



MAESTRÍA DE GERENCIA DE SISTEMAS Y TECNOLOGÍAS DE LA INFORMACIÓN

**Formulación de un Marco de Trabajo y Ensayo de una Metodología para Inteligencia de
Negocios que Identifique Grupos de Contribuyentes con Importancia Fiscal del Ente
Administrador de la Gestión Tributaria en el Ecuador.**

**Trabajo de Titulación presentado en conformidad con los requisitos establecidos
para optar por el título de Magister en Gerencia en Sistemas y Tecnologías de la
Información**

Profesor Guía

Ing. Jaime Vinuesa Trujillo, MBA

Autora

Ing. Ana María Paredes Castro

Año

2014

DECLARACION DEL PROFESOR GUÍA

"Declaro haber dirigido este trabajo a través de reuniones periódicas con el estudiante, orientando sus conocimientos y competencias para un eficiente desarrollo del tema escogido y dando cumplimiento a todas las disposiciones vigentes que regulan los Trabajos de Titulación"

**Ing. Jaime Vinueza, MBA
C.C.1716028509**

DECLARACION DE AUTORIA DE ESTUDIANTE

“Declaro que éste trabajo es original, de mi autoría, que se han citado las fuentes correspondientes y que en su ejecución se respetaron las disposiciones legales que protegen los derecho de autor vigentes”

**Ana María, Paredes Castro
C.C.0401149703**

AGRADECIMIENTOS

A mis padres, familia y amigos por su apoyo y comprensión.

A Jaime Vinuesa y Rolando Mantilla quienes con su conocimiento, han sido los guías, para que éste emprendimiento pueda concluir.

DEDICATORIA

A mis padres, quienes una vez más, han apoyado el alcance de una nueva meta que me he propuesto.

A mi hermano y amigos quienes siempre han estado ahí para extenderme esa mano cuando más he necesitado.

RESUMEN

Actualmente, la Administración Tributaria del Ecuador se enfrenta a grandes retos dirigidos al control; la tecnología y su globalización, ha dado origen a nuevas formas de evasión y elusión fiscal. En solución a éste inconveniente el Administrador Tributario del Ecuador, Servicio de Rentas Internas, siempre ha estado a la vanguardia buscando diferentes y nuevas soluciones que han permitido que sea reconocido como el país más eficiente del continente, en recaudación de impuesto al valor agregado, de acuerdo al Banco Interamericano del Desarrollo. (Ecuador Inmediato.com, 2014)

Alineado a ésta premisa y buscando siempre una mejora continua, el objetivo principal del presente estudio es conocer a los Contribuyentes, Sociedades, clasificados de acuerdo a su Importancia Fiscal.

En sí, el desarrollo de la investigación, se enfoca en dos pilares: La estructuración de un concepto de Importancia Fiscal contextualizado en el control, basado netamente en la consolidación de criterios de los diferentes expertos tributarios a nivel país y la aplicación de una técnica de minería de datos que permita clasificar a los contribuyentes, Sociedades de acuerdo a su Importancia fiscal previamente establecida.

Finalmente los resultados indican que la clasificación realizada en base a la estructuración del concepto de Importancia Fiscal y la técnica de minería de datos ha dado como resultado tres grupos: Importancia Alta, media y baja, en conclusión, los casos de cada uno de los grupos servirán de apoyo para la toma de decisiones en relación a una Administración Tributaria contextualizada en el control.

ABSTRACT

Currently the Tax Administration of Ecuador faces great challenges in terms of control, technology, and globalization, which have led to new forms of tax evasion and avoidance. To address this issue, the Tax Administrator of Ecuador, Servicio de Rentas Internas, has always been at the forefront in searching for different and new solutions which have allowed the country to be recognized as the most efficient one in the continent in the collection of value added tax, in accordance with the Banco Interamericano del Desarrollo. (Ecuador Inmediato.com, 2014)

In line with this premise and always aiming at continuous improvement, the main objective of this study is to know the taxpayers and societies classified by their fiscal relevance.

The development of this research is focused on two main axes: the structuring of the fiscal relevance concept in the context of control, entirely based on the consolidation of criteria from different tax experts nationwide and the application of a data mining technique. Thus, taxpayers and societies will be classified in accordance with their previously established fiscal relevance.

Finally, the results show that the classification based on the structuring of the fiscal relevance concept and the data mining technique has led to three groups: High, medium, and low importance. Such groups help decision support making for a Tax Administration in the context of control.

ÍNDICE

1. INTRODUCCIÓN	1
1.1 Antecedentes	1
1.2 Justificación de la investigación	4
1.3 Objetivos de la investigación	7
1.3.1 Objetivo general	7
1.3.2 Objetivos específicos.....	7
2. MARCO DE REFERENCIA	8
2.1 Marco Teórico	8
2.1.1 La Administración tributaria y la importancia Fiscal	8
2.1.2 Descubrimiento de Conocimiento en Base de Datos	14
2.1.3 Minería de Datos	27
2.1.3.1 Importancia de la Minería de Datos.....	28
2.1.3.2 Técnicas de Minería de Datos	30
2.1.3.3 Arquitectura de Minería de Datos	47
2.1.3.4 Herramientas de Minería de Datos	49
2.1.4 Tablero de Mandos.....	52
2.1.4.1 Importancia de Tablero de Mandos	53
2.1.4.2 Tipos de Tablero de Mandos	54
Cuadro de Mando Operativo (CMO):	54
Cuadro de Mando Integral (CMI):.....	54
2.2. Marco Conceptual	56
3. METODOLOGÍA	59
3.1. Metodología de Inteligencia de Negocios	59
3.2. Fases	59
3.2.1 Fase de comprensión del negocio o problema	59
3.2.2 Fase de comprensión de los datos.....	66

3.2.3 Fase de preparación de los datos	73
3.2.4 Fase de modelado	77
3.2.5 Fase de evaluación	91
3.3. Entregables de la Metodología.....	92
4. CONCLUSIONES, RECOMENDACIONES	102
4.1. Conclusiones.....	102
4.2. Recomendaciones	106
REFERENCIAS.....	108
ANEXOS.....	112

Índice de Figuras

Figura 1. Modelo Integral de Gestión Estructural de Riesgo por Procesos	9
Figura 2. Fases de la Gestión de Riesgos	12
Figura 3. Metodologías de Minería de Datos.....	17
Figura 4. Proceso de Descubrimiento en Base de Datos KDD	18
Figura 5. Fases CRISP - DM.....	20
Figura 6. Técnicas de la Minería de Datos.....	31
Figura 7: Ejemplo de un Árbol de Decisión	34
Figura 8. Ejemplo de regresión logística	35
Figura 9. Ejemplo de Análisis Discriminante	36
Figura 10. Estructura de una Red Neuronal.....	37
Figura 11. Ejemplo de Mapas de Kohonen	38
Figura 12. Ejemplo de Agrupamiento, Kmedias	39
Figura 13. Ejemplo Vecino más cercano.....	40
Figura 14. Ejemplo de Correlación	41
Figura 15. Ejemplo de un Clasificador Bayes.....	42
Figura 16. Ejemplo de Regresión Lineal	43
Figura 17. Ejemplo de Maquinas de Soporte Vectorial	44
Figura 18. Ejemplo de Algoritmos Genéticos	45
Figura 19. Relación entre modelos y algoritmos	46
Figura 20. Arquitectura de Minería de Datos.....	48
Figura 21. Modelo de Entendimiento de Datos.....	69
Figura 22. Total de Activos.....	70
Figura 23: Verificación de Anomalías e Imputación de Datos en Blancos	75
Figura 24: Código CIIU, aplicado en el Ecuador	79
Figura 25. Calidad de Conglomerado K medias.....	81
Figura 26. Clusters K medias	85
Figura 27. Evaluación de la precisión de los Modelos Predictivos, Algoritmo Discriminante	86
Figura 28. Evaluación de la precisión de los Modelos Predictivos, Algoritmo Redes Neuronales.....	87

Figura 29. Evaluación de la precisión de los Modelos Predictivos, Algoritmo de Regresión.....	89
Figura 30. Evaluación de la precisión de los Modelos Predictivos, Algoritmo Árbol de Decisión C5.....	90
Figura 31. Modelo Clasificador de Sociedades	95
Figura 32. Importancia Fiscal por Regional, Datos.....	97
Figura 33. Importancia Fiscal por Regional, Porcentual.....	98
Figura 34. Código de cada Actividad.....	101

Índice de Tablas

Tabla 1. Fases utilizadas por cada una de las Metodologías	16
Tabla 2. Evolución de las tecnologías de base de datos.....	29
Tabla 3. Cuadrante Mágico para Plataformas Analíticas e Inteligencia de Negocios	51
Tabla 4. Perspectiva Financiera.....	55
Tabla 5. Riesgos y Contingencias.....	62
Tabla 6. Costo - Beneficio	63
Tabla 7. Beneficios de Implementación.....	64
Tabla 8. Descripción Variables.....	68
Tabla 9. Detalle Descriptivo de las Variables	72
Tabla 10. Casos Anómalos reportados por la Herramienta.....	76
Tabla 11. Código CIU, aplicado en el Ecuador.....	77
Tabla 12. Variables y sus respectivas Medias.....	82
Tabla 13. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Discriminante.....	83
Tabla 14. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Discriminante.....	84
Tabla 15. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Neuronales	85
Tabla 16. Matriz de Contingencias:Entrenamiento y Prueba, Algoritmo de Neuronales.....	86
Tabla 17. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Regresión	87
Tabla 18. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Árbol de Decisión	88
Tabla 19. Importancia Fiscal por Regional.....	
Tabla 20. Importancia Fiscal con relación a su Actividad Económica.....	99
Tabla 21. Letra de la Actividad con su respectivo significado.....	100

CAPÍTULO I: INTRODUCCIÓN

1.1. Antecedentes

Una retrospectiva de la historia visualiza que la gestión tributaria tiene su origen en los impuestos, no se sabe a ciencia cierta cuándo fue pero sí se puede citar que coincide con los inicios de la misma humanidad.

Con relación a la recaudación, los impuestos difieren de los actuales puesto que dependían del criterio de los caciques, soberanos o jefes que existían en esa época, pero en sí no se basaba en un procedimiento específico que garantice un cobro equitativo de acuerdo a los ingresos que cada persona percibía.

En el Ecuador, la gestión tributaria se origina también con los impuestos en la época de la Colonia, siglo XVI - XIX, de la misma forma el criterio para la recaudación se trataba de una imposición del rey a sus vasallos por la explotación de los recursos, a diferencia de la actualidad que se trata de una contribución que los ciudadanos realizan a cambio de dotación equitativa y eficiente de todos los bienes y servicios estatales. (Los Impuestos y la Colonia, 2012)

El responsable de la administración tributaria también ha sido un factor que se ha diferenciado en estos años en el Ecuador, de un no centralizado a una responsabilidad estatal, que según el artículo 1 del año 1997 de la Ley de creación del Servicio de Rentas Internas adquiere el nombre de Servicio de Rentas Internas SRI, conceptuado como:

“Una entidad técnica y autónoma que tiene la responsabilidad de recaudar los tributos internos establecidos por Ley mediante la aplicación de la normativa vigente. Su finalidad es la de consolidar la cultura tributaria en el país a efectos de incrementar sostenidamente el cumplimiento voluntario de

las obligaciones tributarias por parte de los contribuyentes” (Servicio de Rentas Internas Sitio Oficial, 2013)

Bajo esta premisa y alineado a uno de los objetivos del SRI que es:

“Hacer bien al país por cumplir a cabalidad la gestión tributaria, disminuyendo significativamente la evasión, elusión y fraude fiscal” (Servicio de Rentas Internas Sitio Oficial, 2013)

Cabe mencionar que en la actualidad los cambios que han existido en todos los ámbitos principalmente en la gestión fiscal tributaria han sido significativos gracias a un mundo globalizado y los cambios tecnológicos que éste conlleva, los mismos que han permitido que se tenga a la mano más herramientas que permitan garantizar una mejor administración tributaria, con mayor equidad en las contribuciones y evitando en mayor grado la evasión tributaria, es así como la tecnología se ha convertido en un ente estratégico y no solo un apoyo, facilitando el manejo de cada uno de los procedimientos necesarios para una recaudación más eficiente y eficaz.

Los diferentes países del mundo se han concienciado del gran apalancamiento que brinda la tecnología con relación a la Administración tributaria, un ejemplo de ello es el FAT, Foro sobre la Administración Tributaria, donde los países que la conforman poseen una página web que contiene un gran repositorio de información de todo lo tratado en las reuniones realizadas anualmente, tales como: mejores prácticas, innovaciones, investigaciones e intercambios de conocimientos acerca de una administración efectiva de impuestos cuyo objetivo primordial es la mejora de los servicios al contribuyente y el cumplimiento de las obligaciones tributarias. (OECD Organization for Economic Co-operation and Development, 2009)

Todo esto ha dado origen a una gran cantidad de información donde los países a nivel mundial puedan plantearse nuevos objetivos cuyo principal recurso utilizado son todos esos datos que se posee y que fueron ingresados,

intercambiados o gestionados, con los cuáles se puedan identificar nuevos patrones que ayuden a mejorar la administración tributaria.

La mayoría de países que conforman el G8 que ahora se denomina G20 han sido parte activa de reuniones como el CIAT, Centro Interamericano de Administraciones Tributarias o FAT donde entre los temas tratados se han planteado mejoras a la administración tributaria con ayuda de la tecnología, entre las estrategias que se han planteado resalta la segmentación de contribuyentes con ayuda de herramientas de minería de datos con la cual se ha facilitado el entendimiento de todo ese universo de contribuyentes y ha brindado resultados como en Jamaica en donde la unidad de inteligencia de minería de Datos se enfocó en grupos comunes en base a un análisis de riesgo, identificando a más de 5000 nuevos contribuyentes, entre personas que han dejado de declarar o que reportaban menos ingresos.

En el Ecuador se posee una Administración Tributaria alineada al Plan Nacional del Buen Vivir, formada por un modelo integral de gestión estructural de riesgo que permite un trabajo organizado por componentes donde cada una de éstos son los pilares que permiten un funcionamiento integral enfocados en la visión y misión empresarial.

La estructura del SRI permite consolidar dos modelos gestión por procesos y riesgo tributario, en el primero se identifica y gestiona procesos interrelacionados, la mejora continua está presente en los resultados, y la erradicación de errores es fundamental, en cambio con el riesgo tributario en un contexto operativo se identifica, analiza, determina, prioriza, trata y evalúa los riesgos de incumplimiento de las obligaciones tributarias por parte de los contribuyentes. (Servicio de Rentas Internas, 2012)

La gestión de riesgos se enfoca en las acciones efectivas con estrategias adecuadas a las características y comportamientos de los contribuyentes, bajo éste lineamiento se busca recopilar factores relativos a los mismos centrándose específicamente en Sociedades donde se determine la priorización o grado de

importancia dentro de las funciones que realiza la Gestión tributaria, que en el contexto de la presente tesis se lo va a tratar como **Importancia Fiscal**.

La **Importancia Fiscal** juega un papel determinante en la ejecución de estudios de mejora, cuyo fin es hacer cumplir los objetivos de la gestión de tributos, actualmente ésta variable queda a juicio del analista tributario. (Impacto de Notificaciones sobre la Omisidad, 2011)

En consecuencia, proponer un criterio de **Importancia Fiscal** en base a parámetros de negocio y explotación de la información que permita tomar decisiones con información certera ayudará a generar valor al negocio y evitará un proceso empírico en su consecución.

1.2. Justificación de la investigación

Actualmente las soluciones tecnológicas alineadas al negocio han tenido un crecimiento trascendental y la parte de tributación no ha sido la excepción.

Es oportuno considerar que “El avance fundamental ha sido en el uso de la informática para sistematizar los procedimientos internos de las administraciones tributarias: registro de contribuyentes, procesamiento de declaraciones y pagos, control de la recaudación, apoyos al área de fiscalización y cobranza coactiva, entre otros.” (Baer, 2011, pág. 15)

El CIAT, Centro Interamericano de Administraciones Tributarias, ha citado la importancia de la tecnología en la actualidad, en la resolución de la 43^a Asamblea General del CIAT, donde recomienda “el desarrollo de tecnologías de la información y las comunicaciones al servicio de las administraciones tributarias”.

Alineados a ésta premisa y tomando en cuenta que se está viviendo en un mundo que ha sufrido una transformación por la invasión de la globalización, la

nueva economía, los medios de comunicación, la era post industrial de los negocios, los mismos que dan origen a una nueva forma de ver al mercado donde lo tradicional ya no está en auge; la tecnología, creatividad e innovación están ganando terreno a pasos agigantados, los mismos que conducen a razonar en nuevas estrategias focalizadas más en las necesidades de los clientes.

La era de la nueva economía ha traído consigo una gran cantidad de datos almacenados consecuencia de la fusión de esa red de redes de alcance mundial.

El Ecuador como parte de ese mundo globalizado y específicamente el SRI, es una institución que cuenta con una gran cantidad de datos, 14 terabytes y solo se conoce que se usan 4 de ellos, la información que se puede obtener de aquellos datos no procesados ayudaría en la obtención de información más precisa basada en indicadores y no en criterios empíricos que pueden ocasionar un mal uso de tiempo y recursos, por un mal direccionamiento del enfoque del verdadero inconveniente.

A pesar que en el SRI, existe un área específica de BI, Business Intelligence, que se encarga de proyectos que buscan optimizar con criterios fundamentados los procesos inherentes en la recaudación, existen parámetros que aún se esquematizan de acuerdo a criterios propios de cada analista, como es el caso de la **Importancia Fiscal**, medida que es utilizada en diferentes áreas, para realizar análisis y cumplir así con los objetivos de gestión tributaria.

En la actualidad en el SRI, no existe un concepto de **Importancia Fiscal** consensuado en criterios de negocio y explotación de la información, es decir dependen del criterio aplicado por el analista de turno que se encuentra realizando el estudio.

Por éste motivo es indispensable la construcción de un concepto de **Importancia Fiscal** donde el origen de análisis sea toda la información que genera el contribuyente y con ésta información se pueda estructurar un conocimiento acertado que canalice criterios que contribuyan a la toma de decisiones alineadas al objetivo gubernamental que posee la recaudación que es el financiamiento de educación, salud y seguridad social.

Las herramientas propuestas para llegar a parametrizar ese concepto de **Importancia Fiscal** son aquellas relacionadas a la excavación, manejo y administración de información, algo que se encuentra en auge en ésta temática es la minería de datos, la misma que ha permitido a nivel mundial crear conocimiento valioso dirigido a esa cantidad de información generada día a día por ese crecimiento exponencial que la tecnología ha presentado.

La minería de datos ha presentado grandes beneficios en diferentes temáticas donde ha permitido estudiar patrones comunes, se puede resaltar el caso de la campaña electoral realizada por el actual Presidente de los EEUU, donde gracias a la información que compró sobre los ciudadanos le permitió dirigir la misma, a la población indecisa. (Franganillo, 2009)

Esa característica de la minería de datos de definir patrones ocultos en comportamiento y/o decisiones, y menos en nombres, direcciones, colores, etc. ayudará a segmentar con mayor fiabilidad aquellos grupos de contribuyentes con relación a su **Importancia Fiscal**.

A partir de la segmentación propuesta, el organismo fiscalizador mejorará su labor, haciendo más certero el proceso de validación de la información que declara el contribuyente.

1.3. Objetivos de la investigación

1.3.1 Objetivo general

Identificar grupos de contribuyentes con Importancia Fiscal en sociedades privadas activas

1.3.2 Objetivos específicos

- Construir en base a una metodología un concepto de Importancia Fiscal.
- Valorar, depurar, explotar las variables y criterios de análisis tomando como base el Modelo Analítico más adecuado.
- Determinar y segmentar los grupos de interés a partir de técnicas de minería de datos.
- Viabilizar una herramienta que permita parametrizar las diferentes variables que son parte del concepto de Importancia Fiscal.

CAPITULO II: MARCO DE REFERENCIA

2.1 Marco Teórico

2.1.1 La Administración tributaria y la Importancia Fiscal

La Administración Tributaria se define como una entidad de servicio al cliente que tiene como actividad la aplicación y cobranza de impuestos que están obligados a realizar por ley. (Alink & Van Kommer, 2000)

A nivel mundial posee un objetivo general que es la recaudación de tributos, de acuerdo al país, el ente recaudador varía y también el tipo de tributo que administra.

En el Ecuador, el Servicio de Rentas Internas (SRI), es el encargado de la recaudación de tributos internos, exceptuando los de carácter aduanero como los derechos, IVA, ICE sobre importaciones, cuyo organismo recaudador es la Corporación Aduanera Ecuatoriana (CAE).

La gestión de la Administración Tributaria en el SRI se resume en una estrategia basada en procesos cuyo modelo de gestión se ha esquematizado por medio de componentes, los cuales se encuentran alineados a la visión y misión institucional.

La Figura 1 que se cita a continuación, se indica en forma gráfica el modelo de gestión esquematizado por componentes.

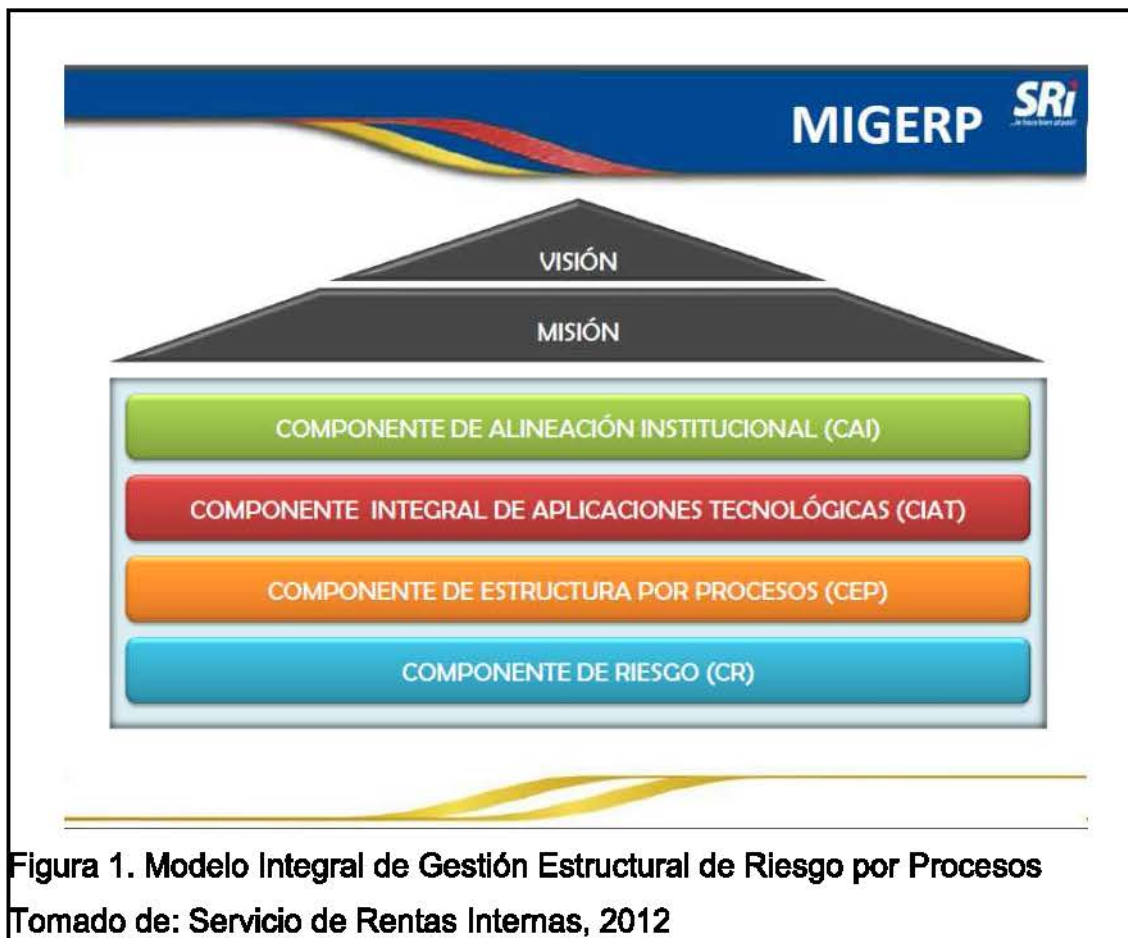


Figura 1. Modelo Integral de Gestión Estructural de Riesgo por Procesos
Tomado de: Servicio de Rentas Internas, 2012

El componente de Alineación Institucional (CAI): Como su nombre lo indica es el que se basa netamente en el cumplimiento de los objetivos y hacia donde se encuentra alineada en el futuro. Se encuentra estructurada por tres niveles: estratégico, operativo y táctico.

- **Estratégico:** La parte estratégica se encuentra definida por los objetivos estratégicos del SRI, los cuales son 6:
 - Incrementar el uso eficiente del presupuesto.
 - Incrementar el desarrollo del talento humano.
 - Incrementar la eficiencia operacional.
 - Incrementar la conciencia de la ciudadanía acerca de sus deberes y derechos fiscales.
 - Incrementar la aplicación de los principios constitucionales

- en materia tributaria.
- Incrementar la eficiencia y efectividad en los procesos de asistencia y control enfocados al cumplimiento tributario, sustentados en un modelo de gestión de riesgos.
- **Operativo:** Optimiza las tareas y actividades que se alinean en el cumplimiento de las metas de la Institución.
- **Táctico:** Esquematiza que se debe hacer y qué objetivos se desean alcanzar alineados a condiciones racionales basándose en una realidad actual y futura.

Componente Integral de Aplicaciones Tecnológicas (CIAT): Se encuentra formada de la estructura informática y aplicaciones integradas, es la que comprende la situación actual orientada a respuestas de los procesos de la cadena de valor. Como algunas de las principales características podemos citar: Flexibilidad, Alta disponibilidad, Velocidad.

Componente de Estructura por Procesos (CEP): Es el que por medio de la estructura por procesos fusiona dos modelos riesgo tributario y de gestión por procesos, el mismo que le permite identificar y gestionar procesos enfocados a una mejora continua de los resultados con la minimización de errores.

Cuando se refiere a éste componente se hace alusión que el SRI posee procesos gobernantes, de cadena de valor y Habilitantes.

- **Procesos Gobernantes:** Determina las directrices de la Institución, analizando las necesidades y expectativas del ciudadano.
- **Procesos habilitantes de asesoría:** Se refiere a lo legislativo y normativo, complementando de esa forma los procesos de la cadena de valor.
- **Procesos habilitantes de Apoyo:** Son los que proveen de recursos

a las actividades que posee el SRI.

- **Cadena de Valor:** Es la operatividad y ejecución del negocio, se ejemplariza la manera cómo la organización provee valor a sus clientes e interesados.

Componente de Riesgo (CR): Se focaliza en dos perspectivas Administración de Riesgos y la Operación de la Gestión de Riesgo.

Administración de Riesgos: Se encarga netamente de las amenazas, las cuales son identificadas y evaluadas con relación a lo que pueda causar el posible no cumplimiento de las obligaciones de los contribuyentes.

Operación de la Gestión de Riesgo: Proceso mediante el cual se identifica, analiza, determina, prioriza, trata y evalúa riesgos que incurren en el incumplimiento de las obligaciones de los contribuyentes. (Servicio de Rentas Internas, 2012)

El Componente de Riesgo en general posee una estructura alineada en Identificación y Clasificación, Reducción, Detección, Selección, Cobertura y Evaluación.

La Figura 2 que se cita a continuación, se indica las fases que presenta éste componente.



Figura 2. Fases de la Gestión de Riesgos

Tomado de: Servicio de Rentas Internas, 2012

Identificación y Clasificación: La investigación es parte fundamental de esta fase, la cual permite obtener información y realizar actividades de control que identifique los hechos y eventos relacionados con los riesgos. Adicionalmente la clasificación junto con la priorización es el siguiente paso, donde el objetivo es buscar una ponderación.

Reducción: Es tratar de prevenir o mitigar la ocurrencia del riesgo, para eso se utiliza una detección antes del momento o en el momento inmediatamente posterior.

Detección: Es establecer si se ha producido o producirá el riesgo.

Se clasifica en:

- **Detección Proactiva:** Se realiza antes que suceda el evento relevante.
- **Detección Corriente:** Se realiza en el momento inmediatamente posterior que suceda el evento relevante.
- **Detección Reactiva:** Se realiza luego de la declaración de impuestos.
- **Detección Activa:** Se presenta cuando se realiza visitas al

contribuyente.

- **Detección Pasiva:** Cruces de información del contribuyente y terceros.

Selección: Es la realización de acciones de control sobre un número limitado de contribuyentes, en ésta etapa se asegura la eficacia del sistema de control fiscal, ya que se optimiza la utilización de los recursos que han sido asignados.

Cobertura: Es la estrategia que en función a la capacidad disponible, el nivel de capacitación y el desarrollo del personal determina el grado de atención o de intensidad que se va a aplicar a cada riesgo,

Evaluación: Ésta fase es la que recalca la importancia de la etapa de Selección, es la que provee información respecto a las estrategias, políticas y planes de auditoría, causando un impacto directo al modelo.

Centrándose en el Componente de Riesgos (CR) y su concatenación con la Administración Tributaria, la lucha contra la evasión fue una política de gobierno que se asumió desde el primer día, en la cuál, se hizo y se hace indispensable la focalización de estrategias que ayuden a incrementar la recaudación, la Ciudadanía Fiscal, la obtención de Información nueva o actualizada de contribuyentes y disminuya la elusión y evasión fiscal, garantizando la equidad tributaria

Alineados a lo anteriormente citado se puede hacer referencia al concepto de **Importancia Fiscal**, el cuál a nivel mundial no posee un significado específico, ya que el mismo depende de lo que se considere como importante en el País con respecto a su realidad Tributaria. En el SRI no es la excepción para la elaboración del concepto se ha esquematizado un criterio estandarizado de acuerdo a diferentes reuniones realizadas con los expertos tributarios, cuyo resultado se alineó al alcance de la presente Tesis, el cuál se lo va a

determinar inicialmente como la priorización o grado de valor dentro de las funciones que realiza la Gestión tributaria para fines de control tributario alineadas a las características y comportamientos de los contribuyentes, específicamente en Sociedades.

Bajo éste contexto la **Importancia Fiscal** juega un papel determinante en la ejecución de cada uno de los procedimientos necesarios para cumplir la gestión de tributos, actualmente ésta variable queda a juicio del analista tributario, en consecuencia proponer un criterio cuantificado de **Importancia Fiscal** en base a parámetros de negocio y explotación de la información con ayuda de herramientas tecnológicas, apoya sustancialmente a la estructuración de reportes más precisos que apoyarán cada uno de los lineamientos que permiten que el SRI no se desvíe de los objetivos planteados institucionalmente.

2.1.2 Descubrimiento de Conocimiento en Base de Datos

La necesidad de encontrar conocimiento en datos, de acuerdo a, Yves-Michel Marti en sus estudios de economía inteligente, relata que se remonta a fines del siglo XVI, donde la Reina Elizabeth I, determinó que la base de la fuerza inglesa era la información y comercio, solicitando al filósofo Francis Bacon que cree un sistema dinámico de información, y de acuerdo al autor el mismo fue ampliamente utilizado por los Ingleses. (Bryant, Krzanich, & James, 2013)

Bajo éste contexto y con el paso del tiempo la gran necesidad de almacenar grandes cantidades de datos, y obtener información de las mismas surge en los años 70 donde se cita un concepto más estructurado por Edgar Frank Codd, quien define el modelo relacional con sus respectivas reglas para la evaluación de sistemas de bases de datos relacionales.

Actualmente, gracias a un mundo globalizado, los datos que se producen día a día han crecido a pasos agigantados, a nivel mundial se genera información

consciente e inconscientemente; consciente, cuando se utiliza la tarjeta de crédito e inconsciente cuando se conduce un auto y a través de dispositivos que toman información de la velocidad, distancia, tiempo, entre otras da lugar a una aglomeración de datos que si multiplicamos por cada uno de las personas a nivel mundial y por cada una de las actividades adicionales que son captadas y almacenadas se puede citar como un hecho que la cantidad de información crece exponencialmente y que el tratamiento que se le da a esos datos que son recolectados deben ser automatizados para que se direccionen a generar conocimiento de una manera más rápida y eficaz.

Bajo éste contexto, el descubrimiento de conocimiento en la base de datos surge con la necesidad de poseer acceso a datos más precisos en una amplia cantidad de información, el cual ha tomado cada día más fuerza coexistiendo procesos y metodologías para su construcción; procesos, que citan el qué hacer, y metodologías que se direccionan más al cómo hacerlo.

Es así que a inicios del año 1996, Knowledge Discovery in Databases KDD fue un modelo aceptado por la comunidad científica que estableció las etapas principales de explotación de información como: Selección, pre-procesamiento, transformación, minería de datos y evaluación. A partir del año 2000 se dio a conocer nuevas formas de hacer frente al tratamiento de datos con enfoque sistemático como: SEMMA, acrónimo de Sample, Explore, Modify, Model and Assess, desarrollado por el Instituto SAS; CRISP-DM, acrónimo de Cross-Industry Estándar Process for Data Mining, desarrollada por un grupo de empresas Europeas; CATALYST, desarrollada por Dorian Pyle. (Moine, 2013)

La Tabla 1 que se cita a continuación, se muestra claramente sus similitudes para entablar cada una de las fases.

Tabla 1. Fases utilizadas por cada una de las Metodologías

Fases	KDD	CRISP – DM	SEMMA	CATALYST
Análisis y comprensión del negocio	Comprensión del dominio de aplicación	Comprensión del negocio		Modelado del negocio
Selección y preparación de los datos	Crear el conjunto de datos	Entendimiento de los datos	Muestreo Comprensión	Preparación de los datos
	Limpieza y pre-procesamiento de los datos	Preparación de los datos	Modificación	
	Reducción y proyección de los datos			
Modelado	Determinar la tarea de minería	Modelado	Modelado	Selección de herramientas y modelado inicial
	Determinar el algoritmo de minería			
	Minería de datos			
Evaluación	Interpretación	Evaluación	Valoración	Refinamiento del modelo
Implementación	Utilización del nuevo conocimiento	Despliegue		Comunicación

Tomado de: Moine, 2013

La diferencia principal se cita en cada una de las tareas y actividades que conforman cada fase. Mientras que CATALIST y CRISP-DM se asemeja más a una metodología por el detalle que toma en cada una de las mismas; KDD y SEMMA, se considera un proceso por la generalidad al explicarlas. (Moine, 2013)

Se citará cada una de las tareas y actividades con mayor detalle, solo de KDD y CRISP-DM, la primera por ser el modelo inicial que fue aceptado para el desarrollo de proyectos de búsqueda de conocimiento en base de datos, y la segunda por ser una de las metodologías más difundidas, presenta independencia de la herramienta a utilizar, detalla cada una de sus fases haciendo entendible el desarrollo, además se comenta que existen alrededor de 200 miembros activos que la utilizan, por otra parte se señala que no ha sido construida de manera teórica, más bien se basa en casos reales de proyectos de minería de datos. (LLombart, 2010)

La Figura 3 que se cita a continuación se puede observar que CRISP – DM es una de las metodologías más utilizadas a nivel mundial de acuerdo a una encuesta realizada por KDnuggets, en los últimos años.



KDD:

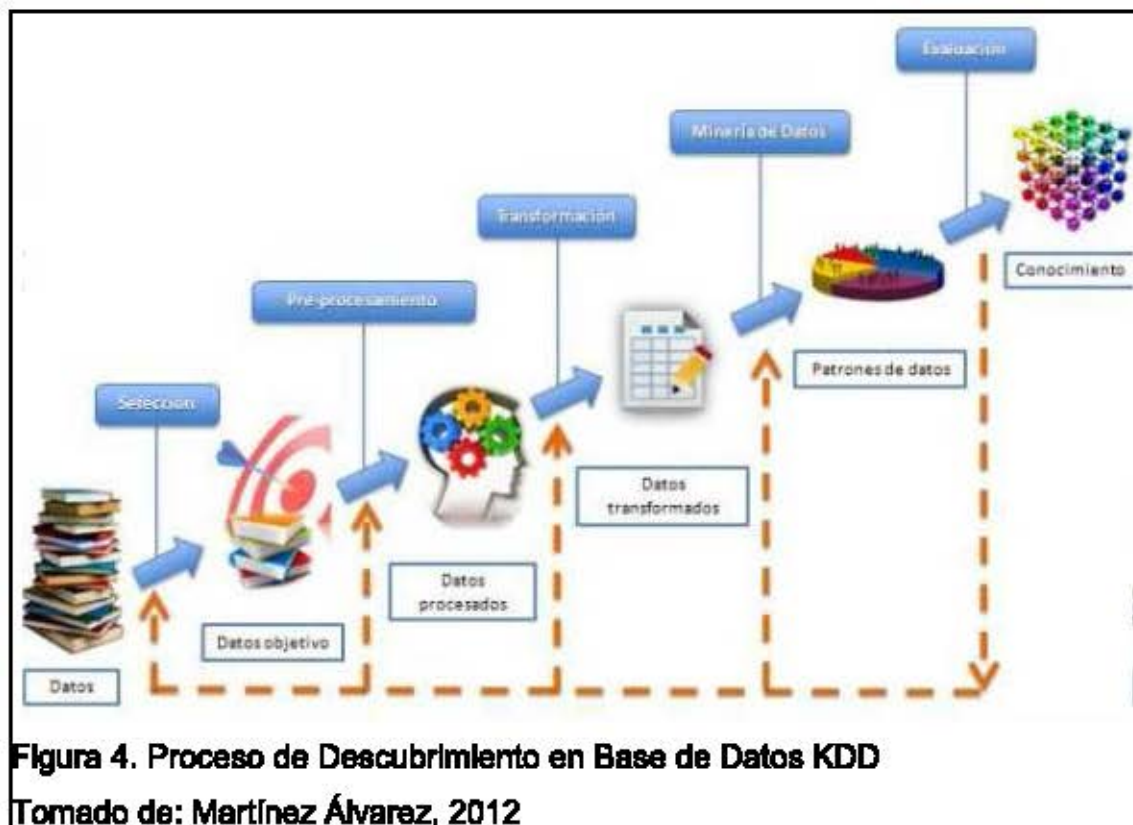
KDD, acrónimo de Knowledge Discovery In Databases, constituye el primer modelo en ser utilizado a Inicios del año 1996, se lo define como: "Proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y en última instancia entendibles en los datos." (Fayyad, 1996)

Cabe recalcar que algunos autores lo consideran como un proceso pero otros lo consideran una metodología y la utilizan como si lo fuera.

En el desarrollo de ésta tesis se lo va a considerar como un proceso, por la descripción general de las actividades sin entrar a detalles del cómo realizar cada una de ellas.

El proceso KDD está conformado por fases de forma iterativa e Interactiva: Iterativa ya que permite regresar a una etapa anterior para cualquier ajuste, e interactiva porque el experto debe estar presente para la validación de datos,

La figura 4 que se cita a continuación se puede observar las fases del proceso KDD.



Identificación del problema

- Establecer el objetivo claro para el problema a resolver
- Entender las metas del proceso y cuáles son las preguntas que se quieren responder.
- Analizar si un objetivo es inútil

Pre-procesamiento o Selección de los Datos

- Verificar la ubicación de los datos.
- Confirmar qué tipo de datos se encuentran estructurados o no.
- Verificar disponibilidad y significado de los datos.
- Analizar si los datos están alineados a los objetivos de negocio.

- Valorar y esquematizar si es posible realizar muestras al azar para reducir el volumen.
- Clasificar las variables y registros apropiados como datos de entrada para el proceso de minería.

Transformación y almacenamiento de los datos

- Reducir o agrupar los datos en las características de interés.
- Consolidar la información y escoger una arquitectura acorde a las necesidades del problema

Selección y aplicación de algoritmos de Minería de Datos

- Aplicar la técnica de minería de datos: Descriptiva o predictiva.
- Generar modelos de minería de datos y descubrir patrones

Interpretación y evaluación de los patrones encontrados

- Analizar de los patrones resultantes desde diferentes perspectivas.
- Documentar los procedimientos que generen propuestas de valor al negocio.

CRISP – DM:

CRISP-DM, acrónimo de Cross-Industry Estándar Process for Data Mining fue propuesta en el año de 1999 por un grupo importante de empresas europeas, entre ellas, Daimler Chrysler AG de Alemania, NCR Systems Engineering Copenhague de Dinamarca, SPSS de Inglaterra y OHRA de Holanda. (Valencia, 2006)

CRISP-DM se caracteriza por el detalle en cada uno de los procedimientos, en el mismo se puede encontrar dos documentos el modelo de referencia y la guía

del usuario. El primero describe las fases y tareas generales, por otro lado la guía del usuario como su nombre lo indica es la aplicación práctica del modelo de referencia a proyectos de minería de datos específicos. (Valencia, 2006)

CRISP – DM está conformado por fases de manera cíclica e iterativa y se encuentra estructurado de lo general a lo específico, cada fase posee sus tareas Generales, el mismo que posee sus tareas específicas y cada una de ellas posee instancias de proceso.

En la figura 5 se ilustra las fases de CRISP - DM

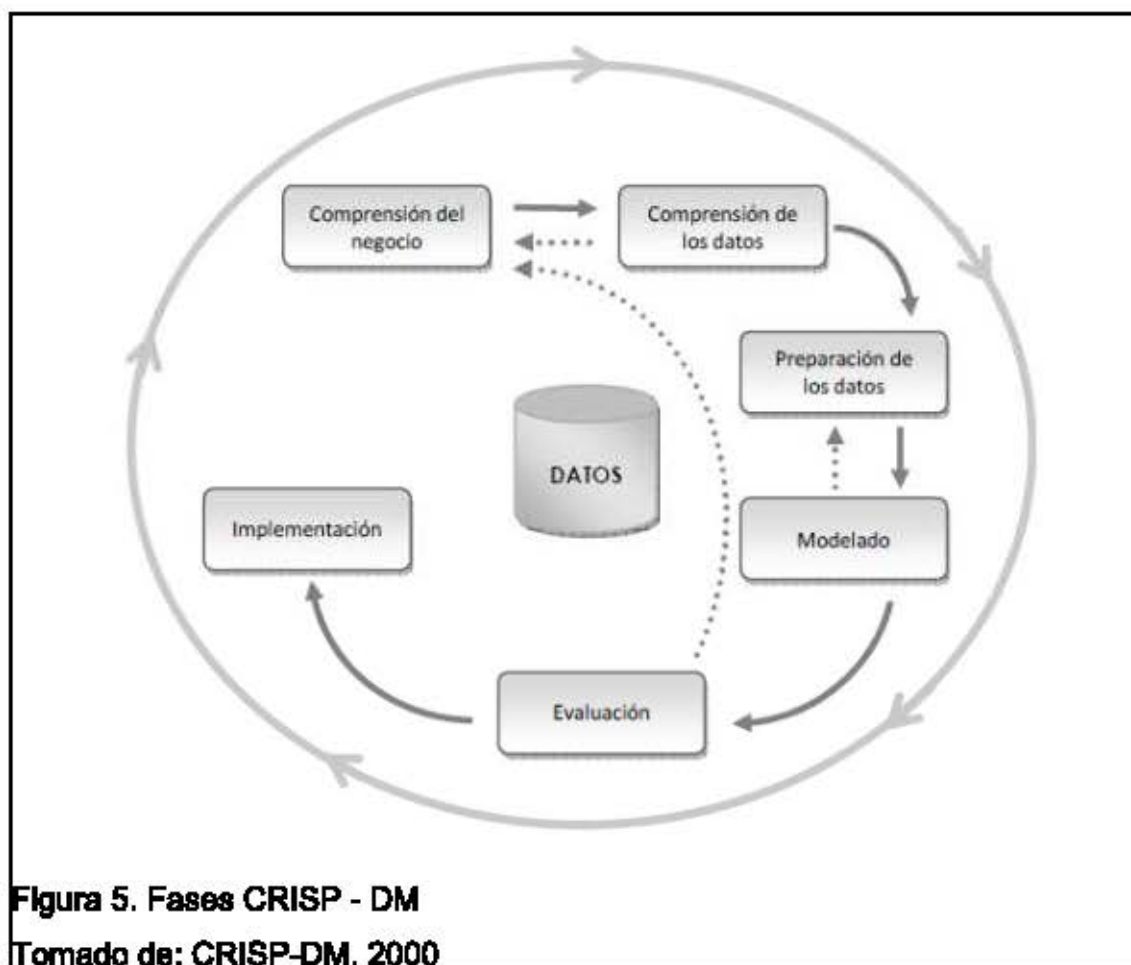


Figura 5. Fases CRISP - DM
Tomado de: CRISP-DM, 2000

Fase de comprensión del negocio o problema

Se considera la más importante ya que con la misma se garantiza el alineamiento del proyecto con los objetivos del negocio. Es visualizar el mismo desde una perspectiva empresarial, incrementando la fiabilidad del procedimiento a aplicar y garantizando cumplir con todas las perspectivas del proyecto.

Ésta fase posee las siguientes tareas con sus respectivas actividades:

Determinar los objetivos del negocio.

- Determinar problema a resolver
- Validar por qué es necesario utilizar Minería de Datos
- Determinar los objetivos del negocio
- Determinar los criterios de éxito

Evaluación de la situación.

- Analizar el estado actual del negocio antes de la Minería de Datos
- Definir requisitos del problema; negocio y minería
- Analizar riesgos y contingencias
- Determinar la terminología
- Analizar Coste – Beneficio de Minería de Datos

Determinación de los objetivos de DM.

- Determinar los objetivos de la Minería de Datos
- Determinar los criterios de éxito de la Minería de Datos

Realizar el plan proyecto

- Desarrollar un plan de proyecto
- Describir y evaluar las herramientas y técnicas en el plan proyecto

Fase de comprensión de los datos

Es el primer contacto con el problema, aquí se parte con la primera recolección de datos identificando su calidad y estableciendo relaciones evidentes; las consultas, modificaciones a la base de datos son innumerables es por eso que se recomienda poseer una base de datos ad-Hoc.

Ésta fase posee las siguientes tareas con sus respectivas actividades:

Recolección de datos iniciales.

- Recolectar datos iniciales y adecuar para el futuro procesamiento
- Realizar un reporte de la obtención de los datos iniciales, localización, técnicas, problemas y soluciones inherentes

Descripción de los datos

Describir los datos iniciales, identificación, significado de cada campo y descripción del formato inicial

Exploración de datos.

- Encontrar una estructura general de datos, aplicación de pruebas estadísticas básicas.
- Realizar un reporte de la Exploración de Datos, que posea tablas de frecuencia y gráficos de distribución

Verificación de la calidad de los datos.

- Determinar la consistencia de los valores de los campos, cantidad y distribución de valores nulos.
- Encontrar valores fuera de rango para evitar ruido en el proceso.
- Realizar un reporte de la calidad de los datos

Fase de preparación de los datos

Es la adaptación de los datos a la técnica de minería de datos, preparación de los mismos en la técnica de modelado, limpieza, generación de variables adicionales, integración de diferentes orígenes, se interrelaciona cercanamente con la fase de modelado.

Ésta fase posee las siguientes tareas con sus respectivas actividades:

Selección de datos.

Seleccionar un subconjunto de datos con su debida justificación de la inclusión / Exclusión

Limpieza de los datos.

- Preparar los datos para la fase de modelación utilizando técnicas para optimizar la calidad de los datos, normalización de datos, soluciones para valores ausentes, etc.
- Realizar un reporte de Limpieza de Datos

Estructuración de los datos.

- Preparación de los datos generando atributos derivados
- Integrar registros generados con atributos existentes

Integración de los datos.

Crear nuevas estructuras, datos combinados a partir de datos seleccionados, tablas resumen.

Formateo de los datos.

Transformar los datos sin alterar su significado reorganizarlos acorde a las limitaciones de la herramienta

Fase de modelado

En esta fase se selecciona las técnicas a ejecutar y los modelos más óptimos que se alinean al negocio.

Ésta fase posee las siguientes tareas con sus respectivas actividades:

Selección de la técnica de modelado.

- Escoger la técnica de minería de datos, la misma va a depender del tipo de problema a resolver si es de clasificación, predicción, segmentación, etc.
- Realizar el Modelamiento

Generación del plan de prueba.

Diseñar y Probar la calidad y validez del modelo utilizado

Construcción del Modelo.

- Ejecutar la técnica en los datos previamente mencionados generando uno o más modelos, configuración de los parámetros del modelo que determinen sus características.

- Describir el Modelo, con su respectiva interpretación del resultado obtenido.

Evaluación del modelo.

- Interpretar el modelo correlacionándolo con los criterios de éxito acordados previamente.
- Juzgar el modelo bajo criterios tanto de expertos en Minería de datos como expertos en el negocio con referencia a los objetivos planteados inicialmente.

Fase de evaluación

Es el cumplimiento de los criterios de éxito de los datos donde se realiza el respectivo análisis, se debe evaluar minuciosamente los resultados para evitar cometer errores y en el caso que se presenten se debe revisar el proceso desde la fase Compresión del negocio parametrizando excepciones o ausencia de datos y su respectivo tratamiento.

Ésta fase posee las siguientes tareas con sus respectivas actividades:

Evaluación de los resultados.

- Evaluar el modelo en relación con los objetivos del negocio.
- Evaluar el modelo con relación a otros objetivos que difieran de los originales con la finalidad de revelar información adicional

Proceso de revisión.

- Revisar el Proceso
- Calificar el proceso en busca de mejoras

Determinación de futuras fases.

- Lista de Posibles Acciones en función a los resultados obtenidos.
- Decidir en función a los resultados partir desde cero o repetir otra iteración desde la fase de la preparación

Fase de implementación

Es la ejecución de los resultados obtenidos, se puede recomendar al negocio cambios de acuerdo a las observaciones de los modelos.

Ésta fase posee las siguientes tareas con sus respectivas actividades:

- **Plan de implementación.**

Desplegar los resultados y realizar una estrategia para la implementación, donde incluye la respectiva documentación.

- **Monitorización y Mantenimiento.**

Realizar un Plan de monitoreo y mantenimiento

- **Informe Final.**

Realizar un reporte final, donde cite el resumen del proyecto o una presentación que ilustre los resultados obtenidos.

- **Revisión del proyecto:**

Documentar las experiencias tanto positivas como negativas.

En sí todo el proceso de descubrimiento de conocimiento en la base de datos con su respectiva metodología sea propia o guiada por una marco referencial ha permitido que sea aplicado en diferentes áreas como: Detección de Fraudes, Mercadeo, Manufactura y producción, Internet, Administración de Redes, Evasión Fiscal entre otras, a cada una de ellas ha permitido entablar patrones que se alinean con el objetivo del negocio y genere valor al mismo.

2.1.3 Minería de Datos

Los procesos y metodologías que se han creado con la finalidad de apoyar a las decisiones de alta gerencia respaldadas por información certera, se hace posible gracias a la minería de datos, la cual es una fase denominada modelamiento en ese conjunto de pasos necesarios para crear conocimiento, cuyo objetivo es encontrar aquellos patrones y relaciones no evidentes.

Es muy usual que muchos autores conceptúen de la misma forma al proceso KDD y la Minería de datos, exceptuando que ésta última es una fase que compone todo el descubrimiento de conocimiento de la base de datos, esto se presenta porque tradicionalmente KDD fue el primer modelo aceptado por el mundo científico como se mencionó anteriormente, otro punto adicional para que suceda esto es porque la minería de datos no alcanzaría su objetivo de crear nuevo conocimiento si no posee unos datos depurados, parametrizados y de calidad.

Bajo éste contexto, mezclar los dos conceptos quedaría en un criterio tradicional, actualmente se puede decir que la minería de datos se basa en algoritmos que buscan características comunes en una gran cantidad de datos encontrando patrones desde diferentes perspectivas y abstrayendo información que genere valor al negocio.

Se debe tener claro que la minería de datos es un apoyo a los analistas y no elimina la necesidad de validación, interpretación y análisis de los patrones

resultantes, tampoco elimina la necesidad del estudio de factibilidad para su implementación en el mundo real. (Ríos Perez, 2012)

En sí la Minería de Datos es conceptualizada como:

“El proceso de descubrir correlaciones significativas, patrones y tendencias socavando a través de grandes cantidades de datos almacenados en repositorios. Minería de Datos emplea tecnologías de reconocimiento de patrones, así como las técnicas matemáticas y estadísticas” (Gartner, 2013)

2.1.3.1 Importancia de la Minería de Datos

El crecimiento de la información ha sido contundente, terabytes ya no es suficiente para su almacenamiento, ahora es necesario medidas con mayores características como: petabytes, exabytes y zetabytes, toda ésta inundación de información es favorable si se direcciona a un buen tratamiento caso contrario el análisis se vuelve incontrolable, complejo y la generación de conocimiento disminuye.

La evolución de las tecnologías ha ido direccionando a las empresas a ser más exigentes en los requerimientos que necesitan, el conocimiento está en auge y la competitividad de conocer lo futuro es implacable, el fin es buscar y tomar más decisiones proactivas que reactivas, evitando lo empírico en la mayoría de los casos.

La Tabla 2 que se cita a continuación, se puede visualizar claramente cómo las necesidades del negocio han ido cambiando.

Tabla 2. Evolución de las tecnologías de base de datos

Hito histórico	Pregunta de Negocio	Tecnología que lo posibilita	Suministrador	Característica principal
Data Collection (1960s)	¿Cuáles fueron mis ingresos en los últimos 5 años?	Ordenadores, cintas, discos, DBMS jerárquicos (IMS) y en red	IBM, CDC	Datos históricos
Data Access (1980s)	¿Cuántas unidades vendí el mes pasado en España?	Bases de datos relacionales (RDBMS, SQL, ODBC)	Oracle, Sybase, Informix, IBM, Microsoft	Datos dinámicos a nivel de registro (histórico)
Data Warehousing & Decision Support (1990s)	¿Cuántas unidades vendí el mes pasado en España en relación con Europa?	On-line analytic processing (OLAP), gestores multidimensionales	Cognos, Business Objects, Microstrategy, NCR, SPSS, Comshare, etc.	Datos dinámicos en múltiples niveles o jerarquías (histórico)
Data Mining (2000s)	¿Cuáles serán las ventas del próximo mes en Europa?	algoritmos avanzados (data stream, weblog, bio-data...), RDBMS	SPSS/Clementine, Lockheed, IBM, SGI, SAS, NCR, Oracle, etc.	Datos de prospección (análisis de mercado, de riesgos, ...)

Tomado de: Zorrilla, 2010

Las preguntas de negocio que antes tomaban mucho tiempo ahora con la Minería de datos pueden ser resueltas de manera eficiente y eficaz.

Es aquí donde se puede citar la gran importancia que conlleva saber esquematizar un buen procedimiento para sistematizar los datos donde nos puedan generar información y ésta a su vez conocimiento que finalmente conducirá a la acción.

Con relación a ésta premisa juega un papel súper importante el descubrimiento de conocimiento, donde con datos recopilados, seleccionados y depurados, se puede aplicar la técnica de minería de datos la misma que va a direccionar a patrones contextualizados, los cuales luego de ser evaluados e interpretados darán como resultado el conocimiento que al ser difundido es el soporte para que se tome nuevas decisiones o rectificar las antiguas.

Por tanto, en función a la misma podemos citar que la minería de datos es la fase del descubrimiento de conocimiento que permite modelar la mejor solución para extraer patrones que se encuentran inicialmente ocultos, los cuáles puedan estar al alcance del personal que lo necesite en el momento adecuado

para que cumpla con la actividad que fue asignado.

Concluyendo se puede citar que, la toma de decisiones inteligentes tiene su origen en información más robusta, escalable, portable, significativa y precisa basada en umbrales predefinidos por el negocio, donde de acuerdo a su criticidad se vuelve indispensable y no solo necesaria tenerla a la mano lo más rápido posible.

2.1.3.2 Técnicas de Minería de Datos

Actualmente el camino a seguir y el cómo se realice la búsqueda de conocimiento que genere resultados, los cuáles sean un apoyo para una mejor evaluación, predicción y selección en la toma de las decisiones del negocio, ha adquirido una importancia crucial, pues de éste depende que se cumpla con el objetivo que inicialmente se propuso en el proyecto de Minería de Datos.

La técnica de Minería de Datos se basa en algoritmos que nos permiten recuperar la información que no se encuentra a la vista y que es estratégicamente útil, la elección de ésta depende del estudio que vamos a realizar y de los datos que se posea, no existe un mejor o peor método, todo depende del objetivo que queremos llegar y cuál se adapta mejor a nuestras necesidades.

Bajo éste contexto se la puede conceptuar como: Un filtro donde al introducir los datos se produce un conjunto de patrones relevantes útiles para dar inicio al nuevo conocimiento. (Molero Castillo, 2008)

En la figura 6 que se cita a continuación se puede observar las entradas y salidas para que sea posible realizar la respectiva técnica de minería de datos.

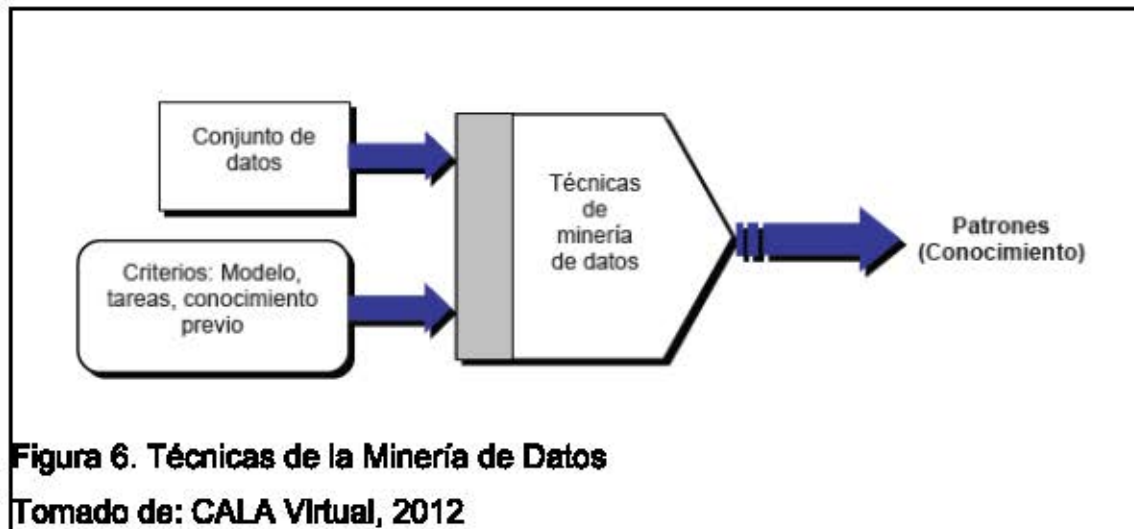


Figura 6. Técnicas de la Minería de Datos
 Tomado de: CALA Virtual, 2012

Este filtro recibe un conjunto de datos por medio de modelos o un conocimiento previo.

Los modelos de aprendizaje que se pueden presentar son: Predictivos o Supervisados y Descriptivos o No Supervisados.

Predictivos o Supervisados: Se hace referencia a predecir, es decir responde a preguntas de datos futuros basados en su comportamiento del pasado, posee una variable desconocida, denominada objetivo o variable dependiente y la función de éste modelo es determinarla en base a un conjunto de datos previamente entrenados o variables independientes. (CALA Virtual, 2012)

Los datos son entrenados con otro conjunto cuya variable objetivo o independiente es conocida y comparable, es por éste motivo que también adquieren el nombre de supervisados, pues su procedimiento se debe ajustar y verificar con relación al resultado conocido que se tiene como objetivo. (Martínez Álvarez, 2012)

Los modelos predictivos o supervisados pueden ser : Clasificación, Regresión y Priorización o Pronóstico.

Clasificación: Es aquel que ordena las variables de entrada en categorías verificando si el valor resultante coincide con el real.

En éste tipo también se puede hacer énfasis en la determinación del grado de certeza.

Las variables con las que se trabaja son discretas. Por ejemplo: Predecir si una persona adquirirá o no el producto que estamos ofertando en una campaña de envío de correo directo.

Regresión: Es encontrar el valor numérico de la variable objetivo, basados en las variables de entrada. Si solo se posee una variable se trata de regresión simple, mientras que si se dispone de algunas variables es regresión múltiple.

Las variables con las que se trabaja son continuas. Por ejemplo: Predicción del tiempo basado en la temperatura, presión del aire y humedad.

Preferencia o Priorización, Pronóstico: Es el que en base a algoritmos asigna técnicamente probabilidades para estimar valores futuros.

Por ejemplo: Proyectar la producción de acuerdo a demanda futura de productos.

Descriptivos o No Supervisados: Es el que identifica patrones que sintetizan el conjunto de datos ingresados, adquiriendo conocimientos de las relaciones que existen entre ellos sin necesidad de influencia externa, no posee un resultado con el cual compararse, se basa más en argumentos heurísticos para verificar resultados. (Martínez Álvarez, 2012)

Los modelos predictivos o no supervisados son: Agrupamiento (Clustering), Correlaciones y Reglas de Asociación Secuenciales

Agrupamiento (Clustering): También conocidos como métodos de optimización, segmenta un conjunto de datos, validando que cada uno de los

miembros del grupo (cluster) sean lo más similar posible en base a una función de distancia, y que cada cluster que compone el universo sea lo más diferente posible. Existe dos tipos de agrupamientos: Jerárquico, similar a un algoritmo tipo árbol, con desarrollo recursivo, y el Particional que organiza al mismo en k grupos. Por ejemplo: Segmentar el mercado en base a comportamientos de clientes.

Correlaciones: Son las posibles relaciones, es el grado de similitud entre acciones o sucesos que se miran a primera vista independientes, pero poseen algo en común. Se lo mide con el coeficiente de correlación, cuya medida va ente -1 y 1, si posee el valor de 1 son totalmente correlacionadas, en caso que sea el valor -1 estan correlacionadas negativamente, con ayuda de un diagrama de dispersión, se puede visualizar de mejor manera la correlación.

Reglas de Asociación Secuenciales: Expresan patrones de comportamiento indistintamente pero cercanos en el tiempo. Por ejemplo: Si un cliente compra un blue ray es probable que a la semana siguiente compre películas de ese formato.

Cada uno de estos modelos, posee un conjunto de pasos organizados, que deben seguir para dar solución a un problema específico, el cual se le denomina algoritmo, los mismos, que se dividen y se relacionan de acuerdo al modelo.

Entre los algoritmos que se pueden utilizar son:

Árboles de Decisión: Son estructuras que admiten variables discretas y continuas, realiza divisiones sucesivas, utilizando criterios de selección, son tolerantes al ruido, alta inteligibilidad, discrimina la información en función a atributos.

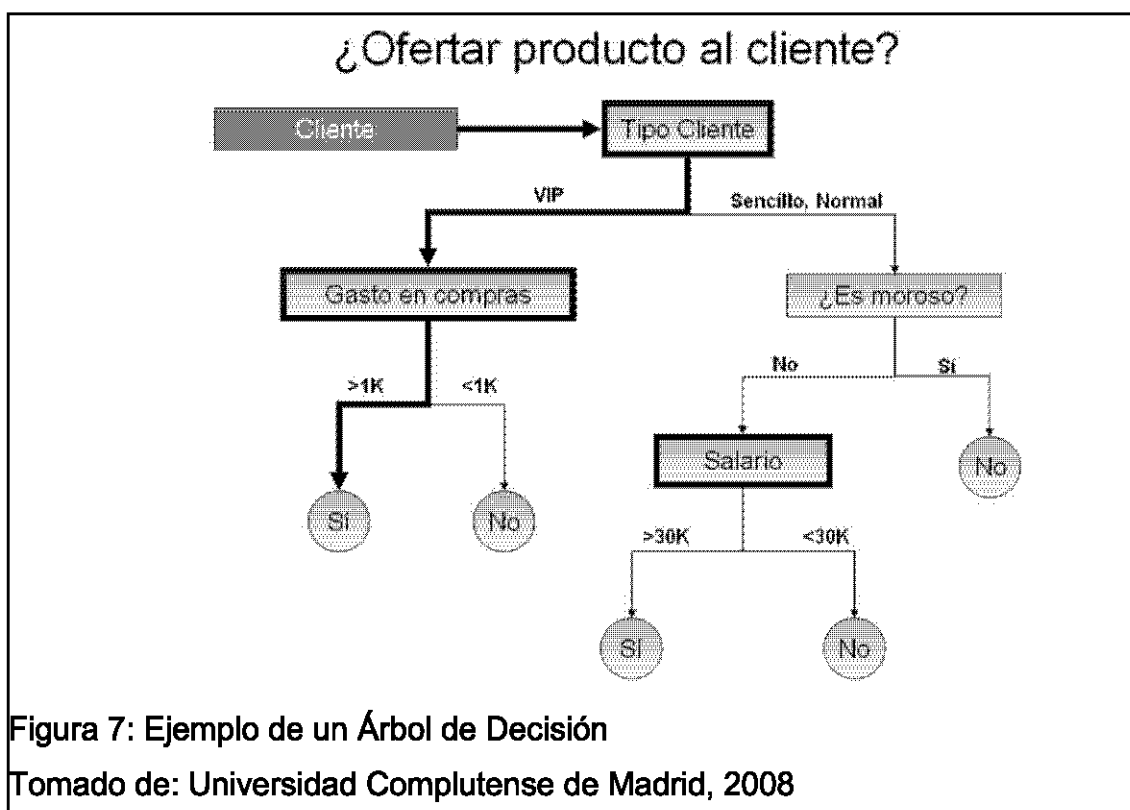
Cada nodo representa una propiedad que puede tomar diferentes valores que

da origen a una rama que se constituye de un subconjunto. El tronco del árbol representa la decisión inicial, las ramas representan las variables de decisión, los nodos hojas representan las clasificaciones finales, en sí éstos últimos son los grupos de datos resultantes.

El número de subconjuntos varía de acuerdo al número de valores que pueda tomar la variable y son disjuntos en todo su estructura de distribución de clases.

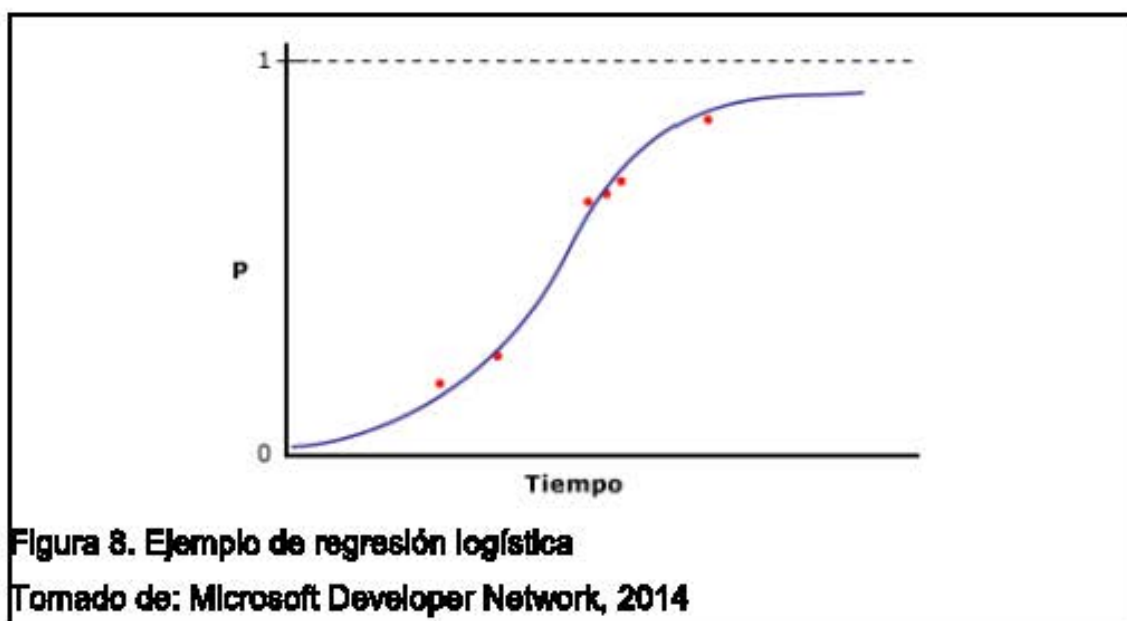
Lo podemos aplicar en decisiones que poseen varias alternativas.

La Figura 7 que se cita a continuación, se puede observar un árbol de decisión de un ofertar producto a un cliente.



Regresión logística: Especie de regresión para predecir el resultado de una variable categórica (variable que puede adoptar número ilimitado de categorías) en función de las variables predictoras.

La Figura 8 que se cita a continuación, se puede observar un ejemplo de cómo se pueden dar los resultados de una aplicación de regresión logística.



Discriminante: Técnica de regresión donde el objetivo es obtener una función capaz de clasificar a un nuevo individuo a partir del conocimiento de los valores de ciertas variables discriminadoras, analiza si existen diferencias significativas entre grupos de objetos respecto a un conjunto de variables medidas proporcionando procedimientos de clasificación sistemática de nuevas observaciones de origen desconocido en uno de los grupos analizados.

La Figura 9 que se cita a continuación, se puede observar un ejemplo del análisis discriminante, donde la línea recta separa dos regiones.

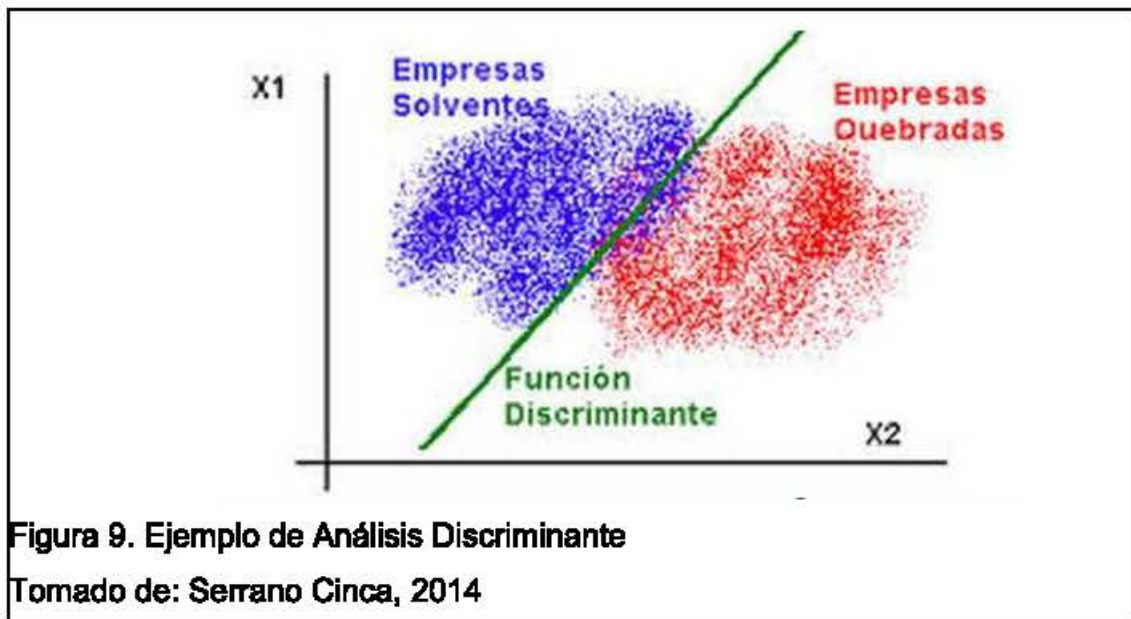


Figura 9. Ejemplo de Análisis Discriminante

Tomado de: Serrano Circa, 2014

Redes Neuronales: Similar a los árboles de decisión, resuelve inconvenientes de clasificación y regresión, detectan y aprenden patrones complejos y características dentro de los datos, que son difíciles de detectar manualmente. Su estructura se basa en el funcionamiento del cerebro, tiene entradas 1 y -1, cada una de las entradas tienen un peso que llega a una función no lineal entre 1 y -1 que no incluye éstos valores, y da como resultado la suma de la multiplicación de las entradas por los pesos, aprenden de la experiencia y del pasado permiten modelar con muchas variables predictoras que direccionan a la resolución de problemas nuevos. Trabajan muy bien con datos complejos y hasta a veces contradictorios, su aprendizaje es asociativo y el mismo puede ser supervisado y no supervisado.

Se construyen en capas de entrada, procesamiento u oculta y salida, permitiendo trabajar en paralelo. Se puede aplicar en las mismas tareas de los árboles de decisión, reconocimiento de patrones.

La Figura 10 que se cita a continuación, se puede observar un la estructura de una red neuronal.

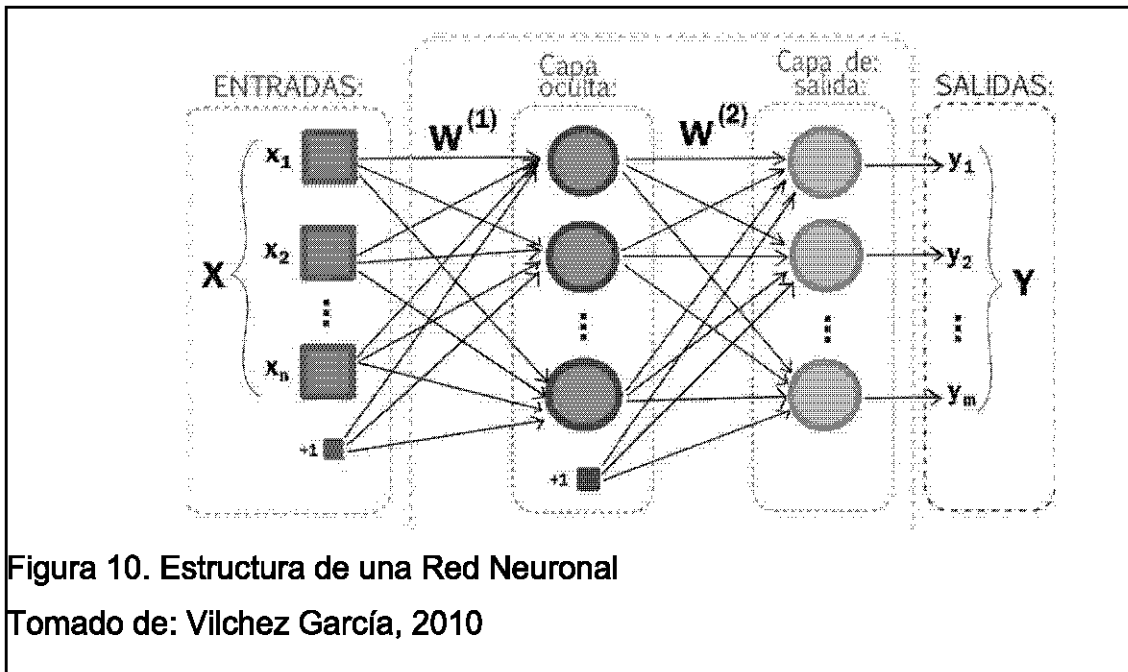


Figura 10. Estructura de una Red Neuronal

Tomado de: Vilchez García, 2010

Mapas de Kohonen (SOM): Es un conjunto de vectores de dimensión n , cada uno de ellos definen un vecindario de 8 o 6, Se usa para conglomerar conjunto de datos en grupos distintos, cuando se presentan juntos son similares, operan con entrenamiento y mapeo, el primero construye el mapa y el segundo clasifica nuevas variables, cada grupo poseerá las mismas salidas.

La Figura 11 que se cita a continuación, se puede observar un ejemplo de empresas quebradas en rojo y empresas abiertas sin color.

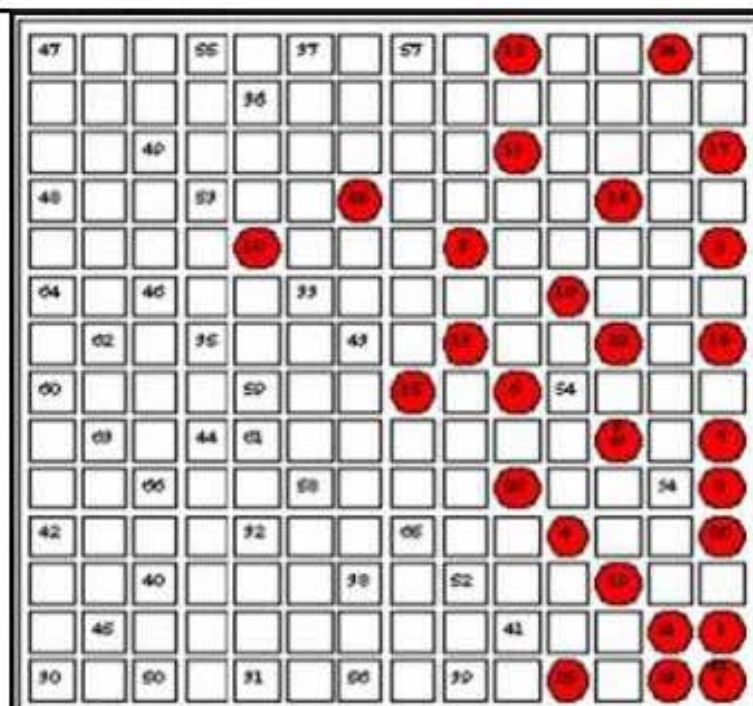


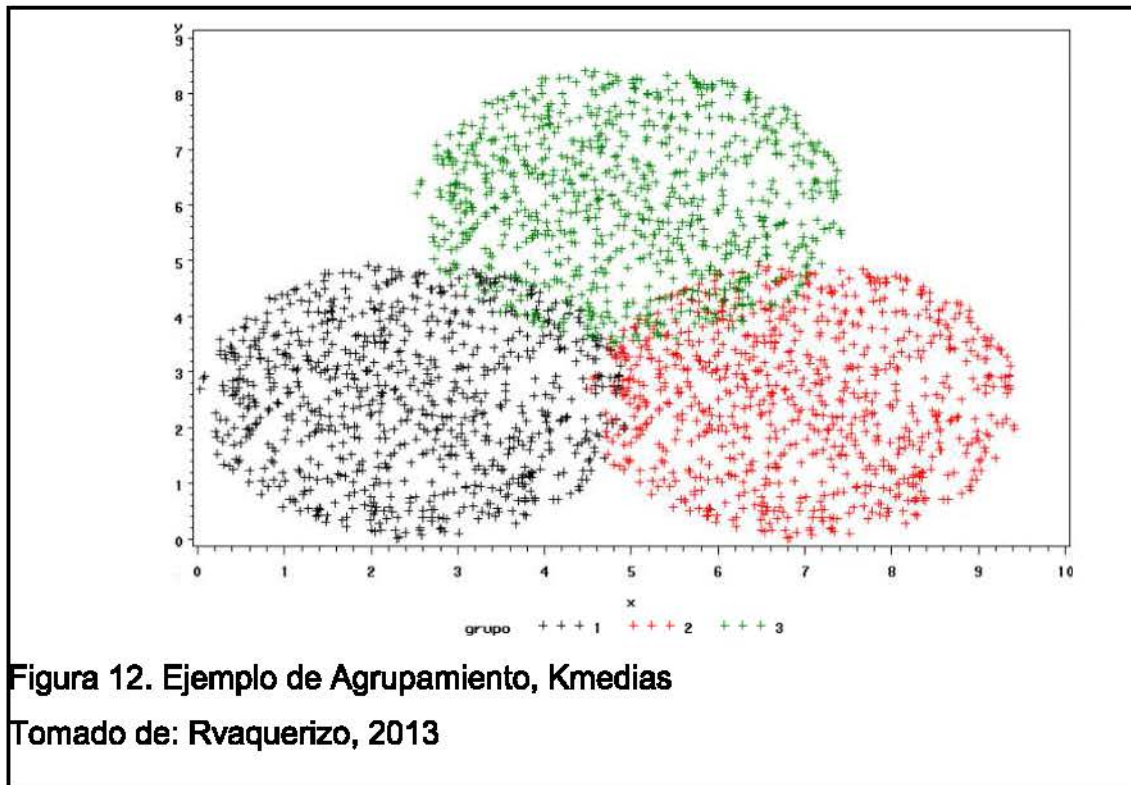
Figura 11. Ejemplo de Mapas de Kohonen

Tomado de: Flores, 2010

K-medias: Es uno de los métodos más utilizados, cuando las variables son cuantitativas, su función es iterativa, distribuyendo los datos ingresados en diferentes clusters, siempre analizando con relación al que se encuentre más cercano en términos de la media, centroide.

En sí, se toman al azar k cluster iniciales, se distribuye los datos hasta que pertenezcan a un grupo de acuerdo a la media. Se verifica cuáles son las clases que se presente más veces y con ésta se clasifica al nuevo caso. El proceso se termina hasta que todos posean su respectivo cluster.

La Figura 12 que se cita a continuación, se puede observar un ejemplo de agrupamiento.



Vecino más cercano o próximo: Modelo de clasificación, reconoce patrones de ubicación, dentro de k casos detecta cuál es la variable más cercana de acuerdo a una localización específica, para evaluar la cercanía depende del tipo de casos que se hace uso, por ejemplo: Parques de diversiones más cercanos a la casa.

La Figura 13 que se cita a continuación, se puede observar un ejemplo de cómo funciona el algoritmo Vecino más cercano.

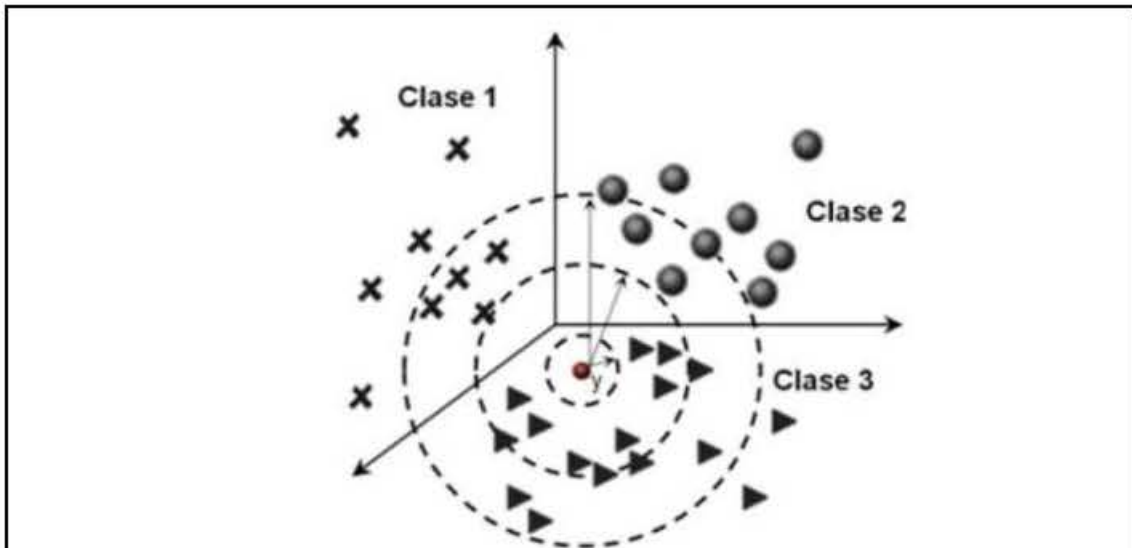


Figura 13. Ejemplo Vecino más cercano

Tomado de: Bernal de Lázaro, Moreno, Oerestes Llanes, & García Moreno, 2011

Análisis de Correlación: Grado o intensidad de asociación de dos o más variables, cuando el grado de asociación es grande el coeficiente de correlación es grande caso contrario tiende a cero, si poseemos dos valores los mismos se encuentran correlacionados si son directamente proporcionales.

La Figura 14 que se cita a continuación se puede observar un ejemplo de correlación existente entre las Px medidas.

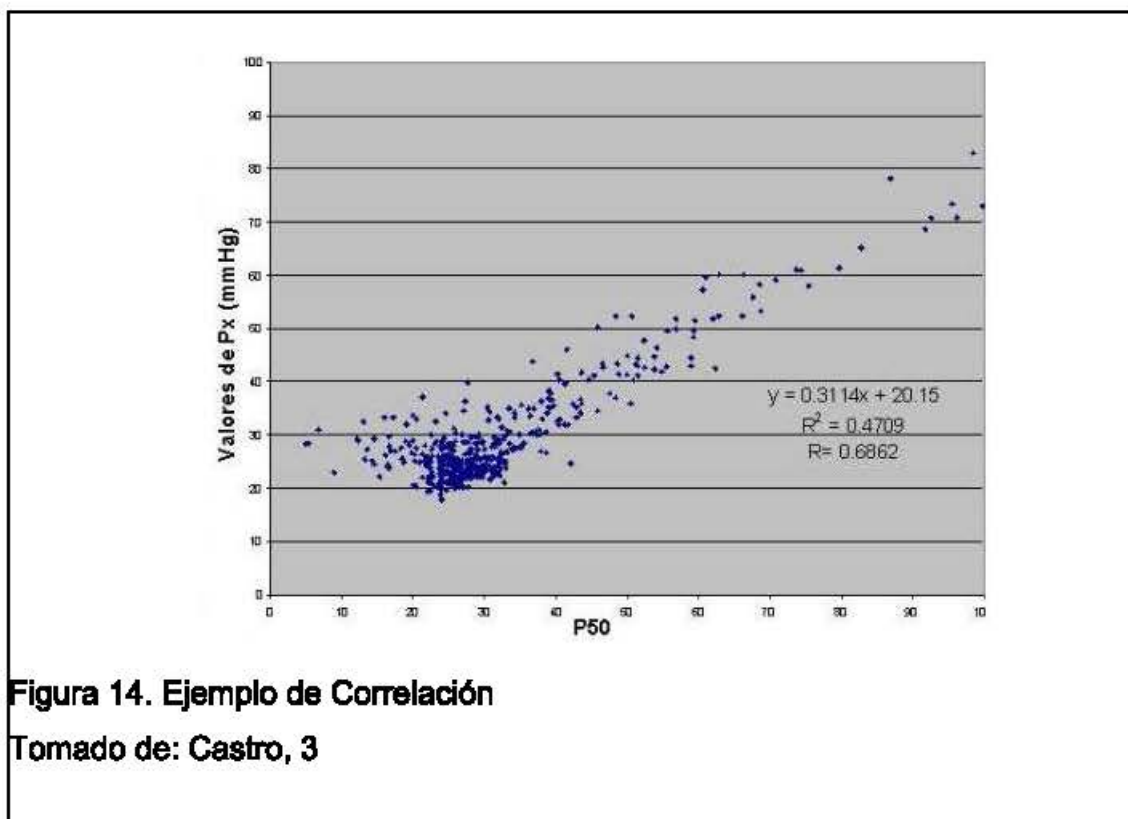


Figura 14. Ejemplo de Correlación

Tomado de: Castro, 3

Bayes: Se puede construir modelos predictivos y descriptivos busca correlaciones entre las variables se los utiliza en aprendizaje basado en la evidencia, combinando probabilidades para llegar a calcular la hipótesis, se lo utiliza comunmente para exploración inicial de los datos. Se basa en el teorema de Bayes, el cuál predice la probabilidad que un caso pertenezca a una clase determinada.

La Figura 15 que se cita a continuación se puede observar una imagen de un clasificador naive Bayes

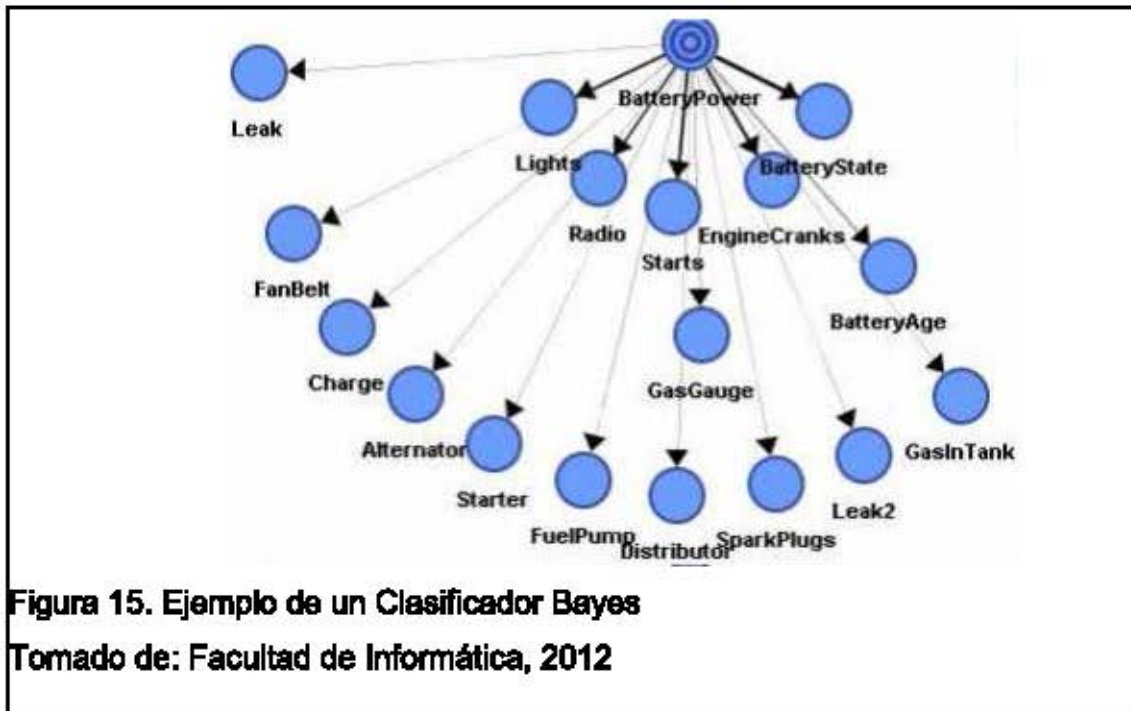


Figura 15. Ejemplo de un Clasificador Bayes

Tomado de: Facultad de Informática, 2012

Regresión Lineal: Es la cuál se modelan los datos usando una línea recta, los coeficientes de la fórmula que utiliza se calcula por el método de mínimos cuadrados, los cuales minimizan entre los datos reales y la línea estimada, existe también la regresión múltiple donde posee más de una variable predictora. Para aplicarla se debe poseer una relación de variables lineal.

La Figura 16 que se cita a continuación se puede observar un ejemplo de regresión lineal simple, que corresponde a 20 pares observaciones: x representa la temperatura fijada en un recinto cerrado e Y el ritmo cardíaco de un vertebrado.

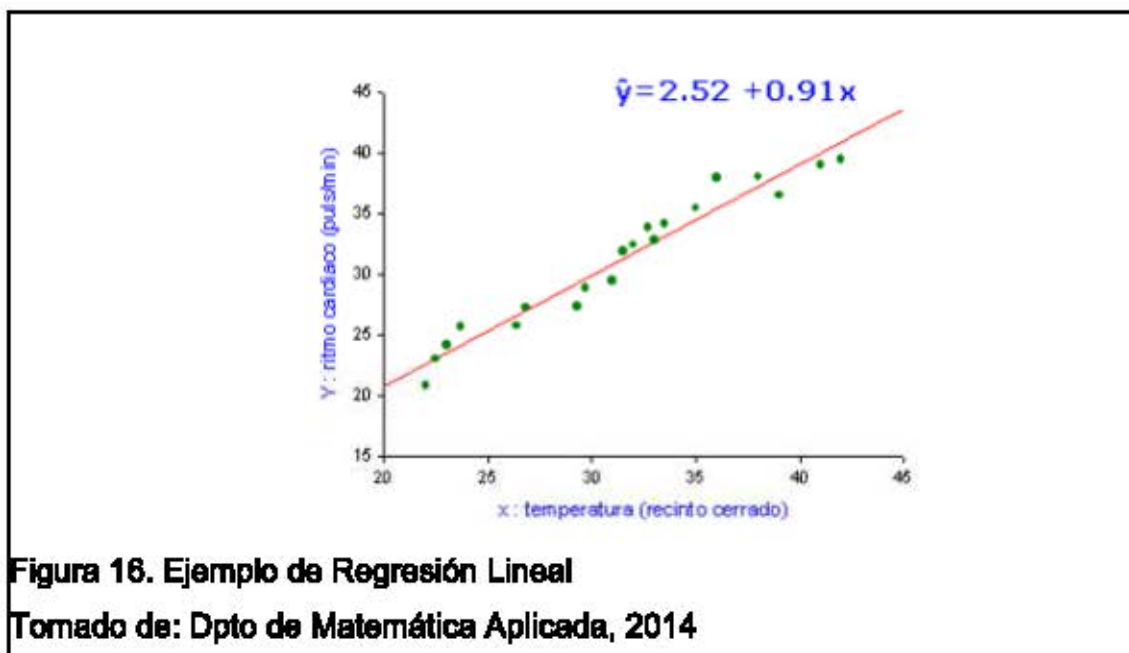


Figura 16. Ejemplo de Regresión Lineal

Tomado de: Dpto de Matemática Aplicada, 2014

Máquinas de Soporte Vectorial: Es el que predice de acuerdo a los datos de entrada las posibles categorías, maximizando el margen entre los grupos formados, construye un hiperplano en un espacio de dimensionalidad alta, es muy utilizado en problemas de clasificación y regresión.

La Figura 17 que se cita a continuación se puede observar un ejemplo de regresión Soporte Vectorial.

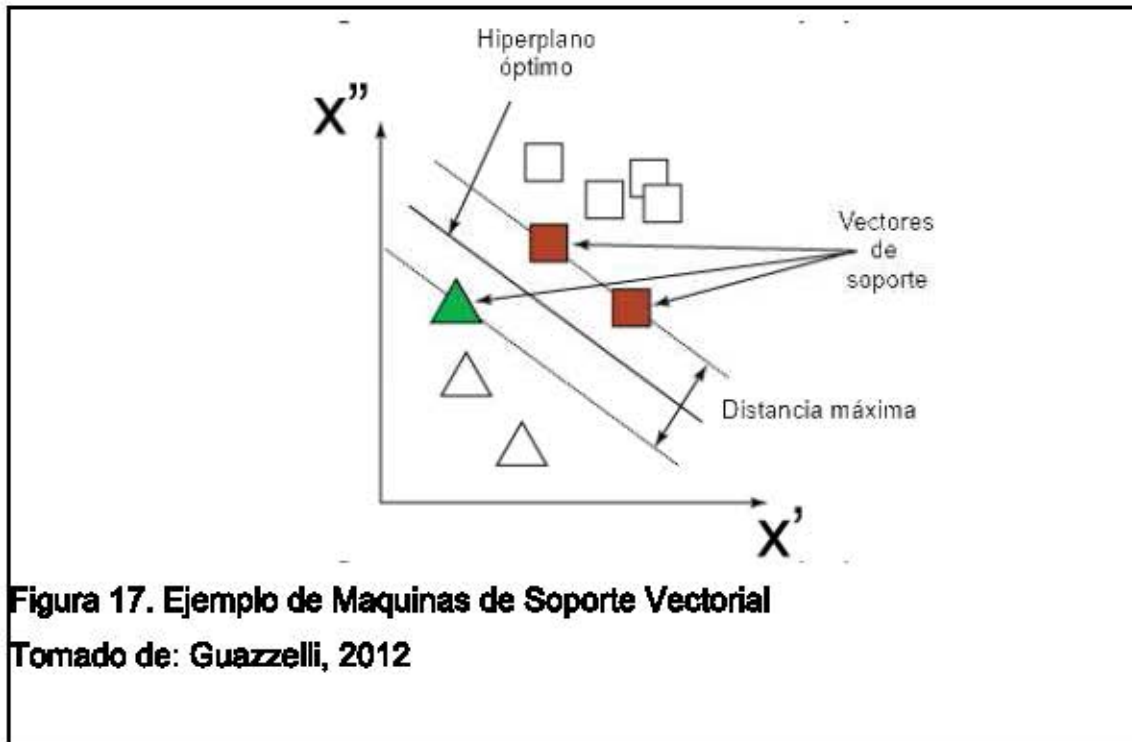


Figura 17. Ejemplo de Maquinas de Soporte Vectorial
Tomado de: Guazzelli, 2012

Algoritmos Genéticos y evolutivos: Al igual que las redes neuronales, se basan en la anatomía humana, evolución de las especies, hacen uso de técnicas biológicas de reproducción mutación y cruce, para ser utilizadas en problemas de búsqueda y optimización, su uso no es tan generalizado como otras técnicas, se basan en el principio de supervivencia, donde solo los más aptos sobreviven. (Luzón Calderon, 2004)

La figura 18 que se cita a continuación se puede divisar un ejemplo de reconocimiento facial realizados gracias a algoritmos genéticos.

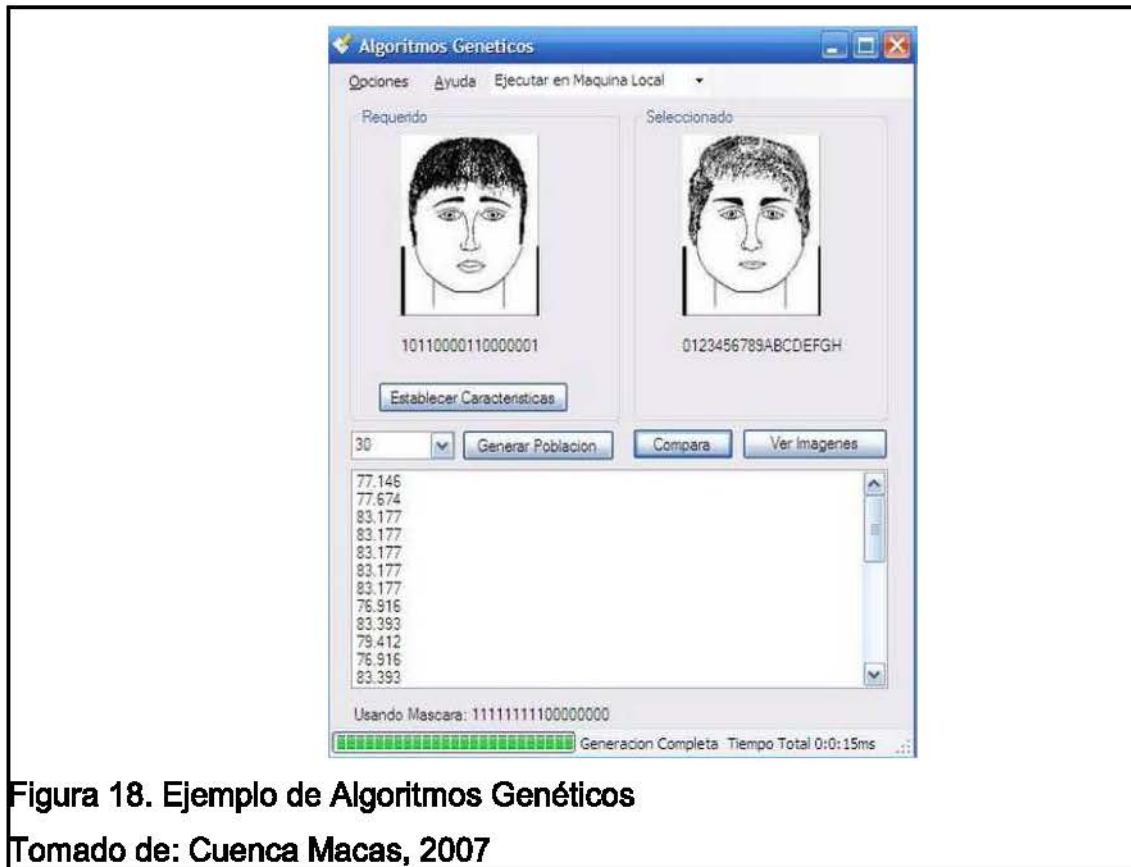


Figura 18. Ejemplo de Algoritmos Genéticos

Tomado de: Cuenca Macas, 2007

La figura 19 que se cita a continuación se puede divisar las relaciones que existen entre los modelos y algoritmos.

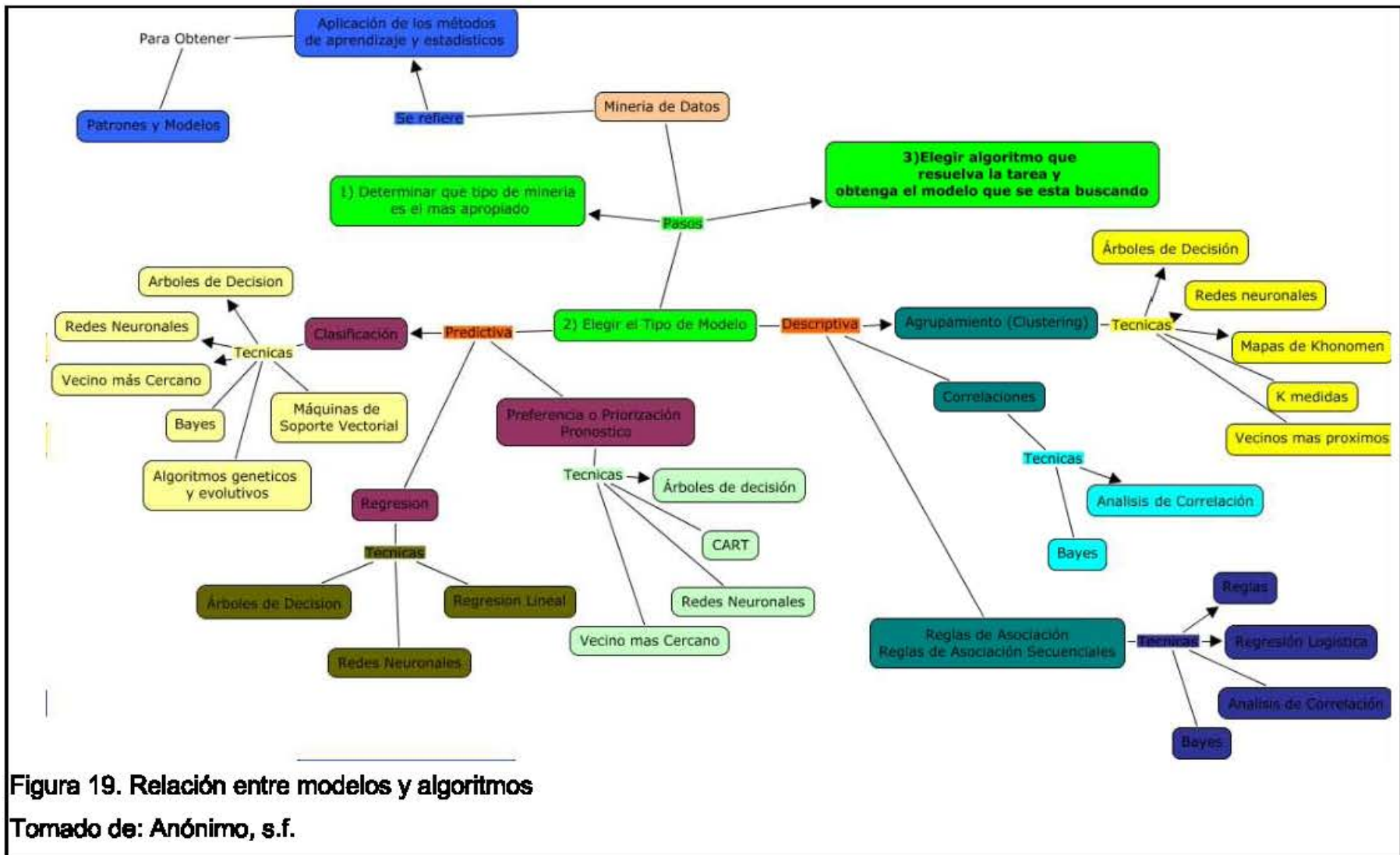


Figura 19. Relación entre modelos y algoritmos
 Tomado de: Anónimo, s.f.

Aunque se realizó un esquema genérico de cómo se puede relacionar los modelos con las técnicas a aplicar se debe tomar en cuenta que no existe limitación en escoger el algoritmo a utilizar, el análisis depende del tipo del dato y el objetivo al que se quiere llegar, se verifica en los diferentes ejemplos encontrados y realizados por analistas experimentados que existen casos que usan un algoritmo para determinar las mejores variables de entrada y después se guían de otro algoritmo para direccionarse al resultado.

2.1.3.3 Arquitectura de Minería de Datos

La arquitectura de Minería de Datos depende mucho de la empresa a la que se va a aplicar, la cantidad de datos que posee, el objetivo al que se desee llegar, la escalabilidad que se desea tener.

Guiándonos en el concepto de lo que significa minería de datos, se conceptualiza y esquematiza la Arquitectura de Minería de Datos con su base fundamental, los datos, éstos como paso inicial deberían encontrarse en un almacén o bodega de datos, aunque la minería no tiene limitantes con el origen de los mismos, es más aconsejable que se encuentre en un repositorio en que los hechos y dimensiones se han determinado, juntamente con una limpieza de datos que garantice su calidad. (Mallikarjunan, 2013)

La Figura 20 que se cita a continuación, se puede observar los procesos genéricos que debe poseer una arquitectura.

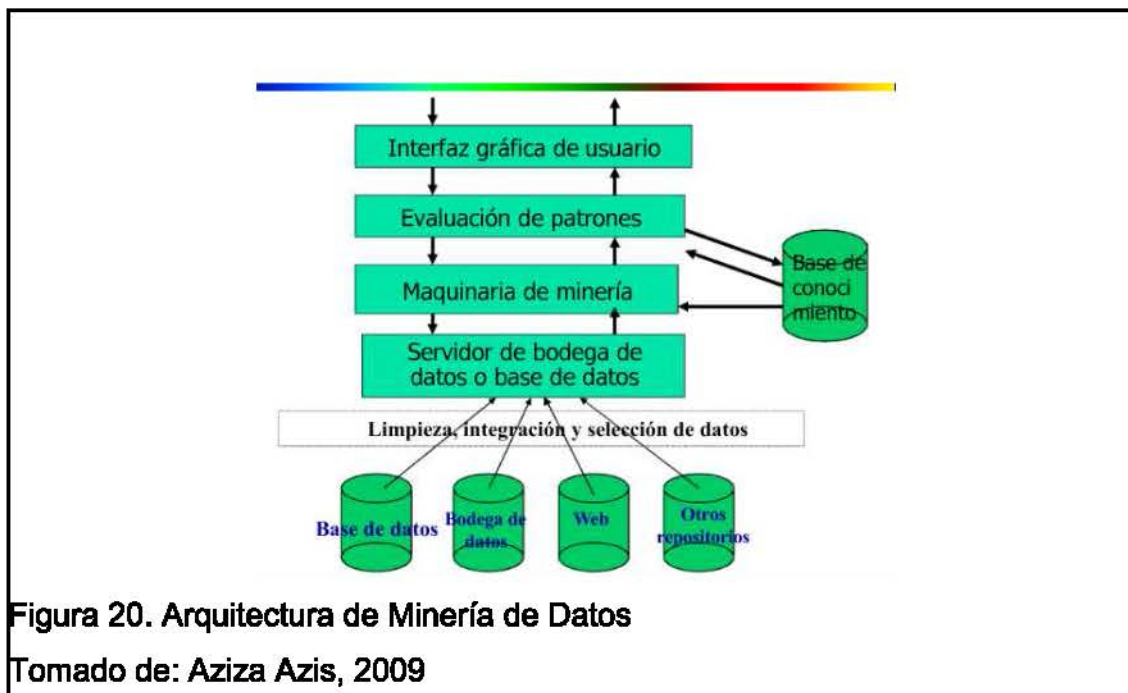


Figura 20. Arquitectura de Minería de Datos

Tomado de: Aziza Azis, 2009

La estructura, interacción y funcionamiento de la arquitectura se encuentra configurado de:

Servidor de Bodega de Datos o Base de Datos: Lugar centralizado de la Organización donde se encuentran información digitalizada la cuál puede ser capturada, almacenada, ordenada, actualizada, preservada y redistribuida, es utilizado para obtener información necesaria de acuerdo al proyecto de minería de datos que se va a realizar, cuya fuente de origen puede ser: Base de datos de un ERP o CRM, data warehouse, bodega de datos, aplicaciones, web, hoja de cálculo u otra clase de repositorio.

Los datos que van a ser parte del servidor de bodega de datos deben pasar previamente un proceso de limpieza, integración y selección.

Limpieza: Es el paso donde se elimina el ruido, como errores y casos anómalos, es el proceso donde también se suprime datos incompletos, inconsistentes, duplicados e irrelevantes.

Integración: Consiste en crear datos y valores en los mismos términos o formatos

Selección: Es suprimir casos atípicos, eligiendo datos de calidad, tanto para elegir la variable objetivo como las variables para el cálculo o proceso.

Maquinaria de Minería: Es el algoritmo de la minería de datos a aplicar, con la cual se puede obtener diferentes patrones que son necesarios para la evaluación e implementación de los resultados requeridos como parte del proyecto que se va a realizar. En el ítem, 2.1.3.1. Técnicas de Minería de Datos, se puede divisar a más detalle los algoritmos de Minería de Datos.

Base de Conocimiento: Umbrales que guían la búsqueda, detalle de los datos que se espera obtener, conocimiento previo y verificación de los patrones con relación al objetivo del proyecto.

Módulo de Evaluación: Interactúa con el algoritmo de Minería de Datos direccionando la búsqueda a patrones que se alineen con el objetivo que se desea alcanzar.

Interfaz Gráfica: Es la capa de interacción con el usuario, visualiza los patrones desde diferentes ángulos.

2.1.3.4 Herramientas de Minería de Datos

Tomando como base que en el 2011 existían 1,8 exabytes de información y el pronóstico para el 2020 es 35000 exabytes, muchas compañías se han preocupado por el tratamiento que se le dan a los datos, con los cuales el apoyo de las herramientas para efectuar esa conversión a información relevante generando valor es indispensable. (Ugalde, 2014)

Entre las empresas que se han preocupado por las tendencias tecnológicas se encuentra Gartner, la cual es una compañía que realiza consultoría e investigación de mercado, centrándose en representaciones gráficas de la situación de un producto tecnológico, que toma el nombre de Cuadrante Mágico de Gartner, el cual se encuentra formado de cuatro secciones: líderes, aspirantes, visionarios y nichos específicos.

Líderes: Son aquellos que poseen gran capacidad de ejecución como: satisfacen la demanda del mercado, tienen una gran cantidad de clientes satisfechos, tamaño y fortaleza financiera. Y una alta Integridad de la Visión, solución adaptada a nuevas tendencias, junto con una estrategia de mercado.

Aspirantes: Son aquellos que tienen una fuerte capacidad de ejecución pero carecen de visión y comprensión de las nuevas tendencias del mercado.

Visionarios: Son los que necesitan apoyo financiero su fortaleza se basa en la visión de la evolución del mercado, débiles en distribución y ventas.

Nichos Específicos: Se concentran en segmentos y nichos de mercado, lo que permite superar a otros proveedores, limitada capacidad de ejecución, sin innovación.

La Tabla 3 que se cita a continuación se puede divisar el cuadrante mágico de Gartner donde se puede visualizar las herramientas para Inteligencia de Negocios y Plataformas Analíticas, distribuidas en cuatro cuadrantes.

Tabla 3. Cuadrante Mágico para Plataformas Analíticas e Inteligencia de Negocios



Tomado de: Sallam, Tapadinhas, Parenteau, Yuen, & Hostmann, 2014

Con relación a ésta premisa y especificando netamente el uso de herramientas para minería de datos se puede citar que las herramientas que se encuentran en el cuadrante de líderes son una buena opción para cumplir con el objetivo planteado inicialmente.

En el grupo de líderes nos vamos a alinear específicamente a IBM, cuyo producto SPSS Modeler, Statistical Package for the Social Sciences, es utilizada actualmente por el SRI.

IBM SPSS Modeler, es una herramienta de Minería de Datos que permite el desarrollo de modelos predictivos de forma rápida e intuitiva, su diseño sigue el modelo CRISP – DM, es decir brinda diferentes procedimientos donde permite

tener una data depurada como input (entrada) e información relevante como output (salida).

Adicionalmente posee una interfaz visual, que permite encontrar relaciones en un gran conjunto de datos con mayor facilidad, en comparación a los métodos estadísticos tradicionales, en éste modelo no se conoce necesariamente los que se está buscando al iniciar la investigación, se exploran datos con apoyo de otros modelos hasta que encuentre la información útil en el campo a tratar. (IBM, 1994, 2011)

2.1.4 Tablero de Mandos

También conocido como Cuadro de Mando Integral (CMI), Balance Scorecard (BSC) o Dashboard, es el que permite traducir y poner en práctica la estrategia de una empresa colocándola en términos operativos, la cual debe ser comprensible para todos los miembros de la organización. (Sinnexus Business Intelligence Informática, 2013)

Su orientación es el seguimiento de indicadores alineados a la estrategia empresarial, se considera como un modelo estratégico pues posee un mapa de causa y efecto adicionalmente es integrado ya que se encuentra esquematizado por cuatro perspectivas que permiten ver a la empresa como un todo.

Sus inicios fueron en el año 1990, cuando Nolan Norton Institute, patrocinó un proyecto de un año de duración: "La medición de los resultados en la empresa del futuro", el líder del estudio fue David Norton, Director General de Nolan Norton, y Roberth Kaplan, como asesor académico.

Luego de varias reuniones el proyecto se basaba en un cuadro integral organizado en cuatro perspectivas, la financiera, la del cliente, la interna, y la de innovación y formación, alineado específicamente a los indicadores de

desempeño de la empresa. A medida que se fue utilizando el cuadro de mando, existieron nuevos aportes como el de Norman Chambers y Brady quienes miraron al cuadro de mando integral como algo más que indicadores, netamente se enfocó a una alineación con las estrategias de la empresa, para su comunicación y respectiva gestión. (Deinsa, 2013)

Actualmente se direcciona específicamente a crear valor al negocio, muchas empresas que lo han implementado han sincronizado el funcionamiento de todos los procesos claves que lo componen, como una sinergia que se ha convertido en la base principal que alinea todas o la mayoría de las perspectivas de la empresa hacia los objetivos proyectados.

2.1.4.1 Importancia de Tablero de Mandos

El poseer un conocimiento de cómo y qué necesitamos realizar para cumplir con el objetivo empresarial es fundamental para que la compañía no se desvíe de su horizonte.

El análisis del mercado donde se quiere sobresalir ofreciendo productos y servicios alineados a las necesidades conocidas y aún las desconocidas por el usuario necesita de un seguimiento continuo, medible, y una estrategia de acción donde se identifique claramente los parámetros que se necesitan seguir para subsistir como una empresa que genere valor al negocio y se fundamente en productos innovadores en un mercado competitivo.

Conocer donde se está trabajando en qué fase se encuentra la empresa es indispensable, por lo tanto poseer indicadores que representen medidas cuantitativas de la funcionalidad de un sistema en un específico tiempo es determinante.

El tablero de Mandos que es una estructura de administración y operación que se fundamenta en una mejora continua y en la consolidación de procesos

basados en una estrategia, respaldada por indicadores, es una herramienta que se convierte en un soporte que abstrae un panorama cuantificado y general de la situación actual de la empresa permitiendo que el cumplimiento de la misión y visión sea más eficiente y eficaz.

2.1.4.2 Tipos de Tablero de Mandos

Existen dos tipos de tablero de Comandos, su ejecución depende de la necesidad de la empresa y el objetivo que se desea llegar.

Cuadro de Mando Operativo (CMO): Se enfoca al seguimiento de variables operativas de área o departamentos de la empresa a analizar, su periodicidad varía de acuerdo a la necesidad de la empresa y representan en su mayoría a procesos. Siempre se encuentra ligado a un Sistema de Soporte a Decisiones, cuyo propósito es el análisis minucioso de información. (Sinnexus Business Intelligence Informática, 2013)

Cuadro de Mando Integral (CMI): Es la representación integral de la estrategia de la compañía enfocada desde la Dirección General, la cuál obligatoriamente debe estar involucrada en todas las fases, se convierte en el modelo de la estrategia desarrollada por la empresa, su estructura básica se fundamenta en cuatro perspectivas: Financiera, cliente, interna, innovación y aprendizaje.

Perspectiva Financiera: La perspectiva financiera depende de la fase que se encuentre pasando el negocio, cuál es nuestra carta de presentación hacia los accionistas y propietarios con relación a un perfil financiero exitoso.

Se basa en tres parámetros de acuerdo al ciclo de vida de la empresa, puede ser considerado en: crecimiento, sostenimiento y recolección.

Crecimiento: Es la fase más temprana del ciclo de vida de productos y servicios de una empresa, el objetivo que busca ésta etapa es direccionar recursos con la finalidad de repotenciar el desarrollo de éstos, para disminuir la operatividad de cash flows (Flujos de efectivo) negativos y aumentar el rendimiento sobre el capital invertido.

Sostenimiento: Es la fase que busca atraer las inversiones con buenos resultados y beneficios, éste se difiere de la fase de crecimiento pues se preocupa más de ampliar la capacidad y realizar la mejora continua, el objetivo que se busca es generar una rentabilidad.

Cosecha o Recolección: Es la última fase que se puede encontrar en una empresa muy pocas son las empresas que han llegado a ésta fase, el objetivo que se busca es garantizar su operatividad.

La Tabla 4 que se cita a continuación se puede divisar las relaciones que existen entre las fases del ciclo de vida de una empresa y sus respectivas estrategias en cada etapa.

Tabla 4. Perspectiva Financiera

	Fases	Temas Estratégicos		
		Crecimiento y diversificación de los ingresos	Reducción de costos / mejora de la productividad	Utilización de los activos
Estrategia de la unidad de negocio	Crecimiento	Tasa de crecimiento de las ventas por segmento Porcentaje de los ingresos de nuevos productos, servicios y clientes	Ingresos / empleados	Inversiones (% de ventas) I + D (% de ventas)
	Sostenimiento	Cuota de cuentas y clientes seleccionados Venta cruzada Porcentaje de ingresos de nuevas aplicaciones	Costo frente a competidores Tasas de reducción de costos Gastos indirectos	Ratios de capital circulante (ciclo de maduración) ROCE por categorías de activos clave Tasas de utilización de los activos
	Recolección	Rentabilidad de la línea de productos y clientes Porcentaje de clientes no rentables	Costos por unidad (por und. de output por transacción)	Periodo de recuperación (pay-back) <i>Throughput</i>

Tomado de: Kaplan & Norton, 2000

Perspectiva del Cliente: La perspectiva del Cliente es como se presenta ante los clientes tanto internos como externos, es en sí el posicionamiento en el segmento que se escoge competir, las empresas los identifican con relación al consumidor y el mercado, todo enfocado al objetivo empresarial.

Se orienta en aquellos productos que son alineados con las preferencias de los clientes cuantificando el valor añadido que se busca entregar a éstos, y alineándolo específicamente a algo innovador y/o un servicio de calidad.

Perspectiva Interna del Negocio: Es la identificación de procesos críticos en la búsqueda de objetivos empresariales que satisfagan las necesidades de accionistas, clientes internos o externos y proveedores, identificando con eficacia una cadena de valor real que se ajuste a los objetivos empresariales

Perspectiva de Aprendizaje y Crecimiento: Es el proceso clave, que se encarga de verificar que las tres perspectivas anteriormente citadas lleguen a cumplir su objetivo, por medio de una infraestructura bien estructurada, dirigida a las capacidades de los empleados, sistemas de información, motivación. (Deinsa, 2013)

2.2. Marco Conceptual

Ciudadanía Fiscal: “En el Ecuador la Ciudadanía Fiscal es la conciencia de las personas, grupos e instituciones, acerca de su responsabilidad y compromiso con la construcción de una sociedad cohesionada, democrática y justa, a través del ejercicio de sus derechos y deberes fiscales” (Marx Carrasco, 2010).

CIIU: Clasificador determinado por las Naciones Unidas con la finalidad de establecer uniformidad en las actividades económicas productivas.

Cohesión social: Se consolida en un solo concepto que es la ciudadanía fiscal, la actitud consciente y voluntaria de mujeres y hombres en el cumplimiento de sus obligaciones fiscales, como condición para el ejercicio pleno de sus derechos. (Carrasco, 2011)

Clustering: Agrupamiento cuyo proceso consiste en la división de los datos en grupos de objetos similares. (Conocimiento con todos y para todos EcuRed, 2014)

Elusión: Vía legal para evitar el pago de impuestos.

Evasión: La evasión fiscal o tributaria es el impago voluntario y premeditado de los impuestos que determina la Ley.

Evento Relevante: Alta Afectación al Core (Núcleo o centro) del Negocio.

Heurístico: Es aquel algoritmo no respaldado por un procedimiento matemático que busca soluciones innovadoras más sencillas y rápidas guiados más por la experiencia a un problema planteado, no siempre garantiza encontrar el resultado óptimo. (Rodríguez Ortiz, 2009)

Plan Nacional del Buen Vivir: Buen Vivir, es la forma de vida que permite la felicidad y la permanencia de la diversidad cultural y ambiental; es armonía, igualdad, equidad y solidaridad. No es buscar la opulencia ni el crecimiento económico infinito. (Embajada del Ecuador, 2014)

Empresas o Sociedades Instrumentales: Son aquellas que generalmente solo existen en papel, no desarrollan actividades mercantiles significativas y sus activos son insignificantes. Su objetivo de creación es la confidencialidad que ofrecen a sus accionistas como una sociedad anónima, ocultando o disimulando a sus verdaderos propietarios.

Tipo Impositivo Efectivo: Es el impuesto a la renta causado sobre el Total de ingresos, se encuentra 0 y 24% (Porcentaje que se graba a la base imponible 2012)

CAPÍTULO III: METODOLOGÍA

3.1. Metodología de Inteligencia de Negocios

La metodología de Inteligencia de Negocios que se va a aplicar en la Institución, Servicio de Rentas Internas, es CRISP-DM, acrónimo de Cross-Industry Estándar Process for Data Mining, la cual de acuerdo al análisis realizado en el Capítulo 2, es una de las más utilizadas actualmente, puesto que brinda un detalle minucioso de lo que se considera una buena práctica en la búsqueda de patrones válidos en un conjunto de datos, cuyo objetivo es brindar información y conocimiento que sirvan de soporte en la toma de decisiones alineándose siempre a los objetivos institucionales.

3.2. Fases

3.2.1 Fase de comprensión del negocio o problema

El Servicio de Rentas Internas posee objetivos estratégicos, los cuales son un lineamiento que solidifica su institucionalidad, contribuye a la cohesión social y promueve la generación de riqueza a favor del país, a corto, mediano y largo plazo.

Contextualizando podemos citar:

Determinación de los objetivos del negocio

El Problema a resolver se alinea en la creación de un concepto de **Importancia Fiscal**, el cuál será construido inicialmente con los diferentes criterios expertos de los Servidores de Gestión Tributaria, a los cuáles se les enviará una encuesta, Anexo 1, donde las respuestas que se obtengan de la misma serán la guía para poseer una base inicial y común entre las diferentes áreas que son partes activas del

departamento en mención, las cuáles darán como resultado la variable objetivo necesaria para la investigación, adicionalmente con apoyo de los datos que se encuentran almacenados en el SRI, se socavará y entrenará en comparación a ésta variable objetivo inicial la información necesaria para poseer una **Importancia Fiscal** basado en conocimiento y no solo en un criterio establecido por el analista evaluador de ese momento.

El SRI posee una gran cantidad de datos almacenados, el descubrimiento de conocimiento adquiere importancia, ya que el mismo permite extraer patrones válidos que inicialmente se encuentran ocultos. La Minería de Datos, la cual es una técnica que apoya a un manejo centralizado de información, está direccionada a crear éste nuevo conocimiento tomando como input datos depurados y direccionando a un resultado que servirá como base para la rectificación o la toma de nuevas decisiones.

La toma de nuevas decisiones siempre se va alinear a los objetivos estratégicos del SRI, los cuales son:

- Incrementar el uso eficiente del presupuesto.
- Incrementar el desarrollo del talento humano.
- Incrementar la eficiencia operacional.
- Incrementar la conciencia de la ciudadanía acerca de sus deberes y derechos fiscales.
- Incrementar la aplicación de los principios constitucionales en materia tributaria.
- Incrementar la eficiencia y efectividad en los procesos de asistencia y control enfocados al cumplimiento tributario, sustentados en un modelo de gestión de riesgos.

Con ésta inferencia, una medida clave de éxito de ésta tesis es tener una herramienta que identifique de acuerdo a los parámetros asignados grupos de sociedades clasificados por su **Importancia Fiscal**.

Evaluación de la situación.

El SRI es una institución estructurada en componentes bajo una misión y visión general, con relación a éste esquema posee diferentes procesos con sus respectivos procedimientos, cuya calidad de operación siempre se ha alineado a una mejora continua, pero a pesar que su organización es muy bien constituida, se ha detectado que todavía algunos criterios operacionales dependen específicamente de cada uno de los analistas, la **Importancia Fiscal** podemos citar como parte de éstos, el cuál debería ser un concepto de negocio estandarizado, pero de acuerdo a criterios recopilados en diferentes reuniones se ha verificado que su estructuración depende de un criterio particularizado y basado en la experiencia y lineamientos de los diferentes expertos tributarios, bajo ésta realidad, se está buscando un criterio generalizado como parte de ésta investigación en las áreas de Auditoría y Control Tributario alineados a los objetivos de toda la institución como fase inicial y alcance de la presente investigación.

El concepto de **Importancia Fiscal** va a ser ratificado con apoyo de encuestas, (ver en el Anexo 1), las cuales van a ser direccionadas a los expertos del Departamento de Gestión Tributaria del Servicio de Rentas Internas, consecuentemente con apoyo de herramientas de minería de datos más el concepto anteriormente mencionado se clasificará a los contribuyentes, específicamente sociedades.

Por lo tanto, lo que se necesita es manejar la **Importancia Fiscal** de forma generalizada con la finalidad de orientar la segmentación de los contribuyentes, basado netamente en información y criterios expertos con apoyo de herramientas tecnológicas que nos permitan minar de mejor manera las relaciones entre las diferentes variables.

Adicionalmente se ha enfatizado la investigación en reuniones con los funcionarios responsables de las Áreas de Gestión Tributaria, con la finalidad de identificar los principales riesgos que se pueden presentar de acuerdo a la realidad operativa del SRI.

En la Tabla 5 se puede visualizar los principales riesgos y contingencias que se dedujo de las diferentes reuniones mantenidas.

Tabla 5. Riesgos y Contingencias

RIESGOS	CONTINGENCIAS
Construir un concepto de Importancia Fiscal particularizado por área o analista responsable.	Se realizará reuniones generalizadas con los responsables y directores del Negocio del área a evaluar, donde se concluirá con un concepto común.
Poseer variables que no se alineen al 100% a la realidad, darían como resultado un modelo que no cumplirá con el objetivo para lo que fue creado.	Se escogerá las variables más aptas que cumplan con el objetivo que se quiere llegar con ayuda de los expertos y responsables tributarias del área de Gestión Tributaria

Por otro lado el Costo - Beneficio del proyecto se lo va a enfocar al costo que implica realizarlo y ejecutarlo con relación netamente a la automatización de sueldos del personal, no se evaluó una posible consecución de aumento en el valor de la recaudación, pues no se posee datos concretos con los cuales realizar el análisis con un detalle exhaustivo.

En la Tabla 6 se visualiza el Costo – Beneficio del Proyecto donde se puede verificar que el primer año se puede recuperar la inversión, con un flujo positivo de \$3867.

Tabla 6. Costo - Beneficio

	Inversión Inicial	Año 1	Año 2	Año 3	
INGRESOS					
Salarios que se dejan de abonar: Jefe Departamental, Experto y 2 horas de Analista		52.200	52.200	52.200	
TOTAL INGRESOS		52.200	52.200	52.200	
EGRESOS					
Mano de Obra	17.400				
Experto BI (3 meses)	7.200				
Expertos tributarios (3 meses)	7.200				
Analista (3 meses)	3.000				
Personal de operacion y Mto. (6 horas/dia)		9.000	9.000	9.000	
Entrenamiento (16 horas)	3.000				
Instructor tecnico y operativo de la herramienta	1.000				
movilizacion	2.000				
Equipo y Sistemas	18.100				
SPSS Modeler	15.000		3.150	3.150	
Equipo de Computación	2.500				
Software Base	600				
Depreciación Equipo de Computación		833	833	833	
Total Egresos	38.500	9.833	12.983	12.983	
UTILIDAD NETA	-38.500	42.367	39.217	39.217	
Ahorro por año		3.867	43.084	82.301	AHORRO ACUMULADO EN LOS TRES AÑOS
Salario del Flujo de obtención de Datos Actualmente*					
Jefe Departamental	2.400				
Experto	1.700				
Analista	1.000				
Sueldo y Salario Fijo Mensual	5.100				
*Remuneraciones tomadas de la tabla salarial SRI					

En la tabla 6, los ingresos se basa netamente en la comparación del número de funcionarios que hacen posible la obtención de información, en la situación actual, se necesita tres funcionarios: jefe departamental, experto y analista, en cambio en la situación que se propone solo se necesitaría un analista.

Con relación a los egresos, en el ítem equipo y sistemas, se evalúa el software con costo y el hardware necesario asimilando con un equipo de escritorio de cuarta generación, Core i7, 1T en almacenamiento, el mismo que se seleccionó en función a los recursos necesarios para el correcto funcionamiento del aplicativo, tomando en cuenta que sólo se va hacer un llamado a la base de datos donde se necesita extraer la información por lo tanto no es necesario la utilización de todos los recursos que nos brindaría un servidor si no solamente se utilizaría un porcentaje de ellos.

En la Tabla 7 se puede visualizar los beneficios en forma general de la herramienta que se va a implementar

Tabla 7. Beneficios de Implementación

DESCRIPCION	MODELO TRADICIONAL	MODELO CON MINERÍA DE DATOS
Construcción del concepto de Importancia Fiscal dirigida al control	Personal	Estandarizado previamente en base a criterios expertos
Nivel de conocimiento para reportear variables que conforman parte de la Importancia Fiscal	experto	cualquier persona
Interpretación de resultados	Experiencia y su realidad local	Reporteador/Herramienta de Data Mining
Automatización en el análisis de Importancia Fiscal	Criterio experto	Clasificador automático
Herramientas que se utilizan para la obtención de los datos	Herramientas de bd y office	SPSS Modeler
Visualización de reportes	Diferentes herramientas	Saiku Reporting
Tecnologías que permiten la selección de los datos	selección de acuerdo a la necesidad	Algoritmos avanzados
Metodologías que se utilizan para la resolución del problema	selección de acuerdo a la necesidad	CRISP-DM
Personal encargado para realizar la obtención de la información	Expertos en el negocio	Cualquier servidor o funcionario del SRI
Alcance de la Información, quien puede verla y manipularla	Experto tributario, Gerente	Operativo, táctico y estratégico
Flujo de trabajo para la obtención de los datos	Analista – Extrae la información Experto – valida y analiza la información de acuerdo a su criterio. Jefe departamental - Aprueba.	Requiere
Tiempo para la obtención de la información	2 días*	minutos

Nota: *Se organizan planes de control, donde se reúnen directores, jefes departamentales, jefes de área a nivel nacional y expertos tributarios para líneas estratégicas tácticas y operativas, dentro de éste tipo de talleres se consolidan los criterios para determinar sectores económicos, criterios de riesgo, y contribuyentes de mayor importancia.

Determinación de los objetivos de Data Mining.

Lo que se busca al aplicar minería de datos es:

- Valorar, depurar, explotar las variables y criterios de análisis de la data que fue seleccionada en el Servicio de Rentas Internas.
- Determinar y segmentar los grupos de interés, Contribuyentes, Sociedades Activas a partir de técnicas de Minería de datos.
- Viabilizar una herramienta que permita visualizar los resultados que se obtengan de la Minería de Datos.

El criterio de éxito de la Minería de Datos es socavar información que apoye a la toma de decisiones para determinar el grado de Importancia de las Sociedades Activas.

Realizar el plan proyecto

Las herramientas que se van a utilizar para la realización de éste proyecto es SPSS Modeler la cual es una herramienta adquirida por el Servicio de Rentas Internas.

SPSS Modeler ofrece diferentes técnicas a seguir que se basan en procedimientos alineados en la metodología CRISP - DM.

Las técnicas que se ofrecen se encuentran en base a algoritmos, los cuales son necesarios para obtener información que a simple vista no se visualiza y que son fundamentales para la toma de decisiones, la técnica a escoger depende de la necesidad que posee el Servicio de Rentas Internas, que es esquematizar un concepto de **Importancia Fiscal** y de acuerdo a éste criterio clasificar las sociedades activas del año 2012, bajo ésta premisa, se escogerá algoritmos no supervisados para obtener la variable objetivo y en base a la misma se aplicará algoritmos supervisados, para la clasificación de las Sociedades.

3.2.2 Fase de comprensión de los datos

Los datos que se han identificado como parte de la fase inicial de la investigación son aquellos que constan como características de cada uno de los Contribuyentes, Sociedades, año 2012, la selección de los mismos en una forma generalizada se lo realizó en base a la solidificación de diferentes criterios expertos del negocio específicamente en el ámbito de la Gestión Tributaria, tomando como lineamiento el Control Tributario.

Contextualizando podemos citar:

Recolección de datos iniciales.

Los datos iniciales se han recolectado de la información que se ingresa o se actualiza cada día de cada Contribuyente, el mismo que se lo consolida en un Data Warehouse. Adicionalmente la selección y pre procesamiento de los datos para nuestro objetivo que es la clasificación de Sociedades en base a su **Importancia Fiscal** se lo ha realizado de acuerdo a criterios basados tanto en las respuestas de las preguntas realizadas en el Anexo 1 como entrevistas, Anexo 3, donde se deduce cuáles son las variables más relevantes que van a ser parte de éste proceso; como primera segregación, se ha tomado el periodo fiscal, año 2012, en el entorno específico de la Gestión Tributaria, enfocada en el ámbito del control.

Alineado a toda la información recolectada en el Anexo 1 y 3 se ha seleccionado las siguientes variables:

- RUC
- Prelación
- Regional

- Tipo
- Clase
- Código Actividad CIIU Nivel 6
- Pertenencia a Grupo Económico
- Activos (Declaración de renta 2012)
- Pasivos (Declaración de Renta 2012)
- Impuesto a la Renta Causado (Declaración de Renta 2012)
- Número de empleados
- De Transaccionalidad (2012):
 - Máximo Ingresos Internos (fuentes propias y de terceros)
 - Máximo Ingresos Externos (fuentes propias y SENA)
 - Máximo compras Internas Reportadas (fuentes propias y terceros)
 - Máximo compras externas
 - Máximo pagos por Sueldos y Salarios
 - Ranking de Transaccionalidad.

Descripción de los datos

Cada una de las variables se las describe como se visualiza en la Tabla 8.

Tabla 8. Descripción Variables

VARIABLES	IDENTIFICACIÓN
RUC	Registro Único de Contribuyentes*
Prelación	Clasificación de las Sociedades en: Gran Contribuyente Especial Sociedad Bajo el Control de Superintendencias Sector Público Grupo Económico Naturales Obligados Otra Sociedad con fin de lucro Sociedad sin fin de Lucro Naturales no Obligados Rise
Regional	Lugar de pertenencia del Contribuyente
Tipo	División de Contribuyentes en naturales y jurídicos
Clase	División de Contribuyentes en especiales y otros
Código Actividad CIU Nivel 6	Clasificación industrial internacional uniforme de las actividades económicas
Pertenencia a Grupo Económico	Concatenación de una sociedad con otra o más sociedades
Activos	Ingresos Declarados
Pasivos	Egresos Declarados
Impuesto a la Renta Causado (2012)	23% de la base imponible
Número de empleados	Cantidad de trabajadores que posee la Sociedad
Máximo Ingresos Internos	Ingresos internos fuentes propias
Máximo Ingresos Externos	Ingresos internos fuentes externas y de terceros
Máximo compras Internas Reportadas	Compras internas realizadas
Máximo compras externas	Compras externas realizadas
Máximo pagos por Sueldos y Salarios	Sueldos y Salarios
Ranking de Transaccionalidad.	Nivel de Transaccionalidad, máximo de ingresos por sus diferentes fuentes

Nota: *Para objeto de la presente investigación solo se analizará Sociedades, no incluye Personas Naturales.

Exploración de datos.

Para proceder con el entendimiento de los datos se ha realizado gráficos con ayuda de la herramienta SPSS Modeler donde podemos identificar si los mismos son congruentes o no con la realidad.

En La Figura 21 que se cita a continuación se puede divisar el modelamiento de entendimiento de datos.

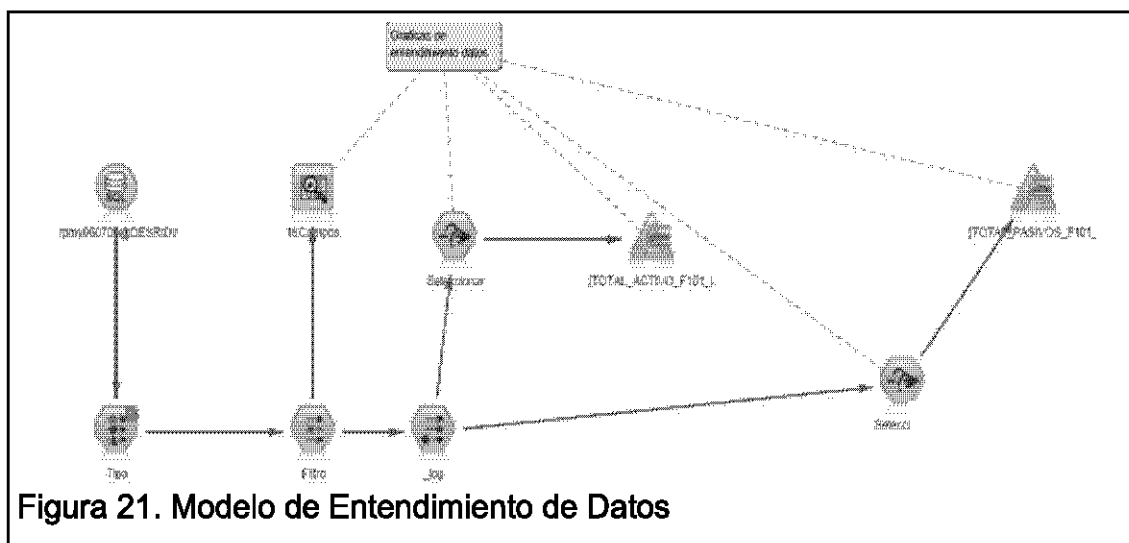
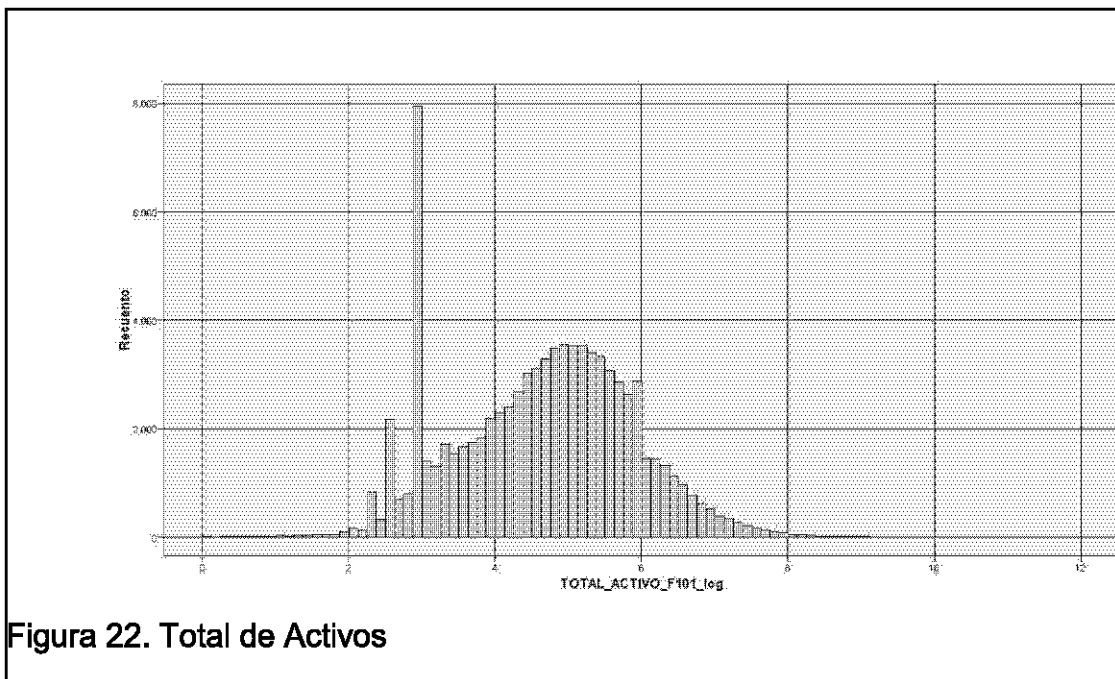


Figura 21. Modelo de Entendimiento de Datos

A continuación se cita el procedimiento que se realizó, para evaluar cada una de las variables.

En la Figura 22 que se cita a continuación se puede divisar en base logarítmica, el total de activos que posee cada Contribuyente, Sociedades año 2012.



La gráfica en mención se ha analizado que el pico que se presenta para cerca de 8000 sociedades se encuentra acorde a la realidad puesto que el monto de capital mínimo para crear sociedades por ley está entre 400 y 800 es por eso que se presume que las sociedades poseen el monto mínimo de creación, como total de ingresos, claro que éste debería ser un ítem de análisis más profundo de acuerdo al objetivo que se desea alcanzar pero en el caso de la presente investigación éste factor en forma individual no afecta directamente.

El caso de las demás variables se puede citar que no se presenta un desfase que sea objeto de análisis profundo particular, y que cada una de sus gráficas posee un comportamiento normal.

Verificación de la calidad de los datos.

Cuando se refiere a la calidad de datos, se puede evaluar que la mayoría de sus campos poseen consistencia, que se debe realizar un tratamiento específico con algunas variables que poseen valores vacíos pero en si no se encontró valores fuera de rango que generaran ruido, se puede concluir que la

data posee una buena calidad y no existe valores desfasados de forma incorrecta.

En la Tabla 9 se visualiza los resultados del tipo de datos que se posee, en el mismo se puede verificar que existe valores nulos en variables: grupo económico, ranking, los cuales se les dará tratamiento en la fase preparación de datos.

Tabla 9. Detalle Descriptivo de las Variables

	Clase	Metodo	Valores tipicos	Externos	Modo	Instituciones	Metodo	# Contabilidad	Registros y datos	Valor min	Cadena words	Exposicion Blanco	Valor max
1	ANO FISCAL	Normal	--	--	--	Nunca	Eje	100.000	145184	0	0	0	0
2	PREFENCIA A GELNO ECONOMICO	Normal	--	--	--	Nunca	Eje	1334	163	14251	0	0	0
3	DESCRIPCION PRELACION	Normal	--	--	--	Nunca	Eje	100.000	145184	0	0	0	0
4	DESCRIPCION CLASE CONTAB	Normal	--	--	--	Nunca	Eje	100.000	145184	0	0	0	0
5	DESCRIPCION REGIONAL	Normal	--	--	--	Nunca	Eje	100.000	145184	0	0	0	0
6	TOTAL ACTIVO F101	Continua	70	96	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
7	TOTAL PASIVOS F101	Continua	66	83	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
8	IMPUESTO RENTA CARGADO F101	Continua	7	12	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
9	INGRESOS TRX INTERNAS	Continua	65	65	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
10	INGRESOS TRX EXTERNAS	Continua	20	13	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
11	COSTOS GASTOS INTERNOS PROPIOS	Continua	1	2	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
12	COSTOS GASTOS EXTERNOS PROPIOS	Continua	101	161	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
13	COSTOS GASTOS TRX INTERNAS	Continua	1	2	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
14	COSTOS GASTOS TRX EXTERNAS	Continua	104	164	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
15	PAGO SEGUROS SALARIOS PROPIOS	Continua	16	25	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
16	PAGO SEGUROS SALARIOS TERCEROS	Continua	220	312	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
17	RANKING	Normal	--	--	--	Nunca	Eje	89.292	129697	15447	0	0	0
18	DESCRIPCION BANKING	Normal	--	--	--	Nunca	Eje	89.292	129697	15447	0	0	0
19	ACTIVIDAD	Normal	--	--	--	Nunca	Eje	100.000	145184	0	0	0	0
20	TOTAL ACTIVO F101_log	Continua	0	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
21	TOTAL PASIVOS F101_log	Continua	2	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
22	IMPUESTO RENTA CARGADO F101_log	Continua	559	1	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
23	INGRESOS TRX INTERNAS_log	Continua	0	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
24	INGRESOS TRX EXTERNAS_log	Continua	2333	2626	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
25	COSTOS GASTOS INTERNOS PROPIOS_log	Continua	2	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
26	COSTOS GASTOS EXTERNOS PROPIOS_log	Continua	3937	31	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
27	COSTOS GASTOS TRX INTERNAS_log	Continua	2	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
28	COSTOS GASTOS TRX EXTERNAS_log	Continua	5465	8	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
29	PAGO SEGUROS SALARIOS PROPIOS_log	Continua	17	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0
30	PAGO SEGUROS SALARIOS TERCEROS_log	Continua	0	0	Napagus	Nunca	Eje	100.000	145184	0	0	0	0

3.2.3 Fase de preparación de los datos

La limpieza, generación de variables adicionales, integración de diferentes orígenes, en general, la adaptación de los datos a la técnica de minería, se direcciona directamente a un modelo que se enfoque al objetivo que es la clasificación de Sociedades en base a su **Importancia Fiscal**.

Contextualizando podemos citar:

Selección de datos.

El subconjunto de datos fue seleccionado en función a un consolidado de criterios expertos, basado tanto en encuestas como en entrevistas, Anexo 1, 2 y 3; cuyo objetivo principal es la construcción de un concepto de Importación Fiscal el mismo que se lo va a utilizar como guía para la clasificación de contribuyentes, Sociedades, en función a sus características.

Es decir se excluyó cada una de las variables que no se consideró parte de una **Importancia Fiscal**, enfocada directamente al Control Tributario es así como se llegó a seleccionar las siguientes:

- Prelación
- Regional
- Tipo
- Clase
- Código Actividad CIIU Nivel 6
- Pertenencia a Grupo Económico
- Activos (Declaración de renta 2012)
- Pasivos (Declaración de Renta 2012)
- Impuesto a la Renta Causado (Declaración de Renta 2012)
- Número de empleados

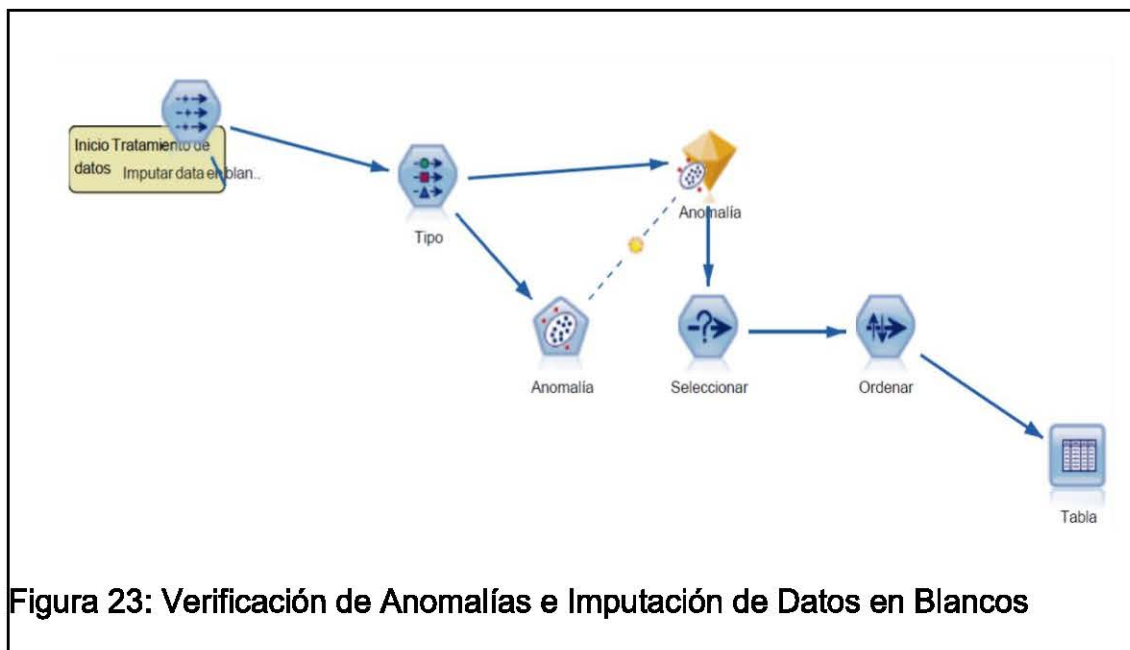
- De Transaccionalidad (2012):
 - Máximo Ingresos Internos (fuentes propias y de terceros)
 - Máximo Ingresos Externos (fuentes propias y SENA)
 - Máximo compras Internas Reportadas(fuentes propias y terceros)
 - Máximo compras externas
 - Máximo pagos por Sueldos y Salarios
 - Ranking de Transaccionalidad.


Limpieza de los datos.

La preparación de los datos que se obtuvo en la selección, se lo ha realizado de acuerdo a un análisis minucioso de cada una de las variables, su optimización se basa solamente en el tratamiento de valores vacíos puesto que no se encontró valores desfasados del rango normal, en comparación a la Transaccionalidad del Servicio de Rentas Internas.

La imputación de datos en blancos y la verificación de anomalías se lo realizaron con la herramienta de SPSS Modeler.

En la Figura 23 se visualiza el flujo construido para la verificación de anomalías y tratamientos de datos en blancos.



En la Figura 23 se puede divisar el nodo rellenar , el cual se lo ha aplicado a todos los campos que no poseían información y se encontraban vacíos, se determina que no se trata de inconsistencias puesto que algunas variables como por ejemplo, **Grupo Económico**, se encuentra sin datos porque la **Sociedad** no posee grupo económico, en este caso se ha colocado la palabra “no”, para evitar valores vacíos.

Por otro lado en la verificación de normalización, existen datos que de acuerdo a lo que nos indica la herramienta presentan anomalías, se los analiza detenidamente y se concluye que individualmente como variables independientes y en comparación de variables en otros casos, pueden estar alejados de la media común pero con respecto a su contexto general de acuerdo a todas las características del contribuyente **sociedades** la variable es un valor razonable que no cae en un valor anómalo.

En la Tabla 10 se visualiza el resultado los cuatro casos que la herramienta lo reporta como los anómalos más relevantes.

Tabla 10. Casos Anómalos reportados por la Herramienta

ITEM	PERTENECE_A_GRUPO_ECONOMICO	DESCRIPCION_PRELACION	DESCRIPCION_CLASE_CONTRIB	TOTAL_ACTIVO_F101	TOTAL_PASIVOS_F101	\$O-Anomaly	\$O-AnomalyIndex
1	NO	ESPECIAL	ESPECIAL	683.559.050	68.577.690	T	155.190
2	SI	CONTROL DE SL	OTROS	26.656.568.770	398.063.570	T	49.108
3	NO	CONTROL DE SL	OTROS	31.887.734.530	29.530.031.640	T	48.538
4	SI	CONTROL DE SL	OTROS	9.477.526.320	9.431.667.760	T	42.129

ITEM	\$O-Field-1	\$O-FieldImpact-1	\$O-Field-2	\$O-FieldImpact-2	\$O-Field-3	\$O-FieldImpact-3
1	DESCRIPCION_CLASE_CONTRIB	0.499	DESCRIPCION_PRELACION	0.499	PERTENECE_A_GRUPO_ECONOMICO	0.001
2	PERTENECE_A_GRUPO_ECONOMICO	0.763	TOTAL_ACTIVO_F101	0.236	TOTAL_PASIVOS_F101	0.000
3	TOTAL_PASIVOS_F101	0.653	TOTAL_ACTIVO_F101	0.343	PERTENECE_A_GRUPO_ECONOMICO	0.004
4	PERTENECE_A_GRUPO_ECONOMICO	0.890	TOTAL_PASIVOS_F101	0.076	TOTAL_ACTIVO_F101	0.034

Los casos anómalos que la herramienta presenta y que son visualizados en la Tabla 9, de acuerdo a la realidad se ha verificado que son valores correctos, que la Sociedad con ítem 1, de acuerdo al Listado de Contribuyentes Especiales Marzo 2014, es un contribuyente especial y no pertenece a grupo económico alguno de acuerdo a la declaración realizada, la misma evaluación se tuvo con los tres ítems siguientes, con lo cual se concluye que la data no posee valores anómalos.

Estructuración e Integración de los datos.

Al referirse a la estructuración e integración de los datos se puede citar que en el caso de la variable **tipo impositivo efectivo**, se debe realizar la operación; división, entre el impuesto a la renta causado sobre el Total de ingresos, por otro lado se generó un atributo derivado de la variable **Código de la actividad CIUU**, puesto que el ítem que necesitamos es el primer dígito, el cuál representa el clasificador determinado por las Naciones Unidas.

En la Tabla 11 se visualiza el clasificador determinado por las Naciones y utilizadas en el Ecuador.

Tabla 11. Código CIU, aplicado en el Ecuador

Código Sección	Descripción Sección
A	AGRICULTURA, GANADERIA, CAZA Y SILVICULTURA
B	PESCA
C	EXPLOTACION DE MINAS Y CANTERAS
D	INDUSTRIAS MANUFACTURERAS.
E	SUMINISTROS DE ELECTRICIDAD, GAS Y AGUA
F	CONSTRUCCION
G	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACION DE VEHICULOS AUTOMOTORES, MOTOCICLETAS, EFECTOS PERSONALES Y ENSERES
H	HOTELES Y RESTAURANTES
I	TRANSPORTE, ALMACENAMIENTO Y COMUNICACIONES
J	INTERMEDIACION FINANCIERA
K	ACTIVIDADES INMOBILIARIAS, EMPRESARIALES Y DE ALQUILER
L	ADMINISTRACION PUBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACION OBLIGATORIA
M	ENSEÑANZA
N	ACTIVIDADES DE SERVICIOS SOCIALES Y DE SALUD
O	OTRAS ACTIVIDADES COMUNITARIAS SOCIALES Y PERSONALES DE TIPO SERVICIOS
P	HOGARES PRIVADOS CON SERVICIO DOMESTICO
Q	ORGANIZACIONES Y ORGANOS EXTRATERRITORIALES
R	BAJO RELACION DE DEPENDENCIA SECTOR PRIVADO
S	BAJO RELACION DE DEPENDENCIA SECTOR PUBLICO
T	SIN ACTIVIDAD ECONOMICA - CIU

Formateo de los datos.

La reorganización de datos sin alterar su significado no fue necesario aplicarlo puesto que el origen de datos que se ingresó en la herramienta SPSS Modeler era una base de datos cuyos valores eran totalmente compatibles a la misma.

3.2.4 Fase de modelado

Las técnicas que se han seleccionado y los modelos más óptimos son aquellos que al ejecutarlos se han alineado con mayor coherencia a los objetivos de la presente investigación.

Contextualizando podemos citar:

Selección de la técnica de modelado.

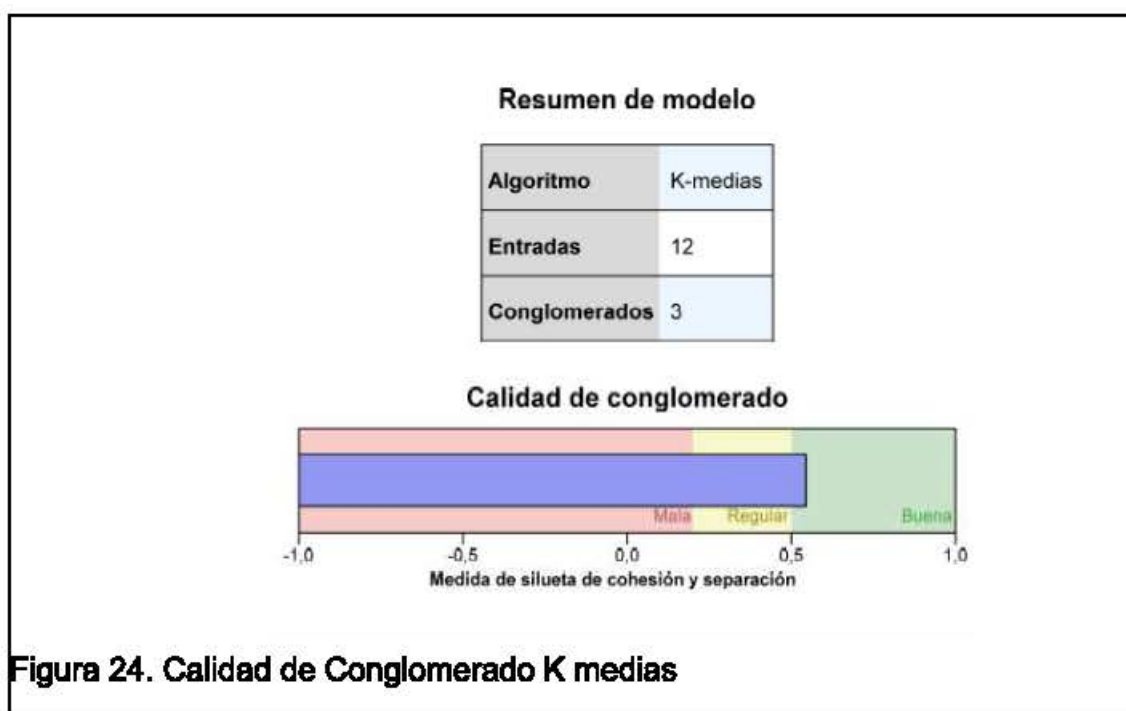
En el ítem 2.1.3.2 de la presente investigación, se puede visualizar cada uno de las técnicas de la minería de datos, que se podría aplicar.

Con relación a las mismas, y de acuerdo al objetivo que se desea llegar que es la clasificación de sociedades con relación a su **Importancia Fiscal**, se va a escoger algoritmos no supervisados y supervisados; no supervisados, porque se requiere encontrar patrones similares entre los casos Contribuyentes, Sociedades, que conforman el grupo y lo más diferentes posibles si los casos son parte de los diferentes grupos, con la finalidad de clasificarlos, con el mejor criterio posible alineados con los objetivos Institucionales. Adicionalmente se aplicará algoritmos supervisados, porque se desea que la clasificación obtenida sea la base de un aprendizaje para la distribución de nuevos casos.

En función a los algoritmos no supervisados se ha determinado que la técnica que se alinea mejor, es el algoritmo **K medias**, el cual al ser comparado con los demás algoritmos que nos facilita la herramienta, Kohonen, Vecinos más próximos, se ha verificado por análisis de la clasificación de la data, tomando como base las encuestas y entrevistas realizadas, que el algoritmo **K medias** se observa una mayor calidad en la conformación de los conglomerados donde se garantiza la homogeneidad intragrupo y heterogeneidad entre grupo, dando como resultado clusters o agrupaciones que se encuentran más acorde a la realidad y al objetivo planteado en la investigación.

Adicionalmente para la selección de la técnica anteriormente mencionada se ha evaluado su sensibilidad por medio de la herramienta SPSS Modeler.

En la Figura 24 se visualiza el resumen que la herramienta nos facilita, donde nos presenta el nombre del algoritmo utilizado **K medias**, el número de entradas y los grupos que se generó al aplicarlo, adicionalmente se visualiza la precisión del modelo que se ha escogido, para éste caso se encuentra entre 0,5 a 1, donde lo considera como buena.



Cada uno de los clusters, que son parte del resultado del algoritmo K medias, van a ser alineados a la tabulación de respuestas de la encuesta Anexo 1, con la finalidad de determinar que cluster representa la importancia alta, media y baja, con esto posteriormente se reconocerá, una variable objetivo que servirá como parte de un algoritmo supervisado, Arbol de Decisión C5, el mismo que se lo ha escogido en comparación del grado de exactitud que presenta en función a los otros algoritmos utilizados: discriminante, redes neuronales, regresión Multilogística con el cual se garantizará que el modelo posea entradas correctas para un posterior aprendizaje idóneo.

Generación del plan de prueba.

El diseño y prueba de la calidad y validez del modelo se lo va a referenciar al conjunto de respuestas tabuladas en la encuesta Anexo 1, con las cuales se realizará una comparación de los resultados del modelo con criterios consolidados de las encuestas realizadas a los expertos Tributarios, determinando de esa forma con mayor exactitud si la variable **Sociedad** es de importancia alta, media y baja.

Construcción del Modelo.

La construcción del modelo se direcciona específicamente a la extracción de la data desde una tabla, la cual fue construida con la concatenación de las diferentes consultas realizadas a la base de datos del Servicio de Rentas Internas, donde la selección de variables se basó netamente en criterios recopilados en las diferentes reuniones con los expertos tributarios, Anexo 3.

Del conjunto de casos de sociedades conformadas con las variables que se hace referencia, se ha tomado como muestra el 30%, los cuales constituyen los datos de entrada para la aplicación de la técnica de minería de datos K medias.

El 30% de muestra se considera suficiente para obtener niveles de confianza mayores al 99% y márgenes de error menores 1%, fórmula

$$\frac{k^2 Npq}{e^2(N-1) + k^2pq}$$
, el 30% representa una representatividad adecuada del total sociedades. (Torres, Paz, & Salazar, s.f.)

En la Figura 25 que se visualiza a continuación se observa el número de clusters resultado de la aplicación de la minería, k medias

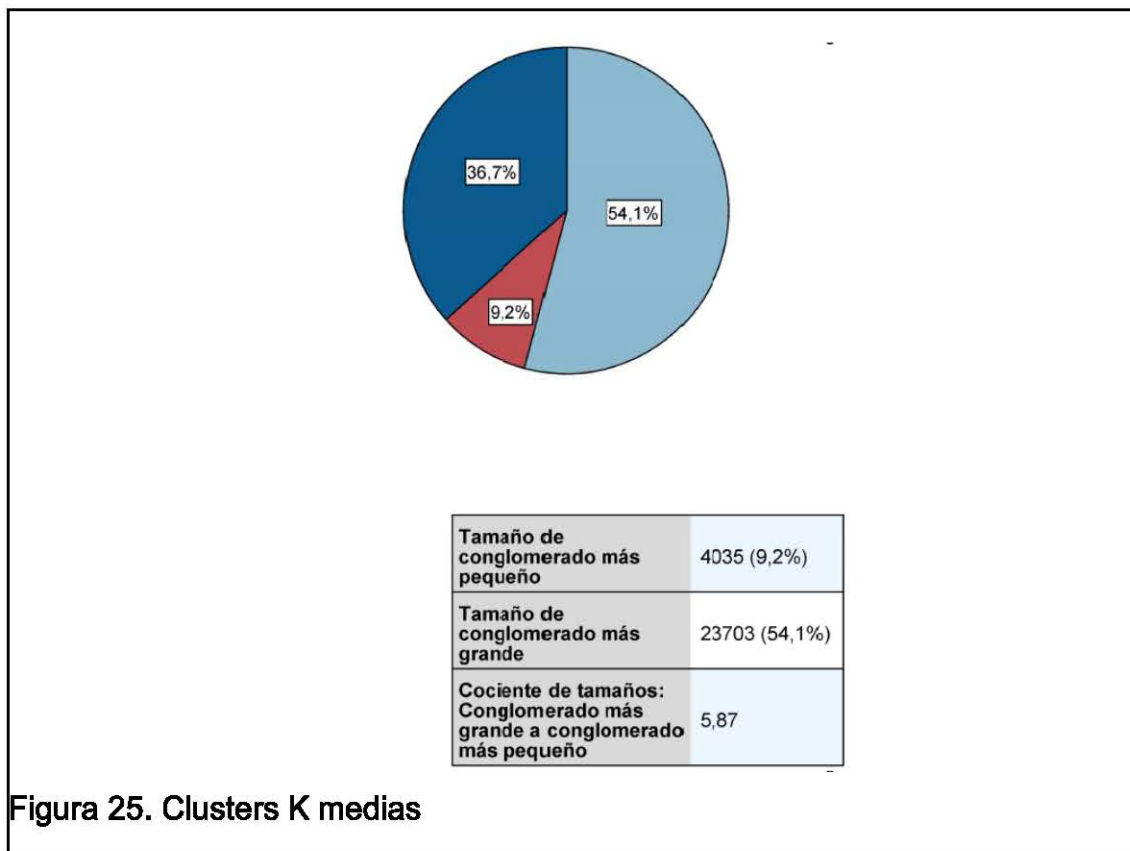


Figura 25. Clusters K medias

El número de clusters, que se ha escogido es 3 porque al colocar un número mayor no permite discriminar grupos con mayor exactitud, los resultados que se obtuvo se relacionó con las respuestas de la encuesta Anexo 2 con la finalidad de determinar qué cluster o grupo representa importancia alta, media o baja.

En la Tabla 12 que se visualiza a continuación se observa las medias de cada una de las variables donde se verifica claramente su distinción y su clasificación en importancia alta, media, baja.

Tabla 12. Variables y sus respectivas Medias

Campo	Importancia Fiscal Alta*	Importancia Fiscal Baja*	Importancia Fiscal Media*
TOTAL_ACTIVOS_F101	24847447.970	54243.446	624700.669
TOTAL_PASIVOS_F101	15180986.889	10990.170	360818.008
IMPUESTO_RENTA_CAUSADO_F101	543999.004	10.293	4372.292
INGRESOS_TRX_INTERNAS	23651808.864	8047.666	510600.408
INGRESOS_TRX_EXTERNAS	4511977.119	1143.542	29716.299
COSTOS_GASTOS_INTERNOS_PROPIOS	17484615.051	10908508.630	1003782.193
COSTOS_GASTOS_EXTERNOS_PROPIOS	3943157.322	106.362	66143.766
COSTOS_GASTOS_TRX_INTERNAS	20874044.111	11080922.791	1289855.868
COSTOS_GASTOS_TRX_EXTERNAS	4083462.324	341.458	67815.039
PAGO_SUELDOS_SALARIOS_PROPIOS	2713357.682	165039.620	252583.274
PAGO_SUELDOS_SALARIOS_TERCEROS	2156257.247	17072.466	70084.724
Presión Fiscal	0.011	0.002	0.012

Con la clasificación obtenida, Importancia alta, media y baja, se esquematiza la variable objetivo del nuevo algoritmo supervisado, donde el porcentaje de datos utilizados es de 67% para el entrenamiento y 33% para las pruebas, cuyas proporciones han sido seleccionadas en base a una media obtenida en proyectos similares que han sido ejecutados en el SRI junto con criterios expertos de funcionarios que han trabajado en la ejecución de los mismos.

Con relación a la selección de la técnica de minería de datos supervisada se ha realizado la comparación entre cuatro algoritmos con la finalidad de escoger el más Idóneo.

El análisis se lo ha realizado basándose en los resultados del nodo matriz de contingencias y en gráficos que calculan la precisión de los modelos predictivos.

En la Tabla 13 se visualiza el primer algoritmo a evaluar, Discriminante, el cual al ejecutarlo se observa los resultados en una matriz de contingencias, tanto para el fase de aprendizaje como la fase de prueba.

Tabla 13. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Discriminante

		Objetivo_IF		
SD-Objetivo_IF		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	369	4	27
	% de filas	92.250	1.000	6.750
Importancia Fiscal Baja	Recuento	431	16306	6810
	% de filas	1.830	69.249	28.921
Importancia Fiscal Media	Recuento	349	340	4490
	% de filas	6.739	6.565	86.696

Tabla 14. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Discriminante

		Objetivo_IF		
\$D-Objetivo_IF		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	177	0	15
	% de filas	92.188	0.000	7.812
Importancia Fiscal Baja	Recuento	205	7965	3487
	% de filas	1.759	68.328	29.913
Importancia Fiscal Media	Recuento	175	175	2160
	% de filas	6.972	6.972	86.056

En la Tabla 13 se puede visualizar la relación entre campos, el Objetivo_IF, es el real, mientras que el \$D-Objetivo_IF es el predicho, la intersección de ambos nos permite determinar el grado de precisión del modelo en forma cuantitativa y porcentual, con lo cual podemos concluir que el modelo en la matriz de entrenamiento posee un bajo porcentaje de precisión pues es un 69%, en menor proporción ocurre con la Importancia Fiscal Media 86.696, y un porcentaje aceptable para la Importancia Fiscal Alta 92 %; como era de suponerse en la matriz prueba posee similar comportamiento.

En la Figura 26 se visualiza la precisión del modelo para la variable Importancia Fiscal Alta, en la fase de entrenamiento y prueba, donde la línea de color lila es la idónea y la de color verde es la que nos presenta en el modelo.

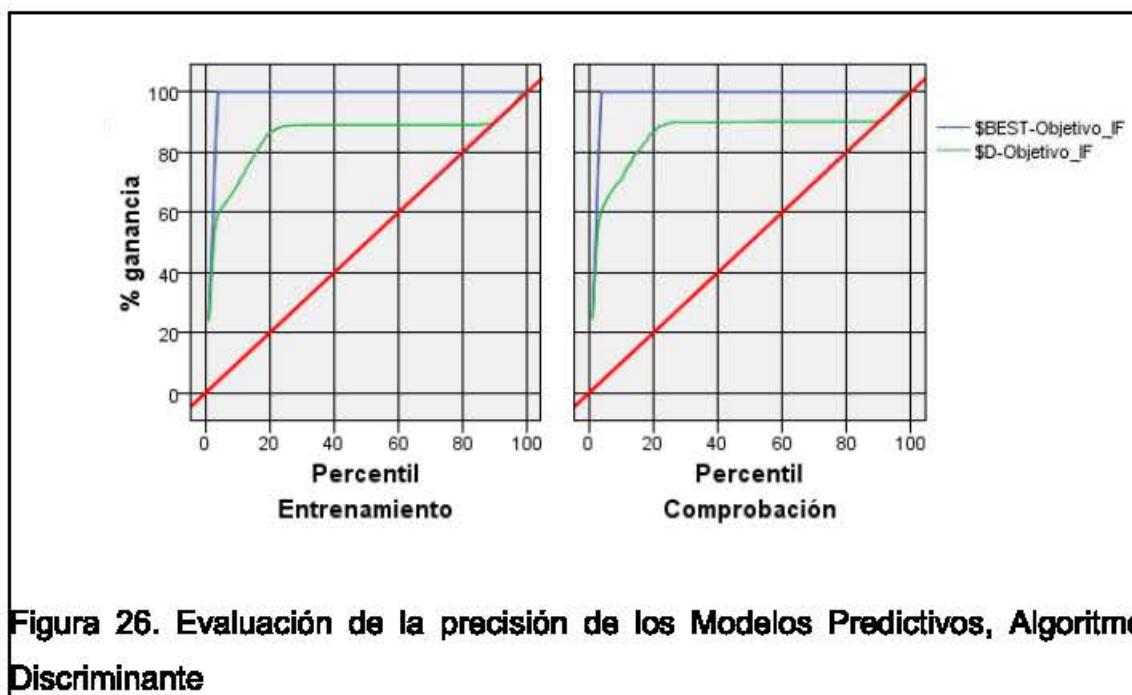


Figura 26. Evaluación de la precisión de los Modelos Predictivos, Algoritmo Discriminante

Un análisis parecido sucede con los algoritmos, Redes Neuronales y Multilogística donde los resultados no poseen una precisión idónea de acuerdo a lo que se requiere.

En la Tabla 14 y Figura 27 se puede visualizar tanto en forma cuantitativa, porcentual y gráfica los resultados de la precisión del Algoritmo Redes Neuronales.

Tabla 15. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Redes Neuronales

SN-Objetivo_IF		Objetivo_IF		
		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	555	2	1
	% de filas	99.462	0.358	0.179
Importancia Fiscal Baja	Recuento	0	7467	538
	% de filas	0.000	93.279	6.721
Importancia Fiscal Media	Recuento	2	671	5123
	% de filas	0.035	11.577	88.389

Tabla 16. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Redes Neuronales

		Objetivo_IF		
SN-Objetivo_IF		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	1137	0	5
	% de filas	99.562	0.000	0.438
Importancia Fiscal Baja	Recuento	0	15303	1065
	% de filas	0.000	93.493	6.507
Importancia Fiscal Media	Recuento	12	1347	10257
	% de filas	0.103	11.596	88.301

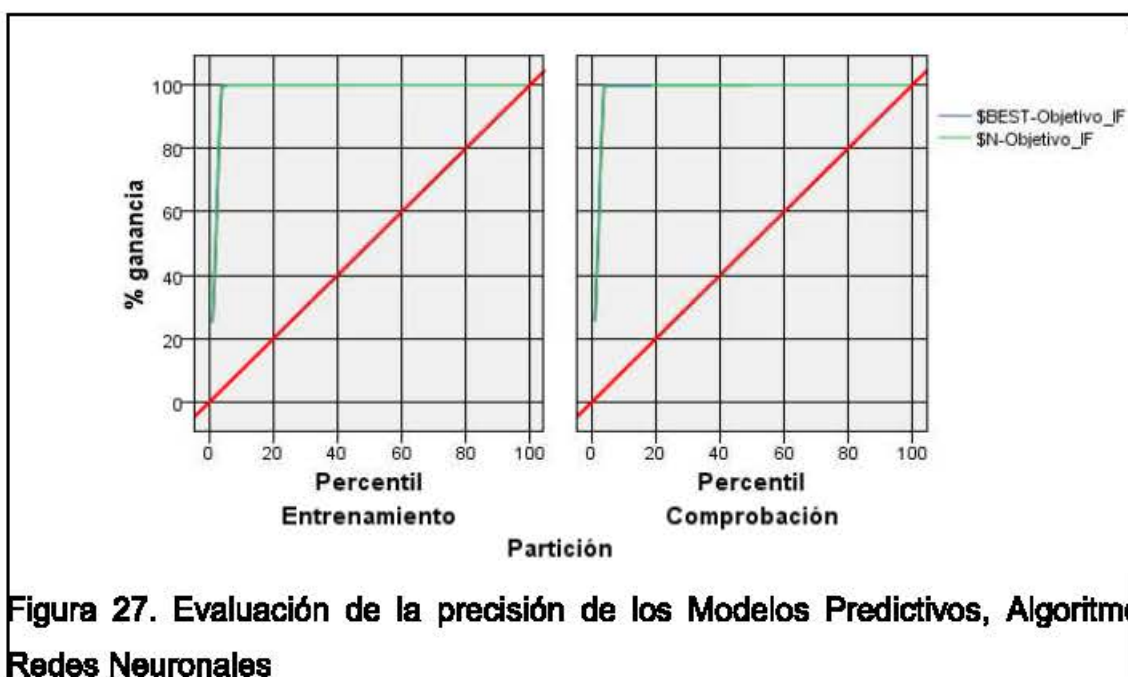


Figura 27. Evaluación de la precisión de los Modelos Predictivos, Algoritmo Redes Neuronales

En el modelo de Redes Neuronales, la Importancia Fiscal Media es 88% y alta es 99%, con lo que se puede concluir que más reconoce a los de importancia alta mientras que deja rezagados a los de Importancia Media.

Adicionalmente en la Tabla 15 y Figura 28 se puede visualizar tanto en forma cuantitativa, porcentual y gráfica los resultados de la precisión del Algoritmo de Regresión Multilogística.

Tabla 17. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo de Regresión

Objetivo_IF				
\$L-Objetivo_IF		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	1149	2	0
	% de filas	99.826	0.174	0.000
Importancia Fiscal Baja	Recuento	0	15230	1295
	% de filas	0.000	92.163	7.837
Importancia Fiscal Media	Recuento	0	1418	10032
	% de filas	0.000	12.384	87.616

Objetivo_IF				
\$L-Objetivo_IF		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	557	2	0
	% de filas	99.642	0.358	0.000
Importancia Fiscal Baja	Recuento	0	7449	638
	% de filas	0.000	92.111	7.889
Importancia Fiscal Media	Recuento	0	689	5024
	% de filas	0.000	12.060	87.940

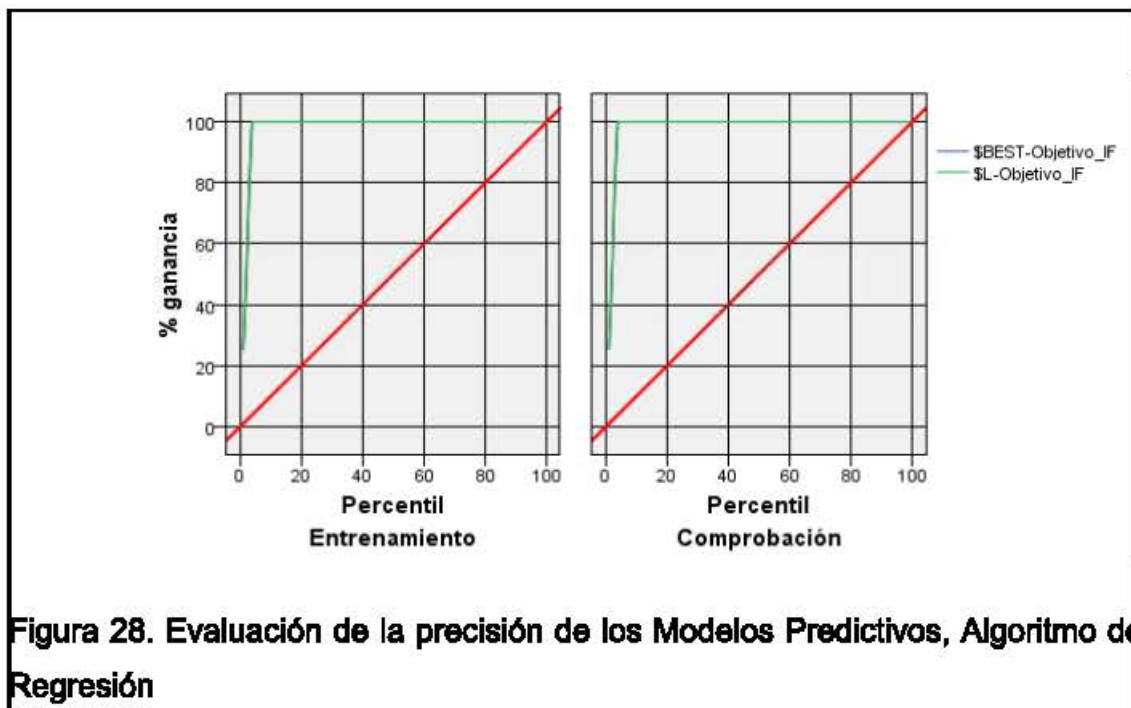


Figura 28. Evaluación de la precisión de los Modelos Predictivos, Algoritmo de Regresión

En el modelo Regresión Multilogística, la Importancia Fiscal Alta es 99% y la media es 87%, con lo que se puede concluir que más reconoce a los de Importancia Alta mientras que deja rezagados a los de Importancia Media, al igual que el algoritmo de Redes Neuronales.

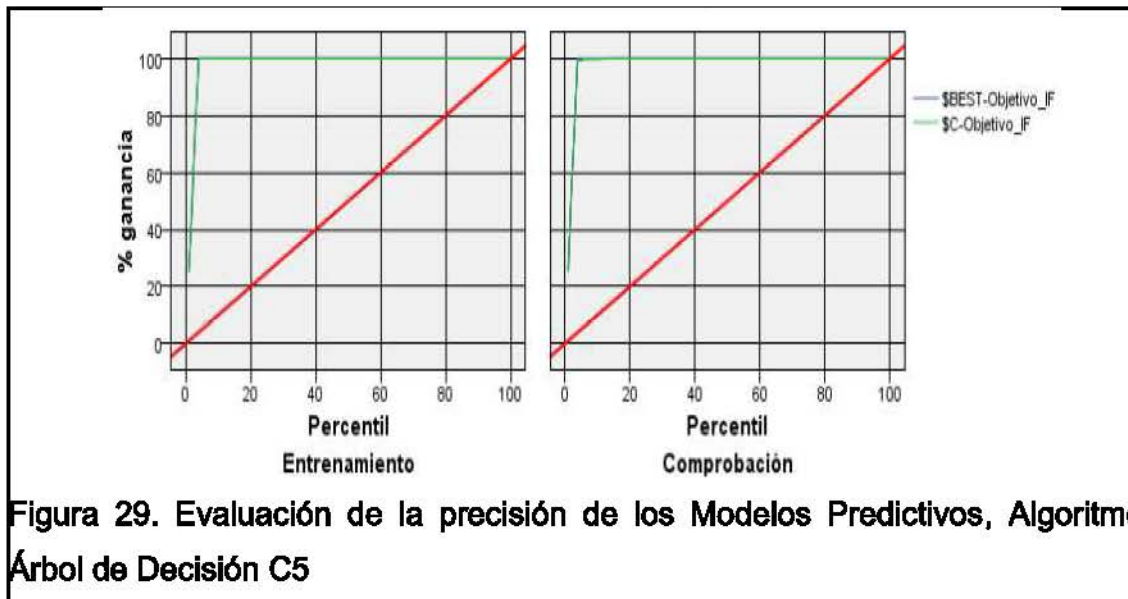
Finalmente citamos el algoritmo Árbol de Decisión C5, el cual se va a escoger por presentar mayor precisión tanto en el aprendizaje como en la fase de prueba.

En la Tabla 16 y Figura 29 se puede visualizar tanto en forma cuantitativa, porcentual y gráfica los resultados de la precisión del Algoritmo, Árbol de Decisión C5, el cual al ejecutarlo se observa los resultados en una matriz de contingencias, tanto para el fase de aprendizaje como la fase de prueba.

Tabla 16. Matriz de Contingencias: Entrenamiento y Prueba, Algoritmo Árbol de Decisión

		Objetivo_IF		
SC-Objetivo_IF		Import. Alta	Import. Baja	Import. Media
Importancia Fiscal Alta	Recuento	553	0	2
	% de filas	99.640	0.000	0.360
Importancia Fiscal Baja	Recuento	2	8014	132
	% de filas	0.025	98.355	1.620
Importancia Fiscal Media	Recuento	2	126	5528
	% de filas	0.035	2.228	97.737

		Objetivo_IF		
SC-Objetivo_IF		Import. Alta	Import. Baja	Import. M
Importancia Fiscal Alta	Recuento	1149	0	1
	% de filas	99.913	0.000	0.087
Importancia Fiscal Baja	Recuento	0	16552	127
	% de filas	0.000	99.239	0.761
Importancia Fiscal Media	Recuento	0	98	11199
	% de filas	0.000	0.867	99.133



El Árbol de Decisión C5, es el que se observa en la evaluación de su precisión con mayor exactitud, su importancia alta, media y baja tanto en entrenamiento como en prueba, se presenta sobre el 90%, lo que significa que el aprendizaje es muy bueno y no tiende solo por una opción si no que equilibra las tres opciones, lo que no ocurre con los demás algoritmos, es por éste motivo que se ha escogido ésta opción como la más idónea.

En la Figura 30 que se presenta a continuación se puede observar la estructura del modelo en forma completa.

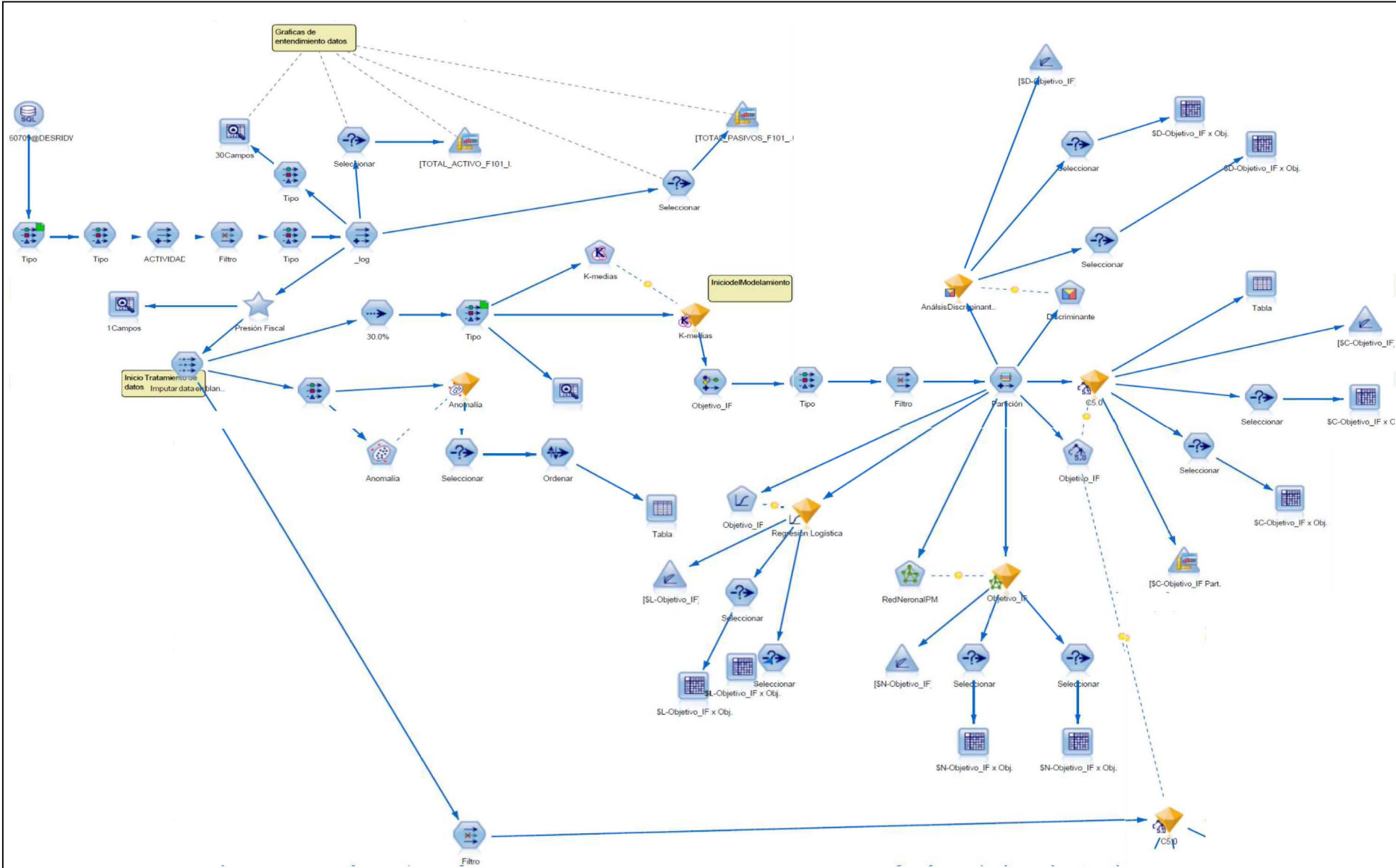


Figura 30 . Modelo Clasificador de Sociedades

Evaluación del modelo.

Con relación a la evaluación del modelo, y comparándolo con el criterio de éxito que nos hemos planteado inicialmente que es socavar información que apoye a la toma de decisiones para determinar el grado de Importancia de las Sociedades Activas, se puede citar que el árbol de decisión C5, es el que clasifica con mayor precisión a las sociedades, permitiendo la posibilidad de poseer con ayuda de un visor la manipulación de variables de acuerdo al objetivo que se desea alcanzar.

3.2.5 Fase de evaluación

Evaluación de los resultados.

Los resultados del modelo se alinean a los objetivos del negocio, puesto que el poseer un concepto estructurado de **Importancia Fiscal**, cuyas variables se puedan manipular de acuerdo a la necesidad, permitirá poseer una visión estratégica de clasificación de las Sociedades con la cual se pueda tomar decisiones de acuerdo a parámetros basados en información más detallada que apoye la asistencia y control, incremente la eficiencia operacional y en consecuencia aumente el uso eficiente del presupuesto.

Proceso de revisión.

Para la realización del modelo se hizo la revisión de cada una de las variables para verificar si poseían coherencia como parte de las entradas que se iban a utilizar para el algoritmo que se escogió al final del proceso. Se probó con diferentes modelos para seleccionar el más idóneo y en función a encuestas realizadas a expertos tributarios y a entrevistas personales se validó la coherencia de la realización del procedimiento y de sus respectivas respuestas.

En el ámbito de realizar mejoras al proceso planteado podemos citar, que se puede construir variables adicionales con ayuda de la gestión de información, donde con un análisis minucioso se podría buscar la forma de comparar el nuevo modelo con el que teníamos inicialmente y de esa manera encontrar nuevo conocimiento que apoye a la toma de las decisiones.

Adicionalmente con relación a la construcción de la encuesta, la misma puede ser más generalizada en función a los criterios expertos de las diferentes direcciones y no solo enfocada al control tributario, con la finalidad que la **Importancia Fiscal** sea estructurada de manera más global buscando alinearse con cada uno de los componentes que son parte de la institución Servicio de Rentas Internas.

Determinación de futuras fases.

Cuando nos referimos a futuras fases si nos enfocamos a la parte estratégica, la presente investigación va a ser una guía para poder aplicar una Gestión Tributaria más direccionada para toma de decisiones en el ámbito del Control.

Adicionalmente en función a los resultados se puede concluir que los mismos se alinean en su totalidad al objetivo que se deseaba alcanzar con lo que no es necesario repetir alguna iteración, pues se posee un visor el cual permite manipular las variables de acuerdo a la necesidad.

Entregables de la Metodología

Entre los entregables tenemos: la estructuración de un concepto de Importancia Fiscal y la esquematización de un Marco de Trabajo.

Con relación a la estructuración de la Importancia Fiscal, se lo ha enfocado como resultado de las entrevistas realizadas a los Directores y Expertos Tributarios seleccionados para la investigación, Anexo 3, y la tabulación de

respuestas a las encuestas realizadas en el Anexo 2 alineado específicamente al control, el cuál es uno de los cuatro pilares que constituye el Modelo Integral de Gestión Estructural del Servicio de Rentas Internas, los otros pilares a los que se hace referencia son los que se cita en la Figura 1, y cuya descripción se indica en el capítulo 2.1.1.

Bajo ésta premisa se puede deducir que para estructurar un concepto de Importancia Fiscal se lo puede analizar en diferentes aristas como: recaudación, control, conciencia Fiscal e información.

Centrándose específicamente al control, direccionado solo a sociedades, se puede citar que la Importancia Fiscal en el Servicio de Rentas Internas es: Aquella que basándose en los objetivos estratégicos de la Institución, consolida criterios que permite determinar la priorización de las sociedades para fines de Control Tributario.

Por otro lado con relación a la esquematización de un Marco de Trabajo, el cuadro sinóptico que se cita a continuación permite visualizar las principales etapas que se realizó para la aplicación de una minería de datos, en el Servicio de Rentas Internas, en la cual se resalta la estructuración de un concepto de **Importancia Fiscal** lo cual fue la fase de iniciación del proyecto es por esto que se lo ha clasificado como fases preliminares, porque el mismo es indispensable para que se pueda realizar la clasificación de Sociedades.

A continuación se cita una fase de validación de políticas institucionales, la cual se ha propuesto, pues el desconocimiento de la misma puede ser un obstáculo para el proceso o culminación del proyecto.

La siguiente Fase son los acuerdos a los que se llega con la Institución, que son cada uno de los compromisos mutuos que son necesarios para que el proyecto camine.

Por último la Fase de desarrollo, capacitación e implementación son necesarias para que el proyecto culmine con un buen resultado donde el usuario cumpla con todas sus expectativas iniciales.

En la Figura 31 que se presenta a continuación se puede observar la estructura del Marco de Trabajo.

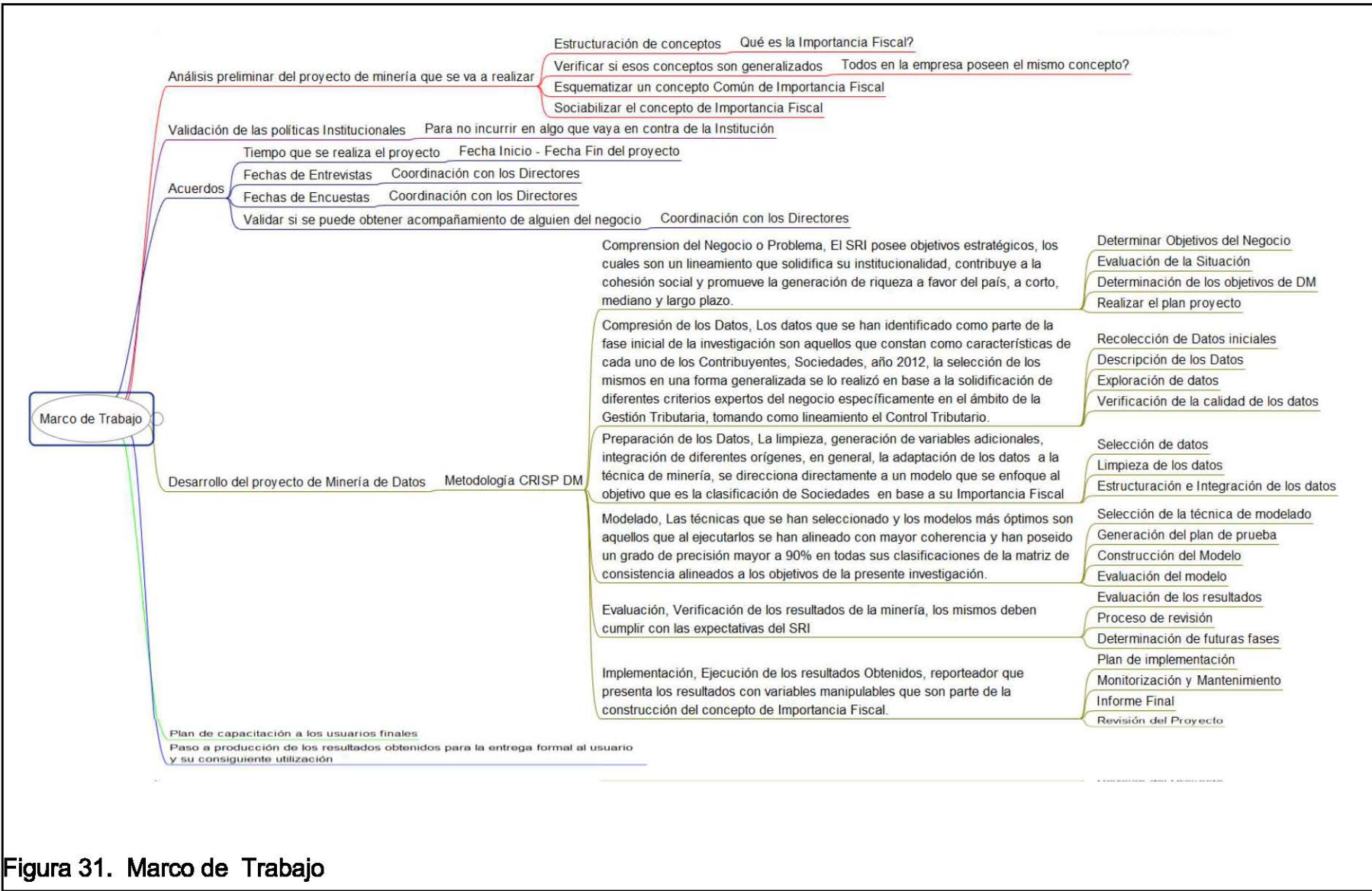


Figura 31. Marco de Trabajo

Para la manipulación de resultados de acuerdo a la necesidad, se utilizó, Saiku Reporting, el cual es una herramienta de Business Intelligence (Inteligencia de Negocios) que permite visualizar diferentes reportes que son esenciales de acuerdo a la toma de las decisiones, que desea realizar, en función a la necesidad del negocio.

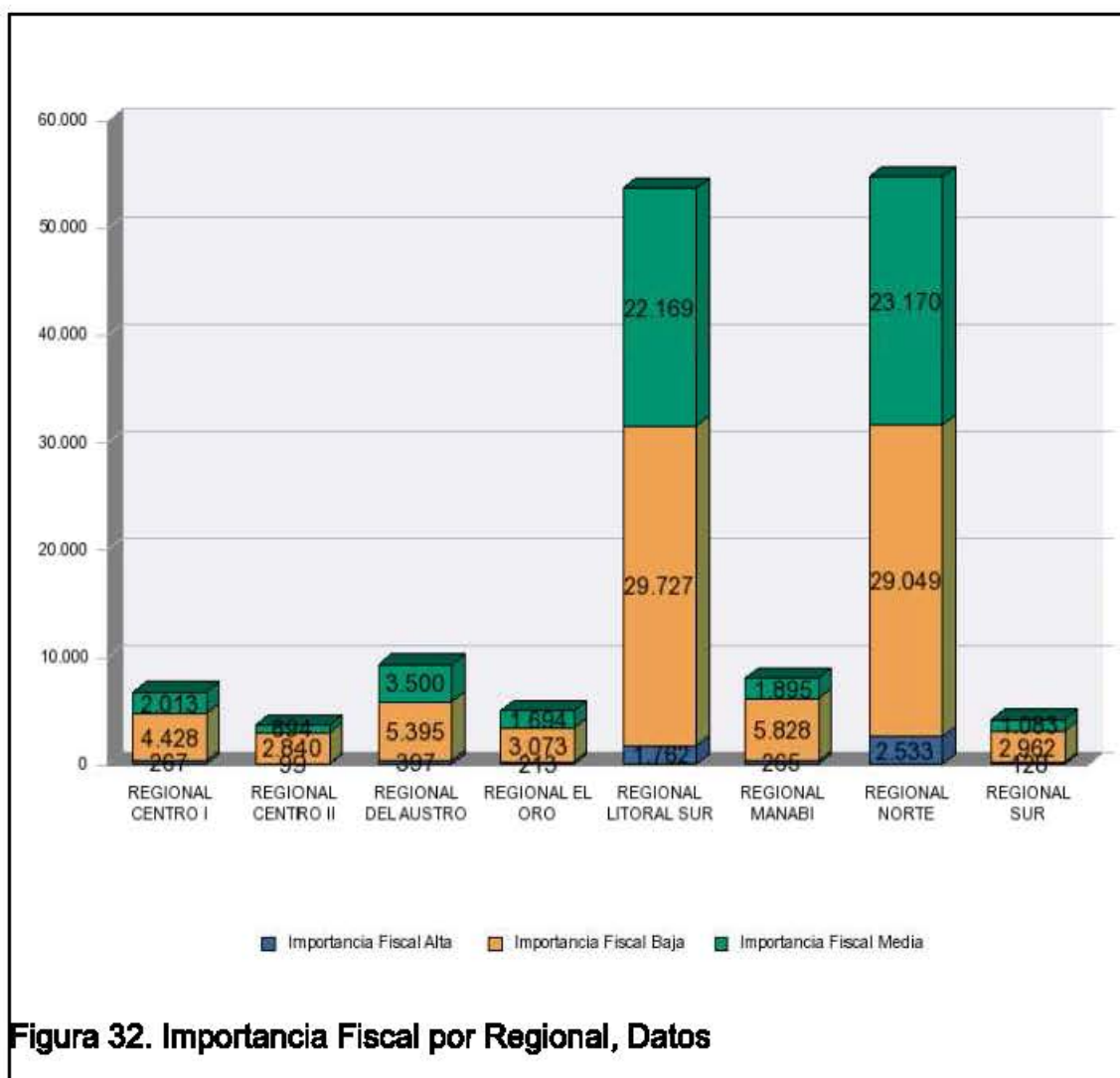
En la Tabla 17 y Figura 32 se puede visualizar la Importancia Fiscal que se presenta con relación a cada Regional.

Tabla 19. Importancia Fiscal por Regional

	Importancia Fiscal Alta	Importancia Fiscal Baja	Importancia Fiscal Media	Suma:
REGIONAL CENTRO I	18	259	133	410
REGIONAL CENTRO II	24	492	163	679
REGIONAL DEL AUSTRO	99	852	801	1.752
REGIONAL EL ORO	50	444	294	788
REGIONAL LITORAL SUR	61	387	255	703
REGIONAL MANABI		2	2	4
REGIONAL NORTE	41	467	150	658
REGIONAL SUR	1	2	3	6
Suma:	294	2.905	1.801	5.000

De acuerdo a los resultados la Regional Norte es la que representa la Importancia Fiscal Alta en mayor cantidad le sigue la Regional Litoral Sur y la Regional del Austro.

Se puede citar que los resultados obtenidos en la Tabla 16, se alinean correctamente a la recaudación del 2012, documento RECAUDACIÓN POR TIPO DE IMPUESTO, PROVINCIA, CANTÓN, MES Y AÑO, donde la provincia con mayor recaudación fue Quito con \$ 5 959 686 451 (55%), la cual pertenece a la Regional Norte, adicionalmente la provincia que le sigue es \$ 3 280 185 144, 8 (30%) de un total de recaudación de 10 788 923 463, 9 (100%). (Servicio de Rentas Internas, 2014)



En función al gráfico citado, con el enfoque de Importancia Fiscal Alta, se puede apoyar en el mismo para direccionar con más precisión las sociedades que tendrán prioridad para entablar una gestión tributaria dirigida al control.

Adicionalmente en la Figura 33 se puede visualizar el gráfico **Importancia Fiscal por Regional**, con datos porcentuales.

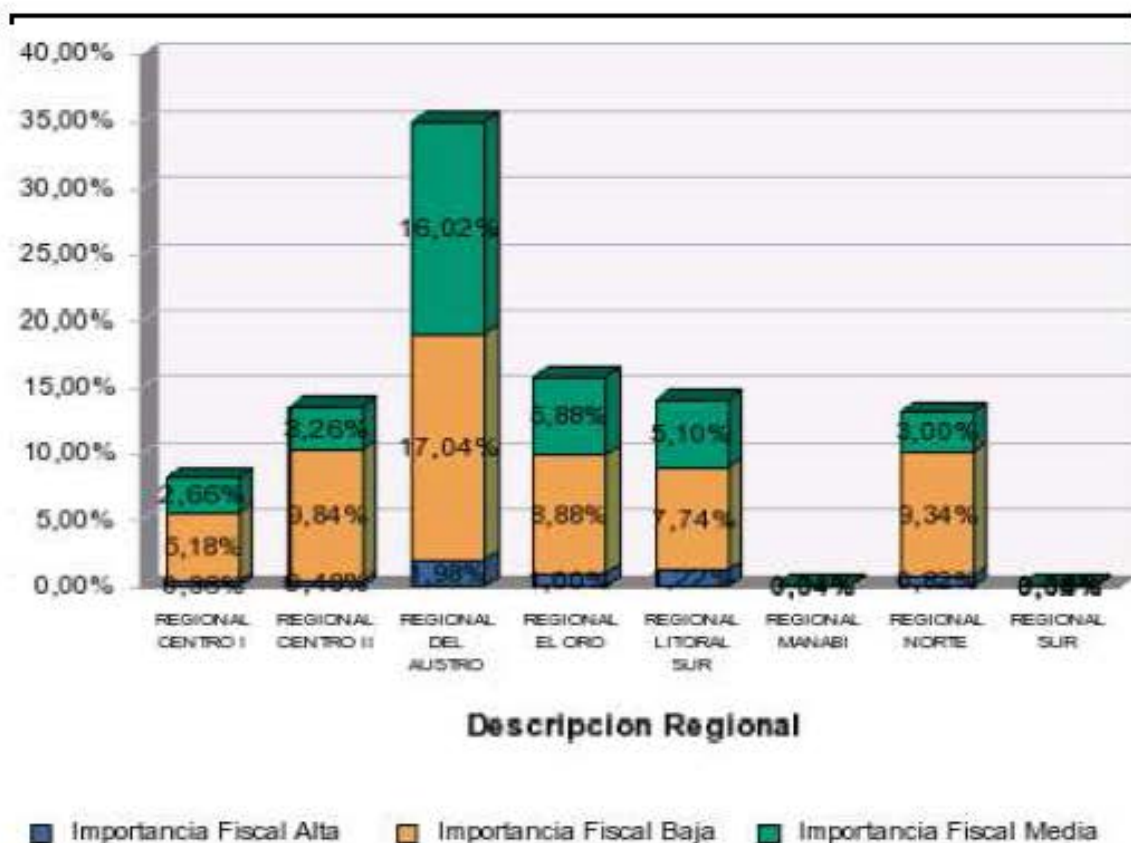


Figura 33. Importancia Fiscal por Regional, Porcentual

En la Figura 33 se puede verificar claramente que el porcentaje más alto en Importancia Fiscal Alta se presenta en la Regional Norte 1,74%.

Con ayuda del visor podemos también enfocar la Importancia Fiscal con relación a la actividad.

En la tabla 18 se puede visualizar la información que se presenta al tabular las variables de Importancia Fiscal vs. El sector económico.

Tabla 20. Importancia Fiscal con relación a su Actividad Económica

	Importancia Fiscal Alta	Importancia Fiscal Baja	Importancia Fiscal Media	Suma:
R		0,00%	0,00%	0,00%
A	0,18%	1,56%	1,51%	3,25%
B	0,07%	0,44%	0,55%	1,06%
C	0,08%	0,70%	0,35%	1,13%
D	0,62%	1,94%	2,63%	5,19%
E	0,02%	0,15%	0,06%	0,24%
F	0,17%	3,30%	2,83%	6,29%
G	1,41%	7,71%	8,77%	17,89%
H	0,05%	0,56%	0,71%	1,33%
I	0,19%	4,14%	4,00%	8,33%
J	0,16%	2,47%	1,72%	4,34%
K	0,33%	12,41%	9,67%	22,40%
L	0,33%	2,94%	0,97%	4,23%
M	0,08%	2,00%	1,17%	3,25%
N	0,12%	3,13%	1,32%	4,56%
O	0,10%	13,85%	2,45%	16,40%
P			0,00%	0,00%
Q	0,00%	0,05%	0,02%	0,07%
T		0,02%		0,02%
Suma:	3,90%	57,38%	38,72%	100,00%

En la tabla 18 se puede verificar que la actividad que influye más en la **Importancia Fiscal Alta**, es el literal G, Comercio al por Mayor y al por Menor, preparación de Vehículos Automotores, Motocicletas y Enseres.

Donde el significado de cada letra se ilustra en el siguiente Gráfico.

Tabla 21. Letra de la Actividad con su respectivo significado

Codigo Sección	Descripcion Sección
A	AGRICULTURA, GANADERIA, CAZA Y SILVICULTURA
B	PESCA
C	EXPLORACION DE MINAS Y CANTERAS
D	INDUSTRIAS MANUFACTURERAS.
E	SUMINISTROS DE ELECTRICIDAD, GAS Y AGUA
F	CONSTRUCCION
G	COMERCIO AL POR MAYOR Y AL POR MENOR; REPARACION DE VEHICULOS AUTOMOTORES, MOTOCICLETAS, EFECTOS PERSONALES Y ENSERES
H	HOTELES Y RESTAURANTES
I	TRANSPORTE, ALMACENAMIENTO Y COMUNICACIONES
J	INTERMEDIACION FINANCIERA
K	ACTIVIDADES INMOBILIARIAS, EMPRESARIALES Y DE ALQUILER
L	ADMINISTRACION PUBLICA Y DEFENSA; PLANES DE SEGURIDAD SOCIAL DE AFILIACION OBLIGATORIA
M	ENSEÑANZA
N	ACTIVIDADES DE SERVICIOS SOCIALES Y DE SALUD
O	OTRAS ACTIVIDADES COMUNITARIAS SOCIALES Y PERSONALES DE TIPO SERVICIOS
P	HOGARES PRIVADOS CON SERVICIO DOMESTICO
Q	ORGANIZACIONES Y ORGANOS EXTRATERRITORIALES
R	BAJO RELACION DE DEPENDENCIA SECTOR PRIVADO
S	BAJO RELACION DE DEPENDENCIA SECTOR PUBLICO
T	SIN ACTIVIDAD ECONOMICA - CIU

Adicionalmente los datos se los puede representar gráficamente como se indica en la Figura 34.

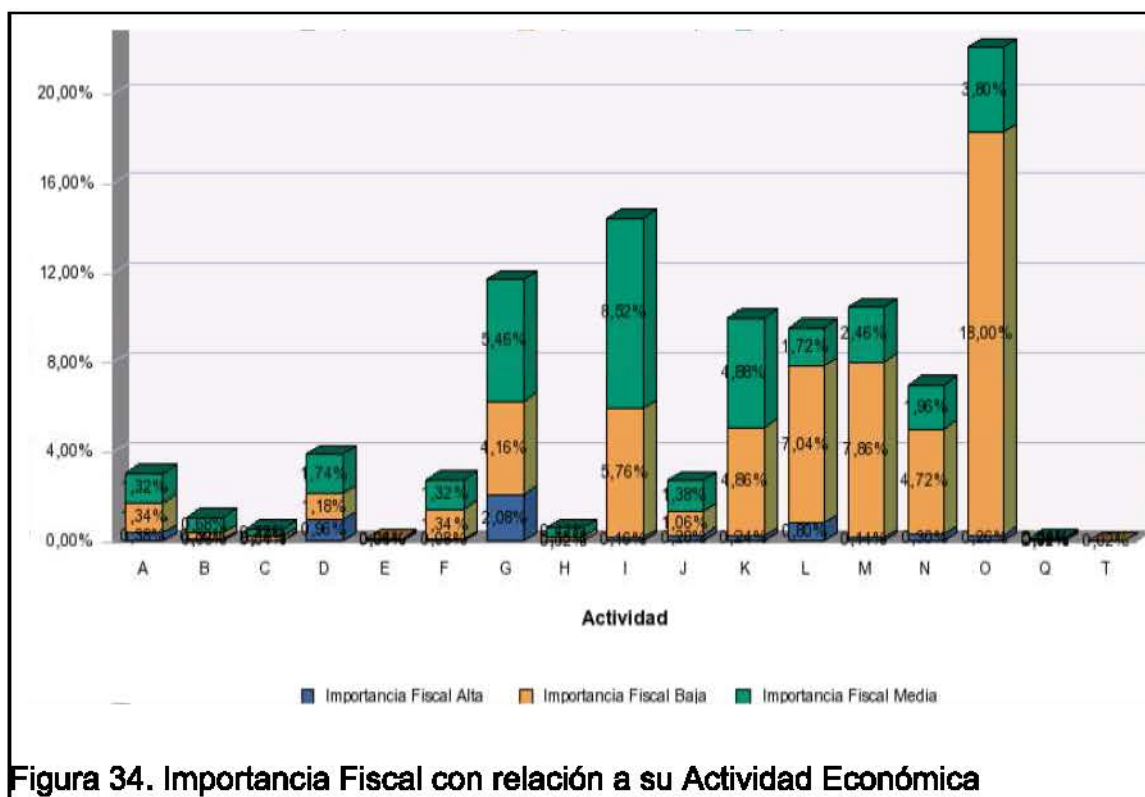


Figura 34. Importancia Fiscal con relación a su Actividad Económica

Finalmente, se nota claramente que la actividad que influye más en la Importancia Fiscal Alta es el literal G, que representa el 1,41%, Comercio al por Mayor y al por Menor, preparación de Vehículos Automotores, Motocicletas y Enseres.

Por otro lado caber recalcar, que la herramienta utilizada posee una gran facilidad para realizar reportes de acuerdo a la necesidad y lo que se desea validar (Ad Hoc).

Su alta granularidad permite conocer específicamente al detalle la Sociedad y su respectiva clasificación, dando paso a un análisis trascendental para la toma de las decisiones con mayor probabilidad de acertividad.

CAPÍTULO IV:

CONCLUSIONES Y RECOMENDACIONES

4.1 Conclusiones

- Con apoyo de la metodología CRISP – DM, se identificó tres grupos de contribuyentes con referencia a la Importancia Fiscal de Sociedades Activas 2012, adicionalmente el análisis de la data al aplicar el algoritmo K-medias se deduce que las medias de las variables que intervienen en cada caso con mayor influencia en la **Importancia Fiscal Alta** poseen las siguientes características: Total Activo \$24 847 447. 970, Total Pasivos \$15 180 986. 889, Impuesto Renta Causado 543 999. 004, Ingresos transacciones Internas \$23 651 808. 864, Ingresos Transacciones Externas \$4 511 977. 119, Costos Gastos Internos Propios \$17 484 615, 051, Costos Gastos Externos Propios 3 943 157, 322, Costos Gastos Transacciones Internas \$20 874 044. 111, Costos Gastos Transacciones Externas \$4 083 462. 324, Pago Salarios Propios \$2 713 357. 682, Pago Salario Terceros \$2 156 257. 247, Presión Fiscal 0,011; por otro lado en la **Importancia Fiscal Media** poseen las siguientes características: Total Activo \$624 700. 669, Total Pasivos \$360 818. 008, Impuesto Renta Causado 4 372. 292, Ingresos transacciones Internas 510 600. 408, Ingresos Transacciones Externas 29 716. 299, Costos Gastos Internos Propios 1 003 782, 193, Costos Gastos Externos Propios 66 143. 766, Costos Gastos Transacciones Internas \$1 289 855. 868, Costos Gastos Transacciones Externas \$6 7815. 039, Pago Salarios Propios 252 583. 274, Pago Salario Terceros 70 084. 784, Presión Fiscal 0. 012; Finalmente en la **Importancia Fiscal Baja** poseen las siguientes características: Total Activo \$54 243. 446, Total Pasivos \$1 090. 170, Impuesto Renta Causado 10. 293, Ingresos transacciones Internas 8 047. 666, Ingresos Transacciones Externas 1 143. 542, Costos Gastos Internos Propios 10 908 508. 630, Costos Gastos Externos Propios 106. 362, Costos Gastos

Transacciones Internas 11 080 922. 791, Costos Gastos Transacciones Externas 341. 458, Pago Salarios Propios 165 039. 620, Pago Salario Terceros 17 072. 466, Presión Fiscal 0. 002.

- La valoración, depuración, explotación de variables y criterios de análisis más idóneos se han realizado con la ayuda de la herramienta SPSS Modeler, con apoyo de criterios sustraídos en encuestas y las diferentes reuniones con los expertos tributarios, con las cuales se obtuvo una selección de variables acorde a la realidad, que se han basado en un algoritmo de clusterización que demostró que los grupos resultantes son diferentes y admiten tener segmentación de contribuyentes, Sociedades, con homogeneidad intragrupo y heterogeneidad entre grupo, que guían el enfoque óptimo de Acciones de Gestión para el Servicio de Rentas Internas.
- La determinación y segmentación de los grupos de interés se lo ha realizado con ayuda de técnicas de minerías de datos seleccionadas en base a una comparación de los resultados de la matriz de contingencias y gráficos, las cuales, calculan la precisión de los modelos predictivos donde el resultado con mayor exactitud tanto en el entrenamiento como en la prueba fue el obtenido por el algoritmo Árbol de Decisión C5, donde la clasificación de Importancia Fiscal Alta, dio como resultado 99,9% para el entrenamiento y 99,640% para la prueba; en el caso de la Importancia Fiscal Media, 99,133% para el entrenamiento y 97,737% para la prueba; y en el caso de Importancia Fiscal Baja fue 99,239 para el entrenamiento y 98,355% para la prueba es decir que en comparación a los algoritmos Redes Neuronales, Regresión Multilogística, Discriminante, los cuales presentaban resultados sesgados, el algoritmo que más se alinea al objetivo y posee más precisión es Árbol de Decisión C5.
- La aplicación de la metodología CRISP-DM, ayudó a estructurar a

detalle cada una de las instancias del proceso de minería de datos que se utilizó para la clasificación de Sociedades basado en la Importancia Fiscal, evitando la pérdida del objetivo, optimizando la comprensión y preparación detallada de los datos, abstrayendo de una forma ágil e iterativa el modelo más idóneo que se alineó más a la necesidad del negocio.

- El marco de trabajo planteado junto con la metodología aplicada consolidó un ahorro en tiempo y esfuerzo pues los mismos se alinearon de acuerdo a las políticas, procesos y procedimientos del Servicio de Rentas Internas, evitando re-procesos que obstaculizan el avance del proyecto de minería de datos.
- Se ha viabilizado una herramienta visor, que permite parametrizar las diferentes variables de la presente investigación de acuerdo al análisis que se requiere realizar, determinando cruces de información de Importancia Fiscal con la granularidad que se necesite, con tiempos cortos de obtención de datos.
- La estructuración de un Marco de Trabajo y el ensayo de la metodología CRISP DM Modeler enfocada a la construcción del concepto de Importancia Fiscal, direccionada al Control, se encuentra enfocada directamente a las variables que conforman el Concepto de Importancia Fiscal, pues las mismas son la base de la investigación y si no se encuentran correctamente parametrizados, los resultados se desfazarán de los objetivos institucionales.
- De acuerdo al análisis costo - beneficio se puede visualizar el ahorro significativo en recursos humanos para la obtención de reportes, actualmente se necesita de varios actores, analista que recolecta la información, el experto que valida y genera el reporte y el responsable quién es el que autoriza, con apoyo de la herramienta y

su correcta aplicación metodológica y el marco de referencia propuesto en la presente tesis se minimizaría a una persona, el Analista.

- Los datos al ser depurados, canalizados, verificados y explotados se han convertido en información que al procesarla son un soporte importante y vital para la toma de decisiones.
- La Minería de Datos en empresas que tengan madurez en cuanto a datos históricos, limpios, organizados y procesados, con objetivos claros y analistas con conocimiento del área, es fundamental para que el proyecto alcance su objetivo.
- Aplicar minería de datos junto a una metodología, es sumamente importante pues se adquiere una estrategia y dirección sistemática para realizar las actividades y procesos de acuerdo a estándares establecidos y propuestos, con lo cual la factibilidad del proyecto de minería de datos se incrementa.
- A partir de la segmentación propuesta, identificando y reconociendo los diferentes grupos de Importancia Fiscal, se pueden direccionar las diferentes estrategias que posee el Servicio de Rentas Internas en función a la Recaudación de tributos y consolidación de la cultura tributaria.

4.2 Recomendaciones

- Complementar la investigación realizada en el SRI, con variables adicionales que aporten la construcción del concepto de Importancia Fiscal enfocada al control, las cuales a pesar de ser detectadas no han sido tomadas en cuenta actualmente debido a la disponibilidad de la información para obtenerlas, ya sea por su ubicación o porque simplemente no se posee un dato específico de las mismas, esto sería necesario para darle una orientación más globalizada de la construcción del concepto.
- Organizar, Canalizar, verificar y explotar datos se debe realizar de forma periódica en el Servicio de Rentas Internas, porque es una Institución que posee un conglomerado de información que al procesarla se convertiría en un soporte importante y vital en la toma de decisiones.
- Ampliar el concepto de Importancia Fiscal en el Servicios de Rentas Internas para abarcar criterios de un mayor número de componentes y no solo basarse en el componente de riesgo alineado al control. Otro parámetro podría ser el riesgo alineado a la recaudación, conciencia fiscal e
- información, con ello se puede evaluar en diferentes aristas la construcción de la Importancia Fiscal para cada uno de las direcciones alineándose siempre con los objetivos institucionales, con la finalidad de direccionar las diferentes estrategias con mayor exactitud.
- Actualizar el modelo creado basado en el Servicio de Rentas Internas, con una periodicidad anual con el fin de representar la naturaleza dinámica de los Contribuyentes y de la Economía.

- Seleccionar una herramienta de minería de datos acorde a las necesidades es obligatorio puesto que se va a trabajar con la misma todo el tiempo del proyecto, hay diferentes herramientas para trabajar con minería de datos, la selección se debe basar en productos reconocidos por estudios realizados, una consultora reconocida a nivel mundial y que puede ayudar con la selección del producto de software es Gartner, empresa consultora, que posee una representación gráfica basada en la investigación de la situación del mercado de un producto tecnológico en un momento determinado.
- Analizar y Verificar la técnica de minería de datos que se va a utilizar es indispensable que se lo haga previamente antes de la selección de la herramienta, porque puede darse el caso que no exista la técnica que se desea aplicar dentro del Software/herramienta selección.

REFERENCIAS

- Alink, M., & Van Kommer, V. (Julio de 2000). Manual para las Administraciones Tributarias del CIAT. *Estructura Organizacional y Gerencia de las Administraciones Tributarias 2000*. 2000 Ministerio de Finanzas, Países Bajos.
- Anónimo. (s.f.). Obtenido de <http://cmapspublic.ihmc.us/rid=1LZ7SDQ4Z-185JDN9-24QH/Mineria%20de%20Datos.cmap>
- Bryant, A. D., Krzanich, B. M., & James, R. J. (11 de Noviembre de 2013). *Nextgeneration center Intel*. Obtenido de Nextgeneration center Intel: http://www.tacticsoftware.com/CRM-CURSO/Business_Intelligence.pdf
- Carrasco, C. M. (4 de abril de 2011). Definición, Alcance y Objetivos del concepto de Moral Tributaria. Quito, Pichincha, Ecuador.
- Centro de Estudios Fiscales. (27 de Octubre de 2013). *CEF*. Obtenido de CEF: <https://cef.sri.gob.ec/virtualcef/mod/book/view.php?id=1145&chapterid=846>
- Deinsa. (2013). *Cuadro de Mando Integral*. Obtenido de Introducción: <http://www.deinsa.com/cmi/index.htm?MenuState=AXcAAAEABBAQ77SD77SKAEAAQe-0gxVU77SF77SD77STw7BVFUBxFAHvtJM=>
- Embajada del Ecuador. (2014). *Plan Nacional del Buen Vivir 2013 - 2017*. Obtenido de <http://www.embassyecuador.eu/site/index.php/es/noticias-2/216-plan-nacional-del-buen-vivir-2013-2017>
- Franganillo, J. (2009). Implicaciones "éticas de la minería de datos". 320-324.
- Gartner. (2013). *Gartner IT Glossary* . Obtenido de <http://www.gartner.com/it-glossary/data-mining/>
- Giraldo, J. A. (s.f.). *Minería de Datos*. Obtenido de Minería de Datos: <http://ingesystem.galeon.com/index.html>
- IBM. (1994, 2011). Manual de minería interna de bases de datos de IBM SPSS Modeler 14.2.
- Impacto de Notificaciones sobre la Omisidad. (22 de Julio de 2011). *Evidencia de un Experimento de campo en el SRI(1)*. Quito, Pichincha, Ecuador.

- Instituto de Estadísticas de la Rioja. (Octubre de 2010). Business Intelligence en la Estadística. Obtenido de <http://www.jecas.org/ponencias/jueves/tarde/desarrollosII/businessintelligenceLaRioja.pdf>
- Jaramilo Villegas, E. J. (01 de Diciembre de 2013). <http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060024/Lecciones/Capitulo%20I/conceptos.htm>. Obtenido de <http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060024/Lecciones/Capitulo%20I/conceptos.htm>: <http://www.virtual.unal.edu.co/cursos/sedes/manizales/4060024/Lecciones/Capitulo%20I/conceptos.htm>
- Kaplan, R., & Norton, D. (2000). Cuadro de Mando Integral (The Balanced Scorecard).
- KDnuggets. (Agosto de 2007). KDnuggets : Polls : Data Mining Methodology (Aug 2007). Obtenido de KDnuggets : Polls : Data Mining Methodology (Aug 2007): http://www.kdnuggets.com/polls/2007/data_mining_methodology.htm
- Leyes, Derecho y Legalidad. (5 de Mayo de 2008). Obtenido de Gestión Tributaria: <http://leyes.tv/articulo/gestion-tributaria/>
- LLombart, O. A. (16 de Enero de 2010). Metodología de Data Mining CRISP. Obtenido de Metodología de Data Mining CRISP: <http://www.slideshare.net/oalonso/metodologa-de-data-mining-crisp>
- Luzón Calderon, V. (18 de Junio de 2004). Implantación de un proyecto de Knowledge Center con una herramienta comercial Synera. Obtenido de Implantación de un proyecto de Knowledge Center con una herramienta comercial Synera: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/442/1/27899ffc.pdf>
- Mallikarjunan, A. (2013). Corporación Sudamericana para el Desarrollo. Obtenido de Sistemas de Inteligencia Empresarial. La verdad sobre Minería de Datos: <http://grupocsl.org/sin-categoria/sistemas-de->

inteligencia-empresarial-la-verdad-sobre-la-mineria-de-datos-informe-de-anna-mallikarjunan/

- Martínez Álvarez, C. A. (Octubre de 2012). Aplicación de Técnicas de Minería de Datos para mejorar el proceso de Control de Gestión de ENTEL. Aplicación de Técnicas de Minería de Datos para mejorar el proceso de Control de Gestión de ENTEL. Santiago de Chile, Santiago de Chile, Chile.
- Moine, J. M. (Abril de 2013). Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo. Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo. La Plata, Buenos Aires, Argentina: Universidad Nacional de la Plata.
- Morales Carrasco, C., Ruiz Chang, F., & Ycaza Pesantes, W. J. (s.f.). Análisis de la Evasión Fiscal en el Impuesto a la Renta del Ecuador. Obtenido de Análisis de la Evasión Fiscal en el Impuesto a la Renta del Ecuador: <http://www.dspace.espol.edu.ec/bitstream/123456789/745/1/1396.pdf>
- OECD Organization for Economic Co-operation and Development. (2009). Foro sobre la Administración Tributaria: Subgrupo de Servicios a los Contribuyentes. Francia.
- Ríos Perez, S. A. (Octubre de 2012). Aplicación de Técnicas de Minería de Datos para mejorar el proceso de Control de Gestión de ENTEL. Aplicación de Técnicas de Minería de Datos para mejorar el proceso de Control de Gestión de ENTEL. Santiago de Chile, Chile: Universidad de Chile.
- Sallam, R. L., Tapadinhas, J., Parenteau, J., Yuen, D., & Hostmann, B. (20 de Febrero de 2014). Gartner. Obtenido de Magic Quadrant for Business Intelligence and Analytics Platforms: <http://www.gartner.com/technology/reprints.do?id=1-1QLGACN&ct=140210&st=sb>
- Servicio de Rentas Internas. (1998-2002). Memoria Institucional. Quito.
- Servicio de Rentas Internas. (2012). Circular No. NAC-DGECCGC12-00011. En Resoluciones del Servicio de Rentas Internas. Quito.

- Servicio de Rentas Internas. (17 de Julio de 2012). Plan Estratégico 2012-2015. Quito, Pichincha, Ecuador: Dirección Nacional de Planificación y Coordinación.
- Servicio de Rentas Internas. (Julio de 2012). Plan Estratégico Institucional. Obtenido de Servicio de Rentas Internas: www.sri.gob.ec
- Servicio de Rentas Internas. (2013). Equidad y Desarrollo Libro del futuro contribuyente - Bachillerato. Quito: Departamento de Servicios Tributarios Dirección Nacional - SRI.
- Servicio de Rentas Internas. (03 de 11 de 2013). Presentación de Anexos. Obtenido de Presentación de Anexos: <http://www.sri.gob.ec/web/10138/104>
- Servicio de Rentas Internas. (2013). Programación Anual de la Política Pública 2013. Quito: SRI.
- Sinnexus Business Intelligence Informática. (2013). Business Intelligence. Obtenido de Business Intelligence: http://www.sinnexus.com/business_intelligence/cuadro_mando_integral.aspx
- SRI. (2013). Instructivo Formulario Ruc. Inscripción y Actualización de la Información General del RUC. Quito, Pichincha, Ecuador.
- Ugalde, E. (26 de Febrero de 2014). Mundo Contact. Obtenido de Mundo Contact: <http://mundocontact.com/tendencias-en-ti-y-almacenamiento-de-datos-para-2014/>
- Valencia, E. (2006). Aplicación de Redes Neuronales a la Minería de Datos. Aplicación de Redes Neuronales a la Minería de Datos. México, México, México: Universidad Nacional Autónoma de México.
- Zorrilla, M. (Noviembre de 2010). Introducción al Bussiness Intelligence. Introducción al Bussiness Intelligence. Cantabria, España: Universidad de Cantabria.

ANEXOS

1.9 Anexos

1.10 Anexo 1: Encuesta

DATOS GENERALES:

La encuesta que se cita a continuación fue realizada el 19 de Noviembre de 2013.

El número de personas a las cuales se realizó la encuesta es 30, que representa el universo.

Las personas que fueron encuestadas, son funcionarios del SRI, del departamento de Gestión Tributaria, cuyos cargos son Responsables Departamentales y Jefes de Área.

El tipo de encuesta es un Censo.

Los criterios que se tomaron en cuenta para la realización de Qué, Cómo, Cuándo y a Quiénes se realizará la encuesta fueron deducidos en diferentes reuniones con el Director, Responsable de Gestión Tributaria, Jefe Nacional de Centro de Estudios Fiscales, y Expertos Tributarios.

ESTRUCTURA DE LA ENCUESTA

La presente encuesta busca recopilar aquellos factores relativos a los contribuyentes Sociedades que determinan su priorización o grado de importancia dentro de las funciones que realiza su unidad para fines de Control Tributario.

La finalidad de este cuestionario contempla el alcance de proveer a la Administración Tributaria una herramienta informática que permita la generación de reportes y estadísticas a partir de los criterios expertos recopilados en esta encuesta.

***Obligatorio**

Principio del formulario

Fecha_*

(*) Esta pregunta es obligatoria.

Dirección o Regional a la que pertenece:*

Departamento al que pertenece:

Área al que pertenece:

Ej: Control de Gestión, Infracciones

Años en la Institución:*

Ej: 1,5

1.- De los siguientes factores, cuáles considera que son más relevantes para priorizar a las sociedades para fines de Control Tributario?_*

	Alto	Medio - Alto	Medio	Bajo	No Aplica
Clase de Contribuyentes (Sociedades y Otros)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Tipo Impositivo Efectivo (Impuesto a la Renta Causado/ Ingresos)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

	Alto	Medio - Alto	Medio	Bajo	No Aplica
Totales)					
Sector Económico	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nivel de Activos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Nivel de Ingresos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

2.- Liste los factores adicionales que Ud. considera que son relevantes en la priorización de las sociedades en el contexto del Control Tributario y que no fueron tomados en cuenta en la pregunta 1, junto con el peso que Ud. considera (Alto/ Medio / Bajo)

Ej: Sector Económico Alto

3.- En caso de considerar que la Clase de Contribuyente es relevante, cuál de sus elementos es el principal para su unidad?

- a. Especial
- b. Sector Público
- c. Sociedad sin fin de Lucro
- d. Otro:

4.- En caso de considerar que el Sector Económico es relevante, cuál es el sector más importante por montos de recaudación o deuda general en el trabajo de su unidad?

Ej: Petróleo

5.- En caso de considerar que el Nivel de Activos es relevante, a partir de qué cantidad aproximadamente se considera importante para su unidad?

Ej: Mayor que 1 millón USD

6.- En el caso de considerar que el Nivel de Ingresos es relevante, a partir de qué cantidad aproximada se considera importante para su unidad?

Ej: Mayor que 5 millones USD

7.- Si Ud. respondió la pregunta 2 por favor detalle cuáles de sus elementos son los más importantes.

8.- Qué acciones orientan en su unidad los factores citados anteriormente en el contexto el contexto del Control Tributario en las sociedades?*

- e. Controles
 - f. Capacitación / Asesoría
 - g. Planificación Operativa
 - h. Otro:
-

9.- Qué grado de importancia considera que tienen los relacionados o interesados (accionistas, representantes legales, contadores y proveedores) para el Control Tributario de sociedades en su unidad?_*

1 2 3 4

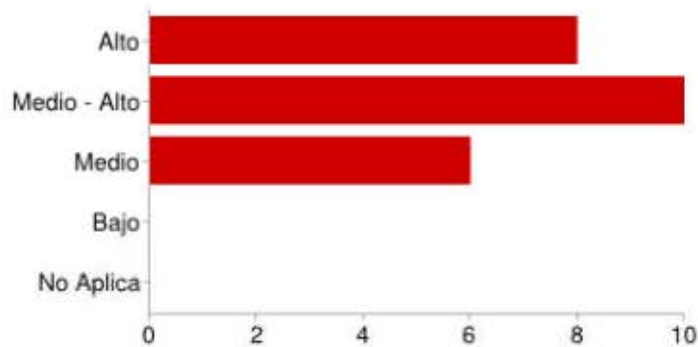
Menor Mayor

10.- Cite cuál relacionado o interesado es el más importante de acuerdo a su criterio?

1.11 Anexo 2: Resultados

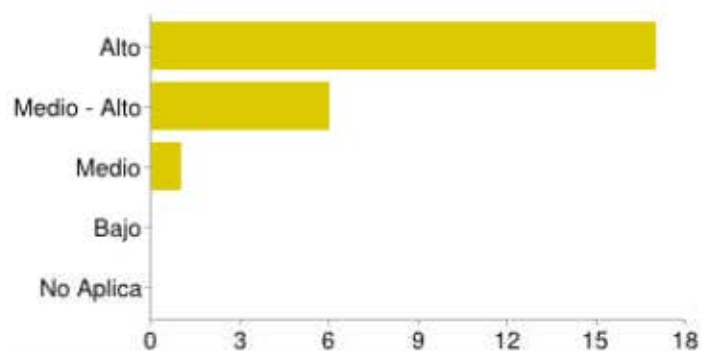
1. De los siguientes factores, cuáles considera que son más relevantes para priorizar a las sociedades para fines de Control Tributario?

- **Clase de Contribuyentes (Sociedades y Otros)**



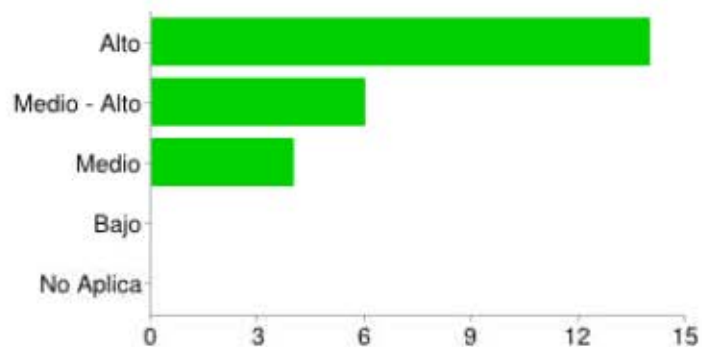
Criterio		
Alto	8	33%
Medio - Alto	10	42%
Medio	6	25%
Bajo	0	0%
No Aplica	0	0%

- Tipo Impositivo Efectivo (Impuesto a la Renta Causado/ Ingresos Totales)**



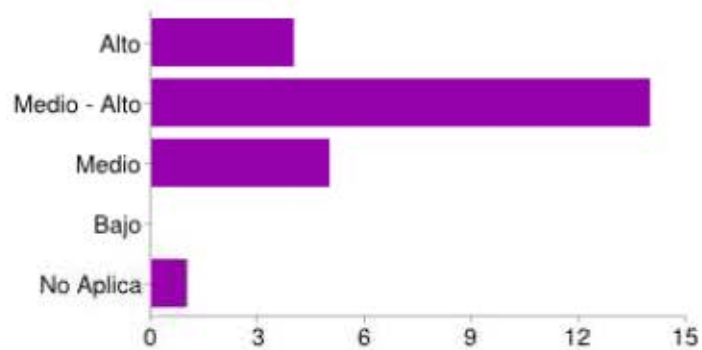
Alto	17	71%
Medio - Alto	6	25%
Medio	1	4%
Bajo	0	0%
No Aplica	0	0%

- Sector Económico**



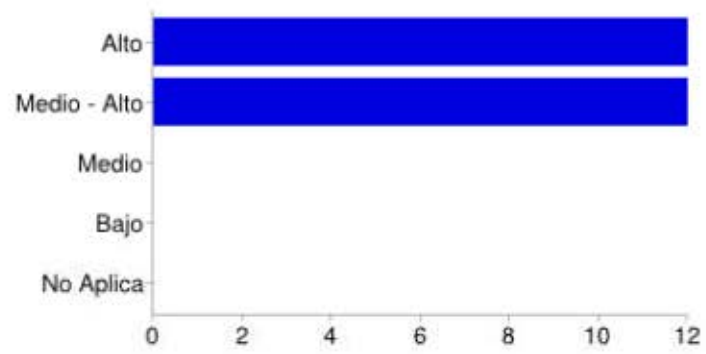
Alto	14	58%
Medio - Alto	8	25%
Medio	4	17%
Bajo	0	0%
No Aplica	0	0%

- **Nivel de Activos**



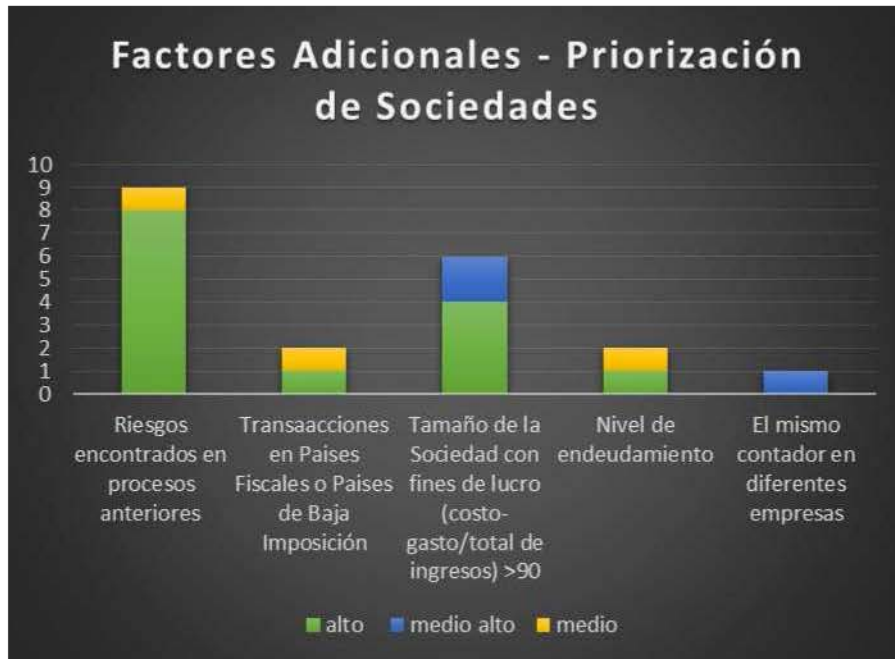
Alto	4	17%
Medio - Alto	14	58%
Medio	5	21%
Bajo	0	0%
No Aplica	1	4%

- **Nivel de Ingresos**



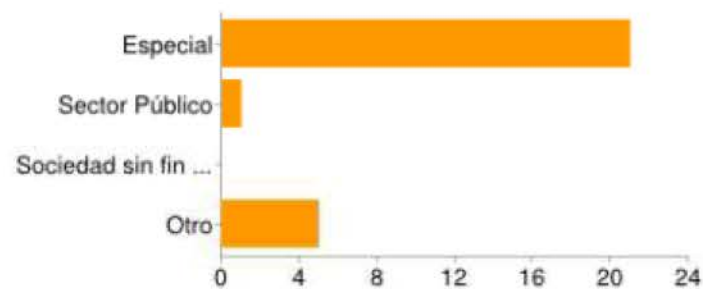
Alto	12	50%
Medio - Alto	12	50%
Medio	0	0%
Bajo	0	0%
No Aplica	0	0%

2. Liste los factores adicionales que ud. considera que son relevantes en la priorización de las sociedades en el contexto del Control Tributario y que no fueron tomados en cuenta en la pregunta 1, junto con el peso que ud. considera (Alto/ Medio / Bajo)



Riesgos encontrados en procesos anteriores	9	45%
Transacciones en Países Fiscales o Países de Baja Imposición	2	10%
Tamaño de la Sociedad con fines de lucro (costo-gasto/total de ingresos) >90	6	30%
Nivel de endeudamiento	2	10%
El mismo contador en diferentes empresas	1	5%

3. En caso de considerar que la Clase de Contribuyente es relevante, cuál de sus elementos es el principal para su unidad



Especial	21	78%
Sector Público	1	4%
Sociedad sin fin de Lucro	0	0%
Otro	5	19%

4. En caso de considerar que el Sector Económico es relevante, cuál es el sector más importante por montos de recaudación o deuda general en el trabajo de su unidad?



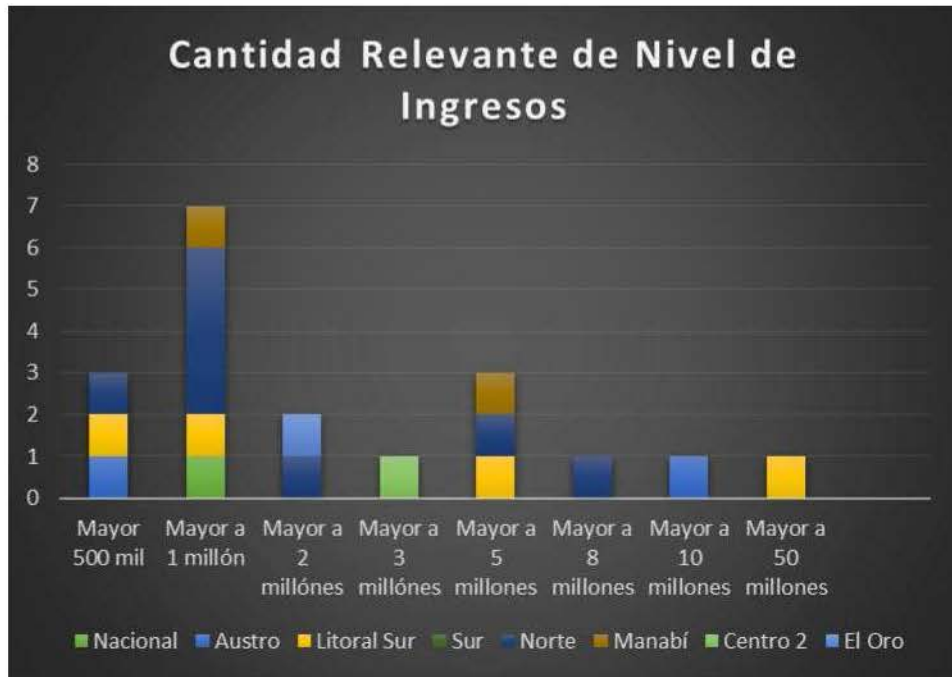
Telecomunicación	1	4%
Comercial	9	38%
Industrial	8	33%
Agrícola	3	13%
Construcción	3	13%

5. En caso de considerar que el Nivel de Activos es relevante, a partir de qué cantidad aproximadamente se considera importante para su unidad?



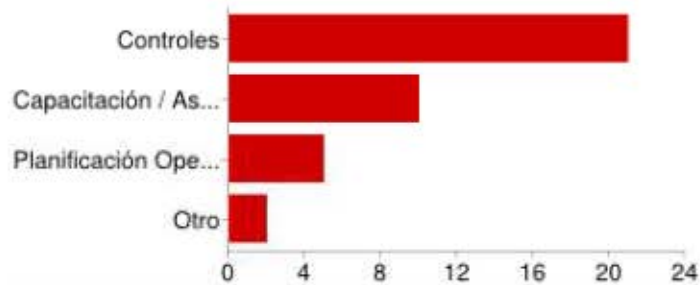
Mayor a 100 mil	1	6%
Mayor 500 mil	7	39%
Mayor 800 mil	1	6%
Mayor a 1 millón	5	28%
Mayor a 3 millones	1	6%
Mayor a 5 millones	3	17%

6. En el caso de considerar que el Nivel de Ingresos es relevante, a partir de qué cantidad aproximada se considera importante para su unidad?



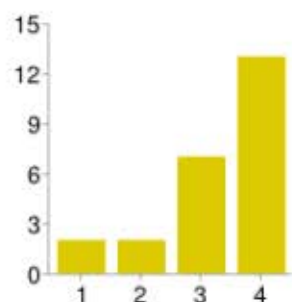
Mayor a 300 mil	2	10%
Mayor 500 mil	3	14%
Mayor a 1 millón	7	33%
Mayor a 2 millones	2	10%
Mayor a 3 millones	1	5%
Mayor a 5 millones	3	14%
Mayor a 8 millones	1	5%
Mayor a 10 millones	1	5%
Mayor a 50 millones	1	5%

7. Qué acciones orientan en su unidad los factores citados anteriormente en el contexto el contexto del Control Tributario en las sociedades?



Controles	21	55%
Capacitación / Asesoría	10	26%
Planificación Operativa	5	13%
Otro	2	5%

8. Qué grado de importancia considera que tienen los relacionados o interesados (accionistas, representantes legales, contadores y proveedores) para el Control Tributario de sociedades en su unidad?



1	2	8%
2	2	8%
3	7	28%
4	13	54%

9. Cite cuál relacionado o interesado es el más importante de acuerdo a su criterio?



Accionistas	9	38%
Representante Legal	6	25%
Proveedores	4	17%
Contador	4	17%
Clientes	1	4%

1.12 Anexo 3: Entrevistas

Las entrevistas fueron realizadas netamente a:

ENTREVISTADO	CARGO
1	Director de Planificación y Coordinación
2	Director de Gestión Tributaria
3	Jefe Nacional del CEF
4	Jefe Nacional de Gestión Tributaria
5	Experto Tributario
6	Experto Tributario

La entrevista netamente abarcaba parámetros para la definición de la encuesta con la finalidad de estructurar el Concepto de Importancia Fiscal, se seleccionó realizar a éste grupo la entrevista por su experiencia en la Gestión Tributaria