

# TECNOLOGÍAS DE SISTEMAS PARA PYMES

# PANORAMA



# Editorial

---

En Colombia actualmente se discute sobre los temas de transferencia de tecnología, de innovación y de su relación con la actividad investigativa. Dentro de la agenda de entidades como el Ministerio de Educación, Colciencias, el Departamento Nacional de Planeación y otras entidades del gobierno, se plantea la necesidad de vincular la universidad y sus procesos de investigación aplicada con las necesidades reales del medio empresarial. Estos organismos demuestran la imperativa necesidad de este tipo de alianzas para generar un mayor desarrollo económico y social del país y para empezar a construir esquemas de cooperación que nos permitan aprovechar nuestros recursos de conocimiento.

Esta publicación es una evidencia de que tal proceso de integración es posible y de los resultados que produce. El Politécnico Grancolombiano y la Corporación de la Microempresa - ADMICRO han realizado una alianza en un convenio de cooperación a través del cual se han elaborado una serie de trabajos investigativos sobre el tema de la gestión tecnológica de las microempresas y de los modelos más adecuados para el desarrollo de *software* de gestión de negocios de bajo costo y amplia aplicabilidad en las microempresas.

El resultado de este proyecto de cooperación técnica es el *software* denominado AGIL. Este permitirá realizar actividades de gestión de negocios a partir de información consignada en bases de datos como Excel.

A lo largo de la publicación se podrán encontrar los referentes conceptuales así como algunos ejemplos y discusiones que los investigadores han considerado en el desarrollo del *software*. Esperamos que su lectura pueda permitir el debate académico sobre la adaptación de tecnologías del conocimiento a las realidades de la empresa colombiana.

**Diana Forero**  
Directora de Investigación Aplicada

# Flujo de caja.

## Herramienta clave para la toma de decisiones en la microempresa

JAIME FERNANDO RODRÍGUEZ ROCHA

### RESUMEN

Este artículo describe la experiencia de poner en contacto el entorno productivo (la empresa) con el del conocimiento (la academia o universidad) mediante grupos interdisciplinarios de académicos y empresarios para desarrollar un sistema de información sobre el manejo de los flujos de dinero que maneja una microempresa, y de esta manera optimizar la toma de decisiones en la misma.

### PALABRAS CLAVE

Microempresa, flujo de caja, flujo de efectivo, flujo de dinero, programas de computador, herramientas de gestión.

### ABSTRACT

This article describes the experience of contacting the productive environment (the company) with the one of knowledge (the academy or university) through interdisciplinary groups of scholars and businessmen in order to develop an information system about the management of money flow within a micro enterprise. This way, the decision making is optimized.

### KEY WORDS

Micro enterprise, cash flow, money flow, computer programs, management tools.

### I. INTRODUCCIÓN

Los profundos cambios estructurales de los negocios han modificado la naturaleza de la competencia en las grandes, pequeñas y microempresas, lo cual hace imperiosa la necesidad de poder organizar, controlar y manejar dichas empresas. El que no organiza no puede controlar; el que no controla no puede manejar y el que no maneja su empresa no sobrevive.

La base o materia prima para esto es la información interna y externa. Sin embargo, aun hoy existe una gran cantidad de empresas que toman decisiones simplemente con datos o sin ellos, en vez de resolver sus necesidades con base en información, entendida esta como datos procesados y analizados que nos dan un conocimiento para poder decidir.

Hoy, más que nunca, la tenencia de información del día a día sobre inventarios, flujos de efectivo o caja, ventas, vendedores, cartera, proveedores, etcétera determinará la diferencia entre las posibilidades de éxito o un inminente fracaso de la empresa.

Si la empresa carece de sistemas de información y análisis sobre sus clientes, de operaciones internas y de desempeño de sus productos y de sus trabajadores, las decisiones se tomarán con base en las sensaciones, con todas las consecuencias y riesgos que esto trae consigo.

Ante la posibilidad de poder aplicar una metodología para soportar la operación diaria en el manejo del efectivo disponible de acuerdo con la necesidad de la empresa basada en la información almacenada en el sistema Admoney<sup>®1</sup>, se realizó un proyecto conjunto con la Corporación de la Microempresa – Admicro<sup>2</sup> con el fin de brindarle a los empresarios de la micro-empresa, usuarios del sistema, la posibilidad de tener un instru-

1 Admoney. El Administrador de su negocio; Programa de computador para manejar la operación de una microempresa. Admoney es una marca registrada de Internet & Contenido Ltda.

2 Corporación de la Microempresa – Admicro. Entidad sin ánimo de lucro para apoyar a la microempresa.

mento útil que les permita tomar mejores decisiones, que garanticen una mayor productividad entendida como la optimización de los recursos y la mejora de la competitividad basada en tener ventajas distintivas para diferenciarse de sus pares y generar valor agregado al sistema productivo en general de la empresa.

Colombia se encuentra en una fase de iniciación de este proceso; por lo tanto, requiere de lineamientos por parte de expertos que puedan fijar un rumbo de desarrollo seguro, al igual que la unión de esfuerzos para lograr fines comunes prioritarios en el avance efectivo del tema. Para ello, se debe convocar a los empresarios de la microempresa, empresarios independientes, profesionales independientes y pequeños negocios de nuestro país, para capacitarlos en áreas administrativas y de gestión con el fin de organizar, controlar y manejar sus empresas mediante herramientas tecnológicas tales como computadores, programas de computador, conexiones a internet y diferentes instrumentos tecnológicos. Además enseñarles el funcionamiento de los programas y de las conexiones en beneficio de la empresa y del empresario a fin de liberarlos de la presión de tener que efectuar todas las operaciones diarias, sin que les quede tiempo ni fuerzas para conseguir nuevos negocios, disfrutar de su trabajo, compartir con los amigos, estar en familia y descansar.

La microempresa no cuenta con herramientas de información financiera que le permitan la toma de decisiones adecuadas y oportunas, especialmente en lo referente al manejo de los recursos financieros que son base de la gestión y existencia de la misma. Al no contar con dicha información, el manejo no es eficiente ni competente. Dichas herramientas tienen un alto costo y un nivel de complejidad que no permiten que la microempresa tenga acceso a ellas.

Por lo tanto, el mencionado proyecto de investigación aplicada ha unido el entorno del conocimiento representado por la academia a través del Politécnico Gran-colombiano, el entorno productivo representado por la empresa a través de la Corporación de la Microempresa, Admicro, y puntualmente, la herramienta Admo-

ney®, “El Administrador de su Negocio” como medio de articulación de la información para diseñar y desarrollar un modelo sistemático basado en las operaciones diarias de la microempresa y registradas en un sistema de información transaccional (SIT) que proveerá los datos necesarios de la microempresa para llevarlos a otra herramienta como puede ser Excel, Staroffice o Pentaho donde se generará el flujo de caja periódico que permitirá manejar los recursos de tesorería de la misma de una manera segura y eficiente.

## II. EL ESTADO DEL ARTE Y SU ENTORNO

A continuación se describe el estado de las microempresas en Colombia, mostrando entre otros el tipo de apoyo institucional que reciben, cómo se encuentran segmentadas y por qué tienen un alto índice de fracaso.

### 2.1. La microempresa en Colombia

En Colombia existen aproximadamente 5.000.000 de empresas registradas, formales e informales, o entes productivos. Registradas y formalizadas ante las Cámaras de Comercio, aproximadamente 768.000 de las cuales 707.000 son microempresas, y de estas, un 26% se encuentran en Bogotá D.C. Estas microempresas se dedican a diferentes actividades industriales, de servicios y comerciales, dependiendo de la zona en que se encuentren ya sea nacional o local.

La microempresa nace de una necesidad familiar de generación de ingresos. Se definen como aquellas empresas con menos de diez empleados y un nivel de activos menor a 150 millones de pesos. La microempresa ha evolucionado y participa tímidamente dentro del tejido empresarial, pero tiene una gran importancia por su participación. Son aproximadamente el 90% de las empresas en el país, el 50% de los salarios, el 25% de las exportaciones no tradicionales y el 5% del PIB. Con la internacionalización de la economía corre graves riesgos de desaparecer<sup>3</sup>.

3 Cámara de Comercio de Bogotá (2005). *Observatorio Económico de Bogotá*. No. 18, Bogotá, CCB, Págs. 2 y 3.

## 2.2. Marco legal

Como se mencionó, solo aproximadamente un 14% de las microempresas se encuentran formalizadas, lo cual es un reto para el estado y los empresarios si se quiere organizar la economía creando valor general<sup>4</sup>.

## 2.3. Marco tecnológico

De las microempresas debidamente registradas, aproximadamente un 70% de sus fundadores tienen formación universitaria que garantiza en un alto porcentaje el uso de tecnología en todas sus formas y particularmente, de herramientas informáticas con programas de uso general como contables, administrativos y de escritorio<sup>5</sup>.

## 2.4. Segmentación por sectores, crecimiento y desaparición de las microempresas

Según datos de la Corporación para el Desarrollo de la Microempresa al año 2006, la microempresa ha crecido aproximadamente en un 95% en los últimos años y se podría decir que en el país, un 13% están dedicadas al sector industrial, un 55% al comercio y un 32% a servicios. Maneja buenos márgenes de utilidad por sus bajas infraestructuras, lo cual también las hace vulnerables a los cambios<sup>6</sup>.

De acuerdo con la Cámara de Comercio de Bogotá (CCB), un 95.5% de las empresas se quiebran y desaparecen al primer año; en un 83% las de servicios, 11% las de industria y 3% las de construcción<sup>7</sup>.

Para el Banco Mundial, globalmente, el porcentaje es del 90% y las razones principales para su desaparición son:

- Falta de oportunidades de negocio y calidad de las mismas (ventas).
- Falla en facilitar a los empresarios experiencias efectivas y no solo teóricas en gestión empresarial.
- Falta de acceso a crédito o financiamiento<sup>8</sup>.

### Jaime Fernando Rodríguez Rocha.

Ingeniero Industrial con experiencia en compañías privadas del sector financiero y entidades sin ánimo de lucro en asuntos administrativos, financieros y comerciales. Especialista en Administración de la Universidad del Rosario, corredor de Bolsa del Incolda y de la Bolsa de Bogotá. Consultor en emprendimiento, formación y gestión de microempresas. Profesor del Gimnasio Moderno y de la Universidad Javeriana. Jefe de Cambios, Comercio Exterior Banco de Bogotá, Subgerente, Gerente y Director del Departamento Financiero del Banco Industrial Colombiano. Gerente de Visor Cía. Ltda. Presidente de Internet & Contenido S.A. Gestor de la Corporación de la Microempresa - ADMICRO. Director de Proyección Social y Empresarial del Politécnico Gran Colombiano.

Para contactar al autor: [jfrodrig@poligran.edu.co](mailto:jfrodrig@poligran.edu.co)

En general, los empresarios de las microempresas adolecen de ser empresarios lo que implica ser integrales y competentes en el manejo de la empresa y específicamente en sus diferentes áreas funcionales tales como la administrativa, la financiera, la comercial y la productiva. Asimismo la empresa se mueve al ritmo de su área funcional que va más lenta, para lo cual se debe manejar un adecuado equilibrio.

Lo anterior no siempre lo conoce el empresario naciente, y a veces tampoco el que lleva ya tiempo, lo cual los pone en serios riesgos de fracaso. Y no lo saben porque sean torpes o ignorantes, sino porque no se lo han enseñado, en el mejor de los casos, de una manera real, práctica y ejecutable sino de una manera irreal, difícil y teórica no aplicable en la práctica. Se requieren herramientas útiles, fáciles, que apoyen en el día a día a los empresarios a manejar y gestionar sus empresas con información pertinente, actualizada y veraz de la situación de las mismas.

En lo pertinente al manejo del flujo de efectivo en el tiempo esta herramienta se desarrolló con base en la

4 Ibid.

5 Ramírez, Juan Carlos (2005). *Pymes más competitivas*. Bogotá: Mayol Ediciones, 2005. Pág. 8.

6 Corporación para el Desarrollo de la Microempresa, II Foro Nacional de la Microempresa. *La República*. Bogotá. 22, mayo, 2007, Pág. Pymes.

7 Cámara de Comercio de Bogotá, Congreso de microempresa. *La República*. Bogotá. 24, marzo, 2006, Pág. 18.

8 Banco Mundial. "El 90% de las nuevas firmas desaparece en el primer año". *La República*. Bogotá. 17, diciembre, 2005, Pág.19.

necesidad por parte del empresario, de contar con información sobre el dinero disponible para cumplir con las obligaciones y requerimientos diarios de la empresa para su normal desempeño. Al no contar con dicha información se presentan incumplimientos en el pago a proveedores, empleados, y en general, a terceros relacionados con la empresa generando desconfianza, requerimientos innecesarios y hasta problemas legales que desenfocan al empresario de su objetivo principal. De la misma manera, no contar con la información descrita impide anticiparse a posibles inconvenientes de flujo de efectivo, a poder tramitar un crédito, una inyección de capital por parte de los socios o una gestión comercial que subsane a estos.

Mediante Admoney®, “El Administrador de su Negocio” se pueden hacer las transacciones diarias que muestran el ciclo y la operación en la empresa y estas, a su vez, ser representadas en documentos contables así: compras reflejadas en entradas de inventarios a almacén así como una salida de dinero inmediata o futura, producción en salidas de almacén de materia prima y entradas de productos terminados; venta en salidas de almacén de producto terminado o la venta de un servicio representada en una factura que genera una entrada de dinero inmediata o futura y, finalmente, un pago representado en un comprobante de egreso que refleja la salida de dinero para el pago de una obligación. Llevar las salidas y entradas de dinero a una tabla y mostrar cuándo ocurren en un periodo determinado de tiempo (día, semana, mes, etc.), manejar el neto de entradas y salidas para conocer un disponible o un faltante, mover los pagos a fechas diferentes, ingresar dinero por créditos obtenidos o inyecciones de capital de socios o inversionistas temporales, permite conocer con bastante certeza cómo se desempeñará la empresa en cuanto a entradas, salidas y requerimientos o excedentes de dinero y tomar las decisiones más pertinentes.

Si los sistemas de información para validar gestión y estado del negocio, basados en sus sistemas de información de gestión diaria, ante las entidades crediticias son deficientes o inexistentes, se cierra cualquier

oportunidad de crédito posible en el sistema formal o bancario y se tiene que acudir al sistema informal o peor aún, cerrar y liquidar la empresa<sup>9</sup>.

## 2.5. Apoyo institucional y gubernamental a las microempresas

Existen apoyos estatales para la creación de nuevas empresas (Fondo Emprender, Fomipyme), programas (Bogotá Emprende, CCB-Alcaldía) que incentivan la creación de empresas, pero con los resultados enunciados en el numeral 2.4.

En el año 2007 fue aprobado el documento Conpes 3484 de 2007<sup>10</sup> que crea las condiciones para el desarrollo de la microempresa y las pymes enfocándolas a su modernización mediante la utilización de la tecnología y alineándolas con el Plan Visión Colombia 2019 del Departamento Nacional de Planeación<sup>11</sup>.

## 2.6. Sistema Admoney® y las microempresas

Todas las empresas requieren organización, control y manejo de sus áreas funcionales (administrativa-financiera, comercial y de producción) y, en consecuencia, de sistemas de gestión de información para la toma de decisiones.

Asimismo, el Estado a través de la DIAN hace un seguimiento a los evasores e informales, los cuales, al formalizarse, requieren de herramientas y asesorías para manejar sus obligaciones tributarias.

Admoney® es una herramienta de *software* (Enterprise Resource Planning, ERP por sus siglas en inglés) o un planeador de recursos empresariales que permite manejar, controlar y organizar la operación diaria de las microempresas, negocios u oficinas de una manera fácil, económica y eficiente.

Definido de una manera más simple y práctica, es un programa de computador para manejar un negocio o una microempresa; o definido para conocedores del

9 Asociación Bancaria. “La Pyme tiene dolientes”. *Revista Latinpyme*. Bogotá, noviembre, 2005.

10 Ministerio de Comercio, Industria y Turismo.

<<http://www.dnp.gov.co/PortalWeb/CONPES/ConpesEcon%C3%B3micos/2007/tabid/245/Default.aspx>>

11 Departamento Nacional de Planeación. Visión Colombia 2019. <[www.dnp.gov.co](http://www.dnp.gov.co)>

tema y lenguaje tecnológico avanzado, es una solución o *software* multiplataforma para Windows o Linux cuya función es generar operatividad, conectividad, transaccionalidad bidireccional, trazabilidad e integración de aplicaciones entre usuarios.

**OPERATIVIDAD:** Facilita labores diarias como: elaboración de facturas, comprobantes de egreso, entradas a almacén, etcétera. Produce informes y reportes de ventas, inventarios, cartera y otros datos base para elaborar el flujo de caja.

**CONECTIVIDAD:** Une, mediante el servicio de transferencia, almacenamiento y administración de datos, si es requerido, a las microempresas, negocios u oficinas con profesionales de cualquier disciplina que presten servicios de consultaría por internet en áreas administrativas, financieras y comerciales.

**TRANSACCIONALIDAD:** Permite que entre microempresas, negocios u oficinas y profesionales de cualquier disciplina exista una comunicación bidireccional permanente por internet para el envío y recibo de mensajes, documentos y/o archivos, de una manera ágil y segura.

**INTEGRACIÓN:** Desde un solo programa, facilita la operación y el manejo de diferentes aplicaciones que se requieren diariamente tales como: Office, StarOffice, Internet Explorer, etcétera. Asimismo, la exportación de información a estas para ser manejada como se requiera.

**UTILIDAD:** Brinda herramientas de escritorio como calculadora, calendario, agenda y base de datos para el manejo de las relaciones con clientes y proveedores.

Basa su operatividad en el ciclo del negocio el cual se explica en la Figura 1.

Dicho ciclo comienza con la compra de materia prima para producir insumos o mercancía para comercializar. Luego produce, si es del caso para luego vender, cobrar y pagar lo comprado si fue a crédito; ahí recomienza el ciclo. Todas estas transacciones se representan en documentos contables para cada una de ellas: comprar – entrada a almacén, producir – salida de almacén de materia prima, entrada a almacén de producto terminado, vender – facturar, cobrar – recibo de caja, pagar – comprobante de egreso. En el caso de una empresa comercial, no produce y en una de

prestación de servicios no requiere compra de materia prima ni producir.

### III. HIPÓTESIS DE LA NECESIDAD Y SU SOLUCIÓN

Ante el cambio de la economía y en la forma de hacer las cosas, la microempresa requiere basarse en la información y en el conocimiento, apoyada en TICs (Tecnologías de la Información y la Comunicación), pensar globalmente y basarse en esquemas asociativos que la fortalezcan.

Para el caso específico que nos ocupa, según lo analizado en el numeral 2.6 Ciclo del negocio, este se soporta principalmente en el manejo del flujo del dinero para mantener la empresa en marcha. Para que esta maneje sus flujos de efectivo de una manera ágil y eficiente requiere de herramientas o sistemas, preferiblemente basados en soluciones informáticas. Lo anterior permitirá tomar decisiones ciertas, oportunas y actualizadas basadas en información de la empresa y no en presentimientos sin sustentación.

### IV. EL FLUJO DE CAJA

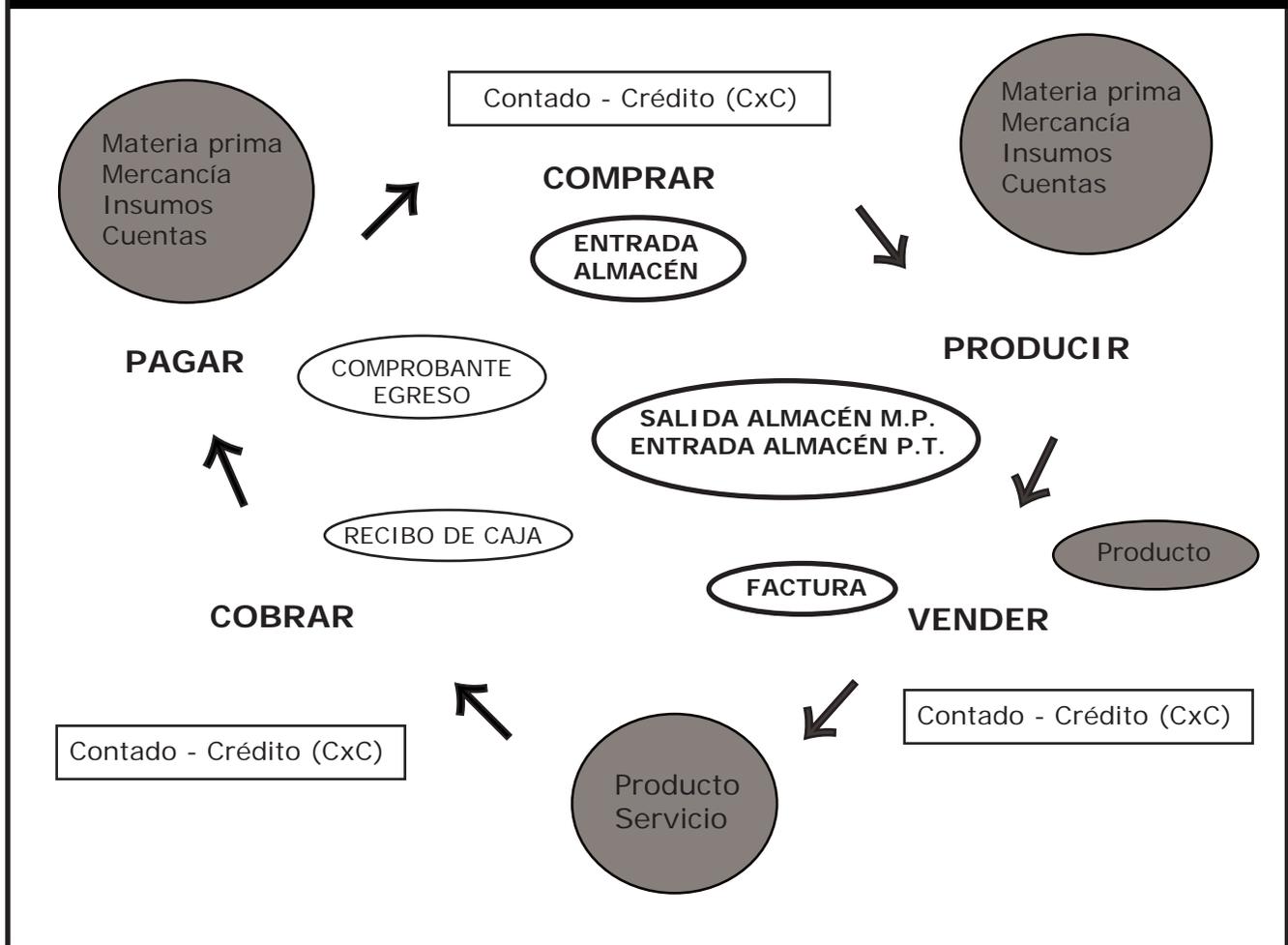
El objetivo de la implementación y utilización de un modelo de flujo de caja a partir de los datos generados por la operación diaria del sistema Admoney®, es permitir examinar detalladamente la información de los flujos de efectivo en un periodo de tiempo (diario, semanal, mensual, etcétera), para que los gerentes y directivos de las microempresas tengan un mejor análisis de la información, una organización financiera, un control del dinero y una ayuda en la toma de decisiones sobre factores económicos de la empresa.

El modelo muestra el saldo inicial del dinero, los ingresos y egresos de dinero con su respectivo saldo final. Tanto ingresos como egresos se dividen en operativos y no operativos y se subdividen de acuerdo con la actividad inherente a cada empresa (Ver Tabla 1).

### V. METODOLOGÍA

La metodología utilizada para la elaboración del flujo de caja que conduzca sistémicamente y paso a paso al logro de los objetivos fijados es la de comprender el

**Figura N°. 1. Ciclo de negocio**



negocio para ofrecer una solución real y clara, **comprender los datos** a analizar, a fin de no dejar ninguno por fuera, o si fuese el caso por no encontrarse en Admoney®, anexarlos al modelo (numeral 2.6), **preparar los datos** de la forma como se van a utilizar, **modelar el sistema** para adecuarlo a las necesidades de información del usuario y de las herramientas que se van a utilizar con el mismo, **evaluar el modelo** con usuarios o empresas manufactureras, comerciales y de servicios para medir su utilidad en cada una de ellas así como los ajustes requeridos, y finalmente, **divulgar y soportar** el modelo para llegar con él al usuario final

y poder cumplir con el objetivo de una investigación aplicada al segmento de la microempresa que deberá recibir un soporte permanente para cumplir con la meta de un modelo dinámico que se pueda ajustar a futuras necesidades.

### 5.1. Metodología tecnológica utilizada para mostrar el flujo de caja

Para la publicación del flujo de caja es necesario contar con herramientas de *software* que permitan efec-

## Figura N°. 2. Modelo de flujo de caja

	Enero	Febrero	Marzo	Abril
<b>Saldo inicial de caja</b>	\$ 0	\$ 52,646,359	\$ 118,544,542	\$ 117,444,523
<b>Ingresos operativos</b>				
Cuentas por cobrar	\$ 40,155,034	\$ 54,742,358	\$ 44,944,156	\$ 30,309,742
<b>Total ingreso operativo</b>	\$ 40,155,034	\$ 54,742,358	\$ 44,944,156	\$ 30,309,742
<b>Egresos operativos</b>				
Sueldos de producción	\$ 4,355,825	\$ 4,355,825	\$ 4,355,825	\$ 4,355,825
Otros pagos de producción	\$ 435,500	\$ 0	\$ 0	\$ 0
Gastos administrativos	\$ 3,000,000	\$ 3,000,000	\$ 3,000,000	\$ 3,000,000
Gastos generales	\$ 200,000	\$ 200,000	\$ 200,000	\$ 200,000
Compra por pagar	\$ 1,000,000	\$ 0	\$ 1,000,000	\$ 0
Cuenta por pagar	\$ 500,000	\$ 600,000	\$ 400,000	\$ 700,000
<b>Total egreso operativo</b>	\$ 9,491,325	\$ 8,155,825	\$ 8,955,825	\$ 8,255,825
<b>Ingresos no operativos</b>				
Venta de valores	\$ 0	\$ 0	\$ 0	\$ 0
Venta activos fijos	\$ 0	\$ 0	\$ 0	\$ 0
Venta inversiones	\$ 0	\$ 0	\$ 0	\$ 0
Capital	\$ 0	\$ 0	\$ 0	\$ 0
Préstamos recibidos	\$ 0	\$ 0	\$ 0	\$ 0
<b>Total ingreso no operativo</b>	\$ 0	\$ 0	\$ 0	\$ 0
<b>Egresos no operativos</b>				
Pago intereses	\$ 1,000,000	\$ 1,000,000	\$ 1,000,000	\$ 1,000,000
Pago impuestos	\$ 2,000,000	\$ 2,000,000	\$ 2,000,000	\$ 2,000,000
<b>Total egresos no operativos</b>	\$ 3,000,000	\$ 3,000,000	\$ 3,000,000	\$ 3,000,000
<b>Nuevo saldo</b>	<b>\$ 52,646,359</b>	<b>\$ 118,544,542</b>	<b>\$ 175,444,523</b>	<b>\$ 217,010,090</b>

tuar ciertos procesos como: captura de datos (Admoney), almacenamiento de datos en una base de datos (H2), preparar los datos para llevarlos a cubos olap y análisis dinámicos (Kettle), crear los *datamart* y los cubos (Mondrian), publicar los datos mediante *dashboards* o tableros de mando (Pentaho). Estas herramientas permiten finalmente mostrar un flujo de caja dinámico como el indicado en la Figura 2. Varias de las herramientas utilizadas son de código abierto lo cual facilita su utilización, redistribución y acceso al mismo por parte del usuario final.

## 5.2. Metodología utilizada para conocer las necesidades del usuario

La metodología utilizada maneja un concepto de clientelización, desarrolla lo que el usuario necesita para obtener un producto de acuerdo con las necesidades del mismo. Para tal fin, se efectuó una alianza academia (Politécnico Grancolombiano)–Empresa (Corporación de la Microempresa – ADMICRO), se intercambiaron conceptos y necesidades reales, los cuales se plasmaron en el modelo de flujo de caja y su operatividad y funcionalidad condujo a un mejor resultado final.

Se determinó que las microempresas no tienen sistemas para manejar datos y menos sistemas de información, por lo que se hace necesario y oportuno dotarlos de algunos. Se estructuró el sistema con la herramienta de soporte de operaciones Admoney®, que recopila los datos en una base de datos y puede ser utilizada para los fines propuestos. Allí se analizaron los datos y se prepararon.

Identificadas las necesidades y los datos disponibles, se procedió a diseñar el modelo de sistema que requirió conocer sobre flujos de caja y su función, para luego trasladar esto a una solución informática que cumpliera con las necesidades de actualizar, ser dinámica, consultar bases de datos, crear cubos y publicar los resultados.

La prueba inicial con empresas mostró la necesidad real del sistema así como otros requerimientos adicionales que se incluyeron dentro de la aplicación.

Con excepción de Admoney® todas las restantes herramientas son *opensource* o de código abierto ya que las

existentes con licenciamiento comercial son muy costosas. Como ventaja adicional todas las enunciadas se encuentran en una solución asociada por ser cada una un desarrollo independiente, pero asociado para dar una solución integral y completa.

## VI. CONCLUSIONES

Una vez finalizadas las fases y las tareas planteadas aparecieron necesidades de datos adicionales los cuales se suman a Admoney® para hacer más entendible al usuario final, el identificar lo realmente útil para la microempresa en cuanto a un modelo de flujo de caja, para finalmente, tener un prototipo de sistema informático que muestre al empresario las fechas, las cifras y la descripción de su flujo de dinero en un periodo definido que le permita tomar acciones y decisiones anticipadamente para controlar futuros inconvenientes.

Asimismo en el transcurso de las fases se requirió de búsqueda, análisis y elección de las herramientas informáticas con personal de la universidad, y externo a la misma, para llegar finalmente a la utilización de programas ya desarrollados que se pudieran utilizar por la optimización de sus costos, tiempo de desarrollo e implementación.

La mayor experiencia obtenida fue la de haber conformado equipos interdisciplinarios con personal de la universidad y la empresa: profesores, empresarios, ingenieros de sistemas, contadores, comunicadores, diseñadores, y especialistas en mercadeo. Esta posibilidad de poner a trabajar la universidad y a la empresa a través de sus actores genera un valor inmenso a la solución de necesidades e innovación al sector productivo y permite a la universidad cumplir realmente su papel de aportar conocimiento aplicado y real a quien lo necesita. Todo lo anterior es vital para la universidad quien brindará a sus estudiantes modelos de investigación aplicada que luego ellos podrán replicar en las empresas a las cuales vayan a trabajar, ya sean estas familiares, propias o ajenas. Para esto la universidad ha generado espacios físicos y herramientas tecnológicas de divulgación y formación de los estudiantes que cursen materias afines o pertinentes al tema.

El modelo se basa en una investigación aplicada que va de la microempresa a la docencia para ser imparti-

da a los alumnos quienes la proyectarán a la sociedad que finalmente se empoderará de una herramienta tecnológica facilitadora de toma de decisiones en condiciones de menor incertidumbre, mayor oportunidad y más fiables.

Las próximas fases contemplan la culminación de los informes, los tableros de mando o *dashboards* y la fase de prueba en varias empresas manufactureras, comerciales o de servicios para posteriormente concluir con la fase de implementación masiva de la aplicación.

## BIBLIOGRAFÍA

- CORNELLA, Alfonso (1996). *Información digital para la empresa*. Barcelona: Alfaomega.
- DÁVILA LADRÓN DE GUEVARA, Fernando (2002). *Administración y gestión de la Información*. Bogotá: Alfaomega-Politécnico Gracolombiano.
- DÁVILA LADRÓN DE GUEVARA, Fernando (2002). *Gestión de la información*. Bogotá: Cambio-Alfaomega-Politécnico Gracolombiano.
- ESTRADA, Gustavo (1998). *La riqueza de la información*. Bogotá: Gustavo Estrada.
- ESTUPIÑAN, Rodrigo (2004). *Estados de flujo de efectivo*. Bogotá: ECOE Ediciones.
- HERNÁNDEZ A., Jorge (2007). *Vendiendo tecnología a las pymes*. Bogotá: Mc Graw Hill, Décimonovena edición.
- RAMÍREZ G., Juan Carlos (2007). *Pymes más competitivas*. Bogotá: Rayol Ediciones, Segunda edición.

## Páginas web

- Admoney, El Administrador de su Negocio:  
<<http://www.admoney.biz/subseccion.asp?id=1048>>
- Corporación de la Microempresa - ADMICRO:  
<[www.admicro.org](http://www.admicro.org)>
- Pentaho:  
<[www.pentaho.com](http://www.pentaho.com)>

# Herramientas para la planeación estratégica en microempresas

RODOLFO VARGAS VALLECILLA

## RESUMEN

El siguiente artículo trata en forma general el proceso para desarrollar una planeación estratégica enfocada a resultados en pequeñas empresas y microempresas. Con ejemplos muestra cómo se debe recoger la información que se necesita para realizar un análisis acertado en la toma de decisiones.

## PALABRAS CLAVE

Planeación estratégica, mercado, competencia, costos, estrategia, objetivos estratégicos, misión, plan de mercadeo.

## ABSTRACT

The following article is an overview about the process used to develop a strategic planning focused on results for small companies and micro enterprises. Through examples, it shows how to gather the information that is necessary to make an accurate analysis in the decision making.

## KEY WORDS

Strategic planning, market, competition, costs, strategic objectives, strategy, mission, marketing plan.

Años atrás se creía que la planeación estratégica era solo para grandes empresas, de hecho, hoy en día todavía encontramos libros especializados en el tema con ejemplos en su mayoría aplicados a macroempresas<sup>1</sup>.

Philip Kotler hace una somera mención de la importancia de la planeación estratégica en la pequeña empresa en su libro de *Fundamentos de marketing*<sup>2</sup>, pero no entra en detalles de cómo se debe hacer.

Sin embargo, para fortuna de todos aquellos microempresarios con visiones claras de su negocio, la planeación estratégica es una herramienta para todo tipo de empresas, y su aplicación busca a través de un análisis profundo de la empresa y de su entorno, encontrar estrategias efectivas que permitan cumplir los objetivos que se propongan.

Tras años de experiencia e investigación sobre el tema, y trabajando en una compañía de consultoría para pymes, se desarrolló una metodología de planeación estratégica sencilla y dinámica, cuya condición de contar con menos personal y un presupuesto justo para operar requiere de pasos firmes y seguros de crecimiento.

A continuación relaciono todos los pasos que se deben aplicar para la planeación estratégica efectiva:

1. Análisis de la situación de la empresa en los últimos años.
2. Análisis del mercado.
3. Análisis de la competencia.
4. Análisis de costos.
5. Análisis del entorno macro.

1 Ferrel, O. C. y Hartline, Michael D. (2006). *Estrategia de marketing*. México: Thompson, Tercera edición, Pág. 22 y Apéndice B.

2 Kotler, Philip y Armstrong, Gary (2003). *Fundamentos de marketing*. México: Pearson Educación, Sexta edición, Pág. 59.

6. Pensamiento de los dueños.
7. Objetivos estratégicos.
8. Estrategia de la organización.
9. Planes por área de trabajo.

## 1. ANÁLISIS DE LA SITUACIÓN DE LA EMPRESA EN LOS ÚLTIMOS AÑOS

Para este análisis se requiere recopilar información numérica de la empresa en por lo menos tres o máximo cinco años. Lo más sencillo es tener la información en Excel en tablas que registren:

- Ventas de la empresa
- Ventas por ciudad
- Ventas por línea de productos
- Ventas por producto
- Ventas por vendedor
- Ventas por distribuidor

La clave para todo el análisis es tener información en bases de datos que permitan manejarla en forma eficiente. Fernando Dávila en su libro *Gestión de la información* lo ilustra de la siguiente forma:

La manera natural para el manejo de la información son las bases de datos. Para lograr un eficiente uso de ellas en Excel, es fundamental plantear las soluciones a los problemas de una manera organizada y con un diseño correcto, que nos permita aprovechar todas las posibilidades de Excel<sup>3</sup>.

Un ejemplo de cómo se recoge la información se muestra en la Tabla 1.

La forma más fácil de analizar esta información es graficándola como se muestra en el Gráfico 1. Así se verían las ventas de un vendedor.

Una vez se ha recopilado toda la información para el análisis de la situación actual, se procede con el siguiente paso del análisis del mercado.

En el trabajo que realizamos en el Politécnico Grancolombiano en el 2007, usamos el *software* Admone<sup>4</sup> que facilitó la recolección de toda la información comercial necesaria para el análisis.

## 2. ANÁLISIS DEL MERCADO

Cuando decimos análisis del mercado, nos referimos a los compradores tanto actuales como potenciales de los productos que vende la empresa.

Este análisis debe responder las siguientes preguntas básicas:

1. ¿Tenemos clientes o compradores?
2. ¿Cuánto compran de cada producto? (Tamaño del mercado).
3. ¿Cómo crece el mercado por producto?
4. ¿Qué características los identifican: demográficas, geográficas, sectores, etcétera?
5. ¿En el proceso de compra hay compradores, usuarios o influenciadores?
6. ¿Qué importancia tiene nuestro producto para los clientes? ¿Es fundamental para ellos o es un producto complementario?
7. ¿Cómo compran: mensual, semestral, anual, o de otra forma?
8. ¿Dónde lo compran: directamente con nosotros o a través de los distribuidores?
9. ¿Por qué eligen los clientes los productos de la empresa?
10. ¿Hay fidelidad de los clientes?
11. ¿Por qué los posibles clientes no eligen los productos de la empresa?
12. ¿Qué planes al futuro tienen nuestros clientes, donde nosotros podamos participar?

### Rodolfo Vargas Vallecilla.

Ingeniero Mecánico de la Universidad Nacional, Especialista en Administración de Empresas de la Universidad del Rosario, Especialista en Comercio Internacional de la Universidad Jorge Tadeo Lozano, Gerente de Proyecto de Business Partner Consulting, catedrático del Politécnico Grancolombiano en las materias de Plan de Mercadeo y Gerencia de Mercadeo.

Para contactar al autor: rvargasv@poligran.edu.co o rodolfovargasv@hotmail.com

<sup>3</sup> Dávila Ladrón de Guevara, Fernando (2002). *Gestión de la información, guía práctica con Excel*. Colombia: Alfaomega Colombiana, Primera edición, Pág. 1.

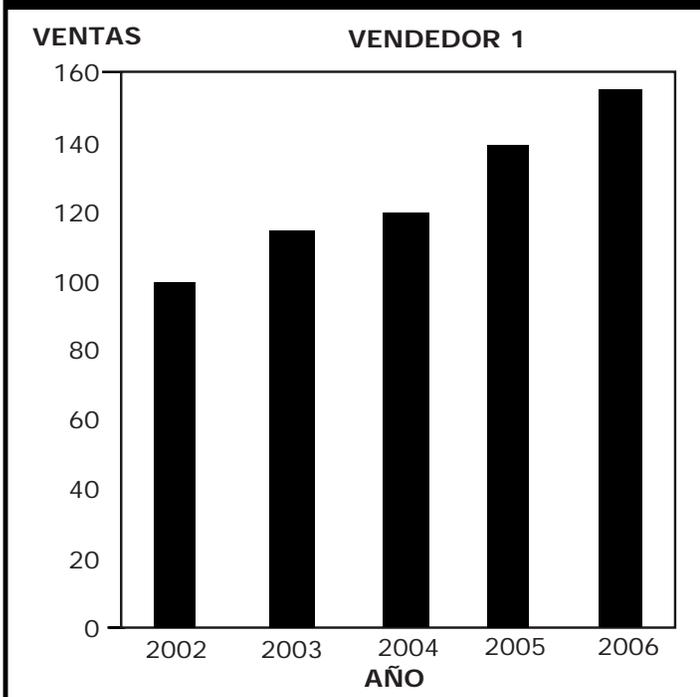
<sup>4</sup> Programa que factura, hace control de inventarios, almacena costos de productos, vendedores y clientes y tiene otras herramientas adicionales para manejar administrativa y comercialmente una pequeña empresa.

**Tabla N°. 1. Ejemplo de cómo se debe tabular la información de ventas por vendedor**

VENTAS POR VENDEDOR (Millones)					
Vendedores	Ventas				
Nombre*	2002	2003	2004	2005	2006
Vendedor 1	100	115	120	140	150
Vendedor 2	50	87	90	100	120
Vendedor 3	45	56	65	85	96
Vendedor 4	84	95	100	115	130
Vendedor 5	65	65	70	78	89
Vendedor 6	25	90	95	111	160
Vendedor 7	84	85	95	120	140
<b>TOTAL</b>	<b>453</b>	<b>593</b>	<b>635</b>	<b>749</b>	<b>885</b>

\* En orden alfabético

**Gráfico N°. 1. Ejemplo de cómo se grafican las ventas por vendedor**



Estas son las preguntas iniciales para hacer un análisis del mercado. Dependiendo del producto o negocio, puede ser posible que surjan nuevas preguntas para completar el análisis.

### 3. ANÁLISIS DE LA COMPETENCIA

De igual forma, para realizar el análisis de la competencia, debemos responder una serie de preguntas que nos faciliten el proceso:

1. ¿Quiénes son nuestros principales competidores? Nombre, productos, marcas, tamaño (ventas).
2. ¿Qué estrategias usan?
3. ¿Cuál es su participación en el mercado?
4. ¿Es fácil que entren nuevos competidores?

Seguramente también surgirán otras preguntas en la dinámica del análisis y harán más eficaz los resultados.

### 4. ANÁLISIS DE COSTOS

Para la realización de este punto, también es necesario conseguir, en tablas de Excel, información interna de la empresa que nos facilite el análisis, y, de igual forma, usamos el *software* Admoney para extraer la información.

Los costos que deben ir en las tablas son los variables, de forma que podamos obtener la utilidad bruta que genera cada producto y su margen bruto de rentabilidad. En la Tabla 2 vemos un ejemplo de esto. De esta forma es fácil percibir cuáles son los productos más rentables para la organización.

### 5. ANÁLISIS DEL ENTORNO MACRO

Aquí se revisan qué variables de tipo económico, político, cultural o tecnológico pueden afectar positiva o negativamente a la empresa. Toda la información para este análisis se recopila de fuentes secundarias. Por ejemplo, ¿qué influencia puede tener el T.L.C. en la compañía?

## 6. PENSAMIENTO DE LOS DUEÑOS

Con los dueños de la empresa se analizará si tienen clara la Misión de la organización, cuál es el rumbo que quieren para la empresa, qué proyectos tienen, y si quieren lanzar nuevos productos. La reunión con los dueños servirá para definir los objetivos estratégicos de la empresa.

## 7. OBJETIVOS ESTRATÉGICOS

En este aparte definimos cuáles van a ser los objetivos estratégicos a corto y mediano plazo que quieren los dueños de la empresa. Los objetivos se deben especificar en términos medibles, alcanzables y que tengan un tiempo determinado. Se debe tener claro qué tipo de objetivos quieren los dueños de la organización: rentabilidad, crecimiento, posicionamiento, etc. Por ejemplo, una vez estimado el potencial del mercado, se puede plantear un objetivo estratégico de crecimiento en ventas para uno o dos años.

## 8. ESTRATEGIA DE LA ORGANIZACIÓN

Una vez que se definen los objetivos estratégicos, se debe estudiar cuál debe ser la estrategia de la organización y cómo se van a lograr los objetivos estratégicos propuestos.

La estrategia de la empresa se analiza a la luz de la matriz producto-mercado, cuyo uso lo explica O. C. Walker Jr. en su libro *Marketing Estratégico*<sup>5</sup>.

Con base en la estrategia que se determine, se debe realizar el plan de mercadeo para cumplir con los objetivos estratégicos. La matriz producto - mercado se ilustra en el Gráfico 2.

## 9. PLANES POR ÁREA DE TRABAJO

Una vez definida la estrategia de la organización, se debe plasmar en los diferentes planes por realizar para lograr los objetivos.

Para el caso específico del área de mercadeo y ventas de la empresa, el plan que se debe llevar a cabo es un plan de mercadeo.

Un plan de mercadeo debe tener los siguientes elementos mínimos:

1. Objetivo del plan.
2. Estrategia de mercadeo.
3. Plan de trabajo.
4. Presupuesto.
5. Sistema de control.

Los objetivos del plan de mercadeo generalmente se especifican en términos de ventas anuales, aunque pueden haber otros. Estas ventas tienen que ser congruentes con los objetivos estratégicos de la empresa.

La estrategia de mercadeo será un conjunto de acciones que se deberán realizar sobre el producto, los precios, la comunicación y sobre la forma de distribución del producto.

Para que la ejecución de la estrategia sea fácil de implantar, es necesario plasmarla en un plan de trabajo que especifique qué acción se va a realizar, cómo se debe realizar la acción, quién es el responsable, cuánto tiempo dura y cuánto costaría ejecutarla. Una guía para desarrollar un plan de trabajo se muestra en la Tabla 3.

Una vez que se determine el plan de trabajo, es más sencillo establecer el presupuesto de gastos de mercadeo, dado que en el plan de trabajo se estima el costo de ejecución de las acciones que se van a llevar a cabo.

Para que el plan funcione, es necesario desarrollar un sistema de control que garantice la correcta implantación del mismo. Para controlar la ejecución, se deben desarrollar básicamente dos tipos de indicadores: indicadores de gestión que van a monitorear la ejecución de las acciones y los de resultado que van a monitorear el logro de los objetivos propuestos.

5 Walker, O. C. Jr., Boyd, Harper, Mullins, John y Larréché, Jean-Claude (2004). "Decisiones de estrategia corporativa y sus implicaciones de marketing". En: *Marketing Estratégico*. México: McGraw-Hill, capítulo IV, Cuarta edición, Pág. 49.

**Tabla N°. 2. Ejemplo de tabulación de cifras para analizar rentabilidad**

Producto	Precio de venta	Costo variable	Utilidad bruta	Margen bruto
Producto 1	\$ 1.000	\$ 700	\$ 300	30 %
Producto 2	\$ 2.500	\$ 1.500	\$ 1.000	40 %
Producto 3	\$ 1.800	\$ 1.500	\$ 300	17 %

**Gráfico N°. 2. Matriz de producto - mercado\***

	Productos actuales	Nuevos productos
Mercado actual	Estrategia de penetración de mercado	Estrategia de desarrollo de producto
Nuevos mercados	Estrategia de desarrollo de mercado	Estrategia de diversificación

\* Fuente: Ansoff, Igor (1976). *La estrategia de la empresa*. Pamplona: McGraw Hill, Pág.123 y 128.

**Tabla N°. 3. Ejemplo de un plan de trabajo**

Acción	Cómo se hace	Responsable	Tiempo de ejecución	Costo
Acción 1				
Acción 2				
Acción ...				

**BIBLIOGRAFÍA**

- DÁVILA L. DE G. Fernando (2002). *Gestión de la información*. Colombia: Alfaomega Colombiana, Primera edición.
- WALKER, O. C. Jr.; BOYD, Harper; MULLINS, John y LARRÉCHÉ, Jean-Claude (2004). *Marketing Estratégico*. México: McGraw Hill, Cuarta edición.
- FERRELL, O. C. y HARTLINE, Michael (2006). *Estrategia de marketing*. México: Thompson, Tercera edición.
- GUILTINAN, Joseph, GORDON, Paul y MADDEN, Thomas J. (1998). *Gerencia de marketing, estrategias y programas*. Colombia: McGraw Hill, Sexta edición.
- KOTLER, Philip y ARMSTRONG, Gary (2003). *Fundamentos de marketing*. México: Pearson Educación, Sexta edición.
- ANSOFF, Igor (1976). *La estrategia de la empresa*. Pamplona: McGraw Hill.

# Diseño de un sistema de costos para pymes

EDILMA PINEDA MARÍN

## RESUMEN

La pequeña empresa colombiana, con limitaciones en recursos financieros, tecnológicos y humanos requiere de apoyo administrativo y contable que faciliten gestionar sus propios procesos de crecimiento. Las ofertas de valor que las empresas proponen a los consumidores requieren ser costeadas de manera tal que el valor percibido por estos, permita a las organizaciones la fijación de precios y la toma de decisiones en los diferentes ámbitos ya sea financiero, de mercadeo o de producción, entre otros.

## PALABRAS CLAVE

Costo, sistemas de costeo, rastreo, acumulación, asignación, *full costing*, *direct costing*.

## ABSTRACT

Colombian small enterprises, limited in financial, technological, and human resources, require administrative and accounting support that allows them to manage their own growing processes. Value offers proposed to consumers require to be afforded in such a way that the value taken from these allows companies to establish prices and make decisions regarding the financial, marketing, and production environments, among others.

## KEY WORDS

Cost, costing systems, tracking, accumulation, allocation, *full costing*, *direct costing*.

## 1. INTRODUCCIÓN

Este documento presenta algunos conceptos básicos de la teoría de costos que deben ser tenidos en cuenta por las empresas cuando se trata de establecer un sistema primario de acumulación de costos para bienes y servicios que conduzcan al costeo de estos. El costo de adquisición o de producción de bienes y servicios, es uno de los tantos parámetros que se deben tomar como referencia a la hora de establecer o fijar precios de venta, de realizar análisis de rentabilidad de productos, de tomar decisiones frente a qué producir o comercializar.

El texto inicialmente plantea una conceptualización básica de costos, enmarcada dentro de la metodología denominada costeo total o de absorción (*full costing*), y la acumulación de costos bajo el sistema de costos por órdenes de fabricación. Seguidamente se presenta a los lectores, los documentos de entrada de información de costos que luego es procesada para la rendición de informes finales, tales como el costo reproducción de cada lote u orden de producción, el costo unitario, y finalmente, la presentación del estado de costo de producción.

## 2. COSTEANDO COSTOS

En la actualidad la información de costos se vuelve cada vez más crucial, pues los precios de los productos y servicios tienden a ser fijados por el mercado y no por los productores. La presión ejercida por la globalización desencadena guerras de precios que al final se convierten en guerras de costos haciendo que los negocios exitosos sean aquellos que tienen la capacidad de mantener o reducir sus costos a una velocidad mayor que la de sus competidores; o bien aquellos negocios

que logran diferenciar sus productos y servicios de manera tal que sus precios no son fijados por el mercado. De hecho, en materia de costeo de bienes y servicios se han presentado importantes cambios en la última mitad del siglo XX, ya que dentro de los componentes del costo de hacer negocios o de producir bienes y servicios, los costos indirectos han llegado a representar una gran parte del costo de estos.

En respuesta a los cambios de mercado, las compañías han invertido en tecnología y revisado las formas de administrar sus negocios, de manera tal, que con la introducción de la robótica en los procesos, costear bienes y servicios se ha convertido en una tarea básica, pero ha cobrado mayor importancia la administración de esos costos y la administración de la información como base para la toma de decisiones, que permitan a las empresas direccionar sus negocios en pos de la consecución de una posición de ventaja competitiva basada en costos.

### 3. COSTEAR ¿PARA QUÉ?

Los cambios en el mundo empresarial desde mediados de los años setenta del siglo pasado, disparados por la competencia global y las innovaciones tecnológicas, han conducido a innovaciones sorprendentes en la utilización de la información financiera y no financiera en las organizaciones. El nuevo entorno exige una información más precisa sobre los costos y la forma de proceder en cuanto a actividades, procesos, productos, servicios y clientes de la organización. Las empresas líderes están utilizando sus sistemas de costos para:

- Diseñar productos y servicios que satisfagan las expectativas de los clientes y, al mismo tiempo, puedan ser producidos y entregados con un beneficio.
- Detectar dónde hay que realizar mejoras continuas o reingeniería en calidad, eficiencia o rapidez; en actividades de aprendizaje.
- Guiar las decisiones de inversión y de mix (oferta) de producto.
- Elegir entre proveedores alternativos.
- Negociar con los clientes el precio, las características del producto, la calidad, las condiciones de entrega y el servicio a satisfacer.
- Estructurar unos procesos eficientes y eficaces de distribución y servicio para los segmentos objetivos de mercado y de clientes<sup>1</sup>.

No obstante lo anterior, nuestras pymes no están preparadas para generar y obtener una posición de ventaja competitiva sostenible basada en costos, debido entre otras causas a la ausencia de estructuras organizacionales y administrativas, a la carencia de sistemas de contabilidad adecuados y por ende, a la no implementación de sistemas de costos propios de la nueva era tecnológica que los priva de información relevante para valorar sus inventarios, determinar el monto de los recursos invertidos en la generación de sus propuestas de valor, y por tanto, evaluar el proceso de gestión de la firma y de sus administradores.

### 4. CÓMO HEMOS EVOLUCIONADO EN MATERIA DE COSTOS DURANTE EL ÚLTIMO SIGLO

La mayoría de los procedimientos de contabilidad interna y de costos de productos usados en el siglo pasado progresaron entre 1880 y 1925. Es interesante advertir que muchos de los primeros aspectos desarrollados (hasta casi 1914) se referían a la administración del costeo del producto (rentabilidad de la empresa a productos individuales y uso de esta información para tomar decisiones estratégicas).

Hacia 1925, este énfasis se había abandonado en favor de un costeo de inventarios (o sea, distribución de los costos de producción a los productos), así que los costos de inventario se pudieran reportar a los usuarios externos de los estados financieros de la empresa.

El reporte financiero se volvió la fuerza más importante que impulsaba el diseño de los sistemas de contabilidad de costos. Los gerentes y las empresas estaban dispuestos a aceptar información del costo promedio agregado sobre productos individuales. Al parecer no era necesaria una información de costos más detallada y exacta acerca de los productos individuales.

Las dos primeras décadas del siglo XX, según Chatfield: “se han caracterizado en la industria por la adopción de sistemas uniformes de costos y de contabilidad. Algunos de estos bastante burdos, en tanto que otros presentaban un trabajo excelente, preparado por los mejores cerebros contables de las industrias respectivas”<sup>2</sup>.

1 Kaplan, Robert y Cooper, Robin (1998). *Coste y efecto*. Barcelona: Ediciones Gestión, Segunda edición, Pág. 13.

2. Chatfield, Michael (1979). *Estudios contemporáneos sobre la evolución del pensamiento contable*. México: ECASA. Pág. 166.

Para las dos décadas siguientes, se evidencian mejoras y refinamientos en los principios, procedimientos y técnicas de la contabilidad de costos, los cuales incluyen todas las mejoras y adelantos en el desarrollo y utilización de los costos predeterminados y estándar; en el cálculo y distribución de los gastos indirectos; en el manejo de los materiales, compras, recepción, almacenaje, salidas y fijación de precios, en muchos de los casos debido a la creciente utilización de máquinas de cálculo, como registradoras y computadoras. Toda esta información de costos es empleada para fines de control y como base en el presupuesto flexible.

Algunos esfuerzos por mejorar la utilidad de los sistemas de costos convencionales se dieron en los años 50 y 60 del siglo XX; pero se centraron en darle más utilidad a la contabilidad financiera para los usuarios internos, en lugar de producir un conjunto de información y procedimientos nuevos, fuera del sistema de reporte externo.

Al respecto, Kaplan y Cooper afirman:

Muchas empresas, reconociendo la naturaleza arbitraria de las asignaciones de los costos indirectos de fabricación en un sistema de valoración de las existencias, han pasado a sistemas de costos directos para facilitar las decisiones de la dirección<sup>3</sup>.

La metodología del costeo directo, creada por Harris en 1934 en Estados Unidos y de Harrison en 1935 en Inglaterra, ignora totalmente los costos indirectos a la hora de calcular los costos de fabricación de productos y servicios y de los clientes. Sólo asignan los costos de materiales directos y los costos directos del personal a los productos. Los métodos de costos directos son adecuados si los costos indirectos y de estructura son una pequeña fracción de los costos totales, o si, como sostienen quienes abogan por este sistema, se trata de costos fijos. Sin embargo, las organizaciones han aprendido que no sólo los costos indirectos y de estructura no son fijos, sino que ni siquiera son variables. Para muchas empresas, estos costos son súper variables, los cuales se incrementan a una tasa más rápida que el volumen de producción o de ventas. Y los sistemas de cálculo de costos directos no pueden asignar el monto creciente de costos de *marketing*, ventas, distribución, desarrollo del producto y administrativos a clientes, canales y divisiones.

A medida que aumentaba la competencia y a medida que las bases de la misma se alejaban de la utilización eficiente de la mano de obra y de las máquinas, los directivos necesitaban una información más fiel respecto a los costos de los procesos, productos y clientes de la que podían obtener través del sistema utilizado para la contabilidad externa.

Los sistemas de costos basados en las actividades (ABC) emergieron a mediados de los años ochenta para satisfacer la necesidad de información fidedigna respecto al costo de los recursos asignables a los productos, servicios, clientes, canales de distribución. Los sistemas ABC permitieron que los costos indirectos y de estructura fueran conducidos, primero hasta las actividades y procesos y luego a los productos, servicios y clientes. Estos sistemas proporcionaron a los directivos una imagen mucho más clara de los costos de sus operaciones<sup>4</sup>.

Hay dos avances importantes que se relacionan con la tecnología de la información. Uno está íntimamente vinculado con la industrialización integrada a los sistemas. Con un proceso automatizado de fabricación, las computadoras se emplean para vigilar y controlar operaciones. Como se emplea una computadora, se puede captar una vasta cantidad de información sobre lo que está sucediendo en planta, casi en el mismo momento en que ocurre y pasarla a los administradores. Es posible rastrear productos a medida que se van moviendo de un lugar a otro en la fábrica e informar en tiempo real detalles como unidades producidas, material utilizado, mermas y costo de productos.

#### **Edilma Pineda Marín.**

Contadora Pública Titulada egresada de la Universidad Externado de Colombia con estudios de postgrado en Gerencia de Mercadeo. Profesora e investigadora de la Universidad Externado de Colombia, profesora de las universidades Jorge Tadeo Lozano y Central. Profesor adscrito al Área Contable de la Fundación Universitaria Politécnica Grancolombiano. Presidente ejecutiva de la Asociación de Contadores Públicos Externadistas Ascopex. Así mismo ejerce como consultora empresarial y Auxiliar de la Justicia. Para contactar a la autora: [epinedam@poligran.edu.co](mailto:epinedam@poligran.edu.co).

3 Kaplan, Robert y Cooper, Robin. Op. cit., Pág. 15.

4 Ídem. Págs. 15 y 16.

El segundo avance proporciona las herramientas requeridas: la disponibilidad de computadores personales (PC), *software* de hojas de cálculo y paquetes gráficos. El PC sirve como vínculo de comunicación con el sistema de información de la compañía, en tanto que las hojas de cálculo y los paquetes gráficos ofrecen la capacidad analítica para usar esa información. Máquinas y paquetes están al alcance de los administradores en todo tipo de organizaciones; estos les permiten efectuar gran parte de su propio análisis y reducir su dependencia de un departamento centralizado de información. Si una computadora actúa como terminal y está conectada a la base de datos de la organización, el acceso a la información se efectúa con mayor rapidez y permite preparar muchos reportes propios.

## 5. EL PROCESO DE PRODUCCIÓN

En las empresas manufactureras el proceso de producción se inicia con los requerimientos de materia prima de los departamentos de producción (etapa 1). Dentro de esos departamentos, la materia prima es procesada utilizando la mano de obra y otros recursos conocidos como costos indirectos (etapa 2) hasta obtener, al final del proceso, un producto terminado (etapa 3).

Cuando las condiciones económicas del negocio o del mercado no son buenas, o bien cuando los costos están por encima del precio de venta, la información de costos y la adecuada diferenciación de los costos fijos y variables es esencial para generar estrategias o acciones que ayuden a aumentar las ventas y disminuir en lo posible las pérdidas. Cuando se tiene certeza de que las condiciones económicas desfavorables al negocio persistirán a largo plazo, la empresa debe tomar decisiones sobre si seguir produciendo, cambiar de giro, diversificar los productos o en un caso extremo, cerrar el negocio, siendo la información de costos pilar para la toma de decisiones.

Bajo condiciones económicas favorables en las que los precios de mercado están por arriba del costo de producción, la información de costos puede ser utilizada para planear ofertas y otras tácticas para ganar mercado y aumentar utilidades.

## 6. CONCEPTOS DE COSTOS

Mallo y Kaplan se refieren a la existencia de diversas versiones de costo o, ya que su definición interesa a múltiples disciplinas como la economía, el derecho, la ingeniería y la contabilidad dicen: "...el concepto económico del costo se ha utilizado en dos versiones generales: la primera, en sentido de consumo o sacrificio de recursos de factores productivos, y la segunda en el sentido de costo alternativo o de oportunidad"<sup>5</sup>, conceptos estos que ya habían sido tratados por los economistas clásicos, principalmente por David Ricardo.

Baker y Jacobsen definen el costo "...como una suma de erogaciones, es decir, el costo inicial de un activo o servicio adquirido se refleja en el desembolso de dinero en efectivo y otros valores, o sea un pasivo incurrido"<sup>6</sup>. Estos dos autores incorporan el concepto de "factores de costo" cuando afirman que:

...además del precio de adquisición de un activo, se puede incurrir en otros costos preliminares para permitir que el activo rinda los servicios esperados, estando incluidos en esta categoría los cargos por transporte, por recepción de materiales y equipos, los costos de instalación, etc.<sup>7</sup>

Finalmente, concluyen que una organización incurre en costos con el propósito de generar ingresos.

Ahora bien, para Baker y Jacobsen cuando se trata de diferenciar costos, gastos y pérdidas, El costo, está definido como la porción del precio de adquisición de bienes, artículos, propiedades o servicios, que ha sido diferida o que todavía no se ha aplicado a la realización de ingresos (inventarios, propiedad planta y equipo), en tanto que los gastos son costos que se han aplicado contra el ingreso de un periodo determinado (salarios de administración). De otro lado, las pérdidas son reducciones en la participación de la empresa por las que no se ha recibido ningún valor compensatorio, sin incluir los retiros de capital.

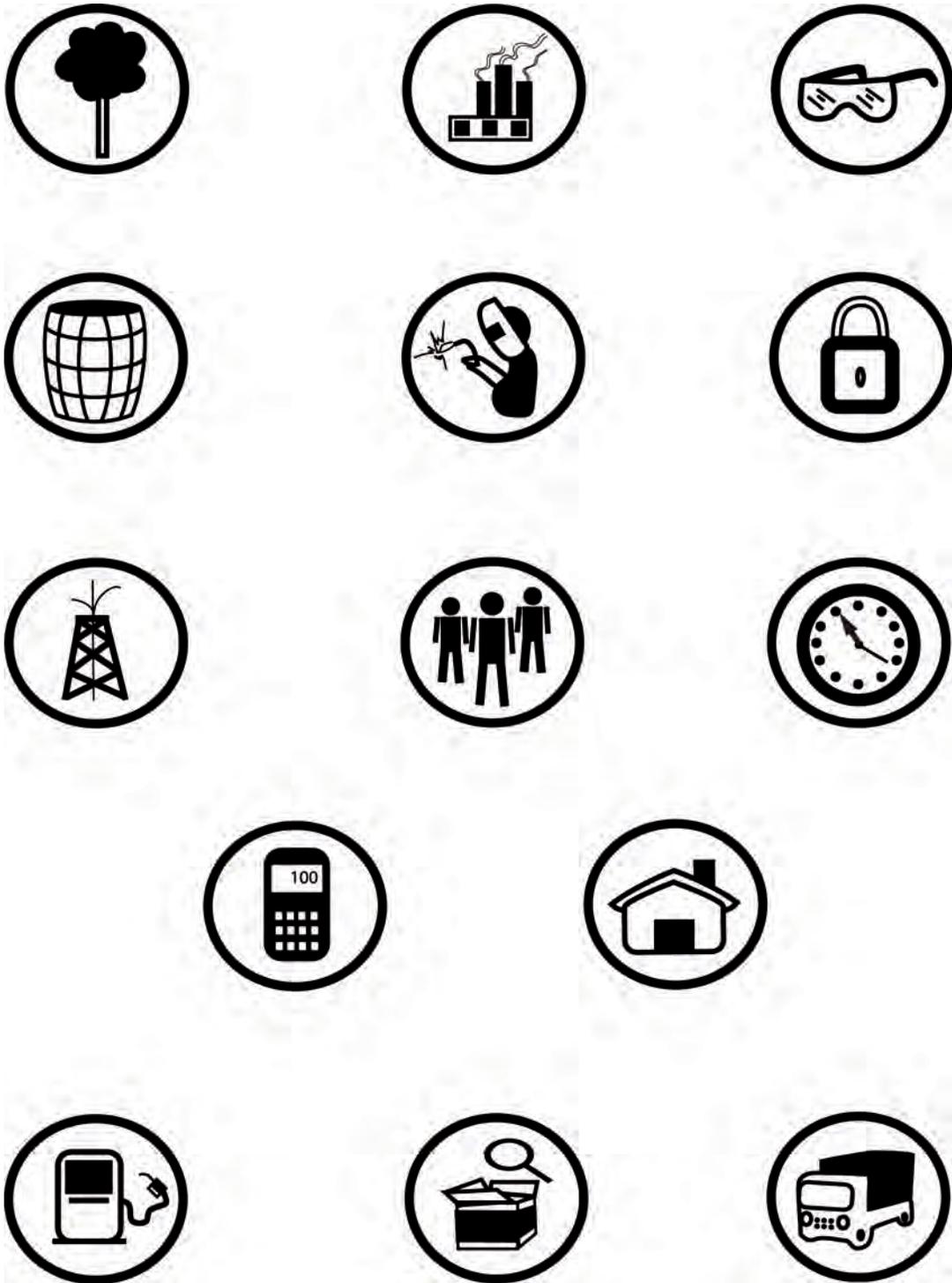
Los autores anteriormente mencionados presentan una clasificación primaria de los costos que afectan el estado de resultados de las organizaciones:

5 Mallo, Carlos, Kaplan, Robert y Meljem, Silvia (2000). *Contabilidad de costos y estratégica de gestión*. Madrid: Prentice Hall.

6 Baker, Morton, Jacobsen, Lyle y Padilla, David Noel (1988). *Contabilidad de costos*. México D.F.: McGraw Hill Interamericana, Segunda edición, Pág. 3.

7 Ídem. Pág. 4.

## Gráfico N°. 1. Etapas del proceso de producción

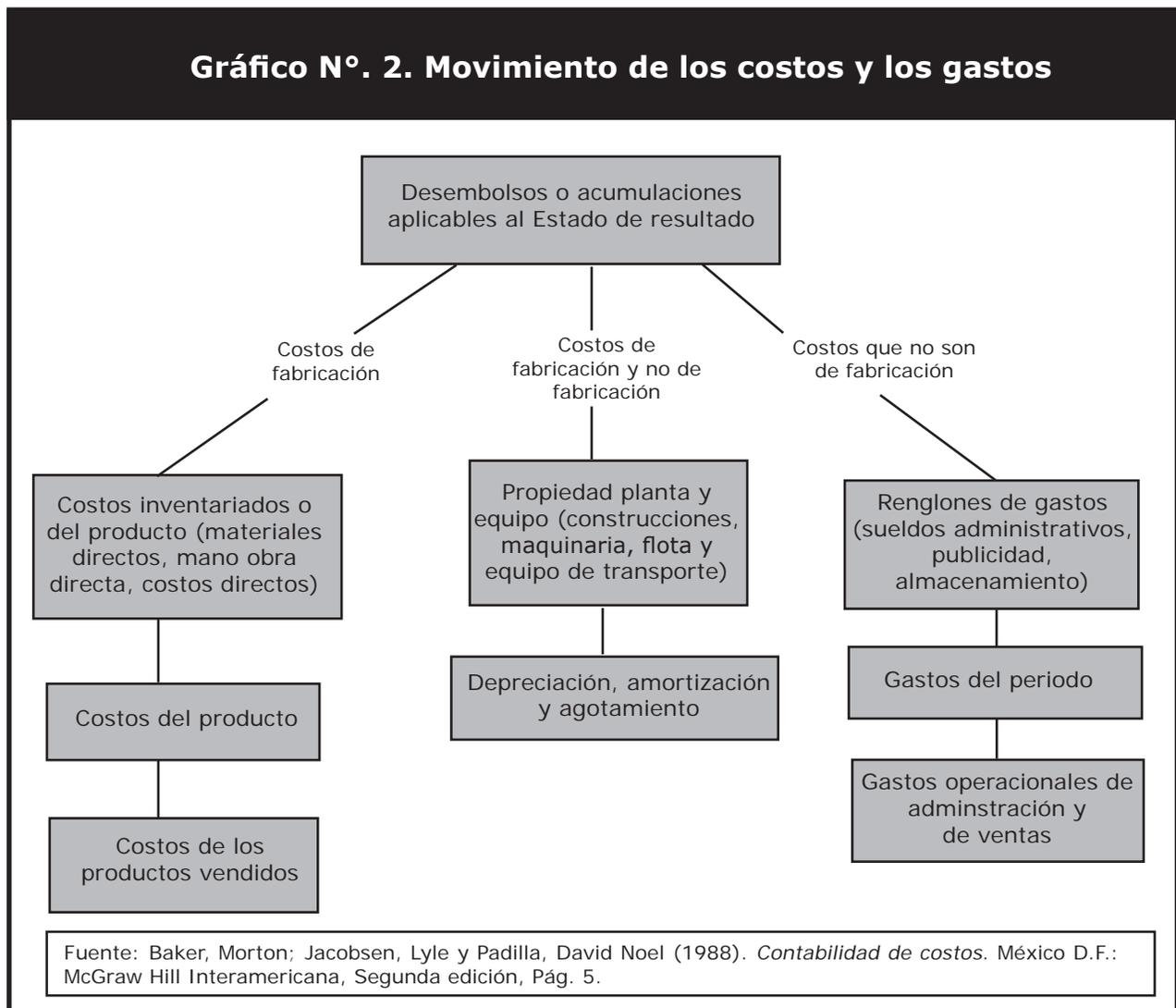


- a. Los desembolsos pueden gastarse y compararse con los ingresos del periodo (gastos operacionales de administración y ventas).
- b. Los desembolsos pueden ser capitalizados como activos y posteriormente ser susceptibles de depreciarse, amortizarse o agotarse (propiedad planta y equipo, activos diferidos).
- c. Los desembolsos pueden ser inventariados o ser tratados como costos del producto, que cuando se vendan, se traducirán en el costo de ventas. (Ver Gráfico N°. 2).

## 7. CLASIFICACIÓN DE LOS COSTOS

Retornando a Mallo, dentro del ámbito de la contabilidad y posteriormente de la economía de empresa, los costos han sido clasificados atendiendo a diversos criterios. Analicemos la clasificación presentada por Ramírez Padilla<sup>8</sup> y por Baker y Jacobsen<sup>9</sup>.

### 1. De acuerdo con la función en la que incurren:



8 Ramírez Padilla, David Noel (2005). *Contabilidad Administrativa*. México, D.F.: Mc Graw-Hill Interamericana, Séptima edición, Pág. 35-41.  
9 Op. cit.

## **Costos de producción**

Definidos como aquellos que intervienen directamente en el proceso de transformación de la materia prima en productos terminados. Se clasifican en:

1. Costo de la materia prima: son aquellos integrados al producto, tales como la madera en la fabricación de muebles.
2. Costo de la mano de obra: integrado por el valor que se paga a quienes intervienen en la transformación directa de los materiales (en Colombia, se incorpora el valor de los salarios, las prestaciones sociales y los aportes parafiscales o patronales).
3. Costos indirectos: todos aquellos costos, diferentes a los dos anteriores, en los que se incurre para la elaboración del producto tales como los salarios de supervisión, depreciaciones y mantenimiento, entre otros.

## **Costos de distribución o venta**

Generados en el departamento de mercadeo, el cual es el encargado de la comunicación y de la entrega de valor a los clientes, tales como la publicidad, los salarios de los vendedores, etcétera.

## **Costos de administración**

Originados en el área administrativa (salarios, depreciaciones, servicios públicos e impuestos sobre propiedad).

## **Costos de financiamiento**

Corresponden al costo que se paga por el uso de recursos provenientes de fuentes alternas de financiación. Vale la pena destacar que tanto los costos de distribución o venta, como los costos de administración y de financiamiento se tratan como gastos del periodo, los cuales podrán ser enfrentados a los ingresos de ese periodo para obtener bien sea la utilidad o la pérdida del ejercicio contable.

2. De acuerdo con su identificación con una actividad, departamento o producto:

### **Costo directo**

Es el que se identifica plenamente con la actividad, departamento o producto; es decir, “con el objeto de costo en particular, siendo por tanto perfectamente rastreable desde el punto de vista económico”<sup>10</sup>. Por ejemplo, los salarios del personal de ventas para el departamento de mercadeo, así como la mano de obra cancelada a los operarios, es un costo directo de producción.

### **Costo indirecto**

Este tipo de costo, está relacionado con un objeto de costo particular, pero no puede rastrearse de manera económicamente factible. Tal es el caso de la depreciación de los activos involucrados en el proceso y la remuneración de los supervisores.

3. De acuerdo con el tiempo en que fueron calculados:

### **Costos históricos**

Denominados así pues ocurrieron en un determinado periodo.

### **Costos predeterminados**

Su cálculo se hace empleando métodos estadísticos y normalmente se emplean en la presupuestación.

4. De acuerdo con su comportamiento:

### **Costos variables**

Estos costos cambian en función directa con el nivel de actividad o volumen.

### **Costos fijos**

Estos permanecen constantes dentro de un periodo de tiempo, independientemente del nivel de actividad.

10 Horngren, Charles; Datar, Srikant y Foster, George (2007). *Contabilidad de costos*. México: Pearson Educación, Decimosegunda edición, Pág. 27.

De lo anterior, se concluye que los costos se definen como variables; o fijos, con respecto a una actividad específica y por un periodo determinado.

Además de las clasificaciones anteriores, se encuentran otras formas que corresponden a criterios como la importancia en la toma de decisiones, el tipo de costo incurrido y el cambio originado por las variaciones en el nivel de actividad.

## 8. EL COSTO DE FABRICACIÓN

Los tres términos para describir los costos de fabricación son: costos de materiales directos, costos de mano de obra directa de fabricación y costos indirectos de fabricación<sup>11</sup> (ver Gráfico 3).

1. Los materiales directos son aquellos insumos que entran directamente en la producción, perfectamente identificables dentro del producto, pues lo constituyen tanto física como económicamente. Los costos de los materiales directos incorporan no sólo los costos de adquisición, sino todos aquellos factores o causantes del costo, tales como fletes, gastos de nacionalización, bodegajes, seguros, etcétera. Nótese también cómo el costo de los materiales directos consumidos en el periodo estarán determinados por el método de valuación de inventarios que la organización haya escogido (PEPS, UEPS, promedios).

2. Los costos de la mano de obra directa incluyen la remuneración pagada a los operarios de producción quienes son los que realizan directamente la transformación de la materia prima en productos terminados. En Colombia el valor de la mano de obra directa se integra con tres componentes a saber: los salarios devengados, el valor de las prestaciones sociales y los aportes a cargo del empleador.

3. Los costos indirectos de fabricación, como se mencionó anteriormente, están relacionados con el proceso de producción, pero son diferentes a la materia prima directa y a la mano de obra directa. “Los costos indirectos

pueden dividirse en tres categorías según el objeto del gasto así: materiales indirectos, mano de obra indirecta y los otros costos indirectos de fabricación”<sup>12</sup>.

## 9. SISTEMA DE COSTOS POR ÓRDENES DE PRODUCCIÓN

### 9.1. Concepto de sistema

Antes de adentrarnos en el sistema de costeo por órdenes, se hace necesario puntualizar algunos aspectos básicos de todo sistema de costeo a saber: Hansen y Mowen<sup>13</sup> definen el sistema como una serie de piezas interrelacionadas que desempeñan un proceso o más, a fin de alcanzar objetivos específicos; y los sistemas contables tienen el mismo patrón general: 1) partes interrelacionadas; 2) procesos, y 3) objetivos.

Las partes interrelacionadas consisten en puntos como registro de órdenes y ventas, facturación, inventarios, siendo cada uno de estos un sistema en sí, y un subsistema de la información contable. Los procesos implican el desarrollo de actividades de recepción, clasificación, ordenamiento, codificación, registro y en general, manejo de datos. El objetivo de ese sistema de información es suministrar información tanto a usuarios internos como externos para la toma de decisiones.

Ahora bien: los sistemas de costeo, en lo referente a la función de producción se definen como un:

Conjunto de procedimientos, técnicas, registros e informes estructurados sobre la base de la partida doble y otros principios técnicos, que tienen por objeto la determinación de los costos unitarios de producción y el control de las operaciones fabriles efectuadas<sup>14</sup>.

“Un sistema de costeo por lo general representa los costos en dos etapas básicas: la acumulación, seguida de la asignación”<sup>15</sup> anotan Horngren, Datar y Foster. La acumulación del costo es la recopilación de información de costos en forma organizada a través de un sistema contable. La asignación del costo es un término general

11 Ídem. Pág. 37.

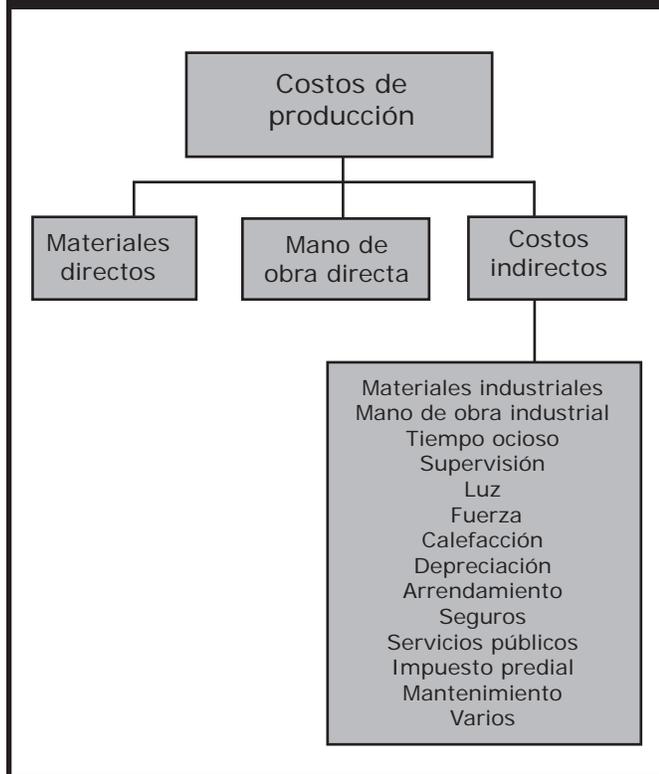
12 Baker, Morton; Jacobsen, Lyle y Padilla, David Noel (1988). Op. cit., Pág. 138.

13 Hansen, Don y Mowen, Maryanne (1996). *Administración de costos*. México, D.F.: Internacional Thomson Editores, S.A., Primera edición, Pág. 30.

14 Ortega Pérez de León, Armando (1994). *Contabilidad de costos*. México, D.F.: Editorial Limusa, S.A. de C.V. Quinta edición, Pág. 33.

15 Horngren, Charles, Datar, Srikant, Foster, George (2007). Op. cit., Pág. 27.

### Gráfico N°. 3. Costos de producción



que abarca: 1) el rastreo de los costos acumulados que tienen una relación directa con el objeto del costo, y 2) el prorrateo de costos acumulados que tienen una relación indirecta con el objeto del costo.

“En la construcción lógica de un sistema de costos es necesario utilizar con criterio muy estricto el principio de causalidad”<sup>16</sup>, dicen Mallo, Kaplan y Meljem; esto es, que la actividad comercial o industrial representa la causa y el fin por y para el que existen otras funciones en la empresa; por tanto, los portadores de los costos finales son portadores industriales o comerciales. Continúa aclarando Mallo, que es conveniente advertir que, aunque sea por extensión, puede considerarse portador industrial a todo aquel que refleje el costo de transformación cuyo valor quede incorporado en obje-

tos materiales cuantificables, y que el portador comercial es generalizable a aquel costo cuyo valor no permanece almacenado, sino que expira en el periodo en que se produce, y debe ser reintegrado por los ingresos del mismo periodo. Ahora bien, las condiciones muy particulares de cada tipo de organización demandarán la adopción de sistemas de costos igualmente particulares que se ajusten a esos requerimientos, siendo en todo caso menos complejos los costos de las comercializadoras habida consideración de no poseer actividad industrial.

El establecimiento de un sistema de costos para empresas, va de la mano con las características del proceso de fabricación que se tenga. Un buen sistema de información, flexible y confiable, debe dar información suficiente para una variedad de propósitos y debe responder a una variedad de preguntas.

Veamos algunos aportes de tratadistas contables :

“La asignación del costo es un término general que abarca: (1) el rastreo de los costos acumulados que tienen una relación directa con el objeto del costo, y (2) el prorrateo de los costos acumulados que tienen una relación indirecta con el objeto del costo”<sup>17</sup>.

“Los sistemas de costeo registran el costo de los recursos adquiridos, tales como materiales, mano de obra y equipo, e identifican cómo se utilizan esos recursos para producir y vender productos y servicios”<sup>18</sup>.

“Por objeto de costo entenderemos todo aquello para lo que sea necesaria una medida de costos, llámese producto o servicio”<sup>19</sup>. Los centros de responsabilidad, otro objeto de costo, son partes, segmentos o subunidades de una organización cuyos gerentes y responsables de actividades específicas, son los departamentos o subdepartamentos, algunos de los ejemplos más comunes.

Teniendo en cuenta que dentro de la industria, la producción puede darse en forma ininterrumpida (en serie) o bien, en forma interrumpida, lotificada o diversificada (por órdenes), los sistemas contables deben adecuarse a las necesidades de la organización, dando origen a dos métodos o sistemas de acumulación de costos como son, los costos por procesos y los costos

16 Mallo, Carlos, Kaplan, Robert, Meljem, Silvia (2000). Op. cit., Pág. 57.

17 Horngren, Charles, Datar, Srikant, Foster, George (2007). Op. cit., Pág. 27.

18 Ídem. Pág. 98.

19 Ídem. Pág. 30.

por órdenes de producción. En el primer caso, los costos por procesos se usan en organizaciones donde la producción no está sujeta a las interrupciones y diversificaciones particulares de un sistema por pedidos específicos, sino que se desarrolla en forma continua, mediante la afluencia constante de materiales a los procesos de transformación, tal como en la industria de medicamentos y de producción de vidrio, entre otras. El sistema de costos por pedidos u órdenes, se emplea en industrias de producción en función de instrucciones específicas de los clientes, como sucede en las empresas editoriales, de artes gráficas y astilleros.

## 9.2. Modelos contables de costos

Al respecto, los tratadistas contables emplean en forma indistinta los términos “modelos contables de costos”, “métodos de costeo”, “sistemas de costeo”, para referirse a la filosofía, práctica o al tratamiento contable de los costos.

Ramírez Padilla señala:

...en el periodo de la industrialización, sólo se daba importancia al valor de los materiales y la mano de obra directas (costo primo) habida consideración de ser los únicos que se inventariaban, pero es con la expansión de los centros de producción y el desarrollo de las organizaciones que surge el concepto de costos indirectos de fabricación, haciéndose notable la diferencia entre costos fijos y variables de producción y los costos directos e indirectos de producción<sup>20</sup>.

El modelo de costeo empleado para la época y empleado hoy todavía es el **full o absorption costing** (costeo total o completo o por absorción), el cual centra su atención en el costeo de los productos o inventarios. A mediados de la década de los treinta del siglo pasado aparece una metodología denominada **direct costing** (costeo directo o variable o costeo parcial), propuesta por Harris y Harrison<sup>21</sup>.

Cada uno experimentó con su método, el primero en la Dervey and Almy Chemical Company, de Cambridge, y sus resultados fueron publicados en el boletín de la Asociación Nacional de Contadores de Costos en 1936 en el artículo “Cuánto ganamos en el mes pasado”. En-

tre tanto, Harrison en 1935 motu proprio en su folleto *Vino joven en botellas viejas. ¿Por qué la mayoría de los estados financieros están equivocados?* presenta las bases del nuevo método.

Finalmente, “Los modelos contables que actualmente contempla la literatura parten de la combinación, con diferente dosificación de diversos criterios clasificativos”<sup>22</sup>, los cuales se presentan en Tabla N°. 1.

### 9.2.1. Costeo directo y costeo absorbente

El costeo total o por absorción, básicamente es un sistema de inventarios que incorpora al costo del producto, tanto los costos fijos como variables de producción, con el propósito de costear sus inventarios, excluyendo todos los costos que no son de fabricación.

El costeo variable o directo es un sistema de costeo de inventarios en el que todos los costos variables de fabricación se incluyen como inventariables; es decir, que al costo del producto sólo se cargan los costos variables ya que en estos costos se incurre de manera directa en la fabricación del producto, en tanto que los costos fijos de producción se consideran gastos del periodo ya que están asociados con la capacidad para producir o vender independientemente del nivel de actividad que desarrolle la organización.

Veamos un resumen paralelo entre estos dos métodos en la Tabla N°. 2.

## 9.3. Aspectos preliminares del sistema de costos por órdenes de producción.

Este sistema conocido como sistema de costos por lotes específicos de producción, pedidos específicos u órdenes específicas de producción, es propio de organizaciones que desarrollan productos o servicios a la medida de las necesidades de los clientes, es decir, siguiendo las especificaciones muy particulares de estos para dichos bienes o servicios. Aldo Torres señala igualmente que existen a su juicio “dos factores que justifican la fabricación de una orden específica de trabajo cuales son: la falta de disponibilidad en el mer-

20 Ramírez Padilla, David Noel (2005). Op. cit., Pág. 220.

21 “Jonathan Harris (1934) en los Estados Unidos y G. Charter Harrison (1935) en Inglaterra, descubrieron la mencionada técnica del costo directo (materiales y labor directos), la cual no incluye obviamente, los costos indirectos de producción”. Del Río González, Cristóbal (2000). *Costos III*. México: ECAFSA, Tercera edición, Pág. 1-6.

22 Mallo, Carlos, Kaplan, Robert, Meljem, Silvia (2000). Op. cit., Pág. 56.

Tabla N°. 1. Modelos contables

Actividad económica de la empresa	Sistema de costeo	Estructura organizacional de la empresa	Acumulación y rastreo de costos	Objeto de costos
<ul style="list-style-type: none"> <li>• Comercial</li> <li>• Industrial</li> <li>• Servicios</li> <li>• Agrícola</li> <li>• Ganadera</li> <li>• Extractiva</li> <li>• Otras</li> </ul>	<ul style="list-style-type: none"> <li>• <i>Full costing</i> o costo completo</li> <li>• <i>Direct costing</i> o costeo parcial</li> </ul>	<ul style="list-style-type: none"> <li>• Orgánico</li> <li>• Por procesos</li> <li>• Por responsabilidad</li> <li>• Por actividad</li> <li>• Por factores</li> <li>• Inorgánico</li> </ul>	<ul style="list-style-type: none"> <li>• Histórico</li> <li>• Estimado</li> <li>• Estándar</li> <li>• Presupuesto global</li> </ul>	<ul style="list-style-type: none"> <li>• Unidad</li> <li>• Actividad</li> <li>• Pedido u orden</li> <li>• Producción en masa</li> </ul>

Tabla N°. 2.

Costeo absorbente <i>Total full costing</i>	Costeo directo o variable <i>Direct costing</i>
<ol style="list-style-type: none"> <li>1. Los costos fijos de producción hacen parte del costo del producto.</li> <li>2. Los costos del producto incluyen tanto costos fijos como variables.</li> <li>3. Se emplea para fines externos: Al costear inventarios, determina costo de artículos vendidos y por tanto, la utilidad del periodo que se reporta en sección patrimonial del balance general como estado financiero básico.</li> </ol>	<ol style="list-style-type: none"> <li>1. Los costos fijos de producción son considerados gastos del periodo.</li> <li>2. El costo de los productos (inventarios) se integra con los costos directos de producción.</li> <li>3. Se emplea para fines internos: la gerencia lo usa como instrumento de toma de decisiones, tales como el modelo costo-volumen-utilidad.</li> </ol>

cado y/o el cumplimiento de necesidad específica<sup>23</sup> y aclara: “Este sistema también es propio de empresas que producen sus artículos con base en el ensamblaje de varias partes hasta obtener un producto final, en el cual los diferentes productos pueden identificarse fácilmente por unidades o por lotes individuales<sup>24</sup>.”

Al respecto Ortega Pérez de León anota:

El sistema de costos por pedidos, se caracteriza por la posibilidad de notificar y subdividir la producción de acuerdo con las necesidades graduales establecidas por la dirección de la fábrica o, más concretamente, por el departamento de planeación de la producción y control de inventarios<sup>25</sup>.

Lo cual requiere entonces, de una planeación cuidadosa para lograr la utilización más económica del potencial humano y de la maquinaria.

## 10. PROCEDIMIENTO ADMINISTRATIVO ADMINISTRACIÓN DE LOS COSTOS

Parfraseando a Baker y Jacobsen, “la planeación de la producción comienza con el recibo de un pedido por parte del cliente. Esta es generalmente la base para la preparación y emisión a la fábrica de una orden de producción<sup>26</sup>.”

23 Torres, Aldo (2002). *Contabilidad de costos*. México: McGraw Hill Interamericana editores, S.A., Segunda edición, Pág. 50.

24 Gómez Bravo, Oscar (2005). *Contabilidad de costos*. Bogotá, D.C.: McGraw- Hill Interamericana, Quinta edición, Pág. 22.

25 Ortega Pérez de León, Armando (1994). Op. cit., Pág. 159.

26 Baker, Morton, Jacobsen, Lyle, Padilla, David Noel (1988). Op. cit., Pág. 221.

La orden de producción es un instrumento de acumulación y control de costos para cada pedido específico. En ella se van registrando los valores de los elementos del costo de producción; es decir, las materias primas directas empleadas, el valor de la mano de obra directa y los costos directos aplicados, información esta que nos permite determinar el costo total de producción de cada lote o pedido específico y, por consiguiente, el costo de producción de cada unidad fabricada. Este documento, entre otros datos, contiene información e instrucciones detalladas acerca del tipo de producto, el proceso de transformación (ruta de proceso), la tecnología necesaria en la transformación, tiempos de proceso, etcétera. El anexo 1 presenta un modelo de hoja de costos.

El flujo de los costos debe ir paralelo con el flujo de producción, dado que los costos se asignan a los productos como resultado de las actividades de manufactura.

El ciclo de los costos se inicia cuando desde el almacén de materiales se les solicita a los proveedores las diferentes cantidades de materia prima que han de surtir al proceso productivo. Para estos menesteres, la empresa dispone de un registro de proveedores quienes suministrarán las referencias, cantidades, precios y oportunidad de entrega de estos insumos. Procedimentalmente, el almacén emitirá una orden de pedido con las especificaciones respectivas en original y varias copias, las cuales se distribuirán así: original para el proveedor, copia para almacén y otras para dependencias interesadas (contabilidad, etc.).

Cuando el proveedor despacha las mercancías, remite al comprador la factura de venta de estas; el almacén procede a la verificación de lo facturado con lo solicitado y a la recepción de las materias primas. En este paso, el almacén procede a elaborar una orden de entrada de mercancías y a su disposición en el almacén de materias primas. (Ver modelo de orden de entrada a almacén).

La factura de compra suministrada por el proveedor (original y dos copias) es remitida al departamento financiero para proceder a su contabilización (cuentas por pagar a proveedores e inventarios). Es pues, en el departamento contable donde se procede a determinar el valor de la entrada al almacén, teniendo en cuenta no sólo el valor de la factura, sino la adición de otros

factores de costo en los cuales se ha incurrido para dejar en condiciones de incorporar al proceso productivo los materiales tales como: gastos de nacionalización, fletes, seguros y bodegaje, entre otros. Es también en el departamento de contabilidad donde se determina tanto el método como el sistema de registro y valuación de los inventarios.

Para dar inicio al proceso de producción, la fábrica solicita al almacén las materias primas necesarias mediante una requisición de materiales. El anexo 2 es un modelo de requisición de materiales. Vale la pena destacar que el departamento de producción, y teniendo en cuenta las políticas de inventario, previamente ha solicitado la adquisición de los materiales e insumos necesarios para el proceso de producción de acuerdo con el portafolio de producto ofrecido. Los anexos 3, 4 y 5 nos presentan los modelos del portafolio de materias primas, del portafolio de productos y de la orden de compra de materiales. Posteriormente, el almacén procede a despachar lo solicitado, empleando para ello las órdenes de salida de almacén. Ver anexo 6.

La aplicación y el control de la mano de obra empleada en el proceso se hacen teniendo en cuenta las disposiciones legales que en materia laboral existen en nuestro país. Al respecto anotaremos que el valor de la materia prima está integrado por tres componentes a saber: el valor devengado por los trabajadores y que constituyen el salario; los aportes de ley (parafiscales o aportes patronales) y finalmente, el componente prestacional. El anexo 7 nos presenta los instrumentos mediante los cuales se hace el proceso de manejo de la mano de obra.

Finalmente, para los costos indirectos de fabricación, se emplea el procedimiento de la aplicación de los costos, esto ante la imposibilidad de rastrear y asignar cada costo indirecto de fabricación a cada orden o pedido específico, se aplica teniendo en cuenta bases predeterminadas: el costo de los materiales directos consumidos, el valor de la mano de obra directa utilizada, o el costo primo. Una vez se conozca, al final del periodo, el monto de los costos indirectos reales, se procederá a determinar la sobre o subaplicación de los costos indirectos, la cual se llevará como menor o mayor valor del costo de los artículos vendidos respectivamente.

## 11. RUTA DEL PROCESO

A continuación se presenta un resumen de la ruta que sigue el proceso de costos en el acompañamiento empresarial.

### 1. Diseño y diligenciamiento de los documentos de entrada:

- Diligenciar portafolio de producto.
- Diligenciar portafolio de materias primas.
- Diligenciar base de datos de proveedores.
- Diligenciar base de datos de clientes.
- Diligenciar base de datos de empleados.
- Diligenciar datos de entrada de la nómina.
- Diligenciar hoja de costos por órdenes.
- Diligenciar datos de operarios de fábrica.

### 2. Diseño y diligenciamiento de información de kardex:

- Hoja de kardex por materiales.
- Hoja de kardex por productos.
- Hoja de control de kardex por productos.
- Hoja de control de costos por materiales.

### 3. Hoja de control de costos por órdenes

### 4. Rutina de materiales directos:

- Diligenciar hoja de orden de compra de materiales.
- Diligenciar hoja de requisición de materiales.
- Diligenciar hoja de salida de materiales.
- Reflejar su efecto sobre el kardex y sobre la hoja de costos por órdenes y sobre la hoja de control.
- Reflejar el traslado de la orden terminada en la hoja de kardex por producto terminado.

### 5. Rutina de mano de obra directa:

- Diligenciar la base de datos de empleados.
- Diligenciar las tarjetas de tiempo o de reloj.
- Totalizar el movimiento mensual, quincenal o semanal por empleado para llevarlo a la hoja de liquidación de nómina.
- Liquidación de la nómina.
- Liquidación de los aportes de ley y de las prestaciones sociales.
- Consolidar la información del devengado con aportes de ley y prestaciones sociales.

g. Distribuir la nómina en función del número de horas trabajadas en cada una de las órdenes (podría ser en forma porcentual).

h. Cargar el valor de la mano de obra directa en la hoja de costos por cada orden afectada y consolidarse en la hoja control de costos por órdenes.

### 6. Rutina de costos indirectos. Este elemento del costo se aplicará a cada orden de producción así:

- Se escogerá una base de aplicación :
  - Materiales directos empleados en cada orden
  - Mano de obra directa utilizada
  - Costo primo
- Se determinará el porcentaje (%) de aplicación más adecuado según el tipo de producto (industria) y la experiencia del empresario.
- Se aplicará el porcentaje (%) a la base y este valor se llevará a cada orden de fabricación y, por ende, al control de órdenes de producción.
- Nota: La base se aplicará uniformemente para las órdenes de fabricación.

### 7. Rutina de las ordenes de producción:

- Cada hoja de costos se ha ido alimentando de la información de materiales directos, mano de obra directa y costos indirectos aplicados.
- Una vez concluida la orden, se totaliza cada elemento del costo y se determina el costo total de producción, así como el costo unitario de fabricación.
- Determinado el costo unitario de producción, éste será uno de los varios criterios que el empresario puede emplear en la fijación de precios.

## 12. EL ESTADO DE COSTO DE PRODUCCIÓN

Finalmente, y para concluir, la información de costos queda plasmada en el estado de costo de producción y ventas.

El estado de costo de producción y ventas, es un informe de propósito especial que tiene por objetivo presentar, en forma clara, el monto de los recursos financieros que la empresa ha invertido en la producción de bienes en un período determinado. En este estado financiero se reflejan los costos de la materia prima directa con-

sumida, el valor de la mano de obra directa empleada en el período, así como los costos indirectos involucrados (materia prima indirecta, mano de obra indirecta y otros costos indirectos tales como depreciaciones, servicios públicos, mantenimiento, etc.).

Los dos primeros elementos del costo son denominados directos por su fácil identificación en el producto, en tanto que los costos indirectos no se pueden identificar directamente, pero son igualmente necesarios en la elaboración del bien. Estos costos indirectos normalmente se clasifican en fijos y variables, siendo llamados los primeros también costos de capacidad, en atención a que la empresa tiene que asumirlos independientemente del nivel de actividad, en tanto que los segundos o sea los costos indirectos variables, como su nombre lo indica, sólo se presentan en función de ese nivel de producción.

En el estado de costo de producción y ventas vemos cómo la sumatoria de la materia prima directa, la mano de obra directa y los costos indirectos se denominan **cargos a producción**, los cuales se le suma el monto del inventario inicial de productos en proceso y se le resta el valor del inventario final de productos en proceso para hallar el valor del **costo de producción**.

Al costo de producción se le adiciona el valor del inventario inicial de productos terminados, obteniéndose el **costo de la mercancía disponible para la venta**, al cual se le resta el valor del inventario final de productos terminados, que da como resultado el **costo de la mercancía vendida o costo de ventas**.

El modelo del estado de costo de producción y ventas se presenta en el Gráfico 4.

## BIBLIOGRAFIA

- BAKER, Jacobsen y otros (1998). *Contabilidad de costos. Un enfoque administrativo para la toma de decisiones*. México D.F.: Mc Graw Hill, Segunda edición.
- CHATFIELD, Michael (1979). *Estudios contemporáneos sobre la evolución del pensamiento contable*. México: ECASA.
- DEL RÍO GONZÁLEZ, Cristóbal (1999). *Costos III*. ECAFSA., Tercera edición.
- GÓMEZ, Oscar (2005). *Contabilidad de costos*. México D.F.: Mc Graw Hill, Quinta edición.
- HANSEN y MOWEN (1996). *Administración de costos*. México, D.F.: Internacional Thomson Editores, S.A., Primera edición.
- HORNGREN, Charles; DATAR, Srikant y otros (2007). *Contabilidad de costos*. México: Pearson Educación, Decimosegunda edición.
- KAPLAN, Robert y COOPER, Robin (1998). *Coste y efecto*. Barcelona: Gestión 2000, Segunda edición.
- MALLO, Carlos; KAPLAN, Robert y otros (2000). *Contabilidad de costos y estratégica de gestión*. Madrid: Prentice Hall, Primera edición.
- ORTEGA PÉREZ DE LEÓN, Armando (1994). *Contabilidad de costos*. México, D.F.: Editorial Limusa, S.A. de C.V. Quinta edición.
- RAMÍREZ PADILLA, David Noel (2005). *Contabilidad administrativa*. México, D.F.: Mc Graw Hill Interamericana, Séptima edición.
- TORRES, Aldo (2002). *Contabilidad de costos*. México: Mc Graw Hill, Segunda edición.
- SINISTERRA, Gonzalo (2006). *Contabilidad de costos*. Bogotá: Ecoe Ediciones.

## ANEXOS

Ver página 33.

## Gráfico N°. 4. Modelo del estado de costo de producción y ventas

COMPAÑÍA PRODUCTORA S.A.  
ESTADO DE COSTOS DE VENTAS (en millones de pesos)  
PERÍODO COMPRENDIDO ENTRE 01-01- AL 31-12 DE 200X

### Materia prima directa consumida

Inventario inicial de materia prima	XXXXX
Compras netas de materia prima	XXXXX
Materia prima disponible	XXXXX
Inventario final de materia prima	(XXXXX)
Materia prima consumida	XXXXX
Materiales indirectos utilizados	(XXXXX)
Materia prima directa consumida	XXXXX

**Mano de obra directa utilizada** XXXXX

**Costos indirectos** XXXXX

**Cargos a producción** XXXXX

Inventario inicial de productos en proceso XXXXX

Inventario final de productos en proceso XXXXX

### Costos de producción

Inventario inicial de productos terminados XXXXX

Compras de mercancías XXXXX

Mercancía disponible para la venta XXXXX

Inventario final de productos terminados (XXXXX)

**Costos de ventas** XXXXX

\_\_\_\_\_  
Gerente General

\_\_\_\_\_  
Contador

\_\_\_\_\_  
Revisor Fiscal

## Anexo N°. 1. Modelo de la hoja de costos por orden de producción

### HOJA DE COSTOS POR ORDEN

#### NOMBRE DE LA COMPAÑÍA

Cliente \_\_\_\_\_ Nit \_\_\_\_\_ Dirección \_\_\_\_\_  
 Teléfono \_\_\_\_\_ Fax \_\_\_\_\_  
 Artículo \_\_\_\_\_ Referencia \_\_\_\_\_  
 Cantidad \_\_\_\_\_  
 Fecha de inicio \_\_\_\_\_ Orden de trabajo N°. \_\_\_\_\_  
 Fecha de entrega \_\_\_\_\_

Fecha	Concepto	Materiales directos	Mano de obra directa	Costos indirectos	Total
<b>Total</b>					

RESUMEN DEL COSTO	\$
Materiales directos	
Mano obra directa	
Costos indirectos	
Costo total	
Costo unitario	

**Responsables** \_\_\_\_\_

\_\_\_\_\_

**Auditoría** \_\_\_\_\_



**Anexo N°. 4. Portafolio de producto**

**PORTAFOLIO DE PRODUCTO**

**NOMBRE DE LA COMPAÑÍA**

Código interno	Código barras	Categoría o clase de producto	Línea de productos	Tipo de producto	Artículo o referencia

Diseño de un sistema de costos para pymes

**Anexo N°. 5. Solicitud de compra de materiales**

**SOLICITUD DE COMPRA DE MATERIALES**

**NOMBRE DE LA COMPAÑÍA**

Ciudad \_\_\_\_\_ Fecha \_\_\_\_\_ N°. \_\_\_\_\_

Proveedor \_\_\_\_\_ Nit \_\_\_\_\_ Dirección \_\_\_\_\_

Descripción artículo	Referencia	Unidad	Cantidad	Fecha de entrega	Observaciones

Solicitó \_\_\_\_\_ Aprobó \_\_\_\_\_ Auditoría \_\_\_\_\_

# Predicción con series de tiempo y regresión

EFRAÍN MORENO SARMIENTO

## 1. INTRODUCCIÓN

Diariamente los dueños de empresa se ven enfrentados al problema de pronosticar el nivel futuro de su actividad económica. Pronósticos de ventas, utilidades, inversiones y costos entre otros, se constituyen en elementos importantes de la planificación de los negocios; sin embargo para lograrlos, se necesita conocer y disponer de información confiable y suficiente del pasado; estudiarla y detectar tendencias, variaciones cíclicas y

estacionales, con el fin de poderlas reproducir a través de modelos estadísticos. Este campo de los modelos estadísticos ha crecido al mismo ritmo del desarrollo de los computadores de tal forma, que existe una gran cantidad de modelos de pronósticos. Unos sencillos, como los que inicialmente se consideran en este documento y de gran despliegue en los textos clásicos de la Estadística básica, los tradicionales de la regresión

### RESUMEN

En el siguiente documento se desarrollan algunos modelos básicos de pronóstico con base en series cronológicas, regresión lineal, exponencial y parabólica de amplio despliegue en los textos clásicos de la Estadística, utilizando en todos los casos las herramientas que ofrece la hoja electrónica Excel. Este trabajo resalta la importancia de conocer la evolución y tendencia de las variables principales de los negocios, conocimientos que manejan de manera intuitiva los dueños de pequeñas y medianas empresa, pero que algunas veces es necesario traducirlos a lenguajes abstractos, mediante modelos estadísticos, que permitan generalizar y puedan ser utilizados de manera fácil y económica como instrumento para la toma de decisiones en situaciones futuras.

### PALABRAS CLAVE

Predicción con series de tiempo, valores de tendencia mensual, valores desestacionalizados, índices estacionales en la predicción, modelos de regresión lineal, exponencial y parabólico.

### ABSTRACT

This document develops some basic models of forecast based on time series, linear regression and exponential and parabolic regression, known widely in classic statistics textbooks, and their usages are supported with an Excel worksheet.

The importance of knowing the evolution and tendency of the main variables in businesses are highlighted in this paper. These concepts are managed intuitively for owners of median and small businesses, however it is necessary to translate them to abstract models by using statistical models. This type of models generalizes different kind of situations and can be used easy and economical way, as a tools to make decisions in future events.

### KEY WORDS

Forecast with times series, monthly tendency data, unseasonalized data, seasonal indexes in the forecast; parabolic, exponential, and linear regression models.

múltiple, otros más sofisticados como los modelos de series temporales propuestos por Box and Jenkins<sup>1</sup>; modelos que involucran redes neurales, o las propuesta de los premios Nobel de economía del año 2003, entre ellos, Robert F. Engle, de series temporales con volatilidad variante en el tiempo (modelos ARCH) y los de Clive W. J. Granger desarrollados con tendencias comunes (modelos de cointegración). En esta línea aparecen también los modelos generalizados GARCH estudiados por Bollersley (1986) y los modelos de volatilidad estocástica (SV) como alternativa a las falencias del uso de los modelos anteriores.

De todas formas estos modelos tienen unos denominadores comunes: la dependencia de una información confiable, adecuadamente caracterizada, normalizada y almacenada, que unida a la experiencia y pericia del investigador, puede generar buenas o malas estimaciones siempre y cuando las condiciones históricas sean muy similares a las que se pretenden modelar en el futuro.

## 2. DESCRIPCIÓN DEL PROBLEMA

Pronosticar es fundamental en las pequeñas y medianas empresas pymes puesto que las predicciones se pueden incorporar al proceso de toma de decisiones inteligentes. En particular, estas pequeñas empresas requieren pronósticos de muchos hechos y situaciones en todas las fases de su producción. En el sitio de trabajo del dueño, deben estar disponibles pronósticos confiables de las ventas para poder planificar las estrategias de la producción. En un programa de producción se requieren las predicciones de la demanda de cada producto. Estas proyecciones se elaboran para periodos concretos que pueden ser días específicos, semanas o meses. Estas estimaciones permiten a los propietarios poder planificar la producción y el mantenimiento del inventario. Los pronósticos de la demanda para cada producto se pueden leer en pronósticos de cantidades necesarias de materia prima, de modo que las compras se puedan planificar.

## 3. MARCO TEÓRICO

Como se señaló en la introducción, no hay un solo modelo para establecer pronósticos que sea el mejor. De hecho, existen bastantes métodos para realizar estimaciones que se puedan usar. Estos métodos se dividen en dos tipos clásicos: métodos cualitativos y cuantitativos.

### 3.1. Métodos cualitativos para establecer pronósticos

Son aquellos que valoran la experiencia y usan su opinión para establecer pronósticos que predicen, en forma subjetiva, hechos futuros. Son claves sobre todo cuando los datos históricos no están disponibles o son pocos. Las técnicas cualitativas son usadas a menudo para predecir cambios en los patrones de datos históricos.

Entre las técnicas cualitativas más representativas se encuentra el ajuste de curva subjetivo basado en el “ciclo de vida de un producto” de acuerdo con Bowerman, O`Connell y Koehler (2007)<sup>2</sup>. Se considera que este ciclo de vida consta de una primera etapa: el crecimiento, en el que las ventas comienzan muy lentamente, luego rápidamente. En la segunda etapa se consigue una estabilidad, sigue el aumento, pero a menor velocidad hasta estabilizarse, y en la última, aparece la declinación. La construcción subjetiva de estas curvas no es sencilla; requiere gran experiencia y criterio.

#### **Efraín Moreno Sarmiento.**

Estadístico de la Universidad Nacional, docente de tiempo completo e Investigador del Politécnico Grancolombiano, ex funcionario del sector financiero colombiano.

Para contactar al autor:  
emorenos@poligran.edu.co

1 Box, G. E. P. y Jenkins G. M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden Day Inc. Pág. 1.

2 Bowerman, B. L.; O`Connell, R. T. y Koehler, A. B. (2007). *Pronósticos, series de tiempo y regresión*. México: Torzón. Pág. 10.

El método Delphi se basa en el concurso de un grupo de expertos que de forma aislada resuelve un cuestionario sobre las actividades a pronosticar, luego se resumen las opiniones. Se espera que después de varias rondas de cuestionarios, las respuestas del grupo lleguen a un consenso que se utilizará como pronóstico. Rowe y Wright (2001)<sup>3</sup> desarrollan al detalle dicha metodología.

Una tercera técnica cualitativa requiere comparaciones técnicas independientes del tiempo; es decir, las transformaciones en un área se predicen mediante el monitoreo de cambios que tienen lugar en un área distinta, pero que están relacionadas (Gerstenfeld, 1971)<sup>4</sup>.

### 3.2. Métodos cuantitativos para establecer pronósticos

Estas técnicas necesitan el estudio de información histórica para estimar los valores futuros de la variable de interés. Estos modelos se pueden agrupar en dos clases: univariados y causales.

#### 3.2.1. Modelos univariados

Predicen el futuro de una serie con base en su comportamiento histórico propio; son muy útiles si el patrón detectado en el pasado se mantiene hacia el futuro, de lo contrario no son aconsejables. Los modelos *integrated autoregressive moving average model* (ARIMA *model*) son representativos de este grupo<sup>5</sup>.

#### 3.2.2. Modelos causales

Requieren la identificación de otras variables que se relacionan de la manera causa efecto con la variable que se desea predecir. Una vez identificadas estas variables relacionadas, se construye un modelo estadístico que pretende describir la relación entre estas variables y la variable que se desea pronosticar. Los modelos de regresión lineal simple y los modelos de regresión lineal múltiple son los más conocidos de este grupo.

## 4. METODOLOGÍA

### 4.1. Naturaleza de las series de tiempo

Una serie de tiempo es una serie de registros realizados en diversos periodos de tiempo (días, semanas, meses, trimestres, años). Los registros son valores numéricos que varían en el tiempo. Un aspecto básico del estudio de las series de tiempo requiere analizar la naturaleza de estas variaciones<sup>6</sup>.

Las variaciones de una serie de tiempo –ST– se clasifican en sistemáticas y aleatorias; las sistemáticas ocurren con regularidad y se pueden modelar; las aleatorias son causadas por situaciones aisladas como terremotos, huelgas y en consecuencia, son difíciles de modelar.

Los economistas han identificado tres diferentes tipos de variaciones: tendencia secular, variaciones cíclicas y variaciones estacionales.

La *tendencia secular* describe la naturaleza general de la serie en periodos largos de tiempo y son debidas a fuerzas importantes como crecimiento de la población, cambios en tecnología y en los hábitos de consumo de los consumidores, entre otras. Estas situaciones pueden desembocar en tendencias ascendentes, descendentes e incluso ninguna.

*Variaciones estacionales* se efectúan cuando las observaciones son realizadas en intervalos inferiores a un año (semanas, meses, trimestres); estas pueden reflejar comportamientos estacionales que se repiten de la misma manera y con la misma regularidad año tras año. El término estacional proviene de los países con estaciones; estas pueden influir en la evolución de los negocios; en nuestro medio dependen de las costumbres y hábitos que se manifiestan durante el año.

*Variaciones cíclicas* corresponden a fluctuaciones a largo plazo, más o menos periódicas que se repiten regularmente cada cierto número de años, en las actividades económicas en periodos de crisis y de recuperación.

3 Rowe, G. y Wright, G. (2001). "Expert opinions in forecasting: The role of the Delphi technique". En: *Principles of forecasting: A handbook for researchers and practitioners*. Boston: Kluwar.

4 Gerstenfeld, Arthur (1971). "Technological forecasting". En: *Journal of Business*. Vol. 44, No. 1.

5 Pindick, R. S. y Rubinfeld, D. L. (1991). *Econometric models and economic forecasts*. Sinagpore: McGraw Hill. Ver introducción. Pág. XVII.

6 Zuwaylif, F. H. (1977). *Estadística general aplicada*. México: Fondo Educativo Interamericano. Pág. 333.

*Variaciones aleatorias* se refieren a los factores que se presentan en forma accidental y son difíciles de predecir.

## 4.2. Tendencia secular y variaciones estacionales aplicadas a la predicción

### 4.2.1. Tendencia secular

Como ya se comentó, son aquellas variaciones suaves y constantes que se suceden en un período relativamente extenso. El período debe ser largo, mínimo de cinco años como para establecer una línea de tendencia significativa. Antes de tomar cualquier decisión se debe hacer un estudio exploratorio de la información que involucre gráficos de dispersión en el tiempo, de tal forma que se pueda visualizar algún tipo de relación. Las relaciones pueden ser diversas; este trabajo está desarrollado bajo el supuesto de que la relación es lineal.

### 4.2.2. Tendencia lineal

El modelo de regresión lineal es el apropiado para realizar pronósticos cuando la tendencia es lineal.

El objetivo del modelo de regresión de dos variables, el lineal, es la comprensión de la naturaleza probabilística del modelo de regresión de tal manera que, a partir de un valor observado de X (variable independiente), logremos observar varios valores posibles de Y (variable dependiente). El ejemplo desarrollado por Zuwaylif considera las ventas hipotéticas efectuadas en millones de pesos mensualmente. Estas probablemente variarán cada año, debido a la situación cambiante de los negocios, las preferencias de los clientes, la competencia, y en general, a los hechos que acompañan el desarrollo normal de las empresas. A no ser que se disponga de información adicional, supondremos que para cada observación de X (años), los comportamientos de Y (ventas) diferirán de forma aleatoria. Usamos  $\epsilon$  como componente de "error" aleatorio.

Para tal efecto se tomó una serie con periodicidad mensual del valor hipotético de las ventas mensuales de

la empresa XX entre enero de 1999 y diciembre del 2003. Ver Anexo N°.1.

A continuación se obtuvo el total de las ventas anuales en millones de pesos.

<b>Año</b>	<b>Ventas en millones de \$</b>
31/1/99	
31/12/99	218,7
30/12/00	235,3
30/12/01	246,3
31/12/02	261,6
31/12/03	283,9

Mediante un diagrama de dispersión se visualiza la posible relación lineal (ver Gráfico N°.1).

### 4.2.2.1. Obtención del modelo de regresión lineal

A partir de los datos obtenidos que se presentan en la Tabla N°.1 y realizando un cambio de base para la variable tiempo, como es tomar el año 1999 como período cero "0", el año 2000 como 1 y así sucesivamente hasta el último período de estudio, se obtiene la ecuación de regresión lineal, utilizando el menú Herramientas/Analizar Datos/Regresión que posee la hoja Excel. Los resultados se presentan en el Anexo N°. 2, en la hoja Ec\_t\_a.

Del anexo N°.2 obtenemos los parámetros de regresión para construir el modelo de regresión o la ecuación de tendencia anual (Ecta) para los diversos años de la serie:

#### Ecuación de tendencia anual

(1)

$$\hat{Y}_{\text{Ventas\_mills}} = \beta_0 + \beta_1 * X_{\text{años}} + \epsilon$$

Donde:

- $\hat{Y}_{\text{Ventas\_mills}}$  : Variable dependiente que se pretende estimar; en este caso, valor de las ventas anuales.
- $\beta_0$  : Parámetro de regresión correspondiente al valor estimado de las ventas en el año cero.
- $\beta_1$  : Parámetro de regresión correspondiente al crecimiento o disminución (depende del signo) del valor.
- $X_{\text{años}}$  : Variable independiente para el caso años.
- $\varepsilon$  : Componente de “error” aleatorio.

Mediante la ecuación (1), se pretende la comprensión de la naturaleza probabilística del modelo de regresión, en este caso, la ecuación de tendencia anual de manera que, cada observación Y es una variable aleatoria, el valor de X es fijo (conocido por el investigador) y  $\varepsilon$  es la componente de “error” aleatorio cuyo comportamiento se basa en una distribución de probabilidad subyacente<sup>7</sup>.

El objetivo de obtener esta ecuación radica en que a partir de los datos reales mensuales podamos obtener los resultados anuales, con el fin de estudiar y verificar la tendencia secular; es decir, el movimiento de la serie de tiempo en el periodo analizado y de esta manera predecir las ventas anuales en años futuros. Estas estimaciones serán razonables si existe una regularidad estadística que haya prevalecido en el pasado y continúe manifestándose aproximadamente de la misma forma en el futuro. Estas predicciones deben ser consideradas como información aproximada, razón de peso para la construcción de los respectivos intervalos de confianza dentro de los cuales se espera que esté el valor real de la variable de estudio, con una confianza que establezca el propio investigador.

Es importante que cada ecuación de tendencia esté acompañada del nombre y de la unidad de medida de la variable tanto dependiente como de la independiente.

### Cambio de base anual a base mensual (2)

$$\hat{Y}_{\text{Ventas\_men\_mills}} = \beta_0/12 + \beta_1/144 * X_{\text{mes\_jun/jul}} + \varepsilon$$

El objetivo de este cambio es necesario cuando exploremos las variaciones estacionales de la serie de tiempo, la fórmula (2) incorpora la tendencia secular anual observada en los datos, a cada uno de los meses, obteniéndose la ecuación de tendencia modificada mensual con origen en el mes promedio teórico que recoge 15 días de junio y 15 de julio y las correspondientes ventas mensuales calculadas por el modelo.

### Cambio de base a inicio de año (3)

$$\hat{Y}_{\text{Ventas\_men\_mills}} = (\beta_0/12 + \beta_1/144 * X_{\text{mes\_jun/jul}}) * (-5,5) + \beta_1/144 * X_{\text{mes\_enero}} + \varepsilon$$

Con esta ecuación determinamos los valores de tendencia mensual haciendo un cambio de base X=0 para enero de 1999; X=1 para febrero de 1999; X=59 para diciembre del 2003. Ver Anexo No. 3 columna **VTM<sub>i</sub>** valor de tendencia mensual. En atención a que la ecuación anterior (2) nos llevó a un mes que estaba compuesto por días de junio y julio, por comodidad se trasladada al mes inicio del año, enero.

En resumen la ecuación 3 permite obtener los valores de tendencia mensual de la variable de estudio, calculados por el modelo de acuerdo con los parámetros de regresión  $\beta_0$ ,  $\beta_1$  modificados.

### División de los valores reales de cada mes $y_i$ por el respectivo valor de tendencia mensual $VTM_i$

Se toman uno a uno los valores reales entre enero de 1999 y diciembre de 2003 y se dividen por sus correspondientes valores de tendencia mensual para obtener la columna  $y_i / VTM_i$ . Ver Anexo N.º. 3 columna

<sup>7</sup> Pindick, R. S. y Rubinfeld, D. L. (1991). Op. cit. Pág. 46.

$y_i / VTM_i$ . Este procedimiento elimina la variación por tendencia; en otras palabras, los valores reales de cada mes se expresan como porcentaje del valor de tendencia del respectivo mes.

### Determinación de la razón promedio de cada mes $rpm_i$

Las variaciones aleatorias de la serie se eliminan promediando las cinco razones de cada mes.

$$\overline{rpm}_i = \sum_{j=1}^n \sum_{i=1}^12 \frac{y_i}{VTM_{ij}} / n \quad (4)$$

Donde:

$\overline{rpm}_i$  : Se define como la razón promedio de cada mes.

$n$  : Número de años.

$i$  : Varía de 1 = enero a 12 = diciembre.

Ver Tabla N°. 2.

### Determinación de los valores desestacionalizados $VD_i$ mensuales

Es el resultado de dividir cada uno de los valores reales por su correspondiente índice estacional:

$$VD_i = y_i / ie_i \quad (5)$$

Ver anexo N°. 3 columna identificada con  $VD$ .

### Ventas desestacionalizadas a tasa anual $vdata_i$

$$vdata_i = VD_i * 12$$

**Tabla N°. 2.**

Mes	Razón media $rm$	Índice estacional $ie_i$
1	0,90155105	0,90250628
2	0,85278366	0,85368723
3	0,97347949	0,97451094
4	0,99387255	0,99492561
5	1,03139259	1,0324854
6	1,02382464	1,02490943
7	0,9948741	0,99592822
8	0,99492287	0,99597704
9	0,95329536	0,95430543
10	1,02658246	1,02767018
11	1,02144751	1,02252978
12	1,21927259	1,22056447
<b>TOTAL</b>	<b>11,9872989</b>	<b>12</b>

Corresponde a la multiplicación de cada uno de los valores mensuales desestacionalizados por los 12 meses del año. Estos valores desestacionalizados se utilizan para hacer análisis más rigurosos del comportamiento real de la variable que se esté pronosticando.

### Uso de los índices estacionales en la predicción

A partir del mes que se desee estimar y con la ecuación de tendencia mensual de las ventas obtenida en la fórmula (3) se encuentra el valor de tendencia para el mes deseado. Ahora bien, si en seguida este valor lo multiplicamos por el respectivo índice estacional registrado en la tabla 2, se obtiene la respectiva predicción. Al final del Anexo 3, soporte de la Tabla 2, se obtiene la respectiva predicción.

#### 4.2.3. Modelos de regresión exponencial

Si en el análisis exploratorio de la información se presentan crecimientos geométricos, que se pueden verificar con la correspondiente gráfica de dispersión, se deben utilizar los modelos o ajustes exponenciales. Ejemplos típicos: la población de un país, región o ciudad y aquellas variables que guardan relación con el

comportamiento de la población o la evolución de una enfermedad (propagación de virus). Las variables financieras, captación, préstamos que involucran el cálculo de rendimientos e intereses periódicos compuestos.

La ecuación general del modelo de regresión exponencial está dada por la siguiente expresión:

Donde:

$$\hat{Y} = \beta_0 \beta_1^X + \varepsilon$$

- $Y$  : Variable dependiente a estimar o valor futuro.
- $\beta_0$  : Parámetro de regresión o valor inicial o valor presente.
- $\beta_1$  : Parámetro de regresión o tasa de crecimiento o decrecimiento.
- $X$  : Variable independiente.
- $\varepsilon$  : Término de error.

Si se aplican logaritmos a ambos lados de la ecuación, el modelo se transforma en lineal, de tal forma, que la estimación de los parámetros de regresión y la obtención de las medidas para medir la bondad de ajuste se resuelven por los diferentes métodos que existen en la regresión lineal.

En el Anexo N.º. 4 se presenta el desarrollo para los dos modelos: el exponencial y por supuesto, el lineal con base en la misma serie de ventas mensuales del período enero de 1999 hasta diciembre del 2003<sup>8</sup>. En ambos casos se muestra el desarrollo manual utilizando el método de los mínimos cuadrados, y el desarrollo automático a través de las funciones de análisis de datos que posee la herramienta.

#### 4.2.4. Modelos de regresión parabólica

Si el diagrama de dispersión presenta comportamientos parabólicos o la línea poligonal presenta una parte ascendente y en seguida una descendente, o lo contra-

rio, situaciones de incertidumbre con las utilidades y los ingresos, se acostumbra trabajar con una curva de segundo grado.

La ecuación general del modelo de regresión parabólica es:

$$\hat{Y} = \beta_0 + \beta_1 * X + \beta_2 X^2 + \varepsilon$$

Las ecuaciones normales en este caso serían:

$$\sum y = n\beta_0 + \beta_1 \sum X + \beta_2 \sum X^2$$

$$\sum xy = \beta_0 \sum X + \beta_1 \sum X^2 + \beta_2 \sum X^3$$

$$\sum X^2 y = \beta_0 \sum X^2 + \beta_1 \sum X^3 + \beta_2 \sum X^4$$

Se estiman los parámetros de regresión  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$  con la ayuda de algún método algebraico y se reemplazan en la ecuación general del modelo parabólico de regresión para realizar las correspondientes estimaciones. El procedimiento se ilustra en el Anexo N.º. 5.

#### 4.2.5. Resultados

Los métodos de regresión lineal, exponencial y parabólico, así como los de predicción con series cronológicas, se constituyen en la metodología clásica utilizada por la estadística para realizar estimaciones. Son muy usados gracias a su fácil comprensión e implementación. Asimismo poseen unas medidas que permiten valorar la calidad de las estimaciones, entre las que se encuentran el coeficiente de correlación, el coeficiente de determinación y el error estándar de estimación, entre otros. Estos estadígrafos que miden la dependencia mutua entre las variables se deben calcular e interpretar de tal manera que se constituyen en el indicador principal para seleccionar el modelo a utilizar.

<sup>8</sup> Zuwaylif, F. H. (1977). Op. cit. Pág. 354.

## 5. CONCLUSIONES Y DISCUSIÓN

El presente trabajo pretende implementar la teoría básica de los modelos de regresión desarrollada en los textos, utilizando herramientas de fácil adquisición y utilización como la hoja de cálculo Excel, con el fin de democratizar estas metodologías y ponerlas al alcance de los dueños de las microempresas que no cuentan con muchos recursos para la compra de *software* especializado.

Estas herramientas son útiles siempre y cuando la información con la que se cuenta sea confiable, de lo contrario, se sugieren métodos de tipo cualitativo. Por último se espera despertar en la comunidad de las pequeñas y medianas empresas el interés por llevar un buen sistema de información que sea de su entera confianza a través de los años.

## BIBLIOGRAFÍA

- BOX, G.E.P. y JENKINS, G. M. (1976). *Time series analysis, forecasting and control*. San Francisco: Holden Day Inc.
- BOWERMAN, B. L.; O`CONNELL, R. T. y KOEHLER, A. B. (2007). *Pronósticos, series de tiempo y regresión*. México: Torzón.
- ROWE, G. y WRIGHT, G. (2001). "Expert opinions in forecasting: The role of the Delphi Technique". En: *Principles of forecasting: A handbook for researchers and practitioners*. Boston: Kluwar.
- GERSTENFELD, Arthur (1971). "Technological forecasting". En: *Journal of Business*. Vol. 44. N°. 1.
- PINDICK, R. S. y RUBINFELD, D. L. (1991). *Econometric models and economic forecasts*. Sinagore: McGraw Hill.
- ZUWAYLIF, F. H. (1977). *Estadística General Aplicada*. México: Fondo Educativo Interamericano.
- BENCARDINO, C. M. (2002). *Estadística y muestreo*. Bogotá: Ecoe.

### Anexo N°. 1. Ventas

Fecha	Ventas en millones de \$
31/1/99	15,8
28/2/99	15
31/3/99	17,9
30/4/99	17,4
31/5/99	18,5
30/6/99	18,9
31/7/99	17,9
31/8/99	18,3
30/9/99	18,1
31/10/99	18,8
30/11/99	19,2
31/12/99	22,9
31/1/00	17
28/2/00	16
30/3/00	19
29/4/00	19,3
30/5/00	20,2
29/6/00	20,3
30/7/00	19,1
30/8/00	19,9
29/9/00	18,9

Fecha	Ventas en millones de \$
30/10/00	20,6
29/11/00	20,9
30/12/00	24,1
30/1/01	18,3
27/2/01	17,1
30/3/01	19,7
29/4/01	20,5
30/5/01	21,2
29/6/01	20,7
30/7/01	20,5
30/8/01	21
29/9/01	19,2
30/10/01	21,5
29/11/01	21,5
30/12/01	25,1
30/1/02	19,2
28/2/02	18,8
31/3/02	20,5
30/4/02	21,2
31/5/02	22,5
30/6/02	22,2

Fecha	Ventas en millones de \$
31/7/02	22,1
31/8/02	21,8
30/9/02	21,3
31/10/02	22,6
30/11/02	21,7
31/12/02	27,7
31/1/03	20,6
28/2/03	19,6
31/3/03	21,9
30/4/03	23,5
31/5/03	23,8
30/6/03	23,8
31/7/03	24,1
31/8/03	23
30/9/03	22,7
31/10/03	25,1
30/11/03	25,2
31/12/03	30,6

Fuente: Fadil H. Zuwaylif. *Estadística General Aplicada*. México: Fondo Educativo Interamericano. Pág. 353

## Anexo N°. 2.

## Resumen

Estadísticas regresión	
Coefficiente correlación	0,993
Coefficiente determinación	0,986
R <sup>2</sup> ajustado	0,982
Error típico	3,378
Observaciones	5

Análisis de los residuales			
Observaciones	Pronóstico y	Residuos	Rdos. estándares
1	217,82	0,88	0,300852503
2	233,49	1,81	0,618798898
3	249,16	-2,86	-0,977770635
4	264,83	-3,23	-1,104265437
5	280,5	3,4	1,162384671

Análisis de varianza					
	Gras libertad	Suma cuadrados	Pro. de cuadrados	F	Vr crítico F
Regresión	1	2455,489	2455,489	215,2490	0,0007
Residuos	3	34,223	11,40766667		
Total	4	2489,712			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%
Intercepción	217,82	2,616218645	83,25756733	3,819E-06	209,4940	226,146
Variable x <sub>1</sub>	15,67	1,068066789	14,67136714	0,00069	12,2709	19,069

## Ecuación de tendencia anual Ec\_t\_a

$$\hat{Y} = \beta_0 + \beta_1 X + \varepsilon$$

$$\beta_0 = 217,82$$

$$\beta_1 = 15,67$$

x <sub>i</sub>	y <sub>i</sub>	$\hat{y}_i$	Residuos
0	218,7	217,82	0,88
1	235,3	233,49	1,81
2	246,3	249,16	-2,86
3	261,6	264,83	-3,23
4	283,9	280,5	3,4

Origen: 31/12/99 31/12/99

Unidad de x: anual

y: ventas anuales

## Anexo N° . 2.

### Cambio de base anual a base mensual

$$\hat{Y}' = \beta_0 / 12 + (\beta_1 / 144)X + \varepsilon$$

$$\beta_0 / 12 = 18,1516667$$

$$\beta_1 / 12 = 0,10881944$$

Origen: 16/06/99 16/07/99

Unidad de x: mensual

y: ventas mensuales

### Cambio de base de origen

$$\hat{Y}' = (\beta_0 / 12 + (\beta_1 / 144)(-5,5)) + \beta_1 / 144 X + \varepsilon$$

Ecuación de tendencia mensual

$$\beta_0 = 17,5531597$$

$$\beta_1 = 0,10881944$$

Factor cambio origen enero: 5,5

Origen: 31/01/99

Unidad de x: mensual

y: ventas mensuales

### Anexo N°. 3. Determinación del valor de tendencia mensual VTM<sub>i</sub> ; y<sub>i</sub> / VTM<sub>i</sub> ; VD ; vdata

Periodo	Fecha	VTM <sub>i</sub>	y <sub>i</sub> / VTM <sub>i</sub>	VD	vdata
0	31/01/1999	17,5531597	0,90012284	17,5068033	210,08164
1	28/02/1999	17,6619792	0,84928194	17,5708381	210,850057
2	31/03/1999	17,7707986	1,00727043	18,3681878	220,418254
3	30/04/1999	17,8796181	0,97317515	17,4887448	209,864937
4	31/05/1999	17,9884375	1,02843841	17,9179289	215,015147
5	30/06/1999	18,0972569	1,04435717	18,4406538	221,287845
6	31/07/1999	18,2060764	0,98318823	17,9731828	215,678194
7	31/08/1999	18,3148958	0,99918668	18,3739175	220,487011
8	30/09/1999	18,4237153	0,98242942	18,9666741	227,600089
9	31/10/1999	18,5325347	1,0144322	18,2938071	219,525686
10	30/11/1999	18,6413542	1,02996809	18,7769593	225,323511
11	31/12/1999	18,7501736	1,22132202	18,7618111	225,141733
12	31/01/2000	18,8589931	0,90142671	18,836434	226,037208
13	28/02/2000	18,9678125	0,84353428	18,7422273	224,906727
14	30/03/2000	19,0766319	0,99598294	19,4969591	233,96351
15	29/04/2000	19,1854514	1,0059706	19,3984353	232,781223
16	30/05/2000	19,2942708	1,04694291	19,5644413	234,773295
17	29/06/2000	19,4030903	1,0462251	19,8066281	237,679538
18	30/07/2000	19,5119097	0,97888932	19,1780889	230,137067
19	30/08/2000	19,6207292	1,01423346	19,9803803	239,764563
20	29/09/2000	19,7295486	0,957954	19,8049801	237,659762
21	30/10/2000	19,8383681	1,03839186	20,0453419	240,544102
22	29/11/2000	19,9471875	1,04776676	20,4395026	245,274031
23	30/12/2000	20,0560069	1,201635	19,7449627	236,939553
24	30/01/2001	20,1648264	0,90752083	20,2768672	243,322406
25	27/02/2001	20,2736458	0,84345954	20,0307554	240,369065
26	30/03/2001	20,3824653	0,96651704	20,2152682	242,583218
27	29/04/2001	20,4912847	1,00042532	20,6045556	247,254667
28	30/05/2001	20,6001042	1,02912101	20,532978	246,395736
29	29/06/2001	20,7089236	0,99956909	20,1969065	242,362878
30	30/07/2001	20,8177431	0,98473691	20,5838127	247,005753
31	30/08/2001	20,9265625	1,0035093	21,0848234	253,017881
32	29/09/2001	21,0353819	0,91274787	20,1193449	241,432139
33	30/10/2001	21,1442014	1,01682724	20,9211092	251,053311
34	29/11/2001	21,2530208	1,0116209	21,0262825	252,31539
35	30/12/2001	21,3618403	1,1749924	20,5642558	246,77107
36	30/01/2002	21,4706597	0,8942436	21,2740901	255,289082
37	28/02/2002	21,5794792	0,87119804	22,0221171	264,265405
38	31/03/2002	21,6882986	0,94521015	21,0361927	252,434313
39	30/04/2002	21,7971181	0,97260564	21,3081258	255,69751
40	31/05/2002	21,9059375	1,02711879	21,7920757	261,504908 →

### Anexo N°. 3.

→ 41	30/06/2002	22,0147569	1,00841449	21,6604505	259,925406
42	31/07/2002	22,1235764	0,99893433	22,1903542	266,284251
43	31/08/2002	22,2323958	0,98055109	21,8880548	262,656657
44	30/09/2002	22,3412153	0,95339487	22,3198982	267,838779
45	31/10/2002	22,4500347	1,00667996	21,9914916	263,897899
46	30/11/2002	22,5588542	0,96192829	21,2218759	254,66251
47	31/12/2002	22,6676736	1,22200454	22,6944178	272,333013
48	31/01/2003	22,7764931	0,90444126	22,8253259	273,90391
49	28/02/2003	22,8853125	0,8564445	22,9592284	275,510741
50	31/03/2003	22,9941319	0,95241691	22,4728108	269,673729
51	30/04/2003	23,1029514	1,01718606	23,6198564	283,438277
52	31/05/2003	23,2117708	1,02534185	23,0511734	276,614081
53	30/06/2003	23,3205903	1,02055736	23,221564	278,658768
54	31/07/2003	23,4294097	1,02862173	24,1985311	290,382373
55	31/08/2003	23,5382292	0,97713383	23,0929018	277,114822
56	30/09/2003	23,6470486	0,95995066	23,7869338	285,443206
57	31/10/2003	23,7558681	1,05658105	24,4241787	293,090144
58	30/11/2003	23,8646875	1,05595349	24,6447591	295,737109
59	31/12/2003	23,9735069	1,276409	25,0703676	300,844412

Mes	Razón media $rm_i$	Índice estacional $ie_i$
1	0,90155105	0,90250628
2	0,85278366	0,85368723
3	0,97347949	0,97451094
4	0,99387255	0,99492561
5	1,03139259	1,0324854
6	1,02382464	1,02490943
7	0,9948741	0,99592822
8	0,99492287	0,99597704
9	0,95329536	0,95430543
10	1,02658246	1,02767018
11	1,02144751	1,02252978
12	1,21927259	1,22056447
TOTAL	11,9872989	12

Uso de los índices estacionales en la predicción

$t = 5$  años (valor del denominador utilizado para el cálculo de la razón media de cada mes durante 5 años)

PRONÓSTICO

71    12/01/2004    25,2793403    22,8147634\*

\* Valor estimado de las ventas para el periodo  $x = 71$ , es decir, 12 de enero de 2004, por ejemplo:

$$\hat{Y} = \beta_0 + \beta_1 X + \varepsilon$$

$$25,2793 = 17,5531 + (0.1088 * 71) + \varepsilon$$

Cálculo del valor desestacionalizado estimado de las ventas en enero de 2004

$$22,8147 = \hat{Y} + ie_i$$

$$22,8147 = 25,2793 * 0,9025$$

## Anexo N°. 4.

## Modelo de regresión lineal

## A. CÁLCULO MANUAL

$$b1 = ((n * \text{suma\_xy}) - (\text{suma\_x} * \text{suma\_y})) / ((n * \text{suma\_x}) - (\text{suma\_y} * \text{suma\_y}))$$

$$b0 = (\text{suma\_y} / n) - (b1 * (\text{suma\_x} / n))$$

$$r = ((n * \text{suma\_xy}) - (\text{suma\_x} * \text{suma\_y})) / (((n * \text{suma\_x}^2) - ((\text{suma\_x})^2)) * ((n * \text{suma\_y}^2) - ((\text{suma\_y})^2)))^{(1/2)}$$

n	60,00	b1	0,123457
SUMA(xy)	40218,50	b0	16,998
SUMA(x)	1830,00	r	0,76131
SUMA(y)	1245,80	r^2	0,57959
SUMA(x**2)	73810,00	Syx	3,43
SUMA(y**2)	26340,18		

Estimación para x = 60,00  
y estimado será: 24,41

FECHA	x	y	x^2	x*y	y^2
31/1/99	1	15,8	1	16	249,64
28/2/99	2	15,0	4	30	225,00
31/3/99	3	17,9	9	54	320,41
30/4/99	4	17,4	16	70	302,76
31/5/99	5	18,5	25	93	342,25
30/6/99	6	18,9	36	113	357,21
31/7/99	7	17,9	49	125	320,41
31/8/99	8	18,3	64	146	334,89
30/9/99	9	18,1	81	163	327,61
31/10/99	10	18,8	100	188	353,44
30/11/99	11	19,2	121	211	368,64
31/12/99	12	22,9	144	275	524,41
31/1/00	13	17,0	169	221	289,00
28/2/00	14	16,0	196	224	256,00
30/3/00	15	19,0	225	285	361,00
29/4/00	16	19,3	256	309	372,49
30/5/00	17	20,2	289	343	408,04
29/6/00	18	20,3	324	365	412,09

## Anexo N°. 4.



30/7/00	19	19,1	361	363	364,81
30/8/00	20	19,9	400	398	396,01
29/9/00	21	18,9	441	397	357,21
30/10/00	22	20,6	484	453	424,36
29/11/00	23	20,9	529	481	436,81
30/12/00	24	24,1	576	578	580,81
30/1/01	25	18,3	625	458	334,89
27/2/01	26	17,1	676	445	292,41
30/3/01	27	19,7	729	532	388,09
29/4/01	28	20,5	784	574	420,25
30/5/01	29	21,2	841	615	449,44
29/6/01	30	20,7	900	621	428,49
30/7/01	31	20,5	961	636	420,25
30/8/01	32	21,0	1024	672	441,00
29/9/01	33	19,2	1089	634	368,64
30/10/01	34	21,5	1156	731	462,25
29/11/01	35	21,5	1225	753	462,25
30/12/01	36	25,1	1296	904	630,01
30/1/02	37	19,2	1369	710	368,64
28/2/02	38	18,8	1444	714	353,44
31/3/02	39	20,5	1521	800	420,25
30/4/02	40	21,2	1600	848	449,44
31/5/02	41	22,5	1681	923	506,25
30/6/02	42	22,2	1764	932	492,84
31/7/02	43	22,1	1849	950	488,41
31/8/02	44	21,8	1936	959	475,24
30/9/02	45	21,3	2025	959	453,69
31/10/02	46	22,6	2116	1040	510,76
30/11/02	47	21,7	2209	1020	470,89
31/12/02	48	27,7	2304	1330	767,29
31/1/03	49	20,6	2401	1009	424,36
28/2/03	50	19,6	2500	980	384,16
31/3/03	51	21,9	2601	1117	479,61
30/4/03	52	23,5	2704	1222	552,25
31/5/03	53	23,8	2809	1261	566,44
30/6/03	54	23,8	2916	1285	566,44
31/7/03	55	24,1	3025	1326	580,81
31/8/03	56	23,0	3136	1288	529,00
30/9/03	57	22,7	3249	1294	515,29
31/10/03	58	25,1	3364	1456	630,01
30/11/03	59	25,2	3481	1487	635,04
31/12/03	60	30,6	3600	1836	936,36
<b>TOTALES</b>	<b>1830</b>	<b>1245,8</b>	<b>73810</b>	<b>40219</b>	<b>26340,18</b>

## Anexo N° . 4.

## B. CÁLCULO EXCEL

## Resumen

Análisis de varianza					
	Gra libertad	Suma cuadrados	Pro. de cuadrados	F	Vr crítico d
Regresión	1,00	274,27	274,27	79,96	0,00
Residuos	58,00	198,95	3,43		
Total	59,00	473,22			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%
Intercepción	17,00	0,48	35,10	0,00	16,03	17,97
Variable $x_1$	0,12	0,01	8,94	0,00	0,10	0,15

Estadísticas regresión	
Coefficiente correlación	0,76
Coefficiente determinación	0,58
R <sup>2</sup> ajustado	0,57
Error típico	1,85
Observaciones	60,00

Análisis de los residuales		
Observaciones	Pronóstico $y^*$	Residuos
1,00	17,12	-1,32
2,00	17,24	-2,24
3,00	17,37	0,53
4,00	17,49	-0,09
5,00	17,62	0,88
6,00	17,74	1,16
7,00	17,86	0,04
8,00	17,99	0,31
9,00	18,11	-0,01
10,00	18,23	0,57
11,00	18,36	0,84
12,00	18,48	4,42
13,00	18,60	-1,60
14,00	18,73	-2,73
15,00	18,85	0,15
16,00	18,97	0,33
17,00	19,10	1,10
18,00	19,22	1,08
19,00	19,34	-0,24
20,00	19,47	0,43



## Anexo N° . 4.

→	21,00	19,59	-0,69
	22,00	19,71	0,89
	23,00	19,84	1,06
	24,00	19,96	4,14
	25,00	20,08	-1,78
	26,00	20,21	-3,11
	27,00	20,33	-0,63
	28,00	20,45	0,05
	29,00	20,58	0,62
	30,00	20,70	0,00
	31,00	20,83	-0,33
	32,00	20,95	0,05
	33,00	21,07	-1,87
	34,00	21,20	0,30
	35,00	21,32	0,18
	36,00	21,44	3,66
	37,00	21,57	-2,37
	38,00	21,69	-2,89
	39,00	21,81	-1,31
	40,00	21,94	-0,74
	41,00	22,06	0,44
	42,00	22,18	0,02
	43,00	22,31	-0,21
	44,00	22,43	-0,63
	45,00	22,55	-1,25
	46,00	22,68	-0,08
	47,00	22,80	-1,10
	48,00	22,92	4,78
	49,00	23,05	-2,45
	50,00	23,17	-3,57
	51,00	23,29	-1,39
	52,00	23,42	0,08
	53,00	23,54	0,26
	54,00	23,66	0,14
	55,00	23,79	0,31
	56,00	23,91	-0,91
	57,00	24,03	-1,33
	58,00	24,16	0,94
	59,00	24,28	0,92
	60,00	24,41	6,19

**Pronóstico y\* tiene implícita la tendencia**

Resultados de datos de probabilidad	
Percentil	y
0,83	15,00
2,50	15,80
4,17	16,00
5,83	17,00
7,50	17,10
9,17	17,40
10,83	17,90
12,50	17,90
14,17	18,10
15,83	18,30
17,50	18,30
19,17	18,50
20,83	18,80
22,50	18,80
24,17	18,90
25,83	18,90
27,50	19,00
29,17	19,10
30,83	19,20
32,50	19,20
34,17	19,20
35,83	19,30
37,50	19,60
39,17	19,70
40,83	19,90
42,50	20,20
44,17	20,30
45,83	20,50
47,50	20,50
49,17	20,50
50,83	20,60
52,50	20,60
54,17	20,70
55,83	20,90
57,50	21,00

→	59,17	21,20
	60,83	21,20
	62,50	21,30
	64,17	21,50
	65,83	21,50
	67,50	21,70
	69,17	21,80
	70,83	21,90
	72,50	22,10
	74,17	22,20
	75,83	22,50
	77,50	22,60
	79,17	22,70
	80,83	22,90
	82,50	23,00
	84,17	23,50
	85,83	23,80
	87,50	23,80
	89,17	24,10
	90,83	24,10
	92,50	25,10
	94,17	25,10
	95,83	25,20
	97,50	27,70
	99,17	30,60



## Anexo N°. 4.

## Modelo de regresión exponencial

## A. CÁLCULO MANUAL

$$\hat{Y} = \beta_0 + \beta_1^x + \varepsilon_i$$

$$\text{Log } \hat{y} = \text{Log } \beta_0 + x \text{ Log } \beta_1 + \varepsilon_i$$

n	60,00	LOGB1	0,0026	B1	1,0059
SUMA(xlogy)	2449,84	LOGB0	1,2351	B0	17,1847
SUMA(x)	1830,00	r	0,7749		
SUMA(logy)	78,81	r <sup>2</sup>	0,6005	Tasa	0,59%
SUMA(x**2)	73810,00	Syx	0,0014	VR PRESENTE	17,18
SUMA(logy**2)	103,71	Estimación	0,0369		
		Para X =	60,0000		60,0000
		LOGY			
		ESTIMADO	1,3892*	24,50**	24,50263

$$* 1,3892 = 1,2351 + (0,0026 * 60)$$

$$** 24,50 = 17,1847 * (1,0059)^{60}$$

Fecha	x	y	Logy	xlogy	logy <sup>2</sup>	x <sup>2</sup>
31/1/99	1	15,8	1,1987	1,199	1,437	1
28/2/99	2	15,0	1,1761	2,352	1,383	4
31/3/99	3	17,9	1,2529	3,759	1,570	9
30/4/99	4	17,4	1,2405	4,962	1,539	16
31/5/99	5	18,5	1,2672	6,336	1,606	25
30/6/99	6	18,9	1,2765	7,659	1,629	36
31/7/99	7	17,9	1,2529	8,770	1,570	49
31/8/99	8	18,3	1,2625	10,100	1,594	64
30/9/99	9	18,1	1,2577	11,319	1,582	81
31/10/99	10	18,8	1,2742	12,742	1,623	100
30/11/99	11	19,2	1,2833	14,116	1,647	121
31/12/99	12	22,9	1,3598	16,318	1,849	144
31/1/00	13	17,0	1,2304	15,996	1,514	169
28/2/00	14	16,0	1,2041	16,858	1,450	196
30/3/00	15	19,0	1,2788	19,181	1,635	225
29/4/00	16	19,3	1,2856	20,569	1,653	256
30/5/00	17	20,2	1,3054	22,191	1,704	289
29/6/00	18	20,3	1,3075	23,535	1,710	324
30/7/00	19	19,1	1,2810	24,340	1,641	361
30/8/00	20	19,9	1,2989	25,977	1,687	400

## Anexo N° . 4.

→	29/9/00	21	18,9	1,2765	26,806	1,629	441
	30/10/00	22	20,6	1,3139	28,905	1,726	484
	29/11/00	23	20,9	1,3201	30,363	1,743	529
	30/12/00	24	24,1	1,3820	33,168	1,910	576
	30/1/01	25	18,3	1,2625	31,561	1,594	625
	27/2/01	26	17,1	1,2330	32,058	1,520	676
	30/3/01	27	19,7	1,2945	34,951	1,676	729
	29/4/01	28	20,5	1,3118	36,729	1,721	784
	30/5/01	29	21,2	1,3263	38,464	1,759	841
	29/6/01	30	20,7	1,3160	39,479	1,732	900
	30/7/01	31	20,5	1,3118	40,664	1,721	961
	30/8/01	32	21,0	1,3222	42,311	1,748	1024
	29/9/01	33	19,2	1,2833	42,349	1,647	1089
	30/10/01	34	21,5	1,3324	45,303	1,775	1156
	29/11/01	35	21,5	1,3324	46,635	1,775	1225
	30/12/01	36	25,1	1,3997	50,388	1,959	1296
	30/1/02	37	19,2	1,2833	47,482	1,647	1369
	28/2/02	38	18,8	1,2742	48,418	1,623	1444
	31/3/02	39	20,5	1,3118	51,158	1,721	1521
	30/4/02	40	21,2	1,3263	53,053	1,759	1600
	31/5/02	41	22,5	1,3522	55,439	1,828	1681
	30/6/02	42	22,2	1,3464	56,547	1,813	1764
	31/7/02	43	22,1	1,3444	57,809	1,807	1849
	31/8/02	44	21,8	1,3385	58,892	1,791	1936
	30/9/02	45	21,3	1,3284	59,777	1,765	2025
	31/10/02	46	22,6	1,3541	62,289	1,834	2116
	30/11/02	47	21,7	1,3365	62,814	1,786	2209
	31/12/02	48	27,7	1,4425	69,239	2,081	2304
	31/1/03	49	20,6	1,3139	64,379	1,726	2401
	28/2/03	50	19,6	1,2923	64,613	1,670	2500
	31/3/03	51	21,9	1,3404	68,363	1,797	2601
	30/4/03	52	23,5	1,3711	71,296	1,880	2704
	31/5/03	53	23,8	1,3766	72,959	1,895	2809
	30/6/03	54	23,8	1,3766	74,335	1,895	2916
	31/7/03	55	24,1	1,3820	76,011	1,910	3025
	31/8/03	56	23,0	1,3617	76,257	1,854	3136
	30/9/03	57	22,7	1,3560	77,293	1,839	3249
	31/10/03	58	25,1	1,3997	81,181	1,959	3364
	30/11/03	59	25,2	1,4014	82,683	1,964	3481
	31/12/03	60	30,6	1,4857	89,143	2,207	3600
	<b>TOTALES</b>	<b>1830</b>	<b>1245,8</b>	<b>78,8077</b>	<b>2449,843</b>	<b>103,708</b>	<b>73810</b>

## B. CÁLCULO EXCEL

### Resumen

Análisis de varianza					
	Gra libertad	Suma cuadrados	Pro. de cuadrados	F	Vr crítico d
Regresión	1,00	0,12	0,12	87,19	0,00
Residuos	58,00	0,8	0,00		
Total	59,00	0,20			

	Coefficientes	Error típico	Estadístico t	Probabilidad	Inferior 95%	Superior 95%
Intercepción	1,24	0,01	128,06	0,00	1,22	1,25
Variable $x_1$	0,00	0,00	9,34	0,00	0,00	0,00

Estadísticas regresión	
Coefficiente correlación	0,77
Coefficiente determinación	0,60
R <sup>2</sup> ajustado	0,59
Error típico	0,04
Observaciones	60,00

Análisis de los residuales		
Observaciones	Pronóstico $y^*$	Residuos
1,00	1,24	-0,04
2,00	1,24	-0,06
3,00	1,24	0,01
4,00	1,25	0,00
5,00	1,25	0,02
6,00	1,25	0,03
7,00	1,25	0,00
8,00	1,26	0,01
9,00	1,26	0,00
10,00	1,26	0,01
11,00	1,26	0,02
12,00	1,27	0,09
13,00	1,27	-0,04
14,00	1,27	-0,07
15,00	1,27	0,01
16,00	1,28	0,01
17,00	1,28	0,03
18,00	1,28	0,03
19,00	1,28	0,00
20,00	1,29	0,01



## Anexo N°. 4.

→	21,00	1,29	-0,01
	22,00	1,29	0,02
	23,00	1,29	0,03
	24,00	1,30	0,09
	25,00	1,30	-0,04
	26,00	1,30	-0,07
	27,00	1,30	-0,01
	28,00	1,31	0,00
	29,00	1,31	0,02
	30,00	1,31	0,00
	31,00	1,31	0,00
	32,00	1,32	0,00
	33,00	1,32	-0,04
	34,00	1,32	0,01
	35,00	1,33	0,01
	36,00	1,33	0,07
	37,00	1,33	-0,05
	38,00	1,33	-0,06
	39,00	1,34	-0,02
	40,00	1,34	-0,01
	41,00	1,34	0,01
	42,00	1,34	0,00
	43,00	1,35	0,00
	44,00	1,35	-0,01
	45,00	1,35	-0,02
	46,00	1,35	0,00
	47,00	1,36	-0,02
	48,00	1,36	0,08
	49,00	1,36	-0,05
	50,00	1,36	-0,07
	51,00	1,37	-0,03
	52,00	1,37	0,00
	53,00	1,37	0,01
	54,00	1,37	0,00
	55,00	1,38	0,01
	56,00	1,38	-0,02
	57,00	1,38	-0,03
	58,00	1,38	0,02
	59,00	1,39	0,01
	60,00	1,39	0,10

**Pronóstico y\* dados en logaritmos**

## Anexo N°. 5.

## Modelo de regresión parabólica

## CÁLCULO MANUAL

## Ecuaciones normales para la regresión parabólica

n	60,00
SUMA(xy)	38972,70
SUMA(x)	1770,00
SUMA(y)	1245,80
SUMA(x <sup>2</sup> )	70210,00
SUMA(x <sup>3</sup> )	3132900,00
SUMA(x <sup>4</sup> )	149111998,00
SUMA(x <sup>2</sup> *y)	1592559,70

		B <sub>2</sub>	B <sub>1</sub>	B <sub>0</sub>
1	1246 =	70210	1770	60
2	38973 =	3132900	70210	1770
3	1592560 =	149111998	3132900	70210

## Valores estimados de los parámetros de la regresión

$$B_2 = 0,00086$$

$$B_1 = 0,07297$$

$$B_0 = 17,6094$$

$$X = 2004,00 \quad 71,00$$

$$\text{Estimación} = 27,10$$

$$y = B_2 X^2 + B_1 X + B_0$$

$$Y_{\text{EST}} = 0,00086 (71)^2 + 0,07297(71) + 17,61$$

$$\text{Ventas estimadas dic-04} = 27,10$$

Fecha	x	y	x <sup>2</sup>	x*y	y <sup>2</sup>	x <sup>3</sup>	x <sup>4</sup>	(x <sup>2</sup> )*y
31/1/99	0	15,8	0	0,0	249,6	0	0	0,0
28/2/99	1	15,0	1	15,0	225,0	1	1	15,0
31/3/99	2	17,9	4	35,8	320,4	8	16	71,6
30/4/99	3	17,4	9	52,2	302,8	27	81	156,6
31/5/99	4	18,5	16	74,0	342,3	64	256	296,0
30/6/99	5	18,9	25	94,5	357,2	125	625	472,5
31/7/99	6	17,9	36	107,4	320,4	216	1296	644,4
31/8/99	7	18,3	49	128,1	334,9	343	2401	896,7
30/9/99	8	18,1	64	144,8	327,6	512	4096	1158,4
31/10/99	9	18,8	81	169,2	353,4	729	6561	1522,8
30/11/99	10	19,2	100	192,0	368,6	1000	10000	1920,0
31/12/99	11	22,9	121	251,9	524,4	1331	14641	2770,9
31/1/00	12	17,0	144	204,0	289,0	1728	20736	2448,0
28/2/00	13	16,0	169	208,0	256,0	2197	28561	2704,0
30/3/00	14	19,0	196	266,0	361,0	2744	38416	3724,0
29/4/00	15	19,3	225	289,5	372,5	3375	50625	4342,5
30/5/00	16	20,2	256	323,2	408,0	4096	65536	5171,2
29/6/00	17	20,3	289	345,1	412,1	4913	83521	5866,7
30/7/00	18	19,1	324	343,8	364,8	5832	104976	6188,4

## Anexo N°. 5.

→ 30/8/00	19	19,9	361	378,1	396,0	6859	130321	7183,9
29/9/00	20	18,9	400	378,0	357,2	8000	160000	7560,0
30/10/00	21	20,6	441	432,6	424,4	9261	194481	9084,6
29/11/00	22	20,9	484	459,8	436,8	10648	234256	10115,6
30/12/00	23	24,1	529	554,3	580,8	12167	279841	12748,9
30/1/01	24	18,3	576	439,2	334,9	13824	331776	10540,8
27/2/01	25	17,1	625	427,5	292,4	15625	390625	10687,5
30/3/01	26	19,7	676	512,2	388,1	17576	456976	13317,2
29/4/01	27	20,5	729	553,5	420,3	19683	531441	14944,5
30/5/01	28	21,2	784	593,6	449,4	21952	614656	16620,8
29/6/01	29	20,7	841	600,3	428,5	24389	707281	17408,7
30/7/01	30	20,5	900	615,0	420,3	27000	810000	18450,0
30/8/01	31	21,0	961	651,0	441,0	29791	923521	20181,0
29/9/01	32	19,2	1024	614,4	368,6	32768	1048576	19660,8
30/10/01	33	21,5	1089	709,5	462,3	35937	1185921	23413,5
29/11/01	34	21,5	1156	731,0	462,3	39304	1336336	24854,0
30/12/01	35	25,1	1225	878,5	630,0	42875	1500625	30747,5
30/1/02	36	19,2	1296	691,2	368,6	46656	1679616	24883,2
28/2/02	37	18,8	1369	695,6	353,4	50653	1874161	25737,2
31/3/02	38	20,5	1444	779,0	420,3	54872	2085136	29602,0
30/4/02	39	21,2	1521	826,8	449,4	59319	2313441	32245,2
31/5/02	40	22,5	1600	900,0	506,3	64000	2560000	36000,0
30/6/02	41	22,2	1681	910,2	492,8	68921	2825761	37318,2
31/7/02	42	22,1	1764	928,2	488,4	74088	3111696	38984,4
31/8/02	43	21,8	1849	937,4	475,2	79507	3418801	40308,2
30/9/02	44	21,3	1936	937,2	453,7	85184	3748096	41236,8
31/10/02	45	22,6	2025	1017,0	510,8	91125	4100625	45765,0
30/11/02	46	21,7	2116	998,2	470,9	97336	4477456	45917,2
31/12/02	47	27,7	2209	1301,9	767,3	103823	4879681	61189,3
31/1/03	48	20,6	2304	988,8	424,4	110592	5308416	47462,4
28/2/03	49	19,6	2401	960,4	384,2	117649	5764801	47059,6
31/3/03	50	21,9	2500	1095,0	479,6	125000	6250000	54750,0
30/4/03	51	23,5	2601	1198,5	552,3	132651	6765201	61123,5
31/5/03	52	23,8	2704	1237,6	566,4	140608	7311616	64355,2
30/6/03	53	23,8	2809	1261,4	566,4	148877	7890481	66854,2
31/7/03	54	24,1	2916	1301,4	580,8	157464	8503056	70275,6
31/8/03	55	23,0	3025	1265,0	529,0	166375	9150625	69575,0
30/9/03	56	22,7	3136	1271,2	515,3	175616	9834496	71187,2
31/10/03	57	25,1	3249	1430,7	630,0	185193	10556001	81549,9
30/11/03	58	25,2	3364	1461,6	635,0	195112	11316496	84772,8
31/12/03	59	30,6	3481	1805,4	936,4	205379	12117361	106518,6
<b>TOTALES</b>	<b>1770</b>	<b>1245,8</b>	<b>70210</b>	<b>38972,7</b>	<b>26340,2</b>	<b>3132900</b>	<b>149111998</b>	<b>1592559,7</b>

## Anexo N° . 5.

**Modelo general de regresión parabólico**

$$y_{est} = B_2X^2 + B_1X + B_0 + \varepsilon$$

$$\text{suma}(y) = B_2 * \text{suma}(x^2) + B_1 * \text{suma}(x) + n * B_0$$

$$\text{suma}(xy) = B_2 * \text{suma}(x^3) + B_1 * \text{suma}(x^2) + B_0 * \text{suma}(x)$$

$$\text{suma}(x^2 * y) = B_2 * \text{suma}(x^4) + B_1 * \text{suma}(x^3) + B_0 * \text{suma}(x^2)$$

Reemplazando por sus equivalentes tenemos las ecuaciones normales para la regresión parabólica:

1	1246	=	70210	B2	1770	B1	60,00	B0
2	38973	=	3132900	B2	70210	B1	1770,00	B0
3	1592560	=	149111998	B2	3132900	B1	70210,00	B0

Multiplicamos por -29,5 la primera ecuación y la sumamos a la segunda  
-29,5

1	-36751	=	-2071195	B2	-52215	B1	-1770	B0
2	38973	=	3132900	B2	70210	B1	1770	B0

SUMA 2221,60 = 1061705,00 B2 17995,00 B1 0,00 B0 \_4\_

Multiplicamos por -1170,16 la primera ecuación y la sumamos a la tercera  
-1170,16

1	-1457793,63	=	-82157401,67	B2	-2071195,00	B1	-70210,00	B0
3	1592559,70	=	149111998,00	B2	3132900,00	B1	70210,00	B0

SUMA 134766,07 = 66954596,33 B2 1061705,00 B1 0,0 B0 \_5\_

\_4\_ 2221,60 = 1061705,00 B2 17995,00 B1 0,00 B0

\_5\_ 134766,07 = 66954596,33 B2 1061705,00 B1 0,00 B0

Multiplicamos por -59,00 la cuarta ecuación y la sumamos a la quinta  
-59,00

4	-131074,40	=	-62640595,00	B2	-1061705,00	B1		
5	134766,07	=	66954596,33	B2	1061705,00	B1		

SUMA 3691,67 = 4314001,33 B2 0,00 B1

Despejamos B2

Luego B2 0,000856

Tomamos la ecuación 4 y le reemplazamos el valor de B2

Tomamos 4 2221,60 = 1061705,00 B2 17995,00 B1

Reemp B2 2221,60 = 908,54 17995,00 B1

Despejamos B1

luego B1 = 0,07

Tomamos la ecuación 4 y le reemplazamos el valor de B2 y B1

Tomamos 1 1245,80 = 70210,00 B2 1770,00 B1 60,00 B0

r por b1 b2 1245,80 = 60,08 129,15 60,00 B0

Despejamos B0

Luego B0 = 17,61

Obtenidos B2, B1 y B0 se reemplazan en el modelo del ajuste parabólico y se estima para x deseado en este caso:

x 2004,0 x= 71,00 Diciembre 2004

Estimación 27,10393

Y = B2X<sup>2</sup>+B1X+B0

Y estimado 0,00086 X<sup>2</sup>+ 0,07 X+ 17,61

Ventas estimadas dic 2004 = 27,10

# Descubrimiento de conocimiento en los negocios

JOHANY ARMANDO CARREÑO GAMBOA

## LOS EMPRESARIOS FRENTE A LA INFORMACIÓN

El procesamiento de los datos ha sido siempre un tema importante, agradable e interesante para muchos empresarios, aunque lo consideren un gran reto. Generar información novedosa para la toma de decisiones implica arte e ingeniería. La comprensión de los datos y su traducción en modelos eficaces es un proceso artístico. Transformar los datos en información es un proceso de ingeniería<sup>1</sup>.

Los datos aumentan exponencialmente cada año, pero la información tiende a decrecer. Por lo tanto, Las organizaciones o empresas deben estar y tener la capacidad de crear y llevar a cabo planes de gestión de los datos (*data management*) y descubrimiento de conocimiento (KDD), con el objeto de realizar consultas de los mismos, generar informes/reportes y más específicamente, llevar a cabo todo el procesamiento para

### RESUMEN

Ante la internacionalización de la economía, las organizaciones requieren basarse en la información y el conocimiento, apoyadas en tecnologías de la información y comunicación (TIC), pensar globalmente en políticas integrales y basarse en economías en red bajo esquemas asociativos que las fortalezcan. El crecimiento de empresas en los últimos años, hace prioritario tratar de obtener conocimiento útil desde los propios datos y dar un paso más allá en el apoyo a la toma de decisiones más acertada. A tal fin, se ofrece en el documento información básica acerca de la minería de datos, se reconocen sus diferentes etapas y se determina su relación con otras disciplinas. Además se da a conocer el funcionamiento del tipo de algoritmo “árboles de decisión” y, se utiliza la herramienta “Weka” para ajustar modelos a conjuntos de datos.

### PALABRAS CLAVE

Dato, información, descubrimiento de conocimiento, minería de datos, toma de decisiones, análisis de decisiones.

### ABSTRACT

In view of the internationalization of economy, companies need to rely on information and knowledge. Also, they have to get support from information and communication technologies (TIC), think globally about comprehensive policies, and be based on network economies under associative schemes that strengthen them. The growth of companies in the last years prioritizes the achievement of getting useful knowledge from own data and go beyond when supporting the most appropriate decision making. In order to do this, we offer in the document basic information about data mining, its different stages, and its relationship with other disciplines. Besides, we present the functioning of an algorithm “decision trees” and the use of the tool “Weka” in order to adjust models to data groups.

### KEY WORDS

Datum, information, knowledge discovery, data mining, decision making, decision analysis.

---

1 Turban, E. (2008). *Business intelligence: a managerial approach*. Upper Saddle River, N. J.: Pearson Prentice Hall.

traducir la lógica de los negocios a la lógica de sistemas empresariales y procesar información útil, válida y relevante para tomar decisiones<sup>2</sup>.

En el ámbito empresarial todos los días se toman decisiones, pero ¿cuál es el grado de acierto de éstas? Los administrativos conocen como estratégica, táctica y operativamente se comporta cada una de sus organizaciones. Además, saben quiénes son sus empleados; se sienten confiados porque tienen buenos sistemas de información; están seguros porque poseen un buen parque tecnológico; pero la realidad es otra, en el mundo de los negocios actuales hay que estar preparado para descubrir conocimiento.

Las empresas se cierran y muchos culpan a las políticas gubernamentales, sin tener en cuenta que sus decisiones han sido consideradas bajo criterios pobres de extracción, depuración y transformación de los datos. No basta con invertir en nuevas plantas, lanzar nuevos productos, seleccionar e implantar tecnologías de punta, entre otras, sin haber realizado un estudio de las características comunes (reversibilidad, replicación, riesgo, impacto, futuro, etc.) de las decisiones que se toman día a día<sup>3</sup>.

Con todo y lo anterior y a pesar de su importancia, las decisiones no siempre se toman o evalúan utilizando métodos, herramientas y procedimientos apropiados. Aún más preocupante es que, en general, las empresas invierten cantidades considerables de dinero en publicidad y estudios de mercadeo, por ejemplo, y muy poco en métodos y herramientas de soporte a la toma de decisiones<sup>4</sup>.

Para finalizar este aparte, cabe resaltar que según los requerimientos de información y su funcionalidad, existen procesos que permiten el descubrimiento de conocimiento, que pueden ser aplicados en cada uno de los niveles de la organización. Las soluciones de gestión de los datos proporcionan un fácil acceso a los datos críticos dentro de la empresa, necesarios para el análisis,

así como un medio para integrar los datos corporativos con los procesos de toma de decisiones estratégicas y tácticas; también permite a la empresa afinar la toma de decisiones cotidiana, asegurando que cada grupo operativo tenga acceso a la información necesaria para contestar preguntas específicas y distribuir dicha información a todos los niveles de la organización.

## EL PROCESO DE DESCUBRIMIENTO DE CONOCIMIENTO (KDD)

KDD se ha definido como el “proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia, comprensibles a partir de los datos”<sup>5</sup>.

Se trata de interpretar grandes cantidades de datos y encontrar relaciones o patrones. Para conseguirlo las

### Johany Armando Carreño Gamboa.

Ingeniero de Sistemas de la Universidad Autónoma de Bucaramanga; consultor y asesor en implantación de proyectos de Gestión Tecnológica en empresas nacionales; analista y diseñador de *software* para gestión de datos y toma de decisiones. Docente investigador de tiempo completo adscrito a la Facultad de Ingeniería y Ciencias Básicas del Politécnico Grancolombiano Institución Universitaria. Coordinador de los semilleros de investigación en “Descubrimiento de Conocimiento y Minería de Datos” (<http://si.kdmind.googlepages.com/home>) y “Diseño y Desarrollo de Aplicaciones para Dispositivos Móviles” (<http://sid.mobi.googlepages.com/home>). Actualmente, hace parte del grupo de Expertos en Redes NGN Colombia, en calidad de representante del Politécnico Grancolombiano. Para contactar al autor: [jcarreno@poligran.edu.co](mailto:jcarreno@poligran.edu.co).

2 Williams, Steve y Williams, Nancy (2006). *The profit impact of business intelligence*. Amsterdam, Boston: Morgan Kaufmann.

3 Castillo Hernández, Mario (2006). *Toma de decisiones en las empresas: entre el arte y la técnica: metodologías, modelos y herramientas*. Bogotá: Universidad de los Andes.

Vitt, E.; Luckevich, M.; y Misner, S. (2003). *Business intelligence: técnicas de análisis para la toma de decisiones. IT/Tecnología y Empresa*. Madrid: McGraw Hill Interamericana.

T. Moss, L., y Atre, S. (2003). *Business intelligence roadmap: The complete project lifecycle for decision-support applications*. Boston, MA: Addison-Wesley Professional.

4 Castillo Hernández, Mario (2006). Op. cit. 15.

5 Fayyad, U.; Piatetsky-Shapiro, G. y Smyth, P. (1996). *Advances in knowledge discovery and data mining*. Massachusetts: MIT Press. Pág. 1-34.

investigaciones, en estos temas, incluyen técnicas de aprendizaje (*machine learning*), bases de datos, análisis estadístico de datos, técnicas de representación de conocimiento, razonamiento basado en casos (*case based reasoning* –CBR-), filtrado colaborativo, razonamiento aproximado, adquisición de conocimiento, redes neuronales, visualización de datos, algoritmos genéticos, recuperación de información y computación de altas prestaciones, entre otros. Tareas comunes en KDD son los problemas de clasificación y agrupamiento (*clustering*), el reconocimiento de patrones, estimación/regresión, el modelado predictivo, la detección de dependencias, análisis de *links*, análisis de canastas de mercado e inducción de reglas, entre otras<sup>6</sup>.

Cuando se tratan temas de descubrimiento de conocimiento en el ámbito de los negocios, es importante entender que hay que focalizarse en aprender de lo que ha sucedido a través del tiempo y que aprender no puede hacerse de la nada. Por lo tanto, la aplicación de técnicas de KDD requiere la realización de una serie de actividades previas encaminadas a preparar los datos de entrada, debido a que en muchas ocasiones los datos provienen de fuentes heterogéneas, no tienen el formato adecuado o contienen ruido. Por otra parte, es necesario interpretar y evaluar los resultados obtenidos.

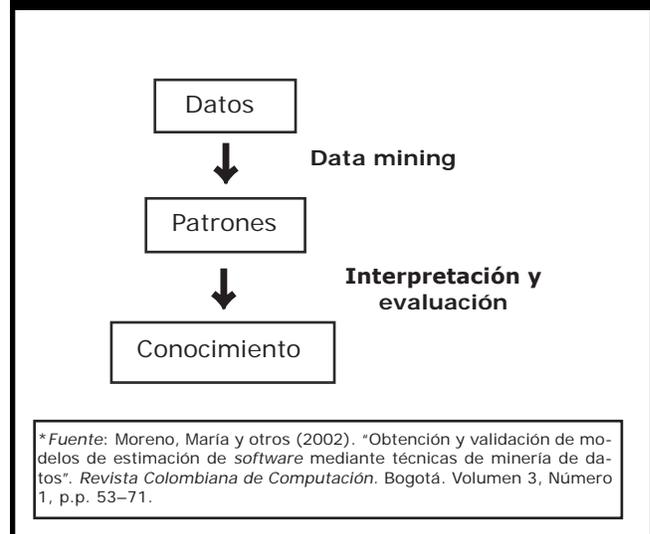
El proceso completo de KDD consta de las siguientes etapas<sup>7</sup> y se puede observar en la Figura 1.

1. Determinar las fuentes de información que pueden ser útiles y dónde conseguirlas. Para llevar a cabo esta tarea es necesario tener dominio del negocio y del problema. También es importante contar con sistemas de procesamiento de transacciones que permitan capturar la interacción de la empresa con sus clientes.

2. Diseñar el esquema de un almacén de datos (*data warehouse*) que almacene de manera operativa toda la información histórica exacta del comportamiento de los clientes, de modo que la aplicación de la minería de datos pueda hacer uso de ésta.

Cabe aclarar, que no es necesario siempre implementar un *data warehouse*, simplemente lo que se requiere es

**Figura N° . 1. Proceso de extracción de conocimiento\***



preparar, estructurar y unificar los datos en el formato especializado, requerido e idóneo para la herramienta de minería de datos.

3. Implantación del conjunto de datos que permita la "exploración" y visualización previa de sus datos, para discernir qué aspectos pueden interesar para ser estudiados.

4. Fase de selección, limpieza y transformación de los datos que se van a analizar. En esta fase se eliminan o corrigen los datos incorrectos y se decide la estrategia a seguir con los datos incompletos. Además, con el objetivo de hacer más fácil la tarea propia de minería y para que los resultados de la misma sean más útiles, se proyectan los datos para considerar únicamente aquellas variables o atributos que van a ser relevantes. La selección incluye tanto una criba o fusión horizontal (filas/registros) como vertical, de columnas o atributos.

5. Seleccionar el método de minería de datos apropiado (clasificar, agrupar, etc.) y aplicarlo. La minería de datos tiene como objetivo primordial convertir la historia en planes de acción útiles para el futuro.

6 Berry, Michael J. A. y Linoff, Gordon (2004). *Data mining techniques: for marketing, saes, and customer relationship management*. Segunda edición. Indianapolis: Wiley.

7 Lavrac, N. "Selected techniques for data mining in medicine". En: *Artificial Intelligence in Medicine*. Tecklenburg, West Germany: Burgberlag. Vol. 16 (1), Pág. 3-23, 1999.

6. Interpretación, transformación y representación de los patrones extraídos. En esta etapa se evalúan los patrones y se analizan por los expertos, y si es necesario, se vuelve a las fases anteriores para una nueva iteración. El objetivo primordial es diseñar óptimos planes estratégicos de interacción y administración de relaciones con el cliente.

7. Fase de difusión y uso del nuevo conocimiento. En esta fase los empresarios deben hacer uso del nuevo conocimiento y hacer partícipes de él a todos los posibles usuarios.

Aunque las etapas anteriores se realizan en el orden en que aparecen, el proceso es altamente iterativo, y se establece retroalimentación entre los mismos. Además, no todos los pasos requieren el mismo esfuerzo; generalmente la etapa de preprocesamiento o preparación de los datos es la más costosa ya que representa aproximadamente el 60% del esfuerzo total, mientras que la etapa de minería solo representa el 20%.

Además de las fases descritas, frecuentemente se incluye una fase previa de análisis de las necesidades de la organización y definición del problema<sup>8</sup>, en la que se establecen los objetivos de minería de datos. Por ejemplo, un objetivo de negocio de una empresa comercial sería encontrar patrones en los datos que le ayuden a conservar los buenos clientes; para ello, se podría tener varios objetivos de minería de datos: construir un modelo para predecir clientes rentables, y un segundo modelo para identificar los clientes que probablemente dejarán de hacerlo.

## Dominios

Los dominios hacen referencia a las actividades del día a día en las organizaciones. Cuando se decide iniciar un proyecto orientado al descubrimiento de conocimiento hay que tener en cuenta que éste es un campo multidisciplinar que se ha desarrollado en paralelo con otras tecnologías. Por ello es importante nutrir los proyectos con bases de datos, técnicas de visualización, sistemas para la toma de decisiones, herramientas de recuperación de información, personal especializado o experto y todas las disciplinas que se requieran para profundizar en el área relacionada con el dominio de estudio.

Los negocios de la distribución y la publicidad dirigida han sido tradicionalmente las áreas en las que más se han empleado los métodos de descubrimiento de conocimiento. Pero éstas no son las únicas áreas a las que se pueden aplicar. De hecho, se pueden encontrar un gran número de dominios: financieros, científicos (medicina, farmacia, astronomía, psicología, etc.), políticas económicas, sanitarias o demográficas, educación, policiales, procesos industriales, turismo, tráfico, deportes, recursos humanos, *web*, entre otros.

A continuación se incluye una lista de ejemplos de algunas de las áreas antes referidas para exponer en qué dominios se pueden usar técnicas de KDD:

### Medicina:

- Identificación de patologías. Diagnóstico de enfermedades.
- Detección de pacientes con riesgo de sufrir una patología concreta.
- Gestión hospitalaria y asistencial. Predicciones temporales de los centros asistenciales para el mejor uso de los recursos, consultas, salas y habitaciones.
- Recomendación priorizada de fármacos para una misma patología.

### Análisis de mercado, distribución y, en general, comercio:

- Análisis de la cesta de compra (compras conjuntas, secuenciales, ventas cruzadas, señuelos, etc.).
- Evaluación de campañas publicitarias.
- Análisis de la fidelidad de los clientes. Reducción de fuga.
- Segmentación de los clientes.
- Estimación de existencias (*stocks*), de costes, de ventas, etc.

### Aplicaciones financieras y de banca:

- Obtención de patrones de uso fraudulento de tarjetas de crédito.
- Determinación del gasto en tarjeta de crédito por grupos.
- Cálculo de correlaciones entre indicadores financieros.
- Identificación de reglas de mercado de valores a partir de históricos.
- Análisis de riesgos en créditos<sup>9</sup>.

<sup>8</sup> Hernández Orallo, José y otros (2004). *Introducción a la minería de datos*. Madrid: Pearson Educación S.A.

<sup>9</sup> *Ibid.*

Todos estos ejemplos muestran la gran variedad de aplicaciones donde el uso de técnicas de descubrimiento de conocimiento puede ayudar a entender mejor el entorno en el que se desenvuelve la organización y, en definitiva, la repercusión en calidad de servicio y de disminución de costes que pueden ser altamente significativos. Sin embargo, el uso de estas técnicas todavía es reducido, especialmente en Colombia<sup>10</sup>.

## MINERÍA DE DATOS COMO HERRAMIENTA DE APOYO EN LA TOMA DE DECISIONES

El término “minería de datos” está relacionado con la extracción de información relevante (no trivial, implícita, desconocida y con potencial utilidad) en grandes volúmenes de datos. El objetivo de la minería de datos es proporcionar el potencial de análisis necesario para explotar grandes volúmenes de información transaccional con el fin de obtener conocimiento de apoyo en la toma de decisiones (normas, regularidades, patrones, restricciones). Witten, Frank, C. y Boswell (2000) definen la minería de datos como el proceso de extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos<sup>11</sup>.

La información transaccional de las organizaciones ha sido recogida a lo largo del tiempo por sistemas automáticos o de forma manual. Información que cuanto más precisa sea, más completo será el registro que proporcione de las interacciones de los distintos subsistemas de la empresa entre sí y de la empresa con el exterior: proveedores, clientes, aseguradoras, entidades promotoras de salud, centros de investigación, hoteles, jugadores, etc.

La minería de datos es una forma de aprendizaje inductivo, que permite seleccionar las regularidades y reglas más plausibles soportadas por los datos. Los sistemas cognitivos intentan entender su ambiente usando una simplificación del mismo, llamado modelo, que consiste de clases que representan objetos similares en el ambiente y reglas que describen los cambios en el mismo<sup>12</sup>.

El proceso de minería de datos implica ajustar modelos o determinar patrones a partir de datos. Este ajuste normalmente es de tipo estadístico, en el sentido en que se permite un cierto ruido o error dentro del modelo.

Los algoritmos de minería de datos realizan, en general, tareas de descripción (de datos y patrones), de predicción (de datos desconocidos) y de segmentación (de datos). Otras, como análisis de dependencias e identificación de anomalías, se pueden utilizar tanto para descripción como para predicción.

Los algoritmos de minería de datos se clasifican en dos grandes categorías: supervisados o predictivos y no supervisados o de descubrimiento de conocimiento<sup>13</sup>. Los modelos predictivos pretenden estimar valores futuros o desconocidos de variables de interés, que denominamos variables objetivo o dependientes, usando otras variables o campos de la base de datos, a las que nos referiremos como variables independientes o predictivas. Por ejemplo, un modelo predictivo sería aquel que permite estimar la demanda de un nuevo producto en función del gasto de publicidad<sup>14</sup>. Esta forma de trabajar se conoce como aprendizaje supervisado y se desarrolla en dos fases: entrenamiento (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y prueba (prueba del modelo sobre el resto de los datos).

Cuando una aplicación no es lo suficientemente madura, no tiene el potencial necesario para una solución predictiva, en ese caso hay que recurrir a los métodos no supervisados que descubren patrones y tendencias en los datos actuales (no utilizan datos históricos). Estos modelos descriptivos, en cambio, identifican patrones que explican o resumen los datos, es decir, sirven para explorar las propiedades de los datos examinados, no para predecir nuevos datos. Por ejemplo, una agencia de viajes desea identificar grupos de personas con unos mismos gustos, con el objeto de organizar diferentes ofertas para cada grupo y poder así remitirles esta información; para ello analiza los viajes que han realizado sus clientes e infiere un modelo descriptivo que caracterize a estos grupos<sup>15</sup>.

10 Ibid.

11 Witten, I. H. y Frank, E. (1999). *Data mining: practical machine learning tools and techniques with Java implementations*. San Francisco: Morgan Kaufmann.

12 Hernández Orallo, J.; Ramírez Quintana, M. J. y Ferri Ramírez, C. (2004). *Introducción a la minería de datos*. Madrid: Pearson Educación. Pág. 3-18.

13. Witten, H., y Frank, E. Op. cit.

14. Hernández Orallo, José y otros. Op. cit.

15. Ídem.

Agregando a lo anterior, algunos problemas de interés intelectual, económico y de negocios, pueden ser expresados en términos de las siguientes tareas: clasificación, estimación, predicción, reglas de asociación, *clustering*, descripción y optimización de parámetros (*profiling*).

En la Tabla 1 se muestran algunas de las técnicas de minería en ambas categorías<sup>16</sup>, que pueden ser utilizadas como estrategias de análisis de datos.

Hay que tener en cuenta una gran variedad de técnicas, combinaciones y nuevas variantes aparecidas recientemente, debido al interés del campo. Los sistemas de KDD incorporan la mayor cantidad de técnicas, así como la heurística para determinar o asesorar al usuario sobre qué método es mejor para distintos problemas. En varios autores (ver bibliografía) se puede encontrar información relevante y de interés sobre cada una de las técnicas de minería de datos.

**Tabla N°. 1. Clasificación de las técnicas de minería de datos**

Supervisados	No supervisados
Árboles de decisión	Detección de desviaciones
Inducción neuronal	Segmentación
Regresión	<i>Clustering</i>
Series temporales	Reglas de asociación
	Patrones secuenciales

Después de todo, a nadie se le escapa que la calidad de un trabajo de minería de datos está por demás condicionada al estado de la información sobre la que dicho trabajo se aplica: datos incompletos o distorsionados sólo producen resultados deficientes, por más sofisticadas que sean las herramientas con que se procesan. La naturaleza inicial de los datos del negocio no es, en

este sentido, la más atractiva para un proyecto impecable de minería de datos, pues pueden presentar (y a menudo lo hacen) problemas en su captura, codificación, etc. No obstante, el uso de una metodología como CRISP-DM (CROSS Industry Standard Process for Data Mining) a la que nos referimos en el siguiente aparte hace que, en un momento determinado, la información alcance una “madurez” en términos de procesamiento informático a nuestro parecer suficiente para, ofrecer resultados de calidad.

## UNA METODOLOGÍA PARA EL DESCUBRIMIENTO DE CONOCIMIENTO EN GRANDES CONJUNTOS DE DATOS

Existen varias metodologías de trabajo para la elaboración de un proyecto de minería de datos, que especifican diferentes etapas o fases, permitiendo tener una secuencia clara, ordenada y controlada de las mismas y facilitando la planeación y ejecución del proyecto. Un grupo de empresas europeas pioneras en enfrentar problemas de minería de datos son Teradata, SPSS, Daimler-Chrysler y OHRA y propusieron en 1999 la guía de referencia denominada CRISP-DM (Cross-Industry Standard Process for Data Mining)<sup>17</sup>.

A grandes rasgos, se describe CRISP-DM una de las más relevantes y que puede ser incorporada, adaptada y especializada por las empresas, de tal manera que los gerentes y directivos en general puedan tener una mejor gestión y análisis de la información.

Dicha metodología consiste en un modelo estándar, no propietario y de libre distribución. Se describe como un modelo jerárquico, consistente en un conjunto de tareas con cuatro niveles de abstracción: 1. fases; 2. para cada fase un conjunto de tareas genéricas; 3. las tareas generan unas situaciones específicas (por ejemplo: limpieza de datos), y por último, 4. instancias o procesos (decisiones y resultados)<sup>18</sup>.

Esta metodología, como se muestra en la Figura 2, brinda una ruta clara permitiendo determinar qué actividades desarrollar en qué fase, de tal manera que se logren cumplir los objetivos del proyecto de minería de datos. Las tareas pueden ejecutarse en diferente orden

16 MORENO, María y otros. “Obtención y validación de modelos de estimación de *software* mediante técnicas de minería de datos”. Bogotá: *Revista Colombiana de Computación*. Volumen 3, número 1, pág. 53-71, 2002.

17 CHAPMAN, P. y otros (1999). *The crisp-dm process model*. Technical Report, CRISPDM Consortium.

CRISP-DM; “Cross Industry Standard Process for Data Mining”, [en línea] <<http://www.crisp-dm.org>> [Consultado 05 de julio de 2007].

18 *Ibid.*

e incluso pueden repetirse al conocer nuevos resultados obtenidos a partir de otras tareas (*backtrack*). Los procesos posteriores se benefician de las experiencias de los anteriores. El aprendizaje durante los procesos puede desencadenar nuevas tareas, más centradas en los objetivos de gestión.

A continuación se hace una breve descripción de las fases de la metodología<sup>19</sup>:

### **Fase 1. Comprensión del negocio**

Incluye la comprensión de los objetivos y requerimientos del proyecto desde una perspectiva empresarial, con el fin de convertirlos en objetivos técnicos y en un plan para lograr éstos objetivos.

### **Fase 2. Comprensión de los datos**

Abarca la recolección inicial de datos para identificar la calidad de los mismos estableciendo las relaciones más evidentes y permitiendo tener un acercamiento a las primeras hipótesis.

### **Fase 3. Preparación de los datos**

Incluye las tareas de selección de los datos a los que se les va a aplicar la técnica de modelado (variables y muestras), limpieza de los datos, generación de variables adicionales, integración de diferentes orígenes de datos y cambios de formato. Esta fase se encuentra estrechamente relacionada con la fase de modelamiento ya que, de acuerdo con la técnica de modelado que se vaya a utilizar, los datos necesitarán ser procesados y formateados de maneras diferentes.

### **Fase 4. Modelamiento**

En esta fase se seleccionan las técnicas de modelado más apropiadas para el proyecto específico de minería de datos. Dichas técnicas se seleccionan en función de los siguientes criterios:

1. Ser apropiado al problema.
2. Disponer de datos adecuados.

3. Cumplir los requerimientos del problema.
4. Conocimiento de la técnica.

### **Fase 5. Evaluación de negocios**

En esta fase se procede a la generación y evaluación del modelo, no desde el punto de vista de los datos, sino del cumplimiento adecuado de los objetivos empresariales. Se debe revisar el proceso seguido, teniendo en cuenta los resultados obtenidos, para poder repetir alguno de los pasos en los que se hayan podido cometer errores. Si el modelo generado es válido en función de los objetivos establecidos en la primera fase, se procederá al despliegue del modelo.

### **Fase 6. Despliegue**

La generación y despliegue del modelo no determinan el final del proyecto. Incluso si el objetivo del modelo es “aumentar el conocimiento” de los datos, éstos deberán ser organizados y presentados de manera que la empresa pueda utilizarlos de manera adecuada. Dependiendo de los requisitos, la fase de despliegue puede ser tan simple como la generación de un informe de resultados o tan complejo como la aplicación y revisión nuevamente de todo el proceso de minería de datos.

En suma, CRISP-DM por no estar ligada a ninguna herramienta de software y por ser un estándar de amplia utilización, permite su libre aplicación como gestor de proyectos de minería de datos sobre diferentes herramientas orientadas a la inteligencia de negocios.

Conviene observar, sin embargo, que CRISP-DM no es la única guía que ha sido propuesta. También existen otras apropiadas o abiertas, como la desarrollada por la empresa SAS, denominada SEMMA (*sample, explore, modify, model, assess*)<sup>20</sup>, DMAMC<sup>21</sup> o las cinco aes<sup>22</sup>. Todas estas metodologías, sin embargo, adolecen de métodos o técnicas que permitan tomar adecuadamente los requisitos del proyecto. Más concretamente, aún no existe un proceso maduro que pueda calificarse como una metodología sólida, pues si bien, por ejemplo, CRISP-DM que establece un conjunto de tareas y actividades que deben ser ejecutadas en el proyecto, no establece con qué técnicas o modelos se debe implementar cada

19 Ibid.

20 Portal [www.sas.com](http://www.sas.com). “Descripción de la metodología SEMMA”, [en línea] <<http://www.sas.com/technologies/analytics/datamining/miner/semma.html>> [Consulta: 19 de abril de 2006].

21 Portal [www.isixsigma.com](http://www.isixsigma.com). “Consulta sobre metodología 6-Sigma” [en línea] <[http://www.isixsigma.com/sixsigma/six\\_sigma.asp](http://www.isixsigma.com/sixsigma/six_sigma.asp)> [Consulta: 23 de junio de 2006].

22 Laudon, K. C. (2002). *Sistemas de información gerencial, organización y tecnología de la empresa conectada en red*. Mexico: Ed. Prentice Hall.

actividad. Por lo anterior, se recomienda que en cada proyecto se adapte una metodología y se complemente con los recursos (*hardware, software* y humanos), técnicas y actividades propias de cada tarea.

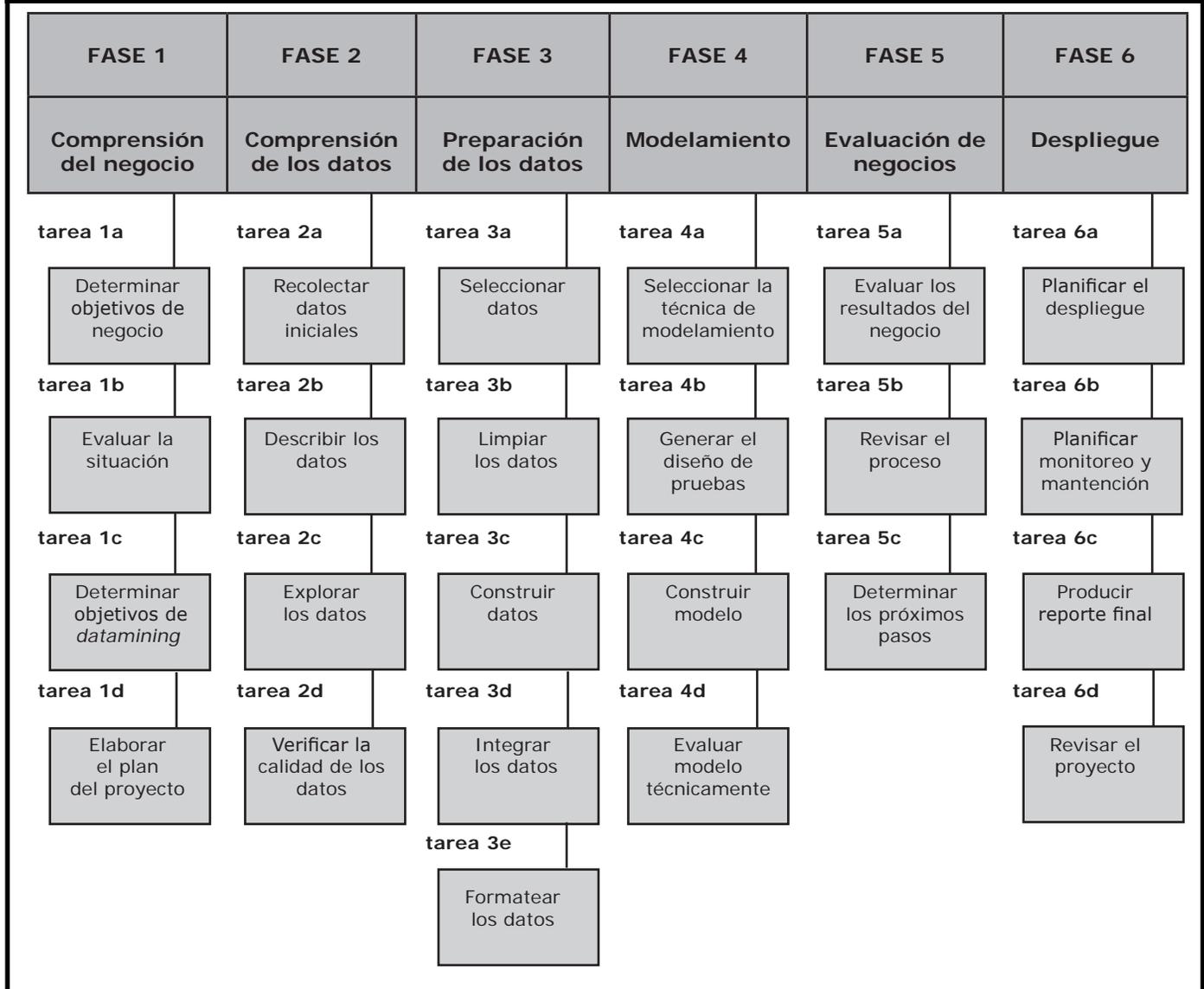
## DESCUBRIMIENTO DE CONOCIMIENTO EN DATOS CLÍNICOS

A través de un ejemplo y siguiendo los siguientes numerales, se generaliza el inicio y desarrollo de un proyecto de minería de datos:

La integración de las técnicas de minería de datos en el dominio de la medicina se está convirtiendo en algo habitual.

Otra característica importante es que los usuarios de estos nuevos sistemas son profesionales de la medicina que, aunque tienen ciertos conocimientos de estadística obligatorios en su formación, no tienen conocimientos de aprendizaje de modelos ni de la mayoría de técnicas presentadas en el punto anterior. Por tanto, los sistemas deben ser sencillos de manejar; los patrones descubiertos deben ser fáciles de entender (ya sean simbólicos o

**Figura N°. 2. Fases y tareas metodológicas CRISP-DM**



visuales) y la interrelación con el resto de sistemas informáticos de adquisición de datos, visualización y gestión de los centros asistenciales, debe ser transparente para el usuario.

Al llegar a este punto los nuevos sistemas de una manera cómoda y eficaz deben permitir<sup>23</sup>:

1. Asociación de síntomas y clasificación diferencial de patologías.
2. Estudio de factores (genéticos, precedentes, hábitos alimenticios, etc.) de riesgo/salud en distintas patologías.
3. Segmentación de pacientes para una atención más inteligente según su grupo.
4. Predicciones temporales de los centros asistenciales para mejor uso de los recursos, consultas, salas y habitaciones.
5. Estudios epidemiológicos, análisis de rendimientos de campañas de información, prevención, sustitución de fármacos, entre otros.

Finalmente, el objetivo primordial de aplicar técnicas de minería de datos es la adquisición, el descubrimiento y el mantenimiento de gran parte del conocimiento de una manera automática. La inclusión manual del mismo a partir de expertos u otras fuentes debería minimizarse, siempre que hubiera alternativa automática. Para ello, como se ha visto anteriormente, surge el KDD.

Los nuevos objetivos del KDD en medicina son<sup>24</sup>:

1. La interpretación comprensiva de los datos de los pacientes de una manera contextual y la presentación de tales interpretaciones de una manera visual o simbólica.
2. La extracción (descubrimiento) de conocimiento médico a partir del diagnóstico, revisiones médicas, seguimientos, terapias o tareas globales de gestión de los pacientes.

Dentro de la extracción de conocimiento se incluye el uso de conocimiento previo, ya sea para su refinamiento o particularización, o para ayudar al descubrimiento de nuevos patrones o modelos.

## A. Metodología

El ejercicio consiste en realizar un tipo de análisis de detección de patrones de comportamiento asociados al dominio de hepatitis, sobre un conjunto de datos con un número de 20 atributos incluido el atributo clase o etiquetado y 155 instancias de datos nominales y numéricos normales recolectados por la Carnegie-Mellon University. Cabe aclarar que las instancias poseen valores perdidos de algunos atributos.

Antes de seguir adelante es importante nombrar algunas herramientas de *software* comercial existentes en el mercado, tales como: SPSS Clementine, Salford CART/MARS/TreeNet/R, SAS, Angoss, KXEN, entre otras, y algunas herramientas de *software* libre como Yale, Weka, R, Orange y Knime. El *software* utilizado para el análisis de los datos clínicos en este documento es WEKA (Waikato Environment for Knowledge Analysis), que contiene una biblioteca de clases de aprendizaje en Java, con interfaces gráficas muy sencillas de utilizar y con las que se pueden aplicar y evaluar un gran número de algoritmos de minería a grandes conjuntos de datos. En <http://www.cs.waikato.ac.nz/ml/weka/><sup>25</sup> se puede encontrar toda la documentación necesaria y relevante sobre esta herramienta.

El formato de los datos en WEKA es como se describe en un archivo con extensión ARFF (Attribute-Relation File Format) donde se especifican los atributos y los datos, los cuales pueden tomar valores nominales y numéricos<sup>26</sup>.

Por otra parte, también se encuentran disponibles en internet un gran número de archivos ARFF para los investigadores que estén interesados en el estudio de datos médicos. Estos conjuntos de datos son el resultado de evaluaciones de datos clínicos realizadas por la Universidad de Waikato y la Universidad de California<sup>27</sup>.

Como estrategias de análisis de descubrimiento de conocimiento en datos médicos, se pueden utilizar métodos estadísticos, métodos basados en conocimiento (sistemas expertos) y métodos de aprendizaje automático supervisado (redes neuronales, reglas de asocia-

23 Hernández, O. José y otros (2004). Op. cit.

24 Lavrac, N. Op. cit.

25 Machine Learning Project. "Minería de Datos" 2006. [en línea] <<http://www.cs.waikato.ac.nz/ml/weka/>> [Consultado 02 de septiembre 2006].

26 Portal <<http://www.cs.waikato.ac.nz/ml/weka/>>, "Consulta sobre la herramienta WEKA, archivos ARFF, algoritmos y técnicas de minería de datos" [en línea], <<http://www.cs.waikato.ac.nz/~ml/publications.html>> [Consulta: 23 de junio de 2006].

27 Ibid.

ción, sistemas inmunes, árboles de decisión) y no supervisado (aprendizaje bayesiano, vecino más cercano, *fuzzy logic*, *support vector machines* y redes neuronales). Para el ejemplo descrito en el modelo de estudio se escogieron árboles de decisión, que son utilizados fundamentalmente para clasificación; además es uno de los más sencillos y fáciles de implementar y a su vez, de los más poderosos<sup>28</sup>.

Resumiendo y para entender mejor lo descrito en el modelo de estudio, a continuación se describe brevemente la inducción de árboles de decisión:

Un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final se puede determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. Una de las ventajas de los árboles de decisión es que, en su forma más general, las opciones posibles a partir de una determinada condición, son excluyentes. Esto permite analizar una situación y, siguiendo el árbol de decisión apropiadamente, llegar a una sola acción o decisión. Estos algoritmos se llaman algoritmos de partición o algoritmos de “divide y vencerás”.

Un árbol de decisión toma de entrada un objeto o situación descrita por un conjunto de atributos y regresa una decisión de verdadero / falso. Cada nodo interno corresponde a una prueba en el valor de los atributos y las ramas están etiquetadas con los posibles valores de la prueba. Cada hoja especifica el valor de la clase.

Los árboles de decisión están limitados a hablar de un solo objeto; es decir, son esencialmente proposicionales, siendo cada prueba de atributo una proposición. Por lo tanto, cualquier función booleana (función matemática cuyas variables son binarias y están unidas mediante los operadores del álgebra de Boole suma lógica (+), producto lógico (•) o negación (‘)) puede ser descrita por un árbol de decisión.

Para muchas funciones, los árboles son relativamente pequeños. Sin embargo, para otras funciones se puede requerir de un árbol exponencialmente grande. Para  $n$  atributos, hay  $2n$  filas. La salida se puede considerar como una función definida por  $2n$  bits. Con esto hay  $(2^2 \wedge n)$  posibles funciones diferentes para  $n$  atributos (para 6 atributos, hay  $2 \times 10^{19}$ ).

Cuando se realiza inducción de árboles de decisión a partir de ejemplos, un ejemplo es descrito por los valores de los atributos y el valor del predicado meta. Al valor del predicado meta se le llama “la clasificación del ejemplo”. Si el predicado es verdadero, entonces el ejemplo es positivo, si no, el ejemplo es negativo. En caso de existir más clases, los ejemplos de una sola clase son positivos y el resto de los ejemplos se consideran negativos.

Para elegir qué atributos y en qué orden aparecen en el árbol, se utiliza una función de evaluación: ganancia de entropía. La entropía es la cantidad de bits, en promedio, que harían falta para codificar los ejemplos; es equivalente a la medida de la “cantidad de información” representada. Si se tienen  $p$  ejemplos positivos y  $n$  ejemplos negativos, entonces la entropía o medida de la “cantidad de información” representada en el conjunto (S) es:

$$Et(S) = -p \log_2(p) - n \log_2(n) \quad (1)$$

Caso binario donde:  $p$  y  $n$  = proporciones de ejemplos positivos/negativos.

Un atributo normalmente no proporciona toda esta información, pero se puede estimar viendo cuánta información se necesita después de utilizarlo; cada atributo  $A$  divide los ejemplos del conjunto de entrenamiento en subconjuntos  $E_1, E_2, \dots, E_v$  de acuerdo con los valores del mismo.

Cada subconjunto  $E_i$  tiene  $p_i$  ejemplos positivos y  $n_i$  negativos, por lo que para cada rama se necesita:

$$I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right) \quad (2)$$

La fórmula (2) describe la cantidad de información para responder a una pregunta.

<sup>28</sup> Ibid.

Si todos los ejemplos son positivos o negativos, por ejemplo, pertenecen todos a la misma clase, la entropía será 0. Una posible interpretación de esto, es considerar la entropía como una medida de ruido o desorden en los ejemplos. Se define la ganancia de información como la reducción de la entropía causada por particionar un conjunto de entrenamiento S, con respecto a un atributo A:

$$E(A) = \sum_{i=1}^v \frac{p_i + n_i}{p + n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right) \quad (3)$$

La cantidad de información que se gana al seleccionar un atributo está dada por:

$$\text{Ganancia}(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - E(A) \quad (4)$$

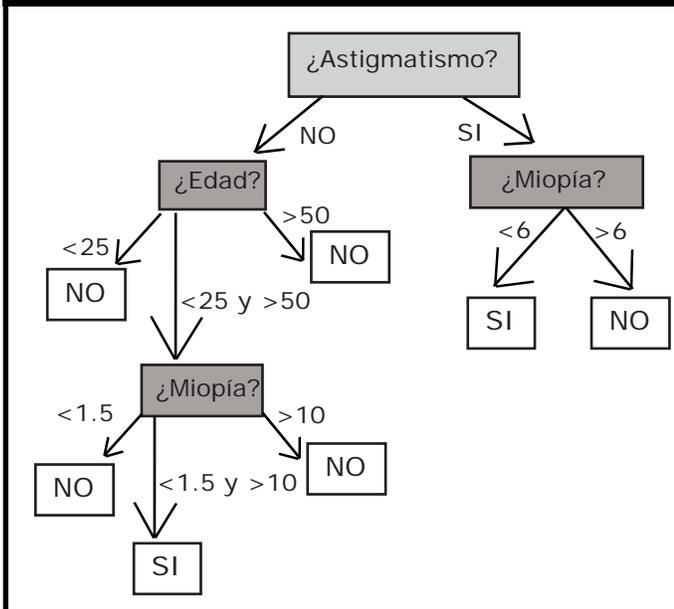
La ganancia de A dice el número de bits que se ahorran para responder a la pregunta clase de un ejemplo, dado que se conoce el valor del atributo A.

Dicho de otra forma, mide qué tan bien un atributo separa los ejemplos de entrenamiento de acuerdo con la clase meta o etiquetada.

La función evaluadora escoge el atributo con mayor ganancia de entropía en cada iteración.

Por ejemplo<sup>29</sup>, un hospital público en el que se realizan operaciones de cirugía refractiva (LASIK) a los miopes que lo soliciten. Evidentemente, dichas operaciones no están indicadas en muchos casos, y algunos podrían ser excluidos en una primera fase, con el objetivo de evitar riesgos potenciales o efectos secundarios. Aunque la indicación o no de dicha cirugía requiere un examen minucioso por parte del servicio de oftalmología del hospital, existen algunas condiciones claras por las cuales se puede determinar si, en principio, una persona está indicada para el estudio detallado (tensión ocular y paquimetría) y, finalmente, para la operación. En la figura 3 se muestra un ejemplo del árbol de decisión que se utiliza para admitir solicitudes.

**Figura N°. 3. Árbol de decisión para determinar recomendación de cirugía ocular**



Como se puede observar en la figura, es sencillo aplicar el árbol de decisión a un nuevo paciente para saber si se le ha de recomendar o no para dicha operación. Basta realizar las preguntas y seguir las respuestas hasta alguna de las hojas del árbol, catalogadas con un “no” o un “sí”. Este árbol de decisión en concreto funciona como un “clasificador”; es decir, dado un nuevo individuo, lo clasifica en una de las dos clases posibles: “no” o “sí”.

Los árboles de decisión se pueden expresar como un conjunto de reglas, con respecto al ejemplo de la cirugía refractiva. En las siguientes líneas se muestra un ejemplo del árbol de decisión anterior expresado en forma de reglas:

¿Operación?

1. Si astigmatismo = No y 25 < edad <= 50 y 1.5 < miopía <= 10 Entonces sí
2. Si astigmatismo = sí y miopía <= 6 entonces sí
3. En otro caso no

29 Hernández, O. José y otros (2004). Op. cit.

La representación en forma de reglas suele ser, en general, más concisa que la de los árboles, ya que permite incluir condiciones y permite el uso de reglas por defecto, como la que comienza por “en otro caso”, en el ejemplo anterior.

Para terminar, cabe recalcar que uno de los aspectos más importantes en los sistemas de aprendizaje de árboles de decisión es el denominado criterio de partición, ya que una mala elección de la partición (especialmente en las partes superiores del árbol) generará un peor árbol.

## B. Modelo de estudio

A continuación se hace una breve descripción del trabajo realizado en el área de análisis de datos clínicos a través de técnicas y del uso de herramientas de minería de datos. La limitación de espacio con la que se cuenta obliga a dedicar estas líneas sólo a las aportaciones más relevantes.

Una de las formas de usar WEKA es aplicar un método de aprendizaje (clasificador), inducción de árboles de decisión al conjunto de datos hepatitis.arff y analizar la salida para extraer información sobre los datos. Lo primero que se debe hacer es estructurar los datos y normalizarlos, con el objeto de hacer la carga del archivo en el *software* WEKA. Los datos estructurados y normalizados para utilizar en el ejercicio son (atributo: instancias o valores que puede tomar el atributo en un momento determinado)(ver columna siguiente):

Antes de continuar se sugiere que el lector esté familiarizado con el dominio que se evalúa aquí en este aparte, para tal fin, se recomienda revisar: Hepatitis - Guías y Revisiones de la Fundación Ginebrina para la Formación y la Investigación Médica, Falla Hepática Aguda, Hepatitis en Urgencias, Encefalopatía Hepática y Health-Cares.net, que son temas, documentos y enlaces que enriquecen el conocimiento en cuanto a la evaluación y manejo de la hepatitis, así como presentan una gran cantidad de lecturas recomendadas al respecto<sup>30</sup>.

1. Class: die, live
2. Age: 10, 20, 30, 40, 50, 60, 70, 80
3. Sex: male, female
4. Steroid: no, yes
5. Antivirals: no, yes
6. Fatigue: no, yes
7. Malaise: no, yes
8. Anorexia: no, yes
9. Liver big: no, yes
10. Liver firm: no, yes
11. Spleen palpable: no, yes
12. Spiders: no, yes
13. Ascites: no, yes
14. Varices: no, yes
15. Bilirubin: 0.39, 0.80, 1.20, 2.00, 3.00, 4.00
16. Alk phosphate: 33, 80, 120, 160, 200, 250
17. Sgot: 13, 100, 200, 300, 400, 500
18. Albumin: 2.1, 3.0, 3.8, 4.5, 5.0, 6.0
19. Prottime: 10, 20, 30, 40, 50, 60, 70, 80, 90
20. Histology: no, yes

Weka permite la aplicación de diversos algoritmos de clasificación, y para este ejercicio se prestó especial atención en el algoritmo J4.8, que presenta los mejores resultados de precisión, cobertura y coste computacional para el ejemplo descrito en este aparte. También se ha escogido este algoritmo por ser quizás el más representativo y uno de los algoritmos más utilizados en la práctica<sup>31</sup>.

La herramienta Weka también permite aplicar cuatro tipos de pruebas distintas:

**Use training set:** mide la calidad del clasificador para predecir la clase de las instancias en las que ha sido entrenado. Útil cuando se tienen pocas muestras en el conjunto.

**Supplied test set:** evalúa la calidad del clasificador para predecir la clase de un conjunto de instancias cargadas desde un archivo.

**Cross-validation:** evalúa la calidad del clasificador mediante validación cruzada, usando el número de grupos que se especifiquen. Este tipo de prueba es la utilizada para la evaluación del caso de estudio.

30 Valcarcel Asencios, Violeta (2004). *Data mining y el descubrimiento del conocimiento*. Lima: Universidad Nacional Mayor de San Marcos. Cios, K. J.(2001). *Medical data mining and knowledge discovery*. New York: Physica-Verlag Heidelberg.

31 Portal <<http://www.cs.waikato.ac.nz/ml/weka/>>. Op. cit.

**Percentage split:** evalúa la calidad del clasificador según lo bien que clasifique un porcentaje de los datos que se reservan para “test”<sup>32</sup>.

Realizado el corrimiento del clasificador J4.8 se observa la siguiente información<sup>33</sup>:

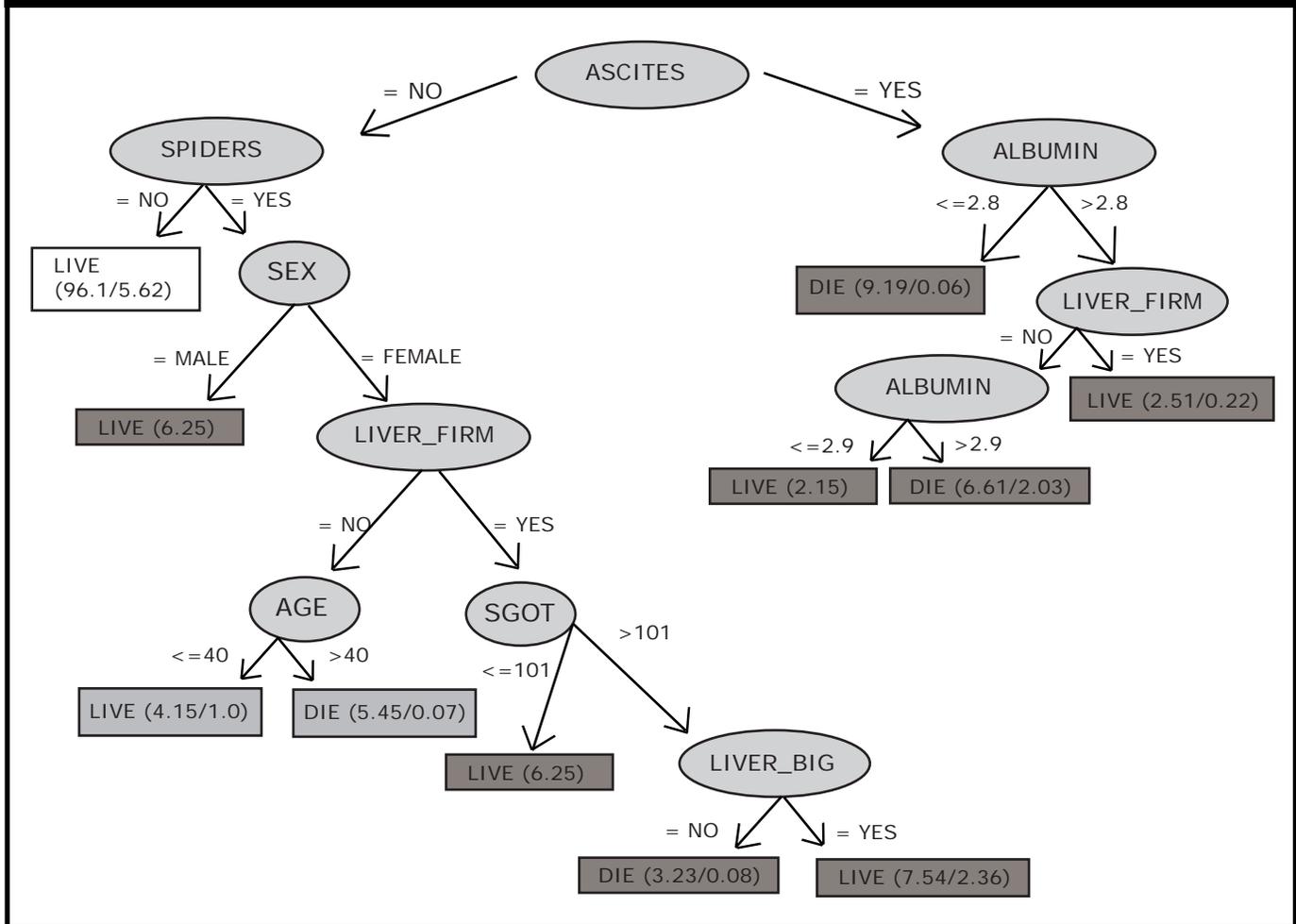
**Run Information:** muestra el esquema utilizado para tratar los datos (el nombre de la clase empleada y los parámetros usados), el número de instancias, una lista de los atributos presentes y el modo de “test” para las cuatro anteriores.

```

=== Run information ===
Scheme: weka.classifiers.trees.j48 -c 0.25 -m 2
Relation: hepatitis
Instances: 155
Attributes: 20
        [list of attributes omitted]
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===
    
```

**Figura N° . 4. Árbol de decisión para determinar esperanza de vida en pacientes con hepatitis**



32 Witten, I. H., & Frank, E. (1999). Op. cit.  
 33 Machine Learning Project. Op. cit.

**Descripción del modelo del clasificador:** depende del algoritmo utilizado; aquí se presenta la estructura del árbol de clasificación resultante (ver Figura 4). En cada hoja y rama se especifica el criterio de división, y en las hojas finales aparece para la clase que se especifica el número de casos correctamente clasificados y el número de casos mal clasificados. Si todos los datos pertenecen a la clase correcta, solo aparece dicho número. Las ramas que aparecen son aquellas que clasifican el mayor número de casos con el menor error posible (menor número de casos erróneos).

En este caso, el primer atributo es "ASCITES" porque es el que produce la división de los datos con entropía mínima en ese nivel, y análogamente, con el resto. Para clasificar un ejemplo nuevo se sigue el árbol de arriba abajo; la hoja final es la categoría inferida. Los caminos, desde la raíz hasta los nodos hoja, se pueden ver como reglas, donde el antecedente está formado por la intersección de los pares atributo-valor de los caminos.

Para todos los casos se han realizado validaciones cruzadas para crear los conjuntos de pruebas y entrenamiento, tomando los datos y dividiéndolos de forma aleatoria en diez subconjuntos mutuamente excluyentes del mismo tamaño aproximadamente, y utilizando nueve de ellos para la inducción del modelo y uno para la prueba. Cuando estamos en el primer tipo de prueba se hace una evaluación de los datos, *evaluation on training set* y *stratified cross-validation* en el resto. El proceso de inducción se repite diez veces, de manera que en cada iteración se elige un subconjunto distinto como conjunto de prueba, utilizando los restantes para el entrenamiento. Posteriormente se han realizado, también de forma aleatoria, particiones diferentes y se ha vuelto a inducir el modelo de forma iterativa.

A continuación se presentan las estadísticas: porcentaje de instancias clasificadas correctamente (precisión), estadísticas kappa (mide lo que se ajusta la predicción a la clase real, 1.0 significa ajuste total), error medio, error cuadrático medio, error relativo y error cuadrático relativo (más útiles cuando se hace regresión que en clasificación). Es necesario recalcar que Kappa es siempre menor o igual a 1. Un valor de 1 implica que la predicción se ajusta a la clase real; un valor menor de 1 representa un pobre ajuste. Una posible interpretación de la estadística Kappa puede ser<sup>34</sup>:

- Ajuste pobre = Menor de 0.20
- Acuerdo justo = 0.20 a 0.40
- Ajuste moderado = 0.40 a 0.60
- Buen ajuste = 0.60 a 0.80
- Muy de acuerdo = 0.80 a 1.00

#### === Stratified cross-validation ===

##### === Summary ===

Correctly Classified Instances	130	83.871 %
Incorrectly Classified Instances	25	16.129 %
Kappa statistic	0.436	
Mean absolute error	0.2029	
Root mean squared error	0.363	
Relative absolute error	61.4384 %	
Root relative squared error	89.6358 %	
Total Number of Instances	155	

## C. Evaluación

Los procesos y medidas de evaluación son los mismos para todos los experimentos: dada la colección de datos, una parte de la misma es considerada como conjunto de entrenamiento y el resto como conjunto de examen. Así, los modelos aprenden del conjunto de entrenamiento y tratan de inferir las categorías de los ejemplos del conjunto de prueba. Puesto que las categorías de éstos últimos son conocidas, se pueden validar las inferencias de los modelos. Así pues, esta validación se realiza para cada categoría mediante tres medidas típicas: precisión, cobertura y medida-F<sup>35</sup>.

Los parámetros de exactitud para cada clase son los siguientes:

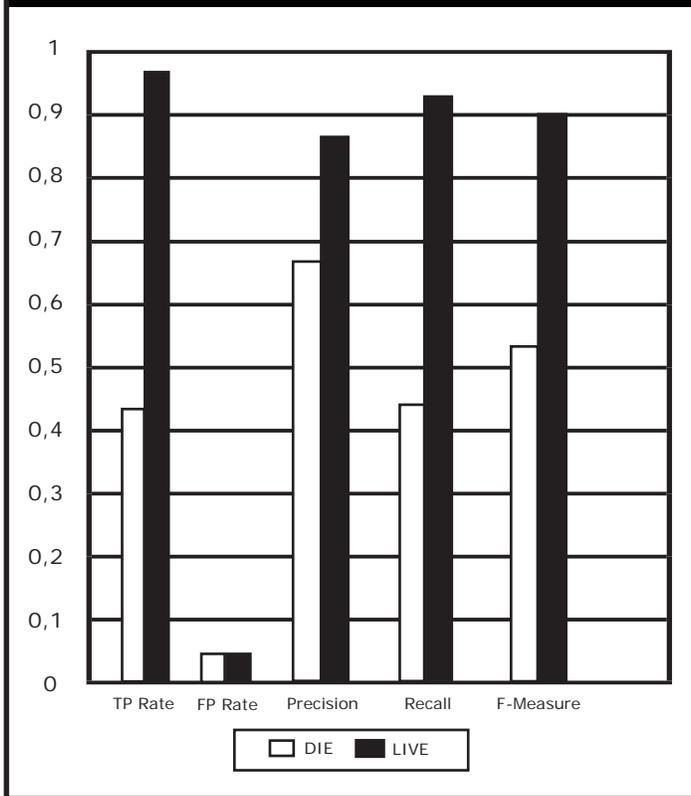
#### === Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
0.438	0.057	0.667	0.438	0.528	DIE
0.943	0.563	0.866	0.943	0.903	LIVE

<sup>34</sup> Hernández, O. José y otros (2004).

<sup>35</sup> Machine Learning Project. Op. cit.

**Figura N°. 5. Gráfica de comparación de medidas de exactitud para las clases "die" y "live"**



A continuación se presentan los detalles de exactitud (ver Figura 5) para cada clase (*detailed accuracy y class*) y finalmente se procede a realizar la evaluación del modelo inducido.

**True positive (TP) rate:** es la proporción de ejemplos que fueron clasificados como clase x ("die", "live"), de entre todos los ejemplos que realmente tienen clase x; es decir, la cantidad de la clase que ha sido capturada. En la matriz de confusión, es el valor del elemento de la diagonal dividido por la suma de la fila relevante.

**False positive (FP) rate:** es la proporción de ejemplos que fueron clasificados como clase x, pero en realidad pertenecen a otra clase de entre todos los ejemplos que no tienen clase x. En la matriz de confusión, es la suma de la columna menos el valor del elemento de la diagonal dividido por la suma de las filas de las otras clases.

**Precision:** precisión es la proporción de ejemplos que de veras tienen clase x entre todos los que fueron clasificados como tal. En otras palabras el indicador de precisión es el resultado que se presenta en forma de clasificación o estimación, medido a través del porcentaje de predicciones que son correctas. Cuando se habla de clasificación se emplea el porcentaje de casos bien clasificados, y para la estimación se emplea el porcentaje de registros con una estimación que el clasificador considere correcta. En la matriz de confusión, la precisión es el elemento de la diagonal dividido por la suma de la columna relevante.

**Recall:** cobertura es el resultado de dividir el número de ejemplos recuperados que son relevantes, sobre el total de elementos que son relevantes. Es el coeficiente que mide el porcentaje de registros en los cuales se puede aplicar una regla. Se corresponde con el denominador que se ha empleado para el cálculo del coeficiente de confianza y/o precisión. Por ejemplo, si en el conjunto de datos que se ha utilizado en el desarrollo de este ejercicio se dispone de 15 registros, de los cuales el atributo "albumin": albúmina se presenta en 10 ocasiones con un valor menor de 2.8, que es precisamente la condición impuesta por la regla "if (albumin <= 2.8) then die". Así, el coeficiente de cobertura se sitúa para dicha regla en el 0.66 (10/15).

**F-measure:** es simplemente el resultado de la ecuación:

$$\frac{2 * Precision * Recall}{Precision + Recall} \quad (5)$$

Una medida combinada de "precision" y "recall". Representa, en cierto modo, la intersección entre los ejemplos implicados en la precisión y la cobertura, normalizada mediante la suma de ambas.

#### D. Validación del modelo inducido

Para realizar la validación del modelo inducido con todos los datos de entrenamiento, se aplica la técnica de

las matrices de confusión, o también la tabla de contingencia que muestran el tipo de las predicciones correctas e incorrectas cuando se aplica el modelo sobre el conjunto de prueba. La matriz de confusión ayuda a conocer la bondad del modelo para predecir y también para descubrir en qué casos se producen errores<sup>36</sup>.

=== Confusion Matrix ===			
a	b	<-- Classified as	
14	18		a = DIE
7	116		b = LIVE

Una tabla de contingencia está formada por tantas filas y columnas como clases hay. El número de instancias clasificadas correctamente es la suma de la diagonal principal de la matriz; el resto están clasificadas de forma incorrecta e indican el tipo de error cometido (qué valor ha predicho el modelo y cuál es el valor verdadero). En el ejemplo que se está tratando no se clasifican correctamente todos los registros, aunque el peso de las clasificaciones incorrectas es muy pequeño; además, éstas se encuentran próximas a la diagonal, lo que indica que el error cometido no es demasiado elevado.

Como se ha visto, la información que se da aquí y en el apartado anterior es la misma expresada de otra forma. Por tanto, se presentaron solamente los resultados de la matriz de confusión con 130 instancias bien clasificadas equivalentes al 83.871% del número total de instancias del modelo, ya que comprobando el número de elementos no nulos (25 instancias mal clasificadas equivalentes al 16.129% del número total de instancias, y que se pueden visualizar sumando los valores que se encuentran en la diagonal secundaria de la matriz de confusión,) se puede demostrar que J4.8 es un buen algoritmo para este caso. El número de personas que en este análisis tienen posibilidades de vivir es de 116 y de morir, un total de 14.

¿LIVE?

1. **IF** (ASCITES=no) **THEN** LIVE
2. **IF** (ASCITES=no) **AND** (SPIDERS=no) **THEN** LIVE
3. **IF** (ASCITES=no) **AND** (SPIDERS=yes) **AND** (SEX=MALE) **THEN** LIVE
4. **IF** (ASCITES=no) **AND** (SPIDERS=yes) **AND** (SEX=FEMALE) **AND** (LIVER\_FIRM=no) **AND** (AGE<=40) **THEN** LIVE
5. **IF** (ASCITES=no) **AND** (SPIDERS=yes) **AND** (SEX=FEMALE) **AND** (LIVER\_FIRM=yes) **AND** (SGOT<=101) **THEN** LIVE
6. **IF** (ASCITES=no) **AND** (SPIDERS=yes) **AND** (SEX=FEMALE) **AND** (LIVER\_FIRM=yes) **AND** (SGOT>101) **AND** (LIVER\_BIG=yes) **THEN** LIVE
7. **IF** (ASCITES=yes) **AND** (2.8<ALBUMIN<=2.9) **THEN** LIVE
8. **IF** (ASCITES=yes) **AND** (ALBUMIN>2.9) **AND** (PROTIME>52) **THEN** LIFE
9. **IN ANOTHER CASE** DIE

## E. Interpretación

Los patrones asociados al comportamiento de la hepatitis se pueden representar mediante reglas, como se muestra en el siguiente recuadro.

Interpretando el árbol de decisión en el modelo de estudio y las reglas descritas anteriormente, lo que resalta, desde luego, es que el mayor problema asociado a las personas que fueron diagnosticadas es la ascitis “ascites”, que es una de las complicaciones de la cirrosis y consiste en la acumulación de líquido en la cavidad abdominal. La presencia de ascitis en una persona enferma del hígado se considera una indicación de trasplante hepático<sup>37</sup>. Las causas de ascitis son muy variadas, desde infecciones hasta insuficiencia cardiaca. Sin embargo, la causa más frecuente es la cirrosis hepática. Asociados a la ascitis existen otros atributos que pueden complicar o no la valoración del paciente, como se puede observar si un paciente presenta cuadro de ascitis y ausencia de proteínas específicas (albúmina: “albumin” y protrombina: “prottime”) o enzimas (“sgot”: enzima glutamato oxaloacetato deshidrogenada; un aumento de la actividad de esta enzima es indicador de muerte celular o daño severo en el hígado) para meta-

36 Cabena, P. y otros (1997). Op. cit.

37 Chen, H. (2005). *Medical informatics: knowledge management and data mining in biomedicine. Integrated series in information systems*. New York: Springer.

Aseervatham, S. y Osmani A. *Mining short sequential patterns for hepatitis type detection*. Université de Paris-Nord, Laboratoire LIPN-CNRS UMR 7030F-93430 Villetaneuse Cedex, France: ECML/PKDD Discovery Challenge, 2005.

bolizar diferentes sustancias en el hígado, las posibilidades que se tienen de sobrevivir son muy pocas.

En el conjunto de datos la mayoría de los pacientes son diagnosticados entre los 40 y los 60 años y se presenta un comportamiento asociado muy interesante como lo es el atributo "spiders" (arañas vasculares: dilatación de capilares sanguíneos a modo de patas de araña), que como caso típico se encuentra fuertemente coligado al género femenino, porque aunque no es exclusivo de las mujeres, sí es cierto que a ellas, las pequeñas lesiones en las venas suelen afectarlas más que a los hombres.

Evidentemente, muchas personas desarrollan síntomas tales como pérdida de apetito, malestar general, fiebre, náuseas y vómitos, dolor en los músculos, fatiga, dolor de cabeza, agrandamiento del hígado, etc., dependiendo de la enfermedad, antes de ser diagnosticados y asistidos por especialistas. Como aporte a la medicina, las técnicas de minería de datos, ayudan a los médicos en la detección de patrones de comportamiento de enfermedades, infecciones, posibles complicaciones, entre otros, que son relevantes en el momento de tomar una decisión inherente al tratamiento de los mismos<sup>38</sup>.

Para finalizar, es necesario resaltar que el número de veces que se ejecutan los procesos y por el cual está enmarcada la métrica, debe ser determinado por un grupo de especialistas en medicina. Estas métricas son diseñadas dependiendo de los criterios relevantes y asociados a los patrones de comportamiento de la gran variedad de enfermedades existentes actualmente, así como hay que tener en cuenta las actividades desarrolladas por el organismo de salud y la infraestructura tecnológica, entre otros.

## CONCLUSIONES

La aplicación de técnicas de minería de datos en las empresas ofrece una excelente oportunidad a las directivas en general de tener una visión amplia del negocio para hacer proyecciones y descubrimientos de patrones en el comportamiento de sus dominios, y, en definitiva, garantizar más competitividad y lograr así su permanencia y crecimiento en el mercado. Además existe un buen número de herramientas de despliegue

de los modelos de minería basados en *software* libre, lo cual rompe con la limitación de muchas de ellas en cuanto a la inversión de programas costosos, con requerimientos e infraestructura que muchas veces están fuera de su alcance.

Las técnicas estadísticas son fundamentales a la hora de validar hipótesis y analizar datos, por lo cual la estadística desempeña un papel importante en KDD. Asimismo, destacamos que, en particular, los sistemas de KDD pretenden automatizar el proceso completo de análisis de datos y descubrir modelos que permitan generar nuevas posibilidades de descubrimiento de conocimiento en grandes conjuntos de datos.

Con la aplicación de técnicas de descubrimiento de conocimiento, los estudios realizados y validados pueden ser enfocados a la mejora de cualquier organización de la que se pueda extraer un gran número de datos estructurados.

Queda pendiente de momento emprender más estudios reales y verídicos en conjuntos de datos aportados por los sistemas transaccionales de diferentes organizaciones que mediante la aplicación de las diferentes técnicas de minería de datos generen resultados predictivos útiles para prever el comportamiento futuro de algún tipo de comportamiento o dominio, con el único objeto de mejorar la calidad de los sistemas organizacionales. Este documento ofrece información general sobre el descubrimiento de patrones de comportamiento asociados al dominio de la Medicina, específicamente a un conjunto de datos de hepatitis. No quiere esto decir que puede ser considerado un consultor, y se sugiere comprobar siempre con un especialista cuando se tenga una pregunta sobre la forma de gestionar y administrar la información.

Se pretende motivar a los administradores y expertos del comercio a usar herramientas especializadas de descubrimiento de conocimiento a partir de conjuntos de datos estructurados de sus sistemas, para apoyar las tareas de toma de decisiones. Además de inducirlos a que se vayan familiarizando con las nuevas herramientas y sistemas que, bien serán cada día más potentes y posibles de manejar.

38 Pizzi, L. C., Ribeiro, M. X., Vieira, M. T. P. (2005). *Analysis of hepatitis dataset using multirelational association rules*. Department of Computer Science, Federal University of São Carlos, São Carlos, Brazil: ECML/PKDD Discovery Challenge, 2005.

## BIBLIOGRAFÍA

- TURBAN, E. (2008). *Business intelligence: A managerial approach*. Upper Saddle River, N.J.: Pearson Prentice Hall.
- WILLIAMS, S., y WILLIAMS, N. (2007). *The profit impact of business intelligence*. Amsterdam: Elsevier/Morgan Kaufmann.
- CASTILLO HERNÁNDEZ, Mario (2006). *Toma de decisiones en las empresas: entre el arte y la técnica: metodologías, modelos y herramientas*. Bogotá: Ediciones Uniandes.
- VITT, E.; LUCKEVICH, M. y MISNER, S. (2003). *Business intelligence: técnicas de análisis para la toma de decisiones*. IT/Tecnología y Empresa. Madrid: McGraw-Hill Interamericana.
- MOSS, L. T. y ATRE, S. (2003). *Business intelligence roadmap: the complete project lifecycle for decision support applications*. Boston, MA: Addison-Wesley.
- FAYYAD, U., PIATESKY-SHAPIRO, G. y SMYTH, P. (1996). *Advances in knowledge discovery and data mining*. Massachusetts: MIT Press.
- BERRY, M. J. A., y Linoff, G. (2004). *Data mining techniques for marketing, sales, and customer relationship management*. Indianapolis: Wiley.
- LAVRAC, N. "Selected techniques for data mining in medicine". *Artificial Intelligence in Medicine*. Vol. 16 (1), pp. 3-23, 1999.
- MORENO, María y otros. "Obtención y validación de modelos de estimación de software mediante técnicas de minería de datos". Bogotá: *Revista Colombiana de Computación*. Volumen 3, número 1, pág. 53-71, 2002.
- HERNÁNDEZ, O. José y otros (2004). *Introducción a la minería de datos*. Madrid: Pearson Educación S.A.
- WITTEN, I. H. y FRANK, E. (1999). *Data mining: Practical machine learning tools and techniques with Java implementations*. San Francisco: Morgan Kaufmann.
- WITTEN, I. H., & FRANK, E. (2005) *Data Mining: Practical machine learning tools and techniques*. San Francisco: Morgan Kaufmann, Segunda edición.
- PÉREZ, César y SANTÍN, Daniel (2006). *Data mining: soluciones con enterprise miner*. Madrid: Alfaomega Ra-Ma.
- CHAPMAN, P. y otros (1999). *The crisp-dm process model*. Technical Report, CRISPDM Consortium.
- CRISP-DM; "Cross Industry Standard Process for Data Mining", [en línea] <<http://www.crisp-dm.org>> [Consultado 05 de julio de 2007].
- Portal [www.sas.com](http://www.sas.com), "Descripción de la metodología SEMMA", [en línea] <<http://www.sas.com/technologies/analytics/datamining/miner/semma.html>> [Consulta: 19 de abril de 2006].
- Portal [www.isixsigma.com](http://www.isixsigma.com), "consulta sobre metodología 6-Sigma" [en línea] <[http://www.isixsigma.com/sixsigma/six\\_sigma.asp](http://www.isixsigma.com/sixsigma/six_sigma.asp)> [Consulta: 23 de junio de 2006].
- LAUDON, K. C. (2002). *Sistemas de información gerencial, organización y tecnología de la empresa conectada en red*. Mexico: Ed. Prentice Hall.
- Machine Learning Project. "Minería de Datos" 2006. [en línea] <<http://www.cs.waikato.ac.nz/ml/weka/>> [Consultado 02 de septiembre 2006].
- Portal <http://www.cs.waikato.ac.nz/ml/weka/>, "Consulta sobre la herramienta WEKA, archivos ARFF, algoritmos y técnicas de minería de datos" [en línea], <<http://www.cs.waikato.ac.nz/~ml/publications.html>> [Consulta: 23 de junio de 2006].
- CABENA, P. y otros (1997). *Discovering data mining: from concept to implementation*. New York: Editorial Prentice Hall, Upper Saddle River.
- VALCARCEL ASENCIOS, Violeta (2004). *Data mining y el descubrimiento del conocimiento*. Lima: Universidad Nacional Mayor de San Marcos.
- CIOS, K. J.(2001). *Medical data mining and knowledge discovery*. New York: Physica-Verlag Heidelberg.
- Two Crows Corporation. (1999). *Introduction to data mining and knowledge discovery*. Potomac, MD: Two Crows Corp.
- CHEN, H. (2005). *Medical informatics: knowledge management and data mining in biomedicine. Integrated series in information systems*. New York: Springer.
- ASEERVATHAM, S. y OSMANI A. *Mining short sequential patterns for hepatitis type detection*. Université de Paris-Nord, Laboratoire LIPN-CNRS UMR 7030F-93430 Villetaneuse Cedex, France: ECML/PKDD Discovery Challenge, 2005.
- PIZZI, L. C.; RIBEIRO, M. X. yVIEIRA, M. T. P. *Analysis of hepatitis dataset using multirelational association rules*. Department of Computer Science, Federal University of São Carlos, São Carlos, Brazil: ECML/PKDD Discovery Challenge, 2005.