METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTO PARA LA VALORACIÓN DEL RIESGO DE DESERCIÓN DE LOS APRENDICES DEL SENA

ALEJANDRO GAMEZ SERNA

NARCISO CANTILLO ALVAREZ

UNIVERSIDAD SIMÓN BOLIVAR

MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

BARRANQUILLA - COLOMBIA

2020

Trabajo de investigación

METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTO PARA LA VALORACIÓN DEL RIESGO DE DESERCIÓN DE LOS APRENDICES DEL SENA

ALEJANDRO GAMEZ SERNA

NARCISO CANTILLO ALVAREZ

Trabajo de investigación presentado como requisito parcial para optar el título de

Magister en Ingeniería de Sistemas y Computación

DIRECTOR: PhD, JUAN CARLOS CALABRIA SARMIENTO

CODIRECTOR: Msc, KARINA PAOLA JIMENEZ VEGA

UNIVERSIDAD SIMÓN BOLÍVAR

MAESTRÍA EN INGENIERÍA DE SISTEMAS Y COMPUTACIÓN

BARRANQUILLA - COLOMBIA

2020

TABLA DE CONTENIDO

LISTA DE FIGURAS……8

CAPÍTULO I

1 INTRODUCCIÓN…….....9

1.1 IDENTIFICACIÓN (DESCRIPCIÓN) Y FORMULACIÓN DEL PROBLEMA……....9

1.2 JUSTIFICACIÓN DEL PROBLEMA Y/O DE LA PROPUESTA……16

1.3 OBJETIVO GENERAL…….....18

1.3.1 OBJETIVOS ESPECÍFICOS……18

1.4 REVISIÓN DE LA LITERATURA…….....18

1.4.1 DESERCIÓN EN LAS INSTITUCIONES DE EDUCACION SUPERIOR…….....18

1.4.2 ENFOQUE PSICOLÓGICO……20

1.4.3 ENFOQUE SOCIOLÓGICO……22

1.4.4 ENFOQUE ECONÓMICO……23

1.4.5 ENFOQUE ORGANIZACIONAL……..24

1.4.6 ENFOQUE DE INTERACCIONES…….……24

1.5 EL PANORAMA EDUCATIVO COLOMBIANO……25

1.6 EL PROBLEMA DE LA DESERCIÓN ESTUDIANTIL……....26

1.7 SPADIES……28

1.8 METODOLOGIAS DE EXTRACCIÓN DE CONOCIMIENTO…….....28

1.9 MINERÍA DE DATOS…….....32

1.9.1 EVOLUCIÓN HISTORICA DE LA MINERIA DE DATOS…….....32

1.9.2 MODELO DE LAS TÉCNICAS DE MINERIA DE DATOS……..34

1.9.2.1 MODELO DESCRIPTIVO……34

1.9.2.2 MODELO PREDICTIVO…….....35

1.9.2.3 TAREAS DE MINERIA DE DATOS……....36

1.9.3 TIPOS DE TAREA EN LA MINERIA DE DATOS……36

1.9.3.1 RESUMEN……...36

1.9.3.2 CLASIFICACIÓN……..…..36

1.9.3.3 AGRUPACIÓN……..36

1.9.3.4 ANALISIS DE LAS TENDENCIAS……..…37

1.9.4 LA MINERIA DE DATOS EN LA EDUCACIÓN……37

1.9.5 LA APLICACIÓN DE LA MINERIA DE DATOS A LA DESERCION ESTUDIANTIL…….39

2. METODOLOGIAS DE EXTRACCION DE CONOCIMIENTO…….40

2.1 FASES METODOLOGÍA CRISP-DM …….....40

2.1.1 PRIMERA FASE ……...41

2.1.2 SEGUNDA FASE ……..…41

2.1.3 TERCERA FACE ……..41

2.1.4 CUARTA FASE ……41

2.1.5 QUINTA FASE……42

2.1.6 SEXTA FASE ……42

2.2 METODOLOGÍA SEMMA ……..42

2.2.1 ETAPAS METODOLOGÍA SEMMA……...42

2.2.2 PRIMERA ETAPA……..42

2.2.3 SEGUNDA ETAPA …….....43

2.2.4 TERCERA ETAPA ……...43

2.2.5 CUARTA ETAPA …….…43

2.2.6 QUINTA ETAPA …….....43

2.2.7 GRÁFICO METODOLOGÍA SEMMA …….....44

2.3 COMPARACIÓN DE LAS METODOLOGÍAS CRISP-DM Y SEMMA ……..…….44

2.3.1 GRÁFICO COMPARATIVA DE LA METODOLOGÍA CRISP-DM Y SEMMA…….45

2.4 LA MEDICIÓN DE LA DESERCIÓN SENA VS SPADIES…...……...…….45

2.5 EL SERVICIO NACIONAL DE APRENDIZAJE SENA Y SUS CARACTERISTICAS…….....46

2.6 METODOLOGÍA…….....47

2.6.1 TIPO DE INVESTIGACIÓN……..47

2.6.2 MÉTODOS DE INVESTIGACIÓN……...47

2.6.3 TÉCNICAS PARA LA RECOPILACIÓN DE INFORMACIÓN……48

CAPITULO II

3. APLICACIÓN METODOLOGIA DE EXTRACCION DE CONOCIMIENTO CRISP-DM……..48

3.1 COMPRENSIÓN DEL NEGOCIO……..……48

3.2 DETERMINAR LOS OBJETIVOS DEL NEGOCIO……..……...49

3.3 EVALUACIÓN DE LA SITUACION……..50

3.3.1 LISTADO DE RECURSOS…….……..50

3.3.2 RESTRICCIONES……51

3.3.3 RIESGOS Y CONTINGENCIAS:……...……..51

3.3.4 BENEFICIOS:……51

3.4 DETERMINACION DE LOS OBJETIVOS DE LA MINERIA DE DATOS…….……..51

3.5 PRODUCIR EL PLAN DEL PROYECTO……..52

3.6.COMPRESIÓN DE DATOS ……52

3.6.1 RECOLECCIÓN DE DATOS INICIALES……...52

3.6.2 DESCRIPCIÓN DE LOS DATOS……...53

3.6.3 VERIFICACIÓN DE LOS DATOS……...53

3.6.4 PREPARACIÓN DE LOS DATOS……53

3.7 MODELADO……..54

3.7.1 SELECCIÓN DE LA TÉCNICA DE MODELADO……54

3.7.2 EVALUACIÓN …….....56

4. APLICACIÓN METODOLOGIA DE EXTRACCION DE CONOCIMIENTO SEMMA……57

4.1 MUESTREO DE LOS DATOS……...……57

4.2 EXPLORACIÓN DE LOS DATOS……..58

4.3 MODIFICACIÓN DE LOS DATOS……58

4.4 MODELADO Y VALORACIÓN……..…...59

CAPITULO III

5. METODOLOGÍA DE EXTRACCION DE CONOCIMIENTO PARA LA VALORACION DEL RIESGO DE DESERCION DE LOS APRENDICES DE EL SENA…….…….....59

5.1 ANALISIS DE DATOS UTLIZANDO HERRAMIENTAS DE INTELIGENCIA DE NEGOCIO……..63

5.2 EVALUACIÓN……69

5.3 ORIENTACIÓN DE TOMA DE DECISIONES……..70

6. CONCLUSIONES……73

7. PERSPECTIVAS FUTURAS……74

8. BIBLIOGRAFÍA…….....75

9. ANEXOS

ANEXO 1.....80

ANEXO 2……80

ANEXO 3……81

ANEXO 4……81

ANEXO 5……82

ANEXO 6……..82

ANEXO 7……...83

ANEXO 8……83

ANEXO 9……...84

ANEXO 10……84

ANEXO 11……85

ANEXO 12……85

ANEXO 13……86

ANEXO 14……86

ANEXO 15……87

ANEXO 16……..87

ANEXO 17……...88

ANEXO 18……..88

ANEXO 19……..89

ANEXO 20……89

ANEXO 21……90

ANEXO 22……90

ANEXO 23……91

ANEXO 24……..91

ANEXO 25……...92

ANEXO 26……...92

ANEXO 27……..93
ANEXO 28……...93
ANEXO 29……....94
ANEXO 30…….....94
ANEXO 31……...95
ANEXO 32 ……..95
ANEXO 33……...96
ANEXO 34……..96
ANEXO 35……97
ANEXO 36 …….....97
ANEXO 37…….....98
ANEXO 38…….....98

LISTA DE FIGURAS

Figura 1**:** Esquema de la teoría de la acción razonado……....21

Figura 2**:** El modelo de Ethington. (1990) …….....23

Figura 3**:** Modelo de Spady (1970) …….....24

Figura 4**:** Etapas del proceso KDD ……..33

Figura 5**:** Metodología CRISP-DM (Arancibia, 2010) ……41

Figura 6**:** Metodologia SEMMA……....44

Figura 7**:** Metodología CRISP-DM Y SEMMA……...45

Figura 8: Arbol de decisiones……...55

Figura 9: Arbol de decisiones (2)……..……..56

Figura 10: Arbol de decisiones (3)……...……..……..56

Figura 11: Regla de asociaciones……57

Figura 12: Fase Comprensión del negocio metodología MECDE……..60

Figura 13: Fase Comprensión de los datos metodología MECDE……..61

Figura 14: Fase Preparación de los datos metodología MECDE……62

Figura 15: Fase Modelado metodología MECDE……..……63

Figura 16: Metodología para la valoración del riesgo de deserción de los aprendices del Sena…….……66

Figura 17: Patrones de deserción estudiantil SENA……..67

Figura 18: Deserción estudiantil SENA……67

Figura 19: Deserción estudiantil SENA……68

Figura 20: Deserción estudiantil por programa técnico SENA …….....69

Figura 21: Deserción estudiantil por programa técnico SENA……69

Figura 22: Deserción estudiantil por programa técnico SENA……70

CAPÍTULO I

1. INTRODUCCIÓN

1.1 IDENTIFICACIÓN (DESCRIPCIÓN) Y FORMULACIÓN DEL PROBLEMA.

En los procesos de formación, las instituciones de educación superior estructuran programas para acoger a los estudiantes e iniciar con ellos un proceso de adaptabilidad y estabilidad durante toda su carrera, es así como cada institución educativa establece un sistema específico para atender, hacer seguimiento y garantizar la permanencia de los estudiantes en todo el proceso desde su admisión hasta su graduación.

Pero existen factores y variables que en el proceso de formación pueden afectar el desempeño y la aprobación exitosa de todos los ciclos de formación de los estudianteshasta llegar, incluso, a desistir de la profesión. En este sentido la deserción estudiantil ha sido foco de atención de muchas entidades de educación superior y el ministerio de educación , pues impacta a las instituciones de forma negativa.

En razón de lo anterior, se puede señalar como una primera causa de éstas dificultades el aspecto académico acompañado del bagaje cultural y académico con el que entran los estudiantes a la educación superior. Adicionalmente, las variables: socioeconómicas, politicas institucionales, afrontamiento de nuevas normas, relaciones con los profesores y con los otros estudiantes, la particularidad de los programas, programas de acompañamiento, todas afectan la estructura de personalidad de los jóvenes (Barragán y Patiño, 2013). Es así como los estudiantes estarían ante una situación que podría ser estresante, desmotivante y en efecto estimularía una toma de decisiones en relación con su permanencia académica en la institución de educación superior.

Siguiendo con este mismo factor que en detalle es lo socioeconómico, es decir, pobreza, influye dejando en una condición de vulnerabilidad a los estudiantes, por eso Sánchez, Quirós, Reverón y Rodríguez (2002) afirman que existe la necesidad de comprender en profundidad e inspeccionar las desigualdades sociales y educativas de

los estudiantes que se gradúan de las instituciones de educación media, y centrar la atención en la condición de vida de las familias, pues es una situación que repercute sobre la posibilidad de competir por el acceso a las entidades de educación superior y que, posteriormente va a incidir sobre el rendimiento académico y el posible riesgo de abandono del sistema educativo.

Y es que el problema de la deserción al ser vista en su mayor amplitud, y desde muchas investigaciones, dejan al descubierto varios aspectos, sobre todo en una mirada desde la desigualdad social; las bajas oportunidades de acceso educativo, las bajas cifras de graduación, en donde los grupos sociales menos favorecidos reflejan situaciones alejadas de la igualdad de derechos y oportunidades (Tapasco, Ruiz, Osorio y Ramírez, 2019). Es decir, que son éstas clases sociales las que siempre reflejan mayor participacion en la desercion escolar, dado los factores socioeconomicos que le rodea.

Pero el trabajo de Blanco (2014) hace fuerza en que la política del estado y el mercado, han regalado terreno a la democratización de la educación y en esto defiende que el Estado intenta mejorar estas circunstancias mediante el crédito educativo, puesto en marcha por el Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior. Del mismo modo Melguizo, Sánchez y Jaime (2011), acompaña esta misma idea al afirmar que los prestamos financieros, en cierta forma, ayudan a reducir los indices de deserción, mas que todo en estudiantes que comienzan una carrera universitaria. Esta situacion, para muchos, mejora su estabilidad y permanencia, pero no todos son elegibles para tales prestamos educativos.

Otros autores siguen como Gómez y Celis (2009), aseguran que los créditos de educación terminan endeudando a los estudiantes de condiciones más desfavorables y, a su vez, las acciones afirmativas que realizan las universidades solo se encaminan a asuntos económicos, desconociendo, en ambas situaciones, la historia formativa de los

estudiantes, la cual no les permite tener las competencias necesarias para un rendimiento adecuado y mejorar las tasas de permanencia y graduación.

Por otro lado, la deserción como fenómeno, desde el aporte teórico de Henríquez y Escobar (2016) se desglosa en tres perspectivas: una que aborda las dificultades en la pertinencia con los procesos asociados a la enseñanza y el aprendizaje, tales como su calidad, la estructura de las mallas curriculares y la formación de los docentes encargados de los procesos de formación. La segunda se enfoca en procesos externos asociados a lo educativo como los estilos de aprendizaje, financiamiento que se ofrecen a quienes ingresan al sistema. Por último, la que esta relacionada con las interacciones de múltiples factores provenientes del sistema educativo, la estructura social y su economía, que permiten comprender los cambios que han tenido éstas organizaciones en la actualidad.

Otro aspecto a detallar, es el proceso cognitivo que haya desarrollado el estudiante desde su educacion básica en el que el fenómeno de la deserción se da con mayor probabilidad, es decir, la transición que realizan los estudiantes desde el nivel secundario al nivel universitario y muchas veces se manifiestan serios problemas de adaptación, y con ello el reconocimiento de una primera dificultad que incide en la decision de abandono, ya que el cambio de un ambiente conocido a un espacio aparentemente impersonal junto con las debilidades básicas que son percibidas en los primeros semestres dejan en desventaja para asumir un buen desarrollo en sus aprendizajes.

Pero en realidad siendo empáticos, el solo hecho de ser admitido muchas veces genera todo tipo de expectativas y al enfrentarse después a las condiciones de vida estudiantil que, al no ser satisfechas, pueden conducir a la decepción inicial que favorece la decisión de abandono, ya sea temporal o permanente. Adicionalmente, pocas instituciones educativas hacen un efectivo control para brindar las herramientas apropiadas para superar las deficiencias académicas y esto a la postre abre las puertas de aumentar la dejación de la carrera (Tinto, 1989).

Otro punto de análisis sobre esta problemática realizada por Vázquez y Rodríguez (2007) determinan ciertas situaciones como determinantes para la deserción:

-La integración social y compromiso institucional e individual, relacionada con el papel que desempeña el estudiante en su lugar de estudios, donde el mismo puede sentirse comprometido en mayor o menor grado, dependiendo de la calidad de su aporte de trabajo y la satisfacción que reciba en función de su desempeño.

- La capacidad intelectual, compromiso académico e identificación profesional, considerada desde el proceso educativo, la misma se vincula con el desarrollo de la inteligencia, con la capacidad de adaptación y la solución de problemas frente a los desafíos académicos.

- El factor socio-económico, educativos y demográficos, los cuales agrupan una gran cantidad de elementos complejos tales como: nivel de estudios ambicionado; aspectos socioculturales y políticos del medio; la influencia familiar, la edad que determina un grado de madurez frente a lo que se está haciendo, entre otros.

Por las razones anteriormente expuestas, La Madriz (2016) reflexiona que, al momento de analizar el rol actual de las instituciones educativas entorno al tema en estudio, es importante que éstas autoevaluen sus estructuras, métodos de enseñanza – aprendizaje y la gestión de los conocimientos, para encontrar un modelo que responda a los nuevos requerimientos con una prospectiva de las renovaciones por hacer. Es desde estos parámetros que el tema de la permanencia y disminucion de la desercion debe estar a la mesa para encontrar estrategias efectivas que consoliden la formacion de los educandos y sus proyectos de vida.

En un estudio realizado (Reyes, Castañeda y Pabón 2012) con estudiantes que desertaron durante los años 2000, 2001, 2002 y 2003, se analizaron las causas económicas, personales y académicas, y como resultado el factor económico fue el más predominante, sin embargo, se menciona la importancia de las dificultades en el aprendizaje que los estudiantes presentan para estudiar y comprender textos, pues al mismo tiempo esto se convierte en una desventaja para su rendimiento académico.

De acuerdo a los resultados del Modelo de indicadores de desempeño de educación realizado por la MINISTERIO DE EDUCACION NACIONAL (2017), en materia de calidad, se evidencio un aumento en la cobertura en la educación superior del 31.2% al 51.2%. Sin embargo, estos logros en la cobertura implican nuevos desafíos para las IES ya que se debe garantizar la sostenibilidad de recursos para ciencia y tecnología de manera que se puedan cerrar las brechas sociales que existen en el país, teniendo en cuenta que para el año 2015, la tasa de deserción en programas universitarios fue del 9.3% mientras que en los programas técnicos y tecnológicos llegó al 32.6% y 18.0% respectivamente.

Pero la preocupación sobre el tema es compartida por los sistemas educativos de diferentes países del mundo. Para medir la dimensión del problema existen dos indicadores principales que evidencian la magnitud del problema: las tasas de deserción anual y por cohorte. Al analizar el porcentaje de estudiantes que estaban matriculados un año antes y que figuran como desertores un año después, en Colombia muestra un registro de una tasa de 10.4% al cierre de 2013, cifra superior a la del Reino Unido, 8.6% e inferior a la de Brasil (18%) y Estados Unidos (18.3%).

Por su parte, la tasa de deserción por cohorte muestra la magnitud del tema de no culminación de estudios en tanto ilustra la cantidad de estudiantes que desertan de cada 100 que ingresan a algún programa universitario. Colombia alcanza una cifra de 44.9%, lo que la deja en una posición intermedia entre México con 42%, Argentina con 43%, Venezuela con un 52% y Chile con 54%.

Un estudio realizado por el SPADIES (Herramienta para hacer seguimiento sobre las cifras de deserción de estudiantes de la educación superior) en el año 2017, permitió obtener los siguientes datos de deserción académica en las instituciones de educación superior en Colombia.

Las cifras obtenidas muestran un mayor porcentaje de deserción por periodo en los programas Técnicos con un 26,1%, seguido de los tecnológicos en un nivel de 16,7 % y

en la puntuación más baja las carreras universitarias con un 9 % de deserción (Ministerio de educación nacional, SPADIES).

El Servicio Nacional de Aprendizaje - SENA, foco de atención para el presente trabajo de investigacion, es un establecimiento público del orden nacional, con personería jurídica, patrimonio propio e independiente, y autonomía administrativa; Adscrito al Ministerio del Trabajo de Colombia. Ofrece formación gratuita a millones de colombianos que se benefician con programas técnicos, tecnológicos y complementarios que, enfocados en el desarrollo económico, tecnológico y social del país, entran a engrosar las actividades productivas de las empresas y de la industria, para obtener mejor competitividad y producción con los mercados globalizados.

La Institución está facultada por el Estado para la inversión en infraestructura necesaria para mejorar el desarrollo social y técnico de los trabajadores en las diferentes regiones, a través de formación profesional integral que logra incorporarse con las metas del Gobierno Nacional, mediante el cubrimiento de las necesidades específicas de recurso humano en las empresas, a través de la vinculación al mercado laboral -bien sea como empleado o subempleado-, con grandes oportunidades para el desarrollo empresarial, comunitario y tecnológico (Servicio Nacional de aprendizaje, 2020).

Esta entidad funciona en permanente alianza entre Gobierno, empresarios y trabajadores, desde su creación hace 60 años, con el firme propósito de lograr la competitividad de Colombia a través del incremento de la productividad en las empresas y regiones, sin dejar de lado la inclusión social, en articulación con la política nacional: Más empleo y menos pobreza. Por tal razón, se generan continuamente programas y proyectos de responsabilidad social, empresarial, formación, innovación, internacionalización y transferencia de conocimientos y tecnologías.

El Sena Centro Industrial y de Aviación, ubicado en la ciudad de barranquilla, es una de las treinta y tres sedes que tiene el Sena a nivel nacional. Los programas a nivel técnico y tecnólogo ofrecidos por el Sena Industrial y de Aviación son:

A. Artes gráficas

B. Automotriz
C. Aviación
D. Calzado y marroquinería
E. Confecciones
F. Construcción
G. Gestión
H. Madera
I. Salud Ocupacional
J. Sistemas

Un estudio realizado por el Sena (oficina de bienestar al aprendiz), en el año 2018 muestra unas cifras de deserción en el Centro Industrial y de Aviación así: el mayor número de deserciones en el año 2018 pertenecen al programa de Tecnólogo en Construcción con 55 aprendices desertados, seguido de instalaciones eléctricas con 28 aprendices, seguridad ocupacional con 26 aprendices, técnico en control de calidad con 22 aprendices y los demás programas por debajo de 20 aprendices que abandonan los estudios. Adicionalmente, se destacan con 1 aprendiz los programas de técnico en operación de maquinaria, tecnólogo en análisis y desarrollo y tecnólogo en desarrollo gráfico.

Las posibles causas de deserción según los estudios realizados por el Sena Centro Industrial y de Aviación son las siguientes:

A. Razones laborales
B. Situación económica.
C. Situación Residencial (Inseguridad en los barrios, residen en zona rural…)
D. Situación familiar

E. Dificultades de salud (Embarazo / Maternidad)

F. Desorientación vocacional (Desconocimiento del Programa de Formación, Falta de interés en la especialidad).

G. Falta de oportunidad para realizar las prácticas.

H. Fallas en el Proceso de Ingreso, Selección y Matrícula.

I. Desconocimiento de las normas Institucionales Sena.

J. La información de la Oferta web, podría estar incompleta (en una oferta, se podría indicar, por ejemplo: “una vez ingrese al Programa de formación, No se aceptan cambios de programa” …)

K. Dificultades académicas.

L. Dificultades personales.

De acuerdo con la situación expresada se enuncia el interrogante que direcciona el proceso de investigación:

¿Qué metodología de extracción de conocimientos sería pertinente para el contexto educativo a fin de prevenir la deserción de los aprendices del Sena?

1.2 JUSTIFICACIÓN DEL PROBLEMA Y/O DE LA PROPUESTA.

El presente trabajo es una apuesta a buscar caminos para incentivar la convergencia de diversas alternativas en torno a la problemática de la deserción estudiantil que en la actualidad tiene tazas altas en las instituciones de educación superior. Por lo tanto, la propuesta que se construye tiene como propósito el desarrollo de una metodologia de extracción de conocimientos para valorar el riesgo de deserción de los aprendices del Sena. Todos estos datos obtenidos desde el mismo momento de la admisión. Como es bien sabido en las metodologias, de acuerdo a sus características permiten predecir o describir datos que permitirían evaluar e interpretar esos datos para generar nuevos conocimientos y en si tomar decisiones (Hasperue,2013).

De forma específica para el Sena, escenario de formulación del estudio, este proceso puede traer muchos beneficios dado que aparte del sistema que ya tiene, se suma una metodología con una finalidad más enfocada en la deserción que va a segmentar factores con mayor claridad y, se agrega a esto, la sugerencia de permanencia que arroja el mismo proceso. En este sentido las dependencias especializadas para el acompañamiento de los aprendices tendrán un insumo importante para tomar decisiones en favor de la estabilidad y permanencia de ellos.

Esta metodología podría servir de instrumento para las instituciones de educación superior e incluso de educación básica para conocer de primer mano aquellos factores que, muy probablemente, puedan afectar los procesos de permanencia y de esta forma, tomar acciones para mejorar en mayor medida lo relacionado con los desarrollos formativos. Ésto porque las instituciones de educación superior se ven en la necesidad de contar con este tipo de mecanismos que ayuden a disminuir la deserción en las diferentes carreras que se ofrecen (Amaya, Barrientos y Heredia, 2015).

En general con lo que se propone, se complementan las metodologías que en buena medida habían contribuido a obtener una visión o un panorama sobre la deserción en ciertas instituciones, pero con éstos nuevos componentes la mirada hacia quienes desertan y sus factores específicos se hace mucho más nítida en el plano de las condiciones socioeconómicas, contexto estudiantil y de personalidad.

Finalmente, desde lo teórico los resultados de este trabajo sentarían bases para atender a la población en riesgo de deserción con la formulación de nuevas estrategias y planes de atención que favorezcan la instrucción y replanteamiento a tanto de las políticas emanadas del Estado para menguar las desigualdades sociales como de las dependencias especializadas al interior de las instituciones de educación. Como ejemplo se supone afrontar variables como la pertinencia de los planes de estudios, ciclos educativos, la oferta educativa, subsidios, ambientes universitarios y escolares, entre otros (Sáenz, Malagón, Quintero, Vélez y Parra, 2010).

1.3 OBJETIVO GENERAL

- Proponer una metodología de extracción de conocimiento para la valoración del riesgo de deserción de los aprendices del Sena.

1.3.1 OBJETIVOS ESPECÍFICOS

- Diagnosticar el estado de los procesos académicos administrativos asociados con la deserción estudiantil en los programas de educación técnicos y tecnológicos.
- Definir relaciones, funciones y características de las diferentes categorías identificadas.
- Describir una metodología basada en procesos de extracción de conocimiento para el análisis y predicción de la deserción a través de procesos de minería.

1.4 REVISION DE LA LITERATURA

1.4.1 DESERCIÓN EN LAS INSTITUCIONES DE EDUCACIÓN SUPERIOR.

En forma inicial es importante delimitar, en su sentido más amplio, definiciones sobre deserción especificadas por diversos autores. Es así, como en una primera acepción la palabra deserción se deriva del vocablo “desertar” y etimológicamente viene del latín “desertare” que significa abandonar. De acuerdo a Rodríguez (1974), citado por Vélez y López (2004) existen cuatro tipos de deserción estudiantil: deserción o mortalidad estudiantil absoluta, en relación a retiros del estudiante por motivos académicos o de otra naturaleza; deserción o mortalidad estudiantil relativa, se evidencia la proporción de estudiantes que se retiran y el total de matriculados.

En realidad, el termino deserción estudiantil está en discusión pues tiene muchas aristas y variables que intervienen en su realización, sin embargo, de acuerdo con Castaño, Gallón, Gómez y Vásquez (2006) se puede especificar que hay un abandono voluntario, debido a causas que pueden ser de origen individual, institucional y nacional-estatal.

Otras definiciones agregadas por Diaz (2008) estiman que la deserción es como una situación a la que se enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo. Luego, es posible considerar como desertor a aquel individuo que siendo estudiante de una institución de educación superior no presenta actividad académica durante tres semestres académicos consecutivos.

Por su parte, Cabral (2005) enfatiza factores como: lo socioeconómico y cultural y las horas de dedicación al estudio. También son relevantes los problemas motivacionales, personales y psicoafectivos que pueden repercutir en el fracaso académico, tales como los sentimientos de frustración, desorientación vocacional, baja autoestima, la posible y poca adaptación a los entornos universitarios.

En esta misma línea de definiciones Giovagnoli (2002) expresa que la deserción se puede definir como una situación afrontada por el estudiante por no culminar o aprobar su proyecto educativo, en este sentido se convierte en desertor, es decir, aquel individuo que pese a estar vinculado a una institución educativa no tiene ninguna actividad académica durante los dos semestres académicos siguientes.

Otro autor como Anuies (2007) establece que la deserción estudiantil se debe a multiples factores que afectan a los estudiantes. Entre éstos se encuentran:

A. Las malas condiciones económicas de los estudiantes.

B. El bajo nivel cultural de la familia a la que pertenece el estudiante.

C. Las expectativas que tiene el estudiante con respecto a la importancia de la educación.

D. No poder dedicar tiempo al trabajo y al estudio al mismo tiempo.

E. Las características personales del estudiante, por ejemplo, la falta de actitud de logro.

F. El poco interés por los carrera, estudios y la institución.

G. Las características previas del estudiante, como los bajos promedios obtenidos en la educación media que reflejan la insuficiencia de los conocimientos y las habilidades con que egresan los estudiantes, en relación con los requeridos para mantener las exigencias académicas del nivel universitario.

Pero haciendo un detenido análisis sobre los componentes para estudiar la deserción se pueden esgrimir los siguientes enfoques clasificados por Braxton et al. (1997) especifican:

A. Psicológicos
B. Sociológicos
C. Económicos
D. Organizacionales
E. Interacciones

1.4.2 ENFOQUE PSICOLOGICO

Los enfoques psicológicos de la deserción estudiantil se enfatizan en las características de personalidad que diferencian a los estudiantes que finalizan sus estudios de aquellos que no lo logran. Según Fishbein y Ajzen (1975), las intenciones de una persona son el resultado de sus creencias, las que influyen sobre sus actitudes y llevan a manifestar un comportamiento. Lo anterior implica que la personalidad es condicionada por el entorno social en el que están inmerso los individuos.

Adicionalmente, estos autores señalan que la decisión de estar en un programa está determinada por algunas conductas previas, las actitudes con respecto a la deserción y las normas subjetivas acerca de estas acciones, llevando a la conformación y generando un sistema de comportamiento, la conducta. La deserción es, entonces, el resultado de un debilitamiento de las intenciones primarias y, la persistencia, como un fortalecimiento de las mismas.

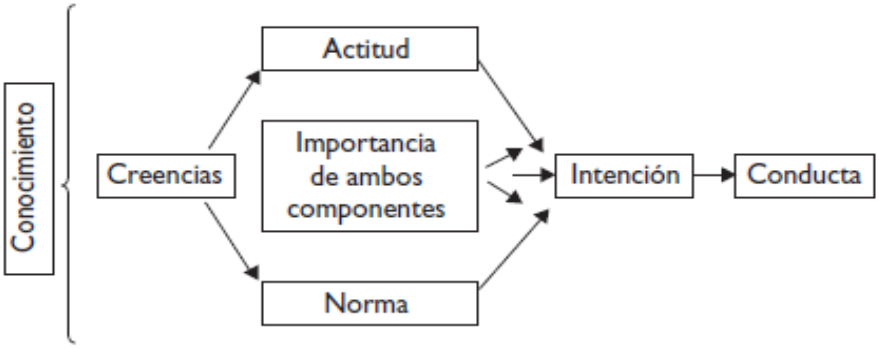


Figura 1. Esquema de la teoría de la acción razonado

Sin embargo, un estudio realizado por Ethington (1990) determina que el modelo anterior está incompleto y se apoya en una teoría más general sobre las conductas de logro. Es así como es menester señalar en primer lugar que la “conducta del logro” según Eccles, Adler, Futterman, Goff, , Kaczala, Meece, y Midgeley (1983), se compone de la perseverancia, la elección y el desempeño. Otro aporte importante de este autor es que el rendimiento académico previo influencia el desempeño futuro, al actuar sobre él al tener injerencia sobre el concepto que tenga el estudiante de sí mismo, sus metas, valores y expectativas del éxito.

En síntesis, Ethington (1990), descubrió en la práctica que el nivel de aspiraciones tenía una influencia directa sobre los valores. Además, pudo observar que las expectativas de éxito fueron explicitas por el auto concepto académico y la percepción de las dificultades de los estudios.

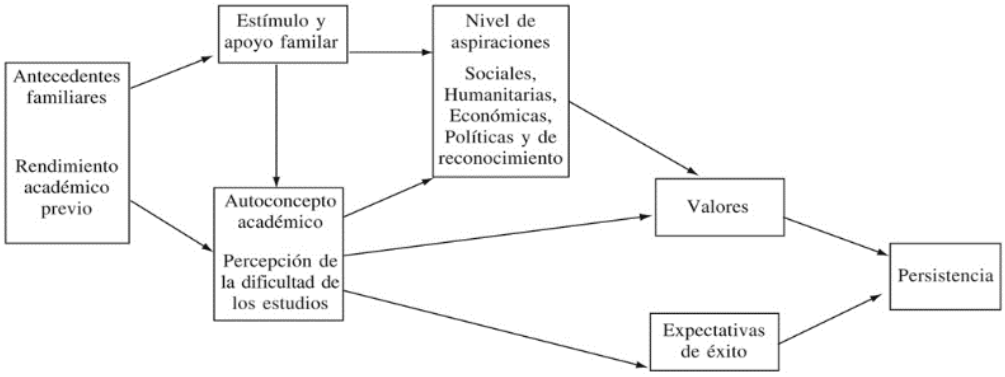


Figura 2: El modelo de Ethington. (1990)

1.4.3 ENFOQUE SOCIOLÓGICO

Los enfoques sociológicos hacen hincapié sobre los factores externos al individuo, la relación con las instituciones y el entorno social. En este ámbito de las relaciones con la deserción educativa se puede resaltar los trabajos realizados por Spady (1970) quien sostiene que existen unos elementos de la integración que son los que directamente afectan la deserción estudiantil y señala, entre otras cosas que este fenómeno es el resultado de la falta de integración y adaptación de los estudiantes a la educación superior.

Otro de los factores que también afecta a los estudiantes en situación de deserción según este autor es la familia, dado que desde ésta se afectan las expectativas y demandas teniendo repercusiones directas e indirectas en el proceso de adaptación e integración a la educación superior.

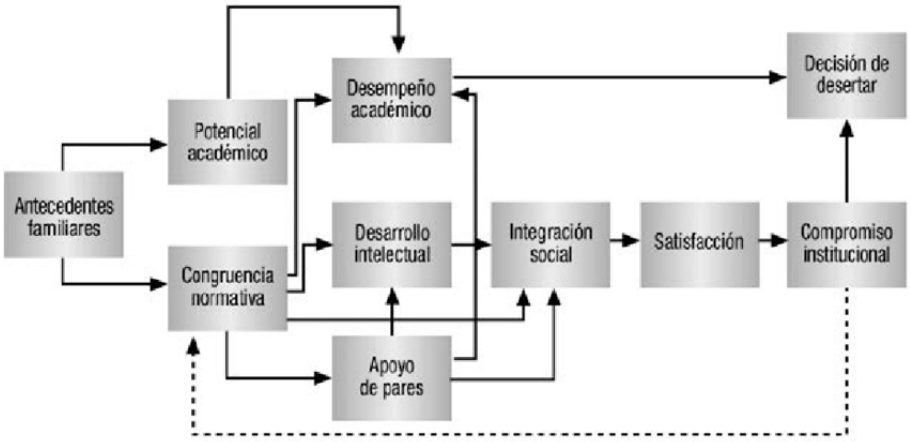


Figura 3: Modelo de Spady (1970)

Las probabilidades de que un estudiante decida abandonar los estudios dependen de que las influencias ejercidas no se lleven a cabo en la dirección positiva, es decir, que exista un rendimiento académico bajo, un bajo nivel de integración social, de satisfacción y de compromiso institucional. Por el contrario, si el estudiante logra un desarrollo académico y social acorde a sus expectativas y con las institucionales, entonces permanece en el programa y en la institución hasta lograr culminar sus estudios.

1.4.4 ENFOQUE ECONÓMICO

Otro factor de la deserción es el enfoque económico, que adopta un enfoque costo-beneficio. Cuando los beneficios sociales y económicos asociados a los estudios en la educación superior son percibidos como mayores que los derivados de actividades alternas, como por ejemplo un trabajo, el estudiante prefiere mantenerse estudiando. Un factor crítico para un estudiante es la capacidad o incapacidad de solventar los costos de su proceso educativo.

Este enfoque privilegia el impacto efectivo de los beneficios estudiantiles sobre la deserción, dejando de lado las percepciones acerca de la adecuación de dichos beneficios o el grado de ajuste de éstos a los costos de los estudios.

1.4.5 ENFOQUE ORGANIZACIONAL

El enfoque organizacional trata la deserción desde las peculiaridades de las instituciones de educación superior, en cuanto a los beneficios que ofrece a sus estudiantes. Desde esta perspectiva se retoma la forma como Tinto (1975) divulga diferentes trabajos sobre la deserción y retención basado en el modelo Spady (1970), pero al mismo tiempo vincula la teoría de intercambio de Nye (1979) la cuál se fundamenta en el principio de que todos los seres humanos evitan las conductas que implican un costo de algún tipo para ellos y buscan recompensas en las relaciones, interacciones y estados emocionales.

Otro postulado significativo de Tinto (1975), es que considera que los estudiantes actúan de acuerdo con la teoría del intercambio en la construcción de su integración social y académica. Si los beneficios de permanecer en la institución son percibidos como por encima de los costos personales, entonces estos individuos permanecerán en las entidades de educación. Pero si otras actividades son reconocidas como fuentes de mayores recompensas, el estudiante tenderá a desertar.

Pero aparte de la conducta en si misma del estudiante en el plano y contexto organizacional otros factores como la calidad de la docencia, pues las experiencias de aprendizaje de los estudiantes jugaran un papel preponderante para su permanencia y/o deserción puesto que las mejores experiencias de aprendizaje van a generar posibilidades de aumentar las expectativas y crecer en el plano intelectual y personal (Braxton, Milem, Sullivan, 2000). Para complementar fortalezas institucionales que permitan dirimir la deserción se puede señalar la disponibilidad y dotación de recursos educativos como los recursos bibliográficos, laboratorios, cantidad de estudiantes (Tillman, 2002).

1.4.6 ENFOQUE DE INTERACCIONES

A medida que el estudiante transita a través de la educación superior, múltiples variables ayudan a reforzar su adaptación a la institución que seleccionó, ya que ingresa a ella con unas características que influyen sobre su experiencia en la educación. Estas

características comprenden antecedentes familiares, nivel socio económico y cultural de la familia, los valores de la familia, atributos personales y la experiencia académica. Dichos rasgos se combinan para influir sobre el compromiso inicial con la institución, igualmente para la consecución de su meta que es la titulación.

1.5 EL PANORAMA EDUCATIVO COLOMBIANO

La educación colombiana se identifica por tener permanentemente procesos de calidad, y tiene como propósito el de fomentar y promover una cultura de mejoramiento continuo, a fin de aumentar la satisfacción de la comunidad educativa, en un ciclo que parte del mejoramiento de la calidad de cada proceso. De esta forma alcanzar la certificación, luego la acreditación de los programas y por último la acreditación institucional (Timaran et al., 2016). Por lo anterior es bueno precisar que existen líneas claras en relación con las asignaciones de tareas para establecer un trabajo de desarrollo institucional sólido que parte desde las primeras orientaciones hacia los estudiantes hasta las estrategias para lograr que los estudiantes culminen exitosamente sus estudios.

Es en el escenario anteriormente planteado donde fijar la atención sobre la permanencia y la deserción es fundamental para medir la escolaridad, pertinencia, contundencia, flexibilidad y utilidad de los programas de educación superior. Profundizando sobre la escolaridad, se descubre que Colombia viene de tener índices muy bajos, sin embargo, los esfuerzos hechos a partir los años sesenta del siglo pasado, cuyo adelanto fue lento, escaso e inequitativo.

Con mayor detalle, los avances han fluctuado de manera parcial en relación con el sube y baja de la economía. Mientras el desarrollo medio de los años sesenta y setenta fue del 2.2%, de los años ochenta a el año 2000 se contrajo al 1%. En tanto, los últimos 50 años del siglo pasado, la tasa de escolaridad de la población colombiana aumentó sólo en 4.4 años, llegando, en el umbral del siglo XXI, la tasa medio de escolaridad de la

población colombiana sólo a 7.3 años de estudio y en el sector campestre no alcanzaba los 5 años de primaria que en aquel tiempo se identificaba con la educación básica.

Por otro lado, la inserción del país a la globalización y la entrada a la sociedad de la información y el conocimiento no solo fue pertinente con un bajo nivel de desarrollo educativo en el país, no obstante, los avances, han sido continuos pero no han superado los promedios de cada región. Por ejemplo, en el año 2007, según lo estipulado por la UNESCO, mientras en la región latinoamericana la matrícula promedio a la educación básica y educació media llegaban al 95%, la tasa colombiana de escolaridad neta alcanzaba el 89,38% de acuerdo con la información suministrada por el Ministerio de Educación Nacional, y ésta tasa es más baja en las zonas rurales.

1.6 EL PROBLEMA DE LA DESERCION ESTUDIANTIL

En el caso de Colombia la situación de la deserción estudiantil tiene unas cifras muy altas y se constituye un fenómeno que está afectando los procesos universitarios de muchas instituciones de educación superior.

La deserción estudiantil, es una expresión de debacle tanto del individuo como de la institución, pues afecta la autoestima y el desempeño de los estudiantes en la sociedad, y deja entrever la falta de eficiencia e inequidades del sistema. Por otro lado, tiene un referente negativo referente a los bienes que se invierten en la formación de las personas, y sobre el progreso en todas las estructuras de la sociedad.

Las estadísticas en Colombia con respecto a la educación son preocupantes. De acuerdo con datos suministrados por el Ministerio de Educación Nacional, en el año 2004 por cada 100 estudiantes que accedieron al colegio, sólo 47 pudieron acabar el bachillerato. De éstos, sólo 27 entraron a la educación superior. Y, de los pocos que ingresaron a la educación superior, sólo un poco más de la mitad terminaron de manera exitosa sus estudios y se graduaron en cualquiera de las modalidades de educación superior.

En la estadística realizada por el Ministerio de Educacion Nacional (2017) refleja que ingresaron 2.394.434 estudiantes al medio universitario desde las carreras técnicas hasta doctorado, desertaron, a nivel nacional el 9%, cifras que han disminuido en relación con la deserción presentada desde el 2010 que inicio con 12,9%.

La deserción es un fenómeno complejo, cuyos factores deben ser analizados en su contexto y con diferentes variables. Esto se puede hacer a partir de la psicología y sociología, atravesando por la económica, incluso la institucional. En Colombia le corresponde al Ministerio de Educación Nacional colocar en marcha estrategias y mecanismos en los diversos niveles para dirimir los factores asociados con este fenómeno.

Desde las instituciones de educación superior, también se han construido estrategias y mecanismos integrales enfocados a las variables que generan la deserción. Uno de estos es los cambios en relación con la adaptación de los nuevos estudiantes a la vida universitaria, Se hacen cada vez más acciones orientadas hacia los factores académicos, se han realizado programas de orientación profesional, de tutoría, talleres de nivelación y mejoramiento de competencias básicas, foros y mecanismos de divulgación de experiencias exitosas. Adicional a ello, se han promulgado programas de crédito académico especialmente para estudiantes en riesgo de deserción, pues se busca garantizar el aspecto académico.

Pero indudablemente se han puesto en marcha herramientas de información, una de ellas es denominada SPADIES (Sistema de Prevención de la deserción en las instituciones de Educación superior), la cual funciona para identificar y hacer seguimiento a los posibles desertores. El sistema toma en consideración la que se denomina “primera deserción; es decir, a quienes por dos períodos académicos consecutivos no asisten a la institución. Sin embargo, no todas las entidades han compartido la información al SPADIES (SPADIES, 2015).

1.7 SPADIES

El Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior (SPADIES), suministra una herramienta informática que busca mantener actualizado el diagnóstico sobre la deserción reportando por las instituciones de educación superior, por medio de mediciones, seguimiento a la evolución de sus causas, al impacto y a los resultados de los programas de becas, tutorías y apoyos financieros, entre otros (SPADIES, 2015). En la actualidad el sistema está instalado en el 99% de las instituciones de educación superior y hace seguimiento a más de 3 millones de estudiantes que serán considerados desertores cuando no se encuentren matriculados en el sistema por dos o más semestres consecutivos; el sistema considera dos formas de medir la deserción:

A. La deserción por periodo, la cual hace parte de los indicadores del plan de desarrollo, que corresponde a la proporción de estudiantes que estando matriculados dos semestres atrás son clasificados como desertores un año después.

B. La otra mide la deserción por cohorte que contabiliza la deserción acumulada en cada semestre para un grupo de estudiantes que ingresaron a primer curso en un mismo periodo académico (cohorte), la cual observa para el nivel universitario al décimo semestre, mientras que para los niveles técnico profesional y tecnológico se toma a sexto semestre. (SPADIES, 2015).

1.8 METODOLOGIAS DE EXTRACCION DE CONOCIMIENTO

Como punto de partida de estas metodologías, se parte de la definición de lo que es el conocimiento y las formas como éste puede ser usado a favor de diferentes procesos sociales, económicos, políticos y científicos. es así como Anderson & Krathwohl (2001) analizan el concepto emitido por el Psicólogo Bloom, quien considera el conocimiento como la capacidad de recordar hechos específicos universales, métodos, procesos, esquemas, estructuras o marcos de referencia sin elaboración de ninguna especie,

puesto que cualquier cambio ya implica un proceso de nivel superior. En consideración estipula varios elementos que hacen parte del conocimiento:

La comprensión: capacidad de comprender o aprehender; en donde el estudiante sabe qué se le está comunicando y hace uso de los materiales o ideas que se le presentan, sin tener que relacionarlos con otros materiales o percibir la totalidad de sus implicaciones. El material requiere de un proceso de transferencia y generalización, lo que demanda una mayor capacidad de pensamiento abstracto.

Esto se demuestra cuando se presenta la información de otra forma, se transforma, se buscan relaciones, se asocia, se interpreta (explica o resume); o se presentan posibles efectos o consecuencias. También determina la importancia de crear, mejorar, optimizar o hacer algo original, guardándolo en la memoria para recordar, reconocer información específica (hechos, sucesos, fechas, nombres, símbolos, teorías, definiciones y otros). Además, los autores especifican varios elementos que hacen parte de los conocimientos como:

La aplicación: se guía por los mismos principios de la comprensión y la única diferencia perceptible es la cantidad de elementos novedosos en la tarea por realizar. Requiere el uso de abstracciones en situaciones particulares y concretas.

El análisis: consiste en descomponer un problema dado en sus partes y descubrir las relaciones existentes entre ellas. En general, la eventual solución se desprende de las relaciones que se descubren entre los elementos constituyentes. Implica el fraccionamiento de una comunicación en sus elementos constitutivos de tal modo, que aparezca claramente la jerarquía relativa de las ideas y se exprese explícitamente la relación existente entre éstas.

La síntesis: es el proceso de trabajar con fragmentos, partes, elementos, organizarlos, ordenarlos y combinarlos para formar un todo, un esquema o estructura que antes no estaba presente de manera clara.

La evaluación: se refiere a la capacidad para evaluar; se mide a través de los procesos de análisis y síntesis. Requiere expresar juicios sobre el valor de materiales y métodos,

de acuerdo con determinados propósitos. Incluye los juicios cuantitativos y cualitativos de acuerdo a los criterios que se sugieran (los cuales son asignados).

Los autores Formia y Lanzarini (2012) proponen unas fases de trabajo que da lugar a un proceso de naturaleza cíclica.

Fase 1. Comprensión del Dominio. Esta fase involucra pasos clave como determinar los objetivos, comprender la situación, determinar el papel en el proyecto y visualizar un plan de trabajo.

Fase 2. Recopilación e integración de datos. Esta fase se inicia con la obtención de los datos, se procede a familiarizarse con ellos e identificar su procedencia. En esta etapa se trabaja en recolectar los datos, describirlos, explorarlos y verificar su calidad.

Fase 3. Preparación de los datos. Es necesario seleccionar y preparar el subconjunto de datos a minar, denominado vista minable. Esta fase extiende todas las actividades para construir el conjunto final de los datos que serán utilizados en las herramientas de modelado, incluye la selección de archivos, registros y atributos, así como la transformación y limpieza de los datos.

Fase 4. Modelado. También denominada Minería de Datos, por ser la más característica del KDD, es la fase en la que se seleccionan y aplican diferentes técnicas de modelado, configurando sus parámetros para la obtención de resultados. En este es donde se produce conocimiento nuevo, construyendo modelos a partir de los datos recopilados.

Fase 5. Interpretación y evaluación. Los modelos obtenidos en la fase anterior son interpretados y evaluados con la finalidad de comprobar si cumplen los objetivos planteados en las fases preliminares. Aquí es crítico determinar si partes importantes de la realidad han sido lo suficientemente consideradas y se debe decidir sobre la utilización de los resultados del proceso.

Fase 6. Difusión y uso de los resultados. La creación del modelo no implica la finalización del proyecto. El conocimiento obtenido debe ser organizado y presentado de manera que pueda ser comprendido y utilizado por el usuario final. La tarea

importante de esta fase consiste en que el usuario comprenda los resultados y pueda utilizar los modelos creados.

Fase 7. Implementación de medidas basadas en el conocimiento obtenido. Cuando la fase de uso de resultados genera una clase de conocimiento que habilita al usuario a ejecutar acciones en pos de resolver el problema planteado originalmente, se produce una etapa de implementación de medidas que debe llevar a cabo la organización.

Fase 8. Medición de resultados. Luego de la implementación de las medidas de la fase anterior, es posible la utilización del Data Mining para medir los resultados alcanzados por esas acciones. En esta fase se pueden volver a ejecutar los modelos para compararlos con los obtenidos en la primera iteración.

En la misma perspectiva Timarán, Hernández, Caicedo, Hidalgo y Alvarado **(2016)** proponen un proceso más interactivo e integrativo e involucra numerosos pasos con la intervención del usuario en la toma de muchas decisiones. Se resume en las siguientes etapas:

- Selección.
- Preprocesamiento/limpieza.
- Transformación/reducción.
- Minería de datos (data mining).
- Interpretación/evaluación.

Además, lo ilustran en la siguiente figura:

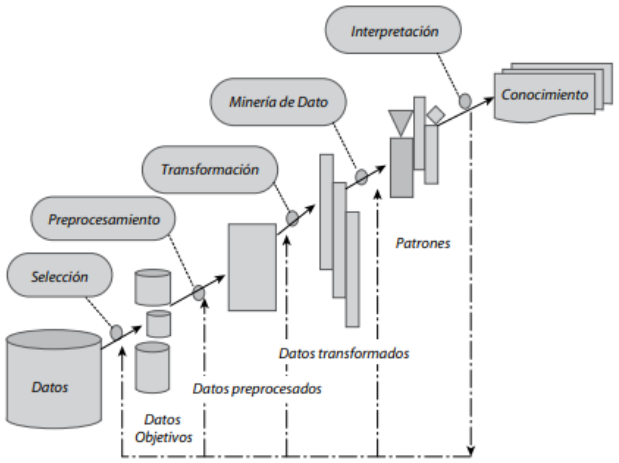


Figura 4: Etapas del proceso KDD

1.9 MINERIA DE DATOS

1.9.1 EVOLUCION HISTORICA DE LA MINERIA DE DATOS

Aunque los componentes claves de la Minería de datos (DM) empiezan desde hace varias décadas con investigaciones en áreas como la inteligencia artificial, la estadística o el aprendizaje automático, se puede aseverar que ahora estas técnicas en el manejo de datos han alcanzado mayor trayectoria. Todo esto ha permitido su uso tanto en la investigación científica como en el mundo empresarial.

Pero una revisión de cuando inicio el proceso de acuerdo a Beltrán (2016) nos llevan a los años 50, ya que la división de informática hacían los resúmenes de la información, especialmente a modo comercial, que se encontraban en el computador principal, con el propósito de posibilitar la tarea administrativa. De esta forma los sistemas de información mostraban datos, que en sí, eran demasiado grandes, poco flexibles, y difíciles de entender para los no informáticos.

Mas adelante, en los años 60, nacen los sistemas gestores de base de datos que aún se mostraban poco flexibles y carecían de maleabilidad para realizar consultas. Posteriormente aparecieron los motores relacionales resolviendo éstos problemas,

aunque los informes resultaban muy difíciles de organizar y depurar, perdiéndose relevancia por su bajo nivel de actualización.

Pero en los años 80 el Data Warehouse (DW) viene a solucionar muchos problemas informáticos. La existencia de DW ha estimulado el desarrollo de los enfoques de Data Mining, en los que las tareas de análisis se automatizan y hacen un considerable avance hacia la extracción de conocimiento inductivo.

Ahora gracias a los procesos de globalización y economía mundial y aumento de la competitividad, los datos son cruciales y se han convertido en información importante para el momento de la toma de decisiones. Por este motivo, las empresas han avanzado y han querido agregarle valor al gran volumen de información que tienen guardada en sus bases de datos. Es por ésto que se han interesado en hacer los procesos de manera automática y asi poder descubrir información importante, que de una u otra manera hubiese sido desperdiciada.

Con la velocidad muy rápida con que ha venido avanzando la tecnología, las empresas de hoy en día tienen herramientas de software y hardware cada vez más avanzadas que permiten el almacenamiento de mucha información y el análisis de ésta. El avance tecnológico y adicionalmente la aparición de mercados cada vez más competitivos, sugiere que las empresas estén en un proceso de mejoramiento continuo de sus esquemas de administración y toma de decisiones, utilizando una de las más grandes fuentes de competitividad como lo es la información.

Existen diferentes técnicas que posibilitan el aprovechamiento de los datos, descubriendo información que no se ve a simple vista. Una de estas técnicas es la denominada Minería de Datos. Ésta combina técnica semiautomáticas de inteligencia artificial, análisis estadístico, bases de datos y visualización gráfica, para la obtención de información que no esté representada explícitamente en los datos.

El propósito de la Minería de Datos es descubrir relaciones, tendencias, desviaciones, comportamientos anormales, patrones y trayectorias ocultas, a fin de soportar los procesos de toma de decisiones con mayor conocimiento. Igualmente, la mineria de

datos se puede ubicar en el nivel más alto de la evolución de los procesos tecnológicos de análisis de datos.

La Minería de Datos debe su nombre a la analogía entre una montaña y el gran volumen de datos almacenado en cualquier empresa. Dentro de la montaña, encubierto entre las piedras y la tierra, existen diamantes de gran valor que mediante procesos de minería son encontrados y explotados.

Recientemente la minería de datos ha sido utilizada en el campo educativo, y está definida como el área de la investigación científica enfocada en el desarrollo de métodos para hacer extracción de datos que provienen de los entornos, a fin de usar esos métodos para mejorar los procesos educativos.

Igualmente, la minería de datos tiene sus métodos para extraer conocimiento útil. Este conocimiento se consigue mediante el análisis de datos y recurriendo a modelos que permitan encontrar relaciones que se desconocían. El propósito de un esfuerzo de minería de datos es crear un modelo descriptivo o un modelo predictivo (Velickov y Solomatine, 2000).

1.9.2 MODELO DE LAS TECNICAS DE MINERIA DE DATOS

1.9.2.1 MODELO DESCRIPTIVO

De acuerdo a Jain y Srivastava (2013) un modelo descriptivo presenta, en forma precisa, los datos fundamentales del conjunto de datos. Específicamente es un resumen de los datos, lo que genera posibilidades de su estudio. Por lo general, un modelo descriptivo se encuentra por medio de la minería de datos no dirigida; es decir, un enfoque de abajo hacia arriba donde los datos hablan por sí mismos. La minería de datos no dirigida encuentra parámetros en el conjunto de datos, pero deja la interpretación de los parámetros al investigador.

1.9.2.2 MODELO PREDICTIVO

El objetivo final de un modelo predictivo es predecir un valor desconocido de una variable específica; La variable objetivo. Si el valor objetivo es uno de un número predefinido de etiquetas discretas (clase), la tarea de minería de datos se denomina clasificación. Si la variable de destino es un número real, la tarea es la regresión.

El modelo predictivo se crea a partir de valores de variables conocidas dadas, posiblemente incluyendo valores previos de la variable objetivo. Los datos de entrenamiento consisten en pares de medidas, cada una consiste en un vector de entrada x (i) con un valor objetivo correspondiente y (i). El modelo predictivo es una estimación de la función y = f (x; q) capaz de predecir un valor y, dado un vector de entrada de valores medidos x y un conjunto de parámetros estimados q para el modelo f. El proceso de encontrar los mejores valores q es el núcleo de la técnica de extracción de datos (Jain y Srivastava, 2013).

Según los autores anteriores algunas técnicas de extracción de datos obtienen directamente la información al realizar una partición descriptiva de los datos. Pero, es muy usual, que en las técnicas de minería de datos se utilicen datos almacenados para construir modelos predictivos.

Es fundamental considerar que las técnicas deben tener un alto rendimiento. Se entiende que este criterio, para el modelo predictivo, significa que la técnica debe producir modelos que se generalizarán bien, por ello son modelos que tienen una alta precisión al realizar predicciones basadas en datos novedosos. La clasificación y la predicción son dos perspectivas de análisis de datos que se pueden usar para extraer modelos que describen las clases de datos importantes o para predecir las tendencias futuras de los datos.

1.9.2.3 TAREAS DE MINERIA DE DATOS

Las tareas de la minería de datos son muy diversas porque existen muchos patrones en una gran base de datos. Es importante aplicar diferentes métodos y técnicas para hallar diferentes tipos de patrones.

1.9.3 TIPOS DE TAREA EN LA MINERIA DE DATOS

De acuerdo con Luo (2008) las siguientes tareas son fundamentales para hallar diferentes tipos de patrones:

1.9.3.1 RESUMEN

El resumen es la abstracción de los datos. Un conjunto de datos relevantes para la tarea se resume y se resume. Esto resulta en un menor conjunto que aporta una visión general de los datos, generalmente con información sustancial agregada.

1.9.3.2 CLASIFICACION

La clasificación tiene su origen de una función o modelo que determina la clase de un objeto en función de sus atributos. Un conjunto de objetos se da como el conjunto de entrenamiento. En éste, cada objeto está representado por un vector de atributos junto con su clase. Una función o modelo de clasificación se construye analizando la relación entre los atributos y las clases de los objetos en el conjunto de entrenamiento. Esta función o modelo puede clasificar objetos futuros. Esto nos ayuda a desarrollar una mejor comprensión de las clases de los objetos en la base de datos.

1.9.3.3 AGRUPACION

La agrupación identifica clases, también llamadas agrupaciones o grupos, para un conjunto de objetos cuyas clases son desconocidas. Los objetos se agrupan para maximizar las similitudes interclases y para minimizar las similitudes interclases. Esto se aplica con base a algunos criterios definidos en los atributos de los objetos. Una vez que se deciden los grupos, los objetos se etiquetan con sus grupos correspondientes. Las

características comunes de los objetos en un clúster se resumen para formar la descripción de la clase.

1.9.3.4 ANALISIS DE LAS TENDENCIAS

Los datos de series de tiempo son registros acumulados en el tiempo. Por ejemplo, las ventas de una empresa, las transacciones con tarjeta de crédito de un cliente y los precios de las acciones son datos de series temporales. Dichos datos pueden verse como objetos con un atributo de tiempo. Los objetos son instantáneas de entidades con valores que cambian con el tiempo. Encontrar los patrones y las regularidades en las evoluciones de datos a lo largo de la dimensión del tiempo puede ser fascinante.

1.9.4 LA MINERIA DE DATOS EN LA EDUCACIÓN

En los últimos años la minería de satos ha sido desplazada a la educación y se ha desarrollado tanto que va paralelo con otras tecnologías; es usada también para datos en la estadística, aprendizaje de máquina, y algoritmos.

En realidad, la gran utilidad de la minería de datos en el contexto educativo es de mucha ventaja para mejorar los procesos educativos por lo que obtiene información relevante que permiten hacer planes de mejora y aumentar la calidad educativa (Romero y Ventura, 2005)

Específicamente la apuesta por la minería de datos en la educación es principalmente para abordar la gestión de contenidos de aprendizaje y los sistemas educativos inteligentes basados en la web. Los autores agregan que después del preprocesamiento de los datos en cada uno de éstos sistemas se aplican las diferentes técnicas de minería de datos: estadísticas y visualización, agrupamiento y clasificación, reglas de asociación y la minería de datos.

Según Oñate (2016) la aplicación de la minería de datos en la educación es un campo de investigación reciente, el objetivo es proporcionar información a los docentes y administrativos para mejorar el aprendizaje de los estudiantes y organizar los recursos educativos de una manera más eficiente.

Los autores Araque, Roldan y Salguero (2009) realizaron un estudio sobre los factores que inciden en la deserción estudiantil, a partir de la creación de un modelo de predicción que permitiera medir el riesgo de abandono de un estudiante con la información socioeconómica y los registros académicos, a través de la técnica de árbol de decisión y regresión logística, para cuantificar los estudiantes con alto riesgo de abandonar los estudios.

Por otro lado, otra problemática entorno a la deserción estudiantil se está presentando mucho en las universidades que ofrecen educación a distancia, dado que muchos estudiantes no terminan el curso, los autores Kotsiantis, Pierrakeas y Pintelas (2004) presentan el estudio de un algoritmo de aprendizaje para la predicción de la deserción de los estudiantes que abandonan sus estudios. El trabajo de campo para aplicar este modelo se realizó con los datos proporcionados por la universidad en estudio, en este se compararon los algoritmos árbol de decisión, red neuronal, Naive Bayes, regresión logística y máquinas de vectores.

Los autores Kuna, García y Villatoro (2010) en su trabajo “El descubrimiento de conocimiento obtenido mediante el proceso de Inducción de árboles de decisión (Sigla en inglés TDIDT, Top Down Induction of Decision Trees)”, utilizaron árboles de decisión para modelar las clasificaciones de los datos. Uno de los principales resultados obtenidos fue la caracterización de los estudiantes con alto riesgo de abandonar sus estudios universitarios.

Kovacic (2012) estudió las variables socioeconómicas como la edad, el género, la etnia, la discapacidad, la situación laboral y el programa de estudio a distancia. El objetivo de la investigación fue identificar a los estudiantes con alto riesgo de abandonar sus estudios. Se utilizaron las técnicas de minería de datos árboles de decisión y la regresión logística, presentando ambos algoritmos una correcta clasificación.

Los autores Yadav, Bharadwaj y Pal (2012) presentaron un proyecto de minería de datos para generar modelos predictivos e identificar a los estudiantes con alto riesgo de deserción teniendo en cuenta los registros de los estudiantes en la primera inscripción. La calidad de los modelos de predicción se examinó con los algoritmos ID3, C4.5 y ADT

de las técnicas de árbol decisión. Los algoritmos de aprendizaje automático ADT pueden aprender de modelos predictivos con los datos de los estudiantes de los años anteriores.

Los autores Zhang y Oussena (2010) propusieron la construcción de un sistema de gestión basado en minería de datos (Sigla en inglés MCMS, Mining Course Management Systems), una vez procesados los datos en el sistema de gestión, los autores identificaron las características de los estudiantes que no tuvieron éxito en el semestre, utilizaron tres algoritmos de clasificación, máquina de vectores, Naive Bayes y árbol de decisión. La mayor precisión en la clasificación se presentó con el algoritmo Naive Bayes, mientras que el árbol de decisión obtuvo uno de los valores más bajos.

En Colombia, el autor Timarán (2010) explicó el proceso de descubrimiento de conocimiento que se hizo en la Universidad Antonio Nariño para establecer en la comunidad universitaria los perfiles de bajo rendimiento académico y deserción estudiantil, para lo cual se utilizó la base de datos histórica de los estudiantes de pregrado. El objetivo de la investigación fue la búsqueda y descubrimiento de patrones de bajo rendimiento académico y deserción estudiantil aplicando técnicas de minería de datos, como clasificación, agrupamiento y reglas de asociación.

1.9.5 LA APLICACIÓN DE LA MINERIA DE DATOS A LA DESERCIÓN ESTUDIANTIL

En el estudio de la deserción estudiantil, el uso de la minería de datos brinda una solución para el análisis de los fenómenos no evidentes en bases de datos. Es mirada como una herramienta que permite examinar bases de datos robustas de forma más minuciosa a partir de información suministrada por los educandos ya sea en los procesos de matrícula o por encuestas. Todo esto facilita la toma de decisiones.

En este sentido, el objetivo principal de la mineria de datos es incorporar, distinguir los datos y separar modelos que estructuran determinados patrones a partir de los datos analizados, con los que permite conseguir una descripción del comportamiento de los datos, tendencias y correlaciones. La Minería de Datos constituye la etapa de descubrimiento en el proceso de KDD (Knowledge Discovery in Databases) , el cuál consiste en el uso de algoritmos concretos, los cuales generan una enumeración de

patrones a partir de los datos anteriormente procesados, apoyándose con algoritmos de Aprendizaje Automático. Todos los datos y patrones estarían en pro de identificar vulnerabilidades en el área socioeconómica y personal de los estudiantes (Aguirre., et al 2015).

2. METODOLOGIAS DE EXTRACCION DE CONOCIMIENTO

2.1 FASES METODOLOGIA CRISP-DM

El modelo de referencia CRISP-DM es el estándar de Data Mining más utilizado que divide el ciclo de vida de un ejercicio de minería de datos en seis fases diferentes. También define las tareas a realizar en cada fase y el resultado esperado. El siguiente paso depende del resultado del paso anterior, igualmente le permite retroceder a un paso anterior y es menos rígido que un modelo de cascada tradicional utilizado en el desarrollo de software.

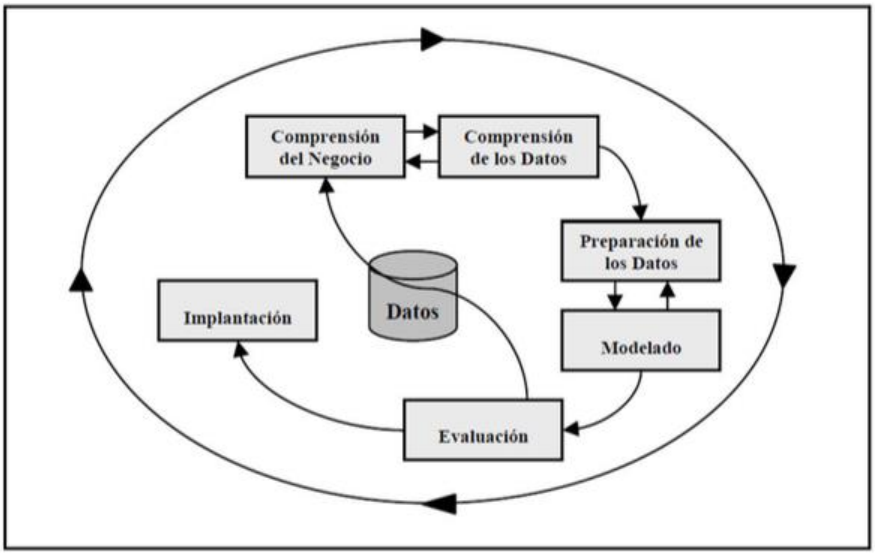


Figura 5: Metodología CRISP-DM (Arancibia, 2010)

2.1.1 PRIMERA FASE

La primera fase es la comprensión comercial o del negocio, se centra en comprender los objetivos comerciales y evaluar la situación para derivar un problema de minería de datos que está alineado con los objetivos comerciales. El resultado es un plan preliminar diseñado para alcanzar los objetivos del proyecto. Probablemente es la etapa más importante.

2.1.2 SEGUNDA FASE

La segunda fase es la fase de comprensión de datos que proporciona una comprensión de los datos que deben analizarse. Permite al experto en minería de datos familiarizarse con los datos y ayuda a identificar posibles problemas de calidad. Después de obtener una comprensión de los datos, los datos deben prepararse para el modelado.

2.1.3 TERCERA FASE

La tercera fase es la preparación de datos incluye decidir qué se debe incluir en el conjunto de datos, limpiar los datos y todas las demás actividades que se deben realizar para procesar los datos que sirven como una entrada para la herramienta de modelado en el paso siguiente. En la práctica, muchas de las tareas pueden ser realizadas en un orden diferente, y a menudo será necesario volver a hacer tareas anteriores repetidamente y repetir ciertas acciones.

2.1.4 CUARTA FASE

La cuarta fase es la fase de modelado. En esta fase, se seleccionan y aplican diversas técnicas de modelado, y sus parámetros se referencian a valores inmejorables. Por lo regular, existen diversas técnicas para el mismo de problema de minería de datos. Algunas de estas técnicas requieren formatos de datos específicos. Hay un vínculo estrecho entre la fase de preparación de datos y la fase de modelado. Con frecuencia, nos damos cuenta de los problemas que tenemos con los datos al modelar o surgen nuevas ideas para construir nuevos datos.

2.1.5 QUINTA FASE

La quinta fase es la fase de evaluación, es donde el modelo se evalúa y los resultados se analizan en relación con el criterio de éxito empresarial. Si el modelo y los resultados no están alineados con los objetivos comerciales, se inicia un nuevo ciclo de CRISP-DM.

2.1.6 SEXTA FASE

La sexta fase es la fase de implantación. La producción del modelo regularmente no es el final del proyecto. Por lo general, el conocimiento adquirido deberá organizarse y presentarse de manera que el cliente pueda utilizarlo. Dependiendo de los requisitos, la fase de implantación puede ser tan simple como generar un informe o tan compleja como implementar un proceso de minería de datos repetible. En muchos casos, será el usuario, no el analista de datos, quien llevará a cabo los pasos de implementación. En cualquier caso, es importante comprender por adelantado qué acciones deberán llevarse a cabo para poder utilizar realmente los modelos creados.

2.2 METODOLOGIA SEMMA

La metodología SEMMA fue desarrollada por el Instituto SAS (Principal fabricante de software de inteligencia empresarial). El acrónimo SEMMA significa Muestra, Explorar, Modificar, Modelar, Evaluar y se refiere al proceso de llevar a cabo un proyecto de minería de datos (SEMMA SAS Institute, 2008).

2.2.1 ETAPAS METODOLOGIA SEMMA

2.2.2 PRIMERA ETAPA

La primera etapa es la etapa de Muestra. Esta etapa consiste en muestrear los datos mediante la extracción de una parte de un conjunto de datos grandes, lo suficientemente grande como para contener información significativa, pero lo suficientemente pequeña como para manipular rápidamente. Esta etapa se señala como opcional.

2.2.3 SEGUNDA ETAPA

La segunda etapa es la etapa de exploración. Esta etapa consiste en la exploración de los datos mediante la búsqueda de tendencias y anomalías no anticipadas para obtener comprensión e ideas.

2.2.4 TERCERA ETAPA

La tercera etapa es la etapa de modificar. Esta etapa consiste en la modificación de los datos mediante la creación, selección y transformación de las variables para enfocar el proceso de selección del modelo.

2.2.5 CUARTA ETAPA

La cuarta etapa es la etapa de modelar. Esta etapa consiste en modelar los datos al permitir que el software busque automáticamente una combinación de datos que predice de manera confiable el resultado deseado.

2.2.6 QUINTA ETAPA

La quinta etapa es la etapa de evaluación. Esta etapa consiste en evaluar los datos mediante la evaluación de la utilidad y confiabilidad de los hallazgos del proceso de extracción de datos y estimar qué tan bien se desempeñan. SEMMA ofrece un proceso fácil de entender que permite un desarrollo y mantenimiento adecuado y organizado de los proyectos de DM. Por lo tanto, confiere una estructura para su concepción, creación y evolución, ayudando a presentar soluciones a los problemas de negocios, así como a encontrar los objetivos de negocio de DM. (Azevedo y Santos, 2008).

2.2.7 GRÁFICO METODOLOGÍA SEMMA

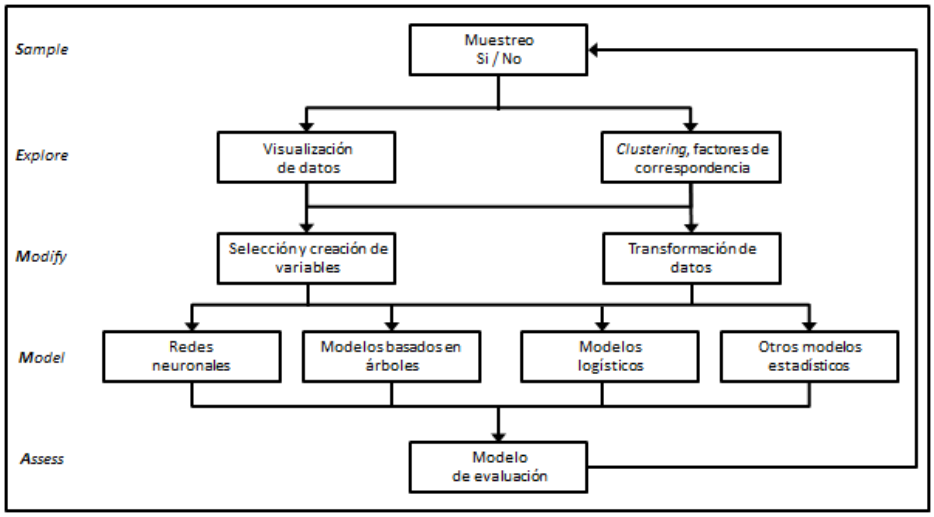


Figura 6: Metodologia SEMMA.

2.3 COMPARACION DE LAS METODOLOGIAS CRISP-DM Y SEMMA

Las metodologías CRISP-DM y SEMMA comparten la misma estructura, segmentando el proyecto de mineria de datos en fases que se encuentran relacionadas entre sí, convirtiendo el proyecto de mineria de datos en un proceso iterativo e interactivo.

La metodología SEMMA se enfoca más en las características técnicas del progreso del proyecto, mientras que la metodología CRISP-DM, tiene una visión más amplia en relación a los objetivos empresariales del proyecto. Esta diferencia se puede observar desde la primera fase del proyecto de Mineria de datos, ya que la metodología SEMMA empieza ejecutando un muestreo de datos, y la metodología CRISP-DM empieza realizando un análisis del problema empresarial para transformarlo en un problema técnico. A partir de ese punto de vista más completo se puede pensar que la metodología CRISP-DM está más cercana al concepto real del proyecto, logrando ser integrada con una Metodología de Gestión de Proyectos especifica que completaría las tareas técnicas y administrativas.

2.3.1 GRÁFICO COMPARATIVA DE LA METODOLOGÍA CRISP-DM Y SEMMA

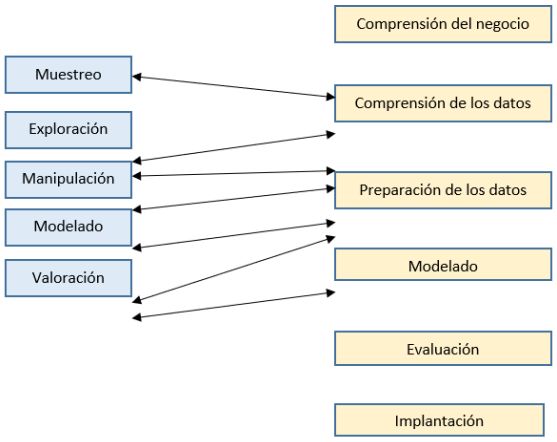


Figura 7: Metodología CRISP-DM Y SEMMA

Otra diferencia entre las metodología SEMMA y CRISP-DM es su conexión con herramientas comerciales. La metodología SEMMA sólo es abierta en sus aspectos generales porque está muy orientada a las empresas SAS en que se encuentra ejecutada. Pero la metodología CRISP-DM ha sido diseñada como una metodología indeterminada respecto a la herramienta que se utilice para el desarrollo del proyecto de mineria de datos y su distribución es libre.

2.4 LA MEDICIÓN DE LA DESERCION SENA VS SPADIES

El SPADIES (Sistema de Prevención y Análisis de la Deserción en las Instituciones de Educación Superior en Colombia) es una herramienta desarrollada por el MEN (Ministerio de educación nacional) para hacerle seguimiento especializado al fenómeno de la deserción estudiantil en las Instituciones de Educación Superior del país y dado que el SENA está suscrita al Ministerio del trabajo, y trabaja con programas técnicos y tecnológicos, no está supervisado por este sistema, no obstante, de manera interna se hacen trabajos específicos para disminuir la deserción con diferentes estrategias propias de la entidad.

En este sentido, los estudios parten de un seguimiento que el SENA realiza al fenómeno de la deserción, en relación con la identificación de las variables, la unicidad de manejo de las mismas y de las estrategias de prevención coherentes con las necesidades particulares de los aprendices, que en su mayoría corresponden a población de escasos recursos. Estas mejoras son soportadas en investigaciones internas sobre las variables que deben ser(citaciones) consideradas en la medición del fenómeno y al mismo tiempo en modelos de medición que puedan ser adaptados al contexto Sena; buscando que los aprendices culminen sus procesos de formación, lo que redundará en mejores oportunidades laborales y económicas.

2.5 EL SERVICIO NACIONAL DE APRENDIZAJE SENA Y SUS CARACTERISTICAS

Como institución de formación para el trabajo que oferta programas de formación con una duración variable, el Sena tiene una manera particular de medir el fenómeno la deserción, según los parámetros establecidos en el reglamento del aprendiz, capítulo VII, artículo 22, punto 4:

Se considera deserción en el proceso de formación:

A. Cuando el aprendiz injustificadamente no se presente por tres (3) días consecutivos al Centro de Formación o empresa en su proceso formativo.
B. Cuando al terminar el periodo de aplazamiento aprobado por el SENA el aprendiz no reingresa al programa de formación.
C. Cuando transcurridos dos (2) años contados a partir de la fecha de terminación de la etapa lectiva del programa, el aprendiz no ha presentado la evidencia de la realización de la etapa productiva. (SENA, 2019. p. 17).

De la misma manera se establece una figura adicional, el retiro voluntario que se define como la solicitud formal que el aprendiz eleva, a través de un oficio radicado en el centro de formación y se registra en el sistema de gestión de formación para retirarse definitivamente del programa de formación, por ello no es claro si los datos de deserción incluyen los retiros voluntarios o no. La información consolidada sobre la deserción se

gestiona a través de la plataforma SOFIA PLUS, y de allí se derivan mensualmente las cifras de deserción discriminando, regionales, centros de formación, tipos de formación (técnico, tecnólogo, especializaciones tecnológicas) y/o educación complementaria (cursos especiales y eventos). Por el momento esta plataforma solo reporta datos estadísticos, pero no permite generar un diagnóstico sobre el fenómeno, ni monitorear el impacto de las estrategias implementadas para su mitigación (SENA, 2019).

2.6 METODOLOGÍA DE INVESTIGACIÓN

2.6.1 TIPO DE INVESTIGACIÓN

El presente estudio es cuantitativo, descriptivo y expostfacto no experimental enfocado en el manejo de bases de datos desde la cual se establece una nueva propuesta metodológica de extracción de conocimeinto para la valoración y análisis de los posibles riesgos de deserción en la población estudiantil.

2.6.2 MÉTODOS DE INVESTIGACIÓN

El método de investigación que se utilizará para el desarrollo del objetivo central será sistémico estructural dando un enfoque holístico. Que conlleva a contemplar el eje problemático como un todo. El Método Sistémico Estructural expresa la lógica o sucesión de procedimientos seguidos por el investigador en la construcción del conocimiento, a través del esquema de categorías, sus definiciones, relaciones y funciones asociadas con la seguridad de la información. Por otra parte, el modelo holístico proporciona una visión compleja para el desarrollo de la investigación, con un proceso global, integrador, evolutivo, concatenado y organizado, reforzadando el primero método ya mencionado (De Barrera y Morales, 2000).

Su meta no se limita a la recolección de datos, sino a la predicción e identificación de los riesgos de deserción estudiantil de acuerdo al manejo de los datos. Se recogerá los datos sobre la base de los paradigmas investigativos y problemáticas planteadas previamente, se expondrá y resumirá la información de manera cuidadosa y luego se analizan minuciosamente los resultados, a fin de proponer el modelo más propicio (Van

Dalen y Meyer, 2006). Para el contexto en cuestión es la determinación de una nueva metodología que responda a las características de la educación.

2.6.3 TÉCNICAS PARA LA RECOPILACIÓN DE INFORMACIÓN

Para formar una idea solidad del estudio de esta investigación, se recurre a la necesidad de utilizar la técnica de: la observación, análisis de documentos históricos, la recopilación documental y bibliográfica; de la misma manera, se recurrirán a técnicas directas como la entrevista directa a los aprendices del Sena, que brinden información sobre el tema de investigación, por medio de un cuestionario bien estructurado para todas las entrevistas.

CAPITULO II

3. APLICACIÓN METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTO CRISP-DM

3.1 COMPRENSIÓN DEL NEGOCIO

El Servicio Nacional de Aprendizaje (SENA) cuenta con un número determinado de centros de formación ubicados en puntos estratégicos de la ciudad de Barranquilla, para que los aprendices puedan acceder a ellos. Cada centro de formación cuenta con su oferta académica de acuerdo a su especialidad.

El Sena cuenta con un grupo de instructores idóneos en cada programa de formación, los cuales permite a los aprendices recibir una formación académica de calidad. Sin embargo, el problema de la deserción estudiantil no es ajeno al Sena. Existen unos factores de deserción ya establecidos por el Sena Centro Industrial y de Aviación que se han estudiado a través de los años, los cuales son determinados por los instructores es su ambiente de formación.

Cuando un instructor nota la ausencia de un aprendiz lo reporta a la oficina de bienestar al aprendiz, pero en algunos casos no se puede determinar la causa exacta de su deserción ya que es imposible contactar al aprendiz. Cuando esto sucede el instructor

da como causa de deserción “Falta de interés en la formación”, sin saber si en realidad el aprendiz desertó por esa causa.

En esta fase se requiere comprender a que problema se le va a realizar la solución utilizando minería de datos, para la recolección de datos necesarios para interpretar los resultados encontrados. Los datos proporcionados por el Servicio Nacional de Aprendizaje (SENA) en los programas ofertados en los años 2015, 2016, 2017 y 2018, se observó que hasta el año 2018 desertaron un total de 1311 aprendices correspondiente a un índice de deserción del 21%.

Todo esto ha llevado que el problema de deserción en las instituciones de nivel técnico y tecnológico se convirtiera en algo por resolver con procesos de minería de datos, con el fin de detectar patrones de deserción estudiantil.

3.2 DETERMINAR LOS OBJETIVOS DEL NEGOCIO

El Servicio Nacional de Aprendizaje - SENA, es un establecimiento público del orden nacional, con autonomía administrativa; cuenta un varias carreras técnicas, tecnológicas y complementarias que benefician a la comunidad en general, con una formación gratuita, para cuando terminen sus estudios logren desempeñarse en actividades productivas en las diferentes empresas que lo requieran, logrando así una mejor competividad en el mercado globalizado.

El servicio nacional de aprendizaje – SENA, cuenta con varias sedes a nivel nacional, pero este estudio se realiza en el nodo del Sena Centro Industrial y de Aviación ubicado en la ciudad de barranquilla. Sus aulas están dotadas de todas las herramientas de aprendizaje requeridas de acuerdo a la especialidad, para el ingreso a este nodo educativo se realiza un proceso de selección y luego se reubican después del proceso de matrícula de acuerdo el programa escogido por el aspirante, su enseñanza es de forma presencial, las clases son de lunes a viernes en las diferentes jornadas establecidas por el Sena, todos los estudiantes deben cumplir con las reglas establecidas tanto con el uniforme y la puntualidad para el ingreso a las aulas, los docentes realizan

sus actividades de aprendizaje generando conocimiento complementadas con talleres en clases. Después de todo este complemento se requiere realizar los siguientes objetivos de negocio del proyecto.

A. Establecer perfiles de los aprendices de acuerdo a su programa de formación en el Sena Centro Industrial de y Aviación.

B. Realizar estadísticas de todos los aprendices matriculados en los programas que ofrece el Sena Centro Industrial y de Aviación para realizar un seguimiento en todas sus actividades tanto académica como asistencial.

C. Establecer cuáles son los programas con los mayores índices de deserción en el Sena Centro Industrial y de Aviación.

3.3 EVALUACION DE LA SITUACIÓN

Ya obtenidos y recopilados los objetivos del negocio del proyecto se procede a darle inicio a la evolución de toda la información conocida.

3.3.1 LISTADO DE RECURSOS

A. **PERSONAL:**

Dos ingenieros de sistemas

B. **DATOS:**

Archivos de Excel donde se encuentran almacenadadas las causas relevantes de deserción de los aprendices del Sena Centro Industrial y de Aviación.

C. **SOFTWARE:**

Se hará el proceso de transformación, extracción y carga de los datos obtenidos en el software minería de datos PENTAHO. Igualmente, el proceso de modelado de los datos se desarrollará con el software Rapid Miner.

3.3.2 RESTRICCIONES:

A. El Centro Industrial y de Aviación tiene sus propios factores de deserción, lo cual nos restringe a utilizar otra fuente de datos distinta a ésta.

3.3.3 RIESGOS Y CONTINGENCIAS:

Los riesgos que tiene nuestro proyecto son los que se presentan a continuación:

A. Los datos que tengan los campos utilizados, contengan información vacía o duplicada ya que se encuentran en archivos de Excel.

B. No haya calidad en los datos suministrados por Centro Industrial y de Aviación, es decir, se encuentren que los datos estén desactualizados o que no contengan la información correcta.

3.3.4 BENEFICIOS:

Elaboración de una metodología de extracción de conocimiento para la valoración del riesgo de deserción de los aprendices del Sena Centro Industrial y de Aviación, y asi contribuir a la toma de decisiones para ayudar a disminuir las deserciones de los aprendices, acompañados por bienestar estudiantil y los docentes, dándole un buen uso a las herramientas ofrecidas por el grupo investigador.

3.4 DETERMINACION DE LOS OBJETIVOS DE LA MINERIA DE DATOS

Los objetivos de mineria de datos de nuestra propuesta son los siguientes:

A. Establecer los diferentes mecanismos de seguimientos realizados por los docentes y bienestar estudiantil a los aprendices de el Sena Centro Industrial y de aviación.

B. Identificar cuáles son los patrones más relevantes de la deserción de los aprendices de Sena Centro Industrial de y de aviación.

C. Utilizar las técnicas de regla de asociaciones y árbol de decisiones con el fin de explorar los datos de los aprendices.

D. Evaluar el conjunto de datos obtenidos de la metodología a proponer, para la concepción de predicciones**.**

E. Predecir la deserción de los aprendices de acuerdo a variables socioeconómicas y académicas.

3.5 PRODUCIR EL PLAN DEL PROYECTO

Ver anexo 1 y anexo 2.

3.6. COMPRENSION DE LOS DATOS

3.6.1 RECOLECCION DE DATOS INICIALES

Los datos que se recolectaron para este proyecto fueron obtenidos de diversos archivos de Excel que hacen referencia a la información con respecto a la deserción de los aprendices en el Sena Centro Industrial y de Aviación. Sin embargo, para la obtención de los datos se tuvieron algunas dificultades ya que el departamento de Bienestar al aprendiz del Centro Industrial y de Aviación no tiene de manera organizada y actualizada los reportes de deserción de los últimos años.

Los datos recolectados son:

A. Nombres y apellidos del aprendiz
B. Documento de identificación
C. Programa de Formación
D. Causa de deserción

3.6.2 DESCRIPCIÓN DE LOS DATOS

Los datos mencionados anteriormente, contienen 1 o varios campos asignados a la base de datos y en diversas tablas. A continuación, se hace una descripción de dicha información.

ARCHIVO AÑO 2015: Este archivo contiene la información correspondiente a la deserción en el año 2015 en el Centro Industrial y de Aviación. Ver anexo 3.

ARCHIVO AÑO 2016: Este archivo contiene la información correspondiente a la deserción en el año 2016 en el Centro Industrial y de Aviación. Ver anexo 4.

ARCHIVO AÑO 2017: Este archivo contiene la información correspondiente a la deserción en el año 2017 en el Centro Industrial y de Aviación. Ver anexo 5.

ARCHIVO AÑO 2018: Este archivo contiene la información correspondiente a la deserción en el año 2018 en el Centro Industrial y de Aviación. Ver anexo 6.

3.6.3 VERIFICACION DE LOS DATOS

Al realizar una exploración minuciosa de los datos recolectados se observa que en todos los años hay datos que son inconsistentes, los cuales se listan a continuación.

A. En los años 2015, 2016. 2017 y 2018 se presentan datos donde la causa de deserción es DESERTADO, DESERTO o está vacía, por lo tanto, toca unificar esa causa de deserción o buscar la causa exacta por el cual el aprendiz desertó.

3.6.4 PREPARACIÓN DE LOS DATOS

Para la construcción de los datos se utilizó la herramienta PENTAHO, lo cual es una herramienta de transformación, extracción y carga de datos. Ver anexo 7 al 38.

3.7 MODELADO

3.7.1 SELECCIÓN DE LA TÉCNICA DE MODELADO

Para la selección de la técnica de modelado se utilizó la herramienta RAPIDMINER STUDIO en la cual se introdujeron los datos del periodo 2018 como datos de entrenamiento, teniendo en cuenta que nuestra a variable a predecir es la deserción.

Es importante destacar que el software Rapidminer es una plataforma para el desarrollo de sofisticados modelos estadísticos y el análisis predictivo de grandes paquetes de datos, dispone de un potente entorno visual para el encadenamiento de operadores, extensiones para importar y manejar series financieras, compatibilidad con R y una abundantísima documentación mediante video tutoriales y aplicaciones prediseñadas para diversos escenarios.

Los operadores disponibles cubren una inmensa variedad de tareas; desde conectividad con bases de datos y otras aplicaciones estadísticas hasta algoritmos avanzados de clasificación, segmentación, predicción, text mining y análisis de procesos. Otro de sus grandes atractivos es el ser una herramienta *open source* muy empleada en el mundo académico y gratuita para usos no comerciales (García, 2017). En la aplicación de esta técnica se obtuvieron los siguientes resultados:

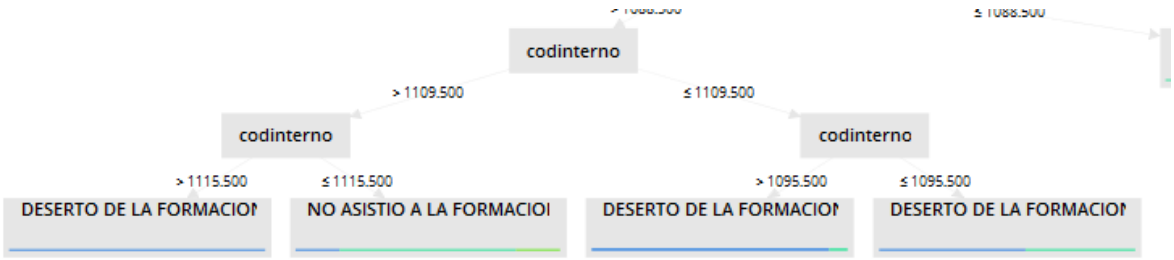


Figura 8: Arbol de decisiones (1)

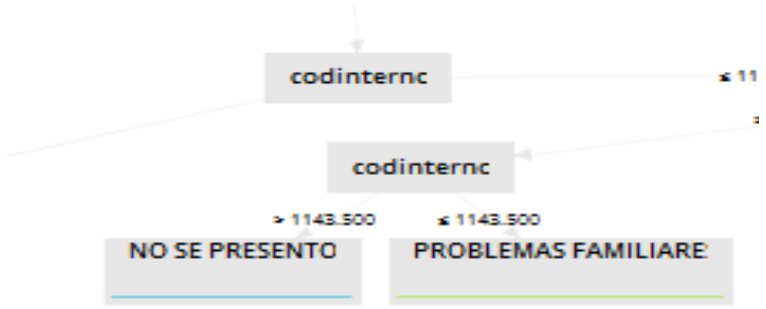


Figura 9 : Arbol de decisiones (2)

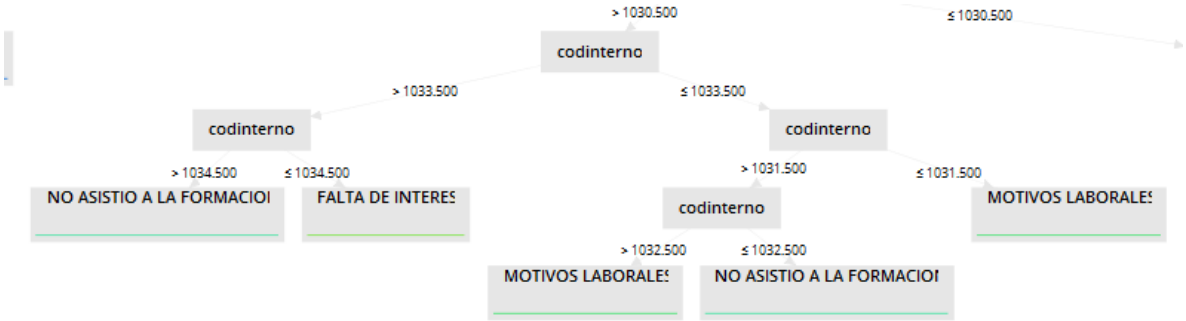
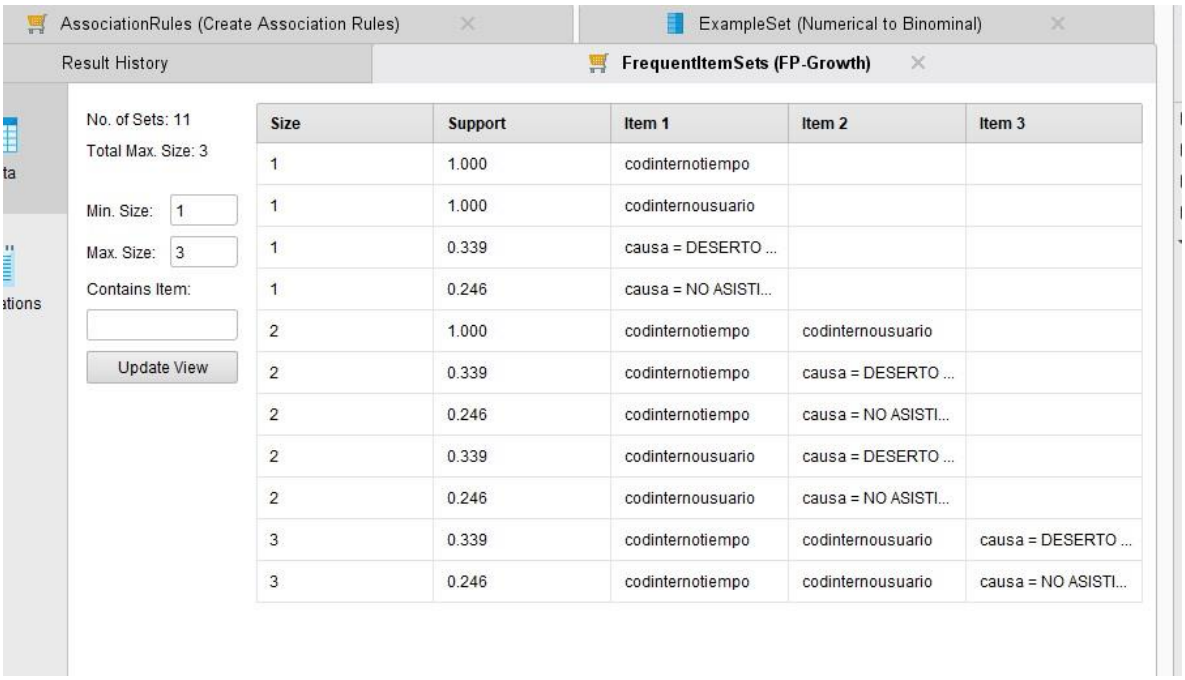


Figura 10 : Arbol de Decisiones (3)



Size	Support	Item 1	Item 2	Item 3
1	1.000	codinternotiempo		
1	1.000	codinternousuario		
1	0.339	causa = DESERTO ...		
1	0.246	causa = NO ASISTI...		
2	1.000	codinternotiempo	codinternousuario	
2	0.339	codinternotiempo	causa = DESERTO ...	
2	0.246	codinternotiempo	causa = NO ASISTI...	
2	0.339	codinternousuario	causa = DESERTO ...	
2	0.246	codinternousuario	causa = NO ASISTI...	
3	0.339	codinternotiempo	codinternousuario	causa = DESERTO ...
3	0.246	codinternotiempo	codinternousuario	causa = NO ASISTI...

Figura 11: Regla de asociaciones

3.7.2 EVALUACION

De acuerdo a la tabla anterior, se puede apreciar que los valores que tienen solo un ítem, sus dos primeros (conditiempo y condinternosaurios) presentan un porcentaje del 100% y en los posteriores disminuye del 30 % al 20% de frecuencia en los ítems causa= DESERTO y causa=NO ASISTIO respectivamente. No presentan relación alguna con los demás ítems.

Los valores que tienen dos ítems se puede ilustrar que el primero (codinterusuario) presenta un valor de 100% con una relación con el segundo (condigointerusuario). De manera posterior los porcentajes de los ítems oscilan entre el 30% y el 20% relacionados con los segundos ítems.

Los valores que presentáron tres relaciones tienen un porcentaje del 30%, códinternotiempo relacionado con codinternousuario y causadeserto. La última con un porcentaje del 20% y una relación con codinternousuario y causa=NO ASISTIO.

En el análisis del árbol de decisiones se ilustra (figura 8) que menor del 11% tiene como causa de deserción DESERTO DE LA FORMACIÓN, frente al 11% que 1 o 1% mayor o igual tiene como causa de deserción NO ASISTIO A LA FORMACION. Por otro lado, se comparan los códigos internos en el que menos del 10% tiene como causa de deserción DESERTO DE LA FORMACION frente a menos del 10 % que tiene como causa DESERTO DE LA FORMACION.

En la figura 9 se aprecia que un 11% tiene como causa de deserción NO SE PRESENTO A LA FORMACIÓN, en comparación con un porcentaje mayor del 11% que tuvo como causa PROBLEMAS FAMILIARES

En la figura 10 se ilustra menos del 10% de esta ramificación tiene como causa de deserción NO ASISTIO A LA FORMACION, en contraparte con más del 10% de los aprendices que tuvo como causa FALTA DE INTERES EN EL CURSO. Seguido a ello los MOTIVOS LABORALES tuvieron menos del 10% en comparación de la causa NO ASISTIO A LA FORMACION.

4. APLICACIÓN METODOLOGIA DE EXTRACCION DE CONOCIMIENTO SEMMA

4.1 MUESTREO DE LOS DATOS

El servicio nacional de aprendizaje SENA, es un establecimiento educativo del orden nacional con autonomía administrativa, donde los aprendices que se encuentran realizando estudios técnicos o tecnológicos son, la mayoria de ellos, de bajos recursos económicos, pero a lo largo de sus carreras, algunos aprendices presentan inconvenientes de diferente índole lo que conllevan a corto o a largo plazo en una deserción educativa. Este estudio se está realizando en el Sena Centro Industrial y de Aviación, ubicado en la ciudad de barranquilla, y se deben tomar como objeto de estudio las causas de deserción para que nos ayuden a disminuir esta problemática. Esta información se encuentra almacenada en hojas de cálculo y no se lleva un seguimiento apropiado, por ende, se procede a tomar muestras para su posterior estudio. Las muestras que se tomarán son las siguientes:

A. Tomar muestras más relevantes sobre las causas de deserción de los aprendices.
B. Tomar muestras sobre todos los programas que presenten mayor grado de inasistencias.

4.2 EXPLORACION DE LOS DATOS

Después de haber recopilado todas las muestras que se necesitan para el proceso de estudio, se pasa a revisar cada una de las muestras recopiladas por programas de formación y por todas las variables socioeconómicas y académicas que impiden un buen desempeño de los estudiantes, cada uno de estos componentes será revisado por una serie de personas idóneas que realizaran una exploración más afondo utilizando técnicas visuales y estadísticas. Para la exploración de los datos se requiere:

A. Verificar que los datos obtenidos de las hojas de cálculo no sean erróneos.
B. Revisar que las muestras de datos obtenidos de las hojas de cálculos no se encuentren datos repetidos.
C. Realizar una estadística de todos los datos más significativos recopilados en las hojas de cálculo.
D. Realizar un listado de todas las muestras más reiterantes por los aprendices.

4.3 MODIFICACION DE LOS DATOS

Una vez realizado un estudio más a fondo de muestreo de variables y exploración de variables obtenidas de las hojas de cálculo donde se encontraban realizando el control de los aprendices en estado de abandono educativo, se pasa a verificar la información obtenida por el personal delegado a revisar todas las muestras y notamos una reducción de variables más significativas, donde los estudiantes presentan el mayor grado de deserción estudiantil.

4.4 MODELADO Y VALORACIÓN

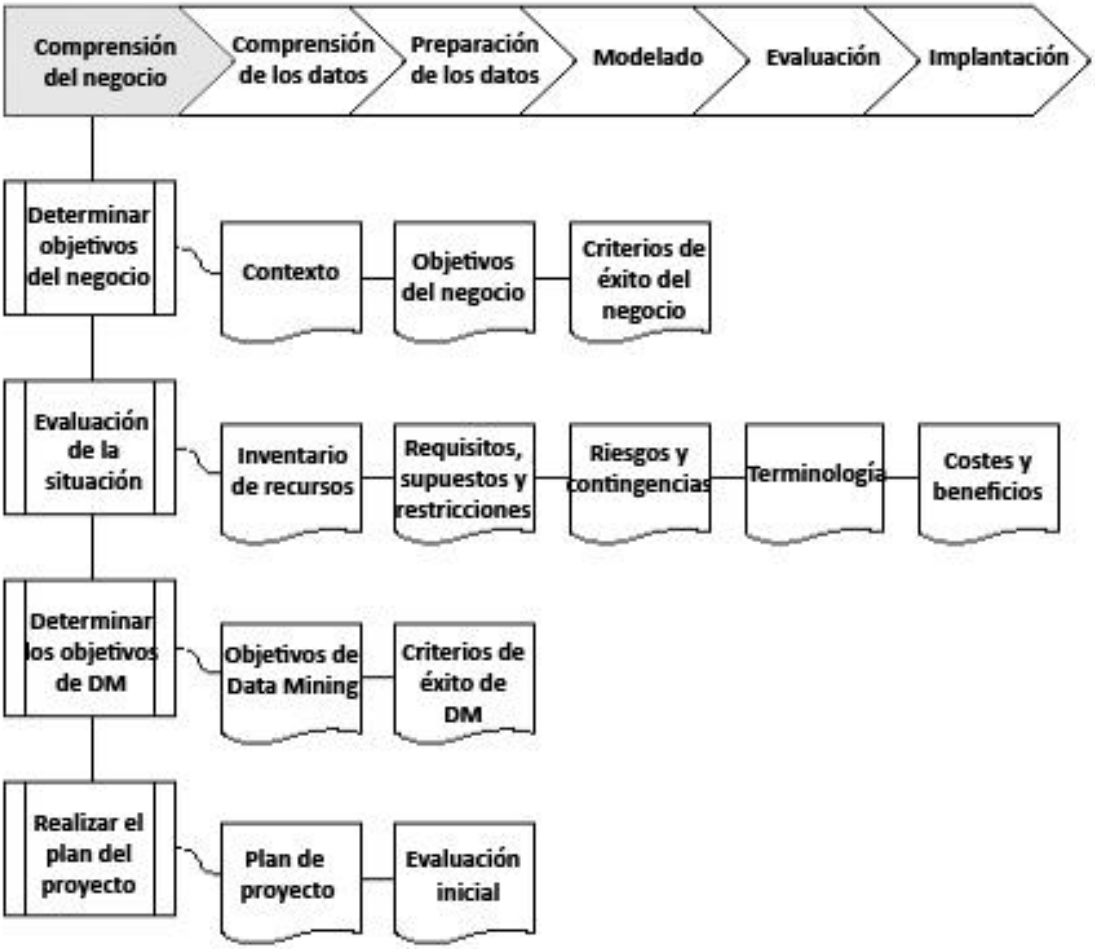
En el manejo de los datos se podrían aplicar los mismos procedimientos y técnicas al igual que la metodología CRISP y se obtienen, por lo tanto, que sus resultados no varían.

CAPITULO III

5. METODOLOGÍA DE EXTRACCIÓN DE CONOCIMIENTO PARA LA VALORACIÓN DEL RIESGO DE DESERCIÓN DE LOS APRENDICES DE EL SENA (MECDE)

Los datos que fueron recopilados con la metodología CRISP-DM en sus cuatro primeras fases, las cuales corresponden a las fases de comprensión del negocio, compresión de los datos, preparación de los datos y modelado, encontramos que sus datos no varían porque provienen de la misma fuente de información refiriéndonos a los archivos de Excel, los cuales fueron suministrados por la oficina de bienestar al aprendiz del Sena Centro Industrial y de Aviación, y éstas fases mencionadas se encuentran en la nueva metodología de extracción de conocimiento llamada MECDE.

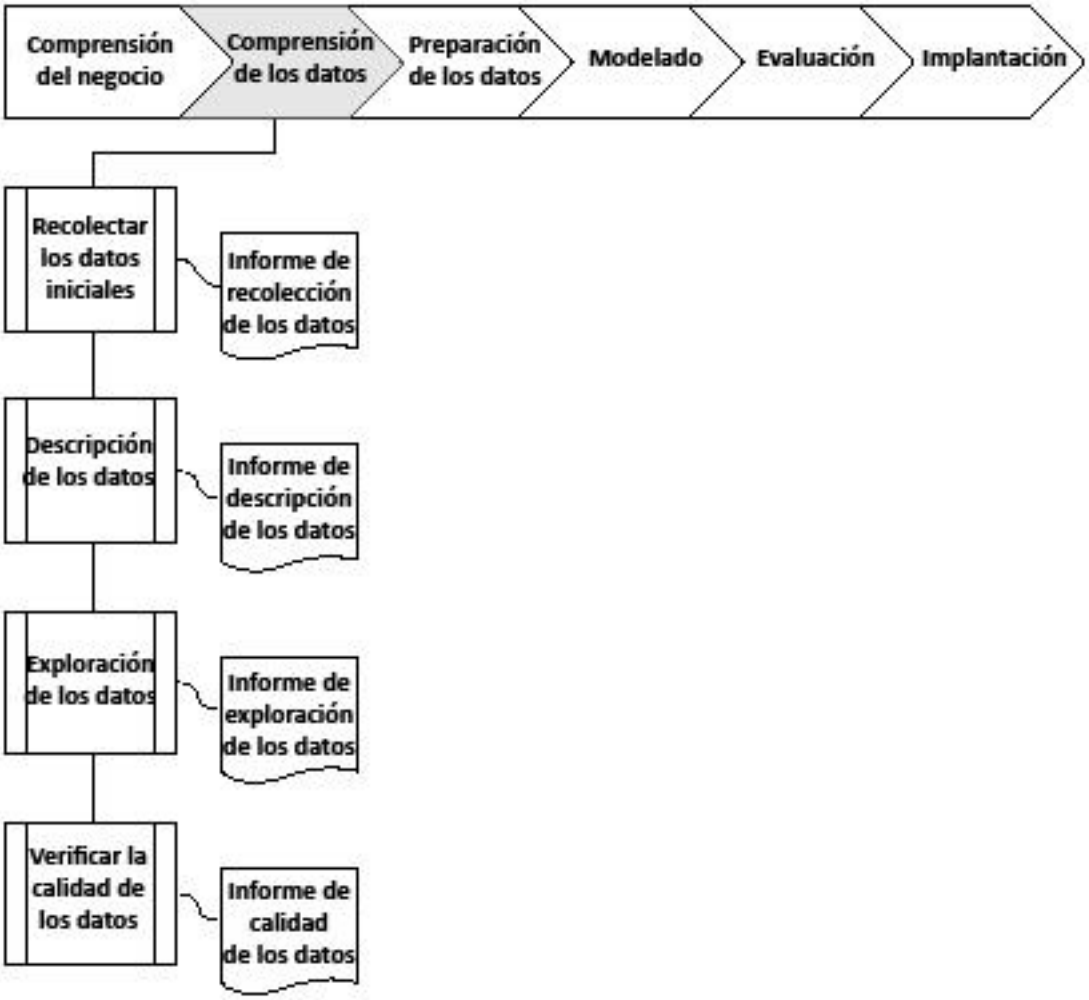
En la fase de comprensión del negocio, se hace una explicación de cómo se encuentra conformado y estructurado el Sena Centro Industrial y Aviación, y los factores de deserción estudiantil que lo afectan desde hace muchos años. Este proceso inicia desde la vinculación del aprendiz al momento que se matricula y escoge el programa de formación a iniciar, hasta poder detectar los niveles socioeconómicos y culturales, para poder hacer un seguimiento asistencial y académico y hacer una estadística de los programas con mayor índice de deserción estudiantil.



Gráfica fase comprensión del negocio metodología MECDE

Figura 12: Fase comprensión del negocio metodología MECDE

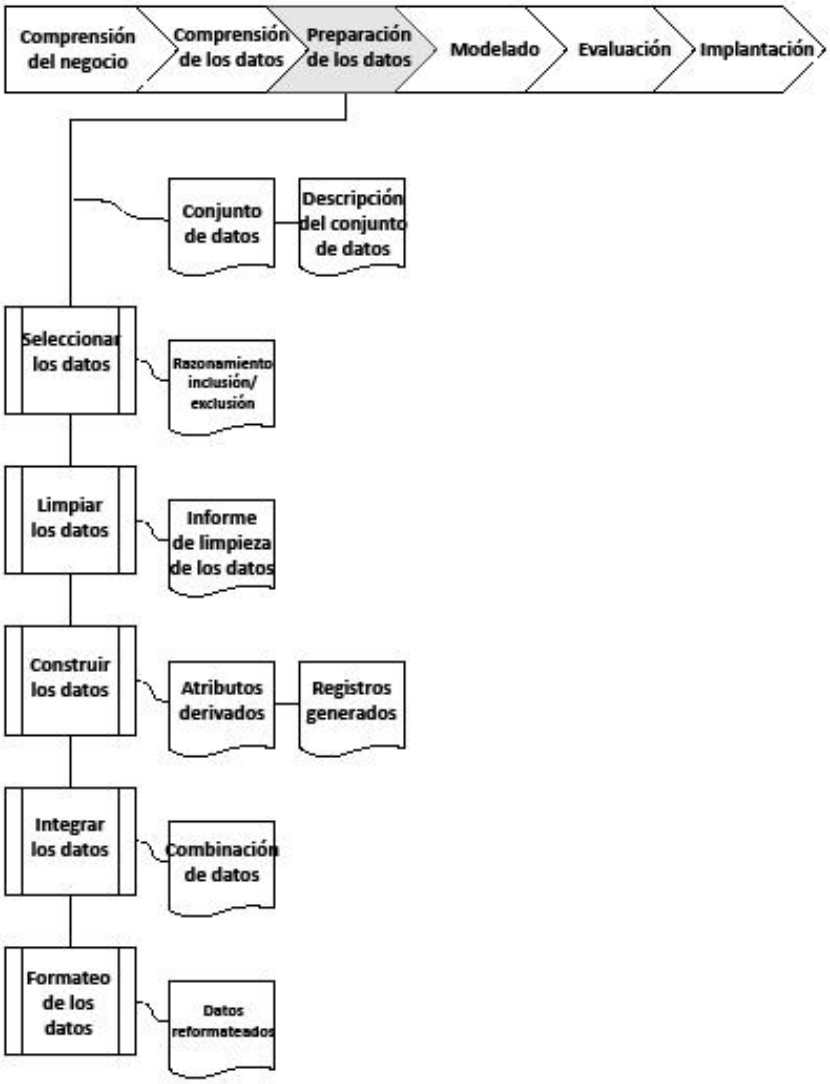
En la fase de compresión de los datos, hubo algunas dificultades en la recopilación de la información ya que el departamento de bienestar al aprendiz del Sena Centro Industrial y Aviación no tenía la información actualizada y la presentaba de una manera desorganizada y en archivos de Excel, lo cual se limitó a la escogencia de unos ítems específicos de la información básica de los aprendices.



Gráfica fase comprensión de los datos metodología MECDE

Figura 13: Fase comprensión de los datos metodología MECDE

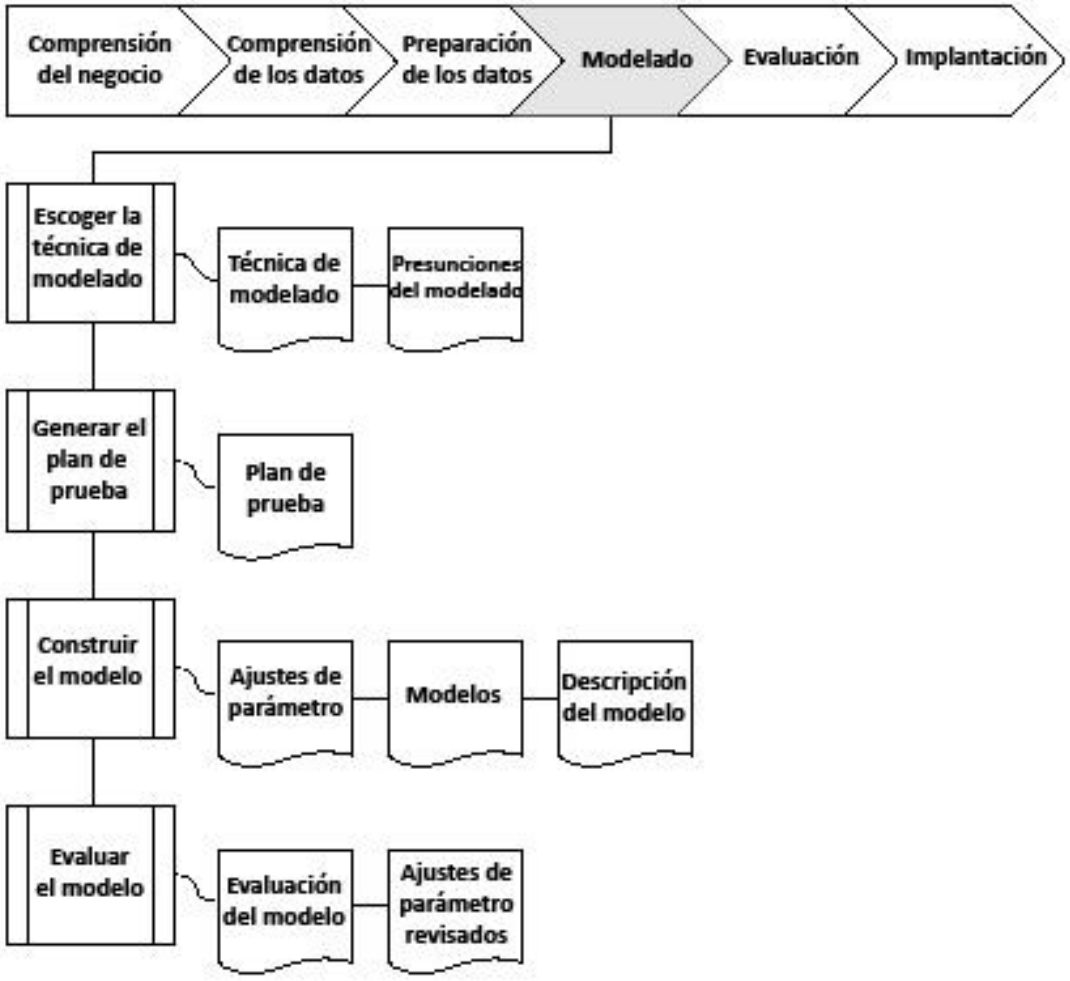
En la fase de preparación de los datos, se tomaron los archivos de Excel suministrados por el departamento de bienestar al aprendiz desde los años 2015 hasta el año 2018, donde se encontraron campos vacíos o unos datos no muy claros sobre el factor de deserción del aprendiz, lo cual se tomó la decisión de unificar algunos conceptos y buscar el factor exacto de deserción, para luego utilizar el Software PENTAHO para hacer la extracción, transformación y carga de los datos.



Gráfica fase preparación los datos metodología MECDE

Figura 14: Fase preparación de los datos metodología MECDE

Posteriormente, en la fase de modelado se aplica la técnica de modelado de los datos con la herramienta de minería de datos a utilizar, los resultados de esta aplicación sirven para obtener una técnica de modelado que muestre los factores de riesgo de deserción más determinantes, pero en una proporción generalizada de los estudiantes objeto de estudio por año en el Sena Centro Industrial y de Aviación.



Gráfica fase modelado metodología MECDE

Figura 15: Fase modelado metodología MECDE

5.1 ANÁLISIS DE DATOS UTLIZANDO HERRAMIENTAS DE INTELIGENCIA DE NEGOCIO

En el análisis realizado de las dos anteriores metodologías (CRISP-DM Y SEMMA) queda claro que tienen procesos fundamentales para extraer información sobre las causas de la deserción de estudiantes mediante la minería de datos, sin embargo, ambas

no fueron diseñadas con especificidades respecto de la educación. En este sentido las ilustraciones estadísticas y resultados no se ajustan a las peculiaridades de esta misma, es decir, que tienen lineamientos más ligados al campo empresarial.

Esta metodologia es una apuesta a identificar las dificultades mas recurrentes en los estudiantes en su proceso educativo. Esto va desde la asistencia a las instituciones a recibir sus clases, ausencias, limitaciones de recursos y motivacion. En este sentido, la construccion de la metodología de extracción de conocimientos es una herramienta crucial para identificar y abordar a los estudiantes con posibilidades de riesgos de desistir de su proyecto académicos. Es asi que este proyecto debe tener una vinculacion con otras dependencias de la institucion de educacion superior para poder hacer una sincronizacion de los casos y tomar las medidas pertinentes para garantizar la permanencia en el contexto de educacion superior.

Por lo anterior se puede determinar que la educación está impregnada de procesos humanos y por lo tanto se precisa de un análisis más cualitativo para abordar fenómenos como la deserción. Con este soporte es menester establecer estadísticas más descriptivas que pueda evidenciar causa y niveles desarrollo de la misma. En beneficio de esta metodología se podría mejorar en la toma de decisiones para solucionar con estrategias contundentes la deserción estudiantil. Con esto último también se intenta adicionar propuestas y sugerencias orientadoras que se encaminen a mejorar el estado de las cosas. Por consiguiente, esta nueva metodología adopta matices propios de la educación que le permite contextualizarse más a los contextos de formación.

Gráfica metodología extracción de conocimiento para la valoración del riesgo de deserción de los aprendices del Sena.

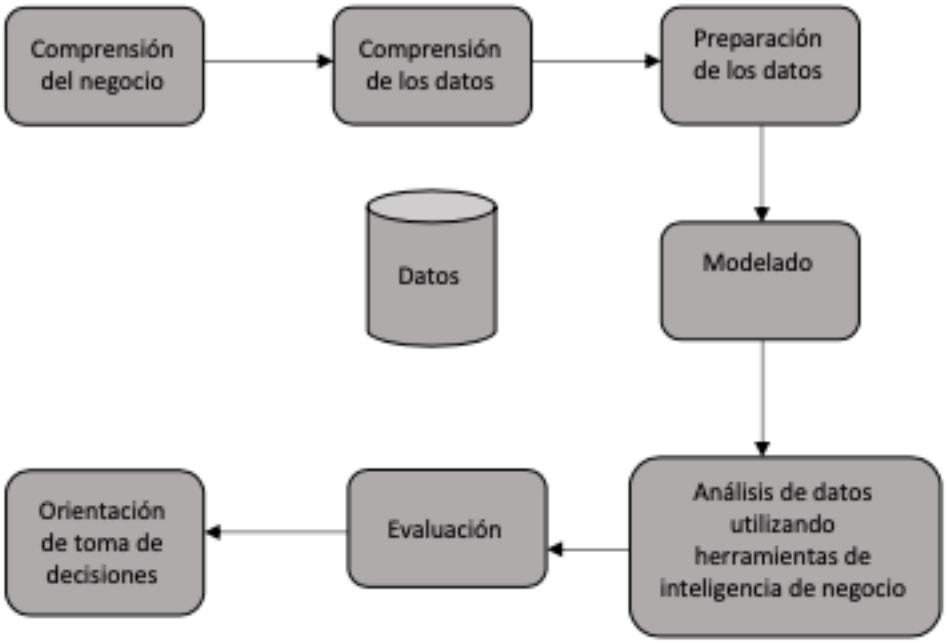
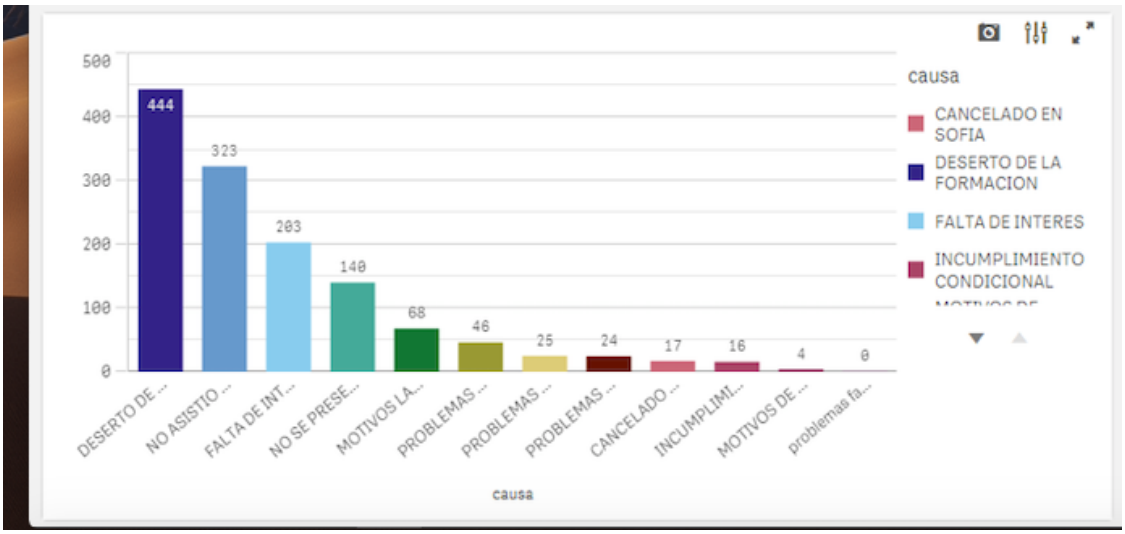


Figura 16: Metodología de extracción de conocimiento para la valoración del riesgo de deserción de los aprendices del Sena.

Para el manejo de los datos se arrojan resultados con el software Qlik sense. Ésta es una aplicación de Data Discovery la cuál busca entre datos sencillos y complejos. El objetivo es descubrir las relaciones ocultas entre ellos. Con esta herramienta cualquier miembro de la organización puede crear con facilidad visualizaciones flexibles e interactivas y tomar decisiones fundamentadas.

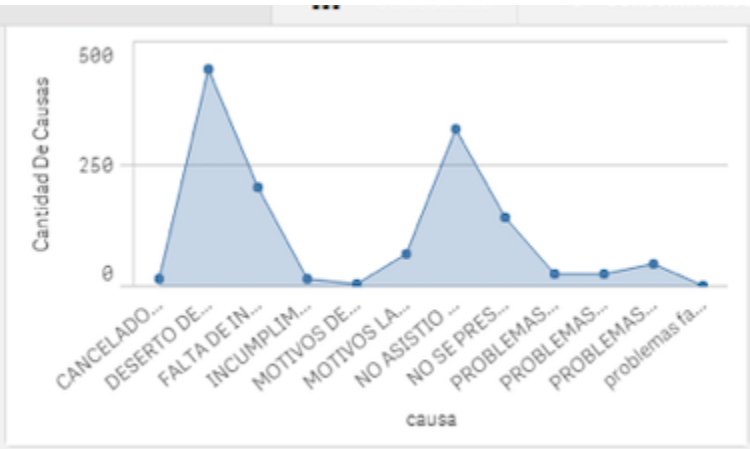
En el caso de la deserción estudiantil es importante obtener en las ilustraciones la cantidad de aprendices objeto de estudio, además graficas de barra, sectores y de líneas. Este último, ilustra cantidad de causas de deserción, el primero la cantidad y estado de cada causa de deserción y la segunda el manejo de los porcentajes de cada causa de deserción.

Figura 17: Patrones de desercion estudiantil SENA.



La presente figura evidencia a la causa de deserción desertor de formación con la cantidad más alta de aprendices (444) y seguido de éste la causa no asistió a la formación con (323) aprendices y por último la causa falta de interés en el curso en un tercer nivel con 203 aprendices. Por otro lado, las causas con cantidades más bajas fueron los problemas familiares con 0 aprendices y en segundo lugar 6 estudiantes en la categoría de motivos laborales.

Figura 18: Desercion estudiantil SENA.



De acuerdo a la figura se ilustra la curva de comportamiento de cada una de las causas de deserción, en esto guarda mucha similitud con la gráfica anterior solo con un enfoque en las líneas más altas y bajas.

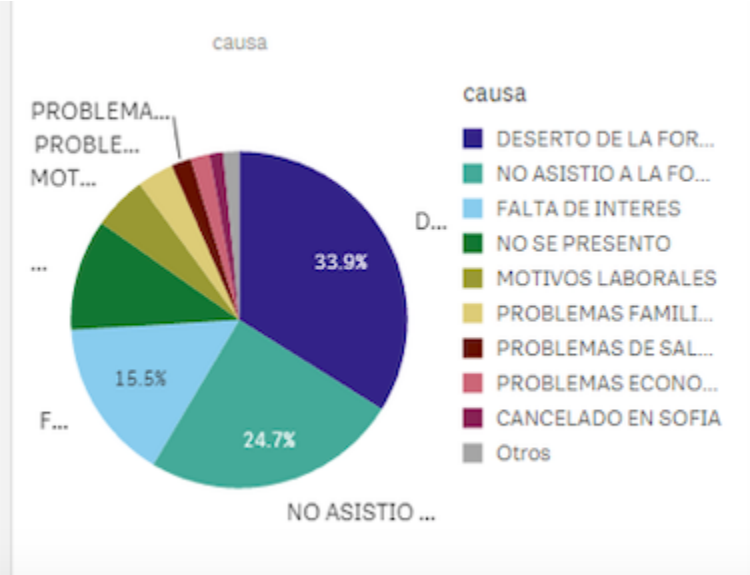


Figura 19: Desercion estudiantil SENA.

Esta figura se enfoca al igual que las dos anteriores, solo que ilustra las causas por porcentajes, es decir, que para el caso de la causa con mayor porcentaje de estudiantes se destaca desertor de la formación con un 33.9% de los aprendices, en segunda instancia la causa no asistió a la formación con un 24.7% y 15.5% la causa falta de interés. Entre las que obtuvieron menor porcentaje son relevantes problemas familiares, motivos laborales, y problemas de salud.

Figura 20: Desercion estudiantil por programa tecnicó SENA.

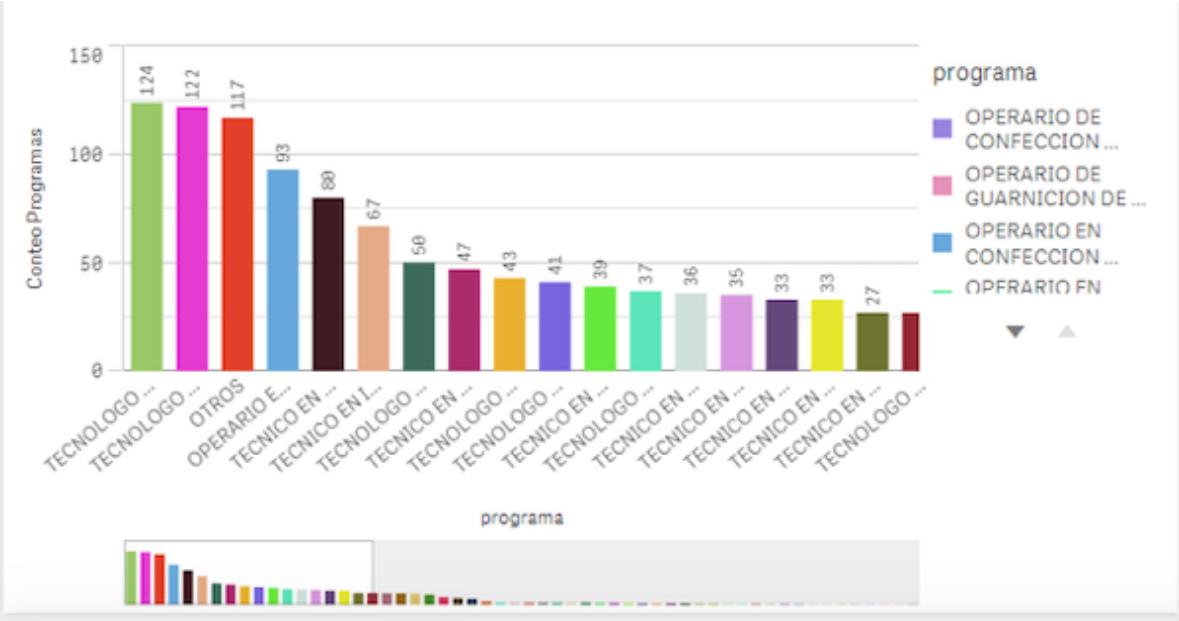
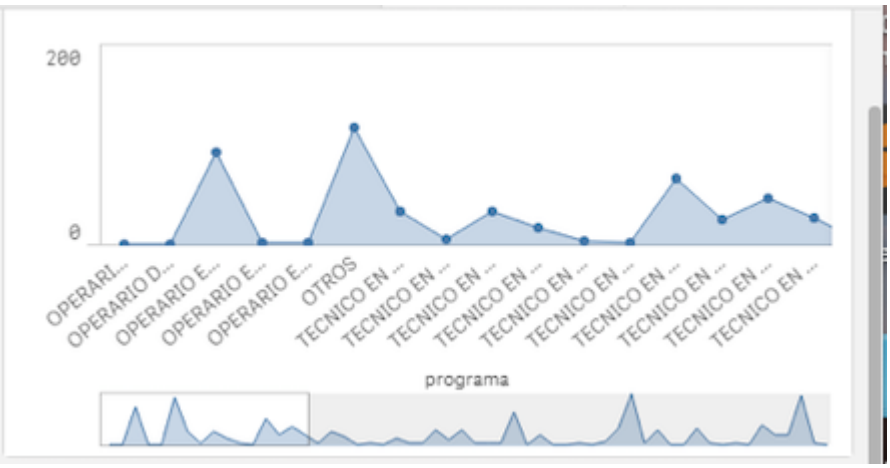


Figura 21: Desercion estudiantil por programa tecnicó SENA.



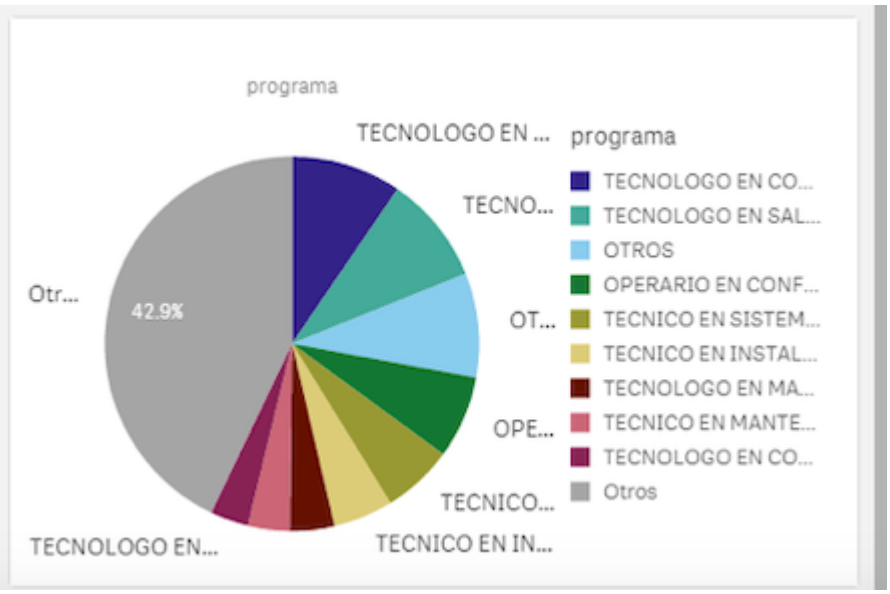


Figura 22: Deserción estudiantil por programa técnico SENA.

5.2 EVALUACIÓN

La evaluación dentro de esta nueva metodología queda enfocada al manejo de la información ya transformada desde la base de datos a considerar mediante software de inteligencia de negocios.

Lo importante en esta forma de valoración de los datos es que es flexible y se acomoda a los parámetros para analizar el fenómeno de la deserción de aprendices. En este sentido los datos a verificar son los concernientes a los que confirman posibilidades de abandono de los estudios de los aprendices. Por lo tanto para la educación estos indicadores podrian tener varias variables que afectan y es alli donde hay que detener el análisis para emitir los resultados con su debida interpretación.

Para garantizar la rigurosidad de la metodología es menester hacer un proceso de evaluación que implique pasar por los siguientes procedimientos:

- Evaluar los resultados que resultaron de la aplicación de la metodología que deben atender a los objetivos establecidos en relación con la deserción de los aprendices. Se debe garantizar que los datos obtenidos de cada aprendiz se realicen de forma adecuada y si los resultados obtenidos corresponde a tal

proceso de evaluación. En general se establecen reglas claras que deben ser verificadas en el momento de aplicar las pruebas para extraer los resultados.

- Como un segundo componente, es importante que en éstas últimas fases se haga un análisis detallado de los resultados en materia de las variables más generales y especificas que estan vinculadas con la deserción y la permanencia académica. Detenerse a aglomerar las que son más proclives a la deserción, determinar casos especificos por carreras técnicas y/o tecnólogica y establecer la relación con el fenómeno del abandono escolar. Es decir, que los resultados deben ser valorados a la luz de los factores que más generan esta situación de fracaso escolar y, si fuere necesario, las dimensiones académicas donde más se presentan.
- A partir del análisis de los factores y variables que afectan la permanencia de los aprendices dentro de esta metodología, los resultados deben deben sugerir pautas y decisiones dentro de un plan, a fin de mejorar la permanencia académica y garantizar las condiciones necesarias para que quienes inicien un proyecto académico lo puedan culminar pese a las adversidades que se presenten. De ahí que los procesos posteriores son tomar decisiones y encausar los medios posibles, asi sean externos, para ayudar a la culminación de una carrera técnica o tecnológica.

5.3 ORIENTACIÓN DE TOMA DE DECISIONES

En el desarrollo técnico de los resultados, a cada caso que es objeto de la muestra y de acuerdo a cada característica en especial se construye una fase de apoyo a la toma de decisiones para orientar en relación de los factores de riesgos de deserción que puedan comprometer a corto, mediano y largo plazo la permanencia de los aprendices de las diferentes carreras de formación del Sena Centro Industrial y de Aviación. En este sentido el trabajo no se estanca en solo detectar las variables, sino que va más allá al arrojar posibles decisiones para la dependencia responsable de determinar orientaciones pertinentes y asertivas hacia el aprendiz.

En este sentido, de acuerdo a los procesos internos en ésta institución del estado las intervenciones a los aprendices se realizan desde la oficina de “Bienestar al aprendiz”, lo cual, brindan un apoyo psicosocial a los aprendices con un equipo interdisciplinario compuesto por psicólogos y trabajadores sociales. Desde ésta dependencia se lideran proyectos de complemento a la formación y seguimiento a aprendices.

Todo lo anterior con el fin de coadyuvar a los procesos de formación de manera paralela, es así como muchos aprendices con dificultades desde lo personal reciben ayuda y por ende se mejora la estabilidad de los estudiantes. Para el trabajo en cuestión, se han clasificado varias categorías de insumo en la toma de decisiones, así: orientación psicológica, visita domiciliaria, contacto telefónico, talleres motivacionales (hábitos, técnicas de estudio y fortalecimiento de personalidad), apoyo de sostenimiento, apoyo académico y tutorías docentes.

Este proceso de toma de decisiones es de gran beneficio para los estudiantes en la medida en que la institucion educativa realice un proceso sincronizado desde el inicio del ingreso con un debido seguimiento sobre las alertas tempranas para tomar las decisiones pertinentes y poder asi brindarle el apoyo adecuado en lo aspectos que estan afectando al aprendiz para tener una estabilidad y de este modo culminar con éxito su proceso de desempeño en la educacion de carácter superior.

A continuación, una tabla que direcciona la toma de decisiones al frente de los factores de riesgo.

Factores de riesgo	Toma de decisiones
Desertor de la formación	Visita domiciliaria
No asistió	Contacto telefónico
Falta de interés	Talleres motivacionales
Motivos laborales	Orientación psicológica
Problemas familiares	Orientación psicológica
Problemas de salud	Contacto telefónico
Problemas económicos	Apoyo de sostenimiento
Cancelado en Sofia	Apoyo administrativo
Incumplimiento condicional	Orientación psicológica

6. CONCLUSIONES

En el desarrollo de este trabajo de investigación se le da relevancia a la propuesta emergente para mejorar los procesos de análisis de datos a partir de diferentes herramientas y metodologías que sirven de soporte y apoyo a la labor de desempeño de los componentes de la formación, es por ello que se consideran las siguientes deducciones:

A. La metodologia de como la propuesta en este trabajo son pertinentes para el campo educativo dado que van mas allá de la mera extraccion datos y pues, tienen un alcance hasta las orientaciones y toma de decisiones que podrian servir para mejorar en lo indices de deserción educativa.

B. Se puede concluir que las metodologias de extraccion conocimientos con peculiaridades educativas tienen indicadores de deserción (factores de riesgo) que posibilitan la coherencia en la toma de decisiones, cuya decisiones son proporcionales a los factores de riesgos, posibilitan la prevencion de la desercion y un manejo a tiempo con las circunstancias que rodean a los aprendices.

C. Las metodologías de extracción de conocimiento son importantes en la medida que las instituciones manejen la información para valorar las condiciones psicosociales que podrían afectar de forma negativa el proceso de permanencia de los estudiantes en el transcurrir de sus estudios.

D. Con el analisis de los factores de riesgo y las variables expuestas como resultado de metodologias más vinculantes con la educación, como la que se propuso en este trabajo de investigación, se puede tener una vez analisis mas cercano y veraz sobre las situaciones que ponen en riesgo la permanencia de los educandos en las instituciones de formación superior.

E. Los estudios de deserción estudiantil tienen que ser complementados con apoyos de tipo académico y de psicorientación, además de la pertinencia

de las acciones para mejorar la estabilidad de los aprendices en el contexto donde recibe la formación con éxito y buen nivel de adaptación.

7. PERSPECTIVAS FUTURAS

Un futuro estudio sobre las metodologías de extracción de conocimientos orientadas a la deserción educativa sería muy significativo si se realiza como un enfoque de investigación aplicada dentro de un contexto de evaluación de la metodología propuesta en esta investigación. En este sentido el método de trabajo seria mixto en el que se coloque en funcionamiento la metodología para la valoración del riesgo de deserción de estudiantes de otras instituciones de formación superior.

Por otro lado, un trabajo con la propuesta de diseño de un software que integre toda la metodología y se mejoren los procedimientos, fusionando de forma integral todos los procesos y procedimientos desde la etapa inicial hasta el final del proceso, en este mismo aspecto se debe fortalecer los siguientes puntos:

1. La estadística con mayor enfoque en el ámbito educativo que posibilite un análisis y unas lecturas comparadas frente a las problemáticas que se estén presentando en el contexto de educación superior y que ésto represente un insumo importante para la toma de medidas por parte de las autoridades o directivas de las instituciones de educación superior.
2. Evolucionar la tabla de toma de decisiones con más opciones de recomendaciones frente a nuevos factores de deserción que aparezcan en el escenario de educación superior.
3. Una mayor sincronización de todo el proceso que involucra la prevención en la deserción universitaria, es decir, unos tiempos específicos de acción frente a cada momento del proceso para que sirva de apoyo técnico a los programas de permanencia educativa.

8. BIBLIOGRAFÍA

Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. Paper presented at the IADIS European Conf. Data Mining. Recuperado de trier.de/db/conf/iadis/dm2008.html#AzevedoS08

Aguirre, J., Valdovinos, R., Velázquez, J., Eleuterio, R., y Marcial, J. (2015). Análisis de deserción escolar con minería de datos. *Research in Computing Science 93*, 71-82.

Amaya, Y., Barrientos, E., & Heredia, D. (2015). Mining Techniques, 13(9), 3127–3134.

Anderson, L.; & Krathwohl (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. New York: Longman.

Anuies (2007). Retención y deserción en un grupo de Instituciones Mexicanas de Educación Superior. Biblioteca de la educación superior. México.

Araque. F., Roldan. C., Salguero. A. (2009). Factors influencing university drop out rates". Computers & Education. Vol. 53. pag. 563–574. España

Barragán, D. y Patiño, L. (2013). Elementos para la comprensión del fenómeno de la deserción universitaria en Colombia. Más allá de las mediciones. Cuadernos

Blanco Suárez, J. O. (2014). Educación superior, crédito educativo y equidad en Colombia 1992-2011. Diálogos de Saberes, 40, 139-155. DOI: 10.18041/0124-0021/dialogos.40.214

Beltrán, M. (2016). *Minería de Datos*. Benemérita Universidad Autónoma de Puebla. Puebla de Zaragoza, México.

Bhise, Ramesh. (2013). "Importance of Data Mining in Higher Education System". IOSR Journal of Humanities and Social Science. 6. 18-21. 10.9790/0837-0661821.

Braxton, J. M., Johnson, R. M. y Shaw-Sullivan, A. (1997), "Appraising Tinto´s theory of college student departure"; en J. C. Smart (Ed.), Higher Education Handbook of theory and research, Vol. 12, Agathon Press, NY, EE.UU.

Braxton, J. M., Milem, J. F. Y Sullivan, A. S. (2000) The influence of active learning onthe college student departure process: Toward a revision of Tinto's theory. *Journal of Higher Education*, 71(5)

Cabral, M. (2005). Deserción en la educación superior pública en República Dominicana. [Material en línea]. Disponible en: http://www.psicologiacientifica.com/bv/psicologia218-1-desercion-en-la-educacion-superior-publica-en-republica-domi.html [Consulta: 2012, Diciembre 12].

Castaño, E, Gallón, S., Gómez, K y Vásquez, J. (2006). Análisis de los factores asociados a la deserción y graduación estudiantil universitaria. Lecturas Económicas, 65, 9 – 36. Universidad de Antioquia, Medellín.

Conference on Enterprise data. Research in Higher Education Journal. New Zealand desertion based on data mining," advances and applications in statistical

Hurtado de Barrera, Jacqueline & Morales, Marcos. (2000). Metodología de la investigación holística.

Díaz, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios Pedagógicos XXXIV*, (2), 65-86.

Eccles, J., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., y Midgeley, C. (1983). Expectancies, values and academic behaviours. En H.J.T. Spence (Ed.), Achievement and achievement motivation. San Francisco: Freeman economic systems volume 2, issue 2, Argentina Education – a Case Study". ICEIS - 12th International

Ethington, C. A. (1990). A psychological model of student persistence. *Research in Higher Education* 31 (3): 266-269.

Fishbein, M., & Ajzen, I. (1975). Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, MA: Addison-Wesley.

Formia, S. & Lanzarini, L. (2012). *Evaluación de técnicas de Extracción de Conocimiento en Bases de Datos y su aplicación a la deserción de alumnos universitarios.* (tesis de grado). Universidad Nacional de la Plata. La plata, Buenos Aires.

García, J. (2017). Minería de datos: Modelos predictivos con RapidMiner. Reflexión

Giovagnoli, P. (2002). Determinantes de la deserción y graduación universitaria: una aplicación utilizando modelos de duración, Documento de Trabajo 37, Universidad Nacional de la Plata.

Gómez, V. y Celis, J. (2009). Crédito educativo, acciones afirmativas y equidad social en la educación superior en Colombia. Revista de Estudios Sociales, 30, 106-117. DOI: 10.7440/res33.2009.09

Hasperué, W.,(2014) Extracción de conocimiento en grandes bases de datos utilizando estrategias adaptativas. - 1a ed. - La Plata: EDULP,210 p.

Henriquez, C., y Escobar, R (2016) construcción de un modelo de alerta temprana para la detección de estudiantes en riesgo de deserción de la universidad metropolitana de ciencias de la educación vol. 21, núm. 71, pp. 1221-1248.

H. kuna, r. García, f. Villatoro. (2010). "Pattern discovery in university students Higher Higher Education System, IOSR Journal of Humanities and Social

ISECO SOFTWARE, QLIK SENSE, BI, BIGDATA, BUSINESS. Recuperate de https://www.isecosoftware.es/2016/11/15/que-es-qlik-sense/

Journal of Computer Science and Information Security (IJCSIS). Vol. junio.2010. pp. 121-130 Universidad de San Buenaventura. Cali. Colombia

Kovacic, Z. (2012). Predicting student success by mining enrolment data. Research in Higher Education Journal. New Zealand

Jain, N. y Srivastava, V. (2013). Data mining techniques: a survey paper. International Journal of Research in Engineering and Technology. Vol.2 (11)

Kotsiantis, S. Pierrakeas. C. and Pintelas. P. (2004). Preventing student dropout in distance learning systems using machine learning techniques. in Proc. Int. Conf. Knowl.-Based Intell. Inf.learning techniques." in Proc. Int. Conf. Knowl. -Based Intell. Inf. Eng. Syst.. Oxford. U.K.,3–5.

La Madriz, J. y Parra, J. (2016). "Competencias personales y profesionales aplicadas por gerentes bajo escenarios de incertidumbre económica", Negotium, 11(33), pp. 69-98.

La Madriz, jenniz(2016). factores que promueven la deserción del aula virtual Orbis. Revista Científica Ciencias Humanas, vol. 12, núm. 35, noviembre, 2016, pp. 18-40

Luo G, et al. (2008) The sphingolipid long-chain base-Pkh1/2-Ypk1/2 signaling pathway regulates eisosome assembly and turnover. J Biol Chem 283(16):10433-44

Melguizo, T., Sánchez, F. y Jaime, H. (2011). The association between financial aid availability and the college dropout rates in Colombia. Higher Education, 62, 231-247. DOI: 10.1007/s10734-010-9385-8

Ministerio de Educación Nacional (MEN). (2009). Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico, y elementos para su prevención. Disponible en: https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_desercion.pdf

Ministerio de Educacion Nacional (MEN). (2017). Educacion superior.Boletin N° 1. Disponible en: https://snies.mineducacion.gov.co/1778/articles-393225_boletin_dic_2017.pdf

Ministerio de Educación Nacional. (2017). *Ministerio de Educación Nacional.* Recuperado el 5 de julio de 2018, de Estadísticas Deserción y graduación 2015:

https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-357549_recurso_3.pdf

R.B Bhise, S.S Thorat, A.K Supekar. (2013). Importance of Data Mining in Predict Student's Retention: A Comparative Study. International

Reyes, L., Castañeda, E., & Pabón, D. (2012). Causas psicosociales de la deserción universitaria. Revista Logos Ciencia & Tecnología. 4(1), 164-168. Recuperado de http://connection.ebscohost.com/c/articles/87411278/causas-psicosociales-de-ladeserci-n-universitaria.

Sanchez, F., Quiroz, M.,Reveron, C. y Rodriguez, A. (2002). "Equidad social en el acceso y permanencia en la universidad pública determinantes y factores asociados". Bogotá: Cede.

SAS Enterprise Miner – SEMMA. SAS Institute. Accessed from http://www.sas.com/technologies/analytics/datamining/miner/semma.html, on May 2008

Romero. C. and Ventura. S. 2007. Educational Data Mining: A Survey from 1995 to 2005. Expert Systems with Applications Volume, Issuse 1, July 2007, Pages 135-146.

Sanz, Malagón, Quintero, Vélez, & Parra. (2010). Identificar y realizar un análisis de los factores asociados a la permanencia y deserción escolar de las instituciones educativas oficiales del país. Bogotá: Universidad Nacional de Colombia.

Servicio nacional de aprendizaje- SENA. (2020). Interventoría técnica, administrativa, financiera, jurídica, contable y ambiental para realizar los estudios, diseños, licenciamiento y construcción de cubiertas, graderías y obras complementarias para las canchas múltiples de diferentes regionales y centros de formación del SENA.

Servicio nacional de aprendizaje- SENA. (2019). Reglamento del aprendiz del servicio nacional de aprendizaje. Dirección de formación profesional. Bogotá.

S. K. Yadav. B. Bharadwaj and S. Pal. (2012). Mining Educational Data to

Sapdy, W. (1970). Dropouts from Higher Education: An interdisciplinary review and synthesis interchance sciences proceedings of the iv meeting on dynamics of social and Science , 1, p. 64-85

SPADIES (2015) Deserción estudiantil. Recuperado de: http://www.sapiencia.gov.co/wp-content/uploads/2017/07/BOLETIN_ODES_DESERCION_EN_LA_EDUCACION_SUPERIOR.pdf

Tapasco-Alzate, O. A., Ruiz-Ortega, F. J., Osorio-García, D. y Ramírez-Ramírez, D. (2019). Deserción estudiantil: incidencia de factores institucionales relacionados con los procesos de admisión. Educación y Educadores 22(1), 81-100. DOI: 10.5294/edu.2019.22.1.5

Tillman, C.A. Sr. (2002). Barriers to student persistence in higher education. Retrieved October 3, 2002 from: https://www.whdl.org/sites/default/files/v2n1_Tillman.pdf

Timarán, S., Hernández, I., Caicedo, S., Hidalgo, A. y Alvarado, J. (2016). El proceso de descubrimiento de conocimiento en bases de datos*. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional* (pp. 63-86).

Timarán-Pereira, S. R., Hernández-Arteaga, I., Caicedo-Zambrano, S. J., Hidalgo-Troya, y Alvarado Pérez, J. C. (2016). El proceso de descubrimiento de conocimiento en bases de datos. En Descubrimiento de patrones de desempeño académico con árboles de decisión en las competencias genéricas de la formación profesional (pp. 63-86). Bogotá: Ediciones Universidad Cooperativa de Colombia. doi: http://dx.doi.org/10.16925/9789587600490

Tinto, V. (1975). Dropout from Higher Education: A theoretical synthesis of recent research. Review of Educational Research 45: 89-125.

Tinto (1989) Una reconsideración de las teorias de la deserción estudiantil. En trayectoria escolar en la educación superior. ANUIES-SEP. México

Van Dalen, D., y Meyer, W. (2006): Manual de técnica de la Investigación educacional. Paidós. México

Vasquez, C., y Rodriguez, M. (2007). La deserción estudiantil en educación superior a distancia: Perspectiva teórica y factores de incidencia. Revista Latinoamericana de Estudios Educativos, Vol. XXXVII, Núm. 3-4, sin mes, 2007, Pp. 107-122. México.

Vélez, A. y López, D. (2004). Estrategias para vencer la deserción universitaria. *Revista Educación y Educadores*, 7, 177-203.

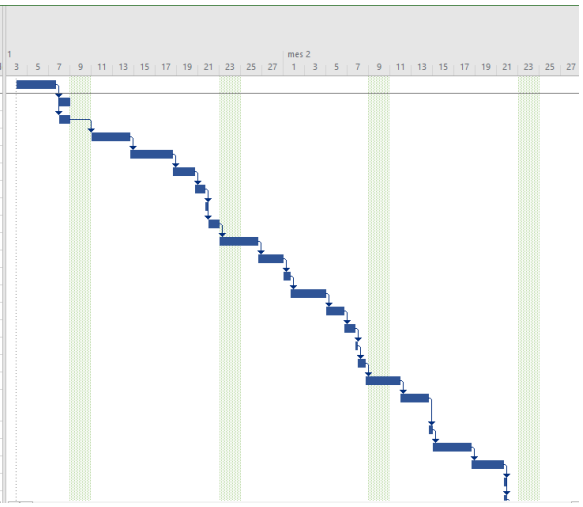
Velickov, S. y Solomatine, D. (2000). Predictive Data Mining: Practical Examples. *Artificial Intelligence in Civil Engineering. Proc. 2nd Joint*

Zhang, S. Oussena. “Use Data Mining to Improve Student Retention in Zlatko J. Kovacic. (2010). Predicting student success by mining.

9. ANEXOS

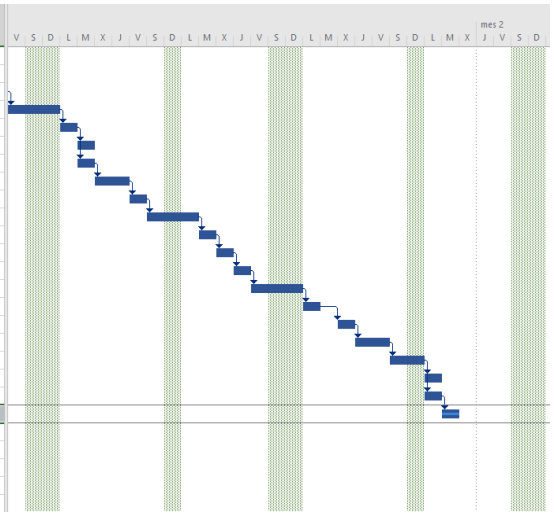
Anexo 1: Plan de proyecto.

Nombre de tarea	Duración	Comienzo	Fin	Pred
Introduccion	5 días?	lun 4/02/19	jue 7/02/19	
Reunion Director Proyecto	1 día?	vie 8/02/19	vie 8/02/19	1
Reunion Asesor Metodologico	1 día?	vie 8/02/19	vie 8/02/19	1
Correciones a Realizar	5 días?	lun 11/02/19	jue 14/02/19	3
Justificacion Problema	4 días?	jue 14/02/19	lun 18/02/19	4
Objetivo General	3 días?	lun 18/02/19	mié 20/02/19	5
Objetivo Especificos	2 días?	mié 20/02/19	jue 21/02/19	6
Reunion Asesor Tecnico	1 día?	jue 21/02/19	vie 22/02/19	7
Reunion Asesor Metodologico	1 día?	vie 22/02/19	vie 22/02/19	7
Correciones a Realizar	3 días?	sáb 23/02/19	mar 26/02/19	9
Revision de la Literatura	4 días?	mar 26/02/19	vie 1/03/19	10
Reunion Asesor Tecnico	1 día?	vie 1/03/19	vie 1/03/19	11
Correciones a Realizar	3 días?	vie 1/03/19	mar 5/03/19	12
Panorama Educativo	2 días?	mar 5/03/19	mié 6/03/19	13
Problema Desercion Estudiantil	2 días?	mié 6/03/19	jue 7/03/19	14
Spadies	1 día?	jue 7/03/19	vie 8/03/19	15
Reunion Asesor Tecnico	1 día?	vie 8/03/19	vie 8/03/19	16
Correciones a Realizar	2 días?	vie 8/03/19	lun 11/03/19	17
Metodologia Extracion de Conocimientos	4 días?	mar 12/03/19	jue 14/03/19	18
Reunion Asesor Tecnico	1 día?	jue 14/03/19	vie 15/03/19	19
Correciones a Realizar	3 días?	vie 15/03/19	lun 18/03/19	20
Mineria de Datos	5 días?	lun 18/03/19	jue 21/03/19	21
Reunion Asesor Tecnico	1 día?	jue 21/03/19	vie 22/03/19	22
Reunion Asesor Metodologico	1 día?	jue 21/03/19	vie 22/03/19	22



Anexo : 2 Plan de proyecto II.

Nombre de tarea	Duración	Comienzo	Fin
Aplicación Metodologia Crisp-DM	3 días?	lun 1/07/19	mar 2/07/19
Comprencion del Negocio	2 días?	mar 2/07/19	mar 2/07/19
Determinar Objetivos Negocio	2 días?	jue 4/07/19	vie 5/07/19
Recoleccion de los Datos	3 días?	vie 5/07/19	lun 8/07/19
Descripcion de los Datos	2 días?	lun 8/07/19	mar 9/07/19
Reunion Asesor Tecnico	1 día?	mar 9/07/19	mar 9/07/19
Reunion Asesor Metodologico	1 día?	mar 9/07/19	mar 9/07/19
Correciones a Realizar	5 días?	mié 10/07/19	jue 11/07/19
Verificacion de los Datos	3 días?	vie 12/07/19	sáb 13/07/19
Transformacion de los Datos (PENTAHO)	6 días?	sáb 13/07/19	mar 16/07/19
Reunion Asesor Tecnico	1 día	mar 16/07/19	mar 16/07/19
Correciones a Realizar	2 días?	mié 17/07/19	jue 18/07/19
Modelado (selección de Tecnica)	2 días?	jue 18/07/19	vie 19/07/19
Modelado (Generacion Tecnica)	4 días?	vie 19/07/19	lun 22/07/19
Evaluacion del Modelado	2 días?	lun 22/07/19	mar 23/07/19
Correciones a Realizar	2 días?	mié 24/07/19	jue 25/07/19
Modelado Metodologia Semma	5 días?	jue 25/07/19	vie 26/07/19
Valoracion Metodologia semma	2 días?	sáb 27/07/19	lun 29/07/19
Reunion Asesor Tecnico	1 día?	lun 29/07/19	lun 29/07/19
Reunion Asesor Medologico	1 día?	lun 29/07/19	lun 29/07/19
Correciones a Realizar	2 días?	mar 30/07/19	mié 31/07/19



Anexo 3: Informacion desercion en el año 2015 en el centro industrial y de aviacion.

3	NOMBRES	APELLIDOS	DOCUMENTO	FORMACION	ACTA	
4	ALFREDO	MENDOZA MARTINEZ	1143116584	TGO. EN MANTENIMIENTO MECATRONICO DE AUTOMOTORES	77	DESERCION
5	HEIMY PATRICIA	VALENCIA TOLEDO	1007117641	TECNOLOGO EN CONSTRUCCION	91	Se dedico a tros estudios
6	VIVIANA PATRICIA	BUZON MACIAS	1140849010	TECNOLOGO EN CONSTRUCCION	91	Se dedico a otros estudios
7	JEISON RICARDO	PEREZ RIVERA	1042348194	TECNOLOGO EN ENTRENAMIENTO DEPORTIVO	44	Problemas familiares y personales
8	MARIA ALEJANDRA	RAMIREZ BARRIOS	1140847714	TECNOLOGO CONFECCION INDUSTRIAL	25	Por Motivos de la enfermedad del esposo.
9	GISELL CAROLINA	PEREZ RIVERA	1043127685	FALSO	91	falta de interes en el curso
10	DONALDO RAFAEL	SANTANA HERRERA	96123021069	FALSO	91	falta de interes en el curso.
11	MILAGRO PAOLA	TRON HERRERA	1042445969	TECNOLOGO EN CONFECCION INDUSTRIAL	25	Por problemas economicos.
12	CINDY PAOLA	DE LA ROSA GONZALEZ	1143243775	TECNOLOGO EN SALUD OCUPACIONAL	40	DESERTO POR FALTA DE INTERES EN EL CURSO
13	KEVIN ANDRES	DIAZ ROBLES	1143255975	TECNOLOGO EN CONSTRUCCION	91	Falta de Interes en el curso
14	JOSE IVAN	DORIA MEZA	1082881003	TECNOLOGO EN CONSTRUCCION	91	falta de interes en la formacion
15	EDINSON ANDRES	FALSO	97091215725	TECNOLOGO EN CONSTRUCCION	91	Falta de interes en el curso
16	PABLO RAFAEL	CAJAR ROMERO	1143141215	TECNOLOGO EN CONSTRUCCION	91	Problemas Familiares
17	FRAIS GABRIEL	ORTIZ VELAZQUEZ	1143134884	TECNOLOGO EN CONSTRUCCION	91	Problemas laborales
18	BRAYAN RAFAEL	ROMERO ARIAS	96120525805	TECNOLOGO EN DESARROLLO GRAFICO DE PROYECTOS DE ARQUITECTURA E INGENIERIA	16	el aprendiz abandono la formaci?n para dedicarse a otras actividades.
19	DIEGO ANDRES	MARTINEZ BARRIOS	1140833564	TECNOLOGO EN CONSTRUCCION	73	abandono de la formacion
20	RICARDO JESUS	BABILONIA HERNANDEZ	72162072	TECNICO EN MANTENIMIENTO DE MAQUINAS DE CONFECCION INDUSTRIAL	76	DESERCION.
21	OSWALDO RAFAEL	BOLIVAR ARCIERI	72208087	TECNICO EN MANTENIMIENTO DE MAQUINAS DE CONFECCION INDUSTRIAL	76	DESERCION.
22	LUIS ALBERTO	CARRILLO DONADO	1042444268	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	16	el aprendiz abandono la formaci?n por dificultades laborales, se cruzaban los horarios con las clases
23	JORGE LUIS	VILLAR RAMOS	1045702034	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	91	Por bajo rendimiento intelectual
24	JHON ALEXANDER	PACHECO ALVARADO	97070217701	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	91	Problemas de domicilio
25	MAYKEL DAVID	SANCHEZ RODRIGUEZ	1045676129	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	91	Bajo rendimiento intectual
26	DAYNER JESUS	SANCHEZ GARCIA	1139614028	TECNOLOGO EN CONSTRUCCION	16	EL APRENDIZ ABANDONO LA FORMACION
27	JOESID	AGAMEZ RUZ	1046404492	TECNOLOGO EN CONSTRUCCION	16	EL APRENDIZ ABANDONO LA FORMACION
28	ELBERT JAVIER	BECERRA NIEVES	1101546078	TECNOLOGO EN CONSTRUCCION	18	se inicia la reunion revisado los formatos de desercion enviados por los instrutores el instrutor catalino sua
29	SAMIR ERNESTO	FALSO	1048210515	TECNOLOGO EN CONSTRUCCION	16	abandono la formaci?n y no se conocen las causas.
30	MIGUEL ANGEL	PULIDO GONZALEZ	1042444226	TECNOLOGO EN CONSTRUCCION	73	abandono de la formacion
31	DARLING ISABEL	FERRER PERALTA	55221411	FALSO	69	Abandono la formacion
32	EDILMA JUDITH	ROYERO CORTINA	39068037	FALSO	69	Abandono la formacion.
33	MAGALYS ESTHER	GUTIERREZ SALAZAR	32610122	FALSO	69	Abandono la formacion.
34	OLGA LUCIA	BETANCOURT SUAREZ	32777570	FALSO	69	Abandono la formacion.
35	RUBEN DARIO	ARGEL PORTOCARRERO	72004208	FALSO	69	Abandono la formacion.
36	ANDRES FELIPE	MIRANDA AMARIS	1045736840	TECNICO EN IMPRESION OFFSET	16	aprendiz abandono la formaci?n y no se conocen las causas.
37	SANDRO JUNIOR	GARCIA FERRER	1045731395	TECNOLOGO EN MANTENIMIENTO MECATRONICO DE AUTOMOTORES	71	DESERCION.
38	YULBER LEYDE	CASTELLAR BLANCO	1045725878	TECNOLOGO EN MANTENIMIENTO MECATRONICO DE AUTOMOTORES	72	Desercion

Desercion2015 +

Anexo 4: Informacion desercion en el año 2016 en el centro industrial y de aviacion.

3	NOMBRES	APELLIDOS	DOCUMENTO	FORMACION	ACTA	
4	KELLYS SALLURYS	VARGAS CHARRIS	1143436749	TECNOLOGO EN SALUD OCUPACIONAL	418	LA APRENDIZ DESERTO DE LA FORMACI?N, SE LE BRINDO LA OPORTUNIDAD PARA APLAZAR Y REINTEGRAR PERO NO REALIZO EL I
5	MARIA ALEJANDRA	MELENDREZ MIRANDA	1143141336	TECNOLOGO EN SALUD OCUPACIONAL	418	Aprendiz Condicionada, no cumpli? los planes de mejoramiento establecido, abandono proceso psicol?gico. Se le brindaron c
6	MARTHA CECILIA	FALSO	55308903	TECNOLOGO EN SALUD OCUPACIONAL	418	DESERTADA. DEJO DE ASISTIR DESDE. EL PRINCIPIO DE LA FORMACION. 11 DE JULIO DEL 2016.
7	OMAR YARID	GUTIERREZ MONTESINO	1143141743	TECNOLOGO EN SALUD OCUPACIONAL	418	APRENDIZ SE RECOMIENDA DESERCI?N, DEJO DE ASISTIR A LA FORMACI?N. FECHA DESERCI?N: 18 DE FEBRERO DEL 2016.
8	MARIA CONCEPCION	VENGOECHEA MONTERO	1047349254	TECNOLOGO EN MANTENIMIENTO MECATRONICO	562	desercion no asiste desde el 17. de Noviembre de 2016, por problemas familiares
9	JORGE ALIRIO	DELGADO VARGAS	1001940984	TECNOLOGO EN SALUD OCUPACIONAL	504	DESERTADO DE LA FORMACION
10	EXEQUIEL JULIO	RODRIGUEZ ROJAS	72433669	TECNICO EN MANTENIMIENTO DE MOTORES GASOLINA Y GAS	565	BAJO RENDIMIENTO INTELECTUAL Y/O PRACTICO
11	CARLOS ALBERTO	HOYOS IMITOLA	1002026157	TECNICO EN MANTENIMIENTO DE MOTORES GASOLINA Y GAS	565	BAJO RENDIMIENTO INTELECTUAL Y/O PRACTICO
12	MARIO ANDRES	MENDOZA PADILLA	1140845251	TECNICO EN MANTENIMIENTO DE MOTORES GASOLINA Y GAS	565	BAJO RENDIMIENTO INTELECTUAL Y/O PRACTICO
13	LUIS ALFONSO	DOMINGUEZ GALLARDO	1007372476	TECNICO EN MANTENIMIENTO DE MOTORES GASOLINA Y GAS	565	BAJO RENDIMIENTO INTELECTUAL Y/O PRACTICO
14	DAVID ALFONSO	GUTIERREZ URUETA	99012211143	TECNICO EN MANTENIMIENTO DE MOTORES GASOLINA Y GAS	565	BAJO RENDIMIENTO INTELECTUAL Y/O PRACTICO
15	DANIELA LUCIA	DURAN JIMENEZ	1001873803	TECNOLOGO EN SALUD OCUPACIONAL	471	DESERTADA DE LA FORMACION; DEJO DE ASISTIR DESDE EL PRIMER TRIMESTRE.
16	JORGE LUIS	FIGUEROA DIAZ	72348860	TECNOLOGO EN SALUD OCUPACIONAL	471	EL APRENDIZ DESERTO DE LA FORMACION, DEJO DE ASISTIR DESDE EL MES DE JULIO. CAUSA DE LA DESERCION LA VOCERA EXPR
17	CAROLINA	DE LA ROSA REYES	1140863477	TECNOLOGO EN SALUD OCUPACIONAL	405	El aprendiz deserto del proceso de formaci?n. Solo se present? 1 vez. Causa deserci?n: 01- falta de inter?s en el curso MEDIDA: I
18	INIRIDA	SOTO ORTEGA	1043437745	TECNOLOGO EN SALUD OCUPACIONAL	405	El aprendiz deserto del proceso de formaci?n. El ?ltimo trimestre no asisti? sin justificaci?n. MEDIDA: DESERCI?N
19	JANETH	FALSO	55242165	TECNOLOGO EN SALUD OCUPACIONAL	405	El aprendiz deserto del proceso de formaci?n. No se present? a la formaci?n Causa deserci?n: 01- falta de inter?s en el curso ME
20	MAILETH MARIA	MARTINEZ BERMUDEZ	1143160105	TECNOLOGO EN SALUD OCUPACIONAL	405	El aprendiz deserto del proceso de formaci?n. No se present? a la formaci?n Causa deserci?n: 01- falta de inter?s en el curso ME
21	ABEL SEGUNDO	WILCHEZ BLANCO	1047216652	TECNOLOGO EN CONSTRUCCION	519	DESERTADO Y REGISTRADO EN EL APLICATIVO
22	DANIELA	FALSO	1140875721	TECNOLOGO EN CONSTRUCCION	519	DESERTADO Y REGISTRADO EN EL APLICATIVO
23	KATIA YOLANIS	ESTRADA VILLA	1002293928	TECNOLOGO EN CONSTRUCCION	519	DESERTADO Y REGISTRADO EN EL APLICATIVO
24	PABLO FRANCISCO	RODRIGUEZ BENITEZ	4978922	TECNOLOGO EN CONSTRUCCION	519	DESERTADO Y REGISTRADO EN EL APLICATIVO
25	VALENTINA	BUZON MORENO	55304271	TECNOLOGO EN CONFECCIÓN INDUSTRIAL	498	EL APRENDIZ ABANDONO LA FORMACION
26	ORIOLA ESTHER	BARRERA SOLAR	22444478	TECNOLOGO EN CONFECCIÓN INDUSTRIAL	498	EL APRENDIZ ABANDONO LA FORMACION POR PROBLEMAS DE SALUD
27	KEYLA	PINILLA TEHERAN	1143261110	TECNOLOGO EN CONFECCIÓN INDUSTRIAL	565	PROBLEMAS ECONOMICOS
28	DEYVIS JOSE	ROJAS VALENCIA	72336574	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	483	EL APRENDIZ ABANDONO LA FORMACION POR PROBLEMAS DE TRABAJO.
29	JESUS DAVID	ESCALANTE MARTINEZ	1043933409	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	483	EL APRENDIZ ABANDONO LA FORMACION POR FALTA DE INTERES EN EL CURSO.
30	ELMER JESUS	GONZALEZ ENCISO	1143120746	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	443	EL APRENDIZ ABANDONO POR PROBLEMAS LABORALES.
31	CESAR ALBERTO	ARAGON ORTIZ	1080013511	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	443	EL APRENDIZ ABANDONO POR PROBLEMAS LABORALES.
32	LUIS FERNANDO	ESTRADA PEREZ	1042454009	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	399	DESERTADO
33	DANIEL ALBERTO	TERAN VILORIA	1045743756	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	399	DESERTADO
34	JORDAN DAVID	FRANCIA QUINTERO	1143158996	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	399	DESERTADO
35	NESTOR ORLANDO	TORRES MONTESINO	1143460913	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	399	DESERTADO
36	JUAN CARLOS	HOYOS GUERRERO	1007909063	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	399	DESERTADO
37	LUIS CARLOS	CARRILLO ALVAREZ	98101050181	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	399	DESERTADO
38	ELA JUDITH	ANGULO MARCHENA	32779300	TECNICO EN PRODUCCION DE MARROQUINERIA	570	Por enfermedad (Pendiente por entregar)

Desercion +

Anexo 5: Informacion desercion en el año 2017 en el centro industrial y de aviacion.

NOMBRES	APELLIDOS	DOCUMENTO	FORMACION	
ALFREDO	MENDOZA MARTINEZ	1143116584	TGO. EN MANTENIMIENTO MECATRONICO DE AUTOMOTORES	Se dedico a tros estudios
HEIMY PATRICIA	VALENCIA TOLEDO	1007117641	TECNOLOGO EN CONSTRUCCION	Se dedico a tros estudios
VIVIANA PATRICIA	BUZON MACIAS	1140849010	TECNOLOGO EN CONSTRUCCION	Se dedico a otros estudios
JEISON RICARDO	PEREZ RIVERA	1042348194	TECNOLOGO EN ENTRENAMIENTO DEPORTIVO	Problemas familiares y personales
MARIA ALEJANDRA	RAMIREZ BARRIOS	1140847714	TECNOLOGO CONFECCION INDUSTRIAL	Por Motivos de la enfermedad del esposo.
GISELL CAROLINA	PEREZ RIVERA	1043127685	TECNOLOGO CONFECCION INDUSTRIAL	falta de interes en el curso
DONALDO RAFAEL	SANTANA HERRERA	96123021069	TECNOLOGO CONFECCION INDUSTRIAL	falta de interes en el curso.
MILAGRO PAOLA	TRON HERRERA	1042445969	TECNOLOGO EN CONFECCION INDUSTRIAL	Por problemas economicos.
CINDY PAOLA	DE LA ROSA GONZALEZ	1143243775	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO POR FALTA DE INTERES EN EL CURSO
KEVIN ANDRES	DIAZ ROBLES	1143255975	TECNOLOGO EN CONSTRUCCION	Falta de Interes en el curso
JOSE IVAN	DORIA MEZA	1082881003	TECNOLOGO EN CONSTRUCCION	falta de interes en la formacion
EDINSON ANDRES	FALSO	97091215725	TECNOLOGO EN CONSTRUCCION	Falta de interes en el curso
PABLO RAFAEL	CAJAR ROMERO	1143141215	TECNOLOGO EN CONSTRUCCION	Problemas Familiares
FRAIS GABRIEL	ORTIZ VELAZQUEZ	1143134884	TECNOLOGO EN CONSTRUCCION	Problemas laborales
BRAYAN RAFAEL	ROMERO ARIAS	96120525805	TECNOLOGO EN DESARROLLO GRAFICO DE PROYECTOS DE ARQUITECTURA E INGENIERI	el aprendiz abandono la formaci?n para dedicarse a otras actividades.
DIEGO ANDRES	MARTINEZ BARRIOS	1140833564	TECNOLOGO EN CONSTRUCCION	abandono de la formacion
RICARDO JESUS	BABILONIA HERNANDEZ	72162072	TECNICO EN MANTENIMIENTO DE MAQUINAS DE CONFECCION INDUSTRIAL	DESERCION.
OSWALDO RAFAEL	BOLIVAR ARCIERI	72208087	TECNICO EN MANTENIMIENTO DE MAQUINAS DE CONFECCION INDUSTRIAL	DESERCION.
LUIS ALBERTO	CARRILLO DONADO	1042444268	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	el aprendiz abandono la formaci?n por dificultades laborales, se cruzaban los horarios con las clases
JORGE LUIS	VILLAR RAMOS	1045702034	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	Por bajo rendimiento intelectual
JHON ALEXANDER	PACHECO ALVARADO	97070217701	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	Problemas de domicilio
MAYKEL DAVID	SANCHEZ RODRIGUEZ	1045676129	TECNICO EN INSTALACIONES ELECTRICAS RESIDENCIALES	Bajo rendimiento intectual
DAYNER JESUS	SANCHEZ GARCIA	1139614028	TECNOLOGO EN CONSTRUCCION	EL APRENDIZ ABANDONO LA FORMACION
JOESID	AGAMEZ RUZ	1046404492	TECNOLOGO EN CONSTRUCCION	EL APRENDIZ ABANDONO LA FORMACION
ELBERT JAVIER	BECERRA NIEVES	1101546078	TECNOLOGO EN CONSTRUCCION	se inicia la reunion revisado los formatos de desercion enviados por los instrutores el instrutor catalino suarez rep
SAMIR ERNESTO	FALSO	1048210515	TECNOLOGO EN CONSTRUCCION	abandono la formaci?n y no se conocen las causas.
MIGUEL ANGEL	PULIDO GONZALEZ	1042444226	TECNOLOGO EN CONSTRUCCION	abandono de la formacion
DARLING ISABEL	FERRER PERALTA	55221411	TECNOLOGO EN CONSTRUCCION	Abandono la formacion
EDILMA JUDITH	ROYERO CORTINA	39068037	TECNOLOGO EN CONSTRUCCION	Abandono la formacion.
MAGALYS ESTHER	GUTIERREZ SALAZAR	33610122	TECNOLOGO EN CONSTRUCCION	Abandono la formacion

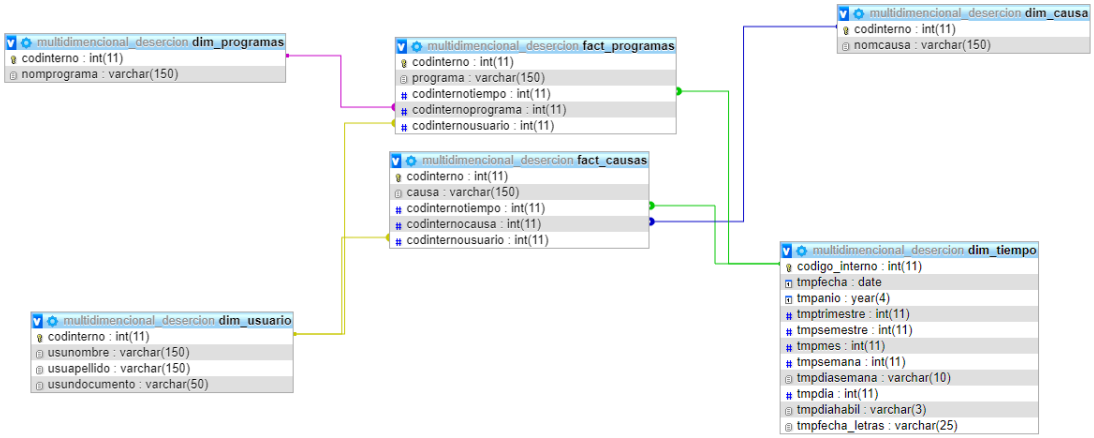
Desercion | TOTALES | Hoja38 | Hoja36 | Hoja35 | Hoja34 | Hoja33 | Hoja32 | Hoja31 | Hoja30 | Hoja29 | Hoja28 | Hoja27 | Hoja26

Anexo 6: Informacion desercion en el año 2018 en el centro industrial y de aviacion.

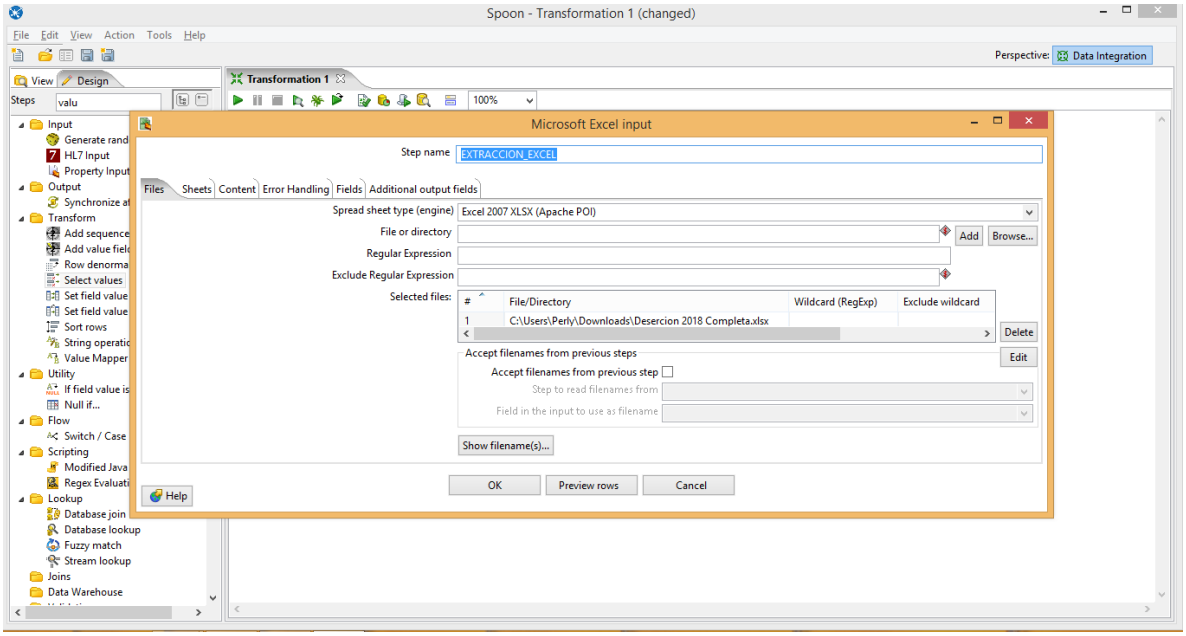
NOMBRES	APELLIDOS	DOCUMENTO	FORMACION	CAUSA
LUIS MIGUEL	URREGO SANABRIA	80902367	OPERARIO EN OPERACION DE EQUIPOS DE SERVICIO INTERMUNICIPAL DE PASAJERO	Desercion de la formacion
JUAN CARLOS	DELGADO ALVAREZ	72210582	OPERARIO EN OPERACION DE EQUIPOS DE SERVICIO INTERMUNICIPAL DE PASAJERO	No se present? al Sena despu?s del periodo de aplazamiento
NAYELYS MILENA	POLO CARRETERO	1043441638	TECNOLOGO EN CONFECCION INDUSTRIAL	Dersercion
CARMEN ROSA	DONADO DE LA ROSA	1042450108	TECNOLOGO EN CONFECCION INDUSTRIAL	Dersercion
FREDDY ANTONIO	RUIZ RESTREPO	1129490644	TECNOLOGO EN CONSTRUCCION	No se presento al sena despues del periodo de aplazamiento
ANTONY RICHARD	RICARDO CANTILLO	1143155188	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	No se present? al Sena despu?s del periodo de aplazamiento
JANN CARLOS	VIZCAINO CRESPO	1045728749	TECNICO EN MANTENIMIENTO DE EQUIPOS DE COMPUTO.	No se present? al Sena despu?s del periodo de aplazamiento
GISELLA CAROLINA	VARGAS ZAPATA	1140858007	FALSO	REALIZAR DESERCION
ANDERSON WILLIAM	CASTILLO GUTIERREZ	1010040370	TECNICO EN MANTENIMIENTO DE MOTORES GASOLINA Y GAS	No se present? al Sena despu?s del periodo de aplazamiento
JUAN CAMILO	ELJACH AVILA	98083173963	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	No se present? al Sena despu?s del periodo de aplazamiento
ALEX SAMIR	OSORIO GARCIA	1143159226	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	Realizar deserci?n en Sofia
CARLOS ANDRES	DE LAS SALAS PEDRAZA	1045727064	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	CANCELADO EN SOFIA
JOHN JAIRO	MAURY PIEDRAHITA	72253484	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	Realizar deserci?n en el aplicativo
JOSE ALFREDO	CAMPO SALAS	1045738325	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	Cancelado en el aplicativo
MAIKOL ANTONIO	DE LA CRUZ CARO	72099444	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	Cancelado en el aplicativo
NEYDER JAVIER	SIMANCA CASSERES	1001913123	TECNOLOGO EN PRODUCCION DE MULTIMEDIA	Cancelado en el aplicativo
EMPERATRIZ IRENE	HERAZO HOYOS	32845536	TECNICO EN PATRONAJE INDUSTRIAL DE PRENDAS DE VESTIR	No se present? al Sena despu?s del periodo de aplazamiento
LEIDY JOHANNA	MAYORGA OROSTEGUI	55312235	TECNOLOGO EN SALUD OCUPACIONAL	No se present? al Sena despu?s del periodo de aplazamiento
HOLLMAN DANIEL	NORIEGA DUNCAN	1143138735	TECNOLOGO EN SALUD OCUPACIONAL	DESERTADO
JEAN CARLOS	AGUDELO RANGEL	1007616158	TECNOLOGO EN SALUD OCUPACIONAL	DESERTADO
JHONATAN CARLOS	CASTRO FLOR	97012124680	TECNOLOGO EN SALUD OCUPACIONAL	DESERTADO
RAUL DE JESUS	GUERRA MENDOZA	1101814302	TECNOLOGO EN SALUD OCUPACIONAL	DESERTADO
ALFREDO JOSE	DOMINGUEZ DELGADO	98052159369	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO
JORGE ARMANDO JUNIOR	PINIILA AVILA	1044425755	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO
JUAN JOSE	GUTIERREZ CASTRO	1143156224	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO
KEVIN RAFAEL	MAESTRE MEJIA	1140877988	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO
KEVIN STHY	PEREZ CERVANTES	97100717984	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO
MIYER NEIL	GOMEZ OJEDA	1085280642	TECNOLOGO EN SALUD OCUPACIONAL	DESERTO
KENIA	DE LA OSSA AMARIS	52932709	TECNOLOGO EN SALUD OCUPACIONAL	No se present? al Sena despu?s del periodo de aplazamiento
SONIA EDITH	CANTILLO GUZMAN	32733860	TECNOLOGO EN DECORACION DE ESPACIOS INTERIORES	No se present? al Sena despu?s del periodo de aplazamiento

Desercion | Hoja36 | Hoja35 | Hoja34 | Hoja33 | Hoja31 | Hoja32 | Hoja30 | Hoja29 | Hoja28 | Hoja27 | Hoja26 | Hoja25 | Hoja24

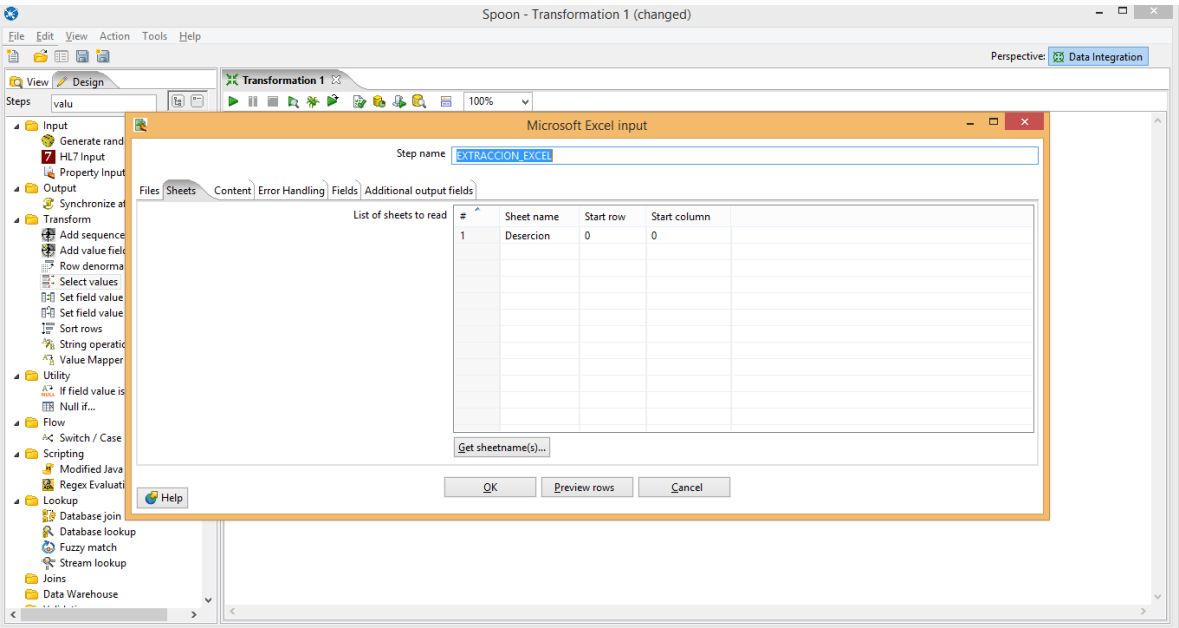
Anexo 7: Base de datos multidimencional.



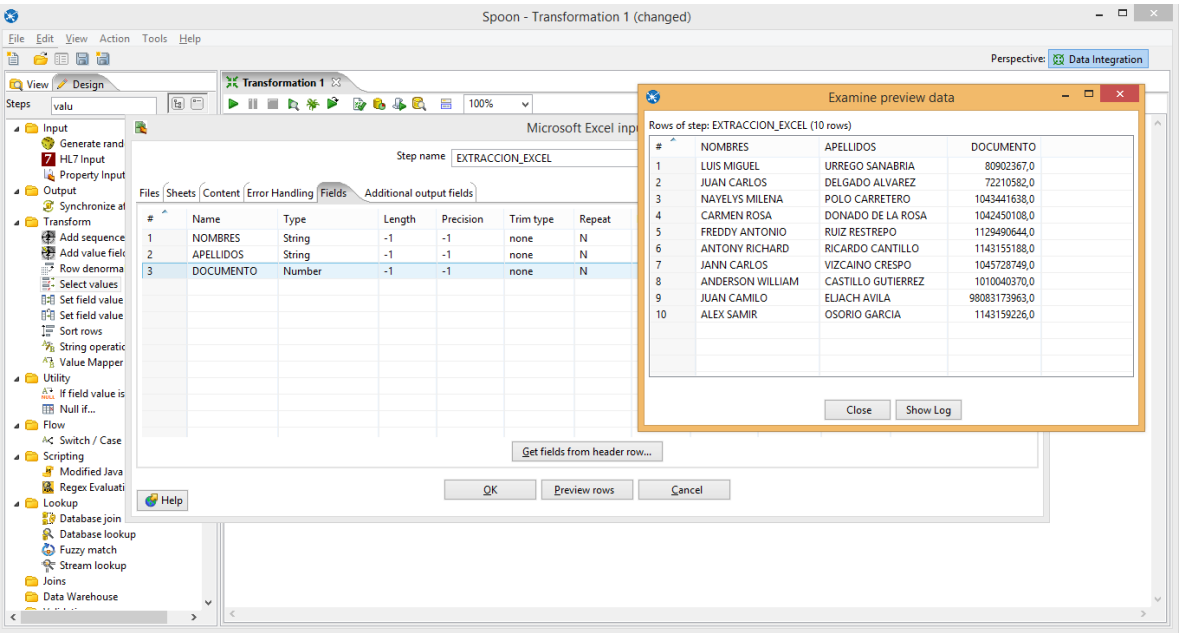
Anexo 8: Proceso de creación de la dimensión Usuario.



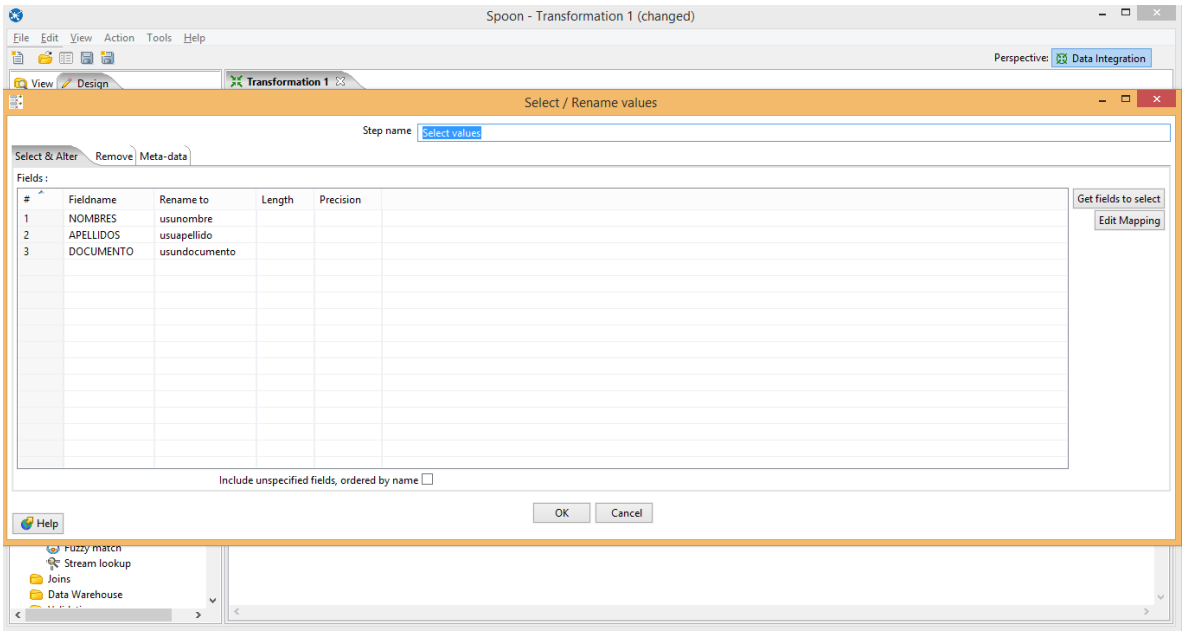
Anexo 9: la extracción y la posición de lectura.



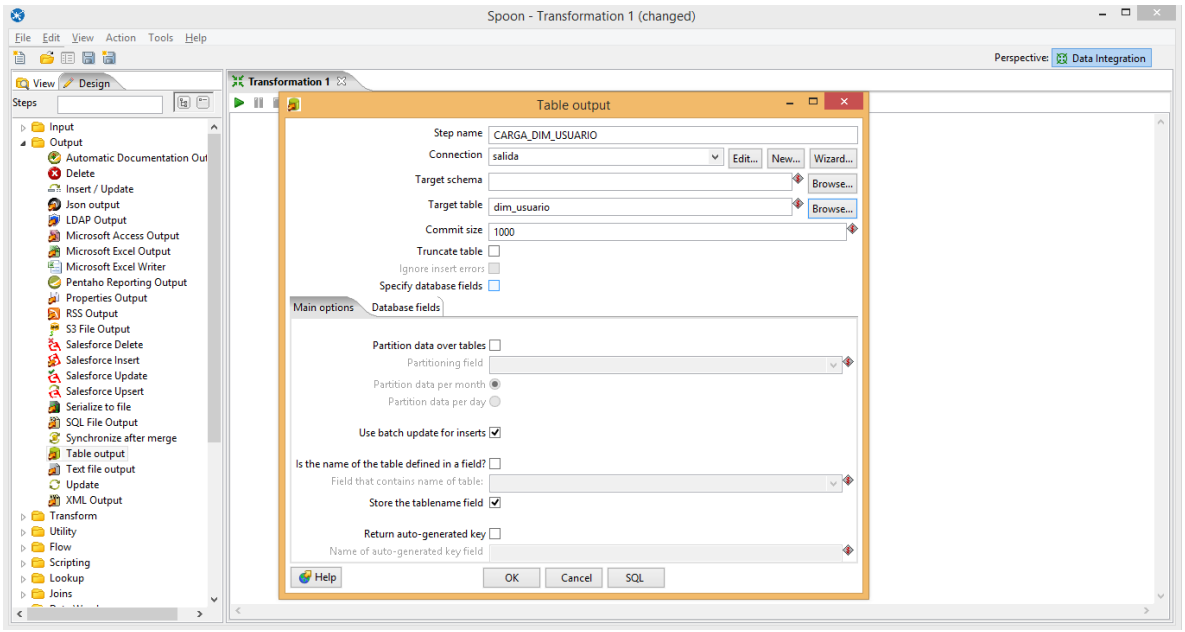
Anexo 10: muestra de los datos.



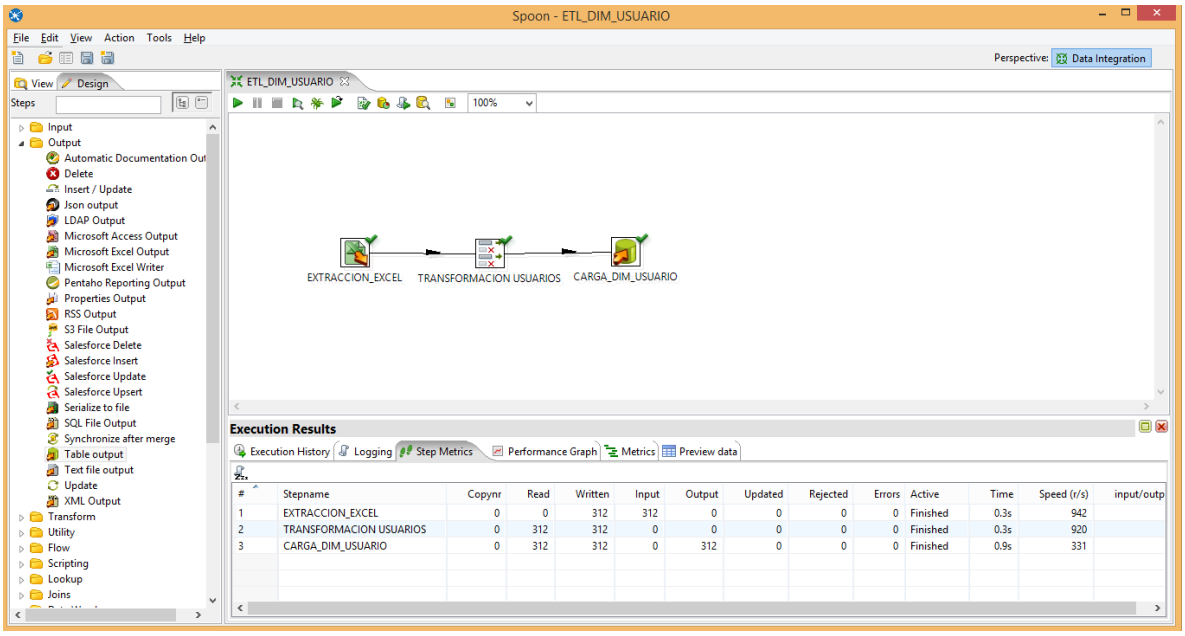
Anexo 11: Nuevos valores y dimension de usuario.



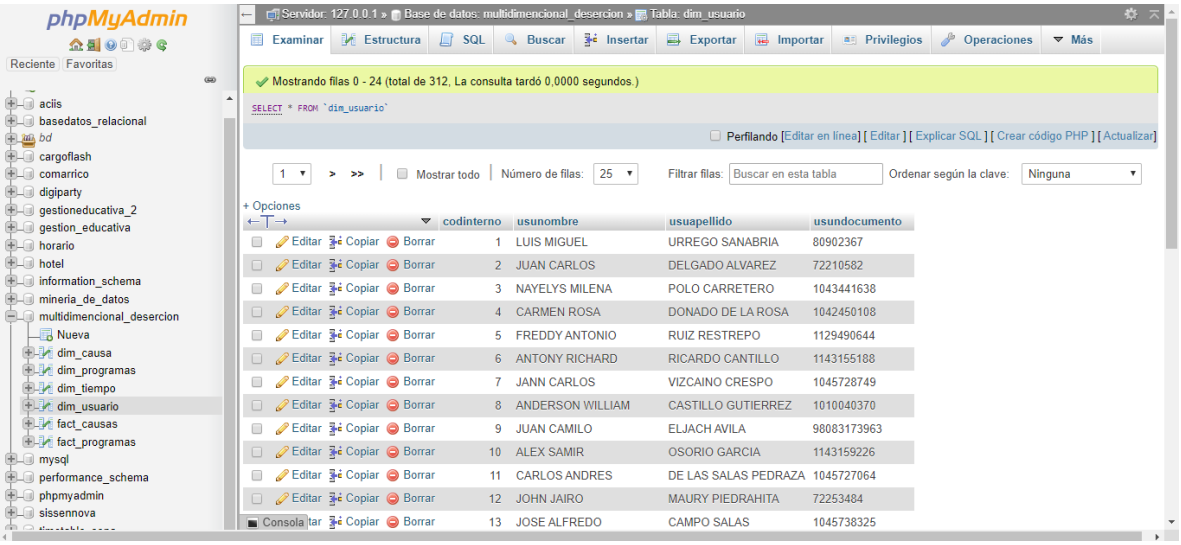
Anexo 12: Resultado de la dimension de usuario.



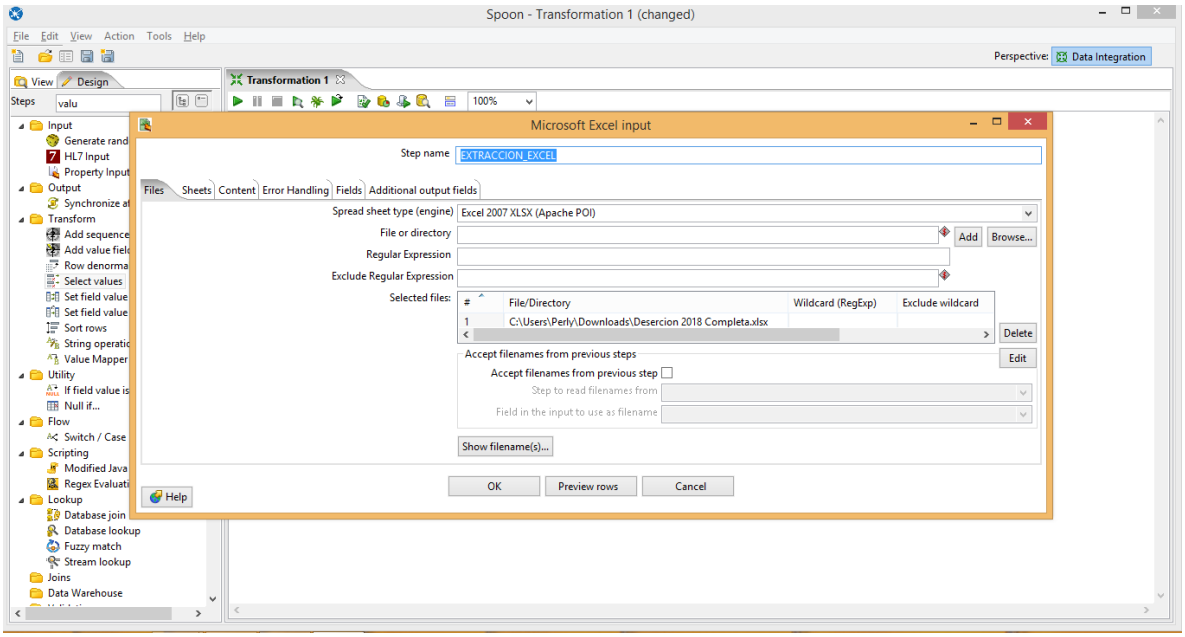
Anexo 13: extracción transformación y carga (ETL).



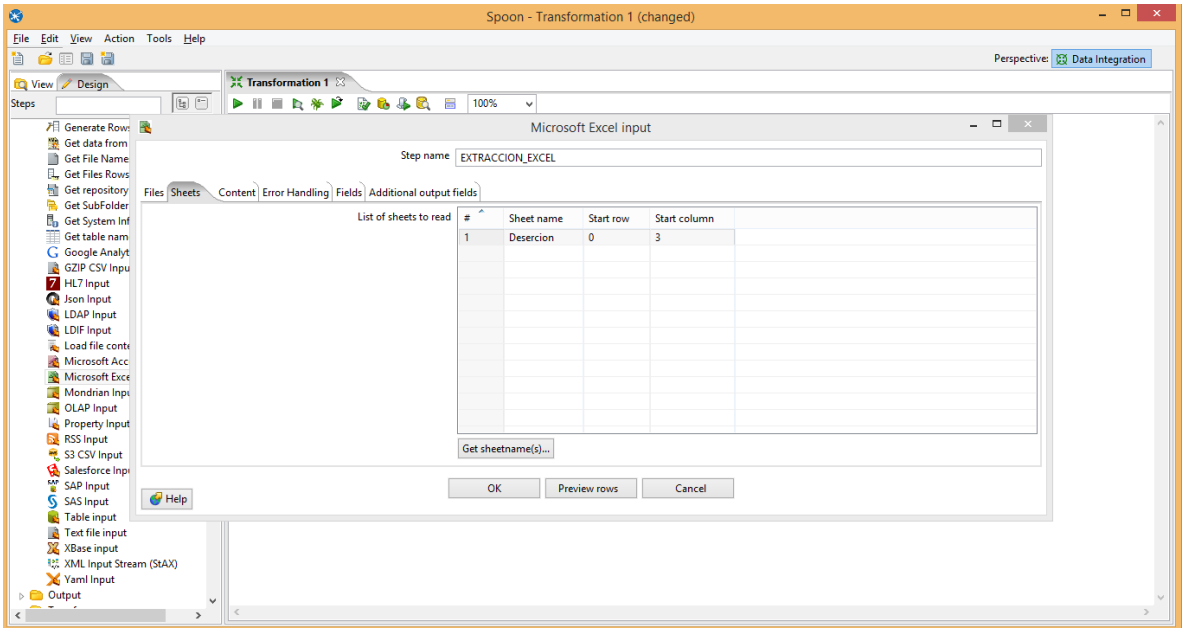
Anexo 14: Verfificacion almacenado de base de datos.



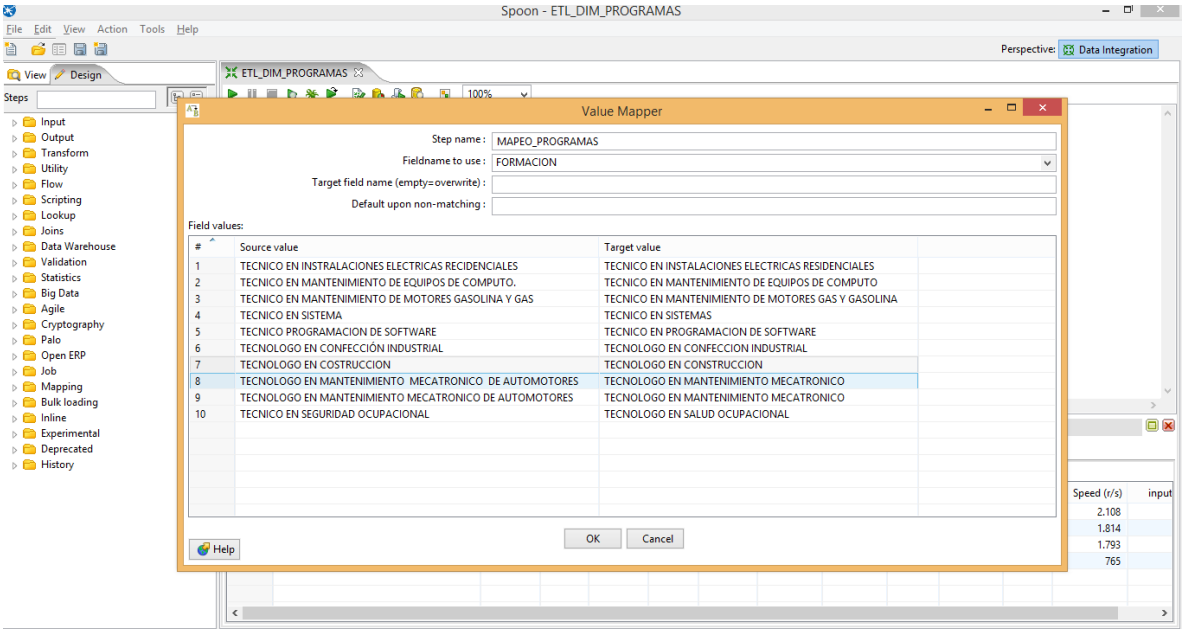
Anexo 15: Proceso de la creacion de la dimension de programas .



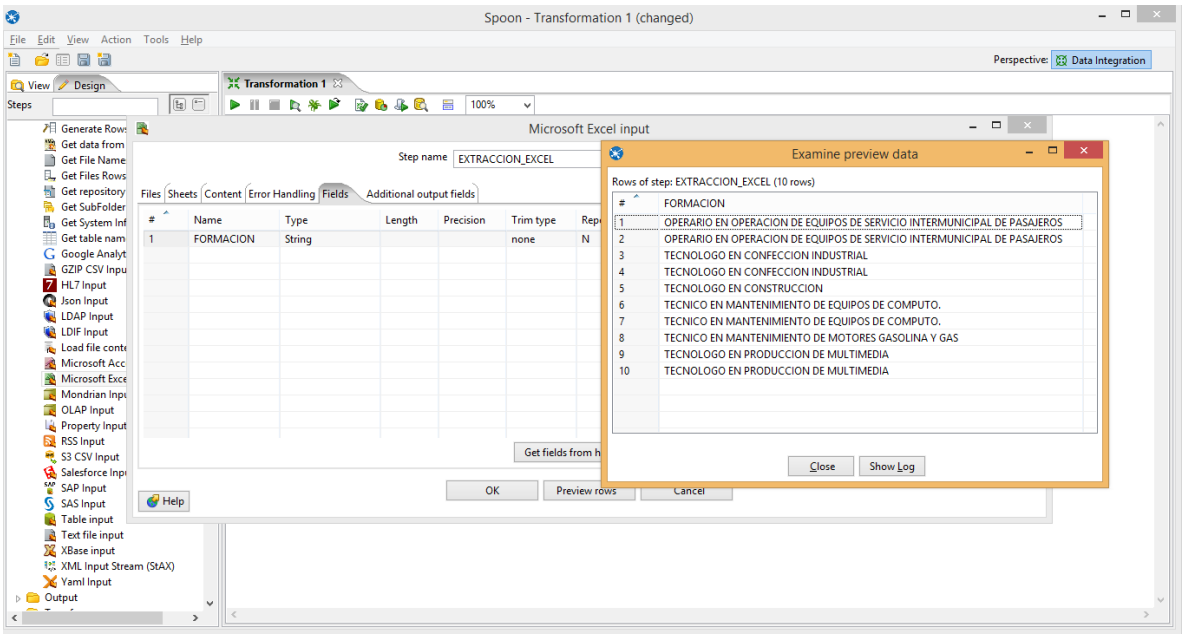
Anexo 16: Extraccion y posicion de la lectura.



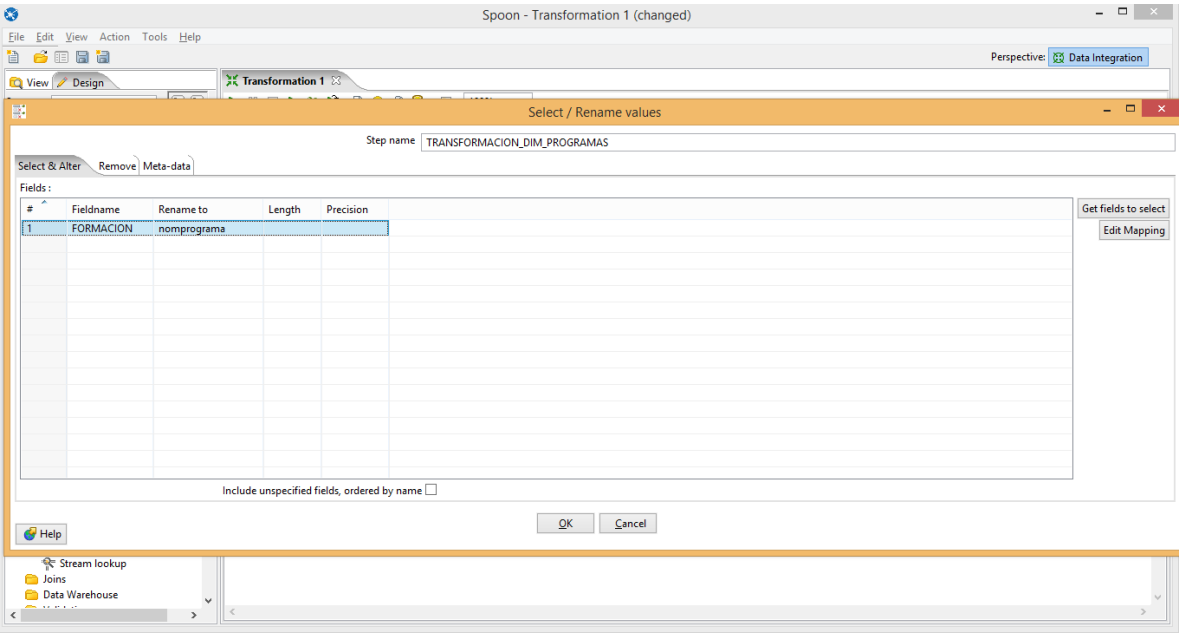
Anexo 17: Mapeo de programas para estandarizar y evitar programas duplicados.



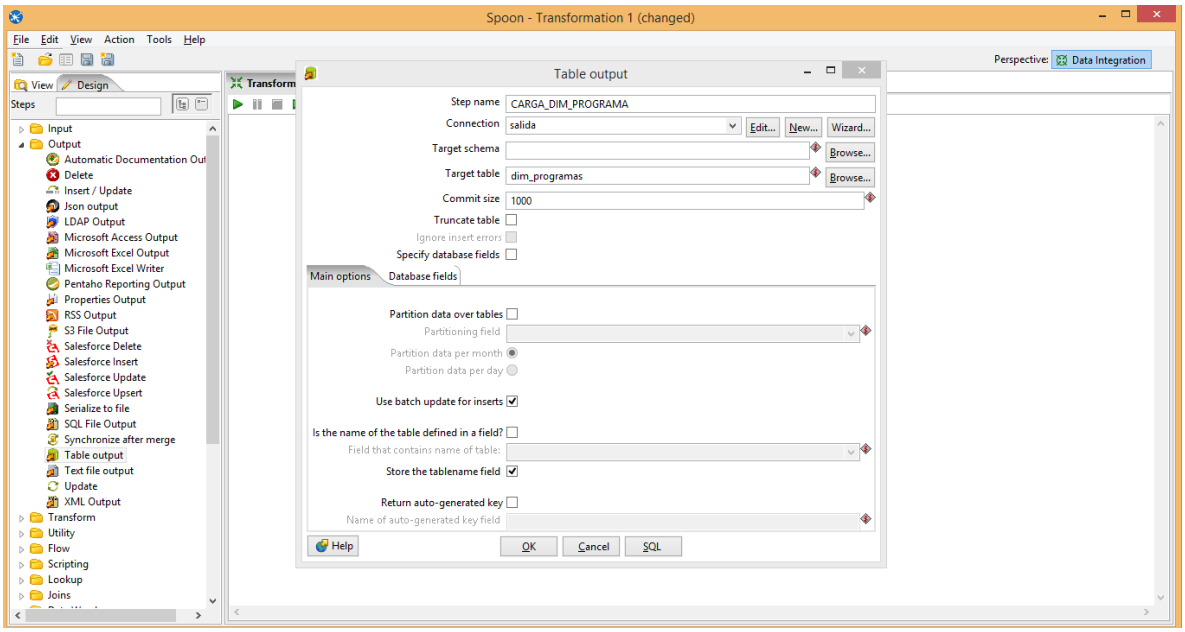
Anexo 18: muestra de los valores.



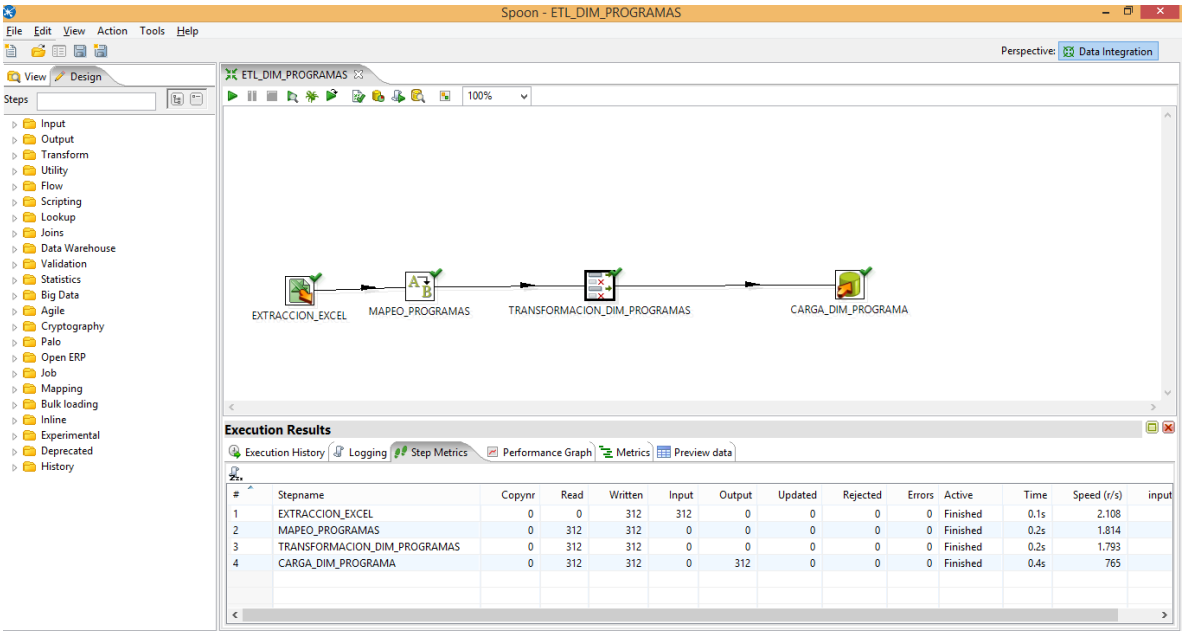
Anexo 19: Cambio de los campos de acuerdo a la dimension del programa.



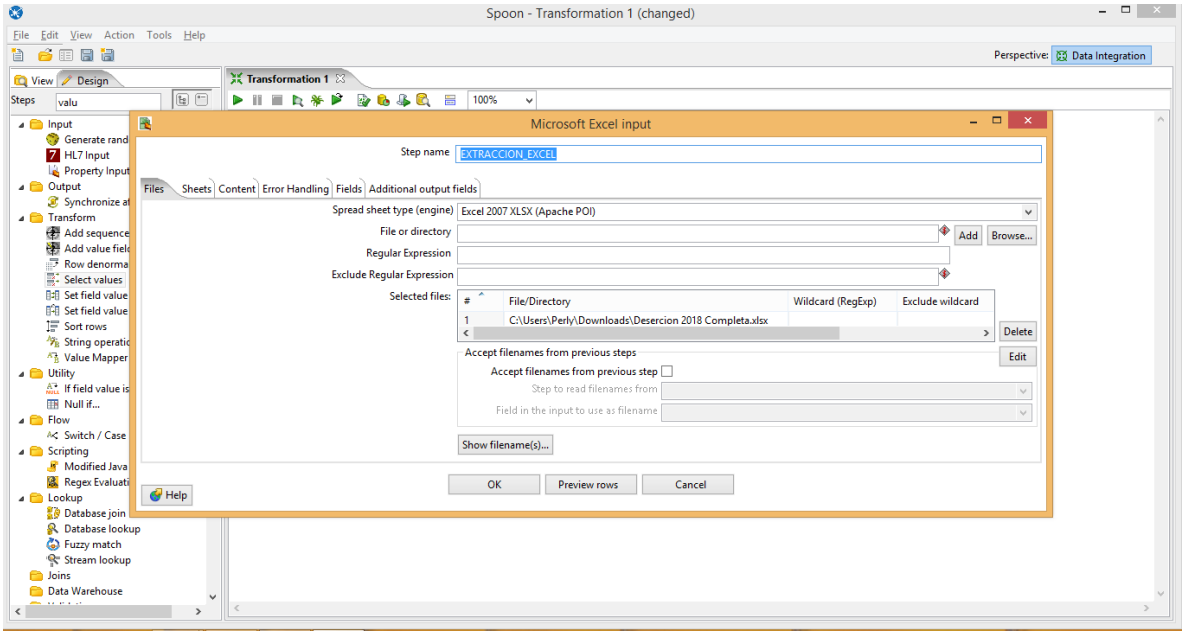
Anexo 20: Salida de la dimension del programa.



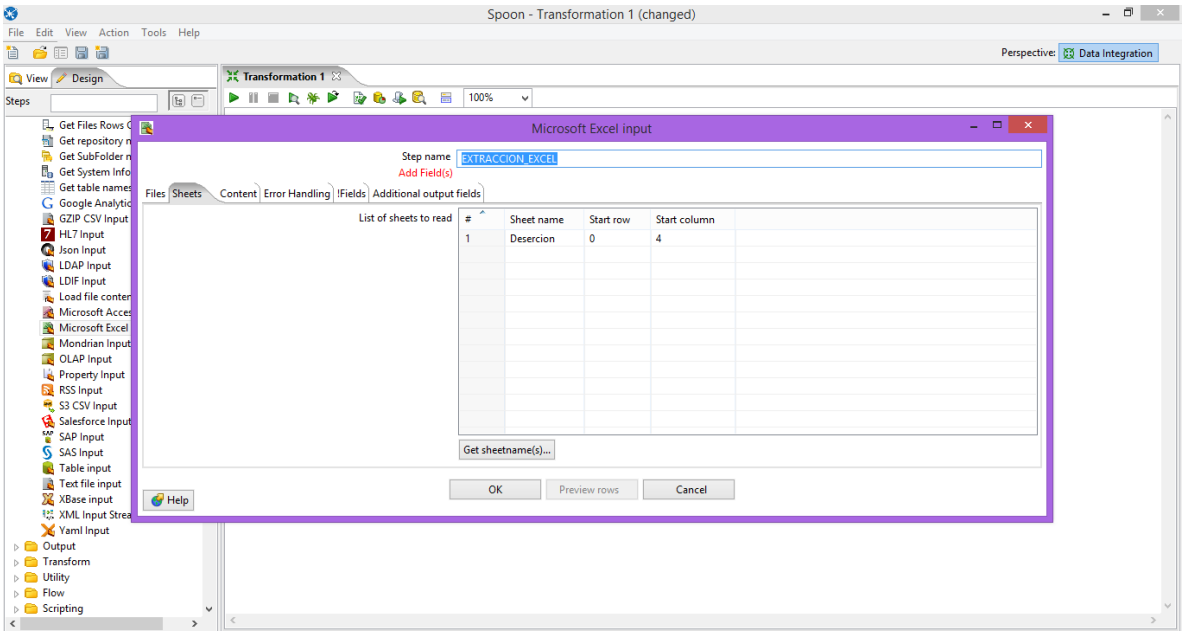
Anexo 21: proceso de extracción, transformación y carga (ETL).



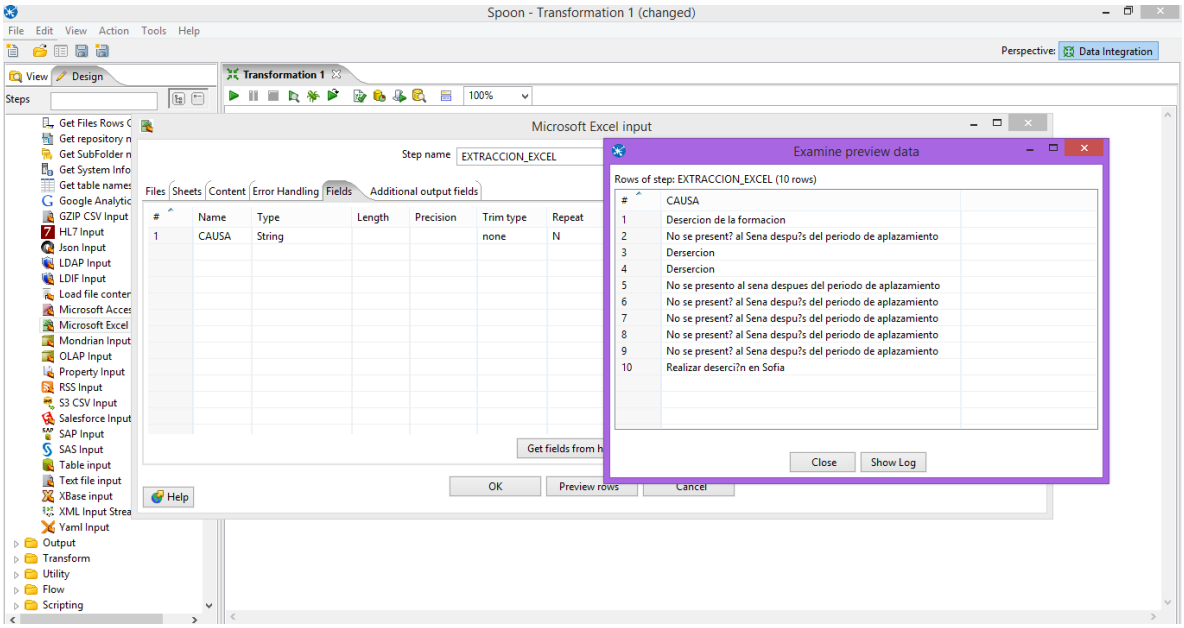
Anexo 22: Proceso de creación de la dimensión causa.



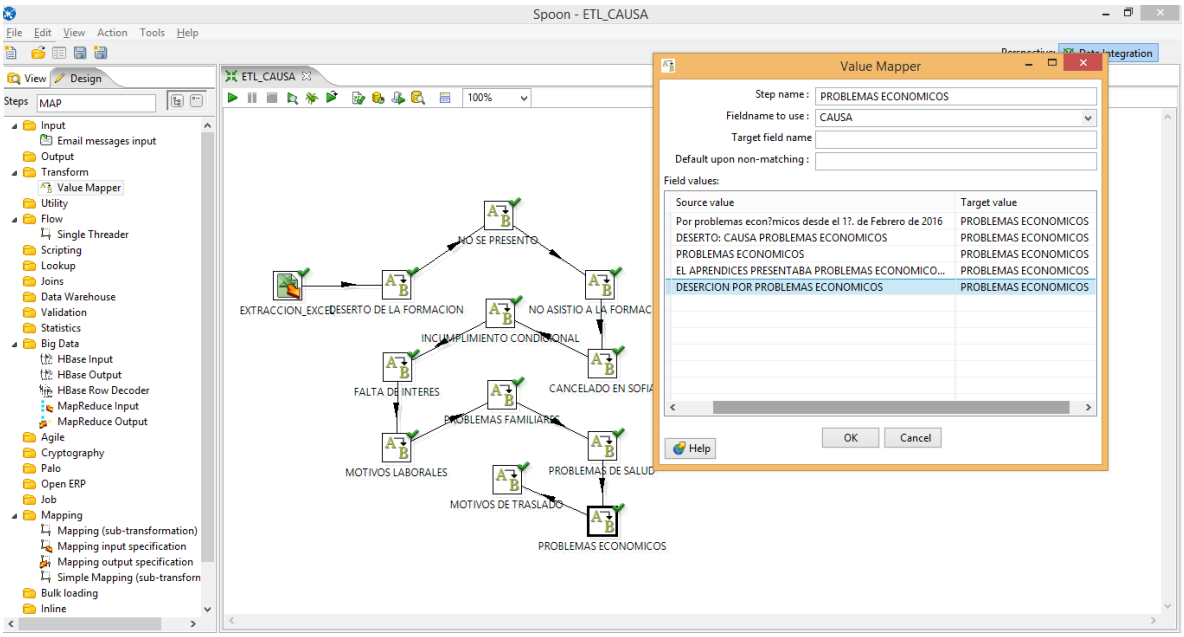
Anexo 23: extracción y la posición de lectura.



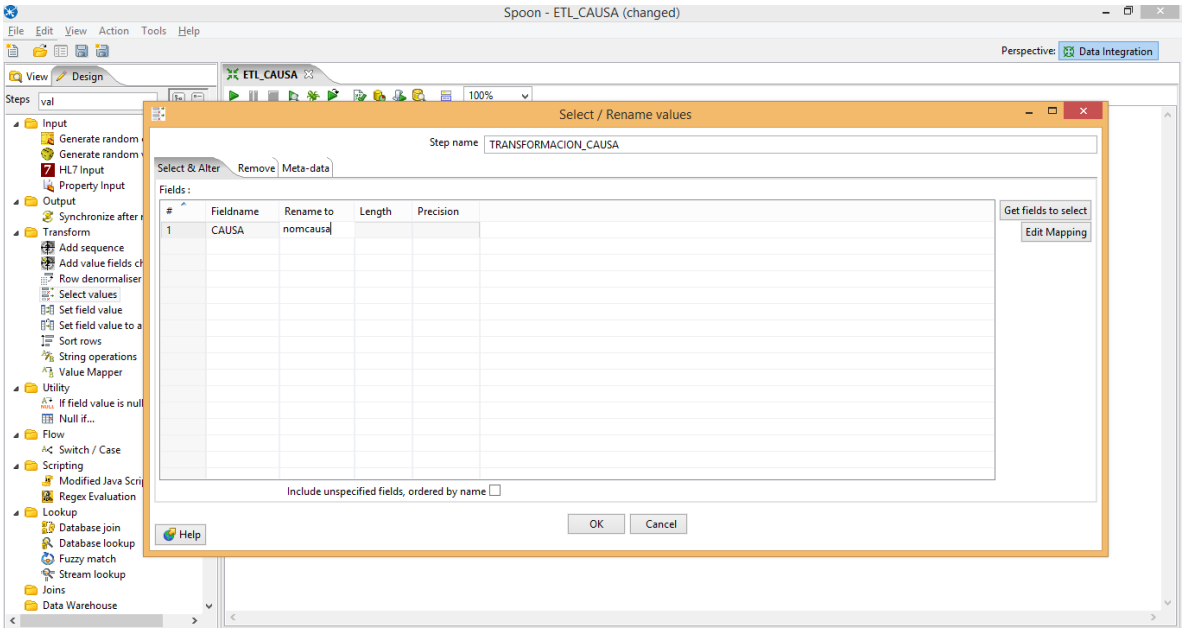
Anexo 24: Visualizacion de la muestra de valores.



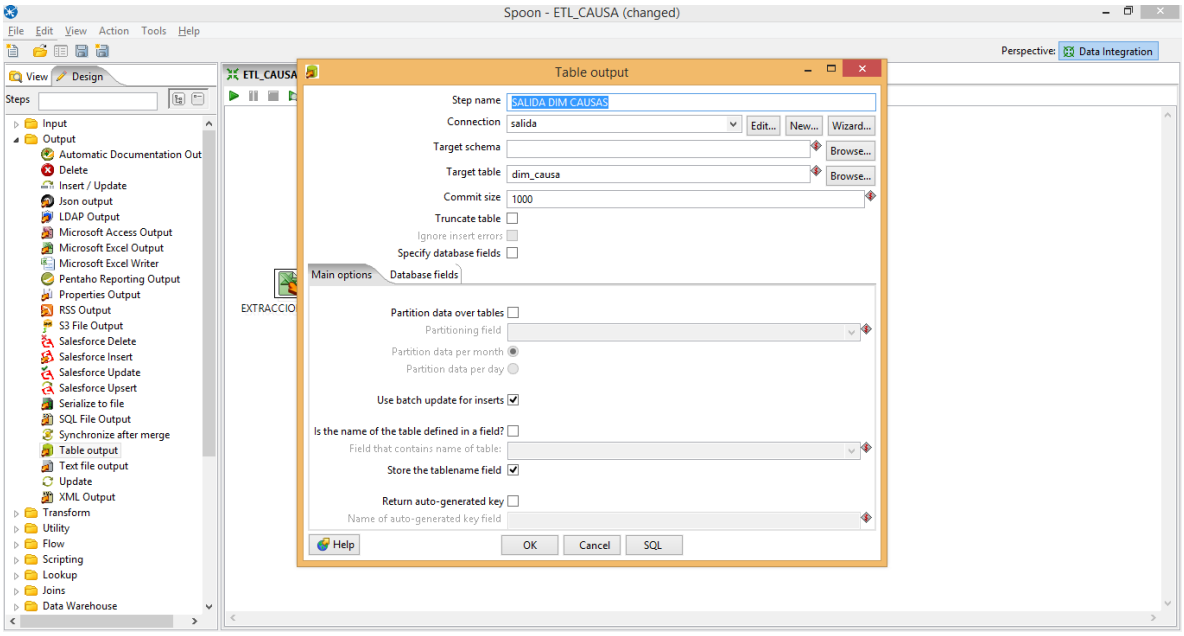
Anexo 25: Mapeo de las causas de desercion.



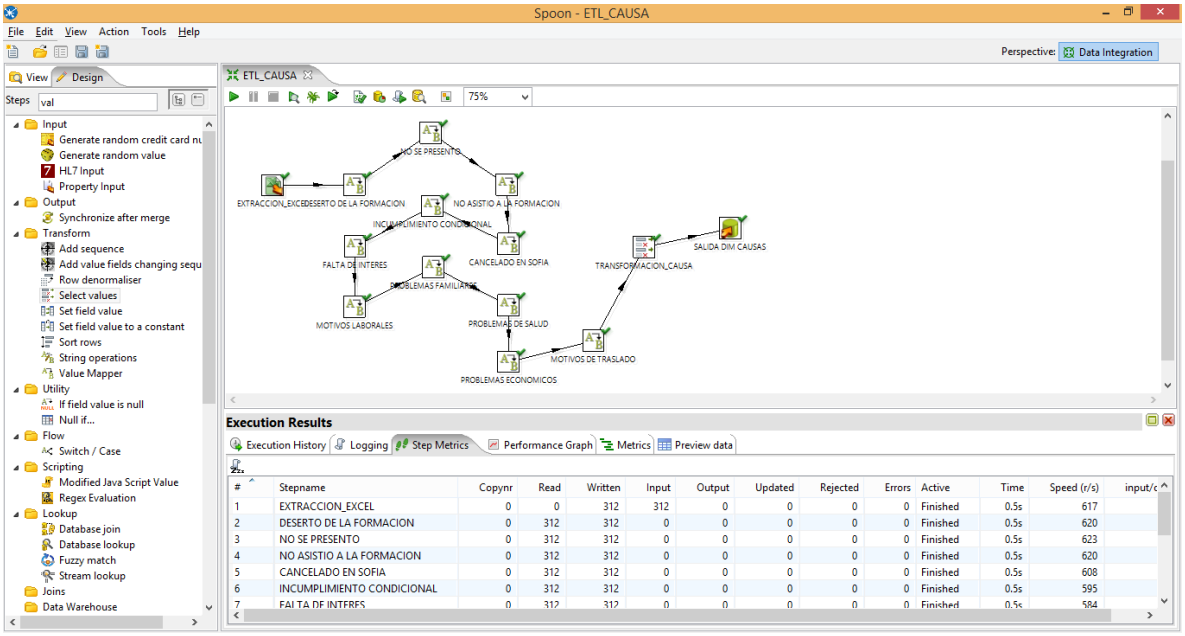
Anexo 26: Cambio de los campos de acuerdo a la dimension causa.



Anexo 27: salida para la dimensión Causa.

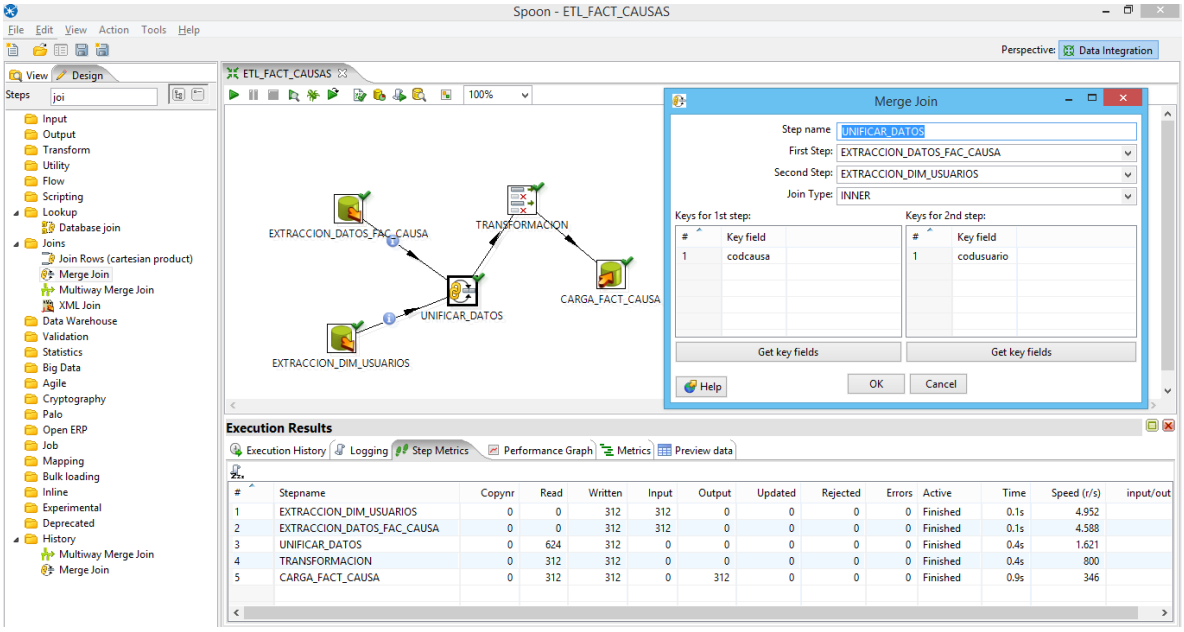


Anexo 28: proceso de extracción, transformación y carga (ETL).

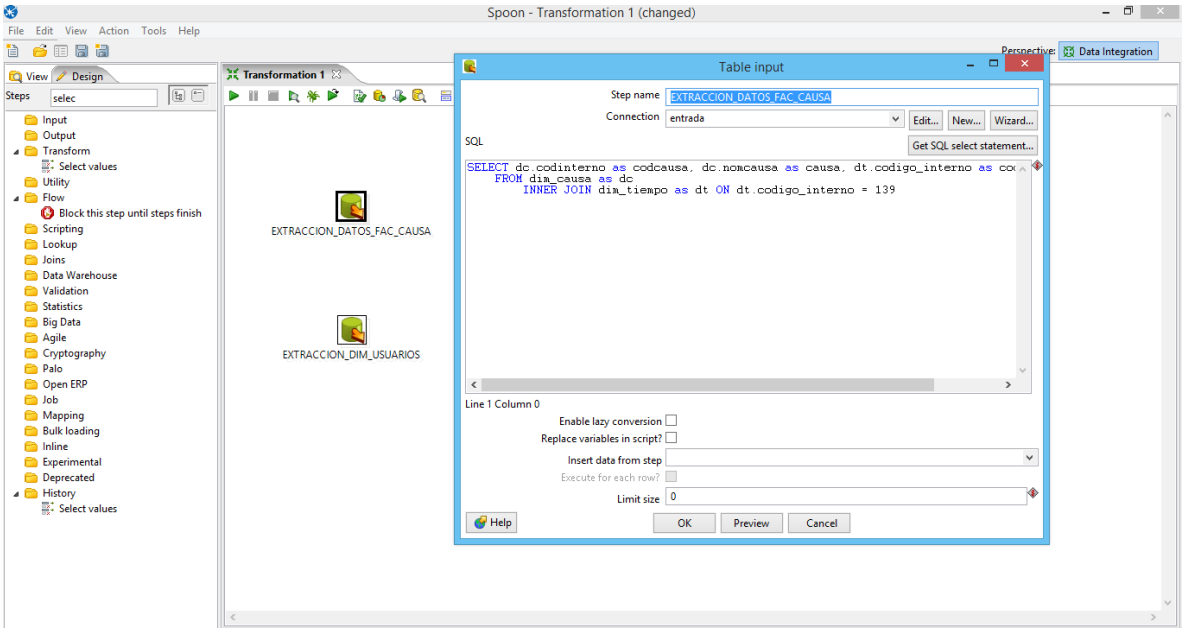


#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/c
1	EXTRACCION_EXCEL	0	0	312	312	0	0	0	0	Finished	0.5s	617	
2	DESERTO DE LA FORMACION	0	312	312	0	0	0	0	0	Finished	0.5s	620	
3	NO SE PRESENTO	0	312	312	0	0	0	0	0	Finished	0.5s	623	
4	NO ASISTIO A LA FORMACION	0	312	312	0	0	0	0	0	Finished	0.5s	620	
5	CANCELADO EN SOFIA	0	312	312	0	0	0	0	0	Finished	0.5s	608	
6	INCUMPLIMIENTO CONDICIONAL	0	312	312	0	0	0	0	0	Finished	0.5s	595	
7	FALTA DE INTERES	0	312	312	0	0	0	0	0	Finished	0.5s	584	

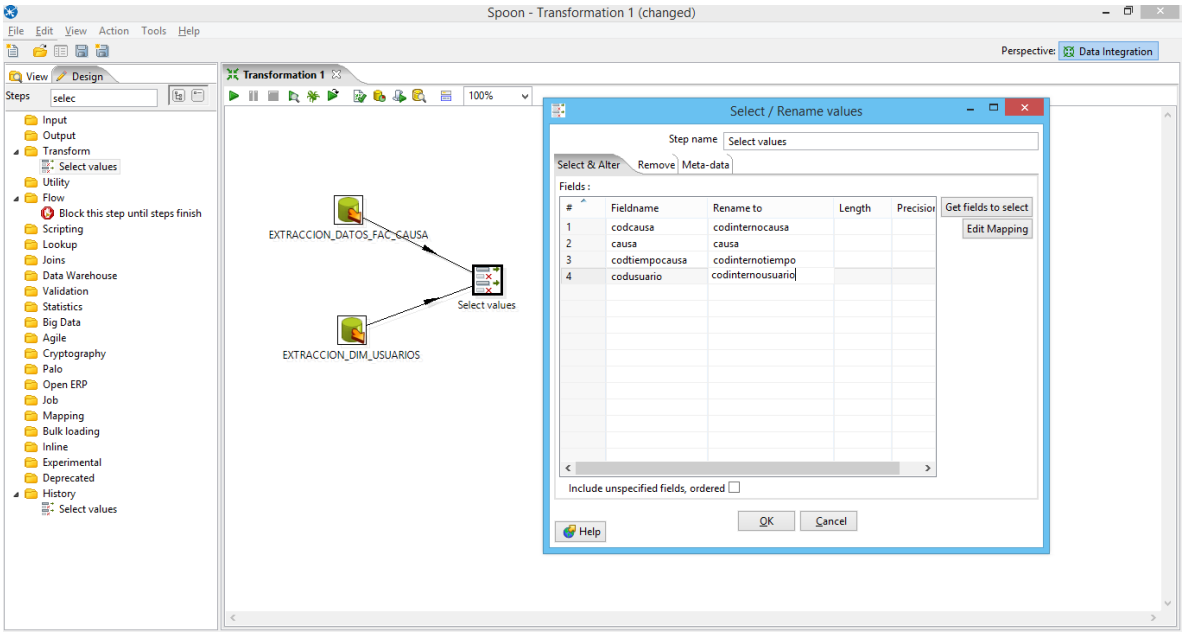
Anexo 29: Proceso de creación de la “fact table” Causas.



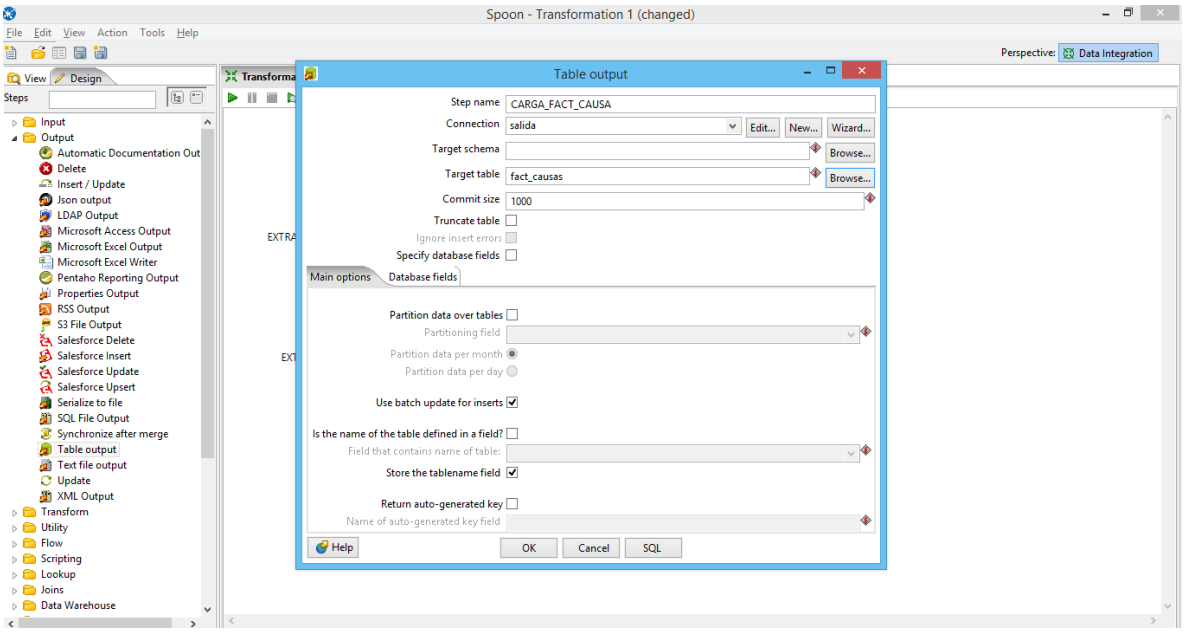
Anexo 30: Extraciion de los datos de las dimensiones para cargar la “fact table” Causa.



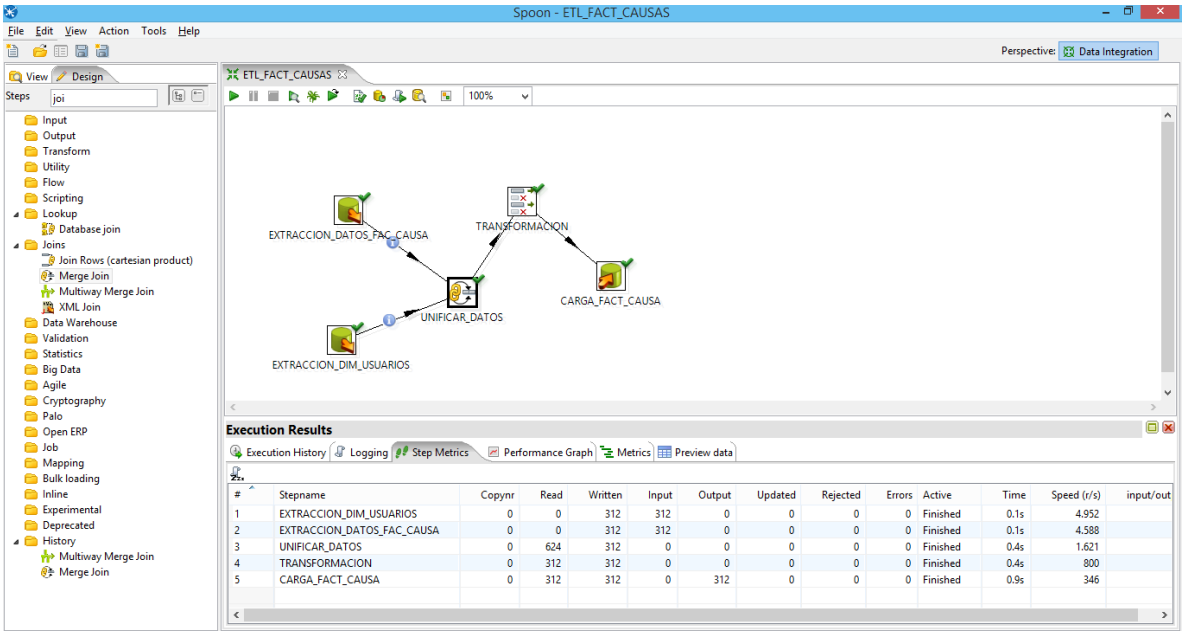
Anexo 31: Trasnformacion de los nombres según la “fact Table” Causa.



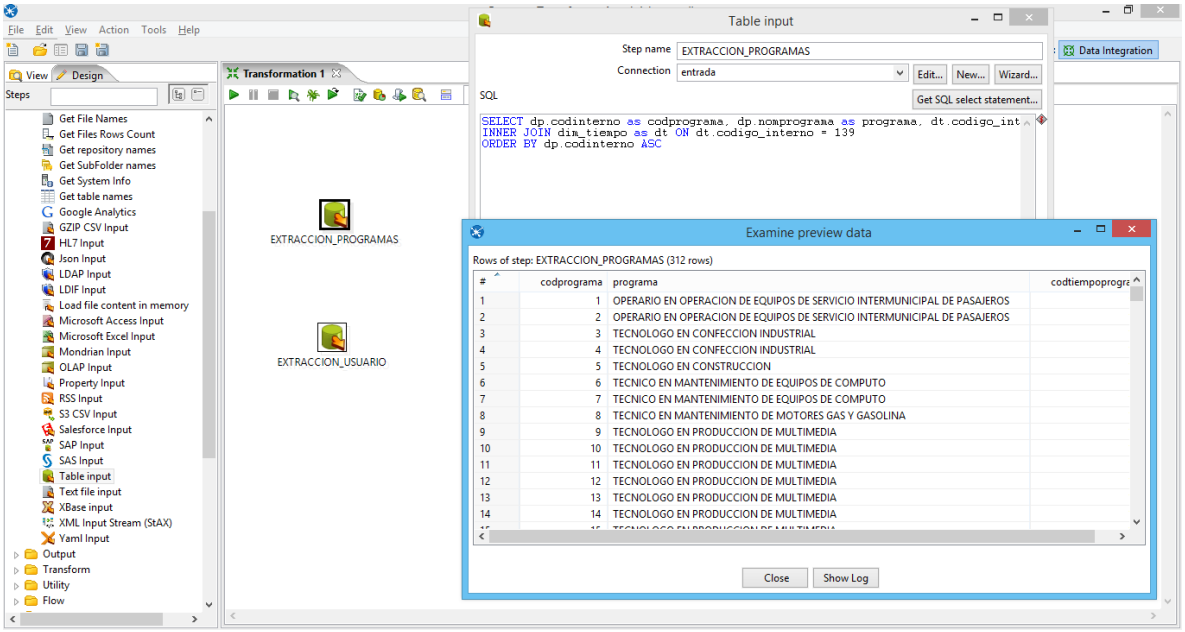
Anexo 32: salida para cargar datos a la “fact table” Causas.



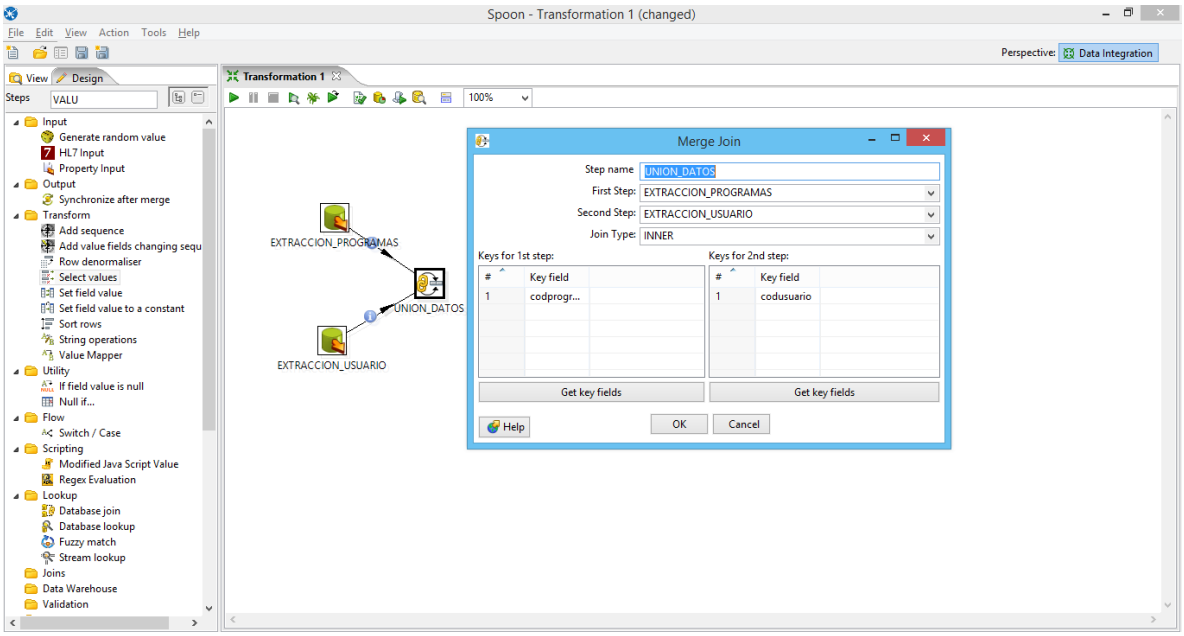
Anexo 33: proceso de extracción, transformación y carga (ETL).



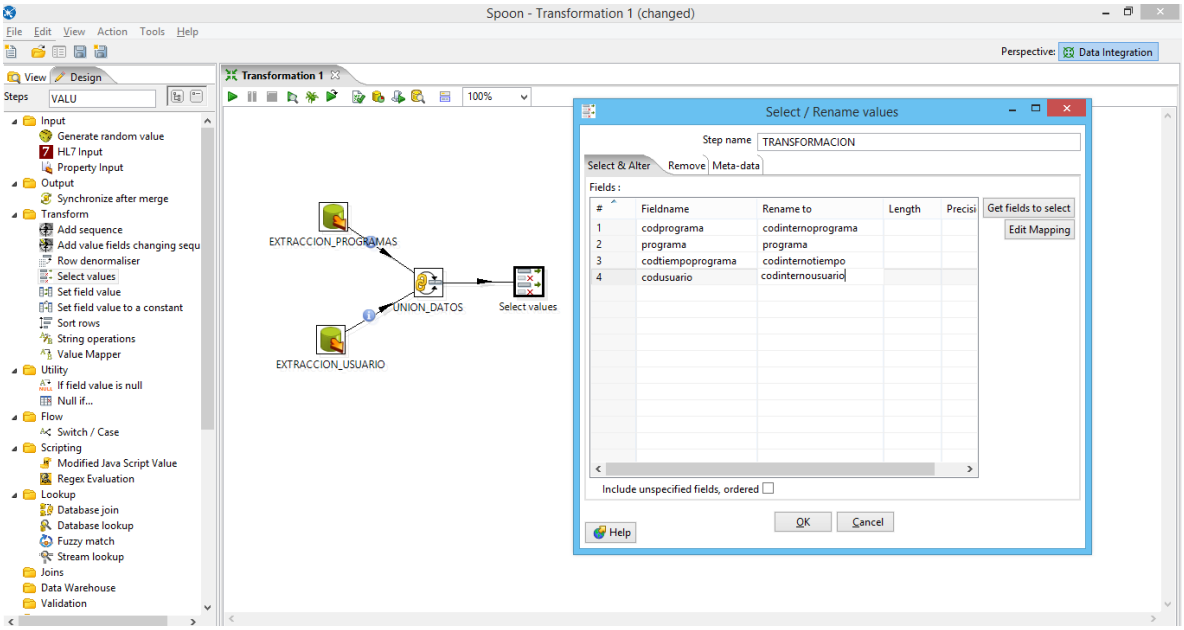
Anexo 34: Proceso de creación de la “fact table” Programas.



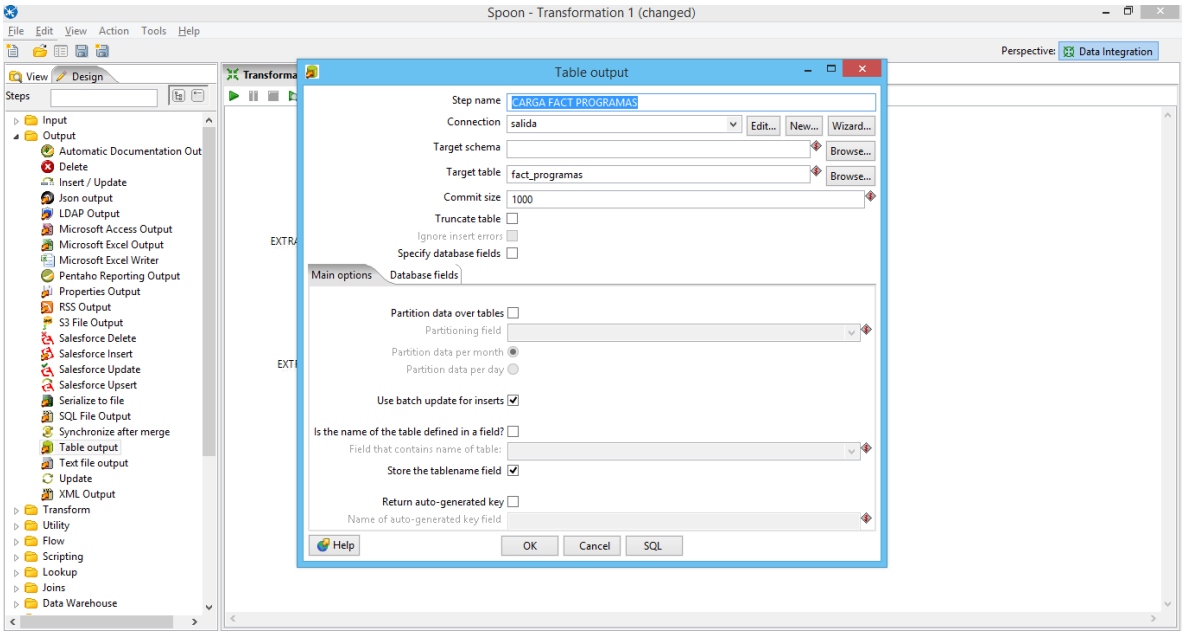
Anexo 35: Realizacion de un “merge join” para unificar los datos extraídos.



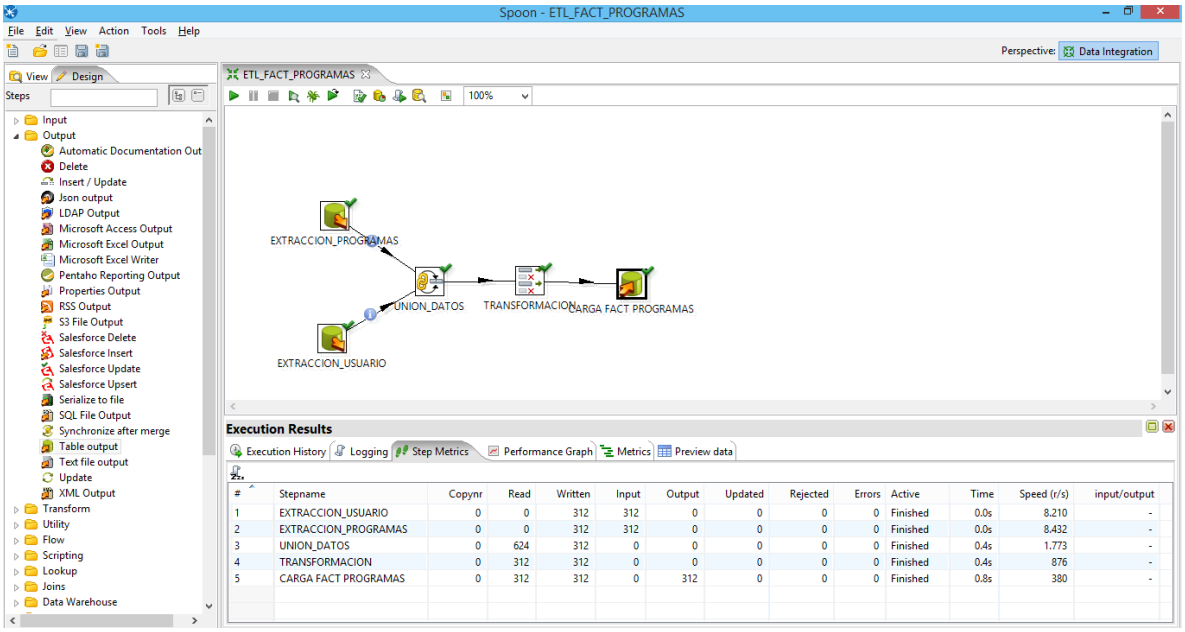
Anexo 36: transformacion de los nombres según la “fact table” Programas.



Anexo 37: salida para cargar datos a la “fact table” Programas.



Anexo 38: proceso de extracción, transformación y carga (ETL).



#	Stepname	Copynr	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	input/output
1	EXTRACCION_USUARIO	0	0	312	312	0	0	0	0	Finished	0.0s	8.210	-
2	EXTRACCION_PROGRAMAS	0	0	312	312	0	0	0	0	Finished	0.0s	8.432	-
3	UNION_DATOS	0	624	312	0	0	0	0	0	Finished	0.4s	1.773	-
4	TRANSFORMACION	0	312	312	0	0	0	0	0	Finished	0.4s	876	-
5	CARGA FACT PROGRAMAS	0	312	312	0	312	0	0	0	Finished	0.8s	380	-