

**APLICACIÓN DE TÉCNICAS DE VISUALIZACIÓN DE DATOS PARA LA
INTELIGENCIA DE NEGOCIOS**

**CARLOS GERMAN CHAUCANES MONTENEGRO
RIDER ALEXANDER SUÁREZ ALFARO**



**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2015**

**APLICACIÓN DE TÉCNICAS DE VISUALIZACIÓN DE DATOS PARA LA
INTELIGENCIA DE NEGOCIOS**

**CARLOS GERMAN CHAUCANES MONTENEGRO
RIDER ALEXANDER SUÁREZ ALFARO**

**Trabajo de grado presentado como requisito parcial para optar al título de
Ingeniero de Sistemas**

**DIRECTOR:
SILVIO RICARDO TIMARÁN PEREIRA, Ph.D.**

**UNIVERSIDAD DE NARIÑO
FACULTAD DE INGENIERÍA
PROGRAMA DE INGENIERÍA DE SISTEMAS
SAN JUAN DE PASTO
2015**

NOTA DE RESPONSABILIDAD

Las ideas y conclusiones aportadas en este Trabajo de Grado son Responsabilidad de los autores.

Artículo 1 del Acuerdo No. 324 de octubre 11 de 1966, emanado del honorable Concejo Directivo de la Universidad de Nariño.

“La Universidad de Nariño no se hace responsable de las opiniones o resultados obtenidos en el presente trabajo y para su publicación priman las normas sobre el derecho de autor”.

Artículo 13, Acuerdo N. 005 de 2010 emanado del Honorable Consejo Académico.

Nota de Aceptación:

Firma del Presidente del Jurado

Firma del Jurado

Firma del Jurado

San Juan de Pasto, Mayo de 2015

AGRADECIMIENTOS

Quiero Agradecer en primer lugar a Dios, por haber hecho posible la culminación de uno de mis sueños, por todo su apoyo en los momentos más difíciles y por ser el guía de mi vida. GRACIAS DIOS.

Agradezco a mis padres: CECILIA MONTENEGRO y CARLOS CHAUCANES, por confiar en mis capacidades, por apoyarme en todo lo necesario en este proceso de formación, por ser mis primeros educadores y el ejemplo a seguir. GRACIAS QUERIDOS PADRES.

A mis hermanas: CLAUDIA y JANETH, por brindarme apoyo en aquellos momentos de dificultad, por haber cuidado de mis padres mientras estaba ausente y por estimular mis logros. GRACIAS QUERIDAS HERMANAS.

Agradezco la confianza, apoyo y dedicación a mis profesores: RICARDO TIMARÁN PEREIRA, quien hizo posible mi vinculación a un grupo de investigación, en el cual adquirí muchas competencias investigativas; al profesor ANDRÉS OSWALDO CALDERON, quien con sus conocimientos, me apoyo en el proceso de desarrollo de esta investigación y en general a todos mis profesores, quienes hicieron de mí una persona con grandes conocimientos y virtudes.

A mis compañeros del grupo de investigación aplicada en sistemas GRIAS, por compartir sus conocimientos conmigo y por brindarme su apoyo incondicional.

A mi mejor amiga Vicky, por sus valiosos consejos que fueron de gran ayuda para lograr esta meta.

A mis compañeros de carrera, quienes también fueron mis amigos, les agradezco por haber compartido momentos agradables y sobre todo por la ayuda incondicional en los momentos de dificultad.

CARLOS GERMAN CHAUCANES MONTENEGRO

Le agradezco a Dios por ser siempre mi guía y compañía en cada momento de mi vida; dándome así, la fuerza necesaria para alcanzar mis metas con éxito.

A mis padres ANA CECILIA ALFARO y JOSÉ PORFIRIO SUÁREZ por su apoyo incondicional, por haberme dado la oportunidad de tener una buena educación, por sus consejos, por creer siempre en mí y por ser un ejemplo a seguir.

Le agradezco a mi hermano JOSE GERARDO SUÁREZ ALFARO por ser una parte importante en mi vida, por creer en mí, por darme sus consejos que me ayudaron a ser cada día mejor y poder sacar a adelante mi carrera.

Al profesor RICARDO TIMARAN PEREIRA, Ph.D. por haber sido el asesor de mi trabajo de grado e inculcarme el deseo e interés por la investigación.

Al profesor ANDRES CALDERON por su apoyo incondicional en el desarrollo de este trabajo de grado y en cada momento de mi carrera.

Le agradezco a la Universidad de Nariño por permitirme desarrollar mi carrera, formándome como una persona íntegra con valores humanos y conocimiento profesional.

A todos mis amigos, profesores y familiares que creyeron en mí, que con su apoyo y consejos me dieron la fuerza para concluir con esta etapa de mi vida profesional.

RIDER ALEXANDER SUÁREZ ALFARO

DEDICATORIA

Este trabajo de grado se lo dedico a mis padres, quienes me apoyaron en todo lo necesario para lograr este gran sueño y por ser mis primeros formadores tanto en lo personal como profesional.

De igual manera se lo dedico a mis dos abuelitos que están en el cielo, que desde pequeño siempre me apoyaron y confiaron en mis capacidades.

CARLOS GERMAN CHAUCANES MONTENEGRO

Dedico este trabajo a Dios, a mi mamá Ana Cecilia Alfaro, a mi papá José Porfirio Suárez y a mi hermano José Gerardo Suárez Alfaro.

RIDER ALEXANDER SUÁREZ ALFARO

RESUMEN

Las técnicas de visualización de datos desempeñan un papel muy importante en las organizaciones y permiten mediante el uso de imágenes y gráficos visualizar los datos para hacerlos más comprensibles, atractivos, manejables y útiles.

En este documento se presenta el análisis de la aplicación de técnicas de visualización de datos para inteligencia de negocios, algunas fueron implementadas en la herramienta VisualBITool. Esta herramienta, permite la visualización de datos para las operaciones de análisis multidimensional y las tareas de minería de datos tales como, clasificación basada en árboles de decisión, reglas de asociación y agrupamiento.

La arquitectura de VisualBITool está conformada por tres módulos: Interfaz gráfica, kernel y conexión. El módulo de interfaz gráfica, permite la interacción entre la herramienta y el usuario. En el kernel se encuentran implementados los algoritmos, técnicas de visualización y el módulo de conexión, el cual permite la recuperación de información desde archivos y bases de datos construidas con el sistema gestor de bases de datos PostgreSQL.

Esta herramienta desarrollada bajo software libre está dirigida a las PYMES Y MIPYMES que deseen utilizarla para ayudarse a soportar la toma de decisiones y de esta manera volverse más competitivas.

ABSTRACT

Data visualization techniques play a very important role in organizations and allow through the use of images and graphics display data to make them more understandable, attractive, manageable and useful.

This paper presents the analysis of the application of data visualization techniques for business intelligence, some were implemented in the VisualBITool tool. This tool allows the data visualization for multidimensional analysis operations and data mining tasks such as classification based on decision trees, association rules and Clustering.

VisualBITool architecture is comprised of three modules: GUI, kernel and connection. The GUI module, allows for interaction between the tool and the user. In the kernel are implemented algorithms, visualization techniques and connection module, which allows information retrieval from files and databases built with the management system PostgreSQL database.

This tool developed under free software is aimed at PYMES and MIPYMES, who wish to use to help support decision making and thus become more competitive.

CONTENIDO

	Pág.
INTRODUCCIÓN	24
1. FUNDAMENTOS TEÓRICOS	27
1.1 VISUALIZACIÓN.....	27
1.1.1 Visualización de datos.	28
1.1.2 Importancia de la visualización de la información.	28
1.2 EL MODELO UNIFICADO DE VISUALIZACIÓN	28
1.3 TÉCNICAS DE VISUALIZACIÓN.....	30
1.3.1 Relaciones entre los puntos de datos.	30
1.3.2 Comparar un conjunto de valores.	38
1.3.3 Seguimiento del sube y baja con el tiempo.....	47
1.3.4 Ver las partes de un todo.....	47
2. SELECCIÓN DE LAS TÉCNICAS DE VISUALIZACIÓN DATOS PARA LA HERRAMIENTA VISUALBITOOL	59
3. METODOLOGÍA.....	73
3.1 SELECCIÓN DE LA METODOLOGÍA	73
3.1.1 SCRUM.....	74
3.1.2 UML..	76
3.2 IDENTIFICACIÓN DE REQUERIMIENTOS.....	77
3.2.1 Requerimientos funcionales.....	77
3.3 CASOS DE USO.....	85
3.3.1 Descripción de los casos de uso.....	85
3.3.2 Diagramas de los casos de uso..	90
3.4 DIAGRAMAS DE PAQUETES	95
3.4.1 Paquete principal..	95
3.4.2 Paquete File.....	96

3.4.3	Paquete QualityData.....	96
3.4.4	Paquete DataMining.	97
3.4.5	Paquete Dwh..	98
3.5	DIAGRAMAS DE CLASES	98
3.5.1	Clases del Paquete File.	98
3.5.2	Clases del Paquete QualityData..	99
3.5.3	Clases del Paquete DataMining.....	102
3.5.4	Clases del Paquete Dwh.....	105
4.	ARQUITECTURA DE VISUALBITOOL.....	109
4.1	ARQUITECTURA GENERAL.....	109
4.2	ARQUITECTURA INTERNA.....	110
4.2.1	Módulos de la aplicación.....	111
4.2.2	Descripción de los paquetes por módulo.	112
5.	PRUEBAS Y EVALUACIÓN DE FUNCIONALIDAD.....	118
5.1	PRUEBAS PARA LA TAREA DE ANÁLISIS PREVIO DE ATRIBUTOS...118	
5.1.1	Evaluación de las pruebas de funcionalidad de atributos del módulo calidad de datos.	123
5.2	PRUEBAS PARA LA TAREA DE ANÁLISIS PREVIO DE DATOS	123
5.2.1	Evaluación de las pruebas de funcionalidad de los datos (instancias) del módulo calidad de datos.	126
5.3	PRUEBAS PARA LA TAREA DE FILTROS, APLICADOS SOBRE LOS DATOS.	127
5.3.1	Evaluación de las pruebas de funcionalidad para el módulo de filtro.....	129
5.4	PRUEBAS PARA LA TAREA DE ÁRBOLES DE DECISIÓN	130
5.4.1	Evaluación de las pruebas de funcionalidad de árboles de decisión	132
5.5	PRUEBAS PARA LA TAREA DE AGRUPAMIENTO (CLUSTERING).....	133
5.5.1	Evaluación de las pruebas de funcionalidad para agrupamiento	137
5.6	PRUEBAS PARA LA TAREA DE REGLAS DE ASOCIACIÓN.	137
5.6.1	Evaluación de las pruebas de funcionalidad para reglas de asociación	140

5.7	PRUEBAS PARA LA TAREA DE ANÁLISIS OLAP	140
5.7.1	Evaluación de las pruebas de funcionalidad para el proceso de análisis OLAP.....	149
6.	CONCLUSIONES.....	150
7.	RECOMENDACIONES	151
	REFERENCIA BIBLIOGRÁFICA	152
	ANEXOS	158

LISTA DE FIGURAS

	Pág.
Figura 1. Modelo unificado de visualización	29
Figura 2. Diagrama de dispersión	31
Figura 3. Diagrama de dispersión simple.....	32
Figura 4. Diagrama de dispersión superpuesto	32
Figura 5. Diagrama matricial	33
Figura 6. Diagrama matricial.....	33
Figura 7. Diagrama Dispersión con sólo marcadores	34
Figura 8. Diagrama Dispersión con líneas suavizadas y dispersión con líneas suavizadas y marcadores	34
Figura 9. Diagrama Dispersión con líneas rectas y dispersión con líneas rectas y marcadores	35
Figura 10. Diagrama de Voronoi	35
Figura 11. Gráfico de matriz.....	36
Figura 12. Gráfico de diagrama de red	37
Figura 13. Gráfico de columnas	38
Figura 14. Gráfico de Columnas agrupadas y columnas agrupadas en 3D	39
Figura 15. Gráfico de Columnas apiladas y columnas apiladas en 3D	40
Figura 16. Gráfico de Columnas 3D.....	40
Figura 17. Gráfico de Cilindro, cono y pirámide	41
Figura 18. Gráfico de Barra agrupada y barra agrupada en 3D.....	41
Figura 19. Gráfico de Barra apilada y barra apilada en 3D	42
Figura 20. Gráfico de Cilindro, cono y pirámide horizontales.....	42
Figura 21. Gráficos de cotizaciones	43
Figura 22. Gráficos de Máximos, mínimos y cierre	43
Figura 23. Gráficos de Máximos, mínimos y cierre	44

Figura 24. Gráficos de Volumen, máximos, mínimos y cierre	44
Figura 25. Gráficos de Volumen, apertura, máximos, mínimos y cierre.....	44
Figura 26. Gráficos de caja y bigotes.....	45
Figura 27. Histograma de bloques	46
Figura 28. Gráfico de burbujas.....	47
Figura 29. Gráfico circular.....	48
Figura 30. Gráfico Circular y circular en 3D	49
Figura 31. Gráfico Circular con subgráfico circular y circular con subgráfico de barras	49
Figura 32. Gráfico Circular seccionado y circular seccionado en 3D	50
Figura 33. Gráfico de embalaje de círculos	50
Figura 34. Gráfico anillos	51
Figura 35. Gráfico de anillos	51
Figura 36. Gráfico Anillos seccionados	52
Figura 37. Gráfico de árbol	52
Figura 38. Gráfico de árbol de sangría	53
Figura 39. Gráfico de mapa de calor.....	54
Figura 40. Gráfico de mapa de datos.....	54
Figura 41. Gráfico radial	55
Figura 42. Diagrama de árbol horizontal.....	55
Figura 43. Diagrama de árbol vertical	56
Figura 44. Diagrama de árbol vertical	56
Figura 45. Nube de etiquetas.....	57
Figura 46. Tabla de información	57
Figura 47. Gráfico de cascada	58
Figura 48. Diagrama de tamiz.....	58
Figura 49. Ciclo de desarrollo ágil.....	76
Figura 50. Ejecutar aplicación.....	91
Figura 51. Cargar archivo	91

Figura 52. Analizar atributos	92
Figura 53. Visualizar instancias	92
Figura 54. Procesar árboles de decisión.....	93
Figura 55. Procesar agrupamiento.....	93
Figura 56. Generar reglas de asociación	94
Figura 57. Procesar conexión	94
Figura 58. Procesar análisis OLAP	95
Figura 59. Paquete principal	96
Figura 60. Paquete File.....	96
Figura 61. Paquete QualityData.....	97
Figura 62. Paquete DataMining	97
Figura 63. Paquete Dwh	98
Figura 64. Clases del Paquete File	99
Figura 65. Clases del paquete Attribute.....	100
Figura 66. Clases del paquete Data.....	101
Figura 67. Clases del paquete Filters	101
Figura 68. Clases uno del paquete Clustering	102
Figura 69. Clases dos del paquete Clustering	102
Figura 70. Clases tres del paquete Clustering	103
Figura 71. Clases del paquete Rules	104
Figura 72. Clases para Tree	105
Figura 73. Clases para Connection.....	106
Figura 74. Clases uno del paquete Cube.....	106
Figura 75. Clases dos del paquete Cube.....	107
Figura 76. Clases del paquete Table	108
Figura 77. Arquitectura general de VISUALBITOOL.....	109
Figura 78. Arquitectura interna de VISUALBITOOL.....	110
Figura 79. Lista de atributos	119

Figura 80. Tabla de valores del atributo Sodio (Na).....	119
Figura 81. Gráfico de área para el atributo Sodio (Na)	120
Figura 82. Gráfico de Barras para el atributo Sodio (Na).....	120
Figura 83. Gráfico de Circulo para el atributo Sodio (Na)	121
Figura 84. Tabla de valores del atributo fármaco (Drug).....	121
Figura 85. Gráfico de barras para el atributo fármaco (Drug)	122
Figura 86. Gráfico de círculo para el atributo fármaco (Drug)	122
Figura 87. Parámetros distribución de datos	124
Figura 88. Gráfico de dispersión	124
Figura 89. Instancia con un valor anómalo o con errores	125
Figura 90. Gráfico de barras con datos anómalos o con errores	126
Figura 91. Selección de atributos.....	127
Figura 92. Archivo de datos nuevo	128
Figura 93. Interfaz para la discretización de atributos.....	128
Figura 94. Resultado Discretización edad (Age) y sodio (Na)	129
Figura 95. Parámetros del árbol de decisión	131
Figura 96. Árbol jerárquico horizontal	131
Figura 97. Árbol de sangría	132
Figura 98. Parámetros para Cluster horizontal	133
Figura 99. Tabla con los diferentes grupos (Clusters)	134
Figura 100. Archivo Excel con los diferentes grupos (Clusters).....	134
Figura 101. Gráfica de grupos (Clusters).....	135
Figura 102. Valores de un grupo (Cluster)	135
Figura 103. Valores de una instancia con su respectivo valor de confianza.....	136
Figura 104. Mapa de árbol para grupo (Cluster)	136
Figura 105. Parámetros de las reglas de asociación	138
Figura 106. Tabla de reglas de asociación generadas	138
Figura 107. Gráfico de red que permite visualizar las reglas de asociación	139

Figura 108. Información del antecedente y consecuente de una regla.....	139
Figura 109. Estructura tipo estrella del datamart Cantidad_estudiantes.....	141
Figura 110. Estructura del cubo udenar_cantidad	143
Figura 111. Parámetros de conexión	144
Figura 112. Mensaje de conexión exitosa.....	144
Figura 113. Tabla de selección de dimensiones y niveles	145
Figura 114. Imagen del cubo con las dimensiones	145
Figura 115. Tabla con el resultado de una consulta	146
Figura 116. Imagen del cubo con el uso de dos dimensiones	146
Figura 117. Tabla con el resultado de una consulta	147
Figura 118. Imagen del cubo con un menú para filtros	147
Figura 119. Tabla para seleccionar un filtro.....	148
Figura 120. Tabla con el resultado de una consulta con filtros	148
Figura 121. Archivo Excel con los resultados de una consulta	149
Figura 122. Ventana que permite ver la estructura de una consulta.....	149

LISTA DE TABLAS

	Pág.
Tabla 1. Formato de los datos para un diagrama de dispersión	31
Tabla 2. Formato de los datos para un gráfico de matriz	36
Tabla 3. Formato de los datos para un diagrama de red	38
Tabla 4. Formato de los datos para un gráfico de barras y columnas	39
Tabla 5. Formato de los datos para un histograma.....	46
Tabla 6. Formato de los datos para un gráfico de burbujas.....	47
Tabla 11. Formato de los datos para un gráfico circular	48
Tabla 12. Formato de los datos para un gráfico de árbol.....	53
Tabla 13. Análisis de los requerimientos de visualización	60
Tabla 14. Análisis comparativo entre los requerimientos y las técnicas de visualización	62
Tabla 15. Evaluación de las características de un componente reutilizable	65
Tabla 16. Resumen de la evaluación de las características de un componente reutilizable	69
Tabla 17. Resumen de las técnicas de visualización seleccionadas	72
Tabla 18. Características de tres metodologías de desarrollo ágil de software	73
Tabla 19. Requerimientos funcionales de VISUALBITOOL	77
Tabla 20. Caso de uso Ejecutar aplicación.....	85
Tabla 21. Caso de uso Cargar archivo	86
Tabla 22. Caso de uso Analizar atributos	86
Tabla 23. Caso de uso Visualizar instancias.....	87
Tabla 24. Caso de uso Procesar árboles de decisión.....	87
Tabla 25. Caso de uso Procesar agrupamiento.....	88
Tabla 26. Caso de uso Generar reglas de asociación	89
Tabla 27. Caso de uso Procesar conexión	89
Tabla 28. Caso de uso Procesar análisis OLAP	90

Tabla 29. Descripción del archivo Drug.arff	118
Tabla 30. Descripción del archivo Autos.arff	123
Tabla 31. Descripción del archivo Drug.arff	130
Tabla 32. Descripción del archivo Drug.arff	133
Tabla 33. Descripción del archivo Titanic.arff	137
Tabla 34. Datos del datamart cantidad_estudiantes	142
Tabla 35. Tabla de hechos del datamart.....	143

LISTA DE ANEXOS

	Pág.
Anexo1. Documentos.....	158
Anexo2. Software.....	158
Anexo3. Archivo WAR.....	158
Anexo4. Repositorio de datos.....	158

GLOSARIO

ALGORITMO A PRIORI: usado en minería de datos para encontrar elementos frecuentes dentro de un conjunto de datos, los cuales son expresados mediante reglas de asociación (Rodas, 2012).

ALGORITMO J48: de acuerdo con Vizcaino (2008,18), es una implementación del algoritmo C4.5, uno de los algoritmos de minería de datos más utilizado para generar árboles de decisión.

ALGORITMO K-MEDIAS: según García & Aránzazu (2014), es uno de los algoritmos más utilizados para encontrar grupos de elementos dentro de un conjunto de datos. Cada grupo está representado por la media ponderada o centroide de sus puntos.

ÁRBOLES DE DECISIÓN: Vizcaino (2008,11) lo define como: "Son un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que la decisión final a tomar se pueda determinar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas".

ARFF: (AttributeRelation File Format), Es un archivo de texto que describe una lista de casos que comparten un conjunto de atributos, es el formato de archivo utilizado por Weka.

CLUSTERING O AGRUPAMIENTO: es una tarea encaminada a agrupar los elementos que están siendo objeto de estudio de acuerdo a unas características en común. El objetivo de esta tarea, es lograr que los objetos de un mismo grupo maximicen su similitud y a la vez incrementen su diferencia con los miembros de otros grupos.

CSV: (comma-separated values), es un archivo de texto que permite representar datos en un formato tipo tabla, donde los valores de las columnas se separan por comas.

DATA WAREHOUSE (BODEGA DE DATOS): es una Colección de datos pertenecientes a una organización, los cuales son orientados a un tema, están integrados, recolectados en el tiempo y no son volátiles, que posteriormente se utilizarán para la toma de decisiones.

DATAMART: es una parte de la bodega de datos, en el cual se almacena información específica de un área, por ejemplo puede existir un DataMart para notas el cual solo se especifica en recibir información pertinente a notas.

DIMENSIÓN: representa las diferentes caras que posee el cubo, una dimensión permite describir las medidas que se están analizando, por ejemplo en un cubo académico se puede contar con una dimensión para alumnos y otra dimensión para materias , las cuales se podrán relacionar para realizar un análisis determinado.

HIERARCHY (Jerarquía): representa la jerarquía de las relaciones entre uno o varios datos de una dimensión. Por ejemplo la dimensión Tiempo se puede representar en años, estos se puede dividir en meses y finalmente en días.

LEVEL (NIVEL): representa hasta qué punto de detalle se puede ver la información que se encuentra en una dimensión.

MEASURES (MEDIDAS): son los valores numéricos de un cubo, sobre los cuales se realizará operaciones aritméticas dependiendo del análisis que se desea hacer.

MINERÍA DE DATOS: es el proceso de análisis y extracción de información sobre grandes cantidades de datos, es decir, obtener información valiosa que permita ser utilizada para la toma de decisiones en las organizaciones.

OLAP: acrónimo de Procesamiento Analítico en Línea (On-line Analytical Proccesing). Es una tecnología software que facilita a los usuarios acceder a los datos de forma fácil, rápida e interactiva, es decir, mediante diferentes tipos de visualizaciones gráficas, brindarle al usuario una interfaz más amigable para manipular la información de datos sin procesar y con esto facilitar la toma de decisiones.

REGLAS DE ASOCIACIÓN: de acuerdo con Rodas (2012) las define así: “Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos”.

INTRODUCCIÓN

En los procesos de análisis multidimensional en línea (OLAP) y minería de datos (data mining) es necesario visualizar los resultados de tal manera que sean fácilmente entendibles para los usuarios finales que generalmente son directivos de las organizaciones y que son los encargados de la toma de decisiones.

Desde hace un buen tiempo, la información ha pasado a ser uno de los principales activos de las PYMES y MIPYMES, sin embargo; recopilar, integrar, transformar, analizar y presentar los datos que genera dicha información para muchas empresas es un proceso que causa muchos problemas, conllevando a toma de decisiones lentas y a veces con poco acierto.

Actualmente las PYMES y las MIPYMES dentro de los procesos que se llevan a cabo en sus dependencias, no presentan un manejo adecuado de la información, es decir, no poseen un sistema o herramienta que permita extraer, analizar, resumir y estructurar la información que a diario ingresa o se genera, situación que conlleva a retrasos y dificultades en la toma de decisiones. Las empresas manejan grandes cantidades de información, pero esta no es explotada al máximo, por eso se requiere de una manipulación más efectiva y eficaz de la misma, en el sentido que sea útil para la toma de decisiones. Los datos y la información son la materia prima para la toma de decisiones, el pilar de una organización y el objeto de análisis, por ello, el gran interés para las empresas en buscar alternativas o estrategias para recolectar, organizar y presentar la información de acuerdo con un modelo que facilite su explotación y su análisis con el fin de ser más competitivas.

La gran mayoría de herramientas de visualización inteligente de datos necesitan de costosas licencias para su utilización que desborda el presupuesto de las PYMES y MIPYMES, dificultando la toma de decisiones ya que se deben hacer de manera intuitiva, basándose únicamente en la experiencia del personal directivo.

Por esta razón, en este proyecto se plantea una investigación que permita determinar las mejores técnicas de visualización de datos aplicadas a la inteligencia de negocios y específicamente a la minería de datos y análisis multidimensional que ayuden a soportar las decisiones del personal directivo de las micro, pequeñas y medianas empresas PYMES y MIPYMES.

La importancia del presente proyecto de investigación radica en dotar a las PYMES y MIPYMES de una herramienta de visualización de datos libre que permita aplicar diferentes técnicas con el objetivo de brindar un excelente soporte gráfico a los datos y de esta manera facilitar el análisis de la información y la toma de decisiones en estas organizaciones.

Esta herramienta ofrecerá diversas opciones de visualización que garanticen un correcto análisis de los datos, al incorporar nuevas formas gráficas de representación de la información, que puedan relacionar diferentes factores que se involucren en el caso de estudio.

La investigación se centrará en analizar diferentes técnicas de visualización de datos que puedan ser aplicadas en la inteligencia de negocios para mostrar los resultados de manera gráfica específicamente en los procesos de minería de datos y análisis multidimensional. En minería de datos, estas técnicas se utilizaron para visualizar las reglas de asociación, clasificación por árboles de decisión y Clustering particional. Además, se limitó a utilizar tres algoritmos de minería uno para cada técnica utilizada, para reglas de asociación se utilizó A priori, para clasificación por árboles de decisión se utilizó j48 y para Clustering particional K-Medias. En análisis multidimensional OLAP se utilizaron para visualizar los diferentes cubos.

Los objetivos de este proyecto fueron:

El objetivo general fue determinar las técnicas de visualización de datos más adecuadas para aplicarlas en la Inteligencia de Negocios con el fin de dotar a las PYMES y MIPYMES de una herramienta de visualización inteligente de datos que les facilite la toma de decisiones.

Los objetivos específicos fueron:

- a) Apropiar el conocimiento sobre Inteligencia de Negocios y las técnicas de visualización de datos.
- b) Identificar las técnicas de visualización de datos más adecuadas para la Inteligencia de Negocios.
- c) Implementar las técnicas de visualización de datos mediante la construcción de una herramienta de visualización inteligente de datos.
- d) Realizar pruebas de funcionalidad a la herramienta VisualBITool.

Dentro de los resultados de esta investigación está la construcción de VISUALBITOOL, una herramienta de visualización de datos para la inteligencia de negocios débilmente acoplada con el SGBD PostgreSQL utilizando software libre. Su funcionamiento fue evaluado mediante conjuntos de datos de prueba y reales.

La metodología contempló cuatro fases: en la primera fase se apropió el conocimiento sobre Inteligencia de Negocios y las técnicas de visualización de datos, en la segunda fase se identificaron las técnicas de visualización de datos más adecuadas para la Inteligencia de Negocios, en la tercera fase se implementaron las técnicas de visualización de datos mediante la construcción de una herramienta de visualización inteligente de datos, cuyo desarrollo se realizó con la metodología ágil Scrum y finalmente en la cuarta fase se realizaron pruebas de funcionalidad a la herramienta VisualBITool y se evaluaron los resultados.

El documento está organizado en cinco capítulos, en los cuales se da a conocer las diferentes etapas de desarrollo del proyecto.

En el capítulo uno, se encuentra los fundamentos teóricos, en el cual se plasman definiciones y conceptos sobre la visualización de datos, las diferentes técnicas de visualización y entre otros temas pertinentes al proyecto que se ha desarrollado.

En el capítulo dos, se hace la selección de los diferentes tipos de visualización que serán implementados en el producto final de este proyecto, se hace un análisis de los requerimientos de visualización por cada tipo de análisis en minería de datos y en análisis OLAP. Además, de seleccionar las bibliotecas bajo licencia libre que permitan involucrar las diferentes gráficas al software.

En el capítulo tres, se presenta la metodología empleada para desarrollar el proyecto.

En el capítulo cuatro, se encuentra la descripción de la arquitectura interna y externa de la herramienta desarrollada VisualBITool.

En el capítulo cinco, están las pruebas de funcionamiento realizadas a la herramienta VisualBITool y sus respectivos resultados.

En los capítulos seis y siete, se encuentran las conclusiones y recomendaciones sobre el trabajo realizado, finalmente en el capítulo diez están las referencias sobre trabajos que contribuyeron al desarrollo de este proyecto.

1. FUNDAMENTOS TEÓRICOS

1.1 VISUALIZACIÓN

La visualización hace más fácil el proceso de interpretar los datos, para obtener la información necesaria que posteriormente se analizará para tomar una decisión. Hassan (2010,2), la define como:

Hace referencia simultáneamente a un fenómeno cognitivo y a una acción práctica. En el primer caso, se define como la capacidad de formar mentalmente imágenes visuales de conceptos abstractos. En el segundo caso la visualización es entendida como la acción de representar gráficamente fenómenos no visibles. Por tanto, visualizar puede referirse tanto a una acción mental como a una acción práctica, en consecuencia se refiere a nuestra capacidad de pensar visualmente acerca de los datos que no son visibles, como al proceso de representarlos gráficamente.

Haciendo uso de la visualización se puede mostrar la información en detalle, de datos que son complejos de observar a simple vista, como lo dice Hassan (2010,2):

La visualización de la información se basa en datos no físicos, estructuras no visibles como la colaboración entre investigadores, la similitud semántica entre documentos de una colección, los patrones de flujo de operaciones financieras etc.

Así la visualización se convierte en una herramienta muy útil para obtener información de datos que en su estado inicial, es decir, almacenados en una base de datos, en una hoja de cálculo y en algún documento de cuentas o registros, no serían comprensibles a simple vista. Castro, Larrea & Más (2011,1), afirman que:

El principal objetivo de la Visualización es la representación perceptual adecuada tanto de los datos con parámetros múltiples como de las tendencias y las relaciones subyacentes que existen entre ellos. Su propósito no es la creación de las imágenes en sí mismas, sino la asimilación rápida de información o monitoreo de grandes cantidades de datos. La Visualización es parte de los nuevos medios hechos posible debido al desarrollo de la visualización en computadoras en tiempo real.

De acuerdo con la anterior afirmación, si se desea representar la información mediante el uso de la visualización, el usuario final tendrá la facilidad de interactuar con interfaces o gráficas que representen organizadamente la interacción entre varios datos, lo que permitirá asimilar de forma rápida y fácil, la información que se presenta.

1.1.1 Visualización de datos. De acuerdo con Laskowski (2012), es un término que se le da a las diferentes herramientas tecnológicas que ayudan a los usuarios finales ver los datos de forma fácil y comprensible, lo cual facilita la obtención de la información que será utilizada para tomar decisiones que beneficien a una organización.

1.1.2 Importancia de la visualización de la información. El almacenamiento y la obtención de datos cada día crece a pasos gigantescos, lo cual se convierte en un problema en el momento que se desee manipularlos para obtener información, como lo dice Martínez (2011):

Uno de los principales obstáculos para que los ciudadanos y empresas se beneficien del acceso a los datos públicos, es la dificultad que suele entrañar comprender unas informaciones que generalmente adoptan la forma de complejas y complicadas tablas, cuadros o listas. Lo que dificultaría un análisis adecuado de los datos para tomar una decisión.

Si no se toman medidas sobre este problema, el manejo de los datos se podría salir de las manos y en vez de servir como una fuente de información se convertirían en un obstáculo más para una organización, he aquí la importancia de la visualización, como lo describe Martínez (2011):

La visualización de la información juega un papel muy importante, al buscar nuevos métodos o desarrollar proyectos que permitan de una forma más atractiva mostrar la información almacenada utilizando herramientas que representan mediante imágenes y gráficos los datos para hacerlos más comprensibles, atractivos, manejables y útiles.

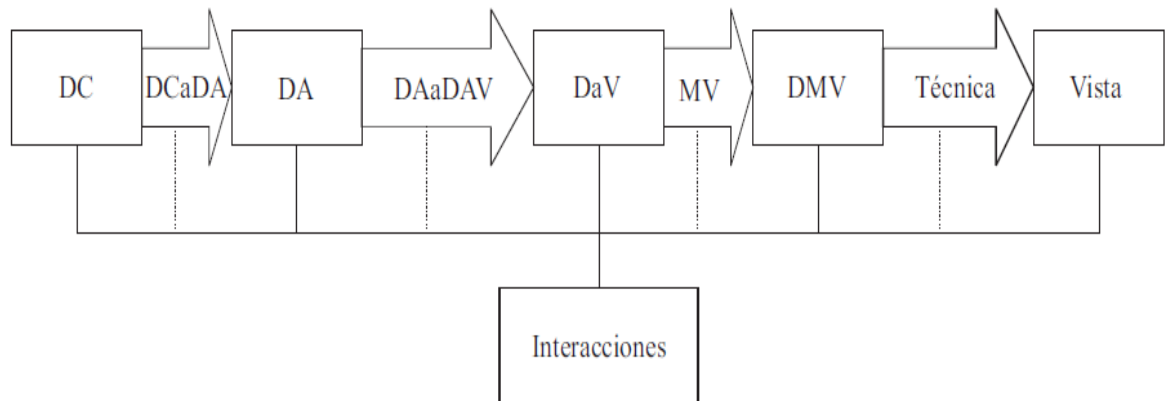
1.2 EL MODELO UNIFICADO DE VISUALIZACIÓN

Se encuentra definido por Urribarri, Castro & Martig (2014,762-763) de la siguiente manera:

Es un modelo aplicable a cualquier visualización, independientemente del campo de origen de los datos. En este modelo se ven representados los diferentes procesos que afectan al conjunto de datos (transformaciones) y las etapas por las que esos datos atraviesan (estados) desde su origen hasta que son finalmente visualizados.

El modelo unificado de visualización presenta unas etapas, que en la figura 1 se dan a conocer, definidas por Urribarri, Castro & Martig (2014, 762-763).

Figura 1. Modelo unificado de visualización (Urribarri, Castro & Martig, 2014,762)



Estado: Datos Crudos (DC): se encuentran los datos en su estado inicial, es decir, como se encuentran en una base de datos o en cualquier otro medio utilizado para recolectar información.

→ **Transformación: Datos Crudos – Datos Abstractos (DCaDA):** el usuario dependiendo del análisis que desea hacer, puede seleccionar de la fuente externa de recolección de datos, cuales quiere utilizar para que sean visualizados. "Esta transformación se encarga de llevar los datos a la representación interna del sistema de visualización" (Urribarri, Castro & Martig, 2014,762).

Estado: Datos Abstractos (DA): en este estado se encuentra el conjunto de datos que el usuario ha seleccionado para ser visualizados, de los cuales puede optar por visualizarlos todos o solo una parte de ellos.

→ **Transformación: Datos Abstractos – Datos a Visualizar (DAaDAV):** en este estado el usuario selecciona exactamente de los datos abstractos, cuáles serán visualizados.

Estado: Datos a Visualizar (DaV): en el conjunto de los datos que el usuario seleccionó para ser visualizados, es decir, los datos que cumplieron con su expectativa o necesidad para realizar el análisis que desea.

→ **Transformación: Mapeo visual (MV):** en esta parte, el usuario analiza como desea visualizar los datos, definiendo así, su estructura visual, que lo definen Urribarri, Castro & Martig (2014,763) como: "está definida por el sustrato espacial, los elementos visuales y los atributos gráficos de los elementos visuales".

De acuerdo con Urribarri, Castro & Martig (2014,763), el sustrato espacial, define como se organizarán los elementos en el espacio determinado para la visualización. Además, los elementos visuales y los atributos gráficos, son las representaciones que mediante estructuras geométricas representan el comportamiento de los datos para ser visualizados en pantalla.

Estado: Datos Mapeados Visualmente (DMV): se encuentran los datos asociados a una técnica de visualización, para ser visualizados por un usuario final en pantalla.

→**Transformación: de Visualización (Técnica):** la definen Urribarri, Castro & Martig (2014,763), así:

Es la aplicación de una técnica que soporte el mapeo visual anterior. Además determina todos los demás elementos que componen la escena: colores, luces, etc. elementos que a pesar de ser parte de la escena, son extras a la visualización de los datos.

Estado: Datos Visualizados (Vista): es la representación final que se le dio a los datos, en la cual el usuario, puede visualizar en pantalla mediante un gráfico la información, así, el usuario puede interactuar con la visualización para explorar los datos y poder tomar decisiones.

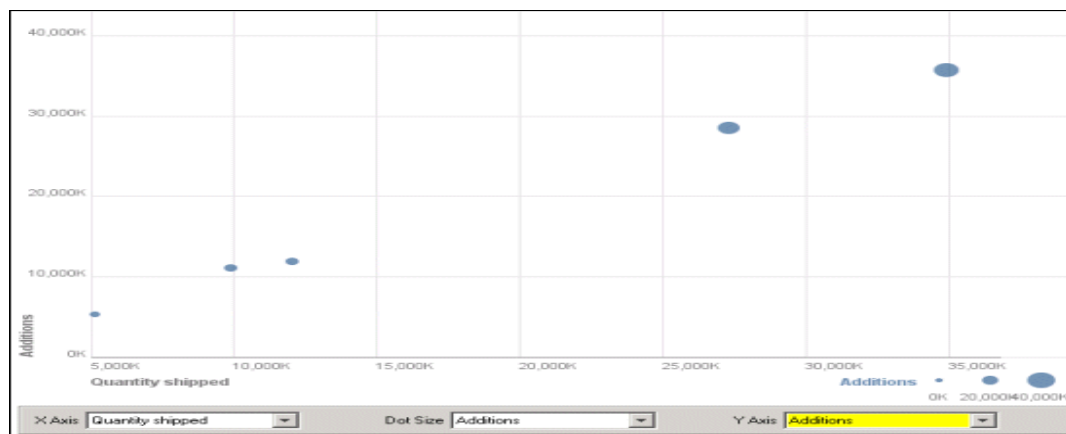
1.3 TÉCNICAS DE VISUALIZACIÓN

En esta sección se presentan las diferentes técnicas de visualización de datos categorizadas según el tipo de representación que se desea generar con los datos.

1.3.1 Relaciones entre los puntos de datos. La representación gráfica más útil para describir el comportamiento de un conjunto de dos o tres variables son los diagramas de puntos, donde cada caso aparece representado como un punto en el plano definido por las variables.

1.3.1.1 Diagrama de dispersión. Permite representar relaciones entre dos o tres variables de tipo numérico de acuerdo con Scatterplot Visualization Guide (2014), que lo define como “Un gráfico de dispersión es un diagrama estadístico clásico que permite visualizar la relación entre las variables numéricas”. Además, da un ejemplo para su mejor interpretación: “si usted tiene una tabla de datos sobre las ciudades, podría utilizar un gráfico de dispersión para ver si hay una relación entre los niveles de población y de la delincuencia”, tal y como se muestra en la figura 2.

Figura 2. Diagrama de dispersión (Scatterplot Visualization Guide, 2014)



Datos requeridos. Según Scatterplot Visualization Guide (2014): “Un gráfico de dispersión es uno de los métodos más flexibles para la visualización de datos numéricos”. Para la generación de este tipo de grafico se necesita cualquier tabla que tenga al menos dos columnas numéricas que representen tanto el eje X como el eje Y y una tercera que represente las etiquetas, como se muestra en la tabla 1.

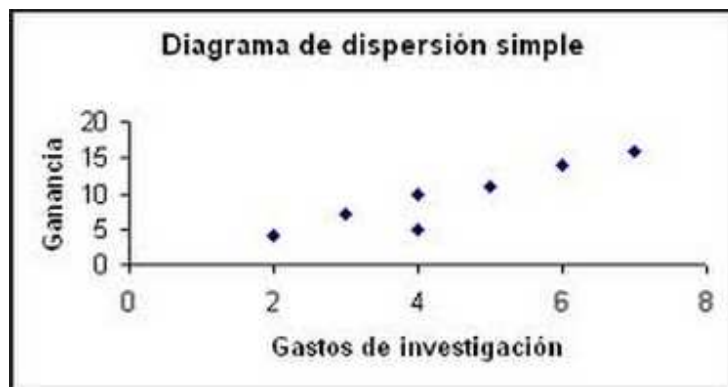
Tabla 1. Formato de los datos para un diagrama de dispersión (Scatterplot Visualization Guide, 2014)

Línea de productos	Costo	Ingresos	Cantidad
Tiendas de campaña	44	210	1000
Sacos de dormir	17	150	100

Tipos de diagramas de dispersión

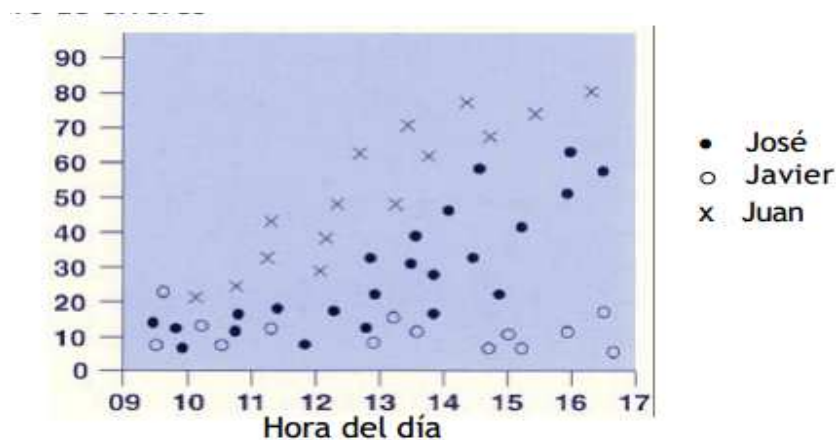
Simple. Un gráfico de dispersión simple solo permite relacionar dos variables de tipo numérico ya que según se define en Diagrama de Dispersión (2014): “este tipo de representación permite relacionar el comportamiento simultáneo de dos variables, una definida en el eje X (abscisas) y la otra en el eje Y (ordenadas)”. En el eje X se selecciona la variable que se considera independiente y en el eje Y la dependiente, tal y como se muestra en la figura 3.

Figura 3. Diagrama de dispersión simple (Bravo ,2014)



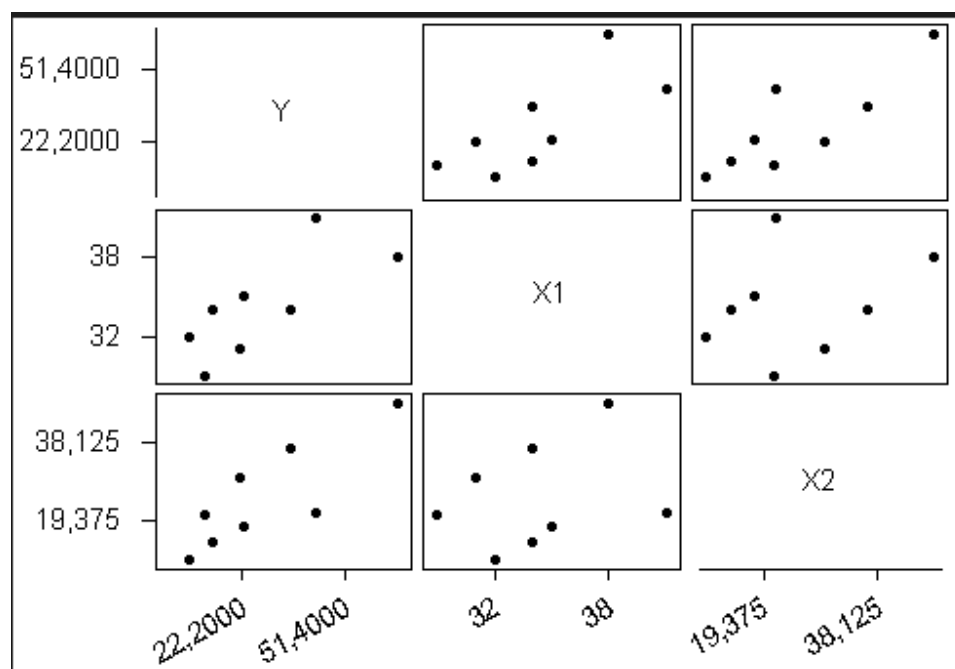
Superpuestos. Este tipo de gráfico permite relacionar más de dos variables en un mismo plano. Por lo general una tercera variable puede ser representada mediante colores, etiquetas o símbolos, como se ve en la figura 4.

Figura 4. Diagrama de dispersión superpuesto (FUNDIBEQ, 2014,6)



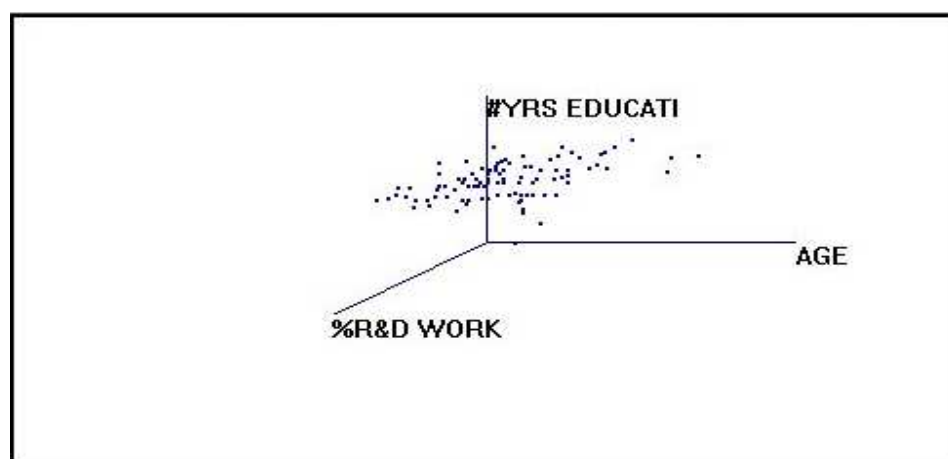
Matricial. Este tipo de diagrama permite relacionar gráficos de dispersión simples en un mismo plano. Según Matrix Chart Guide (2014): “esta representación ofrece una matriz de diagramas de dispersión simples de todos los pares y todas las ordenaciones posibles que se pueden formar con las variables seleccionadas”, ver figura 5.

Figura 5. Diagrama matricial (Matrix Chart Guide, 2014)



Dispersión en 3D. Esta técnica de visualización de datos también puede ser presentada en un plano de tres dimensiones X, Y y Z, tal y como lo afirma UNESCO (2014). En la figura 6 se muestra su representación gráfica.

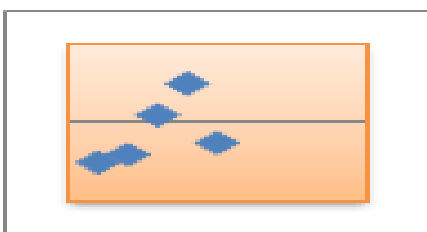
Figura 6. Diagrama matricial (UNESCO ,2014).



Los gráficos de líneas tienen los siguientes subtipos de gráfico:

- **Dispersión con sólo marcadores.** Estos tipos de gráficos (figura 7) comparan pares de valores y para representarlos de forma gráfica utilizan marcadores tales como barras, puntos y símbolos que representan un único valor. Los marcadores que están relacionados conforman una serie de datos. En estos diagramas se suele utilizar las líneas para determinar las conexiones o relaciones entre los puntos (Microsoft-office ,2007).

Figura 7. Diagrama Dispersión con sólo marcadores (Microsoft-office, 2007)



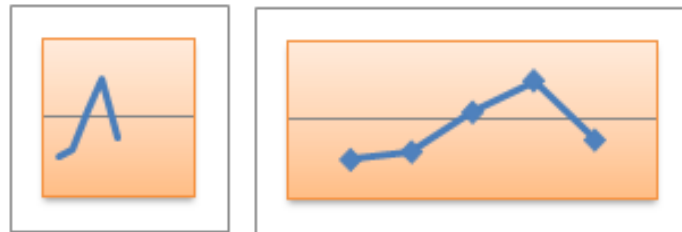
- **Dispersión con líneas suavizadas y dispersión con líneas suavizadas y marcadores.** Este tipo de gráfico que se muestra en la figura 8, representa una curva suavizada que conecta los puntos de datos. Según Microsoft-office (2007): “Las líneas suavizadas se pueden mostrar con o sin marcadores. Use una línea suavizada sin marcadores si hay muchos puntos de datos”.

Figura 8. Diagrama Dispersión con líneas suavizadas y dispersión con líneas suavizadas y marcadores (Microsoft-office, 2007)



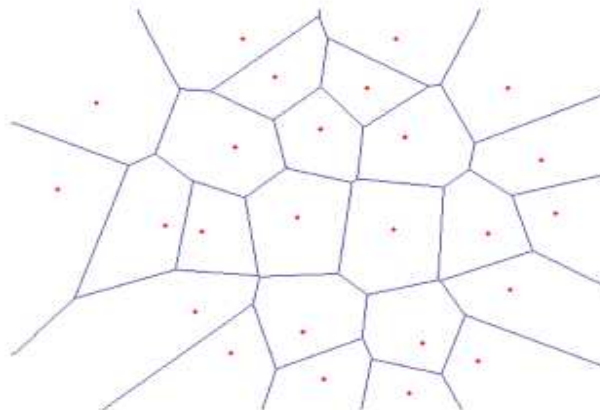
- **Dispersión con líneas rectas y dispersión con líneas rectas y marcadores.** Este tipo de gráfico permite conectar puntos mediante líneas según Microsoft-office (2007), la figura 9 representa su forma gráfica.

Figura 9. Diagrama Dispersión con líneas rectas y dispersión con líneas rectas y marcadores (Microsoft-office, 2007)



1.3.1.2 Diagrama de Voronoi. Representa regiones del plano que están más cercanas a los puntos. Con este tipo de gráficos se puede determinar qué tan lejos o cercano está un punto con respecto a otro. Según Voronoi (2014): “es la subdivisión del plano en regiones formadas por los lugares más próximos a cada uno de los puntos”, la figura 10 muestra su representación gráfica.

Figura 10. Diagrama de Voronoi (Voronoi, 2014)



1.3.1.3 Gráfico de matriz. El gráfico de matriz que se observa en la figura 11, resume un conjunto de datos en una cuadrícula multidimensional.

Figura 11. Gráfico de matriz (Matrix Chart Guide, 2014)



Para la creación de este tipo de gráficos, se utilizar un plano dividido en forma de cuadrícula, donde las filas son los valores de cada columna de tipo texto. Cada celda de la matriz puede estar representada de un tipo de gráfico, que puede ser círculos o barras, los cuales representaran el valor que se genera mediante el cruce de fila y columna (Matrix Chart Guide, 2014).

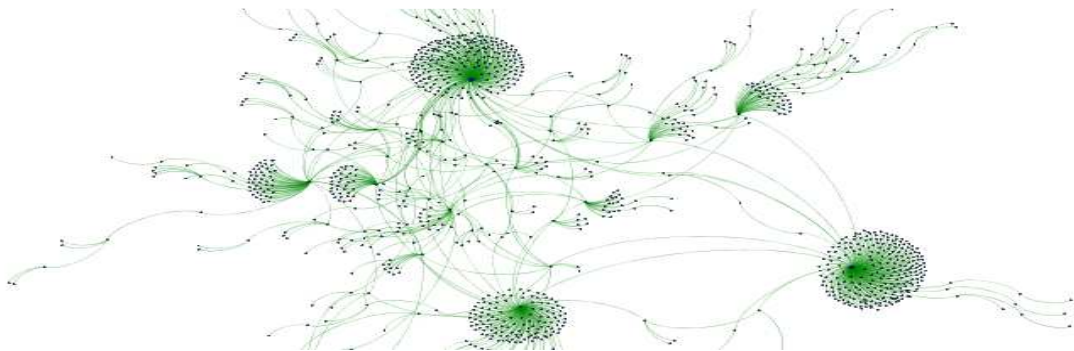
Datos requeridos. Un gráfico de matriz toma los datos en un formato como lo muestra la tabla 2 con al menos dos columnas de texto, de los X(elementos) y Y(categorías). Si hay una tercera columna de texto que puede ser utilizada para determinar colores. Si no hay columnas numéricas, entonces las burbujas o barras simplemente muestran la cantidad de cada combinación de categoría (Matrix Chart Guide, 2014).

Tabla 2. Formato de los datos para un gráfico de matriz (Matrix Chart Guide, 2014)

Línea de productos	Producto	País	Año	Suma Devolución
Equipo de campamento	Tiendas de campaña	Australia	2006	1000000
Equipo de campamento	Linternas	Australia	2006	1000
Equipo de campamento	Sacos de dormir	Austria	2007	10000
El equipaje de golf	Las pelotas de golf	Bélgica	2006	5000

1.3.1.4 Diagrama de Red. Este tipo de gráficos pretende representar relaciones entre elementos tales como cosas, personas y artículos. Actualmente las redes sociales son un ejemplo de este tipo de representación, ya que las personas permanecen conectadas entre sí. Según Network DiagramGuide (2014): “En un diagrama de red, las entidades se representan como nodos y sus relaciones mediante enlaces”, la figura 12 muestra un ejemplo de este tipo de diagrama.

Figura 12. Gráfico de diagrama de red (Network DiagramGuide, 2014)



Network DiagramGuide (2014) define lo siguiente:

Una red se compone de un conjunto de objetos llamados vértices conectados por bordes. La visualización de la red se optimiza para mantener los elementos fuertemente relacionados en estrecha proximidad entre sí. De esta manera, la disposición general de los vértices de la red es muy descriptivo de la estructura de las conexiones entre los vértices (vértices que están lejos están débilmente relacionadas entre sí). En esta visualización, el tamaño de un vértice es proporcional al número de bordes que emanan de ella.

Datos requeridos. Para generar un diagrama de red, es necesario contar con un formato de datos donde existan al menos dos columnas, cada fila represente una única relación entre dos elementos y la información de las columnas representan las etiquetas de visualización de los elementos. Según Network DiagramGuide (2014): “esta técnica no es muy adecuada para redes en las que muchos nodos tienen un gran número de vecinos”.

Por ejemplo, la tabla 3, muestra un conjunto de datos en un formato adecuado para un diagrama de red.

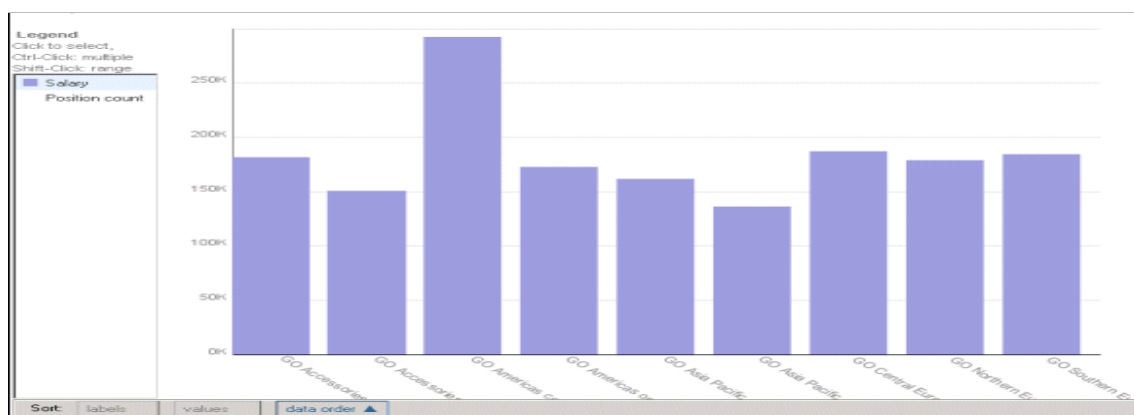
Tabla 3. Formato de los datos para un diagrama de red (Network Diagram Guide, 2014)

Persona	Conoce
Franco	Fernanda
Jesse	Franco
Martín	Jesse
Jesse	Mate
Mate	Fernanda

1.3.2 Comparar un conjunto de valores. A esta categoría pertenecen las representaciones gráficas que permiten mostrar comparaciones entre valores o variables, dando una imagen más clara y comprensible de la distribución de los datos.

1.3.2.1 Gráfico de barras y columnas. Un gráfico de barras generalmente es utilizado para realizar comparaciones entre valores numéricos de uno o más conjunto de variables (Bar Chart Guide, 2014). En la figura 13 se muestra un ejemplo de este tipo de gráfico.

Figura 13. Gráfico de columnas (Bar Chart Guide, 2014)



Datos requeridos. La tabla 4, muestra el formato de datos adecuado para un gráfico de barras, donde las categorías o grupos son comúnmente representados en el eje horizontal y una escala numérica se traza en el eje vertical (Bar Chart Guide, 2014).

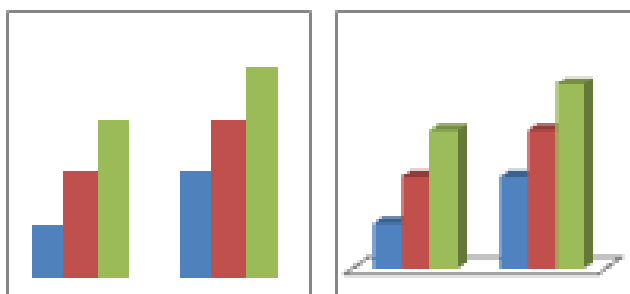
Tabla 4. Formato de los datos para un gráfico de barras y columnas (Bar Chart Guide, 2014)

Animal	Duración de vida	Envergadura
Gryphon	100	12
Fénix	1000	6
Pegaso	50	20

1.3.2.2 Tipos de gráficas en columnas. A continuación se describen los diferentes tipos de gráficos de columnas.

- **Columnas agrupadas y columnas agrupadas en 3D.** Los gráficos de columnas agrupadas, comparan valores entre categorías y muestra valores en rectángulos verticales en 2D. En este tipo de gráficos no se usa un eje de profundidad, solo puede ser visto como una perspectiva en tres dimensiones (Microsoft-office, 2007), ver figura 14.

Figura 14. Gráfico de Columnas agrupadas y columnas agrupadas en 3D (Microsoft-office, 2007)



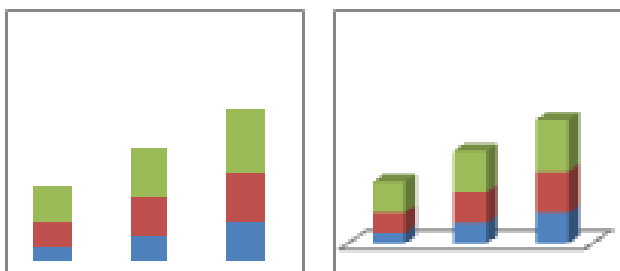
Microsoft-office (2007) define cuando se puede utilizar un tipo de gráfico de columna agrupada, de la siguiente manera:

- ✓ Rangos de valores (por ejemplo, recuentos de elementos).
- ✓ Disposiciones de escalas específicas (por ejemplo, una escala de Likert con entradas como: totalmente de acuerdo, de acuerdo, neutral, en desacuerdo, totalmente en desacuerdo).
- ✓ Nombres que no se encuentran en ningún orden específico (por ejemplo, nombres de artículos, nombres geográficos o los nombres de personas).

Columnas apiladas y columnas apiladas en 3D. Como se observa en la figura 15, este tipo de grafico facilita hacer comparaciones entre valores totales de las categorías y también por medio de una tercera variable, permite mediante colores representar y hacer comparaciones más detalladas de sus valores.

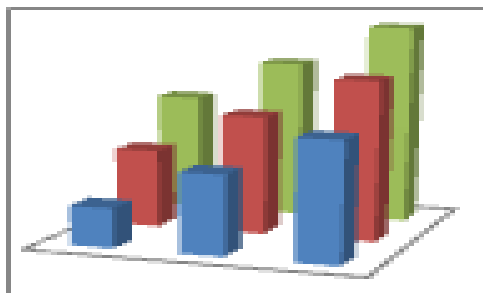
En este tipo de gráficos no se usa un eje de profundidad, solo puede ser visto en una perspectiva de tres dimensiones según Microsoft-office (2007).

Figura 15. Gráfico de Columnas apiladas y columnas apiladas en 3D (Microsoft-office, 2007)



- **Columnas 3D.** Los diagramas de columnas también pueden usar tres dimensiones eje horizontal, eje vertical y un eje de profundidad para comparar valores que pueden ser representados por barras, columnas, líneas y puntos que se denominan marcadores de datos, los cuales al ser representados mediante colores, constituyen los que comúnmente se denomina como serie de datos según Microsoft-office (2007). Un ejemplo de este tipo de gráfico se puede observar en la figura 16.

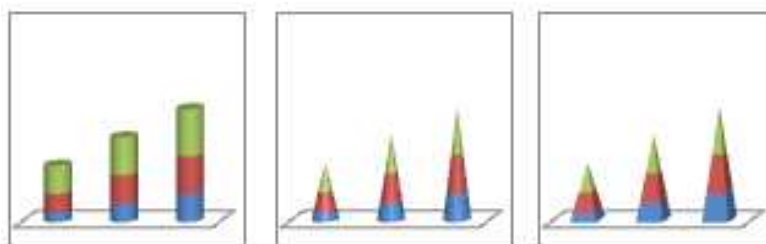
Figura 16. Gráfico de Columnas 3D (Microsoft-office, 2007)



Puede utilizar un gráfico de columnas 3D cuando desee comparar del mismo modo datos entre categorías y entre series, ya que este tipo de gráfico muestra categorías a lo largo de los ejes horizontales y de profundidad, mientras que el eje vertical muestra los valores según Microsoft-office (2007).

- **Cilindro, cono y pirámide.** Según lo planteado por Microsoft-office (2007): “los gráficos de cilindros, conos y pirámides están disponibles en los mismos tipos de gráficos agrupados, apilados, 100% apilados y en 3D proporcionados para gráficos de columnas rectangulares”, ver figura 17.

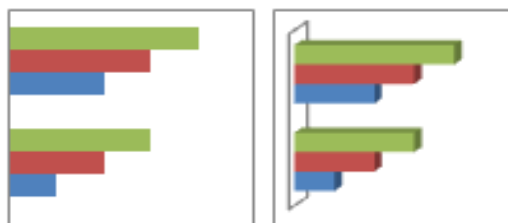
Figura 17. Gráfico de Cilindro, cono y pirámide (Microsoft-office, 2007)



1.3.2.3 Tipos de gráficas en barras. A continuación se dan a conocer los diferentes tipos de gráficas de barras con los que se pueden representar los datos.

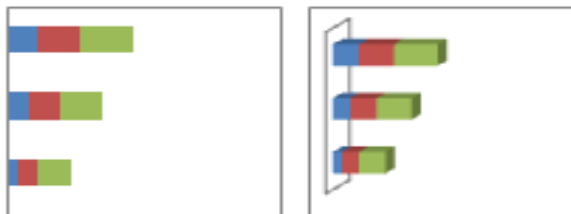
- **Barra agrupada y barra agrupada en 3D.** La figura 18, muestra los gráficos de barras agrupadas que comparan valores entre categorías. En un gráfico de barras agrupadas, las categorías se suelen organizar a lo largo del eje vertical, mientras que los valores lo hacen a lo largo del eje horizontal. Un gráfico de barras agrupadas en 3D muestra rectángulos horizontales en formato 3D, por lo cual no se considera que los datos se están comparando en tres dimensiones o tres ejes del plano según como lo afirma Microsoft-office (2007).

Figura 18. Gráfico de Barra agrupada y barra agrupada en 3D (Microsoft-office, 2007)



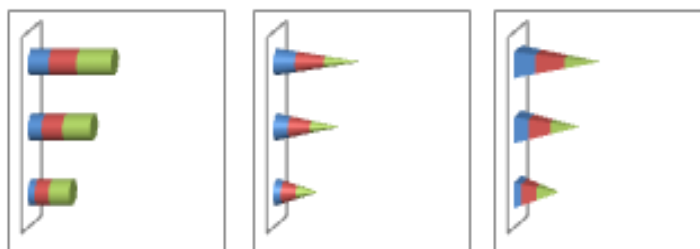
- **Barra apilada y barra apilada en 3D.** En la figura 19, se muestran gráficos de barras apiladas, que permiten representar categorías o series para la representación de relaciones entre elementos de un conjunto. Cuando se hace referencia a 3D no se trata de representar los datos en un plano de tres dimensiones, solo la forma en este caso una barra apilada en 3D según lo que plantea Microsoft-office (2007).

Figura 19. Gráfico de Barra apilada y barra apilada en 3D (Microsoft-office, 2007)



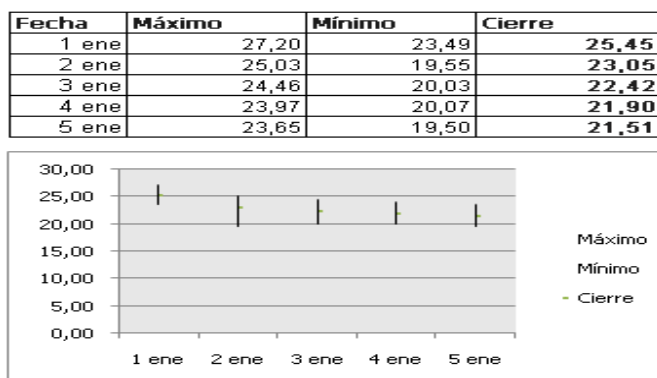
- **Cilindro, cono y pirámide horizontales.** Según Microsoft-office (2007): "Estos gráficos están disponibles en los mismos tipos de gráficos agrupados, apilados y 100% apilados que se proporcionan para los gráficos de barras rectangulares". Muestran y comparan los datos de la misma forma. La única diferencia es que estos tipos de gráfico muestran formas cilíndricas, cónicas y piramidales en lugar de rectángulos horizontales, tal y como se observa en la figura 20.

Figura 20. Gráfico de Cilindro, cono y pirámide horizontales (Microsoft-office, 2007)



2.3.2.4 Gráficos de cotizaciones. Según lo planteado por Microsoft-office (2007), actualmente este tipo de gráficos (ver figura 21) son utilizados específicamente para datos financieros y científicos, pero se utilizan con mayor frecuencia para mostrar la variación de los precios de las acciones. En la parte científica por ejemplo pueden ser usados para mostrar variaciones de la temperatura en periodos de tiempo. Según Microsoft-office (2007): "para crear un gráfico sencillo de cotizaciones de máximos, mínimos y cierre, debería organizar los datos en columnas, con Máximos, Mínimos y Cierre como encabezados de columnas siguiendo ese orden".

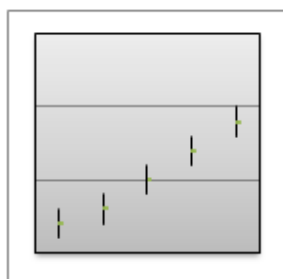
Figura 21. Gráficos de cotizaciones (Microsoft-office, 2007)



Los gráficos de cotizaciones tienen los siguientes subtipos de gráfico:

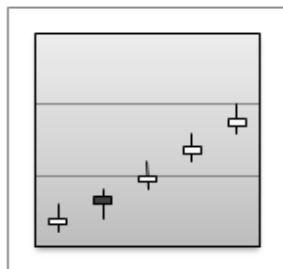
Máximos, mínimos y cierre. Este diagrama se suele utilizar para la representación de los precios de los valores, como se ve en la figura 22. Usa tres series que son máximos, mínimos y cierre respectivamente, los cuales permiten determinar la variación de una determinada variable, en este caso el precio de los valores (Microsoft-office, 2007).

Figura 22. Gráficos de Máximos, mínimos y cierre (Microsoft-office, 2007)



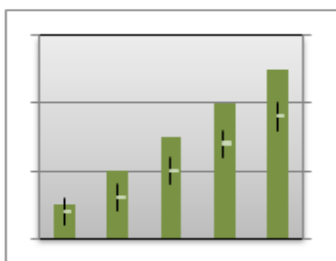
Apertura, máximos, mínimos y cierre. Este diagrama necesita cuatro categorías o series de valores: apertura, máximos, mínimos y cierres, los cuales deben cumplir este orden para una correcta visualización de resultados según lo planteado por Microsoft-office (2007). La figura 23, muestra un ejemplo de este tipo de gráfico.

Figura 23. Gráficos de Máximos, mínimos y cierre (Microsoft-office, 2007)



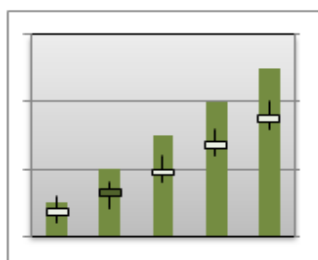
Volumen, máximos, mínimos y cierre. Como se puede ver en la figura 24 este tipo de diagrama necesita cuatro categorías o series de valores en el orden correcto volumen, máximos, mínimos y cierre. Según Microsoft-office (2007): "mide el volumen mediante dos ejes de valores: uno para las columnas que miden el volumen y otro para el precio de los valores".

Figura 24. Gráficos de Volumen, máximos, mínimos y cierre (Microsoft-office, 2007)



Volumen, apertura, máximos, mínimos y cierre. Según Microsoft-office (2007): "Este tipo de gráfico de cotizaciones necesita cinco series de valores en el orden correcto (volumen, apertura, máximos, mínimos y cierre)". La figura 25 muestra un ejemplo de este tipo de diagrama.

Figura 25. Gráficos de Volumen, apertura, máximos, mínimos y cierre (Microsoft-office, 2007)

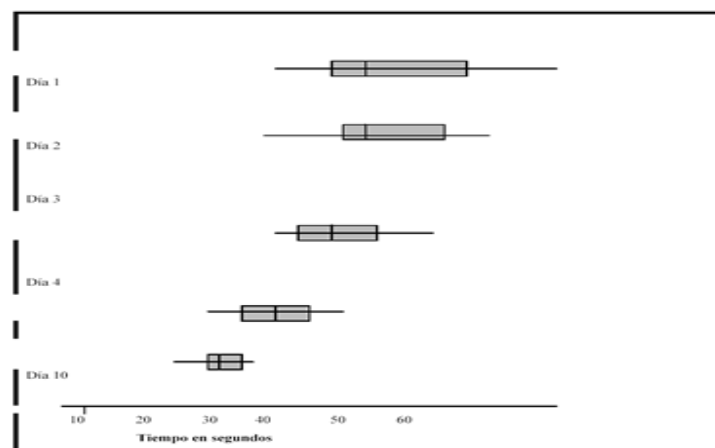


1.3.2.5 Diagrama de Caja y Bigotes. Es un diagrama utilizado para describir características tales como la dispersión y la simetría. Por ejemplo, puede ser muy útil para determinar qué tan cerca está un valor con respecto a su media, según lo que plantea CopyleftTitapg (2008), ver figura 26.

Según CopyleftTitapg (2008), un gráfico de este tipo se puede entender de la siguiente manera:

Una gráfica de este tipo consiste en una caja rectangular, donde los lados más largos muestran el recorrido intercuartílico. Este rectángulo está dividido por un segmento vertical que indica donde se posiciona la mediana y por lo tanto su relación con los cuartiles primero y tercero (recordemos que el segundo cuartil coincide con la mediana). Esta caja se ubica a escala sobre un segmento que tiene como extremos los valores mínimo y máximo de la variable. Las líneas que sobresalen de la caja se llaman bigotes. Estos bigotes tienen un límite de prolongación, de modo que cualquier dato o caso que no se encuentre dentro de este rango es marcado e identificado individualmente.

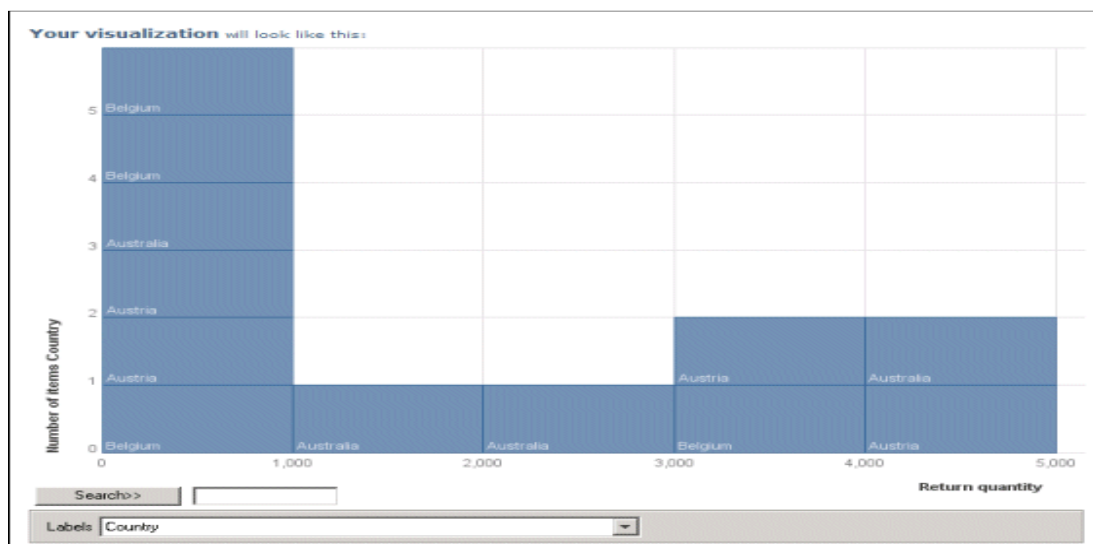
Figura 26. Gráficos de caja y bigotes (CopyleftTitapg, 2008)



1.3.2.6 Histograma de bloques. Esta clase de diagramas permite representar la distribución de rangos de valores y a la vez establecer comparaciones entre ellos.

Según Block HistogramGuide (2014), el eje x se divide en contenedores, los cuales corresponden a los valores de cada intervalo y se dibujan bloques apilados sobre cada contenedor para representar el número de valores que hay en cada intervalo, ver figura 27.

Figura 27. Histograma de bloques (Block HistogramGuide ,2014)



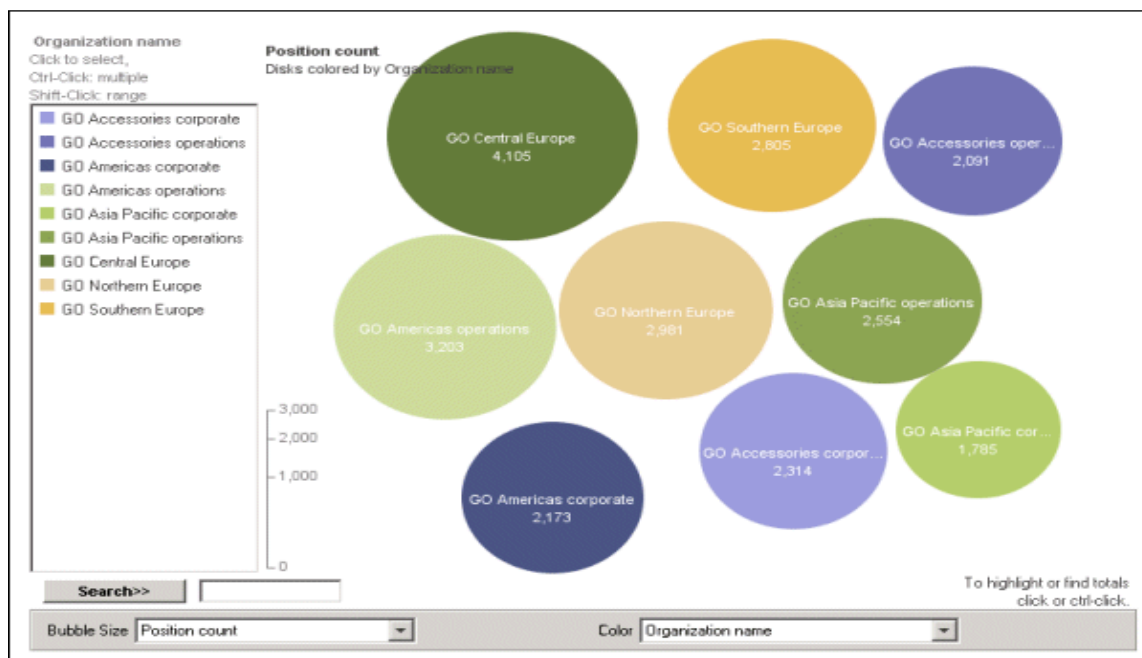
Datos requeridos. Según Block HistogramGuide (2014), para elaborar un gráfico de este tipo, se necesita un formato de datos con estructura de una tabla, la cual posea una columna tipo texto para generar etiquetas y al menos una columna numérica para la representación de los valores. La tabla 5, muestra la estructura de datos de entrada para esta clase de diagrama.

Tabla 5. Formato de los datos para un histograma (Block HistogramGuide, 2014)

Persona	Altura (m)
Adán	6
Sarah	5
David	6
Goliath	11

1.3.2.7 Gráfico de burbujas. Según Bubble Chart Guide (2014), este tipo de diagrama permite representar y comparar valores numéricos por medio de círculos y es muy utilizado para representar grandes cantidades de datos con valores numéricos que difieren en su magnitud, la cual permite establecer un orden dentro de su área, la figura 28, muestra un ejemplo.

Figura 28. Gráfico de burbujas (Bubble Chart Guide, 2014)



Datos requeridos. Según Bubble Chart Guide (2014), la tabla 6 muestra un formato de datos ejemplo para la generación de un gráfico de burbujas. La columna de tipo texto "Condición" es la etiqueta y la columna "Número de personas infectadas" indica el valor.

Tabla 6. Formato de los datos para un gráfico de burbujas (Bubble Chart Guide, 2014)

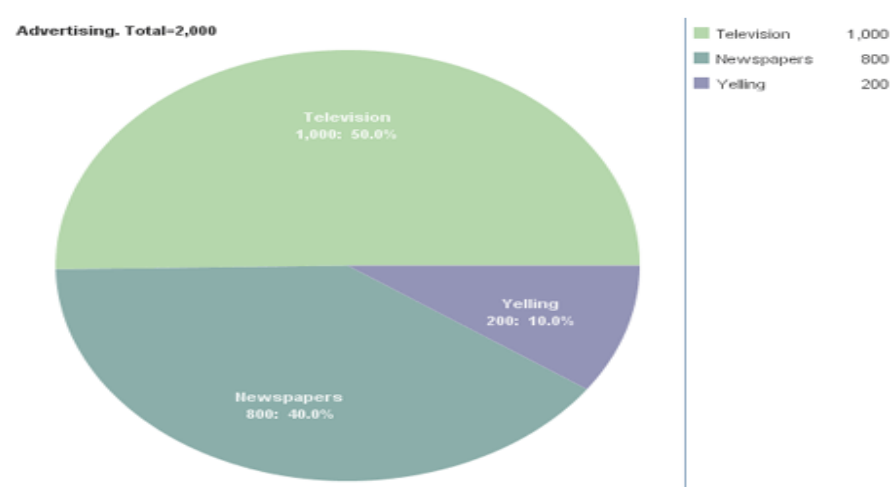
Condición	Número de personas infectadas
Optimismo incurable	1000000
Ataque de risa	50000
Inexplicable sonriendo	6000

1.3.3 Seguimiento del sube y baja con el tiempo. Este tipo de gráficos muestran el seguimiento de los cambios que se dan en los valores de los datos de acuerdo a un determinado tiempo, a continuación se dan a conocer los diferentes tipos de gráficas para esta categoría.

1.3.4 Ver las partes de un todo. Este tipo de gráficos permiten ver como se distribuyen los valores en un conjunto de dato en su totalidad, a continuación se dan a conocer los diferentes tipos de gráficas que pertenecen a este tipo.

1.3.4.1 Gráfico circular. La finalidad principal de un gráfico circular, es representar valores mediante proporciones. Según Pie Chart Guide (2014), estos valores pueden ser porcentajes, donde el total del círculo indica el 100% y las divisiones indica los diferentes porcentajes que se desea representar. Este tipo de grafico es muy semejante a un pastel, donde cada rebanada es un valor o porcentaje. La figura 29, muestra un ejemplo de este diagrama circular.

Figura 29. Gráfico circular (Pie Chart Guide, 2014)



Datos requeridos. De acuerdo con la tabla 7, el gráfico circular necesita una columna de tipo texto, para la asignación de una etiqueta a cada rebanada y al menos una columna numérica para los valores de cada porción del círculo.

Tabla 7. Formato de los datos para un gráfico circular (Pie Chart Guide, 2014)

Presupuesto (USD)	2012	2016	2016
Televisión	45	55	10
Periódicos	55	45	10
Esquinas de las calles	0	0	80

Los gráficos circulares tienen los siguientes subtipos de gráfico:

- **Circular y circular en 3D.** Según Microsoft-office (2007), los gráficos circulares pueden ser representados en una estructura 2D o 3D, esto depende de cómo el usuario quiera visualizarlos, ver figura 30.

Figura 30. Gráfico Circular y circular en 3D (Microsoft-office, 2007)



- **Circular con subgráfico circular y circular con subgráfico de barras.** Estos tipos de diagramas según Microsoft-office (2007), son utilizados para visualizar con más detalle los sectores pequeños de un gráfico circular principal, a través de un gráfico secundario, el cual muestra la información con más claridad. La figura 31, muestra un ejemplo de estos tipos de gráficos.

Figura 31. Gráfico Circular con subgráfico circular y circular con subgráfico de barras (Microsoft-office, 2007)



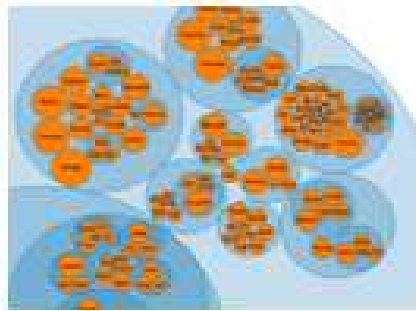
- **Circular seccionado y circular seccionado en 3D.** Un diagrama circular se puede presentar también por secciones. Según Microsoft-office (2007), cada sección no se pueden separar o mover, ya que hacen parte de un círculo principal. En la figura 32, se puede observar que estos tipos de gráficos pueden ser representados tanto en 2D como en 3D.

Figura 32. Gráfico Circular seccionado y circular seccionado en 3D (Microsoft-office, 2007)



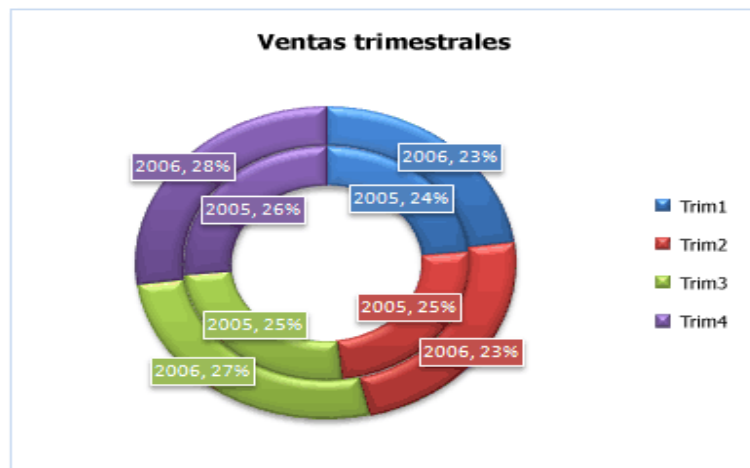
- **Gráfico de embalaje de círculos.** Es la representación visual, de círculos agrupados dentro de otros círculos, ver figura 33.

Figura 33. Gráfico de embalaje de círculos (Data-driven documents, 2013)



- **1.3.4.2 Gráfico de anillos.** La forma de representar los datos es muy similar a un gráfico circular, ya que el grafico de anillos relaciona todas sus partes como un todo, pero la ventaja de este tipo de grafico es que permite adicionar categorías o series, donde cada una de ellas se diferencia mediante un color y pueden considerarse como la leyenda del gráfico, según lo planteado por Microsoft-office (2007). En un gráfico de anillos, se pueden representar datos organizados únicamente en columnas o en filas como una hoja de cálculo, ver figura 34.

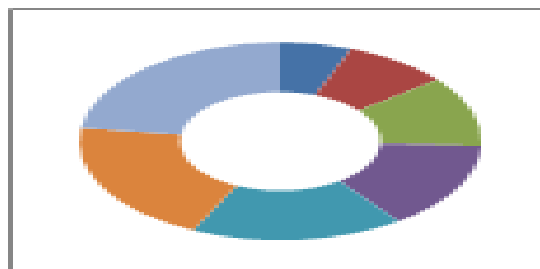
Figura 34. Gráfico anillos (Microsoft-office, 2007)



Según Microsoft-office (2007): "los gráficos de anillos no son fáciles de leer. A veces se puede utilizar un gráfico de columnas apiladas o un gráfico de barras apiladas en su lugar". Los gráficos de anillos tienen los siguientes subtipos:

- **Anillos.** Por lo general, los anillos para poder diferenciar sus valores, hacen uso de los colores. Cada porción representa una parte de un todo, el cual viene siendo un 100% según lo afirmado por Microsoft-office (2007), ver figura 35.

Figura 35. Gráfico de anillos (Microsoft-office, 2007)



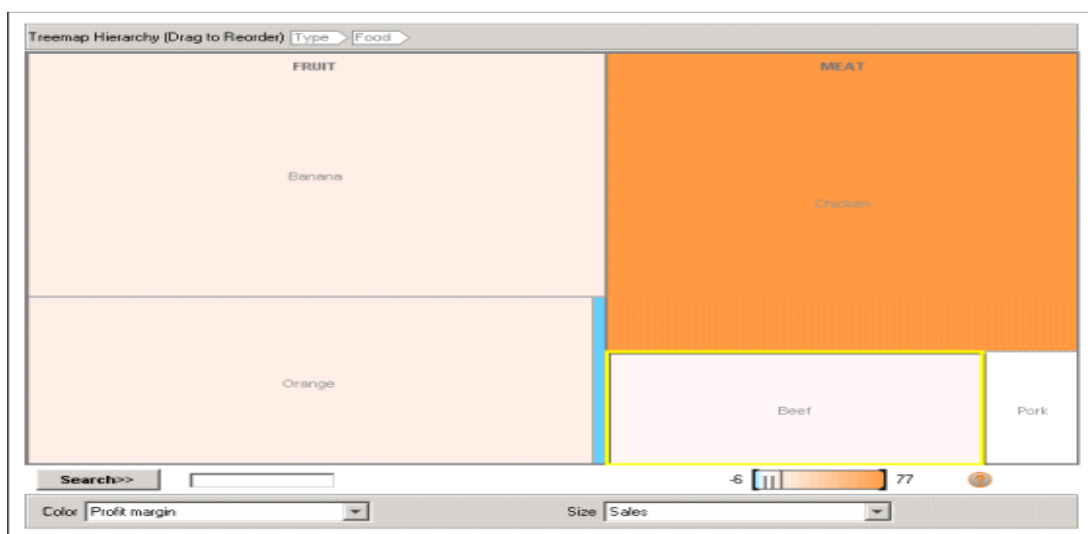
- **Anillos seccionados.** Este tipo de gráficos, permite visualizar sus particiones de forma separada de manera muy similar a los gráficos circulares por secciones. La figura 36, muestra un ejemplo de este tipo de gráficos.

Figura 36. Gráfico Anillos seccionados (Microsoft-office, 2007)



1.3.4.3 Mapa de árbol. Gráfico muy utilizado para la representación de jerarquías. Según TreemapGuide (2014), este tipo de representación trabaja con nodos, los cuales se agrupan en estructuras jerárquicas. Estos permiten describir un camino, ya que a medida que se desplaza por un nodo, se van desplegando ramas hasta llegar finalmente a las hojas. En algunos mapas de árbol sus niveles pueden ser diferenciados por colores. Estos gráficos, por lo general, son usados para la identificación de patrones o información interesante para la toma de decisiones. En la figura 37, se puede visualizar un ejemplo de este tipo de gráfico.

Figura 37. Gráfico de árbol (TreemapGuide, 2014)



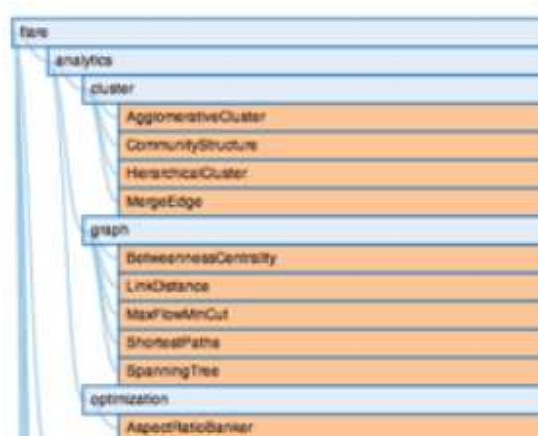
Datos requeridos. Para la generación de un mapa de árbol se requiere una estructura de datos parecida a la que presenta la tabla 8. Según TreemapGuide (2014), las categorías y subcategorías se tomarán de las columnas que están de izquierda a derecha. Es decir, si una columna está más a la izquierda, se considera categoría y las que siguen serán las subcategorías.

Tabla 8. Formato de los datos para un gráfico de árbol (TreemapGuide, 2014)

Tipo	Comida	Venta	Margen de beneficio
Fruta	Plátano	1456	12
Fruta	Limón	23	-6
Fruta	Naranja	981	12
Carne	Cerdo	111	2
Carne	Carne de res	442	6
Carne	Pollo	1456	77

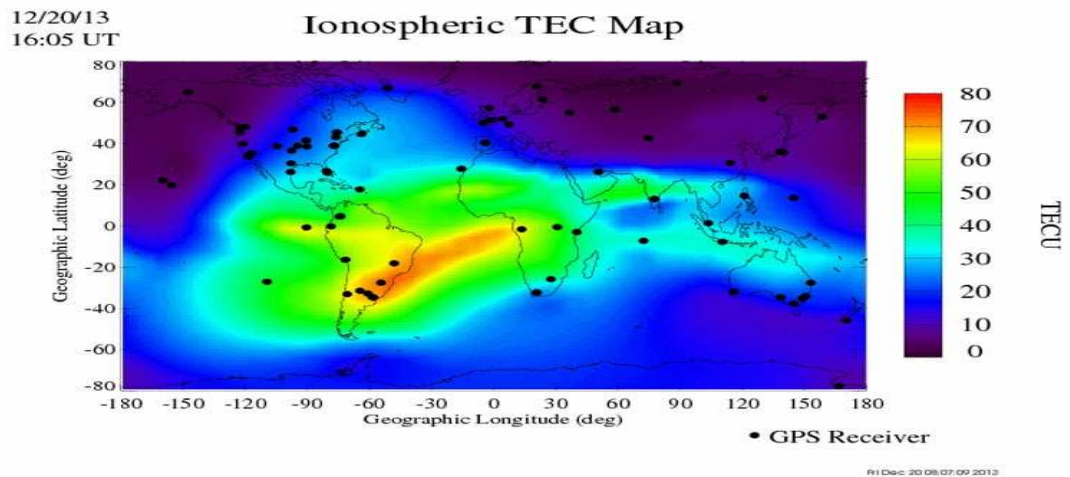
1.3.4.4 Gráfico de árbol de sangría. Es un gráfico que muestra en rectángulos continuos y desplegables en una estructura jerárquica , ver figura 38.

Figura 38. Gráfico de árbol de sangría (Indented Tree, 2014)



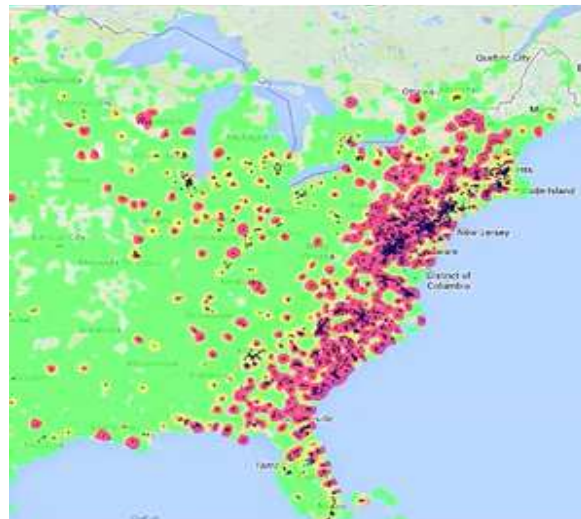
1.3.4.5 Mapa de calor. Este tipo de gráficos es una forma de representar comparaciones de datos categóricos mediante colores. Según Tableau software (2014): “Los mapas de calor permiten ver variaciones en los datos a través de variaciones en color”. Se utiliza una gama de colores organizados jerárquicamente, donde los puntos de mayor interés se representan con colores más cálidos como el rojo, el naranja o amarillo y mientras que las zonas de menos iteres se utiliza colores más fríos como el verde y el azul, según como se afirma en Tableau software (2014), ver figura 39.

Figura 39. Gráfico de mapa de calor (Tableau software, 2014)



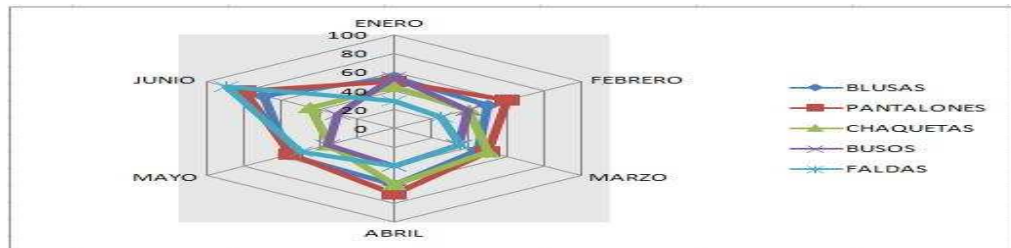
1.3.4.6 Gráfico de mapas. Se usa para representar datos geográficos como por ejemplo un mapa de un país, continente o región, para diferenciar los datos que está representando usa colores que distinguen valores asociados con cada una de las ubicaciones del mapa, de acuerdo con GRAFMAPA (2014), ver figura 40.

Figura 40. Gráfico de mapa de datos (MAPASDATOS, 2013)



1.3.4.7 Gráfico radial. Estos diagramas son una alternativa para visualizar una o más variables en un gráfico de dos dimensiones. Según RADIAL (2014), cada radio corresponde a una variable. La línea que pasa por todos los radios une a todos los puntos que representan los datos. La figura 41, muestra un ejemplo de este tipo de gráfico.

Figura 41. Gráfico radial (RADIAL ,2014)



Formato de los datos. Este formato puede incluir datos en una o en varias columnas. En la primera columna (opcional), se puede introducir datos cualitativos, que según la figura 41, aparecen en la parte externa del círculo. El resto de las columnas deben incluir datos numéricos que representarán los puntos de datos en cada uno de los radios de este gráfico según RADIAL (2014).

1.3.4.8 Diagrama de árbol. Un diagrama de árbol es un método gráfico para identificar todas las partes necesarias para alcanzar un objetivo final. Los diagramas de árbol se utilizan generalmente para identificar todas las tareas necesarias para implantar una solución (DIAGRAMÁRBOL ,2014).

Diagrama de árbol horizontal. Este tipo de gráfico permite representar en una estructura jerárquica las partes necesarias para lograr un objetivo o realizar una tarea, su estructura se representa de manera horizontal como se ve en la figura 42.

Figura 42. Diagrama de árbol horizontal (DIAGRAMÁRBOL ,2014)

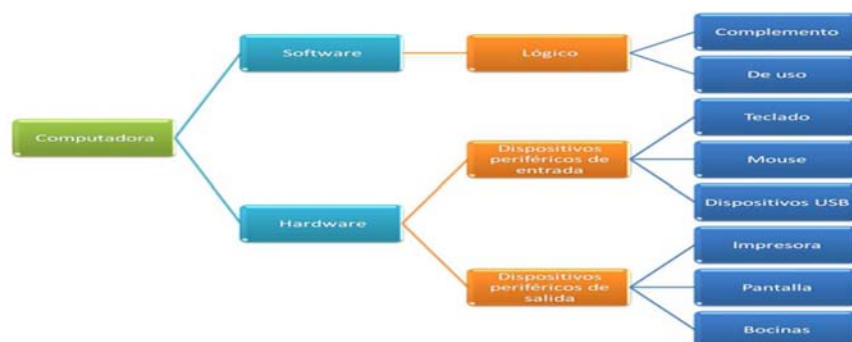


Diagrama de árbol vertical. Este tipo de gráfico permite representar en una estructura jerárquica las partes necesarias para lograr un objetivo o realizar una tarea, su estructura se representa de manera vertical como se ve en la figura 43.

Figura 43. Diagrama de árbol vertical (DIAGRAMÁRBOL ,2014)



1.3.4.9 Dendrograma. Es undiagrama de datos en forma de árbol que organiza los datos en subcategorías que se van dividiendo en otras hasta llegar al nivel de detalle deseado. Esto es muy parecido a las ramas de un árbol que se van dividiendo en otras hasta llegar finalmente a las hojas. La ventaja de este diagrama es que se puede apreciar con claridad las relaciones de agrupación según lo planteado por Dendrograma (2013), ver figura 44.

Figura 44. Diagrama de árbol vertical (Cruz et al ,2007)

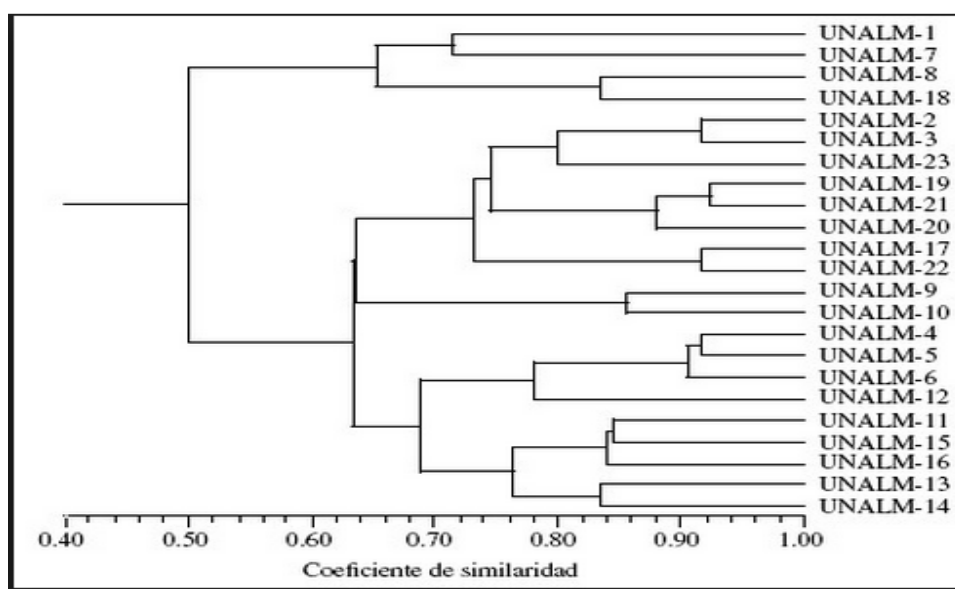


Figura 45. Nube de etiquetas (ONO ,2010)

Figura 46. Tabla de información (Tablas ,2014)

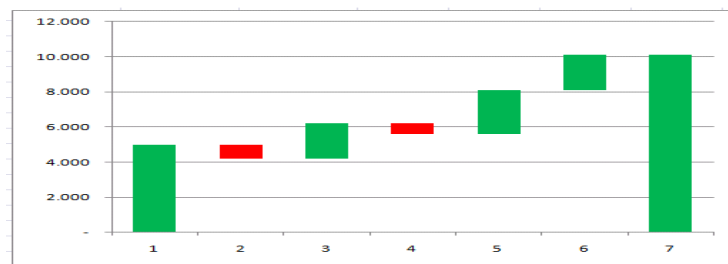
PAIS	Superficie	Población	Densidad
El Salvador	21,040	6,948,073	330.2
Guatemala	108,890	12,728,111	116.8
Costa Rica	51,100	4,509,290	70.8
Honduras	112,492	7,483,763	66.7

1.3.4.12 Gráfico de cascada. Ccance (2014) lo define como:

Un gráfico de cascada (a veces denominado gráfico de puente) permite visualizar un cálculo de suma con subtotales. Los valores positivos generan segmentos ascendentes; los valores negativos crean segmentos descendentes. Los subtotales, es decir, segmentos que descienden hasta la línea de base del gráfico.

Un ejemplo de este tipo de gráfico se muestra en la figura 47.

Figura 47. Gráfico de cascada (Masexcel, 2010)

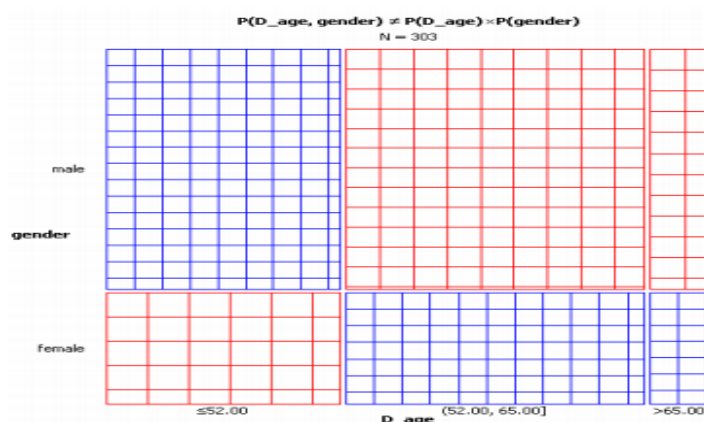


1.3.4.13 Diagrama de tamiz. La definición de este tipo de gráfico según Demšar (2010) es la siguiente:

También llamado diagrama de Parquet, es una visualización de tablas de frecuencia de contingencia de doblez, similar a los utilizados en chi-cuadrado. El rectángulo que representa el conjunto de datos se divide horizontalmente y verticalmente de acuerdo a la proporción de valores de dos atributos discretos.

Un ejemplo de este grafico se ve en la figura 48.

Figura 48. Diagrama de tamiz (Demšar, 2010)



2. SELECCIÓN DE LAS TÉCNICAS DE VISUALIZACIÓN DATOS PARA LA HERRAMIENTA VISUALBITOOL

Para poder aplicar las técnicas de visualización de datos en inteligencia de negocios, fue necesario realizar un estudio sobre los resultados generados por las herramientas de análisis de datos , esto con el fin de determinar los diferentes requerimientos de visualización de cada técnica de análisis de datos y con esto adaptar técnicas de visualización de datos apropiadas.

La tabla 9, muestra el análisis general de los diferentes resultados por cada tarea de análisis de datos proporcionados por clases java, las cuales utilizan según su funcionalidad los Apis Weka y Olap4j para la creación de métodos que faciliten la generación de estos resultados.

Se usó diferentes repositorios de datos para probar las clases en NetBeans IDE 7.0.1 y mediante consola observar los resultados y así poder determinar de forma general los patrones de conocimientos y la forma como representarlos y entenderlos, para ello se determinó los requerimientos de visualización.

Las clases java que se crearon para facilitar este análisis, fueron utilizadas posteriormente para el desarrollo de la herramienta de visualización de datos propuesta. Para más detalle acerca de estas clases, ir al capítulo 3, sección 3.5.

Tabla 9. Análisis de los requerimientos de visualización

Tarea de Análisis de Datos	Herramienta utilizada	Herramienta de Análisis de datos	Clase Java	Resultados	Requerimientos de visualización
Atributos	Weka	Iteración	StatisticalValues.java	<ul style="list-style-type: none"> • Número de atributos. • Nombres de atributos. • Tipos de atributos. • Valores de atributos. • Lista de atributos. • Índices. 	Gráfico(s) que permita(n) visualizar la distribución de valores de cada atributo.
Instancias	Weka	Iteración	ViewInstance.java	<ul style="list-style-type: none"> • Nombre de relación. • Número de instancias. • Valores por instancia. • Lista de instancias. • Índices. 	Gráfico(s) que permita(n) visualizar la distribución de instancias.
Clustering	Weka	Algoritmo k-means.	KMean.java	<ul style="list-style-type: none"> • Número de grupos. • Valores por grupo. • Instancias por grupo. • Lista de grupos. • Distancia euclidiana de un grupo con respecto a una instancia. 	Gráfico(s) que permitan(n) visualizar e identificar los grupos generados y también que permitan observar la distribución de cada instancias con respecto al grupo al cual pertenece.
Reglas de asociación	Weka	Algoritmo a priori.	ClassApriori.java	<ul style="list-style-type: none"> • Antecedentes. • Consecuentes. • Número de reglas generadas. • Factor de Confianza por regla. 	Gráfico(s) que permita(n) visualizar el antecedente, consecuente y factor de confianza de cada regla de asociación generada.

Arboles de decisión	Weka	Algoritmo j48.	ClassJ48.java	<ul style="list-style-type: none"> • Número de Niveles. • Tamaño del árbol. • Nodos. • Relaciones entre nodos. • Árbol. 	Grafico(s) que permita(n) visualizar las relaciones de cada nodo de un árbol de decisión en una estructura jerárquica.
Análisis OLAP	Olap4J	Consultas MDX, SQL	QueryTable.java	<ul style="list-style-type: none"> • Esquema. • Cubo. • Dimensiones • Niveles. • Medidas. • Consultas MDX. • Atributos. • Tuplas (filas) 	Grafico(s) que permita(n) visualizar los resultados de una consulta MDX.

Los resultados y requerimientos obtenidos en la tabla 9, permitieron establecer relaciones entre los requerimientos y las técnicas de visualización, con el objetivo de determinar si la técnica de visualización analizada permite adaptarse a los requerimientos deseados para cada técnica de análisis tal como lo muestra la tabla 10.

Tabla 10. Análisis comparativo entre los requerimientos y las técnicas de visualización

Técnica de visualización	Tipos de análisis					
	Atributos	Instancias	Agrupamiento	Árboles de Decisión	Reglas de Asociación	Multidimensional
Gráfico de caja y bigotes	Si	No	No	No	No	No
Gráfico de burbujas	No	No	Si	No	No	No
Dendograma	No	No	Si	Si	No	No
Diagrama de red	No	No	Si	Si	Si	No
Gráfico de Embalaje de Círculos	No	No	Si	No	No	No
Gráfico de área	Si	No	No	No	No	No
Gráfico Circular Particionado	Si	No	No	No	No	No
Mapa de árbol	No	No	No	Si		
Diagrama de Voronoi	No	No	Si	No	No	No
Gráfico de coordenadas paralelas	No	No	No	No	No	Si
Diagrama de dispersión matricial	No	Si	Si	No	No	No
Gráfico de barras horizontales	Si	No	No	No	No	No
Diagrama de árbol	No	No	No	Si	No	No
Gráfico de dispersión	No	Si	Si	No	No	No
Gráfico de árbol de sangría	No	No	No	Si	No	No
Gráfico circular	Si	No	No	No	No	No
Gráfico de línea	Si	No	No	No	No	No
Gráfico de barras	Si	No	No	No	No	No
Gráfico de anillos	Si	No	No	No	No	No
Tablas	Si	Si	Si	No	Si	Si

Para la implementación de las técnicas de visualización, se consultaron bibliotecas de visualización de datos teniendo en cuenta las siguientes propiedades.

- Orientadas a la web
- Licencia libre
- Alto contenido en técnicas de visualización de datos.

De acuerdo a las características anteriores, se seleccionaron cuatro bibliotecas D3.js (Data-DrivenDocuments, 2013), InfoVis.js (Belmonte, 2013), Vis.js (Vis.js un sistema de interacción visual, 2014) y Chart.js (Chart.js Documentation, 2014)

Para seleccionar un determinado gráfico y su respectiva biblioteca de implementación, se hizo un análisis de selección, el cual se basó principalmente en cuatro características que debe cumplir un componente gráfico para ser utilizado en la herramienta VisualBITool. Los resultados de este análisis se presentan en la tabla 11.

Montilva et al. (2014,3), señalan cuatro características de un componente reutilizable, de la siguiente manera:

Genérico. Es reutilizable, su contenido puede ser implementado o usado en otras aplicaciones web.

Auto contenido. El componente que se desea reutilizar debe ser fácil de integrar a la aplicación que se desea realizar, por lo tanto no debe depender por completo de otros componentes,” para cumplir su función de forma tal que pueda ser desarrollado, probado, optimizado, utilizado, entendido y modificado individualmente”. (Montilvaet al., 2014,3)

Mantenido. El componente debe ser fácil de manipular, es decir, debe permitir adaptar su código a las necesidades del desarrollador y realizar mejoras continuas, dando paso a que las nuevas versiones incluyan correcciones, optimizaciones y nuevas características.

Independiente de la plataforma. El componente que se desea reutilizar, debe funcionar en diferentes plataformas como Hardware, sistemas operativos y si es web en los diferentes navegadores.

En la tabla 11, se presenta la evaluación realizada a las diferentes técnicas de visualización de datos de acuerdo a las características de un componente reutilizable definidas anteriormente.

El proceso de evaluación se realizó de la siguiente manera:

- Se seleccionó cada técnica de visualización de datos (ver tabla 10) que luego se investigaron en la documentación de las cuatro bibliotecas gráficas definidas anteriormente, con el fin de obtener su respectiva implementación.
- Al código de implementación de la biblioteca gráfica seleccionada, se le evaluó el cumplimiento de las cuatro características de un componente reutilizable.
- Si la biblioteca gráfica cumplía con las cuatro características, era seleccionada para ser implementada en la herramienta de visualización VisualBITool.
- Si la biblioteca gráfica no cumple alguna de las cuatro características de un componente reutilizable, en la columna de observaciones, de manera general, se da a conocer la razón del incumplimiento.

Tabla 11. Evaluación de las características de un componente reutilizable

Técnica de visualización	Biblioteca gráfica	CARACTERISTICAS				Observación
		Genérico	Auto contenido	Mantenido	Independiente de la plataforma	
Diagramas de Caja y Bigotes	D3.js	Si	Si	No	Si	Las variables de entrada son limitadas a dos.
Gráfico de burbujas	D3.js	Si	Si	Si	Si	
Dendograma	D3.js	Si	Si	NO	Si	No permite modificar el código fuente para adicionar nuevas variables de entrada.
Diagrama de red	InfoVis.js	Si	Si	No	Si	Cuando el número de nodos supera a 100, el gráfico se torna confuso, es decir no se puede distinguir la información que presenta
	D3.js	Si	Si	No	Si	Cuando el número de nodos supera a 100, el gráfico se torna confuso, es decir no se puede distinguir la información que presenta.
	Vis.js	Si	Si	No	Si	No es eficiente procesando grandes cantidades de datos
Gráfico de Embalaje de Círculos	D3.js	Si	Si	No	Si	No es posible adicionar etiquetas sobre los puntos de cada círculo
Gráfico de área	D3.js	Si	Si	No	Si	Cuando los datos son mayor a 100,La organización de las etiquetas en los dos ejes se vuelve confusa,es decir no se puede distinguir la información que presenta
	Vis.js	Si	Si	No	Si	No es posible adicionar etiquetas sobre los puntos que forman el area.
	chart.js	Si	Si	Si	Si	

Gráfico Circular particionado	InfoVis.js	Si	Si	No	Si	Cuando la cantidad de datos excede los 100, se genera un gráfico de difícil comprensión es decir las etiquetas de información empiezan a superponerse entre ellas, dificultando así distinguirlas.
	D3.js	Si	Si	No	Si	Gráfico confuso para ser analizado, es decir no se puede distinguir la información que presenta. No es óptimo para visualizar más de 100 datos.
Mapa de árbol	InfoVis.js	Si	Si	No	Si	Permite visualizar solo aquellos árboles que poseen como máximo ocho niveles.
	D3.js	Si	Si	Si	Si	
Diagrama de Voronoi	D3.js	Si	Si	No	Si	No permite visualizar etiquetas o información de los datos de entrada.
Gráfico de coordenadas paralelas	D3.js	Si	Si	No	Si	Gráfico confuso, es decir cuando la cantidad de datos superan los 500, las líneas que relacionan los datos no se pueden distinguir.
Diagrama de dispersión matricial	D3.js	Si	Si	No	Si	Es factible solo para representación de datos numéricos.
Gráfico de barras horizontales	D3.js	Si	Si	No	Si	No permite extender su área de graficación cuando ingresan nuevos datos.

Diagrama de árbol	Infovis.js	Si	Si	No	Si	No se permite adicionar etiquetas en las relaciones entre nodos.
	D3.js	Si	Si	Si	Si	
Gráfico de dispersión	D3.js	Si	Si	Si	Si	
Gráfico de árbol de sangría	D3.js	Si	Si	Si	Si	
Gráfico circular	InfoVis.js	Si	Si	Si	Si	
	D3.js	Si	Si	No	Si	No permite colocar etiquetas de información en cada fracción del círculo para su mayor comprensión.
	chart.js	Si	Si	No	Si	No permite extender su área de graficación de forma dinámica cuando el número de variables a visualizar aumente.
Gráfico de línea	D3.js	Si	Si	No	Si	No permite extender su área de graficación cuando ingresan nuevos datos.
	Vis.js	Si	Si	No	Si	No permite colocar etiquetas de información en cada punto que forma la línea.

Gráfico de barras	D3.js	Si	Si	No	Si	No permite extender su area de gráficación cuando ingresan nuevos datos.
	InfoVis.js	Si	Si	No	Si	
	Vis.js	Si	Si	No	Si	Codigo reducido, lo cual limita mejorar la presentacion del grafico, por ejemplo, colores, etiquetas y estilos
	Chart.js	Si	Si	Si	Si	
Gráfico de anillos	D3.js	Si	Si	No	Si	No permite colocar etiquetas de información en cada fracción del circulo para su mayor comprensión.
	chart.js	Si	Si	No	Si	No permite extender su area de gráficación de forma dinámica cuando el número de variables a visualizar aumenta.
Tablas	HTML	Si	Si	Si	Si	

De acuerdo con la tabla 11, en la tabla 12, se presentan el resumen de los resultados de la evaluación realizada a las diferentes técnicas de visualización de datos, acorde con las características de un componente reutilizable, definidas anteriormente, donde la columna de la tabla 12 “**Componente gráfico reutilizable**”, es la que determina si un componente es reutilizable, dependiendo de la información plasmada en la tabla 11.

Tabla 12. Resumen de la evaluación de las características de un componente reutilizable

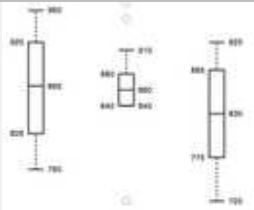

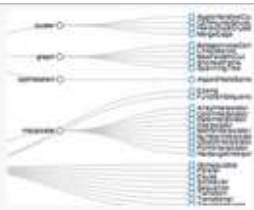
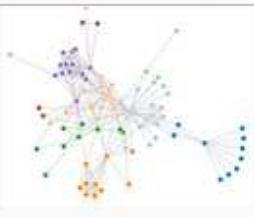
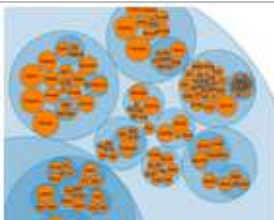
Técnica de visualización	Imagen gráfico	Biblioteca gráfica	Formato de datos	Componente gráfico reutilizable
Diagramas de Caja y Bigotes		D3.js	CSV	No
Gráfico de burbujas		D3.js	Json	Si
Dendograma		D3.js	Json	No
Diagrama de red		InfoVis.js	Json.	No
		D3.js	Json	No
		Vis.js	JavaScript	Si
Gráfico de Embalaje de Círculos		D3.js	Json	No

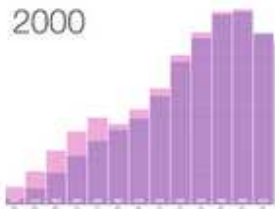
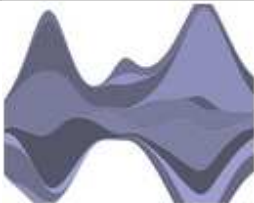
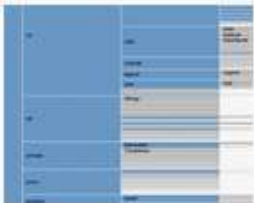
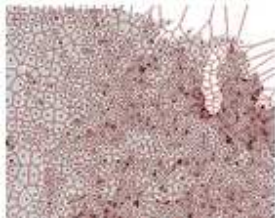
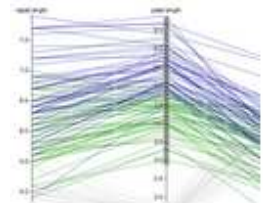
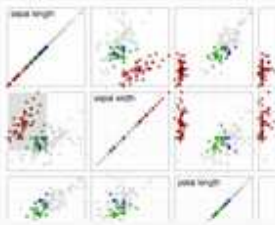
Gráfico de Barras		InfoVis.js	Json	No
		chart.js	JavaScript	Si
		D3.js	CSV	No
		Vis.js	JavaScript	No
Gráfico de área		D3.js	JavaScript	No
		Vis.js	JavaScript	No
		chart.js	JavaScript	Si
Mapa de árbol.		InfoVis.js	Json	No
		D3.js	Json	Si
Diagrama de Voronoi		D3.js	CSV	No
Gráfico de coordenadas paralelas		D3.js	CSV	No
Diagrama de dispersión matricial		D3.js	CSV	No

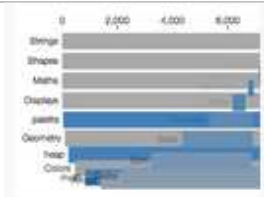
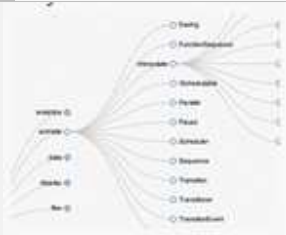
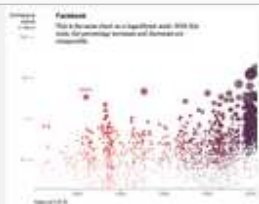
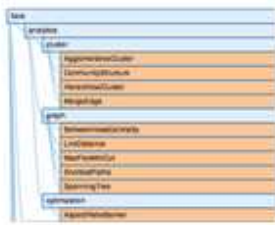
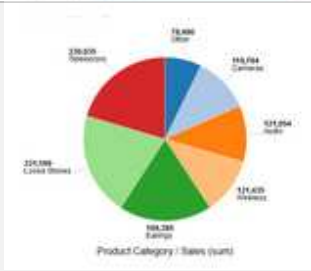
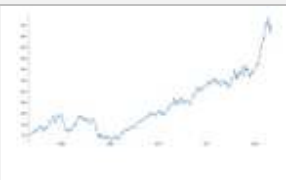



Gráfico de Barras horizontales		D3.js	Json	No
Gráfico de árbol		infovis	Json	No
		D3.js	Json	Si
Gráfico de dispersión		D3.js	CSV	Si
Gráfico de árbol de sangría.		D3.js	Json	Si
Gráfico circular		InfoVis.js	Json.	Si
		D3.js	CSV	No
		chart.js	JavaScript	No
		Vis.js	JavaScript	Si
Gráfico de línea		D3.js	tsv	No
		Vis.js	JavaScript	No
Gráfico de anillos		D3.js	CSV	No

		chart.js	JavaScript	No
Gráfico circular particionado		InfoVis.js	Json	Si
		D3.js	JavaScript	No
		chart.js	JavaScript	No
Tablas		HTML	etiquetas HTML	Si

En la tabla 13, se muestra el resumen de gráficos que fueron seleccionados como componentes reutilizables para ser implementados en las diferentes tareas que realiza la herramienta VisualBITool.

Tabla 13. Resumen de las técnicas de visualización seleccionadas

Tarea de análisis de datos	Técnica de visualización de datos	Biblioteca gráfica
Atributos	Gráfico de barras	Chart.js
	Gráfico de área	Chart.js
	Gráfico Circular particionado	InfoVis.js
	Tabla	HTML
Instancias	Gráfico de dispersión	D3.js
Árboles de decisión	Diagrama de árbol	D3.js
	Gráfico de árbol de sangría	D3.js
Agrupamiento	Gráfico de burbujas	D3.js
	Mapa de árbol	D3.js
	Tabla	HTML
Reglas de asociación	Gráfico de red	Vis.js
	Tabla	HTML
AnálisisOLAP	Tabla	HTML

3. METODOLOGÍA

Para el desarrollo de la herramienta de visualización de datos (VisualBITool), fue necesario seleccionar una metodología de desarrollo de proyectos ágil, que permita entregar resultados de forma rápida y así aumentar la productividad, sin descuidar la calidad de la herramienta.

3.1 SELECCIÓN DE LA METODOLOGÍA

La tabla 14, muestra la comparación de tres metodologías de desarrollo ágil para la construcción de software.

Tabla 14. Características de tres metodologías de desarrollo ágil de software (Bustos, 2014, 50)

AGILE	SCRUM	XP
Enfocado en la administración	Enfocado en la administración	Enfocado en los desarrolladores
Permite cambios en el producto	Permite cambios en el producto	Permite cambios en el producto
Produce software de calidad	Produce software de calidad	Produce software de calidad aceptable
Reduce el riesgo	Reduce el riesgo	El riesgo depende de la administración del proyecto
Reduce el costo	Reduce el costo	Reduce el costo
Facil de entender	Facil de entender	Moderadamente sencilla de entender
Entregas cortas	Entregas cortas	Entregas cortas
Requiere multiples roles	Requiere pocos roles	Requiere múltiples roles
Es una metodologia completa	Simplifica a AGILE	Es una metodologia completa
Control continuo sobre el desarrollo del software	Control continuo sobre el desarrollo del software	Interés en las personas y no en el procesos como elemento principal
Tiempo de entrega corto	Tiempo de entrega de dos a cuatro semanas	Tiempo de entrega de tres semanas
Cuando el dueño del producto muestra conformidad con las tareas realizadas estas no se vuelven a tocar	Cuando el dueño del producto muestra coonformidad con las tareas realizadas estas no se vuelven a tocar	Cuando el dueño del producto muestra conformidad con las tareas realizadas estas son susceptibles a cambios futuros.

De acuerdo a la información obtenida en la tabla comparativa entre las tres metodologías de desarrollo ágil (Ver tabla 14), se ha optado por SCRUM, porque es la metodología que mejor se adapta al desarrollo de la herramienta VisualBITool y no es tan compleja para su aplicación. Además, garantiza la calidad de software, el cumplimiento de los requerimientos establecidos por los clientes y ahorro de tiempo en el proceso de desarrollo de software. Esta metodología es una versión simplificada de una metodología completa como lo es AGILE, lo que facilita obviar algunos procesos que requieren tiempo y que no son necesarios.

3.1.1 SCRUM. Bustos (2014,34), la define como:

Es una metodología que consiste en aplicar una serie de buenas prácticas, como separación de roles, reuniones de equipo y revisiones, dentro de un proceso iterativo grupal, en el que se hacen entregas parciales, que se han priorizado con el objetivo de entregar resultados rápidamente.

SCRUM permite mantener una constante interacción con el cliente, lo cual beneficia al quipo de desarrollo en garantizar que lo que se está haciendo va acorde con el gusto y necesidad del usuario que dispondrá del producto final que se está desarrollando, con esto se puede ahorrar tiempo, costos y calidad del producto software.

SCRUM permite fijar tiempos de entrega de resultados parciales, en diferentes reuniones que se efectúan entre los integrantes del equipo de trabajo, lo cual ayuda a administrar el tiempo para la ejecución de tareas ya definidas por los desarrolladores, así se lleva un control de cumplimiento en el desarrollo del producto final, como lo señala Bustos (2014,49):

Es una metodología de desarrollo de proyectos ágil, que permite entregar resultados rápidamente y aumentar la productividad sin descuidar la calidad, por medio de periodos cortos e iterativos. Se caracteriza por ser una metodología ligera, fácil de aprender pero compleja de dominar, para su funcionamiento utiliza distintos roles que se pueden distribuir en un equipo, que los interrelaciona por medio de eventos, reglas y artefactos.

SCRUM se fundamenta en tres aspectos principales planteados por Schwaber & Sutherland (2013,5):

Transparencia. Los miembros del equipo de desarrollo del proyecto, deben manejar los mismos términos que se usan o serán usados en el transcurso de la elaboración del producto final que será entregado, definiendo así los estados del proceso que representaran lo mismo para todos los encargados, con el fin de evitar confusiones que lleven a demoras o errores en la ejecución del proyecto.

Inspección. Los avances o productos parciales del proyecto, se deben revisar frecuentemente por el equipo responsable y por el cliente final, lo cual permitirá detectar errores o posibles cambios en el proceso. Estos tiempos de revisión deben ser acordados por el equipo encargado del proyecto teniendo en cuenta que no sean muy prolongados, pero también que no interfiera con el trabajo que se esté realizando.

Adaptación. Cuando se han detectado problemas o errores en la ejecución del proyecto, que pueden interferir en logra el objetivo, se deberá tomar medidas correctivas; con el fin de retornar el rumbo, adaptando así, el desarrollo del proyecto a los nuevos requerimientos que garanticen un buen término, es decir, que el producto final cumpla con los requerimientos iniciales.

SCRUM como es una metodología de desarrollo ágil, implica crear ciclos breves para el desarrollo, que en términos comunes se le llama iteraciones pero que SCRUM se llaman “Sprints” como lo señalan Schwaber & Sutherland (2013,9).

El ciclo de desarrollo de SCRUM, que se muestra en la figura 49, está compuesto por cinco fases, señaladas por Trigas & Domingo (2014, 33). Estas fases son:

Concepto. Se definen las características generales que tendrá el producto acorde a las necesidades del cliente, y de acuerdo a esas características definidas, se seleccionará el equipo que se encargará de su desarrollo.

Especulación. El equipo de desarrollo con la información recolectada sobre el producto que se desea desarrollar, empieza a organizarse en cuanto a tiempo y costos, de acuerdo con Trigas & Domingo (2014, 33) que indican:

Se construirá el producto a partir de las ideas principales y se comprueban las partes realizadas y su impacto en el entorno.

Estas fases se repiten en cada iteración y consiste, en rasgos generales, en:

- Desarrollar y revisar los requisitos generales.
- Mantener la lista de las funcionalidades que se esperan.
- Plan de entrega. Se establecen las fechas de las versiones, hitos e iteraciones.
- Se medirá el esfuerzo en el proyecto.

Exploración. En esta fase, el equipo desarrolla el proyecto con las ideas que se plantearon en la fase de especulación.

Revisión. El equipo encargado del desarrollo del producto, lleva un control sobre la funcionalidad del desarrollo parcial, revisando así, lo que ya está construido y contrastando con los requerimientos deseados, con el fin de evaluar que lo que se está haciendo va dirigido al objetivo deseado.

Cierre. Esta fase no quiere decir que el desarrollo del producto ya se encuentra finalizado, ya que en esta fase, se entregan versiones del producto en unas fechas determinadas, lo cual conlleva a su revisión, que traerá consigo algunos cambios denominados “mantenimiento”, hecho que garantizará, que el producto final cumpla con el objetivo deseado.

Figura 49. Ciclo de desarrollo ágil (Trigas & Domingo, 2014,34)



3.1.2 UML. Son las iniciales de Lenguaje Unificado de Modelado (Unified Modeling Language), que mediante el uso de diagramas facilita la representación de lo que un software tiene o deberá tener en su estructura y funcionamiento, como lo definen Rumbaugh, Jacobson & Booch (1999), en el capítulo 1, “Es un lenguaje de modelado visual que se usa para especificar, visualizar, construir y documentar artefactos de un sistema de Software”.

Diagramas de clases. El diagrama de clases permite mediante representaciones estáticas, representar la estructura de un sistema orientado a objetos, en la cual, cada clase que se representa con un rectángulo es un objeto, dando a conocer así, sus atributos y las relaciones que este tiene con otros objetos representados en clases (Rumbaugh, Jacobson & Booch, 1999,37-53).

Diagrama de paquetes. El Diagrama de paquetes, permite agrupar los elementos que conforman a un sistema orientado a objetos de acuerdo a una función común (Rumbaugh, Jacobson & Booch, 1999,87-90), por ejemplo, en un paquete denominado persona, se pueden agrupar todas las clases que cumplan funciones con procesar información de clientes o usuarios humanos y en otro paquete se pueden agrupar clases que procesen información diferente a la de las personas, esto ayudara a organizar el trabajo de desarrollo que se desea hacer.

Diagrama de casos de uso. Los diagramas de caso de uso, permiten representar la interacción del usuario con el sistema, dando a conocer las diferentes acciones necesaria para generar un resultado deseado por un usuario o como se lo denomina en estos diagramas, un actor (Rumbaugh, Jacobson & Booch, 1999, 24).

3.2 IDENTIFICACIÓN DE REQUERIMIENTOS

En esta sección, se presentan los requerimientos funcionales, los cuales definen las funciones que el sistema VisualBITool será capaz de realizar.

3.2.1 Requerimientos funcionales. A continuación, se presentan los diferentes requerimientos funcionales, descritos en la tabla 15.

Tabla 15. Requerimientos funcionales de VISUALBITOOL

Identificación	RF01
Nombre	Ejecución de la herramienta
Características	La aplicación deberá ejecutarse.
Descripción	La aplicación deberá ejecutarse y mostrar la interfaz de usuario.
Prioridad	Alta

Identificación	RF02
Nombre	Interfaz gráfica de usuario.
Características	Una interfaz amigable para el usuario.
Descripción	Se desarrollará una interfaz gráfica que permita visualizar los diferentes módulos de la herramienta.
Prioridad	Alta

Identificación	RF03
Nombre	Mensajes
Características	Mostrar mensajes de ayuda, respuesta o error cuando sea necesario.
Descripción	La herramienta mostrará mensajes de ayuda para que el usuario tenga una guía para manipular la GUI, La herramienta mostrará mensajes de respuesta o error al cargar los diferentes archivos, ingreso de parámetros y realización de consultas.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF04
Nombre	Selección de Archivo.
Característica	Permitir seleccionar y cargar un archivo
Descripción	La herramienta permitirá cargar archivos de tipo ARFF, CSV para minería de datos y XML para el análisis OLAP.
Prioridad	Alta

Identificación	RF05
Nombre	Selección de tarea.
Característica	Permitir seleccionar una tarea para realizar
Descripción	Mediante el uso de un menú, el usuario podrá seleccionar la tarea que desea desarrollar.
Prioridad	Alta

Identificación	RF06
Nombre	Atributos del archivo.
Características	Permitir mostrar los atributos del archivo seleccionado.
Descripción	La herramienta permitirá visualizar los atributos de los archivos ARFF y CSV
Prioridad	Alta

Identificación	RF07
Nombre	Gráficos de los atributos.
Características	Permitir graficar los valores de cada atributo
Descripción	La herramienta permitirá seleccionar el tipo de gráfico en que desea visualizar los valores de cada atributo dependiendo si es categórico o numérico.
Prioridad	Alta

Identificación	RF08
Nombre	Distribución de datos.
Características	Permitir visualizar la distribución de instancias mediante un gráfico de dispersión.
Descripción	La herramienta permitirá generar un gráfico de dispersión, con el cual se relacionen tres variables o atributos.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF09
Nombre	Filtro eliminar atributos
Características	Permitir eliminar atributos irrelevantes del repositorio de datos.
Descripción	La herramienta permitirá seleccionar los atributos irrelevantes para posteriormente ser eliminados y generar un nuevo repositorio de datos.
Prioridad	Alta

Identificación	RF10
Nombre	Filtro discretización de atributos
Características	Permitir discretizar atributos de tipo numérico.
Descripción	La herramienta permitirá seleccionar en una lista, los atributos de tipo numérico que serán categorizados.
Prioridad	Alta

Identificación	RF11
Nombre	Parámetros de un árbol de decisión.
Características	Permitir ingresar y editar los parámetros en los árboles de decisión.
Descripción	La herramienta permitirá ingresar o editar el formulario de parámetros correspondientes a los árboles de decisión.
Prioridad	Alta

Identificación	RF12
Nombre	Información árbol de decisión.
Características	Mostrar información general del árbol de decisión.
Descripción	La herramienta permitirá visualizar mediante una tabla las características generales del árbol de decisión resultante.
Prioridad	Alta

Identificación	RF13
Nombre	Gráfico del árbol de decisión.
Características	Mostrar gráficamente el árbol de decisión.
Descripción	La herramienta permitirá visualizar gráficamente la estructura del árbol de decisión resultante.
Prioridad	Alta

Identificación	RF14
Nombre	Parámetros de agrupamiento(Clustering)
Características	Permitir ingresar el número de grupos a procesar.
Descripción	La herramienta permitirá ingresar la cantidad de grupos (Clusters) que se desea procesar.
Prioridad	Alta

Identificación	RF15
Nombre	Valores de los grupos.
Características	Mostrar los valores de cada grupo (Cluster) resultante.
Descripción	La herramienta permitirá visualizar mediante una tabla los valores correspondientes a cada grupo (Cluster).
Prioridad	Alta

Identificación	RF16
Nombre	Número de instancias.
Características	Permitir seleccionar el número de instancias a visualizar por cada grupo (Cluster).
Descripción	La herramienta permitirá ingresar la cantidad de instancias que se desea visualizar por cada grupo.
Prioridad	Alta

Identificación	RF17
Nombre	Gráfico de grupos (Clusters).
Características	Permitir graficar los grupos (Clusters).
Descripción	La herramienta permitirá construir la gráfica de los grupos (Clusters) generados.
Prioridad	Alta

Identificación	RF18
Nombre	Valores de una instancia.
Características	Mostrar los valores de una instancia de un grupo (Cluster) determinado.
Descripción	La herramienta permitirá visualizar mediante una tabla los valores de una instancia determinada al hacer clic sobre ella.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF19
Nombre	Características generales de una instancia.
Características	Mostrar información general de una instancia de un grupo (Cluster) determinado.
Descripción	La herramienta permitirá visualizar las características generales de una instancia determinada al colocar el puntero del mouse sobre la instancia.
Prioridad	Alta

Identificación	RF20
Nombre	Porcentaje de confianza de una Instancia.
Características	Mostrar porcentaje de confianza de una instancia con respecto al grupo en que pertenece.
Descripción	La herramienta permitirá visualizar el porcentaje de confianza de la instancia al final de la tabla de valores de esta.
Prioridad	Alta

Identificación	RF21
Nombre	Parámetros de las reglas de asociación.
Características	Permitir ingresar y editar los parámetros de las reglas de asociación.
Descripción	La herramienta permitirá ingresar y editar los campos que corresponden a los parámetros de las reglas de asociación.
Prioridad	Alta

Identificación	RF22
Nombre	Reglas de asociación generadas.
Características	Permitir mostrar las reglas resultantes con su respectivo factor de confianza.
Descripción	La herramienta permitirá visualizar mediante una tabla las reglas de asociación generadas con su respectivo factor de confianza.
Prioridad	Alta

Identificación	RF23
Nombre	Selección de reglas de asociación.
Características	Permitir seleccionar las reglas a visualizar.
Descripción	La herramienta permitirá seleccionar las reglas que se desee visualizar.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF24
Nombre	Visualización reglas de asociación.
Características	Mostrar gráfica de las reglas seleccionadas.
Descripción	La herramienta permitirá visualizar mediante un gráfico la estructura de cada una de las reglas.
Prioridad	Alta

Identificación	RF25
Nombre	Visualización reglas de asociación.
Características	Permitir mostrar antecedente y consecuente de una regla determinada.
Descripción	La herramienta permitirá visualizar mediante una tabla los datos del antecedente y consecuente de una regla seleccionada.
Prioridad	Alta

Identificación	RF26
Nombre	Conexión al repositorio de datos.
Características	Permitir generar conexión con el repositorio de datos.
Descripción	La herramienta permitirá efectuar la conexión al repositorio de datos, ingresando los parámetros de conexión solicitados.
Prioridad	Alta

Identificación	RF27
Nombre	Edición de la conexión.
Características	Permitir ingresar y editar los campos para realizar la conexión.
Descripción	La herramienta permitirá editar o ingresar los datos necesarios para efectuar la conexión.
Prioridad	Alta

Identificación	RF28
Nombre	Información del cubo.
Características	Permitir mostrar información del cubo seleccionado.
Descripción	La herramienta permitirá mediante un formulario, mostrar el nombre, las dimensiones, niveles y medidas del cubo.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF29
Nombre	Selección de dimensiones.
Características	Permitir seleccionar solo 3 dimensiones.
Descripción	La herramienta permitirá seleccionar las dimensiones deseadas, restringiéndolas a solo tres.
Prioridad	Alta

Identificación	RF30
Nombre	Selección de niveles.
Características	Permitir seleccionar un nivel por cada dimensión.
Descripción	La herramienta permitirá seleccionar un nivel por cada dimensión
Prioridad	Alta

Identificación	RF31
Nombre	Selección de medidas
Características	Permitir seleccionar las medidas.
Descripción	La herramienta permitirá seleccionar las medidas deseadas para realizar una consulta
Prioridad	Alta

Identificación	RF32
Nombre	Mostrar las dimensiones en un cubo
Características	Mostrar las dimensiones seleccionadas en un cubo.
Descripción	La herramienta permitirá visualizar el esquema de un cubo en el cual la dimensión seleccionada se ira ubicando en sus respectivos lados.
Prioridad	Alta

Identificación	RF33
Nombre	Ejecutar consultas
Características	Permitir ejecutar consultas al dar clic sobre la imagen del cubo.
Descripción	La herramienta permitirá ejecutar una consulta solo al dar clic en la imagen que representa el cubo multidimensional.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF34
Nombre	Escoger las dimensiones
Características	Permitir seleccionar las dimensiones para ejecutar consultas.
Descripción	La herramienta permitirá seleccionar las dimensiones que se involucraran en la consulta que se desea realizar, bien sea tres, dos o una.
Prioridad	Alta

Identificación	RF35
Nombre	Seleccionar los filtros
Características	Permitir crear filtros por cada consulta.
Descripción	La herramienta permitirá seleccionar los diferentes atributos de cada nivel que serán utilizados como filtros en la consulta que será ejecutada.
Prioridad	Alta

Identificación	RF36
Nombre	Visualización de una consulta
Características	Permitir visualizar los resultados de una consulta en tablas.
Descripción	La herramienta permitirá visualizar los resultados de una consulta en una tabla.
Prioridad	Alta

Identificación	RF37
Nombre	Consultas MDX
Características	Visualizar el formato de la consulta MDX.
Descripción	La herramienta permitirá visualizar la estructura de la consulta MDX que genere los datos mostrados en una tabla.
Prioridad	Alta

Identificación	RF38
Nombre	Crear consultas MDX
Características	Permitir que el usuario realice consultas MDX.
Descripción	La herramienta permitirá que el usuario pueda construir una consulta MDX y poderla ejecutar.
Prioridad	Alta

Continuación tabla 15.

Identificación	RF39
Nombre	Guardar la información
Características	Permitir guardar información de la visualización.
Descripción	La herramienta permitirá que el usuario pueda guardar la información que está visualizando, permitiéndole exportarlo a formato (xls) Excel.
Prioridad	Alta

3.3 CASOS DE USO

En esta sección se presentan los diferentes casos de uso de VisualBITool, empezando con una serie de tablas que describen cada caso de uso y luego se muestran sus esquemas.

3.3.1 Descripción de los casos de uso. A continuación se presenta una serie de tablas (mirar tablas 16 a la 24) en las cuales se describe paso a paso cómo se desarrolla un caso de uso en específico. Además, en la parte inferior de las tablas, se muestran los diferentes casos alternos que se pueden presentar por fuera del desarrollo normal del caso de uso.

Tabla 16. Caso de uso Ejecutar aplicación

Caso de uso: Ejecutar aplicación
1. El usuario ejecuta la aplicación
2. El sistema presenta la ventana principal GUI
3. El sistema muestra un menú con las diferentes tareas que se pueden realizar
4. El sistema mediante mensajes guía al usuario en el manejo de la interfaz.
Casos Alternos
<ul style="list-style-type: none"> Si no se ha inicializado el servidor Tomcat, el navegador muestra un mensaje de alerta y no se ejecuta la aplicación.

Tabla 17. Caso de uso Cargar archivo

Caso de uso: Cargar archivo
1. El usuario presiona el botón de examinar archivos.
2. El sistema presenta una ventana de explorador de archivos.
3. El usuario selecciona el archivo
4. El usuario presiona el botón de subir archivo.
5. El sistema muestra un mensaje de éxito al subir el archivo.
Casos Alternos
Si el usuario no ha seleccionado un archivo y presiona el botón subir archivo, el sistema muestra un mensaje de alerta y no permite que se ejecute la aplicación.

Tabla 18. Caso de uso Analizar atributos

Caso de uso: Analizar atributos.
1. El usuario presiona la opción calidad de datos
2. El sistema le muestra dos opciones de calidad de datos
3. El usuario selecciona la opción atributos.
4. El sistema presenta una lista de los atributos del archivo seleccionado.
5. El sistema muestra una tabla con los valores del atributo seleccionado y un icono para visualizar los datos en una gráfica.
6. El usuario con la opción filtros puede discretizar y eliminar atributos
7. El usuario presiona el icono.
8. El sistema muestra una ventana con la gráfica de los datos.
9. El usuario selecciona la opción Filtros.
10. El usuario selecciona un filtro de atributos según su necesidad.
11. El sistema aplica el filtro seleccionado sobre el repositorio de datos.
12. El sistema crea un nuevo repositorio de datos.
Casos Alternos
<ul style="list-style-type: none"> Si el usuario selecciona un archivo diferente a ARFF y CSV, el sistema muestra un mensaje de alerta y no permite continuar con el proceso.

Tabla 19. Caso de uso Visualizar instancias

Caso de uso: Visualizar Instancias.
1. El usuario presiona la opción calidad de datos
2. El sistema le muestra tres opciones de calidad de datos
3. El usuario selecciona la opción Instancias.
4. El sistema presenta parámetros que limitan el número de instancias a visualizar en un gráfico de dispersión.
5. El sistema muestra un gráfico de dispersión.
Casos Alternos
<ul style="list-style-type: none"> • Si el usuario selecciona un archivo diferente a ARFF y CSV, el sistema muestra un mensaje de alerta y no permite continuar con el proceso.

Tabla 20. Caso de uso Procesar árboles de decisión

Caso de uso: Procesar árboles de decisión.
1. El usuario presiona la opción Minería de datos.
2. El sistema le presenta tres opciones de Minería de datos.
3. El usuario selecciona la opción árboles de decisión.
4. El sistema presenta un formulario con los parámetros del árbol de decisión.
5. El usuario selecciona la opción editar y actualiza los campos de los parámetros del árbol de decisión.
6. El sistema muestra una tabla con la información general del árbol de decisión y un icono para visualizar el árbol de decisión.
7. El usuario presiona el icono.
8. El sistema muestra una ventana con el gráfico del árbol de decisión.
Casos Alternos
<ul style="list-style-type: none"> • Si el usuario selecciona un archivo diferente a ARFF y CSV, el sistema muestra un mensaje de alerta y no permite continuar con el proceso.

Tabla 21. Caso de uso Procesar agrupamiento

Caso de uso: Procesar agrupamiento (Clustering).
1. El usuario presiona la opción Minería de datos.
2. El sistema le presenta tres opciones de Minería de datos.
3. El usuario selecciona la opción de agrupamiento.
4. El sistema presenta un campo para ingresar la cantidad de grupos a procesar.
5. El usuario ingresa la cantidad de grupos a procesar.
6. El usuario presiona el botón ejecutar.
7. El sistema muestra una tabla con los valores de cada grupo (Cluster) y un campo para ingresar la cantidad de instancias por grupo a visualizar.
8. El usuario ingresa la cantidad de instancias.
9. El usuario presiona el botón ver gráfico
10. El sistema muestra un gráfico de los grupos procesados.
11. El usuario presiona sobre un punto de un grupo determinado.
12. El sistema muestra una tabla con la información de la instancia seleccionada.
Casos Alternos
<ul style="list-style-type: none"> • Si el usuario selecciona un archivo diferente a ARFF y CSV, el sistema muestra un mensaje de alerta y no permite continuar con el proceso. • Si el usuario no ingresa datos en los campos que se solicitan, el sistema muestra un mensaje de alerta y no permite continuar con el proceso.

Tabla 22. Caso de uso Generar reglas de asociación

Caso de uso: Generar reglas asociación.	
1	El usuario presiona la opción Minería de datos.
2	El sistema le presenta tres opciones de Minería de datos.
3.	El usuario selecciona la opción de reglas de asociación.
4.	El sistema presenta un formulario para ingreso de parámetros de reglas de asociación
5.	El usuario selecciona la opción editar y actualiza los parámetros de las reglas de asociación.
6.	El usuario presiona el botón ejecutar.
7.	El sistema muestra una tabla con la información de cada regla de asociación.
8.	El usuario selecciona las reglas de asociación que desea visualizar.
9.	El usuario presiona el botón ver gráfico
10.	El sistema muestra un gráfico de las reglas seleccionadas.
11.	El usuario selecciona una regla de asociación.
12.	El sistema muestra la información de la regla seleccionada.
Casos Alternos	
<ul style="list-style-type: none"> • Si el usuario selecciona un archivo diferente a ARFF y CSV el sistema muestra un mensaje de alerta y no permite continuar con el proceso. • Si el usuario no ingresa datos en los campos que se solicitan, el sistema muestra un mensaje de alerta y no permite continuar con el proceso. • Si el archivo contiene atributos numéricos, el sistema muestra un mensaje de alerta y no permite continuar con el proceso. 	

Tabla 23. Caso de uso Procesar conexión

Caso de uso: Procesar Conexión.	
1.	El usuario presiona la opción análisis de datos
2.	El sistema le presenta dos opciones.
1.	El usuario selecciona la opción de Conexión.
2.	El sistema presenta un formulario para ingreso de parámetros de la conexión.
3.	El usuario selecciona la opción editar y actualiza los parámetros de la conexión.
4.	El usuario presiona el botón conectar.
5.	El sistema muestra un mensaje (Conexión exitosa).
Casos Alternos	
<ul style="list-style-type: none"> • Si la base de datos no existe, el sistema muestra un mensaje de alerta y no permite continuar con el proceso. • Si el archivo no es de extensión XML, el sistema muestra un mensaje de alerta y no permite que se ejecute la aplicación. 	

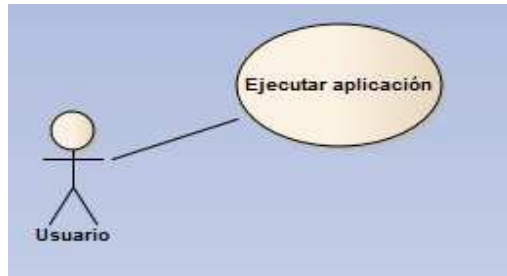
Tabla 24. Caso de uso Procesar análisis OLAP

Caso de uso: Procesar Análisis OLAP.
1. El usuario presiona la opción análisis de datos
2. El sistema le presenta dos opciones.
3. El usuario selecciona la opción de Análisis OLAP.
4. El sistema presenta un formulario con información del cubo y una imagen de un cubo.
5. El usuario selecciona tres dimensiones y un nivel por cada dimensión.
6. El usuario selecciona una o varias medidas.
7. El sistema coloca las dimensiones y los niveles seleccionados en los lados de la imagen del cubo.
8. El usuario en el cubo selecciona las dimensiones con las que desea efectuar la consulta.
9. El usuario presiona sobre la dimensión y selecciona la opción filtro.
10. El sistema muestra una ventana con los valores del nivel a filtrar.
11. El usuario selecciona los filtros.
12. El usuario presiona el botón ok.
13. El usuario presiona sobre la imagen del cubo.
14. El sistema muestra una ventana con el resultado de la consulta realizada en una tabla.
Casos Alternos
<ul style="list-style-type: none"> • Si no se ha efectuado la conexión, el sistema muestra un mensaje y no permite continuar con el proceso.

3.3.2 Diagramas de los casos de uso. En esta sección se representa gráficamente la estructura de los casos de uso de VisualBITool. En la cual se da a conocer los diferentes actores, usuarios y su relación con el sistema.

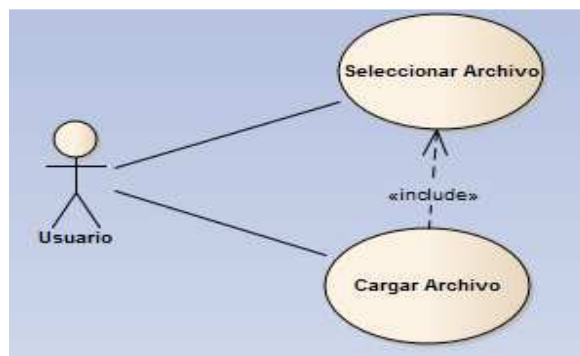
Caso de uso Ejecutar aplicación. La figura 50, muestra el diagrama de este caso de uso.

Figura 50. Ejecutar aplicación



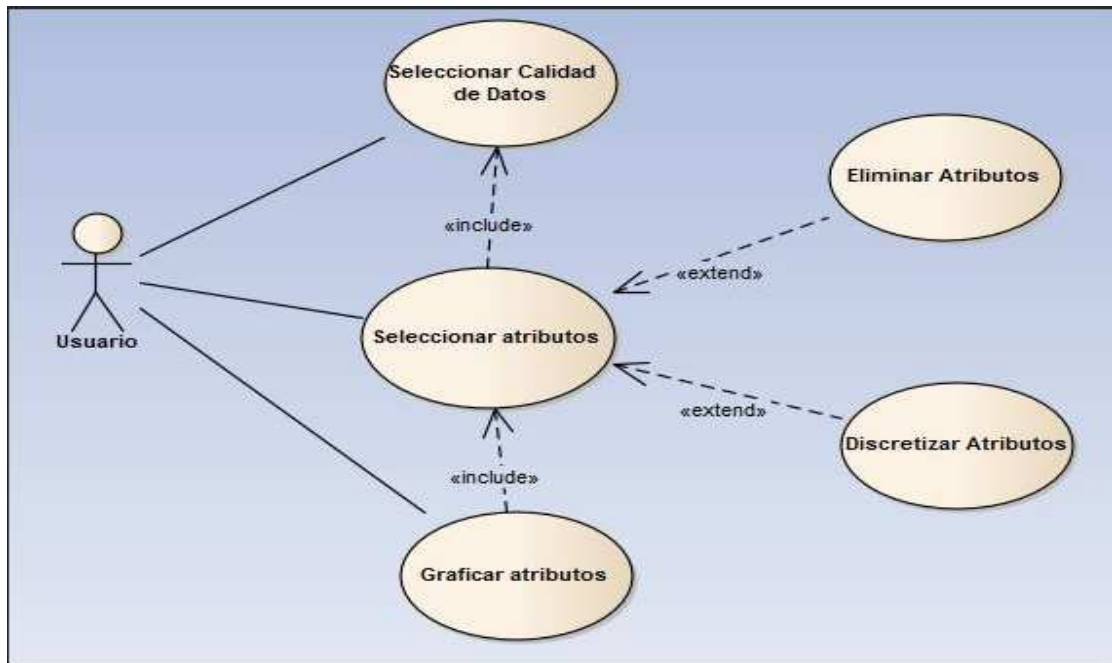
Caso de uso Cargar archivo. La figura 51, muestra el diagrama de este caso de uso.

Figura 51. Cargar archivo



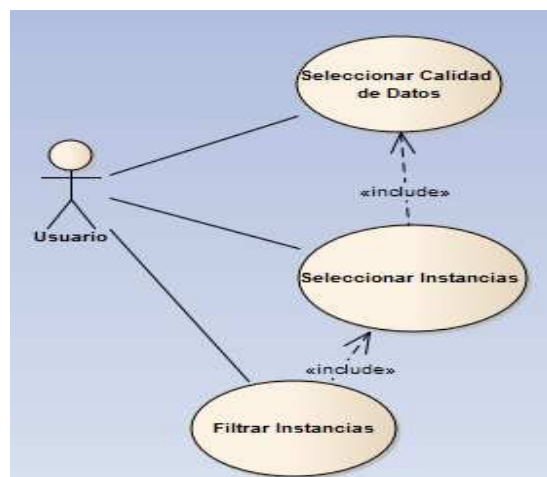
Caso de uso Analizar atributos. La figura 52, muestra el diagrama de este caso de uso.

Figura 52. Analizar atributos



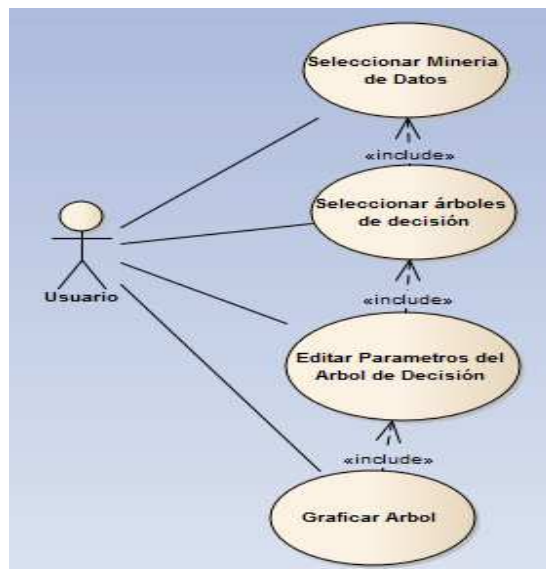
Caso de uso Visualizar instancias. La figura 53, muestra el diagrama de este caso de uso.

Figura 53. Visualizar instancias



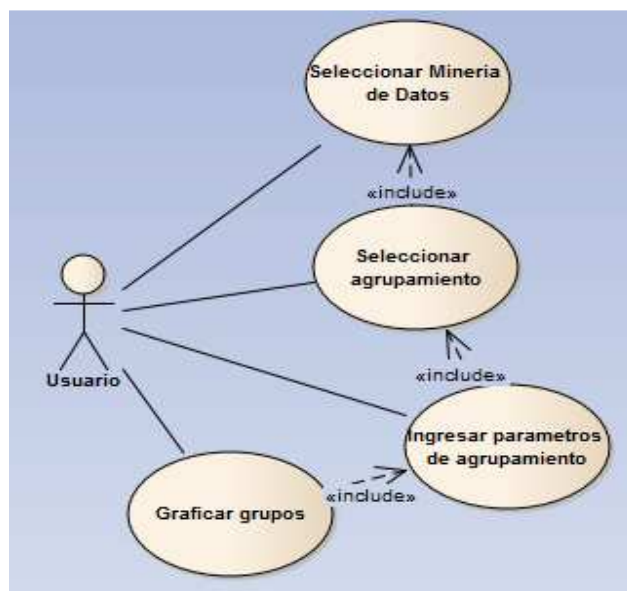
Caso de uso Procesar árboles de decisión. La figura 54, muestra el diagrama de este caso de uso.

Figura 54. Procesar árboles de decisión



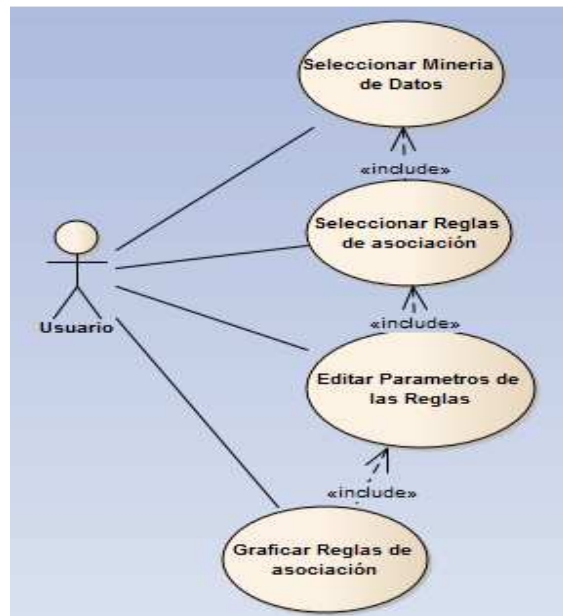
Caso de uso Procesar agrupamiento. La figura 55, muestra el diagrama de este caso de uso.

Figura 55. Procesar agrupamiento



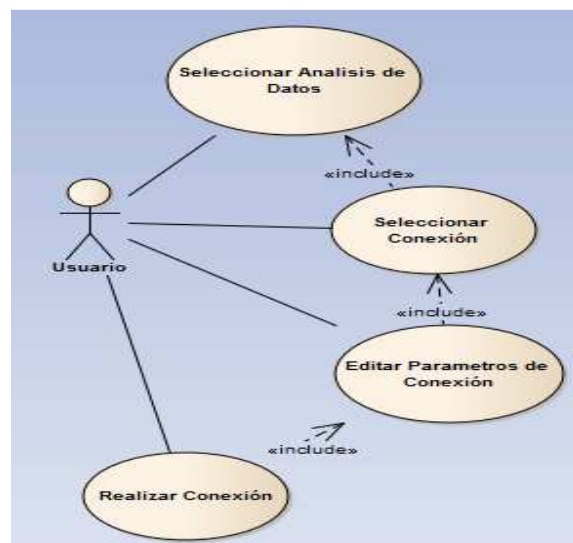
Caso de uso Generar reglas de asociación. La figura 56, muestra el diagrama de este caso de uso.

Figura 56. Generar reglas de asociación



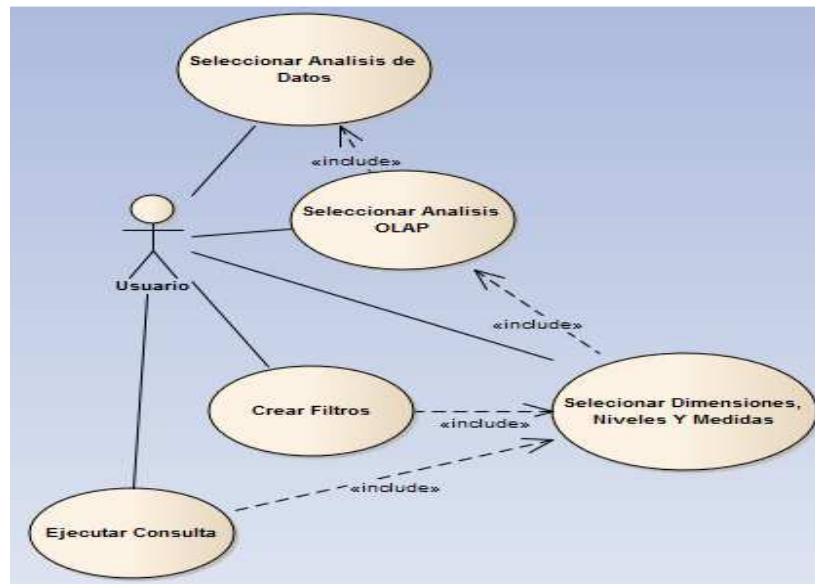
Caso de uso Procesar conexión. La figura 57, muestra el diagrama de este caso de uso.

Figura 57. Procesar conexión



Caso de uso Procesar análisis OLAP. La figura 58, muestra el diagrama de este caso de uso.

Figura 58. Procesar análisis OLAP



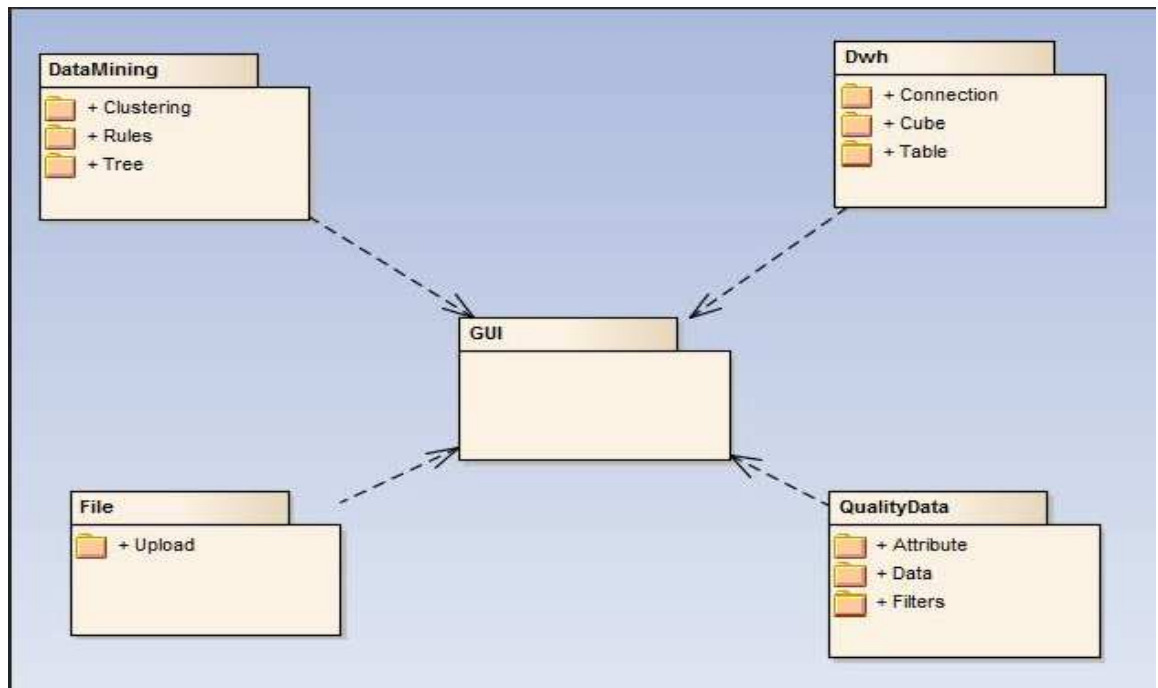
3.4 DIAGRAMAS DE PAQUETES

En esta sección se presenta los diagramas de los diferentes paquetes que integran la herramienta VisualBITool. Para una descripción más detallada de cada paquete, ir al capítulo 4 que trata de la arquitectura de VisualBITool en la sección 4.2.2 de este documento.

3.4.1 Paquete principal. En la figura 59, se representa el diagrama del paquete principal con sus respectivos paquetes que lo conforman.

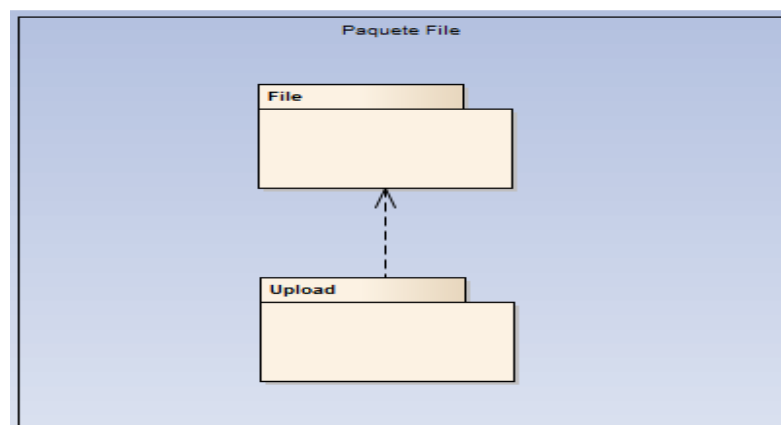
Paquete GUI. Representa la interfaz gráfica con la cual el usuario va a interactuar con la herramienta VisualBITool, ver figura 59.

Figura 59. Paquete principal



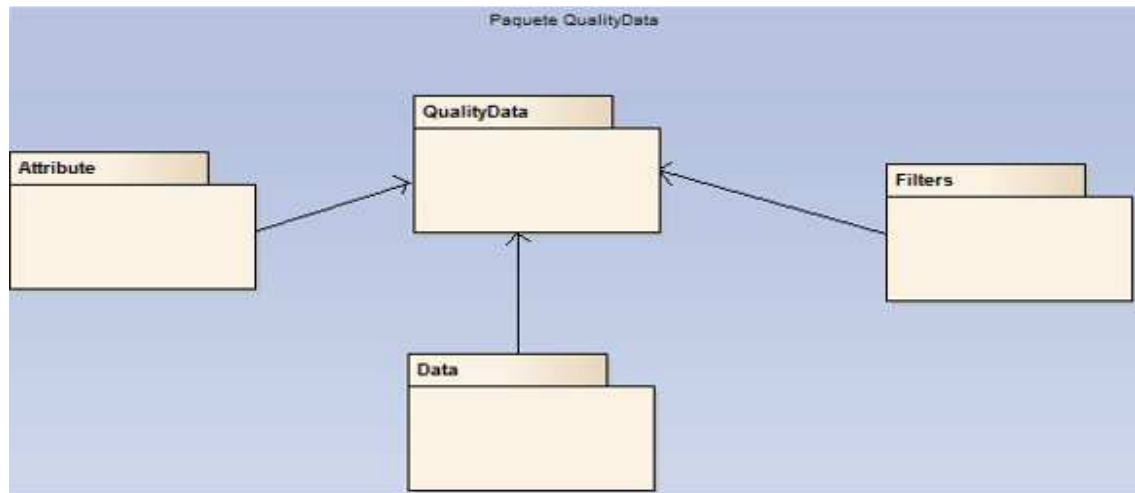
3.4.2 Paquete File. Es el encargado seleccionar los archivos idóneos para el funcionamiento de los diferentes módulos de la herramienta VisualBITool. Su diagrama se puede ver en la figura 60.

Figura 60. Paquete File



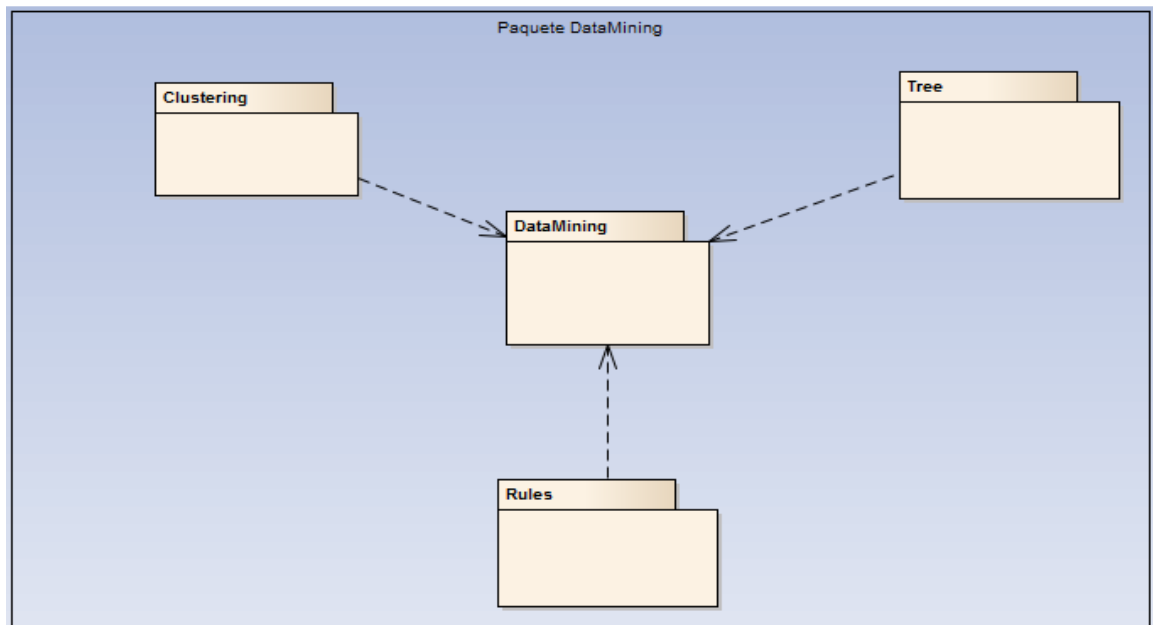
3.4.3 Paquete QualityData. Es el encargado de realizar las tareas de pre-procesamiento de los datos. Su diagrama se puede ver en la figura 61.

Figura 61. Paquete QualityData



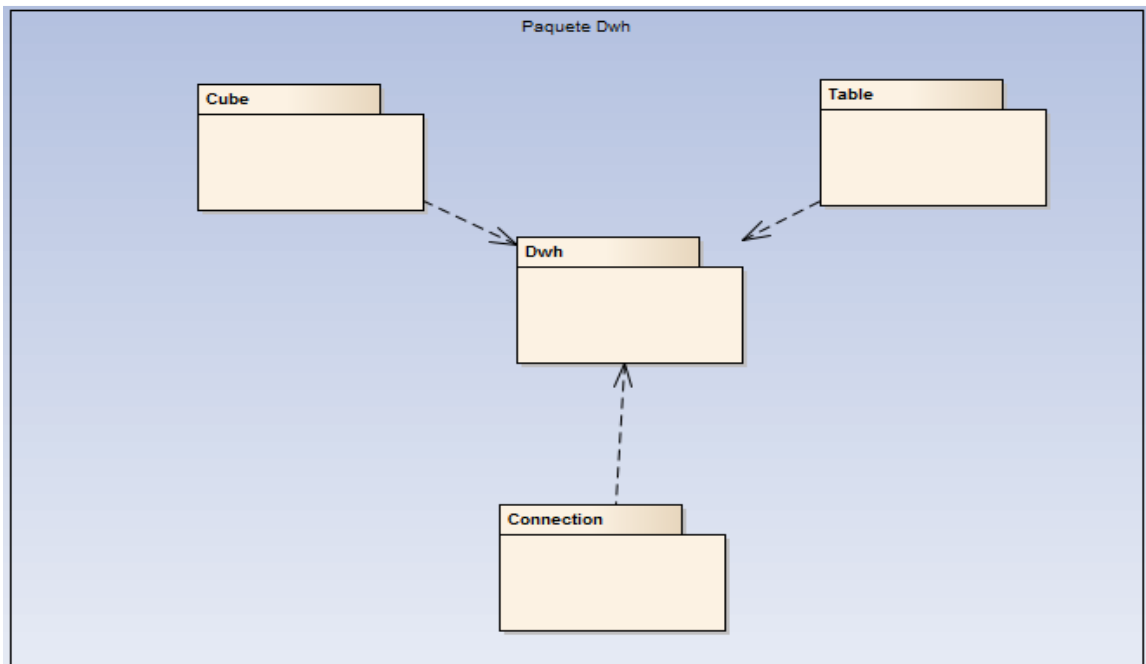
3.4.4 Paquete DataMining. Es el encargado de procesar las tareas que involucran las técnicas de minería de datos. La figura 62 muestra el diagrama de este paquete.

Figura 62. Paquete DataMining



3.4.5 Paquete Dwh. Es el encargado de realizar las tareas de conexión y de análisis multidimensional. Su diagrama se puede ver en la figura 63.

Figura 63. Paquete Dwh

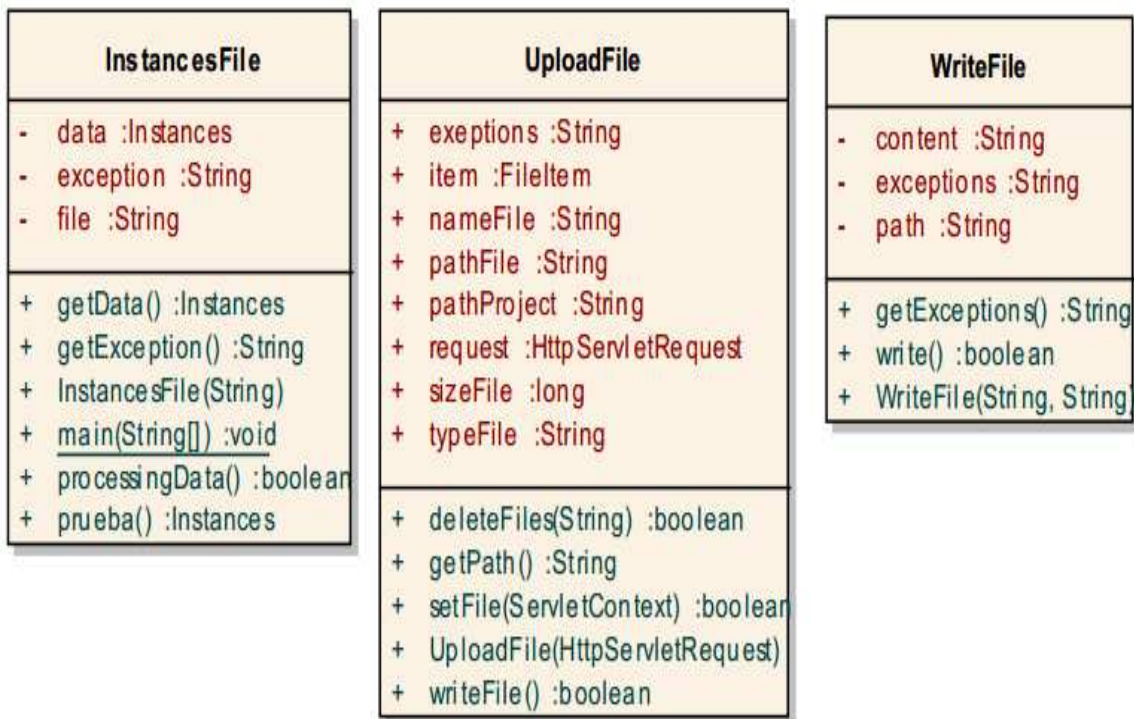


3.5 DIAGRAMAS DE CLASES

En esta sección se presentan los diagramas de clases por cada paquete. Para una descripción más detallada de cada clase, ir al capítulo 4, que trata de la arquitectura de VisualBITool en la sección 4.2.2 de este documento.

3.5.1 Clases del Paquete File. En la figura 64, se presenta los diagramas de clase de este paquete.

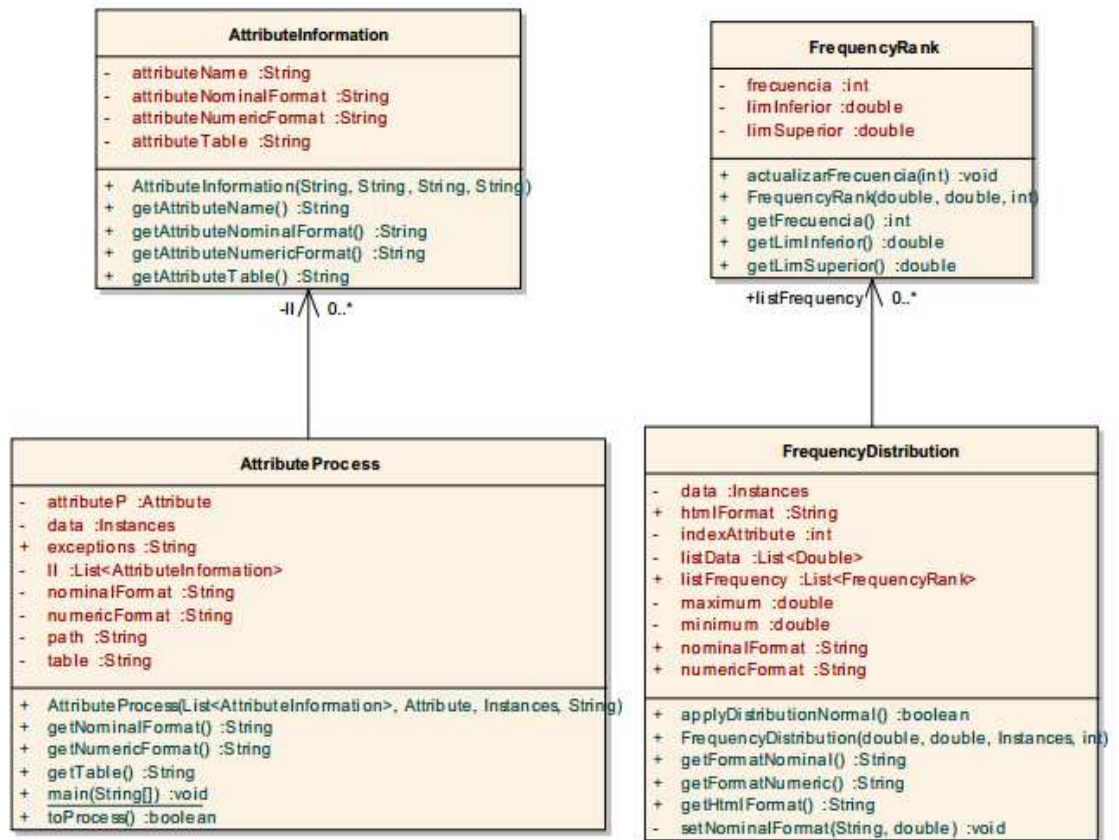
Figura 64. Clases del Paquete File



3.5.2 Clases del Paquete QualityData. En esta sección se presentan los diagramas de clases pertenecientes a este paquete.

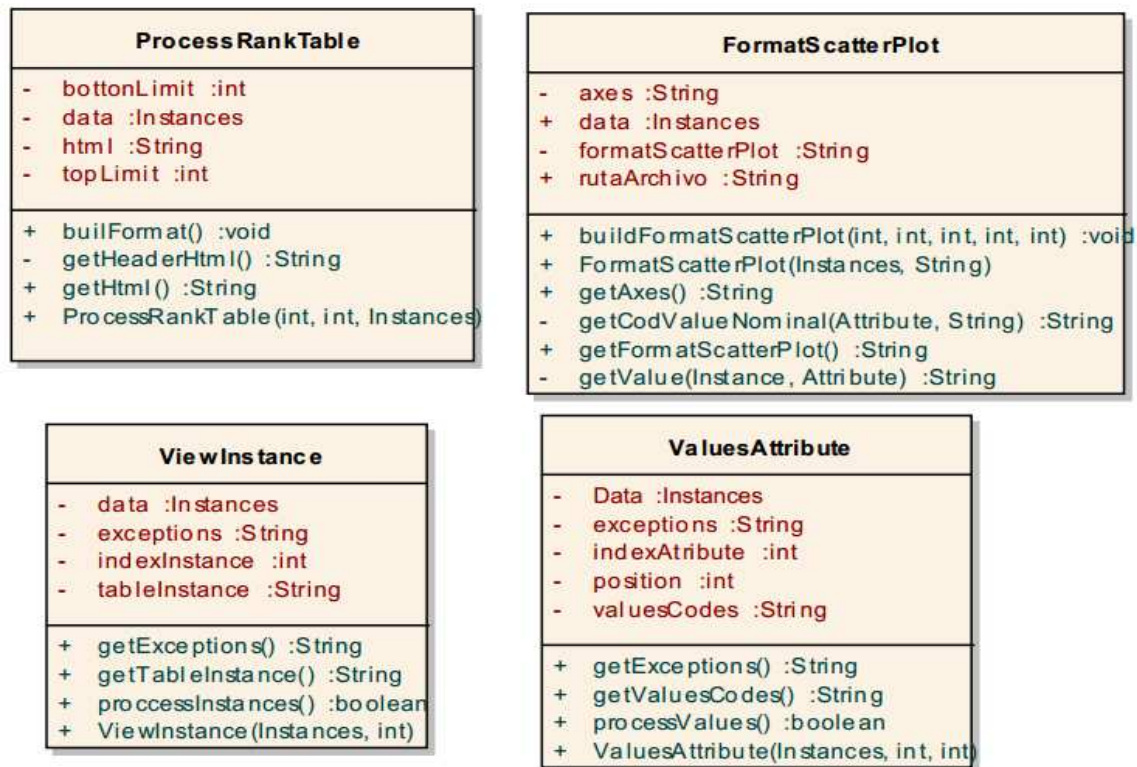
3.5.2.1 Clases de Attribute. En la figura 65, se presenta los diagramas de clase de este paquete.

Figura 65. Clases del paquete Attribute



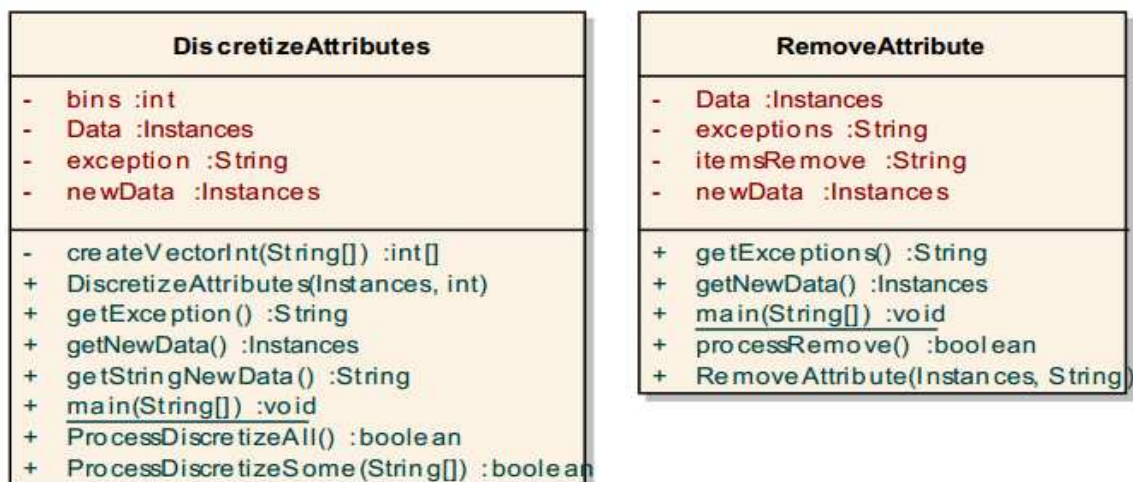
3.5.2.2 Clases de Data. En la figura 66, se presenta los diagramas de clase de este paquete.

Figura 66. Clases del paquete Data



3.5.2.3 Clases de Filters. En la figura 67, se presenta los diagramas de clase de este paquete.

Figura 67. Clases del paquete Filters



3.5.3 Clases del Paquete DataMining. En esta sección se presentan los diagramas de clases que pertenecen a este paquete.

3.5.3.1 Clases para Clustering. En la figura 68, 69 y 70, se presenta los diagramas de clase de este paquete.

Figura 68. Clases uno del paquete Clustering

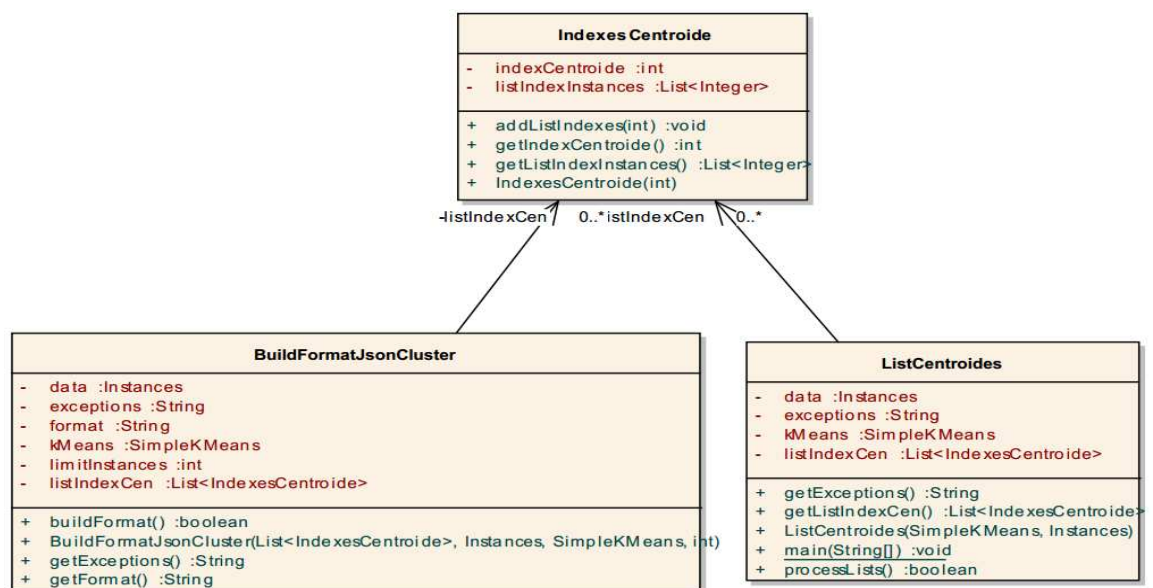


Figura 69. Clases dos del paquete Clustering

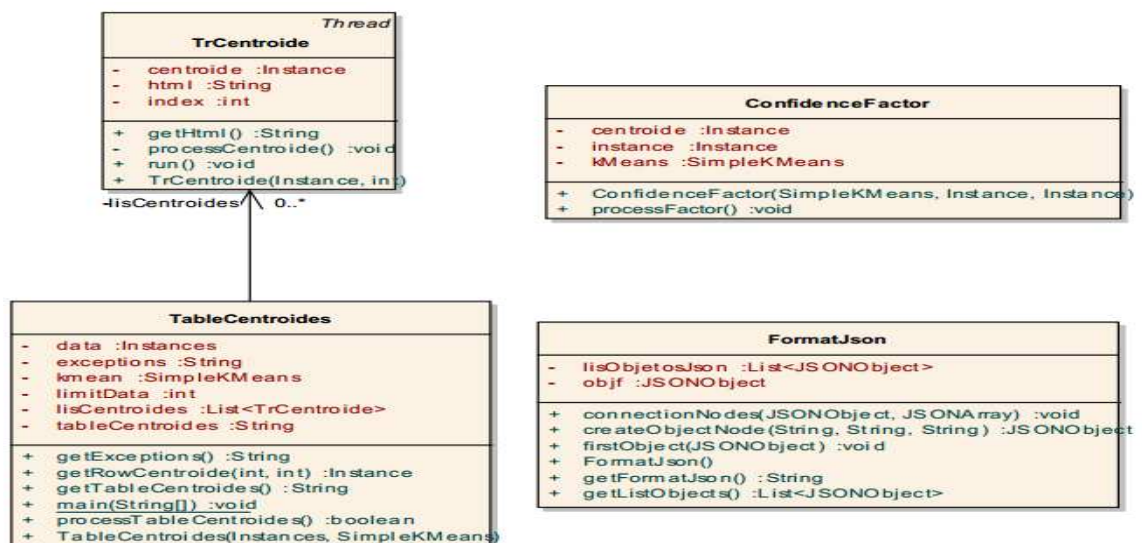
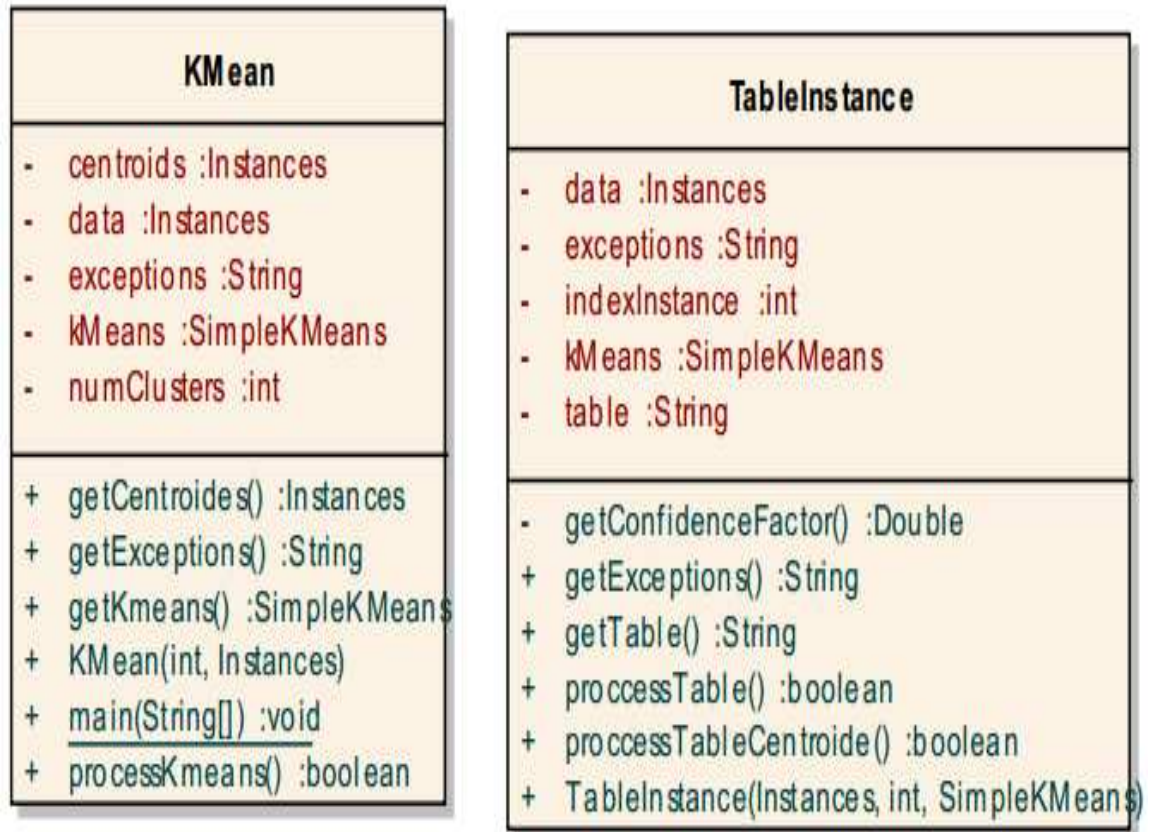
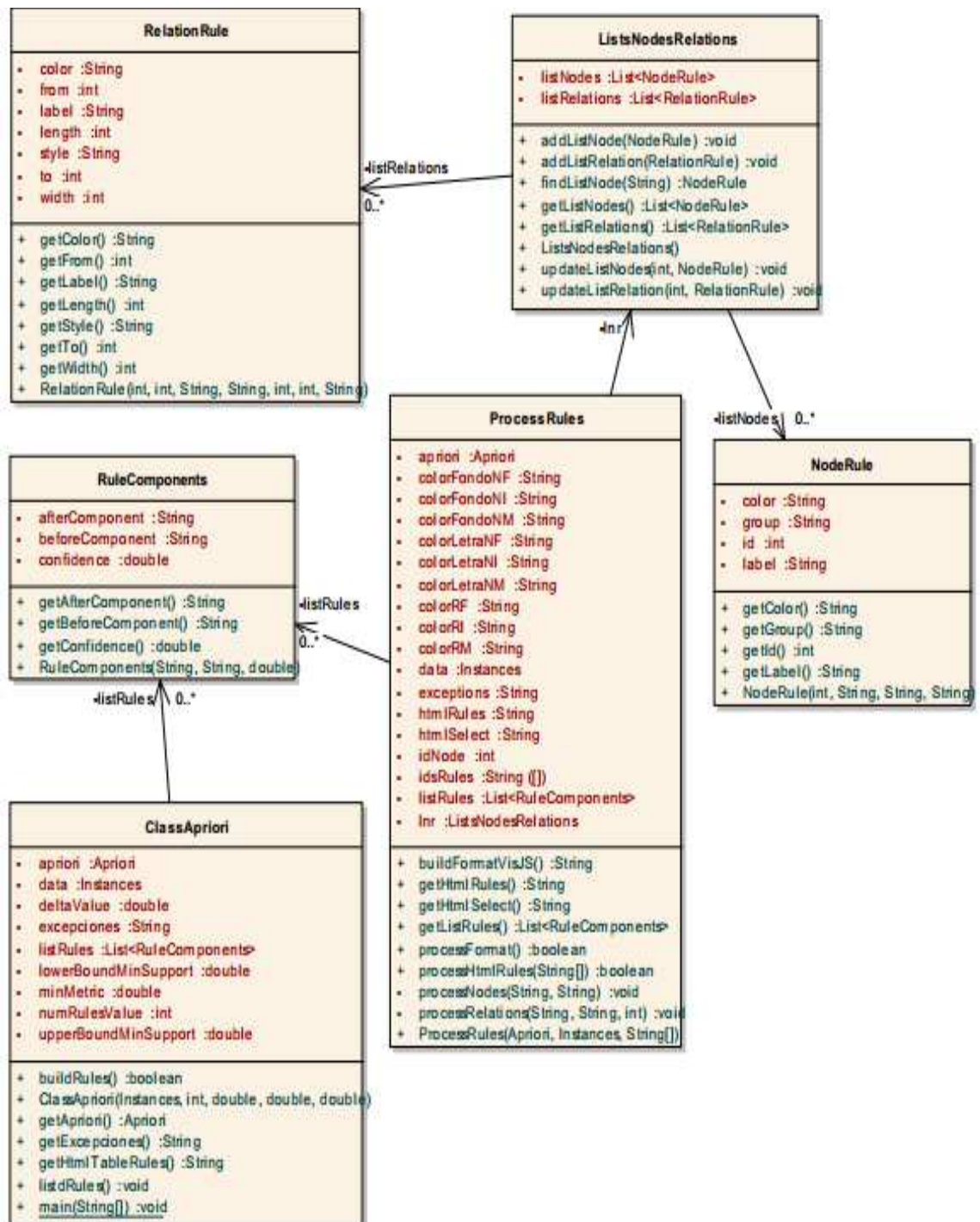


Figura 70. Clases tres del paquete Clustering



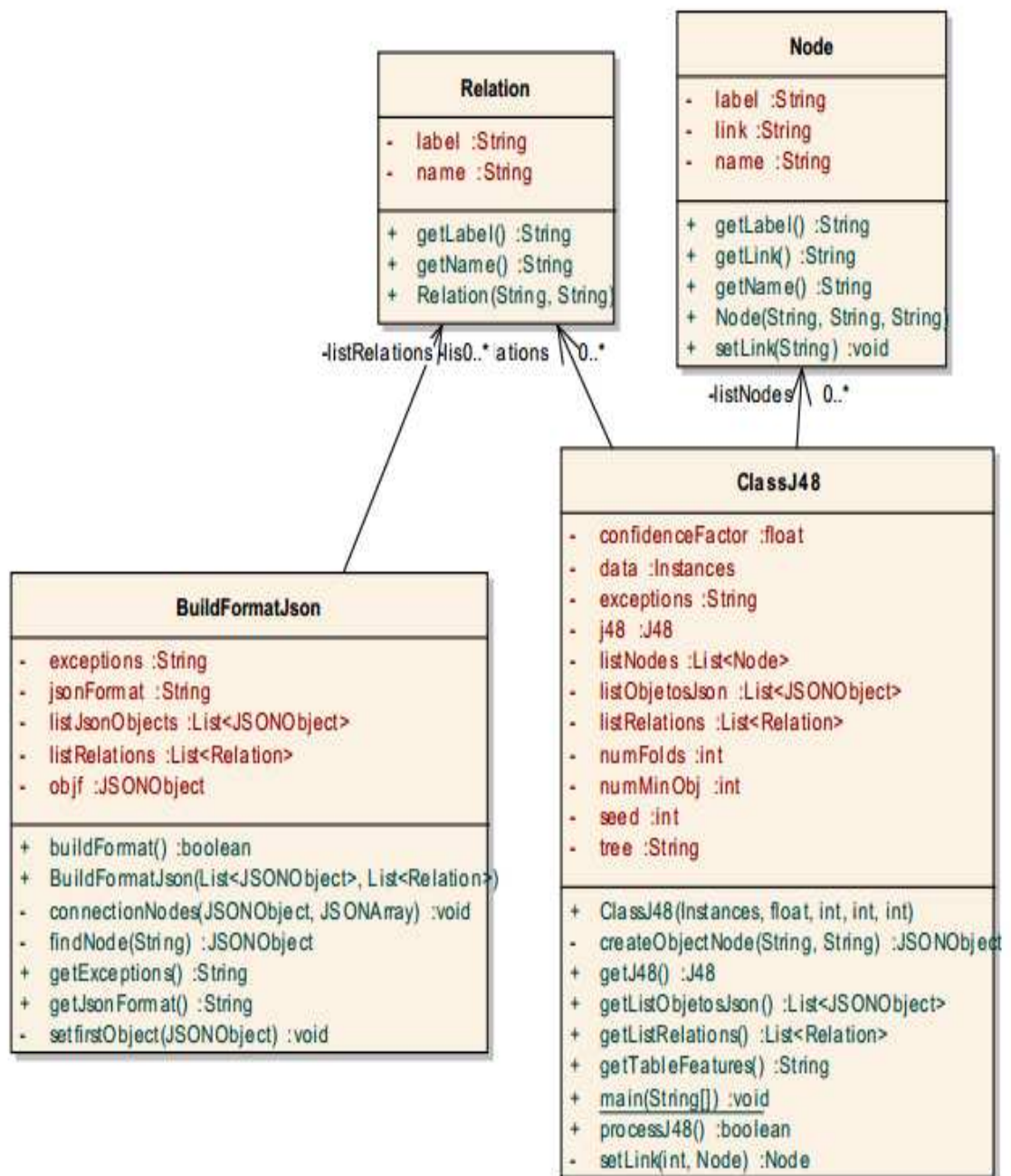
3.5.3.2 Clases para Rules. En la figura 71, se presenta los diagramas de clase de este paquete.

Figura 71. Clases del paquete Rules



3.5.3.3 Clases para Tree. En la figura 72, se presenta los diagramas de clase de este paquete.

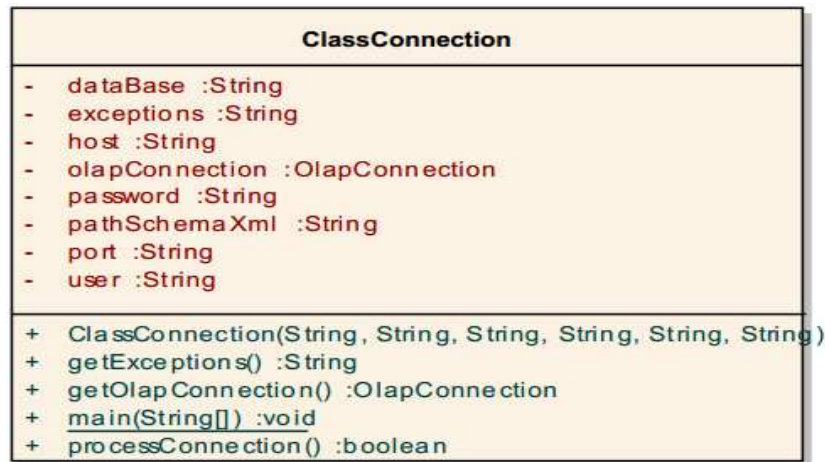
Figura 72. Clases para Tree



3.5.4 Clases del Paquete Dwh. En esta sección se presentan los diagramas de las clases pertenecientes a este paquete.

3.5.4.1 Clases para Connection. En la figura 73, se presenta los diagramas de clase de este paquete.

Figura 73. Clases para Connection

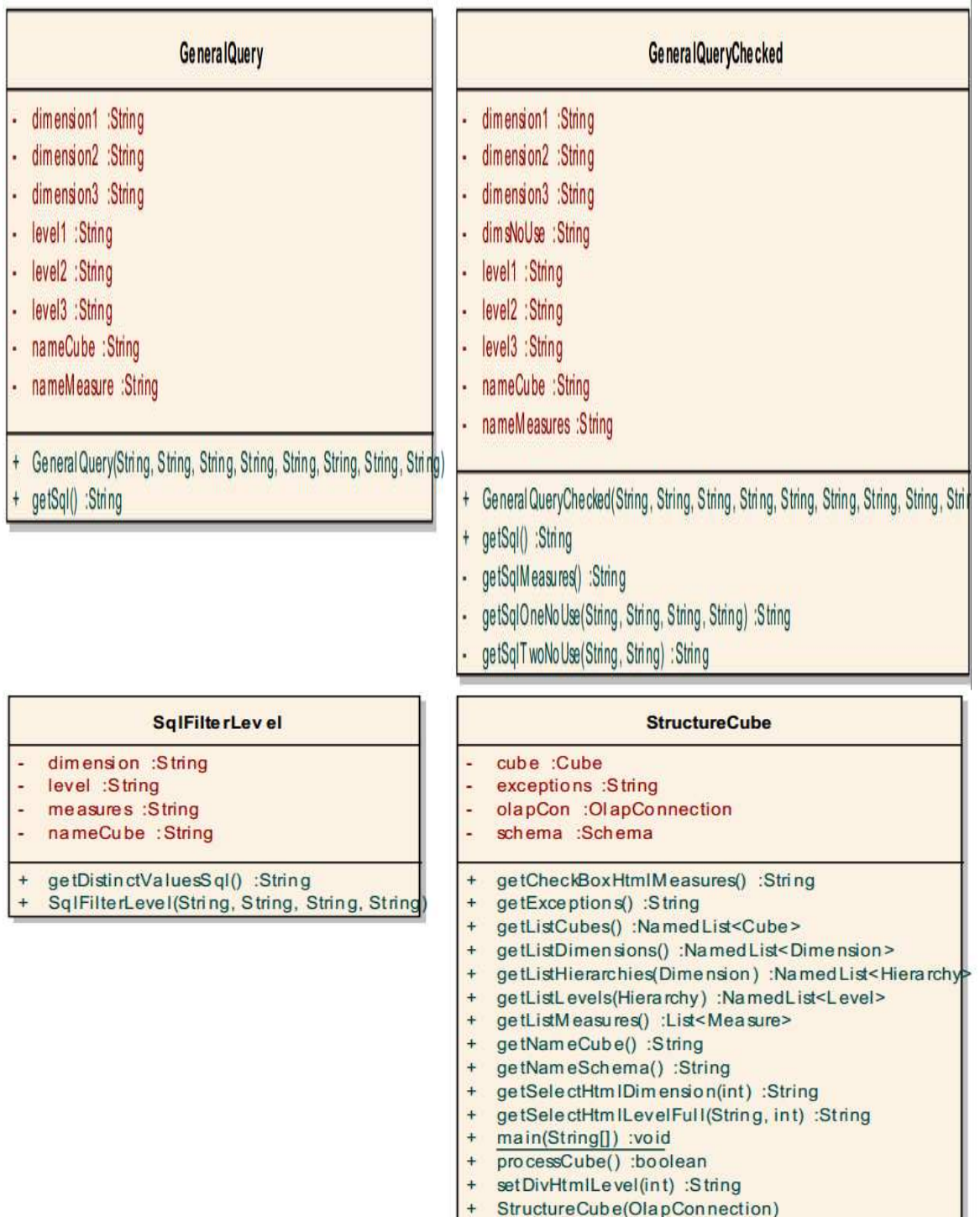


3.5.4.2 Clases para Cube. En la figura 74 y 75, se presenta los diagramas de clase de este paquete.

Figura 74. Clases uno del paquete Cube

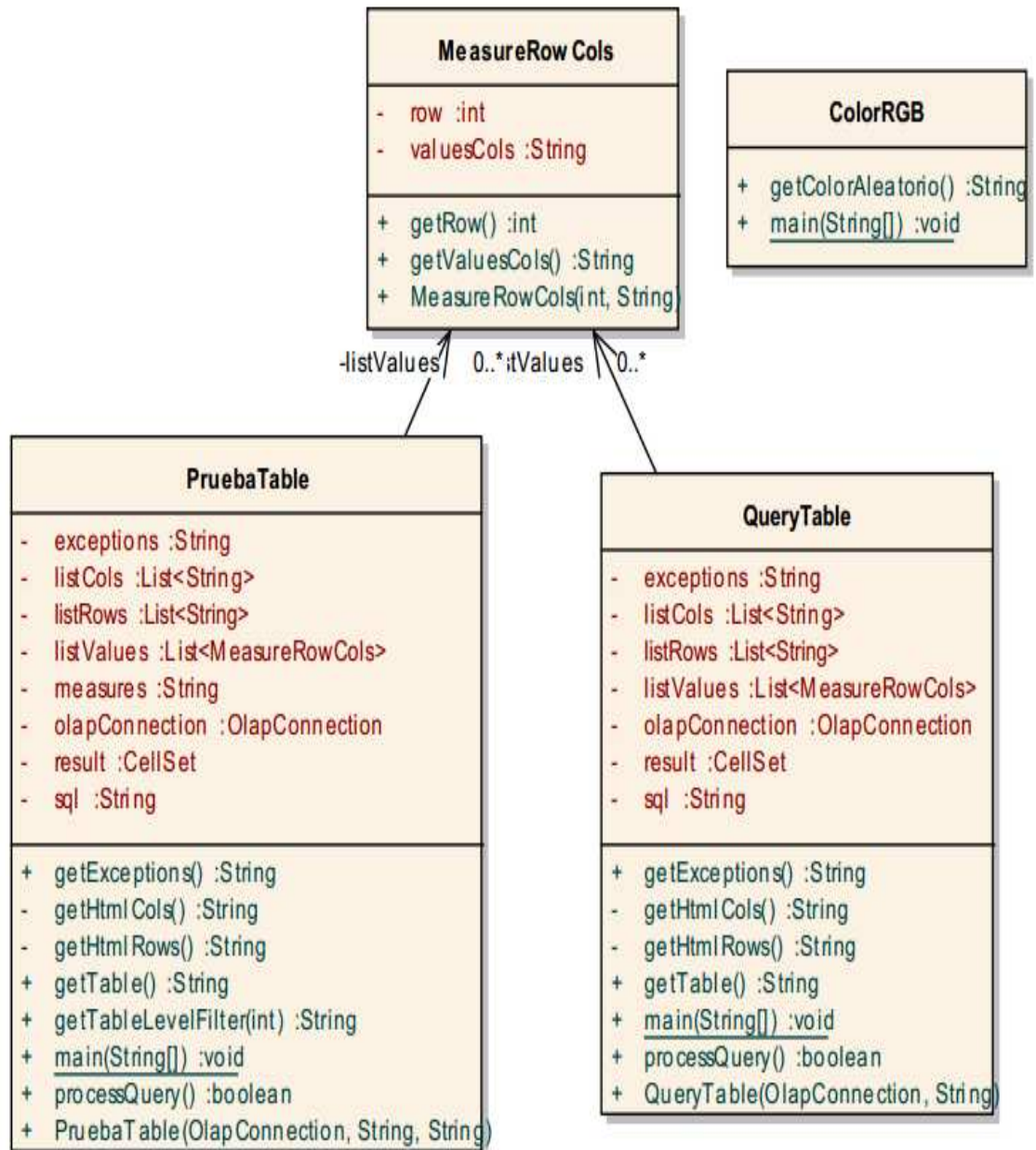


Figura 75. Clases dos del paquete Cube



3.5.4.3 Clases para Table. En la figura 76, se presenta los diagramas de clase de este paquete.

Figura 76. Clases del paquete Table



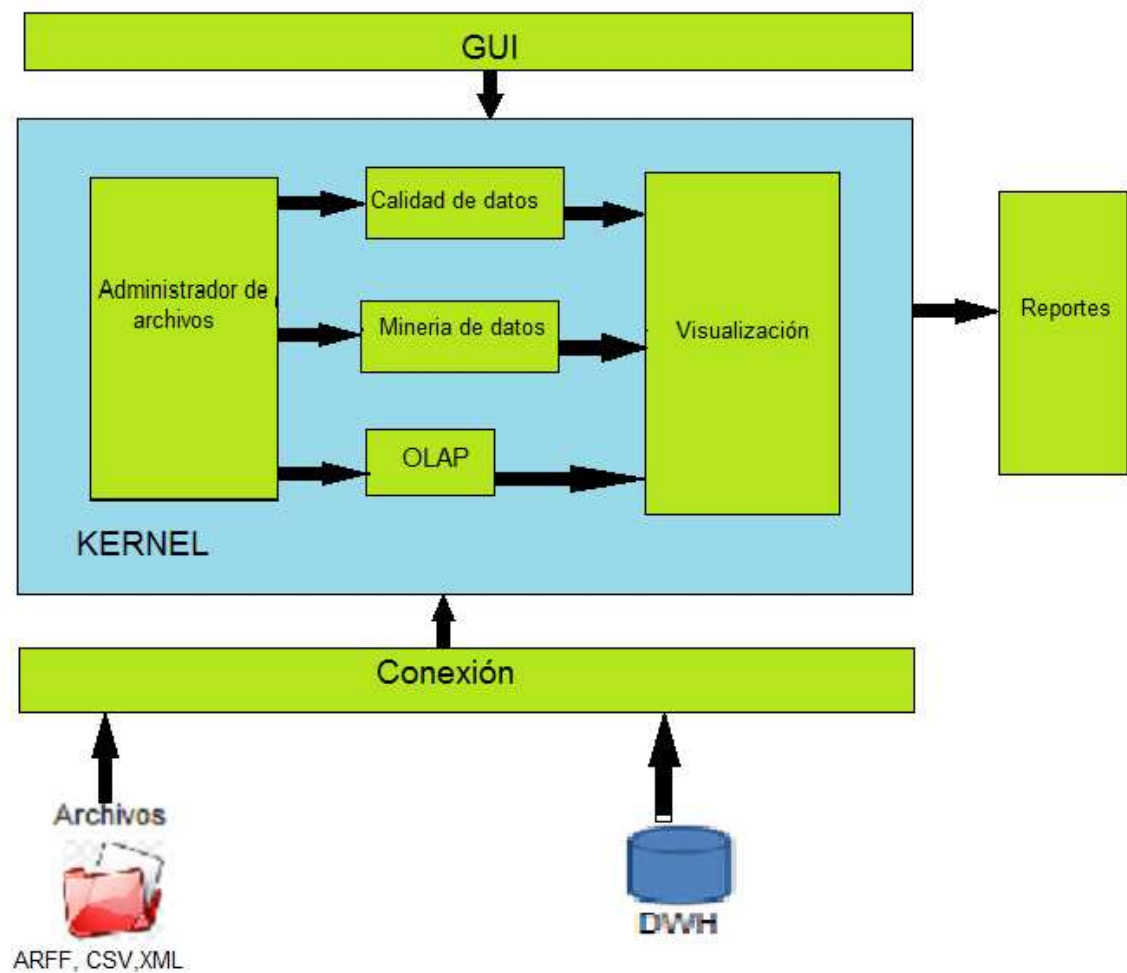
4. ARQUITECTURA DE VISUALBITOOL

En esta sección se presenta la arquitectura general e interna de la herramienta VisualBITool.

4.1 ARQUITECTURA GENERAL

La figura 77, ilustra la arquitectura general de VisualBITool con sus respectivos módulos tales como el Kernel, módulo de conexión y el módulo de interfaz de usuario GUI.

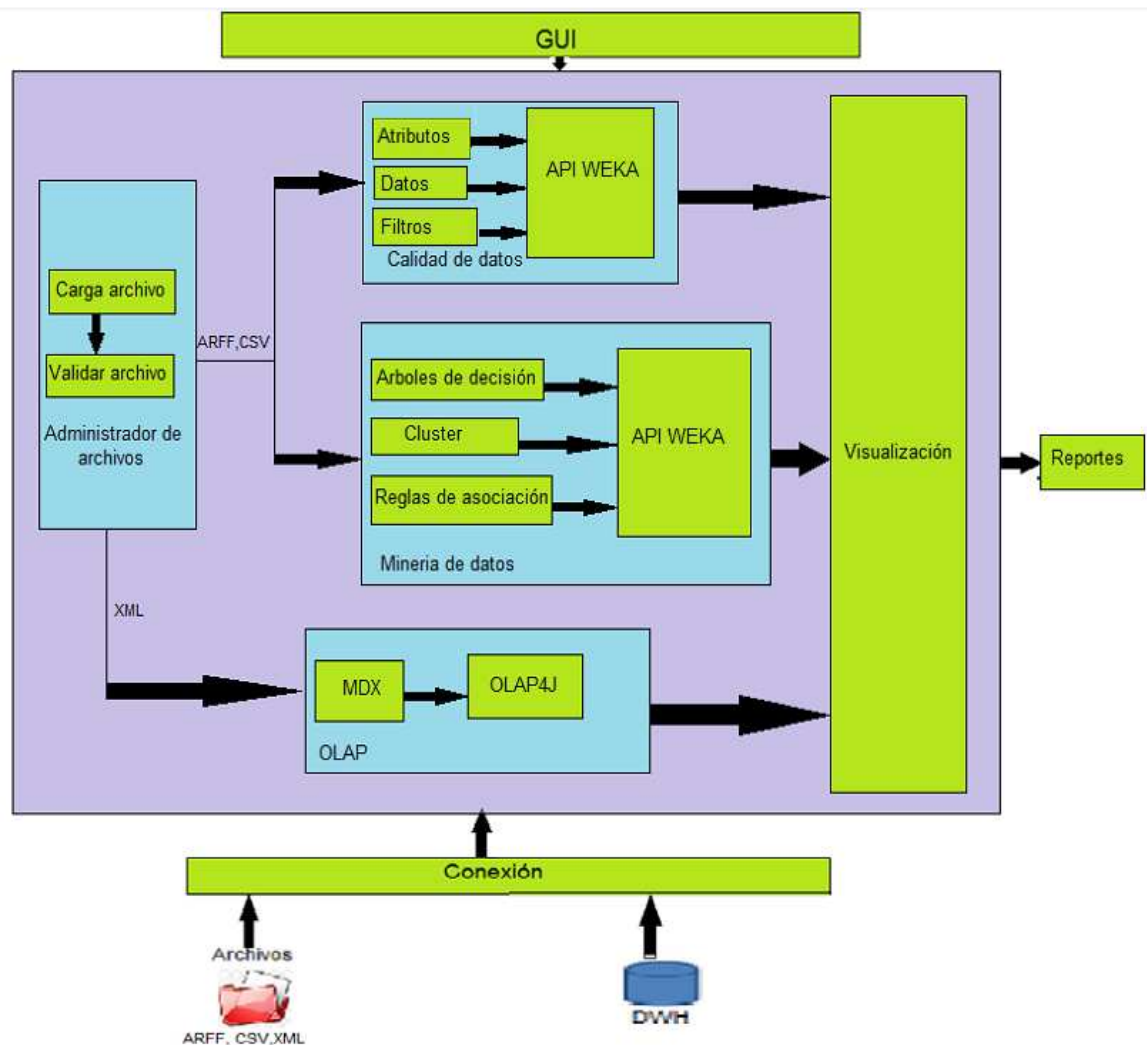
Figura 77. Arquitectura general de VISUALBITOOL



4.2 ARQUITECTURA INTERNA

La figura 78, ilustra la estructura interna de la herramienta VisualBITool, en la cual se puede apreciar los diferentes módulos con sus componentes internos que permiten su funcionamiento.

Figura 78. Arquitectura interna de VISUALBITOOL



A continuación, se da a conocer de manera detallada los diferentes módulos y sub- módulos de la herramienta VisualBITool, con los paquetes y clases que los conforman.

4.2.1 Módulos de la aplicación. Los módulos de software de la herramienta VisualBITool son:

Administrador de archivos. Este módulo se encarga de realizar dos tareas:

- Cargar un archivo, en la cual permite examinar los archivos almacenados en las diferentes carpetas y unidades del equipo, para poder seleccionar el archivo que se desea subir a la herramienta.
- Validar archivo, en la cual se verifica la extensión del archivo (ARFF, CSV, XML) permitidas por la herramienta.

Calidad de datos. Este módulo se encarga de las siguientes tareas:

- Verificar el formato de los datos y la existencia de los datos en un archivo.
- Visualizar los diferentes atributos en forma gráfica que se encuentran en el archivo, con el fin de verificar que los datos se encuentren en buen estado, es decir, que no existan datos nulos o con errores en su estructura, formato y contenido, para su respectivo análisis.

Minería de datos. Este módulo es el encargado de gestionar las técnicas de minería de datos trabajadas en la investigación, por lo tanto se divide en tres tareas:

- Árboles de decisión, se encarga de realizar el análisis extrayendo patrones y relaciones que serán visualizadas mediante un árbol jerárquico.
- Clúster o agrupamiento, se encarga de analizar, generar y representar mediante grupos las características comunes entre las diferentes instancias.
- Reglas de asociación, se encarga de analizar y extraer reglas para luego ser visualizadas.

Conexión. Este módulo se encarga de gestionar el procesamiento de datos multidimensionales. Y realiza las siguientes tareas:

- Efectuar la conexión de la herramienta con PostgreSQL mediante el uso de la biblioteca OLAP4J.
- Muestra la información respectiva al cubo que se está analizando como el nombre, el esquema, las dimensiones y niveles.

- Permite realizar diferentes consultas MDX para poder visualizar los datos requeridos.

Visualización. Este módulo es el encargado de realizar las diferentes gráficas para poder visualizar la información. Además, permite generar reportes en formato (xls) Excel.

GUI. Este módulo es el encargado de manejar la interfaz gráfica de la herramienta VISUALBITOOL.

4.2.2 Descripción de los paquetes por módulo. En esta sección se describen los paquetes pertenecientes a cada módulo de la herramienta VisualBITool.

4.2.2.1 Modulo administrador de archivos. Los paquetes que integran este módulo son:

- **Paquete File.** Implementa todas las clases para realizar un pre procesamiento o tratamiento previo a los archivo de entrada.

El paquete file tiene el siguiente subpaquete:

Paquete Upload. Contiene las clases para subir y validar un archivo.

El paquete Upload contiene las siguientes clases:

- **instancesFile.** Esta clase es la encargada de la gestionar y obtener los atributos e instancias de un archivo.
- **UploadFile.** Esta clase es encargada de subir el archivo y obtener sus propiedades.
- **WriteFile.** Esta clase se encarga de escribir el archivo seleccionado en una ruta específica del servidor Tomcat.

4.2.2.2 Módulo de GUI. Este módulo es administrado mediante archivos JSP y HTML.

4.2.2.3 Calidad de datos. Los paquetes que conforman este módulo son:

Paquete QualityData. Contiene las clases para el análisis de las instancias y los atributos, este se divide en los siguientes paquetes:

- **Paquete Attribute.** Implementa todas las clases para realizar tareas estadísticas con los diferentes atributos y procesar el formato de datos para visualizar los valores de cada atributo.

El paquete Attribute contiene las siguientes clases:

- **AttributeInformation.** Esta clase se encarga de extraer la información pertinente a cada atributo y la muestra en una tabla.
- **AttributeProcess.** Esta clase se encarga de construir el formato de los datos para su respectiva visualización.
- **FrequencyRank.** Esta clase se encarga de analizar la distribución normal de los valores de un atributo en caso de que este sea numérico.
- **FrequencyDistribution.** Esta clase se encarga de obtener datos estadísticos como la media, el valor máximo y mínimo. Además, la desviación estándar de un atributo en caso de que este sea numérico.
- **StaticalValues.** Esta clase se encarga de gestionar los valores estadísticos de atributos numéricos y nominales.
- **Paquete Data.** Implementa todas las clases para visualizar la distribución de las instancias con respecto a sus atributos y procesar el formato de datos para visualizar los valores.

El paquete data contiene las siguientes clases:

- **FormatScatterPlot.** Esta clase se encarga de cruzar los tres atributos seleccionados y de establecer un formato de datos idóneo para la visualización mediante un gráfico de dispersión.
- **ProcessRankTable.** Esta clase permite establecer intervalos de instancias con el fin de que sean visualizadas en una tabla HTML.
- **ValuesAttribute.** Esta clase se encarga de mostrar los valores de un atributo organizados en una tabla HTML.
- **ViewInstance.** Esta clase se encarga de visualizar los valores de una instancia en específico al momento en que se da clic sobre la gráfica que representa una instancia.

- **Paquete Filters.** Implementa las clases para realizar dos tipos de transformación sobre los datos que se desean analizar los cuales son: remover y discretizar atributos.

El paquete data contiene las siguientes clases:

- **RemoveAttribute.** Esta clase se encarga de aplicar el filtro eliminar (remove) en los atributos, permitiendo así quitar un atributo deseado.
- **DiscretizeAttributes.** Esta clase se encarga de aplicar el filtro discretizar (discretize) en los atributos, permitiendo categorizar o discretizar un atributo de tipo numérico.

4.2.2.3 Módulo de minería de datos. Los paquetes que integran este módulo son:

Paquete DataMining. Contiene las clases para el análisis de los datos aplicando las técnicas de minería de datos mediante el uso del api de Weka para java, este se divide en los siguientes paquetes.

- **Paquete Clustering.** Contiene las clases para extraer patrones de agrupamiento y construye el formato de datos para su visualización. El paquete Clustering contiene las siguientes clases:
 - **BuildFormatJsonCluster.** Esta clase se encarga de gestionar los datos para su posterior transformación a formato Json.
 - **ListCentroides.** Esta clase se encarga de listar cada centroide y obtener sus características.
 - **IndexesCentroide.** Esta clase se encarga de gestionar los índices de cada instancia con respecto a un centroide.
 - **TableCentroides.** Esta clase se encarga de construir una tabla en HTML con los valores obtenidos de cada centroide.
 - **TrCentroide.** Esta clase se encarga de generar una fila en HTML con los valores de cada centroide.
 - **FormatJson.** Esta clase se encarga de construir el formato de datos de tipo Json para la visualización.
 - **KMean.** Esta clase se encarga de implementar el algoritmo K-mean del api de Weka.

- **TableInstance.** Esta clase se encarga de crear una tabla HTML con las propiedades de una instancia.
- **Paquete Tree.** Contiene las clases para el análisis de los datos y la generación del árbol de decisión. Además, construye el formato de datos para su visualización.

El paquete Tree contiene las siguientes clases:

- **ClassJ48.** Esta clase se encarga de implementar el algoritmo J48 del api de Weka.
- **Node.** Esta clase se encarga de gestionar las características de cada nodo.
- **Relation.** Esta clase se encarga de gestionar la relación entre nodos.
- **BuildFormatJsonCluster.** Esta clase se encarga de gestionar los datos para su posterior transformación a formato Json.
- **Paquete Rules.** Contiene las clases para analizar los datos y extraer reglas de asociación. Además, construye el formato de datos para su visualización.

El paquete Rules contiene las siguientes clases:

- **ClassApriori.** Esta clase se encarga de implementar el algoritmo Apriori del api de Weka.
- **RuleComponents.** Esta clase se encarga de obtener los antecedentes y consecuentes de cada regla con su respectivo factor de confianza.
- **NodeRule.** Esta clase se encarga de designar características visuales a cada nodo como el color, un identificador y una etiqueta.
- **RelationRule.** Esta clase se encarga de designar características visuales a cada relación como el color, un identificador y una etiqueta.
- **ListaNodeRelations.** Esta clase se encarga de generar listas de los nodos y sus respectivas relaciones.

- **ProcessRules.** Esta clase se encarga de construir el formato de datos para su respectiva visualización.

4.2.2.4 Módulo de conexión. Los paquetes que integran este módulo son:

Paquete DWH. Contiene las clases para la gestión de visualización de los datos multidimensionales mediante un cubo OLAP. Este se divide en los siguientes paquetes.

- **Paquete Cube.** Contiene las clases para la manipulación de la información del cubo y la ejecución de consultas MDX.

El paquete Cube contiene las siguientes clases:

- **FilterQuery.** Esta clase se encarga del manejo de los filtros para realizar las diferentes consultas MDX.
- **FilterQueryChecked.** Esta clase se encarga de gestionar las dimensiones que se involucraran en la consulta a realizar y a estas poder aplicarle filtros.
- **GeneralQuery.** Esta clase se encarga de gestionar las diferentes consultas mdx sin filtros.
- **GenerlQueryChecked.** Esta clase se encarga de gestionar las dimensiones que se involucraran en la consulta a realizar sin utilizar filtros.
- **SqlFilterLevel.** Esta clase se encarga de gestionar los filtros por un nivel en específico.
- **StructureCube.** Esta clase se encarga de gestionar las propiedades del cubo OLAP con el que se está trabajando, en la cual se puede visualizar los niveles y dimensiones.

- **Paquete Conexión.** Contiene las clases para establecer una conexión OLAP con el repositorio en Postgres y el cubo OLAP.

El paquete Conexión contiene las siguientes clases:

- **Classconnection.** Esta clase es la encargada de gestionar la conexión con el repositorio de datos en PostgreSQL y el cubo OLAP.
- **Paquete Table.** Contiene las clases para representar la información organizada en una tabla.

El paquete Table contiene las siguientes clases:

- **ColorRGB.** Esta clase se encarga de generar color de forma aleatoria.
- **MeasureRowCols.** Esta clase se encarga de gestionar la cantidad de filas y columnas que resulta de una consulta.
- **QueryTable.** Esta clase se encarga de generar una tabla HTML con el resultado de una consulta.

5. PRUEBAS Y EVALUACIÓN DE FUNCIONALIDAD

Las pruebas de funcionamiento de la herramienta se desarrollaron utilizando un computador Intel(R) Core(TM) IV i3-2310M , Disco Duro de 298.09 GB, memoria de 3 Gigas , tarjeta de video de 32MB y se utilizó el navegador Chrome.

Para realizar las pruebas de funcionalidad se hizo uso de archivos con extensión ARFF. Archivos extraídos de (ARFF, 2014).

El objetivo de las siguientes pruebas fue determinar la funcionalidad de las diferentes técnicas de visualización implementadas en la herramienta VisualBITool aplicada a la minería de datos en cuanto a calidad de los datos, árboles de decisión, agrupamiento y reglas de asociación.

5.1 PRUEBAS PARA LA TAREA DE ANÁLISIS PREVIO DE ATRIBUTOS

El propósito de esta prueba fue visualizar los valores de cada atributo mediante cuatro clases de gráficos. Para esto se utilizó el siguiente archivo, con características que se muestran en la tabla 25.

Tabla 25. Descripción del archivo Drug.arff

Nombre del archivo	Atributos	Número de Instancias	Tamaño	Descripción
Drug.arff	7	200	9KB	Este archivo contiene información sobre la administración de fármacos a pacientes afectados de rinitis alérgica según distintos parámetros/variables.

Lo primero que se realizó fue seleccionar dos atributos de diferente tipo es decir un nominal y un numérico (ver figura 79). Se inició seleccionando el atributo Sodio 'Na' que es de tipo numérico y automáticamente se generaron cuatro gráficos diferentes, (ver figuras 80, 81, 82 y 83), en este caso, por ser numérico se tomaron como valores de este atributo las frecuencias de cada rango generado.

Figura 79. Lista de atributos



Figura 80. Tabla de valores del atributo Sodio (Na)

X=Na		
Característica	Valor	
Mínimo	0.500169	
Máximo	0.896056	
Media	0.6970952199999999	
Desviación estándar	0.11890733362979117	
0.500169 <=X < 0.54415643	28	
0.54415643 <=X < 0.5881439	23	
0.5881439 <=X < 0.63213134	19	
0.63213134 <=X < 0.6761188	18	
0.6761188 <=X < 0.72010624	10	
0.72010624 <=X < 0.76409364	34	
0.76409364 <=X < 0.8080811	23	
0.8080811 <=X < 0.85206854	24	
0.85206854 <=X < 0.896056	21	
Descargar		

Figura 81. Gráfico de área para el atributo Sodio (Na)

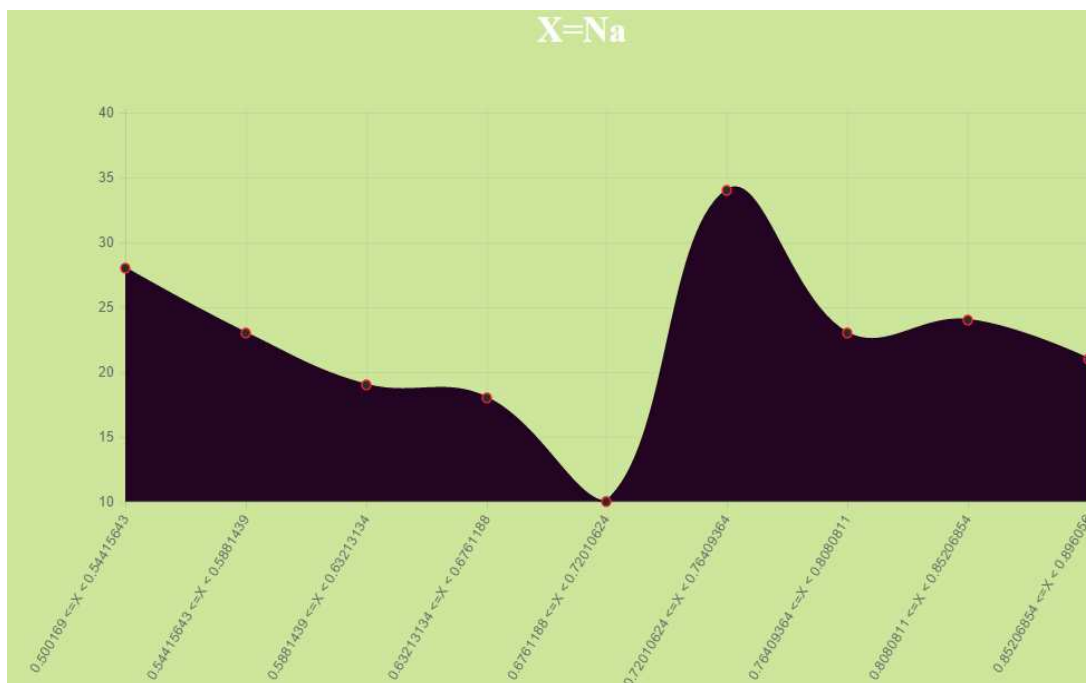


Figura 82. Gráfico de Barras para el atributo Sodio (Na)

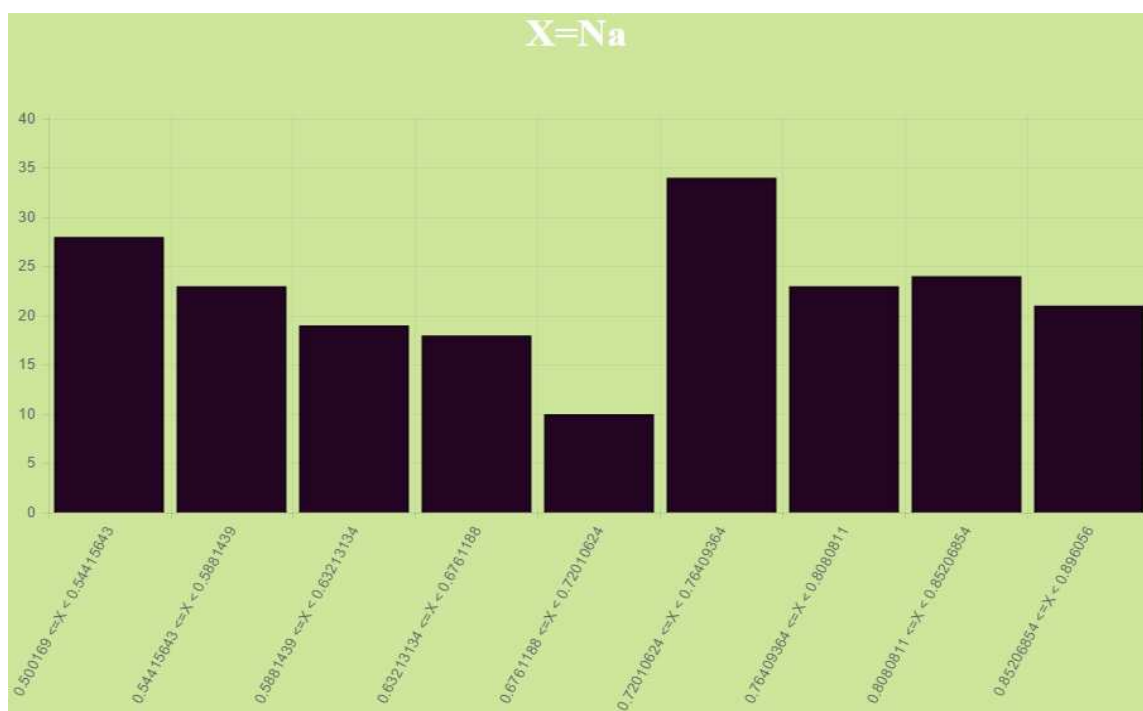
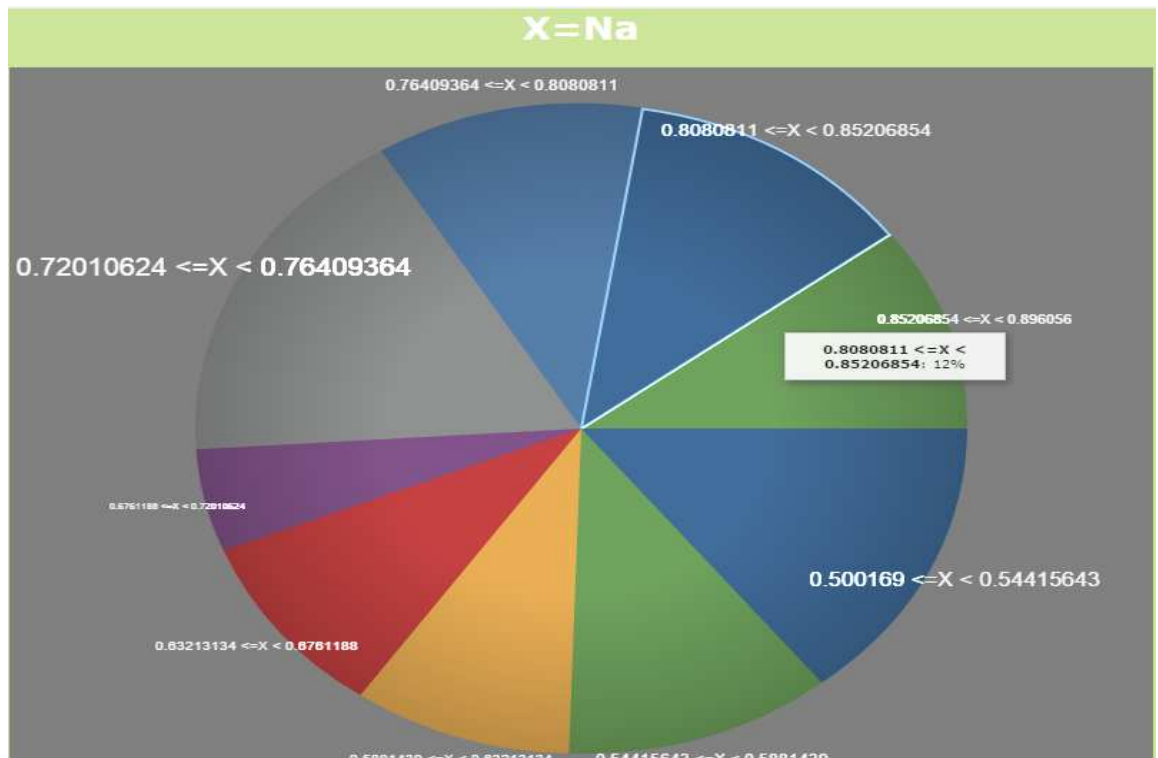


Figura 83. Gráfico de Circulo para el atributo Sodio (Na)



Despues se seleccionó un segundo atributo de prueba, pero ahora de tipo nominal llamado Drug, el cual hace referencia a cinco fármacos posibles: DrugA, DrugB, DrugC, DrugX, DrugY. Se efectuó el mismo procedimiento anterior, solo que para este tipo de atributos solo se generaron tres tipos de gráficos, (ver figuras 84, 85 y 86).

Figura 84. Tabla de valores del atributo fármaco (Drug)

X=Drug

Característica	Valor
drugY	91
drugC	16
drugX	54
drugA	23
drugB	16

Descargar

Figura 85. Gráfico de barras para el atributo fármaco (Drug)

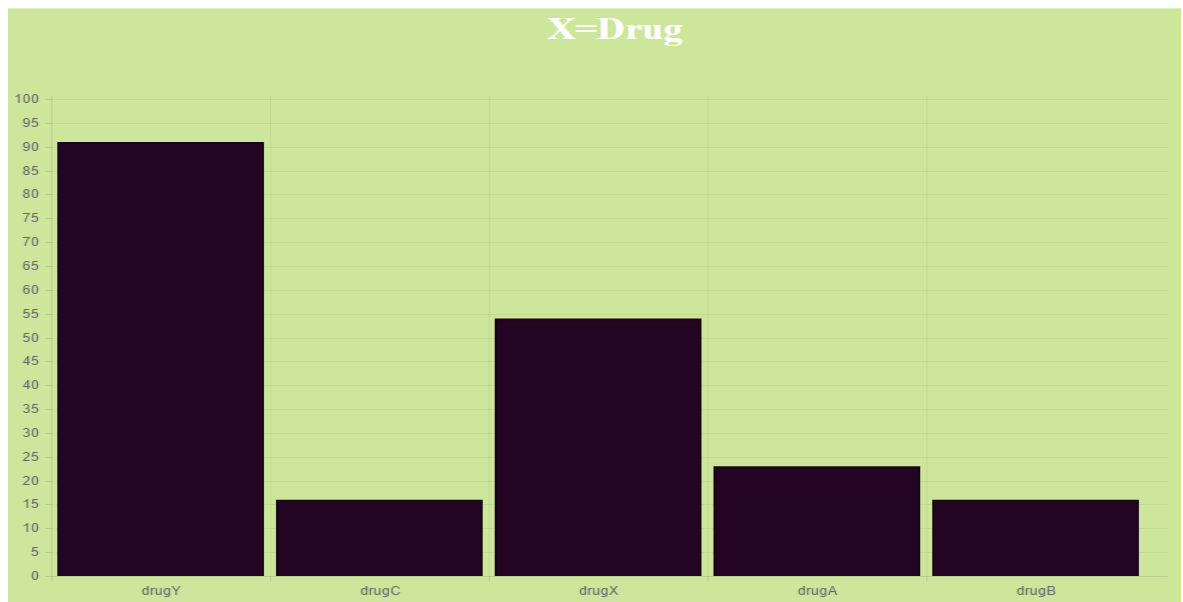
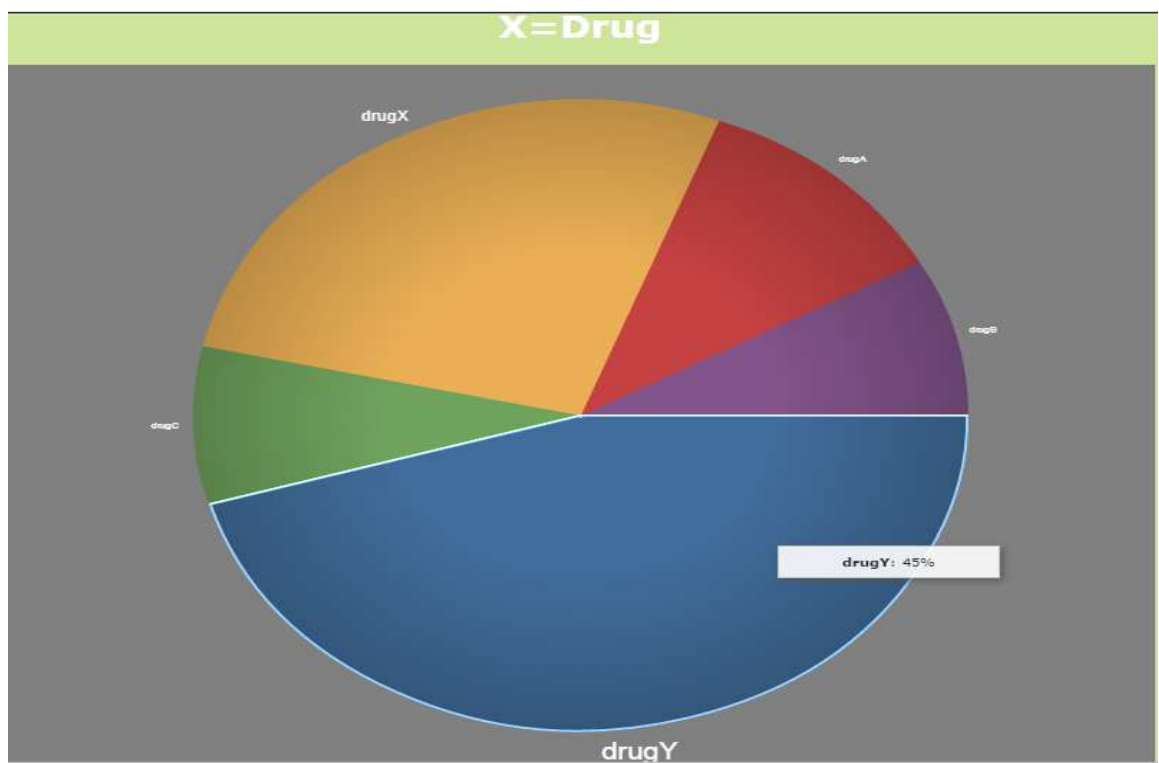


Figura 86. Gráfico de círculo para el atributo fármaco (Drug)



5.1.1 Evaluación de las pruebas de funcionalidad de atributos del módulo calidad de datos. Los cuatro tipos de gráficos anteriormente visualizados, permiten no solo hacer un resumen de las características de cada atributo y entender los datos, sino que también pueden ser de gran ayuda para detectar valores anómalos en los datos.

Cuando un atributo es de tipo numérico, el gráfico de área es una herramienta útil para representar la distribución de sus valores con respecto a una media (Distribución Normal). Esto generalmente ayuda a entender mejor la distribución de los datos y permite determinar, si bajo esa distribución, los datos extremos se consideran estadísticamente anómalos.

La información sobre media, desviación estándar, máximo y mínimo organizada en una tabla proporciona información sobre los atributos numéricos. Estos valores también pueden ser representados en un gráfico de barras, ya que cada barra representa un intervalo y su frecuencia respectiva.

Si se desea visualizar información de un atributo más generalizada, el gráfico de círculo permite representar las frecuencias de cada valor del atributo representadas en porcentajes.

5.2 PRUEBAS PARA LA TAREA DE ANÁLISIS PREVIO DE DATOS

Distribución de datos. El propósito de la prueba fue cruzar tres atributos y visualizar el resultado mediante un gráfico de dispersión. Para esta prueba, se tomó el archivo de datos “*autos.arff*”, cuyas características se encuentran en la tabla 26. Para realizar esta tarea se seleccionó un intervalo de instancias y tres atributos, ver figura 87.

Tabla 26. Descripción del archivo Autos.arff

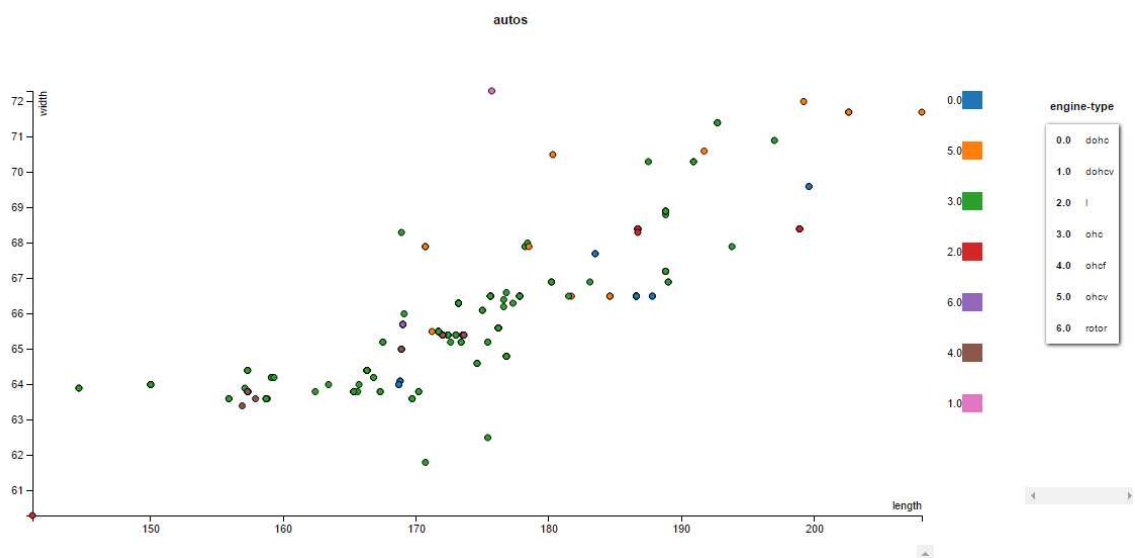
Nombre del archivo	Atributos	Número de Instancias	Tamaño	Descripción
Autos.arff	26	205	30KB	Este archivo de datos contiene información de diferentes clases de autos que se disponen a ser vendidos por una comercializadora.

Figura 87. Parámetros distribución de datos



Antes de generar el gráfico fue necesario procesar los parámetros seleccionados mediante el botón **“Procesar”** y luego presionar sobre la imagen **“Gráfico de dispersión”** se generó un gráfico de dispersión(ver figura 88). En esta técnica de visualización se muestran dos variables numéricas la longitud(length) en el eje de las X, el ancho(width) en el eje de las Y , y una variable categórica representada por un color que representa cada valor de esta variable(**“Engine-type”**). En nuestro caso es el tipo de motor que tiene un determinado auto.

Figura 88. Gráfico de dispersión



Según la figura 88, existen valores erróneos en el atributo “**Engine-type**” (tipo de motor) al mirar en la parte derecha de la gráfica en las etiquetas del atributo “**Engine-type**” que representa el tipo de motor del auto, se observa que en la etiqueta 2.0 se encuentra un símbolo erróneo diferente al tipo de datos de las demás etiquetas, esto quiere decir que las instancias que se encuentran de color rojo, color que se le asigna a la etiqueta 2.0, que poseen valores incorrectos, y al presionar sobre un “**punto rojo**” se pudo verificar los valores de una instancia con un valor que no corresponde al valor de ese atributo, ver figura 89.

Figura 89. Instancia con un valor anómalo o con errores

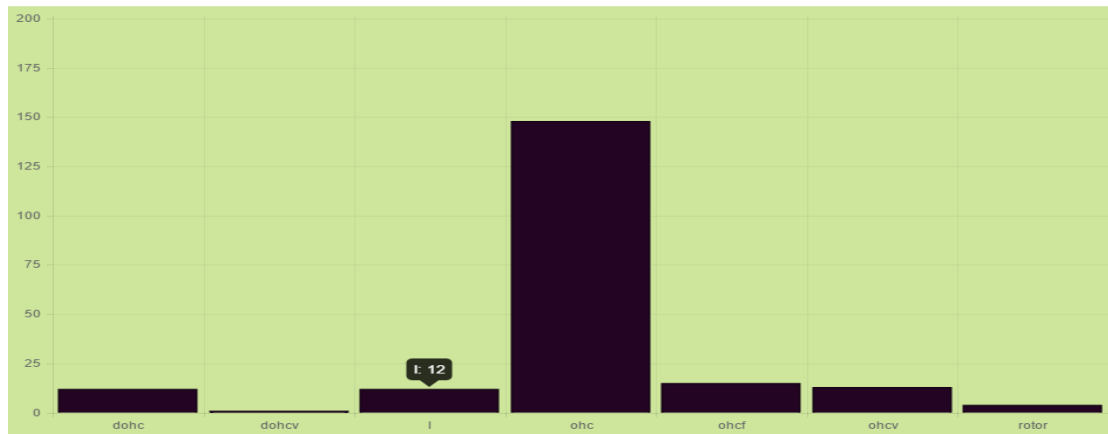


VisualBI Tool (Instancia)	
wheel-base	107.9
length	186.7
width	68.4
height	56.7
curb-weight	3252.0
engine-type	l
num-of-cylinders	four
engine-size	152.0

Salir

En la figura 90, se puede observar que hay 12 instancias con valores incorrectos, datos que representan un problema en cuanto a calidad de datos y para la generación de modelos inteligentes de datos.

Figura 90. Gráfico de barras con datos anómalos o con errores



5.2.1 Evaluación de las pruebas de funcionalidad de los datos (instancias) del módulo calidad de datos. Esta técnica de visualización de datos puede ser muy útil precisamente para detectar comportamientos anómalos, que pueden sugerir, fraudes, fallos, intrusos o comportamientos diferenciados. Con esta herramienta gráfica se pueden encontrar aquellas instancias que no son similares a ninguna(o muy pocas) de las otras instancias. Por ejemplo, cuando en un gráfico de dispersión hay zonas de mayor densidad de instancias quiere decir que estas asumen comportamientos similares y si por el contrario, la densidad es menor, puede ser que algunos de los valores de las instancias contengan valores anómalos o diferentes con respecto a las demás.

También puede ayudar a inferir de forma general respuestas a preguntas que estén relacionadas con los atributos que se está cruzando. Por ejemplo si se desea saber de forma general la respuesta a esta pregunta: ¿De qué tipo de motor (Engine-type) hay más automóviles a la venta?, y pues la respuesta es sencilla, ya que existen muchos puntos de color verde, los cuales corresponden a automóviles con tipo de motor (Engine-type) “**ohc**” siendo así la respuesta a la pregunta planteada.

5.3 PRUEBAS PARA LA TAREA DE FILTROS, APLICADOS SOBRE LOS DATOS.

El propósito de la prueba fue verificar la funcionalidad de la implementación de dos tipos de filtros eliminar y discretizar (**remove y discretize**). Para ello, se utilizó el archivo Drug.arff, cuyas características se pueden ver en la tabla 26.

Filtro para eliminar atributos. En esta parte de la prueba, el objetivo fue verificar la funcionalidad de la selección de atributos a eliminar mediante el filtro remove (remove) y luego observar el archivo nuevo que se generó, el cual contiene los atributos idóneos sin los atributos eliminados, para posteriormente ser analizados con alguna técnica de minería de datos, según sea el caso, ver figura 91.

Figura 91. Selección de atributos



Resultado de la eliminación de atributos. Al presionar la imagen que representa una flecha, como se observa en la figura 91, se obtuvo como resultado el nuevo archivo, ver figura 92.

Figura 92. Archivo de datos nuevo

```

1 @relation DRUG1n
2
3 @attribute Age numeric
4 @attribute Sex {F,M}
5 @attribute BP {HIGH,LOW,NORMAL}
6 @attribute Cholesterol {HIGH,NORMAL}
7 @attribute Drug {drugY,drugC,drugX,drugA,drugB}
8
9 @data
10 23,F,HIGH,HIGH,drugY
11 47,M,LOW,HIGH,drugC
12 47,M,LOW,HIGH,drugC
13 28,F,NORMAL,HIGH,drugX
14 61,F,LOW,HIGH,drugY
15 22,F,NORMAL,HIGH,drugX
16 49,F,NORMAL,HIGH,drugY
17 41,M,LOW,HIGH,drugC
18 60,M,NORMAL,HIGH,drugY
19 43,M,LOW,NORMAL,drugY
20 47,F,LOW,HIGH,drugC

```

Discretización de atributos: el propósito de la prueba fue evaluar el funcionamiento de la tarea de discretización de VisualBITool, para realizarla se obtuvo los datos de tipo numérico, se seleccionaron y luego se discretizaron, creando así un nuevo archivo con datos discretizados. Esta prueba se aplicó sólo a atributos de tipo numérico, en este caso se seleccionó dos atributos edad (Age) y Sodio (Na), como se ve en la figura 93.

Figura 93. Interfaz para la discretización de atributos



Discretización de atributos Cerrar

Num.Bins

5

<input checked="" type="checkbox"/>	Age
<input checked="" type="checkbox"/>	Na
<input type="checkbox"/>	K

Todo Ninguno Discretizar

El Número de Bins, ver figura 93, es el número de intervalos con el cual se desea discretizar un atributo. Al hacer clic sobre el botón **“Discretizar”** y después de mostrar un mensaje de éxito, se prosiguió a visualizar el resultado, haciendo clic sobre la imagen de descarga, como se ve en la figura 93, se obtuvo el siguiente resultado, ver figura 94.

Figura 94. Resultado Discretización edad (Age) y sodio (Na)

```

1 @relation 'DRUG1n-weka.filters.unsupervised.attribute.Discretize-unset-class-temporarily-B5-M-1.0-R1,5'
2
3 @attribute Age {'-inf-26.8','26.8-38.6','38.6-50.4','50.4-62.2','62.2-inf'}
4 @attribute Sex {F,M}
5 @attribute BP {HIGH,LOW,NORMAL}
6 @attribute Cholesterol {HIGH,NORMAL}
7 @attribute Na {'-inf-0.579346','0.579346-0.658524','0.658524-0.737701','0.737701-0.816879','0.816879-inf'}
8 @attribute K numeric
9 @attribute Drug {drugY,drugC,drugX,drugA,drugB}
10
11 @data
12 '-inf-26.8',F,HIGH,HIGH,'0.737701-0.816879',0.031258,drugY
13 '38.6-50.4',M,LOW,HIGH,'0.737701-0.816879',0.056468,drugC
14 '38.6-50.4',M,LOW,HIGH,'0.658524-0.737701',0.068944,drugC
15 '26.8-38.6',F,NORMAL,HIGH,'-inf-0.579346',0.072289,drugX
16 '50.4-62.2',F,LOW,HIGH,'-inf-0.579346',0.030998,drugY
17 '-inf-26.8',F,NORMAL,HIGH,'0.658524-0.737701',0.078647,drugX
18 '38.6-50.4',F,NORMAL,HIGH,'0.737701-0.816879',0.048518,drugY
19 '38.6-50.4',M,LOW,HIGH,'0.737701-0.816879',0.069461,drugC
20 '50.4-62.2',M,NORMAL,HIGH,'0.737701-0.816879',0.05123,drugY
21 '38.6-50.4',M,LOW,NORMAL,'-inf-0.579346',0.027164,drugY
22 '38.6-50.4',F,LOW,HIGH,'0.816879-inf',0.076147,drugC
23 '26.8-38.6',F,HIGH,NORMAL,'0.658524-0.737701',0.034782,drugY
24 '38.6-50.4',M,LOW,HIGH,'0.579346-0.658524',0.040746,drugY
25 '62.2-inf',F,LOW,HIGH,'0.737701-0.816879',0.037851,drugY
26 '38.6-50.4',F,NORMAL,HIGH,'0.816879-inf',0.065166,drugX
27 '-inf-26.8',F,HIGH,NORMAL,'0.816879-inf',0.053742,drugY
28 '62.2-inf',M,LOW,NORMAL,'0.816879-inf',0.074111,drugX
29 '38.6-50.4',M,HIGH,HIGH,'0.579346-0.658524',0.046979,drugA
30 '-inf-26.8',M,LOW,HIGH,'-inf-0.579346',0.076609,drugC
31 '26.8-38.6',F,HIGH,NORMAL,'0.579346-0.658524',0.024773,drugY
32 '50.4-62.2',M,LOW,NORMAL,'-inf-0.579346',0.028061,drugY

```

5.3.1 Evaluación de las pruebas de funcionalidad para el módulo de filtro.

Gracias a esta tarea se puede realizar algunas transformaciones sobre los valores de los atributos como eliminarlos o discretizarlos, estas transformaciones son de gran utilidad como técnicas de pre-procesamiento de los datos que serán utilizados para un posterior análisis.

El filtro de eliminación permite seleccionar de forma coordinada los datos más relevantes del proceso y el filtro de discretización facilita la transformación de los datos numéricos a nominales, ya que sistemas de aprendizaje automático como las reglas de asociación basadas principalmente en el algoritmo de Apriori, requiere atributos de tipo simbólico o nominales para la extracción de reglas y patrones de conocimiento.

5.4 PRUEBAS PARA LA TAREA DE ÁRBOLES DE DECISIÓN

El objetivo de la prueba fue predecir el tipo de fármaco (Drug) que se debe administrar a un paciente afectado por rinitis alérgica según distintos parámetros o variables. Se utilizó el archivo “**Drug.arff**”, cuyas características se ven en la tabla 27.

Tabla 27. Descripción del archivo Drug.arff

Nombre del archivo.	Número de registros	Tamaño	Descripción
Drug.arff	200	9KB	Este archivo contiene información sobre la administración de fármacos a pacientes afectados de rinitis alérgica según distintos parámetros/variables.

Las variables que se recogen en los historiales clínicos de cada paciente son:

Age: edad

Sex: sexo

BP (Blood Pressure): Tension sanguínea

Cholesterol: nivel de colesterol

Na: Nivel de sodio en la sangre

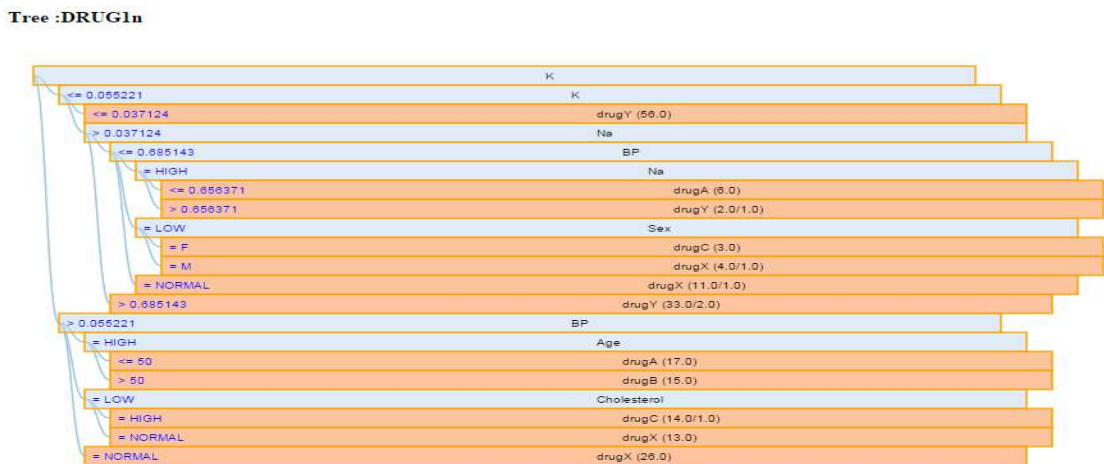
K: Nivel de potasio en la sangre

Hay cinco fármacos posibles, DrugA, DrugB, DrugC, DrugX, DrugY. Se han recogido los datos para del medicamento idóneo para muchos pacientes en cuatro hospitales. Se pretende para nuevos pacientes, determinar el mejor medicamento a probar.

Se inició con la Tabla de parámetros para la construcción de un árbol de decisión, en la figura 95, se muestra una tabla en la cual se pueden editar los parámetros básicos para la construcción del árbol de decisión. Además, se muestra un campo denominado “Atributo clase” con el cual se establece el atributo sobre el cual se determina una decisión. Para más información acerca de los parámetros, revisar el manual de usuario de VisualBITool.

Al seleccionar la gráfica de árbol de sangría se pudo observar las diferentes ramificaciones que tiene un nodo para la toma de decisiones como se muestra en la figura 97.

Figura 97. Árbol de sangría



El gráfico de sangría que se observa en la figura 97, es otra forma de representar el árbol de decisión en el cual los diferentes niveles se representan en rectángulos en un orden jerárquico. Además, los colores permiten diferenciar las ramas y las hojas. Las línea de color azul son las particiones, los rectángulos de color azul son las ramas y los de color naranja son las hojas.

5.4.1 Evaluación de las pruebas de funcionalidad de árboles de decisión. El árbol jerárquico horizontal es una forma de representar un árbol de decisión el cual permite seguir una trayectoria que describa una decisión final de una forma organizada. Además, permite observar la estructura de una regla en la cual los nodos de cada rama que compone una trayectoria conforman el antecedente y la hoja final es el consecuente. El gráfico de sangría facilita observa la estructura de una regla, ya que mediante colores permite diferenciar las ramas de color azul que representan el antecedente y las hojas de color naranja que representan el consecuente, y así brindando una forma más comprensible para tomar una decisión final.

5.5 PRUEBAS PARA LA TAREA DE AGRUPAMIENTO (CLUSTERING)

El propósito de la prueba fue tratar de encontrar grupos de personas a quienes se les suministra un tipo de fármaco (Drug) que se debe administrar a un paciente afectado por rinitis alérgica según distintos parámetros o variables. Se utilizó el archivo “**Drug.arff**”, cuyas características se describen en la tabla 28.

Tabla 28. Descripción del archivo Drug.arff

Nombre del archivo.	Número de registros	Tamaño	Descripción
Drug.arff	200	9KB	Este archivo contiene información sobre la administración de fármacos a pacientes afectados de rinitis alérgica según distintos parámetros/variables.

En la tabla de la figura 98, se pueden editar los parámetros para realizar la gráfica de los diferentes grupos con sus respectivas instancias.

Figura 98. Parámetros para Cluster horizontal

Limite inferior	0
Limite superior	200
Num.Instancias/Cluster	30
<button>Procesar Instancias</button>	

La tabla de la figura 98, presenta tres parámetros, el límite inferior y superior, los cuales permiten establecer un rango de instancias a visualizar en este caso de 0 a 200 instancias. El número de instancias es el número máximo de instancias por cada grupo que serán visualizadas.

Al seleccionar “**ver tabla de grupos**”, se muestra una Tabla que permite visualizar los valores de cada Centroide. En la tabla de la figura 99 se muestra los diferentes grupos (Clusters) con los valores generales de cada centroide.

Figura 99. Tabla con los diferentes grupos (Clusters)

Tabla de grupos

5 records

Search:

	Age	Sex	BP	Cholesterol	Na	K	Drug
Grupo0	43.47727272727273	F	LOW	NORMAL	0.6690185681818182	0.06249104545454545	drugX
Grupo1	34.86363636363637	M	HIGH	HIGH	0.6469599545454545	0.06113722727272726	drugA
Grupo2	48.48888888888889	M	LOW	HIGH	0.7519618	0.04422724444444443	drugY
Grupo3	40.59574468085106	F	HIGH	NORMAL	0.7475557446808512	0.03825229787234042	drugY
Grupo4	42.51851851851852	M	NORMAL	NORMAL	0.6074721851851851	0.04388214814814816	drugY

Showing 1 to 5 of 6 entries

<

1

2

>

Descargar

Al hacer clic en el botón de descargar, que se encuentra en la parte inferior de la figura 99, se creó un archivo Excel con la información presentada en la tabla, ver figura 100.

Figura 100. Archivo Excel con los diferentes grupos (Clusters)

tableClusters - Excel

ARCHIVO INICIO INSERTAR DISEÑO DE PÁGINA FÓRMULAS DATOS REVISAR VISTA Nitro PDF Prof

Pegar

Calibri 10 A⁺ A⁻

N K S

Ajustar texto

Combinar y centrar

General

\$ % 000 0.00 0.00

Portapapeles Fuente Alineación Número

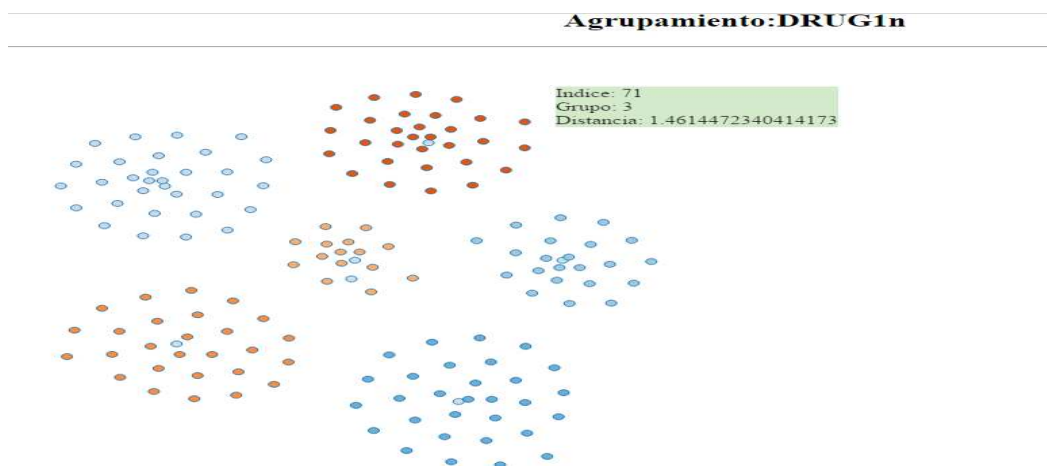
A1

X ✓ fx

	A	B	C	D	E	F	G	H	I
1		Age	Sex	BP	Cholesterol	Na	K	Drug	
2	Grupo0	4.347.727.272.727.270	F	LOW	NORMAL	0.6690185681818182	0.06249104545454545	drugX	
3	Grupo1	3.486.363.636.363.630	M	HIGH	HIGH	0.6469599545454545	0.06113722727272726	drugA	
4	Grupo2	4.848.888.888.888.880	M	LOW	HIGH	0.7519618	0.04422724444444443	drugY	
5	Grupo3	4.059.574.468.085.100	F	HIGH	NORMAL	0.7475557446808512	0.03825229787234042	drugY	
6	Grupo4	4.251.851.851.851.850	M	NORMAL	NORMAL	0.6074721851851851	0.04388214814814816	drugY	
7	Grupo5	63.0	M	HIGH	HIGH	0.6915971999999999	0.0644836	drugB	
8									
9									

Al seleccionar la gráfica de burbujas se pudo observar como las instancias se agruparon respecto a un centroide, como se ve en la figura 101.

Figura 101. Gráfica de grupos (Clusters)



Al seleccionar un grupo (Cluster) se muestra una tabla que permite visualizar los valores de un grupo (Cluster) determinado. En la figura 102, se puede seleccionar un grupo (Cluster) en específico y visualizar sus valores en una tabla.

Figura 102. Valores de un grupo (Cluster)

Grupos

Cluster0 ▼

Age	43.47727272727273
Sex	F
BP	LOW
Cholesterol	NORMAL
Na	0.6690185681818182
K	0.06249104545454545
Drug	drugX

Al hacer clic sobre un punto de un determinado grupo, como se observa en la figura 101. Se genera una tabla que permite comparar los valores de una instancia con respecto a su grupo y su porcentaje de confianza, el cual permite determinar el nivel de similitud de una instancia con respecto a su centro, ver figura 103.

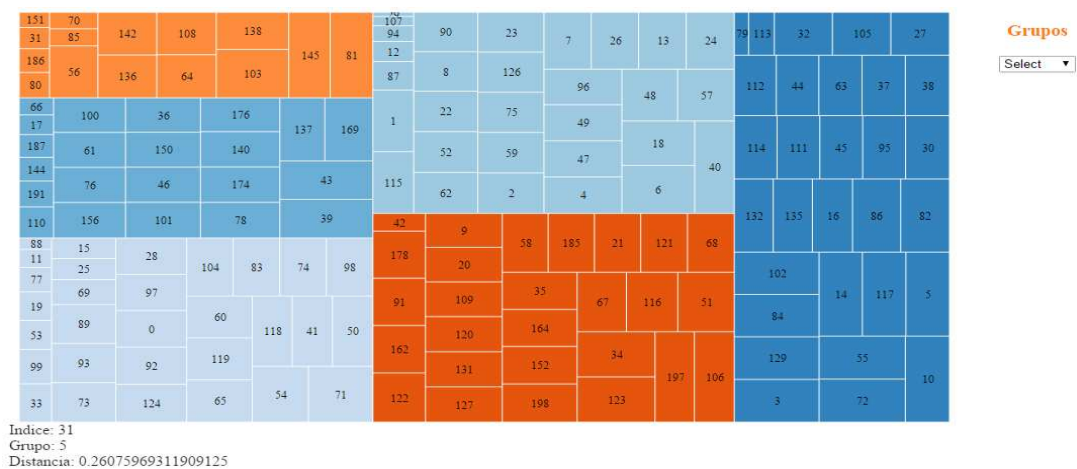
Figura 103. Valores de una instancia con su respectivo valor de confianza

VisualBiTool(Instancia)		
Cholesterol	NORMAL	HIGH
Na	0.749717	0.6469599545454545
K	0.06678	0.06113722727272726
Drug	drugA	drugA
Factor de Confianza	42.0%	

Salir

Al seleccionar la gráfica de mapa de árbol se puede observar como las instancias se agruparon respecto a un centroide y los diferentes grupos se ven en cada rectángulo diferenciados por colores, como se muestra en la figura 104.

Figura 104. Mapa de árbol para grupo (Cluster)



5.5.1 Evaluación de las pruebas de funcionalidad para agrupamiento. El gráfico de burbujas permite observar como las instancias se agrupan de acuerdo a unas características a un centroide. Las instancias se organizan teniendo en cuenta una distancia euclidiana en forma de espiral, la instancia que esté más cerca al centroide es de menor longitud.

Los colores es una forma de diferenciar las instancias que pertenecen a un determinado grupo y se puede observar en los dos tipos de gráficos, lo cual facilita el análisis y visualización de cada grupo.

El gráfico de mapa de árbol permite observar los diferentes grupos mediante rectángulos y la utilización de diferentes colores. Además, permite enfocar la imagen en un solo grupo permitiendo así el análisis de las instancias de un solo grupo en específico.

5.6 PRUEBAS PARA LA TAREA DE REGLAS DE ASOCIACIÓN.

El propósito de la prueba fue determinar los sobrevivientes del Titanic, teniendo en cuenta cuatro variables.

- **Class (1st, 2nd, 3rd, crew).** Clase social de la tripulación.
- **Age (adult, child).** Edad de la tripulación. Adultos y niños.
- **Sex (male, female).** El género de la tripulación.
- **Survived (yes, no).** Tripulación sobreviviente.

El archivo que se utilizó fue "Titanic.arff", cuyas características se pueden ver en la tabla 29.

Tabla 29. Descripción del archivo Titanic.arff

Nombre del archivo	Número de registros	Tamaño	Descripción
Titanic.arff	2201	42KB	Archivo que contiene el registro de las características de los 2201 pasajeros del titanic, estos datos son reales y se han obtenido de "ReportontheLoss of the 'Titanic' (SS)" (1990), British Board of TradeInquiryReport_ (repring), Gloucester, UK: Allan Sutton Publishing.

Con los datos de la tabla 29, se generaron diferentes reglas de asociación. En la figura 105 se muestra una tabla en la cual se puede editar los parámetros básicos para la construcción de las reglas de asociación. Para mayor información de los parámetros que se muestran en la figura 105, consultar manual de usuario de VisualBITool.

Figura 105. Parámetros de las reglas de asociación

Número de reglas	10
Métrica mínima	0.9
Soporte límite inferior	0.1
Soporte límite superior	1.0

Execute ☐ Editar

La tabla de la figura 105, permite configurar el algoritmo de aprendizaje de reglas de asociación A priori; la opción de soporte límite superior indica el límite superior de cobertura requerido para aceptar a un conjunto de ítems. Si no se encuentran conjuntos de ítems para generar las reglas requeridas se va disminuyendo el límite hasta llegar al límite inferior en este caso a la opción soporte límite inferior, con la opción Métrica mínima (u otras métricas dependiendo del criterio de ordenación) que se considerara para mostrar una regla de asociación y finalmente la opción número de reglas en el cual se indica el número de reglas que se desea obtener.

Al dar clic sobre ejecutar, se muestra una tabla con los nombres de las reglas generadas con su respectivo factor de confianza además de una lista de chequeo para seleccionar las reglas que se desean visualizar como se muestra en la figura 106, en este caso se desea graficar 4 reglas, de las cuales 2 reglas tendrán factor de confianza 1 y 2 con factor de confianza menor a 1.

Figura 106. Tabla de reglas de asociación generadas

<input checked="" type="checkbox"/>	Regla0	1
<input checked="" type="checkbox"/>	Regla1	1
<input type="checkbox"/>	Regla2	0.974
<input type="checkbox"/>	Regla3	0.974
<input type="checkbox"/>	Regla4	0.974
<input type="checkbox"/>	Regla5	0.974
<input checked="" type="checkbox"/>	Regla6	0.965
<input type="checkbox"/>	Regla7	0.963
<input type="checkbox"/>	Regla8	0.974

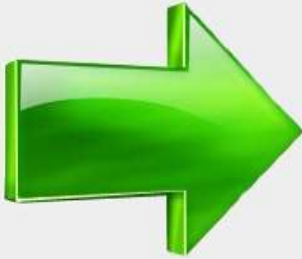
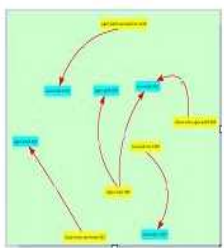
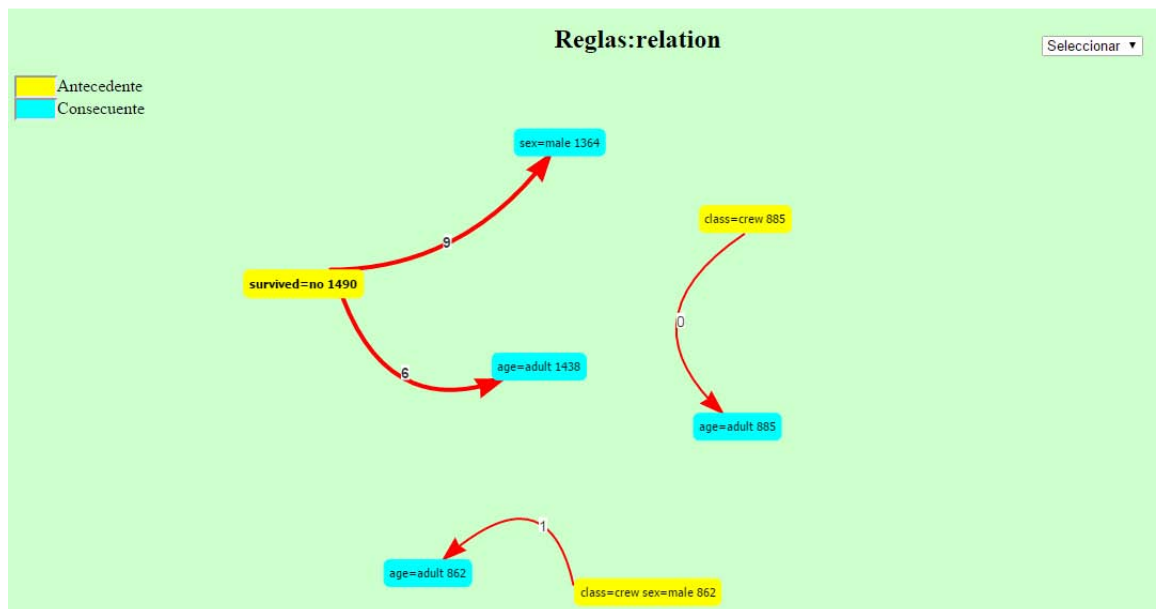


Gráfico de Red



Se seleccionó el gráfico de red, el cual permitió observar la estructura de las reglas con su antecedente y consecuente, como se ve en la figura 107.

Figura 107. Gráfico de red que permite visualizar las reglas de asociación



En cada regla visualizada en la figura 107, se observa la estructura general de una regla, en la cual mediante colores se puede diferenciar el antecedente de color amarillo, el consecuente de color azul y la relación de color rojo.

La figura 108, muestra una tabla que permite visualizar en la parte superior derecha de la gráfica de la figura 107, el antecedente, consecuente y factor de confianza de una regla determinada.

Figura 108. Información del antecedente y consecuente de una regla

Regla6 ▼
Antecedente
survived=no 1490
Consecuente
age=adult 1438
Factor de confianza
0.965

5.6.1 Evaluación de las pruebas de funcionalidad para reglas de asociación.

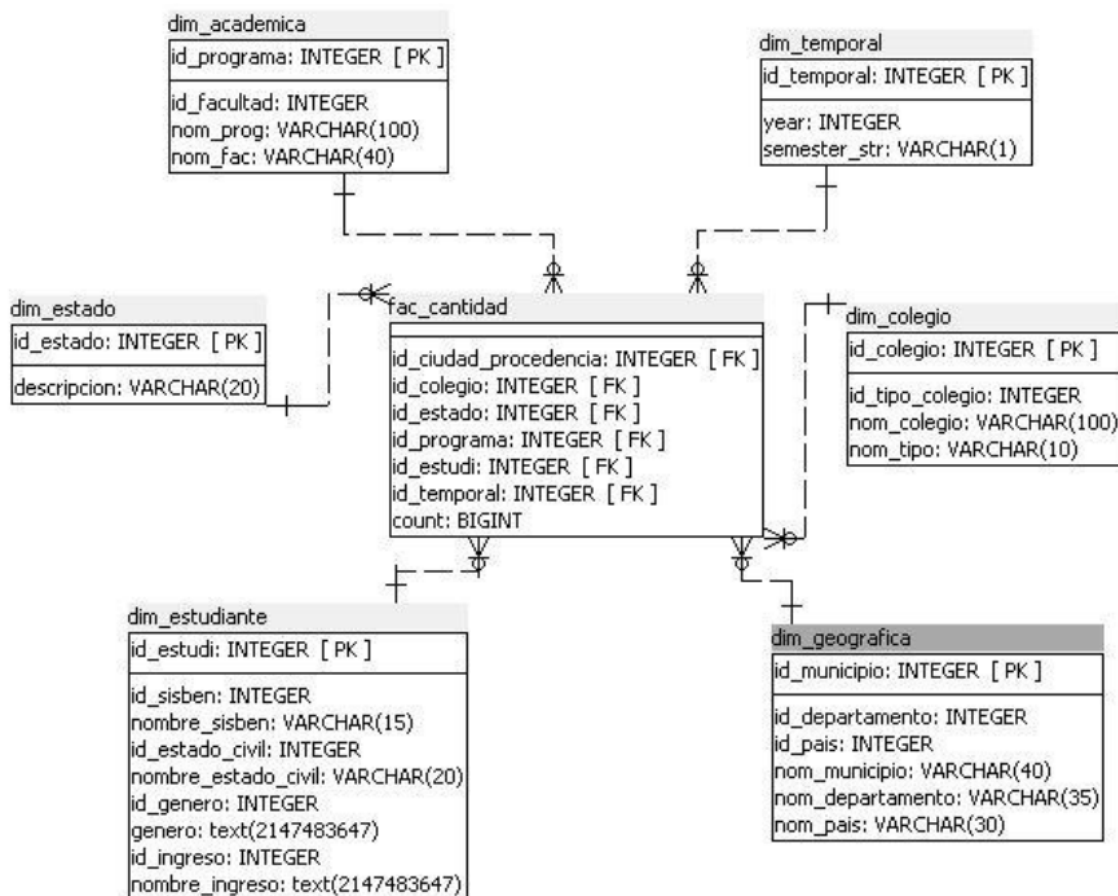
El gráfico de red es una forma de visualizar las reglas de asociación permitiendo mirar su estructura y cuantas reglas se han generado independientes. Además, mediante el uso de colores facilita distinguir el consecuente, antecedente y relación, lo cual hace más cómodo el análisis y la toma de decisiones.

Mediante la personalización de los parámetros para la construcción de las regla de asociación, permite encontrar reglas dependiendo de un factor de confianza, el cual puede ser ingresado por el usuario dependiendo de la necesidad o el caso de estudio que se esté desarrollando.

5.7 PRUEBAS PARA LA TAREA DE ANÁLISIS OLAP

El propósito de la prueba fue realizar consultas sobre la cantidad de estudiantes del repositorio de la Universidad de Nariño denominado, “**repositorio_dwh**”. Para la realización de las pruebas de funcionalidad se hizo uso del repositorio de deserción estudiantil del sistema académico de la Universidad de Nariño, con información desde el año 2004 al 2012, con el cual se construyó el data mart “**cantidad_estudiantes**”, diseñado con esquema en estrella. Su estructura se muestra en la figura 109.

Figura 109. Estructura tipo estrella del data mart Cantidad_estudiantes



El data mart “**cantidad_estudiantes**” de la figura 109, se encuentra formado por seis dimensiones que se describen a continuación:

- **dim_academica.** Contiene información acerca del programa y la facultad en la que se encuentran matriculados los estudiantes.
- **dim_temporal.** Contiene información acerca del año y periodo (A o B) en el cual se matriculó o terminó materias un estudiante.
- **dim_colegio.** Contiene información acerca de la institución educativa en la cual realizaron su bachiller académico.
- **dim_geográfica.** Contiene información acerca de la procedencia de los estudiantes.

- **dim_estudiante.** Contiene información acerca de los estudiantes.
- **dim_estado.** Contiene información acerca de las condiciones en que se encuentra el estudiante, si esta graduado o no entre otras.

En la tabla 30, se describe los datos por cada dimensión del data mart “**cantidad_estudiantes**”.

Tabla 30. Datos del data mart cantidad_estudiantes

Relación	Atributos	Descripción atributo
dim_academica	Id_programa	Identificador del programa
	Id_facultad	Identificador de la facultad
	Nom_prog	Nombre del programa
	Nom_fac	Nombre de la facultad
dim_temporal	Id_temporal	Identificador de cada año
	Year	año
	Semester_str	Semestre matriculado
dim_colegio	Id_colegio	Identificador del colegio
	Id_tipo_colegio	Identificador del tipo del colegio
	Nom_colegio	Nombre del colegio
	Nom_tipo	Nombre del tipo del colegio
dim_geográfica	Id_municipio	Identificador del municipio
	Id_departamento	Identificador del departamento
	Id_pais	Identificador del país
	Nom_municipio	Nombre del municipio
	Nom_departamento	Nombre del departamento
	Nom_pais	Nombre del país
dim_estudiante	Id_estudi	Identificador del estudiante
	Id_sisben	Identificador del nivel del sisben
	Nombre_sisben	Nombre del nivel del sisben
	Id_estado_civil	Identificador del estado civil
	Nombre_estado_civil	Nombre del estado civil
	Id_genero	Identificador del genero
	Genero	Genero de los estudiantes
	Id_ingreso	Identificador de los ingresos
	Nombre_ingreso	Ingresos de los estudiantes
dim_estado	Id_estado	Identificador del estado
	descripción	Nombre del estado del estudiante

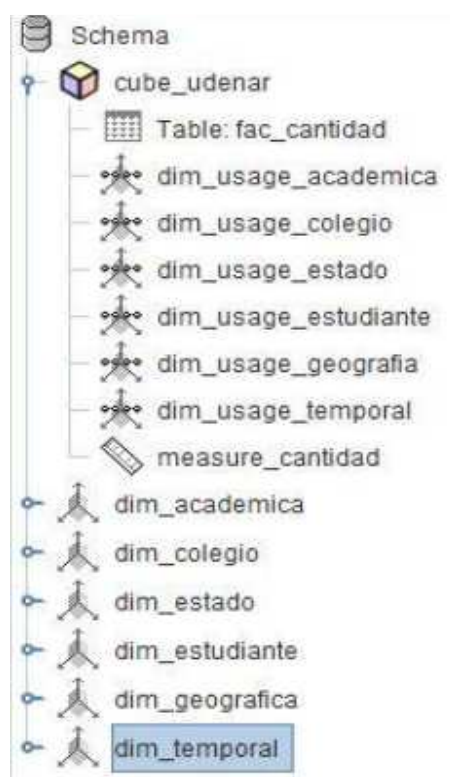
La tabla de hechos del datamart “**cantidad_estudiantes**” se diseñó con la métrica cantidad, para poder saber la cantidad de estudiantes dependiendo el análisis que se desea hacer. En la tabla 31, se describe los datos de la tabla de hechos “**fac_cantidad**” del data mart “**cantidad_estudiantes**”.

Tabla 31. Tabla de hechos del data mart

Fac_cantidad	Id_ciudad_procedencia	Identificador de la ciudad de procedencia del estudiante
	Id_colegio	Identificador del colegio
	Id_estado	Identificador del estado
	Id_programa	Identificador del programa
	Id_estudi	Identificador del estudiante
	Id_temporal	Identificador del año
	Count	Cantidad de estudiantes

De acuerdo a la estructura del data mart se diseñó el cubo “**udenar_cantidad**”, su estructura se conoce en la figura 110.

Figura 110. Estructura del cubo udenar_cantidad



Con el “**cube_udenar**” se hicieron las pruebas para su manipulación en la visualización en la herramienta VisualBITool.

La conexión se realizó con el cubo “**esquema_udenar_cantidad**” y la bodega de datos “**repositorio_dwh**” en PostgreSQL, ver figura 111.

Figura 111. Parámetros de conexión



The image shows a window titled "PostgreSQL" with a table of connection parameters. The table has two columns: the parameter name and its value. The parameters are: Esquema (schema_udenar_cantidad.xml), Servidor (127.0.0.1), Puerto (5432), Base de Datos (repositorio_dwh), Usuario (postgres), and Contraseña (masked with dots). At the bottom left is a blue "Conectar" button, and at the bottom right is a checkbox labeled "Editar" which is checked.

Parámetro	Valor
Esquema	schema_udenar_cantidad.xml
Servidor	127.0.0.1
Puerto	5432
Base de Datos	repositorio_dwh
Usuario	postgres
Contraseña

☒ Editar

Al hacer clic sobre el botón conectar, se observó que el programa si gestiona la conexión, con un mensaje de éxito como se ve en la figura 112.

Figura 112. Mensaje de conexión exitosa



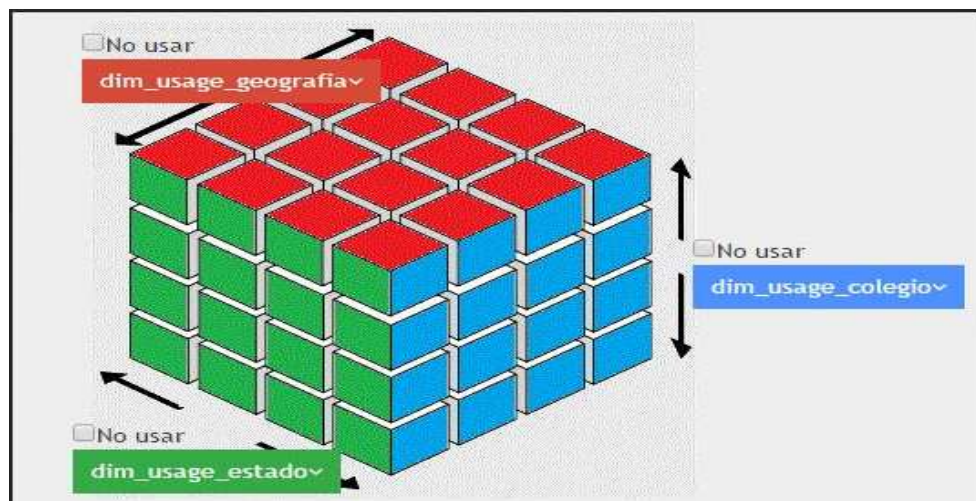
La figura 113, muestra la estructura del cubo además de su nombre y el del esquema. En la tabla se puede seleccionar las tres dimensiones con las que se realizó la consulta y sus respectivos niveles, como también las medidas a seleccionar.

Figura 113. Tabla de selección de dimensiones y niveles

Schema	schema_udenar_cantidad
Cube	cube_udenar
Dimension1	dim_usage_colegio
Level1	level_colegio
Dimension2	dim_usage_estado
Level2	level_estado
Dimension3	dim_usage_geografia
Level3	level_municipio
Measures	<input checked="" type="checkbox"/> measure_cantidad

En la figura 114, se muestra la imagen de un cubo con las tres dimensiones seleccionadas.

Figura 114. Imagen del cubo con las dimensiones



Al hacer clic sobre la imagen del cubo se realizó la consulta de los tres niveles, en este caso se hizo la consulta para conocer la cantidad de estudiantes de cada municipio, estado del estudiante y a que colegio pertenece, como se observa en la figura 115.

Figura 115. Tabla con el resultado de una consulta

Cube VisulaBITools

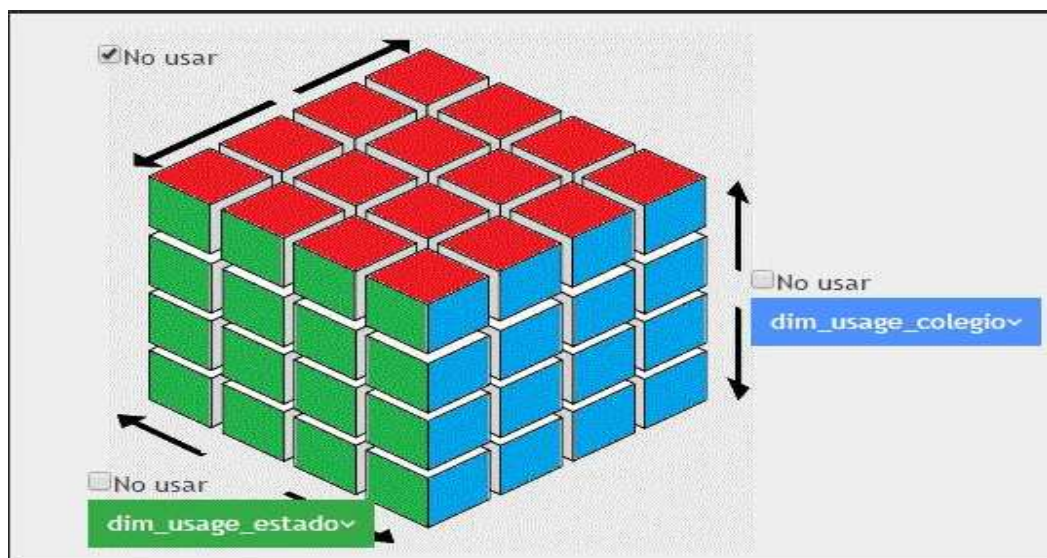
dwhQueryGeneral.jsp

level_colegio	level_estado	level_municipio	measure_cantidad
ATENEO HORIZONTE	EGRESADO SIN TITULO	PASTO	14
	NORMAL	PASTO	14
BACHILLERATO ACADEMICO DEPARTAMENTAL NOCTURNO LICEO BARBACOAS	EGRESADO SIN TITULO	BARBACOAS	14
BACHILLERATO ACADEMICO PARA ADULTOS ANTONIO NARIÑO	EGRESADO SIN TITULO	PUERTO ASIS	11
	NORMAL	PUERTO ASIS	28
BACHILLERATO FEMENINO SAN AGUSTIN	EGRESADO SIN TITULO	PASTO	8
BACHILLERATO NOCTURNO ESCIPION JARAMILLO	EGRESADO SIN TITULO	PASTO	4
	EGRESADO SIN TITULO	TUMACO	5

Descargar SQL Salir

En la siguiente consulta, no se usó el nivel municipio, solo se necesitó saber el colegio y el estado de los estudiantes, ver figura 116.

Figura 116. Imagen del cubo con el uso de dos dimensiones



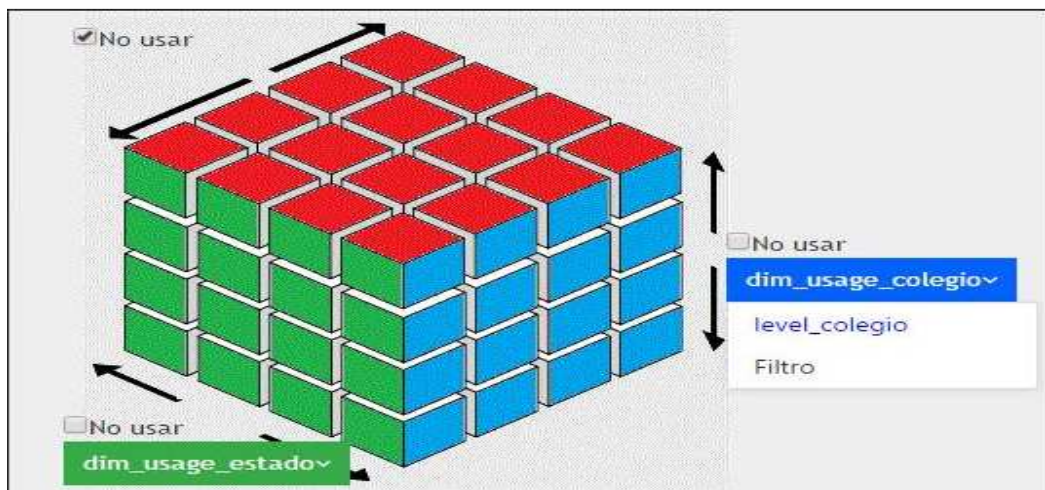
Al hacer clic sobre el cubo se realizó la consulta con el resultado que se puede visualizar en la figura 117.

Figura 117. Tabla con el resultado de una consulta

level_colegio	level_estado	measure_cantidad
ATENEO HORIZONTE	EGRESADO SIN TITULO	14
	NORMAL	14
BACHILLERATO ACADEMICO DEPARTAMENTAL NOCTURNO LICEO BARBACOAS	GRADUADO	14
BACHILLERATO ACADEMICO PARA ADULTOS ANTONIO NARIÑO	EGRESADO SIN TITULO	11
	NORMAL	28
BACHILLERATO FEMENINO SAN AGUSTIN	NORMAL	8
BACHILLERATO OFICIAL FRANCISCO ANTONIO DE ULLOA	NORMAL	9
BACHILLERATO OFICIAL FRANCISCO ANTONIO DE ULLOA	NORMAL	33
CENTRO ACADEMICO ASED POPAYAN	NORMAL	12
CENTRO DE FORMACION PARA LA EDUCACION DE LA COMUNIDAD	NORMAL	11

En la siguiente consulta, se hizo uso de los filtros, se pretendió saber en el colegio Ateneo Horizonte, cuantos estudiantes hay. Además, del estado en que se encuentra, ver figura 118.

Figura 118. Imagen del cubo con un menú para filtros



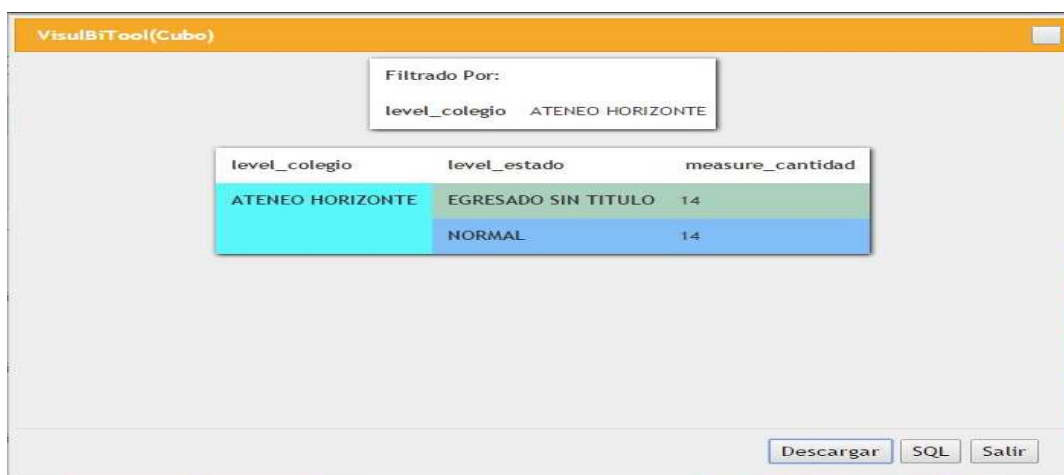
Al seleccionar la opción “Filtro”, que se observa en la figura 118, se pudo seleccionar el colegio por el cual se desea realizar la consulta, ver figura 119.

Figura 119. Tabla para seleccionar un filtro



Después de seleccionar el filtro, se hace clic en “Listo” y luego sobre la imagen del cubo, que se observa en la figura 128 y finalmente se abre una ventana de forma automática con el resultado de la consulta tal y como se muestra en la figura 120.

Figura 120. Tabla con el resultado de una consulta con filtros



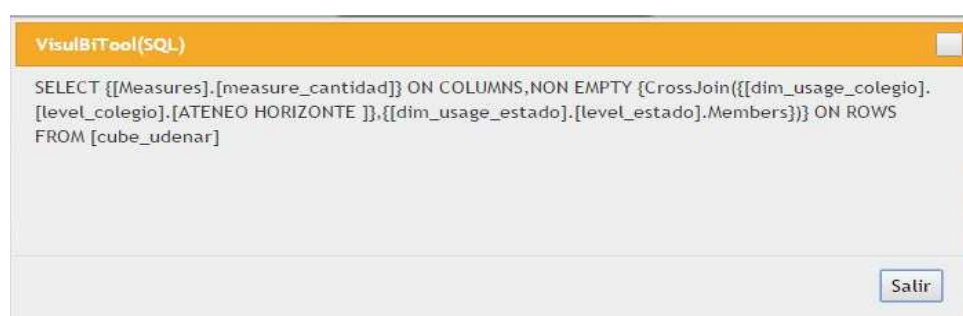
La herramienta VisualBITool, permite descargar en formato Excel (xls), los resultados de las consultas, ver figura 121.

Figura 121. Archivo Excel con los resultados de una consulta

	A	B	C	D	E	F	G	H
1	Filtrado Por:							
2	level_colegio	ATENEO HORIZONTE						
3	level_colegio	level_estado	measure_cantidad					
4	ATENEO HORIZONTE	EGRESADO SIN TITULO	14					
5		NORMAL	14					
6								

Además, al hacer clic sobre el botón “SQL”, que se observa en la figura 120, se pudo observar la estructura de la consulta realizada, ver figura 122.

Figura 122. Ventana que permite ver la estructura de una consulta



5.7.1 Evaluación de las pruebas de funcionalidad para el proceso de análisis OLAP. Los resultados de la prueba de funcionalidad de la tarea de análisis OLAP de la herramienta VisualBITool con el cubo udenar_cantidad demuestran que está funcionando correctamente y que los datos que arrojan las consultas son confiables.

El uso de la herramienta VisualBITool para análisis OLAP, permite un manejo sencillo a los usuarios finales para generar sus reportes y análisis acorde a las necesidades de la organización o negocio, a diferencia a otras herramientas como las hojas de cálculo, permitiendo disminuir la cantidad de datos al limitar la cantidad de dimensiones y niveles que se usan en una consulta a solo tres.

6. CONCLUSIONES

En este proyecto se analizaron las diferentes técnicas de visualización de datos, de las cuales, algunas de ellas se implementaron en la herramienta VisualBITool, que permitieron visualizar de manera gráfica los resultados de las técnicas de minería de datos, clasificación basada en árboles de decisión, reglas de asociación, agrupación y cubos OLAP.

En la herramienta VisualBITool, para el análisis de calidad de datos se implementaron las técnicas de visualización de datos tales como: tabla de datos, gráfico de área, gráfica de barras y gráfico de círculos. En clasificación con árboles de decisión se implementaron las técnicas de visualización de datos tales como árbol jerárquico y árbol de sangría. Para reglas de asociación se implementó la técnica de gráfico de red. Igualmente, para agrupamiento se implementaron las técnicas de gráfico de burbujas y mapa de árbol. Y finalmente, para análisis multidimensional se implementó la tabla de datos.

Las pruebas de funcionamiento de la herramienta VisualBITool, realizadas con repositorios reales y otros disponibles en la web, demostraron que esta funciona correctamente y da solución a las necesidades planteadas en los requerimientos.

Con VisualBITool, las PYMES y MiPYMES cuentan con una herramienta desarrollada bajo software libre para que puedan utilizarla sin ninguna restricción, sin necesidad de conocimientos técnicos profundos, con el fin de explorar y entender su información y convertirla en conocimiento para soportar la toma de decisiones.

Este proyecto permitió a los autores, fortalecer los conocimientos adquiridos en el programa de Ingeniería de Sistemas de la Universidad de Nariño, enfrentar e iniciarse en un proceso investigativo, cuyo resultados se plasman en este informe final.

7. RECOMENDACIONES

Usar la herramienta VisualBITool en las electivas de análisis multidimensional y minería de datos contempladas en el plan curricular del programa de Ingeniería de Sistemas de la Universidad de Nariño como un aporte del grupo de investigación GRIAS KDD al cual pertenecen sus autores.

Implementar esta herramienta en una empresa de la región para evaluar sus resultados y su funcionalidad en una etapa de producción.

REFERENCIA BIBLIOGRÁFICA

- ARFF (2014). Index of /Datasets/UCI/arff.[En línea]. < [http:// repository. seasr. org/ Datasets/UCI/arff/](http://repository.seasr.org/Datasets/UCI/arff/)>. [Consulta: 28/02/2014].
- BAR CHART GUIDE (2014).[En línea]. New York (USA): IBM's Corporate, Many eyes <[http://www-958.ibm.com/ software/ analytics/ manyeyes/ page/ Bar_Chart.HTML](http://www-958.ibm.com/software/analytics/manyeyes/page/Bar_Chart.HTML)>. [Consulta: 05/03/2014].
- BELMONTE, Nicolás (2013).Kit de herramientas de JavaScript InfoVis. [En línea]. SenchaLabs. < <http://philogb.github.io/jit/docs.html>>.[Consulta: 29/07/2014].
- BLOCK HISTOGRAM GUIDE (2014).[En línea]. New York (USA): IBM's Corporate, Many eyes <[http://www-958.ibm.com/ software/analytics/ manyeyes/page/ Block_Histogram.HTML](http://www-958.ibm.com/software/analytics/manyeyes/page/Block_Histogram.HTML)>. [Consulta: 05/03/2014].
- BRAVO, Ariel (2014).Propuesta de guía para el estudio de la estadística en las carreras de ciencias sociales. [En línea].Monografías. <[http:// www. Monografías .com/trabajos80/estudio-estadistica-carreras-ciencias-sociales/estudio-estadistica-carreras-ciencias-sociales2.shtml](http://www.Monografías.com/trabajos80/estudio-estadistica-carreras-ciencias-sociales/estudio-estadistica-carreras-ciencias-sociales2.shtml)>. [Consulta: 29/03/2014].
- BUBBLE CHART GUIDE (2014).[En línea]. New York (USA): IBM's Corporate, Many eyes <[http://www-958.ibm.com/software/ analytics/ manyeyes/ page/ Bubble_Chart.HTML](http://www-958.ibm.com/software/analytics/manyeyes/page/Bubble_Chart.HTML)>. [Consulta: 05/03/2014].
- BUSTOS, Sergio Iván (2014). Metodología SCRUM aplicada a través de un software de consulta de las listas OFAC y ONU de TOPBLS en las PYME colombianas. [En línea]. Universidad Nacional Abierta y a Distancia, Escuela de Ciencias Básicas, Tecnología e Ingeniería, Ingeniería de Sistemas, Bogotá, D.C, Colombia. <[http://repository.unad.edu.co /bitstream/10596/2610 /1 /80258140.pdf](http://repository.unad.edu.co/bitstream/10596/2610/1/80258140.pdf)>. [Consulta: 24/10/2014].
- CASTRO, S & LARREA, M & MAS, R (2011).Visualización. [En línea]. Universidad Nacional de la Plata, Facultad de Informática. < [http://postgrado.info.unlp.edu.ar/Carreras/Especializaciones/Computacion_ Gráfica_ Imagenes_y_Vision_por_Computadora/Pdf/Plan_de_estudios_201 2_Computacion_Gráfica_ Imagenes_y_Vision_por_Computadora_- _Visualizacion.pdf](http://postgrado.info.unlp.edu.ar/Carreras/Especializaciones/Computacion_Gráfica_Imagenes_y_Vision_por_Computadora/Pdf/Plan_de_estudios_2012_Computacion_Gráfica_Imagenes_y_Vision_por_Computadora_-_Visualizacion.pdf)>. [Consulta: 05/03/2014].

- CCANCE, Salomón (2014). Tabla Cascad. [En línea].CcanseWebsite. < http://ccance.net/manuales/thinkcell/think-cell_cap106_tabla_cascada.pdf>. [Consulta: 24/11/2014].
- CHART.JS DOCUMENTATION (2014). [En línea]. <[http:// www. chartjs. org/docs /#radar-chart](http://www.chartjs.org/docs/#radar-chart)>./>.[Con- sulta: 28/12/2014].
- COPYLEFT TITAPG (2008). Estadística para todos: Diagrama de Caja y Bigotes. [En línea]. < <http://www.estadisticaparatodos.es/taller/gráficas/cajas.HTML>>. [Consulta: 25/11/2014].
- CRUZ, José & CAMARENA, Mayta & PIERRE, Jean & HUARINGA, Joaquín & BLAS, Raúl (2007).Evaluación agromorfológica y caracterización molecular de la ñuña (*PHASEOLUS VULGARIS* L.).[En línea]. Universidad Nacional Agraria La Molina, Ap.456, Lima, Perú. <[http://www.scielo.cl/scielo .php?script=sci_arttext&pid = S0718-34292009000100005](http://www.scielo.cl/scielo.php?script=sci_arttext&pid = S0718-34292009000100005)>. [Consulta: 25/11/2014].
- DATA-DRIVEN DOCUMENTS (2013). [En línea]. Mike Bostock. < <http://d3js.org/>>. [Consulta: 29/07/2014].
- DEMŠAR, Janez (2010).Data Mining: Material for lectures on Data mining. Kyoto University, Dept. of Health Informatics. <[http://eprints.fri.uni-lj.si/ 1150/ 1/ DataMining-Kyoto.pdf](http://eprints.fri.uni-lj.si/1150/1/DataMining-Kyoto.pdf)>. [Consulta: 24/11/2014].
- DENDROGRAMA (2013).Dendrograma. [En línea].Wikipedia. < [http:// es.wikip edia.org/ wiki/Dendrograma](http://es.wikipedia.org/wiki/Dendrograma)>./>. [Consulta: 25/11/2014].
- DIAGRAMA DE DISPERSIÓN (2014). [En línea]. < [http:// www.ub .edu / aplica_infor/spss/cap3-7.htm](http://www.ub.edu/aplica_infor/spss/cap3-7.htm)>. [Consulta: 15/03/2014].
- DIAGRAMÁRBOL (2014). Nubes de Etiquetas (TagClouds). [En línea].dgplades.salud.gob.mx. <[http:// www.dgplades.salud.gob.mx/ descargas/ dhg/DIAGRAMA_ÁRBOL.pdf](http://www.dgplades.salud.gob.mx/descargas/dhg/DIAGRAMA_ÁRBOL.pdf)>. [Consulta: 07/03/2014].
- DÜRSTELER, Juan C. (2008). Diagrama de Árbol. En: La revista digital de InfoVis.net. No. 196 (oct.). [En línea].<[http:// www.infovis.net /printMag.php?num =196&lang=1](http://www.infovis.net/printMag.php?num=196&lang=1)>. [Consulta: 07/03/2014].
- DÜRSTELER, Juan C. (2010).Coordenadas Paralelas. [En línea]. Revista digital de InfoVis.net. < <http://www.infovis.net/printMag.php?num=201&lang=1>>. [Consulta: 25/11/2014].

- FUNDIBEQ (2014).DIAGRAMA DE DISPERSIÓN. [En línea].FUNDIBEQ. <http://www.fundibeq.org/opencms/export/sites/default/PWF/downloads/gallery/methodology/tools/diagrama_de_dispersion.pdf >. [Consulta: 28/03/2014].
- GARCIA, María &ARÁNZAZU, Álvarez (2014).Análisis de Datos en WEKA – Pruebas de Selectividad. [En línea].Universidad Carlos III: Ingeniería de Telecomunicaciones. <<http://www.it.uc3m.es/jvillena/irc/practicass/06-07/28.pdf>>. [Consulta: 05/11/2014].
- GRÁFICOS DE EXCEL (2010). [En línea]. <[http:// gráficosdeexcel. blogspot. com/2010/05/gráficos-de-lineas_5783.html](http://gráficosdeexcel.blogspot.com/2010/05/gráficos-de-lineas_5783.html) >. [Consulta: 28/05/2014].
- GRAFMAPA (2014).Gráfico de mapa. [En línea]. < [https:// support. google. com/docs/answer/1409802?hl=es](https://support.google.com/docs/answer/1409802?hl=es) >. [Consulta: 28/05/2014].
- HASSAN MONTERO, Y. (2010).Visualización de Información Persona-Operador: Propuesta Algorítmica para la ordenación espacial de Grafos. [En línea].Tesis de Grado. Universidad de Granada. Facultad de Comunicación y Documentación. < <http://hera.ugr.es/tesisugr/18876390.pdf> >. [Consulta: 19/07/2014].
- INDENTED TREE (2014). [Enlínea].GitHub. < <https://github.com/mbostock/d3/wiki/Gallery>>. [Consulta: 25/11/2014].
- LASKOWSKI, N. (2012).Atrasos en IT impulsan la compra de herramientas de visualización de datos. [En línea].TechTarget. <<http://searchdatacenter.techtarget.com/es/noticias/2240172302/Atrasos-en-IT-impulsan-la-compra-de-herramientas-de-visualizacion-de-datos>>. [Consulta: 28/03/2014].
- LINE GRAPH GUIDE (2014).[Enlínea]. New York (USA): IBM's Corporate, Many eyes [http://www-958.ibm.com/ software/ analytics/ manyeyes/ page/ Line_Graph.HTML](http://www-958.ibm.com/software/analytics/manyeyes/page/Line_Graph.HTML) >. [Consulta: 05/03/2014].
- MAPASDATOS (2013).Chips gráficos logran procesar conjuntos de 'big data' en milisegundos. [En línea].Madrid, España: MIT TechnologyReview. <http://www.technologyreview.es/read_article.aspx?id=44073 >. [Consulta: 30/03/2014].
- MARTINEZ, Miguel A. (2011). La importancia de la visualización de la información. [En línea].Blog.San Juan de Mozarrifar (España): CEPYME Aragón. < <http://cepymearagon.blogspot.com/2011/04/la-importancia-de-la-visualizacion-de.html> >. [Consulta: 19/06/2014].

- MASEXCEL (2010). Mejoras para el gráfico de Cascada o Waterfall. <<http://masexcel.blogspot.com/2010/08/mejoras-para-el-gráfico-de-cascada-o.html>>. [Consulta: 24/11/2014].
- MATRIX CHART GUIDE (2014). [En línea]. New York (USA): IBM's Corporate, Many eyes <http://www-958.ibm.com/software/analytics/manyeyes/page/Matrix_Chart.html>. [Consulta: 06/03/2014].
- MICROSOFT-OFFICE (2007). Tipos de gráficos disponibles. [En línea]. Redmond, WA, Estados Unidos: Microsoft Corporation. <<https://support.office.com/es-hn/article/Tipos-de-gr%C3%A1ficos-disponibles-a6187218-807e-4103-9e0a-27cdb19afb90?ui=es-ES&rs=es-HN&ad=HN>>. [Consulta: 07/03/2014].
- MONTILVA, Jonás & ARAPÉ, Nelson & COLMENARES, Juan. (2014). Desarrollo de Software Basado en Componentes. [En línea]. Universidad de los Andes: Mérida (Venezuela). Universidad del Zulia: Maracaibo (Venezuela). <<http://webdelprofesor.ula.ve/ingenieria/jonas/Productos/Publicaciones/Congresos/CAC03%20Desarrollo%20de%20componentes.pdf>>. [Consulta: 22/12/2014].
- NETWORK DIAGRAM GUIDE (2014). [En línea]. New York (USA): IBM's Corporate, Manyeyes <http://www-958.ibm.com/software/analytics/manyeyes/page/Network_Diagram.HTML>. [Consulta: 05/03/2014].
- ONO (2010). Diseña una nube de tags original y personalizada con Tagxedo.com. [En línea]. Oblog. <<http://blog.ono.es/2010/08/disena-una-nube-de-tags-original-y-personalizada-con-tagxedo-com/>>. [Consulta: 30/03/2014].
- PIE CHART GUIDE (2014). [En línea]. New York (USA): IBM's Corporate, Many eyes. <http://www-958.ibm.com/software/analytics/manyeyes/page/Pie_Chart.HTML>. [Consulta: 05/03/2014].
- RADIAL (2014). Gráfico radial. [En línea]. Mountain View, CA, Estados Unidos: Google. <<https://support.google.com/docs/answer/1047430?hl=es>>. [Consulta: 06/03/2014].
- RODAS, Janina (2012). Data Mining. [En línea]. Prezi. <<https://prezi.com/clccpxbjotj/data-mining/>>. [Consulta: 15/08/2014].
- RUMBAUGH, James; JACOBSON, Ivar & BOOCH, Grady (1999). El Lenguaje Unificado de Modelado, Manual de Referencia. [En línea]. Addison Wesley. <<https://ingenieriasoftware2011.files.wordpress.com/2011/07/el-lenguaje-unificado-de-modelado-manual-de-referencia.pdf>>. [Consulta: 09/04/2015].

- SCATTERPLOT VISUALIZATION GUIDE (2014). [En línea]. New York (USA): IBM's Corporate. <<http://www-958.ibm.com/software/analytics/manyeyes/page/Scatterplot.html>>. [Consulta: 06/03/2014].
- SCHWABER, Ken & SUTHERLAND, Jeff (2013). La Guía de Scrum. [En línea]. <<http://www.scrumguides.org/docs/scrumguide/v1/Scrum-Guide-ES.pdf>>. [Consulta: 24/10/2014].
- STACK GRAPH GUIDE (2014).[En línea]. New York (USA): IBM's Corporate, Many eyes. <http://www-958.ibm.com/software/analytics/manyeyes/page/Stack_Graph.html>. [Consulta: 05/03/2014].
- STACK GRAPHS FOR CATEGORIES (2014).[En línea]. New York (USA): IBM's Corporate, Many eyes. <http://www-958.ibm.com/software/analytics/manyeyes/page/Stack_Graph_for_Categories.HTML>. [Consulta: 05/03/2014].
- TABLAS (2014).Animando tablas en PowerPoint. [En línea].Tutoriales al día. <<http://www.presentyppt.tutorialesaldia.com/animando-tablas-en-powerpoint/>>.[Consulta: 25/11/2014].
- TABLEAU SOFTWARE (2014). Mapa de calor. [En línea]. Onlinehelp. <http://onlinehelp.tableau.com/v7.0/pro/online/es-es/buildexamples_heatmap.html>. [Consulta: 05/03/2014].
- TREEMAP GUIDE (2014).[En línea]. New York (USA): IBM's Corporate, Manyeyes.<<http://www-958.ibm.com/software/analytics/manyeyes/page/Treemap.HTML>>. [Consulta: 05/03/2014].
- TRIGAS, Manuel & DOMINGO, Ana C. (2014).Gestión de Proyectos Informáticos. [En línea] .TFC.<<http://openaccess.uoc.edu/webapps/o2/bitstream/10609/17885/1/mtrigasTFC0612memoria.pdf>>. [Consulta: 24/10/2014].
- UNESCO (2014). Exploración gráfica de datos. [En línea].Unesco. <<http://www.unesco.org/webworld/portal/idams/html/spanish/S1graph.htm>>. [Consulta: 28/03/2014].
- URRIBARRI, Dana K. & CASTRO, Silvia M. & MARTIG, Sergio R. (2014). Interacciones sobre coordenadas paralelas en el marco del Modelo Unificado de Visualización. [En línea]. Secretaria General de Ciencia y Tecnología, Universidad Nacional del Sur, Bahía Blanca, Argentina. <http://sedici.unlp.edu.ar/bitstream/handle/10915/22463/Documento_completo.PDF?sequence=1>. [Consulta: 24/10/2014].

VIS.JS UN SISTEMA DE INTERACCIÓN VISUAL (2014). [En línea].Almende B.V. <<http://visjs.org/>>. [Consulta: 29/07/2014].

VIZCAINO, Paula (2008). Aplicación de técnicas de inducción de árboles de decisión a problemas de clasificación mediante el uso de WEKA (WaikatoEnvironmentforKnowledgeAnalysis). [En línea].Bogotá-Colombia: Fundación Universitaria Konrad Lorenz, Facultad de Ingeniería de Sistemas. <http://www.konradlorenz.edu.co/images/stories/suma_digital_sistemas/2009_01/final_paula_andrea.pdf>. [Consulta: 15/08/2014].

VORONOI (2014).Diagramas de Voronoi. [En línea]. < [http:// www.dma.fi.upm.es/mabellanas/voronoi/voronoi/voronoi.html](http://www.dma.fi.upm.es/mabellanas/voronoi/voronoi/voronoi.html)>. [Consulta: 25/11/2014].

ANEXOS

La carpeta denominada **Anexos-VisualBITool** adjunta en el CD, contiene todos los archivos pertinentes al proyecto desarrollado, los cuales se organizan de la siguiente manera.

Anexo1.Documentos: Contiene los diferentes documentos elaborados en el desarrollo de la investigación en formato Word y PDF, los documentos son:

- Documento de prueba y evaluación de resultados, el cual contiene las pruebas de funcionalidad de la herramienta VisualBITool con datos reales referentes a la Pruebas Saber Pro realizadas a los estudiantes de la Universidad de Nariño.
- Manual de usuario, el cual contiene información sobre la configuración y uso de la herramienta VisualBITool.
- Documento de Herramientas de desarrollo, el cual presenta información sobre las bibliotecas, lenguajes de programación y plataforma de desarrollo de VisualBITool.

Anexo2.Software: contiene todos los archivos de funcionamiento y el código de desarrollo de la herramienta VisualBITool.

Anexo3.Archivo WAR: contiene el archivo ejecutable de la herramienta VisualBITool.

Anexo4.Repositorio de datos: contiene los diferentes repositorios de datos en formato .ARFF, archivo XML del cubo de datos. Además, el Backup del repositorio de datos, utilizado en las pruebas de funcionalidad para análisis multidimensional.