MODELO DE CARACTERIZACIÓN Y AGRUPACIÓN DEL SECTOR TIC DE UN GRUPO EMPRESARIAL

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL CALI

FABIAN DAVID ROLDAN PARRA

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL CALI

FACULTAD DE INGENIERÍA

PREGRADO INGENIERIA DE SISTEMAS

SANTIAGO DE CALI

2014

MODELO DE CARACTERIZACIÓN Y AGRUPACIÓN DEL SECTOR TIC DE UN GRUPO EMPRESARIAL

FABIAN DAVID ROLDAN PARRA

Proyecto de Grado presentado como requisito para optar al título de Ingeniero de Sistemas

Director de Proyecto

MSc. Emilia Rocío Segovia Jiménez

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL CALI

FACULTAD DE INGENIERÍA

PREGRADO INGENIERIA DE SISTEMAS

SANTIAGO DE CALI

2014

Contenido

RESUMEN.....7

CAPITULO 1. INTRODUCCIÓN.....9

1.1. DEFINICION DEL PROBLEMA 11

1.2. JUSTIFICACIÓN..... 12

1.3. OBJETIVOS 14

1.3.1. GENERAL 14

1.3.2. ESPECIFICOS..... 14

CAPITULO 2. MARCO CONCEPTUAL Y ESTADO DEL ARTE 15

2.1. ESTUDIO LOCAL DEL LIDIS 16

2.2. MODELO DE EMPRENDIMIENTO DE LA INDUSTRIA DE TI PARA COLOMBIA [3]..... 17

2.3. GENERACIÓN DE ESTRATEGIAS PARA EL DESARROLLO TECNOLOGICO Y DE MERCADO DEL SECTOR SOFTWARE Y SERVICIOS DE TI MEDIANTE LA APLICACIÓN DE LA VIGILANCIA TECNOLOGICA Y PROSPECTIVA. [4] 19

2.4. ESTUDIO DEL ALCANCE DE LA IMPLANTACIÓN DE TECNOLOGIAS DE LA INFORMACIÓN, COMO APOYO AL MEJORAMIENTO DE LOS PROCESOS, EN LAS PEQUEÑAS Y MEDIANAS EMPRESAS DEL SECTOR MANUFACTURERO DE BOGOTÁ. [5] 20

2.5. HERRAMIENTAS PARA LA IMPLEMENTACIÓN DE ALGORITMOS DE CLUSTERING...... 21

2.5.1. ORACLE DATA MINING..... 22

CAPITULO 3. METODOLOGIA DE LA INVESTIGACIÓN..... 23

CAPITULO 4. COMPARACIÓN DE ESTUDIOS REFERENTES CON EL MODELO DE ESTUDIO DE CARACTERIZACIÓN DEL LIDIS..... 25

4.1. MODELO DE COMPARACION DE REFERENTES 27

4.2 FACTORES Y VARIABLES DE LOS ESTUDIOS ECTI-LIDIS, CTI-FITI, CPYME-BOGOTÁ Y GETIC-SOFTSERVI. 28

Tabla 2. Factores y Variables Caracterización del emprendimiento TI del FITI. 29

4.3. TABLAS COMPARATIVAS DE LOS ESTUDIOS...... 32

4.3.1. MODELO RESULTANTE UTILIZADO EN LA ENCUESTA. 38

4.4. MODELO DE LA HERRAMIENTA DE CLASIFICACIÓN. 40
4.4.1. Responder Encuesta 40
4.4.2. Extracción y Mapeo de Respuestas 40
4.4.3. Clasificación de datos de las empresas 44
4.5. DISEÑO DE CONSULTAS Y REPORTES. 53
4.5.2. PROTOTIPOS DE CONSULTAS Y REPORTES 53
5. CONCLUSIONES 57
6. RECOMENDACIONES PARA TRABAJOS FUTUROS. 58
ANEXOS 59
BIBLIOGRAFIA 60
WEBGRAFIA 61

LISTA DE TABLAS

Tabla 1. Factores y Variables del Estudio de Caracterización LIDIS. 28

Tabla 2. Factores y Variables Caracterización del emprendimiento TI del FITI. 29

Tabla 3. Factores y Variables de Características de las pymes de Bogotá. 30

Tabla 4. Factores y Variables del Estudio de Generación de Estrategias para el desarrollo tecnológico de software y servicios del PTP. 31

Tabla 5 Comparación de Estudio del FITI con Estudio del LIDIS. 32

Tabla 6. Tabla de Comparación del Estudio de Pymes de Bogotá con el Estudio de Caracterización del LIDIS. 34

Tabla 7. Comparación del Estudio del PTP con el Estudio del LIDIS. 35

Tabla 8 Porcentaje resultante de la comparación del estudio de caracterización de pymes de Bogotá con el estudio del FITI. 37

Tabla 9. Modelo Resultante de las comparaciones entre los referentes. 39

Tabla 10. Tabla de Agrupaciones de las Empresas con el modelo de características y algoritmo k-Means. 48

Tabla 11. Tabla Comparativa de las características con los Cluster del algoritmo O-Cluster. 51

LISTA DE FIGURAS

Figura 1. Modelo de referencia entre los estudios de caracterización. 27

Figura 2. Modelo del proceso de clasificación de las empresas. 40

Figura 3. Tabla con información de las encuestas a las empresas. 41

Figura 4. Vista de los datos conectando al proceso de clustering. 45

Figura 5. Árbol de Decisión del Algoritmo K-Means. 46

Figura 6. Árbol de Decisión implementado por el Algoritmo O-Cluster. 49

Figura 7. Prototipo de Consulta y/o reporte: Tipo y actividad de la empresa. 54

Figura 8. Prototipo de consulta y/o reporte: clasificación del perfil del gerente. 55

Figura 9. Prototipo de consulta y/o reporte: porcentaje de las líneas tecnológicas de las empresas. 56

RESUMEN

Colombia es un país cuya economía se soporta en empresas de pequeña escala. Las pymes, junto con las microempresas, representan al menos el 90 por ciento del parque empresarial nacional y generan[1] el 73 por ciento del empleo y el 53 por ciento de la producción bruta de los sectores industriales, comercial y de servicios.

Investigadores interesados en promover estas empresas han realizado estudios que identifican características generales y específicas de las mismas. El objetivo de este proyecto es brindar un modelo computacional que permita agrupar las mipymes relacionadas con Tecnologías de la Información y comunicaciones TIC de un Grupo Empresarial, para facilitar a los entes interesados, visualizar datos correspondientes a actividades principales, líneas tecnológicas, tipos de productos y servicios, entre otras características que presentan una visión de cómo se encuentran conformadas las empresas del grupo empresarial.

Este proyecto plantea en primera instancia un modelo de caracterización de las pymes a partir de la armonización[1] de los diferentes modelos existentes aplicados al sector de las TIC, realizando la identificación y unificación de factores y variables de los referentes. Tomando el modelo definido se diseña la encuesta: "Estudio de caracterización de las empresas de tecnologías de la información y comunicaciones del sector TIC de un Grupo Empresarial de Santiago de Cali".

Esta encuesta fue aplicada a un grupo empresarial y los datos obtenidos en este proceso se almacenaron en una tabla minable, la cual se utilizó en la ejecución de los algoritmos de clustering (Algoritmo K-Means y Algoritmo O-Cluster) con los que se obtuvieron agrupamientos suceptibles de ser analizados por expertos para la participación en convocatorias.

[1] Según ANIF, 2003.

Finalmente se presentan los agrupamientos de empresas de acuerdo con sus características como resultado del proceso de clustering y las apreciaciones de los empresarios con respecto a esta información.

CAPITULO 1. INTRODUCCIÓN

En el primer capítulo de este proyecto se define el problema u objeto de estudio, explica por qué es importante para el sector, y se justifica el por que la caracterización y la importancia que puede tener la herramienta de clasificación para las empresas pertenecientes al sector TIC de un grupo empresarial, por último, se define también el alcance general y específico del proyecto.

En el capitulo 2, se refiere al marco conceptual y estado del arte, donde se introduce al lector en los estudios de caracterización de las empresas. En este se describe, cómo se definen las tecnologías de la información en la empresa y los referentes seleccionados para el presente trabajo. Luego se definen las herramientas de trabajo para la clasificación y análisis de resultados. Se detalla la definición de la encuesta que se aplicó a los empresarios seleccionados.

En el capítulo 3, se define la metodología de trabajo que sirvió de base en este estudio, definiendo las etapas de exploración e investigación de referentes, creación y diseño de modelo de clasificación, diseño del modelo de recolección de datos.

En el capitulo 4, se realiza la comparación del modelo de factores y variables utilizando un referente, donde se efectúa una lectura sobre la armonización de las múltiples normas y modelos [1] y se utilizan estas normas y modelos como líneas de base para el modelo resultante.

En el capitulo 5, se explica y visualiza detalladamente el diseño de la solución, donde se define un modelo de Data WareHouse y el uso de la técnica de Clustering para la generación de clasificación de empresas.

Para finalizar, se visualizan tres prototipos de presentación de la información para que un experto empresarial pueda realizar un análisis que lo oriente en la toma de decisiones, de acuerdo a los requerimientos que se soliciten en una convocatoria o licitación publica y/o privada, así se permitirá seleccionar las empresas mejores capacitadas para dicha convocatoria o licitación.

1.1. DEFINICION DEL PROBLEMA

El Grupo Multisectorial es una asociación empresarial conformada hace tres años, el cual tiene diferentes empresas afiliadas de la región de diferentes sectores entre ellas el comercio, transporte, hotelería, entretenimiento y turismo, TIC, agroindustrial, construcción, salud, emprendimiento, textiles y confecciones, entre otras independientes. Desde hace unos años atrás fue conformado para agremiar a las organizaciones y poder ordenarlas de acuerdo a distintas características, servicios y productos que ellas ofrecen. Sin embargo, con el transcurrir del tiempo, se ha identificado que ese enfoque se ha perdido y no ha sido posible poder ofrecer un plan de negocio ordenado y caracterizado de acuerdo a lo que desarrolla cada empresa, esto ha dificultado el proceso de recomendación o de selección para la participación de empresas en licitaciones o cuando vienen organizaciones externas a contratar un servicio o un producto.

Por otra parte, otro de los objetivos del grupo es dinamizar sus mesas sectoriales, iniciando por la Mesa de TIC, en donde los empresarios formen una cultura gremial activa para trabajar conjuntamente en proyectos e ideas de su conocimiento que beneficie a las empresas y el gremio.

Por lo tanto, el problema del sector TIC es no contar con una clasificación que identifique las características de las empresas, que a pesar de ser un sector importante para el desarrollo de la ciudad, no se cuenta aun con el reconocimiento ni con la fuerza a nivel gremial para integrarlas y agruparlas, además algunas no cuentan con una percepción de sus servicios en la región.

El grupo multisectorial no tiene una herramienta de información que le provea una visión global de las empresas, definiendo sus características y a la vez permita agruparlas estratégicamente para participar en convocatorias públicas o privadas

o bien incluirlas en proyectos internos que las beneficie tanto a ellas como a la región.

1.2. JUSTIFICACIÓN

Los estudios de empresas relacionadas con TIC existen y abordan diferentes características, específicamente a desarrollo de software, pero en estos no hay un modelo que combine otras características con el propósito de agruparlas o contar con datos relevantes del sector. Por lo tanto, al contar con una caracterización y agrupación de este tipo, donde se reflejen las características de un pequeño grupo de empresas del sector TIC, se observará un avance hacia el fortalecimiento y reconocimiento de los nichos a los que se dedican y bajo qué medios contribuyen a la innovación de la región.

Este estudio es un primer acercamiento a establecer la visión del sector e igualmente permitirá continuar con un análisis mas profundo en la medida en que se mantenga y compare periódicamente las características identificadas, lo cual se puede obtener con el uso frecuente, aplicación y almacenamiento de la información obtenida con la encuesta aquí propuesta.

Teniendo en cuenta lo anterior, se llevó cabo el análisis, diseño e implementación de un modelo que permite agrupar a las empresas pertenecientes al sector TIC de un grupo empresarial de la ciudad de Cali. De acuerdo a ciertas características como: servicios web, desarrollo de software, Business Process Outsourcing (BPO), hardware, seguridad y redes, consultoría de negocios, integración de sistemas, soporte de hardware y software, entre otras.

La caracterización y agrupación va a permitir conocer de las empresas de un grupo empresarial, qué tipo de productos ofrecen, que clase de servicios y cuál es el mercado relacionado a dichos productos y servicios, como también ver el perfil del gerente de una compañía y cuál puede ser esa línea tecnológica que lidera el sector TIC de un grupo de pymes. Esto facilitará una selección adecuada para la participación de las empresas del sector, en convocatorias impulsadas por el estado, licitaciones del gobierno, desarrollo de convenios, productos, servicios, entre otros.

A la vez la agrupación identificará a las empresas con negocios específicos o generales, así como también se puede establecer y analizar con los resultados, posibles necesidades y formular soluciones o proyectos.

1.3. OBJETIVOS

1.3.1. GENERAL

Llevar a cabo la caracterización de las empresas del sector TIC de un grupo empresarial para agruparlas de acuerdo a sus características.

1.3.2. ESPECIFICOS

- Armonizar los diferentes modelos de los estudios de caracterización empresarial, para definir un modelo estándar que describa las características que las empresas desarrollan.

- Diseñar un modelo de Agrupación (Clustering) basado en referentes de caracterización de empresas.

- Validar el modelo de caracterización y agrupación obtenido, con el criterio de un experto empresarial.

- Diseñar prototipos de consulta orientados a mejorar el proceso de análisis y búsqueda de información para las empresas TIC.

CAPITULO 2. MARCO CONCEPTUAL Y ESTADO DEL ARTE

Las tecnologías de la información y comunicaciones, TIC son un conjunto de servicios, redes, software y aparatos que tienen como fin la mejora de la calidad de vida de las personas dentro de un entorno.[2] Hoy en día en las empresas no hace falta hablar de un computador cuando se hace referencia al procesamiento de la información. Internet puede formar parte de ese procesamiento, que se realiza de manera distribuida y remota.[3]

Las telecomunicaciones también se relacionan con las TIC ya que es el estudio y aplicación de la técnica que diseña sistemas que permitan comunicación a larga distancia a través de la transmisión y recepción de señales, igualmente se puede afirmar que en la telecomunicación se incluyen muchas tecnologías como la radio, televisión, teléfono y telefonía móvil, comunicaciones de datos y redes informáticas, como internet. Estas tecnologías son de vital importancia en el contexto socioeconómico actual. Hoy en día existen empresas que se dedican a uno o varios desarrollos de sistemas o prestación de los servicios mencionados con anterioridad.

En la actualidad existen estudios de caracterización de empresas que se enfocan en el desarrollo de aplicaciones de software. Estos se basan por ejemplo: En los modelos de calidad que implementan las empresas desarrolladoras de software.

Para el proyecto se tomaron como referentes los siguientes estudios, Estudio Local del Lidis, Caracterización FITI, Fortalecimiento de la Industria de TI, Generación de Estrategias para el desarrollo tecnológico y de mercado del sector software y servicios de TI, Estudio del alcance de la implantación de tecnologías de la información en las pequeñas y medianas empresas de Bogotá y el modelo de la medición de las tecnologías de la información y comunicaciones - TIC.

[2] http://www.monografias.com/trabajos67/tics/tics.shtml
[3] Malbernat, Lucía Rosario (2010). «Tecnologías educativas e innovación en la Universidad». *LaCapitalmdp.com*.

2.1. ESTUDIO LOCAL DEL LIDIS

A continuación se presenta un análisis del estado del arte y la definición de los factores que utilizaron para clasificar la información en el proyecto que planteó el Ingeniero Luis Merchán, uno de los investigadores en la caracterización de las empresas pertenecientes a la industria emergente del software del sur occidente colombiano y que se tomará como referente para esta investigación.

A partir de los resultados presentados en [2], se encontraron algunas investigaciones previas y trabajos relacionados donde el propósito fue definir una clasificación de las pequeñas empresas desarrolladoras de software del sur occidente del país y establecer un conjunto de factores o criterios para orientar el proceso de mejoramiento de aquellas empresas.

La lectura realizada muestra que fueron seleccionadas 49 empresas pequeñas desarrolladoras de software. Una encuesta para la recolección de datos se aplicó a las empresas de la red de parques, Parquesoft en Cali, Manizales, Palmira, Pasto, Pereira, Popayán y Tuluá. Las empresas fueron seleccionadas aleatoriamente para el estudio. La composición total resultó bastante representativa en términos sectoriales, geográficos y función del tamaño de la empresa. En la lectura y el documento de este referente se puede ver las características de la muestra del estudio, análisis cuantitativo de la estructura de las variables comunes entre los estudios y los resultados de la recolección de datos mediante un cuestionario.

En resumen, es posible observar a partir de los trabajos referenciados en la caracterización de las empresas pertenecientes a la industria emergente del software del suroccidente colombiano, que desde los últimos 5 años los empresarios de la región han mostrado un crecimiento con sus empresas a nivel

local y regional, donde se ha podido identificar la situación y estructura de las empresas emergentes de Parquesoft. Así mismo, dentro de la estructura encontrada se puede diferenciar que estas empresas poseen características similares respecto a la cantidad limitada en recursos de personal, a los procesos que realizan, a la metodología que aplican y a la forma en que dirigen y organizan sus actividades.

Esta caracterización realizada en [2], concluye que ha permitido reconocer factores predominantes en la industria de software en Colombia así como: el tipo de software, los canales de distribución del producto, los países a los que exporta y los tipos de empresas de software que predominan según el modelo de referencia que tomaron para el estudio local del LIDIS.

2.2. MODELO DE EMPRENDIMIENTO DE LA INDUSTRIA DE TI PARA COLOMBIA [3]

A partir de una lectura en [3] y citando parte de la metodología de modelado: “El diseño de un modelo de emprendimiento para la industria TI en Colombia exige la generación de una dinámica metodológica que parte de la especificidad del sector y del contexto geográfico en el que se enmarca.” con esto es posible observar cómo se lleva a cabo la caracterización que se realiza al sector TI en Colombia como una investigación de carácter deductivo y se extiende al análisis de 20 emprendedores reconocidos de la industria, con el fin de entender desde estudio de casos (método inductivo), las buenas prácticas y los factores claves que les permitieron convertirse en emprendimientos exitosos. De acuerdo al plan de trabajo que concluyen en [3], al caracterizar el sector y sus emprendedores, desarrollan la bitácora de buenas prácticas de emprendimiento TI y concretan en el estudio la fase de inicio de análisis, a partir de este proceden al desarrollo del diseño de un modelo deseado, planteando un modelo de ecosistema digital de

emprendimiento TI y un modelo de acompañamiento y seguimiento a empresarios que desearon intensificar su dinámica emprendedora en un ecosistema holístico y plurinstitucional.

El Plan de trabajo utilizado en este referente para el desarrollo de los objetivos incluye una fase de caracterización, análisis, diseño e implementación de una bitácora de buenas practicas de emprendimiento TI.

De acuerdo a los indicadores analizados en este estudio se llega a la siguiente conclusión: “Se evidencia un comportamiento estable en los indicadores analizados, el orden de estos se mantiene anualmente, lo cual aduce que la industria TI tiene un comportamiento muy homogéneo, de allí que algún factor que afecte favorable o desfavorablemente el sector, genera un efecto generalizado u orientado por tendencias.”[4]

En el estudio de caso realizado en [3], se diseñó una herramienta específica que mediante entrevista estructurada y no estructurada[5] aplicada directamente a los propietarios o gerentes permitió obtener información de carácter cuantitativo y cualitativo.

En el análisis descriptivo de esta información afirman que la innovación es considerada un factor de alta importancia para las compañías encuestadas y así logran consolidar un portafolio de proyectos, donde mediante una tabla muestran la gestión de la innovación.

Para finalizar en este estudio [3] se observa la información de las diferentes empresas y emprendedores, como características empresariales, estas son

[4] Capítulo 2.5, Conclusiones respecto a las medidas estadísticas del sector, pagina 17.
[5]Son las entrevistas en las que a todos los entrevistados se les hacen las mismas preguntas con la misma formulación y en el mismo orden. En la entrevista no estructurada, ni siquiera el contenido de las preguntas está preestablecido, y puede variar en función del sujeto.

formación del gerente, posición tecnológica de la empresa, la calidad como factor, la capacidad comercial de la organización, la formación del personal y la innovación.

2.3. GENERACIÓN DE ESTRATEGIAS PARA EL DESARROLLO TECNOLOGICO Y DE MERCADO DEL SECTOR SOFTWARE Y SERVICIOS DE TI MEDIANTE LA APLICACIÓN DE LA VIGILANCIA TECNOLOGICA Y PROSPECTIVA. [4]

A partir de la lectura en [4], el objetivo fundamental de este referente fue, la generación de estrategias para el desarrollo tecnológico y de mercado del sector Software y Servicios de TI.

En el proceso de realización de este, resaltan unas actividades, entregables y actores que intervienen en el desarrollo del proyecto, a esto lo llaman modelo de intervención.

La metodología que utilizan para el desarrollo del proyecto consiste en la aplicación de vigilancia tecnológica y prospectiva para la validación de líneas tecnológicas, para esto definen un indicador denominado FCV, Factor Critico de Vigilancia, y son las líneas tecnológicas del sector software y servicios de TI.
En [4] definen una priorización de líneas tecnológicas tomando como base los resultados de la vigilancia tecnológica.

Los resultados y las conclusiones son concluidos por los actores del referente, de acuerdo a varios indicadores tecnológicos y los resultados que estos adquirieron en la definición de las líneas tecnológicas o los factores críticos de vigilancia.

2.4. ESTUDIO DEL ALCANCE DE LA IMPLANTACIÓN DE TECNOLOGIAS DE LA INFORMACIÓN, COMO APOYO AL MEJORAMIENTO DE LOS PROCESOS, EN LAS PEQUEÑAS Y MEDIANAS EMPRESAS DEL SECTOR MANUFACTURERO DE BOGOTÁ. [5]

A partir de la lectura en [5], se observa un diagnostico del uso de tecnología de la información en las pymes industriales de Bogotá como apoyo al manejo de sus principales procesos, igualmente el impacto de la tecnología de la información en sus procesos y en la cultura organizacional. Finalizando este estudio el autor concluye de acuerdo a una serie de graficas que en las pymes manufactureras de Bogotá, existe una tendencia generalizada a manifestar cierto rechazo a la inversión o adquisición de tecnologías de la información y recomiendan que las pymes por su naturaleza y por el hecho de tener en su mayoría un tipo de producción por pedido, deben centrar la restructuración y el mejoramiento de las actividades, en el proceso de producción.

Seguidamente afirman que las tecnologías de la información son herramientas que efectivamente contribuyen al mejoramiento en la velocidad y exactitud en el desarrollo de transacciones cotidianas y permiten la integración de la información entre diferentes procesos.

Para finalizar concluyen que el apoyo de las tecnologías de la información es realmente efectivo cuando se busca la integración de todas las áreas de la empresa, o por lo menos las áreas que se encuentran estrechamente relacionadas, ya que la sistematización separada de unas pocas áreas implica el desaprovechamiento de la funcionalidad de estas herramientas informáticas, y algunos casos puede ocasionar duplicidad en la información manejada.

2.5. HERRAMIENTAS PARA LA IMPLEMENTACIÓN DE ALGORITMOS DE CLUSTERING.

En esta sección, se realizó una búsqueda de herramientas que trabajen e implementen algoritmos de clustering para analizar los datos obtenidos en la encuesta de caracterización. A continuación se presentarán una descripción de cada herramienta.

Tableau Public 8.0, La cual permite realizar la visualización de estadísticas mediante datos almacenados en diferentes bases de datos o una simple hoja de Excel, donde le va a permitir crear graficas interactivas y publicarlas en la web. Esta herramienta no implementa algoritmos de clustering, pero sin embargo, permite realizar análisis de negocios fáciles, mezclar y combinar varias fuentes de datos en una sola vista, trabajar con bases de datos y hojas de cálculo de cualquier tamaño.

Data Mining Add-ins, es un complemento de Excel que implementa clasificación y clustering a partir de la información almacenada en una base de datos myslq server, Esta herramienta requiere de licencias para trabajar sobre ella ya que es un complemento de Microsoft, permite realizar análisis predictivos. Con la plataforma de SQL Server Analysis Server permite construir modelos analíticos de alto rendimiento que pueden ser utilizados para el análisis interactivo de datos, generación de informes y visualización, sin duda es una herramienta soportable para implementar inteligencia de negocios.

Este complemento de Excel permite a los usuarios avanzados a pasar por el ciclo de vida de desarrollo completo para el modelo de minería de datos en Excel, utilizando cualquier fuente de datos como hojas de cálculo o bases de datos SQL Server.

2.5.1. ORACLE DATA MINING

La GUI libre de Oracle Data Miner es una extensión de Oracle SQL Developer que permite a los analistas trabajar directamente dentro de la base de datos, explorar gráficamente, construir y evaluar varios modelos de minería de datos, aplicar modelos de minería de datos, entre esos, Clustering.

Permite crear predicciones y perspectivas de minería de datos Oracle, con los datos existentes. Los flujos de trabajo de Oracle Data Miner capturan y documentan la metodología de análisis del usuario, y pueden ser guardados y compartidos con otras personas para automatizar las metodologías analíticas avanzadas.

Oracle DataMiner crea modelos predictivos que los desarrolladores de aplicaciones pueden integrar en aplicaciones, para automatizar el descubrimiento y distribución de inteligencia de negocios y patrones en toda la empresa, Oracle Data Miner es un complemento de Sql Developer.

Oracle Data Mining soporta dos algoritmos de clustering: una versión mejorada de k-means, y un algoritmo patentado llamado Oracle Partitioning Clustering ortogonal (O-Cluster). Ambos algoritmos son de agrupación jerárquica.[6]

El algoritmo K-means construye modelos de una manera jerárquica, es decir, un modelo de arriba hacia abajo con separaciones binarias y el refinamiento de todos los nodos al final. Este enfoque del algoritmo evita la necesidad de la construcción de múltiples modelos de k-means y proporciona agrupando los resultados que son consistentemente superiores a k-means tradicionales.

El algoritmo O-Cluster crea un modelo de agrupación jerárquica basada en la red, es decir, que crea particiones eje-paralelas (ortogonal) en el espacio de atributo de entrada.

[6] http://www.oracle.com/technetwork/database/options/odm/dataminerworkflow-168677.html

Los grupos descubiertos por O-Cluster se utilizan para generar un modelo de probabilidad bayesiana, el modelo de probabilidad generada es un modelo de mezcla en la que los componentes de la mezcla están representados por un producto de distribuciones normales independientes para los atributos numéricos y multinomiales para los atributos categóricos.

Se selecciona esta herramienta, porque se descubre que trabaja con algoritmos de clustering y clasificación que no requieren muchos datos para identificar y clasificar características de algo específico.

CAPITULO 3. METODOLOGIA DE LA INVESTIGACIÓN

A nivel de la etapa 0, correspondiente a la exploración del referente guía, se llevó a cabo una revisión profunda donde permitió identificar el trabajo relacionado y analizar a fondo sus diferencias, relaciones, debilidades y fortalezas. Los resultados obtenidos, permitieron definir y mejorar el modelo de este proyecto respecto a los trabajos existentes.

A nivel de la etapa 1, se analizó y definió con claridad los factores que deben ser estudiados y aplicados al cuestionario que se realizó a las empresas.

En la etapa 2 se realiza el diseño y desarrollo de los componentes necesarios para el modelo de recolección de datos, el cual es el diseño de una encuesta y almacenamiento de datos que consiste en la creación de un modelo en estrella de datos minables sobre una base de datos, que serán utilizados en la ejecución de los algoritmos de clustering y el diseño del prototipo de consulta, con relación a las condiciones o capacidades que debe cumplir en el análisis.

Las ***disciplinas de gestión***, se encargan de hacer que el proyecto no se desvíe de su curso, con una buena ***planificación***, un estricto ***seguimiento y control*** de las actividades a desarrollar y con ***aseguramiento de la calidad*** por parte del director del proyecto.

Fuentes de Información

Para cumplir con los objetivos del proyecto fue necesario utilizar el envío masivo de información a los empresarios sobre la encuesta que debían realizar via web ó recibir una visita por parte del encuestador donde este mismo se encargó de formular personalmente las preguntas y así fue contestando la encuesta, de acuerdo a las respuestas del empresario.

Delimitación Temporal

El tiempo de desarrollo de la presente investigación corresponde a 12 meses calendario equivalentes a dos semestres académicos iniciando el mes de Agosto de 2013.

Delimitación Espacial

La presente investigación se llevó a cabo con la colaboración de algunas empresas del Grupo Multisectorial, Asociación Multisectorial de Empresas del Valle del Cauca.

CAPITULO 4. COMPARACIÓN DE ESTUDIOS REFERENTES CON EL MODELO DE ESTUDIO DE CARACTERIZACIÓN DEL LIDIS

Durante la investigación del proyecto, se lograron identificar y analizar 4 referentes: un referente local que es el estudio de caracterización de empresas pertenecientes a la red de parques, Parquesoft, en adelante ECTI-LIDIS, un referente regional que es el Estudio del Alcance de la implantación de las tecnologías de información, como apoyo al mejoramiento de los procesos, en las pequeñas y medianas empresas del sector manufacturero en Bogotá, en adelante Cpyme-Bogotá y dos referentes nacionales, el primero es el modelo de emprendimiento de la industria de las tecnologías de la información (TI) para Colombia, en adelante CTI-FITI (Fortalecimiento de la Industria TI de Colombia) y el segundo referente nacional tiene como titulo Generación de estrategias para el desarrollo tecnológico y de mercado del sector software y servicios de TI mediante la aplicación de vigilancia tecnológica y prospectiva, en adelante GETIC-SOFTSERVI.

Para este proyecto se definió el ECTI-LIDIS como el referente principal y por esto se tomó como punto de partida para la comparación con los 3 restantes.

Como se visualiza en la figura 1, primero se definió un modelo que agrupe las características de los referentes, eliminando duplicidades, se realiza la comparación de los factores y variables de los referentes restantes con el principal, ECTI-LIDIS, en esta, se analiza que en el referente nacional quedan muchas variables sin comparar, por lo cual se decidió pasar a una segunda etapa, en donde se comparó el referente regional, Cpyme-Bogotá, con los dos restantes, donde sí se hallaron variables que coincidían.

Al final del proceso de comparación se definieron los factores y las variables que se tomaron en cuenta para realizar las preguntas en la encuesta que se aplicó a

las diferentes empresas y emprendimientos de un grupo empresarial de la ciudad de Cali.

La principal razón por la que se tomó el estudio de Cpyme-Bogotá como objeto de comparación con los dos restantes, consiste en que los factores y variables de este estudio eran más amplios y es donde nace el modelo de clasificación a partir de estos, junto con una encuesta que contiene las principales características de una empresa relacionada con las tecnologías de la información y comunicaciones.

4.1. MODELO DE COMPARACION DE REFERENTES

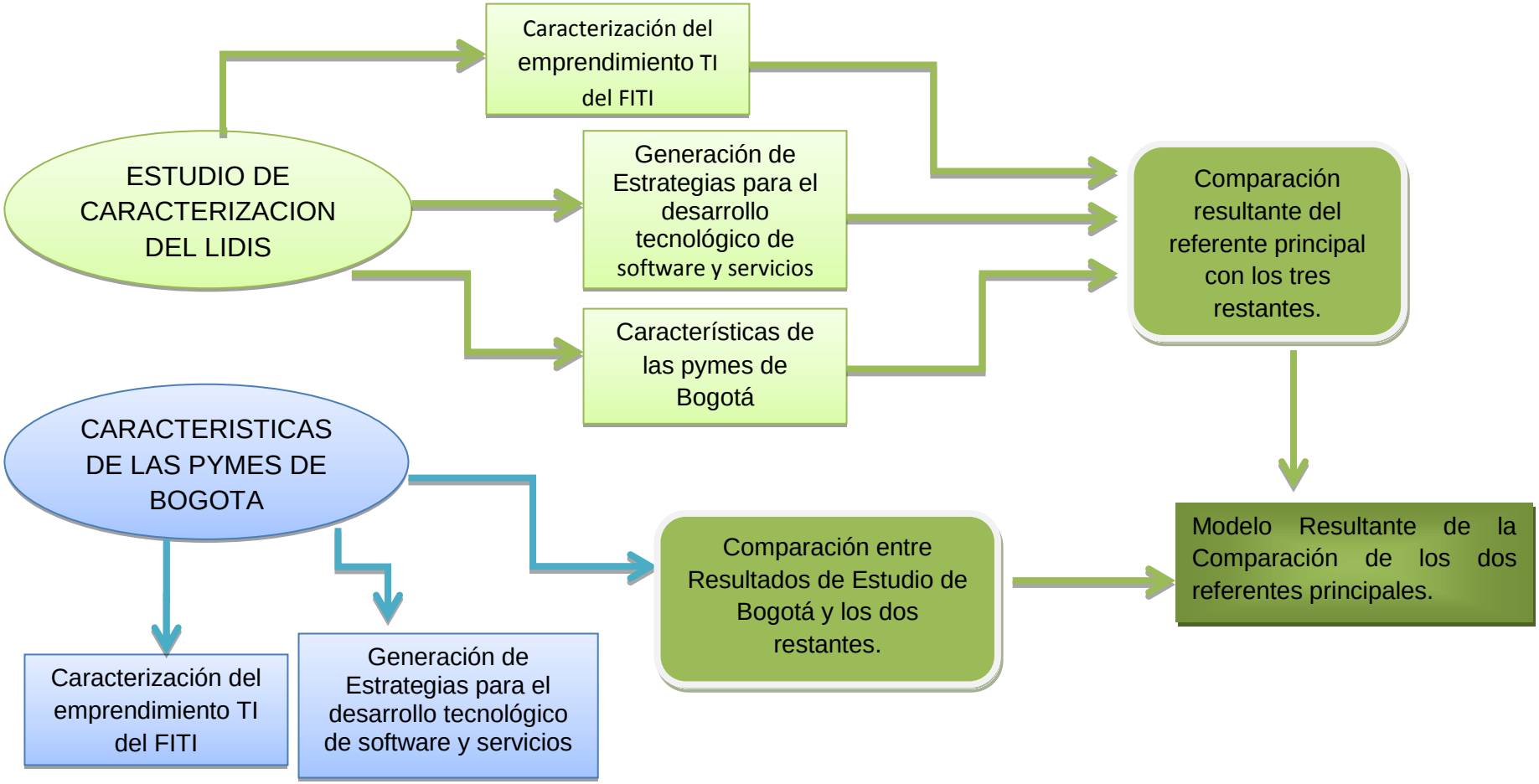

Figura 1. Modelo de referencia entre los estudios de caracterización.

4.2 FACTORES Y VARIABLES DE LOS ESTUDIOS ECTI-LIDIS, CTI-FITI, CPYME-BOGOTÁ Y GETIC-SOFTSERVI.

Principal Actividad de la Empresa	Canales de Distribución	Tipos de Empresas	Exportación del Software	Clasificación de la Industria del Software
Desarrollo de Software a la Medida	Ventas Directas	Desarrolladoras de Software	Ventas al Exterior	Integración de Sistemas
Desarrollo de Software Genérico	Representantes	Productoras de Hardware		Desarrollo de Aplicaciones de Software
	Distribuidores	Distribuidoras		Consultoría de IT
	Sucursales			Soporte de Hardware y Software
	Internet			Outsourcing de IT

Tabla 1. Factores y Variables del Estudio de Caracterización LIDIS.

Perfil Gerente/Empresario	Posición tecnológica de la empresa	Calidad	Gestión del Mercadeo y las Ventas	Gestión de la Innovación	Formación del Personal
Habilidades de carácter tecnológico	Número de patentes	Aplican modelos de calidad dentro de la organización	Mercadeo estratégico	Análisis del mercado	Desarrollo humano
Habilidades en mercadeo y comercialización, contabilidad y finanzas.	Red de relaciones externas		Mercadeo Operativo	Vigilancia tecnológica	Formación y desarrollo del talento humano
					Capacitación por competencias
					Compensaciones

Tabla 2. Factores y Variables Caracterización del emprendimiento TI del FITI.

Características de la Gerencia	Características en la Relación con los Clientes	Características de la Planeación	Características del Manejo de la Calidad	Características de la Competencia	Características del Manejo Financiero
Administración de Tipo Familiar	Flexibilidad en las relaciones comerciales.				
Procesos para el Manejo de Talento Humano					
Actitud Cautelosa y Resistencia a la Colaboración de otras Empresas.					
Participación de los Empleados en las Decisiones					

Tabla 3. Factores y Variables de Características de las pymes de Bogotá.

Líneas Tecnológicas del Sector Software y Servicios de TI
Desarrollo de Software
Integración de Sistemas
Consultoría en Aplicaciones
Servicio de Soporte y Mantenimiento
Outsourcing

Tabla 4. Factores y Variables del Estudio de Generación de Estrategias para el desarrollo tecnológico de software y servicios del PTP.

4.3. TABLAS COMPARATIVAS DE LOS ESTUDIOS.

En esta primera comparación entre los estudios, se visualiza que el perfil del empresario se relaciona con la actividad de exportación de la empresa, ya que podemos afirmar si las habilidades de mercadeo y comercialización del gerente han influido en las ventas fuera del país de la empresa.

El perfil del empresario se relaciona con las variables del factor Clasificación de la industria del software, ya que podemos definir cual es la principal actividad en la que tiene experiencia el gerente de la compañía. En la siguiente tabla se visualiza como se definen los factores y variables abordados, con estos dos estudios y que se tienen en cuenta en la encuesta aplicada a los empresarios.

			LIDIS							
	LIDIS FITI		Principal Actividad de la Empresa		Exportacion del Software	Clasificacion de la industria del software				
			Desarrollo de software a la medida.	Desarrollo de software generico	Ventas al exterior.	Integracion de Sistemas	Desarrollo de aplicaciones de software	Consultoria de IT	Soporte de HW y SW	Outsourcing de IT.
FITI	Perfil Gerente/Empresario	Habilidades de carácter tecnologico								
		Habilidades en mercadeo			X	X	X	X	X	X
	Posicion tecnologica	Número de Patentes	X		X					
	Gestion de la Innovacion	Analisis del Mercado			X					
		Vigilancia Tecnologica								

Tabla 5 Comparación de Estudio del FITI con Estudio del LIDIS.

En la siguiente comparación, se visualiza que las características en la relación con los clientes coinciden con variables de uno de los factores del LIDIS, denominado canales de distribución. Se puede relacionar de la siguiente forma, cuales son los canales de distribución y el porcentaje de efectividad de cada uno de los clientes

de la empresa. Los factores y las variables que no coinciden se presentarán más adelante en la comparación con los demás estudios.

	LIDIS						
PYMES BOGOTA	LIDIS/PYMES BOGOTA		Principal Actividad de la Empresa	Canales de distribucion			
			Ventas Directas	Representantes.	Distribuidores	Sucursales	Internet
	Caracteristica de la gerencia	Flexibilidad en las relaciones comerciales.	X	X	X	X	X
	Caracteristica en la relacion con los clientes						
	Caracteristicas de la planeacion						

Tabla 6. Tabla de Comparación del Estudio de Pymes de Bogotá con el Estudio de Caracterización del LIDIS.

			LIDIS						
Sector TIC y SW. PTP	LIDIS/SECTOR TIC Y SW. PTP		Principal Actividad de la Empresa		Clasificacion de la industria del software				
			Desarrollo de software a la medida.	Desarrollo de software generico	Integracion de Sistemas	Desarrollo de aplicaciones de software	Consultoria de IT	Soporte de HW y SW	Outsourcing de IT.
	Líneas tecnológicas del sector Software y Servicios de TI	Desarrollo de software	X	X		X			
		Integracion de sistemas			X				
		Consultoria en aplicaciones					X		
		Servicio de soporte y mantenimiento						X	
		Outsourcing							X

Tabla 7. Comparación del Estudio del PTP con el Estudio del LIDIS.

Al realizar la comparación de las variables del ECTI-LIDIS con el Estudio de GETIC-SOFTSERVI se puede observar que en el factor correspondiente a la Clasificación del Software con el factor de las Líneas Tecnológicas, se observa que la mayoría de las variables de ambos estudios coinciden significativamente, lo cual quiere decir que hay variables que pueden utilizarse en la definición de varias preguntas en la encuesta, seguidamente se observa que hay otro factor que tiene coincidencias y es la principal actividad de la empresa. Esto ayuda a definir de manera más completa la información en la encuesta de los empresarios.

Al analizar las anteriores comparaciones del estudio del LIDIS se observa que en este proceso se puede tomar un segundo referente guía, Este es el Estudio de las pymes manufactureras de Bogotá, por ser un estudio más amplio en características, que pueden aportar una variedad importante en el proyecto. Por este motivo se realizan las comparaciones necesarias pero enfocando y tomando como objeto de comparación con los restantes. Al realizar estas ultimas comparaciones se logra contar con la relación de factores y variables necesarias para obtener datos e información de las empresas que van hacer parte del objeto de estudio y las características principales de las mismas.

Analizando los datos se puede observar como cada factor relacionado tiene varias palabras coincidentes lo que los hace aun más compatibles, por ejemplo veamos, el primer factor de pymes Bogotá es Características de la gerencia y el primer factor del FITI es perfil del gerente/empresario, se puede ver que ambos contienen variables relacionados con el perfil de un empresario o gerente de una compañía.

			FITI													
			Caracteristicas de la Gerencia		Caracteristica en la relacion con los clientes				Caracteristicas del Manejo de la Calidad				Caracteristicas del Manejo Financiero			
PYMES BOGOTA	BOGOTA/FITI		Procesos para el manejo de talento humano	Actitud cautelosa y resistencia a la colaboracion de otras empresas.	Participacion de los empleados en las decisiones.	Herramientas de investigacion de mercados	Relacion con el cliente basada mas en la confianza que en la calidad del servicio.	Objetivos centrados en los problemas que van apareciendo.	Importancia a la informacion del entorno.	Control de Calidad del proceso del producto.	Politicas de acciones preventivas.	Herramientas de investigacion de la competencia.	Inversion en la infraestructura de la empresa.	Inversion en Investigacion y Desarrollo.	Inversion en el mejoramiento tecnologico de los procesos.	Planeacion financiera.
	Posicion tecnologica de la empresa	Vigilancia Tecnologica		X		X										
		Inteligencia Competitiva		X			X									
		Gastos relacionados en investigacion y desarrollo											X	X	X	X
	Calidad	Aplicacion de Modelos de Calidad								X	X					
	Formacion del Personal	Desarrollo humano	X		X											
		Capacitacion por competencias	X		X											X
		Formacion y desarrollo del talento humano	X		X											
	Gestion de la Innovacion	Benchmarking		X		X						X				
		Trabajo en red	X	X												
		Funcionamiento en equipo	X		X				X							
		Gestion del cambio				X		X			X					

Tabla 8 Porcentaje resultante de la comparación del estudio de caracterización de pymes de Bogotá con el estudio del FITI.

4.3.1. MODELO RESULTANTE UTILIZADO EN LA ENCUESTA.

En la siguiente tabla se visualizan los factores y variables que surgieron de las comparaciones elaboradas de los referentes, para la realización de la encuesta que diligenciaron las empresas y se logra obtener información importante del sector TIC de la ciudad de Cali y las empresas pertenecientes al Grupo Multisectorial.

En la selección de los factores y variables y en el proceso de comparación de los estudios, se realizó una unificación coherente para la definición de la encuesta, como a su vez, la encuesta es la unión de variables similares.

En la realización de la encuesta se mezclan algunos factores y variables para la construcción de una pregunta especifica preferiblemente con múltiples respuestas, este método facilitó hacer preguntas en donde se podía analizar mas de un factor, teniendo en cuenta no superar la barrera de 30 preguntas en la encuesta, ya que, por definición una encuesta de más de 30 preguntas pierde efectividad.

El resultado de este modelo para la encuesta, se obtiene después de la armonización de los estudios referentes en este proyecto y los factores, variables e indicadores resultantes de esta. Donde se puede observar detalladamente en el modelo de clasificación y las tablas detalladas anteriormente presentadas.

Factores resultantes	Variables Resultantes
Principal Actividad de la Empresa	Nombre y Tipo de Empresa.
Características de la Gerencia	Profesión del Gerente/Empresario de la Empresa. Principal actividad del gerente/empresario de la empresa.
Clasificación de la Industria TI	Actividades TIC de la Empresa
Exportación de productos	Análisis del Mercado.
Formación del Personal	Desarrollo humano. Participación de los empleados en la toma de decisiones e ideas de la empresa.
Posición tecnológica de la empresa	Patentes registradas por la empresa.
Características en la relación con los clientes.	Número de Clientes de la Empresa.
Gestión del Mercadeo.	Tipo de Mercadeo de la Empresa. Canales de Distribución
Gestión de la Innovación.	Habilidades del Gerente/Empresario relacionados con las ventas del producto y/o servicio fuera del país.
Características del Manejo de la Calidad.	Implementación de modelos de Calidad.
Línea Tecnológica de la Empresa.	Actividades TIC de la Empresa.
Características del Manejo Financiero de la Empresa.	Canales de Distribución.
Características de la Competencia.	Cadenas productivas con otras empresas.
Características de la Planeación.	Plan de Inversión dela Empresa.

Tabla 9. Modelo Resultante de las comparaciones entre los referentes.

4.4. MODELO DE LA HERRAMIENTA DE CLASIFICACIÓN.

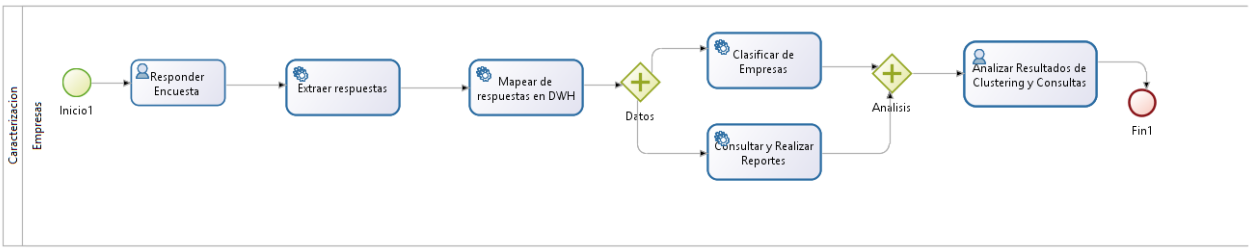

Figura 2. Modelo del proceso de clasificación de las empresas.

4.4.1. Responder Encuesta

En esta primera actividad del proceso, el empresario debe responder las preguntas que se encuentren en la encuesta, para facilitar este proceso se creó desde la página web www.e-encuesta.com, la definición de la encuesta. Posteriormente se realizó un envío masivo a los empresarios del link para que realizaran el registro de su información, respondiendo cada pregunta y en algunos casos debido a la falta de disponibilidad de los empresarios encuestados se optó por realizar la encuesta mediante un trabajo de campo, es decir, vía telefónica o de manera presencial. Para el presente estudio la base de datos con los resultados de la encuesta se encuentra almacenada en dicha página web.

4.4.2. Extracción y Mapeo de Respuestas

En este proceso se extrae la información de la base de datos de la encuesta y se almacena en la tabla de datos minables sobre el cual se continuará el proceso. La extracción, transformación y carga se realizó mediante Microsoft Excel 2010, con lo cual se generaron las instrucciones insert sobre una base de datos Oracle.

A continuación se visualiza el modelo creado para el almacenamiento de la información de las empresas y los datos minables pertenecientes al estudio y que se debe encontrar creado en Oracle y Oracle Data Miner.

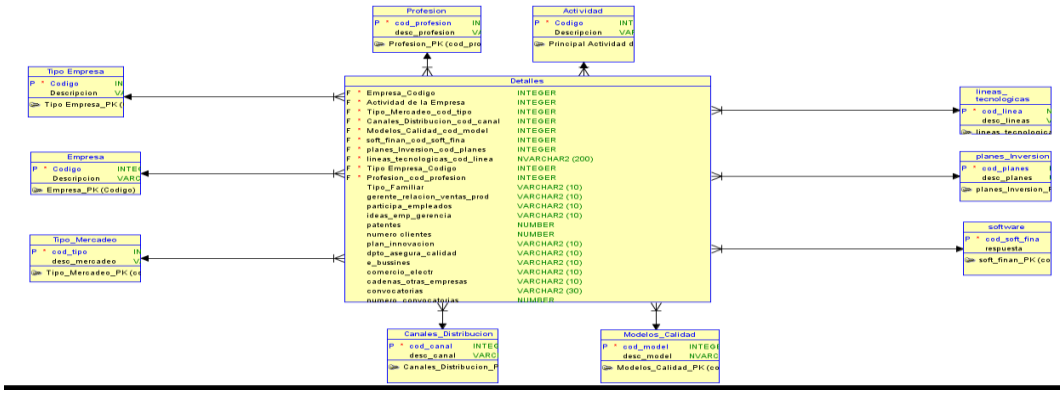

Figura 3. Tabla con información de las encuestas a las empresas.

En este modelo se tienen 10 tablas que se relacionan y almacenan su respectiva llave foránea en la tabla principal, en adelante tabla detalles. Las 10 tablas almacenan las respuestas múltiples y/o simples que el empresario debe de contestar en la encuesta.

Las 10 tablas son:

- Tipo de Empresa:

  En esta tabla se almacenan los diferentes tipos de empresas como: micro, pequeña y mediana.

- Profesión del Empresario:

  En esta tabla se almacenan las profesiones que pueden tener una persona, estas profesiones son: Ingeniero de Sistemas, Ingeniero de Multimedia, Administrador de Empresas, Ingeniero Electrónico,

Especialista en Negocios y Finanzas, Mercadologo, Economista e Ingeniero Industrial.

- Actividades de la empresa:

  En esta tabla se almacena la principal actividad o principales actividades que realice una empresa para satisfacer las necesidades de sus clientes y/o la sociedad en general, estas actividades son: Desarrollo de Aplicaciones, Cloud Computing, Servicios de Contact Center, Marketing Digital, Consultoría de IT, Integración de Sistemas, Soporte de Hardware y Software, Outsourcing de IT, Seguridad Informática, Seguridad de Redes y Calidad de Software.

- El Nombre de la Empresa:

  En esta tabla se almacena el nombre y/o razón social de las empresas que diligenciaron correctamente la encuesta diseñada con base al modelo de comparación.

- Tipo de Mercadeo:

  En esta tabla se almacenan los dos tipos de mercadeo que existen para este estudio y para la encuesta diseñada, los tipos de mercadeo existentes son: Mercadeo Operativo y Mercadeo Estratégico.

- Las líneas tecnológicas de la empresa:

  En esta tabla se almacena las líneas tecnológicas en que se enfoca una empresa, estas líneas tecnológicas son: Desarrollo de Aplicaciones, Cloud Computing, Servicios de Contact Center, Marketing Digital, Consultoría de IT, Integración de Sistemas, Soporte de Hardware y Software, Outsourcing de IT, Seguridad Informática, Seguridad de Redes y Calidad de Software.

- Los planes de inversión: La pregunta que se encuentra relacionada con este componente y se dirige a, ¿Su empresa tiene planes de inversión? Se encuentra como columna en la tabla detalles y sus respuestas pueden ser: Si ó No.

- El tipo de software que se maneja dentro de la organización: La pregunta para responder esta característica de la empresa es: ¿Que tipo de Software Financiero Maneja en su empresa? Las respuestas que puede contestar el encuestado son: Software Empresarial, Software Local, Software Financiero, Excel, Sql, etc. Estas respuestas se encuentran almacenadas en la tabla tipo_software.

- Los diferentes modelos de calidad que se pueden implementar en una organización:

  La tabla en el diseño tipo estrella, que almacena las respuestas a esta característica es: modelos_calidad, las respuestas que están almacenadas en esta tabla son: NTC6001, CMM-SW, CMMI, SPICE, ISO9001, PSP/TSP, No ha implementado, No sabe.

- Los canales de distribución:

  La Tabla que almacena las respuestas de esta característica es Canales_Distribucion, las respuestas que están almacenadas en esta tabla son: Ventas Directas, Representantes y Distribuidores, Sucursales, Internet, Otro.

La tabla de detalles o hechos, además de contener las llaves foráneas de las tablas de dimensiones mencionadas anteriormente, también tiene una serie de atributos que contesta a varias preguntas cerradas de la encuesta, como por ejemplo: ¿La empresa es de tipo familiar? Aquí las opciones de respuesta serán: Si, No ó no sabe.

Otro de los atributos de la tabla hechos es: ¿Si los empleados participaban activamente con sus ideas en proyectos que beneficie a la compañía? En esta pregunta el empresario deberá contestar: Si, No o No sabe. También se encuentra en este tipo de preguntas, ¿Si la compañía tiene departamento de calidad? ¿Si la compañía implementa comercio electrónico? ¿Si la compañía tiene cadenas productivas con otras empresas?

Se pueden observar atributos cuantitativos como: el número de patente de una compañía, el número de convocatorias en las que ha participado una compañía, el número de clientes de la compañía, si la compañía implementa planes de innovación. Así se completa la tabla de hechos del presente modelo en estrella de datos minables.

4.4.3. Clasificación de datos de las empresas

Se selecciona la herramienta: Oracle DataMiner, después de analizar las anteriores referencias en el marco conceptual, y se determina que es esta la que trabaja con los dos algoritmos de clustering.

Para el proceso de agrupamiento de la información se crea una vista con todos los datos minables desde las diferentes dimensiones de la tabla de datos y las columnas predefinidas en esta.

Al tener esta vista con los datos de las empresas, se utiliza los algoritmos de clustering, donde se encuentran los algoritmos K-Means y O-Cluster, como pueden observar en la Figura 4, en esta se encuentra un objeto representando la vista con los datos, conectándose directamente con otro objeto representando el proceso de clustering de la herramienta.

Al ejecutar el proceso de agrupación, los algoritmos generan dos árboles de decisión, con base en esto, al observar el árbol de decisión del algoritmo k-

Means, se realiza el análisis y se selecciona los grupos generados del clúster que se encuentran en el nivel medio, ya que, se llega a la conclusión, que es donde se almacena y clasifica la mayor parte de la información de las empresas. Estos agrupamientos son: Cluster2, Cluster8, Cluster9, Cluster12 y Cluster13.

Figura 4. Vista de los datos conectando al proceso de clustering.

En la figura 5, encontrará los resultados del proceso ejecutado con el algoritmo K-Means y donde se pueden observar los clúster con la información relevante de las empresas y que el algoritmo considera objeto de clasificación y fácil análisis, cabe afirmar que con un número más amplio de empresas e información los niveles de clasificación pueden tender a cambiar.

Se Concluye que la característica que más influye en el agrupamiento de la información es la actividad principal de la empresa, ya que, en cada nivel medio del árbol se notan cambios radicales e influyentes, que obliga a crear un nivel diferente de acuerdo a esta actividad.

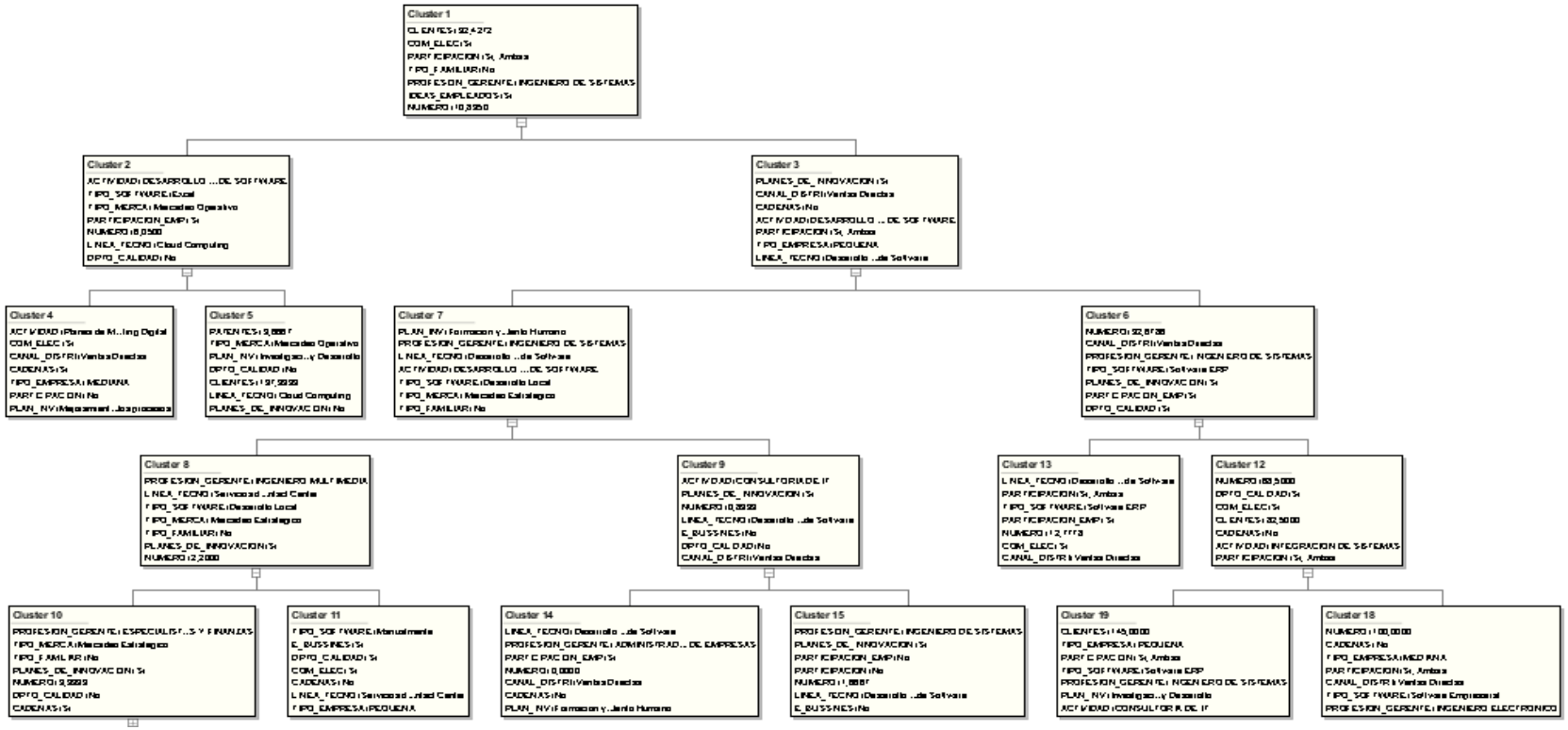

Figura 5. Árbol de Decisión del Algoritmo K-Means.

A continuación se explica los clúster generados:

Cluster 2: Detalla que la profesión del empresario es Ingeniero de Sistemas, Tipo de Empresa es Micro y Actividad de la Empresa es Cloud Computing y Desarrollo de Aplicaciones.

Cluster 8: Detalla que la profesión del empresario es Ingeniero Multimedia, el Tipo de Empresa es Pequeña y que la Actividad de la Empresa se basa en Desarrollo de Aplicaciones.

Cluster 9: Detalla que la profesión del empresario es Ingeniero de Sistemas, el Tipo de la Empresa es Micro y la actividad de la empresa se basa en Desarrollo de Aplicaciones, Calidad de Software y Consultoría de IT.

Cluster 12: Detalla que la profesión del empresario es Ingeniero de Sistemas, el Tipo de Empresa es Pequeña y la actividad de la empresa se basa en Integración de Sistemas y Outsourcing de IT.

Cluster 13: Detalla que la profesión del empresario es Administrador de Empresas y/o Ingeniero de Sistemas, el Tipo de Empresa es Pequeña y no evalúa Actividad de la empresa.

Con la descripción de las variables mas influyentes de los algoritmos se puede afirmar que la actividad dentro de las tecnologías de la información y comunicaciones en una empresa del grupo multisectorial, es el desarrollo de aplicaciones, teniendo en cuenta, igualmente que no todas se dedican al desarrollo pero si es una de las más destacadas dentro de este pequeño grupo de empresas.

Algoritmo K-Means					
Caracteristicas / Clusters	Cluster 2	Cluster 8	Cluster 9	Cluster 12	Cluster13
Profesión del Gerente/Empresario de la Empresa.	Ingeniero de Sistemas	Ingeniero Multimedia	Ingeniero de Sistemas	Ingeniero de Sistemas	Ingeniero de Sistemas y Administrador
Tipo de Empresa	Micro	Pequeña	Micro	Pequeña	Pequeña
Empresa de Tipo Familiar		No			
Principal actividad del gerente/empresario de la empresa.					
Principal Actividad de la Empresa	Cloud Computing y Desarrollo de Aplicaciones	Desarrollo de Aplicaciones	Desarrollo de Aplicaciones, Consultoria de IT y Calidad de Software	Integracion de Sistemas, OutSourcing de IT	
Habilidades del Gerente/Empresario relacionados con las ventas del producto y/o servicio fuera del país.					
Participación de los empleados en la toma de decisiones e ideas de la empresa.	Si				Si
Patentes registradas por la empresa.	5				
Canales de Distribución.	Ventas Directas		Ventas Directas	Ventas Directas	Ventas Directas
Número de Clientes de la Empresa.	353			82	
Tipo de Mercadeo de la Empresa.	Mercadeo Operativo	Mercadeo Estrategico			
Departamento de Aseguramiento de Calidad	No		No	Si	
Implementación de modelos de Calidad.					
Actividades TIC de la Empresa.	Desarrollo de Aplicaciones	Desarrollo de Aplicaciones	Desarrollo de Aplicaciones	Consultoria de IT	Desarrollo de Aplicaciones
Realiza E-Bussines en la empresa			No		
Tipo de Software en el Manejo Financiero	Excel	Desarrollo Local			Software ERP
Cadenas productivas con otras empresas.			No	No	Si
Planes de Innovacion		Si	Si		
Plan de Inversión dela Empresa.		Formacion y Desarrollo del Talento Humano		Investigacion y Desarrollo	
Realiza Comercio Electronico en la Empresa		Si		Si	Si
Número de Convocatorias Ejecutadas		3	0		12
Participacion en Convocatorias		Si, Pública y Privada	No	Si, Pública y Privada	Si, Pública y Privada
Total Empresas Clasificadad	3	4	5	4	5

Tabla 10. Tabla de Agrupaciones de las Empresas con el modelo de características y algoritmo k-Means.

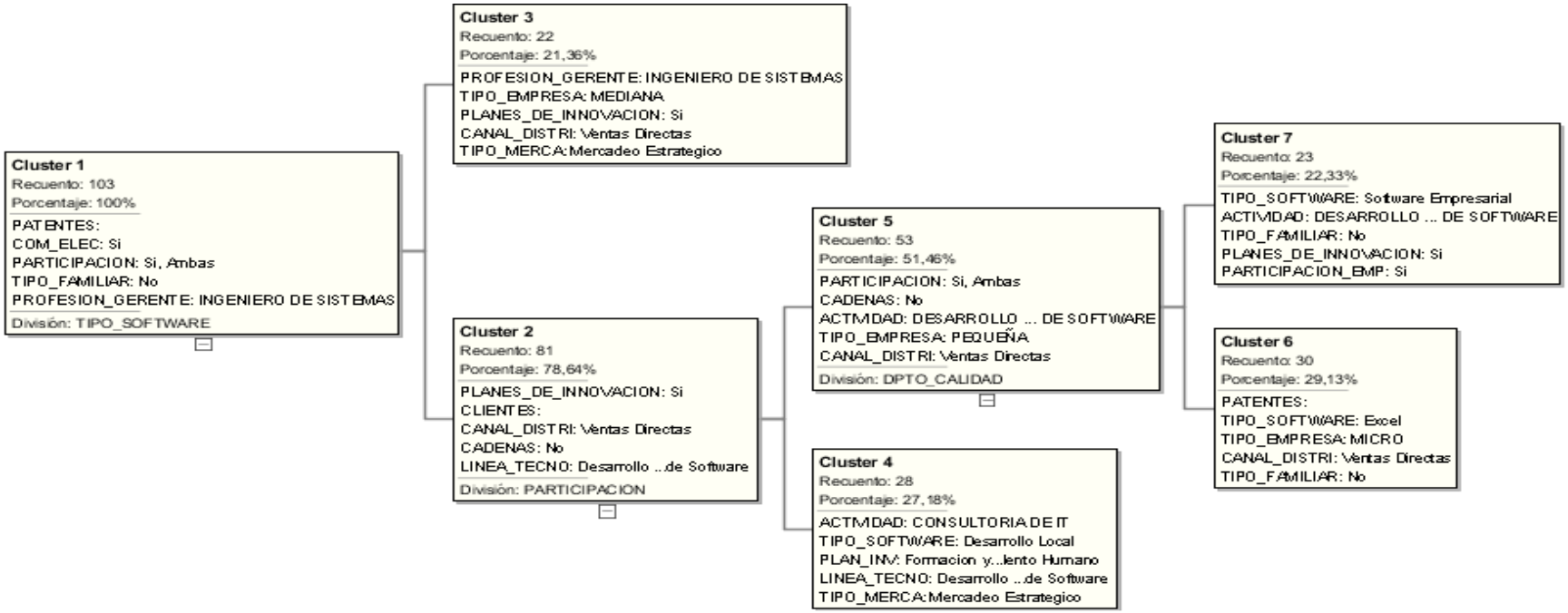

Figura 6. Árbol de Decisión implementado por el Algoritmo O-Cluster.

En la figura 6, se observa el árbol de decisión generado con el algoritmo O-Cluster, en esta figura se ve que se generan menos clúster que con el otro algoritmo, sin embargo, se visualiza que la característica dentro del modelo más influyente es la actividad principal de la empresa y/o la línea tecnológica, es allí, donde el algoritmo toma como base la actividad, la línea tecnológica y la profesión del gerente, y que se puede concluir que guían el agrupamiento de esta información.

Se realiza una tabla comparativa de los Clústeres seleccionados (Cluster 3, 4, 6,7) y donde se observa que existe una gran clasificación de datos y se puede analizar cuáles son los datos más importantes que influye en este pequeño grupo de empresas pertenecientes a un grupo empresarial. Para una visualización de los datos de clasificación, observar la tabla No. 11.

Algoritmo O-Cluster				
Caracteristicas / Clusters	Cluster 3	Cluster 4	Cluster 6	Cluster 7
Profesión del Gerente/Empresario de la Empresa.	Ingeniero de Sistemas	Administrador de Empresas o Ingeniero de Sistemas	Ingeniero de Sistemas, Multimedia o Industrial	Ingeniero de Sistemas
Tipo de Empresa	Mediana	Pequeña	Micro	
Empresa de Tipo Familiar			No	No
Principal actividad del gerente/empresario de la empresa.				
Principal Actividad de la Empresa	Desarrollo de Aplicaciones, Consultoria de IT y Cloud Computing	Desarrollo de Aplicaciones, Consultoria de IT, Calidad de Software	Desarrollo de Aplicaciones, Consultoria de IT, Calidad de Software	Desarrollo de Aplicaciones
Habilidades del Gerente/Empresario relacionados con las ventas del producto y/o servicio fuera del país.				
Participación de los empleados en la toma de decisiones e ideas de la empresa.	Si	Si		Si
Patentes registradas por la empresa.	0			
Canales de Distribución.	Ventas Directas	Ventas Directas	Ventas Directas	
Número de Clientes de la Empresa.				
Tipo de Mercadeo de la Empresa.	Mercadeo Estrategico	Mercadeo Estrategico		
Departamento de Aseguramiento de Calidad			No	Si
Implementación de modelos de Calidad.				
Lineas Tecnologicas de la Empresa.	Desarrollo de Aplicaciones, Consultoria de IT y Cloud Computing	Desarrollo de Aplicaciones de Software	Desarrollo de Apps, Consultoria de IT, Integracion de Sistemas, Soporte Hardware y Software	Outsourcing de IT, Calidad de Software, Integracion de Sistemas, Servicios de Contact Center
Realiza E-Bussines en la empresa				Si
Tipo de Software en el Manejo Financiero	Software ERP	Desarrollo Local	Excel	Software Empresarial
Cadenas productivas con otras empresas.				
Planes de Innovacion	Si	Si		Si
Plan de Inversión de la Empresa.		Formacion y Desarrollo del Talento Humano	Investigacion y Desarrollo	
Realiza Comercio Electronico en la Empresa				Si
Número de Convocatorias Ejecutadas		0		
Participacion en Convocatorias		No	Si, Convocatoria Pública	Si, Convocatoria pública y privada
Total Empresas Clasificadas	4	6	6	5

Tabla 11. Tabla Comparativa de las características con los Cluster del algoritmo O-Cluster.

Los clúster generados por el algoritmo O-Clúster son:

- ✓ Clúster 3: Muestra que la profesión del empresario es Ingeniero de Sistemas, El Tipo de Empresa es Mediana y la actividad de la empresa es Desarrollo de Aplicaciones, Consultoría de IT y Cloud Computing.
- ✓ Clúster 4: Muestra que la profesión del empresario puede ser Administrador de Empresas o Ingeniero de Sistemas, el tipo de empresa es pequeña y la actividad de la empresa es Desarrollo de Aplicaciones, Consultoría de IT y/o Calidad de Software.
- ✓ Clúster 6: Muestra que la profesión del empresario puede ser Ingeniero de Sistemas, Multimedia o Industrial, el Tipo de Empresa es Micro y las actividades de la empresa es Desarrollo de Aplicaciones, Consultoría de IT y/o Calidad Software.
- ✓ Clúster 7: Muestra que la profesión del empresario es Ingeniero de Sistemas, El Tipo de Empresa el algoritmo no lo toma como característica de división principal y las actividades tecnológicas es Desarrollo de Aplicaciones.

De acuerdo a las divisiones de estos agrupamientos se puede afirmar que el algoritmo toma en cuenta los tres primeros datos minables, el tipo de empresa, priorizó las empresas que se dedican al desarrollo de software y áreas relacionadas a estas, como calidad de software y consultoría de IT. Esto significa que el algoritmo está agrupando la información lógicamente, dado que estas áreas van por la misma línea tecnológica y/o dentro de una compañía que se dedica a la construcción de software es fundamental estas tres áreas ya que se encuentran contenidas en el proceso de desarrollo de un producto de software.

4.5. DISEÑO DE CONSULTAS Y REPORTES.

Estas consultas y reportes permiten mostrar otra perspectiva de la información, empleando conceptos de Business Intelligence.

Las gráficas presentadas en esta sección se realizaron con una herramienta, Tableau Public 8.0, que permite realizar una conexión desde una hoja de cálculo o un archivo de una base de datos externa y crear visualizaciones de datos interactivos para que sean visualizados y analizados en la web, cabe afirmar que restringiendo y permitiendo quien ve la información que se va generando.

4.5.2. PROTOTIPOS DE CONSULTAS Y REPORTES

El tipo y la actividad de la empresa como se presenta en la figura 7, observa la clasificación de los diferentes tipos de empresas (micro, pequeña y mediana) y la actividad principal que más desarrollan en el mercado de las tecnologías de la información y comunicaciones.

Una de las preguntas que se encuentran en la encuesta son:

- ¿A que tipo de empresa pertenece la compañia?

La tabla detalles diseñada soporta esta consulta ya que le sugiere al empresario respuestas sencillas como:

Micro – Pequeña – Mediana

Otra de las preguntas relacionadas en la figura 4 es:

- ¿Cuál es la principal actividad de la empresa?

Sugiriéndole al empresario respuestas como:

Integración de Sistemas – Desarrollo de aplicaciones de software – Consultoría de IT – Soporte de Hardware y Software – Outsourcing de IT y Calidad de Software.

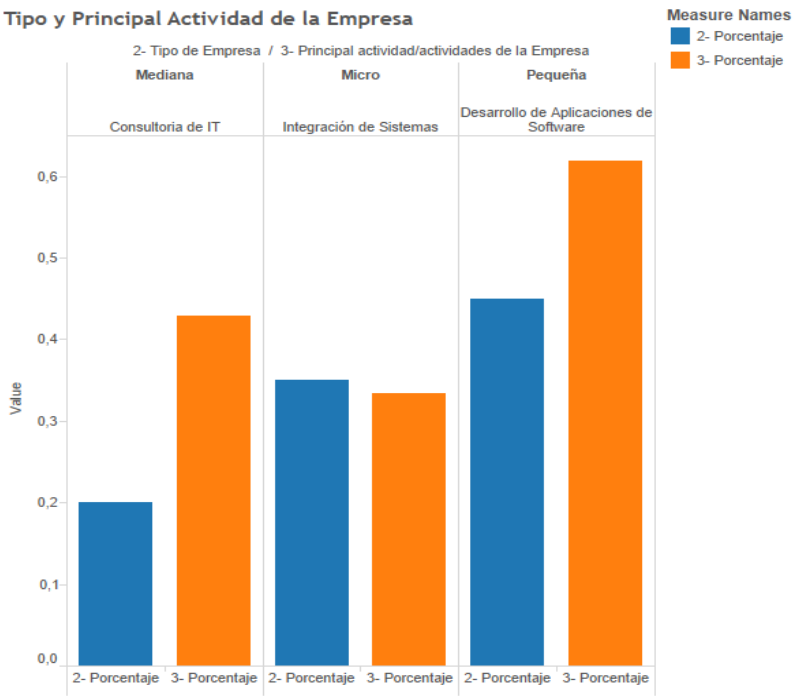

Figura 7. Prototipo de Consulta y/o reporte: Tipo y actividad de la empresa.

La figura 8, presenta la clasificación del perfil del gerente de la empresa de acuerdo a su nivel profesional, aquí se puede ver una clasificación de datos totalmente diferente. La visualización se realiza con el perfil del empresario y el porcentaje de cada perfil de acuerdo a las empresas que hayan contestado por ejemplo que el perfil del empresario es Ingeniero de Sistemas.

Para realizar este tipo de grafica, se realiza la siguiente pregunta a los empresarios y que se encuentra en la encuesta como anexo en este documento.

- Profesión del Gerente y/o Empresario de la empresa

Esta pregunta se deja abierta, para que el empresario tenga la libertad de contestar su título profesional, técnico o tecnológico. Esto se realiza para que haya múltiples respuestas dentro del estudio y/o se defina que dentro de una compañía relacionada con las tecnologías de la información y comunicaciones tenga prioridad profesiones técnicas, tecnológicas e ingenierías relacionadas.

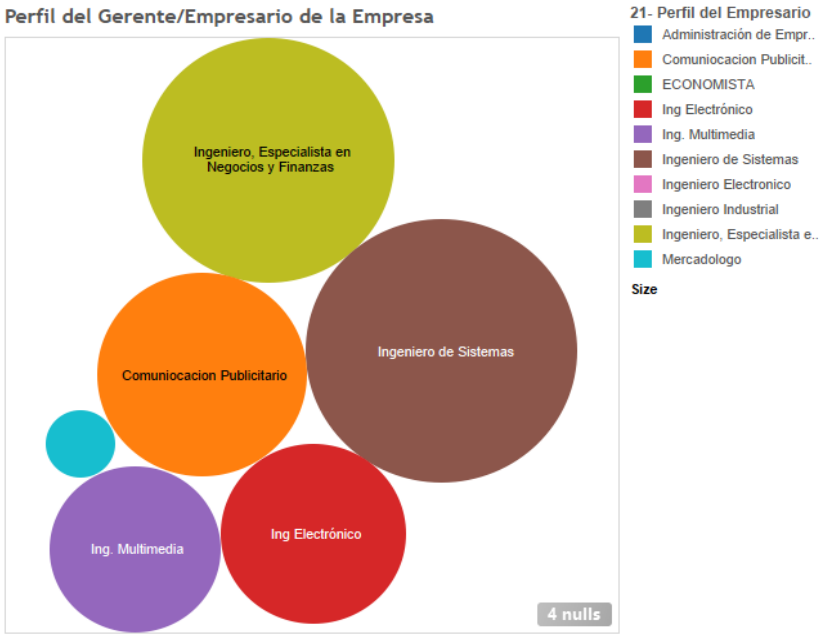

Figura 8. Prototipo de consulta y/o reporte: clasificación del perfil del gerente.

En la figura 9, se puede observar otra de las distintas graficas que nos ofrece la herramienta y que en forma de columnas presenta los datos de las líneas tecnológicas a las que se dedican las empresas encuestadas. Para realizar esta grafica se utiliza la pregunta: ¿cuál o cuáles son las líneas tecnológicas a la que se dedica su empresa? Sugiriéndole al empresario respuestas múltiples como:

Seguridad informática, servicio de redes, calidad de software, Outsourcing de IT soporte de hardware y software, integración de sistemas, consultoría de IT, desarrollo de aplicaciones de software y un campo abierto para que el empresario

escriba una respuesta que no se encuentre dentro de las sugerencias, en la grafica se muestra el porcentaje de cada línea tecnológica, otra de las cosas que se puede realizar es presentar los tipos de empresas y las líneas tecnológicas de las empresas en este estudio.

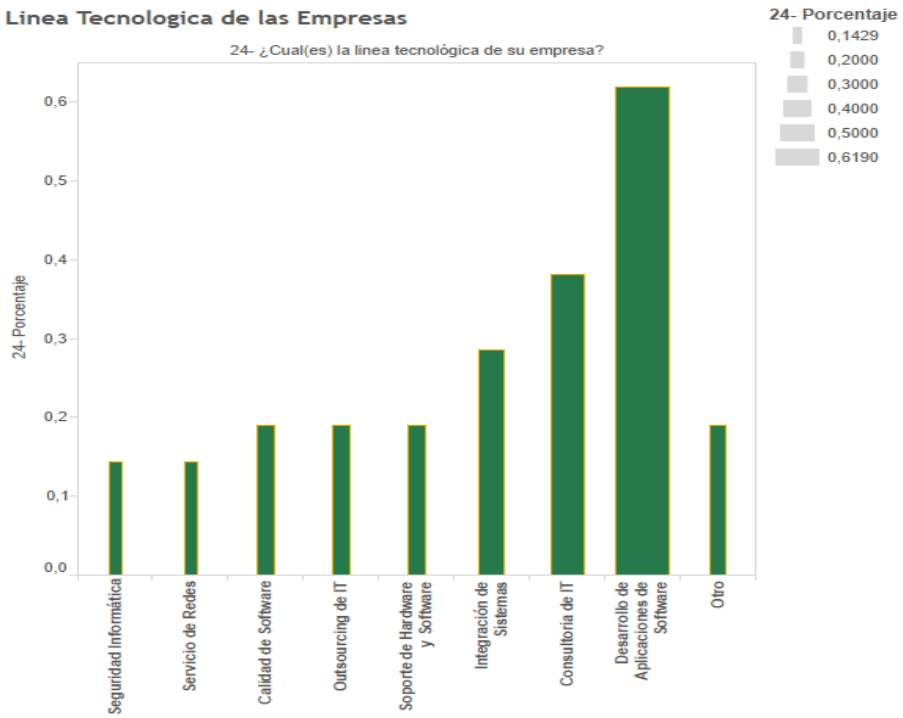

Figura 9. Prototipo de consulta y/o reporte: porcentaje de las líneas tecnológicas de las empresas.

5. CONCLUSIONES

- ✓ Se logró definir una encuesta utilizada en la obtención de datos de las empresas tic, agrupando los factores y variables resultantes del modelo de caracterización empresarial basado en la armonización de los estudios previos.
- ✓ Los análisis realizados de los datos cuando se interpretan con un experto empresarial, tienen coherencia porque este puede evaluar la clasificación obtenida. **(ver anexo)**
- ✓ Los prototipos y consultas obtenidos, se analizan con un experto empresarial obteniendo como resultado comentarios positivos del despliegue de la información en las gráficas obtenidas. La agrupación empresarial generada en este estudio permite a cada empresa que aporto sus datos en el proceso identificar en que sector TIC se encuentra y conocer su entorno al tomar decisiones en la participación de una convocatoria o licitación.
- ✓ El modelo de agrupación realizado puede interesarle a entidades como el ministerio de tics, el ministerio de trabajo y el programa de transformación productiva del gobierno por qué se puede implementar a empresas y grupos de empresarios nacionales.
- ✓ Este proyecto puede ser un modelo de referente para estudios futuros con empresas del sector TIC y otros sectores empresariales.
- ✓ El modelo de agrupación propuesto en este proyecto le permite al grupo multisectorial conocer más sobre herramientas de agrupación de información y los beneficios que de ellas pueden obtener.

6. RECOMENDACIONES PARA TRABAJOS FUTUROS.

- ✓ Se debe realizar una actualización semestral o anual de la recolección de datos de las empresas pertenecientes a este estudio y las nuevas que deseen saber hacia dónde van dirigiendo su organización en las diferentes líneas tecnológicas del sector para poder extender y agrupar las características de las mismas.
- ✓ Se puede desarrollar una herramienta automatizada basada en el modelo de agrupación planteado en este trabajo.
- ✓ El modelo de agrupación planteado se puede utilizar para empresas de sectores ajenos a las TICs y otros grupos empresariales.
- ✓ Con este modelo de agrupación se puede hacer una herramienta que pueda servir para apoyar la toma de decisión y la participación en convocatorias.

ANEXOS

- Encuesta de Estudio de Caracterización de Empresas de Tecnologías de la Información y Comunicaciones.

- Scripts de Creación de Diseño de Base de Datos, Tipo Estrella.

- Comparación de Factores y Variables en Excel.

- Encuestas a Expertos Empresariales sobre resultados del Estudio de Caracterización y Agrupación.

BIBLIOGRAFIA

[1] C. Pardo, F. García, F. J. Pino, M. Piattini, and M. T. Baldassarre, “Método de integración para soportar la armonización de múltiples modelos y estándares.”

[2] U. A. Ms. U. de S. B. S. C. Merchan Luis, MSc., “Caracterizacion de las Empresas Pertenecientes a la Industria Emergente de Software del Sur Occidente Colombiano Caso Red de Parques PARQUESOFT,” vol. 4, no. 2, 2007.

[3] C. González, Gómez Hugo and J. Castellanos, Giraldo David, “Modelo de Emprendimiento de la Industria de las Tecnologias de la Informacion (TI) para Colombia.,” p. 213, 2012.

[4] F. Programa de Transformación Productiva, Ministerio de Comercio Industria y Turismo, ESIeCenter SinerTic, “Generacion de Estrategias Para El Desarrollo Tecnologico y de Mercado Del Sector Software y Servicios de TI Mediante la Aplicacion de Vigilancia Tecnologica y Prospectiva.,” p. 44, 2011.

[5] I. Industrial, “Estudio del alcance de la implantación de tecnologías de información, como apoyo al mejoramiento de los procesos, en las pequeñas y medianas empresas del sector manufacturero en bogotá.,” 2003.

WEBGRAFIA

- ✓ http://www.microsoft.com/en-us/server-cloud/solutions/business-intelligence/predictive-analytics.aspx#fbid=vCouGblU6Bw

- ✓ http://office.microsoft.com/en-us/excel-help/data-mining-add-ins-HA010342915.aspx

- ✓ http://www.tableausoftware.com/public/

Encuesta a Experto Empresarial sobre Estudio de Caracterización y Clasificación de las empresas de un grupo empresarial de Santiago de Cali.

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL CALI

Experto Empresarial

Carlos Octavio Muñoz

Ms. Ingeniería de Sistemas

Grupo Multisectorial

1. ¿Qué piensa de la Comparación de los diferentes estudios de caracterización empresarial contenida en este proyecto?

   R/ Considero que efectivamente para este proyecto se tomaron como base diferentes modelos, algunos que ya se han utilizado en el tema de las TICs específicamente, otros que se han tomado a nivel de estudios de mipymes en otras ciudades importantes del país y sirvieron precisamente como base para definir el modelo final de caracterización empresarial.

2. ¿Qué piensa del modelo de clasificación utilizado en el proyecto?

   R/ el modelo en realidad lo que hace es resaltar la información relevante que se requiere para tener un conocimiento claro en el sector de las TIC aunque mucha de esta información también es válido para cualquier otro sector.

3. ¿Cree que los prototipos de consulta elaborados en este proyecto, le dan una visión de las empresas del grupo?

R/ la realidad es que ya teniendo una caracterización, una información completa de las empresas del grupo multisectorial si permite en realidad tener una vision cuál es la situación actual y hacia donde tenemos que llevar el sector TIC. Por ende es de suma importancia contar con estos prototipos.

4. ¿La clasificación obtenida mediante los algoritmos de clustering le da una visión de las empresas para la participación en convocatorias?

R/ Estos algoritmos de clustering básicamente en algunos casos lo que hace es darme una visión para ubicar en que sector específico de TIC tengo que involucrarme, si estoy en consultoría o a nivel de desarrollo y obviamente si yo como empresario tengo claro en qué sector estoy pues me va a facilitar mucho mirar cuales son las convocatorias que en realidad puedo presentar.

5. ¿Este proyecto aporta al grupo a tener nuevas características de las empresas para clasificarlas?

R/ Considero que si, definitivamente si tenemos unos sectores mucho más claros y una clasificación mucho más definida podemos concentrar a las diferentes empresas y sacar en realidad el mayor provecho de ellas para obtener unos resultados de crecimiento y de visión internacional de productos y servicios de estas empresas.

Recomendaciones:

Como recomendación, considero importante de que esta caracterización que se hace a nivel del grupo multisectorial se puede aplicar a nivel un poco más macro, es decir, aplicar a nivel de un cluster de pacifitic que tiene un mayor número de empresas o inclusive mirar que el ministerio de las tic con el programa FITI o ministerio de comercio a través del programa PTP puedan utilizar este modelo de caracterización empresarial para aplicarlo a los sectores TIC y BPO.

Encuesta a Experto Empresarial sobre Estudio de Caracterización y Clasificación de las empresas de un grupo empresarial de Santiago de Cali.

UNIVERSIDAD DE SAN BUENAVENTURA SECCIONAL CALI

Experto Empresarial

José William González

secretario@grupomultisectorial.com

Director de Proyectos

Grupo Multisectorial

1. ¿Qué piensa de la armonización de los diferentes estudios de caracterización empresarial contenida en este proyecto?

   Es una buena comparación de modelos que permite tener referentes importantes y así definir uno integral y mejorado desde los conceptual.

2. ¿Qué piensa del modelo de clasificación utilizado en el proyecto?

   Permite tener desde diferentes aspectos una visión clara del sector Tic y su realidad

3. ¿Cree que los prototipos de consulta elaborados en este proyecto, le dan una visión de las empresas del grupo?

   No solo una visión, nos permite conocer hacia donde se dirige este sector.

4. ¿La clasificación obtenida mediante los algoritmos de clustering le da una visión de las empresas para la participación en convocatorias?

   Si, nos da algunos elementos que permiten saber si aplican o no a las diferentes convocatorias

5. ¿Este proyecto aporta al grupo a tener nuevas características de las empresas para clasificarlas?

   Sí, claro nos da diferentes características para el análisis

Recomendaciones:

Que este trabajo se enmarque como un referente de estudio