



**ESPE**  
UNIVERSIDAD DE LAS FUERZAS ARMADAS  
INNOVACIÓN PARA LA EXCELENCIA

**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN**

**MAESTRÍA EN SISTEMAS DE INFORMACIÓN E  
INTELIGENCIA DE NEGOCIOS**

**TRABAJO DE TITULACION, PREVIO A LA OBTENCION DEL  
TITULO DE MAGISTER EN SISTEMAS DE INFORMACION E  
INTELIGENCIA DE NEGOCIOS**

**TEMA: MANEJO BÁSICO DE INDICADORES DE  
PARTICIPACIÓN CIUDADANA UTILIZANDO ALGORITMOS Y  
TÉCNICAS DE MINERÍA DE DATOS COMO APOYO EN LA  
GESTIÓN COMUNITARIA, CASO DE ESTUDIO QUITUMBE**

**AUTOR: SINCHIGUANO VIZUETE MARICELLA ALEXANDRA**

**DIRECTOR: SANCHO ARIAS JOSE ALBERTO**

**SANGOLQUI**

**2017**



VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

PROGRAMA DE MAESTRÍA EN GESTIÓN DE SISTEMAS DE  
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

CERTIFICACIÓN

Certifico que el trabajo de titulación, "MANEJO BÁSICO DE INDICADORES DE PARTICIPACIÓN CIUDADANA UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS COMO APOYO EN LA GESTIÓN COMUNITARIA, CASO DE ESTUDIO QUITUMBE" realizado por la señorita MARICELLA ALEXANDRA SINCHIGUANO VIZUETE, ha sido revisado en su totalidad y analizado por el software anti-plagio, el mismo cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de Fuerzas Armadas ESPE, por lo tanto me permito acreditarlo y autorizar a la señorita MARICELLA ALEXANDRA SINCHIGUANO VIZUETE para que lo sustente públicamente.

Sangolquí, 20 de septiembre del 2017

JOSÉ ALBERTO SANCHO ARIAS

**DIRECTOR**



**VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA**

**CENTRO DE POSGRADOS**

**PROGRAMA DE MAESTRÍA EN GESTIÓN DE SISTEMAS DE  
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS**

**AUTORÍA DE RESPONSABILIDAD**

Yo, MARICELLA ALEXANDRA SINCHIGUANO VIZUETE, con cédula de identidad N° 1719482224, declaro que este trabajo de titulación “MANEJO BÁSICO DE INDICADORES DE PARTICIPACIÓN CIUDADANA UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS COMO APOYO EN LA GESTIÓN COMUNITARIA, CASO DE ESTUDIO QUITUMBE” ha sido desarrollado considerando los métodos de investigación existentes, así como también se ha respetado los derechos intelectuales de terceros considerándose en las citas bibliográficas.

Consecuentemente declaro que este trabajo es de mi autoría, en virtud de ello me declaro responsable del contenido, veracidad y alcance de la investigación mencionada.

Sangolquí, 20 de septiembre del 2017

MARICELLA ALEXANDRA SINCHIGUANO VIZUETE

C.C. 1719482224



VICERRECTORADO DE INVESTIGACIÓN INNOVACIÓN Y  
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

PROGRAMA DE MAESTRÍA EN GESTIÓN DE SISTEMAS DE  
INFORMACIÓN E INTELIGENCIA DE NEGOCIOS

#### AUTORIZACIÓN

Yo, MARICELLA ALEXANDRA SINCHIGUANO VIZUETE, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar en la biblioteca Virtual de la institución el presente trabajo de titulación "MANEJO BÁSICO DE INDICADORES DE PARTICIPACIÓN CIUDADANA UTILIZANDO ALGORITMOS Y TÉCNICAS DE MINERÍA DE DATOS COMO APOYO EN LA GESTIÓN COMUNITARIA, CASO DE ESTUDIO QUITUMBE" cuyo contenido, ideas y criterios son de mi autoría y responsabilidad.

Sangolquí, 20 de septiembre del 2017

MARICELLA ALEXANDRA SINCHIGUANO VIZUETE

C.C. 1719482224

## **DEDICATORIA**

A Dios por darme sabiduría de seguir adelante con mis metas.

A mi esposo Wiliam por su apoyo y confianza depositada en mí; gracias por ser parte de mis triunfos y mis fracasos en la etapa de la Maestría; ya que sin tu tiempo y comprensión no hubiese sido posible este sueño. Fuiste, un excelente padre mientras yo no estaba en casa con ustedes y el mejor amigo mientras asistía al proceso de clases. Gracias por creer en mí y ser parte de este sueño cumplido.

Gracias Wiliam, Sebastián y Daniel.

Maricella Alexandra Sinchiguano Vizuite

## **AGRADECIMIENTOS**

Quiero agradecer a mis padres y a mi familia, por estar siempre a mi lado y ser la fuerza que necesité; cuando pensé que no lo lograría.

A mi mami que siempre ha sido y será el mejor regalo que Dios me pudo dar, por creer en mí y mis sueños.

A mis hijos por ser el complemento de mi vida y los que me motivan para ser cada día mejor y seguir adelante luchando.

A mis suegros que siempre estuvieron ahí junto a mis hijos cuando no podía estar con ellos, gracias por el tiempo, cariño, amor y paciencia.

Agradezco a mi tutor Msc. Ing. José Sancho por su confianza y su constante revisión a mi tesis de maestría; sin contar con el tiempo dedicado a cada una de mis dudas y por su ayuda incondicional.

Maricella Alexandra Sinchiguano Vizuite

## ÍNDICE DE CONTENIDOS

DEDICATORIA .....	v
AGRADECIMIENTOS .....	vi
ÍNDICE DE CONTENIDOS .....	vii
ÍNDICE DE FIGURAS .....	x
ÍNDICE DE TABLAS .....	xii
RESUMEN .....	xiii
ABSTRACT .....	xiv
CAPÍTULO I .....	1
GENERALIDADES .....	1
1.1.  Introducción .....	1
1.2.  Antecedentes .....	1
1.3.  Delimitación Geográfica .....	2
1.4.  Justificación e Importancia .....	2
1.5.  Planteamiento del problema .....	3
1.6.  Formulación del problema .....	3
1.7.  Objetivos General .....	4
1.8.  Objetivos Específicos .....	4
CAPITULO II .....	5
MARCO TEORICO .....	5
2.1.  Antecedentes Investigativos .....	5
2.2.  Estado del Arte en la Minería de Datos .....	5
2.3.  Antecedentes Conceptuales y Referencias de Minería de Datos. ....	7
2.3.1.  Base de datos Estáticas: .....	8
2.3.2.  Bases De Datos Dinámicas: .....	8
2.3.3.  Modelos de Bases de Datos: .....	8
2.3.4.  Gestión de bases de datos distribuida (SGBD): .....	11
2.3.5.  Potencialidad de las Bases de Datos: .....	11
2.3.6.  Minería de Datos: .....	15
2.4.  Herramientas de Software .....	17
2.4.1.  Weka .....	17
2.4.2.  Rapid Miner: (YALE =yet another learning Environment) .....	17
2.4.3.  Orange: .....	17
2.4.4.  KNIME: Konstanz Information Miner .....	18
2.4.5.  JHepWork: .....	18
2.4.6.  Data Mining Add-In for Excel: .....	18

2.5.	Metodologías para Minería de Datos .....	18
2.5.1.	SEMMA.....	18
2.5.2.	DMAMC .....	19
2.5.3.	CRISP-DM.....	19
2.5.4.	CATALYST.....	19
2.6.	Cubos de Información.....	20
2.6.1.	OLAP .....	20
2.6.2.	OLTP.....	21
2.6.3.	ROLAP .....	21
2.6.4.	HOLAP .....	21
2.7.	Definición de técnicas de minería de datos.....	22
2.7.1.	Redes Neuronales.....	22
2.7.2.	Árboles de Decisión .....	22
2.7.3.	Algoritmos Genéricos .....	23
2.7.4.	Clustering.....	23
2.8.	Metodología para llevar a cabo una encuesta .....	23
CAPITULO III .....		25
ANALISIS DEL CASO DE ESTUDIO .....		25
3.1.	Aplicación de la Metodología CRISP-DM en el caso de Estudio.....	25
3.2.	Comprensión del negocio.....	29
3.2.1.	Objetivo del Negocio .....	29
3.2.2.	Evaluación del Negocio .....	29
3.2.3.	Objetivos de la Minería de Datos.....	30
3.2.4.	Realizar el Plan del Proyecto .....	30
3.2.5.	Evaluación Inicial de Herramientas y Técnicas .....	31
3.3.	Comprensión de datos.....	31
3.3.1.	Recolección de los datos .....	32
3.3.2.	Describir los datos.....	35
3.3.3.	Explorar los Datos.....	39
3.3.4.	Verificar la calidad de los Datos .....	62
3.4.	Preparación de datos .....	65
3.4.1.	Selección de Datos .....	65
3.4.2.	Limpieza de Datos .....	66
3.4.3.	Construir Datos .....	66
3.4.4.	Integrar Datos.....	67
3.4.5.	Formatear Datos.....	67



3.5	. Modelado .....	67
3.5.1.	Selección de la Técnica de Modelado.....	68
3.5.2.	Generación de la prueba de diseño.....	69
3.5.3.	Construcción del Modelo .....	70
3.5.4.	Evaluación del Modelo .....	78
3.6.	Evaluación.....	80
3.6.1.	Evaluación de los Resultados.....	80
3.6.2.	Revisión del negocio.....	81
3.6.3.	Determinación de los Resultados.....	81
3.7.	Implantación .....	82
3.7.1.	Planear la Implantación.....	82
3.7.2.	Planear la Monitorización y Mantenimiento.....	82
3.7.3.	Informe Definitivo del Producto .....	83
CAPITULO IV .....		87
CONCLUSIONES Y LINEAS DE TRABAJO FUTURO.....		87
4.1	Conclusiones.....	87
4.2	Líneas de Trabajo .....	87
BIBLIOGRAFIA .....		88
GLOSARIO DE TERMINOS .....		91

## ÍNDICE DE FIGURAS

Figura 1. Localización Geográfica.....	2
Figura 2. Fases del proceso KDD .....	13
Figura 3. Esfuerzo requerido para el desarrollo de cada una de las fases.....	14
Figura 4. Fases de la Metodología .....	27
Figura 5. Tabla Persona.....	36
Figura 6. Entidades Responsables.....	37
Figura 7. Tabla de Requerimientos .....	38
Figura 8. Ubicación Geográfica .....	39
Figura 9. Gráfica Histograma de la Primera Pregunta - EDAD.....	41
Figura 10. Gráfica Histograma de la Variable Género .....	41
Figura 11. Gráfica Histograma Nivel de Educación .....	42
Figura 12. Gráfica Uso de la Tecnología vs. Nivel Económico.....	43
Figura 13. Gráfica Histograma uso de la tecnología.....	43
Figura 14. Gráfica Tipo de Tecnología .....	44
Figura 15. Gráfica Nivel de Educación vs. Uso Tecnología.....	45
Figura 16. Gráfica tipo de Tecnología vs. Género .....	45
Figura 17. Gráfica de Árbol de decisión de Uso de la Tecnología .....	46
Figura 18. Gráfica de la Descripción del Árbol de Uso de la Tecnología .....	47
Figura 19. Gráfica de los Rangos de Edad .....	49
Figura 20. Gráfica del Género.....	49
Figura 21. Gráfica del Nivel de Educación.....	50
Figura 22: Árbol de Decisión Del Nivel de Ingreso Económico .....	51
Figura 23. Gráfica de Descripción Del Árbol de Nivel de Ingreso.....	51
Figura 24. Gráfica de la Tabla de Encuestas de la Pregunta Tres.....	52
Figura 25. Gráfica de Servicios Público vs. Nivel de Ingresos.....	53
Figura 26. Gráfica de Servicios Públicos .....	53
Figura 27. Gráfica del Árbol de Decisión Genero vs. Servicios Públicos .....	54
Figura 28. Gráfica Árbol de Decisión - Rangos de Edad vs. Nivel de Ingreso .....	54
Figura 29. Gráfica Arboles De Decisión.....	55
Figura 30. Grafica Ordenamiento Territorial .....	56
Figura 31. Grafica Facilidad de procesos.....	57
Figura 32. Grafica de Facilidad Procesos .....	58
Figura 33. Árboles de Decisión Facilidad de proceso .....	58
Figura 34. Grafica Facilidad Procesos vs.Obra vial.....	60
Figura 35. Grafica Obra Publica vial .....	60
Figura 36. Árboles de Decisión Facilidad de Procesos.....	61
Figura 37. Grafica Facilidad Procesos vs. Nivel Económico - Nivel de Educación..	62
Figura 38. Grafica desarrollo en Rapidminer.....	62
Figura 39. Graficas de Valores Atípicos Rangos Edad.....	63
Figura 40. Grafica de Valores Atípicos Nivel ingresos .....	64
Figura 41. Grafica de Valores Atípicos Nivel de Educación .....	64
Figura 42: Gráfica Proceso vs. Análisis .....	66
Figura 43. Modelo Vista Presentación .....	69
Figura 44. Estructura / arquitectura del componente EJB Module .....	69

Figura 45. Gráfica Nivel de Educación.....	72
Figura 46. Gráfica Nivel de Ingreso Económico .....	74
Figura 47. Gráfica Nivel de Ingreso Económico vs. Edad.....	75
Figura 48. Gráfica de Uso de la Tecnología .....	76
Figura 49. Gráfica que tipo de Tecnología Usa .....	76
Figura 50: Gráfica que tipo de Participación en el Portal Web con la ciudadanía ....	77
Figura 51. Gráfica Servicios Públicos vs. Género .....	78

## ÍNDICE DE TABLAS

Tabla 1: Fases del Proceso de Minería de Datos .....	28
Tabla 2: Actividad del Proyecto.....	30
Tabla 3: Total Pobladores de la Sector Quitumbe .....	32
Tabla 4: Cálculo de la Muestra .....	34
Tabla 5: Plan de Recolección y Técnicas.....	35
Tabla 6: Tabla de Encuestas de la Primera Pregunta. ....	40
Tabla 7: Encuesta de la Pregunta Dos.....	48
Tabla 8. Encuesta Pregunta Cinco .....	59
Tabla 9: Indicadores y Variables.....	67
Tabla 10: Representation regression Lineal – Pregunta 1.....	71
Tabla 11: Gráfica de Datos Homologados .....	73
Tabla 12: Representation regression Lineal – Pregunta 2.....	73

## **RESUMEN**

La Inteligencia de Negocios es una herramienta bajo la cual diferentes tipos de organizaciones, pueden soportar la toma de decisiones, basadas en información precisa y oportuna. La investigación comienza con la obtención de los indicadores planteados por la Concejalía del Municipio de Quito del área de Ordenamiento Territorial; los mismos que nos permiten determinar los niveles de satisfacción e interacción de los ciudadanos, mediante parámetros que permitan obtener registros ya sea por su Género, Edad, Nivel Económico, Nivel de Instrucción, sabiendo que el rango de edad esta entre 20 y 65 años, y los demás indicadores se determinan según su grado alto medio y bajo. El Municipio de Quito es consciente de la problemática que existe hoy en día, los ciudadanos desconocen el tiempo promedio que se tarda en resolver un requerimiento pedido para su barrio, lo que se refleja en las encuestas el momento de analizar su satisfacción por agilidad de respuesta a sus peticiones. Además, las personas indican que a mayor tiempo se prolongue una petición realizada, el barrio seguirá con deficiencia en seguridad, servicios básicos u otro requerimiento realizado al Municipio. El objetivo es crear un conjunto de datos significativo y manejable en cuanto al tamaño, para poder detectar valores anómalos y eliminarlos a través de la herramienta de Minería de Datos que vamos a utilizar. Dentro de esta fase entran las operaciones de transformación y reducción de datos, ya que en muchas ocasiones estas operaciones marcan el éxito o fracaso del proceso.

### **PALABRAS CLAVES:**

- **INTELIGENCIA DE NEGOCIOS**
- **INDICADORES**
- **MINERIA DE DATOS**
- **VALORES ANÓMALOS**
- **REDUCCION DE DATOS**
- **LIMPIEZA DE DATOS**

## **ABSTRACT**

Business Intelligence is a tool under which different types of organizations can support decision making, based on accurate and timely information. The investigation begins with the obtaining of the indicators proposed by the Council of the Municipality of Quito the area of Territorial Ordering; The same ones that allow us to determine the levels of satisfaction and interaction of the citizens, by means of parameters that allow to obtain registers either by their Gender, Age, Level of Instruction. The Municipality of Quito is aware of the problems that exist today, citizens are unaware of the average time it takes to resolve an order request for their neighborhood, which is reflected in the polls when analyzing their satisfaction for agility of response To their requests. In addition, people indicate that a longer request will be extended, the neighborhood will continue to be deficient in security, basic services or other requirements made to the Municipality. The goal is to create a data set that is meaningful and manageable in order to be able to detect anomalous values and eliminate them through the Data Mining tool that we are going to use. Within this phase, data transformation and reduction operations enter, as these operations often mark the success or failure of the process.

### **KEYWORDS:**

- **BUSINESS INTELLIGENCE**
- **INDICATORS, DATA MINING**
- **ANOMALOUS VALUES**
- **DATA REDUCTION**
- **DATA CLEANING.**

# **CAPÍTULO I**

## **GENERALIDADES**

### **1.1. Introducción**

Actualmente la minería de datos es un complemento natural al proceso de explorar y entender los datos a través de BI tradicional. Los algoritmos automáticos pueden procesar cantidades de datos muy grandes y detectar patrones y tendencias que, de lo contrario, estarían ocultos.

Para realizar minería de datos, se debe realizar las siguientes preguntas:

"¿Quiénes son mis clientes?" o "¿Qué productos adquirieron?", y aplicar después un algoritmo para encontrar correlaciones estadísticas en los datos. Los patrones y las tendencias que se detectan en el análisis se almacenan en forma de modelo de minería de datos.

En la época actual la complejidad de la base de datos crece al igual que la enorme cantidad de registros y se vuelven más difíciles de interpretar y analizar para la organización que desea utilizarlos para su propio beneficio.

### **1.2. Antecedentes**

La participación ciudadana, desde hace varias décadas, la intervención más o menos directa de los ciudadanos y ciudadanas para la toma de decisiones públicas; se ha convertido en un objetivo deseable en los discursos oficiales de gobiernos de todo nivel e instituciones internacionales de todo tipo.

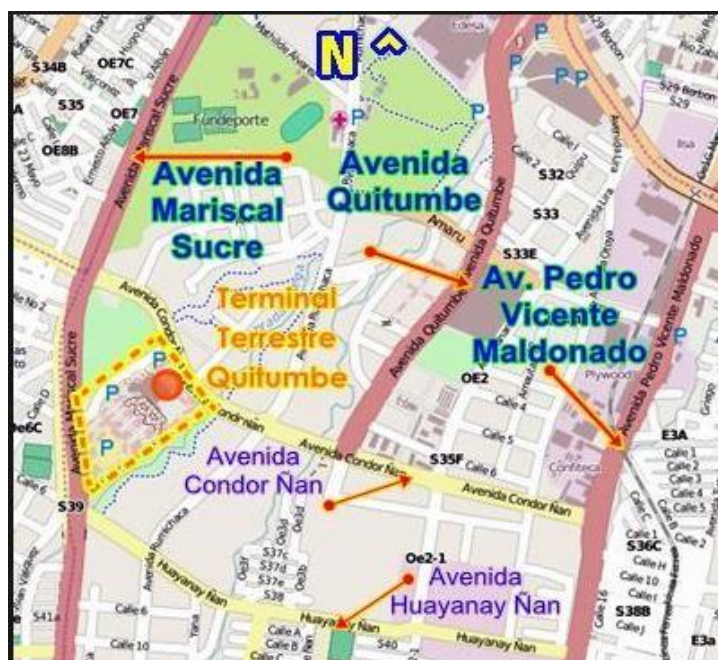
Esta noción políticamente correcta de participación ciudadana, sin embargo, ampara una importante polisemia que se refleja en formas y actuaciones muy diversas que pretenden atender a la preocupación por el involucramiento de las personas en las decisiones que los afectan. Mediante el desarrollo de este proyecto se pretende tener una aproximación con los ciudadanos, esto es, sus diferentes grados y tipos, así como los distintos propósitos que los orientan, lo que puede resultar de utilidad como un marco para el análisis.

El actual interés por la participación se remonta, en buena medida, a los debates y propuestas de los movimientos sociales y la contestación política de los años sesenta, que somete a crítica la lógica institucional formal y burocrática de los Estados.

En otros casos, las fórmulas de participación establecidas apuntan a permitir la intervención ciudadana, de manera consultiva, resolutive y/o fiscalizadora, en procesos de gestión pública, esto es, en el diseño e implementación de programas y proyectos, como también de grandes planes de desarrollo (planificación estratégica, desarrollo urbano, etc.) que, por lo demás, se sitúan a medio camino con el tipo de decisiones de política pública.

### 1.3. Delimitación Geográfica

La Localización geográfica del proyecto y área de influencia, se encuentra en: Provincia Pichincha, Cantón Quito, Sector Quitumbe, como se puede ver en la Figura No.1.



**Figura 1.** Localización Geográfica

**Fuente:** Turistica-y-ruidosa-ciudad-de-quito

### 1.4. Justificación e Importancia

Este estudio se puede justificar con la aplicación de técnicas y algoritmos de minería de datos que permitan entregar indicadores que generen patrones de



comportamiento para obtener la medición de los niveles de participación en el portal web y uso de medios tecnológicos, permitiendo a los ciudadanos y a las autoridades tener la información rápida y conformar una herramienta de toma de decisiones, análisis de la información y modelos predictivos que proyecten niveles de aceptación y satisfacción en la resolución de los requerimientos ciudadanos.

Adicionalmente permitirá contar con personal administrativo capacitado y generar procesos de gestión a partir de los datos obtenidos y procesados inteligentemente a través de la minería de datos y la inteligencia de negocios BI.

### **1.5. Planteamiento del problema**

Las Instituciones cada vez tienen más necesidad de interactuar con la ciudadanía y para esto se ha considerado, la variable uso de la tecnología, dentro de la participación ciudadana, y se generan requerimientos fundamentales en los gobiernos locales y el Municipio del Distrito Metropolitano, no es la excepción. En la problemática se ha identificado que las variables básicas de referencia se pueden integrar funcionalmente con el uso de la tecnología, dentro de la interacción entre gobierno local y comunidad, de manera que pueda ser, a través de procesos de aplicaciones tecnológicas, un recurso de fácil acceso a los ciudadanos y llegar a mejores niveles de colaboración, en la construcción de una mejor ciudad, conforme lo promueve la vigencia de la Ordenanza No. 102 que dice: “Promueve y regula el sistema metropolitano de participación ciudadana y control social”, los que se materializan en los mecanismos antes mencionados y señalados por la ley. Por esta razón se tiene la necesidad de aportar con técnicas de minería de datos que faciliten de manera técnica, el poder realizar procesos de toma de decisiones en base a patrones de comportamiento de la información.

### **1.6. Formulación del problema**

Inicialmente los gobiernos locales tienen un nivel de dificultad en la formulación del uso de la tecnología en la participación ciudadana dando como resultado una necesidad prioritaria de:

Falta de existencia de un sistema de minería de datos para ayudar en los niveles de medición del uso de la tecnología en relación a variables básicas de análisis y procesamiento de la información.

### **1.7. Objetivos General**

Generar procesos de análisis en el manejo básico de indicadores de participación ciudadana, orientados al uso de la tecnología, con soporte en sistemas de información geográfica como apoyo en la gestión comunitaria.

### **1.8. Objetivos Específicos**

- Evaluar la utilización y aprovechamiento de la integración de Sistemas de información e indicadores, orientadas al uso de la tecnología, así como del uso de sus recursos y materiales para el procesamiento de información.
- Considerar las Variables que se encuentran en el portal como referencia de la línea base.
- Definir los indicadores de Participación Ciudadana para cada uno de los procesos agrupándolas en categorías y subcategorías.
- Elaborar procesos de análisis de variables e indicadores como herramienta de tomas de decisiones.
- Aplicar algoritmos para elaborar patrones de comportamiento la relación con el uso de la tecnología como medio para incrementar la participación ciudadana.

## **CAPITULO II**

### **MARCO TEORICO**

#### **2.1. Antecedentes Investigativos**

La minería de datos emplea de forma sistemática diversas técnicas de análisis de datos en los procesos de toma de decisiones empresariales utilizando la información oculta en grandes bancos de datos, que diariamente se generan en la actividad económica, con posibilidad de aumentar el beneficio, pero también con graves riesgos para preservar la intimidad de las personas.

Los métodos de tratamiento de la información en la empresa se iniciaron hace muchos años con la automatización de los procesos repetitivos y administrativos. Los sistemas informáticos centralizados se difundieron en las décadas de los sesenta y setenta en las grandes corporaciones. La aparición de los mini-ordenadores permitió la incorporación en medianas empresas de procesos automatizados, y, finalmente la difusión masiva de los ordenadores personales en los ochenta y de las redes de comunicación generalizaron el uso de los procesos informáticos y obligaron a cambiar las estructuras centralizadas de los centros de proceso de datos. Las tecnologías de la información están orientadas hoy día, no sólo a los procesos de tratamiento administrativo, sino también hacia la gestión de datos y el soporte en los procesos de toma de decisiones. La difusión de redes de ordenadores, incluyendo los equipos personales, origina una descentralización de la información que dificulta la integración en su uso en la gestión de la empresa. Por otra parte, la aparición de nuevas herramientas está facilitando esta integración y uso más eficiente a través de dos tipos de desarrollos tecnológicos: los denominados Data Warehouse (DW) o almacén de datos, y Data Mining (DM) o minería de datos.

#### **2.2. Estado del Arte en la Minería de Datos**

##### **Estado del arte**

El estado del arte dentro de la investigación documental hace referencia al estudio del conocimiento acumulado, sea a través de texto escrito, publicaciones, contribuciones, documentos científicos dentro de un área específica del conocimiento” (Molina Nancy, 2005).

En el estado del arte se consideran quienes han investigado, en sentido de experticia mundial en diferentes áreas de la ciencia. Que tanto han investigado, partiendo de un objeto de estudio e investigación, para proporcionar nuevos aportes a la ciencia y el conocimiento. Se almacenan en bases de datos documentales, que pueden ser procesadas a través de revisiones sistemáticas de la literatura científica. Se considera que vacíos o necesidades o requerimientos nuevos generan a través de futuras investigaciones o trabajos futuros, y que son sugeridos tanto por los investigadores como por los revisores expertos, considerados como verdaderas autoridades de criterio y referencia de altos estándares a nivel internacional. En este contexto un círculo virtuoso se va generando en niveles de mayor o menor impacto, local, regional o mundial y de acuerdo a esos niveles de impactos se acceden a congresos donde las publicaciones y los investigadores de las publicaciones comparten y enriquecen el estado del arte.

Los pasos fundamentales para definir un estado del arte se basan en la contextualización de un problema de estudio o investigación, las delimitaciones, el material documental y los criterios de investigación de los investigadores. Luego la clasificación a través de una revisión sistemática de la información y sus fuentes de alto rigor científico, dentro de sus líneas y sub líneas de investigación. Luego la categorización donde la jerarquización y generación de claves para el tratamiento de la información de forma interna (documentación) y externa (interacción).

El estado del arte de la minería informática se ha diversificado, en diferentes dimensiones, así se obtienen los estados del arte, de la minería de datos, minería web, minería móvil, minería de procesos, minería de intención, minería de grafos y otras sub áreas de investigación.

La minería de datos se puede conceptualizar como una abstracción para la construcción de modelos que ajustados a los datos proporcionan un conocimiento, donde por un lado se elige el modelo y por otro se procede al ajuste final del modelo a los datos. Se tienen condicionantes que son los tipos de datos y el objetivo a obtener. (Riquelme, Ruiz, Sevilla, 2006).

En la relación modelo objetivo, la literatura sugiere un catálogo de diferentes modelos para orientarlos a diversos objetivos. El autor así sugiere a través de la ejemplificación que ante un problema de clasificación se podrían utilizar máquinas de vectores de soporte o también árboles de decisión. Y si el problema fuera de regresiones se sugiere aplicar árboles de regresiones o redes neuronales. Si el objeto central fuera de clustering se podría aplicar modelos jerárquicos o inter relacionados y adaptados. (Riquelme, Ruiz, Sevilla, 2006).

La fase de aprendizaje en la minería de datos representa un factor fundamental en sus procesos intrínsecos para ajustar un modelo hacia un problema particular. Así una red neuronal implica definir arquitecturas y ajustar valores en los pesos de sus conexiones. En rectas de regresión se aplican coeficientes, fijación de métricas para ajustes de modelos. Además, es importante el tratamiento de los niveles de incertidumbre en relación a técnicas como algoritmos evolutivos, búsquedas dispersas y búsquedas tabú. Igualmente, los sistemas de razonamiento se pueden aplicar diferentes lógicas, como lógica difusa, rough sets. Se hace referencia también a un conjunto de técnicas computacionales como referencia de contraparte, tales como la lógica borrosa, o los probabilísticos, evolutivos. (Riquelme, Ruiz, Sevilla, 2006).

También se consideran estudios comparativos de las diferentes metodologías de minería de datos, donde se hace análisis comparativos desde KDD (Knowledge Discovery in Databases) como primer modelo aceptado por la comunidad científica, y luego surgen otros modelos que proponen enfoques de tipo sistemático para llevar a cabo el proceso como son: SENNA, CRISPDM, Catalyst o P3TQ (Moine, Haedo, Buenos Aires, 2013).

### **2.3. Antecedentes Conceptuales y Referencias de Minería de Datos.**

Las bases de datos pueden clasificarse de varias maneras, de acuerdo al contexto que se esté manejando, o la utilidad de la misma.

Según la variabilidad de los datos almacenados:

### **2.3.1. Base de datos Estáticas:**

Estas son bases de datos de solo lectura, utilizadas primordialmente para almacenar datos históricos que posteriormente se pueden utilizar para estudiar el comportamiento de un conjunto de datos a través del tiempo, realizar proyecciones y tomar decisiones.

### **2.3.2. Bases De Datos Dinámicas:**

Estas son bases de datos donde la información almacenada se modifica con el tiempo, permitiendo operaciones como actualización, borrado y adición de datos, además de las operaciones fundamentales de consulta.

### **2.3.3. Modelos de Bases de Datos:**

Se clasifica por la función de las bases de datos, estas también se pueden clasificar de acuerdo a su modelo de administración de datos.

Algunos modelos con frecuencia utilizados en las bases de datos:

#### **Bases De Datos Jerárquicas:**

Éstas son bases de datos que, como su nombre indica, almacenan su información en una estructura jerárquica. En este modelo los datos se organizan en una forma similar a un árbol (visto al revés), en donde un nodo padre de información puede tener varios hijos. El nodo que no tiene padres es llamado raíz, y a los nodos que no tienen hijos se los conoce como hojas.

Las bases de datos jerárquicas son especialmente útiles en el caso de aplicaciones que manejan un gran volumen de información y datos muy compartidos permitiendo crear estructuras estables y de gran rendimiento.

Una de las principales limitaciones de este modelo es su incapacidad de representar eficientemente la redundancia de datos.

### **Base De Datos De Red:**

Éste es un modelo ligeramente distinto del jerárquico; su diferencia fundamental es la modificación del concepto de nodo: se permite que un mismo nodo tenga varios padres (posibilidad no permitida en el modelo jerárquico).

Fue una gran mejora con respecto al modelo jerárquico, ya que ofrecía una solución eficiente al problema de redundancia de datos; pero, aun así, la dificultad que significa administrar la información en una base de datos de red ha significado que sea un modelo utilizado en su mayoría por programadores más que por usuarios finales.

### **Bases De Datos Transaccionales:**

Son bases de datos cuyo único fin es el envío y recepción de datos a grandes velocidades, estas bases son muy poco comunes y están dirigidas por lo general al entorno de análisis de calidad, datos de producción e industrial, es importante entender que su fin único es recolectar y recuperar los datos a la mayor velocidad posible, por lo tanto la redundancia y duplicación de información no es un problema como con las demás bases de datos, por lo general para poderlas aprovechar al máximo permiten algún tipo de conectividad a bases de datos relacionales.

### **Bases De Datos Relacionales:**

Éste es el modelo utilizado en la actualidad para modelar problemas reales y administrar datos dinámicamente. Tras ser postulados sus fundamentos en 1970, de los laboratorios IBM en San José, no tardó en consolidarse como un nuevo paradigma en los modelos de base de datos. Su idea fundamental es el uso de "relaciones". Estas relaciones podrían considerarse en forma lógica como conjuntos de datos llamados "tuplas". Pese a que ésta es la teoría de las bases de datos relacionales creadas por Codd, la mayoría de las veces se conceptualiza de una manera más fácil de imaginar.

En este modelo, el lugar y la forma en que se almacenen los datos no tienen relevancia (a diferencia de otros modelos como el jerárquico y el de red). Esto tiene la considerable ventaja de que es más fácil de entender y de utilizar para un usuario esporádico de la base de datos. La información puede ser recuperada o almacenada

mediante "consultas" que ofrecen una amplia flexibilidad y poder para administrar la información.

El lenguaje más habitual para construir las consultas a bases de datos relacionales es SQL, Structured Query Language o Lenguaje Estructurado de Consultas, un estándar implementado por los principales motores o sistemas de gestión de bases de datos relacionales.

Durante su diseño, una base de datos relacional pasa por un proceso al que se le conoce como normalización de una base de datos

Durante los años 80 la aparición de la base produjo una revolución en los lenguajes de programación y sistemas de administración de datos. Aunque nunca debe olvidarse que base no utilizaba SQL como lenguaje base para su gestión.

### **Bases De Datos Multidimensionales:**

Son bases de datos ideadas para desarrollar aplicaciones muy concretas, como creación de cubos OLAP. Básicamente no se diferencian demasiado de las bases de datos relacionales (una tabla en una base de datos relacional podría serlo también en una base de datos multidimensional), la diferencia está más bien a nivel conceptual; en las bases de datos multidimensionales los campos o atributos de una tabla pueden ser de dos tipos, o bien representan dimensiones de la tabla, o bien representan métricas que se desean estudiar.

### **Bases De Datos Orientadas a Objetos:**

Este modelo, bastante reciente, y propio de los modelos informáticos enfocado a objetos, trata de almacenar en la base de datos los objetos completos (estado y comportamiento).

En bases de datos orientadas a objetos, los usuarios pueden definir operaciones sobre los datos como parte de la definición de la base de datos. Los programas de aplicación de los usuarios pueden operar sobre los datos invocando a dichas operaciones a través de sus nombres y argumentos, sea cual sea la forma en la que se



han implementado. Esto podría denominarse independencia entre programas y operaciones.

#### **2.3.4. Gestión de bases de datos distribuida (SGBD):**

La base de datos y el software SGBD pueden estar distribuidos en múltiples sitios conectados por una red.

Hay de dos tipos:

1. Distribuidos homogéneos: utilizan el mismo SGBD en múltiples sitios.
2. Distribuidos heterogéneos: Da lugar a los SGBD federados o sistemas multibase de datos en los que los SGBD participantes tienen cierto grado de autonomía local y tienen acceso a varias bases de datos autónomas preexistentes almacenados en los SGBD, muchos de estos emplean una arquitectura cliente-servidor.

#### **2.3.5. Potencialidad de las Bases de Datos:**

Los SIG se aplican principalmente en el análisis y resolución de problemáticas en la que interviene la relación entre la sociedad y el espacio geográfico en una gran cantidad de ciencias, por lo cual desde un comienzo fue considerada una herramienta multidisciplinaria.

A nivel computacional esta situación se resuelve mediante la creación de bases de datos gráficas (contienen las formas geométricas) y bases de datos alfanuméricas (contienen los atributos medidos en estas formas). Los SIG integran ambas bases de datos actuando como núcleo de un campo de gran amplitud denominado Geo informática.

KDD (Knowledge Discovery in Databases) KDD se denomina a los procesos no triviales para la identificación de patrones a partir de datos. Sus atributos o características principales se basan en: datos válidos, datos novedosos, datos útiles parcialmente, y datos comprensibles.

Para la obtención del conocimiento se parte de los datos, a través de la selección en fuentes de información, se analiza, pre procesa, y el dato procesado es sujeto a transformación, se lo somete a la minería de datos, para obtener patrones y modelos, y con la interpretación analítica de la información y fases de evaluación se obtiene el conocimiento.

El tipo de conocimiento, entre otros, puede ser descriptivo o predictivo. Es descriptivo, cuando muestra la relación entre las variables del objeto de estudio, para mejorar los modelos, generar reglas, relaciones y correlaciones. Es predictivo, cuando se basa o fundamenta en el modelo que gobierna un sistema de información para proyectar patrones de comportamiento de hechos o procesos o resultados futuros. Sus soluciones se basan en redes neuronales y algoritmos para ofrecer resultados o visiones que orientan el análisis inteligente de la información y los resultados estadísticos por diferentes tipos y categorías de datos.

Las fases del proceso KDD inician con la obtención de datos, hasta la obtención del conocimiento y la toma de decisiones que contempla 5 pasos, como se puede ver en la Figura No.2.

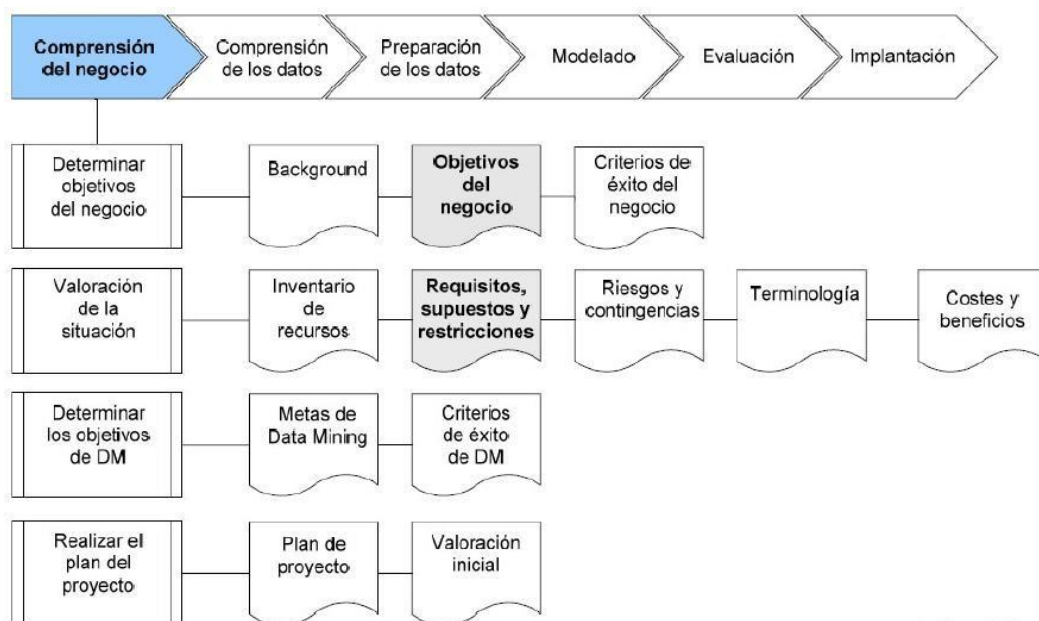
Paso 1: integración y recopilación de información.

Paso 2: selección, limpieza y transformación de la información

Paso 3: minería de datos o data Mining

Paso 4: evaluación e interpretación de resultados

Paso 5: difusión, publicación y uso de la información y el conocimiento para la toma de decisiones



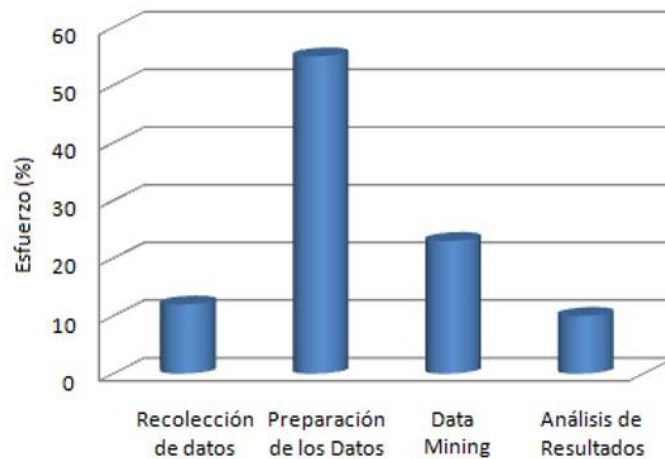
**Figura 2.** Fases del proceso KDD

**Fuente:** Compression-de-negocio-fig-1

#### Fases del proceso de KDD

Las fases del proceso de KDD dependen de un objetivo propuesto o fijado en base al objeto de estudio o análisis en referencia a un parámetro de comparación, como se puede observar en la Figura No.3.

- Recolección de información y datos para el manejo, análisis, procesamiento y administración inteligente de la información
- Preparación de los datos: consiste en la optimización de datos, de la información guardada o almacenada, donde se toma como referencia los datos faltantes, valores extremos, inconsistencias y ruidos. Como actividades a realizar están la limpieza de datos (atributos, variables), de inconsistencias y la resolución de outsiders. Así como la transformación de datos, tales como la discretización.
- Minería o data Mining: se aplican técnicas y métodos de minería de datos
- Análisis de resultados para la toma de decisiones



**Figura 3.** Esfuerzo requerido para el desarrollo de cada una de las fases

**Fuente:** prediccionpaciente.htm

Variables: tiene su origen en el latín variable y se caracteriza por ser algo que se puede medir y que a su vez varía en base a un objeto de estudio. Se puede decir que es un elemento de un conjunto dado, denominando conjunto universal o universo y cada elemento de dicho conjunto resulta ser un valor de la variable.

Las variables se las puede clasificar de algunas formas. Según el tipo de variable, pueden ser independientes y dependientes. Según las características de la información pueden ser cuantitativas (escala nominal) o cualitativas (escala de intervalo). Según el tipo o rango de valores pueden ser discretas (valores finitos, con números enteros) o continuas (valores infinitos, con infinitos decimales): Según las escalas pueden ser nominales, ordinales y de intervalo.

Indicadores constituyen una medición cuantificable de variables o condicionantes, a través de las cuales es posible entender, comprender o interpretar la realidad a través de un fenómeno particular durante un período de tiempo, donde se analizan sus procesos y sus relaciones pueden ser cambiantes en el tiempo y determinar e identificar su evolución.

### **2.3.6. Minería de Datos:**

Es el análisis de archivos y bitácoras de transacciones, trabaja a nivel del conocimiento con el fin de descubrir patrones, relaciones, reglas, asociaciones o incluso excepciones útiles para la toma de decisiones. La MD está muy ligada a los Data Warehouse.

Aplicar y convertir la información obtenida en conocimiento que ayude en la toma de decisiones estratégicas y operacionales.

Generar y mantener actividades de investigación, desarrollo y consultoría en el área de Data Mining, como preámbulo a una actividad investigadora en Estadística Aplicada.

Introducir conceptos de Inteligencia de Negocios y las técnicas relacionadas incluyendo Data Warehousing, Data Mining y Transacciones de Procesos On line (OLTP).

Explorar procesos, contenidos y contextos relativos a las técnicas de decisión en marketing.

Mejora de procesos a partir de la inteligencia de negocios.

El proceso más importante es el de Minería de Datos o DM (Data Mining), es el proceso automático para el descubrimiento de información útil en grandes cantidades de datos.

Este proceso es un campo multidisciplinario, en el que se pretende predecir resultados y/o descubrir relaciones entre los diferentes datos. Las diferentes tareas que puede realizar la DM son:

- **Clasificación:** Mediante la clasificación se busca encontrar un modelo que pueda predecir el comportamiento de una variable a partir de sus características.
- **Análisis de Asociaciones:** Estas técnicas pretenden sacar patrones de las relaciones que hay entre diferentes rasgos de los datos.

- Detección de Anomalías: El objetivo de la detección de anomalías es encontrar aquellos elementos o características que son significativamente diferentes del resto de los datos.

Algunas de las áreas que más se pueden beneficiar de la minería de datos son las siguientes:

Ámbitos financieros y de negocios: índices de producción y coste, marketing, detección de fraudes, descubrir perfiles de clientes.

- Salud: modelos de diagnóstico a partir de información almacenada en hospitales, gestión de tratamientos, comprobación de la adecuación de tratamientos.
- Sistemas informáticos: control del sistema y detección de ataques.
- Ciencia: observaciones astronómicas, genómica, análisis de datos biológicos.

Business Intelligence (BI): Según Howard Dresner, (1989). Conjunto de conceptos y métodos para mejorar la toma de decisiones en los negocios, utilizando sistemas de apoyo basados en hechos. La inteligencia de negocio actúa como un factor estratégico para una empresa u organización, generando una potencial ventaja competitiva, que no es otra que proporcionar información privilegiada para responder a los problemas de negocio: entrada a nuevos mercados, promociones u ofertas de productos, eliminación de islas de información, control financiero, optimización de costes, planificación de la producción, análisis de perfiles de clientes, rentabilidad de un producto concreto.

Los principales productos de Business Intelligence que existen hoy en día son:

- Cuadro de mando Integral (CMI)
- Sistemas de Soporte a la Decisión (DSS)
- Sistemas de Información Ejecutiva (EIS)

Por otro lado, los principales componentes de orígenes de datos en el Business Intelligence que existen en la actualidad son:

- Datamart

- Datawarehouse

Los sistemas y componentes del BI se diferencian de los sistemas operacionales en que están optimizados para preguntar y divulgar sobre datos. Esto significa típicamente que, en un datawarehouse, los datos están desnormalizados para apoyar consultas de alto rendimiento, mientras que en los sistemas operacionales suelen encontrarse normalizados para apoyar operaciones continuas de inserción, modificación y borrado de datos. En este sentido, los procesos ETL (extracción, transformación y carga), que nutren los sistemas BI, tienen que traducir de uno o varios sistemas operacionales normalizados e independientes a un único sistema desnormalizado, cuyos datos estén completamente integrados.

## **2.4. Herramientas de Software**

Las siguientes herramientas se pueden aplicar directamente para Minería de Datos.

### **2.4.1. Weka**

(WEKA: Waikato Environment for Knowledge Analysis) es un conjunto de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos o bien se pueden aplicar directamente a un conjunto de datos o llamadas de su propio código Java; contiene herramientas para el procesamiento previo de datos, clasificación, regresión, clustering, reglas de asociación, y la visualización.

### **2.4.2. Rapid Miner: (YALE =yet another learning Environment)**

Es un entorno para aprendizaje mecánico y experimentos de data Mining, utilizada tanto en investigación como en tareas de día a día. Produce sus resultados en archivos XML y cuentan con la interface gráfica del mismo programa. Proveen más de 500 operadores para los principales procesos de aprendizaje en máquina y al tiempo combina esquemas y atributos de evaluación.

### **2.4.3. Orange:**

Es un componente de minería de datos y también es un software de aprendizaje de máquina, que permite una programación visual, rápida y versátil para un análisis exploratorio de datos. También permite pre procesamiento, filtros de información, modelación de datos, evaluación de modelos y técnicas de exploración.

#### **2.4.4. KNIME: Konstanz Information Miner**

Es un software de integración de datos amigable, intuitivo y fácil de usar, que permite el procesamiento, análisis y exploración de datos, desde su plataforma. Permite crear visualmente flujos de datos, ejecutar análisis selectivamente, estudiar los resultados, modelar y generar vistas interactivas, para facilitar la toma de decisiones a nivel gerencial.

#### **2.4.5. JHepWork:**

Diseñado para científicos, ingenieros y estudiantes, es una herramienta gratuita y de uso libre, que permite el análisis de datos mediante la creación de un entorno comprensible, amigable y adaptable a programas comerciales.

#### **2.4.6. Data Mining Add-In for Excel:**

Es una herramienta que se puede instalar a la suite de Office, que incluye herramientas de análisis de tablas y una función adicional llamada Data Mining Client, pueden obtener patrones y tendencias que existen en datos complejos, visualizar los patrones de los gráficos para obtener presentaciones de análisis de negocios.

### **2.5. Metodologías para Minería de Datos**

Son diversos los modelos de proceso que han sido propuestos para el desarrollo de proyectos de Minería de Datos tales como:

#### **2.5.1. SEMMA**

SEMMA (Sample, Explore, Modify, Model, Assess) [SAS, 2003], la metodología SEMMA se encuentra enfocada especialmente en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando. Fue propuesta especialmente para trabajar con el software de minería de datos de la compañía SAS. Este producto organiza sus herramientas (llamadas “nodos”) en base a las distintas fases que componen la metodología. Es decir, el software proporciona un conjunto de herramientas especiales para la etapa de muestreo, otras para la etapa de exploración, y así sucesivamente. Sin embargo, el usuario podría hacer uso del



mismo siguiendo cualquier otra metodología de minería de datos (como CRISP-DM por ejemplo).

### **2.5.2. DMAMC**

DMAMC (Definir, Medir, Analizar, Mejorar, Controlar) [Sixsigma, 2005], es un acrónimo (por sus siglas en inglés: Define, Measure, Analyze, Improve, Control) de los pasos de la metodología: Definir, Medir, Analizar, Mejorar y Controlar.

Es una herramienta de la metodología Seis Sigma, enfocada en la mejora incremental de procesos existentes.

La herramienta es una estrategia de calidad basada en estadística, que da mucha importancia a la recolección de información y a la veracidad de los datos como base de una mejora.

Cada paso en la metodología se enfoca en obtener los mejores resultados posibles para minimizar la posibilidad de error.

### **2.5.3. CRISP-DM**

CRISP-DM, creada por el grupo de empresas SPSS, NCR y Daimler Chrysler en el año 2000, es actualmente la guía de referencia más utilizada en el desarrollo de proyectos de minería de datos. Estructura el proceso en seis fases: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implantación. La sucesión de fases, no es necesariamente rígida. Cada fase se descompone en varias tareas generales de segundo nivel.

### **2.5.4. CATALYST**

CATALYST En el año 2003, Dorian Pyle propone en su libro “Business modelling and data Mining” [26] una metodología para el proceso de extracción de conocimiento en bases de datos llamada “Catalyst”. A pesar de ser una metodología muy completa, actualmente no tiene tanto éxito y difusión como CRISP-DM. Pyle recomienda que el proceso de minería de datos siempre debería colaborar con una situación organizacional, como un problema u oportunidad. Recomienda no trabajar directamente con los datos sino establecer de antemano la problemática que se

aborda, el personal involucrado y las expectativas y necesidades de los usuarios. Este punto resulta de gran importancia para justificar la realización del proyecto, ya que difícilmente una organización compre una herramienta si no sabe la función que cumplirá. Para proyectos donde el problema u oportunidad de negocio no está definido, se recomienda comenzar analizando las relaciones P3TQ - Product (Producto), Place (Lugar), Price (Precio), Time (Tiempo) y Quantity (Cantidad) - que existen en la cadena de valor organizacional. Las relaciones P3TQ se refieren a tener el producto correcto, en el lugar adecuado, en el momento adecuado, en la cantidad correcta y con el precio correcto.

## **2.6. Cubos de Información**

### **2.6.1. OLAP**

OLAP, OnLine Analytical Processing o procesamiento Analítico en Línea, término acuñado por Edgar Frank Codd de EF Codd & Associates, encargado por Arbor Software (en la actualidad Hyperion Solutions)

Los cubos son subconjuntos de datos de un almacén de datos, organizado y sumariado dentro de una estructura multidimensional. Los datos se sumarizan de acuerdo a factores de negocio seleccionados, proveyendo el mecanismo para la rápida y uniforme tiempo de respuesta de las complejas consultas. La definición del cubo, es el primero de tres pasos en la creación de un cubo. Los otros pasos son, el especificar la estrategia de sumarización diseñando las agregaciones (elementos precalculados de datos), y la carga del cubo para procesarlo. Para definir un cubo, seleccione una tabla objetivo y seleccione las medidas (columnas numéricas de interés a los usuarios del cubo) dentro de esta tabla. Entonces seleccione las dimensiones, cada compuesta de una o más columnas de otra tabla. Las dimensiones proveen la descripción categórica por el cual las medidas son separadas para su análisis por los usuarios del cubo.

En el mundo de las soluciones para Business Intelligence, una de las herramientas más utilizadas por las empresas son las aplicaciones OLAP, ya que las misma han sido creadas en función a bases de datos multidimensionales, que permiten procesar

grandes volúmenes de información, en campos bien definidos, y con un acceso inmediato a los datos para su consulta y posterior análisis.

### **2.6.2. OLTP**

OLTP, es la sigla en inglés de Procesamiento de Transacciones En línea (OnLine Transaction Processing) es un tipo de procesamiento que facilita y administra aplicaciones transaccionales, usualmente para entrada de datos y recuperación del procesamiento de transacciones (gestor transaccional). Los paquetes de software para OLTP se basan en la arquitectura cliente-servidor ya que suelen ser utilizados por empresas con una red informática distribuida.

MOLAP, Los datos fuente del cubo son almacenados junto con sus agregaciones (Sumarizaciones) en una estructura multidimensional de alto rendimiento. El almacenaje de MOLAP, provee excelente rendimiento y compresión de datos. Como se dice, todo va en el cubo. Tiene el mejor tiempo de respuesta, dependiendo solo en el porcentaje y diseño de las agregaciones del cubo. En general este método, es muy apropiado para cubos con uso frecuente por su rápida respuesta.

### **2.6.3. ROLAP**

ROLAP, Toda la información del cubo, sus datos, su agregación, sumas etc., son almacenados en una base de datos relacional. ROLAP no almacena copia de la base de datos, acceda a las tablas originales cuando necesita responder a preguntas, es generalmente, mucho más lenta que las otras dos estrategias de almacenaje. Típicamente ROLAP se usa, para largos conjuntos de datos que no son frecuentemente buscados, tales como datos históricos de los años más recientes.

### **2.6.4. HOLAP**

HOLAP, (Hybrid OLAP) combina atributos de MOLAP y ROLAP, la agregación de datos es almacenada en una estructura multidimensional usada por MOLAP, y la base de datos fuentes, en una base de datos relacional. Para procedimientos de búsqueda que accedan datos sumarizados, HOLAP es equivalente a MOLAP, por el contrario, si estos procesos accederán datos fuentes como los Drill Down, estos deben

de buscar los datos en la base de datos relacional y esto no es tan rápido comparado a si los datos estuvieran almacenados en una estructura MOLAP.

Los cubos almacenados en HOLAP, son más pequeños que los MOLAP y responden más rápidos que los ROLAP.

## **2.7. Definición de técnicas de minería de datos**

La definición formal de data Mining o minería de datos sería, la extracción no trivial de información implícita, previamente desconocida y potencialmente útil a partir de datos. Otra manera de definirlo podría ser: la exploración y el análisis -por medios automáticos o semiautomáticos- de grandes cantidades de datos con el fin de descubrir patrones con significado.

### **2.7.1. Redes Neuronales**

Esta técnica de inteligencia artificial, en los últimos años se ha convertido en uno de los instrumentos de uso frecuente para detectar categorías comunes en los datos, debido a que son capaces de detectar y aprender complejos patrones, y características de los datos.

Una de las principales características de las redes neuronales, es que son capaces de trabajar con datos incompletos e incluso paradójicos, que dependiendo del problema puede resultar una ventaja o un inconveniente. Además, esta técnica posee dos formas de aprendizaje: supervisado y no supervisado.

### **2.7.2. Árboles de Decisión**

Esta técnica se encuentra dentro de una metodología de aprendizaje supervisado. Su representación es en forma de árbol en donde cada nodo es una decisión, los cuales a su vez generan reglas para la clasificación de un conjunto de datos.

Los árboles de decisión son fáciles de usar, admiten atributos discretos y continuos, tratan bien los atributos no significativos y los valores faltantes. Su principal ventaja es la facilidad de interpretación.

### **2.7.3. Algoritmos Genéricos**

Los algoritmos genéticos imitan la evolución de las especies mediante la mutación, reproducción y selección, como también proporcionan programas y optimizaciones que pueden ser usadas en la construcción y entrenamiento de otras estructuras como es el caso de las redes neuronales. Además, los algoritmos genéticos son inspirados en el principio de la supervivencia de los más aptos.

### **2.7.4. Clustering**

Agrupan datos dentro de un número de clases preestablecidas o no, partiendo de criterios de distancia o similitud, de manera que las clases sean similares entre sí y distintas con las otras clases. Su utilización ha proporcionado significativos resultados en lo que respecta a los clasificadores o reconocedores de patrones, como en el modelado de sistemas. Este método debido a su naturaleza flexible se puede combinar fácilmente con otro tipo de técnica de minería de datos, dando como resultado un sistema híbrido.

Un problema relacionado con el análisis de clúster es la selección de factores en tareas de clasificación, debido a que no todas las variables tienen la misma importancia a la hora de agrupar los objetos. Otro problema de gran importancia y que actualmente despierta un gran interés es la fusión de conocimiento, ya que existen múltiples fuentes de información sobre un mismo tema, los cuales no utilizan una categorización homogénea de los objetos. Para poder solucionar estos inconvenientes es necesario fusionar la información a la hora de recopilar, comparar o resumir los datos.

## **2.8. Metodología para llevar a cabo una encuesta**

Las encuestas, como instrumentos sirven para coleccionar información de la realidad, son parte del diagnóstico y no son un fin en sí mismo. En el caso de las encuestas de percepción, el método probado y más efectivo consiste en llevar a cabo encuestas directas y simultáneas a las diferentes partes involucradas. Para ello, se pueden tener en cuenta los siete pasos principales del ciclo o proceso de implementación de una encuesta:

- Definición del objetivo, que significa responder a la pregunta ¿qué se desea investigar y con qué propósito? Es el ¿qué y el para qué? De esto dependerá en gran parte los siguientes pasos a seguir
- Diseño muestral, el cual dependerá del universo seleccionado y la cobertura que se pretende. Implica responder a la pregunta de ¿quién proveerá la información que se necesita?
- Diseño del instrumento, es el cuestionario a utilizar; lo que implica responder a la pregunta de ¿cómo se captará la información? Además, se define el tipo de preguntas, si serán cerradas o abiertas, debe utilizarse un lenguaje claro, sencillo y directo, y adaptado a la idiosincrasia del encuestado.
- Ejecución de la encuesta, el encuestador debe tener credibilidad, habilidad para realizar las entrevistas y contar con un buen respaldo institucional que permita reducir los márgenes de posibles resistencias.
- Procesamiento de la información colectada, estará en función a la metodología definida previamente, donde debe prevalecer un enfoque objetivo e imparcial a fin de evitar problemas de manipulación o sesgo en este trabajo, base del informe final.
- Análisis de los resultados de las encuestas, se requiere independencia del poder de la institución que encomendó y/o financió el trabajo de la encuesta. No se debe ocultar información “negativa”. Se debe garantizar la transparencia, objetividad y consistencia técnica en el informe final.
- Difusión del resultado, todos los que han participado y contribuido con la encuesta deben tener acceso a los resultados de la misma. La divulgación debe ser amplia, clara, educativa y de fácil acceso. Esto generará credibilidad para acciones posteriores, como la realización de talleres de diagnósticos participativos y luego la implementación del plan de acción.

## **CAPITULO III**

### **ANALISIS DEL CASO DE ESTUDIO**

Actualmente los gobiernos locales tienen un nivel de dificultad en la formulación del uso de la tecnología en la participación ciudadana dando como resultado una necesidad prioritaria de un sistema de minería de datos para ayudar en los niveles de medición del uso de la tecnología en relación a variables básicas de análisis y procesamiento de la información.

Se ha tomado como caso de estudio el sector de Quitumbe, el mismo que nos permitirá determinar cuál es el porcentaje de participación ciudadana tomando en cuenta procesos de análisis de variables e indicadores como herramienta de tomas de decisiones.

Además, se aplicará algoritmos para elaborar patrones de comportamiento, para cada una de las variables directas que son Nivel de Educación, Nivel económico, Género, Edad en relación a las variables indirectas que son Uso de la Tecnología y Tipo de Tecnología.

Tomando en cuenta que el Nivel de Educación y Nivel Económico son factores preponderantes en el nivel de participación ciudadana entre el gobierno local y la comunidad utilizando la Tecnología.

Como complemento tenemos las variables edad y género que determinan mayor o menor apertura al uso de la tecnología y colaboración en la opinión ciudadana.

#### **3.1. Aplicación de la Metodología CRISP-DM en el caso de Estudio**

El estándar incluye un modelo y una guía, estructurados en seis fases, algunas de estas fases son bidireccionales, lo que significa que permitirán revisar parcial o totalmente las fases anteriores.

La metodología CRISP-DM hace referencia a niveles de abstracción organizados jerárquicamente, y que involucran la consecución de tareas que van desde aspectos más generales hasta los más específicos. El nivel de mayor generalidad se organiza

en seis fases, donde cada una de ellas tiene tareas de segundo nivel o sub fases. Es decir que las tareas generales se proyectan hacia las tareas específicas. En un segundo nivel se realiza la limpieza de datos. Luego en su tercer nivel se desarrolla para un caso específico, como lo propone el autor una referencia de limpieza a datos de tipo numérico o limpieza categórica de datos. En un cuarto nivel se recoge un conjunto de acciones, también decisiones y por supuesto, resultados de la explotación del proyecto de su información específica.

De modo que existe una inter acción de proyección entre un modelo genérico y un modelo específico, desde el cual las fases se van desglosando en sus tareas generales, específicas hasta las instancias de los procesos. (Britos Paola, La Plata, 2008).

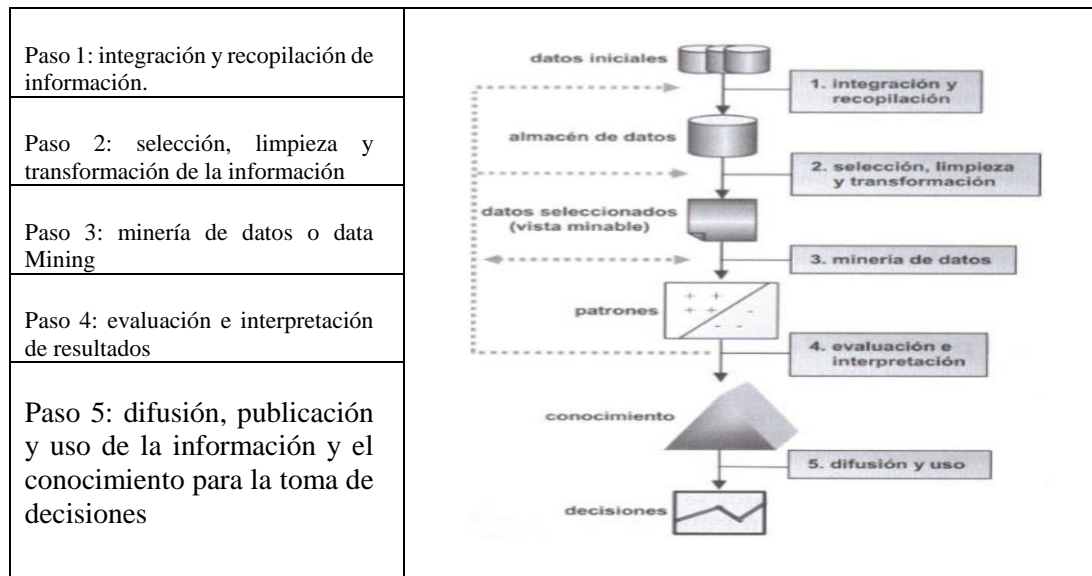
Se parte desde el objeto de estudio para los procesos de aplicación de la metodología, para esto se puede orientar en secuencias del proceso de investigación:

La sistematización del proceso de minería de datos es un punto importante para la planificación y ejecución de este tipo de proyecto. Algunas organizaciones implementan el proceso KDD, mientras que otras aplican un estándar más específico como CRISP-DM. Si la organización ha adquirido productos de la empresa SAS, tiene a su disposición una metodología especialmente desarrollada para los mismos, la metodología SEMMA. Por otro lado, la metodología Catalyst (conocida como P3TQ) está ganando cada vez mayor popularidad debido a su completitud y flexibilidad para adaptarse en distintos escenarios.

Como se puede observar en la Figura No.4, CRISP-DM se ha convertido en la metodología más utilizada, según un estudio publicado en el año 2007 por la comunidad KDnuggets (Data Mining Community's Top Resource).

Algunos modelos profundizan en mayor detalle sobre las tareas y actividades a ejecutar en cada etapa del proceso de minería de datos (como CRISP-DM), mientras que otros proveen sólo una guía general del trabajo a realizar en cada fase (como el proceso KDD o SEMMA).





**Figura 4.** Fases de la Metodología

**Fuente:** Documento Completo.pdf

La metodología SEMMA se encuentra enfocada especialmente en aspectos técnicos, excluyendo actividades de análisis y comprensión del problema que se está abordando. Fue propuesta especialmente para trabajar con el software de minería de datos de la compañía SAS.

La metodología Catalyst, conocida como P3TQ (Product, Place, Price, Time, Quantity), fue propuesta por Dorian Pyle en el año 2003. Esta metodología plantea la formulación de dos modelos: el Modelo de Negocio y el Modelo de Explotación de Información. El Modelo de Negocio (MII), proporciona una guía de pasos para identificar un problema de negocio (o la oportunidad del mismo) y los requerimientos reales de la organización. Contempla diferentes ámbitos para el proyecto de minería de datos, explicitando acciones específicas según el escenario desde el cual se parte. Para proyectos donde el problema u oportunidad de negocio no está definido, se recomienda comenzar analizando las relaciones P3TQ que existen en la cadena de valor organizacional, es decir, aquellas relaciones precio/lugar/producto/tiempo/cantidad que son importantes para la empresa. El Modelo de Explotación de Información (MIII), proporciona una guía paso para la construcción y ejecución de modelos de minería de datos a partir del Modelo de

Negocio (MII). El foco que la metodología Catalyst propone en su Modelo de Negocio sobre la cadena de valor organizacional, hizo que sea difundida en la comunidad científica como metodología “P3TQ”, aunque ésta no sea su denominación original. La metodología Catalyst, en sus dos modelos, está compuesta por una serie de pasos llamados “boxes”. El concepto es que luego de llevar a cabo una acción, se deben evaluar los resultados y determinar cuál es el próximo paso (box) a seguir. La secuencia y la interacción entre los distintos pasos permiten una flexibilidad muy grande, y una amplia variedad de caminos posibles, como podemos ver en la Figura No.5, la comparación de las metodologías.

**Tabla 1:**  
Fases del Proceso de Minería de Datos

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

**Fuente:** Documento\_completo\_metodología.pdf

## **3.2. Comprensión del negocio**

El presente estudio está centrado en el enfoque cuantitativo y cualitativo realizado al Sector de Quitumbe, dicho nos permitirán analizar indicadores que pueden ser implementadas en el portal Web del Municipio de Quito, para obtener el diferente punto de vista de los ciudadanos a través de la medición de las experiencias y las expectativas de los pobladores visualizadas en datos reales que ayuden a asumir con claridad la problemática.

### **3.2.1. Objetivo del Negocio**

Nuestra investigación se apoya en encuestas con información que será de utilidad para el cumplimiento de los objetivos de la investigación. En nuestro caso se realizará el estudio en los barrios del Sector Quitumbe para tener un contacto directo con la realidad y obtener información que será de utilidad para determinar la utilidad del portal web y su interacción con la ciudadanía.

También se realizará análisis estadísticos, basados en las encuestas realizadas que serán de gran ayuda para la investigación y así poder determinar el nivel de satisfacción de la ciudadanía, utilizando medios tecnológicos.

### **3.2.2. Evaluación del Negocio**

El Municipio de Quito, el área de ordenamiento Territorial es consciente de las necesidades de la ciudadanía, el tiempo que se demoran en solventar un requerimiento y la satisfacción de los mismos; ya que la relación entre costos y tiempo es grande.

Teniendo en cuenta estas necesidades nuestra investigación se centra en los indicadores que nos permitirán determinar mediante el uso de la tecnología la participación y satisfacción de la ciudadanía con el portal web.

La institución cuenta con información suficiente del barrio Quitumbe, que nos ha sido provista para realizar nuestro estudio.

### 3.2.3. Objetivos de la Minería de Datos

- Verificar la existencia de la relación entre las variables que se determinó en la encuesta relacionada con los medios tecnológicos y la ciudadanía.
- Analizar el modelo predictivo, que nos ayudara a saber la correlación de los indicadores, que intervienen en el estudio del sector Quitumbe.
- Además aproximar la relación de dependencia entre una variable dependiente y analizar el comportamiento de los indicadores, utilizando un modelo de Regresión Lineal.

### 3.2.4. Realizar el Plan del Proyecto

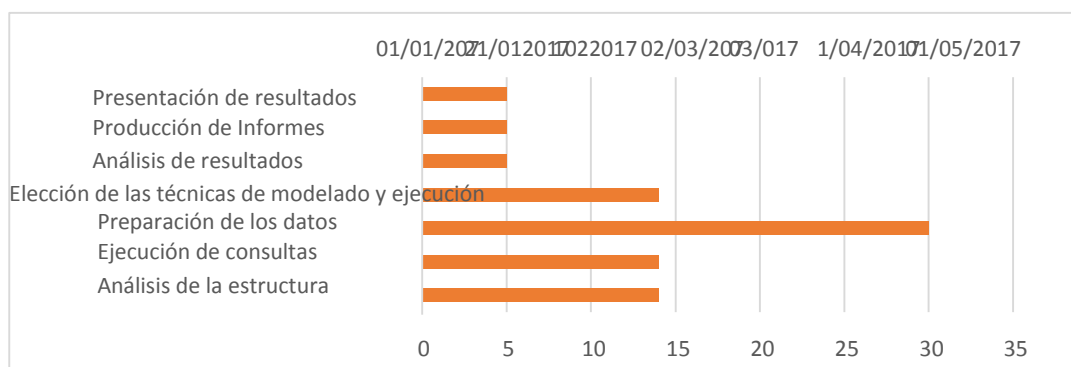
El proyecto se dividirá en las siguientes etapas para facilitar su organización y estimar el tiempo de realización del mismo, podemos ver en la Figura No.6 y Figura No.7; que están a continuación.

**Tabla 2:**  
Actividad del Proyecto

ACTIVIDAD	INICIO	DURACION	FIN
Análisis de la estructura	01/01/2017	14	20/01/2017
Ejecución de consultas	22/01/2017	14	10/02/2017
Preparación de los datos	13/02/2017	30	03/03/2017
Elección de las técnicas de modelado y ejecución	06/03/2017	14	13/03/2017
Análisis de resultados	14/03/2017	5	20/03/2017
Producción de Informes	21/03/2017	5	07/04/2017
Presentación de resultados	10/04/2017	5	24/04/2017

**Fuente:** Autora

**Figura 5:** Encuesta relacionada con uso de la Tecnología



**Fuente:** Autora

### 3.2.5. Evaluación Inicial de Herramientas y Técnicas

**Encuesta.** - Dirigida a los habitantes del Sector de Quitumbe, cantón Quito, cuyo instrumento es el cuestionario, elaborado con preguntas que permitieron recolectar la información sobre las variables de estudio.

**Validez y confiabilidad.** - La validez de los instrumentos se determina por la técnica llamada “Juicio de Expertos”, mientras que su confiabilidad será por medio del software RapidMiner, con la muestra total de pobladores de Quitumbe.

Las herramientas que se va a utilizar para llevar a cabo este proyecto de minería de datos son: Rapid Miner, Excel y SQL Server ya que se adaptan bien a la metodología que estamos empleando. Se facilita la lectura de los datos de la base de datos de SQL Server y así poder realizar nuestro análisis de la información.

### 3.3. Comprensión de datos

Se realizó la recolección de datos para establecer un primer análisis a cada uno de los indicadores que intervienen en el estudio, de esta manera relacionar las variables e identificar el impacto en el portal.

La metodología CRISP-DM nos permite establecer un primer contacto con el problema, familiarizarse con los datos y averiguar su calidad.

### 3.3.1. Recolección de los datos

Los datos utilizados en este proyecto son datos que se obtuvieron mediante las encuestas realizadas a la muestra establecida en el sector de Quitumbe donde se establecieron varios indicadores:

- Género
- Edad
- Nivel de Instrucción
- Nivel Económico

La información solo estará disponible para el análisis de los datos. Con esta información se podrá obtener la tendencia en base a un modelo y determinar el nivel de satisfacción de los ciudadanos e interacción con el portal. Debido a una muestra robusta es necesario hacer un trabajo de minería de datos con éxito. A continuación, la encuesta como técnica de recolección de información, la que se complementa con la observación, entrevistas, y el trabajo de campo.

La población de Quitumbe tiene un total de 32.645, como podemos observar en la Tabla No.1.

**Tabla 3:**  
Total Pobladores de la Sector Quitumbe

QUITUMBE	Población	Total población
	PUEBLO SOLO PUEBLO	4.796,00
	EJERCITO NACIONAL	4.141,00
	SALVADOR ALLENDE	4.248,00
	EL VERGEL	3.111,00
	NUEVOS HORIZONTES DEL SUR	1.445,00
	PUEBLO UNIDO	2.129,00
	QUITUMBE	6.600,00
	QUITUMBE	6.175,00
	<b>Tamaño Total Muestra</b>	<b>32.645,00</b>

**Fuente:** Autora

La encuesta realizada al sector de Quitumbe. **Ver Anexo 1.**

**Población.** - El universo o población de la presente investigación está compuesto por todas las personas pertenecientes al sector Quitumbe de los barrios que a continuación se describen:

- Pueblo solo pueblo
- Ejército Nacional
- Salvador Allende
- El Vergel
- Nuevos Horizontes del Sur
- Pueblo Unido
- Quitumbe

Los mismos que participan del estudio, que nos permitirá analizar el porcentaje de satisfacción de los usuarios con respecto a los servicios obtenidos por el municipio al igual que su participación con medios tecnológicos.

**Muestra.** - Es una parte o porción de la población estadística, a la que se aplicarán las encuestas. En este caso, al tener un universo poblacional mayor a 1000 individuos, se aplican ciertas parroquias, como podemos observar en la Tabla No.2.

Para nuestro cálculo utilizaremos la siguiente fórmula:

$$n = \frac{Z^2 \sigma^2 N}{e^2 (N - 1) + Z^2 \sigma^2}$$

Donde:

n = el tamaño de la muestra.

N = tamaño de la población.

Desviación estándar de la población que, generalmente cuando no se tiene su valor, suele utilizarse un valor constante de 0,5.

$Z$  = Valor obtenido mediante niveles de confianza. Es un valor constante que, si no se tiene su valor, se lo toma en relación al 95% de confianza equivale a 1,65 (como más usual) o en relación al 99% de confianza equivale 2,58, para la distribución normal estándar

$e$  = Límite aceptable de error muestral que, generalmente cuando no se tiene su valor, suele utilizarse un valor que varía entre el 1% (0,01) y 9% (0,09), (valor que queda a criterio del encuestador)

Por lo que en nuestros cálculos obtendríamos:

$$N = 32.645$$

$$Z = 1,96$$

$$e = 0,05$$

Reemplazando valores de la fórmula se tiene:

Los cálculos se realizaron también en Excel, como podemos ver en la Tabla No.4.

**Tabla 4:**  
Cálculo de la Muestra

E	0,05	
N	32645	
$\Sigma$	0,5	
Confianza	95	
Área a La izquierda de $-z=-1.96$	0.025	según tabla de distribución Normal Estándar
Z	1,96	
n	379,70336	

**Fuente:** Autora

**Descripción del Universo Investigado**



En todo este proceso es importante la aplicación tanto de métodos (inducción, deducción, análisis, síntesis, abstracción, etc.) como técnicas (encuesta, entrevista, observación, consulta documental, consulta de expertos, trabajo de campo) de la investigación del objeto de aprendizaje de saberes.

Las técnicas de recolección tradicional de datos aplicadas, se basan en la observación, el trabajo de campo y las encuestas y entrevistas procesadas dentro de los procesos de recolección de datos y validación de la información, como se puede observar en la Tabla No.4; las Variables de estudio.

**Tabla 5:**  
Plan de Recolección y Técnicas

Para qué?	Para alcanzar los objetivos de la investigación
Pobladores de la Sector Quitumbe ¿Sobre qué aspectos?	Manejo de indicadores de participación ciudadana
¿Sobre qué aspectos?	Participación ciudadana, en relación con el uso de la tecnología, caso de estudio en el sector Quitumbe.
¿Quién? ¿Quiénes?	Investigadora
¿Qué técnicas de recolección?	Encuesta, direccionada a los pobladores de la Sector Quitumbe; y procesamiento mediante la herramienta Rapidminer.

Fuente: **Autora**

### 3.3.2. Describir los datos

Los datos se encuentran almacenados en la Tabla Persona, que se puede visualizar en la Figura No. 5 donde hemos utilizado la herramienta gratuita online “dbdesigner.net”.

persona		
<u>per_id</u>	<u>SERIAL</u>	<pk>
per_apellidos	character varying(200)	
per_celular1	character varying(15)	
per_celular2	character varying(15)	
per_direccion	character varying(500)	
per_email	character varying(200)	
per_fecha_nacimiento	date	
per_nombres	character varying(200)	
per_numero_documento	character varying(15)	<ak>
per_referencia_web	character varying(200)	
per_sexo	character varying(1)	
per_telefono1	character varying(15)	
per_telefono2	character varying(15)	
cat_tipo_persona_id	bigint	
ubi_id	bigint	

**Figura 5.** Tabla Persona

**Fuente:** Autora

**Tabla Persona.** - Esta tabla es la tabla central, es decir donde se registran las personas que se han tomado como muestra para nuestro estudio. La tabla tiene un total de 32.645 registros. Los campos de cada registro de esta tabla son los que continuación describiremos uno a uno, como podemos observar en la Figura No.8

**per\_id:** Tipo numérico. Este campo es el identificador auto numérico de la tabla.

**per\_apellidos y per\_nombres:** Este campo contiene el nombre de las personas que tomamos con muestra en nuestra encuesta realizada al sector de Quitumbe. El tipo de Dato de los campos indicados es un varchar.

**per\_celular1 y per\_celular2:** Es de tipo varchar, este campo nos permite tener una mejor manera de localización para el municipio.

**per\_direccion, per\_email:** Este tipo de dato es varchar, nos permite agregar la dirección del domicilio con el mail, para poder realizar cualquier gestión en el barrio.

**per\_fecha\_nacimiento:** Es de tipo date, nos permite registrar su edad, de esta manera podremos controlar la veracidad de los datos.

**per\_referencia\_web:** Esta opción nos permite escoger entre varios medios de comunicación, por los cuales los moradores del sector Quitumbe, que podrían opinar acerca de la gestión municipal o gestión pública.

**per\_sexo:** Es de tipo varchar, nos permite determinar con qué frecuencia y qué cantidad de la población tomada de la muestra es la que más participa en las encuestas y por ende en la gestión de la participación ciudadana.

entidades_reponsables		
enre_id	SERIAL	<pk>
enre_direccion	character varying(200)	
enre_estado	boolean	
enre_nombre	character varying(200)	
enre_responsable	character varying(200)	
enre_telefono	character varying(200)	

**Figura 6.** Entidades Responsables

**Fuente:** Autora

**Tabla Entidades\_Responsables:** Esta tabla permite que el proyecto de participación ciudadana se extienda a otras entidades, de esta manera incorporar al estudio de indicadores en relación con el uso de la tecnología, a otras instituciones, así como podemos visualizar en la Figura No.9.

Los campos de cada registro de esta tabla son los que continuación describiremos:

**enre\_id:** Tipo numérico. Este campo es el identificador auto numérico de la tabla.

**enre\_direccion:** Tipo de dato varchar. En este campo ingresamos la dirección de las entidades que desean implementar la participación ciudadana en sus proyectos.

**enre\_estado:** Tipo de dato, booleano; es decir si está vigente o cancelado.

**enre\_nombre:** Tipo de dato varchar, en este campo ingresamos el nombre de la institución que estamos realizando el estudio en nuestro caso el Municipio de Quito.

**enre\_responsable:** Tipo de dato varchar, en este caso ingresamos el nombre del responsable del área.

**enre\_telefono:** Tipo de dato varchar, en este campo ingresamos los numero de teléfono de la empresa.

requerimientos		
<u>req_id</u>	<u>SERIAL</u>	<pk>
req_area	character varying(200)	
req_descripcion	character varying(200)	
req_documentos	character varying(500)	
req_estado	boolean	
req_nombre	character varying(200)	

**Figura 7.** Tabla de Requerimientos

**Fuente:** Autora

**Tabla de Requerimientos:** en esta tabla podemos determinar los requerimientos de la ciudadanía, entre los cuales está la participación ciudadana y que medios tecnológicos serían de fácil interacción con la ciudadanía para obtener información, o conocer acerca de los servicios que ofrece el municipio.

Los campos que contiene la tabla son los que describiremos según la Figura No.10

**req\_id:** Tipo numérico. Este campo es el identificador auto numérico de la tabla.

**req\_area:** Tipo de datos varchar, en este campo tenemos una breve descripción del área a la que se solicita los requerimientos; en este caso Participación Ciudadana del municipio de Quito.

**req\_descripcion:** Tipo de Dato varchar, aquí describimos las solicitudes realizadas al Municipio o administración zonal de Quitumbe.

**req\_documentos:** Tipo de Dato varchar, en este campo ingresamos que tipo de documento fue ingresado al municipio, es decir si fue una solicitud respondida por parte del municipio o un oficio realizadas para convocatorias a las reuniones.

**req\_estado:** Tipo de dato booleano, en este campo podemos determinar en qué estado se encuentra la solicitud ya sea vigente, en espera, o rechazado.

**req\_nombre:** Tipo de Dato varchar, en este campo se registra el nombre de la persona que está encargada de la solicitud.

ubicacion_geografica		
<u>ubi_id</u>	<u>SERIAL</u>	<pk>
ubi_nombre	character varying(200)	
ubi_poblacion	character varying(200)	

### **Figura 8. Ubicación Geográfica**

**Fuente:** Autora

**Tabla Ubicación Geográfica:** En esta tabla describiremos el sector de la muestra que estamos tomando para nuestro estudio, en este caso es para el sector de Quitumbe de la zona sur de Quito, así como podemos observar en la Figura No.11.

**ubi\_id:** Tipo numérico. Este campo es el identificador auto numérico de la tabla.

**ubi\_nombre:** Tipo de datos varchar, en este campo describiremos el lugar de donde se tomó la muestra.

**ubi\_poblacion:** Tipo de dato varchar, aquí describimos el número de habitantes por zona.

#### **3.3.3. Explorar los Datos**

Una vez que se han descrito los datos, se procede a explorarlos, esto implica aplicar combinaciones de las variables que se han elegido para el estudio las mismas que revelarán si el nivel de satisfacción de las solicitudes realizadas al municipio es buena, mala o excelente. Al igual que predomina en la intervención de las encuestas, es decir mujeres u hombres y con qué medio tecnológico pueden interactuar con mayor facilidad, de esta manera se crearán tablas y gráficos de distribución de los datos.

En la Tabla No.6, podemos observar nuestra muestra de datos, procesada en la aplicación:

**Tabla 6:**  
Tabla de Encuestas de la Primera Pregunta.

ExampleSet (81 examples, 0 special attributes, 7 regular attributes) Filter (81 / 81 examples):

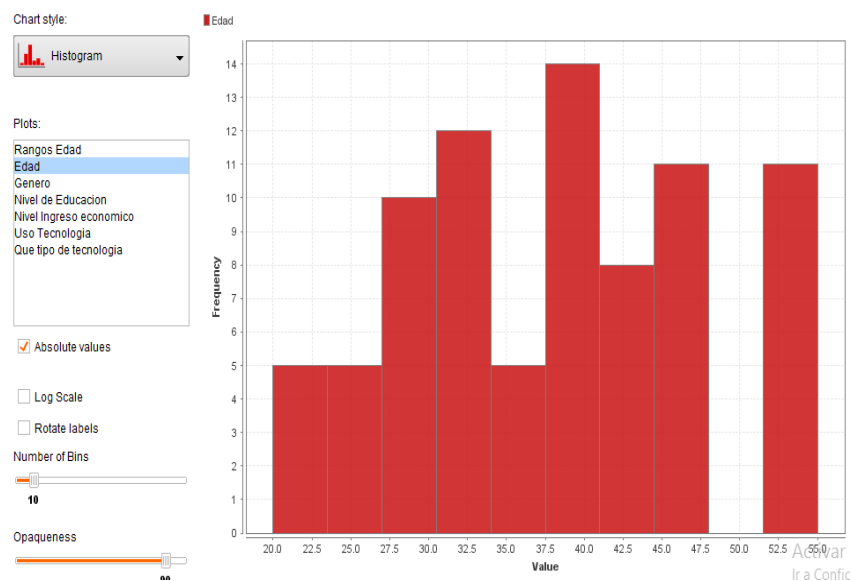
Row No.	Rangos Edad	Edad	Genero	Nivel de Edu...	Nivel Ingres...	Uso Tecnolo...	Que tipo de t...
1	JOVEN	20	MUJER	SECUNDARIA	SUELDO BA...	SI	COMPUTAD...
2	JOVEN	24	MUJER	SECUNDARIA	SUELDO BA...	SI	COMPUTAD...
3	JOVEN	23	HOMBRE	SECUNDARIA	ENTRE 500 Y...	SI	COMPUTAD...
4	JOVEN	22	HOMBRE	SUPERIOR	ENTRE 500 Y...	SI	COMPUTAD...
5	JOVEN	23	HOMBRE	SUPERIOR	ENTRE 500 Y...	SI	CYBER
6	ADULTO	25	MUJER	SUPERIOR	MAYOR A 1200	SI	CYBER
7	ADULTO	26	MUJER	SUPERIOR	MAYOR A 1200	SI	COMPUTAD...
8	ADULTO	30	MUJER	ESPECIALID...	MAYOR A 1200	SI	CYBER
9	ADULTO	30	MUJER	ESPECIALID...	MAYOR A 1200	SI	COMPUTAD...
10	ADULTO	28	HOMBRE	SUPERIOR	MAYOR A 1200	SI	CYBER
11	SENIOR	31	HOMBRE	SECUNDARIA	ENTRE 500 Y...	SI	CYBER
12	SENIOR	35	HOMBRE	SECUNDARIA	ENTRE 500 Y...	SI	COMPUTAD...
13	SENIOR	38	HOMBRE	SUPERIOR	MAYOR A 1200	SI	CELULAR
14	SENIOR	37	MUJER	ESPECIALID...	MAYOR A 1200	SI	CELULAR
15	SENIOR	36	MUJER	SUPERIOR	ENTRE 500 Y...	SI	CELULAR
16	MAYOR	45	HOMBRE	ESPECIALID...	ENTRE 500 Y...	SI	COMPUTAD...
17	MAYOR	47	HOMBRE	PRIMARIA	INGRESOS M...	NO	OTROS

**Fuente:** Autora

**Primera Pregunta: Ha utilizado medios tecnológicos para opinar acerca de la Gestión del Municipio u otro organismo del estado:**

**Si, que medios a utilizado.**

La encuesta realizada a los ciudadanos del sector Quitumbe, al relacionar el indicador de la edad con el número de pobladores podemos determinar que la parte más activa en contestar las preguntas e interesarse en el tema son los ciudadanos que oscilan entre 30 y 35 años; pero las personas que están entre los 37 y 42 años son los más interesados en este tipo de encuestas para el bienestar de su barrio, como podemos observar en la Figura No.9.

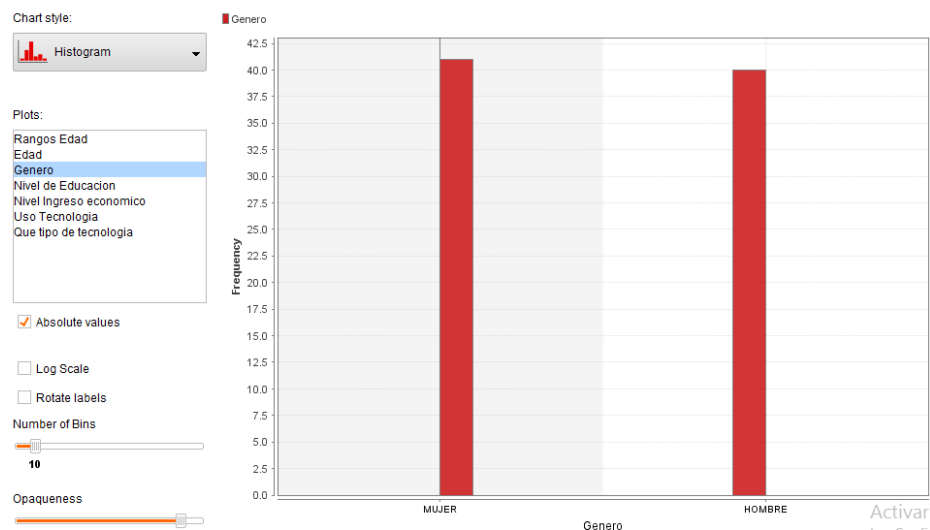


**Figura 9.** Gráfica Histograma de la Primera Pregunta - EDAD

**Fuente:** Autora

Genero

Al relacionar el número de ciudadanos Hombres y Mujeres que más participaron en las encuestas fueron las Mujeres, como podemos observar en la Figura No.10.

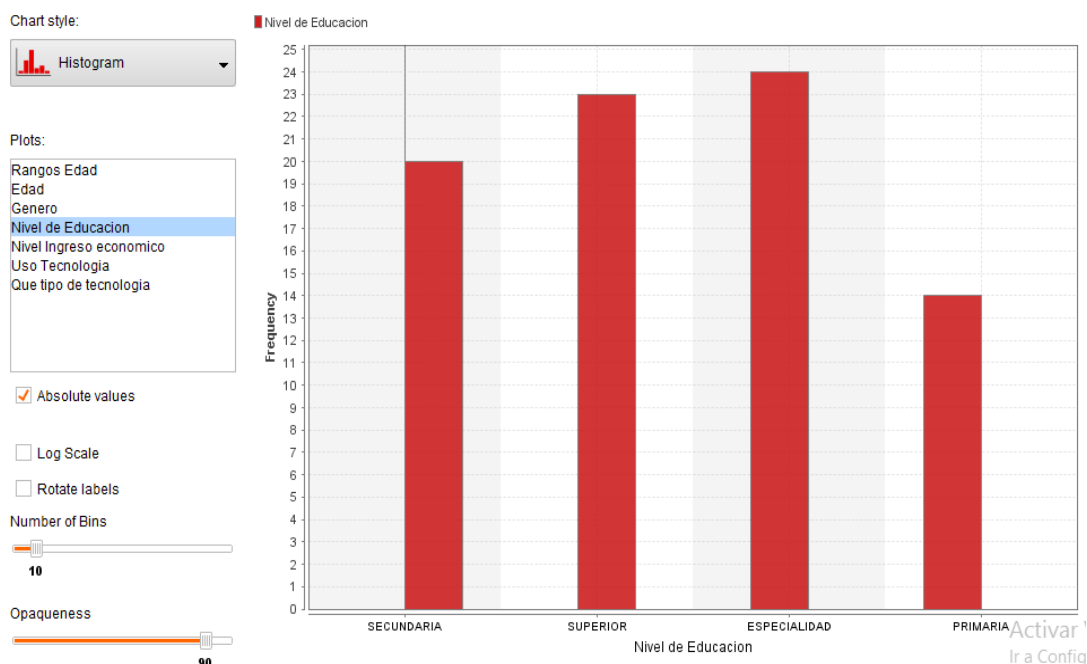


**Figura 10.** Gráfica Histograma de la Variable Género

**Fuente:** Autora

Nivel de Educación

La muestra nos permite determinar el nivel de Educación de los ciudadanos del sector Quitumbe, obteniendo de esta manera el nivel de estudios que poseen los moradores, como podemos observar en la Figura No. 11.



**Figura 11.** Gráfica Histograma Nivel de Educación

**Fuente:** Autora

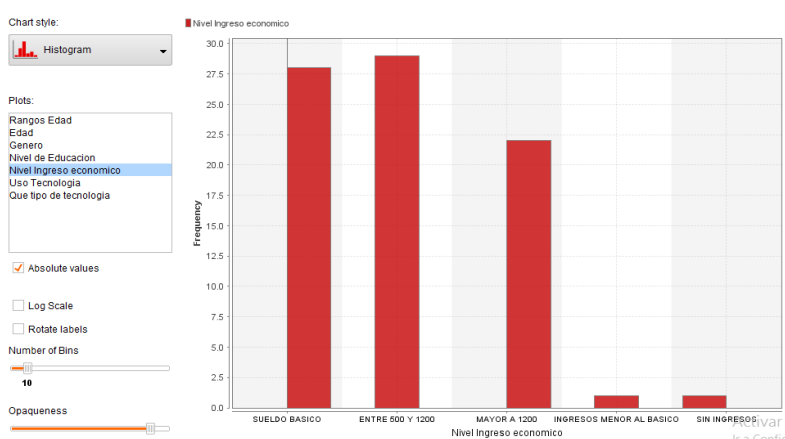
### Nivel Económico

Podeos determinar el nivel socioeconómico de los ciudadanos; los cuales les hemos clasificado en cuatros aspectos:

- Sueldo Básico
- Entre 500 y 1200
- Mayor a 1200
- Ingresos Menor al básico

Como podemos observar en la Figura No.12.



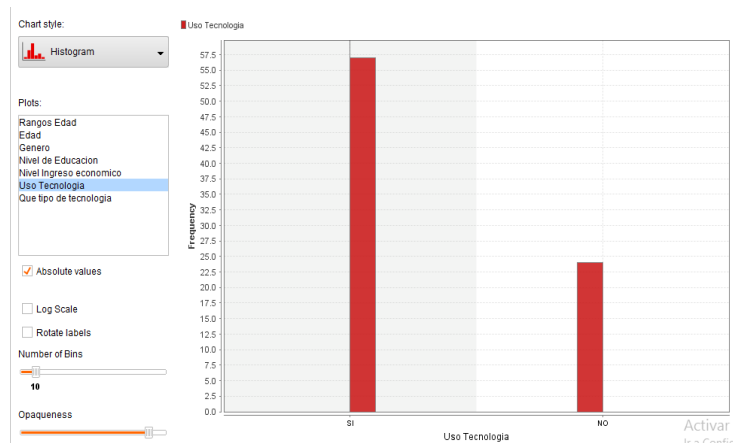


**Figura 12.** Gráfica Uso de la Tecnología vs. Nivel Económico

**Fuente:** Autora

### Uso de la Tecnología

Podemos verificar que la ciudadanía tiene una participación activa y fácil manejo de dispositivos electrónicos, como podemos observar en la Figura No.13.

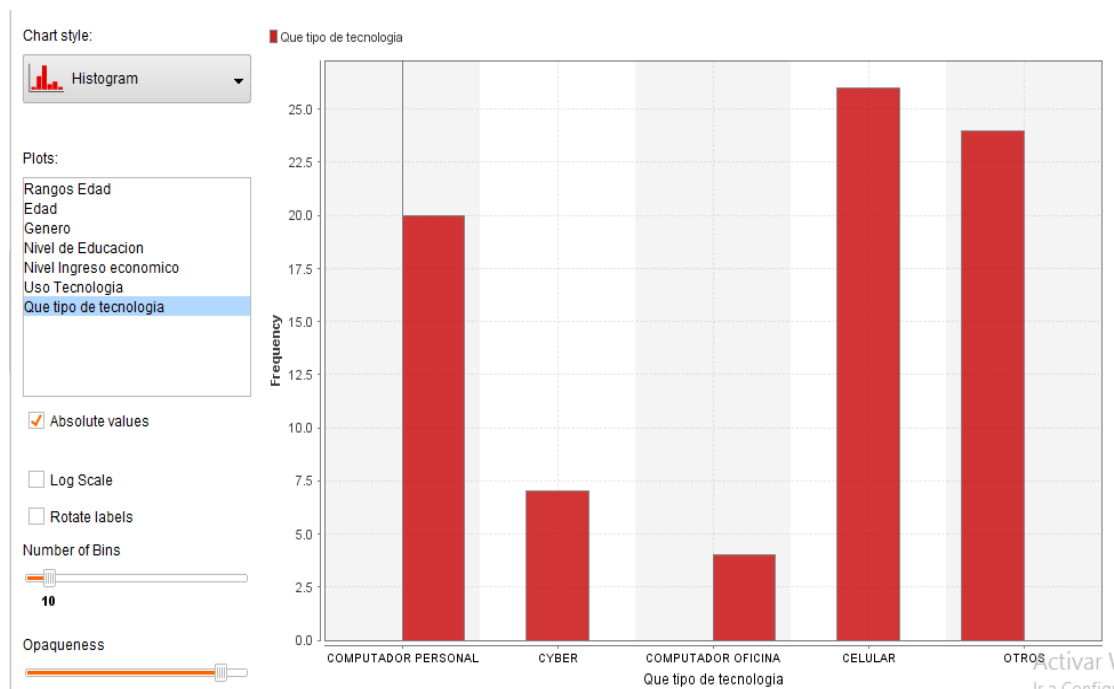


**Figura 13.** Gráfica Histograma uso de la tecnología

**Fuente:** Autora

Tipo de Tecnología que usa, podemos determinar que la ciudadanía tiene más acogida por los dispositivos móviles siendo ese el de mayor relevancia, seguido por Otros, que se clasifico como televisión, radio, prensa, y seguidamente esta la computadora personal que también tiene buena apertura.

Sin contar con los Cybers café o la computadora de la oficina, como podemos observar en la Figura No.14.

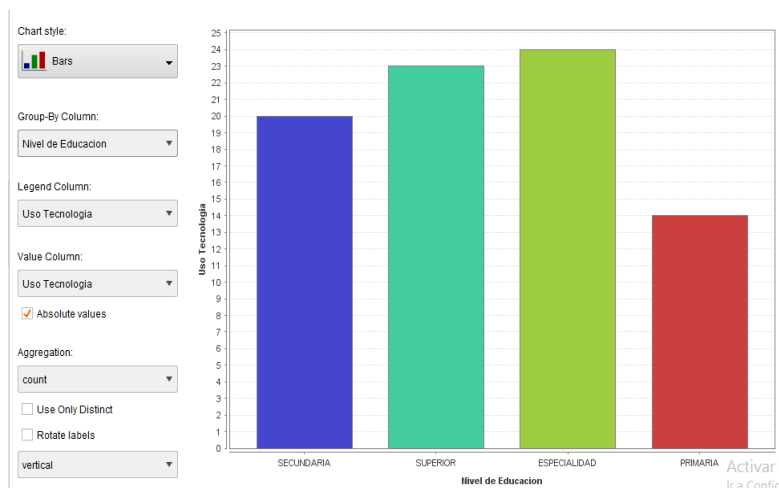


**Figura 14.** Gráfica Tipo de Tecnología

**Fuente:** Autora

### Nivel de Educación vs Uso de la Tecnología

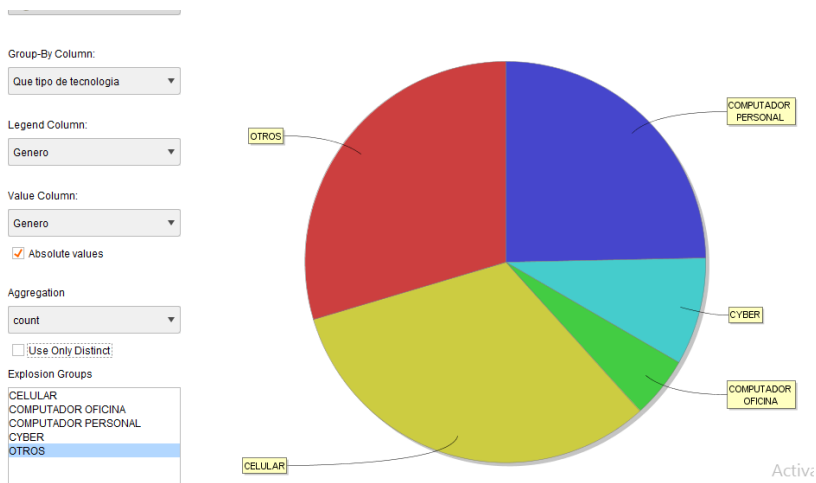
El diagrama de Barras que vemos a continuación relacionamos el Nivel de Educación referente al uso de la tecnología; lo que nos permite demostrar que las personas que tienen una educación ya sea Especialidad y Superior son más apegados a los medios tecnológicos e interactúan de mejor manera con los dispositivos electrónicos, como podemos observar en la Figura No. 15.



**Figura 15.** Gráfica Nivel de Educación vs. Uso Tecnología

**Fuente:** Autora

Según los datos relacionados de la muestra tomada en Quitumbe se puede observar que los medios tecnológicos más utilizados son el celular seguido de otros donde aquí participa la población Mayor y seguido del Computador personal, como podemos observar en la Figura No.16.



**Figura 16.** Gráfica tipo de Tecnología vs. Género

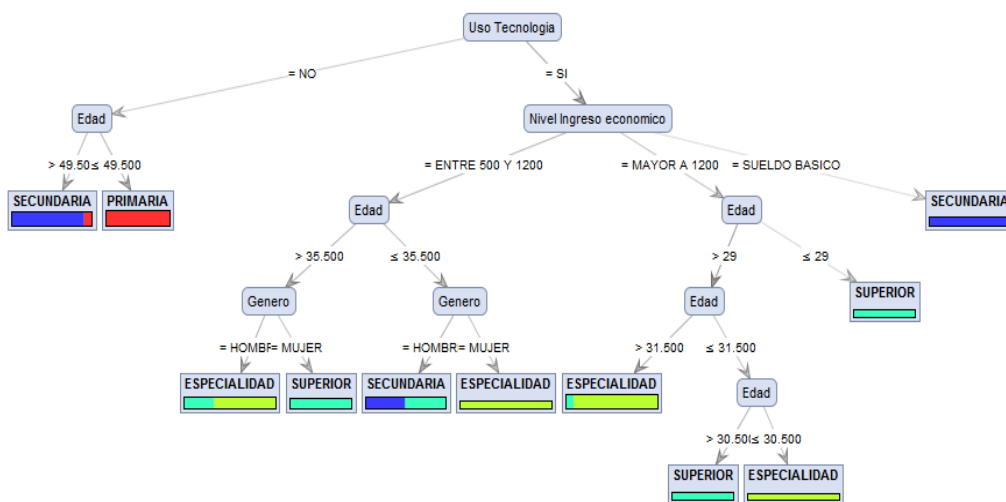
**Fuente:** Autora

Árbol de Decisión

Al obtener la información del siguiente gráfico podemos determinar que el uso de la tecnología con respecto de la Edad no tiene apertura para las personas que tienen un nivel de estudios básico.

Por otro lado, al tomar la variable Nivel de ingreso económico, este se divide en dos partes donde se identifica a personas por la edad y el Género siendo así que las personas que tienen un salario que oscila entre 500 y 1200 y tienen una edad mayor a 35 años y son mujeres con nivel de instrucción Superior utilizan la tecnología.

Las personas que tienen un salario básico y tienen secundaria posiblemente sea participes de la tecnología independientemente de su edad mientras que las personas que ganan más de 1200 y están en una edad mayor a 29 años con un nivel de instrucción alto manejan bien los dispositivos, como podemos observar en la Figura No.17.



**Figura 17.** Gráfica de Árbol de decisión de Uso de la Tecnología

**Fuente:** Autora

Descripción por parte de la herramienta, como podemos observar en la Figura No.18.

## Tree

```

Uso Tecnologia = NO
| Edad > 49.500: SECUNDARIA {SECUNDARIA=10, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=1}
| Edad ≤ 49.500: PRIMARIA {SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=13}
Uso Tecnologia = SI
| Nivel Ingreso economico = ENTRE 500 Y 1200
| | Edad > 35.500
| | | Genero = HOMBRE: ESPECIALIDAD {SECUNDARIA=0, SUPERIOR=3, ESPECIALIDAD=6, PRIMARIA=0}
| | | Genero = MUJER: SUPERIOR {SECUNDARIA=0, SUPERIOR=7, ESPECIALIDAD=0, PRIMARIA=0}
| | Edad ≤ 35.500
| | | Genero = HOMBRE: SECUNDARIA {SECUNDARIA=4, SUPERIOR=4, ESPECIALIDAD=0, PRIMARIA=0}
| | | Genero = MUJER: ESPECIALIDAD {SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=5, PRIMARIA=0}
| Nivel Ingreso economico = MAYOR A 1200
| | Edad > 29
| | | Edad > 31.500: ESPECIALIDAD {SECUNDARIA=0, SUPERIOR=1, ESPECIALIDAD=10, PRIMARIA=0}
| | | Edad ≤ 31.500
| | | | Edad > 30.500: SUPERIOR {SECUNDARIA=0, SUPERIOR=4, ESPECIALIDAD=0, PRIMARIA=0}
| | | | Edad ≤ 30.500: ESPECIALIDAD {SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=3, PRIMARIA=0}
| | Edad ≤ 29: SUPERIOR {SECUNDARIA=0, SUPERIOR=4, ESPECIALIDAD=0, PRIMARIA=0}
| Nivel Ingreso economico = SUELDO BASICO: SECUNDARIA {SECUNDARIA=6, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=0}

```

**Figura 18.** Gráfica de la Descripción del Árbol de Uso de la Tecnología

**Fuente:** Autora

### **Segunda Pregunta: Cuantas veces ha participado en el portal (opción múltiple)**

La pregunta dos se procesaron la información en el software y obtuvimos los siguientes resultados, como podemos observar en la Tabla No.7.

**Tabla 7:**  
Encuesta de la Pregunta Dos

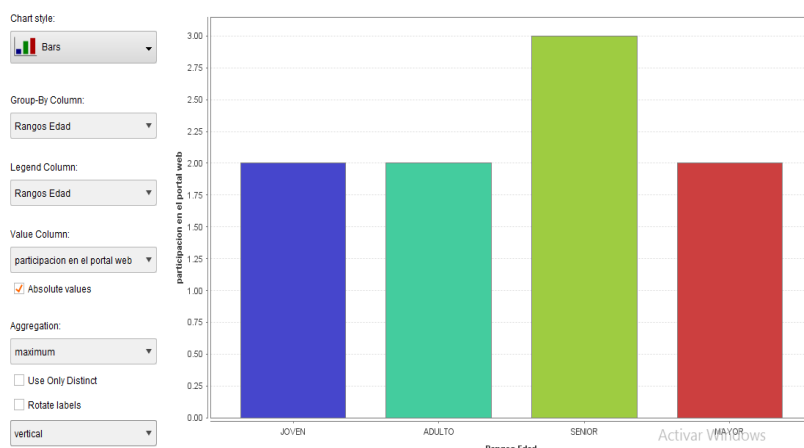
ExampleSet (81 examples, 0 special attributes, 5 regular attributes)

Row No.	Rangos Edad	Edad	Genero	Nivel Ingres...	participacio...
1	JOVEN	20	MUJER	SUELDO BA...	NINGUNA
2	JOVEN	24	MUJER	SUELDO BA...	NINGUNA
3	JOVEN	23	HOMBRE	ENTRE 500 Y...	NINGUNA
4	JOVEN	22	HOMBRE	ENTRE 500 Y...	NINGUNA
5	JOVEN	23	HOMBRE	ENTRE 500 Y...	NINGUNA
6	ADULTO	25	MUJER	MAYOR A 1200	MEDIA
7	ADULTO	26	MUJER	MAYOR A 1200	NINGUNA
8	ADULTO	30	MUJER	MAYOR A 1200	BAJA
9	ADULTO	30	MUJER	MAYOR A 1200	NINGUNA
10	ADULTO	28	HOMBRE	MAYOR A 1200	NINGUNA
11	SENIOR	31	HOMBRE	ENTRE 500 Y...	BAJA
12	SENIOR	35	HOMBRE	ENTRE 500 Y...	MEDIA
13	SENIOR	38	HOMBRE	MAYOR A 1200	ALTA
14	SENIOR	37	MUJER	MAYOR A 1200	BAJA
15	SENIOR	36	MUJER	ENTRE 500 Y...	NINGUNA
16	MAYOR	45	HOMBRE	ENTRE 500 Y...	BAJA
17	MAYOR	47	HOMBRE	INGRESOS M...	NINGUNA

**Fuente:** Autora

Al relacionar las variables Participación en el portal Web vs Rangos de Edad, se clasifico mediante la edad sus diferentes categorías, según la tabla de categorías establecidas por edad en este caso el Rango de Edad Sénior son las personas que su edad oscila entre los 31 y 40 años.

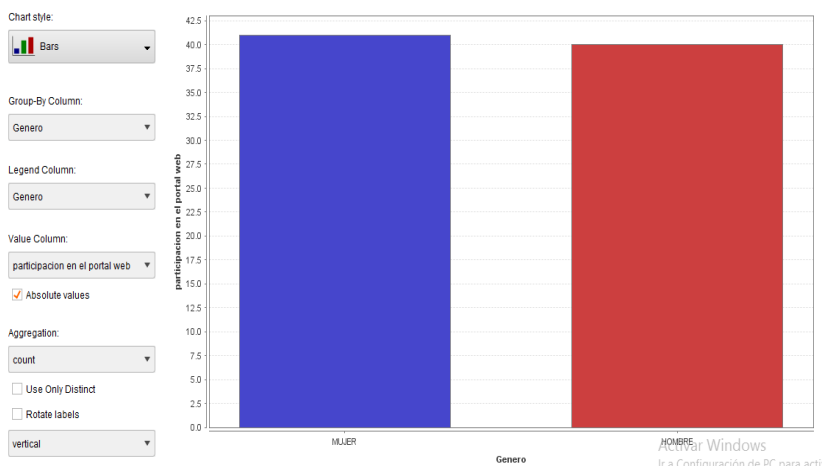
Dichos son las personas que participan activamente en el portal Web, como podemos observar en la Figura No.19.



**Figura 19.** Gráfica de los Rangos de Edad

**Fuente:** Autora

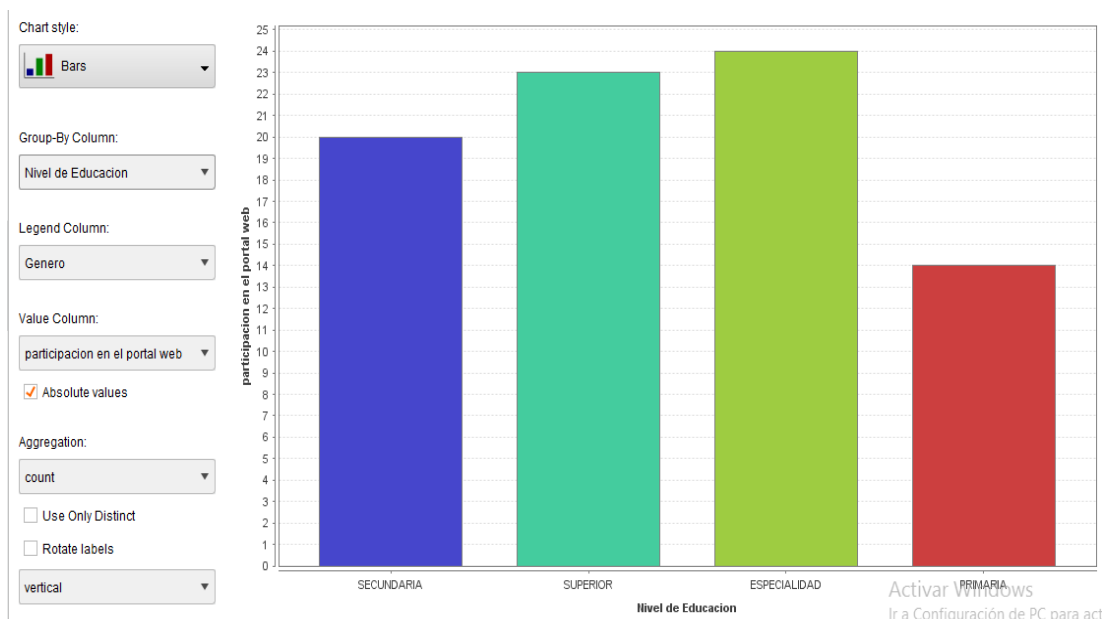
Al relacionar la variable participación en el portal web vs el género podemos determinar que la población activa son las mujeres, como podemos observar en la Figura No.20.



**Figura 20.** Gráfica del Género

**Fuente:** Autora

La relación de las variables Nivel de Educación vs participación en el portal Web se puede determinar que las personas con nivel de instrucción más alto son proporcionales a más interacción con el portal, como podemos observar en la Figura No.21.



**Figura 21.** Gráfica del Nivel de Educación

**Fuente:** Autora

### Arboles de Decisión

Podemos determinar que la Edad es la variable mandataria ya que al ser la variable independiente nos permite determinar que las personas que tienen secundaria y su edad es mayor a 49 años participan activamente mientras que al integran la variable nivel de Ingreso Económico nos permite clasificar entre sueldo básico, mayor a 1200 y salario entre 500 y 1200.

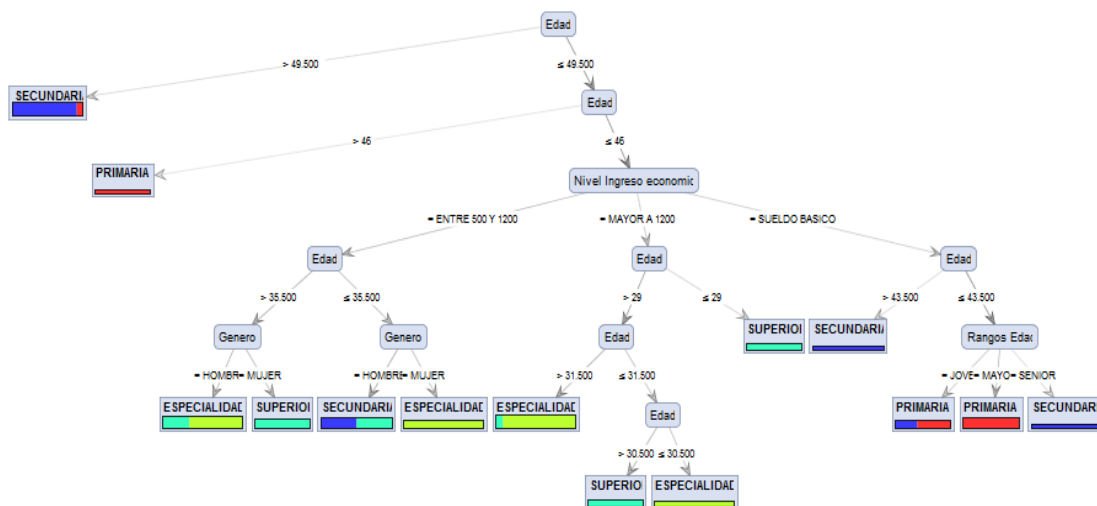
Luego de relacionar estas variables interviene la variable género que nos indica que las Mujeres con estudios superiores participan activamente en el portal.

Mientras que al relacionar la variable Rango de edad que se clasifico por:

- Joven
- Mayor
- Sénior



Las personas que tienen Primaria y son Mayores no participan en el portal; mientras que las personas Sénior que tienen Secundaria participan en el portal, como podemos observar en la Figura No.22.



**Figura 22:** Árbol de Decisión Del Nivel de Ingreso Económico

**Fuente:** Autora

Descripción de la herramienta, como podemos observar en la Figura No.23.

```

Tree
Edad > 49.500: SECUNDARIA (SECUNDARIA=10, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=1)
Edad <= 49.500
| Edad > 46: PRIMARIA (SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=2)
| Edad <= 46
| | Nivel Ingreso economico = ENTRE 500 Y 1200
| | | Edad > 35.500
| | | | Genero = HOMBRE: ESPECIALIDAD (SECUNDARIA=0, SUPERIOR=3, ESPECIALIDAD=6, PRIMARIA=0)
| | | | Genero = MUJER: SUPERIOR (SECUNDARIA=0, SUPERIOR=7, ESPECIALIDAD=0, PRIMARIA=0)
| | | | Edad <= 35.500
| | | | | Genero = HOMBRE: SECUNDARIA (SECUNDARIA=4, SUPERIOR=4, ESPECIALIDAD=0, PRIMARIA=0)
| | | | | Genero = MUJER: ESPECIALIDAD (SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=5, PRIMARIA=0)
| | | Nivel Ingreso economico = MAYOR A 1200
| | | | Edad > 29
| | | | | Edad > 31.500: ESPECIALIDAD (SECUNDARIA=0, SUPERIOR=1, ESPECIALIDAD=10, PRIMARIA=0)
| | | | | Edad <= 31.500
| | | | | | Edad > 30.500: SUPERIOR (SECUNDARIA=0, SUPERIOR=4, ESPECIALIDAD=0, PRIMARIA=0)
| | | | | | Edad <= 30.500: ESPECIALIDAD (SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=3, PRIMARIA=0)
| | | | Edad <= 29: SUPERIOR (SECUNDARIA=0, SUPERIOR=4, ESPECIALIDAD=0, PRIMARIA=0)
| | | Nivel Ingreso economico = SUELDO BASICO
| | | | Edad > 43.500: SECUNDARIA (SECUNDARIA=2, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=0)
| | | | Edad <= 43.500
| | | | | Rangos Edad = JOVEN: PRIMARIA (SECUNDARIA=2, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=3)
| | | | | Rangos Edad = MAYOR: PRIMARIA (SECUNDARIA=0, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=8)
| | | | | Rangos Edad = SENIOR: SECUNDARIA (SECUNDARIA=2, SUPERIOR=0, ESPECIALIDAD=0, PRIMARIA=0)
    
```

**Figura 23.** Gráfica de Descripción Del Árbol de Nivel de Ingreso

**Fuente:** Autora

**Tercera Pregunta: Que tema considera usted que es el más importante, en el cual le gustaría opinar (opción múltiple).**

Al ingresar los datos nos indica la siguiente información, como podemos observar en la Figura No.24.

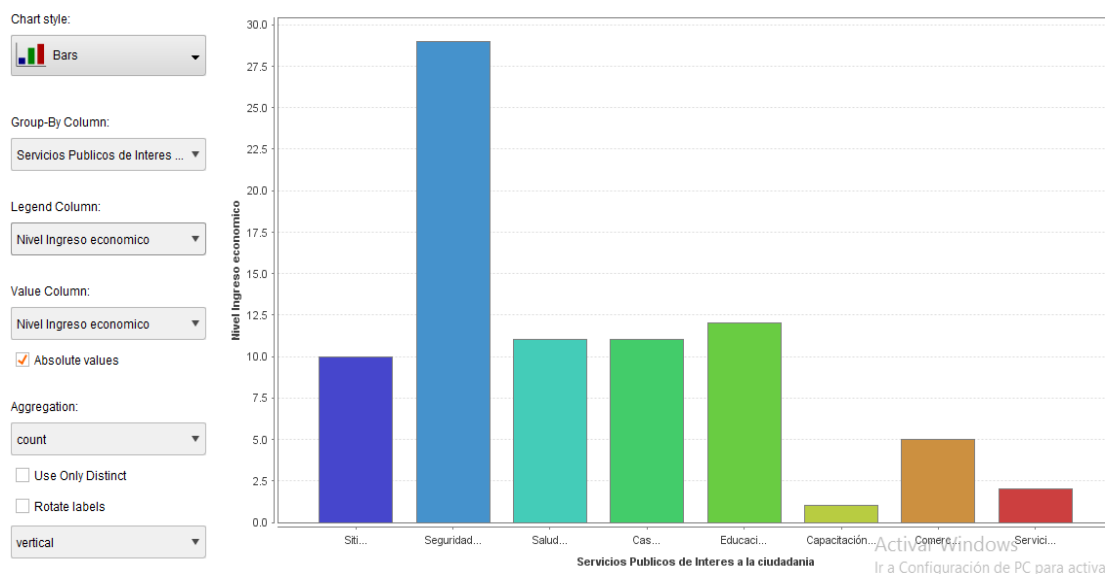
ExampleSet (81 examples, 1 special attribute, 5 regular attributes)

Row No.	Nivel de Edu...	Rangos Edad	Edad	Genero	Nivel Ingres...	Servicios Pu...
1	SECUNDARIA	JOVEN	20	MUJER	SUELDO BA...	Sitios Recrea...
2	SECUNDARIA	JOVEN	24	MUJER	SUELDO BA...	Sitios Recrea...
3	SECUNDARIA	JOVEN	23	HOMBRE	ENTRE 500 Y...	Seguridad y p...
4	SUPERIOR	JOVEN	22	HOMBRE	ENTRE 500 Y...	Seguridad y p...
5	SUPERIOR	JOVEN	23	HOMBRE	ENTRE 500 Y...	Sitios Recrea...
6	SUPERIOR	ADULTO	25	MUJER	MAYOR A 1200	Seguridad y p...
7	SUPERIOR	ADULTO	26	MUJER	MAYOR A 1200	Seguridad y p...
8	ESPECIALID...	ADULTO	30	MUJER	MAYOR A 1200	Seguridad y p...
9	ESPECIALID...	ADULTO	30	MUJER	MAYOR A 1200	Seguridad y p...
10	SUPERIOR	ADULTO	28	HOMBRE	MAYOR A 1200	Salud y Acce...
11	SECUNDARIA	SENIOR	31	HOMBRE	ENTRE 500 Y...	Casas Comu...
12	SECUNDARIA	SENIOR	35	HOMBRE	ENTRE 500 Y...	Educación (c...
13	SUPERIOR	SENIOR	38	HOMBRE	MAYOR A 1200	Educación (c...
14	ESPECIALID...	SENIOR	37	MUJER	MAYOR A 1200	Casas Comu...
15	SUPERIOR	SENIOR	36	MUJER	ENTRE 500 Y...	Capacitación ...
16	ESPECIALID...	MAYOR	45	HOMBRE	ENTRE 500 Y...	Comercio Inf...
17	PRIMARIA	MAYOR	47	HOMBRE	INGRESOS M...	Salud y Acce...

**Figura 24.** Gráfica de la Tabla de Encuestas de la Pregunta Tres

**Fuente:** Autora

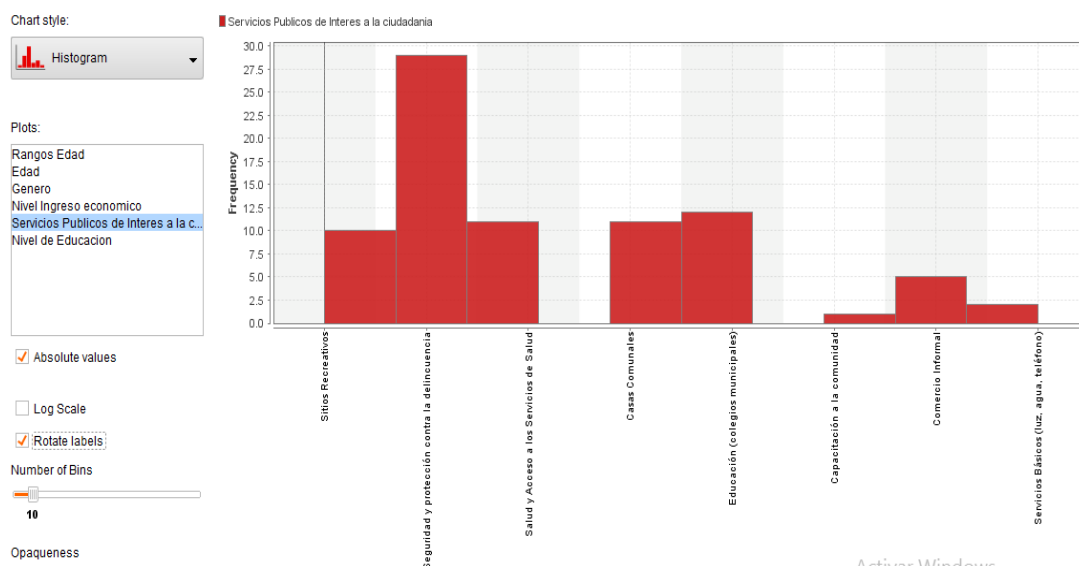
Al analizar las variables Nivel de Ingreso Económico versus Servicios Públicos de Interés a la ciudadanía nos indica que la comunidad está preocupada por la seguridad que no es tan buena en el sector de Quitumbe, además la educación que sería un factor importante en el estudio de nuestro proyecto, como podemos observar en la Figura No.26.



**Figura 25.** Gráfica de Servicios Público vs. Nivel de Ingresos

**Fuente:** Autora

Al determinar nuestra muestra versus los servicios Públicos de Interés a la ciudadanía podemos determinar que la seguridad sigue siendo el principal requerimiento que el barrio pide a la concejalía del Municipio de Quito, como podemos observar en la Figura No.27.

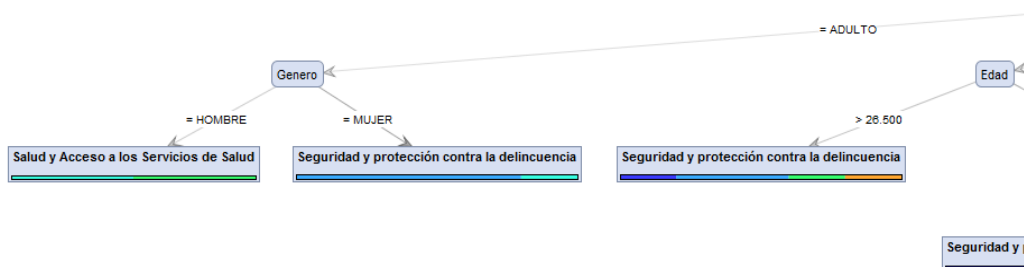


**Figura 26.** Gráfica de Servicios Públicos

**Fuente:** Autora

Árbol de Decisión:

Podemos determinar que el factor de la edad cuando las personas están en un rango de edad Adulto y su género es Hombre los mismos que se preocupan por la salud y acceso a los servicios, mientras que las mujeres se preocupan por la Seguridad y Protección contra la delincuencia, como podemos observar en la Figura No.27.



**Figura 27.** Gráfica del Árbol de Decisión Genero vs. Servicios Públicos

**Fuente:** Autora

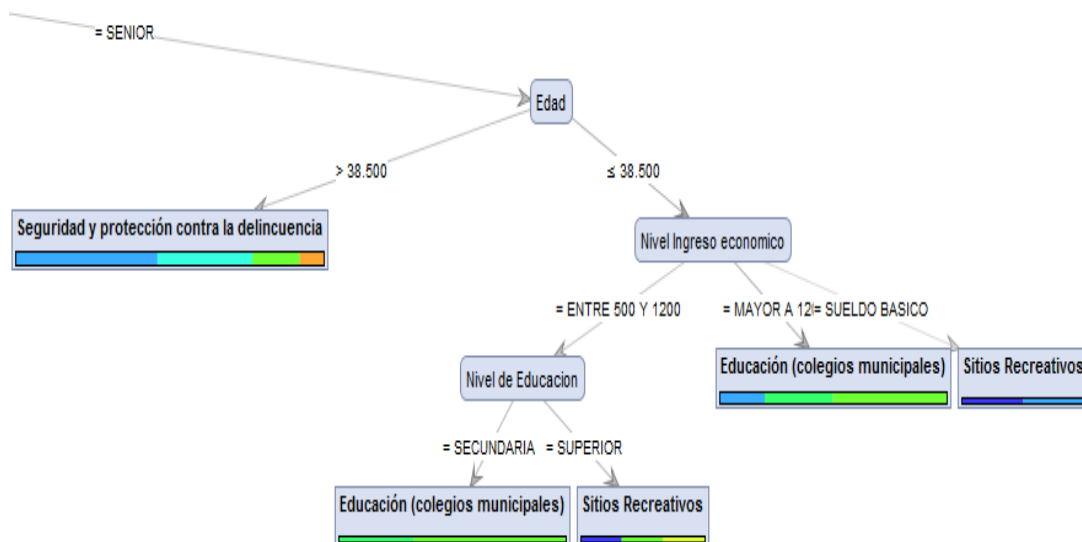
Mientras que cuando estamos en un rango de edad Joven y el nivel de ingreso económico está entre 500 y 1200 la preocupación es la misma, Seguridad mientras que cuando sus ingresos son menores su preocupación es por sitios recreativos, como podemos observar en la Figura No.28.



**Figura 28.** Gráfica Árbol de Decisión - Rangos de Edad vs. Nivel de Ingreso

**Fuente:** Autora

Cuando el rango de edad está en Senior es decir mayores de 38 y sus ingresos están entre 500 y 1200 o mayor a 1200; su preocupación es por la Educación, como podemos observar en la Figura No.29.



**Figura 29.** Gráfica Arboles De Decisión

**Fuente:** Autora

**Cuarta Pregunta: Como usted calificaría el nivel de satisfacción de las solicitudes realizadas al municipio o administración zonal Quitumbe.**

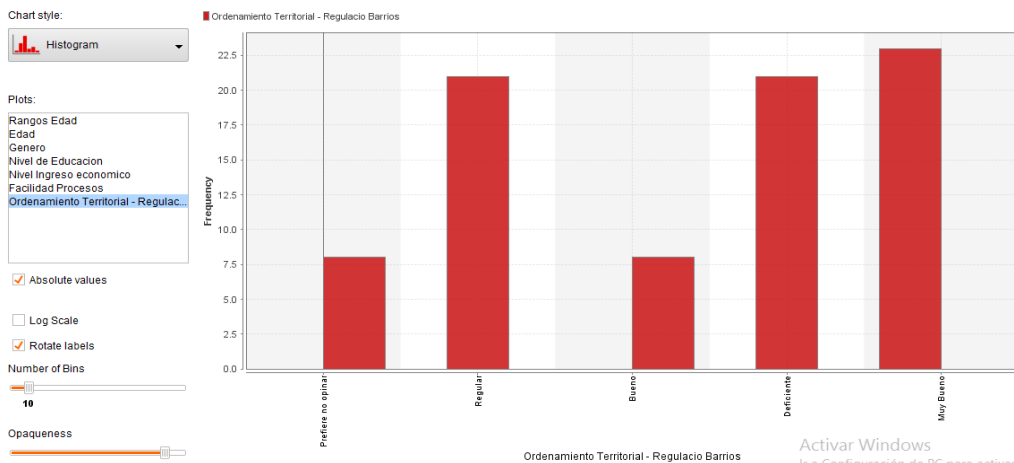
Al procesar la información de la pregunta cuatro obtenemos lo siguiente, como podemos observar en la Tabla No.8.

**Tabla 8:** Pregunta Cuatro

Row No.	Ordenamien...	Rangos Edad	Edad	Genero	Nivel de Edu...	Nivel Ingres...	Facilidad Pr...
1	Prefiere no o...	JOVEN	20	MUJER	SECUNDARIA	SUELDO BA...	4
2	Prefiere no o...	JOVEN	24	MUJER	SECUNDARIA	SUELDO BA...	4
3	Regular	JOVEN	23	HOMBRE	SECUNDARIA	ENTRE 500 Y...	2
4	Regular	JOVEN	22	HOMBRE	SUPERIOR	ENTRE 500 Y...	2
5	Bueno	JOVEN	23	HOMBRE	SUPERIOR	ENTRE 500 Y...	3
6	Regular	ADULTO	25	MUJER	SUPERIOR	MAYOR A 1200	3
7	Regular	ADULTO	26	MUJER	SUPERIOR	MAYOR A 1200	3
8	Deficiente	ADULTO	30	MUJER	ESPECIALID...	MAYOR A 1200	1
9	Deficiente	ADULTO	30	MUJER	ESPECIALID...	MAYOR A 1200	1
10	Deficiente	ADULTO	28	HOMBRE	SUPERIOR	MAYOR A 1200	1
11	Regular	SENIOR	31	HOMBRE	SECUNDARIA	ENTRE 500 Y...	1
12	Regular	SENIOR	35	HOMBRE	SECUNDARIA	ENTRE 500 Y...	3
13	Bueno	SENIOR	38	HOMBRE	SUPERIOR	MAYOR A 1200	3
14	Bueno	SENIOR	37	MUJER	ESPECIALID...	MAYOR A 1200	2
15	Muy Bueno	SENIOR	36	MUJER	SUPERIOR	ENTRE 500 Y...	2
16	Muy Bueno	MAYOR	45	HOMBRE	ESPECIALID...	ENTRE 500 Y...	4
17	Muy Bueno	MAYOR	47	HOMBRE	PRIMARIA	INGRESOS M...	4

**Fuente:** Autora

La información luego de ser procesada nos permite analizar obteniendo como resultado que el nivel de satisfacción de la ciudadanía con respecto a los servicios que presenta la ordenanza territorial es muy bueno, aunque las personas que no están conformes con los requerimientos también están en gran porcentaje, como podemos observar en la Figura No.30.



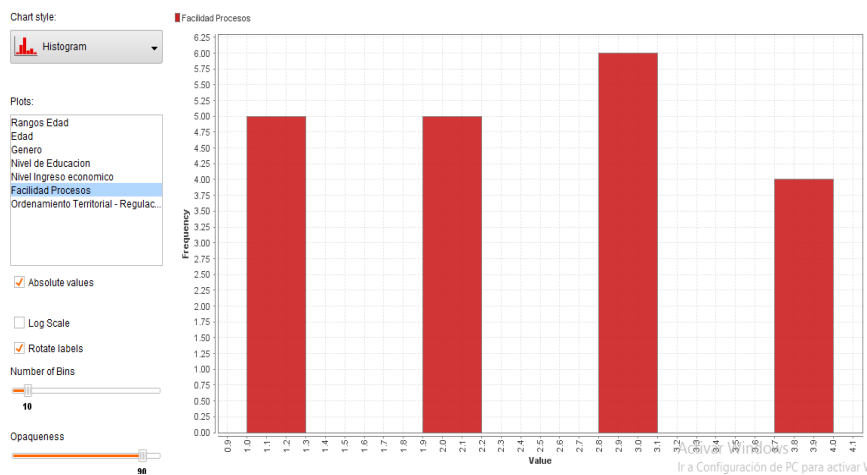
**Figura 30.** Grafica Ordenamiento Territorial

**Fuente:** Autora

La facilidad de procesos la comparamos directamente con la muestra de ciudadanos que escogimos donde podemos indicar que existen 4 tipos de procesos entre ellos están:

1. requisitos
2. Documentos
3. tramites
4. Coordinación

Por lo que los Trámites son lo que demandan más tiempo para la ciudadanía al no tener respuesta rápida y ágil, como podemos observar en la Figura No.31.

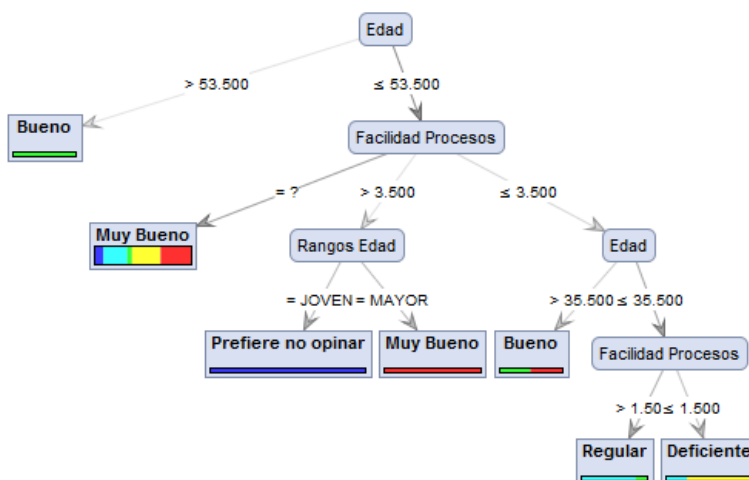


**Figura 31.** Grafica Facilidad de procesos

**Fuente:** Autora

Arboles de Decisión:

Las personas de Rango de Edad Mayor es decir mayor a 50 años indican q los procesos son muy Buenos, mientras que los Jóvenes prefieren no Opinar con respecto a los servicios brindados por la Concejalía del municipio, como podemos observar en la Figura No.34.



**Figura 32.** Grafica de Facilidad Procesos

**Fuente:** Autora

Descripción que es procesada en la herramienta, como podemos observar en la Figura No.33.

**Tree**

```

Edad > 53.500: Bueno {Prefiere no opinar=0, Regular=0, Bueno=2, Deficiente=0, Muy Bueno=0}
Edad ≤ 53.500
| Facilidad Procesos = ? : Muy Bueno {Prefiere no opinar=6, Regular=15, Bueno=3, Deficiente=18, Muy Bueno=19}
| Facilidad Procesos > 3.500
| | Rangos Edad = JOVEN: Prefiere no opinar {Prefiere no opinar=2, Regular=0, Bueno=0, Deficiente=0, Muy Bueno=0}
| | Rangos Edad = MAYOR: Muy Bueno {Prefiere no opinar=0, Regular=0, Bueno=0, Deficiente=0, Muy Bueno=2}
| Facilidad Procesos ≤ 3.500
| | Edad > 35.500: Bueno {Prefiere no opinar=0, Regular=0, Bueno=2, Deficiente=0, Muy Bueno=2}
| | Edad ≤ 35.500
| | | Facilidad Procesos > 1.500: Regular {Prefiere no opinar=0, Regular=5, Bueno=1, Deficiente=0, Muy Bueno=0}
| | | Facilidad Procesos ≤ 1.500: Deficiente {Prefiere no opinar=0, Regular=1, Bueno=0, Deficiente=3, Muy Bueno=0}
    
```

**Figura 33.** Árboles de Decisión Facilidad de proceso

**Fuente:** Autora

**Quinta Pregunta:** Cuales serían los medios por los cuales le gustaría obtener información o conocer acerca de los servicios que ofrece el municipio. (En el portal web).



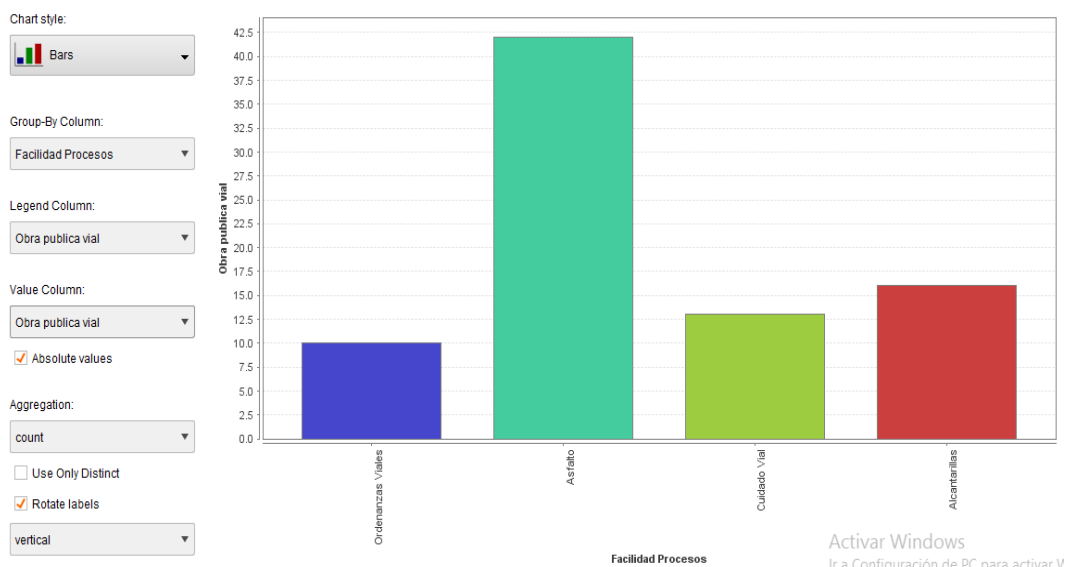
Luego de cargar los datos obtenemos el siguiente cuadro, como podemos observar en la Tabla No.8.

**Tabla 8.**  
Encuesta Pregunta Cinco

Row No.	Facilidad Pr...	Rangos Edad	Edad	Genero	Nivel de Edu...	Nivel Ingres...	Obra public...
1	Ordenanzas ...	JOVEN	20	MUJER	SECUNDARIA	SUELDO BA...	Prefiere no o...
2	Asfalto	JOVEN	24	MUJER	SECUNDARIA	SUELDO BA...	Prefiere no o...
3	Cuidado Vial	JOVEN	23	HOMBRE	SECUNDARIA	ENTRE 500 Y...	Regular
4	Asfalto	JOVEN	22	HOMBRE	SUPERIOR	ENTRE 500 Y...	Regular
5	Asfalto	JOVEN	23	HOMBRE	SUPERIOR	ENTRE 500 Y...	Bueno
6	Asfalto	ADULTO	25	MUJER	SUPERIOR	MAYOR A 1200	Regular
7	Asfalto	ADULTO	26	MUJER	SUPERIOR	MAYOR A 1200	Regular
8	Cuidado Vial	ADULTO	30	MUJER	ESPECIALID...	MAYOR A 1200	Deficiente
9	Asfalto	ADULTO	30	MUJER	ESPECIALID...	MAYOR A 1200	Deficiente
10	Asfalto	ADULTO	28	HOMBRE	SUPERIOR	MAYOR A 1200	Deficiente
11	Asfalto	SENIOR	31	HOMBRE	SECUNDARIA	ENTRE 500 Y...	Regular
12	Alcantarillas	SENIOR	35	HOMBRE	SECUNDARIA	ENTRE 500 Y...	Regular
13	Alcantarillas	SENIOR	38	HOMBRE	SUPERIOR	MAYOR A 1200	Bueno
14	Alcantarillas	SENIOR	37	MUJER	ESPECIALID...	MAYOR A 1200	Bueno
15	Ordenanzas ...	SENIOR	36	MUJER	SUPERIOR	ENTRE 500 Y...	Muy Bueno
16	Ordenanzas ...	MAYOR	45	HOMBRE	ESPECIALID...	ENTRE 500 Y...	Muy Bueno
17	Cuidado Vial	MAYOR	47	HOMBRE	PRIMARIA	INGRESOS M...	Muy Bueno

**Fuente:** Autora

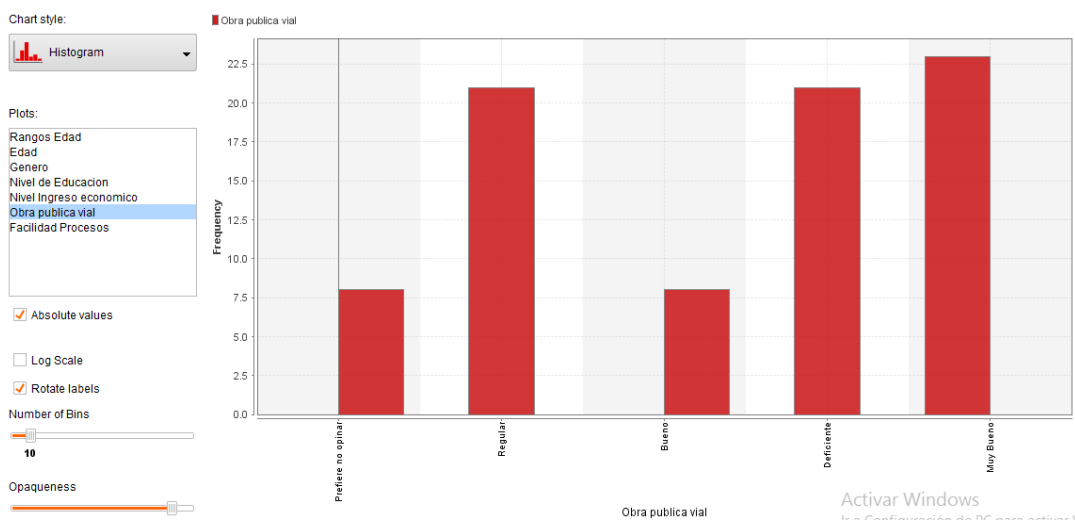
Al determinar las variables de Obra pública vial versus las facilidades de procesos podemos determinar que la necesidad de la ciudadanía es el Asfalto, seguido de las alcantarillas y en tercer lugar irían las ordenanzas y cuidado vial, como podemos observar en la Figura No.34.



**Figura 34.** Grafica Facilidad Procesos vs.Obra vial

**Fuente:** Autora

Al determinar un histograma con la variable obra pública vial versus la muestra tomada de la ciudadanía se puede observar que las opiniones de la satisfacción de los servicios solicitados son proporcionales entre tres criterios, muy bueno, regular y Deficiente; siendo el más bajo la opción de no dar su criterio, como podemos observar en la Figura No.35.

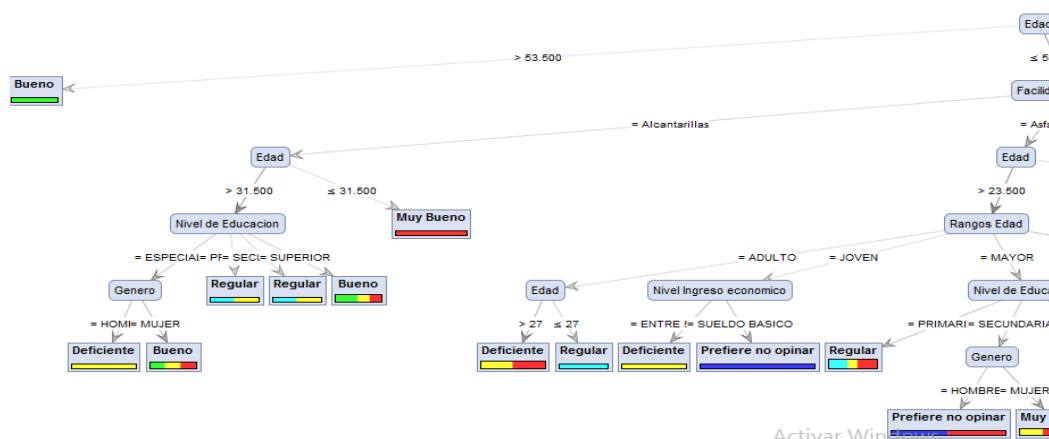


**Figura 35.** Grafica Obra Publica vial

**Fuente:** Autora

Arboles de Decisión:

Los resultados nos permiten determinar que las personas menores a 31 años en el rango de nivel de instrucción están los que tienen primaria y secundaria creen que el alcantarillado está en un rango de bueno; mientras los que tienen nivel de instrucción superior, consideran que es regular y bueno, como podemos observar en la Figura



No.36.

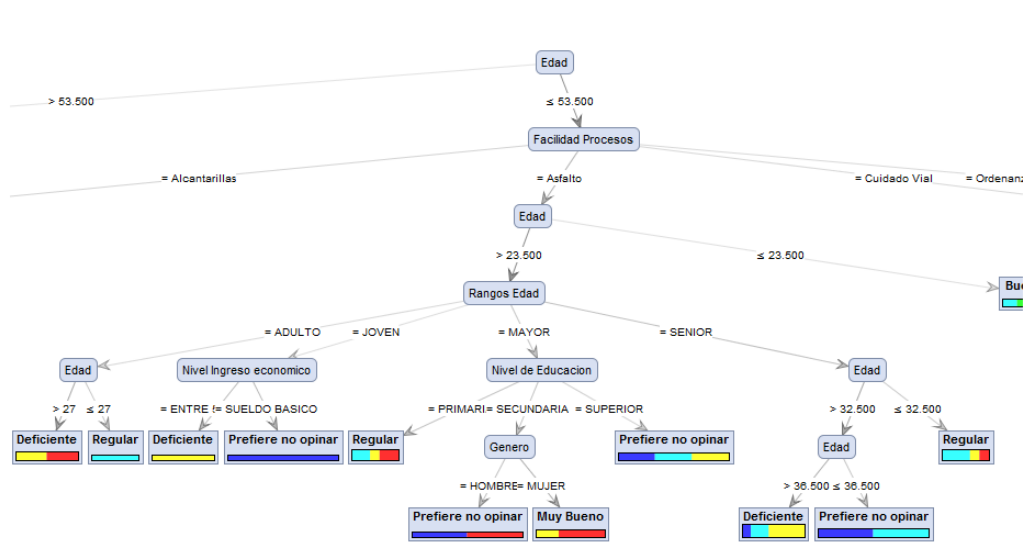
**Figura 36.** Árboles de Decisión Facilidad de Procesos

**Fuente:** Autora

El igual la factibilidad de procesos en cuanto a Alcantarillado a las personas que están dentro de los rangos de edad adulto joven mayor y sénior; hace que interactúen con las variables nivel económico básico; que son quienes no prefieren opinar o piensan que es deficiente.

Según el nivel de educación primaria, secundaria y superior piensan que la facilidad de procesos es regular o prefieren no opinar.

En cuanto al género los hombres prefieren no opinar, mientras que las mujeres piensan que es muy bueno, como podemos observar en la Figura No.37.

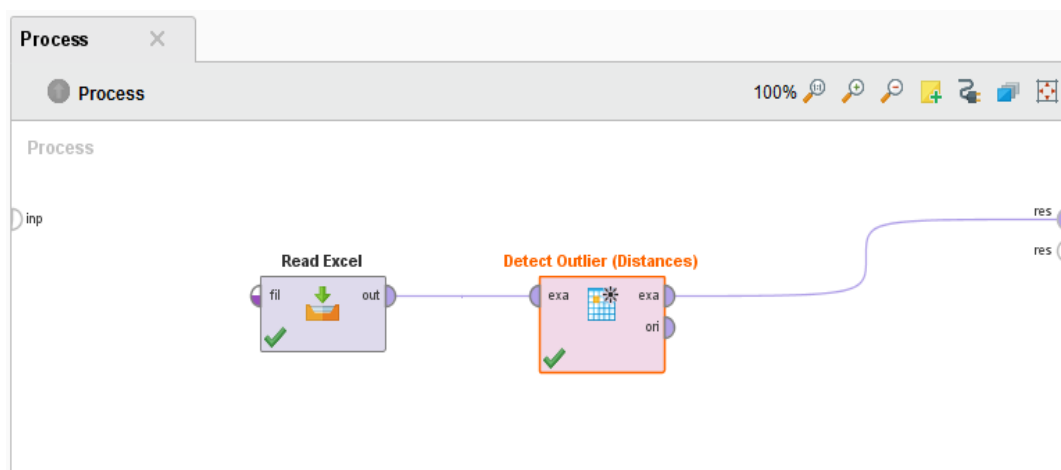


**Figura 37.** Grafica Facilidad Procesos vs. Nivel Económico - Nivel de Educación

**Fuente:** Autora

### 3.3.4. Verificar la calidad de los Datos

Para verificar la calidad de los datos se determina mediante valores atípicos los cuales son detectados mediante la herramienta con el operador outlier, como podemos observar en la Figura No.38.

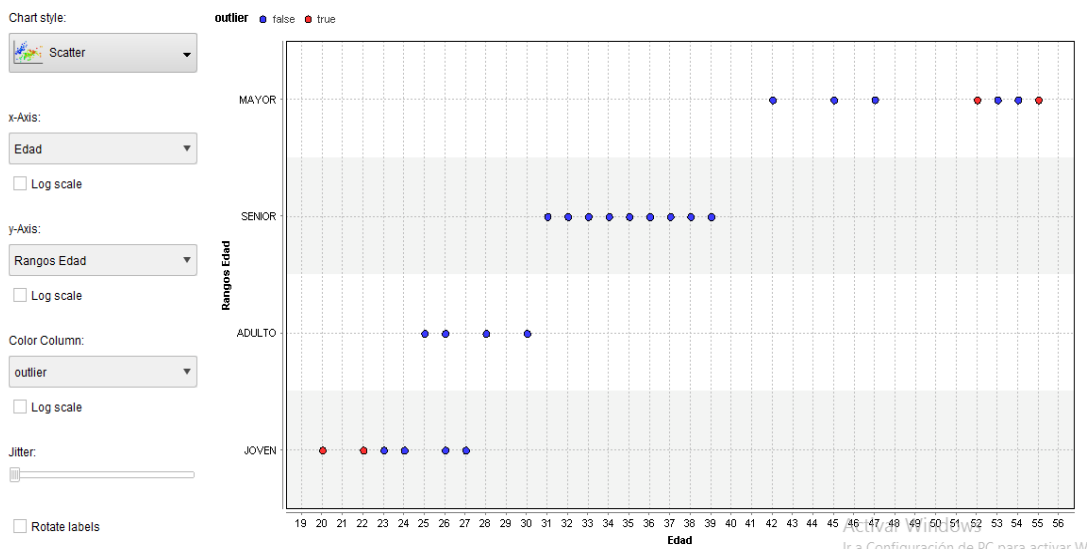


**Figura 38.** Grafica desarrollo en Rapidminer

**Fuente:** Autora

Al realizar la verificación mediante dicho operador obtuvimos los siguientes valores atípicos.

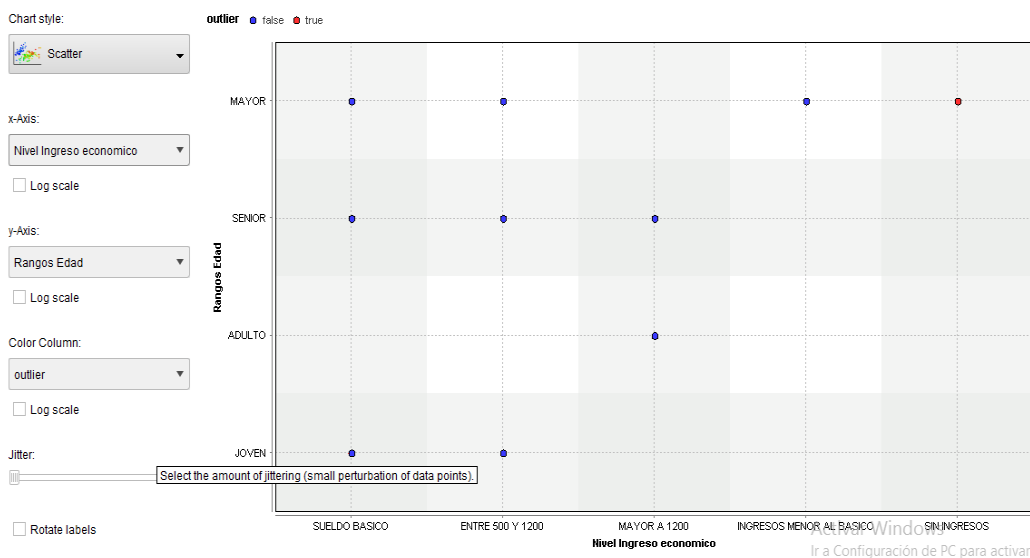
Con respecto a la Edad los valores tienen una similitud a excepción de las personas que están en el rango de Edad Mayores que tienen edades de 52 y 55, como podemos observar en la Figura No.39.



**Figura 39.** Graficas de Valores Atípicos Rangos Edad

**Fuente:** Autora

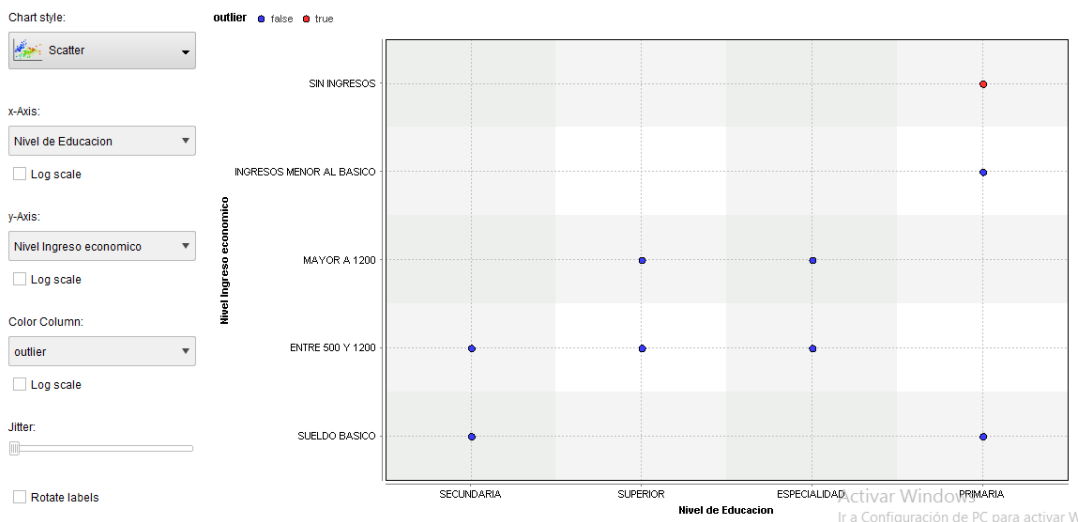
Al determinar la variable Nivel de Ingreso Económico versus la Edad se puede determinar que el rango de edad no varía de su rango Mayor y este se relaciona con un nivel de ingresos bajos, como podemos observar en la Figura No.40.



**Figura 40.** Grafica de Valores Atípicos Nivel ingresos

**Fuente:** Autora

También relacionamos el nivel de educación con el nivel de Ingresos, de lo que se pudo determinar que las personas sin ingresos y que tienen una instrucción primaria son valores fuera de los rangos normales, como podemos observar en la Figura No.41.



**Figura 41.** Grafica de Valores Atípicos Nivel de Educación

**Fuente:** Autora

Pero estos valores atípicos no varían la muestra de una manera considerable ya que la mayor parte de la población si está dentro de las características por lo que los resultados que arroja la herramienta son cercanos a la realidad.

### **3.4. Preparación de datos**

En esta fase se trata de tener los datos listos para adecuarlos a las técnicas de minería de datos que se van a emplear sobre ellos. Esto implica seleccionar la muestral que se va a utilizar, limpiarlos para mejorar su calidad, añadir nuevos datos a partir de los existentes y darles el formato requerido por la herramienta de modelado.

#### **3.4.1. Selección de Datos**

Se van a utilizar todos los registros de la tabla Persona que compone la base de datos, ya que al ser ésta una base de datos específicamente creada para este proyecto, el número de registros que se han insertado ha sido el de la muestra tomada del sector Quitumbe. Sin embargo, hay campos dentro de estos registros que no son necesarios para nuestros objetivos de minería de datos, por lo que se puede prescindir de algunos de ellos.

Estos son los campos mas utilizados para la mineria de datos, ya que nos permiten cumplir con los objetivos inicialmente mencionados.

#### **Tabla Persona**

- per\_nombres
- per\_referencia\_web
- per\_nivel\_instruccion
- per\_tipo\_de\_ingreso
- per\_fecha\_nacimiento
- per\_genero

### 3.4.2. Limpieza de Datos

La muestra tomada en el sector Quitumbe, contiene toda la información necesaria para cumplir los objetivos de la minería de datos. Por lo tanto no hay necesidad de hacer una limpieza más profunda sobre ellos, ya que se ha determinado los objetivos específicos dentro de las preguntas de la encuesta.

Las inconsistencias descubiertas, modificadas o eliminadas se han realizado en el proceso de ingreso de las encuestas validando la integridad de los datos.

Esta grafica nos permitirá reflejar la limpieza de nuestra información, Figura No.42.



**Figura 42:** Gráfica Proceso vs. Análisis

**Fuente:** Limpieza-y-transformacin-de-datos

### 3.4.3. Construir Datos

Los Atributos derivados nos permiten determinar el tipo de datos de los campos, y realizar comparaciones entre variables e indicadores; pero en nuestras tablas no se realizó ningún tipo de transformación ni tampoco se agregaron campos adicionales.

Los datos se relacionaron entre variables dependientes e independientes verificando de esta manera el uso de la tecnología en la ciudadanía versus el nivel de Estudios, y otras relaciones.

Para los Registros generados no ha sido necesario generar nuevos atributos ni integrar nuevos registros a la base de datos ya que ésta completa y ha sido creada específicamente a partir de las encuestas, para su uso en este proyecto.



### 3.4.4. Integrar Datos

Se crearon tablas a partir de las encuestas donde se interactúa con las variables e indicadores, bosquejo de la correlación de variables, como podemos observar en la Tabla No.9.

**Tabla 9:**  
Indicadores y Variables

	A	B	C	D	E	F	G	H	I	J	K	L	M
		GENERO		PARTICIPACION DE LA TECNOLOGÍA PARA OPINAR									
	población	M	F	SI_TM	NO_TM	SI_EP	NO_EP	V1	V2				
15-20	30	15	15	3	27	1	29	30	30				
21-30	200	120	80	25	175	5	195	200	200				
31-45	180	80	100	7	173	5	175	180	180				
46-55	25	10	15	2	23	2	23	25	25				
56-70	15	7	8	2	13	1	14	15	15				
	450	232	218	39	411	14	436	450	450				
					450		450						
		GENERO		NIVEL DE INSTRUCCIÓN			PARTICIPACION DE LA TECNOLOGÍA PARA OPINAR						
	población	SA	P	SEC	SUP	ESP	SI_TM	NO_TM	SI_EP	NO_EP	V1	V2	
15-20	30	6	10	14	0	0	3	27	1	29	30	30	
21-30	200	20	120	40	20	0	25	175	5	195	200	200	
31-45	180	5	105	30	35	5	7	173	5	175	180	180	
46-55	25	5	7	5	4	4	2	23	2	23	25	25	
56-70	15	6	8	1	0	0	2	13	1	14	15	15	
	450	42	250	90	59	9	39	411	14	436	450	450	

### 3.4.5. Formatear Datos

No fue necesario modificar el orden de los campos, ni tampoco de sus tablas ya que al utilizar el software Rapidminer la información fue admitida sin ningún error, nos permitió elegir el formato ya obtenido y procesado luego de las encuestas.

## 3.5. Modelado

En esta fase de la metodología se escogerá la técnica (o técnicas) más apropiadas para los objetivos marcados de la minería de datos, una vez realizado un plan de prueba para los modelos escogidos, se procederá a aplicar dichas técnicas sobre los datos para generar el modelo y por último se obtendrá los resultados con éxito o no.

### 3.5.1. Selección de la Técnica de Modelado

En esta fase de la metodología se escogerá la técnica (o técnicas) más apropiadas para los objetivos marcados de la minería de datos. A continuación, y una vez realizado un plan de prueba para los modelos escogidos, se procederá a aplicar dichas técnicas sobre los datos para generar el modelo y por último se tendrá que evaluar si dicho modelo ha cumplido los criterios de éxito o no.

Lo más laborioso de esta última etapa fue determinar los operadores que intervienen en cada fase del proyecto, el tipo de variable entre ellas polinomial, entera, id, numerica.

Investigar el uso de las funciones se extendió un poco el tiempo de la culminación de la exploración de los datos. Sin embargo, se tuvo puntos positivos porque la herramienta al ser muy flexible permitiendo realizar graficas de modelos cosas que en otros programas tomaría mucho tiempo en realizarlos.

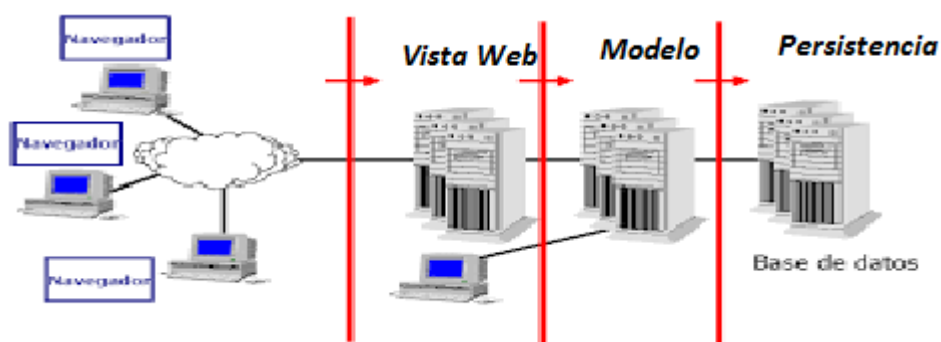
Software de apoyo y soporte para la automatización de la información se ha implementado para complementar los procesos de enlace y vinculación de la información y su tratamiento

Esto contempla plataformas JEE Java Enterprise Edition, basados en el patrón MVC Modelo Vista Controlador que contiene la arquitectura, como podemos observar en la Figura No.45.

Vista que hace referencia a la capa de presentación que utiliza tecnología web JSF Java Server Face con el framework primeface

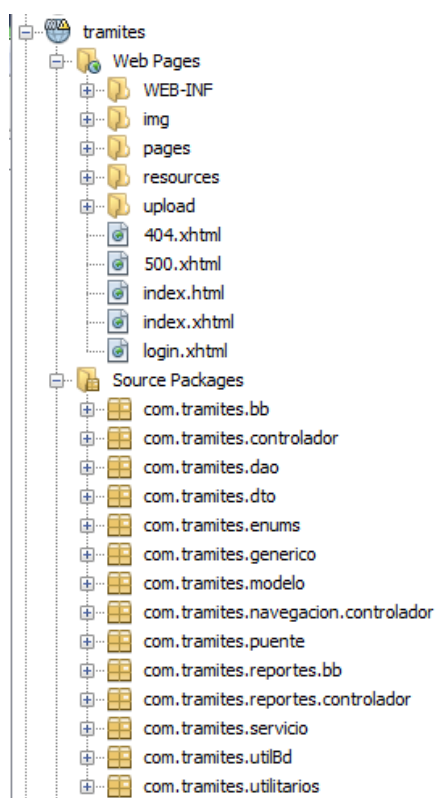
Controlador conformado por tecnología EJB Enterprise Java Beans donde están plasmados los procesos principales de inter acción entre la Lógica de Negocio y Reglas de Negocio

Modelo basado en sistema de entidades y servicios donde se incluyen componentes de persistencia de datos en arquitecturas JEE, como podemos observar en la Figura No.43.



**Figura 43.** Modelo Vista Presentación

**Fuente:** modelo-vista-controlador



**Figura 44.** Estructura / arquitectura del componente EJB Module

**Fuente:** Autora

### 3.5.2. Generación de la prueba de diseño

Para la generación de la prueba de diseño se utilizó la herramienta Excel, que nos permite crear un modelo de regresión lineal con nuestros indicadores, Género, Nivel

de Educación, Nivel Económico, Nivel de Instrucción; entre otros que cumple las siguientes reglas.

- Que la relación entre las Variables sea lineal.
- Que los errores en la medición de las variables explicativas sean independientes entre sí.

Toda función de la forma:  $Y_i = B_0 + X_i B_i$  determina, al representarla en el plano una línea recta, donde X e Y son variables y a y b son constantes.

Al analizar las distintas graficas obtenidas en la que los puntos de un diagrama de dispersión se encontraran en una recta podría decir que la relación entre los indicadores es dependiente uno del otro.

Al realizar los diagramas, la información que se obtuvo para la determinación de la muestra fue del Consejo Nacional Electoral, donde nos indica cual es la población dividida por sectores y por género. **Ver Anexo2.**

### **3.5.3. Construcción del Modelo**

En este apartado se describirán los ajustes de parámetros del modelo que se eligen en la herramienta de minería de datos, así como la salida de información y su descripción.

#### **Regresión Lineal entre Edad y Nivel de Educación.**

La regresión lineal o ajuste lineal es un modelo matemático usado para aproximar la relación de dependencia entre una variable dependiente Y, las variables independientes X y un término aleatorio.

En siguiente cuadro podemos determinar la relación que existe entre las variables Edad y Nivel de Educación y como la Línea de tendencia se relaciona con varios puntos de la tabla.

**Tabla 10:**  
Representation regression Lineal – Pregunta 1

Rangos Edad	Edad	Genero	Nivel de Educación	Nivel Ingreso económico
20 - 25	20	1	3	3
	24	1	3	3
	23	2	3	4
	22	2	4	4
	23	2	4	4
25 - 30	25	1	4	5
	26	1	4	5
	30	1	5	5
	30	1	5	5
	28	2	4	5
30 - 40	31	2	3	4
	35	2	3	4
	38	2	4	5
	37	1	5	5
	36	1	4	4
40 - 60	45	2	5	4
	47	2	2	2
	52	2	3	3
	55	1	2	1
	54	1	3	3

**Fuente:** Autora

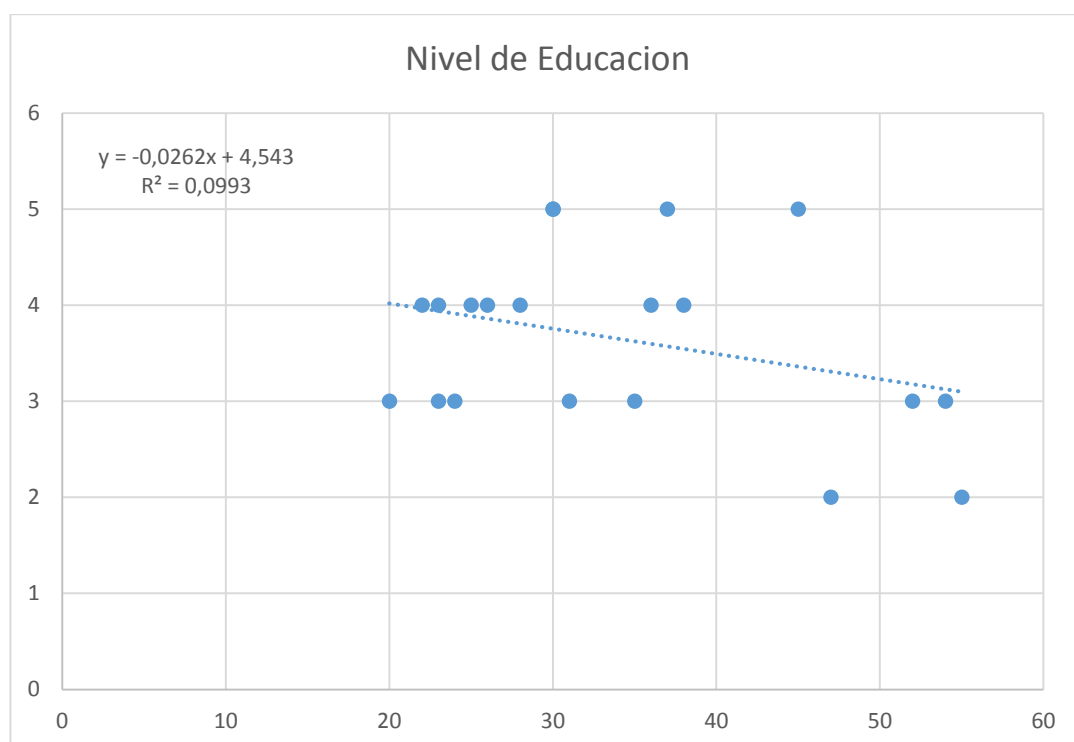
La ecuación que obtuvimos es:

$$y = -0,0262x + 4,543$$

$$R^2 = 0,0993$$

El coeficiente de regression lineal nos permite determinar que la relacion de las variables esta dentro del rango establecido que es de -1,1.

Al tener 0,0993 podemos determinar que las variables no tiene una estrecha relación, es decir no son la edad no se relaciona en un 100% con el nivel de educación, como podemos ver en la figura 45.



**Figura 45.** Gráfica Nivel de Educación

**Fuente:** Autora

### **Regresión Lineal Nivel de Educación vs. Nivel de Ingreso Económico**

Aquí relacionamos estas dos variables dando nombres a los parámetros homologados así como podemos ver en la Tabla No.11

**Tabla 11:**  
Gráfica de Datos Homologados

N. Homologado	Nombre	N. Homologado	Nombre
1	Sin Alfabetización	1	sin Ingresos
2	Primaria	2	Ingresos Menor al Básico
3	Secundaria	3	Sueldo Básico
4	Superior	4	Entre 500 y 1200
5	Especialidad	5	Mayor a 1200

**Fuente:** Autora

La Tabla No.12 que a continuación se muestra nos permite obtener los datos a relacionarse.

**Tabla 12:**  
Representation regression Lineal – Pregunta 2

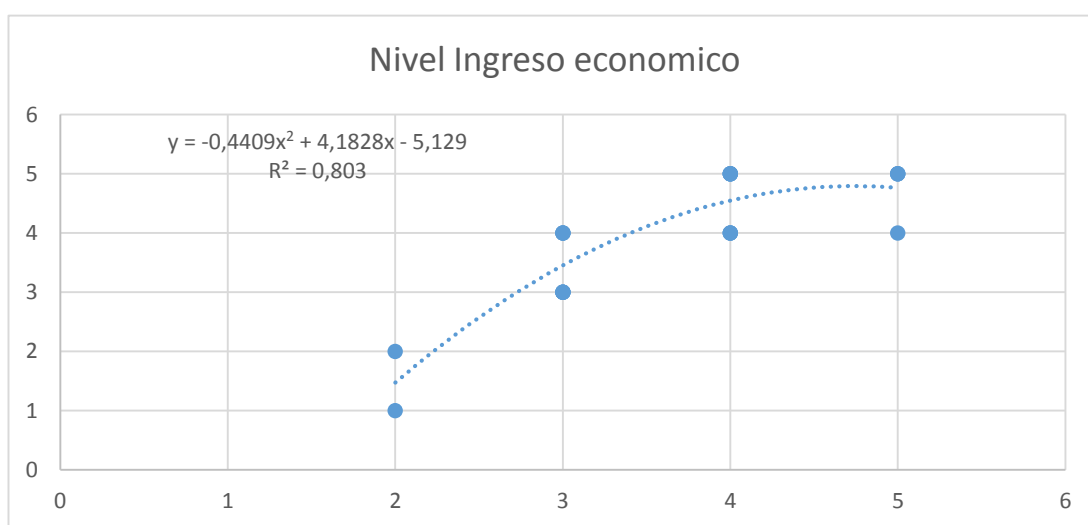
Rangos Edad	Edad	Genero	Nivel de Educación	Nivel Ingreso económico
20 - 25	20	1	3	3
	24	1	3	3
	23	2	3	4
	22	2	4	4
	23	2	4	4
25 - 30	25	1	4	5
	26	1	4	5
	30	1	5	5
	30	1	5	5
	28	2	4	5
30 - 40	31	2	3	4
	35	2	3	4
	38	2	4	5
	37	1	5	5
	36	1	4	4

40 - 60	45	2	5	4
	47	2	2	2
	52	2	3	3
	55	1	2	1
	54	1	3	3

**Fuente:** Autora

En la gráfica podemos observar que la ecuación y el coeficiente de regresión están altamente relacionados:

Es decir podemos ver que el coeficiente de regresión lineal es 0,803 lo que dentro del rango de -1,1 mantienen una estrecha relación, además que el gráfico adopto una forma polinómica, así como podemos ver en la Figura No.46.



**Figura 46.** Gráfica Nivel de Ingreso Económico

**Fuente:** Autora

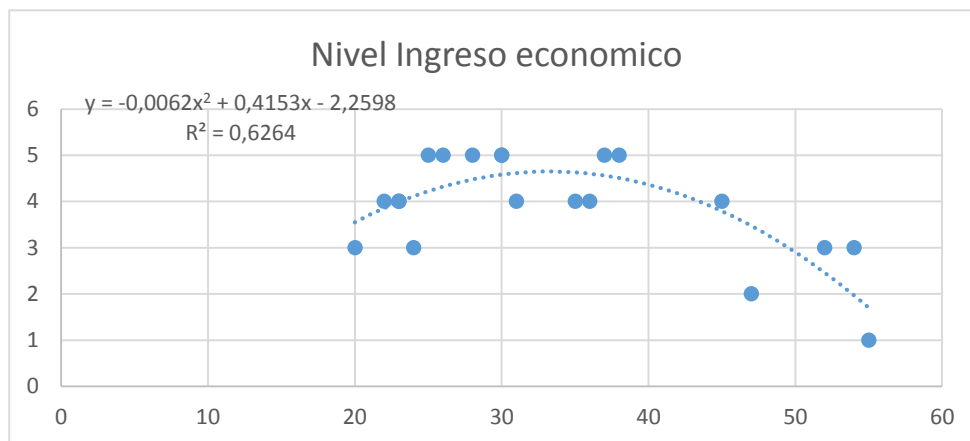
La misma Tabla relacionada con diferentes variables:

- Edad
- Nivel de Ingreso Económico.

La relación no es tan estrecha pero si se puede describir que a medida que la edad esta entre el rango de 20 a 45 años el ingreso económico es bueno. Al igual el gráfico se



adaptó a una línea de tendencia polinómica, así como podemos observar en la Figura No.47.



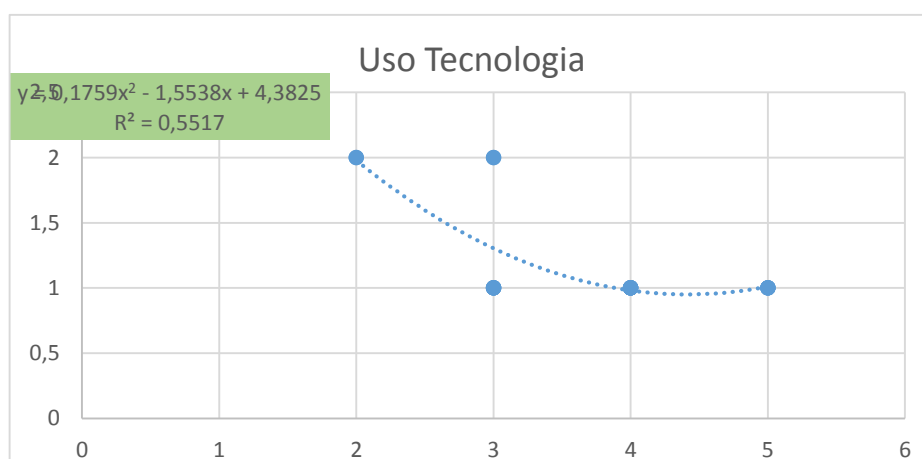
**Figura 47.** Gráfica Nivel de Ingreso Económico vs. Edad

**Fuente:** Autora

La regresión Lineal entre las variables Uso de la Tecnología y Nivel de Educación, siendo las variables homologadas las siguientes:

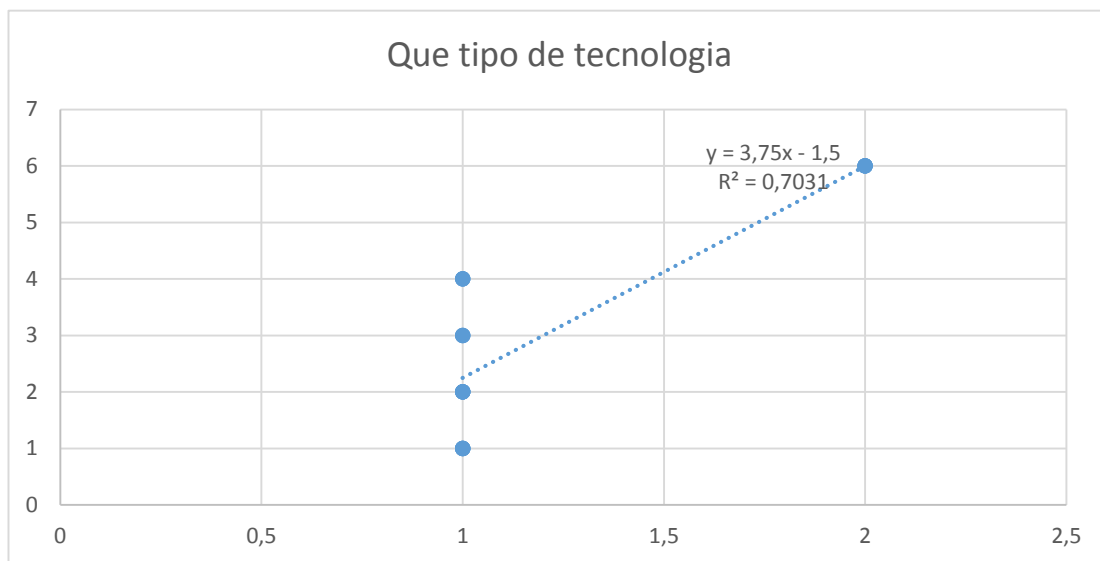
- 1 SI
- 2 NO

Y Nivel de educación las mencionadas en la Figura 48. Podemos determinar que la educación está relacionada en más de un 50% con la variable Uso de la tecnología, lo que demostraría que nuestro estudio está bien relacionado tanto con las encuestas y con la herramienta de minería de datos.



**Figura 48.** Gráfica de Uso de la Tecnología**Fuente:** Autora

Al relacionar la Variable Uso de la Tecnología con el tipo de tecnología que utiliza la ciudadanía podemos determinar que están estrechamente relacionados, así como podemos ver en la Figura No.49.

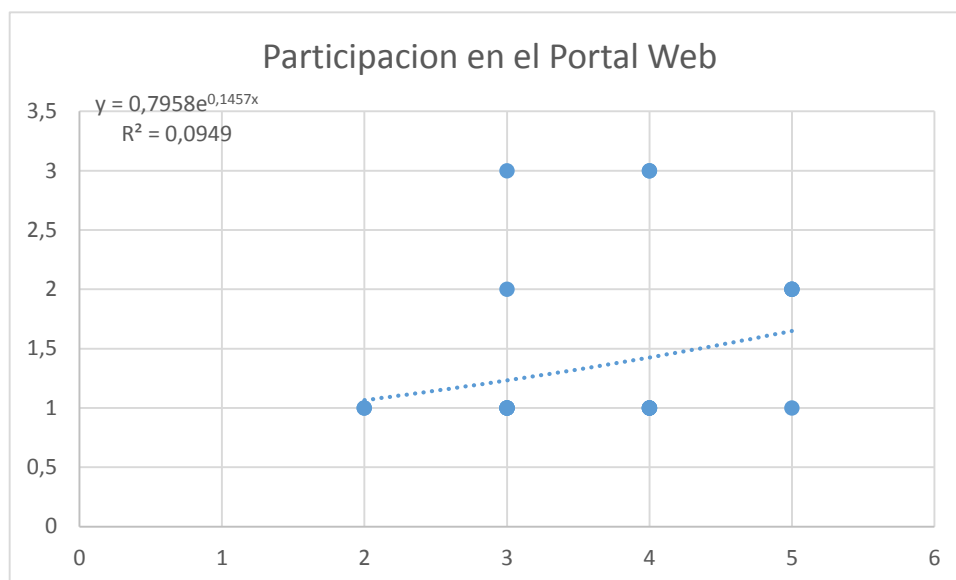
**Figura 49.** Gráfica que tipo de Tecnología Usa**Fuente:** Autora

Al relacionar el portal web con el uso de la tecnología podemos determinar que la ciudadanía no tiene mucho conocimiento de esta herramienta ni de su funcionamiento.

Sus datos homologados son:

- 1 Ninguna
- 2 1 – 3
- 3 3 – 5
- 4 5 a mas

La grafica que acontinuacion observamos nos determina el poco conocimiento que la ciudadanía tienen sobre el portal web que maneja la consejalia de Quito, asi como podemos observar en la Figura No. 50.



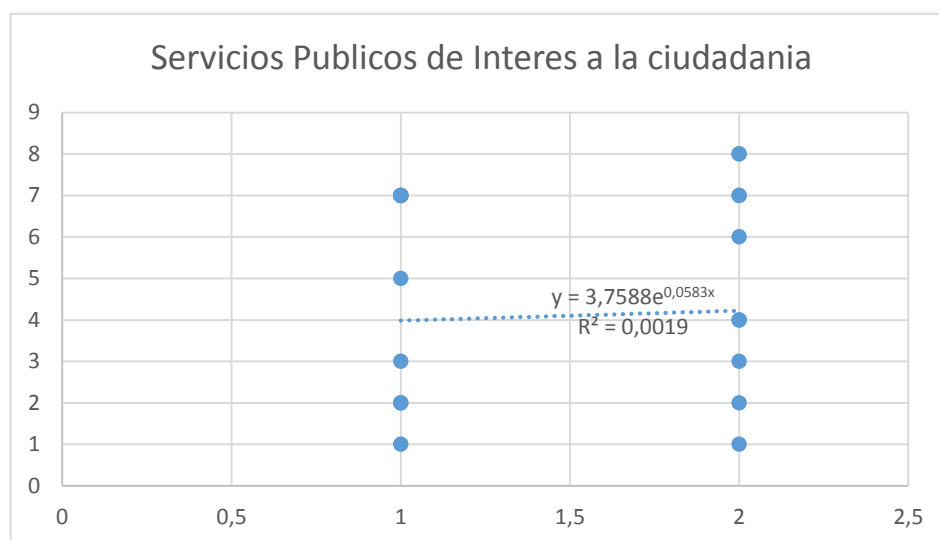
**Figura 50:** Gráfica que tipo de Participación en el Portal Web con la ciudadanía

**Fuente:** Autora

Al relacionar las variables Género y Servicios Públicos de Interés a la Ciudadanía se puede determinar que ninguno de los ciudadanos está interesado en los servicios que presenta el portal web, así como podemos observar en la Figura No.51.

- 1 Servicios Básicos (luz, agua, teléfono)
- 2 Sitios Recreativos
- 3 Casas Comunales
- 4 Educación (colegios municipales)
- 5 Capacitación a la comunidad
- 6 Comercio Informal
- 7 Seguridad y protección contra la delincuencia

- 8 Salud y Acceso a los Servicios de Salud
- 9 Ordenamiento territorial (barrios)
- 10 Obra pública vial de calles (asfalto)
- 11 Otros



**Figura 51.** Gráfica Servicios Públicos vs. Género

**Fuente:** Autora

### 3.5.4. Evaluación del Modelo

La primera Figura.48, nos indica la muestra contabilizada en la herramienta Excel, de donde tomamos las variables Edad vs Nivel de Educación y obtuvimos la siguiente ecuación  $Y = -0.0262x + 4.543$ , nos indica que 4.543 es el punto en el que la recta corta el eje vertical y -0.0262 nos indica el decremento que existe entre las variable Edad y Nivel de Educación mostrando así que las variables no están relacionadas.

En el diagrama de dispersión donde el eje X es la Edad y el eje Y es el Nivel de Educación, a simple vista podemos determinar que las personas que están entre los 20 y 30 años de edad tienen una relación positiva entre ambas variables, pero no dependientes.

Al ver que existe poca dispersión entre las variables el nivel de Educación con respecto a la Edad, se trazó una recta de regresión y ajuste; lo que nos permitió determinar con exactitud la relación entre las dos variables obteniendo el coeficiente de determinación igual a:

$R^2 = 0.0993$ , dicho valor es una medida estandarizada que toma valores entre cero y uno, 0 cuando son independientes y 1 cuando existe relación entre ellas; y cómo podemos observar el valor es demasiado bajo por lo que se confirma que las variables analizadas no están relacionadas.

Al analizar la Figura.49 de las variables Nivel de Educación vs. Ingresos Económicos se determinó la siguiente ecuación polinómica  $Y = -0.4409x^2 + 4.1828x - 5.129$  que es una parábola abierta para abajo, permite que el diagrama de dispersión; reúne algunos puntos que están relacionados.

Se analizó las diferentes formas que puede tomar la gráfica siento la forma polinómica la mejor, ya que al calcular el coeficiente de determinación :  $R^2 = 0.803$ , nos indica que tienen una relación estrecha las variables mencionadas.

Al analizar la Figura.50 de las variables Nivel Económico vs. Edad se obtuvo la siguiente ecuación,  $Y = -0.0062x^2 + 0.4153x - 2.2598$ , dicha tomó la forma polinómica para obtener un diagrama de dispersión donde se pudo verificar que la concentración de puntos esta entre los rangos de edad de 20 a 30 años.

Al calcular el coeficiente de determinación  $R^2 = 0.6264$  se puede determinar que las variables no están relacionadas directamente, es decir son independientes.

Al analizar la Figura.50 de las variables uso de la Tecnología vs. Nivel de la Educación se obtuvo la siguiente Ecuación  $Y = -0.1759x^2 - 1.5538x + 4.3825$ ; la misma nos permitió determinar que las personas con estudios Superiores y Especialidades son las que más utilizan la tecnología; a diferencia de las otras personas que tienen otro tipo de instrucción.

Al calcular el coeficiente de determinación  $R^2 = 0.5517$  se puede determinar que las variables no están relacionadas directamente, pero que si marcan una diferencia, entre los diferentes niveles de Educación.

### **3.6. Evaluación**

En esta fase de la metodología se intentan evaluar los modelos generados pero en esta ocasión la evaluación se hace desde el punto de vista de los objetivos de negocio en lugar de los objetivos de minería de datos. Una vez realizada esta evaluación, se debe decidir si los objetivos han sido cumplidos y de ser así se puede avanzar a la fase de implantación.

#### **3.6.1. Evaluación de los Resultados**

En esta parte del proyecto pasamos a la parte práctica, donde se irá aplicando cada una de las fases de la metodología CRISP-DM al problema práctico que nos planteamos, que es la extracción y explotación de datos.

##### Comprensión del negocio

El municipio de Quito es uno de los entes principales del estado ecuatoriano, es el encargado de cumplir la tarea legislativa para la aprobación de ordenanzas, resoluciones y acuerdos en el Distrito Metropolitano de Quito, además ofrece un portal WEB que permite a los usuarios fomentar la participación de la ciudadanía y satisfacer sus necesidades.

El objetivo de la minería de datos que se va a aplicar en este proyecto es el de hacer predicciones lo más fiables posible a partir de los datos que se obtienen de las encuestas realizadas a la muestra obtenida del Sector Quitumbe.

El objetivo es determinar un modelo predictivo mediante el análisis de las solicitudes realizadas al Municipio de Quito para determinar el nivel de satisfacción y la interacción de la comunidad con el portal web.

### **3.6.2. Revisión del negocio**

El Municipio de Quito es consciente de la problemática que existe hoy en día, los ciudadanos desconocen el tiempo promedio que se tarda en resolver un requerimiento pedido para su barrio, lo que se refleja en las encuestas el momento de analizar su satisfacción por agilidad de respuesta a sus peticiones.

Además, las personas indican que a mayor tiempo se prolongue una petición realizada, el barrio seguirá con deficiencia en seguridad, servicios básicos u otro requerimiento realizado al Municipio.

Adicionalmente se tiene que agregar otros factores que intervienen en el proceso tal como las aprobaciones de las autoridades y el tiempo que se demoran en ser ejecutadas las obras.

### **3.6.3. Determinación de los Resultados**

Verificar el nivel de satisfacción de los ciudadanos mediante las encuestas realizadas a la muestra tomada del Sector de Quitumbe.

Formular un modelo predictivo que permita saber el nivel de interacción con el portal web del Municipio de Quito para obtener información de sus peticiones realizadas.

Determinar los niveles de satisfacción e interacción de los ciudadanos, mediante indicador que permitan obtener registros ya sea por su Género, Edad, Nivel de Instrucción.

El desarrollo del modelo de la minería de datos puede ser de mucha utilidad a la hora de aplicar nuevas técnicas ya que nos va permitir mejorar el servicio de interacción del portal con la ciudadanía y de esta manera mejorar la satisfacción de las peticiones realizadas. Todo esto permitirá mejorar la calidad de los servicios ofrecidos por la institución.

Las herramientas que se va a utilizar para llevar a cabo este proyecto de minería de datos son: SQL Server, Rapidminer ya que se adaptan bien a la metodología que estamos empleando.

### **3.7.Implantación**

Esta es la última fase de la metodología CRISP-DM y su objetivo es explicar a los ciudadanos como poner en funcionamiento el proyecto que se ha construido en las fases anteriores, así como exponer los resultados obtenidos a la ciudadanía de forma que lo pueda entender fácilmente. Otro objetivo de esta fase es el de crear una estrategia para el mantenimiento del proyecto y producir un informe en el que se incluyan posibles mejoras para el futuro y un listado de las dificultades encontradas a la hora de realizarlo.

#### **3.7.1. Planear la Implantación**

Para poder implantar este proyecto en el negocio real sería necesario en primer lugar tener acceso directo a las bases donde se encuentra la información sectorizada, y con el número de personas que conforman el barrio.

A partir de ahí, los pasos a seguir serían los mismos que se han seguido en este documento desde la comprensión del negocio hasta la implantación. Si bien, cabe decir que habrá algunas fases, que se deba procesar nuevamente por el volumen de información ya que tendrán muchos más registros.

#### **3.7.2. Planear la Monitorización y Mantenimiento**

Como plan de supervisión y mantenimiento se podría establecer los siguientes procesos:

- Almacenamiento de requerimientos enviados por la ciudadanía que luego serán ingresados en hojas de datos Excel para luego ser procesados por las herramientas.
- Distribución de los datos en función de los modelos de software de minería de datos a trabajar.



- Los archivos de la explotación de datos deberán ser guardados, por ejemplo, almacenándolos en carpetas ordenadas por tipo de requisitos.
- Los resultados obtenidos en cada explotación de datos deberán ser llevados a formato de hoja de cálculo y generar gráficas de distintos tipos para una mejor visualización e interpretación de los resultados obtenidos en cada periodo.
- El uso de la metodología CRISP-DM en este proyecto ha permitido encontrar un comportamiento predictivo de cómo los ciudadanos interactúan con el portal web siendo la variable más importante el nivel de instrucción ya que luego de todas las pruebas realizadas se determine que a mayor grado de instrucción más visitado será el portal web.

### **3.7.3. Informe Definitivo del Producto**

**Nombre de la Organización:** Universidad de las Fuerzas Armadas Espe

**Título del Proyecto:** Manejo de Indicadores de Participación Ciudadana con el Uso de la Tecnología

**Socios Involucrados en la Implementación del Proyecto:**

**Fecha de Informe (Mes/Año):** mayo 2017

#### **COMENTARIOS INICIALES**

En la problemática se ha identificado que las variables básicas de referencia se pueden integrar funcionalmente con el uso de la tecnología, dentro de la interacción entre gobierno local y comunidad, de manera que pueda ser, a través de procesos de aplicaciones tecnológicas, un recurso de fácil acceso a los ciudadanos y llegar a mejores niveles de colaboración, en la construcción de una mejor ciudad, conforme lo promueve la vigencia de la Ordenanza No. 102 que dice: “Promueve y regula el sistema metropolitano de participación ciudadana y control social”, los que se materializan en los mecanismos antes mencionados y señalados por la ley. Por esta razón se tiene la necesidad de aportar con técnicas de minería de datos que faciliten de manera técnica, el poder realizar procesos de toma de decisiones en base a patrones de comportamiento de la información.

Inicialmente los gobiernos locales tienen un nivel de dificultad en la formulación del uso de la tecnología en la participación ciudadana dando como resultado una necesidad prioritaria de:

Falta de existencia de un sistema de minería de datos para ayudar en los niveles de medición del uso de la tecnología en relación a variables básicas de análisis y procesamiento de la información.

### **LOGRO DEL PROPÓSITO DEL PROYECTO**

Generar procesos de análisis en el manejo básico de indicadores de participación ciudadana, orientados al uso de la tecnología, como apoyo en la gestión comunitaria.

- Evaluar la utilización y aprovechamiento de la integración de Sistemas de información e indicadores, orientadas al uso de la tecnología, así como del uso de sus recursos y materiales para el procesamiento de información.
- Considerar las Variables que se encuentran en el portal como referencia de la línea base.
- Definir los indicadores de Participación Ciudadana para cada uno de los procesos agrupándolas en categorías y subcategorías.
- Elaborar procesos de análisis de variables e indicadores como herramienta de tomas de decisiones.
- Aplicar algoritmos para elaborar patrones de comportamiento, la relación con el uso de la tecnología como medio para incrementar la participación ciudadana.

<b>Indicador</b>	<b>Real a la Finalización</b>
Identificar la mayor relevancia de las variables	Se determina que el nivel de Educación es el factor más relevante en relación al uso de la tecnología.

Identificar las variables directas e indirectas en orden de prioridad.	Se identifica como variables directas:  Nivel de Educación, Nivel económico, Género, Edad en relación a las variables indirectas que son Uso de la Tecnología y Tipo de Tecnología.
	Nivel de Educación y Nivel Económico son factores preponderantes en el nivel de participación ciudadana entre el gobierno local y la comunidad utilizando la Tecnología.
	La Edad y el Genero tienen también su importancia fundamental como complemento especialmente por rangos de edad y género que determinan mayor o menor apertura al uso de la tecnología y colaboración en la opinión ciudadana.
Patrones de Comportamiento	Se realiza el proceso de Minería de Datos determinándose diferentes tipos de patrones de comportamiento que determinan instancia de proyección y predicción para fortalecer la participación ciudadana; en este estudio de tipo piloto en relación a un estrato de la ciudad.
Procesos de Verificación en la Aplicación del Uso de la Metodología en relación a posteriores impactos de participación Ciudadana	Se realizan técnicas de procesamiento, análisis de los sistemas de información en relación al uso de la tecnología con diferentes medios recursos y dispositivos como herramienta de toma de decisiones después de la aplicación de estándares de verificación y validación de la información.
	Se determina sugerencias y recomendaciones para posteriores estudios inmediatamente subsecuentes para completar ciclos con la tendencia a una normativa de compartimiento institucional en relación al uso de la tecnología en la participación ciudadana.

### Impacto Positivo:

Una visión de manejo que se ajusta al contexto actual indispensable en toda institución y gobierno local en conseguir como meta el incremento de la participación ciudadana.

Si bien es cierto se ha tomado se ha tomado una pequeña referencia en relación a una muestra segmentada como plan piloto, pero recoge una iniciativa completamente positiva e ilustradora para complementar y ampliar el estudio en otras muestras de la ciudad e incorporando otras variables.

Este plan piloto puede servir de referencia no solo al contexto del gobierno local actual sino para referencia a otros gobiernos locales que podrían tomar como modelo de referencia para aumentar la participación ciudadana, considerando como una iniciativa técnica y un aporte positivo de como la tecnología es una herramienta de toma de decisiones.

## **DISCUSIÓN Y RESULTADOS**

Actualmente los portales Web de los gobiernos locales son informativos de la gestión e interactivos en relación a instancia de e-goverment para la administración financiera y política; pero tienen deficiencia en una real participación ciudadana basadas en estándares y tecnologías de inteligencia de negocios dentro de los sistemas de información.

Si bien es cierto el departamento de sistemas de tecnología se encuentra desarrollando e implementando técnicas se sugiere que este estudio se complemente de manera inmediata desde la planificación de nuevas fases con nuevos actores y con el apoyo significativo de cada una de las secretarías del gobierno local.

Se recomienda que los equipos de trabajo sean multidisciplinarios en relación a los objetivos de incremento de participación ciudadana a través de la tecnología y se podría incorporar un componente importante y masificador como son los diferentes medios de comunicación.

## CAPITULO IV

### CONCLUSIONES Y LINEAS DE TRABAJO FUTURO

#### 4.1 Conclusiones

Al realizar los procesos en la herramienta de minería de Datos Rapidminer se determinó que la población entre el rango de 25 y 45 años no utiliza frecuentemente los medios tecnológicos, También se realizó un análisis con la herramienta Excel que nos permitió determinar si existe una estrecha relación entre las variables estudiadas es decir Nivel Económico, Nivel de Instrucción, género que al relacionarles; la brecha de aprendizaje de la ciudadanía con el uso de la tecnología es muy grande, ya que tanto hombres como mujeres no participan activamente con el portal web.

La limpieza de datos y selección de los mismos son de mucha importancia para las predicciones, y determinación de los resultados.

La metodología CRISP-DM es una guía muy detallada que permite llevar el control y alcanzar de los objetivos de la minería de datos paso a paso de esta manera la corrección de los datos atípicos permitió tener predicciones más acertadas a la realidad.

Rapidminer nos proporciona una interfaz de visualización y procesamiento de datos para realizar operaciones de entrada y salida, con los indicadores ya mencionados durante el estudio.

Mediante algoritmos que permiten tener una mejor calidad de datos y así mejores resultados, el programa Rapidminer es una herramienta interactiva que nos permite determinar un buen análisis de los datos,

#### 4.2 Líneas de Trabajo

Se recomienda por tanto utilizar minería de datos para realizar la proyección de todos los datos, en nuestro caso la proyección de cada una de las preguntas realizadas en la encuesta ya que al realizar el cálculo sobre la población total y obtener la muestra nos permitió saber cuál es la satisfacción de la ciudadanía a través del portal sobre los servicios.

Si se desea implementar este tipo de reportes para la Concejalía del Municipio de Quito se debería obtener cada 3 o 6 meses información de las últimas solicitudes ingresada para poder ver cuál es la frecuencia del uso del portal al igual que la satisfacción de la ciudadanía.

## **BIBLIOGRAFIA**

(TALAYA, Águeda Esteban, GARCIA DE MARIAGA, Jesús, NARROS, María José, OLARTE, Cristina, REINARES, Eva y SACO, Manuela, 2006;

TALAYA, Águeda Esteban, GARCIA DE MARIAGA, Jesús, NARROS, María José, OLARTE, Cristina, REINARES, Eva y SACO, Manuela, 2006).

(Groth, Robert Prentice Hall. Han, Jiawei; Micheline Kamber, 1999-2000).

(Albrech, 1979)

(María José Ramírez Quintana José Hernández Orallo., 2003)

(Félix)

(Albrieu, C; Ferrari, S; y Navarro, V; et al., 2014)

(Implementación de los Sistemas de Información Geográfica en la gestión de los Espacios Naturales Protegidos)

(Universidad Politécnica de Puebla UPP.)

(Procesamiento de Transacciones En Línea (OnLine Transaction Processing))

(Herramientas gratuitas para análisis de datos.)

(Minería de Datos Herramientas y Técnicas Prácticas de Aprendizaje de Máquinas.) (datos.)

(Sinnexus Business Intelligence Informática Estratégica.)

(K. Aerts, 2006.)

(AGUILERA ONTIVEROS & LÓPEZ PAREDES, 2017)

(FINITAS)

(Minería de datos)

(Informe del Relator Especial sobre la situación de los derechos humanos y las libertades fundamentales de los indígenas, James Anaya”. En “Promoción y protección de todos los derechos humanos, civiles, políticos, económicos, sociales y culturales, incluid, 2009)

(Abril D., Pérez J., 2007)

(ESTADO DEL ARTE DE LA MINERÍA DE DATOS.)

(Oracle Big Data SQL: One fast query, on all your data.)

(La minería de datos o exploración de datos)

(Fayyad, 1996)

(Tipos de variables)

(Ing. Juan Miguel Moine Dra. Ana Silvia Haedo Dra. Silvia Gordillo)

( Ing. Juan Miguel Moine Ing. Juan Miguel Moine Dra. Silvia Gordillo (UNLP) )

(OCERIN)

( Business Intelligence)



## **GLOSARIO DE TERMINOS**

### Glosario de Terminología de Minería de Datos

**Data cleansing:** Proceso de asegurar que todos los valores en un conjunto de datos sean consistentes y correctamente registrados.

**Data Mining:** La extracción de información predecible escondida en grandes bases de datos.

**Análisis prospectivo de datos:** Análisis de datos que predice futuras tendencias, comportamientos o eventos basado en datos históricos.

**Análisis exploratorio de datos:** Uso de técnicas estadísticas tanto gráficas como descriptivas para aprender acerca de la estructura de un conjunto de datos.

**Outlier:** Un ítem de datos cuyo valor cae fuera de los límites que encierran a la mayoría del resto de los valores correspondientes de la muestra. Puede indicar datos anormales. Deberían ser examinados detenidamente, pueden dar importante información.

**Clustering (agrupamiento):** Proceso de dividir un conjunto de datos en grupos mutuamente excluyentes de tal manera que cada miembro de un grupo esté lo "más cercano" posible a otro, y grupos diferentes estén lo "más lejos" posible uno del otro, donde la distancia está medida con respecto a todas las variables disponibles.

**Modelo analítico:** Una estructura y proceso para analizar un conjunto de datos. Por ejemplo, un árbol de decisión es un modelo para la clasificación de un conjunto de datos.

**Modelo lineal:** Un modelo analítico que asume relaciones lineales entre una variable seleccionada (dependiente) y sus predictores (variables independientes).

**Modelo no lineal:** Un modelo analítico que no asume una relación lineal en los coeficientes de las variables que son estudiadas.

Modelo predictivo: Estructura y proceso para predecir valores de variables especificadas en un conjunto de datos.

Regresión lineal: Técnica estadística utilizada para encontrar la mejor relación lineal que encaja entre una variable seleccionada (dependiente) y sus predicados (variables independientes).

RMSE: Error cuadrático medio y sirve para evaluar un modelo de regresión

MAE: Error absoluto medio y sirve para evaluar la calidad en los modelos de regresión.

