacional con tendencia lineal, con un

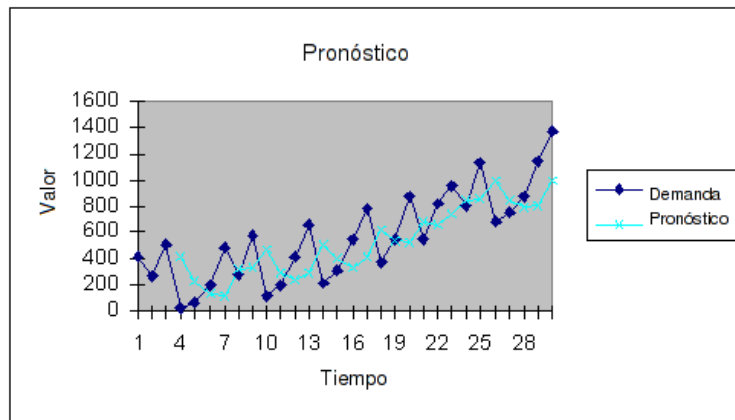


FIGURA 22. Comparativa de valor real contra el pronóstico obtenido por el método de promedios móviles ponderados.

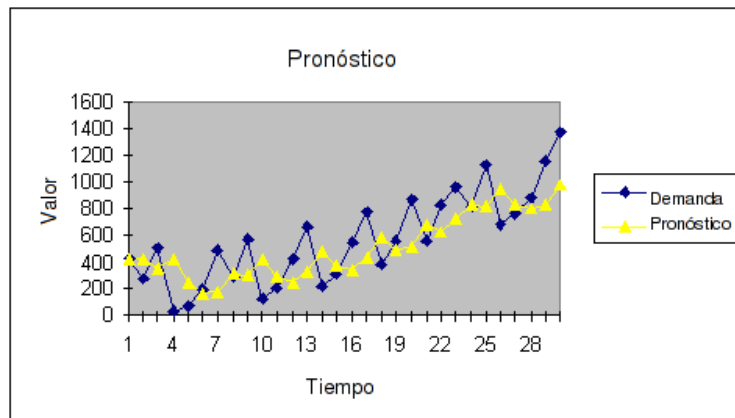


FIGURA 23. Pronóstico mediante suavizamiento exponencial en un ambiente estacional ascendente.

valor de 25.46 como error cuadrático medio y 35.77 pronóstico para el siguiente periodo.

- *Suavizamiento exponencial*. Comparativa del valor real contra el pronóstico obtenido por el método de suavizamiento exponencial. Error cuadrático medio igual a 23.85 y pronóstico para el siguiente periodo de 34.77. Figura 23.

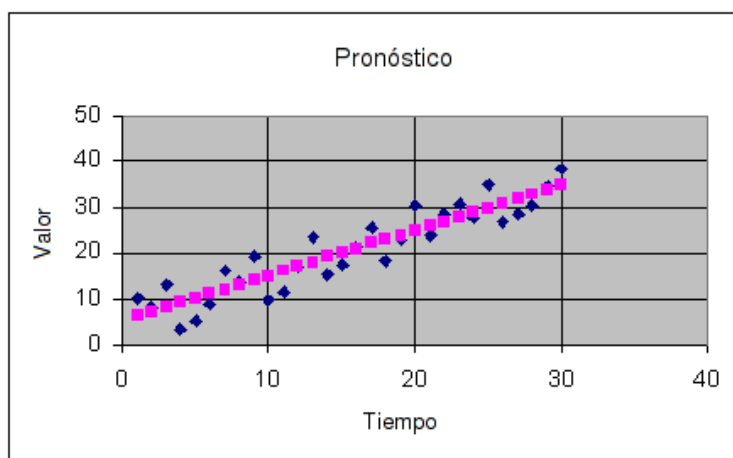


FIGURA 24. Análisis de regresión lineal sobre un diagrama de dispersión estacional con tendencia lineal ascendente

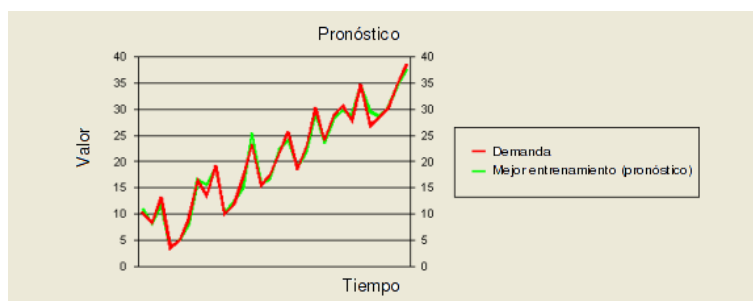


FIGURA 25. Resultado obtenido mediante programación genética para realizar pronósticos sobre una serie de datos estacionales con tendencia lineal.

- *Regresión Lineal.* Con un valor de error cuadrático medio de 13.38 y mediante la línea estimada se encontró un valor de pronóstico para el siguiente periodo de 35.84 . Figura 24.
- *Programación genética.* En este ambiente el método propuesto obtuvo un error de 2.31 y un pronóstico para el periodo siguiente de 35.56 . Figura 25.

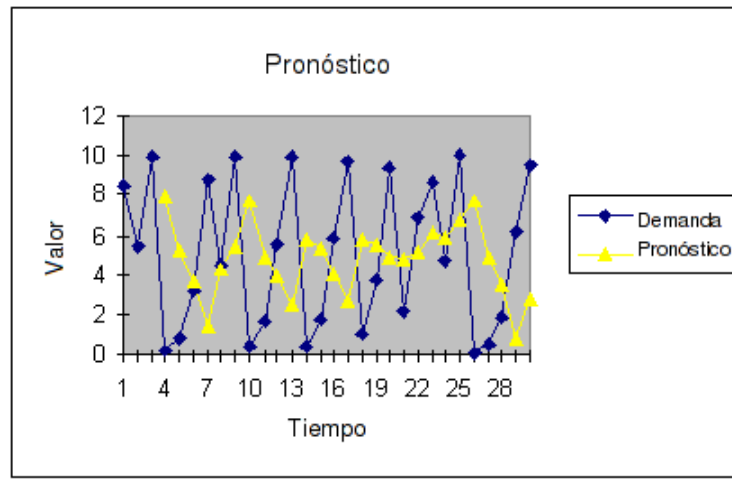


FIGURA 26. Análisis de pronósticos mediante promedios móviles en un ambiente no lineal.

1.1.5. *Ambiente de incertidumbre.* La serie de datos de la figura 5 se utilizó para realizar los siguientes análisis:

- *Promedios móviles.* El error cuadrático medio fue 21.83 unidades y el pronóstico para el siguiente periodo es 5.83 unidades. Figura 26.
- *Promedios móviles ponderados.* En la figura 27 se observa el análisis de pronósticos mediante promedios móviles ponderados en un ambiente de incertidumbre, con un valor de 22.2 como error cuadrático medio y 7.3 pronóstico para el siguiente periodo.
- *Suavizamiento exponencial.* Comparativa del valor real contra el pronóstico obtenido por el método de suavizamiento exponencial. Error cuadrático medio igual a 19.38 y pronóstico para el siguiente periodo de 6.48. Figura 28.

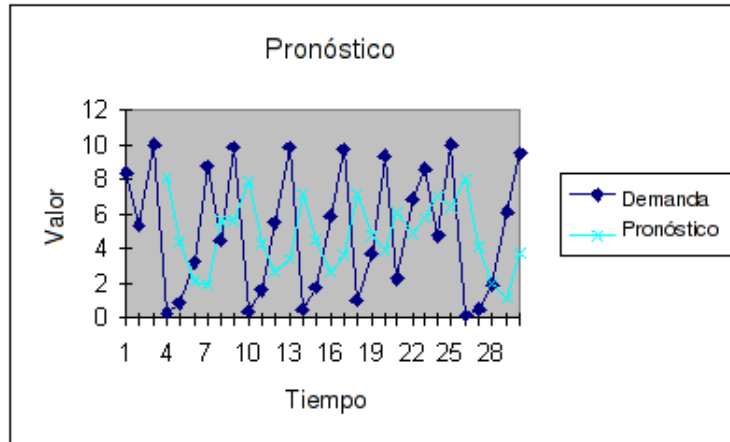


FIGURA 27. Comparativa de valor real contra el pronóstico obtenido por el método de promedios móviles ponderados.

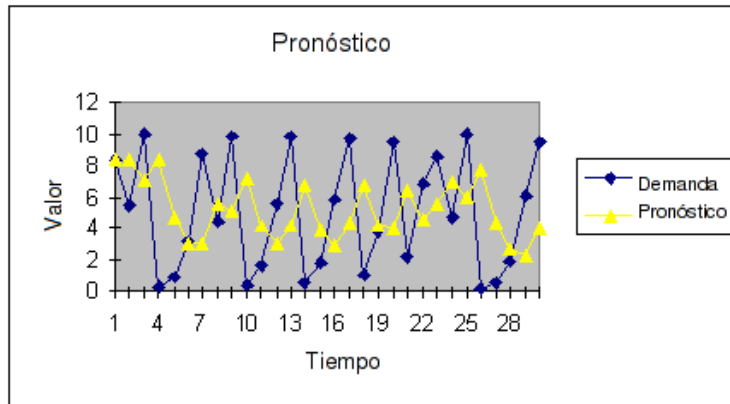


FIGURA 28. Pronóstico mediante suavizamiento exponencial en un ambiente sin tendencia.

- *Regresión Lineal*. Con un valor de error cuadrático medio de 12.94 y mediante la línea estimada se encontró un valor de pronóstico para el siguiente periodo de 5 . Figura 29.
- *Programación genética*. En la figura 30 se tiene la aproximación encontrada con Magister a los datos sin tendencia. Se encontró

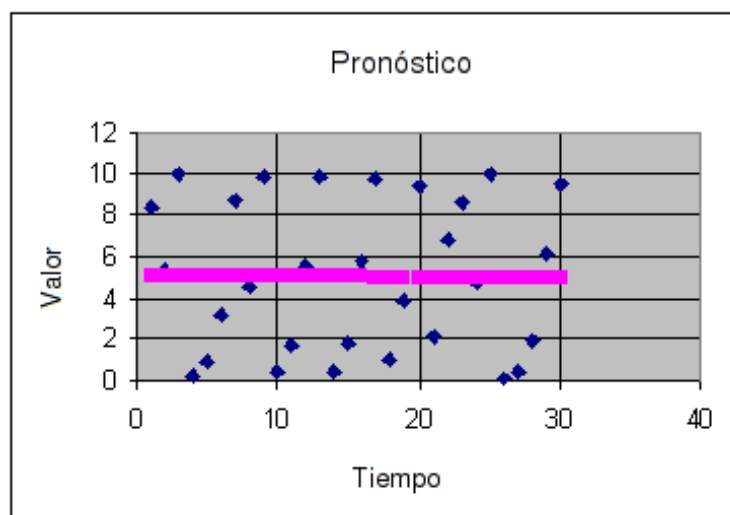


FIGURA 29. Análisis de regresión lineal sobre un diagrama de dispersión sin tendencia alguna.

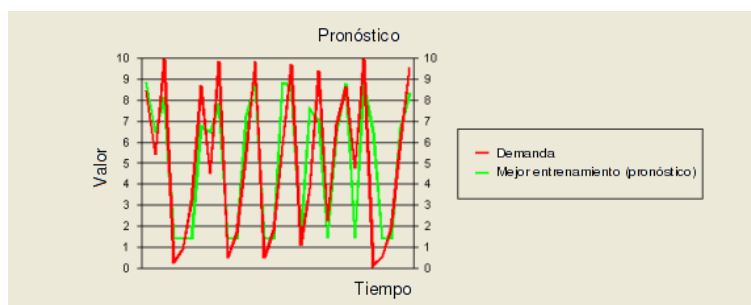


FIGURA 30. Resultado obtenido mediante programación genética para realizar pronósticos sobre una serie de datos sin tendencia.

un error cuadrático medio de 3.80 y un pronóstico para el periodo siguiente de 2.02 .

1.2. Análisis de resultados sobre el primer caso de estudio. En la figura 31 se muestran los resultados obtenidos con las técnicas tradicionales para realizar pronósticos y la propuesta para los diferentes ambientes de trabajo.

Escenario	Técnica	ECM	Pronóstico	Valor Real	% Error
Lineal	PM	25.51	34.83	32.89	5.90%
Lineal	PMP	24.48	36.44	32.89	10.79%
Lineal	SE	23.67	35.26	32.89	7.21%
Lineal	RL	2.94	36.00	32.89	9.46%
Lineal	PG	2.77	35.90	32.89	9.15%
No lineal	PM	61386.66	1133.36	1055.58	7.37%
No lineal	PMP	60655.89	1216.18	1055.58	15.21%
No lineal	SE	55262.18	1154.81	1055.58	9.40%
No lineal	RL	39171.93	1045.79	1055.58	0.93%
No lineal	PG	14891.33	1298.42	1055.58	23.01%
Estacional	PM	1.29	10.12	9.78	3.48%
Estacional	PMP	1.06	10.40	9.78	6.34%
Estacional	SE	0.86	10.60	9.78	8.38%
Estacional	RL	0.63	10.04	9.78	2.66%
Estacional	PG	0.06	10.09	9.78	3.17%
Est. Asc	PM	26.21	34.37	32.48	5.82%
Est. Asc	PMP	25.46	35.77	32.48	10.13%
Est. Asc	SE	23.85	34.77	32.48	7.05%
Est. Asc	RL	13.38	35.84	32.48	10.34%
Est. Asc	PG	2.31	35.56	32.48	9.48%
Incertidumbre	PM	21.83	5.83	1.89	208.47%
Incertidumbre	PMP	22.20	7.30	1.89	286.24%
Incertidumbre	SE	19.38	6.48	1.89	242.86%
Incertidumbre	RL	12.94	5.00	1.89	164.55%
Incertidumbre	PG	3.80	2.02	1.89	6.88%

PM = Promedios Móviles
PMP = Promedios Móviles Ponderados
SE = Suavizamiento Exponencial
RL = Regresión Lineal
PG = Programación Genética

FIGURA 31. Resultados del primer caso experimental

En este caso en particular se observa que el método tiene un error cuadrático medio menor que todos los métodos tradicionales utilizados. Sin embargo, el método no realiza el mejor pronóstico para el siguiente periodo. En los únicos casos en que se comporta mejor es en los ambientes no lineal y de incertidumbre.

1.3. Análisis de resultados sobre los diferentes casos de prueba. El mismo procedimiento realizado sobre el primer caso de prueba, se realizó sobre los otros casos de prueba diseñados en el procedimiento de la investigación, hasta completar un total de 25 casos diferentes. Después se obtuvo la media para todos los experimentos.

Escenario	Técnica	Promedio %error
Lineal	PM	5.07%
Lineal	PMP	5.41%
Lineal	SE	5.81%
Lineal	RL	9.86%
Lineal	PG	18.19%
No lineal	PM	10.75%
No lineal	PMP	10.78%
No lineal	SE	12.15%
No lineal	RL	13.23%
No lineal	PG	6.30%
Estacional	PM	2.49%
Estacional	PMP	2.28%
Estacional	SE	3.31%
Estacional	RL	4.25%
Estacional	PG	4.75%
Est. Asc	PM	5.15%
Est. Asc	PMP	5.36%
Est. Asc	SE	5.80%
Est. Asc	RL	10.55%
Est. Asc	PG	8.05%
Incertidumbre	PM	79.25%
Incertidumbre	PMP	90.21%
Incertidumbre	SE	86.89%
Incertidumbre	RL	83.48%
Incertidumbre	PG	12.47%

PM = Promedios Móviles
PMP = Promedios Móviles Ponderados
SE = Suavizamiento Exponencial
RL = Regresión Lineal
PG = Programación Genética

FIGURA 32. Promedio de los porcentajes de error de los diferentes casos de prueba.

En la figura 32 se presenta el resumen de los resultados obtenidos sobre todos los casos de prueba.

De acuerdo a los resultados obtenidos, se concluye que para los casos de prueba diseñados los resultados fueron los mismos que para el primer caso de prueba. Es decir, el método propuesto para realizar

Escenario	Normalidad
Lineal	Si
No lineal	No
Estacional	Si
Est. Asc	Si
Incertidumbre	No

FIGURA 33. Resultados sobre la prueba de normalidad aplicados a los diferentes escenarios.

pronósticos se tuvo un menor porcentaje de error para los ambientes no lineales y de incertidumbre.

Estos resultados hacen pensar en la razón por la cual el método propuesto fue mejor en dichos ambientes. Para encontrar la respuesta a esta pregunta se realizaron pruebas de normalidad a los diferentes casos de estudio. La prueba de normalidad fue la de Shapiro & Wilk. Los resultados se muestran en la figura 33, de donde se concluye que el método propuesto es mejor para el caso en que los datos que representan al ambiente no tienen una distribución normal.

2. Casos de estudio

Para enseñar la posible utilidad del método propuesto sobre ambientes de alta incertidumbre, se sometieron a prueba tres diferentes casos de la vida real. El primer caso es el tipo de cambio del peso mexicano con respecto al dólar americano, el segundo caso es el IPC de la bolsa de valores mexicana y finalmente un ambiente de muy alta

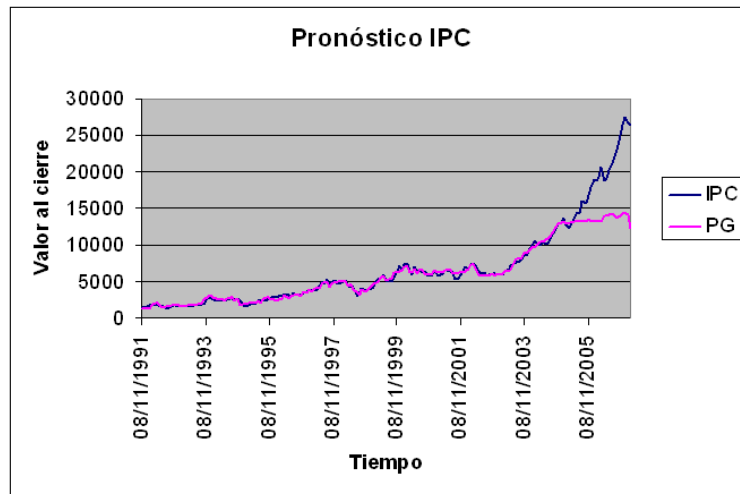


FIGURA 34. Pronóstico del año 2006 utilizando la serie de 1991 al 2005.

incertidumbre, el índice NASDAQ.

2.1. Índice mexicano IPC. En el caso del índice mexicano IPC se utilizó una serie temporal desde el año 1991 hasta el 2005 con una muestra semanal para entrenar el método propuesto y la serie del 2006 (ya conocida) para realizar el pronóstico y comprobación de la utilidad del método propuesto.

El resultado obtenido es el mostrado en la figura 34. A pesar de que el entrenamiento del método propuesto se acerca al valor real del IPC, es fácilmente observable que el valor pronosticado se aleja demasiado del valor real del IPC, es decir, el error cuadrático medio es muy grande para este caso.

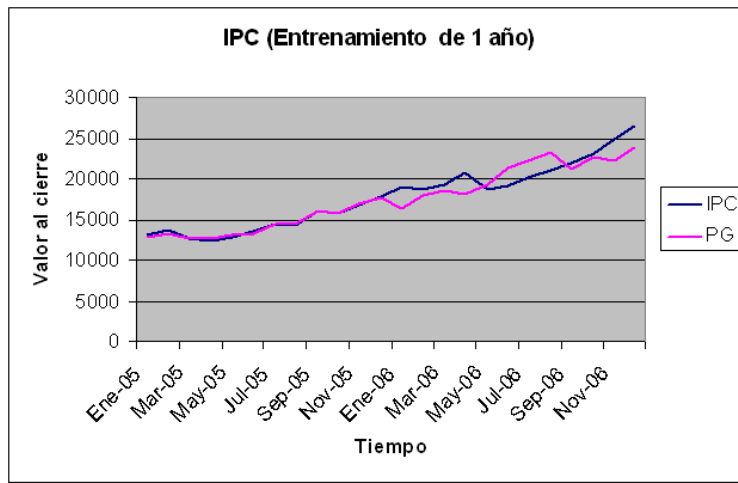


FIGURA 35. Pronóstico del IPC entrenando el método propuesto con un año de anterioridad.

Para buscar tener un mejor pronóstico de la serie IPC, se utilizó la serie con una año de anterioridad al 2006 para entrenar el método propuesto. Después se utilizaron dos años de anterioridad y así sucesivamente hasta encontrar el tiempo óptimo para realizar un pronóstico del índice IPC.

El resultado para predecir el último año utilizando un año antes, se muestra en la figura 35. En este caso el error cuadrático medio del pronóstico fue de 3,644,869.

Para el caso en que se utilizó un entrenamiento de dos años (figura 36), el error cuadrático medio fue de 2,637,246. Dicho error fue menor que el obtenido al utilizar un año de entrenamiento.

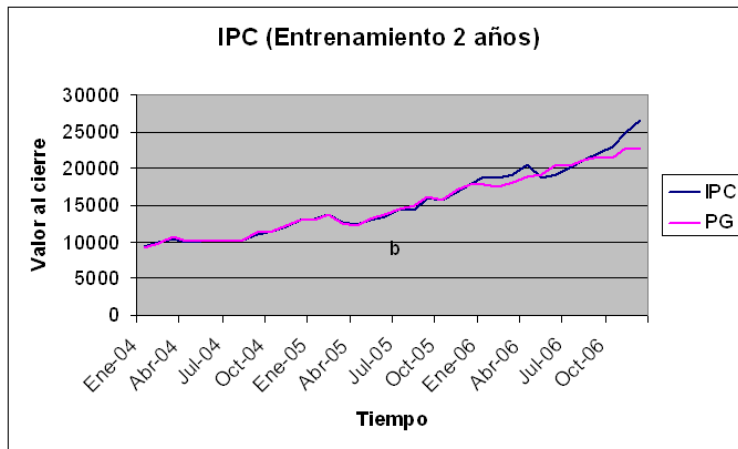


FIGURA 36. Pronóstico del IPC entrenando el método propuesto con 2 años de anterioridad.

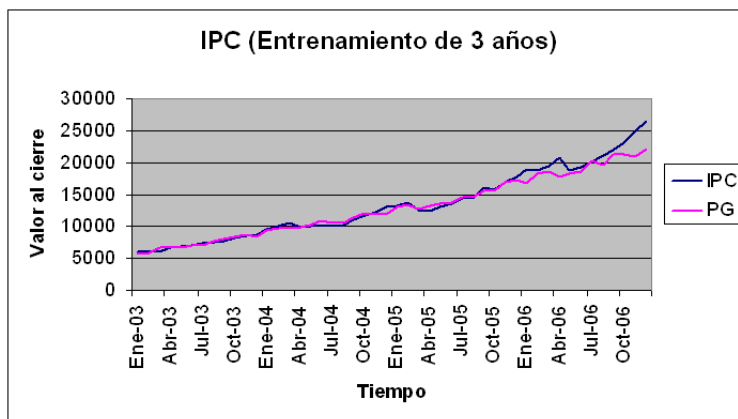


FIGURA 37. Pronóstico del IPC entrenando el método propuesto con 3 años de anterioridad.

El error cuadrático medio para el entrenamiento con 3 años de anterioridad para el IPC (figura 37), subió a 3,687,101. Este error es similar al entrenar el método con un año de anterioridad.

Para el siguiente entrenamiento (figura 38), correspondiente a cuatro años de anterioridad, el error cuadrático medio medio vuelve a

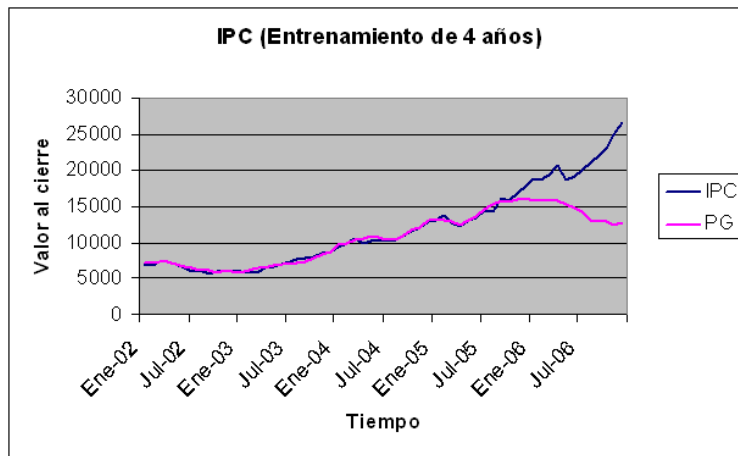


FIGURA 38. Pronóstico del IPC entrenando el método propuesto con 4 años de anterioridad.

incrementarse. El valor del ECM fue de 59,925,440.

Como último caso, se realizó el pronóstico con 5 años de anterioridad de entrenamiento. El error cuadrático medio fue de 20,543,554. Un ECM mucho mayor que para los tres primeros casos. El resultado se muestra en la figura 39.

Se puede concluir que para realizar un pronóstico del IPC mexicano con el método de programación genética, la mejor forma es utilizar dos años de información pasada.

2.2. Índice NASDAQ. El índice NASDAQ es una serie temporal que tiene una gran componente aleatoria (incertidumbre). Para este caso de estudio se tomó la serie en un periodo de tiempo igual que para el índice IPC. Es decir, para realizar el pronóstico del intervalo de enero

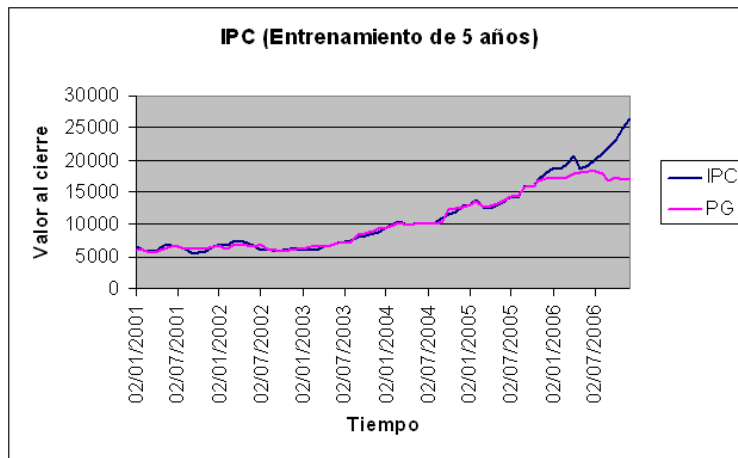


FIGURA 39. Pronóstico del IPC entrenando el método propuesto con 5 años de anterioridad.

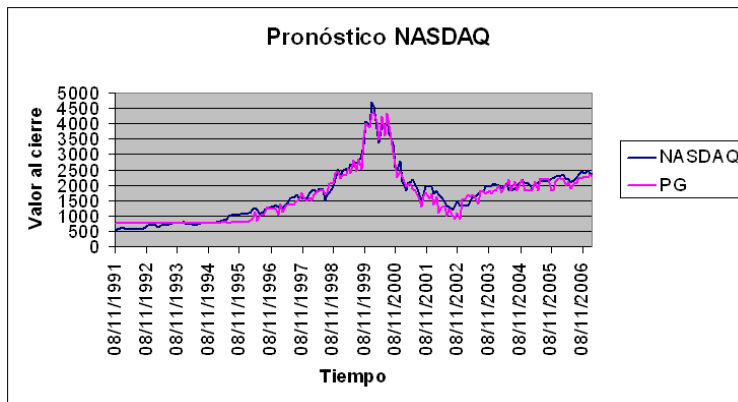


FIGURA 40. Pronóstico del índice NASDAQ.

del 2006 a enero del 2007, se utilizó la serie desde el año 1991 hasta el año 2005, como lo muestra la figura 40. El error cuadrático medio del pronóstico fue de 22,136. Pero al igual que para el índice IPC se buscó el intervalo de tiempo óptimo para realizar un pronóstico más acertado.

En la figura 41 se muestra la gráfica correspondiente al pronóstico del año 2006 obtenido al entrenar el método con un año de información.

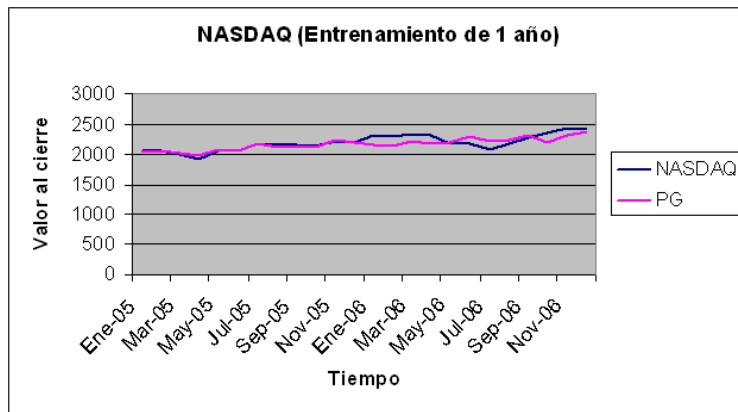


FIGURA 41. Pronóstico del NASDAQ entrenando el método propuesto con 1 año de anterioridad.

En este caso se obtuvo un error cuadrático medio de 13,429, un ECM menor que el obtenido con los 15 años de entrenamiento.

Para el caso del entrenamiento con dos años de anterioridad (figura 42) se obtuvo un ECM igual a 9,014. Este valor es menor que para el entrenamiento con un año de anterioridad. Esto asegura un mejor pronóstico del índice NASDAQ.

En el índice NASDAQ, y a diferencia del índice IPC, el error cuadrático medio baja para cuando se utilizó tres años de entrenamiento. El ECM fue de 8,593, mejorando el pronóstico. Figura 43.

Para el caso del entrenamiento con 4 años, el error cuadrático medio aumenta a 13,838. Figura 44.

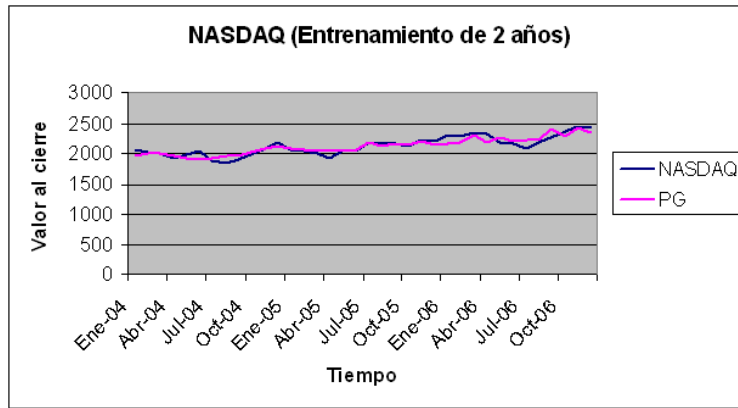


FIGURA 42. Pronóstico del NASDAQ entrenando el método propuesto con 2 años de anterioridad.

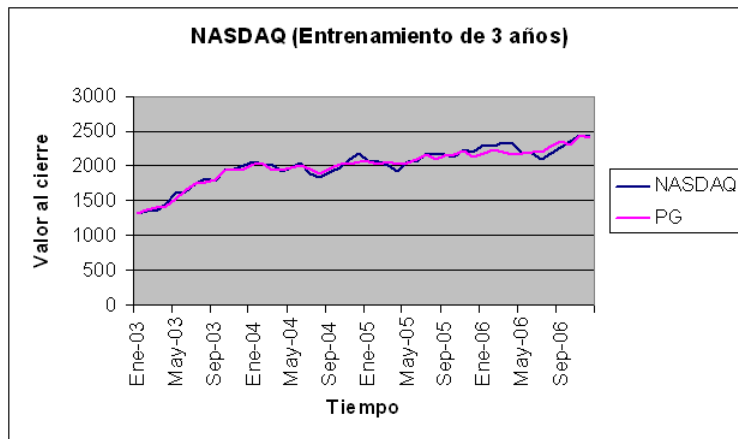


FIGURA 43. Pronóstico del NASDAQ entrenando el método propuesto con 3 años de anterioridad.

Finalmente, se obtuvo un ECM de 24,994 para el pronóstico del NASDAQ con un entrenamiento de 5 años. Este valor vuelve a aumentar con respecto a los anteriores entrenamientos.

Para el caso de estudio del índice NASDAQ, el mejor pronóstico se obtiene al utilizar 3 años de información pasada.

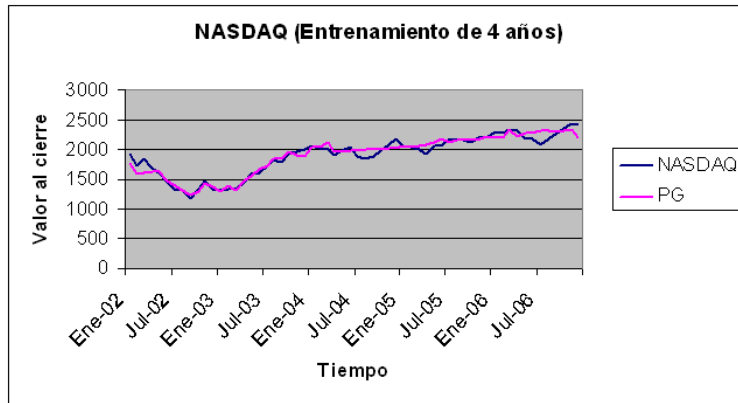


FIGURA 44. Pronóstico del NASDAQ entrenando el método propuesto con 4 años de anterioridad.

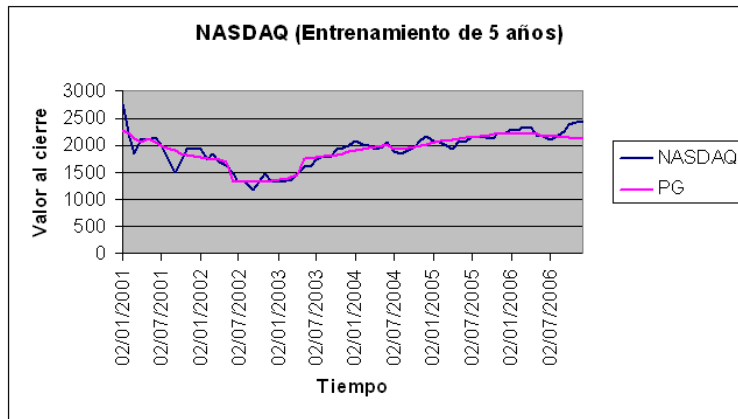


FIGURA 45. Pronóstico del NASDAQ entrenando el método propuesto con 5 años de anterioridad.

CAPÍTULO 5

Conclusiones

En la administración de una organización es de suma importancia el tomar las decisiones correctas para lograr el objetivo de la empresa. Para tomar las mejores decisiones es necesario evaluar los posibles escenarios futuros. Actualmente, los escenarios cambian rápidamente debido a la gran globalización que existe, por lo que realizar un pronóstico bajo estas condiciones es bastante complicado.

Es por ello que en este trabajo, se estudió y analizó el uso de un método basado en programación genética para realizar pronósticos bajo elementos de difícil previsión, bajo ambientes de alta incertidumbre. Para lograr el objetivo, se realizó un análisis comparativo entre las técnicas más utilizadas (tradicionales) contra el método basado en programación genética. Se utilizaron dichas técnicas para realizar pronósticos sobre diferentes casos experimentales. Finalmente la técnica nueva se utilizó para tratar de realizar pronóstico sobre casos reales.

1. Resumen

En el capítulo "Introducción" se explicó el problema a resolver en este trabajo. También se explicó el proceso que se siguió para realizar

la investigación.

En el capítulo "Marco teórico" se realizó una investigación sobre la forma en que se realizan los pronósticos en la administración y la economía. Así mismo, sobre que tipos de técnicas existen para realizar pronósticos y cuáles son las más utilizadas. Por otro lado, debido a que se propuso utilizar la programación genética para realizar pronósticos, se realizó una investigación sobre los conceptos básicos de la misma. También fue importante diferenciar entre los conceptos de programación genética y algoritmos genéticos, ya que el nombre de ellos se presta a confusión y a creer que son técnicas iguales. Para ello se explicó el origen de dichas dos técnicas, proviniendo de la inteligencia artificial.

En el capítulo "Procedimiento de la investigación" se planteó de forma concreta el problema a resolver y el método o experimento que se siguió para resolverlo. También se decidió en cómo se seleccionaría la muestra para realizar el experimento, cómo se recolectarían los datos para el experimento y cómo se procedería para analizar los datos. Para analizar los datos, se diseñó y se construyó un sistema de información basado en programación genética, llamado Magister.

En el capítulo "Análisis de resultados finales" se presentaron los resultados obtenidos del experimento diseñado, llegando a las siguientes conclusiones: El método para realizar pronósticos basado en programación genética tiene mayor eficiencia que los métodos tradicionales

en series temporales en las que su componente irregular predomina, es decir, en ambientes de alta incertidumbre.

2. Conclusiones

Es importante hacer notar que las siguientes conclusiones fueron obtenidas sobre ambientes univariantes y simples, es decir, en ambientes con una sola variable dependiente y una sola variable independiente. Y dado que existen una infinidad de posibles ambientes, las conclusiones también se limitan a los casos de estudio del presente trabajo. En otras palabras, a ambientes lineales, no lineales, estacionarios y de incertidumbre.

Los resultados obtenidos en el capítulo "Análisis de resultados finales" arrojan las siguientes conclusiones.

La técnica basada en programación genética se comporta mejor en ambientes no lineales y de incertidumbre, que las técnicas tradicionales. En el apartado "Análisis de resultados sobre los diferentes casos" se concluye que el método propuesto trabaja mejor que las técnicas tradicionales probadas en los ambientes no lineales y de incertidumbre. En ambientes lineales y con tendencia estacionaria es mejor utilizar las técnicas tradicionales. Sin embargo, es importante hacer notar, que el método propuesto también puede ser utilizado para los ambientes lineales y con tendencia estacionaria, ya que su porcentaje de error no es

mucho mayor que el de las otras técnicas.

La técnica propuesta se comporta mejor en una serie de datos que no tenga una distribución normal. Otra conclusión importante sobre estos resultados es que la distribución estadística de los datos está ligada al tipo de ambiente. Al realizar una prueba de normalidad sobre los datos de los diferentes ambientes, se encontró que aquellos ambientes que no tienen una distribución normal son los ambientes no lineales y de incertidumbre. Mientras que aquellos que si tienen una distribución normal son los ambientes lineales y con tendencia estacionaria. De este hecho, se puede recomendar que para realizar un análisis de datos con la finalidad de hacer pronósticos, se debería realizar una prueba de normalidad a los datos. Si no pasa la prueba de normalidad, se recomienda utilizar el método propuesto en esta investigación.

No es necesario utilizar mucha información pasada para realizar buenos pronósticos con el método propuesto. En el apartado "Análisis de casos de estudio" se encontró que para los casos particulares del índice IPC y NASDAQ no es necesario utilizar gran cantidad de información pasada para realizar el pronóstico con un porcentaje de error pequeño. De los diferente análisis realizados se concluye que para estos dos casos en particular, los mejores pronósticos se realizan cuando se tiene información pasada de entre dos y tres años. Tener menos información o mayor información pasada se deriva en pronósticos más

errados.

En el método propuesto no es necesario analizar previamente la información. Finalmente es importante señalar que una de las principales ventajas del método propuesto sobre los tradicionales es que no se necesita hacer un pre-análisis de los datos para utilizarlo. Es decir, en los métodos tradicionales se recomienda graficar las series de tiempo para tratar de deducir si la serie temporal tiene componentes lineales, no lineales, estacionarios, ciclos o irregulares, y de esa forma escoger el método a utilizar para realizar los pronósticos. En el caso del método propuesto sólo se utilizan los datos sin tener información previa de ellos.

La técnica propuesta para realizar pronósticos no necesita de la experiencia para realizar el pronóstico. De igual forma, en algunos de los métodos tradicionales es necesario utilizar el conocimiento previo y experiencia del analizador para obtener buenos pronósticos. En el método propuesto esta sería una condición útil pero no imprescindible.

3. Recomendaciones y trabajo futuro

En este trabajo de investigación se estudió la aplicación de la programación genética para realizar pronósticos en series temporales con tendencias irregulares.

Esta investigación se limitó al estudio de series temporales univariantes, es decir con series que dependen únicamente del tiempo. Sin

embargo, la mayoría de las variables dentro de la administración, como la oferta, demanda, inventarios, economía, etc., no dependen únicamente del tiempo sino de otras variables, por lo que se propone como trabajo futuro el uso de esta técnica con series multivariantes.

De igual forma, también se propone que este método sea explorado para investigar la posibilidad de que pueda explicar la causalidad entre datos, es decir, si la técnica de programación genética puede indicar qué factores influyen en la variable de estudio (a pronosticar).

La programación genética es una técnica relativamente nueva de la inteligencia artificial. Se está demostrando que ésta técnica es útil y eficaz para resolver problemas en diferentes áreas. No sólo de la administración, sino también en otros campos del saber humano. A continuación se mencionan algunas de ellas que han sido aplicadas sobre diferentes especialidades tales como en:

- Economía y finanzas para análisis de riesgos utilizando el modelo Montecarlo de simulación [**Chidambaran(2003)**], para el diseño de portafolios de inversiones en [**Roberts(2005)**].
- Sistemas de información para la administración del conocimiento y de la información en [**Oren(2002)**], para clasificación de la información en [**Kenneth Hennessy(2005)**], y para sistemas al soporte de las decisiones en [**Wong(2001)**].

- Electrónica para el diseño de amplificadores operacionales de 96 decibeles en [**J. Koza(1997)**], o para la optimización del arreglo en compuertas lógicas de un FPGA en [**J. Koza(1998)**].
- Física para el diseño de lentes ópticos en [**John Koza(2005)**].
- Biología para evolucionar código del ADN en [**Byoung-Tak Zhang(2005)**], para encontrar familias de proteínas en [**Rolv Seehuus(2005)**], para clasificación de cáncer mediante el análisis del ADN en [**Jin-Hyuk Hong(2005)**].
- Computación en inteligencia artificial para clasificación de patrones en [**Will Smart(2004)**], [**Ankur Teredesai(2005)**] y en [**Jeroen Eggermont(2004)**], para teoría de juegos en [**Eleazar Eskin(1999)**], en lenguajes de programación para hacer parsers en [**Jackson(2005)**], para el diseño de redes inalámbricas en [**D. M. Johnson(2005)**].
- Matemáticas para resolver la transformada de Fourier en [**Paul Massey(2005)**], para resolver ecuaciones diferenciales en [**S. J. Kirstukas(2005)**].
- Ecología para predecir el crecimiento de bancos de algas en [**Nitin Muttil(2005)**].
- Investigación de operaciones para predecir la solvencia de las compañías de seguros en [**Sancho Salcedo(2005)**].
- Medicina en [**Celia C. Bojarczuk(2004)**].
- Química en [**F. V. Buontempo(2005)**].
- Robótica para el reconocimiento de huellas digitales en [**X. Tan(2005)**] y para el control de robots en [**S. Harding(2005)**].

Estos son sólo algunos de los ejemplos en los que se está utilizando la programación genética para resolver problemas en diferentes campos. Se está utilizando desde la medicina hasta la ingeniería. Esto muestra la gran potencialidad de esta técnica de la inteligencia artificial para ayudar al desarrollo del conocimiento humano y su aplicación en técnicas que faciliten la vida.

Es por ello que se recomienda seguir haciendo estudios sobre aplicaciones en dónde la programación genética pueda ser útil para resolverlos, en específico en el área de la administración, tal y como se propuso en este trabajo de investigación.

En el presente trabajo se investigó la viabilidad para aplicar un método basado en programación genética para realizar pronósticos dentro de la administración, con la finalidad de ayudar a mejorar la toma de decisiones. Dicha técnica debía ser mejor que los métodos tradicionales para realizar pronósticos sobre ambientes incertidumbres. Se aportó evidencia para confirmar dicha propuesta, llegando a la conclusión que para ayudar a hacer mejores pronósticos en ambientes de difícil previsibilidad, tiene mayor utilidad el método propuesto.

El software Magister (manual de usuario)

El software Magister es un programa computacional que realiza regresión simbólica mediante programación genética. Magister está escrito en lenguaje Scheme y se ejecuta sobre una máquina virtual de java (JVM).

Los requisitos mínimos para poder utilizar Magister son:

- Una computadora PC
- Procesador Pentium I o superior
- Sistema Operativo Windows 95 o superior
- Máquina virtual de java (JVM) de Sun.

Debido a la portabilidad de java, el programa debe poder ejecutar sobre cualquier plataforma que tenga una máquina virtual de java.

4. Ejecución de Magister

Para ejecutar el programa Magister es necesario tener instalado el sistema operativo Windows 95 o superior y una máquina virtual de java.

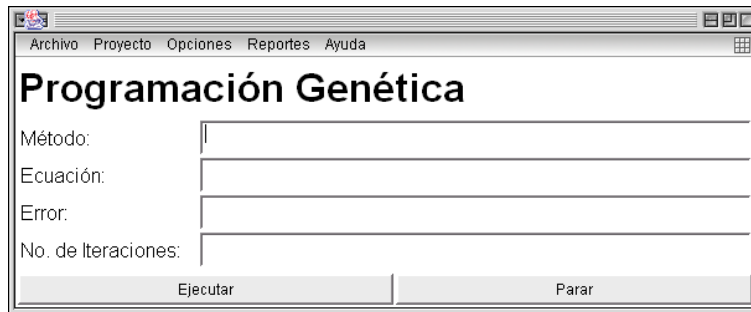


FIGURA 1. Interface gráfica del software Magister.

Copiar el programa `Magister.class` a un subdirectorio, por ejemplo:
`c:/Programas/`

Desde una ventana de MSDos ejecutar el comando de java:

```
c:\Programas\java Magister
```

En la figura 1 se observa una pantalla del programa Magister.

El primero paso a realizar es proporcionar al programa la información de forma adecuada para que la pueda leer. Los datos deben estar en un archivo de texto. Cada una de las variables dependientes están separadas por un espacio o un tabulador finalizando con la variable independiente.

Después es necesario fijar los valores de diferentes parámetros para el algoritmo genético, como son la probabilidad de cruza, de mutación, el método para generar el árbol de expansión, etc.

Y finalmente se ejecuta el programa.

5. El menú ARCHIVO

En el menú Archivo se puede cargar el archivo de texto mediante el submenú Abrir.

El submenú Salir termina la ejecución del programa.

6. El menú PROYECTO

En el menú Proyecto se encuentra el submenú Ejecutar que empieza a realizar los cálculos. Esta operación también se puede realizar a través del botón Ejecutar en la ventana principal.

El submenú Parar y el botón en la ventana principal con el mismo nombre termina la ejecución.

7. El menú OPCIONES

Este menú permite escoger los valores necesarios para utilizar un algoritmo genético.

El submenú Algoritmo Genético en la figura 2 tiene las siguientes opciones:

- Número de iteraciones: Número de veces que se ejecuta el algoritmo genético
- Terminar si el error es menor a: Condición para terminar el programa cuando se alcanza un error menor al especificado

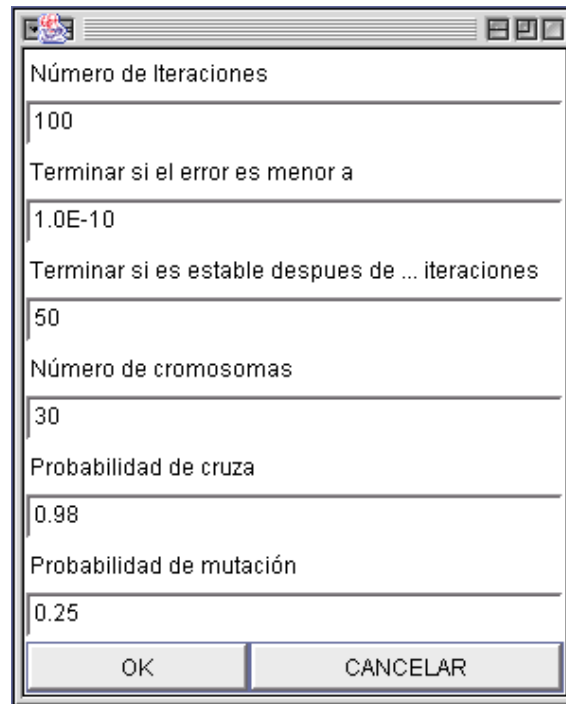


FIGURA 2. Controles para modificar el comportamiento del algoritmo genético en el submenú "Algoritmo Genético".

- Terminar si es estable después de ... iteraciones: Condición para terminar el programa si el error no cambia después del número de iteraciones especificado.
- Número de cromosomas: Número de individuos en la población
- Probabilidad de cruza
- Probabilidad de mutación

El submenú Árbol de expresión en la figura 3 fija los valores para crear la estructura de árbol. Tiene los siguientes puntos:

- Número de nodos en el árbol: El número de nodos en el árbol de expresión
- Número de variables independientes

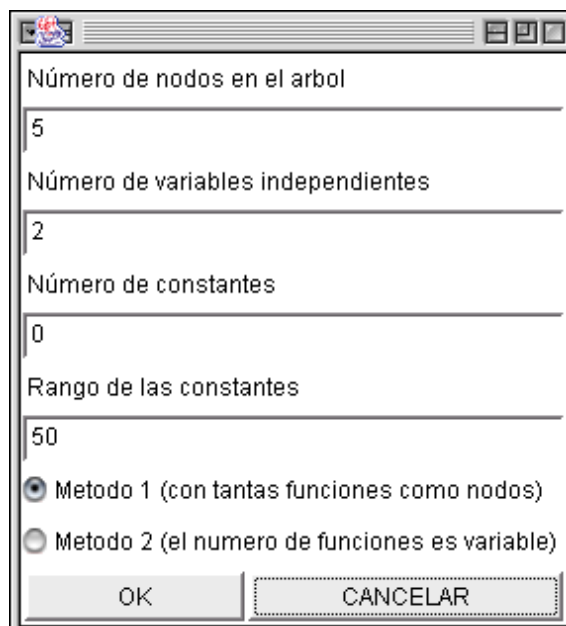


FIGURA 3. Controles para modificar las propiedades del árbol de expresión en el submenú "Árbol de expresión".

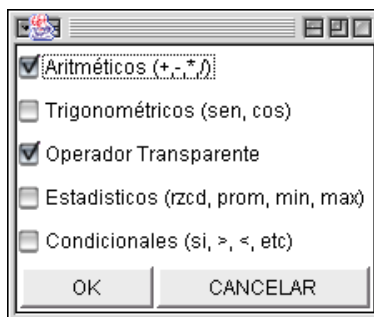


FIGURA 4. Controles para seleccionar el conjunto de operaciones a utilizar del submenú "Set de Operaciones".

- Número de constantes
- Rango de las variables
- Método: Método para generar el árbol

El submenú Set de Operaciones en la figura 4 permite escoger las operaciones o funciones que pueden ser utilizadas por el algoritmo genético.



FIGURA 5. Controles para seleccionar el tipo de codificación a utilizar del submenú "Método".

El submenú Método en la figura 5 permite escoger el método para generar el árbol de expresión. Se tiene cuatro métodos: Prüfer, Neville, Micikevicius y Picciotto.

8. Ejemplo

John Koza propuso obtener la tercera ley de Kepler mediante programación genética [Koza(1992)]. La tercera ley establece una relación entre el periodo orbital de los planetas y la distancia media de ellos hacia el Sol. En términos matemáticos, la ley se escribe:

$$P^2 = d^3$$

Donde P es el periodo orbital medido en años y d es la distancia media del planeta al Sol medida en Unidades Astronómicas que equivalen a 149,597,871km, que es la distancia de la Tierra al Sol.

La siguiente tabla muestra la relación existente entre P y d para los planetas del sistema solar.

Planeta	d (UA)	P (años)
Mercurio	0.38	0.24
Venus	0.72	0.61
Tierra	1.00	1.00
Marte	1.52	1.88
Júpiter	5.20	11.8
Saturno	9.56	29.6
Urano	19.3	84.8
Neptuno	30.2	166.6
Plutón	39.7	250.9

Para obtener la ecuación de la tercera ley de Kepler utilizando Magister, realice los siguientes pasos:

- Crear un archivo que contenga los datos de la siguiente forma:

```
0.387 0.240750084
0.723 0.61476261
1.000 1.00000000
1.524 1.881384018
5.203 11.86808748
9.569 29.60055349
19.309 84.84762177
30.284 166.6555788
39.781 250.9074427
```

- Abrir el archivo con el menú Archivo - Abrir · En el menú Algoritmo Genético introducir los siguientes datos:

Número de Iteraciones = 100

Número de Cromosomas = 20

Probabilidad de Cruza = 0.98

Probabilidad de Mutación = 0.2

- En el menú Árbol de Expresión introducir los datos siguientes:

Número de Operadores = 4

Número de Variables Independientes = 1

Número de Constantes = 0

- En el menú Set de Operaciones seleccionar: operadores aritméticos, operador transparente y operadores estadísticos.
- En el menú Método seleccionar cualquiera de los métodos, por ejemplo: Picciotto.
- Ejecutar el programa con el botón Ejecutar.

Bibliografía

- [Ankur Teredesai(2005)] V. G. Ankur Teredesai. *GP-based secondary classifiers*. Pattern Recognition, Volume 38, Issue 4, April 2005, Pages 505-512, 2005.
- [Barry Render(2004)] J. H. Barry Render. *Principios de Administracion de operaciones*. Pearson, Prentice Hall, 2004.
- [Brameier(2004)] M. Brameier. *On linear genetic programming*. PhD thesis, 2004.
- [Bures(1985)] M. E. Bures. *Metodos de pronosticos aplicados a la administracion*. Instituto Tecnológico y de estudios superiores de monterrey, 1985.
- [Byoung-Tak Zhang(2005)] H.-Y. J. Byoung-Tak Zhang. *Genetic programming: Molecular programming: evolving genetic programs in a test tube*. Proceedings of the 2005 conference on Genetic and evolutionary computation GECCO '05, 2005.
- [Celia C. Bojarczuk(2004)] A. A. F. E. L. M. Celia C. Bojarczuk, Heitor S. Lopes. *A constrained-syntax genetic programming system for discovering classification rules: application to medical data sets*. Artificial Intelligence in Medicine, Volume 30, Issue 1, January 2004, Pages 27-48, 2004.
- [Chidambaran(2003)] N. K. Chidambaran. *Risk analysis: New simulation methodology for risk analysis: genetic programming with monte carlo simulation for option pricing*. Proceedings of the 35th conference on Winter simulation: driving innovation, 2003.
- [D. M. Johnson(2005)] R. T. S. D. M. Johnson, A. M. Teredesai. *Genetic programming in wireless sensor networks*. GENETIC PROGRAMMING, PROCEEDINGS LECTURE NOTES IN COMPUTER SCIENCE, 2005.

- [D. Rivero(2005)] J. R. R. D. Rivero. *Time series forecast with anticipation using Genetic Programming*. COMPUTATIONAL INTELLIGENCE AND BIOINSPIRED SYSTEMS, PROCEEDINGS LECTURE NOTES IN COMPUTER SCIENCE, 2005.
- [Darwin(2002)] C. Darwin. *El origen de las especies*. CAYFOSA-QUEBECOR, 2002.
- [Eleazar Eskin(1999)] E. S. Eleazar Eskin. *Genetic programming applied to Othello: introducing students to machine learning research*. ACM SIGCSE Bulletin , The proceedings of the thirtieth SIGCSE technical symposium on Computer science education, 1999.
- [F. V. Buontempo(2005)] M. M. N. H. A. Y. D. O. F. V. Buontempo, X. Z. Wang. *Genetic programming for the induction of decision trees to model ecotoxicity data*. JOURNAL OF CHEMICAL INFORMATION AND MODELING, 2005.
- [F.E. Emery(1965)] E. T. F.E. Emery. *The causal texture of organization environments*. Human Relations, Vol. 18, 1965.
- [Ferreira(2001)] C. Ferreira. *Gene expression programming: a new adaptive algorithm for solving problems*. Complex Systems, Vol. 13, 87-129, 2001.
- [Greene(1998)] W. H. Greene. *Analisis Econometrico*. Prentice Hall. Tercera Edicion, 1998.
- [Grosan(2004)] C. Grosan. *Evolving mathematical expressions using genetic algorithms*. Paper, 2004.
- [Holland(1975)] J. Holland. *Adaptation in natural and artificial systems*. University of Michigan Press, 1975.
- [Huber(1986)] G. P. Huber. *The Decision-Making Paradigm of Organizational Design*. Management Science, Vol. 32, No. 5, 1986.
- [J. Koza(1997)] A. D. K. M. J. Koza, B. Forrest. *Evolution using genetic programming of a low-distortion, 96 decibel operational amplifier*. Proceedings of the 1997 ACM symposium on Applied computing, 1997.

- [J. Koza(1998)] A. D. K. M. J. Koza, B. Forrest. *Evolving computer programs using rapidly reconfigurable field-programmable gate arrays and genetic programming*. Proceedings of the 1998 ACM/SIGDA sixth international symposium on Field programmable gate arrays, 1998.
- [Jackson(2005)] D. Jackson. *Genetic programming: Parsing and translation of expressions by genetic programming*. Proceedings of the 2005 conference on Genetic and evolutionary computation GECCO '05, 2005.
- [James W. Dean(1996)] M. P. S. James W. Dean. *Does Decision Process Matter? A study of strategic decision-making effectiveness*. The Academy of Management Review, Vol. 39, No. 2, 1996.
- [Jeroen Eggermont(2004)] W. A. K. Jeroen Eggermont, Joost N. Kok. *Evolutionary computation and optimization (ECO): Genetic Programming for data classification: partitioning the search space*. Proceedings of the 2004 ACM symposium on Applied computing, 2004.
- [Jin-Hyuk Hong(2005)] S.-B. C. Jin-Hyuk Hong. *The classification of cancer based on DNA microarray data that uses diverse ensemble genetic programming*. Artificial Intelligence in Medicine, In Press, 2005.
- [John Hanke(2002)] A. R. John Hanke. *Pronosticos en los negocios*. Prentice Hall Hispanoamericana, S.A., 2002.
- [John Koza(2005)] L. W. J. John Koza, Sameer H. Al-Sakran. *Real world applications: Automated re-invention of six patented optical lens systems using genetic programming*. Proceedings of the 2005 conference on Genetic and evolutionary computation GECCO '05, 2005.
- [Joseph E. McCann(1984)] J. S. Joseph E. McCann. *Hyperturbulence and the emergence of type 5 environments*. The Academy of Management Review, Vol. 9, No. 3, 1984.
- [Keijzer(2002)] M. Keijzer. *Scientific discovery using genetic programming*. PhD thesis, 2002.

- [Kenneth Hennessy(2005)] J. C. A. G. R. Kenneth Hennessy, Michael G. Madden. *An improved genetic programming technique for the classification of Raman spectra*. Knowledge-Based Systems, Volume 18, Issues 4-5, August 2005, Pages 217-224, 2005.
- [Koza(1992)] J. Koza. *Genetic programming. On the programming of computers by means of natural selection*. The MIT Press, 1992.
- [M. Oltean(2003)] C. G. M. Oltean. *Evolving evolutionary algorithms using multi expression programming*. The 7th European Conference on Artificial Life, September 14-17, Dortmund, Edited by W. Banzhaf, 2003.
- [M. O'Neill(2001)] C. R. M. O'Neill. *Grammatical evolution*. IEEE Transaction on Evolutionary Computation, Vol 5, No 4, 2001.
- [Maddala(1996)] Maddala. *Introduccion a la econometria*. Prentice Hall. Segunda Edicion, 1996.
- [Nitin Muttill(2005)] J. H. L. Nitin Muttill. *Genetic programming for analysis and real-time prediction of coastal algal blooms*. Ecological Modelling, In Press, 2005.
- [Oren(2002)] N. Oren. *Research papers: data/knowledge management: Reexamining tf.idf based information retrieval with genetic programming*. Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology SAICSIT '02, 2002.
- [P. Raymond(2000)] C. C. P. Raymond, Steven C. Canale. *Numerical methods for engineers with programming and software applications*. Mc-Graw Hill, 2000.
- [Paul Massey(2005)] S. S. Paul Massey, John A. Clark. *Genetic programming: Evolution of a human-competitive quantum fourier transform algorithm using genetic programming*. Proceedings of the 2005 conference on Genetic and evolutionary computation GECCO '05, 2005.
- [Radford(1978)] K. J. Radford. *Decision-Making in a Turbulent Environment*. The Journal of the Operational Research Society, vol. 29, no. 27, 1978.

- [Roberto Hernandez Sampieri(2005)] P. B. L. Roberto Hernandez Sampieri, Carlos Fernandez Collado. *Fundamentos de metodologia de la investigacion*. Mc Graw Hill, 2005.
- [Roberts(2005)] M. C. Roberts. *Technical analysis and genetic programming: Constructing and testing a commodity portfolio*. JOURNAL OF FUTURES MARKETS, 2005.
- [Rolv Seehuus(2005)] O. E. Rolv Seehuus, Amund Tveit. *Biological applications: Discovering biological motifs with genetic programming*. Proceedings of the 2005 conference on Genetic and evolutionary computation GECCO '05, 2005.
- [S. Harding(2005)] J. F. M. S. Harding. *Evolution of robot controller using Cartesian Genetic Programming*. GENETIC PROGRAMMING, PROCEEDINGS LECTURE NOTES IN COMPUTER SCIENCE, 2005.
- [S. J. Kirstukas(2005)] D. A. A. S. J. Kirstukas, K. M. Bryden. *A hybrid genetic programming approach for the analytical solution of differential equations*. INTERNATIONAL JOURNAL OF GENERAL SYSTEMS, 2005.
- [Sancho Salcedo(2005)] M. J. S. C. B. Sancho Salcedo, Jose Luis Fernandez. *Genetic programming for the prediction of insolvency in non life insurance companies*. Computers and Operations Research, Volume 32, Issue 4, April 2005, Pages 749-765, 2005.
- [Spears(2000)] W. Spears. *Evolutionary algorithms. The role of mutation and recombination*. Springer, 2000.
- [W. Banzhaf(2001)] P. N. W. Banzhaf. *Genetic programming. An introduction on the automatic evolution of computer programs and its applications*. Morgan Kaufman Publishers, Inc., 2001.
- [Will Smart(2004)] M. Z. Will Smart. *Applying online gradient descent search to genetic programming for object recognition*. Proceedings of the second workshop on Australasian information security, Data Mining and Web Intelligence, and Software Internationalisation - Volume 32 CRPIT '04, 2004.

- [Wong(2001)] M. L. Wong. *A flexible knowledge discovery system using genetic programming and logic grammars*. Decision Support Systems, Volume 31, Issue 4, October 2001, Pages 405-428, 2001.
- [X. Tan(2005)] Y. L. X. Tan, B. Bhanu. *Fingerprint classification based on learned feature's*. IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS PART C-APPLICATIONS AND REVIEWS, 2005.
- [Zbigniew(1999)] M. Zbigniew. *Genetic algorithms + data structures = evolution programs*. Springer, 1999.