

Análisis de la Deserción Estudiantil de Cecar Utilizando Herramientas de Inteligencia de
Negocios con Licencia Libre

Luis Carlos Canchila Tapias
Javier Andrés Sánchez Bohórquez

Corporación Universitaria del Caribe – Cecar
Facultad de Ciencias Básicas, Ingeniería y Arquitectura
Programa de Ingeniería de Sistemas
Sincelejo
2016

Análisis de la Deserción Estudiantil de Cecar Utilizando Herramientas de Inteligencia de
Negocios con Licencia Libre

Luis Carlos Canchila Tapias
Javier Andrés Sánchez Bohórquez

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

Director:
Jhon Jaime Mendez Alandete

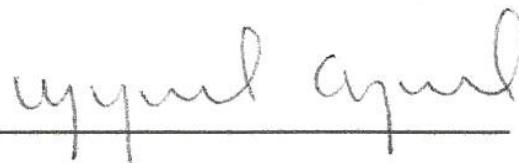
Corporación Universitaria del Caribe – Cecar
Facultad de Ciencias Básicas, Ingeniería y Arquitectura
Programa de Ingeniería de Sistemas
Sincelejo
2016

Nota de aceptación

4,25



Firma del presidente del jurado



Firma del jurado



Firma del jurado

Tabla de Contenido

	Pág.
Resumen	10
Abstract	11
Introducción	12
1. Planteamiento de Problema	14
2. Objetivos	16
2.1. Objetivo General	16
2.2. Objetivos Específicos	16
3. Marco Teórico	18
3.1. Estado del Arte	18
3.2. Sistemas de Información	24
3.2.1. Preguntas Orientadoras en el Contexto de una Organización	24
3.2.2. Sistemas de Información y la Organización	26
3.2.3. Datos, Información y Conocimiento (DIC)	27
3.2.4. Definición de Sistemas de Información	28
3.2.5. Tipos de Sistemas de Información	30
3.2.6. Sistemas de Información Transaccionales (TPS)	30
3.2.7. Sistemas de Información de Apoyo a la Toma de Decisiones	31
3.2.8. Sistemas de Información Estratégicos	32
3.3. Inteligencia de Negocios	33
3.3.1. Fuentes de Datos	37
3.3.2. ETL (Extraction, Transform, Load)	38
3.3.3. Bodegas o Almacenamiento de datos (en Inglés: Datawarehouse. DW)	39
3.3.3.1. Nivel Lógico	40
3.3.3.2. Nivel Conceptual	40
3.3.3.3. Nivel Físico	41
3.3.4. Servidores Intermedios	41
3.3.5. Aplicaciones Frontend	42
3.3.5.1. Olap (Online Analytical processing)	42
3.3.5.1.1. Rolap (Relational Online Analytical Processing)	42
3.3.5.1.2. Molap (Multidimensional Online Analytical Processing)	43
3.3.5.1.3. Holap (Hybrid Online Analytical Processing)	43
3.3.6. Minería de Datos	47
3.3.6.1. Modelos de Minerías de Datos	48
3.3.6.2. Tareas de la Minería de Datos	54
3.3.7. Datamart	54

4. Metodología	56
4.1.Fase I: Caracterización de los Procesos del Programa TAE para el Análisis de la Deserción en la Corporación Universitaria del Caribe - CECAR .	57
4.1.1. Entrevistas	57
4.1.2. Flujo de procesos del programa TAE	58
4.1.3. Matriz DOFA del Programa TAE	59
4.2. Fase II: Implementar Herramientas de Inteligencia de Negocio que Apoyen el Proceso de Análisis de Deserción del Programa de Trayectoria Académica TAE.	60
4.2.1. Diagnóstico de los Sistemas de Información Transaccionales de CECAR	60
4.2.2. Definición de Requerimientos en Base al Proceso de Análisis de Deserción Definidos por TAE	64
4.2.3. Diseño de la Solución en Base a la Metodología de KIMBALL	67
4.2.3.1.Diseño Arquitectónico	67
4.2.3.2.Modelo Dimensional	69
4.2.3.3.Bus Matrix DATAWAREHOUSE	72
4.2.3.4.Modelo Lógico	72
4.2.3.5.Diseño Conceptual	73
4.2.3.6.Diseño Lógico	73
4.3. Fase III: Implementación de Herramientas de Inteligencia de Negocios Seleccionadas	74
4.3.1. Instalación y Configuración del Repositorio de Datos	74
4.3.2. Instalación y Configuración ETL	75
4.3.3. Trabajo Dimensión Programas	75
4.3.4. Trabajo Dimensión Lugar de Residencia	75
4.3.5. Trabajo Dimensión Asignatura	77
4.3.6. Trabajo Dimensión Estudiantes	78
4.3.7. Trabajo Dimensión Tiempo	60
4.3.8. Trabajo Dimensión Docente	80
4.3.9. Trabajo Tabla de Hechos Rendimiento	82
4.3.10. Trabajo Tabla de Hechos Deserciones	82
4.3.11. Soluciones Reporting para Analizar la Deserción	83
4.3.12. Solución OLAP para Analizar la Deserción	85
Cronograma	105
Conclusiones	107
Recomendaciones	109
Referencias Bibliográficas	111
Anexos	117

Lista de Figuras

	Pág.
Figura 1. Elementos que definen un Sistema de Información	27
Figura 2. Datos, Información y Conocimiento – DIC	28
Figura 3. Sistemas de información desde la perspectiva del usuario	31
Figura 4. Arquitectura de Inteligencia de negocios típica	36
Figura 5. Cubo dimensional	41
Figura 6. Proceso Extracción de conocimiento en Bases de datos	46
Figura 7. Matriz DOFA TAE	59
Figura 8. Diseño Arquitectónico y distribución del hardware de CECAR	63
Figura 9. Diseño Arquitectónico	68
Figura 10. Diseño conceptual del modelo dimensional para estudiantes por debajo de 3.0 en el datamart académico	72
Figura 11. Diseño conceptual del modelo dimensional para la cuantificación de la deserción estudiantil en el datamart académico	73
Figura 12. Diseño lógico del modelo dimensional para estudiantes por debajo de 3.0 en el datamart académico	73
Figura 13. Diseño conceptual del modelo dimensional para la cuantificación de la deserción estudiantil en el datamart académico.	75
Figura 14. Trabajo dimensión programas	75
Figura 15. Trabajo dimensión lugar	76
Figura 16. Adaptación de datos para el trabajo dimensión lugar residencia	76
Figura 17. Trabajo dimensión asignatura	78
Figura 18. Trabajo dimensión estudiantes	79
Figura 19. Unión de las diferentes tablas a través del objeto tMap	80
Figura 20. Trabajo dimensión tiempos.	81
Figura 21. Expresión creada para la columna.	81
Figura 22. Trabajo dimensión docente.	83
Figura 23. Trabajo tabla de hecho rendimiento	85
Figura 24. Trabajo tabla de hecho deserciones	86
Figura 25. Regla creada para la columna promedio	87
Figura 26. Login: plataforma APPTAE	88
Figura 27. Vista Historia Clínica (Filtros)	88
Figura 28. Vista Historia Clínica (Resultados)	89
Figura 29. Vista Reporte (Historia Clínica)	90
Figura 30. Vista Medios de Comunicación	90
Figura 31. Vista Detalles Asignatura (filtros)	90
Figura 32. Vista Reporte (Detalles Asignatura)	90

Figura 33. Vista Historial Académico (Filtro)	90
Figura 34. Vista Respuesta de la Consulta (Detalles Asignatura)	92
Figura 35. Vista Reporte para imprimir el historial académico.	93
Figura 36. Vista de las Gráficas (Aprobados y Reprobados)	94
Figura 37. Vista de Reporte de estudiantes.	95
Figura 38. Vista Resultado del Reporte para Imprimir (Reporte)	95
Figura 39 Consulta OLAP con JPivot que discrimina por estudiantes las medidas del cubo deserciones.	96
Figura 40. Consulta OLAP con JPivot que discrimina por sexo las medidas del cubo.	98
Figura 41 Consulta OLAP con JPivot que discrimina por sexo y condiciona por tiempo las medidas del cubo deserciones	99
Figura 42. Consulta OLAP con JPivot que discrimina por periodo, programa y estudiantes las medidas del cubo deserciones.	99

Lista de Tablas

	Pág
Tabla 1. Elementos de un Sistema de Información	29
Tabla 2. Las primeras 5 suite de inteligencia del negocio en el periodo 2011-2012	37
Tabla 3. Relación entre tareas y técnicas	53
Tabla 4. Entrevistas Aplicadas a diferentes actores que intervienen en el programa TAE	57
Tabla 5. Inventario de la infraestructura tecnológica en software y hardware.	62
Tabla 6. Historia de usuario (RESPIS).	64
Tabla 7. Historia de usuario (RECAR).	65
Tabla 8. Historia de usuario (INFACA).	65
Tabla 9. Historia de usuario (INPRO).	67
Tabla 10. Historia de usuario (INDE).	67
Tabla 11. Bus Matrix Data WareHouse	70
Tabla 12. Sentencia SQL para la vista AcaDptoMun_v	76
Tabla 13. Sentencia SQL para la vista AcaBloquesReg_V	77
Tabla 14. Sentencia SQL para las vista AcaMatriculas_V y AcaSniesAspirante_V	78
Tabla 15. Sentencia SQL para las vista AcamatriculasSem_V	80
Tabla 16. Sentencia SQL para las vista AcamatriculasSem_V	82
Tabla 17. Promedio validación cruzada utilizando las particiones con todos los atributos	100
Tabla 18. Promedio validación cruzada utilizando las particiones con mejores atributos	100
Tabla 19. Promedio validación cruzada utilizando las particiones re-balanceadas con mejores atributos	100
Tabla 20. Resultados obtenidos por OneR para los estudiantes desertores	101
Tabla 21. Resultados obtenidos por JRip para los estudiantes desertores	102
Tabla 22. Resultados obtenidos por Ridor para los estudiantes desertores	102

Lista de Anexos

	Pág.
Anexo A. Instrumento # 1	117
Anexo B. Instrumento # 2	121
Anexo C. Instrumento # 3	123
Anexo D. Audio de entrevistas realizadas a coordinares de programas, funcionarios del programa de trayectoria académica exitosa y estudiantes.	124
Anexo E. Script del modelo de datos para el datamart académico.	124
Anexo F. Cubo deserciones y rendimientos.	124
Anexo G. Trabajos ETL generados en WEKA.	124
Anexo H. Reporte de estudiantes en riesgo de desertar en BIRT.	124

Resumen

El presente trabajo desarrolla un prototipo funcional de inteligencia de negocios utilizando herramientas de software libre para el análisis y proyección de la deserción de estudiantes en la Corporación Universitaria del Caribe –CECAR.

El objetivo principal es proveer un sistema de información decisional de tipo Inteligencia de negocios que de soporte a las necesidades de análisis de información y proyección en relación a estudiantes desertores y en riesgos de desertar, con el fin de proponer estrategias para disminuir tal fenómeno.

La solución desarrollada incluye generación de reportes de análisis tipo OLAP que permiten analizar la deserción y rendimiento académico desde diferentes puntos de vista. De igual forma se presenta la proyección de estudiantes en riesgo de desertar y la generación de alertas tempranas.

Para la implementación del prototipo se utilizó la metodología de Ralph Kimball la cual facilita la construcción del sistema y permite lograr la culminación satisfactoria de las necesidades propias de los usuarios. Se especificaron las etapas de gestión de proyectos, planificación, análisis, diseño, implementación y evaluación.

Como futuras ampliaciones se propone el desarrollo del sistema completo agregando la configuración de usuarios del sistema, accesos web, hosting, servidores e implantación. Además de la incorporación de tableros de control, indicadores de desempeño y ampliar la funcionalidad a otras áreas en la Corporación Universitaria del Caribe -CECAR.

Palabras clave: Inteligencia de negocios, deserción académica, minería de datos, OLAP.

Abstract

The present work develops a functional prototype of business intelligence using free software tools for the analysis and projection of student dropout in the University Corporation of the Caribbean -CECAR.

The main objective is a decisional information system of business intelligence type that supports the needs of information analysis and projection in relation to students deserters and in risk of deserting, in order to propose strategies to reduce the phenomenon.

The solution developed includes the generation of OLAP analysis reports that allow analyzing dropout and academic performance from different points of view. Likewise, the projection of students at risk of deserting and the generation of early warnings is presented.

For the implementation of the prototype use the methodology of Ralph Kimball which facilitates the construction of the system and allows to achieve the satisfactory completion of the needs of the users. It specifies the stages of project management, planning, analysis, design, implementation and evaluation.

As future extensions we propose the development of the complete system that adds the configuration of the system users, the web accesses, the hosting, the servers and the implantation. In addition to the incorporation of control panels, performance indicators and expand the functionality of other areas in the University Corporation of the Caribbean -CECAR.

Key words: Business intelligence, academic desertion, data mining, OLAP.

Introducción

La deserción estudiantil es uno de los problemas que enfrenta el sistema de educación superior a nivel mundial. Colombia en los últimos años se ha caracterizado por el aumento de cobertura e ingreso de estudiantes nuevos, el número de alumnos que logra culminar sus estudios superiores no es alto, dejando entrever que una gran parte de éstos abandona sus estudios, principalmente en los primeros semestres. Según estadísticas del Ministerio de Educación Nacional (Superior, 2015) de cada cien estudiantes que ingresan a una institución de educación superior cerca de la mitad no logra culminar su ciclo académico y obtener la graduación. Actualmente CECAR presenta una tasa deserción del 12.92% en los programas presenciales y 15.05% para los programas de educación abierta y distancia, cifras obtenidas en el segundo periodo del 2015 las cuales se incrementaron en un casi 3% en relación al periodo inmediatamente anterior. En aras de disminuir el porcentaje de deserción académica en todos los programas ofrecidos, CECAR implementó el Programa de Trayectoria Académica Exitosa (TAE), que busca disminuir la tasa de deserción estudiantil a través de un acompañamiento a los estudiantes durante su proceso de formación profesional. El programa se encuentra limitado debido a que los sistemas de información transaccionales de CECAR no apoyan los procesos al interior del programa, ni generan información y mucho menos conocimiento en pro de proponer nuevas estrategias para disminuir las tasas de deserción.

Teniendo en cuenta lo anterior, este desarrollo tecnológico hace uso de herramientas con licencias de software libre de inteligencia de negocio para apoyar a los procesos de análisis de la deserción desde el programa TAE. Presentando en este, una caracterización del proceso de análisis de estudiantes en riesgos de desertar realizado desde el programa de trayectoria académica liderado por el departamento de bienestar universitario. Seguidamente y en base a lo anterior, se identifican requerimientos que permitan seleccionar, instalar, configurar y desplegar herramientas de inteligencia de negocio que sirvan para realizar análisis descriptivos y predictivos de estudiantes con riesgo a desertar. Por último, se evalúa la información suministrada por las herramientas y se define un perfil del estudiante desertor bajo el contexto de CECAR:

El desarrollo del sistema utilizó la metodología de Ralph Kimball, la cual propone un desarrollo incremental de los requisitos del sistema identificando actividades claras y concisas en el ciclo de vida, para implementar un prototipo software que servirá de apoyo en la toma de decisiones del programa de trayectoria académica exitosa (TAE), a través de la utilización de herramientas de inteligencia de negocios que permitan la identificación de factores que inciden en la deserción estudiantil, analizados desde el contexto y la problemática del estudiante CECARENSE, además de definir estrategias proactivas encaminadas a la erradicación o mitigación de los principales factores de deserción.

1. Planteamiento del Problema

La deserción es uno de los problemas que aborda la mayoría de las instituciones de educación superior de muchos países a nivel mundial. A través de distintas investigaciones, se da cuenta de un número importante de estudiantes que no logran culminar sus estudios universitarios, con el consecuente costo social asociado a este fenómeno. Por lo anterior, algunos países han comenzado a diseñar profundos procesos de mejoramiento para aumentar la retención en los primeros años de estudios universitarios (UNESCO 2004).

En la Corporación Universitaria del Caribe –CECAR, la tasa de deserción para el segundo periodo de 2015 se situó en el 12.92% en los programas presenciales y 15.05% para los programas de educación abierta y distancia (Academica, 2015). En base a esto y con el fin de disminuir el porcentaje de deserción académica en todos los programas ofrecidos, la institución ha implementado el programa de trayectoria académica exitosa (TAE), que busca disminuir la tasa de deserción estudiantil a través de un acompañamiento a los estudiantes durante su proceso de formación profesional. Actualmente el programa TAE se encuentra limitado debido a que no posee un sistema de información que apoye el análisis de los estudiantes con riesgos a desertar, de forma que se generen patrones descriptivos y predictivos de las características del estudiante desertor y que permiten definir estrategias proactivas y no reactivas, estas últimas implementadas en la actualidad.

Por otra parte CECAR utiliza el software SPADIES, herramienta gratuita suministrada por el Ministerio de Educación a las instituciones de educación superior, para medir y analizar los factores que influyen en la deserción académica. Este software, trabaja en base a características propias de las instituciones (variables institucionales si es pública o privada) y otras de carácter regional (que se relaciona con el entorno macroeconómico regional) para analizar y proyectar la deserción en las instituciones educativas. Lo anterior genera información que algunos casos no son determinantes debido a que se mide en base a los mismos factores y no en función a los contextos de cada institución. Por ejemplo; la definición de desertor que se adopte pueden conllevar a

diferentes resultados, ya que si el desertor es un estudiante que falta por varios periodos (uno, dos o más) al mismo programa académico, esto evidenciara cambios en los resultados debido a que muchos estudiantes que faltan a un periodo pueden regresar al siguiente semestre, lo que no constituye un evento de deserción en sí mismo (teniendo en cuenta la definición), sino una situación de ausencia temporal o si por el contrario el estudiante regresa y se matricula en otro programa académico no se puede medir como deserción sino más bien cambio de programa.

Por lo anterior se hace necesario implementar soluciones tecnológicas a la medida, acorde a las características educativas, socioeconómicas, necesidades y problemáticas de CECAR que permita analizar el fenómeno de la deserción y poder definir estrategias a mediano y largo plazo que contrarresten los índices actuales de deserción.

2. Objetivos

2.1. Objetivo General

Analizar la deserción estudiantil a través de herramientas de la inteligencia de negocio que sirva de apoyo para la generación de estrategias que busquen mitigar las tasas de abandono estudiantil.

2.2. Objetivos Específicos

- Caracterizar los procesos del programa TAE para el análisis de la deserción en la corporación Universitaria CECAR.
- Implementar herramientas de Inteligencia de negocio que apoyen el proceso de análisis de deserción del programa de trayectoria académica TAE
- Evaluar la información del estudiante desertor con el fin de caracterizar y proyectar los estudiantes con riesgo a desertar.

3. Marco Teórico

3.1. Estado del Arte

La deserción es un problema común para todas las entidades educativas, los factores que inciden en el abandono de un programa pueden ser diversos y particulares para cada región e institución educativa de nivel superior. Dentro de los principales factores identificados por el Ministerio de Educación Nacional (MEN) están: situación económica, bajo rendimiento académico, embarazos no deseados, problemas en el interior del núcleo familiar. (Cardozo, 2013). Cambio de carrera, situación laboral y fallos en la orientación vocacional. Pinzón (2011) es importante anotar que algunos factores son ajenos a las instituciones educativas de nivel superior, pero otros pueden ser tratados y proponer soluciones a corto y mediano plazo con el fin de mejorar los índices de deserción, tal es el caso del rendimiento académico, donde las instituciones implantan estrategias de acompañamiento de tipo psicológico y académico a la población estudiantil.

Según el informe presentado por la Ministra de Educación María Fernanda Campo Saavedra (cargo que ocupa desde el 7 de agosto de 2010 hasta la fecha) “Como rinde la educación en Colombia”, las cifras de deserción en Colombia son altas, considerando que por cada 100 estudiantes que ingresan a primer semestre solo 55 se gradúan. Los porcentajes de deserción por niveles de formación mostrados en el informe son: 63.2% técnicas, 52.3% tecnologías y del 45.3% carreras universitarias. De estas últimas, las que presentan una mayor deserción son las ingenierías, arquitectura y afines.

En la Corporación Universitaria del Caribe -CECAR, según el informe de vicerrectoría académica la deserción alcanzó un 12.92% en los programas presenciales, aumentado en un 1,79% con relación al semestre anterior y un 3,62% al primer semestre de 2011, siendo mayor esta tasa de deserción en los programas de ingeniería; especialmente en ingeniería de sistemas

con un 16.35% del total. Para los programas de educación abierta y distancia la tasa de deserción fue de un 15.05%, aumentado un 4,87% con relación al semestre anterior.

Teniendo en cuenta los altos índices de deserción que se presentan en las instituciones educativas del país, los entes de control educativo del estado al igual que las instituciones educativas de nivel superior, no solo en Colombia sino en la mayoría de países, han adoptado por soluciones basadas en las tecnologías de la información con el objetivo de identificar los factores que afectan la deserción, obteniendo una caracterización y proyección eficiente, adecuada y coherente de la población desertora², así como también un análisis del rendimiento académico estudiantil, permitiendo con esto mejorar la planificación y definición de estrategias que logren disminuir las altas tasas de deserción estudiantil.

La inteligencia de negocios juega un papel fundamental para la solución de los problemas mencionados en el anteriormente; debido a que permite orientar los procesos, tecnologías y herramientas para convertir los datos en información y estos en conocimiento con el objeto de mejorar la toma de decisiones en pro de mejorar la eficiencia y calidad de los procesos del negocio.

Data Warehousing Institute. Data Warehousing (2013). Lo anterior se sustenta en primera medida en “promover el éxito de los estudiantes requiere de la implementación de procesos y mecanismos que permitan un estrecho seguimiento de sus actividades académicas” (Piedade & Santos, 2010. P.1-5). y en segunda medida sobre la hipótesis “que las actividades de un docente son esencialmente análogas a la actividades de una empresa” (Pinzón, 2011). Además las actividades empresariales y comerciales se apoyan en el conocimiento y los recursos con los que cuenta la organización. Pinzón (2011). Por ende la inteligencia de negocios se adaptaría perfectamente a los procesos académicos-administrativos de las instituciones educativas. Es importante anotar que la utilización de inteligencia de negocio en el sector educativo no solo se limita a las instituciones de nivel superior, existen muchos casos exitosos a nivel de educación básica primaria y secundaria en países como Estados Unidos. Fiorenza (2013), en donde antes de la implantación de este tipo de sistemas, las estrategias para ayudar a los estudiantes con

dificultades y con ciertas habilidades eran completamente reactivas. Hoy en día, después de la implantación, las estrategias son más proactivas en el tratamiento de deficiencias académicas de los estudiantes, como lo comenta Melinda Dixon, directora de la Escuela primaria de la avenida Livingston en Columbus "nos hacen mucho más eficiente y eficaz la enseñanza de una manera mucho más significativa de conducir" (Marianne, 2013.p. 30). y además el sistema de gestión permite aprovechar el mismo tipo de herramientas motorizadas –Inteligencia de Negocios, que ayudan a los usuarios en el mundo de la empresa a seleccionar mejores sistemas de administración y negocios, y donde los informes ayudan a identificar las debilidades de los estudiantes en áreas específicas del currículo básico, como dificultad para la suma de fracciones, de modo que los maestros conozcan qué tipo de ayuda adicional necesita el joven estudiante. Además como lo expresa. Bichsel (2012). Muchos colegios y universidades han manifestado que el uso de inteligencia de negocio ayuda al avance de áreas estratégicas, como la asignación de recursos y finanzas.

En el mismo sentido, instituciones educativas de nivel superior como la Universidad Estatal de Valdosta ubicada en Georgia (USA), han adoptado soluciones tecnológicas basadas en la Inteligencia de negocios para analizar y proyectar el rendimiento académico, proporcionando a 600 docentes y 100 miembros del personal administrativo una rica información sobre este factor con el objetivo de tomar decisiones oportunas para facilitar el éxito de los estudiantes. Las mejoras en las calificaciones parciales y finales desde la implementación del sistema han sido de un 10% a un 15%, logrando con esto hacer frente a los altos índices de deserción que estaban en un 67% anual, de forma que los estudiantes no abandonen el alma mater por problemas relacionados con el rendimiento académico. Además se ha mejorado la productividad administrativa en más de un 500%, con ayuda de herramientas de la inteligencia de negocios como los cuadros de mando (Dashboard) interactivos proporcionando la posibilidad de crear informes a partir de plantillas existentes y completar solicitudes en un tiempo promedio de 24 horas (mucho más rápido en comparación con el tiempo de respuesta promedio anterior de una a tres semanas) para profesores, jefes de departamento, decanos y la Oficina de becas de ayuda financiera. Shores (2013). La importancia de los cuadros de mando en las toma de decisiones radica en que son herramientas

que permiten realizar consultas ad hoc, visualizar rápidamente datos para exploración dinámica de patrones a través de gráficas. Chaudhuri, Dayal, & Narasayya (2011).

La Universidad de Purdue ubicada en el estado de Indiana (USA) ha implementado y desplegado un proyecto de inteligencia de negocio conocido con el nombre de “Signal”, con el fin de adoptar acciones tempranas sobre bajos desempeños en los alumnos y así evitar la deserción mejorando las tasas de graduación y demostrar la responsabilidad institucional con la sociedad. El objetivo principal del proyecto no busca sustituir al docente, sino más bien ser una herramienta para soportar el proceso de enseñanza-aprendizaje. Kimberly (2013). El sistema se basa en un algoritmo desarrollado en el alma mater denominado “Algoritmo de éxito estudiantil (SSA: Student Success Algorithm)” diseñado para proporcionar a los estudiantes alertas temprana (desde la segunda semana del semestre) de problemas potenciales, proporcionando actualizaciones de estado en tiempo real del rendimiento y esfuerzo en un curso, para que este actúe buscando asesoría o enfocándose en temáticas clave de riesgo académico¹¹. Hoy por hoy, más de 11.000 estudiantes se han visto favorecidos con el proyecto de “Signal”, y más de 50 instructores han utilizado señales en al menos uno los cursos impartidos. Estos resultados han mejorado la manera de identificar a los estudiantes en situación de riesgo y ayudarlos, se espera que el proyecto evolucione para incluir mejores modelos, estrategias de intervención más eficaces y métodos de evaluación más rigurosos.

De igual forma, el ministerio de educación de Argentina se encuentra trabajando conjuntamente con la facultad de ingeniería de la Universidad Nacional de la Plata de este país, con el objeto de definir los requerimientos para el diseño e implementación de cubos de análisis y desagregamiento por cohorte a partir de cursos, periodos, docentes y años académicos, de forma que sean explotados a través de herramientas de inteligencia de negocios como OLAP (On-Line Analytical Processing). Herramienta que ofrece a los usuarios del sistema una vista multidimensional de los datos y le permiten realizar operaciones comunes de la inteligencia de negocios tales como: filtrado, agregación, drill-down y pivotante¹⁰. Al implementar los cubos de análisis y desagregamiento por cohorte, la universidad de la plata logró abordar la problemática de la deserción estudiantil analizada desde el punto de vista del rendimiento académico con el fin

de definir distintas políticas o estrategias para trabajarla. Lo anterior le permitió al alma mater lanzar el Plan Estratégico de Ingeniería 2012 – 2016, presentando como eje principal el mejoramiento de los indicadores académicos utilizando como herramienta de explotación de la información OLAP. Universidad Nacional de La Plata (UNLP, 2012).

La investigación anterior sirvió de base para definir otro proyecto de investigación. Ostojic, Yanina & Kruzylo (2014). Encaminado a identificar patrones y perfiles de estudiantes y egresados con el objetivo de comprender mejor su comportamiento ante la utilización de las tecnologías de información y comunicación (TIC) en la educación (Plataformas virtuales, redes sociales, entre otras) y definir propuestas educativas que apoyen todo el proceso académico disminuyendo en lo posible las tasas de abandono estudiantil. Para lograr esto se utilizará otra herramienta de la inteligencia de negocios conocida como minería de datos, la cual permite un análisis en profundidad de los datos que va mucho más allá de lo ofrecido por OLAP o reportes, proporcionando la capacidad de construir modelos descriptivos y predictivo10.

De igual manera en Chile se desarrolló la investigación “Deserción estudiantil en el pregrado en la Pontificia Universidad Católica de Valparaíso” la cual determinó los patrones que tipifican el perfil de los estudiantes desertores, lo anterior basado en causas externas como las condiciones socioeconómicas (lugar de residencia, nivel de ingresos, nivel educativo de los padres, ambiente familiar, necesidad de trabajar para mantenerse o aportar a la familia), causas internas de la institución educativa (incremento en la matrícula, carencia de mecanismos adecuados de financiamiento, otorgamiento de becas, metodologías de la carreras, ambiente educativo e institucional y carencia de lazos afectivos con la universidad), causas de orden académico (formación académica previa, exámenes de ingreso, nivel de aprendizaje adquirido, excesiva orientación teórica y falta de orientación por parte de los docentes) y causas de orden personal (aspiraciones, motivaciones personales, insuficiente madurez emocional, aptitudes propias de la juventud, grado de satisfacción de la carrera y dedicación). Para lograr lo anterior se utilizó minería de datos a través de un modelo no supervisado -descriptivo (Clúster) y un modelo supervisado –

predictivo (árboles de decisión). Ocaranza & Quiroz (2013). Como resultado se obtuvo un conjunto de causas comunes en los estudiantes desertores.

Por su parte, la Universidad Tecnológica de Izúcar de Matamoros en México, para el año 2009 inició la investigación titulada “Aplicación de técnicas de minería de datos para predecir deserción” con el propósito de identificar y calcular el porcentaje de probabilidad de que un estudiante pueda desertar, desde que inicia su vida estudiantil en la Universidad y de esta manera proponer las estrategias necesarias con mucha anticipación para disminuir el índice de deserción tomando acciones preventivas o remediales Orea, Vargas, & Alonso(2009). Otras investigaciones como “Educational datamining: A case study for predicting dropoutprone students”. Kotsiantis, (2009). Proporcionan un estudio de las diferentes metodologías, basadas en la minería de datos, que se han utilizado para la deserción estudiantil y presenta un estudio experimental “Student Dropout Analysis With Application Of Data Mining Methods” ofreciendo diferentes aplicaciones de la minería de datos en las instituciones de educación superior, generando un modelo que permite la planificación estratégica de mecanismos adicionales con el fin de mejorar la eficiencia de estudio Jadrić, Garača & Čukušić (2010).

El ministerio de Educación en Colombia no es ajeno a las tendencias de utilizar las tecnologías de la información para el estudio de la deserción, en su política de “Ampliación de cobertura en educación superior” se formuló el proyecto “Estrategias para Disminuir la Deserción en las Instituciones de Educación Superior”, en el año 2005 se contrató al CEDE (Centro de Estudios sobre Desarrollo Económico) de la universidad de los Andes para realizar el estudio que tenía con fin determinar cuáles son los factores que afectan en gran medida la deserción de los programas académicos en Colombia. Como resultado se obtuvo un producto de software conocido como SPADIES (Sistema de Prevención y Análisis de la Deserción en las Instituciones de Educación Superior) orientado a prevenir y analizar el fenómeno de la deserción estudiantil en las Instituciones Educativas de nivel Superior. Pérez & Largo (2011). En el 2010 la Escuela de Marketing y Publicidad de la Universidad Sergio Arboleda creo el grupo de investigación Perceptron, centrando su investigación en el marketing educativo, haciendo gestión de

conocimiento en las bases de datos existentes en la universidad, para caracterizar el perfil de los estudiantes que ingresan y desertan en los programas de Publicidad Internacional y Marketing & Negocios Internacionales. El grupo utilizó herramientas de minería de datos para la extracción de conocimiento logrando mediante modelos descriptivos los clúster de estudiantes cursantes, egresados y desertores de los programas de Marketing & Negocios Internacionales y Publicidad Internacional. Como producto de esta investigación, se realizaron cuatro trabajos de grado, uno enfocado en el estudio del perfil de los estudiantes actuales y graduados del programa de Marketing, dos grupos se centraron en caracterizar el perfil de los estudiantes desertores de los dos programas y el último definió un modelo descriptivo para la deserción de los estudiantes del programa de Marketing y Negocios. Pinzón (2011).

Por su parte en la Universidad de Quindío se presentó en el año 2007 el trabajo de tesis doctoral titulado “Una propuesta de dashboard digital del docente como estrategia de gestión personal del conocimiento en el ámbito académico y su aplicación en la universidad de Quindío” como una estrategia tecnológica para la gestión personal de conocimiento docente, con el objetivo de mejorar su desempeño, estableciendo las bases conceptuales necesarias que intervienen en los procesos claves de enseñanza a través de un herramienta de control de mando (Dashboard) para “el manejo de la información, la comunicación, el trabajo colaborativo y la producción profesional, de tal manera que les permita afrontar los nuevos retos de la sociedad del conocimiento y ponerlo en práctica en la Universidad del Quindío” Suárez(2008).

En la Institución Universitaria Antonio José Camacho ubicada en la ciudad Cali se presentó el proyecto de tesis magisterial denominado “Guía metodológica para la implementación de un sistema de información tipo inteligencia de negocios que apoya la toma de decisiones, en el área académica, en la Institución Universitaria Antonio José Camacho” trabajo que sirve de guía para la implementación de un sistema de inteligencia de negocios para el apoyo de la toma de decisiones en el área académica, permitiendo a los usuarios finales definir estrategias para mejorar los procesos internos de la institución educativa. Como resultado se obtuvo una guía completa para implementar a través de las herramientas de inteligencia de negocios un seguimiento y control a

las labores diarias, consolidando la información a través de indicadores de gestión que miden el desempeño de cada actividad en el área académica. De igual forma se presenta la generación de informes consolidados a nivel del personal con responsabilidades en el ámbito estratégico y táctico de la organización y la automatización de labores de servicio a la comunidad académica, docentes y estudiantes (Rodríguez, 2007).

3.2. Sistemas de Información

En la dinámica de las operaciones diarias de las organizaciones el escenario característico se encuentra conformado por personas rodeadas de una infraestructura física y utilizando recursos tecnológicos, económicos, humanos y de diferente índole, realizando actividades inherentes a la naturaleza de la empresa y consecuentes con la misión y visión definidas en la misma.

Atendiendo a lo expresado anteriormente, se plantea las siguientes preguntas y sus posibles respuestas, en el contexto de una organización:

3.2.1. Preguntas Orientadoras en el contexto de una organización.

¿Qué actividad es encomendada a las personas? Entiéndase actividad “Conjunto de operaciones o tareas propias de una persona o entidad” Espanola, R. A. (2013). En este sentido, las empresas mediante contrato encomiendan y responsabilizan a personas a realizar un conjunto de actividades cada una de las cuales se componen de tareas más específicas y hacen parte de actividades más complejas (procesos empresariales – procesos de negocios). Por ejemplo, una empresa del sector productivo puede encomendar a una persona la actividad supervisar un lote de producción, esta hace parte del proceso de producción y se descompone de una serie de tareas tales como recibir lote, verificar el estado de cada uno de los productos que conforman el lote, registrar las observaciones y valoraciones obtenidas y emitir informe final.

¿Por qué son necesarias las actividades encomendadas a las personas? Toda empresa tiene sentido en la medida que provee un producto y/o servicio a la sociedad, para lograr esto diseña un conjunto de acciones y actividades (procesos) debidamente pensadas, reflexionadas y articuladas. Estas actividades se delegan al talento humano, quien apoyado con otros recursos y desplegando su Saber y Saber Hacer se le encarga y responsabiliza su ejecución. Cualquier actividad diseñada por la empresa y que conduzca y/o apoye a la obtención de un producto y/o servicio, es tan importante como la tarea de bombear sangre por parte del Corazón.

¿Qué recursos son necesarios para que las personas realicen y completen las actividades encomendadas? En la realización de las actividades son necesarios y relevantes muchos recursos, por ejemplo, en la actividad de preparar un pedido para un cliente la persona o personas encargadas requieren: herramientas que le permitan anotar los productos y cantidades solicitados por el cliente, herramientas que le permitan verificar si la empresa actualmente cuenta con los productos y la cantidad solicitada, bolsas y cajas para empacar los productos, un facturero para elaborar la respectiva factura, etc. Hasta aquí pareciera que los recursos utilizados son los necesarios y relevantes para culminar la actividad, pero si no se tienen los datos (dirección, teléfonos) que permitan ubicar al cliente para el envío del pedido, no se podría completar la actividad. Es de destacar que existen recursos (económicos, humanos, tecnológicos, logísticos, etc.), que hacen posible realizar una actividad, se destaca aquí un recurso que se origina sin ningún costo, su gestión muy a pesar de requerir cierta inversión de dinero, es superada por la utilidad que brinda para completar una actividad y tomar decisiones acertadas. Los datos se convierten así en un recurso intangible de un valor difícil de cuantificar.

¿Qué beneficios obtiene la empresa cuando las personas realizan y completan las actividades encomendadas? Las actividades que se desencadenan de una solicitud de producto por parte de un cliente, al ser completadas generan ciertos ingresos para la organización, y si son completadas con eficiencia y con altos niveles de calidad, además de los ingresos, se logra la satisfacción del cliente lo que puede garantizar una futura solicitud de productos.

En cada una de las respuestas emergen y se hacen evidentes situaciones, recursos y otros elementos que al primer contacto no son percibidos, y que nos llevan a abstraer elementos de una marcada relevancia en el logro de la homeostasis de la empresa. En este sentido, podemos identificar el concepto de Sistemas de Información.

La siguiente definición es propuesta por Andreu, Ricart y Valor (1991): “Conjunto integrado de procesos, principalmente formales, desarrollados en un entorno usuario-computador, que operando sobre un conjunto de datos estructurados de una organización, recopilan, procesan y distribuyen selectivamente la información necesaria para la operatividad habitual de la organización y las actividades propias de la dirección de la misma”.

3.2.2. Sistemas de información y la organización.

Un Sistema de Información surge en cualquier grupo social formal o informalmente conformado, indistintamente de los objetivos que persiga y la estructura adoptada, pues siempre emergerán datos de la interacción y las actividades de las personas que lo conforman. Los sistemas de Información en el marco de las organizaciones han sido objeto de estudio por muchos años, debido a la importancia que tiene la gestión de los datos para el correcto y eficiente funcionamiento de la organización. Con el advenimiento y desarrollo acelerado de las Tecnologías de la Información (TI, tanto del hardware como del software), los niveles de madurez y de estabilidad y la disminución en sus costos de adquisición, han facilitado que las organizaciones incorporen este tipo de tecnologías para apoyar la gestión de sus datos, siendo muchos los casos de éxito que han logrado las pequeñas, medianas y grandes empresas. La inversión en TI que hacen las organizaciones para apoyar su gestión de datos se justifica por el potencial de esta en la mejora de la eficacia y eficiencia del tratamiento de la información, provocando una nueva revolución en el orden social y económico.

Para ampliar la conceptualización de un Sistema de Información es necesario plantear tres conceptos ligados: Organización, Sistema e Información (ver figura 1).



Figura 1. Elementos que definen un Sistema de Información

Fuente: FUNIBER. Reingeniería, estrategia y dirección de sistemas y TIC [En línea]. <<http://coparmex.funiber.org/cursos/tecnologias-tic/reingenieria-estrategica-direccion-sistemas/>> [Citado el 3 de septiembre de 2013]

3.2.3. Dato, información y conocimiento – DIC.

Sir Francis Bacon, fue un visionario de la revolución social que atraviesa la humanidad y que se conoce como Sociedad de la Información y Sociedad del Conocimiento.

Se puede entonces deducir que la información y el conocimiento juegan un papel clave en las sociedades actuales, asimismo las tecnologías que facilitan su eficaz y eficiente gestión (las TICs). Ante esto, qué papel juegan los datos en todo este concepto, por ello, es necesario precisar las diferencias y relaciones que existen entre estos conceptos (ver figura 2).

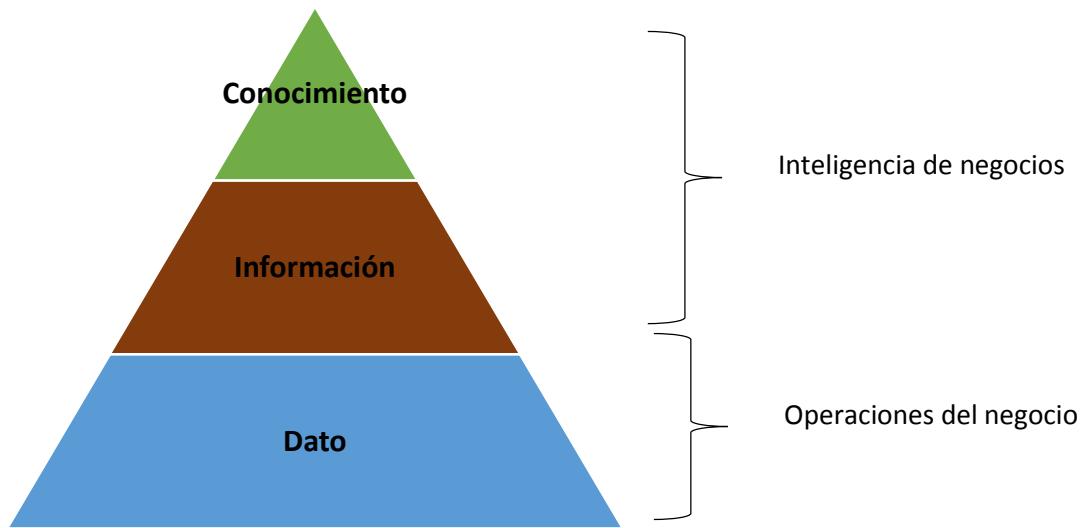


Figura 2. Datos, Información y Conocimiento – DIC
Fuente. Elaboración propia

3.2.4. Definición de Sistema de Información.

Autores han aportado en la formalización del concepto de sistema de información, se recurre a varios de ellos:

- Un sistema de información se puede definir desde el punto de vista técnico como un conjunto de componentes interrelacionados que recolectan (o recuperan), procesan, almacenan y distribuyen información para apoyar la toma de decisiones y el control de en una organización. Además los sistemas de información también pueden ayudar a los gerentes y los trabajadores a analizar problemas, visualizar asuntos complejos y crear nuevos productos.

Laudon (2004).

- Un sistema de información es un conjunto de elementos que interactúan entre sí con el fin de apoyar las actividades de una empresa o negocio. En un sentido amplio, un sistema de información no necesariamente incluye equipo electrónico (hardware). Sin embargo, en la

práctica se utiliza como sinónimo de “sistema de información computarizado” (Cohen, & Asín ,2009).

- An information system (IS) can be any organized combination of people, hardware, software, communications networks, data resources, and policies and procedures that stores, retrieves, transforms, and disseminates information in an organization. (O'BRIEN & MARAKAS, 2010).
- Conjunto de procesos ordenados que tratan información en una organización concreta para mantenerla informada.
- Medio a través del cual se recoge, trata, almacena y distribuye información para satisfacer las necesidades de usuarios, sus actividades de negocio y sus objetivos.
- Conjunto extenso y coordinado de sub-sistemas de información racionalmente integrados que transforman datos en información en una variedad diversa de formas para mejorar la productividad conforme a los estilos y características de los gerentes de una organización. “El conjunto integrado de procesos, principalmente formales, desarrollados en un entorno usuario-ordenador, que operando sobre un conjunto de datos estructurados de acuerdo con las necesidades de una organización, recopilan, elaboran y distribuyen selectivamente la información necesaria para la operatividad habitual de la organización y las actividades propias de la dirección en la misma” (ANDREU, Ricart & VALOR, 1996).

De todas las definiciones se pueden extraer tres elementos comunes
(ver tabla 1).

Tabla 1.

Elementos de un Sistema de Información

Termino	Descripción
Organización	Los sistemas de información obedecen a una estructura y están imbricados en una organización, formando una dualidad

Sistema	Conjunto de elementos con funciones formalmente definidas y son los que se encargan de las tareas básicas de un sistema de información: captura, procesamiento, almacenamiento y diseminación de la información
Información	Elemento intangible del sistema de información, su razón de ser. Para lograr su elaboración los elementos del sistema deben capturar, procesar y almacenar datos, lo que conduce a establecer una serie de relaciones y conexiones que brindan mayor valor semántico.

3.2.5. Tipos de Sistemas de Información.

De acuerdo con los autores Cohen Karen Daniel y Asín Lares Enrique, desde la perspectiva de los usuarios, los Sistema de Información se clasifican en los siguientes tipos:

3.2.6. Sistemas de Información Transaccionales (TPS).

De acuerdo con Laudon, Kenneth & Laudon (2012) este tipo de SI fueron los primeros en desarrollarse al interior de las organizaciones. Desde sus inicios, los TPS surgieron con la idea de registrar transacciones (es decir eventos que ocurren al interior de una empresa o entre la empresa y agentes externos, y que necesariamente genera o requiere datos) ordinarias cotidianas que se requieren para conducción de la empresa. Los TPS logran automatización de procesos operativos dentro de una organización.

Las características principales de estos SI son Cohen, & Lares (2009):

- Permiten ahorro significativo de mano de obra.
- Tienen una intensa entrada y salida de información, los cálculos y proceso que realizan suelen ser poco complejos.
- A través de ellos se cargan las grandes bases de datos para su posterior utilización.
- Son fáciles de justificar ante la dirección general, ya que sus beneficios son visibles y cuantificables en el corto plazo.
- Se pueden adaptar con facilidad a aplicaciones que se ofertan en el mercado

3.2.7. Sistema de Información de Apoyo a la Toma de Decisiones.

La toma de decisiones se realiza por parte de todo directivo de la empresa, independientemente del nivel jerárquico que ocupe, ya que consideramos que decidir es resolver problemas, y problemas surgen en todos los niveles de la empresa. En esta situación los datos, información y conocimientos son recurso valiosos para lograr una toma de decisiones acertada, así mismo, de acuerdo con Andreu, Ricart & Valor (1997). Los procesos de toma de decisiones están basados por lo general en la intuición, la experiencia acumulada por el decisor (construyéndose a partir de pautas repetitivas) y la información. En la empresa red, resulta especialmente significativa la complejidad en la que se han de tomar decisiones, basadas fundamentalmente en la información proveniente del consumidor y/o usuario. Los sistemas de apoyo a decisiones tienen por finalidad soportar el proceso de apoyo a la toma de decisiones al interior de la empresa, en ese sentido se define como: un conjunto de programas y herramientas que permiten obtener oportunamente la información requerida durante el proceso de la toma de decisiones, en un ambiente de incertidumbre.

Las principales características de estos sistemas son:

- Se deben introducir después de la implementación de los sistemas transaccionales.
- La información que generan sirven de apoyo a los mandos intermedios y a la alta administración en el proceso de toma de decisiones
- Suelen ser intensivos y escasos en entrada y salidas de información.
- No ahorran mano de obra.
- Son sistemas interactivos y amigables
- Apoyan tomas de decisión que por su naturaleza, son repetitivas y estructuradas, así como decisiones no repetitivas y no estructuradas

Los sistemas de apoyo a las decisiones se clasifican en:

- DSS (decision support systems), sistema de apoyo a la toma de decisiones.
- GDSS (group decision support systems), sistema para la toma de decisiones en grupo

- EIS (executive information systems) sistema de información para ejecutivos.
- EDDS (expert decision support system), sistemas expertos de apoyo a la toma de decisiones

3.2.8. Sistemas de Información Estratégicos.

De acuerdo con la Figura 3, en el nivel de la cúspide de la pirámide alude a un sistema de información para un tipo de usuario con muchas exigencias en información de calidad. Las empresas maduras y donde las TIC son vistas más como una inversión que como un gasto, pueden dirigir sus esfuerzos para lograr desarrollar sistemas de información estratégicos, que puedan lograr una ventaja competitiva a la organización.

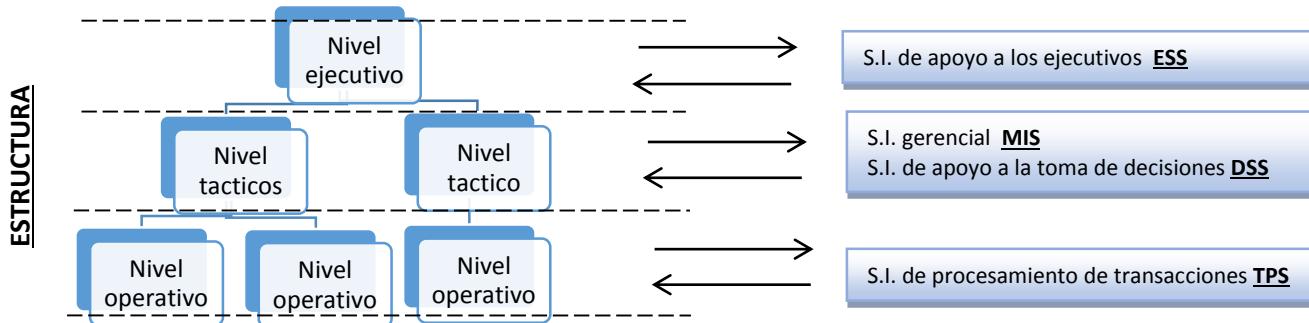


Figura 3. Sistemas de información desde la perspectiva del usuario
Fuente. Elaboración Propia

Los sistemas a nivel estratégico ayudan a los directores a enfrentar y resolver aspectos estratégicos y tendencias a largo plazo, tanto en la empresa como en el entorno externo. Su función principal es compaginar los cambios del entorno externo con la capacidad organizacional existente (Atelin, & Dordogne, 2006).

En los sistemas de información estratégicos se definen como elementos claves para respaldar o dar forma a la estrategia competitiva de la organización.

Al igual que los otros sistemas de información, los sistemas de información estratégicos, poseen una serie de características que lo permiten identificar y dirigir esfuerzos para su desarrollo al interior de la organización:

- Apoyan el proceso de innovación de productos y proceso dentro de la empresa debido a que buscan ventajas respecto a los competidores y una forma de hacerlo en innovando o creando productos y procesos.
- Surgen por la necesidad de integración de procesos y como un resultado de la maduración de la industria del software.
- Surgen por la necesidad de integración de procesos y como un resultado de la maduración de la industria del software.
- Son altamente costosos y de gran alcance.
- Típicamente su forma de desarrollo es a base de incrementos y a través de su evolución dentro de la organización. Se inicia con un proceso o función en particular y a partir de ahí se van agregando nuevas funciones o procesos

3.3. Inteligencia de Negocios

Actualmente, la gran mayoría de organizaciones cuentan con sistemas de información que permiten automatizar sus procesos operativos, proporcionar información que sirve de apoyo en el proceso de toma de decisiones y lograr ventajas competitivas a través de su implantación y uso. Estos sistemas se pueden clasificar en: transaccionales diseñados para almacenar, modificar, eliminar y recuperar todo tipo de información que es generada por las transacciones en una organización. Estos se basan en sistemas OLTP (OnLine Transaction Processing) orientados el procesamiento de un gran número de transacciones (Insert, update, Delete, Select) en tiempo real (Online). Se caracterizan principalmente por la rapidez e integridad de los datos que ofrecen. El proceso transaccional es típico de las bases de datos operacionales. (Maharaja & Sinhji, 2013).

Por otra parte, se encuentran los sistemas de información decisionales cuyo propósito fundamental es el de apoyar el proceso de toma de decisiones al interior de las organizaciones.

Estos sistemas juegan un papel fundamental en la eficiencia y eficacia de la toma de decisiones, logrando ventajas competitivas en las empresas, debido a que el poder competitivo de estas se basa en la calidad y cantidad de información que sea capaz de usar en la toma de decisiones. La inteligencia de negocios proporciona herramientas necesarias para aprovechar los datos almacenados en las base de datos operacionales de los sistemas de información transaccionales, para ser utilizadas en la disminución de la incertidumbre en la toma de decisiones y reducir el impacto negativo que puede conllevar una mala decisión.

El término “*inteligencia de negocios*” (en inglés: “*Business Intelligence*” B.I.) data del año 1958 del artículo publicado por Hans Peter Luhn titulado: “*A Business Intelligence System*” donde la define como “*the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal.*” (Luhn, 1958). Adaptando esta definición a un contexto más actualizado se refiere a la habilidad de aprender a partir de hechos pasados con el objetivo de mejorar la estrategia del negocio para crear ventajas competitivas con respecto a la competencia. Otros autores como. Dresner (1989). La define como to describe a set of methods that support sophisticated analytical decision making aimed at improving business performance. Entendiéndose como un conjunto de métodos dirigidos a apoyar el proceso de toma de decisiones de manera acertada y estratégica para el éxito del negocio, basado en el análisis de la información que se produce en el interior de la organización. Según “*Data Warehouse Institute*”, la inteligencia de negocios es... la combinación de tecnología, herramientas y procesos que me permiten transformar mis datos almacenados en información, esta información en conocimiento y este conocimiento dirigido a un plan o una estrategia comercial. La inteligencia de negocios debe ser parte de la estrategia empresarial, esta le permite optimizar la utilización de recursos, monitorear el cumplimiento de los objetivos de la empresa y la capacidad de tomar buenas decisiones para así obtener mejores resultados.... Por ende, la inteligencia de negocio no es más que un conjunto de tecnologías, herramientas y métodos enfocados a la transformación de los datos acumulados de la empresa en conocimiento, con el objetivo de apoyar las estrategias de la organización para lograr ventajas competitivas ofreciendo a las organizaciones una visión interna de los procesos, con información valiosa para determinar amenazas, fortalezas, debilidades y oportunidades a los

directivos en los niveles tácticos y estratégicos de la pirámide organizacional produciendo una ventaja sobre los competidores (Maureen & Fernández, 2009).

La implementación de herramientas de inteligencia de negocio en la empresa tuvieron un crecimiento del 7% para el año 2012, reportando ingresos de 13.1 millones de dólares las compañías proveedoras de este tipo de herramientas (Ver tabla 2). Este crecimiento evidencia el papel que juegan en las empresas para el éxito de su negocio los sistemas de inteligencia de negocios. Aunque comparado con cifras de los años anteriores ha sufrido una desaceleración producto de las difíciles condiciones económicas y en algunos casos la confusión provocada por palabras que han sonado mucho en la industria como Big Data. Gartner (2013). En la tabla 2 se observa las 5 empresas líderes a nivel mundial en ofrecer sistemas de inteligencia de negocios, encabezando la lista SAP, Oracle e IBM, teniendo una cuota de mercado de 22.1%, 14.9% y 12.4% respectivamente. De la tabla 2 se puede observar que en el caso de SAP solo se tuvo un crecimiento anual 0,6%, según el reporte de Gartner esto se debe en parte a que BusinessObjects está “perdiendo impulso”, aun así SAP tiene un futuro prometedor para la inteligencia de negocios debido a su amplio portafolio de servicios, fidelización de sus clientes, fuerte organización de ventas y un ecosistema de socios.

Una arquitectura típica de Inteligencia de negocios está compuesta por cinco componentes básicos (Ver Figura 4): Como primer elemento se encuentran las fuentes de donde provienen los datos, típicamente a partir de múltiples bases de datos transaccionales, como segundo elemento las diferentes fuentes contienen datos de diversa calidad, representación, códigos y/o formatos, por lo tanto se necesita de una capa de software que permita extraer, transformar y cargarlos a un repositorio esta capa es conocida como ETL. Como tercer elemento un repositorio de datos centralizado denominado Bodega de datos (en inglés: DatawareHouse. DW.). Como cuarto elemento se encuentran servidores de nivel medio que complementan los servidores de bodegas de datos proporcionando funcionalidad especializada para diferentes escenarios BI. Por último, se encuentran aplicaciones a través de las cuales los usuarios realizan las tareas de Inteligencia de negocios: hojas de cálculo, rendimiento de aplicaciones de gestión que permitir a los tomadores

de decisiones clave para el seguimiento de indicadores de desempeño del negocio utilizando cuadros de mando visuales, herramientas que permiten que los usuarios plantean consultas ad hoc, visores para modelos de minería de datos, visualización rápida ad hoc de datos puede permitir la exploración dinámica de patrones, valores atípicos y ayudar a descubrir hechos relevantes para BI. (Data Warehousing Institute, 2013).

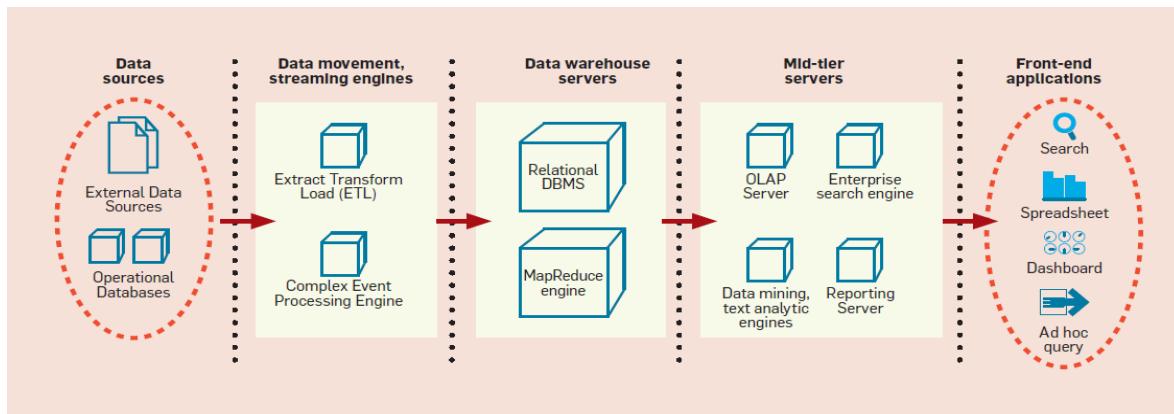


Figura 4. Arquitectura de Inteligencia de negocio típica.

Fuente: Chaudhuri, S., Umeshwar, D., & Narasayya, V. (2011). An overview of Business intelligence technology. Communications of the ACM , 88-98.

Tabla 2.

Las primeras 5 suite de inteligencia del negocio en el periodo 2011-2012

Company	2012 Revenue	2012 Market Share (%)	2011 Revenue	2011-2012 Growth (%)
SAP	2,902.5	22.1	2,884.0	0.6
Oracle	1,952.1	14.9	1,913.5	2.0
IBM	1,625.6	12.4	1,478.8	9.9
SAS	1,599.7	12.2	1,542.9	3.7
Microsoft	1,189.3	9.1	1,059.9	12.2
Others	3,861.90	29.3	3,416.00	13.0
Total	13,131.1	100.0	12,295.1	6.8

Note: SAP reports in Euros, and faced currency head wind that hampered growth in USD.

Fuente. Abello, A., & Samos, J. (17-19 De Julio De 2002). Yam2 (Yet Another Multidimensional Model): An extension of UML. *Database Engineering and Applications Symposium, 2002. Proceedings. International*, 172 - 181.

3.3.1. Fuentes de Datos.

Las fuentes de datos que alimentan el sistema de inteligencia de negocios pueden ser muy diversas, principalmente bases de datos internas de sistemas de información transaccionales de las organizaciones, así como también bases de datos externas (No pertenecientes a las empresas y que puede ser adquiridas gratuitamente (P. E. bases de datos de censos) o través de la cancelación de alguna suma de dinero (P.E. bases de datos de investigaciones de mercado)). Al igual que archivos de texto, bases de datos no relacionales (NoSQL), hojas de cálculo, registros del servidor Web con datos de navegación de usuario, entre otras. (Oracle, 2013).

3.3.2. ETL (Extraction, Transform, Load).

Capa de software que se encarga del proceso de integración y preparación de las diferentes fuentes de datos de una organización, a través de la extracción, transformación y carga de datos hacia un repositorio denominado Bodega o almacén de datos (en inglés: “DatawareHouse”. DW.).

El proceso de carga de datos se lleva a cabo de forma incremental y se puede configurar para que se realice de forma automática en períodos de tiempos definidos. Kimball & Ross (2013). Esto hace eficiente la capacidad de carga de datos y la actualización escalables imperativo para la inteligencia de negocios empresarial.

Por otra parte existe una necesidad cada vez mayor por apoyar las tareas de la inteligencia de negocios en tiempo real, es decir, tomar decisiones de negocio basadas en los propios datos operativos. Debido a esto se han implementado motores especializados conocidos como CEP (Complex Event Processing) para apoyar este tipo de escenarios¹⁰.

3.3.3. Bodegas o Almacenes de Datos (en Inglés: Datawarehouse. DW.).

Bill Inmon fue uno de los primeros autores en definir Bodegas de datos en el año de 1992 como: “...una colección de datos integrados orientados a temas, integrados, no-volátiles y variables en el tiempo, organizados para soportar necesidades empresariales...”, Ralph Kimball la definió en el año 2002 como: “...una colección de datos en forma de una base de datos que guarda y ordena información que se extrae directamente de los sistemas operacionales (ventas, producción, finanzas, marketing, etc.) y de datos externos⁴⁵. Partiendo de lo anterior, las bodegas o almacenes de datos se pueden definir como bases de datos que almacenan, integran, adicionan y clasifican la información que se extrae de los sistemas transaccionales y datos externos (Datos que no se encuentran en la base de datos transaccional sino en formatos de hojas de cálculo, archivos de texto, entre otros). La información almacenada presenta la característica de “No volátil”, es decir, no se modifica ni se elimina la información una vez almacenada, solo se consulta.

Existen en el mercado diferentes opciones de repositorios para bodegas o almacenes de datos, una opción popular son los motores de bases de datos relacionales (RDBS: Relation Data Base System) que en las últimas dos décadas han desarrollado nuevas estructuras de datos (P.E. Arrays), optimizadores de consultas (P.E. Índices de Mapa de bits, índices de Join, técnicas de particionamiento de datos) y mejoras en el procesamiento y tiempo de respuestas de consultas complejas SQL (Structured Query Language) sobre grandes volúmenes de datos, requisito clave

para la inteligencia de negocios. Otra opción son los motores basados en MapReduce. (Dean & Ghemawat, 2008). Paradigma (originalmente construido para el análisis de los documentos Web y registros de consultas de búsqueda web) utilizado recientemente para el análisis empresarial extendido para soportar consultas compleja SQL necesarias para escenarios de bodegas de datos¹⁰. El modelo de datos de una bodega de datos se basa en un modelo multidimensional que permite el almacenamiento y recuperación eficiente de grandes volúmenes de datos, vistos y analizados desde diferentes puntos de vista, denominados dimensiones. El modelo posee un conjunto de medidas numéricas que son los objetos de análisis identificados a través de valores de un dato en particular (P.E Medidas como ventas, presupuesto, ingresos e inventario) y donde cada una de estos está asociado a un conjunto de dimensiones que proporcionan el contexto para medir y describir de las características que lo definen. Fradel, Orlando & Pérez Castillo (2007). Las dimensiones a su vez se describen a partir de un conjunto de atributos relacionados a través de un orden jerárquico¹⁰. El modelo en general se representa vectorialmente: los hechos se especifican en una celda ubicada en la intersección de ciertas coordenadas según el modelo de coordenadas (x, y, z,...), donde cada una representa una dimensión. En la figura 5, se observa la representación de un modelo multidimensional a través de un cubo dimensional, donde las dimensiones se representan por medio de coordenadas. El intersección de las uniones simboliza los hechos y cada hechos está compuesto por medidas.

Actualmente existen varias representaciones para el modelo multidimensional de datos entre las que tenemos: Estrella/ER (Tryfona, Busborg, & Borch Christiansen, 1999), Kimball (Kimball & Ross, 2013), Multidimensional/ER (Sapia, Blaschka, Höfling, & Dinter, 1998), GOLD (Luján-Mora & Trujillo, 2004) (Trujillo, 2000), y YAM2 (ABELLO & SAMOS, 2002) y Husemann (Husemann , Lechtenbörger , & Vossen, 2000). De igual forma que en la arquitectura de los modelos de bases de datos (Elmasri & Navhate, 2011), la estructura del modelo multidimensional está compuesta por tres niveles: lógico, conceptual y físico.

3.3.3.1 Nivel Lógico.

Este nivel se basa en el modelo relacional para el diseño de las tablas de hechos, dimensiones, atributos, jerarquías y su interacción. Además permite definir los procesos de extracción, transformación y carga (ETL) y del mapeo entre las fuentes de datos y la bodega o almacén de datos.

3.3.3.2 Nivel Conceptual.

Este nivel representa las interacciones entre las entidades y relaciones a través de un lenguaje de definición de objetos (en inglés: Object Definition Language ODL) haciendo una abstracción del mundo real de forma que el usuario represente sus ideas.

3.3.3.3 Nivel Físico.

Este nivel representa el código necesario para la generación de las tablas de hechos, dimensiones y las reglas de integridad.

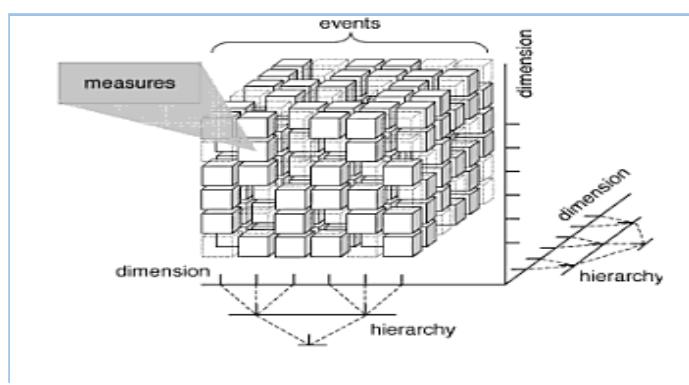


Figura 5. Cubo dimensional

Fuente: Data Warehouses and OLAP: Concepts, Architectures, and Solutions. (2007). Idea Group Inc

3.3.4. Servidores Intermedios.

Las bodegas o almacenes de datos se complementan con un conjunto de servidores de nivel medio (en inglés: “*Mid-tier Servers*”) que proporcionan funcionalidades especializadas para diferentes escenarios de los procesos de la Inteligencia de negocios.

El servidor de procesamiento analítico en línea (en inglés: “*OLAP*”) permite de manera eficiente mostrar una vista multidimensional de los datos para los usuarios o aplicaciones. De igual forma también proporciona la ejecución de operaciones comunes de la Inteligencia de negocios tales como: filtrado (en inglés: filtering), agregación (en inglés: aggregation), drill-down, roll-up, Slice, Dice y pivoting. La tendencia en el mercado es la utilización de “ *motores en memoria*” (en inglés: *in-memory BI*) explotando los grandes tamaño de memoria principal que poseen los equipos de cómputo mejorando drásticamente el rendimiento de consultas multidimensionales.

Los servidores de reportes (en inglés: “*Reporting servers*”) permiten la definición, ejecución y rendición eficiente de informes. Por su parte, los motores de búsqueda empresarial (en inglés: “*Enterprise search engines*”) apoyan la búsqueda de palabras claves sobre texto y/o estructuras de datos en la bodega o almacén de datos (P.E. buscar mensajes de correo electrónico, documentos, historial de compras y/o llamadas de soporte técnico), convirtiéndose en una valiosa herramienta en la última década.

Por último, los servidores de minería de datos (en inglés: “*Data mining engine*”) permiten realizar un análisis en profundidad de los datos que va mucho más allá de lo ofrecido por los servidores de OLAP o de reportes, ofreciendo la capacidad de construir modelos descriptivos y predictivos. Los motores analíticos de texto (en inglés: “*Text analytic engines*”) pueden analizar grandes cantidades de datos de texto (P.E respuestas de encuestas, comentarios de los clientes) y extraer información valiosa para la empresa (P.E. Productos que son mencionados en las respuestas de una encuesta y los temas de discusión en torno a estos).

3.3.5 Aplicaciones FrontEnd.

Existen en el mercado varias aplicaciones que permiten acceder a los datos para realizar tareas de la inteligencia de negocios entre las cuales tenemos: hojas de cálculo (en inglés: Spreadsheets), centros de mandos visuales (en inglés: Dashboards) que ofrecen a los directores información clave para el seguimiento de indicadores de desempeño del negocio, herramientas OLAP que permiten a los usuarios realizar consultas ad hoc sin el conocimiento de modelos de datos o lenguajes de consulta, visores para modelos de minería de datos permitiendo la exploración dinámica de patrones, valores atípicos y ayudar a descubrir y predecir hechos relevantes para el negocio¹⁰. A continuación se describen en detalle cada una de las aplicaciones mencionadas anteriormente.

3.3.5.1. OLAP (*OnLine Analytical Processing*).

En el año 1993 Codd definió: OLAP es un tipo de procesamiento de datos que se caracteriza, entre otras cosas, por permitir el análisis multidimensional de datos” además agregó “Las herramientas OLAP pueden empoderar a los usuarios-analistas para realizar fácilmente los tipos de análisis que previamente no los realizan debido a su aparente complejidad. De lo anterior se puede abstraer que OLAP parte del hecho que para hacer un buen análisis de la información se debe hacer de forma multidimensional, es decir tener en cuenta varias dimensiones o puntos de vista. Además las herramientas permiten realizar análisis de información de forma sencilla sin tener un previo conocimiento en lenguajes de consultas o en la estructura interna de los modelos de datos. Los sistemas OLAP se caracterizan principalmente por poseer un volumen relativamente bajo de transacciones y ser solo de lectura. (Bhavnagar & Krishnakumar, 2013) Tradicionalmente los sistemas OLAP se clasifican en:

3.3.5.1.1. ROLAP (*Relational Online Analytical Processing*).

Implementación OLAP que almacena los datos en una base de datos transaccional proporcionando técnicas para mejorar el rendimiento (índice de mapas de bits, índice de JOIN, técnicas de particionamiento y optimizadores de consultas). La arquitectura está compuesta por 3 niveles: ROLAP se utiliza una arquitectura de tres niveles: una base de datos relacional que maneja el almacenamiento de datos, un motor OLAP, el cual proporciona la funcionalidad analítica y alguna herramienta especializada empleada para el nivel de presentación. Tiene como principal ventaja permitir el análisis de una gran cantidad de datos y como desventaja ofrecer tiempo de respuestas altos comparado con las otras categorías de OLAP.

3.3.5.1.2. MOLAP (Multidimensional Online Analytical Processing).

Implementación de OLAP que almacena los datos una base de datos multidimensional (P. E. Bodega de datos). La información se encuentra precalculada logrando tiempo de respuesta muy bajos comparada con ROLAP. Además disponen de estructuras de almacenamiento específicas (arrays) y técnicas de compactación de datos que favorecen el rendimiento del sistema.

3.3.5.1.3. HOLAP (Hybrid Online Analytical Processing).

Implementación de OLAP hibrida la cual permite almacenar una parte de los datos en motores de bases de datos transaccionales y la otra parte en motores de bases de datos multidimensionales.C

Las principales operaciones ofrecidas por estos sistemas son: slice, dice (Proyección y Selección), drill-down (Información de lo general a específico), drill-up (Información de lo específico a general), roll-accros (Eliminar un criterio de una consulta), drill-across (Aregar un criterio a una consulta) y Pivot (Cambiar el orden de visualización de filas a columnas o viceversa. (Hüsemann, Lechtenbörger & Vossen, 2000.p 1- 6).

3.3.6. Minería de Datos.

- El término “*Minería de datos*” (en inglés: “*DataMining*”) fue definido por primera vez en el año 1988 por la enciclopedia británica Merriam-Wester como: “*the practice of searching through large amounts of computerized data to find useful patterns or trends*”. Fayyad la define como: “*un proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos*”. Por ende la minería de datos se puede entender como la exploración y análisis de grandes volúmenes de datos, con el objetivo de obtener patrones, tendencias y reglas de comportamiento generando nuevos conocimientos a las organizaciones logrando ventajas competitivas.

La minería de datos se basa en un proceso mucho más grande conocido como Extracción de conocimiento en Bases de datos o Descubrimiento de Conocimiento en Bases de Datos, termino (en inglés: “*Knowledge Discovery in Databases*”). KDD) que fue utilizado por primera vez en Workshops “*Knowledge Discovery in Databases*” organizado por Gregory Piatetsky-Shapiro en el año de 1989, donde se expuso el proceso de extraer conocimiento a partir de la bases de datos explicando que el Knowledge discovery in databases poses many interesting problems, especially when databases are large. Such databases are usually accompanied by substantial domain knowledge which can significantly facilitate discovery. Access to large databases is expensive – hence the need for sampling and other statistical methods. Finally, knowledge discovery in databases can benefit from many available tools and techniques from several different fields including expert systems, machine learning, intelligent databases, knowledge acquisition, and statistics. Para Han y Kamber (2006) el proceso de descubrimiento conocido como Knowledge Discovery in Databases (KDD), se refiere al proceso no-trivial de descubrir conocimiento e información potencialmente útil dentro de los datos contenidos en algún repositorio de información. No es un proceso automático, es un proceso iterativo que exhaustivamente explora volúmenes muy grandes de datos para determinar relaciones (p, 43). Por su parte Melanie J. Norton la define como (2000).el Descubrimiento de Conocimiento en Bases de Datos (KDD) hace

referencia a una serie de procesos involucrados en la extracción de información utilizable desde cualquier colección de datos en cualquier formato o medio. Por ende el Descubrimiento de Conocimiento en Bases de Datos es un proceso que consta de una serie de etapas, entre las cuales se encuentra la “*Minería de datos*”. Es importante anotar que algunos autores utilizan sin distinción ambos términos para hacer referencia a lo mismo (Jiawei, Yandong & Nick, 1992). Aunque. Molina, Félix & Carlos, Luis (2013) explica que la diferencia en la utilización de los términos existe debido a que desde punto de vista académico, el término minería de datos se considera como una etapa dentro de un proceso mayor denominado Extracción de conocimiento en Bases de datos o Descubrimiento de Conocimiento en Bases de Datos y en el entorno empresarial ambos términos se usan de manera indistinta.

El resultado del proceso tiene como fin extraer el conocimiento a partir de los datos almacenados en las bases de datos, identificando modelos y patrones válidos, nuevos y potencialmente útiles con el objetivo de mejorar la toma de decisiones en la organización.

En la figura 6, se observa las diferentes etapas del proceso:

- a. Entendimiento del dominio del problema a partir de los requerimientos de los usuarios.
- b. Selección de las fuentes de datos para la realización del proceso de descubrimiento.
- c. Integración, recopilación, Limpieza y transformación de los datos
- d. Minería de datos seleccionando la tarea de descubrimiento a realizar: clasificación, agrupamiento, reglas de asociación, entre otras. Además la selección de algoritmos a aplicar.
- e. Transformación de los datos al formato requerido por el algoritmo escogido de explotación de datos.
- f. Encontrar patrones y modelos interesantes para apoyar los procesos de toma de decisiones.
- g. Evaluación e interpretación de los patrones descubiertos y presentación de los mismos mediante técnicas de visualización.
- h. Gestión del conocimiento descubierto a las personas interesadas.

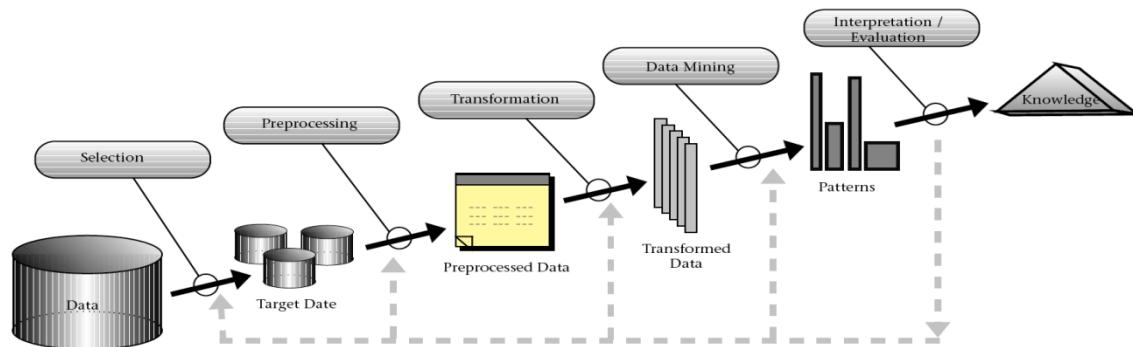


Figura 6. Proceso Extracción de conocimiento en Bases de datos

Fuente: FICA, E. (5 de 06 de 2011). *bit.* Recuperado el 18 de 12 de 2013, de <http://bituchile.com/2011/06/data-mining-concepto-bit/>

3.3.6.1. *Modelos de minería de datos.*

El modelo es una representación de relaciones y patrones entre datos para realizar descripciones que permitan dar explicación a eventos pasados o para realizar predicciones a futuro. En base a esto, todo desarrollo de aplicaciones de Minería de Datos, se identifica bien sea con un modelo descriptivo o uno predictivo y la elección de las tareas y técnicas de Minería de Datos deben poder resolver dicho modelo escogido. Linoff, Gordon y Berry Michael (2010).

Los modelos descriptivos se caracterizan por realizar descripciones de los datos con el objetivo de encontrar patrones, asociaciones y grupos de información proporcionada información de los datos oculta en estos, debido a la complejidad y tamaño o por el modelo de datos de las tablas que los registran. Moreno (2013) afirma que Cuando una aplicación no es lo suficientemente madura no tiene el potencial necesario para una solución predictiva, en ese caso hay que recurrir a

los métodos descriptivos que descubren patrones y tendencias en los datos actuales (no utilizan datos históricos). El descubrimiento de esa información sirve para llevar a cabo acciones y obtener un beneficio (científico o de negocio) de ellas.

Un modelo descriptivo proporciona información sobre las relaciones entre los datos y sus características generando información como por ejemplo:

- Los clientes que compran pañales suelen comprar cerveza.
- El tabaco y el alcohol son los factores determinantes de una enfermedad específica.
- Los clientes sin televisión y con bicicleta tienen características de comportamiento muy diferentes del resto.
- Conocer cuáles son los clientes de una organización (características de los mismos).
- Encontrar los productos que frecuentemente se compran juntos.
- Síntomas de enfermedades que se presentan juntos.

Los modelos predictivos tienen como fin obtener un modelo que permita hacer predicciones de futuros comportamientos de los datos.

Un modelo predictivo responde preguntas sobre datos futuros como:

- ¿Cuáles serán las ventas el año próximo?
- ¿Será esta transacción fraudulenta?
- ¿Qué tipo de seguro es más probable que contrate el cliente X?
- Encontrar el perfil del comprador del producto A
- Encontrar el perfil del cliente que me abandonará el mes siguiente
- Calcular el valor potencial de un cliente
- Probabilidad de que un cliente devuelva un préstamo

Se recomiendan unos aspectos que se deben tener en cuenta en su antes de empezar el proceso de minería de Datos:

- Determinar qué tipo de tarea de Minería de Datos es la más apropiada.
- Elegir el tipo de modelo.
- Elegir el algoritmo de Minería de Datos que resuelva la tarea y obtenga el tipo de modelo que estamos buscando.

3.3.6.2. *Tareas de la Minería de Datos.*

Las tareas en la minería de datos apuntan a la solución de diferentes problemas mediante un algoritmo. Estas se clasifican según el modelo al que pertenezcan. (Perez & Santin, 2010, p, 8-11)

- Tareas predictivas:
 - Clasificación.
 - Regresión.
- Tareas descriptivas:
 - Agrupamiento (o segmentación)
 - Reglas de asociación

Clasificación

Consiste en examinar las características de una entidad nueva y asignarle una clase predefinida. Por ejemplo: Clasificar a un nuevo cliente según su riesgo de crédito (alto, medio, bajo). (Linoff, Gordon & Berry Michael, 2010, p, 195-205).

Teniendo en cuenta lo anterior, la clasificación se puede entender como la asignación de un dato o registro a una clase determinada, a partir de las características que cumpla, esta clase es evalúa con anterioridad y se establece según las necesidades del problema que se pretenda resolver. Un ejemplo de este tipo de problemas sería encontrar un modelo que a la vista de un histórico de clientes clasificados como “buenos”, “regulares” y “malos”, establezca qué tipo de cliente es uno nuevo.”

Básicamente la clasificación establece la construcción de una lógica que tome un registro sin clasificar y luego lo clasifique en una de las clases existentes, de manera que al final cada uno de los registros existentes en la base de datos se encuentre clasificado. La clasificación trata con problemas de salidas discretas (si o no, alto, medio o bajo riesgo, responderá o no responderá).

Regresión

La regresión asigna un registro a una clase especificada con anterioridad, al igual que la clasificación, con la diferencia que la variable es continua. El objetivo es proyectar los valores de una variable continua a partir de la evolución sobre otra variable continua, generalmente el tiempo. (Vieira, Ortiz, Valencia & Segundo, 2009, p ,45-47).

El algoritmo de regresión permite establecer una función real que permite tomar un registro numérico y luego de aplicarlo retorne un valor real que refleje un comportamiento o parámetro.

Agrupamiento (o segmentación)

El objetivo de la tarea es el descubrimiento de grupos de registros que cumplen características o atributos similares, pero basados en los mismos datos o registros y en donde no hay clases predefinidas. Se diferencia de la clasificación en que no se conocen ni las clases ni su número. Boris (2005)

Así el agrupamiento tiene como objetivo el segmentar a un grupo diverso y disperso en un conjunto de subgrupos, los cuales presentan diferencias marcadas.

Reglas de asociación

Una asociación ocurre entre dos atributos cuando la frecuencia que se den dos valores determinados de cada uno conjuntamente es relativamente alta, es decir, determinar la dependencia de eventos. La asociación, tiene como fin identificar relaciones no explícitas entre atributos *categóricos*, reconociendo que existen grupos de atributos que determinan la presencia de otro

grupo. Por ejemplo: que productos debieran ir juntos en un supermercado, recomendaciones de productos.

Técnicas de Minería de Datos

Las técnicas son la particularización de las tareas anteriormente mencionadas. Cada particularización se transforma en un algoritmo de minería de datos. Existen una gran cantidad de algoritmos, muchos conocidos y aplicados en diferentes áreas de las ciencias, pero cada uno de estos puede presentar variaciones que permiten solucionar problemas específicos para la extracción de conocimiento. Por medio de un algoritmo es posible resolver varias tareas, pero habrá ventajas e inconvenientes que permiten identificar cual algoritmo es el más apropiado (ver tabla 3).

Las técnicas de Minería de Datos crean modelos que son predictivos y/o descriptivos. Existen diferentes algoritmos y variaciones de los mismos, así como restricciones que hacen que la efectividad de un algoritmo dependa del dominio de aplicación, no existiendo un método universal aplicable a todo tipo de situación. Estos algoritmos es lo que se conoce como técnica, ya que esta sería la forma de alcanzar la creación del modelo que se está buscando.

Se definen una serie de atributos que se pueden evaluar para hacer una elección de los algoritmos más convenientes: Linoff & Berry (2010)

- **Generalidad:** Se refiere a la posibilidad de aplicar la técnica a múltiples tipos de problemas, afectados por un amplio grupo de variables y con el empleo de varios tipos de datos.
- **Adaptabilidad:** Se refiere a la existencia o no de una fase de preproceso sobre los datos. Ésta es una de las etapas más costosas del proceso de obtención de conocimiento, por ello, resulta conveniente emplear aquella técnica que permita tomar los datos directamente de la base de datos, sin necesidad de cambios previos.
- **Precisión:** Mide la capacidad del algoritmo de obtener al resultado correcto. Es decir, se trata de evaluar el grado de error presentado en la respuesta.

- **Claridad:** Mide la transparencia de una técnica, en cuanto a la interpretación de los resultados potenciales, así como el método en que se ha llegado a los mismos. Una red neuronal es un claro ejemplo de una técnica que genera modelos de caja negra, en los cuales es difícil conocer cómo se producen las transformaciones internas que hacen llegar al resultado final.
- **Utilidad:** La información tiene un valor que decrece con el paso del tiempo, por ello, resulta necesario que la técnica empleada genere información en un formato fácil de entender, con el objetivo de que después de la interpretación del personal encargado de la toma de decisiones se convierta en un recurso de acción para la empresa. En decir, el indicador mide la forma de presentación de los resultados. Ésta se mejora con el empleo de técnicas de visualización, entre otros.
- **Facilidad de construcción:** La construcción del modelo suele ser bastante autónoma con relación a quien tome las decisiones. Sin embargo, también se ha de medir el coste de construcción en base a otros elementos, como el consumo de registros que necesita para el entrenamiento.
- **Rapidez:** Los algoritmos pueden tardar 5 minutos en proporcionar una respuesta al problema planteado o un día o más. Por esto, a partir de las características de dinamismo del mercado virtual, éste es un elemento a tener en cuenta. Depende en gran medida del hardware en el que se ejecute la herramienta.
- **Gestión de memoria:** Mide la necesidad de recursos de computación que son necesarios para que la herramienta pueda actuar correctamente. Generalmente, rapidez y recursos necesarios van de la mano, es decir, aquellos algoritmos más lentos son los que necesitan.
- **Robustez:** ¿En qué medida es capaz el algoritmo de trabajar con datos perdidos o con errores, sin que afecte significativamente a alcanzar el resultado óptimo? Se trata de un atributo relacionado con el preproceso, puesto que reduce la necesidad de éste.
- **Validación:** Se refiere a la facilidad para comprobar que el modelo ha llegado a la solución óptima. Así, las técnicas estadísticas disponen de buenos indicadores, como el coeficiente de determinación (R^2). Sin embargo, en la minería de Datos resulta una tarea más

compleja. Habitualmente cada herramienta construye sus propios indicadores, o se emplea la validación cruzada.

- **Disponibilidad:** Algunas técnicas están más disponibles en los distintos paquetes comerciales que otras. Así, las redes neuronales y los árboles de decisión son algoritmos usuales, mientras que los algoritmos genéticos difícilmente se encuentran.

Las técnicas más comúnmente mencionadas y correspondientes a las tareas de clasificación, regresión, agrupamiento y asociaciones son las siguientes: Linoff, & Berry (2010).

- **Redes neuronales artificiales:** Son modelos predecibles no-lineales que aprenden a través del entrenamiento y semejan la estructura de una red neuronal biológica. Estos son modelos de predicción no lineales que aprenden como detectar un patrón para emparejar un perfil particular a través de un proceso de entrenamiento que envuelve aprendizaje iterativo, utilizando un conjunto de datos que describe lo que se quiere encontrar.
- **Árboles de decisión:** Estructuras en forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Métodos específicos de árboles de decisión incluyen Árboles de Clasificación y Regresión (CART: Classification And Regression Tree) y Detección de Interacción Automática de Chi Cuadrado (CHAI: Chi Square Automatic Interaction Detection).
Un árbol de decisión es una estructura en forma de árbol que visualmente describe una serie de reglas (condiciones) que causan que una decisión sea tomada.
- **Algoritmos genéticos:** Técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño basado en los conceptos de evolución.
Los algoritmos genéticos son técnicas de optimización que pueden ser utilizadas para mejorar otros algoritmos de Minería de Datos obteniendo como resultado el mejor modelo para una serie de datos. El modelo resultante es aplicado a los datos para descubrir patrones escondidos o para realizar predicciones.

- **Regresión Lineal:** Corresponde a una de varias técnicas estadísticas utilizadas para identificar patrones, los cuales pueden ser utilizados para predecir.

Tabla 3.

Relación entre tareas y técnicas

TÉCNICA	PREDICTIVO		DESCRIPTIVO		
	Clasificación	Regresión	Agrupamiento	Reglas de Asociación	Correlación / Factorización
Redes neuronales	✓	✓	✓		
Árboles de decisión ID3, C4.5, C5.0	✓				
Arboles de Decisión CART	✓	✓			
Otros Árboles de Decisión	✓	✓	✓	✓	
Redes de Kahonen			✓		
Regresión lineal y Algorítmica		✓			✓
Regresión logística	✓			✓	
Kmeans			✓		
A priori				✓	
Naïve Bayes	✓				
Vecinos más Próximos	✓	✓	✓		
Análisis Factorial y de Comp. Pples.					✓
Twostep, Cobweb			✓		
Algoritmos Genéticos y evolutivos	✓	✓	✓	✓	✓
Máquinas de vectores soporte	✓	✓	✓		
CN2 rules (cobertura)	✓			✓	
Análisis discriminante multivariante	✓	✓			

3.3.7. Datamart.

Una empresa pequeña, un departamento o área específica de una empresa grande tienen la posibilidad de crear un subconjunto de datos derivado del Data Warehouse conocido como Datamart, diseñado para soportar requerimientos analíticos específicos de una determinada unidad de negocio con contenidos particulares, volumen de datos limitado y una alcance histórico menor.

3.3.8. Inteligencia de negocio el sector de la educación.

Las tecnologías de la información (Del inglés: Technology Information) descritas anteriormente mejoran la calidad de los procesos y proveen de una mejor competitividad a las organizaciones, debido a que estas permiten suministrar información relevante para la toma de decisiones a través de las herramientas de inteligencia de negocios, proporcionando a los usuarios la información que necesitan en cada momento con el fin de responder preguntas de negocio y tomar decisiones en los niveles tácticos y estratégicos.

Por ende, la implementación de sistemas de inteligencia de negocios en las instituciones educativas permitirá la toma de decisiones en la gestión de calidad de los procesos académico – administrativos con el menor grado de incertidumbre posible bajo ambientes controlados, permitiendo a los centros educativos ser más proactivos y menos reactivos, logrando con esto una ventaja competitiva y una actualización de las tecnologías de la información. De igual forma las herramientas de inteligencia de negocios proporcionarían a las instituciones información de calidad relacionada a:

- El análisis de los factores que inciden en la deserción de los estudiantes a través de técnicas de minería de datos.
- La proyección de estudiantes en riesgo de desertar a través de técnicas de minería de datos.
- La evolución del rendimiento estudiantil a través de informes OLAP relacionándolo a variables de: género, situación socioeconómica, estudios previos, grupo racial, núcleo familiar, entre otras.
- La creación de perfiles estudiantiles con el objetivo de definir y predecir causas comportamiento.
- La generación de informes a la medida por los docentes para hacer análisis de metodologías, currículo y estrategias de aprendizaje.

4. Metodológica

El desarrollo tecnológico será dividido en tres fases:

Primera fase: En esta fase se caracterizaran los procesos del programa TAE mediante una entrevista directa con los funcionarios del programa de trayectoria académica (TAE), con el objetivo de identificar las actividades realizadas en el proceso de análisis de estudiantes con riesgo a desertar. Posteriormente, se analizarán las respuestas de las entrevistas de los funcionarios del programa (TAE), con el fin de definir requerimientos que sirvan de base para selección de herramientas de inteligencia de negocios que apoyen el proceso de análisis. Así mismo, se realizará el estudio de la metodología de Ralph Kimball para la construcción del proyecto.

Seguidamente se analizaran las bases de datos de la institución con el fin de definir un prototipo general que será el insumo básico para la investigación, con el propósito de definir la calidad de los datos.

Segunda fase: En base a la caracterización se definen requerimientos del sistema y con estos se inicia el proceso de seleccionar las herramientas de inteligencia negocio, necesarias para dar solución a la problemática planteada. A continuación se instalan, configuran y despliegan. De igual forma, se procederá a realizar los diseños apropiados y definidos por la metodología seleccionada. En este orden de ideas, se configurara la base de datos de la institución para optimizar las consultas, dichos repositorios están en SQL server. Es importante anotar que el diseño gráfico se realizara en base al definido por las aplicaciones de CECAR.

Tercera fase: Por último, se analizara la información obtenida de las herramientas de inteligencia de negocio para definir patrones descriptivos y predictivos del estudiante desertor de CECAR. De igual modo, se generarán los diferentes reportes, graficas, y/o consultas permitan los funcionarios del programa trayectoria académica tomar decisiones proactivas basadas en información.

4.1. Fase I: Caracterizar los procesos del programa TAE para el análisis de la deserción en la corporación Universitaria CECAR

Para el desarrollo de esta fase se realizaron entrevistas a estudiantes y funcionarios del programa trayectoria académica exitosa (TAE) con el fin de comprender el flujo de procesos, analizar fortalezas, debilidades, oportunidades y necesidades que requiere el programa TAE.

4.1.1. Entrevistas.

Para la realización de las entrevistas se diseñaron y aplicaron dos instrumentos, uno para los funcionarios del programa TAE (ver anexo A) y otro para estudiantes que pertenezcan o hayan pertenecido al programa TAE (ver anexo B).

Muestra el personal entrevistado, así como también el tipo de actor y los anexos correspondientes.

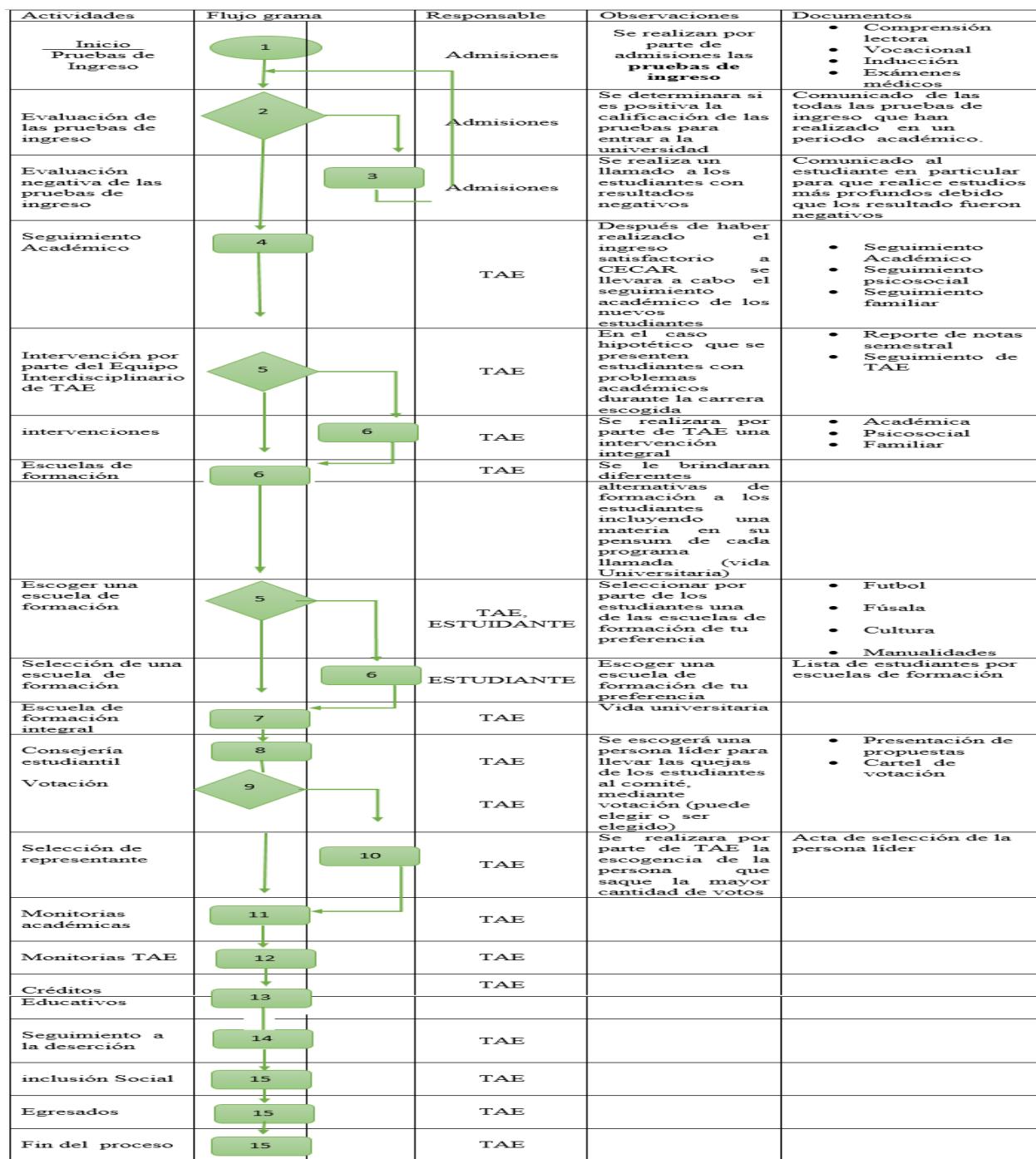
Tabla 4.

Entrevistas Aplicadas a diferentes actores que intervienen en el programa TAE

Personal Entrevistado	Cargo/Función	Instrumento Aplicado	Objetivo	Entrevista
Yaneth Polo Bolaño	Directora del Programa TAE	Instrumento # 1	Comprender el flujo de procesos, analizar fortalezas, debilidades, oportunidades y necesidades que requiere el programa TAE.	
Ángela Erazo Arias	Coordinadora de seguimiento del programa TAE a los programas de ingenierías de la Corporación Universitaria del Caribe CECAR	Instrumento # 1	Comprender el flujo de procesos, analizar fortalezas, debilidades, oportunidades y necesidades que requiere el programa TAE.	
Estudiantes que pertenecen	Estudiante del programa de	Instrumento # 2	Comprender el flujo de procesos, analizar	

o pertenecieron al programa TAE	ingeniería de sistemas.		fortalezas, debilidades, oportunidades y necesidades que requiere el programa TAE.	
---------------------------------	-------------------------	--	--	--

4.1.2. Flujo de Procesos del programa TAE.



4.1.3. Matriz DOFA del programa TAE.



Figura 7. Matriz DOFA TAE

4.2. Fase II: Implementar herramientas de Inteligencia de negocio que apoyen el proceso de análisis de deserción del programa de trayectoria académica TAE

Atendiendo el proceso de desarrollo basado en la metodología de Ralph Kimball, se inicia la fase de implementación del proyecto con un diagnóstico de los sistemas de información transaccionales de CECAR que serán tomados como fuentes de datos para la implementación de las herramientas de inteligencia de negocio.

4.2.1. Diagnóstico de los Sistemas de Información transaccionales de CECAR.

Actualmente CECAR cuenta con un sistema de información transaccional para apoyar los procesos académicos - administrativos de la institución. El sistema se encuentra compuesto por diferentes módulos orientados al área administrativa o académica según el caso. A continuación se mencionan los módulos que pertenecen a cada área:

- **Área Administrativa**
 - Módulo de contabilidad, que a su vez se encuentra integrado por los submódulos de:
 - Contabilidad.
 - Activo Fijos.
 - Módulo de financiera, que a su vez se encuentra integrado por los submódulos de:
 - Tesorería.
 - Presupuesto.
 - Pagaduría.
 - Módulo de recursos humanos, que a su vez se encuentra integrado por los submódulos de:
 - Nomina.
 - Catedra.
 - Módulo de mercadeo y publicidad.
- **Área académica**
 - Academia
 - Catedra

- Snies
- Evaluación docente
- Planta física.

Para efectos del trabajo de desarrollo, se profundiza en el módulo académico con el objetivo de identificar qué información puede ser relevante y necesaria para el desarrollo del prototipo de inteligencia de negocios. Es importante anotar que esta área se encuentra cimentada sobre varias bases de datos implementadas sobre el motor de bases de datos SQLServer, cada una almacena información de los módulos de academia, entrevistas, planta física, catedra y evaluación docente en las bases de datos de academia, Snies, planta, catedra y evaluación respectivamente. Las principales funciones del módulo académico son:

- Gestión estudiantil.
- Gestión docente.
- Gestión de cursos (creación, asignación de docente y estudiantes, horarios y ubicación en el campus universitario).
- Gestión de actas.
- Generación de certificados académicos.
- Evaluación docente.
- Gestión de hojas de vidas de funcionarios y docentes.
- Gestión de matrícula estudiantil.
- Gestión del pensum académico.
- Graduación de estudiantes.
- Gestión de programas académicos.
- Inscripción de estudiantes potenciales.
- Entrevista a estudiantes potenciales.

Por otra parte, el sistema de información se encuentra cimentado sobre una arquitectura cliente – servidor mult capas, distribuida en diferentes servidores de torres (*En inglés: towers*

server), los cuales permiten un escalamiento de tipo horizontal (ver figura 7). Los servicios se ofrecen a clientes de escritorio en la red y en entornos web sobre la Intranet e Internet La accesibilidad de estos últimos servicios se puede realizar desde microcomputadores, notebook, netbook y/o dispositivos móviles. Los componentes de software se encuentran desarrollados sobre software propietario, más específicamente sobre soluciones Microsoft (Ver tabla 5). A nivel de infraestructura de comunicaciones la Corporación cuenta con una red que utiliza como protocolo de capa 2 Ethernet y la arquitectura de protocolos TCP/IP que trabajan en la capa 3 y 4 respectivamente. La red se encuentra compuesta por varios switches de nivel 2 y access point de diferentes proveedores (Alcatel, Cisco, DLink) para las conexiones del borde (*En inglés: edge*), un switch de nivel 3 que actúa como núcleo (*En inglés: Core or backbone*) y permite configurar las diferentes VLAN con las que cuenta la institución. Un router Mikrotic que además de enrutar permite balancear el ancho de banda y actuar como firewall y por ultimo un canal dedicado de conexión a Internet con un ancho de banda de 16MB (Ver inventario de la infraestructura tecnológica en la tabla 5).

Tabla 5.

Inventario de la infraestructura tecnológica en software y hardware.

Servicio	Cantidad Servidores	Software	Proveedor
Web	1	IIS (Internet Information Server)	Microsoft
Bases de datos	1	SQL Server	Microsoft
Firewall y Proxy	1	ForeFront	Microsoft
DHCP, DNS, Archivos, Active Directory ¹	1	Proporcionados por el Windows Server	Microsoft
Aplicaciones	1	Windows Server 2008	Microsoft

Como resultado de la investigación para el diagnóstico de los sistemas de información en CECAR y basado en la investigación titulada “*SISTEMAS DE INFORMACIÓN ORGANIZACIONALES DE CECAR Y SU IMPACTO EN LA TOMA DE DECISIONES*” desarrollada en el año 2013 en CECAR, el sistema de información transaccional implantado en la Corporación solo apoya el nivel operacional de la pirámide organizacional, automatizando las actividades administrativas de los funcionarios, así como también las actividades académicas del

personal docente y estudiantil de la Corporación y brindado apoyo solo en la toma de decisiones estructuradas del nivel operacional.

Por ende, actualmente no existen sistemas de información implantados en CECAR que apoyen la de toma de decisiones en los procesos académicos – administrativos en los niveles tácticos y estratégicos, que permitan, a través de la generación de conocimiento, definir estrategias o políticas específicamente en el área académica que logren mitigar o erradicar el fenómeno de la deserción.

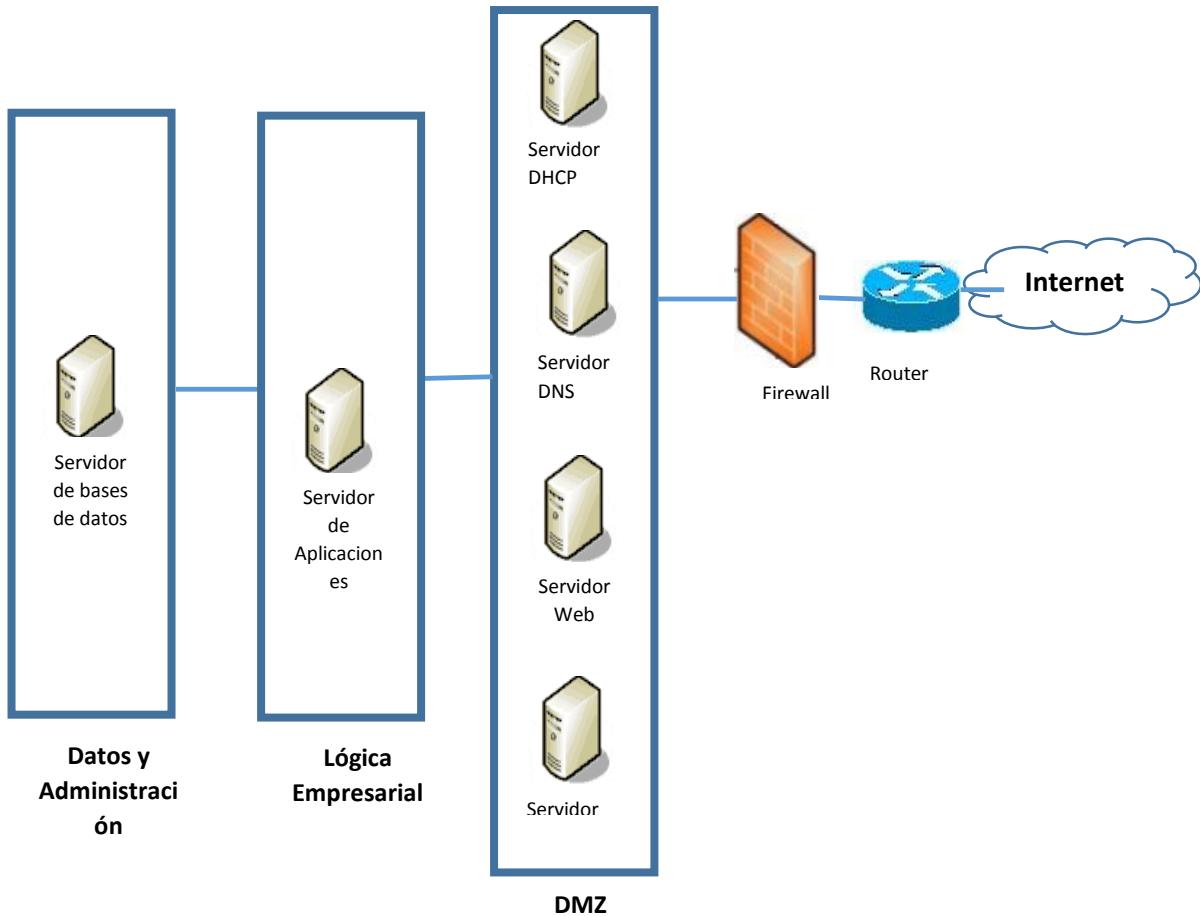


Figura 8. Diseño Arquitectónico y distribución del hardware de CECAR

4.2.2. Definición de requerimientos en base al proceso de análisis de deserción definidos por TAE.

A continuación y en base a la metodología de Kimball se describen los requerimientos del negocio necesarios para analizar el fenómeno de la deserción en CECAR tomando como referencia la caracterización realizada en la fase I.

Es importante aclarar que la metodología no define artefactos de software para los requerimientos y por ende se seleccionó los formatos de historias de usuarios definidos en la metodología XP

Tabla 6.
Historia de usuario (RESPIS).

Identificador del requerimiento:01	Nombre: Reporte de estudiantes que pierden por semestre (REPIS)	Caso de uso/Evento: 01
Descripción: Coordinador TAE entra en el sistema debidamente registrado, escoge el periodo. El sistema valido si ese periodo está activo o es erróneo; en caso de ser erróneo lanzara una advertencia diciendo que ese periodo no es válido. En caso de que el periodo este bien se generara un reporte por semestre de los estudiantes que pierden material clasificados por programa.		
Justificación del requerimiento: Encontrar un modelo de datos en CECAR para apoyar al Programa TAE dentro de CECAR.		
Fuente (que interesado lo propone): Yaneth Del Socorro Polo.	Unidad en la que se origina: Bienestar Universitario.	
Criterios de validación: Encontrar el foco correcto a donde debe apuntar bienestar universitario.		
Dependencias (que requerimientos depende de este):02	Conflictos(que requerimientos son incompatibles o inconsistentes con este):	
Documentos de soporte:	Histórico de cambios:	

Tabla 7.

Historia de usuario (RECAR).

Identificador del requerimiento:02	Tipo de requerimiento: Funcional RECAR	Caso de uso/Evento:	01
Descripción: Coordinador TAE entra en el sistema debidamente registrada. El usuario se le presentaran dos opciones: (Edad, Programa).			
1. Edad: Se generara un reporte totalizado por edad de cada uno de los aspirantes. 2. Programa: se generara un reporte por programa de cada uno de los aspirantes.			
Justificación del requerimiento: Encontrar un modelo de datos en CECAR para apoyar al Programa TAE dentro de CECAR.			
Fuente (que interesado lo propone): Yaneth Del Socorro Polo.	Unidad en la que se origina: Bienestar Universitario.		
Criterios de validación: Encontrar el foco correcto a donde debe apuntar bienestar universitario.			
Dependencias (que requerimientos depende de este):02	Conflictos (que requerimientos son incompatibles o inconsistentes con este):		
Documentos de soporte:	Histórico de cambios:		
Proyecto: TESIS	Analista(as): Luis Carlos Canchila Tapia Javier Andres Sanchez Bohorquez		

Tabla 8.

Historia de usuario (INFACA).

Identificador del requerimiento:03	Tipo de requerimiento: Funcional INFACA	Caso de uso/Evento:	01
Descripción: Generar un informe por medio de una encuesta a los estudiantes de cuáles son los factores que más los afectan en la universidad, los cuales afectan su rendimiento académico.			
Justificación del requerimiento: Encontrar un modelo de datos en CECAR para apoyar al Programa TAE dentro de CECAR.			
Fuente (que interesado lo propone): Yaneth Del Socorro Polo.	Unidad en la que se origina: Bienestar Universitario.		
Criterios de validación: Encontrar el foco correcto a donde debe apuntar bienestar universitario.			
Dependencias (que requerimientos depende de este):02	Conflictos (que requerimientos son incompatibles o inconsistentes con este):		
Documentos de soporte:	Histórico de cambios:		
Proyecto: TESIS	Analista(as): Luis Carlos Canchila Tapia Javier Andres Sanchez Bohorquez		

Tabla 9.

Historia de usuario (INPRO).

Identificador del requerimiento:04	Tipo de requerimiento: Funcional INPRO	Caso de uso/Evento:	02
Descripción: Generar un Informe general por carrera de los estudiantes con bajo promedio.			
Justificación del requerimiento: Con este informe se busca que las personas encargadas en bienestar logren hacer un mejor seguimiento.			
Fuente (que interesado lo propone): Yaneth Del Socorro Polo.	Unidad en la que se origina: Bienestar Universitario.		
Criterios de validación: Con los datos obtenidos del informe lograr hacer información útil para apoyar			
Dependencias (que requerimientos depende de este):01	Conflictos(que requerimientos son incompatibles o inconsistentes con este):		
Documentos de soporte:			
Proyecto: TESIS	Analista(as):Luis Carlos Canchila Tapia Javier Andres Sanchez Bohorquez		

Tabla 10.

Historia de usuario (INDE).

Identificador del requerimiento:05	Tipo de requerimiento: Funcional INDE	Caso de uso/Evento:	03
Descripción: Generar un Informe de posibles desertores.			
Justificación del requerimiento: apoyar a las personas que están en el programa TAE para actúen antes de que ese estudiante deserte de la carrera.			
Fuente(que interesado lo propone): Luis Carlos Canchila Tapia Javier Andres Sanchez Bohorquez	Unidad en la que se origina:		
Criterios de validación: le permitirá ver los posibles desertores por semestre y por carrera con los bajos promedios, y los que no se matriculen en semestre arroje la alamar para ayudar y hacer que ese estudiante no abandone.			
Dependencias (que requerimientos depende de este):	Conflictos(que requerimientos son incompatibles o inconsistentes con este):		
Documentos de soporte:			
Proyecto: TESIS	Analista(as):Luis Carlos Canchila Tapia Javier Andres Sanchez Bohorquez		

A continuación se mencionan las herramientas seleccionadas para dar solución a cada uno de los requerimientos establecidos en la sección anterior.

Historia de Usuario	Herramienta
1	Minería de datos
2 y 3	Reportes
4	OLAP

4.2.3. Diseño de la solución en base a la Metodología de Kimball.

4.2.3.1. *Diseño Arquitectónico.*

La figura 9 muestra los diferentes componentes que integran la solución para el desarrollo del prototipo de inteligencia de negocios desde el punto de vista de la arquitectura del sistema.

Las fuentes de datos están compuestas por cuatro bases de datos transaccionales que dan soporte al sistema de información transaccional de CECAR. La base de datos **Catedra**, la cual gestiona información del cuerpo docente, Academia encargada de gestionar los datos de estudiantes, **Snies** la cual almacena los datos socioeconómicos, núcleo familiar e historial académico del estudiante y **Planta** que almacena la información relacionada a horarios y carga académica de los docentes.

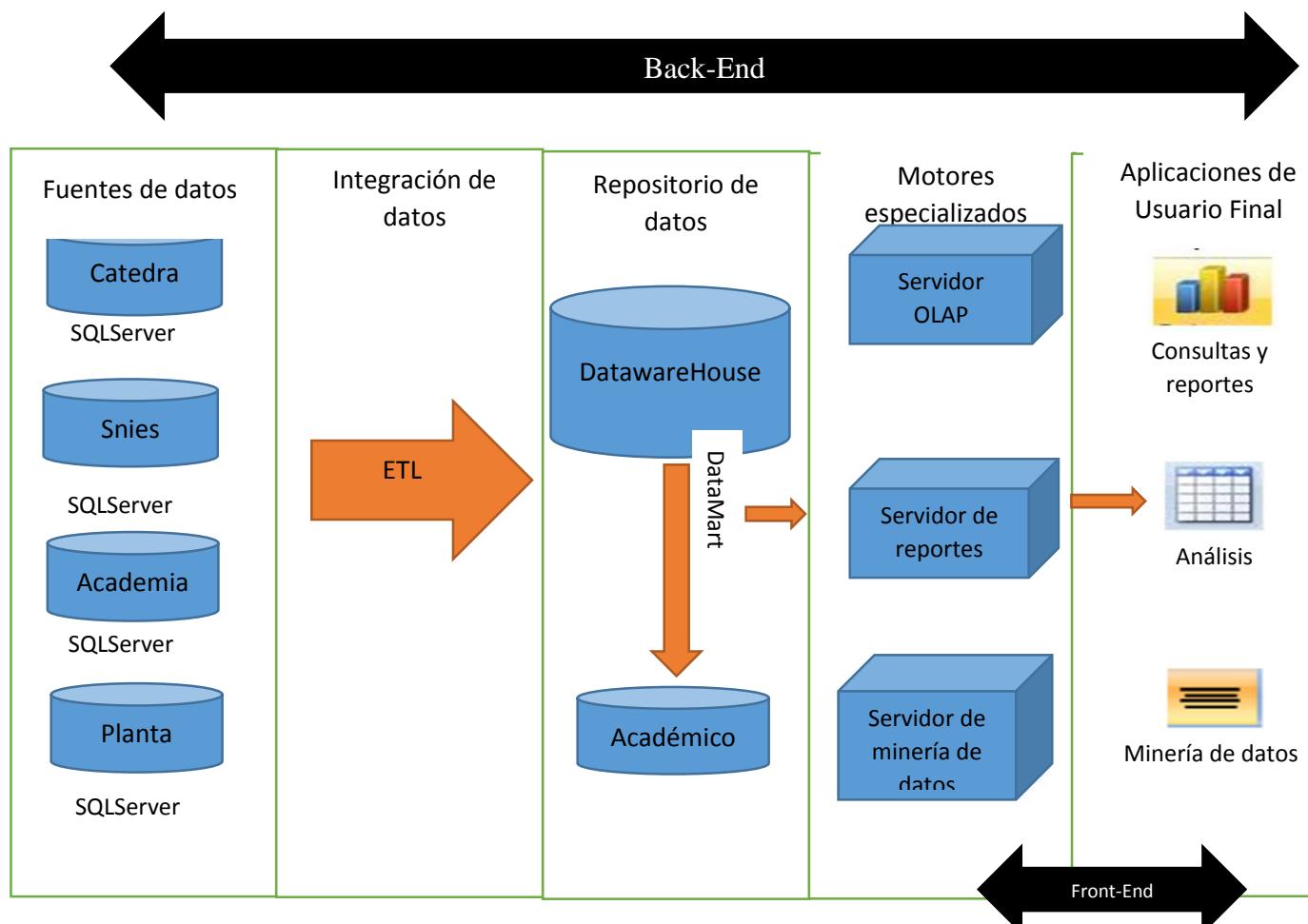


Figura 9. Diseño Arquitectónico
Fuente. Elaboración propia

La integración de datos está conformada por la herramienta ETL encargada de implementar las funcionalidades de Extracción, Transformación y Carga de Datos desde las fuentes de datos al Datawarehouse.

El repositorio de datos se encuentra compuesto por el Datawarehouse sobre el cual se creará un Datamart (base de datos en postgresql con el modelo decisional) denominado académico orientado a dar solución a los requerimientos del programa TAE liderado por el departamento de bienestar académico de la Corporación universitaria del caribe.

Los motores especializados para OLAP, minería de datos y reportes son ofrecidos por la suite de inteligencia de negocios escogida.

Finalmente, el Front-End del sistema contiene los componentes de usuario final, presentando interfaces gráficas para acceder a los servicios de minería de datos, OLAP y reportes.

4.2.3.2. *Modelo Dimensional.*

Bus Matrix DataWareHouse

En la tabla 11 se presenta el Bus Matrix en donde se definen las dimensiones que se comparten entre los diferentes datamart que integran la bodega de datos, la granularidad de los datos que se van a mostrar, las tablas de hechos y las medidas del modelo dimensional.

Tabla 11.
Bus Matrix Data WareHouse

Dimensiones														
Proceso de negocio	Tablas de hecho	Medidas	Granularidad		Dim_Docente	Dim_Tiempo	Dim_Asignatura	Dim_Estudiante	Dim_EstadoCivil	Dim_Programa	Dim_Estrato	Dim_LugarResidencia	Dim_PromedioEstudiantil	Dim_Sexo
TAE	Notas	Nota Primer Corte	Detalle de las notas por debajo de 3.0 en algún	*	*	*	*	*	*					

4.2.3.3. *Modelo Lógico.*

El modelo lógico se realizó siguiendo un esquema en estrella que optimiza el tiempo de respuesta para consultas complejas, siguiendo la metodología de Ralph Kimball se creó un Datamart para el área de bienestar académico denominado académico. De igual forma se definieron (ver 5.2.4.4) las dimensiones *Dim_Docente*, *Dim_Tiempo*, *Dim_Asignatura*, *Dim_Estudiante*, *Dim_EstadoCivil*, *Dim_Programa*, *Dim_Estrato*, *Dim_LugarResidencia*, *Dim_PromedioEstudiantil* y *Dim_Sexo*, y las tablas de hechos redimiendo y deserciones, así también las medidas *nota primer Corte*, *nota Segundo Corte*, *nota Tercer Corte*, *definitiva*, *asignaturas aprobadas*, *asignaturas cursada*, *matriculado* y *potencial desertado*.

A continuación se describe el diseño conceptual y lógico de la solución del prototipo de inteligencia de negocios.

4.2.3.4. *Diseño Conceptual.*

A continuación se describen los diseños conceptuales para la generación de reportes dinámicos para estudiantes por debajo de 3.0 y cuantificación de la deserción estudiantil en el datamart académico (ver figura 10 y 11).

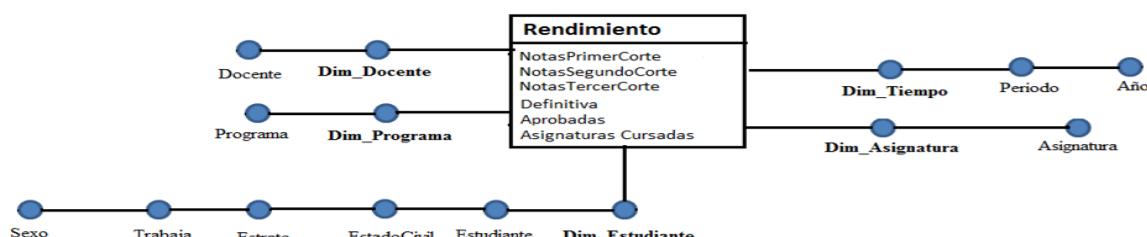


Figura 10. Diseño conceptual del modelo dimensional para estudiantes por debajo de 3.0 en el datamart académico

Fuente. Elaboración propia



Figura 11. Diseño conceptual del modelo dimensional para la cuantificación de la deserción estudiantil en el datamart académico

Fuente. Elaboración propia

4.2.3.5. *Diseño Lógico.*

A continuación se describen los diseños lógicos para la generación de reportes dinámicos para estudiantes por debajo de 3.0 y cuantificación de la deserción estudiantil en el datamart académico (ver figura 12 y 13).

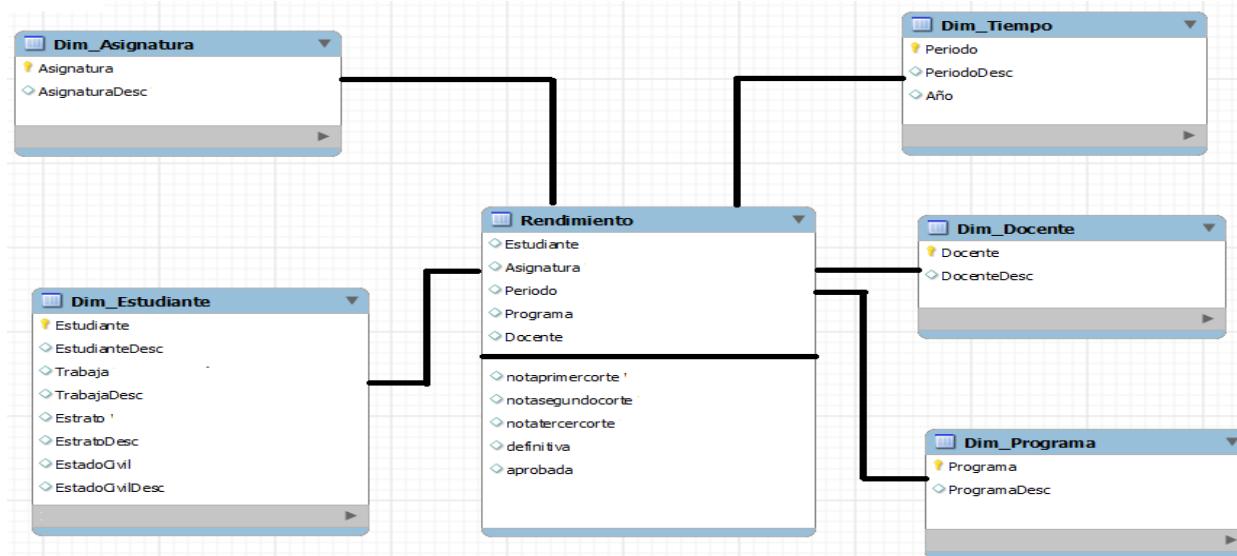


Figura 12. Diseño lógico del modelo dimensional para estudiantes por debajo de 3.0 en el datamart académico

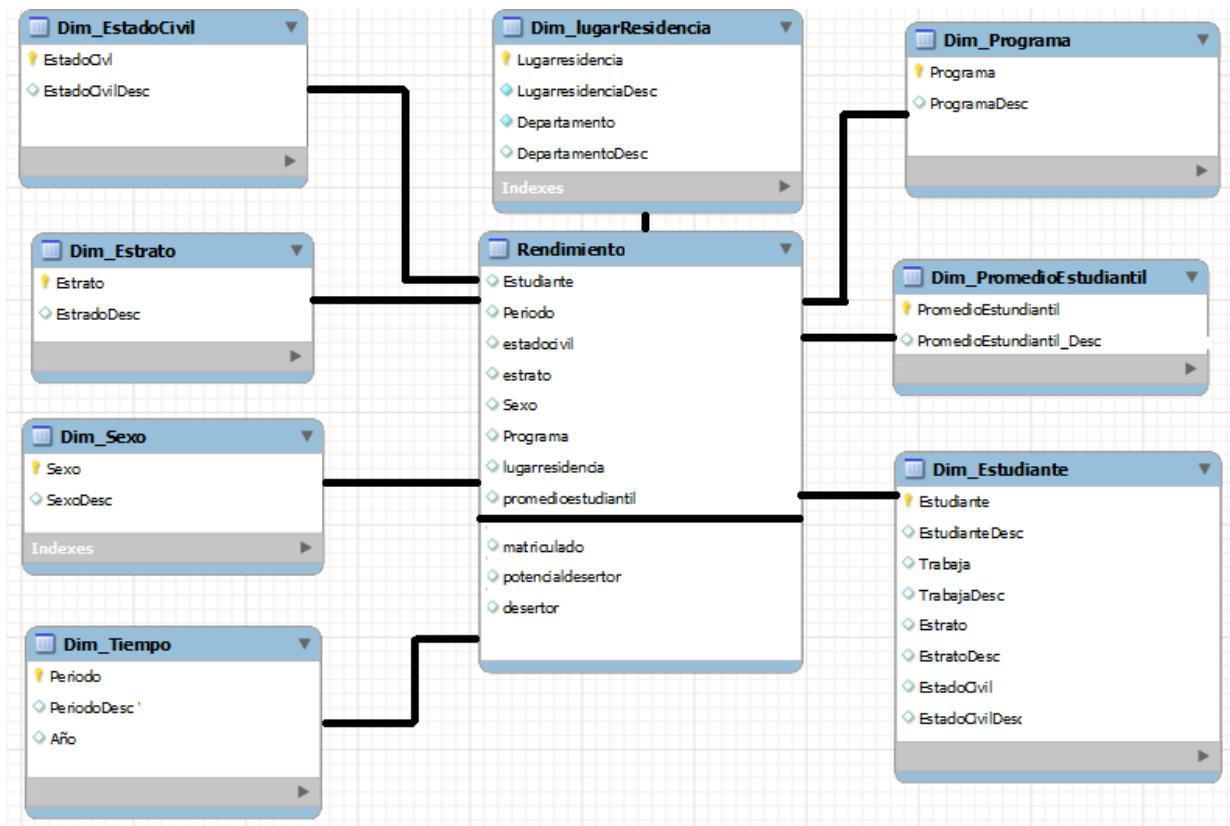


Figura 13. Diseño conceptual del modelo dimensional para la cuantificación de la deserción estudiantil en el datamart académico.

4.3. Fase III: Implementación de Herramientas de Inteligencia de Negocios Seleccionadas

4.3.1. Instalación y configuración del repositorio de datos.

Para el repositorio de datos se escogió como motor de base de datos PostgreSQL en donde se crearan los diferentes datamart que conformaran la bodega de datos. Se escogió este motor debido a las características de capacidad, seguridad, concurrencia, técnicas de optimización para consulta de datos y licenciamiento libre que ofrece.

4.3.2. Instalación y Configuración ETL.

En este proceso se realiza la extracción, limpieza y carga de datos desde las bases de datos Catedra, Snies, Planta y Academia, creadas sobre el motor de base de datos SQLServer, hacia la base de datos académico, creada en PostgreSQL (ver punto anterior).

A continuación se describen los Job configurados para extraer los datos del modelo transaccional.

4.3.3. Trabajo Dimensión Programas.

Trabajo que permite migrar los datos de todos los programas académicos almacenados en la tabla AcaProgramas de la base de datos Academia hacia la tabla Dim_Programa (ver figura 14).

4.3.4. Trabajo Dimensión Lugar de Residencia.

Trabajo que permite migrar los datos de los diferentes departamentos y municipios que conforman el territorio nacional. Para facilitar la migración de estos datos se crea la vista AcaDptoMun_v (ver tabla 12) en la base de datos Academia y se cargan los datos en la tabla Dim_LugarResidencia (ver figura 16). Es importante anotar que se adaptaron los datos para que no se repitiera información, teniendo en cuenta la estructura de la tabla destino (ver figura 14).

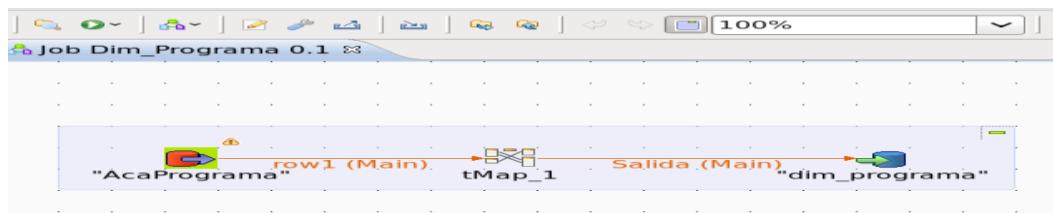


Figura 14. Trabajo dimensión programas

Fuente. Elaboración propia

Tabla 12.

Sentencia SQL para la vista AcaDptoMun_v

Vista	Tablas	Sentencia SQL
AcaDptoMun_v	AcaDptoMun	<pre>SELECT Id_Dpto, (Select Nombre From [Academia].[dbo].[AcaDptoMun] b where a.Id_Dpto = b.Id_Dpto and Id_Municipio = 0) nombreDpto,Id_Municipio,Nombre FROM AcaDptoMun a Where Id_Municipio != 0</pre>

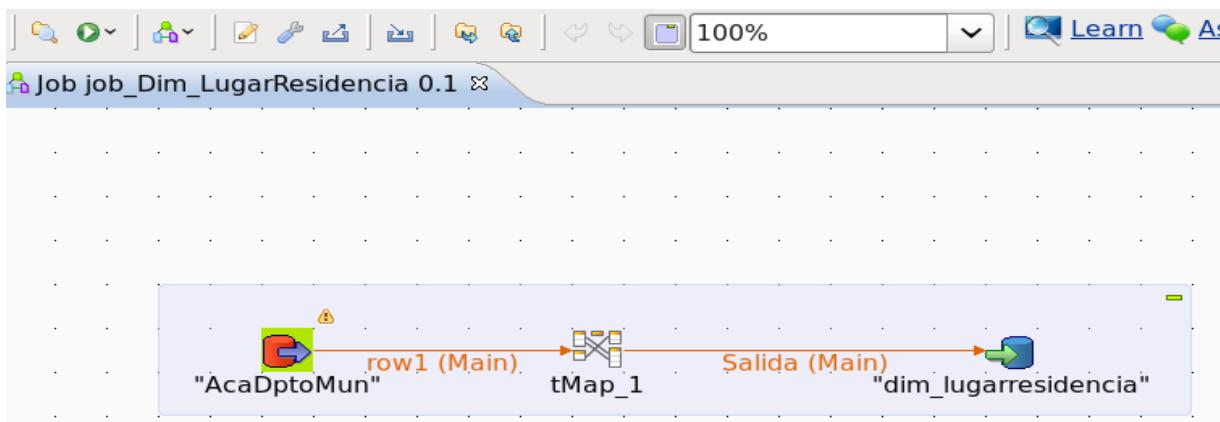


Figura 15. Trabajo dimensión lugar residencia

Fuente. Elaboración propia

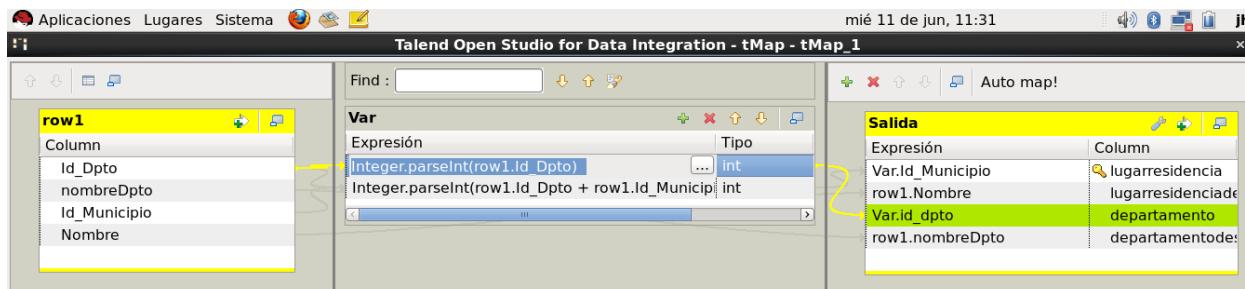


Figura 16. Adaptación de datos para el trabajo dimensión lugar residencia
 Fuente. Elaboración propia

4.3.5. Trabajo Dimensión Asignatura.

Trabajo que permite migrar los datos de las diferentes asignaturas de los programas de pregrado presencial que actualmente ofrece la Corporación universitaria del Caribe (ver figura 17). Estos datos residen en la tabla AcaBloquesReg de la base de datos Academia. Teniendo en cuenta la desnormalización que posee la tabla en mención, se generó una vista denominada AcaBloquesReg_V (ver tabla 13), en donde se elimina la redundancia de datos y se filtran algunas asignaturas que son obsoletas en el pensum de cada programa de CECAR.

Tabla 13.

Sentencia SQL para la vista AcaBloquesReg_V

Vista	Tabla	Sentencia SQL
AcaBloquesReg_V	AcaBloquesReg	<pre> ;WITH InformacionAsignaturas AS (SELECT Regla,Descripcion, ROW_NUMBER() OVER(PARTITION BY Regla ORDER BY Regla) asignatura FROM [Academia].[dbo].[AcaBloquesReg] where SUBSTRING(regla,1,3) <> 'PRE') SELECT Regla,Descripcion FROM InformacionAsignaturas WHERE asignatura = 1 ;WITH InformacionAsignaturas AS (SELECT Regla,Descripcion, ROW_NUMBER() OVER(PARTITION BY Regla ORDER BY Regla) asignatura FROM [Academia].[dbo].[AcaBloquesReg] where oficina = 1 and jornada = 1 and SUBSTRING(regla,1,3) <> 'PRE') SELECT Regla,Descripcion FROM InformacionAsignaturas WHERE asignatura = 1 </pre>

Fuente. Elaboración propia

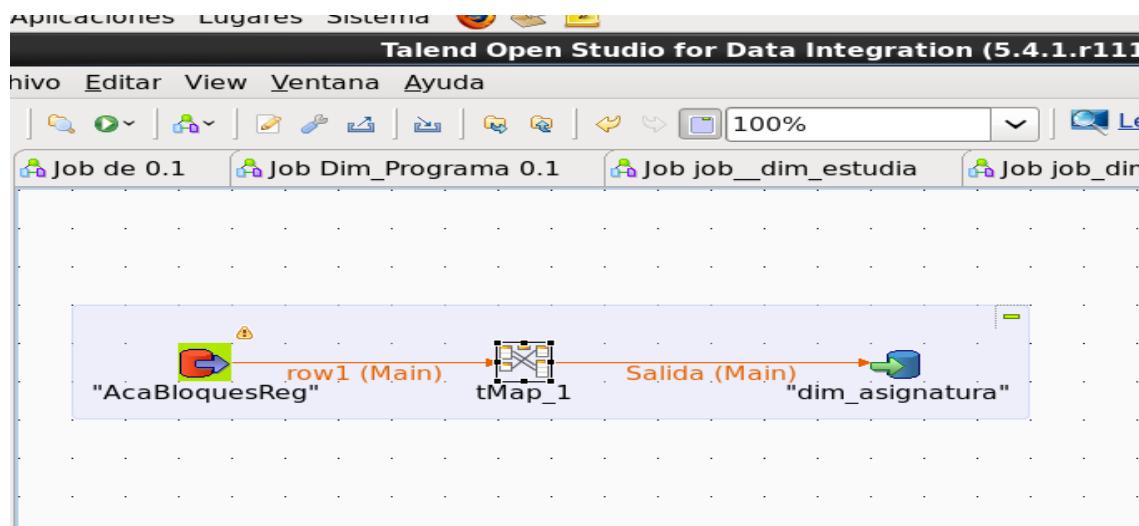


Figura 17. Trabajo dimensión asignatura

Fuente: Elaboración propia

4.3.6. Trabajo Dimensión Estudiantes.

Trabajo encargado de extraer todos los datos personales de los diferentes estudiantes de cada programa presencial que ofrece la Corporación Universitaria del Caribe, para lograr este objetivo se migraron datos de las tablas AcaAlumnos, AcaMatriculas, AcaProgramas, AcaSexos y AcaEstado_Civil contenidas en la base de datos academia. De igual forma se migraron los datos de las entrevistas realizadas a cada estudiante al momento de la inscripción, estos datos son almacenados en las tablas AcaSniesAspirante, SnEstrato, SnEstadoCivil contenidos en la base de datos Snies (ver figura 18 y 19). Además se crean las vistas AcaMatriculas_V, AcaSniesAspirante_V con el objetivo de eliminar datos redundantes en las tablas fuentes de cada tabla (ver tabla 14).

Tabla 14.

Sentencia SQL para las vista AcaMatriculas_V y AcaSniesAspirante_V

Vista	Tabla	Sentencia SQL
AcaMatriculas_V	AcaMatriculas	<pre>SELECT distinct AcaMatriculas.Id_Estudiante, AcaMatriculas.PROG FROM AcaMatriculas AcaMatriculas" AcaMatriculas.Id_Estudiante, AcaMatriculas"</pre>
AcaSniesAspirante_V	AcaSniesAspirante	<pre>;WITH Inscrito AS (SELECT *, ROW_NUMBER() OVER(PARTITION BY [Docidentidad] ORDER BY [Docidentidad]) identidad FROM [Snies].[dbo].[AcaSniesAspirante]) SELECT Docidentidad,estadocivil,trabaja,estratovivienda FROM Inscrito WHERE identidad = 1</pre>

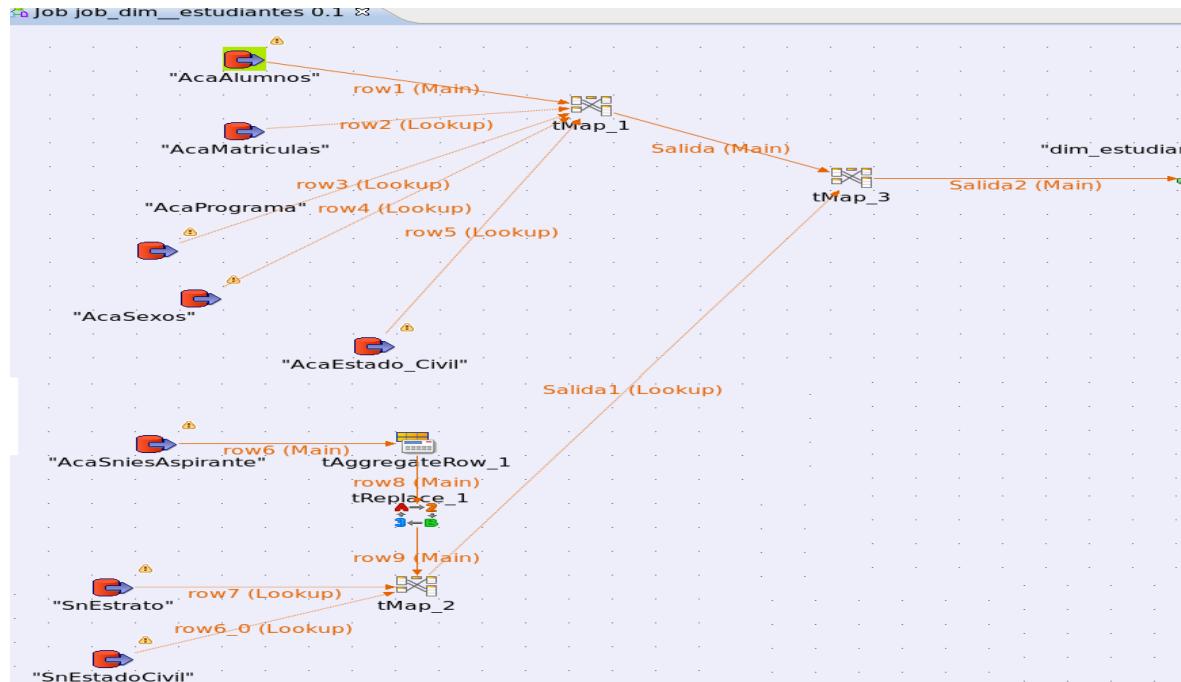


Figura 18. Trabajo dimensión estudiantes

Fuente. Elaboración propia

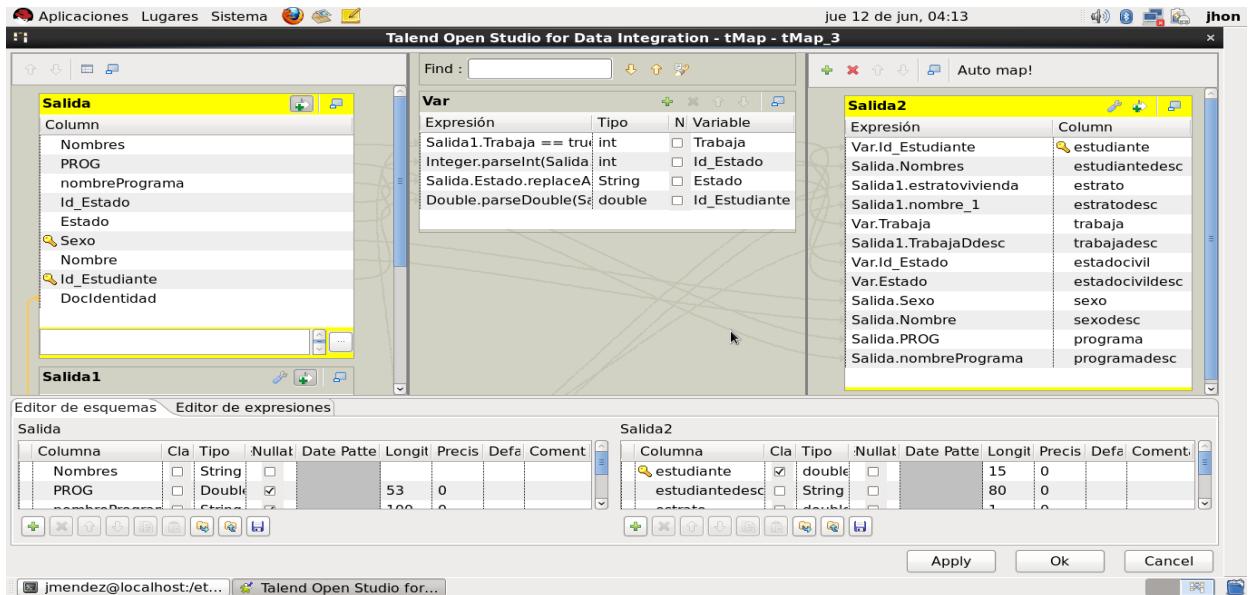


Figura 19. Unión de las diferentes tablas a través del objeto tMap

Fuente. Elaboración propia

4.3.7. Trabajo Dimensión Tiempo.

Trabajo que permite migrar los datos de todos los períodos académicos registrados en el sistema de información, para esto se tomó la tabla AcamaticulasSem contenida en la base de datos Academia (ver figura 17 y 18). Para eliminar la redundancia de datos y filtrar algunos datos erróneos se creó la vista denominada AcamaticulasSem_V (ver tabla 15).

Tabla 15.

Sentencia SQL para las vista AcamaticulasSem_V

Vista	Tablas	Sentencia SQL
AcamaticulasSem _v	AcamaticulasSem	<pre>SELECT distinct AcaMatriculaSem.Periodo FROM AcaMatriculaSem Where substring(periodo,5,5) in ('1','2')</pre>

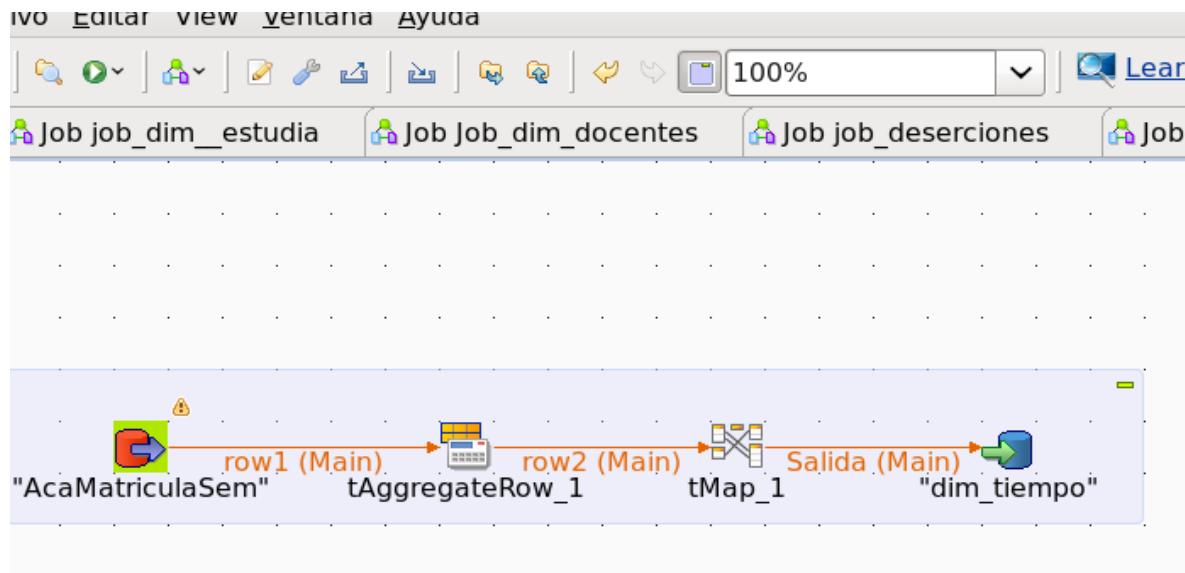


Figura 20. Trabajo dimensión tiempos

Fuente. Elaboración propia

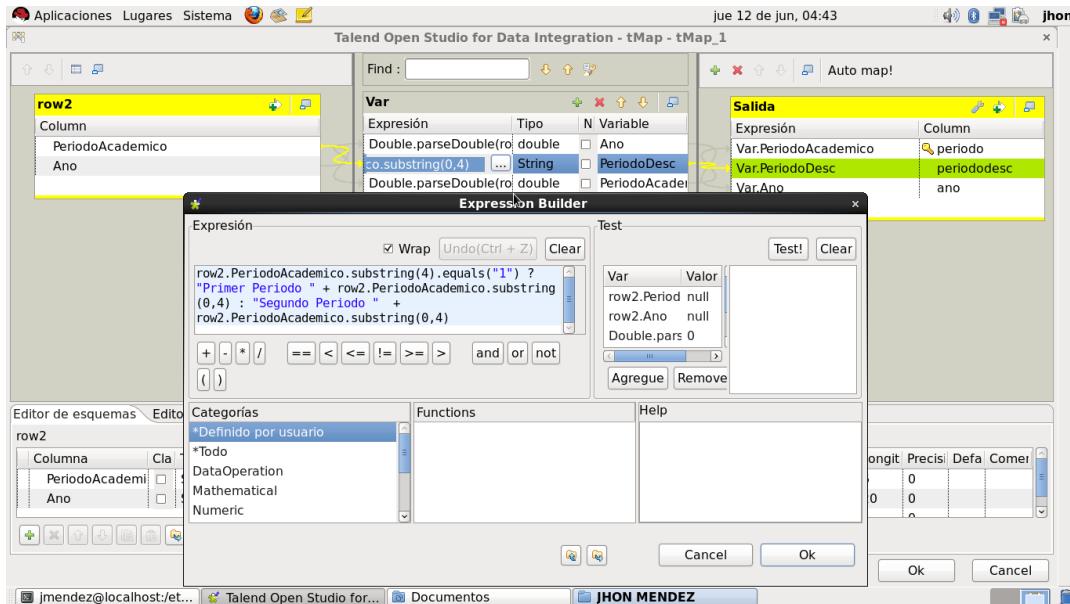


Figura 21. Expresión creada para la columna PeriodoDesc

Fuente. Elaboración propia

4.3.8. Trabajo Dimensión Docente.

Trabajo encargado de extraer los datos de los docentes que imparten asignaturas en la corporación universitaria del Caribe hacia la tabla Dim_docente. Se tomó como base la tabla NomDocente contenida en la base de datos Academia (ver figura 22).

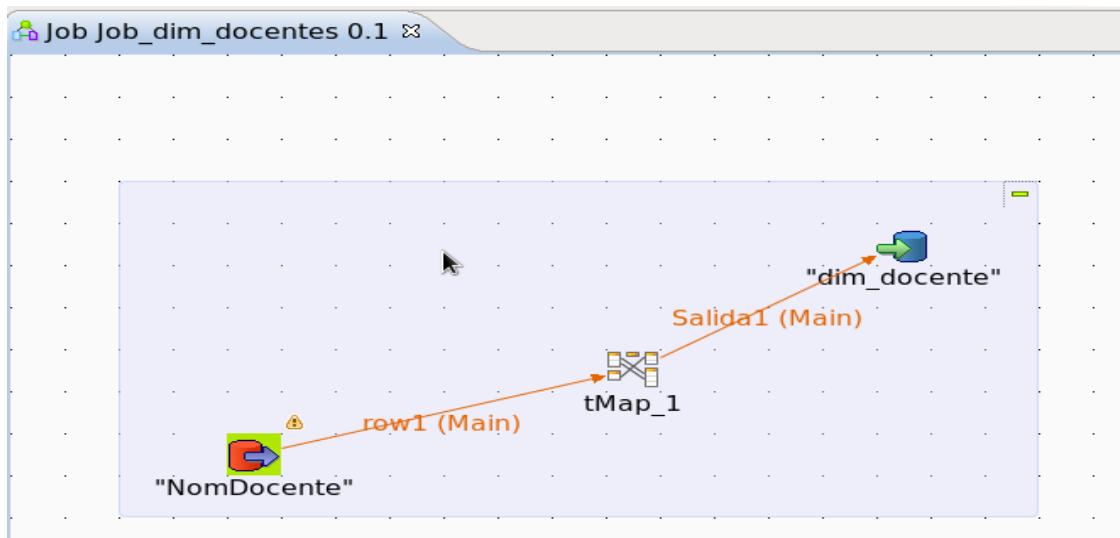


Figura 22. Trabajo dimensión docente

4.3.9. Trabajo Tabla de Hechos Rendimiento.

Trabajo encargado de extraer, filtrar y cargar los datos de la tabla de hechos rendimiento, para esto se tomó como base las tablas AcaNotasSem de la base de datos academia y Asignaturas de la base de datos planta (ver figura 23). Además se creó una regla que catalogaba el promedio estudiantil en bajo, medio y alto a partir del puntaje obtenido de la siguiente forma: si el promedio es menor a 3.0 se considera bajo, si es igual o mayor a 3 y menor a 4 se considera medio e igual y superior a 4 altos.

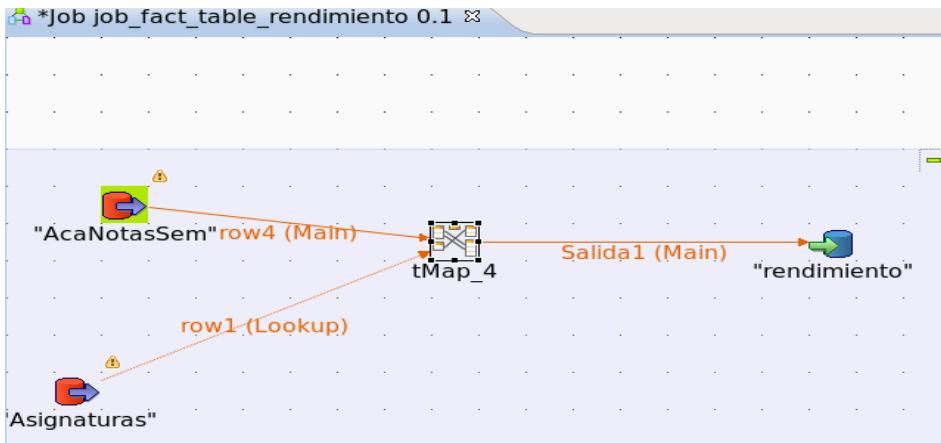


Figura 23. Trabajo tabla de hecho rendimiento

4.3.10. Trabajo tabla de hechos deserciones.

Trabajo encargado de extraer, filtrar y cargar los datos de la tabla de hechos deserciones, para esto se tomó como base las tablas AcaAlumnos, AcaMatriculaSem y PromedioAcumulado de la base de datos academia (ver figura 24). Además se creó la vista AcaAlumnos_V (ver tabla 16) y la regla que catalogaba el promedio estudiantil en bajo, medio y alto a partir del puntaje obtenido de la siguiente forma: si el promedio es menor 3.0 se considera bajo, si es igual o mayor a 3 y menor a 4 se considera medio e igual y superior a 4 altos.

Tabla 16

Sentencia SQL para las vista AcamatriculasSem_V

Vista	Tablas	Sentencia SQL
AcaAlumnos_V	AcaAlumnos	<pre> ;WITH InformacionEstudiantes AS (SELECT a.id_estudiante,a.Sexo, a.DptoRes,a.CiudadRes, a.Estcivil,b.estratovivienda, ROW_NUMBER() OVER(PARTITION BY [id_estudiante] ORDER BY [id_estudiante]) estudiantes FROM AcaAlumnos a left join AcaSniesAspirante b on a.DocIdentidad = b.Docidentidad </pre>

		<pre> Where a.Id_Estudiante not in (Select id_estudiante from AcaGraduadosInformacion)) SELECT id_estudiante,sexo,replace(Isnull(DptoRes,0),- 1,70001),replace(Isnull(CiudadRes,0),- 1,70001),ISNULL(Estcivil,0),ISNULL(estratovivienda ,0) FROM InformacionEstudiantes WHERE estudiantes = 1 and id_estudiante not in (Select id_estudiante from AcaActaGrado) </pre>
--	--	--

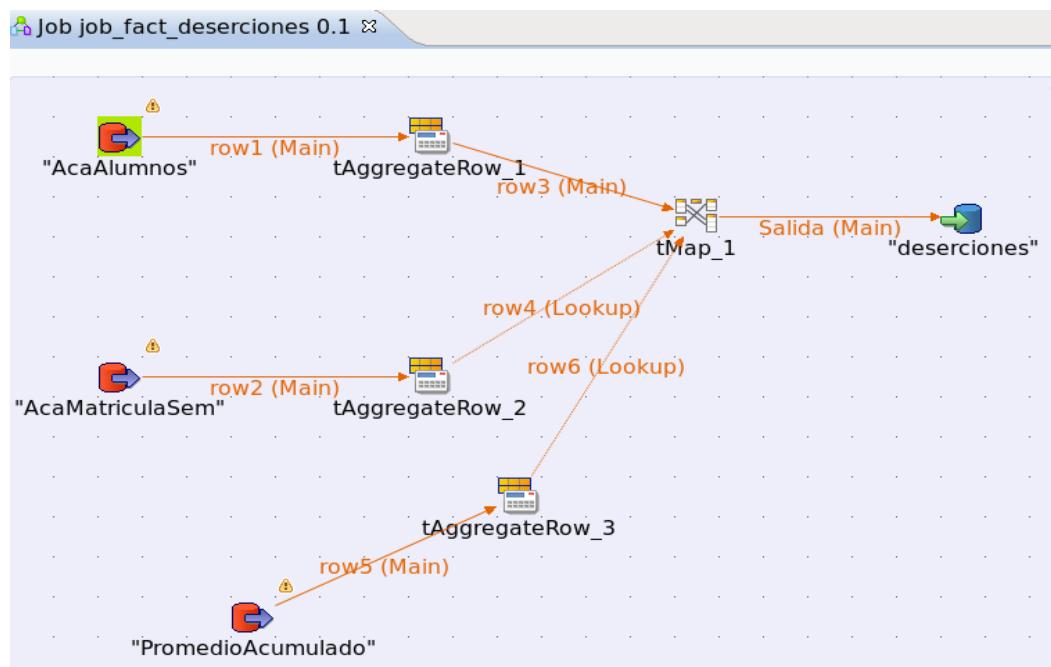


Figura 24. Trabajo tabla de hecho deserciones

Fuente. Elaboración propia

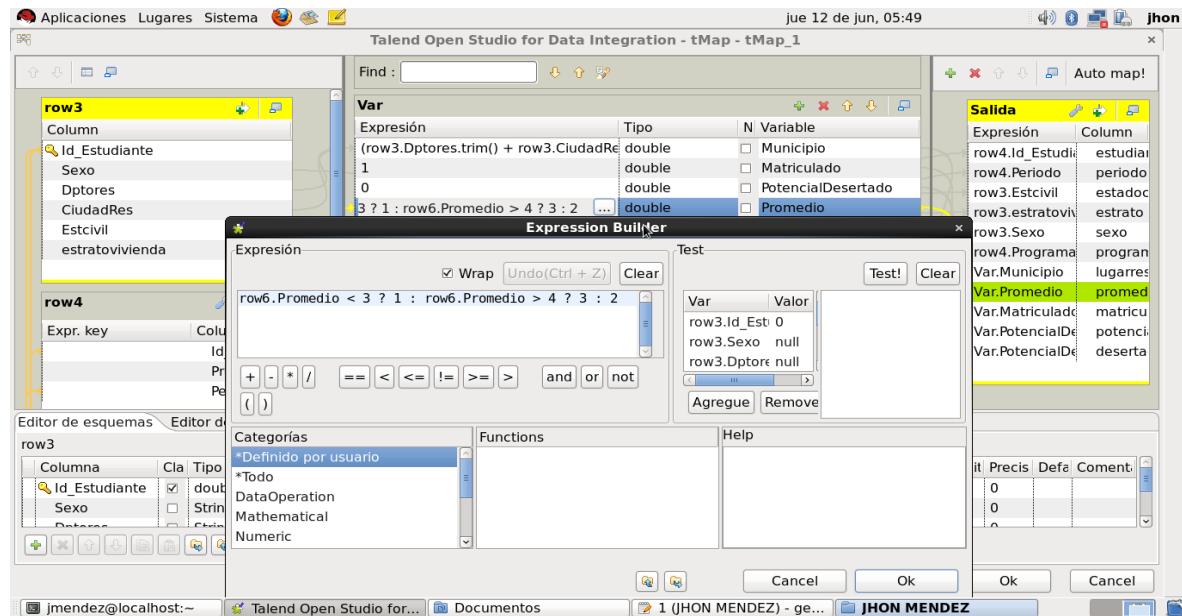


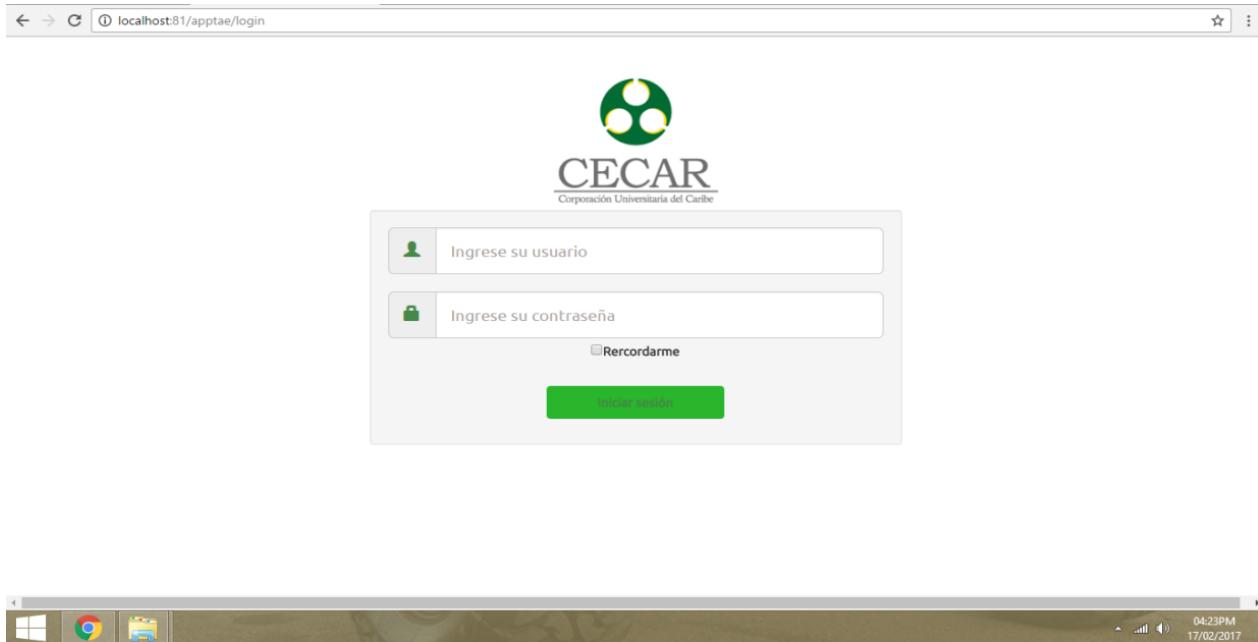
Figura 25. Regla creada para la columna promedio

Fuente. Elaboración propia

4.3.11. Solución Reporting para Analizar la Deserción.

A continuación se muestran los pantallazos para el acceso al módulo de reportes como apoyo al proceso de seguimiento y monitorización de estudiantes con riesgo de abandono académico.

LOGIN: mediante el cual se controla el acceso individual a la plataforma utilizando credenciales provistas por el usuario.



INICIO: donde estarán alojados distintos mensajes de ayuda para el manejo del software, videos y otras noticias importantes.

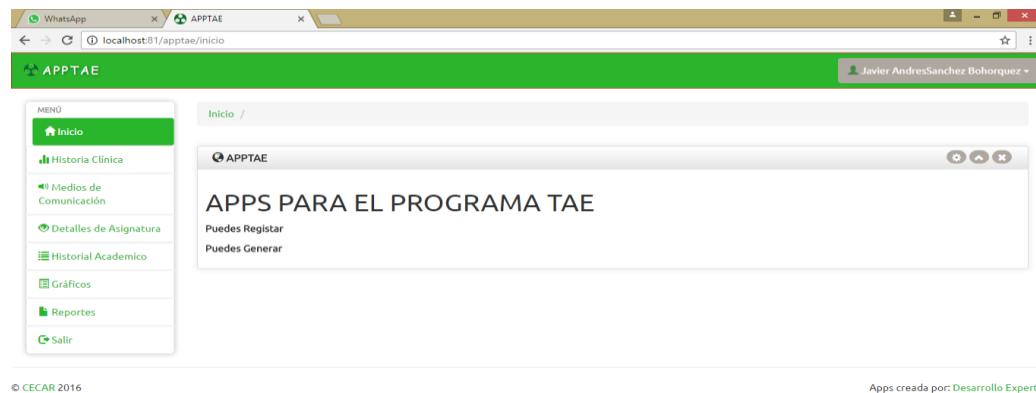


Figura 26. Login: plataforma APPTAE

Es donde se guardaran todas las historias de intervenciones realizadas a estudiantes de dicho programa. Se buscara por los filtros semestre y periodo.

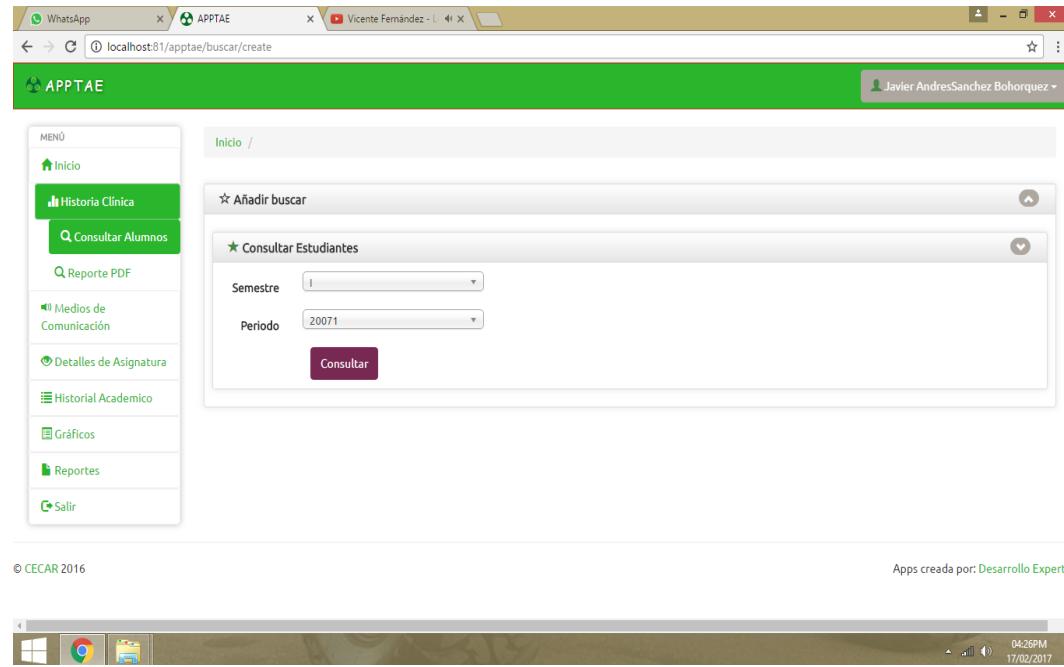
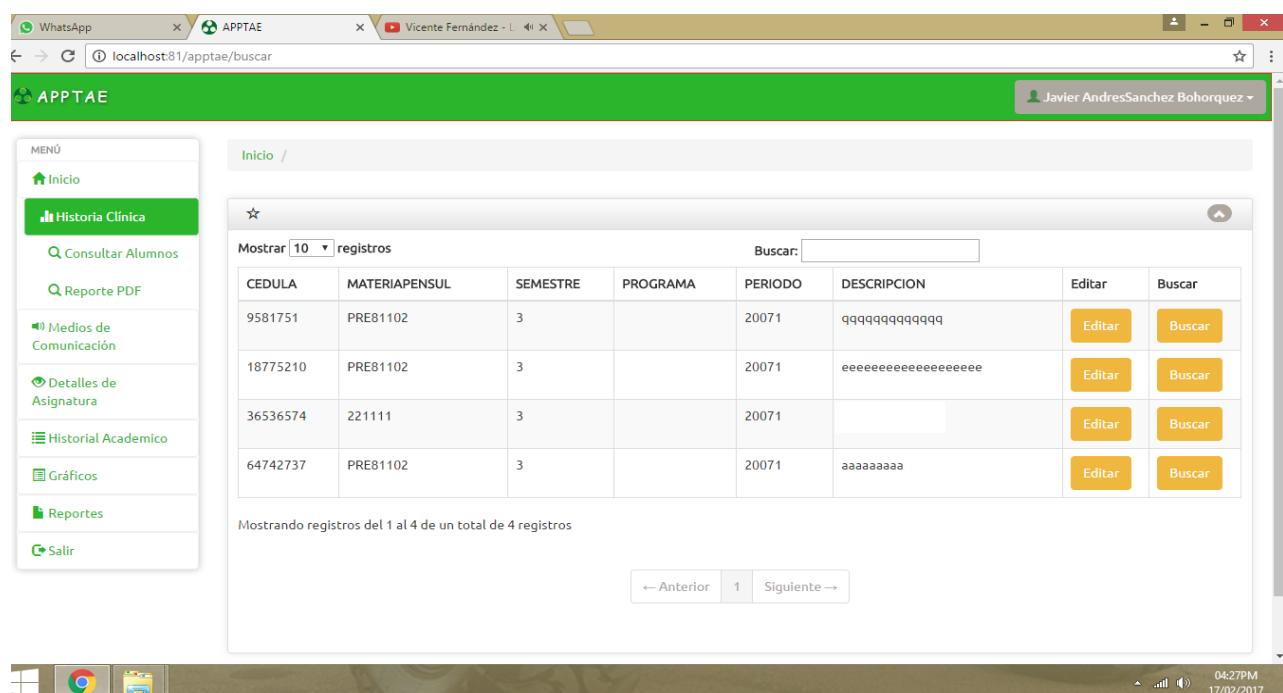


Figura 27. Vista Historia Clínica (Filtros)

Fuente. Elaboración propia

Después de realizar el proceso de filtro debe arrojar la consulta con los criterios de búsqueda.



The screenshot shows a web-based application titled "APPTAE". The left sidebar contains a menu with options like Inicio, Historia Clínica, Consultar Alumnos, Reporte PDF, Medios de Comunicación, Detalles de Asignatura, Historial Académico, Gráficos, Reportes, and Salir. The main content area displays a table of clinical history records. The table has columns for CEDULA, MATERIA/PENSUL, SEMESTRE, PROGRAMA, PERÍODO, DESCRIPCION, Editar, and Buscar. The data in the table is as follows:

CEDULA	MATERIA/PENSUL	SEMESTRE	PROGRAMA	PERÍODO	DESCRIPCION	Editar	Buscar
9581751	PRE81102	3		20071	qqqqqqqqqqqqqq	Editar	Buscar
18775210	PRE81102	3		20071	eeeeeeeeeeeeeeeeee	Editar	Buscar
36536574	221111	3		20071		Editar	Buscar
64742737	PRE81102	3		20071	aaaaaaaaaa	Editar	Buscar

Below the table, it says "Mostrando registros del 1 al 4 de un total de 4 registros". At the bottom, there are navigation buttons for Anterior, 1, and Siguiente.

Figura 29. Vista Reporte (Historia Clínica) (Resultados)



The screenshot shows a web-based application titled "PDFHClinico". The top header includes the university's address, contact information, and legal entity number. The main content area displays a table titled "Reporte Historia Clínica". The table has columns for Cedula, Materia, Semestre, Programa, Período, and Observación. The data in the table is identical to the one in Figure 29:

Cedula	Materia	Semestre	Programa	Período	Observación
36536574	221111	3		20071	holo juete
64742737	PRE81102	3		20071	aaaaaaaaaa
9581751	PRE81102	3		20071	qqqqqqqqqqqqqq
18775210	PRE81102	3		20071	eeeeeeeeeeeeeeeeee

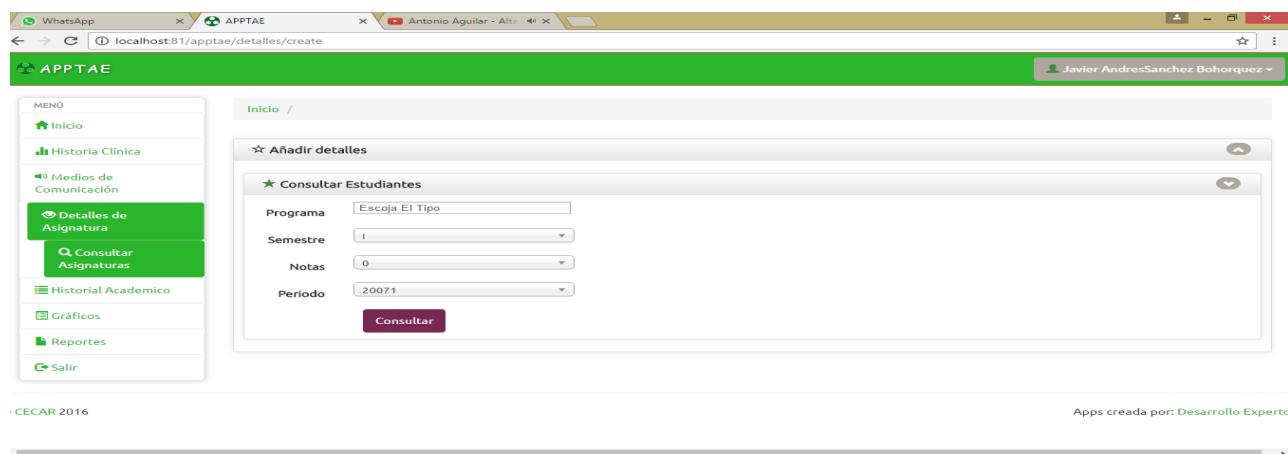
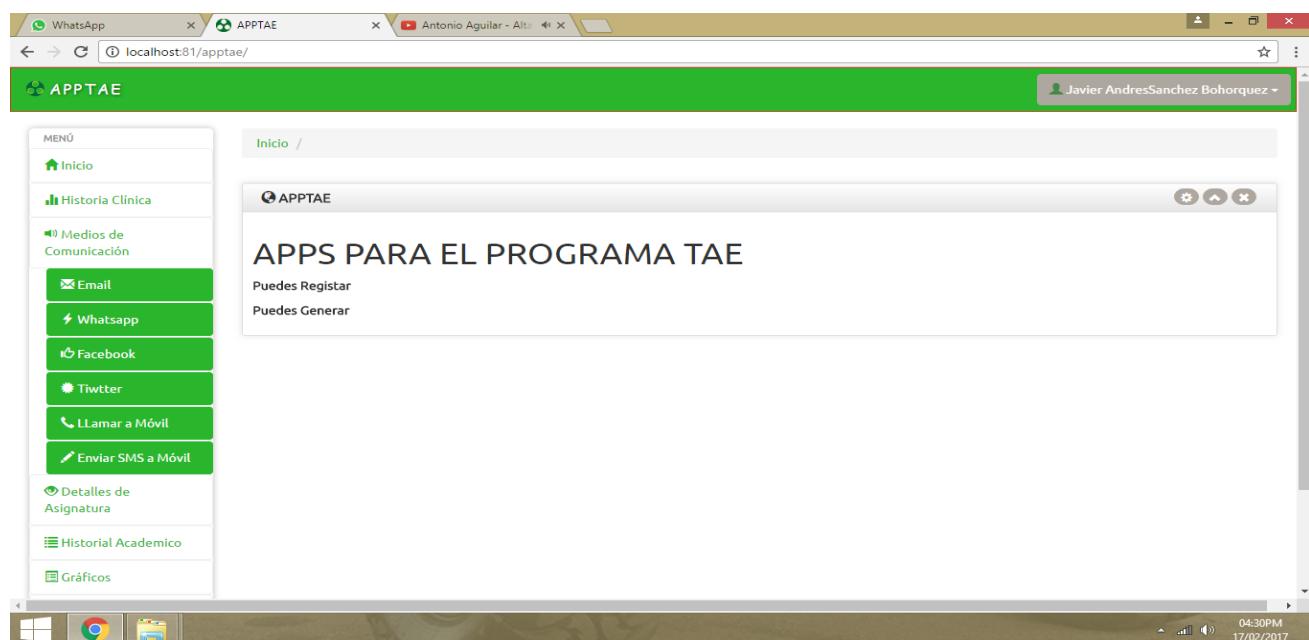
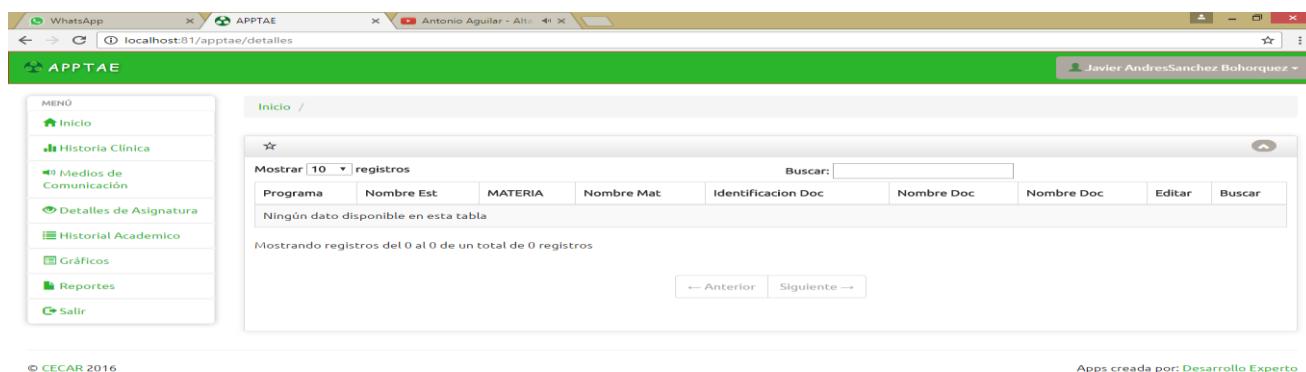


Figura 30. Vista detalle de asignatura

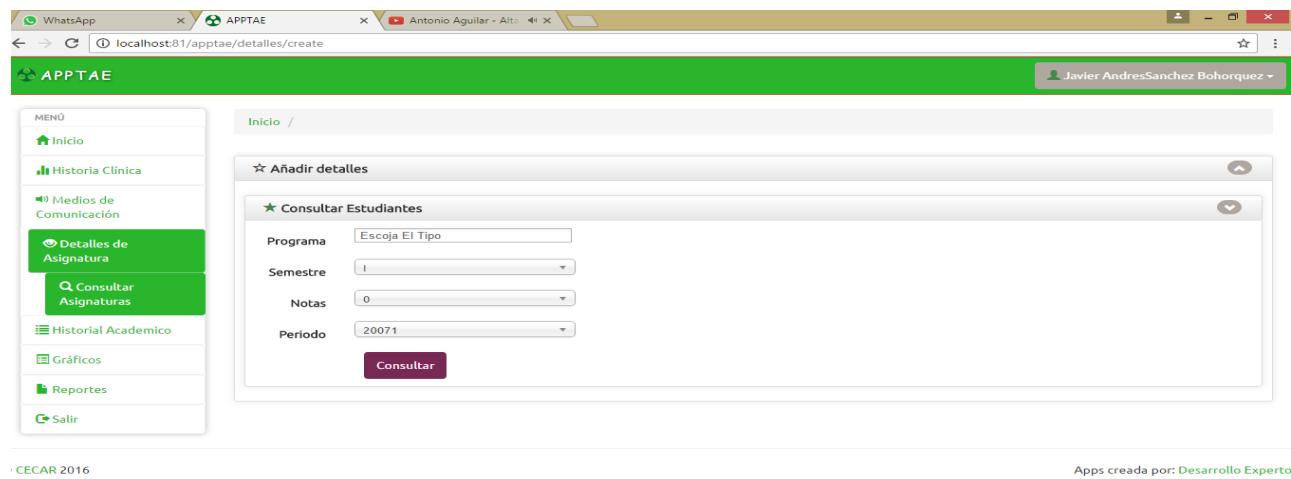


© CECAR 2016 Apps creada por: Desarrollo Experto



Detalles de las asignaturas que correspondan con el criterio establecido mostrando el nombre del profesor (Aquí se debe tener la posibilidad de enviar comunicados al docente), tipo asignatura (Es decir si es teórico práctica, a que área del pensum pertenece).

Fuente. Elaboración propia



© CECAR 2016 Apps creada por: Desarrollo Experto



Figura 31. Vista historial académico (filtro)

ANALISIS DE LA DESERCIÓN ESTUDIANTIL DE CECAR



WhatsApp APPTAE Antonio Aguilar - El A localhost:81/apptae/reprobados APPTAE

MENÚ /

Mostrar 10 registros Buscar:

CEDULA	NOMBRES Y APELLIDOS	MATERIA/PENSUL.	ASIGNATURA	PRIMER CORTE	SEGUNDO CORTE	SEMESTRE	PROGRAMA	Buscar
1082942701	TURIZO ECHEVERRIA YULY	CCO4015	ANALISIS Y DISEÑO II	2.00	1.70	7	100014	<button>Buscar</button>
1101441122	RICARDO ESCUDERO DANIEL ALBERTO	CCO7020	ESTRUCTURA DE DATOS II	2.00	3.10	4	100014	<button>Buscar</button>
1102799471	GUTIERREZ MENESES MARIO RAFAEL	MAT2055	INVESTIGACION DE OPERACIONES I	2.40	3.60	6	100014	<button>Buscar</button>
1577258682	VERCARA CUADRADO LUIS RAMON	EPP1010	ELECTIVA PROFESIONAL II	2.50	3.40	10	100014	<button>Buscar</button>
23030013	BERTEL SIERRA INGRID SOFIA	EPP1010	ELECTIVA PROFESIONAL II	2.30	2.80	10	100014	<button>Buscar</button>
23179913	ATENCIO JULIO MAYRULIN	CCO4010	ANALISIS Y DISEÑO I	2.70	3.20	6	100014	<button>Buscar</button>
258990628	VILLADIEGO RAMOS MARIA KELLY	CCO4010	ANALISIS Y DISEÑO I	2.90	3.60	6	100014	<button>Buscar</button>
258990628	VILLADIEGO RAMOS MARIA KELLY	MAT2055	INVESTIGACION DE OPERACIONES I	2.90	3.20	6	100014	<button>Buscar</button>
328990894	VERCARA ALVAREZ ELEUTH DEL CARMEN	MAT1028	CALCULO I	2.80	2.50	2	100014	<button>Buscar</button>
3414122	RIVERA PEREZ FREDY DE JESUS	EPP1010	ELECTIVA PROFESIONAL II	2.90	3.40	10	100014	<button>Buscar</button>

Mostrando registros del 1 al 10 de un total de 100 registros

— Anterior 1 2 3 4 Siguiente —

Windows Google Chrome Microsoft Edge 04:34PM 17/02/2017

Figura 32. Vista Respuesta de la Consulta (Detalles Asignatura)

Fuente. Elaboración propia

WhatsApp PDFReprobados Antonio Aguilar - El A localhost:81/apptae/PDFReprobados PDFReprobados 1 / 3

T.A.E Trayectoria Académica Exitosa CECAR Corporación Universitaria del Caribe Personería Jurídica N° 7786 MEN-ICFES - NIT. 892.201.263-1

PROGRAMA: SEMESTRE:

Reporte Histórico:

Cedula	Nombres Apellidos	Asignatura	Primer Corte	Segundo Corte	Semestre	Programa
850618-51726	PIZARRO MARQUEZ CARLOS ANDRES	ADMINISTRACION GENERAL	2.70	3.30	8	100014
92032853	CASTILLO GARCIA MANUEL ANTONIO	ELECTRONICA BASICA	2.90	3.30	8	100014
80112166	FONSECA OCHOA GUILLERMO ENRIQUE	ELECTRONICA BASICA	2.80	2.80	6	100014
92549997	GOMEZ PADILLA CARLOS DANIEL	SISTEMAS OPERATIVOS	2.90	4.30	7	100014
92641461	VILLALBA GARRIDO TONY JOSE	INGENIERIA DEL SOFTWARE	2.90	4.20	8	100014
23179913	ATENCIO JULIO MAYERLIN	ANALISIS Y DISEÑO	2.70	3.20	6	100014
92641987	ARROYO BETIN JEISSON ALFONSO	ANALISIS Y DISEÑO	1.80	3.70	8	100014
92549847	LUNA AGOSTA RICARDO JAVIER	ANALISIS Y DISEÑO	2.30	3.60	5	100014
92842071	SALCEDO QUIROZ WILMER MOISES	ANALISIS Y DISEÑO	1.00	4.90	8	100014
860224-34979	MANJARRES MERCADO KAREN SOFIA	LENGUAJES DE PROGRAMACION II	2.40	2.40	4	100014
80112166	FONSECA OCHOA GUILLERMO ENRIQUE	LENGUAJES DE PROGRAMACION II	1.10	2.60	4	100014
871222-63000	JARABA GALVAN JUAN CARLOS	FUNDAMENTOS DE PROGRAMACION II	1.00	00	2	100014
		ESTADISTICA	2.00	2.10	4	100014

Windows Google Chrome Microsoft Edge 04:35PM 17/02/2017

Figura 33. Vista de las Gráficas (Aprobados y Reprobados)

Fuente. Elaboración propia



Figura 34. Gráficas y análisis de las materias que se pierden y se ganan en un periodo.

Fuente. Elaboración propia

Se mostraran los siguientes: Código, Nombres (Completos), numero de faltas a clase en el semestres, numero de mensajes enviados desde TAE, numero de respuestas por parte del estudiante, Número de asignaturas reprobadas 1 corte, Número de asignaturas reprobadas 2 corte, Número de Asignaturas Perdidas En Semestre.

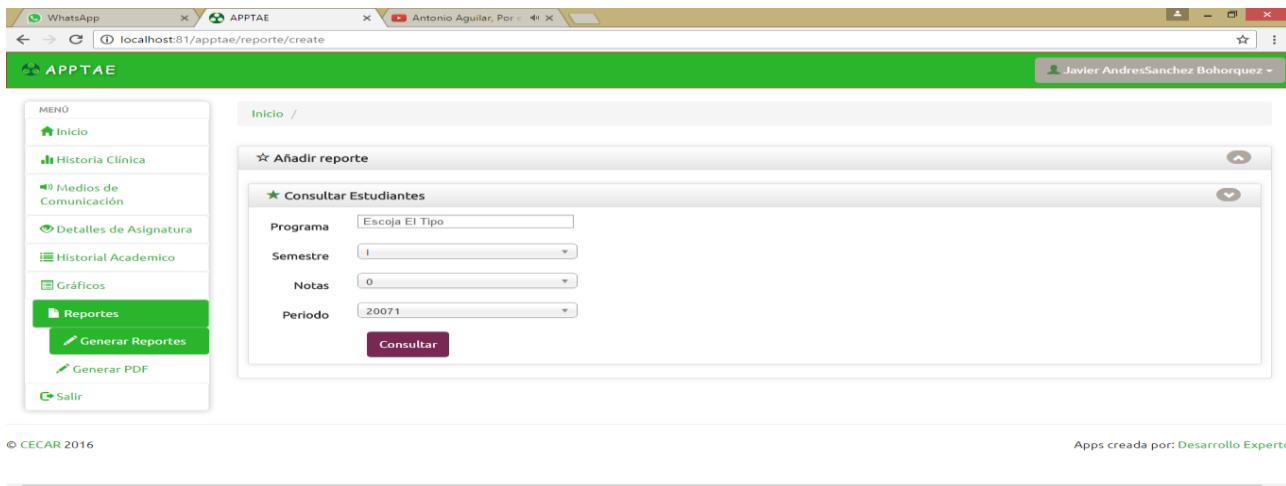
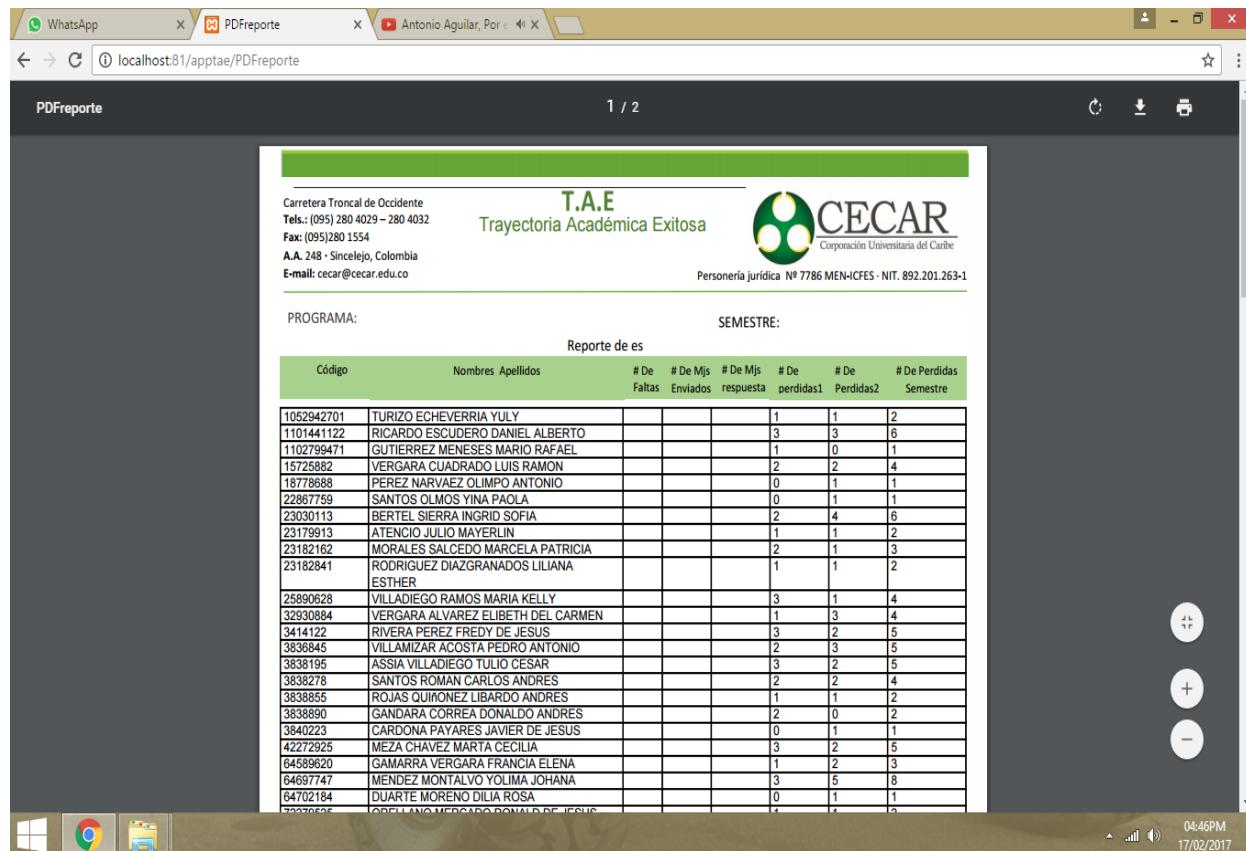


Figura 34.Fuente. Elaboración propia



Código	Nombres Apellidos	# De Faltas	# De Mjs Enviados	# De Mjs respuesta	# De perdidas1	# De Perdidas2	Semestre
1052942701	TURIZO ECHEVERRIA YULY			1	1	2	
1101441122	RICARDO ESCUDERO DANIEL ALBERTO			3	3	6	
1102799471	GUTIERREZ MENESES MARIO RAFAEL			1	0	1	
15725882	VERGARA CUADRADO LUIS RAMON			2	2	4	
18778688	PEREZ NARVAEZ OLIMPO ANTONIO			0	1	1	
22887759	SANTOS OLMO YINA PAOLA			0	1	1	
23030113	BERTEL SIERRA INGRID SOFIA			2	4	6	
23179913	ATENCIO JULIO MAYERLIN			1	1	2	
23182162	MORALES SALCEDO MARCELA PATRICIA			2	1	3	
23182841	RODRIGUEZ DIAZGRANADOS LILIANA ESTHER			1	1	2	
25890628	VILLADIEGO RAMOS MARIA KELLY			3	1	4	
32930884	VERGARA ALVAREZ ELIBETH DEL CARMEN			1	3	4	
3414122	RIVERA PEREZ FREDDY DE JESUS			3	2	5	
3836845	VILLAMIZAR ACOSTA PEDRO ANTONIO			2	3	5	
3838195	ASSIA VILLADIEGO TUJIO CESAR			3	2	5	
3838278	SANTOS ROMAN CARLOS ANDRES			2	2	4	
3838855	ROJAS QUIRONEZ LIBARDO ANDRES			1	1	2	
3838890	GANDARA CORREA DONALDO ANDRES			2	0	2	
3840223	CARDONA PAYARES JAVIER DE JESUS			0	1	1	
42272925	MEZA CHAVEZ MARTA CECILIA			3	2	5	
64589620	GAMARRA VERGARA FRANCIA ELENA			1	2	3	
64697747	MENDEZ MONTALVO YOLIMA JOHANA			3	5	8	
64702184	DUARTE MORENO DILIA ROSA			0	1	1	
72237656	CORELLANO MERCADO ROMALDO DE JESUS			1	1	2	

Figura 35. Vista Resultado del Reporte para Imprimir (Reporte).

Fuente. Elaboración propia

4.3.12. Solución Olap para Analizar la Deserción.

El cubo deserción permite al personal del programa TAE hacer una análisis desde diferentes puntos de vista de la deserción académica en la Corporación Universitaria del Caribe. En la implementación del cubo se tuvo en cuenta la definición de estudiante desertor aportada por la funcionaria de TAE, la cual lo identifica como aquel estudiante que no se ha matriculado durante dos periodos consecutivos.

Teniendo en cuenta lo anterior el cubo brinda la posibilidad al usuario final de analizar los estudiantes desertores desde diferentes puntos de vista, tales como: estado civil, estrato socioeconómico, lugar de residencia, programa académico al cual se encuentra adscrito, promedio

estudiantil, sexo y periodo académico (dimensión tiempo). Por ejemplo el usuario tiene la posibilidad de consultar los estudiantes matriculados, potencialmente desertores y desertores.

En la figura 34 se realiza la consulta anterior utilizando el cliente OLAP JPivot. De aquí podemos observar que el cubo le permite al funcionario del programa TAE consultar la información en forma consolidada o detallada, para el caso en particular, se consolidan los datos de todos los estudiantes de los programas presenciales en el periodo escogido que se encuentren matriculados, potencialmente en riesgo de deserción o hayan desertado

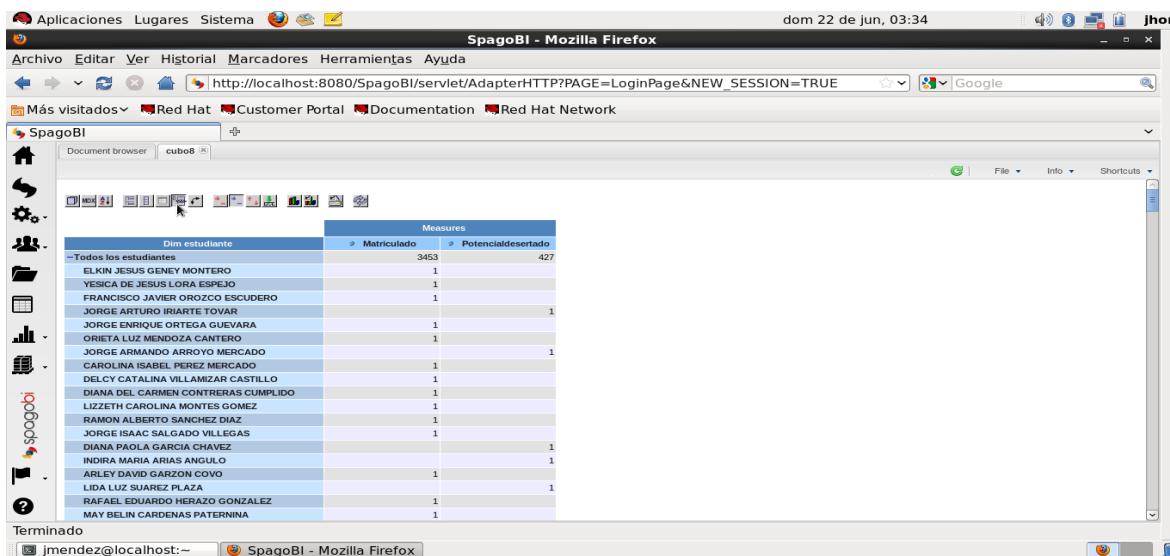
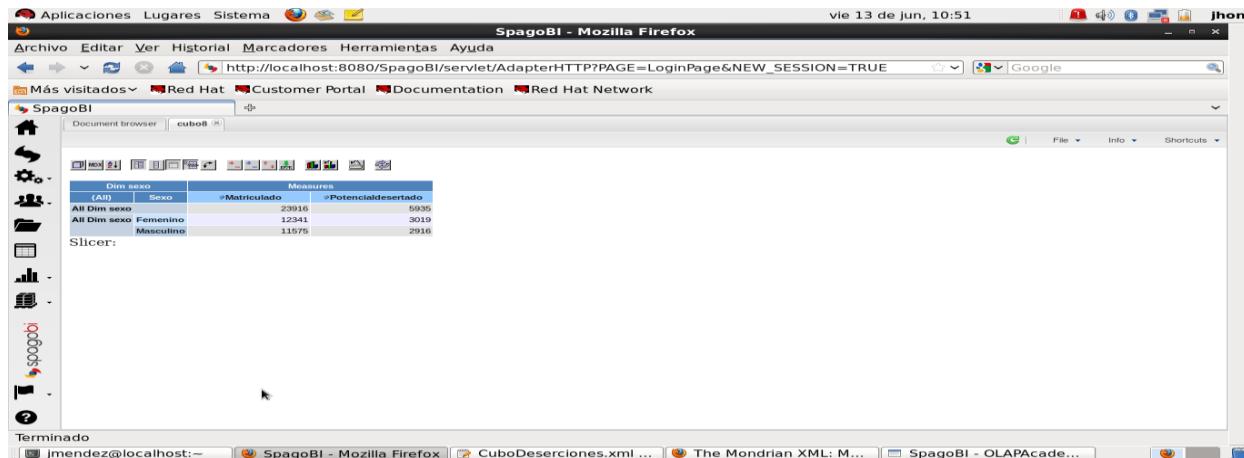


Figura 36. Consulta OLAP con JPivot que discrimina por estudiantes las medidas del cubo deserciones.

Fuente. Elaboración propia

Si el usuario lo desea puede consultar la información de desertados desde la dimensión sexo. La figura 40 muestra la consulta anterior, donde se observa que la mayor tasa de deserción la sufren las personas de sexo femenino.



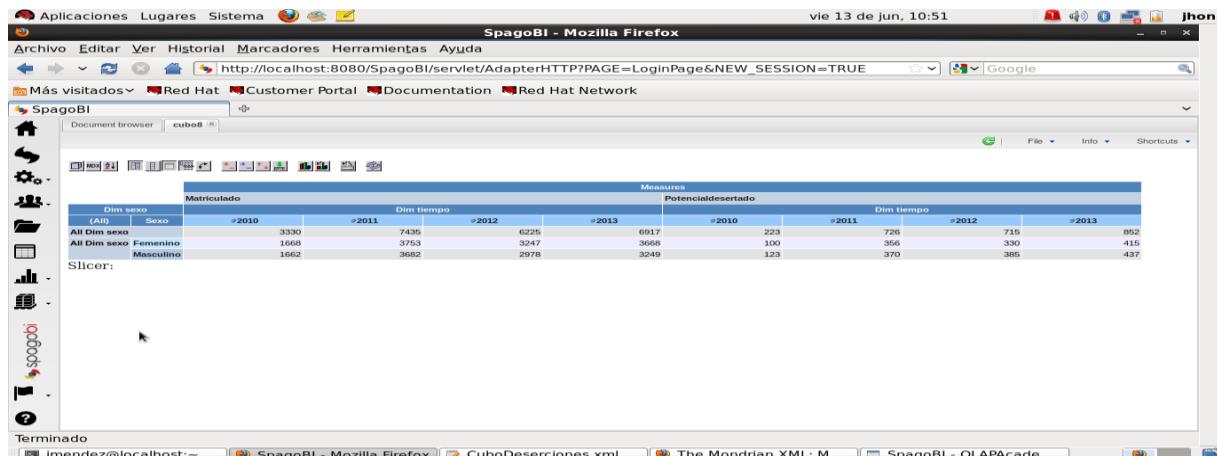
The screenshot shows a Mozilla Firefox browser window titled "SpagoBI - Mozilla Firefox". The address bar displays the URL http://localhost:8080/SpagoBI/servlet/AdapterHTTP?PAGE=LoginPage&NEW_SESSION=TRUE. The main content area shows an OLAP cube query results table. The table has two columns: "Dim sexo" and "Measures". The "Measures" column contains three items: "Matriculado", "Dim tiempo", and "Potencialdeseritado". The "Dim sexo" column has three rows: "(All)", "Femenino", and "Masculino". The "Matriculado" measure shows values: 23916, 12341, and 11575 respectively. The "Dim tiempo" measure shows values: 5935, 3019, and 2916 respectively. The "Potencialdeseritado" measure shows values: 6917, 3668, and 3249 respectively.

Dim sexo	Measures
(All)	Matriculado Dim tiempo Potencialdeseritado
All Dim sexo	23916 5935 6917
All Dim sexo Femenino	12341 3019 3668
All Dim sexo Masculino	11575 2916 3249

Figura 37.Consulta OLAP con JPivot que discrimina por sexo las medidas del cubo deserciones.

Fuente. Elaboración propia

De igual forma, se puede analizar la deserción en los últimos periodos académicos discriminando por sexo y condicionando por los últimos periodos académicos las medidas se el cubo (ver figura 38).



The screenshot shows a Mozilla Firefox browser window titled "SpagoBI - Mozilla Firefox". The address bar displays the URL http://localhost:8080/SpagoBI/servlet/AdapterHTTP?PAGE=LoginPage&NEW_SESSION=TRUE. The main content area shows an OLAP cube query results table. The table has four columns: "Dim sexo", "Measures", "Dim tiempo", and "Potencialdeseritado". The "Measures" column contains three items: "Matriculado", "Dim tiempo", and "Potencialdeseritado". The "Dim sexo" column has three rows: "(All)", "Femenino", and "Masculino". The "Matriculado" measure shows values: 3330, 1668, and 1662 respectively. The "Dim tiempo" measure shows values: 2010, 2011, 2012, and 2013. The "Potencialdeseritado" measure shows values: 6917, 3668, 3249, 726, 100, 356, 123, 715, 852, 415, 385, and 437 respectively.

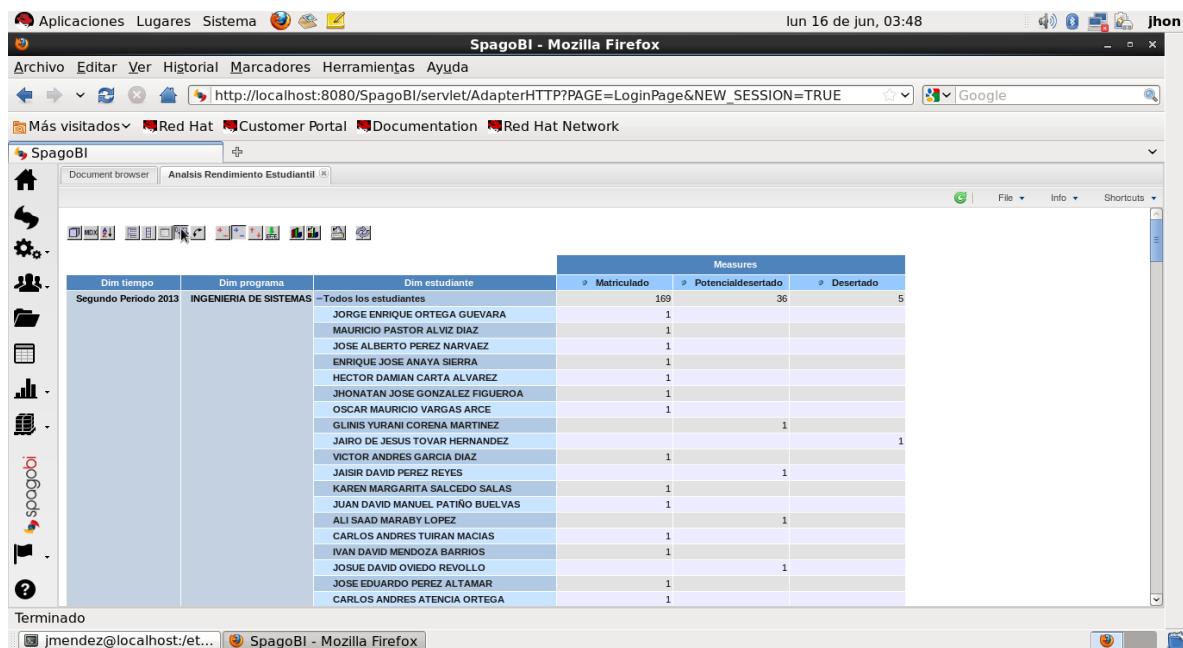
Dim sexo	Measures	Dim tiempo	Potencialdeseritado
(All)	Matriculado Dim tiempo	>2010 >2011 >2012 >2013	6917 3668 3249 726 100 356 123 715 852
All Dim sexo			
All Dim sexo Femenino	3330 1668 1662	3753 3247 3682	7435 6225 2978
All Dim sexo Masculino			

Figura 38.Consulta OLAP con JPivot que discrimina por sexo y condiciona por tiempo las medidas del cubo deserciones.

Fuente. Elaboración propia

En el mismo sentido se pueden realizar consultas a partir de diferentes puntos de vista simultáneamente. Por ejemplo estudiantes desertados por periodo y programa académico (ver imagen 42).

Así como el cubo anterior, la explotación del cubo deserciones queda en disposición de los funcionarios del programa TAE y coordinadores de programa para analizar las diferentes combinaciones de puntos de vista en aras de encontrar patrones que ayuden en la generación de estrategias que apunten a la disminución de las tasas de deserción con las que cuenta la Corporación Universitaria del Caribe CECAR.



The screenshot shows a SpagoBI interface running in Mozilla Firefox. The title bar reads "SpagoBI - Mozilla Firefox". The main content area displays an OLAP cube titled "Analysis Rendimiento Estudiantil". The cube has three dimensions: "Dim tiempo" (Second Period 2013), "Dim programa" (INGENIERIA DE SISTEMAS), and "Dim estudiante". The measures displayed are "Matriculado", "Potencialdesertado", and "Desertado". The data table lists 18 students with their respective counts for each measure. The interface includes a sidebar with various icons for navigation and a toolbar at the top.

Measures			
	Matriculado	Potencialdesertado	Desertado
Todos los estudiantes	169	36	5
JORGE ENRIQUE ORTEGA GUEVARA	1		
MAURICIO PASTOR ALVIZ DÍAZ	1		
JOSE ALBERTO PÉREZ MARVÁEZ	1		
ENRIQUE JOSE ANAYA SIERRA	1		
HECTOR DAMIAN CARTA ALVAREZ	1		
JHONATAN JOSE GONZALEZ FIGUEROA	1		
OSCAR MAURICIO VARGAS ARCE	1		
GLINIS YURANI CORENA MARTINEZ		1	
JAIRO DE JESUS TOVAR HERNANDEZ			1
VICTOR ANDRES GARCIA DIAZ	1		
JAISIN DAVID PÉREZ REYES		1	
KAREN MARGARITA SALCEDO SALAS	1		
JUAN DAVID MANUEL PATIÑO BUELVAS	1		
ALI SAAD MARABY LOPEZ		1	
CARLOS ANDRES TURAN MACIAS	1		
IVAN DAVID MENDOZA BARRIOS	1		
JOSUE DAVID OVIDIO REVOLLO		1	
JOSE EDUARDO PÉREZ ALTAMAR	1		
CARLOS ANDRES ATENCIA ORTEGA	1		

Figura 39. Consulta OLAP con JPivot que discrimina por periodo, programa y estudiantes las medidas del cubo deserciones.

Fuente. Elaboración propia

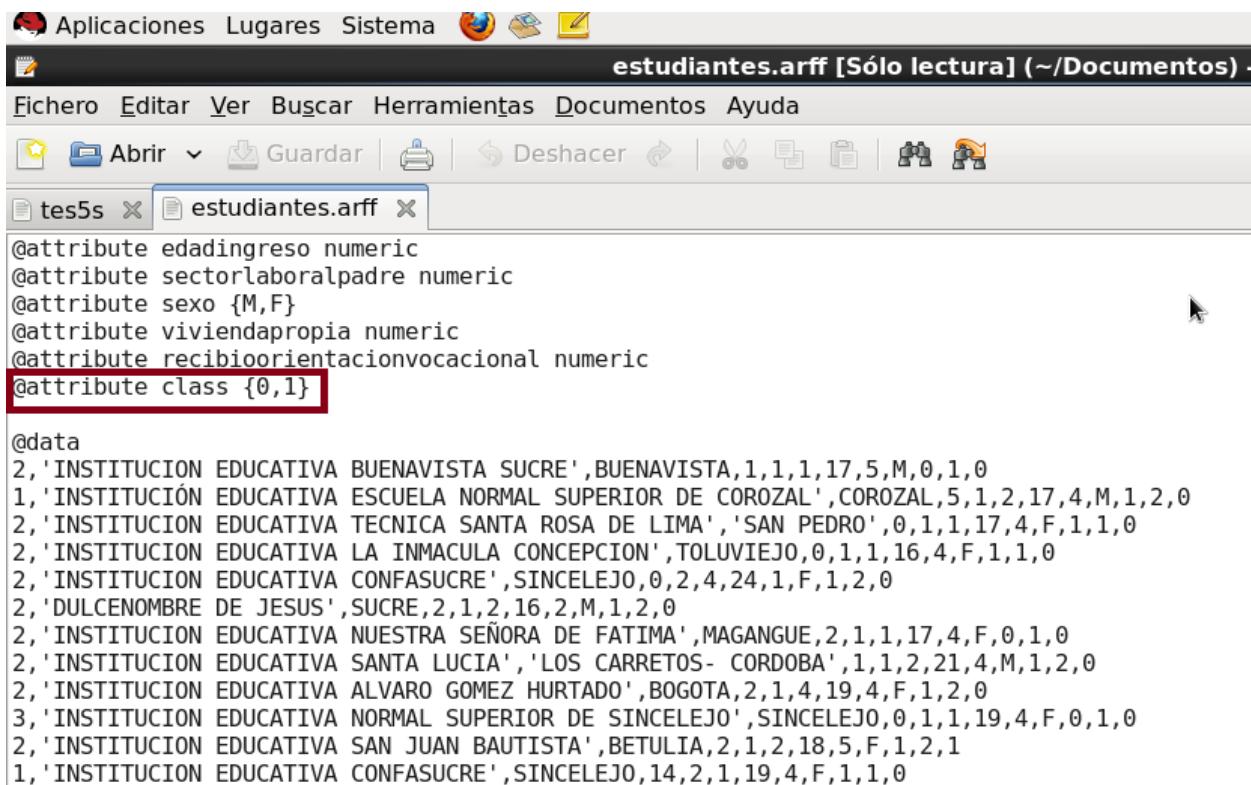
Solución Minería de Datos

Uno de los principales objetivos del desarrollo tecnológico es realizar un análisis de la deserción bajo los parámetros y el contexto del estudiante Cecarence. Para lo anterior se utilizó la herramienta de minería de datos WEKA con el fin de caracterizar el estudiante con riesgo de abandono académico.

Para lo anterior y siguiente el modelo del proceso del KDD (Knowledge Discovery in Databases) se construyó una vista minable con datos personales de cada estudiante tomado de las bases de datos transaccionales de CECAR: Academia y Snies. A continuación se mencionan los datos que fueron tomados en cuenta para la vista:

Datos	<ul style="list-style-type: none">• Identificación• Nombres• Estado civil• Estrato económico• Sexo• Programa que estudia• Ciudad de residencia• Promedio estudiantil• Si trabaja o no• Edad de ingreso• Número de asignaturas reprobadas a los largo de su estadía en la corporación universitaria• Grupo étnico al que pertenece• Sector donde labora el padre• Grupo familiar• Posee vivienda propia o no• Si recibió alguna orientación vocacional• Tipo de orientación vocacional en caso de haberla recibido• Deserto
-------	---

A partir de los datos anteriores se creó un archivo con extensión arff para ser procesado con el motor de minería WEKA el cual está compuesto por 17 campos (las columnas especificadas en la tabla anterior) y 2012 filas (estudiantes de CECAR los cuales cuentan con los datos anteriores). Además se agregó una columna que reflejara si el estudiante ha desertado o no (Ver figura 39).



```

@attribute edadingreso numeric
@attribute sectorlaboralpadre numeric
@attribute sexo {M,F}
@attribute vivienda propia numeric
@attribute recibio orientacion vocacional numeric
@attribute class {0,1}

@data
2,'INSTITUCION EDUCATIVA BUENAVISTA SUCRE',BUENAVISTA,1,1,1,17,5,M,0,1,0
1,'INSTITUCIÓN EDUCATIVA ESCUELA NORMAL SUPERIOR DE COROZAL',COROZAL,5,1,2,17,4,M,1,2,0
2,'INSTITUCION EDUCATIVA TECNICA SANTA ROSA DE LIMA','SAN PEDRO',0,1,1,17,4,F,1,1,0
2,'INSTITUCION EDUCATIVA LA INMACULADA CONCEPCION',TOLUVIEJO,0,1,1,16,4,F,1,1,0
2,'INSTITUCION EDUCATIVA CONFASUCRE',SINCELEJO,0,2,4,24,1,F,1,2,0
2,'DULCENOMBRE DE JESUS',SUCRE,2,1,2,16,2,M,1,2,0
2,'INSTITUCION EDUCATIVA NUESTRA SEÑORA DE FATIMA',MAGANGUE,2,1,1,17,4,F,0,1,0
2,'INSTITUCION EDUCATIVA SANTA LUCIA','LOS CARRETOS- CORDOBA',1,1,2,21,4,M,1,2,0
2,'INSTITUCION EDUCATIVA ALVARO GOMEZ HURTADO',BOGOTA,2,1,4,19,4,F,1,2,0
3,'INSTITUCION EDUCATIVA NORMAL SUPERIOR DE SINCELEJO',SINCELEJO,0,1,1,19,4,F,0,1,0
2,'INSTITUCION EDUCATIVA SAN JUAN BAUTISTA',BETULIA,2,1,2,18,5,F,1,2,1
1,'INSTITUCION EDUCATIVA CONFASUCRE',SINCELEJO,14,2,1,19,4,F,1,1,0

```

Figura 40. Anexo de columna

Fuente. Elaboración propia

Para realizar la prueba se dividió el archivo anterior en dos: uno que se tomó para entrenar el modelo y el otro se utilizara para probar y ajustar el modelo el modelo generado por la herramienta.

Por su parte, después de analizar el número de estudiantes que no desertan (1752) barra azul en relación a los que no desertan (337 barra roja ver imagen 42) se evidenció una alta cantidad de estudiantes que no desertan. Por tal razón, se consideró que los datos de las dos clases se encontraban desbalanceados y al utilizar los algoritmos de clasificación sobre estos datos, provocaría que se tengan en la fase de entrenamiento una clase mayoritaria, lo que llevaría a clasificar en la fase de prueba estudiantes con baja sensibilidad a los elementos de la clase minoritaria.

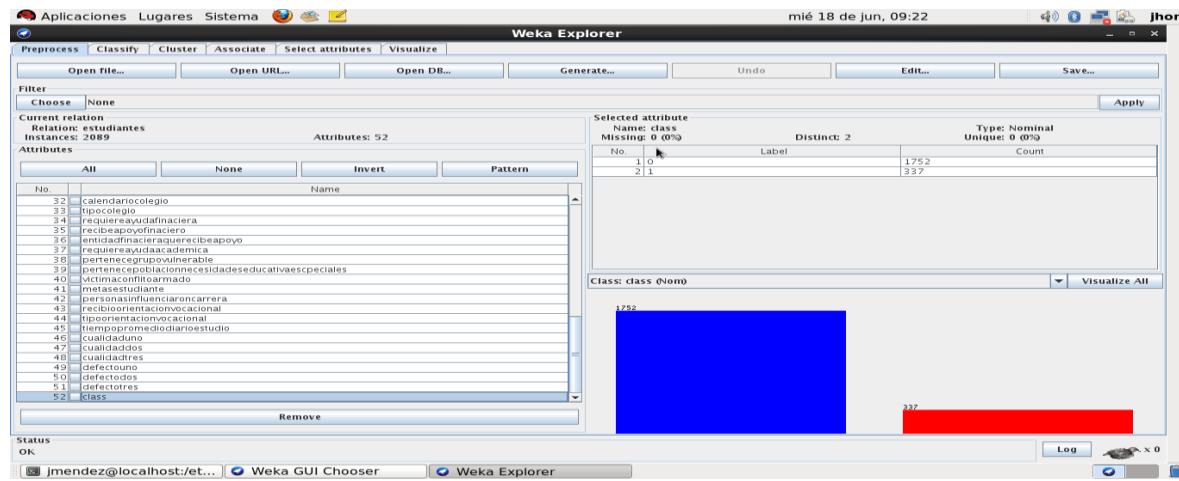


Figura 41. Análisis del número de estudiantes

Fuente. Elaboración propia

Para solucionar el problema anterior se realizó un balanceo de la distribución de clases, utilizando el algoritmo de rebalanceo SMOTE (Synthetic Minority Oversampling Technique) disponible en Weka, de forma que se tuviera el 50% de instancias de estudiantes no desertados y 50% de desertados (ver imagen 45)

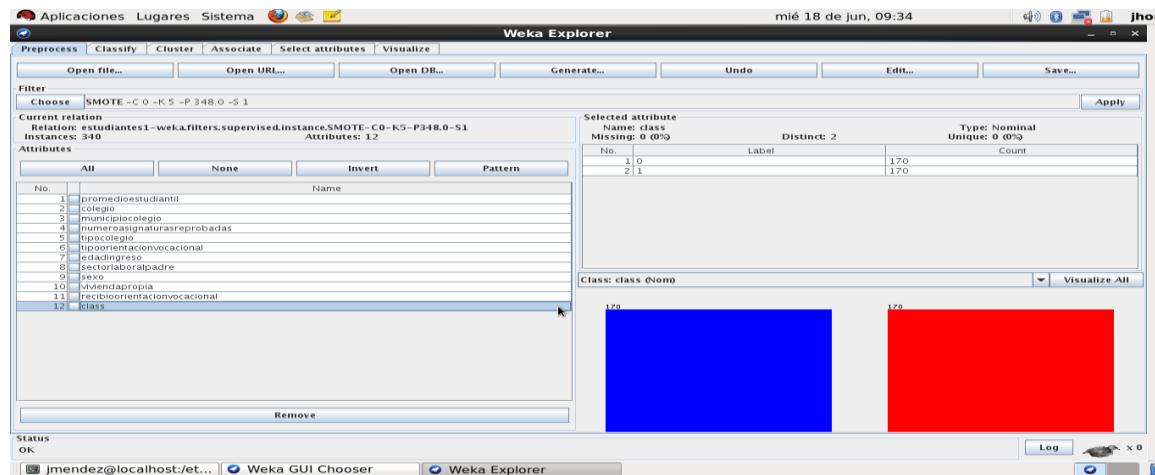


Figura 41. Algoritmo de rebalanceo SMOTE

Fuente. Elaboración propia

Seguidamente se seleccionaron 5 algoritmos de clasificación del tipo “caja blanca”, es decir, aquellos donde se obtiene un modelo de salida comprensible para el usuario, proporcionado reglas de clasificación del tipo “Si – Entonces” o árboles de decisión. Los algoritmos seleccionados son: JRip, NNge, OneR, Ridor y J48.

Seguidamente se realizaron tres experimentos con el objetivo de obtener la máxima exactitud de clasificación obteniendo los siguientes resultados:

Tabla 17.

Promedio validación cruzada utilizando las particiones con todos los atributos

Algoritmo	% instancias correctas	% instancias incorrectas	Coeficiente Kappa	Error absoluto
JRip	82,50001	17,49999	0,25061	0,24953
NNge	70,62501	29,37499	0,02405	0,29375
OneR	80,57693	19,42307	0,00198	0,19422
Ridor	81,68268	18,31732	0,16853	0,18316
J48	81,97116	18,02884	0,14889	0,25892

Tabla 18.

Promedio validación cruzada utilizando las particiones con mejores atributos

	% instancias correctas	% instancias incorrectas	Coeficiente Kappa	Error absoluto
JRip	81,49039	18,50961	0,11652	0,25037
NNge	72,88463	27,11537	-0,02701	0,27115
OneR	82,01922	17,98078	-0,02836	0,17981
Ridor	82,40386	17,59614	0,05771	0,17597
J48	82,35578	17,64422	0,06182	0,25133

Tabla 19.

Promedio validación cruzada utilizando las particiones re-balanceadas con mejores atributos

Algoritmo	% instancias correctas	% instancias incorrectas	Coeficiente Kappa	Error absoluto
JRip	85,051345	14,948655	0,70102727	0,21519091
NNge	83,057227	16,942773	0,66115455	0,16943636

OneR	88,350618	17,2077	0,76700909	0,11648182
Ridor	83,60925	16,39075	0,67219	0,16393
J48	78,0527	21,9473	0,56106364	0,28494545

A continuación se muestran y analizan algunos de los modelos de reglas o árboles de clasificación que han sido generados por los algoritmos que mejores resultados de clasificación obtuvieron en la etapa anterior de experimentación y en el experimento en el que se obtuvieron los mayores porcentajes de instancias clasificadas correctamente, los cuales son: OneR (ver Tabla 20), JRip (ver Tabla 21) y Ridor (ver Tabla 22). Es importante anotar que solo se tendrán en cuenta las reglas o ramas de los árboles relacionadas a estudiantes desertores.

Tabla 20.

Resultados obtenidos por OneR para los estudiantes desertores

Resultados OneR	
LICEO MODERNO DEL LITORAL	-> 1
INTITUCION EDUCATIVA SAN JORGE	-> 1
INSTITUCION EDUCATIVA NUESTRA SEÑORA DE FÀTIMA	-> 1
INSTITUCION EDUCATIVA MIGUEL DE SERVANTES SAAVEDRA	->
1	
CEDI	-> 1
INSTITUCIÓN EDUCATIVA LUIS CARLOS GALÁN	-> 1
INSTITUCION EDUCATIVA SANTIAGO APOSTOL	-> 1
COLEGIO TERCER MILENIO	-> 1
INSTITUCION EDUCATIVA DE MACAJAN	-> 1
INSTITUTO LUIS CARLOS GALAN	-> 1
INSTITUCION EDUCATIVA GABRIEL TABOADA SANTODOMINGO	->
1	
INSTITUTO EDUC. SAN MARCOS	-> 1
INSTUTITUCION EDUCATIVA RURAL PTO CLAVER	-> 1
COLPEC	-> 1
INST. EDUC. ANTONIO PRIETO	-> 1
INSTITUCION EDUCATIVA LICEO 20 DE JULIO	-> 1
INSTITUCION EDUCATIVA LICEO POLITECNICO DEL SINU	-> 1
INSTITUCION EDUCATIVA INDIGENA TECNICO AGROPECUARIA DE ESCOBAR ARRIBA	-> 1
LICEO BARTOLOME DE LAS CASAS	-> 1
LICEO DE CERVANTES SAAVEDRA	-> 1
INSTITUCIÓN EDUCATIVA PARA POBLACIONES ESPECIALES INPES	->
1	

	INSTITUTO SUR ORIENTAL SAN JORGE	-> 1
	INSTITUCION EDUCATIVA TECNICO AGROPECUARIO ARTESANAL	->
1	INST. EDUCATIVA NORMAL SUPERIOR SINCELEJO	-> 1
	INSTITUCIÓN EDUCATIVA COMFASUCRE	-> 1
	INSTITUCION EDUCATIVA JUAN JACOBO ROSEAN-> 1	
	INSTITUCIÓN EDUCATIVA COMFASUCRE	-> 1
	INSTITUCION EDUCATIVA JUAN JACOBO ROSEAN-> 1	
	INSTITUCIÓN EDUCATIVA SAN MARTIN INESAM	-> 1
	INSITUCIÓN EDUCATIVA LICEO MODERNO DEL LITORAL	-> 1
	INSTITUCION EDUCATIVA TECNICA COMERCIAL MARIA INMACULADA	-> 1
	INSTITUCION EDUCATIVA AVANZADO ICSA	-> 1
	INSTITUCION EDUCATIVA NUEVO MILENIO-> 1	
	LICEO MODERNO EL LITORAL	-> 1
	INSTITUCIÓN EDUCATIVA TECNICA ALVARO ULCUE CHOCUE	-> 1
	INSTITUCION EDUCATIVA LIMASOR	-> 1
	INSTITUCIÓN EDUCATIVA CUCUTA	-> 1
	COLEGIO SAN RAFAEL	-> 1
	INSTITUTO TENICO JUAN MEJIA GOMEZ	-> 1
	INSTITUCION EDUCATIVA LOS PALMITOS.	-> 1
	INTITUCION EDUCATIVA IGMNASIO DEL ROSARIO	-> 1
	INSTITUCION EDUCATIVA SAN JOSE DE MAJAGUAL	-> 1

Fuente. Elaboración propia

Tabla 21.

Resultados obtenidos por JRip para los estudiantes desertores

Resultados JRip
(promedioestudiantil <= 1) and (numeroasignaturasreprobadas <= 7) and (tipoorientacionvocacional >= 5) => deserto=1 (38.0/15.0)

Tabla 22.

Resultados obtenidos por Ridor para los estudiantes desertores

Resultados Ridor
deserto = 0 (2089.0/337.0)
Except (promedioestudiantil <= 1.5) and (numeroasignaturasreprobadas <= 7.5) and (edadingreso <= 18.5) and (edadingreso > 17.5) and (tipocolegio > 1.5) and (tipoorientacionvocacional > 2.5) => deserto = 1 (7.0/0.0) [3.0/1.0]

Except (promedioestudiantil <= 1.5) and (numeroasignaturasreprobadas <= 7.5) and
 (edadingreso <= 18.5) and (edadingreso <= 16.5) and (tipocolegio <= 1.5) and
 (recibioorientacionvocacional > 1.5) => deserto = 1 (12.0/1.0) [9.0/4.0]

Except (promedioestudiantil <= 1.5) and (viviendapropia <= 0.5) and
 (situacionlaboralpadre > 1.5) => deserto = 1 (22.0/4.0) [12.0/6.0]

Total number of rules (incl. the default rule): 4

Time taken to build model: 0.08 seconds

Evaluar la información del estudiante desertor con el fin de caracterizar y proyectar los estudiantes con riesgo a desertar.

A continuación se analizan las reglas y árboles de clasificación generados por los mejores algoritmos seleccionados anteriormente. La tabla 40 muestra el resultado obtenido por el algoritmo OneR, este solo descubre como regla los colegios en los cuales el estudiante Cecarense realizó sus estudios de secundaria y en donde existe una probabilidad más alta de deserción. Es importante señalar la importancia de realizar más adelante un estudio a fondo en la relación que existe entre estos colegios y la deserción, de forma que se analicen los tipos de colegios, metodologías, currículo, entre otras variables.

La tabla 21 se muestra una regla generada por el algoritmo JRip, indicando que aquellos estudiantes con riesgo de desertar presentan un promedio bajo, un número de asignaturas reprobadas entre 1 y 7 y recibieron ayuda vocacional de tipo charla o test.

Por su parte la tabla 22 muestra varias reglas generadas por el algoritmo Ridor para estudiantes con riesgo de desertar. La primera indica que aquellos estudiantes con promedio estudiantil bajo, edad de ingreso entre 17 y 18 años, tipo de colegio público y tipo de orientación vocacional de tipo test están en riesgo de abono académico. Otra regla indica que estudiantes con promedio estudiantil bajo, edad de ingreso menor o igual a 18 años, tipo de colegio público y no recibió orientación vocacional corren riesgo de desertar. Por último se tiene que estudiantes con promedio bajo, carecen de vivienda propia y sus padres no poseen trabajo se encuentren en igual medida propensos a abonar los programas académico que se encuentran cursando.

Después de haber realizado las pruebas anteriores se analizaron los resultados obtenidos por los algoritmos seleccionados de clasificación, detectando un consenso sobre factores comunes que influyen en la deserción de los estudiantes de la Corporación Universitaria de Caribe CECAR. Entre esos factores tenemos el promedio Estudiantil y el número de asignaturas reprobadas.

De igual forma se pueden señalar otros factores, que si bien no fueron escogidos en algunos algoritmos vale la pena su estudio y análisis de influencia en el fenómeno de la deserción estudiantil de CECAR, como son: edad de ingreso, si recibió orientación vocacional, tipo de orientación vocacionales que recibió, tipo y colegio donde realizó los estudios de secundaria, si posee vivienda propia y si el padre se encuentra laborando.

Cronograma

Conclusiones

Al culminar el proyecto se concluye lo siguiente:

Se comprobó que la inteligencia de negocios y en particular las herramientas de tipo OLAP, permiten a los usuarios finales no expertos realizar consultas de manera sencilla que involucren el análisis desde diferentes puntos de vista y la generación de gráficos que les brinden nuevas perspectivas para la definición de estrategias que conlleven al éxito de las organizaciones.

Los factores que más influyen en el fenómeno de la deserción académica de la Corporación Universitaria del Caribe –CECAR son el promedio estudiantil y el número de asignaturas cursadas. Otros factores que influyen en menor medida son la edad de ingreso, si recibió orientación vocacional, tipo de orientación vocacional que recibió, tipo y colegio donde realizó los estudios de secundaria, si posee vivienda propia y si el padre se encuentra laborando

Al aplicar en la minería de datos técnicas de clasificación es importante determinar si los datos se encuentran balanceados. En caso que no se encuentren se recomienda balancearlos utilizando algoritmos como el SMOTE (Synthetic Minority Oversampling Technique) debido a que junto con la selección de atributos mejoran los porcentajes de instancias clasificadas correctamente

Recomendaciones

Se recomienda lo siguiente:

Integrar la herramienta con el nuevo sistema de información de gestión académica para que el modelo generado aprenda y genere patrones con una mayor exactitud.

Generar un módulo que integre en tiempo real el patrón y con los datos de estudiantes, de forma que se obtenga un listado de posibles desertores.

Se propone la utilización de tableros de control que permitan a los usuarios tener una visión integrada de la estrategia de negocios de la Corporación Universitaria del Caribe tanto en lo académico como en lo administrativo, estableciendo un sistema de medición del logro de objetivos propuestos a través de indicadores de desempeño.

Referencias Bibliográficas

Alejo, I. Gil A, E. López y J. A. Velásquez, A (2011). *Aplicación de técnicas de minería de datos para predecir deserción*, la Universidad Tecnológica de Izúcar de Matamoros en México.

Allan, A. Alejandra, S. (2005). La deserción estudiantil en la educación superior: el caso de la universidad DE COSTA RICA Disponible en:
http://revista.inie.ucr.ac.cr/uploads/tx_magazine/deserc_01.pdf

Andreu, R; Ricart, J.E. ;VALOR, J.(1996). “*Estrategia y sistemas de Información; Ed. McGraw-Hill” Segunda edición.*

Andreu, R; Ricart J; Valor, J (1997). “*Estrategia y sistemas de Información.McGraw-Hill, Segunda edición,*” Madrid.

Brien, J. Marakas, G (2010). Introduction to Information Systems. Mc Graw-Hill.

Bhavnagar, S y Krishnakumar, M (2013). “*Introduction on Data Warehouse with OLTP and OLAP. International Journal Of Engineering And Computer Science.*” Volumen 2, Numero 8.

Bichsel, J, (2013). Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations. Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations. EDUCAUSE Center for Applied Research. 2013.

Briones, G (1996). “*Metodología de la Investigación Cuantitativa en las ciencias Sociales*”. Barcelona: Gedisa.

Cardozo, A (2013). lunar en educación superior [Versión electrónica] Disponible en: http://www.elcolombiano.com/BancoConocimiento/D/desercion_lunar_en_educacion_superior/desercion_lunar_en_educacion_superior.asp.

Castaño E., Gallón, S., Gómez, K. y Vásquez, J.(2008) análisis de los factores asociados a la deserción estudiantil en la educación superior: un estudio de caso [Versión electrónica] (2008) 255-280, 2009 Disponible en: http://www.revistaeducacion.mec.es/re345/re345_11.pdf

Carmona, E (2008) Una propuesta de dashboard digital del docente como estrategia de gestión personal del conocimiento en el ámbito académico y su aplicación en la universidad de Quindío. Quindío, 2008. Trabajo de grado de Ingeniero Informático. Universidad de las palmas en gran canaria.

Carmona, E. (2007) Una propuesta de dashboard digital del docente como estrategia de gestión personal del conocimiento en el ámbito académico y su aplicación en la universidad de Quindío.

Ceballos, M. Villota, D. (2007). Factores asociados a la deserción estudiantil en la cohorte 2003 periodo B del programa de licenciatura en educación básica con énfasis en humanidades lengua castellana e inglés [Versión electrónica] Disponible en: http://www.colombiaaprende.edu.co/html/micrositios/1752/articles-323185_recurso_10.pdf

Cohen K; Asín, L (2009) Tecnologías de Información en los Negocios. Mc Graw Hill.

Cohen K, D. Asín L, E (2009). “*Tecnología de Información en los Negocios. Quinta Edición. Mc Graw Hill*” México.

Chaudhuri,S, Dayal,U y Narasayy,V (2011) An Overview of Business Intelligence Technology.
En: communications of the acm. Número 8, Volumen 54.

Data Warehousing Institute. Data Warehousing (2013). [Versión electrónica] Disponible en:
<http://tdwi.org/portals/data-warehousing.aspx>.

Data Warehousing Institute (2013) TDWI -The Data Warehousing Institute. [Versión electrónica]
Disponible en: <http://tdwi.org/Home.aspx>.

Dean, J & Ghemaw (2008). “*Sanjay.MapReduce: simplified data processing on large clusters*.Communications of the ACM.” Volumen 51, Numero 1.

Dordogne, A (2006). “*Redes Informáticas Conceptos Fundamentales*”. 1 Edición. Barcelona.

Dresner, H (1989). “*Business Intelligence. Journeys and the Roadmap for change.*” Volumen.
1989.

Fiorenza, Pat (2010). Using Business Intelligence Tools to Improve School Districts. [Versión electrónica] Disponible: <http://www.govloop.com/profiles/blogs/using-business-intelligence-tools-to-improve-school-districts>.

Frade, Abril; O, D; Castillo, Pérez y Nelson,J (2007). “*Estado actual de las tecnologías de bodega de datos y OLAP aplicadas a bases de datos espaciales.*” Ingeniería e Investigación.
Volumen 27, Numero 1

Gartner. P (2013) Business Intelligence. [Versión electrónica] Disponible en: www.com/technology/core/products/research/topics/businessIntelligence.jsp

González, L., & Uribe, D. (2005) Estimaciones sobre la repitencia y deserción superior chilena. Consideraciones sobre sus implicaciones. Santiago.tomado de:
<http://portales.mineduc.cl/usuarios/bmineduc/doc/201209281737360.EVIDENCIASCEM9.pdf>

Husemann, B; Lechtenbörger, J y Vossen, G (2000) “*Conceptual Data Warehouse Desing. In Proc. of the International Workshop on Design and Management of DataWarehouses.*” Volumen 1.

Jadrić, M; G, Ž y Čukušić, M (2010) Student Dropout Analysis with Application of Data Mining Methods. Management: Journal of Contemporary Management Issues.Volumen 15.

Kimball, R y Margy R (2013). “*The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling.*” 3 Edición. Kimball Group.

Kimberly, A (2013) Signals: Applying Academic Analytics [Versión electrónica] Disponible en:
<http://www.educause.edu/ero/article/signals-applying-academic-analytics>

Kotsiantis, S (2009) Educational data mining: a case study for predicting dropout-prone students. International Journal of Knowledge Engineering and Soft Data Paradigms. En: Journal International Journal of Knowledge Engineering and Soft Data Paradigms. Volumen 1. Geneva, SWITZERLAND.

Laudon, K; Laudon, J (2007) Sistemas de Información Gerencial Administración de la Empresa Digital.

Laudon, K y Laudon, J (2012). “*Sistemas de Información Empresarial. Pearson*” Décimo Segunda Edición, 2012. Mexico.

Luhn, P (1958). “*A Business Intelligence System. 1958. IBM Journal.*”. Volume 2, Number 4, Page.

Maharaja K y Sinhji B (2013). “*Introduction on Data Warehouse with OLTP and OLAP.*”. India. International Journal Of Engineering And Computer Science ISSN: 2319-7242. Volumen 2.

Maureen L., Fernández V. (2009) La gestión del valor de la cartera de clientes y su efecto en el valor global de la empresa: diseño de un modelo explicativo como una herramienta para la toma de decisiones estratégicas de marketing. [Versión electrónica] Disponible en: <http://eprints.ucm.es/8064/1/T29976.pdf> Sistema para la prevención de la deserción de la educación superior, (2013). Momento (semestre) en el cual los estudiantes desertores abandonan sus estudios en el nivel de formación universitario año 2013 [Versión electrónica] SPADIES, 2013 2 (3). Disponible en:

<http://www.colombiaaprende.edu.co/html/micrositios/1752/w3-article-343426.html>.

Marianne, M, (2009). Ohio Schools Use Business Intelligence To Improve Student Performance. En: Informationweek. [Versión electrónica] Disponible en: http://www.informationweek.com/ohio-schools-use-business-intelligence-to-improve-student-performance/d/d-id/1080159?page_number=2

Ministerio de educación nacional, (2006). Deserción estudiantil en la educación superior colombiana [Versión electrónica] Mineducación, 2006 Disponible en: http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_despcion.pdf

Ministerio de educación nacional, (2009). Deserción estudiantil en la educación superior colombiana [Versión electrónica] MINEDUCACION, 2009 Disponible en:

http://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_libro_despcion.pdf

Ministerio de educación nacional, (2015). Deserción estudiantil en la educación superior colombiana [Versión electrónica] MINEDUCACION, 2015 Disponible en: <http://www.colombiaaprende.edu.co/html/micrositios/1752/w3-article-350629.html>

Ocaranza,O y Quiroz,M (2006) Deserción Estudiantil del Pregrado en la Pontificia Universidad Católica de Chile. [Versión electrónica] Disponible en: http://aprendeenlinea.udea.edu.co/lms/investigacion/file.php/38/ARCHIVOS_2010/textos/guia_Bibliografia.PDF

Oracle (2013) Data Warehousing Guide. [Versión electrónica] Disponible en: http://docs.oracle.com/cd/E11882_01/server.112/e25555/tdpdw_owb_sources.htm#TDP_DW10500

Pérez, C Y Largo, E (2009) Análisis De La Herramienta SPADIES Diseñada Por El Ministerio De Educación Nacional Y El Cede. Centro De Investigaciones Y Documentación Socioeconómica (Cidse) Universidad Del Valle.

Piedade, M y Santos, M (2010). Business Intelligence in Higher Education: Enhancing the teaching-learning process with a SRM system. In Information Systems and Technologies (CISTI). Pag. 1-5.

Pinzón, Liza (2011). “Aplicando minería de datos al marketing educativo”. Marketing 1. Volumen 1.

Semillero de investigación (2010), *signal*, Universidad de Purdue estado de Indiana (USA).

Rodríguez, C (2007) metodológica para la implementación de un sistema de información tipo inteligencia de negocios que apoya la toma de decisiones, en el área académica, en la Institución Universitaria Antonio José Camacho. Cali, 2007. Trabajo de grado para optar el título de en Dirección Estratégica de Tecnologías de la Información. Fundación Universitaria Iberoamericana.

Ruiz. J Luis, G. M, Pérez (2014). Causas y consecuencias de la deserción escolar en el bachillerato: caso universidad autónoma de sinaloa [Versión electrónica] Disponible en: <http://www.uaimlosmochis.org/ECFD/index.php/2014/2/paper/viewFile/38/58>

Salcedo, A (2010). Deserción universitaria en Colombia [Versión electrónica] Disponible en: http://www.alfaguia.org/alfaguia/files/1319043663_03.pdf

Shores, R (2013) Valdosta State University Improves Student Success and Retention with Oracle's Robust Business Intelligence Environment [Versión electrónica] Disponible en: <http://www.oracle.com/us/corporate/press/1895958>

Sistema para la prevención de la deserción de la educación superior, SPADIES (2012). Momento (semestre) en el cual los estudiantes desertores abandonan sus estudios en el nivel de formación universitario universitario año 2012 [Versión electrónica] SPADIES, 2012 2 (3). Disponible en: <http://www.colombiaaprende.edu.co/html/micrositios/1752/w3-article-325044.html>

Sistemas para la prevención de la deserción de la educación superior (2015). SPADIES [Versión electrónica] Disponible en:

<http://www.mineducacion.gov.co/sistemasdeinformacion/1735/w3-propertyname-2895.html>

Universidad nacional de Colombia, (2006). Caracterización de la deserción estudiantil en la universidad nacional de Colombia sede Medellín [Versión electrónica] Unal medellin , 2006 Disponible en:
<http://www.medellin.unal.edu.co/dirplanea/documentos/EstudioDesercionUnalMed.pdf>

Universidad Nacional de la Plata (2012). “*Plan Estratégico de Ingeniería. En: La universidad de la plata*”. Argentina: La plata.

Valero, S, Salvador, A y García, M (2009) Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. Prolongación Reforma 168. Numero 1.

Anexos**Consentimiento Informado****Anexo A****Instrumento # 1**

Entrevista al personal del programa TAE (Trayectoria Académica Exitosa) y coordinadores de programa de la Corporación Universitaria del Caribe –CECAR

1. ¿Describa el flujo de procesos para el ingreso de estudiantes al programa TAE?
2. ¿Cuáles son las estrategias implantadas para el seguimiento y control de estudiantes que pertenecen al programa TAE?
3. ¿Qué tipo de apoyo (psicológico, académico, etc) reciben los estudiantes que ingresan al programa?
4. ¿Qué mecanismos de comunicación (personal, correo electrónico, mensajería instantánea, llamadas, entre otras) utiliza el programa con los estudiantes?
5. ¿Analice y describa cuales son los puntos fuertes (P.E aceptación, acompañamiento) y débiles (P.E problemas de comunicación o seguimiento estudiantil) del programa?
6. ¿Cuáles cree que son los principales factores de deserción del estudiante Cecarense?
7. El programa cuenta con alguna(s) herramienta(s) software(s) que apoye(n) las actividades de los procesos del programa TAE. En caso afirmativo especifique el nombre, describa la

función que cumple y que actividades apoya la herramienta (seguimiento, control, toma de decisiones, definición de políticas y estrategias del programa TAE, entre otras).

8. En caso que cuenta con una herramienta software responda el siguiente cuestionario

a) ¿El Sistema de Información utilizado provee la información útil?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

b) ¿El Sistema de Información provee los reportes como exactamente usted los requiere?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

c) ¿Obtiene siempre a tiempo la Información que necesita a través del Sistema de Información?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

d) ¿El sistema provee de reportes y consultas de información de forma completa y detallada?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

e) ¿La información que brinda el Sistema se caracteriza por ser consistente y confiable?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

f) ¿Considera usted que la información arrojada por el sistema es presentada de una forma estructurada y organizada?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

g) ¿El sistema es amigable (entendible, vistoso, agradable interfaz gráfica, entre otros)?

Siempre: _____ Con frecuencia: _____ Casi Nunca: _____
A veces: _____ Pocas veces: _____ Nunca: _____

h) ¿El sistema es fácil de usar y le ayuda a responder a sus preguntas o a resolver sus problemas?

Siempre: _____

Con frecuencia: _____

Casi Nunca: _____

A veces: _____

Pocas veces: _____

Nunca: _____

9. ¿Qué actividades desea que una herramienta software le automatice o le apoye?

10. ¿Qué tipo de información desea que el sistema de información le presente?

11. ¿De qué forma le ayudaría al programa una herramienta software que le permita realizar proyecciones de estudiantes en riesgo de desertar?

Entrevista a estudiantes que pertenecen o han pertenecido al programa TAE (Trayectoria Académica Exitosa) de la Corporación Universitaria del Caribe –CECAR

1. ¿Qué semestre cursa actualmente?
2. ¿Hace cuánto ingreso al programa TAE?
3. ¿Pertenece actualmente al programa TAE?
4. Describa como fue el proceso de ingreso al programa
5. ¿Cuáles fueron o son las estrategias implantadas para el seguimiento y control del programa TAE?
6. ¿Qué tipo de apoyo (psicológico, académico, etc) recibió o está recibiendo?
7. ¿Qué mecanismos de comunicación (personal, correo electrónico, mensajería instantánea, llamadas, entre otras) utilizó el programa?
8. ¿Analice y describa cuales son los puntos fuertes (P.E aceptación, acompañamiento) y débiles (P.E problemas de comunicación o seguimiento estudiantil) del programa?
9. Que recomendaciones le surgiría al programa para mejorar
10. Desde su punto de vista como estudiante del programa de ingeniería de sistemas, como las tecnologías de la información y las tecnologías de la información y comunicación pueden ayudar a los estudiantes a mejorar el rendimiento académico y con este lograr disminuir las tasas deserción del programa de ingeniería de sistemas

Anexo B
Instrumento # 2

1. Describa la infraestructura tecnológica con la que cuenta actualmente CECAR, especificando hardware (tipo de servidores (torre, blade, rack), características de los equipos de servidores), software (sistemas operativos, lenguajes de programación, motores de bases de datos, servidor de aplicación, entre otros), red de datos (tipo, protocolos, dispositivos, ancho de banda).
2. ¿Qué procesos apoya el sistema información transaccional implantado en CECAR?
3. ¿Explique la arquitectura de software del sistema de información?
4. ¿Qué módulos del sistema de información transaccional apoyan el área académica?
5. En la primera columna de la tabla # 2 se muestran los datos requeridos para el desarrollo del proyecto de investigación. ¿Mencione las tablas sobre las cuales se almacena información de los datos de la primera columna de la tabla?

Tabla 1. *Datos requeridos para el desarrollo del proyecto*

Datos Requeridos	Tablas del modelo relacional que apoyan
Datos académicos y personales de los docentes (Carga horaria, título máximo, antigüedad, número de horas catedra semanales, fecha de nacimiento, entre otros).	
Datos académica y personales del estudiantes (Entrevista inicial, fecha de nacimiento, estudios de básica primaria y secundaria, datos familiares, entre otros)	
Asignaturas contenidas en el pensum de cada programa	
Estudiantes matriculados	
Asignaturas cursadas por estudiante.	
Docentes de asignaturas.	
Nota obtenida por el estudiante en cada asignatura cursada	

Promedio estudiantil.	
Evaluación, coevaluación y autoevaluación docente.	
Ubicación en la planta física de cada asignatura	

6. El departamento de sistemas tendrá algunas restricciones de seguridad al momento en que se realice la carga de datos desde la base de datos del sistema de información transaccional hasta la bodega de datos que se implementara.

Anexo C
Instrumento # 3

Entrevista a estudiantes que pertenecen o han pertenecido al programa TAE (Trayectoria Académica Exitosa) de la Corporación Universitaria del Caribe –CECAR

1. ¿Qué semestre cursa actualmente?
2. ¿Hace cuánto ingreso al programa TAE?
3. ¿Pertenece actualmente al programa TAE?
4. Describa como fue el proceso de ingreso al programa
5. ¿Cuáles fueron o son las estrategias implantadas para el seguimiento y control del programa TAE?
6. ¿Qué tipo de apoyo (psicológico, académico, etc) recibió o está recibiendo?
7. ¿Qué mecanismos de comunicación (personal, correo electrónico, mensajería instantánea, llamadas, entre otras) utilizó el programa?
8. ¿Analice y describa cuales son los puntos fuertes (P.E aceptación, acompañamiento) y débiles (P.E problemas de comunicación o seguimiento estudiantil) del programa?
9. Que recomendaciones le surgiría al programa para mejorar
10. Desde su punto de vista como estudiante del programa de ingeniería de sistemas, como las tecnologías de la información y las tecnologías de la información y comunicación pueden ayudar a los estudiantes a mejorar el rendimiento académico y con este lograr disminuir las tasas deserción del programa de ingeniería de sistemas

Anexo D

Audio de entrevistas realizadas a coordinares de programas, funcionarios del programa de trayectoria académica exitosa y estudiantes.

Entrevistado	Dirección electrónica
Audio entrevista Funcionaria TAE.	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZlGM8YLReRd1c1YVZuTXFGTEU/view?usp=sharing
Audio entrevista estudiante TAE.	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZlGM8YLReRQ2ktY2hfMGNNZjg/view?usp=sharing
Audio entrevista estudiante TAE.	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZlGM8YLReRdHdGbU52UFBqTVk/view?usp=sharing
Audio entrevista coordinadora TAE.	<ul style="list-style-type: none">https://drive.google.com/a/cecar.edu.co/file/d/0B_ZlGM8YLReRUVBXd0tkZ2xuQ0E/view?usp=sharinghttps://drive.google.com/file/d/0BzZiafAQIO1pTXFHNWxONVJJSkU/view?usp=sharing



Anexo E.

Script del modelo de datos para el datamart académico.



Nombre del Script	Dirección electrónica
Script. SQL	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZIgM8YLReRUIpVSkVJajZSNzg/view?usp=sharing

Anexo F.

Cubo deserciones y rendimientos.



Cubos	Dirección electrónica
CubosDesercion	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZIgM8YLReRd3RpYTFTOG15VTQ/view?usp=sharing
CuboRendimiento	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZIgM8YLReRWmZuQIN1N2FHSDg/view?usp=sharing

Anexo G.

Trabajos ETL generados en WEKA

Weka	Dirección electrónica
ETL generados en WEKA	https://drive.google.com/a/cecar.edu.co/file/d/0B_ZIgM8YLReRMklTY0hrenNHQW8/view?usp=sharing

Anexo H.

Reporte de estudiantes en riesgo de desertar en BIRT

Reporte Birt	Dirección electrónica
Reporte	https://drive.google.com/drive/folders/0B_ZIgM8YLReRV01aU0RfR3Q1YUE?usp=sharing