



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

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

CENTRO DE POSGRADOS

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

**TEMA: “MANEJO DE INDICADORES BÁSICOS FINANCIEROS Y DE
GESTIÓN MEDIANTE LA APLICACIÓN DE UN MODELO DE
GESTIÓN DE DATOS – CASO DE ESTUDIO INSTITUTO MANUEL
CÓRDOVA GALARZA”**

AUTORA: SUNTAXI IMBAQUINGO, TATIANA GABRIELA

DIRECTOR: JOSÉ ALBERTO SANCHO ARIAS

SANGOLQUÍ

2019



ESPE

UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

CERTIFICACIÓN

Certifico que el trabajo de titulación "MANEJO DE INDICADORES BÁSICOS FINANCIEROS Y DE GESTIÓN MEDIANTE LA APLICACIÓN DE UN MODELO DE GESTIÓN DE DATOS – CASO DE ESTUDIO INSTITUTO MANUEL CÓRDOVA GALARZA" fue realizado por la señorita Suntaxi Imbaquingo, Tatiana Gabriela, el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 01 de Abril de 2019

Firma:

José Alberto Sancho Arias
Director



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORIZACIÓN

Yo, **Suntaxi Imbaquingo, Tatiana Gabriela**, autorizé a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Manejo de indicadores básicos financieros y de gestión mediante la aplicación de un modelo de gestión de datos – Caso de Estudio Instituto Manuel Córdova Galarza”**, en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 01 de Abril de 2019

Firma:

A handwritten signature in blue ink, enclosed in a blue oval, which reads 'Tatiana Suntaxi Imbaquingo'.

Tatiana Gabriela Suntaxi Imbaquingo
CC. 1723959258



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

VICERRECTORADO DE INVESTIGACIÓN, INNOVACIÓN Y
TRANSFERENCIA DE TECNOLOGÍA

CENTRO DE POSGRADOS

AUTORIZACIÓN

Yo, **Suntaxi Imbaquingo, Tatiana Gabriela**, autorizó a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **“Manejo de indicadores básicos financieros y de gestión mediante la aplicación de un modelo de gestión de datos – Caso de Estudio Instituto Manuel Córdova Galarza”**, en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 01 de Abril de 2019

Firma:

A handwritten signature in blue ink, enclosed in a blue oval. The signature appears to read 'Tatiana Suntaxi'.

Tatiana Gabriela Suntaxi Imbaquingo
CC. 1723959258

DEDICATORIA

A mi querida familia,

Con todo mi cariño y amor a las personas que han estado conmigo en cada aprendizaje y obstáculo de mi vida, que han hecho todo por lograr mi bienestar en todo los sentidos, señalándome un camino y aconsejándome a cada paso.

A mi padre Carlos, mi madre Sandra y mi hermana Carla, por estar siempre presentes en cada momento, objetivo y meta planteada por su apoyo, motivación y amor, mi familia Uds. son el motor de mi vida por eso les dedico el fruto de nuestro esfuerzo.

Y, a cada una de las personas que han sido parte de este proceso, compañeros y amigos, con los cuales compartí más que un aula de clases, vivimos anécdotas, enseñanzas que me permitieron crecer como persona y profesional, por su ayuda, apoyo y la gran amistad que formamos.

AGRADECIMIENTO

A Dios,

Por la vida que me entrego, la familia y amigos que tengo, por bendecirme para cumplir esta meta planteada años atrás, gracias por brindarme la fuerza para superar los obstáculos y tener una culminación exitosa.

A mis padres,

Quien con todo su cariño, respeto y enseñanzas me han entregado más de lo que he necesitado, impulsando a ser mejor persona y crecer como profesional, mi más sincero agradecimiento ya que sin Uds. no estaría donde estoy.

Al Instituto Manuel Córdova Galarza,

Por la apertura recibida y el apoyo constante en la ejecución del trabajo, gracias a uds. pudimos concluir el proyecto de la mejor manera.

A mi querida Universidad,

En la que he pasado varios años de mi vida, me ha formado no solo como profesional sino como ser humano, me brindo las mejores enseñanzas y los mejores amigos, siempre agradeceré haber tenido la oportunidad de haberme formado entre sus muros.

Finalmente, gracias a todas las personas que de una u otra forma formaron parte de este proceso.

ÍNDICE DE CONTENIDOS

CERTIFICADO DEL DIRECTOR.....	i
AUTORÍA DE RESPONSABILIDAD	ii
AUTORIZACIÓN.....	iii
DEDICATORIA	iv
AGRADECIMIENTO	v
ÍNDICE DE CONTENIDOS	vi
ÍNDICE DE TABLAS.....	ix
ÍNDICE DE FIGURAS.....	x
RESUMEN.....	xiv
ABSTRACT	xv
CAPITULO I	
GENERALIDADES	
1.1. Introducción	1
1.2. Antecedentes	1
1.3. Delimitación Geográfica	4
1.4. Justificación, Importancia y Alcance	5
1.5. Contexto del Problema.....	6
1.6. Planteamiento del Problema.....	7
1.7. Objetivos	8

1.7.1. Objetivo General.....	vii
1.7.1. Objetivo General.....	8

1.7.2. Objetivos Específicos	9
------------------------------------	---

CAPITULO II

MARCO TEORICO

2.1. Antecedentes Investigativos.....	10
---------------------------------------	----

2.2. Estado del Arte.....	10
---------------------------	----

2.3. Antecedentes Conceptuales.....	16
-------------------------------------	----

2.3.1. Business Intelligence.....	16
-----------------------------------	----

2.3.2. Minería de datos	17
-------------------------------	----

2.3.3. Indicadores	28
--------------------------	----

2.3.4. Metodología Design Science Research.....	29
---	----

CAPITULO III

MEMORIA TÉCNICA METODOLÓGICA

3.1. Metodología de la Investigación	30
--	----

CAPITULO IV

DESARROLLO

4.1. Ejecución del proceso de investigación	33
---	----

4.1.1. Conocimiento del Problema.....	33
---------------------------------------	----

4.1.2. Sugerencia.....	35
------------------------	----

4.1.3. Desarrollo.....	36
------------------------	----

	viii
4.1.4. Evaluación.....	40
4.1.5. Conclusión.....	40
4.2. Guía Metodológica.....	40
4.3. Construcción del modelo de Gestión de Datos	42
4.3.1. Análisis y selección de datos.....	42
4.3.2. Preparación de los datos.....	44
4.3.3. Procesos ETL	46
4.3.4. Base de datos.....	54
4.3.5. Modelo de Minería	57
4.3.6. Modelo de Minería de datos.....	65
4.3.7. Evaluación del modelo.....	72
4.3.8. Dashboard.....	73

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones	90
5.2. Recomendaciones.....	92
BIBLIOGRAFÍA.....	93

ÍNDICE DE TABLAS

Tabla 1. <i>Estudios por Grupo de Control</i>	11
Tabla 2. <i>Construcción de Cadena de Búsqueda</i>	14
Tabla 3. <i>Asignación monto de gastos</i>	35
Tabla 4. <i>Datos entregados por Izquierda Democrática</i>	42
Tabla 5. <i>Comparación de herramientas ETL</i>	46
Tabla 6. <i>Cuadro Comparativo de Herramientas de Data Mining</i>	57
Tabla 7. <i>Atributos seleccionados por el Auto Model</i>	65
Tabla 8. <i>Atributos de entrada escogidos</i>	65
Tabla 9. <i>Cuadro comparativo de Herramientas líderes de BI</i>	73
Tabla 10. <i>Cuadro comparativo de Herramientas líderes de BI open source</i>	74

ÍNDICE DE FIGURAS

<i>Figura 1.</i> Organigrama Nacional.....	2
<i>Figura 2.</i> Reseña Histórica de los Presupuestos.....	4
<i>Figura 3.</i> Localización Geográfica.....	5
<i>Figura 4.</i> Cuadrante Mágico de Gartner de Herramientas de BI.....	17
<i>Figura 5.</i> Procesos KDD	20
<i>Figura 6.</i> Estructura de una red neuronal	23
<i>Figura 7.</i> Ejemplo de Clustering	25
<i>Figura 8.</i> Cuadrante Mágico de Gartner para las ciencias de la información	26
<i>Figura 9.</i> Tipos de artefactos.....	29
<i>Figura 10.</i> Design Science Research Process Mode	30
<i>Figura 11.</i> Financiamiento IZ.....	34
<i>Figura 12.</i> Archivo Excel – Balance de Comprobación.....	37
<i>Figura 13.</i> Archivo Excel – Libro Mayor.....	37
<i>Figura 14.</i> Archivo Excel- Plan de cuentas	38
<i>Figura 15.</i> Archivo Excel – Libro bancos	38
<i>Figura 16.</i> Archivo Excel – Estado de Situación financiera	39
<i>Figura 17.</i> Limpieza de datos en Pentaho	39
<i>Figura 18.</i> Proceso de ejecución del trabajo.....	41
<i>Figura 19.</i> Diagrama Entidad – Relación.....	44
<i>Figura 20.</i> Diagrama Modelo Multidimensional.....	45
<i>Figura 21.</i> Proceso ETL dimensión fecha – Fuente destino.....	47
<i>Figura 22.</i> Proceso Extracción y Carga dimensión fecha	48

	xi
Figura 23. Proceso ETL dimensión Plan Cuenta – Fuente Origen	48
Figura 24. Proceso de Extracción y Carga dimensión Plan Cuenta.....	49
Figura 25. Clasificación por cuentas contables.....	49
Figura 26. Clasificación por grupo de indicadores	50
Figura 27. Proceso ETL Plan Cuentas – Fuente Destino	51
Figura 28. Proceso ETL Tabla de hechos – Fuente Origen	51
Figura 29. Proceso de extracción y carga Tabla de Hechos.....	52
Figura 30. Cálculo de monto por cada transacción	52
Figura 31. Montos transaccionales por grupo de indicadores.....	53
Figura 32. Proceso ETL Tabla de hechos – Fuente Destino	53
Figura 33. XAMPP.....	54
Figura 34. Creación de la base de datos en MySQL	54
Figura 35. Conexión con la base de datos MySQL mediante Data Integration de Pentaho	55
Figura 36. Creación de tabla Dim_fecha en base de datos MySQL	55
Figura 37. Creación de tabla Dim_Plan_cuenta en base de datos MySQL	56
Figura 38. Creación de la tabla de hechos en la base de datos MySQL	56
Figura 39. Base de datos “base_datos_tesis_v2_18” con todas las tablas creadas	57
Figura 40. RapidMiner.....	58
Figura 41. Proceso RapidMiner, carga de base de datos.....	59
Figura 42. Tabla de datos	60
Figura 43. Paso 1: Auto Model – Carga de datos	60
Figura 44. Paso 2: Auto Model – Seleccionar tarea.....	61
Figura 45. Paso 3: Auto Model – Preparar objetivo.....	62

	xii
Figura 46. Paso 4: Auto Model – Seleccionar entradas	62
Figura 47. Paso 5: Auto Model – Tipo de Modelos	63
Figura 48. Paso 6: Auto Model – Resultados.....	64
Figura 49. Proceso creado por Auto Model	64
Figura 50. Tabla de Datos “Base_Fac_Mayor_v3_18	66
Figura 51. Selección de datos para el modelo.....	66
Figura 52. Selección de atributo a predecir.....	67
Figura 53. Opciones de la variable a predecir.....	67
Figura 54. Atributos de entrada.....	68
Figura 55. Comparación de algoritmos.....	68
Figura 56. Resultado con algoritmo Naive Bayes.....	69
Figura 57. Gráfico Naive Bayes	69
Figura 58. Resultado con algoritmo Decision Tree	70
Figura 59. Proceso de Evaluación de Modelo.....	72
Figura 60. Matriz de precisión	73
Figura 61. Descarga Power BI Desktop Gratuito.....	75
Figura 62. Carga de datos en Power BI.....	75
Figura 63. Datos en Power BI.....	76
Figura 64. Asignación de monto de gastos	76
Figura 65. Comparación de gastos	77
Figura 66. Análisis de proveedores - Publicaciones	78
Figura 67. Análisis de proveedores – Capacitación e Investigación.....	79
Figura 68. Análisis de proveedores – Formación.....	79

	xiii
Figura 69. Análisis de proveedores – Funcionamiento Institucional	80
Figura 70. Comparación de gastos incurridos vs. los presupuestados	81
Figura 71. Comportamiento de gastos a marzo de 2019	82
Figura 72. Cuentas de Pasivos	83
Figura 73. Cuentas contables – Funcionamiento Institucional 2018	84
Figura 74. Cuentas contables – Funcionamiento Institucional 2019	85
Figura 75. Proveedores por grupo de gasto.....	86
Figura 76. Alerta de consumo de partidas presupuestarias – Capacitación e Investigación.....	87
Figura 77. Alerta de consumo de partidas presupuestarias – Formación.....	87
Figura 78. Alerta de consumo de partidas presupuestarias – Funcionamiento Institucional	88
Figura 79. Alerta de consumo de partidas presupuestarias – Publicaciones.....	89

RESUMEN

Las entidades públicas y privadas tienen como objetivo ser eficientes en la ejecución de sus procesos administrativos-financieros, un medio de evaluación son los indicadores que contemple la revisión de la estrategia del negocio. La gestión que realiza el Instituto Manuel Córdova Galarza de los recursos económicos que el CNE¹ asigna al Partido Izquierda Democrática, entidad que rige bajo la Ley Orgánica de Control del gasto electoral y de la propaganda electoral, la cual dictamina los gastos que les permiten realizar, si la Institución no cumple con la normativa legal serán sancionados con llamados de atención o posible suspensiones, este estudio reduce el riesgo de sanciones que resulten en la disminución del presupuesto asignado lo que provocaría el incumplimiento de los eventos y/o actividades planificadas, con la implementación de un modelo de datos que cumpla con los requerimientos institucionales y legales mediante la aplicación de indicadores dentro de los Sistemas de Información, el proceso de ejecución se lo realizo con la metodología Design Science Research, la cual propone 5 pasos: conocimiento del problema, sugerencia, desarrollo, evaluación y conclusión, con el desarrollo de estas etapas tenemos como resultado un Modelo de Gestión de datos que disminuyen los tiempos de análisis y ejecución de los datos del Partido, agilitando los procesos y tomando las acciones correctivas de ser el caso.

PALABRAS CLAVES:

- **INDICADORES**
- **SISTEMAS DE INFORMACIÓN**
- **MODELO DE GESTIÓN DE DATOS**

¹ Consejo Nacional Electoral: Es el máximo organismo de sufragio en dicho país. Sus funciones son organizar, controlar las elecciones, puede castigar a partidos y candidatos que infrinjan las normas electorales.

ABSTRACT

The public and private entities aim to be efficient in the execution of their administrative-financial processes, a means of evaluation are the indicators that include the review of the business strategy. The management carried out by the Manuel Cordova Galarza Institute of the economic resources that the CNE assigns to the Political Party, entity that governs the Organic Law of Control of electoral spending and electoral propaganda, which dictates the expenses that allow them to perform, if the Institution does not comply with the legal regulations will be sanctioned with calls for attention or possible suspensions, this study reduces the risk of sanctions that result in the reduction of the allocated budget which would cause the non-compliance of the events and/or planned activities, with the implementation of a data model that complies with the institutional and legal requirements through the application of indicators within the Information Systems, the execution process was carried out with the Design Science Research methodology, which proposes 5 steps: problem awareness, suggestion, development, evaluation and conclusion, with the development of these stages we have as a result a Data Management Model that reduces the analysis and execution times of the organization's data, streamlining the processes and taking the corrective actions if necessary.

KEYWORDS:

- **INDICATORS**
- **INFORMATION SYSTEMS**
- **DATA MANAGEMENT MODEL**

CAPITULO I

GENERALIDADES

1.1. Introducción

Las entidades tanto públicas, privadas o de economía mixta tienen como objetivo ser eficientes en la ejecución de sus procesos administrativos-financieros, siendo necesario realizar evaluaciones periódicas con base a la información de cada organización para poder tomar decisiones oportunas o acciones correctivas de ser el caso.

Un medio de evaluación es el uso de los indicadores financieros y de gestión que contemple la revisión desde el nivel corporativo, funcional y la estrategia del negocio en el Caso de estudio – Instituto Manuel Córdova Galarza, dicha entidad rige bajo la Ley Orgánica de Control del gasto electoral y de la propaganda electoral, en la cual dictamina que la Institución responde al proceso obligatorio de formación política en cual esta enlazado a un control administrativo-financiero. El organismo no cuenta con un Sistema de Información específico para el análisis inteligente, elaborar los reportes para los respectivos organismos de control, por lo cual no puede controlar de manera adecuada el riesgo de minimización del presupuesto que es asignado por el Consejo Nacional Electoral (CNE) para la ejecución de sus actividades, si esta asignación fuese reducida el organismo dejaría de cumplir con los eventos y/o actividades planificadas; es por esta razón que el presente estudio responde a una necesidad institucional prioritaria.

1.2. Antecedentes

El Instituto Manuel Córdova Galarza, forma parte del Partido Político Izquierda Democrática, por medio del mismo se gestionan las transacciones económicas que realiza, es decir la asignación de los gastos para su funcionamiento y crecimiento en el mercado; se puede observar en la **Figura 1. Organigrama Nacional**, en el cual el Instituto forma parte del Organismo Político.

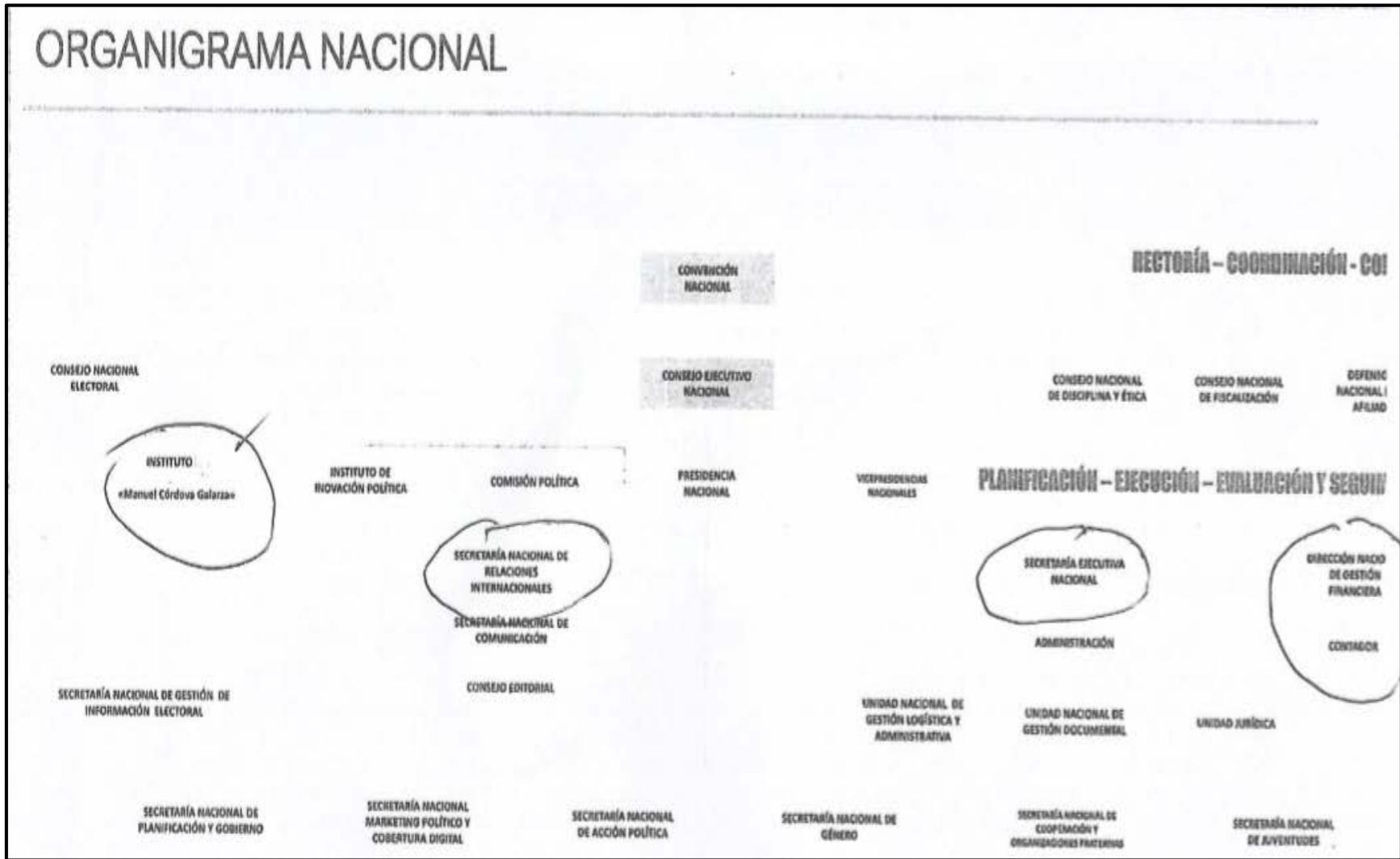


Figura 1. Organigrama Nacional
 Fuente: (ID, 2018)

El Partido Político Izquierda Democrática tiene como objeto generar la participación de los ciudadanos en la toma de decisiones políticas, realizando capacitaciones, a nivel local, provincial, regional y nacional, acorde con la normativa emitida por el CNE. La normativa del CNE indica que los miembros activos, pasivos, simpatizantes y la comunidad en general, deben ser capacitados en los temas fundamentales de conocimiento político, social, cultural, sociológico, laboral y las tendencias de la participación ciudadana como actores fundamentales entre el estado y sus ciudadanos.

Las actividades económicas que realizan se basan en los siguientes ejes:

- Eventos de difusión, análisis y participación de diferentes actores internos y externos para conocer temas de interés nacional.
- Programas de capacitación con diferentes metodologías y formas organizativas, tales como: charlas, conferencias, simposios, talleres, mesas redondas para los afiliados, simpatizantes y la ciudadanía en general
- Procesos de capacitación sistemática de competencias políticas, culturales, sociales y de la cultura popular.
- Capacitación a través del uso de las TICs para masificación de la imagen corporativa a la cual representa el organismo.

La organización cuenta con presupuesto compuesto por recursos públicos y privados que permiten el mantenimiento de la infraestructura, los procesos administrativos y recursos humanos, así como el manejo desde la perspectiva financiera, económica y social de la institución.

De igual forma es importante mencionar los antecedentes de los presupuestos, su manejo, control y evolución que permita entender la importancia de una correcta implementación, ya que esta ayuda a tomar decisiones a la alta gerencia.

El origen de los presupuestos data de 1820, Francia adoptó el sistema de presupuestar en el sector gubernamental y Estados Unidos lo acogió en 1821 como elemento de control de gastos públicos y como herramienta para los funcionarios cuyo trabajo era presupuestar, para garantizar el funcionamiento eficiente de las actividades gubernamentales (Fandom, 2015)

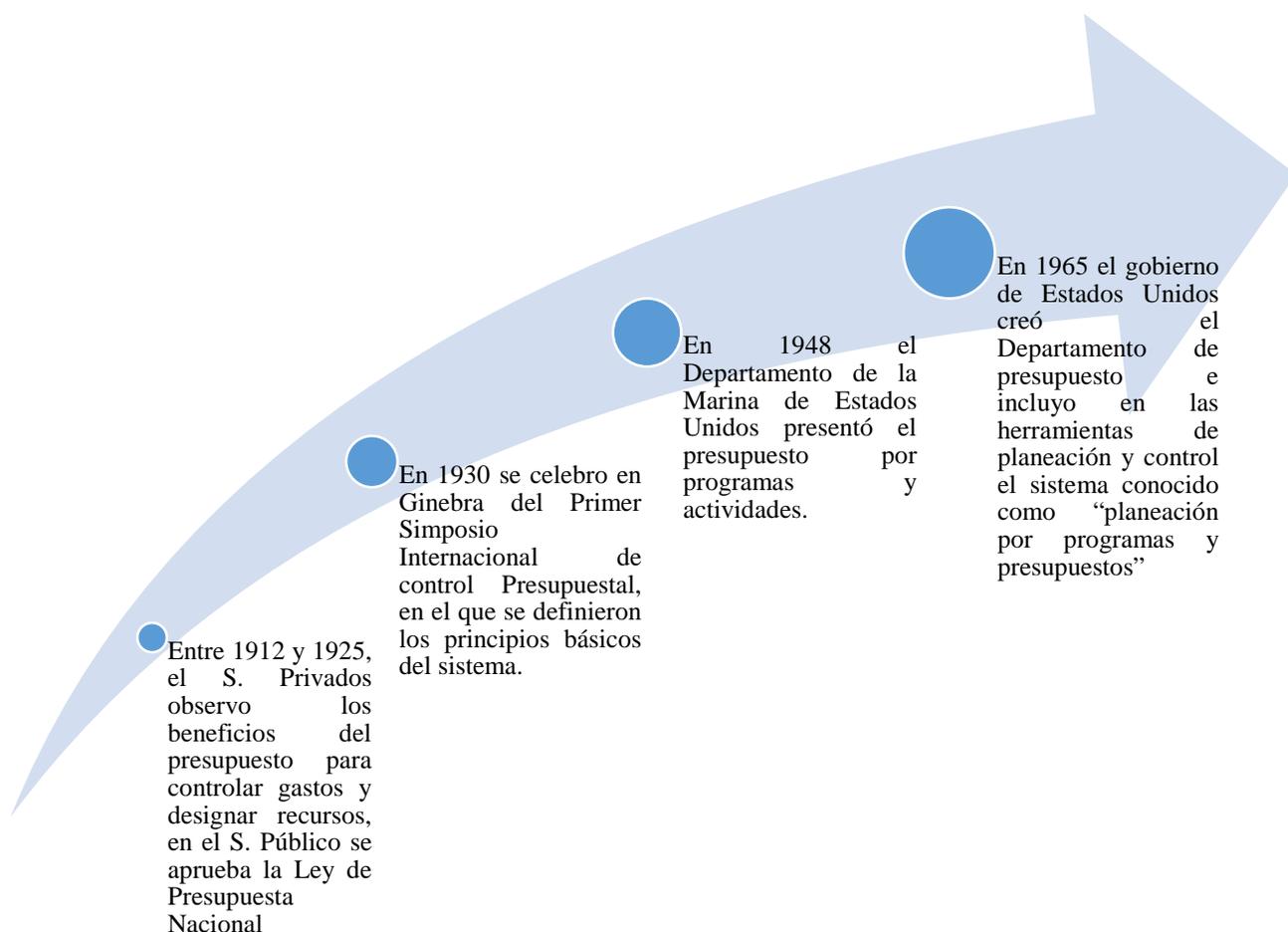


Figura 2. Reseña Histórica de los Presupuestos

Fuente: (Fandom, 2015)

1.3. Delimitación Geográfica

La localización geográfica del proyecto se encuentra en Provincia Pichincha, Cantón Quito, dirección: Polonia N30-83 y, Vancouver, como se muestra en la **Figura 3.** Localización Geográfica

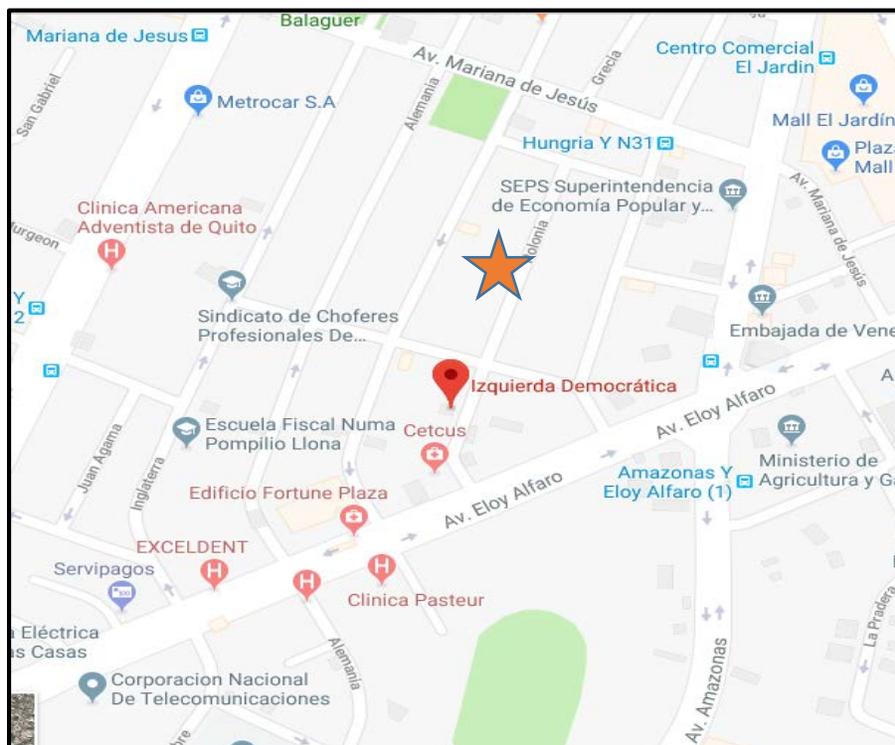


Figura 3. Localización Geográfica

Fuente: Google Maps

1.4. Justificación, Importancia y Alcance

La justificación del tema de estudio se basó en minimizar el riesgo de sanciones por el ente regulador (CNE) debido al incumplimiento de los requisitos impuestos; las cuales son: llamados de atención verbales, suspensión de las actividades y reducción del presupuesto asignado al Partido, lo que impediría el cumplimiento de todos los eventos y/o actividades planificadas.

La necesidad de la organización de obtener información oportuna y adecuada, para analizarla y evaluarla mediante una aplicación de indicadores financieros y de gestión. En base a esta aplicación se buscó dar apoyo a la toma de decisiones de las máximas autoridades, de no realizar esta automatización y análisis de la información las decisiones se seguirían tomando de acuerdo a los hechos o situaciones que se presenten, es decir una vez ocurrido o detectado un problema o desviación se toman acciones y no se prevén adecuadamente los posibles riesgos que pueden ocurrir.

El presente trabajo fue un aporte directo a la Institución, la implantación del Modelo de gestión de datos se realiza un adecuado control interno, se identificó patrones de comportamiento con el objetivo de promover una administración exitosa, además de que se evalúa el cumplimiento de su planificación y la normativa gubernamental que lo rige, este estudio también sirve como base y/o referencia para próximos estudios de instituciones con similar actividad económica.

El alcance de este proyecto es el análisis de indicadores básicos con el apoyo de las herramientas tecnológicas que permitió el desarrollo de los reportes prediseñados para responder a las necesidades de los usuarios internos y externos.

Es importante considerar que en la actualidad el Instituto cuenta con un Modelo de Gestión de datos que permitan realizar un análisis adecuado y oportuno, por lo que la presente propuesta puede ser considerada como un prototipo de modelo analítico de evaluación de desempeño mediante indicadores financieros y de gestión que se realizaría con la información de la organización elegida como caso de estudio, y podrá servir a otras entidades con similares actividades económicas.

1.5. Contexto del Problema

El partido Izquierda Democrática es una entidad de economía mixta, cuyo eje central de capacitación política, social, cultural, debe ser un referente del conocimiento público en estas competencias. Por la normativa a la que se rige “*Reglamento de Fondo partidario permanente CNE- Resolución PLE-CNE-3-3-2018*”, la rendición periódica de cuentas es fundamental a su militancia, parte directriz, a los organismos de control y transparentar su accionar a la comunidad en general e inclusive, a los medios de comunicación.

Para cumplir con estos objetivos cuenta con una estructura orgánico funcional, con una planificación estratégica para los procesos de capacitación y con un presupuesto, recibido por organismos de control público (Consejo Nacional Electoral) y gestionado a través del Instituto

Manuel Córdoba Galarza; y recursos privados que pueden provenir de entes nacionales e internacionales, así como personas naturales y jurídicas. A esto se suman las políticas públicas de transparencia en el manejo de los recursos, la obligatoriedad actual como política de estado, de asegurar la difusión de esta información con medios tecnológicos, que puedan ser accedidos desde cualquier dispositivo, que es obligación de la organización cumplir con adecuada y oportunamente con esta disposición legal ya que de no hacerlos sus actividades serán suspendidas.

La información, por normativa, transparencia y por manejo de estándares locales, nacionales e internacionales, necesita en forma indispensable, el procesamiento de los sistemas de información para cumplimiento de estos requerimientos, si la institución no cumple con los requerimientos gubernamentales sufrirá un recorte de presupuesto lo que pone en riesgo la ejecución de las actividades planificadas.

En este contexto, la parte directriz ha diseñado modelos de procesamiento de información aprobados a nivel nacional y que todas las organizaciones políticas deben cumplirlas y además implementarlas para poder ser competitivas.

La prioridad de este proyecto coincide con la necesidad inmediata a corto plazo de implementarse aplicando estándares de inteligencia en el manejo de la información, o cualquier estrategia tecnológica que con el debido sustento metodológico para la rendición de cuentas periódica y permanente en un contexto del mundo globalizado y que además necesita la transparencia y conocimiento de la comunidad en general.

1.6. Planteamiento del Problema

El Partido Político Izquierda Democrática mediante la gestión del Instituto Manuel Córdoba Galarza cumple con los requerimientos de control de las entidades regulatorias CNE, evaluando su gestión mediante el uso de indicadores, solicitados por:

- Organismos de control público (CNE)
- Entidades de tipo privado nacional
- Entidades de tipo privado internacional

Además, en el contexto de la problemática, la militancia, la dirigencia aprobó como prioridad en su planificación estratégica, la implementación de procesos con la aplicación de la tecnología, basados en estándares nacionales e internacionales y de manejo inteligente de los sistemas de información.

Con los antecedentes descritos, si el Partido político no cumple con lo dispuesto en la normativa legal serán sancionados con: llamados la atención, ser observados y posteriormente suspendidos, lo cual significa el desmedro de la credibilidad interna, externa y pone en riesgo su permanencia local, regional y nacional, además la entidad regulatoria tendría la potestad de disminuir el presupuesto que se asigna en cada periodo determinado (1 año), de existir esta reducción el organismo, estaría limitando la ejecución de las actividades planificadas (eventos, charlas, seminarios y capacitaciones) y no conseguirían la aceptación y crecimiento en su segmento de mercado.

1.7. Objetivos

1.7.1. Objetivo General

Implementar un modelo de datos mediante técnicas de minería de datos para cumplir con los requerimientos institucionales internos y externos, de los organismos de control a través del uso aplicativo de indicadores dentro de los Sistemas de Información para la organización.

1.7.2. Objetivos Específicos

OE1: Analizar la información existente en la Planificación Estratégica y bases de datos del IMCG para determinar los indicadores de mayor impacto y riesgo, verificando la calidad y consistencia de los datos a ser utilizados en el modelo analítico.

OE2: Realizar una revisión de literatura para determinar herramientas y metodologías, estableciendo las más adecuadas e idóneas que aporten a la solución del problema planteado.

OE3: Implementar el modelo analítico para construir los valores de los indicadores financieros y de gestión seleccionados, que permitirán obtener la información consolidada de las operaciones de la empresa en las cuales se basarán las decisiones de las máximas autoridades.

OE4: Evaluar el modelo analítico, a través de la metodología cuantitativa, es decir si cumple o no el objetivo; y técnicas de validación de minería de datos que determine el nivel de confianza de los resultados obtenidos.

CAPITULO II

MARCO TEORICO

2.1. Antecedentes Investigativos

A continuación, se describen varios artículos, tesis y trabajos relacionados con el presente proyecto y los resultados obtenidos en cada uno, lo que permitió enfocar el presente proyecto en un contexto adecuado y utilizar ciertos procedimientos donde se obtuvieron resultados satisfactorios, para lograr esto se realizó el Estado del Arte.

2.2. Estado del Arte

El presente estudio del arte responde a las actividades iniciales de un SMS², que corresponde a los criterios de inclusión y la estrategia de búsqueda para ello se ha revisado bibliografía en los principales repositorios académicos siguiéndolas fases de:

Definición de Objetivo: En esta fase se relacionó el esquema que conducen al objetivo de la investigación planteada.

Definición de los criterios de inclusión y exclusión: Los criterios de inclusión y exclusión ubican los estudios con las características apropiadas al tema planteado. Estos criterios fueron planteados por quien conduce la investigación y discutido con los investigadores.

- Criterios de Inclusión

- Se incluyen trabajos a partir del 2010
- Se realizó filtros directos en los repositorios tomando como prioridad el tipo research article
- Se incluyen libros, tesis, capítulos de libros, artículos y ponencias, siempre y cuando tengan

disponibilidad en la web.

² Systematic Mapping Study (SMS): estudio de alcance que analiza un amplio conjunto de estudios primarios (artículos, publicaciones) para identificar qué y cuantas evidencias hay disponibles sobre un tema determinado.

- Criterios de Exclusión

- Se excluye contenido en idiomas diferentes al inglés o español.

Definición de la estrategia de búsqueda

Revisión Inicial: Se realizó una indagación previa revisando si existen estudios relacionados con las preguntas de investigación propuestas.

Validación cruzada de estudios: La validación cruzada permitió garantizar que los estudios cumplan los criterios de inclusión y exclusión, como resultado se pudo constatar que todos los trabajos cumplen con los criterios de inclusión y exclusión, y se procedió a realizar un listado de integración del grupo de control.

Integración del grupo de Control: El grupo de control conforman los estudios que cumplen con las características de la investigación analizados por los investigadores considerando el título del estudio, resumen y palabras claves. Los estudios del grupo de control analizados fueron los siguientes:

Tabla 1.
Estudios por Grupo de Control

Grupo de Control	Título	Palabras Clave
EC1	Análisis, diseño e implementación de una Solución Business Intelligence para la generación de indicadores y control de desempeño, en la empresa Otecel S.A., utilizando la metodología Hefesto V2.0	Business Intelligence, Análisis de datos, Indicadores, Análisis de información, Estrategias y Nivel de Confianza
EC2	Diseño de un sistema de indicadores de gestión y desempeño sobre el área de ventas basado en un modelo de Inteligencia de Negocios	Sistema de gestión, Indicadores, Proceso, SIPOC y Modelo de Inteligencia de Negocios
EC3	Manejo básico de indicadores de Participación Ciudadana utilizando algoritmos y técnicas de minería de datos como apoyo en la Gestión Comunitaria, Caso estudio Quitumbe	Inteligencia de Negocios, Indicadores, Minería de Datos, Valores anómalos, Reducción de Datos y Limpieza de datos
EC4	Indicadores de Gestión para la toma de decisiones basada en Inteligencia de Negocios	Inteligencia de Negocios, Toma de decisiones, Indicadores, Bodegas de datos y Cuadro de mando integral

Continúa 

Grupo de Control	Título	Palabras Clave
EC5	Metodología de Business Intelligence para la toma de decisiones efectivas en las empresas PYME	Business Intelligence, Metodologías de BI, Toma de decisiones y Análisis de datos
EC6	Solución de inteligencia de negocios y toma de decisiones en la gestión administrativa de boticas	Inteligencia de Negocios, Muestra, Nivel de confianza, estrategias, toma de decisiones y plan de negocios

Elaborado por: Tatiana Suntaxi

Los estudios revisados se detallan a continuación:

- Sofía Bustos y Verónica Mosquera, Análisis, diseño e implementación de una Solución Business Intelligence para la generación de indicadores y control de desempeño, en la empresa Otecel S.A., utilizando la metodología Hefesto V2.0, Universidad de las Fuerzas Armadas-ESPE. Departamento de Ciencias de la Computación. Carrera de Ingeniería de Sistemas e Informática.

Las autoras plantean que el uso de la herramienta BI permite organizar y analizar los datos alojados en bases de datos de distintas fuentes que facilite la comprensión que ayuda a la toma de decisiones, la solución se propone con un software Open Source Pentaho, con el cual se obtiene indicadores en tiempo real que permite llevar un control continuo (Sofía Bustos, 2013).

- Dalton Novoa, Diseño de un sistema de indicadores de gestión y desempeño sobre el área de ventas basado en un modelo de Inteligencia de Negocios, Escuela Superior Politécnica del Litoral, Universidad Santa María.

El autor en su publicación, resalta la importancia de un Sistema de indicadores que permita medir la gestión y desempeño del área de ventas basado en un modelo de inteligencia de negocios es beneficioso y genera un valor empresarial debido a que a través de dicha aplicación se conoce información como: captación de nuevos clientes, ventas al contado, a crédito, niveles de ventas de vendedores, de nuevos productos, etc., información necesaria para la toma de decisiones estratégicas (Noboa, 2010).

- Maricella Sinchiguano, Manejo básico de indicadores de Participación Ciudadana utilizando algoritmos y técnicas de minería de datos como apoyo en la Gestión Comunitaria, Caso estudio Quitumbe, Maestría en Sistemas de la Información e Inteligencia de Negocios. Universidad de las Fuerzas Armadas. Departamento de Ciencias de la Comunicación.

La autora define BI como una herramienta mediante al cual la gerencia basa la toma de sus decisiones, con la implementación adecuada se puede detectar valores anómalos y eliminarlos a través de la herramienta de Minería de datos que se utilizó (Sinchiguano, 2017).

- Fabián Castro, Indicadores de Gestión para la toma de decisiones basada en Inteligencia de Negocios, Google Scholar, Universidad Nacional de Colombia, Ingeniero de Sistemas.

El autor concluye en su artículo de investigación que mediante BI el proceso de toma de decisiones gana en eficiencia ya que facilitan la identificación de tendencias y permitan la realización de predicciones confiables basadas en los resultados obtenidos mediante la aplicación de las herramientas (Castro, 2013).

- Alberto Montaña, Juan Mendieta & Saúl Reyes, Metodología de Business Intelligence para la toma de decisiones efectivas en las empresas PYME, Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas.

Los autores plantean una propuesta metodológica basada en BI, el cual considera diferentes metodologías y herramientas de conocimiento administrativo, que ayuda a las empresas para determinar en qué situación se encuentran en los distintos escenarios en los que intervienen (Alberto Montaña, 2010).

- Ítalo Maldonado Ramírez, Solución de inteligencia de negocios y toma de decisiones en la gestión administrativa de boticas, Universidad César Vallejo, Chiclayo-Perú.

El autor plantea a la Inteligencia de Negocios como apoyo a la toma de decisiones en la gestión de ventas del caso en estudio, determinando puntos estratégicos de ventas y marketing, plan de negocios, todo esto basando un análisis histórico de los datos que maneja la organización, a través de la implementación se logró que la empresa se beneficie conociendo y usando una herramienta tecnológica para obtener mejores decisiones, por lo que la entidad mejoro su capacidad de respuesta, visibilidad y su comprensión el negocio, ya que puede estimar el comportamiento de la empresa en el futuro (Maldonado, 2014).

Construcción de la cadena de Búsqueda: Para la construcción de la cadena de búsqueda se analizó los estudios del grupo de control, encontrando palabras comunes entre estudios y las palabras propias que están direccionadas en el objetivo de la presente investigación, para ellos se formaron los siguientes contextos: Financiero, Inteligencia de Negocios, Datos, Administrativo – Financiero

Tabla 2.

Construcción de Cadena de Búsqueda

Contexto	Palabras Claves	EC1	EC2	EC3	EC4	EC5	EC6	Palabras repetidas
Inteligencia de Negocios	Business Intelligence	x		x	x	x	x	5
	Modelo de Inteligencia de Negocios		X					1
	Minería de Datos			x				1
	Metodologías de BI					x		1
Datos	Análisis de datos	x						1
	Análisis de información	x				x		2
	Reducción de Datos			x				1
	Limpieza de Datos			x				1
	Bodega de datos			x	x			2
	Valores Anómalos				x			1
Administrativo - Financiero	Indicadores	x	x	x	x			4
	Nivel de Confianza	x					x	2
	Sistema de Gestión		x					1
	Procesos		x					1

Continúa 

Contexto	Palabras Claves	EC1	EC2	EC3	EC4	EC5	EC6	Palabras repetidas
Administrativo - Financiero	Toma de Decisiones				x	x		2
	Cuadros de Mando				x		x	2
	Muestra						x	1
	Estrategias	x					x	2
	Plan de Negocios						x	1

Elaborado por: Tatiana Suntaxi

La cadena de búsqueda se formó con la combinación de las palabras que más se repiten en cada contexto para unir la cadena se utilizó el conector OR y para concatenar con el contexto AND, estableciendo la siguiente cadena de búsqueda:

((Inteligencia de Negocios) AND (Indicadores) AND ((Toma de Decisiones) OR

(Estrategias)) AND ((Información) OR (Datos))

((Business Intelligence) AND (Indicators) AND ((Decision Making) OR (Strategies))

AND ((Information) OR (Data))

Esta cadena de búsqueda fue aplicada en Google Scholar, con los criterios de inclusión y exclusión para determinar 6 estudios relacionados a la problemática planteada y la generación de indicadores aplicando Business Intelligence como posible solución para un control de desempeño del Instituto Manuel Córdova Galarza

En conclusión, de acuerdo al estudio de arte revisado, a nivel internacional se propone el uso de herramientas tecnológicas, en este caso específico Business Intelligence como apoyo para la gestión administrativa-financiera; en cuanto a la problemática planteada se revisaron estudios en que la aplicación e BI cumplió con los objetivos mejorando el análisis de la información y generando un apoyo a la alta gerencia en la toma de decisiones o de acciones correctivos de ser el caso, de igual forma no se encontró estudios realizados a entidades del sector público o de economía mixta los estudios encontrados son aplicados a empresas privada.

En consecuencia, el presente estudio contribuye con la generación de conocimiento sobre la evaluación del desempeño de una institución de economía mixta, regida por un organismo de control gubernamental (CNE), contribuyendo con nuevos enfoques de análisis a los estudios actuales y considerando las características de leyes gubernamentales ecuatorianas.

2.3. Antecedentes Conceptuales

2.3.1. Business Intelligence

2.3.1.1. Definición

“Conjunto de estrategias, acciones y herramientas enfocadas a la administración y creación de conocimientos mediante el análisis de datos existentes en una organización o empresa”, dichas estrategias permiten recabar un conjunto de acciones que la empresa puede emprender, mismas que ofrecen ventajas sobre sus competidores” (Tello & Perusquia, 2016).

2.3.1.2. Metodologías

“La metodología de Business Intelligence nos da mayor grado de certidumbre que nuestro proyecto cumplirá los objetivos trazados y en el tiempo convenido desde el punto de vista de un proyecto de Tecnología de Información, pero cabe aclarar que los proyectos de BI tienen 2 perspectivas, la de negocios y la de TI” (Cavazos, 2008).

2.3.1.3. Herramientas

“Las herramientas de software de Business Intelligence son usadas para acceder a los datos de los negocios y proporcionar reportes, análisis, visualizaciones y alertas a los usuarios. La gran mayoría de las herramientas de Business Intelligence son usadas por usuarios finales para acceder, analizar y reportar contra los datos que más frecuentemente residen en data warehouse, data marts y almacenes de datos operacionales” (Meter, 2012).

La Inteligencia de negocios cuenta con varias herramientas tecnológicas, las cuales tienen como objetivo visualizar el resultado final de la aplicación de un BI, en la **Figura 4**. Cuadrante Mágico de Gartner de Herramientas de BI, podemos observar la ubicación en los cuadrantes de las herramientas de BI.



Source: Gartner (February 2019)

Figura 4. Cuadrante Mágico de Gartner de Herramientas de BI

Fuente: (Microsystem, 2019)

2.3.2. Minería de datos

2.3.2.1. Definición

“Data mining es un proceso de descubrimiento que permite al usuario conocer la esencia y las relaciones entre sus datos. Data mining descubre patrones y tendencias en el contenido de información” (Abbey, 1997).

La minería de datos o data mining es un conjunto de técnicas y tecnológicas que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el fin de encontrar

patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto (Lopez, 2015)

2.3.2.2. Relaciones

El descubrimiento de relaciones se ilustra mejor observando hechos diarios, como hábitos de compras; en sistemas operacionales o de data warehouse, el arquitecto de datos y los diseñadores definen meticulosamente entidades y relaciones, en este contexto, una entidad es un conjunto de información que contiene hechos acerca de sus grupos de datos asociado.

En data mining se descubre e identifica las relaciones escondidas en las profundidades de muchas capas de datos corporativos.

2.3.2.3. Patrones

Descubrir patrones de conducta en un negocio agregar valor real al mismo y a la aplicación de data mining, nadie puede predecir si cierta persona va a realizar una actividad en estrecha proximidad que otra diferente, sin embargo, usando técnicas y análisis sistemáticos en datos del warehouse, estas predicciones pueden ser apoyadas gracias al descubrimiento de patrones de conducta.

Los patrones están ligados a los hábitos de las personas; es decir, la probabilidad de que se realice una actividad determinada se descubre en la identificación de un patrón.

Usando data mining como herramienta, los datos son convertidos en información; los datos por sí mismos no significan nada, sin embargo, al agruparlos constituyendo elementos de datos, comienzan a tomar sentido, con data mining es cuando pueden proporcionar información valiosa.

2.3.2.4. Beneficios

Uno de los principales beneficios de Data mining es la habilidad para convertir pensamientos en hechos (Abbey, 1997).

Como ejemplo de lo anteriormente descrito tenemos: Para un grupo de consumidores X el producto A es más atractivo en comparación a otros grupos, esta situación puede llegar a ser un hecho cuando se acompaña de la habilidad de examinar datos corporativos que permitan identificar patrones de conducta y hábitos de los clientes, el beneficio fundamental del data mining es el doble.

2.3.2.5. Metodologías para Minería de datos

Las principales metodologías para el proceso de minería de datos son:

- **Crisp – DM**

Iniciativa financiada por la Comunidad Europea ha unido para desarrollar una plataforma para Minería de datos, el objetivo de esta metodología es fomentar la interoperabilidad de las herramientas a través de todo el proceso de minería de datos, elimina la experiencia misteriosa y costosa de las tareas simples de minería de datos (León, 2015).

Crisp-DM es flexible para tener en cuenta las diferencias, variedades de tipos de datos, diferentes problemas de negocios, es neutral con respecto a herramientas, su proceso se estructura en 6 fases: Comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y Despliegue. La sucesión de fases no es rígida, cada fase se descompone en diversas tareas, pero en ningún momento esta metodología propone como realizarlas.

- **KDD³**

El proceso general es conocido como KDD, el cual está conformado de varias fases en los cuales se encuentra la minería de datos, como podemos observar a continuación en la **Figura 5**. Procesos KDD (Aguirre, 2012).

³ KDD (Knowledge Discovery from data): Descubrimiento de conocimiento a partir de datos.

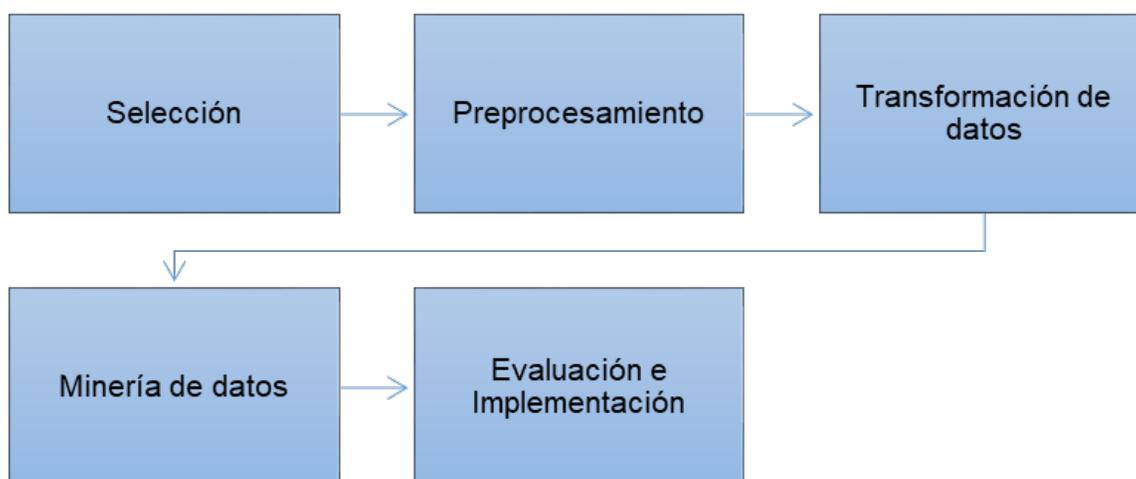


Figura 5. Procesos KDD

Fuente: (Aguirre, 2012)

Formalmente el modelo establece que la minería de datos es la etapa dentro del proceso en el cual se realiza la extracción de patrones a partir de los datos, sin embargo en la actualidad se utiliza el término KDD para hacer referencia al proceso completo de descubrimiento de conocimiento (Juan Moine, 2011).

- **Semma**

Semma es el acrónimo a las 5 fases: Sample, explore, modify, model y assess, la metodología es propuesta por SAS Institute Inc, la define como “proceso de selección, exploración y modelamiento de grandes cantidades de datos para descubrir patrones de negocios desconocidos” (Juan Moine, 2011).

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

del mismo siguiendo cualquier otra metodología de minería de datos (como CRISP-DM por ejemplo) (Juan Moine, 2011).

2.3.2.6. Cubos de Información

- OLAP (Procesamiento Analítico en Línea)

Se entiende por OLAP, o proceso analítico en línea, el método ágil y flexible para organizar datos, especialmente metadatos, sobre un objeto o jerarquía de objetivos como un sistema u organización multidimensional, y cuyo objetivo es recuperar y manipular datos y combinaciones de los mismos a través de consultas o incluso informes (Conesa, 2010).

OLAP forma parte de lo que se conoce como sistemas analíticos, que permiten responder preguntas como: ¿por qué paso? Estos sistemas pueden encontrarse tanto integrados en suites de Business Intelligence y ser simplemente una aplicación independiente (Conesa, 2010).

- OLTP

Los sistemas OLTP son bases de datos orientadas al procesamiento de transacciones. Una transacción genera un proceso atómico (que debe ser validado con un commit, o invalidado con un rollback), y que puede involucrar operaciones de inserción, modificación y borrado de datos. El proceso transaccional es típico de las bases de datos operacionales (Sinnexus, 2018).

- MOLAP (Multidimensional OLAP)

Es la forma clásica de OLAP y frecuentemente es referida con dicho acrónimo. MOLAP utiliza estructuras de base de datos generalmente optimizadas para la recuperación de los mismos. Es lo que se conoce como bases de datos multidimensionales (cubos). En definitiva, se crea un fichero que contiene todas las posibles consultas pre calculadas. A diferencia de las bases de datos relacionales, estas formas de almacenaje están optimizadas para la velocidad del cálculo. También se optimizan a menudo para la recuperación a lo largo de patrones jerárquicos de acceso. Las

dimensiones de cada cubo son típicamente atributos tales como periodo, localización, producto o código de cuenta. La forma en la que cada dimensión será agregada se define por adelantado (Conesa, 2010).

- **ROLAP (Relational OLAP)**

ROLAP trabaja directamente con las bases de datos relacionales, que almacenan los datos base y las tablas dimensionales como tablas relacionales mientras se crean nuevas tablas para guardar la información agregada (Conesa, 2010).

- **HOLAP (Hybrid OLAP)**

HOLAP: no hay acuerdo claro en la industria en cuanto a qué constituye OLAP híbrido, exceptuando el hecho de que es una base de datos en la que los datos se dividen en almacenaje relacional y multidimensional. Por ejemplo, para algunos vendedores, HOLAP consiste en utilizar la tablas relacionales para guardar las cantidades más grandes de datos detallados y utiliza el almacenaje multidimensional para algunos aspectos de cantidades más pequeñas de datos menos detallados o agregados (Conesa, 2010).

2.3.2.7. Técnicas de Data Mining

Data mining utiliza una serie de técnicas para descubrir patrones y tendencias en los datos del data warehouse, a continuación, se describen las más relevantes:

- **Redes Neuronales**

La minería basada en redes neuronales está indicada para identificar patrones y prever tendencias basadas en comportamientos previamente identificados. Una tendencia identifica un movimiento en el hábito basado en el comportamiento anterior, un ejemplo puede ser el precio de las acciones que se cotizan en la bolsa de valores, ya que para predecir su cuota en mercado en las

próximas semanas se analizan sus datos de compra históricos, la evolución del precio en el mercado.

La base de este tipo de procesamientos es lo que se aprendió trabajando con el sistema nervioso central. El conocimiento se puede aprender de una serie de datos ampliamente dispares, complejos o imprecisos. Hay tres capas en la red: la capa inferior recibe los datos de entrada, la capa, la capa oculta (intermedia) realiza el trabajo, y la exterior presenta las salidas al analista; estas tres tipos capas se llaman nodos, son otra forma de llamar a los elementos de procesamiento, que se asemejan a las neuronas del cerebro es por eso que esta técnica se llama redes neuronales (Abbey, 1997).

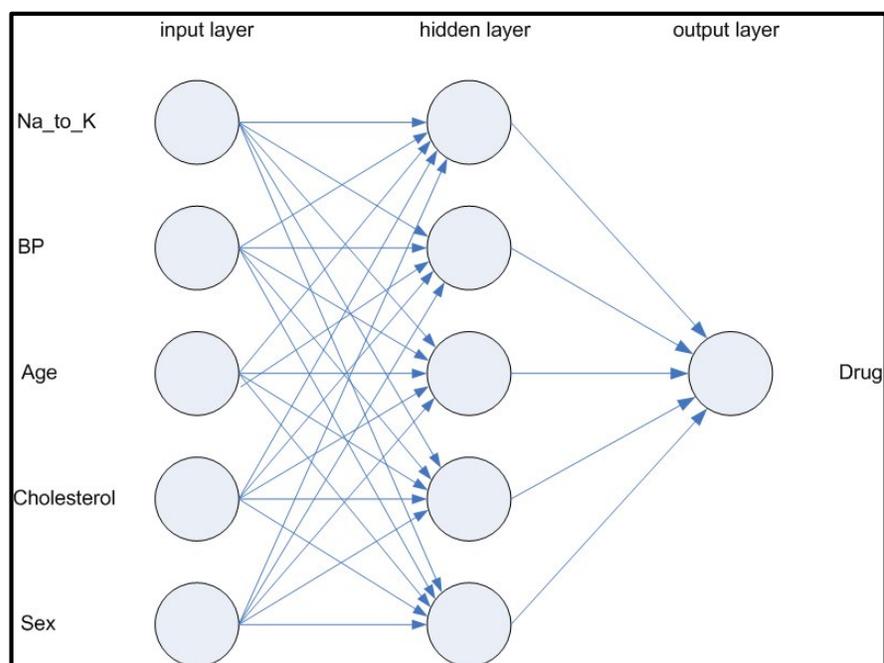


Figura 6. Estructura de una red neuronal

Fuente: (IBM, 2018)

- Descubrimiento de Asociaciones

En esta técnica concierne al estudio de conjuntos de datos e intenta descubrir asociaciones entre las ocurrencias de los atributos dentro de los datos, el descubrimiento de asociaciones intenta descubrir ocurrencias similares de los valores de los datos de los registros, y produce una salida que podría ser expresada como una regla (Abbey, 1997).

La asociación descubierta puede desvelar relaciones entre lo que parece circunstancias absolutamente sin relación. Estas circunstancias pueden afectar profundamente a la forma en que una organización realiza actividades o negocios determinados.

- **Clasificación**

La clasificación consiste en la agrupación de los datos basándose en una serie de características predefinidas por el análisis antes que comience el ejercicio, esta técnica examina los datos ya clasificados y agrupados mediante la aplicación de una regla de pertinencia. Esta regla de pertinencia puede tener un componente temporal, por ejemplo, un periodo fiscal, un año natural o un mes del año.

- **Clustering**

Clustering (agrupación) consiste en agrupar conjuntos de datos similares partiendo de un conjunto más grande de datos. Contrariamente a la clasificación la técnica de clustering descubre las agrupaciones a medida que trabaja con los datos de entrada, se identifican las similitudes que lleven a las segmentaciones en grandes grupos de datos que se parezcan unos a otros. Una vez que las agrupaciones y sus miembros son identificados, se pueden descubrir generalizaciones, patrones y tendencia basándose en las características de los miembros de cada agrupación (Abbey, 1997).

La base de esta técnica es el descubrimiento de que lo que parecían valores de datos y de atributos sin ninguna relación dentro de una serie de registros, de hecho, contiene información acerca de un conjunto claro de similitudes (Abbey, 1997).

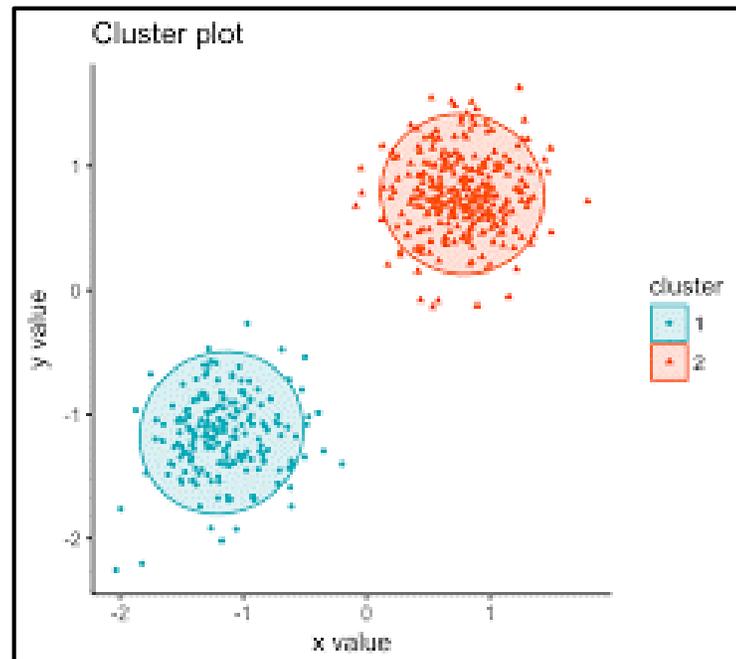


Figura 7. Ejemplo de Clustering
Fuente: (Sthda, 2013)

- Árboles de decisión

Un árbol de decisión es un mapa de los probables resultados de una serie de decisiones relacionadas, con ayuda de esta herramienta se puede comparar posibles acciones entre sí según sus probabilidades, costos y beneficios, se usa para trazar un algoritmo que anticipe la opción idónea (Lucidchart, 2015).

Un árbol de decisión por lo general comienza con nodo y luego se ramifica en los posibles resultados, con cada uno se crean nodos adicionales, es por esto que gráficamente se asemeja a un árbol.

2.3.2.8. Herramientas de Software

Las herramientas de minería de datos sirven para el análisis y exploración de los datos en determinados casos, usando modelos estadísticos, que determinen patrones de comportamiento y tendencias de uso predictivo.

La Minería de datos cuenta con varias herramientas de acuerdo a la **Figura 8**. Cuadrante Mágico de Gartner para las ciencias de la información y las plataformas de aprendizaje automático, podemos observar la ubicación de cada una de estas en los cuadrantes



Figura 8. Cuadrante Mágico de Gartner para las ciencias de la información
Fuente: (RapidMiner, 2018)

- **RapidMiner**

RapidMiner es una plataforma de software para equipos de análisis que combina la preparación de datos, el aprendizaje automático y el despliegue predictivo del modelo. RapidMiner aporta inteligencia artificial a la empresa a través de una plataforma de ciencia de datos abierta y extensible. Creado para equipos de análisis, RapidMiner unifica todo el ciclo de vida de la ciencia de datos, desde la preparación de datos hasta el aprendizaje automático y la implementación del modelo predictivo. 400,000 profesionales de análisis usan productos RapidMiner para generar ingresos, reducir costos y evitar riesgos (Rapidminer, 2017).

- **Weka**

Weka (Waikato Environment for Knowledge Analysis) es una colección de algoritmos de aprendizaje automático para tareas de minería de datos. Los algoritmos se pueden aplicar directamente a un conjunto de datos o llamar desde su propio código Java. Weka contiene herramientas para el pre procesamiento de datos, clasificación, regresión, clustering, reglas de asociación y visualización. También es adecuado para desarrollar nuevos esquemas de aprendizaje automático, una desventaja de este software es su baja potencia en el análisis de clústeres, en el que solo se ofrecen los procedimientos más importantes. Weka es un software de código abierto emitido bajo la Licencia Pública General de GNU (Weka, 2017).

- **Orange**

Aprendizaje automático de código abierto y visualización de datos para principiantes y expertos, flujos de trabajo de análisis de datos interactivos con una gran caja de herramientas.

El software permite la visualización interactiva de datos, programación visual, además cuentan con complementos que amplían su funcionalidad, con los cuales extraen datos de fuentes externas, realizar análisis de red, deducir elementos frecuentes y minar las reglas de asociación. Además, los bioinformáticos y los biólogos moleculares pueden usar Orange para clasificar genes por su expresión diferencial y realizar análisis de enriquecimiento (Orange, 2018).

- **Knime**

KNIME Analytics Platform es el software de código abierto desarrollado por la Universidad de Constanza, crea aplicaciones y servicios de ciencia de datos. Nuevos desarrollos intuitivos, abiertos y de integración continua, KNIME hace que la comprensión de datos y el diseño de flujos de trabajo de ciencia de datos y componentes reutilizables sean accesibles para todos (Knime, 2018).

Entre sus funciones resalta el análisis de datos integrativo, permite la integración de diversos procedimientos de aprendizaje automático y de data mining, adicional presenta eficiencia en el tratamiento previo de los datos, así como su extracción, transformación y carga.

- **Sas**

Se considera que esta herramienta es la más adecuada para el uso en grandes empresas, aunque también es un software de bajo costo económico, con ayuda de Sas se puede realizar pronósticos, visualización interactiva de los datos, se puede realizar con éxito un proceso de minería de datos, cuenta con gran escalabilidad, ya que puede aumentar progresivamente su eficiencia aumentando los recursos de hardware o de cualquier otro tipo (Guide, 2018).

2.3.3. Indicadores

Los KPI o indicadores de desempeño clave es un valor tangible que permite medir y cuantificar si estamos en el camino correcto para lograr nuestro objetivo (CulturaCrm, 2016).

2.3.3.1. Indicadores Financieros

Los Indicadores Financieros o Ratios Financieros son ratios o medidas que tratan de analizar el estado de la empresa desde un punto de vista individual, comparativamente con la competencia o con el líder del mercado (Financiera, 2018).

2.3.3.2. Indicadores de Gestión

Son formulados para hacer seguimiento y evaluar los resultados de la gestión y permiten evaluar el logro de las metas físicas de los objetivos de los componentes, los programas, los proyectos y los procesos del Sistema de Gestión de la Calidad (Colmayor, 2017).

2.3.4. Metodología Design Science Research

Principio Fundamental

El conocimiento, la comprensión de un problema y su solución, se adquieren en la aplicación y construcción de un artefacto (Cataldo, 2015).

Objetivos

Lograr el conocimiento y la comprensión del dominio del problema mediante la construcción y la aplicación de un artefacto diseñado (Cataldo, 2015).

Tipos de artefactos

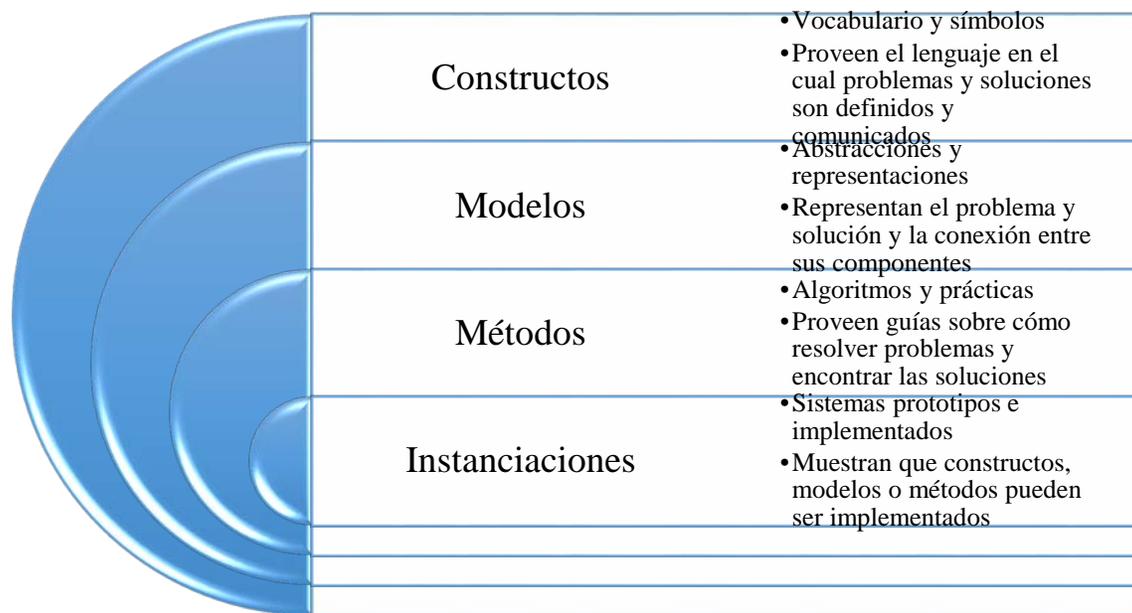


Figura 9. Tipos de artefactos

Fuente: (Cataldo, 2015)

CAPITULO III

MEMORIA TÉCNICA METODOLÓGICA

3.1. Metodología de la Investigación

La investigación del presente trabajo se elaboró y ejecutó utilizando la metodología de investigación Design Science Research.

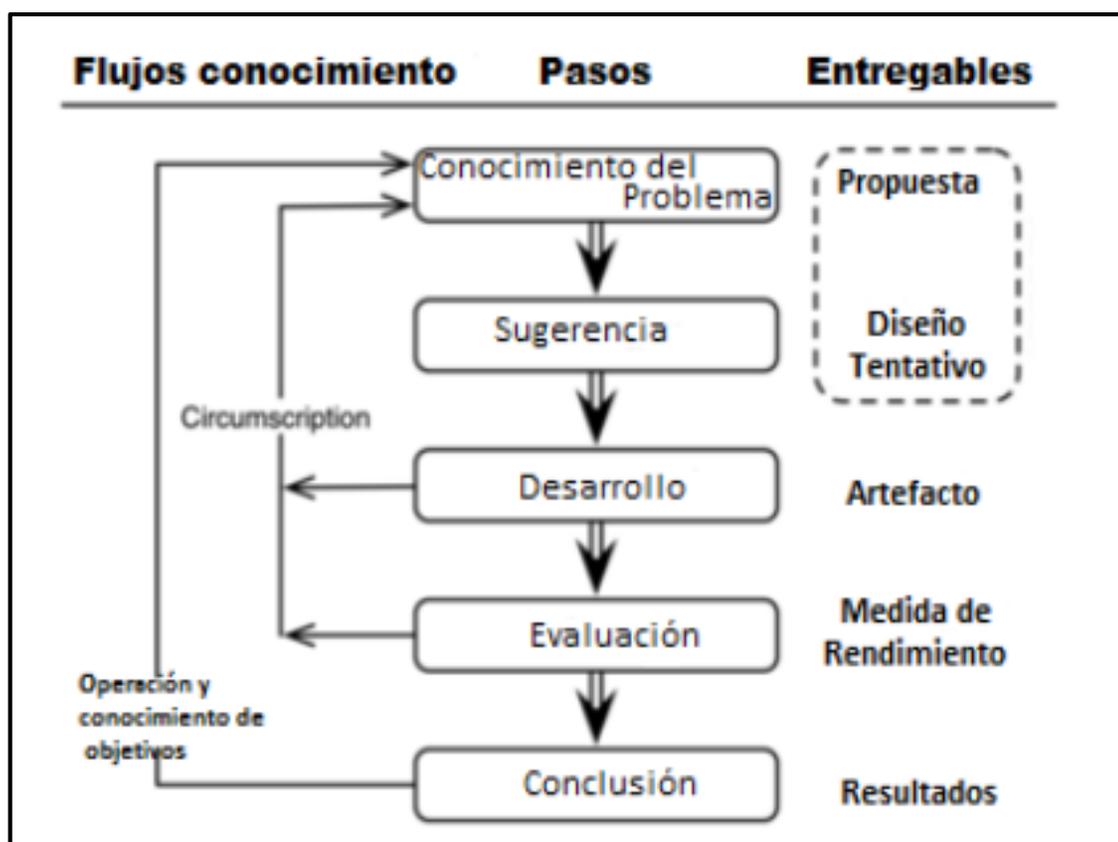


Figura 10. Design Science Research Process Mode

Fuente: (Vijai Vaishnavi, 2015)

El principio fundamental de la metodología Design Science Research es que el conocimiento, comprensión de un problema y su solución, se adquiere en la aplicación y construcción de un artefacto, es decir Investigación Científica basada en el diseño y la intención (Vijai Vaishnavi, 2015).

Se plantea utilizar la investigación científica basada en el diseño con el fin de brindar aportes significativos y brindar soluciones a problemas relevantes.

La metodología Design Science Research propone 5 pasos como se detalla en la **Figura 10**. Design Science Research Process Mode (Vijai Vaishnavi, 2015). Los cuales fueron utilizados en la construcción del modelo, como se detalla a continuación:

- 1. Conocimiento del Problema:** se determina un problema de acuerdo al entorno y a las múltiples fuentes.
- 2. Sugerencia:** sugiere una propuesta o posible solución de diseño en forma de artefacto.
- 3. Desarrollo:** define la implementación del artefacto.
- 4. Evaluación:** el artefacto se evalúa de acuerdo a los criterios predefinidos.
- 5. Conclusión:** Indica la finalización del ciclo de la investigación y sus ideas finales.

Esta metodología plantea como paso inicial el **Conocimiento del Problema**, por lo que se determinó la problemática a abordar, sus causas, las cuales fueron documentadas, de igual forma en esta etapa se establecieron las posibles soluciones, con estos lineamientos previos se siguió con la etapa de **Sugerencia**, se identificaron posibles soluciones para la problemática planteada, se realizó un estudio del arte desarrollado en la sección **2.2. Estado del Arte**, con el cual se conocieron las diversas técnicas, aplicaciones y herramientas tecnológicas que se usaron en casos de estudios con características similares, es decir que involucre el uso de indicadores con aplicaciones de BI.

En la etapa de **Desarrollo**, se elaboró el modelo analítico de indicadores de la organización, se planteó las técnicas y herramientas que se aplicaron, las cuales fueron seleccionadas de acuerdo al grado de cumplimiento que se obtuvo respecto a las necesidades de la entidad.

En la fase de **Evaluación**, se usaron las técnicas de validación de minería de datos, que permita evaluar el modelo analítico de indicadores, determinando el nivel de confianza en los resultados, en esta fase se evaluó la hipótesis con la técnica de investigación definida.

En la etapa final de **Conclusiones** se detalló el aporte del conocimiento en la aplicación del modelo analítico de indicadores financieros y de gestión.

CAPITULO IV

DESARROLLO

4.1. Ejecución del proceso de investigación

El proceso de investigación se realizó con la metodología Design Science Research, descrita anteriormente.

Mediante este proceso se tuvieron 5 ciclos:

4.1.1. Conocimiento del Problema

El presente proyecto se centró en el enfoque cuantitativo y cualitativo realizado al Instituto Manuel Córdova Galarza, lo que permitió analizar los indicadores que fueron implementados para evaluar los ingresos y gastos que maneja el partido político Izquierda Democrática.

Misión

“La misión de Izquierda Democrática es lograr la justicia social en el marco del ejercicio pleno de las libertades. Creemos en la institucionalización de la participación de los ciudadanos en la toma de decisiones políticas, por eso somos un partido político” (ID, 2018).

Visión

“Nuestra visión apunta a posicionar el partido como la primera fuerza política del país y tomar el poder para construir un Ecuador en el que se supere la pobreza y la inequidad, en el que el bienestar sea de todos sin diferencias. Un país con instituciones sólidas y en las cuales la ciudadanía deposita confianza, lejos de las políticas radicales y de confrontación. Con la innovación como base de todas las acciones” (ID, 2018).

Base legal

Al ser una entidad que cuenta con recursos públicos, esta Institución se rige a la siguiente normativa:

- Ley Orgánica de Control del gasto electoral y de la propaganda electoral
- Ley Orgánica Electoral y de Organizaciones Públicas – Código de la democracia
- Reglamento de Fondo Partidario Permanente de Organizaciones Políticas
- Reglamento Interno para la utilización del fondo Partidario Permanente

Fuentes de Financiamiento

La organización se rige al Código de la Democracia – Art. 353, el cual estipula “Las organizaciones políticas reciben financiamiento público y privado”.

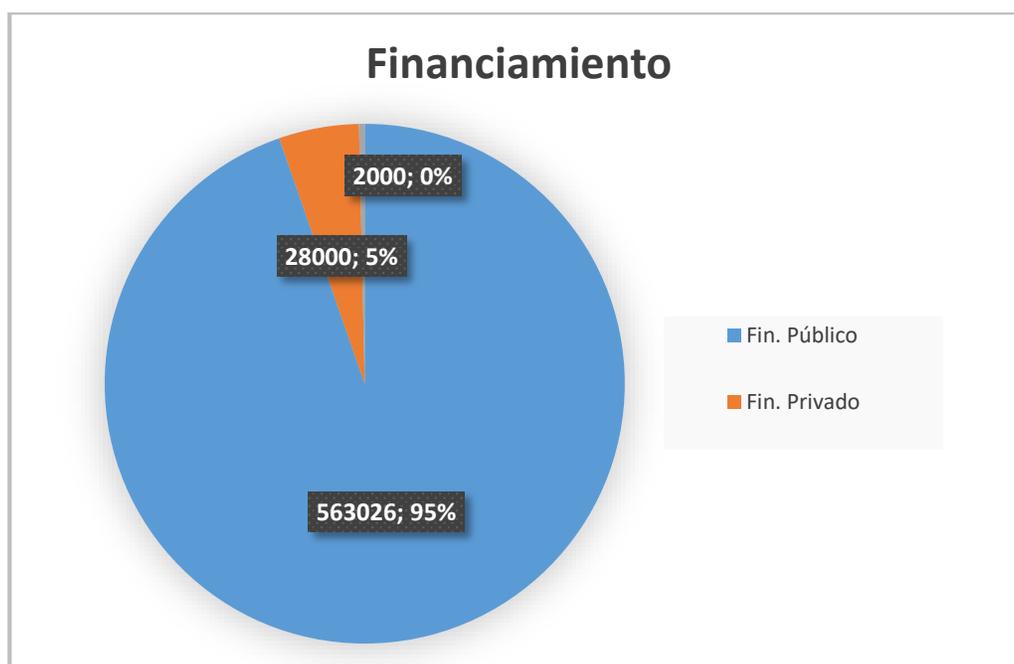


Figura 11. Financiamiento IZ
Fuente: (ID, 2018)

Financiamiento Público

El financiamiento público se da calculando el equivalente al 0,3 por mil de los egresos fiscales contantes en el Presupuesto General del Estado (ID, 2018).

Financiamiento Privado

Comprende contribuciones periódicas de los afiliados o adherentes, los recursos obtenidos de las actividades organizativas promovidas por sus frentes sectoriales (ID, 2018).

Rendimientos Financieros

Corresponden a interés ganados por colocación de pólizas de acumulación en el sistema bancario (ID, 2018).

Programa General de gastos

La normativa legal que rige al partido dictamina “*El fondo partidario permanente entregado por el Consejo Nacional Electoral, será utilizado exclusivamente para propiciar actividades de formación, publicaciones, capacitación e investigación, así como para el funcionamiento institucional de las organizaciones políticas beneficiarias de estos recursos públicos*” (I.D, 2017).

Los montos asignados por el partido para los gastos son:

Tabla 3.

Asignación monto de gastos

GASTOS	DESCRIPCIÓN	MONTO
Programa de formación	Gastos necesarios para implementar y sostener un centro de formación política	40.000,00
Programa Publicaciones	Gastos destinados a la publicación de revistas, guías, periódicos, etc	30.000,00
Programa de Capacitación e Investigaciones	Gastos destinados a procesos educativos.	160.000,00
Programa de Funcionamiento Institucional	Gastos derivados del funcionamiento habitual y permanente, gastos administrativos, arriendo, viáticos, transporte, etc	333.026,00

Fuente: (ID, 2018)

4.1.2. Sugerencia

Una vez identificada la información existente en el partido político se realizó el diseño de una bodega de datos que cumpla con el propósito del trabajo e involucre los indicadores para el análisis deseado, se utilizó Power Designer para documentar el diseño realizado. Adicional, se determinó la herramienta ETL, para realizar las transformaciones, en este caso se escogió la herramienta Data Intergration de Pentaho, ya que ofrece una versión gratuita de su suite que incluye análisis de negocios, así como la integración de datos, un diseñador de informes, de igual forma se identificaron los algoritmos de minerías utilizados en el modelo diseñado en la herramienta

RapidMiner, con el fin de facilitar la comprensión de los resultados y determinar alertas de fácil manejo y visualización se utilizó la herramienta de Power BI.

4.1.3. Desarrollo

En esta fase se creó el modelo de minería de datos mediante los requerimientos obtenidos del entorno del partido político que determine el comportamiento de los gastos que incurre la organización y determinar los límites de los mismos con el objetivo de no incumplir con la normativa que rige al partido y no minimizar los montos asignados por el Consejo Nacional Electoral, de igual forma se determinaron las alertas y las visualizaciones de los indicadores seleccionados.

En el desarrollo se implementó la metodología CRISP-DM la cual propone 6 fases:

Fase I. Comprensión del negocio: Los partidos políticos en el Ecuador cuentan con un presupuesto público el cual es asignado por el Consejo Nacional Electoral de acuerdo a las autoridades que ganen cargos en las votaciones democráticas del país, estos gastos deben ser incurridos únicamente en rubros específico como son: Gastos por formación, gasto por publicaciones, gasto por capacitación e investigación y gasto por funcionamiento institucional, si el partido político de izquierda democrática incurriera en otro tipos de gastos que no sean los permitidos la autoridad competente podría sancionar al partido minimizando el monto de asignación por lo que es importante conocer el comportamiento de los mismos que permitan determinar alertas de consumo del presupuesto asignado.

Fase II. Comprensión de los datos: Se analizó los datos facilitados por el partido de los periodos 2017 y 2018 generados en el sistema contable Latinium, es importante mencionar que el partido reinicio sus actividades en el año 2017 por lo que no existe información de años anteriores, se recibieron 5 archivos en Excel como se detalla a continuación:

- Balance de comprobación: Código, nombre, ma_debe, ma_haber, sdebe, shaber, debe, haber, deudor y acreedor.

	A	B	C	D	E	F	G	H	I	J
1	codigo	nombre	ma_debe	ma_haber	sdebe	shaber	debe	haber	deudor	acreedor
2	1.1.2.01	Bco. Pichincha	0	0	0	0	533908,44	19321,27	514587,17	0
3	1.1.3.02	Anticipos a C	0	0	0	0	3000	0	3000	0
4	1.1.3.04	Otros Anticip	0	0	0	0	200	0	200	0
5	1.1.6.03	Otras cuenta	0	0	0	0	1102,66	984,92	117,74	0
6	1.2.4.01	Equipos de C	0	0	0	0	258,93	0	258,93	0
7	2.1.2.03	IESS por pag	0	0	0	0	0	329,4	0	329,4
8	2.1.2.09	Otras cuenta	0	0	0	0	102,53	224,45	0	121,92
9	2.1.3.01	Décima Terc	0	0	0	0	0	127,08	0	127,08
10	2.1.3.02	Décima Cuar	0	0	0	0	0	128,68	0	128,68
11	2.1.3.03	Vacaciones	0	0	0	0	0	63,56	0	63,56
12	2.2.1.03	Retención 7	0	0	0	0	0	38,52	0	38,52
13	2.2.1.04	Retención 10	0	0	0	0	0	648	0	648
14	2.2.2.02	1% Retenció	0	0	0	0	0	2,15	0	2,15
15	2.2.2.03	2% Retenció	0	0	0	0	0	10,99	0	10,99
16	2.2.2.04	8% Retenció	0	0	0	0	0	32	0	32
17	2.2.2.05	10% Retenci	0	0	0	0	0	500	0	500
18	4.1.1.01	Transferenci	0	0	0	0	0	533026,04	0	533026,04
19	4.1.2.02	Contribucior	0	0	0	0	0	6260,84	0	6260,84
20	5.1.1.1.01	Sueldos	0	0	0	0	1525	0	1525	0
21	5.1.1.1.02	Décima Terc	0	0	0	0	127,08	0	127,08	0
22	5.1.1.1.03	Décima Cuar	0	0	0	0	128,68	0	128,68	0
23	5.1.1.1.04	Vacaciones	0	0	0	0	63,56	0	63,56	0

Figura 12. Archivo Excel – Balance de Comprobación

- Libro mayor: número, tipocom, numcom, centro, código, fecha, nombre, detalle, saldoa, debe, haber, saldo.

numstj	numcom	centro	codigo	fecha	nombre	detalle	saldoa	debe	haber	saldo
000011	I 00008		1.1.2.01	15-Oct-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	ACREDITACION FONDO PARTIDARIO		0	533026,04	0
000010	D 0001		1.1.2.01	28-Oct-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	N/D COSTO CHEQUIFRA, GASTOS RA	533026,04	0	18,44	
000010	F 0000001		1.1.2.01	06-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/001 FONDO ROTATIVO 001-DM	533007,6	0	500	
000010	F 0000002		1.1.2.01	14-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	FONDO ROTATIVO 03-DM	532507,6	0	1500	
000010	E 0000003		1.1.2.01	15-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CHEQUE ANULADO	531007,6	0	0	
000010	E 0000004		1.1.2.01	17-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	FONDO ROTATIVO 03-EH-MC	531007,6	0	4000	
000010	D 003		1.1.2.01	28-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	N/C AJUSTE BANCARIO	527007,6	0,01	0	5
000010	E 0000005		1.1.2.01	30-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	FONDO ROTATIVO 03-DM	527007,61	0	600	5
000010	D 004		1.1.2.01	04-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	N/D CONSUMO INTERNET NOVIEMBRE	526407,61	0	109,95	5
000010	E 0000006		1.1.2.01	17-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/006 FONDO ROTATIVO 04-DM	526297,66	0	1000	5
000010	E 0000007		1.1.2.01	17-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/007 CHEQUE ANULADO	525297,66	0	0	5
000010	E 0000008		1.1.2.01	18-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/008 TALLER DE CAPACITACION	525297,66	0	2000	5
000010	E 0000009		1.1.2.01	21-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	PRIMER ANTICIPO REMODELACION A	523297,66	0	3000	5
000010	E 0000010		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/10 HONORARIOS PROFESIONALES	520297,66	0	4500	5
000010	E 0000011		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/011 ANULADO	515797,66	0	0	5
000010	E 0000012		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/12 SUELDO SEGUN ROL DE PAGO	515797,66	0	339,56	5
000010	E 0000013		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/13 ANULADO	515458,1	0	0	5
000010	E 0000014		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/14 SUELDO SEGUN ROL DE PAGO	515458,1	0	339,56	5
000010	E 0000015		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/15 SUELDO ROL DE PAGOS DICI	515118,54	0	339,56	5
000010	E 0000016		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/16 PAGO POR SERVICIO HOSTIN	514778,98	0	344	5
000010	E 0000017		1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/17 SUELDO SEGUN ROL DE PAGO	514434,98	0	362,2	5
000010	D 010		1.1.2.01	28-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	DEPOSITO DEV. DIF FONDO ROTATI	514072,78	35,14	0	5

Figura 13. Archivo Excel – Libro Mayor

- Plan de cuentas: Código y nombre.

1	codigo	nombre
2	1.	ACTIVO
3	1.1.	ACTIVO CORRIENTE
4	1.1.1.	CAJA
5	1.1.1.01	Caja Chica FPP
6	1.1.2.	BANCO CUENTA CORRIENTE UNICA
7	1.1.2.01	Bco. Pichincha Cta. 2100151007 (F.PUBLICO)
8	1.1.2.02	Bco. Pichincha Cta. 2100151007 (F.PRIVADO)
9	1.1.3.	ANTICIPO DE FONDOS
10	1.1.3.01	Anticipo a Remuneraciones
11	1.1.3.02	Anticipos a Contratos
12	1.1.3.03	Anticipos a Proveedores
13	1.1.3.04	Otros Anticipos
14	1.1.4.	FONDOS ROTATIVOS SEDES PROVINCIALES
15	1.1.4.10	Fondo Rotativo Imbabura
16	1.1.4.17	Fondo Rotativo Provincia Pichincha
17	1.1.5.	FONDO A RENDIR CUENTAS
18	1.1.5.01	Anticipo de Viáticos, pasajes y otros de viaje (OP)
19	1.1.5.02	Anticipo de viáticos, pasajes y otros (proyectos)
20	1.1.5.03	Otros fondos para fines específicos
21	1.1.6.	CUENTAS POR COBRAR
22	1.1.6.01	Cuentas por cobrar años anteriores
23	1.1.6.02	Cuentas por cobrar afiliados y/o adherentes

Figura 14. Archivo Excel- Plan de cuentas

- Libro Bancos: número, tipocom, numcom, centro, código, fecha, nombre, detalle, saldoa, debe, haber, saldo.

A	B	C	D	E	F	G	H	I	J	K	L	M
nume	tij	numdoc	numcor	codigo	fecha	nombre	detalle	saldoa	debe	haber	saldo	
00001	D	0009999	0008	1.1.2.01	15-Oct-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	ACREDITACION FONDO PARTIDAR	0	533026,04	0	533026,04	
00006	N	D	001	1.1.2.01	28-Oct-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	N/D COSTO CHEQUERA , GASTOS E	533026,04	0	18,44	533007,6	
00007	C	E	00000000	00000000	1.1.2.01	06-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/001 FONDO ROTATIVO 001-DM	533007,6	0	500	532507,6
00004	C	E	00000000	00000000	1.1.2.01	14-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/002 FONDO ROTATIVO 02-DM	532507,6	0	1500	531007,6
00007	C	E	00000000	00000000	1.1.2.01	15-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/003 CHEQUE ANULADO	531007,6	0	0	531007,6
00008	C	E	00000000	00000000	1.1.2.01	17-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/004 FONDO ROTATIVO 03-EH-F	531007,6	0	4000	527007,6
00026	N	D	003	1.1.2.01	28-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	N/C AJUSTE BANCARIO	527007,6	0,01	0	527007,61	
00005	C	E	00000000	00000000	1.1.2.01	30-Nov-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/005 FONDO ROTATIVO 03-DM	527007,61	0	600	526407,61
00027	N	D	004	1.1.2.01	04-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	N/D CONSUMO INTERNET NOVIEN	526407,61	0	109,95	526297,66	
00017	C	E	00000000	00000000	1.1.2.01	17-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/006 FONDO ROTATIVO 04-DM	526297,66	0	1000	525297,66
00015	C	E	00000000	00000000	1.1.2.01	17-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/007 CHEQUE ANULADO	525297,66	0	0	525297,66
00014	C	E	00000000	00000000	1.1.2.01	18-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/008 TALLER DE CAPACITACION	525297,66	0	2000	523297,66
00015	C	E	00000000	00000000	1.1.2.01	21-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	PRIMER ANTICIPO REMODELACION	523297,66	0	3000	520297,66
00016	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/10 HONORARIOS PROFESIONA	520297,66	0	4500	515797,66
00016	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/12 SUELDO SEGUN ROL DE PAG	515797,66	0	339,56	515458,1
00015	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/13 ANULADO	515458,1	0	0	515458,1
00020	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/14 SUELDO SEGUN ROL DE PAG	515458,1	0	339,56	515118,54
00021	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/15 SUELDO ROL DE PAGOS DICI	515118,54	0	339,56	514778,98
00022	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/16 PAGO POR SERVICIO HOSTI	514778,98	0	344	514434,98
00025	C	E	00000010	00000010	1.1.2.01	22-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/17 SUELDO SEGUN ROL DE PAG	514434,98	0	362,2	514072,78
00017	C	E	00000010	00000010	1.1.2.01	23-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	CH/011 ANULADO	514072,78	0	0	514072,78
00031	N	D	010	1.1.2.01	28-Dec-1	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	DEPOSITO DEV. DIF FONDO ROTAT	514072,78	35,14	0	514072,92	

Figura 15. Archivo Excel – Libro bancos

- Estado de situación: Código, nombre, debe, haber, saldo, saldom

1	codigo	nombre	debe	haber	saldo	saldom
2	1.	ACTIVO	0	0	0	518163,84
3	1.1.	ACTIVO CORRIENTE	0	0	0	517904,91
4	1.1.2.	BANCO CUENTA CORRIENTE UNICA	0	0	0	514587,17
5	1.1.2.01	Bco. Pichincha Cta. 2100151007 (F.PUBLIC	533908,44	19321,27	514587,17	0
6	1.1.3.	ANTICIPO DE FONDOS	0	0	0	3200
7	1.1.3.02	Anticipos a Contratos	3000	0	3000	0
8	1.1.3.04	Otros Anticipos	200	0	200	0
9	1.1.6.	CUENTAS POR COBRAR	0	0	0	117,74
10	1.1.6.03	Otras cuentas por cobrar	1102,66	984,92	117,74	0
11	1.2.	ACTIVOS FIJOS	0	0	0	258,93
12	1.2.4.	EQUIPOS DE COMPUTACION Y SOFTWARE	0	0	0	258,93
13	1.2.4.01	Equipos de Computación	258,93	0	258,93	0
14	1.9.9.9.9.9.	TOTAL ACTIVOS	0	0	0	518163,84
15	2.	PASIVOS	0	0	0	2002,3
16	2.1.	OBLIGACIONES POR PAGAR	0	0	0	770,64
17	2.1.2.	CUENTAS POR PAGAR	0	0	0	451,32
18	2.1.2.03	IESS por pagar	0	329,4	329,4	0
19	2.1.2.09	Otras cuentas por pagar	102,53	224,45	121,92	0
20	2.1.3.	PROVISIONES BENEFICIOS SOCIALES	0	0	0	319,32
21	2.1.3.01	Décima Tercera Remuneración	0	127,08	127,08	0
22	2.1.3.02	Décima Cuarta Remuneración	0	128,68	128,68	0
23	2.1.3.03	Vacaciones	0	63,56	63,56	0

Figura 16. Archivo Excel – Estado de Situación financiera

Fase III. Preparación de los datos: La información entregada por la administración fueron archivos en Excel, cuyos datos no eran aptos para realizar un análisis adecuado, dichos archivos se encontraban en diferentes hojas de cálculos, tildes en cada vocal, uso de la letra ñ, como se muestra en la **Figura 17**. Limpieza de datos en Pentaho, es por esta razón que se realizó el proceso ETL, para limpiar y organizar los datos y cargarlos en una base MySQL con la cual se pudo trabajar el modelo de gestión de datos.

Replace in string										
Step name: Replace in string										
Fields string										
#	^	In stream field	Out stream field	use RegEx	Search	Replace with	Set empty string?	Replace with field	Whole Word	Case sensitive
1		FMAY_Nombre		N	Ú	o	N		N	N
2		FMAY_Nombre		N	Vi-ticos	Viaticos	N		N	N
3		FMAY_Nombre		N	È	e	N		N	N
4		FMAY_Nombre		N	ì	i	N		N	N
5		FMAY_Nombre		N	ó	o	N		N	N
6		FMAY_Nombre		N	inform-t	informat	N		N	N
7		FMAY_Nombre		N	á	a	N		N	N
8		FMAY_Nombre		N	é	e	N		N	N
9		FMAY_Nombre		N	í	i	N		N	N
10		FMAY_Nombre		N	ú	u	N		N	N
11		FMAY_Nombre		N	aÕos	años	N		N	N
12		FMAY_Nombre		N	pùblicos	publicos	N		N	N

Figura 17. Limpieza de datos en Pentaho

Fase IV. Modelado: Una vez realizado el proceso de ETL, se obtuvo la bodega de datos en la nueva base MySQL, con la cual se procedió a realizar el respectivo modelo de minería de datos, analizando, comparando los modelos en la herramienta RapidMiner que permitió seleccionar el más adecuado para el objeto de estudio.

Fase V. Evaluación: Con el modelo creado en la herramienta, en esta fase se realiza la evaluación respecto a la precisión y porcentaje de error de acuerdo a los factores establecidos. De igual forma se evalúa en nivel de satisfacción de los objetivos planteados en el presente proyecto, los resultados y el proceso de minería de datos.

Fase VI. Despliegue: En esta última fase se brinda recomendaciones acerca del uso del modelo de gestión de datos, generando reportes gráficos que permite obtener la visualización de la situación actual de la organización, el cumplimiento de los montos asignados para cada uno de los gastos, y en base a esto tomar las acciones correctivas de ser el caso.

4.1.4. Evaluación

Dicho modelo fue evaluado a través del operador performance, bondad que RapidMiner ofrece, el cual determinó un nivel de confianza del 100% aceptado dentro del resultado obtenido, como se muestra en la **Figura 60**. Matriz de precisión

4.1.5. Conclusión

En la fase de conclusión se observan los resultados obtenidos con la aplicación del modelo de gestión de datos, se dan las pautas para manejar e interpretar las alertas que se generan en base a la información empresarial ya que en base a estas se determinarían las acciones a tomar por las autoridades de la organización.

4.2. Guía Metodológica

A continuación, se visualiza el proceso que se llevó a cabo en la ejecución del presente trabajo:

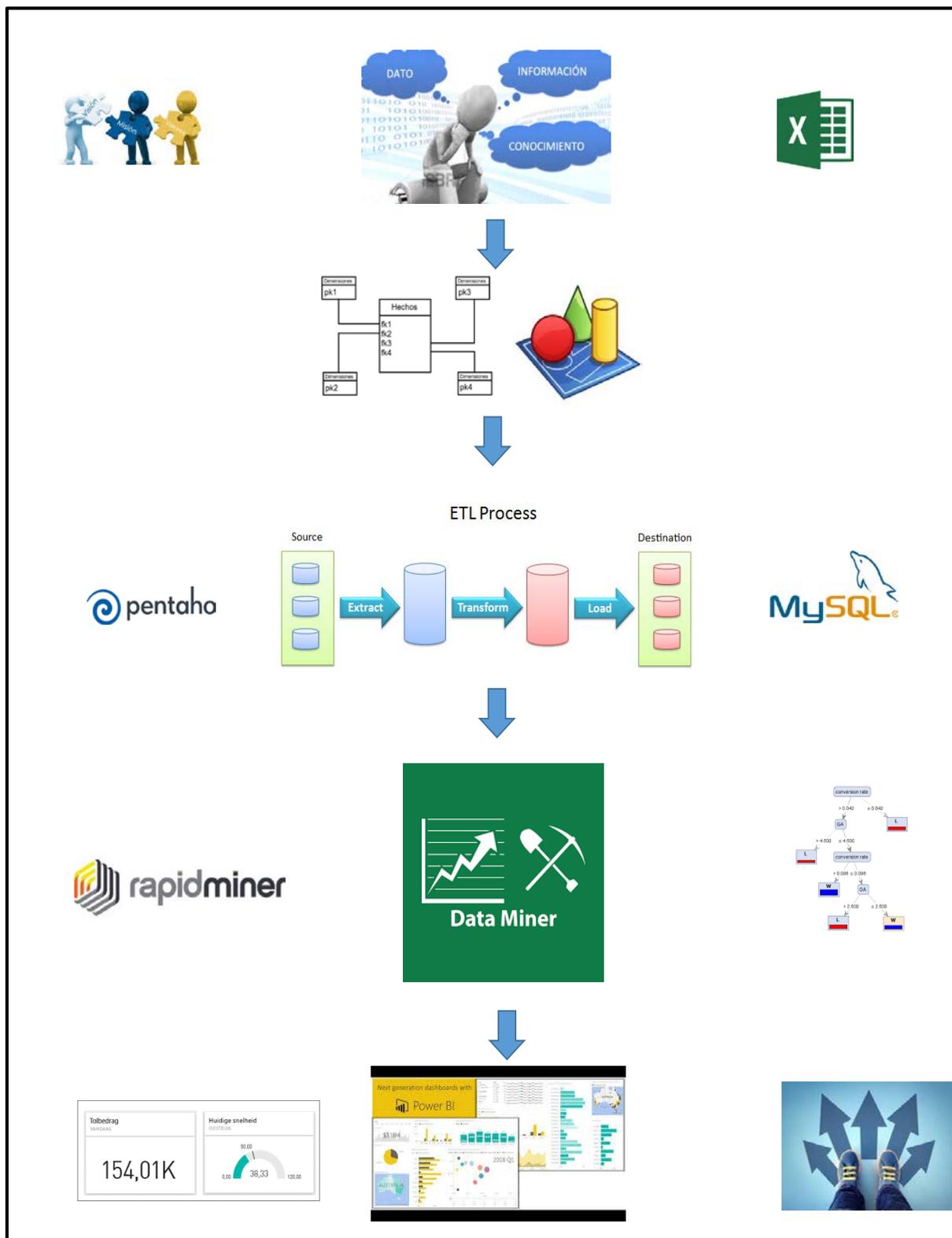


Figura 18. Proceso de ejecución del trabajo

4.3. Construcción del modelo de Gestión de Datos

4.3.1. Análisis y selección de datos

Los registros de la información contable y financiera que la organización otorgó para la realización del presente proyecto corresponde al periodo 2018, dicha información fue la fuente principal para el desarrollo de la investigación.

Las variables a estudiar son: fecha, código plan de cuentas, Nombre de cuenta, Detalle de la transacción, debe, haber; las cuales fueron determinadas de los archivos en Excel facilitados por el departamento financiero de la organización.

La selección de variables se basó en los requerimientos institucionales, la información proporcionada por parte del partido político, considerados relevantes para el estudio de los gastos y corroborar que se esté cumplimiento de las partidas presupuestarias asignadas.

Estas variables fueron entregadas en 5 archivos que son útiles para la creación de la bodega de datos descritos en la **Tabla 4**.

Tabla 4.

Datos entregados por Izquierda Democrática

	Codigo
Balance de Comprobación	Nombre
	ma_debe
	ma_haber
	Sdebe
	Saber
	Debe
	Haber
	Deudor
	Acreeedor

Continúa 

	Numero
Libro Mayor	Tipocom
	Numcom
	Centro
	Código
	Fecha
	Nombre
	Detalle
	Saldoa
	Debe
	Haber
	Saldo
	Plan de Cuentas
Nombre	
Libro Bancos	Numero
	Tipodoc
	Tipocom
	Numdoc
	Numcom
	Código
	Fecha
	Nombre
	Detalle
	Saldoa
	Debe
	Haber
	Saldo
	Estado de Situación
Nombre	
Debe	
Haber	
Saldo	
Saldom	

4.3.2. Preparación de los datos

El análisis y selección de datos fue el primer paso en la creación de la bodega de datos ya que con esta se realizó el modelo entidad relación que se muestra en la **Figura 19**. Diagrama Entidad – Relación

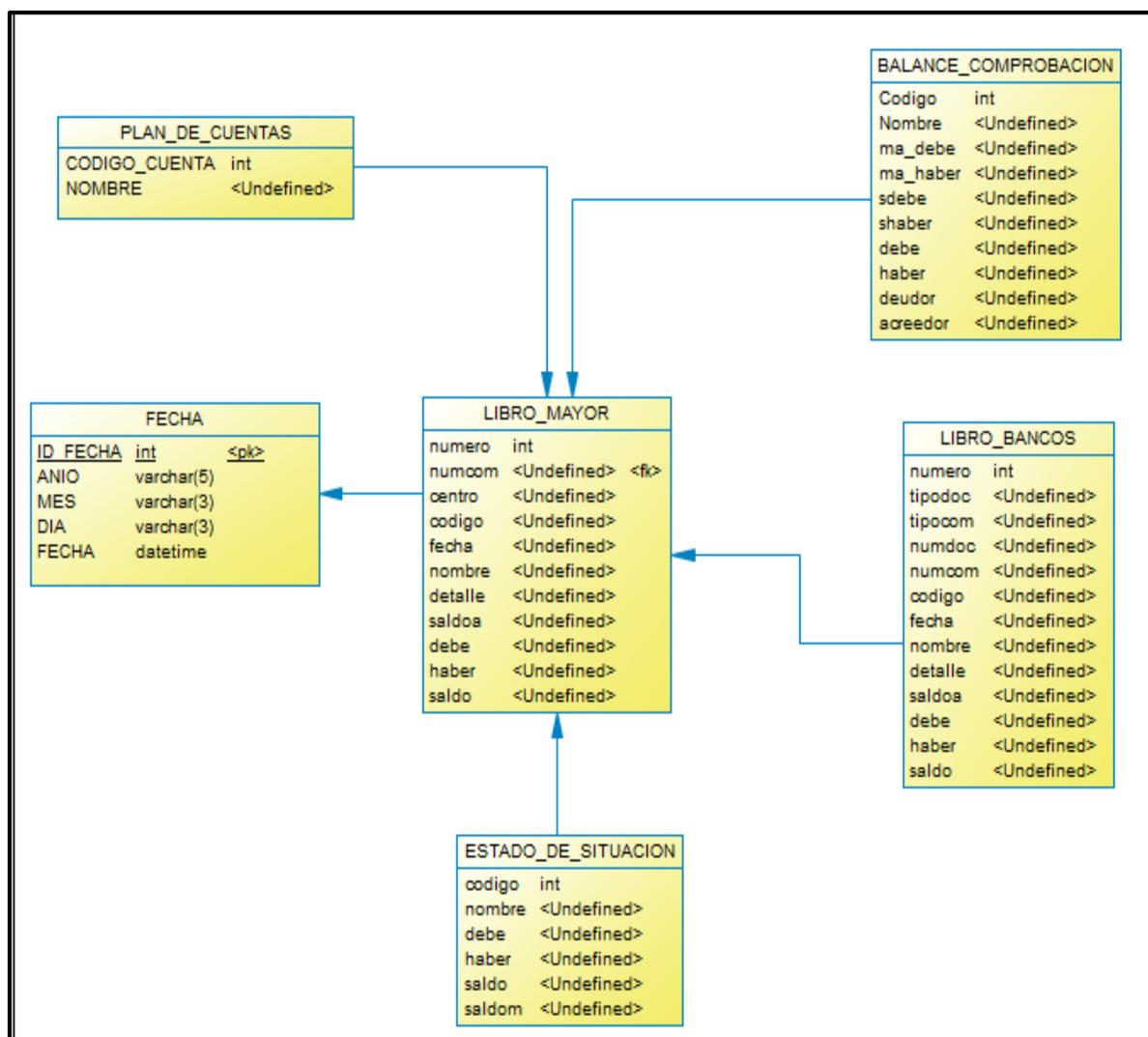


Figura 19. Diagrama Entidad – Relación

Con el objetivo de construir un modelo enfocado al objetivo del presente proyecto, se realizó el diagrama multidimensional como se muestra en la **Figura 20**. Diagrama Modelo Multidimensional

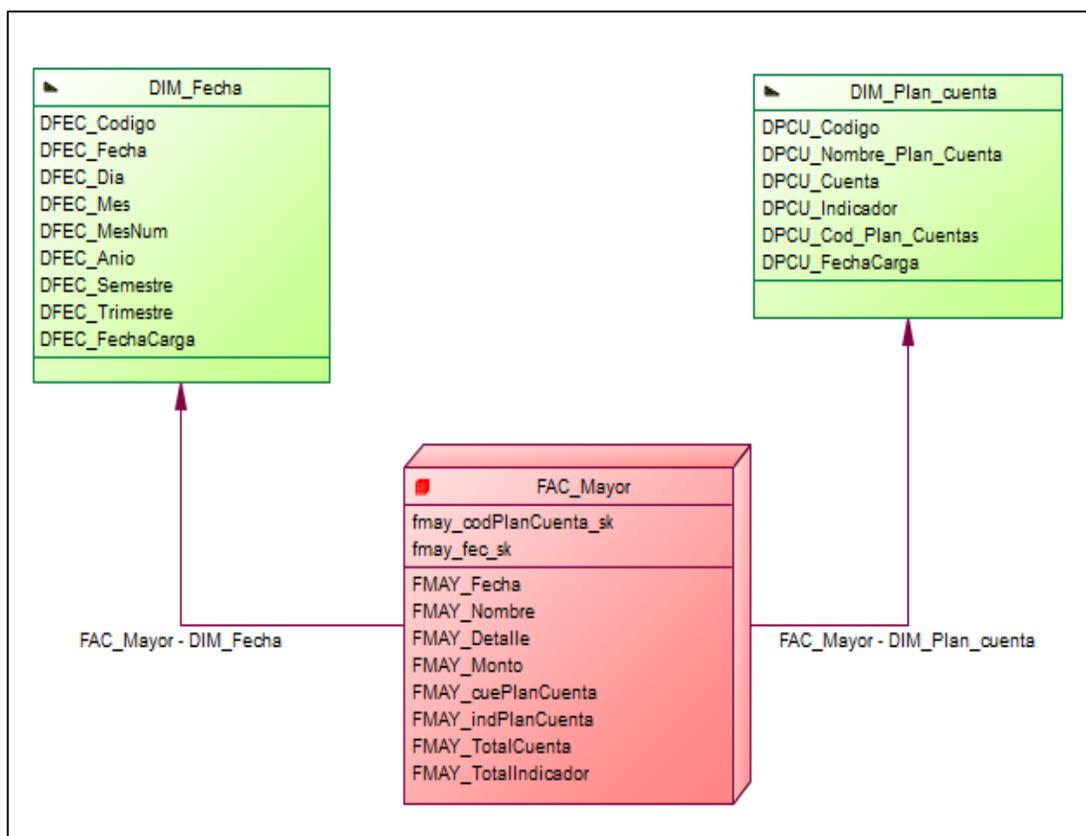


Figura 20. Diagrama Modelo Multidimensional

La bodega de datos está conformada por 2 dimensiones y 1 tabla de hechos, que se describen a continuación:

DIMENSIÓN FECHA (DIM_Fecha): Tabla en la que se referencia los periodos de tiempo de los datos otorgados, a través de esta dimensión se diferencia los datos de acuerdo a los periodos en que se presenta la información contable – financiera.

DIMENSIÓN PLAN CUENTA (DIM_Plan_cuenta): Tabla en la que se almacenan todas las cuentas contables con las que registran transacciones económicas en el partido político Izquierda Democrática.

TABLA DE HECHOS (FAC_Mayor): Tabla central en la cual se almacenan los datos cambiantes, en su mayoría cualitativos, definen las variables a analizar de los gastos en los que incurre el partido político Izquierda Democrática.

4.3.3. Procesos ETL

Se realizó la comparación de las herramientas ETL con el fin de realizar el proceso de extracción, transformación y carga.

Tabla 5.
Comparación de herramientas ETL

	Talend	Inaport	Pentaho	Oracle WareHouse	Server Integration
Costo	●	●	●	●	●
Riesgo	●	●	●	●	●
Facilidad	●	●	●	●	●
Soporte	●	●	●	●	●
Implementación	●	●	●	●	●
Velocidad	●	●	●	●	●
Calidad de la data	●	●	●	●	●
Monitoreo	●	●	●	●	●
Conectividad	●	●	●	●	●

De acuerdo a la **Tabla 5.**, podemos observar que Pentaho tiene varios atributos comparados con otras herramientas como son: costo, riesgo, su facilidad en el manejo, cuenta con páginas web en las cuales los usuarios pueden encontrar diversas ayuda en la ejecución de la herramienta, se puede conectar a diversas bases de datos, en el caso de estudio se realizó la conexión MySQL.

Se seleccionó la herramienta Data Integration de Pentaho, que además de los atributos anteriormente descritos, cumple con el requisito indispensable de ser Software libre ya que el Instituto Manuel Córdova Galarza al gestionar los montos asignados a un organismo público con actividades determinadas no puede invertir en software con altos costos de implementación.

Se inició el proceso de extracción desde los archivos en formato Excel que la administración otorgó, luego la transformación y finalmente la carga, seleccionando los datos útiles para el desarrollo del proyecto y realizando la transformación de los mismos, cargando los datos a la nueva base de datos.

Dimensión Fecha

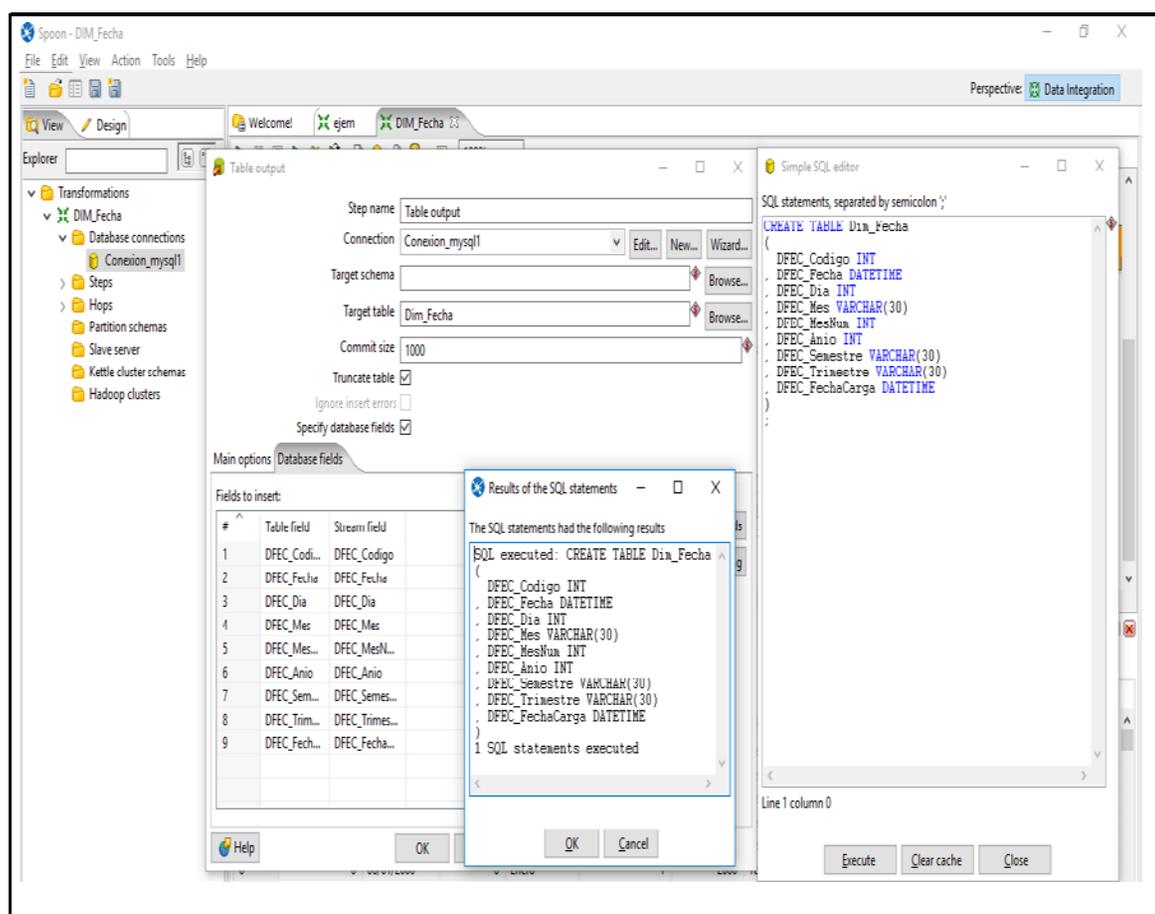


Figura 21. Proceso ETL dimensión fecha – Fuente destino

Como primer paso se agrega 15.000 días a partir del 01-01-2000, por lo que se cargan fechas hasta el 24-01-2041, se realiza el proceso ETL que permite sacar reportes por determinados periodos: mensuales, anuales, etc.

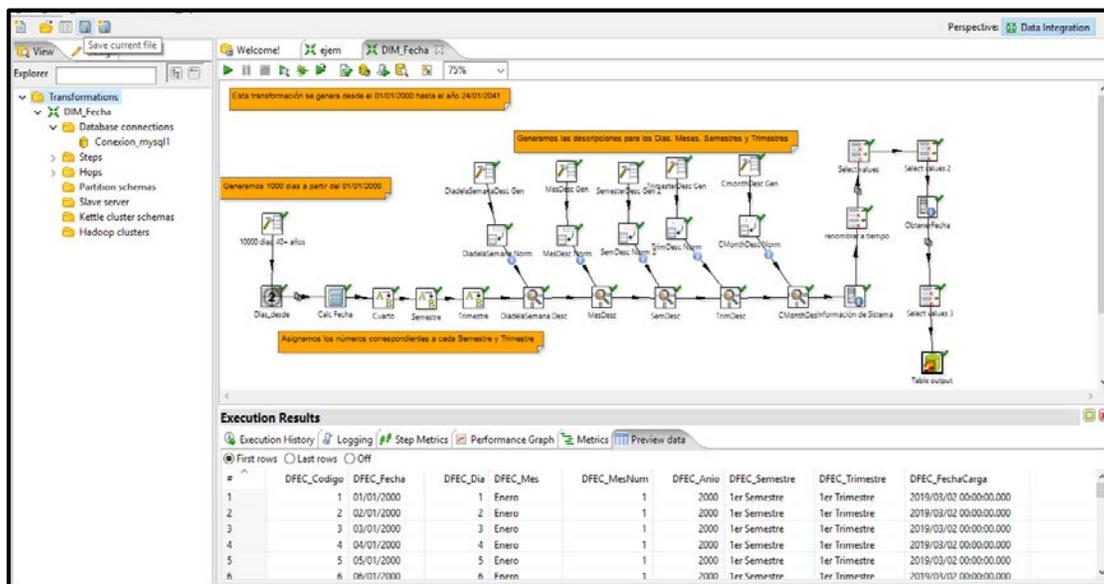


Figura 22. Proceso Extracción y Carga dimensión fecha

El resultado de correr el proceso de la dimensión fecha es la carga de la tabla con todos los registros de fechas, periodos y fecha de carga.

Dimensión Plan Cuenta

Se seleccionó como fuente de origen el archivo en excel Plan de cuentas.

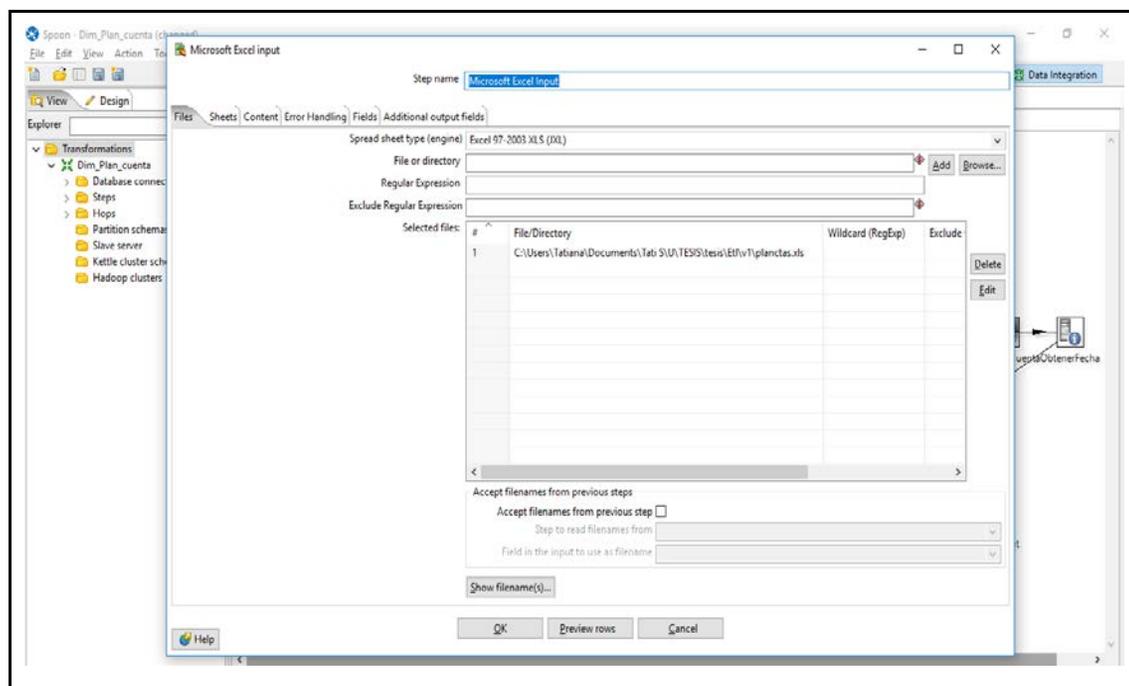


Figura 23. Proceso ETL dimensión Plan Cuenta – Fuente Origen

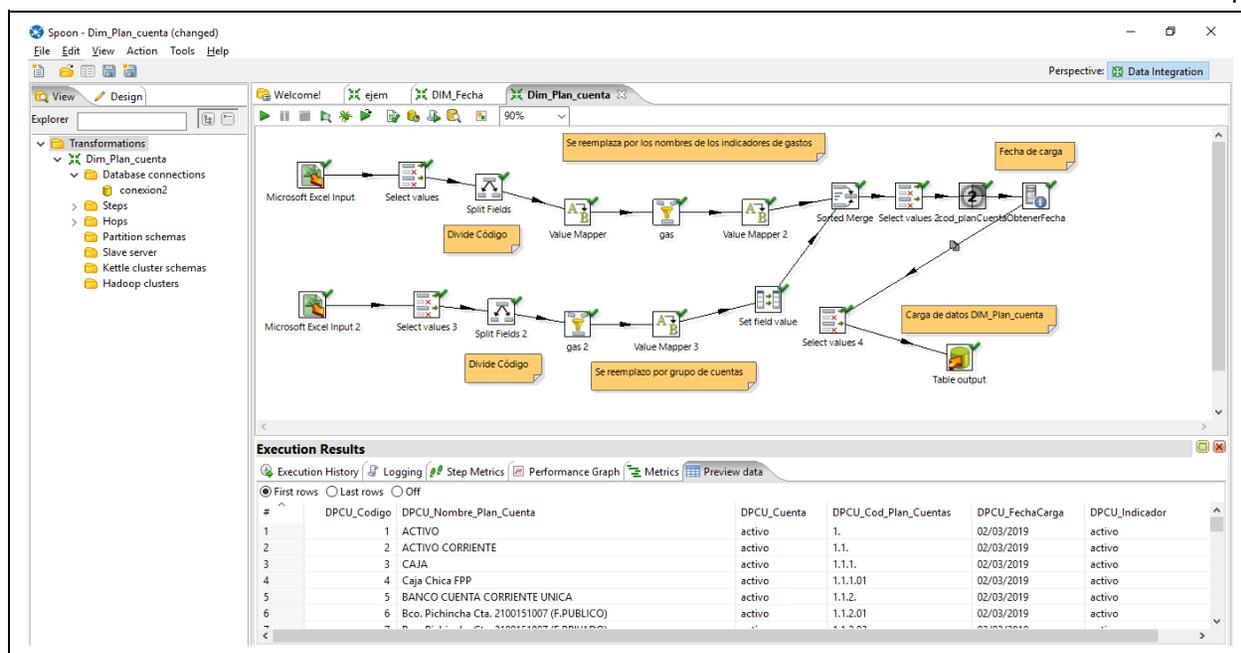


Figura 24. Proceso de Extracción y Carga dimensión Plan Cuenta

En el proceso ETL de la dimensión Plan Cuenta que se observa en la **Figura 24.** Proceso de Extracción y Carga dimensión Plan Cuenta, en la cual se realizaron varias clasificaciones:

- Se clasifico por los grupos de cuentas contables, para generar reportes y análisis por cada grupo.

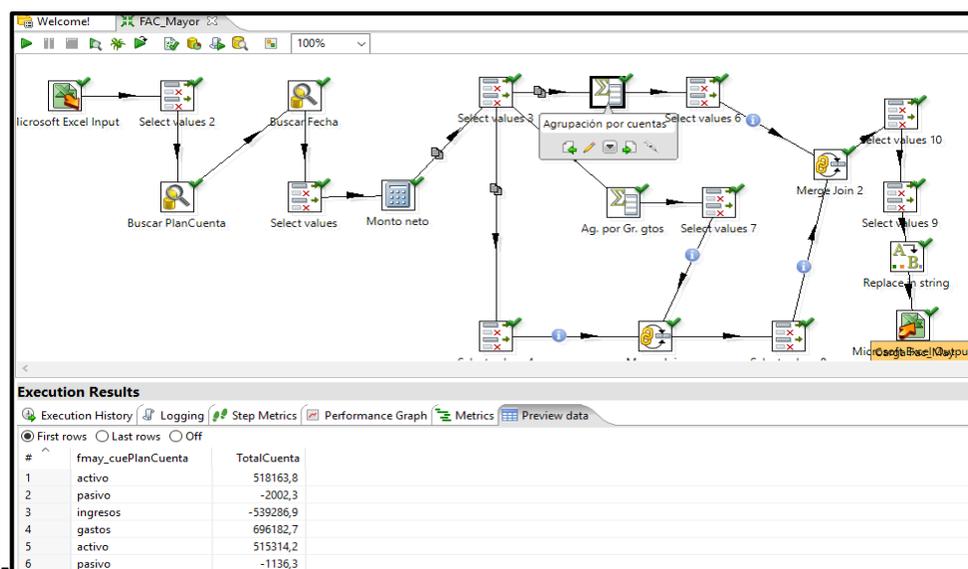


Figura 25. Clasificación por cuentas contables

- Se clasifico por los grupos de indicadores: Funcionamiento Institucional, Publicaciones, Capacitaciones e Investigación y Formación, de acuerdo al código contable asignado, de esta forma se determina los montos incurridos en cada periodo de tiempo, información que se evaluara mediante la aplicación de los indicadores, estableciendo límites de cada asignación y porcentajes de cumplimiento, en las cuales se basaran las decisiones de las autoridades de la organización y tomar acciones correctivas de ser el caso.

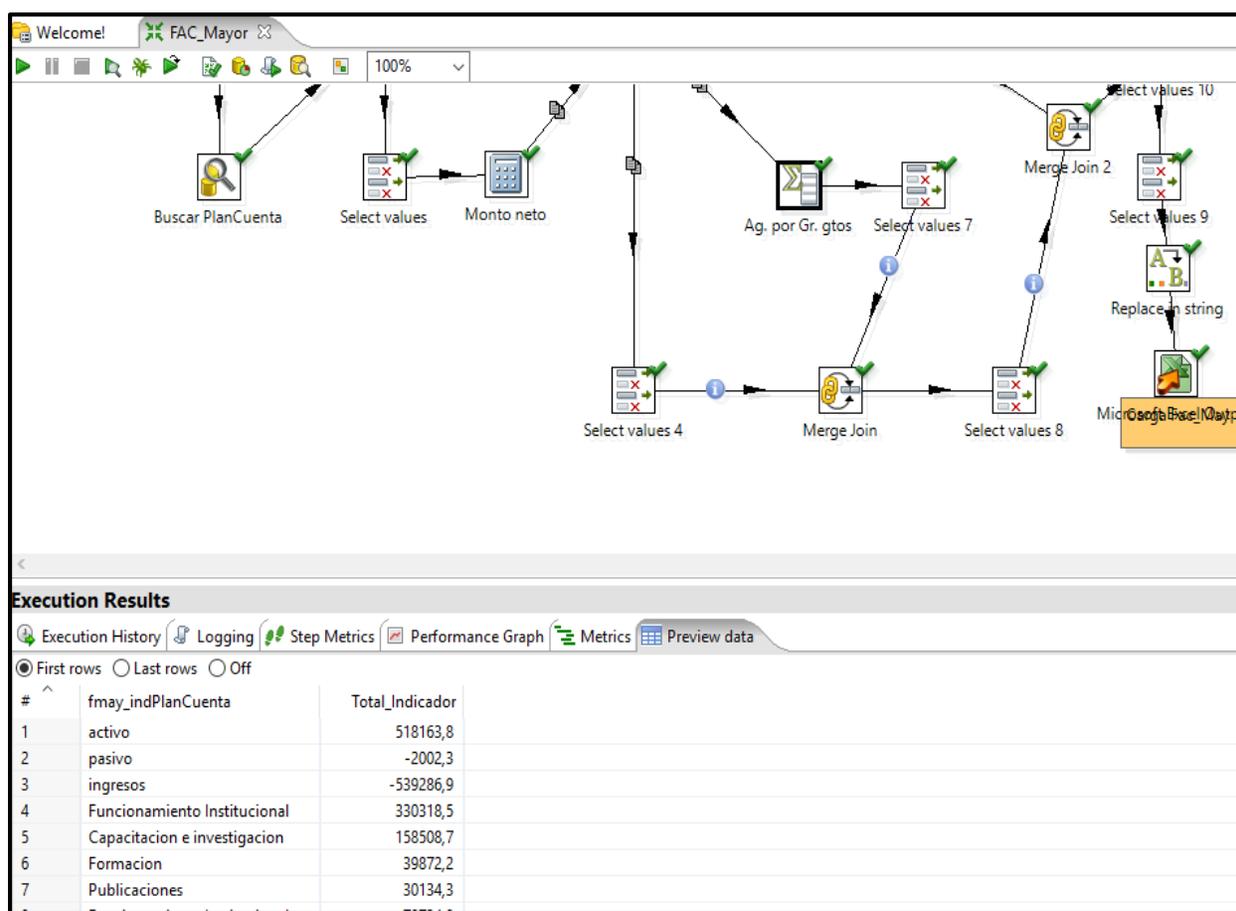


Figura 26. Clasificación por grupo de indicadores

Como fuente destino se generó la tabla dim_plan_cuenta en la base de datos MySql

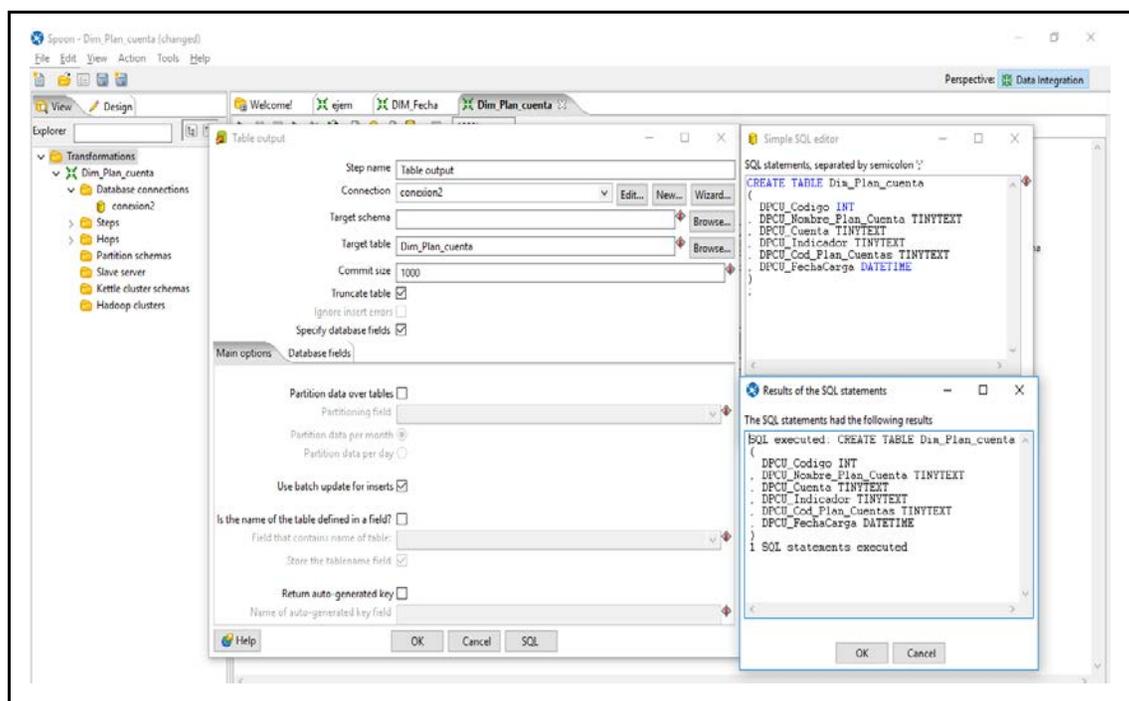


Figura 27. Proceso ETL Plan Cuentas – Fuente Destino
Tabla de hechos Fac Mayor

Se seleccionó como fuente de origen de la tabla de hechos Fac_mayor el archivo Excel Libro Mayor.

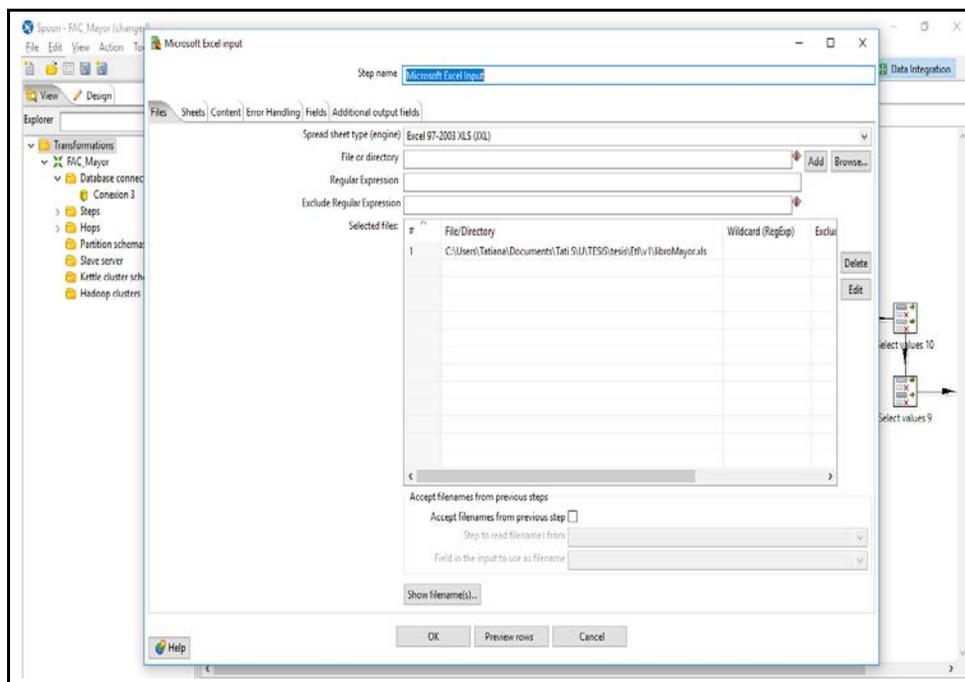


Figura 28. Proceso ETL Tabla de hechos – Fuente Origen

A continuación se observa el proceso ETL de la tabla de hechos Fac_mayor

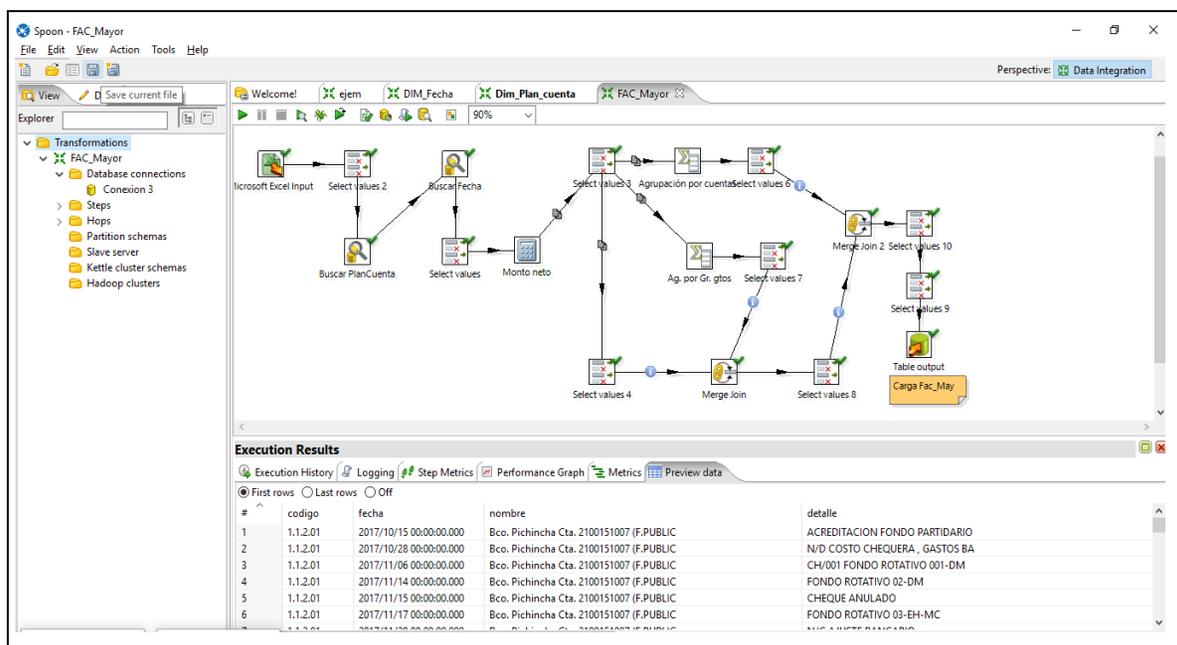


Figura 29. Proceso de extracción y carga Tabla de Hechos

En la tabla de hechos se realizaron varios cálculos y adiciones con el fin de obtener una base de datos idónea para realizar el análisis deseado:

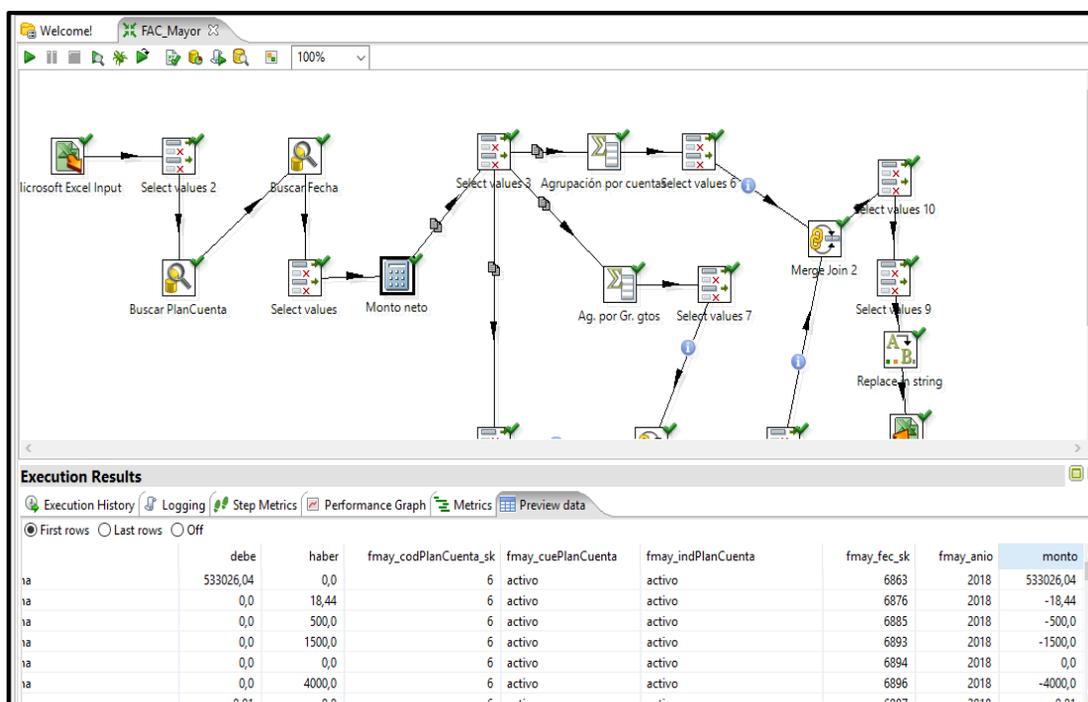


Figura 30. Cálculo de monto por cada transacción

- Se realizó el cálculo de montos por cada transacción: que permitió determinar los montos netos de las transacciones, se restó el debe del haber generado los siguientes resultados: cuentas de gastos y activos son saldos positivos y las cuentas de ingreso, patrimonio y pasivo con saldo negativo como se puede observar en la **Figura 30**. Cálculo de monto por cada transacción.
- Se determinó el monto de las transacciones por grupo contable y grupo de indicadores:

<input checked="" type="radio"/> First rows <input type="radio"/> Last rows <input type="radio"/> Off		
#	fmay_indPlanCuenta	Total_Indicador
1	activo	518163,8
2	pasivo	-2002,3
3	ingresos	-539286,9
4	Funcionamiento Institucional	330318,5
5	Capacitacion e investigacion	158508,7
6	Formacion	39872,2
7	Publicaciones	30134,3

Figura 31. Montos transaccionales por grupo de indicadores

Se seleccionó como fuente destino la tabla fac_mayor en la base de datos MySQL.

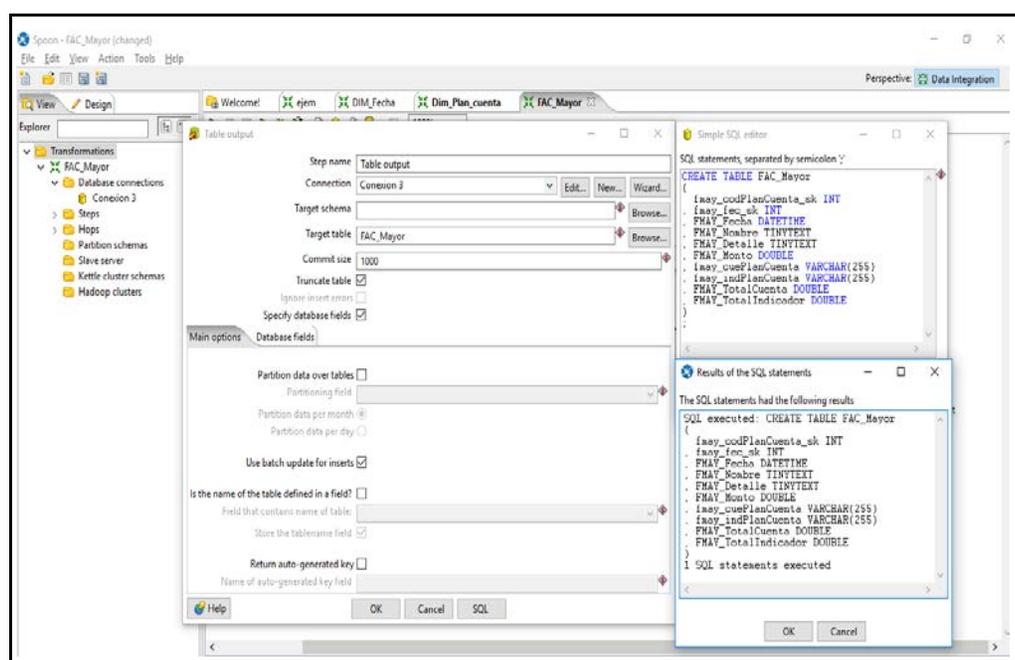


Figura 32. Proceso ETL Tabla de hechos – Fuente Destino

4.3.4. Base de datos

Se creó la base de datos en MySQL designada `base_datos_tesis_v2_18` utilizando la herramienta XAMPP, el cual permite inicializar el servidor gestor de base de datos MySQL y el servidor web Apache, para la configuración de la base de datos en una página web local.

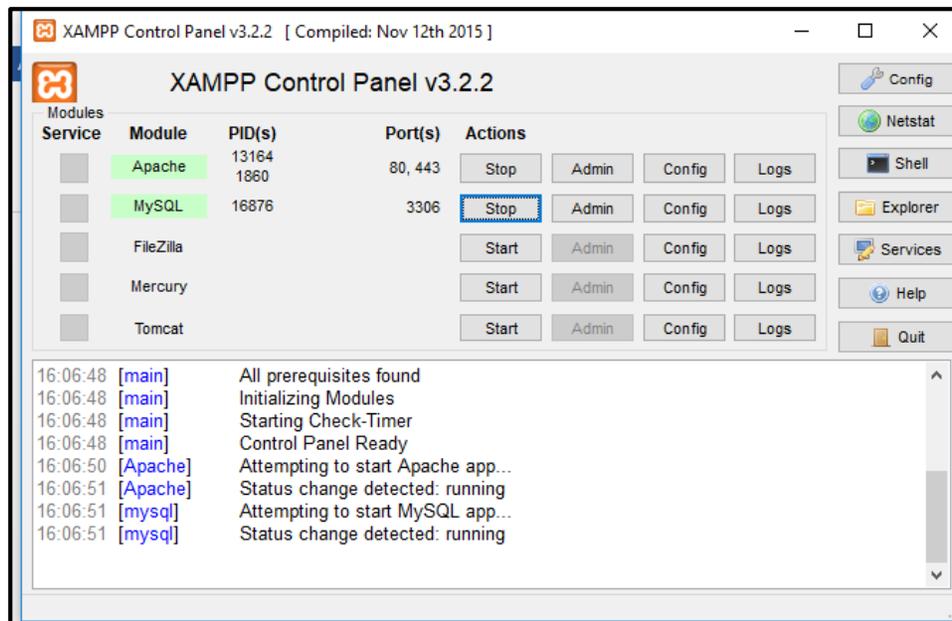


Figura 33. XAMPP

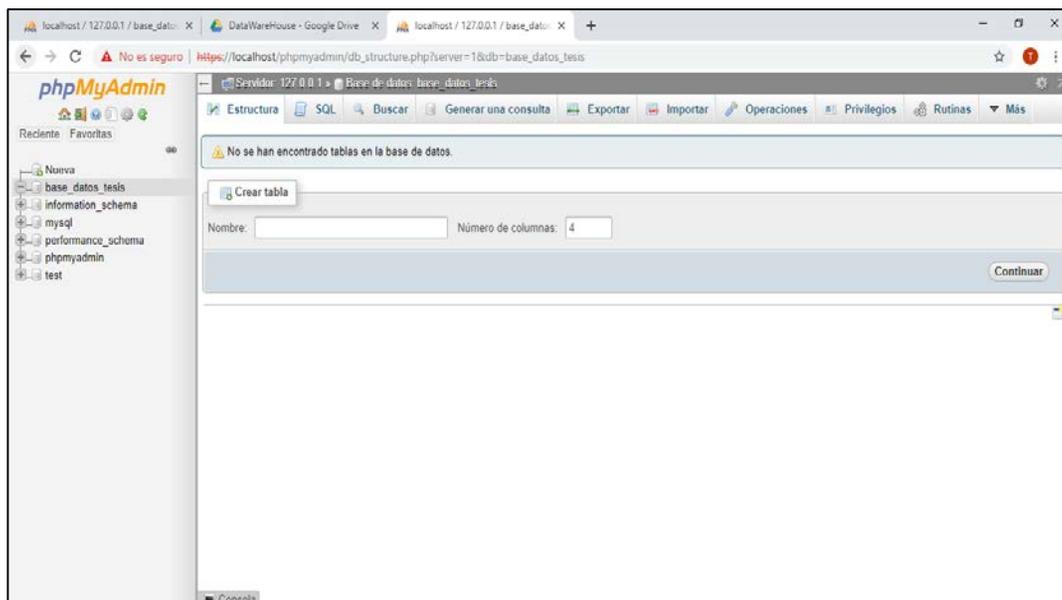


Figura 34. Creación de la base de datos en MySQL

Se cargaron las dimensiones mediante los procesos de ETL a las tablas en la base de datos MySQL, mediante la conexión creada en la herramienta Data Integration de Pentaho, como se muestra a continuación:

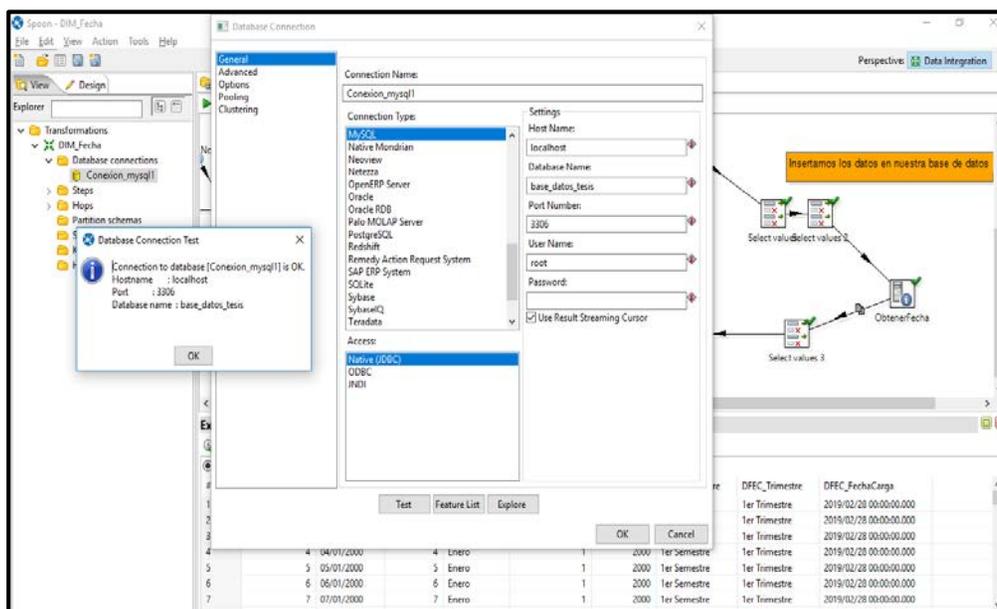


Figura 35. Conexión con la base de datos MySQL mediante Data Integration de Pentaho

Con las conexiones establecidas, cada una de las dimensiones y la tabla de hechos se fueron cargando a la base de datos, mientras se ejecutaban los procesos ETL:

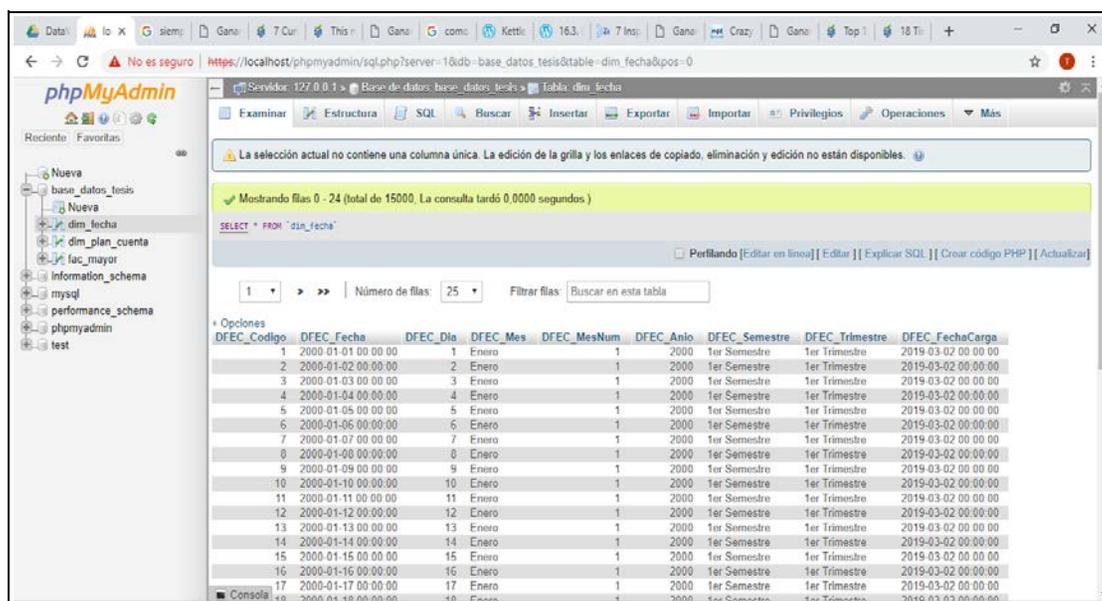


Figura 36. Creación de tabla Dim_fecha en base de datos MySQL

The screenshot shows the phpMyAdmin interface for a MySQL database named 'base_datos_tesis'. The table 'dim_plan_cuenta' is selected, and its structure is displayed. The table has 6 columns: DPCU_Codigo, DPCU_Nombre_Plan_Cuenta, DPCU_Cuenta, DPCU_Indicador, DPCU_Cod_Plan_Cuentas, and DPCU_FechaCarga. The data is as follows:

DPCU_Codigo	DPCU_Nombre_Plan_Cuenta	DPCU_Cuenta	DPCU_Indicador	DPCU_Cod_Plan_Cuentas	DPCU_FechaCarga
1	ACTIVO	activo	activo	1.	2019-03-02 00:00:00
2	ACTIVO CORRIENTE	activo	activo	1.1.	2019-03-02 00:00:00
3	CAJA	activo	activo	1.1.1.	2019-03-02 00:00:00
4	Caja Chica FPP	activo	activo	1.1.1.01	2019-03-02 00:00:00
5	BANCO CUENTA CORRIENTE UNICA	activo	activo	1.1.2.	2019-03-02 00:00:00
6	Bco. Pichincha Cta. 2100151007 (F.PUBLICO)	activo	activo	1.1.2.01	2019-03-02 00:00:00
7	Bco. Pichincha Cta. 2100151007 (F.PRIVADO)	activo	activo	1.1.2.02	2019-03-02 00:00:00
8	ANTICIPO DE FONDOS	activo	activo	1.1.3.	2019-03-02 00:00:00
9	Anticipo a Remuneraciones	activo	activo	1.1.3.01	2019-03-02 00:00:00
10	Anticipos a Contratos	activo	activo	1.1.3.02	2019-03-02 00:00:00
11	Anticipos a Proveedores	activo	activo	1.1.3.03	2019-03-02 00:00:00
12	Otros Anticipos	activo	activo	1.1.3.04	2019-03-02 00:00:00
13	FONDOS ROTATIVOS SEDES PROVINCIALES	activo	activo	1.1.4.	2019-03-02 00:00:00
14	Fondo Rotativo Imbabura	activo	activo	1.1.4.10	2019-03-02 00:00:00
15	Fondo Rotativo Provincia Pichincha	activo	activo	1.1.4.17	2019-03-02 00:00:00
16	FONDO A RENDIR CUENTAS	activo	activo	1.1.5.	2019-03-02 00:00:00
17	Anticipo de Vi-ticos, pasajes y otros de viaje (OP...	activo	activo	1.1.5.01	2019-03-02 00:00:00

Figura 37. Creación de tabla Dim_Plan_cuenta en base de datos MySQL

The screenshot shows the phpMyAdmin interface for a MySQL database named 'base_datos_tesis_v2'. The table 'fac_mayor' is selected, and its structure is displayed. The table has 8 columns: fmay_codPlanCuenta_ek, fmay_fec_ek, FMAY_Fecha, fmay_anio, FMAY_Nombre, FMAY_Detalle, FMAY_Monto, fmay_cuePlanCuenta, and fmay_indPlanCuenta. The data is as follows:

fmay_codPlanCuenta_ek	fmay_fec_ek	FMAY_Fecha	fmay_anio	FMAY_Nombre	FMAY_Detalle	FMAY_Monto	fmay_cuePlanCuenta	fmay_indPlanCuenta
6	6863	2018-10-15 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	ACREDITACION FONDO PARTIDARIO	533026.04	activo	activo
6	6876	2018-10-20 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	N/D COSTO CHEQUERA, GASTOS BA	-10.44	activo	activo
6	6885	2018-11-06 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	CH001 FONDO ROTATIVO 001-DM	-500	activo	activo
6	6893	2018-11-14 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	FONDO ROTATIVO 02-DM	-1500	activo	activo
6	6894	2018-11-15 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	CHEQUE ANULADO	0	activo	activo
6	6896	2018-11-17 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	FONDO ROTATIVO 03-EH-MC	-4000	activo	activo
6	6907	2018-11-28 00:00:00	2018	Bco Pichincha Cta. 2100151007 (F.PUBLICO)	N/C AJUSTE BANCARIO	0.01	activo	activo

Figura 38. Creación de la tabla de hechos en la base de datos MySQL

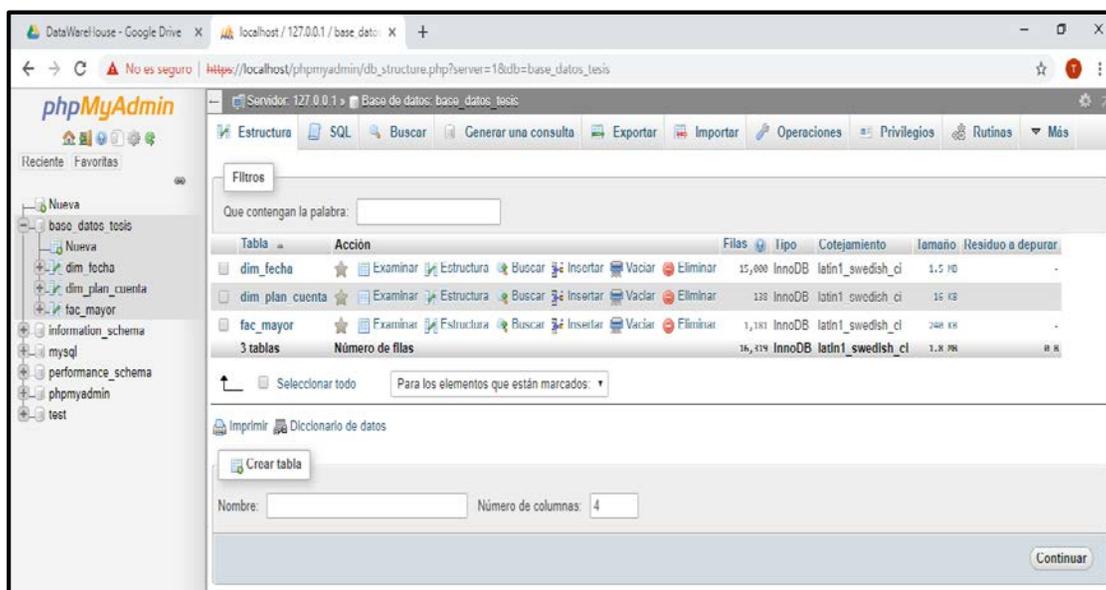


Figura 39. Base de datos “base_datos_tesis_v2_18” con todas las tablas creadas

4.3.5. Selección del Modelo de Minería

Para la realización del modelo de minería se llevó a cabo un estudio comparativo de las herramientas, que nos permita seleccionar la adecuada para el presente estudio, como se observa a continuación:

Tabla 6.
Cuadro Comparativo de Herramientas de Data Mining

	Características	Lenguaje de programación	Sistema operativo	Precio/Licencia
RapidMiner	Apto para todos los procesos. Destaca en el análisis predictivo	Java	Windows, macOS, Linux	Freeware, diferentes versiones de pago
WEKA	Diversos métodos de clasificación	Java	Windows, macOS, Linux	Software libre (GPL)
Orange	Crea una visualización de datos atractiva sin que se requieran muchos conocimientos previos para ello	Python	Windows, macOS, Linux	Software libre (GPL)

Continúa 

	Características	Lenguaje de programación	Sistema operativo	Precio/Licencia
KNIME	Software de código abierto que ha democratizado el acceso a los análisis predictivos	Java	Windows, macos, Linux	Software libre (GPL) (a partir de la versión 2.1)
SAS	Costoso, pero de gran potencia, orientado a grandes empresas	Lenguaje SAS	Windows, macos, Linux	Freeware limitado a instituciones públicas, el precio se establece tras solicitud, diferentes modelos disponibles

Fuente: (Guide, 2018)

La herramienta seleccionada es Rapid Miner, por sus diversas ventajas, destacando que es una herramienta que se encuentra como top en el Cuadrante de Gartner, además de su adición Automodel que permite la creación de varios modelos en similar tiempo de ejecución.

RapidMiner es una plataforma de software para equipos de análisis que integra la preparación de datos, aprendizaje automático y despliegue de modelos predictivos (RapidMiner, 2018), RapidMiner realizó una nueva adición desde la versión 8.1 Auto Model RapidMiner.

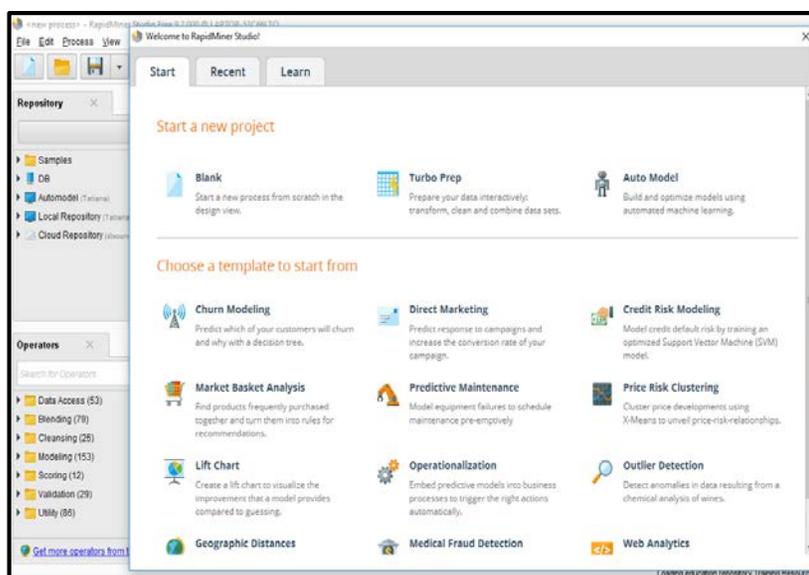


Figura 40. RapidMiner

Auto Model es una extensión de RapidMiner, cuyo objetivo es acelerar los procesos de creación y validación de modelo, ya que los procesos que crea se pueden modificar, adaptándolos a las necesidades propias del análisis que se realiza.

El proceso de minería da inicio tomando los datos de la bodega de datos que fue creada con la ejecución de los procesos ETL en la herramienta de Data Integration Pentaho, con los cual se obtuvo una base en MySQL confiable, que es el punto de partida para el análisis con Auto Model de RapidMiner.

Se creó el Proceso _v2_2018, de acuerdo a la **Figura 41**. Proceso RapidMiner, carga de base de datos, en la cual se conecta a la base de datos MySQL, para obtener la tabla de hechos FAC_Mayor, ya que en esta se encuentran todos los atributos necesarios para el análisis en Auto Model, en el procesamiento se añadió el cálculo del porcentaje de consumo de las partidas presupuestarias, ya que este campo es importante al momento de establecer las alertas.

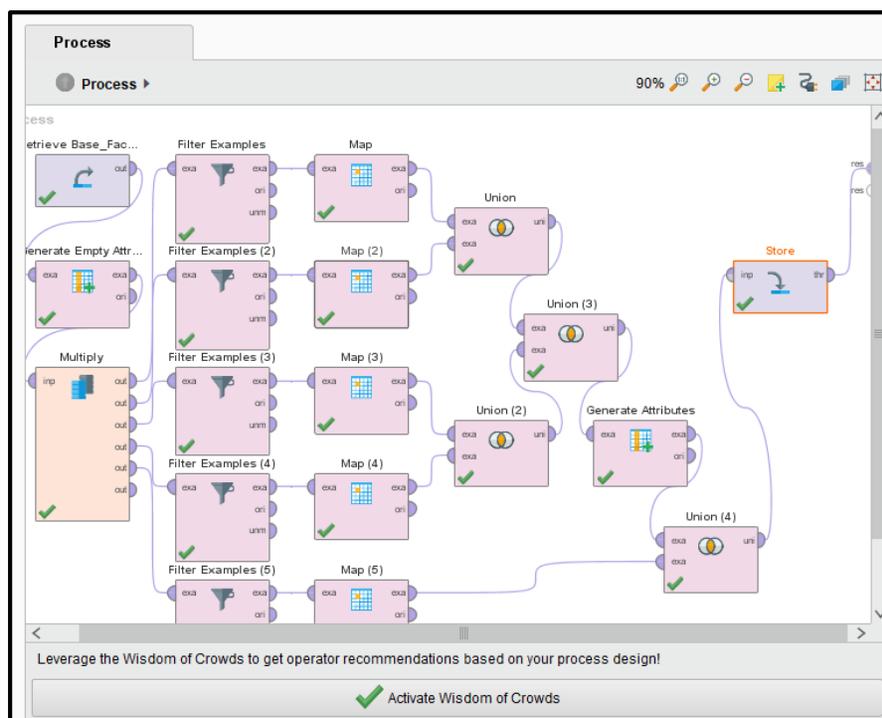
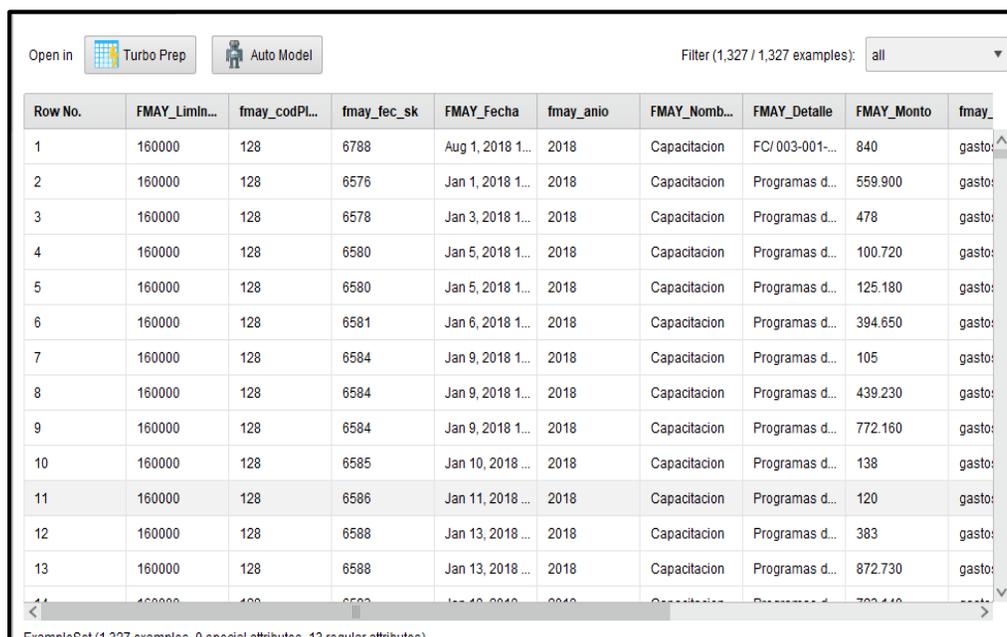


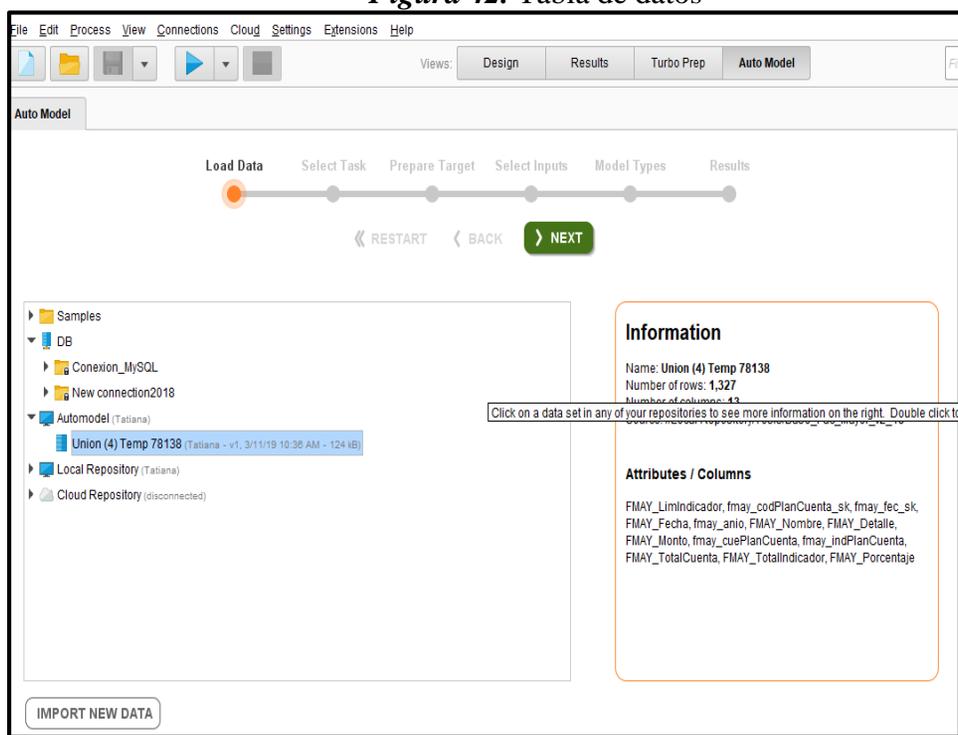
Figura 41. Proceso RapidMiner, carga de base de datos

Cuando se corre el proceso, se obtiene la tabla con los datos para el análisis, se realiza el proceso del Auto Model, que determina las mejores técnicas de minería de datos, de acuerdo a la información y la relación de los atributos.



Row No.	FMay_LimIn...	fmay_codPL...	fmay_fec_sk	FMay_Fecha	fmay_anio	FMay_Nomb...	FMay_Detalle	FMay_Monto	fmay_...
1	160000	128	6788	Aug 1, 2018 1...	2018	Capacitacion	FC/ 003-001-...	840	gasto:
2	160000	128	6576	Jan 1, 2018 1...	2018	Capacitacion	Programas d...	559.900	gasto:
3	160000	128	6578	Jan 3, 2018 1...	2018	Capacitacion	Programas d...	478	gasto:
4	160000	128	6580	Jan 5, 2018 1...	2018	Capacitacion	Programas d...	100.720	gasto:
5	160000	128	6580	Jan 5, 2018 1...	2018	Capacitacion	Programas d...	125.180	gasto:
6	160000	128	6581	Jan 6, 2018 1...	2018	Capacitacion	Programas d...	394.650	gasto:
7	160000	128	6584	Jan 9, 2018 1...	2018	Capacitacion	Programas d...	105	gasto:
8	160000	128	6584	Jan 9, 2018 1...	2018	Capacitacion	Programas d...	439.230	gasto:
9	160000	128	6584	Jan 9, 2018 1...	2018	Capacitacion	Programas d...	772.160	gasto:
10	160000	128	6585	Jan 10, 2018 ...	2018	Capacitacion	Programas d...	138	gasto:
11	160000	128	6586	Jan 11, 2018 ...	2018	Capacitacion	Programas d...	120	gasto:
12	160000	128	6588	Jan 13, 2018 ...	2018	Capacitacion	Programas d...	383	gasto:
13	160000	128	6588	Jan 13, 2018 ...	2018	Capacitacion	Programas d...	872.730	gasto:

Figura 42. Tabla de datos



The screenshot shows the 'Auto Model' interface in a software application. At the top, there are menu options: File, Edit, Process, View, Connections, Cloud, Settings, Extensions, Help. Below the menu is a toolbar with icons for file operations and a 'Views' section with buttons for Design, Results, Turbo Prep, and Auto Model. The main area is titled 'Auto Model' and features a progress bar with five steps: Load Data (highlighted with an orange circle), Select Task, Prepare Target, Select Inputs, Model Types, and Results. Below the progress bar are 'RESTART', 'BACK', and 'NEXT' buttons. On the left, a file explorer shows a tree structure with folders like 'Samples', 'DB', 'Conexion_MySQL', 'New connection2018', 'Automodel (Tatiana)', 'Local Repository (Tatiana)', and 'Cloud Repository (disconnected)'. A data set 'Union (4) Temp 78138' is selected. On the right, an 'Information' panel displays details for the selected data set: Name: Union (4) Temp 78138, Number of rows: 1,327, and Number of columns: 13. Below this, an 'Attributes / Columns' list includes: FMay_LimIndicador, fmay_codPlanCuenta_sk, fmay_fec_sk, FMay_Fecha, fmay_anio, FMay_Nombre, FMay_Detalle, FMay_Monto, fmay_cuePlanCuenta, fmay_indPlanCuenta, FMay_TotalCuenta, FMay_TotalIndicador, and FMay_Porcentaje. A tooltip message reads: 'Click on a data set in any of your repositories to see more information on the right. Double click to...'. At the bottom left, there is an 'IMPORT NEW DATA' button.

Figura 43. Paso 1: Auto Model – Carga de datos

El principal objetivo del proyecto es cumplir los requerimientos institucionales internos y externos de los organismos de control a través del uso aplicativo de indicadores, por lo que se determina que el atributo a predecir son los gastos en los que incurre la organización, los cuales son clasificados en 4 grupos con límites determinados, de acuerdo a lo descrito en la **Tabla 3**.

De esta forma se continuó con el Paso 2 de Auto Model, en el que se indica la tarea a realizar en el proceso de minería de datos, en este caso se busca predecir los valores de la columna FMAY_Porcentaje, ya que está relacionada con el consumo de la partida presupuestaria de los gastos incurridos.

Con la selección del atributo a predecir, la herramienta genera un modelo automático que predice los valores del atributo elegido en función a los valores de las otras columnas.

cha	fmay_anio	FMAY Nombre	FMAY Detalle	FMAY Monto	fmay_cuePla...	fmay_indPla...	FMAY_TotalC...	FMAY_Total...	FMAY_Porce...
	Number	Category	Category	Number	Category	Category	Number	Number	Number
17 12...	2017	Capacitacion	FC/003-001-3...	840	gastos	Capacitacion e...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	559.900	gastos	Capacitacion e ...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	478	gastos	Capacitacion e ...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	100.720	gastos	Capacitacion e ...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	125.180	gastos	Capacitacion e ...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	394.650	gastos	Capacitacion e ...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	105	gastos	Capacitacion e ...	696182.720	158508.730	99.068
17 12...	2017	Capacitacion	Programas de ...	430.220	gastos	Capacitacion e ...	696182.720	158508.730	99.068

1,327 rows - 13 columns (4 nominal, 8 numerical, 1 date)

Figura 44. Paso 2: Auto Model – Seleccionar tarea

En el paso 3 la herramienta permite preparar el objetivo con el fin de tener mejores resultados, el atributo seleccionado es numérico por lo que tendrá como respuesta una regresión con un valor exacto calculado.

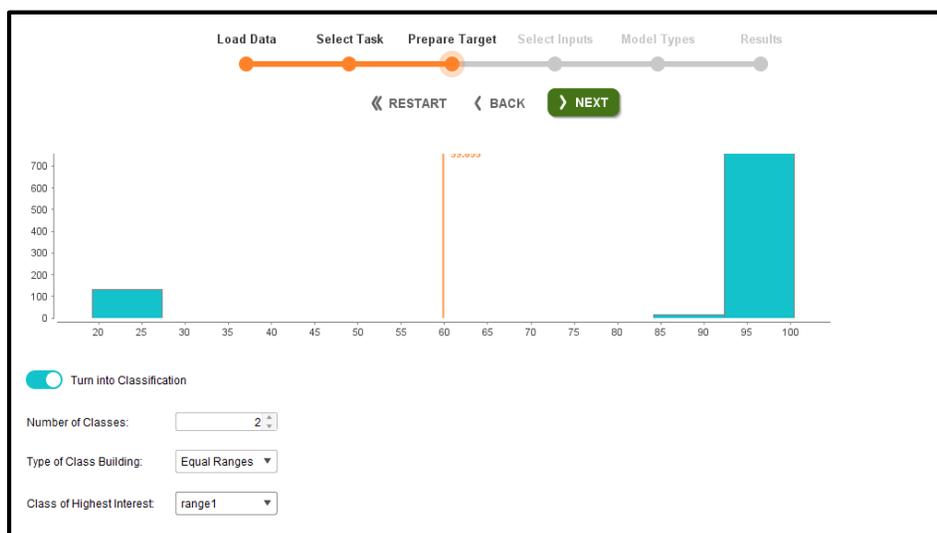


Figura 45. Paso 3: Auto Model – Preparar objetivo

En el paso 4 se selecciona los datos de entrada para el modelo, determinados por la calidad de los mismos y de acuerdo al atributo a predecir, la herramienta sugiere atributos y brinda visualizaciones sobre los porcentajes de correlación con el atributo destino, valores diferentes, valores perdidos y valores idénticos, con esta información se puede determinar cuáles son los atributos que mejor se adaptan al análisis que se está realizando.

Selected	Status	Quality	Name	Correlation	ID-ness	Stability	Missing	Text-ness
<input type="checkbox"/>	Red	Low	FMAI_TotalCuenta	?	?	93.14%	0.00%	0.00%
<input checked="" type="checkbox"/>	Yellow	High	fmay_anio	87.75%	0.15%	89.00%	0.00%	0.00%
<input checked="" type="checkbox"/>	Green	High	FMAI_LimIndicador	0.01%	?	49.96%	0.00%	0.00%
<input checked="" type="checkbox"/>	Green	High	fmay_codPlanCuenta...	0.10%	4.37%	19.59%	0.00%	0.00%

Figura 46. Paso 4: Auto Model – Seleccionar entradas

Una vez seleccionadas las entradas, la herramienta genera visualizaciones de los modelos pertinentes de aprendizaje automático, se pueden seleccionar varios modelos, y Auto Model incluye una comparación de rendimientos.

La herramienta sugiere los modelos que se puede seleccionar de acuerdo a los datos que se cargan, lo que ayuda a comparar y elegir el modelo que mejor se adapte a las necesidades del problema que se quiere resolver.

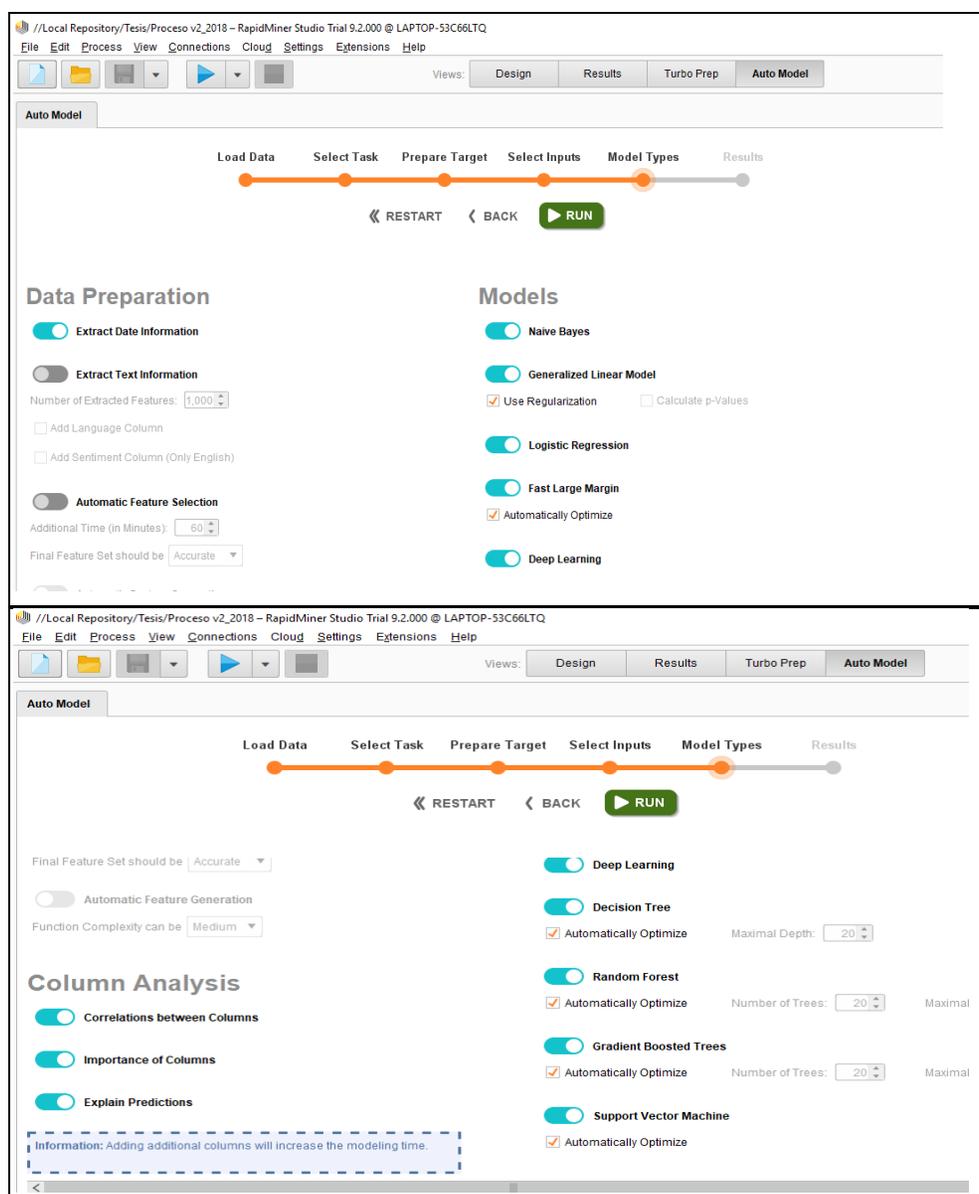


Figura 47. Paso 5: Auto Model – Tipo de Modelos

En el paso final, se selecciona la opción ejecutar, Auto Model presentó los resultados que se obtuvieron en los modelos de manera general e independiente.

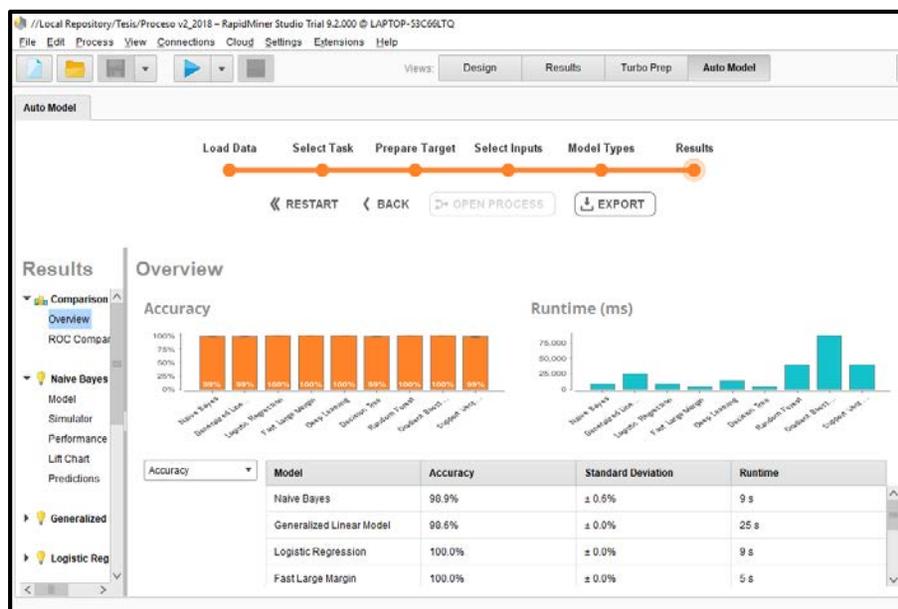


Figura 48. Paso 6: Auto Model – Resultados

La herramienta Auto Model permite observar el proceso de los modelos que se selecciona como se muestra en la **Figura 49**. Proceso creado por Auto Model.

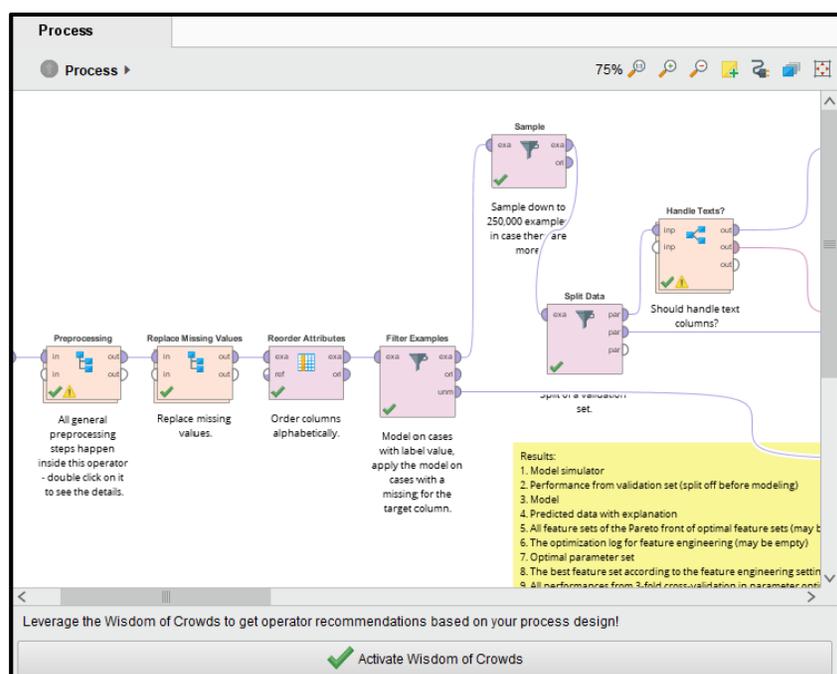


Figura 49. Proceso creado por Auto Model

4.3.6. Modelo de Minería de datos

Con el fin de determinar los atributos idóneos para el modelo, se analizó la tabla de atributos de entrada que sugiere el Auto Model en el paso 4, como se observa a continuación:

Tabla 7.

Atributos seleccionados por el Auto Model

Atributos	Recomendación Herramienta	Correlación	Estabilidad	Datos faltantes
fmay_cue_PlanCuenta	rojo	-	93%	0%
FMAY_TotalCuenta	rojo	-	93%	0%
fmay_anio	amarillo	88%	89%	0%
FMAY_LimIndicado	verde	0%	50%	0%
fmay_codPlanCuenta	verde	0%	20%	0%
fmay_fec_sk	verde	28%	2%	0%
FMAY_Fecha	verde	28%	2%	0%
FMAY_Nombre	verde	3%	20%	0%
FMAY_Detalle	verde	5%	31%	0%
FMAY_Monto	verde	1%	4%	0%
fmay_id_Plan_Cuenta	verde	0%	50%	0%
FMAY_TotalIndicador	verde	23%	45%	0%

Mediante los valores que arroja la herramienta conjuntamente con la experiencia y conocimiento de información financiera, se determinaron los siguientes atributos para el modelo de minería:

Tabla 8.

Atributos de entrada escogidos

Atributos	Recomendación Herramienta	Correlación	Estabilidad	Datos faltantes
FMAY_LimIndicado	verde	0%	50%	0%
fmay_anio	verde	12%	89%	0%
FMAY_Monto	verde	1%	4%	0%
fmay_id_Plan_Cuenta	verde	0%	50%	0%
FMAY_TotalIndicador	verde	23%	45%	0%

Previo al modelado, se creó un proceso con el fin de determinar el consumo de las partidas presupuestarias (porcentaje), con este cálculo se establece las alertas en las que se basarán las decisiones de las autoridades del partido, creando una nueva tabla denominada Base_Fac_Mayor_v3_18.

Fecha	fmay_anio	FMAY_Nomb...	FMAY_Detalle	FMAY_Monto	fmay_cuePL...	fmay_indPla...	FMAY_Total...	FMAY_Total...	FMAY_Porc...
2018 1...	2018	Capacitacion	FC/003-001-...	840	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	559.900	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	478	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	100.720	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	125.180	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	394.650	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	105	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	439.230	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 1...	2018	Capacitacion	Programas d...	772.160	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 ...	2018	Capacitacion	Programas d...	138	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 ...	2018	Capacitacion	Programas d...	120	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 ...	2018	Capacitacion	Programas d...	383	gastos	Capacitacion ...	696182.720	158508.730	99.068
2018 ...	2018	Capacitacion	Programas d...	872.730	gastos	Capacitacion ...	696182.720	158508.730	99.068

Figura 50. Tabla de Datos “Base_Fac_Mayor_v3_18

Se selecciona la tabla para comenzar la creación en el Auto Model:

The screenshot displays the 'Auto Model' workflow. The 'Load Data' step is highlighted, indicating the selection of the 'Base_Fac_Mayor_v3_18' table. The table's metadata is shown in the 'Information' panel, and a preview of its columns is visible in the lower section.

Role	Name	Type	Range	Missings
	fmay_codP...	integer	=[-6 - 138]	= 0
	fmay_fec_sk	integer	=[8576 - 7...	= 0
	FMAY_Fecha	date_time	=[Jan 1, 20...	= 0
	fmay_anio	integer	=[2018 - 2...	= 0

Figura 51. Selección de datos para el modelo

Luego se definió el atributo a predecir fmay_indPlanIndicador

The screenshot shows the 'Auto Model' interface with the 'Select Task' step active. A progress bar at the top indicates the sequence: Load Data, Select Task, Prepare Target, Select Inputs, Model Types, Results. Below the progress bar are buttons for 'RESTART', 'BACK', and 'NEXT'. Three task options are presented: 'Predict' (highlighted in orange), 'Clusters', and 'Outliers'. Below these options is a data table with columns: codPla..., fmay_fec_sk, FMAY_Fecha, fmay_anio, FMAY_Nombre, FMAY_Detalle, FMAY_Monto, fmay_cuePla..., fmay_indPla..., FMAY_TotalC..., FMAY_Total..., and FMAY_Porce... The 'fmay_indPla...' column is highlighted in orange.

codPla...	fmay_fec_sk	FMAY_Fecha	fmay_anio	FMAY_Nombre	FMAY_Detalle	FMAY_Monto	fmay_cuePla...	fmay_indPla...	FMAY_TotalC...	FMAY_Total...	FMAY_Porce...
Number	Number	Date / Time	Number	Category	Category	Number	Category	Category	Number	Number	Number
6788		Aug 1, 2018 12:...	2018	Capacitacion	FC/ 003-001-3...	840	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6576		Jan 1, 2018 12:...	2018	Capacitacion	Programas de ...	559.900	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6578		Jan 3, 2018 12:...	2018	Capacitacion	Programas de ...	478	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6580		Jan 5, 2018 12:...	2018	Capacitacion	Programas de ...	100.720	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6580		Jan 5, 2018 12:...	2018	Capacitacion	Programas de ...	125.180	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6581		Jan 6, 2018 12:...	2018	Capacitacion	Programas de ...	394.650	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6584		Jan 9, 2018 12:...	2018	Capacitacion	Programas de ...	105	gastos	Capacitacion e ...	696182.720	158508.730	99.068
6584		Jan 9, 2018 12:...	2018	Capacitacion	Programas de ...	420.220	gastos	Capacitacion e ...	696182.720	158508.730	99.068

Figura 52. Selección de atributo a predecir

Como siguiente paso se observa las opciones de la variable a predecir

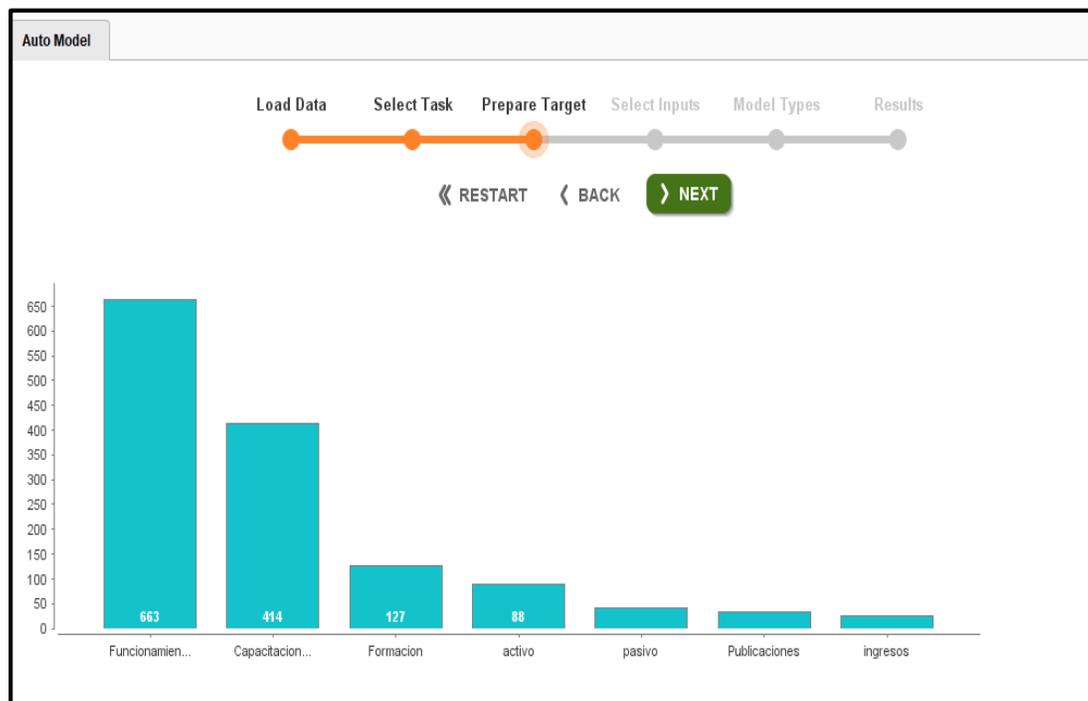


Figura 53. Opciones de la variable a predecir

A continuación, se seleccionó los atributos de entrada:

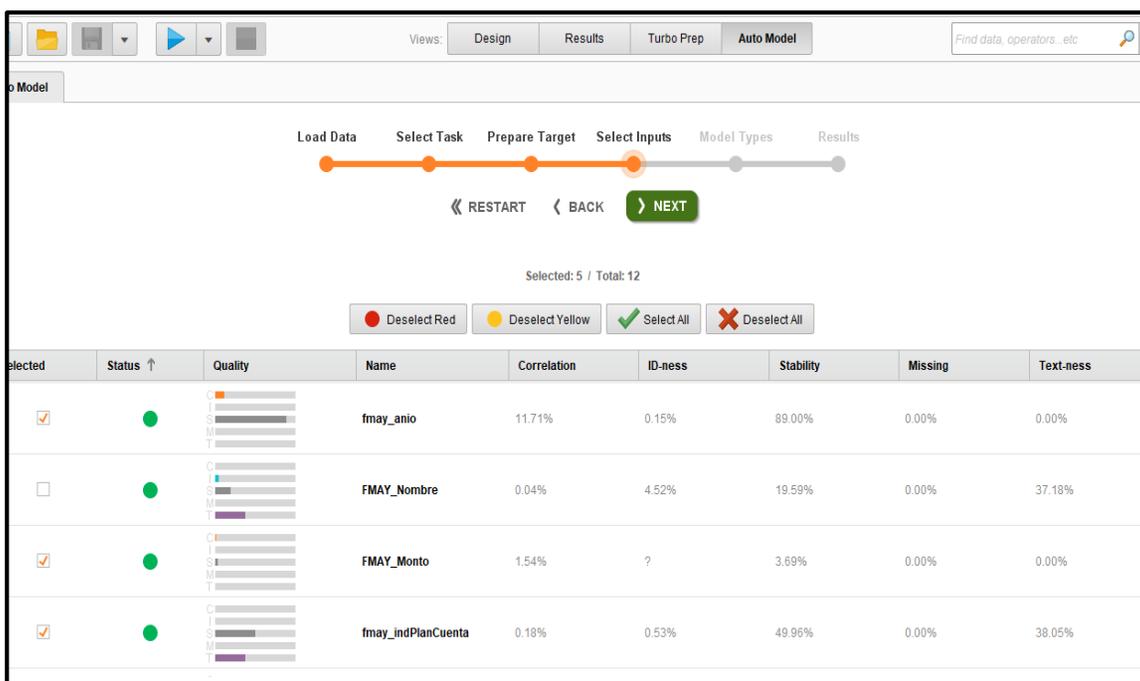


Figura 54. Atributos de entrada

Una vez seleccionados los atributos de entrada, se seleccionan los modelos y se comparan los algoritmos:

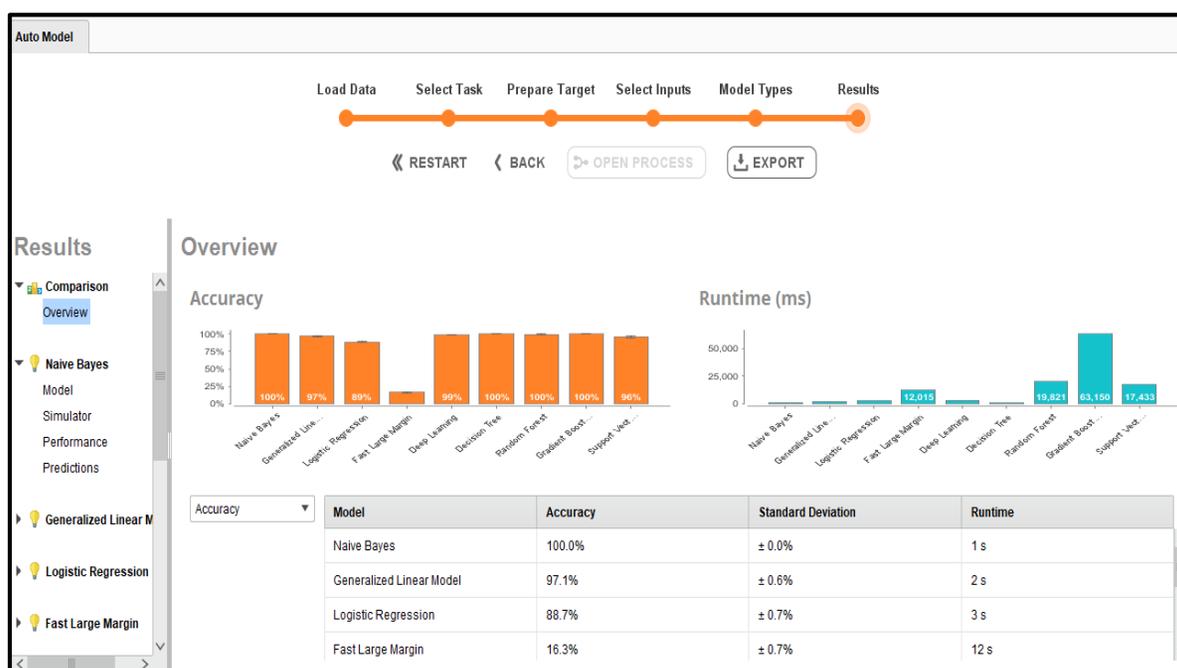


Figura 55. Comparación de algoritmos

Algoritmo Naive Bayes: es un clasificador, en el cual asume que características determinadas no se relacionan con la presencia o ausencia de cualquier otra característica.

Como ejemplo en la herramienta, seleccionamos el algoritmo Naive Bayes, se observan resultados como:

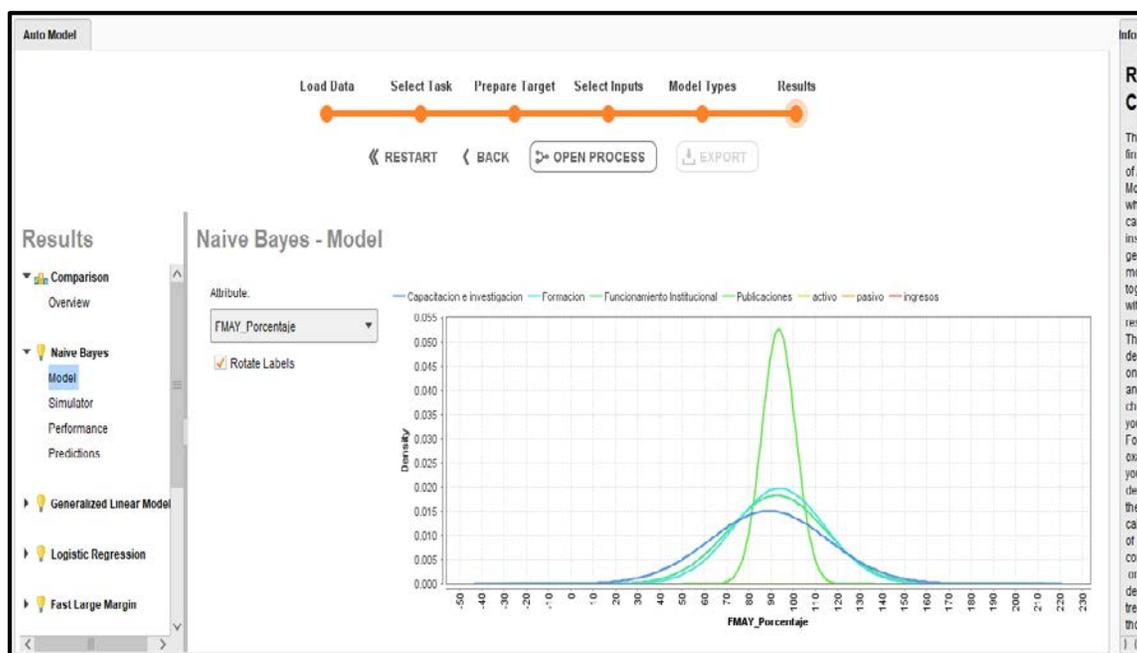


Figura 56. Resultado con algoritmo Naive Bayes

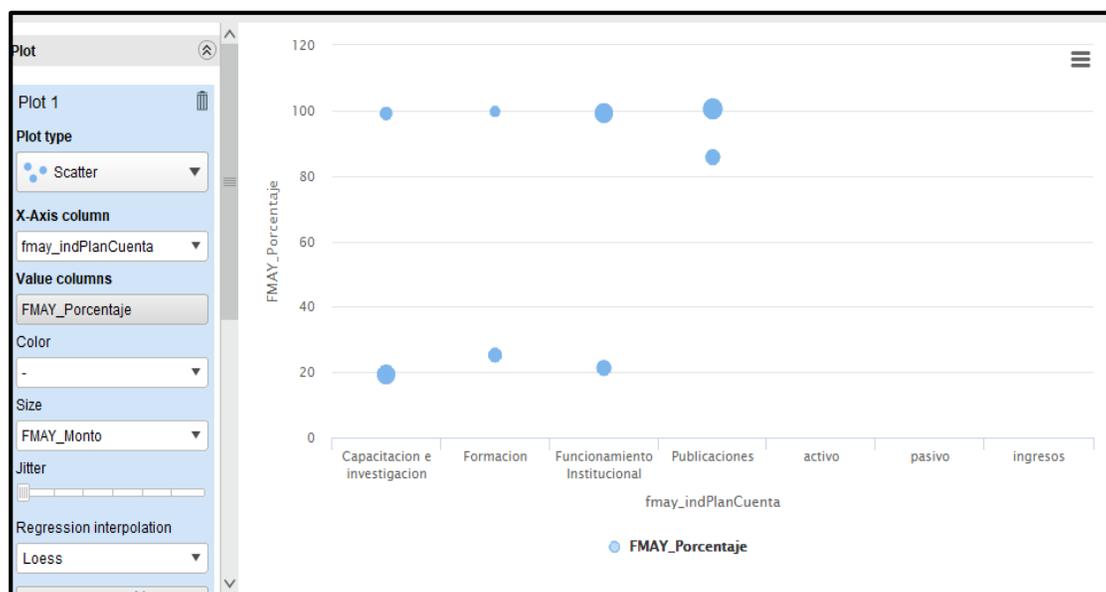


Figura 57. Gáfico Naive Bayes

En la **Figura 57**. Gáfico Naive Bayes, se puede observar que los gastos por publicaciones son los que alcanzan el mayor porcentaje de consumo en las partidas presupuestarias tanto en el año 2018 (100%), como en el 2019 (85.70%), y en contraste los gastos por capacitación e investigación son los que menor consumo tienen en el año 2018 (99%) y el año 2019 (19,26%).

Decision Tree: Un árbol de decisión es un mapa de los posibles resultados de una serie de decisiones relacionadas. Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios (Lucid Software Inc., 2019).

En el modelo Decision Tree – Model la herramienta mostro lo siguiente:

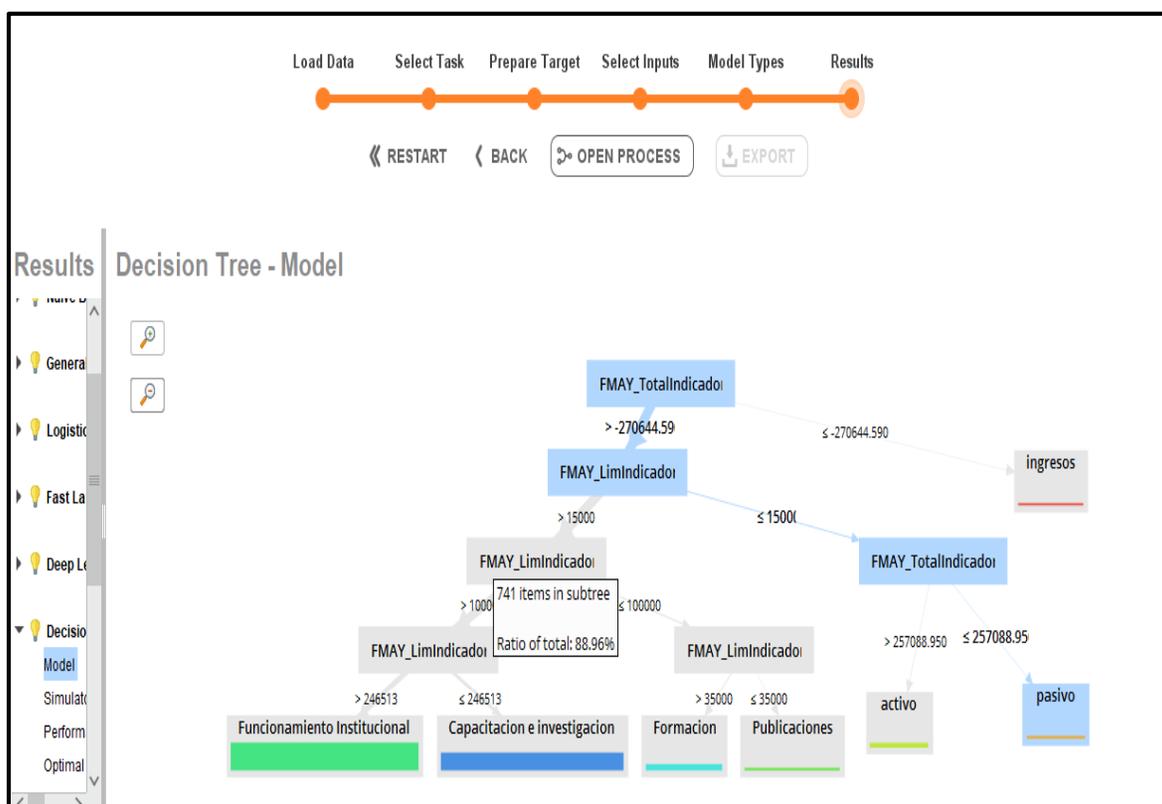


Figura 58. Resultado con algoritmo Decision Tree

En la **Figura 58**. Resultado con algoritmo Decision Tree, se puede estimar un conjunto de reglas de decisión que indica los patrones de comportamiento de la información financiera del partido Izquierdo Democrática, por ejemplo los límites de indicadores positivos corresponden

a los grupos de gastos y activos, mientras que los valores en negativos corresponde a los ingresos y pasivos ya que estas cuentas contables son de naturaleza acreedora, en el modelo también se determina que los gastos que cuentan con mayor asignación presupuestaria son: Formación Institucional y Capacitación e Investigación y en contraste los gastos que menos asignación tiene son : Formación y Publicaciones.

De igual forma con el modelo se puede observar que la empresa cuenta con mayores activos que pasivos.

Se obtuvo las reglas de decisión generadas por el árbol de decisión creado.

Tree

```

FMAY_TotalIndicador > -270644.590
|   FMAY_LimIndicador > 15000
|   |   FMAY_LimIndicador > 100000
|   |   |   FMAY_LimIndicador > 246513: Funcionamiento Institucional
{Capacitacion e investigacion=0, Formacion=0, Funcionamiento Institucional=398,
Publicaciones=0, activo=0, pasivo=0, ingresos=0}
|   |   |   FMAY_LimIndicador ≤ 246513: Capacitacion e investigacion {Capacitacion
e   investigacion=248,   Formacion=0,   Funcionamiento   Institucional=0,
Publicaciones=0, activo=0, pasivo=0, ingresos=0}
|   |   FMAY_LimIndicador ≤ 100000
|   |   |   FMAY_LimIndicador > 35000: Formacion {Capacitacion e investigacion=0,
Formacion=76,   Funcionamiento   Institucional=0,   Publicaciones=0,   activo=0,
pasivo=0, ingresos=0}
|   |   |   FMAY_LimIndicador ≤ 35000: Publicaciones {Capacitacion e
investigacion=0, Formacion=0, Funcionamiento Institucional=0, Publicaciones=19,
activo=0, pasivo=0, ingresos=0}
|   FMAY_LimIndicador ≤ 15000

```

```
| | FMAY_TotalIndicador > 257088.950: activo {Capacitacion e investigacion=0,
Formacion=0, Funcionamiento Institucional=0, Publicaciones=0, activo=53,
pasivo=0, ingresos=0}
```

```
| | FMAY_TotalIndicador ≤ 257088.950: pasivo {Capacitacion e investigacion=0,
Formacion=0, Funcionamiento Institucional=0, Publicaciones=0, activo=0,
pasivo=25, ingresos=0}
```

```
FMAY_TotalIndicador ≤ -270644.590: ingresos {Capacitacion e investigacion=0,
Formacion=0, Funcionamiento Institucional=0, Publicaciones=0, activo=0,
pasivo=0, ingresos=14}
```

4.3.7. Evaluación del modelo

Como paso final se evaluó el modelo analítico mediante el operador Performance que la herramienta RapidMiner contiene.

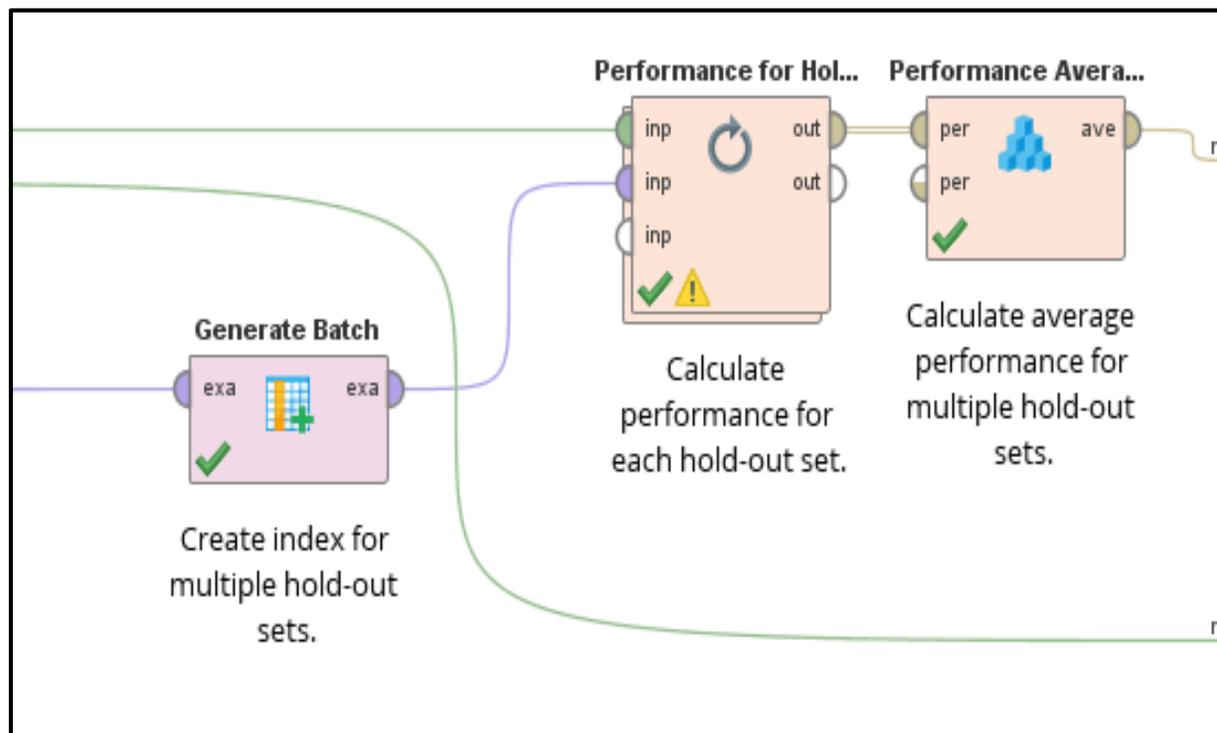


Figura 59. Proceso de Evaluación de Modelo

A través de esta técnica, se comparan los valores predichos por el modelo con los valores reales, de esta forma se determina la precisión del mismo.

Table View Plot View

accuracy: 100.00% +/- 0.00% (micro average: 100.00%)

	true Capacitaci...	true Formacion	true Funcionam...	true Publicacion...	true activo	true pasivo	true ingresos	class precision
pred. Capacitac...	118	0	0	0	0	0	0	100.00%
pred. Formacion	0	37	0	0	0	0	0	100.00%
pred. Funciona...	0	0	189	0	0	0	0	100.00%
pred. Publicacio...	0	0	0	9	0	0	0	100.00%
pred. activo	0	0	0	0	25	0	0	100.00%
pred. pasivo	0	0	0	0	0	11	0	100.00%
pred. ingresos	0	0	0	0	0	0	7	100.00%
class recall	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	

Figura 60. Matriz de precisión

Según la matriz de precisión que se observa tenemos un nivel de confianza del 100%

4.3.8. Dashboard

Una vez realizado el análisis de los datos obtenidos con los procesos ETL y la aplicación de minería de datos, se obtuvo el comportamiento de la asignación presupuestaria en gastos y porcentajes en que se cumplieron las partidas. Como paso final se realizó dashboard con el fin de que las autoridades de la organización puedan visualizar la información relevante de manera gráfica para poder entenderla e interpretarla adecuadamente, ya que en esta se basa las decisiones.

Para la elección de la herramienta se realizó estudios comparativos de las siguientes herramientas:

Tabla 9.

Cuadro comparativo de Herramientas líderes de BI

Logo			
Definición	Herramienta de Microsoft con servicio en la nube con fácil implementación, permite subir, compartir y tener acceso a informes desde cualquier dispositivo, permite acceso a orígenes a datos locales, bases de datos y servicios en la nube.	Herramienta de BI que permite la visualización interactiva de los datos.	Herramienta enfocada al análisis visual de los datos y aplicaciones interactivas cuyo objetivo es mejorar el proceso de acceso a los datos de cara al usuario

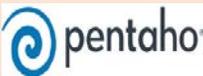
Continúa 

	POWER BI	TABLEAU	QLIK
Ventajas	<ul style="list-style-type: none"> * Incremento de la eficiencia para extraer informes autónomas sin tener conocimiento específico * Presenta una herramienta Quick Insights que determina las correlaciones y patrones dentro de sus datos * Integración y análisis avanzados a través de scripts y objetos visuales de R, Microsoft Azure 	<ul style="list-style-type: none"> * Profundizar y explorar datos sin necesidad de experiencia en programación. * Capacidad de conexión a diversas fuentes de datos. * Si tiene experiencia en programación, Tableau puede funcionar como front-end para realizar inmersiones estadísticas profundas y análisis avanzado. 	<ul style="list-style-type: none"> * Los conocimientos de datos se generan rápidamente a partir de un usuario competente. * El motor asociativo subyacente realiza uniones naturales en tiempo real en función de las selecciones del usuario.

Fuente: (Monzón, 2018)

Tabla 10.

Cuadro comparativo de Herramientas líderes de BI open source

	PENTAHO	RAPIDMINER	BIRT
Logo			
Definición	Pentaho ofrece una versión gratuita de su suite que incluye análisis de negocios, así como la integración de datos, un diseñador de informes y un mercado de plugins	Plataforma de análisis de código libre para data mining, integra un aprendizaje automático y análisis predictivo	Birt es un software de código abierto que proporciona la plataforma Birt, con este se crea visualizaciones de datos e informes que se pueden incrustar en aplicaciones del cliente web, especialmente los basados en Java y Java EE
Ventajas	*Arquitectura flexible y de calidad	* Funciona con una amplia gama de formatos de datos: hadoop, excel, csv, Oracle y MySQL	* Respaldo por una gran comunidad de usuarios
	*Bajo costo	*Se puede ampliar con plugins adicionales	*Soporte comunitario, detectando rápidamente falencias
	*Madurez de la tecnología	*El starter edition es gratuito y permite el acceso a los datos de excel y csv	
	*Alto desempeño, estabilidad y escalabilidad	* Cuenta con versiones personales, profesionales y empresariales desde US\$999	

Fuente: (CulturaCrm, 2016)

Power BI, es una herramienta de análisis de la información detallada que permite tomar decisiones rápidas e informadas (Power BI, 2019). Power BI además cuenta con Power BI Desktop,

que permite una descarga gratuita del programa, característica importante ya que no existe un monto destinado a la inversión en software por parte de la organización.



Figura 61. Descarga Power BI Desktop Gratuito
Fuente: (Power BI, 2019)

Una vez instalado el programa, como primer paso se subió la base de datos del RapidMiner, ya que en esta herramienta se realizaron cálculos (porcentaje de consumo de partidas presupuestarias).

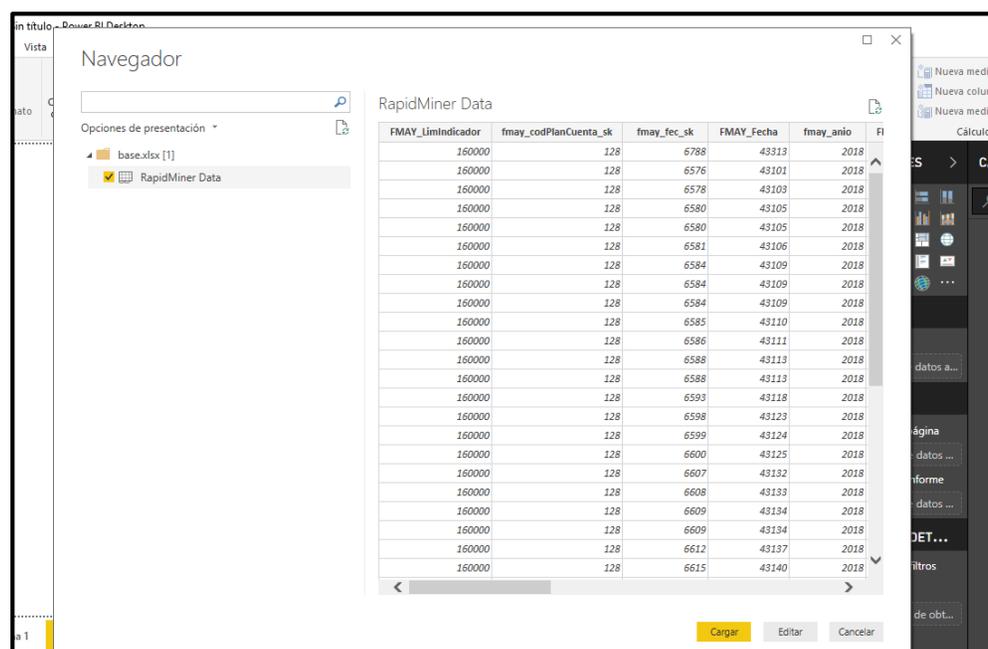


Figura 62. Carga de datos en Power BI

F MAY_LimIndicador	fmay_codPlanCuenta_sk	fmay_fec_sk	F MAY_Fecha	fmay_anio	F MAY_Nombre	F MAY_Detalle	F MAY_Monto	fmay_cuePlanCu
333026	102	6579	43104	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	1799,2	gastos
333026	102	6579	43104	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	500	gastos
333026	102	6579	43104	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	150	gastos
333026	102	6579	43104	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	1718,28	gastos
333026	102	6584	43109	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	1400	gastos
333026	102	6584	43109	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	385	gastos
333026	102	6584	43109	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	439,23	gastos
333026	102	6584	43109	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	385	gastos
333026	102	6584	43109	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	660	gastos
333026	102	6585	43110	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	175	gastos
333026	102	6585	43110	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	240	gastos
333026	102	6585	43110	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	240	gastos
333026	102	6585	43110	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	138	gastos
333026	102	6585	43110	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	1526,49	gastos
333026	102	6586	43111	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	1125	gastos
333026	102	6587	43112	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	272	gastos
333026	102	6596	43121	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	90	gastos
333026	102	6598	43123	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	385	gastos
333026	102	6605	43130	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	170	gastos
333026	102	6606	43131	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	1883,57	gastos
333026	102	6607	43132	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	40	gastos
333026	102	6608	43133	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Juan Margueytoy	32	gastos
333026	102	6608	43133	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Vaca Amable	150	gastos
333026	102	6608	43133	2018	Mantenimiento y Reparacion muebles y equipos de oficina	Correa Maria	676	gastos

Figura 63. Datos en Power BI

Con el comportamiento de la información analizada en la herramienta RapidMiner, se establecieron los siguientes indicadores:

Asignación de montos a incurrir por cada gasto:

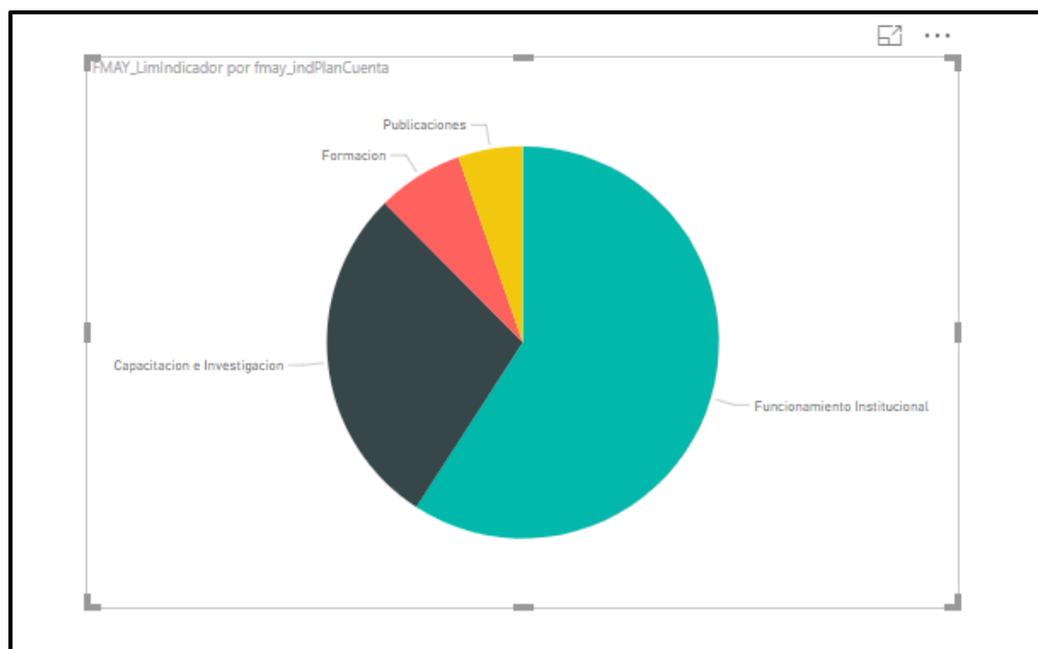


Figura 64. Asignación de monto de gastos

En la **Figura 64**. Asignación de monto de gastos se puede observar que se asigna mayores recursos a los gastos de formación institucional (Servicios básicos, arriendo, viáticos, gastos administrativos, etc), mientras que los gastos en publicaciones son los de menor monto.

Gastos reales comparativos entre los años 2018 vs. 2019:

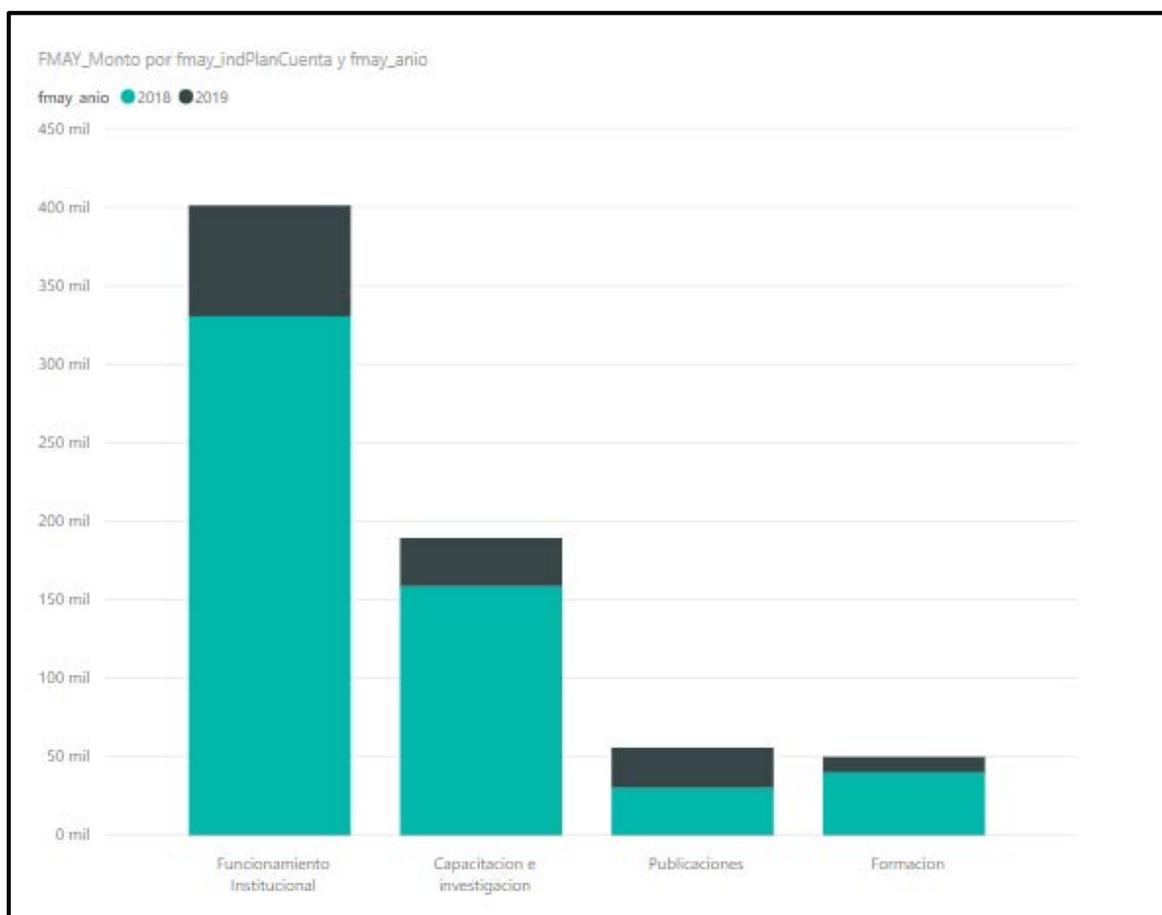


Figura 65. Comparación de gastos

En la **Figura 65**. Comparación de gastos se puede observar en el año 2018 los gastos de mayor monto son los de funcionamiento institucional seguidos de los gastos de capacitación e investigación, mientras en el año 2019 se mantienen los gastos de funcionamiento institucional como los de montos más altos sin embargo en segundo lugar tenemos los gastos por publicaciones esto se origina debido al actual proceso electoral en el que se encuentra el país por lo que el comportamiento es adecuado.

Análisis de proveedores por grupo de gasto:

- Publicaciones



Figura 66. Análisis de proveedores - Publicaciones

En la **Figura 66.** Análisis de proveedores - Publicaciones, se observa que el proveedor con el que la organización trabaja en su mayoría es Mosaico Soluciones Gráficas en el 2018 y 2019, en el 2018 el proveedor con menor contratación fue Lucas Rene y en el 2019 Colocastro & Publicidad S.A.

- Capacitación e Investigación



Figura 67. Análisis de proveedores – Capacitación e Investigación

En la **Figura 67.** Análisis de proveedores – Capacitación e Investigación, se puede observar que en el 2018 el proveedor top fue Suquillo Moreno César Xavier mientras que en el 2019 este proveedor es de los que menos contrataciones tuvo en contraste en el 2019 es Navarrete Arboleda Edmundo Daniel es el proveedor top del partido político.

- Formación

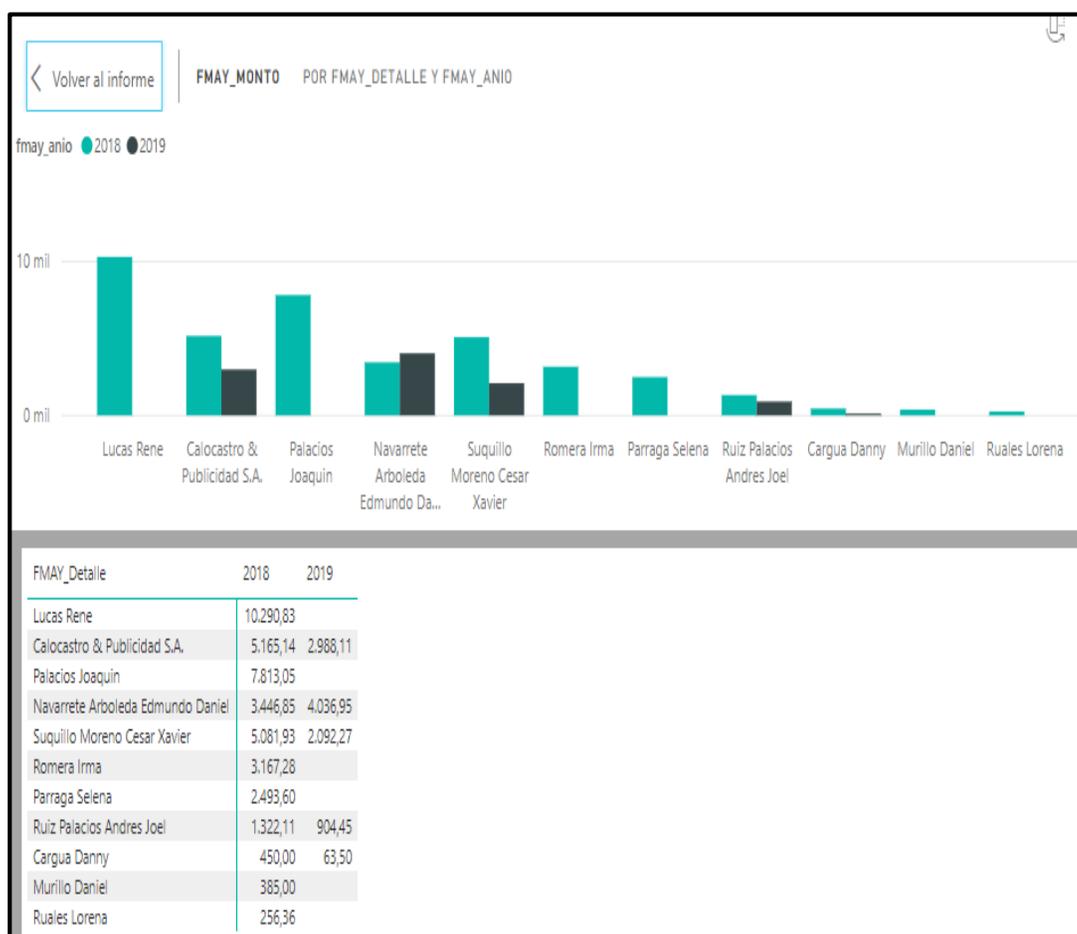


Figura 68. Análisis de proveedores – Formación

En la **Figura 68.** Análisis de proveedores – Formación, se observa que el proveedor con el monto mayor en el 2018 fue Lucas Rene mientras que en el 2019 es Navarrete Arboleda Edmundo

Daniel en contraste en el 2019 el proveedor con menor monto es Cargua Danny y en 2018 fue Ruales Lorena.

- Funcionamiento Institucional

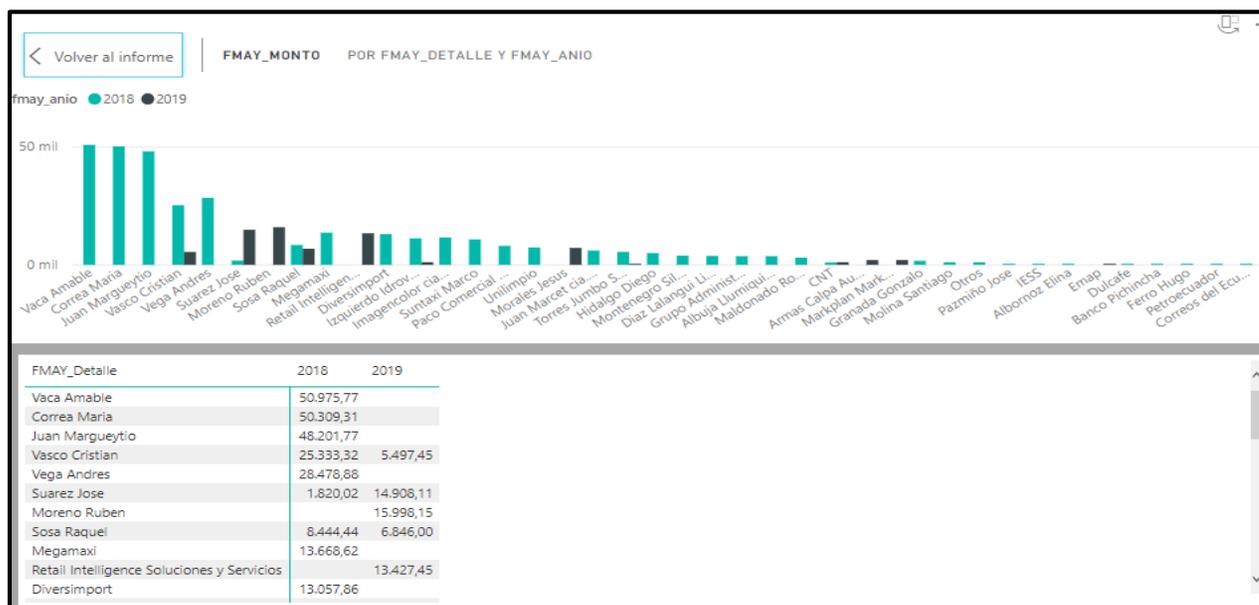


Figura 69. Análisis de proveedores – Funcionamiento Institucional

En la **Figura 69.** Análisis de proveedores – Funcionamiento Institucional se observa que en el 2018 Vaca Amable fue el proveedor top mientras que en el 2019 no se establecieron relaciones comerciales con el mismo, en el 2019 el proveedor top hasta el mes de marzo es Moreno Rubén.

Comparación entre los montos asignados y los gastados en el año 2018:

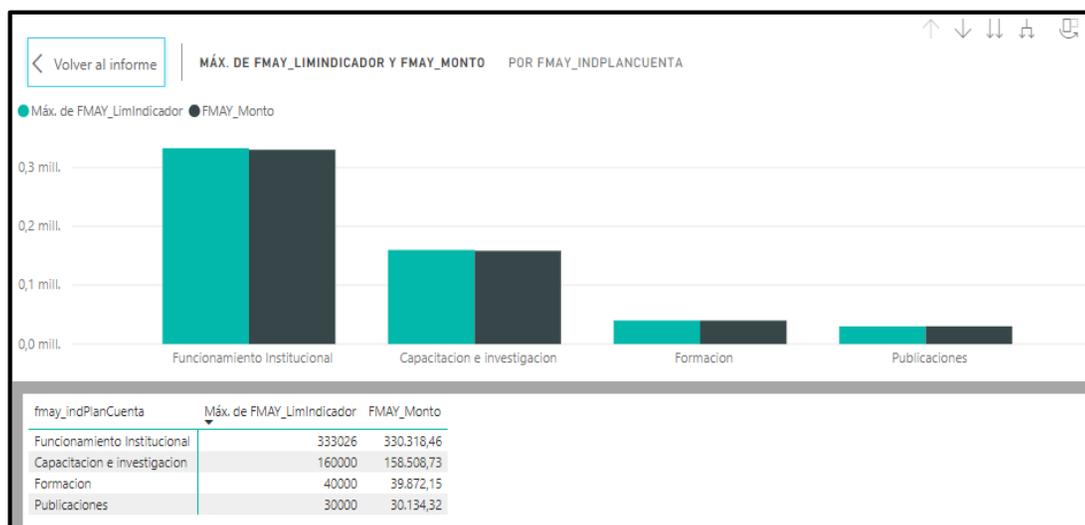


Figura 70. Comparación de gastos incurridos vs. los presupuestados

Como se puede observar en la **Figura 70.** Comparación de gastos incurridos vs. los presupuestados, las partidas presupuestarias fueron consumidas en su totalidad, el rubro con el mayor saldo es el de Capacitación con US\$2.707, en contraste el rubro de Publicaciones excedió el límite asignado en US\$134.32.

A la fecha de elaboración del presenta análisis, el partido no cuenta con los límites de sus gastos en el 2019, por lo que se manejan los del 2018 para un previo análisis, por lo que se analizan los gastos hasta marzo de 2019, como se observa a continuación:



Figura 71. Comportamiento de gastos a marzo de 2019

En la **Figura 71.** Comportamiento de gastos a marzo de 2019, se observa la tendencia de los gastos, dado el proceso electoral por el que pasa el país se observa que el Gasto por Publicaciones se encuentra consumido en un 85%, mientras que las otras partidas no cuentan con desviaciones significativas, en este caso especial se resalta la importancia del análisis de datos ya que con este se puede determinar cambios o comportamientos atípicos, como es que en el primer trimestre del año una partida presupuestaria se encuentra consumida en más del 75%.

A continuación se detalla las cuentas contables con mayor cuantía:



Figura 72. Cuentas de Pasivos

En la **Figura 72.** Cuentas de Pasivos, se puede observar que las cuentas con mayor cuantía en el 2018 son las de impuestos, los pagos al SRI representan la mayor parte de los pasivos que tiene la organización, de igual forma resaltan los beneficios que se paga a los trabajadores.



Figura 73. Cuentas contables – Funcionamiento Institucional 2018

En la **Figura 73.** Cuentas contables – Funcionamiento Institucional, se observa que la cuenta con mayor cuantía en este grupo de gastos en el 2018 son los incurridos por Mantenimiento y reparación de muebles y equipos de oficina, en contraste, los servicios de correos son los de menor monto.

En el año 2019, de acuerdo a la **Figura 74.** Cuentas contables – Funcionamiento Institucional 2019, los gastos con mayor cuantía son los de Mantenimiento y reparación de vehículos seguido

de Mantenimiento y reparación de edificios, en contraste los gastos de agua potable son los menores.



Figura 74. Cuentas contables – Funcionamiento Institucional 2019

La herramienta permite realizar diversos análisis a continuación se observa a los proveedores que prestan sus servicios en diversas actividades económicas:

Como se observa en la **Figura 75**. Proveedores por grupo de gasto, observamos que Colocastro & Publicidad S.A. presta sus servicios a la organización en las actividades económicas de Capacitación e investigación y Formación.

Con esta funcionalidad podemos determinar si hay proveedores cuyos servicios prestados acaparan en demasía las transacciones económica, esto es importante debido a que la empresa debe asegurar iguales condiciones de negociación para todos sus proveedores.



Figura 75. Proveedores por grupo de gasto

Alertas presupuestarias por grupo de gasto:

- Capacitación e Investigación

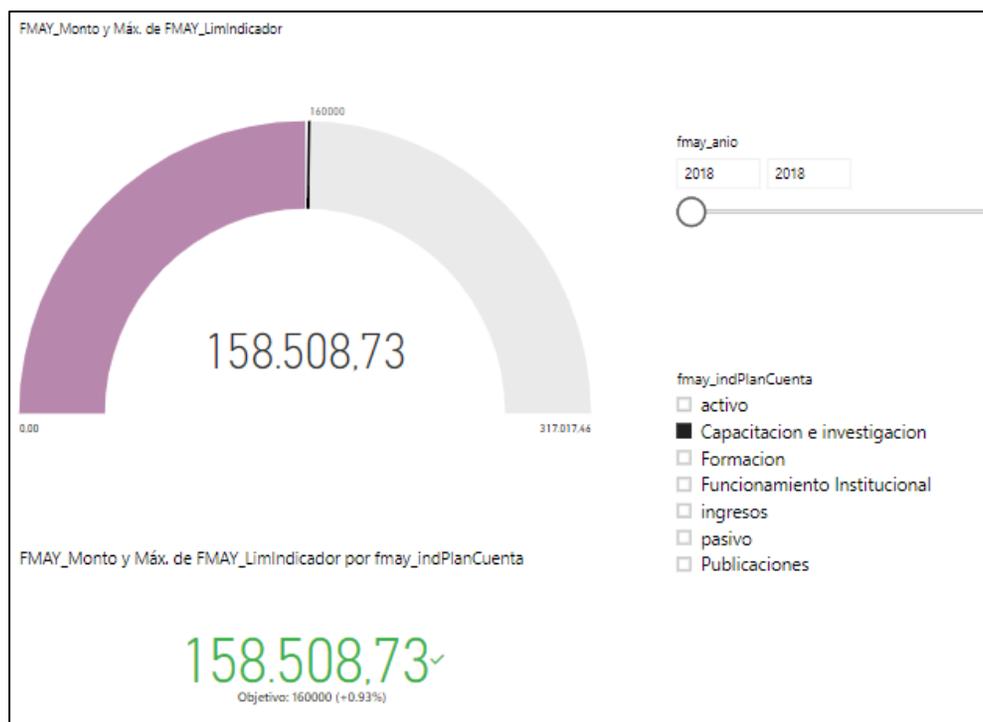


Figura 76. Alerta de consumo de partidas presupuestarias – Capacitación e Investigación

Como se puede observar en la **Figura 76.** Alerta de consumo de partidas presupuestarias – Capacitación e Investigación, en el 2019 esta partida tuvo un saldo a favor de 0,93%.

- Formación

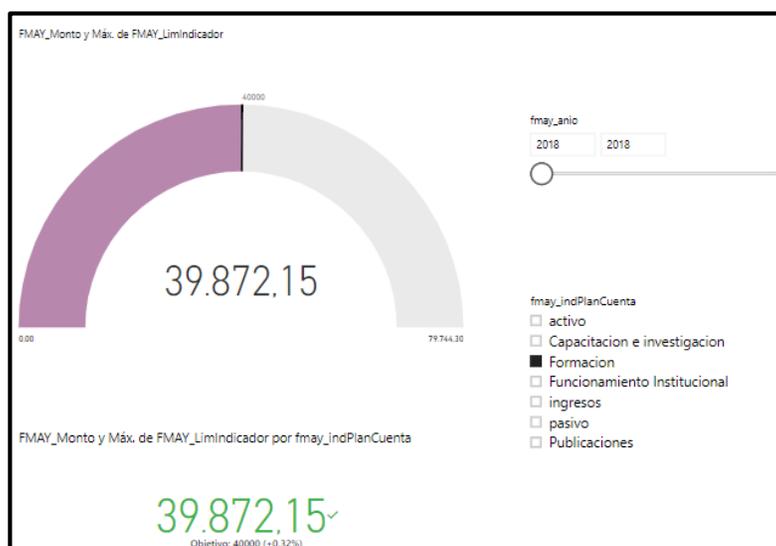


Figura 77. Alerta de consumo de partidas presupuestarias – Formación

En la **Figura 77**. Alerta de consumo de partidas presupuestarias – Formación, en el 2019 la partida formación no fue consumida en su totalidad, quedando el 0,32% de la misma, en términos monetarios US\$1.280

- Funcionamiento Institucional

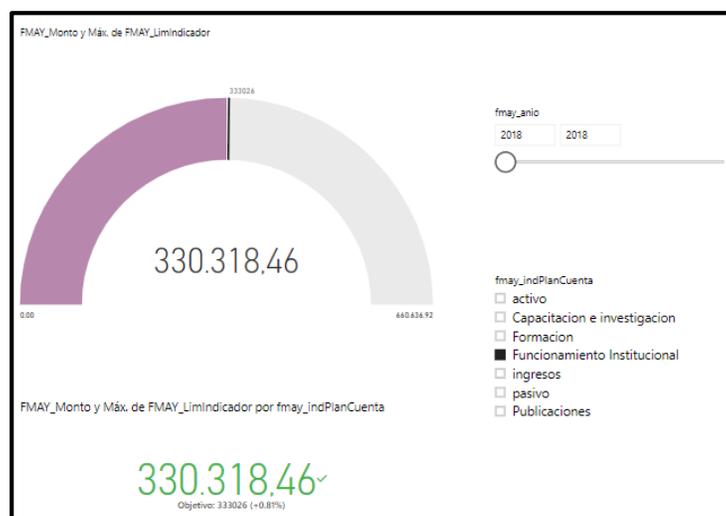


Figura 78. Alerta de consumo de partidas presupuestarias – Funcionamiento Institucional

En la **Figura 78**. Alerta de consumo de partidas presupuestarias – Funcionamiento Institucional, se observa que esta partida finalizó el año 2018 con un saldo del 0,87%.

- Publicación

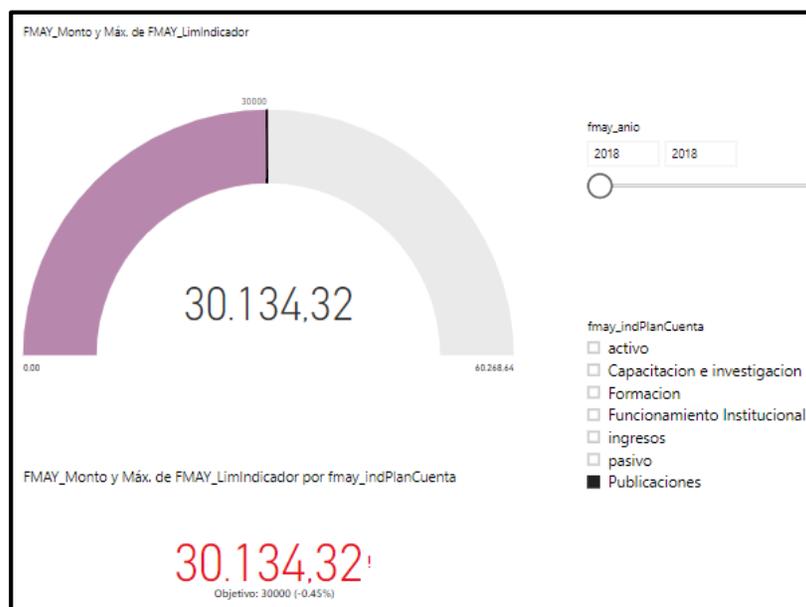


Figura 79. Alerta de consumo de partidas presupuestarias – Publicaciones

En la **Figura 79.** Alerta de consumo de partidas presupuestarias – Publicaciones, se puede observar como funcionan las alertas presupuestarias en la herramienta, ya que en la partida de Publicaciones se sobrepaso el presupuesto en 0,45%, Power BI arroja el resultado en rojo en seal de desviacion.

CAPITULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

El análisis y gestión de datos en las empresas son procesos importantes ya que están directamente relacionadas a la toma de decisiones no solo de los altos ejecutivos, sino también de mandos medios, ya que si bien es cierto los resultados son mostrados a alto nivel jerárquico, las personas que alimentan los modelos son las personas de mando operacional y mando medio.

En el Partido político Izquierda Democrática, los datos son diversos y su volumen de transaccionalidad depende de los procesos electorales en los que se encuentren inmerso el país, se recibió la planificación estratégica del año 2018 en la cual se evaluaron los grupos de gastos que la organización deber realizar, así como sus límites según la normativa; de igual forma la administración otorgo acceso a los datos contables de los periodos 2018 y 2019 hasta el mes de marzo, se realizó la integración de los mismos, diseñando un modelo multidimensional, con el que se establecieron las relaciones entre los datos de la organización. Con los antecedentes anteriormente descritos se realizó el modelo de limpieza y extracción de datos mediante la herramienta ETL.

En la selección de la herramienta ETL, se realizó un estudio comparativo que determino que la más adecuada para el diseño de la bodega de datos es Data Integration Pentaho, con esta herramienta se creó las 2 dimensiones y 1 tabla de hechos que permitieron armar el modelo multidimensional propuesto, obtenido datos óptimos y organizados; ya que esto es fundamental importancia para las predicciones y determinación de resultados.

El modelo creado a través de la herramienta RapidMiner, permitió explorar los datos, de manera automática mediante Automodel, un complemento de dicha herramienta, una vez que los

modelos fueron procesados, se seleccionó DECISION TREE, debido a la facilidad de visualización, presentando un resultado comprensible para los miembros de la organización, no únicamente para los expertos en la materia. Con este modelo se determinó la tendencia que tienen los gastos en los cuales incurre el partido y se determinó posibles desviaciones o cambios que afecten el curso de la institución.

La evaluación del modelo se realizó a través de Automodel de RapidMiner que cuenta con el operador Performance como parte de sus bondades.

Como paso final se crearon los dashboard, en la herramienta Power BI, se seleccionó la herramienta debido a la familiaridad que tiene con los usuarios ya que su manejo es similar a un archivo Excel, por lo que no es necesario incurrir en gastos de uso tecnológico y/o capacitación para las personas que manejen la herramienta. Con Power BI las organizaciones pueden conectar los datos, combinarlos y darles forma mediante la creación de informes que pueden ser compartidos con cada uno de sus miembros.

Las principales alertas que se crearon son para los grupos de gastos: Formación, Publicaciones, Capacitación e Investigación y Funcionamiento Institucional, en las cuales se determina el límite de cada una de las partidas presupuestarias, que son los montos reales incurridos de acuerdo a los registros contables, y en cuantos puntos porcentuales se ha cumplido o se ha sobrepasado, este informe es indispensable para la toma de decisiones ya que lleva un control continuo y pueden tomar decisiones oportunas.

Adicional a las alertas con la herramienta se pueden realizar diversos análisis en los cuales se determinaron los proveedores top de los grupos de gastos, proveedores con los que no se ha realizado negociaciones continuas, tipo de gastos que se realizan en mayor y menor monto, etc.

5.2. Recomendaciones

Es importante para el desarrollo de la solución el conocimiento y comprensión del negocio u organización mantener una relación cordial con los miembros de la organización para obtener la información necesaria e identificar áreas críticas.

Al iniciar el proyecto es importante definir los recursos y la información que se dispone, para determinar el alcance y desarrollo de la solución.

Las herramientas utilizadas fueron seleccionadas por sus diversas cualidades, destacando entre ellas su bajo costo, por lo que servirán como referentes para que organizaciones de similar actividad económica que permitan reducir tiempos de análisis en la obtención de indicadores, basar en este sus mejoras y tomar decisiones de acuerdo a sus necesidades.

Para el correcto análisis es importante mantener una comunicación adecuada y constante con los miembros de la organización, que permitan identificar cambios o adiciones en la información entregada en primer lugar que afecten la carga inicial y por consecuencia los resultados en los cuales se basan las decisiones.

Las bondades del análisis en las herramientas propuestas deben ser llevadas conjuntamente con la persona responsable de la parte operativa, alimentar las bases y ejecutar los análisis, ya que si las herramientas no son correctamente utilizadas no se cumplirán con el objetivo de control continuo.

BIBLIOGRAFÍA

- Abbey, M. C. (1997). *Oracle Data Warehousing*. Madrid: Mercedes Franco Calvo.
- Aguirre, P. (23 de 10 de 2012). *Instituto Politecnico Nacional*. Obtenido de <http://148.204.210.201/tesis/1359572993732PamelaGutirre.pdf>
- Alberto Montaña, J. M. (2010). *Unidad Profesional Interdisciplinaria de Ingeniería y Ciencias Sociales y Administrativas*. Obtenido de <http://tesis.ipn.mx/bitstream/handle/123456789/8403/I7.1896.pdf?sequence=1>
- Castro, F. E. (2013). Indicadores de gestión para la toma de decisiones. *Google Scholar*, 87-99.
- Cataldo, A. (Octubre de 2015). *ResearchGate*. Obtenido de https://www.researchgate.net/publication/283018388_Design_science_research_Una_breve_introduccion
- Cavazos, E. (30 de 09 de 2008). *Gravitar, Información sin límites*. Obtenido de <https://gravitar.biz/bi/metodologia-business-intelligence/>
- Colmayor. (09 de 05 de 2017). *Institución Universitaria Colegio Mayor de Antioquia* . Obtenido de [http://www.colmayor.edu.co/uploaded_files/images/archivos/normograma/manuales/Manualindicadoresversion20\(may8\)10\(1\).pdf](http://www.colmayor.edu.co/uploaded_files/images/archivos/normograma/manuales/Manualindicadoresversion20(may8)10(1).pdf)
- Conesa, J. C. (2010). *Introducción al Business Intelligence*. México: Editorial UOC.
- CulturaCrm. (26 de 12 de 2016). *CulturaCrm*. Obtenido de <http://culturacrm.com/business-intelligence/kpi-business-intelligence/>
- Fandom. (15 de 02 de 2015). *Ligia Balleares*. Obtenido de http://ligiaballeares.wikia.com/wiki/RESE%C3%91A_HISTORICA_DE_LOS_PRESU
PUESTOS

Financiera, E. (2018). *Enciclopedia Financiera*. Obtenido de

<https://www.encyclopediainanciera.com/indicadores-financieros.htm>

Guide, D. (30 de 01 de 2018). *Software de data mining: realiza análisis de datos más efectivos*.

Obtenido de <https://www.1and1.es/digitalguide/online-marketing/analisis-web/software-de-data-mining-las-mejores-herramientas/>

I.D. (2017). *Resolución PLE-CNE-3-3-7-2017*. Quito.

IBM. (14 de 10 de 2018). *IBM*. Obtenido de

https://www.ibm.com/support/knowledgecenter/es/SS3RA7_sub/modeler_mainhelp_client_ddita/components/neuralnet/neuralnet_model.html

ID. (2018). *Planificación Estratégica*. Quito.

Juan Moine, A. H. (09 de 10 de 2011). *Universidad Nacional de Buenos Aires*. Obtenido de

http://sedici.unlp.edu.ar/bitstream/handle/10915/20034/Documento_completo.pdf?sequence=1&isAllowed=y

Knime. (2018). *Knime*. Obtenido de <https://www.knime.com/knime-software>

León, E. (15 de 03 de 2015). *Universidad Nacional de Colombia*. Obtenido de

http://disi.unal.edu.co/~eleonguz/cursos/md/presentaciones/Sesion5_Metodologias.pdf

Lopez, R. (13 de 05 de 2015). *Comunidad IEBS*. Obtenido de

<https://comunidad.iebschool.com/bigdata/2015/05/13/mineria-de-datos-presentacion/>

Lucid Software Inc. (2019). *Lucid Software Inc*. Obtenido de

<https://www.lucidchart.com/pages/es/que-es-un-diagrama-de-arbol-de-decision>

Lucidchart. (15 de 12 de 2015). *Lucidchart*. Obtenido de

<https://www.lucidchart.com/pages/es/qu%C3%A9-es-un-diagrama-de-%C3%A1rbol-de-decisi%C3%B3n>

Maldonado, I. (2 de 07 de 2014). *UCV-HACER Revista de Investigación y Cultura*. Obtenido de <http://www.redalyc.org/pdf/5217/521751976006.pdf>

Meter, W. (27 de 07 de 2012). *Work Meter*. Obtenido de <https://es.workmeter.com/blog/bid/192978/principales-herramientas-de-business-intelligence>

Microsystem. (Febrero de 2019). *Información Inteligente*. Obtenido de <https://www.microsystem.cl/tableau-lidera-gartner-bi-2019/>

Monzón, L. (24 de 04 de 2018). *Arbentia*. Obtenido de <https://www.arbentia.com/comparativa-de-herramientas-de-business-intelligence-microsoft-power-bi/>

Noboa, D. (2010). *Escuela Superior Politécnica de Litoral*. Obtenido de <http://publicaciones.usm.edu.ec/index.php/GS/article/download/35/54>

Orange. (17 de 07 de 2018). *Orange*. Obtenido de <https://orange.biolab.si/#Orange-Features>

Power BI. (23 de 03 de 2019). *Microsoft*. Obtenido de <https://powerbi.microsoft.com/es-es/what-is-power-bi/>

Rapidminer. (01 de 01 de 2017). *Rapidminer*. Obtenido de <https://rapidminer.com>

RapidMiner. (01 de 01 de 2018). *RapidMiner*. Obtenido de <https://rapidminer.com/resource/gartner-magic-quadrant-data-science-platforms/>

Sinchiguano, M. (20 de 09 de 2017). *Repositorio ESPE*. Obtenido de <http://repositorio.espe.edu.ec/bitstream/21000/13529/1/T-ESPE-053888.pdf>

Sinnexus. (01 de 01 de 2018). *Business Intelligence*. Obtenido de https://www.sinnexus.com/business_intelligence/olap_vs_oltp.aspx

Sofía Bustos, V. M. (01 de 03 de 2013). *Repositorio ESPE*. Obtenido de <http://repositorio.espe.edu.ec/bitstream/21000/6305/1/T-ESPE-047033.pdf>

Sthda. (12 de 05 de 2013). *Statistical tools for high-throughput data analysis*. Obtenido de <http://www.sthda.com/english/articles/27-partitioning-clustering-essentials/>

Tello, E., & Perusquia, J. (2016). Inteligencia de negocios: estrategia para el desarrollo de competitividad en empresas de base tecnológica. *Contaduría y Administración*, 127-158.

Vijai Vaishnavi, B. K. (2015). *Google Shoolar*. Obtenido de <http://desrist.org/desrist/content/design-science-research-in-information-systems.pdf>

Weka. (01 de 01 de 2017). *The university of Waikato*. Obtenido de <https://www.cs.waikato.ac.nz/ml/weka/index.html>