



Universidad Autónoma de Querétaro

Facultad de Informática

Maestría en Ingeniería de Sistemas de Información

**MODELO DE APOYO AL RAZONAMIENTO PROSPECTIVO EMPRESARIAL  
UTILIZANDO INTELIGENCIA ARTIFICIAL: CASO DE ESTUDIO  
EN UNA EMPRESA**

Que como parte de los requisitos para obtener el grado de  
Maestro en Ingeniería de Sistemas de Información

Presenta:

**M.A. José Alfredo Acuña García**

Dirigido por:

**M.I.S.D. Carlos Alberto Olmos Trejo**

SINODALES

M.I.S.D. Carlos Alberto Olmos Trejo  
Presidente






Mtra. Ruth Angélica Rico Hernández  
Secretario

Dra. Rosa María Romero González  
Vocal

Mtra. Elieth Velázquez Chávez  
Suplente

Mtro. Gerardo Rodríguez Rojano  
Suplente

  
Mtra. Ruth Angélica Rico Hernández  
Director de la Facultad

  
Firma  
  
Firma  
  
Firma  
  
Firma  
  
Firma

Dr. Luis Gerardo Hernandez Sandoval  
Director de Investigación y  
Posgrado

Centro Universitario  
Querétaro, Qro.  
Mayo de 2011  
México

## RESUMEN

La toma de decisiones práctica dentro de una empresa, es un proceso que normalmente se fundamenta en la información que está generando la propia empresa. Un tomador de decisiones utiliza la información con la que se cuenta en la empresa, de manera empírica e intuitiva, y en la medida de su conocimiento la relaciona con la información del entorno económico nacional en que se desenvuelve, y es así como cualquier empresario toma las decisiones que afectan el desarrollo futuro de la misma, en el mejor de los casos. Idealmente el tomador de decisiones debería contar con una mezcla de datos que reflejen la situación interna de la empresa, los factores externos, los eventos históricos que han hecho a la empresa, además de otros. Esta información debidamente procesada sería la base para apoyar adecuadamente la toma de decisiones. Por medio de metodologías y técnicas informáticas podemos acercarnos a comprender los patrones de comportamiento interno de una entidad organizacional a través de sus datos históricos, y proyectar posibles comportamientos a futuro dentro del ámbito de los datos conocidos.

La presente tesis propone un modelo prospectivo de apoyo a la toma de decisiones fundamentado en la utilización de minería de datos, clustering e inteligencia artificial, que no solo aporta un resultado para la propia toma de decisiones, sino que además permita la razonabilidad del proceso que emitió este resultado, y que en consecuencia fortalezca la seguridad en la decisión que ha de tomar el responsable de la misma. Aunque parece de sentido común, es necesario mostrar la necesidad de utilizar la información de una empresa de manera inteligente, alimentando un modelo de solución adecuado, y no solo que sea información que se archive en el olvido sin tener un papel en la evolución, experiencia y sobrevivencia de una empresa y proponer un modelo que facilite (y en un futuro) automatice su aplicación.

**(Palabras clave:** minería de datos, prospectiva, empresa)

## SUMMARY

A decision maker uses the information that the company has in an empirical and intuitive form, and in the possibility of their knowledge, he relates information with national economic environment in which it operates. Ideally the decision maker should have a mix of data that reflects the internal situation of the company, external factors, and the historical events that have formed. This thesis proposes a prospective model to support decision making based on the use of data mining, clustering and artificial intelligence, which not only provides a result for the decision itself, but also allows the fairness of the process gave this result, and consequently strengthen security in the decision that has to take responsibility for it.

**(Keywords:** data mining, prospective, industry)

## **DEDICATORIA**

A la **Universidad Autónoma de Querétaro**

A mi esposa Gris y mi hija Lizbeth

A mi asesor y amigo Carlos Olmos Trejo

## AGRADECIMIENTOS

Ser agradecido es una actitud de gratitud honesta y de aprecio a quien ha hecho algo por ti. Es además reconocer que sin los demás es imposible lograr aspiraciones o metas.

- A mi Dios por darme la vida y todas mis oportunidades.
- A mi esposa e hija por su comprensión, apoyo y paciencia durante mis estudios e investigación de tesis.
- Sinceramente mi agradecimiento por el apoyo incondicional del M.I.S.D. Carlos Alberto Olmos Trejo en la elaboración de esta investigación.
- Un Agradecimiento especial a los profesores Mtra. Ruth Rico Hernández, Dra. Rosa María Romero González, Mtra. Elieth Velázquez Chávez y al Mtro. Gerardo Rodríguez Rojano, por su apoyo y asesoría durante la formación de la presente tesis.
- A mis compañeros de generación agradezco su interacción y enseñanza durante este proceso de formación.
- A la Universidad Autónoma de Querétaro por ser la institución que permite y da carácter a mi formación.

# ÍNDICE

<b>RESUMEN</b>	<b><i>i</i></b>
<b>SUMMARY</b>	<b><i>iii</i></b>
<b>DEDICATORIA</b>	<b><i>iv</i></b>
<b>AGRADECIMIENTOS</b>	<b><i>v</i></b>
<b>ÍNDICE</b>	<b><i>vi</i></b>
<b>ÍNDICE DE TABLAS Y GRÁFICAS</b>	<b><i>x</i></b>
<b>ÍNDICE DE FIGURAS</b>	<b><i>xii</i></b>
<b>I.INTRODUCCIÓN</b>	<b><i>1</i></b>
<b>II.ANTECEDENTES</b>	<b><i>4</i></b>
1. Marco Histórico	<b><i>4</i></b>
2. Trabajos Relacionados	<b><i>6</i></b>
<b>III.ORGANIZACIÓN EMPRESARIAL Y TECNOLÓGICA</b>	<b><i>9</i></b>
1. Procesos de una empresa	<b><i>9</i></b>
La Planeación	<b><i>9</i></b>
La Metodología	<b><i>14</i></b>
El Plan	<b><i>19</i></b>
Análisis FODA	<b><i>23</i></b>
2. Madurez de sus procesos informáticos	<b><i>23</i></b>
3. Proceso de la toma de decisiones	<b><i>30</i></b>
4. Aplicación de Herramientas de inteligencia en los negocios.	<b><i>36</i></b>
5. Utilidad de la prospectiva empresarial.	<b><i>39</i></b>

<b>IV.ANTECEDENTES DEL OBJETO DE ANÁLISIS</b>	<b>41</b>
1. Acuerdo de Privacidad	41
2. Contexto empresarial	42
3. Entorno informático	43
4. Obtención de datos	44
<b>V.ANÁLISIS DE TÉCNICAS INFORMÁTICAS</b>	<b>46</b>
1. Técnicas para el análisis de datos por clústeres	46
3. Selección de Herramientas informáticas	49
a. Investigación de aplicaciones	49
b. Evaluación comparativa	50
c. Selección de herramientas y justificación	57
<b>VI.ANÁLISIS PRELIMINAR DE DATOS</b>	<b>59</b>
1. 1ª selección de datos a analizar	59
2. Análisis de gráficos y patrones	60
4. Análisis de pertinencia de estandarización	61
5. Datos: Estandarización	62
6. Selección final de datos por modelar	63
7. Datos: Análisis de singularidades	64
8. Pertinencia de Análisis bayesiano y/o de correlaciones	64
9. Análisis grafico de clústeres en tres dimensiones	65
10. Análisis de la elección de Clústeres	70
<b>VII.DEFINICIÓN Y CONSTRUCCIÓN DEL MODELO</b>	<b>72</b>
1. Antecedentes similares y resultados	72
<i>“An Evolutionary Modularized Data Mining Mechanism for Financial Distress Forecasts”</i>	72
<i>“DATA MINING FOR FINANCIAL APPLICATIONS”</i>	73

<i>"Using genetic algorithms to optimize nearest neighbors for data mining"</i>	77
<b>2. Definición de modelo</b>	<b>80</b>
Diagrama General del modelo:	81
Modelo detallado en Pseudocódigo:	81
<b>3. Construcción con aplicaciones seleccionadas</b>	<b>83</b>
<b>VIII.APLICACIÓN DEL MODELO</b>	<b>84</b>
<b>1. Consideraciones de la aplicación</b>	<b>84</b>
<b>2. Caso 1.- Aplicación del modelo con datos históricos</b>	<b>86</b>
<b>3. Caso 2.- Aplicación del modelo con datos incrementales</b>	<b>86</b>
<b>4. Caso 3.-Aplicación del modelo de forma prospectiva</b>	<b>87</b>
<b>IX.ANÁLISIS DE RESULTADOS</b>	<b>88</b>
<b>1. Caso 1.- Análisis de la aplicación del modelo con datos históricos</b>	<b>88</b>
<b>2. Criterios de validación del modelo</b>	<b>92</b>
<b>3. Caso 2.- Análisis de la aplicación del modelo con datos incrementales</b>	<b>97</b>
<b>4. Caso 3.-Análisis de la aplicación del modelo de forma prospectiva</b>	<b>98</b>
<b>5. Conclusiones del análisis</b>	<b>100</b>
<b>X.CONCLUSIÓN</b>	<b>103</b>
<b>1. Conclusión</b>	<b>103</b>
<b>2. Proyección</b>	<b>105</b>
<b>XI.BIBLIOGRAFÍA</b>	<b>107</b>
<b>XII.APÉNDICE A</b>	<b>110</b>
<b>XIII.APÉNDICE B</b>	<b>115</b>
<b>XIV.APÉNDICE C</b>	<b>118</b>
<b>XV.APÉNDICE D</b>	<b>121</b>



Programación en la aplicación Wolfram Mathematica 7.0 (TrialV)	121
<b>XVI.APÉNDICE E</b>	<b>123</b>
Análisis del no de clústeres y cercanos con menos error con resultados conocidos.	123
<b>XVII.APÉNDICE F. CASO 1</b>	<b>126</b>
Resultados Pronóstico 2009/07	126
Resultados Pronóstico 2009/08	132
Resultados Pronóstico 2009/09	138
Resultados Pronóstico 2009/10	144
Resultados Pronóstico 2009/11	150
Resultados Pronóstico 2009/12	156
Resultados Pronóstico 2009/07-12 Incrementales	162
<b>XVIII.APÉNDICE G</b>	<b>165</b>
Resultados Pronóstico 2010/01	165
Resultados Pronóstico 2010/02	168
Resultados Pronóstico 2010/03	171
<b>XVI.GLOSARIO</b>	<b>174</b>

## ÍNDICE DE TABLAS Y GRÁFICAS

<i>Tabla 6.1. Muestra del proceso de estandarización de datos.</i>	62
<i>Gráfica 6.2. Vectores de datos estandarizados de diciembre/2009.</i>	66
<i>Gráfica 6.3. Gráfica de clústeres de diciembre/2009.</i>	66
<i>Gráfica 6.4. Gráfica de agrupaciones de 1 a 12 clústeres de diciembre/2009.</i>	67
<i>Gráfica 6.5. Datos del agrupamiento en 6 clústeres de diciembre/2009.</i>	69
<i>Gráfica 6.6. Gráficas del agrupamiento en 6 clústeres de diciembre/2009.</i>	70
<i>Gráfica 6.7. Gráfica del agrupamiento en 6 clústeres de diciembre/2009.</i>	70
<i>Tabla 7.3. Clasificación de eventos.</i>	83
<i>Tabla 8.1. Análisis sumario de errores porcentualizados.</i>	85
<i>Gráfica 9.1. Análisis sumario de errores porcentualizados.</i>	89
<i>Tabla 9.2. Análisis sumario de errores porcentualizados.</i>	91
<i>Tabla 9.3. Eventos de los vecinos cercanos en los estados N y N-1.</i>	92
<i>Tabla 9.4. Eventos de los vecinos cercanos en el estado N+1.</i>	92
<i>Tabla 9.5. Análisis estadístico de los datos históricos.</i>	93
<i>Gráfica 9.6. Histograma de ventas a clientes.</i>	93
<i>Tabla 9.7. Descripción de clases del Histograma de ventas a clientes.</i>	94
<i>Gráfica 9.8. Histograma de valor del almacén.</i>	94
<i>Tabla 9.9. Descripción de clases del Histograma de valor del almacén.</i>	95
<i>Tabla 9.10. Análisis estadístico de los datos históricos.</i>	96
<i>Tabla 9.11. Análisis sumario del proceso de la prospección incremental.</i>	98
<i>Tabla 9.12. Eventos de los vecinos cercanos en los estados N y N-1.</i>	99
<i>Tabla 9.13. Eventos de los vecinos cercanos en el estado N+1.</i>	99
<i>Tabla 9.14. Eventos de los vecinos cercanos en los estados N y N-1.</i>	100
<i>Tabla A.1. Tablas de datos Originales (muestra)</i>	113
<i>Tabla A.2. Descripción total de campos.</i>	114
<i>Tabla B.1. Ventas de la empresa de estudio.</i>	115
<i>Gráfica B.2. Ventas de la empresa de estudio.</i>	115
<i>Gráfica B.3. Anual de ventas de la empresa de estudio.</i>	116
<i>Tabla B.4. Valor del almacén de la empresa de estudio.</i>	116

<i>Gráfica B.5. Valor del almacén de la empresa de estudio.</i>	117
<i>Gráfica B.6. Anual del valor del almacén de la empresa de estudio.</i>	117
<i>Gráfica C.1. Proceso de Estandarización de datos.</i>	118
<i>Gráfica C.2. Datos estandarizados de ventas.</i>	118
<i>Gráfica C.3. Datos estandarizados de valor del almacén.</i>	119
<i>Gráfica C.4. Gráfica de datos estandarizados de ventas.</i>	119
<i>Gráfica C.5. Gráfica de datos estandarizados de valor de almacén.</i>	120
<i>Código D.1. Programa en de procesamiento de datos.</i>	122
<i>Tabla E.1a. Análisis del número adecuado de vecinos cercanos.</i>	123
<i>Tabla E.1b. Análisis del número adecuado de vecinos cercanos.</i>	124
<i>Tabla E.1c. Análisis del número adecuado de vecinos cercanos.</i>	124
<i>Tabla E.2. Análisis sumario de errores porcentualizados.</i>	125
<i>Tabla E.3. Análisis de lecturas de cercanos independientes.</i>	125
<i>Tabla F1. Análisis porcentual del proceso de pronóstico incremental.</i>	163
<i>Tabla F.2. Análisis sumario del proceso incremental.</i>	164

## ÍNDICE DE FIGURAS

<i>Figura 3.1. Tecnologías de Información y Sistemas de Información.</i>	11
<i>Figura 3.2. Modelos de gobernanca</i>	12
<i>Figura 3.3. Gobernanca legislativa.</i>	13
<i>Figura 3.4. Gobernanca ejecutiva.</i>	14
<i>Figura 3.5. Modelo de Gobernanca.</i>	14
<i>Figura 3.6. Modelo de planeación estratégica en sistemas.</i>	15
<i>Figura 3.7. Metodologías de impacto y de alineamiento.</i>	15
<i>Figura 3.8. Análisis de la cadena de Valor.</i>	16
<i>Figura 3.9. Factores Críticos de Éxito.</i>	17
<i>Figura 3.10. Factores Críticos de Éxito.</i>	18
<i>Figura 3.11. Integración de la ventaja Competitiva.</i>	20
<i>Figura 3.12. Fuentes de la ventaja Competitiva.</i>	21
<i>Figura 3.13. Impacto de la ventaja Competitiva.</i>	22
<i>Figura 3.14. Impacto de la ventaja Competitiva.</i>	23
<i>Figura 3.16. Esquema de una empresa.</i>	24
<i>Figura 3.17. Esquema de necesidades a cubrir.</i>	27
<i>Figura 3.18. Portafolio de Servicios.</i>	28
<i>Figura 3.19. Inversión en TI en México.</i>	29
<i>Figura 3.20. Proceso de toma de decisiones.</i>	32
<i>Figura 3.21. Bestialismo en las decisiones.</i>	34
<i>Figura 3.22. Veleidad en las decisiones.</i>	34
<i>Figura 3.23. Intelectualismo en las decisiones.</i>	34
<i>Figura 3.24. Reglas y formación del Decisor.</i>	35
<i>Figura 3.25. Modelo de Carnegie.</i>	35
<i>Figura 3.26. Integración informática.</i>	36
<i>Figura 3.27. Niveles de una empresa.</i>	39
<i>Figura 5.1. Proceso con SQL Business Intelligence.</i>	50
<i>Figura 5.2. Indirectos del proceso con SQL Business Intelligence.</i>	51
<i>Figura 5.3. Indirectos del proceso con SQL Business Intelligence.</i>	51

<i>Figura 5.4. Análisis de indirectos con Weka.</i>	53
<i>Figura 5.5. Ambiente de trabajo con Wolfram Mathematica.</i>	56
<i>Figura 7.1. Modelo de optimización k-NN.</i>	78
<i>Figura 7.2. Modelo del sistema de apoyo al razonamiento prospectivo.</i>	81
<i>Figura 9.1. Vecindad más cercana a la posición 90.</i>	90
<i>Figura 9.2. Posiciones de vecindad a los estados N y N+1.</i>	90
<i>Figura 9.3. Posiciones más cercanas en los estados siguientes.</i>	91
<i>Figura 9.4. Resultado prospectivo con 3 cercanos.</i>	91

## I. INTRODUCCIÓN

Inscritas en la sociedad, como un gran organismo vivo, existen entidades sociales integradas por personas, organizadas mediante convenios que establecen propósitos comunes. Durante su crecimiento se plantea la necesidad de planear y controlar la dirección que tomara a futuro, los objetivos y metas que definirán su desarrollo y permanencia en el mercado.

Una necesidad intrínseca a este crecimiento, es comprender mediante representaciones numéricas la salud que guarda a cada paso. Con este propósito se utilizan sistemas informáticos que facilitan la operación de la empresa, a la vez que permiten un registro de las operaciones detalladas y acumuladas diarias.

Las herramientas informáticas facilitan los registros durante las operaciones, a la vez que apoyan los procesos que dan vida a la empresa, y se encuentran presentes desde las capas operacionales que realizan el trabajo físico productivo, hasta las capas estratégicas pasando por las administrativas de mandos medios. De esta manera se crea una simbiosis entre la operación de la empresa y los mecanismos informáticos implantados en la misma, de manera tal que es difícil en la actualidad concebir el funcionamiento de uno sin el otro.

La exploración del comportamiento de las organizaciones sociales, es un tema que ha generado múltiples estudios desde diferentes ópticas. Comprender sus patrones de comportamiento como una base para pronosticar la forma en que la organización se desarrollara en un periodo de tiempo futuro cercano, en un momento en particular es uno de los escenarios deseable en la actualidad.

Existen ámbitos dentro de la empresa, que por la necesidad de registro de información son más accesibles para su análisis en cuanto a su comportamiento, debido principalmente a que existe un registro histórico de los eventos que los definen. De esta manera se puede contar con información inicial que describe su evolución. Contar con datos de un lapso de tiempo adecuado registrados y de alguna manera accesible al análisis da las facilidades para iniciar un análisis de comportamiento.

El principal objetivo del presente trabajo es construir un modelo fundamentado en la utilización de minería de datos cuyo resultado proporcione elementos de razonamiento que sean la base de los procesos de toma de decisiones para la empresa, utilizando un acumulado organizado de información de los casos presentados en su historia financiero-organizacional.

Con este propósito principal se realiza inicialmente un análisis comparativo de herramientas informáticas de aplicación a la minería de datos, buscando el reconocimiento de patrones de comportamiento aplicando tecnologías informáticas de inteligencia artificial y minería de datos a información de la pequeña y mediana empresa nacional.

En el proceso se define un modelo básico de apoyo al razonamiento prospectivo como guía de razonamiento en apoyo a un tomador de decisiones empresarial.

La aplicación del modelo requirió de su construcción, y de la obtención de datos objetivo, que se acordó su utilización de manera confidencial con una empresa del sector productivo de Querétaro.

En el análisis de los resultados se realiza la consulta de ejecutivos de la empresa con la finalidad de visualizar la aplicación positiva del modelo y considerando sus aportaciones. De esta forma se obtiene un pronóstico dirigido

aplicando el modelo a los datos caso de estudio relacionado con proyecciones y comportamiento empresarial a nivel nacional.

Se trabaja con el supuesto de que los datos generados por una empresa, permiten un pronóstico válido, fundamentado en información organizada y procesada adecuadamente mediante tecnologías informáticas y algoritmos de inteligencia artificial. Además esta información debe permitir la toma de decisiones fundamentadas, y la sana evolución organizacional y financiera de la empresa.

Un estudio de este tipo, con información real conlleva la responsabilidad de mantener la privacidad y anonimato de la empresa objeto de estudio, por lo que, por acuerdo tomado con los ejecutivos de la misma, no serán detallados datos que sugieran su identificación.

Trabajar con datos reales da una importancia que se debe valorar en análisis de los resultados y aportes que tiene el presente trabajo.

Comprender que existen información que se maneja cotidianamente pero que no pasan de ser un registro en el mejor de los casos consultado por el tomador de decisiones como referencia, sin una metodología que le proporcione inicialmente una validación de su contribución al bienestar financiero de la empresa. Además es importante procesar la información comparativamente con los eventos financieros y organizacionales pasados sugiriendo elementos similares, y tomándolos como base para un futuro planteamiento de pronósticos.

La importancia de poder determinar un modelo útil para el análisis de la empresa y que permita sugerencias basados en escenarios reales pasados, y contar con una herramienta metodológica que sistematice el proceso es una aportación importante para el tomador de decisiones.



## II. ANTECEDENTES

### 1. Marco Histórico

Conocer aunque solo sea un atisbo del futuro, siempre ha sido un empeño humano en las diferentes áreas del quehacer humano. En el área de los negocios, donde la ventaja competitiva principalmente reside en las decisiones que toma el líder de la empresa, y que dependiendo de su asertividad desarrolla o frena la actividad de la misma, conocer proyectivamente algunos de los factores que son afectados por sus decisiones y que esto le permita tomarlas con más elementos informativos es lo deseable.

Debido a que no podemos conocer con total certeza cuál será el comportamiento de las diferentes variables que afectan el desarrollo de una empresa, modelando podemos determinar algunas variables que son más importantes en función al giro de la empresa en cuestión, y que pueden guardar comportamientos cíclicos por periodos de tiempo establecidos.

Siempre que un líder realiza la planeación de alguna empresa, proyecta a futuro, hace prospectiva. Del real diccionario de la lengua española, prospectiva es:

- I. *adj. Que se refiere al futuro.*
- II. *f. Conjunto de análisis y estudios realizados con el fin de explorar o de predecir el futuro, en una determinada materia.*
- III. *La palabra prospectiva viene del griego 'prospekt' y significa el modo de mirar algo.*

Los primeros estudios en prospectiva se realizaron en Estados Unidos en la década de los años veinte. Fueron abandonados a consecuencia de la

grave depresión sufrida por este país en 1929 y la segunda guerra mundial que le siguió.

Años más tarde, Japón, completamente devastado, reinició los estudios de prospectiva como una alternativa de solución y mejora de sus condiciones sociales, industriales, económicas, etc. Varios años de perfeccionamiento le permitieron a este país oriental recobrar el tiempo perdido y convertirse en una de las naciones más poderosas a nivel mundial.

"En este nuevo mundo, la información reina..." afirma Geoffrey A. Moore, Director de Chasm Group.

En el año 1989 Howard Dresner, Un becario de investigación en Garter Group (una empresa de investigación de TI y de asesoramiento en Connecticut) popularizó el término de "Inteligencia de Negocios" (BI - Business Intelligence) como un conjunto de métodos y conceptos para mejorar la toma de decisiones empresariales mediante el uso de un recurso de la empresa, llamado "Datos".

EL siglo XX y XXI se conoce como el siglo de la información y la tecnología (IT), la edad en la que todo depende de la disponibilidad de la información y la innovación de nuevas tecnologías. Esto es así en el caso de las empresas que trabajan con innovaciones tecnológicas, como la World Wide Web e Internet que ha proporcionado una plataforma amplia y abierta a los negocios que se realizan en todo el mundo.

Como el desarrollo de cualquier estrategia de negocios exitosa requiere información y conocimiento adecuado sobre el estado del mercado, la innovación de nuevas tecnologías y conceptos como el de inteligencia de negocios, facilita el acceso y manejo de todo tipo de información en los negocios.

Antes de la llegada de la era de la información, era muy difícil acceder a ella en ausencia de métodos de computación y así todas las decisiones

empresariales se basaban principalmente en la intuición. Pero en el mundo globalizado de hoy la aplicación de las modernas tecnologías de Inteligencia a las empresas se utiliza de manera muy general y poco aterrizada en sus problemáticas particulares.

Como toda organización grande o pequeña requiere de información para promover su negocio en previsión de las tendencias futuras, la información es ahora la principal herramienta para entender las tendencias del mercado y entender su propia posición de mercado y a sus competidores. Es un término amplio asociado con diferentes tecnologías modernas que se refieren a las tendencias cambiantes del mercado, comportamiento de los clientes, la demanda y la cadena de suministro, y varios otros aspectos del negocio para mejorar el rendimiento de la empresa [Laudon, 2007].

## **2. Trabajos Relacionados**

La toma de decisiones practica dentro de una empresa mediana y pequeña, es un proceso que normalmente se fundamenta en la información que está generando la propia empresa, utilizándola de manera empírica e intuitivamente relacionada con la información del ambiente nacional, poco basado en la información actual de la empresa, y de los eventos históricos que conforman su evolución.

Existe información en la empresa que se maneja comúnmente para el registro financiero, contable y fiscal, que conforma su historia, y que siendo utilizada metodológica e informatizadamente, es una base solida para la toma de decisiones y el crecimiento sano de la misma. [Musalem. 2008].

Decidir la información que debe ser la base para la toma de decisiones en una entidad (como instancia también en una empresa) es uno de los elementos importantes a definir en el proceso de construcción de un modelo

[Liang-Hsuan, 2008]. Existen herramientas matemáticas que permitan orientar esta definición. [Po-Chang, 2007].

Poder utilizar la información histórica de una entidad por los eventos pasados, extrayendo de esta historia patrones esenciales para realizar un pronóstico de comportamiento en el área de información financieros-organizacionales es la base para poder determinar el comportamiento que tiene actualmente y establecer un pronóstico confiable de su comportamiento futuro. [Kovalerchuk, 2006].

Si la entidad es una empresa y se cuenta con la información que ha conformado su evolución, se puede construir en un Sistema de Razonamiento Basado en Casos (Case Based Reasoning, CBR). Esta base de casos representa un valor importante para comparar evolutivamente el crecimiento de la empresa y por comparar elementos similares poder pronosticar su comportamiento futuro y proponer escenarios previsibles como apoyo a la toma de decisiones. [Hyunchul, 2008].

Analizar entidades de negocios ha sido materia de estudios académicos durante las últimas tres décadas, ocupando en diferentes modelos de decisión como los CBR, tecnologías que inician desde las puramente estadísticas, hasta las orientadas a la aplicación de la inteligencia artificial, como son Minería de datos (Data Mining, DM), Algoritmos Genéticos (Genetic Algorithms, GA), Vecinos Cercanos (Nearest Neighbors, NN), Predicción Híbrida de Fallas (Hybrid Failure Prediction, HFP), Teoría del Conjunto Rough (Rough Set Theory, RST), Análisis Relacional Grey (Grey Relational Analysis, GRA) y demás técnicas utilizadas en esta área [Rong-Ho, 2009] [Hyunchul, 2008].

Analizar los posibles escenarios buscando optimizar la propuesta en base a consideraciones económicas temporales del entorno empresarial, utilizando técnicas de inteligencia artificial para buscar la aproximación a la mejor sugerencia para la toma de decisiones. [Hyunchul, 2008].

Es importante poder presentar un modelo que no solo aporte un resultado para la toma de decisiones, sino que además permita la razonabilidad del proceso que emitió este resultado, y que en consecuencia fortalezca la seguridad en la decisión que ha de tomar el responsable de la misma o del asesor que realiza el análisis y la sugerencia de decisión. [Kovalerchuk, 2006].

Aunque parece de sentido común, es necesario mostrar la necesidad de utilizar la información de una empresa de manera inteligente, alimentando un modelo de solución adecuado, y no solo que sea información que se archive en el olvido sin tener un papel en la evolución, experiencia y sobrevivencia de una empresa y proponer un modelo que facilite (y en un futuro) automatice su aplicación.

### III. ORGANIZACIÓN EMPRESARIAL Y TECNOLÓGICA

Al analizar la entidad llamada empresa, debemos comprender como es que este organismo vive, y en base a qué mecanismos se desarrolla, para lo que es conveniente puntualizar estos procesos y la vinculación que tienen con los procesos informáticos que se integran en esa entidad.

#### 1. Procesos de una empresa

##### La Planeación

Es el proceso mediante el cual se prepara la organización para conseguir los objetivos y metas de negocio establecidas según su análisis prospectivo y pronóstico. Existen tres tipos de planeación:

- Planeación a Largo Plazo (plan estratégico): Se considera que es a un término de 5 a 10 años.
- Plan Maestro: Establece el *portafolio de aplicaciones* que soportarán el plan estratégico.
- Planes Tácticos: Son los que se utilizan para periodos de 1 año y definen Presupuestos y Plazos específicos para llevar a cabo los proyectos.

El plan estratégico de sistemas de información se entiende, en términos prácticos, como la forma en la que se dispondrán los recursos y se tomarán las decisiones en el área de sistemas de una organización, con el objeto de mantenerlos alineados con la estrategia de la misma; es decir, la manera en la que el área de informática va a contribuir para el logro de los objetivos de la organización o, en una organización madura, convertirse

eventualmente en una ventaja competitiva para ésta, dentro de la cadena de valor definida [Porter, 1996].

En otras palabras, esta es una capa adicional de administración de la tecnología que se encarga de traducir las necesidades y prioridades del negocio en requerimientos técnicos del área de información con el objeto de dar certeza a las inversiones en arquitectura, minimizar riesgos y apoyar el crecimiento sustentable de la organización.

La importancia de la planeación depende de la industria y las circunstancias del entorno. Se suele decir que cuando el negocio va bien, la planeación no se considera tan necesaria (error de consideración), pero cuando el negocio va mal, la planeación es crítica. Como ejemplo se tiene que:

- Las industrias de servicios requieren más planeación e infraestructura que las de manufactura, especialmente las de servicios financieros
- La volatilidad del mercado, las empresas que renuevan sus productos más constantemente, requieren más planeación.
- El grado de sinergia, las empresas que enfatizan la sinergia en su modelo de negocio, son más susceptibles a la planeación
- Integración, las empresas que tienen un mayor grado de integración de IT en la organización para el seguimiento y monitoreo de los alcances de objetivos, tienen mayor planeación [Laudon, 2007].

De aquí se desprende que se tenga que hacer la diferencia entre 'Tecnologías de Información (TI)' y 'Sistemas de Información (SI)'. El primer término se refiere a la entrega de servicios de TI o la operación de los equipos de tecnología y el segundo se refiere a la demanda de servicios y traducción de necesidades del negocio en requerimientos de tecnología, según se puede ver en la Figura 3.1:

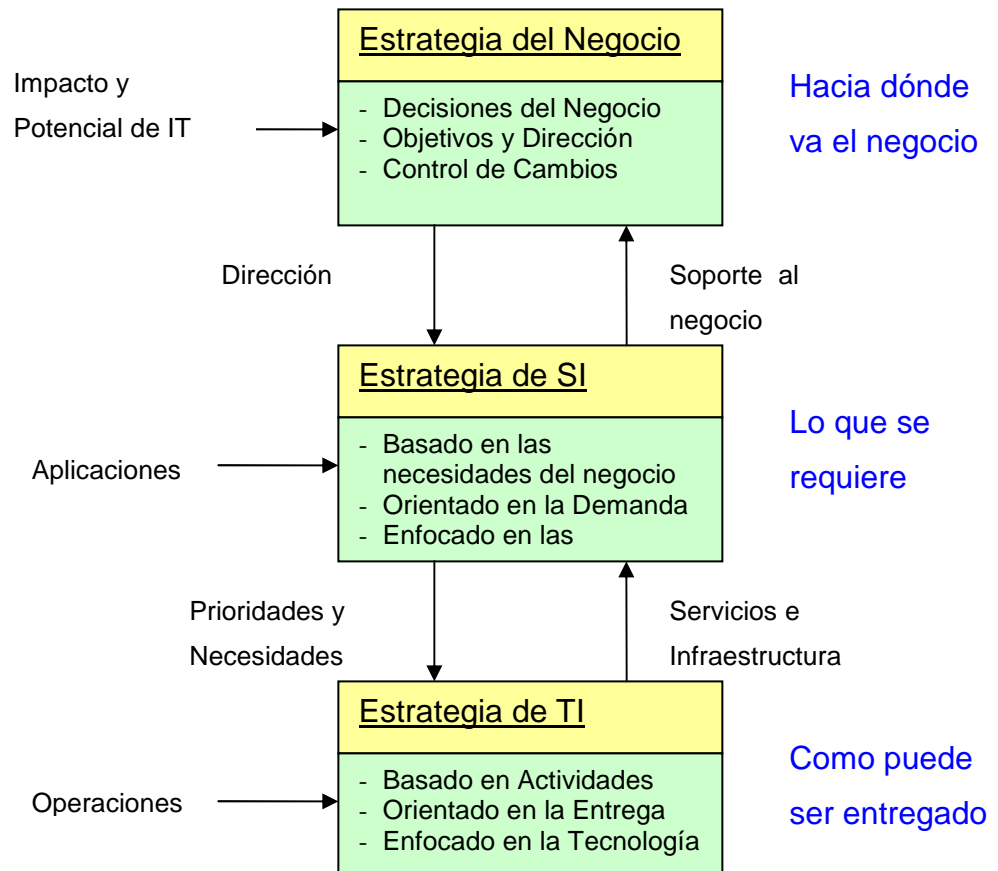


Figura 3.1. Tecnologías de Información y Sistemas de Información.

Basados en el diagrama de la Figura 3.1 se puede decir que los roles de cada una de las áreas son:

- Junta de Dirección: Definir el rumbo del negocio.
- SI: Identificar y Priorizar las necesidades del negocio para traducirlas en términos de tecnología
- TI: Llevar a cabo la ejecución de proyectos y mantener el portafolio de servicios, es decir, operar la infraestructura que apoye los procesos de negocios.



Sin embargo falta una parte que defina los roles, responsabilidades, así como la interacción entre quienes tomarán las decisiones sobre asuntos de tecnología en la organización, es decir: La gobernanza de los sistemas. He aquí los diferentes modelos de gobernanza:

<b>Modelo de Gobernanza</b>	<b>Características</b>
Monarquía del Negocio	La junta directiva toma todas las decisiones en tema de tecnología, y esas decisiones se llevan a toda la empresa
Monarquía de TI	Los expertos de tecnología toman de manera autónoma todas las decisiones
Feudal	Las distintas unidades de negocio, las regiones o funciones toman las decisiones de tecnología en su área de responsabilidad
Federal	Las decisiones se toman de manera coordinada entre la unidad de negocios y el corporativo con representantes de ambos
Duopolio	Las decisiones se toman de manera bilateral entre los expertos de TI y alguna otra entidad de la organización

Figura 3.2. Modelos de gobernanza

La gobernanza se compone de 2 partes: la Legislativa y la Ejecutiva. La parte legislativa define y decide lo que se debe hacer y la ejecutiva se encarga de materializarlo y administrarlo.

La parte legislativa mostrada en la Figura 3.3 se debe integrar por miembros representativos de la Junta de Dirección (EB), Dirección General (CEO) Finanzas (CFO), Operaciones (COO), Informática (CIO), Auditoría (CIA), Gerente de Servicios (SM) responsable de negociar la calidad del servicio o

SLA, así como de cada uno de los responsables de cada función o proceso de negocio, en caso de una organización matricial enfocada a procesos.

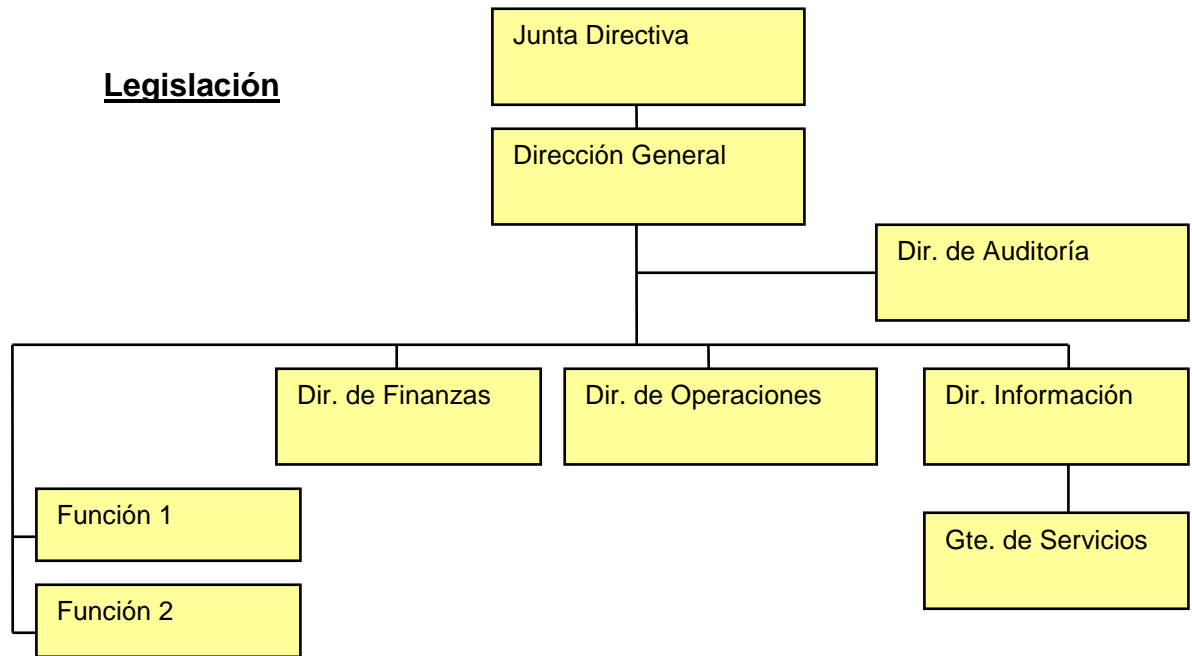


Figura 3.3. Gobernanza legislativa.

Por otra parte, la gobernanza ejecutiva mostrada en la Figura 3.4, se compone del área de informática (CIO), proyectos (PM), servicios (Service Manager), seguridad (SIO), Administrador de BD (DBA), Administrador de Red (NWA) y el responsable de cada una de las aplicaciones involucradas (Application Owner).

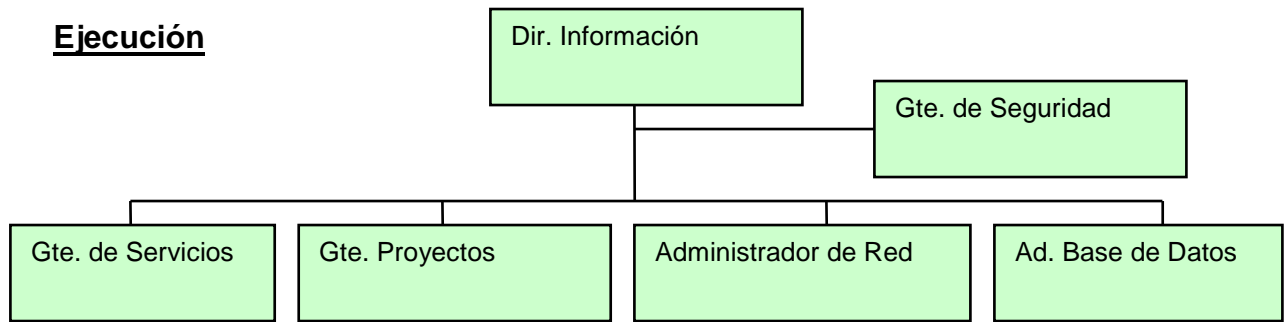


Figura 3.4. Gobernanca ejecutiva.

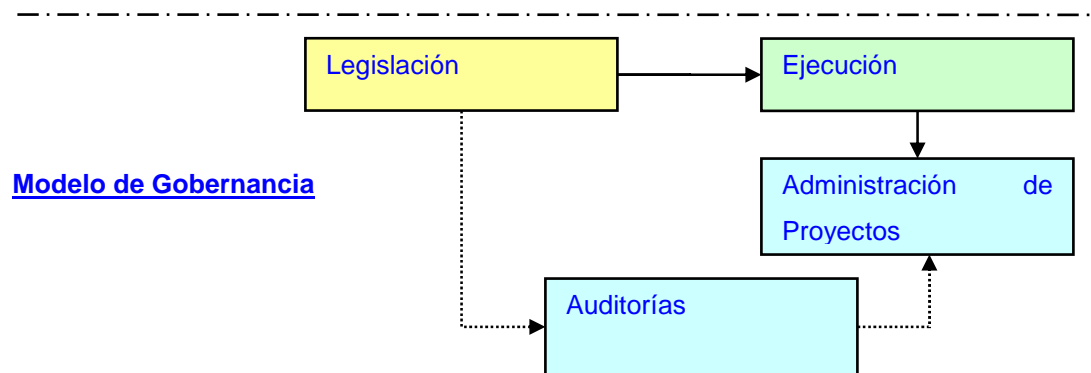


Figura 3.5. Modelo de Gobernanca.

## **La Metodología**

Existen varios modelos para realizar la planeación estratégica de sistemas de información de una organización, cada una con sus ventajas y desventajas, la selección dependerá de varios factores como la industria, el tamaño de la organización y la habilidad de la persona que lo realiza, sin embargo todos tienen en común el modelo general presentado en la Figura 3.6 a continuación:

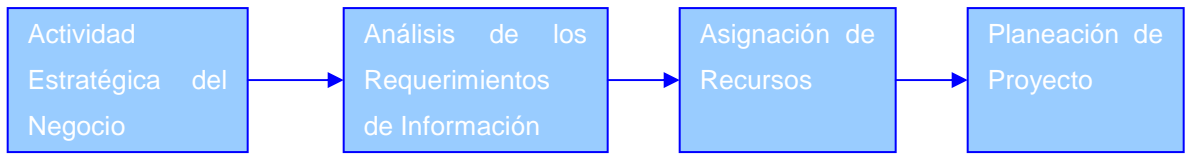


Figura 3.6. Modelo de planeación estratégica en sistemas.

Las metodologías para realizar este ejercicio se dividen en 2 categorías: de *Impacto* y de *Alineamiento* (mostradas en la Figura 3.7). Las de impacto ayudan a crear y justificar nuevos usos de TI como ventajas competitivas, mientras que las de alineamiento sirven para enfocar los objetivos de SI a los objetivos organizacionales, tal como se describe en el siguiente modelo:

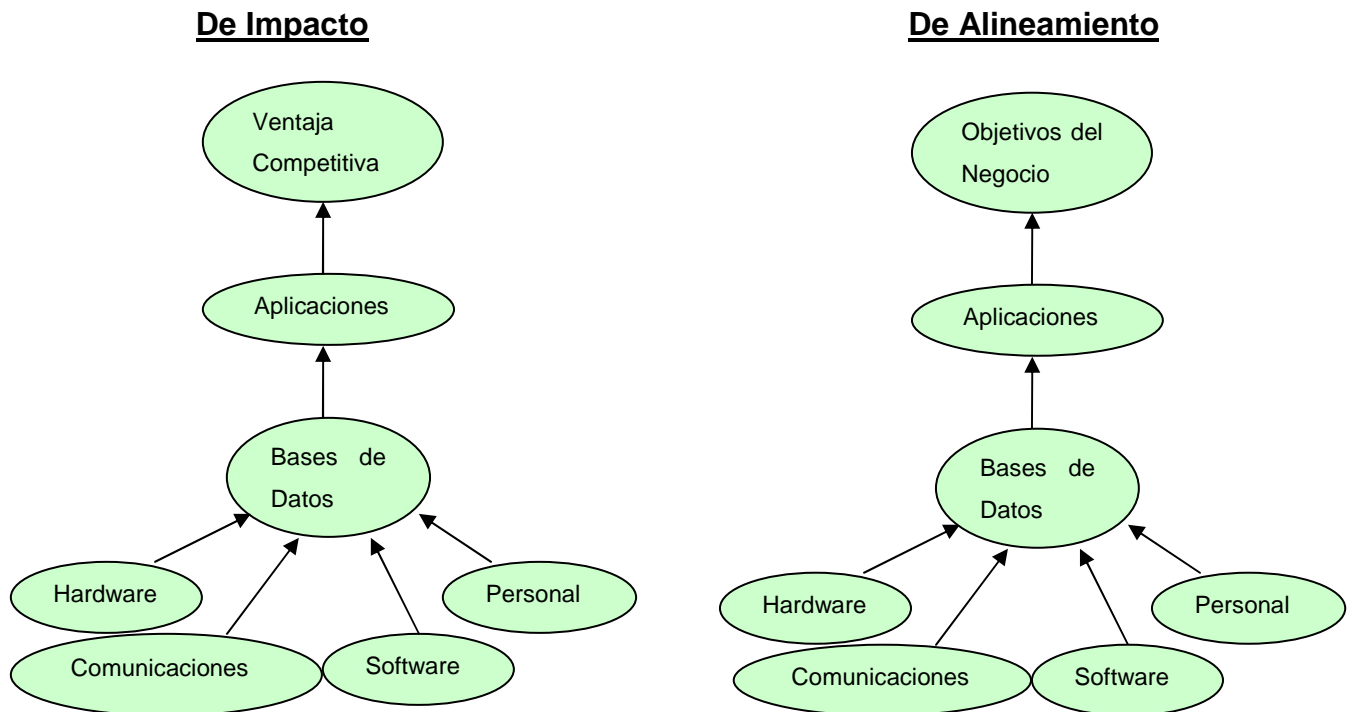


Figura 3.7. Metodologías de impacto y de alineamiento.

### Metodologías de Impacto:

Análisis de la Cadena de Valor (Figura 3.8):

- Ventajas: Se puede identificar perfectamente en que actividades se pueden concentrar los esfuerzos de TI para agregar valor, fácilmente comparable con la competencia.
- Desventajas: No es muy bueno para empresas de servicios cuyo producto es intangible [Porter, 1996].



Figura 3.8. Análisis de la cadena de Valor.

Análisis de los Factores Críticos de Éxito CFS (Rockart, 1979) (Figura 3.9):

- Ventajas: Permite concentrar fácilmente información clave sobre los requerimientos de la industria, organización, unidad de negocio y/o gerentes, es la técnica más ampliamente usada.
- Desventajas: no permite tener un panorama completo sobre la arquitectura o necesidades de soporte a la organización.

## Factores Críticos de Éxito

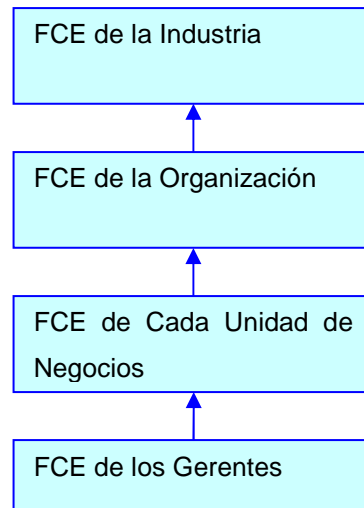


Figura 3.9. Factores Críticos de Éxito.

### Metodologías de Alineamiento:

Sistema de Planeación de Negocio BSP (IBM, 1982) (Figura 3.10):

- Ventajas: Hace un análisis top-down y una implementación bottom-up, es la más completa y mejor aceptada entre los directores de empresas
- Desventajas: Requiere un compromiso firme de la dirección general, se requiere cierta experiencia para aplicarlo y se crean huecos en el análisis e implementación en organizaciones grandes

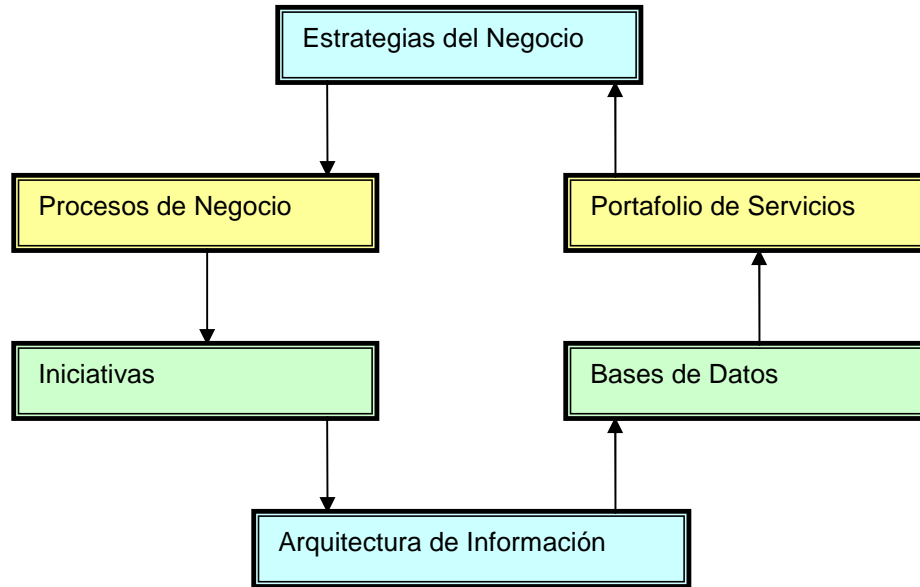


Figura 3.10. Factores Críticos de Éxito.

Sistema de Planeación Estratégica SSP (R. Holland, 1985):

- Ventajas: Dado que está basado en el BSP anterior, son las mismas, salvo que se utiliza un método automatizado para recolectar los datos y permite identificar nuevos sistemas y calendarios para ser implementados
- Desventajas: las mismas que en BSP

Ingeniería de Información IE (Martin, 1982)

Método/I (A. Anderson, 1982):

- Ventajas: Evalúa la organización, entorno, tecnología en 10 segmentos de trabajo específicos, propone acciones para hacerlo y define los entregables
- Desventaja: Es un proceso muy largo y complejo además de que se basa en el modelo de cadena de valor de Porter

Means/Ends Análisis EM (1950)

## El Plan

En esta sección se analizará detalladamente la elaboración del plan estratégico así como sus fases y como identificar la posición de la organización en cada una de las etapas del desarrollo del plan.

Asimilar de manera clara la identidad de la organización:

- Misión, Visión y Valores
- Objetivos y Necesidades
- Hacer un análisis de la situación actual y las metas

Se requiere hacer un análisis de la situación actual y la deseada en puntos como: la organización de TI, la Gobernancia, las aplicaciones, la infraestructura, los procesos y el control de costos de TI.

Identificar el tipo de iniciativas que ayuden al plan de la organización y evaluación del entorno:

- Establecer las prioridades de las necesidades y ver que estén en armonía con la organización, el entorno actual, y que agreguen valor al negocio, es decir, crear ventajas competitivas. Es así como se proponen las iniciativas.
- Fórmula para crear e identificar una ventaja competitiva:

En el diagrama de la Figura 3.11, se puede ver como se integran las ventajas competitivas en la organización.



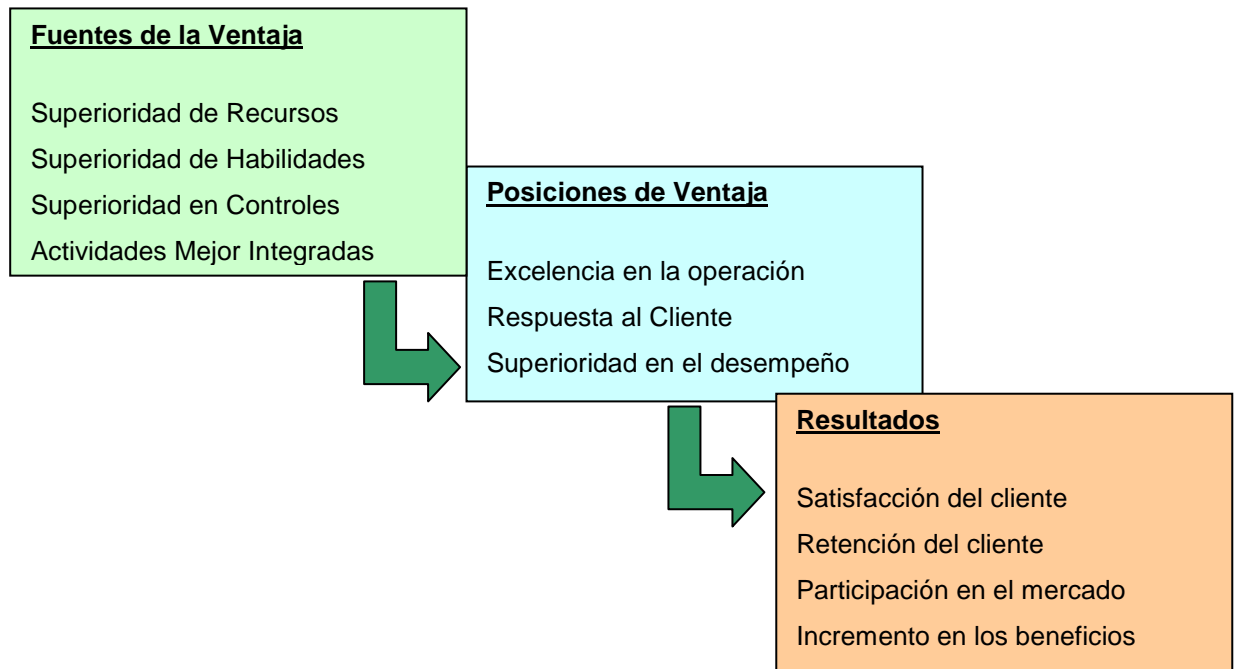


Figura 3.11. Integración de la ventaja Competitiva.

En la tabla de la Figura 3.12, se pueden ver las características de las fuentes de ventaja que nos permita identificar las fortalezas de nuestra organización y de esta manera poder ir enfocando la propuesta del portafolio de soluciones.

<u>Características de las Fuentes de Ventaja</u>			
Superioridad de Recursos	Superioridad de Habilidades	Superioridad en Controles	Actividades Mejor Integradas
Mejor cobertura en la distribución	Conocimiento especializado	Control de costos	Mayor eficiencia y fluidez en cada uno de los eslabones de la cadena de valor
Mayor número de personas en ventas o servicio a clientes	Relaciones cercanas a los clientes	Calidad satisfacción y lealtad de los clientes	
Capacidad en la infraestructura	Mejor diseño	Sistemas para la administración de inventarios	
Mayores inversiones en tecnología	Mejor tiempo de respuesta	Seguimiento al flujo de efectivo	
Mayor gasto en publicidad o promoción	Mayor Flexibilidad	Mayor transparencia en los procesos con el cliente	
Capacidad y estructura financiera, costo de capital	Habilidad para aprovechar los SI		
Mejor posicionamiento de marca	Posibilidad de utilizar tecnologías de BI integradas a los SI		
Más patentes	Atención más cercana a la operación		
Menor costo de insumos	Mejor experiencia en identificar, negociar, y administrar alianzas		
Personal mejor capacitado			

Figura 3.12. Fuentes de la ventaja Competitiva.

En la tabla de la Figura 3.13, se pueden ver las características de las posiciones de ventaja de una empresa, que es en donde impactan las fuentes de ventaja anteriores:

<u>Características de las Posiciones de Ventaja</u>		
Excelencia en la Operación	Respuesta al Cliente	Superioridad en el Desempeño
<ul style="list-style-type: none"> <li>• Alta confiabilidad y consistencia en la calidad</li> <li>• El precio más bajo</li> <li>• Los procesos clave están estandarizados, altamente automatizados e integrados</li> <li>• Sistemas y procesos administrativos top-down</li> <li>• Uso de tecnologías de BI</li> </ul>	<ul style="list-style-type: none"> <li>• Productos y servicios a la medida</li> <li>• Decisiones con base en información detallada de clientes y canales</li> <li>• Los procesos clave son flexibles y facilitan múltiples modos de entregar productos</li> <li>• El personal con contacto con los clientes toma decisiones</li> </ul>	<ul style="list-style-type: none"> <li>• Innovación constante en las características del producto para mejorar su uso</li> <li>• Primeros en entrar al mercado, reducción del ciclo de vida</li> <li>• Estructura organizacional descentralizada</li> <li>• Receptivos a nuevas ideas, interés por experimentar</li> <li>• Los procesos clave enfatizan el conocimiento del mercado</li> </ul>

Figura 3.13. Impacto de la ventaja Competitiva.

## Análisis FODA

Para terminar de delimitar la situación actual de la empresa se tiene que realizar un análisis de Fortalezas, Oportunidades, Debilidades y Amenazas (esquema mostrado en la Figura 3.14). Las Características y las Posiciones de Ventaja nos identifican claramente las fortalezas. El tipo de impacto nos indican las oportunidades. Solamente resta analizar las debilidades de la organización y las Amenazas de la industria en el momento que se realiza el ejercicio:

<u>Fortalezas</u>	<u>Debilidades</u>
<ul style="list-style-type: none"><li>• Características de ventaja</li><li>• Posiciones de ventaja</li></ul>	<ul style="list-style-type: none"><li>• Debilidad de la organización 1</li><li>• Debilidad de la organización 2</li></ul>
<u>Oportunidades</u>	<u>Amenazas</u>
<ul style="list-style-type: none"><li>• Tipos de Impacto</li></ul>	<ul style="list-style-type: none"><li>• Amenaza de la industria 1</li><li>• Amenaza de la industria 2</li></ul>

Figura 3.14. Impacto de la ventaja Competitiva.

La experiencia ha demostrado que el seleccionar las iniciativas basándose en el factor económico solamente, o político, no son la forma más eficiente de hacerlo, ni la que más beneficios trae a la empresa; dado que los conflictos de intereses dentro de las organizaciones son muy comunes, es de vital importancia de SI, en representación de la dirección de la empresa y su equipo cercano, concilien dichas discrepancias.

## 2. Madurez de sus procesos informáticos

Los procesos de una empresa tienen que ser maduros para permitir que la informática los afecte, automatizándolos. Para permitir analizar la

estructura de una empresa, en la Figura 3.16, se muestran estratificadas las actividades de una empresa, desde la base operativa, hasta la punta estratégica, en donde se toman las decisiones que afectan de manera importante a la empresa.

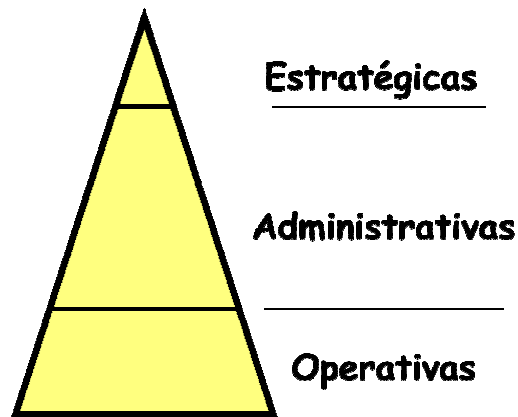


Figura 3.16. Esquema de una empresa.

Los planes de la empresa deben considerar de forma importante los procesos de gobernanza informática otorgándole el lugar de importancia que permita la toma de decisiones orientada adecuadamente. La empresa puede haber iniciado su proceso de automatización de manera progresiva iniciando por automatizar los procesos operativos, para facilitar los procesos diarios, entre los cuales están los productivos, y básicos de registro de la operación diaria, procesos como registro de asistencia, facturación, entregas, cobros diarios, y demás.

Si ha logrado automatizar las áreas administrativas, esto permitirá contar con información acumulada que agilice el registro y recuperación de información en procesos básicos de la empresa, como la contabilidad, nominas, y demás.

En el caso de los procesos estratégicos de la empresa, estos son los que se dictan desde dirección de la empresa, y que conjugan y concentran la información de las áreas administrativas para dar elementos de toma de decisiones a los directivos que así lo requieren con la finalidad de que estas sean tomadas con la mayor información posible. Esta información permite además ser tratada mediante métodos matemáticos y de inteligencia en los negocios, con la finalidad de ofrecer una mejor guía de la información actual y de escenarios posibles a enfrentar.

Definir la gobernanza de las tecnologías y sistemas de información, permite alinearlas con los objetivos de la empresa, y establecer en un plan de acción la arquitectura de TI necesaria para el logro de los objetivos.

Como consecuencia se deben definir las posiciones operativas, administrativas y estratégicas de TI necesarias para cubrir las necesidades como: DBA's, NWA, Help Desk, BI y demás. La siguiente tabla muestra un esquema de definición.

	Centralizado	Distribuido	Cliente/Servidor	Virtualización
Hardware	Compartir recursos caros  Compartir infraestructura (Energía, Seguridad Física, Aire Acond., etc.)  Control de compras	Menos sensible a caídas  Más adaptado a las necesidades locales  Distribución de costo más justa por unidad de negocios o departamento	Se comparte el procesamiento  Independencia entre clientes y servidores	Ahorro de energía  Facilidad de Administración  Reducción de espacio  Facilidad de mantenimiento  Reducción de

	Optimización de Uso Menos duplicidad			caídas Incremento de capacidad
Telecomunicaciones	Control del tráfico Control de salidas y entradas con seguridad Mejor Administración	Costos más bajos de los enlaces Distribución del costo Comunicación más eficiente, evita duplicidad de transmisiones a sitios remotos	Acceso Remoto, permite la comunicación desde cualquier cliente, casi desde cualquier red y tener los servicios y seguridad de la red corporativa	Operadores virtuales, concentran la com. en sus <i>POP's</i> más cercanos a cada sitio y ahí hacen la estrella virtualmente  Ideal en organizaciones internacionales
Aplicaciones	Compatibilidad garantizada Descuento por volumen Fácil de mantener y actualizar	Más personalizado Diferentes opciones de configuración Facilidad de acceso	Independencia de los clientes respecto del servidor  Reducción de costos para mantenimiento y actualizaciones	Independencia de otras aplicaciones  Multiplataformas
Datos	Facilidad para respaldar Facilidad para compartir Mayor consistencia Facilidad para	Datos más individuales, el nivel de colaboración es mínimo  Acceso más rápido  Controles y	No se almacena información en los clientes	No Aplica

	monitorear y administrar la seguridad.	políticas locales		
Personal	Perfiles y desempeño del staff muy similares Especialistas Fácil de controlar y administrar	Más cerca de los usuarios Mejores tiempos de respuesta Mayor integración con otras áreas	Los especialistas están en el centro de datos El usuario tiene un punto de contacto único (SPOC)	Outsourcing, permite dejar la operación a personal externo y enfocarse en el negocio o 'Core business'

Figura 3.17. Esquema de necesidades a cubrir.

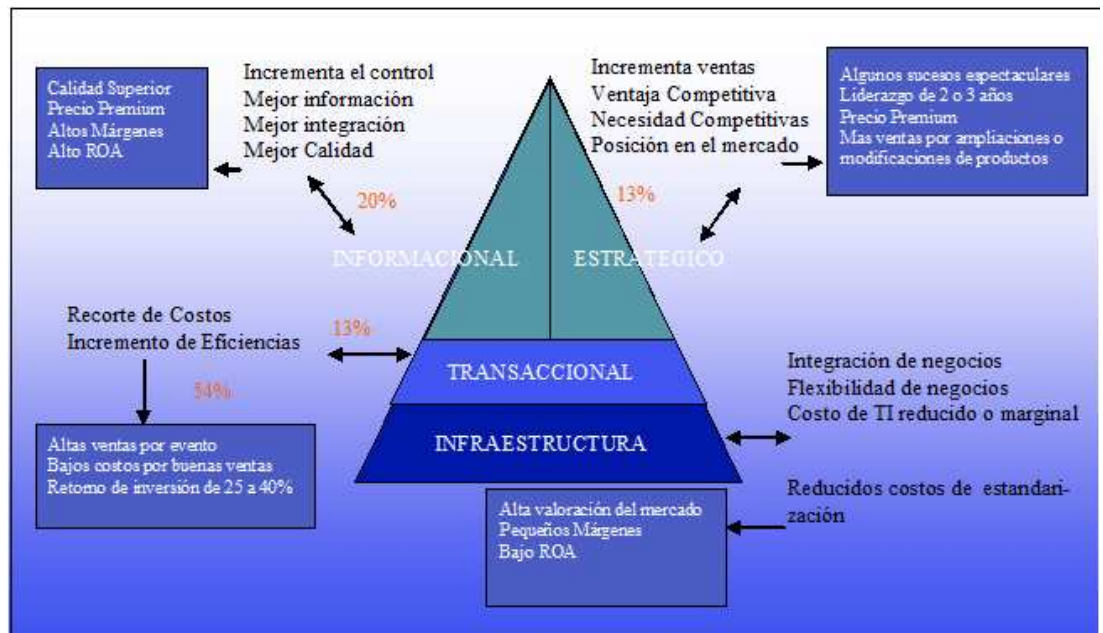
### Diseñar el Portafolio de Servicios

El motivo por el que las organizaciones invierten en TI es alcanzar 4 objetivos básicos, los cuales motivan el desarrollo de TI en el mercado, estos objetivos son:

- De Infraestructura: Necesaria para proveer servicios confiables son la base de la capacidad tecnológica (LAN, WAN, Intranet, Bases de Datos, etc)
- Operativos o Transaccionales: Es la plataforma operativa que procesa y automatiza las tareas básicas repetitivas o peligrosas (Control de Inventarios, Procesamiento de órdenes, Cuentas por cobrar, Cuentas por Pagar, etc.).
- De Información: Necesaria para gestionar y controlar la organización, apoyo a la toma de decisiones, planeación, comunicación y contabilidad (Data Warehouse, Balance Score Card, Administración del Conocimiento, etc.)



- Estratégicos: Son para obtener y mantener la ventaja competitiva que eventualmente evolucionará en transaccional y finalmente infraestructura (Sistemas expertos, Sistemas de BI).



- Fuente: MIT estudio CIRS por Weill & Aral utilizando datos de 1999-2002 de 147 firmas, y el libro "leveraging the new infrastructure: How Market leaders capitalize on IT", Harvard Business School Press, Junio 1998.

Figura 3.18. Portafolio de Servicios.

Tomando como base el diagrama de la Figura 3.18, se puede comparar la distribución del gasto e inversiones en TI de las distintas industrias en México, mostrado en la Figura 3.19:

Tipo de Empresa	Gasto de Infraestructura	Gastos Operativos	Gasto de Información	Gasto en Estrategia
Promedio	8%	2%	6%	4%
Enfocada en Costos	2%	0%	3%	7%
Enfocada en Agilidad	8%	1%	4%	7%
Balance Agilidad-Costos	0%	5%	0%	5%

Figura 3.19. Inversión en TI en México.

Se debe considerar que la forma en que se define el plan estratégico de la empresa en el área de la TI y SI, es importante para la sobrevivencia de la empresa, siendo importante definir claramente la gobernanza de la informática sobre los procesos de la empresa, y el crecimiento gradual de su integración a los niveles de decisión de la misma.

Como un organismo vivo la empresa debe crecer mediante los planes prospectados por los niveles estratégicos, basados en herramientas que gradualmente faciliten la prospección de escenarios mediante las herramientas de BI adecuadas.

### 3. Proceso de la toma de decisiones

¿Qué es tomar decisiones? “Decidir es transformar la voluntad en acción”. Principalmente la toma de decisiones permite obrar al actor que la toma, permite fragmentar los actos de una empresa en tantas competencias que permite los contrapesos y facilitando la libertad dentro del orden.

El proceso de toma de decisiones es el corazón de la administración y normalmente no existe un proceso que se apegue a un modelo de toma de decisiones racional, es decir que este proceso se considera que se encuentra entre los límites de la aplicación de la racionalidad y la irracionalidad, comportamiento común de la sociedad humana.

Las decisiones en una organización deben estar basadas en premisas reales que incluyan elementos normativos y empíricos, las cuales son llamadas premisas de valor y de hecho. En el proceso de decisión estas dos premisas son elegidas considerando cuál de ellas es el mejor medio para alcanzar los objetivos. Estos objetivos son meramente instrumentos para alcanzar los fines de la empresa.

La teoría administrativa de la toma de decisiones de Herbert Simon describe el proceso en tres pasos:

- a) Enlistar todas las estrategias posibles alternativas.
- b) Determinar todas las consecuencias que tendría el tomar cada una de las alternativas posibles.
- c) Evaluar comparativamente estos grupos de consecuencias.

El hombre económico es la representación del hombre racional, aquel que toma decisiones bajo las premisas de hechos. Un ejemplo clásico de esto es cuando se analiza la función de transitividad lógica  $A > B > C$ , en los hechos la

consecuencia transitiva es que  $A > C$ , y nunca pasara bajo estas premisas que  $C > B$ . El hombre económico basado en premisas de hechos siempre elegirá que  $A > C$ , y será la respuesta lógica tomada.

Además se establece la definición del hombre administrativo, que es el hombre limitado por sus habilidades, hábitos y reflexiones inconscientes, con concepciones, valores y propósitos que pueden ser diferentes a los de la organización para la cual está empleado. Este hombre mezcla estas características propias, con los objetivos de la empresa y con su conocimiento de la información de la que dispone, para tomar una decisión. Su comportamiento es en consecuencia limitado a su racionalidad objetiva en al menos tres formas:

- a) La racionalidad requiere de un completo conocimiento de anticipación de las consecuencias de tomar una determinada decisión. De hecho el conocimiento de las consecuencias nunca es completo, siempre hay un factor de incertidumbre.
- b) Debido a la incertidumbre en el futuro, la imaginación de eventos por acontecer debe ser suplida por la experiencia y valores relacionados con ella. Pero estos valores pueden ser solo anticipados de manera imperfecta.
- c) La racionalidad requiere de elegir entre todos los posibles comportamientos y consecuencias alternativas. El comportamiento actual es solo uno de los comportamientos que llego a la mente del decisor.

El hombre administrativo y el hombre económico son diferentes en dos aspectos principales:

- a) El hombre económico maximiza. Selecciona la mejor alternativa. El hombre administrativo satisface necesidades de solución para un curso de acción que es satisfactorio, o al menos “suficientemente buena”.

- b) El hombre económico trata con el mundo real en toda su complejidad. El hombre administrativo reconoce que el mundo que el percibe es un modelo drásticamente simplificado. El toma sus decisiones utilizando una simple fotografía de la situación que toma en cuenta una cantidad limitada de factores y deja de lado una cantidad de elementos que pueden ser cruciales y de mucha relevancia.

El hombre administrativo es el que siempre simplifica un modelo abstrayéndolo de la realidad, viendo una cantidad limitada de alternativas y de información, acerca de las consecuencias de las diferentes alternativas. Solo ocasionalmente una decisión es dirigida a maximizar los beneficios de la misma. Como el hombre administrativo está satisfecho con un limitado conocimiento de la situación a considerar, sus decisiones pueden ser la aplicación de una sencilla heurística, la cual no requiere de un análisis imposible o poco realista.

El proceso de toma de decisiones es lineal, siguiendo las etapas de:

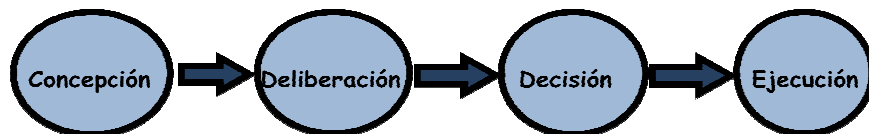


Figura 3.20. Proceso de toma de decisiones.

Inicialmente desde la identificación de la necesidad de tomar la decisión, concebirla como inicio de proceso, es un momento importante en que el decisor deber captar la necesidad de tomar, o no, acciones que le permitan enfrentar el futuro con mayores posibilidades de éxito.

Es un elemento crucial del proceso, durante el cual se deberá hacer una descripción minuciosa de las condiciones que privan al interior y al exterior de la organización y que están provocando la necesidad de actuar para

mantener una posición congruente con el tiempo y circunstancias que se viven y, así mismo de los requerimientos que se deberán satisfacer en la nueva posición. Después se realiza la deliberación en donde en dinámicas grupales o de manera personal se analizan diferentes opciones con la información que los sistemas propios implementados permitan proyectar, y analizando las consecuencias que cada de ellas.

Una vez planteadas y analizadas las opciones, se debe tomar una decisión sobre las bases que esta opción plantea. Es elegir entre todas las alternativas establecidas, aquella que brinde al decisor, un mayor grado de certidumbre para alcanzar la nueva posición deseada.

Finalmente el paso trascendente entre el pensar y el hacer, que actualmente demandan de una gran flexibilidad, comunicación y compromiso dentro de las organizaciones. Llevarla a su implementación en la siguiente etapa.

En base a los resultados que se obtienen de tomar una decisión se puede calificar al tomador de la misma como:

- a) El hombre acertado.
- b) El hombre probable.
- c) El hombre aleatorio.

La toma de decisiones informada y que utiliza métodos de apoyo en la fase de deliberación es la que permite en general acertar en mayor medida a la correcta opción, y existen muchas herramientas de apoyo para este propósito, entre las que se cuentan:

- a) Árboles de decisiones.
- b) Metodologías Maxmin – Minmax.
- c) Análisis de pérdidas de oportunidad.

- d) Reglas de decisión de bayes.
- e) Valor esperado de la información perfecta.
- f) Modelos de colas.
- g) Modelos de teoría de juegos.
- h) Optimización lineal.
- i) Teoría de redes.
- j) Sistemas informáticos de apoyo a la toma de decisiones.

Existen varias formas para la toma de decisiones:

- a) Bestialismo



Figura 3.21. Bestialismo en las decisiones.

- b) Veleidad

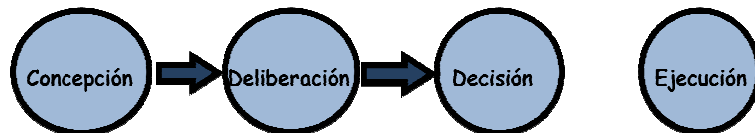


Figura 3.22. Veleidad en las decisiones.

- c) Intelectualismo

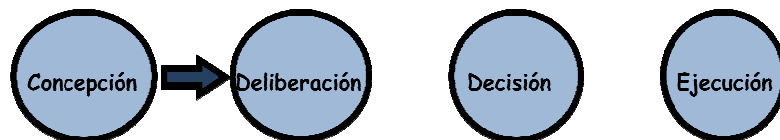


Figura 3.23. Intelectualismo en las decisiones.

Y están en función a las reglas y formación del decisor, analizadas en el esquema de la Figura 3.24:

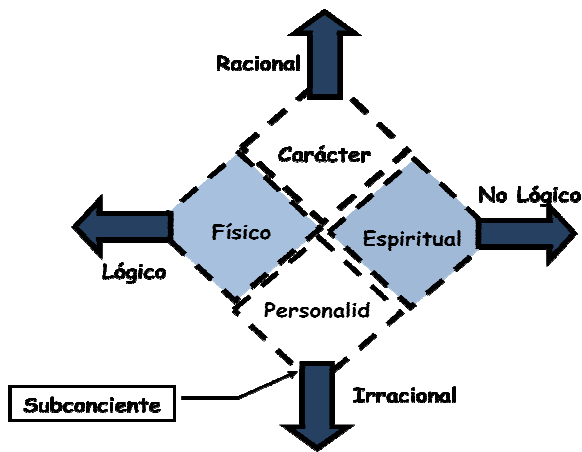


Figura 3.24. Reglas y formación del Decisor.

Según el modelo Carnegie, la toma de decisiones debería seguir un esquema (Figura 3.25) como el que propone: Richard Daft: "Organizations: Theory and Design"

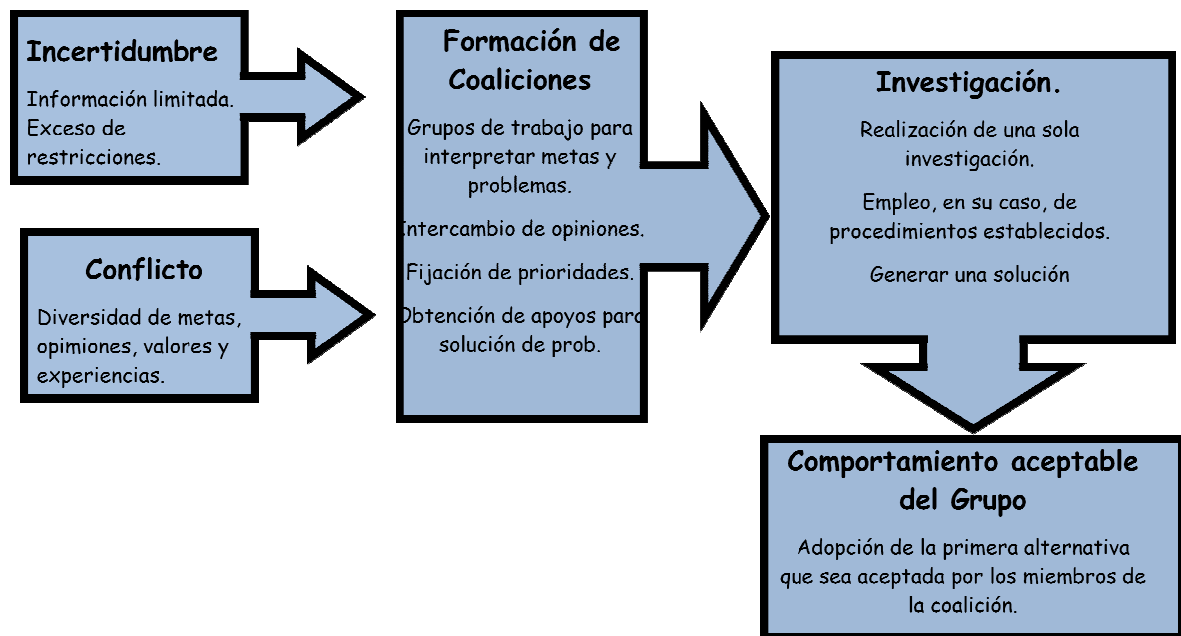


Figura 3.25. Modelo de Carnegie.

Este modelo propone que la toma de decisiones se toma con información limitada, y con un conjunto de restricciones de la misma, que en su



conjunto forma el modelo sobre el cual decide el Hombre Administrativo propuesto por Herbert Simon, pero que al tomar decisiones requiere de hechos, es decir información certera para tomar la mejor opción de las que puede analizar dentro del modelo conocido. Además de que se propone una investigación para poder obtener más información para la correcta elección del decisor.

#### 4. Aplicación de Herramientas de inteligencia en los negocios.

La aplicación de herramientas de inteligencia de negocios a las empresas, esta normado por la infraestructura informática que la empresa haya implementado en sus procesos. La implicación es que si la empresa tiene informatizados sus procesos operativos de organización, cuenta con sistemas tecnológicos y con sistemas que desarrollen su administración, podrá contar con un sistema de información integrado, que a su vez sea el soporte para un sistema de soluciones de negocios, que permita el apoyo a la toma de decisiones informada.

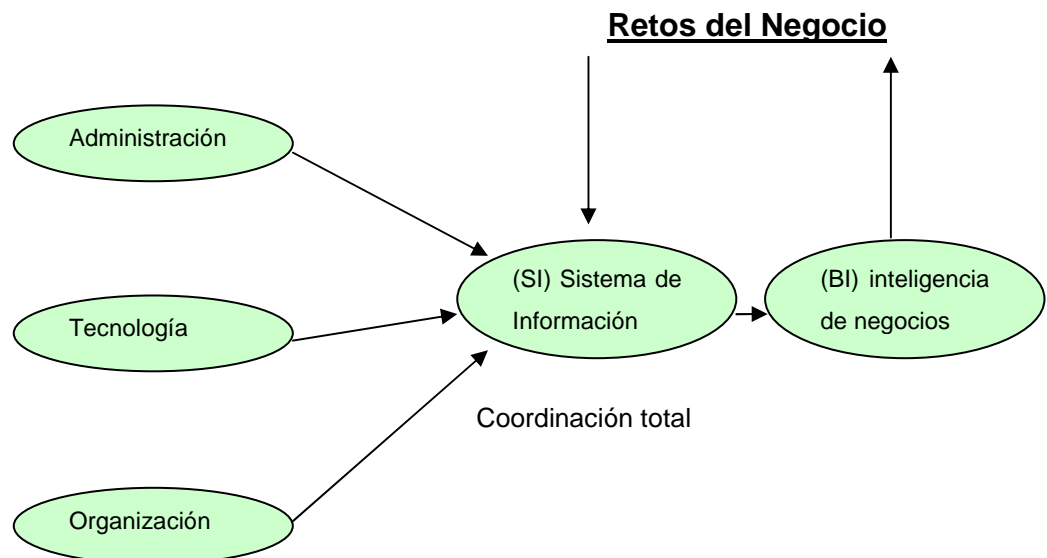


Figura 3.26. Integración informática.

En la Figura 3.26, se muestra el esquema de la integración de la informática en la empresa, y esta dado por entendido que si no se cuenta con la adecuada y suficiente informatización en alguna de las etapas previas a la solución de negocios, es muy difícil poder contar con los beneficios de la aplicación de inteligencia de negocios. Por esta razón es que solamente empresas con alto grado de madurez en sus procesos, y con un alto grado de integración informática actualmente hacen uso de estas herramientas como una ventaja competitiva para agilizar la toma de decisiones, darle mayor asertividad, y tener una respuesta eficiente a los cambios que plantea un mercado cambiante y una competencia competitiva. Es mas en muchas industrias el contar con sistemas de inteligencia en los negocios se hace tan indispensable que de ello depende su sobrevivencia.

Los negocios tienen diferentes tipos de necesidades en cuanto a sistemas de información, para apoyar la toma de decisiones, trabajando en varios niveles de la estructura y con las funciones de cada nivel. Por ejemplo se pueden necesitar sistemas para coordinar la cadena de suministro, para diseño, producción y logística, y esto producirá información disponible sobre estos procesos, que se puede ocupar en capas estratégicas del negocio. Se puede decir que los retos de la administración de una empresa referentes a la informática son:

- a) Integración.- diseñar varios sistemas que sirvan a diferentes niveles y funciones de la empresa, encontrando ventajas en la integración. Sin embargo esta integración es un proceso que cuesta mucho dinero y mucho tiempo. El administrador del área informática deberá determinar con visión el nivel de integración que gradualmente se irá alcanzando en la empresa.
- b) Ampliar el campo del pensamiento administrativo.- la mayoría de los administradores están capacitados para manejar las líneas de producción, divisiones de producción y oficinas. Pocas veces están

entrenados para optimizar procesos en las organizaciones como un todo. Esta capacidad de visualizar el campo de la aplicación de la informática es necesaria para comprender mejor el ambiente de la empresa y su propio comportamiento, además de proyectar su futuro.

La organización está dividida en áreas estratégicas, de administración, de conocimiento y operacionales, y además está dividida en áreas funcionales. Para permitir que el nivel estratégico tome las decisiones basados en información de hechos, con la agilidad necesaria para que represente una ventaja competitiva, los niveles inferiores requieren de tener sus procesos automatizados a un nivel adecuado. Así se muestra en la Figura 3.27. [Laudon, 2007]

En empresas con recursos limitados contar con una integración informática a sus procesos es tan costoso que difícilmente se puede pasar de los niveles operacionales, por lo que lograr que los niveles de conocimiento, de administración o más aun los estratégicos es una tarea difícil. De esta manera las herramientas de inteligencia de negocios fluirán más fácilmente en un ambiente bien integrado, aunque difícil de alcanzar.

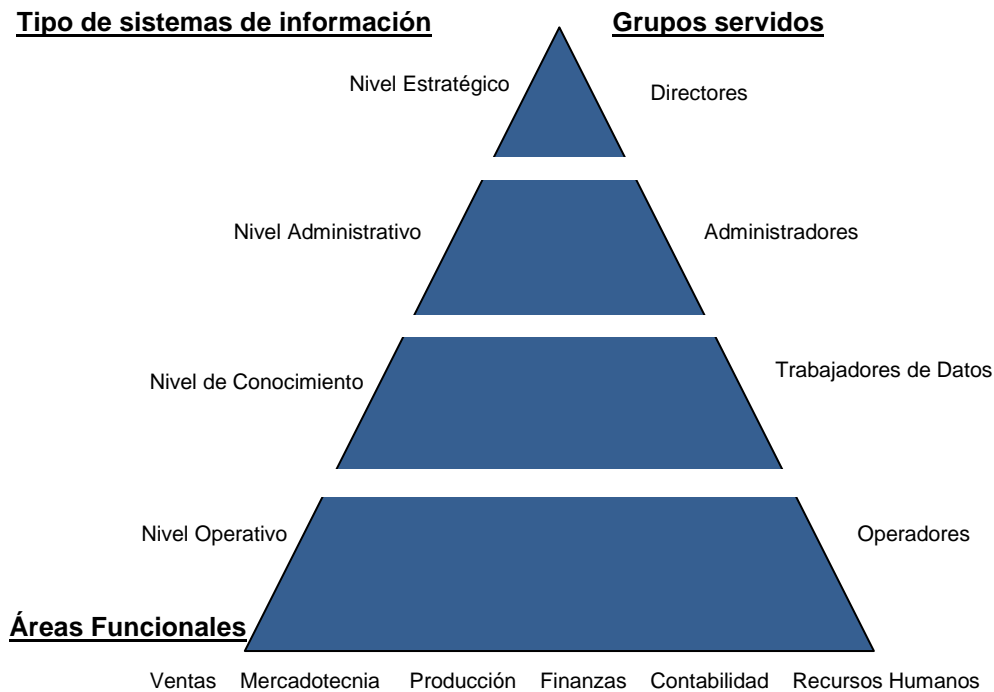


Figura 3.27. Niveles de una empresa.

## 5. Utilidad de la prospectiva empresarial.

La prospectiva es un tipo de aplicación de diferentes herramientas de inteligencia que aplicada a los negocios a través de metodologías matemáticas logra obtener el comportamiento de un área de análisis. Se ocupan metodologías de minería de datos como análisis por clústeres, para encontrar patrones ocultos que agrupen lecturas de datos, y que definan similitudes entre ellos.

Si este comportamiento se modela adecuadamente es posible lograr comprender y predecir a partir de un estado N el comportamiento del estado N+1 de esa área de análisis en particular.

La prospectiva es una herramienta que es fuertemente ambicionada por diferentes campos de aplicación por sus implicaciones en cuanto al

conocimiento de eventos futuros, y el área de negocios no es la excepción. Existen muchos estudios citados en el capítulo II.2. referentes a la utilización de minería con el fin de prospectar un comportamiento.

Los sistemas de Soporte a la toma de decisiones (DSS) sirven tanto al nivel de administración, como a los niveles estratégicos de la organización. La mayoría de los DSS hacen uso de los sistemas de información internos y toman información de fuentes externas como la información de la competencia para diseñar la información que brinda a la toma de decisiones.

Un DSS tiene un poder más analítico que cualquier otro sistema. Son construidos explícitamente con una variedad de modelos de análisis de datos y condensan grandes volúmenes de datos para presentarlos de forma que puedan ser analizados por los tomadores de decisión de la organización. Normalmente son sistemas amigables que permiten la fácil interacción del decisor. [Laudon, 2007].

#### **IV. ANTECEDENTES DEL OBJETO DE ANÁLISIS**

El objeto de análisis de este estudio son los datos de una empresa mediana real del sector productivo. Es una empresa establecida en el estado de Querétaro, con una antigüedad de más de 28 años de operación, que recientemente desde hace 16 años, inicio la integración de herramientas informáticas a sus procesos administrativos, y hace aproximadamente 11 años a sus procesos productivos.

Actualmente la empresa tiene sus oficinas corporativas en el estado de Querétaro, con una planta de producción única para el país, y con delegaciones (o sucursales de distribución) en 14 estados de la república.

##### **1. Acuerdo de Privacidad**

Se estableció un acuerdo de privacidad del total de la información de la empresa objeto del análisis, manteniendo en anonimato su identidad y permitiéndosele ocupar los datos de su sistema administrativo, productivo y contable en su totalidad con la condicionante de poder hacer pública solo información parcial. La información contenida en esta investigación se ocupa de manera integral, aun cuando su publicación pueda ser solo de manera parcial.

Se establece como beneficio para esta investigación el poder contar en consecuencia con información real de trabajo de 8 años de la empresa, pudiendo seleccionar la información que considere conveniente para el estudio, siempre y cuando tenga al tanto de las decisiones a la dirección de la misma. Los datos son extraídos directamente de su sistema y analizados para no comprometer la identidad o posible mal uso de esta información por otras empresas competencia.

Es importante contemplar que como parte del acuerdo se establece el beneficio para la empresa de poder contar con información resultado del presente análisis. Los resultados se entregaron gradualmente en función a lo analizado hasta el momento en que se tuvieron avances, y se realizaron sugerencias derivadas de los mismos.

Además se consideraron casos de análisis histórico que no comprometieran la identidad de la empresa, pero que dieran información del desarrollo histórico de la misma, y sirvan de marco referencial para rastrear el origen de los resultados de la propuesta prospectiva que siguiere el modelo.

## **2. Contexto empresarial**

Para el presente proyecto se extrajo información de 8 años de datos de la empresa objeto de estudio, y se seleccionan los elementos de la misma que tengan representatividad en su desempeño económico (acuerdo inicial tomado con la propia empresa de las variables a considerar para el estudio).

Inicialmente se realizaron entrevistas con la dirección de la misma con la finalidad de platicar de los alcances del proyecto y de los posibles resultados que se pueden obtener en función al proceso que se ocupe para modelar el sistema. La dirección de la empresa mostro interés en poder conocer los resultados del modelo, y la utilización que pudieran dar de esta herramienta prospectiva.

Se le hizo de conocimiento la brecha informática que se tiene en los procesos actuales con respecto a los que se requerirían para poder explotar los datos de la empresa en un proceso prospectivo tan fluido que diera una ventaja competitiva a la empresa. Además se planteo la necesidad de analizar las posibilidades de llevar a cabo en el futuro, una integración gradual de sus procesos actuales, hacia una gobernanca informática que proyecte a la empresa a un liderazgo en el ramo en el que se desempeña.

Se tuvieron acuerdos en cuanto a la privacidad de la identidad, y de la información motivo del modelado, y se inicio una definición de los datos motivo posible de análisis. Es importante ponderar la importancia de la selección de los elementos que han de ser motivo de la minería de datos, buscando los patrones de comportamiento que mas reflejen su situación histórica, porque diferentes datos pueden tener diferentes comportamientos, y es posible que requieran de la aplicación de diferentes parámetros al modelo.

### **3. Entorno informático**

La empresa inicio su esfuerzo de integración informática hace 16 años, con su aplicación a los procesos operativos básicos. En este inicio ocuparon paquetería muy sencilla definida para ser operada de manera mono usuario en cada computadora. 3 años después integraron un sistema en red que permitía el intercambio de información entre las computadoras, operando con el sistema administrativo.

Dos años después para 1997 se instala un sistema administrativo para red, que ya permitía la integración de las operaciones básicas de compras, ventas, inventario, cuentas por cobrar y por pagar y contabilidad de manera más ágil, almacenando los datos en una base de datos centralizada en un servidor. En consecuencia también se integro en la estructura de red un servidor para atender las necesidades del sistema administrativo.

Para el año 2002 las necesidades de la empresa llevan a analizar e instalar a para inicios del 2003 un sistema informático más integrado y con mayores funcionalidades hacia las áreas de Recursos Humanos y Producción, y se implanta un sistema ERP. Este es el sistema de información actualmente ocupado en la empresa para solucionar sus necesidades en las fases operativas y administrativas de la empresa. Durante el año 2002 y buena parte del 2003 se mantuvo el sistema administrativo anterior operando, hasta que el



sistema ERP estuviera adaptado a las necesidades básicas de la operación de la empresa.

Actualmente el sistema está operando dando servicio a mas de 30 equipos de computo locales, y aproximadamente a 35 equipos remotos en el interior de la república mexicana.

Cada uno de los sistemas anteriores ocupa un esquema de base de datos propio (ejecutándose sobre el sistema operativo Windows) que tiene pocas posibilidades de hacer interfaz con los sistemas de bases de datos comerciales. Es por esto su difícil integración a otras herramientas de análisis de datos comerciales al momento presente.

El actual sistema ERP cuenta con una interface que permite exportar la información a bases de datos SQL Server de Microsoft, con lo cual es posible manipular la información hacia herramientas externas a su entorno, pero esta explotación es obteniendo una imagen de los datos, no operándolos o explotándolos en línea.

#### **4. Obtención de datos**

Los datos han sido extraídos principalmente del sistema actual (de los años 2003 al 2009) y del sistema anterior solo 2 años incompletos (2002 y parte de 2003). No se explotaron años anteriores por dos razones principales: la primera es por la dificultad de obtener información del sistema anterior al no contar con muchas posibilidades de hacer una manipulación informática de los datos almacenados en la base de datos.

En segundo lugar porque la empresa comento que los registros de esos años no eran muy fidedignos de ser tomados en cuenta por motivos de cambios en la administración de la empresa que no permitieron estandarizar el registro de la información, y en consecuencia era difícil de considerarla seria para su análisis.

Finalmente se extrajeron 92 datos del sistema de 8 años de operación, de los cuales se presenta una muestra en el apéndice A. Se presenta solo una muestra por cuestiones de privacidad. La extracción se realizó importando información del sistema administrativo anterior, por medio de reportes, debido a los problemas para conectarse. Del sistema ERP se generó una imagen de los datos necesarios en una base de datos SQL Server 2005 express, y desde ahí se manejó la información para obtener los datos que fueron considerados inicialmente como interesantes para el análisis.

Finalmente ya integrados todos los datos se realizaron exportaciones al sistema Excel de Microsoft (Hoja de cálculo), para revisar los datos y manipular ejemplos en tablas y en gráficas buscando comprender mejor los datos. Una de las principales necesidades es hacer interfaces hacia diferentes formatos para varias aplicaciones, con la finalidad de exportar hacia ellas los datos y probar sus métodos de generación de clústeres y evaluar los resultados directos e indirectos que se obtienen.

La formación de clústeres son considerados como los resultados directos, y la información que analiza el resultado de la formación, son considerados los resultados indirectos.

## V. ANÁLISIS DE TÉCNICAS INFORMÁTICAS

Se analizaron varias técnicas que facilitaran encontrar patrones de similitud en muestras de vectores de datos, y la minería de datos proporciona una variedad de posibilidades con este propósito.

Al analizar los estudios similares ocupados por otras investigaciones recientes, citadas en el capítulo de antecedentes, y sobre todo comprendiendo los procesos y resultados obtenidos, se decide por ocupar la técnica de clústeres para agrupar datos similares y en base a este análisis descubrir patrones de comportamiento inmersos en los vectores de datos.

La técnica de análisis de información por clústeres es una de varias ocupadas en la minería de datos para el descubrimiento de información no evidente en un conjunto de datos. En esta técnica se agrupan los datos en grupos que tengan características similares entre ellos. Para evaluar las características que los asemejan se ocupan funciones de similitud en base a las que se definen los más semejantes y se agrupan en unidades llamadas clústeres.

### 1. Técnicas para el análisis de datos por clústeres

Se utiliza la metodología de análisis de clústeres, dentro de las posibles de minería de datos, con la intención de identificar patrones similares entre los vectores de datos seleccionados del objeto de estudio.

Existen varias funciones de similitud para el agrupamiento de los vectores de datos, las cuales se explican a continuación en base a que:

***Sean dos vectores  $[a, b]$***

- i. **Distancia Manhattan**, que es la distancia obtenida entre los elementos del vector tomando como distancia entre ellos la sumatoria de la diferencia en valor absoluto existente entre los miembros ordenados del vector:

$$||a - b|| = \sum_i |a_i - b_i|$$

- ii. **Distancia Euclideana** es una de las opciones considerada como de las mejores aproximaciones para valores numéricos discretos, y que se expresa a través de la siguiente expresión matemática para el vector de datos anterior:

$$||a - b|| = \sqrt{\sum_i (a_i - b_i)^2}$$

- iii. **Raíz Cuadrada de la Distancia Euclideana** es la una de las ocupadas opciones por diferentes aplicaciones matemáticas, y que es considerada como una de las mejores aproximaciones para valores numéricos discretos y que se muestra en la siguiente expresión matemática:

$$||a - b|| = \sum_i (a_i - b_i)^2$$

- i. **K-means** es un método de agrupamiento en donde se determina el numero de clústeres K a definir, ocupando la siguiente expresión:

$$\arg \min_s \sum_{i=1}^k \sum_{a_j \in S_i} ||a_j - b_i||^2$$

Donde se expresa que se agruparan los datos en K clústeres aleatorios generando centroides para cada uno de ellos, y buscando la mínima distancia entre los centroides y los demás puntos geométricos. Después se deben recalculan los centroides y reasignarlos iterativamente

hasta un criterio de paro, por ejemplo la distancia máxima entre los miembros de diferentes clústeres.

Especialmente para versión de SQL Server este modelo es adaptado a usar probabilidades y trabajar con una muestra de dos datos alimentados a la base de datos.

- ii. **C-means** es un método de agrupamiento en donde un punto en lugar de tener un solo grupo de pertenencia, tiene un grado de pertenencia a varios grupos, ocupando un concepto llamado lógica difusa. Es por esto que se puede decir que los puntos en el límite del clúster, pertenecen en cierta medida al clúster más cercano, pero también al cúmulo completo que forma todos los clústeres. De esta manera, el grado en que un punto pertenece a un grupo, más su grado de pertenencia a todos los demás, es igual a uno. Es como manejar una cierta probabilidad de pertenencia.

$$\forall x \left( \sum_{k=1}^N \vartheta_k(x) = 1 \right)$$

Es muy similar al algoritmo K-means, aunque sus resultados no son la pertenecía total, sino graduada a varios clústeres. [lógica difusa].

- iii. **EM Clustering** es un método que ocupa un algoritmo que iterativamente redefine un modelo inicial de clúster para que los datos se ajusten, y determina la probabilidad de que los datos existan en un determinado clúster. Este método ocupa un poco de la lógica difusa según Microsoft comenta, y el algoritmo se basa en encontrar la mayor distancia entre los datos de N clústeres, hasta que iterativamente se encuentre estable. Este es el default que ocupa SQL Server en su herramienta de servicios de análisis.
- iv. **Dice Dissimilarity** es una función de similaridad para valores booleanos. Es equivalente  $(n_{10} + n_{01}) / (2n_{11} + n_{10} + n_{01})$ , where  $n_{ij}$  is the number of corresponding pairs of elements in u and v respectively equal to i and j.

- v. **Jaccard Dissimilarity** es también utilizada para vectores con valores booleanos.
- vi. **Distancia de Hamming** es una función utilizada para vectores booleanos que determina su valor en relación con los elementos del vector que son diferentes tomándolo como una cadena de caracteres. Mide el número mínimo de sustituciones necesarias para cambiar a un miembro de un clúster a otro.
- vii. **Edit Distance** es la distancia de igualdad entre dos vectores y determina diferencias. Es ocupada al agrupar cadenas de caracteres.
- viii. Adicionalmente existen más métodos de agrupación que genera los clústeres por cercanías similares entre ellos están los métodos de: **Agglomerate, QT, hash sensible, Distancia Mahalanobis**, y varios mas.

### **3. Selección de Herramientas informáticas**

#### **a. Investigación de aplicaciones**

Esto no representa un análisis comparativo exhaustivo de las herramientas analizadas, debido a que no se pretende calificar la mejor herramienta de análisis de datos para minería, y específicamente para manejo de clústeres. En vez de eso se utilizaron las mismas para analizar un conjunto de datos muestra, con la finalidad de comprender la información que nos podían dar con el procesamiento de los datos, con el objetivo de aplicarlos posteriormente a los datos de estudio.

En este proceso se investigaron y ocuparon las siguientes herramientas aplicativas:

- SQL Server Analisis Services de Microsoft/ sobre SQL Express 2005 / 2008.
- Aplicación WEKA 3.4.12 /Mysql

- Wolfram Mathematica 7.0 trial Version

## b. Evaluación comparativa

### SQL Server Analysis Services de Microsoft.

Esta es una herramienta que se integra a la base de datos de Microsoft SQL Server, y que existe en versiones para base de datos 2005 y 2008. La herramienta de análisis corre bajo Visual Studio en particular se probó con la versión 2008 y que se ejecuta con la opción de Business Intelligence Development Studio.

Se ocupó el algoritmo de agrupamiento llamado K-mean para analizar resultados, por ser lo más parecido al método de Euclidean Distance, ocupado para comparar.

Los resultados directos obtenidos del proceso de clustering se muestran en la Figura 5.1.

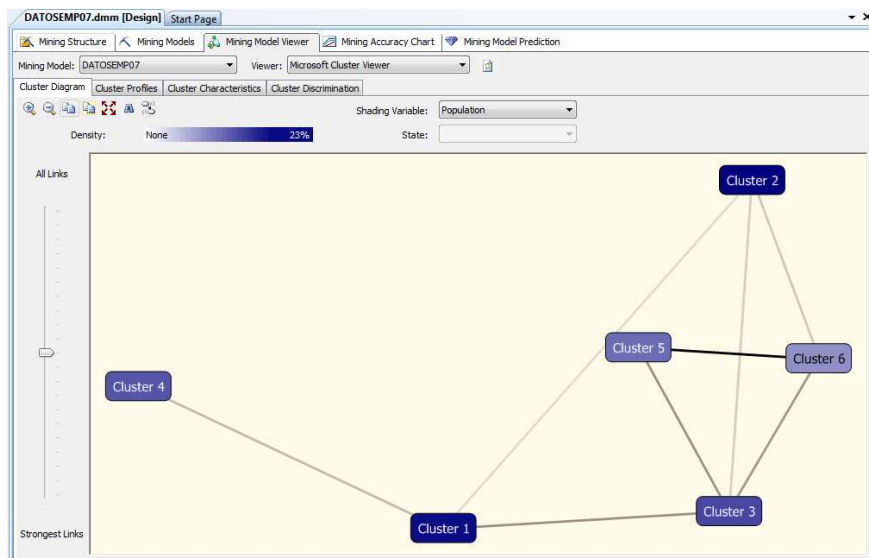


Figura 5.1. Proceso con SQL Business Intelligence.

Los resultados indirectos del proceso de agrupamiento, que son características de los clusters en cuanto a la población que los forman, o

estadísticas que determina como la media o desviación estandar, se presentan como ejemplo a continuación, en la Figura 5.2 y 5.3.

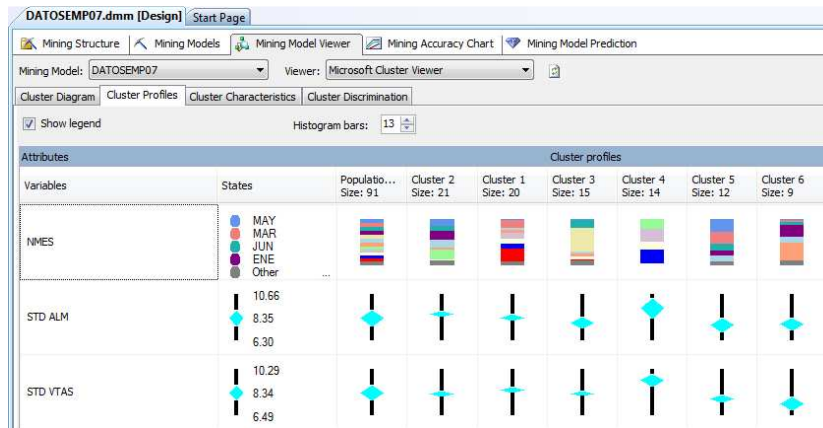


Figura 5.2. Indirectos del proceso con SQL Business Intelligence.

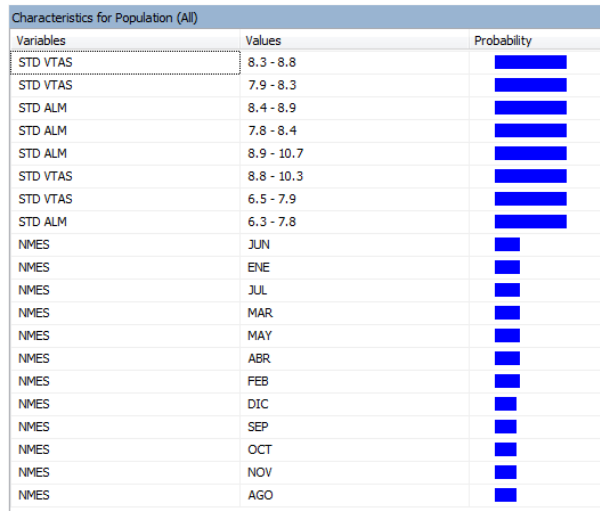


Figura 5.3. Indirectos del proceso con SQL Business Intelligence.

En esta herramienta si se requiere de conocer a los integrantes del clúster, la aplicación muestra una expresión en SQL, llamada un Query con el cual se pueden obtener probabilísticamente los datos integrantes. Un problema es que para conocer los integrantes de cada clúster, es necesario explotar la base de datos con un Query, lo cual ya dificulta este proceso.



Además es necesario considerar que el análisis se puede realizar sobre una muestra y es probabilístico, por lo que es posible que al agrupar en clústeres, en las gráficas que muestra la herramienta, el número de integrantes del clúster sea de N, pero al explotar la base de datos con el Query del clúster, se obtengan M integrantes, donde  $M \neq N$ . No se obtienen los datos seguros integrantes del clúster. Es decir, no se pueden identificar ni con los Querys los integrantes puntuales de un clúster.

Tiene un área de predicción basada en estadística, que se analizó para algunos datos históricos, y la predicción se obtenía con grados de error de un 23% en promedio para 20 ejemplos ejecutados, con respecto a las cifras reales [MacLennan, 2009].

#### Aplicación WEKA 3.4.12

Es una aplicación que agrupa algoritmos de minería de datos. Esta desarrollada en Java por la Universidad de Waikato, en Nueva Zelanda. La aplicación es de open source y contiene herramientas de pre procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. En la Figura 5.4 se muestra la corrida de información que se realizó sobre este aplicativo.

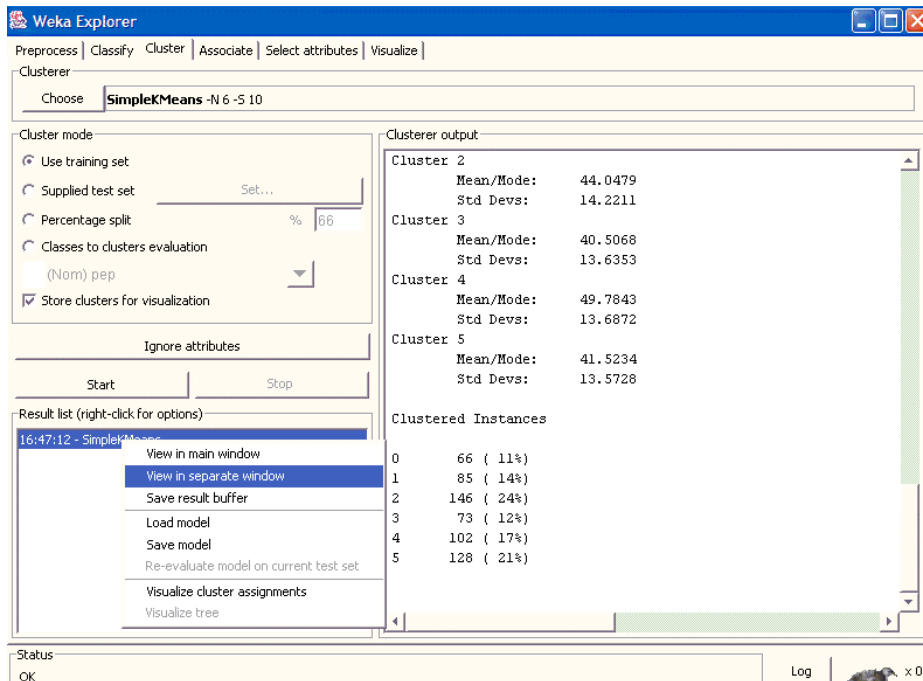


Figura 5.4. Análisis de indirectos con Weka.

Requiere para su ejecución contar con una versión de java para ejecución de código y dependiendo de la versión de WEKA es la versión de java que ocupa.

Para su utilización se instalo la base de datos MySql de software libre para almacenar los datos de trabajo en ella, solo se definió una tabla de trabajo, y en ella se vaciaron los datos. Se realizaron los procedimientos de conexión de la base de datos con WEKA desde la aplicación, y se realizaron pruebas.

Esta herramienta tiene integrados varios algoritmos de minería de datos para generar clústeres, y con la finalidad de comparar resultados se ocupo el método de K-means, ya mencionado, y se obtuvieron resultados con las siguientes observaciones:

- a) Los resultados al generar clústeres por este método, son similares a los obtenidos con la aplicación Wolfram Mathematica, aunque no idénticos, al trabajar con los mismos parámetros de agrupación.
- b) Las herramientas de agrupación de clústeres permiten la clasificación, y adicionalmente obtienen resultados indirectos estadísticos como la media, desviación estándar, y estadísticas de integración de datos.
- c) Es un poco complicado determinar los datos integrantes de cada clúster, porque no los muestra explícitamente.
- d) Contiene muchas herramientas adicionales de visualización para solución de problemas específicos, como por ejemplo visualizadores en cuadros cruzados para cada una de las variables.
- e) No se pueden obtener gráficas de tres dimensiones para visualizar la forma de agrupación de tres dimensiones, con la finalidad de determinar algún comportamiento racional y visual.

*Wolfram Mathematica 7.0 trial Version.*

Esta es una herramienta utilizada en áreas científicas, de ingeniería, matemáticas y áreas computacionales. Originalmente fue concebido por Stephen Wolfram, quien continúa siendo el líder del grupo de matemáticos y programadores que desarrollan el producto en Wolfram Research, compañía ubicada en Champaign, Illinois. Comúnmente considerado como un sistema de álgebra computacional, Mathematica es también un poderoso lenguaje de programación de propósito general. En la Figura 5.5 se muestra la corrida de información ejecutada con esta aplicación.

A grandes rasgos se pueden enlistar las siguientes herramientas de la aplicación:

- Bibliotecas de funciones elementales y especiales para matemáticas.
- Herramientas de visualización de datos en 2D y 3D.

- Matrices y manipulación de datos, así como soporte de matrices tipo "sparse".
- Capacidad de solucionar sistemas de ecuaciones, ya sea ordinarias, parciales o diferenciales, así como relaciones de recurrencia y algebraicas en general.
- Herramientas numéricas y simbólicas para cálculo de variable continua o discreta.
- Estadística multivariable.
- Restringida y no restringida optimización de local y global.
- Lenguaje de programación que soporta programación funcional.
- Un kit de herramientas para añadir interfaces de usuario para cálculos y aplicaciones.
- Herramientas para procesamiento de imágenes.
- Herramientas de análisis y visualización.
- Minería de datos, como análisis de clústeres, alineamiento de secuencias.
- Capacidades de importación y exportación de información de datos, imágenes, video y sonido, así como otros formatos biomédicos y de intercambio de documentos en general.
- Una colección de bases de datos incluidas de matemáticas, ciencia e información socio económica (astronomía, diccionarios, clima, poliedros, países, instrumentos financieros, componentes químicos, el genoma humano, entre otros).
- Soporte para variable compleja, aritmética de precisión infinita y computación simbólica para todas las funciones incluidas.

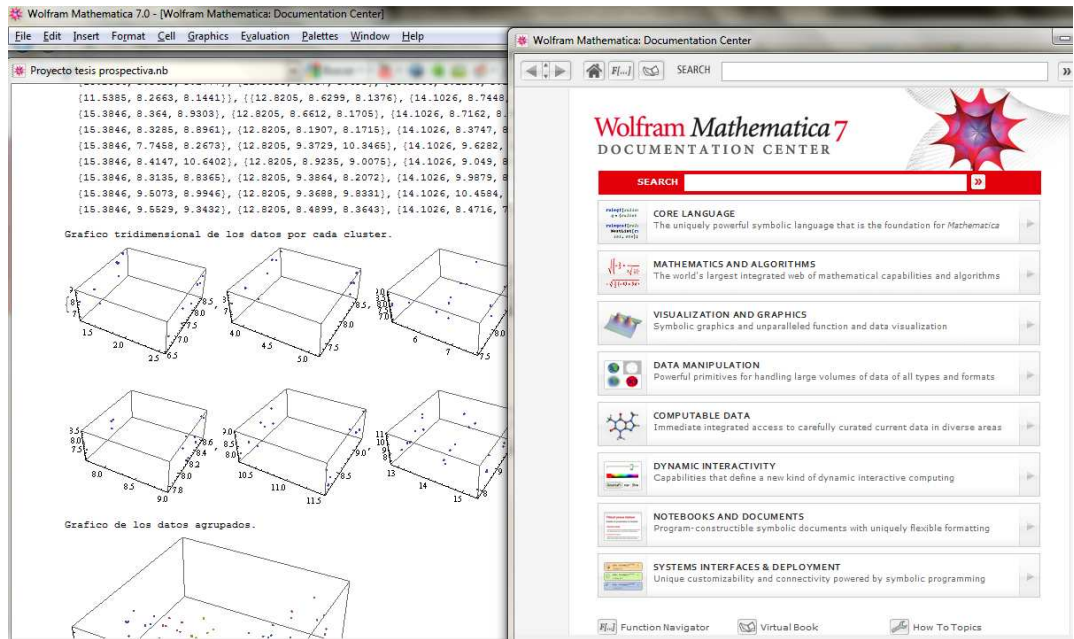


Figura 5.5. Ambiente de trabajo con Wolfram Mathematica.

Esta herramienta cuenta con varios algoritmos de minería de datos para clústeres, y con la finalidad de comparar resultados se ocupó el método de Euclidean Distance, ya mencionado, y se obtuvieron resultados con las siguientes observaciones:

- Existe un grado de similitud entre los resultados obtenidos por mathematica y Weka al ocupar los mismos algoritmos de agrupamiento circunscribiéndose a algoritmos no probabilísticos.
- El medio ambiente de trabajo es a través de primitivas de programación, muy complejas pero de resultados inmediatos.
- La forma de procesar datos tiene como resultado un conjunto de datos directos agrupado que muestra directamente con una notación que visualmente facilita su identificación.
- Cuenta con una forma de visualización en tres dimensiones que facilita la observación de la formación de patrones de comportamiento en los datos.

- e) Ya una vez obtenidas las gráficas es posible manipularlas en su forma gráfica, y cambiar los ejes de posición para observar las lecturas que integran los clústeres desde la perspectiva de diferentes ejes.
- f) Es sencillo permutar diferentes datos de origen para ver su comportamiento con los mismos algoritmos de procesamiento de clústeres, lo cual facilita su elección dentro del universo de posibles combinaciones.
- g) Permite realizar programas que automatizan con cierta facilidad el proceso completo de un algoritmo matemático, facilitando y flexibilizando su operación para realizar varias iteraciones con diferentes datos o grados de avance.
- h) Permite trabajar con algoritmos de generación de clúster determinísticos, evitando los algoritmos probabilísticos debido a que la cantidad de datos que se requieren procesar son finitos y no muy extensos.
- i) Permite determinar con herramientas sencillas las características de los clústeres, y las estadísticas de cada grupo, para conocer su conformación.

### **c. Selección de herramientas y justificación**

La selección de herramientas se realiza en función a la utilidad de cada una de ellas para el modelo, y a la necesidad de hacer más transparente la interface que facilite la migración de información. De esta manera la selección queda de la siguiente manera:

- a) Se escoge el transporte directo a la hoja de cálculo Excel para los datos obtenidos del sistema anterior administrativo de la empresa.

- b) Se escoge para la extracción de la información la base de datos SQL Server de Microsoft, solo para hacer la extracción de los datos del sistema ERP.
- c) Estos datos se concentran finalmente en la hoja de datos de Excel para realizar desde ahí análisis gráfico y estandarización de la información.
- d) Una vez estandarizados los datos se procede a migrarlos a un archivo que permita su importación hacia la aplicación Mathematica para el procesamiento de datos, aplicando herramientas de minería de datos.
- e) La relación de los casos históricos se alimenta y mantiene en Excel solo para relacionarla con los resultados obtenidos.

## **VI. ANÁLISIS PRELIMINAR DE DATOS**

Comprendiendo los objetivos de la empresa de estudio, y de acuerdo con los directivos de la misma, se seleccionaron de los 92 datos extraídos de la base de datos histórica (presentada una muestra en el apéndice A), un conjunto de 9 datos de los que más significantes. Se realizaron con estos datos, tareas de análisis con la finalidad de buscar algunos de los que tenían gráficamente un comportamiento que denotaba la posibilidad de encontrar patrones. Estos dos criterios se utilizaron y se obtuvieron interesantes resultados intermedios para mi experiencia y para el beneficio de la empresa de estudio.

### **1. 1ª selección de datos a analizar**

Se tomaron los 9 datos a analizar por en triadas integradas por un indicador del mes representante de las lecturas, y parejas de datos considerando la selección entre las siguientes:

- a) Valores en bancos
- b) Ventas a clientes
- c) Valor global del almacén
- d) Pasivo a proveedores nacionales
- e) Pasivo de proveedores extranjeros
- f) Capital social
- g) Total del costo de ventas
- h) Utilidad Operativa
- i) Pagos con Notas de crédito (derivado)



Se realiza el análisis de esta forma proyectando que su visualización gráfica directa y su visualización en gráficas de clústeres, es un elemento importante para considerar la posibilidad de encontrar comportamientos ocultos y dar pie a investigaciones más profundas.

En el caso de este último dato se realizó un análisis al comprender la importancia de la cobranza y de la necesidad de monitorear este evento, bajo petición de la empresa objeto de estudio.

## **2. Análisis de gráficos y patrones**

Al analizar las gráficas de cada uno de los datos de forma independiente, se observaron que algunos de ellos tenían comportamientos cíclicos anuales, los cuales denotaban un posible patrón de comportamiento. Es importante hacer notar que hasta este punto los datos se consideran naturales como son obtenidos en bruto del sistema administrativo, por lo que muchas gráficas presentan los incrementos propios del crecimiento en actividad de la empresa de estudio.

El factor de manejar los datos originales, es una barrera al momento de realizar un análisis de clústeres, debido a que cifras similares en cantidades pueden no representar estadios similares en el comportamiento histórico de la empresa. Esto se debe a que por ejemplo las cifras de ventas a clientes de febrero del año 2004 son similares a las cifras de noviembre del año 2003, pero en su representación anual o total (tomando los totales de todos los datos de ese concepto), no tienen la misma representación, y por lo tanto no corresponden al mismo estadio con respecto a las ventas a los clientes.

3. Las tablas de los datos originales de conceptos elegidos y sus gráficas correspondientes se encuentran detalladas en el Apéndice B.

#### 4. Análisis de pertinencia de estandarización

Con la necesidad de analizar estadios similares en los 9 datos acordados de análisis, se realizaron dos tipos de estandarizaciones para evaluar la que mejor resultados nos daba:

- a) Se realizó una porcentualización de los datos de un concepto (ej. bancos), con respecto a los totales de del mismo en toda la historia acotada por datos de la empresa objeto de estudio. Esta estandarización presenta las siguientes características:
  - a. Los datos estandarizados que se obtenían como resultado de cada dato original, estaban representados por cifras muy pequeñas, y con poca diferencia entre ellas.
  - b. Para cada uno de los datos a evaluar prospectivamente se tendría que calcular el total del dato del concepto al aumentar el número total de datos, y los porcentajes de representación también variarían cada vez, haciendo difícil su comprensión y relación con casos históricos.
  - c. Analizando gráficamente las cifras estandarizadas el resultado eran gráficas muy planas y con poca diferenciación.
- b) Se realizó una estandarización porcentualizando los datos de un concepto con respecto a los totales anuales del mismo, en la historia acotada por datos de la empresa objeto de estudio. Esta estandarización presenta las siguientes características:
  - a. Los datos que se obtenían incluyendo el dato de mes de operación, tenían como resultado datos estandarizados representados por cifras mayores que el caso anterior, y con mayor diferenciación.
  - b. Los datos estandarizados a evaluar prospectivamente mantienen los mismos valores al aumentar el número total de datos. En el caso de los años nuevos en donde no se cuenta con una cantidad anual, se

realiza un análisis de regresión lineal para proyectar el valor anual del concepto a analizar y de ahí partir para la estandarización.

- c. Las gráficas resultado del cruce de información del mes con los datos estandarizados de los 9 conceptos de estudio eran más diferenciadas, y se presumió un mejor resultado al analizarlas con las herramientas de generación de clústeres.

## 5. Datos: Estandarización

Se realizó una estandarización porcentualizando los datos de un concepto con respecto a los totales anuales del mismo con las ventajas analizadas en el punto anterior.

De esta manera se logra que la contribución de cada elemento a la función de similitud para el algoritmo de clúster sea la misma entre cada uno de los elementos que forman el vector que se compara. Además así se logra no influir de manera personal los resultados de la función de similitud.

A continuación se muestra la Tabla 6.1 como ejemplo de la forma en que se estandarizaron los datos:

PROCEDIMIENTO PARA OBTENER DATOS ESTANDARIZADOS						
ANO	MES	CLIENTES	ALMACEN	PMES	P.CLIENTES	P.ALMACEN
2006	1	15,898,325	12,387,695	1.28	8.03	7.50
	2	15,565,578	12,337,242	2.56	7.86	7.47
	3	16,155,699	12,906,388	3.85	8.16	7.81
	4	16,464,862	12,567,332	5.13	8.31	7.61
	5	15,712,592	13,158,014	6.41	7.93	7.97
	6	15,539,285	15,099,881	7.69	7.85	9.14
	7	16,083,683	14,175,121	8.97	8.12	8.58
	8	17,159,972	13,786,098	10.26	8.66	8.35
	9	17,407,727	15,033,092	11.54	8.79	9.10
	10	17,672,651	14,877,933	12.82	8.92	9.01
	11	17,921,240	14,247,719	14.10	9.05	8.63
	12	16,464,526	14,595,382	15.38	8.31	8.84
SUMATORIA	78	198,046,139.92	165,171,896.93	100.00	100.00	100.00

Tabla 6.1. Muestra del proceso de estandarización de datos.

Además las tablas de los datos estandarizados de conceptos elegidos y sus gráficas correspondientes se encuentran detalladas en el Apéndice C.

## **6. Selección final de datos por modelar**

Derivado del análisis gráfico de los 9 elementos estandarizados y en función a presentar un resumen a la dirección de la empresa objeto de estudio y tomando en cuenta sus intereses personales por los diferentes conceptos, se determino realizar el modelado de dos conceptos. Es importante notar que en caso de que una vez realizado el modelado de los conceptos, si estos no dieran resultados de prospección aceptables se tendría que tomar la decisión de cambiar de conceptos, y en el último de los casos de cambiar de modelo.

Además de que al analizar los conceptos por modelar, se observaron singularidades que dificultaban su estudio, como por ejemplo:

- a) Datos incompletos (por no haber sido considerados en los meses en cuestión).
- b) Datos con errores por haber sido considerados por su naturaleza diferente, aunque agrupados bajo el mismo nombre.
- c) Datos que históricamente cambiaron su naturaleza intrínseca de definición.
- d) Datos que aparecen en periodos de tiempo recientes que antes no habían sido considerados.
- e) El criterio de la empresa objeto de estudio para seleccionar de los conceptos elegibles, aquellos que les fueran más útiles desde el punto de vista de la prospección dentro de los planes establecidos.

Los 3 elementos seleccionados cuantitativos (el mes de estudio de los datos más dos) son definidos como: mes, ventas a clientes, valor del almacén.

## **7. Datos: Análisis de singularidades**

Existe en la empresa objeto de estudio, datos de conceptos que aun considerando que estén en la medida de lo posible completos, presentan saltos en sus valores, o singularidades que se deben a sucesos especiales de gran impacto que provocan variaciones que lo pueden sacar de el estadio que le corresponde por historicidad. Se contempla mantener estas singularidades sin suavizar por las siguientes razones:

- a) Son casos especiales que se pueden volver a presentar en la historia de la empresa, y que de la singularidad bien documentada se puede obtener elementos de apoyo al razonamiento de eventos pasados y elementos de apoyo para la toma de decisiones de eventos futuros.
- b) Estos casos se pueden agrupar en clústeres de casos especiales que pueden tener un conocimiento intrínseco no descubierto pero que pueden arrojar elementos para el tomador de decisiones como hombre administrativo basado más en hechos que en valores.
- c) Al estar el mayor número de los casos históricos de la empresa registrados en un sistema automatizado, permitirá comprender mejor las singularidades y la presentación de singularidades en el futuro.

## **8. Pertinencia de Análisis bayesiano y/o de correlaciones**

Se realizaron algunas pruebas de correlación de datos entre los diferentes 9 conceptos, para obtener la relación de dependencia entre ellos, y no se encontró una dependencia que determinara el agruparlos para poder abarcar mas conceptos en uno solo. Además de que por la naturaleza de las cifras no se establecía una dependencia razonable.

Es importante para un sistema que apoya el razonamiento de la decisión que tome un decisor de una empresa, que los conceptos que se

sugieren como proyección, sean lo más claros posibles, para que él los pueda comprender en su justa dimensión.

Si además se presentarán estos datos de conceptos como cifras antecedentes y consecuentes de eventos, y cifras sugeridas proyectadas, sería difícil con certeza desintegrar un concepto compuesto por varios conceptos originales, y el grado de posible falla en la manipulación numérica aumentaría errores finales.

## **9. Análisis gráfico de clústeres en tres dimensiones**

Como una necesidad se planteo hacer un análisis de los datos estandarizados en varios clústeres, obteniendo así gráficas de agrupamiento en clústeres desde 2 hasta 12 y observar visualmente su resultado. Para los datos de Diciembre de 2009 se obtienen por ejemplo los siguientes resultados en vectores de tres elementos (mes, ventas a clientes, valor del almacén).

Este análisis se realiza con la aplicación Wolfram Mathematica 7.0, tomado para el mismo los datos del mes de diciembre de 2009 estandarizados y aplicando el algoritmo de agrupación de datos por distancia euclidiana y se obtiene el resultado mostrado en la Gráfica 6.2.

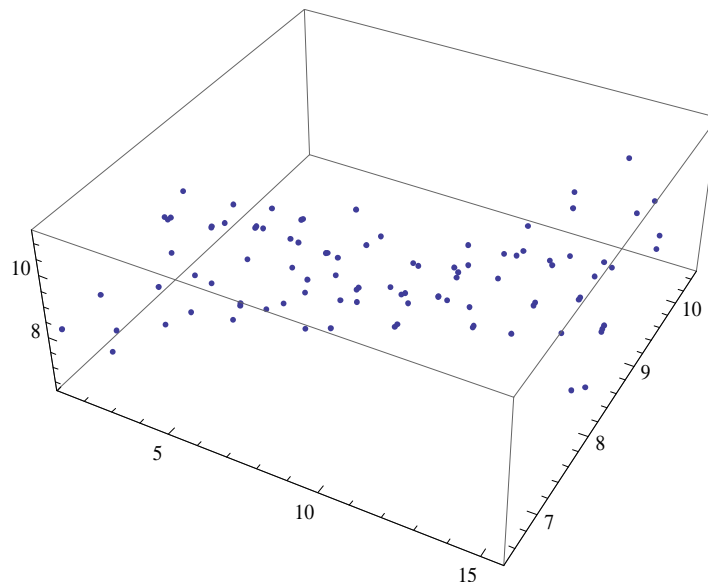
```

In[1]= datos = Import["c:\\md\\dm.dat"]
Out[1]= {{1.2821, 8.0176, 8.6622}, {2.5641, 8.1577, 8.4998}, {3.8462, 8.3259, 8.6036}, {5.1282, 8.4922, 8.8759},
{6.4103, 8.3216, 8.4906}, {7.6923, 8.2182, 7.9435}, {8.9744, 8.1986, 7.0714}, {10.2564, 8.2928, 8.3092},
{11.5385, 8.2367, 7.8971}, {12.8205, 8.6299, 8.1376}, {14.1026, 8.7448, 8.5788}, {15.3846, 8.364, 8.9303},
{1.2821, 8.0626, 8.646}, {2.5641, 8.1631, 8.5342}, {3.8462, 8.3069, 8.5875}, {5.1282, 8.523, 8.8419},
{6.4103, 8.2944, 8.5418}, {7.6923, 8.1833, 7.9521}, {8.9744, 8.2447, 7.0581}, {10.2564, 8.2983, 8.3182},
{11.5385, 8.2179, 7.8823}, {12.8205, 8.6612, 8.1705}, {14.1026, 8.7162, 8.5712}, {15.3846, 8.3285, 8.8961},
{1.2821, 7.9761, 8.8288}, {2.5641, 8.3633, 8.2559}, {3.8462, 8.4302, 8.3239}, {5.1282, 8.4307, 8.2305},
{6.4103, 8.823, 8.9579}, {7.6923, 8.6175, 8.8463}, {8.9744, 8.4397, 7.356}, {10.2564, 8.465, 7.8217},
{11.5385, 8.1433, 8.7216}, {12.8205, 8.1907, 8.1715}, {14.1026, 8.3747, 8.2187}, {15.3846, 7.7458, 8.2673},
{1.2821, 7.1569, 6.6511}, {2.5641, 7.3798, 6.7989}, {3.8462, 7.8743, 6.3015}, {5.1282, 7.8717, 7.1044},
{6.4103, 7.9161, 6.7662}, {7.6923, 7.7433, 7.6053}, {8.9744, 8.5273, 8.5307}, {10.2564, 8.8442, 8.8834},
{11.5385, 9.2705, 9.0591}, {12.8205, 9.3729, 10.3465}, {14.1026, 9.6282, 11.3127}, {15.3846, 8.4147, 10.6402},
{1.2821, 8.0276, 7.4999}, {2.5641, 7.8596, 7.4693}, {3.8462, 8.1575, 7.8139}, {5.1282, 8.3136, 7.6086},
{6.4103, 7.9338, 7.9663}, {7.6923, 7.8463, 9.1419}, {8.9744, 8.1212, 8.582}, {10.2564, 8.6646, 8.3465},
{11.5385, 8.7897, 9.1015}, {12.8205, 8.9235, 9.0075}, {14.1026, 9.049, 8.626}, {15.3846, 8.3135, 8.8365},
{1.2821, 7.0135, 8.1863}, {2.5641, 7.3411, 8.1718}, {3.8462, 7.6018, 8.1535}, {5.1282, 7.4852, 8.0734},
{6.4103, 7.5955, 8.3312}, {7.6923, 7.9101, 8.1832}, {8.9744, 8.3867, 7.8153}, {10.2564, 8.5841, 8.6711},
{11.5385, 9.2005, 8.3652}, {12.8205, 9.3864, 8.2072}, {14.1026, 9.9879, 8.8471}, {15.3846, 9.5073, 8.9946},
{1.2821, 6.4884, 8.1916}, {2.5641, 6.6532, 7.5065}, {3.8462, 7.2818, 7.8703}, {5.1282, 7.4914, 8.1391},
{6.4103, 7.9791, 8.3059}, {7.6923, 8.1848, 7.527}, {8.9744, 8.6229, 8.242}, {10.2564, 8.8613, 8.1777},
{11.5385, 9.057, 8.498}, {12.8205, 9.3688, 9.8331}, {14.1026, 10.4584, 8.3655}, {15.3846, 9.5529, 9.3432},
{1.2821, 8.2746, 9.133}, {2.5641, 8.5156, 8.5972}, {3.8462, 8.5878, 8.6958}, {5.1282, 8.3052, 8.6099},
{6.4103, 8.4916, 7.9838}, {7.6923, 8.3602, 9.0718}, {8.9744, 8.31587, 7.9177}, {10.2564, 8.6368, 8.2235},
{11.5385, 8.6882, 8.4939}, {12.8205, 8.98197, 8.73727}, {14.1026, 9.25103, 8.68143}, {15.3846, 8.0554, 7.624}}

```

Gráfica 6.2. Vectores de datos estandarizados de diciembre/2009.

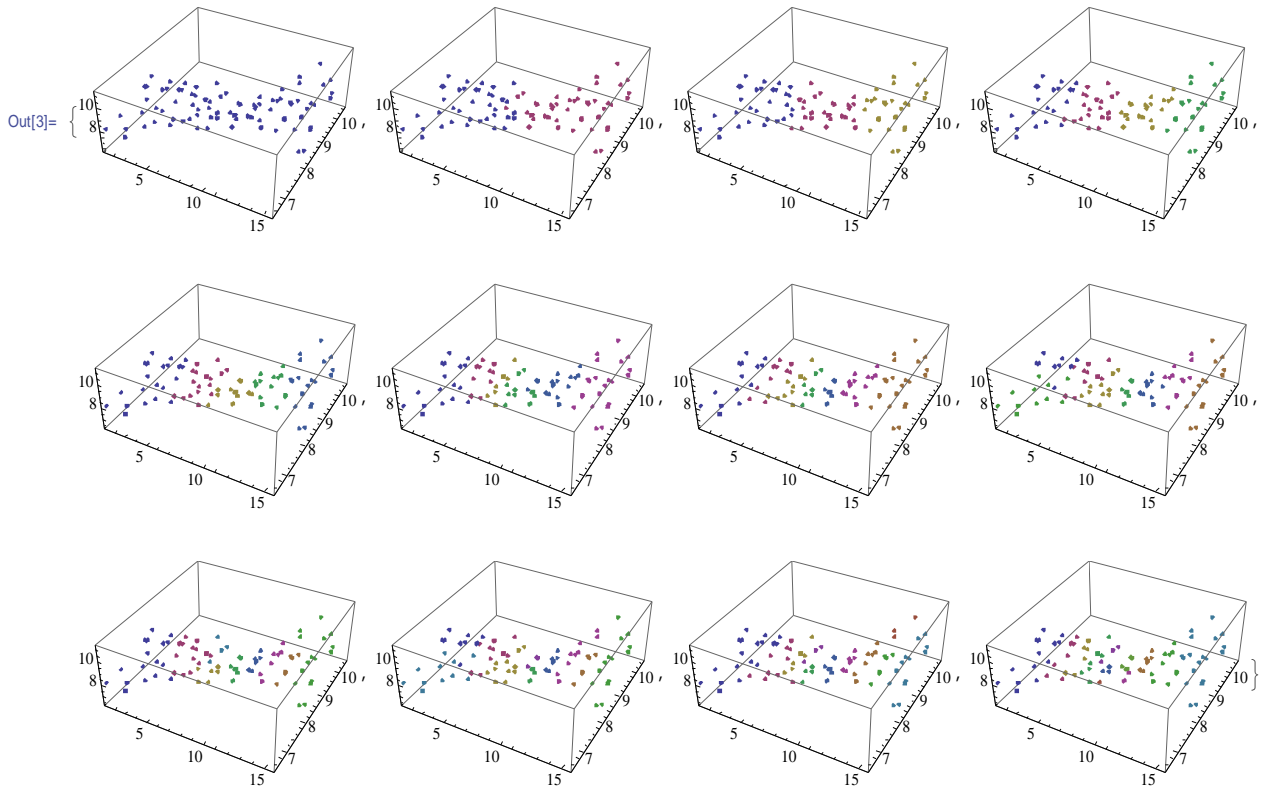
Esta gráfica se obtiene de aplicar la función de graficación en tres dimensiones, de todos los puntos de vectores de datos, la finalidad es observar la dispersión y los resultados se ven en la Gráfica 6.3.



Gráfica 6.3. Gráfica de clústeres de diciembre/2009.

Con los vectores de datos estandarizados se generan ensayos de 12 gráficas de 1 a 12 clústeres en donde se puede analizar visualmente la agrupación los que se muestran en la Gráfica 6.4.

```
In[3]:= Map[ListPointPlot3D[FindClusters[datos, #]] &, Range[12]]
```



Gráfica 6.4. Gráfica de agrupaciones de 1 a 12 clústeres de diciembre/2009.

Analizando las gráficas anteriores se determina que 6 clústeres es lo más adecuado para visualmente reconocer patrones de similitud y agrupamientos consistentes con los grupos que representan, lo cual se corrobora con los datos y gráficos mostrados en la Gráfica 6.5.

```
grupos=FindClusters[datos,6]
{
```



#### Cluster 1

{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{3.8462,7.8743,6.3015},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{3.8462,7.6018,8.1535},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{3.8462,7.2818,7.8703},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}},

#### Cluster 2

{{3.8462,8.3259,8.6036},{5.1282,8.4922,8.8759},{3.8462,8.3069,8.5875},{5.1282,8.523,8.8419},{3.8462,8.4302,8.3239},{5.1282,8.4307,8.2305},{5.1282,7.8717,7.1044},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086},{5.1282,7.4852,8.0734},{5.1282,7.4914,8.1391},{3.8462,8.5878,8.6958},{5.1282,8.3052,8.6099}},

#### Cluster 3

{{6.4103,8.3216,8.4906},{6.4103,8.2944,8.5418},{6.4103,8.823,8.9579},{6.4103,7.9161,6.7662},{6.4103,7.9338,7.9663},{6.4103,7.5955,8.3312},{6.4103,7.9791,8.3059},{6.4103,8.4916,7.9838}},

#### Cluster 4

{{7.6923,8.2182,7.9435},{8.9744,8.1986,7.0714},{7.6923,8.1833,7.9521},{8.9744,8.2447,7.0581},{7.6923,8.6175,8.8463},{8.9744,8.4397,7.356},{7.6923,7.7433,7.6053},{7.6923,7.8463,9.1419},{7.6923,7.9101,8.1832},{8.9744,8.3867,7.8153},{7.6923,8.1848,7.527},{7.6923,8.3602,9.0718},{8.9744,8.31587,7.9177}},

#### Cluster 5

{{10.2564,8.2928,8.3092},{11.5385,8.2367,7.8971},{10.2564,8.2983,8.3182},{11.5385,8.2179,7.8823},{10.2564,8.465,7.8217},{11.5385,8.1433,8.7216},{8.9744,8.5273,8.5307},{10.2564,8.8442,8.8834},{11.5385,9.2705,9.0591},{8.9744,8.1212,8.582},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6711},{11.5385,9.2005,8.3652},{8.9744,8.6229,8.242},{10.2564,8.8613,8.1777},{11.5385,9.057,8.498},{10.2564,8.6368,8.2235},{11.5385,8.6882,8.4939}},

#### Cluster 6

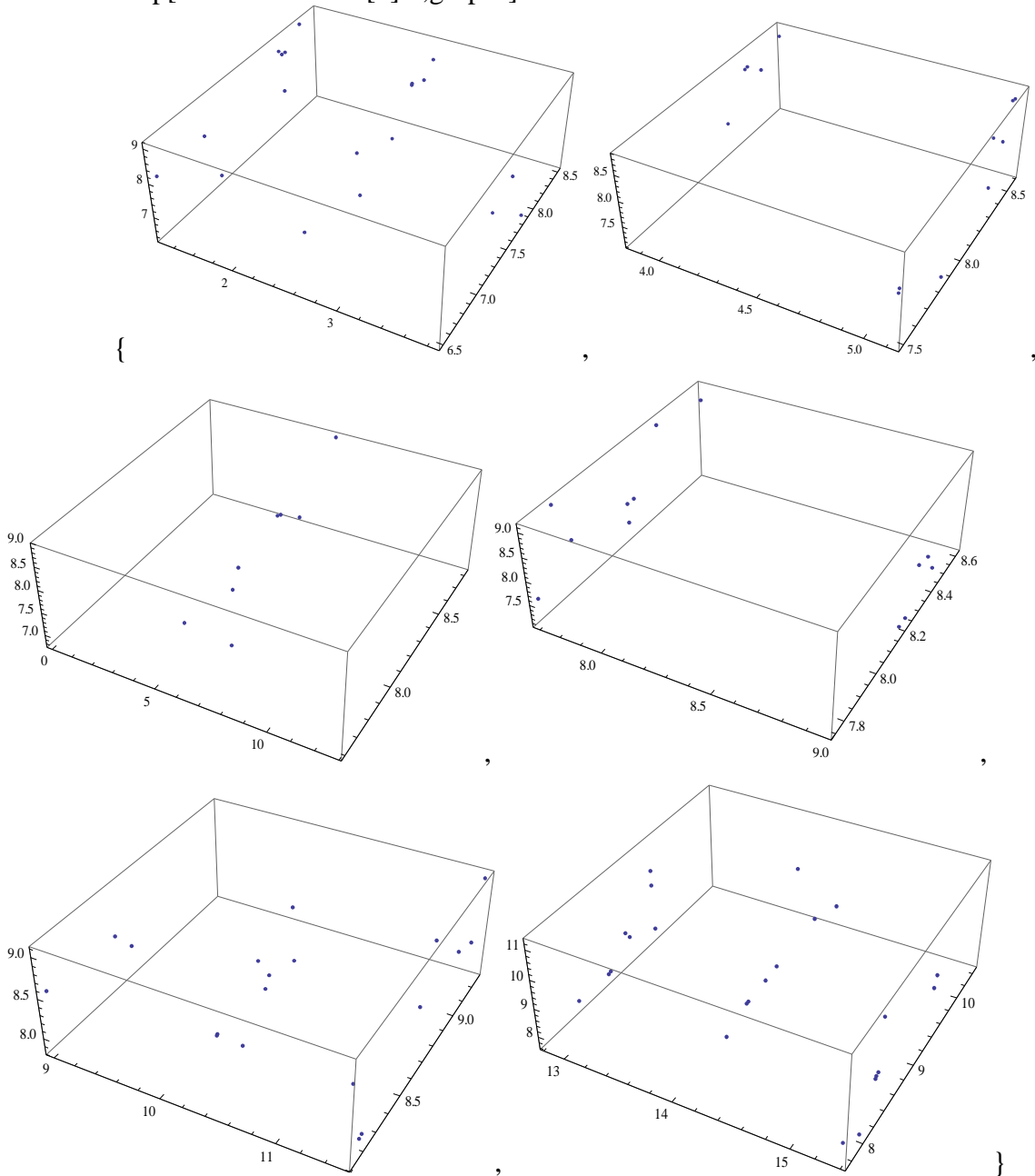
{{12.8205,8.6299,8.1376},{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{12.8205,8.6612,8.1705},{14.1026,8.7162,8.5712},{15.3846,8.3285,8.8961},{12.8205,8.1907,8.1715},{14.1026,8.3747,8.2187},{15.3846,7.7458,8.2673},{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},{14.1026,9.049,8.626},{15.3846,8.3135,8.8365},{12.8205,9.3864,8.2072},{14.1026,9.9879,8.8471},{15.3846,9.5073,8.9946},{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432},{12.8205,8.98197,8.73727},{14.1026,9.25103,8.68143},{15.3846,8.0554,7.624}}

}

Gráfica 6.5. Datos del agrupamiento en 6 clústeres de diciembre/2009.

Ahora se muestran las gráficas de cada uno de los 6 clústeres formados por la aplicación en orden ascendente en la Grafica 6.6:

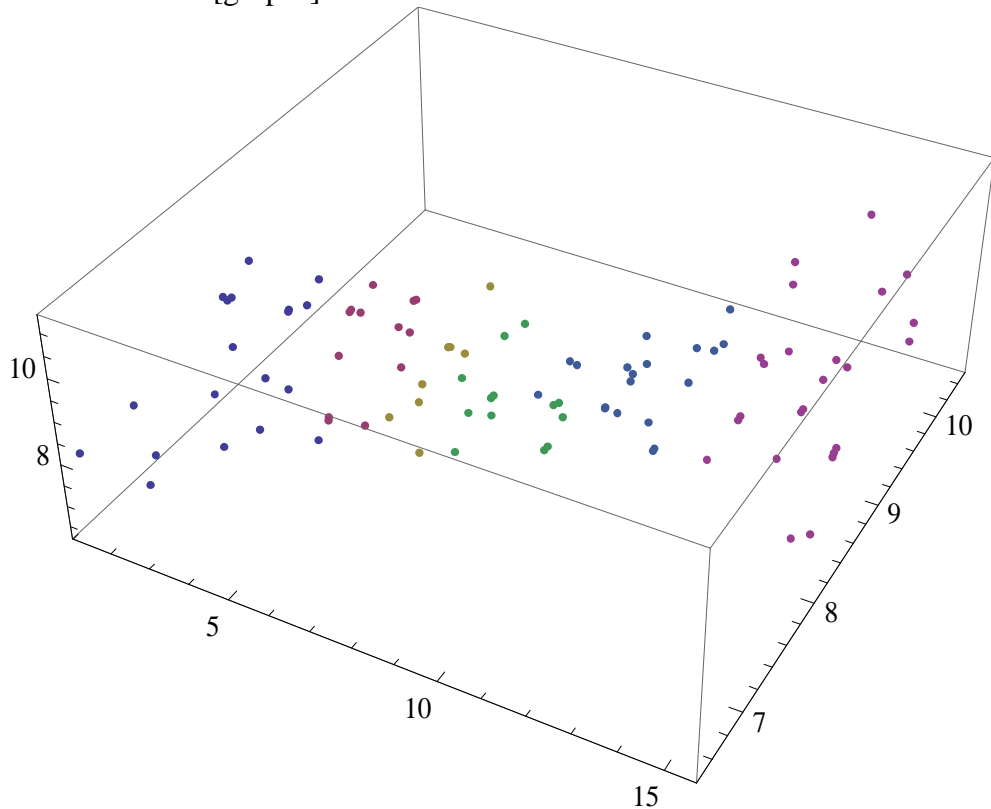
Map[ListPointPlot3D[#]&,grupos]



Gráfica 6.6. Gráficas del agrupamiento en 6 clústeres de diciembre/2009.

Y finalmente los vectores de los 6 clústeres se muestran en la Grafica 6.7, en una sola Gráfica con los diferentes vectores coloreados en función al clúster de pertenencia:

ListPointPlot3D[grupos]



Gráfica 6.7. Gráfica del agrupamiento en 6 clústeres de diciembre/2009.

## 10. Análisis de la elección de Clústeres

Al analizar los resultados del punto 9 de este capítulo, se puede concluir lo siguiente:

- a) La definición de agrupamiento en 6 clústeres representa de la mejor forma posible patrones de comportamiento, tomando en cuenta el ensayo realizado con 12 grupos.
- b) En este agrupamiento vectores con estadios históricos similares quedan agrupados en el mismo clúster, aunque no quedan con el mismo número de miembros, lo cual no representa ningún problema, sino solo una observación.
- c) El clúster 3 según se visualiza en los vectores que lo conforman contiene una variedad más amplia de datos de diferentes estadios históricos (meses), lo cual puede representar un clúster atípico por la integración de sus elementos.
- d) Algunos agrupamientos en más o menos clústeres, no presentan visualmente un mejor agrupamiento en estadios históricos.

Estas son las razones diferenciadas de los otros agrupamientos que nos hacen tomar la decisión de 6 clústeres.

Esta decisión está tomada en función a los datos seleccionados, y no quiere decir que para todos los conceptos de datos sea esta la mejor selección de número de clústeres, sino que en cada caso se debe de realizar un análisis similar y valorar los resultados en función al reconocimiento de ciertos patrones.

## VII. DEFINICIÓN Y CONSTRUCCIÓN DEL MODELO

El modelo es una entidad que permite conceptualizar el evento que analiza y abstraer los elementos de la realidad en unos cuantos conceptos que son tomados en cuenta para la aplicación del mismo. En la construcción de este modelo se aplica la investigación previa al presente capítulo y que define las pautas principales bajo las cuales se regirá el modelo de prospección.

### 1. Antecedentes similares y resultados

#### *“An Evolutionary Modularized Data Mining Mechanism for Financial Distress Forecasts”*

En este estudio, los autores Po-Chang, Ping-Chen Lin, comentan acerca de la importancia que tiene un pronóstico preciso de las finanzas de una corporación en el fortalecimiento de los principios de juicio de un tomador de decisiones. Ellos definen un mecanismo basado en un algoritmo evolutivo para extraer patrones de comportamiento fundamentales de la información financiera.

Muchos profesionales como directivos de bancos, de crédito, accionistas, analistas financieros, oficiales de gobierno, y público en general, toman sus decisiones con las herramientas de información con la que disponen. Un reto importante se plantea de extraer o descubrir información valiosa escondida desde varias homogéneas y heterogéneas bases de datos, de muchas fuentes diferentes, para dar más elementos al decisor.

Su propuesta se basa en establecer varios módulos de procesamiento de información, con funcionalidades como Regresión lógica, Análisis discriminante, y redes neurales para obtener el mejor pronóstico posible. Para

tratar de eliminar resultados irracionales entre los módulos, fueron diseñados criterios de evaluación basados en reglas que todos los módulos deben cumplir.

Los resultados obtenidos son importantes y según sus cifras tienen desviaciones dentro de rangos de aceptabilidad, aunque el poder de cómputo ocupado para el procesamiento de redes neurales resulta ser muy costoso e indispensable para obtener resultados [Po-Chang, 2007].

Observaciones para el presente estudio:

- a) Se otorga una especial importancia a la cantidad de estudios que se han realizado al pronóstico, debido a las repercusiones que realizarlo de manera precisa puede tener en el ámbito financiero y económico.
- b) Además se abre el panorama a la cantidad de herramientas de minería de datos que se pueden aplicar en la explotación de datos de estudio.

#### *“DATA MINING FOR FINANCIAL APPLICATIONS”*

El autor del artículo, Kovalerchuk analiza que el campo de las finanzas a nivel de mercado de valores no es un campo de estudio que pueda dar resultados exactos. En estudios como el presente se presentan aplicaciones con ciertas metodologías y comentando otras que en otros estudios aplican con más o menos éxito.

La evaluación del éxito de las metodologías o algoritmos utilizados a determinada aplicación o problemática no es sencilla de calificar.

Este trabajo describe la aplicación de minería de datos al área de las finanzas, especialmente de los mercados financieros, y se discuten principalmente tareas financieras, técnicas y metodologías. Las Instituciones financieras producen grandes cantidades de hojas de datos que representan problemas enormemente complejos y dinámicos. Años de investigación se han

aplicación a la búsqueda de probables beneficios de extraer conocimiento de estos datos, y se ha utilizado diferentes metodologías en su búsqueda como las siguientes:

- Simulación de Montecarlo
- Análisis de las diferencias finitas
- Transformada de Fourier para análisis de precios
- Minería de datos como un proceso para descubrir patrones frecuentes, correlaciones.
- Modelos de programación lineal y no lineal
- Redes neurales multi-capas
- Aproximación con clúster jerárquico
- Análisis de Vecinos cercanos
- Análisis de árboles de decisión, regresiones, aprendizaje bayesiano, y más...
- Publicaciones han estimado métodos de minería de datos con arquitecturas híbridas de redes neurales, algoritmos genéticos, y teoría de caos.

Se requiere en el área financiera que la minería de datos satisfaga:

- Pronósticos con un alto nivel de ruido en los datos de origen.
- Criterios de eficiencia (maximización de recuperación) en adición a pronósticos.
- Pronóstico multi resolución coordinado en años, meses, días, horas y minutos.
- La necesidad de un modelo que explique el pronóstico y no lo maneje como una caja negra.
- Hacer posible encontrar patrones con tiempos de vida muy cortos.

- Incorporar el impacto de los jugadores del mercado en la normalidad del mercado.
- La minería de datos en el área financiera tiene el mismo reto que la minería de datos en general, que es el cuidado en la selección de los datos que permitan construir el modelo.
- En finanzas esta cuestión está altamente conectada con la selección de las variables objetivo.
- La calidad del modelo de pronóstico basado en minería de datos aplicados a finanzas está definida por la desviación estándar que exista entre el pronóstico y los valores actuales en las pruebas de datos.
- Se busca identificar patrones de comportamiento.

Se establecen una serie de pasos para el modelado de datos con la metodología de minería de datos aplicada al sector financiero:

- Comprensión del problema.
- Recolección de datos.
- Refinamiento de los mismos.
- Construcción del modelo.
- Evaluación del modelo.
- Desarrollo adicionando reglas basadas en expertos cuando se trate de ausencia o insuficiencia de datos.

Se establece una metodología de minería de datos llamada relacional, definida por medio de tres puntos principales:

1. Está basada en los siguientes atributos:
  - i. Cada objeto  $x$  cuenta con un vector de valores  $A1(x), A2(x), \dots, An(x)$
  - ii. Cada objeto cuenta con un tipo de dato relacional, en donde se representa su relación con otros objetos.  $X > Y$ .



2. Utiliza la programación lógica inductiva:
  - i. muestras positivas + muestras negativas + fondo de conocimientos = hipótesis.
3. Busca comprender los datos a través de series de tiempo en donde gráficamente analiza ciclicidad y patrones de comportamiento.

En la práctica de su estudio ellos manejan la información de un portafolio financiero de S&P500 (stándar and pourst) que es una sección de la bolsa de Wall Street y su análisis de basa en la aplicación de un modelo diseñado bajo algoritmos de redes neurales y que se describe a continuación:

1. Recolectar 30-40 factores fundamentales y técnicos de las empresas en que comprenda de 10 a 20 años en cuanto a sus valores de mercado llamando a este universo S1.
2. Construir una red neural para predecir su valor de mercado S1.
3. Repetir el paso 1 y 2 para cada empresa monitoreada. Por ejemplo si 500 empresas son monitoreadas, 500 redes son generadas.
4. Predecir el valor  $i+1$  en tantos días hacia adelante  $S_i(t+k)$  por calcular
  - i.  $NN_i(S_i(t)) = S(t+k)$  donde  $k$  es  $n$  días o semanas adelante.
5. Seleccionar el valor  $NN_i$  con el  $S(t+k)$  más alto.
6. Reajustar los portafolios calculados  $NN_i$  para cada una de las redes restantes.

Algunas de las aplicaciones más importantes de este tipo de estudios esta en Identificación de comportamientos sospechosos en empresas analizadas y de transacciones electrónicas inusuales en S&P 500. Para que la minería de datos resulte adecuada en este campo se requiere de que las

aplicaciones generadas y sus resultados sean rápidamente testeados (monitoreo constante).

Procesos que se deben correr diariamente durante meses para estimar su calidad. El poder informático para el procesamiento de la cantidad de redes neurales propuestas por el estudio vuelve a ser importante, sobre todo si se evalúa que es necesaria una velocidad de respuesta muy breve para actividades de monitoreo de operaciones [Kovalerchuk, 2006].

Observaciones para el presente estudio:

- a) El campo financiero representa un reto para la aplicación de la minería de datos por la necesidad de obtener resultados certeros en tiempos breves, y esto requiere de un poder de cómputo importante.
- b) Se establece la base de un conjunto de pasos para modelar datos financieros al pensar en utilizar minería de datos para extraer información oculta.
- c) Se establece la importancia de reconocer patrones de comportamiento como la base de un pronóstico certero.

*“Using genetic algorithms to optimize nearest neighbors for data mining”*

En el estudio, Hyunchul analiza la utilización de sistemas CBR como una promesa en la implementación de soluciones efectivas en la toma de decisiones complejas y desestructuradas, comprendiendo por desestructuradas, a aquellas decisiones que no pueden ser tomadas con juicios de hechos totalmente conocidos.

En el diseño de un CBR es importante la determinación de los k números de vecinos cercanos (k-NN) trabajando con metodologías de agrupación de datos o vectores de datos en clústeres. De esta forma el estudio

propone el uso de algoritmos genéticos para optimizar el número de vecinos cercanos que deben ser tomados en cuenta para el cálculo interno en un sistema CBR.

Pero cuál es la importancia de la definición óptima de vecinos cercanos tratada como k-NN. Analizando el número de casos para combinar y generar la solución, se debe comprender que si:

- Si k es muy pequeño, es difícil encontrar un patrón general de comportamiento y la exactitud del resultado se reduce.
- Si k es muy grande, se pueden incluir casos irrelevantes o que distorsionan la muestra y reducen la exactitud del resultado también.

El modelo mostrado en la Figura 7.1, que se aplica en el artículo para optimizar el número de k-NN es el siguiente:

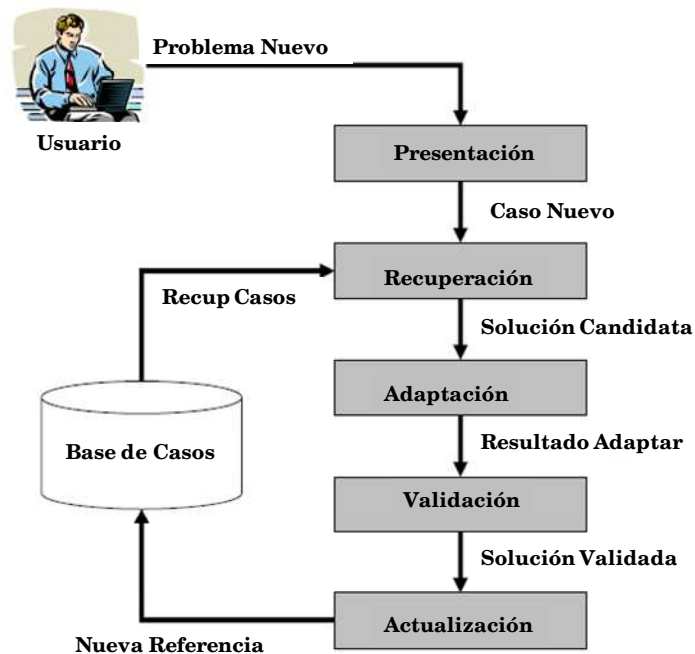


Figura 7.1. Modelo de optimización k-NN.

Para la aplicación del modelo se define una metodología para realizar el cálculo de la similaridad de los elementos cercanos a un dato definido, y la definición sigue la metodología siguiente:

Si los casos son representados por vectores

$$V_i = (C_1, C_2, C_3, C_4, C_5, \dots, C_n)$$

y se calcula el peso con la suma de la distancia (euclidiana ej.) de sus características utilizando la siguiente expresión:

$$\frac{\sum_{i=1}^n W_i \cdot \text{sim}(f_i^I, f_i^R)}{\sum_{i=1}^n W_i}$$

Donde:

- $W_i$  – Pesos de la característica  $i$ .
- $f_i^I$  – El  $i$ ésimo valor característico de ese caso.
- $f_i^R$  – El  $i$ ésimo valor de la base de casos.
- $\text{Sim}()$  – Función de similaridad.

Finalmente se utiliza un algoritmo genético que ocupa una búsqueda estocástica con la posibilidad de escudriñar en grandes y complicados espacios de datos. Este proceso es parecido a las teorías de genética biológica, y los principios de la teoría de la evolución, donde el resultado de la búsqueda sugiere valores a ocupar como posibles valores óptimos. Además se ocupa el término mutación, cruza y selección para que con base a valores conocidos se obtengan valores inesperados, pero posibles que obtengan mejores resultados al ser incluidos en el proceso de búsqueda de k-NN.

El estudio se ocupó prácticamente en la predicción de compra de una tienda online en Corea, y se analiza que personas comprarían pronto artículos

de entrenamiento, de los que la tienda comercializa. Los resultados se consideraron con validez dentro de los criterios de validación establecidos.

Una importante limitación de la que se habla es la potencia de cómputo requerido. Para instancias de cómputo limitado 1NN es más eficiente que GA-kNN. El uso de GA (algoritmos genéticos) puede ser optimizado si se hace una correcta selección de características y pesos de los datos motivo del modelado [Hyunchul, 2008].

Observaciones para el presente estudio:

- a) Se enfoca todo el esfuerzo en obtener una optimización, que es interesante de obtener para los datos de estudio en particular, y que puede variar dependiendo de los datos que en cada caso se estudien.
- b) Se obtienen criterios importantes en cuanto como afecta a un pronóstico el número de vecinos cercanos por los márgenes de errores que con cada uno de ellos puede resultar.
- c) Se hace un énfasis especial en determinar que si la potencia de cómputo no es alcanzable para el estudio, se deben tomar medidas de simplicidad para resolver el problema satisfactoriamente.

## **2. Definición de modelo**

El modelo es obtenido en función al análisis de los datos objeto de la investigación, y a diferentes pruebas realizadas con varios conceptos de los datos de la empresa objeto de estudio y se muestra esquematizado en la Figura 7.2. Además se obtiene como resultado de la búsqueda de poder definir el mejor modelo de prospección.

## Diagrama General del modelo:

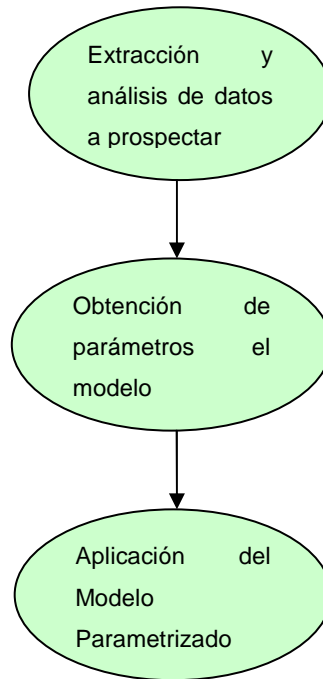


Figura 7.2. Modelo del sistema de apoyo al razonamiento prospectivo.

## Modelo detallado en Pseudocódigo:

- *Se realiza la extracción de datos de la empresa.*
- *Extracción historia de casos relacionados con el desarrollo de la empresa.*
- *Análisis gráfico de datos originales.*
- *Análisis de patrones de comportamiento.*
- *Análisis selectivo de los datos que serán elegidos para la prospección*
  - *Vector de datos originales*
- *Elección de la mejor forma de estandarizar los datos originales.*
- *Análisis de Gráficas de comportamiento de datos estandarizados.*
- *Procesamiento de los datos estándar elegidos mediante métodos de agrupación (Clústeres).*

- *Elección del mejor número de clústeres por métodos gráficos.*
- *Elección del mejor número de vecinos cercanos a cada vector a prospectar por comparación de error en resultados conocidos.*
- *Aplicación del algoritmo para cada caso a prospectar:*
  - *Determinación del dato actual a prospectar.*
  - *Recuperación de los datos históricos inmediatos anteriores al dato a prospectar.*
  - *Procesamiento de los datos en 6 clústeres (caso de estudio).*
  - *Análisis gráfico de los clústeres*
  - *Análisis de los vectores vecinos cercanos al dato estándar a prospectar (3 vecinos cercanos para mejores resultados) (análisis presentado en el Siguiete Capitulo).*
  - *Determinación de los vectores del estado  $N+1$  (3 siguientes).*
  - *Determinación de la media de estos tres vectores  $N+1$ .*
  - *Determinación de valor prospectado  $N+1$ .*
  - *Aplicación del proceso de des estandarización del vector estándar para obtener el vector real prospectado.*
  - *Obtención de los datos de casos relacionados con los estados:*
    - *$N-1$  como antecedentes.*
    - *$N$  como antecedente inmediato a la prospección.*
    - *$N+1$  como histórico de apoyo a la prospección*
  - *Presentación de resultados.*
- *Fin del modelo*

El código resultante programable del proceso anterior corresponde con el pseudocódigo anterior, codificado en la aplicación Wolfram Mathematica 7.0, se encuentra en el apéndice D.

### 3. Construcción con aplicaciones seleccionadas

Para la construcción se confirma la selección previa definida al final del capítulo V y consiste en lo siguiente:

- a) Los datos se concentraron finalmente en la hoja de datos de Excel para realizar desde ahí análisis gráfico y estandarización de la información.
- b) Una vez estandarizados los datos se migran a un archivo plano (en texto con un formato definido) para importarlos hacia la aplicación Wolfram Mathematica 7.0 para ejecutar el proceso agrupación en clústeres, aplicando herramientas de minería de datos.
- c) Se buscan a través de minería de datos los clústeres, vecinos cercanos en el clúster, vectores antecedentes y vectores consecuentes
- d) La relación de los casos históricos cargada en Excel se ocupa solo para relacionarla con los casos históricos de la empresa, y con los casos de los casos más cercanos al prospectado. Esta relación contiene información de casos clasificados en categorías como se muestra en la Tabla 7.3:

Tipos	Descripción
EIF	EVENTOS INTERNOS (FINANCIEROS)
EIO	EVENTOS INTERNOS (ORGANIZACIONALES)
EIX	EVENTOS INTERNOS (ESPECIALES)
ENR	EVENTOS NACIONALES RELACIONADOS
ETR	EVENTOS INTERNACIONALES RELACIONADOS

Tabla 7.3. Clasificación de eventos.



## VIII. APLICACIÓN DEL MODELO

Los parámetros que se han de definir y las bases sobre las que se definen son tan importantes o más que el proceso que las define y que se ocupa para realizar el modelo de prospección que nos ocupa el capítulo anterior.

### 1. Consideraciones de la aplicación

Adicionalmente a la definición de parámetros, es obligado comentar que el poder computacional para aplicar el modelo, está considerado para ser ejecutado en un equipo típico de una pequeña empresa, por lo que además el modelo necesitaba tener respuestas en tiempos considerados accesibles. Comprendiendo esta necesidad de utilizar una computadora con las siguientes características generales de hardware y software:

- a) Procesador de 2 núcleos intel Pentium T4200 de 2.0 Ghz. Y 1 MB de cache L2
- b) Memoria de 2Gbytes DDR de 800Mhz.
- c) Procesador Grafico Intel GMA 4500M a 732MB
- d) Disco duro de 250GBytes a 7200 Rpm
- e) Sistema operativo Windows 7 SE

Existen una serie de parámetros que se han definido gradualmente en el modelo en función a diferentes análisis dentro de los cuales se encuentra uno que ya se concluyó referente al número de clústeres para los vectores del caso de estudio que nos ocupa, definido en 6 para este caso.

Un segundo parámetro importante a definir es el número de vecinos cercanos al vector que se prospecta, y para poder llegar a esta definición se realizaron las siguientes pruebas con el modelo programado en la aplicación:

- a) Se realizó el análisis tomando como base una muestra de 4 años (2006 al 2009 inclusive) 48 lecturas tomadas de manera aleatoria considerando 8 muestras de meses (muestra del 16.6%).
- b) Se realizaron corridas de modelo analizando en cada corrida un diferente número de vectores cercanos (variando entre 2 y 5) y analizando estadísticamente cuál de ellos obtenía el menor error con resultados históricos conocidos.
- c) Los resultados y tablas comparativas se encuentran en el apéndice E, concluyendo que el menor error se obtiene definiendo el vector prospectado con el promedio de 3 vecinos cercanos.

A continuación se muestran los resultados finales la Tabla 8.1:

<b>Análisis de Promedios de Errores</b>		
<b>Ventas</b>	<b>Almacén</b>	<b>Promedio</b>
<b>CON 2 CERCANOS</b>		
-0.68%	0.01%	0.33%
<b>CON 3 CERCANOS</b>		
0.02%	0.61%	0.30%
<b>CON 4 CERCANOS</b>		
-0.18%	-1.29%	-0.74%
<b>CON 5 CERCANOS</b>		
-0.34%	-1.89%	-1.12%

Tabla 8.1. Análisis sumario de errores porcentualizados.

Además se analizaron los errores que se arrojaban al comparar los resultados obtenidos con los tres vecinos cercanos procesando el análisis de errores de manera individual. Este resultado está en el apéndice E tabla E.3. y nos confirma la elección de 3 vectores vecinos cercanos para prospectar con menos errores.

## **2. Caso 1.- Aplicación del modelo con datos históricos**

Teniendo en cuenta que se cuenta con datos desde el año 2002 al 2009 inclusive, se tomaron para realizar pruebas de validación de los resultados del modelo los 6 últimos meses del año 2009.

Esta validación se realiza desde dos puntos de vista:

- a) Desde el punto de vista cuantitativo al ponderar los márgenes de error que se presentan al aplicar el modelo con datos que comprenden desde el inicio del rango, hasta un mes antes del mes que se pretende prospectar. De esta forma se prospecta el vector, y se puede comparar los valores obtenidos por el modelo contra los valores reales de la empresa objeto de estudio para el mes en cuestión.
- b) Desde el punto de vista cualitativo es importante notar que al darse un vector prospectado, el modelo da al tomador de decisiones de los 3 vecinos más cercanos, sus eventos históricos registrados, y adicionalmente los 3 históricos que les antecedieron. Son 6 estadios históricos que proporcionan al tomador de decisiones elementos de razonabilidad sobre el porqué de la cifra cuantitativa propuesta. Adicionalmente se exponen también el caso histórico de los tres eventos para la cifra prospectada como un elemento adicional de apoyo al razonamiento. Se dan los elementos de hechos para el hombre administrativo.

De esta forma se aplica el modelo cada vez con más vectores de datos históricos, incrementalmente, prospectando un mes adelante.

## **3. Caso 2.- Aplicación del modelo con datos incrementales**

Considerando el mismo marco de datos desde el año 2002 al 2009 inclusive, se realiza esta prueba con la finalidad de prospectar incrementalmente, es decir iniciando con proponer el mes de Julio y prospectando Agosto no sobre datos históricos, sino sobre los obtenidos por el

modelo. Este proceso se siguió para los meses de Septiembre, Octubre, Noviembre y Diciembre del año 2009.

Esta validación se realiza con la finalidad de comprender hasta donde y con qué margen de error se puede realizar una proyección partiendo desde un dato histórico y hacia N periodos en el futuro.

#### **4. Caso 3.-Aplicación del modelo de forma prospectiva**

Considerando ahora los datos desde el 2002 hasta el 2009 inclusive, se inicia un proceso de proyección mes con mes desde enero del 2010 hasta marzo del mismo año, con la finalidad de interactuar directamente con los directivos de la empresa objeto de estudio, antes de que las cifras reales se obtengan por los procesos administrativos propios, proyectando valores y proponiendo casos de apoyo al razonamiento del hombre administrativo, y siendo evaluados de dos formas diferentes:

- a) Cuando los datos reales no son conocidos y para el decisor representa una escenario posible de comportamiento y la rastreabilidad del porque se obtiene el resultado. Es una evaluación cualitativa.
- b) Cuando se obtienen los resultados de la empresa y se comparan de forma cuantitativa con los datos proyectados obteniendo los márgenes de error y analizando desviaciones.
- c) La aplicación del modelo en esta fase se realizo mes con mes anticipándose a la obtención de datos reales, desde Enero hasta Marzo del 2010.

Estas son las formas en que se aplicaran los modelos y en que se validaran los resultados obtenidos.

## IX. ANÁLISIS DE RESULTADOS

Se analizaron los resultados desde una perspectiva que llamamos interna, es decir tomado en cuenta criterios matemáticos y de formación del modelo para evaluar los resultados. Una segunda perspectiva se sitúa desde el punto de vista del tomador de decisiones, es decir tomando en cuenta los comentarios de la dirección de la empresa objeto de estudio. Un tercer análisis mezcla los dos puntos de vista analizando los resultados a la luz de ofrecer los datos prospectados antes de que los resultados reales sean emitidos por la empresa, y con el conocimiento previo, comprender el proceso de apoyo al razonamiento basado en la prospectiva y casos históricos registrados.

### 1. Caso 1.- Análisis de la aplicación del modelo con datos históricos

En el apéndice F, se muestran los resultados obtenidos de aplicar el modelo con datos históricos, aplicando el modelo para cada uno de los meses a partir de Julio y hasta Diciembre de 2010, comprendiendo que para el momento del análisis los resultados obtenidos se pueden comparar contra los datos reales.

Para cada una de las aplicaciones del modelo, se define un ámbito en el cual se desempeña el modelo. Por ejemplo para la primera aplicación, haciendo la prospectiva del mes de Julio/2009, se presenta el ámbito:

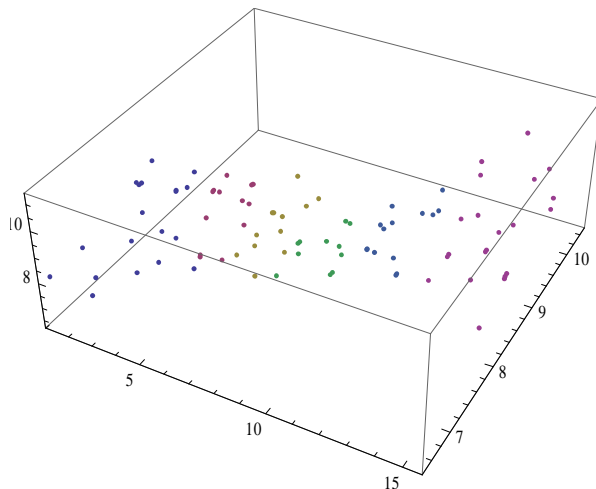
- a) *Datos de 2002/01-2009/06.*
- b) *Clústeres=6.*
- c) *Vecinos cercanos a considerar 3.*
- d) *Pronóstico: 2009/07.*

Que implica que se tomaran los datos de vectores tridimensionales, de los meses de Enero del 2002 al mes de Junio de 2009, por lo que prospectará el mes de Julio del 2009. Además se considera en el proceso aplicar el modelo para 6 clústeres y considerando 3 vecinos cercanos al mes ultimo base del análisis (Junio) como base de la prospección del mes siguiente. El algoritmo de agrupación utilizado es el de Distancia Euclidiana.

Al aplicar el modelo programado, se obtienen los resultados, dentro de los cuales adicionalmente se sitúa el vector del mes ultimo base de la prospección que en este caso es *vector: {7.6923,8.3602,9.0718}*.

Después se presentan los datos agrupados por clústeres, en este caso 6 clústeres y las gráficas tridimensionales una por cada uno de los clústeres, ordenadas de manera secuencial desde el Clúster 1 al Clúster 6.

A continuación se muestra la Grafica 9.1, que muestra de forma tridimensional de todos los clústeres, donde cada vector tridimensional es un punto, y donde los colores distintos muestran cada uno de los clústeres formados.



Grafica 9.1. Análisis sumario de errores porcentualizados.

Ya como parte del modelo donde se busca la vecindad más cercana dentro del clúster donde se encuentra la lectura base de la prospección (Junio) y se presentan los tres más cercanos en la Figura 9.1.

*Datos en la vecindad más cercana dentro del clúster a la posición original {{90}} :*

*{{7.6923,8.3602,9.0718},{7.6923,8.6175,8.8463},{7.6923,7.8463,9.1419},  
{7.6923,7.9101,8.1832}}*

Figura 9.1. Vecindad más cercana a la posición 90.

Se presentan como resultado 4 lecturas, porque se incluye la propia lectura por ser la de distancia euclidiana cero, pero esta se discrimina finalmente para tomar solo las tres más cercanas.

Se buscan en el vector maestro (que incluye todos los vectores históricos), la posición de los vectores cercanos a la posición base de prospección, y un estadio histórico anterior, para hacer la búsqueda de casos históricos apoyo al razonamiento.

*Posiciones en la vecindad actual cercana del vector:*

*{30,54,66}*

*Posiciones del estado N-1 o anterior:*

*{29,53,65}*

Figura 9.2. Posiciones de vecindad a los estados N y N+1.

Además se obtienen los estadios siguientes, que son realmente los datos de prospección histórica, y se localizan los vectores correspondientes a cada posición N+1.

*Posiciones del estado N+1 o siguiente:*  
 {31,55,67}  
*Datos del estado N+1 o siguiente:*  
 {{8.9744,8.4397,7.356},{8.9744,8.1212,8.582},{8.9744,8.3867,7.8153}}

Figura 9.3. Posiciones más cercanas en los estados siguientes.

Como resultado final del proceso prospectivo, se saca el promedio de los valores del vector, dando como resultado:

*Resultado Prospectivo con cercanos igual a 3*  
 {8.9744,8.31587,7.91777}

Figura 9.4. Resultado prospectivo con 3 cercanos.

Con este vector que esta estandarizado, se trabaja en des-estandarizarlo, es decir aplicar el proceso inverso, aplicándolo a el acumulado anual de ventas a clientes (segundo término del vector) y al acumulado anual de valor del almacén (tercer término del vector), y obtener desviaciones con respecto a las lecturas originales conocidas para el mes de Julio/2009.

<b>PRONOSTICO</b>		
<b>A: JUL 2009</b>		
	<b>VENTAS</b>	<b>INVENTARIO</b>
Datos Estándar	8.31587	7.91777
Cifra Pronosticada	29,342,744	18,521,099
Diferencia Real	879,889	-543,106
% Diferencias	3.09%	-2.85%
Promedio % Diferencias		0.12%

Tabla 9.2. Análisis sumario de errores porcentualizados.

Se obtienen la cifra pronosticada, en el tercer renglón la diferencia contra la lectura real del mes de Julio (positiva o negativa), y en el cuarto renglón la misma diferencia expresada en forma porcentual.



Finalmente se obtiene un error promedio de las dos lecturas anteriores, que en este caso es de 0.12% en promedio contra las cifras reales de la empresa objeto de estudio.

El último resultado presentado son los casos históricos relacionados con los estadios previos, y con los estadios N+1 prospectados, como apoyo de hechos a la toma de decisiones.

a) Estado N y N-1 de vecinos cercanos

Posición	Año	Mes	Tipo	Memoria de Casos
29	2004	5	EXR	Unión Europea: se integran Chipre, la República Checa, Eslovaquia, Eslovenia, Estonia, Hungría y tres mas
65	2007	5	EIO	Integración de Sistema Administrativo Kepler Acoplado

Tabla 9.3. Eventos de los vecinos cercanos en los estados N y N-1.

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
67	2007	7	EIO	Integración de Sistema de producción al sistema kepler

Tabla 9.4. Eventos de los vecinos cercanos en el estado N+1.

Y su función es dar razonabilidad a la propuesta prospectada del resultado del mes de Julio 2009.

## 2. Criterios de validación del modelo

Para comprender si este error es aceptable, se tiene que proceder a analizar los datos totales históricos desde el punto de vista estadístico, por medio de la desviación estándar, la media y apoyos mediante graficas.

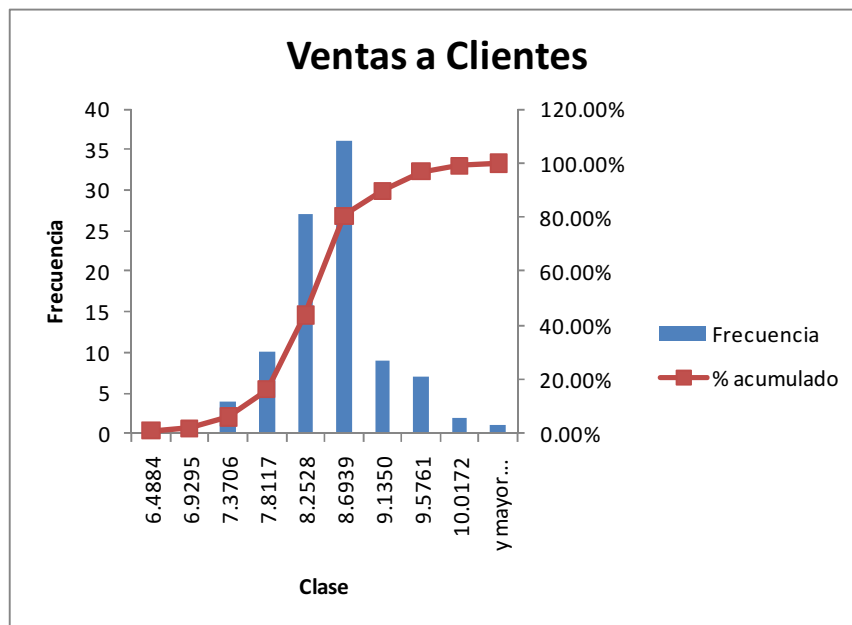
Los datos de la Tabla 9.5. muestran el siguiente comportamiento estadístico:

Concepto	Ventas	Almacen	Error min/max	Obtenido Min/max	Porcentaje
Promedio	8.3084	8.3083			
Desv Std	0.6404	0.7613			
Min	6.4884	6.3015	-20%	-1.20%	6%
Max	10.4584	11.3127	37%	7.60%	21%

Tabla 9.5. Análisis estadístico de los datos históricos.

La desviación estándar de los datos muestra una concentración muy fuerte hacia los valores promedio de cada concepto, por lo que su campana está muy concentrada en ese punto.

Al obtener los histogramas de cada uno de los conceptos se confirma lo comprendido de la Grafica 9.6.



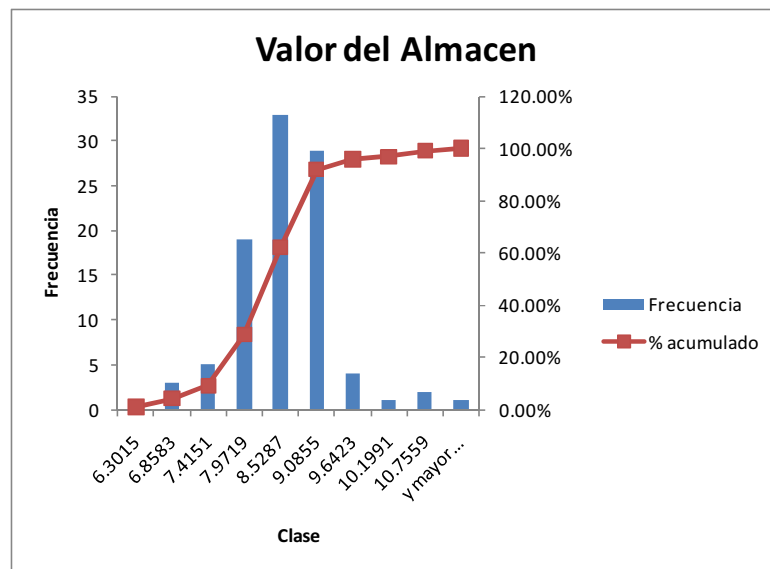
Gráfica 9.6. Histograma de ventas a clientes.

Un análisis de la información de la Gráfica 9.6. considerando las clases que generan las barras, se obtiene lo mostrado en la Tabla 9.7.

Clase	Frecuencia	% acumulado
6.4884	1	1.02%
6.9295	1	2.04%
7.3706	4	6.12%
7.8117	10	16.33%
8.2528	27	43.88%
8.6939	36	80.61%
9.1350	9	89.80%
9.5761	7	96.94%
10.0172	2	98.98%
y mayor...	1	100.00%

Tabla 9.7. Descripción de clases del Histograma de ventas a clientes.

En un análisis acumulado de frecuencias se obtiene la Gráfica 9.8. ahora con referencia a el valor del almacén de la empresa, de la misma forma que se obtuvo la de ventas a clientes.



Gráfica 9.8. Histograma de valor del almacén.

Un análisis de la información de la Gráfica 9.8. considerando las clases que generan las barras, se obtiene lo mostrado en la Tabla 9.9.

Clase	Frecuencia	% acumulado
6.3015	1	1.02%
6.8583	3	4.08%
7.4151	5	9.18%
7.9719	19	28.57%
8.5287	33	62.24%
9.0855	29	91.84%
9.6423	4	95.92%
10.1991	1	96.94%
10.7559	2	98.98%
y mayor...	1	100.00%

Tabla 9.9. Descripción de clases del Histograma de valor del almacén.

Al tomar los valores mínimos y máximos de cada concepto se analiza en promedio hasta donde se puede tener el error porcentual mínimo y máximo y compararlo contra los datos mínimos y máximos obtenidos en el proceso de prospección de los meses de Julio a Diciembre de 2009. Estos datos se encuentran en la columna quinta de “Obtenido Min/Max”.

La columna sexta de “Porcentaje” representa el valor porcentual de los mínimos y máximos errores registrados, comparados entre ellos, y se obtienen valores de errores que fluctúan entre el 6% con respecto a los mínimos (negativos) y el 21% con respecto a los máximos (positivos).

Particularmente se analizó no solo las mínimas y máximas diferencias de los resultados de la prospección, sino los promedios de esas diferencias, que son más representativas de lo esperado para la aplicación del modelo a cada caso. Y se obtuvo una diferencia esperada que realmente se considera muy aceptable para el modelo. Los resultados de esta se muestran en la Tabla 9.10. a continuación:

ERRORES PROMEDIO		
Error min/max	Obtenido Min/max	Porcentaje
-20%	-0.73%	4%
37%	4.76%	13%

Tabla 9.10. Análisis estadístico de los datos históricos.

Se debe aclarar que este análisis se realizó sobre la diferencia promedio y además sobre la diferencia de cada uno de los conceptos de manera independiente, para los meses de julio a Diciembre, obteniendo resultados similares, por lo que se considera representativo trabajar con el promedio de las diferencias para facilitar la comprensión de la validación del modelo.

En el caso específico de la proyección del mes de julio el error promedio es de 0.12% lo cual representa con respecto al error máximo positivo posible, un 0.32% lo cual se considera una cifra muy aceptable de acercamiento con los datos reales.

De esta misma forma se encuentra el análisis de resultados de los meses de Agosto a Diciembre del 2009 en el apéndice F, teniendo algunas observaciones importantes al respecto:

- a) En el mes de julio la diferencia fue de 0.12% aunque las cifras de diferencias de cada concepto se compensaron, y cada una de ellas fue de 3.09% para ventas, y -2.85% para almacén. Consideradas aceptables bajo los lineamientos de validación.
- b) Para el mes de agosto las cifras se mantuvieron en un margen parecido los dos conceptos, y el promedio reflejó ese comportamiento con un 3.81% de diferencias positivas.
- c) En el mes de septiembre la cifra de diferencias resultó sobre -0.26%, muy buena con resultados positivos y negativos de los conceptos por separado. Especialmente se hace notar la necesidad de contar con más datos de casos históricos como apoyo a la razonabilidad de la toma de decisión, porque en este mes en particular no se tuvieron datos relacionados.

- d) El caso de octubre adoleció de igual forma de casos históricos, lo cual se explica porque perteneció al mismo clúster del caso de septiembre. Las diferencias fueron compensatorias, y su promedio fue de -1.20%.
- e) El mes de noviembre presentó diferencias mayores a los meses pasados, con un promedio de 7.51% y por concepto con diferencias de 1.66% para ventas y elevada la cifra de almacén en un 13.36% dentro de los límites de validación. Se volvió a hacer patente la falta de casos históricos, y se hicieron más necesarios por la necesidad de comprender estas diferencias en el caso del valor de almacén.
- f) Diciembre presentó un caso similar a noviembre en cuanto a cifras y falta de casos históricos. Con promedio de diferencias de 7.61% y diferencias en ventas por 1.13% y almacén de 14.09% dentro de límites de validación.

### 3. Caso 2.- Análisis de la aplicación del modelo con datos incrementales

Los resultados de este análisis se presentan en el apéndice F en la sección correspondiente, simplificando la presentación de resultados por comprender que el proceso es muy parecido al de el Caso 1, pero con la diferencia de que los resultados de cada mes se acumulan y se toman en vez de los resultados reales como base del siguiente proceso de prospección.

En la Tabla F.1. se detallan los resultados obtenidos mes con mes de Julio a Diciembre, de forma similar a la ya presentada.

En la siguiente Tabla 9.11. se presenta el resumen de los resultados obtenidos:

Comparación de Casos			
Mes	Caso 1	Caso 2	Diferencias
Julio	0.12%	0.12%	0.00%
Agosto	3.81%	3.56%	-0.25%
Septiembre	-0.26%	4.70%	4.96%
Octubre	-1.20%	5.13%	6.33%
Noviembre	7.51%	12.79%	5.28%
Diciembre	7.61%	13.75%	6.14%

Tabla 9.11. Análisis sumario del proceso de la prospección incremental.

En la Tabla 9.11 se muestra los resultados obtenidos con los datos históricos de forma natural, y los resultados obtenidos con los datos incrementales (Caso 1 y Caso 2). En el Caso los resultados tenían las diferencias analizadas dentro de las cuales estando en límites algunas de ellas, como las de noviembre y diciembre dieron diferencias mayores. Para el Caso 2 se observa cómo se despega gradualmente las diferencias contra los datos reales de la empresa objeto de estudio, llegando en el mes de Diciembre con un porcentaje de diferencia de 13.75% contra el real, y contra el 7.6% obtenido del Caso 1.

En términos reales se puede considerar que en este caso hasta segundo o tercer mes se podría obtener un resultado prospectivo que pueda ser considerado para la toma de decisiones asumiendo las diferencias y los riesgos que conllevan las decisiones tomadas bajo estas condiciones.

#### **4. Caso 3.-Análisis de la aplicación del modelo de forma prospectiva**

Este caso es muy parecido al Caso 1, trabajando con datos históricos, la única diferencia es que se trabajo en una realidad prospectiva con la empresa. Es decir, se obtuvieron resultados prospectivos y casos históricos relacionados, antes de obtener los reales de la empresa, y se presentaron a la dirección, con la finalidad de que analizaran y evaluaran su utilidad.

Los resultados de este caso se encuentran en el apéndice G. se puede comentar con respecto a los resultados las siguientes observaciones analizadas:

- a) El resultado del mes de enero del 2010, presenta cifras con pocas diferencias con respecto a los reales. En promedio la diferencia es de -0.73% y diferencias compensatorias en cada concepto por -2.68% para ventas y 1.22% para almacén.

- b) En enero se volvieron a hacer patentes la falta de casos históricos de apoyo a la racionalidad de la toma de decisiones.
- c) Cuando se dieron a conocer la prospección de enero, era el día 8 del mes por lo que la cifra se reviso y se tomo en cuenta para ver como se desempeñaban los datos reales. Para los días últimos del mes se revisaron los datos reales y se vio su fuerte acercamiento con los datos prospectados.
- d) Enero fue interesante como el primer mes de real prospección y que cumplió las expectativas con un buen margen de asertividad, según la dirección de la empresa.
- e) La aplicación del modelo para febrero presenta mayores diferencias, promedio de 7.35% y por concepto 4.62% para ventas y 10.08% para almacén, las dos positivas.
- f) Los datos prospectivos se presentaron a la dirección el día 10 de febrero, y se revisaron gradualmente conforme avanzo el mes.
- g) En febrero si se presentaron casos históricos, los cuales fueron de utilidad para comprender los resultados, y los casos fueron los siguientes:

a. Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
1	2002	1	EO	Manejo del Sistema informático ASPEL
1	2002	1	ETR	Argentina devalúa el peso en un 28 por ciento.
48	2005	12	EF	Absorción de una empresa competencia

Tabla 9.12. Eventos de los vecinos cercanos en los estados N y N-1.

b. Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
50	2006	2	EO	Adaptación de sistema informático al 80% al proceso

Tabla 9.13. Eventos de los vecinos cercanos en el estado N+1.



- h) Se consideraron de interés la observación del caso 48, por el hecho de que se han integrado nuevas líneas de productos para mercados especiales que no se habían atacado, y que iniciaron hace aproximadamente 3 meses.
- i) Los resultados cuantitativos se consideraron interesantes en cuanto a la dirección que seguían los dos conceptos de estudio.
- j) Para el mes de marzo la diferencia fue de 1.19% promedio, y por conceptos fue compensatoria de 3.58% para ventas y -1.20% para almacén.
- k) Se presento a la dirección el día 10 de marzo, y se considero por la dirección como una buena prospección.
- l) Se consideraron interesantes los casos históricos antecedentes al estadio que se prospecto, por el mismo caso del mes de febrero, y que ahora la prospección se comienza a convertir en historia
  - a. Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
38	2005	2	EIO	Ampliación de planta de producción
50	2006	2	EIO	Adaptación de sistema informático al 80% al proceso

Tabla 9.14. Eventos de los vecinos cercanos en los estados N y N-1.

- m) Este último caso presento a la dirección la interesante disyuntiva de revisar los planes de ventas para tomar decisiones proactivas.

## 5. Conclusiones del análisis

Aparte de lo interesante y emocionante de la aplicación del Caso 3, se consideran muy valorables los resultados obtenidos por el modelo, en cuanto a la aplicación de la minería de datos, en particular de las herramientas de análisis por clústeres, aplicadas a un proceso de prospección.

El caso 1 de aplicación del modelo fue un experimento interesante, que se siguió por la dirección de la empresa objeto de análisis, con las reservas de saber que se estaba validando el modelo de apoyo al razonamiento prospectivo. Los datos obtenidos fueron muy buenos con las diferencias anotadas en noviembre y diciembre del 2009 y que no fue posible tener casos históricos de apoyo a la razonabilidad.

Los casos históricos presentados en el modelo del caso 1 para los primeros meses fueron interesantes como explicación de las cifras, y de las diferencias menores que presentaba el modelo contra los datos reales.

El caso 2 propone límites y posibilidades de generar estadios prospectivos en estados  $N+M$ , donde  $N$  es el estadio actual, y  $M$  la cantidad de estadios futuros a prospectar. Siendo realistas es posible tomarlos en cuenta comprendiendo los riesgos que implica tomarlos como posibilidades muy reales, y en general el tomador de decisiones que se basa en hechos, como el hombre administrativo enunciado por Herbert Simon, puede tener una herramienta de apoyo a la planeación proactiva de la empresa.

El caso 3 planteo un reto muy importante, al dar a conocer las cifras prospectadas antes de que las reales fueran conocidas, y para la empresa fue interesante durante los dos primeros meses de su aplicación, y dio pautas de poder ser tomada en serio para actuar antes de que el fin del tercer mes llegara con las cifras reales.

Es interesante pensar como el conocer los datos prospectivos obtenidos del modelo pudiera cambiar su asertividad porque al tener el conocimiento de ellas, el tomador de decisiones puede tomar acciones que provoquen mejor desempeño de los indicadores de la empresa, pero al mismo tiempo provoquen mayores diferencias en las cifras prospectadas, como una posibilidad.



## X. CONCLUSIÓN

### 1. Conclusión

El modelo producto de esta tesis se aplicó a una empresa objeto de estudio, de tamaño mediana, con procesos administrativos medianamente maduros, pero en donde las herramientas informáticas no están totalmente integradas a los procesos administrativos. Es más, la gobernanza sobre la planeación de la empresa se encuentra más cimentada en las necesidades propias de crecimiento mercadológico y productivo, que en la informática.

Como consecuencia la TI y los SI's no se encuentran en los niveles administrativos y estratégicos de la empresa, como facilitadores de la información que diera la ventaja competitiva que pueden dar a una empresa de su ramo.

El presupuesto que se debe de invertir para llevar la integración informática a estos niveles intermedios es importante, comparándolo con el presupuesto anual que tiene para su operación. Y existe una brecha importante entre lo que se debería destinar a la integración informática, y lo que cada empresa planea anualmente para el área.

Los datos se obtuvieron como se comentó con un proceso que no fue sencillo y que llevó un tiempo considerable por no existir la integración informática en las capas medias, y este es un ejemplo muy representativo de lo que sucede en la empresa mediana de nuestro país. Se debe hacer conciencia de que la gobernanza informática es necesaria para el desarrollo de la empresa y posiblemente para su sobrevivencia. La competencia es cada vez mayor y

grandes empresas compiten con ventajas competitivas realmente importantes de análisis y prospección de datos.

Para algunas empresas conocer lo que puede hacerse si este proceso de contar con información para el soporte de la toma de decisiones, y más para la inteligencia de negocios, debe hacer entrar a la reflexión de planear el futuro considerando seriamente la gobernancia de la informática.

La ventaja competitiva en el caso del modelo de apoyo al razonamiento prospectivo, es realmente importante si se puede procesar en tiempos de respuesta breves, lo que solo se puede lograr contando con la integración informática en las etapas administrativas y estratégicas del modelo piramidal presentado en la fig. 3.27.

Existen muchos más datos que pueden ser motivos del análisis prospectivo. Dentro de los datos extraídos de la empresa objeto de estudio se pueden realizar combinaciones que presenten una utilidad a la planeación de la empresa y a la toma de decisiones.

Al realizar los primeros análisis, estandarizaciones y agrupamientos, y ensayar con algunos otros datos se reconocen patrones de comportamiento. Uno de los patrones de comportamiento que se analizaron derivó en que algunas cifras de pagos de facturas y formas de pago no se comportaban conforme a su comportamiento normal. Este resultado se presentó a la dirección de la empresa, y esta realizó una investigación del caso, llegando a conclusiones de fraude y tomando las medidas contra los responsables. Este ejemplo debe servir para comprender cómo las herramientas de análisis pueden dar además elementos de apoyo a la detección de potenciales problemas administrativos, si es que la integración informática y la gobernancia o co-gobernancia informática se toma en serio en la planeación estratégica de una empresa.

Considerando que las premisas de valor del tomador de decisiones, sus conocimientos vertidos en las premisas de hechos y los casos históricos de la entidad sobre la que se toman las decisiones debidamente automatizados sean los ingredientes para la toma de decisiones, esto proporcionará una decisión razonada, razonable y que aportara una mayor experiencia al tomador de decisiones y a la entidad.

En conclusión se puede decir que el modelo presentado es una herramienta muy útil para el tomador de decisiones que basa sus elecciones en hechos, pero que necesita conocer los hechos para poder ser más asertivo. En consecuencia se puede decir que se requiere de mayor integración informática (consecuentemente mayores velocidades de respuesta) para que realmente la aplicación de este modelo represente una ventaja competitiva para la empresa.

Una de las más importantes aportaciones que considero de este estudio, es el que no solo ofrece un dato frío de estimación, sino que analiza el conocimiento escondido de los datos, y presenta al tomador de decisiones el dato frío más un rastro histórico del porque es que se llega al dato frío. Ese rastro formado por casos históricos, es parte de la experiencia global de la empresa, y puede ser parte de la experiencia personal del tomador de la decisión. Una meta a lograr debería ser el formarse para ser hombres administrativos y mejores líderes.

Se puede decir que *“Una decisión comprendida es una mejor decisión”*.

## **2. Proyección**

El modelo de apoyo al razonamiento prospectivo presentado en este trabajo, tiene la posibilidad de ser utilizado en diferentes áreas, comprendiendo que requiere de analizar previamente los datos para conocer su comportamiento y establecer los mejores parámetros para su prospección.

Algunos elementos que se pueden considerar para futuros trabajos relacionados son:

- a) Contar con las herramientas informáticas idóneas para lograr la integración de las aplicaciones administrativas con las herramientas de análisis de datos propuestas en este trabajo.
- b) En su caso poder desarrollar para una aplicación ya establecida en el mercado y estable desde el punto de vista de su madurez, la aplicación que permita el diagnóstico y modelado en tiempos que realmente den la ventaja de el proceso a la empresa.
- c) Sería muy interesante poder comparar los resultados de algunas herramientas con el modelo propuesto, en la búsqueda de su mejora y comprensión.
- d) Especialización de ramos de la información.

## XI. BIBLIOGRAFÍA

Aching Guzmán, C. 2006. Matemáticas financieras para toma de decisiones empresariales. serie MYPES.

Ben Klemens. 2009. Modeling with Data: Tools and Techniques for Scientific Computing. PRINCETON UNIVERSITY PRESS.

Calva, Alberto. 2009. Datos económicos y financieros 16/marzo/2009. Acus Consultores (<http://www.acus.com.mx/>)

Deloitte. 2010.

[http://www.deloitte.com/view/en\\_BE/be/services/aers/riskandgovernance/it-governance/index.htm](http://www.deloitte.com/view/en_BE/be/services/aers/riskandgovernance/it-governance/index.htm)

Deloitte. 2005. From Compliance to Sustainable IT Governance: Where to Start. Deloitte, Nov. 2005.

Hamparsum, Bozdogan. 2004. Statistical Data mining and Knowledge discovery. Editorial Chapman & Hall CRC.

Hyunchul Ahn ·Kyoung-jae Kim. 2008. Using genetic algorithms to optimize nearest neighbors for data mining. Published online: 6 March 2008. © Springer Science+Business Media, LLC 2008



Jamie MacLennan, ZhaoHui Tang, Bogdan Crivat. 2009. Data Mining with Microsoft SQL Server 2008. Wiley Publishing Inc.

Jie Wang 1 and Hong Wang2. 2008. Application of Data Mining in the Financial Data Forecasting. © Springer-Verlag Berlin Heidelberg.

Kenneth C. Laudon, Jane P. Laudon. 2007. Management Information Systems: Managing the digital Firm. Prentice Hall Inc.

Klemens, Ben. 2009. Modeling with Data. PRINCETON UNIVERSITY PRESS.

Kovalerchuk, Boris, Evgenii Vityaev. 2006. Data mining for financial applications. <http://www.springerlink.com/content/x7701876pq815457/Año> 2006/  
springerlink.

Liang-Hsuan Chen, Huey-Der Hsiao. 2008. Feature selection to diagnose a business crisis by using a real GA-based support vector machine: An empirical study. ACM.

Musalem, Erik. 2008. La cabeza vs el estomago. Soy Entreprenur, Marzo 2008.

Po-Chang (P.C.) KO' and Ping-Chen (P.C.) Lin. 2007. An Evolutionary Modularized Data Mining Mechanism for Financial Distress Forecasts. Evolutionary computation in data mining, Springer.

Rong-Ho Lin, Yao-Tien Wang. 2009. Developing a business failure prediction model via RST, GRA and CBR. ACM.

Sadaaki Miyamoto, Hidetomo Ichihashi, Katsuhiro Honda. 2008. Algorithms for Fuzzy Clustering. Springer-Verlag Berlin Heidelberg.

Simon, Herbert. 1997. Administrative Behavior. Computer Science, Roskilde University, 4th Edition.

Watson, Ian. 2004. Applying Knowledge Management: Techniques for Building Corporate Memories. Elsevier Science Inc. New York, NY, USA.

Witten, Ian H. 2005. Data Mining: Practical Machine Learning Tools and Techniques. MORGAN KAUFMANN PUBLISHERS.

## XII. APÉNDICE A

Se presentan por cuestiones de privacidad solo los datos comprendidos de marzo de 2006 a Septiembre de 2007, mostrando todos los datos contemplados para este análisis al inicio del mismo. Los datos completos comprenden de Enero de 2002 al mes de Marzo de 2010 inclusive.

		ACTIVO								
AÑO	MES	CAJA	BANCOS	INVER	CLIENTES	ALMACEN	MER TRANS	DEUD DIV	EMPLE	IVA XACR
2006	3	48,061	1,075,467	0	16,155,699	12,906,388	2,337,616	540,301	198,648	10,177,796
2006	4	48,511	1,007,945	0	16,464,862	12,567,332	2,313,265	923,012	174,764	10,771,558
2006	5	48,961	1,086,650	0	15,712,592	13,158,014	622,391	1,411,718	160,182	11,575,529
2006	6	49,411	991,236	0	15,539,285	15,099,881	618,772	1,801,902	158,828	12,613,383
2006	7	49,861	1,068,848	0	16,083,683	14,175,121	1,964,929	2,163,580	178,919	13,241,840
2006	8	50,311	972,905	0	17,159,972	13,786,098	2,341,337	2,208,369	180,517	13,996,631
2006	9	50,761	907,625	0	17,407,727	15,033,092	1,834,254	696,531	184,596	14,976,060
2006	10	51,211	1,364,537	0	17,672,651	14,877,933	1,845,103	762,665	163,262	15,771,297
2006	11	51,661	1,468,153	0	17,921,240	14,247,719	1,650,227	724,811	157,951	16,428,780
2006	12	52,111	946,997	0	16,464,526	14,595,382	827,393	324,257	103,208	17,186,314
2007	1	52,500	2,862,941	0	15,445,193	15,464,787	1,144,712	-217,690	134,477	18,016,294
2007	2	52,500	1,580,217	0	16,166,575	15,437,528	1,256,719	-224,714	167,982	18,739,897
2007	3	54,500	1,362,155	0	16,740,592	15,402,860	1,212,424	499,953	155,021	19,490,826
2007	4	54,500	1,788,377	0	16,483,779	15,251,597	1,925,135	565,902	160,880	20,093,957
2007	5	54,500	1,768,452	0	16,726,727	15,738,562	1,615,913	628,390	180,109	20,983,303
2007	6	57,500	1,207,009	0	17,419,476	15,458,953	1,131,955	693,497	165,476	21,719,564
2007	7	58,000	851,572	0	18,469,148	14,763,904	2,708,887	626,611	182,215	22,498,919
2007	8	59,700	1,540,418	0	18,903,780	16,380,751	1,727,961	670,042	174,885	23,646,235
2007	9	60,200	1,466,432	0	20,261,358	15,802,884	2,183,239	700,468	143,450	24,395,855

ANT PROV	ACT CIR	VAL HIST	TERR/nave	EQP MAQ	EQP TRAN	EQP OFNA	EQP LAB	EQP COM	EQP INC
365,966	43,805,942	3,170,691	8,229,393	3,737,173	3,807,679	700,947	488,454	557,735	444,383
365,966	44,637,214	3,170,691	8,229,393	3,737,173	3,807,679	713,286	488,454	562,221	444,383
365,966	44,142,004	3,170,691	8,229,393	3,738,903	3,807,679	713,286	489,457	569,790	444,383
365,966	47,238,664	3,170,691	8,229,393	3,738,903	3,807,679	713,286	489,457	572,674	444,383
365,966	49,292,747	3,170,691	8,229,393	3,748,903	3,881,592	715,891	489,457	577,853	444,383
365,966	51,062,105	3,170,691	8,229,393	3,778,897	4,104,114	718,497	489,457	582,724	444,383
365,966	51,456,612	3,170,691	9,735,545	3,904,392	4,104,114	719,765	497,495	582,724	444,383
365,966	52,874,625	3,170,691	9,735,545	3,918,392	4,104,114	720,180	499,607	585,814	444,383
365,966	53,016,508	3,170,691	9,735,545	3,918,392	4,104,114	724,150	499,607	585,814	444,383
365,966	50,866,154	3,170,691	9,735,545	4,345,193	4,086,624	753,469	519,588	620,446	462,395
365,966	53,269,179	3,170,691	9,735,545	4,350,073	4,086,624	762,031	526,817	620,446	462,395
365,966	53,542,670	3,170,691	9,735,545	4,350,073	4,086,624	762,378	526,817	628,006	462,395
365,966	55,284,298	3,300,103	9,735,545	4,372,603	4,086,624	762,378	526,817	628,006	462,395
365,966	56,690,092	3,310,367	9,735,545	4,372,603	4,086,624	762,378	538,395	628,006	462,395
365,966	58,061,922	3,317,083	9,735,545	4,372,603	4,086,624	762,378	538,395	628,006	462,395
365,966	58,219,395	3,317,083	9,735,545	4,396,237	4,086,624	768,279	538,395	628,608	462,395
365,966	60,525,223	3,317,083	9,735,545	4,482,371	4,283,842	784,379	538,395	628,608	462,395
365,966	63,469,739	3,317,083	9,735,545	4,500,171	4,283,842	784,379	542,928	635,727	462,395
365,966	65,379,852	3,317,083	9,735,545	4,500,171	4,283,842	784,379	542,928	637,552	462,395

HERR	D MAQ	D MOB	D TRAN	D COM	DA COM	D LAB	DA MEN	ACT FIJ
1,969	-2,258,819	-449,496	-2,785,730	-464,416	-304,159	-315,728	-489	14,559,589
1,969	-2,280,054	-453,155	-2,824,763	-467,184	-306,447	-318,002	-506	14,505,137
1,969	-2,301,289	-456,814	-2,863,795	-469,953	-308,736	-320,277	-524	14,444,164
1,969	-2,322,523	-460,473	-2,902,828	-472,721	-311,025	-322,552	-541	14,375,771
1,969	-2,343,758	-464,132	-2,941,861	-475,490	-313,314	-324,827	-559	14,396,193
16,739	-2,365,178	-468,656	-2,982,434	-480,232	-315,603	-328,949	-576	14,593,266
18,304	-2,386,848	-472,548	-3,018,962	-488,014	-317,892	-331,224	-717	16,161,209
19,474	-2,408,518	-476,439	-3,055,490	-495,797	-320,181	-333,499	-858	16,107,419
21,702	-2,432,397	-476,335	-3,057,116	-500,157	-322,469	-331,876	-1,024	16,083,025
28,728	-2,614,288	-512,713	-3,077,205	-519,725	-356,337	-362,023	-1,382	16,279,006
28,728	-2,632,574	-515,956	-3,108,553	-527,114	-358,626	-363,763	-14,188	16,222,575
28,728	-2,651,664	-519,193	-3,142,193	-534,503	-360,915	-365,412	-14,425	16,162,953
28,728	-2,670,568	-522,428	-3,173,540	-542,081	-363,204	-366,915	-14,588	16,249,874
28,728	-2,690,008	-525,660	-3,204,887	-549,471	-365,493	-368,419	-14,751	16,206,353
28,728	-2,708,912	-528,891	-3,236,235	-556,860	-367,782	-370,165	-14,915	16,147,999
28,728	-2,727,816	-532,123	-3,267,582	-564,249	-370,071	-371,910	-15,078	16,113,066
28,728	-2,747,828	-535,436	-3,298,929	-572,220	-372,359	-373,414	-15,606	16,345,554
28,728	-2,768,020	-538,859	-3,334,385	-579,813	-374,648	-375,387	-15,769	16,303,916
28,728	-2,788,279	-542,087	-3,369,841	-587,584	-376,937	-377,170	-16,005	16,234,720

PAG ANT	DEP GAR	GST PRE	GST INS	AMO GPRE	AMO ACMGINS	AMO GIND	AMO OTROS	ACT DIF	ACT TOT
245,176	291,915	60,866	529,256	-60,866	0	-202,894	249,441	1,112,893	59,478,424
243,427	295,277	60,866	529,256	-60,866	0	-204,458	243,306	1,106,807	60,249,158
240,715	295,277	60,866	529,256	-60,866	0	-206,023	253,130	1,112,355	59,698,522
334,417	295,277	60,866	529,256	-60,866	0	-206,023	253,130	1,206,057	62,820,491
360,878	295,277	60,866	529,256	-60,866	0	-207,587	300,899	1,278,722	64,967,662
402,851	295,277	60,866	600,587	-60,866	0	-209,151	294,764	1,384,326	67,039,697
393,160	295,277	60,866	780,401	-60,866	0	-210,716	288,628	1,546,749	69,164,571
417,723	295,277	60,866	919,342	-60,866	0	-212,280	282,493	1,702,554	70,684,598
411,789	295,977	60,866	1,086,234	-60,866	0	-213,844	276,358	1,856,512	70,956,045
61,306	78,901	60,866	1,167,275	-60,866	0	-245,207	281,964	1,344,238	68,489,399
40,143	80,316	60,866	1,167,275	-60,866	0	-246,772	275,828	1,316,791	70,808,546
132,826	80,316	60,866	1,184,958	-60,866	0	-248,336	307,827	1,457,591	71,163,213
195,488	84,344	60,866	1,216,481	-60,866	0	-249,900	301,692	1,548,104	73,082,276
244,996	84,344	60,866	1,267,892	-60,866	0	-251,465	301,417	1,647,184	74,543,629
270,347	84,344	60,866	1,277,495	-60,866	0	-253,029	349,555	1,728,712	75,938,633
252,203	91,244	60,866	1,382,827	-60,866	0	-254,593	343,318	1,814,999	76,147,460
322,719	91,244	60,866	1,416,833	-60,866	0	-256,158	337,081	1,911,719	78,782,496
368,854	87,321	60,866	1,417,401	-60,866	0	-257,722	330,844	1,946,698	81,720,353
375,781	87,321	60,866	1,453,004	-60,866	0	-259,286	324,607	1,981,426	83,595,999

PASIVO								
PROV NAL	PROV EXT	ACR DIV	IMP X PAG	PRES BANC	IVA X PAG	PAS CIR	DOC X PAG	PAS LP
3,741,016	4,598,234	2,696,566	211,035	-7,859	12,374,948	23,613,939	16,031,069	16,031,069
3,959,013	3,970,954	2,648,767	296,280	-9,765	13,101,389	23,966,639	16,031,069	16,031,069
4,330,271	1,756,277	2,871,498	232,272	-9,765	13,976,761	23,157,315	15,953,272	15,953,272
4,814,365	2,815,409	2,987,877	276,380	590,235	14,662,309	26,146,576	15,703,766	15,703,766
5,097,581	3,276,425	3,071,761	219,156	590,235	15,555,502	27,810,660	15,781,563	15,781,563
5,225,942	3,540,467	3,124,257	291,134	843,235	16,143,137	29,168,171	15,781,563	15,781,563
5,792,925	4,215,515	3,508,900	205,042	350,649	16,995,385	31,068,416	15,537,867	15,537,867
6,436,137	3,501,384	3,618,919	289,711	130,871	17,938,689	31,915,711	15,537,867	15,537,867
5,913,412	2,819,034	3,658,200	212,007	290,982	18,650,909	31,544,544	15,537,867	15,537,867
6,698,788	6,514,042	1,189,092	579,231	806,094	19,015,995	34,803,243	15,286,593	15,286,593
8,743,277	6,870,603	924,785	220,785	128,556	19,674,841	36,562,846	15,286,593	15,286,593
7,933,724	7,140,367	1,206,618	334,603	118,667	20,080,824	36,814,804	15,286,593	15,286,593
8,061,253	6,639,474	1,260,418	277,741	1,208,778	21,043,118	38,490,782	15,286,593	15,286,593
8,787,430	7,423,271	1,444,434	374,221	598,889	21,761,736	40,389,981	14,948,409	14,948,409
8,801,673	6,971,076	1,590,836	272,188	1,089,000	22,484,633	41,209,406	14,948,409	14,948,409
8,553,371	5,898,197	1,633,764	343,240	1,261,202	23,344,967	41,034,741	14,948,409	14,948,409
8,954,648	6,901,092	1,956,362	277,235	1,426,315	24,295,214	43,810,867	14,388,980	14,388,980
9,663,826	7,092,105	2,041,177	390,974	1,803,453	25,224,455	46,215,991	14,388,980	14,388,980
9,977,425	7,586,643	1,708,968	293,189	1,859,817	26,270,663	47,696,706	14,388,980	14,388,980



UTL OPER	OTROS ING	PROD FIN	OTROS PRO	GST FIN	GST OTRO	REEXP	UTIL AIMP
852,665	0	-8,041	-1,800	118,791	0	0	743,716
503,084	0	0	-2,000	87,662	0	0	417,422
424,430	0	-0	-2,500	88,895	0	0	338,035
545,849	0	0	-2,350	166,435	0	0	381,764
446,851	0	-16,656	-2,700	42,040	0	0	424,167
716,325	0	-41,302	-2,500	71,257	0	0	688,870
511,949	0	-15,029	-2,000	53,837	0	0	475,141
713,353	0	-6,183	0	47,254	0	0	672,282
686,743	0	-14,696	0	57,931	0	0	643,508
-4,965,421	0	-2,123	-55,406	351,338	0	0	-5,259,230
619,221	0	-21,065	0	81,290	0	0	558,996
147,064	0	-1,406	0	49,571	0	0	98,899
429,538	0	-5,562	0	204,583	0	0	230,517
8,640	0	-339	-290	121,492	0	0	-112,223
594,201	0	-13,620	-18,768	63,563	0	0	563,026
391,810	0	-52,299	-300	73,522	0	0	370,887
475,653	0	-22,281	0	94,537	0	0	403,397
613,284	0	-6,345	-200	99,648	0	0	520,181
467,126	0	-14	0	84,760	0	0	382,380

Tabla A.1. Tablas de datos Originales (muestra)

AREA	DESCRIPCION	RESUMIDO	AREA	DESCRIPCION	RESUMIDO
<b>GENERALES</b>			<b>ESTADO DE RESULTADOS</b>		
	AÑO	AÑO		VENTAS	VENTAS
	MES	MES		DEVOLUCIONES Y REBAJAS	DEV REB
<b>BLANZA GENERAL</b>				VENTA NETA	VENT NET
	CAJA	CAJA		<b>COSTO DE VENTAS</b>	COST VEN A
	BANCOS	BANCOS		COSTO DE VENTAS	COST VENT
	INVERSIONES EN VALORES	INVER		COMPRAS	COMPRAS
	CLIENTES	CLIENTES		DEVOLUCIONES Y REBAJAS S/COMPRAS	DEV REB SC
	ALMACEN	ALMACEN		MANO DE OBRA	MANO OBRA
	MERCANCIAS EN TRANSITO	MER TRANS		<b>TOTAL COSTO DE VENTAS</b>	TOT CV
	DEUDORES DIVERSOS	DEUD DIV		<b>UTILIDAD BRUTA</b>	UTIL BRUT
	FUNCIONARIOS Y EMPLEADOS	EMPLE		<b>GASTOS DE OPERACION</b>	GST OPER
	IVA POR ACREDITAR	IVA X ACR		GASTOS DE ADMINISTRACION	GST ADM
	ANTICIPO A PROVEEDORES	ANT PROV		GASTOS DE FABRICACION	GST FAB
	<b>ACTIVO CIRCULANTE</b>	<b>ACT CIR</b>		GTOS DE VENTA MEXICO	GST OM
	VALOR HISTORICO	VAL HIST		GASTOS DE VENTA OFNA. CENTRA	GST OC
	ACT DE TERRENO	TERR		GASTOS DE VENTA DELEGACION I - SLP	GST D1
	MAQUINARIA Y EQUIPO	EQP MAQ		GASTOS DE VENTA OFNA. AGUSCALIENTES	GST D2
	EQUIPO DE TRANSPORTE	EQP TRAN		GASTOS DE VENTA DELEGACION MONTERREY	GST D3
	MOBILIARIO Y EQ. DE OFNA.	EQP OFNA		GASTOS DE VENTA DELEGACION DELICIAS	GST D4
	EQUIPO DE LABORATORIO	EQP LAB		GASTOS VENTA GUADALAJARA	GST D5
	EQUIPO DE COMPUTO	EQP COM		GASTOS DE VENTA DEL VI	GST D6
	EQUIPO CONTRA INCENDIOS	EQP INC		UTILIDAD OPERATIVA	UTL OPER
	HERRAMIENTAS MENORES	HERR		OTROS INGRESOS Y GASTOS	OTROS ING
	DEPRECIACION ACUM. DE MAQ. Y	D MAQ		PRODUCTOS FINANCIEROS	PROD FIN
	DEPRECIACION DE MOBILIARIO Y	D MOB		OTROS PRODUCTOS	OTROS PROD
	DEPRECIACION EQUIPO DE TRANS	D TRAN		GASTOS FINANCIEROS	GST FIN
	DEPRECIACION EQUIPO DE COMPU	D COM		OTROS GASTOS	GST OTRO
	DEPRECIACION ACUMULADA EQ. C	DA COM		CORRECCION POR REEXPRESION	REEXP
	DEPRECIACION EQ. DE LABORATO	D LAB		UTILIDAD ANTES DE IMPUESTOS	UTIL AIMP
	DEP ACUM HTAS MENORES	DA MEN			
	<b>ACTIVO FIJO</b>	<b>ACT FIJ</b>			
	PAGOS ANTICIPADOS	PAG ANT			
	DEPOSITOS EN GARANTIA	DEP GAR			
	GASTOS PREOPERATIVOS	GST PRE			
	GTOS. DE INSTALACION (AMPL.	GST INS			
	AMORTIZACION DE GASTOS PREOP	AMO GPRE			
	AMORTIZACION ACUM. GTOS. INS	AMO ACMGINS			
	AMORTIZACION DE GASTOS DE IN	AMO GIND			
	OTROS GASTOS POR AMORTIZAR	AMO OTROS			
	<b>ACTIVO DIFERIDO</b>	<b>ACT DIF</b>			
	<b>ACTIVO TOTAL</b>	<b>ACT TOT</b>			
	PROVEEDORES NACIONALES	PROV NAL			
	PROVEEDORES EXTRANJEROS	PROV EXT			
	ACREEDORES DIVERSOS	ACR DIV			
	IMPUESTOS POR PAGAR	IMP X PAG			
	PRESTAMOS BANCARIOS	PRES BANC			
	I.V.A. POR PAGAR	IVA X PAG			
	<b>PASIVO CIRCULAN</b>	<b>PAS CIR</b>			
	DOCUMENTOS POR PAGAR	DOC X PAG			
	PASIVO A LARGO PLAZO	PAS LP			
	CAPITAL SOCIAL	CAP SOC			
	EXCESO Y/O INSUF. EN APORTACION DE PASIVO	EXSO APOR			
	RESULTADO DE EJERCICIOS ANTERIOR	EJER ANT			
	RESULTADO DEL EJERCICIO	EJER ANT			
	ACTUALIZACION RES EJERCICIO 1998	EJER ANT			
	ACTUALIZACION RES EJERCICIO 1999	EJER ANT			
	ACTUALIZACION RES EJERCICIO 2000	EJER ANT			
	ACTUALIZACION RES EJERCICIO 2001	EJER ANT			
	RESULTADO DEL EJERCICIO 2000	EJER ANT			
	RESULTADO DEL EJERCICIO 2001	EJER ANT			
	RESULTADO DEL EJERCICIO 2002	EJER ANT			
	RESULTADO DEL EJERCICIO 2003	EJER ANT			
	RESULTADO DEL EJERCICIO 2004	EJER ANT			
	UTILIDAD DEL EJERCICIO	UTL EJER			
	<b>CAPITAL CONTABLE</b>	<b>CAP CONT</b>			
	<b>PASIVO Y CAPITAL</b>	<b>PAS Y CAP</b>			

Tabla A.2. Descripción total de campos.

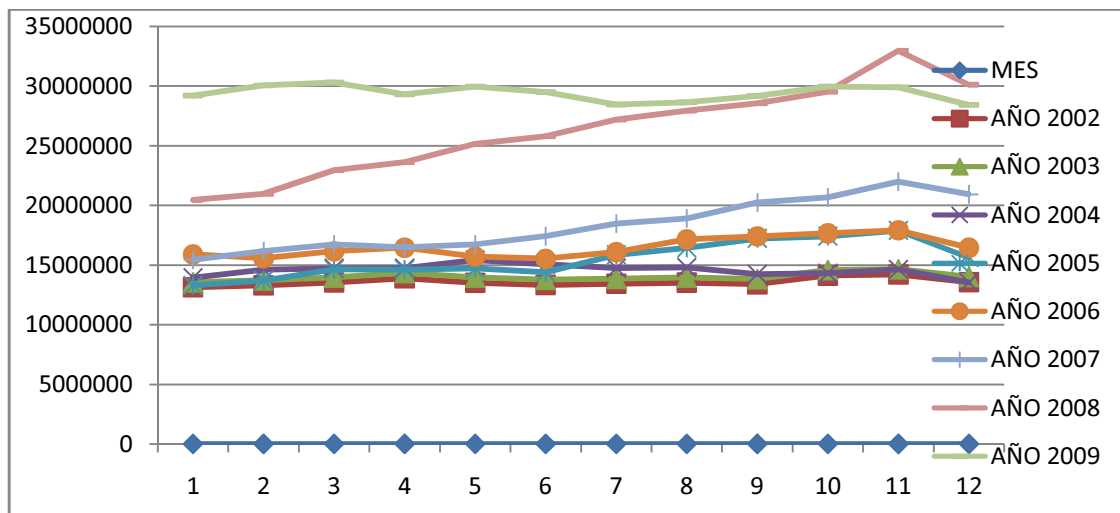
### XIII. APÉNDICE B

Análisis grafico de datos definidos de estudio.

a) Análisis de la variable ventas de la empresa (tabla, gráfica y gráfica anualizada).

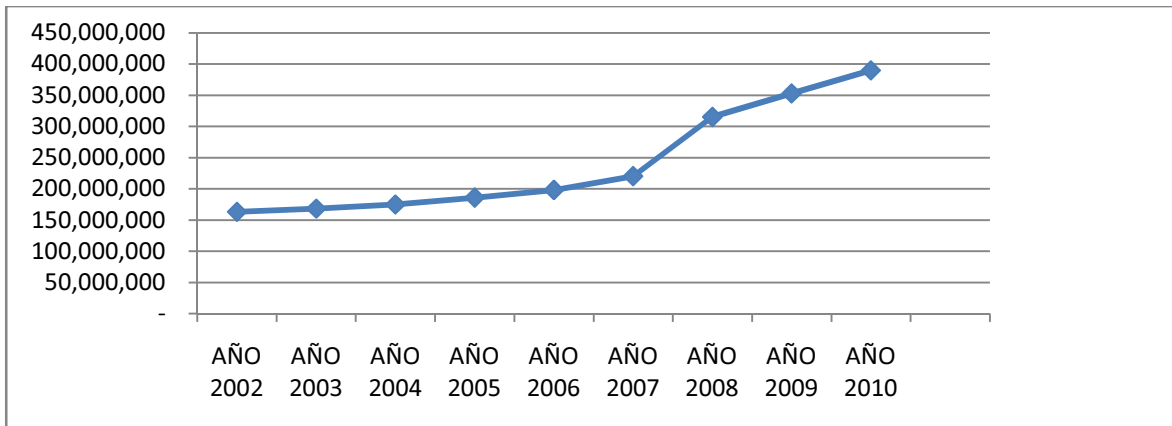
INFORMACION DE CLIENTES DE LA EMPRESA (VENTAS)								
MES	AÑO 2002	AÑO 2003	AÑO 2004	AÑO 2005	AÑO 2006	AÑO 2007	AÑO 2008	AÑO 2009
1	13,150,263	13,556,972	13,935,385	13,304,690	15,898,325	15,445,193	20,453,883	29,197,062
2	13,314,141	13,725,919	14,611,877	13,719,157	15,565,578	16,166,575	20,973,593	30,047,369
3	13,548,736	13,967,769	14,728,690	14,638,451	16,155,699	16,740,592	22,955,214	30,302,187
4	13,901,191	14,331,124	14,729,567	14,633,458	16,464,862	16,483,779	23,615,907	29,305,105
5	13,528,320	13,946,721	15,415,111	14,716,000	15,712,592	16,726,727	25,153,308	29,962,681
6	13,347,176	13,759,976	15,055,992	14,394,774	15,539,285	17,419,476	25,801,573	29,499,206
7	13,447,234	13,863,128	14,745,431	15,852,388	16,083,683	18,469,148	27,182,795	28,462,855
8	13,534,730	13,953,330	14,789,584	16,441,432	17,159,972	18,903,780	27,934,380	28,635,309
9	13,403,541	13,818,084	14,227,591	17,233,942	17,407,727	20,261,358	28,551,135	29,167,839
10	14,126,678	14,563,586	14,310,373	17,424,333	17,672,651	20,670,649	29,534,078	29,956,956
11	14,216,345	14,656,026	14,631,794	17,898,857	17,921,240	21,995,195	32,968,810	29,892,223
12	13,584,049	14,004,175	13,532,999	15,643,046	16,464,526	20,937,000	30,114,353	28,423,569
	163,102,405	168,146,809	174,714,395	185,900,528	198,046,140	220,219,471	315,239,031	352,852,361

Tabla B.1. Ventas de la empresa de estudio.



Gráfica B.2. Ventas de la empresa de estudio.



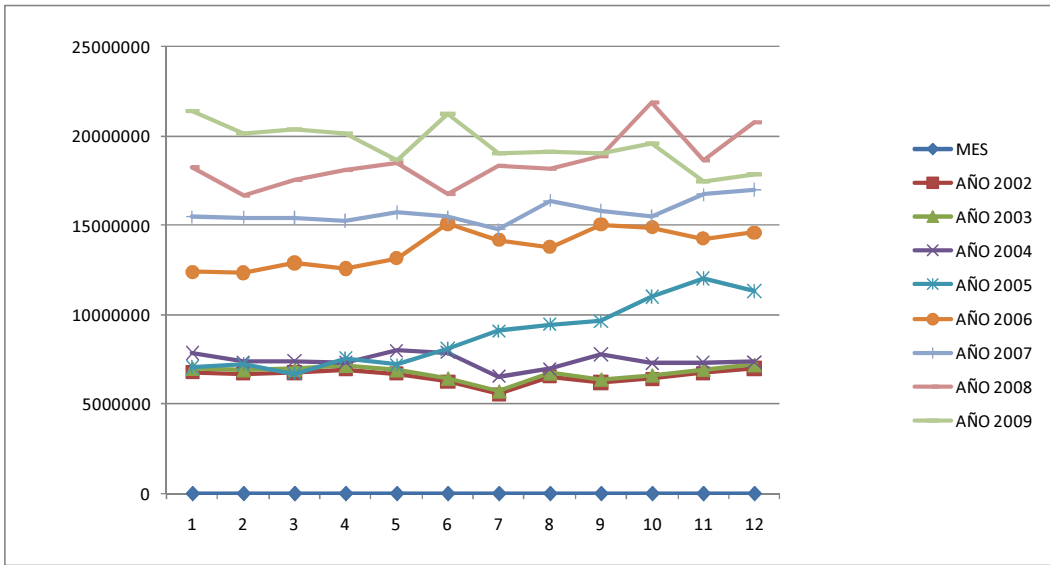


Gráfica B.3. Anual de ventas de la empresa de estudio.

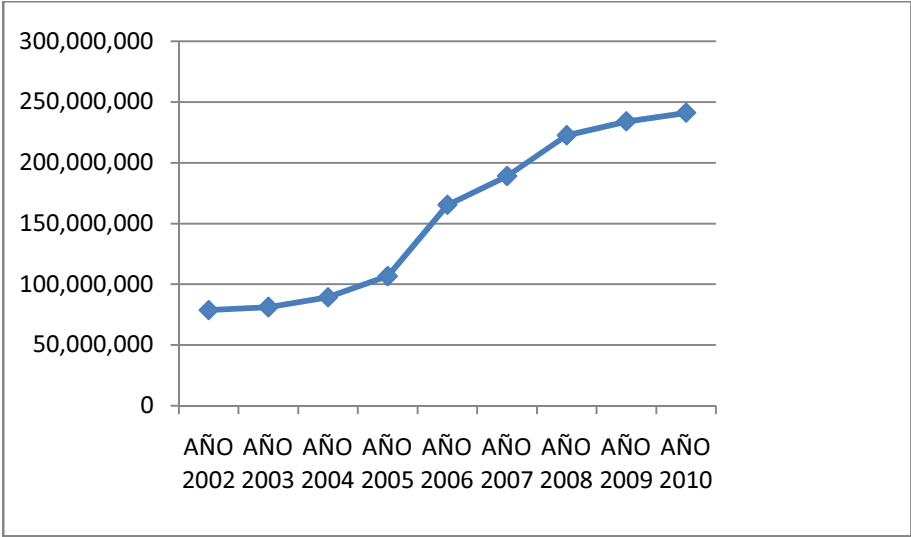
b) Análisis de la variable almacén de la empresa (tabla, gráfica y gráfica anualizada).

INFORMACION DE ALMACEN DE LA EMPRESA (VALOR DE INVENTARIO)								
MES	AÑO 2002	AÑO 2003	AÑO 2004	AÑO 2005	AÑO 2006	AÑO 2007	AÑO 2008	AÑO 2009
1	6,796,037	7,006,224	7,867,077	7,079,607	12,387,695	15,464,787	18,222,384	21,363,629
2	6,708,189	6,915,659	7,356,632	7,237,005	12,337,242	15,437,528	16,698,381	20,110,504
3	6,750,067	6,958,832	7,417,178	6,707,531	12,906,388	15,402,860	17,507,641	20,340,964
4	6,950,057	7,165,007	7,333,974	7,562,156	12,567,332	15,251,597	18,105,632	20,140,199
5	6,714,163	6,921,817	7,982,157	7,202,200	13,158,014	15,738,562	18,476,596	18,675,556
6	6,250,616	6,443,934	7,882,717	8,095,361	15,099,881	15,458,953	16,743,967	21,220,513
7	5,547,931	5,719,516	6,554,737	9,080,298	14,175,121	14,763,904	18,334,514	19,064,206
8	6,538,363	6,740,581	6,969,749	9,455,738	13,786,098	16,380,751	18,191,447	19,102,984
9	6,195,782	6,387,404	7,771,562	9,642,747	15,033,092	15,802,884	18,903,871	19,050,533
10	6,422,276	6,620,903	7,281,395	11,013,121	14,877,933	15,504,313	21,873,771	19,565,636
11	6,737,248	6,945,617	7,323,470	12,041,568	14,247,719	16,713,249	18,609,119	17,449,460
12	6,992,620	7,208,886	7,366,742	11,325,703	14,595,382	16,991,861	20,784,001	17,833,939
	78,603,350	81,034,381	89,107,390	106,443,035	165,171,897	188,911,249	222,451,323	233,918,124

Tabla B.4. Valor del almacén de la empresa de estudio.



Gráfica B.5. Valor del almacén de la empresa de estudio.



Gráfica B.6. Anual del valor del almacén de la empresa de estudio.

## XIV. APÉNDICE C

### Estadandarización y análisis de datos del caso de estudio

a) Procedimiento de estandarización de datos.

PROCEDIMIENTO PARA OBTENER DATOS ESTANDARIZADOS						
AÑO	MES	CLIENTES	ALMACEN	PMES	P.CLIENTES	P.ALMACEN
2006	1	15,898,325	12,387,695	1.28	8.03	7.50
	2	15,565,578	12,337,242	2.56	7.86	7.47
	3	16,155,699	12,906,388	3.85	8.16	7.81
	4	16,464,862	12,567,332	5.13	8.31	7.61
	5	15,712,592	13,158,014	6.41	7.93	7.97
	6	15,539,285	15,099,881	7.69	7.85	9.14
	7	16,083,683	14,175,121	8.97	8.12	8.58
	8	17,159,972	13,786,098	10.26	8.66	8.35
	9	17,407,727	15,033,092	11.54	8.79	9.10
	10	17,672,651	14,877,933	12.82	8.92	9.01
	11	17,921,240	14,247,719	14.10	9.05	8.63
	12	16,464,526	14,595,382	15.38	8.31	8.84
SUMATORIA	78	198,046,139.92	165,171,896.93	100.00	100.00	100.00

Gráfica C.1. Proceso de Estandarización de datos.

b) Datos de trabajo estandarizados.

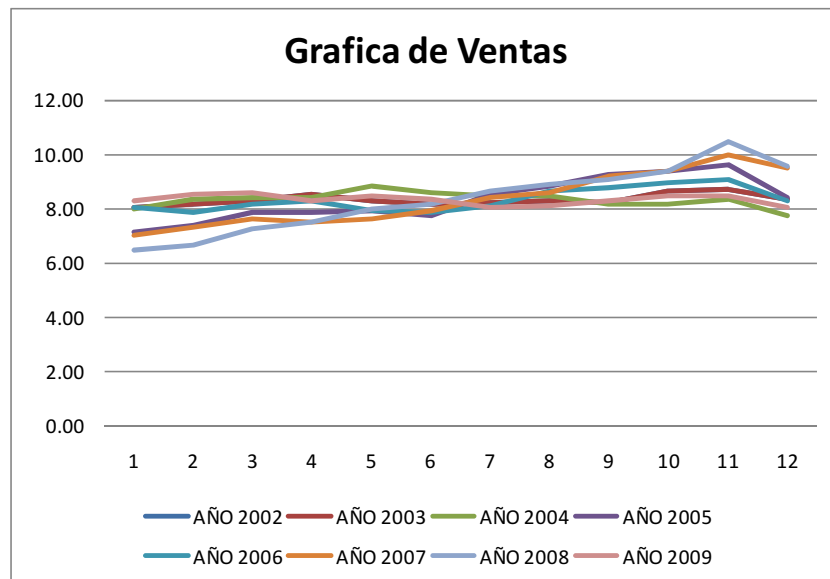
VENTAS								
MES	AÑO 2002	AÑO 2003	AÑO 2004	AÑO 2005	AÑO 2006	AÑO 2007	AÑO 2008	AÑO 2009
1	8.06	8.06	7.98	7.16	8.03	7.01	6.49	8.27
2	8.16	8.16	8.36	7.38	7.86	7.34	6.65	8.52
3	8.31	8.31	8.43	7.87	8.16	7.60	7.28	8.59
4	8.52	8.52	8.43	7.87	8.31	7.49	7.49	8.31
5	8.29	8.29	8.82	7.92	7.93	7.60	7.98	8.49
6	8.18	8.18	8.62	7.74	7.85	7.91	8.18	8.36
7	8.24	8.24	8.44	8.53	8.12	8.39	8.62	8.07
8	8.30	8.30	8.47	8.84	8.66	8.58	8.86	8.12
9	8.22	8.22	8.14	9.27	8.79	9.20	9.06	8.27
10	8.66	8.66	8.19	9.37	8.92	9.39	9.37	8.49
11	8.72	8.72	8.37	9.63	9.05	9.99	10.46	8.47
12	8.33	8.33	7.75	8.41	8.31	9.51	9.55	8.06

Gráfica C.2. Datos estandarizados de ventas.

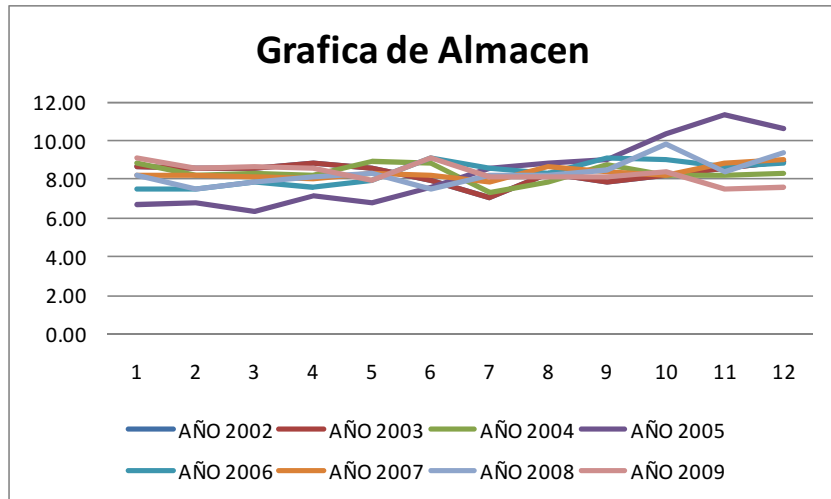
ALMACEN								
MES	AÑO 2002	AÑO 2003	AÑO 2004	AÑO 2005	AÑO 2006	AÑO 2007	AÑO 2008	AÑO 2009
1	8.65	8.65	8.83	6.65	7.50	8.19	8.19	9.13
2	8.53	8.53	8.26	6.80	7.47	8.17	7.51	8.60
3	8.59	8.59	8.32	6.30	7.81	8.15	7.87	8.70
4	8.84	8.84	8.23	7.10	7.61	8.07	8.14	8.61
5	8.54	8.54	8.96	6.77	7.97	8.33	8.31	7.98
6	7.95	7.95	8.85	7.61	9.14	8.18	7.53	9.07
7	7.06	7.06	7.36	8.53	8.58	7.82	8.24	8.15
8	8.32	8.32	7.82	8.88	8.35	8.67	8.18	8.17
9	7.88	7.88	8.72	9.06	9.10	8.37	8.50	8.14
10	8.17	8.17	8.17	10.35	9.01	8.21	9.83	8.36
11	8.57	8.57	8.22	11.31	8.63	8.85	8.37	7.46
12	8.90	8.90	8.27	10.64	8.84	8.99	9.34	7.62

Gráfica C.3. Datos estandarizados de valor del almacén.

c) Gráficas de los datos estandarizados.



Gráfica C.4. Gráfica de datos estandarizados de ventas.



Gráfica C.5. Gráfica de datos estandarizados de valor de almacén.

## XV. APÉNDICE D

### Programación en la aplicación Wolfram Mathematica 7.0 (TrialV)

```
datos=Import["c:\\md\\dm.dat"];
nocercanos=3;
datoactual=Last[datos];
nogrupos=6;
nosgrupos=3;
Print["*****"];
Print["Modelo de análisis prospectivo."]
Print[" Datos de trabajo - vector: ",datoactual," con ",nogrupos," clusters, y ",nocercanos," vecinos
cercanos."];
Print["*****"]
Print[" Datos agrupados por clusters en ",nogrupos," clusters."]
grupos=FindClusters[datos,nogrupos]
Print[" Grafico tridimensional de los datos por cada cluster."]
Map[ListPointPlot3D[#]&,grupos]
Print[" Grafico de los datos agrupados."]
ListPointPlot3D[grupos]
cercanos=Nearest[datos,datoactual,nocercanos+1];
posicionoriginal=Position[datos,cercanos[[1]]];
Print[" Datos en la vecindad más cercana dentro del cluster a la posición original",posicionoriginal,"
",cercanos];
posicioncerca1=Position[datos,cercanos[[2]]];
posicioncerca2=Position[datos,cercanos[[3]]];
If[nocercanos>2,posicioncerca3=Position[datos,cercanos[[4]]];
If[nocercanos>3,posicioncerca4=Position[datos,cercanos[[5]]];
If[nocercanos>4,posicioncerca5=Position[datos,cercanos[[6]]];
If[nocercanos==2,posicionescercanas={posicioncerca1,posicioncerca2};
If[nocercanos==3,posicionescercanas={posicioncerca1,posicioncerca2,posicioncerca3};
If[nocercanos==4,posicionescercanas={posicioncerca1,posicioncerca2,posicioncerca3,posicioncerca4};
```

```

If[nocercanos==5,posicionescercanas={posicioncerca1,posicioncerca2,posicioncerca3,posicioncerca4,po
sicioncerca5}];
Print[" Posiciones en la vecindad actual cercana del vector: "];
posicionescercanas=Flatten[posicionescercanas]
Print[" Posiciones del estado N-1 o anterior: "];
posicionesanteriores= posicionescercanas -1
Print[" Posiciones del estado N+1 o siguiente: "];
posicionessiguientes= posicionescercanas +1
Print[" Datos del estado N+1 o siguiente: "];
estadomas1=datos[[posicionessiguientes]]
Print[" Resultado Prospectivo con cercanos igual a ",nocercanos];
pronos=Mean[estadomas1]

```

Código D.1. Programa en de procesamiento de datos.

## XVI. APÉNDICE E

### Análisis del no de clústeres y cercanos con menos error con resultados conocidos.

Análisis tomando como base una muestra de 4 años (2006 al 2009 inclusive) 48 lecturas tomadas de manera aleatoria considerando 8 muestras de meses (muestra del 16.6%).

PRONOSTICO A: JUL 2009			PRONOSTICO A: MAR 2009			PRONOSTICO A: SEP 2008		
	Ventas	Almacen		Ventas	Almacen		Ventas	Almacen
<b>CON 2 CERCANOS</b>			<b>CON 2 CERCANOS</b>			<b>CON 2 CERCANOS</b>		
Datos Estándar	8.2845	7.969	Datos Estándar	8.3164	8.59555	Datos Estándar	8.4665	8.91155
Cifra Pronosticada	29,232,054	18,640,935	Cifra Pronosticada	29,344,614	20,106,549	Cifra Pronosticada	26,689,713	19,823,861
Diferencia Real	769,199	-423,270	Diferencia Real	-957,573	-234,415	Diferencia Real	-1,244,668	1,632,414
% Diferencias	2.70%	-2.22%	% Diferencias	-3.16%	-1.15%	% Diferencias	-4.46%	8.97%
Prom % Diferencias		0.24%	Prom % Diferencias		-2.16%	Prom % Diferencias		2.26%
<b>CON 3 CERCANOS</b>			<b>CON 3 CERCANOS</b>			<b>CON 3 CERCANOS</b>		
Datos Estándar	8.31587	7.91777	Datos Estándar	8.35433	8.505	Datos Estándar	8.71117	8.72943
Cifra Pronosticada	29,342,744	18,521,099	Cifra Pronosticada	29,478,451	19,894,736	Cifra Pronosticada	27,461,008	19,418,733
Diferencia Real	879,889	-543,106	Diferencia Real	-823,736	-446,228	Diferencia Real	-473,372	1,227,286
% Diferencias	3.09%	-2.85%	% Diferencias	-2.72%	-2.19%	% Diferencias	-1.69%	6.75%
Prom % Diferencias		0.12%	Prom % Diferencias		-2.46%	Prom % Diferencias		2.53%
<b>CON 4 CERCANOS</b>			<b>CON 4 CERCANOS</b>			<b>CON 4 CERCANOS</b>		
Datos Estándar	8.29808	7.70285	Datos Estándar	8.1662	8.41712	Datos Estándar	8.58785	8.51765
Cifra Pronosticada	29,279,971	18,018,362	Cifra Pronosticada	28,814,629	19,689,169	Cifra Pronosticada	27,072,255	18,947,625
Diferencia Real	817,116	-1,045,843	Diferencia Real	-1,487,558	-651,795	Diferencia Real	-862,125	756,178
% Diferencias	2.87%	-5.49%	% Diferencias	-4.91%	-3.20%	% Diferencias	-3.09%	4.16%
Prom % Diferencias		-1.31%	Prom % Diferencias		-4.06%	Prom % Diferencias		0.54%
<b>CON 5 CERCANOS</b>			<b>CON 5 CERCANOS</b>			<b>CON 5 CERCANOS</b>		
Datos Estándar	8.27818	7.57656	Datos Estándar	8.2314	8.50888	Datos Estándar	8.51762	8.39354
Cifra Pronosticada	29,209,754	17,722,947	Cifra Pronosticada	29,044,689	19,903,812	Cifra Pronosticada	26,850,863	18,671,541
Diferencia Real	746,899	-1,341,258	Diferencia Real	-1,257,498	-437,152	Diferencia Real	-1,083,517	480,094
% Diferencias	2.62%	-7.04%	% Diferencias	-4.15%	-2.15%	% Diferencias	-3.88%	2.64%
Prom % Diferencias		-2.21%	Prom % Diferencias		-3.15%	Prom % Diferencias		-0.62%

Tabla E.1a. Análisis del número adecuado de vecinos cercanos.



PRONOSTICO A: MAYO 2008			PRONOSTICO A: DICIEMBRE 2007			PRONOSTICO A: JULIO 2007		
Ventas	Almacen		Ventas	Almacen		Ventas	Almacen	
<b>CON 2 CERCANOS</b>			<b>CON 2 CERCANOS</b>			<b>CON 2 CERCANOS</b>		
Datos Estándar	8.20925	8.64455	Datos Estándar	8.33875	8.8834	Datos Estándar	8.22165	7.06475
Cifra Pronosticada	25,878,760	19,229,916	Cifra Pronosticada	18,363,551	16,781,742	Cifra Pronosticada	18,105,674	13,346,107
Diferencia Real	725,453	753,320	Diferencia Real	-3,631,643	68,493	Diferencia Real	686,198	-2,112,845
% Diferencias	2.88%	4.08%	% Diferencias	-16.51%	0.41%	% Diferencias	3.94%	-13.67%
Prom % Diferencias		3.48%	Prom % Diferencias		-8.05%	Prom % Diferencias		-4.86%
<b>CON 3 CERCANOS</b>			<b>CON 3 CERCANOS</b>			<b>CON 3 CERCANOS</b>		
Datos Estándar	8.11743	8.11743	Datos Estándar	8.33533	8.88763	Datos Estándar	8.32353	7.5534
Cifra Pronosticada	25,589,308	18,057,330	Cifra Pronosticada	18,356,020	16,789,733	Cifra Pronosticada	18,330,034	14,269,222
Diferencia Real	436,000	-419,266	Diferencia Real	-3,639,175	76,484	Diferencia Real	910,558	-1,189,730
% Diferencias	1.73%	-2.27%	% Diferencias	-16.55%	0.46%	% Diferencias	5.23%	-7.70%
Prom % Diferencias		-0.27%	Prom % Diferencias		-8.04%	Prom % Diferencias		-1.23%
<b>CON 4 CERCANOS</b>			<b>CON 4 CERCANOS</b>			<b>CON 4 CERCANOS</b>		
Datos Estándar	8.0671	8.0054	Datos Estándar	8.74848	8.8775	Datos Estándar	8.27295	7.81055
Cifra Pronosticada	25,430,648	17,808,118	Cifra Pronosticada	19,265,856	16,770,596	Cifra Pronosticada	18,218,647	14,755,008
Diferencia Real	277,340	-668,478	Diferencia Real	-2,729,338	57,347	Diferencia Real	799,171	-703,945
% Diferencias	1.10%	-3.62%	% Diferencias	-12.41%	0.34%	% Diferencias	4.59%	-4.55%
Prom % Diferencias		-1.26%	Prom % Diferencias		-6.03%	Prom % Diferencias		0.02%
<b>CON 5 CERCANOS</b>			<b>CON 5 CERCANOS</b>			<b>CON 5 CERCANOS</b>		
Datos Estándar	8.118	8.10244	Datos Estándar	8.80858	8.8272	Datos Estándar	8.3063	7.71964
Cifra Pronosticada	25,591,105	18,023,985	Cifra Pronosticada	19,398,208	16,675,574	Cifra Pronosticada	18,292,090	14,583,268
Diferencia Real	437,797	-452,611	Diferencia Real	-2,596,986	-37,675	Diferencia Real	872,614	-875,684
% Diferencias	1.74%	-2.45%	% Diferencias	-11.81%	-0.23%	% Diferencias	5.01%	-5.66%
Prom % Diferencias		-0.35%	Prom % Diferencias		-6.02%	Prom % Diferencias		-0.33%

Tabla E.1b. Análisis del número adecuado de vecinos cercanos.

PRONOSTICO A: ABRIL 2007			PRONOSTICO A: OCTUBRE 2006		
Ventas	Almacen		Ventas	Almacen	
<b>CON 2 CERCANOS</b>			<b>CON 2 CERCANOS</b>		
Datos Estándar	8.4183	8.22525	Datos Estándar	8.7818	9.259
Cifra Pronosticada	18,538,736	15,538,422	Cifra Pronosticada	17,392,016	15,293,266
Diferencia Real	1,798,144	135,562	Diferencia Real	-280,635	415,333
% Diferencias	10.74%	0.88%	% Diferencias	-1.59%	2.79%
Prom % Diferencias		5.81%	Prom % Diferencias		0.60%
<b>CON 3 CERCANOS</b>			<b>CON 3 CERCANOS</b>		
Datos Estándar	8.42243	8.227	Datos Estándar	8.9447	9.19237
Cifra Pronosticada	18,547,831	15,541,728	Cifra Pronosticada	17,714,633	15,183,212
Diferencia Real	1,807,239	138,868	Diferencia Real	41,982	305,279
% Diferencias	10.80%	0.90%	% Diferencias	0.24%	2.05%
Prom % Diferencias		5.85%	Prom % Diferencias		1.14%
<b>CON 4 CERCANOS</b>			<b>CON 4 CERCANOS</b>		
Datos Estándar	8.43987	8.38923	Datos Estándar	8.866	8.92868
Cifra Pronosticada	18,586,237	15,848,199	Cifra Pronosticada	17,558,771	14,747,670
Diferencia Real	1,845,645	445,339	Diferencia Real	-113,881	-130,263
% Diferencias	11.02%	2.89%	% Diferencias	-0.64%	-0.88%
Prom % Diferencias		6.96%	Prom % Diferencias		-0.76%
<b>CON 5 CERCANOS</b>			<b>CON 5 CERCANOS</b>		
Datos Estándar	8.27226	8.34208	Datos Estándar	8.82504	8.77704
Cifra Pronosticada	18,217,127	15,759,127	Cifra Pronosticada	17,477,651	14,497,203
Diferencia Real	1,476,535	356,267	Diferencia Real	-195,000	-380,730
% Diferencias	8.82%	2.31%	% Diferencias	-1.10%	-2.56%
Prom % Diferencias		5.57%	Prom % Diferencias		-1.83%

Tabla E.1c. Análisis del número adecuado de vecinos cercanos.

Análisis sumario de errores porcentualizados.

Análisis de Promdios de Errores		
Ventas	Almacen	Promedio
<b>CON 2 CERCANOS</b>		
-0.68%	0.01%	-0.33%
<b>CON 3 CERCANOS</b>		
0.02%	-0.61%	-0.30%
<b>CON 4 CERCANOS</b>		
-0.18%	-1.29%	-0.74%
<b>CON 5 CERCANOS</b>		
-0.34%	-1.89%	-1.12%
** Menor error encontrado con 3 cercanos		

Tabla E.2. Análisis sumario de errores porcentualizados.

Análisis de Lecturas cercanas independientes.

PRONOSTICO A: JUL 2009			PRONOSTICO A: MAR 2009		
	Ventas	Almacen		Ventas	Almacen
<b>LECTURAS INDEPENDIENTES</b>			<b>LECTURAS INDEPENDIENTES</b>		
<b>PRIMERA</b>			<b>PRIMERA</b>		
Datos Estándar	8.4397	7.356	Datos Estándar	8.3069	8.5875
Cifra Pronosticada	29,779,681	17,207,017	Cifra Pronosticada	29,311,093	20,087,719
Diferencia Real	1,316,826	-1,857,188	Diferencia Real	-991,094	-253,245
% Diferencias	4.63%	-9.74%	% Diferencias	-3.27%	-1.25%
Prom % Diferencias		-2.56%	Prom % Diferencias		-2.26%
<b>SEGUNDA</b>			<b>SEGUNDA</b>		
Datos Estándar	8.1212	8.582	Datos Estándar	8.3259	8.6036
Cifra Pronosticada	28,655,846	20,074,853	Cifra Pronosticada	29,378,135	20,125,380
Diferencia Real	192,991	1,010,648	Diferencia Real	-924,052	-215,584
% Diferencias	0.68%	5.30%	% Diferencias	-3.05%	-1.06%
Prom % Diferencias		2.99%	Prom % Diferencias		-2.05%
<b>TERCERA</b>			<b>TERCERA</b>		
Datos Estándar	8.3867	7.8153	Datos Estándar	8.4302	8.323
Cifra Pronosticada	29,592,669	18,281,403	Cifra Pronosticada	29,746,160	19,469,005
Diferencia Real	1,129,814	-782,802	Diferencia Real	-556,027	-871,959
% Diferencias	3.97%	-4.11%	% Diferencias	-1.83%	-4.29%
Prom % Diferencias		-0.07%	Prom % Diferencias		-3.06%

Tabla E.3. Análisis de lecturas de cercanos independientes.

## XVII. APÉNDICE F. CASO 1

### Caso 1.- Análisis de la aplicación del modelo con datos históricos

#### Resultados Pronóstico 2009/07

Ámbito definido:

- e) Datos de 2002/01-2009/06.
- f) Clústeres=6.
- g) Vecinos cercanos a considerar 3.
- h) Pronóstico: 2009/07.

Resultados:

\*\*\*\*\*

*Modelo de análisis prospectivo.*

*Datos de trabajo - vector: {7.6923,8.3602,9.0718} con 6 clústeres, y 3 vecinos cercanos.*

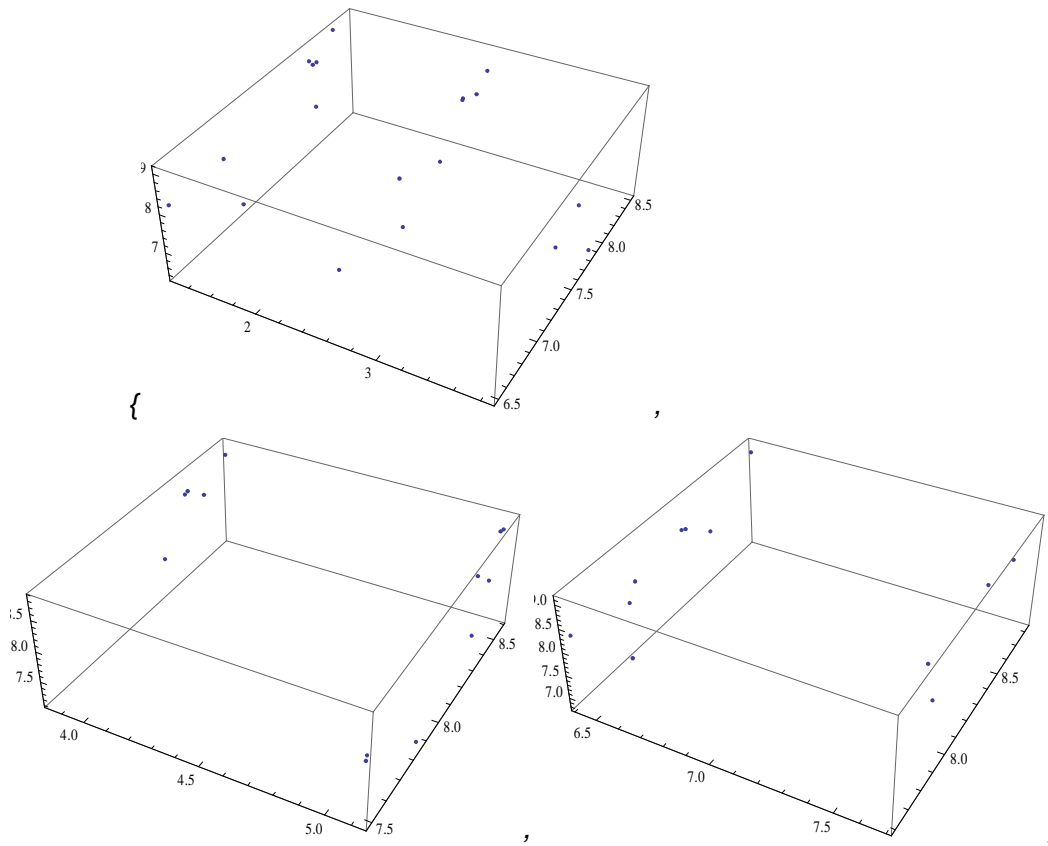
\*\*\*\*\*

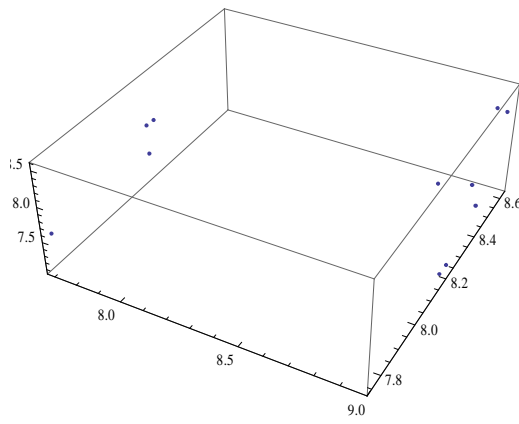
*Datos agrupados por clústeres en 6 clústeres.*

*{{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{3.8462,7.8743,6.3015},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{3.8462,7.6018,8.1535},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{3.8462,7.2818,7.8703},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}},{3.8462,8.3259,8.6036},{5.1282,8.4922,8.8759},{3.8462,8.3069,8.5875},{*

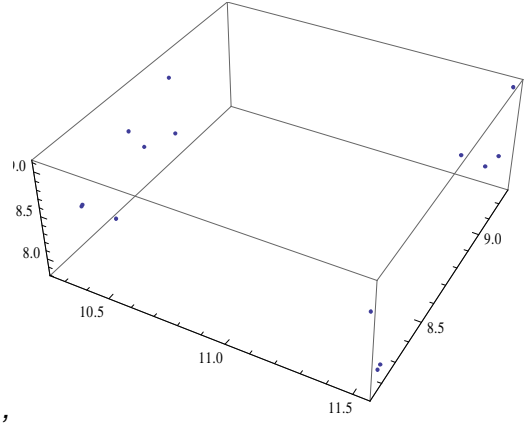
5.1282,8.523,8.8419},{3.8462,8.4302,8.3239},{5.1282,8.4307,8.2305},{5.1282,7.8717,7.1044},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086},{5.1282,7.4852,8.0734},{5.1282,7.4914,8.1391},{3.8462,8.5878,8.6958},{5.1282,8.3052,8.6099}},{6.4103,8.3216,8.4906},{6.4103,8.2944,8.5418},{6.4103,8.823,8.9579},{7.6923,8.6175,8.8463},{6.4103,7.9161,6.7662},{6.4103,7.9338,7.9663},{7.6923,7.8463,9.1419},{6.4103,7.5955,8.3312},{7.6923,7.9101,8.1832},{6.4103,7.9791,8.3059},{6.4103,8.4916,7.9838},{7.6923,8.3602,9.0718}},{7.6923,8.2182,7.9435},{8.9744,8.1986,7.0714},{7.6923,8.1833,7.9521},{8.9744,8.2447,7.0581},{8.9744,8.4397,7.356},{7.6923,7.7433,7.6053},{8.9744,8.5273,8.5307},{8.9744,8.1212,8.582},{8.9744,8.3867,7.8153},{7.6923,8.1848,7.527},{8.9744,8.6229,8.242}},{10.2564,8.2928,8.3092},{11.5385,8.2367,7.8971},{10.2564,8.2983,8.3182},{11.5385,8.2179,7.8823},{10.2564,8.465,7.8217},{11.5385,8.1433,8.7216},{10.2564,8.8442,8.8834},{11.5385,9.2705,9.0591},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6711},{11.5385,9.2005,8.3652},{10.2564,8.8613,8.1777},{11.5385,9.057,8.498}},{12.8205,8.6299,8.1376},{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{12.8205,8.6612,8.1705},{14.1026,8.7162,8.5712},{15.3846,8.3285,8.8961},{12.8205,8.1907,8.1715},{14.1026,8.3747,8.2187},{15.3846,7.7458,8.2673},{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},{14.1026,9.049,8.626},{15.3846,8.3135,8.8365},{12.8205,9.3864,8.2072},{14.1026,9.9879,8.8471},{15.3846,9.5073,8.9946},{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432}}

*Grafico tridimensional de los datos por cada clúster.*

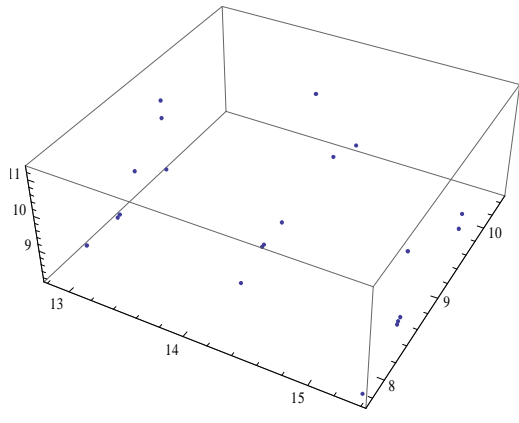




,

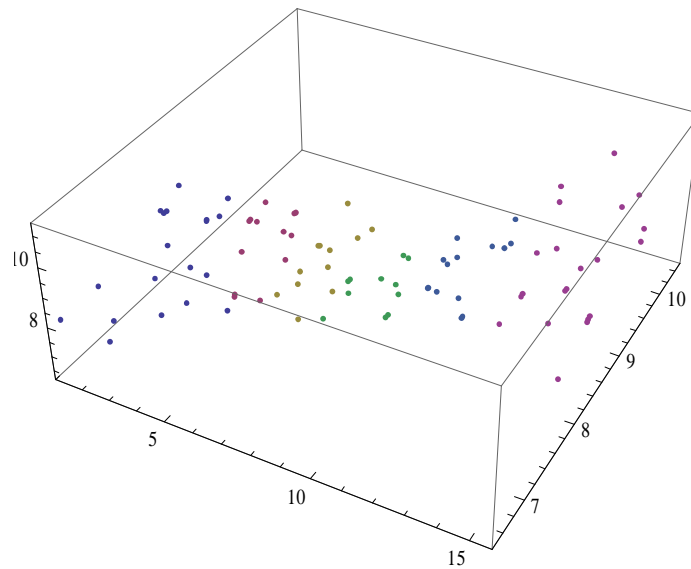


,



}

*Grafico de los datos agrupados.*



*Datos en la vecindad más cercana dentro del clúster a la posición original  $\{90\}$  :*  
 $\{(7.6923, 8.3602, 9.0718), (7.6923, 8.6175, 8.8463), (7.6923, 7.8463, 9.1419), (7.6923, 7.9101, 8.1832)\}$

*Posiciones en la vecindad actual cercana del vector:*

$\{30, 54, 66\}$

*Posiciones del estado N-1 o anterior:*

$\{29, 53, 65\}$

*Posiciones del estado N+1 o siguiente:*

$\{31, 55, 67\}$

*Datos del estado N+1 o siguiente:*

$\{(8.9744, 8.4397, 7.356), (8.9744, 8.1212, 8.582), (8.9744, 8.3867, 7.8153)\}$

*Resultado Prospectivo con cercanos igual a 3*

$\{8.9744, 8.31587, 7.91777\}$

## Pronostico resultante

PRONOSTICO A: JUL 2009		
	VENTAS	INVENTARIO
Datos Estándar	8.31587	7.91777
Cifra Pronosticada	29,342,744	18,521,099
Diferencia Real	879,889	-543,106
% Diferencias	3.09%	-2.85%
Promedio % Diferencias		0.12%

## Casos de eventos antecedentes

c) Estado N y N-1 de vecinos cercanos

Posición	Año	Mes	Tipo	Memoria de Casos
29	2004	5	EXR	Unión Europea: se integran Chipre, la República Checa, Eslovaquia, Eslovenia, Estonia, Hungría y tres mas
65	2007	5	EIO	Integración de Sistema Administrativo Kepler Acoplado

d) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
67	2007	7	EIO	Integración de Sistema de producción al sistema kepler



## Resultados Pronóstico 2009/08

Ámbito definido:

- a) Datos de 2002/01-2009/07.
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2009/08.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

Datos de trabajo - vector: {8.9744,8.0665,8.1499} con 6 clusters, y 3 vecinos cercanos.

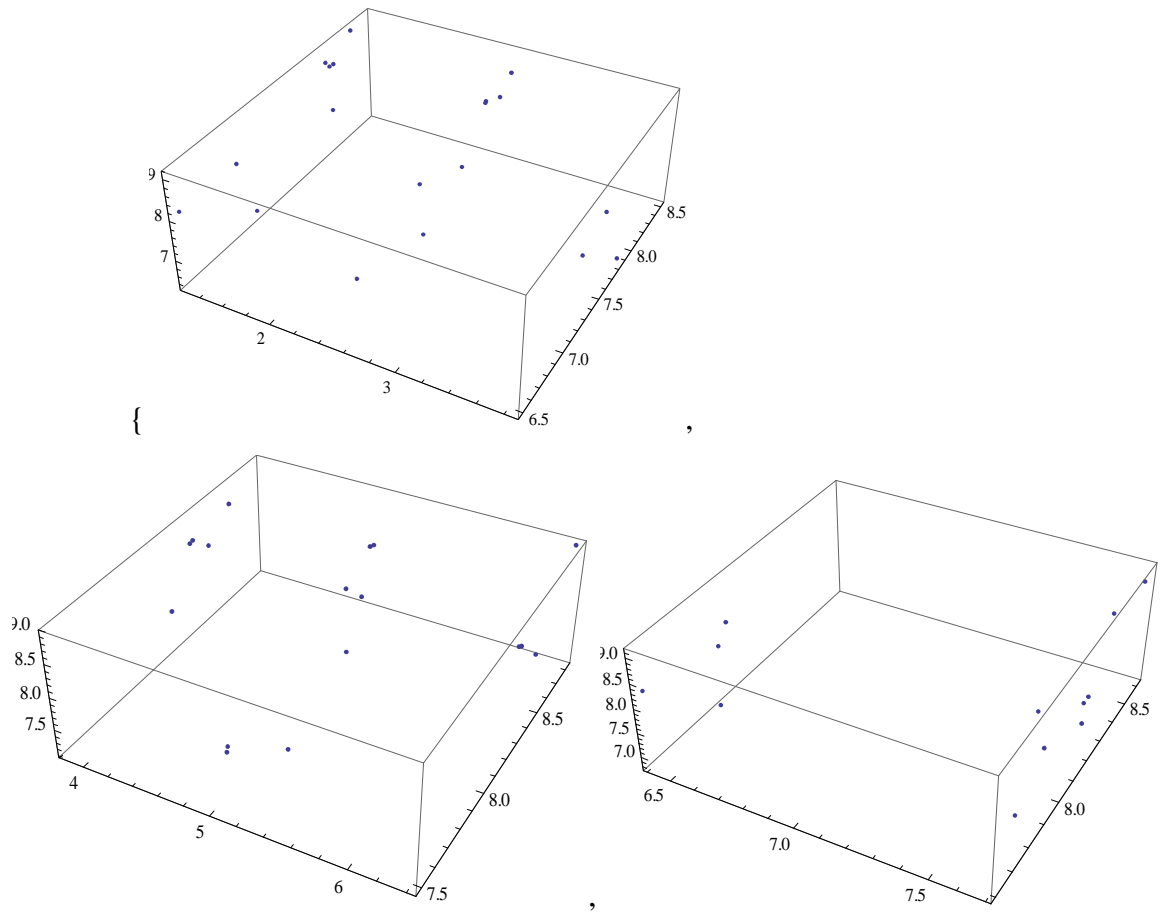
\*\*\*\*\*

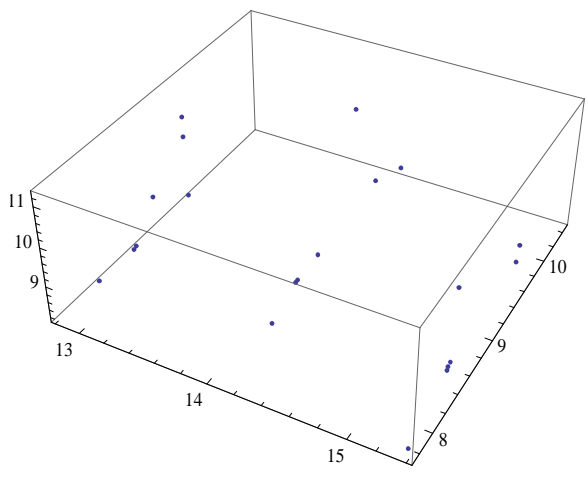
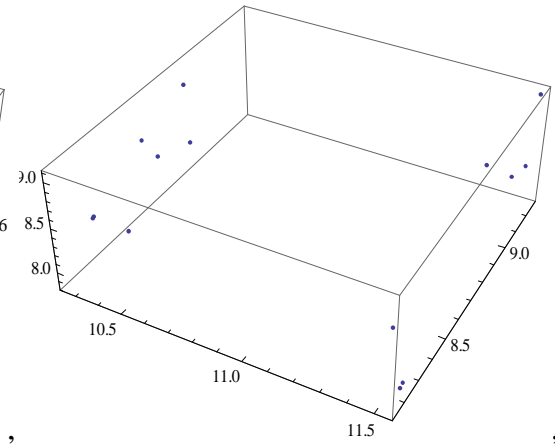
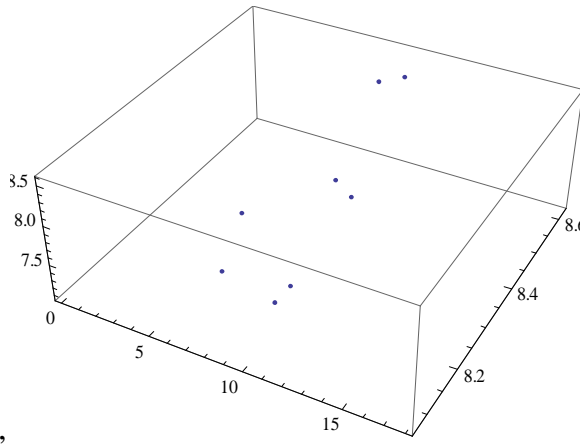
Datos agrupados por clusters en 6 clusters.

{{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{3.8462,7.8743,6.3015},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{3.8462,7.6018,8.1535},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{3.8462,7.2818,7.8703},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}},{3.8462,8.3259,8.6036},{5.1282,8.4922,8.8759},{6.4103,8.3216,8.4906},{3.8462,8.3069,8.5875},{5.1282,8.523,8.8419},{6.4103,8.2944,8.5418},{3.8462,8.4302,8.3239},{5.1282,8.4307,8.2305},{6.4103,8.823,8.9579},{5.1282,7.8717,7.1044},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086},{5.1282,7.4852,8.0734},{5.1282,7.4914,8.1391},{3.8462,8.5878,8.6958},{5.1282,8.3052,8.6099},{6.4103,8.4916,7.9838}},{7.6923,8.2182,7.9435},{7.6923,8.1833,7.9521},{7.6923,8.6175,8.8463},{6.4103,7.9161,6.7662},{7.6923,7.7433,7.6053},{6.4103,7.9338,7.9663},{7.6923,7.8463,9.1419},{6.41

03,7.5955,8.3312},{7.6923,7.9101,8.1832},{6.4103,7.9791,8.3059},{7.6923,8.1848,7.527},  
{7.6923,8.3602,9.0718}},{{8.9744,8.1986,7.0714},{8.9744,8.2447,7.0581},{8.9744,8.4397,7.356},  
{8.9744,8.5273,8.5307},{8.9744,8.1212,8.582},{8.9744,8.3867,7.8153},{8.9744,8.6229,8.242},  
{8.9744,8.0665,8.1499}},{{10.2564,8.2928,8.3092},{11.5385,8.2367,7.8971},{10.2564,8.2983,8.3182},  
{11.5385,8.2179,7.8823},{10.2564,8.465,7.8217},{11.5385,8.1433,8.7216},{10.2564,8.8442,8.8834},  
{11.5385,9.2705,9.0591},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6711},  
{11.5385,9.2005,8.3652},{10.2564,8.8613,8.1777},{11.5385,9.057,8.498}},{{12.8205,8.6299,8.1376},  
{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{12.8205,8.6612,8.1705},{14.1026,8.7162,8.5712},  
{15.3846,8.3285,8.8961},{12.8205,8.1907,8.1715},{14.1026,8.3747,8.2187},{15.3846,7.7458,8.2673},  
{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},  
{14.1026,9.049,8.626},{15.3846,8.3135,8.8365},{12.8205,9.3864,8.2072},{14.1026,9.9879,8.8471},  
{15.3846,9.5073,8.9946},{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432}}}

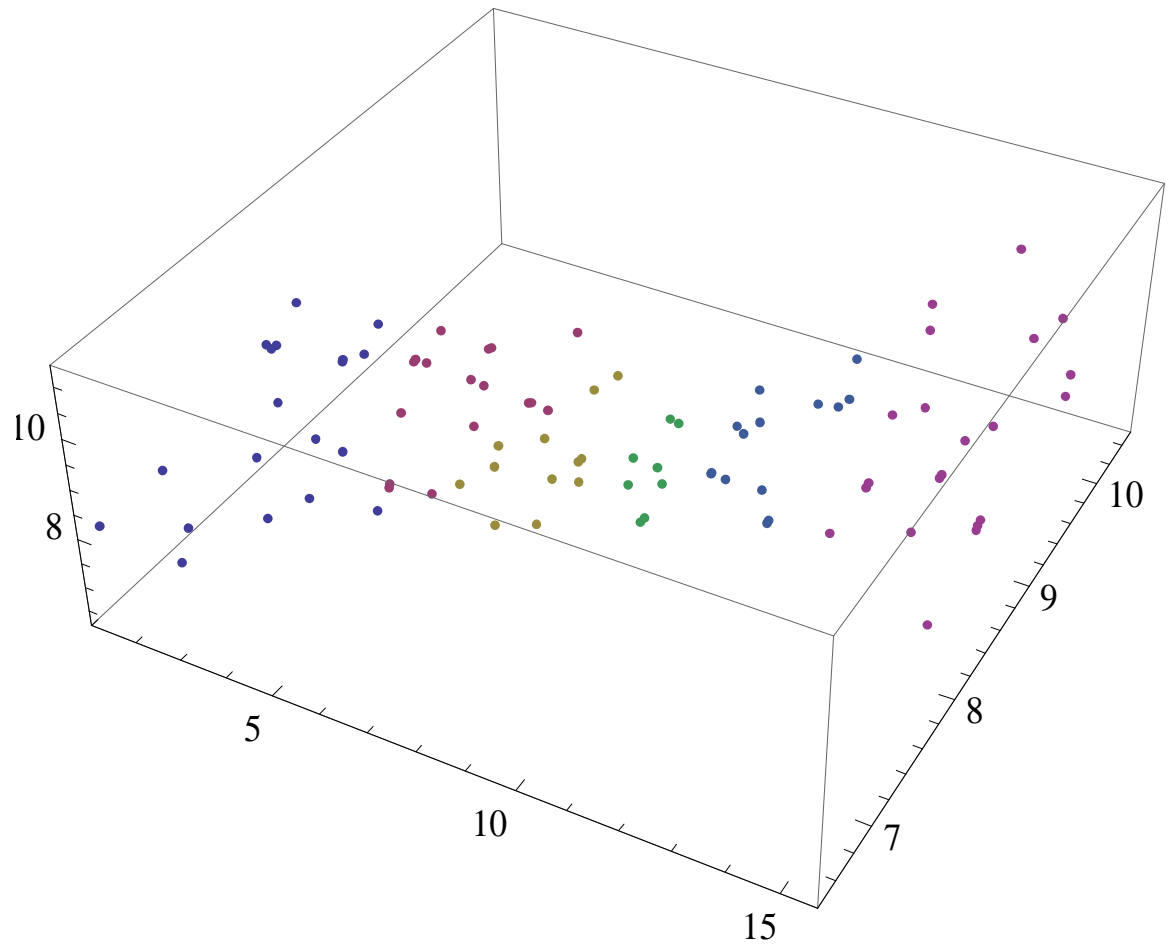
Grafico tridimensional de los datos por cada cluster.





}

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{91\}$  :  
 $\{8.9744, 8.0665, 8.1499\}, \{8.9744, 8.1212, 8.582\}, \{8.9744, 8.3867, 7.8153\}, \{8.9744, 8.6229, 8.242\}$

Posiciones en la vecindad actual cercana del vector:

$\{55, 67, 79\}$

Posiciones del estado N-1 o anterior:

$\{54, 66, 78\}$

Posiciones del estado N+1 o siguiente:

{56,68,80}

Datos del estado N+1 o siguiente:

{{10.2564,8.0646,8.3465},{10.2564,8.5841,8.6711},{10.2564,8.8613,8.1777}}

Resultado Prospectivo con cercanos igual a 3

{10.2564,8.50333,8.39843}

### Pronostico resultante

PRONOSTICO A:AGO 2009		
	VENTAS	INVENTARIO
Datos Estándar	8.50333	8.39843
Cifra Pronosticada	30,004,201	19,645,450
Diferencia Real	1,368,892	542,466
% Diferencias	4.78%	2.84%
Promedio % Diferencias		3.81%

### Casos de eventos antecedentes

#### a) Estado N y N-1 de vecinos cercanos

Posición	Año	Mes	Tipo	Memoria de Casos
67	2007	7	EIO	Integración de Sistema de producción al sistema kepler
78	2008	6	EIO	Lanzamiento de barnices sobre polímeros

#### b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
70	2007	10	EXR	España entra en recesión

## Resultados Pronóstico 2009/09

Ámbito definido:

- a) Datos de 2002/01-2009/08.
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2009/09.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

Datos de trabajo - vector: {10.2564,8.1154,8.1665} con 6 clusters, y 3 vecinos cercanos.

\*\*\*\*\*

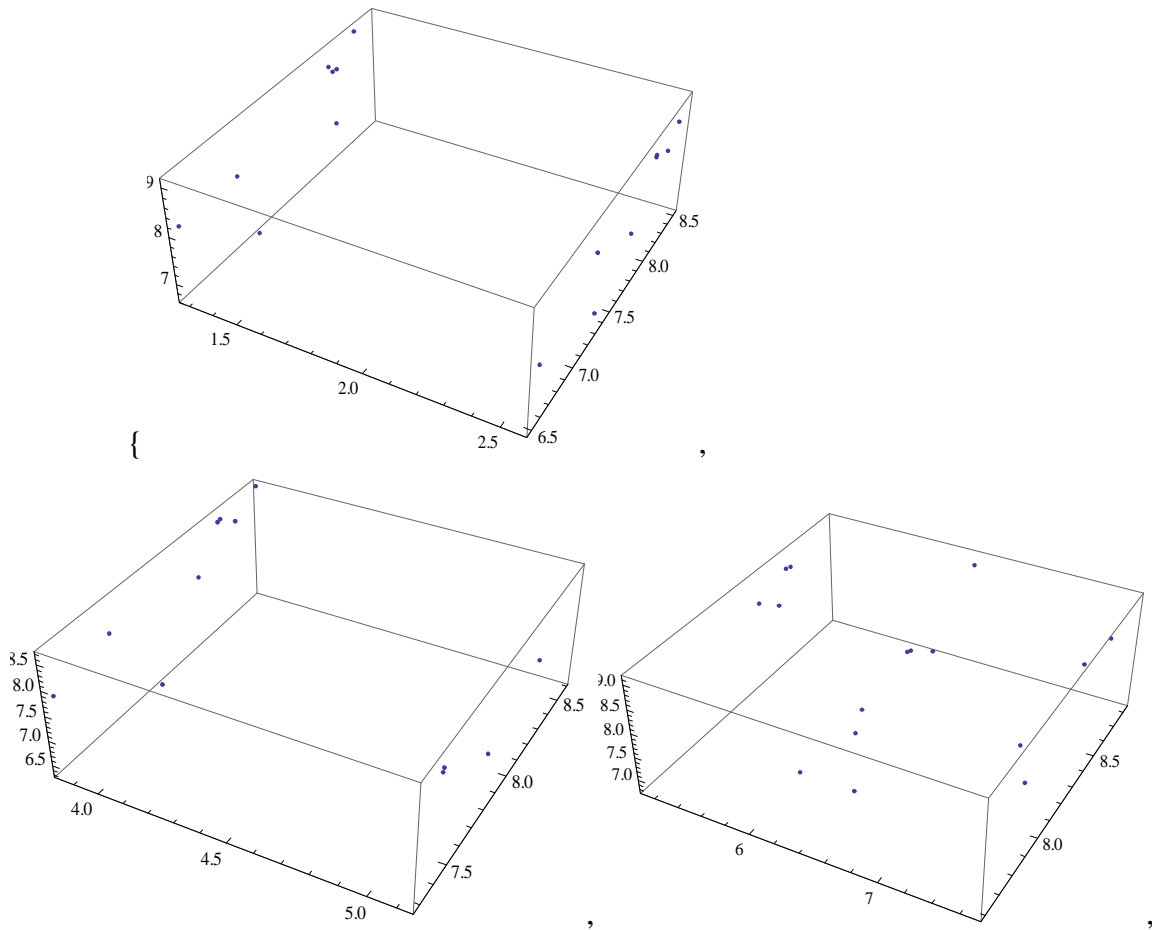
Datos agrupados por clusters en 6 clusters.

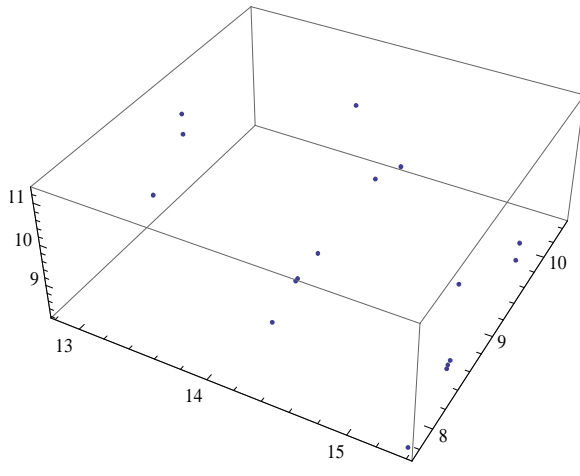
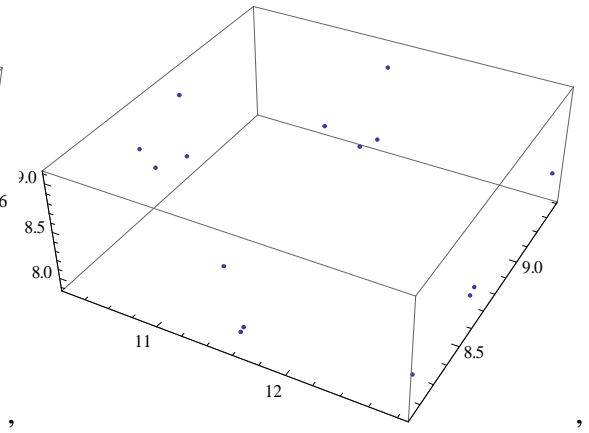
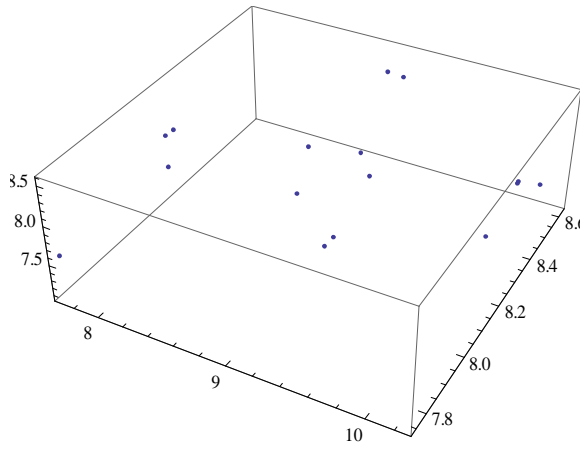
{{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}},{3.8462,8.3259,8.6036},{3.8462,8.3069,8.5875},{3.8462,8.4302,8.3239},{3.8462,7.8743,6.3015},{5.1282,7.8717,7.1044},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086},{3.8462,7.6018,8.1535},{5.1282,7.4852,8.0734},{3.8462,7.2818,7.8703},{5.1282,7.4914,8.1391},{3.8462,8.5878,8.6958}},{5.1282,8.4922,8.8759},{6.4103,8.3216,8.4906},{5.1282,8.523,8.8419},{6.4103,8.2944,8.5418},{5.1282,8.4307,8.2305},{6.4103,8.823,8.9579},{7.6923,8.6175,8.8463},{6.4103,7.9161,6.7662},{6.4103,7.9338,7.9663},{7.6923,7.8463,9.1419},{6.4103,7.5955,8.3312},{7.6923,7.9101,8.1832},{6.4103,7.9791,8.3059},{5.1282,8.3052,8.6099},{6.4103,8.4916,7.9838},{7.69

23,8.3602,9.0718}},{{7.6923,8.2182,7.9435},{8.9744,8.1986,7.0714},{10.2564,8.2928,8.3092},{7.6923,8.1833,7.9521},{8.9744,8.2447,7.0581},{10.2564,8.2983,8.3182},{8.9744,8.4397,7.356},{10.2564,8.465,7.8217},{7.6923,7.7433,7.6053},{8.9744,8.5273,8.5307},{8.9744,8.1212,8.582},{8.9744,8.3867,7.8153},{7.6923,8.1848,7.527},{8.9744,8.6229,8.242},{8.9744,8.0665,8.1499},{10.2564,8.1154,8.1665}},{{11.5385,8.2367,7.8971},{12.8205,8.6299,8.1376},{11.5385,8.2179,7.8823},{12.8205,8.6612,8.1705},{11.5385,8.1433,8.7216},{12.8205,8.1907,8.1715},{10.2564,8.8442,8.8834},{11.5385,9.2705,9.0591},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6711},{11.5385,9.2005,8.3652},{12.8205,9.3864,8.2072},{10.2564,8.8613,8.1777},{11.5385,9.057,8.498}},{{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{14.1026,8.7162,8.5712},{15.3846,8.3285,8.8961},{14.1026,8.3747,8.2187},{15.3846,7.7458,8.2673},{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},{14.1026,9.049,8.626},{15.3846,8.3135,8.8365},{14.1026,9.9879,8.8471},{15.3846,9.5073,8.9946},{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432}}}



Grafico tridimensional de los datos por cada cluster.

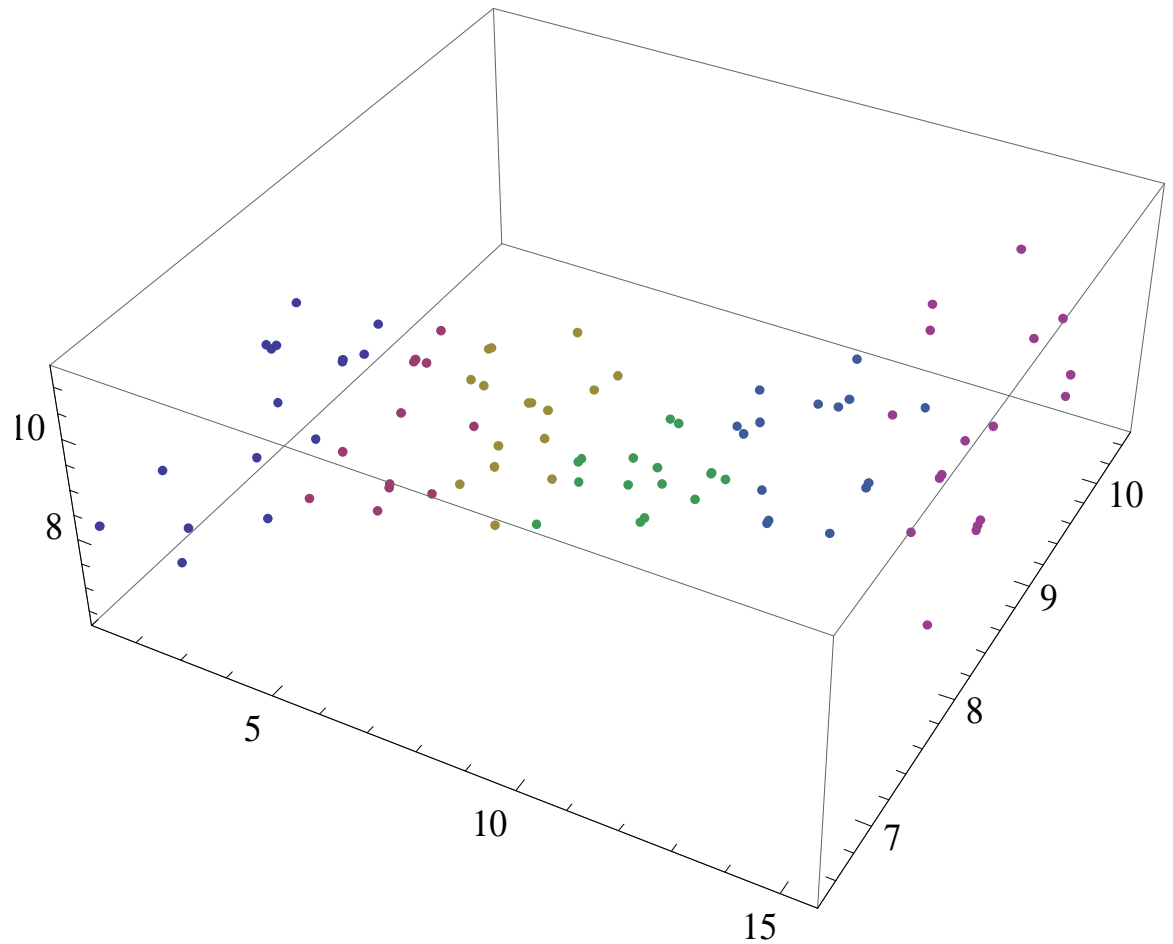




,

}

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{\{92\}\}$  :  
 $\{\{10.2564, 8.1154, 8.1665\}, \{10.2564, 8.2928, 8.3092\}, \{10.2564, 8.2983, 8.3182\}, \{10.2564, 8.465, 7.8217\}\}$

Posiciones en la vecindad actual cercana del vector:

$\{8, 20, 32\}$

Posiciones del estado N-1 o anterior:

$\{7, 19, 31\}$

Posiciones del estado N+1 o siguiente:

{9,21,33}

Datos del estado N+1 o siguiente:

{{11.5385,8.2367,7.8971},{11.5385,8.2179,7.8823},{11.5385,8.1433,8.7216}}

Resultado Prospectivo con cercanos igual a 3

{11.5385,8.1993,8.167}

### Pronóstico resultante

PRONOSTICO A:SEP 2009		
	VENTAS	INVENTARIO
Datos Estándar	8.1993	8.167
Cifra Pronosticada	28,931,424	19,104,093
Diferencia Real	-236,415	53,560
% Diferencias	-0.81%	0.28%
Promedio % Diferencias		-0.26%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

## Resultados Pronóstico 2009/10

Ámbito definido:

- a) Datos de 2002/01-2009/09.
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2009/10.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

Datos de trabajo - vector: {11.5385,8.2663,8.1441} con 6 clusters, y 3 vecinos cercanos.

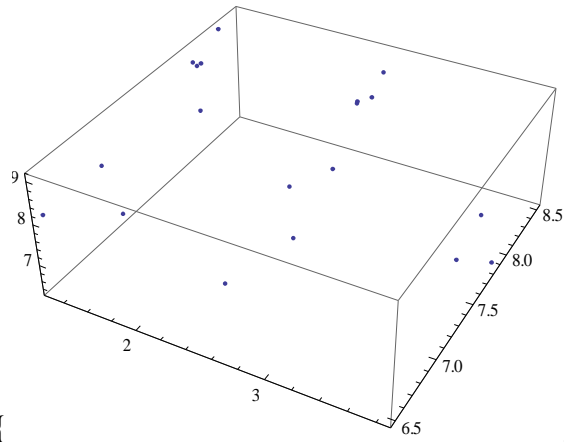
\*\*\*\*\*

Datos agrupados por clusters en 6 clusters.

{{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{3.8462,7.8743,6.3015},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{3.8462,7.6018,8.1535},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{3.8462,7.2818,7.8703},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}},{3.8462,8.3259,8.6036},{5.1282,8.4922,8.8759},{3.8462,8.3069,8.5875},{5.1282,8.523,8.8419},{3.8462,8.4302,8.3239},{5.1282,8.4307,8.2305},{5.1282,7.8717,7.1044},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086},{5.1282,7.4852,8.0734},{5.1282,7.4914,8.1391},{3.8462,8.5878,8.6958},{5.1282,8.3052,8.6099}},{6.4103,8.3216,8.4906},{6.4103,8.2944,8.5418},{6.4103,8.823,8.9579},{7.6923,8.6175,8.8463},{6.4103,7.9161,6.7662},{6.4103,7.9338,7.9663},{7.6923,7.8463,9.1419},{6.4103,7.5955,8.3312},{7.6923,7.9101,8.1832},{6.4103,7.9791,8.3059},{6.4103,8.4916,7.9838},{7.69

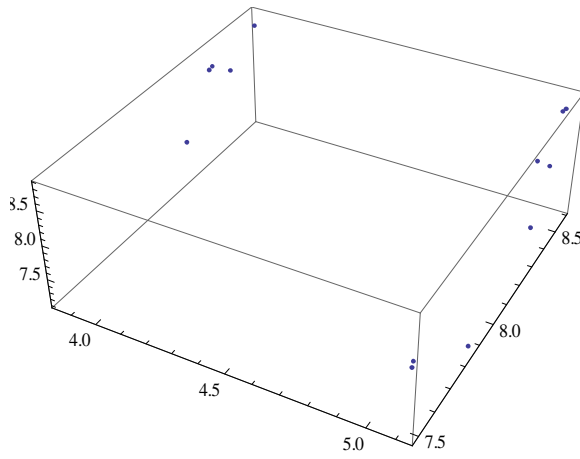
23,8.3602,9.0718}},{{7.6923,8.2182,7.9435},{8.9744,8.1986,7.0714},{10.2564,8.2928,8.3092},{7.6923,8.1833,7.9521},{8.9744,8.2447,7.0581},{10.2564,8.2983,8.3182},{8.9744,8.4397,7.356},{10.2564,8.465,7.8217},{7.6923,7.7433,7.6053},{8.9744,8.5273,8.5307},{8.9744,8.1212,8.582},{8.9744,8.3867,7.8153},{7.6923,8.1848,7.527},{8.9744,8.6229,8.242},{8.9744,8.0665,8.1499},{10.2564,8.1154,8.1665}},{{11.5385,8.2367,7.8971},{12.8205,8.6299,8.1376},{11.5385,8.2179,7.8823},{12.8205,8.6612,8.1705},{11.5385,8.1433,8.7216},{12.8205,8.1907,8.1715},{10.2564,8.8442,8.8834},{11.5385,9.2705,9.0591},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6711},{11.5385,9.2005,8.3652},{12.8205,9.3864,8.2072},{10.2564,8.8613,8.1777},{11.5385,9.057,8.498},{11.5385,8.2663,8.1441}},{{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{14.1026,8.7162,8.5712},{15.3846,8.3285,8.8961},{14.1026,8.3747,8.2187},{15.3846,7.7458,8.2673},{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},{14.1026,9.049,8.626},{15.3846,8.3135,8.8365},{14.1026,9.9879,8.8471},{15.3846,9.5073,8.9946},{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432}}}

Grafico tridimensional de los datos por cada clúster.

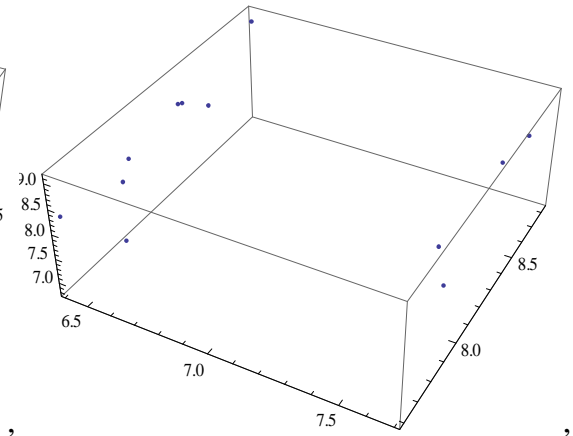


{

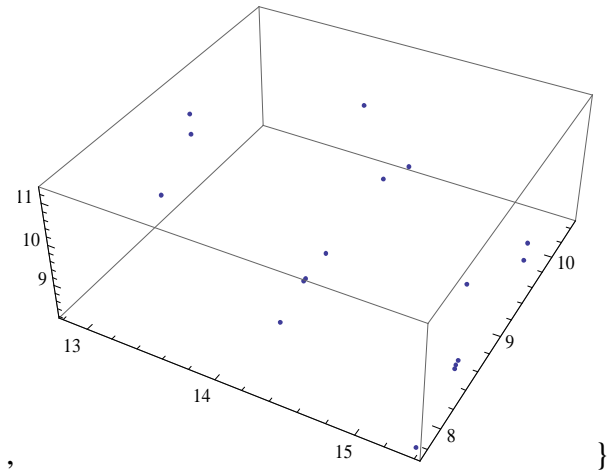
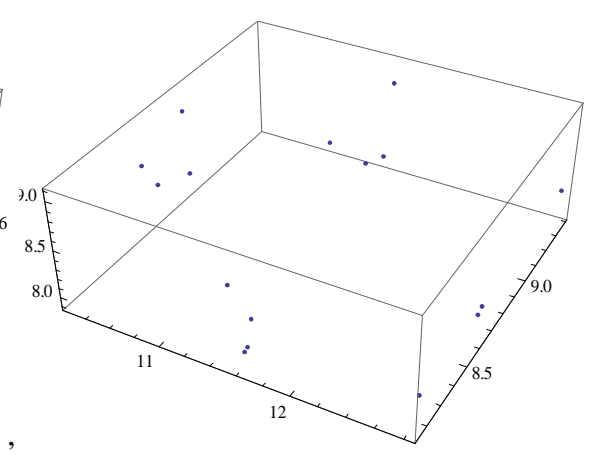
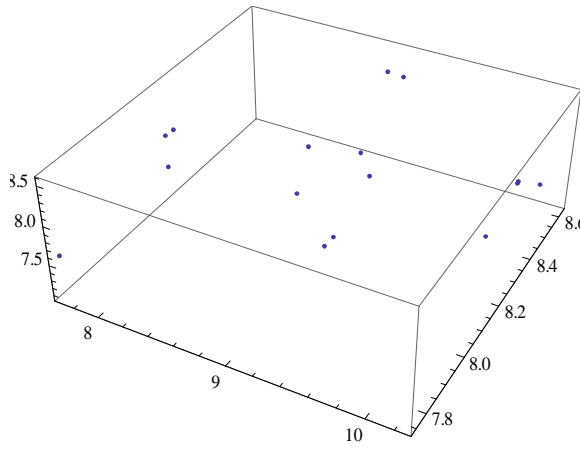
,



,



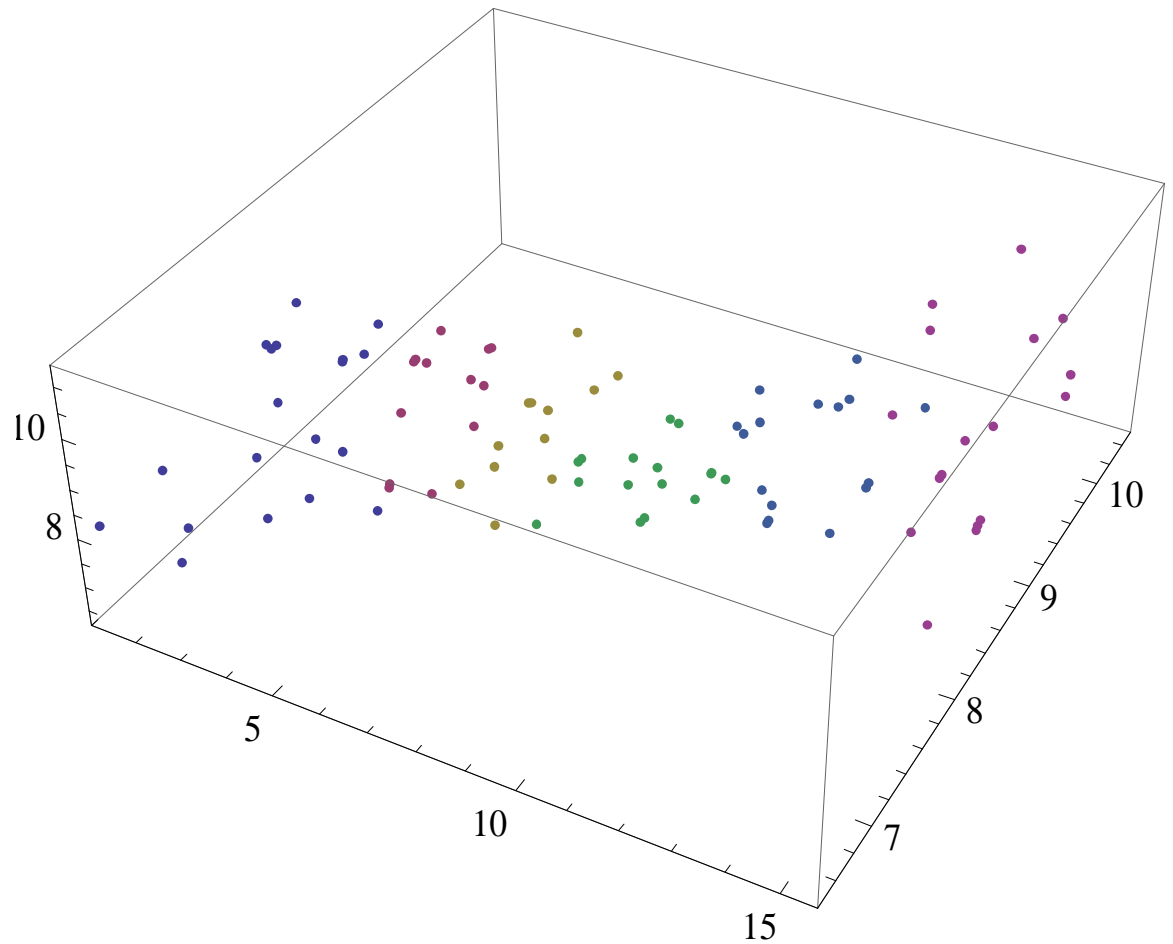
,



,  
 },



Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{\{93\}\}$  :  
 $\{\{11.5385, 8.2663, 8.1441\}, \{11.5385, 8.2367, 7.8971\}, \{11.5385, 8.2179, 7.8823\}, \{11.5385, 8.1433, 8.7216\}\}$

Posiciones en la vecindad actual cercana del vector:

$\{9, 21, 33\}$

Posiciones del estado N-1 o anterior:

$\{8, 20, 32\}$

Posiciones del estado N+1 o siguiente:

{10,22,34}

Datos del estado N+1 o siguiente:

{{12.8205,8.6299,8.1376},{12.8205,8.6612,8.1705},{12.8205,8.1907,8.1715}}

Resultado Prospectivo con cercanos igual a 3

{12.8205,8.49393,8.15987}

### Pronóstico resultante

PRONOSTICO A:OCT 2009		
	VENTAS	INVENTARIO
Datos Estándar	8.49393	8.15987
Cifra Pronosticada	29,971,033	19,087,415
Diferencia Real	14,076	-478,221
% Diferencias	0.05%	-2.44%
Promedio % Diferencias		-1.20%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

## Resultados Pronóstico 2009/11

Ámbito definido:

- a) Datos de 2002/01-2009/10.
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2009/11.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

Datos de trabajo - vector: {12.8205,8.4899,8.3643} con 6 clusters, y 3 vecinos cercanos.

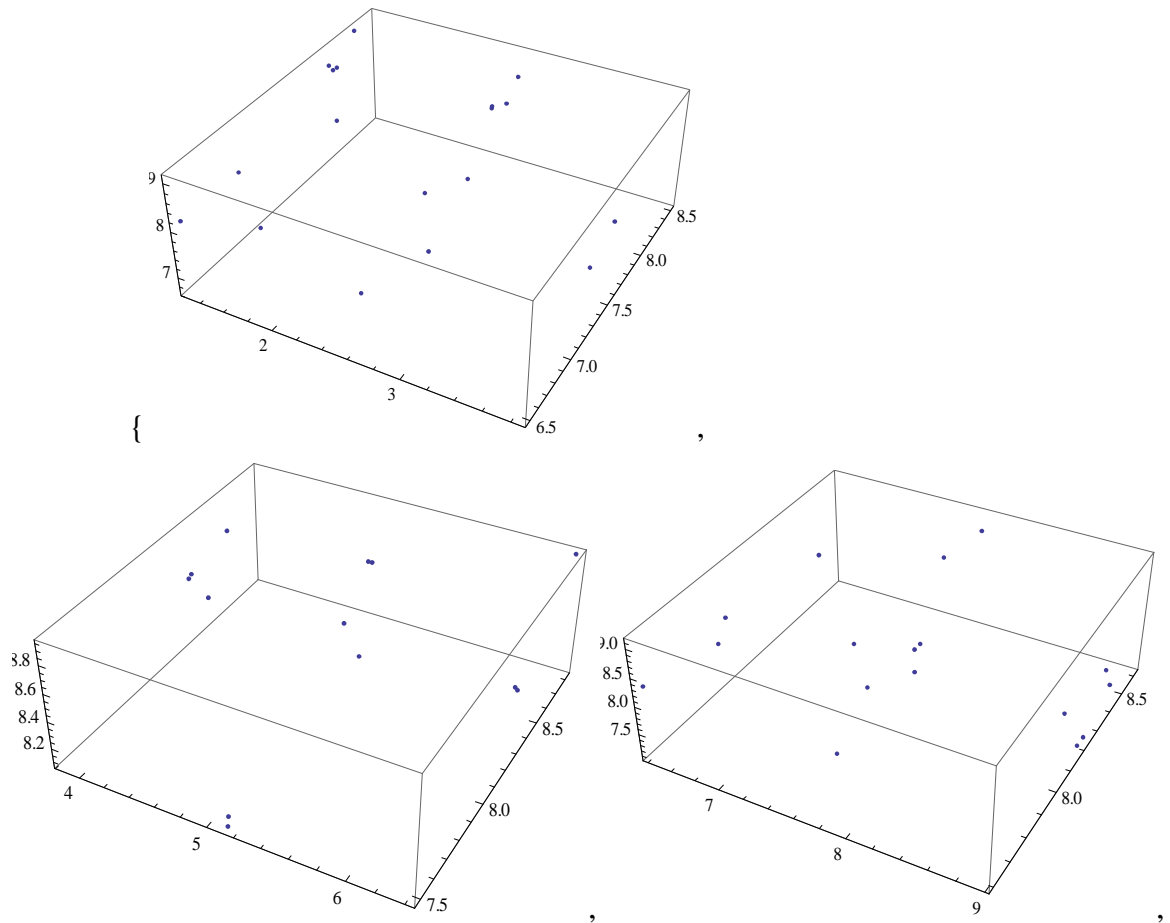
\*\*\*\*\*

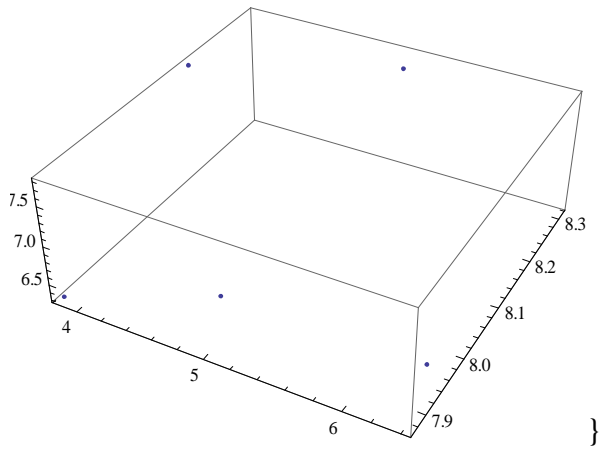
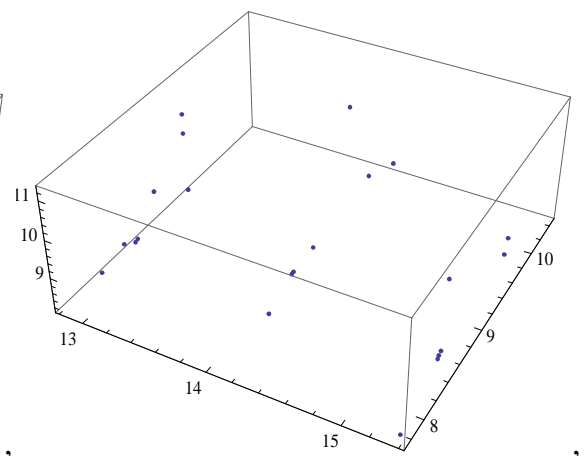
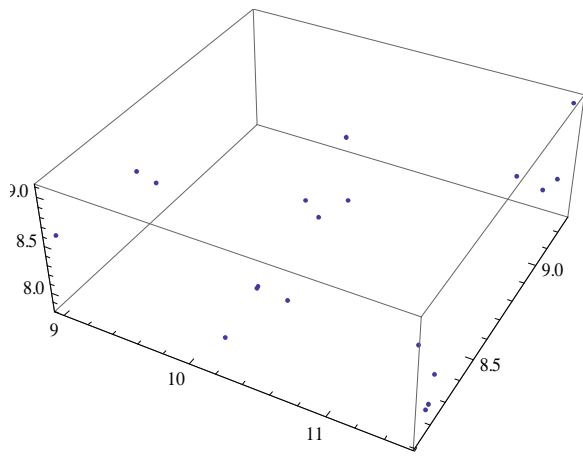
Datos agrupados por clusters en 6 clusters.

{{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{3.8462,7.6018,8.1535},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{3.8462,7.2818,7.8703},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}}},  
{{{3.8462,8.3259,8.6036},{5.1282,8.4922,8.8759},{6.4103,8.3216,8.4906},{3.8462,8.3069,8.5875},{5.1282,8.523,8.8419},{6.4103,8.2944,8.5418},{3.8462,8.4302,8.3239},{5.1282,8.4307,8.2305},{6.4103,8.823,8.9579},{5.1282,7.4852,8.0734},{5.1282,7.4914,8.1391},{3.8462,8.5878,8.6958},{5.1282,8.3052,8.6099}}},  
{{{7.6923,8.2182,7.9435},{8.9744,8.1986,7.0714},{7.6923,8.1833,7.9521},{8.9744,8.2447,7.0581},{7.6923,8.6175,8.8463},{8.9744,8.4397,7.356},{7.6923,7.7433,7.6053},{6.4103,7.9338,7.9663},{7.6923,7.8463,9.1419},{6.4103,7.5955,8.3312},{7.6923,7.9101,8.1832},{8.9744,8.3867,7.8153},{6.410

3,7.9791,8.3059},{7.6923,8.1848,7.527},{6.4103,8.4916,7.9838},{7.6923,8.3602,9.0718  
{8.9744,8.0665,8.1499}},{10.2564,8.2928,8.3092},{11.5385,8.2367,7.8971},{10.256  
4,8.2983,8.3182},{11.5385,8.2179,7.8823},{10.2564,8.465,7.8217},{11.5385,8.1433,8.7  
216},{8.9744,8.5273,8.5307},{10.2564,8.8442,8.8834},{11.5385,9.2705,9.0591},{8.974  
4,8.1212,8.582},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6  
711},{11.5385,9.2005,8.3652},{8.9744,8.6229,8.242},{10.2564,8.8613,8.1777},{11.538  
5,9.057,8.498},{10.2564,8.1154,8.1665},{11.5385,8.2663,8.1441}},{12.8205,8.6299,8.  
1376},{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{12.8205,8.6612,8.1705},{14.1  
026,8.7162,8.5712},{15.3846,8.3285,8.8961},{12.8205,8.1907,8.1715},{14.1026,8.3747,  
8.2187},{15.3846,7.7458,8.2673},{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},  
{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},{14.1026,9.049,8.626},{15.3846,8.3  
135,8.8365},{12.8205,9.3864,8.2072},{14.1026,9.9879,8.8471},{15.3846,9.5073,8.9946  
{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432},{12.820  
5,8.4899,8.3643}},{3.8462,7.8743,6.3015},{5.1282,7.8717,7.1044},{6.4103,7.9161,6.7  
662},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086}}

Grafico tridimensional de los datos por cada cluster.



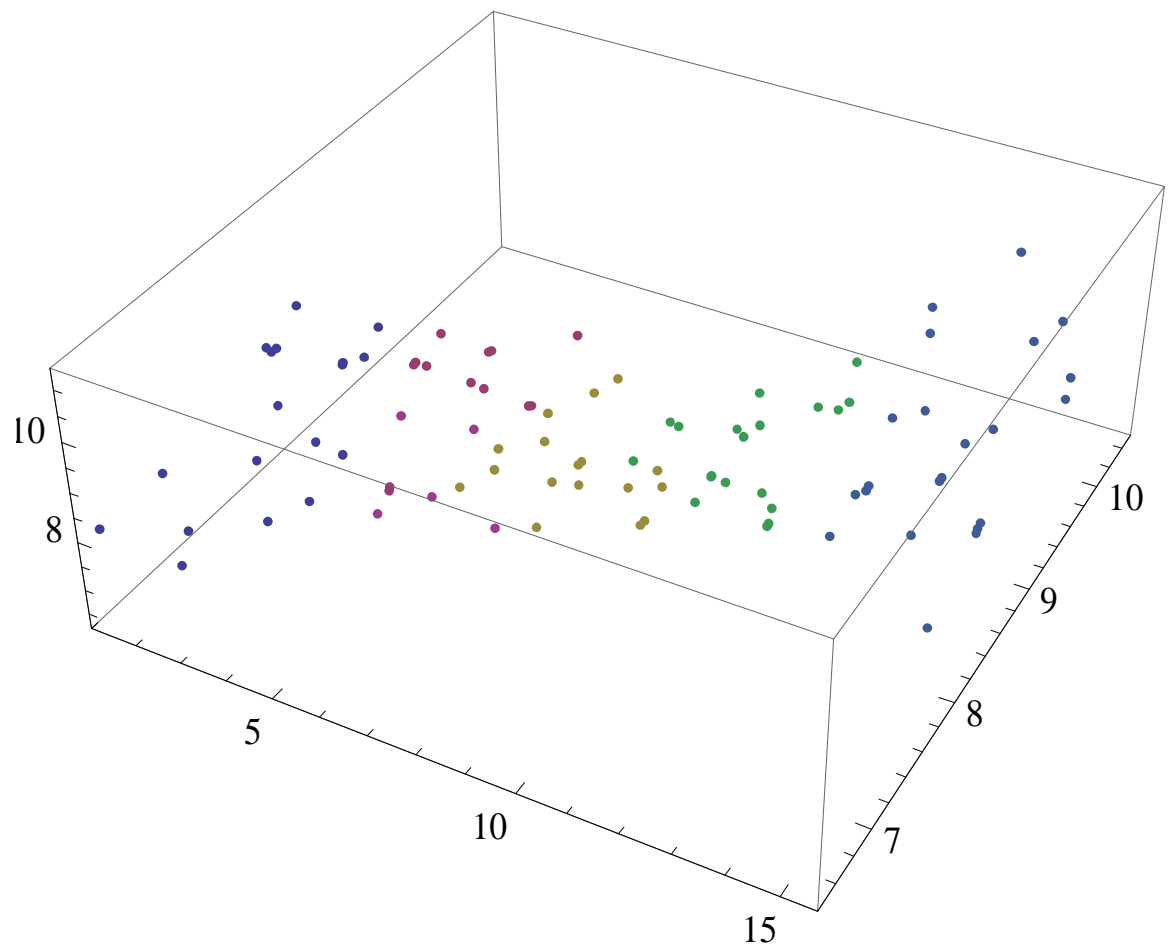


,

,

}

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{\{94\}\}$  :  
 $\{\{12.8205, 8.4899, 8.3643\}, \{12.8205, 8.6612, 8.1705\}, \{12.8205, 8.6299, 8.1376\}, \{12.8205, 8.1907, 8.1715\}\}$

Posiciones en la vecindad actual cercana del vector:

$\{22, 10, 34\}$

Posiciones del estado N-1 o anterior:

$\{21, 9, 33\}$

Posiciones del estado N+1 o siguiente:

{23,11,35}

Datos del estado N+1 o siguiente:

{{14.1026,8.7162,8.5712},{14.1026,8.7448,8.5788},{14.1026,8.3747,8.2187}}

Resultado Prospectivo con cercanos igual a 3

{14.1026,8.6119,8.45623}

### Pronóstico resultante

PRONOSTICO A:NOV 2009		
	VENTAS	INVENTARIO
Datos Estándar	8.6119	8.45623
Cifra Pronosticada	30,387,292	19,780,655
Diferencia Real	495,069	2,331,195
% Diferencias	1.66%	13.36%
Promedio % Diferencias		7.51%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS



## Resultados Pronóstico 2009/12

Ámbito definido:

- a) Datos de 2002/01-2009/11.
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2009/12.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

Datos de trabajo - vector: {14.1026,8.4716,7.4596} con 6 clusters, y 3 vecinos cercanos.

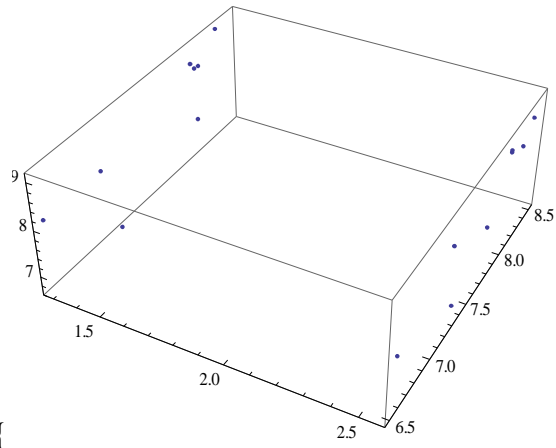
\*\*\*\*\*

Datos agrupados por clusters en 6 clusters.

{{{1.2821,8.0176,8.6622},{2.5641,8.1577,8.4998},{1.2821,8.0626,8.646},{2.5641,8.1631,8.5342},{1.2821,7.9761,8.8288},{2.5641,8.3633,8.2559},{1.2821,7.1569,6.6511},{2.5641,7.3798,6.7989},{1.2821,8.0276,7.4999},{2.5641,7.8596,7.4693},{1.2821,7.0135,8.1863},{2.5641,7.3411,8.1718},{1.2821,6.4884,8.1916},{2.5641,6.6532,7.5065},{1.2821,8.2746,9.133},{2.5641,8.5156,8.5972}},{3.8462,8.3259,8.6036},{5.1282,8.4922,8.8759},{3.8462,8.3069,8.5875},{5.1282,8.523,8.8419},{3.8462,8.4302,8.3239},{5.1282,8.4307,8.2305},{3.8462,7.8743,6.3015},{3.8462,8.1575,7.8139},{5.1282,8.3136,7.6086},{3.8462,7.6018,8.1535},{3.8462,7.2818,7.8703},{3.8462,8.5878,8.6958},{5.1282,8.3052,8.6099}},{6.4103,8.3216,8.4906},{6.4103,8.2944,8.5418},{6.4103,8.823,8.9579},{7.6923,8.6175,8.8463},{5.1282,7.8717,7.1044},{6.4103,7.9161,6.7662},{6.4103,7.9338,7.9663},{7.6923,7.8463,9.1419},{5.1282,7.4852,8.0734},{6.4103,7.5955,8.3312},{7.6923,7.9101,8.1832},{5.1282,7.4914,8.1391},{6.4103,7.9791,8.3059},{6.4103,8.4916,7.9838},{7.6923,7.9101,8.1832}}

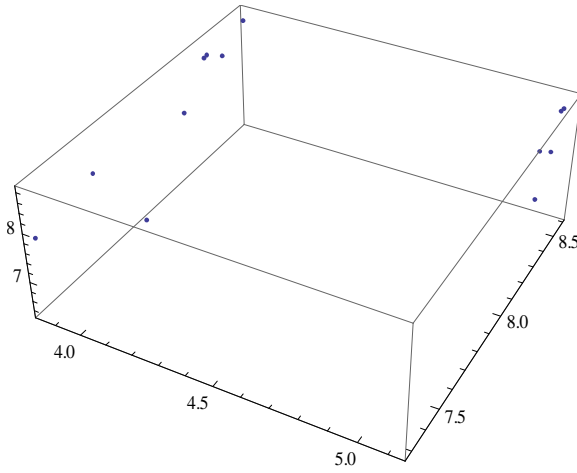
3,8.3602,9.0718}},{{7.6923,8.2182,7.9435},{8.9744,8.1986,7.0714},{7.6923,8.1833,7.9521},{8.9744,8.2447,7.0581},{8.9744,8.4397,7.356},{7.6923,7.7433,7.6053},{8.9744,8.5273,8.5307},{8.9744,8.1212,8.582},{8.9744,8.3867,7.8153},{7.6923,8.1848,7.527},{8.9744,8.6229,8.242},{8.9744,8.0665,8.1499}},{{10.2564,8.2928,8.3092},{11.5385,8.2367,7.8971},{10.2564,8.2983,8.3182},{11.5385,8.2179,7.8823},{10.2564,8.465,7.8217},{11.5385,8.1433,8.7216},{10.2564,8.8442,8.8834},{11.5385,9.2705,9.0591},{10.2564,8.6646,8.3465},{11.5385,8.7897,9.1015},{10.2564,8.5841,8.6711},{11.5385,9.2005,8.3652},{10.2564,8.8613,8.1777},{11.5385,9.057,8.498},{10.2564,8.1154,8.1665},{11.5385,8.2663,8.1441}},{{12.8205,8.6299,8.1376},{14.1026,8.7448,8.5788},{15.3846,8.364,8.9303},{12.8205,8.6612,8.1705},{14.1026,8.7162,8.5712},{15.3846,8.3285,8.8961},{12.8205,8.1907,8.1715},{14.1026,8.3747,8.2187},{15.3846,7.7458,8.2673},{12.8205,9.3729,10.3465},{14.1026,9.6282,11.3127},{15.3846,8.4147,10.6402},{12.8205,8.9235,9.0075},{14.1026,9.049,8.626},{15.3846,8.3135,8.8365},{12.8205,9.3864,8.2072},{14.1026,9.9879,8.8471},{15.3846,9.5073,8.9946},{12.8205,9.3688,9.8331},{14.1026,10.4584,8.3655},{15.3846,9.5529,9.3432},{12.8205,8.4899,8.3643},{14.1026,8.4716,7.4596}}}

Grafico tridimensional de los datos por cada cluster.

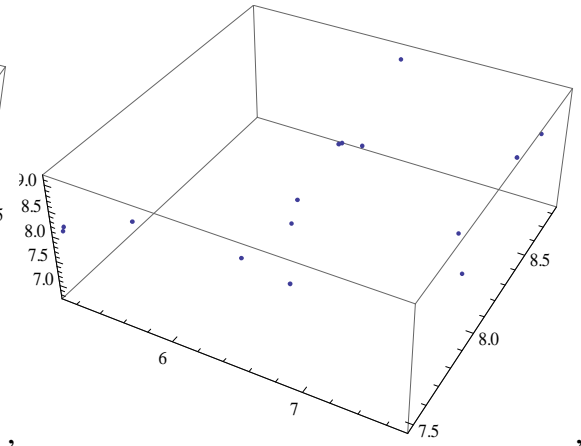


{

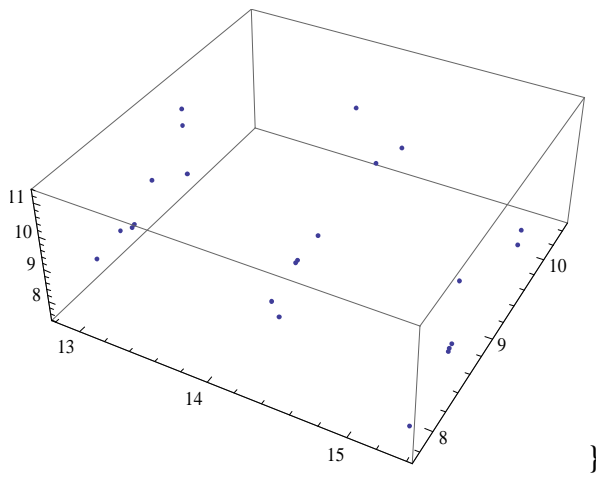
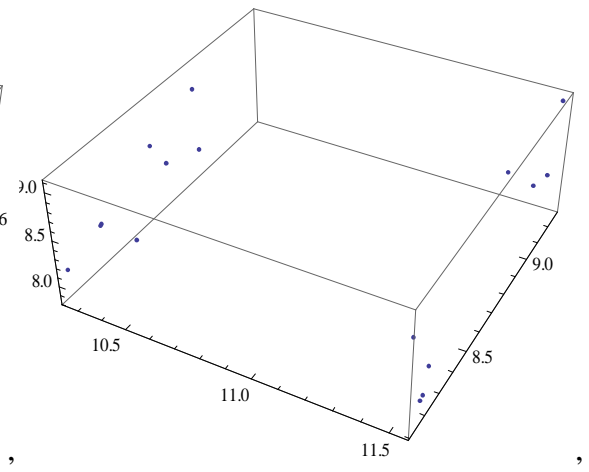
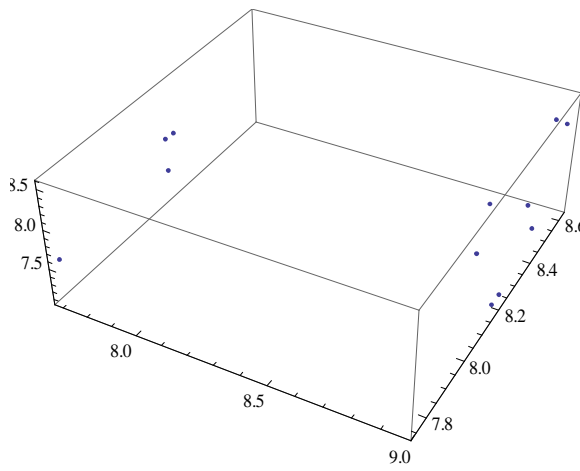
,



,

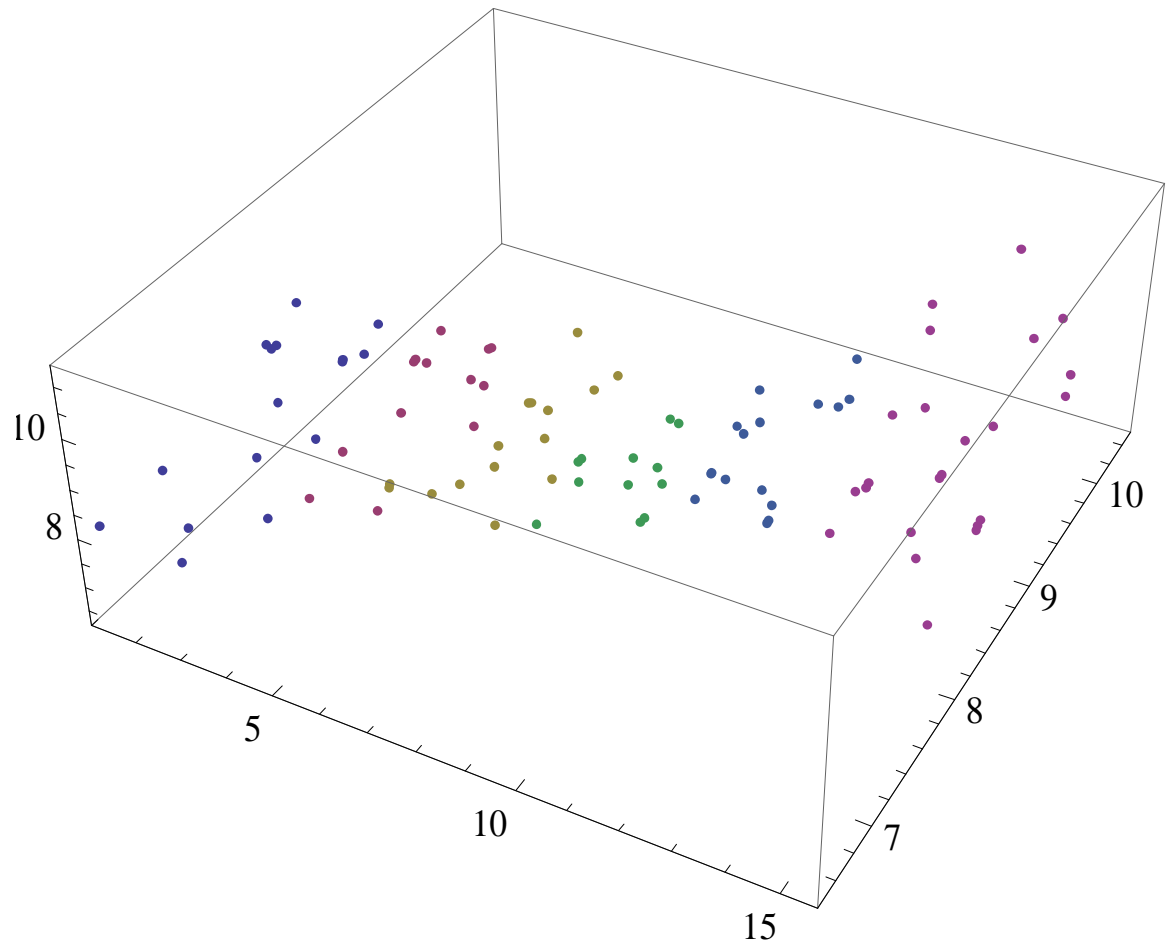


,



,  
}

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{\{95\}\}$  :  
 $\{\{14.1026, 8.4716, 7.4596\}, \{14.1026, 8.3747, 8.2187\}, \{14.1026, 8.7162, 8.5712\}, \{14.1026, 8.7448, 8.5788\}\}$

Posiciones en la vecindad actual cercana del vector:

$\{35, 23, 11\}$

Posiciones del estado N-1 o anterior:

$\{34, 22, 10\}$

Posiciones del estado N+1 o siguiente:

{36,24,12}

Datos del estado N+1 o siguiente:

{{ 15.3846,7.7458,8.2673},{ 15.3846,8.3285,8.8961},{ 15.3846,8.364,8.9303}}

Resultado Prospectivo con cercanos igual a 3

{15.3846,8.1461,8.6979}

### Pronóstico resultante

PRONOSTICO A:DIC 2009		
	VENTAS	INVENTARIO
Datos Estándar	8.1461	8.6979
Cifra Pronosticada	28,743,706	20,345,965
Diferencia Real	320,137	2,512,025
% Diferencias	1.13%	14.09%
Promedio % Diferencias		7.61%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

## Resultados Pronóstico 2009/07-12 Incrementales

Ejemplo de corrida de 2009/08 (Agosto 2009)

\*\*\*\*\*

Modelo de análisis prospectivo.

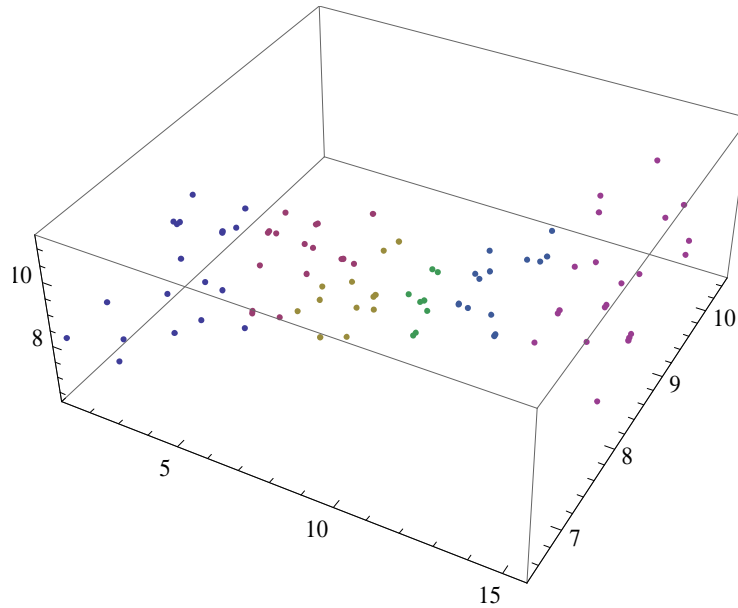
Datos de trabajo - vector: {8.9744,8.31587,7.9177} con 6 clusters, y 3 vecinos cercanos.

\*\*\*\*\*

Datos agrupados por clusters en 6 clusters.

Grafico tridimensional de los datos por cada cluster.

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original {{91}} :  
{{8.9744,8.31587,7.9177},{8.9744,8.3867,7.8153},{8.9744,8.6229,8.242},{8.9744,8.439  
7,7.356}}

Posiciones en la vecindad actual cercana del vector:

{67,79,31}

Posiciones del estado N-1 o anterior:

{66,78,30}

Posiciones del estado N+1 o siguiente:

{68,80,32}

Datos del estado N+1 o siguiente:

{{10.2564,8.5841,8.6711},{10.2564,8.8613,8.1777},{10.2564,8.465,7.8217}}

Resultado Prospectivo con cercanos igual a 3

{10.2564,8.6368,8.2235}

Tablas de resultados del proceso incremental de pronóstico:

RESULTADOS CON DATOS PROSPECTIVOS INCREMENTALES								
PRONOSTICO A: JUL 2009			PRONOSTICO A: AGO 2009			PRONOSTICO A: SEP 2009		
	VENTAS	INVENTARIO		VENTAS	INVENTARIO		VENTAS	INVENTARIO
Datos Estándar	8.31587	7.91777	Datos Estándar	8.6368	8.2235	Datos Estándar	8.6882	8.49393
Cifra Pronosticada	29,342,744	18,521,099	Cifra Pronosticada	30,475,153	19,236,257	Cifra Pronosticada	30,656,519	19,868,842
Diferencia Real	879,889	-543,106	Diferencia Real	1,839,844	133,273	Diferencia Real	1,488,680	818,308
% Diferencias	3.09%	-2.85%	% Diferencias	6.43%	0.70%	% Diferencias	5.10%	4.30%
Promedio % Diferencias		0.12%	Promedio % Diferencias		3.56%	Promedio % Diferencias		4.70%
PRONOSTICO A: OCT 2009			PRONOSTICO A: NOV 2009			PRONOSTICO A: DIC 2009		
	VENTAS	INVENTARIO		VENTAS	INVENTARIO		VENTAS	INVENTARIO
Datos Estándar	8.98197	8.73727	Datos Estándar	9.25103	8.68143	Datos Estándar	8.93533	8.88763
Cifra Pronosticada	31,693,093	20,438,058	Cifra Pronosticada	32,642,478	20,307,438	Cifra Pronosticada	31,528,523	20,789,777
Diferencia Real	1,736,137	872,422	Diferencia Real	2,750,254	2,857,978	Diferencia Real	3,104,953	2,955,838
% Diferencias	5.80%	4.46%	% Diferencias	9.20%	16.38%	% Diferencias	10.92%	16.57%
Promedio % Diferencias		5.13%	Promedio % Diferencias		12.79%	Promedio % Diferencias		13.75%

Tabla F1. Análisis porcentual del proceso de pronóstico incremental.



<b>Comparacion de Casos</b>			
Mes	Caso 1	Caso 2	Diferencias
Julio	0.12%	0.12%	0.00%
Agosto	3.81%	3.56%	-0.25%
Septiembre	-0.26%	4.70%	4.96%
Octubre	-1.20%	5.13%	6.33%
Noviembre	7.51%	12.79%	5.28%
Diciembre	7.61%	13.75%	6.14%

Tabla F.2. Análisis sumario del proceso incremental.

## XVIII. APÉNDICE G

### Resultados Pronóstico 2010/01

Ámbito definido:

- a) Datos de 2002/01-2009/12
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2010/01.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

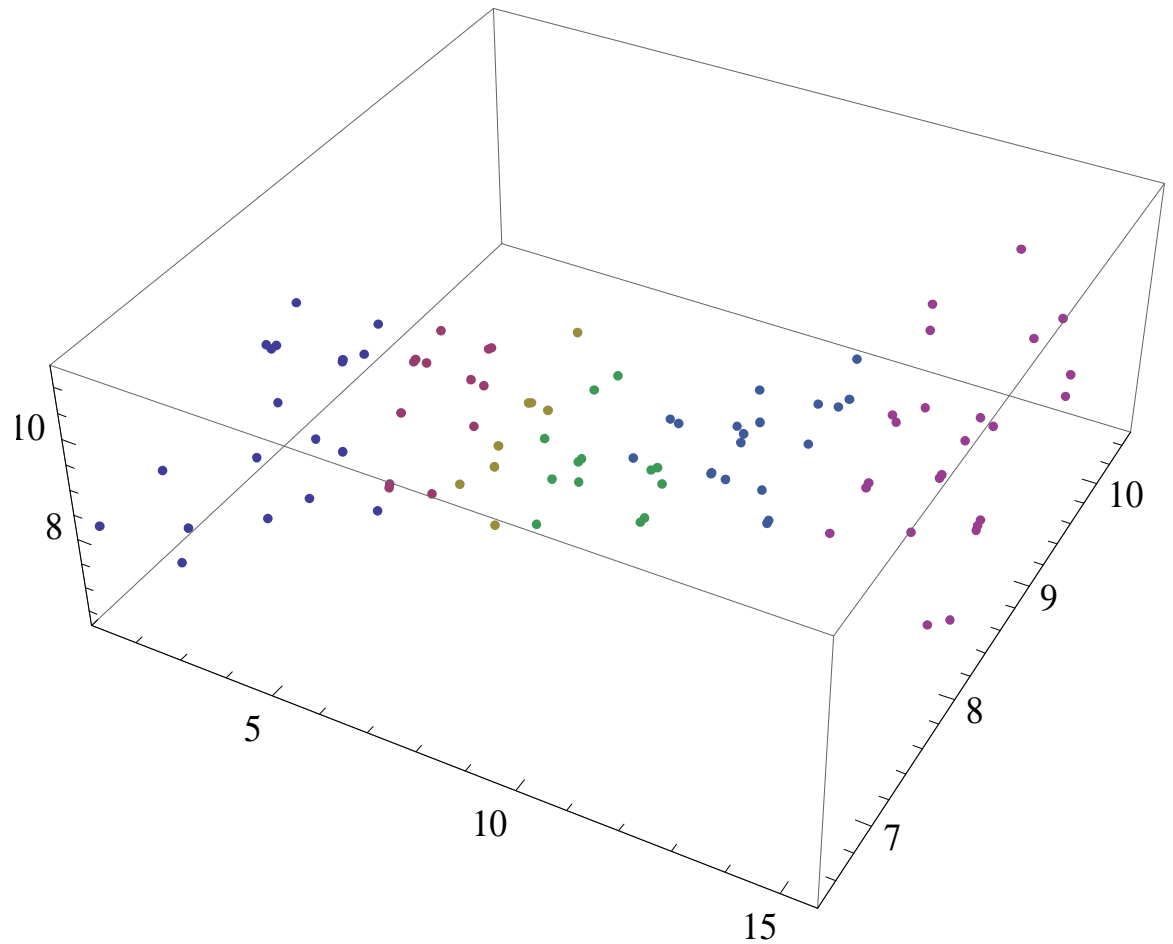
Datos de trabajo - vector: {15.3846,8.0554,7.624} con 6 clusters, y 3 vecinos cercanos.

\*\*\*\*\*

Datos agrupados por clusters en 6 clusters.

Grafico tridimensional de los datos por cada cluster.

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{96\}$  :  
 $\{15.3846, 8.0554, 7.624\}, \{15.3846, 7.7458, 8.2673\}, \{15.3846, 8.3135, 8.8365\}, \{15.3846, 8.3285, 8.8961\}$

Posiciones en la vecindad actual cercana del vector:

$\{36, 60, 24\}$

Posiciones del estado N-1 o anterior:

$\{35, 59, 23\}$

Posiciones del estado N+1 o siguiente:

{37,61,25}

Datos del estado N+1 o siguiente:

{{1.2821,7.1569,6.6511},{1.2821,7.0135,8.1863},{1.2821,7.9761,8.8288}}

Resultado Prospectivo con cercanos igual a 3

{1.2821,7.38217,7.88873}

### Pronóstico resultante

PRONOSTICO A: ENE 2010		
	VENTAS	INVENTARIO
Datos Estándar	7.38217	7.88873
Cifra Pronosticada	28,774,257	19,012,925
Diferencia Real	-793,247	229,627
% Diferencias	-2.68%	1.22%
Promedio % Diferencias		-0.73%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

## Resultados Pronóstico 2010/02

Ámbito definido:

- e) Datos de 2002/01-2010/01
- f) Clústeres=6.
- g) Vecinos cercanos a considerar 3.
- h) Pronóstico: 2010/02.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

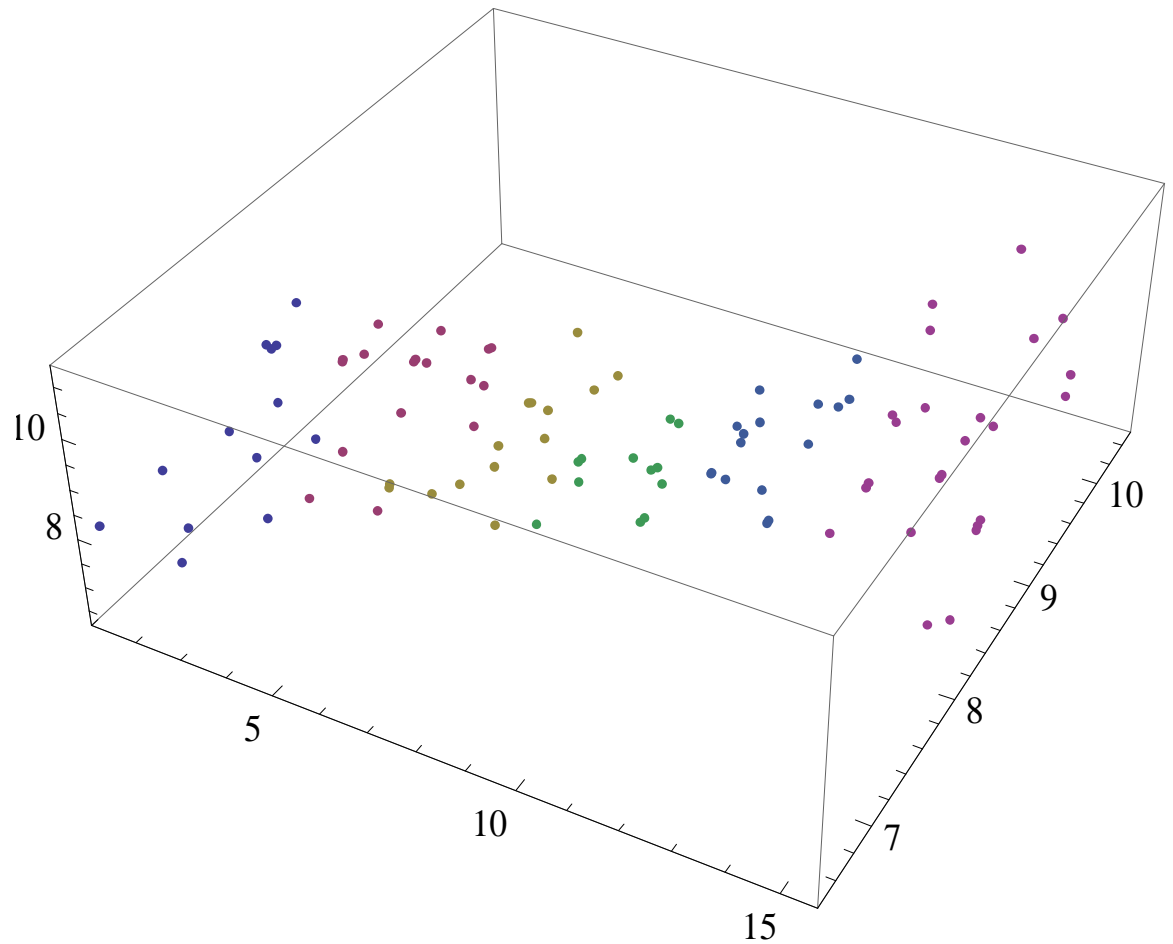
Datos de trabajo - vector: {1.2821,7.5857,7.7935} con 6 clusters, y 3 vecinos cercanos.

\*\*\*\*\*

Datos agrupados por clusters en 6 clusters.

Grafico tridimensional de los datos por cada cluster.

Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{\{97\}\}$  :  
 $\{\{1.2821,7.5857,7.7935\},\{1.2821,8.0276,7.4999\},\{1.2821,7.0135,8.1863\},\{1.2821,8.0176,8.6622\}\}$

Posiciones en la vecindad actual cercana del vector:

$\{49,61,1\}$

Posiciones del estado N-1 o anterior:

$\{48,60,0\}$

Posiciones del estado N+1 o siguiente:

{50,62,2}

Datos del estado N+1 o siguiente:

{{2.5641,7.8596,7.4693},{2.5641,7.3411,8.1718},{2.5641,8.1577,8.4998}}

Resultado Prospectivo con cercanos igual a 3

{2.5641,7.78613,8.04697}

### Pronóstico resultante

PRONOSTICO A: FEB 2010		
	VENTAS	INVENTARIO
Datos Estándar	7.78613	8.04697
Cifra Pronosticada	30,348,814	19,394,306
Diferencia Real	1,339,962	1,776,577
% Diferencias	4.62%	10.08%
Promedio % Diferencias		7.35%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
1	2002	1	EIO	Manejo del Sistema informático ASPEL
1	2002	1	ETR	Argentina devalúa el peso en un 28 por ciento.
48	2005	12	EIF	Absorción de una empresa competencia

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
50	2006	2	EIO	Adaptación de sistema informático al 80% al proceso

## Resultados Pronóstico 2010/03

Ámbito definido:

- a) Datos de 2002/01-2010/02
- b) Clústeres=6.
- c) Vecinos cercanos a considerar 3.
- d) Pronóstico: 2010/03.

Resultados:

\*\*\*\*\*

Modelo de análisis prospectivo.

Datos de trabajo - vector: {2.5641,7.4424,7.3098} con 6 clusters, y 3 vecinos cercanos.

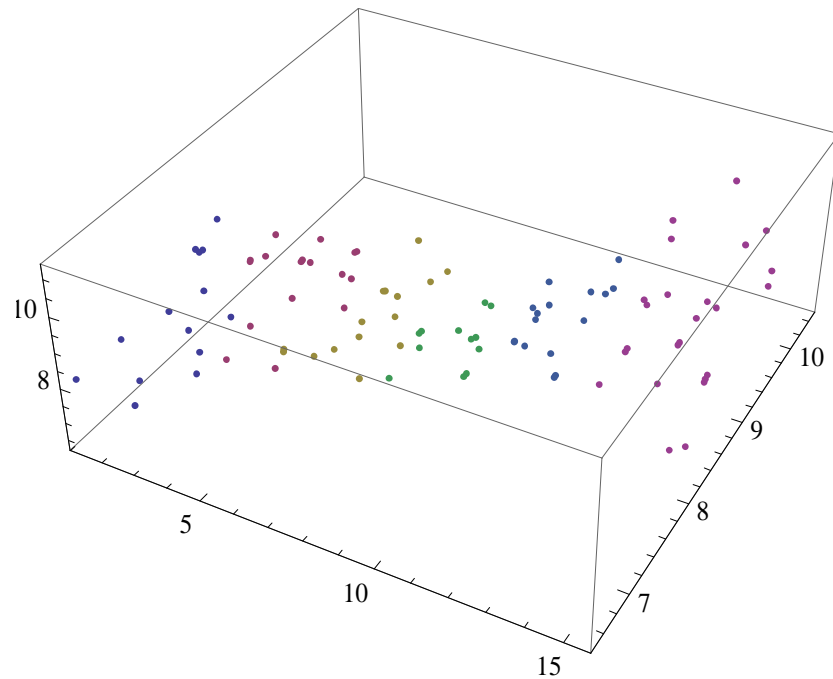
\*\*\*\*\*

Datos agrupados por clusters en 6 clusters.

Grafico tridimensional de los datos por cada cluster.



Grafico de los datos agrupados.



Datos en la vecindad más cercana dentro del cluster a la posición original  $\{\{98\}\}$  :

$\{\{2.5641, 7.4424, 7.3098\}, \{2.5641, 7.8596, 7.4693\}, \{2.5641, 7.3798, 6.7989\}, \{2.5641, 6.653$   
 $2, 7.5065\}\}$

Posiciones en la vecindad actual cercana del vector:

$\{50, 38, 74\}$

Posiciones del estado N-1 o anterior:

$\{49, 37, 73\}$

Posiciones del estado N+1 o siguiente:

$\{51, 39, 75\}$

Datos del estado N+1 o siguiente:

$\{\{3.8462, 8.1575, 7.8139\}, \{3.8462, 7.8743, 6.3015\}, \{3.8462, 7.2818, 7.8703\}\}$

Resultado Prospectivo con cercanos igual a 3

{3.8462,7.7712,7.32857}

### Pronóstico resultante

PRONOSTICO A: MAR 2010		
	VENTAS	INVENTARIO
Datos Estándar	7.7712	7.32857
Cifra Pronosticada	30,290,620	17,662,863
Diferencia Real	1,046,183	-213,883
% Diferencias	3.58%	-1.20%
Promedio % Diferencias		1.19%

### Casos de eventos antecedentes

a) Estado N y N-1 de vecinos cercanos.

Posición	Año	Mes	Tipo	Memoria de Casos
38	2005	2	EIO	Ampliación de planta de producción
50	2006	2	EIO	Adaptación de sistema informático al 80% al proceso

b) Estado N+1 de vecindad cercana.

Posición	Año	Mes	Tipo	Memoria de Casos
				NO HAY DATOS

## XVI. GLOSARIO

**Algoritmos genéticos:** Son algoritmos computacionales que se inspiran en la teoría de la evolución biológica para generar soluciones a problemas, basados en la cruce de soluciones posibles y mutaciones para dar resultados.

**Arquitectura Centralizada:** Todos los recursos se ubican en un solo sitio físico

**Arquitectura Cliente/Servidor:** Los servidores consolidan los datos de manera central y los clientes realizan el proceso de aplicaciones de manera distribuida

**Arquitectura Distribuida:** Los recursos se ubican en cada unidad de negocio

**Arquitectura de Virtualización:** Una forma de administración que aparenta ser centralizada, que permite la ubicación física en uno o varios lugares sin que esto impacte al usuario

**Balance Score Card:** Técnica de control de gestión basada en la medición del desempeño de 4 factores: Financiero, Clientes, Conocimiento y Procesos.

**BI: Business Intelligence.** Es una técnica computacional aplicable a problemas de negocios, basada en el análisis de datos a través de herramientas matemáticas, y que elabora propuestas de apoyo la toma de decisiones.

**Cadena de valor:** Es un modelo teórico que permite describir el desarrollo de las actividades de una organización empresarial generando valor al cliente final.

CBR.- Es un modelo que propone la resolución de nuevos problemas sobre la base de las soluciones de problemas similares del pasado.

CEO: Chair Executive Officer, Director General

CIO: Chair Information Officer, Director de Informática

CFO: Chair Financial Officer, Director Financiero

CIA: Chair Internal Audit, Director de Auditoría Interna

COO: Chair Operations Officer, Director de Operaciones

Clúster: Es un conjunto de datos o vectores de datos que por sus características propias guardan niveles de semejanza que los agrupan en un conjunto.

DBA: Database Administrator, Administrador de Bases de Datos

DSS (Decision support systems): Sistemas de soporte a la toma de decisiones. Es un sistema de información basado en herramientas computacionales, que incluye sistemas basados en conocimiento, y que apoya los procesos de toma de decisiones.

Economía: es la ciencia que se ocupa de la manera en que se administran unos recursos o el empleo de los recursos existentes con el fin de satisfacer las necesidades que tienen las personas y los grupos humanos.

Empowered: Característica de las organizaciones que consiste en otorgar capacidad de decisión e iniciativa a sus funcionarios

ERP (Enterprise resource planning): Sistema empresarial de planeación de recursos. Es un sistema utilizado para administrar y coordinar todos los recursos de una empresa, información y funciones de negocios.

FODA: Análisis empresarial cuyas siglas son Fortalezas, Oportunidades, Debilidades y Amenazas.

**Fuentes de Ventaja:** Son las características internas de una organización que permiten tener una identidad de superioridad frente a los competidores.

**Gobernancia:** Es el establecimiento de una nueva forma de gobernar, con eficacia, calidad y buena orientación. Es el conjunto de metodologías, procesos, personas y herramientas que componen la fuerza que asegura la sustentabilidad de la organización alineando estos esfuerzos a la estrategia y cumplimiento de los objetivos.

**Help Desk:** Punto de contacto de los usuarios para obtener soporte técnico

**Heurística:** Es la capacidad de un sistema para realizar de forma inmediata innovaciones positivas para sus fines. La capacidad heurística es un rasgo característico de los humanos, desde cuyo punto de vista puede describirse como el arte y la ciencia del descubrimiento y de la invención o de resolver problemas mediante la creatividad y el pensamiento lateral o pensamiento divergente.

**Inteligencia artificial:** Es una área de la ciencia computacional dedicada al desarrollo de agentes racionales no vivos, construyendo modelos de procesos que simulan la inteligencia humana.

**Minería de Datos: DM, Data Mining.** Es una tecnología informática que busca encontrar información no trivial que reside internamente en datos, a través de varias metodologías matemáticas.

**NWA: Network Administrator, Administrador de Redes**

**Outsourcing:** Técnica de administración donde se solicita la operación de una o varias funciones de la operación a un tercero

**Portafolio de Servicios (portafolio de soluciones o aplicaciones):** Es la gama de servicios, soluciones o aplicaciones de tecnología que se integran a los procesos de operación y gestión de la organización.

PM: Project Manager, Gerente de Proyecto

POP: Point of Presence, sitio de entrada de un operador de comunicaciones

Posiciones de Ventaja: Son las formas en las que se manifiestan las fuentes de ventaja

Pronóstico: Señal por donde se conjetura o adivina algo futuro. Es establecer la probabilidad que un evento ocurra en el futuro, basándose en análisis y en consideraciones de juicio.

Prospectiva: Conjunto de análisis y estudios realizados con el fin de explorar o de predecir el futuro, en una determinada materia, aportando elementos rastreables de predicción.

Racionalidad: Es una capacidad humana que permite pensar, evaluar y actuar de acuerdo a ciertos principios de optimización y consistencia, para satisfacer algún objetivo.

ROA: Return On Assets, medida financiera que muestra el porcentaje de cuan rentable son los activos de una organización

Singularidad: En el área matemática se definen como eventos que tienen comportamientos extraños e inesperados.

SIO: Security Information Officer, Director de Seguridad

SM: Service Manager, Gerente de Servicios

SPOC: Single Point Of Contact, punto de contacto único para cualquier problema de soporte técnico.

S&P500 (stándar and pourst): Es una sección de la bolsa de Wall Street que analiza 500 empresas de mejor desempeño histórico y presente, entre todas las que intervienen en el mercado financiero americano.