

# UNIVERSIDAD TÉCNICA DE AMBATO



## FACULTAD DE INGENIERÍA EN SISTEMAS ELECTRÓNICA E INDUSTRIAL

### CENTRO DE ESTUDIOS DE POSGRADO MAESTRÍA EN GESTIÓN DE BASES DE DATOS III VERSIÓN

---

**TEMA:** Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

---

Trabajo de Investigación previo a la obtención del Grado Académico de Magíster  
en Gestión de Bases de Datos

**AUTOR:** Ing. Cristina Alexandra Frutos Ortega

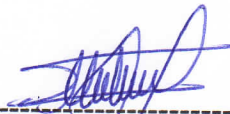
**DIRECTOR:** Ing. Edwin Hernando Buenaño Valencia Mg.

Ambato – Ecuador

2017

A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas,  
Electrónica e Industrial

El Tribunal receptor del Trabajo de Investigación presidido por la Ingeniera Elsa Pilar Urrutia Urrutia Magíster, e integrado por los señores Ingeniero Carlos Israel Núñez Miranda Magister, Ingeniero Franklin Oswaldo Mayorga Mayorga Magister, Ingeniero Jaime Bolívar Ruiz Banda Magister, designados por la Unidad Académica de Titulación de la Universidad Técnica de Ambato, para receptor el Trabajo de Investigación con el tema: “LA TASA DE RETENCIÓN Y SU INCIDENCIA EN LA DETECCIÓN DE PATRONES DE DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DE AMBATO”, elaborado y presentado por la señora Ing. CRISTINA ALEXANDRA FRUTOS ORTEGA, para optar por el Grado Académico de Magister en Gestión de Bases de Datos; una vez escuchada la defensa oral del Trabajo de Investigación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la UTA.



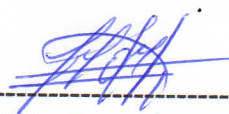
---

Ing. Elsa Pilar Urrutia Urrutia, Mg.  
Presidente del Tribunal



---

Ing. Carlos Israel Núñez Miranda Mg.  
Miembro del Tribunal



---

Ing. Franklin Oswaldo Mayorga Mayorga Mg.  
Miembro del Tribunal

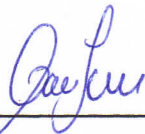


---

Ing. Jaime Bolívar Ruiz Banda Mg.  
Miembro del Tribunal

## AUTORÍA DEL TRABAJO DE INVESTIGACIÓN

La responsabilidad de las opiniones, comentarios y críticas emitidas en el Trabajo de Investigación presentado con el tema: “La Tasa de Retención y su Incidencia en la Detección de Patrones de Deserción Estudiantil en la Universidad Técnica de Ambato”, le corresponde exclusivamente a la: Ingeniera Cristina Alexandra Frutos Ortega, Autor bajo la Dirección del Ingeniero Edwin Hernando Buenaño Valencia Magister, Director del Trabajo de Investigación; y el patrimonio intelectual a la Universidad Técnica de Ambato.

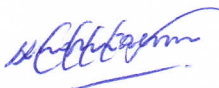


---

Ing. Cristina Alexandra Frutos Ortega

cc. 1804286340

AUTOR



---

Ing. Edwin Hernando Buenaño Valencia, Mg

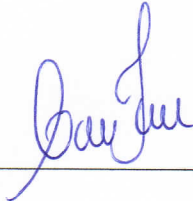
cc. 1802662955

DIRECTOR

## DERECHOS DE AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el Trabajo de Investigación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos de mi trabajo, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad.



---

Ing. Cristina Alexandra Frutos Ortega

c.c. 1804286340

## ÍNDICE GENERAL DE CONTENIDOS

Portada .....	I
A la Unidad Académica de Titulación .....	II
Autoría del Trabajo de Investigación.....	III
Derechos de Autor.....	IV
Índice General de Contenidos .....	V
Índice de Figuras .....	X
Índice de Tablas .....	XIV
Agradecimiento .....	XVI
Dedicatoria .....	XVII
Resumen Ejecutivo.....	XVIII
Executive Summary .....	XX
Introducción .....	1
CAPÍTULO I.....	3
EL PROBLEMA DE INVESTIGACIÓN.....	3
1.1. Tema de Investigación .....	3
1.2. Planteamiento del Problema.....	3

1.2.1. Contextualización .....	3
1.2.2. Análisis crítico.....	4
1.2.3. Prognosis .....	5
1.2.4. Formulación del problema.....	5
1.2.5. Preguntas Directrices.....	5
1.2.6. Delimitación del Objeto de investigación .....	5
1.2. Justificación .....	6
CAPÍTULO II .....	7
MARCO TEÓRICO.....	7
2.1. Antecedentes Investigativos.....	7
2.2. Fundamentación Filosófica .....	8
2.3. Fundamentación Legal .....	8
2.4. Categorías Fundamentales .....	11
2.4.1. Categorías fundamentales de la Variable Independiente.....	14
Tasa de Retención Estudiantil .....	14
Evaluación de la Calidad de las Carreras.....	17
Evaluación de Carreras .....	19
2.4.2. Categorías fundamentales de la Variable Dependiente .....	21
Descubrimiento de Conocimiento en Bases de Datos .....	21
Minería de Datos.....	23
Métodos de Minería de Datos .....	27
Almacén de Datos.....	29
Deserción Estudiantil.....	34
2.5. Hipótesis.....	36
2.6. Señalamiento de Variables .....	36

CAPÍTULO III.....	37
METODOLOGÍA .....	37
3.1. Enfoque .....	37
3.2. Modalidad Básica de Investigación .....	37
3.3. Nivel o tipo de investigación .....	37
3.4. Población y Muestra.....	38
3.5. OPERACIONALIZACIÓN DE LAS VARIABLES.....	39
3.6. Recolección de Información .....	41
3.7. Procesamiento y Análisis de la Información.....	41
CAPÍTULO IV.....	43
ANÁLISIS E INTERPRETACIÓN DE RESULTADOS.....	43
4.1. Análisis e Interpretación de Resultados .....	43
4.2. Verificación de Hipótesis.....	52
4.2.1. Planteamiento de la Hipótesis .....	52
4.2.2. Frecuencias Observadas .....	55
4.2.3. Frecuencias Esperadas.....	55
4.2.4. Prueba Chi – Cuadrado.....	56
4.2.5. Grados de Libertad .....	56
4.2.6. Decisión estadística .....	57
CAPÍTULO V .....	58
CONCLUSIONES Y RECOMENDACIONES.....	58
5.1 Conclusiones .....	58

5.2. Recomendaciones.....	59
CAPÍTULO VI.....	60
LA PROPUESTA .....	60
6.1. Datos Informativos .....	60
6.1.1. Título .....	60
6.1.2. Institución Ejecutora.....	60
6.1.3. Beneficiarios .....	60
6.1.4. Ubicación.....	60
6.1.5. Equipo Técnico Responsable .....	60
6.2. Antecedentes de la Propuesta.....	61
6.3. Justificación .....	61
6.4. Objetivos .....	61
Objetivo General .....	61
Objetivos Específicos .....	61
6.5. Análisis de Factibilidad.....	62
6.5.1. Factibilidad Técnica .....	62
6.5.2. Factibilidad Organizacional.....	62
6.5.3. Factibilidad Económica .....	62
6.5.4. Factibilidad Operativa .....	62
6.6 Fundamentación .....	62
6.6.1. Análisis Tasa de Retención .....	63
6.6.2 KDD (Knowledge Discovery in Databases).....	69
6.6.3. CRISP-DM .....	72
6.6.4. SEMMA .....	73
6.6.5. Herramientas.....	74
SQL- Server .....	74
Integration Services .....	75



WEKA .....	76
6.6.6. Técnicas de Minería de Datos .....	76
REDES NEURONALES .....	77
ÁRBOLES DE DECISIÓN .....	77
AGRUPAMIENTO O CLUSTERING.....	77
6.7. Metodología .....	78
6.7.1. Metodología para procesos de Minería de Datos .....	78
6.7.2 Selección de Metodología .....	78
6.7.3. Aplicación de Metodología .....	78
6.7.3.1. Selección de Datos.....	79
6.7.3.2. PREPROSESAMIENTO .....	82
6.7.3.3. Transformación.....	97
6.7.3.4. Minería de Datos.....	99
6.7.3.5. Etapa de Interpretación/ Evaluación de datos.....	110
BIBLIOGRAFÍA .....	124
ANEXOS .....	127

## ÍNDICE DE FIGURAS

Figura 1. Árbol de Problemas .....	4
Figura 2. Categorías Fundamentales .....	11
Figura 3. Constelación de Ideas Variable Independiente.....	12
Figura 4. Constelación de Ideas Variable Independiente.....	13
Figura 5. Estructura de árbol del modelo genérico para la evaluación del entorno de aprendizaje de la carreras .....	16
Figura 6. Etapas del Proceso DCBD .....	22
Figura 7. Data WareHouse .....	30
Figura 8. Carga de almacén de datos .....	32
Figura 9. Arquitectura de Data WareHouse .....	32
Figura 10. Proceso ETL .....	33
Figura 11. Cálculo de la Tasa de Retención Estudiantil .....	44
Figura 12. Análisis de la Tasa de Retención Estudiantil.....	45
Figura 13. Estrategias para maximizar el índice de Retención Estudiantil.....	46
Figura 14. Importancia para la institución, la Retención Estudiantil.....	47
Figura 15. Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil.....	48
Figura 16. Importancia para la institución identificar los Patrones de Deserción Estudiantil .....	49

Figura 17. Reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil .....	50
Figura 18. La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato .....	51
Figura 19. Distribución Chi – Cuadrado .....	56
Figura 20. Prueba Chi- Cuadrado .....	57
Figura 21. Sistema Integrado de Reportes .....	63
Figura 22. Tasa de Retención Selección de Datos .....	64
Figura 23. Estudiantes del Período Inicio de Estudios.....	64
Figura 24. Estudiantes Retenidos .....	65
Figura 25. Formula Tasa de Retención .....	66
Figura 26. Tasa de Retención Sistemas período Abril – Septiembre 2016.....	66
Figura 27. Tasa de Retención Sistemas período Octubre 2016 - Marzo 2017.....	67
Figura 28. Tasa de Retención Electrónica período Abril - Septiembre 2016 .....	67
Figura 29. Tasa de Retención Electrónica período Octubre 2016 - Marzo 2017..	68
Figura 30. Tasa de Retención Industrial período Octubre 2016 - Marzo 2017.....	68
Figura 31. Tasa de Retención Industrial período Octubre 2016 - Marzo 2017.....	69
Figura 32. Pasos que componen el proceso KDD.....	71
<b>Figura 33.</b> Esquema de los 4 niveles de CRISP-DM .....	72
Figura 34: Modelo de proceso CRISP–DM ([CRISP-DM, 2000]).....	73
Figura 35. SEMMA.....	74

Figura 36. SQL Server 2008 .....	75
Figura 37. Técnicas de Minería de Datos.....	76
Figura 39. Creación de la Base de Datos .....	82
Figura 40. Asignación del Nombre al Data Warehouse.....	83
Figura 41. Creación de nuevo Proyecto .....	84
Figura 42. Creación de Proyecto de Integration Services .....	84
Figura 43. Nuevo Administrador de conexiones.....	85
Figura 44. Conexión con la Base de Datos Transaccional.....	85
Figura 45. Tarea de Flujo de Datos .....	86
Figura 46. Asignación del Nombre a la Tarea de Flujo de Datos .....	86
Figura 47. Origen de Datos .....	87
Figura 48. Configuración de la conexión para obtener los datos.....	87
Figura 49. Propiedades del flujo de datos permite obtener datos del proveedor. .	88
Figura 50. Columna Derivada.....	89
Figura 51. Variación Lenta .....	90
Figura 52. ETL Colegios.....	91
Figura 53. ETL Carrera.....	92
Figura 54. ETL Materias .....	93
Figura 55. ETL Profesor .....	94
Figura 56. ETL Periodos .....	95

Figura 57. ETL Desertores .....	96
Figura 58. Proyecto ETL.....	97
Figura 59. Diagrama Desertor .....	98
Figura 60. Creación del Archico .CSV .....	98
Figura 61. Reemplazo Carácter “;” .....	99
Figura 62. Weka .....	99
Figura 63. Carga de archivo .csv.....	100
Figura 64. Selección Algoritmo Kmedia .....	101
Figura 65. Instancias agrupadas .....	101
Figura 66. Resultados Clouster .....	110
Figura 67. Patrón de Deserción Estudiantil N°1 .....	111
Figura 68. Patrón de Deserción Estudiantil N°2 .....	112
Figura 69. Patrón de Deserción Estudiantil N°3 .....	112
Figura 70. Incidencia de Desercion por Facultad.....	113
Figura 71. Incidencia de Desercion por Carrera .....	114
Figura 72. Incidencia de Desercion por Género.....	117
Figura 73. Incidencia de Desercion por Etnia.....	118
Figura 74. Incidencia de Desercion por Discapacidad.....	119
Figura 75. Incidencia de Desercion por Provincia .....	120
Figura 76. Incidencia de Desercion por Período.....	121

## ÍNDICE DE TABLAS

Tabla 1: Población y muestra.....	38
Tabla 2: Operacionalización de las variables (Variable Independiente).....	39
Tabla 3. Operacionalización de las variables (Variable Independiente).....	40
Tabla 4. Recolección de información.....	41
Tabla 5. Cálculo de la Tasa de Retención Estudiantil.....	44
Tabla 6. Análisis de la Tasa de Retención Estudiantil.....	45
Tabla 7. Estrategias para maximizar el índice de Retención Estudiantil.....	46
Tabla 8. Importancia para la institución, la Retención Estudiantil.....	47
Tabla 9. Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil.....	48
Tabla 10. Importancia para la institución identificar los Patrones de Deserción Estudiantil.....	49
Tabla 11. Reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil.....	50
Tabla 12. La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato.....	51
Tabla 13. Evaluación Respuestas Pregunta 2.....	54
Tabla 14. Evaluación Pregunta 7.....	54
Tabla 15. Frecuencias Observadas.....	55
Tabla 16. Frecuencias Observadas.....	55

Tabla 17. Prueba Chi – Cuadrado .....	56
Tabla 18. Metodología para procesos de Minería de Datos.....	78
Tabla 19. Vista Estudiantes.....	79
Tabla 20. Vista Colegios .....	79
Tabla 21. Vista Especialidad.....	80
Tabla 22. Vista Materias .....	80
Tabla 23. Vista Profesor.....	80
Tabla 24. Vista Período.....	80
Tabla 25. Vista Deserción .....	81
Tabla 26. Full Data.....	103
Tabla 27. Cluster 0 .....	104
Tabla 28. Cluster 1 .....	106
Tabla 29. Cluster 2 .....	107
Tabla 30. Cluster 3 .....	108
Tabla 31. Cluster 4 .....	109

## **AGRADECIMIENTO**

Agradezco a Dios estar siempre junto a mí, por brindarme salud y fuerzas para poder seguir adelante.

A mi familia, que en todo este tiempo, me brindo su comprensión y apoyo total para culminar esta etapa de mi carrera profesional.

A mi tutor por su apoyo en el trabajo realizado.

A la Universidad Técnica de Ambato por darme la oportunidad de realización del trabajo de investigación.

*Cristina Alexandra Frutos Ortega*



## DEDICATORIA

Este trabajo lo dedico especialmente a Dios, quien ha bendecido mi vida, en cada momento me acompaña y me da fuerza para seguir adelante.

A mis madrecitas por brindarme su amor, confianza y apoyo incondicional.

A mi esposo Marcelo por ser un pilar fundamental en mi vida.

A mis hermosas hijas Danna y Victoria por llenar mi vida de felicidad.

*Cristina Alexandra Frutos Ortega*

**UNIVERSIDAD TÉCNICA DE AMBATO**  
**FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E**  
**INDUSTRIAL**  
**DIRECCIÓN DE POSGRADO MAESTRÍA EN GESTIÓN DE BASES DE**  
**DATOS**

**TEMA**

“Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato”.

**AUTOR:** Cristina Alexandra Frutos Ortega

**DIRECTOR:** Ing. Edwin Hernando Buenaño, Mg.

**FECHA:** 07 de junio del 2017

**RESUMEN EJECUTIVO**

La deserción estudiantil es un problema social que no es ajeno a ninguna institución de educación superior a nivel nacional, en el Ecuador las universidades e institutos tecnológicos se encuentran en proceso de evaluación y acreditación realizado por el Consejo de Evaluación, Acreditación y Aseguramiento (CEACCESS), siendo de relevancia el indicador tasa de retención, de aquí la importancia de generar estrategias que permitan maximizar el índice de estudiantes que culminen su carrera.

La investigación sobre el Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil se lo realizó en la Universidad Técnica de Ambato, el compromiso de la investigación es detectar los Patrones de Deserción Estudiantil en cada una de las carreras de la Universidad.

Mediante esta investigación se plantea una solución al problema con el uso de técnicas y algoritmos de Minería de Datos para la búsqueda y descubrimiento de patrones de interés, aplicando la metodología KDD que nos muestra el proceso de Selección, Preprocesamiento, Transformación, Minería de Datos y la Interpretación de resultados.

Para el descubrimiento de los patrones de Deserción se aplica el algoritmo K-medias, basado en agrupamiento numérico, siendo un algoritmo clasificado como método de particionado y recolocación, además es el método más utilizado en aplicaciones científicas e industriales. Este algoritmo nos permitirá establecer e identificar patrones tanto sociales como académicos que aquejen directamente a la Deserción Estudiantil.

EL propósito de esta investigación tiene como objetivo de brindar esta información a las Autoridades de la Institución para que se puedan establecer estrategias administrativas con el fin de mitigar de alguna manera los factores que afectan directamente a la deserción estudiantil.

**Descripciones:** Deserción Estudiantil, algoritmos, minería de datos, metodología KDD, Selección, Preprocesamiento, Transformación, Interpretación, K-Medias, particionada, recolocación.

**UNIVERSIDAD TÉCNICA DE AMBATO**  
**FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E**  
**INDUSTRIAL**  
**DIRECCIÓN DE POSGRADO MAESTRÍA EN GESTIÓN DE BASES DE**  
**DATOS**

**THEME**

“Analysis of the retention rate and its incidence in the detection of patterns of student dropout at the Technical University of Ambato”.

**AUTHOR:** Cristina Alexandra Frutos Ortega

**DIRECTED BY:** Ing. Edwin Hernando Buenaño, Mg.

**DATE:** June 07, 2017

**EXECUTIVE SUMMARY**

The student desertion is a social problem that is not alien to any Higher Education Institution at the national level, in Ecuador the universities and technological institutes are in the process of evaluation and accreditation carried out by the Evaluation, Accreditation and Assurance Council (CEACCESS), With relevance being the Retention Rate Indicator, hence the importance of generating strategies to maximize the rate of students who complete their career.

The research on the Retention Rate Analysis and its incidence in the detection of student desertion patterns was carried out at the Technical University of Ambato, the research commitment is to detect the Desertion Patterns in students of each of their careers.

This research proposes a solution to the problem with the use of techniques and algorithms of Data Mining for the search and discovery of patterns of interest, applying the methodology KDD that shows us the process of Selection, Preprocessing, Data Mining Transformation and Interpretation of results.

For the discovery of Desertion patterns, the K-media algorithm, based on numerical grouping, is applied, being an algorithm classified as a method of partitioning and relocation, in addition it is the most used method in scientific and industrial applications. This algorithm will allow us to establish and identify both social and academic patterns that directly affect Student Dropout.

The purpose of this is to provide this information to the Authorities of the Institution so that administrative strategies can be established in order to mitigate in some way the factors that directly affect student dropout.

**Descriptors:** Student Desertion, Algorithms, Data Mining, KDD Methodology, Selection, Preprocessing, Transformation, Interpretation, K-Stockings, Partitioned, Replacement.

## INTRODUCCIÓN

El presente trabajo de investigación tiene como tema: Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

Su importancia radica en la necesidad de aplicar técnicas de minería de datos que permitan Detectar Patrones de Deserción Estudiantil en la Universidad Técnica de Ambato, para poder establecer estrategias que permitan maximizar la Tasa de Retención Estudiantil.

La presente investigación está organizada en los siguientes capítulos:

**Capítulo I**, denominado como EL PROBLEMA, describe el problema que es objeto de la investigación, este contiene: el tema de investigación, el planteamiento del problema, su contexto, análisis crítico, prognosis, delimitación, justificación y objetivos.

**Capítulo II**, denominado como MARCO TEÓRICO, en este se detalla: antecedentes de la investigación, fundamentación filosófica, fundamentación legal, categorías fundamentales, formulación de hipótesis y señalamiento de variables.

**Capítulo III**, denominado como METODOLOGÍA, incluye: el enfoque investigativo, modalidad básica de la investigación, nivel o tipo de investigación, población y muestra, operacionalización de variables, técnicas e instrumentos, plan de recolección de información y plan de procesamiento de la información. El

**Capítulo IV**, denominado como ANÁLISIS E INTERPRETACIÓN DE RESULTADOS en este se detalla el análisis e interpretación de la información recolectada a través de los diferentes métodos utilizados en la presente investigación.

**El Capítulo V**, denominado como CONCLUSIONES Y RECOMENDACIONES, expone de forma clara y concisa las conclusiones y recomendaciones de la investigación realizada.

**El Capítulo VI**, denominado como LA PROPUESTA, contiene información relacionada a los antecedentes de la Institución, justificación, objetivos, factibilidad, proceso de selección de metodologías, plataformas y arquitectura, herramientas para la implementación, desarrollo y puesta en producción de la propuesta realizada a la comunicación en línea del proceso de recuperación de cartera de créditos en campo, la misma que ayudará de manera positiva en la Detección de Patrones de Deserción Estudiantil.

## **CAPÍTULO I**

### **EL PROBLEMA DE INVESTIGACIÓN**

#### **1.1. Tema de Investigación**

Análisis de la tasa de retención y su incidencia en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

#### **1.2. Planteamiento del Problema**

##### **1.2.1. Contextualización**

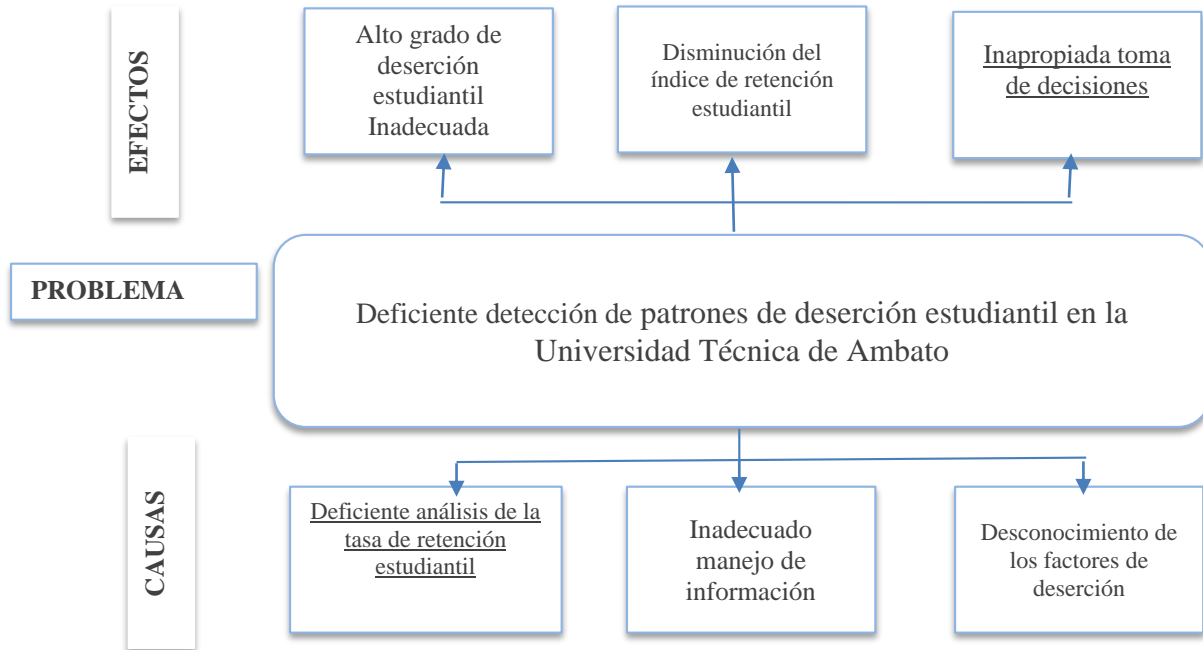
En Latinoamérica, la educación superior presenta altas tasas de deserción estudiantil, especialmente en los primeros semestres académicos, hecho que conlleva a efectos de tipo financiero, académico y social tanto para las Instituciones de Educación Superior (IES) como para el estudiante, la región, el país y el Estado (*Javier Jiménez Toledo, 2013*).

La Deserción Universitaria es un fenómeno que afecta a todas las instituciones educativas (*Darío Alejandro Viteri Castro, Mario Andrés Uquillas Narváez, Quito 2011*), En nuestro país el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior (CEAACES) es el organismo técnico, público y autónomo encargado de ejercer la rectoría política para la evaluación, acreditación y el aseguramiento de la calidad de las Instituciones de Educación Superior, sus programas y carreras, esto realizando procesos continuos de evaluación y acreditación que evidencien el cumplimiento de las misiones, fines y objetivos de las mismas.



Por esta razón la Universidad Técnica de Ambato requiere realizar un estudio para determinar las causas que influyen en el retiro de estudiantes universitarios, y poder implementar un modelo de gestión apropiado de detectar los factores de deserción con el fin de disminuir el índice de deserción estudiantil e incrementar la tasa de retención que es un indicador de evaluación y acreditación.

En la figura 1.1 se sintetiza el árbol de problemas.



**Figura 1.** Árbol de Problemas

**Elaborado por:** Investigador

### 1.2.2. Análisis crítico

En la Universidad Técnica de Ambato se ha detectado que el deficiente análisis de la tasa de retención estudiantil ha ocasionado como principal inconveniente, la inadecuada detección de patrones de deserción estudiantil. Otra de las causas importantes es el inadecuado manejo de la información ya que con el uso de herramientas apropiadas se podría descubrir conocimiento de gran relevancia respecto a lo factores de deserción estudiantil.

Uno de los mayores motivos es el desconocimiento de los factores de deserción provocando una inapropiada toma de decisiones por parte de las autoridades.

### **1.2.3. Prognosis**

En caso de que la inadecuada detección de patrones de deserción estudiantil de la Universidad Técnica de Ambato persiste, los índices de deserción estudiantil podrían maximizarse.

### **1.2.4. Formulación del problema**

¿Incide el análisis de la tasa de retención en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato?

### **1.2.5. Preguntas Directrices**

- ¿Existe un análisis de la tasa de retención estudiantil en la Universidad Técnica de Ambato?
- ¿Se detectan los patrones de deserción estudiantil en la Universidad Técnica de Ambato?
- ¿Se puede proponer una solución factible al problema planteado?

### **1.2.6. Delimitación del Objeto de investigación**

- Delimitación del contenido
  - ✓ Campo: Minería de datos
  - ✓ Área: Base de datos
  - ✓ Aspecto: Detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato
- Delimitación espacial
  - ✓ Universidad Técnica de Ambato
- Delimitación temporal
  - ✓ Períodos académicos comprendidos desde el año 2008 al 2016.

## **1.2. Justificación**

Este trabajo es importante ya que la deserción estudiantil es un problema social que no es ajeno a ninguna Institución de Educación Superior a nivel nacional. La Universidad Técnica de Ambato se encuentran en proceso de Evaluación y Acreditación de cada una de sus carreras, siendo de vital importancia el indicador *Tasa de Retención*, por lo que se requiere establecer procesos que permita la detección de patrones de deserción estudiantil, para que de esta manera se puedan establecer estrategias que permitan maximizar el índice de Retención estudiantil en nuestra institución y a su vez minimizar la Deserción Estudiantil.

Técnicamente es factible porque la Universidad Técnica de Ambato, posee el hardware y software requerido, además la investigación cuenta con las facilidades de la información histórica del record académico y social de los estudiantes.

Operativamente es factible la investigación porque se basa en la experiencia y conocimientos del investigador y del personal involucrado en los procesos.

Económicamente es factible porque el investigador cubrirá todos los gastos relacionados con la investigación.

## **1.4. Objetivos**

### **1.4.1. Objetivo General**

Determinar la incidencia del análisis de la tasa de retención en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

### **1.4.2. Objetivo Específico**

- Analizar la tasa de retención estudiantil en la Universidad Técnica de Ambato.
- Detectar los patrones de deserción de la Universidad Técnica de Ambato.
- Proponer una solución factible al problema en caso de existir.

## **CAPÍTULO II**

### **MARCO TEÓRICO**

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

Una vez revisados varios Repositorios de las diferentes Universidades que ofertan la Carrera de Ingeniería en Sistemas, Electrónica e Industrial se pudo encontrar la siguiente información que servirá como un antecedente investigativo para la presente investigación:

- En la Pontificia Universidad Católica del Ecuador, Facultad de Psicología, se realizó una investigación denominada: Estudio sobre la deserción estudiantil en la Pontificia Universidad Católica del Ecuador - Matriz, en los niveles 1ro, 2do y 3ero de todas las Facultades y Escuelas del primer semestre del año académico 2007-2008, investigada por Darío Alejandro Viteri Castro y Mario Andrés Uquillas Narváez en el año 2011, uno de los objetivos de esta investigación fue “Determinar las causas que han generado el abandono de los estudios universitarios en la PUCE-Q estableciendo cuáles son las más incidentes”, obteniendo como conclusión: Luego de culminado el estudio podemos afirmar con un grado de significación del 0,05, con un nivel de confiabilidad del 95% y en relación a los estudiantes encuestados que de cada 10 estudiantes que desertan de la PUCE - Matriz, siete lo hacen por motivos académicos personales o económicos.
  
- En la Universidad Politécnica Salesiana Sede Guayaquil, Facultad de Ciencias Administrativas y Económicas, carrera: Contabilidad y Auditoría., se realizó una investigación denominada: “Análisis sobre la Deserción Estudiantil en la Universidad Politécnica Salesiana, sede Guayaquil: Caso de las carreras de Administración de

Empresas y Contabilidad y Auditoría periodo de aplicación 2007 – 2012”, investigada por Mabel Arlene Narváez Barros y Geraldine Estefanía Barragán Reyes en el año 2015, uno de los objetivos de esta investigación fue “Determinar las variables que afectan la deserción estudiantil en las carreras de Administración de Empresas y Contabilidad y Auditoría de la Universidad Politécnica Salesiana sede Guayaquil, mediante un análisis cuantitativo y cualitativo para proponer correctivos”, obteniendo como conclusión: El problema sobre Deserción Estudiantil no obedece a un solo factor, dentro de nuestro análisis manifestamos que las causas de la deserción estudiantil en las carreras de Administración de Empresas y Contabilidad y Auditoría tienen mayor impacto en los problemas económicos con un porcentaje del 50% para Contabilidad y Auditoría y más del 60 % para la carrera de Administración de Empresas, teniendo como consecuencias problemas para poder escalar laboralmente en puestos de jefaturas, pierden sus aspiraciones e interés por continuar estudiando, es evidente que este factor es relevante ya que en épocas de exámenes notamos largas filas para efectuar convenios de pagos, la situación económica familiar es muy baja.

## **2.2. Fundamentación Filosófica**

La presente investigación se enmarca en el paradigma crítico propositivo, es crítico porque realiza un Análisis crítico del problema, y es propositivo porque busca proponer una solución factible al problema.

## **2.3. Fundamentación Legal**

Como base legal para la siguiente investigación se menciona los siguientes artículos:

Reglamento de Régimen Académico del Sistema Nacional de Educación Superior, 22 de enero de 2009.

- **Normativa Sobre la Autoevaluación de acuerdo a la Ley Orgánica de Educación Superior**

Art. 95.- Acreditación.- La Acreditación es una validación de vigencia quinquenal realizada por el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior, para certificar la calidad de las instituciones de educación superior, de una carrera o programa educativo, sobre la base de una evaluación previa.

La Acreditación es el producto de una evaluación rigurosa sobre el cumplimiento de lineamientos, estándares y criterios de calidad de nivel internacional, a las carreras, programas, postgrados e instituciones, obligatoria e independiente, que definirá el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior. El procedimiento incluye una autoevaluación de la propia institución, así como una evaluación externa realizada por un equipo de pares expertos, quienes a su vez deben ser acreditados periódicamente.

Art. 98.- Planificación y ejecución de la autoevaluación.- La planificación y ejecución de la autoevaluación estará a cargo de cada una de las instituciones de educación superior, en coordinación con el Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior. [...]

Art. 99.- La autoevaluación.- La Autoevaluación es el riguroso proceso de análisis que una institución realiza sobre la totalidad de sus actividades institucionales o de una carrera, programa o posgrado específico, con amplia participación de sus integrantes, a través de un análisis crítico y un diálogo reflexivo, a fin de superar los obstáculos existentes y considerar los logros alcanzados, para mejorar la eficiencia institucional y mejorar la calidad académica.

- **En La Constitución de la República del Ecuador**

Art. 16.- Todas las personas, en forma individual o colectiva, tiene derecho a: 2. El acceso universal a las tecnologías de información y comunicación.

Art. 18.- Todas las personas en forma individual o colectiva tienen derecho a:

1. Buscar, recibir, intercambiar, producir y difundir información veraz, verificada, oportuna, contextualizada, plural, sin censura previa acerca de los hechos, acontecimientos y procesos de interés general, y con responsabilidad ulterior.

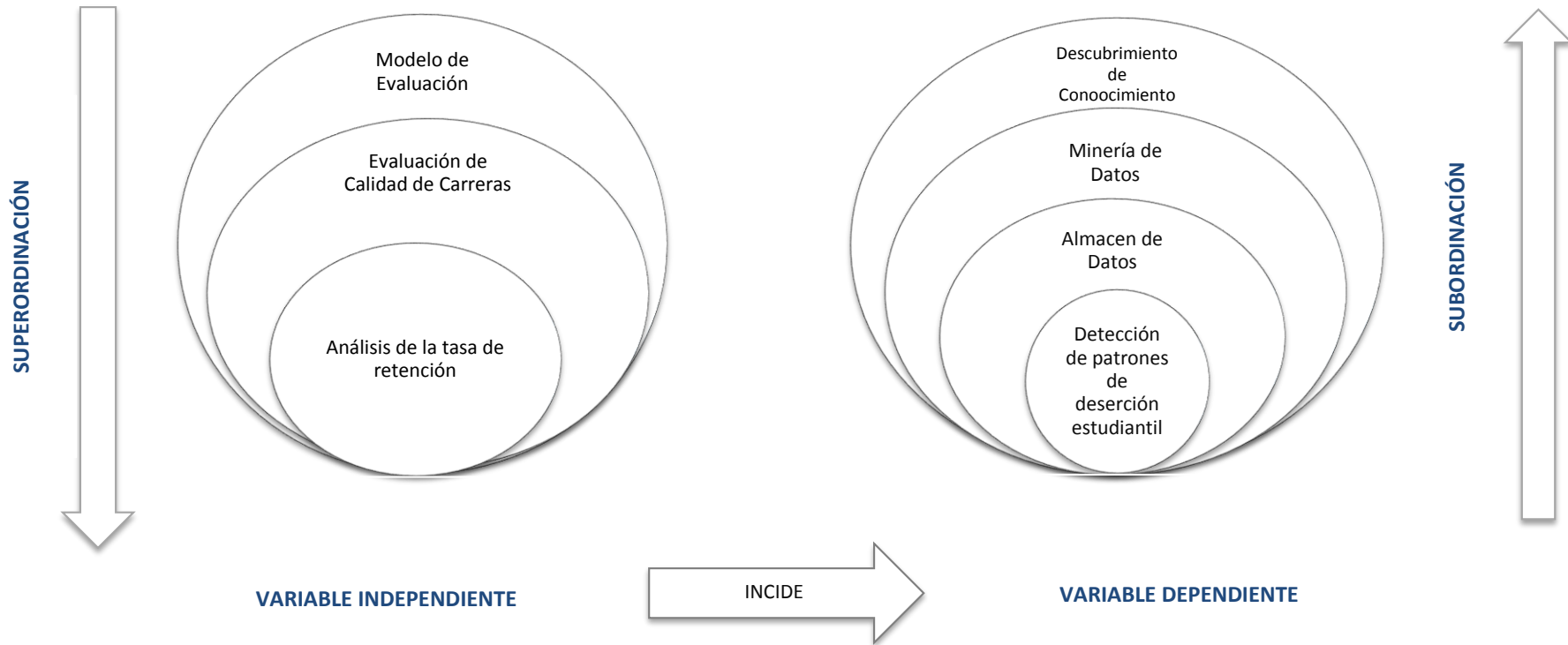
2. Acceder libremente a la información generada en entidades públicas, o en las privadas que manejen fondos del Estado o realicen funciones públicas. No existirá reserva de información excepto en los casos expresamente establecidos en la ley. En caso de violación a los derechos humanos, ninguna entidad pública negará la información

▪ **En la Ley Orgánica de Educación Superior (LOES)**

Art 12 inciso b) manifiesta: “Promover la creación, desarrollo, transmisión y difusión de la ciencia, la técnica, la tecnología y la cultura”.

Art. 32.- “Asignación de recursos para investigación, ciencia y tecnología e innovación.- Las instituciones del Sistema de Educación Superior podrán acceder adicional y preferentemente a los recursos públicos concursables de la pre asignación para investigación, ciencia, tecnología e innovación establecida en la Ley correspondiente”.

## 2.4. Categorías Fundamentales

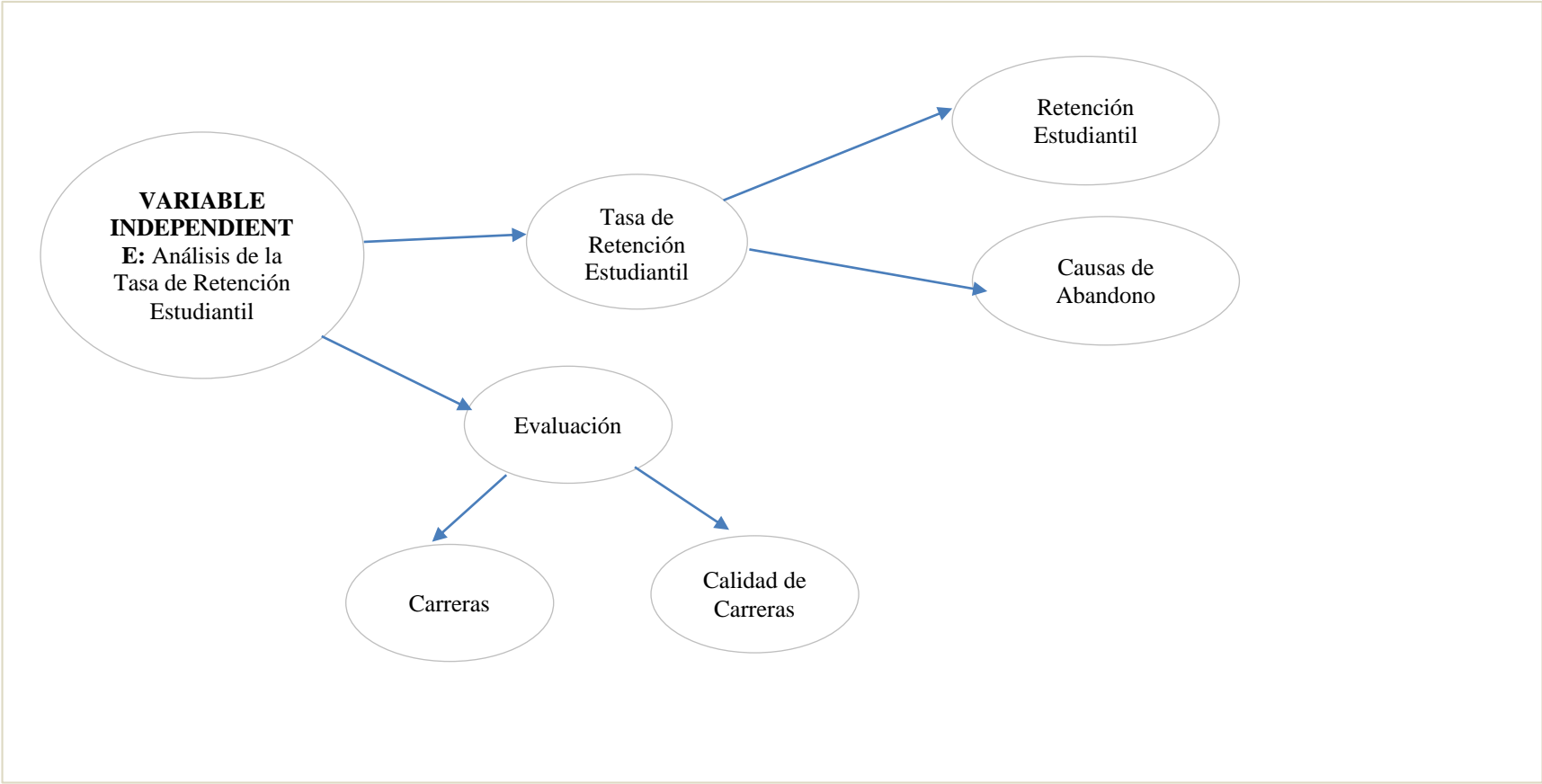


**Figura 2.** Categorías Fundamentales

**Elaborado por:** Investigador



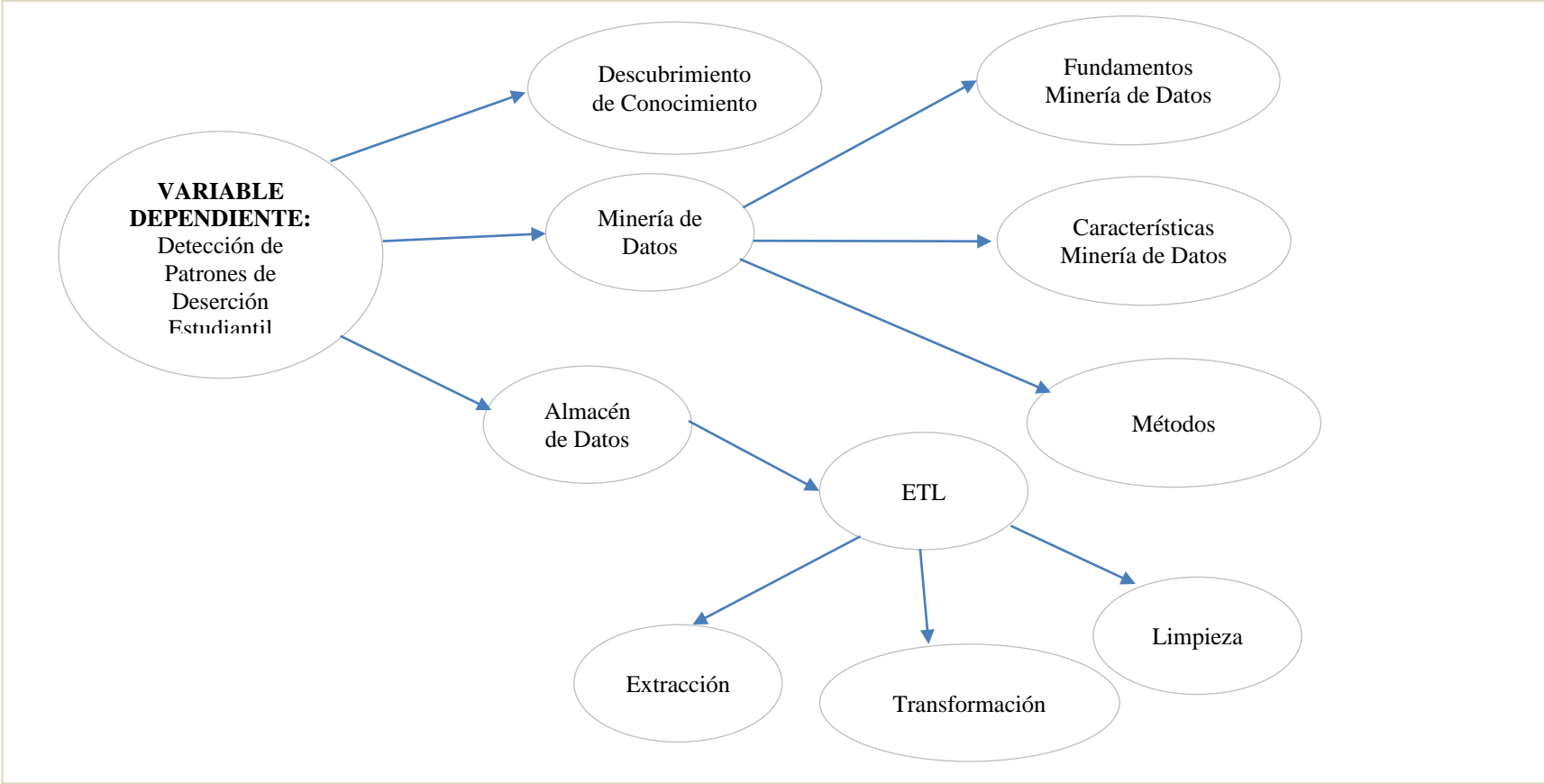
**Constelación de Ideas Variable Independiente**



**Figura 3.** Constelación de Ideas Variable Independiente

**Elaborado por:** Investigador

**Constelación de Ideas Variable Dependiente**



**Figura 4.** Constelación de Ideas Variable Independiente

**Elaborado por:** Investigador

### **2.4.1. Categorías fundamentales de la Variable Independiente**

#### **▪ Tasa de Retención Estudiantil**

La retención estudiantil ha cobrado relevancia en el campo educativo, debido a la sentida necesidad de generar alternativas para el fomento de la permanencia y graduación de los estudiantes del sistema de educación superior.

El término retención se deriva del latín retentio, -onis, que significa acción y efecto de retener. Un programa de retención estudiantil comprende las acciones que desarrolla el aparato educativo en una institución para garantizar el acompañamiento al estudiante durante su trayectoria académica, de tal manera que pueda culminarla de forma exitosa. Su objetivo es proporcionar las herramientas necesarias para la terminación de los diferentes ciclos y etapas en los tiempos establecidos, y adicionalmente asegurar el conocimiento necesario y el desarrollo de competencias y actitudes indispensables para desenvolverse en la vida (OEA, Agencia Interamericana para la Cooperación y el Desarrollo [AICD], 2006). Si bien es cierto que el término se emplea en la literatura a nivel mundial, en la actualidad las instituciones tienden a denominar estas acciones como “programas de fomento de la permanencia y el éxito académico”. Con este nombre se pretende exaltar la idea de potenciar al estudiante versus retenerlo, lo cual implica un carácter de menor voluntad y participación del alumno. (Pineda-Báez & Pedraza-Ortiz & Moreno, 2011)

### **PROCESO DE EVALUACIÓN INSTITUCIONAL**

La evaluación institucional de universidades y escuelas politécnicas tiene como objetivo determinar el grado de cumplimiento de los estándares de calidad definidos en el modelo de evaluación vigente.

El CEAACES ejecuta procesos de evaluación quinquenal con fines de acreditación a todas las universidades y escuelas politécnicas del Sistema de Educación Superior ecuatoriano, para determinar la categorización de las instituciones según lo determina la Ley Orgánica de Educación Superior – LOES –.

Hasta el momento se han definido cuatro categorías A, B, C y D. Las tres primeras corresponden a instituciones que han aprobado satisfactoriamente la evaluación y han obtenido la acreditación con vigencia quinquenal. Las instituciones en categoría D se encuentran en proceso de acreditación, estas podrán acreditarse en una nueva evaluación que realice el CEAACES; para ese propósito existen plazos definidos en el reglamento respectivo. En caso de que estas instituciones no aprueben la evaluación, pasarán al grupo de las instituciones no acreditadas.

La acreditación es obligatoria y necesaria para que una institución pertenezca al Sistema de Educación Superior ecuatoriano.

### **MODELO DE EVALUACIÓN**

El modelo de evaluación contiene los criterios y estándares que se exigen para asegurar niveles mínimos de calidad y su objetivo es generar una cultura de excelencia. El modelo es establecido luego de jornadas de discusión mantenidas con las instituciones a ser evaluadas. Además, se definen otros instrumentos técnicos que apoyan a la ejecución de los procesos: manuales, guías, sistemas informáticos, etc.

### **PROCESO DE EVALUACIÓN**

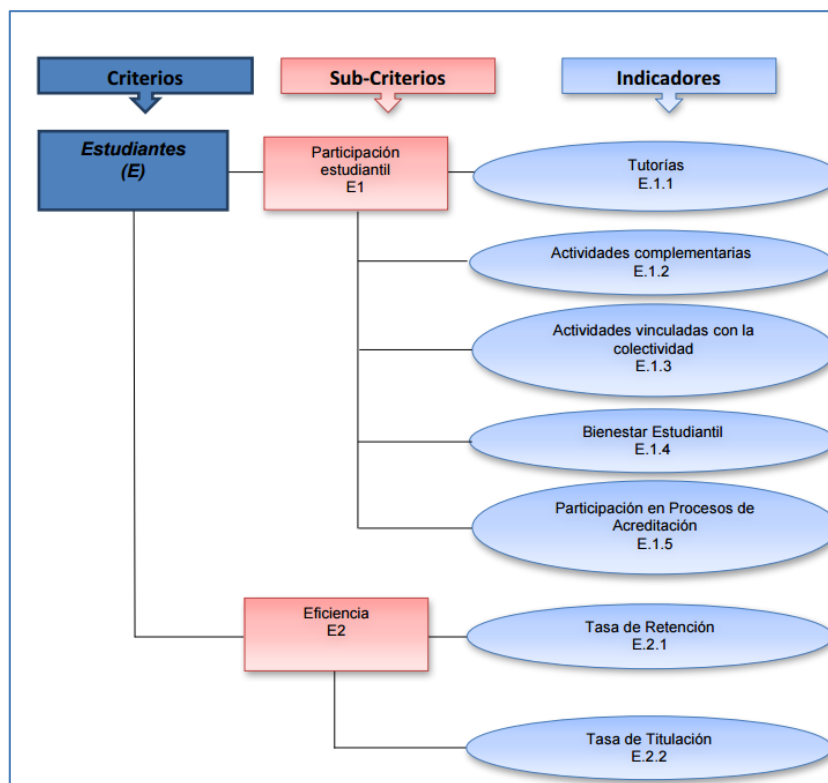
El proceso de evaluación está conformado de las siguientes etapas:

- ✓ Autoevaluación,
- ✓ Carga de información,
- ✓ Evaluación documental,
- ✓ Visita in situ,
- ✓ Rectificaciones,
- ✓ Apelaciones e
- ✓ Informe final.

(<http://www.ceaaces.gob.ec/sitio/proceso-de-recategorizacion-institucional-2/>)

## La Tasa de Retención según criterio de Evaluación y Acreditación

En la figura 5 se muestra la estructura de árbol del modelo genérico para la evaluación del entorno de aprendizaje de carreras



**Figura 5.** ESTRUCTURA DE ÁRBOL DEL MODELO GENÉRICO PARA LA EVALUACIÓN DEL ENTORNO DE APRENDIZAJE DE LA CARRERAS

**FUENTE:** <http://www.ceaaces.gob.ec/sitio/proceso-de-recategorizacion-institucional-2/>

La Tasa de Retención se encuentra en el subcriterio **Eficiencia** dentro del criterio **Estudiantes**, este indicador evalúa la relación de los estudiantes de la carrera que fueron admitidos dos años antes del período de evaluación y que se encuentran matriculados a la fecha.

Además es el coeficiente que expresa la relación entre los estudiantes que han permanecido en la carrera hasta concluir su formación académica y el total de los matriculados, en una misma cohorte.

**Estándar.-** Se fijará a través de la función de utilidad, considerando parámetros de calidad con la información obtenida del proceso de evaluación.

**Cálculo del indicador y escala:**

$$TR = 100. \frac{NEMA}{NTEA}$$

*TR: Tasa de retención.*

*NEMA: Número de estudiantes matriculados en la carrera en el último semestre o el último año concluido antes del inicio del proceso de evaluación que fueron admitidos dos años antes.*

*NTEA: Número total de estudiantes que fueron admitidos en la carrera dos años antes del periodo de evaluación.*

**Variable:**

*Número de estudiantes matriculados en la carrera en el último semestre o el último año concluido antes del inicio del proceso de evaluación que fueron admitidos dos años antes.*

**Variable:**

*Número total de estudiantes que fueron admitidos en la carrera dos años antes del periodo de evaluación.*

▪ **Evaluación de la Calidad de las Carreras**

El Modelo Genérico de Evaluación de la Calidad de las Carreras es un proceso que consiste en realizar una valoración de éstas, comparándolas y contextualizándolas en un marco histórico, económico y social, que se determina de acuerdo al enfoque teórico utilizado para definir la calidad, pero que es, en esencia, un proceso complejo en el que intervienen múltiples factores que interactúan de formas diversas y poco predecibles. En este proceso, por carrera se ha adoptado la definición que utilizó el CONESUP, que señala que la carrera representa la unidad básica objeto de la presente propuesta de evaluación del entorno de aprendizaje, y se caracteriza como: "Una organización curricular que prepara a los estudiantes para la obtención de un título de Licenciado o un Título Profesional Universitario o Politécnico, luego de que éstos han demostrado haber alcanzado una formación que les permite solucionar problemas a través de la aplicación de conocimientos científicos, habilidades y

destrezas, procedimientos y métodos, dentro de un área científico-tecnológica determinada" (CONESUP, 2009).

Lo cual complementa a lo definido por el Consejo de Educación Superior (CES) en su Reglamento para Aprobación de Proyectos de Carreras y Posgrados de las universidades y escuelas politécnicas del Ecuador (2012) para una carrera, como: "Conjunto de actividades educativas conducente al otorgamiento de un grado académico o título profesional de tercer nivel, orientadas a la formación de una disciplina, o al ejercicio de una profesión". Es decir, las actividades educativas caracterizadas en la definición, se entienden como las suficientes para la obtención del título de Licenciado o de Profesional Universitario o Politécnico.

A partir de todas estas consideraciones y concepciones teóricas y metodológicas se ha formulado un modelo genérico de evaluación de las carreras que se lo presenta en este documento. Para ello, desde la óptica de la TDM se han determinado cinco criterios básicos que cubren prácticamente la totalidad de las funciones y actividades de las carreras, como son: la pertinencia de la carrera en el entorno, la formación que se ofrece a los futuros profesionales, la calidad de su planta de docentes, el ambiente en el que se desenvuelven las actividades académicas, y la relación y participación de los estudiantes en esas actividades. De esta manera, los criterios correspondientes a estas actividades y funciones se han establecido como:

- Pertinencia
- Plan Curricular
- Academia
- Ambiente Institucional y
- Estudiantes.

## ▪ **Evaluación de Carreras**

Uno de los procesos para ejecutar la “evaluación de desempeño institucional de las universidades y escuelas politécnicas del Ecuador”, como ordena el Mandato Constituyente No 14 de 2008, consiste en la evaluación de las carreras presenciales y semi-presenciales ofrecidas por estas instituciones de educación superior (IES); dicho proceso demanda la elaboración de un modelo genérico a partir del cual se construirá modelos específicos complementarios para las carreras. La necesidad de llevar a cabo estas evaluaciones, y en particular la de carreras, deviene del derecho fundamental que tiene la ciudadanía de recibir una educación superior de calidad, conforme la constitución de la república y las normas supranacionales que integran el bloque de constitucionalidad del sistema jurídico ecuatoriano; además, la educación es un servicio público y como tal, debe ser llevado a cabo con ciertos parámetros de calidad establecidos por el organismo competente que, en el caso de la educación superior, es el CEAACES. Todo esto necesariamente implica que la evaluación de la educación superior está relacionada inseparablemente con la búsqueda de la calidad de los procesos que la constituyen, y de su aseguramiento.

A partir de las premisas expuestas, es necesario enmarcar el concepto de calidad con el que se quiere evaluar las carreras y, además, los criterios que se utilizarán para tal evaluación. Ahora bien, los procesos implícitos en el funcionamiento y la realización de una carrera son harto complejos e involucran a un número considerable de actores, lo que ha exigido la adopción, por parte del CEAACES, de una conceptualización de calidad proveniente de múltiples enfoques, tales como la tendencia a la calidad, a la perfección, al cumplimiento y ajuste de propósitos y objetivos, a la producción de valor a menor costo, o a la transformación de un sistema. Por separado, ninguno de estos enfoques agotaría la complejidad de los procesos universitarios, y algunos podrían incluso ser incompatibles con ciertos postulados básicos que rigen los procesos de democratización de la universidad ecuatoriana; sin embargo, en su conjunto, y con ciertas particularizaciones, estos enfoques pueden acercarse



satisfactoriamente a una concepción acorde a la deseable para la evaluación de dichos procesos. Por otra parte, en lo que tiene relación con la selección de los criterios a ser utilizados en el modelo, también se ha tomado en cuenta la complejidad de la situación, la cual implica la evaluación de diversas actividades de diferente índole, en cuya caracterización pueden intervenir elementos que demanden una evaluación más o menos cualitativa.

Si bien es cierto que, en su mayoría, los indicadores seleccionados revisten un carácter estrictamente cuantitativo, considerando la situación señalada, en la propuesta de evaluación se ha adoptado la Teoría de la Decisión Multicriterio (TDM), que constituye un fundamento metodológico coherente con el problema planteado, ya que contempla un proceso de toma de decisiones para resolver problemas en “sistemas complejos”, caracterizados por incertidumbre y conflicto de intereses. Esta metodología es especialmente utilizada en decisiones que incluyen aspectos intangibles a evaluar, y sirve para tratar problemas de toma de decisiones en los que se presentan “diversos objetivos o criterios que simultáneamente deben incorporarse”

(Ávila Mogollón, 20003)

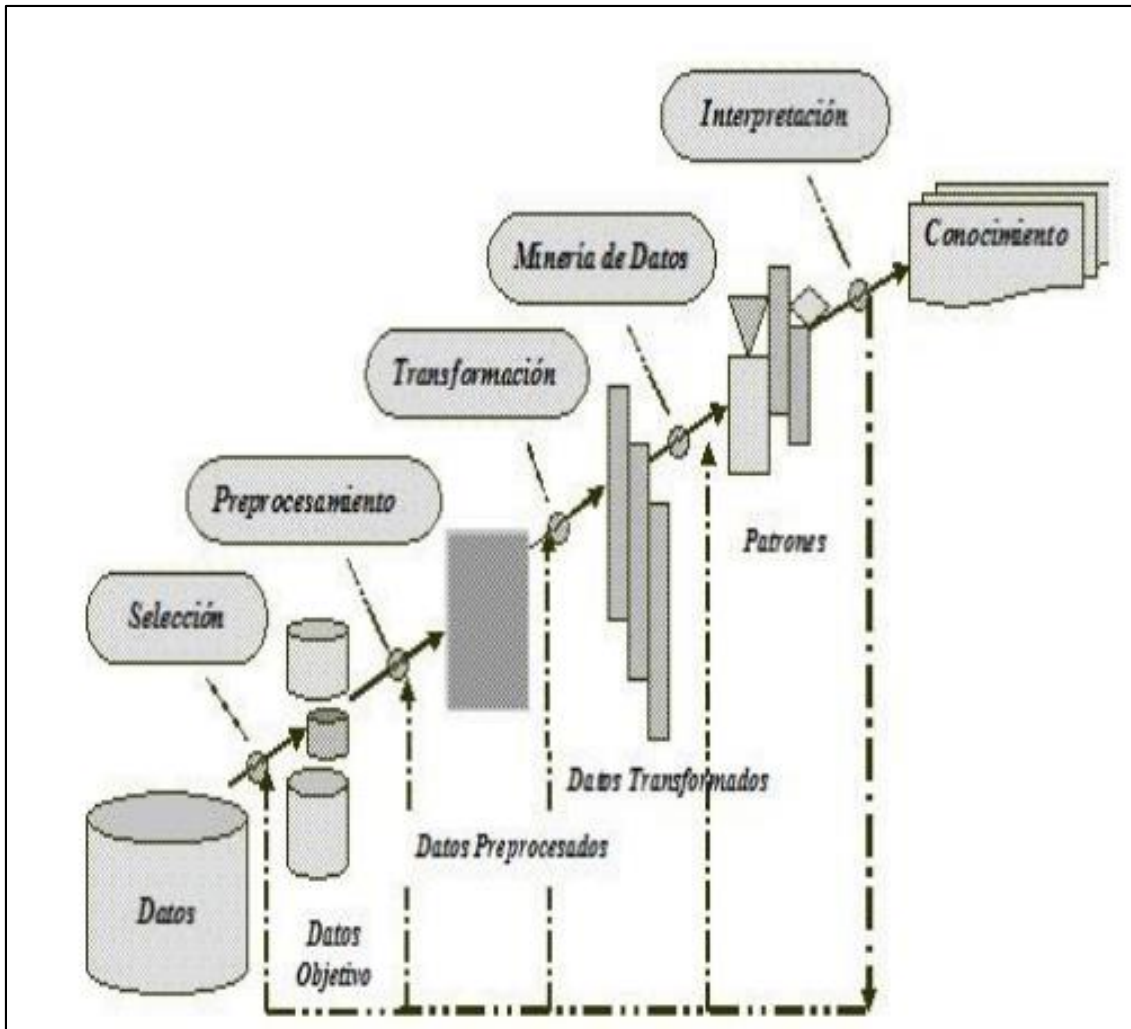
## **2.4.2. Categorías fundamentales de la Variable Dependiente**

- **Descubrimiento de Conocimiento en Bases de Datos**

Es la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de un conjunto de datos. (Beltrán Martínez)

El Descubrimiento de Conocimiento en Bases de Datos (DCBD) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice. Esta tarea implica generalmente preprocesar los datos, hacer minería de datos y presentar resultado. Este proceso es interactivo e iterativo, involucra numerosos pasos con la intervención del usuario en la toma de muchas decisiones. En la Figura 6 se muestran las etapas del proceso DCBD.

(Pereira, 2009)



**Figura 6.** Etapas del Proceso DCBD

**Fuente:** (Pereira, 2009)

Un sistema de descubrimiento será un programa que toma como entrada el conjunto de hechos y extrae las regularidades existentes. Cuando el conocimiento se extrae partiendo de los datos de una base de datos, se tiene KDD. (Beltrán Martínez)

- **Minería de Datos**

Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas (César Pérez López, 2007)

La data mining es la etapa de descubrimiento en el proceso de KDD: Paso consistente en el uso de algoritmos concretos que generan una enumeración de patrones a partir de los datos preprocesados (Vallejos, 2006).

### **Predictivas**

- **Clasificación o discriminación (en estadística)**

La clasificación asume que hay un conjunto de objetos caracterizados por algún atributo o rasgo que pertenece a distintas clases. La etiqueta de clase es un valor discreto y es conocido para cada objeto.

El objetivo de esta tarea es asignar la etiqueta de clase correcta a objetos nuevos y sin etiqueta dados los valores de sus atributos. La clasificación es una de las tareas más comunes en la minería de datos.

Un ejemplo de clasificación puede ser clasificar un mensaje de correo electrónico como spam o no.

- **Clasificación suave**

Esta tarea es igual que la clasificación pero se introduce una función que estima el grado de certeza de la predicción.

Un ejemplo sería clasificar un mensaje de correo electrónico como spam o no, proporcionando además la certeza de la clasificación.

- Estimación de probabilidad de clasificación  
 Es una generalización de la clasificación suave. El problema a resolver es el mismo que para la clasificación y clasificación suave.  
 La diferencia está en que en esta tarea el resultado es un conjunto de probabilidades de que el objeto pertenezca a una clase u otra.  
 Por ejemplo, si se quisiera clasificar entre varios medicamentos cuál es el mejor para una determinada patología, esta tarea proporcionaría la probabilidad de que sea cada uno de los medicamentos escogidos.
- Categorización  
 En esta tarea a diferencia de las clasificaciones donde a cada objeto le corresponde una y sólo una clase, a un objeto le puede corresponder n clases.  
 Por ejemplo, dado un conjunto de documentos, asignar categorías de los temas que trata cada documento.
- Regresión  
 Esta tarea es muy parecida a la clasificación ya que a cada elemento se le asigna únicamente un valor de salida, con la diferencia de que este valor de salida es un valor numérico, es decir, puede ser un valor entero o real.  
 Un ejemplo muy sencillo sería estimar las ventas de un determinado producto para un determinado año.

## **Descriptivas**

- Agrupamiento (clustering)  
 El objetivo de esta tarea es agrupar los elementos de un conjunto de objetos, de tal manera que los elementos que formen parte de un mismo grupo tengan características similares. La diferencia con respecto a la clasificación es que en el agrupamiento son los grupos y la pertenencia a dichos grupos lo que se quiere determinar, y además a priori no se sabe cuántos grupos hay ni cómo son.  
 Un ejemplo podría ser agrupar clientes en grupos diferenciados para luego estudiar que grupos se comportan mejor ante determinados productos.

- **Correlaciones y factorizaciones**  
El objetivo es ver si dos o más atributos numéricos están correlacionados linealmente o relacionados de algún modo.
- **Reglas de asociación**  
Esta tarea es parecida a la de las correlaciones y factorizaciones pero en este caso se utilizan atributos nominales en lugar de numéricos.  
Se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.  
Un ejemplo muy conocido sobre las reglas de asociación es el de la “cerveza y los pañales” que está en el comportamiento de los compradores en el supermercado. Se descubrió que muchos hombres compraban pañales por encargo de sus esposas, por lo que una cadena de supermercados decidió colocar la cerveza junto a los pañales y el resultado fue un aumento en la venta de cerveza. (Galán Cortina, 2016)

### **Los Fundamentos del Data Mining**

Las técnicas de Data Mining son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y más recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. Data Mining toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva. Data Mining está listo para su aplicación en la comunidad de negocios porque está soportado por tres tecnologías que ya están suficientemente maduras:

- Recolección masiva de datos.
- Potentes computadoras con multiprocesadores.
- Algoritmos de Data Mining

## **Principales características y objetivos de la Minería de Datos**

- Explorar los datos se encuentran en las profundidades de las bases de datos, como los almacenes de datos, que algunas veces contienen información almacenada durante varios años.
- En algunos casos, los datos se consolidan en un almacén de datos y en mercados de datos; en otros, se mantienen en servidores de Internet e Intranet.
- El entorno de la minería de datos suele tener una arquitectura cliente servidor.
- Las herramientas de la minería de datos ayudan a extraer el mineral de la información enterrado en archivos corporativos o en registros públicos, archivados.
- El minero es, muchas veces un usuario final con poca o ninguna habilidad de programación, facultado por barrenadoras de datos y otras poderosas herramientas indagatorias para efectuar preguntas adhoc y obtener rápidamente respuestas.
- Hurgar y sacudir a menudo implica el descubrimiento de resultados valiosos e inesperados.
- Las herramientas de la minería de datos se combinan fácilmente y pueden analizarse y procesarse rápidamente.
- Debido a la gran cantidad de datos, algunas veces resulta necesario usar procesamiento en paralelo para la minería de datos.
- La minería de datos produce cinco tipos de información:
  - Asociaciones.
  - Secuencias.
  - Clasificaciones.
  - Agrupamientos.
  - Pronósticos.
- Los mineros de datos usan varias herramientas y técnicas.

(Vallejos, 2006)

- **Métodos de Minería de Datos**

Para poder resolver cualquiera de las tareas anteriores se necesitan unas técnicas, algoritmos o métodos. Es importante saber que una tarea se puede resolver con diferentes métodos y no sólo uno, y que un método puede servir para resolver más de una tarea. A continuación veremos los distintos tipos de métodos que existen para poder resolver las tareas anteriormente mencionadas:

- ✓ Técnicas algebraicas y estadísticas

El objetivo de estas técnicas es expresar modelos y patrones mediante fórmulas algebraicas, funciones lineales, funciones no lineales, distribuciones o valores agregados estadísticos tales como medias, varianzas, correlaciones, etc. Generalmente estas técnicas obtienen un patrón a partir de un modelo predeterminado del cual se estiman unos coeficientes o parámetros. Ejemplos de algoritmos dentro de estas técnicas son: regresión lineal (global o local), regresión logarítmica y regresión logística.

- ✓ Técnicas bayesianas

El objetivo es estimar la probabilidad de pertenencia a una clase o grupo mediante la estimación de las probabilidades condicionales inversas o a priori, para lo cual utilizan el teorema de Bayes. Ejemplos de algoritmos dentro de este grupo son: el clasificador bayesiano naive, los métodos basados en máxima verisimilitud y el algoritmo EM.

- ✓ Técnicas basadas en conteos de frecuencias y tablas de contingencia

Se trata de contar la frecuencia con la que dos o más sucesos se dan conjuntamente. Cuando el conjunto de sucesos posibles es muy grande, sólo se cuentan aquellos casos en los que las frecuencias conjuntas superan un determinado umbral.



- ✓ Técnicas basadas en árboles de decisión y sistemas de aprendizaje de Reglas  
Estas técnicas además de sus representaciones en forma de reglas, también se basan en algoritmos del tipo “divide y vencerás” (por ejemplo el ID3/C4.5 o el CART) y “separa y vencerás” (por ejemplo el CN2).

- ✓ Técnicas relacionales, declarativas y estructurales

Estas técnicas representan el modelo mediante lenguajes declarativos, como los lenguajes lógicos, funcionales o lógico-funcionales.

- ✓ Técnicas basadas en redes neuronales artificiales

Son técnicas que aprenden un modelo mediante el entrenamiento de los pesos que conectan un conjunto de nodos o neuronas. La topología de la red y los pesos de las conexiones determinan el patrón aprendido. Existen multitud de variantes de organización con muchos algoritmos diferentes para cada organización, pero el más conocido probablemente sea el de retropropagación.

- ✓ Técnicas basadas en núcleo y máquinas de soporte vectorial

Son técnicas que intentan maximizar el margen entre los grupos o las clases formadas. Para ello se basan en unas transformaciones que pueden aumentar la dimensionalidad, que se llaman núcleos o kernels.

- ✓ Técnicas estocásticas y difusas

En este grupo se incluyen la mayoría de las técnicas que junto a las redes neuronales forman lo que se denomina computación flexible (soft computing). Son técnicas en las que o bien los componentes aleatorios son fundamentales, como el simulated annealing, los métodos evolutivos y genéticos, o bien al utilizar funciones de pertenencia difusas (fuzzy).

- ✓ Técnicas basadas en casos, en densidad o distancia

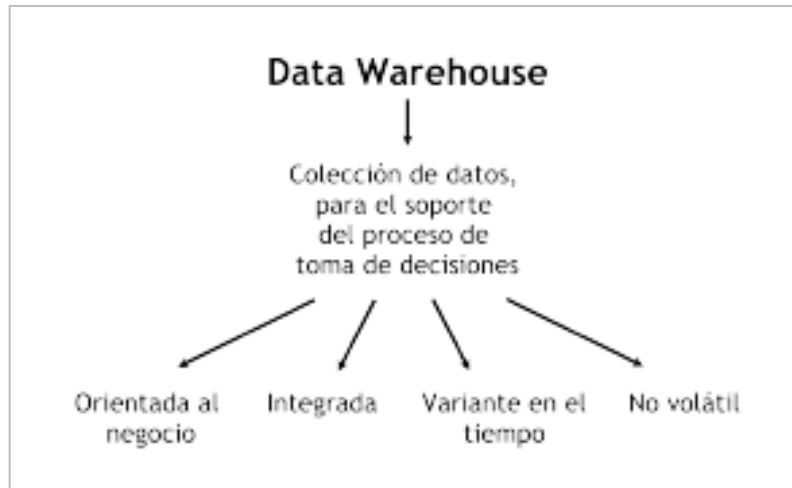
Son métodos que se basan en distancias al resto de elementos, ya sea directamente, como los vecinos más próximos, o de una manera más sofisticada, mediante la estimación de funciones de densidad. Algunos algoritmos muy conocidos son los jerárquicos, como Two-step o COBWEB, y los no jerárquicos como K medias.

(Galán Cortina, 2016)

- **Almacén de Datos**

Los Data Warehouse o Almacén de Datos es una herramienta útil que permiten recoger la información de las diferentes Bases de Datos Transaccionales que posee una Institución, su principal objetivo es filtrar y procesar la información para su almacenamiento, con la finalidad de brindar una plataforma de datos consolidados e históricos que permitan mejorar el análisis de la información para una adecuada toma de decisiones (Carrera Sepúlveda, Varas Contreras, & Urrutia Sepúlveda, 2010).

Para extraer conocimiento de la información histórica almacenada en una Institución se debe utilizar los Data Warehouse o Almacenes de Datos, en vista de que permiten disponer de Sistemas de Información de apoyo a la toma de decisiones (DSS o Decision Support Systems) (Pérez López & Santín González, Minería de datos: técnicas y herramientas, 2007).



**Figura 7.** Data WareHouse

**Fuente:** <http://www.businessintelligence.info/docs/hefesto-v2.pdf>

### **Características:**

- **Orientada al negocio**

La primera característica del DW, es que la información se clasifica en base a los aspectos que son de interés para la organización. Esta clasificación afecta el diseño y la implementación de los datos encontrados en el almacén de datos, debido a que la estructura del mismo difiere considerablemente a la de los clásicos procesos operacionales orientados a las aplicaciones.

- **Integrada**

La integración implica que todos los datos de diversas fuentes que son producidos por distintos departamentos, secciones y aplicaciones, tanto internos como externos, deben ser consolidados en una instancia antes de ser

agregados al DW, y deben por lo tanto ser analizados para asegurar su calidad y limpieza, entre otras cosas. A este proceso se lo conoce como Integración de Datos, y cuenta con diversas técnicas y subprocesos para llevar a cabo sus tareas. Una de estas técnicas son los procesos ETL: Extracción, Transformación y Carga de Datos.

- **Variante en el tiempo**

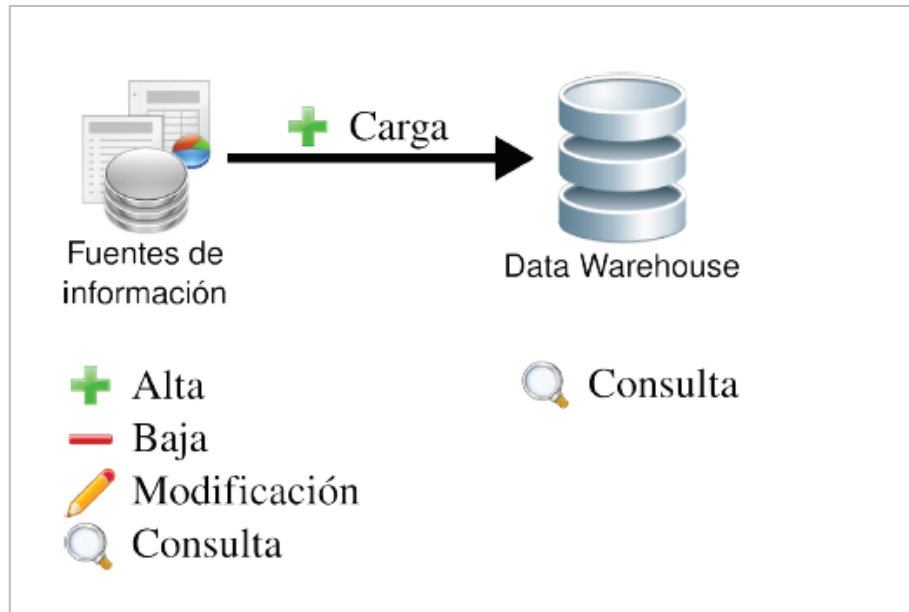
Debido al gran volumen de información que se manejará en el DW, cuando se le realiza una consulta, los resultados deseados demorarán en originarse. Este espacio de tiempo que se produce desde la búsqueda de datos hasta su consecución es del todo normal en este ambiente y es, precisamente por ello, que la información que se encuentra dentro del depósito de datos se denomina de tiempo variable.

- **No volátil**

La información es útil para el análisis y la toma de decisiones solo cuando es estable. Los datos operacionales varían momento a momento, en cambio, los datos una vez que entran en el DW no cambian. La actualización, o sea, insertar, eliminar y modificar, se hace de forma muy habitual en el ambiente operacional sobre una base, registro por registro, en cambio en el depósito de datos la manipulación básica de los datos es mucho más simple, debido a que solo existen dos tipos de operaciones: la carga de datos y el acceso a los mismos.

Por esta razón es que en el DW no se requieren mecanismos de control de concurrencia y recuperación.

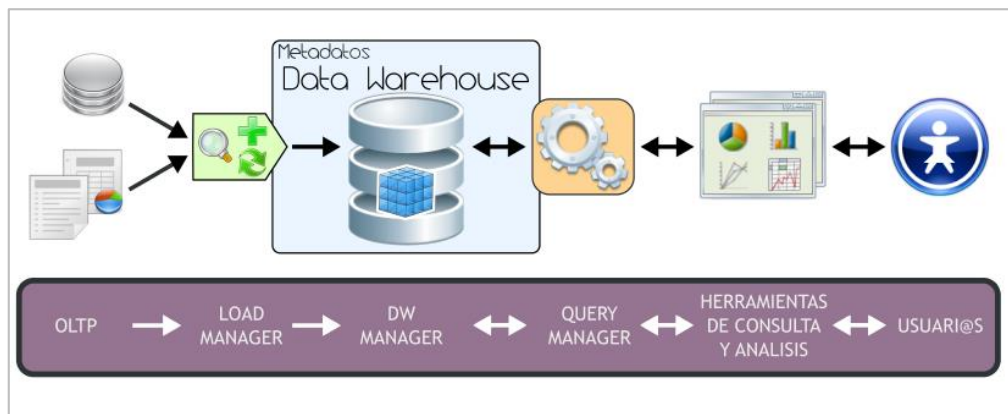
(Bernabeu, 2010).



**Figura 8. Carga de almacén de datos**

Fuente: (Bernabeu, 2010)

### Arquitectura Data WAREHOUSING



**Figura 9. Arquitectura de Data Warehouse**

Fuente: (Bernabeu, 2010)

- **OLTP**

OLTP (On Line Transaction Processing), representa toda aquella información transaccional que genera la empresa en su accionar diario, además, de las fuentes externas con las que puede llegar a disponer.

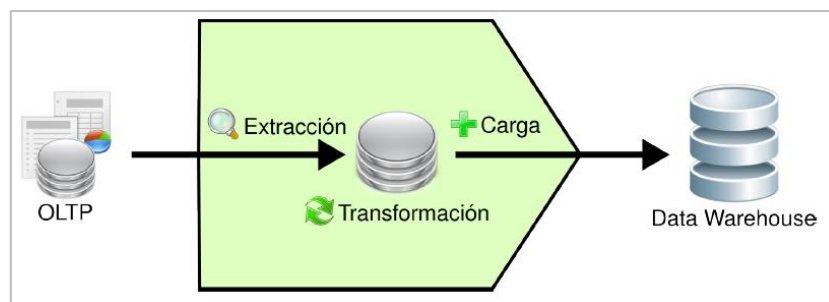
Como ya se ha mencionado, estas fuentes de información, son de características muy disímiles entre sí, en formato, procedencia, función, etc.

Entre los OLTP más habituales que pueden existir en cualquier organización se encuentran: Archivos de textos, Hipertextos, Hojas de cálculos, Informes, Bases de datos transaccionales.

- **Load Manager**

Para poder extraer los datos desde los OLTP, para luego manipularlos, integrarlos y transformarlos, para posteriormente cargar los resultados obtenidos en el DW, es necesario contar con algún sistema que se encargue de ello. Precisamente, la Integración de Datos es quien cumplirá con tal fin.

La Integración de Datos agrupa una serie de técnicas y subprocesos que se encargan de llevar a cabo todas las tareas relacionadas con la extracción, manipulación, control, integración, depuración de datos, carga y actualización del DW, etc. Es decir, todas las tareas que se realizarán desde que se toman los datos de los diferentes OLTP hasta que se cargan en el DW.



**Figura 10.** Proceso ETL

**Fuente:** (Bernabeu, 2010)

- **Carga**

Esta función se encarga, por un lado de realizar las tareas relacionadas con: Carga Inicial (Initial Load).

Actualización o mantenimiento periódico (siempre teniendo en cuenta un intervalo de tiempo predefinido para tal operación).

La carga inicial, se refiere precisamente a la primera carga de datos que se le realizará al DW. Por lo general, esta tarea consume un tiempo bastante considerable, ya que se deben insertar registros que han sido generados aproximadamente, y en casos ideales, durante más de cinco años.

- **Deserción Estudiantil**

Para definir la deserción estudiantil, es necesario comenzar por el significado del término deserción, que como se encuentra en el artículo publicado en internet “Definición de Deserción Escolar” se define como, abandonar las obligaciones y separarse de las concurrencias que se solían frecuentar, y que la deserción escolar se utiliza para referirse a aquellos alumnos que dejan de asistir a clase y quedan fuera del sistema educativo. (Castro & Uquillas, 2011).

**Causas del abandono de los estudios universitarios.**

Son varias las razones por las cuales los alumnos deciden, de manera voluntaria, abandonar los estudios universitarios, que pueden ser de índole académica, personal académico, personal, familiar, institucional y social.

José María Álvarez en su libro “Etiología de un sueño o el abandono de la Universidad por parte de los estudiantes por factores no académicos”, las clasifica en las siguientes categorías tomando en cuenta a algunos autores:

- **Razones personales**

Se refiere a las causas que influyeron de manera particulares en cada uno de los individuos en la decisión de abandonar sus estudios universitarios. Estas van de la mano con el estado anímico, los focos de interés, la motivación frente al trabajo y las relaciones intrafamiliares del sujeto.

- ✓ Factores motivacionales.
- ✓ Factores psicológicos o emocionales
- ✓ Expectativa no satisfecha del estudiante.
- ✓ Infelicidad por factores no académicos.
- ✓ Choque de horario entre trabajo
- ✓ Influencias del contexto familia.
- ✓ Influencias del contexto externo a la familia.
- ✓ Ausencia de Orientación Vocacional.
- ✓ Mala elección de carrera o de institución.

- **Razones académicas.**

Son aquellas causas que hacen referencia al nivel de instrucción con la que inician sus estudios universitarios y a la metodología y técnicas de estudio que se utilizan para el aprendizaje.

- ✓ Ausencia de aptitud académica.
- ✓ Bajo rendimiento académico.
- ✓ Reprobación de cursos
- ✓ Deficiente formación previa.
- ✓ Ausencia de disciplina y método de estudio.



- **Razones institucionales.**

Éstas hacen referencia a las características que posee la institución educativa, en la metodología de enseñanza y a la propuesta educativa de la misma, que no cumplen con las expectativas del estudiante.

- ✓ Incómoda distancia entre la vivienda y la Universidad.
- ✓ Desfavorable ambiente en la residencia.
- ✓ Deficiencias en el proceso educativo propiamente hablando.
- ✓ Ausencia de métodos enseñanza.
- ✓ Deficiencias en los programas académicos.
- ✓ Influencias negativas ejercidas por profesores y personas de los centros educativos. (Castro & Uquillas, 2011).

## **2.5. Hipótesis**

El análisis de la tasa de retención si incide en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

## **2.6. Señalamiento de Variables**

**Variable Independiente:** Análisis de la tasa de retención

**Variable Dependiente:** La detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

## **CAPÍTULO III**

### **METODOLOGÍA**

#### **3.1. Enfoque**

Esta investigación va a tener dos enfoques, el primero es cualitativo porque el análisis de la tasa de retención influye en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato, y el segundo es cuantitativo porque va a obtener parámetros de medición con el análisis de la tasa de retención estudiantil.

#### **3.2. Modalidad Básica de Investigación**

- **Investigación Bibliográfica**

La investigación será bibliográfica porque utilizara fuentes como libros, revistas, documentos para la construcción del marco teórico tanto de análisis de la tasa de retención como en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

- **Investigación Campo**

La investigación también tendrá la modalidad de campo por que se buscara obtener la información análisis de la tasa de retención y en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

#### **3.3. Nivel o tipo de investigación**

Esta investigación es investigación exploratoria, descriptiva, explicativa.

La investigación correlacionar por que buscara medir el grado de relación existente entre el análisis de la tasa de retención y en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

### **3.4. Población y Muestra**

El presente proyecto trabaja con la población total que es el grupo de profesionales encargados de las coordinadores UPE de cada una de las carreras de la Universidad Técnica de Ambato, como se puede apreciar en la tabla 1

<b>PERSONAL</b>	<b>NUMERO (FRECUENCIA)</b>	<b>%</b>
Docentes UPE	40	100 %
<b>Total</b>	<b>40</b>	<b>100%</b>

**Tabla 1:** Población y muestra

**Elaborado por:** Investigador

### 3.5. OPERACIONALIZACIÓN DE LAS VARIABLES

#### Variable Independiente: Tasa de Retención

CONCEPTUALIZACIÓN	DIMENSIONES	INDICADORES	ÍTEMS BÁSICOS	TÉCNICAS E INSTRUMENTOS
Un programa de retención estudiantil comprende las estrategias que desarrolla el aparato educativo en una institución para garantizar el acompañamiento al estudiante durante su trayectoria académica, de tal manera que pueda culminarla de forma exitosa. Su objetivo es proporcionar las herramientas necesarias para la terminación de los diferentes ciclos y etapas en los tiempos establecidos, y adicionalmente asegurar el conocimiento necesario y el desarrollo de competencias y actitudes indispensables para desenvolverse en la vida	<p>Programa de retención estudiantil.</p> <p>Estrategias que desarrolla el aparato educativo</p> <p>Garantizar el acompañamiento al estudiante</p>	<p>Retención Estudiantil</p> <p>Estrategias</p>	<p>¿El Sistema Informático de la Universidad Técnica de Ambato genera algún reporte la Tasa de Retención Estudiantil?</p> <p>¿Se realiza un análisis de la Tasa de Retención Estudiantil?</p> <p>¿Existen estrategias para maximizar el índice de Retención Estudiantil?</p> <p>¿Es importante para la institución, la Retención Estudiantil?</p>	<p>Encuesta</p> <p>Docentes Responsables UPE</p>

**Tabla 2:** Operacionalización de las variables (Variable Independiente)

**Elaborado por:** Investigador

**Variable Dependiente: Detección de Patrones de Deserción Estudiantil**

CONCEPTUALIZACIÓN	DIMENSIONES	INDICADORES	ITEMES BÁSICOS	TÉCNICAS E INSTRUMENTOS
<p>El Descubrimiento de Conocimiento en Bases de Datos (DCBD) es básicamente un proceso automático en el que se combinan descubrimiento y análisis. El proceso consiste en extraer patrones en forma de reglas o funciones, a partir de los datos, para que el usuario los analice</p>	<p>Descubrimiento de Conocimiento.</p> <p>Proceso automático en el que se combinan descubrimiento y análisis.</p> <p>Extraer patrones en forma de reglas o funciones, a partir de los datos</p>	<p>Conocimiento</p> <p>Análisis</p> <p>Patrones</p>	<p>¿El Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil?</p> <p>¿Es importante para la institución identificar los Patrones de Deserción Estudiantil?</p> <p>¿Los reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil?</p> <p>¿La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato?</p>	<p>Encuesta</p> <p>Docente UPE.</p>

**Tabla 3.** Operacionalización de las variables (Variable Independiente)

**Elaborado por:** Investigador

### 3.6. Recolección de Información

La recolección de información se puede observar en la tabla 4.

PREGUNTAS BÁSICAS	EXPLICACIÓN
¿Para qué?	Para cumplir los objetivos de la investigación
¿De qué personas u objetos?	A quien se va entrevistar: <ul style="list-style-type: none"><li>▪ Docentes UPE por cada carrera.</li></ul>
¿Sobre qué aspectos?	Sobre los indicadores expuestos en la matriz
¿Quién, Quiénes?	Cristina Alexandra Frutos Ortega
¿Cuándo?	Segundo semestre 2016
¿Dónde?	Universidad Técnica de Ambato
¿Cuántas veces?	Una vez para la obtención de la información de la investigación
¿Qué técnicas de recolección?	Encuesta
¿Con qué?	Encuesta
¿En qué situación?	Condiciones normales mientras se cumple funciones

**Tabla 4.** Recolección de información

**Elaborado por:** Investigador

### 3.7. Procesamiento y Análisis de la Información

#### 3.7.1. Plan de Procesamiento de la Información

Los datos obtenidos con las técnicas e instrumentos utilizados, serán clasificados y tabulados siguiendo ciertos procedimientos que ayuden a que los resultados presentados sean lo más reales posibles.

Los procedimientos a seguir serán:

- Revisión crítica y minuciosa de la información obtenida, a través de esta se deberá separar la información defectuosa; es decir, contradictoria, incompleta, no pertinente, etc.

- Repetición de la recolección de la información en caso que esta se muestre muy obsoleta.
- Tabulación de la información obtenida, en base a los criterios de las variables de la hipótesis.
- Presentación de resultados estadísticos.

### **3.7.2. Plan de Análisis e Interpretación de Resultados.**

- Análisis de los resultados estadísticos, resaltando tendencias o relaciones basándonos en los objetivos planteados e hipótesis.
- Interpretación de los resultados, con apoyo del marco teórico.
- Verificación de la hipótesis.
- Establecimiento de conclusiones y recomendaciones

## **CAPÍTULO IV**

### **ANÁLISIS E INTERPRETACIÓN DE RESULTADOS**

#### **4.1. Análisis e Interpretación de Resultados**

Los resultados que se muestran a continuación son basados en la encuesta realizada a los Docentes UPE de cada una de las carreras de la Universidad Técnica de Ambato.

#### **Encuesta realizada a los Docentes UPE de las Carreras de Universidad Técnica de Ambato.**

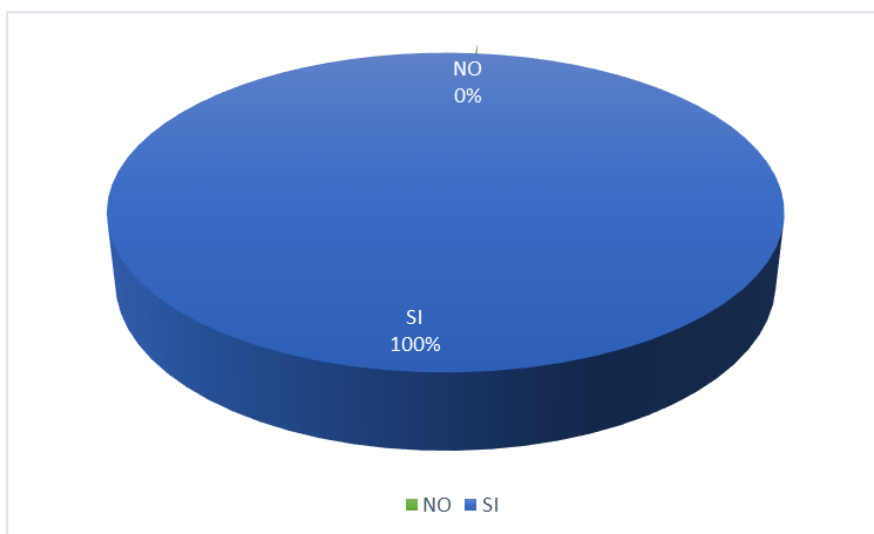
La encuesta consta de 10 preguntas y fue aplicada a 40 Docentes Responsables de UPE y 4 Funcionarios de la DEAC.



**Pregunta 1:** ¿El Sistema Informático de la Universidad Técnica de Ambato genera el reporte “La Tasa de Retención Estudiantil”?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	40	100%
NO	0	0%
<b>TOTAL</b>	40	100%

**Tabla 5.** Cálculo de la Tasa de Retención Estudiantil  
**Elaborado por:** Investigador



**Figura 11.** Cálculo de la Tasa de Retención Estudiantil  
**Elaborado por:** Investigador

### **Análisis**

De los 40 encuestados (100%), el 100% afirman que el Cálculo de la Tasa de Retención Estudiantil SI se obtiene del Sistema Informáticos de la Universidad Técnica de Ambato

### **Interpretación**

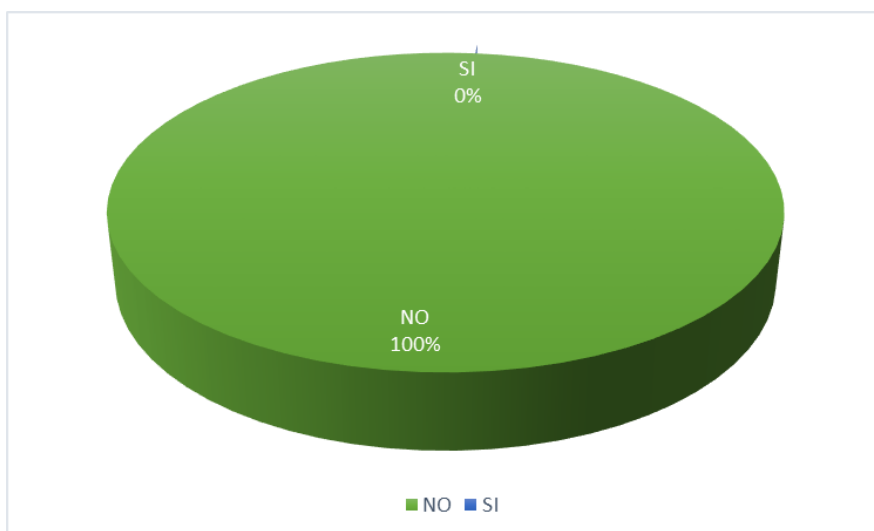
En la Encuesta realizada a los docentes responsable de UPE de la Universidad Técnica de Ambato, se puede, se observa que el 100% de los docentes afirman que la Tasa de Retención Estudiantil actualmente se obtiene de un sistema de información. Claramente se puede observar que el Sistema Integrado de la Universidad Técnica de Ambato permite obtener este indicador a través de un Sistema de Información.

## Pregunta 2:

¿Se realiza un análisis de la Tasa de Retención Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	0	0%
NO	44	100%
<b>TOTAL</b>	44	100%

**Tabla 6.** Análisis de la Tasa de Retención Estudiantil  
**Elaborado por:** Investigador



**Figura 12.** Análisis de la Tasa de Retención Estudiantil  
**Elaborado por:** Investigador

### Análisis

De los 40 encuestados (100%), el 100% de los encuestados indican que no se realiza un análisis de la Tasa de Retención Estudiantil.

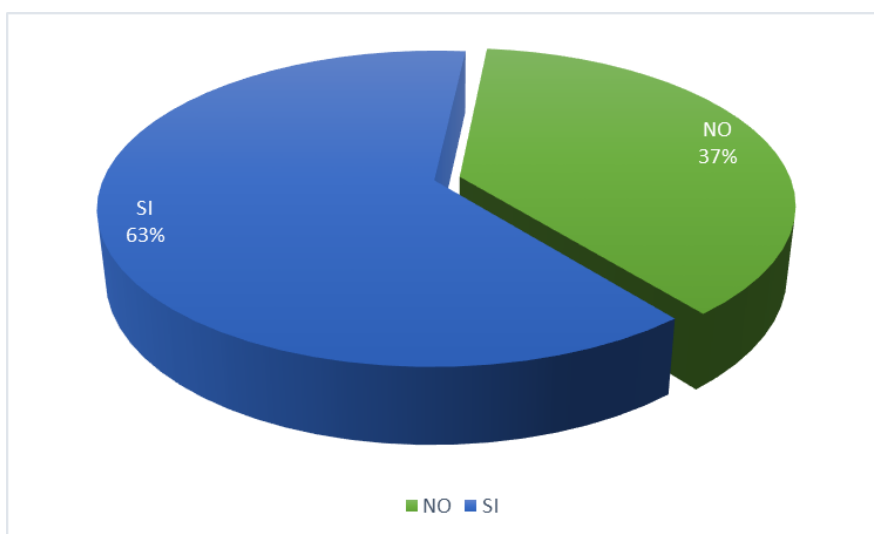
### Interpretación

Los docentes UPE de la carreras de la Universidad Técnica de Ambato únicamente obtienen la Tasa de Retención Estudiantil por ser un indicador de evaluación y acreditación.

**Pregunta 3:** ¿Existen estrategias para maximizar el índice de Retención Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	25	63%
NO	15	37%
<b>TOTAL</b>	40	100%

**Tabla 7.** Estrategias para maximizar el índice de Retención Estudiantil  
**Elaborado por:** Investigador



**Figura 13.** Estrategias para maximizar el índice de Retención Estudiantil  
**Elaborado por:** Investigador

### **Análisis**

De los 40 encuestados (100%), el 63% de los encuestados indican que la institución SI cuenta con estrategias para maximizar el índice de Retención Estudiantil, mientras que 37% de los encuestados afirman que NO.

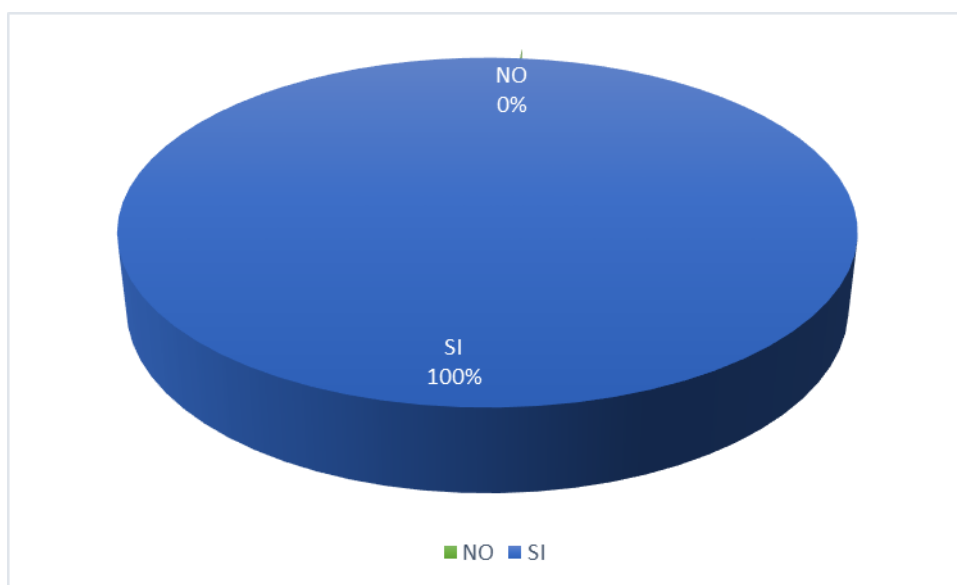
### **Interpretación**

El 63% de los docentes UPE consideran que la funciones del Departamento de Bienestar Estudiantil realice gestiones que aportan con la retención Estudiantil, mientras que el 37% consideran que NO se realizan actividades que permitan mejorar el índice que Retención de los Estudiantes.

**Pregunta 4:** ¿Es importante para la institución, la Retención Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	40	100%
NO	0	0%
<b>TOTAL</b>	40	100%

**Tabla 8.** Importancia para la institución, la Retención Estudiantil  
**Elaborado por:** Investigador



**Figura 14.** Importancia para la institución, la Retención Estudiantil  
**Elaborado por:** Investigador

#### **Análisis**

De los 40 encuestados (100%), el 100% de los encuestados indican que es de vital importancia la Retención o permanencia de los estudiantes de la Universidad Técnica de Ambato.

#### **Interpretación**

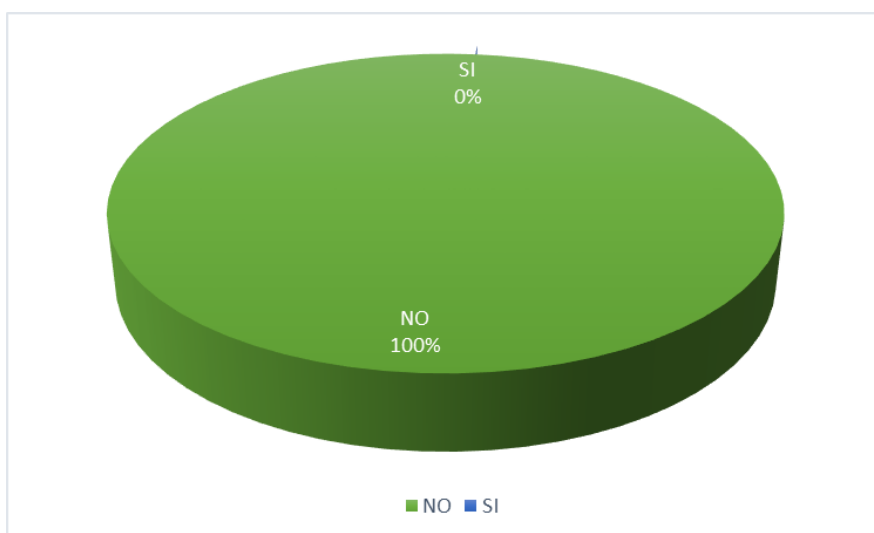
Claramente se puede observar que la Retención de los Estudiantes es un factor importante en las Instituciones de Educación.

**Pregunta 5:**

¿El Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	0	0%
NO	40	100%
<b>TOTAL</b>	40	100%

**Tabla 9.** Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil  
**Elaborado por:** Investigador



**Figura 15.** Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil  
**Elaborado por:** Investigador

**Análisis**

De los 40 encuestados, el 100% de los docentes UPE indican que el Sistema Integrado de la UTA NO permite obtener el Índice de Deserción Estudiantil.

**Interpretación**

El Sistema Integrado de la Universidad NO dispone de un Aplicación de la cual se obtenga el Índice de Deserción Estudiantil.

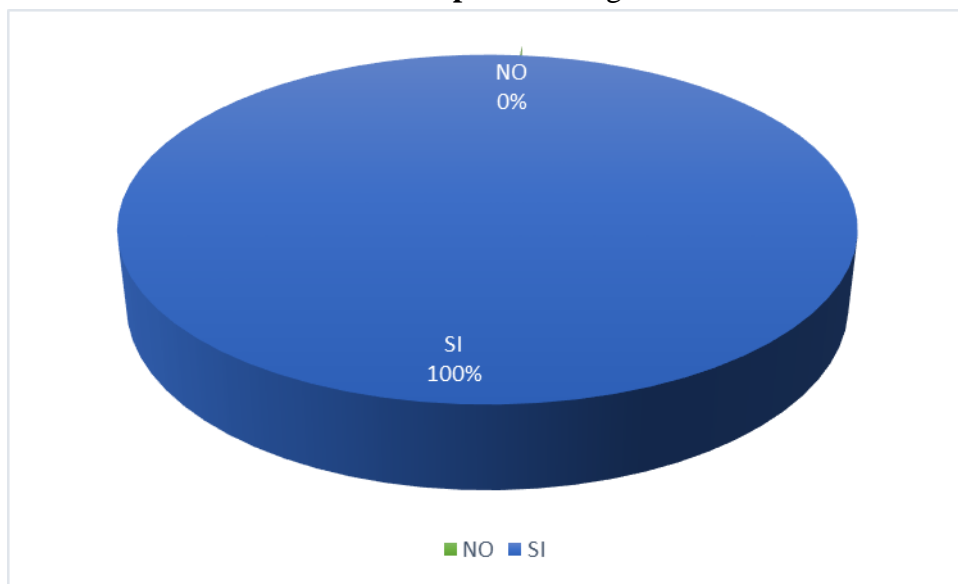
**Pregunta 6:**

¿Es importante para la institución identificar los Patrones de Deserción Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	40	100%
NO	0	0%
<b>TOTAL</b>	40	100%

**Tabla 10.** Importancia para la institución identificar los Patrones de Deserción Estudiantil

**Elaborado por:** Investigador



**Figura 16.** Importancia para la institución identificar los Patrones de Deserción Estudiantil

**Elaborado por:** Investigador

**Análisis**

De los 40 encuestados, el 100% de los docentes UPE indican que es importante para la institución el identificar los Patrones de Deserción Estudiantil.

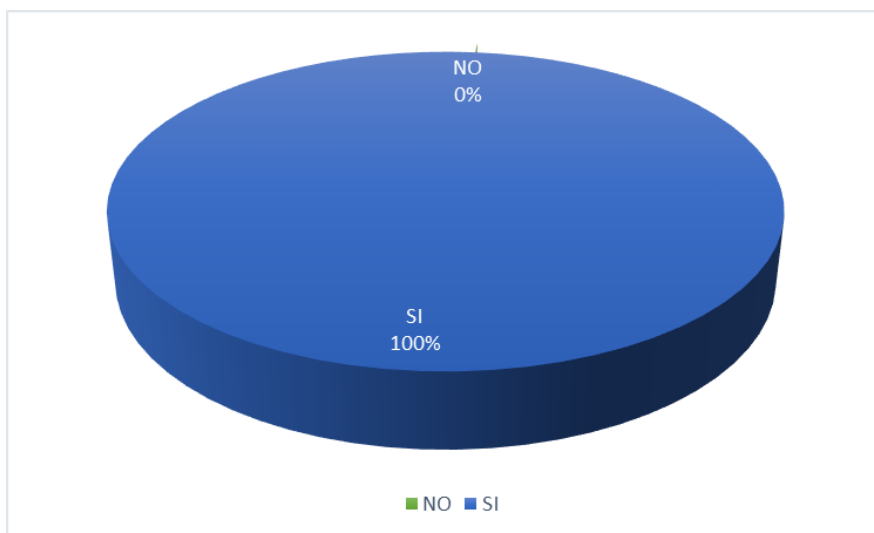
**Interpretación**

La importancia de identificar los Patrones de Deserción Estudiantil en la Universidad Técnica de Ambato, aportaría de manera significativa en la Toma de Decisiones.

**Pregunta 7:** ¿La Detección de Patrones de Deserción Estudiantil permitirá establecer procesos para maximizar la Tasa de Retención Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	40	100%
NO	0	0%
<b>TOTAL</b>	40	100%

**Tabla 11.** Reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil  
**Elaborado por:** Investigador



**Figura 17.** Reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil  
**Elaborado por:** Investigador

### Análisis

De los 40 encuestados, el 100% de los docentes UPE indican Reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil.

## Interpretación

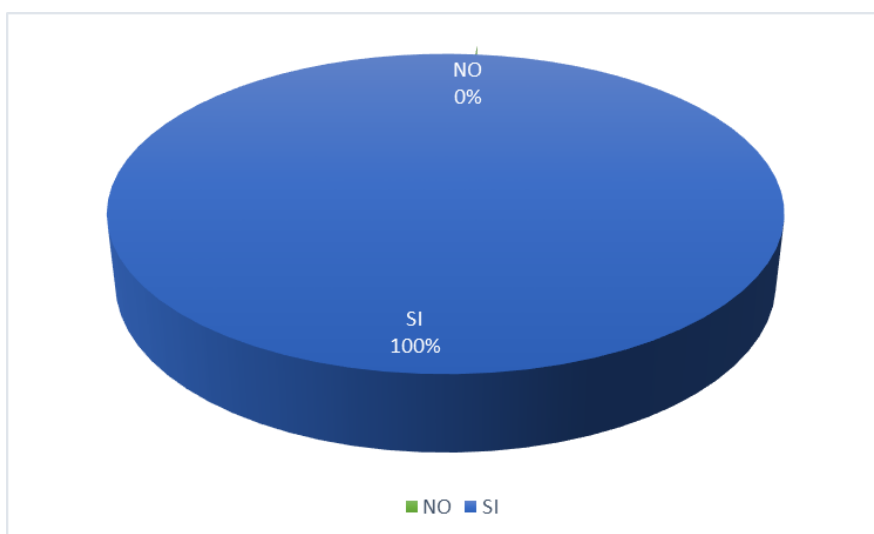
Los Docentes encargados UPE, indican que el identificar los Patrones de Deserción Estudiantil permitirá establecer procesos en bienestar de los estudiantes y maximizará la Tasa de Retención Estudiantil que es un Indicador de Evaluación y Acreditación institucional.

**Pregunta 8:** ¿La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	40	100%
NO	0	0%
<b>TOTAL</b>	40	100%

**Tabla 12.** La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato

**Elaborado por:** Investigador



**Figura 18.** La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato

**Elaborado por:** Investigador



## **Análisis**

De los 40 encuestados, el 100% de los docentes UPE indican Reportes de Patrones de Deserción Estudiantil permitirán una oportuna toma de decisiones por parte de la Autoridades de la Universidad Técnica de Ambato.

## **Interpretación**

Reportes que identifiquen Patrones de Deserción Estudiantil es fundamental en la toma de Decisiones, por parte de las autoridades, que claramente beneficiaran a los estudiantes

### **4.2. Verificación de Hipótesis**

#### **4.2.1. Planteamiento de la Hipótesis**

El análisis de la tasa de retención si incide en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

- **Modelo Lógico**

H<sub>0</sub> El análisis de la tasa de retención **NO** incide en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

H<sub>1</sub> El análisis de la tasa de retención **SI** incide en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

- **Modelo Matemático**

$$H_0: O = E$$

$$H_1: O \neq E$$

- **Modelo Estadístico**

En este trabajo se utilizó la prueba estadística Chi - Cuadrado, la que nos permitirá determinar si existe o no, la relación entre las dos variables.

Para lo cual se utilizó la siguiente fórmula:

$$x^2 = \sum \left[ \frac{(O-E)^2}{E} \right]$$

En donde:

$X^2$  = Chi – Cuadrado

O = Frecuencias Observadas

E = Frecuencias Esperadas

$\Sigma$  = Sumatoria

- **Nivel de Significancia**

$\alpha = 0,05$  Nivel de significancia y con un nivel de confianza del 95 %

Para la verificación de la hipótesis, se seleccionó una pregunta por cada variable para determinar la relación que existe entre las dos, Análisis de la Tasa de Retención como variable independiente y de manera positiva la Detección de Patrones de Deserción Estudiantil como variable dependiente, las preguntas y respuestas seleccionadas fueron:

- **Análisis de la Tasa de Retención**

Pregunta 2: ¿Se realiza un análisis de la Tasa de Retención Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	0	0%
NO	44	100%
<b>TOTAL</b>	<b>44</b>	<b>100%</b>

**Tabla 13.** Evaluación Respuestas Pregunta 2

**Elaborado por:** Investigador

- **Detección de Patrones de Deserción Estudiantil**

Pregunta 7: ¿La Detección de Patrones de Deserción Estudiantil permitirá establecer procesos para maximizar la Tasa de Retención Estudiantil?

ALTERNATIVA	GESTORES DE CAMPO	
	FRECUENCIA	%
SI	40	100%
NO	0	0%
<b>TOTAL</b>	<b>40</b>	<b>100%</b>

**Tabla 14.** Evaluación Pregunta 7

**Elaborado por:** Investigador

#### 4.2.2. Frecuencias Observadas

VARIABLE	PREGUNTAS	FRECUENCIA		
		SI	NO	TOTAL
VI	¿Se realiza un análisis de la Tasa de Retención Estudiantil?	0	40	40
VD	¿La Detección de Patrones de Deserción Estudiantil permitirá establecer procesos para maximizar la Tasa de Retención Estudiantil?	40	0	40
TOTAL		40	40	80

**Tabla 15.** Frecuencias Observadas

**Elaborado por:** Investigador

#### 4.2.3. Frecuencias Esperadas

Para calcular la frecuencia esperada se utiliza la siguiente ecuación:

$$fe = \frac{(total\ de\ la\ fila) * (total\ de\ la\ columna)}{N}$$

V	PREGUNTAS	FRECUENCIA		
		SI	NO	TOTAL
VI	¿Se realiza un análisis de la Tasa de Retención Estudiantil?	$(40*40)/80=20$	$(40*40)/80=20$	40
VD	¿La Detección de Patrones de Deserción Estudiantil permitirá establecer procesos para maximizar la Tasa de Retención Estudiantil?	$(40*40)/80=20$	$(40*40)/80=20$	40
TOTAL		40	40	80

**Tabla 16.** Frecuencias Observadas

**Elaborado por:** Investigador

#### 4.2.4. Prueba Chi – Cuadrado

ALTERNATIVAS		CÁLCULOS			
		O	E	(O-E) <sup>2</sup>	(O-E) <sup>2</sup> /E
Pregunta 2	SI	0	20	400	20
	NO	40	20	400	20
Pregunta 7	SI	40	20	400	20
	NO	0	20	400	20
$\chi^2_c =$					80

**Tabla 17.** Prueba Chi – Cuadrado

**Elaborado por:** Investigador

#### 4.2.5. Grados de Libertad

$$gl = (f - 1)(c - 1)$$

$$gl = (3 - 1)(2 - 1)$$

$$gl = 1 * 1$$

$$gl = 1$$

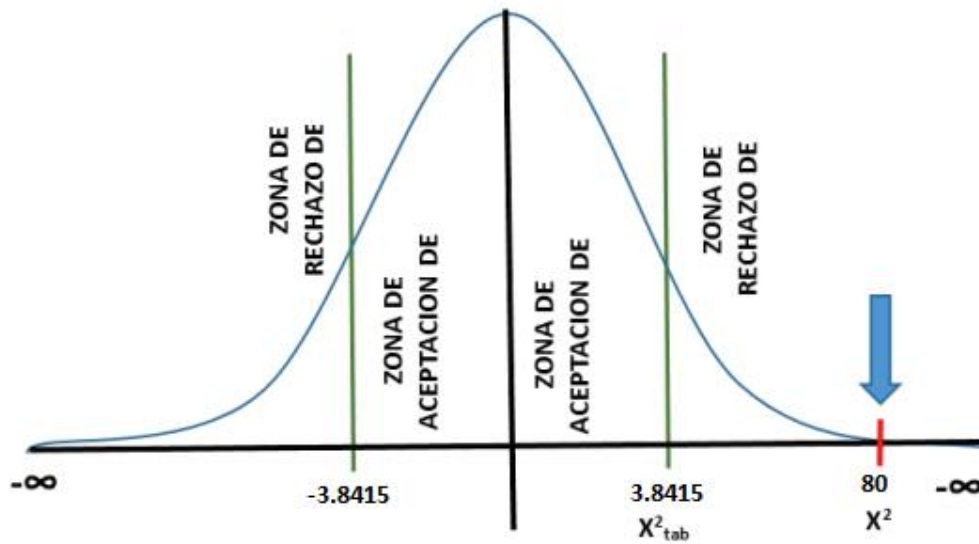
Según la tabla de distribución Chi – Cuadrado y los grados de libertad el valor obtenido para es: **3,8415**

<b>TABLA 3-Distribución Chi Cuadrado <math>\chi^2</math></b>										
P = Probabilidad de encontrar un valor mayor o igual que el chi cuadrado tabulado, v = Grados de Li										
v/p	0,001	0,0025	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25
1	10,8274	9,1404	7,8794	6,6349	5,0239	<b>3,8415</b>	2,7055	2,0722	1,6424	1,3233
2	13,8150	11,9827	10,5965	9,2104	7,3778	5,9915	4,6052	3,7942	3,2189	2,7726

**Figura 19.** Distribución Chi – Cuadrado

**Fuente:** [http://labrad.fisica.edu.uy/docs/tabla\\_chi\\_cuadrado.pdf](http://labrad.fisica.edu.uy/docs/tabla_chi_cuadrado.pdf)

#### 4.2.6. Decisión estadística



**Figura 20.** Prueba Chi- Cuadrado

**Elaborado:** Investigador

$\chi^2_{tab}$       Corresponde al valor obtenido de la tabla de distribución de Chi - Cuadrado

$\chi^2$           Corresponde al valor obtenido del cálculo de Chi – Cuadrado

Si  $\chi^2 > \chi^2_{tab}$  se rechaza  $H_0$  y se acepta  $H_1$ , por lo tanto  $80 > 3.8415$

Entonces el análisis de la tasa de retención **SI** incide en la detección de patrones de deserción estudiantil en la Universidad Técnica de Ambato.

## **CAPÍTULO V**

### **CONCLUSIONES Y RECOMENDACIONES**

#### **5.1 Conclusiones**

Una vez realizado el análisis de la encuesta aplicada a los docentes responsables UPE de cada una de las carreras de la Universidad Técnica de Ambato, con el fin de conocer el problema de la presente investigación, se concluye que:

- La tasa de Retención Estudiantil es un indicador de gran relevancia para la Universidad Técnica de Ambato.
- El análisis de la Tasa de Retención permitirá identificar si este indicador ha maximizado o ha minimizado en cada período académico.
- La detección de Patrones de Deserción Estudiantil permitirá a las Autoridades de la Universidad Técnica de Ambato en la toma de decisiones.
- La Detección de Patrones de Deserción Estudiantil permitirá que las autoridades creen estrategias para mitigar los factores de Deserción Estudiantil.

## 5.2. Recomendaciones

- Identificados los Patrones de Deserción Estudiantil, tales como: Facultad, Carrera, Semestre, Materia, Docente, Período Académico, Género, Etnia, Estado Civil.
- Crear un Almacén de Datos que permita realizar el proceso de extracción, transformación y carga de datos.
- Identificar Herramientas de Software Libre que permitan aplicar Técnicas de Minería de Datos.
- Diseñar un Modelo de predicción basado en técnicas de minería de datos(clouster) para predecir una posible Deserción Estudiantil



## **CAPÍTULO VI**

### **LA PROPUESTA**

#### **6.1. Datos Informativos**

##### **6.1.1. Título**

“DATAMINING PARA LA DETECCIÓN DE PATRONES DE DESERCIÓN ESTUDIANTIL EN LA UNIVERSIDAD TÉCNICA DE AMBATO”

##### **6.1.2. Institución Ejecutora**

UNIVERSIDAD TÉCNICA DE AMBATO

##### **6.1.3. Beneficiarios**

- Comunidad Universitaria

##### **6.1.4. Ubicación**

- Provincia: Tungurahua.
- Cantón: Ambato.

##### **6.1.5. Equipo Técnico Responsable**

- Investigador: Cristina Alexandra Frutos Ortega.

## 6.2. Antecedentes de la Propuesta

- “La minería de datos como un método innovador para la detección de patrones de deserción estudiantil en programas de pregrado en instituciones de educación superior” (Timarán & Calderón, 2013); en esta investigación se realiza a partir de datos:

- ✓ Socioeconómicos
- ✓ Académicos
- ✓ Disciplinarios
- ✓ Institucionales

Basándose en el proceso de descubrimiento de conocimiento en bases de datos, Creando un repositorio de datos para la limpieza de la información.

Aplicando la técnica de clasificación basada en árboles de decisión usando R, con los resultados esperados

## 6.3. Justificación

## 6.4. Objetivos

### Objetivo General

Aplicar técnicas de Minería de Datos para la Detección de Deserción Estudiantil de la Universidad Técnica de Ambato.

### Objetivos Específicos

- Seleccionar información relevante de los estudiantes desertores de la Universidad Técnica de Ambato.
- Extraer, Transformar y Cargar información del almacén de datos.
- Aplicar Técnicas de Minería de Datos para detectar patrones de deserción estudiantil de la Universidad Técnica de Ambato.

## **6.5. Análisis de Factibilidad**

La propuesta planteada es factible ya que la Universidad Técnica de Ambato cuenta con los elementos necesarios para desarrollar la investigación propuesta:

### **6.5.1. Factibilidad Técnica**

- La Universidad Técnica de Ambato cuenta con el Gestor de Base de Datos SQL Server 2008 con sus respectivas Licencias.
- La herramienta Integration Services
- Weka Software libre para la minería de datos

### **6.5.2. Factibilidad Organizacional**

Puesto que para la Detección de Patrones de Deserción Estudiantil en la Universidad Técnica de Ambato se necesita de información del Sistema Integrado de la Institución, se cuenta con la aprobación y disponibilidad del personal de la Dirección de Comunicación y Tecnología a cargo de la información.

### **6.5.3. Factibilidad Económica**

La Universidad Técnica de Ambato cuenta con los recursos tecnológicos necesarios, y el desarrollo por parte del investigador cubrirá con los gastos relacionados para el desarrollo de la solución.

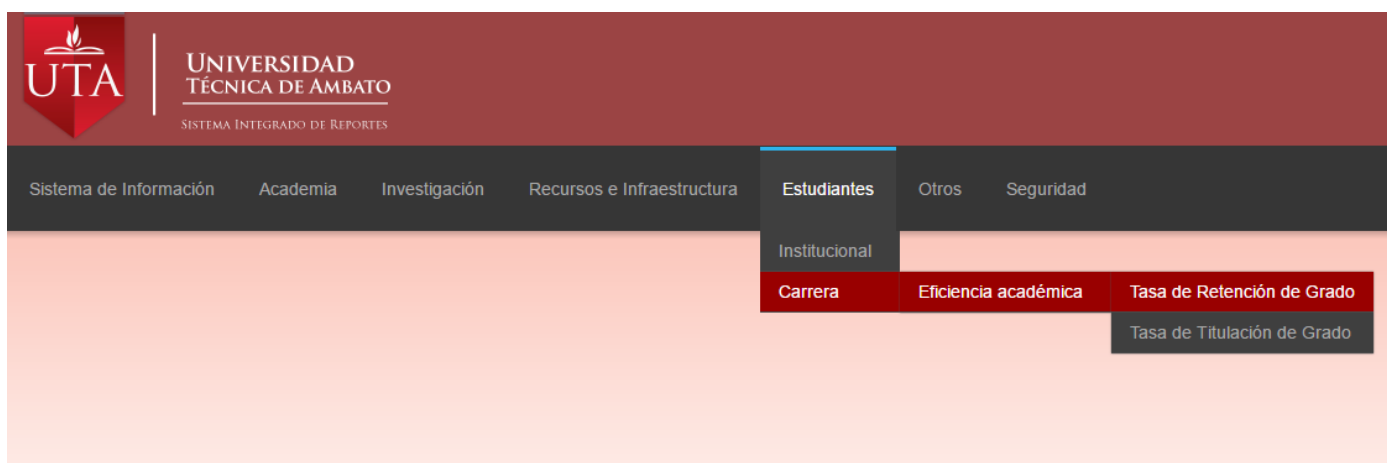
### **6.5.4. Factibilidad Operativa**

Se cuenta con el recurso humano de desarrollo, puesto que el investigador pertenece a la Institución.

## **6.6 Fundamentación**

### 6.6.1. Análisis Tasa de Retención

El Sistema Integrado de la Universidad Técnica de Ambato permite obtener este indicador automáticamente como un reporte, en el menú Estudiantes, submenú Carrera, opción Tasa de Retención de Grado como se muestra en la siguiente figura.



**Figura 21.**Sistema Integrado de Reportes

**Elaborado:** Investigador

Como se puede observar en la Figura 22, se debe seleccionar la facultad, la carrera, período inicio estudios y retenidos al período, de la siguiente manera:

Como muestra se toma las carreras de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial

Carrera: Ingeniería en Sistemas Computacionales e Informáticos

Período Inicio de Estudios: Abril – Septiembre 2014

Retenidos al Período: Abril – Septiembre 2016

**Tasa de Retención Grado** Parametrización

Facultad \*

Carrera \*

Periodo Inicio Estudios \*

Retenidos al periodo: \*

\* Campos obligatorios

**Figura 22.** Tasa de Retención Selección de Datos

**Elaborado:** Investigador

La primera pestaña permite visualizar el listado de estudiantes pertenecientes al período de inicio seleccionado. Ver Figura 23

N° de Estudiantes Admitidos en el Periodo (NTEA)		N° de Estudiantes Retenidos al Periodo (NEMA)		Fórmula de Cálculo							
Periodo Académico	Apellido 1	Apellido 2	Nombres	Identificación	Género	Especialidad	Matrícula	Folio	Creditos Aprobados	Modalidad	Certificado
ABR/14-SEP/14	ALTAMIRANO	GUERRA	MARCOS ANTONIO	1850361773	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0219	0110	24	PRESENCIAL	
ABR/14-SEP/14	VILLA	VEGA	VIVIANA VERONICA	1805371455	FEMENINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0194	0097	13	PRESENCIAL	
ABR/14-SEP/14	ALCIVAR	VARGAS	MARCOS ANDRES	1723451132	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0210	0105	27	PRESENCIAL	
ABR/14-SEP/14	CRUZ	IZA	EDISON WLADIMIR	0503852030	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0209	0105	0	PRESENCIAL	
ABR/14-SEP/14	SAILEMA	INTRIAGO	ERICK GERMAN	1803676111	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0232	0116	6	PRESENCIAL	
ABR/14-SEP/14	TORRES	TIGSELEMA	DIEGO ARMANDO	1803792553	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0193	0097	16	PRESENCIAL	
ABR/14-SEP/14	JARRIN	NUÑEZ	ALEXIS ROBERTO	1804370466	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0213	0107	23	PRESENCIAL	
ABR/14-SEP/14	BARROS	MANZANO	IVETH CRISTINA	1804783619	FEMENINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0201	0101	24	PRESENCIAL	
ABR/14-SEP/14	BARROS	MANZANO	MARILYN VALERIA	1804783627	FEMENINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0199	0100	23	PRESENCIAL	
ABR/14-SEP/14	LLERENA	MANOBANDA	EDGAR MIGUEL	1804788121	MASCULINO	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0229	0115	10	PRESENCIAL	

Ir a  Ver

[1 a 10 de 41]

**Figura 23.** Estudiantes del Período Inicio de Estudios

**Elaborado:** Investigador

La segunda pestaña permite visualizar el listado de estudiantes retenidos según los períodos seleccionados. Ver Figura 24

N° de Estudiantes Admitidos en el Período (NTEA)		N° de Estudiantes Retenidos al Período (NEMA)			Fórmula de Cálculo						
Período Académico	Carrera	Identificación	Apellido 1	Apellido 2	Nombres	Género	Matricula	Folio	Modalidad	Créditos Aprobados	Certificado
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0503594285	CORNEJO	JIMENEZ	EDISON MAURICIO	MASCULINO	0103	0052	PRESENCIAL	80	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0504138785	SANDOVAL	UNAPUCHA	MARCELO DANIEL	MASCULINO	0062	0031	PRESENCIAL	125	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	0550063986	QUIMBITA	TERAN	ERIKA TATIANA	FEMENINO	0139	0070	PRESENCIAL	92	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1803748647	SHULCA	ANDRADE	JONATHAN JAVIER	MASCULINO	0079	0040	PRESENCIAL	118	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1803792553	TORRES	TIGSELEMA	DIEGO ARMANDO	MASCULINO	0269	0135	PRESENCIAL	53	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1804336707	GAVILANEZ	MARIÑO	RICARDO FABRICIO	MASCULINO	0088	0044	PRESENCIAL	113	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1804348009	YAGUARGOS	CASTRO	SANTIAGO DAMIAN	MASCULINO	0070	0035	PRESENCIAL	125	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1804363552	MOYA	FREIRE	JOSE ANDRES	MASCULINO	0083	0042	PRESENCIAL	118	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1804370466	JARRIN	NUÑEZ	ALEXIS ROBERTO	MASCULINO	0091	0046	PRESENCIAL	118	
ABR/16-SEP/16	ING. EN SISTEMAS COMPUTAC.E INFORMATICOS	1804454815	ULLOA	LOPEZ	LENIN PAUL	MASCULINO	0145	0073	PRESENCIAL	65	

Ir a 1 Ver 10 [1 a 10 de 27]

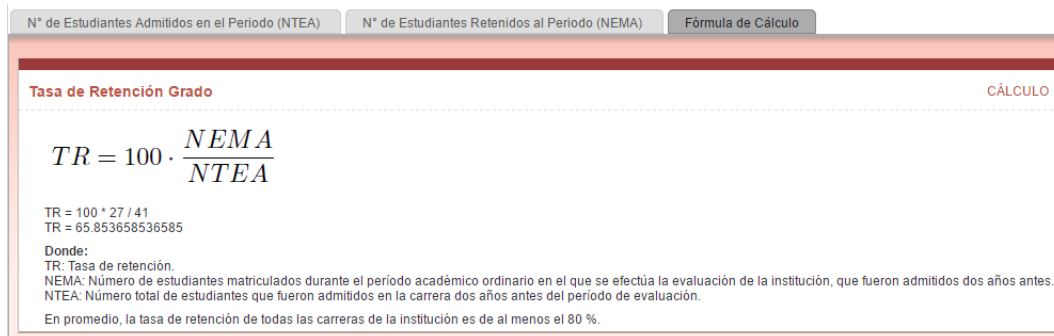
**Figura 24.** Estudiantes Retenidos

**Elaborado:** Investigador

La siguiente imagen nos muestra la fórmula de Tasa de Retención de Estudiantes, en donde

**NEMA:** Representa el número de estudiantes matriculados durante el período académico ordinario en el que se efectúa la evaluación de la institución, que fueron admitidos dos años antes.

**NTEA:** Representa el número total de estudiantes que fueron admitidos en la carrera dos años antes del período de evaluación.



**Figura 25.** Formula Tasa de Retención

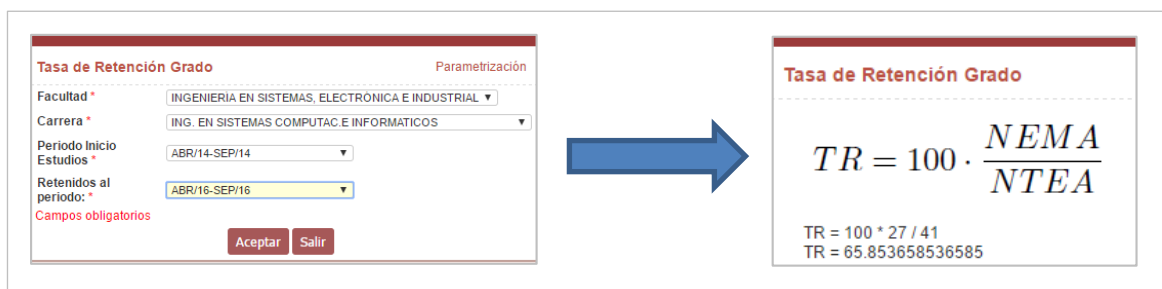
**Elaborado:** Investigador

## RETENCIÓN ESTUDIANTIL

Con el fin de mostrar el funcionamiento de la aplicación se tomará como ejemplo las carreras de la Facultad de Ingeniería en Sistemas, electrónica e Industrial.

## CARRERA: INGENIERÍA EN SISTEMAS COMPUTACIONALES E INFORMÁTICOS

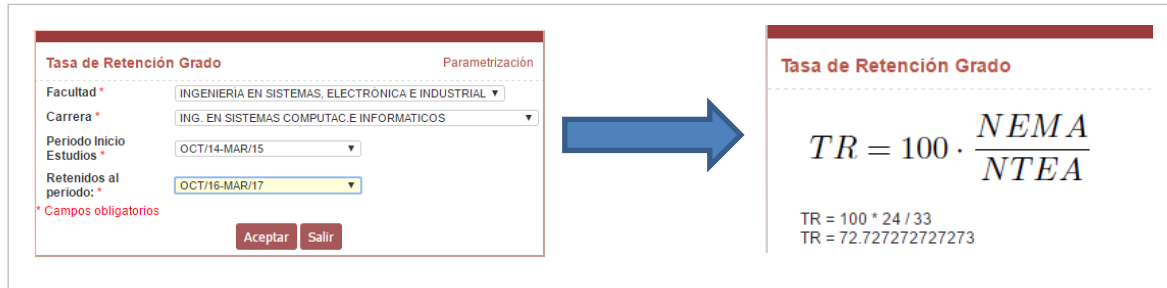
En la figura 26 se muestran el reporte obtenido de la aplicación de la fórmula de Tasa de Retención estudiantil correspondientes a la carrera de Ingeniería en Sistemas Computacionales e Informáticos en el período Abril – Septiembre 2016.



**Figura 26.** Tasa de Retención Sistemas período Abril – Septiembre 2016

**Elaborado:** Investigador

En la figura 27 se muestran la Tasa de Retención estudiantil correspondientes a la carrera de Ingeniería en Sistemas Computacionales e Informáticos en el período Octubre 2016 – Marzo 2017



**Figura 27.** Tasa de Retención Sistemas período Octubre 2016 - Marzo 2017

**Elaborado:** Investigador

### **CARRERA: INGENIERÍA ELECTRÓNICA Y COMUNICACIONES**

En la figura 28 se muestran el reporte obtenidos de la aplicación de la fórmula de Tasa de Retención estudiantil correspondientes a la carrera de Ingeniería Electrónica y Comunicaciones en el período Abril – Septiembre 2016.

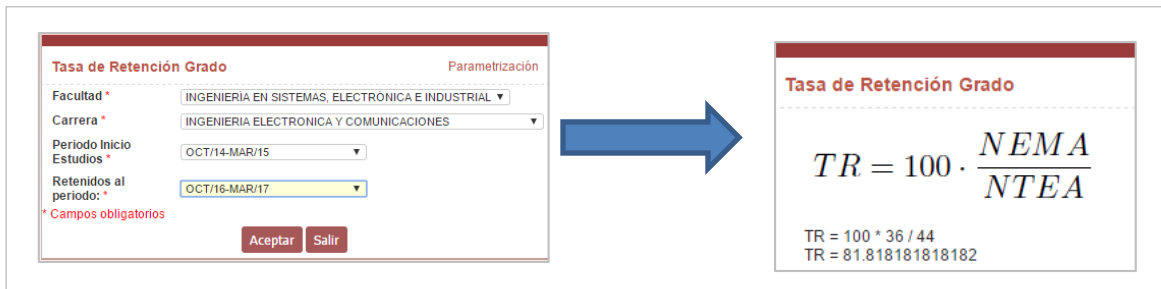


**Figura 28.** Tasa de Retención Electrónica período Abril - Septiembre 2016

**Elaborado:** Investigador



En la figura 29 se muestran la Tasa de Retención estudiantil correspondientes a la carrera de Ingeniería Electrónica y comunicaciones en el período Octubre 2016 – Marzo 2017

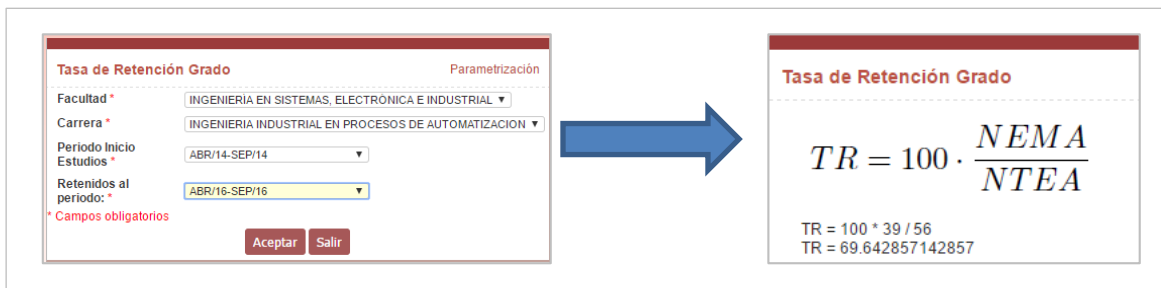


**Figura 29.** Tasa de Retención Electrónica período Octubre 2016 - Marzo 2017

**Elaborado:** Investigador

### **CARRERA: INGENIERÍA INDUSTRIAL Y PROCESOS DE COMUNICACIÓN**

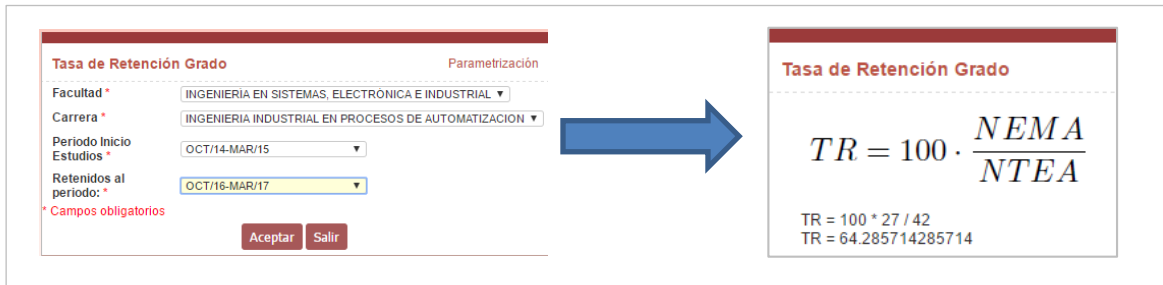
En la figura 30 se muestran el reporte obtenidos de la aplicación de la fórmula de Tasa de Retención estudiantil correspondientes a la carrera de Ingeniería Industrial y Procesos de Automatización en el períodos Abril – Septiembre 2016.



**Figura 30.** Tasa de Retención Industrial período Octubre 2016 - Marzo 2017

**Elaborado:** Investigador

En la figura 31 se muestran la Tasa de Retención estudiantil correspondientes a la carrera de Ingeniería Industrial y Procesos de Automatización en el período Octubre 2016 – Marzo 2017



**Figura 31.** Tasa de Retención Industrial período Octubre 2016 - Marzo 2017

**Elaborado:** Investigador

### 6.6.2 KDD (Knowledge Discovery in Databases)

Descubrimiento de Conocimiento en Base de Datos, es un nivel abstracto, el campo KDD se ocupa del desarrollo de métodos y técnicas para dar sentido a los datos. El problema básico abordado por el proceso KDD es el de cartografiar datos de bajo nivel (que suelen ser demasiado voluminosos para comprender y digerir fácilmente) en otras formas que podrían ser más compactas.

Minería de Datos y KDD Históricamente, la noción de encontrar patrones útiles en datos se ha dado una variedad de nombres, incluyendo minería de datos, extracción de conocimiento, descubrimiento de información, recolección de información, arqueología de datos y procesamiento de patrones de datos. El término minería de datos ha sido utilizado principalmente por estadísticos, analistas de datos y las comunidades de sistemas de información de gestión (SIG). También ha ganado popularidad en el campo de base de datos.

La frase descubrimiento de conocimiento en bases de datos fue acuñada en el primer taller de KDD en 1989 (Piatetsky-Shapiro 1991) para enfatizar que el conocimiento es el producto final de un descubrimiento basado en datos. Se ha popularizado en los campos

de la Inteligencia Artificial y del aprendizaje automático. (Fayyad & Piatetsky-Shapiro & Smyth, 1996)

- **Etapa de Selección de datos**

El objetivo de esta etapa es obtener las fuentes internas y externas de datos que sirven de base para el proceso de minería de datos. (Timarán & Calderón, 2013).

- **Etapa de procesamiento de datos**

Esta etapa es de larga duración y consume gran cantidad de recursos, por ende, requiere de mayor tiempo en el desarrollo de un proyecto de minería de datos. Aproximadamente el 70% o más de la duración estimada de un proyecto, se gasta en el análisis de datos. (Medina, 2011).

El objetivo de esta etapa es obtener datos limpios, sin valores nulos o anómalos que permitan obtener patrones de calidad. (Timarán & Calderón, 2013).

- **Etapa de Transformación de datos**

El objetivo de esta fase es transformar la fuente de datos en un conjunto listo para aplicar las técnicas de minería de datos.

- **Etapa de Minería de Datos**

El objetivo de esta etapa es la búsqueda y descubrimiento de patrones de interés aplicando técnicas de minería de datos para el descubrimiento de nuevo conocimiento, tales como clasificación, etc.

- **Etapa de Interpretación/ Evaluación de datos**

Etapa en la cual se interpretan y evalúan los patrones descubiertos con el propósito de consolidar el conocimiento descubierto para su posterior incorporación en otro sistema para confrontarlo con conocimiento descubierto con anterioridad. (Timarán & Calderón, 2013).

Una visión general de los pasos que componen el proceso KDD.

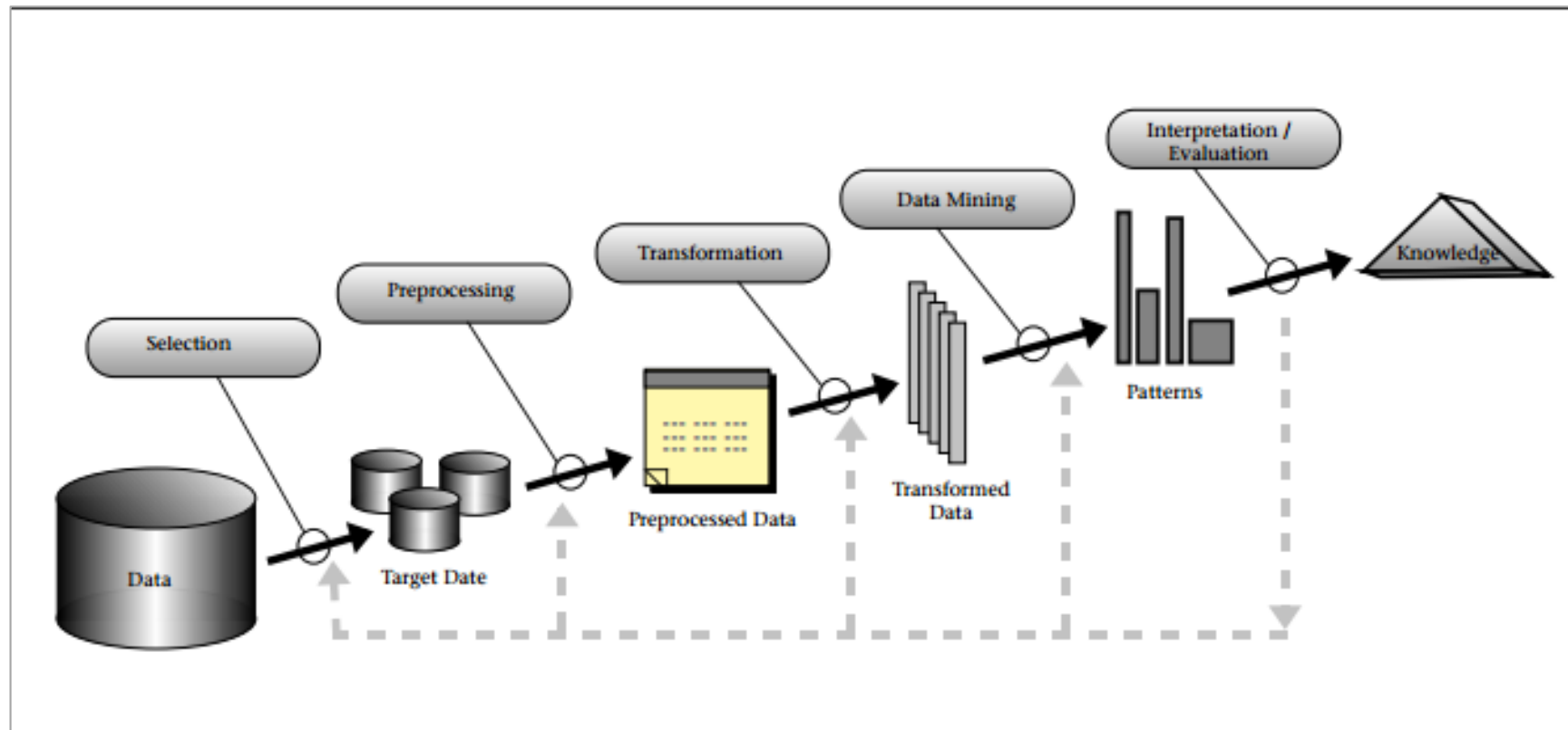


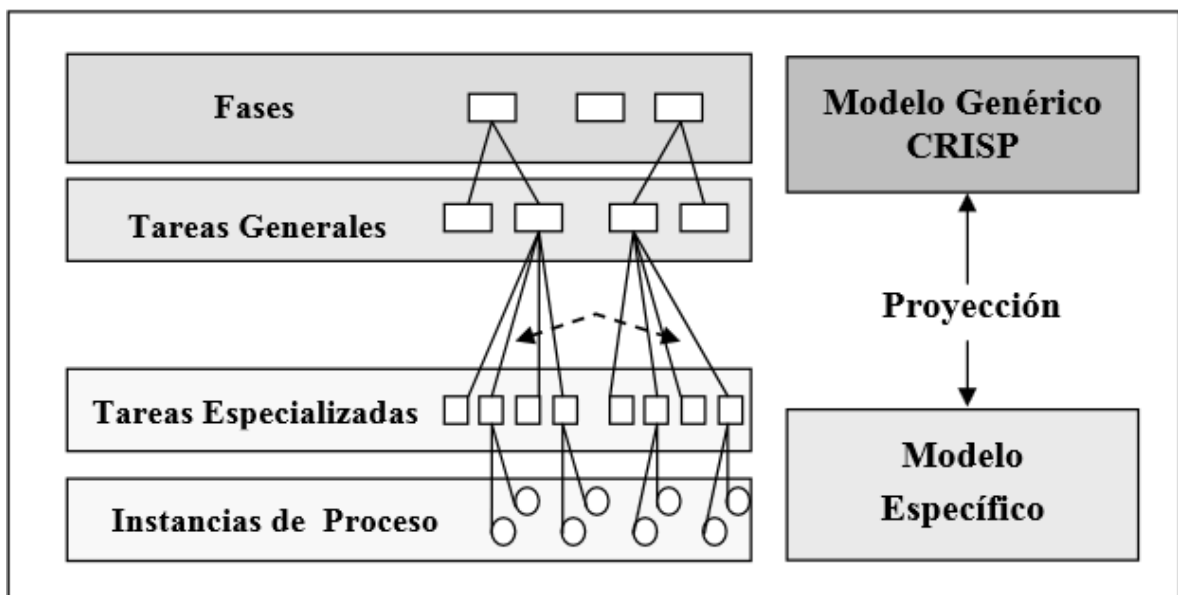
Figura 32. Pasos que componen el proceso KDD.

Fuente: (Fayyad & Piatetsky-Shapiro & Smyth, 1996)

### 6.6.3. CRISP-DM

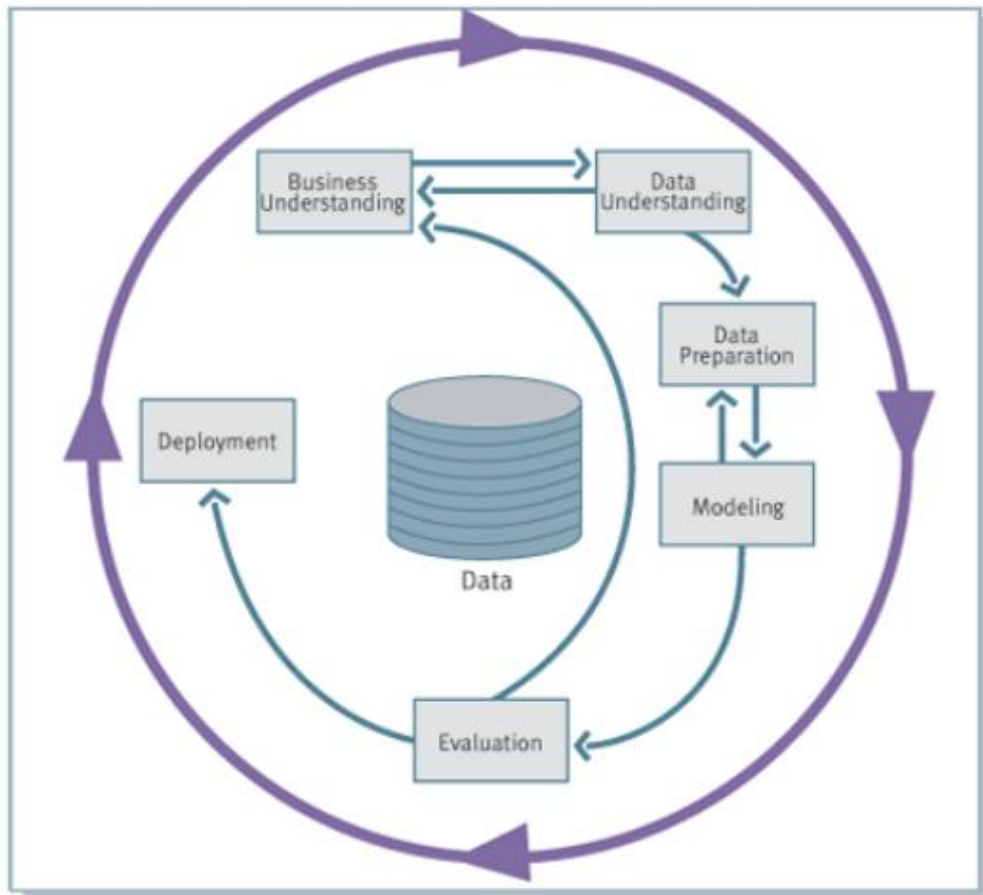
La Metodología CRISP-DM (CRoss-Industry Standard Processfor Data Mining), ha sido diseñada como una metodología neutra respecto a la herramienta que se utilice para el desarrollo del proyecto de explotación de datos, cuentan con un nivel de detalle mayor, describen las tareas en cada fase, utiliza un proceso jerárquico formado por varias tareas que ofrece a las organizaciones la estructura necesaria para obtener resultados rápidos y eficientes, además mantiene como foco central los objetivos empresariales del proyecto, razón por la cual CRISP-DM comienza realizando un análisis del problema empresarial para su transformación en un problema técnico de minería de datos. (Cobos, Zuñiga, Guarín, León, & Mendoza, 2010)

CRISP-DM, está dividida en 4 niveles de abstracción organizados de forma jerárquica (Figura N° 26) y organiza el desarrollo de un proyecto de Data Mining, en una serie de seis fases algunas de las cuales son bidireccionales, es decir que de una fase en concreto se puede retornar a la fase anterior para poder revisar la(Imagen N° 9 y N° 10): (Gallardo Arancibia, 2009)



**Figura 33.** Esquema de los 4 niveles de CRISP-DM

**Fuente:** (Gallardo Arancibia, 2009)

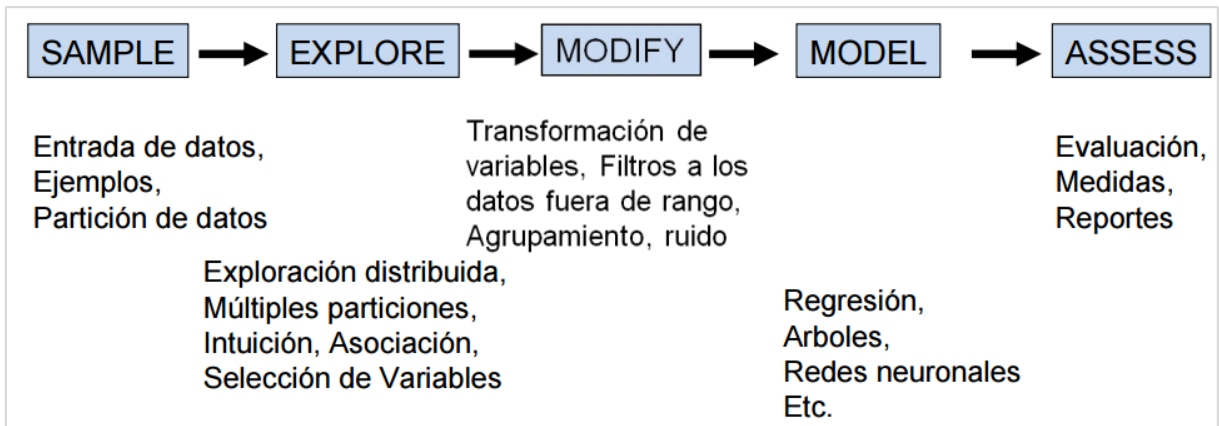


**Figura 34:** Modelo de proceso CRISP–DM ([CRISP-DM, 2000]).

**Fuente:** (Gallardo Arancibia, 2009)

#### 6.6.4. SEMMA

SEMMA es el acrónimo a las cinco fases: (Sample, Explore, Modify, Model, Assess) La metodología es propuesta por SAS Institute Inc, la define como: “... proceso de selección, exploración y modelamiento de grandes cantidades de datos para descubrir patrones de negocios desconocidos...”[2].



**Figura 35. SEMMA**

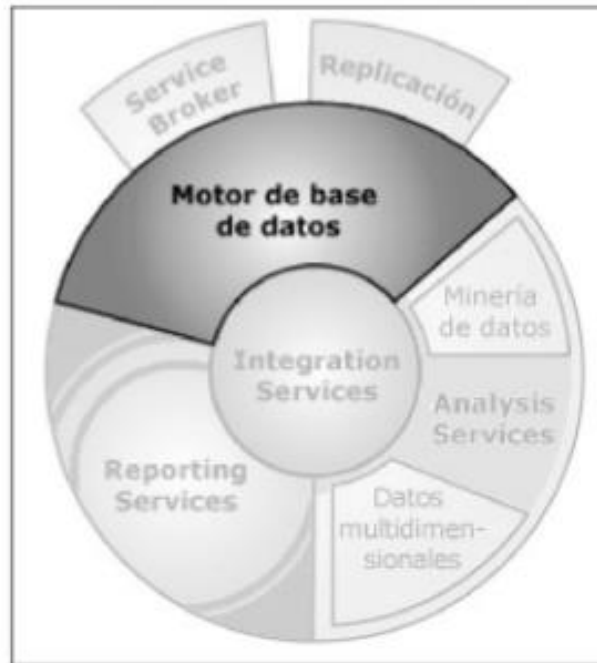
**Fuente:** <http://www.crisp-dm.org/CRISPWP-0800.pdf>

### 6.6.5. Herramientas

#### SQL- Server

Es el sistema de Base de Datos profesional de Microsoft. Contiene una variedad de características y herramientas que se pueden utilizar para desarrollar y administrar bases de datos y soluciones de todo tipo basados en ellas.

La figura N° 29 muestra los componentes de SQL server 2008 R.



**Figura 36.** SQL Server 2008

**Fuente:**(Pérez, 2011)

El motor de bases de datos es el servicio principal para almacenar, procesar y proteger los datos. Asimismo proporciona acceso controlado y procesamiento rápido de transacciones de base de datos más exigentes. (Pérez, 2011)

### **Integration Services**

Es una plataforma para generar soluciones de integración de datos de alto rendimiento, lo que incluye paquetes que proporcionan procesamientos de extracción, transformación y carga (ETL) para almacenamiento de datos.

A través de Integration Services, SQL Server 2008 proporciona una nueva capacidad de Business Intelligence de autoservicio a través de la interacción de productos y tecnologías de Sharepoint. (Pérez, 2011)



## WEKA

WEKA (Waikato Environment for Knowledge Analysis) es una herramienta que permite la experimentación de análisis de datos mediante la aplicación, análisis y evaluación de las técnicas más relevantes de análisis de datos, principalmente las provenientes del aprendizaje automático, sobre cualquier conjunto de datos del usuario. (Jiménez & Sierra, 2010)

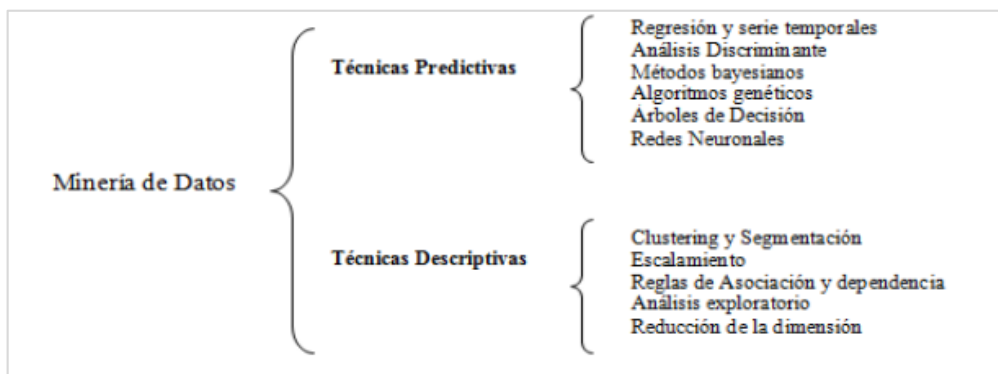
WEKA se distribuye como software de libre distribución desarrollado en Java. Está constituido por una serie de paquetes de código abierto con diferentes técnicas de preprocesado, clasificación, agrupamiento, asociación, y visualización, así como facilidades para su aplicación y análisis de prestaciones cuando son aplicadas a los datos de entrada seleccionados. (Corso, 2009)

### 6.6.6. Técnicas de Minería de Datos

Las técnicas de la minería de datos provienen de la Inteligencia artificial y de la estadística, dichas técnicas, no son más que algoritmos, más o menos sofisticados que se aplican sobre un conjunto de datos para obtener resultados.

En la fase de minería de datos, se decide cuál es la tarea a realizar (clasificar, agrupar etc.) y se elige la técnica descriptiva o predictiva que se va a utilizar.

En la figura N° 30 se observa las principales Técnicas de Minería de Datos



**Figura 37.** Técnicas de Minería de Datos

**Fuente:** (Corso, 2009)

## **REDES NEURONALES**

Se trata de un sistema de interconexión de neuronas en una red que colabora para producir un estímulo de salida. Algunos ejemplos de red neuronal son: El Perceptrón, El Perceptrón multicapa. Los Mapas Autoorganizados, también conocidos como redes de Kohonen.

## **ÁRBOLES DE DECISIÓN**

Un árbol de decisión es un modelo de predicción utilizado en el ámbito de la inteligencia artificial, dada una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.

Ejemplos:

- ✓ Algoritmo ID3
- ✓ Algoritmo C4.5.

## **AGRUPAMIENTO O CLUSTERING**

Es un procedimiento de agrupación de una serie de ítems según criterios habitualmente de distancia; se tratará de disponer los vectores de entrada de forma que estén más cercanos aquellos que tengan características comunes.

Ejemplos:

- ✓ Algoritmo K-means
- ✓ Algoritmo K-medoids.

(Corso, 2009)

## 6.7. Metodología

### 6.7.1. Metodología para procesos de Minería de Datos

KDD	CRISP-DM	SEMMMA
Selección	Conocimiento del Negocio	Muestra
Preprocesamiento	Preparación de Datos	Exploración
Transformación		Modificación
Minería de Datos	Modelamiento	Modelo
Interpretación y Evaluación	Evaluación y Difusión	Evaluación

**Tabla 18.** Metodología para procesos de Minería de Datos

**Elaborado:** Investigador

### 6.7.2 Selección de Metodología

En base a la Tabla 18. Metodología para procesos de Minería de Datos y considerando aspectos como la complejidad del proyecto y experiencia del investigador se decidió utilizar la metodología KDD.

### 6.7.3. Aplicación de Metodología

### 6.7.3.1. Selección de Datos

El origen de Datos es el UTAMATICO que se encuentra en el Gestor de Base de Datos SQL-Server 2008, base de datos que contiene datos académicos y personales de los estudiantes de cada una de las carreras de la Universidad Técnica de Ambato.

A continuación se detalla las vistas necesarias para detectar los estudiantes desertores de la Universidad Técnica de Ambato:

VISTA ESTUDIATES	
CEDULA	Campo que permitirá identificar el estudiante.
GENERO	Campo que especifica el género del estudiante.
ETNIA	Campo que especifica la etnia del estudiante.
DISCAPACIDAD	Campo que especifica si el estudiante posee alguna discapacidad.
FECHA_NACIMIENTO	Campo que identifica la Fecha de Nacimiento del estudiante.
ESTADO_CIVIL	Campo que identifica el Estado Civil del estudiante.
TITULO_COLEGIO	Campo que identifica el Título del estudiante obtenido en el colegio.
ESPECIALIDAD_COLEGIO	Campo que identifica la Especialidad del Colegio en el que se graduó el estudiante
PROMEDIO_COLEGIO	Campo que identifica la Nota de Graduación del Estudiante.
COLECODIGO	Campo que permite obtener detalles del Colegio en el que se graduó el Estudiante.

**Tabla 19. VISTA ESTUDIATES**

**Elaborado:** Investigador

VISTA COLEGIOS	
COLECODIGO	Campo que identifica al colegio.
COLEGIO	Campo que contiene el nombre del colegio.
CIUDAD	Campo que contiene el nombre de la ciudad del colegio.
CANTON	Campo que contiene el catón del colegio.
PROVINCIA	Campo que contiene la provincia del colegio.

**Tabla 20. VISTA COLEGIOS**

**Elaborado:** Investigador

<b>VISTA ESPECIALIDAD</b>	
ESPCODIGO	Campo que identifica la Carrera.
CARRERA	Campo que contiene el nombre de la Carrera.
FACULTAD	Campo que contiene el nombre de la Facultad a la que pertenece la Carrera.

**Tabla 21. VISTA ESPECIALIDAD**

**Elaborado:** Investigador

<b>VISTA MATERIA</b>	
ID_MATERIA	Campo que identifica la Materia.
MATERIA	Campo que identifica la el nombre de la materia.
HORAS	Campo que identifica el número de horas de la materia.

**Tabla 22. VISTA MATERIAS**

**Elaborado:** Investigador

<b>VISTA PROFESOR</b>	
PROCEDULA	Campo que identifica el profesor.
PROFESOR	Campo que identifica el nombre del profesor.

**Tabla 23. VISTA PROFESOR**

**Elaborado:** Investigador

<b>VISTA PERÍODO</b>	
ID_PERIOD	Campo que identifica la el período académico.
PERIODO	Campo que identifica el nombre del período.
AÑO	Campo que identifica el año según la fecha de inicio del período.

**Tabla 24. VISTA PERIODO**

**Elaborado:** Investigador

<b>VISTA DESERCIÓN</b>	
ID_DESERTOR	Campo que identifica el id de deserción.
PERIODO	Campo que identifica el período de deserción.
HDESTCODIGO	Campo que identifica la cedula del estudiante.
HDNUMVECES	Campo que identifica el número de veces que el estudiante a tomado la materia que deserta
HDASISTENCIA	Campo que identifica la asistencia del estudiante.
HDARRASTRE	Campo que identifica si la materia perdida es un arrastre.
HDPROFECEDULA	Campo que identifica el profesor de la materia.
HDESPECIALIDAD	Campo que identifica la carrera de deserción.
HDMATCODIGO	Campo que identifica la Materia.
HDCURCODIGO	Campo que identifica el curso.
HFECHAMATRICULA	Campo que identifica la fecha de matrícula

**Tabla 25.** VISTA DESERCIÓN

**Elaborado:** Investigador

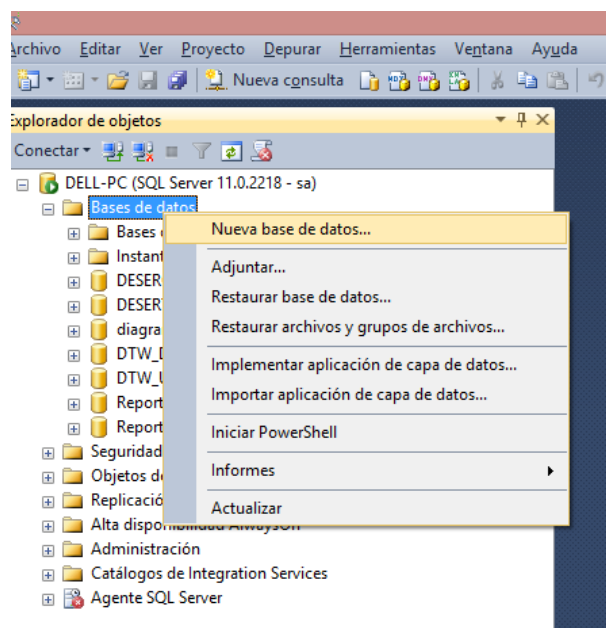
### 6.7.3.2. PREPROSESAMIENTO

En esta etapa se debe obtener un Almacén de datos limpios y sin valores nulos.

Para lo cual realizaremos el proceso del ETL.

#### 1. Creación de la Base de Datos

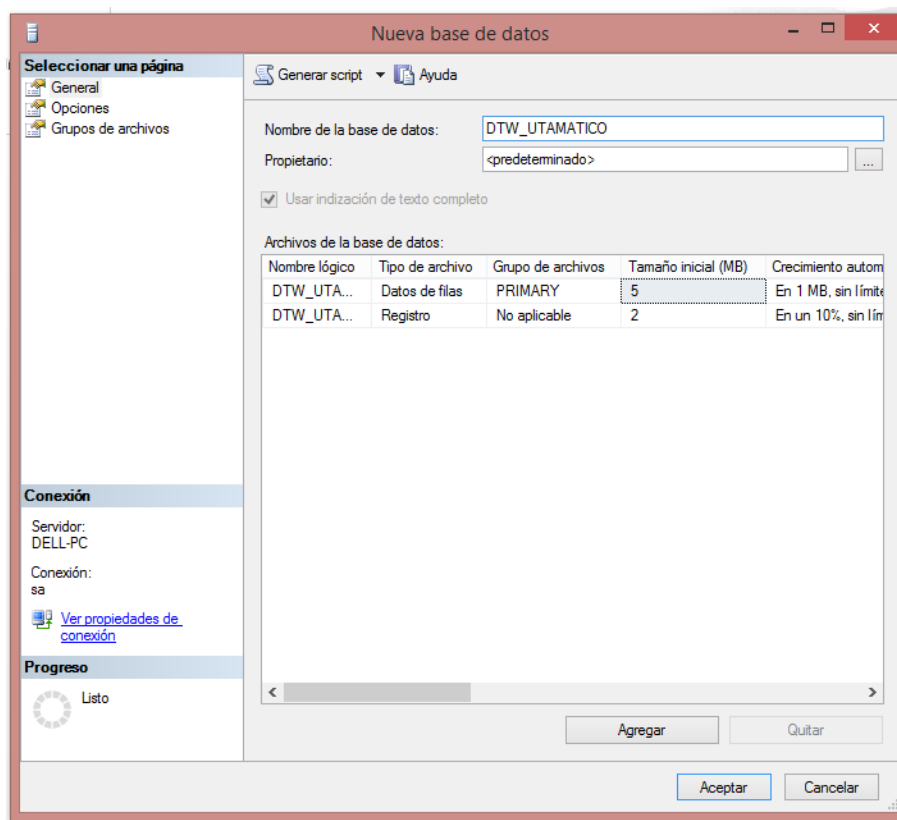
Con el Gestor de Base de Datos SQL-SERVER creamos una Base de Datos, la misma que dispone de la estructura necesaria para el Data Warehouse o Almacén de Datos. Como muestra la Figura 39.



**Figura 38.** Creación de la Base de Datos

**Elaborado:** Investigador

- Se procede con la asignación del nombre que para el Data Warehouse o Almacén de Datos, como muestra la Figura 40.



**Figura 39.** Asignación del Nombre al Data Warehouse

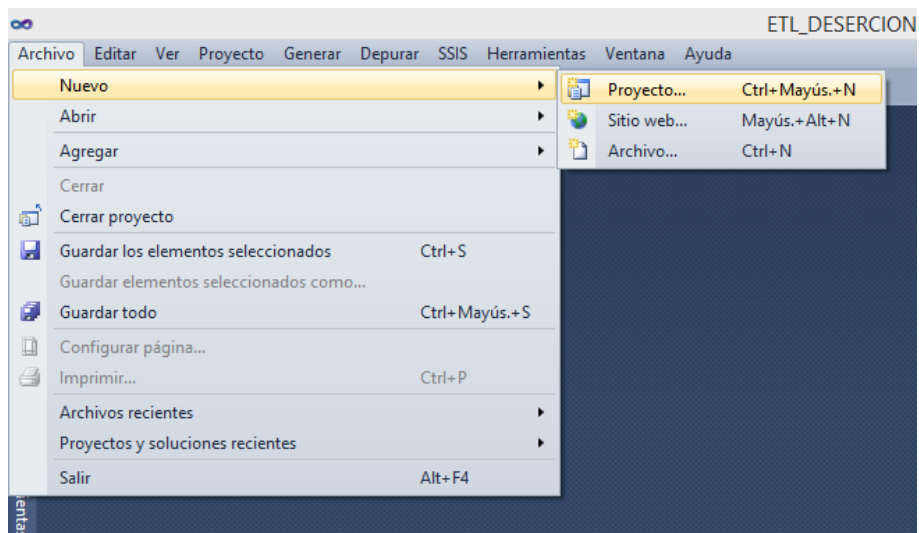
**Elaborado:** Investigador

## 2. Creación de un ETL (extracción, transformación y carga)

Para la creación del ETL utilizaremos la herramienta de Business Intelligence de Visual Studio:

- Creación de nuevo Proyecto, en el menú archivo, opción Nuevo Proyecto, Figura 41.

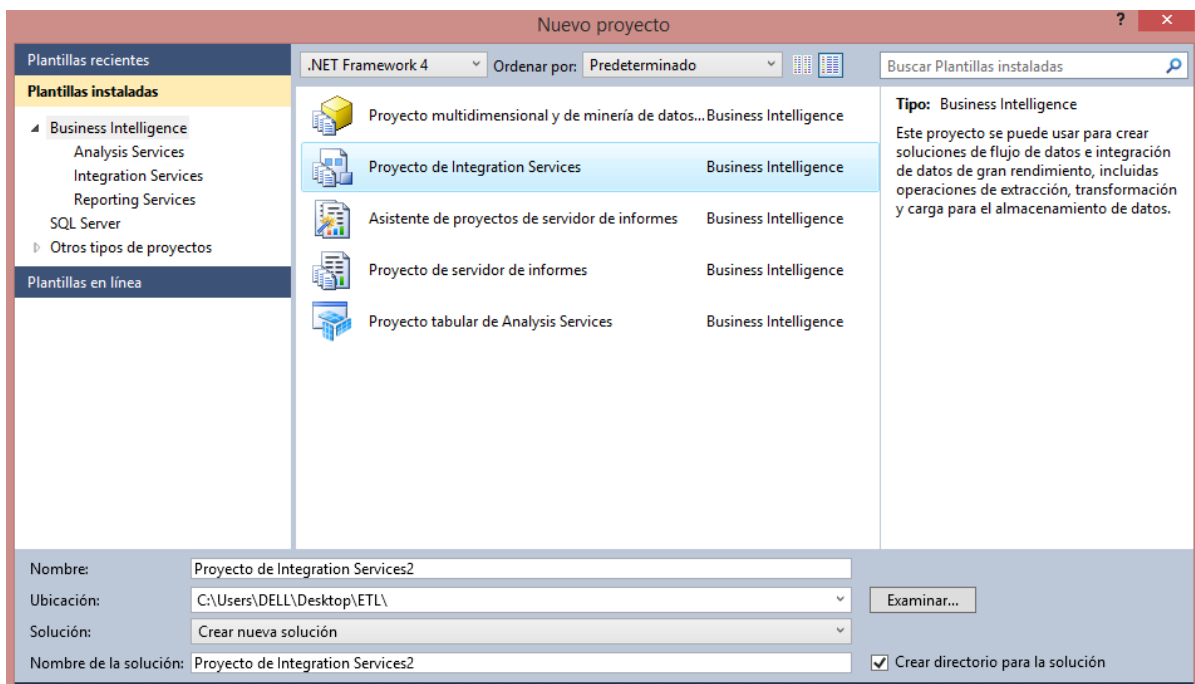




**Figura 40.** Creación de nuevo Proyecto

**Elaborado:** Investigador

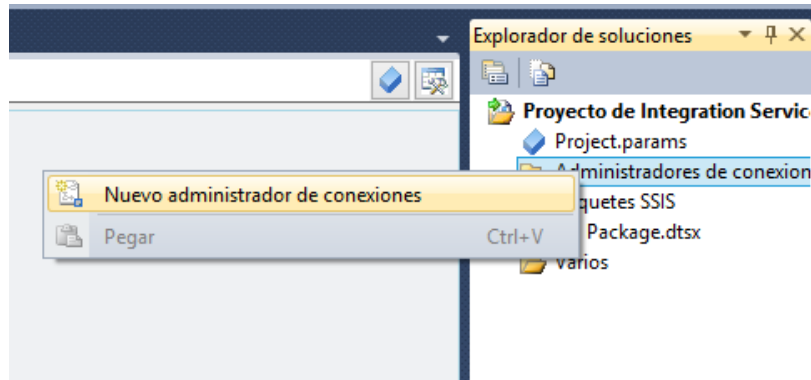
- Se procede con la creación de un Nuevo Proyecto de Integration Services, y la asignación del nombre al proyecto, como se muestra en la Figura 42.



**Figura 41.** Creación de Proyecto de Integration Services

**Elaborado:** Investigador

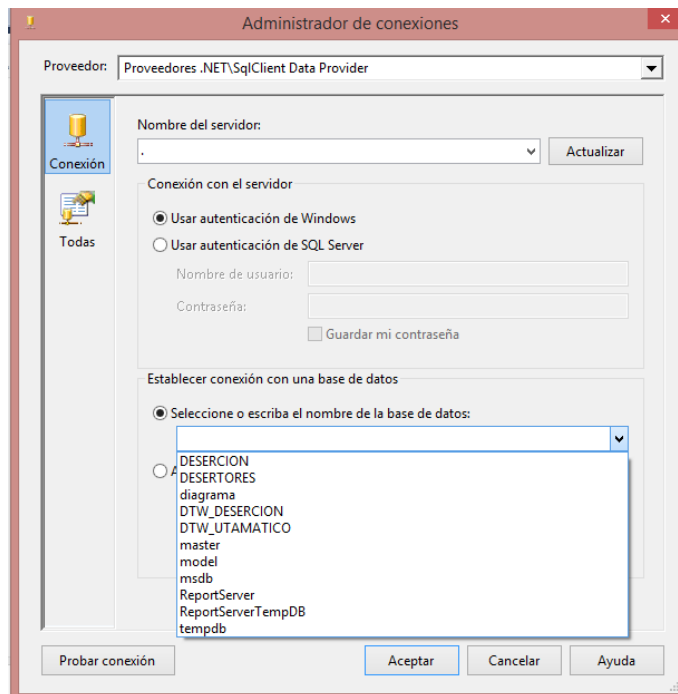
- Se configura la conexiones con las que se necesite trabajar, Para lo cual se procede con la creación de un Nuevo Administrador de conexiones, como se muestra en la Figura 43.



**Figura 42.** Nuevo Administrador de conexiones

**Elaborado:** Investigador

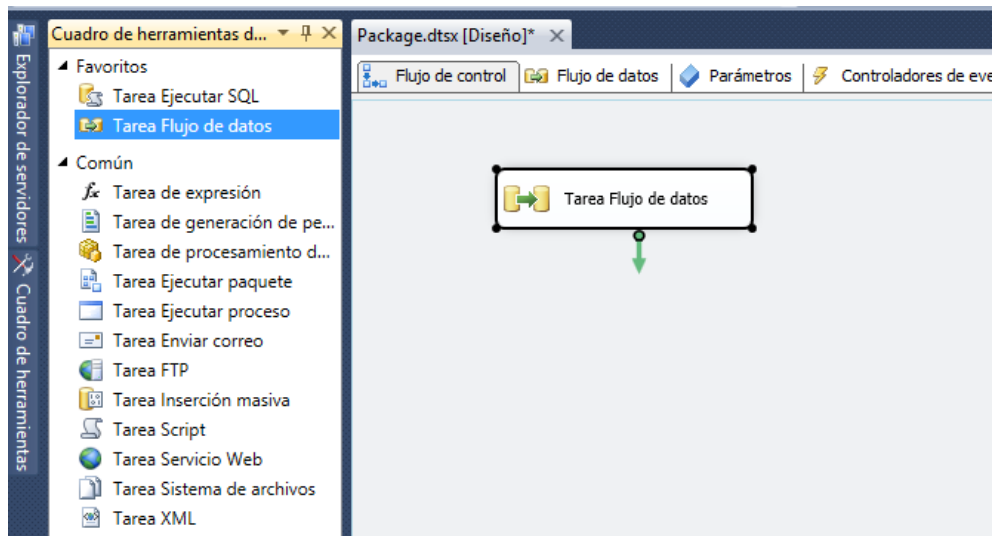
- Se establece conexión con la Base de Datos Transaccional. Como lo muestra en la Figura 44.



**Figura 43.** Conexión con la Base de Datos Transaccional.

**Elaborado:** Investigador

- Una vez establecidas conexiones, se procede con la selección de Tarea de Flujo de Datos, Observar Figura 45.



**Figura 44.** Tarea de Flujo de Datos

**Elaborado:** Investigador

- Se procede con la Asignación del Nombre a la Tarea de Flujo de Datos, como lo muestra la Figura 46.

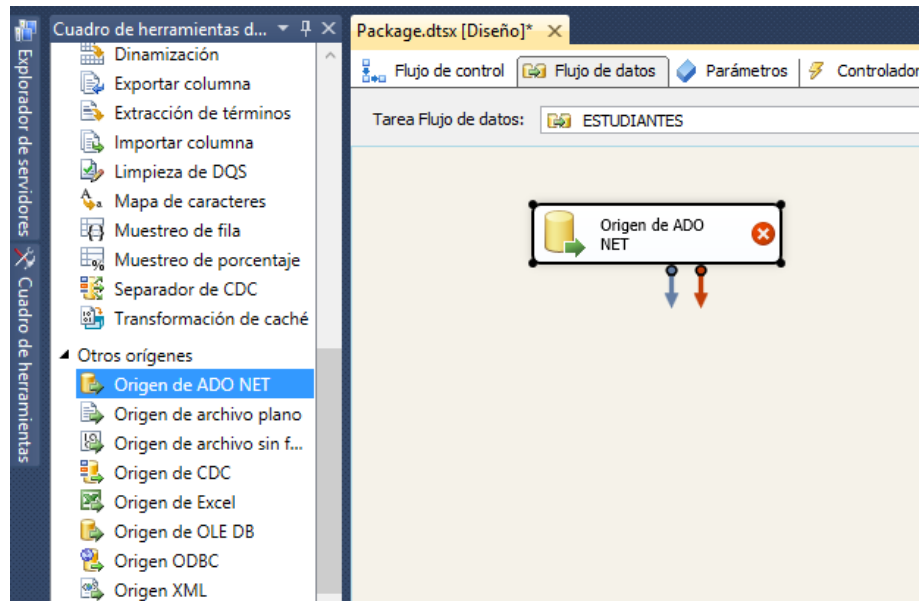


**Figura 45.** Asignación del Nombre a la Tarea de Flujo de Datos

**Elaborado:** Investigador

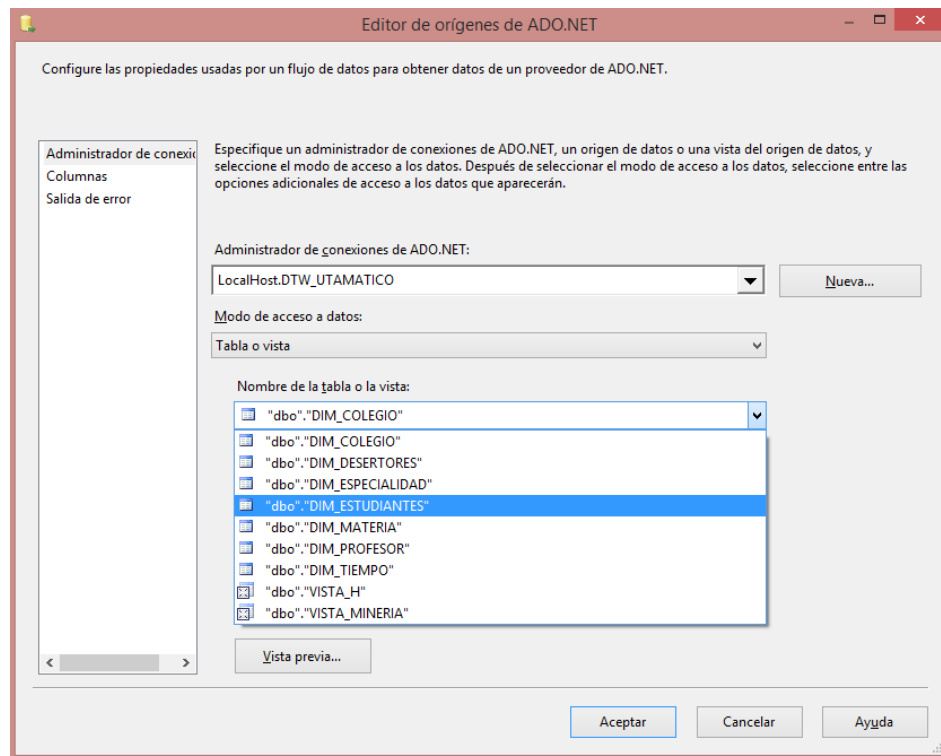
- Seleccionar la pestaña flujo de datos, en donde se selecciona un Origen de Datos, como se observa en la Figura 47 y consiguientemente se procede con la

configuración de la conexión para obtener los datos, como se observa en la Figura 48.



**Figura 46.** Origen de Datos

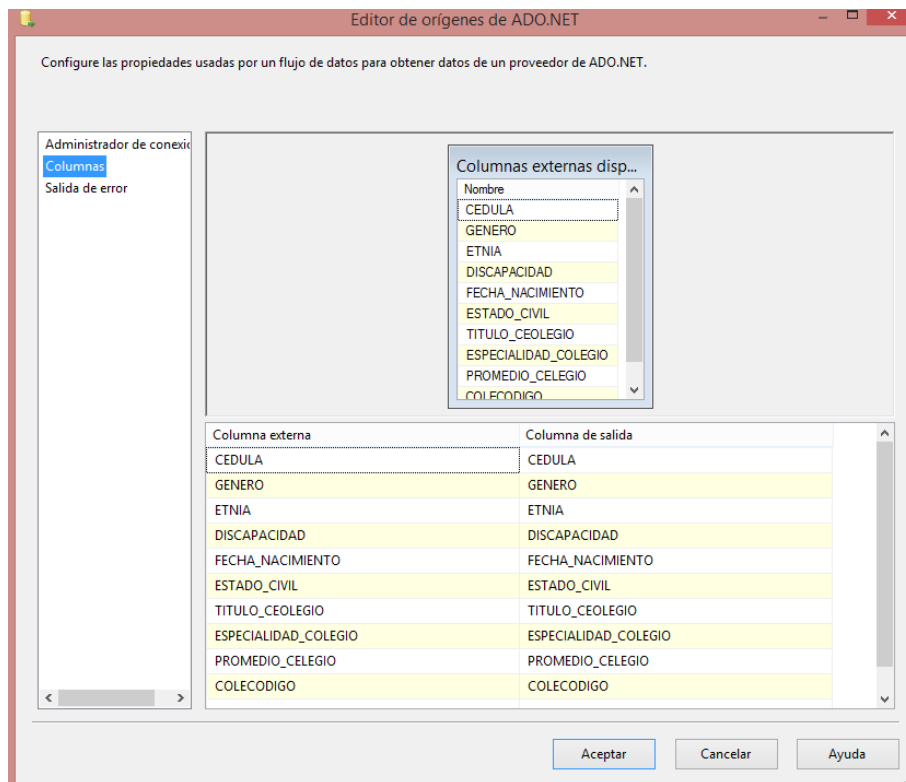
**Elaborado:** Investigador



**Figura 47.** Configuración de la conexión para obtener los datos

**Elaborado:** Investigador

- En la siguiente figura se observa como las propiedades del flujo de datos permite obtener datos del proveedor.

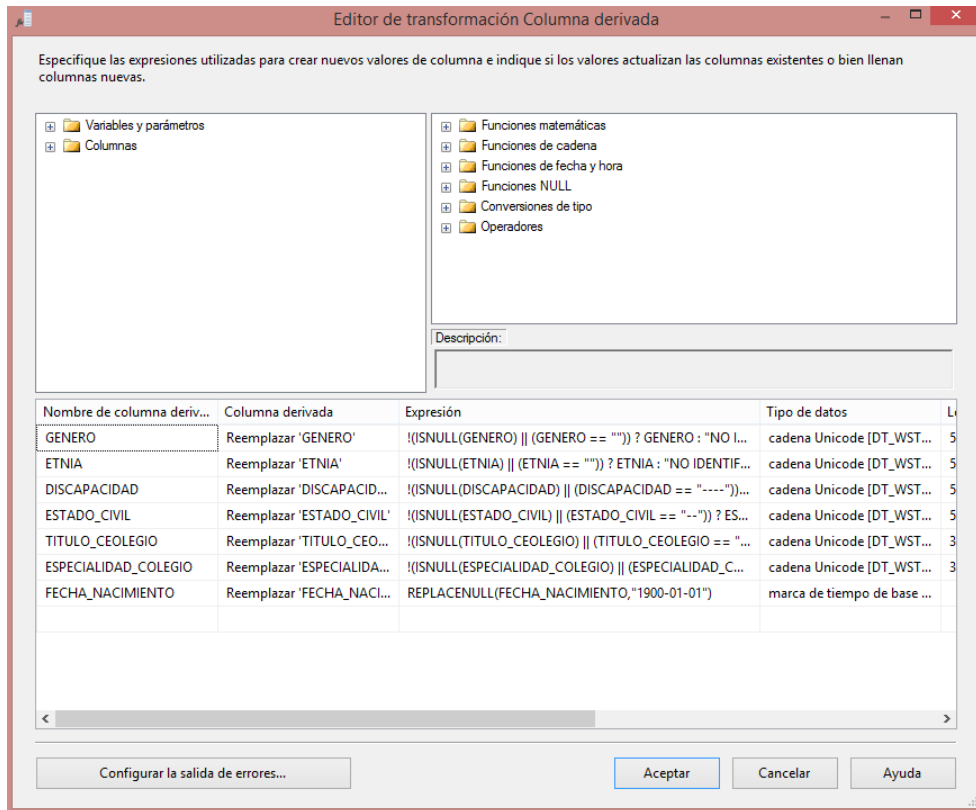


**Figura 48.** Propiedades del flujo de datos permite obtener datos del proveedor.

**Elaborado:** Investigador

- Se procede con el uso del componente de Columna Derivada, que permite realizar las reglas necesarias para la limpieza de los datos en blanco, nulos, según Figura 50.

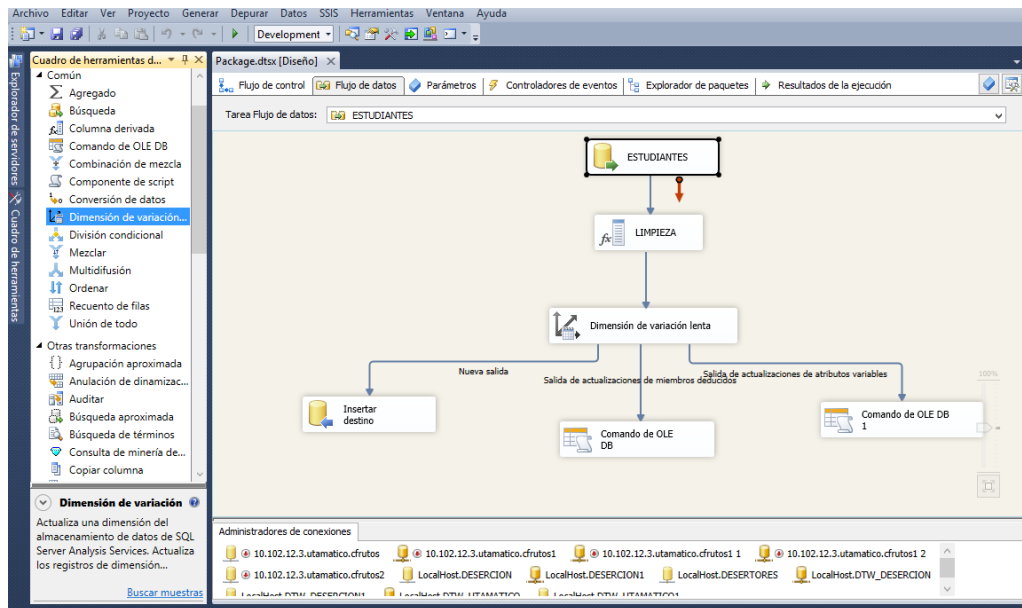
Ejemplo: *!(ISNULL(ETNIA) || (ETNIA == "")) ? ETNIA : "NO IDENTIFICADO"*



**Figura 49.** Columna Derivada

**Elaborado:** Investigador

- Para la carga de datos se selecciona el componente de Variación Lenta el que permite Insertar o Actualizar los datos en la tabla destino, Figura 51.

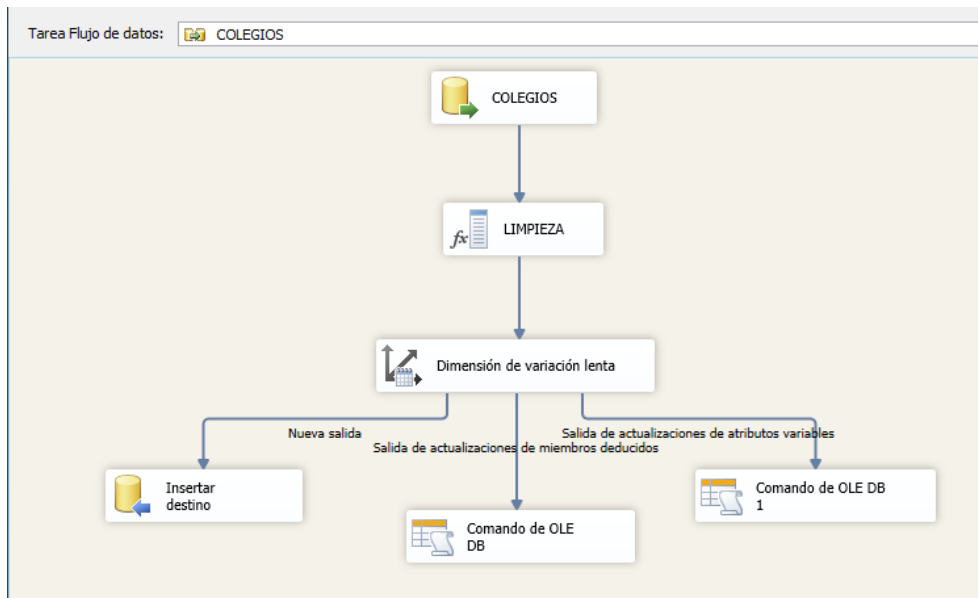


**Figura 50.** Variación Lenta  
**Elaborado:** Investigador

- Se debe realizar el mismo procedimiento con cada una de las vistas creadas en la Base de Datos Transaccional, a continuación se observa el ETL para cada una de la vista mencionadas para obtener los Datos de los Estudiantes Desertores.

## ETL COLEGIOS

En la siguiente figura se muestra la tarea de flujo que permite realizar el proceso de extracción, transformación y carga al almacén de datos de la vista Colegios. Ver la figura 52.



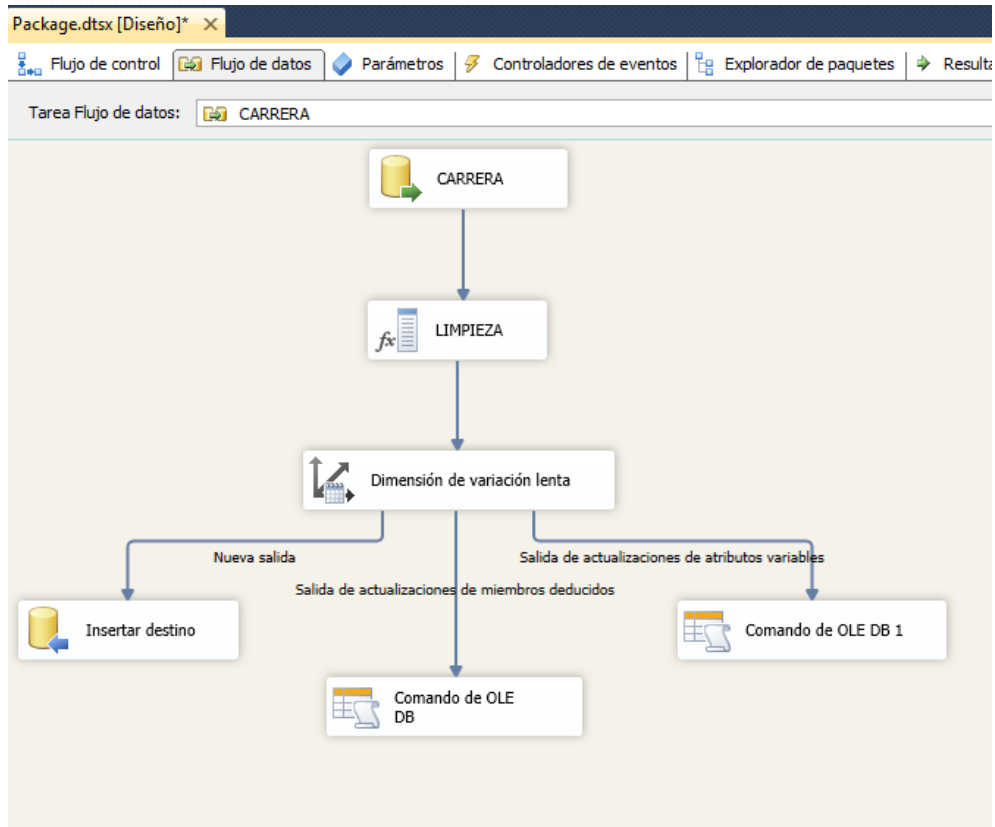
**Figura 51.** ETL COLEGIOS

**Elaborado:** Investigador



## ETL CARRERA

En la siguiente figura se muestra la tarea de flujo que permite realizar el proceso de extracción, transformación y carga al almacen de datos de la vista Carreras. Ver la figura 53.



**Figura 52. ETL CARRERA**

**Elaborado:** Investigador

## ETL MATERIAS

En la siguiente figura se muestra la tarea de flujo que permite realizar el proceso de extracción, transformación y carga al almacén de datos de la vista Materias.

Ver la figura 54.



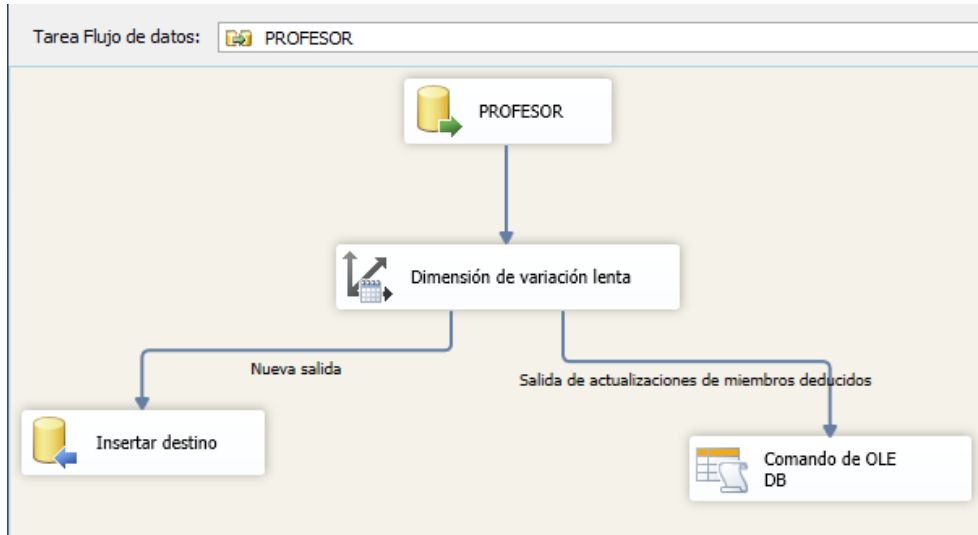
**Figura 53.** ETL MATERIAS

**Elaborado:** Investigador

## ETL PROFESOR

En la siguiente figura se muestra la tarea de flujo que permite realizar el proceso de extracción, transformación y carga al almacén de datos de la vista Profesor.

Ver la figura 55.



**Figura 54.** ETL PROFESOR

**Elaborado:** Investigador

## ETL PERIODOS

En la siguiente figura se muestra la tarea de flujo que permite realizar el proceso de extracción, transformación y carga al almacén de datos de la vista Tiempo.

Ver la figura 56.



**Figura 55.** ETL PERIODOS

**Elaborado:** Investigador

## ETL DESERTORES

En la siguiente figura se muestra la tarea de flujo que permite realizar el proceso de extracción, transformación y carga al almacén de datos de la vista Desertores.

Ver la figura 57.

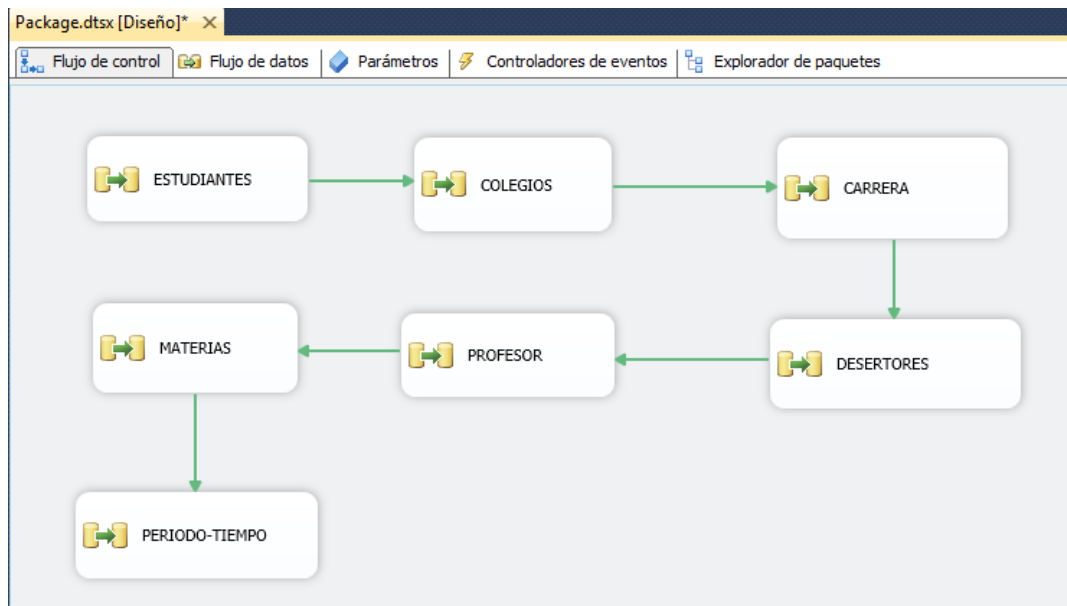


**Figura 56. ETL DESERTORES**

**Elaborado:** Investigador

## CARGA COMPLETA

Por ultimo se procede con la ejecución del proyecto completo, como lo muestra la Figura 58, en donde se procede con la extracción, transformación y carga de los datos limpios de los estudiantes desertores de la Universidad Técnica de Ambato.

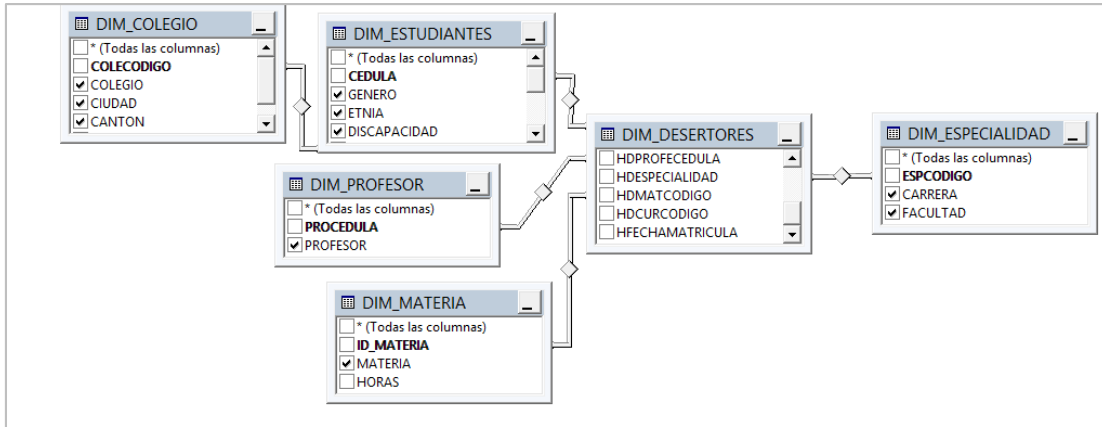


**Figura 57.** Proyecto ETL  
**Elaborado:** Investigador

### 6.7.3.3. Transformación

Esta fase debe transformar la fuente de datos en un conjunto listo para aplicar las técnicas de minería de datos.

- Previo a la transformación de la fuente es necesario la creación de la vista que permita obtener los datos completos de todos los estudiantes desertores. Lo cual lo realizamos desde el Data Warehouse o Almacén de Datos. Con el Siguiete Vista. Figura 59.



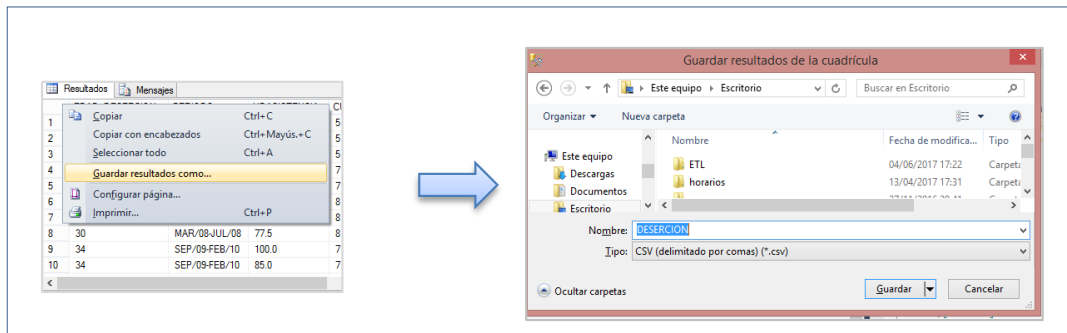
**Figura 58. Diagrama Desertor**  
**Elaborado: Investigador**

```

SELECT CAST(DATEDIFF(dd, dbo.DIM_ESTUDIANTES.FECHA_NACIMIENTO, dbo.DIM_DESERTORES.HFECHAMATRICULA) / 365 AS int) AS EDAD_DESERCION, dbo.DIM_DESERTORES.PERIODO,
dbo.DIM_DESERTORES.HDASISTENCIA, SUBSTRING(dbo.DIM_DESERTORES.HDCURCODIGO, 6, 1) AS CURSO, dbo.DIM_ESTUDIANTES.GENERO, dbo.DIM_ESTUDIANTES.ETNIA,
dbo.DIM_ESTUDIANTES.DISCAPACIDAD, dbo.DIM_ESTUDIANTES.ESTADO_CIVIL, dbo.DIM_ESTUDIANTES.TITULO_COLEGIO, dbo.DIM_ESTUDIANTES.PROMEDIO_COLEGIO, dbo.DIM_ESPECIALIDAD.CARRERA,
dbo.DIM_ESPECIALIDAD.FACULTAD, dbo.DIM_COLEGIO.COLEGIO, dbo.DIM_COLEGIO.CIUDAD, dbo.DIM_COLEGIO.CANTON, dbo.DIM_COLEGIO.PROVINCIA, dbo.DIM_MATERIA.MATERIA,
dbo.DIM_PROFESOR.PROFESOR
FROM
dbo.DIM_DESERTORES INNER JOIN
dbo.DIM_ESTUDIANTES ON dbo.DIM_DESERTORES.HDESTCODIGO = dbo.DIM_ESTUDIANTES.CEDULA INNER JOIN
dbo.DIM_ESPECIALIDAD ON dbo.DIM_DESERTORES.HDESPECIALIDAD = dbo.DIM_ESPECIALIDAD.ESPCODIGO INNER JOIN
dbo.DIM_COLEGIO ON dbo.DIM_ESTUDIANTES.COLECODIGO = dbo.DIM_COLEGIO.COLECODIGO INNER JOIN
dbo.DIM_MATERIA ON dbo.DIM_DESERTORES.HDMATCODIGO = dbo.DIM_MATERIA.ID_MATERIA INNER JOIN
dbo.DIM_PROFESOR ON dbo.DIM_DESERTORES.HDPROFECEDULA = dbo.DIM_PROFESOR.PROCEDULA

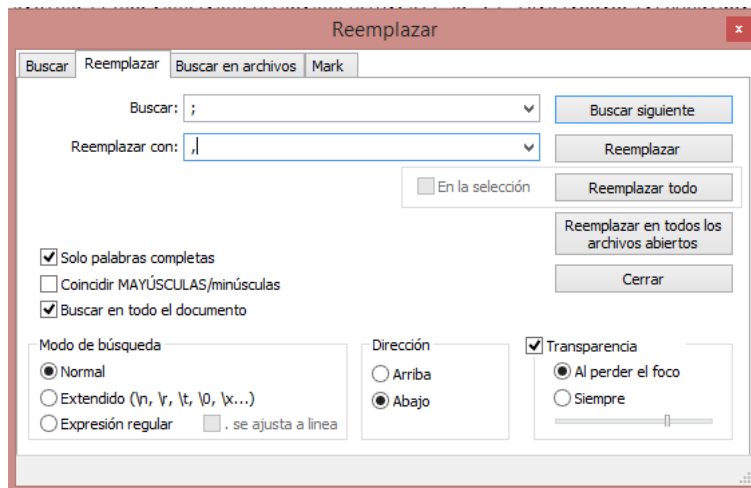
```

- Con la vista lista se procede con la creación del Archivo .csv, como se observa en la Figura 60.



**Figura 59. Creación del Archivo .CSV**  
**Elaborado: Investigador**

- Como se muestra en la Figura 61. Se procede con el reemplazo del carácter “;” por el carácter “,”



**Figura 60.** Reemplazo Carácter “;”  
**Elaborado:** Investigador

#### 6.7.3.4. Minería de Datos

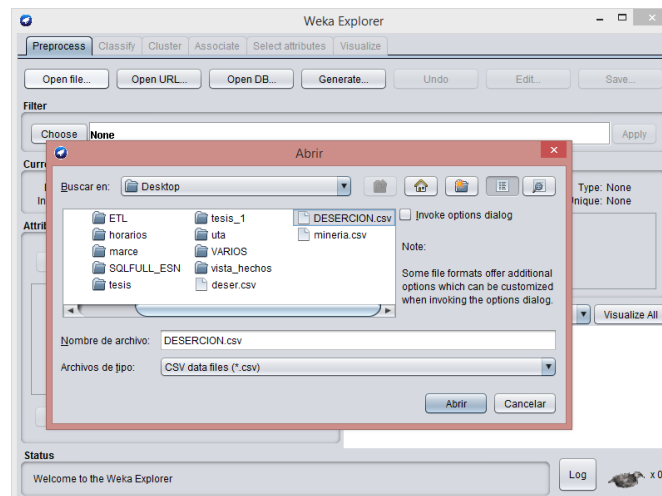
Para la fase de Minería de Datos usaremos la herramienta WEKA. Figura 62.



**Figura 61.** Weka  
**Elaborado:** Investigador



- En el menú de aplicaciones que contiene Weka se procede con la selección de “Explorer”, y se procede con la carga del archivo .csv



**Figura 62.** Carga de archivo .csv  
**Elaborado:** Investigador

- Ahora procedemos a aplicar la Minería de Datos mediante **Agrupamiento**

Los algoritmos de agrupamiento buscan grupos de instancias con características similares, según un criterio de comparación entre valores de atributos de las instancias definidos en los algoritmos.

### **Agrupamiento Numérico**

Algoritmo K-Medias: Se trata de un algoritmo clasificado como Método de Particionado y Recolocación.

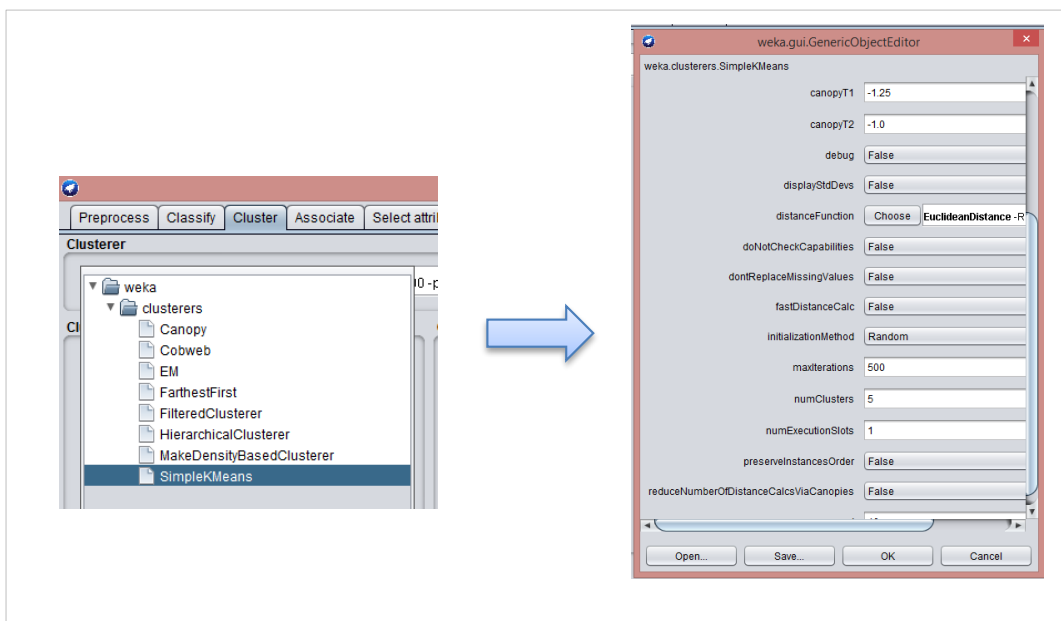
Este método es hasta ahora el más utilizado en aplicaciones científicas e industriales. El nombre le viene porque representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide. Este método únicamente se puede aplicar a atributos numéricos, y los outliers le pueden afectar muy negativamente.

Sin embargo, la representación mediante centroides tiene la ventaja de que tiene un significado gráfico y estadístico inmediato. La suma de las discrepancias entre un punto y su centroide, expresado a través de la distancia apropiada, se usa como función objetivo. La función objetivo, suma de los cuadrados de los errores entre

los puntos y sus centroides respectivos, es igual a la varianza total dentro del propio cluster.

La suma de los cuadrados de los errores se puede racionalizar, como el negativo del log-likelihood, para modelos mixtos que utilicen distribuciones normales. Vamos a aplicar el algoritmo de agrupamiento Kmedias, por ser uno de los más veloces y eficientes. (Jiménez, M. G., & Sierra, 2010)

Para la aplicación de este algoritmo se procede con la selección del algoritmo SimpleKMeans con un número de clusters igual a 5, como se observa en la Figura 64.



**Figura 63.** Selección Algoritmo Kmedia  
**Elaborado:** Investigador

A continuación observamos el número de instancias agrupadas:

Clustered Instances		
0	10300	( 10%)
1	29995	( 28%)
2	28387	( 27%)
3	8542	( 8%)
4	28782	( 27%)

**Figura 64.** Instancias agrupadas  
**Elaborado:** Investigador

## Resultados k-Media

Initial starting points (random):

### Cluster 0:

22,OCT/15-MAR/16,79,8,MASCULINO,MESTIZA,NINGUNA,SOLTERO,'TECNICO INDUSTRIAL',17,'PSICOLOGIA EDUCATIVA','CIENCIAS HUMANAS Y DE LA EDUCACION','INS TEC IND RAMON BARBA NARANJ',LATACUNGA,LATACUNGA,COTOPAXI,'TALLER DE NEUROLINGUISTICA'

### Cluster 1:

21,SEP/11-FEB/12,93.5,7,MASCULINO,MESTIZA,NINGUNA,SOLTERO,CIENCIAS,17,'INGENIERIA CIVIL','INGENIERIA CIVIL Y MECANICA','COL JUAN FRANCISCO MONTALVO',AMBATO,AMBATO,TUNGURAHUA,'HORMIGON I'

### Cluster 2:

18,SEP/06-FEB/07,80,1,FEMENINO,NINGUNO,'NO IDENTIFICADO',SOLTERO,CIENCIAS,16,MEDICINA,'CIENCIAS DE LA SALUD','COL EXP AMBATO',AMBATO,AMBATO,TUNGURAHUA,'INTRODUCCION A LA MEDICINA'

### Cluster 3:

23,SEP/10-FEB/11,79,0,FEMENINO,NINGUNO,'NO IDENTIFICADO',SOLTERO,'POLIVALENTE EN ADMINISTRACION',18,'ING. MARKETING Y GESTION DE NEGOCIOS','CIENCIAS ADMINISTRATIVAS','INS TEC FEM LUIS FERNANDO RUIZ',LATACUNGA,LATACUNGA,COTOPAXI,'TECNICAS ELABORACION DE TESIS'

### Cluster 4:

20,MAR/10-AGO/10,100,1,MASCULINO,MESTIZA,NINGUNA,SOLTERO,'TECNICO EN INFORMATICA',17,GENERAL,'INGENIERIA EN SISTEMAS ELECTRONICA E INDUSTRIAL','INS TEC SUP BOLIVAR',AMBATO,AMBATO,TUNGURAHUA,ALGEBRA

FULL DATA	
Attribute	Full Data (106006.0)
=====	=====
i»¿EDAD_DESERCION	21.2689
PERIODO	MAR/12-AGO/12
HDASISTENCIA	87.7517
CURSO	3.5061
GENERO	MASCULINO
ETNIA	MESTIZA
DISCAPACIDAD	NINGUNA
ESTADO_CIVIL	SOLTERO
TITULO_COLEGIO	CIENCIAS
PROMEDIO_COLEGIO	16.0173
CARRERA	INGENIERIA CIVIL
FACULTAD	INGENIERIA EN SISTEMAS ELECTRONICA E INDUSTRIAL
COLEGIO	INS TEC SUP BOLIVAR
CIUDAD	AMBATO
CANTON	AMBATO
PROVINCIA	TUNGURAHUA
MATERIA	FISICA I

**Tabla 26. FULL DATA**

**Elaborado:** Investigador

Como se puede observar en la tabla 26. FULL DATA el número de total de atributos analizados es 106006.

- ✓ En donde la edad promedio que registra mayor incidencia en deserción es de 21 años.
- ✓ El período que registra mayor número de deserciones es el período académico Marzo. Agosto 2012.
- ✓ El porcentaje de asistencia de los estudiantes desertores es del 87%
- ✓ Los estudiantes de etnia mestiza son los que mayor número de incidencia en deserción.
- ✓ Los estudiantes que tienen ningún tipo de discapacidad son aquellos que mayor índice de deserción presentan
- ✓ Los estudiantes con estado civil Soltero son aquellos que inciden más en deserción
- ✓ Los estudiantes con título de colegio en Ciencias son aquellos que tienen mayor incidencia en deserción.
- ✓ Los estudiantes que como promedio de colegio es “16” son aquellos que mayor índice de deserción presentan.

- ✓ Los estudiantes de la carrera de Ingeniería en Civil son los que mayor número de incidencia en deserción.
- ✓ La facultad que mayor número de incidencia en deserción presenta es la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.
- ✓ Los estudiantes graduados en el colegio Instituto Tecnológico Superior Bolívar es el que mayor número de incidencia en deserción.
- ✓ Los estudiantes que residen en la ciudad y cantón Ambato son aquellos con mayor número de incidencia en deserción.
- ✓ La materia que mayor número de incidencia en deserción presenta es Física I.

CLOUSTER 0	
Attribute	Cluster#
	0
	(12015.0)
=====	-----
i»¿EDAD_DESERCION	20.6553
PERIODO	OCT/15-MAR/16
HDASISTENCIA	89.9958
CURSO	3.3605
GENERO	MASCULINO
ETNIA	MESTIZA
DISCAPACIDAD	NINGUNA
ESTADO_CIVIL	SOLTERO
TITULO_CEOLEGIO	CIENCIAS
PROMEDIO_CEOLEGIO	16.4957
CARRERA	INGENIERIA MECANICA
FACULTAD	INGENIERIA CIVIL Y MECANICA
COLEGIO	INS TEC IND RAMON BARBA NARANJ
CIUDAD	LATACUNGA
CANTON	LATACUNGA
PROVINCIA	COTOPAXI
MATERIA	LENGUAJE Y COMUNICACION

**Tabla 27. CLOUSTER 0**

**Elaborado:** Investigador

El resultado del CLOUSTER 0 con 12015 atributos representa el cuarto en nivel de incidencia de cinco niveles, según la tabla 27.

En donde el primer patrón detectado con mayor incidencia en deserción es:

1. Edad de 20 años.
2. Período académico Octubre 2015 – Marzo 2016.
3. Curso Tercer semestre
4. Género Masculino
5. El porcentaje de asistencia de los estudiantes es el 89.99%
6. Etnia mestiza
7. Discapacidad ninguna
8. Estado civil Soltero
9. Título de colegio en Ciencias
10. Promedio de graduación de colegio es 16.49
11. Carrera de Ingeniería Mecánica
12. Facultad de Ingeniería Civil y Mecánica
13. Colegio Instituto Tecnológico Industrial Ramón Barba Naranjo
14. Cantón Latacunga
15. Provincia de Cotopaxi
16. Materia Lenguaje y Comunicación.

CLOUSTER 1	
Attribute	1 (29232.0)
=====	=====
i»¿EDAD_DESERCION	21.9222
PERIODO	SEP/11-FEB/12
HDASISTENCIA	87.5962
CURSO	4.5356
GENERO	MASCULINO
ETNIA	MESTIZA
DISCAPACIDAD	NINGUNA
ESTADO_CIVIL	SOLTERO
TITULO_COLEGIO	CIENCIAS
PROMEDIO_COLEGIO	16.4301
CARRERA	INGENIERIA CIVIL
FACULTAD	INGENIERIA CIVIL Y MECANICA
COLEGIO	INS TEC SUP BOLIVAR
CIUDAD	AMBATO
CANTON	AMBATO
PROVINCIA	TUNGURAHUA
MATERIA	EMPLEO DE NTICS II

**Tabla 28.** CLOUSTER 1

**Elaborado:** Investigador

El resultado del CLOUSTER 1 con un total de 29232 atributos representa el primero y más importante patrón de deserción de incidencia de cinco niveles, según la tabla 28.

En donde el patrón de mayor relevancia detectado con mayor incidencia en deserción es:

1. Edad de 21 años.
2. Período académico Septiembre 2011 – Febrero 2012.
3. Curso cuarto semestre
4. Género masculino
5. El porcentaje de asistencia de los estudiantes es el 87%
6. Etnia mestiza
7. Discapacidad ninguna
8. Estado civil Soltero
9. Título de colegio en Ciencias
10. Promedio de graduación de colegio es 16.43
11. Carrera de Ingeniería Civil
12. Faculta de Ingeniería Civil y Mecánica

13. Colegio Instituto Tecnológico Superior Bolívar

14. Cantón Ambato

15. Provincia de Tungurahua

16. Materia Empleo de Ntic's II

CLOUSTER 2	
Attribute	2
-----	-----
i»¿EDAD_DESERCION	21.5761
PERIODO	SEP/07-FEB/08
HDSISTENCIA	86.3478
CURSO	3.6192
GENERO	FEMENINO
ETNIA	NINGUNO
DISCAPACIDAD	NO IDENTIFICADO
ESTADO_CIVIL	SOLTERO
TITULO_CELEGIO	CIENCIAS
PROMEDIO_CELEGIO	15.2484
CARRERA	MEDICINA
FACULTAD	CIENCIAS DE LA SALUD
COLEGIO	COL EXP AMBATO
CIUDAD	AMBATO
CANTON	AMBATO
PROVINCIA	TUNGURAHUA
MATERIA	TECNICAS DE ESTUDIO

**Tabla 29.** CLOUSTER 2

**Elaborado:** Investigador

El resultado del CLOUSTER 2 con un total de 28020 atributos representa el segundo en orden de relevancia de deserción esto de cinco niveles, según la tabla 29.

En donde el segundo patrón de mayor relevancia detectado con mayor incidencia en deserción es:

1. Edad de 21 años.
2. Período académico Septiembre 2007 – Febrero 2008.
3. Curso Tercero
4. Género Femenino



5. El porcentaje de asistencia de los estudiantes es el 86%
6. Etnia mestiza
7. Discapacidad ninguna
8. Estado civil Soltero
9. Título de colegio en Ciencias
10. Promedio de graduación de colegio es 16.43
11. Carrera de Ingeniería Civil
12. Faculta de Ingeniería Civil y Mecánica
13. Colegio Instituto Tecnológico Superior Bolívar
14. Cantón Ambato
15. Provincia de Tungurahua
16. Materia Empleo de Ntic's II

CLOUSTER 3	
Attribute	3
	(7893.0)
=====	=====
i»¿EDAD_DESERCION	20.987
PERIODO	SEP/10-FEB/11
HDASISTENCIA	84.9411
CURSO	2.9582
GENERO	FEMENINO
ETNIA	NINGUNO
DISCAPACIDAD	NO IDENTIFICADO
ESTADO_CIVIL	SOLTERO
TITULO_COLEGIO	CIENCIAS
PROMEDIO_COLEGIO	16.2539
CARRERA	ING. MARKETING Y GESTION DE NEGOCIOS
FACULTAD	CIENCIAS ADMINISTRATIVAS :
COLEGIO	INS TEC VICTORIA VASCONEZ CUVI
CIUDAD	LATACUNGA
CANTON	LATACUNGA
PROVINCIA	COTOPAXI
MATERIA	TECNICAS DE ESTUDIO

**Tabla 30. CLOUSTER 3**  
**Elaborado:** Investigador

El resultado del CLOUSTER 3 con un total de 7893 atributos representa el quinto y último en orden de relevancia de deserción esto de cinco niveles, según la tabla 30.

En donde el patrón que se detalla es el quinto de cinco niveles:

1. Edad de 20 años.
2. Período académico Septiembre 2010 – Febrero 2011.
3. Curso segundo semestre
4. Género Femenino
5. El porcentaje de asistencia de los estudiantes es el 84%
6. Etnia mestiza
7. Discapacidad ninguna
8. Estado civil Soltero
9. Título de colegio en Ciencias
10. Promedio de graduación de colegio es 16.25
11. Carrera de Ingeniería Marketing y Gestión de Negocios
12. Facultad de Ciencias Administrativas
13. Colegio Instituto Victoria Vascones
14. Cantón Latacunga
- 15. Provincia de Cotopaxi**
- 16. Materia Técnicas de Estudio**

CLOUSTER 4	
Attribute	4 (28846.0)
=====	=====
i»¿EDAD_DESERCION	20.641
PERIODO	MAR/12-AGO/12
HDAASISTENCIA	89.1073
CURSO	2.5635
GENERO	MASCULINO
ETNIA	MESTIZA
DISCAPACIDAD	NINGUNA
ESTADO_CIVIL	SOLTERO
TITULO_CEOLEGIO	COMERCIO Y ADMINISTRACION
PROMEDIO_CEOLEGIO	16.082
CARRERA	GENERAL
FACULTAD	INGENIERIA EN SISTEMAS ELECTRONICA E INDUSTRIAL
COLEGIO	INS SUP TEC HISPANO AMERICA
CIUDAD	AMBATO
CANTON	AMBATO
PROVINCIA	TUNGURAHUA
MATERIA	FISICA I

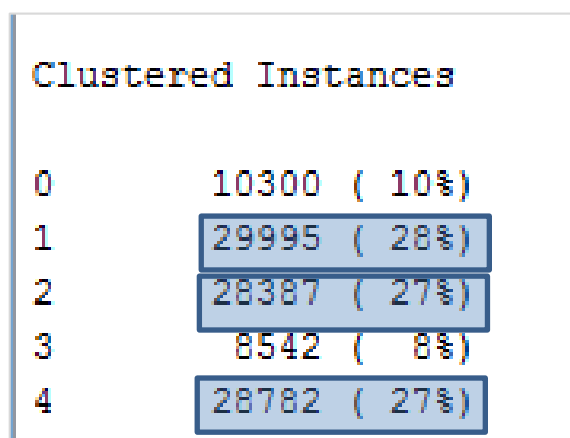
**Tabla 31. CLOUSTER 4**  
**Elaborado:** Investigador

El resultado del CLOUSTER 4 con un total de 28846 atributos representa el tercero en orden de relevancia de deserción esto de cinco niveles, según la tabla 29.

En donde el tercer patrón de mayor relevancia detectado con mayor incidencia en deserción es:

1. Edad de 20 años.
2. Período académico Marzo – Agosto 2012
3. Semestre segundo
4. Género masculino
5. El porcentaje de asistencia de los estudiantes es el 89%
6. Etnia mestiza
7. Discapacidad ninguna
8. Estado civil Soltero
9. Título de colegio en Comercio y Administración
10. Promedio de graduación de colegio es 16
11. Carrera General
12. Facultad de Ingeniería en Sistemas Electrónica e Industrial
13. Colegio Instituto Tecnológico Superior Hispano América
14. Cantón Ambato
- 15. Provincia de Tungurahua**
16. Materia Física I

#### 6.7.3.5. Etapa de Interpretación/ Evaluación de datos



**Figura 65.** Resultados Cluster  
**Elaborado:** Investigador

## PATRONES DE DESERCIÓN ESTUDIANTIL

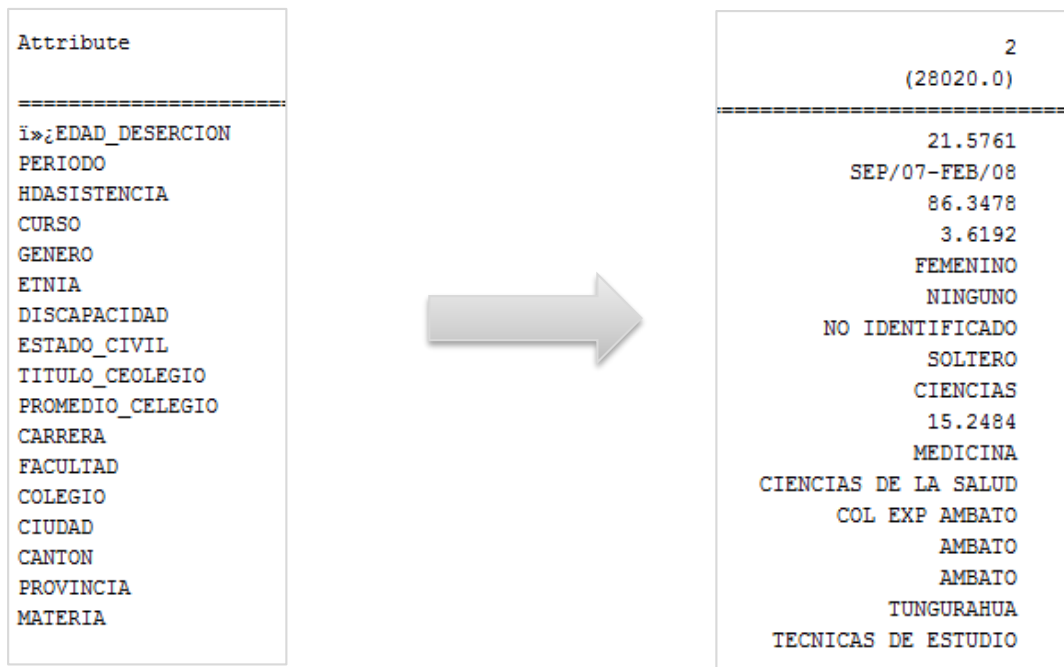
El clouster 1 muestra mayores instancias y porcentaje de incidencias, por tanto es el primer patrón de Deserción Estudiantil

Attribute	
=====	
ï»¿EDAD_DESERCION	1
PERIODO	(29232.0)
HDASISTENCIA	
CURSO	21.9222
GENERO	SEP/11-FEB/12
ETNIA	87.5962
DISCAPACIDAD	4.5356
ESTADO_CIVIL	MASCULINO
TITULO_CEOLEGIO	MESTIZA
PROMEDIO_COLEGIO	NINGUNA
CARRERA	SOLTERO
FACULTAD	CIENCIAS
COLEGIO	16.4301
CIUDAD	INGENIERIA CIVIL
CANTON	INGENIERIA CIVIL Y MECANICA
PROVINCIA	INS TEC SUP BOLIVAR
MATERIA	AMBATO
	AMBATO
	TUNGURAHUA
	EMPLEO DE NTICS II

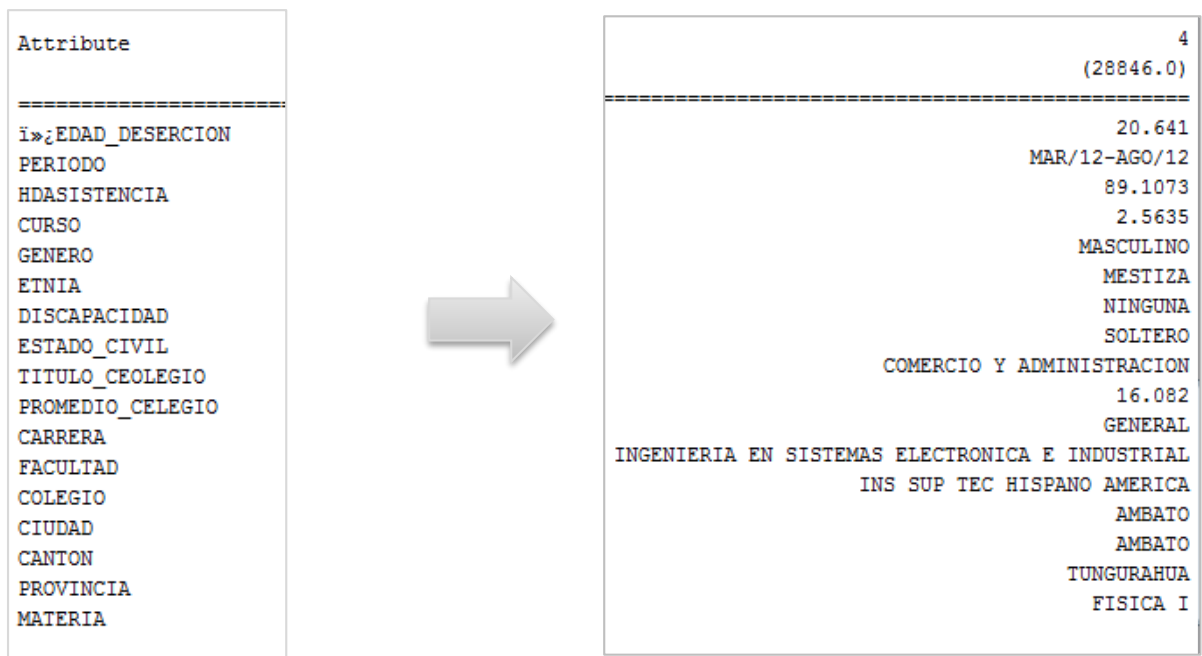
**Figura 66.** Patrón de Deserción Estudiantil N°1

**Elaborado:** Investigador

El cluster 2 y 4 muestran consiguientes las mayores instancias y porcentaje de incidencias, por tanto el segundo y tercer patrón de Deserción Estudiantil



**Figura 67.** Patrón de Deserción Estudiantil N°2  
**Elaborado:** Investigador

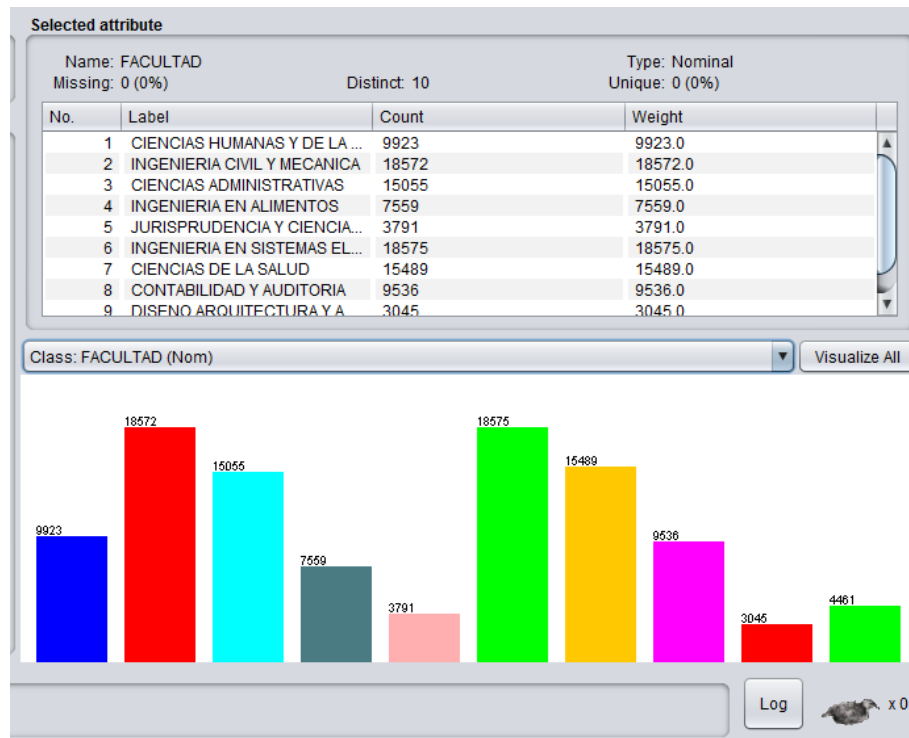


**Figura 68.** Patrón de Deserción Estudiantil N°3  
**Elaborado:** Investigador

La herramienta de visualización de WEKA permite representar gráficas en 2D

## INCIDENCIA DE DESERCIÓN POR FACULTAD

En la siguiente figura se muestra el índice de deserción por cada una de las facultades de la Universidad Técnica de Ambato



**Figura 69.** INCIDENCIA DE DESERCIÓN POR FACULTAD

**Elaborado:** Investigador

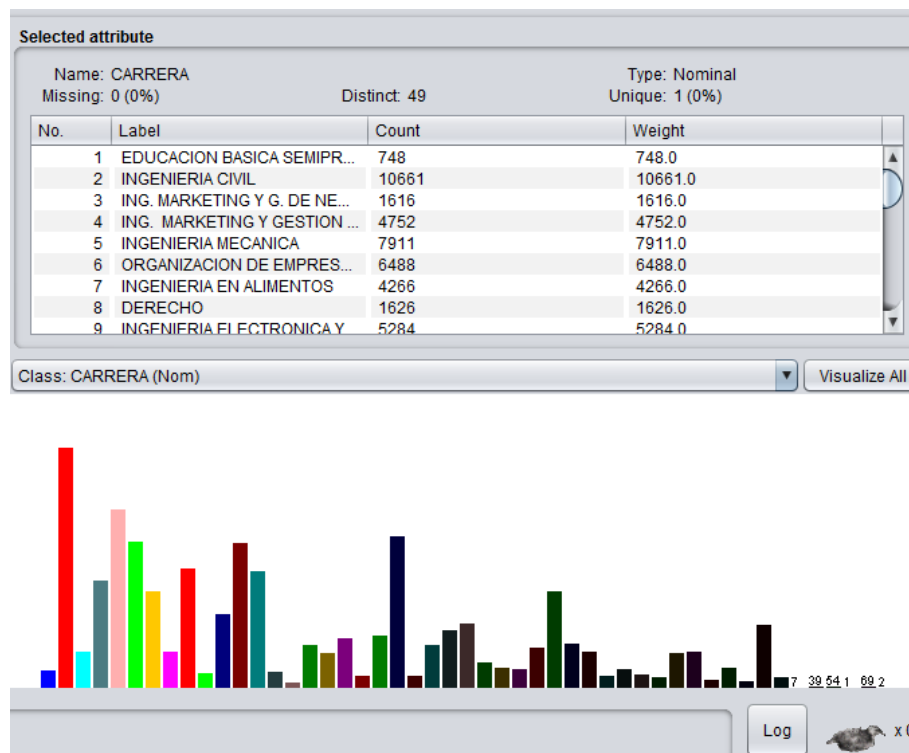
A continuación se detalla el orden de facultades según el índice de deserción que cada una presenta:

1. Facultad de Ingeniería en Sistemas Electrónica e Industrial (18575).
2. Facultad de Ingeniería Civil y Mecánica (18572)
3. Ciencias de la Salud (18489)
4. Ciencias Administrativas (15055)
5. Ciencia Humanas y de la Educación (9923)
6. Contabilidad y Auditoría (9536)
7. Ingeniería en Alimentos ( 7559)

8. Ciencias Agropecuarias (4661)
9. Jurisprudencia y Ciencias Sociales (3791)
10. Diseño, Arquitectura y Artes (3045)

### INCIDENCIA DE DESERCIÓN POR CARRERA

Como se observa en la siguiente figura la carrera que mayor índice de deserción presenta es la de Ingeniería Civil



**Figura 70.** INCIDENCIA DE DESERCIÓN POR CARRERA

**Elaborado:** Investigador

A continuación de detalla el orden de deserción por carreras de la Universidad Técnica de Ambato

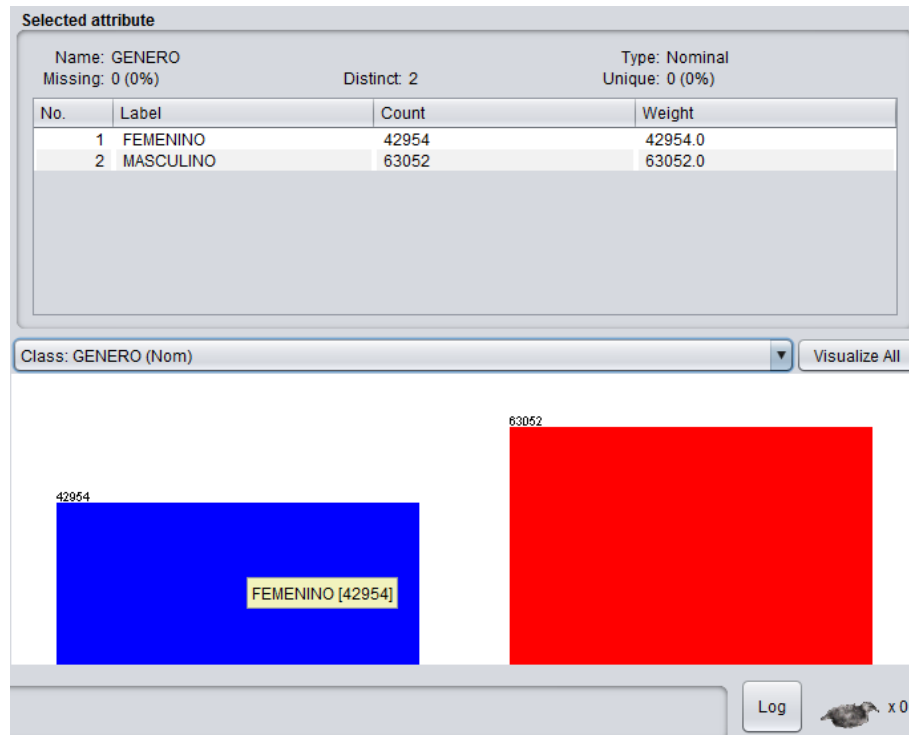
1. Ingeniería Civil	10661
2. Ingeniería Mecánica	7911
3. General	6711
4. Organización de empresas	6488
5. Ingeniería Bioquímica	6403
6. Ingeniería en Electrónica y comunicación	5284
7. Ingeniería en Marketing y G. de Negocios	4752
8. Ingeniería Industrial y Procesos de Automatización	4295
9. Ingeniería en Alimentos	4266
10. Ingeniería en Sistemas Computacionales e Informáticos	2883
11. Medicina Veterinaria	2773
12. Laboratorio Clínico	2571
13. Terapia Física	2327
14. Turismo y Hotelería	2220
15. Economía	1951
16. Idiomas	1923
17. Formación Básica	1901
18. Enfermería	1769
19. Derecho	1626
20. Ingeniería Agronómica	1619
21. Ingeniería en Marketing y G. de Negocios semipresencial	1616
22. Ingeniería Financiera	1579
23. Comunicación Social	1559
24. Psicología Clínica	1530
25. Diseño Gráfico Publicitario	1104
26. Estimulación Temprana	889
27. Diseño de Espacios Arquitectónicos	872
28. Psicología	823
29. Contabilidad y Auditoría Semipresencial	805



30. Educación Básica semipresencial	748
31. Contabilidad y Auditoria	690
32. Docencia Informática	681
33. Psicología Industrial	581
34. Psicología Educativa	559
35. Trabajo Social	552
36. Educación Básica	547
37. Medicina	521
38. Cultura Física	504
39. Disnea de Modas	471
40. Educación Parvulario	349
41. Mercadotecnia	297
42. Cultura Física Semipresencial	250
43. Ingeniería Agropecuaria	69
44. Comunicación Social Semipresencial	54
45. Secretaria en Español Semipresencial	39
46. Antropología social y cultural	7
47. Cultura Estética Semipresencial	2
48. Ingeniería en Recursos Humanos	1

## INCIDENCIA DE DESERCIÓN POR GÉNERO

Como se muestra en la siguiente figura el Género Masculino tiene un índice de deserción mayor que el Género Femenino.



**Figura 71.** INCIDENCIA DE DESERCIÓN POR GÉNERO

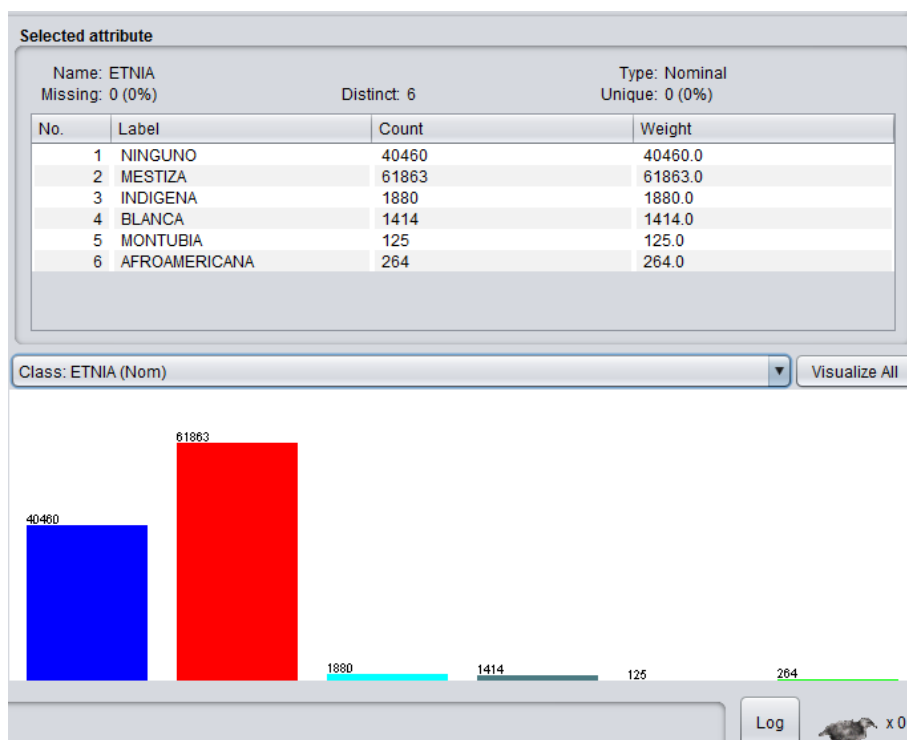
**Elaborado:** Investigador

A continuación se detalla el orden de deserción por género:

1. Femenino 42954
2. Masculino 63052

## INCIDENCIA DE DESERCIÓN POR ETNIA

Como se muestra en la Figura 65, la Etnia Mestiza presenta mayor índice de deserción



**Figura 72.** INCIDENCIA DE DESERCIÓN POR ETNIA

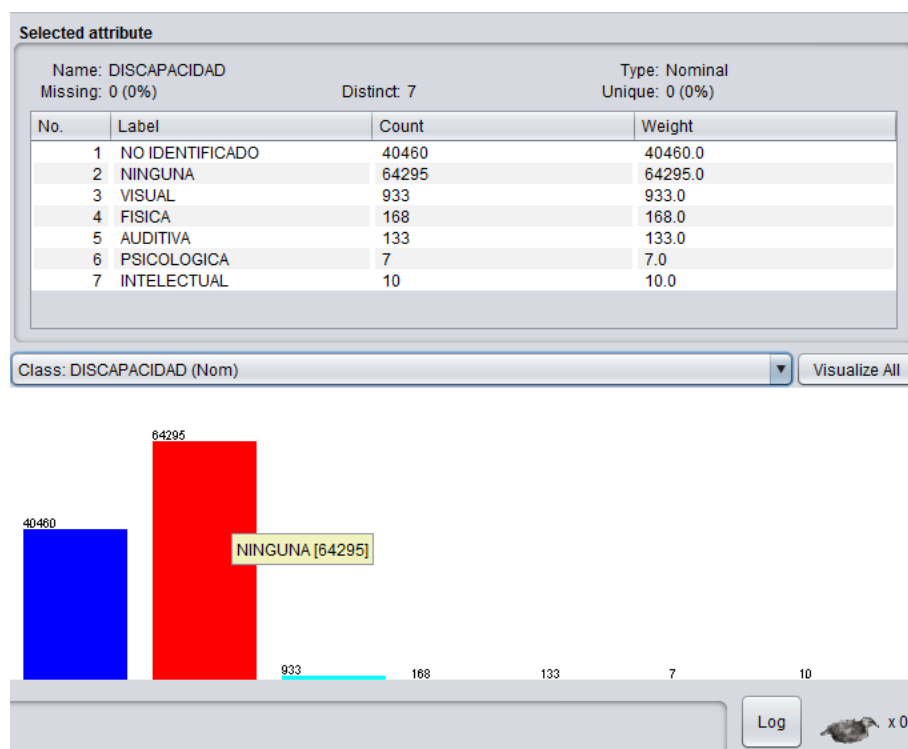
**Elaborado:** Investigador

A continuación se detalla el orden de deserción por Etnia:

1. Mestiza 61863
2. Ninguno 44460
3. Indígena 1880
4. Blanca 1414
5. Montubia 125
6. Afroamericana 264

## INCIDENCIA DE DESERCIÓN POR DISCAPACIDAD

Como se muestra en la Figura 66, los estudiantes que no poseen ningún tipo de discapacidad presentan un índice mayor de deserción sin considerar que un gran número de estudiantes no especifican si poseen o no algún tipo de discapacidad.



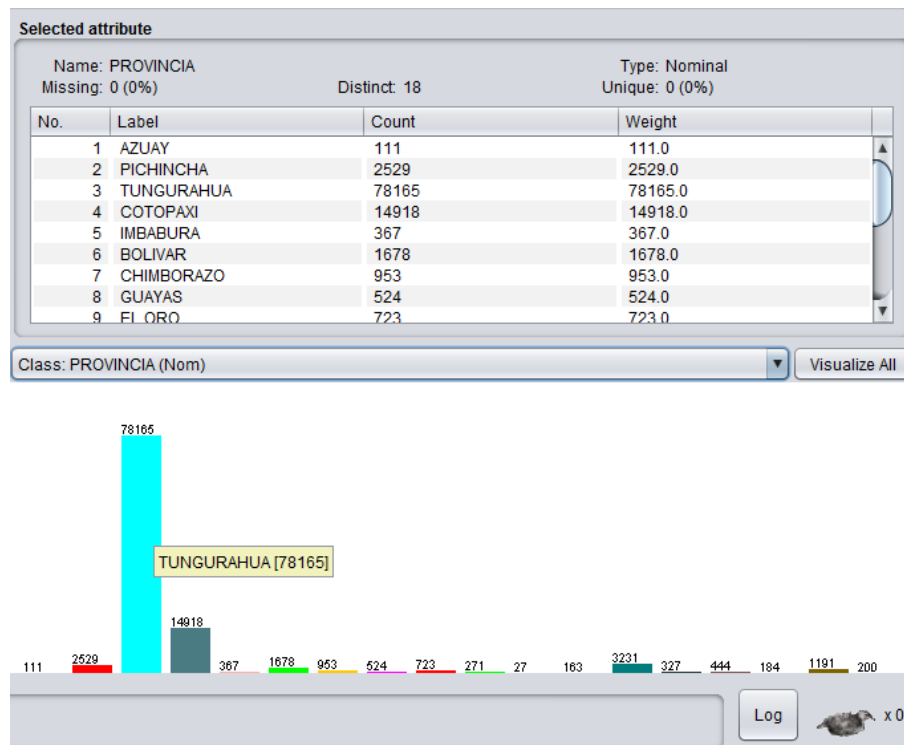
**Figura 73.** INCIDENCIA DE DESERCIÓN POR DISCAPACIDAD  
**Elaborado:** Investigador

A continuación de detalla el orden de deserción por Discapacidad:

1.	Ninguna	64295
2.	No identificado	40460
3.	Visual	933
4.	Física	168
5.	Auditiva	133
6.	Intelectual	10
7.	Psicológica	7

## INCIDENCIA DE DESERCIÓN POR PROVINCIA

Como se muestra en la Figura 67, la provincia de Tungurahua y seguida por la provincia de Cotopaxi

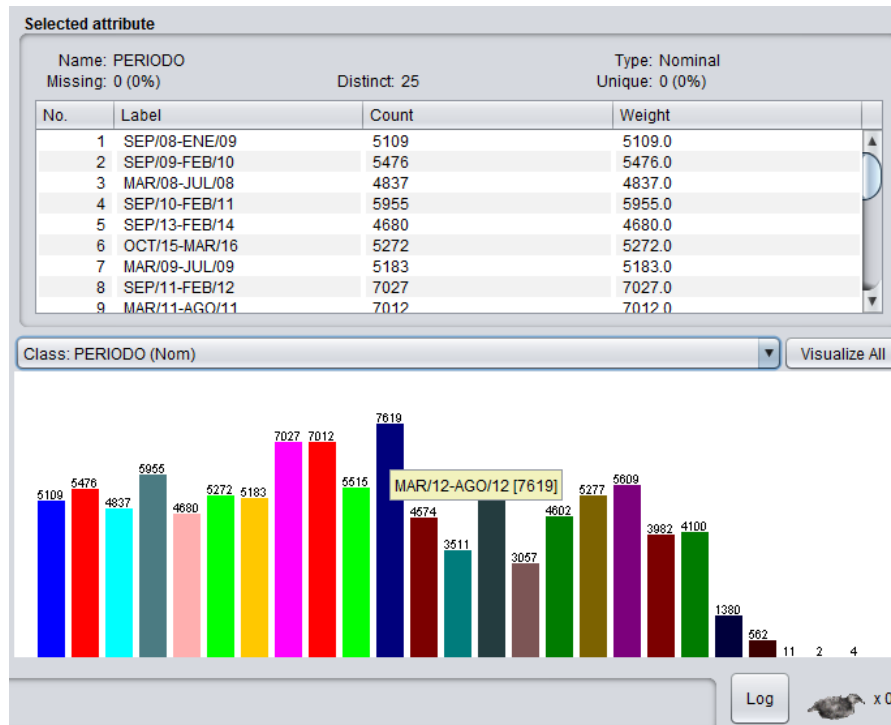


**Figura 74.** INCIDENCIA DE DESERCIÓN POR PROVINCIA

**Elaborado:** Investigador

## INCIDENCIA DE DESERCIÓN POR PERÍODO

Como se muestra en la Figura 68, el período que mayor índice de retención es el Marzo - Agosto 2012



**Figura 75.** INCIDENCIA DE DESERCIÓN POR PERÍODO

**Elaborado:** Investigador

## 6.8 Conclusiones

Al finalizar la propuesta se concluye que:

- La minería de datos permitió descubrir patrones de deserción estudiantil en la Universidad Técnica de Ambato.
- El desarrollo de este trabajo de investigación permitió seleccionar la información relevante de los estudiantes desertores de la Universidad Técnica de Ambato, en un almacén de datos que permitirá implementar nuevos modelos para la detección de patrones de Deserción Estudiantil.
- El proceso de ETL( Extracción, Transformación y Carga) , permitió transformar todos los datos crudos en datos limpios para la Minería de Datos
- El uso de Weka (Herramienta Software Libre) permitió descubrir conocimiento de Base de Datos para la Detección de Patrones de Deserción Estudiantil.
- La aplicación del Algoritmo K- Media permitió descubrir Patrones de Deserción Estudiantil, identificando categorías tanto académicas como sociales.

## **6.9. Recomendaciones**

- Se recomienda integrar datos correspondientes a la Dirección de Bienestar Estudiantil (DIBESAU), con el fin de disponer de datos socio-económicos de los estudiantes de la Universidad Técnica de Ambato.
- Se recomienda se alimente de información socio-económica de los estudiantes en el proceso de matrícula, para mantener actualizada la información de los estudiantes de la Universidad Técnica de Ambato.
- Se recomienda profundizar en métodos de minería de datos para encontrar el algoritmo más óptimo para el descubrimiento de conocimiento de datos académicos de la institución.
- Una vez detectados los patrones de deserción estudiantil en la Universidad Técnica se recomienda difundir la información detectada a las Autoridades, con el fin de establecer estrategias administrativas para mitigar de alguna manera los factores que afectan directamente en la deserción estudiantil.



## BIBLIOGRAFÍA

- ✓ Bernabeu, R. D. (2010). *HEFESTO. DATA WAREHOUSING: Investigación y Sistematización de Conceptos. HEFESTO: Metodología para la Construcción de un Data Warehouse*. Córdoba.
- ✓ Carrera Sepúlveda, S., Varas Contreras, M., & Urrutia Sepúlveda, A. (17 de Agosto de 2010). *Transformación de Esquemas Multidimensionales Difusos desde el nivel Conceptual al nivel Lógico*. (C. V. González, Ed.) *Ingeniare. Revista Chilena de Ingeniería*, 18(2), 165-175.
- ✓ Cobos, C., Zuñiga, J., Guarín, J., León, E., & Mendoza, M. (2010). *CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos* (Vol. 30). *Ingeniería e Investigación*.
- ✓ Corso, C. L. (2009). *Aplicación de algoritmos de clasificación supervisada usando Weka*. Córdoba: *Universidad Tecnológica Nacional, Facultad Regional Córdoba*.
- ✓ Curto Díaz, J. (2012). *Introducción al Business Intelligence* (ISBN: 8497889797 ed.). Barcelona: UOC.
- ✓ Galán Cortina, V. (2016). *Aplicación de la metodología CRISP-DM a un proyecto de minería de datos en el entorno universitario* (Bachelor's thesis).
- ✓ Gallardo Arancibia, J. (2009). *Metodología para la definición de requisitos en Proyectos de Data Mining*.
- ✓ Jiménez, M. G., & Sierra, A. Á. (2010). *Análisis de datos en WEKA—pruebas de selectividad*. línea] disponible en <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/28.pdf>.

- ✓ Osvaldo M. Sposito, & Martín E. Etcheverry, & Hugo L. Ryckeboer, & Julio Bossero, (), *Aplicación de técnicas de minería de datos para la evaluación del rendimiento académico y la deserción estudiantil*
- ✓ Pereira, *Detección de Patrones de Bajo Rendimiento Académico y Deserción Estudiantil con Técnicas de Minería de Datos*, 2009
- ✓ Pérez López, C. (2013). *Modelos Con Herramientas de Minería de Datos. Ejercicios Con Modeler y SAS Miner.*
- ✓ Pérez López, C., & Santín González, D. (2007). *Minería de datos: técnicas y herramientas* (Vol. 2). Madrid: Thomson.
- ✓ Pineda-Báez, C., Pedraza-Ortiz, A., & Moreno, I. D. (2011). Efectividad de las estrategias de retención universitaria: la función del docente. *Educación y educadores*, 14(1), 119-135.
- ✓ Proceso de evaluación institucional, Ceaaces.gob.ec,  
URL: <http://www.ceaaces.gob.ec/sitio/proceso-de-recategorizacion-institucional-2/>
- ✓ Ramez, E., & Shamkant, B. N. (2010). *Fundamentals of Database Systems*. Boston: Pearson Education.
- ✓ Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, *From Data Mining to Knowledge Discovery in Databases*, AI Magazine Volume 17 Number 3 (1996)
- ✓ Timarán, R., & Calderón, A. (2013). *La minería de datos como método innovador para la detección de patrones de deserción estudiantil en programas de pregrado en Instituciones de Educación Superior. WEEF* , 2-9.

- ✓ Viteri Castro, D. A., & Uquillas Narváez, M. A. (2011). *Estudio sobre la deserción estudiantil en Pontificia Universidad Católica del Ecuador-Matriz, en los niveles 1ro, 2do y 3ero de todas las Facultades y Escuelas del primer semestre del año académico 2007-2008* (Bachelor's thesis, QUITO/PUCE/2011).

## ANEXOS

### ENCUESTA DIRIGIDA A LOS DOCENTES RESPONSABLES UPE DE CARRERAS DE LA UNIVERSIDAD TÉCNICA DE AMBATO

**Objetivo:** Obtener información acerca del Indicador Tasa de Retención Estudiantil.

**Instrucciones:** Maque con una X la respuesta

**Preguntas:**

<b>Preguntas:</b>	<b>SI</b>	<b>NO</b>
1) ¿El Sistema Informático de la Universidad Técnica de Ambato genera algún reporte la Tasa de Retención Estudiantil?		
2) ¿Se realiza un análisis de la Tasa de Retención Estudiantil?		
3) ¿Existen estrategias para maximizar el índice de Retención Estudiantil?		
4) ¿Es importante para la institución, la Retención Estudiantil?		
5) ¿El Sistema Informático de la Universidad Técnica de Ambato permite obtener el índice de Deserción Estudiantil?		
6) ¿Es importante para la institución identificar los Patrones de Deserción Estudiantil?		
7) ¿Los reportes de Patrones de Deserción Estudiantil permitirán establecer procesos para maximizar la Tasa de Retención Estudiantil?		
8) ¿La generación de reportes de Patrones de Deserción Estudiantil permitirá una oportuna Toma de Decisiones en la Universidad Técnica de Ambato?		

**¡...GRACIAS POR SU COLABORACIÓN...!**