UNIVERSIDAD
NACIONAL
DE COLOMBIA

# Descubrimiento de patrones en interacciones entre estudiantes y plataformas virtuales de educación mediante el uso de analíticas de aprendizaje

**Mauricio Giraldo Ocampo**

Universidad Nacional de Colombia

Facultad de Minas

Departamento de Ciencias de la Computación y de la Decisión

Medellín, Colombia

2017

# Descubrimiento de patrones en interacciones entre estudiantes y plataformas virtuales de educación mediante el uso de analíticas de aprendizaje

**Mauricio Giraldo Ocampo**

Tesis de investigación presentada como requisito parcial para optar al título de:

**Magister en Ingeniería de Sistemas**

Director (a):

**Ph.D. Néstor Darío Duque Méndez**

Codirector (a):

**Ph.D. Demetrio Arturo Ovalle Carranza**

Línea de Investigación:

**Inteligencia Artificial**

Grupo de Investigación:

**Grupo de Ambientes Inteligentes Adaptativos – GAIA**

**Grupo de Investigación y Desarrollo en Inteligencia Artificial – GIDIA**

Universidad Nacional de Colombia

Facultad de Minas

Departamento de Ciencias de la Computación y de la Decisión

Medellín, Colombia

2017

*Dedicatoria*

*En estos años de trabajo mi familia y mi novia han sido la fuerza para continuar luchando por mis metas a pesar de los momentos de cansancio y desmotivación; por eso a ellos les dedico esta tesis.*

# Agradecimientos

Agradezco a mis padres por su apoyo absoluto y por ser mis primeros formadores, por sus consejos y cariño. A mi novia, quien ha sido más que incondicional en este proceso, ella ha luchado junto a mí por el logro de este objetivo, con gran paciencia y recordándome el horizonte cuando me desenfocaba.

Mi gratitud al profesor Néstor Darío Duque porque durante todo el tiempo de mi maestría, incluso desde el pregrado, ha sido un guía y me ha acompañado no solo desde la parte académica sino también brindándome una formación integral.

Al profesor Demetrio Arturo Ovalle, mi codirector, agradezco por sus horas de dedicación y por sus consejos en el desarrollo de esta tesis, también por acompañar mí proceso investigativo.

# Resumen

En esta tesis se plantea un acercamiento a un modelo de descubrimiento de patrones basado en la observación y análisis de datos educativos y de las interacciones existentes entre los estudiantes y las plataformas virtuales de aprendizaje con el uso de minería de datos educativos y de analíticas de aprendizaje, que permita hacer algunas recomendaciones para fortalecer el proceso enseñanza-aprendizaje, de manera que este se pueda adaptar y posiblemente personalizar de acuerdo a las características propias de los estudiantes y de sus interacciones. El modelo fue validado con un caso de estudio donde se contó con datos académicos de estudiantes de la Universidad Nacional de Colombia Sede Manizales para los periodos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015 y sus interacciones en plataformas virtuales de aprendizaje para el segundo semestre de 2013 y primer semestre de 2014.

**Palabras clave:** Analíticas de Aprendizaje, Educación Virtual, Minería de Datos, Plataformas Virtuales de Educación, Proceso Enseñanza-Aprendizaje, Reconocimiento de Patrones.

# Abstract

This thesis proposes an approach to a model of discovery of patterns based on observation and analysis of educational data and the interactions between the students and virtual platforms of learning with the use of educational data mining and analytical learning, allowing to make some recommendations to strengthen the teaching and learning process, so that it can adapt and possibly customized according to the characteristics of the students and their interactions. The model was validated with a case study which included academic data of students of the Universidad Nacional de Colombia in Manizales for the periods between the first semester of 2009 and the first semester of 2015 and their interactions in virtual learning platforms for the second semester of 2013 and first semester of 2014.

**Keywords:** Learning Analytics, Data Mining, Pattern Recognition, Teaching-Learning Process, Virtual Education, Virtual Education Platform.

# Contenido

**Pág.**

**Resumen** ..... **IX**

**Abstract** ..... **X**

**Lista de figuras** ..... **XIV**

**Lista de tablas** ..... **XVI**

**Introducción** ..... **1**

**1. Presentación de la tesis** ..... **3**
1.1 Motivación ..... 3
1.2 Problemática ..... 4
1.3 Preguntas de investigación ..... 5
1.4 Objetivos de la tesis ..... 6
1.4.1 Objetivo general ..... 6
1.4.2 Objetivos específicos ..... 6
1.5 Alcance de la tesis ..... 7
1.6 Metodología ..... 8
1.7 Cumplimiento de los objetivos ..... 9
1.8 Principales contribuciones logradas ..... 11
1.9 Productividad Académica ..... 12
1.9.1 Artículos en revistas ..... 12
1.9.2 Artículos en conferencias ..... 12
1.9.3 Capítulos de libro ..... 13
1.9.4 Participación en proyectos de investigación ..... 13
1.10 Organización del documento ..... 13

**2. Marco teórico** ..... **15**
2.1 e-Learning y Plataformas Virtuales de Aprendizaje ..... 15
2.1.1 Virtual Learning Environment – VLE ..... 17
2.1.2 Learning Management System – LMS ..... 17
2.1.3 Características de los LMS ..... 19
2.1.4 Tipos de LMS ..... 21
2.1.5 Tipos de actividades educativas en algunas plataformas virtuales de aprendizaje ..... 21
2.2 Sistemas de información académicos ..... 24
2.3 Analíticas de aprendizaje ..... 24
2.3.1 Objetivos de las analíticas de aprendizaje ..... 27
2.4 Minería de Datos ..... 27

2.4.1 Algunas técnicas de Minería de Datos ..... 30
2.4.2 Minería de Datos Educativos (EDM) ..... 32
2.5 Conclusiones del capítulo ..... 34

**3. Estado del arte ..... 35**
3.1 Primer grupo: artículos de revisión – contextualización ..... 35
3.2 Segundo grupo: artículos con modelos propuestos ..... 43
3.3 Conclusiones de capítulo ..... 57

**4. Modelo Propuesto ..... 59**
4.1 Componente 1: Fuentes de datos ..... 59
4.1.1 Datos académicos ..... 60
4.1.2 Datos de las interacciones de los estudiantes con las plataformas virtuales de aprendizaje ..... 60
4.2 Componente 2: Proceso de ETL ..... 61
4.3 Componente 3: Bodega de datos académica ..... 61
4.3.1 Modelo de datos del estudiante ..... 62
4.3.2 Modelo de datos académicos ..... 62
4.3.3 Modelo de datos de las interacciones con LMS ..... 62
4.4 Componente 4: Aplicación de analíticas de aprendizaje ..... 63
4.4.1 El enfoque descriptivo ..... 64
4.4.2 El enfoque diagnóstico ..... 64
4.4.3 El enfoque predictivo ..... 65
4.4.4 El enfoque prescriptivo ..... 65
4.5 Conclusiones del capítulo ..... 66

**5. Validación del modelo ..... 67**
5.1 Caso de estudio ..... 67
5.2 Componente 1: Fuentes de datos ..... 68
5.2.1 Datos académicos ..... 68
5.2.2 Datos de las interacciones de los estudiantes en las plataformas Blackboard y Moodle ..... 74
5.2.3 Diagnóstico del uso de plataformas virtuales de aprendizaje ..... 76
5.3 Componente 2: Proceso de ETL ..... 77
5.4 Componente 3: Bodega de datos académica ..... 81
5.5 Componente 4: Aplicación de analíticas de aprendizaje ..... 85
5.5.1 Enfoque descriptivo ..... 85
5.5.2 Enfoque diagnóstico ..... 87
5.5.3 Enfoque predictivo ..... 105
5.5.4 Enfoque prescriptivo ..... 113
5.6 Conclusiones del capítulo ..... 114

**6. Conclusiones y trabajos futuros ..... 117**
6.1 Conclusiones ..... 117
6.2 Trabajos futuros ..... 119

**A. Anexo: Encuesta para el diagnóstico del uso de plataformas virtuales de aprendizaje 122**

**B. Anexo: Diseño de la aplicación para realizar ETL ..... 133**

**C. Anexo: Diccionario de datos de la bodega de datos académica .....141**

**Bibliografía .....167**

# Lista de figuras

**Pág.**
Figura 2-1: Conceptos principales ..... 15
Figura 2-2: Etapas del proceso de descubrimiento del conocimiento en bases de datos. (KDD). Tomado de (Quiroz Gil & Valencia, 2012) ..... 28
Figura 2-3: Minería de datos educativos y áreas relacionadas. Adaptada de (Romero & Ventura, 2013) ..... 32
Figura 3-1: Factores que impulsan el desarrollo de LA. Adaptado de (Ferguson, 2012) . 37
Figura 3-2: Flujo del modelo de analíticas de aprendizaje. Adaptado de (Picciano, 2012) ..... 38
Figura 3-3: Niveles de aplicación de las analíticas de aprendizaje. Adaptado de (MacNeill, Campbell, & Hawksey, 2014). ..... 40
Figura 3-4: Tendencias en Minería de Datos Educativos. Adaptado de (Peña-Ayala, 2014). ..... 42
Figura 3-5: Framework para Datasets Educativos. Tomado de (Verbert et. al., 2012)..... 45
Figura 3-6: Proceso de generación de indicadores. Tomado de (Dyckhoff et. al. 2012) .. 47
Figura 3-7: Arquitectura de LeMo. Adaptada de (Beuster et. al. 2013) ..... 48
Figura 3-8: Framework para LA en mundos educativos virtuales 3D. Tomado de (Fernández-Gallego et. al., 2013) ..... 51
Figura 3-9: Arquitectura del módulo ALAS-KA ..... 55
Figura 4-1: Modelo propuesto ..... 60
Figura 4-2: Enfoques de las analíticas de aprendizaje ..... 64
Figura 5-1: Proceso de ETL implementado ..... 78
Figura 5-2: Interfaz de la herramienta de ETL desarrollada ..... 79
Figura 5-3: Modelo de datos de la bodega de datos académica ..... 82
Figura 5-4: Distribución de estudiantes por programa académico ..... 89
Figura 5-5: Distribución de estudiantes por género ..... 90
Figura 5-6: Distribución de estudiantes por estrato ..... 91
Figura 5-7: Promedio de duración en horas de las interacciones por mes ..... 98
Figura 5-8: Promedio de duración en horas, minutos y segundos de las interacciones por mes ..... 98
Figura 5-9: Cantidad de interacciones por plan de estudios ..... 99
Figura 5-10: Cantidad de interacciones por Facultad ..... 101
Figura 5-11: Promedio de PAPA por plan de estudios ..... 102
Figura 5-12: Promedio del PAPA por Facultad ..... 103
Figura 5-13: Promedio de PAPA por año ..... 104
Figura 5-14: Promedio de PAPA por semestre ..... 105

Figura 5-15: Correlación entre las variables num_interacciones y duracion_acumulada .....108
Figura 5-16: Distribución de los Clúster respecto al delta_papa y el promedio_tiempo_interacciones para K=2 .....110
Figura 5-17: Distribución de los Clúster respecto al delta_papa y el promedio_tiempo_interacciones para K=3 .....111
Figura 5-18: Distribución de los Clúster respecto al delta_papa y el promedio_tiempo_interacciones para K=4 .....112
Figura 6-1: Nivel de formación de los estudiantes encuestados .....123
Figura 6-2: Conocimiento de la existencia de LMS .....123
Figura 6-3: Uso de plataformas virtuales en los programas académicos .....124
Figura 6-4: Plataformas usadas por los estudiantes en la Universidad .....124
Figura 6-5: Actividades realizadas por los estudiantes en los LMS.....125
Figura 6-6: Frecuencia de uso de LMS en los programas académicos.....125
Figura 6-7: Uso de LMS en otros procesos de formación .....126
Figura 6-8: Plataformas usadas en otros procesos de formación .....126
Figura 6-9: Frecuencia de uso de los LMS en otros procesos de formación.....127
Figura 6-10: Conocimiento de la existencia de LMS .....128
Figura 6-11: Uso de los LMS por parte de los docentes .....128
Figura 6-12: Plataformas usadas por los docentes.....129
Figura 6-13: Actividades que emplean los docentes en sus cursos en los LMS .....130
Figura 6-14: Frecuencia de uso de los LMS por parte de los docentes .....130
Figura 6-15: Estrategias de mantenimiento de cursos .....131
Figura 6-16: Modelo de la bodega de datos académica .....141

# Lista de tablas

**Pág.**
Tabla 1-1: Metodología: Fases, objetivos y actividades ..... 8
Tabla 2-1: Comparativo entre LMS y LCMS. Tomada de (Boneu, 2007)..... 18
Tabla 2-2: Comparativo entre Blackboard y Moodle en cuanto a herramientas de soporte. Adaptado de (Kumar, Gankotiya, & Dutta, 2011) ..... 22
Tabla 2-3: Comparación de algunos LMS en cuanto a herramientas de aprendizaje. Adaptada de (Kumar et al., 2011) ..... 22
Tabla 2-4: Analíticas de aprendizaje y académicas. Adaptado de (Siemens et al., 2011)25
Tabla 2-5: Clasificación de las técnicas de minería de datos. Tomada de (Moreno García, et.al., 2001)..... 29
Tabla 2-6: Tipos de algoritmos de MD según su tarea. Construido a partir de (Microsoft, 2016) ..... 30
Tabla 2-7: Analíticas de aprendizaje Vs. EDM. Tomada de (Rojas Castro, 2015)..... 33
Tabla 3-1: Las variables extraídas de uso de Moodle por parte de los estudiantes. Adaptada de (Kotsiantis et. al., 2014) ..... 52
Tabla 3-2: Comparativo de trabajos revisados en el segundo grupo ..... 56
Tabla 5-1: Resumen de los registros presentes en el SIA ..... 68
Tabla 5-2: Identificación de variables del SIA..... 69
Tabla 5-3: Consolidado de la base de datos de Blackboard..... 74
Tabla 5-4: Número de registros cargados en la bodega de datos ..... 80
Tabla 5-5: Indicadores de estadística descriptiva para los datos académicos..... 85
Tabla 5-6: Indicadores de estadística descriptiva para las interacciones ..... 86
Tabla 5-7: Distribución de estudiantes por programa académico ..... 88
Tabla 5-8: Distribución de estudiantes por género ..... 89
Tabla 5-9: Distribución de estudiantes por estrato ..... 91
Tabla 5-10: Número de estudiantes por rango de PAPA para el 2015-I ..... 92
Tabla 5-11: Clasificación de estudiantes por promedio y carrera para el 2015-I..... 93
Tabla 5-12: Ingreso de estudiantes en los últimos 10 años..... 94
Tabla 5-13: Clasificación de acuerdo al número de interacciones..... 95
Tabla 5-14: Número de estudiantes por número de cursos..... 96
Tabla 5-15: Relación plan de estudios y nombres de programa..... 99
Tabla 5-16: Distribución de planes por Facultad ..... 101
Tabla 5-17: Centroides en el Clustering sobre el dataset inicial ..... 107
Tabla 5-18: Correlaciones entre los atributos del dataset..... 108
Tabla 5-19: Resultados del K-means para K=2..... 109

Tabla 5-20: Resultados del K-means para K=3 .....110
Tabla 5-21: Resultados del K-means para K=4 .....111

# Introducción

La utilización de plataformas virtuales de aprendizaje ha generado grandes cambios en la manera de ofrecer, direccionar y ejecutar los procesos de enseñanza-aprendizaje. Lo anterior ha conllevado a que las instituciones de educación den un giro a la formulación de sus programas académicos y currículos, incluyendo nuevas herramientas, cursos y ayudas pedagógicas basadas en plataformas de educación virtual. Todo esto ha generado la necesidad de empezar a estudiar las diferentes formas de aprender de los estudiantes y a desarrollar estrategias que permitan adaptar los procesos educativos.

La educación virtual es una tendencia cada vez más acogida en las diferentes instituciones de educación, esto debido a que permite llegar a un mayor número de personas con los mismos recursos, así como también atender al público que tiene limitaciones para asistir a una institución presencial. Por medio de las plataformas de educación virtual, se ofrecen cursos complementarios e incluso acciones de formación completas, como carreras profesionales; lo anterior hace que el número y tipo de interacciones que se da entre dichas plataformas y los estudiantes sea cada vez más creciente y variado.

Existiendo una gran variedad en las interacciones que pueden darse entre cada estudiante y una misma plataforma virtual de aprendizaje, se hace necesario analizar dichas interacciones para lograr establecer y configurar las herramientas que se adaptan mejor a cada estudiante y así poder proporcionar los recursos que contribuyan a éste en su proceso de aprendizaje.

El volumen de datos generado por las interacciones entre los estudiantes y las plataformas virtuales de aprendizaje se convierte en un gran material de estudio, que posibilita la aplicación de técnicas y herramientas de análisis de datos, orientadas a extraer modelos o patrones desde los datos que permitan tomar decisiones tanto académicas como administrativas.

La presente tesis muestra el desarrollo de un modelo para el descubrimiento de patrones a partir del análisis de datos académicos y datos de las interacciones entre los estudiantes y las plataformas virtuales de aprendizaje aplicando analíticas de aprendizaje. Para lo anterior se hizo el diseño de una bodega de datos académica donde se combinan estas dos fuentes de datos y se utiliza un esquema hibrido para poder cubrir todas las necesidades. Se aplicaron análisis a partir de algoritmos y herramientas existentes y se reportan los resultados obtenidos para un caso de estudio que comprende varios periodos académicos de la Universidad Nacional de Colombia – Sede Manizales.

# 1. Presentación de la tesis

En este capítulo se presenta una descripción general de la tesis, esto por medio de la motivación, problemática, preguntas de investigación, objetivos, alcance, metodología y logros alcanzados.

## 1.1 Motivación

Desde hace varios años, como parte de las líneas de investigación abordadas por los grupos de investigación a los que hacen parte el investigador y sus asesores (GAIA – Grupo de Investigación en Ambientes Inteligentes Adaptativos y GIDIA – Grupo de Investigación y Desarrollo en Inteligencia Artificial), se ha trabajado en temas relacionados a la educación virtual y a la inteligencia artificial en sistemas adaptativos; de allí nace el interés inicial de abordar el problema de investigación que se trata en esta tesis. Lo anterior, asociado a una revisión bibliográfica que resaltó la necesidad de explorar más en temas relacionados con los grandes volúmenes de datos producidos por las herramientas de educación virtual, y mostró en las analíticas de aprendizaje y la minería de datos educativos un camino a recorrer para tal fin.

Cabe anotar en este apartado, una breve descripción del estado actual del problema que se profundizará más adelante. Una de las formas que permite adaptar y personalizar ciertas actividades y contenidos desarrollados para el proceso de aprendizaje seguido por los estudiantes a través de las plataformas de educación virtual, es el estudio de las interacciones que surgen en este proceso y la extracción de información a partir de estas y de los datos propios del estudiante y de su relación con la institución educativa. Lo anterior llevó a identificar una oportunidad en la aplicación de analíticas de aprendizaje y minería de datos educativos sobre los datos resultantes de dichas interacciones, entre los estudiantes y las plataformas virtuales, con el fin de ayudar a determinar posibles tendencias y patrones de comportamiento.

## 1.2 Problemática

En este momento la personalización de los cursos para los estudiantes de diferentes niveles de educación representa un reto para la comunidad investigativa, que busca la generación de modelos encaminados a la caracterización de los procesos de enseñanza-aprendizaje, mediante el análisis de grandes cantidades de datos, donde además de identificar información relevante, se busca extraer patrones que ayuden en la formulación de dichos modelos de personalización y que se conviertan en una base sobre la cual se puedan tomar decisiones.

El análisis de los datos y de la información recolectada sobre las interacciones que se dan entre los estudiantes y las plataformas virtuales permite, entre tantas cosas, apoyar la generación de estrategias para adaptar los contenidos y las actividades educativas, procurando así que cada estudiante reciba la enseñanza y los recursos que son necesarios (Siemens et al., 2011). De allí también, que no sólo se tengan que analizar las interacciones resultantes de los procesos de enseñanza-aprendizaje, sino que sea relevante tener en cuenta las recomendaciones resultantes, de tal manera que sea de utilidad el trabajo realizado (Ferguson & Buckingham Shum, 2012). Como ejemplo de esto, Apple e IBM, introdujeron el IBM Watson Element for Educators, la cual es una aplicación que de manera intuitiva mide el interés, la experiencia de aprendizaje y el comportamiento de los estudiantes en el proceso educativo (Innovation Group, 2016).

En diferentes áreas del conocimiento se utilizan grandes cantidades de datos para ser analizados y llevar a cabo procesos de toma de decisiones, un ejemplo de esto es la inteligencia de negocios (Business Intelligence) donde se utilizan datos empresariales para generar recomendaciones y apoyar a la dirección a la hora de tomar las decisiones. Así mismo, con los datos educativos se pueden llevar a cabo este tipo de procesos (Long & Siemens, 2011).

A partir de lo anterior, se puede evidenciar la importancia que tiene el análisis de los datos para llegar a obtener una posible adaptación en el trascurso de los procesos de enseñanza-aprendizaje; además, estos análisis realizados en un periodo de tiempo prudente y contando con los datos necesarios, pueden ayudar a realizar una mejor

personalización, de manera que la adaptación se realice durante el proceso educativo y no después.

Es por ello que el análisis de los datos en determinado momento, debe estar en concordancia con los objetivos curriculares, con la práctica docente llevada a cabo y los resultados esperados de la actividad de enseñanza-aprendizaje, llevando a proporcionar interpretaciones más acertadas acerca de los resultados obtenidos (Siemens et al., 2011).

Complementando lo anterior, Long & Siemens (2011) aseguran que los datos producidos por el alumno, son una valiosa fuente de información sobre lo que está sucediendo en el proceso de enseñanza-aprendizaje y sugieren formas en que los educadores pueden realizar mejoras al proceso, de manera que aumenten los niveles de comprensión, interés y demás aspectos que sean benéficos para el estudiante. También este análisis puede dar un panorama sobre los estudiantes que están en peligro de deserción escolar o requieren apoyo para aumentar su éxito.

En la educación superior se reúnen grandes cantidades de datos sobre los estudiantes y su proceso; sin embargo, estos no siempre se usan eficientemente en procesos como la planificación y asignación de recursos a las diferentes actividades educativas virtuales (Long & Siemens, 2011). Por esto se planteó trabajar en el desarrollo de un modelo que permita tomar diferentes fuentes de datos académicos y de interacciones en plataformas virtuales, llevarlos a un esquema de almacenamiento centralizado y a partir de allí realizar algunos análisis para extraer información correspondiente a posibles tendencias o patrones de comportamiento que puedan ser posteriormente usados en la planeación de estrategias de mejoramiento o adaptación de procesos de enseñanza-aprendizaje.

## 1.3 Preguntas de investigación

De acuerdo al problema de investigación planteado y después de la revisión bibliográfica donde se identificaron las fortalezas, limitaciones y oportunidades de trabajos previos, surgieron las siguientes preguntas de investigación:

¿Es posible elaborar un modelo de descubrimiento de patrones sobre datos educativos y de las interacciones de los estudiantes con las plataformas virtuales utilizando analíticas de aprendizaje?

¿Cómo se pueden incorporar técnicas de minería de datos en el modelo basado en analíticas de aprendizaje, para poder hacer recomendaciones que mejoren los procesos educativos apoyados en plataformas virtuales?

¿Se cuenta con algún modelo o plataforma que provea guías sobre la implementación de analíticas de aprendizaje para el descubrimiento de patrones en procesos de educación virtual?

## 1.4 Objetivos de la tesis

A continuación, se presenta el objetivo general de esta tesis, el cual surge a partir de la problemática y preguntas de investigación. A su vez, el objetivo general fue desglosado por medio de los objetivos específicos.

### 1.4.1 Objetivo general

Proponer un modelo basado en analíticas de aprendizaje sobre datos educativos e interacciones de los estudiantes con las plataformas virtuales de aprendizaje para apoyar la personalización de actividades educativas.

### 1.4.2 Objetivos específicos

- Caracterizar los elementos asociados al modelo, incluyendo analíticas de aprendizaje, minería de datos, ambientes virtuales de aprendizaje, sistemas de gestión de aprendizaje y personalización del aprendizaje.

- Identificar los distintos enfoques que se pueden aplicar por medio de las analíticas de aprendizaje, que se adecuan al modelamiento de patrones comportamentales en procesos educativos que son realizados a través de plataformas virtuales de enseñanza-aprendizaje.

- Diseñar un modelo que incluya la aplicación de analíticas de aprendizaje a través de las diferentes fases que se comprenden desde la extracción de datos hasta las recomendaciones de mejora de los procesos educativos virtuales.

- Implementar y validar un prototipo del modelo basado en analíticas de aprendizaje propuesto, haciendo su aplicación en un caso de estudio.

## 1.5 Alcance de la tesis

Como parte del alcance de esta tesis de maestría se plantean las siguientes consideraciones:

- El modelo de descubrimiento de patrones sobre datos educativos e interacciones entre estudiantes y plataformas virtuales de aprendizaje, abarcará la aplicación de algunas técnicas de Minería de Datos, algunas de las cuales son descritas en el marco teórico.

- Para el proceso de extracción de conocimiento (KDD), se emplearán algoritmos de minería de datos, los cuales serán buscados entre los existentes, no se creará un algoritmo particular para el modelo.

- Las recomendaciones generadas a partir de la aplicación del modelo de descubrimiento de patrones no llegarán a ser probadas mediante su aplicación, simplemente se evaluará si se ajustan a la información recolectada de las interacciones. La validación práctica en entornos educativos reales de dichas recomendaciones hace parte de un trabajo futuro.

Se debe aclarar que los puntos que sobrepasen el alcance de esta tesis serán considerados como trabajos futuros.

## 1.6 Metodología

La metodología propuesta para dar desarrollo a la tesis se divide en una serie de actividades agrupadas en fases que corresponden a la ejecución de cada uno de los objetivos planteados. La Tabla 1-1 contiene la descripción y actividades de cada fase.

**Tabla 1-1:** Metodología: Fases, objetivos y actividades

| OBJETIVO | FASE | ACTIVIDADES |
|---|---|---|
| Caracterizar los elementos asociados al modelo, incluyendo analíticas de aprendizaje, minería de datos, ambientes virtuales de aprendizaje, sistemas de gestión de aprendizaje y personalización del aprendizaje. | **Fase 1:** Construcción del marco teórico y estado del arte. | **1.** Revisión de la literatura sobre educación virtual (e-learning), analíticas de aprendizaje, plataformas de educación virtual, minería de datos y personalización.<br>**2.** Revisión de la literatura y construcción del estado del arte sobre trabajos realizados en cuanto a la formulación de modelos o sistemas de descubrimiento de patrones en procesos educativos.<br>**3.** Comparación de las técnicas utilizadas para la construcción de modelos de reconocimiento de patrones. |
| Identificar los distintos enfoques que se pueden aplicar por medio de las analíticas de aprendizaje, que se adecuan al modelamiento de patrones comportamentales en procesos educativos que son realizados a través de plataformas virtuales de enseñanza-aprendizaje. | **Fase 2:** Caracterización de las analíticas de aprendizaje asociadas al modelo. | **1.** Revisión bibliográfica acerca de analíticas de aprendizaje empleadas en sistemas y procesos educativos.<br>**2.** Construcción de la estrategia de analíticas de aprendizaje a emplear en el modelo. |

| | | |
|---|---|---|
| Diseñar un modelo que incluya la aplicación de analíticas de aprendizaje a través de las diferentes fases que se comprenden desde la extracción de datos hasta las recomendaciones de mejora de los procesos educativos virtuales. | **Fase 3:** Diseño del modelo propuesto. | **1.** Concepción teórica y estructural del modelo basado en analíticas de aprendizaje.<br>**2.** Análisis y diseño del modelo de descubrimiento de patrones utilizando analíticas de aprendizaje.<br>**3.** Definir de la arquitectura adecuada para la integración del modelo con los mecanismos de recolección de los datos de las interacciones de los estudiantes en las plataformas de educación virtual. |
| Implementar y validar un prototipo del modelo basado en analíticas de aprendizaje propuesto, haciendo su aplicación en un caso de estudio. | **Fase 4:** Implementación y validación del prototipo. | **1.** Caracterización de la tecnología que se adecue al desarrollo e implementación del prototipo del modelo.<br>**2.** Desarrollo del modelo propuesto.<br>**3.** Búsqueda y selección del caso de estudio para la validación del prototipo del modelo.<br>**4.** Ejecución y validación del prototipo para el caso de estudio seleccionado.<br>**5.** Evaluación de los resultados arrojados en la ejecución del prototipo. |

## 1.7 Cumplimiento de los objetivos

Seguidamente se presenta de forma general las estrategias con las que se abordaron cada uno de los objetivos de la tesis con el fin de darles desarrollo y cumplimiento

**Objetivo 1:** *Caracterizar los elementos asociados al modelo, incluyendo analíticas de aprendizaje, minería de datos, ambientes virtuales de aprendizaje, sistemas de gestión de aprendizaje y personalización del aprendizaje.*

En la fase asociada a este objetivo se realizó una revisión bibliográfica tanto para la construcción del marco teórico, presentado en el capítulo 2, como para la identificación de

fortalezas, limitaciones y oportunidades de los trabajos relacionados que componen el estado del arte, el cual se presenta en el capítulo 3.

**Objetivo 2:** *Identificar los distintos enfoques que se pueden aplicar por medio de las analíticas de aprendizaje, que se adecuan al modelamiento de patrones comportamentales en procesos educativos que son realizados a través de plataformas virtuales de enseñanza-aprendizaje.*

Para dar cumplimiento a este objetivo se realizó una búsqueda de trabajos relacionados con el uso de analíticas de aprendizaje en el campo del e-learning, aplicadas a plataformas de educación virtual. A partir de esto se seleccionaron las más adecuadas para el desarrollo de esta tesis.

**Objetivo 3:** *Diseñar un modelo que incluya la aplicación de analíticas de aprendizaje a través de las diferentes fases que se comprenden desde la extracción de datos hasta las recomendaciones de mejora de los procesos educativos virtuales.*

En esta fase se hizo la definición de la estructura general para el modelo propuesto, partiendo de los avances logrados en los objetivos previos. El modelo se presenta en el capítulo 4, donde se describen los componentes asociados al modelo y la forma como se integraron cada uno de estos.

**Objetivo 4:** *Implementar y validar un prototipo del modelo basado en analíticas de aprendizaje propuesto, haciendo su aplicación en un caso de estudio.*

Con el desarrollo de la fase asociada a este objetivo se consiguió llevar el modelo propuesto a un prototipo y se hizo la validación del mismo por medio de la aplicación a un caso de estudio con datos de estudiantes de la Universidad Nacional de Colombia, sede Manizales para los periodos comprendidos desde el primer semestre de 2009 hasta el primer semestre de 2015 en cuanto a datos académicos y para las interacciones en las plataformas se contó con datos del segundo semestre de 2013 y primer semestre de 2014. Los datos fueron suministrados por la Dirección Académica de esta institución y se llevaron a un modelo de almacenamiento representado en una Bodega de Datos Académica

basada en un enfoque multidimensional en estrella ampliado. Los resultados obtenidos de los análisis aplicados de acuerdo a los enfoques definidos para las analíticas, muestran que para los datos suministrados se dificulta establecer patrones certeros para el comportamiento de las interacciones en relación con los datos de rendimiento académico. Sin embargo, se hizo una aplicación completa de los enfoques presentes en el modelo propuesto.

## 1.8 Principales contribuciones logradas

La investigación realizada en esta tesis se enfocó en brindar una contribución en la aplicación de analíticas de aprendizaje para el descubrimiento de patrones en datos educativos y en las interacciones de estudiantes con las plataformas virtuales de aprendizaje; para esto se propuso un esquema de almacenamiento representado en una bodega de datos académica y un modelo para análisis de los datos que permite extraer información relevante por medio de la aplicación de consultas relaciones y técnicas de minería de datos.

Las principales contribuciones de esta tesis son:

- Desde lo conceptual se desarrolló un modelo de almacenamiento representado en una bodega de datos académica, que contempla diversas dimensiones y un esquema híbrido, incluyendo el tratamiento de datos generados tanto por sistema de información académica como en plataformas virtuales de aprendizaje. Este esquema es bastante novedoso y lograr unir ventajas de un esquema relacional con uno multidimensional.
- Desde la investigación aplicada se implementó un prototipo basado en el modelo propuesto y se hizo la evaluación del desempeño de este por medio de la aplicación a un caso de estudio con registros de estudiantes de la Universidad Nacional de Colombia – Sede Manizales.
- Desde el aporte a la institución, el proceso de cargue y limpieza de datos académicos en la bodega de datos, crea una herramienta de apoyo a la Dirección Académica de la Sede con quienes se han socializado los avances y se trabaja en

un proyecto conjunto para la consolidación y aprovechamiento de los resultados obtenidos.

## 1.9 Productividad Académica

Como parte del proceso investigativo realizado durante el desarrollo de esta tesis, se logró difundir conocimiento con las siguientes publicaciones:

### 1.9.1 Artículos en revistas

- Méndez, N. D. D., Carranza, D. A. O., & Ocampo, M. G. (2015). Representación Ontológica de Perfiles de Estudiantes para la Personalización del Aprendizaje. Revista Educación en Ingeniería, 10(19), 105-115.
- Méndez, N. D. D., Carranza, D. A. O., & Ocampo, M. G. CSCL-Agents, as Useful Components to Handle Collaborative Activities within a VLE. Journal of Cases on Information Technology (JCIT). Enviado.

### 1.9.2 Artículos en conferencias

- Giraldo, M., Duque, N., & Hernández, E., “Sistema de Recuperación de Objetos de Aprendizaje usando un Chatterbot con Agentes Inteligentes en la Federación FROAC”, Evento: Primera Conferencia Ibero-Americana de Ambientes de Aprendizaje Futuros – CIAAF 2015. Oporto, Portugal. 2015

- Rodríguez, P., Giraldo, M., Tabares, V., Duque, N., & Ovalle, D. “Recommendation System of Educational Resources for a Student Group”. Event: International Conference on Practical Applications of Agents and Multi-Agent Systems – MASLE 2016. (pp. 419-427). Springer International Publishing. Sevilla, España. 2016

- Tabares, V., Duque, N., Giraldo, M., Rodríguez, P., & Ovalle, D. “Plataforma Adaptativa para la Búsqueda y Recuperación de Recursos Educativos Digitales”.

Event: XI Latin American Conference on Learning Objects and Technology, LACLO 2016. San Carlos, Costa Rica. 2016

- Giraldo, M., Hernández, E., Espinosa, D., Duque, N., & Moreno, J. "Application of a technical tool to support the inclusion process of people with visual impairment in an educational web platform". Event: XI Latin American Conference on Learning Objects and Technology, LACLO 2016. San Carlos, Costa Rica. 2016.

### 1.9.3 Capítulos de libro

Esquema de almacenamiento y recuperación de objetos u-CSCL. Un chatterbot con agentes inteligentes en la FROAC. 2016. Bogotá, Colombia. Entornos Ubicuos y Colaborativos (U_CSCL) para Ambientes de Enseñanza – Aprendizaje de Competencias Profesionales. ISBN: 978-958-8785-83-7. Editorial Bonaventuriana. Editores: César Alberto Collazos, José Luis Jurado Muñoz y Luis Merchán Paredes.

### 1.9.4 Participación en proyectos de investigación

- 2015 – 2016. Proyecto de COLCIENCIAS: "RAIM: Implementación de un framework apoyado en tecnologías móviles y de realidad aumentada para entornos educativos ubicuos, adaptativos, accesibles e interactivos para todos" con código 111956934172 y desarrollado por la Universidad Nacional de Colombia.
- 2015. Proyecto de investigación – Joven Investigador: "RA-200-COLCIENCIAS-Modelo de Sistema Adaptativo u-CSCL para Aprendizaje Virtual"

## 1.10 Organización del documento

El resto de este documento está organizado de la siguiente manera, en el capítulo 2 se presenta el marco teórico que abarca los principales conceptos y temáticas asociadas a la problemática. En el capítulo 3 se muestra el estado del arte construido a partir de la revisión bibliográfica de trabajos previos relacionados. Ya en el capítulo 4 se describe el modelo

propuesto, mientras que en el capítulo 5 se presenta la implementación y evaluación del prototipo implementado. Finalmente, en el capítulo 6 se presentan las conclusiones y se plantean trabajos futuros.

# 2. Marco teórico

En este capítulo se abordarán los conceptos más relevantes relacionados con la investigación y que sirven para dar soporte al modelo de descubrimiento de patrones en datos educativos e interacciones entre estudiantes y plataformas virtuales de aprendizaje mediante el uso de analíticas de aprendizaje y técnicas de minería de datos, que es propuesto en esta tesis.

En la Figura 2-1 se presenta la relación entre los conceptos principales que van a ir siendo abordados y detallados en este capítulo, los cuales se enmarcan dentro de las analíticas de aprendizaje, las plataformas virtuales de aprendizaje, las técnicas de análisis de datos para el descubrimiento de patrones y la personalización del aprendizaje.

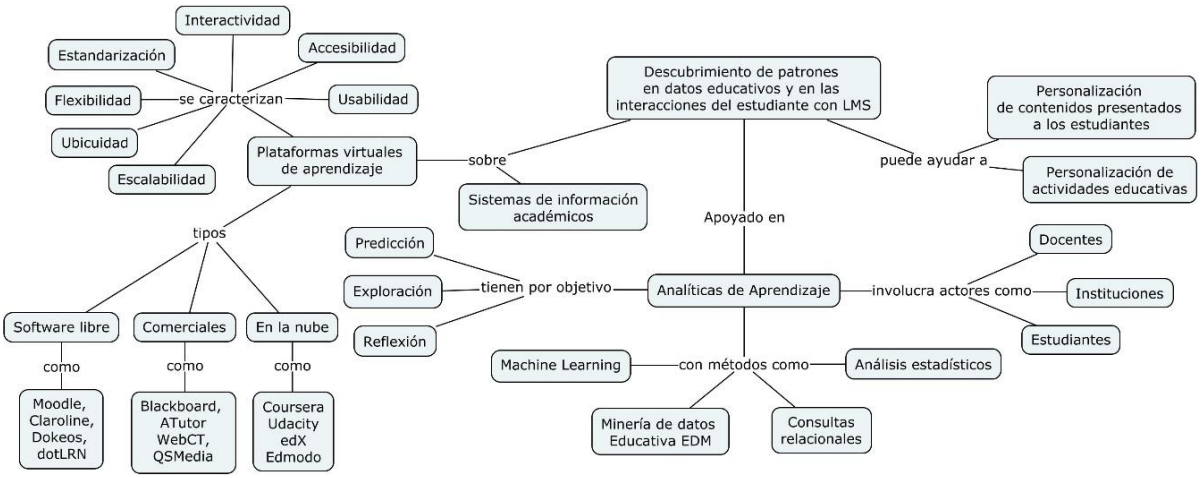



**Figura 2-1:** Conceptos principales

## 2.1 e-Learning y Plataformas Virtuales de Aprendizaje

El aprendizaje virtual (e-learning) puede tener diferentes definiciones, determinadas por la connotación que se le dé dentro de un ámbito de aplicación específico. Sin embargo, se

puede dar una definición del e-learning como la “capacitación no presencial que, a través de plataformas tecnológicas, posibilita y flexibiliza el acceso y el tiempo en el proceso de enseñanza-aprendizaje, adecuándolos a las habilidades, necesidades y disponibilidades de cada discente, además de garantizar ambientes de aprendizaje colaborativos mediante el uso de herramientas de comunicación síncrona y asíncrona, potenciando en suma el proceso de gestión basado en competencias” (García Peñalvo, 2005). El e-learning tiene dos componentes, uno pedagógico y otro tecnológico; el primero hace que estos sistemas no lleguen a convertirse simplemente en almacenamiento de información digitalizada, sino que mantengan un modelo pedagógico que garantice un proceso de enseñanza-aprendizaje acorde al contexto donde se está implementando. El segundo componente, hace referencia al soporte que da la tecnología, principalmente por medio de aplicaciones de software desarrolladas para ambientes web, a dichos procesos de enseñanza-aprendizaje, con lo que se garantiza que el proceso formativo sea más flexible y adaptativo, marcando con esto una diferencia clara frente a la formación tradicional presencial.

Un aspecto importante a considerar, cuando se habla de e-learning, es la plataforma usada para soportar el proceso de aprendizaje, ya que suele darse una dependencia de la metodología y de los contenidos con la herramienta o entorno usado para su implementación, que incluyen tanto a los estudiantes que hacen uso de ésta como al docente que configura un material y/o dirige un curso. Lo anterior hace necesario considerar el uso de estándares para facilitar la independencia entre los recursos, la metodología y la aplicación donde se ejecutan (Burgos, Tattersall, & Koper, 2005). Una plataforma virtual se entiende como un conjunto de herramientas y aplicaciones informáticas (síncronas o asíncronas) que están pensadas para apoyar y facilitar la administración y desarrollo de actividades formativas, generalmente cursos virtuales, a través de la web (Clarenc et. al. 2013). Existen diferentes denominaciones que se asociación al término de plataforma virtual y algunos de ellos pueden llegar a usarse indistintamente; estos son, Entorno de Aprendizaje Virtual (Virtual Learning Environment - VLE), Sistema de Gestión de Aprendizaje (Learning Management System - LMS), Sistema de Gestión de Cursos (Course Management System - CMS), Plataforma de Aprendizaje (Learning Plataform - LP).

### 2.1.1 Virtual Learning Environment – VLE

VLE corresponde a la sigla en inglés de Virtual Learning Environment. En español se traduce como un Entorno Virtual de Aprendizaje (EVA), y hace referencia a un sistema para la entrega de materiales de aprendizaje a los estudiantes a través de la web. Estos sistemas incluyen herramientas de evaluación, de seguimiento, de colaboración y de comunicación. Se puede acceder tanto dentro como fuera de la institución educativa, es decir, pueden apoyar el aprendizaje de los estudiantes fuera de las aulas de clase, incluso las 24 horas del día, siete días a la semana. Esto permite que las instituciones educativas puedan enseñar no sólo a estudiantes tradicionales, los cuales están disponibles a tiempo completo, sino también enseñar a aquellos que no pueden visitar regularmente las instalaciones físicas, debido a las restricciones geográficas o de tiempo (Oxford University Press, 2015).

Los VLE pueden presentar en su arquitectura cinco áreas (Rojas Castro, 2015):

- Información: organización de información como anuncios, noticias, diferentes tipos de datos, resultados, entre otros.
- Contenido: pueden disponer de diferentes tipos de contenidos como textos, formatos multimedia (audio, imágenes, animaciones, videos), presentaciones, etc.
- Comunicación: ya se comunicación síncrona o asíncrona, se dispone de diferentes herramientas para ello.
- Evaluación: se dispone de espacios para realizar pruebas evaluativas como test, encuestas, cuestionarios, tareas, entre otras.
- Gestión de estudiantes: se dispone de recursos para gestionar el ingreso de los estudiantes, el acceso a los contenidos, los tiempos de uso, entre otras.

### 2.1.2 Learning Management System – LMS

Los LMS corresponden a la sigla en inglés de learning management system, en español se pueden considerar como sistemas de gestión de aprendizaje. Estos surgen a partir de los CMS (content management systems) y brindan un ambiente que posibilita la actualización, mantenimiento y ampliación de la web con la colaboración de múltiples usuarios. Los LMS están dirigidos al aprendizaje y a la educación, por medio de

herramientas para la administración de los contenidos académicos, buscando así mejorar las competencias de las personas que toman los cursos y la intercomunicación entre estos. Todo lo anterior en un ambiente que permita hacer posible las adaptaciones de la formación a los requisitos de la organización o institución. Los LMS cuentan con herramientas que facilitan la distribución de cursos, recursos, noticias y contenidos con la formación general (Boneu, 2007). La Tabla 2-1 muestra un comparativo entre los LMS y los LCMS (integración de un CMS con un LMS), incluyendo dieciséis características fundamentales.

**Tabla 2-1:** Comparativo entre LMS y LCMS. Tomada de (Boneu, 2007)

| Usos | LMS | LCMS |
|---|---|---|
| **Usuarios a los que va dirigido** | Responsables de los cursos, administradores de formación, profesores o instituciones | Diseñadores de contenidos, diseñadores instruccionales, directores de proyectos |
| **Proporciona** | Cursos, eventos de capacitación y está dirigido a estudiantes | Contenidos para el aprendizaje, soporte en el cumplimiento y usuarios |
| **Manejo de clases, formación centrada en el profesor** | Si (pero no siempre) | No |
| **Administración** | Cursos, eventos de capacitación y estudiantes | Contenidos para el aprendizaje, soporte en el cumplimiento y usuarios |
| **Análisis de competencias – habilidades** | Si | Si (en algunos casos) |
| **Informe del rendimiento de los participantes en el seguimiento de la formación** | Enfoque principal | Enfoque secundario |
| **Colaboración entre usuarios** | Sí | Sí |
| **Mantiene una base de datos de los usuarios y sus perfiles** | No siempre | No siempre |
| **Agenda de eventos** | Sí | No |
| **Herramientas para la creación de contenidos** | No | Sí |
| **Organización de contenidos reutilizables** | No siempre | Sí |

| Usos | LMS | LCMS |
|---|---|---|
| **Herramientas para la evaluación integrada para hacer exámenes** | Sí (la mayoría de los LMS tienen esta capacidad) | Sí (la mayoría tienen esta capacidad) |
| **Herramientas de flujo de trabajo** | No | Sí (en algunas ocasiones) |
| **Comparte datos del estudiante con un sistema ERP (enterprise requeriment planning)** | Sí | No |
| **Evaluación dinámica y aprendizaje adaptativo** | No | Sí |
| **Distribución de contenido, control de navegación e interfaz del estudiante** | No | Sí |

La función principal de un LMS se concentra en el hecho de dar soporte a docentes y estudiantes en sus procesos de enseñanza-aprendizaje. Sin embargo, según Clarenc et. al., 2013, las funciones de un LMS pueden abarcar un campo más amplio, que incluye hacer que un ambiente de aprendizaje se convierta en una verdadera experiencia, por medio de la integración de materiales didácticos, herramientas de comunicación, colaboración y gestión educativa; algunas de las funciones se listan a continuación:

- Administrar los usuarios, los recursos, los contenidos y las actividades para la enseñanza de un tema en particular; calendarizar, organizar y ordenar eventos.
- Administrar el acceso; controlar y hacer seguimiento del proceso de aprendizaje; contar con herramientas para evaluar.
- Generar los informes de avances; gestionar servicios de comunicación (como foros de discusión y videoconferencias, entre otros);
- Permitir colaboración entre usuarios y posibilitar la conversación en línea.

### 2.1.3 Características de los LMS

Según Clarenc et. al. (2013) las principales características que deben cumplir las plataformas de e-Learning son las listadas a continuación.

- **Interactividad:** hace referencia a la comunicación bidireccional entre el receptor y el emisor. Esta característica se debe ver reflejada en los LMS en la medida que

los estudiantes tengan acceso a diversidad de información, material, recursos, entre otros; convirtiéndose en el protagonista de su proceso de aprendizaje.

- **Flexibilidad:** esta es una condición que se refiere al poder que tienen algunas cosas materiales o inmateriales a sufrir adaptaciones a los cambios. Se considera un LMS flexible cuando este no se mantiene rígido a los planes de estudio, sino que se adapta a la pedagogía y a los contenidos.
- **Escalabilidad:** es la propiedad de aumentar la capacidad de trabajo de un sistema sin perder funcionamiento o calidad. En los LMS la escalabilidad hace que la plataforma funcione con la misma calidad independiente de la cantidad de usuarios activos.
- **Estandarización:** cuando se emplea de un método aceptado, establecido y seguido para efectuar una actividad o función se dice que se sigue un estándar, lo cual implica también cumplir con ciertas reglas para obtener resultados esperados y aprobados para una actividad o función. En un LMS la estandarización es importante para efectos de utilizar cursos y/o materiales realizados independientemente.
- **Usabilidad:** hace referencia a la rapidez y facilidad para realizar las tareas por parte de las personas, logrando los objetivos con: efectividad, eficiencia y satisfacción. En un LMS, deben ser los usuarios los que determinan su usabilidad.
- **Funcionalidad:** en un LMS la funcionalidad se refiere a las características que permiten que una plataforma sea funcional a los requerimientos y necesidades de los usuarios, y está relacionada a su capacidad de escalabilidad.
- **Ubicuidad:** el termino ubicuo significa “en todas partes”. Para un LMS esta característica se refleja en la capacidad de la plataforma de hacer sentir al usuario omnipresente, es decir, que en ella encuentra lo necesario para su proceso de aprendizaje y que la puede acceder desde diferentes lugares, horas y dispositivos.
- **Accesibilidad:** se refiere a los medios que permiten que personas con capacidades distintas puedan acceder a la información online. Para cumplir con esta característica, las páginas web deben estar diseñadas de tal manera que puedan ser accedidas incluso por personas con deficiencia visual.

### 2.1.4 Tipos de LMS

Una manera de clasificar las plataformas virtuales de aprendizaje o LMS es según el tipo de licencia de uso que presenten. Según esto, Clrenc et. al. (2013) señalan que puede haber tres tipos: de uso comercial (o propietario), de software libre y en la nube.

- **Comercial:** Son de uso licenciado, es decir que para su uso hay que pagar a la empresa propietaria (quien lo desarrolló) o a la empresa que lo distribuye. Son sistemas generalmente robustos y bastante documentados. Cuentan con varias funcionalidades que pueden expandirse de acuerdo con las necesidades y presupuesto. Es decir que, cuanto más completo sea el paquete que se adquiera, más servicios se recibirán a cambio. Entre las más conocidas se encuentran Blackboard, WebCT, OSMedia, Saba, eCollege, Fronter, SidWeb, e-ducativa y Catedr@, entre otras.
- **Software libre:** Surgieron como una alternativa para disminuir los costos de los proyectos de formación en línea. Generalmente los LMS de este tipo han sido desarrollados por instituciones educativas o por personas que están vinculadas al sector educativo. Algunas de estas plataformas son de tipo “Open Source”, lo que las hace de libre acceso, permitiendo que el usuario sea autónomo para manipular el software, es decir, que una vez obtenido se pueda usar, estudiar, cambiar y redistribuir libremente. Algunas cuentas con una variada gama de funcionalidades, lo que hace que puedan equipararse o superar a las comerciales, mientras que otras sólo cuentan con funcionalidades básicas. Entre las más usadas están ATutor, Dokeos, Claroline, dotLRN, Moodle, Ganesha, ILIAS y Sakai.

- **En la nube:** no llegan a ser consideradas plataformas LMS propiamente dichas, ya que su mayor utilidad es la de permitir el apoyo a la clase presencial, así como el desarrollo de MOOC (Cursos online abiertos y masivos). Las más populares son Udacity, Coursera, Udemy, edX, Ecaths, Wiziq y Edmodo, entre otras.

### 2.1.5 Tipos de actividades educativas en algunas plataformas virtuales de aprendizaje

A continuación, se muestra en la Tabla 2-2 un comparativo entre la plataforma Blackboard y Moodle en cuanto a herramientas de soporte.

**Tabla 2-2:** Comparativo entre Blackboard y Moodle en cuanto a herramientas de soporte. Adaptado de (Kumar, Gankotiya, & Dutta, 2011)

| Característica | BlackBoard V.7 | Moodle |
|---|---|---|
| Autenticación | 1 | 1 |
| Autorización | 1 | 1 |
| Intercambio de archivos | 1 | 1 |
| Integración de registro | 1 | 1 |
| Tipos de prueba | 1 | 1 |
| Gestión automatizada | 1 | 1 |
| Soporte automatizado | 1 | 1 |
| Gestión de cursos | 1 | 1 |
| Calificación en linea | 1 | 1 |
| Seguimiento de estudiantes | 1 | 1 |
| Accesibilidad | 1 | 1 |
| Contenidos compartidos | 0 | 1 |
| Plantillas de curso | 1 | 1 |
| Aspecto y comportamiento | 1 | 1 |
| Diseño | 1 | 1 |
| Estándar instruccional | 1 | 1 |
| Total de características | 16 | 16 |
| Características disponibles | 15 | 16 |
| Características no-disponibles | 1 | 0 |

Las diferentes plataformas virtuales de aprendizaje ofrecen una variedad de actividades y herramientas para la organización de los cursos. A continuación, se presenta una comparativa entre unos de los LMS más populares en la Tabla 2-3.

**Tabla 2-3:** Comparación de algunos LMS en cuanto a herramientas de aprendizaje. Adaptada de (Kumar et al., 2011)

| Herramientas | Blackboard V.7 | Moodle | Claroline 1.6 | Dokeos 2.1.1 | eCollege |
|---|---|---|---|---|---|
| Foros de discusión | SI | SI | SI | SI | SI |

| | | | | | |
|---|---|---|---|---|---|
| Gestor de discusiones | SI | SI | SI | SI | SI |
| Intercambio de archivos | SI | SI | SI | SI | SI |
| Correo electrónico interno | SI | SI | SI | SI | SI |
| Revista en línea | SI | SI | NO | SI | NO |
| Chat en tiempo real | SI | SI | SI | SI | SI |
| Servicios de video | NO | SI | NO | NO | NO |
| Tablero | SI | SI | SI | SI | SI |
| Marcadores | NO | SI | NO | NO | NO |
| Calendario | SI | SI | NO | NO | SI |
| Orientación | NO | NO | SI | NO | SI |
| Búsqueda de cursos | SI | SI | SI | SI | SI |
| Trabajo off-line | SI | SI | NO | SI | SI |
| Trabajo grupal | SI | SI | SI | NO | SI |
| Comunidades | SI | SI | NO | SI | SI |
| Portafolios de estudiante | SI | SI | SI | NO | SI |
| Total de características | 16 | 16 | 16 | 16 | 16 |
| Características disponibles | 14 | 15 | 11 | 12 | 14 |
| Características no-disponibles | 2 | 1 | 5 | 4 | 2 |

Se puede apreciar en la tabla que Blackboard y Moodle son dos de las plataformas más completas en cuanto a esta característica de herramientas de aprendizaje. Ambas presentan una gran variedad de posibilidades para los docentes a la hora de definir las actividades de aprendizaje y configurar sus cursos. Sin embargo, en algunas ocasiones no todas estas son aprovechadas, con lo cual se puede despreciar la posibilidad que ofrecen estas diferentes actividades en las plataformas para atender las diversas formas de aprender de los estudiantes.

Por otro lado, Lerís & Sein-Echaluce (2011) afirman que “los sistemas de gestión del aprendizaje (LMS) se han desarrollado de tal manera que ofrecen entornos de aprendizaje con un amplio abanico de herramientas de aprendizaje, con facilidades para la gestión y

seguimiento de usuarios y para la administración de cursos", de modo que estos tienden a integrar la posibilidad de personalizar el aprendizaje (Lerís & Sein-Echaluce, 2011).

## 2.2 Sistemas de información académicos

Un sistema de información está formado por un grupo de elementos que se orientan a la administración de datos e información de diferentes actores y actividades. En el dominio académico, los sistemas de información procesan datos de estudiantes, docentes, datos propios de cada institución y de los procesos que se llevan a cabo en ella. En los sistemas de información académicos se pueden encontrar registros correspondientes a ingreso de estudiantes a programas curriculares, datos socioeconómicos y familiares, datos de calificaciones obtenidas, entre otros. Es de gran importancia que las instituciones cuenten con este tipo de sistemas para poder realizar análisis y estadísticos por periodos académicos, carreras o cualquier otro criterio que permita mostrar patrones para apoyo a la toma de decisiones o para la formulación de planes de mejoramiento (Kumar, Shobha, & Pal, 2015).

## 2.3 Analíticas de aprendizaje

Se puede tomar como referente la definición adoptada por la Society for Learning Analytics Research (SoLAR), quienes definen que, "las analíticas de aprendizaje son la medición, recopilación, análisis e interpretación de datos sobre los estudiantes y su contexto, con el objetivo de entender y optimizar el aprendizaje y el entorno en el que éste ocurre. Las analíticas de aprendizaje están en gran medida interesadas en mejorar el rendimiento del estudiante" (Siemens et al., 2011).

En ocasiones se suele tomar el término analíticas académicas como sinónimo de analíticas de aprendizaje y usarlos de manera indistinta. Sin embargo, SoLAR también adopta una definición particular para estas, afirmando que "las analíticas académicas hacen referencia al mejoramiento de los procesos organizacionales, los flujos de trabajo, la asignación de recursos y la medición institucional mediante el uso de datos de los estudiantes, datos académicos e institucionales. Las analíticas académicas, se asemejan a las analíticas de negocios, están interesadas en el mejoramiento de la efectividad organizacional" (Siemens

et al., 2011). En la Tabla 2-4 se presenta un cuadro comparativo entre las analíticas de aprendizaje y las analíticas académicas, junto con su nivel, objeto de análisis y beneficiarios.

**Tabla 2-4:** Analíticas de aprendizaje y académicas. Adaptado de (Siemens et al., 2011)

| Tipo de Analítica | Nivel y objeto de análisis | ¿A quién beneficia? |
|---|---|---|
| **Analíticas de aprendizaje** | **Nivel personal:** analíticas sobre el desempeño personal en relación con los objetivos de aprendizaje, recursos y hábitos de estudio de otros compañeros de clase.<br>**Nivel del Curso:** redes sociales, desarrollo conceptual, análisis del discurso, "currículo inteligente". | Estudiantes, educadores y personal docente |
| | **Departamental:** modelos predictivos, patrones de éxito/fracaso | Estudiantes, educadores |
| **Analíticas académicas** | **Institucional:** perfiles de estudiantes, rendimiento académico, flujo de conocimiento, asignación de recursos | Administradores, proveedores de fondos, mercadeo |
| | **Regional (estado/provincia):** comparaciones entre sistemas, calidad y estándares. | Proveedores de fondos, administradores |
| | **Nacional e internacional** | Gobierno nacional, autoridades educativas, UNESCO, OECD |

También existen conceptos como la Minería de Datos Educativos (EDM por sus siglas en inglés) y las Analíticas Web (WA por sus siglas en inglés), los cuales son enfoques cercanos a las analíticas de aprendizaje, pero que con frecuencia se usan como sinónimos. Sin embargo, todos estos enfoques han surgido como respuesta al fenómeno de la masificación de datos educativos, o conformación de Big Data. En el caso de las WA es el paradigma que puede ser considerado como el más genérico, ya que comprende un campo de aplicabilidad mayor, que va desde el análisis de sitios web sencillos hasta el análisis de entornos virtuales complejos y se datan sus orígenes incluso desde el surgimiento de la WWW (World Wide Web) (Rojas Castro, 2015).

Al aplicar las analíticas de aprendizaje en los procesos educativos se pueden obtener beneficios variados e incluyen, (1) la personalización y adaptación de los procesos de aprendizaje y contenidos, lo cual asegura que cada estudiante recibe los recursos y la enseñanza de acuerdo a lo que refleja su estado de conocimiento actual; (2) la ampliación y mejoramiento de los logros de aprendizaje, la motivación y la confianza por medio de la presentación oportuna al estudiante de información sobre su desempeño y el de sus compañeros, así mismo la generación de sugerencias sobre las actividades y contenidos que deben abordar para cubrir las falencias de conocimiento identificadas; (3) la optimización del tiempo del profesor y su esfuerzo al saber que estudiantes requieren información y ayuda adicional; (4) el mejoramiento en la calidad del diseño de los procesos de aprendizaje y en el desarrollo curricular a través de la utilización de los datos generados durante las actividades de instrucción y aprendizaje en tiempo real; (5) las visualizaciones interactivas de información compleja en función de las necesidades identificadas en un contexto de aprendizaje y (6) la rapidez en el logro de las metas de aprendizaje, al brindar a los estudiantes el acceso a herramientas que les ayuden a evaluar su progreso y a determinar qué actividades les llevan a conseguir mejores resultados (Siemens et al., 2011).

Long & Siemens (2011) plantean que existe una similitud entre las analíticas de aprendizaje y la inteligencia de negocio y plantean la necesidad de contar con un modelo o una serie de etapas para el desarrollo de las analíticas de aprendizaje, proponiendo el siguiente ciclo para ello (Long & Siemens, 2011):

**1**. **A nivel de curso:** rutas de aprendizaje, análisis de redes sociales, análisis del discurso.

**2**. **Minería de datos educativos:** modelado predictivo, clustering, minería de patrones

**3. Currículo inteligente:** desarrollo de recursos curriculares semánticamente definidos.

**4. Contenido adaptativo:** adaptación secuencial de contenidos basada en el comportamiento del estudiante, sistemas de recomendación.

**5. Aprendizaje adaptativo:** procesos de aprendizaje adaptativos (interacciones sociales, actividades de aprendizaje, apoyo al estudiante, no sólo contenido).

### 2.3.1 Objetivos de las analíticas de aprendizaje

Las analíticas de aprendizaje han dado lugar a espacios de investigación variados y por ende atienden objetivos diversos. Además han ido evolucionando desde su aparición, en el Horizon Report del año 2011 (Johnson, Smith, Willis, Levine, & Haywood, 2011), donde se presentaban las analíticas de aprendizaje como una tecnología por revisar y con un horizonte a largo plazo (cuatro a cinco años para su adopción) y mostraban como su objetivo principal “movilizar el poder de las herramientas de minería de datos al servicio del aprendizaje y abarcar la complejidad, diversidad y abundancia de información que se puede generar en los entornos de aprendizaje”.

Por su parte, en el Horizon report del año 2016 las analíticas de aprendizaje aparecen como un desarrollo importante en tecnología educativa para la educación superior con un horizonte de corto plazo (un año o menos). Y muestran como su objetivo “construir mejores pedagogías, potenciar el aprendizaje activo, orientar a los estudiantes en riesgo y evaluar los factores que afectan la terminación y el éxito del estudiante”.

## 2.4 Minería de Datos

En los últimos años con el aumento acelerado de la capacidad de almacenamiento y de la potencia computacional de los diferentes sistemas informáticos se ha dado base a la formulación de un conjunto de técnicas que combinan la inteligencia artificial y la estadística y se encierran bajo el apelativo de minería de datos (data mining). Como su nombre lo indica, el fin que tienen estas técnicas es la extracción de información relevante a partir de grandes volúmenes de datos, por medio de la utilización de algoritmos o técnicas que tratan de localizar información no trivial (diferencias, patrones, relaciones significativas, efectos de interacción, entre otros) (Castro & Lizasoain, 2012).

La minería de datos suele estar enmarcada en el proceso de descubrir conocimiento a partir de una base de datos, KDD (siglas en inglés de Knowledge Discovery in Databases), el cual describe una secuencia de etapas bien definidas, siendo cada una primordial para la transformación de los datos en conocimiento. El proceso incluye la presentación de los datos, el análisis estadístico, los algoritmos de minería de datos y la evaluación e interpretación de resultados. Los sistemas de KDD se caracterizan fundamentalmente incluir la aplicación de técnicas de minería de datos, y la diferencia entre unos y otros está

dada es por la implementación y presentación. El resultado final de las etapas de un sistema de KDD suele ser un modelo descriptivo, que luego puede ser convertido en un modelo predictivo si se hace necesario (Quiroz Gil & Valencia, 2012). La Figura 2-2 representa las etapas que se siguen en el proceso de KDD

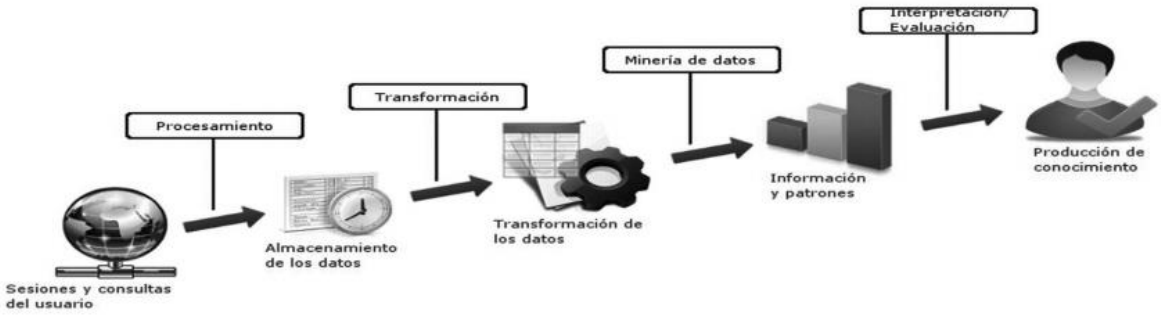



**Figura 2-2:** Etapas del proceso de descubrimiento del conocimiento en bases de datos. (KDD). Tomado de (Quiroz Gil & Valencia, 2012)

Desde la minería de datos se pueden afrontar diversos problemas, siempre que existan datos para afrontarlos. Banet (2001) presenta la siguiente lista de problemas a los que puede hacer frente la minería de datos:

- **Búsqueda de lo inesperado por descripción de la realidad multivariante**. Para describir un fenómeno cuantas más variables se tengan mejor, esto hará más globales y coherentes las descripciones y facilitará la detección de lo inesperado, es decir, aquello que no está previsto y que resulta valioso para entender mejor el comportamiento de algún grupo de individuos.
- **Búsqueda de asociaciones.** Si se tiene un cierto acontecimiento, ¿se puede saber si este está asociado a otro suceso?, ¿se pude inferir que ciertos sucesos ocurren simultáneamente con mayor frecuencia de lo que sería esperado si fuesen independientes?
- **Definición de tipologías.** Detectar tipos de comportamientos, gustos, preferencias, tendencias, entre otros. Por ejemplo, perfiles de compra en los consumidores, tendencia de opinión, condiciones de vida, entre otros.
- **Detección de ciclos temporales.** Detectar los diferentes ciclos y/o fases donde se sitúa un individuo, proceso, evento o situación; con el fin de ayudar a crear modelos y a adecuar mejoras.

- **Predicción.** A menudo se precisa la necesidad de generar pronósticos, que atiendan a preguntas como: ¿cuál es la probabilidad de perder un cliente?, ¿cuál es el precio de una vivienda concreta?, ¿lloverá mañana? Si la variable de respuesta a estas preguntas es continua se tratará de un problema de regresión, mientras que si la variable de respuesta es categórica se tratará de un problema de clasificación.

Se propone en la literatura una clasificación de los algoritmos de minería de datos en dos grandes grupos, los supervisados o predictivos y no supervisados o de descubrimiento del conocimiento. A su vez, las técnicas de minería de datos pueden ser clasificadas bajo los dos grupos anteriores (Moreno García, Quintales, García-Peñalvo, & Polo Martín, 2001), como se aprecia en la Tabla 2-5.

**Tabla 2-5:** Clasificación de las técnicas de minería de datos. Tomada de (Moreno García, et.al., 2001).

| Supervisados | No supervisados |
|---|---|
| Árboles de decisión | Detección de desviaciones |
| Inducción neuronal | Segmentación |
| Regresión | Agrupamiento (“clustering”) |
| Series temporales | Reglas de asociación |
| | Patrones secuenciales |

Los algoritmos que se clasifican como supervisados buscan predecir el valor de un atributo a partir de un conjunto de datos, conociendo otros atributos que son denominados atributos descriptivos. La predicción se hace entonces mediante la inducción de la relación entre el atributo y los atributos descriptivos; esto se suele conocer bajo la denominación de aprendizaje supervisado y consta de dos fases, una de entrenamiento, en la cual se construye un modelo usando un subconjunto de datos y otra fase de prueba, en la cual se utiliza el resto de datos para verificar el modelo. Por su parte, los algoritmos o métodos no supervisados (también llamados de descubrimiento de conocimiento) encuentran patrones y tendencias en datos actuales, no requieren de un conocimiento histórico a priori (Moreno García et al., 2001).

Otra forma de clasificar los algoritmos de minería de datos es según el tipo de tarea que desempeñan. En la Tabla 2-6 se presenta una descripción de algunos de ellos.

**Tabla 2-6:** Tipos de algoritmos de MD según su tarea. Construido a partir de (Microsoft, 2016)

| Tipo de algoritmo | Descripción |
|---|---|
| Algoritmos de clasificación | Se basan en atributos presentes en un conjunto de datos para predecir una o más variables discretas |
| Algoritmos de regresión | Se basan en atributos presentes en un conjunto de datos para predecir una o más variables numéricas continuas |
| Algoritmos de segmentación | Toman el conjunto de datos y lo dividen en grupos de datos que tienen elementos de similares características |
| Algoritmos de asociación | Toman el conjunto de datos y buscan la existencia de correlaciones entre los diferentes atributos |
| Algoritmos de análisis de secuencias | Tratan la presencia de secuencias o eventos frecuentes en el conjunto de datos |

### 2.4.1 Algunas técnicas de Minería de Datos

A continuación, se describirán algunas de las principales técnicas de minería de datos, tanto supervisadas como no supervisadas.

**Arboles de decisión**

Esta técnica hace parte del aprendizaje supervisado, consiste en una representación de una función multivariada. Un árbol de decisión puede ser interpretado como una serie de reglas que se han compactado bajo una representación en forma de árbol. Al contar con un conjunto de ejemplos, los cuales han sido estructurados como vectores de pares ordenados atributo-valor, el concepto que estos sistemas adquieren durante el proceso de aprendizaje consiste en un árbol. Cada eje está etiquetado con un par atributo-valor y las hojas con una clase, de forma que la trayectoria que determinan desde la raíz los pares de un ejemplo de entrenamiento alcanza una hoja etiquetada -normalmente- con la clase del ejemplo. La clasificación de un ejemplo nuevo del que se desconoce su clase se hace con la misma técnica, solamente que en ese caso al atributo clase, cuyo valor se desconoce, se le asigna de acuerdo con la etiqueta de la hoja a la que se accede con ese ejemplo (Molina & García, 2006).

**Clustering**

El clustering o también conocido como agrupamiento, permite la identificación de tipologías o grupos donde los elementos guardan gran similitud entre sí y muchas diferencias con los de otros grupos. Esta técnica puede ser usada para segmentar un grupo de clientes, un conjunto de valores e índices financieros, una serie de observaciones astronómicas, un conjunto de zonas forestales, un grupo de empleado, entre otros. El clustering ha empezado a despertar gran interés en las comunidades académicas y científicas, dadas las ventajas que aporta al permitir el tratamiento de grandes colectivos o conjuntos de datos que requieren ser agrupados o segmentados (Molina & García, 2006).

**Reglas de asociación**

Este tipo de técnica no supervisada tiene como objetivo establecer las posibles relaciones o correlaciones entre distintos sucesos supuestamente independientes; contando con la capacidad de reconocer como la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros. Son utilizadas cuando el objetivo es realizar análisis exploratorios, buscando relaciones dentro del conjunto de datos. Las asociaciones identificadas pueden usarse para predecir comportamientos, y permiten descubrir correlaciones y co-ocurrencias de eventos. Para hacer uso de estas es necesario disponer de información de los sucesos ocurridos en un determinado período temporal. Generalmente esta forma de extracción de conocimiento se basa en técnicas estadísticas, como los análisis de correlación y de variación y uno de los algoritmos más utilizado es el algoritmo A priori (Molina & García, 2006).

**Patrones secuenciales**

Esta técnica consiste en realizar un análisis que permite encontrar patrones similares en los datos de transacciones ejecutadas durante un período dado. Los analistas pueden usar estos patrones para identificar relaciones entre los datos. Los modelos matemáticos son patrones secuenciales detrás de la lógica normativa, la lógica difusa u otras. En las etapas de la minería de datos, se pueden estudiar varias secuencias similares para identificar tendencias futuras en el desarrollo de dichas transacciones. Este enfoque es útil en el tratamiento de bases de datos con características de series de tiempo (Dueñas Reyes, 2010).

### 2.4.2 Minería de Datos Educativos (EDM)

Como una rama específica de la minería de datos surge la minería de datos educativos o EDM (por sus siglas en inglés Educational Data Mining). Se puede definir la EDM como la investigación aplicada a métodos para detectar algunos patrones, ya sea de comportamiento o relacionales, en conjuntos de datos de gran tamaño o Big Data educativos (Romero & Ventura, 2013). La EDM y las analíticas de aprendizaje comparten unidades de análisis y herramientas, son dos enfoques cercanos, algunos autores caracterizan a la EDM con una concentración hacia los datos y sus relaciones, pero alejada de aspectos de tipo cualitativo como por ejemplo el contexto (Rojas Castro, 2015). En la Figura 2-3 se muestran las áreas relacionadas a la minería de datos educativos y como comparten en algunos espacios.

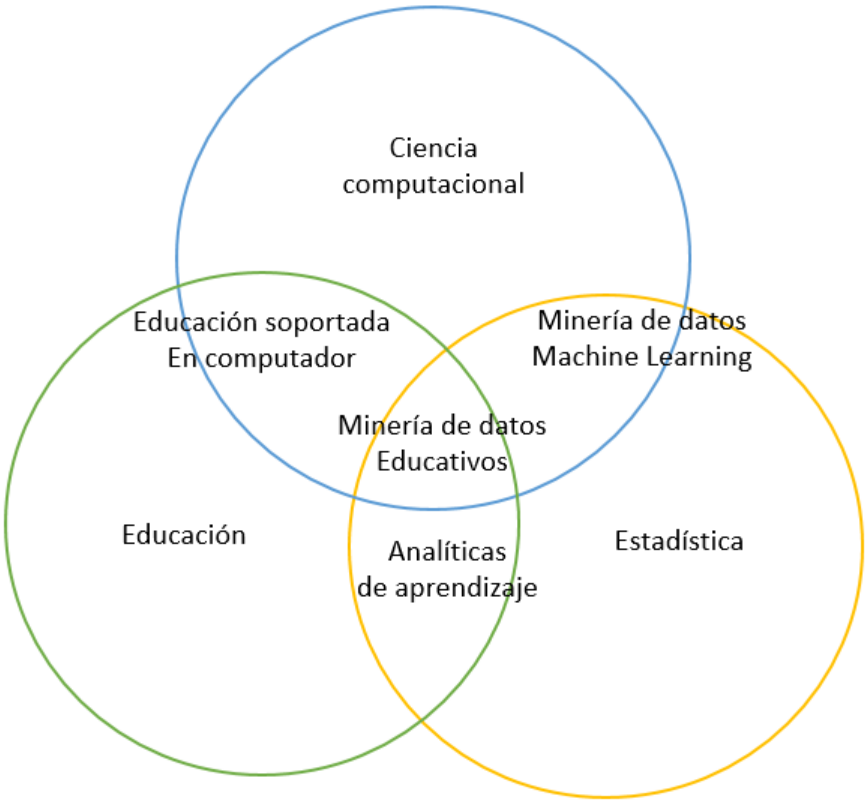



**Figura 2-3:** Minería de datos educativos y áreas relacionadas. Adaptada de (Romero & Ventura, 2013)

En el caso particular de las analíticas de aprendizaje y la Minería de Datos Educativos, se puede revisar específicamente sus puntos en común y diferencias a través de la comparativa en la Tabla 2-7.

**Tabla 2-7:** Analíticas de aprendizaje Vs. EDM. Tomada de (Rojas Castro, 2015)

| Aspecto | Analíticas de Aprendizaje | Minería de Datos Educativos |
|---|---|---|
| Técnicas | Técnicas estadísticas, Visualización de datos, análisis de redes sociales, análisis de opiniones, del discurso, conceptual, construcción de modelos de sentido | Técnicas de clasificación, clustering (identificación, tipificación de grupos), modelos bayesianos. |
| Orígenes | Se basa en el concepto de la web semántica, análisis curricular (curriculum inteligente), intervención sistémica. | Software educacional, modelamiento computacional, predicción de resultados. |
| Énfasis | - Descripción de datos y resultados.<br><br>- Busca la comprensión en la totalidad del sistema<br><br>- Hay lugar para respuestas personalizadas | - Descripción y comparación utilizando técnicas de data mining.<br>- Reducción en componentes para la comprensión<br>- Las respuestas buscan ser automatizadas y generales |
| Tipos de descubrimientos | El juicio y criterio humano es la llave principal, los procesos automatizados son herramientas | Procesos automatizados de descubrimiento son la llave principal, se complementa con el juicio humano |
| Métodos computacionales | Utiliza los métodos informático computacionales que existen | Desarrolla nuevos métodos informático computacionales de acuerdo a las necesidades que se presenten |
| Objetivos | Analizar datos y procesos que ocurren en los VLE. Reflexionar sobre las interacciones y aprendizaje | Aplicar técnicas de análisis para mejorar proceso de enseñanza aprendizaje |
| Nivel de análisis | Individual / Grupal | Grupal |

## 2.5 Conclusiones del capítulo

En este capítulo se introdujo la definición de los principales conceptos que serán utilizados a lo largo del desarrollo de este documento, ya que están conceptos están directamente asociados a la investigación realizada y soportan el modelo propuesto para el descubrimiento de patrones en datos académicos y en las interacciones entre estudiantes y plataformas virtuales de educación mediante el uso de analíticas de aprendizaje. En el capítulo siguiente se presentarán algunos trabajos investigativos relacionados con el área temática de esta tesis.

# 3. Estado del arte

En este capítulo se muestran algunos trabajos de investigación revisados y que son afines a esta tesis; para ello se resaltan las fortalezas, limitaciones y oportunidades de cada uno. La revisión de los trabajos se presenta en dos grupos, un primer grupo incluye los trabajos de revisión y contextualización acerca de la temática y un segundo grupo recoge los trabajos en los cuales existe una propuesta de modelo en el tópico de interés. Para la búsqueda de los artículos revisados en el estado del arte, se utilizó una herramienta suministrada por la dirección nacional de bibliotecas de la Universidad Nacional de Colombia llamada “Descubridor”, el cual está disponible para estudiantes y docentes de la universidad para realizar búsqueda de material bibliográfico en reconocidas bases de datos académicas. Además, se utilizaron bases de datos como Scopus, Web of Science, IEEE Xplore y ScienceDirect.

## 3.1 Primer grupo: artículos de revisión – contextualización

El primer grupo recoge los trabajos que hacen una revisión o contextualización del estado actual de las investigaciones en temas como analíticas de aprendizaje, minería de datos educativos, grandes volúmenes de datos educativos e interacciones en plataformas de educación virtual.

Ecuación de búsqueda: “Learning” + “analytics”
Restricción de años: 2012 – 2016
Tipo de documento: Documento científico

**Título:** Learning analytics: drivers, developments and challenges
**Autor:** Ferguson
**Fuente:** International Journal of Technology Enhanced Learning

**Año:** 2012

En este artículo la autora presenta la minería de datos educativos – EMD y las analíticas académicas como precursores de las analíticas de aprendizaje; para ello realiza la definición de cada, vislumbrando su verdadero potencial de desarrollo y posteriormente presentando las áreas de desarrollo emergentes y algunos retos futuros que se deben afrontar. Las analíticas de aprendizaje se distinguen por su preocupación por proporcionar valor a los estudiantes, ya sea en contextos formales, informales o combinados, siendo empleadas para entender y optimizar el aprendizaje y los entornos en los que se lleva a cabo el proceso educativo. A pesar de ser un área de investigación que se considera reciente y que incluye una amplia labor en áreas relacionadas, ya se han desarrollado una serie de herramientas y métodos que ofrecen gran potencial (Rebecca Ferguson, 2012).

**Fortalezas:**

- Hace una caracterización de los factores que dan impulso al tema de LA (ver Figura 3-1), lo que permite tener un panorama más claro sobre su alcance a la hora de la implementación en un proyecto de investigación.
- Se enuncian los retos futuros a los que se van a enfrentar las analíticas de aprendizaje, dando así una perspectiva sobre los temas que más relevancia van a tener a la hora de realizar una investigación científica. Los retos enunciados son:
    - Construir conexiones con las ciencias del aprendizaje.
    - Construir métodos que permitan manejar grandes cantidades de datos, de manera que se optimicen los ambientes de aprendizaje.
    - Centrarse en las perspectivas de los estudiantes.
    - Desarrollar y aplicar un conjunto de guías éticas claras.

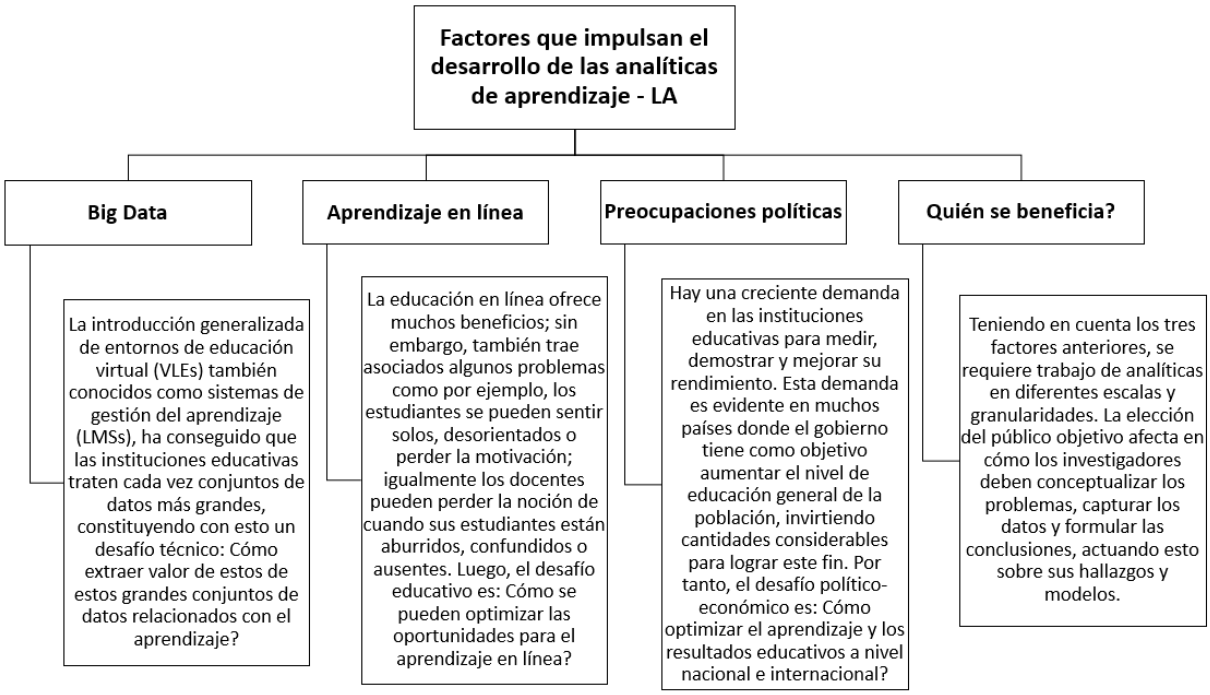



**Figura 3-1:** Factores que impulsan el desarrollo de LA. Adaptado de (Ferguson, 2012)

**Oportunidades:**

- Teniendo en cuenta los retos futuros enunciados en las fortalezas, se puede encontrar una posible contribución en la formulación de un modelo basado en analíticas de aprendizaje, que permita utilizar la gran cantidad de datos que son recolectados por medio de las interacciones propias de los procesos de enseñanza-aprendizaje en entornos virtuales.

**Título:** The Evolution of Big Data and Learning Analytics in American Higher Education
**Autor:** Picciano
**Fuente:** Journal of Asynchronous Learning Networks
**Año:** 2012

En este artículo, se examina la evolución de Big Data y de las analíticas de aprendizaje en la educación superior americana, viendo estos dos conceptos como parte de una solución integral a algunos problemas y decisiones que se deben afrontar desde el nivel administrativo de las instituciones educativas. Se plantea como objetivo del trabajo la revisión de la naturaleza de estos conceptos, considerando posibles aplicaciones e identificando las preocupaciones que surgen en su implementación y crecimiento en los

entornos educativos. Picciano afirma que la educación superior en América está pasando por un periodo de crisis, el cual se puede ver como una oportunidad para examinar y mejorar los procesos. Actualmente, la tecnología se ha convertido en el centro de muchas actividades; con el internet y su ubicuidad, se ha permeado cada aspecto de la sociedad, incluyendo la educación, evidenciado esto en el creciente número de estudiantes que optan por un aprendizaje en línea. Se resalta la posibilidad de integrar las transacciones instruccionales generadas en CMS o LMS con otros recursos como los datos de los sistemas de información universitaria (estudiantes, cursos, facultades) así como las experiencias de profesores y asesores, la intuición y conocimientos; los cuales, al ser revisados por medio de un software de análisis, por ejemplo, con árboles de decisión, pueden ser usados para elaborar directrices y acciones de mejora para cursos siguientes. La Figura 3-2 representa el flujo del modelo planteado. (Picciano, 2012).

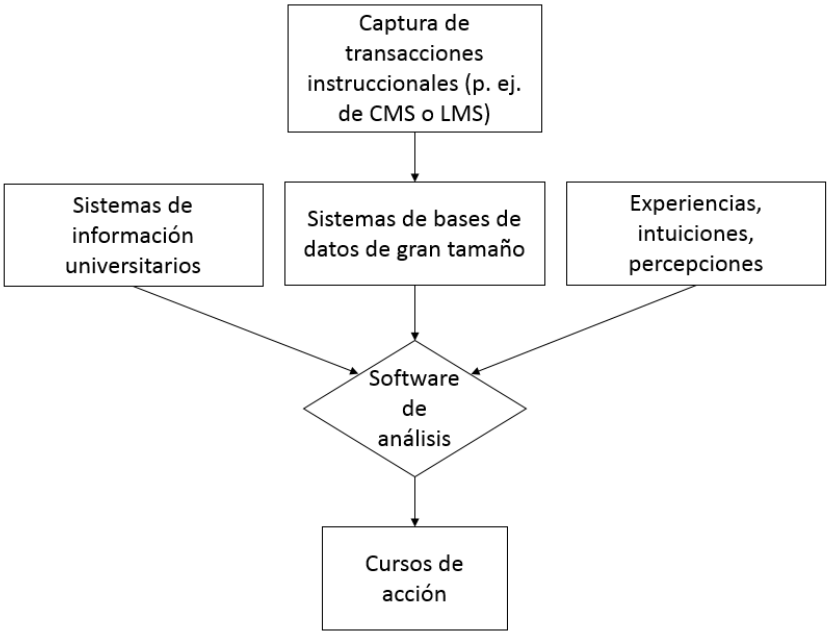



**Figura 3-2:** Flujo del modelo de analíticas de aprendizaje. Adaptado de (Picciano, 2012)

**Fortalezas:**

- Presentan la importancia de los sistemas gestores de contenido (CMS por sus siglas en inglés) y sistemas gestores de aprendizaje (LMS por sus siglas en inglés), como fuente principal de recolección de datos sobre las interacciones que generan los estudiantes y profesores en los procesos de enseñanza-aprendizaje.

- También se menciona la importancia de recolectar los datos de los procesos educativos sobre cursos en línea, ya que los cursos presenciales necesitarían un gran esfuerzo en cuanto a conversión de datos se refiere, para poder registrar qué eventos se están presentando en el proceso de enseñanza-aprendizaje.
- Además, se menciona un punto a tener en cuenta, la privacidad de los individuos que hacen parte de los procesos educativos, ya que, si bien es necesario recolectar gran cantidad de datos sobre estos procesos, se debe tener en cuenta hasta qué punto están estos dispuestos a ser monitoreados.

**Oportunidades:**

- Si bien es necesario generar un modelo de analíticas de aprendizaje para el análisis de los datos recolectados en las interacciones de los individuos en los procesos educativos, también es necesario tener en cuenta qué tan cómodos se sienten estos al ser monitoreados constantemente.

**Título de artículo:** Analytics for Education
**Autores:** MacNeill, Campbell, & Hawksey
**Fuente:** Journal of Interactive Media in Education
**Año:** 2014

En este artículo se presenta una visión sobre el estado del desarrollo y uso de analíticas en contextos educativos. Mediante el uso de los niveles de analíticas provisto por Buckingham Shum, los autores presentan un análisis crítico de los desarrollos que actualmente se realizan en el dominio de las analíticas de aprendizaje, además de contrastar el valor potencial de la investigación y desarrollo de las analíticas con la implementación y práctica en el mundo educativo real. Por otra parte, este artículo se centra en el desarrollo de contenido analítico, considerando las implicaciones legales y éticas de la recolección y tratamiento de datos educativos (MacNeill, Campbell, & Hawksey, 2014).

**Fortalezas:**

- Describen los niveles del ámbito educativo (ver Figura 3-3) en los cuales se pueden aplicar las analíticas de aprendizaje (análisis de nivel macro, análisis de nivel meso y análisis de nivel micro), con el fin mejorar los procesos de enseñanza-aprendizaje.

- Presentan la diferencia entre minería de datos para la educación y analíticas de aprendizaje, lo que ayuda a enfocar realmente el alcance y enfoque que se le quiere dar a la investigación.
- Mencionan las diferentes motivaciones que pueden tener los interesados en la aplicación de analíticas de aprendizaje, lo que ha creado discrepancias en su uso en entornos reales.
- Mencionan los inconvenientes éticos que se pueden llegar a tener con la utilización de los datos recolectados en las plataformas de educación virtual, provenientes de los diferentes actores involucrados en el proceso de enseñanza-aprendizaje.

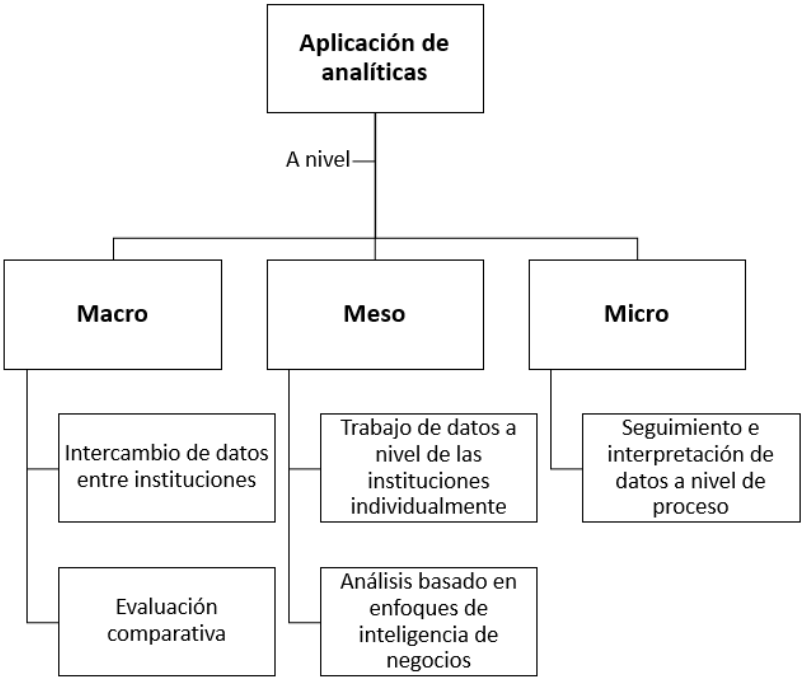



**Figura 3-3:** Niveles de aplicación de las analíticas de aprendizaje. Adaptado de (MacNeill, Campbell, & Hawksey, 2014).

**Limitaciones:**

- Hacen una descripción de los diferentes temas relacionados con las analíticas de aprendizaje; sin embargo, no propone un posible modelo que, utilizando estos temas, ayude a buscar una solución sobre la utilización efectiva de los datos recolectados en las interacciones del proceso de enseñanza-aprendizaje.

**Oportunidades:**

- Se puede proponer un modelo de analíticas de aprendizaje, el cual permita, mediante la utilización de los conceptos y tecnologías mencionadas en este artículo, aprovechar de manera positiva, los datos generados en los procesos de enseñanza-aprendizaje.

**Título:** Educational data mining: A survey and a data mining-based analysis of recent works
**Autor:** Peña-Ayala
**Fuente:** Expert Systems with Applications
**Año:** 2014

Peña-Ayala hace una revisión de los trabajos realizados en el campo de la minería de datos educativos (EMD por sus siglas en inglés), teniendo en cuenta su perfil y enfoque. En los resultados se incluyen la selección y análisis de 240 trabajos de EMD, entre los cuales se encuentran enfoques y herramientas. Se muestra que la EMD es un campo de investigación aún incipiente y da pie a que en esta primera década del siglo se recorran los pasos iniciales y sea el surgimiento de investigaciones más a fondo en el campo. La EMD surge como un paradigma orientado al diseño de modelos, tareas, métodos y algoritmos para explorar datos provenientes de entornos educativos; esto con el propósito de encontrar patrones y hacer predicciones que caractericen el comportamiento y logros de los estudiantes, los contenidos en dominios de conocimiento, las evaluaciones y las funcionalidades y aplicaciones educativas. El autor propone algunas temáticas (ver Figura 3-4) para el desarrollo de futuras investigaciones de manera que se hagan aportes reales a la solución de los retos encontrados (Peña-Ayala, 2014).

**Fortalezas:**

- Facilita una descripción de la línea de base en Minería de Datos que soporta el enfoque de la Minería de Datos Educativos - EDM.
- Se define la naturaleza y línea de base del enfoque de EDM.
- Se dan patrones para diseñar enfoques de EMD, que son útiles como referencia para desarrollar versiones de modelos descriptivos y predictivos.
- Los hallazgos presentados permiten enfocar la investigación de EMD hacia los temas que realmente pueden generar un impacto positivo en la comunidad académica.

**Oportunidades:**

- Teniendo en cuenta los aportes realizados por el autor, se tiene la oportunidad de utilizar los enfoques de EMD para obtener un mayor desarrollo en futuras investigaciones, de tal manera que se hagan aportes significativos en el desarrollo de esta línea de investigación.

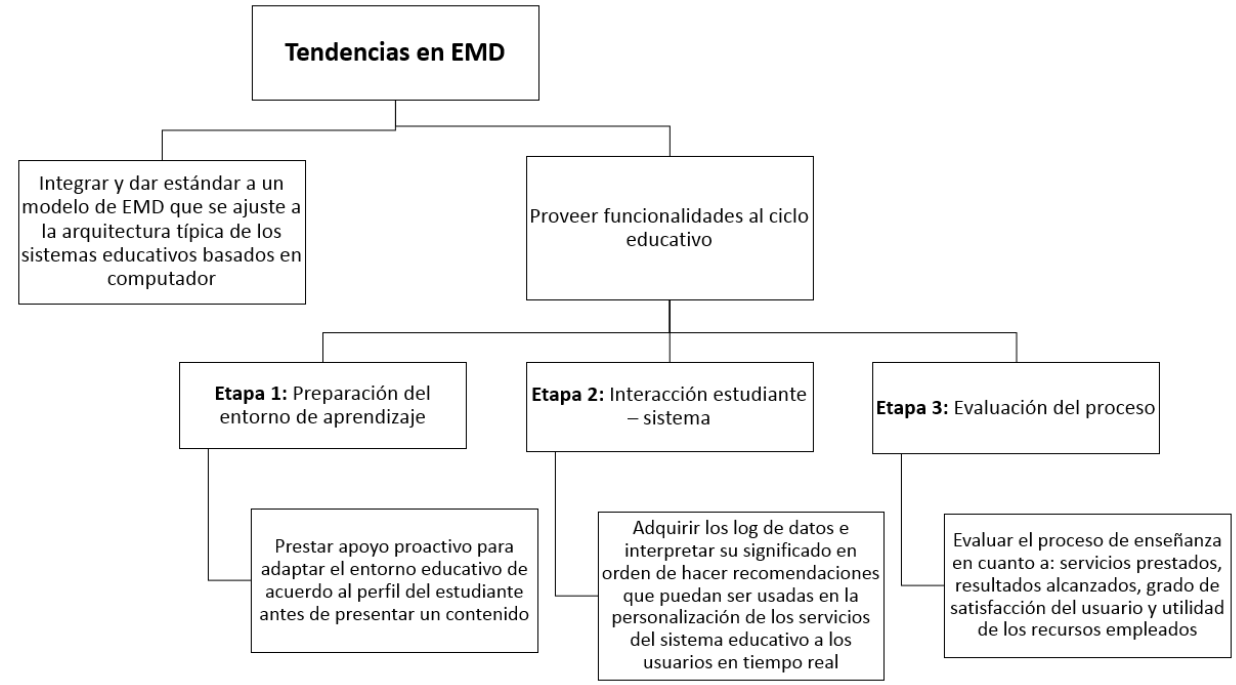



**Figura 3-4:** Tendencias en Minería de Datos Educativos. Adaptado de (Peña-Ayala, 2014).

**Título:** Horizon Report 2016

**Autores:** Johnson, Adams Becker, Cummins, Estrada, Freeman, & Hall, C.

**Fuente:** The New Media Consortium

**Año:** 2016

Este informe es generado anualmente y hace un reporte completo de las tendencias, retos y tecnologías en desarrollo en temas de educación, agrupando en horizontes de tiempo, corto, mediano y largo plazo. En su última edición, correspondiente al año 2016, el Horizon Report muestra a las analíticas de aprendizaje como una tecnología en adopción en un año o menos y las asocia directamente con el aprendizaje adaptativo y con el Machine Learning como técnica que puede ayudar a hacer adaptaciones para cada estudiante en tiempo real. Se muestra que las instituciones educativas, tanto colegios como

universidades ya cuentan con acceso a herramientas y grandes conjuntos de datos, requiriendo ahora, iniciar procesos de personalización de las experiencias de aprendizaje por medio del análisis de estos datos, que permita descubrir las particularidades de cada estudiante. Se muestran las analíticas de aprendizaje como una tecnología que se ha desarrollado en tres etapas, la primera caracterizada por la descripción de resultados, una segunda donde sobresale el diagnóstico y la tercera se concentra en la predicción de lo que sucederá. Todas las etapas son importantes dentro de los modelos que aplican analíticas de aprendizaje. Se indica además que “a pesar de que existe un creciente interés mundial en utilizar el aprendizaje analítico y las tecnologías de aprendizaje adaptativo para mejorar la enseñanza y el aprendizaje, no hay muchos estudios que detallan los resultados concretos” (Johnson et al., 2016).

**Fortalezas:**

- Deja ver como las analíticas de aprendizaje han ido evolucionando desde un horizonte de tiempo de alrededor de 5 años. Mostrando las etapas que han ido abarcando y los distintos enfoques que pueden tomar.
- Se destaca en este último reporte del consorcio, que las analíticas de aprendizaje han formado una fuerte relación con el aprendizaje adaptativo, lo que implica que se han venido convirtiendo en fuente para apoyar procesos como la personalización de actividades o espacios educativos.

**Oportunidades:**

- El reporte presenta una serie de enlaces que permiten acceder a proyectos donde se provee ejemplos del uso de analíticas y aprendizaje adaptativo en la educación superior. Se destacan iniciativas de comunidades europeas como la Universidad Internacional de La Rioja (UNIR) en España, con su programa iLime, un sistema de aprendizaje adaptativo que se apoya en las interacciones formales e informales de los estudiantes.

## 3.2 Segundo grupo: artículos con modelos propuestos

El segundo grupo recoge los trabajos en los cuales existe una propuesta de modelo en el tópico de interés.

Ecuación de búsqueda: “Learning” + “analytics”, “Learning” + “application”, “Learning” + “analytics” + “model”

Restricción de años: 2012 – 2016

Tipo de documento: Documento científico

**Título:** Dataset-Driven Research to Support Learning and Knowledge Analytics

**Autores:** Verbert, Manouselis, Drachsler, & Duval

**Fuente:** Journal of Educational Technology & Society

**Año:** 2012

Verbert et. al. (2012) presentan un análisis de diferentes conjuntos de datos (datasets), los cuales pueden ser usados para la investigación en el tema de Analíticas de Aprendizaje y de Conocimiento (LAK por sus siglas en inglés). Como primer punto, presentan un marco de trabajo (ver Figura 3-5) mediante el cual los investigadores pueden cuestionar la pertinencia de un conjunto de datos para el desarrollo de su investigación, posteriormente analizan conjuntos de datos existentes mediante la utilización del marco de trabajo construido, y finalmente describen los retos de recolectar y compartir conjuntos de datos educativos (Verbert, Manouselis, Drachsler, & Duval, 2012).

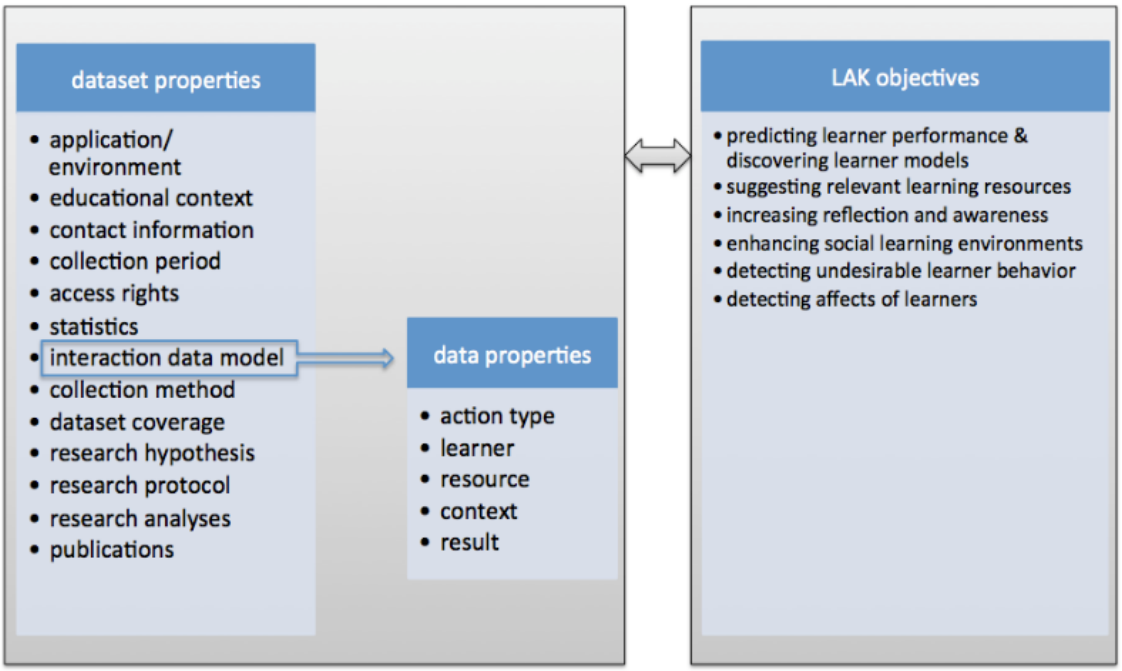


**Figura 3-5:** Framework para Datasets Educativos. Tomado de (Verbert et. al., 2012)

**Fortalezas:**

- Presentan un marco de trabajo, con el cual se pueden analizar conjuntos de datos educativos, de tal manera que los investigadores puedan cuestionar acerca de la utilidad potencial que tiene el conjunto de datos en la labor investigativa.
- Presentan algunos conjuntos de datos, los cuales pueden ser utilizados para la validación de las investigaciones realizadas en torno a analíticas de aprendizaje y conocimiento (LAK) y que están disponibles para su uso.
- Presentan algunos de los retos que tiene la obtención de conjuntos de datos educativos, entre los cuales podemos encontrar:
    - Los derechos de autor bajo los cuales pueden ser utilizados los conjuntos de datos.
    - Falta de heterogeneidad por la ausencia de estándares en la representación de los datos de las interacciones, lo que dificulta compartir y reusar datos entre sistemas.
    - La identificación de datos relevantes sobre los estudiantes y demás entidades involucradas en la investigación.
    - El desarrollo de sensores de datos que permitan recolectar datos.

**Oportunidades:**

- A partir de los conjuntos de datos y el marco de trabajo propuesto se pueden generar investigaciones, en las cuales se puedan solucionar algunos de los retos propuestos por el autor como los son:
    - El primer reto propuesto, tiene que ver con la falta de un formato predefinido, que permita tener las propiedades requeridas bien documentadas.
    - El segundo reto propuesto, está relacionado con la heterogeneidad de los datos, puesto que la falta de estándares en la representación, no permite compartir ni reusar los datos entre diferentes sistemas.
    - El tercer reto propuesto, está relacionado con la necesidad de poder identificar de los conjuntos de datos, los elementos que son relevantes y útiles para la investigación en analíticas de aprendizaje y de conocimiento.

- El cuarto reto propuesto, está relacionado con el desarrollo de sensores de datos, que permitan una recolección de datos más ágil y que además tengan más relevancia para el proceso investigativo.

**Título:** Design and Implementation of a Learning Analytics Toolkit for Teachers
**Autores:** Dyckhoff, Zielke, Bültmann, Chatti & Schroeder
**Fuente:** Educational Technology & Society
**Año:** 2012

Dyckhoff et. al. (2012) presentan las analíticas de aprendizaje como potentes herramientas para de los docentes, las cuales pueden apoyarlos en el proceso iterativo de mejora de la eficacia de sus cursos y en el mejoramiento del desempeño de sus estudiantes. En su trabajo presentan los fundamentos teóricos, el diseño, implementación y evaluación de "eLAT" (siglas de exploratory Learning Analytics Toolkit), un set de herramientas de analíticas de aprendizaje, que fue construido con el fin de permitir a los docentes explorar y correlacionar el uso de objetos de aprendizaje, las características y comportamiento del usuario y sus resultados de evaluación por medio de indicadores gráficos. El objetivo principal eLAT es procesar grandes conjuntos de datos con gran rapidez (microsegundos) con el fin de ayudar a los docentes a auto-reflexionar sobre sus escenarios de enseñanza-aprendizaje e identificar las posibilidades de intervenir y mejorar los procesos (Dyckhoff, Zielke, Bültmann, Chatti, & Schroeder, 2012). En la Figura 3-6 se muestra el proceso de generación de indicadores.

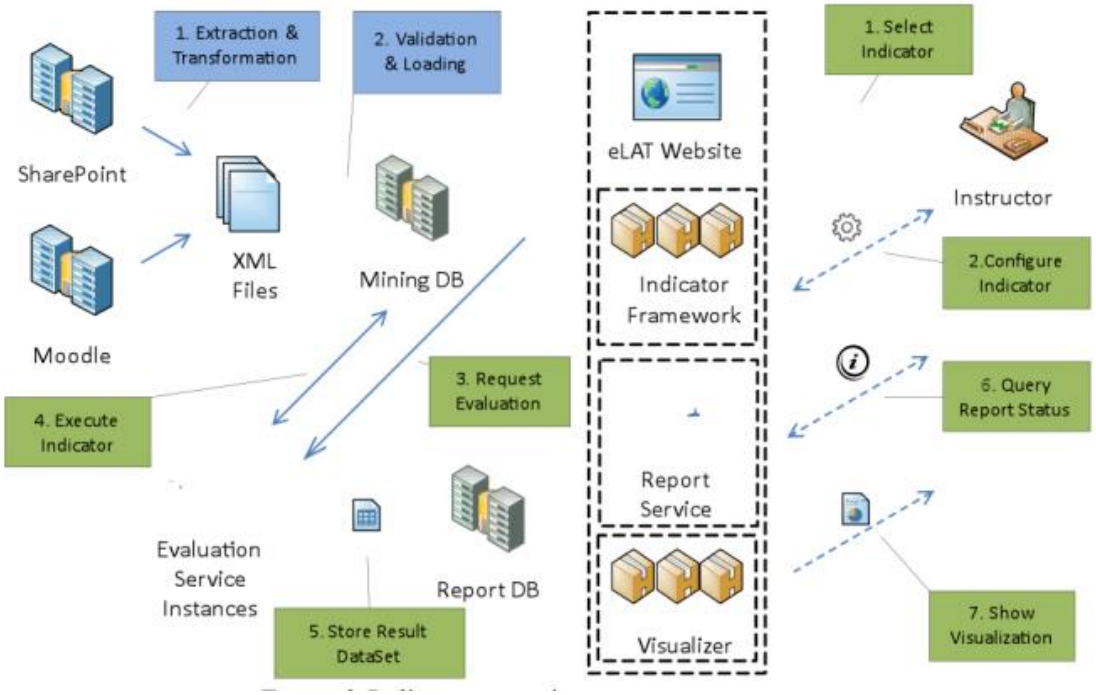



**Figura 3-6:** Proceso de generación de indicadores. Tomado de (Dyckhoff et. al. 2012)

**Fortalezas:**

- La herramienta proporciona una interfaz clara, sencilla, fácil de interpretar y usable para los docentes; al mismo tiempo es suficientemente poderosa y flexible para la exploración de datos e información.
- El prototipo fue probado revisando los criterios como usabilidad, interoperabilidad, extensibilidad, operación en tiempo real y privacidad de los datos, presentando un cumplimiento aceptable en estos requerimientos.

**Limitaciones:**

- El prototipo fue probado con datos de cuatro cursos y cumple con el criterio de interoperabilidad, pero no se describe que haya sido integrado a alguna VLE
- No se incluye en el modelo un componente que revise características propias de los perfiles de aprendizaje de los estudiantes, con el ánimo de poder hacer posteriores recomendaciones.
- Los indicadores presentados a los docentes son estadísticas sencillas, no se tienen todavía integrados algunos análisis más completos que incluyan el uso de minería de datos.

**Oportunidades:**

- Mejorar la herramienta con la integración de un componente de recomendación inteligente y aprovechar su interoperabilidad para pruebas en diferentes VLEs

**Título:** LeMo: A learning analytics application focussing on user path analysis and interactive visualization

**Autores:** Fortenbacher, Beuster, Elkina, Kappe, Merceron, Pursian, Schwarzrock & Wenzlaff

**Fuente:** IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)

**Año:** 2013

En Fortenbacher et. al. (2013) se presenta el prototipo de una aplicación para analíticas de aprendizaje llamada "LeMo", su arquitectura está en la Figura 3-7. Los datos analizados por la aplicación son obtenidos de las actividades de los estudiantes en diferentes plataformas de aprendizaje en línea. LeMo se conecta con varias plataformas, entre ellas Moodle y Clix, de igual manera con la enciclopedia online Chemgapedia. El modelo de datos contiene entidades para los diferentes objetos o recursos de aprendizaje, como cursos, wikis, pruebas y la relación entre estos. El principal objetivo de esta aplicación es el análisis de la trayectoria del usuario, lo cual se une a un potente componente de visualización interactiva (Fortenbacher et al., 2013).

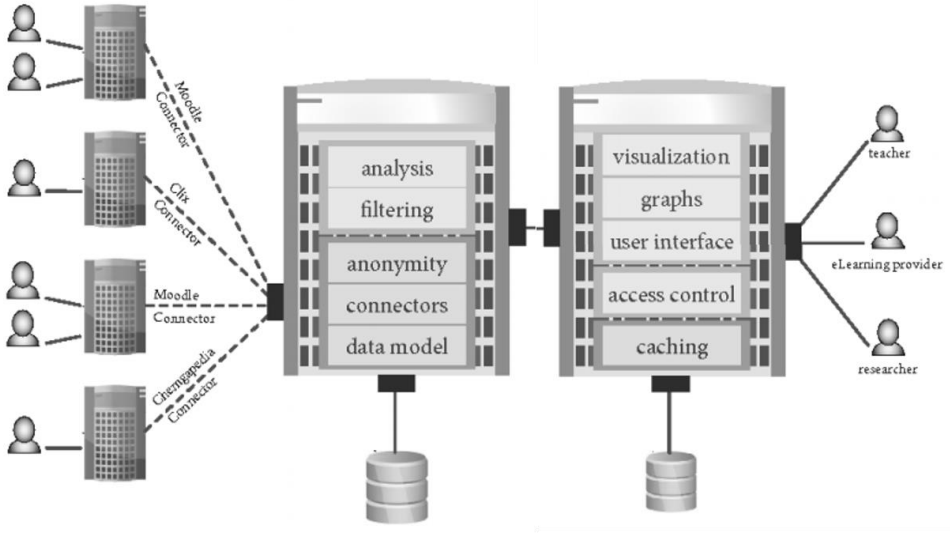



**Figura 3-7:** Arquitectura de LeMo. Adaptada de (Beuster et. al. 2013)

**Fortalezas:**

- Se tiene especial cuidado en el tratamiento de los datos en cuanto al tema de privacidad de los usuarios.
- Se presentan diferentes formas de visualizar los análisis, incluyendo gráficas de líneas, barras, grafos, entre otras.

**Limitaciones:**

- No se hace uso de algoritmos de minería de datos para la extracción de información que permita hacer recomendaciones al proceso educativo.
- La conectividad entre la aplicación y las LMS es limitada.

**Oportunidades:**

- Extender la aplicación para su integración con otras LMS y generar además de los indicadores asociados a la trayectoria del usuario, otros que tengan en cuenta los grupos de usuarios, el rendimiento de los estudiantes y la comunicación y colaboración entre estos.

**Título:** Learning Analytics Framework for Educational Virtual Worlds

**Autores:** Fernández-Gallego, Lama, Vidal, & Mucientes

**Fuente:** Procedia Computer Science

**Año:** 2013

En este artículo se presenta un marco de trabajo para el descubrimiento de flujos de aprendizaje en mundos educativos virtuales 3D (3D EVWs), los cuales cuentan con una naturaleza inmersiva que anima a los estudiantes (como avatares) a realizar actividades de aprendizaje que no son previstas inicialmente por los docentes; actividades como la interacción con otros estudiantes a través de chats, la búsqueda de nuevos contenidos multimedia o el uso de artefactos físicos 3D para aprender acerca de un tema determinado. Los docentes deben conocer el flujo real de las actividades de aprendizaje que los estudiantes han seguido durante un curso en el 3D EVW, esto con el fin de mejorar el diseño del proceso de aprendizaje de dicho curso y a su vez, poder evaluar su eficiencia. De acuerdo a este objetivo, el uso de analíticas de aprendizaje se ve como una buena solución, analizando los datos generados por los avatares se puede descubrir el comportamiento de los estudiantes en la ejecución del proceso de aprendizaje. El marco de trabajo desarrollado se basa en la plataforma OPENET4VE, que a su vez está compilada con la especificación IMS LD del IMS Global Learning Consortium, y tiene la capacidad de monitorear y registrar los eventos generados por los estudiantes y docentes.

Utilizando los datos recolectados algoritmos de minería de datos procesan la información de la cual se extrae el flujo de aprendizaje del curso, lo que permite introducir cambios en el momento en que sea necesario. (Fernández-Gallego, Lama, Vidal, & Mucientes, 2013). El marco de trabajo se puede apreciar en la Figura 3-8.

**Fortalezas:**

- Proponen un marco de trabajo, el cual está construido en una plataforma denominada OPENET4VE, y ésta se fundamenta en la especificación IMS LD (lenguaje de modelado educativo para la representación de las unidades de aprendizaje), lo que le da al framework un nivel de robustez, ya que sigue los lineamientos de una especificación validada en ámbitos educativos y mundos virtuales.
- El marco de trabajo propuesto, permite extraer los flujos de aprendizaje de los cursos a partir de los registros de eventos generados, además de permitir a los docentes realizar cambios en los flujos de aprendizaje según las necesidades encontradas.
- Los 3D EVWs sobre los cuales se centra este trabajo, permiten una interacción continua y dinámica entre los estudiantes y profesores, lo que favorece el proceso de enseñanza-aprendizaje.

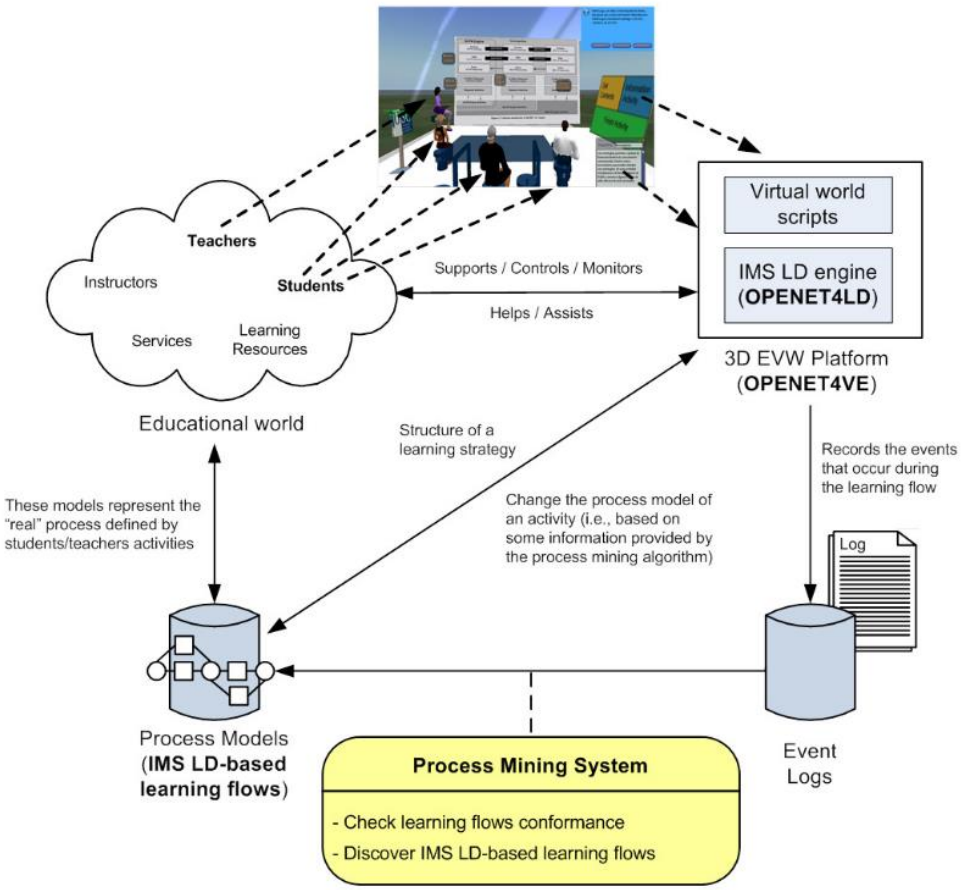



**Figura 3-8:** Framework para LA en mundos educativos virtuales 3D. Tomado de (Fernández-Gallego et. al., 2013)

**Limitaciones:**

- El marco de trabajo está diseñado únicamente para mundos virtuales de educación 3D, dejando de lado entornos de educación convencionales como los ambientes de educación virtual (VLEs) o los cursos en línea masivos y abiertos (MOOCs), los cuales son más utilizados en las instituciones educativas.
- El sistema no realiza recomendaciones concretas para tratar de personalizar el proceso de aprendizaje seguido por los estudiantes en el 3D EVW.

**Oportunidades:**

- Proponer un modelo basado en analíticas de aprendizaje que permita ser aplicado a cualquier tipo de proceso enseñanza-aprendizaje de manera que los datos resultantes de las interacciones en este tipo de proceso generen un impacto en la construcción de cursos educativos.

**Título:** Using learning analytics to identify successful learners in a blended learning course

**Autor:** Kotsiantis, Tselios, Filippidi & Komis

**Fuente:** International Journal of Technology Enhanced Learning

**Año:** 2013

Kotsiantis et. al. (2013) presentan el análisis de un caso de estudio del uso de una plataforma de educación virtual (Moodle) por parte de un grupo de 337 estudiantes. Los datos recogen las interacciones registradas durante tres periodos (años) de un curso con componente presencial y virtual. Para el análisis de los datos se usaron técnicas de minería de datos y análisis estadístico (visualización de datos, árboles de decisión, reglas de asociación y clustering). Los resultados muestran que el bajo rendimiento de los estudiantes en el curso está asociado principalmente con la actitud y percepción hacia la plataforma Moodle; mientras que un alto rendimiento, está asociado al incremento en el uso del sistema de gestión del aprendizaje. El trabajo realizado por los autores sigue cuatro etapas, la recolección de datos, un pre-procesamiento de los datos, la aplicación de métodos de análisis (minería) y la interpretación de los resultados (Kotsiantis, Tselios, Filippidi, & Komis, 2013). En la Tabla 3-1 se pueden apreciar las variables que son consideradas por los autores para el estudio.

**Tabla 3-1:** Las variables extraídas de uso de Moodle por parte de los estudiantes. Adaptada de (Kotsiantis et. al., 2014)

| Variable | Tipo | Descripción |
|---|---|---|
| Computer_at_home | Nominal | El estudiante posee un computador |
| Internet_at_home | Nominal | El estudiante tiene internet en casa |
| Computer_use_per_week | Ordinal (1-5) | Frecuencia semanal de uso del computador por parte del estudiante |
| Ease_of_Moodle_use_ perceptions | Interval (1–5) | Percepción del estudiante de la usabilidad de Moodle |
| Moodle_use_capability_ perceptions | Interval (1–5) | Percepción de los estudiantes acerca de la utilidad del material entregado a través de Moodle |
| Attitude_about_Moodle | Interval (1–5) | Actitud de los estudiantes frente a Moodle |
| Perceived_Moodle_Useful ness_ lesson | Interval (1–5) | Percepción de los estudiantes acerca de la utilidad del material entregado a través de Moodle |

| Variable | Tipo | Descripción |
|---|---|---|
| Perceived_Usefulness_ assignment | Interval (1–5) | Percepción de los estudiantes acerca de la utilidad de las tareas que se les asignaron a través de Moodle |
| Total_Of_ id | Ratio (0–X) | Número total de acciones por estudiante |
| Assignmet_view | Ratio (0–X) | Número de acciones en Moodle en la sección tareas |
| course_view | Ratio (0–X) | Cuántas veces el estudiante accede a la descripción y el material básico de cada sesión de laboratorio semanal |
| forum_add_pos | Ratio (0–X) | Número de publicaciones en el foro por estudiante |
| forum_view | Ratio (0–X) | Cuántas veces el estudiante accede a la sección del foro |
| glossary_view | Ratio (0–X) | Cuántas veces el estudiante accede a la sección del glosario |
| Questionnaire_view | Ratio (0–X) | Cuántas veces el estudiante accede a la sección de cuestionario que contiene rúbricas de evaluación referentes al software educativo |
| resource_view | Ratio (0–X) | Cuántas veces el estudiante accede al servicio, que contiene material complementario y recursos de aprendizaje adicionales |
| user_view | Ratio (0–X) | Cuántas veces el estudiante accede al servicio que contiene el perfil de cada usuario |
| user_view_all | Ratio (0–X) | Cuántas veces el estudiante accede al servicio que contiene la vista general de todos los perfiles de usuario |
| Final note | Ratio (0–X) | Calificación final del estudiante |

**Fortalezas:**

- Se presenta la importancia que tienen los VLEs o LMSs en el apoyo a procesos educativos combinados (educación presencial y virtual), representado esto en la obtención de datos sobre dichos procesos; los cuales pueden ser utilizados para su análisis y descubrimiento de conocimiento posterior.
- Se utilizan métodos de minería de datos, los cuales son predictores significativos del rendimiento de los estudiantes en la plataforma de educación virtual.

**Limitaciones:**

- No se tiene en cuenta dentro de las variables ningún aspecto que aborde el perfil de aprendizaje de los estudiantes.

- No se llegan a hacer recomendaciones basadas en los resultados obtenidos para unas posibles mejoras o personalización ya sea en los contenidos o metodología del curso.

**Oportunidades:**

- Se puede implementar un modelo que, utilizando los métodos de minería de datos aquí propuestos, este en capacidad de analizar estos resultados para posteriormente generar recomendaciones sobre posibles cambios a realizar en el proceso enseñanza-aprendizaje.
- Se puede implementar un sistema de recolección de datos educativos, que además de capturar los datos de las interacciones con las plataformas virtuales, también obtenga datos de las interacciones en las clases presenciales, teniendo así datos de dinámicas educativas diferentes, los cuales pueden enriquecer los análisis para toma de decisiones y fortalecer el proceso de enseñanza-aprendizaje

**Título:** ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform

**Autores:** Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Delgado Kloos, C.

**Fuente:** Computers in Human Behavior

**Año:** 2015

Los autores presentan el desarrollo de un módulo llamado ALAS-KA que se diseñó para extender las analíticas de aprendizaje soportadas por la plataforma “Khan Academy” (KA). La plataforma KA soporta cursos online donde los estudiantes pueden ver videos, resolver ejercicios, alcanzar resultados, entre otros. Esta plataforma ya contaba con un módulo de analíticas de aprendizaje con varias visualizaciones útiles; sin embargo, el trabajo presentado corresponde a una mejora a ese módulo, incluyendo nuevas visualizaciones a nivel grupal e individual. En el caso de las analíticas de tipo individual, permiten verificar los estilos de aprendizaje basados en los indicadores disponibles. Los indicadores implementados apoyan a los docentes y estudiantes en la toma de decisiones, la propuesta fue probada en un caso de estudio con cursos de Física, Química y Matemáticas diseñados para estudiantes de primer año de la Universidad Carlos III de Madrid, que fueron tomados por más de 300 estudiantes. Los dato de las interacciones de los estudiantes con la plataforma con generados y almacenados en la Google App Engine Datastore (Ruipérez-

Valiente, Muñoz-Merino, Leony, & Delgado Kloos, 2015). La arquitectura de ALAS-KA es presentada en la Figura 3-9.

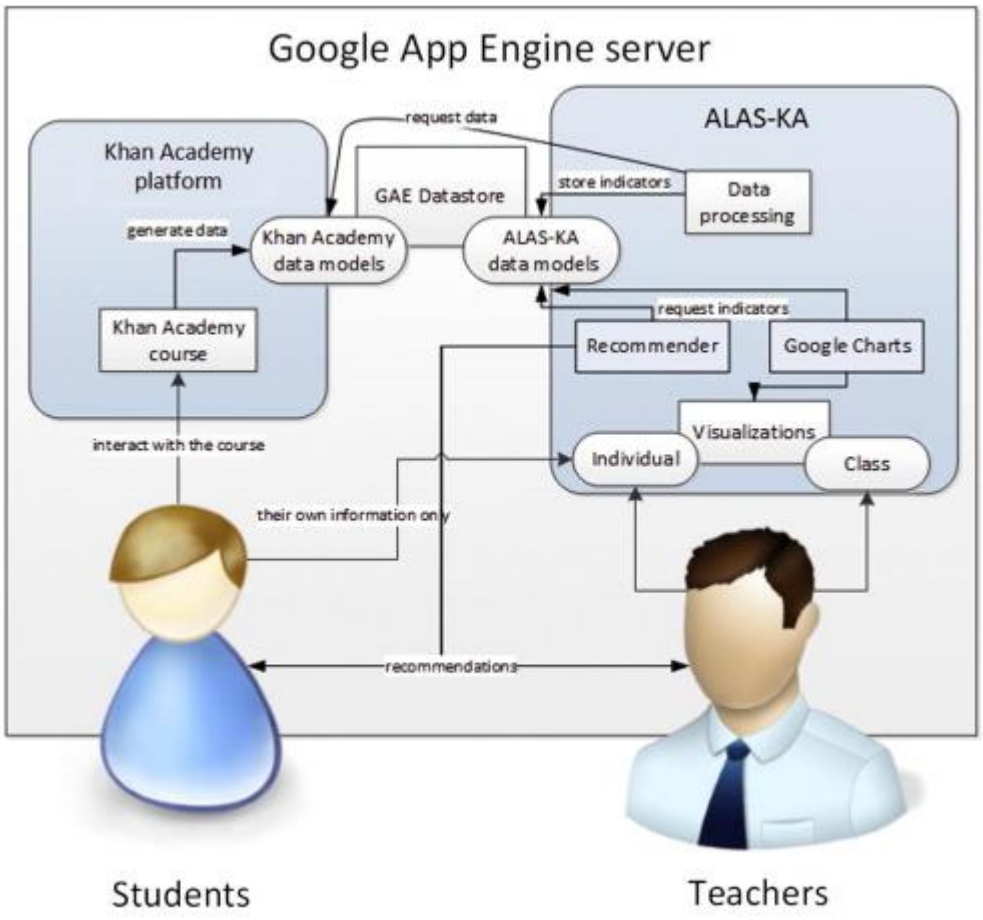



**Figura 3-9:** Arquitectura del módulo ALAS-KA

**Fortalezas:**

- Se mejora el módulo de analíticas de aprendizaje usado por una plataforma virtual de cursos en línea Khan Academy, por medio de la adopción de nuevos indicadores, permitiendo tanto a docentes como estudiantes revisar su rendimiento y apoyar la toma de decisiones.
- En el caso particular de los reportes individuales, se permite a los estudiantes conocer resultados de sus estilos de aprendizaje y qué tan acorde son las actividades que realiza de acuerdo a esto.

**Limitaciones:**

- El módulo ha sido construido para una plataforma particular, lo cual no permite determinar si es fácilmente replicable para otras plataformas de alto uso en la educación superior, como por ejemplo Moodle.
- No se describe con exactitud los datos que maneja la plataforma y sobre los cuales se generan los indicadores

**Oportunidades:**

- Se puede revisar con detenimiento la arquitectura planteada por los autores, y determinar si es posible replicarla para otras plataformas virtuales de aprendizaje.

Para sintetizar los trabajos presentados en este segundo grupo, se muestra en la Tabla 3-2 un comparativo de estos trabajos en cuanto a cinco criterios que se consideran relevantes para la temática abordada.

**Tabla 3-2:** Comparativo de trabajos revisados en el segundo grupo

| Trabajo relacionado | Uso de LMS y cuál | Uso de datos de interacciones | Uso de datos educativos y de contexto | Recomendaciones para personalización del aprendizaje | Enfoque con analíticas de aprendizaje |
|---|---|---|---|---|---|
| Verbert et. al., 2012 | NA | NA | Si | No | Si |
| Dyckhoff et. al 2012 | Si - Moodle | Si | No | No | Si |
| Fortenbacher et. al. 2013 | Si - Moodle | Si | No | No | Si |
| Fernández-Gallego et. al., 2013 | NA | Si | No | No | No es claro |
| Kotsiantis et. al., 2013 | Si - Moodle | Si | No | No | No |
| Ruipérez-Valiente et. al., 2015 | Si - Khan Academy | Si | No | Si | Si |

## 3.3 Conclusiones de capítulo

Después de hacer esta revisión del estado del arte se encontró que las temáticas que se pretenden tratar en esta tesis son campos de investigación emergentes, que han venido tomando fuerza en los últimos años, despertando el interés de la comunidad académica por abordarlos. Se logró encontrar en estos trabajos relacionados que existen buenas fortalezas que ayudan a determinar con claridad los avances que se han dado en el campo; a pesar de que existen también algunas limitaciones, si estas son unidas a las oportunidades planteadas, se encuentra que existen espacios que pueden ser atacados en parte, por medio de esta tesis. Estos espacios principalmente se encuentran asociados a la necesidad de contar con un modelo que permita no solo aplicar las analíticas de aprendizaje para extraer indicadores de las interacciones entre los estudiantes y las plataformas de educación virtual, sino que a su vez, lleve a la generación de algunas recomendaciones para el mejoramiento de los procesos y orientar el tipo de actividades de aprendizaje según los patrones o modelos obtenidos a partir del uso de técnicas de minería de datos. Teniendo en cuenta lo anterior, en el siguiente capítulo se presenta el modelo propuesto que pretende el descubrimiento de patrones en las interacciones entre estudiantes y plataformas virtuales de educación mediante el uso de analíticas de aprendizaje y técnicas de minería de datos.

# 4. Modelo Propuesto

En este capítulo se presenta y describe el modelo propuesto para el descubrimiento de patrones en datos académicos e interacciones de estudiantes con plataformas virtuales de educación con el uso de analíticas de aprendizaje. En primer lugar, se hace una identificación de cada uno de los componentes del modelo y luego se ahondará, describiendo de manera detallada las particularidades de cada componente.

Para la construcción del modelo para el descubrimiento de patrones que se propone en esta tesis se decidió considerar varias formas de aplicar las analíticas de aprendizaje, esto con el fin de tener resultados desde varios enfoques y lograr un mejor entendimiento de los datos fuente que se analizarán.

En la Figura 4-1 se puede observar el modelo propuesto y a continuación se describen los elementos que hacen parte de cada uno de los componentes considerados.

## 4.1 Componente 1: Fuentes de datos

En este componente se encuentran los datos del proceso educativo y de las interacciones con las plataformas virtuales de aprendizaje. Para el alcance definido en esta tesis, se considera información educativa representada en datos de tipo académico y datos que representan las interacciones con las plataformas virtuales de aprendizaje. Es necesario considerar la mayor cantidad de información académica posible, ya que de esta manera se pueden lograr resultados confiables y que permitan generar conocimiento para soportar la toma de decisiones eficaces por parte de las directivas universitarias o incluso de los docentes que imparten las asignaturas, representadas en la formulación de estrategias relacionadas con el mejoramiento de los procesos de enseñanza-aprendizaje soportados o apoyados con el uso de plataformas virtuales de aprendizaje. A continuación, se detallarán las fuentes de datos consideradas.

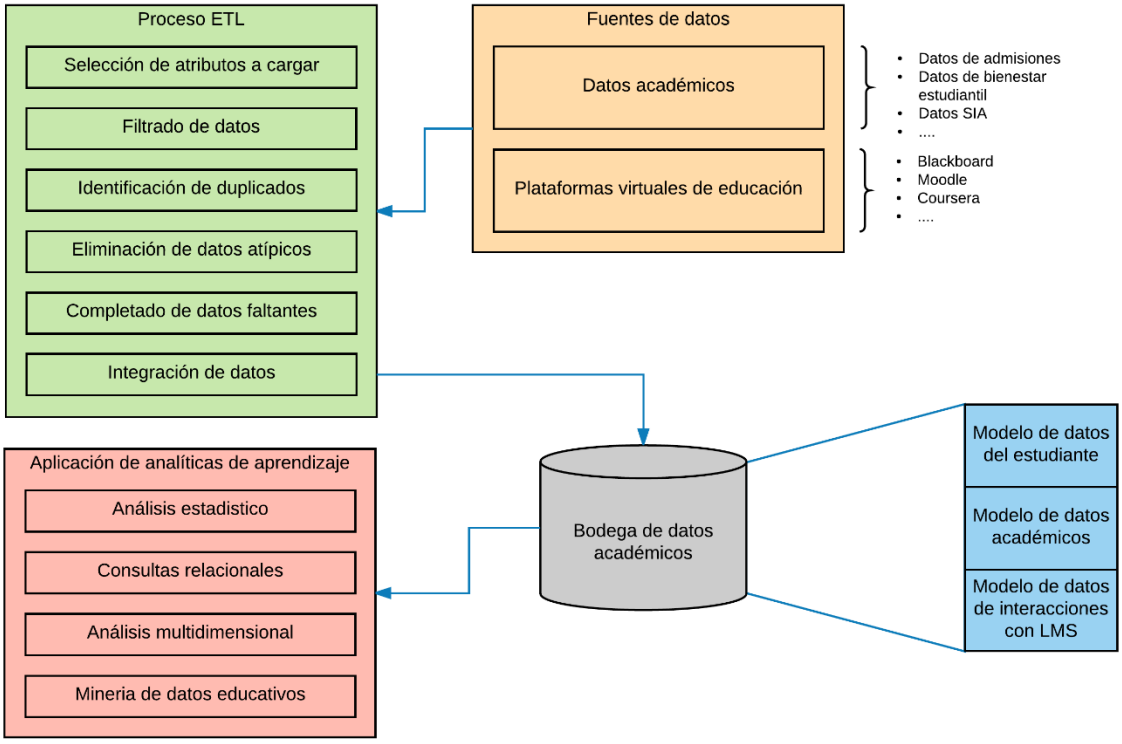



**Figura 4-1:** Modelo propuesto

### 4.1.1 Datos académicos

Esta fuente de datos puede incluir bases de datos que almacenan información personal y académica de los estudiantes pertenecientes a una institución de educación superior. Esto incluye datos socioeconómicos, datos familiares, datos de procedencia, datos del proceso de admisión, datos de los resultados obtenidos en el transcurso del proceso de formación, entre otros.

### 4.1.2 Datos de las interacciones de los estudiantes con las plataformas virtuales de aprendizaje

Esta fuente de datos puede incluir bases de datos de diferentes plataformas virtuales de educación utilizadas dentro del proceso de formación de los estudiantes. Estas plataformas suelen almacenar una cantidad considerable de datos que corresponden a registros de diferentes procesos ejecutados en ellas. Principalmente se pueden encontrar datos

correspondientes a: ingresos a la plataforma representado en fechas y horas de acceso y salida, cursos inscritos, contenidos de los cursos, resultados o notas obtenidas en las actividades, participación de los estudiantes en las diferentes acciones, datos personales y de registro, roles dentro de la plataforma, entre otras.

## 4.2 Componente 2: Proceso de ETL

Si bien es cierto que una gran cantidad de datos representa un beneficio para la aplicación de algunas de las técnicas o algoritmos dentro del modelo de descubrimiento de patrones, es necesario considerar también que se requiere de datos que sean confiables para obtener resultados confiables. Por lo anterior se considera dentro del modelo un componente correspondiente al proceso de extracción, transformación y carga de datos (ETL por sus siglas en inglés). En este componente se busca atender tareas de este tipo de procesos como identificación de datos duplicados, filtrado de datos, integración de datos, selección de atributos a cargar, eliminación de datos atípicos y completado de algunos datos faltantes.

Para realizar los procesos de ETL existen diferentes herramientas que ayudan a automatizar el proceso y que son de uso libre; sin embargo, en algunos casos, se requiere del desarrollo de soluciones propias debido a la complejidad del dominio de datos que se trabaja y de las características de los mismos, como tipos de formatos, medios y estructura de almacenamiento.

No obstante, ya sea que se decida utilizar una herramienta disponible en el medio o desarrollar una solución particular, es necesario garantizar que se cubran todos los requerimientos necesarios para obtener datos confiables. Estos procesos de ETL suelen requerir del apoyo de personal experto tanto en el área de sistemas como en el tratamiento de los datos de sistemas de información académicos y de las plataformas virtuales de aprendizaje.

## 4.3 Componente 3: Bodega de datos académica

Este componente corresponde al esquema de almacenamiento que dará soporte a la aplicación de las analíticas. Para ello se ha concebido una bodega de datos académica en la cual se abstraen tres modelos de datos a saber, el modelo de datos del estudiante, el modelo de datos académicos y el modelo de datos de las interacciones con los LMS. Como se ha indicado anteriormente, es necesario considerar un dominio de datos educativos completo para poder articular las analíticas a aplicar y para ello se requiere poder centralizar los datos bajo un esquema. A continuación, se describen los tres modelos de datos incluidos.

### 4.3.1 Modelo de datos del estudiante

Corresponde al esquema para el almacenamiento de la información propia del estudiante, esto es datos personales y datos familiares, como procedencia, condiciones socioeconómicas del estudiante y su familia, datos del colegio del que proviene el estudiante.

### 4.3.2 Modelo de datos académicos

Corresponde al esquema para el almacenamiento de la información del proceso del estudiante desde que ingresa a la institución de educación superior, esto es datos del ingreso, datos de asignaturas cursadas, de notas alcanzadas, créditos cursados, avances en el plan de estudios, etc.

### 4.3.3 Modelo de datos de las interacciones con LMS

Corresponde al esquema para el almacenamiento de la información de los procesos de interacción del estudiante con las plataformas virtuales de aprendizaje que soportan los cursos de su programa académico. Incluye datos de los cursos inscritos, contenidos de estos cursos, ingresos a la plataforma, tiempo de estos ingresos, actividades desarrollas, entre otros.

## 4.4 Componente 4: Aplicación de analíticas de aprendizaje

Después de descritos los componentes correspondientes a las fuentes de datos académicos y de interacciones en las plataformas virtuales de aprendizaje, al tratamiento inicial de los mismos y su respectivo almacenamiento en una bodega de datos con un diseño hibrido, ahora se hará la descripción del componente final, que corresponde a la aplicación de las analíticas de aprendizaje.

Para la construcción del modelo, en primer lugar, es necesario caracterizar las estrategias para la aplicación de las analíticas de aprendizaje. Una vez realizada la revisión del estado del arte y el marco teórico, se identificó que existen diferentes propuestas y enfoques para la aplicación de estas analíticas, claramente se encuentra, que los autores proponen modelos y soluciones que se van ajustando de acuerdo al tipo de datos educativos que desean usar y a los datos que recolecten las plataformas que soportan los procesos de aprendizaje virtual de las instituciones objeto de estudio. Por lo anterior, y sin identificar una metodología para la implementación de analíticas de aprendizaje que claramente esté enfocada a los datos que se desea analizar con el modelo de esta tesis, se decide hacer un agrupamiento de los análisis bajo cuatro enfoques.

Estos enfoques van desde un estado de descripción de los datos hasta un estado de recomendación de estrategias para mejorar el proceso. En la Figura 4-2 se representan los enfoques que se van a adoptar para la aplicación de las analíticas.

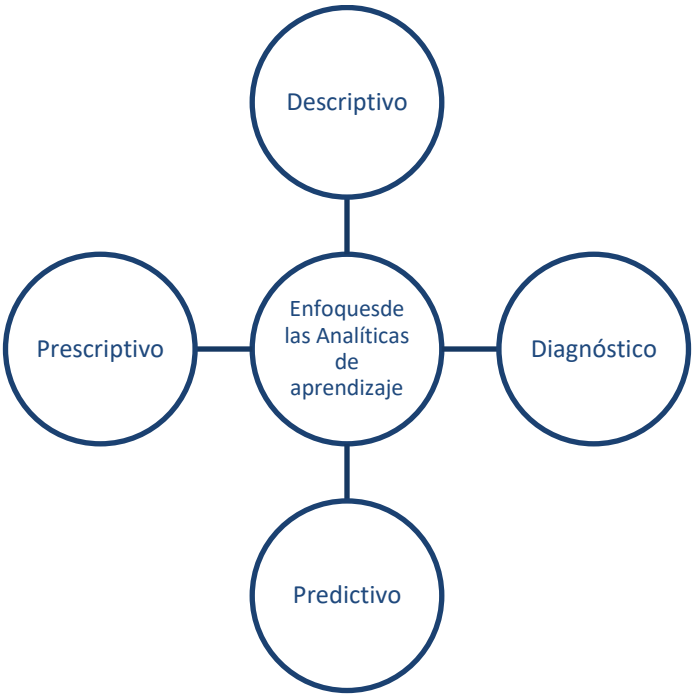



**Figura 4-2:** Enfoques de las analíticas de aprendizaje

A continuación, se define cada uno de los enfoques.

### 4.4.1 El enfoque descriptivo

Este se concentra en el análisis inicial de los datos, para contextualizar el dominio que se trabaja, los datos con que se cuenta y cómo se comportan. Se puede hacer por medio de la aplicación de estadísticos básicos. En este enfoque se hace una caracterización por medio de indicadores estadísticos de forma global de las diversas variables tratadas con el fin de establecer una identificación de tendencias en los datos. Se realizan cálculos para identificar las medidas de tendencia central, medidas de dispersión y medidas de posición.

### 4.4.2 El enfoque diagnóstico

Tiene por objeto conseguir una serie de resultados iniciales, puede estar representado en la ejecución de consultas relacionales o multidimensionales sobre los datos. Las consultas relacionales corresponden a la extracción de conocimiento desde cualquier tipo de almacenamiento de datos de tipo de relacional. El lenguaje de programación usado para

la generación de este tipo de consultas es el SQL. Las consultas relacionales usan las operaciones fundamentales del algebra relacional como: selección, proyección, producto cartesiano, unión y diferencia de conjuntos. Otras operaciones son: la unión, intersección y operación de división, las cuales se expresan en términos de las siguientes operaciones básicas: count, sum, avg, min y max (Ramírez Romero, Patiño Ortiz, & Patiño Ortiz, 2015). El análisis multidimensional corresponde a la revisión y categorización de la información basada en múltiples dimensiones y medidas. Este tipo de análisis ayuda en el modelamiento de los procesos cuando se involucran conjuntos de datos complejos y de gran tamaño (Hira & Deshpande, 2015). Para el análisis multidimensional se usan operaciones como Roll up (drill up) que se ocupa de escalar la jerarquía o reducir dimensión, Drill down (roll down) permite ir desde un alto nivel de resumen a un nivel más detallado, slice permite hacer cortes o proyecciones, dice se usa para seleccionar y rotate para cambiar dimensiones unas con otras (Araníbar, 2003) (Feliciano Morales, Cuevas Valencia, & Martínez Castro, 2016).

### 4.4.3 El enfoque predictivo

Este se concentra en tratar de mostrar lo que podría suceder a partir del análisis de los datos que se tienen y por medio de la aplicación de técnicas más complejas, como por ejemplo la minería de datos educativos (Educational Data Mining), el aprendizaje máquina (Machine Learning) e incluso el aprendizaje profundo (Deep Learning); en estas técnicas se encuentra gran variedad de algoritmos supervisados y no supervisados que permiten atender tareas fundamentales como la selección de atributos y el descubrimiento o predicción de comportamientos.

### 4.4.4 El enfoque prescriptivo

Corresponde a las recomendaciones sobre qué se debe hacer después de los hallazgos obtenidos por medio de la adopción de estrategias que permitan atacar los puntos identificados como débiles o incentivar los aspectos positivos. Para llegar a la aplicación de este enfoque de las analíticas de aprendizaje es necesario contar con un análisis por parte de personal experto, personal encargado de tomar o apoyar la toma de decisiones dentro de la institución educativa.

## 4.5 Conclusiones del capítulo

En este capítulo se presentó una visión general del modelo para el descubrimiento de patrones en datos académicos e interacciones en plataformas virtuales de aprendizaje. Las principales características del modelo son:

- El modelo se encuentra orientado tanto a los datos de las interacciones de los estudiantes en las plataformas virtuales de aprendizaje como a los datos académicos y datos del estudiante recolectados por las instituciones de educación superior.
- El modelo presenta cuatro componentes principales que son: las fuentes de datos, el proceso de ETL, el almacenamiento bajo una bodega de datos académica y la aplicación de analíticas de aprendizaje.
- Se consideran cuatro enfoques para la estrategia de aplicación de las analíticas de aprendizaje. El enfoque descriptivo, que tiene en cuenta el análisis exploratorio inicial de los datos; el enfoque diagnóstico que tiene a cargo el descubrimiento de relaciones y correlaciones iniciales y empieza a dar resultados de aspectos a tener en cuenta y hechos que se están presentando; el enfoque predictivo que considera la aplicación de técnicas más complejas para determinar posibles comportamientos a repetir, acciones que se replicarán; el enfoque prescriptivo viene acompañado de recomendaciones sobre estrategias a tomar indicando posibles medidas por adoptar de acuerdo a los resultados obtenidos.

Considerando este modelo, se pueden adelantar nuevos proyectos mejorando algunos de los elementos involucrados en los diferentes componentes, por ejemplo, incluyendo nuevas fuentes de datos, nuevos modelos de datos dentro de la bodega (por medio de tablas de hechos y/o dimensiones), nuevos enfoques o técnicas para la aplicación de las analíticas de aprendizaje.

# 5. Validación del modelo

En este capítulo se presenta una aplicación del modelo de descubrimiento de patrones en datos académicos y de interacciones entre estudiantes y plataformas virtuales de aprendizaje a un caso de estudio específico, con el objeto de validar y evaluar la viabilidad de su implementación; comprobando también, la pertinencia del modelo y de las analíticas de aprendizaje para datos académicos de una institución de educación superior. A continuación, se seguirán los pasos descritos en el modelo en la sección anterior.

## 5.1 Caso de estudio

Para validar la propuesta se tomó como caso de estudio una serie de datos académicos suministrada por la Dirección Académica de la Universidad Nacional de Colombia – Sede Manizales, los cuales fueron extraídos del Sistema de Información Académica (SIA) de dicha institución y que corresponden a los periodos comprendidos entre el primer semestre de 2009 al primer semestre de 2015. La institución en menciona registra en este sistema datos de cada estudiante desde su proceso de admisión e ingreso. También se contó con otra serie de datos correspondiente a datos extraídos de las interacciones de los estudiantes de esta misma institución con dos plataformas virtuales de aprendizaje, Blackboard y Moodle. Estos datos corresponden a lo recolectado en el segundo semestre de 2013 y primer semestre de 2014 para el caso de Blackboard y de 2014 para Moodle.

Un proyecto de investigación realizado en conjunto entre la Dirección Académica de la Universidad y el Grupo de Investigación en Ambientes Inteligentes Adaptativos – GAIA de la misma institución, permitió el acceso a la información bajo políticas de protección y privacidad de los datos.

## 5.2 Componente 1: Fuentes de datos

Ahora se describirán de forma detallada estas fuentes de datos, representadas en los datos académicos y los datos de las interacciones de los estudiantes en las plataformas virtuales de aprendizaje usadas en la institución educativa.

### 5.2.1 Datos académicos

Los datos académicos brindados por la institución, como se mencionó anteriormente comprenden desde el primer semestre del año 2009 y el primer semestre del año 2015. En la Tabla 5-1 se presenta un resumen de la cantidad de registros recibidos por cada periodo académico en cuanto a los extraídos del SIA.

**Tabla 5-1:** Resumen de los registros presentes en el SIA

| Año | Semestre | Número de registros | Registros Acumulados |
|---|---|---|---|
| 2009 | 1 | 4.094 | 4.094 |
| 2009 | 2 | 4.018 | 8.112 |
| 2010 | 1 | 4.162 | 12.274 |
| 2010 | 2 | 4.012 | 16.286 |
| 2011 | 1 | 4.167 | 20.453 |
| 2011 | 2 | 4.134 | 24.587 |
| 2012 | 1 | 4.385 | 28.972 |
| 2012 | 2 | 4.475 | 33.447 |
| 2013 | 1 | 4.678 | 38.125 |
| 2013 | 2 | 4.645 | 42.770 |
| 2014 | 1 | 4.563 | 47.333 |
| 2014 | 2 | 4.714 | 52.047 |
| 2015 | 1 | 4.857 | 56.904 |

Luego de una revisión de estos registros contenidos en la base de datos del SIA, se identificó un total de 91 variables, de las cuales se establece la existencia de variables descriptivas y variables indicadoras. Se hizo una agrupación inicial de dichas variables en

seis categorías: identificativas o datos personales, admisión académica, socio-económicas, antecedentes académicos, familiares y de desempeño académico. En la Tabla 5-2 se presentan estas categorías y las variables identificadas para cada una, mostrando también el tipo de variable (cualitativa nominal politómica, cualitativa ordinal politómica, cualitativa nominal dicotómica, cualitativa ordinal dicotómica, cuantitativa discreta y cuantitativa continua) y una columna de observaciones donde se establecen algunos valores paramétricos de las mismas.

**Tabla 5-2:** Identificación de variables del SIA

| Categoría | Nombre variable | Tipo | Observación |
|---|---|---|---|
| Identificativo o datos personales | Tipo de documento de identidad | Cualitativa Nominal Politómica | Se establecen 4 posibles valores identificados así: E –; O -; T – Tarjeta de identidad; C – Cédula de ciudadanía |
| | Número del documento de identidad | Cualitativa Nominal Politómica | |
| | Apellido 1 | Cualitativa Nominal Politómica | Primer apellido del estudiante |
| | Apellido 2 | Cualitativa Nominal Politómica | Segundo apellido del estudiante |
| | Nombre | Cualitativa Nominal Politómica | Nombre completo del estudiante |
| | Correo electrónico institucional | Cualitativa Nominal Politómica | Correo electrónico (generalmente institucional) |
| | Fecha de nacimiento | Cuantitativo Discreto | Fecha de nacimiento, separada por tres atributos (Año, mes, día) |
| | Sexo | Cualitativa Nominal Dicotómica | Con dos posibles valores: F – Femenino; M – Masculino |
| | Estado civil | Cualitativa Nominal Politómica | Se toma uno de ocho opciones: Casado; Nulo, Soltero, Unión libre, Viudo, Separado, Divorciado, No definido |
| | Distrito Militar | Cualitativa Nominal Politómica | Códigos no estandarizados. |
| | Libreta Militar | Cualitativa Nominal Politómica | Valores numéricos es interpretado como existencia de esta. |
| | Código departamento | Cualitativa Ordinal Politómica | Lugar de nacimiento. Código propio del sistema de información |
| | Departamento | Cualitativa Nominal Politómica | Lugar de nacimiento. Identificación exponiendo el nombre |
| | Código municipio | Cualitativa Ordinal Politómica | Lugar de nacimiento. Código propio del sistema de información |
| | Municipio | Cualitativa Nominal Politómica | Lugar de nacimiento. Identificación exponiendo el nombre |

| Categoría | Nombre variable | Tipo | Observación |
|---|---|---|---|
| | Nacionalidad 1 | Cualitativa Ordinal Politómica | Propios códigos no identificados |
| | Nacionalidad 2 | Cualitativa Ordinal Politómica | Códigos no identificados |
| | Extranjero | Cualitativa Nominal Dicotómica | Existen dos opciones: N – no es extranjero, S - es extranjero. |
| | Lugar de residencia | Cualitativa Nominal Politómica | Existen cuatro opciones, de las cuales se identifican las siguientes cantidades de registros: Especial= 482, SIN INFO=28, Dentro=6495, Fuera=3773 |
| Admisión académica | Código Facultad | Cualitativa Ordinal Politómica | Identificador institucional |
| | Facultad | Cualitativa Nominal Politómica | Nombre de la facultad |
| | Plan | Cualitativa Ordinal Politómica | Código del plan educativo |
| | Carrera | Cualitativa Ordinal Politómica | Código de la carrera |
| | Programa | Cualitativa Nominal Politómica | Nombre de la carrera |
| | Código estudiante | Cualitativa Nominal Politómica | Código estudiantil |
| | Fecha de creación | Cualitativa Ordinal Politómica | Fecha de creación en el sistema académico |
| | Fecha inicio de estudios | Cualitativa Ordinal Politómica | Periodo del primer semestre estudiado. Periodo académico de inicio de clases, o primer semestre de estudios |
| | Tipo de acceso | Cualitativa Ordinal Politómica | Código del tipo de acceso o admisión a la universidad |
| | Acceso | Cualitativa Nominal Politómica | Se identifica forma de calificar en la admisión (Examen, Traslado, no se identifica) |
| | Tipo de sub acceso | Cualitativa Ordinal Politómica | Código de programa de acceso |
| | Sub acceso | Cualitativa Nominal Politómica | Nombre del programa con el que se accedió a la universidad, se logran identificar 10 opciones |
| | Puntaje | Cuantitativo Continuo | Valor calificación obtenida en el examen de admisión |
| | Puesto | Cualitativa Ordinal Politómica | Posición adquirida según el puntaje de admisión. |
| Socio-económico | Tipo estrato | Cualitativa Ordinal Politómica | Código del estrato social |
| | Estrato | Cualitativa Ordinal Politómica | Nombre del estrato social. Se identifican 8 escalas, entre ellas "No estratifica", "No informa" |

| Categoría | Nombre variable | Tipo | Observación |
|---|---|---|---|
| | Becado | Cualitativa Nominal Dicotómica | Toma uno de tres posibles valores: Vacío, N, S |
| | Pensión | Cuantitativo Discreto | Valor |
| | Tipo pensión | Cualitativa Nominal Politómica | Se toma uno de tres valores: Vacío, M, A |
| | Año pensión | Cuantitativo Discreto | Año de pensión |
| | Tipo de vivienda | Cualitativa Ordinal Politómica | Cinco posibles valores: Vacío, 1, 2, 3, 4. |
| | Vivienda | Cualitativa Nominal Politómica | Nombre del tipo de vivienda: Sin propiedad raíz, Pagando crédito hipotecario, Vivienda propia, No informa |
| | PBM[1] Calculado | Cuantitativo Discreto | Valor PBM obtenido de análisis datos de económicos de admisión |
| | PBM Colegio | Cuantitativo Discreto | PBM identificado de la institución educativa donde proviene el estudiante |
| | Factor | Cuantitativo Continuo | |
| | PBM Asignado | Cuantitativo Discreto | Valor PBM asignado (permanente) |
| | PBM Periodo | Cuantitativo Discreto | Valor PBM asignado en el semestre (permite identificar cuando el estudiante solicita generar nuevamente el cálculo de este factor) |
| | Recibo Pagado | Cuantitativo Discreto | Valor total a pagar para inscripción de semestre |
| | Matricula académica | Cuantitativo Discreto | Valor por derechos de matricula |
| | Bienestar universitario | Cuantitativo Discreto | Valor por derechos de bienestar universitario |
| | Sistematización | Cuantitativo Discreto | Valor por derechos de sistematización |
| | Seguro | Cuantitativo Discreto | Valor del costo del seguro universitario |
| | Icetex | Cualitativa Nominal Dicotómica | Préstamo ICETEX[2]: Vacío=10692, Si=86 |
| | Pensión | Cuantitativo Continuo | |
| | Colegio | Cuantitativo Continuo | |
| | Estrato | Cuantitativo Continuo | |
| | Lugar residencia | Cuantitativo Continuo | |
| | Vivienda | Cuantitativo Continuo | |

[1] El Puntaje Básico de Matrícula PBM es el indicador que utiliza la Universidad Nacional de Colombia para hacer la clasificación socioeconómica de los estudiantes y asignar la matrícula.
[2] ICETEX: El Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior, es una entidad del Estado Colombiano que promueve la Educación Superior a través del otorgamiento de créditos educativos y su recaudo, con recursos propios o de terceros, a la población con menores posibilidades económicas y buen desempeño académico.

| Categoría | Nombre variable | Tipo | Observación |
|---|---|---|---|
| | Ingresos | Cuantitativo Continuo | |
| | Número de hijos | Cuantitativo Discreto | |
| | EPS[3] | Cualitativa Nominal Politómica | Nombre de la entidad prestadora del servicio de salud al que está afiliado el estudiante |
| Antecedentes académicos | Titulo pregrado | Cualitativa Nominal Dicotómica | Trata de identificar si el estudiante ya posee un título de pregrado. Nulos=5180, S=3283, N=2315 |
| | Plantel | Cualitativa Ordinal Politómica | Código del colegio en el que estudió el estudiante |
| | Tipo colegio | Cualitativa Nominal Politómica | Son siglas, de las cuales se asume: O – Colegio Oficial P – Colegio Público T – Colegio Técnico N – Colegio Normalista |
| | Jornada colegio | Cualitativa Nominal Politómica | Identifica la jornada académica de la institución. Los valores pueden ser: Sin Registro Mañana - Completa U Ordinaria Tarde - Noche - NV - Sabatina - Dominical |
| | Característica colegio | Cualitativa Nominal Politómica | Se identifica el tipo de plantel del colegio: Plantel Privado - Sin Registro, Plantel Privado - Urbana, Plantel Oficial - Urbana, Plantel Nocturno - Urbana, Oficial Plantel Oficial -, Plantel Oficial - Rural, Otros Planteles, Plantel Privado -, Plantel Privado - Rural, Plantel Oficial, Plantel Nocturno, Vacío, Plantel Nocturno - Rural |
| | Modalidad académica | Cualitativa Nominal Politómica | Identificación del tipo de modalidad académica: Vacío, No Oficial, Oficial, NV |
| | Nombre del colegio | Cualitativa Nominal Politómica | Nombre de la institución |
| | Calendario | Cualitativa Nominal Politómica | B – Septiembre/Junio F – A – Febrero/Noviembre NV –. |
| | Código municipio del colegio | Cualitativa Ordinal Politómica | Código del Municipio donde se encuentra el colegio |
| | Municipio del colegio | Cualitativa Nominal Politómica | Nombre del Municipio donde se encuentra el colegio |
| | Código departamento del colegio | Cualitativa Ordinal Politómica | Código del Departamento donde se encuentra el colegio |
| | Departamento del colegio | Cualitativa Nominal Politómica | Nombre del Departamento donde se encuentra el colegio |

[3] EPS: En Colombia, EPS es la sigla usada para referirse a la Entidad Promotora de Salud.

| Categoría | Nombre variable | Tipo | Observación |
|---|---|---|---|
| Familiar | Número de hermanos | Cuantitativo Discreto | Cantidad de hermanos del estudiante |
| | Número de hermanos en la UNAL[4] | Cuantitativo Discreto | Cantidad de hermanos del estudiante que estudian en la universidad |
| | Municipio Familiar | Cualitativa Nominal Politómica | Nombre del Municipio donde vive la familia |
| | Dirección Familiar | Cualitativa Nominal Politómica | Dirección del domicilio familiar |
| | Teléfono Familiar | Cualitativa Nominal Politómica | Teléfono del domicilio familiar |
| | Municipio Actual | Cualitativa Nominal Politómica | Nombre del Municipio en el que vive el estudiante |
| | Dirección Actual | Cualitativa Nominal Politómica | Dirección del domicilio del estudiante |
| | Teléfono 1 Actual | Cualitativa Nominal Politómica | Teléfono del domicilio del estudiante |
| | Teléfono 2 Actual | Cualitativa Nominal Politómica | Otro teléfono del domicilio del estudiante |
| Desempeño académico dentro de la UNAL | PAPA[5] | Cuantitativo Continuo | Promedio aritmético ponderado acumulado |
| | Graduado | Cualitativa Nominal Dicotómica | Se indica si el estudiante finalizo o no sus estudios. |
| | Aprobadas | Cuantitativo Discreto | Cantidad de asignaturas aprobadas |
| | Homologadas | Cuantitativo Discreto | Cantidad de asignaturas homologadas |
| | Inscritas | Cuantitativo Discreto | Cantidad de asignaturas inscritas |
| | Créditos aprobados | Cuantitativo Discreto | Cantidad de créditos aprobadas |
| | Créditos adicionales | Cuantitativo Discreto | Cantidad de créditos adicionales |
| | Créditos homologados | Cuantitativo Discreto | Cantidad de créditos homologados |
| | Créditos inscritos | Cuantitativo Discreto | Cantidad de créditos inscritos |
| | Periodo académico | Cualitativa Ordinal Politómica | Se indica el año seguido por el código del periodo académico, del cual se identifica: 01 – Febrero/Mayo 02 – Agosto/Noviembre |

[4] UNAL: Sigla empleada para referirse a la Universidad Nacional de Colombia.
[5] PAPA: Promedio Aritmético Ponderado Acumulado, utilizado para medir el rendimiento académico de los estudiantes de la Universidad Nacional de Colombia.

## 5.2.2 Datos de las interacciones de los estudiantes en las plataformas Blackboard y Moodle

Con respecto a los datos de las interacciones se recibieron registros de las plataformas virtuales de aprendizaje Blackboard y Moodle.

### 5.2.2.1 Blackboard

En cuanto a la plataforma educativa Blackboard (Blackboard Inc, 2015), se contó con un número considerable de bases de datos. En la Tabla 5-3 se presenta un consolidado de estas bases de datos, donde se indica el número de tablas que contiene cada una.

A través de una revisión de la documentación presente en internet respecto a la plataforma Blackboard y de la visualización de los datos, se logra identificar que se cuenta con diversos esquemas y diversas bases de datos para los mismos periodos académicos, es decir, se tienen registros repetidos en diferentes copias de seguridad. Estas bases de datos varían según las funcionalidades que se requieren, muchas de ellas son para operaciones administrativas y otras tantas para operaciones funcionales de la plataforma como herramienta educativa.

**Tabla 5-3:** Consolidado de la base de datos de Blackboard

| ESQUEMAS BB | NOMBRE DE LA BASE DE DATOS | VERSION PLATAFORMA BB | VERSION MOTOR BD | CANTIDAD DE TABLAS | | |
|---|---|---|---|---|---|---|
| | | | | SIN DATOS | CON DATOS | TOTAL |
| As_core | Bb_bb60_backup_201406180100 | 8.4 | 2005 | | | |
| | bb_bb60_backup_201406200300 | 8.7 | 2005 | 271 | 167 | 438 |
| | BBLEARN_backup_2014_10_03_030006_0926210 | 9.1 | 2008 | 362 | 294 | 656 |
| Stats | Bb_bb60_stats_backup_201406170100 | 8.4 | 2005 | | | |
| | bb_bb60_stats_backup_201406190300 | 8.7 | 2005 | 9 | 11 | 20 |
| | BBLEARN_stats_backup_2014_10_03_030006_3266912 | 9.1 | 2008 | 59 | 74 | 133 |
| Admin | Bbadmin_backup_201410200100 | 8.4 | 2005 | | | |
| | bbadmin_backup_201410150300 | 8.7 | 2005 | | | |
| | BBLEARN_admin_backup_2014_10_03_030006_3066852 | 9.1 | 2008 | 7 | 6 | 13 |
| INDEFINIDO | BBLEARN_cms_backup_2014_10_03_030006_3136873 | 9.1 | 2008 | 81 | 29 | 110 |
| | BBLEARN_cms_doc_backup_2014_10_03_030006_3196891 | 9.1 | 2008 | 12 | 15 | 27 |

Es importante resaltar que las bases de datos de Blackboard proporcionan datos del segundo periodo académico de 2013 y primer periodo académico del 2014. También se identifica que las copias de seguridad correspondientes a las bases de datos de esta

plataforma educativa están en SQLServer como motor de base de datos. Este es un software que requiere de licenciamiento, lo que dificultó el proceso de revisión ya que las versiones que se ofrecen para demostración del producto por la compañía propietaria tienen limitaciones en capacidad de procesamiento y almacenamiento en la carga de las bases de datos y debido a que estas bases cuentan con un peso aproximado de 25 GB, estas versiones de prueba no permiten su carga.

Se hizo un proceso de exploración detallado de estas bases de datos y para determinar los registros que realmente podrían aportar valor al modelo, esto llevo un buen tiempo y esfuerzo por la cantidad de tablas contenidas en el esquema de almacenamiento de Blackboard. En el apartado correspondiente al proceso de ETL se mostrará la consulta que extrae los datos considerados pertinentes para el modelo propuesto.

#### 5.2.2.2 Moodle

La base de datos suministrada del Moodle (Moodle Project, 2013) está compuesta por un total de 325 tablas, las cuales contienen información del segundo semestre del año 2014. Se logró identificar una existencia de 172.279 usuarios, entre los cuales hay estudiantes, docentes y administrativos. Después de una exploración detallada se identificaron un aproximado de 80 tablas con una cantidad de registros insuficientes para realizar un análisis relevante. Adicional a esto, al aplicar herramientas de apoyo para identificar la estructura de las tablas de esta plataforma, se identificó que no existen relaciones entre las tablas; sin embargo, cada tabla tiene un atributo identificador auto-incremental.

De forma general se logra identificar que la base de datos cuenta con un diseño enfocado a la operación funcional de la plataforma. Se resalta también la existencia de rangos jerárquicos modulares, donde una tabla almacena la información general según el tipo de acción o funcionamiento que se requiere para Moodle. Hay varias tablas carentes de datos o que presentan información paramétrica propia del funcionamiento de la plataforma.

Para efectos de las pruebas realizadas como parte de la validación del modelo, se utilizaron los datos del Blackboard, dado que estos presentaban mayor cantidad y se contaba con más de un periodo académico, mientras que para el caso de Moodle no se tenía sino un periodo y además se encontraron varias inconsistencias en el proceso de ETL, lo cual se está revisando en detalle con personal de la Dirección Académica de la

sede para identificar si las copias de respaldo suministradas fueron generadas erróneamente.

### 5.2.3 Diagnóstico del uso de plataformas virtuales de aprendizaje

Para obtener una referencia respecto al estado actual del uso de plataformas virtuales de aprendizaje en la Universidad del caso de estudio, se realizó una encuesta a estudiantes del plantel educativo y otra a docentes del mismo. Los instrumentos fueron inicialmente validados con un grupo de diez personas entre estudiantes de pregrado y posgrado y dos docentes para verificar la comprensión de las preguntas y validar la coherencia de sus respuestas. Una vez realizada esta validación, los instrumentos fueron refinados y llevados a un formato digital a través de la herramienta de formularios de Google. Una vez realizada su difusión a través de correo electrónico institucional se logró tener una muestra de 98 estudiantes y 25 docentes. En el anexo A se presentan los instrumentos aplicados con su respectiva tabulación.

***Conclusiones de la encuesta aplicada a los estudiantes***

Se encuentra que la mayoría de los estudiantes encuestados conoce y hace uso de las plataformas virtuales de aprendizaje tanto de sus programas académicos actuales como para otros procesos de formación externos a la universidad.

Para el caso del uso de las plataformas dentro de la Universidad, se concentra en actividades de descarga de contenidos y de realización de tareas que se suelen asociar a la entrega de documentos. Lo anterior deja ver que las plataformas son usadas principalmente como mecanismo de intercambio de archivos entre docentes y estudiantes.

Por lo tanto, permite vislumbrar que se está desaprovechando las ventajas de estas plataformas para la realización de una mayor variedad de actividades de aprendizaje que contribuyan a detectar patrones de comportamiento entre las formas de aprender de los estudiantes y su relación con el uso de estas plataformas.

Finalmente, los comentarios dejados por algunos de los estudiantes, muestran el desinterés que pueden llegar a tener estos, ya que no ven en las plataformas un valor agregado a sus procesos de formación dentro de la Universidad, sino que las perciben

como un medio de comunicación más. Además, el bajo uso de las plataformas puede estar asociado, como lo indica uno de los comentarios, en el poco uso de estas por parte de los docentes.

***Conclusiones de la encuesta aplicada a los docentes***

De las respuestas obtenidas del grupo de docentes se puede determinar que a pesar que la mayoría conoce que es una plataforma virtual de educación, las usan en menos proporción que los estudiantes.

Se corrobora con las respuestas de los docentes que las dos principales actividades realizadas son la carga de contenidos y materiales para los estudiantes y la creación de actividades o tareas relacionadas con la entrega de documentos.

Los docentes indican, en la mayoría de los casos, que realizan actualización de los cursos cada semestre de acuerdo a nuevos conocimientos adquiridos, desempeño y retroalimentaciones de los estudiantes.

Finalmente, es de rescatar de uno de los comentarios, que los docentes requieren ser capacitados para dar un mejor aprovechamiento a las diferentes actividades que ofrecen las plataformas virtuales para la creación y configuración de los cursos.

## 5.3 Componente 2: Proceso de ETL

Para el proceso de ETL sobre los datos del caso de estudio se desarrolló una herramienta en JAVA que está en la capacidad de tomar los datos tanto académicos como de las interacciones realizadas en la plataforma virtual. En la Figura 5-1 se muestra el esquema general del proceso de ETL.

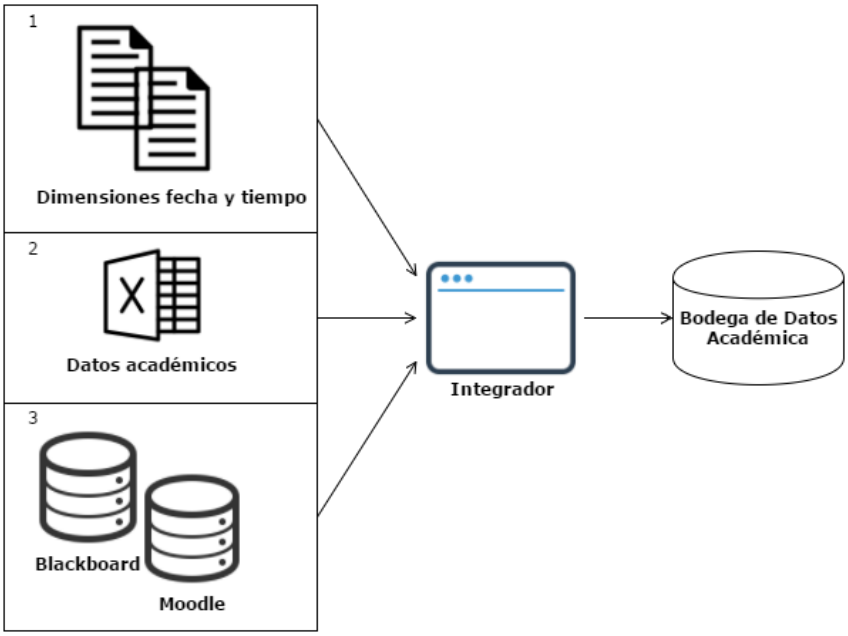



**Figura 5-1:** Proceso de ETL implementado

Para el caso de los datos académicos, estos son tomados de archivos en Excel, formato en el cual fueron suministrados por la dirección académica. Por su parte, los datos de las interacciones son tomados de las bases de datos de Blackboard y Moodle, estas bases de datos están en SQL Server y MySQL respectivamente. Para las dimensiones fecha y tiempo, que se contemplan en la bodega de datos y que no están dentro de los datos suministrados, se emplearon archivos planos creados previamente para otra bodega de datos construida en el Grupo de Investigación GAIA y que cumplen con los requerimientos que se tenían para este caso en particular (Duque Méndez, Hernández Leal, Pérez Zapata, Arroyave Tabares, & Espinosa Gómez, 2016).

La función de la aplicación que realiza la integración de los datos fue, en primera instancia, cargar los registros correspondientes a las dimensiones fecha y tiempo. Esto se realizó previo a la carga de los demás datos, ya que para realizar las relaciones dentro del esquema de la bodega con estas dimensiones es necesario tenerlas almacenadas. Esta carga de las dimensiones se hace tomando el archivo de texto plano y mapeando con los campos de la bodega.

En segunda instancia se hace el procesamiento de los datos académicos, en este caso se toma cada registro de los archivos de Excel y se mapean las columnas donde se encuentra

cada atributo hacia cada una de las tablas que componen el modelo de la bodega de datos académica. Una vez se toman estos atributos se realiza la verificación de los datos para evitar ingresar datos duplicados en el componente relacional del modelo de la bodega, además, se hacen algunas tareas de completado de datos, por ejemplo, para el caso de estudiantes que no reportan la ciudad de procedencia, este campo se completa con un valor designado dentro del sistema, en este caso “No informa”. Finalmente se hace la carga de los datos en la bodega.

El tercer paso corresponde al procesamiento de los datos de las interacciones, los cuales son tomados desde las bases de datos suministradas. En primer lugar, se hace una conexión JDBC (por sus siglas en inglés de Java Database Connectivity) y se extraen los datos consultando por cada uno de los estudiantes ya ingresados en la bodega, si el estudiante presenta registros de interacciones dentro de la plataforma estos se extraen mediante una consulta SQL que selecciona los atributos que se determinaron después del proceso de pre-procesamiento y caracterización inicial. Se analiza cada registro y se verifica si el curso donde el estudiante tuvo interacción está almacenado en la bodega y se crea o recupera el registro, según sea el caso; luego se consultan los identificadores de las dimensiones fecha y tiempo, se hace la relación y se guarda el registro de la interacción.

En la Figura 5-2 se presenta la interfaz de la herramienta para la carga de los datos a cada uno de los modelos de datos de la bodega.

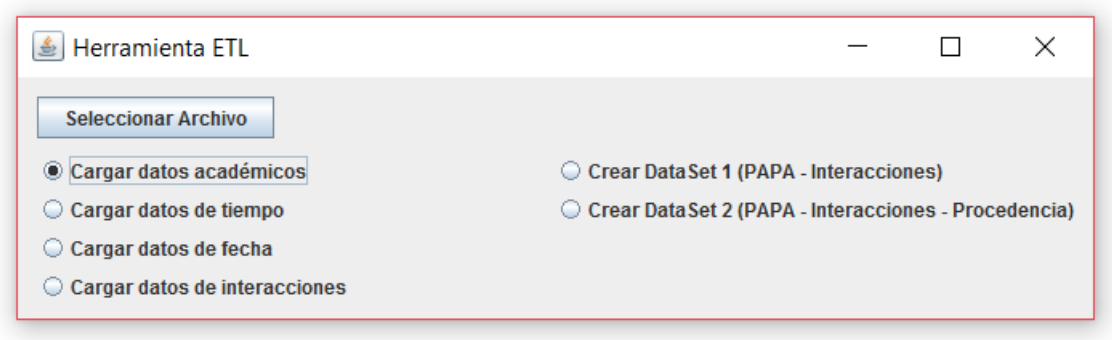



**Figura 5-2:** Interfaz de la herramienta de ETL desarrollada

La consulta utilizada en el proceso para cargar los datos de las interacciones es la siguiente:

```
SELECT
users.pk1 AS id_usuario_plataforma, users.user_id AS nombre_usuario,
users.firstname AS nombre,
users.lastname AS apellidos, users.student_id AS documento, users.email AS
correo,
users.last_login_date AS ultimo_ingreso,
ods_aa_session_activity.login_time AS hora_acceso_sesion,
ods_aa_session_activity.logout_time AS hora_salida_sesion,
ods_aa_session_activity.user_pk1 AS id_sesion_usuario,
ods_aa_course_activity.initial_datetime_access AS hora_acceso_curso,
ods_aa_course_activity.course_access_minutes AS duracion_acceso_curso,
ods_ls_course.pk1 AS id_curso, ods_ls_course.course_name AS nombre_curso
FROM users
INNER JOIN ods_aa_session_activity
ON ods_aa_session_activity.user_pk1=users.pk1
INNER JOIN ods_aa_course_activity
ON ods_aa_course_activity.session_pk1=ods_aa_session_activity.session_pk1
INNER JOIN ods_ls_course
ON ods_ls_course.pk1=ods_aa_course_activity.course_pk1
ORDER BY users.pk1
```

Una vez finalizado el proceso de ETL se logró tener una bodega de datos académica poblada con la cantidad de registros que se muestran en la Tabla 5-4.

**Tabla 5-4:** Número de registros cargados en la bodega de datos

| Tabla de Hechos | No. de registros |
|---|---|
| fact_academica | 56.904 |
| fact_personal | 56.904 |
| fact_familiar | 56.904 |
| fact_interaccion | 105.609 |

## 5.4 Componente 3: Bodega de datos académica

Este componente para el caso de estudio se ve reflejado en la Bodega de datos Académica que se ha construido y poblado en el motor de bases de datos PostgreSQL (Global Development Group, 2016). La Figura 5-3 presenta el modelo específico de la Bodega de Datos Académica, basado en un enfoque multidimensional en estrella ampliado, donde se puede ver la relación entre las dimensiones y las tablas de hechos.

Como se aprecia, el modelo de datos no cuenta con una única tabla de hechos, ya que el dominio de datos estudiado es bastante complejo y exigió la construcción de un modelo hibrido entre una bodega de datos relacional y una bodega multidimensional en estrella, la cual, según las necesidades identificadas en cuanto a la naturaleza de los datos, cubre las necesidades de almacenamiento y posterior recuperación para aplicar técnicas de análisis de datos que permitan encontrar información que ayude a los procesos de toma de decisiones dentro del plantel educativo.

Se ha considerado el diseño de una bodega de datos de manera hibrida con el fin de aprovechar las ventajas de los dos modos de almacenar los datos en una bodega de datos: el modelo relacional y el modelo multidimensional. Los modelos relacionales presentan beneficios como la integridad, compatibilidad, estandarización, fiabilidad, garantía de independencia y facilidad de conectividad con los lenguajes de programación, además hay grandes posibilidades para la realización de consultas (Jaramillo Valbuena & Londoño, 2015). Por su parte, el modelo multidimensional como técnica de diseño lógico de las bodegas de datos se caracteriza por presentar los datos de forma intuitiva y con alto desempeño (Duque Méndez, Vélez Upegui, & Orozco Alzate, 2015). Al contar con una solución que integra estas dos tipologías de almacenamiento, se puede tener beneficios de las características de cada una y permite atender las necesidades que presentan los diferentes tipos de datos y estructuras de las que provienen.

En el anexo A se ampliará el modelo con su diccionario de datos detallado.

**fact_informacion_academica**

En esta tabla se registran los datos del comportamiento académico del usuario, es decir, se registra la relación entre cada ingreso del estudiante y cada materia cursada y como se

va comportando el promedio académico, con base en esto es posible identificar tendencias y posibles problemas de fundamentación que conlleve a bajo rendimiento en ciertos cursos.

**fact_informacion_personal**

En esta tabla se registran los cambios de la información socioeconómica del usuario, como puede ser el tipo de residencia, la dirección, el estrato, el municipio, ingresos, etc., esto con el objetivo de identificar si estos cambios afectaron el rendimiento académico del usuario. Cruzando los datos del comportamiento académico con los datos registrados en esta tabla, se pueden tomar decisiones, pensado en realizar un acompañamiento para identificar los posibles problemas y tratar de encontrar soluciones.

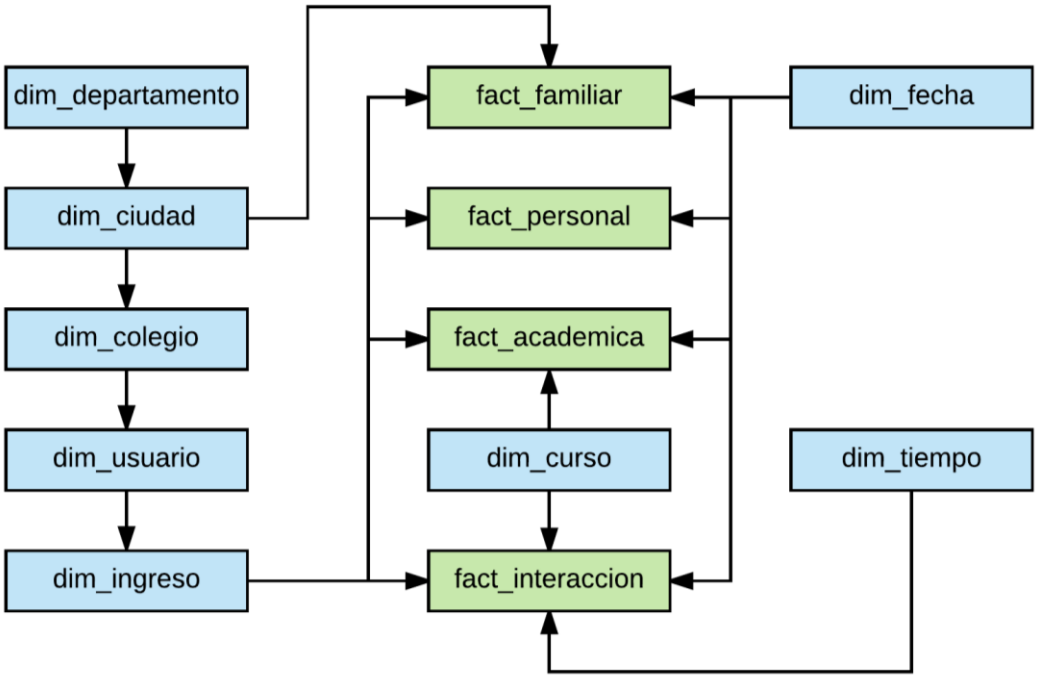



**Figura 5-3:** Modelo de datos de la bodega de datos académica

**fact_informacion_familiar**

En esta tabla se registran los cambios en la información socioeconómica familiar del usuario en aspectos como la situación económica familiar, tipo de vivienda familiar, número de integrantes de la familia y demás aspectos que pueden presentar cambios del entorno

familiar y que pueda afectar de alguna manera el comportamiento y por ende el rendimiento académico del usuario en la institución.

**fact_interaccion**

En esta tabla se registran las interacciones que el usuario realiza en cada curso con las plataformas de educación virtual disponibles en la institución, esto con el objetivo de determinar en qué medida la utilización de dichas plataformas incide en el rendimiento académico que el usuario puede presentar durante su proceso de formación. Los campos que se registran en esta tabla son el tipo de interacción que el usuario realizó en la plataforma, la duración de la interacción, el curso en el cual realizó la interacción, la hora de la interacción, estos datos pueden ayudar a identificar hábitos de estudio y que tan beneficiosos son para el usuario teniendo en cuenta los resultados obtenidos en las calificaciones.

**dim_departamento**

Esta es una dimensión de ubicación, donde se encuentran los datos de los departamentos del país, con atributos como el código y nombre del departamento.

**dim_ciudad**

En esta tabla se encuentran los datos de los municipios ubicados en el país, con lo cual se determina la ubicación de cada uno de los actores presentes en el sistema, esto es de gran relevancia para determinar tendencias en comportamientos en la región, esta tabla cuenta con atributos como, código de municipio, nombre del municipio y una relación con dim_departamento para determinar en qué departamento se encuentra el municipio.

**dim_colegio**

En esta tabla se guardan los datos de los colegios o instituciones de educación media de los cuales se tenga registro para relacionarlos con cada uno de los estudiantes que se ingresen en la bodega, lo cual puede ayudar a determinar de qué colegios son los estudiantes que tiene determinado rendimiento académico y con base en esto se pueden tomar decisiones para dar acompañamiento especial a estudiantes de determinadas instituciones y así mejorar la calidad académica, esta tabla tiene atributos como: código del colegio, nombre del colegio, jornada académica, modalidad, calendario y relación con la tabla municipio para indicar en qué lugar se encuentra ubicado.

**dim_usuario**

En esta tabla se guardan todos los datos básicos del usuario, los cuales no presentan cambios en el tiempo o no son de interés para el análisis académico, como pueden ser el nombre, el número de documento, el sexo, correo electrónico, fecha de nacimiento, estado civil, distrito militar donde obtuvo la libreta militar (en el caso de los varones), entre otros; además esta tabla tiene relación con dim_colegio para determinar de qué institución educativa proviene el estudiante.

**dim_ingreso**

En esta tabla se guardan datos acerca del ingreso del estudiante a la universidad, esto se guarda aparte, ya que presenta datos muy específicos como son el puntaje de ingreso, el tipo de ingreso del estudiante a la institución, entre otros. Además de esto, esta tabla puede presentar varios registros por estudiante, ya que, en algunas instituciones, los estudiantes pueden optar por modalidades de doble titulación, o puede presentar diferentes ingresos si este ha solicitado traslado a otro programa académico y por cada uno de estos los datos relacionados con los cursos y promedio académico pueden variar.

**dim_curso**

En esta tabla se guardan los datos de los cursos que van atendiendo los usuarios, estos datos son necesarios para registrar el comportamiento académico que van teniendo estos a través del tiempo y como se va afectando el promedio académico, con base en el cual se pueden buscar tendencias comportamentales.

**dim_tiempo**

En esta tabla se registran las fechas en las cuales van sucediendo los eventos de cada estudiante, posteriormente se usa la referencia entre esta tabla, la tabla dim_estudiante y las fact donde es necesario registrar cambios en el tiempo.

## 5.5 Componente 4: Aplicación de analíticas de aprendizaje

La aplicación de los análisis sobre los datos llevados al almacenamiento en la bodega de datos académica ha sido realizada según los cuatro enfoques de analíticas de aprendizaje definidos en el modelo propuesto. A continuación, se presentan los resultados en cada enfoque.

### 5.5.1 Enfoque descriptivo

Inicialmente se busca generar indicadores estadísticos de forma global de las diversas variables tratadas con el fin de establecer una identificación de tendencias de los datos, mediante la realización de cálculos se identificaron las medidas de tendencia central, medidas de dispersión y medidas de posición.

En la Tabla 5-5 se aprecian los indicadores estadísticos en el cálculo de las medidas de tendencia central, de dispersión y posición para las principales variables del modelo de datos académicos.

**Tabla 5-5:** Indicadores de estadística descriptiva para los datos académicos

| Atributos | Medidas de tendencia central | | | Medidas de posición | | | | |
|---|---|---|---|---|---|---|---|---|
| | Promedio | Mediana | Moda | Varianza | Desviación estándar | Cuartiles | | |
| | | | | | | 25% | 50% | 75% |
| PAPA | 3,65 | 3,70 | 3,70 | 0,21 | 0,45 | 3,40 | 3,70 | 3,90 |
| Créditos Inscritos | 16,31 | 17,00 | 18,00 | 28,54 | 5,34 | 14,00 | 17,00 | 20,00 |
| Créditos Homologados | 6,76 | 0,00 | 0,00 | 275,54 | 16,60 | 0,00 | 0,00 | 6,00 |
| Créditos Adicionales | 7,51 | 9,00 | 12,00 | 22,73 | 4,77 | 3,00 | 9,00 | 12,00 |
| Créditos Aprobados | 119,04 | 138,00 | 168,00 | 2.961,97 | 54,42 | 75,00 | 138,00 | 165,00 |
| Cantidad de Asignaturas Inscritas | 4,88 | 5,00 | 5,00 | 2,96 | 1,72 | 4,00 | 5,00 | 6,00 |
| Cantidad de Asignaturas Homologadas | 2,16 | 0,00 | 0,00 | 26,02 | 5,10 | 0,00 | 0,00 | 2,00 |
| Cantidad de Asignaturas Aprobados | 38,05 | 44,00 | 52,00 | 288,08 | 16,97 | 24,00 | 44,00 | 52,00 |
| PBM Periodo | 34,93 | 32,00 | 30,00 | 292,71 | 17,11 | 21,00 | 32,00 | 44,00 |
| PBM Permanente | 30,35 | 27,00 | 20,00 | 210,30 | 14,50 | 20,00 | 27,00 | 37,00 |
| PBM Colegio | 44,74 | 43,00 | 42,00 | 107,87 | 10,39 | 39,00 | 43,00 | 51,00 |
| PBM Calculado | 25,89 | 17,00 | 12,00 | 458,89 | 21,42 | 10,00 | 17,00 | 40,00 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Cantidad Hermanos UNAL | 0,15 | 0,00 | 0,00 | 0,29 | 0,54 | 0,00 | 0,00 | 0,00 |
| Cantidad Hermanos | 2,37 | 3,00 | 3,00 | 2,05 | 1,43 | 1,00 | 3,00 | 3,00 |
| Puesto de Admisión | 27,66 | 26,00 | 0,00 | 440,24 | 20,98 | 12,00 | 26,00 | 41,00 |

Se puede apreciar al analizar las medidas de dispersión, que el PAPA (promedio aritmético ponderado acumulado) presenta una varianza y una desviación estándar pequeñas, indicando una “estabilidad” en los datos. El número de créditos y de asignaturas inscritas poseen una mayor dispersión de los datos, es decir que hay una alta probabilidad de encontrar numerosos datos atípicos respecto a las medidas de tendencia central.

En cuanto a las medidas de posición: El 25% de los estudiantes obtuvo un PAPA menor o igual a 3.4, también hay un 25% de estudiantes que posee un PAPA entre 3.4 y el promedio que es 3.7, lo cual es igual que decir que el 50% de estudiantes tiene un PAPA inferior o igual al 3.7, pero también hay un 25% de estudiantes que tiene un PAPA entre el promedio y un puntaje de 3.9, es decir que el 75% de los estudiantes en la muestra tienen un PAPA igual o menor de 3.9, y hay un 25% restante el cual posee una nota superior de 3.9 puntos. Teniendo en cuenta que se tuvo un total de 10778 estudiantes se puede deducir entonces que aproximadamente 2694 estudiantes tienen un promedio aritmético acumulado superior a 3.9 puntos. Pero al realizar una revisión general de los cuartiles, se logra identificar una linealidad entre el Q1 y Q2 para los créditos y cantidad de asignaturas homologadas, en estas variables se aprecia que para el 75% de los estudiantes, es decir que por cada 8.08 estudiantes se homologan seis créditos, equivalentes aproximadamente a dos asignaturas.

En la Tabla 5-6 se aprecian los indicadores estadísticos en el cálculo de las medidas de tendencia central y de posición para las principales variables del modelo de datos de las interacciones.

**Tabla 5-6:** Indicadores de estadística descriptiva para las interacciones

| Atributos | Medidas de tendencia central | | | Medidas de posición | |
|---|---|---|---|---|---|
| | Mediana | Moda | Promedio | Varianza | Desviación estándar |
| Duración en horas de las interacciones | 0,00 | 0,00 | 1,48 | 3,14 | 1,77 |
| Duración en minutos de las interacciones | 12,00 | 0,00 | 18,43 | 353,18 | 18,79 |

| | | | | | |
|---|---|---|---|---|---|
| Duración en segundos de las interacciones | 26,00 | 2,00 | 27,20 | 301,71 | 17,37 |
| Duración en total de las interacciones en minutos | 56,62 | 0,03 | 107,86 | 13494,06 | 116,16 |
| Número de interacciones por estudiante | 17,00 | 1,00 | 28,74 | 1103,58 | 33,22 |
| Número de cursos por estudiante | 2,00 | 1,00 | 3,15 | 5,60 | 2,37 |

Se puede apreciar que los datos de las interacciones son bastantes variables, se presenta una varianza alta en cuanto al número de interacciones por estudiante e igualmente en la duración total en minutos de las mismas, esto deja en evidencia que existen estudiantes con mucha más interacción en la plataforma que otros. Para el número total de cursos por estudiante se presenta la misma situación. Lo anterior muestra que posiblemente pueden existir datos atípicos que se han saltado el proceso de ETL ya que es difícil determinar casos como, por ejemplo, un estudiante que ha dejado su sesión en la plataforma abierta, pero que no está realizando como tal una acción concreta.

## 5.5.2 Enfoque diagnóstico

En este enfoque se llevó a cabo un análisis por medio de consultas relacionales y un análisis multidimensional. A continuación, se traen a colación los resultados obtenidos en estos dos pasos.

### 5.5.2.1 Consultas relacionales

Como primera instancia se ejecutaron una serie de consultas relacionales sobre los datos almacenados en la bodega de datos académica, con el fin de tener información acerca del contenido de cada una de las tablas del modelo y su relación con las demás, con lo cual se obtuvo conocimiento acerca del comportamiento de los datos y como estos se relacionan entre sí. A continuación, se presentan algunas de las consultas SQL ejecutadas y sus respectivas gráficas generadas en Excel a partir de los resultados de dichas consultas:

***Consulta:*** Estudiantes por carrera entre los periodos académicos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015. Los resultados se muestran en la Tabla 5-7 y Figura 5-4. Se tiene que la mayor cantidad de estudiantes se encuentra en el

programa Administración de Empresas, el cual cuenta con 2188 estudiantes, además el programa con menor cantidad de estudiantes es Construcción y esto se debe a que éste ya no es ofertado por la universidad.

```
SELECT COUNT(dim_usuario.id), dim_ingreso.programa
FROM dim_usuario, dim_ingreso
WHERE dim_usuario.id=dim_ingreso.id
GROUP BY dim_ingreso.programa
```

**Tabla 5-7:** Distribución de estudiantes por programa académico

| N° estudiantes | Programa |
|---:|---|
| 1058 | Ingeniería Industrial |
| 898 | Administración De Sistemas Informáticos |
| 1008 | Arquitectura |
| 1020 | Ingeniería Electrónica |
| 2188 | Administración De Empresas |
| 319 | Matemáticas |
| 359 | Gestión Cultural Y Comunicativa |
| 1191 | Ingeniería Civil |
| 539 | Ingeniería Física |
| 1193 | Ingeniería Eléctrica |
| 991 | Ingeniería Química |
| 14 | Construcción |

**Figura 5-4:** Distribución de estudiantes por programa académico

***Consulta:*** Estudiantes por género entre los periodos académicos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015. Los resultados se muestran en la Tabla 5-8 y en la Figura 5-5. Se aprecia que la proporción de estudiantes de género masculino frente a los estudiantes de género femenino es casi 2 a 1.

*SELECT COUNT(dim_usuario.id), dim_usuario.sexo*
*FROM dim_usuario*
*GROUP BY dim_usuario.sexo*

**Tabla 5-8:** Distribución de estudiantes por género

| N° estudiantes | Género |
|---:|---|
| 7147 | M |
| 3631 | F |

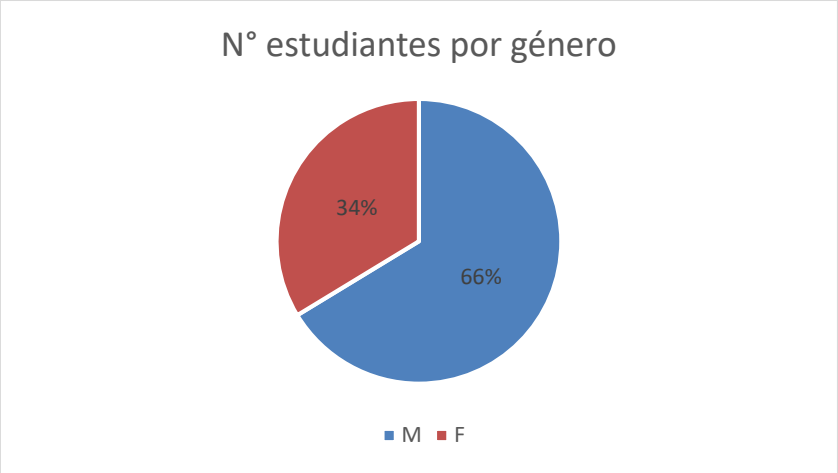



**Figura 5-5:** Distribución de estudiantes por género

***Consulta:*** Estudiantes por estrato entre los periodos académicos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015. Los resultados se muestran en la Tabla 5-9 y en la Figura 5-6. La mayoría de estudiantes pertenecen a los estratos 1, 2 y 3, los habitantes de estos estratos son los de menores ingresos en el país.

```
SELECT COUNT(dim_usuario.id), informacion_familiar_2.estrato FROM dim_usuario INNER JOIN
(SELECT fact_familiar.estrato, informacion_familiar_1.id_usuario, informacion_familiar_1.periodo FROM fact_familiar
INNER JOIN dim_ingreso
ON fact_familiar.id_ingreso=dim_ingreso.id
INNER JOIN (SELECT dim_ingreso.id_usuario, MAX(fact_familiar.periodo) AS periodo FROM fact_familiar, dim_ingreso
WHERE fact_familiar.id_ingreso=dim_ingreso.id
GROUP BY dim_ingreso.id_usuario
ORDER BY dim_ingreso.id_usuario) informacion_familiar_1
ON dim_ingreso.id_usuario=informacion_familiar_1.id_usuario AND fact_familiar.periodo=informacion_familiar_1.periodo) informacion_familiar_2
ON dim_usuario.id=informacion_familiar_2.id_usuario
GROUP BY informacion_familiar_2.estrato
ORDER BY informacion_familiar_2.estrato
```

**Tabla 5-9:** Distribución de estudiantes por estrato

| N° estudiantes | Estrato |
|---|---|
| 27 | - |
| 1265 | Estrato 1 |
| 3385 | Estrato 2 |
| 4250 | Estrato 3 |
| 1170 | Estrato 4 |
| 282 | Estrato 5 |
| 251 | Estrato 6 |
| 22 | No estratificada |
| 126 | No informa |

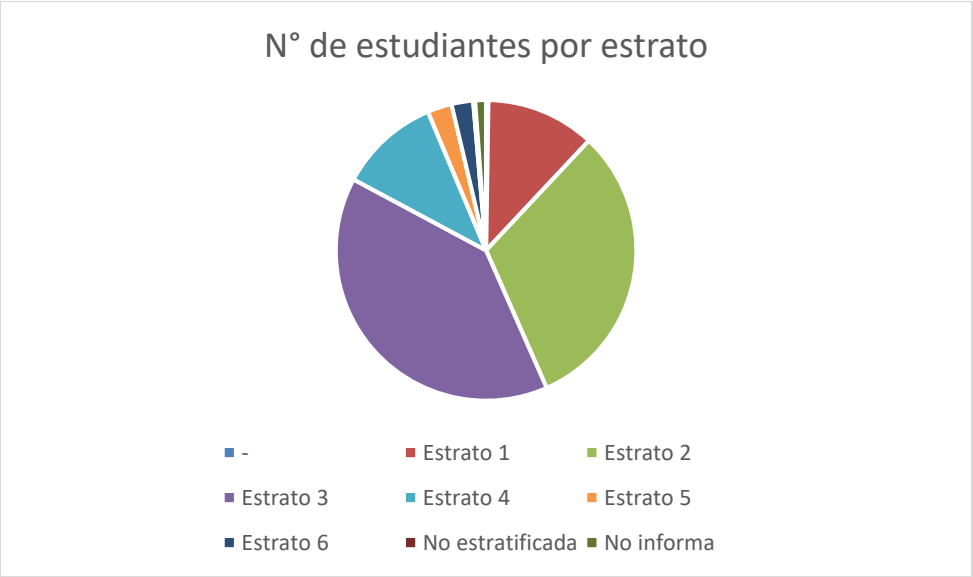



**Figura 5-6:** Distribución de estudiantes por estrato

***Consulta:*** Estudiantes por promedio por periodo para los periodos académicos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015.

```
SELECT COUNT(dim_ingreso.id_usuario), fact_academica.periodo,
fact_academica.papa
FROM fact_academica
INNER JOIN dim_ingreso
ON fact_academica.id_ingreso=dim_ingreso.id
```

*GROUP BY fact_academica.periodo, fact_academica.papa*
*ORDER BY fact_academica.periodo*

La totalidad de los resultados no son presentados por el gran tamaño que tienen. Sin embargo, como un balance general de la consulta, se tiene en la Tabla 5-10 la distribución de los estudiantes por rango de PAPA para el último semestre registrado (2015-I). De acuerdo a los resultados, se puede ver que la mayor cantidad de estudiantes, tiene su PAPA entre 3.0 y 4.0 estando concentrada la mayoría con promedio superior a 3.6.

**Tabla 5-10:** Número de estudiantes por rango de PAPA para el 2015-I

| **Rango PAPA** | [0 - 3.0] | [3.1 - 3.5] | [3.6 - 4.0] | [4.1 - 4.5] | [4.6 - 5.0] |
|---|---|---|---|---|---|
| **No. estudiantes** | 299 | 1522 | 2095 | 859 | 68 |

***Consulta:*** Estudiantes por promedio, por programa y por periodo para los periodos académicos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015.

*SELECT informacion_academica.periodo, dim_ingreso.programa,*
*informacion_academica.papa, COUNT(informacion_academica.numero_usuarios)*
*FROM dim_ingreso INNER JOIN*
*(SELECT COUNT(dim_ingreso.id_usuario) AS numero_usuarios,*
*fact_academica.papa, fact_academica.periodo, fact_academica.id_ingreso*
*FROM fact_academica*
*INNER JOIN dim_ingreso*
*ON fact_academica.id_ingreso=dim_ingreso.id*
*GROUP BY fact_academica.periodo, fact_academica.papa,*
*fact_academica.id_ingreso*
*ORDER BY fact_academica.periodo) informacion_academica*
*ON dim_ingreso.id=informacion_academica.id_ingreso*
*GROUP BY dim_ingreso.programa, informacion_academica.papa,*
*informacion_academica.periodo*
*ORDER BY informacion_academica.papa, informacion_academica.periodo*

La totalidad de los resultados no son presentados por el gran tamaño que tienen. En la Tabla 5-11 se muestra un consolidado de los promedios por programa académico para el último semestre registrado (2015-I), el cual da un estimado general del comportamiento de estos datos. De acuerdo a los resultados, se puede ver que los programas de Ingeniería Civil e Ingeniería Industrial tienen la mayor cantidad de estudiantes con promedios superiores a 4.0, además, en los programas Administración de Sistemas Informáticos y Matemáticas, una mayor cantidad de estudiantes tienen promedios entre 3.1 y 3.5.

**Tabla 5-11:** Clasificación de estudiantes por promedio y carrera para el 2015-I

| **Rango PAPA** / **Programa** | **[0 - 3.0]** | **[3.1 - 3.5]** | **[3.6 - 4.0]** | **[4.1 - 4.5]** | **[4.6 - 5.0]** |
|---|---|---|---|---|---|
| Administración de Empresas | 59 | 252 | 474 | 127 | 4 |
| Administración de Sistemas Informáticos | 34 | 189 | 118 | 28 | - |
| Arquitectura | 12 | 73 | 251 | 129 | - |
| Gestión Cultural y Comunicativa | 11 | 37 | 68 | 39 | 3 |
| Ingeniería Civil | 14 | 194 | 248 | 140 | 13 |
| Ingeniería Eléctrica | 29 | 228 | 245 | 62 | 5 |
| Ingeniería Electrónica | 43 | 158 | 174 | 82 | 8 |
| Ingeniería Física | 31 | 66 | 95 | 42 | 9 |
| Ingeniería Industrial | 7 | 110 | 169 | 133 | 13 |
| Ingeniería Química | 35 | 185 | 198 | 65 | 8 |
| Matemáticas | 24 | 30 | 21 | 12 | 5 |

***Consulta:*** Estudiantes ingresados por periodo (tomando último ingreso registrado) entre los periodos académicos comprendidos entre el primer semestre de 2009 y el primer semestre de 2015.

```
SELECT COUNT(dim_ingreso.id), dim_ingreso.periodo_inicio_estudio FROM
dim_ingreso INNER JOIN
(SELECT dim_ingreso.id_usuario, MAX(dim_ingreso.id) AS id_ingreso
FROM dim_ingreso
```

*GROUP BY dim_ingreso.id_usuario) ingreso*
*ON dim_ingreso.id=ingreso.id_ingreso*
*GROUP BY dim_ingreso.periodo_inicio_estudio*
*ORDER BY dim_ingreso.periodo_inicio_estudio*

Los resultados de esta consulta mostraron que se tienen registros de ingresos desde el segundo semestre de 1984 hasta el primer semestre de 2015. A continuación, se presentan en la Tabla 5-12 los ingresos de los últimos 10 años registrados. De acuerdo a los resultados, se puede ver que la tendencia en la universidad ha sido un aumento en el número de estudiantes admitidos en cada semestre, además se puede ver que en general ingresan más estudiantes en el primer periodo de cada año que en el segundo periodo. Lo primero debido a que la universidad ha ido ampliando su cobertura y lo segundo responde a los calendarios del sistema educativo colombiano, donde la mayor cantidad de estudiantes de colegio se gradúan terminando el año para ingresar en el primer semestre del año siguiente.

**Tabla 5-12:** Ingreso de estudiantes en los últimos 10 años

| No. estudiantes | Período académico |
|---|---|
| 326 | 2005-01 |
| 283 | 2005-03 |
| 344 | 2006-01 |
| 325 | 2006-03 |
| 369 | 2007-01 |
| 277 | 2007-03 |
| 337 | 2008-01 |
| 313 | 2008-03 |
| 460 | 2009-01 |
| 368 | 2009-03 |
| 622 | 2010-01 |
| 474 | 2010-03 |
| 648 | 2011-01 |
| 511 | 2011-03 |
| 643 | 2012-01 |
| 562 | 2012-03 |
| 679 | 2013-01 |
| 513 | 2013-03 |

| No. estudiantes | Período académico |
|---|---|
| 475 | 2014-01 |
| 579 | 2014-03 |
| 668 | 2015-01 |

***Consulta:*** Registros de interacciones por estudiante para los periodos académicos comprendidos entre el segundo semestre de 2013 y el primer semestre de 2014.

```
SELECT dim_usuario.id, COUNT(fact_interaccion.id_interaccion) FROM
dim_ingreso
INNER JOIN dim_usuario
ON dim_usuario.id=dim_ingreso.id_usuario
INNER JOIN fact_interaccion
ON fact_interaccion.id_ingreso=dim_ingreso.id
GROUP BY dim_usuario.id
```

En la Tabla 5-13 se han clasificado los resultados de la consulta en rangos. Se puede apreciar que la mayoría de los estudiantes tiene menos de 100 interacciones registradas para los periodos académicos comprendidos entre el segundo semestre de 2013 y el primer semestre de 2014. De acuerdo a los resultados, se puede ver que la mayor cantidad de estudiantes, han realizado entre 0 y 100 interacciones con las plataformas de educación virtual, lo cual evidencia un bajo uso de éstas durante los procesos educativos que llevan a cabo.

**Tabla 5-13:** Clasificación de acuerdo al número de interacciones

| Rango No. Interacciones | [0 - 100] | [101 - 200] | [201 - 300] | [301 o más |
|---|---|---|---|---|
| **Estudiantes** | 3.509 | 157 | 8 | 1 |

***Consulta:*** Número de estudiantes por cantidad de cursos inscritos en la plataforma Blackboard para los periodos académicos comprendidos entre el segundo semestre de 2013 y el primer semestre de 2014. Los resultados se muestran en la Tabla 5-14. De acuerdo a los resultados, se ve que la mayor cantidad de estudiantes que presentan interacciones con las plataformas de educación virtual, tienen menos de 4 cursos con

presencia en dichas plataformas, lo cual es un número bajo, teniendo en cuenta que se tienen datos de 2 semestres y que en promedio un estudiante inscribe 6 cursos por semestre, por lo tanto solo entre 1 y 4 de los 12 cursos potenciales que inscriben estos tienen presencia en las plataformas virtuales de educación suministradas por la universidad.

```
SELECT COUNT(fact_interaccion_2.id_usuario) AS num_estudiantes,
num_cursos FROM
(SELECT COUNT(fact_interaccion_1.id_curso) AS num_cursos, dim_usuario.id
AS id_usuario FROM dim_usuario
INNER JOIN
(SELECT dim_usuario.id AS id_usuario, COUNT(fact_interaccion.id_interaccion)
AS num_interacciones, fact_interaccion.id_curso FROM dim_ingreso
INNER JOIN dim_usuario
ON dim_usuario.id=dim_ingreso.id_usuario
INNER JOIN fact_interaccion
ON fact_interaccion.id_ingreso=dim_ingreso.id
GROUP BY dim_usuario.id, fact_interaccion.id_curso) fact_interaccion_1
ON dim_usuario.id=fact_interaccion_1.id_usuario
GROUP BY dim_usuario.id) fact_interaccion_2
GROUP BY num_cursos
```

**Tabla 5-14:** Número de estudiantes por número de cursos

| N° cursos | N° de estudiantes |
|---|---|
| 1 | 1164 |
| 2 | 765 |
| 3 | 484 |
| 4 | 328 |
| 5 | 328 |
| 6 | 234 |
| 7 | 161 |
| 8 | 99 |
| 9 | 41 |
| 10 | 29 |
| 11 | 22 |

| N° cursos | N° de estudiantes |
|---|---|
| 12 | 9 |
| 13 | 3 |
| 14 | 4 |
| 15 | 3 |
| 17 | 1 |

Las consultas realizadas como primera parte del análisis permiten tener una visión general de algunos aspectos relevantes de los datos que seguirán siendo objeto de estudio. A continuación, se traen los resultados del análisis multidimensional.

#### 5.5.2.2 Análisis multidimensional

En segunda instancia se hizo un análisis multidimensional, para ello se empleó una herramienta construida en el Grupo de Investigación en Ambientes Inteligentes Adaptativos – GAIA como resultado de un trabajo de grado. La herramienta es llamada GAIA-OLAP y se puede acceder a ella a través del enlace: froac.manizales.unal.edu.co/gaia_olap

A continuación, se muestran algunas de las consultas OLAP realizadas en la herramienta, las cuales muestran el comportamiento de los datos almacenados en las tablas de hechos “fact_interaccion” y “fact_academica” con relación a algunas dimensiones. La herramienta usada para este análisis multidimensional presenta por medio de gráficas, los resultados de las consultas construidas.

***Consulta:*** Nivel de agregación promedio de duración en horas de las interacciones teniendo en cuenta la dimensión fecha con un nivel de profundización mes. (Tipo de operación: Drill Down). En la Figura 5-7 se muestra gráficamente los resultados de la consulta. Según los resultados representados en la gráfica, se puede ver que durante los primeros meses de cada semestre (febrero para el primer semestre del año y agosto para el segundo semestre del año) se suele hacer un bajo uso de las plataformas virtuales de educación y que este empieza a aumentar en el segundo mes de cada semestre.

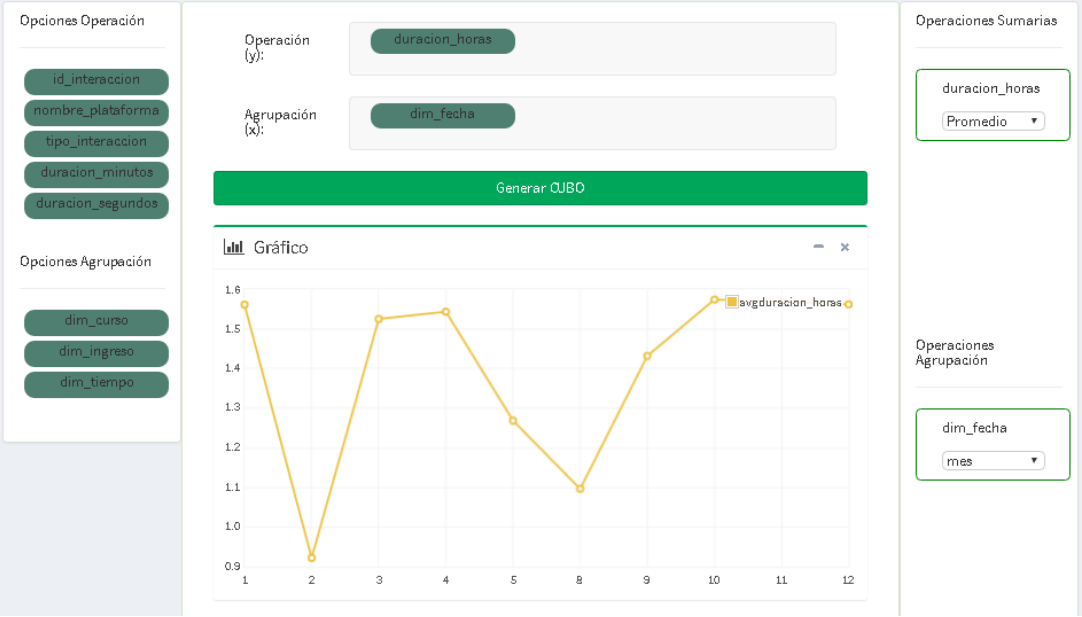



**Figura 5-7:** Promedio de duración en horas de las interacciones por mes

***Consulta:*** Nivel de agregación promedio de duración en horas, minutos y segundos de las interacciones teniendo en cuenta la dimensión fecha con un nivel de profundización mes. (Tipo de operación: Drill Down). La Figura 5-8 muestra gráficamente los resultados de la consulta.

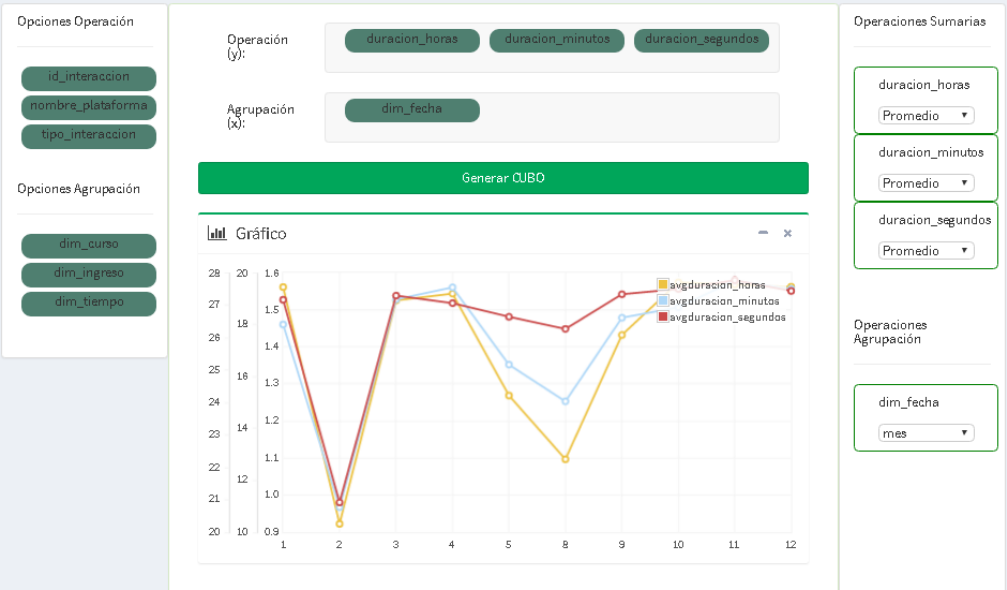



**Figura 5-8:** Promedio de duración en horas, minutos y segundos de las interacciones por mes

***Consulta:*** Nivel de agregación contar interacciones teniendo en cuenta la dimensión ingreso con un nivel de profundización “plan de estudios”. (Tipo de operación: Drill Down) En la Figura 5-9 se muestra gráficamente los resultados de la consulta. Como el campo plan de estudios corresponde a un código, en la Tabla 5-15 se presenta su correspondencia con el nombre de cada uno de los programas académicos. Según los resultados representados en la gráfica, se puede ver que los planes de estudio con mayor cantidad de interacciones tienen código 4024 y 4029, que corresponden a Ingeniería Industrial y Administración de Sistemas Informáticos, este último es el plan de la universidad que tiene relación directa con la rama de la computación. Además, se puede ver que los planes con código 4025, 4032 y 4033, los cuales corresponden a Ingeniería Civil, Matemáticas y Gestión Cultural, son los que presentan menor cantidad de interacciones, para los últimos dos, esto se puede explicar por la baja cantidad de estudiantes, 319 y 359 respectivamente, en comparación con los 1119 estudiantes de Ingeniería Civil.

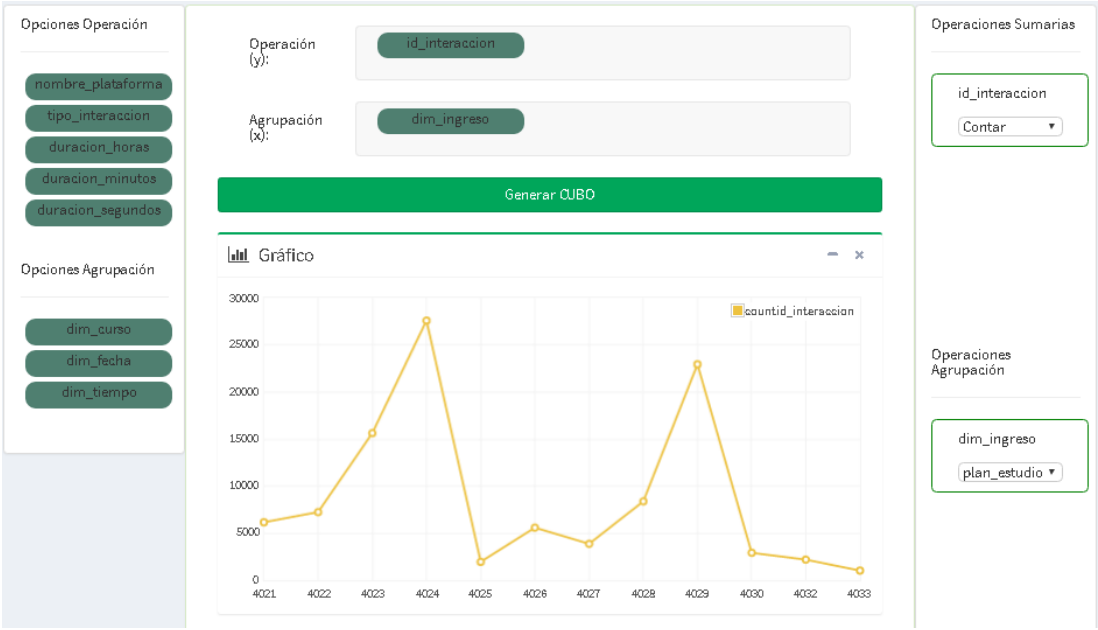



**Figura 5-9:** Cantidad de interacciones por plan de estudios

**Tabla 5-15:** Relación plan de estudios y nombres de programa

| Plan de estudios | Programa |
|---|---|
| 4021 | INGENIERIA CIVIL |
| 4022 | INGENIERIA ELECTRICA |

| Plan de estudios | Programa |
|---|---|
| 4023 | INGENIERIA QUIMICA |
| 4024 | INGENIERIA INDUSTRIAL |
| 4025 | ARQUITECTURA |
| 4026 | ADMINISTRACION DE EMPRESAS |
| 4027 | ADMINISTRACION DE EMPRESAS |
| 4028 | INGENIERIA ELECTRONICA |
| 4029 | ADMINISTRACION DE SISTEMAS INFORMATICOS |
| 4030 | INGENIERIA FISICA |
| 4031 | CONSTRUCCION |
| 4032 | MATEMATICAS |
| 4033 | GESTION CULTURAL Y COMUNICATIVA |

***Consulta:*** Nivel de agregación contar interacciones teniendo en cuenta la dimensión ingreso con un nivel de profundización “facultad” con base a la construcción anterior. (Tipo de operación: Drill Up). La Figura 5-10 contiene los resultados de la consulta de forma gráfica. Según los resultados representados en la gráfica, se puede ver que la facultad que mayor cantidad de interacciones presenta es la Facultad de Ingeniería y Arquitectura y la que menos interacciones presenta es la Facultad de Ciencias Exactas y Naturales, esto se debe a que la primera de ellas agrupa la mayor cantidad de programas académicos de la sede con seis carreras, mientras que las segunda facultad mencionada solo agrupa dos programas académicos los cuales además tienen una baja cantidad de estudiantes.

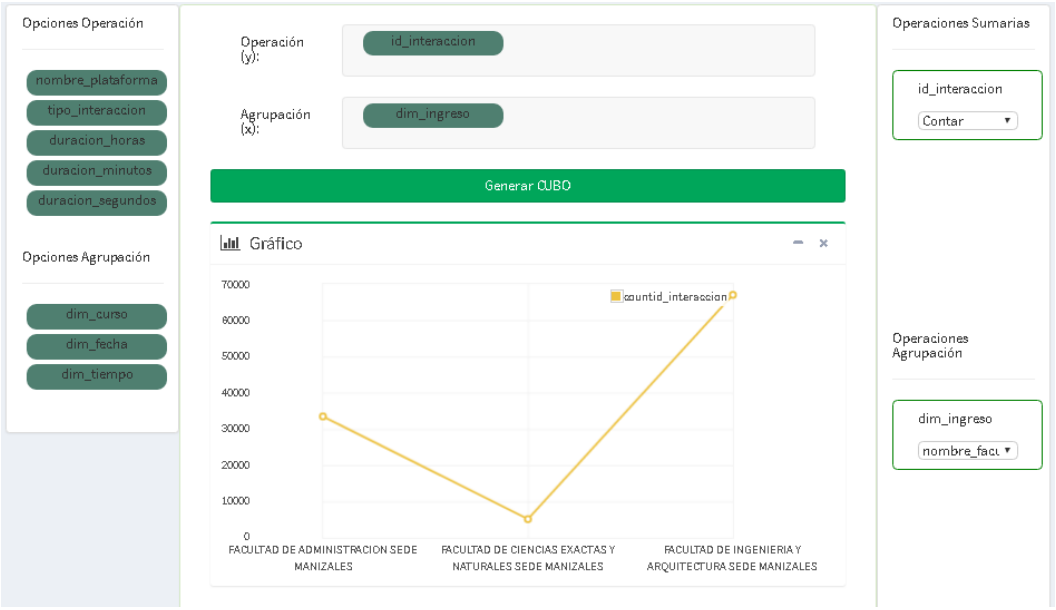



**Figura 5-10:** Cantidad de interacciones por Facultad

**Tabla 5-16:** Distribución de planes por Facultad

| Facultad | Programa |
|---|---|
| FACULTAD DE ADMINISTRACION SEDE MANIZALES | ADMINISTRACION DE EMPRESAS |
| | GESTION CULTURAL Y COMUNICATIVA |
| | ADMINISTRACION DE SISTEMAS INFORMATICOS |
| FACULTAD DE CIENCIAS EXACTAS Y NATURALES SEDE MANIZALES | INGENIERIA FISICA |
| | MATEMATICAS |
| FACULTAD DE INGENIERIA Y ARQUITECTURA SEDE MANIZALES | INGENIERIA CIVIL |
| | INGENIERIA ELECTRICA |
| | INGENIERIA QUIMICA |
| | INGENIERIA INDUSTRIAL |
| | ARQUITECTURA |
| | INGENIERIA ELECTRONICA |

***Consulta:*** Nivel de agregación promedio para el PAPA teniendo en cuenta la dimensión ingreso con un nivel de profundización “plan de estudio”. (Tipo de operación: Drill Down). En la Figura 5-11 se muestra gráficamente los resultados de la consulta. Como el campo plan de estudios corresponde a un código, se remite nuevamente a la Tabla 5-15 donde se tiene la correspondencia con el nombre de los programas. Según los resultados representados en la gráfica, se puede observar que el plan de estudios con mayor promedio académico es el 4024, el cual corresponde a Ingeniería Industrial, por otra parte, el plan de estudios 4032 correspondiente a Matemáticas, presenta el promedio más bajo.

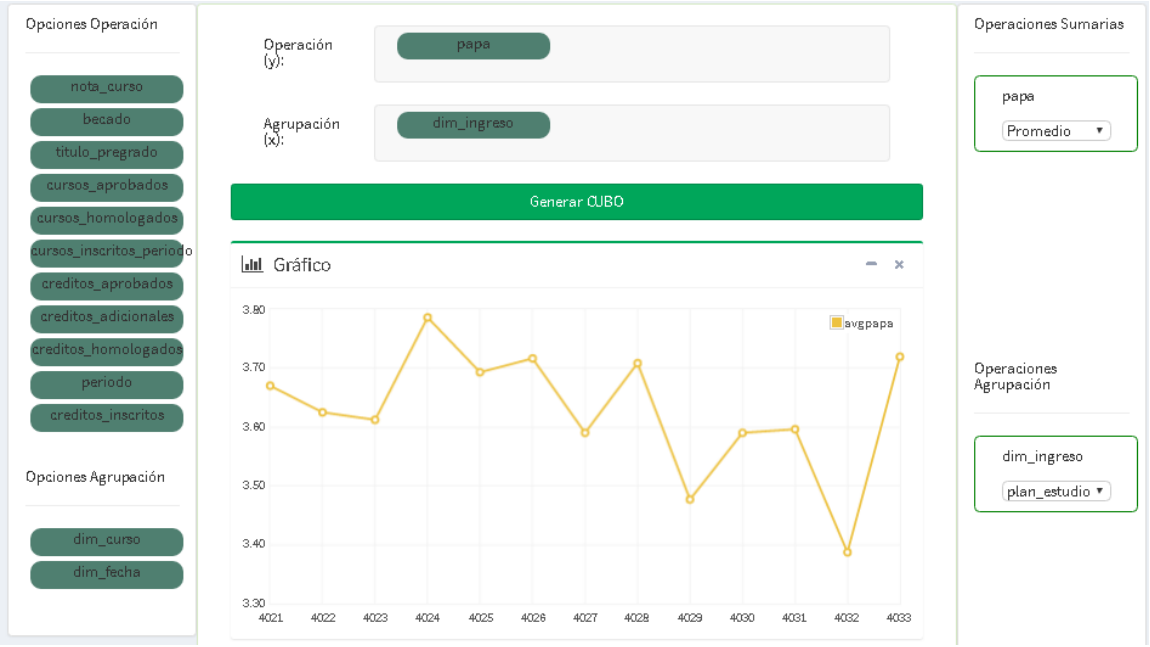



**Figura 5-11:** Promedio de PAPA por plan de estudios

***Consulta:*** Nivel de agregación promedio para el PAPA teniendo en cuenta la dimensión ingreso con un nivel de profundización “facultad” con base a la construcción anterior. (Tipo de operación: Drill Up). En la Figura 5-12 se muestra la gráfica arrojada por la consulta. Según los resultados representados en la gráfica, se puede observar que la Facultad de Ingeniería y Arquitectura presenta el mayor promedio académico y que la Facultad de Ciencias Exactas y Naturales, presenta el menor promedio académico de las facultades estudiadas.

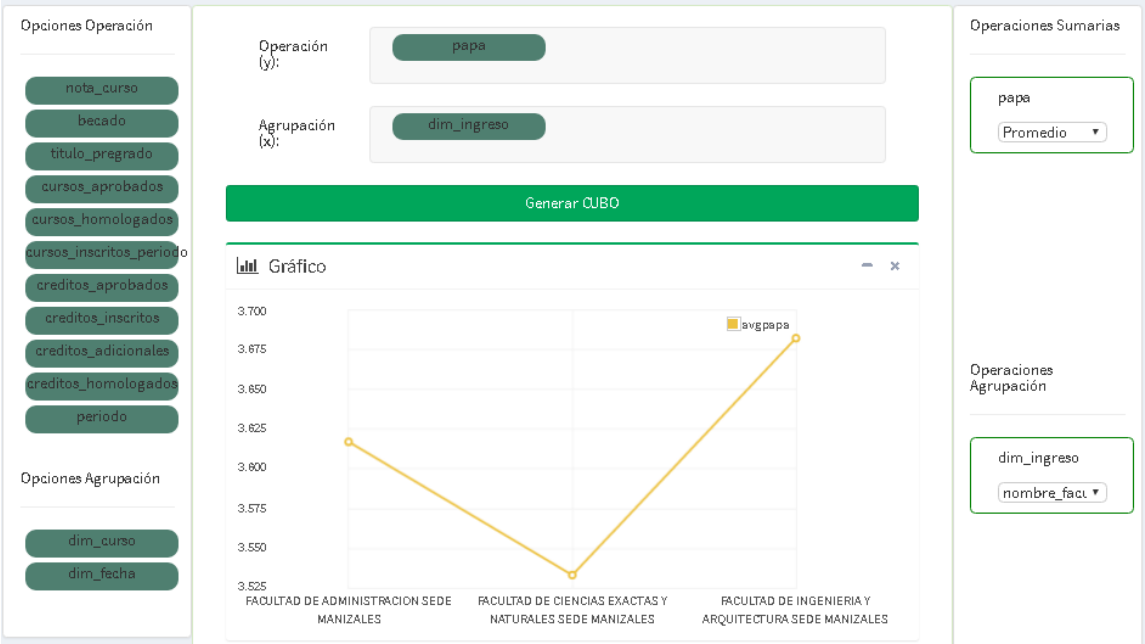



**Figura 5-12:** Promedio del PAPA por Facultad

***Consulta:*** Nivel de agregación promedio para el PAPA teniendo en cuenta la dimensión fecha con un nivel de profundización “año” con base a la construcción anterior. (Tipo de operación: Drill Anywhere). Esto se evidencia en la Figura 5-13. Según los resultados representados en la gráfica, se puede concluir que en general el promedio de los estudiantes ha ido en ascenso desde el 2010 al 2015, sin embargo, en 2015 se observa una disminución de la tendencia, pero esto se puede deber a que solo se tienen datos del primer semestre de ese año.

**Figura 5-13:** Promedio de PAPA por año

***Consulta:*** Nivel de agregación promedio para el PAPA teniendo en cuenta la dimensión fecha con un nivel de profundización “semestre”. (Tipo de operación: Drill Down). La consulta con sus resultados gráficos se muestra en la Figura 5-14. Según los resultados representados en la gráfica, se puede concluir que el promedio académico tiende a disminuir en el segundo semestre de cada año, esto se puede deberse a que en este periodo suelen ingresar estudiantes que en el primer periodo no pudieron ingresar a la universidad, ya que el puntaje de admisión no fue suficiente.

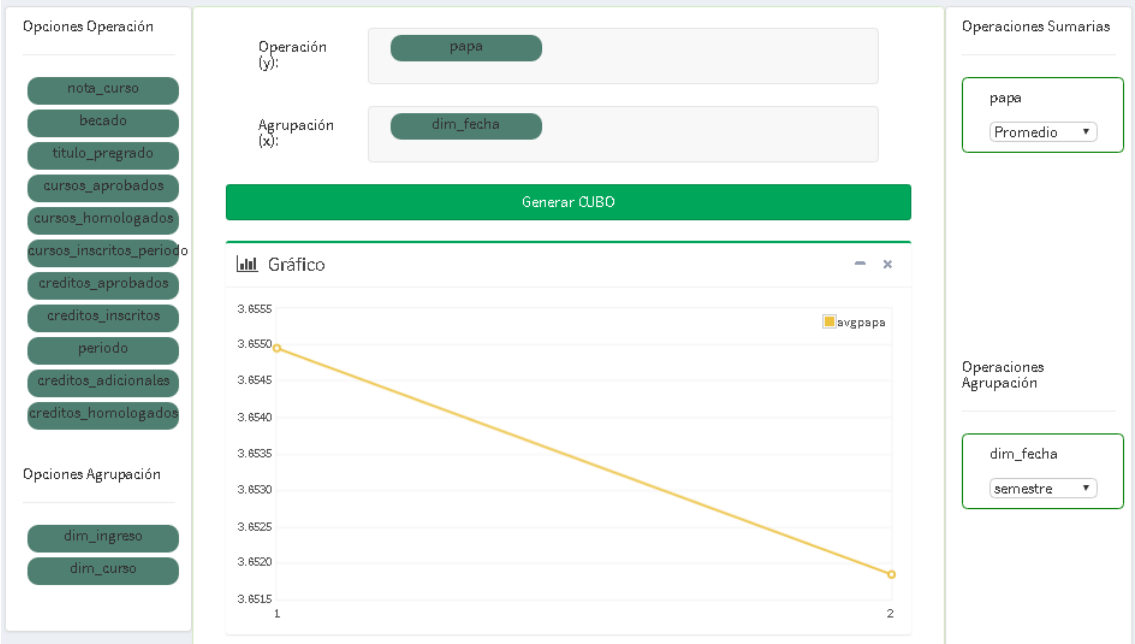



**Figura 5-14:** Promedio de PAPA por semestre

El análisis multidimensional realizado sobre las tablas de hechos “fact_academica” y “fact_interaccion”, permiten entender cómo se comportan y relacionan los datos almacenados en cada una de estas cuando se utilizan los mismos factores de agrupación, ya que las gráficas tienden a tener las mismas disposiciones, lo cual ayuda a identificar posibles relaciones entre las variables.

### 5.5.3 Enfoque predictivo

Para la realización de los análisis correspondientes al enfoque predictivo se creó inicialmente un dataset conformado por algunos de los atributos considerados relevantes para encontrar un patrón de comportamiento respecto a las interacciones de los estudiantes en las plataformas virtuales de aprendizaje y su rendimiento académico. Este dataset fue conformado con los siguientes campos: año, semestre, numero de interacciones por usuario, duración acumulada de las interacciones por usuario, promedio aritmético ponderado acumulado (PAPA) del periodo de las interacciones, cantidad de cursos inscritos en el periodo de las interacciones, número de cursos con presencia en la plataforma para el periodo de las interacciones, año del periodo académico anterior, semestre del periodo académico anterior, PAPA del periodo académico anterior, variación del PAPA respecto al periodo académico anterior. A continuación, se presenta la consulta

con la cual se extrajo el dataset inicial. La aplicación del algoritmo de Clustering sobre los diferentes dataset construidos fue realizada en Rapidminer y con esta misma herramienta se graficaron los resultados presentados.

```
SELECT dim_usuario.id AS id_usuario, dim_ingreso.id AS id_ingreso,
fact_academica_1.ano, fact_academica_1.semestre,
SUM(fact_interaccion_1.duracion_horas) AS duracion_horas_acumulado,
SUM(fact_interaccion_1.duracion_minutos) AS duracion_minutos_acumulado,
SUM(fact_interaccion_1.duracion_segundos) AS duracion_segundos_acumulado,
COUNT(fact_interaccion_1.id_interaccion) AS num_interacciones,
SUM(fact_interaccion_1.duracion) AS duracion_acumulada,
MAX(fact_academica_1.papa) AS papa,
MAX(fact_academica_1.cursos_inscritos_periodo) AS cursos_inscritos_periodo,
COUNT(DISTINCT fact_interaccion_1.id_curso) AS num_cursos_plataforma
FROM dim_usuario
INNER JOIN dim_ingreso
ON dim_ingreso.id_usuario=dim_usuario.id
INNER JOIN (SELECT fact_academica.id_ingreso, fact_academica.papa,
fact_academica.cursos_inscritos_periodo,
dim_fecha.id AS id_fecha, dim_fecha.ano, dim_fecha.semestre
FROM fact_academica
INNER JOIN dim_fecha
ON dim_fecha.id=fact_academica.id_fecha) fact_academica_1
ON fact_academica_1.id_ingreso=dim_ingreso.id
INNER JOIN (SELECT fact_interaccion.id_interaccion, fact_interaccion.id_ingreso,
fact_interaccion.id_fecha, fact_interaccion.id_curso,
fact_interaccion.duracion_horas, fact_interaccion.duracion_minutos,
fact_interaccion.duracion_segundos,
(((fact_interaccion.duracion_horas * 60)::FLOAT) +
((fact_interaccion.duracion_minutos)::FLOAT) +
((fact_interaccion.duracion_segundos / 60.0)::FLOAT)) AS duracion,
dim_fecha.id, dim_fecha.ano, dim_fecha.mes, dim_fecha.dia, dim_fecha.semestre
```

*FROM fact_interaccion*
*INNER JOIN dim_fecha*
*ON fact_interaccion.id_fecha=dim_fecha.id) fact_interaccion_1*
*ON fact_interaccion_1.id_ingreso=dim_ingreso.id AND*
*fact_interaccion_1.ano=fact_academica_1.ano AND*
*fact_interaccion_1.semestre=fact_academica_1.semestre*
*GROUP BY dim_usuario.id, dim_ingreso.id, fact_academica_1.ano, fact_academica_1.semestre*

Mediante la aplicación hecha en java se extrae el PAPA en el semestre anterior para cada estudiante recuperado con la consulta anterior, además el proceso de refinamiento del dataset se realizó cambiando los atributos que se exportan en la aplicación anteriormente mencionada.

Teniendo en cuenta que en los datos que conforman el dataset contienen registros de interacciones de estudiantes con la plataforma Blackboard para los períodos académicos segundo semestre de 2013 y primer semestre de 2014 y que los centroides de los clúster están en torno a estos dos períodos, se puede apreciar en la Tabla 5-17 que estos valores están concentrando los puntos del clúster, por lo tanto se toma la decisión de excluir del conjunto de datos estos atributos, de manera que estos no sean los elementos determinantes a la hora de conformar los clúster.

**Tabla 5-17:** Centroides en el Clustering sobre el dataset inicial

| Atributo | Clúster 1 | Clúster 2 |
|---|---|---|
| año | 2014 | 2013 |
| semestre | 1 | 2 |
| num_interacciones | 16,1855 | 21,3063 |
| duracion_acumulada | 1759,9973 | 2297,2315 |
| papa | 3,6796 | 3,6023 |
| cursos_inscritos_periodo | 5,2155 | 5,1965 |
| num_cursos_plataforma | 2,7194 | 2,3701 |
| año-1 | 2013,2792 | 2013 |
| semestre-1 | 1 | 1,1986 |
| papa-1 | 3,6722 | 3,6117 |
| delta_papa | 0,0074 | -0,0094 |

| Resultados clúster dataset 1.1 |
|---|
| Clúster 0: 2808 ítems |
| Clúster 1: 2351 ítems |

Fue entonces construido el dataset 1.2 al retirar los atributos (año, año-1, semestre y semestre-1). Con este nuevo dataset se hizo un análisis de correlación entre las variables y como se puede observar en la Figura 5-15 existe una correlación marcada entre las variables num_interacciones y duracion_acumulada dada la distribución de los puntos.

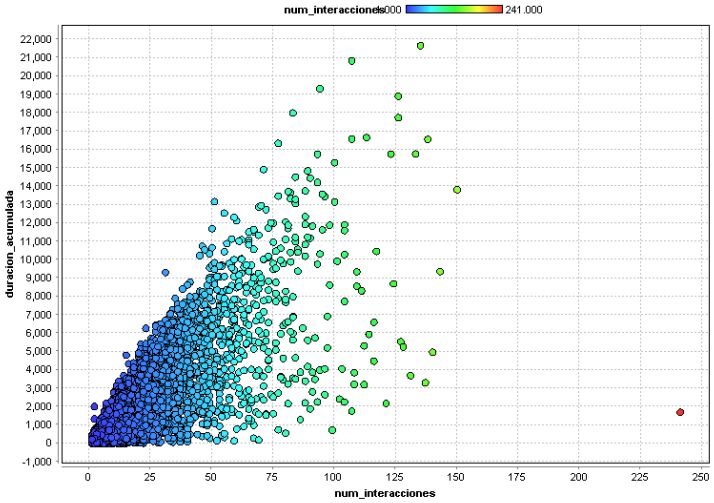



**Figura 5-15:** Correlación entre las variables num_interacciones y duracion_acumulada

Además, en la Tabla 5-18, se observa que estas dos variables presentan una correlación de 0.795, con lo cual se puede decir que ésta es alta, por lo tanto, se tomó la decisión de calcular con base a estos dos atributos un nuevo atributo llamado promedio_tiempo_interacciones, el cual fue calculado con la siguiente ecuación:

$$\text{promedio_tiempo_interacciones} = \frac{\text{duracion_acumulada}}{\text{num_interacciones}}$$

**Tabla 5-18:** Correlaciones entre los atributos del dataset

| Atributos | num_interacciones | duracion_acumulada | papa | cursos_inscritos_periodo | num_cursos_plataforma | papa-1 | delta_papa |
|---|---|---|---|---|---|---|---|
| num_interacciones | 1 | 0,795 | 0,093 | 0,138 | 0,571 | 0,088 | 0,012 |
| duracion_acumulada | 0,795 | 1 | 0,088 | 0,112 | 0,41 | 0,089 | -0,007 |
| papa | 0,093 | 0,088 | 1 | 0,268 | 0,019 | 0,939 | 0,142 |
| cursos_inscritos_periodo | 0,138 | 0,112 | 0,268 | 1 | 0,112 | 0,24 | 0,071 |
| num_cursos_plataforma | 0,571 | 0,41 | 0,019 | 0,112 | 1 | 0,017 | 0,005 |
| papa-1 | 0,088 | 0,089 | 0,939 | 0,24 | 0,017 | 1 | -0,206 |
| delta_papa | 0,012 | -0,007 | 0,142 | 0,071 | 0,005 | -0,206 | 1 |

A partir de este nuevo hallazgo se decidió construir el dataset 1.3, en el cual se redujeron los atributos al combinar mediante la aplicación de la ecuación los atributos num_interacciones y duracion_acumulada. Con este nuevo dataset se ejecutó el algoritmo de Clustering K-means para el K=2, K=3 y K=4.

A continuación, en la Tabla 5-19 y en la Figura 5-16 se muestran los resultados de los centroides para el K=2 en donde se puede observar una tendencia de varianza negativa del PAPA al tener mayor promedio de tiempos de interacción y una varianza positiva del PAPA al tener menor promedio de tiempos de interacción.

**Tabla 5-19:** Resultados del K-means para K=2

| Atributo | Clúster 1 | Clúster 0 |
|---|---|---|
| promedio_tiempo_interacciones | 174,831 | 54,012 |
| papa | 3,666 | 3,623 |
| cursos_inscritos_periodo | 5,175 | 5,238 |
| num_cursos_plataforma | 2,289 | 2,827 |
| papa-1 | 3,671 | 3,619 |
| delta_papa | -0,004 | 0,003 |
| Resultados clúster dataset 1.3 K=2 | | |
| Clúster 0: 2601 ítems | | |
| Clúster 1: 2558 ítems | | |

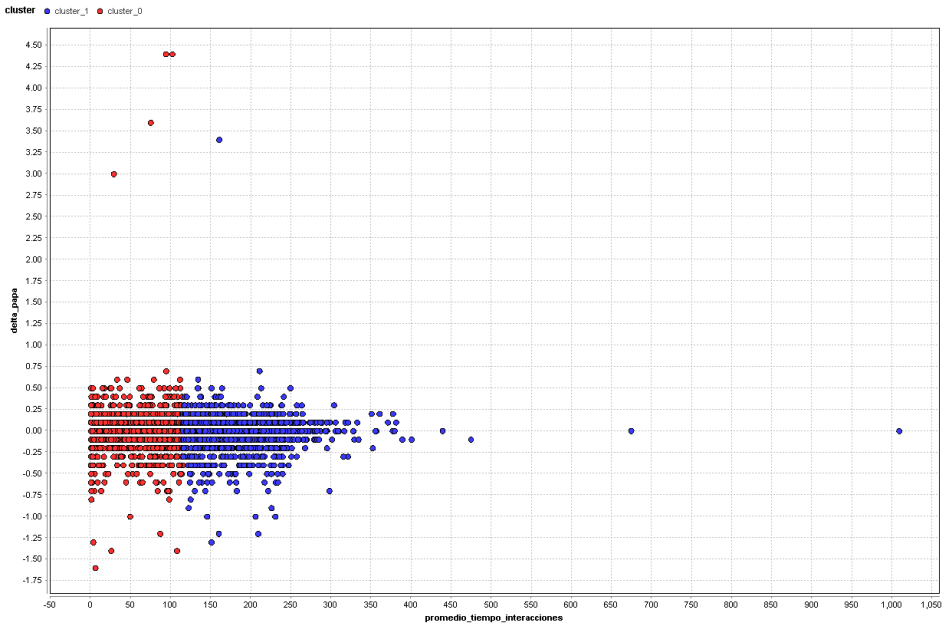



**Figura 5-16:** Distribución de los Clúster respecto al delta_papa y el promedio_tiempo_interacciones para K=2

Para corroborar esta tendencia se hizo la corrida del mismo algoritmo para un K=3. Los resultados muestran que la tendencia continúa. Se puede observar que los centroides del delta del PAPA disminuyen a medida que el tiempo promedio de interacción aumenta. Esto se puede apreciar en la Tabla 5-20 y en la Figura 5-17.

**Tabla 5-20:** Resultados del K-means para K=3

| Atributo | Clúster 0 | Clúster 1 | Clúster 2 |
|---|---|---|---|
| promedio_tiempo_interacciones | 210,073 | 122,817 | 33,37 |
| papa | 3,676 | 3,643 | 3,623 |
| cursos_inscritos_periodo | 5,045 | 5,329 | 5,176 |
| num_cursos_plataforma | 1,742 | 2,919 | 2,719 |
| papa-1 | 3,683 | 3,641 | 3,621 |
| delta_papa | -0,007 | 0,002 | 0,003 |
| Resultados clúster dataset 1.3 K=3 | | | |
| Cluster 0: 1275 items | | | |
| Cluster 1: 2127 items | | | |
| Cluster 2: 1757 items | | | |

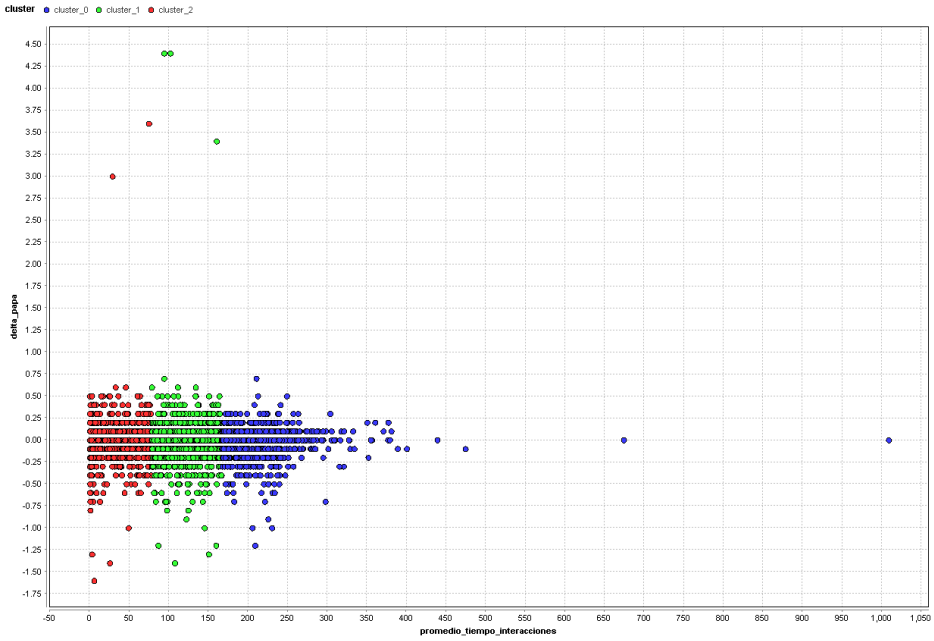



**Figura 5-17:** Distribución de los Clúster respecto al delta_papa y el promedio_tiempo_interacciones para K=3

Para finalizar este análisis se corrió el algoritmo para un K=4. En este caso la tendencia que se venía observando en las ejecuciones anteriores de una disminución del delta del PAPA a medida que el tiempo promedio de interacción aumentaba se rompió, ya que se puede observar que en el clúster 3 el delta del PAPA es menor al delta del PAPA del Clúster 1 donde el promedio de tiempo de interacciones es mayor. Esto se muestra en la Tabla 5-21 y en la Figura 5-18.

**Tabla 5-21:** Resultados del K-means para K=4

| Atributo | Clúster 0 | Clúster 2 | Clúster 1 | Clúster 3 |
|---|---|---|---|---|
| promedio_tiempo_interacciones | 228,422 | 154,652 | 92,667 | 23,489 |
| papa | 3,664 | 3,676 | 3,62 | 3,623 |
| cursos_inscritos_periodo | 4,964 | 5,246 | 5,375 | 5,129 |
| num_cursos_plataforma | 1,432 | 2,647 | 3,059 | 2,599 |
| papa-1 | 3,672 | 3,681 | 3,613 | 3,622 |
| delta_papa | -0,007 | -0,005 | 0,007 | 0,001 |
| Resultados clúster dataset 1.3 K=3 | | | | |
| Cluster 0: 808 items | | | | |
| Cluster 1: 1455 items | | | | |

| |
|---|
| Cluster 2: 1527 items |
| Cluster 3: 1369 items |

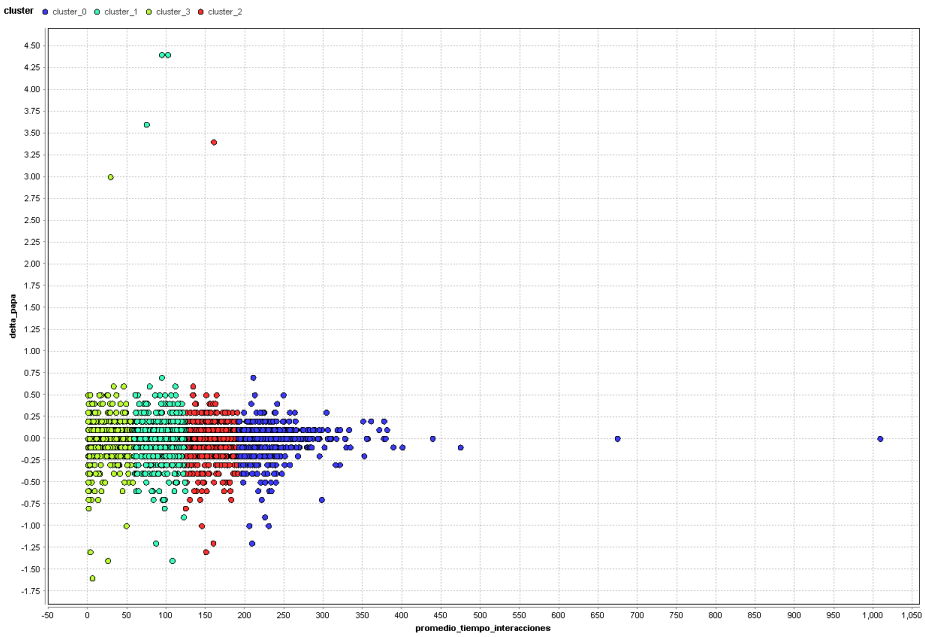



**Figura 5-18:** Distribución de los Clúster respecto al delta_papa y el promedio_tiempo_interacciones para K=4

De acuerdo a lo anterior y en vista que el patrón registrado para los clústeres formados para los K=2 y K=3 no se continuó evidenciando en los clústeres del K=4.

Se concluye que con los datos disponibles en el caso de estudio con el cual se realizó la validación del modelo propuesto no se puede demostrar que exista una relación directa expresada en un patrón de comportamiento que indique que los estudiantes que dedican mayor tiempo en las plataformas virtuales de aprendizaje obtienen mejores promedios en sus programas académicos dentro de la Universidad.

Los datos disponibles también han limitado la aplicación de otras analíticas, que incluyan por ejemplo el análisis de las interacciones por curso respecto a las notas obtenidas en los mismos o el análisis de las interacciones de acuerdo al tipo de contenido que se presente en cada curso.

### 5.5.4 Enfoque prescriptivo

Teniendo en cuenta los resultados obtenidos en la aplicación de los enfoques descriptivo, diagnóstico y predictivo, se obtienen las siguientes conclusiones:

- Se evidencia que los datos a los cuales se ha tenido acceso no son suficientes para encontrar información útil para el proceso de recomendación, ya que los datos de la información académica y de interacciones con las plataformas de educación virtual, tienen granularidad diferente, los primeros presentan los datos registrados cada semestre, sin incluir las notas obtenidas en cada curso inscrito por el estudiante, mientras que los segundos tienen los datos de las fechas y horas en las que los estudiantes realizaron las actividades dejadas por el profesor, lo cual no permite diferenciar muy bien si un estudiante que interactúa más con la plataforma tiene mejor rendimiento en este curso. Es por ello, que se hace necesario igualar el nivel de granularidad de los datos, obteniendo datos académicos más detallados, con lo cual es posible obtener mejores resultados a la hora de ejecutar los análisis que se consideren pertinentes.
- La baja cantidad de cursos que tienen presencia en las plataformas virtuales de educación, hace que los datos recolectados sean pocos y explica por qué existen estudiantes con tan bajo uso de las plataformas virtuales de educación. Por lo tanto, es necesario impulsar el uso de las plataformas virtuales de aprendizaje, mostrando a estudiantes y docentes que su uso beneficia los procesos educativos y puede ser un factor que potencie las habilidades de cada estudiante.
- En general los profesores solo utilizan la plataforma como mecanismo de intercambio de documentos, por lo que no hay muchos datos acerca de qué tipo de actividades beneficia más el proceso educativo del estudiante, lo cual puede llegar a constituir una fuente de datos significativa para realizar recomendación de actividades a los estudiantes. Por lo anterior, es importante impulsar el uso de otro tipo de actividades en la plataforma mostrando las ventajas que ofrece tanto a profesores como estudiantes.
- Para poder clasificar y predecir el comportamiento de los estudiantes con relación a las actividades en las plataformas se requiere tener las notas obtenidas en cada curso y periodo; a partir de allí poder especificar las preferencias de acción

orientadas a aprovechar las plataformas y las estrategias educativas que proporcionan para mejorar la efectividad de los procesos de enseñanza y aprendizaje.

- Es necesario tener datos académicos de los cursos y las notas obtenidas por cada estudiante en cada semestre y de esta manera verificar que tan provechoso es para el estudiante dedicar tiempo en realizar las actividades dejadas por el profesor en las plataformas virtuales de educación.

## 5.6 Conclusiones del capítulo

En este capítulo se presentaron los resultados de las pruebas realizadas con el caso de estudio, donde se diseñaron una serie de experimentos en los cuales se aplicaron los pasos del modelo propuesto en esta tesis. Con estas pruebas se demuestra que es posible implementar un modelo mediante el cual sea posible el descubrimiento de patrones en interacciones entre estudiantes y plataformas virtuales de educación mediante el uso de analíticas de aprendizaje, siempre y cuando se cuente con los datos necesarios y suficientes que permitan esbozar tendencias comportamentales de los estudiantes durante su proceso educativo. Además, se muestra que es posible crear un modelo de almacenamiento que permita almacenar datos de una forma coherente y organizada lo cual permite a través del uso de técnicas de minería de datos extraer información cruzando los datos almacenados en las diferentes tablas de hechos que hacen parte del esquema de almacenamiento, lo que facilita el descubrimiento de tendencias y relaciones que tienen los datos allí almacenados.

Sin embargo, dentro de los análisis realizados se encontró un caso particular, donde al aplicar el análisis multidimensional del enfoque diagnóstico, encontramos que al realizar la agrupación de los datos de PAPA de los estudiantes por facultad, la gráfica tiene un comportamiento muy parecido a la agrupación del número de interacciones con la plataforma de los estudiantes por facultad, lo que tratamos de corroborar durante la aplicación del enfoque predictivo con la creación de clústeres, no obstante, al tener más variables presentes, los resultados demuestran que esta tendencia no se cumple, dando a entender que las variables cambian su comportamiento al estar en presencia de otras.

Dado los resultados encontrados durante la aplicación del enfoque predictivo, para el enfoque prescriptivo, se concluye que al no contar con datos que permitieran la generación de patrones de comportamiento, no se puede determinar con precisión las situaciones que se presentan en esta institución respecto al uso de las plataformas y su influencia en el rendimiento académico. Se espera al contar con más datos, poder establecer estas relaciones y finalmente producir las recomendaciones necesarias.

# 6. Conclusiones y trabajos futuros

## 6.1 Conclusiones

Las instituciones educativas recolectan y generan gran cantidad de datos sobre sus estudiantes, no solo sobre su rendimiento académico, condición socioeconómica y situación sociocultural, sino también sobre sus hábitos de estudio, esto mediante la recolección de datos sobre las horas de entrada, salida y tiempos de permanencia en las plataformas virtuales de educación. Sin embargo, esta gran cantidad de datos se maneja de forma aislada, lo cual dificulta aprovechar la información que allí se encuentra oculta.

A continuación, se responden las preguntas de investigación planteadas al inicio de la tesis.

*¿Es posible elaborar un modelo de descubrimiento de patrones sobre datos educativos y de las interacciones de los estudiantes con las plataformas virtuales utilizando analíticas de aprendizaje?*

Si, ya que este gran conjunto de datos educativos, fue en parte, el punto de partida para el desarrollo de esta tesis la cual se enfoca en relacionar las fuentes de datos académicos con las interacciones en plataformas virtuales.

Con lo anterior, se buscó encontrar patrones comportamentales de los estudiantes mediante la creación y aplicación de un modelo de analíticas de aprendizaje enfocado en la recopilación, procesamiento, almacenamiento y análisis de los datos de este dominio.

Para apoyar la toma de decisiones que fomenten la mejora en los procesos de enseñanza y aprendizaje de los estudiantes y en lo posible, contribuir a la generación de estrategias para el mejoramiento de la calidad en la educación.

*¿Cómo se pueden incorporar técnicas de minería de datos en el modelo basado en analíticas de aprendizaje, para poder hacer recomendaciones que mejoren los procesos educativos apoyados en plataformas virtuales?*

Se pueden incorporar técnicas de minería de datos en el modelo a través de la aplicación de los cuatro enfoques que se desarrollaron en la tesis, a saber: el enfoque descriptivo (permite conocer con que datos se cuenta), el enfoque diagnóstico (permite conocer el comportamiento de los datos), el enfoque predictivo (permite encontrar información útil para la toma de decisiones) y el enfoque prescriptivo (permite tomar las decisiones con base en los hallazgos) se da la posibilidad de entender el comportamiento de los estudiantes, sus interacciones y los factores externos que están influyendo y afectando los procesos educativos del mismo.

Como aportes logrados con esta tesis se destacan:

- La creación de una bodega de datos académica, la cual integra datos educativos de diferentes fuentes provenientes de la Universidad Nacional de Colombia Sede Manizales y servirá como insumo para procesos de análisis llevados a cabo por la Dirección Académica de este plantel educativo.
- El esquema de almacenamiento propuesto para el modelo conjuga dos visiones de las bodegas de datos, al integrar un componente relacional y un componente multidimensional atendiendo de esta manera las necesidades presentes en las diferentes fuentes de datos.
- La construcción de un modelo para el análisis de datos educativos que contempla no solo la aplicación de analíticas de aprendizaje, sino también fue diseñado atendiendo a las diferentes fases que se requieren en un proceso de KDD.
- El modelo propuesto fue aplicado a un caso de estudio particular combinando diferentes fuentes de datos, logrando que el modelo de bodega de datos hibrida fuera útil para la extracción de datos y llegando a cubrir los tres primeros enfoques de las analíticas de aprendizaje.
- Al realizar el diagnóstico del uso de las plataformas virtuales de educación con estudiantes y docentes de la Universidad del caso de estudio, se pudo corroborar la falta de explotación de las plataformas como Moodle y Blackboard, para las diferentes actividades educativas que estas ofrecen y que pueden llegar a incentivar el aprendizaje desde los diversos estilos.

*¿Cómo se pueden incorporar técnicas de minería de datos en el modelo basado en analíticas de aprendizaje, para poder hacer recomendaciones que mejoren los procesos educativos apoyados en plataformas virtuales?*

Basado en la revisión del estado del arte, se revisaron varios modelos propuestos por diferentes autores, sin embargo, no se encontró ninguno que contemplara todas las fases de descubrimiento de patrones usando analíticas de aprendizaje.

## 6.2 Trabajos futuros

Como un trabajo en marcha, actualmente se encuentra realizando la revisión de los datos de la plataforma Moodle con expertos de la institución educativa para poder incorporarlos a la bodega de datos académica y realizar la aplicación de los cuatro enfoques de las analíticas sobre ellos. Como trabajos futuros se plantea:

- Cruzar los datos de las interacciones recolectados con los datos socioeconómicos y sociodemográficos y de esta manera tratar de encontrar comportamientos relacionados con las condiciones de vida de los estudiantes para generar sistemas de apoyo para quien lo necesite antes de presentarse problemas como bajo rendimiento académico.

- Extraer desde la web los datos de la prueba Saber 11 presentada por el estudiante y agregar estos al modelo, lo cual puede contribuir a determinar en qué áreas del conocimiento el estudiante requiere mayor apoyo y así mejorar su proceso educativo. También tomar los datos detallados del examen de admisión del estudiante para identificar falencias y poder fortalecerlas.

- Agregar al modelo del estudiante dentro del esquema de almacenamiento, datos relacionados con el estilo de aprendizaje, de tal manera que permita identificar características comportamentales de cada uno de los estilos que pueden presentar los estudiantes y así poder identificar el estilo de aprendizaje del estudiante sin necesidad de aplicar pruebas para tal fin.

- Crear una ontología que permita la representación del conocimiento relacionado con los componentes del modelo propuesto.

- Agregar dentro de la aplicación desarrollada en JAVA para el proceso de ETL, un módulo que permita la generación de las gráficas de análisis estadístico y relacional de los datos.

# A. Anexo: Encuesta para el diagnóstico del uso de plataformas virtuales de aprendizaje

A continuación, se presentan los instrumentos aplicados para el diagnóstico del uso de las plataformas virtuales de aprendizaje por parte de los estudiantes y docentes de la Universidad del caso de estudio y su respectiva tabulación.

## *Encuesta sobre el uso de plataformas virtuales de aprendizaje para estudiantes*

El siguiente formulario corresponde a una encuesta dirigida a estudiantes de pregrado y posgrado de la Universidad Nacional de Colombia - Sede Manizales como parte de un proyecto de investigación, que tiene por objetivo construir un estado actual del uso de las plataformas virtuales de aprendizaje en las asignaturas impartidas en los diferentes programas de formación. Solicitamos su colaboración diligenciando el formulario y agradecemos sus respuestas, al finalizar hemos dispuesto un espacio para observaciones respecto a la comprensión de las preguntas o comentarios que nos puedan ayudar a mejor el instrumento.

***Preguntas preliminares***

***Pregunta 1:*** Nivel de formación actual

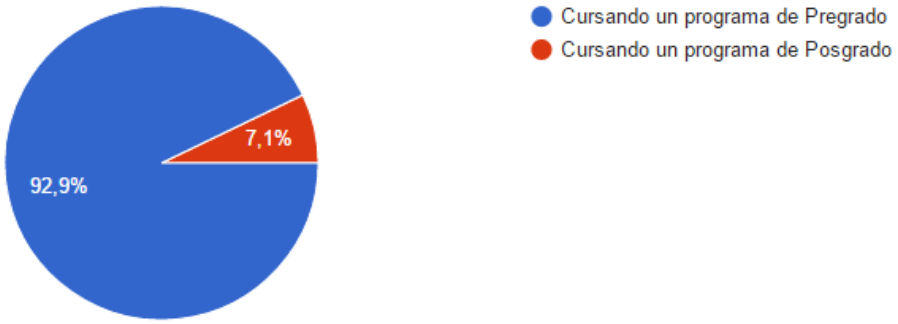



**Figura 6-1:** Nivel de formación de los estudiantes encuestados

***Pregunta 2:*** Conoce qué es una plataforma virtual de aprendizaje (plataformas e-Learning) o LMS (Learning Management System)? (Ejemplo: Moodle, Blackboard, Udacity, Caroline, Coursera, entre otras)

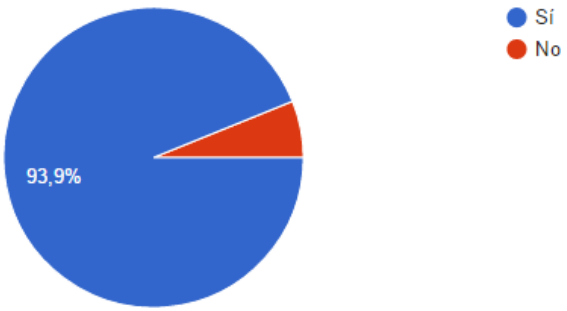



**Figura 6-2:** Conocimiento de la existencia de LMS

***Pregunta 3:*** En las asignaturas que ha cursado o está cursando en su programa académico actual, ¿ha hecho uso de alguna plataforma virtual de aprendizaje?

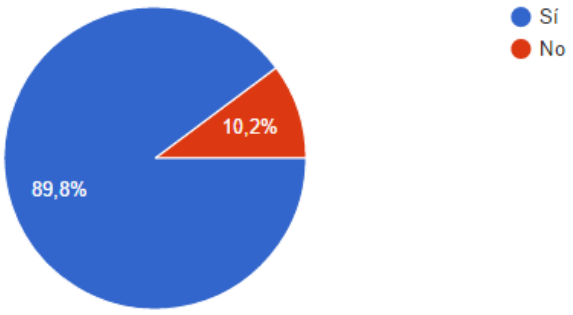



**Figura 6-3:** Uso de plataformas virtuales en los programas académicos

***Uso de las plataformas virtuales en su programa académico actual***

***Pregunta 1:*** Si su respuesta anterior fue afirmativa, ¿Cuál (es) plataforma (s) ha usado o se encuentra usando?

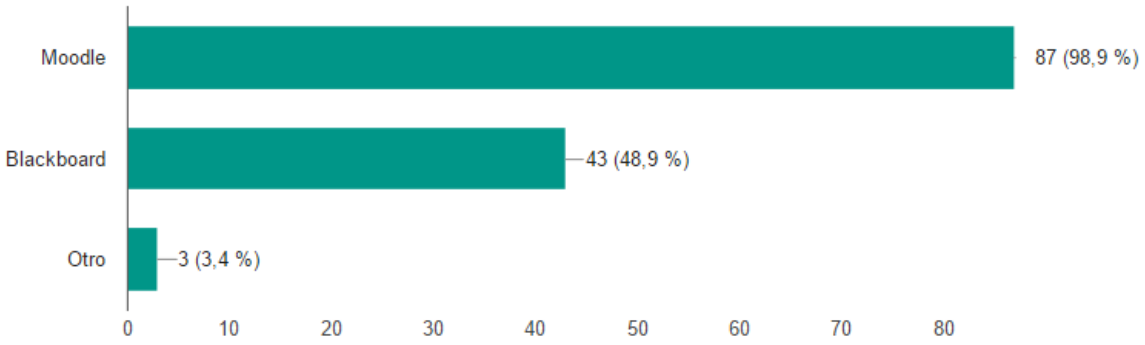



**Figura 6-4:** Plataformas usadas por los estudiantes en la Universidad

***Pregunta 2:*** En los cursos de su programa académico actual, para cuáles de las siguientes acciones suele usar o ha usado las plataformas virtuales de aprendizaje:

1. Participación en Foros de discusión con compañeros y docente
2. Revisión y descarga de materiales educativos como archivos pdf, presentaciones de diapositivas, videos, entre otras
3. Evaluación de su desempeño
4. Autoevaluaciones o evaluación del docente
5. Chat con compañeros o docente
6. Videoconferencias con compañeros o docente

7. Comunicación por medio del envío de mensajes
8. Desarrollo de tareas o actividades
9. Otro

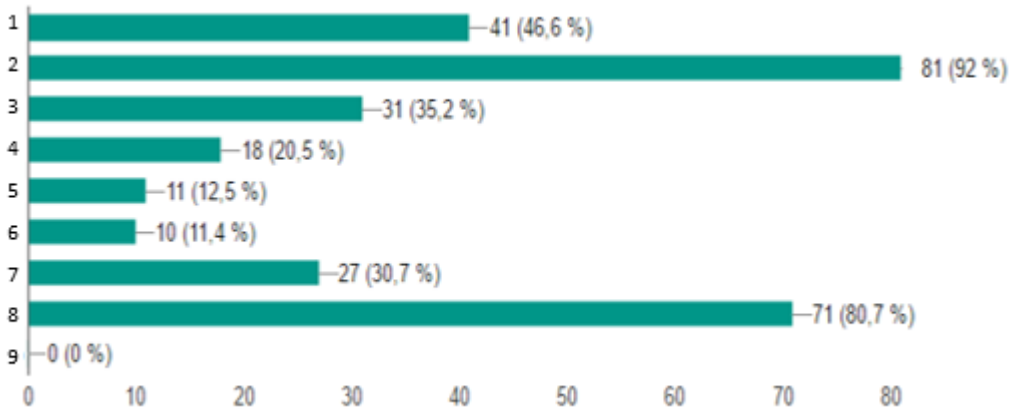



**Figura 6-5:** Actividades realizadas por los estudiantes en los LMS

***Pregunta 3:*** En promedio, ¿con qué frecuencia hace uso de las plataformas virtuales de aprendizaje en las asignaturas que cursa en su programa académico actual?

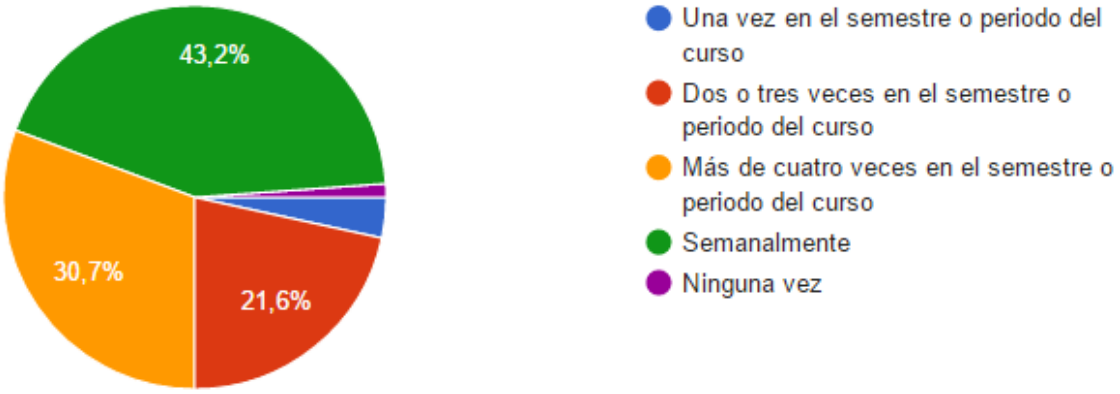



**Figura 6-6:** Frecuencia de uso de LMS en los programas académicos

***<u>Uso de las plataformas virtuales en otros procesos de formación</u>***

***Pregunta 1:*** En otros procesos de formación, diferentes a su programa académico actual (por ejemplo, cursos de corta duración, cursos complementarios, capacitaciones, certificaciones, entre otros) ¿ha hecho uso de alguna plataforma virtual de aprendizaje?

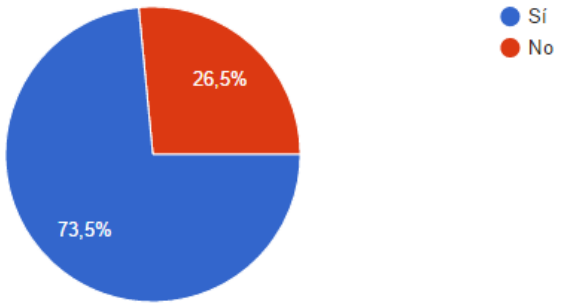



**Figura 6-7:** Uso de LMS en otros procesos de formación

***Pregunta 2:*** Si su respuesta anterior fue afirmativa, ¿Cuál (es) plataforma (s) ha usado?

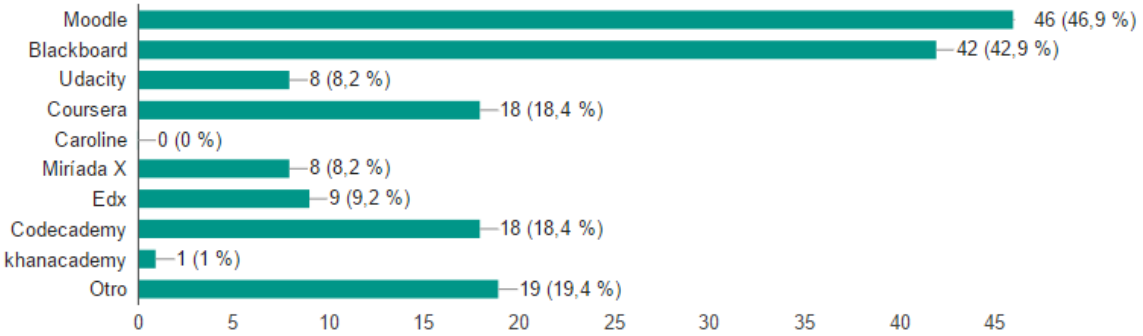



**Figura 6-8:** Plataformas usadas en otros procesos de formación

***Pregunta 3:*** En promedio, ¿con qué frecuencia hace uso de las plataformas virtuales de aprendizaje en procesos de formación diferentes a su programa académico dentro de la Universidad?

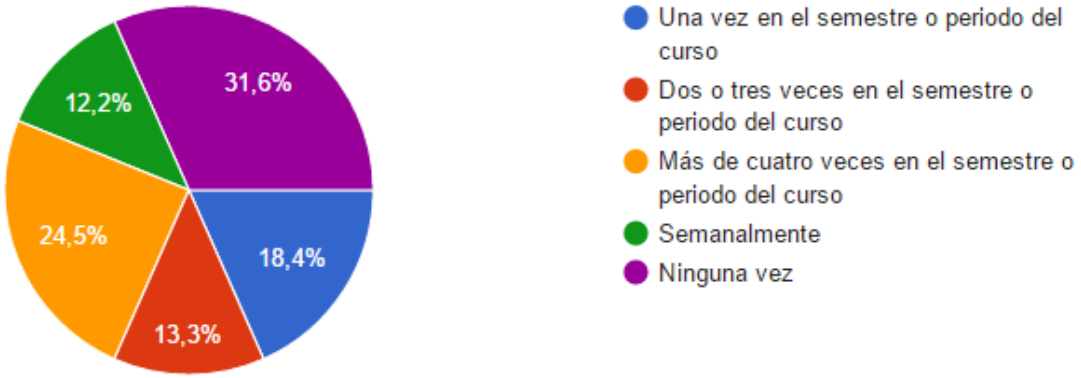



**Figura 6-9:** Frecuencia de uso de los LMS en otros procesos de formación

***Algunas observaciones sobre la encuesta***

*"Los profesores deberian de incitar más a los estudiantes a usar esa plataforma, actualmente de 5 materia que veo este semestre solo una materia utiliza moodle"*

*"He notado que en la universidad no se hace uso de todo el potencial y herramientas que ofrecen las plataformas virtuales. Los profesores generalmente se limitan a subir contenido; me pregunto si es por falta de interés o desconocimiento"*

*"la mejor plataforma de aprendizaje actualmente es youtube todo los estudiantes que he analizado estudian ahi para preparar examenes de cualquier indole, asi que para mí no estaria nada mal que le prestaran más atención a esa plataforma"*

*"NO suelo usarlas mucho no me siento comoda con ellas porque olvido que suben actividades y no las notifica"*

## *Encuesta sobre el uso de plataformas virtuales de aprendizaje para docentes*

El siguiente formulario corresponde a una encuesta dirigida a docentes de la Universidad Nacional de Colombia - Sede Manizales como parte de un proyecto de investigación, que tiene por objetivo construir un estado actual del uso de las plataformas virtuales de aprendizaje en las asignaturas impartidas en los diferentes programas de formación.

Solicitamos su colaboración diligenciando el formulario y agradecemos sus respuestas, al finalizar hemos dispuesto un espacio para observaciones respecto a la comprensión de las preguntas o comentarios que nos puedan ayudar a mejor el instrumento.

***Pregunta 1:*** Conoce qué es una plataforma virtual de aprendizaje (Plataformas e-Learning) o LMS (Learning Management System)? (Ejemplo: Moodle, Blackboard, Udacity, Caroline, Coursera, entre otras)

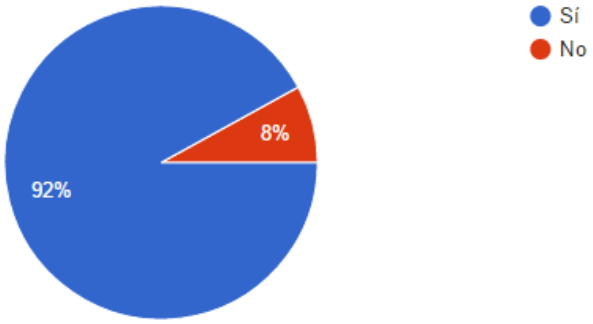



**Figura 6-10:** Conocimiento de la existencia de LMS

***Pregunta 2:*** En la (s) asignatura (s) que imparte, ¿hace uso de alguna plataforma virtual de aprendizaje?

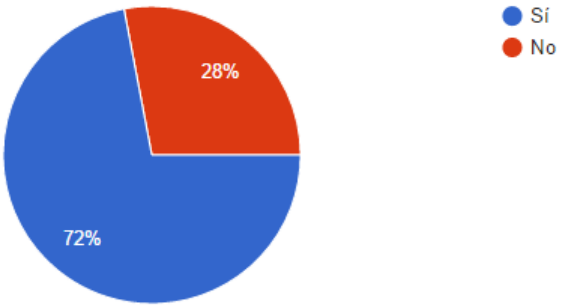



**Figura 6-11:** Uso de los LMS por parte de los docentes

***Pregunta 3:*** En caso de que la respuesta sea afirmativa, ¿Cuál (es) plataforma (s) ha utilizado?

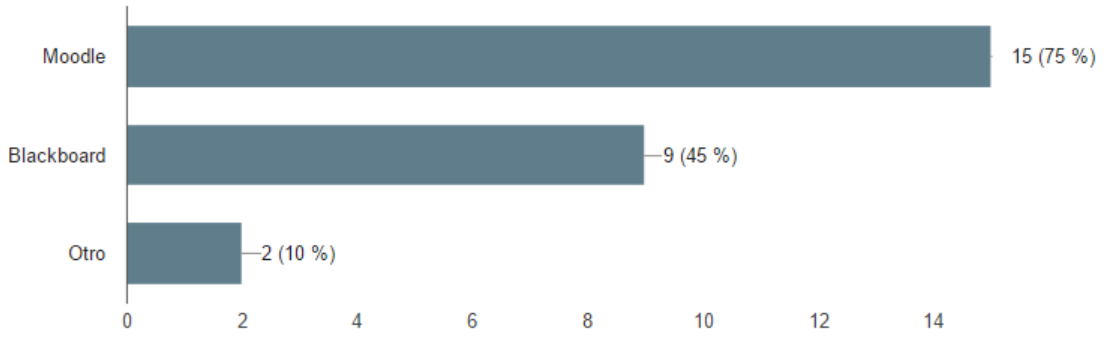



**Figura 6-12:** Plataformas usadas por los docentes

***Pregunta 4:*** Para cuáles de las siguientes acciones suele usar la plataforma:

1. Realización de foros de discusión con los estudiantes
2. Carga de materiales educativos como archivos pdf, presentaciones de diapositivas, videos, entre otras
3. Evaluación de desempeño de los estudiantes
4. Autoevaluaciones por parte de los estudiantes
5. Chat con y entre estudiantes
6. Videoconferencias con y entre estudiantes
7. Generación de reportes de uso de la plataforma para su curso
8. Comunicación por medio del envío de mensajes
9. Carga de tareas o actividades
10. Otro

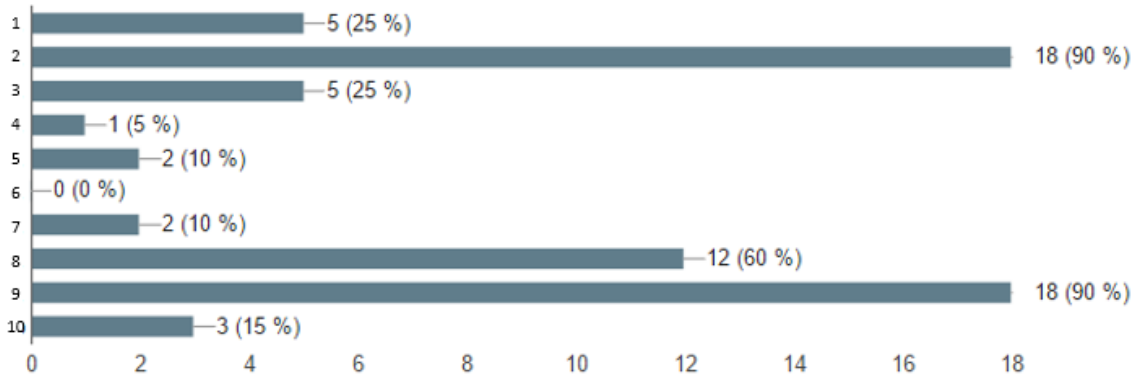



**Figura 6-13:** Actividades que emplean los docentes en sus cursos en los LMS

***Pregunta 5:*** En promedio, ¿con qué frecuencia hace uso de la plataforma virtual de aprendizaje en los cursos que imparte?

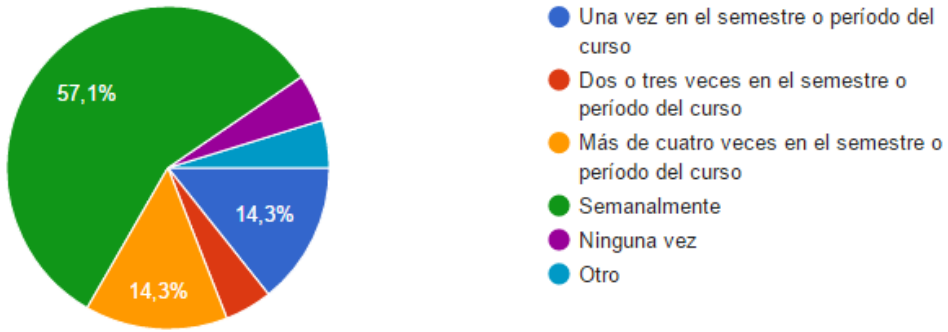



**Figura 6-14:** Frecuencia de uso de los LMS por parte de los docentes

***Pregunta 6:*** Si usted ha impartido un curso o asignatura por más de un semestre y para ello utiliza como apoyo una plataforma virtual de aprendizaje, normalmente suele:

1. Mantener los mismos materiales (lecturas, presentaciones, videos, etc)
2. Actualizar los materiales y/o actividades de acuerdo al desempeño de los estudiantes
3. Actualizar los materiales y/o actividades de acuerdo a la exploración que usted hace en la red o a la actualización de sus conocimientos
4. Actualizar los materiales y/o actividades de acuerdo a retroalimentaciones de los estudiantes respecto al curso
5. Actualizar los materiales y/o actividades de acuerdo a reportes generados por la plataforma usada

6. Otro

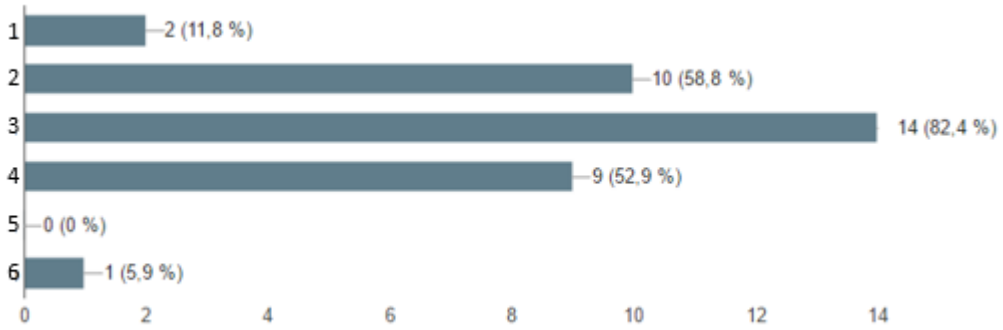



**Figura 6-15:** Estrategias de mantenimiento de cursos

***Algunas observaciones sobre la encuesta***

*"Preguntar si el docente conoce las ventajas de la web 2.0 que ofrece los LMS, y si las aprovecha :)"*

*"Bien, Muy compacta"*

*"Importante motivar el uso de las plataformas virtuales"*

*"Es importante conocer que herramientas son las que más usan los docentes de la plataforma, y cuales le gustaría saber manejar mejor".*

# B. Anexo: Diseño de la aplicación para realizar ETL

A continuación, se detalla brevemente algunos diagramas de diseño de la aplicación construida en JAVA para la realización del proceso ETL.

**Diagrama de casos de uso de alto nivel:**

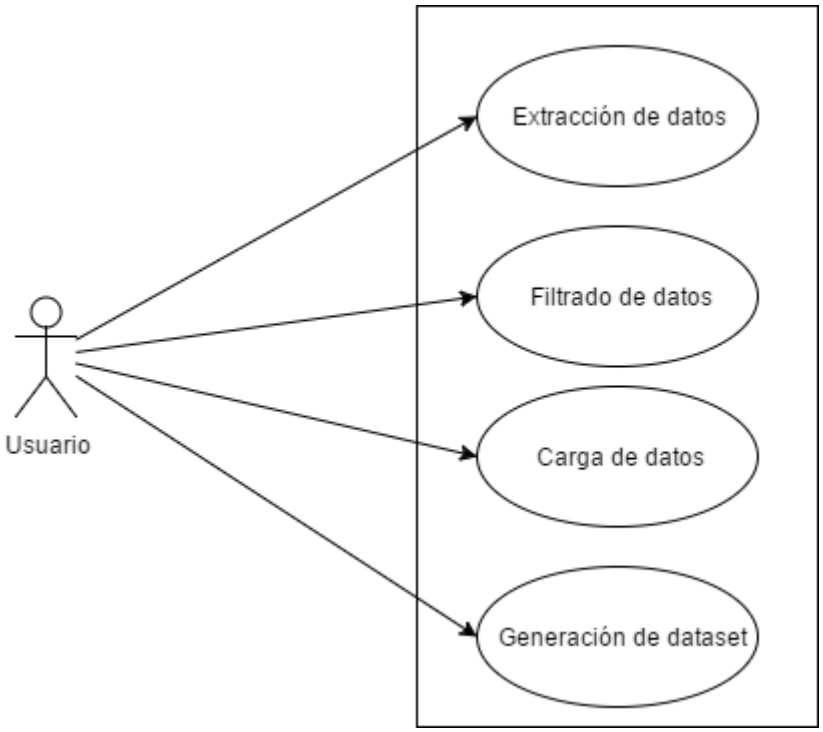


**Diagrama detallado caso de uso – Extracción de datos:**

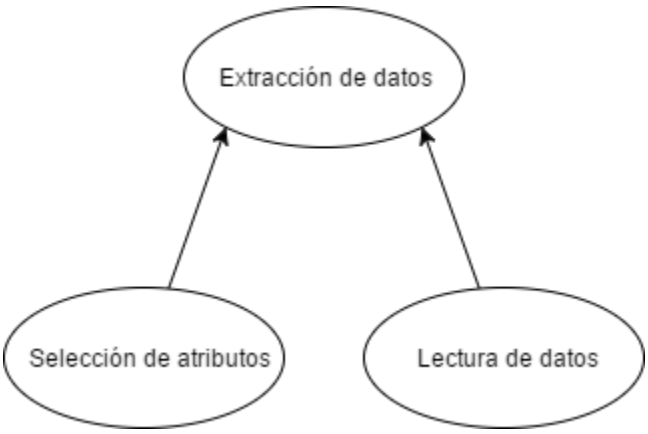



**Diagrama detallado caso de uso – Filtrado de datos:**

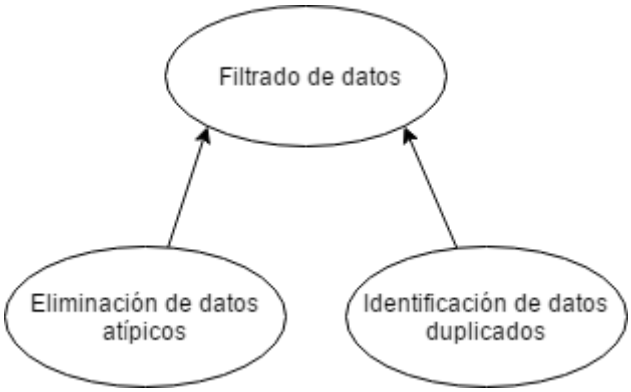



**Diagrama detallado caso de uso – Carga de datos:**

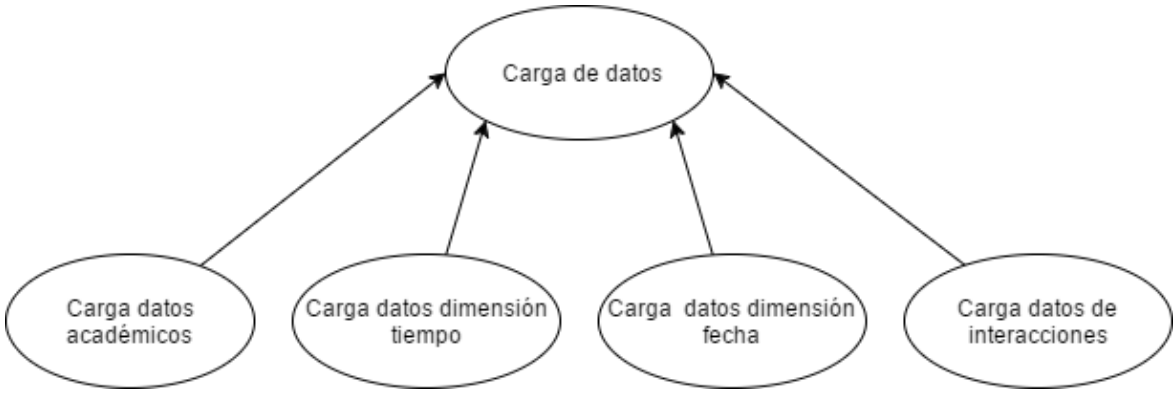
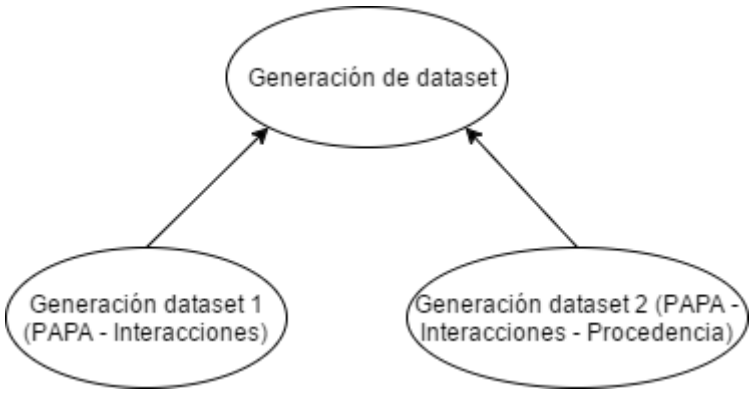


**Diagrama detallado caso de uso – Generación de dataset:**




A continuación, se adjuntan las instrucciones SQL utilizadas para la extracción de datos a la bodega de datos académica construida para el caso de estudio.

**Consulta para extraer interacciones de Blackboard:**

Las líneas que tienen doble slash (//) son comentarios.

```
SELECT
//Campos extraidos tabla users
users.pk1 AS id_usuario_plataforma, users.user_id AS nombre_usuario, users.firstname AS nombre,
users.lastname AS apellidos, users.student_id AS documento, users.email AS correo,
users.last_login_date AS ultimo_ingreso,
//Campos extraidos de la tabla ods_aa_session_activity
ods_aa_session_activity.login_time AS hora_acceso_sesion,
ods_aa_session_activity.logout_time AS hora_salida_sesion,
ods_aa_session_activity.session_pk1 AS id_sesion_usuario,
//Campos extraidos tabla ods_aa_course_activity
ods_aa_course_activity.initial_datetime_access AS hora_acceso_curso,
ods_aa_course_activity.course_access_minutes AS duracion_acceso_curso,
//Campos extraidos tabla ods_ls_course
ods_ls_course.pk1 AS id_curso, ods_ls_course.course_name AS nombre_curso
FROM users
```

```
//Join con tabla ods_aa_session_activity para obtener los datos de las sesiones del usuario
INNER JOIN ods_aa_session_activity
ON ods_aa_session_activity.user_pk1=users.pk1
//Join con tabla ods_aa_course_activity para obtener los datos de los ingresos a los cursos
INNER JOIN ods_aa_course_activity
ON ods_aa_course_activity.session_pk1=ods_aa_session_activity.session_pk1
//Join con tabla ods_ls_course para obtener los datos de los cursos
INNER JOIN ods_ls_course
ON ods_ls_course.pk1=ods_aa_course_activity.course_pk1
WHERE users.email='correo usuario'
```

**Consulta para extraer el dataset 1.1 de la bodega de datos académica:**

```
SELECT dim_usuario.id AS id_usuario, dim_ingreso.id AS id_ingreso,
fact_academica_1.ano, fact_academica_1.semestre,
SUM(fact_interaccion_1.duracion_horas) AS duracion_horas_acumulado,
SUM(fact_interaccion_1.duracion_minutos) AS duracion_minutos_acumulado,
SUM(fact_interaccion_1.duracion_segundos) AS duracion_segundos_acumulado,
COUNT(fact_interaccion_1.id_interaccion) AS num_interacciones,
SUM(fact_interaccion_1.duracion) AS duracion_acumulada,
MAX(fact_academica_1.papa) AS papa,
MAX(fact_academica_1.cursos_inscritos_periodo) AS cursos_inscritos_periodo,
COUNT(DISTINCT fact_interaccion_1.id_curso) AS num_cursos_plataforma
FROM dim_usuario
INNER JOIN dim_ingreso
ON dim_ingreso.id_usuario=dim_usuario.id
INNER JOIN (SELECT fact_academica.id_ingreso, fact_academica.papa,
fact_academica.cursos_inscritos_periodo,
dim_fecha.id AS id_fecha, dim_fecha.ano, dim_fecha.semestre
FROM fact_academica
INNER JOIN dim_fecha
ON dim_fecha.id=fact_academica.id_fecha) fact_academica_1
ON fact_academica_1.id_ingreso=dim_ingreso.id
```

```
INNER JOIN (SELECT fact_interaccion.id_interaccion, fact_interaccion.id_ingreso,
fact_interaccion.id_fecha, fact_interaccion.id_curso,
fact_interaccion.duracion_horas, fact_interaccion.duracion_minutos,
fact_interaccion.duracion_segundos,
(((fact_interaccion.duracion_horas * 60)::FLOAT) +
((fact_interaccion.duracion_minutos)::FLOAT) +
((fact_interaccion.duracion_segundos / 60.0)::FLOAT)) AS duracion,
dim_fecha.id, dim_fecha.ano, dim_fecha.mes, dim_fecha.dia, dim_fecha.semestre
FROM fact_interaccion
INNER JOIN dim_fecha
ON fact_interaccion.id_fecha=dim_fecha.id) fact_interaccion_1
ON fact_interaccion_1.id_ingreso=dim_ingreso.id AND
fact_interaccion_1.ano=fact_academica_1.ano AND
fact_interaccion_1.semestre=fact_academica_1.semestre
GROUP BY dim_usuario.id, dim_ingreso.id, fact_academica_1.ano,
fact_academica_1.semestre
```

**Consulta para extraer el dataset 1.2 de la bodega de datos académica:**

```
SELECT
academica_interaccion.id_usuario, academica_interaccion.id_ingreso,
academica_interaccion.ano, academica_interaccion.semestre,
academica_interaccion.duracion_acumulada, academica_interaccion.num_interacciones,
((academica_interaccion.duracion_acumulada /
academica_interaccion.num_interacciones)::FLOAT) AS promedio_tiempo_interacciones,
academica_interaccion.papa, academica_interaccion.cursos_inscritos_periodo,
academica_interaccion.num_cursos_plataforma,
dim_colegio.modalidad, dim_colegio.calendario, dim_colegio.jornada,
dim_colegio.tipo, dim_colegio.categoria, dim_ciudad.codigo, dim_ciudad.nombre
FROM dim_usuario
INNER JOIN (SELECT dim_usuario.id AS id_usuario, dim_ingreso.id AS id_ingreso,
fact_academica_1.ano, fact_academica_1.semestre,
SUM(fact_interaccion_1.duracion_horas) AS duracion_horas_acumulado,
SUM(fact_interaccion_1.duracion_minutos) AS duracion_minutos_acumulado,
```

```
SUM(fact_interaccion_1.duracion_segundos) AS duracion_segundos_acumulado,
COUNT(fact_interaccion_1.id_interaccion) AS num_interacciones,
SUM(fact_interaccion_1.duracion) AS duracion_acumulada,
MAX(fact_academica_1.papa) AS papa,
MAX(fact_academica_1.cursos_inscritos_periodo) AS cursos_inscritos_periodo,
COUNT(DISTINCT fact_interaccion_1.id_curso) AS num_cursos_plataforma
FROM dim_usuario
INNER JOIN dim_ingreso
ON dim_ingreso.id_usuario=dim_usuario.id
INNER JOIN (SELECT fact_academica.id_ingreso, fact_academica.papa,
fact_academica.cursos_inscritos_periodo,
dim_fecha.id AS id_fecha, dim_fecha.ano, dim_fecha.semestre
FROM fact_academica
INNER JOIN dim_fecha
ON dim_fecha.id=fact_academica.id_fecha) fact_academica_1
ON fact_academica_1.id_ingreso=dim_ingreso.id
INNER JOIN (SELECT fact_interaccion.id_interaccion, fact_interaccion.id_ingreso,
fact_interaccion.id_fecha, fact_interaccion.id_curso,
fact_interaccion.duracion_horas, fact_interaccion.duracion_minutos,
fact_interaccion.duracion_segundos,
(((fact_interaccion.duracion_horas * 60)::FLOAT) +
((fact_interaccion.duracion_minutos)::FLOAT) +
((fact_interaccion.duracion_segundos / 60.0)::FLOAT)) AS duracion,
dim_fecha.id, dim_fecha.ano, dim_fecha.mes, dim_fecha.dia, dim_fecha.semestre
FROM fact_interaccion
INNER JOIN dim_fecha
ON fact_interaccion.id_fecha=dim_fecha.id) fact_interaccion_1
ON fact_interaccion_1.id_ingreso=dim_ingreso.id AND
fact_interaccion_1.ano=fact_academica_1.ano AND
fact_interaccion_1.semestre=fact_academica_1.semestre
GROUP BY dim_usuario.id, dim_ingreso.id, fact_academica_1.ano,
fact_academica_1.semestre) academica_interaccion
ON academica_interaccion.id_usuario=dim_usuario.id
```

```
INNER JOIN dim_colegio
ON dim_colegio.id=dim_usuario.id_colegio
INNER JOIN dim_ciudad
ON dim_ciudad.id=dim_colegio.id_ciudad
```

**Consulta para extraer el PAPA en el semestre anterior al dato que se está recuperando:**

```
SELECT dim_ingreso.id, fact_academica.id_fecha, dim_fecha.ano, dim_fecha.semestre,
fact_academica.papa
FROM dim_ingreso
INNER JOIN fact_academica
ON fact_academica.id_ingreso=dim_ingreso.id
INNER JOIN dim_fecha
ON dim_fecha.id=fact_academica.id_fecha
WHERE fact_academica.id_ingreso=id_ingreso
AND dim_fecha.ano=año-1 AND dim_fecha.semestre=semestre-1
```

# C. Anexo: Diccionario de datos de la bodega de datos académica

A continuación, se presenta el diccionario de datos de la bodega de datos académica, donde se describe de forma detallada cada uno de los campos de las tablas de hechos y de las tablas de dimensiones.

La Figura 6-16 presenta el modelo de la bodega, por su extensión no se puede apreciar en detalle, por lo cual se ha ampliado cada una de las tablas colocándolas al inicio del diccionario correspondiente.

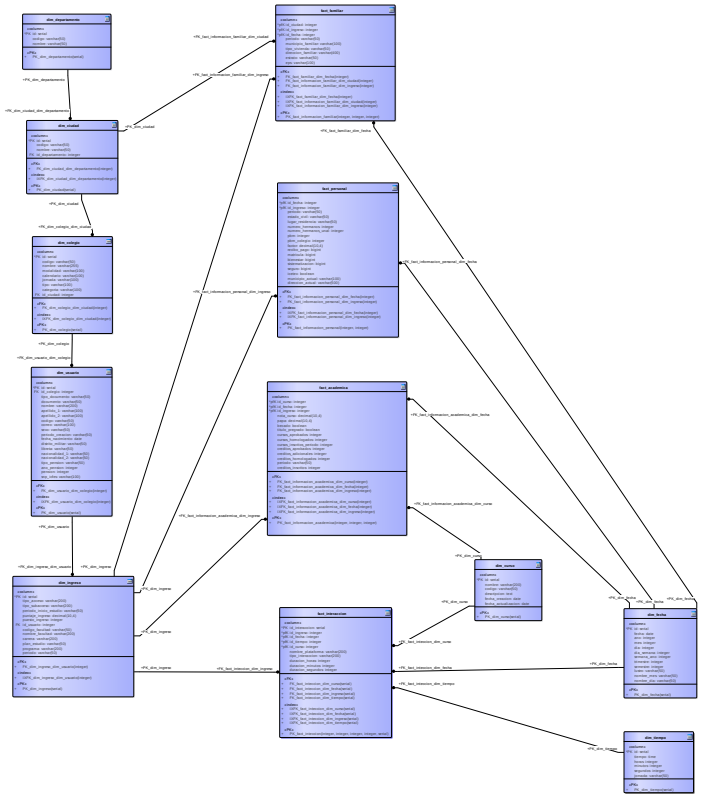
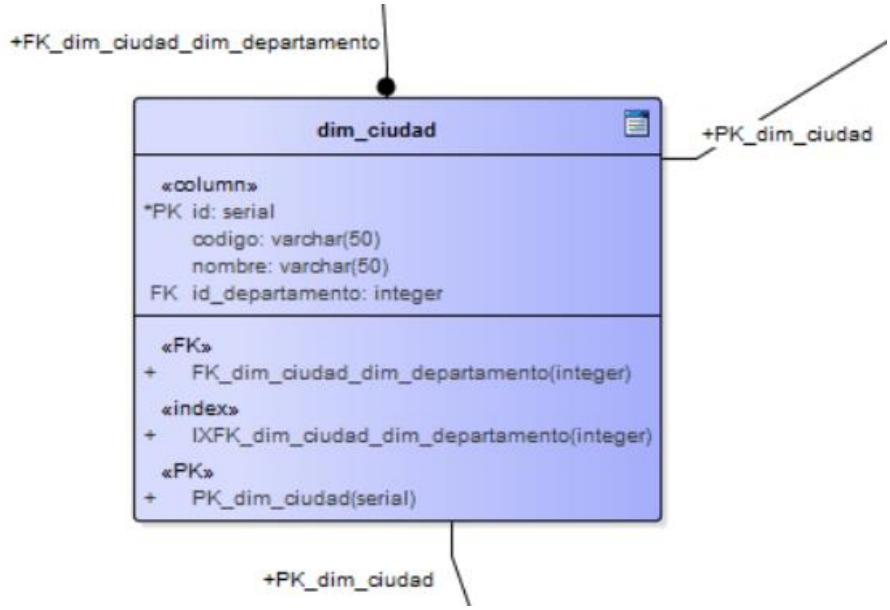


**Figura 6-16:** Modelo de la bodega de datos académica

## dim_ciudad



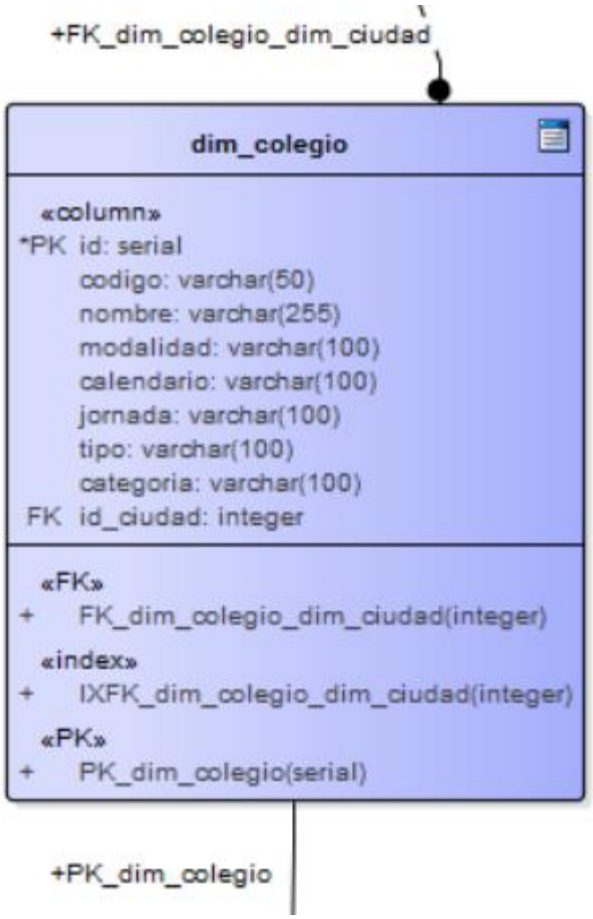

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| codigo | varchar(50) | False | Campo que contiene el código designado para el municipio |
| nombre | varchar(50) | False | Campo que contiene el nombre del municipio |
| id_departamento | integer | False | Campo que contiene el id de relación con la tabla dim_departamento |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_ciudad | id | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_dim_ciudad_dim_departamento | id_departamento | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_dim_ciudad_dim_departamento | id_departamento | public.dim_departamento(id) |

**dim_colegio**




| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| | | | |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| codigo | varchar(50) | False | Campo que contiene el código designado para el colegio |
| nombre | varchar(255) | False | Campo que contiene el nombre del colegio |
| modalidad | varchar(100) | False | Identificación del tipo de modalidad académica: Vacío, No Oficial, Oficial, NV |
| calendario | varchar(100) | False | B – Septiembre/Junio F – A – Febrero/Noviembre NV –. |
| jornada | varchar(100) | False | Identifica la jornada académica de la institución. Los valores pueden ser: Sin Registro Mañana - Completa U Ordinaria Tarde - Noche - NV - Sabatina - Dominical |
| tipo | varchar(100) | False | Son siglas, de las cuales se asume: O – Colegio Oficial P – Colegio Público T – Colegio Técnico N – Colegio Normalista |
| categoria | varchar(100) | False | Se identifica el tipo de plantel del colegio: Plantel Privado - Sin Registro, Plantel Privado - Urbana, Plantel Oficial - Urbana, Plantel Nocturno - Urbana, Oficial Plantel Oficial -, Plantel Oficial - Rural, Otros Planteles, Plantel Privado - , Plantel Privado - Rural, Plantel Oficial , Plantel Nocturno , Vacío, Plantel Nocturno - Rural |
| id_ciudad | integer | False | Campo que contiene el id de relación con la tabla dim_ciudad que identifica el municipio al que pertenece el colegio |

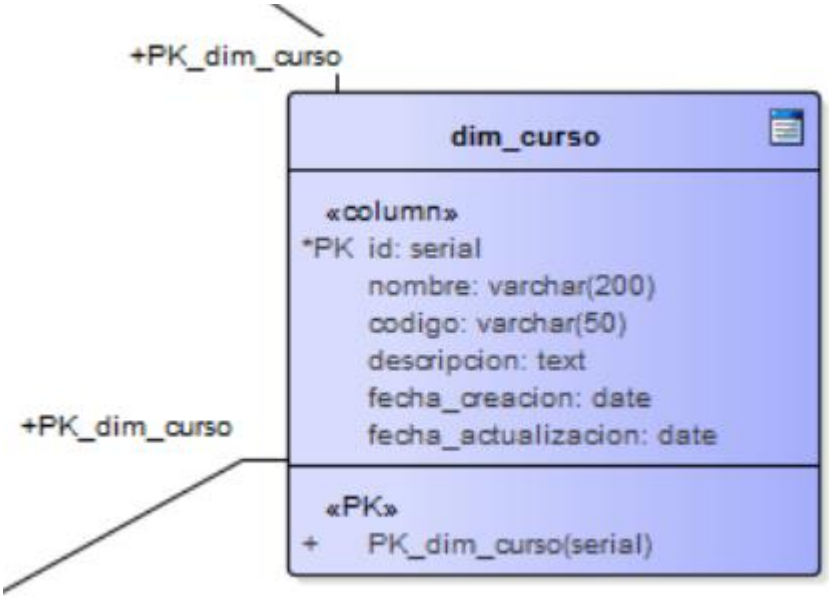
| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_colegio | id | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_dim_colegio_dim_ciudad | id_ciudad | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_dim_colegio_dim_ciudad | id_ciudad | public.dim_ciudad(id) |

**dim_curso**

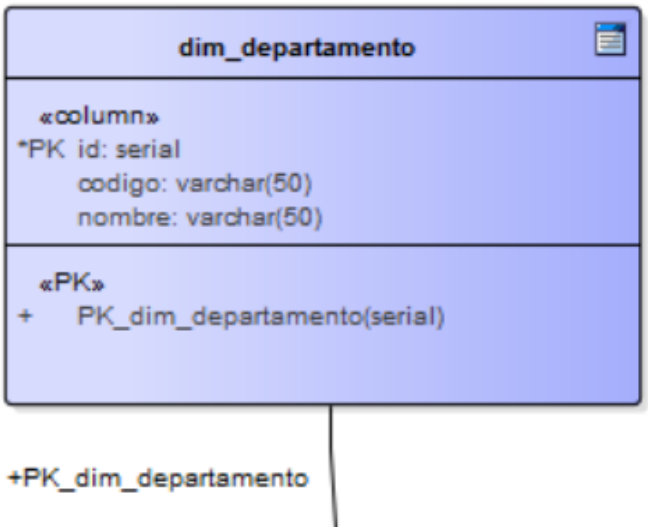



| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |
| nombre | varchar(200) | False | Campo que contiene el nombre del colegio |
| codigo | varchar(50) | False | Campo que contiene el código designado para el colegio |
| descripcion | text | False | |
| fecha_creacion | date | False | Fecha de creación del curso |
| fecha_actualizacion | date | False | |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_curso | id | |

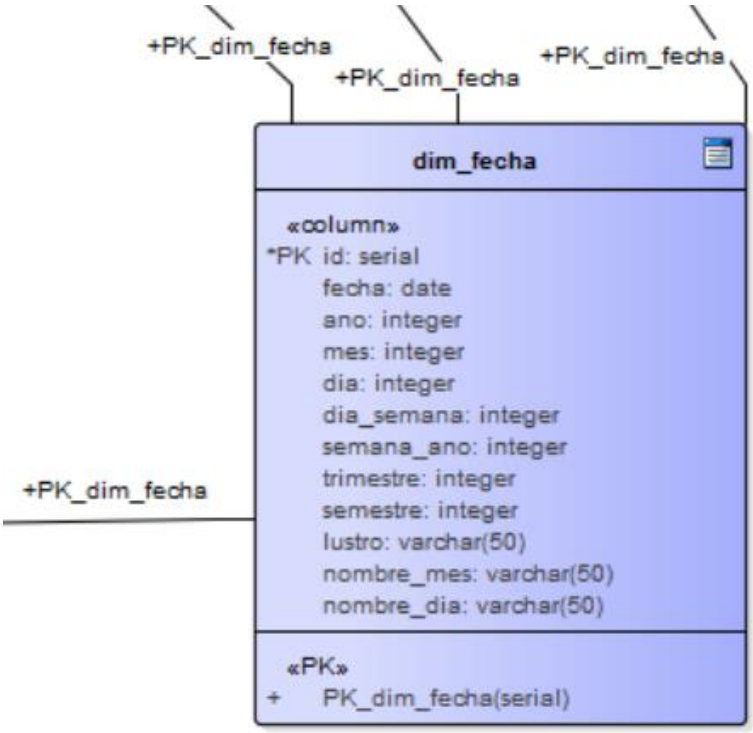
**dim_departamento**



| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| codigo | varchar(50) | False | Campo que contiene el código designado para el departamento |
| nombre | varchar(50) | False | Campo que contiene el nombre del departamento |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_departamento | id | |

**dim_fecha**

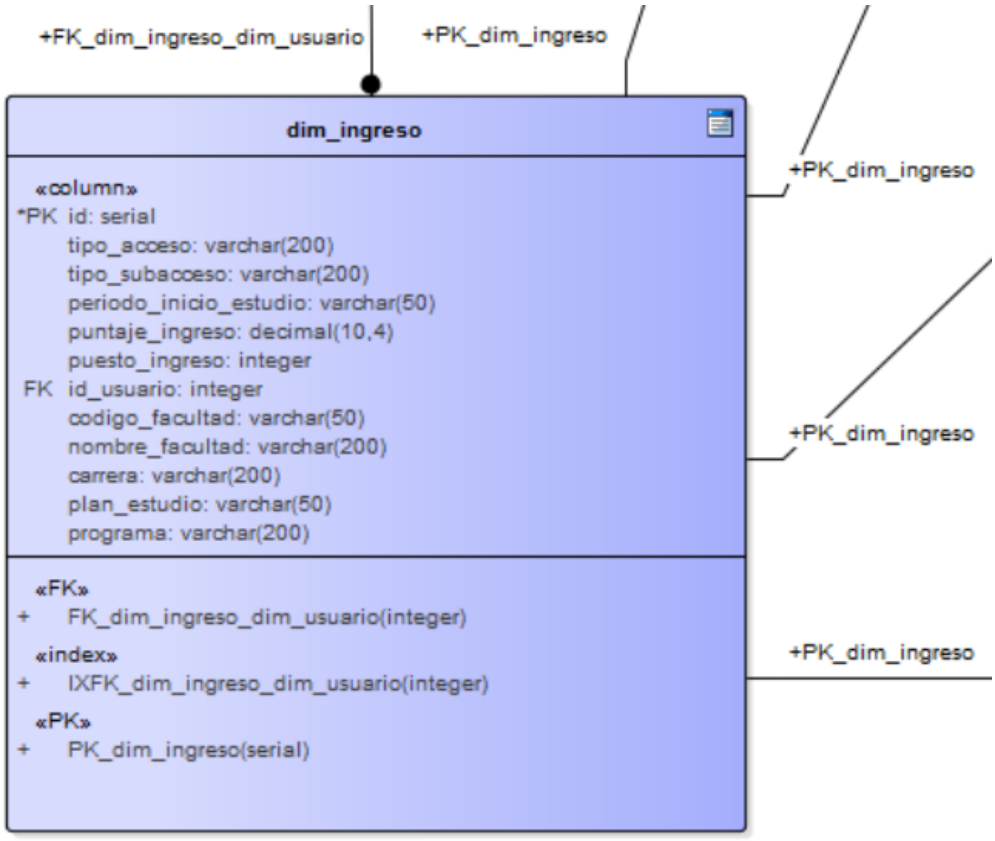


| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| fecha | date | False | |
| ano | integer | False | |
| mes | integer | False | |
| dia | integer | False | |
| dia_semana | integer | False | Número del día contenido en la fecha |
| semana_ano | integer | False | Semana del año al cual pertenece la fecha |
| trimestre | integer | False | Trimestre del año al cual pertenece la fecha |
| semestre | integer | False | Semestre del año al cual pertenece la fecha |
| lustro | varchar(50) | False | Perido de 5 años al cual pertenece la fecha |
| nombre_mes | varchar(50) | False | Nombre del mes de la fecha |
| nombre_dia | varchar(50) | False | Nombre del dia de la fecha |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_fecha | id | |

**dim_ingreso**




| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| tipo_acceso | varchar(200) | False | Identifica el mecanismo por el cual ingresó el estudiante a la universidad |

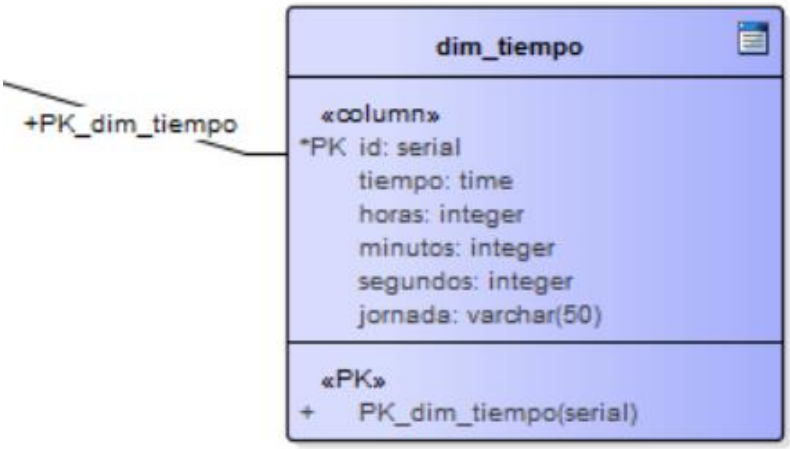
| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| tipo_subacceso | varchar(200) | False | Identifica si el estudiante ingresó por algún mecanismo especial |
| periodo_inicio_estudio | varchar(50) | False | Periodo (año-semestre) de inicio de los estudios en la universidad |
| puntaje_ingreso | decimal(10,4) | False | Valor calificación obtenida en el exámen de admisión |
| puesto_ingreso | integer | False | Posición adquirida según el puntaje de admisión |
| id_usuario | integer | False | Campo que contiene el id de relación con la tabla dim_usuario que identifica a que estudiante pertenece el ingreso |
| codigo_facultad | varchar(50) | False | Identificador institucional |
| nombre_facultad | varchar(200) | False | Nombre de la facultad |
| carrera | varchar(200) | False | Código de la carrera dentro de la facultad |
| plan_estudio | varchar(50) | False | Código del plan educativo |
| programa | varchar(200) | False | Nombre de la carrera |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_ingreso | id | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_dim_ingreso_dim_usuario | id_usuario | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_dim_ingreso_dim_usuario | id_usuario | public.dim_usuario(id) |

## dim_tiempo

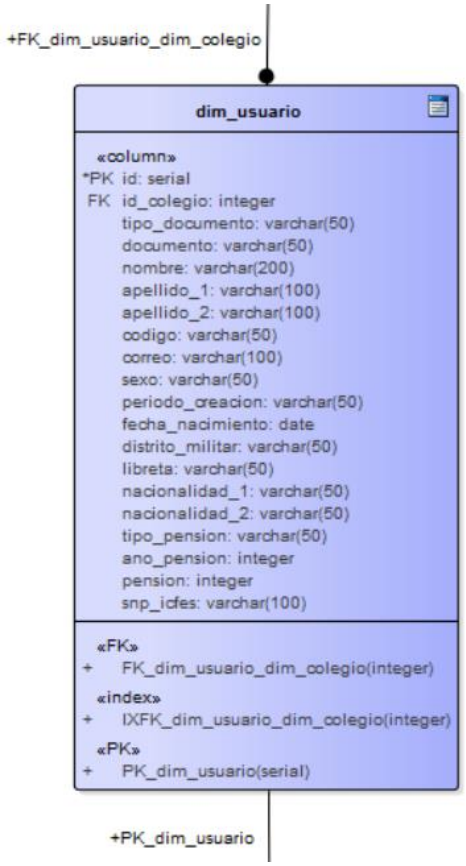



| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| tiempo | time | False | Tiempo |
| horas | integer | False | Horas del tiempo |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |
| minutos | integer | False | Minutos del tiempo |
| segundos | integer | False | Segundos del tiempo |
| jornada | varchar(50) | False | Jornada a la cual pertenece el tiempo (Madrugada, mañana, tarde y noche) |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_tiempo | id | |

**dim_usuario**




| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id | serial | True | Campo que contiene el id consecutivo de generación del registro |
| id_colegio | integer | False | Campo que contiene el id de relación con la tabla dim_colegio que identifica el colegio del cual el estudiante se graduó |
| tipo_documento | varchar(50) | False | Se establecen 4 posibles valores identificados así: E –; O -;T – Tarjeta de identidad; C – Cedula de ciudadanía |
| documento | varchar(50) | False | Contiene el número del documento que identifica al usuario |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| nombre | varchar(200) | False | Nombre completo del estudiante |
| apellido_1 | varchar(100) | False | Primer apellido del estudiante |
| apellido_2 | varchar(100) | False | Segundo apellido del estudiante |
| codigo | varchar(50) | False | Código asignado en la universidad para el estudiante |
| correo | varchar(100) | False | Correo electrónico asignado por la institución educativa |
| sexo | varchar(50) | False | Con dos posibles valores: F – Femenino; M – Masculino |
| periodo_creacion | varchar(50) | False | Periodo (año-semestre) de creación en el sistema académico |
| fecha_nacimiento | date | False | Fecha de nacimiento, separada por tres atributos (Año, mes, día) |
| distrito_militar | varchar(50) | False | |
| libreta | varchar(50) | False | |
| nacionalidad_1 | varchar(50) | False | |
| nacionalidad_2 | varchar(50) | False | |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |
| tipo_pension | varchar(50) | False | Periodicidad con la que pagaba la pensión en el colegio, se toma uno de tres valores: vacío, M - Mensual, A - Anual |
| ano_pension | integer | False | |
| pension | integer | False | |
| snp_icfes | varchar(100) | False | Código que identifica la prueba Saber 11 que presentó el estudiante durante el último año de colegio |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_dim_usuario | id | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_dim_usuario_dim_colegio | id_colegio | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_dim_usuario_dim_colegio | id_colegio | public.dim_colegio(id) |

**fact_academica**

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id_curso | integer | True | Campo que contiene el id de relación con la tabla dim_curso |
| id_fecha | integer | True | Campo que contiene el id de relación con la tabla dim_fecha que identifica el periodo en que se creó el registro |
| id_ingreso | integer | True | Campo que contiene el id de relación con la tabla dim_ingreso |
| nota_curso | decimal(10,4) | False | Nota asociada al curso relacionado |
| papa | decimal(10,4) | False | Promedio académico ponderado calculado en la universidad para evaluar el rendimiento académico |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |
| becado | boolean | False | |
| titulo_pregrado | boolean | False | |
| cursos_aprobados | integer | False | Número de cursos aprobados durante la carrera |
| cursos_homologados | integer | False | Número de cursos homologados durante la carrera |
| cursos_inscritos_periodo | integer | False | Número de cursos inscritos en el periodo de creación del registro |
| creditos_aprobados | integer | False | Número de creditos aprobados durante la carrera |
| creditos_adicionales | integer | False | Número de creditos adicionales |
| creditos_homologados | integer | False | Número de creditos homologados durante la carrera |
| periodo | varchar(50) | False | Identifica el periodo en que se creó el registro |
| creditos_inscritos | integer | False | Número de créditos inscritos en el periodo |

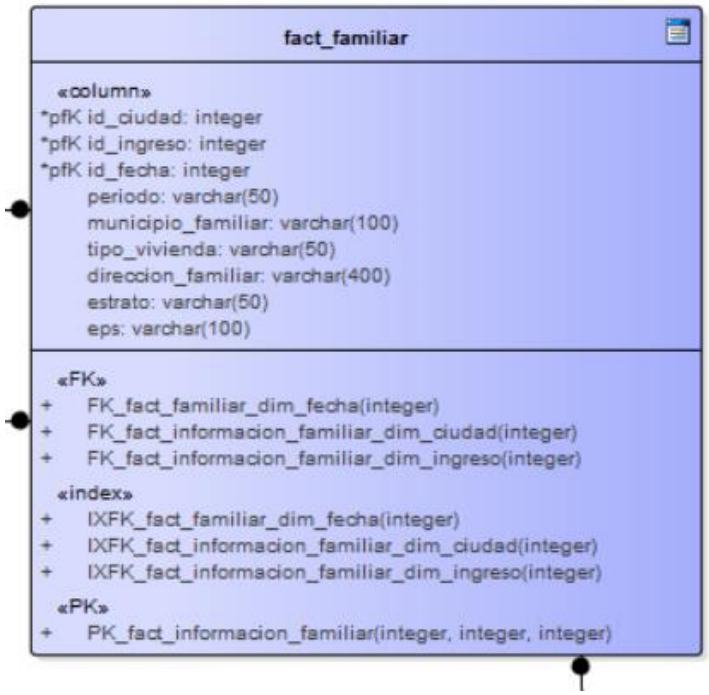
| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|

| | | |
|---|---|---|
| PK_fact_informacion_academica | id_curso, id_fecha, id_ingreso | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_fact_informacion_academica_dim_curso | id_curso | |
| «index» IXFK_fact_informacion_academica_dim_fecha | id_fecha | |
| «index» IXFK_fact_informacion_academica_dim_ingreso | id_ingreso | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_fact_informacion_academica_dim_ingreso | id_ingreso | public.dim_ingreso(id) |
| FK_fact_informacion_academica_dim_fecha | id_fecha | public.dim_fecha(id) |
| FK_fact_informacion_academica_dim_curso | id_curso | public.dim_curso(id) |

**fact_familiar**




| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id_ciudad | integer | True | Campo que contiene el id de relación con la tabla dim_cuidad la procedencia de la familia |
| id_ingreso | integer | True | Campo que contiene el id de relación con la tabla dim_ingreso |
| id_fecha | integer | True | Campo que contiene el id de relación con la tabla dim_fecha que identifica el periodo en que se creó el registro |
| periodo | varchar(50) | False | Identifica el periodo en que se creó el registro |
| municipio_familiar | varchar(100) | False | |
| | | | |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| tipo_vivienda | varchar(50) | False | |
| direccion_familiar | varchar(400) | False | Dirección del lugar de residencia de la familia |
| estrato | varchar(50) | False | Estrato socioeconómico del lugar de residencia de la familia |
| eps | varchar(100) | False | EPS (empresa prestadora de salud) en la que está afiliado el estudiante |

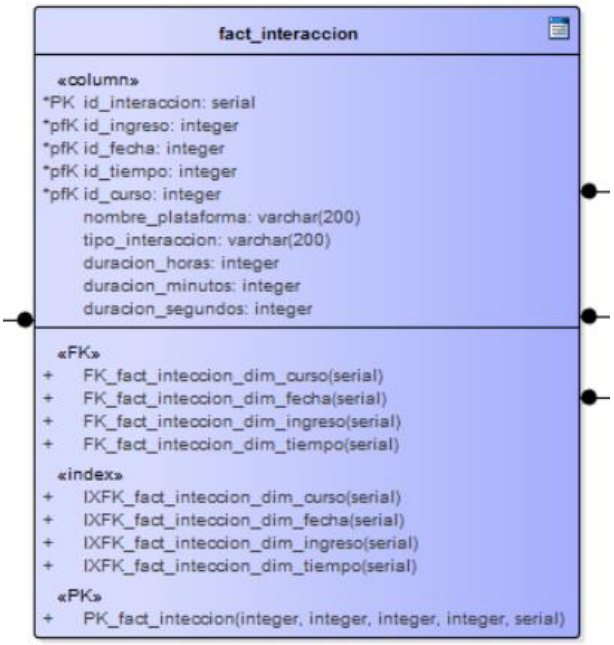
| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_fact_informacion_familiar | id_ciudad, id_ingreso, id_fecha | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_fact_familiar_dim_fecha | id_fecha | |
| «index» IXFK_fact_informacion_familiar_dim_ciudad | id_ciudad | |
| «index» IXFK_fact_informacion_familiar_dim_ingreso | id_ingreso | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_fact_informacion_familiar_dim_ingreso | id_ingreso | public.dim_ingreso(id) |
| FK_fact_informacion_familiar_dim_ciudad | id_ciudad | public.dim_ciudad(id) |

| | | |
|---|---|---|
| FK_fact_familiar_dim_fecha | id_fecha | public.dim_fecha(id) |

## fact_interaccion




| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id_interaccion | serial | True | Consecutivo de creación del registro |
| id_ingreso | integer | True | Campo que contiene el id de relación con la tabla dim_ingreso |
| id_fecha | integer | True | Campo que contiene el id de relación con la tabla dim_fecha que identifica el periodo en que se creó el registro |
| id_tiempo | integer | True | Campo que contiene el id de relación con la tabla dim_tiempo que identifica la hora, munito y segundo de creación del registro |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| | | | |
| id_curso | integer | True | Campo que contiene el id de relación con la tabla dim_curso e identifica con que curso se realizó la interacción |
| nombre_plataforma | varchar(200) | False | Nombre de la plataforma de donde se extrajo el registro de interacción |
| tipo_interaccion | varchar(200) | False | Tipo de interacción realizada con la plataforma |
| duracion_horas | integer | False | Duración en horas de la interacción con la plataforma |
| duracion_minutos | integer | False | Duración en minutos de la interacción con la plataforma |
| duracion_segundos | integer | False | Duración en segundos de la interacción con la plataforma |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_fact_inteccion | id_ingreso,<br>id_fecha,<br>id_tiempo,<br>id_curso,<br>id_interaccion | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index»<br>IXFK_fact_inteccion_dim_curso | id_curso | |
| | id_fecha | |

| | | |
|---|---|---|
| «index» IXFK_fact_inteccion_dim_fecha | | |
| «index» IXFK_fact_inteccion_dim_ingreso | id_ingreso | |
| «index» IXFK_fact_inteccion_dim_tiempo | id_tiempo | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|
| FK_fact_inteccion_dim_ingreso | id_ingreso | public.dim_ingreso(id) |
| FK_fact_inteccion_dim_tiempo | id_tiempo | public.dim_tiempo(id) |
| FK_fact_inteccion_dim_fecha | id_fecha | public.dim_fecha(id) |
| FK_fact_inteccion_dim_curso | id_curso | public.dim_curso(id) |

## fact_personal

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| id_fecha | integer | True | Campo que contiene el id de relación con la tabla dim_fecha que identifica el periodo en que se creó el registro |
| id_ingreso | integer | True | Campo que contiene el id de relación con la tabla dim_ingreso |
| periodo | varchar(50) | False | Identifica el periodo en que se creó el registro |
| estado_civil | varchar(50) | False | Se toma uno de ocho opciones: Casado; Nulo, Soltero, Unión libre, Viudo, Separado, Divorciado, No definido |
| lugar_residencia | varchar(50) | False | Existen cuatro opciones, de las cuales se identifican las siguientes cantidades de registros: Especial, SIN INFO, Dentro, Fuera |
| numero_hermanos | integer | False | |
| numero_hermanos_unal | integer | False | |
| pbm | integer | False | Valor PBM obtenido de análisis datos de económicos de admisión |
| pbm_colegio | integer | False | PBM identificado de la institución educativa donde proviene el estudiante |

| ATRIBUTO | TIPO DE DATO | NO NULL | COMENTARIOS |
|---|---|---|---|
| factor | decimal(10,4) | False | |
| recibo_pago | bigint | False | |
| matricula | bigint | False | |
| bienestar | bigint | False | |
| sistematizacion | bigint | False | |
| seguro | bigint | False | |
| icetex | boolean | False | Préstamo ICETEX |
| municipio_actual | varchar(100) | False | |
| direccion_actual | varchar(500) | False | |

| LLAVE PRIMARIA | COLUMNA | COMENTARIOS |
|---|---|---|
| PK_fact_informacion_personal | id_fecha, id_ingreso | |

| TIPO / NOMBRE | COLUMNA | COMENTARIOS |
|---|---|---|
| «index» IXFK_fact_informacion_personal_dim_fecha | id_fecha | |
| «index» IXFK_fact_informacion_personal_dim_ingreso | id_ingreso | |

| LLAVE FORANEA | COLUMNA | REFERENCIA |
|---|---|---|

| | | |
|---|---|---|
| FK_fact_informacion_personal_dim_fecha | id_fecha | public.dim_fecha(id) |
| FK_fact_informacion_personal_dim_ingreso | id_ingreso | public.dim_ingreso(id) |

# Bibliografía


Araníbar, J. C. (2003). Inteligencia de negocios. *Revista Ciencia y Cultura*, (12), 95-101.

Blackboard Inc. (2015). Blackboard. Recuperado 16 de octubre de 2016, a partir de http://lac.blackboard.com/sites/international/globalmaster/

Boneu, J. M. (2007). Plataformas abiertas de e-learning para el soporte de contenidos educativos abiertos - Dialnet. *Universities and Knowledge Society Journal*, *4*(1), 36-47.

Burgos, D., Tattersall, C., & Koper, R. (2005). Utilización de estándares en el aprendizaje virtual. Presentado en II Jornadas Campus Virtual, Madrid: Universidad Complutense de Madrid. Recuperado a partir de http://dspace.ou.nl/handle/1820/488

Castro, M., & Lizasoain, L. (2012). Las técnicas de modelización estadística en la investigación educativa: minería de datos, modelos de ecuaciones estructurales y modelos jerárquicos lineales. *Revista española de pedagogía*, *70*(251), 131-148.

Clarenc, C. A., Castro, S. M., López de Lenz, C., Moreno, M. E., & Tosco, N. B. (2013). Analizamos 19 plataformas de eLearning: Investigación colaborativa sobre LMS. Presentado en Congreso Virtual Mundial de e-Learning. Recuperado a partir de http://cooperacionib.org/191191138-Analizamos-19-plataformas-de-eLearning-primera-investigacion-academica-colaborativa-mundial.pdf

Dueñas Reyes, M. X. (2010). Minería de datos espaciales en búsqueda de la verdadera información. *Ingenieria y Universidad*, *13*(1). Recuperado a partir de http://revistas.javeriana.edu.co/index.php/iyu/article/view/953

Duque Méndez, N. D., Hernández Leal, E. J., Pérez Zapata, Á. M., Arroyave Tabares, A. F., & Espinosa Gómez, D. A. (2016). Modelo para el proceso de extracción, transformación y carga en bodegas de datos. Una aplicación con datos ambientales. *Ciencia e Ingeniería Neogranadina*, *26*(2), 95-109. https://doi.org/10.18359/rcin.1799

Duque Mendez, N. D., Vélez Upegui, J. J., & Orozco Alzate, M. (2015). Análisis multidimensional de datos ambientales. En *Entendimiento de fenómenos ambientales mediante análisis de datos* (Primera, pp. 115-132). Manizales, Colombia: Universidad Nacional de Colombia -Sede Manizales.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and Implementation of a Learning Analytics Toolkit for Teachers. *Educational Technology & Society*, *15*(3), 58–76.

Feliciano Morales, A., Cuevas Valencia, R. E., & Martínez Castro, J. M. (2016). Procesamiento Analítico con Minería de Datos / Analytical Processing with Data Mining. *RECI Revista Iberoamericana de las Ciencias Computacionales e Informática*, *5*(9), 22-43.

Ferguson, Rebbeca, & Buckingham Shum, S. (2012). Social learning analytics: five approaches (pp. 23-33). Presentado en International Conference on Learning Analytics & Knowledge, Vancouver, British Columbia, Canada. Recuperado a partir de http://oro.open.ac.uk/32910/

Ferguson, Rebecca. (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, *4*(5/6), 304-317.

Fernández-Gallego, B., Lama, M., Vidal, J. C., & Mucientes, M. (2013). Learning Analytics Framework for Educational Virtual Worlds. *Procedia Computer Science*, *25*, 443-447. https://doi.org/10.1016/j.procs.2013.11.056

Fortenbacher, A., Beuster, L., Elkina, M., Kappe, L., Merceron, A., Pursian, A., … Wenzlaff, B. (2013). LeMo: A learning analytics application focussing on user path analysis and interactive visualization. En *2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS)* (Vol. 02, pp. 748-753). https://doi.org/10.1109/IDAACS.2013.6663025

García Peñalvo, F. J. (2005). Estado actual de los sistemas e-learning. *Teoría de la Educación: Educación y Cultura en la Sociedad de la Información*, *6*(2). Recuperado a partir de http://dialnet.unirioja.es/servlet/articulo?codigo=1302878

Global Development Group. (2016). PostgreSQL: The world's most advanced open source database. Recuperado 11 de junio de 2016, a partir de https://www.postgresql.org/

Hira, S., & Deshpande, P. S. (2015). Data Analysis using Multidimensional Modeling, Statistical Analysis and Data Mining on Agriculture Parameters. *Procedia Computer Science*, *54*, 431-439. https://doi.org/10.1016/j.procs.2015.06.050

Innovation Group. (2016). The Future 100. Recuperado a partir de https://www.jwtintelligence.com/trend-reports/future-100-2017/

Jaramillo Valbuena, S., & Londoño, J. M. (2015). Sistemas para almacenar grandes volúmenes de datos. *Revista Gerencia Tecnológica Informática*, *13*(37), 17-28.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC Horizon Report: 2016 Higher Education Edition*. Austin, Texas: The New Media Consortium. Recuperado a partir de http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 Horizon Report* (p. 36). Austin, Texas: The New Media Consortium. Recuperado a partir de http://www.nmc.org/pdf/2011-Horizon-Report.pdf

Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2013). Using Learning Analytics to Identify Successful Learners in a Blended Learning Course. *Int. J. Technol. Enhanc. Learn.*, *5*(2), 133–150. https://doi.org/10.1504/IJTEL.2013.059088

Kumar, S., Shobha, & Pal, S. K. (2015). A new sustainable prototype USP for education information system. En *2015 International Conference on Futuristic Trends on Computational Analysis and Knowledge Management (ABLAZE)* (pp. 174-179). https://doi.org/10.1109/ABLAZE.2015.7154988

Lerís, D., & Sein-Echaluce, M. L. (2011). La personalización del aprendizaje: un objetivo del paradigma educativo centrado en el aprendizaje. *ARBOR Ciencia, Pensamiento y Cultura*, *187*(Extra 3), 123-134. https://doi.org/10.3989/arbor.2011.Extra-3n3135

Long, P. D., & Siemens, G. (2011). Penetrating the Fog: Analytics in Learning and Education. *Educause Review*, *46*(5), 30-32.

MacNeill, S., Campbell, L. M., & Hawksey, M. (2014). Analytics for Education. *Journal of Interactive Media in Education*, 1-12.

Microsoft. (2016). Algoritmos de minería de datos (Analysis Services: Minería de datos). Recuperado 20 de noviembre de 2016, a partir de https://msdn.microsoft.com/es-es/library/ms175595.aspx

Molina, J. M., & García, J. (2006). Técnicas de Minería de Datos basadas en Aprendizaje Automático. En *Técnicas de Análisis de Datos* (pp. 96-157). Madrid. Recuperado a partir de http://www.giaa.inf.uc3m.es/docencia/II/ADatos/apuntesAD.pdf

Moodle Project. (2013). Database schema introduction - MoodleDocs. Recuperado 26 de octubre de 2016, a partir de https://docs.moodle.org/dev/Database_schema_introduction

Moreno García, M. N., Quintales, L., García-Peñalvo, F. J., & Polo Martín, M. J. (2001). Aplicación de Técnicas de Minería de Datos en la Construcción y Validación de Modelos Predictivos y Asociativos a Partir de Especificaciones de Requisitos De Software. Presentado en II ADIS Workshop on Decision Support in Software Engineering, Almagro, Ciudad Real, Spain. Recuperado a partir de http://ceur-ws.org/Vol-84/paper4.pdf

Oxford University Press. (2015). Learn about Virtual Learning Environment/Course Management System content. Recuperado 19 de diciembre de 2015, a partir de http://global.oup.com/uk/orc/learnvle/

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, *41*(4, Part 1), 1432-1462. https://doi.org/10.1016/j.eswa.2013.08.042

Picciano, A. G. (2012). The Evolution of Big Data and Learning Analytics in American Higher Education. *Journal of Asynchronous Learning Networks*, *16*(3), 9-20.

Quiroz Gil, N. L., & Valencia, C. A. (2012). Aplicación del proceso de KDD en el contexto de bibliomining: El caso Elogim. *Revista Interamericana de Bibliotecología*, *35*(1), 97-108.

Ramírez Romero, T. A., Patiño Ortiz, J., & Patiño Ortiz, M. (2015). Comparativa entre Data Warehouse y generación dinámica de consultas en SQL. Presentado en 5° Congreso Nacional de Ingeniería Electromecánica y de Sistemas (CNIES), México, D.F.

Rojas Castro, P. (2015). Paradigmas analíticos en entornos virtuales y de aprendizaje: una revisión de sus principales puntos de encuentros y diferenciaciones teóricas y de enfoque. *Revista educación y tecnología*, (7), 91-106.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *3*(1), 12-27. https://doi.org/10.1002/widm.1075

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Delgado Kloos, C. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform. *Computers in Human Behavior*, *47*, 139-148. https://doi.org/10.1016/j.chb.2014.07.002

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., … Baker, R. (2011, julio). Open Learning Analytics: an integrated & modularized platform. Society for Learning Analytics Research. Recuperado a partir de http://cmapspublic3.ihmc.us/rid=1KC16KK3Y-1DGTX1Y-H2/KG-%20OpenLearningAnalytics.pdf

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-Driven Research to Support Learning and Knowledge Analytics. *Journal of Educational Technology & Society*, *15*(3), 133-148.