



Universitat d'Alacant
Universidad de Alicante

Metodología para procesos de Inteligencia de Negocios con mejoras en la extracción y transformación de fuentes de datos, orientado a la toma de decisiones

Santiago Leonardo Morales Cardoso



Tesis Doctorales

www.eltallerdigital.com

UNIVERSIDAD de ALICANTE



Instituto Universitario de Investigación en Informática

Escuela Politécnica Superior

**“Metodología para procesos de Inteligencia de Negocios
con mejoras en la extracción y transformación de fuentes
de datos, orientado a la toma de decisiones”**

SANTIAGO LEONARDO MORALES CARDOSO

**Tesis presentada para aspirar al grado de
DOCTOR POR LA UNIVERSIDAD DE ALICANTE**

DOCTORADO EN INFORMÁTICA

Dirigida por:

Dr. Ramón Rizo Aldeguer

Marzo 2019

ÍNDICE GENERAL

ÍNDICE GENERAL	III
ÍNDICE DE FIGURAS	V
ÍNDICE DE TABLAS	VII
AGRADECIMIENTOS	IX
RESUMEN	XI
ABSTRACT	XIII
RESUM.....	.XV
PRÓLOGOXVII
1. CAPÍTULO I: INTRODUCCIÓN	1
1.1. JUSTIFICACIÓN DE LA INVESTIGACIÓN.....	1
1.2. PROBLEMA DE INVESTIGACIÓN	3
1.3. MOTIVACIÓN, OBJETIVOS GENERALES Y ESPECÍFICOS.	4
1.4. METODOLOGÍA DE LA INVESTIGACIÓN E HIPÓTESIS.....	5
1.5. CICLO DE VIDA DE LOS PROYECTOS I.N.	7
1.6. BENEFICIOS ESPERADOS	11
1.7. ESTRUCTURA DEL DOCUMENTO DE TESIS.	12
2. CAPÍTULO II: ESTADO DEL ARTE.....	15
2.1. CONCEPTOS Y TÉRMINOS.	17
2.2. DE LA INTELIGENCIA DE NEGOCIOS.....	27
2.3. SOLUCIONES TRADICIONALES.....	29
2.4. LA EVOLUCIÓN DE LA INTELIGENCIA DE NEGOCIOS (IN).	30
2.4.1. NO DATAWAREHOUSE.	33
2.4.2. IN DE AUTOSERVICIO.	34
2.4.3. REDES SOCIALES.....	35
2.4.4. MOVILIDAD (WEB 3.0)	36
2.5. ESTUDIO DE METODOLOGÍAS DE IMPLEMENTACIÓN Y DESPLIEGUE.	37
2.5.1. METODOLOGÍAS GENÉRICAS.	38
2.5.1.1. PROJECT MANAGEMENT INSTITUTE (PMI).....	38
2.5.1.2. NORMA ISO 21500	41
2.5.1.3. PROYECTOS EN AMBIENTES CONTROLADOS PRINCE2.....	45
2.5.2. METODOLOGÍAS ESPECÍFICAS.....	50
2.5.2.1. “BUILDING THE DATA WAREHOUSE” POR BILL INMON.	50
2.5.2.2. “THE DATA WAREHOUSE TOOLKIT” POR RALPH KIMBALL.	52
3. CAPÍTULO III. ANÁLISIS METODOLÓGICO.....	57
3.1. COMPARACIÓN DE METODOLOGÍAS.	57
3.1.1. METODOLOGÍAS GENÉRICAS	58
3.1.2. METODOLOGÍAS ESPECÍFICAS.....	59
3.2. ANÁLISIS DE HERRAMIENTAS TECNOLÓGICAS IN	61
3.2.1. GARTNER GROUP	61
3.2.2. FORRESTER RESEARCH	65
3.3. CÍRCULOS DE PROCESOS.....	73
3.3.1. APLICACIÓN DE CÍRCULOS DE CALIDAD EN E.T.C,.....	73
3.4. TEORÍA DE LA DECISIÓN.	74
3.5 PROBLEMAS.....	79
3.5.1 SITUACIÓN ACTUAL DE LA INTELIGENCIA DE NEGOCIOS.....	79
3.5.2. PROBLEMAS NO RESUELTOS Y LIMITACIONES.....	80
3.5.3 ¿DÓNDE AVANZAR?	81

4. CAPÍTULO IV. METODOLOGÍA M3S.....	85
4.1. MODELO M3S.....	85
4.1.1. PRINCIPIOS Y REQUISITOS.....	86
4.1.2. PRESENTACIÓN DE M3S	87
4.2. PROPUESTA DE LA METODOLOGÍA M3S.....	88
4.2.1. FASES DE M3S.....	90
4.2.2 ETAPAS DE LA METODOLOGÍA DE DECISIÓN	94
5. CAPÍTULO V: PRUEBAS METODOLOGÍA M3S.....	101
5.1. CASO EMPRESA ELÉCTRICA QUITO (E.E.Q).....	103
5.1.1. APLICACIÓN METODOLOGÍA DE M3S EN EEQ:.....	104
5.1.2. DISCUSIÓN DE RESULTADOS	121
5.1.3. CONCLUSIONES CASO E.E.Q.	123
5.2. CASO CLUB CASTILLO AMAGUAÑA (C.C.A.)	125
5.2.1. APLICACIÓN METODOLOGÍA DE BI M3S EN C.C.A.	126
5.2.1.1 APLICACIÓN METODOLOGÍA DE DECISIÓN:.....	129
5.2.2. DISCUSIÓN DE RESULTADOS	136
5.2.3. CONCLUSIÓN CASO C.C.A.....	137
5.3. TEST DE RESULTADOS	139
5.4. MEDICIÓN DE RESULTADOS.....	140
5.5. DISCUSIÓN DE RESULTADOS.....	159
6. CAPÍTULO VI: CONCLUSIONES Y LÍNEAS FUTURAS	161
6.1. RESUMEN Y CONSIDERACIONES FINALES	161
6.2. CONTRIBUCIÓN REALIZADA.....	163
6.2.1 APORTACIONES GENERALES.....	163
6.2.2 APORTACIONES TEÓRICAS.....	166
6.2.3 APORTACIONES DE TRANSFERENCIA Y DIFUSIÓN DEL CONOCIMIENTO	167
6.3. LÍNEAS FUTURAS.....	170
6.3.1. DE INVESTIGACIÓN.....	171
ANEXOS	173
ANEXO I. ACTA DE CONSTITUCIÓN DE PROYECTO (IN)	175
ANEXO II. INFRAESTRUCTURA HARDWARE Y SOFTWARE (IN)	176
ANEXO III. CRONOGRAMA (IN).....	177
ANEXO IV. ANÁLISIS DEL NEGOCIO (IN).....	178
ANEXO V. FLUJOGRAMA METODOLOGIA DE DECISIÓN (IN).....	179
ANEXO VI. PSEUDOCÓDIGO METODOLOGÍA DE DECISIÓN (IN).....	180
ANEXO VII. INFORMACIÓN FUNCIONAL Y TÉCNICA (IN)	181
ANEXO VIII. ACTA DE CIERRE PROYECTO BI (IN).....	182
ANEXO IX. ASISTENCIA	184
ANEXO X. DATOS ALGORITMO DE DECISION (IN)	185
ANEXO XI. ENCUESTA DE MADUREZ EMPRESARIAL EN PROYECTOS BI.....	186
ANEXO XII. ENCUESTA COMPARATIVA DE METODOLOGÍAS DE BI.....	188
ANEXO XIII. DATOS TABULACIÓN ENCUESTA DE MADUREZ.....	190
ANEXO XIV. DATOS TABULACIÓN ENCUESTA COMPARATIVA - CUALITATIVA.....	192
ANEXO XV. DATOS TABULACIÓN ENCUESTA COMPARATIVA - CUANTITATIVA.....	196
BIBLIOGRAFÍA	199

ÍNDICE DE FIGURAS

<i>Figura 1. Proceso de Investigación.</i>	6
<i>Figura 2. Ciclo de vida de los proyectos I.N.</i>	7
<i>Figura 3. Ciclo de Vida de Metodologías IN. M3S.</i>	11
<i>Figura 4. Pirámide Niveles de Sistemas Información.</i>	27
<i>Figura 5. Evolución de IN en el soporte a la decisión. (Zarate Gallardo, 2013)</i>	32
<i>Figura 6. Concepto PMI.</i>	38
<i>Figura 7. Procesos y Áreas de Conocimiento de PMBOK. (Guarin, 2016)</i>	40
<i>Figura 8. Grupos de Procesos ISO (Brenes, 2013)</i>	42
<i>Figura 9. Directrices Dirección y Gestión Proyectos ISO21500 (Verástegui, 2014)</i>	42
<i>Figura 10. Directrices UNE-ISO 21500:2013 (AENOR, 2013)</i>	44
<i>Figura 11. Los 4 elementos de Prince2. (Vila Grau, 2015)</i>	46
<i>Figura 12. Enfoque de Bill Inmon.</i>	52
<i>Figura 13. Enfoque Kimball.</i>	53
<i>Figura 14. Ciclo de Vida de Metodología Kimball. (Rivaderra G., 2010)</i>	54
<i>Figura 15. Círculos de Calidad en ETC.</i>	73
<i>Figura 16. Pilares de M3S.</i>	86
<i>Figura 17. Parroquias del Distrito Metropolitano de Quito.</i>	108
<i>Figura 18. Clientes por Sectores del Distrito Metropolitano de Quito.</i>	108
<i>Figura 19. Nodo de Primer Nivel Árbol de Decisión. Caso EEQ.</i>	118
<i>Figura 20. Nodo de Segundo Nivel Árbol de Decisión. Caso EEQ.</i>	119
<i>Figura 21. Nodo de Tercer Nivel Árbol de Decisión. Caso EEQ.</i>	121
<i>Figura 22. Resultado Caso de Estudio EEQ.</i>	123
<i>Figura 23. Clientes por Sectores del Distrito Metropolitano de Quito.</i>	129
<i>Figura 24. Nodo de Primer Nivel Árbol de Decisión. Caso C.C.A.</i>	133
<i>Figura 25. Nodo de Segundo Nivel Árbol de Decisión Caso. C.C.A.</i>	134
<i>Figura 26. Nodo de Tercer Nivel Árbol de Decisión. Caso C.C.A.</i>	135
<i>Figura 27. Resultado Caso de Estudio C.C.A.</i>	136
<i>Figura 28. Pregunta 1 – MADUREZ-Objetivos</i>	141
<i>Figura 29. Pregunta 2 – MADUREZ-Eficiencia</i>	142
<i>Figura 30. Pregunta 3 – MADUREZ-Estandar Recolección</i>	142
<i>Figura 31. Pregunta 4 – MADUREZ-Estandar Externo</i>	143
<i>Figura 32. Pregunta 5 – MADUREZ-Hitos.</i>	143
<i>Figura 33. Pregunta 6 – MADUREZ-Metodológicos.</i>	144
<i>Figura 34. Pregunta 7 – MADUREZ-Certificación.</i>	144
<i>Figura 35. Pregunta 8 – MADUREZ-Inversión.</i>	145
<i>Figura 36. Pregunta 9 – MADUREZ-Retroalimentación.</i>	145
<i>Figura 37. Pregunta 10 – MADUREZ-Uso Metodologías.</i>	146
<i>Figura 38. Pregunta 11 – MADUREZ-Presupuesto.</i>	146
<i>Figura 39. Pregunta 12 – MADUREZ-Cambios.</i>	147
<i>Figura 40. Pregunta 13 – MADUREZ-Procesos Mejora.</i>	147
<i>Figura 41. Pregunta 14 – MADUREZ-Estandar Software.</i>	148
<i>Figura 42. Pregunta 1-CCL-Control Financiero.</i>	149

<i>Figura 43. Pregunta 2-CCL-Percepción.</i>	149
<i>Figura 44. Pregunta 3-CCL-Aceptación.</i>	150
<i>Figura 45. Pregunta 4-CCL-Acogimiento.</i>	150
<i>Figura 46. Pregunta 5-CCL-Valor.</i>	151
<i>Figura 47. Pregunta 6-CCL-Debilidades.</i>	152
<i>Figura 48. Pregunta 7-CCL-Fortalezas.</i>	152
<i>Figura 49. Pregunta 8-CCL-Robustez.</i>	153
<i>Figura 50. Pregunta 1-CCU-Personal.</i>	154
<i>Figura 51. Pregunta 2-CCU-Extracción.</i>	154
<i>Figura 52. Pregunta 3-CCU-Depuración.</i>	155
<i>Figura 53. Pregunta 4-CCU-Construcción.</i>	155
<i>Figura 54. Pregunta 5-CCU-Validación.</i>	156
<i>Figura 55. Pregunta 6-CCU-Costos.</i>	156
<i>Figura 56. Pregunta 7-CCU-Costo Hardware.</i>	157
<i>Figura 57. Pregunta 8-CCU-Usuarios.</i>	157
<i>Figura 58. Pregunta 9-CCU-Usuarios Activos.</i>	158
<i>Figura 59. Pregunta 10-CCU-Aplicativos.</i>	158
<i>Figura 60. Pregunta 11-CCU-Solicitudes.</i>	159
<i>Figura 61. Etapas de Metodología de Decisión.</i>	165



Universitat d'Alacant
Universidad de Alicante

ÍNDICE DE TABLAS

Tabla 1: Ciclo de Vida de Proyectos IN. (Vanrell, Bertone, & Ramón, 2010)	9
Tabla 2: Casos de Proyectos IN (Propia, 2016)	13
Tabla 3. Redes sociales más populares en Ecuador. (Marboleda, BRANDEC, 2017)	36
Tabla 4. Usuarios en redes sociales. Ecuador-2017 (Multiplicalia, Multiplicalia, 2017).....	36
Tabla 5. Cronología de Origen Norma ISO 21500 (Coquillat, 2014) 	43
Tabla 6. Los 7 principios de Prince2. (Aqueda Barrero, 2014)	47
Tabla 7. Las 7 Temáticas de Prince 2. (Vila Grau, 2015).....	47
Tabla 8. Los 7 Procesos de Prince 2. (Propia, 2016).....	48
Tabla 9. Adaptación de diferentes elementos de Prince 2. (Vila Grau, 2015)	49
Tabla 10. Etapas Metodología Kimball.....	55
Tabla 11. Cuadro Comparativo. Metodología Genéricas (Cruz Caicedo, 2012).....	58
Tabla 12. Comparativo Modelos Inmon y Kimball (Propia, 2016)	60
Tabla 13. Interpretación Gartner BI 2009.....	63
Tabla 14. Interpretación Gartner BI 2013.....	64
Tabla 15. Interpretación Gartner BI 2016.....	65
Tabla 16. Interpretación Forrester Research Inc. 2008.....	66
Tabla 17. Interpretación Forrester Research Inc. 2012.....	67
Tabla 18. Interpretación Forrester Research Inc. 2015.....	68
Tabla 19. Cuadro Comparativo de Herramientas de IN (Propia, 2016).....	70
Tabla 20. Cuadro Comparativo Soluciones DW vs. In Memory. (Propia, 2016).....	72
Tabla 21: Ciclo de Vida de Metodología M3S.....	90
Tabla 22. Etapas de Metodología de Decisión. (Propia, 2016).....	97
Tabla 23. Reglas de Negocio en base a Consumo y Otros.	110
Tabla 24. Tabla de Entrenamiento Caso E.E.Q.	111
Tabla 25. Cálculo de Entropía y Ganancia de Información.....	112
Tabla 26. Hechos Reales Caso EEQ.	112
Tabla 27. Cálculo Entropía y Ganancia Nodo: Zona Campaña.	114
Tabla 28. Cálculo Primer Nivel Entropía y Ganancia Nodo: Tipo Identificación.	115
Tabla 29. Cálculo Primer Nivel Entropía y Ganancia Nodo: Excede Consumo Promedio	116
Tabla 30. Cálculo Primer Nivel Entropía y Ganancia Nodo: Sector Campaña.	117
Tabla 31. Resumen de Cálculos Primer Nivel EEQ.	117
Tabla 32. Cálculo Segundo Nivel Entropía y Ganancia Nodo: Zona Campaña.	118
Tabla 33. Resumen de Cálculos Segundo Nivel EEQ.	119
Tabla 34. Cálculo Tercer Nivel Entropía y Ganancia Nodo: Zona_Campaña-Tipo Identificación	120
Tabla 35. Resumen de Cálculos Tercer Nivel EEQ.	120
Tabla 36. Reglas de Negocio. C.C.A.	130
Tabla 37. Tabla de Entrenamiento CLUB.	131
Tabla 38. Hechos Reales Caso C.C.A.	132
Tabla 39. Resumen de Cálculos Primer Nivel C.C.A.	133
Tabla 40. Resumen de Cálculos Segundo Nivel C.C.A.	134
Tabla 41. Resumen de Cálculos Tercer Nivel C.C.A.	134
Tabla 42. Empresas de Estudio para M3S.	139
Tabla 43. Empresas Encuestadas	140

AGRADECIMIENTOS

El haber llegado a esta instancia académica en mi vida se basa en primer lugar al amor de mi familia; mi esposa Anabell el pilar más fuerte de mi hogar, quien me apoya siempre en cada una de las actividades profesionales y este doctorado más aún es el resultado de mucho esfuerzo personal y un gran sacrificio familiar. A mi hija Anita Belén y mis hijos Juan Esteban y Mateo Santiago, a quienes les quité mucho tiempo de vivencia en familia, sin embargo, me siento convencido que dejé una semilla que ya ha iniciado a dar frutos en su vida.

A los señores maestros de la Universidad de Alicante, y de manera especial a Ramón Rizo Aldeguer, quien me brindó primero su amistad, su confianza y sobre todo su experiencia, con la que pude llevar adelante esta investigación.

Estoy seguro que una persona de más de 50 años, regresa a ver a su lado y no cuenta con verdaderos amigos y compañeros, no trascendió en su vida. En este trabajo queda sentado mi agradecimiento a todas las personas con las cuales tuve la oportunidad de transitar en la mejor institución de educación superior del país como es la Universidad Central del Ecuador y en las empresas privadas a las cuales también he tenido el honor de servir.

RESUMEN

Con la idea de orientar a quienes están al frente de diferentes tipos de organizaciones para la mejor y oportuna toma de decisiones, hemos analizado y consideramos que uno de los retrasos en proyectos de Inteligencia de Negocios (IN) conocido en el idioma inglés como Business Intelligence (BI) se produce en los procesos de Extracción, Transformación y Carga de Datos (ETC), para lo cual se propone un algoritmo de decisión basado en ID3, pero incluyendo mejoras en la búsqueda de sus nodos por medio de atributos ubicados en diferentes fuentes de datos estructurados o no, los mismos que pasan por procesos de limpieza especial y a los cuales se les da en este estudio una denominación de "puros". Así obtenidos los nodos estos pasan a ser clasificados en base a los cálculos recursivos de entropía y ganancia de información para lograr una estructura robusta que tome decisiones para cualquier dato que ingrese en un determinado caso de estudio.

M3S, fundamenta, representa y articula en la práctica la implementación de proyectos de IN usando modelos matemáticos y explotando datos disponibles para generar conocimiento tomando en cuenta el tiempo del proyecto, madurez de la información, introducción de técnicas de Inteligencia Artificial como los árboles de decisión y aprendizaje automático; todos estos en conjunto ayudarán a mejorar los procesos de extracción, transformación y carga de datos, de tal forma que aseguremos mejores decisiones en cualquier tipo de actividad empresarial.

M3S, al recopilar datos de diferentes fuentes heterogéneas basados en la experiencia real de los expertos, consideramos que es más robusta y puede tener mayor certeza luego de su procesamiento de pureza de información, a diferencia de datos planos que se toman de equipos como lo que sucede en diferentes disciplinas y que generan variedad de escalas de tiempo, las mismas que determinan los parámetros de integración o identifican los acoplamientos entre estas series, como bien se lo menciona en la Entropía de Permutación.

Es por eso que es pertinente la propuesta de M3S, al analizar desde otra perspectiva distinta a los datos, de cómo se los ve en la entropía de permutación que si bien es cierto proporciona un método simple para estimar la complejidad de las series de tiempo; M3S considera una variedad de tipos de datos que se conjugan para un estudio específico.

Palabras Clave:

Entropía, Clasificación, Limpieza de Datos, BI, Árboles de decisión.

ABSTRACT

With the idea to guide those who are in charge of different types of organizations for the sake of making better decisions. We have analyzed and considered that one of the delays in Business Intelligence BI projects occurs in the processes of Extraction, Transformation and Data Load (ETL), an algorithm based on ID3 is proposed, but including improvements in the research of its nodes by located attributes in different sources of structured data or not, they go through special cleaning processes and they are given in this study a denomination of "pure". Once obtained the nodes they are classified based on the recursive calculations of entropy and information gain to achieve a robust structure that makes decisions for any data that enters in a case study.

M3S, gathers data from different sources based on the real experience of the experts, we consider that it is more robust and it can have greater certainty after processing the purity of the information. Unlike plain data that is taken from equipment like what happens in different disciplines and that generate a variety of time scales. They determine the parameters of integration or identify the links between these series, as it is mentioned in the Permutation Entropy.

That is the reason why the proposal of M3S is relevant, when analyzing from another perspective the data as what is seen in the permutation entropy. Although, it is true it provides a simple method to estimate the complexity of time series; M3S considers a variety of data types that are combined for a specific study.

Keywords:

Entropy, Classification, Cleaning Data, BI, Decision Tree.

RESUM

Amb la idea d'orientar als que estan al capdavant de diferents tipus d'organitzacions per a la millor i oportuna presa de decisions, hem analitzat i considerem que un dels retards en projectes d'Intel·ligència de Negocis (IN) conegut en l'idioma engonals com Business Intelligence (BI) es produïx en els processos d'Extracció, Transformació i Càrrega de Dades (ETC), per a la qual cosa es proposa un algoritme de decisió basat en ID3, però incloent millores en la cerca dels seus nodes per mitjà d'atributs ubicats en diferents fonts de dades estructurats o no, els mateixos que passen per processos de neteja especial i als quals se'ls dóna en este estudi una denominació de "puros". Així obtinguts els nodes estos passen a ser classificats basant-se en els càlculs recursius d'entropia i guany d'informació per a aconseguir una estructura robusta que prenga decisions per a qualsevol dada que ingresse en un determinat cas d'estudi.

M3S, fonamenta, representa i articula en la pràctica la implementació de projectes d'IN usant models matemàtics i explotant dades disponibles per a generar coneixement tenint en compte el temps del projecte, maduresa de la informació, introducció de tècniques d'Intel·ligència Artificial com els arbres de decisió i aprenentatge automàtic; tots estos en conjunt ajudaran a millorar els processos d'extracció, transformació i càrrega de dades, de tal forma que assegurem millors decisions en qualsevol tipus d'activitat empresarial.

M3S, al recopilar dades de diferents fonts heterogènies basats en l'experiència real dels experts, considerem que és més robusta i pot tindre major certesa després del seu processament de pureza d'informació, a diferència de dades planes que es prenen d'equips com el que succeix en diferents disciplines i que generen varietat d'escales de temps, les mateixes que determinen els paràmetres d'integració o identifiquen els adaptaments entre estos sèries, com bé se'l menciona en l'Entropia de Permutació.

És per això que és pertinent la proposta de M3S, a l'analisar des d'una altra perspectiva diferent de les dades, de com se'ls veu en l'entropia de permutació que si bé és cert proporciona un mètode simple per a estimar la complexitat de les sèries de temps; M3S considera una varietat de tipus de dades que es conjuguen per a un estudi específic.

Paraules Clau:

Entropia, Classificació, Neteja de Dades, BI, Arbres de decisió.

PRÓLOGO

La experiencia en desarrollo de sistemas transaccionales durante más de veinte años en pequeñas, medianas y grandes empresas, dentro de diferentes líneas como el administrativo, financiero, comercial, producción y sobre todo en el sector agropecuario, me facilitó la observación, el estudio de sus características y necesidades de información en el pasar del tiempo; por lo que se ha generado paralelamente a la docencia el interés de investigar y profundizar mucho más desde varios aspectos en la Inteligencia de Negocios (IN).

Las Tecnologías de la Información y Comunicación (TIC), me han permitido tomar variada información y convertirla en mayor y mejor conocimiento en diferentes tipos de negocios, así que actualmente considero como una obligación personal desde la academia, aportar científicamente en el área de mi especialidad.

Al realizar proyectos de IN, pude visualizar que en un 90% de ellos los procesos de extracción y transformación de datos, se tomaron demasiado tiempo de los expertos informáticos; por lo que considero que en esta parte podría mejorarse mucho este tipo de procesos, de tal forma que ese tiempo optimizado se transforme en mejor y mayor rentabilidad para las empresas; esto hará que el futuro en IN sea mayormente difundido a nivel de consultorías y nuevos emprendimientos de servicios tecnológicos, para que de esta forma sea más explotada la información por todos los niveles de las empresas.

Una vez finalizado este trabajo, me considero con las competencias necesarias para de igual forma dirigir nuevas investigaciones dentro del ámbito en el cual me desenvuelvo y de esta forma colaborar y mejorar el nivel de investigación que debe aportar la universidad para el desarrollo del país

Esa es una más de mis tareas y así nació esta investigación

1. CAPÍTULO I: INTRODUCCIÓN.

1.1. JUSTIFICACIÓN DE LA INVESTIGACIÓN.

Según un estudio publicado por (Wixom & Watson, 2014), indica que cada vez más las nuevas tecnologías se están usando para dirigir y habilitar Big Data, y que ellas son significativamente diferentes de las tecnologías de datos heredadas (antiguas o tradicionales). Señala a la analítica In-Memory como una opción cada vez más viable, por lo que crecientemente las empresas la están usando para aplicaciones analíticas específicas.

La Inteligencia de Negocios, está creando un cambio de paradigma con el movimiento de datos. Las organizaciones se están moviendo para traer los datos hacia el procesamiento (entiéndase como servidores) empujando éste hacia la data. “El análisis de datos In-Memory, con las arquitecturas distribuidas y el procesamiento paralelo masivo (MPP) está ganando cada vez mayor ‘momentum’ ” (Wixom & Watson, 2014).

De esta forma buscando identificar los principales elementos que caracterizan los proyectos de (IN), se encuentra que el contexto actual está marcado por fuertes, rápidos

cambios tecnológicos y propuestas de software para sus mejores implementaciones, “al mismo tiempo existe una amplia como diversa literatura sobre el tema que manifiesta que IN, despierta actualmente el máximo interés en las medianas y grandes empresas, generalmente reconociendo su valor estratégico para la toma de decisiones” (Fuentes & Ricardo, 2010). De igual manera, “se reconoce el papel que juegan las medianas y grandes empresas para el movimiento de la economía y el desarrollo social en las actividades productivas en todas partes del mundo, más aún se encuentra ahí el impacto que producen las decisiones en base a procesos bien establecidos de IN” (Calzada & Abreu, 2009). Definitivamente en la actualidad se resalta el rol protagonista que tienen las nuevas tecnologías de software en el desarrollo de proyectos, percibiéndose a la IN como un fenómeno que ha recortado brechas entre las grandes y medianas empresas.

Reconocemos algunas diferencias de implementación en el sector público como en el privado; al igual que entre las pequeñas y medianas empresas (PYMES), ya sea por su misma forma de operación, por su dinámica financiera diferente en la cual se desarrollan. Debido a esto de igual forma se reconocen factores limitadores para el desarrollo de proyectos de IN en todo tipo de empresas dentro de los principales destacamos la dispersión del cuerpo teórico actualmente existente lo que limita la comprensión de los conceptos, características, patrones de comportamiento y principios, que repercuten en el inicio de proyectos; la falta de propuestas claras que modelen mejor en ciertos procesos como la extracción y transformación de datos; la necesidad de herramientas asequibles que ayuden a las PYMES a gestionar sus procesos de IN, lo que generaría nuevos tipos de consultorías.

En la práctica empresarial, se puede afirmar que hay cuatro aspectos fundamentales en el marco de esta investigación que deben ser atendidos para un correcto desarrollo de los nuevos proyectos de IN. Disponer de un mejor proceso para la extracción de información que permita optimizar los tiempos en la consecución de estos proyectos; luego la creación

de formatos que describan procesos claros y ágiles para transformar la información a una data específica que se procese en un software de IN; colocar un método concreto que facilite la explotación del modelo y ayude a la empresa en la gestión de conocimientos; y por último fundamentar la metodología propuesta M3S para el desarrollo de proyectos de IN, en la cual incluya el uso de árboles de decisión dentro del proceso de extracción y transformación.

Buscamos sustentar una justificación de una investigación, recalando que “el iniciar proyectos de este tipo planteará siempre la problemática de establecer elementos que desarrollen la capacidad de fortalecer el conocimiento que las empresas adquieran a través de acciones centradas en los sistemas de información, la innovación y el proceso de la toma de decisiones” (Ahumada T & Perusca V, 2016), por esta razón en el caso puntual de esta investigación, la justificamos, en su intento de contribuir y avanzar en la resolución de algunos problemas enunciados, por su conveniencia para las empresas, las economías y las sociedades en su conjunto.

1.2. PROBLEMA DE INVESTIGACIÓN

A pesar de la gran importancia que tiene la IN con la toma de decisiones consideramos que aún se puede trabajar más en metodologías o más aún en ciertos puntos de este tipo de proyectos y de esta forma atendiendo a elementos que configuran la situación sobre el tema propuesto, se destaca como la línea de investigación la siguiente:

La necesidad de construir una metodología que fundamente, represente y articule en la práctica la implementación de proyectos de IN; tomando en cuenta el tiempo del proyecto, madurez de la información, introducción de técnicas de Inteligencia artificial como los árboles de decisión, aprendizaje automático; todos estos en conjunto ayudarán a mejorar

los procesos de extracción y transformación de datos, de tal forma que aseguremos mejores decisiones en cualquier tipo de negocio.

1.3. MOTIVACIÓN, OBJETIVOS GENERAL Y ESPECÍFICOS.

Considerando la importancia que ha tomado en la actualidad la IN y las diversas tecnologías de información, las mismas que ayudan a lograr una toma de decisiones efectiva en todos los ámbitos empresariales, considero de gran valor fortalecer procesos dentro de metodologías que engloben a estos proyectos; de esta forma conseguiré que cada vez se cuente con una mayor formalidad y estructuración para la ejecución y seguimiento apropiado de los mismos.

Esta investigación propone como **objetivo general** el siguiente:

El objetivo de este trabajo doctoral es conseguir que se disminuyan los tiempos de desarrollo de los proyectos de IN, por medio de una metodología orientada a la toma de decisiones, mejorando principalmente los procesos de extracción y transformación de fuentes de datos heterogéneos con la ayuda de árboles de decisión que aplican aprendizaje automático en las decisiones.

Este objetivo general se soporta en los siguientes **objetivos específicos o instrumentales:**

- Revisar procesos y metodologías que se hayan ejecutado en diferentes organizaciones en las cuales se usaron este tipo de tecnologías.
- Desarrollar una metodología que permita establecer de manera óptima tanto en tiempo, como en recursos soluciones IN con un alto grado de confiabilidad en el conocimiento obtenido.
- Contribuir con una nueva metodología formal en IN, para la generación de conocimiento en las organizaciones, particularmente en lo concerniente a manejo de indicadores relevantes de un determinado negocio o actividad de la vida.

- Usar técnicas de aprendizaje automático para mejorar los procesos de decisión.
- Usar todo tipo de herramientas informáticas y actividades como: foros, congresos, correo electrónico, redes sociales, aulas virtuales, entre otras; para apoyar la construcción de la nueva metodología a proponer.
- Validar la metodología propuesta, aplicándola al menos a dos organizaciones que contengan conocimiento y procesos clave muy bien definidos.
- Analizar diferentes herramientas de software en el mercado, y determinar aquellas que pueden apoyar de mejor manera a la metodología propuesta.

1.4. METODOLOGÍA DE LA INVESTIGACIÓN E HIPÓTESIS.

Dentro de la investigación la metodología que se usó fue de tipo mixta: En forma cualitativa usando instrumentos como encuestas, así como entrevistas personales para medir variables como la satisfacción, madurez de la información, calidad del entorno; y en forma cuantitativa estableciendo variables para determinar el nivel de mejoramiento en tiempos de extracción y transformación de datos, valorados en formularios que se crearon para este efecto. A través del estudio de casos se obtuvieron resultados que ayudaron a lograr el objetivo de mejorar el rendimiento de procesos de extracción y transformación, con el uso de aprendizaje automático a través de árboles de decisión.

De esta forma se establece que la Hipótesis dentro del presente trabajo de investigación es la siguiente:

Desarrollar una metodología integradora y óptima que mejore las propuestas actuales en la extracción y transformación de datos, dentro de un proyecto de IN.



Figura 1 Proceso de Investigación.

El proceso de investigación como se mostró en la Figura 1, se encuentra dividido en cuatro fases, las mismas que se detallan a continuación:

1. Aspectos de avanzada, en donde se realiza un estudio y análisis de IN, con definiciones aportadas por diferentes autores y su evolución a través de la historia. Se identifican algunos procesos, metodologías y herramientas existentes para realizar IN, buscando la manera de detectar los aspectos no contemplados y los cuales pueden mejorarse.
2. Propuesta de Nueva Metodología IN: Una vez adquiridos los conocimientos en cuanto a los procesos extracción, transformación y carga de datos, así como el desarrollo de varios casos de estudio, se identificará en qué parte de estos se pueden realizar mejoras lo que se reflejará en la creación de una metodología con aprendizaje automático para potencializar a dicho proceso en proyectos de IN.
3. Prueba y Desarrollo de la Metodología: Desarrollaremos varios casos de estudio en los cuales, dentro de las etapas se llegue a implementar la metodología propuesta para establecer mejoras en el diseño y construcción, optimizando el tiempo de ejecución del proceso de extracción y transformación de datos dentro del proyecto.
4. Validación de Resultados: Se realizará un comparativo de los resultados obtenidos en cada caso de estudio y se aplicarán más datos para verificar las salidas en cada caso.

1.5. CICLO DE VIDA DE LOS PROYECTOS I.N.

Antes de proponer una solución es importante revisar “las fases de metodologías que plantean en forma general el ciclo de vida de un proyecto de IN, en las cuales se hablan de etapas claras, así como la relación entre datos, información y conocimiento” (Fernandez A, 2006); en esta sección presentamos dos de ellas. En la Figura 2, se muestra un Ciclo de Vida general utilizado no solo en proyectos IN, sino en general para proyectos informáticos.



Figura 2. Ciclo de vida de los proyectos I.N.

Fase 1: Dirigir y Planear: Es definir los requerimientos del manejo de la empresa, mediante un grupo de analistas, se formulan preguntas e hipótesis para alcanzar el objetivo y cubrir las necesidades.

Fase 2: Recolección de Información: Se debe identificar la información en base de datos, sin importar que provengan de un sistema independiente, con la finalidad de obtener los resultados esperados.

Martin James define una base de datos como “una colección de datos interrelacionados almacenados en conjunto sin redundancias perjudiciales o innecesarias; su finalidad es servir a una o más aplicaciones de la mejor forma posible; los datos se almacenan de modo que resulten independientes de los programas que los usan; se emplean métodos bien determinados para incluir nuevos datos y para modificar o extraer los datos almacenados” (Martin, 1989).

Según Carlos Coronel y Rob Peter dicen que: “Una base de datos se puede definir como un conjunto de información relacionada que se encuentra agrupada o estructurada. Desde el punto de vista de la informática, la base de datos es un sistema formado por un conjunto de datos almacenados en discos que permiten el acceso directo a ello y un conjunto de programas que manipulen ese conjunto de datos.” (Coronel Carlos, 2011).

Fase 3: Procesamiento de Datos: Corresponde a integrar los datos en un mismo formato para que puedan ser luego analizados, interpretados y procesados.

Fase 4: Análisis y Producción: Se crean reportes personalizados, que trazarán los cursos de acción a seguir, y proporciona datos o indicadores relevantes para la toma de decisiones.

Fase 5: Difusión: “Se despliegan los reportes esperados, para su interpretación de tal forma que se generen planes de mejora y acción” (Medina Chicaiza, 2004).

Desde otro análisis en la Tabla 1, se determina otra consideración de ciclo de vida usado en proyectos de Inteligencia de Negocios.

Tabla 1: Ciclo de Vida de Proyectos IN. (Vanrell, Bertone, & Ramón, 2010)

SUBPROCESO	TAREA	SALIDA
ENTENDIMIENTO DEL NEGOCIO	<ul style="list-style-type: none"> Determinar las metas del Data 	<ul style="list-style-type: none"> Criterios de Éxito
ENTENDIMIENTO DE LOS DATOS	<ul style="list-style-type: none"> Reunir datos iniciales Describir datos Explorar los datos 	<ul style="list-style-type: none"> Reporte de Datos iniciales, descripción y exploración Reporte de Calidad
PREPARACIÓN DE LOS DATOS	<ul style="list-style-type: none"> Tareas preparatorias Seleccionar datos Limpiar datos Construir los datos 	<ul style="list-style-type: none"> Datasets descritos Justificación de inclusión y exclusión Reporte de limpieza
MODELADO	<ul style="list-style-type: none"> Seleccionar técnica de Modelado 	<ul style="list-style-type: none"> Técnica de modelado Suposiciones de Modelado Diseño de test Establecer parámetros Modelo y descripción
EVALUACIÓN	<ul style="list-style-type: none"> Evaluar Resultados Revisar proceso Determinar próximos pasos 	<ul style="list-style-type: none"> Evaluación resultados hacia los criterios de éxito. Modelos Aprobados Revisión del proceso Lista de decisiones
ENTREGA	<ul style="list-style-type: none"> Producir un reporte final 	<ul style="list-style-type: none"> Reporte final Presentación final

Así entonces, partiendo de algo general se quiere mejorar una metodología que proveerá las herramientas y las técnicas necesarias para optimizar principalmente el proceso de extracción y transformación de información de un proyecto IN, incluyendo los aportes que nos pueden hacer los árboles de decisión en un proceso de este tipo. Para esto es necesario también dentro de una organización considerar puntualizaciones técnicas que se han fortalecido con la experiencia y las que deberán ser incluidas como apoyo dentro de una metodología formal:

1. Reconocimiento del entorno y objetivos de la organización
 - Conocer los objetivos estratégicos del negocio
 - Analizar el entorno y los retos a los que se ve enfrentado
 - Conocer la visión de los ejecutivos principales.
2. Insumos para el análisis
 - Sitio web de la institución
 - Entrevistas con gerentes de área para revisión de objetivos corporativos.
 - Análisis de sus principales indicadores y metas
3. Análisis por área

- En cada una de las divisiones o gerencias se llevan a cabo entrevistas con los principales ejecutivos y analistas de negocio
- Se revisan las hojas de trabajo (ppt, xls, otros) en relación con la información que presentan a sus gerencias superiores
- Se analizan los métodos de extracción, cálculo y presentación de la información.

4. Calificación de madurez

- En base a un conjunto de parámetros se determinarán los niveles de datos o la madurez en el proceso de análisis de información de la organización.

5. Conclusiones y ruta a seguir

- Se sugiere un conjunto de pasos a seguir para llegar a una meta deseada, bajo un esquema metodológico de IN.

Así entonces como se muestra en la Figura 3, considero que se puede plantear la Metodología M3S basada en las siguientes etapas generales, las mismas que se explicarán a detalle en el Capítulo IV.

Universitat d'Alacant
Universidad de Alicante

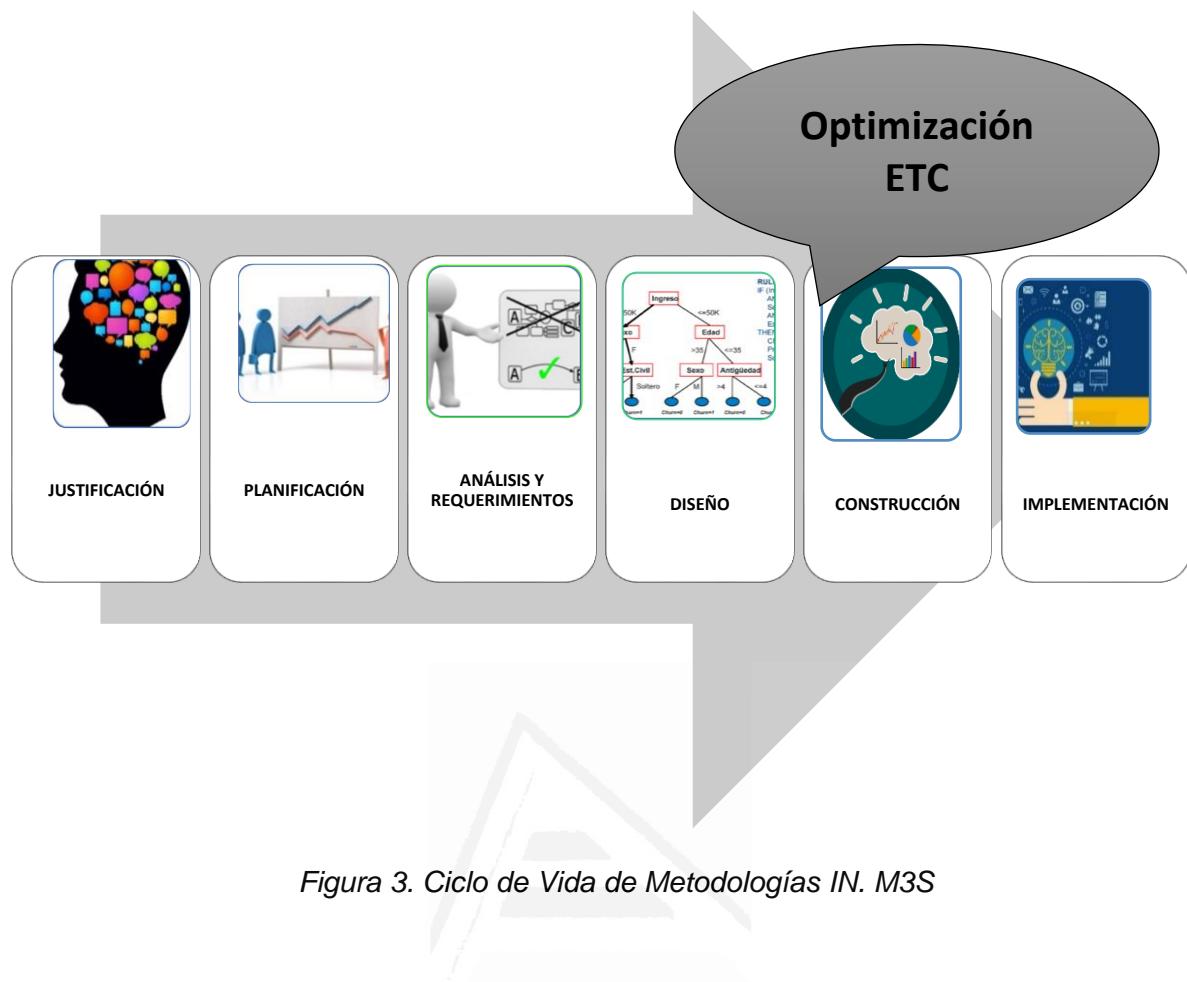


Figura 3. Ciclo de Vida de Metodologías IN. M3S

1.6. BENEFICIOS ESPERADOS

Dentro de lo planteado en esta investigación pretendo aportar en lo que se refiere a la estructuración e implementación de una metodología que proponga la optimización de los procesos de extracción, transformación y carga (ETC) en cualquier tipo de negocio, inmersa siempre en un ciclo normal dentro de la IN; de otra forma se quiere mejorar los tiempos de ejecución en cuanto a la extracción y transformación de información con una propuesta novedosa y eficiente, lo que finalmente aumentará la rentabilidad en los proyectos.

Sería importante unificar los distintos niveles de extracción y transformación de datos, con un algoritmo de toma de decisiones que permita manejarlos con mayor eficiencia y de manera más organizada en los proyectos de IN.

1.7. ESTRUCTURA DEL DOCUMENTO DE TESIS.

En los primeros capítulos se revisa la literatura sobre los conceptos asociados a la IN y las distintas metodologías con los que se ha intentado resolver un problema de negocio a lo largo de las últimas décadas, luego se plantea una nueva alternativa desarrollando varios casos reales y posteriormente se analizan los resultados.

En el capítulo I, correspondiente a la *Introducción* se expone el problema científico que justifica la procedencia de la presente investigación; se establecen sus objetivos y se describe de forma sintética la misma.

Dentro del capítulo II, se sustenta el *Estado del Arte de IN*, en donde se ofrece un análisis de las principales acepciones asociadas a IN, aportadas por diferentes autores y su evolución en los últimos tiempos, adicionalmente se revisa la IN vista desde los distintos niveles que existen en los sistemas de información de una organización; se describen las soluciones tradicionales y las metodologías tanto genéricas como específicas que han sido desarrolladas para realizar un proyecto aceptable en esta tendencia.

A partir del capítulo III, se inicia todo lo que se refiere a las aportaciones que se pueden considerar para construir una nueva metodología de IN, iniciamos con un comparativo de ellas, así como de herramientas actuales que se encuentran en el mercado; se revisan teorías de control y seguimiento de procesos, además de la teoría de la decisión. Por último se revisa la situación actual de IN, los problemas que existen en cuanto a éste ámbito y se plantea donde se puede avanzar dentro del gran campo de oportunidad que nos ofrece este tipo de tecnologías.

En el capítulo IV, se propone la estructuración de la metodología M3S para proyectos de IN, además de la especificación de cada uno de sus procesos elaborados para cumplir

el objetivo inicial de dicha metodología. Los énfasis especiales tienen las etapas de diseño y construcción de ETC, con la implementación de árboles de decisión para el aprendizaje automático de datos.

Para el capítulo V, se preparan dos casos para pruebas del estudio en empresas de proyectos de IN en los cuales se probará M3S, y dentro de la Tabla 2, se muestra las instituciones y el proyecto presentado en cada una de ellas.

Tabla 2: Casos de Proyectos IN (Propia, 2016)

INSTITUCIÓN	PROYECTO
Empresa Eléctrica Quito	Indicadores ubicación en cambios de boquillas LED. (Nueva Matriz productiva Ecuador)
Club Castillo de Amaguaña	Tablero para el departamento de Cobranzas

En el capítulo VI, de *Conclusiones y líneas futuras*, se describirá el principal aporte de la investigación, el modelo M3S es una realidad para la realización de proyectos de IN, centrándose en la optimización de procesos ETC. Este estudio nos lleva a un escenario donde se pretende impulsar el uso del análisis de información (con herramientas de IN modernas y con implementaciones metodológicamente adaptadas) en empresas de segmentos medianos y pequeños. A su vez, esto se medirá en el futuro a través de estudios de niveles de madurez en el uso de herramientas IN para este tipo de empresas, lo que deja pendiente un espectro importante para nuevas investigaciones que se pueden resolver en este campo.

2. CAPÍTULO II: ESTADO DEL ARTE.

Actualmente, casi la totalidad de las organizaciones cuenta con un sistema de información que soporta sus actividades diarias propias del sector de sus negocios, este sistema puede ser sencillo o robusto todo depende de las exigencias del mismo y los niveles de información gubernamental que se deba entregar; “con el tiempo las aplicaciones llegan a tener la historia de la organización y los datos almacenados en las bases de datos, pueden ser utilizados para argumentar la decisión que se quiera tomar ante cualquier aspecto para mejora en la empresa”. (Rosado Gomez, 2010).

El ser humano toma decisiones en cada uno de los aspectos que giran alrededor de su vida con experiencias y hechos pasados; como en actividades cotidianas como la de un ganadero que debe saber cuándo chequear a sus animales, o un floricultor que debe tener conocimiento de cuál es el mejor momento para pinchar los tallos de sus flores, o en una empresa pública o privada, contar con los indicadores que le permitan tomar decisiones en un momento crítico de la empresa en búsqueda de mejores oportunidades para su línea de negocio; en estos ejemplos se ve que los proyectos de explotación de información poseen características muy distintas a las de los proyectos de desarrollo de software

tradicionales. “Las clásicas etapas de análisis, diseño, desarrollo, integración y testeo, no encajan ni son suficientes para brindar conocimiento, como lo son las etapas naturales de los procesos de desarrollo de este tipo de proyectos de mayor peso con la información” (García M, y otros, 2011).

“La inteligencia de negocios, es un enfoque estratégico para orientar sistemáticamente el seguimiento, la comunicación y la transformación relacionada al débil conocimiento de la información procesable en la cual se basa la toma de decisiones”. (KAMEL. Rouibah y SAMIA, 2002). En la época pasada los Data Warehouse (DW) han sido ampliamente adoptados en los negocios, proveían análisis multidimensional con muchos datos históricos del negocio, ayudando a crear nuevas decisiones. “Sin embargo en el 20% de las empresas se encuentran los datos numéricos y el otro 80% están escondidos en NO numéricos y hay que recuperarlos para realizar un análisis más completo de BI”. (Tseng & H, 2006)

“Las recomendaciones para implementar un proyecto de Inteligencia de Negocios basado en un DW, debe analizarse desde diferentes aspectos: Económico (Costos cuantificables e intangibles), Social (Ambiente organizacional, Apoyo, Aspectos Administrativos, Compromiso), y Técnico (Infraestructura, Metodologías)”. (Villalon M, 2006). Cada organización maneja la Inteligencia de Negocios de acuerdo a las exigencias de la misma, es aquí que se ve la necesidad de crear una metodología en cuanto al proceso ETC para empezar a definir estándares, ya que los softwares que existen actualmente en el mercado no necesariamente nos ofrecen las mejores alternativas. “En el año 2003 ya se decía que la demanda de Sistemas de Inteligencia de Negocios estaba creciendo, pero la producción de software en general caminaba más lento”. (Negash, 2004). De esta manera otro autor considera que “la investigación y la gestión del conocimiento es una idea básica, por lo que el conocimiento que se genere esté compartido a la comunidad”. (Krogh & von, 2002).

El crecimiento en los servicios de consultoría de Reingeniería de Procesos de Negocio (BPR) ha llevado a una proliferación de métodos para conducir BPR. Se han hecho investigaciones revisando Métodos, Técnicas y Herramientas (MTTs), y en base a encuestas exhaustivas de estas técnicas de uso común de BPR, se forma una base de conocimientos para mejorar la práctica de cambio de procesos empresariales y proporciona una base para la investigación BPR futuro. (Kettinger, Teng, & Subashish, 1997)

En este capítulo de la investigación se inicia con un análisis de las principales acepciones asociadas a BI, aportadas por diferentes autores en todos los tiempos; se analizan diferentes tipos de metodologías aplicables a proyectos de BI, para el seguimiento de calidad nos enmarcamos a círculos de procesos para finalmente, y apoyado en diversas experiencias, se efectúa un análisis comparativo de los modelos más representativos que se han venido proponiendo en los últimos años, desvelando sus puntos fuertes, débiles y aspectos no abordados a mejorar.

2.1. CONCEPTOS Y TÉRMINOS.

Dentro de muchas acepciones realizadas sobre la inteligencia de negocios, vamos a presentar algunas de ellas propuestas por diferentes autores los que dicen que: “Se entiende por IN al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización”. (Díaz J. C., 2010), (Vilchez, 2011); cada una de ellas se enfocan en la toma de decisiones como también las que mencionan que: “El objetivo primario de la Inteligencia de Negocios es contribuir a tomar decisiones que mejoren el desempeño de la empresa y promover su ventaja competitiva en el mercado. Este concepto se requiere analizar desde tres perspectivas: Tomar mejores decisiones

más rápido, convertir datos en información, y usar una aplicación relacional para la administración". (Calzada & Abreu, 2009) . "La Inteligencia de Negocios se puede definir como el cúmulo de modelos matemáticos y metodologías de análisis que explotan los datos disponibles para generar información y conocimientos útiles para los complejos procesos de toma de decisiones". (Garcia Reyes, 2012). Así también se definen dos aseveraciones más que indican que: "La Inteligencia de Negocios BI (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; garantizando la generación del conocimiento necesario que permita escoger la alternativa que sea más conveniente para el éxito de la empresa". (Rosado & Alveiro, 2010). Debido a la importancia de esta definición en nuestro estudio consideramos algunos autores más que indican: "Inteligencia de Negocio es un concepto que "integra como solución el almacenamiento y procesamiento de enormes cantidades de datos e información para transformarla en conocimiento y decisiones en tiempo real a través de una fácil explotación". (Salinas La Rosa, 2010); el mismo investigador indica que: "BI es un término que engloba las aplicaciones, infraestructura, plataformas, herramientas y mejores prácticas que permiten acceder y analizar la información para optimizar las decisiones y gestionar el rendimiento empresarial" (Salinas La Rosa, 2010). "Se entiende por Inteligencia de Negocios al conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización".(Díaz J. C., 2012). Desde otro punto de vista consideramos una definición en idioma inglés que indica: "La inteligencia empresarial es una categoría de aplicaciones y tecnologías para recopilar, proporcionar acceso y analizar datos con el fin de ayudar a los usuarios empresariales a tomar mejores decisiones" (Ranjan, 2009). Por último revisamos que la inteligencia de negocios se define como "la habilidad corporativa para tomar decisiones; esto se logra mediante el uso de metodologías, aplicaciones y tecnologías que permiten reunir, depurar, transformar datos, y aplicar en ellos técnicas

analíticas de extracción de conocimiento” (Parr & Sons, 2009), “los datos pueden ser estructurados para que indiquen las características del área de interés” (Stackowiak, Greenwald, & Rayman, 2007), “generando el conocimiento sobre los problemas y oportunidades del negocio para que pueden ser corregidos y aprovechados respectivamente” (Ballard, 2006) y (Grossman, 1998), mencionan que IN, “es un término general para un conjunto de conceptos y métodos para mejorar la toma de decisiones empresariales, utilizando sistemas basados en hechos”.

En vista de la relación existente con la *Minería de Datos*, tomamos algunas definiciones sobre esta que indican lo siguiente: “La minería de datos se refiere a la extracción de conocimiento a partir de grandes cantidades de datos, sobre los que se aplican métodos inteligentes con el fin de extraer patrones interesantes. Dichos datos pueden estar almacenados en bases de datos, almacenes de datos u otros repositorios de información”; (Garcia Reyes, 2012); otros autores indican que: “La etapa de minería de datos puede definirse sobre la base de un conjunto de primitivas diseñadas especialmente para facilitar un descubrimiento de conocimientos eficiente y fructífero. Tales primitivas incluyen: la especificación de las porciones de la base de datos o del conjunto de datos en los que se quiere trabajar; la clase de conocimiento a ser descubierto; los conocimientos existentes que podrían resultar útiles para guiar el proceso de Knowledge Discovery in Databases (KDD); las métricas de interés para llevar a cabo la evaluación de patrones en los datos analizados; y finalmente, las formas en que el conocimiento descubierto podría ser visualizado” (Duran & Costaguta, 2007). “La minería de datos es una subdisciplina de las ciencias de la computación que ha logrado mucho reconocimiento en los últimos años, principalmente porque puede ser usada para diferentes propósitos como herramienta de apoyo en las demás disciplinas de las ciencias”. Su fortaleza radica en el hecho de que forma parte del proceso de descubrimiento del conocimiento, cuyo objetivo es la búsqueda de patrones de datos que sean válidos, novedosos, potencialmente útiles y comprensibles” (Rodallegas Ramos Erika, 2010). Adicionalmente encontramos que: “La Minería de Datos

o Data Mining, se ve como el soporte a las decisiones en las actividades de negocio, requiere mucho más que la aplicación de sofisticadas técnicas de redes neuronales o árboles de decisión sobre las tablas de datos". "Proceso que consta de diferentes fases, en las cuales se utilizan como apoyo, técnicas relacionadas con la estadística, el reconocimiento de patrones y algoritmos de aprendizaje, entre otras" (Dueñas-Reyes, 2009), y por último tomado textualmente en inglés: "La minería de datos es el proceso de descubrir patrones y conocimientos interesantes a partir de grandes cantidades de datos. Las fuentes de datos pueden incluir bases de datos, almacenes de datos, la web, otros repositorios de información o datos que se incorporan dinámicamente en el sistema". (Han & Kamber, 2009). "La minería de datos puede definirse inicialmente como un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias al examinar grandes cantidades de datos. La disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas ha transformado el análisis de datos orientados hacia determinadas técnicas especializadas englobadas bajo el nombre de minería de datos o Data Mining" (Pérez Cesar, 2008).

Consideramos de igual forma algunas definiciones de *Big Data* como: "Es el conjunto de herramientas informáticas destinadas a la manipulación, gestión y análisis de grandes volúmenes de datos de todo tipo los cuales no pueden ser gestionados por las herramientas informáticas tradicionales. "Big data es un término de origen inglés cuya traducción equivale a "Datos masivos", la tecnología Big Data tiene por objetivo analizar datos e información de manera inteligente que ayuden a una correcta toma de decisión" (Aaker, 2015); de igual forma podría definirse como: "La ciencia que se basa en el tratamiento de grandes volúmenes de información con técnicas matemáticas e informáticas y que permite recoger datos, procesarlos y visualizarlos obteniendo una gran velocidad en el análisis, pudiendo anticipar tendencias, con el objetivo de comprender y optimizar ciertos servicios en función del comportamiento del usuario, para satisfacer necesidades tanto en tiempo real, como elaborar estrategias de primer orden en un sector determinado" (Gallego

Calonge, 2012) y por último: consideramos cuatro citas: "Big data se refiere a nuestra nueva habilidad de hacer cálculos con una gran cantidad de información, analizar al instante, y sacar conclusiones a veces sorprendentes de ella" (Jara, 2012); "Big Data es el término utilizado para describir a grandes volúmenes de datos no estructurados o semi-estructurados que son creados por diversas compañías o empresas los cuales tendrán demasiado costo a nivel de procesamiento, dinero y tiempo como para ser tratados de la forma actual, cargándolo en bases de datos transaccionales para su análisis". (Jara, 2012)

. Y según McKinsey Global Institute BigData, es el conjunto de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar(Joyanes Aguilar, 2013). Por último (Grossman, 1998) menciona que se conoce como Big Data al "procesar grandes cantidades de datos, la velocidad de los datos, a la frecuencia con la que nuevos los datos ingresan, al proceso de integración y análisis, la variedad de datos, la diversidad de datos y la veracidad de los datos".

Uno de los términos que se ha trabajado en IN, es el de: *Almacenes de Datos ó Data Warehouse* en el idioma inglés y se dice lo siguiente: "Informed decisions faster. When better decisions making is the goal of the data warehouse, it will be succesfull" (Westerman, 2001); de igual forma: "Data Warehouse: integra toda la información interna y externa disponible del cliente que debe organizarse en función de las necesidades de la empresa para que pueda desarrollar las adecuadas estrategias de marketing" (González & M., 2001) ; otra de las definiciones consideradas: "Es un almacén de datos, es el principal repositorio de los datos disponibles para desarrollar arquitecturas de inteligencia de negocios y sistemas de apoyo a las decisiones. Existen tres categorías principales de datos que alimentan un DW: datos internos, externos y datos" (Garcia Reyes, 2012). En otra cita de (Cámara Nuñez, 2010), dice que: "Es una colección de información creada para soportar las aplicaciones de toma de decisiones"; y por último "The concept of data warehousing is really quite simple. Data from older system dedicated entirely to analyzing that data.

Normally, the data warehouse will store a substantial amount of historical data. Users of this system are able to continuously ask or query it to retrieve data for analysis." (Westerman, 2001).

Uno de los términos importantes es el de *Base de Datos Departamentales (BDD)* y se dice que: "Denominada como Data Management en el idioma inglés, es una base de datos especializada en el almacenamiento de la información de un área específica del negocio. Se caracteriza por disponer la estructura óptima de datos para analizar la información al detalle desde todas las perspectivas que afecten a los procesos de dicho departamento. Una Base de Datos puede ser alimentada desde un Data Warehouse, o integrar por sí mismo un compendio de distintas fuentes de información" (Santana, 2011); dentro de este término se definen también a las *BDD de Procesamiento Analítico en Línea(OLAP)* que según algunos autores se dice que: "Se basan en los populares cubos de procesamiento analítico en línea, que se construyen agregando, según los requisitos de cada área o departamento, las dimensiones y los indicadores necesarios de cada cubo relacional. El modo de creación, explotación y mantenimiento de los cubos de procesamiento analítico en línea es muy heterogéneo, en función de la herramienta final que se utilice" (García, Espinoza, Castañeda, Ostos, & Salazar, 2016); A su vez es definido como "Procesamiento analítico en línea es el proceso de mantener, analizar y realizar informes sobre datos, añadiendo que los datos en cuestión son percibidos y manejados como si estuvieran almacenados en un arreglo multidimensional." (Date, 2001);y de igual forma : "(On-line analytical processing): It refers to the way in which business users can slice and dice their way through data using sophisticated tools that allow for the navigation of dimensions such as time or hierarchies". Contemplamos también otra definición como: BDD de Transacciones en Línea(OLTP), que dice: "Pueden basarse en un simple extracto del almacén de datos, no obstante, lo común es introducir mejoras en su rendimiento (las agregaciones y los filtrados suelen ser las operaciones más usuales) aprovechando las características particulares de cada área de la empresa. Las estructuras más comunes en

este sentido son las tablas report, que vienen a ser tablas reducidas (que agregan las dimensiones oportunas), y las vistas materializadas, que se construyen con la misma estructura que las anteriores, pero con el objetivo de explotar la reescritura de consultas". (García, Espinoza, Castañeda, Ostos, & Salazar, 2016).

Consideramos la definición de *Multidimensionalidad*: "La información proviene de diversas fuentes tales como: hojas de cálculo, bases de datos, ya que la herramienta debe poder reunir información completa aunque se encuentre separada para resumirla y tenga un alcance de análisis profundo, con bases sólidas y datos actualizados". (Zarate Gallardo, 2013)

Otro término que ha influido de manera importante en el manejo de información es el de *Modelo Entidad Relación (E-R)*: "Constituye una forma de representar conceptualmente la realidad basada en la representación de esta mediante su abstracción en entidades y relaciones. De esta forma intentamos representar el mundo que nos rodea, los datos de nuestro problema mediante una serie de entidades que representan objetos o conceptos, así como las relaciones que se dan entre ellos tales como su uso, composición, y otros" (Ullman, Widom, & Miguel, 1999), y también se tomó en cuenta la siguiente: "El modelo de datos entidad-relación (E-R) está basado en una percepción del mundo real que consta de una colección de objetos básicos llamados Entidades, y de relaciones entre estos objetos. Una entidad es una «cosa» u objeto en el mundo real que es distingible de otros. Por ejemplo, cada persona es una entidad, y las cuentas bancarias pueden ser consideradas entidades. Las entidades se describen en una base de datos mediante un conjunto de atributos" (Silberschatz, y otros, 2002).

Se define como *Datos a*: "La representación, codificación estructurada de entidades primarias individuales, así como de las transacciones que implican a dos o más entidades primarias" (Vercellis, 2009); también se conoce que: "La palabra Datos proviene del latín "Dtum" cuyo significado es "lo que se da". "Los datos son la representación simbólica, bien

sea mediante números o letras de una recopilación de información la cual puede ser cualitativa o cuantitativa, que facilitan la deducción de una investigación o un hecho". (Cobo Y, 2007).

Al mismo tiempo manejamos el término Información y se lo define desde dos puntos de vista así: "La información es el resultado de las actividades de extracción y procesamiento llevadas a cabo en los datos, y parece significativo para los que la reciben en un dominio específico" (Vercellis, 2009); y "Cada vez que necesitamos o deseamos conocer algo acerca de acontecimientos, hechos, sucesos, personas, empresas, instituciones ciudades, países, etcétera (ya sea porque necesitamos reducir nuestra incertidumbre para tomar decisiones o simplemente porque queremos incrementar nuestro conocimiento acerca de algo), buscamos aquello que denominamos «información». Y esto, es algo que hacemos, en mayor o menor medida, todos los días". (Thompson, 2008)

Dentro del análisis y diseño de base de datos un proceso importante es la *Normalización* de la cual se dice que : "Es el proceso de refinamiento de una base de datos, lo que incluye la creación de tablas y que establece relaciones entre ellas según reglas diseñadas para proteger los datos y hacer la base de información más flexible al eliminar dos factores; redundancia y dependencia incoherente " (Moreno S. , 2012);A su vez también es definido como "Es un proceso para evaluar y corregir estructuras de tablas a fin de minimizar redundancia de datos, con lo cual se reduce la probabilidad de anomalías de datos." (Coronel Carlos, 2011); así como: "Las formas normales se corresponde a una teoría de normalización iniciada por el propio Codd y continuada por otros autores (entre los que destacan Boyce y Fagin). Codd definió en 1970 la primera forma normal, desde ese momento aparecieron la segunda, tercera, la Boyce-Codd, la cuarta y la quinta forma normal. "Una tabla puede encontrarse en primera forma normal y no en segunda forma normal, pero no al contrario. Es decir los números altos de formas normales son más restrictivos (la quinta forma normal cumple todas las anteriores)" (Sánchez, 2004) .

Al hablar de Data Warehouse, se hablan de términos como *Tabla de Hechos*, que se la conoce como: "Técnicamente, es la tabla central de un modelo en estrella. Denominamos "hechos" a los indicadores del negocio. Por ejemplo, son hechos: las ventas, los pedidos, los envíos, las reclamaciones, las compras, etc. Es decir, son todas aquellas medidas numéricas que incluiremos en nuestro sistema de IN" (Facil, 2009); de igual forma en esta parte hablamos de la *Tabla de Dimensiones* de la que se dice: "En el modelo de almacén de datos multidimensional, se representan cada una de las posibles clasificaciones que se pueden aplicar a filas o columnas. "Estas son tablas de doble entrada, a los datos almacenados de esta manera se les suele llamar cubos. Las tablas de dimensiones definen cómo están los datos organizados lógicamente y proveen el medio para analizar el contexto del negocio, representan los ejes del cubo, y los aspectos de interés, mediante los cuales el usuario está en capacidad de filtrar y manipular la información almacenada en la tabla de hechos". (Ecured, 2012)

En la propuesta metodológica que hacemos una de las acepciones importantes es el *Árbol de Decisión*, del que veremos algunas como: "Is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decisión" (Peng, Chen, & Zhou, 2012); a su vez se define como *Algoritmo de Aprendizaje en Arboles de Decisión*: "Is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree. Decision tree learning is one of the most widely used and practical methods for inductive inference" (Tom, 1997). También es manejado el concepto de *ID3 Básico* como: "The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node". (Peng, Chen, & Zhou, 2012). Desde otro análisis, un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que permite determinar la decisión final que se debe tomar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas.

Varios puntos de vista indican que un árbol de decisión es un conjunto de condiciones organizadas en una estructura jerárquica, de tal manera que permite determinar la decisión final que se debe tomar siguiendo las condiciones que se cumplen desde la raíz del árbol hasta alguna de sus hojas. “Los árboles de decisión se utilizan desde los años cuarenta y son especialmente apropiados para expresar procedimientos médicos, legales, comerciales, estratégicos, matemáticos, lógicos, entre otros, estos se caracterizan por la sencillez de su representación y de su forma de actuar, además de la fácil interpretación, dado que pueden ser expresados en forma de reglas de decisión” (Martínez & R., 2009). Y por último según (Grossman,1998), “un árbol es un gráfico dirigido que contiene un nodo raíz con un grado en 0 y un número de hojas o nodos terminales que tienen un grado de salida 0”. Específicamente, los árboles no contienen ni círculos ni bucles.

En última instancia queremos mencionar el concepto de la *extracción de datos no estructurados* conocido por su traducción en el idioma inglés como *web scraping*, que en resumen significa extraer datos no estructurados y luego pasarlo a una base estructurada. Se enfoca más en la transformación de información desde la web como el formato HTML (Hypertext Markup Language) en datos estructurados que pueden ser almacenados y analizados en una base de datos central. Esta tecnología conduce una cantidad sustancial de los negocios, y la viabilidad de muchas empresas se basa en él. Sin embargo, la controversia puede surgir cuando las empresas comerciales utilizan web scraping para recoger grandes cantidades de datos de sitios web para su propio beneficio. “Alguno de los usos del web scraping son la comparación de precios en tiendas, la monitorización de datos relacionados con el clima de cierta región, la detección de cambios en sitios webs y la integración de datos en sitios webs”. (Gracia & Galán, 2014)

2.2. DE LA INTELIGENCIA DE NEGOCIOS.

La Inteligencia de Negocios nos ayuda a utilizar información necesaria y relevante de la empresa para la toma de decisiones y la mejora continua con diferentes herramientas. A nivel empresarial consideramos que los sistemas pueden ser vistos desde 4 niveles que son: Operativo, Táctico, Estratégico e Interinstitucional.

Las políticas institucionales son transversales a toda la empresa dentro de todas las áreas, sin embargo es común encontrar diferentes sistemas implementados, pero todos ellos deben estar en perfecta armonía para conseguir las metas que se quieren. Estos sistemas tienen visión sobre el DW, unos lo alimentarán y otros lo consultarán. Existe una relación proporcional en el sentido de que, si mayor es la institución y mayor grado de complejidad tenga su actividad, más número de sistemas la gestionarán y los mismos se enmarcan en los siguientes niveles que se muestran en la Figura 4, tanto dentro del manejo de la información, cuanto en el tipo de herramientas que se usa en los mismos.



Figura 4. Pirámide Niveles de Sistemas Información.

Adicionalmente sobre la figura anterior podemos comentar que los Sistemas estratégicos están orientados a facilitar las actividades de la dirección, ayudando a la mejor y acertada toma de decisiones. No tienen una frecuencia periódica, las peticiones al sistema son impredecibles y como ejemplo son los de: Información Gerencial, Información a Ejecutivos, Simulación de Negocios y Sistemas Expertos o de Inteligencia Artificial; como es Delphos conocido como Balanced Score Card, por su denominación en el idioma inglés, los mismos que son esquemas con múltiples dimensiones para describir, implementar y administrar una estrategia en cualquier nivel, mediante la relación de objetivos, iniciativas y mediciones a las estrategias organizacionales. Conducen a la empresa a estrategias definidas que se convierten en planes de acción.

"En el nivel táctico tenemos sistemas de apoyo en la coordinación y manejo de documentación, con el fin de facilitar las consultas, se generan informes para mejorar el trabajo de los actores intermedios de la organización y realizar una gestión efectiva; dentro de muchas herramientas nombramos a Deinsa Función". (Zarate Gallardo, 2013), que ayudan en áreas financieras, auditoría, tesorería, planificación y nos ahorran tiempo en cuanto a la elaboración y generación de un reporte como un estado financiero, permitiendo dedicar más tiempo a su análisis; sin requerir una base de datos multidimensional para operar y transformar datos.

"En los niveles operativos se solucionan las operaciones básicas de obtención masiva de datos y su tratamiento según las tareas predefinidas. Avanzaron gracias a la evolución del mercado, todos aquellos que viven el día a día de las empresas; como ejemplo nombramos a Matrix". (Jimenez Nieto, 2013), y cualquier sistema transaccional que se muestre en una plataforma robusta en base de datos y un despliegue con niveles de facilidad en su aprendizaje y uso. Y por último en la parte baja de la pirámide tenemos los sistemas interinstitucionales que forman parte del nivel más bajo de la organización los mismos que nacen a raíz de la necesidad creada cuando las organizaciones se mueven

desde un punto de vista mucho más global que obliga a establecer una comunicación estrecha entre ellas y el mercado, transfiriendo información y aquí nos referimos a todos los sistemas con capacidad de intercambio de información entre instituciones como por ejemplo: paso de nómina para el pago de sueldos a través de la banca pública o privada, pago a proveedores, registros al pago de pensiones, manejo de dinero electrónico en la banca, información personal en las entidades laborales del país, pago de impuestos y rentas.

2.3. SOLUCIONES TRADICIONALES.

Desde hace dos décadas se ha dado un gran crecimiento de productos y servicios ofrecidos por este tipo de tecnologías BI a la industria y el mundo empresarial en general, esto se ha dado por el descenso de costos para el almacenamiento en forma inicial y también por las opciones de acceso más veloces a la data. “Todo esto ha permitido trabajar en forma más agresiva con los datos, aplicando más granularidad, la misma que posteriormente permite mayor conocimiento para decisiones” (Surajit, Umeshwar, & Narasaya, 2011).

Los sistemas de inteligencia de negocios tradicionales ofrecen diferentes niveles y tipos de análisis de datos estructurados, pero hace poco no estaban diseñados para manejar los datos no estructurados, que es uno de los retos de estas plataformas.

Para estos sistemas Big Data trae grandes problemas debido a los datos que fluyen los que pueden ser estructurados o no; eso les hace enormemente limitados cuando se trata de la entrega de grandes volúmenes de datos. El camino a seguir es un replanteamiento completo de la forma en que usamos IN en términos de cómo se captura los datos, almacena y analiza. Otros problemas vienen con la necesidad de un análisis casi en tiempo

real, esto requiere la capacidad para manejar y procesar datos de alta velocidad, lo que es un desafío para los métodos de implementación de IN tradicionales, que tienen datos de latencia incorporado en su arquitectura.

Las soluciones se han desarrollado para evitar estos problemas y aportar mayor cantidad de datos como sea posible casi en tiempo real, pero estos crean sus propios problemas. La aparición de grandes volúmenes de datos requiere un enfoque radicalmente nuevo de gestión de los mismos. “Las organizaciones ahora necesitan de análisis en tiempo real de los datos estructurados y no estructurados, los enfoques tradicionales de IN que requieren la construcción de DW y datamarts están siendo en el tiempo ya difíciles de mantenerse”. (Capgemini, 2014)

2.4. LA EVOLUCIÓN DE LA INTELIGENCIA DE NEGOCIOS (IN).

Para entender la evolución de IN, haremos un recuento del avance informático en el área de hardware que se ha desarrollado con su propia velocidad, capacidades y costos; de forma diferente al software que tiene su propia tendencia evolutiva. Se conoce como la Primera Generación, en donde la primera computadora fue la máquina analítica creada por Charles Babbage, profesor matemático de la Universidad de Cambridge en el siglo XIX, la idea nació para la elaboración de las tablas matemáticas. En 1823 creó un dispositivo mecánico para efectuar sumas repetidas, luego se dedicó al proyecto de la máquina analítica que se pudiera programar con tarjetas perforadas efectuando cualquier cálculo con una precisión de 20 dígitos. En 1947 se construyó en la Universidad de Pennsylvania una primera computadora electrónica, el equipo de diseño lo encabezaron los ingenieros John Mauchly y John Eckert; esta enorme máquina tenía capacidad de realizar cinco mil operaciones aritméticas en un segundo, proyecto que fue auspiciado por el departamento de defensa de los Estados Unidos cuando se integra a ese equipo el ingeniero y

matemático húngaro John Von Neumann (1903 - 1957), quien es considerado el padre de las computadoras. En 1951 aparece UNIVAC (Universal Computer), que fue la primera computadora comercial, que disponía de mil palabras de memoria central y podían leer cintas magnéticas, se utilizó para procesar el censo de 1950 en los Estados Unidos, también usaban tarjetas perforadas.

Se considera la Segunda Generación cuando se redujo su tamaño y creció la capacidad de procesamiento, fueron construidas con circuitos de transistores y se programaron en nuevos lenguajes llamados de alto nivel, los costos para estos equipos eran muy altos.

Dentro de la Tercera Generación, se inaugura el equipo IBM 360 en abril de 1964, cuya fabricación electrónica está basada en circuitos integrados y su manejo es por medio de los lenguajes de control de los sistemas operativos. A mediados en la década de 1970, aparecen en el mercado las computadoras de tamaño mediano, o minicomputadoras que tenían costos elevados pero disponían de mayor capacidad de procesamiento.

En la Cuarta Generación aparecen los microprocesadores que es un gran adelanto de la electrónica, son circuitos integrados de alta densidad y con una velocidad impresionante, su uso se extiende al mercado industrial. Aquí nacen las computadoras personales que han adquirido proporciones enormes y que han incluido en la sociedad en general la llamada revolución informática.

Por último, en vista de una acelerada marcha de la microelectrónica hacia la sociedad industrial, en la Quinta Generación se dio la tarea de poner también a esa altura el desarrollo del software y los sistemas con que se manejan las computadoras; aparece el procesamiento en paralelo mediante arquitecturas, diseños especiales, circuitos de gran velocidad con manejo de lenguaje natural y sistemas de inteligencia artificial. Con el paso del tiempo y el avance tecnológico de los computadores se desarrollaron sistemas de información ejecutivos, modificando los tradicionales para cubrir las necesidades, apoyar las actividades y funciones de las direcciones ejecutivas de las empresas. Así también en

las comunicaciones y conexiones en red, las herramientas de IN fueron retroalimentadas de tecnología con la finalidad de crear sus rutinas básicas y reportes personalizados.

La siguiente Figura 5, representa la historia de evolución de las hoy conocidas como organizaciones inteligentes y de igual manera se presentan los avances de software que han apoyado en la toma de decisiones.

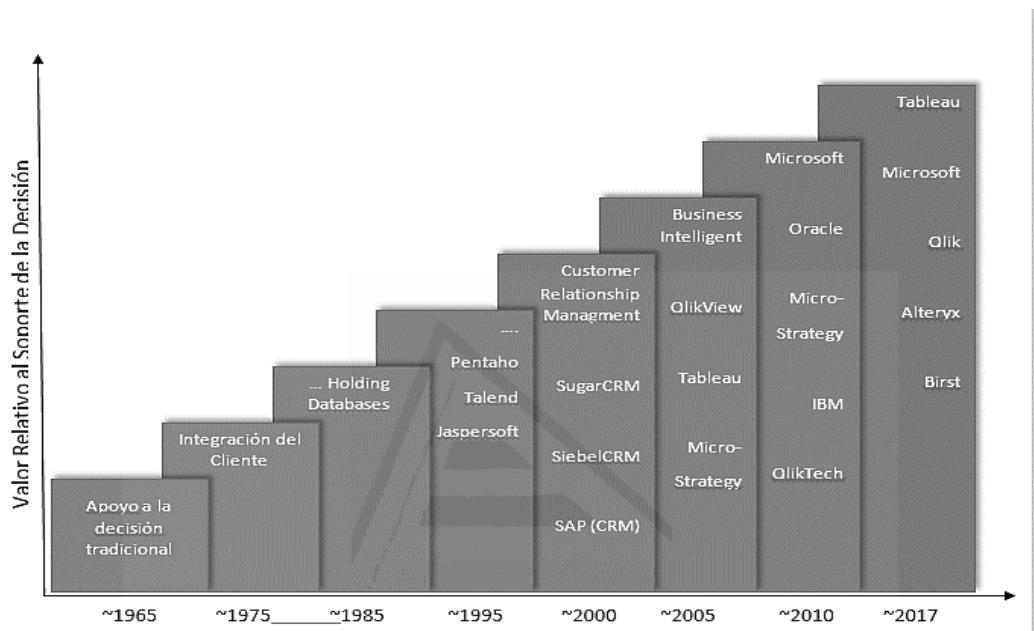


Figura 5. Evolución de IN en el soporte a la decisión. (Zarate Gallardo, 2013)

En un artículo de 1958, el investigador de IBM Hans Peter Luhn usó el término Inteligencia de Negocios, lo definió como: "la habilidad de aprender las interrelaciones entre los hechos concretos de manera que guíen las acciones hacia una meta deseada", lo que también es propuesto con diferentes términos por otros autores.(Salinas La Rosa, 2010). En 1960 los datos se almacenaban en sistemas de archivo y se consultaban con operaciones de acceso, para luego en 1965 se apoyó la decisión tradicional que consistía en reportes generados de un estándar. Para el año de 1975 se tomó en consideración la integración del cliente para la toma de decisiones, es decir se comenzó a incursionar en datos cuantitativos y cualitativos; posteriormente en 1985 la base de

datos se amplió de manera considerable y podía ser utilizada como de uso doméstico. A partir del año 1995 los data warehousing (almacén de datos) eran utilizados por las corporaciones, ya que manejaban grandes cantidades de datos y podían ingresar nuevos usuarios con un tiempo de espera de cero. El acceso a la información podía ser analizada en tiempo real y al mismo instante recibir modificaciones sin necesidad de esperar tiempo para que los cambios se realicen. Los reportes ya eran totalmente personalizados y agilizó la toma de decisiones, pues eran generados con los datos vigentes hasta el momento. En el año 2000 la aplicación del software era el orientado al Manejo de Relaciones con los Clientes conocido como CRM (Customer Relationship Management), la misma que es la combinación de data warehouse, la información de la gestión de ventas y la relación de los clientes con la empresa. “Y finalmente a partir del año 2005 nace el término de Inteligencia de Negocios IN, que será visto detalladamente en el transcurso de la investigación”. (Zarate Gallardo, 2013) , y dentro del cual hay que hacer varias consideraciones a ser tomadas en cuenta para llegar a obtener conocimiento de diferentes áreas de la información y de igual forma tendencias a las que debemos llegar (Movilidad Web 3.0)

2.4.1.NO DATAWAREHOUSE.

Data Warehousing como se expone en varios casos de empresas (Caldeira, 2012), en los cuales se indica que es un repositorio de datos de varias fuentes provenientes de diferentes tipos de sistemas de información y que estos repositorios toman fuerza para quienes pretenden maximizar sus utilidades y la operación de sus empresas. Entonces en un planeta que cada vez se vuelve más inteligente, a la IN se incorpora día a día toda esta gama de posibilidades desde teléfonos, coches con dispositivos y redes eléctricas que los alimentan con su energía, siendo los datos capturados en forma cada vez diferente.

Introducimos el concepto de NO DATAWAREHOUSE, el momento que vemos el mismo escenario que es cierto para las PYMES; donde hay también una gran cantidad de datos, los administradores tienen acceso a los informes de los sistemas Enterprise Resource Planning (ERP), CRM, la parte transaccional o libros mayores, los sistemas de ventas y más. El problema es que estos datos con demasiada frecuencia residen en sistemas separados, en silos, por lo que el reto es ponerlos juntos de manera significativa. Las hojas de cálculo de diferentes departamentos pueden ser difíciles o incluso imposibles de combinar y obtener respuestas a preguntas apremiantes del negocio es a menudo un proceso de tiempo demasiado intenso y de gran mano de obra.

Mientras que las soluciones de (IN) pueden abordar estos retos y proporcionar una ventaja competitiva, muchas compañías de tamaño medio se apartan de las iniciativas de Inteligencia de Negocios, asumiendo que es demasiado caro y complicado, de igual forma otro término equivocado es que la (IN) debe implicar un almacén de datos en toda regla y vemos que no necesariamente es así con los sistemas BI que actualmente presentan su operación en memoria. Entonces de debe actualmente enfocarse en metas exitosas tomando pasos pequeños y discretos, pero en el contexto de un plan de trabajo a largo plazo, así se obtendrán resultados más rápidos que ayuden a la toma de decisiones y se fortalecerán los proyectos de BI más extensos.

2.4.2. IN DE AUTOSERVICIO.

Conocido en el idioma inglés como self service es una capacidad del usuario final que satisface sus necesidades propias requeridas de la información, ya que pueden generar y ejecutar particulares consultas, análisis sin asistencia de un profesional. Los empleados operativos en esta ocasión pueden tomar mejores y propias decisiones de forma rápida sin esperar la larga cola de solicitudes de informes. Con esto los equipos técnicos realizan

menos seguimiento a usuarios finales, centrándose en esfuerzos e iniciativas más estratégicas. Para que sean eficientes los entornos de la inteligencia de negocios en modo de autoservicio, deben ser extremadamente intuitivos y fáciles de manejar para el usuario.

“La mayoría de usuarios comerciales y de negocio de hoy en día no disponen de los conocimientos técnicos necesarios para trabajar con interfaces o herramientas complejas, actualmente se adaptan a una forma fácil y sencilla de acceder a su información personalizada, sin necesidad de una larga formación”. (Horwitt, 2011) .

Existen herramientas para proveer autoservicio de datos, cómputo e información para no especialistas, de tal forma que el usuario logra tener independencia en el procesamiento de datos disponibles. “Con la debida autorización el cliente puede conseguir más conocimiento por medio de la información existente”. (Dictuc S, E.P., & Torres, 2010)

2.4.3. REDES SOCIALES.

“En inteligencia de negocios aplicada a redes sociales que son las más protagonistas hoy en la actualidad, se encuentran opiniones, análisis de sentimientos, libros, tesis, publicaciones, visitas entre otros; adicionalmente en las redes sociales también se han manifestado como una industria de información y entretenimiento, por la nueva manera de comunicación entre personas de diferentes partes del mundo” (Orozco & Soriano, 2016). Los usuarios buscan un vínculo para expresar su sentir en cualquier ámbito de la vida cotidiana, por esta razón en las Tablas 3 y 4, presentamos un estudio realizado en Ecuador sobre las redes sociales más populares y los usuarios registrados:

Tabla 3. Redes sociales más populares en Ecuador. (Marboleda, BRANDEC, 2017)

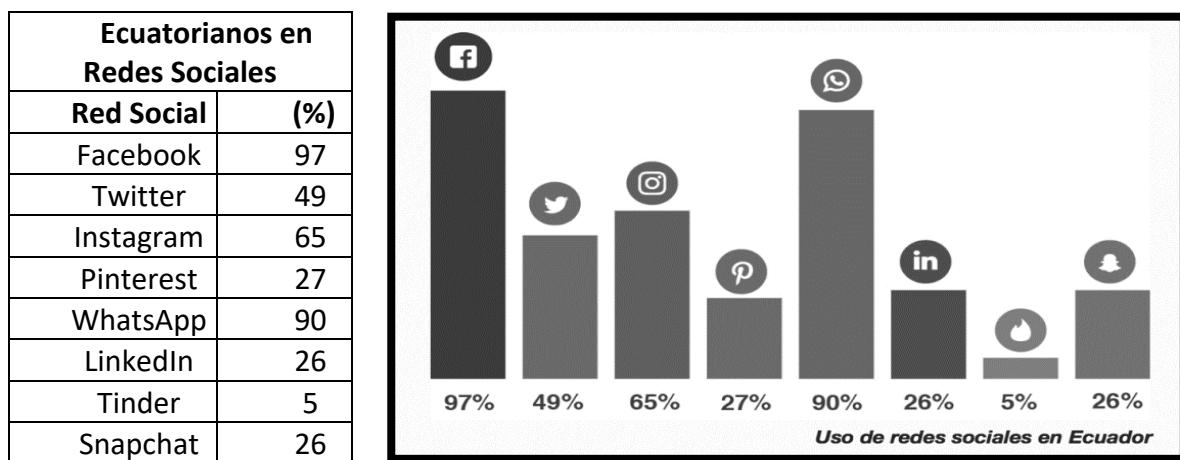
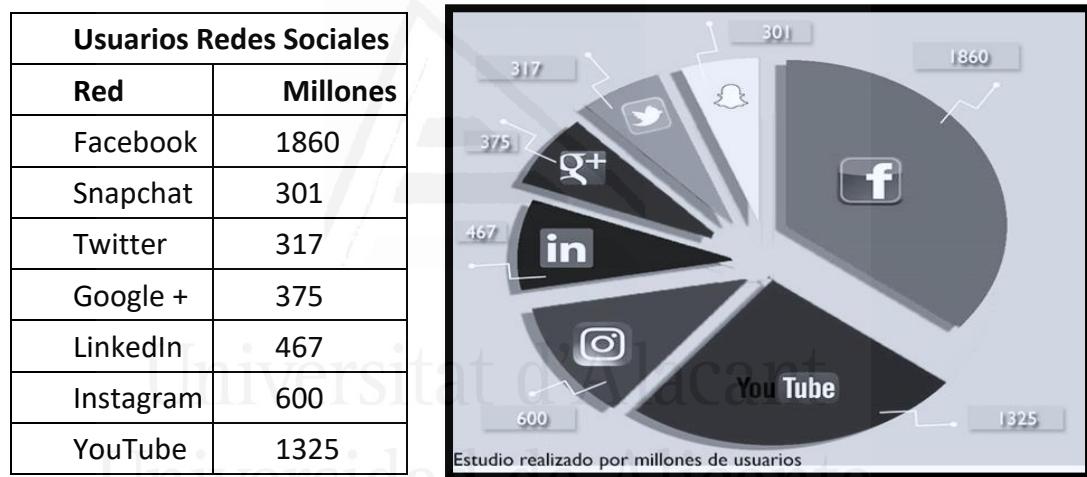


Tabla 4. Usuarios en redes sociales. Ecuador-2017 (Multiplicalia, Multiplicalia, 2017)



2.4.4. MOVILIDAD (WEB 3.0)

Los sistemas actuales de IN obligan a considerar, la posibilidad de la movilidad, disponer de un acceso inmediato a datos e información relevante en tiempo real y en cualquier momento o lugar, por esto es cada vez más necesario que la toma de decisiones se realice de un modo efectivo y con la prontitud requerida en una empresa. La movilidad, más allá de ser un factor a tener en cuenta para mejorar la agilidad de actividades ejecutivas y

comerciales de empresas u organizaciones, es una pieza clave en la implementación de cualquier sistema de IN que pretenda orientar el diseño y la ejecución de estrategias corporativas de un modo integral.

Los dispositivos móviles permiten gran variedad de usos comunicativos, entre los cuales nos permiten integrar de un modo sencillo pero efectivo estas tecnologías en nuestros sistemas de inteligencia de negocios tales como: sistema de alertas, que permiten comunicar situaciones críticas a los usuarios de los dispositivos móviles, manteniéndoles informados sobre cualquier evento o situación particular; así también tenemos los informes y reportes predefinidos, que se ejecutan con regularidad, para ser posteriormente distribuidos entre los usuarios que requieran de la información contenida en ellos para llevar a cabo su actividad; y por último los informes personalizados, que permiten el acceso a información específica mediante la creación de reportes ad hoc, elaborados en función de necesidades puntuales y particulares, que transforma datos en información según los parámetros definidos por el usuario.

2.5. ESTUDIO DE METODOLOGÍAS DE IMPLEMENTACIÓN Y DESPLIEGUE.

Para analizar las metodologías de implementación de proyectos, se debe considerar un ciclo general de vida de la IN, el cual se estableció también en la sección 1.5 de este documento; y en base a los cuales se revisarán otras propuestas válidas para la generación y seguimiento.

2.5.1. METODOLOGÍAS GENÉRICAS.

Las metodologías genéricas como las proporcionadas por el PMI (PMBOK), la ISO 21500 o PRINCE2, permiten gestionar la dirección de un proyecto de cualquier índole basado en el uso integrado de una serie de conocimientos y experiencias que garantizan el éxito del mismo mediante la planificación, programación y control de los principales parámetros especialmente de costo, tiempo y alcance.

2.5.1.1. PROJECT MANAGEMENT INSTITUTE (PMI)

El Project Management Institute (PMI) es una de las asociaciones profesionales de miembros más grandes del mundo en 180 países. PMI, es una organización sin fines de lucro que como se indica en la Figura 6, establece “La profesionalización de la dirección de proyectos a través de estándares, un esquema completo de capacitación y certificaciones reconocidas mundialmente, a través de comunidades de colaboración, de un extenso programa de investigación y de oportunidades de desarrollo profesional”. (PMI, 2017)

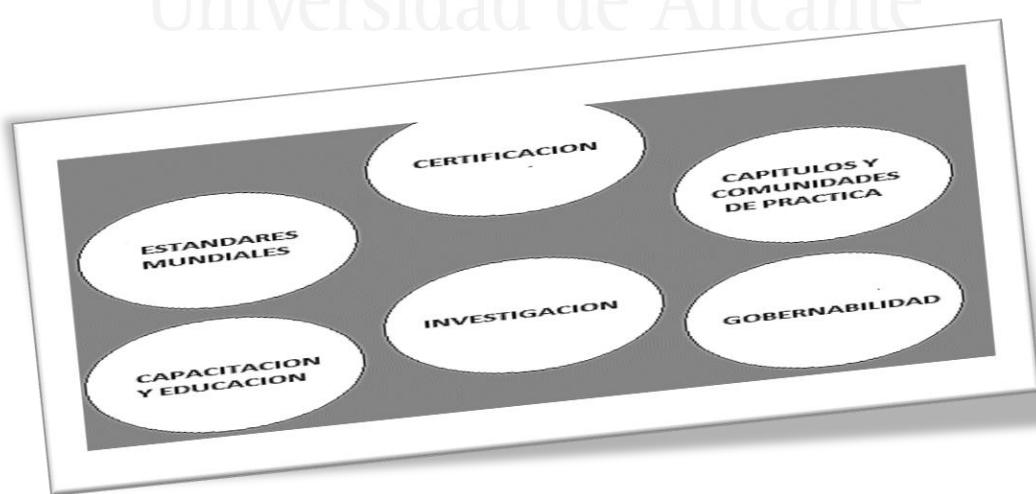


Figura 6. Concepto PMI.

Como antecedentes decimos que mientras el número de proyectos se incrementa día a día, el número de personas calificadas para manejar adecuadamente no crece de la misma manera. De hecho, de los veinte millones de personas involucradas en proyectos alrededor del mundo, solo un millón está formalmente capacitada para trabajarlos adecuadamente.

“La demanda de administradores de proyectos se encuentra en un nivel crítico urgente, pues esto le permite al individuo comunicarse en un lenguaje común, no importa su industria, su geografía o si maneja proyectos o portafolios” (Bonke & Winch, 2002). La gestión de proyectos es la aplicación de conocimientos, habilidades, herramientas y técnicas para satisfacer los requisitos planteados en el mismo. “Reconocemos que el principal valor de los proyectos va más allá de la entrega a tiempo y dentro del presupuesto, pues deben llegar a cumplir los resultados esperados del negocio” (Gartner, 2017).

La administración de proyectos es la planificación, dirección y control de recursos (personas, equipo, material) para cumplir con las restricciones técnicas, de costos y de tiempo de un proyecto; “Es una estructura temporal de la organización diseñada para lograr resultados con ayuda de especialistas de todos los puntos de la empresa y funciona bien si se cumplen los objetivos en común.” (Torres Hernández, 2014).

El Proceso de Certificación (Torres Hernández, 2014) de “PMI ofrece una serie de acreditaciones que reconocen el conocimiento y la competencia, incluyendo la certificación del Profesional en Dirección de Proyectos (PMP) que cuenta con más de 370.000 titulares alrededor del mundo, los salarios y las oportunidades de desarrollo profesional de los individuos demuestran que los empleadores reconocen el valor que representa este grupo humano capacitado”. (PMI, 2017).

En este contexto, PMBOK (Project Management Body of Knowledge), es el más famoso y reconocido producto de PMI, es una guía estándar para la administración de proyectos y puede ser entendida como una colección de sistemas, procesos y áreas de conocimiento como se presenta en la Figura 7, y los mismos son aceptados y reconocidos como los

mejores dentro de la gestión de proyectos; según Barcenas, "Es un término que describe la suma de los conocimientos involucrados en la profesión de la administración de proyectos. " El conocimiento y las prácticas descritas en el PMBOK son aplicables a la mayoría de los proyectos, sin embargo, el equipo administrador es siempre el responsable de determinar lo que es apropiado para cada uno de ellos" (Peña Torres, 2014). "El PMBOK describe procesos de gestión de proyectos, los cuales pertenecen a las Áreas de Conocimiento de Gestión de Procesos y se ejecutan como parte de Grupos de Procesos de Gestión" (Barcenas, 2012); de igual forma se dice que "Brinda un marco de referencia formal para desarrollar proyectos, guiando y orientando a los gerentes sobre la forma de avanzar en los procesos y pasos necesarios para la construcción de resultados y consecución de los objetivos. Sin embargo no debe entenderse como una metodología, ni como un manual cerrado para el desarrollo de proyectos". (Barcenas, 2012).



Figura 7. Procesos y Áreas de Conocimiento de PMBOK. (Guarin, 2016)

Los beneficios de llevar PMBOK nos llevan a lo siguiente:

- Mayor productividad debido a la utilización de una metodología estándar.
- Aumento de la rentabilidad del proyecto debido a la mejora en los procesos de planeamiento y control de los recursos del proyecto.
- Reducción de los recursos aplicados a proyectos sin valor agregado.
- Estandarización de las prácticas en todos los departamentos.
- Sistema estandarizado entre las empresas y los segmentos industriales.
- Su metodología orientada a procesos define el conocimiento necesario para administrar el ciclo de vida de cualquier proyecto, programa y portafolio.

2.5.1.2. NORMA ISO 21500

Constituye el estándar o norma internacional referente de la dirección y gestión de proyectos (DIAZ TOVAR, 2014). “El beneficio de ISO 21500, es que puede ser usado por cualquier tipo de organización, pública o privada y para cualquier segmento de proyecto, independiente de su complejidad, tamaño o duración, se traduce en una mejora de la calidad, en el cumplimiento de los requisitos del producto y/o servicio, en la gestión eficiente de los recursos y en la optimización de costos, los cuales impactan directamente la sostenibilidad, competitividad y satisfacción del cliente, generando mayor eficacia en los resultados que apuntan a los objetivos del negocio” (Calderon C, Iriarte G, & Trejos T, 2014). “Plantea claramente desde el inicio establecer qué información se va a distribuir y a quién, identificando los interesados (stakeholders) y gestionando la información desde un punto de vista concreto para cada parte interesada en el proyecto”. (DIAZ TOVAR, 2014)

Los Procesos contemplados en la Norma ISO21500, se encuentran descritos en la Figura 8, con una explicación sencilla en cada uno de ellos.

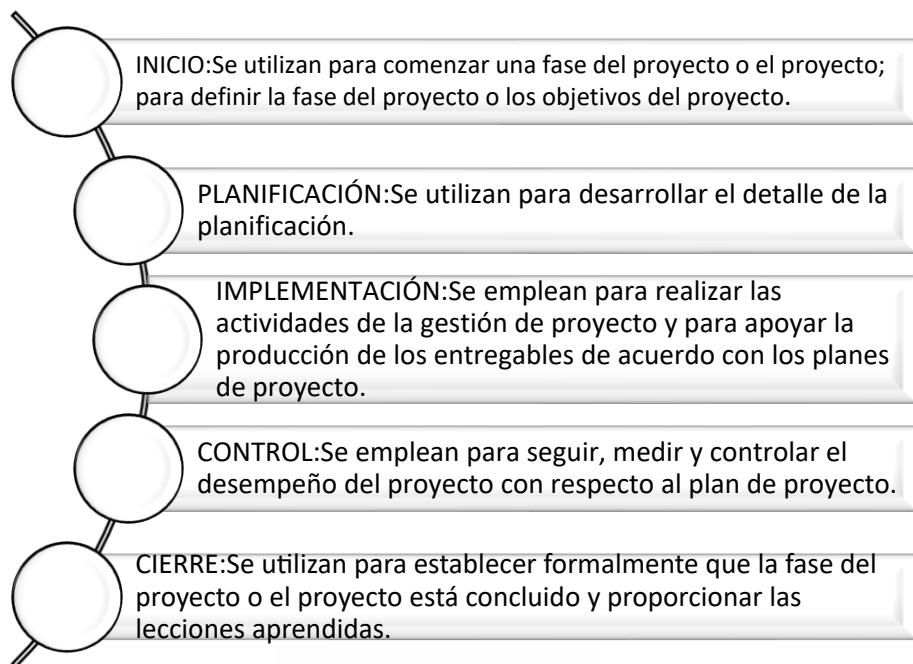


Figura 8. Grupos de Procesos ISO (Brenes, 2013)

Desde el punto de vista de otro autor, presentamos de igual forma en la Figura 9, los Procesos de ISO21500, incluyendo el tipo de documentación que se maneja:

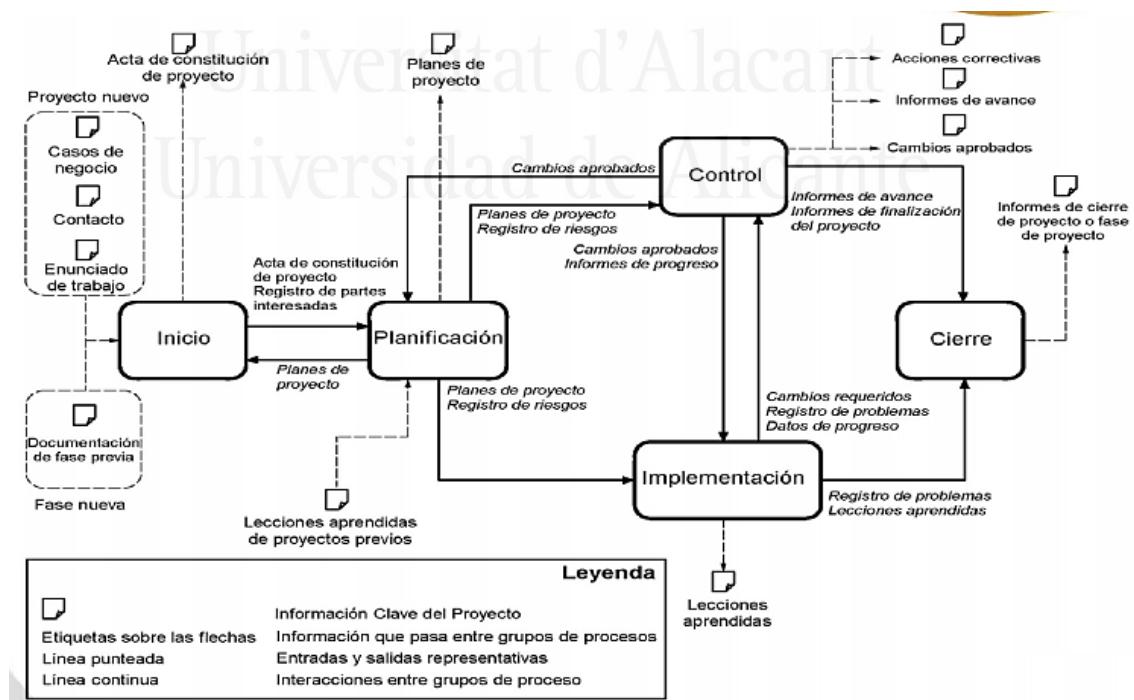


Figura 9. Directrices Dirección y Gestión Proyectos ISO21500 (Verástegui, 2014)

“Si tratamos de establecer una cronología sobre ISO 21500, podemos indicar que la constitución del comité ISO/PC 236 “Project Management” tuvo lugar en Octubre 2007 en Londres.” (Coquillat, 2014), posteriormente 8 de abril de 2013. AENOR Ediciones, la entidad legalmente responsable del desarrollo de las normas técnicas en España, ha publicado la norma Directrices para la dirección y gestión de proyectos (UNE-ISO 21500), la cual fue considerada como la versión española de la norma internacional ISO 21500 (Guidance on Project Management); ésta se editó en 2012 en el seno de ISO (Organización Internacional de Normalización) tras cinco años de trabajos en los que participaron expertos de casi 40 países, entre ellos España, y ha sido incorporada recientemente al catálogo de normas de AENOR.

“Las normas técnicas contienen las buenas prácticas y el consenso del mercado respecto a la mejor forma de abordar procesos importantes para las organizaciones y para la sociedad en general, se elaboran con la participación de las partes implicadas en cada proyecto y se las pone al alcance de todos” (Granero, 2013). En el tiempo ha recibido modificaciones importantes las mismas que según hitos relevantes se los presentan en la Tabla 5, a continuación:

Tabla 5. Cronología de Origen Norma ISO 21500 (Coquillat, 2014)

Octubre 2007 (Londres).	Constitución del comité ISO/PC 236 “Project Management”. Elaboración del Working Draft (WD) 1. (Borrador de Trabajo 1)
Abril 2008 (Washington).	Elaboración del WD 2.
Noviembre 2008 (Munich).	Elaboración del WD 3.
Junio 2009 (Tokio).	Elaboración del WD 4 que dio lugar al CD en Enero 2010.
Julio 2010 (Río de Janeiro).	Elaboración del DIS, aprobado en Septiembre 2011.
Enero 2012 (París).	Elaboración del FDIS, aprobado en Agosto 2012.
Septiembre 2012.	<i>Publicación de la Norma ISO 21500</i>

De igual forma en la Figura 10, se indica el contexto general en el cual se desenvuelve la ISO 21500.

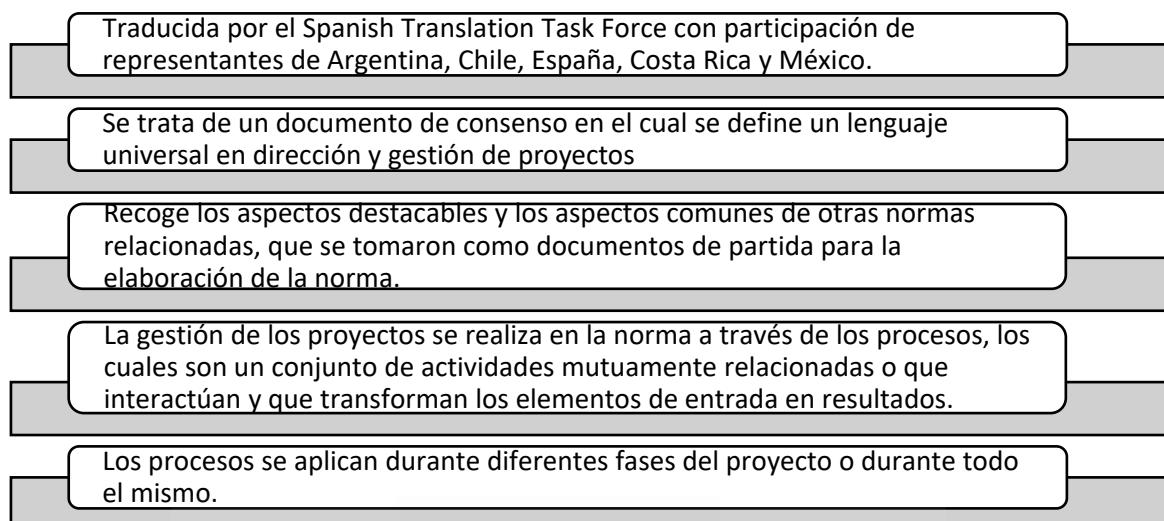


Figura 10. Directrices UNE-ISO 21500:2013 (AENOR, 2013)

El proceso de certificación de la norma puede aplicarse tanto a la acreditación de proyectos individuales como a la certificación de una cartera de proyectos, con un permanente seguimiento de los procesos de certificación con auditorías anuales.

Existe un tipo de Certificación de Proyectos Individuales (GP), en la cual : “Una vez concedida la certificación, se harán auditorías de frecuencia anual comprobando la evolución del proyecto en cuestión hasta la finalización del mismo ”. (Valdez Aguiar, 2016). Por eso se define a un proyecto como: “Un proyecto es un conjunto único de procesos que consta de actividades coordinadas y controladas, con fechas de inicio y fin, que se llevan a cabo para lograr los objetivos del proyecto.” (Brenes, 2013).

De igual manera existe una Certificación de Cartera de Proyectos (GCP) en donde: “Una vez concedida la certificación, se harán auditorías de frecuencia anual comprobando la evolución de los proyectos en ejecución, así como los de nueva incorporación a la cartera. (Valdez Aguiar, 2016). Por eso se considera un Portafolio o cartera de proyectos al: “Grupo

de proyectos de interés para una organización” (Fernández, 2013). De otra forma un Programa de proyectos como un: “Grupo de proyectos que se dirigen de manera coordinada por que comparten objetivos, riesgos, y otros” (Fernández, 2013); y por último el “Triángulo del Proyecto, que es la relación entre coste, plazo y alcance, que normalmente estas suelen ser las variables claves del proyecto y si se modifica alguna de ellas, implicará modificaciones en el resto”. (Fernández, 2013).

Se consideran como Beneficios de la norma ISO las siguientes:

- “Permitir que las empresas conecten sus objetivos con proyectos exitosos.” (Martinez, 2014)
- “Potenciar la mejora continua en las organizaciones, el mejor uso de recursos, la calidad en los proyectos y la eficiencia en los procesos de licitación gracias al uso de terminología coherente de Dirección de Proyectos.” (Martinez, 2014)
- “Tener en cuenta sinergias positivas o identificar posibles conflictos con las partes interesadas desde las etapas iniciales del proyecto.” (Martinez, 2014)
- “Mejorar la empleabilidad y movilidad de los usuarios y su capacidad para trabajar en proyectos internacionales”. (Agenjo, 2014).

2.5.1.3. PROYECTOS EN AMBIENTES CONTROLADOS PRINCE2

“El nombre de PRINCE2 viene de las palabras en inglés “Projects IN Controlled Environments”, es un método de gestión de proyectos que cubre la gestión, el control y la organización de un proyecto” (Lopez, 2014), que fue originalmente desarrollado por la Agencia Central de Informática y Telecomunicaciones (CCTA) en el Reino Unido, que actualmente forma parte de la Oficina Gubernamental de Comercio (OGC). Es un enfoque estructurado a la gestión de proyectos que se publicó en 1996 como un método genérico el que combina la metodología original PROMPT (la cual evolucionó en el método PRINCE) con el MITP de IBM (gestión de la implementación del proyecto total)”. Desde 1989 se

viene usando como un estándar para la gestión de proyectos, sobre todo en el Reino Unido, fue inicialmente desarrollado únicamente para proyectos de Tecnologías de Información y Comunicación (TIC), aunque la última versión es compatible con la gestión de todo tipo de proyectos. La revisión más reciente se publicó el 16 de junio de 2009 por la OGC siendo denominada esta versión como PRINCE2:2009. Este modelo consta de 4 elementos como se tienen establecidos en la Figura 11 y cada uno de ellos tiene sus particularidades, además del entorno que engloba a dos de ellas en los proyectos.

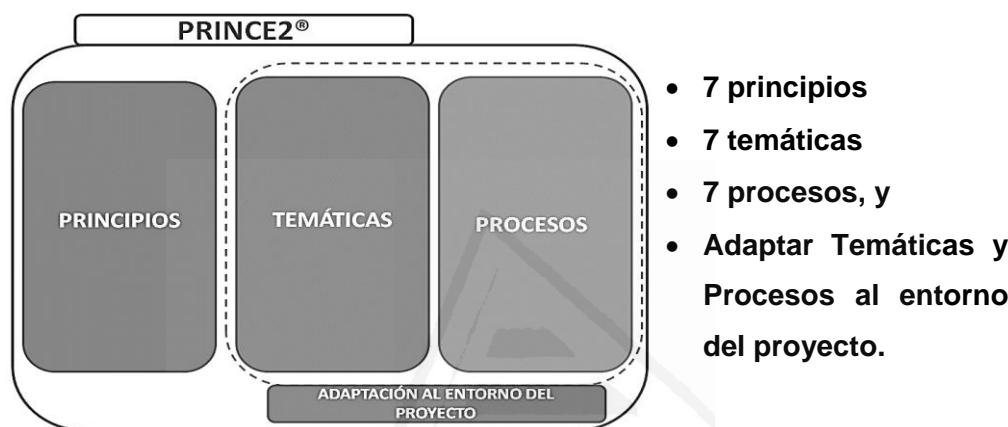


Figura 11.Los 4 elementos de Prince2. (Vila Grau, 2015)

Los 7 Principios se encuentran en la Tabla 6, siendo las obligaciones y buenas prácticas que determinan si un proyecto está siendo gestionado; y más aún son de obligatorio cumplimiento.

Tabla 6. Los 7 principios de Prince2. (Agueda Barrero, 2014)

PRINCIPIOS	DEFINICIÓN
1. Justificación comercial continua	Se requiere un motivo para iniciar un proyecto, que debe mantenerse en el tiempo y documentarse.
2. Aprender de la experiencia	Los equipos de proyectos PRINCE2® aprenden de la experiencia (lecciones aprendidas) a lo largo de toda la vida del proyecto.
3. Roles y Responsabilidades definidos	Se requiere de un equipo de gestión del proyecto con roles y responsabilidades bien definidos.
4. Gestión por fases	Los proyectos PRINCE2® se planifican, supervisan y controlan fase por fase. Al final de cada fase debe ser evaluado el estado del proyecto para asegurar que este se mantiene viable.
5. Gestión por excepción	Se establecen tolerancias para cada objetivo de tiempo, coste, calidad, alcance, riesgo y beneficio del proyecto que permiten delegar autoridad.
6. Enfoque en los productos	Los productos del proyecto se acuerdan y definen antes de realizar las actividades requeridas para producirlos.
7. Adaptación al entorno del proyecto	PRINCE2® se adapta al entorno, tamaño, complejidad y nivel de riesgo del proyecto.

Las 7 temáticas describen los ámbitos o aspectos del proyecto que deben gestionarse y adaptarse a su entorno continuamente durante todo su transcurso, las mismas se encuentran resumidas en la Tabla 7, y básicamente estas responden a los cuestionamientos que se hacen dentro del proyecto.

Tabla 7. Las 7 Temáticas de Prince 2. (Vila Grau, 2015)

TEMÁTICAS	CONTESTA AL.
1. Caso de Negocio	¿Por qué?
2. Organización	¿Quién?
3. Calidad	¿Qué?
4. Planes	¿Cómo? ¿Cuándo? Y ¿Cuánto?
5. Riesgos	¿Qué pasa si...?
6. Cambios	¿Cuál es el impacto?
7. Progreso	¿Dónde estamos ahora? ¿A dónde vamos? ¿Deberías continuar?

Los 7 procesos considerados en PRINCE2® que se suceden a lo largo de la vida del proyecto (Tabla 8), establece para cada uno la lista de actividades, los productos de gestión, y las responsabilidades de cada rol.

Tabla 8. Los 7 Procesos de Prince 2. (Propia, 2016)

PROCESO	DEFINICIÓN
1. Dirección de un Proyecto	DP (Directing a Project): Este proceso es para la Gestión Superior (la Junta de Proyecto controla el proyecto).
2. Puesta en Marcha de un Proyecto	SU (Starting Up a Project): Se trata de un proceso pre-proyecto muy corto que reúne los datos necesarios para comenzar el proyecto.
3. Iniciar un Proyecto	IP (Initiating a Project): El proceso examina la justificación del proyecto y crea la Documentación de Inicio del Proyecto (PID) que incluye el Plan del Proyecto (Project Plan).
4. Control de una Fase	CS (Controlling a Stage): Este proceso describe las tareas diarias de vigilancia y de control que realiza el Jefe de Proyecto sobre el proyecto. Aquí es donde el Jefe de Proyecto pasa la mayor parte de su tiempo en un proyecto.
5. Gestión de los Límites de Fase	SB (Managing a Stage Boundary): Este proceso proporciona una forma controlada de completar una fase y planear la siguiente.
6. Gestión de la Entrega de Productos	MP (Managing Product Delivery): Este es el proceso de entrega de los productos. Es donde los productos (Productos Especializados), que van a ser utilizados por los usuarios, son entregados por los miembros del equipo.
7. Cerrar un proyecto	CP (Closing a Project): Este proceso confirma la entrega de los productos y el Jefe de Proyecto prepara el cierre del proyecto.

La Adaptación al entorno del proyecto se refiere a la necesidad de enmarcar los elementos de PRINCE2® al contexto específico de cada proyecto, con el objetivo de no sobrecargarlo y ofrecer el nivel de control apropiado, como se ejemplifica en la Tabla 9.

Tabla 9. Adaptación de diferentes elementos de Prince 2. (Vila Grau, 2015)

ADAPTACIÓN	PRINCIPIOS	TEMÁTICAS	PROCESOS		
			Actividades	R&R	Productos*
SI		Incorporando las políticas y normas corporativas o del programa pertinentes	-	Incorporando los términos de la organización o programa	Incorporando los términos de la organización o programa y adaptando el contenido
NO	Se aplican siempre	-	Es necesario realizar todas las actividades	El contenido de los Roles, ni las responsabilidades	-

*Productos de gestión

Los beneficios proyectados de esta metodología se consideran que son:

- Un comienzo organizado y controlado, asegurando que así como existe una razón válida para el proyecto y que todas las personas y los recursos están armonizados y listos para el mismo.
- Un desarrollo organizado y controlado, asegurando que, una vez que el proyecto ha comenzado, su validez se mantiene y el trabajo realizado es monitoreado durante toda su duración.
- Un final organizado y controlado es siempre mejor que la tendencia natural a la deriva en la vida operativa del producto de un proyecto; por ello, tenemos que asegurarnos de que la entrega inmediata del proyecto ha sido aceptada por nuestro cliente antes de cerrar el proyecto oficialmente.
- Revisiones periódicas de los progresos.
- Flexibilidad en las decisiones.
- Dirección de control automática en caso de desviación del plan original gracias al principio de “Gestión por Excepción”.
- Buena comunicación entre el, la dirección del proyecto y el resto de la organización.

“La evaluación de los proyectos, antes, durante y después de su desarrollo es importante por cuanto contribuye a que se cumplan las normas de seguridad de los procesos; permite conocer la efectividad, utilidad y eficiencia del sistema, aporta

información real sobre la forma de maximizar el éxito del proyecto y garantizar la continuidad de iniciativas de similares características.” (Martinez A. , 2001)

El Proceso de Certificación para llegar a ser un Practitioner registrado, consiste en realizar el examen PRINCE2 Foundation, con el cual un candidato será capaz de actuar como miembro informado de un equipo de gestión de proyectos; y por otro lado, a continuación el examen PRINCE2 Practitioner, con lo que un candidato está en capacidad de poner en marcha y gestionar un proyecto en su totalidad.

2.5.2. METODOLOGÍAS ESPECÍFICAS.

Otras metodologías específicas proporcionan una guía de construcción de producto, como las propuestas por Bill Inmon y otra que la consideramos fue la realizada por Ralph Kimball, en las cuales se muestran diferentes enfoques como el desarrollo en cascada y consideran diseño de almacenes de datos que especifican prácticas adecuadas estrictamente para un fin. En una de ellas se aplican principios tales como que la colección de datos debe ser orientada a un ámbito, integrado, no volátil y variable en el tiempo.

2.5.2.1. “BUILDING THE DATA WAREHOUSE” POR BILL INMON.

“Bill Inmon fue uno de los primeros autores en escribir sobre el tema de los almacenes de datos, en **1990** publica **“Building the Data Warehouse”**, en **2002** mejora su libro y define una arquitectura como una colección de fuentes dispares en almacenes de datos y variantes en el tiempo” (Inmon W. , 2012) y define un DW bajo las siguientes características: “ Orientado a temas, puesto que los datos, en la BDD están organizados de manera que todos los elementos relativos al mismo evento u objeto del mundo real

queden unidos entre sí; por otro lado lo ve como Variante en el tiempo, puesto que los cambios producidos en los datos a lo largo del tiempo quedan registrados para que los informes que se puedan generar reflejen esas variaciones; así también dice que DW es No volátil, ya que la información no se modifica ni se elimina, una vez almacenado un dato, éste se convierte en información de sólo lectura, y se mantiene para futuras consultas; y por último indica que es Integrado, porque la base contiene los datos de todos los sistemas operacionales de la organización, y los mismos deben ser consistentes” (Pyle, 2003) .

Entonces Inmon, defiende una metodología descendente (top-down) a la hora de diseñar un almacén de datos, ya que de esta forma se considerará mejor toda la implementación corporativa. “En esta metodología los Data Marts se crearán después de haber terminado el DW completo de la organización”. (Pyle, 2003). El DW se nutre de un solo ETL, pero en este caso no está modelado dimensionalmente, sino que está en tercera forma normal (3NF); dicho de otra forma esta propuesta Inmon, consiste en que formar el DW, será el resultado de tomar todas las bases de datos y sistemas de información en una organización como se lo esquematiza en la Figura 12. Así otra definición dice: “Un Data Warehouse o depósito o almacén de datos es una colección de datos: orientado a temas, integrado, no volátil, de tiempo variante, que se usa de soporte al proceso de toma de decisiones”. (Inmon W. H., 2001).

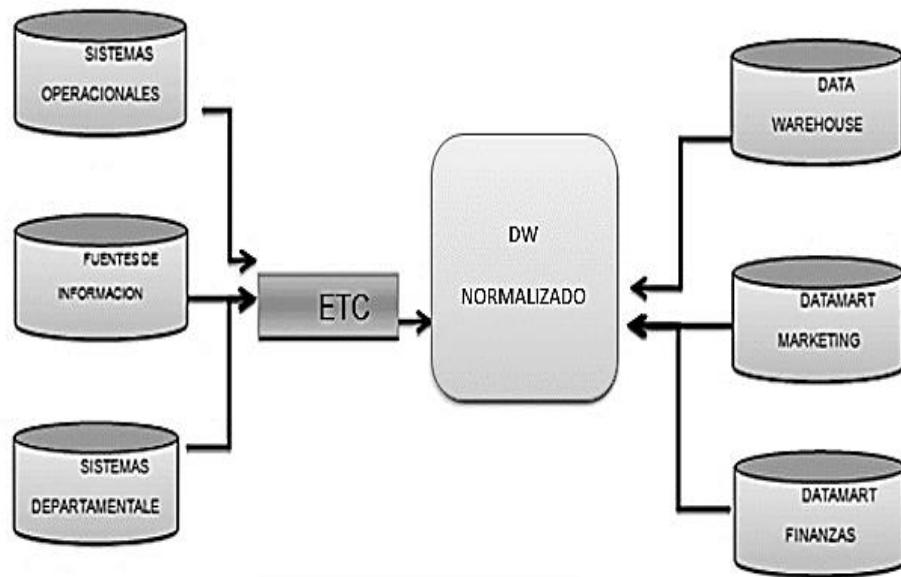


Figura 12. Enfoque de Bill Inmon

A raíz de esto, surge la inteligencia de negocio como un “nuevo concepto orientado al tratamiento inmediato de los datos, la información y en definitiva al conocimiento, con el fin de mejorar los procesos de cualquier organización frente a las exigencias del mercado” (Reinschmidt & Allison, 2000)

2.5.2.2. “THE DATA WAREHOUSE TOOLKIT” POR RALPH KIMBALL.

Cronológicamente hablando Ralph Kimball en 1996 Publica “The Data Warehouse Toolkit”, luego en 1997, declaró que: "... el almacén de datos no es más que la unión de todos los datamarts”; y posteriormente en 2002 mejora su libro y define múltiples bases de datos llamados datamarts que son organizados por procesos de negocio.

“La Metodología Kimball, es una metodología empleada para la construcción de un almacén de datos (DW) que no es más que, una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad que se utiliza” (Rivadera G. R., 2014). Su enfoque se grafica en la Figura 13, y presenta una metodología Bottom-up, pues al final el Datawarehouse Corporativo no es más que la unión de los diferentes datamarts, que están estructurados de una forma común a través de la arquitectura de bus con la que cuenta. “Esta característica le hace más flexible y sencillo de implementar, pues podemos construir un Data Mart como primer elemento del sistema de análisis, y luego ir añadiendo otros que comparten las dimensiones ya definidas o incluyen otras nuevas. En este sistema, los procesos ETC extraen la información de los sistemas operacionales y los procesan igualmente en el área stage, realizando posteriormente el llenado de cada uno de los Data Mart de una forma individual, aunque siempre respetando la estandarización de las dimensiones”. (Espinosa, 2010)

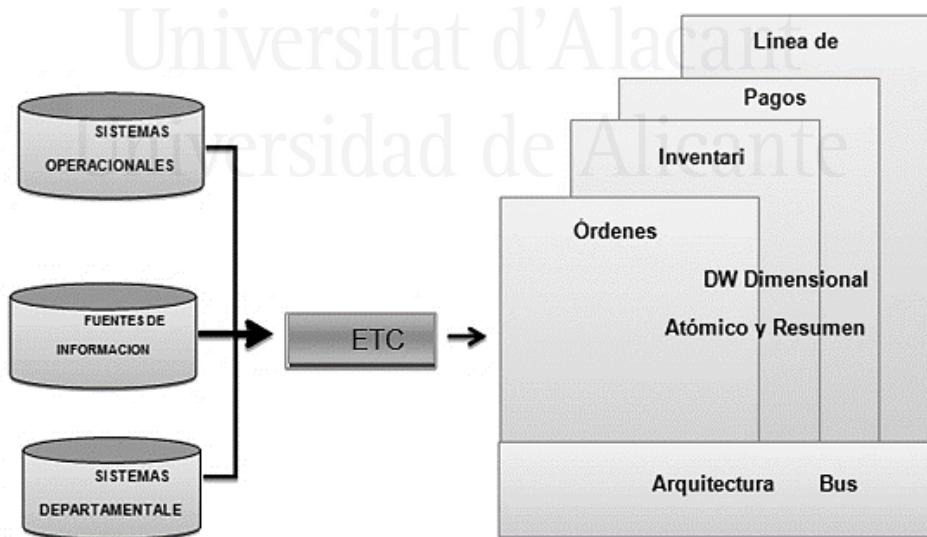


Figura 13. Enfoque Kimball.

El Modelo de datos se constituye con Hechos (contiene métricas), Dimensiones (contiene la normalización) y todo es accesible por el usuario en donde su bus de datos se mueve en el área intermedia. Consideramos el ciclo de vida de Kimball como lo muestra la Figura 14 y las mismas etapas son detallados en la Tabla 10 con una descripción clara e indicando sus actividades y acciones principales.

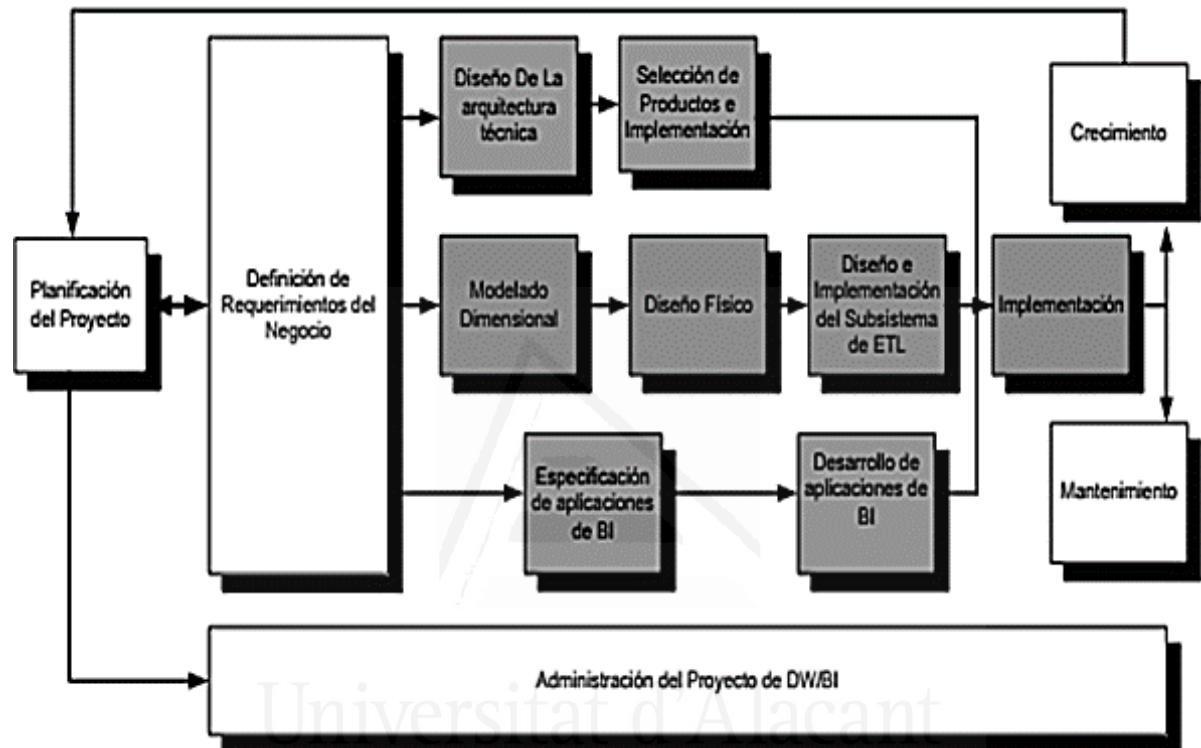


Figura 14. Ciclo de Vida de Metodología Kimball. (Rivaderra G. , 2010)

Tabla 10. Etapas Metodología Kimball.

ETAPA	ACTIVIDADES Y ACCIONES
<p>1. PLANIFICACIÓN: Se determina el propósito del proyecto de DW/BI; Objetivos específicos, Alcance, Riesgos y necesidades de información.</p>	<ul style="list-style-type: none"> • Definir el alcance (entender los requerimientos del negocio). • Identificar las tareas • Programar las tareas • Planificar el uso de los recursos. • Asignar la carga de trabajo a los recursos • Elaboración de documento final. (plan del proyecto)
<p>2. DEFINICIÓN DE REQUERIMIENTOS DEL NEGOCIO: Entrevistar al personal: negocio y técnico. Aprender el negocio, competidores, industria y Clientes. Revisar informes posibles, rastrear documentos estratégicos; entrevistar a los empleados, Conocer los términos y la terminología del negocio.</p>	<p>Cuatro grupos a entrevistar:</p> <ul style="list-style-type: none"> • El directivo que toma las decisiones estratégicas. • Los administradores intermedios y de negocio responsables de explorar alternativas estratégicas y aplicar decisiones • El personal de sistemas, si existe. • El personal que se entrevista por razones políticas.
<p>3. MODELADO DIMENSIONAL: Es un proceso dinámico y altamente iterativo. Comienza con un modelo dimensional de alto nivel obtenido a partir de los procesos priorizados y descritos en la tarea anterior.</p>	<ul style="list-style-type: none"> • Elegir el proceso de negocio: (Dirección) • Establecer el nivel de granularidad: • Elegir las dimensiones: (Atributos posibles candidatos para ser encabezados en los informes, tablas pivot, cubos, o cualquier forma de visualización, unidimensional o multidimensional) • Identificar medidas y las tablas de hechos: Atributo analizar, sumando o agrupando sus datos y usando los criterios de corte conocidos como dimensiones. Las medidas se encuentran en tablas que denominamos tablas de hechos.
<p>4. DISEÑO FÍSICO: En esta etapa se contestan las siguientes preguntas</p>	<p>¿Cómo puede determinar cuán grande será el sistema de DW/I? ¿Cuáles factores de uso para una configuración más grande? ¿Cómo se debe configurar el sistema? ¿Cuánta memoria y servidores se necesitan? ¿Qué tipo de almacenamiento y procesadores? ¿Cómo instalar software en servidores de desarrollo, prueba, producción? ¿Qué necesitan instalar los diferentes miembros en estaciones? ¿Cómo convertir el M. datos lógico en un M.datos físicos? ¿Cómo conseguir un plan de indexación inicial? ¿Debe usarse la partición en las tablas relacionales?</p>
<p>5. DISEÑO E IMPLEMENTACIÓN DEL SUBSISTEMA DE EXTRACCIÓN, TRANSFORMACIÓN Y CARGA (ETC): Base sobre la cual se alimenta el DW.</p>	<p>Si se diseña adecuadamente, puede extraer los datos de los sistemas de origen, aplicar diferentes reglas para aumentar la calidad y consistencia de los mismos, consolidar la información proveniente de distintos sistemas, y finalmente cargar (grabar) la información en el DW en un formato acorde para la utilización por parte</p>

	<p>de las herramientas de análisis, en una periodicidad definida.</p>
6. IMPLEMENTACIÓN: Convergencia de tecnología, datos y aplicaciones de usuarios finales, accesible desde el escritorio.	<p>Existen varios factores extras que aseguran el correcto funcionamiento de todas estas piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación y las estrategias de retroalimentación.</p>
7. MANTENIMIENTO Y CRECIMIENTO DEL DATA WAREHOUSE: Para administrar el entorno del DW actual es importante enfocarse en los usuarios de negocio, los cuales son el motivo de su existencia, medir y proyectar su éxito.	<ul style="list-style-type: none"> • Comunicarse constantemente con los usuarios para establecer un flujo de retroalimentación. • Sentar las bases para el crecimiento y evolución del DW de forma iterativa utilizando el Ciclo de Vida propuesto, y establecer las oportunidades de crecimiento y evolución en orden por nivel de prioridad. • Tareas asociadas al área de Aplicaciones de BI, en esta ruta se encuentran aquellas en las que diseñamos y desarrollamos las aplicaciones de negocios para los usuarios finales.
8. ESPECIFICACIÓN DE APPLICACIONES IN.: En esta tarea se proporciona, a una comunidad de usuarios de forma más estructurada y más fácil, de acceder al almacén de datos; y esto es aplicaciones de BI, que son quienes proporcionan información útil a los usuarios. Las aplicaciones de BI incluyen un amplio espectro de tipos de informes y herramientas de análisis, que van desde resúmenes simples de formato fijo, a sofisticadas aplicaciones analíticas que usan complejos algoritmos e información del dominio.	<p>Según su nivel de sofisticación:</p> <p>a) Informes estándar: Relativamente simples, de formato predefinido, y parámetros de consulta fijos, proporcionan a los usuarios un conjunto básico de información acerca de lo que está sucediendo en un área determinada de la empresa y se utilizan día a día.</p> <p>b) Aplicaciones analíticas: Son más complejas, pueden incluir algoritmos y modelos de minería de datos, que ayudan a identificar oportunidades y el usuario puede pedir cambios en los sistemas transaccionales basándose en los conocimientos obtenidos del uso de la aplicación de IN. Ejemplos:</p> <ul style="list-style-type: none"> • Análisis de la eficacia de las promociones • Análisis de rutas de acceso en un sitio Web • Análisis de afinidad de programas • Planificación del espacio en espacios comerciales • Detección de fraudes • Administración y manejo de categorías de productos.
9. DISEÑO DE LA ARQUITECTURA TÉCNICA: Cubre los procesos y herramientas (back y front room) que se aplican a los datos.	<ul style="list-style-type: none"> • “El back room: es el responsable de la obtención y preparación de los datos, por lo que también se conoce como adquisición de datos, y • El front room: es responsable de entregar los datos a la comunidad de usuario y también se le conoce como acceso de datos”. (Rivaderra G. R., 2014).

3. CAPÍTULO III. ANÁLISIS METODOLÓGICO.

Existen muchas metodologías de diseño y construcción de proyectos, cada fabricante o institución que maneja software de inteligencia de negocios, busca imponer una metodología con sus productos o en base a la experiencia alcanzada en el tiempo. Sin embargo, en esta parte analizaremos parámetros tanto en genéricas como específicas para mejorar aspectos en los procesos ETC incorporando refinamientos en el uso de datos con uso de algoritmos de decisión modificados, todo esto en una propuesta que se plantea y posteriormente la misma será validada en dos casos de estudios reales.

3.1. COMPARACIÓN DE METODOLOGÍAS.

Una vez revisadas diferentes metodologías tanto genéricas como específicas, estas serán analizadas tomando en cuenta ciertos criterios que consideramos importantes para cada una de ellas.

3.1.1. METODOLOGÍAS GENÉRICAS

Luego de revisar las propuestas metodológicas de tipo genérico como fueron ISO21500, PMBOK y PRINCE2; en la Tabla 11, presentamos un análisis sobre características que dentro de esta investigación pueden aportarnos en la definición a la que se llegará con M3S.

Tabla 11. Cuadro Comparativo. Metodología Genéricas (Cruz Caicedo, 2012)

CARACTERÍSTICAS	ISO 21500	PMBOK	PRINCE 2
Tipo	Norma	Marco de Buenas Prácticas	Método de gestión de proyectos
Año de Inicio	2006	Finales de 1990's	1989
Alineación Estratégica, Portafolio, Programa y Proyecto	SI (aunque diferencia entre Portafolio y Programa NO es clara)	SI (más detallado)	SI
Gobierno de Proyectos y Creación de Valor	SI	SI	NO
Ciclo de vida de proyecto y Producto	Solamente ciclo de vida de proyecto	Ambos	Solamente ciclo de vida de proyecto
Roles en la Gerencia de proyecto	Competencias Generales del Personal del proyecto	Rol Gerente Proyecto, Equipo proyecto, Patrocinador, Gerentes funcionales.	Competencias Generales del Personal del proyecto
Restricciones del proyecto	Alcance, Tiempo, Costo, Calidad y Recursos	Adicionalmente Riesgo	Ambos

De lo más relevante de esta tabla es que todas las metodologías son compatibles entre sí y definitivamente pueden ser usadas complementariamente según el caso de proyecto que se esté llevando a cabo, o según los expertos que las apliquen. Por eso dentro de esta

investigación en la metodología propuesta se combinan algunos tópicos de cada una de ellas.

Los procesos y prácticas para gestión de tiempo propuestos por PMI han mostrado ser muy buenos a la hora de generar un plan de proyecto y estimar una fecha de finalización, lo que no se ha logrado conseguir en forma efectiva por las otras dos. ISO21500, tiene problemas a la hora de realizar el control del proyecto a lo largo del ciclo de vida porque igual tiene tareas sin importar el tamaño del mismo, como así también en la gestión de cambios en donde PMI, tiene mejoras y Prince 2 para ser la más robusta en este sentido.

Por sus equipos de trabajo certificados PMI lleva proyectos de software que se caracterizan por ser dinámicos, “las tareas no se realizan en un orden preestablecido y también suelen aparecer nuevas tareas para realizarlas sin problema”. (Freijedo & Grassi, 2012). Por último podemos mencionar que Prince 2, modela de mejor manera la realidad cotidiana del desarrollo de software.

3.1.2. METODOLOGÍAS ESPECÍFICAS

De igual forma una vez revisadas las propuestas específicas aquí mencionadas en lo que se refiere básicamente a sus procesos y etapas, en la Tabla 12, valoramos características a tomarse en cuenta para la propuesta de solución que tendrá esta investigación.

Tabla 12. Comparativo Modelos Inmon y Kimball (Propia, 2016)

CARACTERÍSTICAS	INMON	KIMBALL
Enfoque	Top-down	Bottom up
Complejidad del Método	Alta	Baja
Orientación de la data	Maneja data o áreas	Orientado a procesos
Herramientas	Tradicional (ER y DIS)	Modelación Dimensional
Accesibilidad del usuario final	Bajo	Alto
Objetivos	Ofrecer una buena solución técnica basada en métodos probados	Ofrecer una solución que sea fácil para usuarios finales, consulta de datos y tener una tasa de respuesta razonable
Naturaleza de los requerimientos	Estratégica	Táctica
Requerimientos de integración de datos	Organización	Áreas de negocio individual
Estructura de Datos	No métricas	Métricas de negocio, rendimiento, y Cuadros de mando.
Persistencia de la data	Alta tasa de cambio de sistemas fuentes.	Los sistemas fuentes son relativamente estables.
Requerimientos de personal y habilidades	Equipos grandes de especialistas	Equipos pequeños generalistas.
Tiempo de Entrega	Los requerimientos de la organización permiten la puesta en marcha con más tiempo.	La necesidad de la primera aplicación datawarehouse es urgente.
Costo de Implementación	Altos costos de inicio con menores costos de desarrollo de proyectos posteriores	Menor costo de arranque, con cada proyecto posterior el costo es el mismo.

De la tabla se desprende entre otras cosas que mantienen enfoques diferentes de lo general a lo particular y lo contrario; Kimball muestra mayor facilidad debido a su orientación a los procesos que en algunos casos podría ser modelado en forma más sencilla. Inmon trata de visualizar a toda la organización en su conjunto, mientras que Kimball va sobre un nivel más táctico con lo que consigue de forma más efectiva encontrar indicadores de gestión. Esta orientación influye también en el tiempo que lleva a Kimball a ser más urgente y a su vez sus equipos de trabajo también suelen ser más pequeños. Y por último es muy importante tomar en cuenta que Inmon tiene costos más altos para iniciar el desarrollo de los proyectos, por lo que le hace orientarse hacia un segmento específico de empresas de mayor tamaño.

3.2. ANÁLISIS DE HERRAMIENTAS TECNOLÓGICAS IN

Para analizar la evolución de las herramientas para manejo de IN se revisaron dos empresas consultoras referentes a nivel mundial, de las cuales vemos la forma en la que han calificado a diferentes softwares en este ámbito.

3.2.1. GARTNER GROUP

Es importante señalar que Gartner Group, empresa reconocida en el mundo que proporciona información sobre diferentes ámbitos de la industria en general, basa sus rankings en cuestionarios a clientes y vendedores; toma en cuenta por tanto la opinión de los expertos y de quienes usan estos programas. Cada proveedor es juzgado en su capacidad para ejecutar su estrategia de negocio y la completitud de su visión, que Gartner define como la habilidad de crear valor para los clientes.

El análisis que realiza esta empresa lo plasma en un cuadrante con las siguientes denominaciones: Líderes que son aquellos ubicados arriba en la parte derecha y que tienen la mayor puntuación resultante al combinar su habilidad para ejecutar (lo bien que un vendedor vende y ofrece soporte a sus productos y servicios a nivel global) y el alcance de visión, que se refiere a su potencial. Luego se considera la sección de Retadores, arriba a la izquierda, quienes son caracterizados por ofrecer buenas funcionalidades y un número importante de instalaciones del producto, pero sin la visión de los líderes. A continuación los Visionarios que están abajo a la derecha y son quienes tienen todas las capacidades que ha de ofrecer un gestor de contenido de la empresa de forma nativa, o mediante alianzas con otros socios, con lo cual Inteligencia de Negocio significa un fuerte impulso a la integración de programas y plataformas, así como una habilidad para anticiparse a las necesidades del mercado que ellos no puedan cubrir. Por último, Nichos Específicos, los que están enfocados a determinadas áreas de las tecnologías, pero sin disponer de una suite completa.

En esta clasificación dentro de su cuadrante es importante indicar que todos se mueven bajo coordenadas X y Y; el Eje X, mide la profundidad y amplitud de los objetivos de un proveedor, su conocimiento de los mercados y los clientes que atiende, así como también su posición para hacer frente a futuros escenarios de la industria. El Eje Y, se centra en la gestión de equipo y la estabilidad financiera, los canales de venta, la calidad de la investigación y el desarrollo, la base instalada de productos en el mercado, servicio y soporte de la reputación y trayectoria en la entrega de productos a tiempo.

A continuación, e interpretado bajo las Tablas 13,14 y 15, se disponen de algunos de los últimos años analizados por Gartner Group a partir del 2009, información que se encuentra publicada todos los meses de enero; y esta de hecho es un indicador serio de cómo se han dado los avances en tecnología referente en este caso a Inteligencia de negocios.

Tabla 13. Interpretación Gartner BI 2009

CATEGORIA	SOFTWARE BI
Líderes	<ul style="list-style-type: none"> • IBM (Cognos) • Microsoft (Power BI) • Oracle • Sap (Business Object) • SAS • Information Builders • MicroStrategy
Retadores	<ul style="list-style-type: none"> • No se presentan
Visionarios	<ul style="list-style-type: none"> • QlickTech • Tibco Spotfire
Jugadores de Nicho	<ul style="list-style-type: none"> • Actuate • Panorama • Arcplan • Board International

Los líderes son IBM-Cognos, Oracle, Microsoft, SAS, MicroStrategy, Information Builders y SAP-Business Objects. Los usuarios esperan menos innovaciones al menos de los grandes proveedores, es más probable que estas lleguen de los vendedores de nicho quienes para este año integraron la presentación de reportes IN y las analíticas con los procesos de negocio, que es una característica que para los usuarios crece en importancia. También es posible un avance en las analíticas en memoria por parte de visionarios como Radnor o QlikTech. Se preveía que el mercado crezca un 8,6% anual.

Tabla 14. Interpretación Gartner BI 2013

CATEGORIA	SOFTWARE BI	
Líderes	<ul style="list-style-type: none"> • Microsoft • Tableau Software • QlikView • Oracle • IBM • SAS 	<ul style="list-style-type: none"> • MicroStrategy • Tibco Spotfire • Information Builders • SAP
Retadores	<ul style="list-style-type: none"> • LogIXML 	<ul style="list-style-type: none"> • Birst
Visionarios	<ul style="list-style-type: none"> • No se presentan 	
Jugadores de Nicho	<ul style="list-style-type: none"> • Prognoz • Bifam • Actuate • Board International • Panorama Software • Salient Management Company 	<ul style="list-style-type: none"> • Alteryx • Jaspersoft • Pentaho • Targit • Good Data • Arcplan

Dentro de este análisis que se encuentra en la Tabla 14, recalcamos que: SAP, SAS, Qlikview, IBM, MicroStrategy y Microsoft, constan en el cuadrante de líderes; en el cuadrante de Jugadores de Nicho se matienen principalmente Jaspersoft y Pentaho. “Recalcamos que Tableau y Tibco Spotfire, entran al cuadrante de líderes (eran retadores); y por último Good Data debutó en el cuadrante de jugadores de nicho” (Webmining Consultores, 2013).

Tabla 15. Interpretación Gartner BI 2016.

CATEGORIA	SOFTWARE BI
Líderes	<ul style="list-style-type: none"> • Microsoft • Tableau Software
Retadores	<ul style="list-style-type: none"> • No se presentan
Visionarios	<ul style="list-style-type: none"> • Alteryx • SAS • SAP • MicroStrategy • IBM
Jugadores de Nicho	<ul style="list-style-type: none"> • Birs • Domo • Good Data • Salesforce • Board International • Sisense
	<ul style="list-style-type: none"> • Qlik tech • Logi Analytics • Clear Story data • Pentaho • Tibco Software • Beyond Core • Information Builders • Pyramid Analytics • Yellowfin • Platfora • Datawatch

Desde la Tabla 15, se determina que : Tableau, QlikTech y Microsoft son los únicos tres que aparecen el cuadrante de líderes, moviéndose Microsoft un poco más a la derecha y Tableau bajando de su posición de outlier del año pasado. Como novedad importante es que SAP, SAS, IBM, MicroStrategy y Oracle salen del cuadrante de líderes y se sitúan más abajo en el de visionarios. Pentaho se mueve a la derecha desde jugador de nicho a visionario. Este año no hay ningún retador, todos aquellos retadores del año pasado, pasan a ser competidores de nicho. Se reconoce en este análisis que los nuevos que se agregan a este cuadrante son: Domo, Platfora, Salesforce y Sisense.(Webmining Consultores, 2016)

3.2.2. FORRESTER RESEARCH

De igual forma para fortalecer el criterio de avances en software tecnológico de BI, se toma en cuenta las Tablas 16, 17 y 18, en las cuales se pone en consideración lo analizado por otra consultora mundial importante como es Forrester Research Inc., la que revisa varios criterios importantes de evaluación para empresas de BI en este caso. (Evelson, 2010). Podemos indicar que en su análisis categoriza a las empresas como líderes,

empresas competitivamente fuertes, así también aquellas que quieren tener su mercado importante a las que se les conoce como contenedores, y por último aquellas que quieren ingresar al mercado de producto que se esté analizando. Dentro de estas 4 categorías de igual forma por medio de ejes realiza la siguiente consideración; a través del eje X va puntuando a las organizaciones según su fortaleza, su robustez como producto, luego las que son más estratégicas para su posicionamiento y definitivamente las más débiles. Pero también en su eje vertical o Y, van creciendo desde las más débiles comercialmente, luego aquellas que se encuentran ofertando en la actualidad hasta llegar a las más fuertes quienes ya tienen un posicionamiento en el mercado. Así explicado el análisis, ponemos en consideración 3 años el análisis de Forrester.

Tabla 16. Interpretación Forrester Research Inc. 2008.

CATEGORIA	SOFTWARE BI
Líderes	<ul style="list-style-type: none"> • SAP Business Object • IBM Cognos • SAS • Oracle
Retadores Fuertes	<ul style="list-style-type: none"> • Information Builders • Micro Strategy • Microsoft • Actuate • SAP (NetWeaver BI) • Tibco
Contendientes	<ul style="list-style-type: none"> • Panorama Software • QlikTech
Entrantes (Apostadores Ariesgados)	<ul style="list-style-type: none"> • No se presentan

En la evaluación de los 151 criterios del informe Forrester para las plataformas de Inteligencia de Negocio, IBM Cognos y SAP Business Objects mantienen sus posiciones de líderes del mercado, mientras que Oracle BI y SAS pasaron a entrar en el cuadrante de líderes gracias a la riqueza de sus funcionalidades, escalabilidad, y la completitud de su visión y estrategia corporativa. Otras empresas como: Actuate, Information Builders,

Microsoft, MicroStrategy, SAP, y un nuevo jugador, TIBCO Spotfire, se mantienen como alternativas en muchas elecciones para la gestión de la información y el conocimiento (I&KM). Otras empresas no menos reconocidas y destacadas como Panorama Software y Qlik Tech, en algunos casos concretos mejoran las prestaciones de los líderes.

Tabla 17. Interpretación Forrester Research Inc. 2012.

CATEGORIA	SOFTWARE BI
Líderes	<ul style="list-style-type: none"> • Informática • IBM • SAP • Oracle • Ab Initio • SAS
Retadores Fuertes	<ul style="list-style-type: none"> • Talend • Persuasive Software • May Software • Microsoft
Contendientes	<ul style="list-style-type: none"> • No se presentan
Entrantes (Apostadores Ariesgados)	<ul style="list-style-type: none"> • No se presentan

En los últimos años, el mercado de los servicios de inteligencia de negocio (IN) ha evolucionado de manera significativa lo que se presenta en la Tabla 17, deja ver lo siguiente: que las empresas de consultoría de asesoramiento y gestión han reforzado sus capacidades de ejecución, mientras que el desarrollo tradicional y las empresas de externalización han adquirido rápidamente el talento y han generado consultoría de gestión y prácticas de asesoramiento estratégico. En la evaluación de 53 criterios de Forrester de los proveedores de servicios de inteligencia de negocio, encontramos 12 líderes de élite que todos ofrecen opciones fuertes para los compradores de servicios de IN: Accenture, Capgemini, Cognizant, CSC, Deloitte, HP, Infosys, KPMG, IBM, PwC, TCS, y Wipro. IBM, Accenture, Deloitte, KPMG, CSC, Wipro, Infosys, Capgemini, HP, PwC, TCS, y

Reconociendo liderar el mercado de trabajo de la estrategia de IN, mientras que IBM, Accenture, CSC, Deloitte, Wipro, Infosys, HP, Capgemini, TCS, Consciente, KPMG, PwC y líderes en el mercado en materia tecnológica. Existen numerosas otras opciones en este mercado que es de alguna manera muy maduros sin embargo en otros aspectos todavía en evolución rápida de los integradores de sistemas globales para cientos de empresas de segundo nivel medianas a miles de boutiques. Además de revisar estos resultados de la evaluación, se recomienda la aplicación de nuestra metodología de evaluación (el modelo de Forrester Wave™) para la lista corta sus otros proveedores de servicios de IN en base a sus necesidades específicas y únicas.

Tabla 18. Interpretación Forrester Research Inc. 2015.

CATEGORÍA	SOFTWARE BI
Líderes	<i>Microsoft</i> <i>Qlik</i> <i>Tableau</i> <i>Oracle</i> <i>MicroStrategy</i> <i>SAP</i> <i>Good data</i> <i>Tibco Software</i> <i>SAS</i>
Retadores Fuertes	<i>Panorama</i> <i>Information Builders</i> <i>IBM</i> <i>Birst</i>
Contendientes	<i>No se presentan</i>
Entrantes (Apostadores Ariesgados)	<i>No se presentan</i>

En la evaluación de 60 criterios de Forrester de vendedores de Inteligencia de Negocios (IN), se identificaron los 11 proveedores de software más importantes - IBM , Information Builders, Microsoft, MicroStrategy , OpenText , Oracle , Qlik , SAP , SAS , Tableau Software, Software TIBCO, se han investigado , analizado y anotado sus ofertas actuales del mercado . Este informe detalla nuestras conclusiones acerca de lo bien que cada

proveedor cumpla con los criterios de evaluación de Forrester y cuál es su posición en relación el uno al otro para ayudar al desarrollo y distribución de aplicaciones profesionales y de seleccionar el proveedor adecuado para su plataforma de IN empresarial.

De los dos análisis referentes tanto Gartner Group como Forrester Research Inc, se concluye que las herramientas I.N. que podrían ser analizadas en este estudio son: Tableau, Qlick y Microsoft Power BI, por lo que en la Tabla 19, bajo algunos criterios vamos a revisarlos y sin ningún criterio apegado al mercado y más desde una perspectiva académica expresaremos la posibilidad de contar con una de ellas o simplemente gestionar la posibilidad de establecer convenios para uso exclusivamente académico de estas poderosas herramientas.



Universitat d'Alacant
Universidad de Alicante

Tabla 19. Cuadro Comparativo de Herramientas de IN (Propia, 2016)

	 + a b e a u		
Producto	Tableau Desktop Tableau Server Tableau Online	Qlik Sense Enterprise Qlik Sense Desktop Qlik Sense Cloud Qlik DataMarket QlikView Qlik Analytics Platform	Power BI Power BI Desktop Power BI Mobile Power BI Embedded
Costo	Tableau Desktop Personal \$999 Professional \$1,999	Qlik Sense Cloud Basic. Gratis Qlik Sense Cloud Plus \$20/Usuario/Mes	Power Bi Gratis Power Bi Pro \$9,99/Usuario/Mes
Volumen	Personal Edition Conexión a 6 fuentes de datos, incluidos archivos .csv y de Excel Un año de actualizaciones y soporte incluidos Professional Edition Conexión a cientos de fuentes de datos Compatible con Tableau Server Compatible con Tableau Online 1 año de actualizaciones y soporte incluidos	Qlik Sense Cloud Basic. Comparte 5 usuarios 25 MB Tamaño máximo aplicaciones Máxima de almacenamiento de 250 MB Acceso a Qlik DataMarket paquete gratuito Qlik Sense Cloud Plus Compartir ilimitado 50 MB máximo de aplicaciones 10 GB de almacenamiento Max Acceso a Qlik DataMarket Paquete esencial	Power Bi 1GB/usuario Power BI Pro 10 GB/usuario
Licencia Académica	SI	SI	NO
Soporte Técnico	SI	SI	SI
Servicios	Servicios de consultoría Portal del cliente Base de conocimientos	General Formación Soporte Consultoría	Aprendizaje Guiado Documentación Comunidad
ETL	No cuenta con este proceso	Si cuenta con ese proceso	Si cuenta con ese proceso
OLAP	No cuenta con esta tecnología	No cuenta con esta tecnología	Si cuenta con esta tecnología
Open Source	Software Propietario	Software Propietario	Software Propietario

De las comparaciones realizadas en la Tabla 19, se obtiene que Qlick y Tableau, se acoplan a nuevas tendencias tecnológicas como son el uso de motores indexados o en memoria RAM; lo que posibilitan mayor granularidad y evitan la dependencia de costosos Datawarehouse. En base a esto y según las condiciones que presentan actualmente, recomendamos a la Universidad Central del Ecuador la gestión en la firma de convenios para la utilización de herramientas de este nivel con lo que se conseguiría obtener lo siguiente:

- Los estudiantes de la Universidad Central del Ecuador se familiarizarán con una herramienta actual de BI; independiente de cuál usen en el ámbito profesional.
- Innovadora arquitectura que rompe con los sistemas de IN tradicionales.
- Modificar el análisis dependiendo las necesidades del usuario pudiendo realizar cálculos en tiempo real.
- Ayuda a entender relaciones asociativas disponiendo del conjunto total de los datos.
- Llevar a cabo análisis y presentarlos con rapidez de forma simultánea a usuarios conectados en diferentes localizaciones sin problemas de carga.
- Estas herramientas presentan licencias académicas que benefician en gran medida a docente y alumnos, cuenta con soporte técnico, además genera consultoría lo que permitirá a la universidad brindar este servicio a las empresas de nuestro país.
- Qlik o Tableau, son soluciones completas, que soportan las herramientas IN tradicionales en una misma arquitectura, y permiten consolidar los datos, búsquedas asociadas, ejecutar cálculos y presentar visualmente los resultados mediante cuadros de mando, análisis e informes.

Adicionalmente para complementar este análisis; en el mercado se está manejando el esquema de herramientas de IN, en las cuales se potencializa la rapidez con la que se ejecutan las consultas, se minimizan los tiempos en los procesos ETC y en general se disminuyen los tiempos en proyectos de esta clase. Por esta razón en la Tabla 20, se presenta un análisis sobre las alternativas de software de manejo tradicional bajo un esquema de almacén de datos (DW) y aquellos que muestran su poder en memoria como una de sus mayores ventajas.

Tabla 20. Cuadro Comparativo Soluciones DW vs. In Memory. (Propia, 2016)

SOLUCIONES TRADICIONALES (DATA WAREHOUSE)	SOLUCIONES ACTUALES (TECNOLOGÍAS IN MEMORY)
Requieren mayor tiempo de procesamiento	Lectura más rápida de los datos en memoria que en disco.
Son poco flexibles a la hora de incorporar nuevos requisitos.	Mejor rendimiento de consultas, hace que su modificación sobre la marcha y la navegación sean mucho más fluidas.
Mientras más grande es el volumen de datos mayor es la degradación del rendimiento de las consultas.	Se basa solo en cargar los datos necesarios en memoria.
Realizar un Data Warehouse representa un mayor costo y grado de complejidad	Elimina necesidad de optimizar la B. de datos (índices, agregados entre otros).
El Data Warehouse se alimenta periódicamente	Con los datos en memoria podremos acceder al instante a la información
Para implementar un DW debe basarse en si los beneficios que puede aportar a la organización compensan costes derivados.	Se simplifican las cargas, con lo que el tiempo de desarrollo se ve reducido y el mantenimiento es más sencillo.
El proceso ETC al implementar el DW por lo general toma alrededor del 80% del tiempo de la solución de BI.	Escalable incorporando más memoria al aumentar el volumen de datos.
El almacenamiento del DW se lo hace en discos mecánicos. Haciendo que el acceso tome más tiempo.	Utilizan algoritmos de compresión que reducen considerablemente el tamaño de los datos en la memoria
El Datawarehouse tiene un diseño complejo y multidisciplinario	Implementa los mismos niveles de seguridad (autenticación, encriptación, protocolos de comunicación entre otros).
Son Bases de datos Relacionales (Modelos: estrella o copo de Nieve)	Utilizan la lógica asociativa
Inversión en Hardware y Software (Potente)	La inversión en Hardware (Memorias) lo cual es accesible en cuanto a costos

3.3. CÍRCULOS DE PROCESOS.

Para lograr una mejora permanente es nuestro criterio aplicar a los procesos más importantes en la metodología M3S, las características de círculos de calidad, PHVA, cuyas siglas representan el Planificar, Hacer, Verificar y Actuar; las mismas describen los cuatro pasos esenciales que se deben llevar a cabo de forma sistemática para lograr la mejora continua, entendiendo como tal al introducir mayor calidad en los procesos (disminución de fallos, aumento de la eficacia y eficiencia, solución de problemas, previsión y eliminación de riesgos potenciales). El círculo de Deming lo componen 4 etapas, de forma que una vez acabada la final se debe volver a la primera y repetir el ciclo de nuevo, de forma que las actividades son reevaluadas periódicamente para incorporar nuevas mejoras. La aplicación de esta metodología está enfocada principalmente para ser usada en empresas y organizaciones.” (PDCA, 2013).

3.3.1. APLICACIÓN DE CÍRCULOS DE CALIDAD EN E.T.C.,

Con el propósito de establecer una mejora continua en el proceso E.T.C., se establece en la Figura 15, dentro de un círculo de calidad los pasos más relevantes que deberían ser considerados en las fases de diseño y construcción del proyecto de I.N.

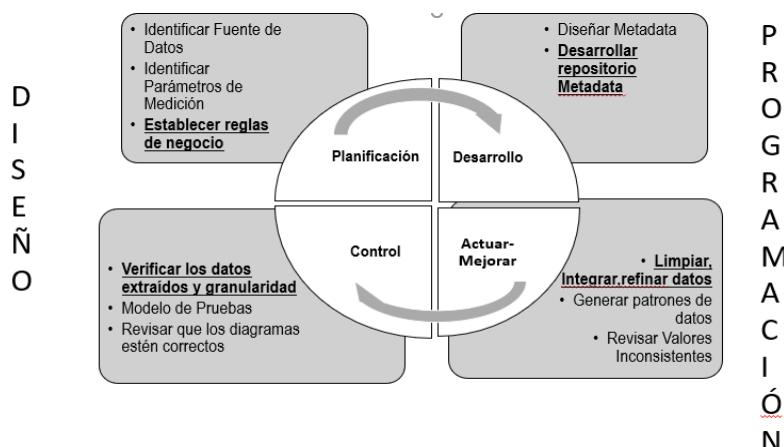


Figura 15. Círculos de Calidad en ETC.

En esta parte establecemos las fases de las etapas de diseño y construcción de proyectos de IN, dentro de un Proceso de Calidad, para tratar de que cada una de ellas sea realizada con éxito, y para lo cual explicamos el esquema de la Figura 15 a continuación:

PLANIFICACIÓN: Fijar objetivos, procesos para obtener resultados acorde a los requisitos de la organización; identificar y definir cómo se extraerá las fuentes de datos; establecer métricas según el negocio; llegar a conocer políticas, normas, restricciones en los objetivos de la organización.

DESARROLLAR: Implementar procesos; usar herramientas para diseño de metadata; desarrollar repositorio de integración de datos para mejor manipulación y acceso.

CONTROLAR: Verificar y validar que se cumplan los objetivos, con lo que se puede diseñar un modelo de pruebas, verificar los datos extraídos, comprobar la granularidad establecida y la respectiva funcionalidad de los diagramas de procesos críticos.

ACTUAR Y MEJORAR: Se requiere que el cliente acepte el producto, de igual forma se puede continuar con limpiezas de datos, verificar la correcta integración y concordancia de la información. También se puede aplicar modelos conceptuales para mejoramiento de nuevos requerimientos y generar patrones de datos dependiendo del objetivo establecido.

3.4. TEORÍA DE LA DECISIÓN.

La vida en general es un permanente afrontar de problemas, lo que consideramos la situación actual y un objetivo. Un objetivo es un resultado, atributo o situación deseados, para cuya consecución se pretende ejecutar alguna acción. “Todo problema plantea alternativas de cursos de acción a fin de lograr el objetivo o al menos acercarse a él, por lo

que es imposible resolver un problema sin tomar una decisión. En sustancia todo el proceso decisorio puede enfocarse como un sistema cuya entrada es la información y su salida es la decisión orientada a su acción; y siempre deberá tomarse en cuenta como en cualquier sistema de información, que la calidad que obtengamos como salida, dependerá de la calidad de su entrada” (Lazzati, 2013).

En el ámbito del fundamento teórico en el cual de desenvuelven las decisiones revisaremos algunas acepciones importantes como el *Aprendizaje automático*, el mismo que “Se refiere a un amplio espectro de situaciones en las cuales el aprendiz incrementa su conocimiento o sus habilidades para cumplir una tarea. El aprendizaje aplica inferencias a determinada información para construir una representación apropiada de algún aspecto relevante de la realidad o de algún proceso.” (Moreno A. , 1994), esto significa desarrollar técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear programas capaces de generalizar comportamientos a partir de una información suministrada (inducción del conocimiento).

El aprendizaje automático tiene una amplia gama de aplicaciones, incluyendo motores de búsqueda, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito, análisis del mercado de valores, clasificación de secuencias de ADN, reconocimiento del habla y del lenguaje escrito, juegos y robótica. Los diferentes algoritmos de aprendizaje automático se agrupan en función de la salida de los mismos, y uno de ellos es el *aprendizaje supervisado*, del cual se dice que es el algoritmo que produce una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema. Un ejemplo de este tipo de algoritmo es el problema de clasificación, donde el sistema de aprendizaje trata de etiquetar (clasificar) una serie de vectores utilizando una entre varias categorías (clases).

Desde la perspectiva de las redes neuronales artificiales es su capacidad de aprender a partir de un conjunto de patrones de entrenamientos, es decir, es capaz de encontrar un

modelo que ajuste los datos. El proceso de aprendizaje también conocido como entrenamiento de la red puede ser supervisado o no supervisado.

De igual forma por otro autor Aprendizaje Supervisado: “Consiste en entrenar la red a partir de un conjunto de datos o patrones de entrenamiento compuesto por patrones de entrada y salida. El objetivo del algoritmo de aprendizaje es ajustar los pesos de la red de manera tal que la salida generada por la Red Neuronal Artificial (ANN) sea lo más cercanamente posible a la verdadera salida dada una cierta entrada. Es decir, “la red neuronal trata de encontrar un modelo al proceso desconocido que generó la salida y este aprendizaje se llama supervisado, pues se conoce el patrón de salida el cual hace el papel de supervisor de la red”. (Salas, 2004)

Según Rodrigo Salas en su publicación de Redes Neuronales Artificiales, Aprendizaje NO Supervisado representa sólo un conjunto de patrones a la ANN, y “el objetivo del algoritmo de aprendizaje es ajustar los pesos de la red de manera tal que la misma encuentre alguna estructura o configuración presente en los datos” (Salas, 2004).

Uno de los enfoques que tiene el aprendizaje automático son los árboles de decisión como modelo predictivo, se mapean observaciones sobre un objeto con conclusiones sobre el valor final de dicho objeto.

Dos formas principales de árboles de decisión es la desarrollada por Quinlan al medir la entropía en cada rama y a su vez podemos decir que son más adecuados cuando:

- Las instancias del concepto son representadas por pares atributo-valor
- La función objetivo tiene valores de salida discretos
- Las descripciones del objeto para analizar son disyuntivas

En diferente literatura se conoce como “cantidad de información mutua” a la Ganancia de información(Quinlan, 1986). En otras definiciones tenemos: “Decisión Tree Learning”, es un método para aproximar funciones objetivo de valores discretos, en el que la función

aprendida está representada por un árbol de decisión. El aprendizaje en el árbol de decisiones es uno de los métodos más prácticos para la inferencia inductiva." (Mitchell, 1997). "Los árboles de decisión, son muy útiles para el planteamiento de problemas secuenciales, pero esta clase de situaciones implica decisiones con resultados hacia el futuro que, en términos de comportamiento del decisor, no se ha definido con claridad cómo manejarlos" (Kilmer & Salinas); y una más de ellas indica que "Un árbol de decisión, es un modelo de predicción cuyo objetivo principal es el aprendizaje inductivo a partir de observaciones y construcciones lógicas. Son muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva para la solución de un problema". (Martínez, y otros, 2009)

Existen acepciones importantes sobre esta temática como decir que: "*El aprendizaje de árboles de decisión*, es uno de los métodos de inferencia inductiva más utilizado en aplicaciones de diagnóstico médico, predicción de fraudes en créditos bancarios, organización personalizada de eventos en una agenda, entre muchos. Esto es debido a que tiene la ventaja de aprender funciones de valores discretos, es robusto al ruido que existe en los datos, y se puede obtener un conjunto de expresiones disyuntivas que es transformable a un conjunto de reglas". (García, 2007), similar criterio presenta (Sempere, 2014).

De igual forma es bueno indicar que un *Modelo de Clasificación ID3*, significa "inducción mediante árboles de decisión" y se define como un sistema de aprendizaje supervisado que aplica la estrategia "divide y vencerás" para hacer la clasificación, implementando métodos y técnicas para la realización de procesos inteligentes, representando el conocimiento y el aprendizaje, con el propósito de automatizar tareas. Podemos decir que su filosofía es:

- Determinar el árbol de decisión mínimo, con la menor entropía para todas las clases.

- El árbol debe estar organizado y entendible para cualquier persona, seleccionando en cada nodo el mejor atributo.
- Realiza una búsqueda de Arriba hacia Abajo, por medio de preguntas
- Está basado en criterio estadístico
- Selecciona los atributos en base a Ganancia de información.

Dentro de los árboles de decisión y el manejo de su estructuración se llegan a encontrar conceptos importantes como la *Entropía*, que se considera a la incertidumbre, impureza, desorden de un conjunto de datos relativo a una clasificación binaria y que matemáticamente se encuentra representada de la siguiente manera:

$$Entropia(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Donde:

- S : es una colección de objetos
- P_i : es la probabilidad de los posibles valores
- i : las posibles respuestas de los objetos
- Si todos los datos pertenecen a una sola categoría la Entropía es cero ($0 * \text{Log}(0)= 0$)
- Si los datos están mezclados (positivos = negativos = 0.5), la Entropía tendrá un valor máximo de 1

De igual forma se especifica la *Ganancia de Información*, cuando se determina el menor valor de entropía esperado (S), al ordenar todos los nodos (S) basándose en cada atributo.

$$Ganancia(S, A) \equiv Entropia(S) - \sum_{v \in Valores(A)} \frac{|S_v|}{|S|} Entropia(S_v)$$

Donde:

S : es una colección de objetos

A : son los atributos de los objetos

$V(A)$: Conjunto de valores que A puede tomar.

3.5 PROBLEMAS.

En el tratamiento de proyectos de BI, generalmente se encuentran algunos problemas que vienen por la falta de conocimiento de los temas que afectan la organización, es por esto que se debe iniciar buscando los inconvenientes que tiene una empresa en su gestión, analizando las herramientas automatizadas que ayudarán a profundizar el nivel de conocimiento de la información para maximizar el rendimiento del negocio y también producir mayor conocimiento en la toma de decisiones. Entre otros problemas se identifican la pérdida de tiempo tanto para generar consultas de gran tamaño, como para la entrega de respuestas, “carencia de una plataforma en la que se pudiera interactuar con la información histórica y también la falta de herramientas que otorgarán reportes en lapsos breves de tiempo” (Fuentes & Ricardo, 2010).

3.5.1 SITUACIÓN ACTUAL DE LA INTELIGENCIA DE NEGOCIOS.

La inteligencia de negocios ayuda a que una empresa sea más intuitiva y competitiva, permitiendo que las personas que están involucradas en la toma de decisiones puedan analizar la información, sin tener que pasar por el arduo trabajo de recopilarla, y ver en qué procesos tienen que realizar cambios o a qué problemáticas tienen que adelantarse.

Como se muestra a lo largo de todo este estudio, IN como herramienta implementada ha generado respuestas a innumerables problemas en diferentes áreas, así pues como los sistemas de información pueden dar solución a cualquier actividad del ser humano, la inteligencia de negocios ayuda a la toma de decisiones en medida que obtenga mayor conocimiento de la información transaccional generada día a día, y así lo vemos en áreas de comercialización, ventas, clientes, análisis financiero, talento humano, inventarios, producción, minería de datos, medicina, aviación entre tantos otros, cubriendo en cada uno de ellos problemas como: rápida detección de desviaciones de presupuesto y stock, identificando las tendencias en las ventas de productos, ahorro de cientos de horas/hombre

en generación y análisis de informes, medición de desempeño, estrategias empresariales para el mejoramiento de la industria farmacéutica, aprovechamiento de los activos de información como ventaja competitiva, aprovechamiento de las oportunidades de negocio al gestionar información completa y actualizada, instrumentos de control orientados a la competencia, para dar prioridades en el mercado energético, disponibilidad de Cuadros comparativos de rendimiento de empleados, asociados y colaboradores.

El reto ya no es solamente el almacenar datos, sino el de almacenarlos de forma tan sistematizada que permita un fácil y rápido acceso a los mismos para la realización de consultas, y mayor precisión para la ejecución de estrategias.

3.5.2. PROBLEMAS NO RESUELTOS Y LIMITACIONES.

La mayor parte de situaciones donde IN no puede ser implementado o definitivamente nos encontramos con problemas y limitaciones para avanzar, se debe a que las organizaciones no disponen de una visión estratégica, esquemas claros de las necesidades y objetivos del negocio que son fundamentales, es decir no precisan las actividades primarias de la cadena de valor de la organización. Otra de las situaciones que nos llevan a estas instancias se presenta cuando no existe un compromiso gerencial y el patrocinio adecuado del proyecto, ya que no todas las organizaciones le dan el verdadero valor que tiene la información y resulta que el almacenamiento de la misma tampoco es de mucho interés, entendemos por su falta de conocimiento y a la final esto les repercute en una toma de decisiones no muy exitosa. Así también en las esferas ejecutivas en algunos casos de las pequeñas y medianas empresas les parece costoso iniciar proyectos de IN, perdiendo la oportunidad de posteriormente gozar de los beneficios que brindan este tipo de proyectos.

3.5.3 ¿DÓNDE AVANZAR?

Las herramientas informáticas desarrolladas en los últimos años en el terreno de la gestión de datos e información han sufrido una serie de cambios, que responden a las demandas que los usuarios de este tipo de aplicaciones producen durante la experiencia al usar dichas herramientas.

La Inteligencia de Negocios, en los últimos tiempos ha vivido las modificaciones e innovaciones de las nuevas tecnologías de software que constantemente llegan al mercado, ofreciendo un servicio superior y una mayor funcionalidad, siendo de lo más relevante la mejora que se tiene en la capacidad de tomar decisiones precisas y de forma rápida, convirtiéndose en una de las claves para que una organización mejore.

Existen muchos programas en los que su diseño se adapta con mayor o menor medida para manejar los datos de la empresa y entonces para superar todas estas limitaciones surge la IN apoyándose en un conjunto de herramientas tecnológicas que facilitan la extracción, la depuración, el análisis y el almacenamiento de los datos generados en una organización, para su explotación directa (reportes, análisis OLAP, minería de datos, etc.) o para su análisis y conversión en conocimiento como soporte a la toma de decisiones sobre el negocio. Debido a las virtudes de esta tecnología, los expertos aseguran que en un futuro a corto plazo se estima que la utilización de IN aumentará en un 40%, ya que las herramientas que brindan esta solución empresarial permitirán conocer las preferencias de los consumidores, incluso a nivel individual. Se prevé que en los próximos años las organizaciones decidan abandonar los métodos tradicionales de manejo de información, para utilizar sistemas IN, con el fin de contar con una herramienta administrativa que ofrezca los elementos necesarios para establecer una estrategia adecuada que responda a los cambios del mercado en tiempo real. Las funcionalidades que ofrecerá IN en el futuro son ilimitadas y después de haber revisado los aspectos vinculantes a la Inteligencia de Negocios, destacamos la posibilidad de avanzar en los siguientes temas:

1. La Integración de Datos Estructurados y No estructurados: La mayoría de los proyectos de Inteligencia de Negocios se centraban en analizar datos estructurados, casi todos ellos almacenados en bases de datos de cualquier tipo; pero gracias a las nuevas herramientas, ahora también son capaces de trabajar con datos desestructurados; aquellos que provienen de cualquier otra fuente que no sea estructurada. Es por eso que M3S, dentro de su proceso de diseño y construcción del proceso ETL, buscará proyectar una solución para centralizar una gran información que pueda venir desde otro tipo de fuentes como páginas web, interacciones en redes sociales que cada vez son más abundantes y estratégicas; y con esto generar una gran base de hechos concretos que ayuden a mejorar las decisiones.
2. La Difusión de IN a otros Niveles Organizacionales: Enfocar las nuevas herramientas metodológicas como M3S a mandos medios como directores financieros, responsables comerciales y demás empleados que puedan apoyar a este tipo de proyectos, lo que en un futuro casi inmediato les permitirá ser capaces de usarlas, mejorando su eficacia en el trabajo y transformándolos en personal más analítico que operativo.
3. El Almacenamiento: Que no debe ser visto más que como una forma de asegurar la información de acuerdo a las políticas de seguridad de cada empresa y esto a su vez como el procesamiento y almacenamiento masivo de datos en servidores que alojen la información del usuario y que tengan la capacidad de ser accesibles de forma instantánea y desde cualquier lugar del mundo y a través tanto de dispositivos móviles (teléfonos inteligentes, tabletas), como de computadoras de escritorio o notebooks, usando canales de transmisión de datos o el mismo internet. Pero aún más importante es que según sea la forma de almacenamiento, esta tendrá o no riesgos de estar disponible o visible ante otras instancias que no sean las de la propia empresa. Por esta razón M3S, es independiente a la forma de seguridad de almacenamiento con el que cuenten las empresas, sin embargo, dejará establecido en forma documentada la estructura de organización de todo tipo de archivo y de aquellos que se generan en base a requerimientos planteados dentro del proyecto para su posterior explotación si fuere del caso.
4. Los Datos Abiertos: Existen tendencias sobre todo desde los organismos gubernamentales, para optar por la reutilización de datos, mediante el desarrollo de aplicaciones con información de las compañías. Se conoce que los movimientos a favor de la transparencia se encuentran impulsando políticas

de fomentar los datos abiertos, pues al estar disponibles se pueden realizar una serie de estudios y análisis de los intereses de los ciudadanos aplicando técnicas y metodologías de IN, por ejemplo se podría medir el grado de interés en la ciudadanía de acuerdo a la intervención de un político dentro de su entorno de acción, la sensibilidad de actuación del gobierno y cualquier análisis de tendencias, entre muchos otros. Esta parte dentro de M3S, no es más que la capacidad que se tiene de llegar a mayor granularidad de información ya sea que se encuentre en fuentes de datos estructuradas totalmente, en forma parcial o definitivamente en aquellas totalmente dispersas o no estructuradas.

5. La Movilidad: Ante el creciente manejo de dispositivos móviles por los usuarios, las empresas se encuentran apostando a herramientas de BI, que pueden usarse desde teléfonos, tablet y otros. “Gartner estimó en un estudio que se dará un 33% de crecimiento en el uso de herramientas de IN en dispositivos móviles para finales del 2013” (Orozco & Soriano, 2016), ya que permiten un sencillo acceso de los usuarios a los datos; y esto se ha dado con mayores escalas hasta la actualidad. La tendencia de nuevas metodologías de BI, deben orientarse a tener bases de conocimientos concentradas que permitan la generación de más conocimiento por su mejor estructuración, para que de esta forma sea más eficiente el tiempo de respuesta, en el momento de consultas por cualquier medio o como en este punto se explica con aparatos móviles.

Universitat d'Alacant
Universidad de Alicante

4. CAPÍTULO IV. METODOLOGÍA M3S.

4.1. MODELO M3S.

Habiendo revisado los avances en el área de herramientas de IN, el buscar un nuevo modelo que nos guie a conseguir soluciones en el ámbito de inteligencia de negocios ya sea considerando soluciones tradicionales o aquellas de las que actualmente se las conoce como In Memory; queremos proponer una metodología de implementación de proyectos IN, explotando y mejorando principalmente el proceso extracción y transformación de datos, pero introduciendo aprendizaje automático a través del uso de algoritmos de árboles de decisión que dentro del estudio se les dará mayores parámetros de valor por lo que les conoceremos como ideales. Esto ayudará a tener soluciones que beneficien la creciente necesidad que tienen los ejecutivos y en general los usuarios de diferentes niveles jerárquicos en las empresas de contar con información tomada de diferentes aristas pero que lleguen limpia, que les permita mejorar sus decisiones en base al mayor conocimiento que se llega a tener de las organizaciones. Tomando en cuenta que en la mayoría de casos la información se encuentra dispersa en varios sistemas y en diferentes fuentes de datos,

esto implica que debe crearse una metodología que logre consolidar la información en base a lo aprendido y en el presente y futuro pueda colaborar con la toma de decisiones certeras en las empresas.

4.1.1. PRINCIPIOS Y REQUISITOS.

La base de la metodología es organizar la información de las áreas críticas de las empresas en una serie de paquetes de trabajo estandarizados para la planificación, luego para la ejecución y finalmente en el control de los proyectos. Al seguir la metodología, el equipo va produciendo una serie de entregables perfectamente definidos, evidenciables y lo menos burocráticos posible, lo que permite cumplir con el alcance y los objetivos de tiempo, costo y calidad establecidos para cada proyecto.

Para cumplir esta misión en la Figura 16., se consideran cuatro factores críticos de éxito dentro de una organización sobre los cuales debería soportarse siempre un proyecto de tecnología y de igual forma lo hará M3S; siendo los siguientes:

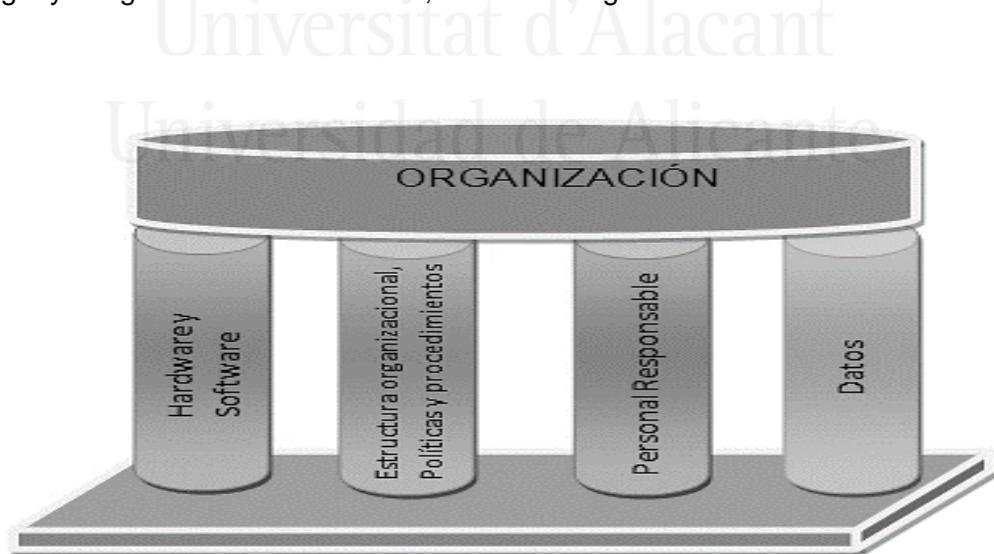


Figura 16. Pilares de M3S.

Con este enfoque se dará respuesta como en algunas metodologías analizadas a las preguntas que se dan en los proyectos:

- ¿Qué criterios se consideran para el éxito?
- ¿Cuál es el cronograma de ejecución?
- ¿Existe aceptación del proyecto por parte de todos en la empresa?
- ¿Se asignarán los recursos humanos, internos y externos necesarios?
- Definir ¿Qué?, ¿Cuándo? y ¿Cómo debería hacerse?

4.1.2. PRESENTACIÓN DE M3S

La metodología de implementación para procesos de IN, cuenta con las técnicas necesarias para el seguimiento de proyectos, así como con una estructura organizacional de los mismos, con lo que se conseguirá una buena rigurosidad en las etapas, de igual forma tiene un control adecuado en todos los procesos y actividades que se vayan ejecutando. Consideramos cinco técnicas y las mismas son aplicables a cualquier proyecto: La primera es la Organización que consiste en escoger personas estratégicas y estructurarlas dentro del esquema del proyecto para su mejor funcionalidad; la segunda es la Ejecución, en donde se establecen las fases, pasos y actividades requeridas para la consecución de los objetivos en cada una de ellas; la tercera es la Dirección, en donde se proveen de herramientas técnicas estructuradas y guías con base de información a los integrantes del equipo; la cuarta el Planeamiento y Control, para lo que se tienen técnicas para programar actividades esenciales para el éxito, asignar recursos para resolver sobrecargas de trabajo, monitorear el avance del proyecto, resolviendo conflictos oportunamente, y concluyendo el proyecto en tiempo y presupuesto previstos. Por último la quinta es la Administración y Seguimiento, en donde se cuenta con hojas de actividades por consultor, reunión de líderes, evaluación de entregables, hitos de control consiguiendo con esto una desburocratización de la documentación en proyectos; el mantener una

estructura organizacional permite analizar el cumplimiento de objetivos y a su vez facilita la gestión de cambio, seguimiento y el debido control en la implantación de la metodología.

Establecemos las siguientes estructuras:

COMITE: Cuya función es la de establecer políticas, facilitar la gestión del cambio, convocar reuniones periódicas para revisar estatus, avance; y está conformada por los Líderes de proyecto del Cliente (responsable de las aprobaciones de definiciones, objetivos y entregables), como del Grupo Consultor; ejecutivos de alto nivel impactados por la implantación del proyecto.

EQUIPO DE PROYECTO: Tiene funciones relacionadas con la ejecución del cronograma, configuración del software, seguimiento y cumplimiento de la metodología de implementación y se estructura con usuarios responsables que tienen conocimiento cabal de los procesos y cuenten con poder de decisión.

CLIENTE: Se debe solicitar un líder o coordinador del proyecto, usuarios claves de las áreas operativas y técnicos de sistemas de información

EQUIPO CONSULTOR: Debe liderar un especialista en implementación de proyectos y consultores especializados, en este caso en M3S.

EQUIPO DE TRABAJO: Son pequeños comités de carácter temporal según el avance del proyecto, se forma por uno o más usuarios claves operativos (pertenecientes al equipo de proyecto); cualquier otro usuario de la organización que tenga conocimiento específico y necesario para la resolución de algún problema presentado en el transcurso del proyecto.

4.2. PROPUESTA DE LA METODOLOGÍA M3S.

En esta sección se propone la metodología que se va a implementar y la misma será documentada en cada una de sus fases con los elementos considerados más apropiados.

En sus etapas de diseño y construcción, dentro de M3S, se encuentra implícito un proceso de árboles de decisión a los que se les llamará “ideales” básicamente por el especial tratamiento a los atributos relevantes que serán considerados en la estructura de decisión y que luego serán los nodos de los que se compone el árbol. Estos nodos que tienen un origen categorizado de dos formas y en cualquier caso con un mayor énfasis y cuidado en la calidad de información que contienen; esto significa un aporte fundamental en la investigación; en este sentido se busca la opción de generar árboles cuya estructura nos permita testear nuevos datos y procesarlos a través de éste en el cual ya se encuentran valorados atributos con las características ideales indicadas y para su complemento se contará con los mejores aspectos para decisión sobre hipótesis propuestas. El criterio principal es mejorar los resultados que se obtendrán con la experiencia de información que contamos y que ayuda a inferir suposiciones en hechos concretos (aprendizaje), todos esto permitirá en lo futuro realizar también un tratamiento de predicciones lo que supondría otra línea a investigar y proponer.

Para conseguir nuestro objetivo utilizaremos el tipo de aprendizaje supervisado proponiendo algunas premisas que deben cumplirse:

- Fijamos atributos relevantes para resolver la hipótesis de dos formas: la primera aquellos datos estructurados que debieron ser validados como los que tienen información de calidad con un alto grado de pureza y que aportarán definitivamente en el estudio; la segunda el tipo de información no estructurada en donde realizamos un análisis de sentimientos.
- Luego con esto tendríamos un subconjunto de datos para los cuales buscamos resultados objetivos, basándonos en experiencias reales, concretas; y
- Por último, queremos una o varias reglas bien definidas y estructuradas que puedan predecir estos resultados.

4.2.1. FASES DE M3S.

Esta metodología tras la revisión que se ha realizado en diferentes experiencias de estudios y planteamientos de varios tipos de proyectos de tecnología; sugerimos como se vio anteriormente en la Figura 3, y en esta sección en la Tabla 21, que el Ciclo de Vida que se llevará en M3S está compuesto de seis etapas que se encuentran definidas de manera literal de la letra A. hasta la F. y dentro de la cual será en la etapa D, en la que haremos el énfasis más importante y en la que se establece una nueva metodología dentro de M3S, que la llamaremos de “Metodología de Decisión”, en donde se aplican como se indicó algoritmos de árboles de decisión conocidos en el estudio como “ideales” y serán explicados en las siguientes partes.

Tabla 21: Ciclo de Vida de Metodología M3S

ETAPAS M3S (INTELIGENCIA DE NEGOCIOS)

A. Justificación	A.1. Evaluación caso de negocio A.2. Acta de Constitución del Proyecto.
B. Planificación	B.1. Evaluación Infraestructura <ul style="list-style-type: none">▪ Técnica▪ No Técnica B.2. Definición del Alcance B.3. Cronograma
C. Análisis y Requisitos de Negocio	C.1. Definición de requisitos de Proyecto C.2. Prototipo de Aplicación
D. Diseño	<i>D.1. Diseño ETC</i> D.1.1. Diseño Metodología de Decisión
E. Construcción	<i>E.1. Desarrollo ETC</i> E.1.1. Obtención de Metadata
F. Implementación	F.1. Puesta en marcha. F.2. Evaluación de la Solución F.3. Cierre del Proyecto

Exponemos la primera etapa (A) de Justificación, en la cual se buscan los argumentos reales que generen la necesidad del negocio para implementar un proyecto de ingeniería dentro del cual tenemos dos sub etapas: La Evaluación del caso de negocio en donde principalmente se debe justificar la inversión mostrando el balance entre el costo, tiempo

de ejecución y los beneficios conseguidos; y la segunda el Acta de Constitución, que es un documento inicial que formaliza la ejecución del proyecto, la misma que se encuentra en el ANEXO I.

La siguiente etapa (B) es la de Planificación, que define la estrategia y planes tácticos a seguir para desarrollar y llevar a cabo el proyecto con éxito, en la que de igual forma se dan tres sub etapas que son: La Evaluación de Infraestructura, que puede verse como una parte técnica, en el sentido que se debe elegir exhaustivamente una plataforma para la aplicación, que garantice la mejor respuesta para la recuperación y acceso a los datos (Hardware, middleware y SGBD) y por otro lado la No técnica, la cual se basa en políticas institucionales, personal que se asigna al proyecto y logística de oficina. Para esta parte se puede usar el ANEXO II. Como segunda sub etapa tenemos la Definición del Alcance, en donde se establece de forma concreta qué se persigue con el proyecto, incluyendo criterios específicos, medibles, realistas y enmarcados en un tiempo establecido, llegando a definirse claramente el objetivo del análisis de IN; y como tercera sub etapa tenemos la fijación del Cronograma, en donde se define cada actividad, designando los responsables incluyendo fechas de hitos importantes a cumplirse y a controlarse. Se debe revisar el ANEXO III, que ayuda a documentar esta parte que será la que guía el tiempo en el cual debe desarrollarse el proyecto.

Continuando con la etapa (C) es el Análisis y Requisitos de Negocio, en la cual se realiza un análisis detallado del problema y oportunidades del negocio, para conseguir una comprensión sólida de los requisitos de la organización que sean necesarios para llegar a una posible solución o producto, consta de dos partes que son: La Definición de Requisitos de Proyecto, que nos llevan a entender claramente la naturaleza del negocio que responden a las preguntas: ¿Qué?, ¿Con quién? y ¿Para qué?; y como segunda parte proponer un prototipo de lo que llegaría a presentarse en los resultados de tableros de medición, para lo cual se puede documentar utilizando el ANEXO IV.

En la cuarta etapa (D) que es el Diseño, concibe una alternativa que resuelva el problema para la toma de decisiones, dentro de la cual tenemos el Diseño ETC en donde los datos de origen para las aplicaciones de IN provienen de varias plataformas heterogéneas, que son gestionadas por una variedad de sistemas operacionales y aplicaciones. Partiendo de la premisa de que los sistemas deben aprender y adaptarse a su entorno, el propósito de los procesos ETC en esta investigación es unir los datos de estas plataformas y transformarlos a un formato que contenga todas las clasificaciones de información necesarias, de tal forma que a partir de estos se tenga capacidad de decidir sobre todo el universo de información actual y la futura que llegue para un determinado objetivo. Por ser una parte esencial de la investigación esta parte tendrá un tratamiento especial luego de la explicación de las dos últimas etapas de M3S.

En M3S la etapa (E) en M3S, es la Construcción, en la cual se aplica la metodología de decisión en un software que realice todo lo diseñado en la etapa anterior (D), dentro de la que se ingresará en un proceso de software las bases necesarias, las reglas definidas a fin de que la tabla de datos pueda ser analizada y poder inferir el árbol de decisión ideal. Es importante para navegar más eficientemente a través del entorno de IN, que la gente de negocio tenga acceso al repositorio de datos o tabla de entrenamiento.

La última etapa (F) es la de Implementación, conocida también como venta del producto acabado, y entonces medir su efectividad para determinar si la solución encontrada, excede o no el resultado esperado; para cumplir con esto se definieron tres sub etapas que son: La Puesta en marcha, pues una vez que la aplicación está construida y probada está lista para ser implementada en el entorno de producción. Esto se puede hacer de forma tradicional, o incremental y según el mejor criterio de los expertos de BI, para lo que puede ayudarnos el ANEXO VII, en donde se refleja cierta información técnica y funcional del proyecto; la siguiente sub etapa es la Evaluación de la Solución, en donde decimos que la construcción de un sistema IN nunca acaba, los objetivos y necesidades cambian, por esta

razón el proyecto seguirá evolucionando. Muy difícil será que la primera solución sea la mejor, normalmente hacen falta varias iteraciones para encontrar la más adecuada. Y como tercera sub etapa tenemos al Cierre del Proyecto, en donde se debe archivar toda la documentación generada en el proyecto para que se encuentre disponible a cualquiera de los niveles empresariales, siendo una buena práctica convocar a una reunión de cierre, formalizando la entrega de todos los aplicativos y aceptación del cliente, se elaborará una lista de pendientes por resolver con responsables y fechas de entrega; y si se considera con los costos respectivos aprobados tanto por los consultores cuanto por la organización, ver ANEXO VIII. En todas las etapas de la metodología M3S puede utilizar como documento de constancia para asistencia o tareas relevantes consideradas por los consultores o representantes del cliente el ANEXO IX.

Como se indicó anteriormente ahora vamos a determinar exactamente en la etapa de Diseño de M3S, la parte metodológica de la misma y la ponemos a consideración en cada una de sus fases.

D.1. Diseño Metodología de Decisión.

Antes de presentar la metodología de decisión, debemos inicialmente realizar ciertas consideraciones como: la extracción es independiente del origen de datos que tengamos, los mismos que pueden ser cualquier motor de base de datos relacionales, un Data Warehouse, archivos planos, hojas Excel, redes sociales entre muchos otros; nuestro centro metodológico se basará en los árboles de decisión ideales, para los que previamente debemos establecer reglas de negocio claras, las mismas que nos lleven a generar una tabla temporal en donde van a ser ubicados los mejores atributos en su categorización establecida y en el número de ellos que sean necesarios, los cuales agregarán mayor y mejor información para la toma de decisiones y en muchos casos para predicciones futuras

Es por esto que al usar M3S, se debe tener claro que IN, lo podemos definir matemáticamente como: $IN = IN + APRENDIZAJE AUTOMÁTICO$

4.2.2 ETAPAS DE LA METODOLOGÍA DE DECISIÓN

Antes de centrarnos en las etapas, revisemos algunas conceptualizaciones importantes que son utilizadas en esta parte y una de ellas es la definición de “datos puros”, que son aquellos que constan en forma definitiva en la tabla de entrenamiento, y quienes pasaron por un proceso de limpieza de datos que incluyó la sustitución de valores nulos o vacíos, reemplazo de cadenas no válidas, homogeneización de datos producto de fuentes heterogéneas. Estos problemas se originan por diferentes razones:

- Errores humanos en digitación. (Entrada de datos)
- Algunos atributos no tenían la validación obligatoria, generando campos nulos.
- Información parcial o totalmente duplicada en uno, grupo o todos los registros.
- Los datos al ser importados pudieron no almacenarse correctamente por diversas razones. (tipos de datos diferentes, ausencia de llaves, inconsistencia de datos).
- Datos traídos de las redes sociales, con motores especiales.

La búsqueda de la llamada pureza de datos es importante por los grandes volúmenes de información que se manejan en la actualidad y los costos que representan los proyectos de Tecnología de Información y que se quiere minimizar en los orientados a IN, y una de las formas es tener la certeza de que esa información que va a ser explotada tenga un porcentaje de confiabilidad muy cercano al 100 %. De nada nos serviría tener una gran tabla de entrenamiento en una data con tuplas que cuando tratamos de consultar algo de esa información resulta ser que la misma no es tan confiable como pensamos.

Los problemas que se generan en las empresas es que pierden diariamente cientos de miles de dólares como por ejemplo por realizar llamadas telefónicas incorrectas desde sus centros de información, mantener datos erróneos, enviar correspondencia a direcciones equivocadas u ofrecer un servicio a un cliente no calificado para un mercado específico; esto produce un aumento de los costos operativos del 40%, hasta un considerable daño de la imagen corporativa. ("Conseguir un nuevo cliente cuesta seis veces más que mantener uno que ya tenemos").

Citamos un par de definiciones de Limpieza de Datos como : "proceso de asegurar que todos los valores en un conjunto de datos sean consistentes y correctamente registrados", (Urdaneta, 2001); y otra autora que señala que el problema consiste en: "remover inconsistencias y errores desde los conjuntos originales de datos, es bien conocida en el área de Sistemas de Soporte de Decisiones y DW" (Galhardas, Florescu, D, & Simón, 2000).

Queremos llegar a que los datos con niveles de pureza que maneja M3S sean consistentes y con alta credibilidad; por lo que proponemos el siguiente flujo de limpieza:

1. Validación de valores nulos o vacíos
2. Tratamiento de registros duplicados.
3. Filtrar los datos semejantes.
4. Homogeneizar los datos de bases heterogéneas que vinieron diferentes orígenes estructurados o definitivamente no estructurado como redes sociales).

Para cumplir con estos pasos se procedió realizando consultas a las bases de datos de la siguiente forma:

- Elegir una tabla.
- Buscar campos vacíos y/o nulos y el tipo de datos de estos.
- Una vez mostrado el tipo de dato, optar por una sugerencia de reemplazo del campo dependiendo del tipo de dato.
- Si no deseamos reemplazar el campo vacío y/o nulo con el dato sugerido por los expertos, tendremos la opción de ingresar el dato que se deseé.

Recordemos también uno de los autores mencionados en la investigación como es Quinlan en su definición de Algoritmo ID3 el cual dentro del estudio será mejorado y conocido como “Ideal”, es un algoritmo potente, que permite elaborar un árbol de decisión como un método para aproximar una función objetivo de valores discretos, que es resistente al ruido en los datos y que es capaz de hallar o aprender de una disyunción de expresiones. “Para construir el árbol, el algoritmo utiliza el análisis de la entropía y la ganancia de información” (Quinlan, 1986).

Este árbol ideal generado por un algoritmo mejorado estará conformado por atributos puros que ya fueron explicados y los cuales deben cumplir con las siguientes posibilidades:

ATRIBUTOS TIPO 1:

- Validados como dato completamente en su ingreso.
- Información llena en todas sus ocurrencias.
- Datos con significado que ayuden a generar conocimiento.
- Datos con hechos referentes, de interés para distintos objetivos.
- Datos que deben tener sentido para los expertos involucrados y para los usuarios finales quienes tomarán decisiones.

ATRIBUTOS TIPO 2:

- Información existente en redes sociales, que inmersos en un análisis de sentimientos, sean válidos para el estudio.
- Datos a los que daremos valores entre 1 y 5 según su polaridad, que serán revisados en un motor de análisis sentimental de tipo caja negra y en el que consideramos solo entradas y salidas.

Y con estas consideraciones especiales de la conformación de los elementos del árbol de decisión ideal, ahora nos basaremos en lo descrito anteriormente sobre el Diseño ETC, poniendo las siguientes etapas que son descritas con cada una de sus actividades en la Tabla 22, para la aplicación del Algoritmo Ideal dentro de la Metodología de Decisión que será implementada en M3S:

Tabla 22. Etapas de Metodología de Decisión. (Propia, 2016).

ETAPAS	ACTIVIDADES
1. Análisis de Calidad de Datos	Analizar datos para un proceso de decisiones, mediante uso de diccionarios de datos, estadísticas, redes sociales y conocimiento de expertos; escogiendo los atributos más relevantes para el estudio, con las características de pureza explicadas anteriormente, sean del tipo 1 (estructurado) o tipo 2 (no estructurado).
2. Hipótesis de Datos	El objetivo que tendrá esta parte metodológica es justamente encontrar una descripción del problema planteado, buscando mediante suposiciones generales a partir de hechos concretos y a raíz de estos clasificar la información para generalizar los casos nuevos.
3. Generación de la tabla para Entrenamiento	Esta tabla es el producto final de los datos limpios y lo más refinados posible según su categoría; así también de la formalización de la o las reglas de negocio previamente establecidas con las cuales se han logrado clasificar la información. A raíz de esta tabla podemos generar el proceso matemático y estadístico para la creación del mejor árbol de decisión de acuerdo con el algoritmo ideal propuesto y a la objetividad que se haya dado en el análisis con la información encontrada.
4. Inferencia del Árbol de decisión	Para generar el árbol de decisión se utilizará inferencia inductiva que es aprendizaje supervisado. Este árbol es una forma de la aplicación del algoritmo ID3 con las mejoras específicas sobre la composición de nodos de dos tipos, el cual se basa en las reglas del negocio formalmente elegido; y del que se prevé tener la mayor ganancia de información, el mismo que se encuentra detallado en la etapa de construcción. Se procederá de forma recursiva de arriba hacia abajo; en cada nodo el atributo “más importante” que discrimina los ejemplos que han llegado hasta este nodo y por último clasificando los casos que siguen para el siguiente nivel.
5. Pruebas de datos sobre el árbol de decisión	Una vez que contamos con el árbol ideal de decisión, el cual con los cálculos de entropía nos genera la mayor ganancia de información; elaboramos un test para verificar que siempre se obtendrán los mejores resultados con nuevos datos que ingresen a evaluarse en el árbol. Se debe tomar en cuenta que, para algunos casos al probar nuevos datos, estos podrían hacer que cambie la clasificación y reglas de la tabla de entrenamiento, lo cual produciría una alteración en el árbol de decisión generado.

A continuación, se representa en forma de un modelo matemático cada uno de los pasos contemplados en la metodología de decisión, representado en base a funciones los datos

y a su vez englobándolos como conjuntos o grupos bajo su pertenencia. Se usan fórmulas como la de entropía y ganancia de Información que definen la posición de cada Nodo (Atributo) dentro del árbol de decisión ideal que se llegó a obtener:

1. Análisis de Calidad de Datos

Dado A conjunto de datos estructurados, y

Dado B conjunto de datos No estructurados ponderados (1..5)

Si $X \in A$ ó $X \in B$

Sea $f : X \rightarrow Y$, en donde:

X son todos los casos posibles de entradas

Y espacio de salidas

2. Hipótesis de Datos (H)

Serán todas las $h \in H$

3. Generación de la tabla para Entrenamiento:

Si $f \in H$ entonces es Aprendizaje considerado

Si $f \notin H$ entonces no será considerado.

4. Inferencia de Árbol de decisión

Usar conocimiento a priori para asegurar que $H \subset f$ en donde la inferencia se lo realiza mediante:

- Cálculo de la entropía

Dados:

- Un problema con dos clases (positiva y negativa)
- S , el conjunto de ejemplos.

$$\text{Entropía}(S) = \sum_i^k p_i \log_2 \left(\frac{1}{p_i} \right) \text{ con } k \text{ clases}$$

- Y también mediante la ganancia que se obtendrá

$$\text{Ganancia}(S, A) = \text{Entropía}(S) - \sum_{v \in \text{valores de } A} \frac{|S_v|}{|S|} \text{Entropía}(S_v)$$

5. Pruebas de datos sobre el árbol de decisión:

Utilizaremos medidas comunes de evaluación que son:

$$\text{promedio positivo verdadero} = \frac{TP}{TP + FN}$$

$$\text{promedio positivo falso} = \frac{FP}{FP + TN}$$

$$\text{precisión} = \frac{TP}{TP + FP}$$

$$\text{corrección} = \frac{TP + TN}{TP + FP + TN + FN}$$

Dónde:

TP: Verdaderos positivos (clasificado: positivo, realidad: positivo)
TN: Verdaderos negativos (clasificado: negativo, realidad: negativo)
FP: falsos positivos (clasificado: positivo, realidad: negativo)
FN: falsos negativos (clasificado: negativo, realidad: positivo)

Así como se describió en la Tabla 22 y matemáticamente el Algoritmo de Decisión Ideal; para una mayor comprensión también se encuentran como ANEXO V la representación bajo un diagrama de flujo y como ANEXO VI el pseudocódigo correspondiente.

Es relevante al finalizar este capítulo indicar que dentro de este proceso de investigación hemos considerado importantes acepciones dentro del estado del arte, sin embargo aquí mencionamos otros análisis desde otras visiones de autores que indican que “Un árbol de decisión es un conjunto de datos donde se fabrican diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema” (Rokach & Maimon, 2008). Como se indicó que los mejores atributos se pueden considerar para toma de decisiones generando aprendizaje automático; se dice también que “es un aspecto crucial del comportamiento inteligente de los seres humanos, que les permite acumular experiencias y adaptarse a entornos cambiantes frente a los que deben utilizarse nuevas estrategias” (Pino, Gómez, & de Abajo, 2001), otro autor dice que “es desarrollar técnicas que permitan a las computadoras aprender, crear programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos, induciendo conocimiento” (Moreno A. , 1994). De igual forma consideramos que aprendizaje supervisado es el algoritmo que produce una función que establece una correspondencia entre las entradas y las salidas deseadas del sistema, como lo es el problema de clasificación, donde el sistema de aprendizaje trata de etiquetar (clasificar) una serie de vectores utilizando una entre varias categorías (clases) (Salas, 2004), como lo es el Modelo de clasificación ID3, que significa "inducción mediante árboles de decisión" que aplica la estrategia "divide y vencerás" para hacer la clasificación, implementando métodos y técnicas para la realización de procesos inteligentes, representando así el

conocimiento y el aprendizaje, automatizando tareas. Recopilando nuevas definiciones de Entropía, la cual toma en cuenta la incertidumbre, impureza, desorden de un conjunto de datos relativo a una clasificación binaria, a su vez se define matemáticamente como la propiedad de que si se asigna $p = 1$ a una de las respuestas, consecuentemente $p = 0$ a todas las demás, en cuyo caso la entropía resulta ser cero. Y si todas las probabilidades son iguales, la entropía es máxima (García R. , 2018); y por último recordamos a la Ganancia de Información, la cual ocurre por el menor valor de entropía esperado (S), al calcular todo el subconjunto de atributos que fueron evaluados según la calidad de información que guardan.

De todo el detalle que encontramos en la revisión de M3S, podemos concluir en primer lugar que las eficiencias de los algoritmos de árboles de decisión dependen en gran medida de la pureza de información que podamos conseguir, el conocimiento que nos entreguen los expertos de negocio, para generar una óptima tabla de entrenamiento (TE), sobre la cual la implementación de ID3-Modificado como lo hemos llamado, nos entrega un árbol de decisión ideal que mantiene la mayor ganancia de información posible, que sin embargo está sujeto a cambios en base a nueva información que puede ingresar al sistema. También es válido indicar que el definir una tabla de entrenamiento que mantiene todas las agrupaciones concretas de datos, garantiza la fiabilidad de las decisiones inducidas, por eso es que como resultados iniciales, se creó una Metodología que aporta como guía dentro de ETC, en la cual lo más relevante es que se complementa un algoritmo de aprendizaje para extraer información valiosa de las empresas.

Finalmente, basados en resultados iniciales, que deberán ser nuevamente validados, considero que los objetivos planteados en esta investigación serán alcanzados en el tiempo previsto.

5. CAPÍTULO V: PRUEBAS METODOLOGÍA M3S

De la idea que toda propuesta debe ser validada, en esta sección lo haremos para la metodología M3S, mediante su aplicación en empresas con diferentes niveles de madurez de información en proyectos de BI, en donde sus usuarios pongan en marcha todas las fases de esta propuesta, haciendo énfasis en la parte de diseño de E.T.C. con la aplicación de la metodología de decisión, al mismo tiempo que se genera la documentación de soporte requerida por M3S. En este contexto seleccionamos dos empresas como grupo de prueba en las que se testeó la metodología en proyectos que ya los ejecutaron en sus instituciones lo que permitió realizar un verdadero comparativo real de las dos propuestas; esto se hizo durante un período de tres meses (12 semanas), puesto que se consideran ya establecidas y en funcionamiento diferentes componentes y plataformas tecnológicas tanto de arquitectura de software como de hardware y otros tipos de recursos (humano, financiero); entonces enfatizamos que este tipo de proyectos los visualizamos como de una perspectiva de mayor nivel informático ya que no son exclusivamente de desarrollo de software, si bien es cierto se analiza, diseña, construye e implementa indicadores de gestión con herramientas sumamente rápidas, la verdadera importancia de todo lo que

involucra la inteligencia de negocios fundamentan su importancia en la explotación de información, que genere conocimiento para la toma de decisiones.

Para llevar a cabo estas pruebas cada empresa designó responsables y con la colaboración de estudiantes del último semestre de la Carrera de Ingeniería Informática, se cumplió con la implementación de M3S en su totalidad, siguiendo la secuencia que a continuación indicamos:

En la primera semana se realizaron dos charlas con los involucrados de cada empresa y los estudiantes en donde se establecieron las definiciones teóricas y se explicó la metodología. Para el seguimiento se revisaron en las próximas cinco semanas la documentación generada y básicamente los formularios de soporte que constan en los anexos de esta investigación.

En la séptima semana realizamos reuniones de los grupos de trabajo en donde se generó la discusión sobre lo ejecutado al momento, como resultado de las mismas se ajustaron ciertos detalles finales necesarios y se avanzó con los proyectos por tres semanas, en donde con una data ya refinada como se planteó anteriormente la pureza de información, se aplicaron herramientas de BI analizadas en el capítulo III, y las cuales sobre todo se encuentran disponibles para pruebas académicas dentro de la universidad, obteniéndose tableros con indicadores críticos de los objetivos planteados en cada caso.

En las dos últimas semanas se cerraron los proyectos y se procesaron las conclusiones obtenidas para la validación de la metodología en los casos realizados; para este efecto se propuso una reunión plenaria en donde se consolidaron las experiencias ganadas y las mismas expuestas para nuevos proyectos.

5.1. CASO EMPRESA ELÉCTRICA QUITO (E.E.Q.)

En la búsqueda de mejorar la efectividad de proyectos en los que se emplea Inteligencia de Negocios, hemos configurado una metodología que creemos puede abordar esta problemática. Para conseguir este objetivo fue probada en empresas con un mediano y alto nivel de madurez de información por lo que se ha realizado un caso de estudio en la Empresa Eléctrica Quito en uno de sus proyectos realizados dentro del área de tecnología, para de esta forma tener un parámetro comparativo entre dos formas metodológicas que llevan a un mismo resultado, o por lo menos muy cercano; a su vez la metodología propuesta en una de sus etapas se sustentará en una aplicación modificada del algoritmo ID3 conocido en la investigación como árboles de decisión ideales, que fueron explicados en el capítulo anterior.

El cambio de la matriz energética iniciada en el gobierno anterior en el Ecuador, posee varios elementos, uno de ellos corresponde a los programas de uso eficiente de energía eléctrica tanto en el sector industrial como el residencial, por ello se trata de que la ciudadanía comprenda y acepte este cambio en su nivel de vida; por lo que una alternativa para reducir el consumo eléctrico y al mismo tiempo el costo de la planilla sin afectar nuestra calidad de vida, es la implementación de focos led, que por una parte son alternativas más apropiadas con el medioambiente y por otra nos permiten una eficiencia energética de hasta un 80% frente a los focos de iluminación tradicionales.

Como objetivo general de este caso de estudio se mencionó lo siguiente: Validar la Metodología M3S en la E.E.Q, identificando a los clientes en las distintas parroquias del Distrito Metropolitano de Quito (capital del Ecuador) para los cuales se pueda generar una

campaña de masificación de uso de focos led; apoyándose en los siguientes objetivos específicos:

- Aplicar y validar la Metodología de decisión en la etapa de diseño de M3S.
- Hacer uso de árboles de decisión para inferir uno de forma ideal, con reglas que nos permitan llegar a cumplir la campaña planteada a los estratos que lo requieran.
- Justificar nuestra decisión basados en el algoritmo de árboles ideales, con el apoyo de expertos en el área de distribución eléctrica.
- Seleccionar las partes de la base de datos adecuadas para aplicar árboles de decisión ideales de tal forma que las predicciones o decisiones que puedan tomarse tengan cada vez menos incertidumbre.

Planteado así el estudio se aplicó y describió la metodología M3S propuesta en el capítulo anterior en cada una de sus etapas, con la información obtenida de la EEQ.

5.1.1. APLICACIÓN METODOLOGÍA DE M3S EN EEQ:

En la aplicación metodológica la importancia de este caso radicó en que se reprodujo un proyecto realizado en meses anteriores por parte de la EEQ y se lo cotejó aplicando M3S que si bien es cierto se lo plantea como el desarrollo de un proyecto de TI, éste no se refiere a producción exclusiva de software sino más bien a la explotación de información lo más refinada para la toma de decisiones. Además estamos seguros que se lo llevó a cabo con fortalezas importantes a saber la unión de información de distintas fuentes de datos con origen totalmente heterogéneo como son el Sistema de Comercialización SIDECOM el mismo que se encuentra soportado en diversas bases de datos como ORACLE, DB2 del AS400 de IBM, archivos tipo hoja de cálculo EXCEL en donde se tienen en algunos casos los pliegos tarifarios aprobados por el Ministerio de Electricidad que son aplicados a los

clientes y por último archivos planos en los cuales contienen en parroquias rurales se tiene el catastro de clientes, es decir se ratificó en este caso la gama totalmente heterogénea en información de diferentes sectores. Para mayor soporte se nos facilitó y explicó la documentación que maneja el área de tecnología como es: Arquitectura de Software, Metodología de Desarrollo de Sistemas y Metodología de Desarrollo de Proyectos; que para el aspecto de estudio nos sirvió en la realización de las comparaciones necesarias.

Y es justamente por lo indicado que es fundamental el hecho de que se quiere documentar de una manera más ágil los proyectos de IN sin que esto sea un tema burocrático para su presentación utilizando formularios sencillos pero que recojan los hitos importantes de cada proyecto; y por último antes de entrar en la rigurosidad de las etapas el hecho de plantear un árbol de decisión ideal de tal forma que implementemos aprendizaje automático significa otro modo de llevar a cabo este tipo de planteamientos en el área. Así esquematizado el proyecto explicamos como se desarrolló cada etapa de M3S en EEQ:

A.) Justificación, con la dirección de tecnologías se definió denominar a este proyecto como: Campaña de utilización de focos LED; el mismo debería entregar información de zonas y clientes en donde podría aplicarse la campaña. Se designó como responsable por parte de la EEQ a la Ing. Cecilia Sánchez y por parte de la U.C.E. al Ing. Santiago Morales C., como apoyo estarán 2 estudiantes de los últimos años de pregrado, cada uno con 10 horas semanales de carga para el proyecto el mismo que se llevó a cabo en un período de 3 meses. Al momento de la evaluación se buscó probar por una parte la pertinencia de la metodología y por otra la disminución de tiempo que podríamos obtener. Todo esto se refleja con datos en el ANEXO I de esta investigación.

B.) En la segunda etapa de Planificación, se analizó la infraestructura de hardware la misma que consistió en una plataforma tecnológica robusta que garantizó un buen

rendimiento al final con los resultados. Por otro lado, se definieron las siguientes políticas de trabajo: Horarios de trabajo del equipo a partir de las 7H:00 hasta las 9H:00, y el personal autorizado a colaborar en diferentes labores, para lo que se usó el ANEXO II. Se definió el alcance del proyecto planteado en su objetivo. El ANEXO III, ayudó a documentar esta parte que tiene parámetros importantes para el seguimiento.

C.) Análisis y Requisitos de Negocio, en esta etapa se recogió información de diferentes partes en primer lugar de una sección de la base de datos del SIDECOM (Sistema de Comercialización Eléctrico). Adicionalmente se analiza el impacto que tiene el consumo de energía en cierto grupo de clientes en la Ciudad de Quito, para lo cual se busca alternativas de reducción del consumo lo que representa igual una menor planilla de pago y todo esto recae en una mejora del problema socio económico en la ciudad. Se definieron con ayuda de técnicos en el área eléctrica la aclaración de preguntas como: ¿Qué vamos hacer?, ¿Con quién lo vamos hacer? ¿Y Para qué lo vamos hacer?. Y en esta etapa también se planteó un prototipo de tableros con indicadores que podrían manejarse y el mismo se presenta en el ANEXO IV.

D.) Diseño, es en la etapa de este estudio en la cual definimos una nueva metodología que trata de optimizar la obtención de la información más importante dentro del ámbito que se está proponiendo y la misma la describiremos como se la llevo a cabo al finalizar las etapas generales de M3S.

E.) Construcción, una vez finalizada la generación de una tabla de entrenamiento con los mejores atributos en términos de pureza que resuelvan este problema y quienes según su ganancia de información establecieron un árbol de decisiones ideal; en este punto nos ayudamos de la herramienta que esté disponible para el cliente en este caso fue TABLEAU, con la que se obtuvieron tableros con indicadores de alta calidad.

F.) Implementación, en esta etapa lo que se consiguió es que nos faciliten datos correspondientes al primer semestre del año 2016 y con los cuales por un lado se realizó la aplicación del árbol de decisión ideal y por otro fueron aplicados bajo los tableros ya construidos, evidenciado los resultados con los mismos criterios obtenidos inicialmente. Así la puesta en marcha se la hizo en forma incremental según obteníamos la información y la misma quedó expresada en el ANEXO VII, en donde se dejó reflejado información técnica y funcional del proyecto. Luego realizamos la evaluación de la solución con la dirección de proyectos de la cual se mantiene un acta de reunión y la misma pudo ratificar los resultados y dejar latente el interés y la posibilidad de usar M3S en futuros estudios. Y por último firmamos el acta de cierre de proyecto con la responsable de la EEQ, documento que servirá exclusivamente como un respaldo académico.

Para todos estos pasos incluido el detalle que vamos a explicar de la etapa de diseño se cuenta con archivos de apoyo que documentan la parte técnica y funcional del proyecto. A continuación se detalla las actividades realizadas en la Metodología de Decisión creada justamente en la etapa D.) de M3S descrita en la página 106 de este documento.

5.1.1.1 APPLICACIÓN METODOLOGÍA DE DECISIÓN:

1: Análisis de Calidad de Datos: El estudio analizó y contempló datos de facturación de las parroquias del Distrito Metropolitano de Quito correspondientes a los 12 meses del año 2016 del sector tipificado en la EEQ. El total de registros utilizados para el estudio inicialmente fue de 12.144.333, de los cuales nos dedicamos exclusivamente aquellos que se encuentran categorizados en el sector Residencial. Dentro de la Figura 17, se cuenta

con información acerca de algunas parroquias en los sectores de Quito a la cual irá dirigido nuestro estudio.

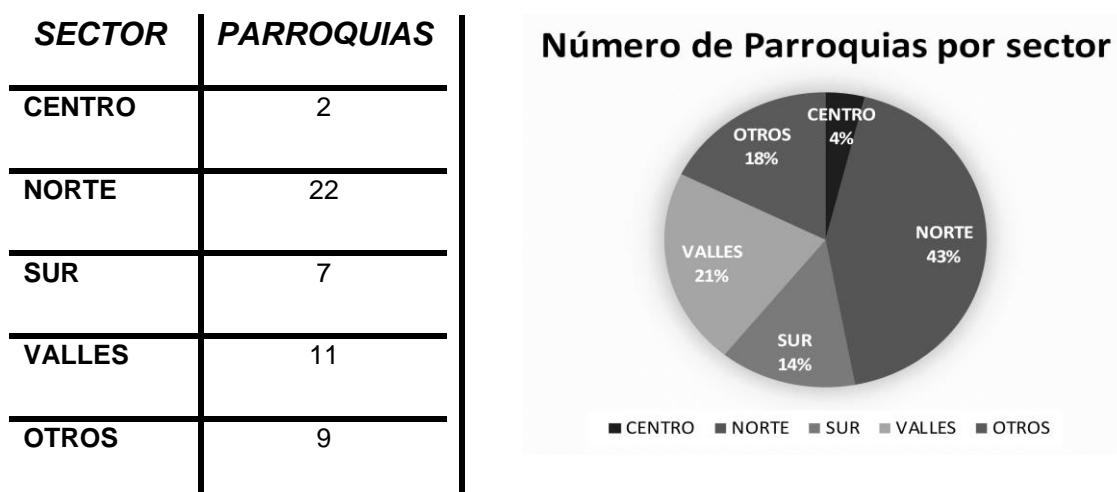


Figura 17. Parroquias del Distrito Metropolitano de Quito

De igual forma la EEQ se tiene en la Figura 18, como información el número de facturaciones de clientes realizadas por el sector de vivienda y son los que participarán en nuestro estudio.

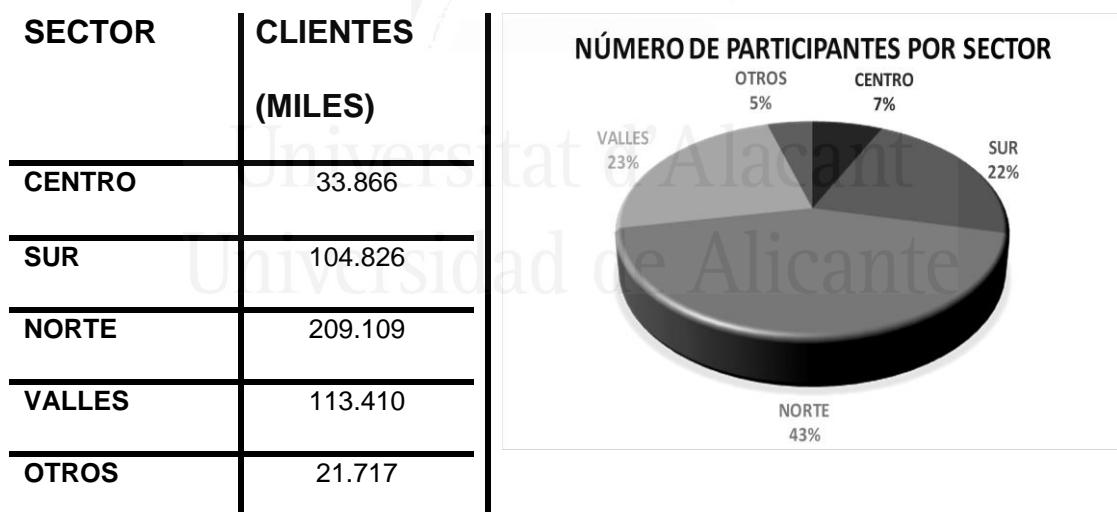


Figura 18. Clientes por Sectores del Distrito Metropolitano de Quito

2: Hipótesis de Datos: Dentro de este caso de estudio se revisaron un gran número de atributos y en base al análisis realizado se identificaron aquellos que podrían generar mayor valor con altos niveles de pureza y a su vez tener posibilidad de dicotomía para la

generación del árbol de decisión ideal; por su importancia dentro de la campaña, dos de ellos están a continuación con su respectiva valoración:

ZONA_CAMPANA toma los valores:

- SI en el caso que SECTOR sea SUR o CENTRO
 - NO en el caso que SECTOR sea cualquier otro.

SECTOR_CAMPANA toma los valores:

- RESIDENCIAL en el caso que TIPO_CLIENTE sea RESIDENCIAL
 - OTROS en el caso que TIPO_CLIENTE sea INDUSTRIAL, COMERCIAL, DIGNIDAD, etc.

3: Generación de la Tabla de Entrenamiento: Esta tabla que contiene los datos de mayor relevancia, se generó en base a la información de expertos que brindando su experiencia en la empresa ayudaron a la creación de las reglas del negocio para focalizar en este caso particular a los clientes hacia quienes queremos llegar con la campaña. Así establecidas las reglas se definió que contamos con cuatro atributos que cumplen la rigurosidad de refinamiento, que según su valoración darán la decisión que deba ser tomada para cada usuario del servicio eléctrico y como ejemplo se encuentran en la Tabla 23, una parte de ellas las que fueron aplicadas en el proceso.

- Zona Campaña= SI, Tipo identificación = C
 - Sector Campaña=RESIDENCIAL, Consumo > 200 Kwh
 - SI en el campo Decisión, cuando cumpla las 4 condiciones.
 - NO, en todas las otras combinaciones.

Tabla 23. Reglas de Negocio en base a Consumo y Otros.

SECTOR	TIPO CLIENTE	PROMEDIO CONSUMO MES(Kwh)	TIPO IDENTIF.	ZONA CAMPANA	SECTOR CAMPANA	DECISION
SUR	COMERCIAL	<= 200Kwh	C	SI	NO	NO
SUR	COMERCIAL	<= 200Kwh	P	SI	NO	NO
SUR	COMERCIAL	<= 200Kwh	R	SI	NO	NO
SUR	COMERCIAL	> 200Kwh	C	SI	NO	NO
SUR	COMERCIAL	> 200Kwh	P	SI	NO	NO
SUR	COMERCIAL	> 200Kwh	R	SI	NO	NO
SUR	RESIDENCIAL	<= 200Kwh	C	SI	SI	NO
SUR	RESIDENCIAL	<= 200Kwh	P	SI	SI	NO
SUR	RESIDENCIAL	<= 200Kwh	R	SI	SI	NO
SUR	RESIDENCIAL	> 200Kwh	C	SI	SI	SI
SUR	RESIDENCIAL	> 200Kwh	P	SI	SI	NO
SUR	RESIDENCIAL	> 200Kwh	R	SI	SI	NO
SUR	RESID-DIGNIDA	<= 200Kwh	C	SI	NO	NO
SUR	RESID-DIGNIDA	<= 200Kwh	P	SI	NO	NO
SUR	RESID-DIGNIDA	<= 200Kwh	R	SI	NO	NO
SUR	RESID-DIGNIDA	> 200Kwh	C	SI	NO	NO
SUR	RESID-DIGNIDA	> 200Kwh	P	SI	NO	NO
SUR	RESID-DIGNIDA	> 200Kwh	R	SI	NO	NO
SUR	INDUS-ARTESA	<= 200Kwh	C	SI	NO	NO
SUR	INDUS-ARTESA	<= 200Kwh	P	SI	NO	NO
SUR	INDUS-ARTESA	<= 200Kwh	R	SI	NO	NO
SUR	INDUS-ARTESA	> 200Kwh	C	SI	NO	NO
SUR	INDUS-ARTESA	> 200Kwh	P	SI	NO	NO
SUR	INDUS-ARTESA	> 200Kwh	R	SI	NO	NO
NORTE	INDUSTRIAL	<= 200Kwh	C	NO	NO	NO
NORTE	INDUSTRIAL	<= 200Kwh	P	NO	NO	NO
NORTE	INDUSTRIAL	<= 200Kwh	R	NO	NO	NO
NORTE	INDUSTRIAL	> 200Kwh	C	NO	NO	NO

Al procesar estas reglas, los clientes que poseían la característica de 'SI' en la decisión, fueron los seleccionados para iniciar con lo previsto en este proyecto y en donde al momento ya se contaría con la tabla de entrenamiento que se presenta como Tabla 24, la cual contiene toda la información necesaria, para poder continuar al siguiente paso.

Tabla 24. Tabla de Entrenamiento Caso E.E.Q.

MES FACT	NOMBRE	TIPO	NÚMERO	TIPO CLIENTE	PARROQUIA	SECTOR	ZONA	SECTOR	MES	EXCÉDE	DECISIÓN
			IDENTIFICACIÓN								
Enero	Zurita Flores	C	1500308760	R	EL Inca	Norte	NO	R	126	NO	NO
Enero	HotelBoutique	R	1792594049001	R	San Roque	Centro	SI	R	199	NO	NO
Enero	RiveraBarragan	C	1714199914	RD	Calderón	Norte	NO	O	97	NO	NO
Enero	Tobar Bedon	C	1000173433	R	Magdalena	Sur	SI	R	408	SI	SI
Enero	Mosquera Vinio	C	1702162452	RD	Magdalena	Sur	SI	O	87	NO	NO
Enero	Coronel Soto	C	1100480597	RD	Magdalena	Sur	SI	O	85	NO	NO
Enero	Perez Ramos	C	1720497120	R	Magdalena	Sur	SI	R	254	SI	SI
Enero	Amaguan Pillaj	C	1709332025	RD	Magdalena	Sur	SI	O	57	NO	NO
Enero	Chulca Pillajo	C	1701480558	R	Magdalena	Sur	SI	R	525	SI	SI
Enero	Quinga Yugch	C	1701343996	RD	Magdalena	Sur	SI	O	48	NO	NO
Enero	Andrade Z Ligia	C	1303272346	R	El Batan	Norte	NO	R	160	NO	NO
Enero	Achig Pillajo	C	1704854692	R	Magdalena	Sur	SI	R	606	SI	SI
Enero	Chulca Correa	C	1700127218	R	Magdalena	Sur	SI	R	225	SI	SI

IDENTIFICACION	DESCRIPCIÓN
C	Cédula
R	RUC

CLIENTE	DESCRIPCIÓN
R	Residencial
RD	Residencial Dignidad

4: Inferencia del Árbol de Decisión ideal (ID3-Mejorado): Al momento contamos con la tabla de entrenamiento de datos y vamos a calcular la entropía mediante el que elegiremos entre todos cuál de los atributos dará mayor ganancia de información tras clasificar los vectores de entrada. Esta elección la haremos basada en los 4 pasos que se explican en la Tabla 25, los que se aplicarán con los atributos Zona Campaña, Tipo de identificación, Tipo Cliente y Consumo en Kwh.

Tabla 25. Cálculo de Entropía y Ganancia de Información.

Paso 1: Calcular la entropía de los registros en la tabla	$I[P(r_1), \dots, P(r_n)] = \sum_i -P(r_i) \cdot \log_2[P(r_i)]$
Paso 2: Para cada atributo calcular la entropía de los subconjuntos	$I_{Ei}[P(r_1 v_i), \dots, P(r_n v_i)] = \sum_j -P(r_j v_i) \cdot \log_2[P(r_j v_i)]$
Paso 3: Para cada valor del atributo A_k , se calcula la cantidad de entropía:	$I(A_k) = \sum_j P(v_j) \cdot I_{Ej}[P(r_1 v_j), \dots, P(r_n v_j)]$
Paso 4: Elegir aquel atributo que obtiene la mayor ganancia de información	$Ganancia(A_k) = I_E - I(A_k)$

Para mejor entendimiento en la siguiente Tabla 26, se describen como ejemplos los posibles casos que se han encontrado luego de todo el análisis de hechos reales sobre los datos de facturación de la EEQ, los mismos que serán los únicos considerados en el momento de la inferencia del árbol ideal.

Tabla 26. Hechos Reales Caso EEQ.

CLIENTE	ZONA CAMPAÑA	SECTOR CAMPAÑA	EXCEDE 200Kwh	TIPO IDENTIFICACION	DECISION
Ejemplo. 1	SI	RESIDENCIAL	SI	C	SI
Ejemplo. 2	SI	COMERCIAL	SI	R	NO
Ejemplo. 3	SI	OTROS	NO	P	NO
Ejemplo. 4	SI	RESIDENCIAL	SI	C	SI
Ejemplo. 5	NO	INDUSTRIAL	NO	C	NO
Ejemplo. 6	SI	RESIDENCIAL	SI	R	NO

Es importante explicar que para este caso de estudio en el primer nivel presentaremos todos los cálculos realizados tanto para la entropía total del nivel, como para cada uno de los atributos; tomando en cuenta que es un proceso recursivo, a partir del segundo nivel solo se mostrará la entropía total y un cuadro con los valores encontrados para cada atributo.

A continuación, aplicamos los 4 pasos descritos anteriormente (Tabla 25.) en los cuadros que van generándose y para cada uno de los atributos, de los cuales el que más ganancia de información obtenga, será el seleccionado como nodo para cada nivel.

ITERACIÓN 1: En primer lugar, se calculó la entropía total (Nivel 1) de todos los atributos aplicando el paso 1 de la Tabla 25, y en este caso se obtiene un valor de 0,06; dado esta primera parte importante para cada iteración, tomamos indistintamente cualquier atributo y lo hacemos en este caso con zona de campaña para el que consideramos su probabilidad de ejecución según el número de registros que cumplan esta condición en lo que se encuentra que en el 31% de los casos se cumple como zona de interés para la campaña y en el 69% no. Como se muestra en la Tabla 27, con esta información se procede al cálculo de la entropía del atributo zona de campaña con las probabilidades obtenidas de ocurrencia generando una entropía del nodo 0,0484 y con este valor aplicando la ganancia de información es decir restado del vector de entrada de todos los atributos en esta iteración se obtiene que para Zona de campaña es de 0,012. presenta a continuación:

Universitat d'Alacant
Universidad de Alicante

Tabla 27. Cálculo Entropía y Ganancia Nodo: Zona Campaña.

Entropia total			$E(Total) = \sum_{t \in C} -p_t * \log_2 p_t$		
Total registros	Total Si	Total no	$E(s)$		
1340166.00	9440.00	1330726.00	0.06		
NODO ZONA DE CAMPAÑA					
V_i		$Prob(X=v_i)$	$Prob(X=v_i) * E(r)$		
NODO	Cantidad de Registros	P(ZonaVenta)	$E(nodo)*P(ZonaVenta)$		
Si	414556.00	0.31	0.048		
NODO	925610.00	0.69	0.000		
			P_1	P_2	$E(r) = \sum_{i \in C} -p_i * \log_2 p_i$
Nodo	Total Si	Total no	P(si)	P(no)	$E(nodo)$
Si	9440.0	405116.00	0.02	0.98	0.16
No	0.00	925610.00	0.0000000	1.00	0.00
$H(VENDER nodo) =$ $\sum Prob(X = v_j) * E(r)$	0.048481633				
$Ganancia =$ $E(Total) - H(VENDER nodo)$	0.012004445				

Universitat d'Alacant
Universidad de Alicante

En segunda instancia tomamos el atributo tipo de identificación para el que consideramos su probabilidad de ocurrencia según el número de registros sean cédula, ruc o pasaporte encontrándose que el 97% de los casos se cumplen para cédula y el 3% para ruc, pasaporte no tiene ocurrencias. Con esta información se procede al cálculo de la entropía del atributo tipo de identificación con las probabilidades obtenidas de ocurrencia generando un valor de 0,06 con este valor aplicando la ganancia de información se obtiene que para tipo de identificación es de 0,00028.

Tabla 28. Cálculo Primer Nivel Entropía y Ganancia Nodo: Tipo Identificación.

Entropía total			$E(Total) = \sum_{t \in C} -p_t * \log_2 p_t$		
Total registros	Total Si	Total no	$E(s)$		
1340166.00	9440.00	1330726.00	0.06		
NODO TIPO DE IDENTIFICACION					
V_i		$Prob(X=v_i)$	$Prob(X=v_i) * E(r)$		
NODO	Cantidad de Registros	$P(ZonaVenta)$	$E(nodo) * P(ZonaVenta)$		
Cédula	1303587.00	0.97	0.060		
RUC	34590.00	0.03	0.000		
Pasaporte	1989.00	0.00	0.000		
			P_1	P_2	$E(r) = \sum_{i \in C} -p_i * \log_2 p_i$
Nodo	Total Si	Total no	$P(si)$	$P(no)$	$E(nodo)$
Cédula	9440.0	1294147.00	0.01	0.99	0.06
RUC	0.00	34590.00	0.0000000	1.00	0.00
Pasaporte	0.00	1989.00	0.00	1.00	0.00
$H(VENDER nodo) =$ $\sum Prob(X = v_j * E(r))$	0.06				
$Ganancia =$ $E(Total) - H(VENDER nodo)$	0.000282236				

En tercera instancia consideramos el atributo consumo tomando su probabilidad de ocurrencia con un 96% a quienes registran menos de 200Kw/h y 4% los que registran más de ese valor. Al realizar el cálculo de la entropía de este atributo con las probabilidades obtenidas se genera un valor de 0,03 y a su vez con este aplicando la ganancia de información se obtiene que para consumo es de 0,0337.

Tabla 29. Cálculo Primer Nivel Entropía y Ganancia Nodo: Excede Consumo Promedio

Entropía total			$E(Total) = \sum_{t \in C} -v_t * \log_2 v_t$
Total registros	Total Si	Total no	$E(s)$
1340166.00	9440.00	1330726.00	0.06
NODO EXCEDE CONSUMO PROM. Kwh			
V_i		$Prob(X=v_i)$	$Prob(X=v_i) * E(r)$
NODO	Cantidad de Registros	P(ZonaVenta)	$E(nodo)*P(ZonaVenta)$
> 200Kwh	53196	0.04	0.03
$\leq 200\text{Kwh}$	1286970	0.96	0.000
			$P_1 \quad P_2 \quad E(r) = \sum_{i \in C} -v_i * \log_2 v_i$
Nodo	Total Si	Total no	$P(si) \quad P(no) \quad E(nodo)$
> 200Kwh	9440.0	43756.00	0.18 0.82 0.62
$\leq 200\text{Kwh}$	0.00	1286970.00	0.0000000 1.00 0.00
H(VENDER nodo)= $\sum_j Prob(X = v_j) * E(r)$	0.03		
Ganancia= $E(Total) - H(VENDER nodo)$	0.033713462		

En cuarta instancia se tomó el atributo sector de campaña en el cual su probabilidad de ocurrencia para residencial se da en un 48% y para otros un 52%. Con estas probabilidades calculamos la entropía del nodo obteniéndose un valor de 0,0529; aplicamos la ganancia de información y se tiene que para sector de campaña es de 0,0075.

Tabla 30. Cálculo Primer Nivel Entropía y Ganancia Nodo: Sector Campaña.

Entropía total			$E(Total) = \sum_{t \in C} -p_t * \log_2 p_t$		
Total registros	Total Si	Total no	E(s)		
1340166.00	9440.00	1330726.00	0.06		
NODO SECTOR DE CAMPAÑA					
V_i		$Prob(X=v_i)$	$Prob(X=v_i) * E(r)$		
NODO	Cantidad de Registros	P{ZonaVenta}	E(nodo)*P{ZonaVenta}		
RESIDENCIAL	639796.00	0.48	0.053		
OTROS	700370.00	0.52	0.000		
$E(r) = \sum_{i \in C} -p_i * \log_2 p_i$					
Nodo	Total Si	Total no	P(si)	P(no)	E(nodo)
RESIDENCIAL	9440.0	630356.00	0.01	0.99	0.11
OTROS	0.00	700370.00	0.0000000	1.00	0.00
$H(VENDER nodo) = \sum Prob(X = v_j) * E(r)$	0.052932683				
Ganancia= $E(Total) - H(VENDER nodo)$	0.007553394				

Tabla 31. Resumen de Cálculos Primer Nivel EEQ.

ATRIBUTO	ENTROPÍA	GANANCIA INFORMACIÓN
ZONA DE CAMPAÑA	0,0484	0,012
CONSUMO	0,03	0,0337
TIPO IDENTIFICACION	0,06	0,00028
SECTOR CAMPAÑA	0,0529	0,0075

De este primer barrido con todos los atributos como se muestra en la Tabla 31, observamos que obtenemos la mejor ganancia de información con el nodo *Consumo (Kw/h)*, el mismo que quedará en la parte inicial del árbol de la siguiente forma:

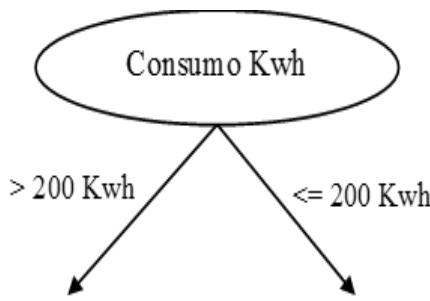


Figura 19. Nodo de Primer Nivel Árbol de Decisión. Caso EEQ.

ITERACIÓN 2: Obtenido ya el nodo de primer nivel, repetimos el procedimiento para la rama >200Kwh, pero esta vez sin tomar en cuenta el nodo *Consumo Kw/h*, y en la Tabla 32, hemos obtenido la entropía total en el segundo nivel de 0,67, y con el cual se procederá a los cálculos respectivos con cada atributo que todavía no ha sido considerado en el ramal superior.

Tabla 32. Cálculo Segundo Nivel Entropía y Ganancia Nodo: Zona Campaña.

Entropía total			$E(\text{Total}) = \sum_{t \in C} -p_t * \log_2 p_t$		
Total registros	Total Si	Total no	$E(s)$		
53196.00	9440.00	43756.00	0.67		
NODO >200Kwh - ZONA DE CAMPAÑA					
V_i		$Prob(X=v_i)$	$Prob(X=v_i) * E(r)$		
NODO	Cantidad de Registros	$P(\text{ZonaVenta})$	$E(\text{nodo}) * P(\text{ZonaVenta})$		
>200Kwh - Si	15263.00	0.29	0.275		
>200Kwh - No	37933.00	0.71	0.000		
$E(r) = \sum_{i \in C} -p_i * \log_2 p_i$					
Nodo	Total Si	Total no	$P(\text{si})$	$P(\text{no})$	$E(\text{nodo})$
>200Kwh - Si	9440.0	5823.00	0.62	0.38	0.96
>200Kwh - No	0.00	37933.00	0.0000000	1.00	0.00
$H(\text{VENDER} \text{nodo}) =$ $\sum Prob(X = v_j) * E(r)$	0.27518564				
$Ganancia =$ $E(\text{Total}) - H(\text{VENDER} \text{nodo})$	0.399296441				

De esta manera y luego ejecutar los cálculos respectivos se define en la Tabla 33, el resultado de los cálculos de entropía y el respectivo de ganancia de información, en la cual se determina que el nodo de segundo nivel será Zona_Campaña y de esta forma en la Figura 20 se presenta la nueva rama definida.

Tabla 33. Resumen de Cálculos Segundo Nivel EEQ.

ATRIBUTO	ENTROPÍA	GANANCIA INFORMACIÓN
ZONA DE CAMPAÑA	0,275	0,399
TIPO IDENTIFICACION	0,666	0,138
SECTOR CAMPAÑA	0,547	0,127

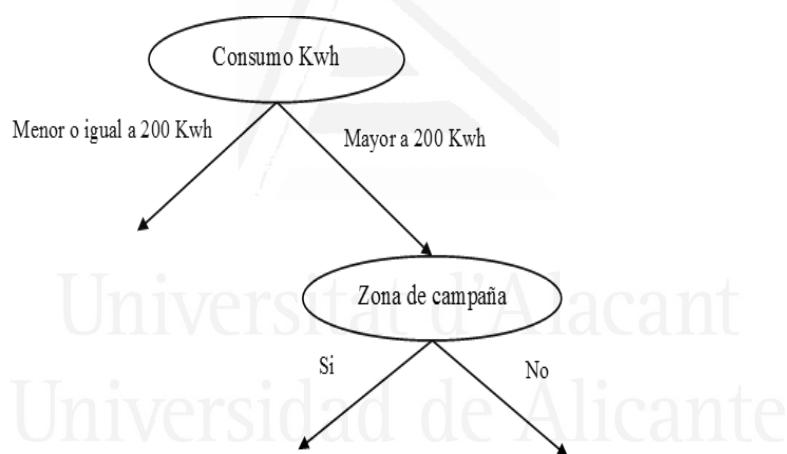


Figura 20. Nodo de Segundo Nivel Árbol de Decisión. Caso EEQ.

ITERACIÓN 3: Al momento contamos con dos niveles, repetimos este proceso en forma recursiva para la rama ‘Si’ de sector de campaña, pero esta vez sin tomar en cuenta los atributos consumo y zona de campaña; así obtenemos la entropía total de 0,96 en el tercer nivel y la Tabla 34 cuenta con los cálculos valores de entropía y ganancia de los atributos para la generación de este nodo.

Tabla 34. Cálculo Tercer Nivel Entropía y Ganancia Nodo: Zona_Campaña-Tipo Identificación

Entropia total			$E(Total) = \sum_{t \in C} -p_t * \log_2 p_t$		
Total registros	Total Si	Total no	$E(s)$		
15263.00	9440.00	5823.00	0.96		
NODO >200Kwh - ZONA DE CAMPAÑA-TIPO DE IDENTIFICACION					
V_i		$Prob(X=v_i)$	$Prob(X=v_i) * E(r)$		
NODO	Cantidad de Registros	P(ZonaVenta)	$E(nodo)*P(ZonaVenta)$		
> 200Kwh - SI - Cédula	14663.00	0.96	0.903		
> 200Kwh - SI - RUC	576	0.04	0.000		
> 200Kwh - SI - Pasaporte	24.00	0.00	0.000		
$E(r) = \sum_{i \in C} -p_i * \log_2 p_i$					
Nodo	Total Si	Total no	$P(si)$	$P(no)$	$E(nodo)$
> 200Kwh - SI - Cédula	9440.0	5223.00	0.64	0.36	0.94
> 200Kwh - SI - RUC	0.00	576.00	0.0000000	1.00	0.00
> 200Kwh - SI - Pasaporte	0.00	24.00	0.00	1.00	0.00
$H(VENDER nodo) =$ $\sum Prob(X = v_j) * E(r)$	0.90				
$Ganancia =$ $E(Total) - H(VENDER nodo)$	0.056548077				

Tabla 35. Resumen de Cálculos Tercer Nivel EEQ.

ATRIBUTO	ENTROPÍA	GANANCIA INFORMACIÓN
TIPO IDENTIFICACION	0,90	0,056
SECTOR CAMPAÑA	0,047	0,911

Como se muestra en la Figura 21, con este resultado se definió que el tercer nivel estaría compuesto por el nodo correspondiente a Sector de campaña, luego del cual por la rama residencial iría el último nodo que sería tipo de identificación; de esta forma la totalidad del árbol estaría estructurado de la siguiente forma:

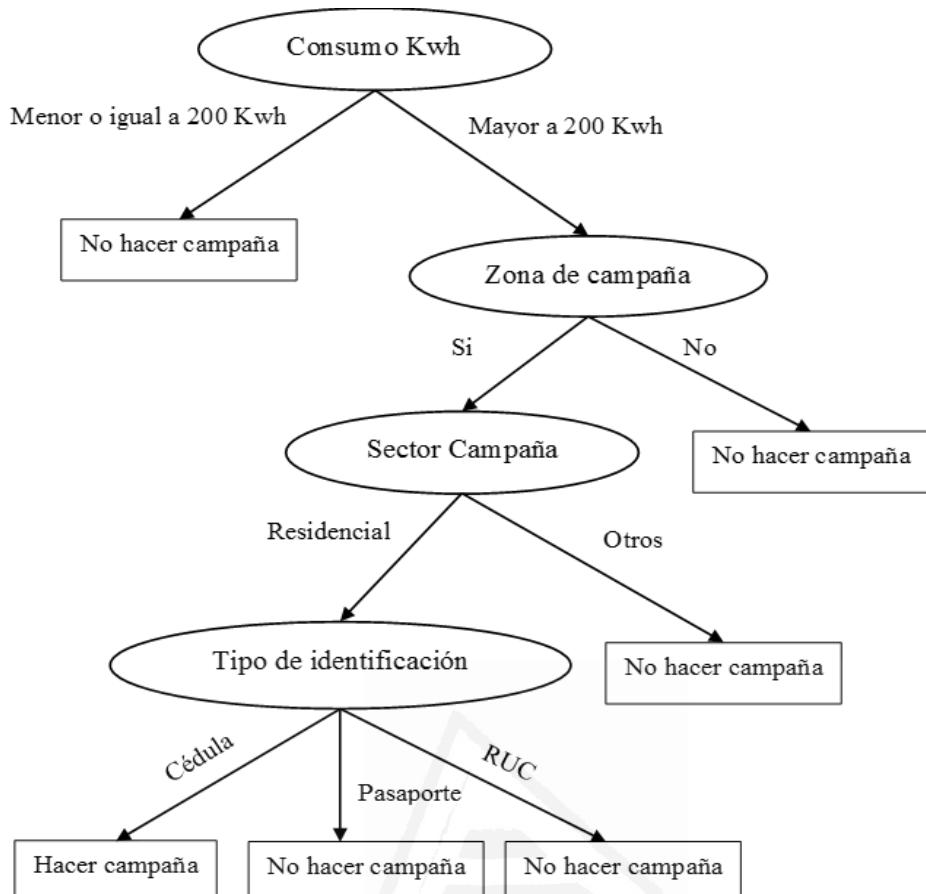


Figura 21. Nodo de Tercer Nivel Árbol de Decisión. Caso EEQ

5: Prueba de Datos sobre el Árbol de Decisión Ideal: Una vez conseguida la estructura del árbol de decisión ideal se realizan las pruebas sobre el mismo, y se llegó a tener información que ayudó a la empresa para tomar las decisiones sobre el objetivo planteado.

5.1.2. DISCUSIÓN DE RESULTADOS

Para la validación dentro de este caso de prueba de M3S, se usaron instrumentos de medición como encuestas las mismas que dentro de la EEQ se obtiene que por ser una empresa grande tiene objetivos claros y medibles, usan estándares externos para mejorar

y medir su gestión incluso a nivel de áreas, cuenta con metodologías aprobadas de desarrollo, infraestructura y de proyectos, presupuesto definido, proceso de mejoramiento continua. Sin embargo parcialmente se han cumplido los tiempos en proyectos, de igual forma manejan estándares de recolección de datos, cuentan con presupuestos de tecnología estos no son claramente designados a cada proyecto y el control de cambios no es tan formal. Con lo que no cuenta la empresa dentro del área de tecnología es una certificación de calidad.

Comparativamente analizando la metodología usada en EEQ contra M3S y haciendo un análisis cualitativo obtenemos lo siguiente: Con M3S se considera que el retorno de la inversión sería más efectivo que la actual, la aceptación es la misma entre lo que manejan actualmente, los clientes de la EEQ al tener alternativas de autoservicio se sentirían mejor atendidos agregándoles mayor valor; sin embargo, que manejan más tiempo la metodología actual sienten que M3S es más entendible

Por último dentro de lo posible cuantitativamente se obtiene que en personal se redujo en un 50%, así como el tiempo en depuración y transformación de datos y más aún en la construcción total del proyecto. El tiempo en extracción de datos se definió en un 40% menos, no se requirió ninguna inversión en hardware, se aumentó en un 30% el número de usuarios de BI; en este corto tiempo se duplicó el número de pedidos de proyectos de este tipo y en consecuencia los costos inherentes al proyecto bajarían en un 50%.

Y por último en la aplicación de M3S nos lleva como se muestra en la Figura 22, a obtener que de un total de 1.302.595 clientes analizados en las zonas planteadas; 87.506 de ellos que equivalen al 7% de la muestra adoptada deberán acogerse a la campaña de implementación de focos led en su domicilio.

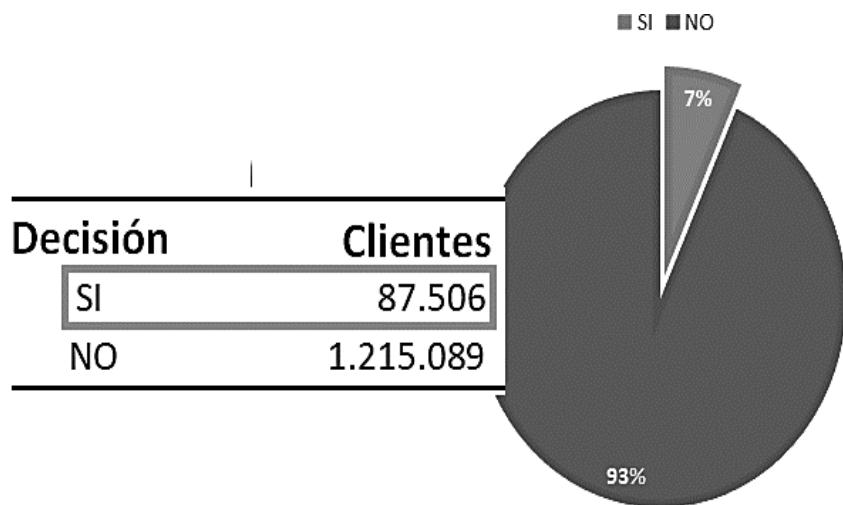


Figura 22. Resultado Caso de Estudio EEQ.

5.1.3. CONCLUSIONES CASO E.E.Q.

Una vez ejecutado todo el proceso metodológico de M3S dentro del área de tecnología de la EEQ y haber probado comparativamente varios aspectos sobre dos metodologías podemos indicar lo siguiente:

- Usando M3S se logró identificar al grupo de clientes hacia quienes se focalizará la campaña prevista por la EEQ.
- M3S dentro de este caso de estudio dio resultados semejantes en un altísimo porcentaje, pero siendo más eficiente tanto en tiempo, como en recursos humanos y por ende se proyectaría a disminuir los costos para su implantación.
- Se validó la utilización de M3S en general y dentro de la etapa de diseño se verificó la utilización del algoritmo de decisión ideal.
- Será muy útil en el sector público, puesto que al ser más ágil puede ayudar a mover el aparato burocrático en el cual se desarrollan las actividades en el sector gubernamental.
- Sin embargo, que este caso se puede volver extremadamente complejo por sus volúmenes de información, M3S logró desarrollar este objetivo planteado sin mayores problemas.

- El algoritmo de decisión ideal se valida nuevamente con buenos resultados tomando en cuenta los atributos escogidos por los expertos para lograr un aprendizaje automático basado en hechos pasados y concretos.
- La propuesta M3S, y su metodología, se ha validado contra los objetivos propuestos en este caso, se ha confrontado contra otra propuesta también validada en EEQ, brindando resultados positivos.

Para futuros estudios de BI, adicionalmente conocemos que el gobierno nacional con este mismo objeto implantó a nivel nacional el proyecto RENOVA que consistió en cambiar las refrigeradoras de los clientes del sector residencial, éste también fue desarrollado a nivel nacional. Siguiendo el mismo objetivo el Ministerio de Electricidad y Energía Renovable en coordinación con todas las empresas de distribución eléctricas del país implementan el proyecto PEC que consiste en el cambio de las cocinas de gas por cocinas de inducción eléctrica que está implantado parcialmente y se espera tenga mayor impulso a futuro.



5.2. CASO CLUB CASTILLO AMAGUAÑA (C.C.A.)

En el presente caso de estudio se consideró una empresa denominada Club Castillo de Amaguaña, de categoría recreativa y del sector privado del país, el cual de igual forma mantiene un convenio de colaboración con la U.C.E, y el mismo que tuvo la apertura de generar la necesidad de un proyecto dentro del ámbito de inteligencia de negocios y autorizar su ejecución. Se desarrolló para el área de cobranzas, área encargada de conseguir nuevos asociados que al incorporarse a los registros del club, ayudan a mejorar el presupuesto financiero de esta organización. El departamento de cobranzas tiene entre sus actividades llevar toda la información de recaudación mensual de los socios activos de la empresa.

Como antecedente el Club Castillo de Amaguaña es un club recreativo privado (IRP), dedicado al entretenimiento y esparcimiento familiar, lleva mediante su Sistema de Cobranzas la relación financiera entre la empresa y sus socios; generando un alto volumen de datos, que se convierten en una importante parte del historial transaccional de la organización.

Una alternativa para mejorar la productividad financiera del Club Castillo de Amaguaña es atraer nuevos socios mediante una campaña publicitaria y así aportar con el crecimiento y sostenibilidad de la empresa, por lo que como objetivo general de este caso de estudio se propone lo siguiente:

Generar una campaña originada en el departamento de cobranzas del Club Castillo de Amaguaña para mejorar el ingreso de nuevos socios cotizantes a partir de los 25 años de edad para incrementar el ingreso mensual de la institución y a su vez usar y validar como metodología de proyecto a M3S.

A su vez este se fortalece con los siguientes objetivos específicos:

- Difundir mayor conocimiento a los directivos con este proyecto de inteligencia de negocios, con el uso de una metodología pertinente a estos proyectos.
- Conseguir socios jóvenes en el club de tal manera de que estos se conviertan en un nuevo pilar financiero.
- Establecer los sectores óptimos donde debe realizar la campaña.
- Hacer uso de árboles de decisión ideales para poder identificar los sectores en donde realizar la campaña.
- Para justificar nuestra decisión nos basaremos en el algoritmo ID3 mejorado y con un fuerte soporte de expertos que conocen el área de cobranzas del club.

5.2.1. APLICACIÓN METODOLOGÍA DE BI M3S EN C.C.A.

El Castillo de Amaguaña mantiene en sus registros un total de alrededor de 12.000 personas como miembros activos en cada una de sus categorías (Cotizantes, No Cotizantes, Dependientes y Honorarios) donde el sistema de cobranzas maneja 3180 socios titulares cotizantes quienes son los que tienen la obligación de pagar una cuota mensual de funcionamiento.

La iniciativa de este análisis se da por la salida de socios debido a diferentes razones: económicas, familiares, profesionales entre otras. Por esta situación la intención es capturar nuevos socios a partir de los 25 años de edad, quienes serán los nuevos aportantes en términos económicos para el club.

Este proyecto tiene como principal fuente de datos al Sistema de Cobranzas, el mismo que se encuentra soportado en MySQL como base de datos, dicho sistema se encuentra dentro de un servidor Hp Proliant G9 con 16 Gb en RAM y un Tera en disco, desarrollado en Power Builder, que se ejecuta bajo un sistema operativo Windows. Además el sistema alimenta a otro llamado PAC, el que controla toda la parte financiera,

administrativa y contable que fue adquirido para el mejoramiento en procesos del club y este a su vez se encuentra en un servidor HP Proliant G6.

Para cada uno de estos sistemas se aplicaron diferentes metodologías de desarrollo y tienen su respectiva arquitectura de software, las mismas que fueron explicadas por el Departamento de Sistemas, esto nos dio las directrices para aplicar de manera más eficiente la documentación y ejecución de cada una de las etapas de este proyecto de BI. Una vez esquematizado el proyecto explicamos cómo se desarrolló cada etapa de M3S en C.C.A.:

A.) Justificación, conjuntamente con el departamento de sistemas, compuesto por cuatro ingenieros informáticos, se definió denominar a este proyecto como: Gestión de Cobranzas en el Club Castillo de Amaguaña; el mismo debería entregar información de zonas y clientes mayores de 25 años en donde podría aplicarse la campaña. Se designó como responsable del proyecto por parte del C.C.A. a la Ing. Ximena Tixi y por la U.C.E. al Ing. Santiago Morales C., como parte de apoyo estarán 3 estudiantes de los últimos semestres de pregrado, cada uno con 10 horas semanales de carga para el proyecto que debe llevarse a cabo en un período de 3 meses. Al momento de la evaluación se busca probar por una parte la pertinencia de la metodología y por otra básicamente la disminución de tiempo que podríamos obtener al aplicar M3S. De igual forma que en el caso anterior se llevó la documentación tal cual el ANEXO I de esta investigación.

B.) Planificación, se analizó la estrategia con el apoyo de la Gerencia además se realizó una revisión de la infraestructura de hardware con el apoyo del equipo técnico informático esta infraestructura consistió de un Servidor HP Proliant. PI180 G9 de 16Gb en RAM, 2 TB en disco Intel xeon, que para los requerimientos actuales cumple con las expectativas de obtención y almacenamiento de información. Se autorizó la disponibilidad de dos de los ingenieros en horario de las 16:00h a 18:00h

para cubrir el proyecto; de igual forma esto se registró en los formularios de M3S en el ANEXO II.

C.) Análisis y Requisitos de Negocio, recopilamos información de diferentes fuentes como el Sistema PAC de socios el mismo que tiene una arquitectura de 3 capas con Java y que se encuentra en una base de datos PostgreSQL; así como también de un sistema de cobranzas que tiene base de datos MySql desarrollado en Visual.net. Paralelamente se definen los tableros con focalización en los potenciales socios a manera de buscar alternativas que los atraigan a partir de los 25 años de edad dentro de los diferentes sectores de la Ciudad de Quito, y por último en esta etapa también se planteó un prototipo de tableros con indicadores para lo cual se uso el ANEXO IV.

D.) Diseño, en la siguiente sección se presenta el detalle de esta etapa con un resumen de cálculos y el árbol de decisión ideal al que se llegó.

E.) Construcción, una vez que se consiguió establecer una tabla de entrenamiento con los mejores atributos que resuelvan este problema y quienes según su ganancia de información establecieron un árbol de decisiones; en este momento se utilizó Qlick Sense, herramienta con la cual se llegó a obtener tableros de excelente composición y ayuda para los directivos de la empresa.

F.) La Implementación, se la ejecutó de forma incremental según obteníamos la información y la misma quedó expresada en el ANEXO VII, en donde se dejó reflejado información técnica y funcional del proyecto. Luego realizamos la evaluación de la solución con el personal de sistemas de la cual se mantiene un acta de reunión y la misma pudo ratificar los resultados que se obtuvieron y creemos que M3S puede ser usada en lo futuro también en esta empresa, con lo que se firmaron actas de cierre de proyecto con la responsable del Departamento de Cobranzas y

de Sistemas, documento que servirá de igual forma que en el estudio anterior como exclusivamente de un respaldo académico.

5.2.1.1 APLICACIÓN METODOLOGÍA DE DECISIÓN:

1: Análisis de Datos: El estudio analiza y contempla datos de facturación de los sectores del Distrito Metropolitano de Quito y sus alrededores correspondientes a los últimos 24 meses del año 2016 y 2017. El total de registro utilizados para el estudio inicialmente fue de 3.133 socios, centrandonos en personas mayores de 25 años Dentro de la Figura 23, se cuenta con información como el número de clientes facturados por sector que participan en nuestro estudio.

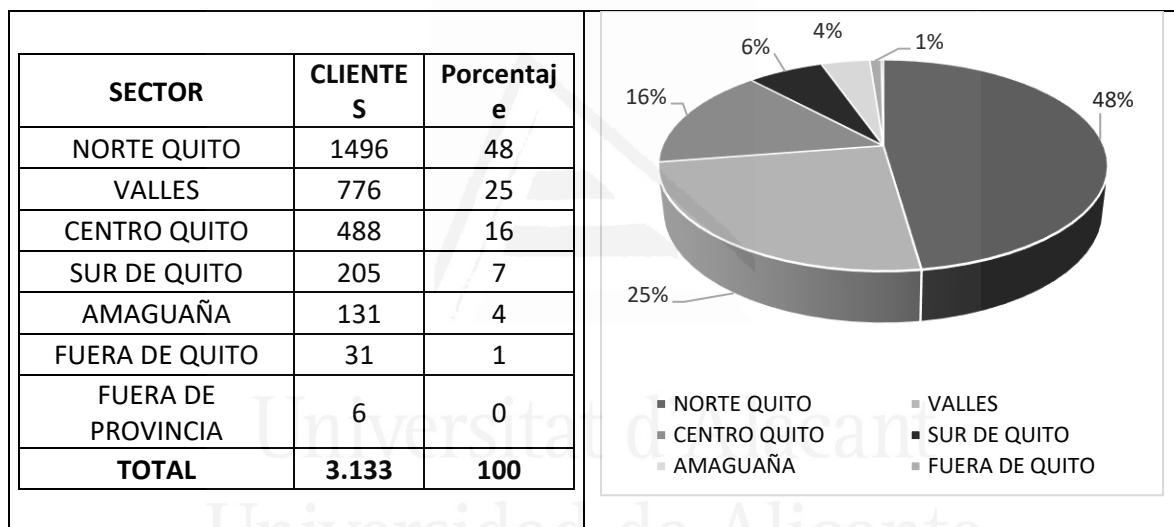


Figura 23. Clientes por Sectores del Distrito Metropolitano de Quito

2: Hipótesis de Datos: Dentro de este caso de estudio se identificaron atributos de gran importancia en niveles de significado y pureza, que deben ser analizados y asignados por un valor dicotómico, por su importancia dentro de la campaña, los mismos están a continuación con su respectiva valoración visualizados en la tabla 36, la cual analiza zonas de cobranza y el porcentaje de morosidad.

Tabla 36. Reglas de Negocio. C.C.A.

Reglas de Negocio	Decisión
ZONA_VENTA	SI en el caso que SECTOR sea Norte, Centro o Valles.
	NO en el caso que SECTOR sea cualquier otro sector.
% MOROSIDAD	NO si el porcentaje de morosidad es mayor al 13%.
	SI en el caso que el porcentaje de morosos sea menor o igual al 13%

3: Generación de la Tabla de Entrenamiento: Esta tabla que contiene los datos de mayor relevancia, se generó en base a la información de expertos que brindando su experiencia al Club ayudaron a la creación de las reglas del negocio para focalizar en este caso a los clientes hacia quienes queremos llegar con la campaña; en este caso los cuatro atributos según su valoración dieron la decisión que debe ser tomada para cada cliente:

- Zona Venta= SI, Estado Civil = C
 - Sector Campaña=Valles Edad > 25-30
- SI en el campo Regla.
NO, en todas las otras combinaciones.

Tabla 37. Tabla de Entrenamiento CLUB.

SOCIO	ZONA VENTA	SECTORES	ESTADO CIVIL	RANGO EDAD	VENDER
1	SI	VALLES	C	>40	NO
2	SI	NORTE QUITO	C	>40	SI
3	SI	VALLES	C	[36-40]	NO
4	SI	NORTE QUITO	C	[36-40]	SI
5	SI	VALLES	C	>40	NO
6	SI	CENTRO QUITO	S	[36-40]	SI
7	NO	OTROS	C	>40	NO
8	SI	NORTE QUITO	C	>40	SI
9	SI	NORTE QUITO	C	[25-35]	NO
10	SI	VALLES	C	>40	NO
11	SI	CENTRO QUITO	D	>40	NO
12	SI	NORTE QUITO	C	[25-35]	NO
13	SI	VALLES	C	[25-35]	SI
14	SI	CENTRO QUITO	C	>40	SI
15	NO	OTROS	S	[36-40]	NO
16	SI	NORTE QUITO	C	>40	SI
17	SI	NORTE QUITO	C	>40	SI
18	SI	CENTRO QUITO	C	>40	SI
19	SI	CENTRO QUITO	C	>40	SI
20	NO	OTROS	C	[36-40]	NO
21	SI	NORTE QUITO	C	>40	SI
22	SI	NORTE QUITO	S	[36-40]	NO
23	SI	NORTE QUITO	C	>40	SI
24	SI	NORTE QUITO	C	>40	SI
25	SI	VALLES	C	>40	NO
26	SI	VALLES	C	>40	NO
3130	SI	VALLES	C	>40	NO
3131	SI	VALLES	C	>40	NO
3132	NO	OTROS	C	>40	NO
3133	SI	VALLES	S	[25-35]	NO

Al procesar estas reglas, los clientes que posean la característica de 'SI' en la decisión, serán los seleccionados para iniciar con lo previsto en este proyecto y en este momento ya se cuenta con la tabla de entrenamiento la cual contiene toda la información necesaria, para poder continuar al siguiente paso.

4: Inferencia del Árbol de Decisión Ideal (ID3-Mejorado): Al momento contamos con la tabla de entrenamiento de datos y vamos a calcular la entropía mediante el que elegiremos el atributo de mayor calidad y que mayor ganancia de información tenga tras clasificar los vectores de entrada.

En la siguiente Tabla 38, se describen como ejemplos los posibles casos que se han encontrado luego de todo el análisis de hechos reales sobre los datos de cobranza del Club Castillo de Amaguaña, los mismos que son considerados en el momento de la inferencia del árbol.

Tabla 38. Hechos Reales Caso C.C.A.

Sector	Estado civil	Edades	Morosidad	Al día	Totales	%	Regla
NORTE QUITO	C	25 – 35	8	47	55	14.55	No
CENTRO	C	36 – 40	1	33	34	3.03	Si
VALLES	C	36 – 40	12	62	74	19.35	No

Aplicamos los 4 pasos anteriores en los cuadros que van generándose y para cada uno de los atributos, de los cuales el que más ganancia de información tenga, será el seleccionado para nodo de ese nivel.

ITERACIÓN 1: En primer lugar, se calculó la entropía total (Nivel 1) de todos los atributos y en este caso se obtiene un valor de 0,98; por lo que en la Tabla 32 se reflejará el cálculo de todos ellos. Dado esta primera parte importante para cada iteración, tomamos el atributo sector para el que consideramos su probabilidad de ejecución según el número de registros que cumplan esta condición en lo que se encuentra que en el sector norte con 48%, sector centro 16%, valle 25% y otros 12%. Con esta información se procedió al cálculo de la entropía del atributo sector con las probabilidades obtenidas de ocurrencia generando una entropía de 0.46 y con este valor aplicando la ganancia de información se obtiene que para Sector es de 0.518. A continuación, se calculó la entropía para el atributo

Zona_Venta la cual llega a un valor de 0.810 y con la que se obtiene una ganancia de información de 0.168; de igual forma para edad se tiene una entropía de 0.92 la que nos lleva a una ganancia de 0.054 y por último con el atributo Estado-Civil se tiene una entropía de 0.89 con ganancia de información de 0.083.

Entropía Total: 0,98

Tabla 39. Resumen de Cálculos Primer Nivel C.C.A.

ATRIBUTO	Entropía ()	GANANCIA INFORMACIÓN
ZONA VENTA	0.810	0.168
EDAD	0.92	0.054
SECTOR	0.46	0.518
ESTADO CIVIL	0.89	0.083

De este primer barrido con todos los atributos, observamos que obtenemos la mejor ganancia de información con el nodo Sector, por lo que en la Figura 23, se muestra que dicho nodo quedará en la parte inicial del árbol.

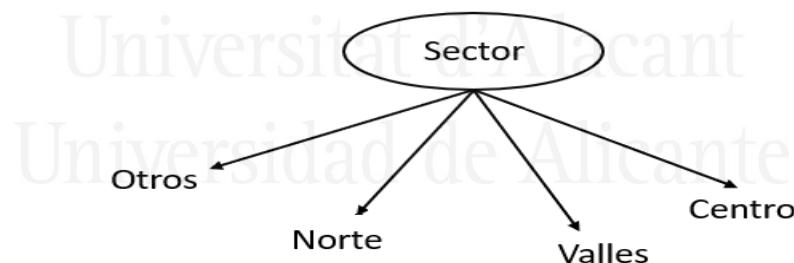


Figura 24. Nodo de Primer Nivel Árbol de Decisión. Caso C.C.A.

ITERACIÓN 2: Obtenido ya el primer nivel, repetimos el procedimiento para la rama Norte, pero esta vez sin tomar en cuenta el nodo Sector, y en la Tabla 40 se despliega que el nodo Zona-Venta brinda mayor ganancia de información para el segundo nivel, comparado con los datos de Edad y Estado_Civil.

Entropía Total (Nivel 2): 0,57

Tabla 40. Resumen de Cálculos Segundo Nivel C.C.A.

ATRIBUTO	Entropía ()	GANANCIA INFORMACIÓN
ZONA VENTA	0.19	0.38
EDAD	0.29	0.284
ESTADO_CIVIL	0.22	0.357

De esta forma en la Figura 24, se genera el segundo nivel del árbol de decisión para este caso de estudio con el nodo Zona_Venta.

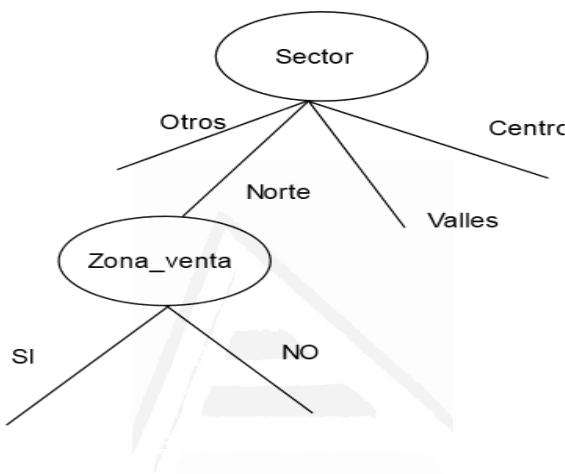


Figura 25. Nodo de Segundo Nivel Árbol de Decisión Caso. C.C.A.

ITERACION 3: Al momento contamos con dos niveles, repetimos este proceso en forma recursiva para la rama SI de zona de campaña, pero esta vez sin tomar en cuenta Sector y Zona_Venta, y vamos a obtener como se muestra en la Tabla 41, los valores de entropía de 0.29 con ganancia de 0.2849 para Edad y una entropía de 0.35 con 0.2267 de ganancia para Estado-Civil.

Entropía Total (Nivel 3): 0,57

Tabla 41. Resumen de Cálculos Tercer Nivel C.C.A.

ATRIBUTO	Entropía ()	GANANCIA INFORMACIÓN
EDAD	0,29	0.2849
ESTADO_CIVIL	0,35	0.2267

Con este resultado se definiría que el tercer nivel estaría compuesto por el nodo correspondiente a Edad, luego del cual por la rama [36 – 40] iría el último nodo que es Estado Civil; de esta forma la totalidad del árbol estaría estructurado de la siguiente forma:

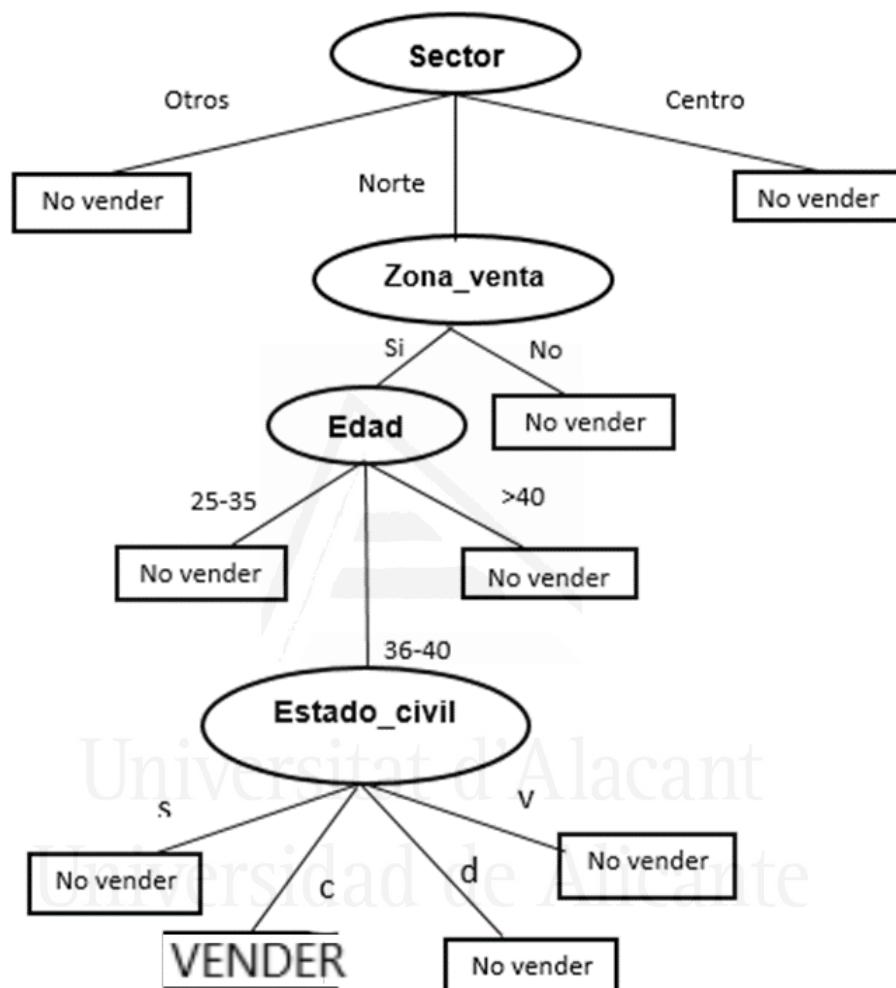


Figura 26. Nodo de Tercer Nivel Árbol de Decisión. Caso C.C.A.

5: Prueba de Datos: De las pruebas realizadas sobre el árbol, como se ve en la Figura 27. se concluye que de un total de 3.133 socios analizados en los sectores propuestos, 1836 de ellos que equivalen al 59% de la muestra deberán seleccionarse para la campaña de venta de membresías del Club Castillo de Amaguaña, lo cual gráficamente se presenta en la siguiente Tabla.

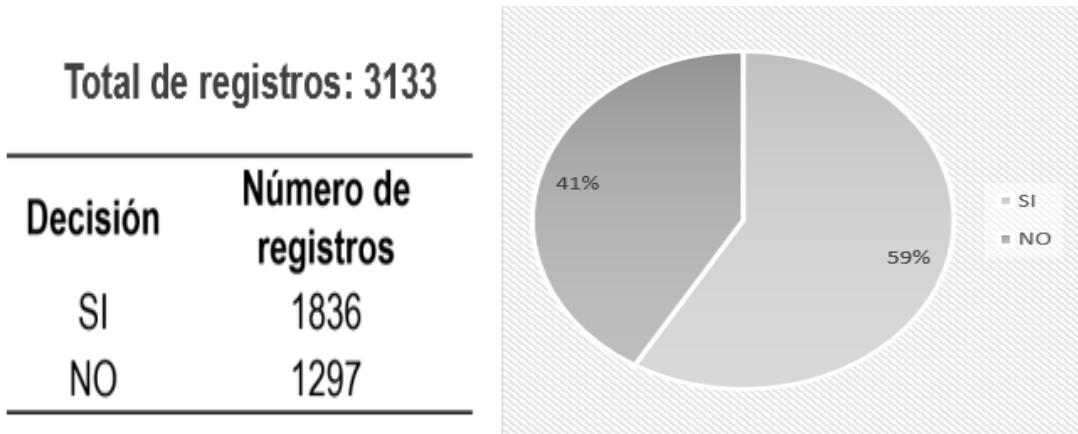


Figura 27. Resultado Caso de Estudio C.C.A.

5.2.2. DISCUSIÓN DE RESULTADOS

Para la validación dentro de este caso de prueba de M3S, se usaron las mismas encuestas que dentro del CLUB se obtiene que por ser una empresa que la hemos catalogado como medianamente grande, en el sentido de que maneja como presupuesto anual \$.2.000.000.- tiene también establecidos sus objetivos de manera clara y medible, no cuentan con una herramienta que les permita medir su gestión, tampoco con metodologías de desarrollo, pero si tienen un presupuesto de tecnología definido el cual se lo ejecuta sin problema. En la parte de proyectos se han cumplido los tiempos, no manejan estándares de recolección de datos, sin embargo, y el control de cambios no es nada formal. La empresa dentro del área de tecnología no tiene una certificación de calidad.

Comparativamente analizando la metodología usada en el Club y la de M3S, haciendo un análisis cualitativo obtenemos lo siguiente: Con M3S se considera que el retorno de la inversión sería más efectiva que la actual, la aceptación fue mayor por la misma capacidad de manejo de los miembros de tecnología.

Por último, dentro de lo posible cuantitativamente se obtiene que en personal se usaron los mismos recursos, pero tomando en cuenta que empezaron a manejar mayor información y generando nuevo conocimiento para la institución, así como una disminución importante en el tiempo de construcción total del proyecto, el tiempo en extracción de datos se definió en por lo menos un 40%, no se requirió ninguna inversión en hardware, se aumentaron 4 usuarios de nivel de jefaturas como usuarios nuevos de BI; en este período se han requerido otros proyectos para el área deportiva y financiera; en consecuencia los costos inherentes al proyecto bajarían en un 50%.

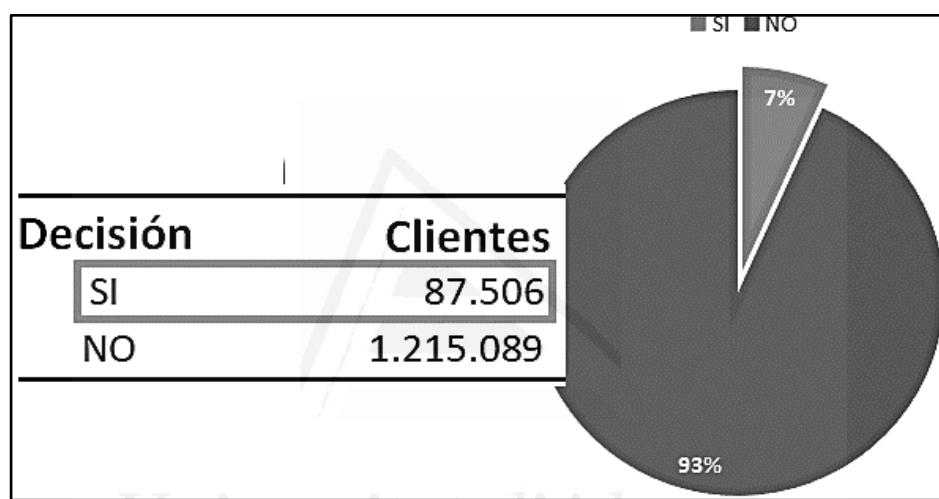


Figura 28. Resultado Caso de Estudio EEQ.

5.2.3. CONCLUSIÓN CASO C.C.A.

Tras haber realizado el proceso metodológico de M3S dentro del área de Sistemas del Club y haber comparado con metodologías realizadas para el mismo fin se puede concluir que:

- Aplicando M3S dentro de este caso de estudio se obtuvo un rendimiento satisfactorio tanto en el factor humano como en tiempos de ejecución, lo que reduce costos en la realización del proyecto de manera significativa.

- A partir de este proyecto se puede dar un inicio a la producción de proyectos BI que soporte necesidades de negocio en cada una de las diferentes áreas administrativas del Club.
- Una vez explorada la base de datos de cobranzas, se podría analizar las demás bases utilizadas en esta organización.
- Al aplicar el algoritmo de decisión se obtiene buenos resultado, lo que indica que el haber incluído una modificación en ID3, aportó con mejores resultados.
- Los atributos escogidos por los responsables de este proyecto fueron correctos, con lo que se logra obtener un muy buen nivel de aprendizaje.

Al realizar este estudio se abrió una gran gama de información para fortalecer la toma de decisiones en el Club y fomentar la creación de proyectos de BI que contemplen cualquier necesidad de negocio en la organización.



5.3. TEST DE RESULTADOS

Para la prueba de nuestro modelo y metodología, como se indicó en el capítulo anterior, se lo aplicó en dos casos de diferente tamaño de empresa, una grande como es la Empresa Eléctrica Quito y una mediana como el Club Castillo de Amaguaña, dentro de los cuales en la Tabla 42, hacemos un resumen de lo que realizan cada una de ellas.

Tabla 42. Empresas de Estudio para M3S.

E.E.Q.	CASTILLO DE AMAGUAÑA
Descripción. - Provee el servicio público de generación, subtransmisión, distribución y comercialización de energía eléctrica a la Ciudad de Quito, siendo una de las empresas modelos de este servicio en todo el país. Actualmente cuenta con 12 millones entre todo su grupo de clientes (Grandes, residenciales, convenios fijos) y funciona en un total de 20 agencias de la Provincia de Pichincha.	Descripción: Tiene prestigio a nivel nacional, desde hace 45 años. Ofrece a sus socios, familiares e invitados, la más grande infraestructura, naturales y acogedoras; que sin perder su estilo medieval hacen que sus visitantes experimenten momentos inolvidables en cada uno de sus espacios. El deporte, la recreación se comparte en las 23Ha que se encuentran distribuidos en tres etapas. Las diversas actividades, el compromiso de sus autoridades, colaboradores y trabajadores fidelizan semana a semana la visita de sus socios.
Objetivo. - Identificar a los clientes en las distintas parroquias del Distrito Metropolitano de Quito que excedan el consumo promedio en el sector Residencial, a quienes el gobierno está interesado en focalizar una campaña de venta masiva a bajos costos de focos LED.	Objetivo. - Implementar una solución de Inteligencia de Negocios, la cual apoye en la toma de decisiones gerenciales mediante indicadores generados en base a los datos obtenidos del Departamento de Cobranzas utilizando el software Qlik Sense

De todas las actividades realizadas, la interpretación en los dos casos de los resultados fue similar en los equipos de trabajo, aunque cada uno había tenido sus propios inconvenientes; sin embargo, se llegó a un consenso en el uso de la metodología siendo compartida y aceptada, con las siguientes consideraciones:

- Se gestiona con éxito en las empresas la prueba de la metodología, por lo que se considera validada como innovación en el área; sin embargo, en las empresas grandes necesita un compromiso más alto y respaldado de los grupos directivos.
- El personal involucrado trabajó sin mayores problemas con M3S, una vez que asistieron a las charlas de inducción. Estimamos que para el personal de diferentes áreas que estuvieron en el aprendizaje de M3S no representa ningún

inconveniente, más aún para personal del área informática que maneja con mayor solvencia este tipo de proyectos.

- El limitado tiempo y en algunos casos la falta de apoyo debido a la carga laboral diaria de trabajo del personal involucrado no permitió culminar con todo éxito algunos pasos, como la documentación completa de formularios; pero estamos seguros que M3S puede aportar en un desempeño más eficaz en los proyectos de la empresa. Debido a esto se refinaron las formas que se utilizan como soporte documental y las mismas constan como anexos en esta tesis.
- M3S, deja latente una mejora en la planificación estratégica a realizarse en las empresas, porque los temas de gestión se representarán más rápido y con mayor conocimiento dentro de la empresa sobre cualquier objetivo planteado.
- En lo que se refiere a costos, M3S ayudó debido a su agilidad en la simplificación de captura y limpieza de datos; y de igual forma en la administrativa con la documentación propuesta. Así garantizamos un menor costo favorable por la disminución de tiempos en la metodología y consecuentemente a favor de la rentabilidad de las empresas en los proyectos.
- En conclusión, podemos afirmar que el proceso de validación de la pertinencia y aplicabilidad de M3S, tuvo éxito desde la experiencia de los usuarios.

5.4. MEDICIÓN DE RESULTADOS.

El objetivo de esta parte de la investigación es obtener información sobre la percepción de M3S en las empresas estableciendo características y parámetros que determinan la satisfacción del uso de la metodología, para esto adicionalmente a las dos empresas en las que se validó la metodología se encuestaron a otras en las que pudimos hacer proyectos de menor tamaño para que se verifique la utilidad y efectividad de M3S.

Tabla 43. Empresas Encuestadas

CARGO SECTOR	EMPRESA	GERENTES	JEFES DEPART.	ING.DESARROLLO
PUBLICO	3	2	3	14
PRIVADO	8	5	5	18
TOTALES	11	7	8	32

Según este resumen se tiene un total de 47 usuarios repartidos en todo este grupo los mismos que tienen funciones diferentes pero en su mayor parte en el área de informática, para quienes se desarrolló una herramienta de medición como las encuestas (Grande & Elena, 2005) las mismas que se definieron con variables del estudio que consideramos medibles, luego realizamos el tratamiento de esos datos y los valoraremos en base a cálculos estadísticos simples.

Las respuestas de cada encuesta representan el consenso de todo el equipo de trabajo en cada empresa, en ese sentido vamos a explicar cada uno de los resultados obtenidos y a continuación lo haremos con las 14 preguntas de la encuesta de Madurez Empresarial en Proyectos BI que la tenemos como ANEXO XI. En la encuesta indicada se han planteado preguntas Cerradas (fáciles de responder y las más usadas, que pueden ser dicotómicas si tienen solo dos respuestas o de elección múltiple como se las planteó en este caso).

En la pregunta 1 cuyo resultado se ve en la Figura 28, la misma que se refiere a que si la organización tiene objetivos claros y medibles, se tuvo que el 91% se fijan metas claras y miden sus resultados, mientras un 9% mantiene una ambigüedad en la medición y obtención de metas.

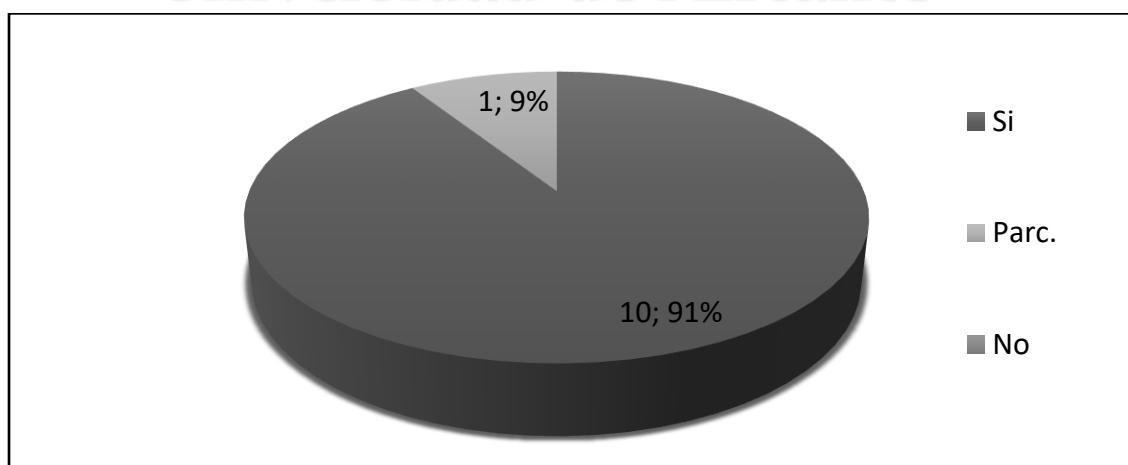


Figura 28. Pregunta 1 – MADUREZ-Objetivos

En la pregunta 2, la cual vemos como resultado en la Figura 29, busca identificar si la eficiencia en costo, calidad y tiempo se han cumplido en proyectos anteriores, con esto se trata de verificar la formalidad con la que se controlan los procesos de manera que un 9% usan herramientas para dar control al seguimiento de metas, mientras que un 82% lo realiza parcialmente y en la misma medida un 9% no ha cumplido con las metas planeadas.

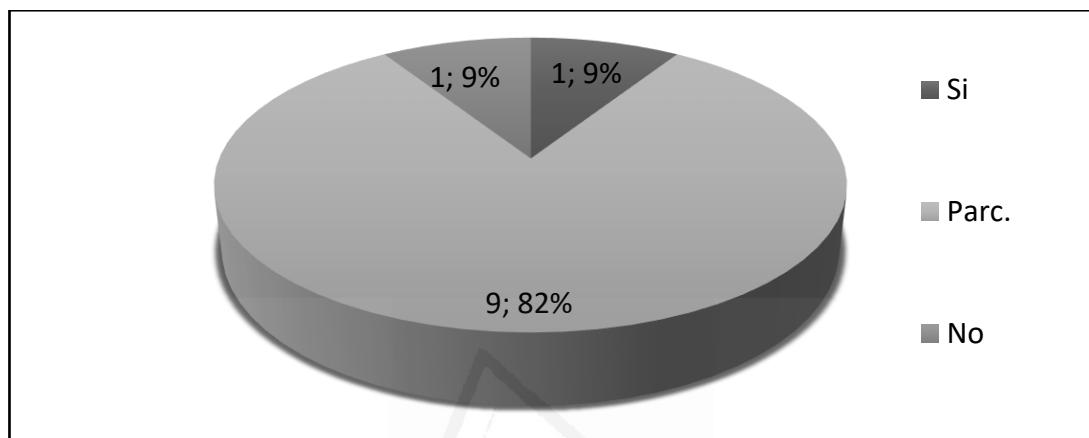


Figura 29. Pregunta 2 – MADUREZ-Eficiencia

En la pregunta 3 que se obtiene respuestas en la Figura 30, identifica si su organización utiliza estándares para la recolección de datos, buscando parámetros para tomar datos desde cualquier fuente de información, teniendo que un 36% manejan un estándar o formato definido, un 46% realiza la recolección de datos más no sigue un proceso estandarizado y finalmente un 18% no realiza la recolección de datos.

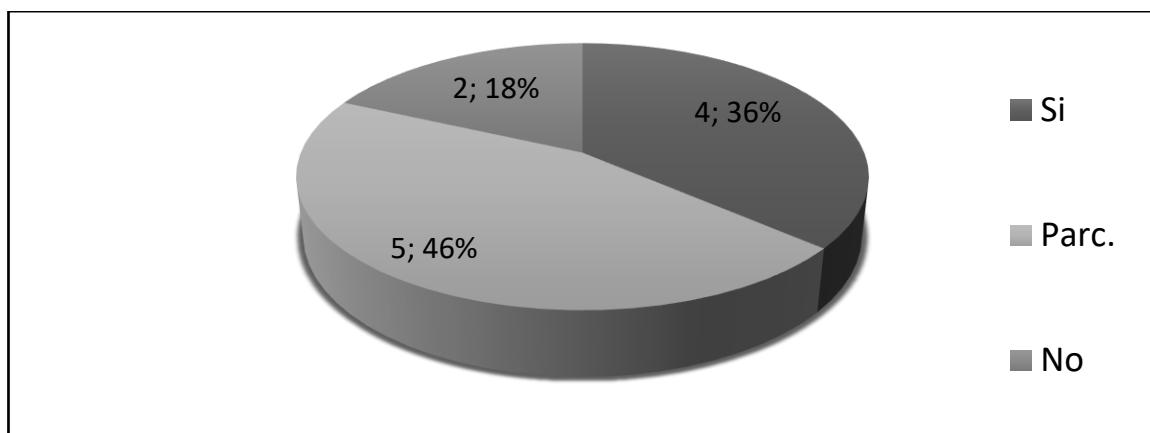


Figura 30. Pregunta 3 – MADUREZ-Estandar Recolección

La pregunta 4 muestra valores en la Figura 31, ésta busca determinar el uso de estándares externos para medir y mejorar su gestión, el fin de la misma es cotejarnos con empresas que mantengan este tipo de normas de lo que obtuvimos que un 9% realiza estas mediciones, un 64% realiza estas comparaciones sin un seguimiento adecuado y un 27% no lo realiza.

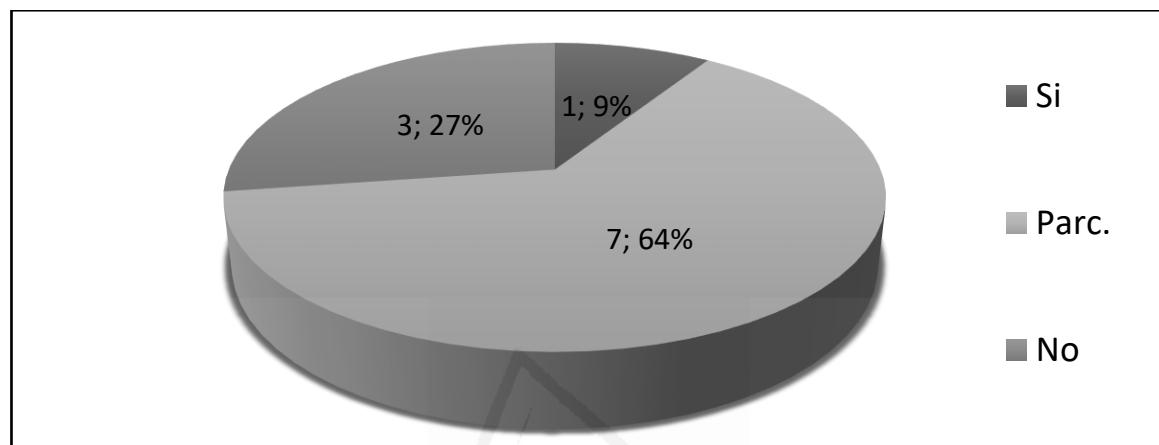


Figura 31. Pregunta 4 – MADUREZ-Estandar Externo

En la Figura 32 se representa la pregunta 5, que trata de identificar si su organización utiliza o tiene hitos definidos, para medir la gestión por áreas con indicadores de gestión; teniendo así que un 18% usan y manejan de manera adecuada indicadores de consecución de hitos, un 64% manejan hitos, pero no se da un seguimiento adecuado a los mismos, 18% no maneja hitos ni miden la gestión por áreas.

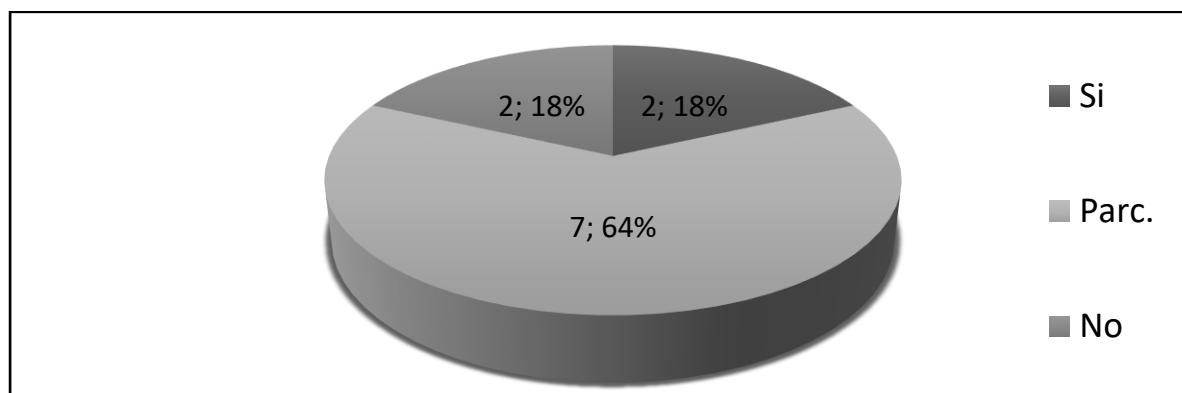


Figura 32. Pregunta 5 – MADUREZ-Hitos.

En la pregunta 6 cuyo resultado se ve en la Figura 33, intenta determinar si su organización utiliza y mantiene un marco de referencia común, metodología y procesos de gestión para todos sus proyectos, con el fin de buscar alguna metodología específica dentro de la organización. Se tiene que apenas un 18% de manejan una metodología, un 27% maneja parcialmente una metodología y más de la mitad (55%) no maneja una metodología.

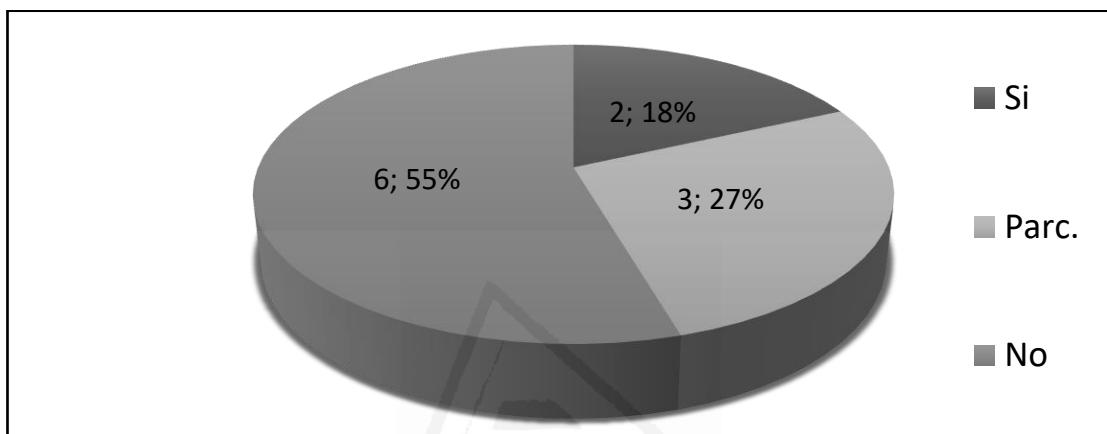


Figura 33. Pregunta 6 – MADUREZ-Metodológicos

En la pregunta 7 de la cual su resultado se ve en la Figura 34, determina si su organización cuenta con alguna certificación de estándares de calidad. Encontramos que un 36% de las organizaciones cuenta con una certificación, un 18% están en capacitación o en obtención de alguna mientras que un 46% no cuenta con certificación alguna.

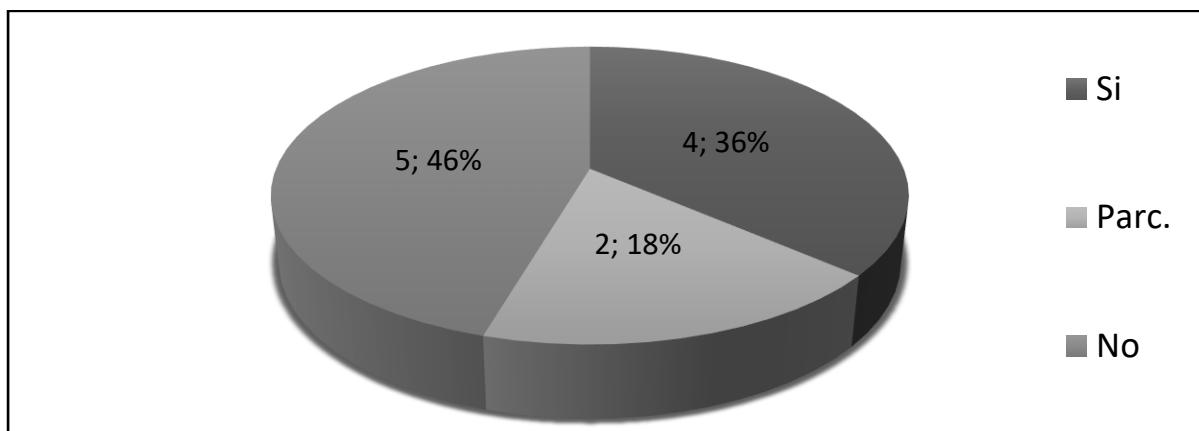


Figura 34. Pregunta 7 – MADUREZ-Certificación.

En la pregunta 8 reflejada en la Figura 35, se busca determinar si la organización presupuesta inversiones en el área tecnológica anualmente, determinando que un 46% asigna un capital fijo, un 45% asigna un capital no exclusivo al área tecnológica y un 9% no asigna un capital para la inversión en el área tecnológica.

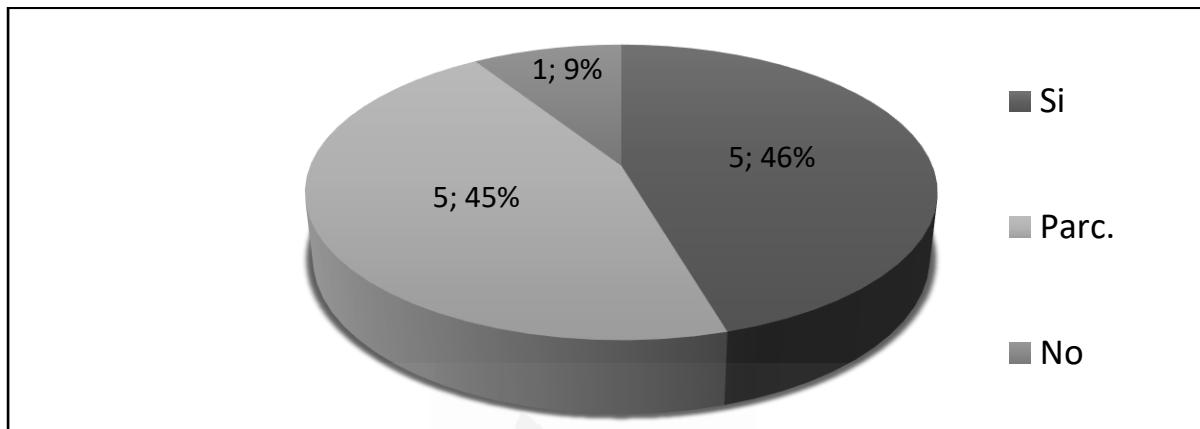


Figura 35. Pregunta 8 – MADUREZ-Inversión.

En la Figura 36 que representa la pregunta 9, determina cómo implementa su organización la retroalimentación de sus proyectos para la mejora continua, con el fin de determinar si se tiene este tipo de procesos. Los resultados muestran que un 18% de las organizaciones utiliza la experiencia obtenida en proyectos anteriores mientras que un amplio 82% no usa la experiencia obtenida en proyectos previos.

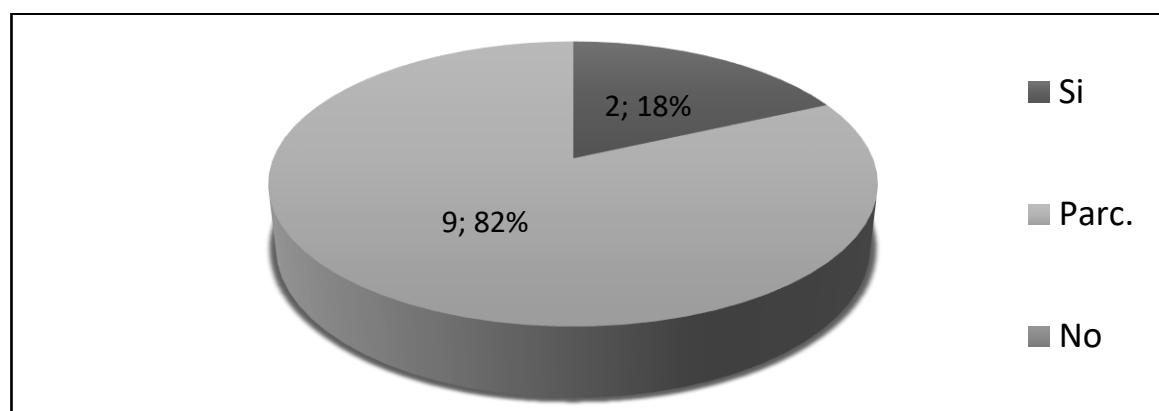


Figura 36. Pregunta 9 – MADUREZ-Retroalimentación.

En la pregunta 10 presentada en la Figura 37, buscamos si se usa algún tipo de metodología de dirección de proyectos de BI, con lo que un 27% usan una metodología o un direccionamiento enfocado en BI, un 55% tiene una noción del uso de BI mientras que un 18% no usa ni administra el BI.

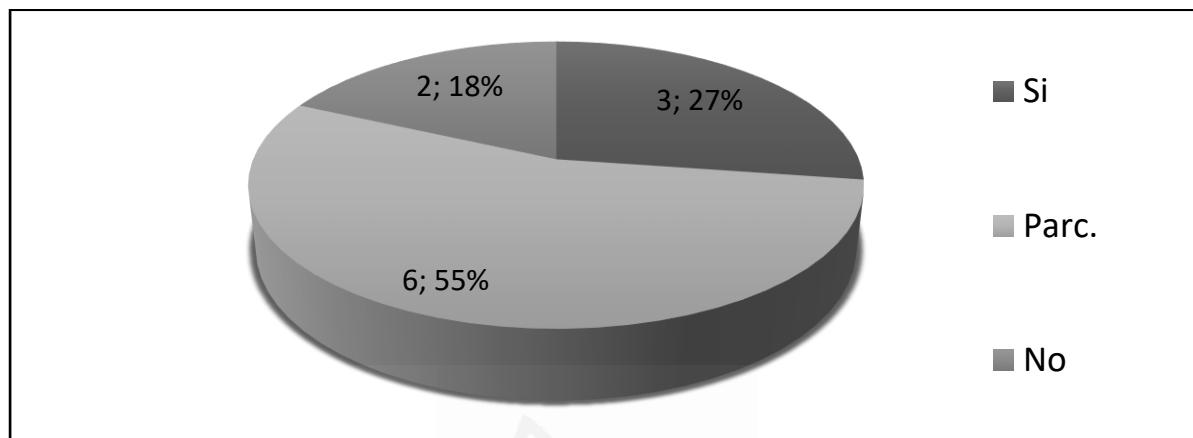


Figura 37. Pregunta 10 – MADUREZ-Uso Metodologías.

En la pregunta 11 del cual su resultado está representado en la Figura 38, la misma que responde a la incógnita ¿Si la aprobación de un plan de proyecto contempla su propio presupuesto?, de lo cual se determinó que un 18% manejan planes de acción que deben ser examinados, estudiados y aprobados por entes internos y/o externos para su ejecución, un 36% maneja a leves rasgos un plan y su aprobación mientras que un 46% no maneja planes de acción.

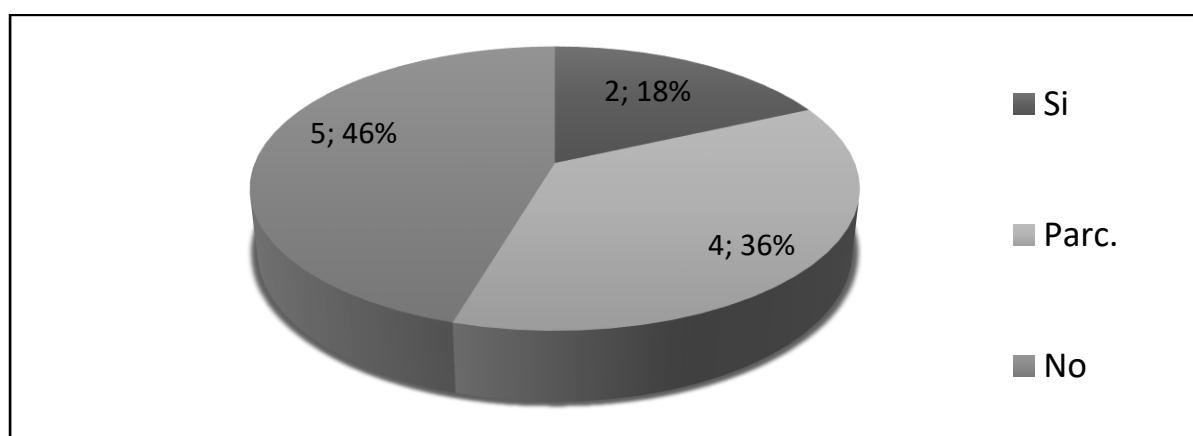


Figura 38. Pregunta 11 – MADUREZ-Presupuesto.

En la pregunta 12 cuyo resultado está representado en la Figura 39, responde a la incógnita de que si ¿Se formaliza administrativamente el control de cambio? Se encuentra que un 36% administra de manera formal los cambios, un 55% maneja de manera arcaica los cambios y un 9% no administra el control de cambios.

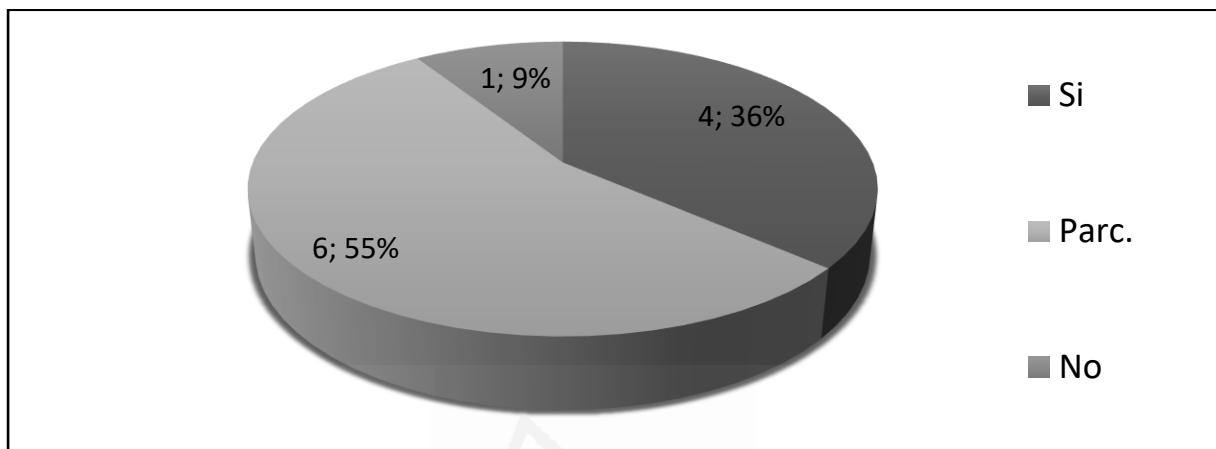


Figura 39. Pregunta 12 – MADUREZ-Cambios.

La pregunta 13 representada en la Figura 40, busca responder a la incógnita ¿Se maneja procesos de mejora continua en los proyectos?. Se tiene que un 18% maneja procesos y/o políticas de mejora continua, un 73% tienen leves nociones del manejo de las mismas y un 9% no lo hace.

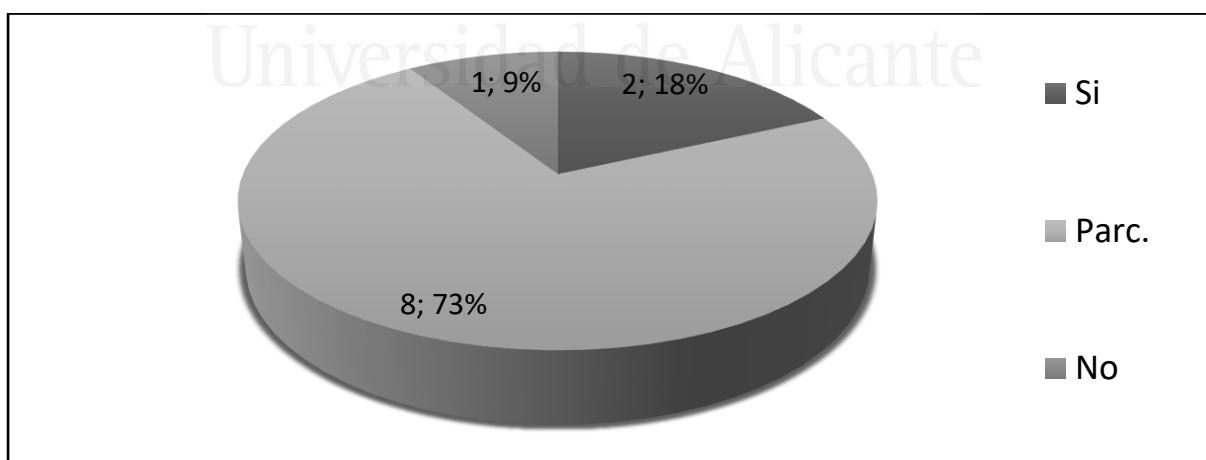


Figura 40. Pregunta 13 – MADUREZ-Procesos Mejora.

Por último, en la encuesta de madurez se representa en la Figura 41, la pregunta 14 que busca estándares de uso de software de dirección de proyectos en los distintos niveles de organización. Encontramos que un 27% maneja estándares, un 27% se guían por los mismos estándares, pero no los aplican a cabalidad y un 46% no maneja patrones o lineamientos.

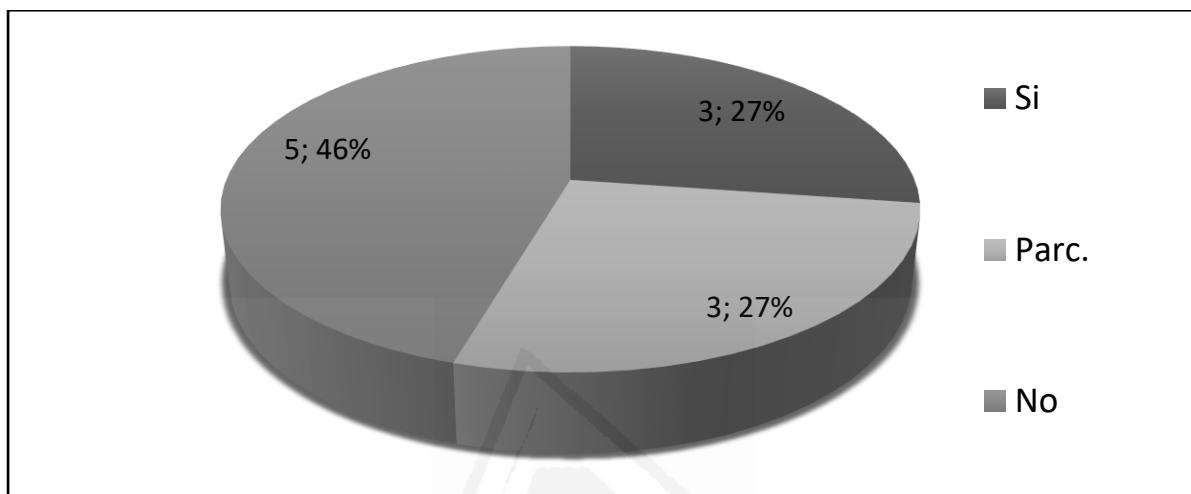


Figura 41. Pregunta 14 – MADUREZ-Estandar Software.

En cuanto a las encuestas comparativas, que constan como ANEXO XI y XII; en la parte cualitativa (CCL) se tienen 8 preguntas, 5 de ellas son cerradas, 2 son abiertas (se da al entrevistado la libertad para responder con sus propias palabras, son difíciles de ser tratadas estadísticamente, son útiles para estudios exploratorios cuando no se conoce mucho sobre las posibles respuestas) y una es Semiabierta (modalidad de pregunta cerrada en la que se añade una pregunta abierta).

En la pregunta 1 cuyo resultado se ve en la Figura 42, determina si existe algún control financiero en la organización desde la implementación de B.I. y a su vez para ver si la aplicación de la misma ha generado un beneficio económico dentro de la organización.

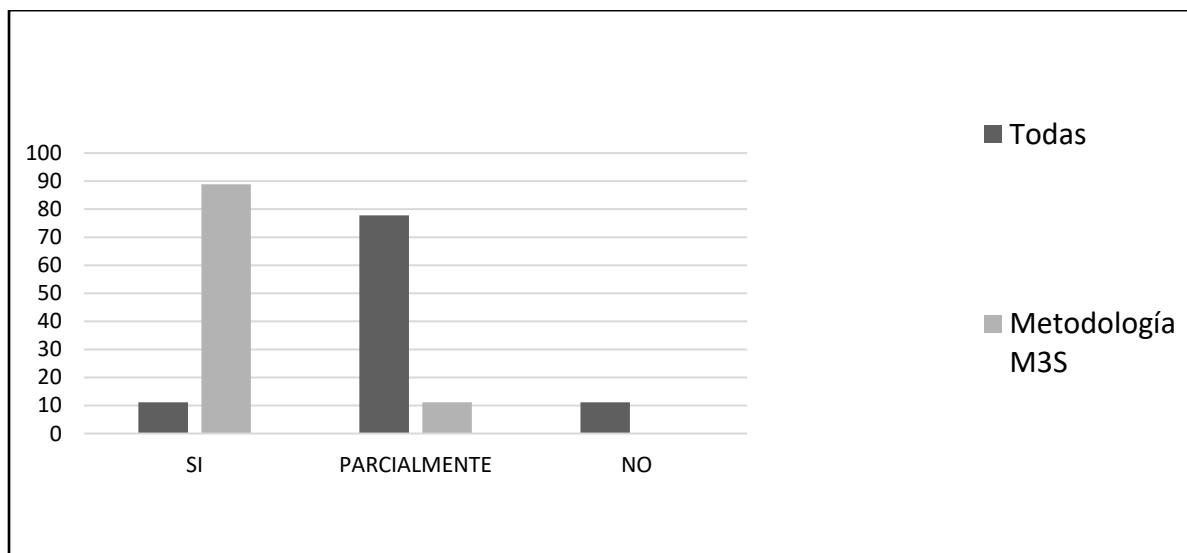


Figura 42. Pregunta 1-CCL-Control Financiero.

La pregunta 2 la cual observamos su resultado en la Figura 43, sirve para determinar si el personal encargado del área de B.I está a gusto con la herramienta y las soluciones implementadas, además si se da el hecho de que a mayor satisfacción se entiende un mejor manejo y comprensión de la misma.

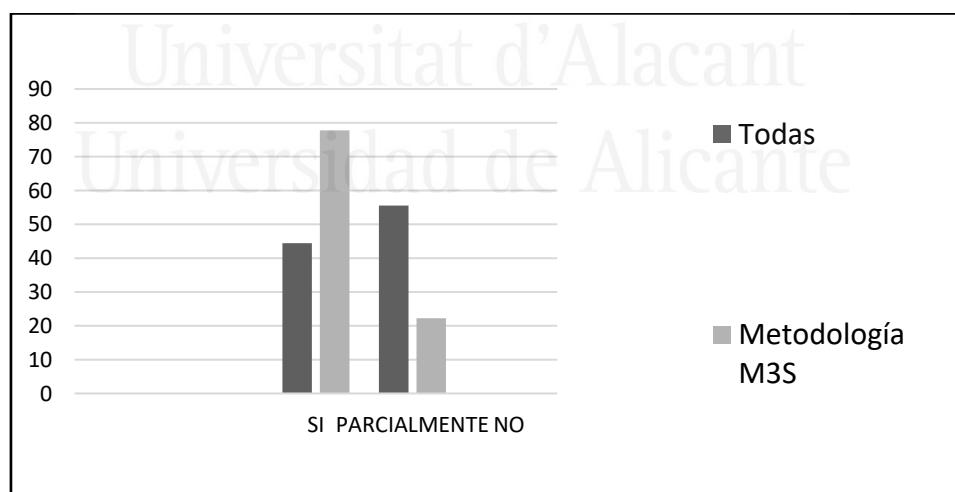


Figura 43. Pregunta 2-CCL-Percepción.

La pregunta 3 de la que vemos su resultado en la Figura 44, sirve para medir el nivel de aceptación de la clientela con la implementación de nuevas tecnologías B.I, es un buen

indicador para la usabilidad del sistema pues mientras mayor sea la aceptación hacia los clientes indica que su uso es sencillo, agradable y con una buena curva de aprendizaje.

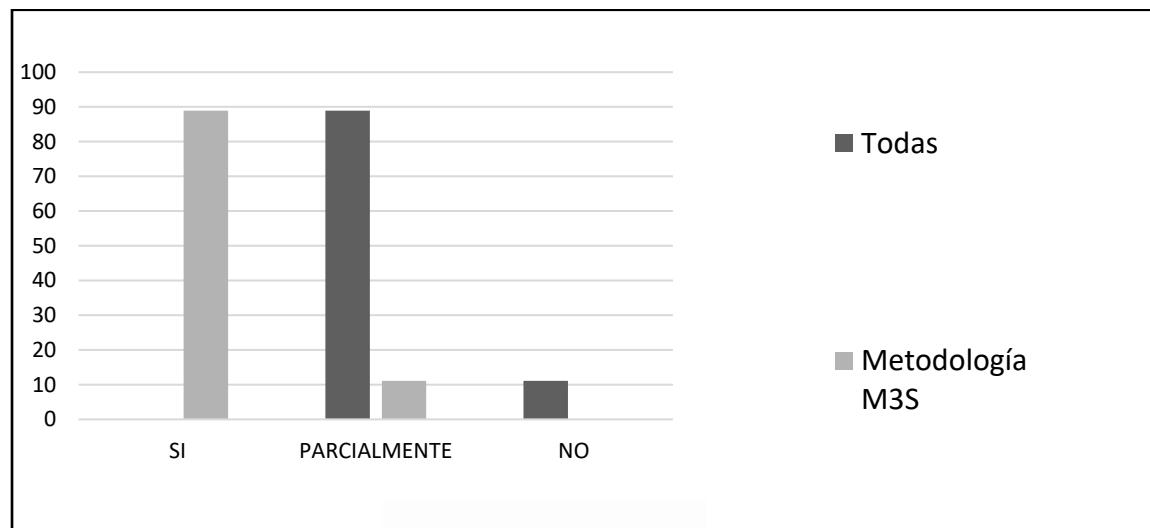


Figura 44. Pregunta 3-CCL-Aceptación.

En la pregunta 4 visualizando su resultado en la Figura 45, sirve para medir el nivel de acogimiento de la metodología B.I, así como también en cuanto a la complejidad del modelo utilizado y determina si la curva de aprendizaje de la organización va en buen camino.

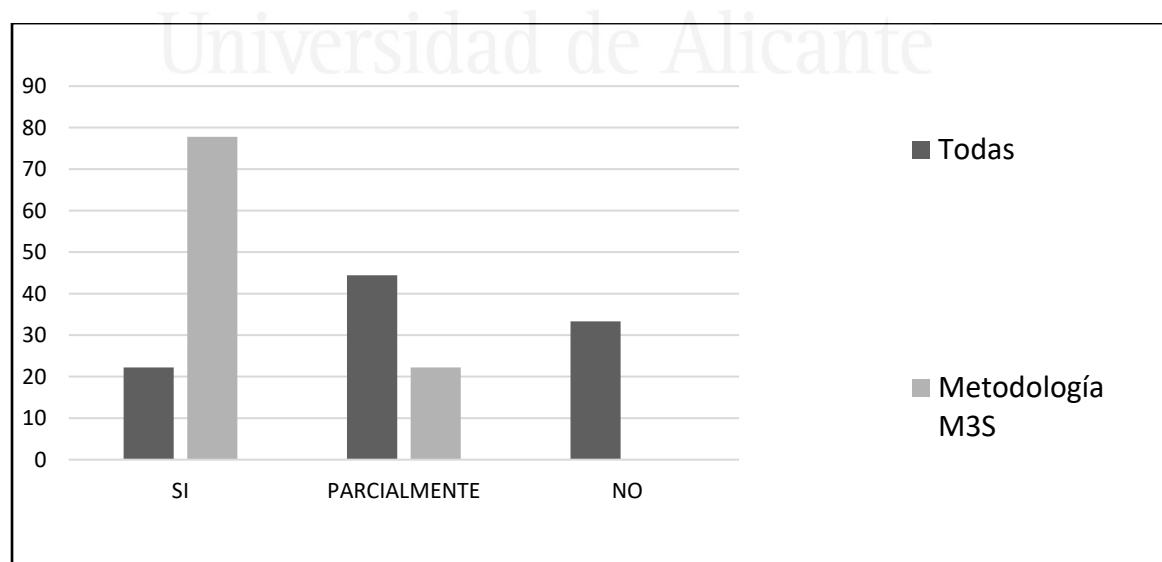


Figura 45. Pregunta 4-CCL-Acogimiento.

La pregunta 5 presentado en la Figura 46, sirve para medir si la metodología de B.I agrega valor en los elementos necesarios en los clientes y si provee de manera adecuada y efectiva al personal con las herramientas para la mejor administración de clientes y con ello obtener mayor ingreso de ellos.

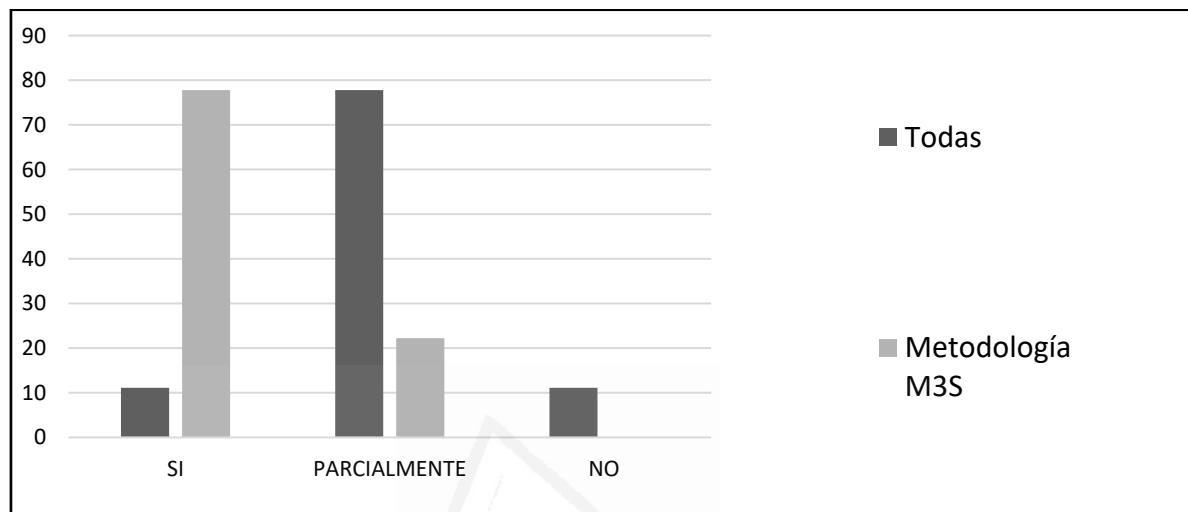


Figura 46. Pregunta 5-CCL. Valor.

La pregunta 6 se tiene representada en la Figura 47, busca identificar las debilidades encontradas por las organizaciones participantes de la encuesta; Destaca con un 43% la pérdida de tiempo por recalibración de procesos que son implementados en primera instancia bajo factores propios del momento, pero que al avanzar el tiempo varían y desencadenan en cambios.

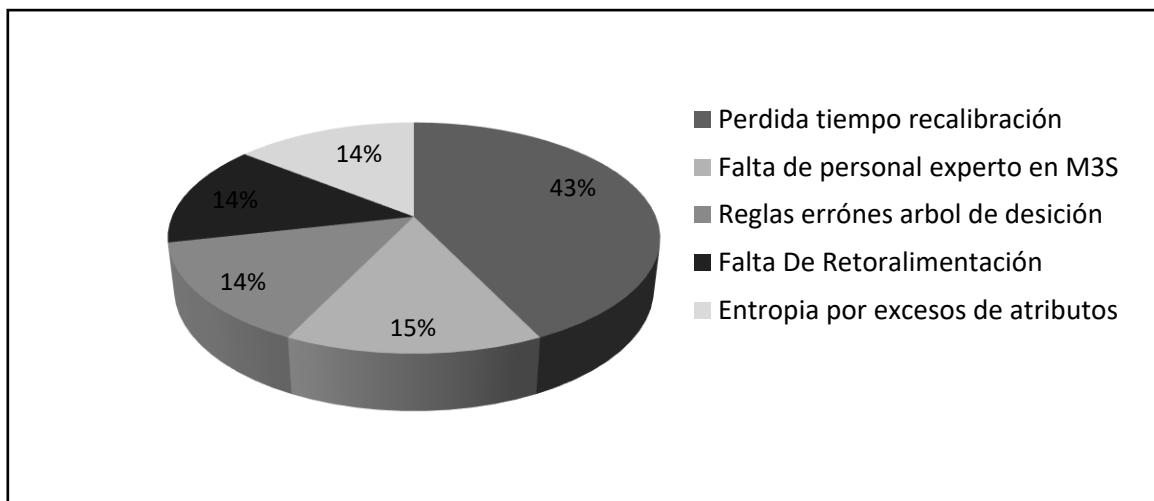


Figura 47. Pregunta 6-CCL-Debilidades.

En la Figura 48 que tiene inmersa a la pregunta 7, busca identificar las fortalezas de la metodología acorde a las empresas participantes en la encuesta; Destacan con 20% su versatilidad de implementación, la extracción de datos, reducción de tiempos, y su documentación ágil y simple; Esto desemboca en una optimización del proceso de implementación de B.I y demuestra que M3S es una metodología que saca ventaja sobre las metodologías usadas en la actualidad.

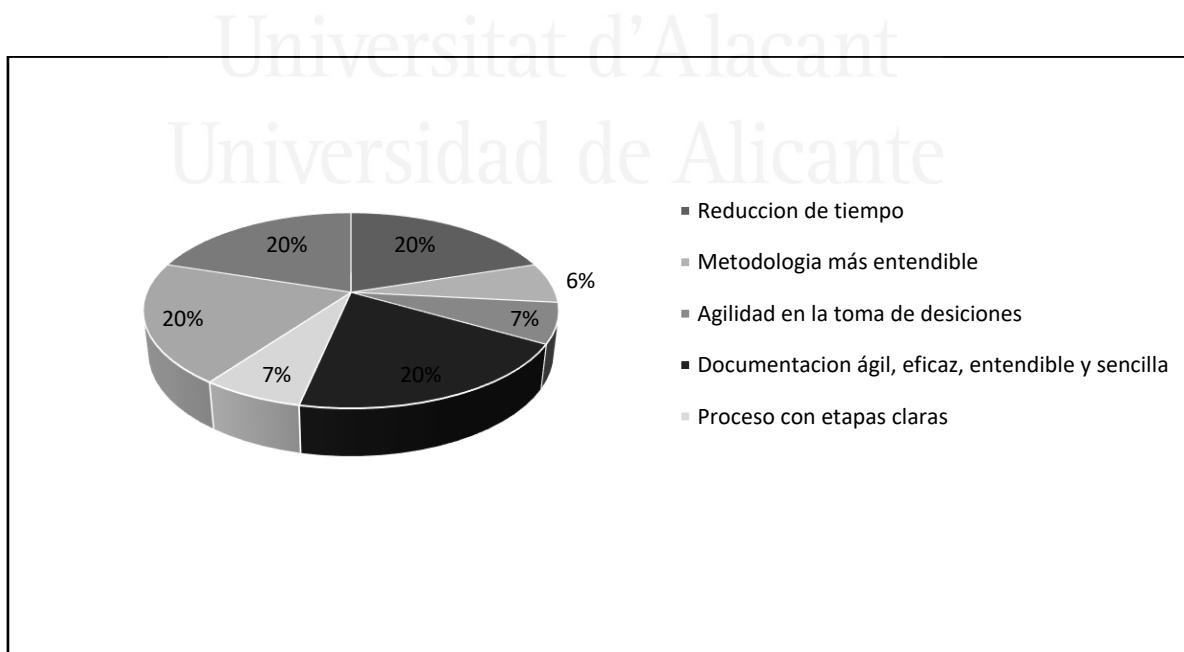


Figura 48. Pregunta 7-CCL-Fortalezas.

La pregunta 8 la cual muestra resultados en la Figura 48, busca determinar la robustez de la metodología y su posible aplicación como nueva política de B.I o como complemento para metodologías aplicadas previamente; De las empresas encuestadas el 60% de ellas estarían dispuestas a un cambio de metodología, mientras que un 40% no realizaría un cambio pues cuentan con políticas claras y bien definidas.

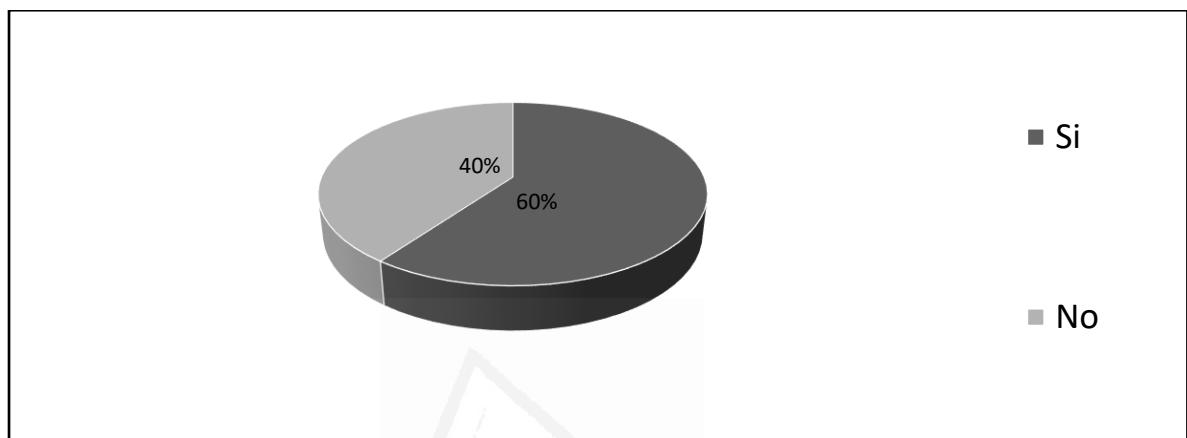


Figura 49. Pregunta 8-CCL-Robustez.

Para la parte cuantitativa (CCU) de las encuestas se tienen once preguntas con variables numéricas las mismas que se evalúan sobre M3S y la metodología usada en cada una de las empresas, a continuación, los resultados encontrados en cada una para el caso de la Empresa Eléctrica Quito.

En la pregunta 1 con resultados en la Figura 50, se colocó el número de personal requerido para la implementación en proyectos BI, en la cual se necesitarán 4 personas en la Metodología Chain Data Approach, mientras que en la metodología M3S se necesitarán dos personas máximo.

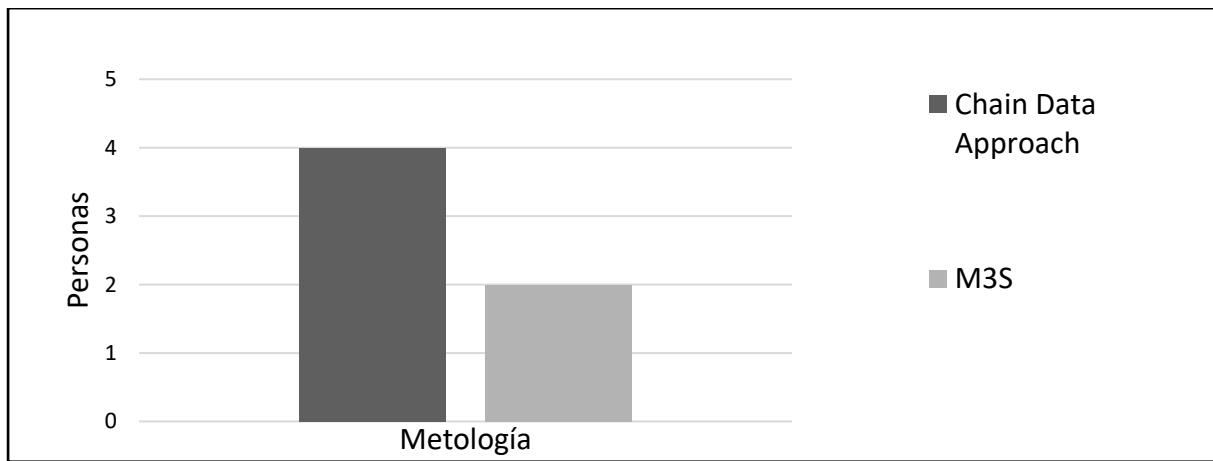


Figura 50. Pregunta 1-CCU-Personal.

En la pregunta 2 cuyo resultado se tiene en Figura 51, se determina el tiempo estimado para la extracción de datos, con un resultado de 240 minutos para la metodología Chain Data Approach, mientras que para la metodología M3S el tiempo estimado es de 120 minutos.

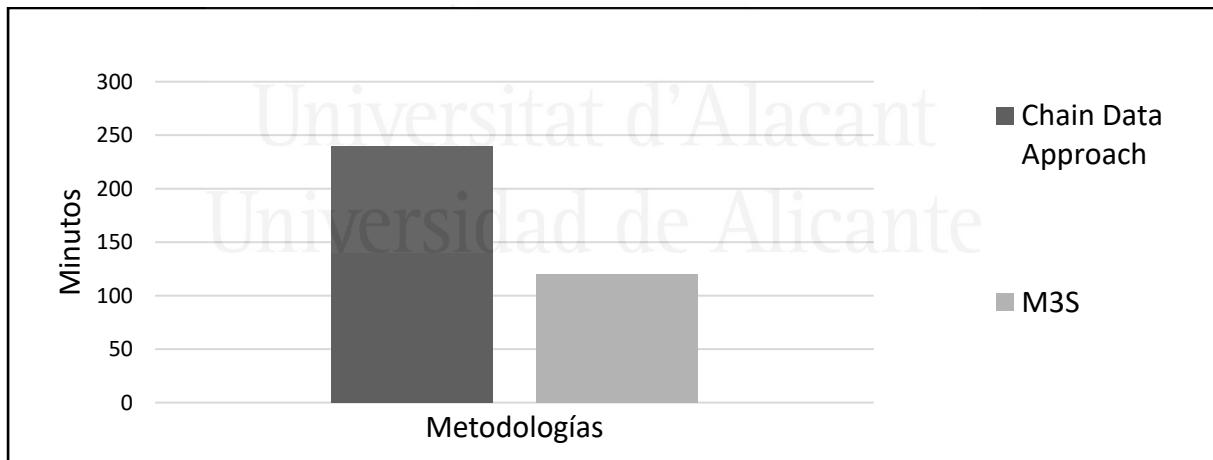


Figura 51. Pregunta 2-CCU-Extracción.

Para la pregunta 3 que se representa en la Figura 52, se muestra el tiempo estimado para la depuración y transformación de datos, en donde la metodología Chain Data

Approach requerirá de 30 días mientras que para la metodología M3S el tiempo estimado es de 17 días.

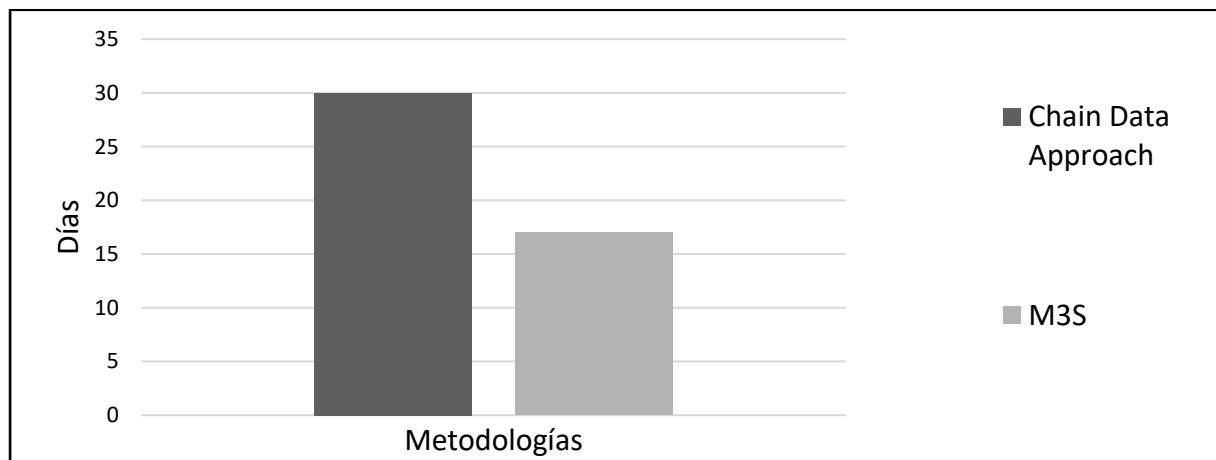


Figura 52. Pregunta 3-CCU-Depuración.

En la pregunta 4 cuyo resultado se representa en la Figura 53, se quiere saber el tiempo estimado para la construcción de la nube de datos, en el cual la metodología Chain Data Approach estima un tiempo de 15 días, mientras que para la metodología M3S 5 días.

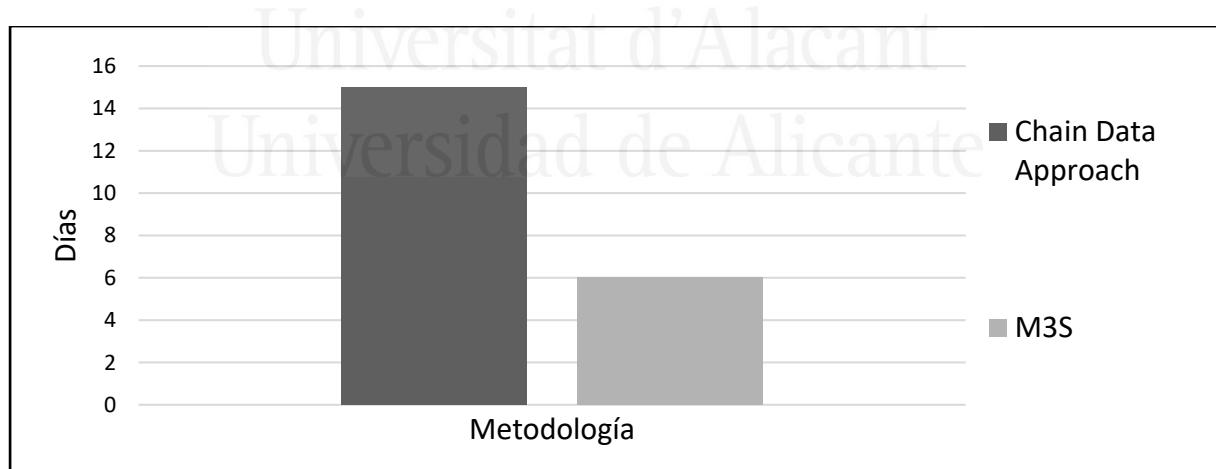


Figura 53. Pregunta 4-CCU-Construcción.

Dentro de la pregunta 5 cuyo resultado se despliega en la Figura 54, se requiere saber el tiempo estimado para la validación de datos. La metodología Chain Data Approach estima alrededor de 19 días, mientras que para la metodología M3S 5 días.

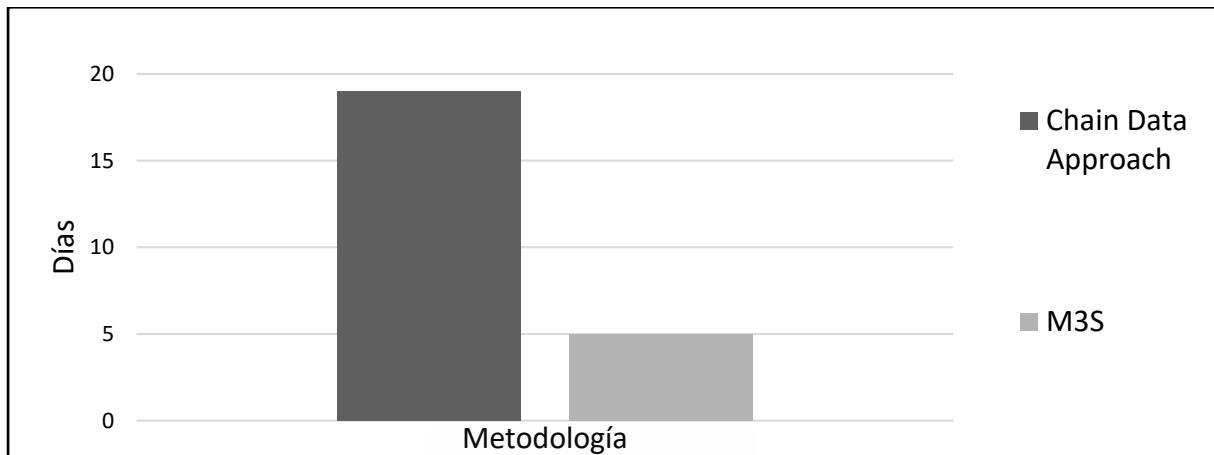


Figura 54. Pregunta 5-CCU-Validación.

En la pregunta 6 los resultados se representan en la Figura 55, con el costo del proyecto BI, \$.28000 dólares requerirá la metodología Chain Data Approach mientras que para la metodología M3S requerirá de \$.19000 dólares.

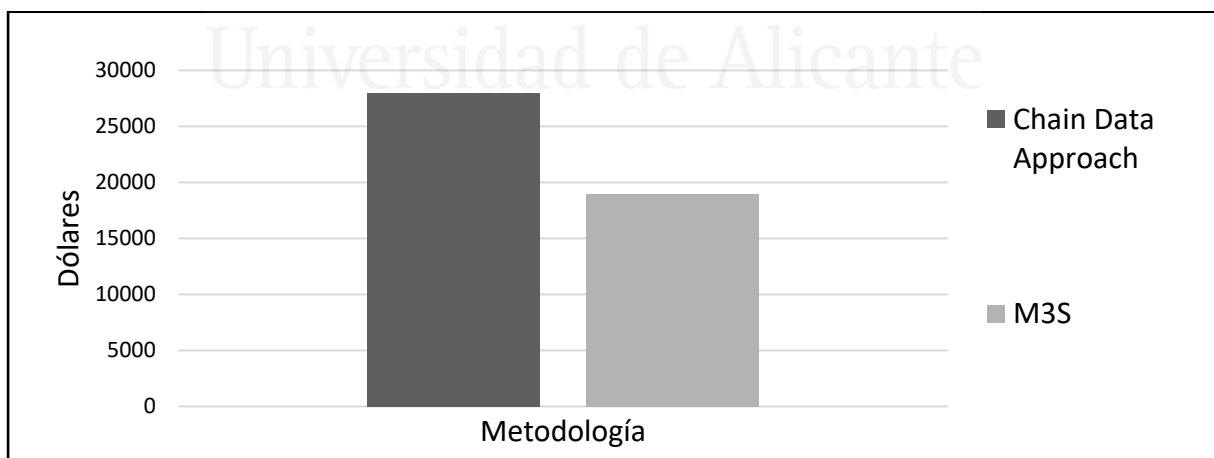


Figura 55. Pregunta 6-CCU-Costos.

En la pregunta 7 representada en la Figura 56 el costo de hardware del proyecto, en el cual se obtuvo que \$.20000 dólares necesita la metodología Chain Data Approach mientras que la metodología M3S de \$.15000 dólares.

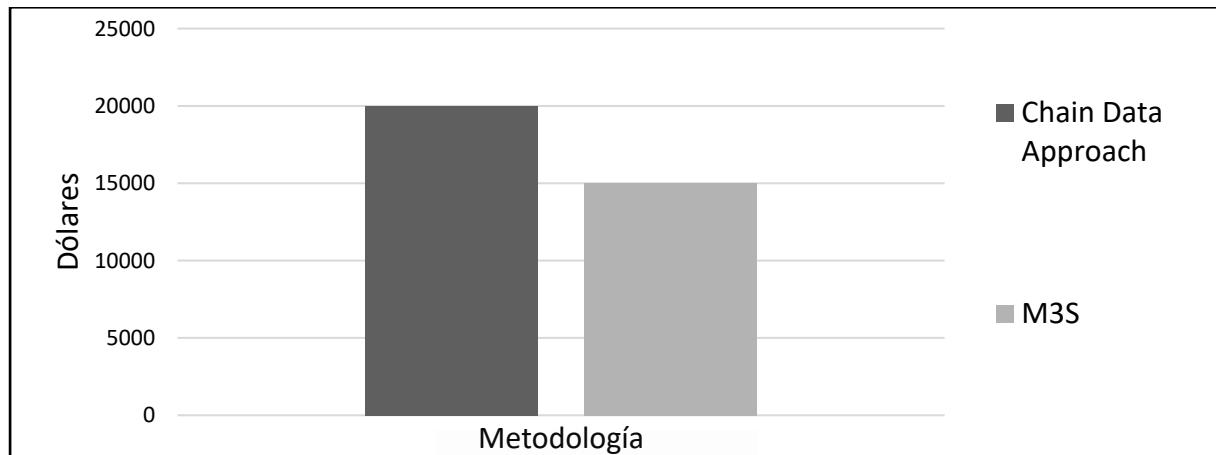


Figura 56. Pregunta 7-CCU-Costo Hardware.

La pregunta 8 resuelta en la Figura 57, representa el número de usuarios definidos para el uso de la solución BI, 5 personas requiere la metodología Chain Data Approach mientras que para la metodología M3S requiere de 15 personas.

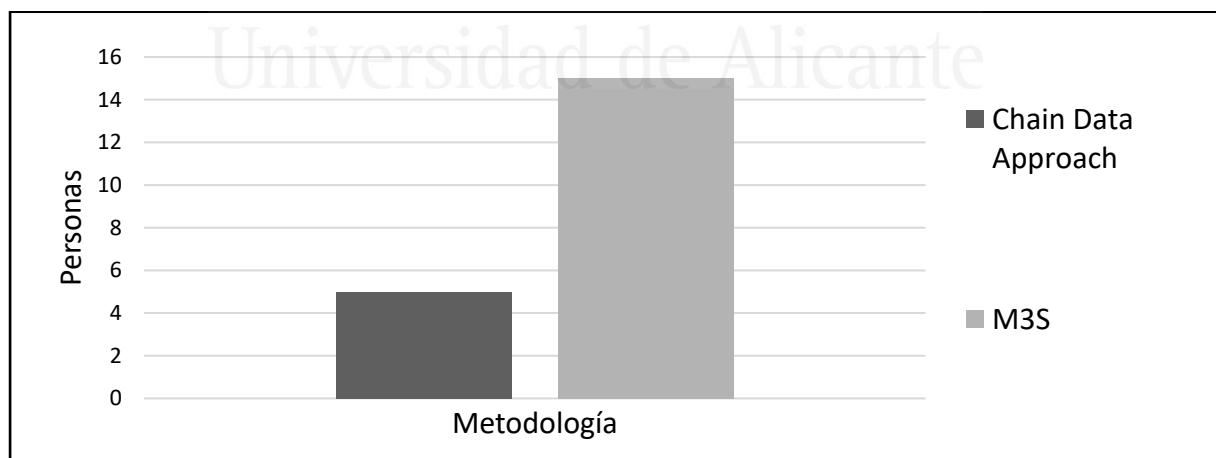


Figura 57. Pregunta 8-CCU-Usuarios

En la pregunta 9 que presenta los resultados en la Figura 58, acerca del porcentaje de usuarios activos en la solución BI, en el cual de 50% representan la metodología Chain Data Approach mientras que para la metodología M3S representa el 100%.

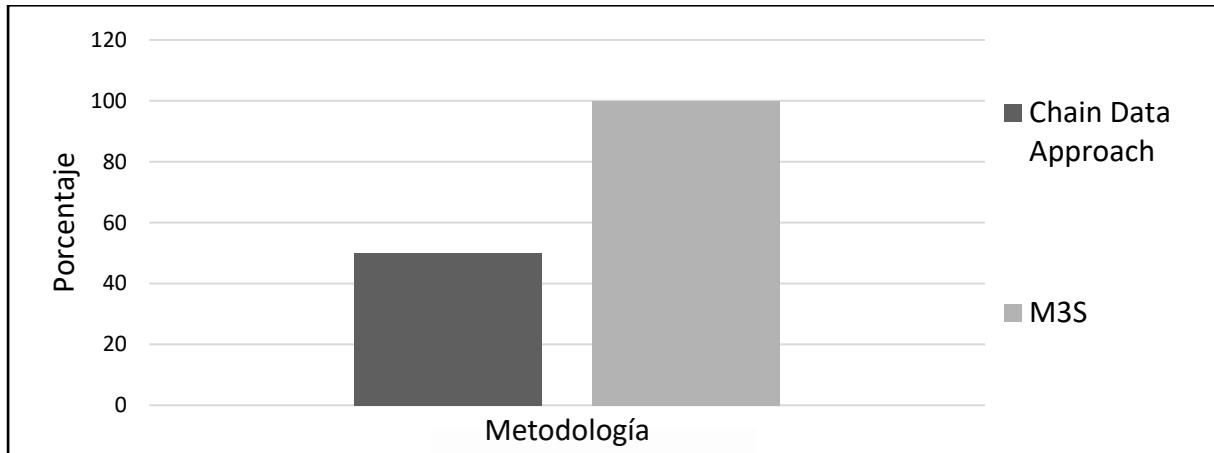


Figura 58. Pregunta 9-CCU-Usuarios Activos.

En la pregunta 10 cuya representación está en la Figura 59, donde se muestran los resultados del número de aplicaciones de BI. La metodología Chain Data Approach utiliza 2 aplicaciones mientras que para la metodología M3S utilizará 1 aplicación.

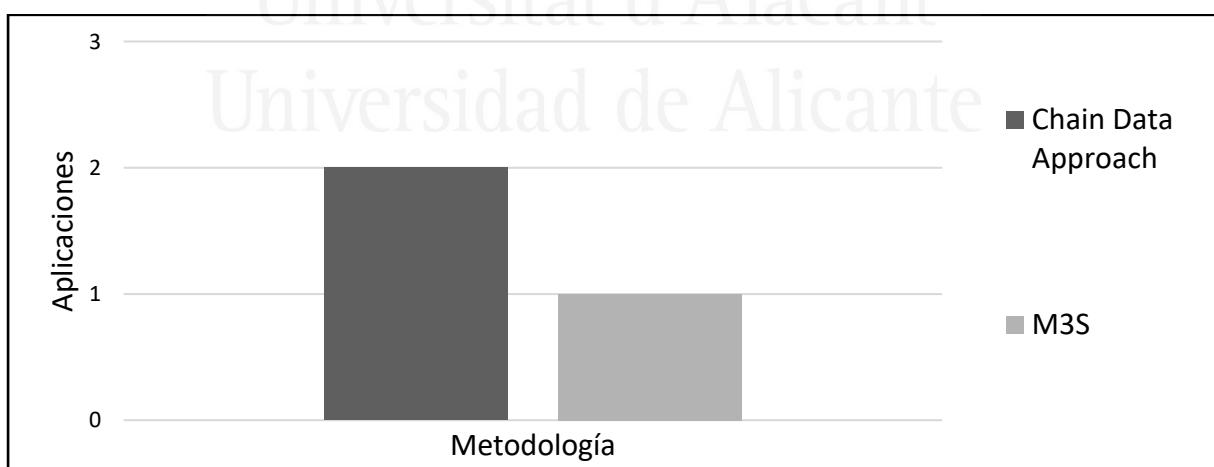


Figura 59. Pregunta 10-CCU-Aplicativos.

En la pregunta 11 gráficamente se muestra los resultados en la Figura 60, refiriéndose al número de solicitudes al año de solución BI. En esta pregunta ambas metodologías tanto la Chain Data Approach como la metodología M3S necesitan de una solicitud.

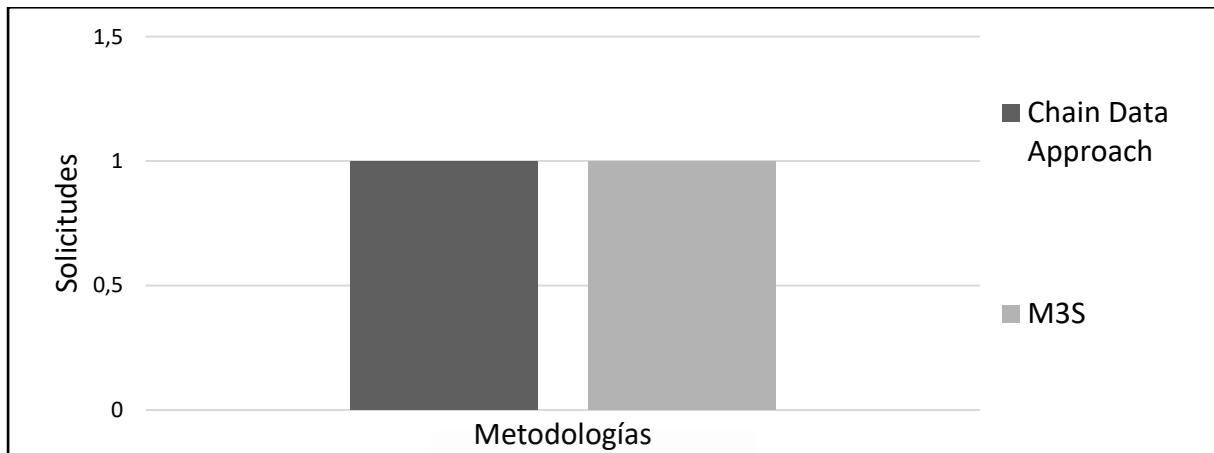


Figura 60. Pregunta 11-CCU-Solicitudes.

5.5. DISCUSIÓN DE RESULTADOS.

La utilización de las encuestas que fueron diseñadas para obtener y valorar ciertos parámetros cualitativos y cuantitativos de las empresas, están tomando cada vez más relevancia entre las investigaciones llevadas a cabo en todo ámbito. En lo que se refiere a nuestro estudio se obtuvo información de dos aspectos importantes como son por un lado la madurez de información que tengan los sistemas transaccionales que operan el día a día de una organización y por otro una comparación entre los procesos metodológicos que llevan en distintos aspectos de la vida de sus propios proyectos. Así establecido estos instrumentos llegamos a obtener los siguientes criterios:

- Los proyectos de tecnología en gran porcentaje se formulan claramente dentro de las empresas, sin embargo que por diferentes circunstancias como las políticas, organizacionales o financieras estos pueden cambiar su propuesta original, lo que influye directamente en la consolidación y

sobre todo en el tiempo de desarrollo total del proyecto.

- La eficiencia en costo, calidad y tiempo, son difíciles de cumplir en proyectos de TI, debido en unos casos a falta de liquidez financiera, procesos claros y standarizados; por lo que estos factores repercuten en una demora en la ejecución de las distintas etapas.
- Se debe trabajar en la difusión y apoyo de certificaciones de calidad, para que los departamentos de tecnología cuenten con indicadores que le permitan tomar fuerza en las empresas; manteniendo niveles internos y externos para medir y justificar su propia gestión.
- El haber llegado a resultados similares a los obtenidos anteriormente, pero a menor costo y tiempo, representa una oportunidad de mejora importante para que M3S se formalice dentro de las empresas y se posicione llevando nuevos proyectos de BI adelante.
- Comparativamente se consigue un mejor rendimiento en el uso de recursos humanos, financieros y más aún una mayor satisfacción al usuario de tecnología que llegó a conocer M3S, dejando atrás desmotivaciones de tipo profesional a los técnicos y en general el personal involucrado.
- En un contexto global se pudieron determinar ciertas debilidades como mal agrupamiento de atributos, reglas erróneas que nos llevan a la creación de una estructura que no ayuda a tomar buenas decisiones; pero a su vez fortalezas importantes parámetros como una disminución de tiempo y recursos; así como dejar latente el uso de este tipo de proyectos en las empresas.

6. CAPÍTULO VI: CONCLUSIONES Y LÍNEAS FUTURAS

6.1. RESUMEN Y CONSIDERACIONES FINALES

En la actualidad y más aún al haber llegado al final de esta investigación ratificamos que uno de los activos más valiosos en una empresa es la información y el valor de esta depende de cómo es gestionada, siendo pieza clave en las organizaciones para obtener mayor conocimiento tomando mejores y eficaces decisiones.

En los casos que fueron analizados se almacenaba enorme cantidad de datos, los cuales pudimos encontrarlos en diferentes tipos de bases de datos y plataformas en muchos de los cuales se tiene un alto grado de heterogeneidad, pero la clave radicó principalmente en que se ganó ventaja competitiva tanto en la velocidad de obtención como en la misma calidad y nivel de pureza de esos datos agrupándolos según la categoría aquí descrita.

Por lo tanto, en los casos de estudio ejecutados, la gestión eficiente de la inteligencia de negocios permitió ampliar la visión estratégica, reducir el riesgo de costo y disminuir la

incertidumbre en la toma de decisiones empresariales; con lo que debemos acotar lo siguiente:

- M3S, luego de haber probado su aplicación se comprueba que es una metodología que apunta a las nuevas posibilidades de BI, con fuerza en las pequeñas y medianas empresas públicas o privadas y de cualquier sector productivo o comercial, de tal forma que soportados en la innovación tecnológica se implantó para mejorar las decisiones en las empresas estudiadas.
- La propuesta M3S y su metodología, se validó contra los objetivos planteados en esta tesis, se ha confrontado contra otras opciones metodológicas; se validó por parte de expertos y fue testeada en dos empresas de diferente tipo con muy buenos resultados en esta investigación.
- M3S se la ha socializado para empresas con un mínimo nivel de madurez en el manejo de la información.
- M3S se construyó en base a la experiencia pragmática de profesionales con experticia de muchos años en proyectos de este tipo, por lo que justamente pudo ser probado en diferentes empresas y con el apoyo necesario.
- Fue ya categorizada como una metodología integradora, que cuenta con buenas prácticas en cada una de sus etapas, en base a la gama de proyectos ejecutados dentro de distintas áreas.
- Luego de los casos de estudio consideramos que se implantó mejor en pequeñas y medianas empresas por la mayor flexibilidad en su estructura plana organizacional, lo que llevó a un mejor apoyo tanto técnico como administrativo.
- M3S, se ejecutó con una clara descripción de los procesos y actividades a seguir en la metodología, haciéndola mucho menos burocrática incluso en su propia documentación, más aún en sus etapas que se definen como más ágiles.
- Se presentó como una metodología con mayor flexibilidad en la medida que la experiencia de los profesionales que estén al frente puedan ejecutarla de manera óptima en cada una de sus etapas.
- M3S, basándose en la fortaleza teórica tanto de metodologías específicas como genéricas existentes y con la disponibilidad de herramientas tecnológicos de alto poder en velocidad de ejecución, se volvió una alternativa práctica y fácil que llevó a cabo proyectos de BI.
- Fue considerada como innovadora, en el sentido que quiere aportar nuevas

optimizaciones en partes críticas de los proyectos de BI, como son los tiempos y formas de extracción de información.

Todo esto permite que con un alto grado de confiabilidad M3S acreditó su pertinencia y su factibilidad, se constituye formalmente como una propuesta a ser considerada en las áreas de informática de las empresas en el país, y finalmente, por lo analizado, descrito en los capítulos anteriores y como conclusión podemos decir que: La Metodología M3S para proyectos de IN, es una nueva alternativa que se pone como una opción dentro de la pequeña y mediana empresa, con objetivos claramente definidos y que se enmarca en un ciclo de vida muy fácil y específico.

6.2. CONTRIBUCIÓN REALIZADA.

En el desarrollo de la investigación se han podido realizar diferentes aportes en varias áreas, en primer lugar dentro de la academia poniendo a consideración de los estudiantes una alternativa metodológica que quedará bajo su propio criterio la implementación según sea el caso de trabajo que se les presente a futuro, se deja establecido opciones de diseño en proyectos de BI para disminución de tiempo y por otro lado se deja en cada evento que vamos participando una propuesta diferente de llevar consultorías de tecnología a otro nivel.

6.2.1 APORTACIONES GENERALES.

El desarrollo de las tecnologías de información ha ido evolucionando, y nosotros lo hemos analizado desde la perspectiva específica en el área de Inteligencia de negocios,

en la que en la actualidad nos permitió buscar una nueva forma de analizar, diseñar, construir y ejecutar proyectos de este tipo, como se ilustra en la Figura 61; por eso es que el aporte principal realizado en esta investigación es poner en consideración una metodología para desarrollo de Proyectos de Inteligencia de Negocios que al mismo tiempo que es ágil en su ejecución brinde nuevo conocimiento sobre la información propia de cada empresa y permita tomar decisiones más eficaces en la vivencia organizacional.

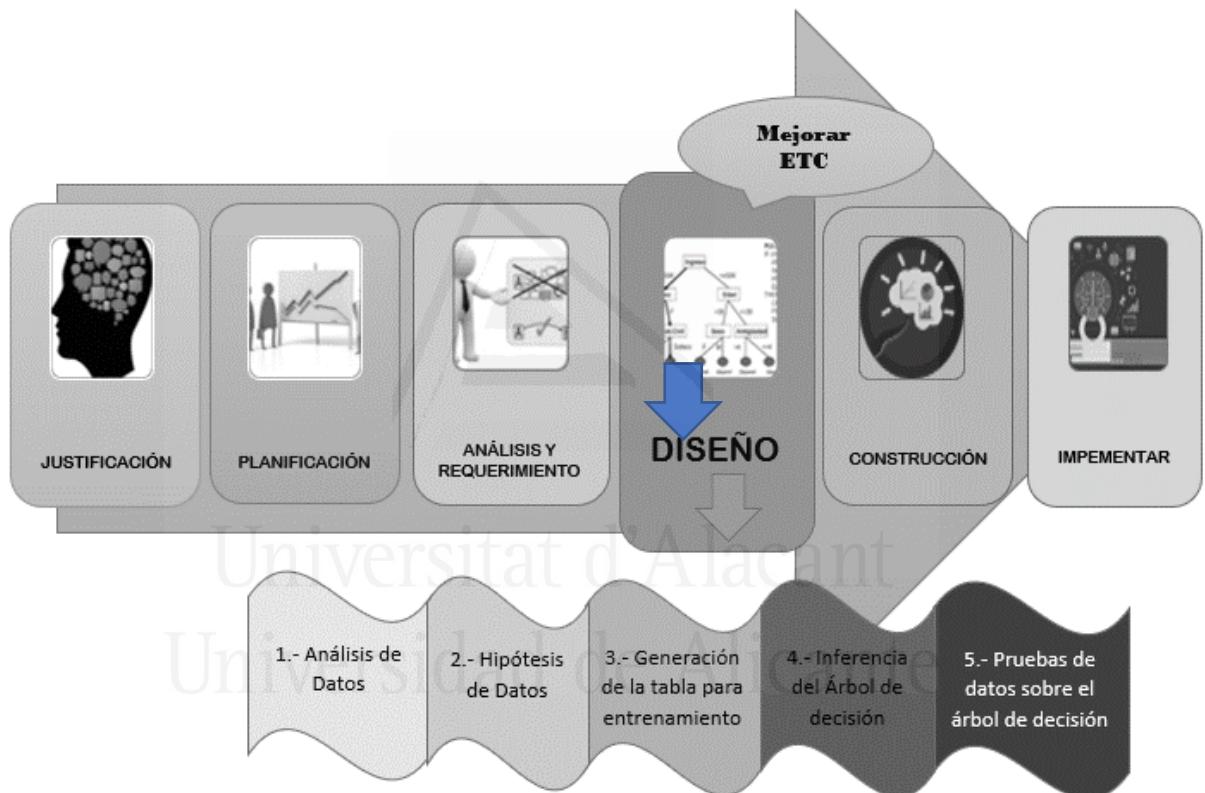


Figura 61. Etapas de Metodología de Decisión.

Se comprobó que dentro de M3S las etapas que aportaron con la mayor disminución en el tiempo invertido fueron la de diseño y la de construcción, por lo que hemos propuesto que dentro de una metodología formal, incluyamos un proceso respaldado a su vez en otra

metodología que dentro del estudio la conocimos como la de “decisión”, la que fue de igual forma perfectamente definida y la cual gráficamente se la muestra en la Figura 62, y la que nos ayudó a generar un árbol de decisión ideal, el mismo que sirvió para la verificación de los datos actuales y de igual forma aportó prediciendo las decisiones en los nuevos que se van incorporando día a día en un determinado estudio. Esto hace que M3S llegue a ser similar a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que ocurren de forma sucesiva, para la resolución de un problema y nosotros dentro de la metodología en la fase de Diseño y Programación, aplicamos Aprendizaje Automático (AA) en el momento en que aplicamos los árboles de decisión.

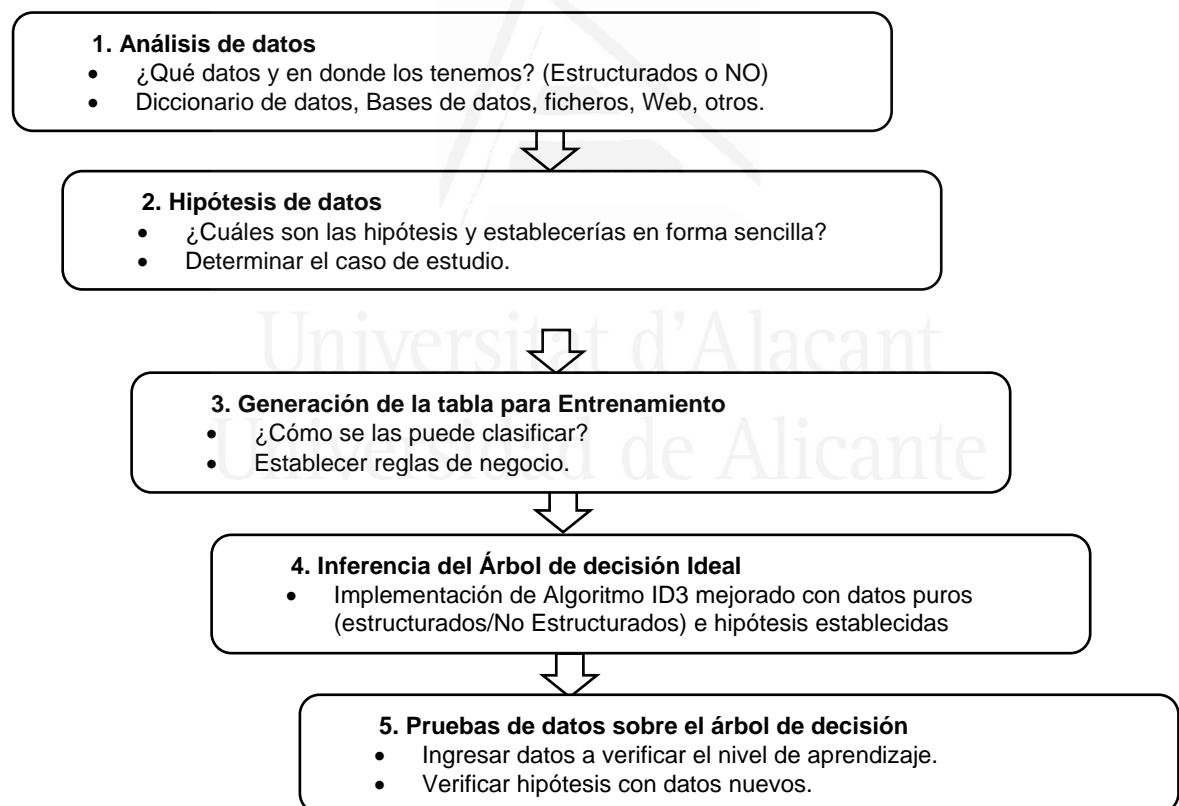


Figura 61. Etapas de Metodología de Decisión.

Adicionalmente, de lo indicado se ratificó con el análisis realizado por medio de instrumentos estadísticos, y en este caso fue tanto con la encuesta de madurez de información empresarial, como en la de comparación de metodologías cuyas tabulaciones se encuentran como ANEXO XIII y XIV, de las que principalmente se obtienen que la mejora en tiempos es sumamente importante, así como en el número de personas que se deben involucrar en los proyectos, haciendo que cualquier estudio de este tipo usando M3S, presente una disminución importante de costos, generando así mayor rentabilidad en el ámbito dentro del cual se esté trabajando.

6.2.2 APORTACIONES TEÓRICAS.

El establecer un orden y distintas prioridades es necesario para cumplir que nuestra actividad tenga éxito y podamos alcanzar el objetivo planteado desde un principio, evitando que este resultado esté condicionado por factores externos y que la cuota del azar o la suerte no sea la más importante a la hora de efectuarlo. Para evitar estos inconvenientes se propuso este modelo y metodología, lo que es aplicable en todo ámbito de estudio y análisis de IN, estableciendo una estructura de orden jerárquico que simplifica y facilita todo el proyecto, llevando a ejecutarlo en menor tiempo y con mejores decisiones a tiempo, permitiendo obtener una mayor rentabilidad a las empresas.

En la práctica de igual forma se concluye que la Metodología M3S, ayudará a las pequeñas y medianas empresas a gestionar sus procesos de innovación tecnológica, las conclusiones obtenidas acerca de M3S en base a su ciclo de vida y todo lo que se gestiona internamente, son frutos de múltiples revisiones de literatura existente en IN, análisis de propuestas, metodologías, discusiones y muchas expectativas que se tiene sobre este tipo de estudios. Si M3S, logra servir para nuevas consultorías que podrán ejecutarlas

profesionales independientes o empresas en esta área importante; estaremos hablando también de un aporte significativo en lo que se refiere a nuevos ingresos de tipo financiero a quienes la manejan y a su vez, se estará contribuyendo dentro del ámbito técnico el cual nos corresponde a dinamizar en gran medida la economía del país.

6.2.3 APORTACIONES DE TRANSFERENCIA Y DIFUSIÓN DEL CONOCIMIENTO

La transferencia de conocimiento práctico en las empresas, deben retornarnos información valiosa para mejorar esta investigación y dentro de cada caso de estudio se tienen nuevas aportaciones a la metodología como:

- La capacidad de adaptabilidad de M3S en el entorno informático de la empresa.
- El impacto y receptividad de diferentes niveles de la información, esto como ratificación de la validez de M3S.
- El aprendizaje sobre M3S de los usuarios de cualquier área inmersos en el proyecto.

Cuando se tenga un grupo de empresas o áreas que requieran de este tipo de proyectos de BI, prioritariamente dentro de la Universidad Central del Ecuador o ya sea en cualquier área que se vincula con la academia, se tendrá la oportunidad de obtener información importante para validar y robustecer a M3S, agregando nuevos indicadores dentro de las encuestas tanto cualitativas como cuantitativas que ratifiquen lo ejecutado en proyectos. Toda esta información que representará el comportamiento de la metodología en nuevos casos reales de estudio será un conocimiento significativo para nuevas investigaciones.

Se ha propuesto compartir los resultados de la investigación a través del blog y páginas web de la Dirección de Investigación de la U.C.E., permitiendo un acceso sencillo, rápido e ilimitado; de igual forma se ha iniciado la organización de charlas informativas, foros interactivos y participación en congresos de investigación ligada a la ingeniería, en donde sea tema principal el planteamiento de esta propuesta. De la forma indicada hemos participado de algunos eventos académicos de distinto tipo con ponencias o asistiendo a los que vamos a indicar:

- Del 7 al 9 del mes diciembre de 2016, dentro de la Universidad Estatal de Milagro, nos presentamos en el [I Congreso Internacional de Software Aplicado a la Agricultura, Turismo e Industria SAATI.](#), en la Ciudad de Milagro-Ecuador, con el tema “Implementación de Metodología de IN, con optimización en el proceso ETL, usando tecnología IN MEMORY”. [Memorias](#). (Morales, Morales, & Rizo, Implementación de Metodología IN, con optimización en ETL usando tecnología In Memory, 2016).
- El 5 de julio de 2017 se publican dos artículos con los títulos de: Metodología de IN con mejoras en la extracción y transformación de fuentes de datos, y por otro lado el tema: Inteligencia de Negocios basada en Bases de Datos In Memory; los mismos que se inscribieron en Publicando, revista que está indexada en Catálogo Latindex, con ISSN: 1390-9304.
- Para el mes de Agosto de 2017, en las Islas Galápagos se expuso en el [Congreso Internacional de Investigación \(CIDE - Ecuador\)](#), el tema: “Tendencias Tecnológicas en Ingeniería, Metodología M3S”, del cual se generaron las memorias científicas y las mismas constan con ISBN: 978-9942-759-18-4. [Memorias](#). (Morales, Morales, & Rizo, Tendencias Tecnológicas en Ingeniería, Metodología M3S., 2017)
- Posteriormente en el mes de Octubre de 2017 entre el 18 y el 20, dentro del [II Congreso Internacional. Ciencia, Sociedad e Investigación Universitaria](#), desarrollado en la Pontificia Universidad Católica del Ecuador, de la Ciudad de Ambato, se propuso el tema “Metodología para los procesos de inteligencia de negocios con mejoras en extracción y transformación de fuentes de datos”. [Memorias](#). (Morales, Morales, & Rizo, Metodología para los procesos de inteligencia de negocios con mejoras en extracción y transformación de fuentes de datos, 2017)

- En los días 22,23 y 24 en el mes de noviembre de 2017 dentro de la Universidad Central del Ecuador, en el II Congreso Internacional de Innovación en la Educación Superior, expusimos nuestro avance con el tema “Metodología para los procesos de inteligencia de negocios con mejoras en extracción y transformación de fuentes de datos”. Memorias. (Morales, Morales, & Rizo, Metodología para los procesos de inteligencia de negocios con mejoras en extracción y transformación de fuentes de datos., 2017)
- Y al final del año 2017 el 14 y 15 de diciembre, en el marco de la I Jornada de Investigación. La Simulación en Ingeniería: Trascendiendo Fronteras, de la Pontificia Universidad Católica del Ecuador en su sede de la Ciudad de Esmeraldas, de igual forma se presentó el avance del trabajo con el tema de nuestra investigación, y las mismas fueron registradas con ISBN: 978-9942-9902-6-6. Memorias. (Morales, Morales, & Rizo, Metodología para los procesos de inteligencia de negocios con mejoras en extracción y transformación de fuente de datos, 2017).
- En lo que va del presente año 2018 se realizaron las primeras Jornadas de “Sistemas de Información e Inteligencia Artificial” en el Instituto de Investigación y Posgrado de la Facultad de Ingeniería Ciencias Físicas y Matemática de la Universidad Central del Ecuador, en las cuales de igual forma tuvimos la presentación de nuestra propuesta dentro del caso de estudio de la E.E.Q. (Morales, Morales, & Rizo, Sistemas de Información e Inteligencia Artificial, 2018).
- Al finalizar el primer semestre desde el 13 al 16 de junio de 2018, asistimos con una ponencia a la 13 ^a Conferencia Ibérica de Sistemas y Tecnologías de Información CISTI’2018 en la Universidad de Extremadura. Cáceres-España, presentando el caso de estudio de la E.E.Q., el mismo que se encuentra registrado con el DOI: 10.23919/CISTI.2018.8399352 e ISBN: 978-989-98434-8-6 y bajo la dirección electrónica <https://ieeexplore.ieee.org/xpl/mostRecentIssue.jsp?punumber=8390719>.

En lo futuro se seguirá revisando la metodología realizando un proceso de socialización más fuerte, incorporando mejoras en cada proceso que se lo pueda hacer, generando así una mejora continua en M3S.

6.3. LÍNEAS FUTURAS.

La metodología M3S, es una alternativa para el avance de proyectos y aportes científicos en la sociedad, por eso su finalidad fue mejorar los tiempos de diseño y desarrollo de un proyecto de BI, para esto es prioritario disponer de la flexibilidad y agilidad necesarios para mejorar los procesos de toma de decisiones en función de los cambios frecuentes que suceden en el contexto de un trabajo específico. El presente estudio combina los principios de las metodologías ágiles en el desarrollo de software, alineados a la ruta metodológica de un proyecto de Inteligencia de Negocios, con el propósito de generar un proceso ágil de extracción de datos, con el fin de obtener resultados en un tiempo relativamente corto a comparación de metodologías tradicionales. La intención es la de dotar de una base sólida a empresas dedicadas a la prestación de servicios de tecnología, para que de este modo se permita llevar a cabo el estudio de los datos a la misma velocidad que estos cambian y se actualizan, permitiendo que la toma de decisiones de las empresas sea la más adecuada y en consecuencia como resultado de éstas sirvan como una mejora en la economía del país.

M3S, en la actualidad persigue generar el aumento de personas que busquen crear su propio emprendimiento dentro o fuera del país; empresas que fomenten la competitividad en el área de tecnología, pero ubicándose en un nivel superior a todo lo que represente software transaccional e incluso ERP; por lo que estamos seguros será una nueva línea de negocio importante. Esto se ratifica en base a que existen una serie de estudios dentro de los cuales se encuentra que a las empresas transnacionales les interesa la implementación de Inteligencia de Negocios, así que los esfuerzos empresariales deberán estar encaminados a conquistar este nivel de organizaciones, si no es posible incorporarlos permanentemente sería factible la tercerización de servicios de ingeniería de software orientados a la propagación de M3S.

6.3.1. DE INVESTIGACIÓN.

En nuestra investigación realizada se enmarcan nuevos proyectos que pueden ser considerados, entre los temas que han surgido y que por su alcance u orientación no han sido tomados en cuenta, los mismos podrían ser los siguientes:

- Riesgos que existen en la Inteligencia de Negocios, tomando en cuenta que la información que se logra obtener puede ser de extrema confidencialidad, lo que debería abrir un esquema de seguridad política y legal importante.
- La rápida y constante evolución de las herramientas de BI, que no permiten ser manejadas adecuadamente y en su totalidad, lo que propone una mayor difusión de estas herramientas en el mercado potencializando niveles de capacitación y consultorías de alto nivel.
- El manejo del término Innovación debe ser fortalecido en las Pymes y concatenarlo con la parte tecnológica de las empresas, como una sección transversal a todo tipo de organización, incorporando herramientas para toma de decisiones lo que permitirá que mejoren las finanzas de las empresas y se dinamice la economía del país.
- Se propondría la automatización completa del Proceso del Algoritmo de Decisión Ideal, para que sea motivo de un testeo más rápido y permita simulaciones de diferentes escenarios con distintos atributos y a su vez se puedan generar patrones automáticos de captura de datos.
- Sería interesante profundizar en el tema aplicaciones “en memoria” (in memory), que puedan dar mayor alcance de usuarios a los sistemas BI, por sus facilidades y rapidez en la ejecución de proyectos.
- M3S se diseñó para que la implementación de proyectos BI, sea más eficaz y vemos que se lo puede lograr en las Pymes con mayor facilidad por el aplanamiento de su organización. Se propone usar con mayor frecuencia software BI con énfasis en memoria, para evaluar puntualmente su impacto.
- La metodología considera la madurez de la información que se tenga en las

empresas para la obtención de mejores resultados, debido a que las que cuentan con este requisito se supone el uso de mejores prácticas metodológicas en proyectos; entonces se puede analizar empresas nuevas en las cuales se aplique M3S y evaluar el impacto que puede tener.

- El proceso de seguimiento de graduados de la U.C.E., es parte del mejoramiento de la gestión en la calidad de los estudiantes, como estrategia para mejorar los indicadores de vinculación con la sociedad; para lo cual existen datos dispersos en diferentes unidades académicas que no han sido explotados en gran medida. Se propone implementar una plataforma de BI bajo la metodología M3S, para integrar, procesar y estandarizar la información que se pretende mejorar.
- A nivel mundial existen procesos electorales que se enmarcan de excelente forma con el tipo de análisis realizado en M3S, por lo que se debe proponer un caso de estudio sobre un proceso dentro de este ámbito político.
- M3S se presentó por etapas, sin embargo estamos seguros que dependiendo la empresa puede transformarse a un modelo de procesos en el cual se describa cada una de las actividades y sus interrelaciones dentro de la realidad de cada proyecto.





ANEXOS.

Universitat d'Alacant
Universidad de Alicante

ANEXO I. ACTA DE CONSTITUCIÓN DE PROYECTO (IN)

FORMA 1. INICIACIÓN DE PROYECTO (IN)		METODOLOGÍA M3S		
 Código de Proyecto: M3SPY-01				
A. INFORMACIÓN GENERAL DEL PROYECTO				
1. Nombre del proyecto: Campaña de utilización de focos LED				
2. Fecha de inicio: 10/10/2017	3. Fecha de Finalización: 10/12/2017			
4. Empresa requirente: Empresa Eléctrica Quito	5. Tableros: Focalización de Clientes Ubicación Geográfica			
B. DEFINICIONES DE PROYECTO				
1. Descripción del proyecto. El proyecto pretende sustituir focos tradicionales por focos led a ciertos clientes del Distrito Metropolitano de Quito que excedan el consumo eléctrico promedio.				
2. Problema/Necesidades a resolver. Excesivo consumo eléctrico en el sector Residencial.				
3. Objetivo(s) del proyecto. Identificar a los clientes en las distintas parroquias del Distrito Metropolitano de Quito para los cuales se pueda generar una campaña de masificación de uso de focos led.				
3.1. Objetivos Específicos. <ul style="list-style-type: none"> * Hacer uso de árboles de decisión para inferir reglas que nos permitan llegar con la campaña planteada a los estratos que lo requieran. * Justificar nuestra decisión basados en el algoritmo ID3, y con un fuerte soporte de expertos. * Seleccionar las partes de la base de datos adecuadas para aplicar árboles de decisión de tal forma que se optimicen las decisiones. 				
4. Entregables por fases <Descripción general de los entregables e hitos>				
Fase	Fecha de entrega	Detalle		
Justificación	16/10/2017	Formalizacion legal, Documento legal de formalización		
Planeación	27/10/2017	Entrega del Cronograma		
Análisis Negocio	31/10/2017	Metricas y Dimensiones		
Diseño	07/11/2017	Tabla de Entrenamiento, Arbol de decisión		
Construcción	26/11/2017	Generacion de aplicativo con QlikSense		
Implementación	05/12/2017	Acta final de entrega		
5. Estimación del personal		Rol	Tiempo	Cantidad
Jorge Morales Cecilia Sánchez Santiago Morales	Daniel Larena Ricardo Lascano	Gerente del proyecto Administración de las TICs UCE	Completo	<Número de recursos>
6. Relación con otros proyectos <Descripción de los proyectos relacionados>				
Código del proyecto	Nombre del proyecto	Comentario <Nombre del proyecto> <Descripción de la relación con el proyecto>		
C. AUTORIZACIÓN DE LA INICIACIÓN DEL PROYECTO				
Nombre:	Santiago Morales			
Cargo:	Gerente del Proyecto			
D. OBSERVACIONES				
	Fecha	Nombre	Firma	
Elaborado por:	16/10/2017	Daniel Larenas		

ANEXO II. INFRAESTRUCTURA HARDWARE Y SOFTWARE (IN)

 FORMA 2. INFRAESTRUCTURA HARDWARE Y SOFTWARE (I N)		METODOLOGÍA M3S	
Código de forma: M3SPY-00			
A. INFORMACIÓN GENERAL DEL PROYECTO			
1. Nombre del proyecto: Sistema de facturación Empresa Eléctrica Quito S.A			
2. Fecha de Inicio:	3. Fecha de Finalización:		
24/07/2017	24/07/2017		
4. Empresa requirente:	5. Tableros:		
Universidad Central del Ecuador			
B. REQUERIMIENTOS MÍNIMOS PARA EQUIPOS DE USUARIOS FINALES			
1. Sistema Operativo:	2. Memoria RAM:		
Windows 10	4 GB (mínimo)		
3. Espacio en Disco Duro:	4. Procesador:		
15 GB	Intel Dual Core (mínimo)		
5. Navegador:	6. Arquitectura (x86/x64):		
Google Chrome	x64		
C. REQUERIMIENTOS MÍNIMOS PARA EQUIPOS DE USUARIOS AVANZADOS			
1. Sistema Operativo:	2. Memoria RAM:		
Windows 10	8 GB (mínimo)		
3. Espacio en Disco Duro:	4. Procesador:		
15 GB	Intel Core I3 (mínimo)		
5. Navegador:	6. Arquitectura (x86/x64):		
Google Chrome	x64		
D. REQUERIMIENTOS MÍNIMOS PARA SERVIDOR DE SOFTWARE BI			
1. Sistema Operativo:	2. Memoria RAM:		
Windows Server 2012	32 GB (mínimo)		
3. Espacio en Disco Duro:	4. Procesador:		
250 GB	CPU de 8 núcleos físicos y 2.0 GHz o más		
5. Navegador:	6. Arquitectura (x86/x64):		
Google Chrome	Procesador de 64 bits		
E. OBSERVACIONES			
Elaborado por:	Fecha	Nombre	Firma
Revisado por:			

ANEXO III. CRONOGRAMA (IN)

FORMA 3. CRONOGRAMA DE ACTIVIDADES (IN)

ANEXO IV. ANÁLISIS DEL NEGOCIO (IN)



**FORMA 4. ÁNALISIS Y REQUISITOS METODOLOGÍA M3S
DEL NEGOCIO (I N)**

Código del Proyecto: M3SPY-00

A. INFORMACIÓN GENERAL DEL PROYECTO

1. Nombre del proyecto:

Sistema de facturación Empresa Eléctrica Quito S.A

2. Fecha de inicio:	3. Fecha de Finalización:
24/02/2017	24/07/2017
4. Empresa requirente:	5. Tableros:
Universidad Central del Ecuador	

B. FUENTE DE DATOS

DMQ Facturación TRB

Parroquias_Sectores

C. PROCESO DE CÁLCULO	INDICADORES (KPI's)

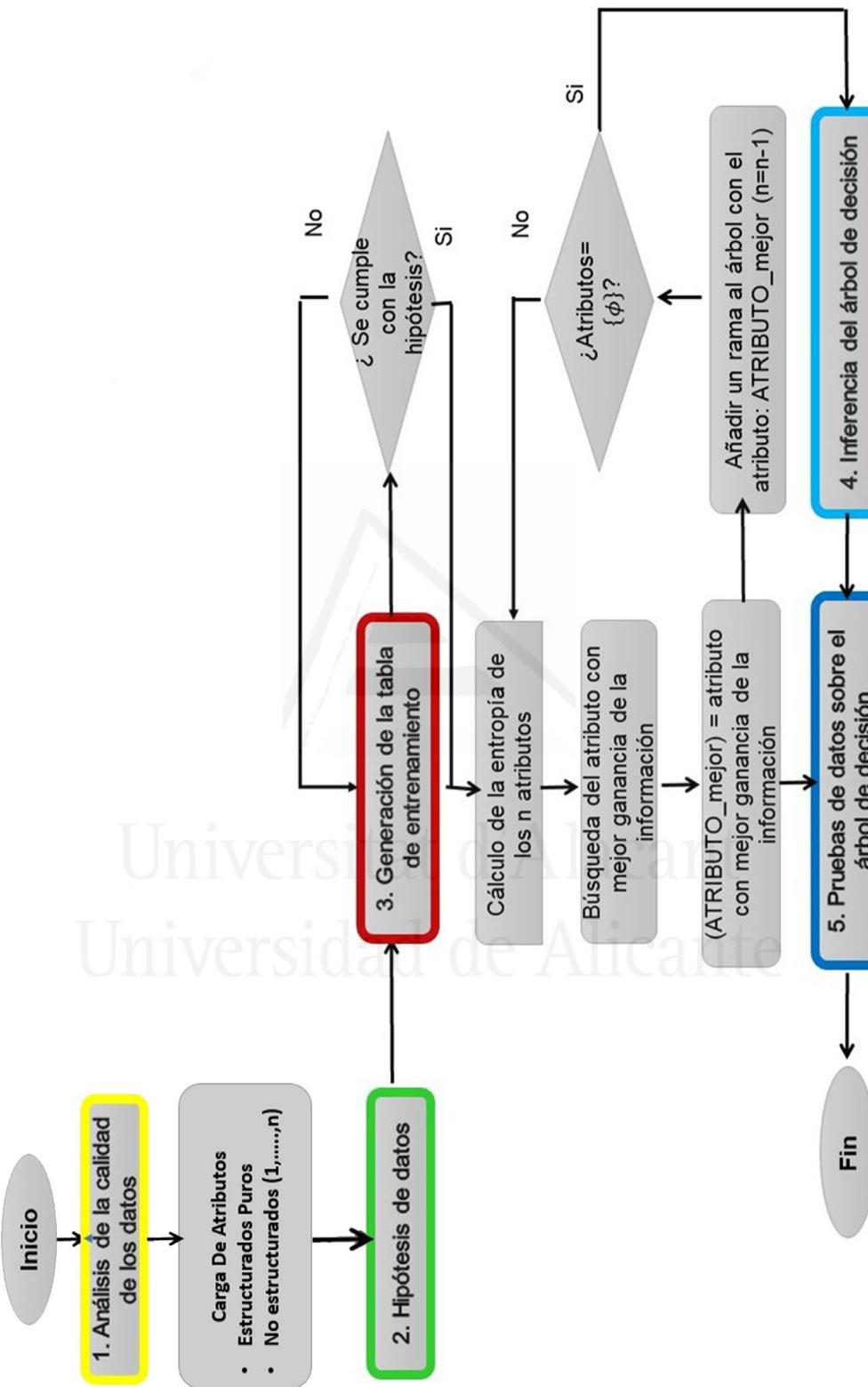
D. DISEÑO DE INTERFAZ

Universitat d'Alacant
Universidad de Alicante

E. OBSERVACIONES

	Fecha	Nombre	Firma
Elaborado por:			
Revisado por:			

ANEXO V. FLUJOGRAMA METODOLOGIA DE DECISIÓN (IN)



ANEXO VI. PSEUDOCÓDIGO METODOLOGÍA DE DECISIÓN (IN)

- 1. ANÁLISIS DE LA CALIDAD DE DATOS:**
 - CONTAR LA CANTIDAD DE ATRIBUTOS QUE SE VAN A PASAR, PARA CREAR LA TABLA TEMPORAL
 - CLASIFICAR TIPO DE ATRIBUTOS EN CASO DE:
 - ATRIBUTO ESTRUCTURADO DE ALTA PUREZA: ALMACENAR
 - ATRIBUTO NO ESTRUCTURADO: PONDERAR(1..5) Y ALMACENAR.
 - CREACIÓN DE TABLAS DE ALMACENAMIENTO, EN FUNCIÓN DE LA CANTIDAD DE ATRIBUTOS SELECCIONADO DE CUALQUIER TIPO SEGÚN SEA EL CASO.
- 2. HIPÓTESIS DE DATOS:**
 - DEFINICIÓN CLARA DEL OBJETIVO.
- 3. GENERACIÓN DE LA TABLA DE ENTRENAMIENTO:**
 - CREACIÓN DE LA TABLA DE REGLAS, AQUÍ SE ALMACENA UNA REGLA POR DEFECTO ADEMÁS DE LAS REGLAS ADICIONALES QUE SE PUEDAN CREAR.
 - SI SE CREA UNA NUEVA REGLA VALIDAMOS LA EFICIENCIA DE LA DICHA REGLA, PARA ESTO EL ALGORITMO VA A CALIFICAR LA NUEVA REGLA INGRESADA, DE LA SIGUIENTE MANERA:
 - SE INGRESA NUEVA REGLA
 - CREAR UN ÁRBOL DE DECISIÓN TEMPORAL (RAÍZ, HOJA DERECHA, HOJA IZQUIERDA)
 - ASIGNA REGLA A RAÍZ.
 - SE CONECTA HACIA EL ORIGEN DE DATOS Y EXTRAЕ EL DATO.
 - CUENTA LOS VALORES QUE CUMPLEN CON LA REGLA A PROBAR.
 - FIN EXTRACCIÓN E INICIO DEL ANÁLISIS DE ENTROPÍA Y GANANCIA
 - ENTREGO INFORMACIÓN
 - SI LA REGLA ES ACEPTADA, LA MISMA ES GUARDADA EN UNA TABLA DE REGLAS.
 - SI NO SE CREA NUEVA REGLA, SE EXTRAЕ LOS DATOS DIRECTAMENTE:
 - SE CREA LA ESTRUCTURA DEL ÁRBOL (RAÍZ, HOJA DERECHA, HOJA IZQUIERDA)
 - NOS CONECTAMOS A LA BASE
 - EXTRAЕMOS DATO DE LA BASE
 - EL DATO EXTRAЕIDO PASA POR EL ÁRBOL DE DECISIÓN, SI CUMPLE CON LA REGLA ESTE ES GUARDADO.
 - EXTRACCIÓN DE DATOS CON MÁS DE UNA REGLA:
 - SE CONECTA A LA BASE Y EXTRAЕ EL DATO
 - COMPARA EL DATO EXTRAЕIDO CON LA REGLA
 - VALIDAR QUE EXISTAN MÁS REGLAS, SI LAS HAY LA REGLA DE LA RAÍZ ES REEMPLAZADA CON LA NUEVA REGLA.
 - CAMBIAMOS DE ATRIBUTO PARA APLICAR LA NUEVA REGLA EN EL VALOR DE FILA.
 - EL PROCESO SE REPITE HASTA QUE EL NÚMERO DE LAS REGLAS ES IGUAL A 1 Y SE HAYAN RECORRIDO TODOS LOS ATRIBUTOS.
 - SE EXTRAЕ EL DATO QUE CUMPLIÓ CON TODAS LAS REGLAS.
- 4. INFERNERIA DEL ÁRBOL DE DECISIÓN**
 - ASIGNAMOS EL ATRIBUTO DE MEJOR GANANCIA DE INFORMACIÓN AL NODO
 - INFERIR EL ÁRBOL. IDEAL
 - REALIZAMOS EL CÁLCULO DE ENTROPÍA DE LOS N-1 ATRIBUTOS Y VERIFICAMOS LA TABLA DE ENTRENAMIENTO HASTA ACABAR CON TODOS
 - REGRESAMOS A LA ETAPA D (RECUSIVIDAD)
- 5. PRUEBAS DE DATOS SOBRE EL ÁRBOL DE DECISIÓN**
 - SELECCIÓN DE SUBCONJUNTO DE DATOS PARA REALIZAR VALIDACIÓN DEL ÁRBOL INFERIDO.

ANEXO VII. INFORMACIÓN FUNCIONAL Y TÉCNICA (IN)



**FORMA 5. INFORMACIÓN FUNCIONAL
Y TÉCNICA (IN)**

METODOLOGÍA M3S

Código de Proyecto: M3S-PY-04

A. INFORMACIÓN GENERAL DEL PROYECTO

1. Nombre del proyecto:

Sistema de facturación Empresa Eléctrica Quito S.A

2. Fecha de inicio:

24/02/2017

3. Fecha de Finalización:

24/07/2017

4. Empresa requirente:

Universidad Central del Ecuador

5. Tableros:

B. DOCUMENTACIÓN FUNCIONAL Y TÉCNICA

C. MÉTRICAS

Qvd's> :

D. DIMENSIONES

E. RESPALDO DE ARCHIVOS

F. OBSERVACIONES

	Fecha	Nombre	Firma
Elaborado por:			

Revisado por:

ANEXO VIII. ACTA DE CIERRE PROYECTO BI (IN)



FORMA 6. ACTA DE CIERRE PROYECTO BI (I N)

METODOLOGÍA M3S

Código de Proyecto: M3S-PY-04

A. INFORMACIÓN GENERAL DEL PROYECTO

1. Nombre del proyecto:

Sistema de facturación Empresa Eléctrica Quito S.A

2. Fecha de inicio:

24/02/2017

3. Fecha de Finalización:

24/07/2017

4. Empresa requirente:

Universidad Central del Ecuador

5. Tableros:

Universidad Central del Ecuador

B. DESCRIPCIÓN FUNCIONAL DEL TABLERO

Universidad Central del Ecuador <Nombre de a persona responsable del proyecto>

1. Objetivo del aplicativo:

Identificar a los clientes en las distintas parroquias del Distrito Metropolitano de Quito para los cuales se pueda generar una campaña de masificación de uso de focos led.

2. Pestañas que contiene el aplicativo:

3. Detalle funcional resumido:

Decisiones de a qué clientes se les aplica el proyecto de implementación de focos led

C. TAREAS PENDIENTES

Costo

Fecha de entrega

<Se deberá enumerar en el caso de que existan tareas pendientes en el aplicativo, y cuál es el responsable de ejecutarlas (Cliente).>

D. DATOS ADICIONALES DE CIERRE

Fecha de Cierre

<dd/mm/aaaa>

Nombre

Firma

Director del proyecto : Consultor BI

<Nombre completo del
requirente>

Director del proyecto : Cliente

<Nombre del director de
proyectos>

E. OBSERVACIONES

Mediante el presente documento se deja constancia de la entrega formal del aplicativo descrito en los párrafos anteriores cumpliendo con los objetivos y alcance establecidos.

	Fecha	Nombre	Firma
Elaborado por:			

Revisado por:			



Universitat d'Alacant
Universidad de Alicante

ANEXO IX. ASISTENCIA

FORMA 7. ASISTENCIA (IN)		METODOLOGÍA M3S	
Código de Proyecto: M3SPY-00			
A. INFORMACIÓN GENERAL DEL PROYECTO			
1. Nombre del proyecto:			
Sistema de facturación Empresa Eléctrica Quito S.A			
2. Fecha de inicio:			
24/02/2017			
3. Fecha de Finalización:			
24/07/2017			
4. Empresa requirente:			
Universidad Central del Ecuador			
B. ASISTENTES			
1. Nombre Evento:			
2. Horas totales:			
3. Facilitador:			
4. Fecha de actividad:			
1. Nombre:		2. Hora Inicio / Hora Fin:	
C. OBSERVACIONES			
		Firma	
		Nombre	
Elaborado por:		Fecha	
Revisado por:			



ANEXO X. DATOS ALGORITMO DE DECISIÓN (IN)**FORMA 8. DOCUMENTACIÓN FUNCIONAL Y TÉCNICA (IN)****Código de Proyecto: M3S-012****A. INFORMACIÓN GENERAL DEL PROYECTO****1. Nombre del proyecto:**

Sistema de facturación Empresa Eléctrica Quito S.A

2. Fecha de inicio:

24/02/2017

4. Empresa requirente:

Universidad Central del Ecuador

5. Objetivo del proyecto.

Describa a que muestra el campo decisión.

3. Fecha de Finalización:

24/07/2017

5. Tableros:**B. MODELO DE DATOS****1. Tabla de base de datos****2. Atributos que participan en el proyecto BI****3. Entropía****4. Ganancia de Información****C. REGLAS DE NEGOCIO****1. Tabla Reglas de Negocio****D. OBSERVACIONES**

Nombre	Fecha	Firma
Elaborado por:		
Revisado por:		

ANEXO XI. ENCUESTA DE MADUREZ EMPRESARIAL EN PROYECTOS BI

INSTITUCIÓN:..... Fecha:.....

RESPONSABLE:

VALORACIÓN					
OPCIÓN		A	B	C	
DETALLE		NO	PARCIALMENTE	NO	
PREGUNTA		A	B	C	
1	¿Los proyectos de su organización tienen objetivos claros y medibles?				
2	¿La eficiencia en costo, calidad y tiempo se ha cumplido en proyectos anteriores?				
3	¿Su organización utiliza estándares para la recolección de datos?				
4	¿Su organización utiliza estándares externos para medir y mejorar su gestión?				
5	¿Su organización tiene hitos definidos, para medir la gestión por áreas?				
6	¿Su organización utiliza y mantiene un marco de referencia común, metodología y procesos de gestión para todos sus proyectos?				
7	¿Su organización cuenta con alguna certificación de estándares de calidad?				
8	¿Su organización presupuesta inversiones en el área tecnológica anualmente?				
9	¿Cómo implementa su organización la retroalimentación de sus proyectos para la mejora continua?				

10. ¿Qué metodología de dirección de proyectos de BI usa?

No tenemos ninguna metodología estandarizada	A
Cada área de negocio desarrolla y aplica su propia metodología	B
Entre 2 y 3 diferencias	C
1 Diferencia	D

11. La aprobación de un Plan de proyecto de BI en mi organización contempla:

Los planes se aprueban sin que se siga ninguna metodología o estandarizada	A	
Un presupuesto y un programa que no están integrados y sin una estructura de desglose de trabajos (Work Breakdown Structure)	B	
Acta del proyecto, WBS, estimados de costo, presupuesto y cronograma	C	
Lo indicado en (c) además del plan de calidad y el plan de adquisiciones	D	
Lo indicado en (d), además de análisis de riesgo, evaluación de participantes (stakeholders), asignación y balanceo de recursos, roles y responsabilidades, y plan para gestión de cambios.	E	

12. ¿De qué forma se administra el control de cambios de los proyectos de BI en su organización?

No se administran los cambios.	A	
Midiendo su impacto para facilitar la autorización de los mismos por los niveles facultados para hacerlo	B	
Lo indicado en (b) y se registra en una bitácora de cambios con los datos más relevantes.	C	
Lo indicado en (c) de acuerdo a una metodología estandarizada de gestión de cambios integrada con una metodología de gestión de la configuración.	D	
Lo indicado en (d), con un repositorio empresarial en una base de datos manejada por una herramienta corporativa de dirección de proyectos en línea que me permite documentar y difundir todos los cambios.	E	

13. En mi organización se manejan las lecciones aprendidas y la mejora continua en Proyectos:

No tenemos un proceso estandarizado para las lecciones aprendidas ni para la mejora continua.	A	
Cada gerente de proyecto guarda los documentos principales de sus proyectos.	B	
Existe un proceso de generación de lecciones aprendidas y se difunde al terminar cada proyecto.	C	

14. Con respecto a la distribución y recopilación de información a todos los Miembros del equipo, Gerentes Funcionales o de Línea, y cualquier participante en los proyectos, en mi Organización existe un estándar de uso de herramientas de Software de Dirección de Proyectos, con vistas y plantillas personalizadas para la empresa (más del 80% de los proyectos lo usan y o presentan igual), para:

No se tiene estándar, cada participante lo usa a su discreción.	A	
Los participantes reciben información de los respectivos Gerentes de Proyectos en formatos estandarizados (papel o electrónicos), que llenan con sus avances y lo regresan al Gerente de Proyecto.	B	
Los participantes tienen acceso directo a las herramientas de software de dirección de proyectos e ingresan sus avances y estado de sus tareas en herramientas individuales por proyecto.	C	
Lo indicado en (c), con un control de horas por persona y de cualquier otro tipo de recurso (interno o externo) que se requiera para el desarrollo de sus tareas, en herramientas individuales por proyecto.	D	
Lo indicado en (d), en una herramienta en línea y en tiempo real (Servidor de Proyectos), con un sistema de autorizaciones por parte de sus Líderes Funcionales y de Proyecto, ligado automáticamente al sistema de correo electrónico de la organización.	E	

Nota: Gracias por ayudarnos con nuestra encuesta.

El éxito de los proyectos BI radica en el involucramiento tanto de clientes internos y externos.

ANEXO XII. ENCUESTA COMPARATIVA DE METODOLOGÍAS DE BI

INSTITUCIÓN: Fecha:

RESPONSABLE:



PARTE 1: CUALITATIVA

OPCIÓN	A	B	C
DETALLE	Satisficho	Parcialmente Satisficho	No Satisficho

Nota: Gracias por ayudarnos con nuestra encuesta.

	Pregunta	Metodología....			Metodología M3S		
		A	B	C	A	B	C
1	¿Ha returned la inversión a su empresa al aplicar la metodología?						
2	¿Existe buena aceptación en los desarrolladores de soluciones BI?						
3	¿Existe aceptación con los clientes?						
4	¿Es entendible esta metodología?						
5	¿Agregan valor a los clientes?						

El éxito de los proyectos BI radica en el involucramiento tanto de clientes internos y externos



PARTE 2: CUANTITATIVA

No.	Pregunta	METODOLOGIA...	METODOLOGIA M3S
1	Personal requerido para implementación en proyectos BI	_____ personas	_____ personas
2	Tiempo estimado en DISEÑO de extracción de datos	_____ min	_____ min
3	Tiempo estimado en depuración y transformación de datos	_____ días	_____ días
4	Tiempo estimado en la construcción de la data.	_____ días	_____ días
5	Tiempo estimado en la validación de datos	_____ días	_____ días
6	Costo del proyecto BI.	_____ Dólares	_____ Dólares
7	Número de usuarios definidos para el uso de la solución BI	_____ personas	_____ personas
8	Porcentaje de usuarios activos en la solución BI	_____ %	_____ %
9	Número de aplicaciones de BI realizadas	_____ soluciones	_____ soluciones
10	Número de solicitudes al año de soluciones BI	_____ solicitudes	_____ solicitudes

Nota: Gracias por ayudarnos con nuestra encuesta.

El éxito de los proyectos BI radica en el involucramiento tanto de clientes internos y externos.

ANEXO XIII. DATOS TABULACIÓN ENCUESTA DE MADUREZ.

Datos Encuesta 1

Pregunta	C.C.A.			UPS			EEQ			Motor Uno			CNT			SIFIZOFT			Seytratec			Facultad Odontología			SEGUROS UNIDOS			KRUGER			NUO		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C			
1	x		x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x				
2	x		x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x				
3		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
4	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
5	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
6	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
7	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
8	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
9	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
10	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
11	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
12	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
13	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			
14	x		x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x		x	x			

Resultados Encuesta 1

Sumatoria			Porcentajes		
A	B	C	Si	Parc.	No
10	1	0	91	9	0
1	9	1	9	82	9
4	5	2	36	45	18
1	7	3	9	64	27
2	7	2	18	64	18
2	3	6	18	27	55
4	2	5	36	18	45
5	5	1	45	45	9
2	9	0	18	82	0
3	6	2	27	55	18
2	4	5	18	36	45
4	6	1	36	55	9
2	8	1	18	73	9
3	3	5	27	27	45
Total:			45	75	34
			29	49	22

ANEXO XIV. DATOS TABULACIÓN ENCUESTA COMPARATIVA - CUALITATIVA.

Castillo De Amaguaña		DC Universidad Politécnica Salesiana		Empresa Eléctrica		Fast Flower		MT-CNT		RA – SIFIZSOFT		Seytratec		RH ODONTOLOGIA		WE SEGUROS UNIDOS		
Metodología A	M3S	Metodología B	M3S	Metodología C	M3S	Metodología D	M3S	Metodología E	M3S	Metodología F	M3S	Metodología G	M3S	Metodología H	M3S	Metodología I	M3S	
P	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
1	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
3	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
5	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x

Resultados Encuesta 2

Sumatoria todas las metodologías vs M3S			Porcentajes		
Todas		M3S	Todas		M3S
A	B	C	A	B	C
1	7	1	8	1	0
4	5	0	7	2	0
0	8	1	8	1	0
2	4	3	7	2	0
1	7	1	7	2	0

Datos Pregunta 6

	DC Universidad Politécnica Salesiana	RA – SIFISOFT	Seytratec	WE SEGUROS UNIDOS	NUO-TECHNOLOGY
Pérdida tiempo re calibración	x			x	x
Falta de personal experto en M3S		x			
Reglas erróneas árbol de decisión		x			
Falta De Retroalimentación		x		x	
Entropía por excesos de atributos					

Resultados Pregunta 6

Sumatoria	Porcentajes
3	43
1	14
1	14
1	14
1	14
7	100

Datos Pregunta 7

	DC Universidad Politécnica Salesiana	RA – SIFIZSOFT	Seytratec	WE SEGUROS UNIDOS	NUO-TECHNOLOGY
Reducción de tiempo	x	x	x		
Metodología más entendible			x		
Agilidad en la toma de decisiones			x		
Documentación ágil, eficaz, entendible y sencilla	x	x		x	
Proceso con etapas claras		x			
Versatilidad en implementación en distintas áreas	x			x	x
Extracción de datos en menor tiempo	x			x	x

Resultado Pregunta 7

Sumatoria	Porcentajes
3	20
1	7
1	7
3	20
1	7
3	20
3	20
15	100

Datos Pregunta 8

DC Universidad Politécnica Salesiana	RA – SIFIZSOFT	Seytratec	WE SEGUROS UNIDOS	NUO-TECHNOLOGY
1	X	X	X	X
2	X		X	

Resultados Pregunta 8

Sumatoria	Porcentajes
3	60
2	40
5	100

ANEXO XV. DATOS TABULACIÓN ENCUESTA COMPARATIVA - CUANTITATIVA.

No.	Pregunta		
1	Personal requerido para implementación en proyectos BI	4	Personas
2	Tiempo estimado en extracción de datos	15	Minutos
3	Tiempo estimado en depuración y transformación de datos	27	Días
4	Tiempo estimado en la construcción de la nube de datos	20	Días
5	Tiempo estimado en la validación de datos	20	Días
6	Costo del proyecto BI	30000	Dólares
7	Costo de hardware	25000	Dólares
8	Número de usuarios definidos para el uso de la solución BI	5	Personas
9	Porcentaje de usuarios activos en la solución BI	55	Porcentaje
10	Número de aplicaciones de BI realizadas	2	Soluciones
11	Número de solicitudes al año de soluciones BI	1	Soluciones

Datos 2da Encuesta Nuo

No.	Pregunta	METODOLOGIA B	M3S	Recursos
1	Personal requerido para implementación en proyectos BI	5	3	Personas
2	Tiempo estimado en extracción de datos	180	120	Minutos
3	Tiempo estimado en depuración y transformación de datos	30	15	Días
4	Tiempo estimado en la construcción de la nube de datos	9	4	Días
5	Tiempo estimado en la validación de datos	15	7	Días
6	Costo del proyecto BI	60535	45456	Dólares
7	Costo de hardware	30267,5	22730	Dólares
8	Número de usuarios definidos para el uso de la solución BI	30	40	Personas
9	Porcentaje de usuarios activos en la solución BI	60%	90%	Porcentaje
10	Número de aplicaciones de BI realizadas	1	2	Soluciones
11	Número de solicitudes al año de soluciones BI	2	4	Soluciones

BIBLIOGRAFÍA

- Aaker, D. A. (2015). Investigación de mercados (2da. ed.). Mexico: McGraw-Hill Interamericana. Obtenido de <http://www.sidalc.net/cgi-bin/wxis.exe/?IsisScript=UCC.xis&method=post&formato=2&cantidad=1&expresion=mfn=030250>
- AENOR. (2013). Dirección y Gestión de Proyectos - Norma UNE-ISO 21500:2013. Obtenido de [iso21500: http://www.idi.es](http://www.idi.es)
- Agenjo, F. E. (2014). Beneficio de la Norma ISO 21500:2013. Obtenido de <http://bpmsat.com/beneficios-de-la-norma-iso-215002013>
- Aqueda Barrero, A. (2014). Prince2 Un buen inicio para Project Manager. Recuperado el 03 de 03 de 2017
- Ahumada T, E., & Perusca V, J. M. (2016). Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica. Contaduría y Administración, 61(1), 127-158.
- ALEXA. (2017). Top Sites in Ecuador. Recuperado el 17 de 06 de 2017, de ALEXA: <http://www.alexa.com/topsites/countries/EC>
- Ballard, C. e. (2006). Improving Business Performance Insight... with Business Intelligence and Business Process Management. International Business Machines (IBM).
- Barcenas, G. (Enero de 2012). ¿Que es el PMI y que es el PMBOK? Recuperado el 08 de 04 de 2017, de <https://formulaproyectosurbanospmipe.wordpress.com/2012/01/18/que-es-el-pmi-y-que-es-el-pmbok/>
- Bernal, J. J. (23 de agosto de 2013). www.pdcahome.com. Recuperado el 16 de enero de 2017, de <http://www.pdcahome.com/5202/ciclo-pdca/>
- Bonke, S., & Winch, G. (2002). Project Stakeholder Mapping: Analyzing the Interests of Project Stakeholders. Frontiers of Project Management Research, 2002, p. 385-405.
- Brenes, C. (06 de 2013). Resumen de Norma ISO21500 Directrices para la Dirección y Gestión de Proyectos. Obtenido de <http://www.ucipfg.com>
- Caldeira, C. (2012). Data Warehousing: Conceitos e Modelos. Brasil. Obtenido de <http://hdl.handle.net/10174/6364>
- Calderon C, M., Iriarte G, A., & Trejos T, E. (2014). Marco de Referencia Implementación ISO21500. (U. d. Buenaventura, Ed.) Biblioteca Digital. Obtenido de <http://hdl.handle.net/10819/2319>
- Calzada, L., & Abreu, J. (09 de 2009). El Impacto de las herramientas de Inteligencia de Negocios. DAENA (International Journal of Good Conscience), 52. Obtenido de [http://www.spentamexico.org/v4-n2/4\(2\)/02016-52.pdf](http://www.spentamexico.org/v4-n2/4(2)/02016-52.pdf)
- Cámara Nuñez, C. (2010). Análisis de los Sistemas Bussines Intelligence y su aplicación práctica en los Proyectos de Software. Madrid, España: Universidad Carlos III de Madrid.
- Capgemini. (2014). Traditional BI vs . Business Data Lake – A comparison The need for new thinking around data storage and analysis. Obtenido de https://www.mx.capgemini.com/resource-file-access/resource/pdf/pivotal_data_lake_vs_traditional_bi_20140805_0.pdf
- Cobo Y, A. (2007). Diseño y Programación de Bases de Datos. Madrid, España: Vision Libros.
- Coquillat, M. (18 de Junio de 2014). Origen de la norma ISO 21500. Obtenido de <http://iso-21500.es>
- Coronel Carlos, R. P. (2011). Concepto de bases de datos. En Bases de datos diseño, . Mexico: Cengage Lerning.

- Cruz Caicedo, L. (2012). Comparativa ISO 21500 y PMBOK Versión 5. i Congreso Internacional Gerencia de Proyectos. Bogotá, Colombia. Obtenido de http://americalatina.pmi.org/media/files/latam/colombia/2012_co_cruz.aspx
- Date, C. (2001). Introducción a los Sistemas de Bases de Datos, Apoyo para la toma de decisiones. Obtenido de Revistas Uis: <http://revistas.uis.edu.co/index.php/revistauisingenierias/article/view/1771/2152>.
- DIAZ TOVAR, Y. (2014). Norma GTC-ISO 21500 en la gestión de la calidad la toma de decisiones en la gestión de proyectos. Bogotá. Obtenido de <http://unimilitar-dspace.metabiblioteca.org/bitstream/10654/13016/1/Ensayo%20Norma%20ISO%2021500.pdf>
- Díaz, J. C. (2010). Introduction Business Intelligence (1era ed., Vol. Primera Edición). (E. C. SA, Ed.) Barcelona, España: UOC. Recuperado el 16 de 07 de 2016, de www.editorialuoc.com
- Díaz, J. C. (2012). Introducción al Inteligencia de Negocios (Vol. 1era.). (E. C. SA, Ed.) Barcelona, España: UOC.
- Dictuc S, A., E.P., C., & Torres, M. (2010). Proyecto Infox divulgación y comercialización de una herramienta para proveer autoservicio de datos, computo e información a personas no especialistas. Santiago, Chile: Innova Chile.
- Dueñas-Reyes, M. (2009). Minería de datos espaciales en búsqueda de la verdadera información. Ingeniería y universidad,. Bogotá, Colombia: Universidad Distrital Francisco José de Caldas. doi:0123-2126
- Duran, E., & Costaguta, R. (2007). Minería de Datos para Descubrir estilos de Aprendizaje. (O. D. OEI, Ed.) Revista Iberoamericana De Educacion ISSN 1681-5653, 42(2), 10.
- Ecured. (2012). Tabla de Dimensión. Recuperado el 09 de 10 de 2017, de http://ecured.cu/Tabla_de_dimension%C3%B3n-bases-de-datos
- Espinosa, R. (2010). Kimball vs Inmon. Ampliación de Conceptos del Modelado Dimensional.
- Evelson, B. (2010). The Forrester Wave™: Enterprise. Business Intelligence. 20. Obtenido de ftp://ftp.support.lotus.com/software/sk/pdf/ForresterWave_for_Enterprise_Business_Intelligence_Q4_2010_Oct10.pdf
- Facil, B. (2009). Tablas de Hecho. Recuperado el 25 de 08 de 2016, de Inteligencia de Negocios Info: <http://www.businessintelligence.info/serie-dwh/tablas-de-hecho-fact-tables.html>
- Fernandez A, V. (2006). Desarrollo de Sistemas de Información. Una metodología basado en el Modelado (1era. ed.). (U. Politécnica, Ed.) Barcelona, España. Obtenido de <https://pdfs.semanticscholar.org/946d/1e19f09e8dac8744f7875c876fe14a73e9b0.pdf>
- Fernández, M. A. (31 de Enero de 2013). La Gestión y Dirección de Proyectos. Obtenido de Red.es: <https://www.aec.es>
- Freijedo, C., & Grassi, J. (2012). Estudio sobre metodologías para la gestión de proyectos de software. Propuesta y Aplicaciones. Argentina: Universidad de San Andrés.
- Fuentes, L., & Ricardo, V. (2010). Incorporación de Elementos de Inteligencia de Negocios en el Proceso de Admisión Y Matrícula de una Universidad Chilena. Ingeniare. Revista chilena de ingeniería, 18(3), 383-394. Obtenido de <http://dx.doi.org/10.4067/s01718-33052010000300012>
- Galhardas, H., Florescu, D., Shasha, D., & Simon, E. (2000, May). AJAX: an extensible data cleaning tool. In ACM Sigmod Record (Vol. 29, No. 2, p. 590). ACM.
- Gallego Calonge, F. (2012). La Medición de la Televisión Social en España. Conceptos y Evolución. Presente y Futuro. III Congreso Internacional de Comunicación 3.0, 13. Obtenido de <http://campus.usal.es/www.communication3punto.com>

- García M, R., Lelli, R., Merlino, H., Cornachia, L., Rodríguez, D., Pytel, P., & Arboleya, H. (2011). Ingeniería de proyectos de explotación de información para PyMEs. Red Universidades de Carreras de Informática. Obtenido de <http://sedici.unlp.edu.ar/handle/10915/20017>
- Garcia Reyes, R. (2012). Minería de Datos para la Toma de Decisiones e Inteligencia de Negocios. México D.F., México: Universidad Nacional Autónoma de México.
- García, C., Espinoza, L., Castañeda, M., Ostos, R., & Salazar, M. (2016). Gestión del Conocimiento.
- Garcia, D. (2007). Algoritmo de discretización de series de tiempo basado en entropía y su aplicación en datos colposcopicos (Doctoral dissertation, tesis para obtener el grado de Maestro en Inteligencia Artificial. Veracruz, México: Universidad Veracruzana. Obtenido de <http://cdigital.uv.mx/bitstream/123456789/32352/1/garcialopezdaniel.pdf>
- García, R. (12 de Febrero de 2018). Repositorio Central Cepal. Obtenido de http://repositorio.cepal.org/bitstream/handle/11362/33747/S8500576_es.pdf?sequence=1
- Gartner. (2017). Gestion de Proyectos. Recuperado el 15 de 02 de 2017, de <http://www.gartner.com/it-glossary/project-management>
- González, R. M., & M., &. d. (2001). Marketing en el Siglo XXI. (5TA ed., Vol. 5TO). (C. d. Financieros., Ed.) España: CEF. Obtenido de www.marketing-xxi.com: <http://cef.eslibros>
- Gracia, V., & Galán, A. (2014). WSM: Metodología De Web Scraping Para Android y Ejemplificación Mediante la Aplicación Upmdroid. Madrid, España.
- Grande, I., & Elena, A. (2005). Análisis de Encuestas. Madrid, España: ESIC.
- Granero, G. (2013). NOTAPRENSA Tabla AEN 6023 1. Recuperado el 18 de 11 de 2016, de <https://www.aenor.es>
- Grossman, W. (1998). Fundamentals of Business Intelligence. Springer.
- Guarin, O. (2016). Gerencia de Proyectos PMBOK (Parte 1). Recuperado el 19 de 04 de 2017
- Han, J. P., & Kamber, M. (2009). Data Mining: Concepts and Techniques (3era. ed., Vol. 3era Edición). (B. L. Data, Ed.) EEUU: Elsevier Inc. Obtenido de <http://mkp.com>
- Horwitt, E. (2011). Self-service BI catches on. Computerworld. Obtenido de <https://scholar.google.es/scholar?hl=es&q=Self-service+BI+catches+on&btnG=&lr=>
- Inmon, W. (2012). Building the Data Warehouse. New York, EEUU: John Wiley & Sons, Inc. New York, NY, USA ©2002. Obtenido de <http://dl.acm.org/citation.cfm?id=560407>
- Inmon, W. H. (2001). Data Warehousing. (berkeley, Ed.) Sao Paulo, Brasil.
- Jara, J. (2012). Big Data & Web Intelligence. Asunción, Paraguay: Universidad Católica Nuestra Señora de la Asuncion. Obtenido de <http://jeuazarro.com>
- Jimenez Nieto, E. (2013). La Inteligencia de Negocios. Recuperado el 16 de 02 de 2017, de <http://www.gestiopolis.com/la-inteligencia-de-negocios/>
- Joyanes Aguilar, L. (2013). Big Data: Análisis de grandes volúmenes de datos en organizaciones. México: AlfaOmega Grupo Editor, S.A de C.V.
- KAMEL. Rouibah y SAMIA, O.-a. (2002). A concept and prototype for linking business intelligence to business strategy. Journal of Strategic Information System Vol.11, 133-152. Obtenido de <http://www.sciencedirect.com/science/article/pii/S0963868702000057>
- Kettinger, W. J., Teng, J. T., & Subashish, a. G. (1997). Business Process Change: A Study of Methodologies, Techniques, and Tools. Management Information Systems Research Center, University of Minnesota, 21(1), 57-80. doi:10.2307/249742

- Kilmer, J., & Salinas, P. (s.f.). Arboles de Decisión.
- Krogh, G., & von. (2002). The Communal Resource and Information System. Journal of Strategic Information System 11, 85-107. Obtenido de <http://www.sciencedirect.com/science/article/pii/S0963868702000069>
- Laudon, K., & Jane, L. (2008). Sistemas de Información Gerencial. Administración de la Empresa Digital (10ma. ed.). (P. Hall, Ed.) México.
- Lazzati, S. (2013). La Toma de Decisiones. Principios, procesos y Aplicaciones. (1era. ed.). Buenos Aires, Argentina: Granica S.A.
- Lopez. (2014). Otra Metodología para la gestión de Proyectos. Recuperado el 2016, de <http://eoи.es/blogs/mcalidadon/2014/12/20/prince2-otra-forma-para-la-gestion-de-proyectos/>
- Marboleda. (2017). Obtenido de <http://www.hablemosdemarcas.com/estadisticas-de-redes-sociales-en-ecuador-2017/>
- Martin, J. (1989). Organización de las Bases de Datos. Obtenido de wpalma: <http://zeus.inf.ucv.cl/~jrubio/docs/2009-01/INF%20340/Capitulo%20I.pdf>.
- Martinez, A. (2001). Bases metodológicas para evaluar la viabilidad y el impacto de proyectos de telemedicina. Madrid, España: Universidad Politécnica de Madrid.
- Martinez, F. (10 de Noviembre de 2014). ISO 21.500 El Lenguaje De La Dirección De Proyectos. Obtenido de <http://www.coaatgr.es>
- Martínez, R. E., Ramírez, N. C., Mesa, H. G., Suárez, I. R., Trejo, M. D., León, P. P., & Morales, S. L. (2009). Arboles de decisión como Herramienta en el Diagnóstico Médico. Veracruz, México. Obtenido de <http://www.medigraphic.com/pdfs/veracruzana/muv-2009/muv092c.pdf>
- Martínez, S., & R., G. (2009). Integración de los algoritmos de minería de datos 1R, PRISM e ID3 a PostgreSQL. Obtenido de Scielo: http://www.scielo.br/scielo.php?pid=S1807-17752013000200389&script=sci_arttext&tlang=pt#B12
- Medina Chicaiza, R. P. (2004). INTERFAZ, INTELIGENCIA DE NEGOCIOS y PYMES. Obtenido de <http://repositorio.pucesa.edu.ec/handle/123456789/1368>
- Mitchell, T. (1997). Machine Learning (Mc Graw Hill ed.). Illinois, EEUU: University of Illinois.
- Morales, S., Morales, M., & Rizo, R. (2016). Implementación de Metodología IN, con optimización en ETL usando tecnología In Memory. I Congreso Internacional de Software Aplicado a la Agricultura, Turismo e industria SAATI. Milagro. Obtenido de <http://cidecuador1.wixsite.com/sistemas-milagro/memorias>
- Morales, S., Morales, M., & Rizo, R. (2017). Metodología para los procesos de inteligencia de negocios con mejoras en extracción y transformación de fuentes de datos. II Congreso Internacional. Ciencia, Sociedad e Investigación Universitaria, desarrollado en la Pontificia Universidad Católica del Ecuador. AMBATO. Obtenido de <http://repositorio.pucesa.edu.ec/bitstream/123456789/2354/1/Metodolog%C3%ada.pdf>
- Morales, S., Morales, M., & Rizo, R. (2018). Sistemas de Información e Inteligencia Artificial. Quito. Obtenido de <http://uce-ing-informatica.blogspot.com/2018/02/i-jornadas-tematicas-en-sistemas-de.html>
- Moreno, A. (1994). Aprendizaje Automático. Barcelona, España: Edicions UPC. doi:ISBN: 9788483019962
- Moreno, S. (2012). Primera Forma Normal. Recuperado el 07 de 12 de 2016, de <http://prezi.com/nanwvndiwhjy/forma-normal/>.
- Multiplicalia. (Febrero de 2017). Multiplicalia. Obtenido de <https://www.multiplicalia.com/redes-sociales-mas-usadas-2017/>
- Negash, S. (2004). Business Intelligence. Communications of the Association for Information Systems, 13, 177-195. Obtenido de <http://aisel.aisnet.org/cgi/viewcontent.cgi?article=3234&context=cais>

- Orozco, J., & Soriano, F. (2016). Study of Intelligence in Business network used more. PMI.
- Parr, O. R., & Sons. (2009). Business Intelligence Success Factors.
- PDCA, C. (2013). El círculo de Deming de mejora continua.
- Peng, W., Chen, J., & Zhou. (2012). An Implementation of ID3. TREE LEARNING ALGORITHM. Sidney, Australia: University of New South Wales, School of Computer Science & Engineering. Obtenido de <http://cis.k.hosei.ac.jp/~rhuang/Miccl/AI-2/L10-src/DecisionTree2.pdf>
- Peña, J. A. (22 de Noviembre de 2012). PMBOK. Obtenido de Universidad Autónoma de la ciudad de México: <http://pmbokuacm.weebly.com/13/post/2012/11/explicacion-acerca-del-pmbok.html>
- Pérez Cesar, S. D. (2008). Conceptos, técnicas y sistemas. En Minería de datos técnicas y herramientas. Madrid.
- Pino, R., Gómez, A., & de Abajo, N. (2001). Aprendizaje automático. En Introducción a la inteligencia artificial(8). Oviedo: Servicios de Publicaciones Universidad de Oviedo.
- PMI. (2017). Que es PMI. Recuperado el 19 de 04 de 2017, de <https://americalatina.pmi.org/latam/aboutus/whatispmi.aspx>
- Pyle, D. (2003). Business Modeling and Data Mining. Morgan Kaufmann Publishers.
- Quinlan, J. (1986). Induction of decision trees. Machine learning. Kluwer Academic Publishers. Obtenido de <https://doi.org/10.1007/BF00116251>
- Ranjan, J. (2009). Business Intelligence: Concepts, components, Techniques and benefits. Journal of Theoretical and Applied Information Technology, 70. Obtenido de <http://www.jatit.org>
- Reinschmidt, J., & Allison, F. (2000). Business Intelligence Certification Guide. California, EEUU: IBM International Technical Support Organisation. doi:SG24-5747-00
- Rivadera, G. (2010). La Metodología de Kimball para el Diseño de Almacenes de Datos.
- Rodallegas Ramos Erika, T. G. (2010). Minería de Datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y los k vecinos más cercanos. Puebla, México: Universidad Tecnológica de Izúcar de Matamoros.
- Rodriguez Mogollón, S. (2009). Gestión de Polizas de Seguros. Un caso práctico de Business Intelligence. Madrid, España: Politecnica Superior. Ingeniería Informática. Obtenido de <https://core.ac.uk/download/pdf/29401348.pdf>
- Rokach, L., & Maimon, O. (2008). Data mining with decision trees: theory and applications. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Rosado Gomez, A. A. (01 de 04 de 2010). Inteligencia de Negocios. (U. T. Pereira, Ed.) Scientict Technica AÑO.VVI. No. 44, 44, 321-326. Obtenido de <http://www.redalyc.org/html/849/84917316060/>
- Salas, R. (2004). Redes Neuronales Artificiales. Valparaiso, Argentina: Departamento de Computación.
- Salinas La Rosa, A. (2010). Inteligencia de Negocio. Auditoría y Control. Prototipo de Herramienta de Calidad (Vol. 1). (D. D. INFORMATICA, Ed.) Leganes, España: UNIVERSIDAD CARLOS III DE MADRID. Obtenido de <https://core.ac.uk/download/pdf/29402111.pdf>
- Sánchez, J. (2004). Principios sobre bases de datos relacionales. California, Eeuu: Creative Commons. Recuperado el 2017, de <http://cursa.ihmc.us/ríd=1H73QYLH3-6LFRCX-JT6/bdrelacional.pdf>
- Santana, B. (2011). Primeros pasos hacia la Inteligencia de Negocios.
- Sempere, J. (2014). Aprendizaje de árboles de decisión. Valencia, España. Recuperado el 18 de 12 de 2016

- Silberschatz, A., Korth, H. F., Sudarshan, S., Pérez, F. S., Santiago, A. I., & &. S. (2002). Fundamentos de Bases de Datos (4ta. ed., Vol. 4ta.). Madrid, España: Concepción fernandez Madrid.
- Stackowiak, R., Greenwald, R., & Rayman, J. (2007). Oracle Data Warehousing and Business Intelligence
- Surajit, C., Umeshwar, D., & Narasaya, &. (2011). BI technologies are essential to running today's business and this technology is going through sea changes. Communications of the ACM.
- Thompson, I. (10 de 2008). Definición de Información. Recuperado el 2017, de https://scholar.google.es/scholar?q=THOMPSON++definicion+de+informacion&btnG=&hl=es&as_sdt=0%2C5:
- Tom, M. (1997). An Overview of Machine Learning. (S. link, Ed.) EEUU: Routger University. Recuperado el 2015, de <https://link.springer.com/book/10.1007/978-3-662-12405-5>
- Torres Hernández, Z. (2014). Administracion de Proyectos (1ra Edición ed.). Mexico. D.F., Mexico: Patria.
- Tseng, F. S., & H, C. A. (2006). The concept of document warehousing for multi-dimensional modeling of textual-based business intelligence. Decision Support System, 727-744.
- Ullman, J. D., Widom, J., & Miguel, E. A. (1999). Introducción a los Sistemas de Base de Datos. (P. Hall, Ed.) doi:No. 001-6-U55Y
- Urdaneta,Elymir (2001). El Data Mining.
- Valdez Aguliar, P. A. (2016). Guia de Gestión de Proyectos. Ciudad de México, México: UPIICSA. Obtenido de <http://tesis.ipn.mx/bitstream/handle/123456789/20367/GUIA%20DE%20GESTION%20DE%20PROYECTOS.pdf>
- Vanrell, J. Á., Bertone, R. A., & Ramón, G. M. (2010). Modelo de Proceso de Operación para Proyectos de Explotación de Información. In XVI Congreso Argentino de Ciencias de la Computación. Universidad De Lanus, Grupo De Investigacion En Sistemas De Informacion. Departamento De Desarrollo Productivo Y Tecnologico., Lanus. Obtenido de <http://sedici.unlp.edu.ar/handle/10915/19328>
- Verástegui, J. M. (22 de Junio de 2014). ISO 21500 Directrices para la Dirección y Gestión de Proyectos. Recuperado el 02 de 02 de 2017, de IV Congreso Internacional de Direccion de Proyectos: <http://www.iso-21500.es>
- Vercellis, C. (2009). Business Intelligence: data Mining and Otimizationfor Decision Making. Milano, Italia: Jhon Wiley & Sons Ltd. Obtenido de <http://wiley.com>
- Vila Grau, J. (2015). El Método PRINCE" en menos de 1000 palabras. Obtenido de <https://es.slideshare.net/benq2011/mp-el-mtodo-prince2-en-menos-de-mil-palabras-46041095>
- Vilchez, M. &. (2011). Desarrollo de un modelo de inteligencia de negocios usando Data Mining para optimizar la toma de decisiones en los procesos de ventas en la empresa San Roque SA Universidad Señor de Sipán Facultad de ingeniería. Escuela Profesional Ingeniería de sistemas. (E. d. Sistemas, Ed.) Chiclayo, Perú: Universidad Señor de Sipán. Obtenido de <http://repositorio.uss.edu.pe/handle/us/387>
- Villalon M, D. M. (2006). Recomendaciones para la Implementación de BI, basado en Data Warehouse. Técnico de Monterrey. Mexico DF, MEXICO. Obtenido de <http://hdl.handle.net>
- Westerman, P. (2001). Data warehousing: using the Wal-Mart model. Morgan Kaufmann. EEUU: Academic Press. Obtenido de <http://www.academicpress.com>
- Wixom, B. H., & Watson, H. J. (2014). The Current state of Business Intelligence Of Academia. (Q. U. Technology, Ed.) Comunication of the Association for Information Systems.
- Zarate Gallardo, E. d. (2013). Inteligencia de Negocios. (<https://www.gestiopolis.com/inteligencia-de-negocios/>, Editor) Recuperado el 20 de 03 de 2017.