ANÁLISIS CORRELACIONAL DEL COMPORTAMIENTO DE ESTUDIANTES EN FUNCIÓN DE LA PARTICIPACIÓN EN UN AVA, HACIENDO USO DE ALGORITMOS KDD

ING. WALBERTO E. MARRUGO ORTEGA
(ESTUDIANTE)

TESIS DE GRADO DE MAESTRIA EN INGENIERIA

ING. OMER SALCEDO
MAGISTER EN INGENIERÍA
(DIRECTOR)

UNIVERSIDAD TECNOLÓGICA DE BOLÍVAR
FACULTAD DE INGENIERÍA, MAESTRÍA EN INGENIERÍA
CARTAGENA DE INDIAS DT Y C.
2014

Agradecimientos

Quiero agradecerle a Dios por brindarme esta oportunidad de promover amor por el conocimiento. Gracias a mis familiares por brindarme su apoyo incondicional. Agradezco por el esfuerzo y tiempo dedicado a mi director de proyecto ing. Omer Salcedo. También quiero agradecer al Director de Registro académico Sr. Hernán Osorio, Jefe Educación Virtual Sra. Elsa Ruiz Ariza, Ing. Erick Torres por colaborar con la información necesaria para la realización de este trabajo.

¡Muchas Gracias a todos!

Resumen

Las Tecnologías de Información y Comunicación (TIC) han cambiado e innovado la forma de trabajar en el sector educativo, como es el caso de los Ambientes Virtuales de Aprendizaje (AVA), sistemas versátiles muy populares e interactivos, pero esta situación hace que los estudiantes tengan una sobrecarga de información que los desorienta en el proceso de aprendizaje, debido a la gran diversidad de recursos didácticos que ofrecen estos AVA.

Una de las problemáticas que se abordan en este trabajo es la implementación del algoritmo de minería de datos k-means, para poder agrupar a los estudiantes de ingeniería de sistema de la Universidad Tecnológica de Bolívar en diferentes grupos, con base en una determinada selección de cursos b-learning, relacionados con las calificaciones finales y las actividades realizadas en el Ambiente Virtual de Aprendizaje SAVIO. El objetivo de este trabajo es analizar patrones de comportamiento de estudiantes en función de su participación en un AVA y su escala de calificación final. La metodología utilizada para llevar a cabo este estudio fue la aplicación del modelo CRISP-DM, una de las metodologías más usadas y pionera en el proceso de descubrimiento de conocimientos en datos (KDD). Entre los resultados obtenidos se tiene un modelo descriptivo de agrupamiento que sirve de indicador para emitir estrategias para la mejora continua del proceso de enseñanza aprendizaje virtual.

Tabla de contenido

Índice de Figuras

Índice de Tablas

Capítulo 1

1. Introducción

La globalización es un fenómeno o proceso mundial que intenta unificar los mercados, tecnologías, sociedades o culturas, este proceso exige que los países promuevan sostenibilidad y calidad en sus productos y mano de obra, en Colombia las instituciones de educación superior tienen como visión o misión promover una educación integral, para que sus egresados cumplan con los estándares que se exigen a nivel mundial, para llevar a cabo esta visión y misión, las instituciones se han apoyado en la implementación de Tecnologías de Información y Comunicación (TIC), que brinden un proceso de enseñanza aprendizaje y evaluación permanente, global, e interactivo, y este es caso de los Ambientes Virtuales de Aprendizaje, sistemas versátiles y flexibles que se pueden adaptar a cualquier proceso académico.

Los AVA son sistemas que ofrecen a sus participantes (profesor - estudiantes) una gran diversidad de recursos didácticos, con el fin de apoyar el proceso de enseñanza aprendizaje en diferentes modalidades e-learning (educación virtual), b-learning (educación semi-virtual), m-learning (educación móvil). Esta gran diversidad de recursos didácticos que ofrecen los AVA, tiene una tendencia a confundir al estudiante en su proceso de aprendizaje.

En los últimos años se están utilizando técnicas de minería de datos para extraer conocimientos de los sitios Web. Algo interesante de esta investigación es implementar algoritmos de minería de datos para entender que es lo que los estudiantes quieren o hacen en los AVA.

Este trabajo consiste en el estudio de los factores que influyen en el proceso de participación de los estudiantes en Ambientes Virtuales de Aprendizaje (AVA), en comparación con sus calificaciones académicas. El escenario problemático contemplado para este estudio son los cursos virtuales con modalidad b-learning de la Universidad Tecnológica de Bolívar (UTB).

Para darle cumplimiento a los objetivos de este trabajo se ha adoptado el modelo CRISP-DM, una de la metodología más usada y la pionera en la extracción de conocimiento en base de datos.

El trabajo se estructura por capítulos de la siguiente forma: en el capítulo 2 presenta el estado del arte de los Ambientes Virtuales de Aprendizajes y la Aplicación de la minería de datos en la Web, en el capítulo 3 se explica el diseño metodológico para llevar a cabo la realización o cumplimiento de cada objetivo, en el capítulo 4 se hace una descripción del caso de estudio y los recursos con los que se cuentan para la investigación, en el capítulo 5 se preparan los datos para la realización del experimento aplicando el algoritmo de minería de datos k-mean, luego se evaluara el modelo obtenido para probar su efectividad, en el capítulo 6 se llevara a cabo dos tipos de análisis con base en los indicadores modelo de agrupamiento obtenido,

seguidamente se emitirán estrategias para el aprovechamiento y buen uso de los AVA, por ultimo las conclusiones, recomendaciones y trabajo futuro.

1.1. Objetivo General

- Analizar los factores que relacionen las calificaciones académicas y las acciones de los estudiantes en los cursos virtuales con modalidad b-learning, implementando algoritmos KDD, que ayuden a la definición de un conjunto de estrategias de apoyo al docente en la mejora continua del proceso de enseñanza aprendizaje virtual.

1.2. Objetivo Especifico

- Identificar variables preliminares que permitan agrupar a los estudiantes de acuerdo a su participación en un Ambiente Virtual de Aprendizaje y la valoración académica.

- Definir el modelo que relacione el comportamiento del estudiante en el uso de los recursos didáctico virtuales con sus respectivas calificaciones, aplicando algoritmos de minería de datos de agrupación, que permitan analizar las evidencias extraídas de los Ambientes Virtuales de Aprendizaje.

- Validar el modelo obtenido con base en el objetivo anterior, haciendo uso de indicadores de desempeño, con el fin de garantizar la efectividad del modelo para posteriores análisis.

- Establecer un conjunto de estrategias que mejoren la participación de los estudiantes en un Ambiente Virtual de Aprendizaje, a partir del análisis del modelo definido con las técnicas de agrupación.

Capítulo 2

2. Estado del Arte

2.1. Ambientes Virtuales de Aprendizaje (AVA)

La implementación de los (AVA) en el proceso educativo ha aumentado, debido a nuevos paradigmas de la educación, como e-learning, b-learning y m-learning, que tienen como propósito la simplificación en el proceso de enseñanza-aprendizaje. Estas nuevos paradigmas de la educación son el factor de motivación de nuevo software encargado de la gestión del trabajo autónomo del estudiante e interacción permanente estudiante-docente, comunicación y trabajo colaborativo.

Los Ambientes Virtuales de Aprendizaje (AVA), también son conocidos como Sistemas de Gestión de Aprendizaje (LMS, Learning Management System), a continuación se describen los sistema de gestión de aprendizaje más usados por las instituciones educativas, como software comercial: TopClass [1] y Blackboard [2]. Y como software libre Dokeos [3], Atutor [4], ILIAS [5], Moodle [6], Claroline [7].

El objetivo de un Sistema de Gestión de Aprendizaje (LMS) es gestionar objetos de aprendizaje, siendo un entorno multiusuario en el que los desarrolladores crean, almacenan, reutilizan, gestionan y distribuyen contenidos a partir de un repositorio central de objetos de aprendizaje [9].

Una de las mayores debilidades de los sistemas de gestión aprendizaje es que proporcionan contenidos de forma estática, a través de los cuales navegan los estudiantes sin importar sus competencias, necesidades y motivaciones. [10] Para contrarrestar estos inconvenientes han surgido en la última década los sistemas hipermedia adaptativos con fines educativos que hacen uso de técnicas de KDD (Knowledge Discovery in Databases) para crear un modelo que permite adaptar el contenido y enlaces del curso al usuario actual. Ejemplos de estos sistemas de hipermedia adaptativos basados en web tenemos: Interbook [11], Elm-Art [12], Aha [13], Indesach [14].

MOODLE Module Object-Oriented Dynamic Learning Environment (Entorno Modular de Aprendizaje Dinámico Orientado a Objetos) está considerado actualmente como uno de los sistemas de gestión de contenido de mayor uso por su adaptabilidad, modularidad, seguridad, flexibilidad y por ser un producto de software libre.

Moodle almacena información detallada sobre la actividad que el estudiante realiza en la plataforma. Esta información es alojada en diferentes tablas de la base de datos de Moodle, y está relacionada con los materiales utilizados y la navegación que hace el estudiante dentro de la plataforma. La herramienta permite el filtrado de estos registros por curso, participante, fecha o actividad específica [14].

La interacción de actores (administrador, docente o estudiante) con Moodle genera gran cantidad de datos, ya sea por la asignación o desarrollo de materiales y actividades en el curso. Moodle proporciona ciertos módulos de estadística para el docente obtener información sobre las actividades llevadas cabo por los estudiantes ya sea de forma individual o grupal, pero resulta difícil obtener información más detallada para determinar patrones de comportamientos en los estudiantes en el desarrollo de cualquier actividad. A nivel de módulo de administración es posible obtener información acerca de registro de actividades por cursos, participante y por fechas. A nivel de sitio Moodle proporciona una estadística global sobre los accesos, pero presenta la dificultad de que estos datos se borran cada cierto tiempo [15].

El trabajo realizado por Marín, Ramírez y Sampedro (2011), tiene como propósito el estudio de la actitud del estudiante con un Ambiente Virtual de Aprendizaje AVA, entre los resultados obtenidos por este trabajo tenemos: que los estudiantes tienden a confundirse o desorientarse en el proceso de aprendizaje, por la variedad de materiales didácticos que establecen dichos AVA [16].

2.2. La Modalidad Blended-Learning (B-Learning)

El aprendizaje combinado (mixto o bimodal), consiste en un proceso que combina una modalidad de enseñanza aprendizaje presencial con una modalidad de enseñanza aprendizaje virtual. B-Learning trata sobre un modelo hibrido a través del cual los tutores pueden hacer uso de sus metodologías de aula para una sesión presencial y al mismo tiempo potenciar el desarrollo de temáticas a través de una plataforma virtual [42].

El trabajo de Fernando Vera (2014). Tiene como interés describir las ventajas de las tecnologías B-Learning en las universidades, entre las conclusiones a las que llega están: La incorporación del aprendizaje combinado o blended learning brindan la posibilidad a las universidades de extender la oferta educativa, mejorar la interacción entre los miembros de una comunidad, y se aumenta la motivación intrínseca de los estudiantes [42].

2.3. Proceso de descubrimiento en base de datos (KDD)

El KDD un proceso iterativo e interactivo. Es iterativo ya que la salida de algunas fases puede hacer volver a pasos anteriores y porque a menudo es necesario varias iteraciones para extraer conocimiento de calidad. Es interactivo porque el usuario o experto en el dominio del problema debe ayudar en la preparación de los datos y validación del conocimiento extraído. El proceso KDD se divide en cinco fases como se describe a continuación: la **fase de integración y recopilación de datos** se determina las fuentes de información utilices. En la **fase de selección, limpieza y transformación** se le da soporte a los datos que contienen valores erróneos o faltantes. La **fase de minería de datos,** se decide cual es la tarea a realizar (calificar, Agrupar, etc) y se decide el modelo a utilizar. **Fase de evaluación e interpretación** se evalúan patrones y se analizan por expertos. Finalmente en la **fase de difusión** se hace uso del nuevo conocimiento [17].

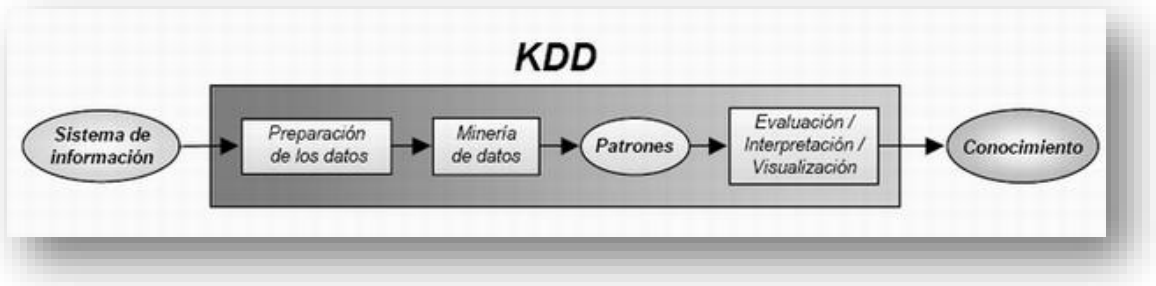

Figura 2.1: Proceso de descubrimiento en base de datos.[17].

2.4. Minería de Datos

La Minería de Datos (MD) es considerad la fase más característica del proceso KDD, y por esta razón, muchas veces se utiliza esta fase para nombrar todo el proceso. El objetivo de la MD es producir nuevo conocimiento que pueda utilizar el usuario. Esto se realiza construyendo un modelo basado en los datos recopilados para este efecto. El modelo es una descripción de los patrones y relaciones entre los datos que pueden usarse para hacer predicciones, para entender mejor los datos o para explicar situaciones pasadas [17].

2.5. Herramientas De Minería de Datos

Existe una gran variedad de herramientas de minería de datos tanto comerciales como no comerciales [18]. A continuación los tipos de herramientas: De las comerciales se destacan: DBMiner [19], SPSS Clementine[20], DB2 Intelligent Miner [21], SAS Enterprise Miner [22], Statistica Data Miner [23] y de las libres: Keel [24], Weka [25], Orange [26], RapidMiner [27], JHepWork [28], Knime [29], entre otras.

2.6. Web Usage Mining

Son herramientas que están siendo aplicadas desde hace varios años, debido a que se especializan en la personalización de sitios Web [17], la Minería del Uso de la Web captura las actividades de los usuarios durante su conexión y extrae patrones de comportamiento que pueden ayudar a comprender las preferencias de navegación de los usuarios.

En el trabajo realizado por [10] proponen un sistema que tiene como propósito detectar posibles problemas de diseño, estructura y contenido de un curso con base en los datos que arrojan la interacción con el curso por parte de los estudiantes.

En el artículo de [14], se propone el estudio de la aplicación de Minería de Datos en un Ambiente Virtual Aprendizaje, logrando como resultado la definición de un modelo que clasifica al usuario en malo o bueno según las acciones en la plataforma de aprendizaje Moodle.

En el artículo de Cristóbal Romero *, Sebastián Ventura, Enrique García, proponen técnicas de minería de datos tales como clustering para el análisis de los datos extraídos del sistema de gestión de aprendizaje Moodle, entre los resultados obtenidos esta un modelo que agrupa a los estudiantes en diferentes grupos, relacionado con las actividades realizadas en el sistema Moodle [14].

La Metodología para la Mejora Continua de Cursos de E-learning (CIECoM, Continuous Improvement of E-learning Courses Methodology) diseñada por [10] consta de 3 etapas: construcción del curso, ejecución del curso y mejora del curso, esta última es una fase de retroalimentación y busca el mantenimiento del curso con base en las bitácoras o log de uso del curso por parte de los estudiantes.

Entre otros trabajos que proponen un enfoque de aplicación Web Usage Mining (WUM) tenemos el de [30], donde se obtiene como resultado el uso de la plataforma virtual clasificada por género y programa, el uso diario de acceso a algunas páginas, entre otros. Entre las conclusiones a las que llega este trabajo es la identificación de patrones de comportamientos de estudiantes de la universidad de la India, tales como el uso más frecuente de las aplicaciones de la institución en periodos de exámenes finales.

2.7. Minería de Datos Educativa (EDM)

La minería de datos educativa (Educational Data Mining, EDM) consiste en la aplicación de técnicas de minería de datos para la exploración, visualización y análisis de información generada en los Ambientes Virtuales de Aprendizaje y datan de hace unos años [15].

La disciplina de aplicación de técnicas de minería de datos en entornos educacionales es conocida como Educational DataMining (EDM), y existe un grupo internacional que se dedica a esta área en concreto, The International Working Group on Educational DataMining (http://www.educationaldatamining.org/). Su objetivo es dar soporte a la colaboración y el desarrollo científico en esta nueva disciplina, a través de la organización de jornadas de trabajo y listas de correo, así como con el desarrollo de recursos para compartir datos y técnicas en esta área [31].

El proceso de la Minería de Datos Educacional convierte a los datos adquiridos desde un AVA en información valiosa que puede causar gran impacto en la práctica e investigación educativa. El conocimiento a extraer podría focalizarse en el estudiante, docente o la institución, dependiendo del tipo de información que se dispone (previa o posterior al curso) y la modalidad educativo aplicada (presencial, a distancia o mixto) [32]. Entre las técnicas

más utilizadas de minería de datos educativa (MDE) tenemos: Reglas de Asociación, Agrupamiento o clustering, Arboles de Decisión, Modelos Estadísticos, Regresión Lineal, Redes Neuronales, Algoritmos Genéticos entre otras.

2.8. El Modelo CRISP-DM

CRISP-DM (Cross-Industry Stándard Process for Data Mining) es un consorcio de empresas (inicialmente bajo una subvención inicial de la comisión Europea), incluyendo SPSS, NCR y DaimlerChrysler. La difusión de este estándar ha sido altísima y al, ser independiente de la plataforma o herramienta, está siendo utilizada por cientos de organizaciones en todo el mundo. El estándar 1.0 incluye un modelo de referencia y una guía para llevar a cabo un proyecto de minería de datos [17]. La guía puede ser muy útil como referencia a la hora de establecer una formulación o planificación de un programa de minería de datos adaptado a las necesidades de una organización. El modelo y la guía se estructuran en seis fases principales, como se muestra a continuación:

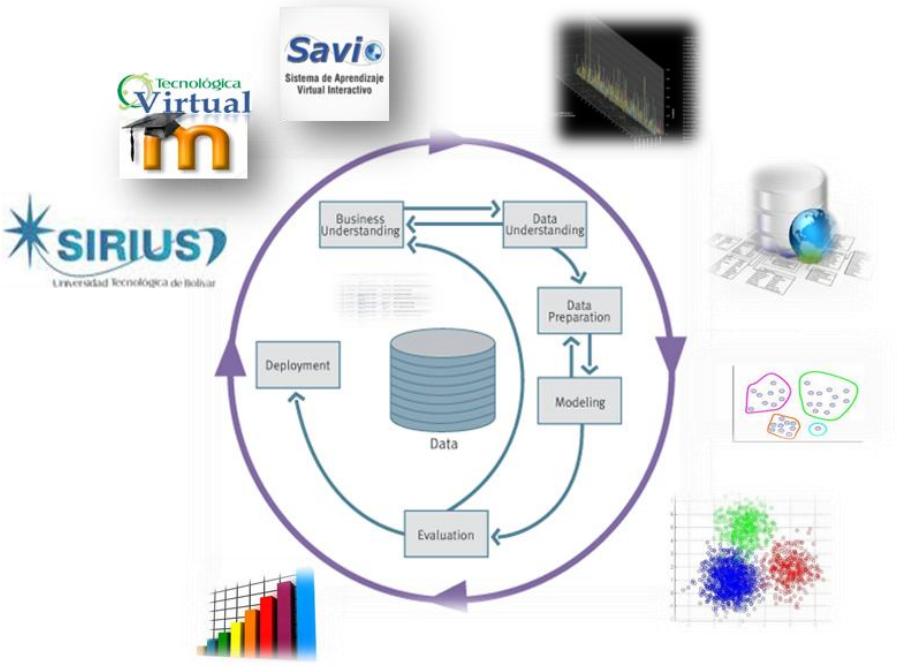

Figura 2.2 Modelo CRISP-DM. Tomado de http://www.rcim.sld.cu/revista_18/articulos_htm/prediccionpaciente.htm, http://www.emagister.com.co/especializacion-gerencia-instituciones-educativas-cursos-2610229.htm, http://letsknowaboutcomputer.blogspot.com/2012/02/mac

Como se puede observar existe realimentación bidireccional entre algunas de las fases, en otras palabras algunas fases pueden obligar a revisar parcial o totalmente otras fases.

Capítulo 3

3. Diseño Metodológico

3.1. Aplicación del Modelo CRISP-DM

En nuestra investigación hacemos uso del modelo CRISP-DM [17], una de las metodologías más usadas y pioneras en la orientación para el descubrimiento de conocimientos en datos. Esta metodología se divide en seis fases: fase conocimiento del negocio, fase de conocimiento de los datos, fase de preparación de los datos, fase modelado, fase evaluación, fase implementación. A continuación se describe cada una de las fases del modelo que permitirán dar cumplimiento a los objetivos del trabajo:

- Fase 1: Conocimiento del Negocio: en la primera fase nos centraremos en los requerimientos del negocio y la evaluación de las herramientas, se debe cumplir las siguientes tareas:

    1. Descripción del problema
    2. Descripción del sistema que utiliza la institución para el proceso de enseñanza aprendizaje y evaluación de estudiante.
    3. Descripción del sistema de gestión de base de datos que utilizan los sistemas antes mencionados.

- Fase 2: Conocimiento de los Datos: en esta fase tratamos de recopilar y familiarizarnos con los datos, a través de esta fase se le da cumplimiento a las siguientes tareas:

    1. Descripción de los periodos de los datos que se van a utilizar.
    2. Selección de los programas académicos y las respectivas asignaturas que se utilizara para el estudio.
    3. Describir e implementar herramientas de diseño para el estudio de los atributos, valores y el comportamiento de los mismos.

- Fase 3: Preparación de los Datos: el objetivo de esta fase es obtener el dataset "Vista minable", aquí se incluye la integración, selección, limpieza y transformación de los datos o evidencias que serán utilizados para el proceso de modelado.

- Fase 4: Modelado: en este espacio se aplicaran técnicas de minería de datos, consta de las siguientes actividades:

    1. Seleccionar una técnicas de modelaje
    2. Generar el diseño de pruebas
    3. Construir, verificar y ajustar el modelo

- Fase 5: Evaluación: en esta fase se evalúa el modelo obtenido con las técnicas de minería de datos en la fase anterior y se revisa la construcción, con fin de comprobar si cumple con los requerimientos del negocio.

- Fase 6: Implementación o Despliegue: en la última fase tratamos de explorar la potencialidad del modelo, presentar el conocimiento obtenido para que el usuario lo pueda utilizar. Otras de las tareas es revisar el proyecto con el fin de mirar debilidades o éxitos para el uso de futuros trabajos.

Observamos que cada una de estas seis fases del modelo CRISP-DM agrupa un conjunto de actividades que dieron cumplimento a los objetivos específicos de este trabajo. La fase 1 conocimiento del negocio y la fase 2 conocimientos de los datos permitieron dar cumplimiento al primer objetivo específico. La fase 3 preparaciones de los datos y fase 4 modelado permitieron dar cumplimento al segundo objetivo específico. La fase 5 evaluación permitió dar cumplimiento al tercer objetivo específico y por último la fase 6 implementación o despliegue le dio cumplimiento al cuarto objetivo específico.

Capítulo 4

4. Conocimiento del Negocio y Conocimiento de los Datos

4.1. Introducción

En el cuarto capítulo se hará una descripción de los sistemas de gestión de aprendizaje y evaluación de estudiantes que implementa la Universidad Tecnológica de Bolívar dándole cumplimiento al primer objetivo específico.

Este trabajo consiste en el estudio de los factores que influyen en el proceso de participación de los estudiantes en Ambientes Virtuales de Aprendizaje (AVA) en comparación con sus calificaciones académicas, resaltando artículos donde se haya trabajado en el fortalecimiento de los AVA. La intención de este de este trabajo es la identificación y descripción de los factores pertinentes al comportamiento del estudiante en los AVA.

4.2. Conocimiento del Negocio

El tipo de investigación que se llevó acabo en este capítulo, fue de tipo documental, basada en artículos o ensayos de revistas.

4.2.1. Descripción del caso de Estudio

Los AVA son medios versátiles que se adaptan a cualquier proceso de enseñanza aprendizaje, poseen una gran variedad de módulos o herramientas didácticas, pero tiene como debilidad la tendencia de sobrecargar de información al estudiante, haciendo que se desoriente en su proceso de aprendizaje según [16]. El problema que se intenta abordar en este trabajo es saber que es los que hacen y quieren los estudiantes de ingeniería de sistema de la Universidad Tecnológica de Bolívar (UTB) en función de la participación del Ambiente Virtual de Aprendizaje SAVIO y su escala de calificación final, esto permitirá personalizar la información cumpliendo con las expectativas o motivaciones de los estudiantes.

4.2.2. Sistema de Aprendizaje Virtual Interactivo (SAVIO)

Sistema de Aprendizaje Virtual Interactivo (SAVIO), es un AVA que ofrece múltiples servicios para el desarrollo de programas virtuales, semi-virtuales y apoyados en las TIC. Esta plataforma es administrada por la Dirección de Educación Virtual (DEV) de la UTB. Entre los recursos didácticos ofrecidos por la plataforma SAVIO, se encuentra: los foros, el chat y correo electrónico, exámenes, cuestionarios, blog, etc. De esta manera la Universidad ofrece una plataforma con amplia cobertura para soportar importantes programas académicos, en los cuales participan docentes y estudiantes de manera interactiva y sin barreras de tiempo y espacio. La plataforma SAVIO está basada en el Sistema de gestión de Aprendizaje Moodle [33].

4.2.3. Sistema información institucional (SIRIUS)

La UTB para complementar su proceso de evaluación hace uso del sistema información institucional SIRIUS [34], el cual ofrece entre otros servicios la valoración de estudiantes, el sistema puede ser usado por docentes y estudiantes vía Web. El sistema SIRIUS aplica concesiones de una, dos o tres fechas de cortes para que todos los docentes carguen sus notas en cada una de las fechas de cortes establecidas en el periodo o semestre académico de acuerdo al "Articulo 75, Reglamento académico estudiantil pregrado de la Universidad Tecnológica de Bolívar". El sistema SIRIUS procesa notas o calificaciones de tipo numérico de dos cifras, expresadas en unidades y décimas, en una escala de valoración de 0 a 5. Cada calificación tiene una equivalencia conceptual según "Parágrafo 1 del artículo 90 del reglamento estudiantil de pregrado Nov. 2002", [35] de acuerdo a la siguiente tabla:

Tabla 4-1: Escala de Calificación. Tomado de Reglamento académico estudiantil pregrado de la Universidad Tecnológica de Bolívar

0.0	a	0.9	Insuficiente
1.0	a	1.9	Muy deficiente
2.0	a	2.9	Deficiente
3.0	a	3.4	Aceptable
3.5	a	4.0	Bueno
4.1	a	4.5	Muy bueno
4.6	a	5.0	Sobresaliente

4.2.4. Características de Moodle

La UTB en su sistema de aprendizaje virtual interactivo (SAVIO) hace uso de la plataforma MOODLE en su versión 2.3. MOODLE (www.moodle.org) es un sistema de gestión de cursos, Inglés Course Management System (CMS) o Sistema de Gestión de Aprendizaje (LMS). Moodle puede ser utilizado en los cursos completamente en línea o servir de complemento a los cursos presenciales. En MOODLE se pueden crear una variedad de cursos, administrados por uno o varios docentes, para el seguimiento a distancia de sus estudiantes. Otros de los potenciales de sistema MOODLE es la creación de "objetos de aprendizaje" o "unidades didácticas" promoviendo el auto aprendizaje y el aprendizaje cooperativo [36][37].

La plataforma SAVIO cuenta con un conjunto de recursos didácticos, que apoyan al docente y estudiante en el proceso de enseñanza aprendizaje (Manual de usuario de SAVIO).

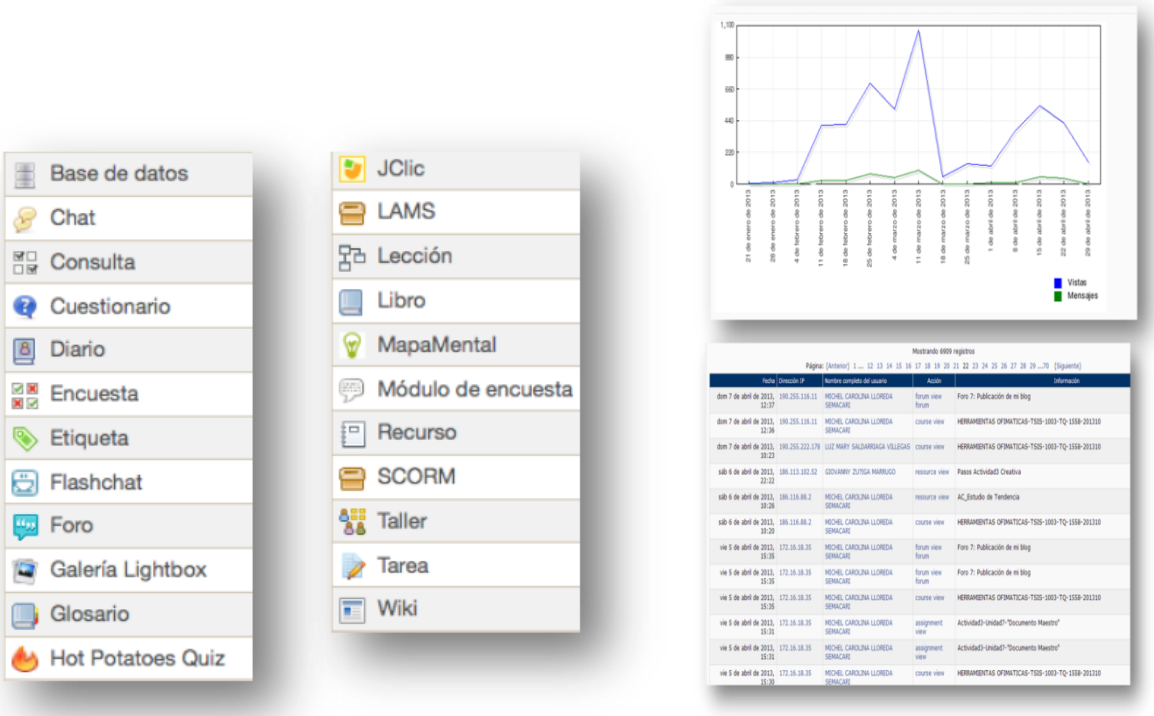

Figura 4.1 Herramientas de Moodle. Tomado de http://www.unitecnologica.edu.co/descargas

Algunas de las características de MOODLE en su versión 2.3 según [6], a continuación:

Tabla 4-2: Características de Moodle. Tomado de http://www.unitecnologica.edu.co/descargas

ASPECTO	**CARACTERÍSTICA**
Diseño general	Promueve una pedagogía constructivista social (colaboración, actividades, reflexión crítica, etc.). Apropiada para el 100% de las clases en línea, así como también para complementar el aprendizaje presencial. Es fácil de instalar en casi cualquier plataforma que soporte PHP. Sólo requiere que exista una base de datos (y la puede compartir).
Administración del sitio	El sitio es administrado por un usuario administrador, definido durante la instalación. Los paquetes de idiomas permiten una localización completa de cualquier idioma. Actualmente hay paquetes de idiomas para 70 idiomas.
Administración de usuarios	Los objetivos son reducir al mínimo el trabajo del administrador, manteniendo una alta seguridad. Base de datos externa: Cualquier base de datos que contenga al menos dos campos puede usarse como fuente externa de autenticación. Cada usuario puede elegir el idioma que se usará en la interfaz de Moodle (Inglés, Francés, Alemán, Español, Portugués, etc.).
Administración de cursos	Se puede elegir entre varios formatos de curso tales como semanal, por temas o el formato social, basado en debates. Ofrece una serie flexible de actividades para los cursos: foros, glosarios, cuestionarios, recursos, consultas, encuestas, tareas, chats y talleres. Todas las calificaciones para los foros, cuestionarios y tareas pueden verse en una única página (y descargarse como un archivo con formato de hoja de cálculo). Registro y seguimiento completo de los accesos del usuario. Se dispone de informes de actividad de cada estudiante, con gráficos y detalles sobre su paso por cada módulo (último acceso, número de veces que lo ha leído) así como también de una detallada "historia" de la participación de cada estudiante,

	incluyendo mensajes enviados, entradas en el glosario, etc. en una sola página.
Módulo de Tareas	Puede especificarse la fecha final de entrega de una tarea y la calificación máxima que se le podrá asignar. Los estudiantes pueden subir sus tareas (en cualquier formato de archivo) al servidor. Se registra la fecha en que se han subido.
Módulo de Chat	Permite una interacción fluida mediante texto síncrono. Todas las sesiones quedan registradas para verlas posteriormente, y pueden ponerse a disposición de los estudiantes
Módulo Foro	Hay diferentes tipos de foros disponibles: exclusivos para los profesores, de noticias del curso y abiertos a todos. Las discusiones pueden verse anidadas, por rama, o presentar los mensajes más antiguos o el más nuevo primero. Si se usan las calificaciones de los foros, pueden restringirse a un rango de fechas. Tipos de foros: • Debate sencillo. Es simplemente un intercambio de ideas sobre un solo tema, todo en un página. Útil para debates cortos y muy concretos. • El foro Normal, para uso general:. Es un foro abierto donde cualquiera puede empezar un nuevo tema de debate cuando quiera. Este es el foro más adecuado para uso general.
Módulo Cuestionario	Los profesores pueden definir una base de datos de preguntas que podrán ser reutilizadas en diferentes cuestionarios. Los cuestionarios se califican automáticamente, y pueden ser recalificados si se modifican las preguntas. Las preguntas y las respuestas de los cuestionarios pueden ser mezcladas (aleatoriamente) para disminuir

	las copias entre los alumnos.
Módulo Recurso	Admite la presentación de cualquier contenido digital, Word, Powerpoint, Flash, vídeo, sonidos, etc. Los archivos pueden subirse y manejarse en el servidor, o pueden ser creados sobre la marcha usando formularios web (de texto o HTML). Se pueden enlazar contenidos externos en web o incluirlos perfectamente en la interfaz del curso.
Módulo de Consulta	Es como una votación. Puede usarse para votar sobre algo o para recibir una respuesta de cada estudiante (por ejemplo, para pedir su consentimiento para algo). El profesor puede ver una tabla que presenta de forma intuitiva la información sobre quién ha elegido qué. Para realizar una encuesta rápida que estimule a los alumnos a reflexionar sobre un tema.
Módulo Taller	Permite la evaluación de documentos entre iguales, y el profesor puede gestionar y calificar la evaluación. Admite un amplio rango de escalas de calificación posibles.

4.3. Conocimiento de los Datos

La administración o gestión de cursos es una de las principales fortalezas de la plataforma MOODLE, permite entre otras cosas, el reporte de los registros o logs de las actividades de los participantes (docentes – estudiantes), definidos a través de atributos demográficos tales como: fecha tiempo, dirección ip, nombre completo de usuario, acción. A continuación se describe los diferentes tipos de valores que pueden tomar el atributo Acción, a partir de la interacción de los participantes con los diferentes módulos establecidos en MOODLE [38].

4.3.1. Tipos de acciones definidas en MOODLE

Tabla 4-3 : Tipos de acciones en MOODLE. Tomado de http://www.upcomillas.es/sifopluspaquetes/Manual%20Completo%20Profesor-%20Moodle.pdf

HERRAMIENTA	***ACCION***	***DESCRIPCIÓN***
COURSE	VIEW	Número de accesos al curso seleccionado
ASSIGNMENT	VIEW	Número de veces que el usuario accede a las tareas enviadas por el profesor.
FORUM	ADD_POST	Número de veces que el usuario envía un POST, a un foro es decir un mensaje que da respuesta a alguna pregunta que realizo el profesor.
FORUM	VIEW_DISCUSSION	Número de veces que el usuario revisa las discusiones dentro de un foro de uso general o Normal.
FORUM	VIEW_FORUM	Número de veces que el usuario revisa o accede a los foros de discusión de debate sencillo.
FORUM	UPDATE_POST	Número de veces que el usuario actualiza el POST que se envió al foro.
FORUM	ADD_ DISCUSSION	Número de veces que el usuario agrega un tema de discusión o debate.
RESOURCE	VIEW	Número de veces que el usuario accede descarga recursos subidos por el profesor.
USER	UPDATE	Número de veces que el usuario actualiza los datos de su perfil.
USER	VIEW	Número de veces que el usuario revisa o accede a su perfil.
UPLOAD	UPLOAD	Número de veces que el usuario sube una tarea con archivos adjuntos.
QUIZ	VIEW	Número de veces que el usuario revisa o accede a los

		cuestionarios.
ASSIGNMENT	UPLOAD	Número de veces que el usuario sube una tarea.
FÓRUM	VIEW FORUMS	Número de veces que el usuario revisa o accede al listado de foros.
FÓRUM	USER REPORT	Número de veces que el usuario revisa o accede a su listado de mensajes publicados en foro

4.3.2. Esquema de Datos de Moodle

La arquitectura de MOODLE está compuesta a nivel de vista o interfaz por lenguaje PHP, a nivel de modelo por los motores de Base de Datos MYSQL, POSTGRE, ORACLE. El módulo de reporte de los registros de actividades de los participantes (estudiantes – docentes), hace uso de las siguientes tablas del esquema de datos de MOODLE:

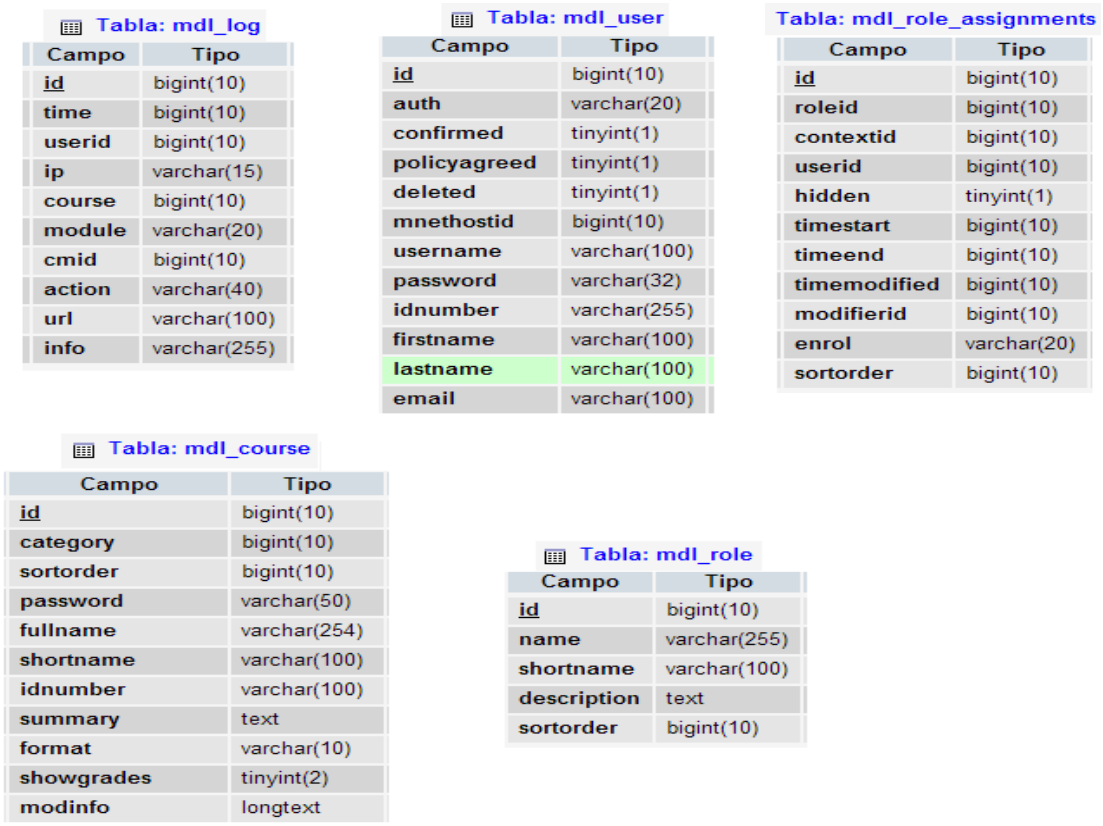

Figura 4.2 Esquema de datos del módulo de reporte de los registros de actividades de AVA MOODLE

Una posible sentencia SQL para obtener los registros de las acciones de los estudiantes en el AVA MOODLE [8] puede ser:

```
SELECT
   u.id AS userid, u.username,
   CONCAT(u.lastname ,' ', u.firstname ) as nombre,
   l.action, COUNT( l.userid ) AS conteo_acciones , r.name,
   l.ip, l.info , c.fullname as n_curso, FROM_UNIXTIME(l.time) as fechatiempo
FROM `mdl_log` AS l
JOIN mdl_user u  ON l.userid = u.id
JOIN `mdl_role_assignments` AS ra ON l.userid = ra.userid
JOIN `mdl_role` AS r ON ra.roleid = r.id
JOIN mdl_course as c ON l.course=c.id
WHERE
(ra.roleid IN (5))
AND c.fullname LIKE '%%'
GROUP BY userid,l.action
ORDER BY n_curso, nombre ASC
```

Figura 4.3 Sentencia SQL para obtener los registros de las acciones de los estudiantes en el AVA MOODLE

Esta sentencia SQL, detalla las acciones o actividades didácticas de los estudiantes y los filtra por curso, a la vez los agrupa por el código de estudiante y el tipo de acción.

4.3.3. Estadísticas sobre el uso de SAVIO

Información suministrada por la Dirección de Educación Virtual (DEV), muestra que durante el primer semestre del 2011, la plataforma SAVIO tuvo 3.643 estudiantes activos, de los cuales 1.630 corresponden a los programas de ingenierías, 492 corresponden a los programas de ciencias sociales y humana, 777 economía y negocios, 744 Estudios Técnicos y Tecnológicos (T&T). En los cursos apoyados en Tecnologías de la Información y Comunicación (TIC), en los programas de ingenierías, tuvo mayor soporte, los cursos del programa de Ingeniería de Sistemas, como se muestra en la siguiente gráfica:

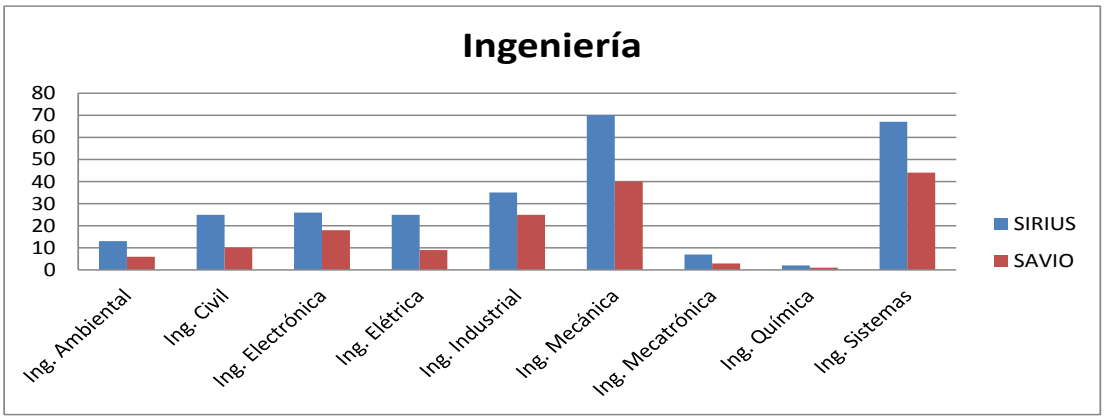

Figura 4.4 Gráfico de Cursos apoyados en TIC, de los programas de ingenierías, Suministrados por la Dirección de Educación Virtual (DEV)

Basados en la estadística realizada por (DEV), se seleccionaron los estudiantes de programa de ingeniería de sistema, puesto que son de los más activos en la participación del Ambiente Virtual de Aprendizaje SAVIO y tiene el mayor número de cursos virtuales, como podemos observar en el grafico 4.4.

4.4. Conclusión

La plataforma MOODLE es una herramienta open source para la gestión de cursos virtuales, que apoya a las instituciones en los procesos de enseñanza aprendizaje y evaluación. Entre las principales características de MOODLE tenemos el soporte de modulo como; chat, cuestionario, glosario, taller entre otros, y el manejo de reportes de registros de las actividades de los participantes (estudiantes – docentes). Los registros de las actividades o interacciones de los participantes del AVA MOODLE está constituido por atributos demográficos que le permiten a los administradores o docentes de cursos virtuales saber cuándo, dónde y que acción realizo el estudiante. Estos registros de actividades pueden determinar el comportamiento del estudiante en el uso de los recursos didácticos virtuales.

La UTB en su sistema de aprendizaje virtual interactivo SAVIO tiene el propósito de apoyar la innovación, flexibilización y virtualización del currículo institucional, la ampliación de cobertura y la proyección social. El sistema SAVIIO hace uso de la herramienta MOODLE en su versión 2.3.

La UTB para complementar su proceso de evaluación implementa el sistema de información institucional SIRIUS, el cual aplica concesiones de una, dos o tres fechas de cortes para que todos los docentes carguen sus notas. El sistema SIRIUS procesa notas o calificaciones de tipo numérico de dos cifras, expresadas en unidades y décimas, en una escala de valoración de 0 a 5, Cada calificación tiene una equivalencia conceptual según “parágrafo del artículo 90 del reglamento estudiantil de pregrado Nov. 2002“.

Capítulo 5

5. Preparación de los Datos, Modelado y Evaluación

5.1. Introducción

A través de este capítulo se le dará cumplimiento al segundo y tercer objetivo específico, se hablara de los resultados obtenidos en la fase de preparación de los datos, modelado y evaluación (Modelo CRISP-DM).

5.2. Preparación de los Datos

Una de las actividades establecidas en la fase de preparación de los datos fue realizar una revisión del estado del arte, para determinar qué tipos de estrategias se están aplicando para extraer conocimiento de los datos extraídos de los AVA.

5.2.1. Retos de la Minería de Datos

Uno de los retos de la Minería de Datos MD es extraer información de la web, puesto que los documentos web contienen datos muy diversos (texto, audio, imagen, etc.), esto hace que existan diferentes formas de minar los sitios web. Existen varias técnicas de minería de datos en la web según [17]: su contenido, estructura y uso.

La internet es uno de los medios más populares e interactivos de difundir información, pero esta situación hace que los usuarios tengan una sobrecarga de información según [16], uno de los problemas relacionados con esta situación, está el de aprender de los usuarios es decir, saber qué es lo que los usuarios hacen y quieren, esto permite personalizar la información cumpliendo con las expectativas o gustos del usuarios [17].

La minería de uso web según [17], Captura las actividades de los usuarios durante su navegación y extrae patrones de comportamiento que pueden ayudar a comprender las preferencia de navegación, el comportamiento de los usuarios o mejorar futuras páginas adaptando las interfaces de los sitios web a los usuarios individuales. Cuando los usuarios interactúan con el sitio web, los datos que registran su comportamiento se almacenan en los servidores web.

Una aproximación para minar los patrones de navegación de los usuarios desde los datos log (bitácoras que registran el comportamiento de usuarios en servidores web) según [3], seria transformar los datos a notación tabular y aplicar técnicas estándar de minería de datos, como las reglas de asociación o agrupación.

Una de las metas principales del trabajo de grado es determinar los factores que relacionan el comportamiento de los estudiantes de ingeniería de sistema de la UTB en función de la participación en el Ambiente virtual de aprendizaje SAVIO y la escala de calificación correspondiente.

Uno de los resultados en esta fase de preparación de los datos, fue obtener el dataset o vista minable con los registros de la participación de los estudiantes de ingeniería de sistema de la UTB en los cursos b-learning (Algoritmo, Estructura de datos, Progresión) con su escala de calificación final. Estos registros fueron suministrados por la dirección de SAVIO y SIRIUS de la UTB, luego aplicando la herramienta Microsoft Excel se llevó a notación tabular, obteniendo de esta manera un dataset con 210 instancias y cada instancia con 56 atributos, correspondiente a los factores de participación en el AVA SAVIO y la escala de calificación final obtenida. Los atributos que representan la participación son de tipo numérico y los que representan la escala de calificación son de tipo categórico. El atributo 'escala de calificación' puede tomar 7 tipos de valores alfabéticos, como se muestra en la siguiente tabla:

Tabla 5-1: Representación Escala de Calificación en letra

Escala	*Representación Escala*
Insuficiente (0.0 a 0.9)	*A*
Muy Deficiente (1.0 a 1.9)	*B*
Deficiente (2.0 a 2.9)	*C*
Aceptable (3.0 a 3.4)	*D*
Bueno (3.5 a 4.0)	*E*
Muy Bueno (4.1 a 4.5)	*F*
Sobresaliente (4.6 a 5)	*G*

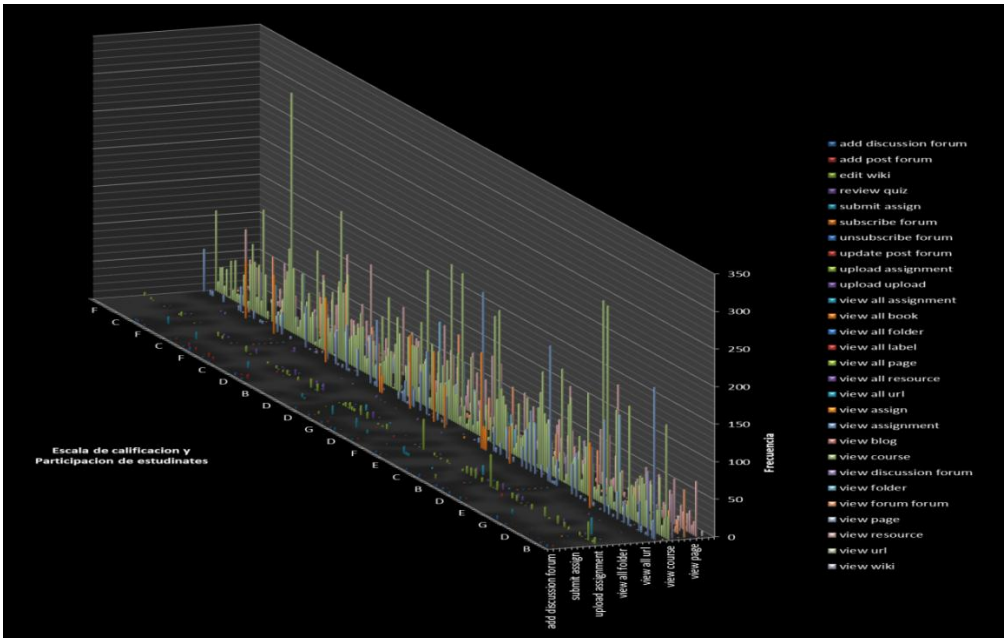

Figura 5.1 Participación de los estudiantes de Ingeniería de Sistema en el AVA SAVIO

La figura 5.1 representa la participación de los estudiantes de ingeniería de sistema de la UTB en los cursos b-learning o semi-virtuales: estructura de datos, algoritmos, programación, en los periodos académicos 2010 y 2011.

5.3. Resultados del proceso de Modelado

El proceso descubrimiento de conocimiento en datos (KDD) tiene como objetivos encontrar patrones o reglas interesantes, precisas y legibles en un conjunto de datos establecido, para llevar a cabo esta misión hace uso de conjunto de técnicas, modelos, metodologías [17].

La minería de datos es una de las fases del proceso de descubrimiento de conocimiento en bases de datos (KDD), está constituida por una suite de algoritmos que permiten realizar un conjunto de tareas [3]. Entre las tareas están: clasificación, regresión, agrupamiento, reglas asociación, factorización. Entre las técnicas o algoritmos están: arboles de decisión (ID3 , c4.5), Redes de Cohonen, regresión lineal y logarítmica, regresión logística, k-means, Apriori, Naive Bayes, Maquinas de vectores de soporte, CN2 rules(cobertura), etc.

5.3.1. Algoritmos de agrupamiento

En la investigación se hará uso de las tareas de agrupamiento, a continuación se describen una serie de técnicas [39], para llevar a cabo estas tareas de minería de datos.

Tabla 5-2: Algoritmo de Agrupamiento

Tipo	***Algoritmo***	***Conceptos***	***Características***
Modelos Probabilísticos	*EM Maximización de la Expectativa*	*Ajustar los datos a un modelo matemático (se supone que los datos provienen de la superposición de varias distribuciones de probabilidades por ejemplo: la distribución normal o campana de gauss). Trabaja en dos pasos:* *Expectativa "Paso E": donde se computa la esperanza de la verosimilitud mediante la inclusión de variables latentes como si fueran observables.* *Maximización "Paso M": se computan estimadores de máxima verosimilitud de los parámetros mediante la maximización de la verosimilitud esperada del paso E.*	*Utilizan la métrica log likelihood para calcular la verisimilitud o error de entropía.*
Por particiones	*K-means*	*Método de agrupamiento, que tiene como objetivo la partición de un conjunto en k grupos en el que cada observación pertenece al grupo más cercano a la media.*	*Algoritmo de agrupamiento por particiones.* *Número de clusters conocido (k).* *Cada cluster tiene asociado un centroide (centro geométrico del cluster).*

			Los puntos se asignan al cluster, cuyo centroide esté más cerca (utilizando cualquier métrica de distancia). *Iterativamente, se van actualizando los centroides en función de las asignaciones de puntos a clusters, hasta que los centroides dejen de cambiar.*
Basado en densidad	*DBSCAN*	*Los algoritmos basados en densidad enfocan el problema de la división de una base de datos en grupos teniendo en cuenta la distribución de densidad de los puntos, de modo tal que los grupos que se forman tienen una alta densidad de puntos en su interior mientras que entre ellos aparecen zonas de baja densidad.* *El primer algoritmo que emplea este enfoque para dividir el conjunto de datos es DBSCAN, en este aparecen los conceptos: punto central, borde y ruido los que serán empleados para determinar los diferentes clusters. Otros algoritmos basados en densidad que siguen la línea de DBSCAN son: OPTICS y GDBSCAN.* *Útiles cuando los clusters tienen formas irregulares, están entrelazados o hay ruido/outliers en los datos.*	*Identifican clusters de formas arbitrarias.* *Robustos ante la presencia de ruido.* *Escalables: Un único recorrido del conjunto de datos*
Jerárquico	*Cobweb*	*son aquellos en los que se va particionando el conjunto de datos por niveles, de modo tal que en cada nivel*	*Dos tipos de técnicas de clustering jerárquico*

		generalmente , se unen o se dividen dos grupos del nivel anterior, según si es un algoritmo aglomerativo o divisivo. *Se basa en la construcción de un árbol en la que las hojas son los elementos del conjunto de ejemplos y el resto de los nodos son subconjuntos de los ejemplos que pueden ser utilizados como particionamiento del especio.*	*Técnicas aglomerativas:* *Comenzar con cada caso como cluster individual.* *En cada paso, combinar el par de clusters más cercano hasta que sólo quede uno (o k).* *Técnicas divisivas:* *Comenzar con un único cluster que englobe todos los casos de nuestro conjunto de datos.* *En cada paso, partir un cluster hasta que cada cluster contenga un único caso.* *Las estrategias jerárquicas más conocidas son Single Link (SL), Average Link (AL) y Complete Link (CL).*

Una de las técnicas de agrupamiento más usadas para el análisis de los datos generados por los sistemas Ambiente Virtuales de Aprendizaje, son los algoritmos k-means y EM, según [41][14].

Por revisión del estado del arte se seleccionó como técnica de agrupamiento k-means para llevar a cabo el experimento en la etapa de modelado. A continuación se resaltan aspectos del algoritmo k-means según [40], tales como problemas, soluciones y variantes

Uno de los posibles problema con k-means

- Hay que elegir a priori el valor de k (a priori, no sabemos cuántos grupos puede haber).

Una de las posibles soluciones para el problema de k-means

- Usar un método jerárquico sobre una muestra de los datos (por eficiencia) para estimar el valor de k.
- Usar un valor de k alto, ver los resultados y ajustar.
- Siempre que se aumente el valor de k, disminuirá el valor SSE.
- Lo normal será ir probando con varios valores de k y comprobar cuándo no hay una mejora significativa en SSE.

Una de las variantes de k-means

- GRASP [Greedy Randomized Adaptive Search Procedure] para evitar óptimos locales.
- k-Modes (Huang'1998) utiliza modas en vez de medias (para poder trabajar con atributos de tipo categórico).
- k-Medoids utiliza medianas en vez de medias para limitar la influencia de los outliers.

Según [17] una de las técnicas o métodos KDD más difíciles de evaluar son los modelos de agrupamiento en comparación con las técnicas de regresión y clasificación. Una primera aproximación es utilizar la verosimilitud (likelihood) para determinar la efectividad del modelo o hipótesis sobre los datos de estudio, en otras palabras las técnicas basadas en la verosimilitud permiten determinar como el modelo o hipótesis obtenida a través de los algoritmos de agrupación describe a los datos de entrenamiento.

Otra técnica de evaluación seria utilizar la suma de los errores cuadráticos SSE [17], para determinar qué tan compacto son los grupos del modelo o hipótesis obtenida, es decir mide la cohesión de los objetos de cada grupo. Otra alternativa para determinar si un modelo de agrupamiento es adecuado para un conjunto de datos consiste en aprender de varios modelos desde ese mismo conjunto utilizando diversas técnicas de aprendizaje. Si los comparamos entre si y coinciden, podemos pensar que el agrupamiento es acertado. Para evaluar la similitud entre dos modelos de agrupamiento podemos utilizar una estrategia similar a la aproximación entrenamiento / prueba de clasificación y regresión, basada en la partición de los datos en dos partes, la primera de ellas se utiliza para construir el modelo de agrupamiento y la segunda para comprobar si los modelos construidos son similares.

5.3.2. Experimentos

Análisis del comportamiento de los estudiantes de Ingeniería de sistema de la UTB Universidad Tecnológica de Bolívar en función de su participación en el AVA SAVIO y su escala de calificación final. A continuación se muestran detalles del experimento realizado:

5.3.3. Pregunta de Investigación

¿Qué grupos de estudiantes de Ingeniería de Sistema de la UTB, tienen mayor interés por el uso de los materiales didácticos virtuales, en función de la frecuencia de sus acciones en la plataforma SAVIO y su escala de calificación?

5.3.4. Hipótesis

Los grupos de estudiantes de ingeniería de sistema de la UTB, que tienen mayor interés en la plataforma SAVIO, son aquellos que obtienen una escala de calificación final aprobatoria.

5.3.5. Muestra

Se tomaron datos de forma aleatoria de 6 cursos B-learning (Estructura de datos, programación, Algoritmos) del programa de Ingeniería en Sistemas de la Universidad Tecnológica de Bolívar UTB, correspondiente a los periodos 2010 y 2011.

5.3.6. Procedimiento

Para llevar a cabo el experimento se reunió toda la información de los cursos b-learning (Algoritmo, Estructura de datos, Programación) del programa de Ingeniería de Sistemas de la UTB en un dataset, para determinar un modelo que describa los factores más significativos en función de las acciones de los estudiantes en la plataforma SAVIO y la escala de calificación obtenida, implementando algoritmos de Minería de Datos no supervisado, con la esperanza de obtener un modelo optimo que sirva de indicador para medir aspectos técnicos de los cursos virtuales.

Otros de los procedimientos aplicados en este experimento, para llevar a cabo el proceso de evaluación de la calidad del modelo generado por la técnica de agrupación k-means, consistió en la partición del dataset (210 instancias), en dos nuevos dataset: uno de entrenamiento con el 80% (168 instancias) y el otro de prueba con el 20% (42 instancias). En el proceso de partición se hizo uso de un filtro de la herramienta WEKA para la generación de los dos nuevos dataset (entrenamiento/prueba).

Tabla 5-3: Atributos del Dataset de Entrenamiento y el Dataset de Prueba

No	Atributo
1	Escala
2	add discussion forum
3	add post forum
4	attempt quiz
5	choose again choice
6	choose choice
7	close attempt quiz
8	continue attemp quiz
9	continue attempt quiz
10	delete discussion forum
11	delete post forum
12	edit wiki
13	mark read discussion
14	recent course
15	review quiz

No	Atributo
16	search forum
17	submission statement accepted assign
18	submit assign
19	subscribe forum
20	unsubscribe forum
21	update post forum
22	update user
23	upload assignment
24	upload upload
25	user report course
26	user report forum
27	view all assignment
28	view all book
29	view all choice
30	view all folder

No	Atributo
31	view all imscp
32	view all journal
33	view all label
34	view all page
35	view all resource
36	view all url
37	view all user
38	view all wiki
39	view assign
40	view assignment
41	view blog
42	view choice
43	view course
44	view discussion forum
45	view folder

No	Atributo
46	view forum forum
47	view forums forum
48	view journal
49	view page
50	view quiz
51	view resource
52	view submit assignment form assign
53	view summary quiz
54	view url
55	view user
56	view wiki

Otras de las estrategias utilizadas para este experimento fue reducir el número de atributos de 56 a 24, esto con el fin de omitir atributos irrelevantes y mejorar la calidad del modelo de agrupamiento [17].

Un panorama alentador que se presentó en este experimento fue trabajar con un numero de cluster igual a k=15 y usando la semilla 150 de números random, logrando con estos criterios bajar la suma de los errores cuadráticos SSE, esto quiere decir que existe una

mayor compactación o cohesión entre los objetos de cada clúster descrito por el algoritmo k-means.

A continuación una explicación pasó a paso el proceso del algoritmo k-mean.

En este caso el algoritmo k-means: inicializando se le asigna K=15 centroides, luego se forman k-grupos asignando cada punto (objeto) al centroide más cercano. En el proceso iterativo se calculan las distancias de todos los puntos a los k-centroides, luego se forman k grupos asignando cada grupo al centroide más cercano, seguidamente se recalculan los nuevos centroides, esta iteración continua mientras los valores de los centroides cambien.

Los valores de los centroides se recalculan de la siguiente forma:

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} d^2(m_i, x)$$

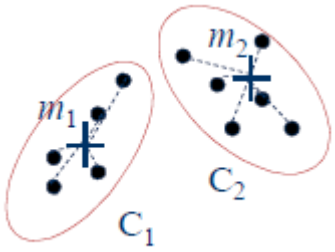

Figura 5.2 Proceso de recalculo de los centroides en el Algoritmo K-mean. Tomado de http://elvex.ugr.es/idbis/dm/slides/4%20Clustering.pdf

Se escoge los valores de m_i que minimicen la función objetivo SSE.

Cuando se utiliza la distancia euclidiana, la suma de los errores cuadráticos (SSE), se minimiza usando la media aritmética (por cada atributo o variable). Cuando se trabaja con atributos nominales, en nuestro caso "escala calificación", se minimizan los valores de m_i con base en la moda de los datos (el valor que más se repita).

Tabla 5-4: Ejemplo del proceso de recalculo de los centroide en el método k media

X_a=(C, 12, 8, 9,…, 5)
X_b=(A, 14, 3, 2,…, 1)
X_c=(C, 10, 9, 2,…, 3)
. . .
m_1=(C, 12, 6.6, 4.3,...,3)

Después de recalcular las nuevas posiciones de los centroides se ajustan los puntos de cada objeto al centroide más cercano, tomando como base todos sus atributos.

A continuación se muestra los resultados del experimento utilizando la técnica k-means con número de cluster k=15 y usando la semilla 150 de números random:

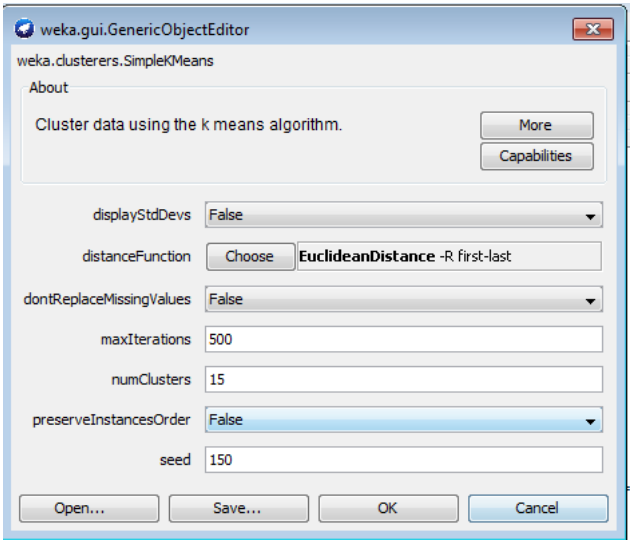

Figura 5.3 Parámetros del Algoritmo K-mean

Se incrementó a 15, el número de cluster para lograr obtener el total de los tipos de escala de calificación, siete niveles (A-F) como se observa en la tabla 5-1.

Se usó la semilla 150 de números random, para tener posibilidades favorables en la ubicación de los centroides.

5.3.7. Resultados del Primer Experimento

```
=== Model and evaluation on training set ===


kMeans
======

Number of iterations: 5
Within cluster sum of squared errors: 25.505747037528362
Missing values globally replaced with mean/mode
```

Tabla 5-5: Modelo de Entrenamiento, formado con K medias

Attribute	Cluster# Full Data	0 (14)	1 (22)	2 (25)	3 (15)	4 (14)	5 (1)	6 (23)	7 (29)	8 (8)	9 (1)	10 (2)	11 (1)	12 (1)	13 (5)	14 (7)
ESCALA	D	G	E	F	B	C	G	C	D	D	F	D	G	B	G	A
add discussion forum	2,3333	2,119	2,424	2,267	2,067	2,5		2,464	2,241	2,125	3	2,667	2	3	3,133	
add post forum	2,2632	2,083	2,239	2,413	2,179	2,387		2,446	2,176	2,04	1		1	1	2,811	
choose choice	1,6667					1,691			1,655							
edit wiki	4								4							
update post forum	1,5							1,5		1,563				1		
upload assignment	5,6667	4,952	5,727	4,427	4,756	5,381	13	5,232	5,701	5,417	41	18,5			5,667	
upload upload	3,7037	3,868	3,985	3,443	3,477	3,788		3,816	3,803	2,94		6,352			2,822	
user report forum	1,9545		2,054	1,84	2,158	2,036		1,83	1,79					8	1,764	1,818
view all assignment	4	3,786	3,818	3,88	3,8	3,786			4,793			2			3,4	
view all resource	1,8235	1,706	1,786	1,798		2,277		1,752	1,767	1,743		1,912		3	1,659	1,849
view assignment	27,9853	23,35	36,44	20,47	18,66	34,49	52	22,21	27,68	28,49	204	75,5	4	2	39,39	20,99
view blog	1,5385		1,514	1,495	1,569	1,462		1,535	1,536					5	1,323	
view choice	9,3333					10,02			9							
view discussion forum	18,234	13,49	19,15	18,52	15,76	16,24		19,2	17,31	19,65	83	5	30	52	21,05	
view folder	17,8077	17,58	17,42	19,46	15,63	18,48	1		16,92		42	9,404	1		26,88	
view forum forum	20,7037	12,59	20,59	20,01	17,34	19,36		23,68	18,63	24,81	75	21,5	34	72	28,94	
view forums forum	3,3333		3,227			3,167			3,563			2,167		6	2,867	
view journal	2					2,071			1,966							
view page	7,1429	6,939	6,812	7,206	6,838	7,061	32	6,652	6,808	10,38	1				5,286	
view quiz	5,1111		5,106	5,227		5,524	4	4,976	5,035			4,056				
view resource	29,6216	21,79	31,13	27,44	26,28	50	14	15,23	18,01	72,38	66	47	9	62	82,8	13,66
view url	4,8387	4,505	4,588	4,364	4,671	5,456	29	4,562	4,953			2,919	5		4,271	5,028
view wiki	8					8			8							

Las celdas en blanco, corresponden a valores missing(vacíos), esto quiere decir que el modelo de agrupamiento definido por el algoritmo k-means, nos indica que algunos grupos de estudiantes de ingeniería de sistema de la UTB presentan ausencia en la participación de ciertas actividades didácticas.

```
Time taken to build model (full training data) : 0.05 seconds

=== Model and evaluation on training set ===

Clustered Instances

 0       14 (  8%)
 1       22 ( 13%)
 2       25 ( 15%)
 3       15 (  9%)
 4       14 (  8%)
 5        1 (  1%)
 6       23 ( 14%)
 7       29 ( 17%)
 8        8 (  5%)
 9        1 (  1%)
10        2 (  1%)
11        1 (  1%)
12        1 (  1%)
13        5 (  3%)
14        7 (  4%)
```

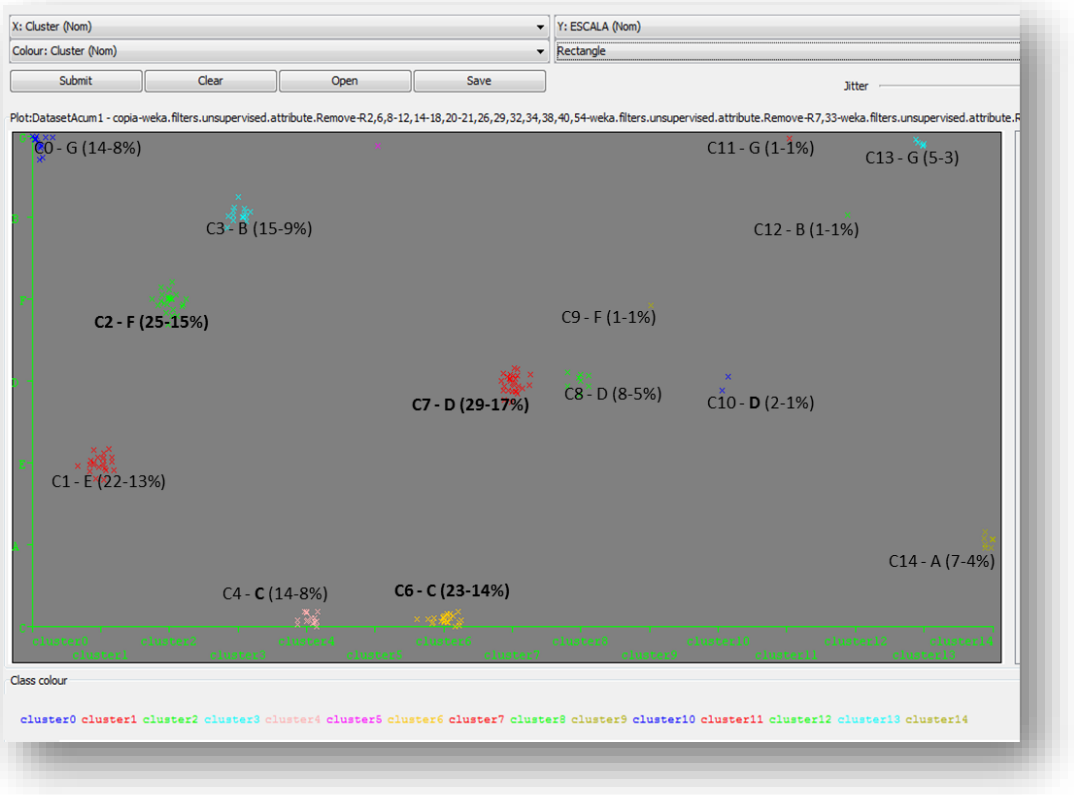

Figura 5.4 Representación de los grupos de estudiantes formados con el método K medias en función de escala de calificación final y la participación en AVA SAVIO.

Se puede observar 15 particiones generadas por el algoritmo Minería de Datos k-means, teniendo mayor cubrimiento de elementos, el clúster número 7, seguido de los cluster 2 y 6.

5.3.8. Resultados del Segundo Experimento

En el primer experimento se definió un modelo de agrupamiento con 15 cluster, para describir el comportamiento de los estudiantes de ingeniería de sistema de la UTB en función de las actividades realizada en la plataforma SAVIO y los siete tipos de escala de calificación. Una de las debilidades del modelo obtenido en el primer experimento es el elevado número de cluster que se determinaron, para mejorar este panorama se reorientaron varios aspectos de la evidencias de estudio y se hizo una revisión del estado del arte.

Con el propósito de mejorar la calidad del modelo de agrupamiento, se realizaron tareas de pre-procesamiento para mejorar la vista minable, en este caso se utilizaron dos tipos de valores (aprobados y no aprobados), para el atributo 'escala de calificación'.

Tabla 5-6: Representación Escala de Calificación en dos tipos

Escala	*Representación Escala*
No Aprobatorio (0.0 a 2.9)	*0*
Aprobatorio (3.0 a 5.0)	*1*

Haciendo revisión del estado del arte se encontró que una de las ventajas de trabajar con k-medias es que tiene un significado gráfico y estadístico inmediato [43]. Al algoritmo k-medias hay que proporcionarle a priori el número de clusters en los que quiere segmentar el número de estudiantes, una forma de obtener este dato es aplicando el algoritmo EM, ya que obtiene este dato de forma óptima [44]. Al aplicar el algoritmo EM con el dataset de entrenamiento se encontraron los siguientes resultados:

```
Time taken to build model (full training data) : 6.78 seconds

=== Model and evaluation on training set ===

Clustered Instances

0        2 (  1%)
1       60 ( 36%)
2       90 ( 54%)
3       16 ( 10%)


Log likelihood: 6.24223
```

Se observa que el algoritmo EM logro una selección automática de 4 grupos de estudiantes. Se procede a la aplicación del algoritmo de K-medias con el dataset de entrenamiento, para un número de cluster de 4 y un valor para la semilla de 10, a continuación los resultados:

```
=== Model and evaluation on training set ===

kMeans
======

Number of iterations: 5
Within cluster sum of squared errors: 31.573633608971978
Missing values globally replaced with mean/mode
```

Tabla 5-7: Modelo de Entrenamiento, formado con K medias y EM

Cluster/Atributos	Full Data (168)	0 (60)	1 (6)	2 (13)	3 (89)
ESCALA	0,6429	0	1	1	1
add discussion forum	2,3333	2,3667	2,5556	2,8205	2,2247
add post forum	2,2632	2,3202	1,8421	2,1053	2,2762
choose choice	1,6667	1,6722			1,6629
edit wiki	4				4
update post forum	1,5	1,4917		1,5385	
upload assignment	5,717	5,245	13	9,2104	5,0339
upload upload	3,7037	3,7099	7,284	3,3647	3,5077
user report forum	1,9545	2,0614	2,6364	1,8811	1,8473
view all assignment	4	3,9	3	3,7692	4,1685
view all resource	1,8235	1,9245	1,7157	1,7104	1,7792
view assignment	27,9853	23,7074	58,6667	63,2986	23,6428
view blog	1,5385	1,5846		1,4556	1,5194
view choice	9,3333	9,4944			9,2247
view discussion forum	18,234	18,0805	9,3333	26,2619	17,765
view folder	17,8077	17,4205	15,0064	23,142	17,4784
view forum forum	20,7037	21,5426	12,6667	30,3248	19,2747
view forums forum	3,3333	3,3389	2,5556	3,1538	3,4082
view journal	2	2,0167			1,9888
view page	7,1429	6,8595		7,9451	7,2167
view quiz	5,1111	5,1556	4,7407		5,1061
view resource	29,6216	26,7018	41,5	88,9231	22,1272
view url	4,8387	4,857	4,1989	4,3499	4,9409
view wiki	8	8			8

```
Time taken to build model (full training data) : 0.02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0       60 ( 36%)
1        6 (  4%)
2       13 (  8%)
3       89 ( 53%)
```

En este caso observamos 4 distribuciones de la siguiente forma: cluster 0 con 60 instancias, cluster 1 con 6 instancias, cluster 2 con 13 instancias, cluster 3 con 89 instancias.

En el segundo experimento encontramos un panorama más alentador, puesto que se definió un modelo de agrupamiento sin necesidad de elevar a un nivel exagerado el número de cluster, para agrupar a los estudiantes de ingeniería de sistema de la UTB en función de su participación y escala de calificación.

El modelo de agrupamiento seleccionado en esta etapa de modelado, fue el obtenido en el segundo experimento, puesto que fue el modelo más preciso y conciso para el análisis de las evidencias o datos de estudio.

5.3.9. Resultado del proceso de evaluación

En la fase de Evaluación del modelo se trabajó con el dataset de 42 instancias. A continuación se muestra el resultado del proceso:

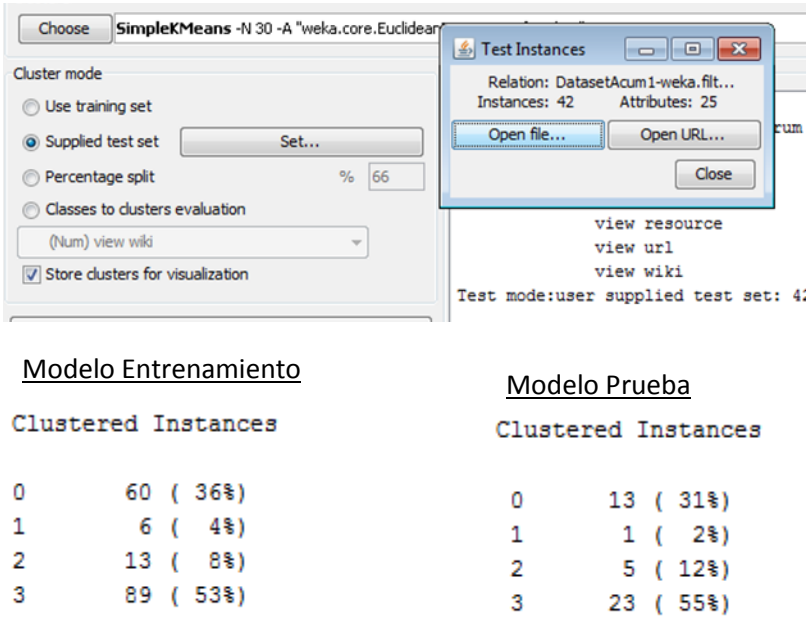

Figura 5.5: Comparación del Modelo de Prueba y Modelo Entrenamiento

Se puede observar los modelos generados por la técnica k-means a la izquierda modelo entrenamiento (168 instancias) y a la derecha modelo de prueba (42 instancias), observamos que el modelo de prueba tiene una alta similitud con el modelo de entrenamiento generado por la misma técnica k-means. En este caso vemos que el modelo de prueba describe un mayor porcentaje en los cluster 0, 2, y 3, el mismo panorama observado en el modelo de entrenamiento.

Con base en el estado del arte se tomaron dos tipos de técnicas para determinar la calidad del modelo de agrupamiento obtenido en la etapa de modelado. La primera con base en la métrica log- likelihood utilizada por el algoritmo EM para la selección automática de los 4 cluster, mostro un valor log- likelihood =6.24223, siendo este valor mayor a cero, esto determina que la mejor agrupación para el dataset de entrenamiento es de 4 [45]. En la segunda se utilizó una estrategia similar a la aproximación entrenamiento / prueba de clasificación y regresión, en este sentido se encontró una alta similitud entre el modelo de entrenamiento y de prueba definido por el algoritmo k-medias.

5.4. Conclusión

Se hizo uso de los algoritmos de agrupamiento k-means y EM para determinar modelos descriptivos que relacionaran el comportamiento de estudiantes en función de su participación en la plataforma SAVIO y su escala de calificación.

Se llevó a cabo dos tipos de experimentos, en el primero se obtuvo un modelo que abarcaba entre los 15 cluster cada uno de los siete niveles de escala de calificación y los factores participación en SAVIO, una de las debilidades encontradas en este modelo fue el elevado número de cluster que se determinaron, para mejorar este panorama se realizó un segundo experimento, donde se aplicó como estrategias la modificación del atributo 'escala de calificación' a dos tipos de valores (Aprobados y No aprobados). Otras de las de las estrategias fue implementar dos tipos de algoritmos de Minería de Datos, EM para determinar de forma automática el número de cluster y K-medias para definir el modelo de agrupamiento a partir del número de cluster obtenidos por la técnica EM. Entre los resultados del experimento se observó un mayor cubrimiento de instancias en los cluster número 3 y 0. El modelo de agrupamiento seleccionado, fue el obtenido en el segundo experimento, puesto que fue el modelo más preciso y conciso para el análisis de las evidencias o datos de estudio.

Aplicando la técnicas de evaluación tales como la métrica log- likelihood y la aproximación particionamiento (Entrenamiento / Prueba) utilizadas en tareas de clasificación y regresión, se puede decir que el modelo de agrupamiento generado a través de los algoritmos k-means y EM en el segundo experimento, es adecuado para el diagnóstico de los datos extraídos del Ambiente Virtual Aprendizaje SAVIO.

Capítulo 6

6. Implementación del modelo

6.1. Introducción

Dándole cumplimiento al objetivo específico número cuatro, se hará un análisis a partir del modelo definido por las técnicas de agrupamiento K-means , en el segundo experimento realizado en la etapa de modelado, con base en este análisis se establecerán estrategias que sirvan de apoyo para la mejora continua del proceso de enseñanza aprendizaje virtual.

En este espacio se analizarán cada uno de los cluster descritos por el algoritmo de minería de datos k-means, tratando de encontrar algún patrón o tendencia que siguen cada uno de los grupos definido.

6.2. Análisis Intra-Cluster del modelo de agrupamiento definido en la etapa de modelado.

A continuación se presenta el análisis de cada uno de los 4 cluster obtenido en el segundo experimento de la etapa de modelado a través de la técnica k-means. En el análisis de los cluster se tuvo en cuanta el top 5 (las cinco acciones más frecuentadas) y bottom 5 (las cinco acciones menos frecuentes), los cluster representan los grupos de estudiantes de ingeniería de sistema de la UTB en función de la participación en el AVA SAVIO y su escala de calificación final.

Tabla 6-1: Análisis del cluster 0 del modelo de entrenamiento, formado con el método k medias

Cluster	Top 5		Bottom 5		Escala
	Accion	Centroide	Accion	Centroide	
0	view resource	26,7018	update post forum	1,4917	No Aprobatoria
	view assignment	23,7074	view blog	1,5846	
	view forum forum	21,5426	choose choice	1,6722	
	view discussion forum	18,0805	view all resource	1,9245	
	view folder	17,4205	view journal	2,0167	

En la tabla 6-1 se muestra el cluster 0, con el top 5 (acciones más frecuentadas) y el Bottom 5 (acciones menos frecuentadas), observamos que en el cluster 0, representados por los estudiantes de ingeniería de sistema con escala de calificación No Aprobatoria (0.0 a 2.9), tienen mayor interacción por participar de las acciones: revisar recursos, tareas, foro de discusión de debate sencillo, foros de discusión de debate general, revisar folder. En su

defecto las instancias que se agruparon en cluster 0 presentaron baja frecuencia en participar de las acciones: actualizar foro, revisar blog, elegir consulta, revisar listado de recursos y revisar diario.

Basado en los trabajos de Norma Scagnoli (2005) [46], Se sugiere a los docentes o administrador de cursos virtuales, las siguientes estrategias para lograr motivar a los estudiantes y mejorar la participación en AVA SAVIO:

- Diseñar Material Educativo y seleccionar recursos de apoyo interesante y variado relacionado con los diversos intereses y necesidades de los estudiantes.
- Evaluar un mismo aprendizaje en varios tipos de actividades virtuales (Foros, Blog, Tareas), esto con fin de darle oportunidad al estudiante que se adapte al proceso de aprendizaje virtual que más se le facilite.
- Adaptar los contenidos al nivel de conocimientos de los estudiantes, a través del uso de actividades didácticas que se complementen con recursos de multimedia tales como voz o video.
- Actuar con paciencia, dando a otros y a si mismo tiempo para procesar la información, puesto que no todos los estudiantes cuentan con una computadora en sus espacios de trabajo o en el hogar.
- Retroalimentar al estudiante de manera permanente para que reconozca sus cualidades y habilidades reales (Satisfacción de necesidad de Estima) a través de la retroalimentación.
- Enviar recordatorios de las actividades o materiales establecidas a los estudiantes, con el fin de alertarlos con la fechas de cumplimiento de las actividades.
- Socializar o mostrar la metodología de trabajo a seguir y sus ventajas, mostrando también cierta flexibilidad para poder realizar posibles cambios en la programación.

Tabla 6-2: Análisis del cluster 1 del modelo de entrenamiento, formado con el método k medias

Cluster	Top 5		Bottom 5		Escala
	Accion	Centroide	Accion	Centroide	
1	view assignment	58,6667	view all resource	1,7157	Aprobatoria
	view resource	41,5	add post forum	1,8421	
	view folder	15,0064	add discussion forum	2,5556	
	upload assignment	13	view forums forum	2,5556	
	view forum forum	12,6667	user report forum	2,6364	

En la tabla 6-2 se muestra el cluster 1, con el top 5 (acciones más frecuentadas) y el Bottom 5 (acciones menos frecuentadas), observamos que en el cluster 1, representados por los estudiantes de ingeniería de sistema con escala de calificación Aprobatoria (3.0 a 5.0), tienen mayor interacción por participar de las acciones: revisar tareas, recursos, folder, enviar

tareas, revisar foros de discusión de debate sencillo. En su defecto las instancias que se agruparon en cluster 1 presentaron baja frecuencia en participar de las acciones: revisar listado de recursos, adicionar respuesta a un foro, adicionar tema a un foro, revisar listado de foros y revisar reporte de mensajes de foros.

Basado en los trabajos de Norma Scagnoli (2005) [46], Se sugiere a los docentes o administrador de cursos virtuales, las siguientes estrategias para lograr motivar a los estudiantes y mejorar la participación en AVA SAVIO:

- Plantear retos, problemas de cierta dificultad, debates. A través del uso de herramientas virtuales como 'foro de debate general' para que los estudiantes puedan crear nuevas preguntas, compartir ideas. El uso de esta estrategia hace que el proceso de enseñanza aprendizaje sea flexible, dándole la oportunidad al estudiante de ser escuchado.
- Ofertar puntos de calificación para estimular a los estudiantes en el uso de las herramientas didácticas: wiki, blog, consulta.

Tabla 6-3: Análisis del cluster 2 del modelo de entrenamiento, formado con el método k medias

Cluster	Top 5		Bottom 5		Escala
	Accion	Centroide	Accion	Centroide	
2	view resource	88,9231	view blog	1,4556	Aprobatoria
	view assignment	63,2986	update post forum	1,5385	
	view forum forum	30,3248	view all resource	1,7104	
	view discussion forum	26,2619	user report forum	1,8811	
	view folder	23,142	add post forum	2,1053	

En la tabla 6-2 se muestra el cluster 2, con el top 5 (acciones más frecuentadas) y el Bottom 5 (acciones menos frecuentadas), observamos que en el cluster 2, representados por los estudiantes de ingeniería de sistema con escala de calificación Aprobatoria (3.0 a 5.0), tienen mayor interacción por participar de las acciones: revisar recursos, tareas, foros de discusión de debate sencillo, foros de discusión de debate general, folder. En su defecto las instancias que se agruparon en cluster 2 presentaron baja frecuencia en participar de las acciones: revisar blog, editar mensajes de foros, revisar listados de recursos, revisar reporte de mensajes de foros y enviar nuevo mensaje a un foro.

Basado en los trabajos de Norma Scagnoli (2005), Se sugiere a los docentes o administrador de cursos virtuales, las siguientes estrategias para lograr motivar a los estudiantes y mejorar la participación en AVA SAVIO:

- Animar y alentar la participación, el diálogo, la intercomunicación, el intercambio, a través del uso de herramientas como foro, correo en el AVA SAVIO.
- Comenzar con actividades moderadamente difíciles en el que el reto pueda ser superado sin complicaciones y continuar con actividades relativamente difíciles.

Tabla 6-4: Análisis del cluster 3 del modelo de entrenamiento, formado con el método k medias

Cluster	Top 5		Bottom 5		Escala
	Accion	Centroide	Accion	Centroide	
3	view assignment	23,6428	view blog	1,5194	Aprobatoria
	view resource	22,1272	choose choice	1,6629	
	view forum forum	19,2747	view all resource	1,7792	
	view discussion forum	17,765	user report forum	1,8473	
	view folder	17,4784	view journal	1,9888	

En la tabla 6-4 se muestra el cluster 3, con el top 5 (acciones más frecuentadas) y el Bottom 5 (acciones menos frecuentadas), observamos que en el cluster 3, representados por los estudiantes de ingeniería de sistema con escala de calificación Aprobatoria (3.0 a 5.0), tienen mayor interacción por participar de las acciones: revisar tareas, recursos, foro de debate sencillo, foro de debate general, revisar folder. En su defecto las instancias que se agruparon en cluster 5 presentaron baja frecuencia en participar de las acciones: revisar blog, elegir opción de consulta (Urna virtual), revisar reporte de mensajes de foro, revisar diario.

Basado en los trabajos de Norma Scagnoli (2005) [46], Se sugiere a los docentes o administrador de cursos b-learning las siguientes estrategias para lograr motivar a los estudiantes y mejorar la participación en AVA SAVIO:

- Intervenir y contestar de manera rutinaria ante cualquier opinión o inquietud de los estudiantes.
- Guiar al alumno a que establezca metas específicas a un mediano y largo plazo durante el transcurso de la materia.
- Promover actividades formativas, que permitan modelar paso a paso la forma de realizar las actividades creativas o investigativas, tratando de lograr de forma significativa cumplir con cada uno de los objetivos de aprendizaje establecidos para la respectiva unidad de estudio.
- Reconocer el esfuerzo de manera grupal e individual de cada uno de los estudiantes es esencial para potencializar las habilidades, además de incrementar la participación.

6.2.1. DETALLES DEL ANÁLISIS INTRA-CLUSTER

Algo interesante para resaltar en este análisis Intra-Cluster es el comportamiento similar entre el cluster 0 (estudiantes con escala de calificación No Aprobatoria) y el cluster 3 (estudiantes con escala calificación Aprobatoria), en relación al top 5 (acciones más frecuentadas) y Botton 5 (acciones menos frecuentes). Este acontecimiento puede ser causado por los siguientes factores:

- Confusión por falta intuición en contenido de los materiales de estudio o las actividades que realizaron los estudiantes con escala de calificación no aprobatoria.
- Las actividades en la que tuvieron mayor interacción, no tenían un peso o porcentaje de valoración significativo para alcanzar una escala de calificación aprobatoria.
- Los cursos virtuales en los que participaron los estudiantes con escala de calificación no aprobatoria, no tuvieron una modalidad 100% B-learning.

Otros de los aspectos a resaltar en esta sección de análisis Intra-Cluster con base en el top 5 y botton 5, se describe a continuación:

- Dentro de las actividades más utilizadas por los grupos estudiantes de Ingeniería de Sistema de la UTB, en el AVA SAVIO se pueden destacar: el revisar recursos, revisar tareas y revisar foros.
- De las actividades menos usadas por los grupos estudiantes de Ingeniería de Sistema de la UTB, en el AVA SAVIO se pueden resaltar: revisar blog, actualizar mensaje en foro, elegir en Urna Virtual.
- Entre los grupos de estudiantes de Ingeniería de Sistema de la UTB, con mayor participación podemos destacar en primer lugar al cluster 2 y en segundo lugar al cluster 1, donde se ubican los estudiantes con escala de calificación aprobatoria.

6.3. Análisis Inter-Cluster del modelo de agrupamiento definido en la etapa de modelado

En esta sesión de análisis Inter-Cluster se comparan los cluster del modelo de agrupamiento en función de los factores de participación con mayor frecuencia de uso del AVA SAVIO.

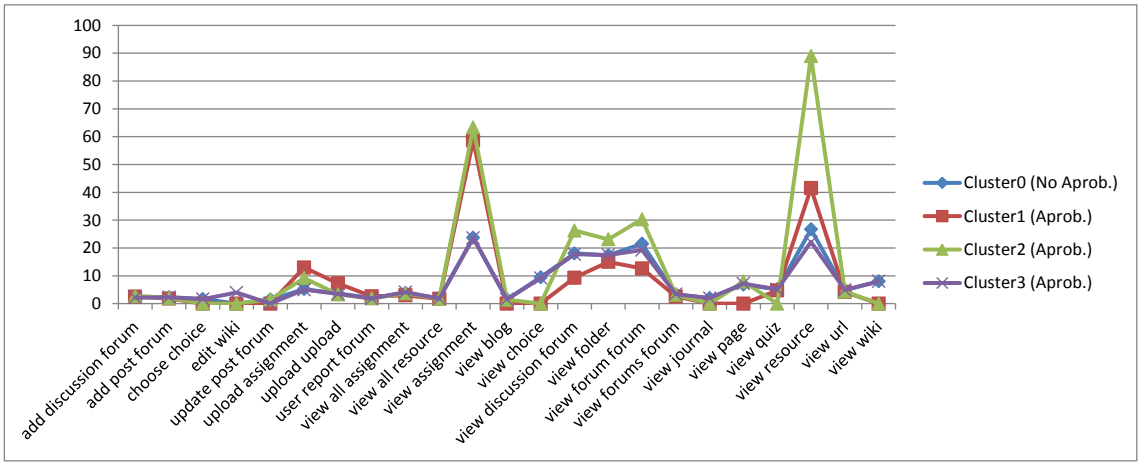

Figura 6.1: Representación de los cluster, seleccionados para el análisis Inter-Cluster.

La figura 6.1 se muestra una selección de los cluster del modelo de entrenamiento formado con el algoritmo K medias en la fase 5 modelado (Metodología CRISP-DM), en función de los factores de participación de los estudiantes de ingeniería de sistema de la UTB en el AVA SAVIO, los valores de los centroides y la escala de calificación final.

En la figura 6.1 se observa que en cada uno de los cluster 0, 1, 2, 3, existe mayor frecuencia por participar en actividades como: enviar tareas (upload assignment), revisar tareas (view assignment), revisar foros de discusión (view discussion fórum), revisar discusiones dentro de un foro (view fórum forum) y revisar recursos (view recurce).

A continuación se muestra un cuadro comparativo tabla 6-16, que permite relacionar los niveles promedio de participación de los estudiantes de ingeniería de sistema (representados por los cluster 0, 1, 2 y 3), en las herramientas didácticas del AVA SAVIO.

La información que se describe en la tabla 6-16, corresponde al modelo generado por la técnica K-medias en la etapa de modelado. En la tabla se analizan 4 cluster que describe el modelo.

Tabla 6-5: Niveles de participación en las herramientas didácticas del AVA SAVIO

Herramientas	Cluster	Factores de Participación	Centroide	Escala de calificación
Tareas	0	*view_assignment*	23,7074	No Aprobatoria
		view_all_assignment	3,9	
		Upload_assignment	5,245	
	1	*view_assignment*	58,6667	Aprobatoria
		view_all_assignment	3	
		Upload_assignment	13	
	2	*view_assignment*	63,2986	Aprobatoria
		view_all_assignment	3,7692	
		Upload_assignment	9,2104	
	3	*view_assignment*	23,6428	Aprobatoria
		view_all_assignment	4,1685	
		Upload_assignment	5,0339	
Foros	0	*view_discussion_forum*	18,0805	No Aprobatoria
		add_discussion_forum	2,3667	
		update_post_forum	1,4917	
		view_forum_forum	21,5426	
		add_post_forum	2,3202	
	1	*view_discussion_forum*	9,3333	Aprobatoria
		add_discussion_forum	2,5556	
		update_post_forum		
		view_forum_forum	12,6667	
		add_post_forum	1,8421	
	2	*view_discussion_forum*	26,2619	Aprobatoria
		add_discussion_forum	2,8205	
		update_post_forum	1,5385	
		view_forum_forum	30,3248	
		add_post_forum	2,1053	
	3	*view_discussion_forum*	17,765	Aprobatoria
		add_discussion_forum	2,2247	
		update_post_forum		
		view_forum_forum	19,2747	
		add_post_forum	2,2762	
Recursos	0	*view_resource*	26,7018	No Aprobatoria
		view_all_resource	1,9245	
	1	*view_resource*	41,5	Aprobatoria
		view_all_resource	1,7157	

	2	*view_resource*	88,9231	Aprobatoria
		view_all_resource	1,7104	
	3	*view_resource*	22,1272	Aprobatoria
		view_all_resource	1,7792	
Wiki	0	*view_wiki*	8	No Aprobatoria
		edit_wiki		
	1	*view_wiki*		Aprobatoria
		edit_wiki		
	2	*view_wiki*		Aprobatoria
		edit_wiki		
	3	*view_wiki*	8	Aprobatoria
		edit_wiki	4	
Cuestionario	0	*view_quiz*	5,1556	No Aprobatoria
	1	*view_quiz*	4,7407	Aprobatoria
	2	*view_quiz*		Aprobatoria
	3	*view_quiz*	5,1061	Aprobatoria
Blog	0	*view_blog*	1,5846	No Aprobatoria
	1	*view_blog*		Aprobatoria
	2	*view_blog*	1,4556	Aprobatoria
	3	*view_blog*	1,5194	Aprobatoria
Folder	0	*view_folder*	17,4205	No Aprobatoria
	1	*view_folder*	15,0064	Aprobatoria
	2	*view_folder*	23,142	Aprobatoria
	3	*view_folder*	17,4784	Aprobatoria
URL	0	*view_url*	4,857	No Aprobatoria
	1	*view_url*	4,1989	Aprobatoria
	2	*view_url*	4,3499	Aprobatoria
	3	*view_url*	4,9409	Aprobatoria
Consulta	0	*view_choice*	9,4944	No Aprobatoria
		choose_choice	1,6722	
	1	*view_choice*		Aprobatoria
		choose_choice		
	2	*view_choice*		Aprobatoria
		choose_choice		
	3	*view_choice*	9,2247	Aprobatoria
		choose_choice	1,6629	
Pagina	0	*view_page*	6,8595	No Aprobatoria
	1	*view_page*		Aprobatoria
	2	*view_page*	7,9451	Aprobatoria

	3	*view_page*	7,2167	Aprobatoria
Diario	0	*view_journal*	2,0167	No Aprobatoria
	1	*view_journal*		Aprobatoria
	2	*view_journal*		Aprobatoria
	3	*view_journal*	1,9888	Aprobatoria

En la tabla 6-5, se presentan los diferentes tipos de acción o participación de los estudiantes de ingeniería de sistema de la UTB, en el Ambiente Virtual de Aprendizaje SAVIO y su respectiva escala de calificación. Se realizó la asignación de los factores de participación (actividades didácticas), con su correspondiente herramienta y escala de calificación en cada cluster. Las celdas vacías en la columna centroides corresponden a valores missing, esto quiere decir a las acciones donde no participaron los estudiantes.

Se puede observar en la tabla 6-16, cuatro tipos de cluster correspondiente a los estudiantes con escala de calificación Aprobatoria y No Aprobatoria. en cada cluster se describe el tipo y frecuencia de uso de las herramientas virtuales asociado a una respectiva escala de calificación.

Podemos observar en la tabla 6-16, que las herramientas didácticas Wiki, Consulta (Choice), Diario (Journal), presentaron poca participación por parte de los estudiantes de ingeniería de sistema de la UTB. En su defecto las herramientas didácticas Recursos, Foro, Tarea, presentaron mayor participación por parte de los estudiantes de ingeniería de sistema de la UTB.

A continuación se hará un análisis de los cluster 0, 1, 2 y 3 del modelo de agrupamiento, a través de gráficos de marcadores y lineales, teniendo en cuanta los factores de acción o participación en el AVA SAVIO. Los gráficos de marcadores estarán representados por los valores de los datos utilizados en el dataset de entrenamiento. Los gráficos lineales estarán representados por los valores de los centroides generados por la técnica k-medias en la etapa de modelado.

En los gráficos de marcadores las instancias estarán etiquetadas con diferentes colores, representando cada color un tipo de escala de calificación, como se muestra en la tabla 6-6.

Tabla 6-6: Representación Escala de Calificación en número y color

Escala	Representación Escala	Color
No Aprobatoria (0.0 a 2.9)	0	Azul
Aprobatoria (3.0 a 5.0)	1	Naranja

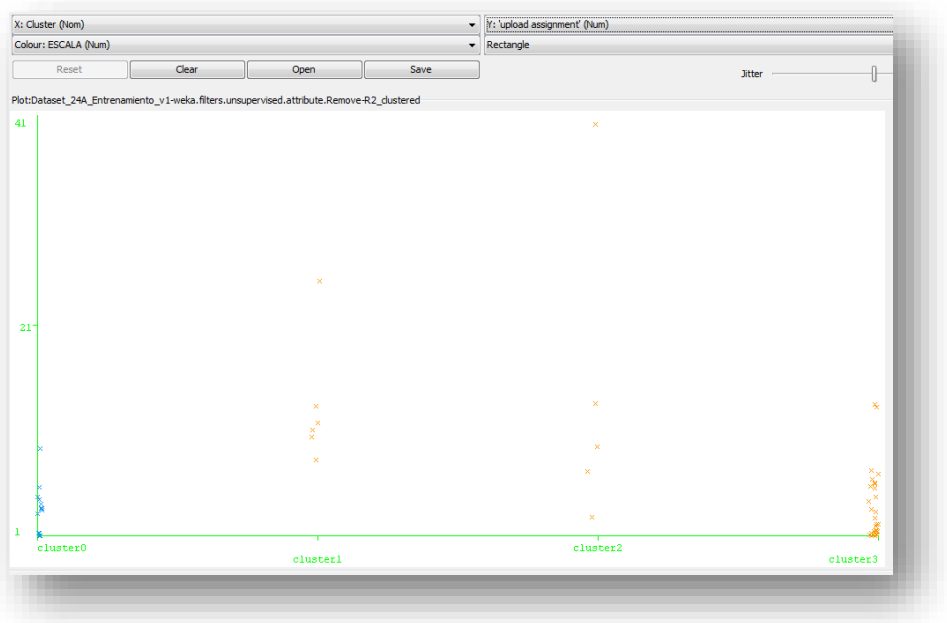

Figura 6.2: Grafico de marcadores del factor de participación 'Enviar tareas'.

Podemos observar en el gráfico de marcador 6.2, en el eje ‘x’ los clúster, que representa los grupos de estudiantes de ingeniería de sistema de la UTB y el eje ‘y’ la frecuencia del factor participación o acción ‘enviar tareas’, las instancias se representan con el color de escala de calificación. Observamos en el grafico aquellos estudiantes con mayor interés o participación en ‘subir tareas’, son aquellos que se ubican en la región o cluster 1 y 2 con escala de calificación ‘Aprobatoria’ (3.0 a 5.0), En este panorama se corrobora la hipostasis nula “los grupos de estudiantes ingeniería de sistema de la UTB que tienen mayor participación en el AVA SAVIO son aquellos que tienen una escala de calificación final aprobatoria”. En el gráfico de marcadores se logra que observar los cluster 0 y 3 contiene mayor número de instancias.

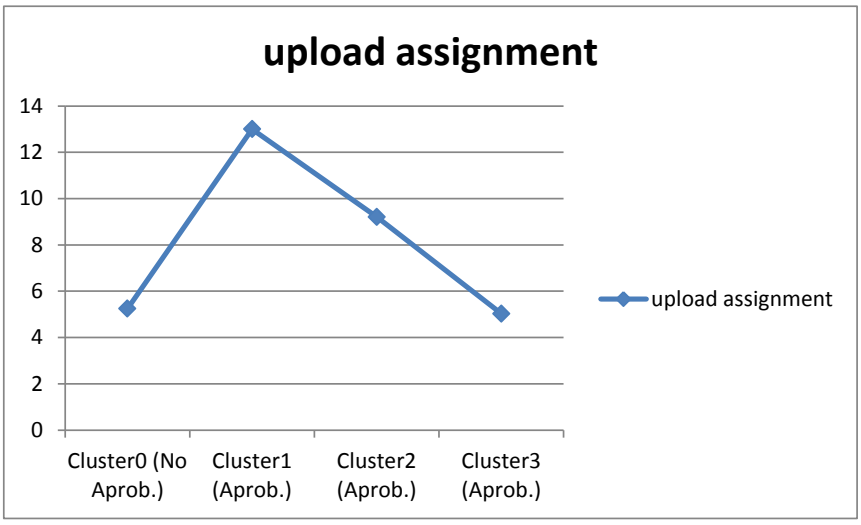

Figura 6.3: Grafico de línea del factor de participación 'Enviar tareas'.

Se observa en el gráfico de líneas con marcador 6.3 generado a través de la herramienta Excel, representando en el eje 'x' los cluster y en el eje 'y' el valor de los centroides para el atributo 'subir tarea', podemos observar en este grafico que los estudiantes que tienen mayor interés o frecuencias son aquellos que se ubican en los cluster 1y 2, se puede ver un comportamiento similar en los cluster representados en el gráfico de marcador.

Otros de los detalles que podemos apreciar en el grafico es la tendencia de los cluster 0 y 3 en participar de forma similar de la actividad 'enviar tarea', una de las causas de este comportamiento puede ser, que los cursos en los que participaron estos grupos (cluster 0 y 3), los docentes tuvieron mayor preferencia por la modalidad de enseñanza aprendizaje presencial que la modalidad de enseñanza aprendizaje virtual.

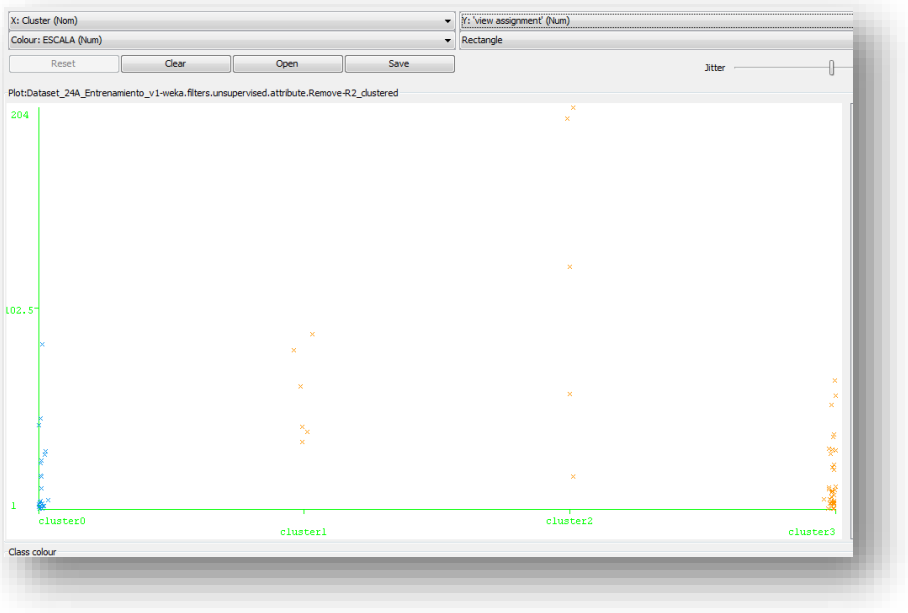

Figura 6.4: Gráfico de marcadores del factor de participación 'Revisar tareas'

Podemos observar en el gráfico de marcador 6.4, en el eje ‘x’ encontramos los clúster, que representa los grupos de estudiantes de ingeniería de sistema de la UTB y el eje ‘y’ la frecuencia del factor participación o acción ‘Revisar ’, las instancias se representan con el color de escala de calificación. Observamos en el grafico aquellos estudiantes con mayor interés o participación en ‘Revisar tareas’, son aquellos que se ubican en la región o cluster 2 y 1 con escala de calificación ‘Aprobatoria’ (3.0 a 5.0). Algo para resaltar en el grafico es la similitud en la frecuencia de participación de las instancias que se ubican en el cluster 0 y cluster 3.

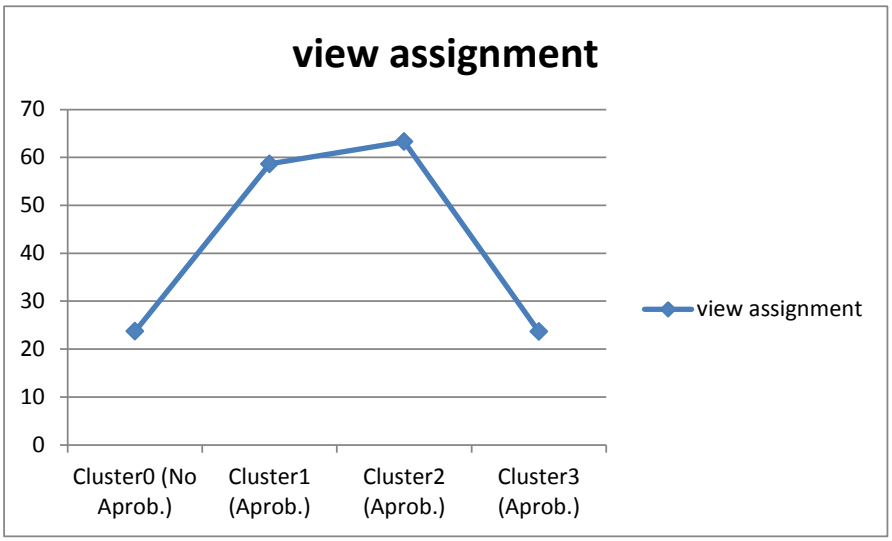

Figura 6.5: Grafico de línea del factor de participación 'Revisar tareas'

Se observa en el gráfico de líneas con marcador 6.5, generado a través de la herramienta Excel, representando en el eje 'x' los cluster y en el eje 'y' el valor de los centroides para el atributo 'Revisar tareas', podemos observar en este grafico que los estudiantes que tienen mayor interés o frecuencias son aquellos que se ubican en los cluster 2 y 1, se puede ver un comportamiento similar en los cluster representados en el gráfico de marcadores 6.4.

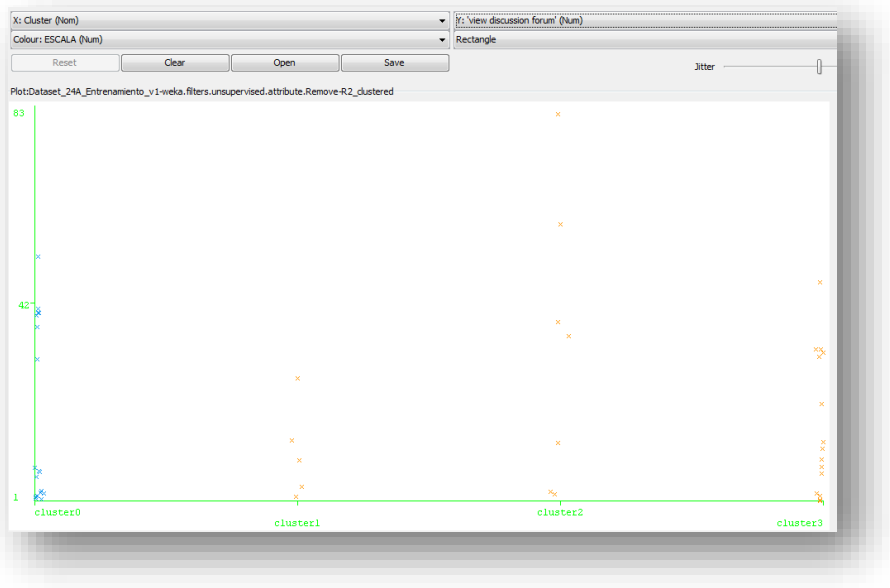

Figura 6.6: Grafico de marcador del factor de participación 'revisar foro de uso general'

Podemos observar en el gráfico de marcador 6.6, en el eje ‘x’ encontramos los clúster, que representa los grupos de estudiantes de ingeniería de sistema de la UTB y el eje ‘y’ la frecuencia del factor participación o acción ‘ver tareas’, las instancias se representan con el color de escala de calificación. Observamos en el grafico aquellos estudiantes con mayor interés o participación en revisar ‘foros de uso general’, son aquellos que se ubican en la región o cluster 2 con escala de calificación ‘Aprobatoria’ (3.0 a 5.0). En el gráfico de marcadores vemos las instancias que se ubican en el cluster 1 con menor participación.

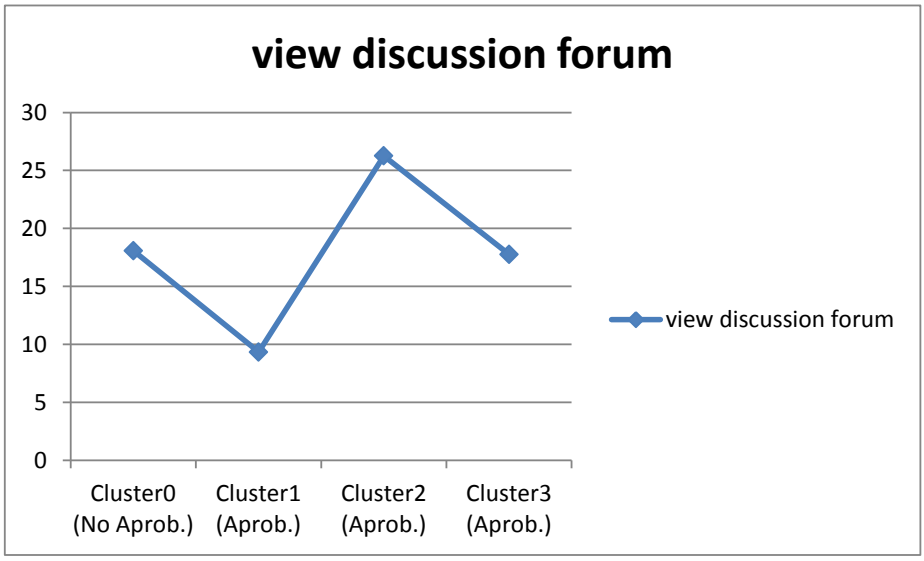

Figura 6.7: Grafico de línea del factor de participación ‘revisar foro de uso general’

Se observa en el gráfico de líneas con marcador 6.7, generado a través de la herramienta Excel, representando en el eje ‘y’ los valores de los centroides del factor ‘revisar foro de uso general’, se muestra que los estudiantes ubicados en los cluster 2, tienen mayor interés o frecuencia por revisar foro de uso general, en el caso contrario vemos que los estudiantes ubicados en el cluster 1 tienen menor interés o frecuencia en revisar foro de uso general, los cluster 0 y 3 mantienen un interés o frecuencia similar. Se puede ver un comportamiento similar en los cluster representados en el gráfico de marcador 6.6.

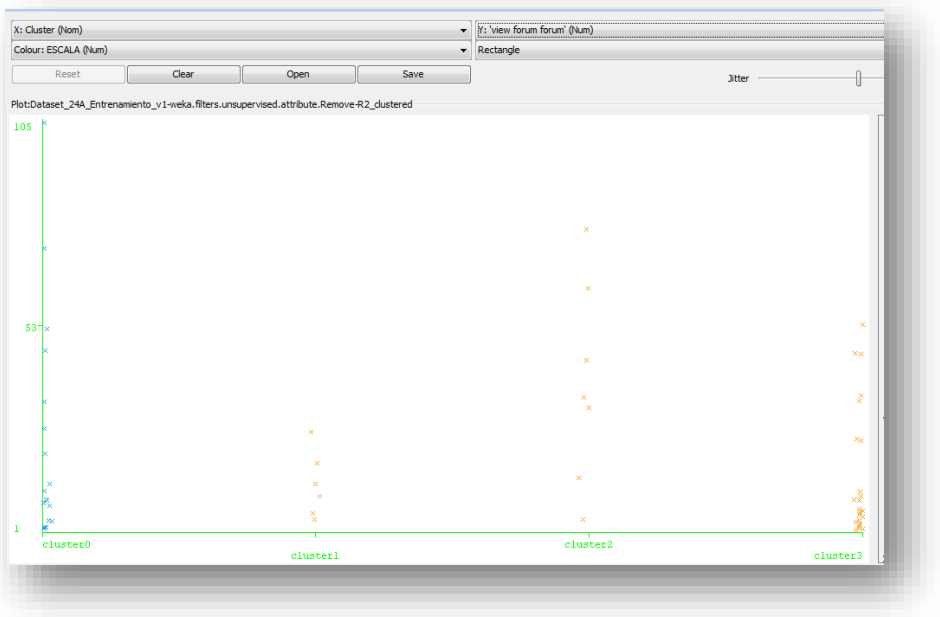

Figura 6.8: Gráfico de marcador del factor de participación 'Revisar foro debate sencillo'

Podemos observar en el gráfico de marcador 6.8, en el eje ‘x’ encontramos los clúster, que representa los grupos de estudiantes de ingeniería de sistema de la UTB y el eje ‘y’ la frecuencia del factor participación o acción ‘ver o revisar foros de debate sencillo’, las instancias se representan con el color de escala de calificación. Observamos en el grafico aquellos estudiantes con mayor interés o participación en ‘revisar foro de debate sencillo ’, son aquellos que se ubican en la región o cluster 2 con escala de calificación ‘Aprobatoria’ (3.0 a 5.0). Algo para resaltar en el grafico es la similitud en la frecuencia de participación de las instancias que conforman el cluster 0 y 3, otro detalle que podemos ver en el grafico es el bajo interés o frecuencia de las instancias que se ubican en el cluster 1.

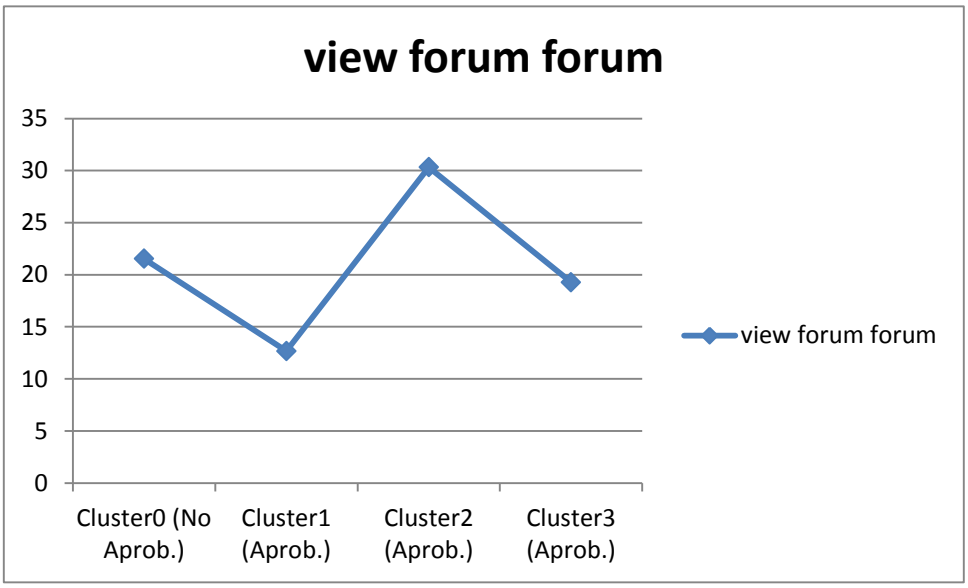

Figura 6.9: Gráfico de línea del factor de participación 'Revisar foro debate sencillo'

Se observa en el gráfico de líneas con marcador 6.9, generado a través de la herramienta Excel, representando en el eje 'x' los cluster y en el eje 'y' el valor de los centroides para el atributo 'ver foro de debate sencillo', podemos observar en este grafico que los estudiantes que tienen mayor interés o frecuencias son aquellos que se ubican en los cluster 13, 6 y 1, se puede ver un comportamiento similar en los cluster representados en el gráfico de marcador 6.8.

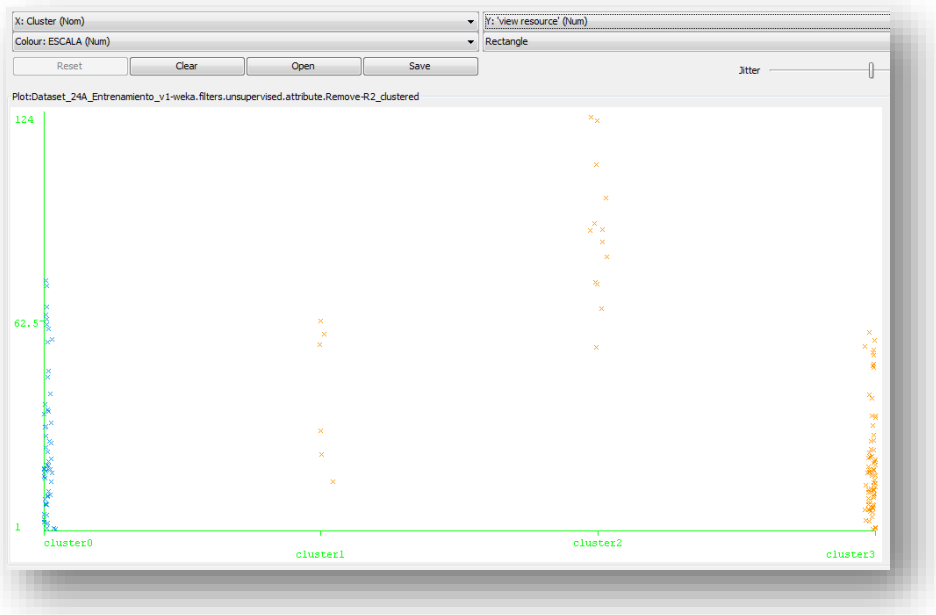

Figura 6.10: Grafico de marcador del factor de participación 'Revisar recursos'

Podemos observar en el gráfico de marcador 6.10, en el eje ‘x’ encontramos los clúster, que representa los grupos de estudiantes de ingeniería de sistema de la UTB y el eje ‘y’ la frecuencia del factor participación o acción ‘ver o revisar recursos’, las instancias se representan con el color de escala de calificación. Observamos en el grafico aquellos estudiantes con mayor interés o participación en ‘revisar recursos’, son aquellos que se ubican en la región o cluster 2 con escala de calificación ‘aprobatoria’ (3.0 a 5.0). Se muestra en el grafico un mayor número de instancias ubicadas en los cluster 0 y 3.

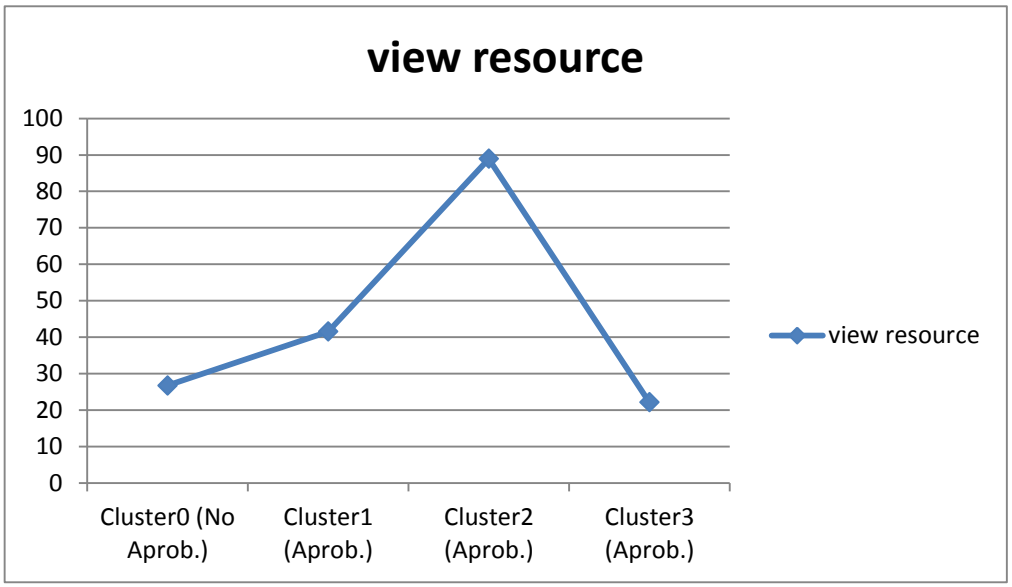

Figura 6.11: Gráfico de línea del factor de participación 'Revisar recursos'

Se observa en el gráfico de líneas con marcador 6.12, generado a través de la herramienta Excel, representando en el eje ‘x’ los cluster y en el eje ‘y’ el valor de los centroides para el atributo ‘ver o revisar recursos’, podemos observar en este grafico que los estudiantes que tienen mayor interés o frecuencias en ‘Revisar recursos’, son aquellos que se ubican en los cluster 2, seguido del cluster 1, se puede ver un comportamiento similar en los cluster representados en el gráfico de marcador 6.11.

6.3.1. DETALLES DEL ANÁLISIS INTER-CLUSTER

Luego de observar todos los gráficos de marcador y líneas que representaban los cluster y la frecuencia en el uso de la plataforma SAVIO, podemos resaltar los siguientes detalles:

- De los cuatro cluster que conforman el modelo de agrupamiento, el cluster 2 representados por los estudiantes de ingeniería de sistema de la UTB con escala de calificación aprobatoria, mantuvo un mayor interés o frecuencia en participar de la plataforma o AVA SAVIO. Este panorama corrobora la hipostasis nula "los grupos de estudiantes ingeniería de sistema de la UTB que tienen mayor partición en el AVA SAVIO, son aquellos que tienen una escala de calificación final aprobatoria".
- En los cluster 0 (representados por los estudiantes de ingeniería de sistema con escala de calificación no aprobatoria) y cluster 3 (representados por los estudiantes de ingeniería de sistema con escala de calificación aprobatoria), se mantuvo un comportamiento similar en participar de la plataforma o AVA SAVIO, una de las causas de este comportamiento puede ser, que los cursos en los que participaron estos grupos (cluster 0 y 3), los docentes tuvieron mayor preferencia por la modalidad de enseñanza aprendizaje presencial que la modalidad de enseñanza aprendizaje virtual.

6.4. ANÁLISIS DE RESULTADOS

Después de analizar a nivel Intra-Cluster e Inter-Cluster el modelo de agrupamiento, se observó que los cluster 0 (conformados por los estudiantes con escala de calificación no aprobatoria) y cluster 3 (conformados por los estudiantes con escala de calificación aprobatoria), tuvieron un mayor cubrimiento en las instancias que representaban a los estudiantes de ingeniería de sistema de la UTB, también se pudo observar que ambos clúster tuvieron una tendencia muy parecida en el uso de la plataforma SAVIO, este escenario nos obliga a pensar en la situación de una falta de correlación entre los factores de participación y la escala de calificación, una posible causa a estos hechos es que los cursos seleccionados para el estudio no aplicaron una modalidad 100% B-learning, en otras palabras, los docentes a cargo de los cursos (Algoritmos, Estructura de datos, Programación), tuvieron mayor preferencia por el proceso de enseñanza aprendizaje presencial , que el proceso de enseñanza aprendizaje virtual.

Se debe tener presente que la fiabilidad y efectividad del modelo de agrupamiento analizado podría ser muy superior, si se analiza una cantidad de instancias considerablemente mayor, y si además se eliminan atributos que manejen rangos muy grandes de valores posibles.

Capítulo 7

7. Conclusiones y Trabajo Futuro

7.1. Conclusiones

En este proyecto se ha definido un modelo descriptivo a través de algoritmos KDD, que permite agrupar a los estudiantes de ingeniería de sistema de la Universidad Tecnológica de Bolívar, en varios grupos, basados en una selección determinada de cursos B-learning (Estructura de Datos, Algoritmos, Programación), relacionados con las actividades realizadas en el AVA SAVIO y las escala de calificación final. Una vez definido el modelo de agrupación, se aplicaron técnicas de evaluación que permitieron determinar su efectividad para la explicación de los datos de estudios.

Las actividades realizadas en la etapa de modelado y evaluación ayudaron en la selección adecuada del modelo agrupamiento y en el análisis del comportamiento de los datos de estudios, representados a través de tablas, gráficos de marcador y de líneas.

Los datos de estudio utilizados en este trabajo corresponden a las acciones realizadas por los estudiantes de ingeniería de sistema de la UTB, en los cursos b-learning (Estructura de datos, Algoritmo, programación) y sus calificaciones finales, esta información fue suministradas por la dirección de SAVIO y de SIRIUS de la Universidad Tecnológica de Bolívar (UTB).

En la etapa de modelado se llevó a cabo dos tipos de experimentos, de los cuales se seleccionó el modelo del segundo experimento, puesto que permitió agrupar de forma acertada, precisa y concisa a los estudiantes de ingeniería de la UTB, en función de los dos tipos de escala de calificación (Aprobatoria y No Aprobatoria) y los factores de participación (actividades didácticas) en el AVA SAVIO. Dentro de las estrategias aplicadas en el segundo experimento estuvo la implementación de los algoritmos de minería de datos, EM para determinar de forma automática el número de cluster y K-medias para definir el modelo de agrupamiento a partir del número de cluster obtenidos por la técnica EM.

En la etapa de evaluación se aplicaron técnicas encontradas en la revisión del estado del arte, tales como la métrica log- likelihood y la aproximación particionamiento (Entrenamiento / Prueba) utilizadas en tareas de clasificación y regresión, a partir de estas técnicas se determinó que el modelo de agrupamiento era adecuado para el diagnóstico de los datos extraídos del Ambiente Virtual Aprendizaje SAVIO.

En la etapa de implementación (Modelo CRISP-DM), se realizó un análisis del modelo agrupamiento a través de varios tipos de representaciones, en primer lugar se describieron cuadros comparativos donde se clasificaban cada uno de los 4 cluster con sus respectivas factores de participación, valores de los centroides y la escala de calificación. En segundo lugar se describió un cuadro comparativo donde se clasificaron cada uno de los tipos de herramientas didácticas con sus respectivos cluster, tipos de acción, valores de los centroides y tipos de escala de calificación.

Después de realizar el análisis Intra-cluster con base en los 4 cluster del modelo agrupamiento, se puede concluir lo siguiente:

- Dentro de las actividades más utilizadas por los grupos estudiantes de Ingeniería de Sistema de la UTB, en el AVA SAVIO se pueden destacar: el revisar recursos, revisar tareas y revisar foros.
- De las actividades menos usadas por los grupos estudiantes de Ingeniería de Sistema de la UTB, en el AVA SAVIO se pueden resaltar: revisar blog, actualizar mensaje en foro, elegir en Urna Virtual.
- Entre los grupos de estudiantes de Ingeniería de Sistema de la UTB, con mayor participación podemos destacar en primer lugar al cluster 2 y en segundo lugar al cluster 1, donde se ubican los estudiantes con escala de calificación aprobatoria.

A través de los gráficos generados por la herramienta de minería de datos WEKA y Excel se llevó a cabo un estudio correlacional del comportamiento de los estudiantes del programa de ingeniería de sistemas de la Universidad Tecnológica de Bolívar (UTB) en función de los factores que representan su participación en los cursos b-learning (Programación, Estructura de Datos, Algoritmos) y la escala de calificación final obtenida.

En el segundo análisis Inter-Cluster realizado en la etapa de implementación (Modelo CRISP-DM) se hizo una comparación de los cluster 0, 1, 2 y 3 del modelo de agrupamiento, se encontró que los estudiantes con mayor interés por participar de las actividades (enviar y revisar tareas, revisar foros y revisar recursos) en el AVA SAVIO son aquellos que se agrupan en el cluster 2 con escala de calificación aprobatoria (3.0 a 5.0), este acontecimiento corrobora lo establecido en la hipótesis de investigación "los grupos de estudiantes de ingeniería de sistema de la UTB que tienen mayor interés por participar del AVA SAVIO, son aquellos que obtienen una calificación final aprobatoria".

En el modelo de agrupamiento generado en la etapa de modelado (Modelo CRISP-DM), se observó varios grupos de estudiantes del programa de ingeniería de sistema de la UTB (Cluster 0 y cluster3), con diferentes escala de calificación y comportamientos similares en función de la participación en el Ambiente Virtual de Aprendizaje SAVIO, una de las posibles causas de este acontecimiento se debe a que los cursos seleccionados para el estudio no aplicaron una modalidad 100% B-learning, en otras palabras, los docentes a cargo de los cursos (Algoritmos, Estructura de datos, Programación), tuvieron mayor preferencia por el proceso de enseñanza aprendizaje presencial, que el proceso de enseñanza aprendizaje virtual.

Se debe tener presente que la fiabilidad y efectividad del modelo de agrupamiento analizado podría ser muy superior, si se analiza una cantidad de instancias considerablemente mayor, y si además se eliminan atributos que manejen rangos muy grandes de valores posibles.

7.2. Recomendación

En el primer análisis realizado en la etapa de implementación, se observó el comportamiento de 4 cluster, que representaban la participación de los estudiantes de ingeniería de sistema de la UTB en el AVA SAVIO y su escala de calificación final. Estas observaciones permitieron establecer una serie de estrategias basadas en los trabajos de Norma Scagnoli (2005) [46], que posiblemente puedan motivar a los estudiantes en el buen uso del AVA SAVIO.

En cuanto al cluster 0 que agrupan a los estudiantes que obtuvieron escala de calificación no aprobatoria, se sugiere que el docente deba implementar las siguientes estrategias:

- Diseñar Material Educativo y seleccionar recursos de apoyo interesante y variado relacionado con los diversos intereses y necesidades de los estudiantes.
- Evaluar un mismo aprendizaje en varios tipos de actividades virtuales (Foros, Blog, Tareas), esto con fin de darle oportunidad al estudiante que se adapte al proceso de aprendizaje virtual que más se le facilite.
- Adaptar los contenidos al nivel de conocimientos de los estudiantes, a través del uso de actividades didácticas que se complementen con recursos de multimedia tales como voz o video.
- Actuar con paciencia, dando a otros y a si mismo tiempo para procesar la información, puesto que no todos los estudiantes cuentan con una computadora en sus espacios de trabajo o en el hogar.
- Retroalimentar al estudiante de manera permanente para que reconozca sus cualidades y habilidades reales (Satisfacción de necesidad de Estima) a través de la retroalimentación.
- Enviar recordatorios de las actividades o materiales establecidas a los estudiantes, con el fin de alertarlos con la fechas de cumplimiento de las actividades.
- Socializar o mostrar la metodología de trabajo a seguir y sus ventajas, mostrando también cierta flexibilidad para poder realizar posibles cambios en la programación.

En cuanto a los cluster 1, 2 y 3 que agrupan a los estudiantes que obtuvieron escala de calificación aprobatoria, se sugiere que el docente deba implementar las siguientes estrategias:

- Plantear retos, problemas de cierta dificultad, debates. A través del uso de herramientas virtuales como 'foro de debate general' para que los estudiantes puedan crear nuevas preguntas, compartir ideas. El uso de esta estrategia hace que el proceso de enseñanza aprendizaje sea flexible, dándole la oportunidad al estudiante de ser escuchado.
- Ofertar puntos de calificación para estimular a los estudiantes en el uso de las herramientas didácticas: wiki, blog, consulta.
- Animar y alentar la participación, el diálogo, la intercomunicación, el intercambio, a través del uso de herramientas como foro, correo en el AVA SAVIO.
- Comenzar con actividades moderadamente difíciles en el que el reto pueda ser superado sin complicaciones y continuar con actividades relativamente difíciles.
- Intervenir y contestar de manera rutinaria ante cualquier opinión o inquietud de los estudiantes.
- Guiar al alumno a que establezca metas específicas a un mediano y largo plazo durante el transcurso de la materia.
- Promover actividades formativas, que permitan modelar paso a paso la forma de realizar las actividades creativas o investigativas, tratando de lograr de forma significativa cumplir con cada uno de los objetivos de aprendizaje establecidos para la respectiva unidad de estudio.
- Reconocer el esfuerzo de manera grupal e individual de cada uno de los estudiantes es esencial para potencializar las habilidades, además de incrementar la participación.

7.3. Trabajo futuro

Para trabajos futuros se tiene pensado realizar un estudio para el desarrollo de un módulo que se pueda adaptar en el Ambiente Virtual de Aprendizaje (AVA) SAVIO, que incluya herramientas de minería de datos de agrupamiento, donde se describa la correlación entre factores de participación de los estudiantes en el AVA y la escala de calificación, apoyando al docente en el análisis y personalización de las herramientas didácticas, con el fin de cumplir con los gustos o expectativas de los estudiantes.

8. Bibliografía

1. WBT Systems. (2012). WBT Systems. Recuperado el 10 de 12 de 2012, de ELearning & Event Management made simple: http://www.wbtsystems.com/product/

2. Blackboard Inc. (1997). Blackboard. Recuperado el 10 de 12 de 2012, de http://www.blackboard.com

3. Dokeos. (2012). Dokeos. Recuperado el 10 de 12 de 2012, de http://www.dokeos.com/

4. ATutor. (2012). ATutor Learning Management Tools. Recuperado el 10 de 12 de 2012, de http://atutor.ca/

5. ILIAS. (2012). ILIAS Open Source e-Learning. Recuperado el 10 de 12 de 2012, de https://www.ilias.de/docu/

6. Moodle. (2012). Moodle. Recuperado el 10 de 12 de 2012, de http://moodle.org/

7. Consortium Claroline. (2012). Claroline - Easy & Flexible Learning Solutions. Recuperado el 10 de 12 de 2012, de http://www.claroline.net/

8. Observatorio de E-Learning (2008). Plataformas LMS vs LCMS, www.aulaglobal.net.ve/ observatorio/, Venezuela.

9. Romero, C., Ventura, S., & García, E. (s.f.). (2008). Data mining in course management systems: Moodle case study and tutorial. Elsevier

10. García, E., Romero, C., Castro, C. D., & Ventura, S. (2006). Usando Minería de datos para la Continua Mejora de cursos de e-learning. Escuela Politécnica Superior Universidad de Córdoba.

11. Brusilovsky, P., Eklund, J., & Schwarz, E. (1998). Web-based education for all: a tool for development adaptive courseware. Computer Networks and ISDN Systems, 291-300.

12. Weber, G., & Brusilovsky, P. (2001). ELM-ART: An Adaptative Versatile System for Web-based Instruction. International Journal of Artificial Intelligence in Education, 351-384.

13. De Bra, P., & Stash, N. (2002). AHA! Adaptative Hypermedia for All. Second International Conference on Adaptative Hypermedia and Adaptative Web- Based Systems, 381-384.

14. Romero, C., Ventura, S., & García, E. (2008a). Data mining in course management systems: Moodle case study and tutorial. Computers & Education, 51(1) 368-384.

15. Aponte Novoa, Hoyos Pineda, Monsalve Pulido (2012). Minería de usabilidad aplicada a plataformas virtuales de aprendizaje. 27-39.

16. Marín, Ramírez y Sampedro, (2011). MOODLE Y ESTUDIANTES UNIVERSITARIOS.DOS NUEVAS REALIDADES DEL EEES.

17. Jose Hernandez Orallo, Jose Martinez Quintana & Cesar Ferri Ramirez: Introducción a la Minería de Datos Pearson Education S.A., Madrid 2004. Bibliografía ISBN 978-84-205-4091-7 [capitulos 1, 2, 4, 5 &16].

18. Romero, C., Ventura, S., & Hervás, C. (2005). Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basados en web. III Taller de Minería de Datos y Aprendizaje, TAMIDA. Thomson.

19. DBMiner Technology Inc. (2012). DBMiner. Recuperado el 10 de 12 de 2012, de http://www.dbminer.com/

20. IBM. (2012). SPSS Modeler. Recuperado el 10 de 12 de 2012, de http://www-01.ibm.com/software/analytics/spss/products/modeler/

21. IBM. (2012). DB2 Intelligent Miner. Recuperado el 10 de 12 de 2012, de http://publib.boulder.ibm.com/infocenter/db2luw/v8/index.jsp?topic=/com.ibm.db2.udb.doc/wareh/getsta06im.htm

22. SAS. (2012). SAS - The Power to Know. Recuperado el 10 de 12 de 2012, de http://www.sas.com/technologies/analytics/datamining/miner/

23. StatSoft Inc. (2012). Statsoft. Recuperado el 10 de 12 de 2012, de http://www.statsoft.com/#

24. Keel. (2004). Knowledge Extraction based on Evolutionary Learning. Recuperado el 10 de 12 de 2012, de http://www.keel.es/

25. The University of Waikato. (2012). WEKA. Recuperado el 10 de 12 de 2012, de http://www.cs.waikato.ac.nz/ml/weka/

26. Orange. (2012). Orange. Recuperado el 21 de 04 de 2012, de http://orange.biolab.si/

27. Rapid-i. (2012). Rapid-i - Report the future. Recuperado el 10 de 12 de 2012, de http://rapid-i.com/content/view/181/196/

28. jHepWork. (2012). jHepWork - Multiplatform environment for scientific computation and data analysis. Recuperado el 10 de 12 de 2012, de http://jwork.org/jhepwork/

29. Knime. (2012). Knime. Recuperado el 10 de 12 de 2012, de http://www.knime.org/

30. Rozita, J. (2010). Differential Internet Behavior's of Students from Gender Groups. Computer Science & Engineering Department.

31. Diego García Saiz y Marta Zorrilla, (2011). Hacia la minería de datos sin parámetros y su aplicación en el campo educativo.

32. Cristóbal Romero Morales, Sebastián Ventura Soto, and Cesar Hervas Martínez. Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. Actas del III Taller Nacional de Minería de Datos y Aprendizaje, TAMIDA2005, pages 49-56, 2005.

33. Plataforma SAVIO (2013). Presentación de la Plataforma SAVIO, recuperado 22 de 07 de 2013, http://www.unitecnologica.edu.co/programas-acad%C3%A9micos/educaci%C3%B3n-virtual

34. Boletín Electrónico UTB (2012), Presentación de la nueva versión del sistema de información institucional SIRIUS II, recuperado 22 de 07 de 2013, http://www.unitecnologica.edu.co/medios-y-boletines/boletines-electr%C3%B3nicos/bolet%C3%ADn-institucional/2012/mayo-2012/semana-del-14-al-18/sirius-ii-llega-la-utb

35. Reglamento UTB (2013). Reglamento Estudiantil para Pregrado Nov de 2012. Recuperado el 22 de 07 de 2013, de http://www.unitecnologica.edu.co/sites/default/files/Reglamento_Estudiantil_pregrado_-_Nov_2012.pdf

36. Ros, I. (2008). Moodle, la plataforma para la enseñanza y organización escolar . Ikastorratza, e- Revista , http://www.ehu.es/ikastorratza/2_alea/moodle.pdf.

37. Paiva Franco. (2010). A Plataforma Moodle como Alternativa para uma Educação Flexível. Ikastorratza, Revista EducaOnline Volume 4 - No 1- Janeiro/Abril de 2010, Universidade Federal do Rio de Janeiro (www.lingnet.pro.br).

38. María Sarango (2012), Aplicación de técnicas de minería de datos para identificar patrones de comportamientos relacionados con las acciones del estudiante con el EVA de la UTPL.

39. Fernando Berzal, Mineria de Datos [en linea] <http://elvex.ugr.es/idbis/dm/slides/4%20Clustering.pdf > [citado en Septiembre de 2013]

40. Clustering (2013). Recuperado el 15 de 12 de 2013, de http://elvex.ugr.es/idbis/dm/slides/4%20Clustering.pdf

41. Minería de Datos en Sistemas Educativos (2013). Recuperado el 15 de 12 de 2013, de http://sci2s.ugr.es/docencia/doctoM6/EducationalDataMining.pdf

42. Fernando Vera (2014). La Modalidad Blended-Learning En La Educación Superior. Recuperado el 2 de 03 de 2014, de http://www.utemvirtual.cl/nodoeducativo/wp-content/uploads/2009/03/fvera_2.pdf

43. M. Garre, J. J. Cuadrado, M. Sicilia, D. Rodríguez, R. Rejas (2007), Comparación de diferentes algoritmos de clustering en la estimación de coste en el desarrollo de software. Recuperado el 9 de 03 de 2014, de https://www.ati.es/IMG/pdf/GarreVol3Num1.pdf

44. Garre M., Cuadrado J.J., Sicilia, M.A., Charro M. y Rodríguez D., "Segmented Parametric Software Estimation Models: Using the EM algorithm with the ISBSG 8 database", Information Technology Interfaces, Croacia, 20-23 junio 2005.

45. D. López De Luise y M. Soffer (2008). Modelización automática de textos en castellano. Recuperado el 9 de 03 de 2014. http://www.palermo.edu/ingenieria/PDFs/2008_ANDESCONManejoTextosCastellano_DLL_MS_v06.pdf

46. Norma Scagnoli (2005). Estrategias para Motivar el Aprendizaje Colaborativo en Cursos a Distancia. Recuperado el 9 de 03 de 2014. https://ideals.illinois.edu/bitstream/handle/2142/10681/aprendizaje-colaborativo-scagnoli.pdf?sequence=2

9. Índice Analítico