

Uso de herramienta libre para la generación de reglas de asociación, facilitando la gestión eficiente de incidentes e inventarios.

Ing. Corso, Cynthia Lorena, Ing. Gibellini Fabián
Departamento de Ingeniería en Sistemas de Información.
Laboratorio de Sistemas de Información.
Facultad Regional Córdoba/Universidad Tecnológica Nacional.
Maestro M. López esq. Cruz Roja Argentina. Ciudad Universitaria
cynthia@bbs.frc.utn.edu.ar
speker@bbs.frc.utn.edu.ar

RESUMEN

El propósito de este trabajo es la presentación de una propuesta metodológica dentro de la rama de la minería de datos, relacionada con la generación de patrones de comportamientos de los incidentes de los equipos informáticos, detectados en el contexto del Laboratorio de Sistemas de Información, perteneciente al Departamento de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Córdoba. A partir de la generación del modelo de conocimiento, el objetivo es brindar información estadística relacionada con los aspectos que tienen mayor incidencia en el reporte de incidentes de equipos informáticos, permitiendo elaborar un plan de prevención para disminuir la generación de reportes de incidentes y recomendaciones para una gestión eficiente de los inventarios relacionado con los componentes y piezas utilizando en el proceso de mantenimiento.

Palabras claves

Minería de Datos, Laboratorios Informáticos, Incidentes de equipos informáticos, Herramientas Informáticas Libre de Aprendizaje Automático, Algoritmos de minería de datos supervisados, Reglas de Asociación.

INTRODUCCIÓN

El laboratorio de Sistemas de Información depende del Departamento de Sistemas de Información, de la Universidad Tecnológica Nacional Facultad Regional Córdoba.

El mismo tiene un conjunto de funciones, entre las que se puede mencionar:

- Brindar la infraestructura adecuada para el dictado de clases en los laboratorios de las distintas cátedras de la carrera de Ingeniería en Sistemas de Información, como así también los cursos de capacitación y extensión que dependen del Departamento de Sistemas.

- O Proporcionar a los alumnos de la carrera de Ingeniería en Sistemas, un espacio denominado prácticas libres. Lo que da la posibilidad a los alumnos de usar las aulas de laboratorios disponibles para la realización de actividades académicas.
- O Otro servicio que brinda el laboratorio es el de impresión de trabajos académicos de los alumnos.
- O Personal del laboratorio brinda asesoramiento técnico tanto a los alumnos como a los docentes de la cátedra.

Las áreas funcionales que están involucradas son:

- O **Encargados de Turno:** El laboratorio de sistemas funciona durante todo el día. El horario de apertura es a las 8:00hs y su cierre es a las 23:00hs. Existen tres encargados asignados a cada turno. Las funciones que cubre es la supervisión a los encargados de aula, como así también la registración de reservas de aulas para cátedras y cursos.
- O **Encargados de Aula:** existen uno o dos encargados por turno, están bajo la supervisión de los encargados de Aula. Dentro de sus funciones: es el acondicionamiento de las aulas para el dictado de cátedras o cursos, impresión de trabajos, reporte y registración de planilla de incidentes.
- O **Operadores de red:** su función principal es lograr que las aulas que dispone el laboratorio de sistemas se encuentren sin ningún inconveniente técnico y asegurar el correcto funcionamiento de la red.
- O **Área Técnica y Mantenimiento:** dentro de las funciones que tiene es el mantenimiento y reparación de cualquier equipamiento informático perteneciente al laboratorio, como así también es el responsable de gestionar todo lo relacionado con los inventarios.
- O **Compras:** esta área es la responsable de gestionar las compras de insumos de uso habitual en el laboratorio de sistemas, como así también todo lo relacionado con los inventarios de los equipos informáticos.

Este trabajo tiene como foco principal atender una de las necesidades detectadas precisamente en el Área Técnica; el personal involucrado manifiesta que en la actualidad no existe ninguna herramienta automatizada que permita la registración de los eventos de fallos y mantenimiento de los mismos. Y en este punto surge otra problemática detectada que está muy relacionada con los fallos que es la gestión de inventarios.

Los inventarios existencias o stock son los componentes o piezas que el laboratorio tiene almacenado con el objetivo de facilitar la continuidad del funcionamiento normal de los laboratorios.

La gestión de inventarios tiene como objetivo determinar la cantidad de existencias que se han de mantener y el ritmo de demanda de los pedidos para cubrir necesidades. Como consecuencia del relevamiento realizado al personal del Área Técnica nos reportan que no existe un plan formal para efectivizar los pedidos de compra de componentes o piezas, en muchas ocasiones en función de los reportes de fallos hacen una estimación aproximada de los que en realidad puede ser un requerimiento de compra.

Otra de las falencias detectadas que en la actualidad no existe ninguna herramienta que permita conocer cuál es el stock real de los componentes o piezas, lo que trae aparejado una serie de desventajas o costos para el laboratorio de sistemas.

[1] **Costos de pedido:** costos de realizar un pedido: administrativos (gestión con proveedores), transporte, descarga, seguros etc. Existe relación inversa al volumen de inventarios, porque cuanto mayor volumen de existencias menor número de pedidos a realizar en el año.

Costos de almacenamiento: costos de mantener las existencias en depósito: espacio, administrativos (personal y sistema gestión), económicos (obsolescencia, depreciación), financieros (intereses de financiar capitales invertidos)

Costos de ruptura de stock: costos que tiene la empresa cuando se queda sin existencias, no puede producir o no puede entregar el pedido a un cliente.

De los requerimientos detectados en esta área en el contexto del laboratorio de sistemas, surge la exposición de una propuesta metodológica relacionada con la rama de la minería de datos que nos permita encontrar relaciones o patrones significativos que permitan facilitar la gestión de inventarios.

METODOLOGÍA

Para llevar a cabo este proyecto se han cumplimentado las siguientes fases:

- 1.** Detección de documentación o reportes: En esta fase se ha identificado los formularios que están involucrados en la gestión de las tareas de mantenimiento y gestión de inventarios. Uno de ellos es la Planilla de Incidentes, que se muestra en la sección siguiente, en la misma se realiza un registro de los movimientos relacionados con los mantenimientos realizados por el personal de área técnica a los recursos informáticos. En lo que respecta a la gestión de inventarios se lleva un registro e En ambos casos esta tarea se realiza de manera manual sin la existencia de ninguna aplicación que facilite la automatización de las mismas.

Incidente		Nro.:	
Fecha:/...../.....		Hora::.....	
Reportado por		Area:	
Apellido y Nombre:		Firma:	
Estación(es):		Aula(s):	
Detalle:			
.....			
.....			
.....			
.....			
.....			
Incluir: cuándo sucedió (hora, cátedra, etc.), cómo sucedió, mensaje de error o del sistema, etc.			
Intervención 1		Fecha:/...../.....	
Area:		Apellido y Nombre:	
Problema resuelto:		SI NO	
Detalle de la acción:			
.....			
.....			
.....			
Intervención 2		Fecha:/...../.....	
Area:		Apellido y Nombre:	
Problema resuelto:		SI NO	
Detalle de la acción:			

Figura 2: Planilla de Incidentes

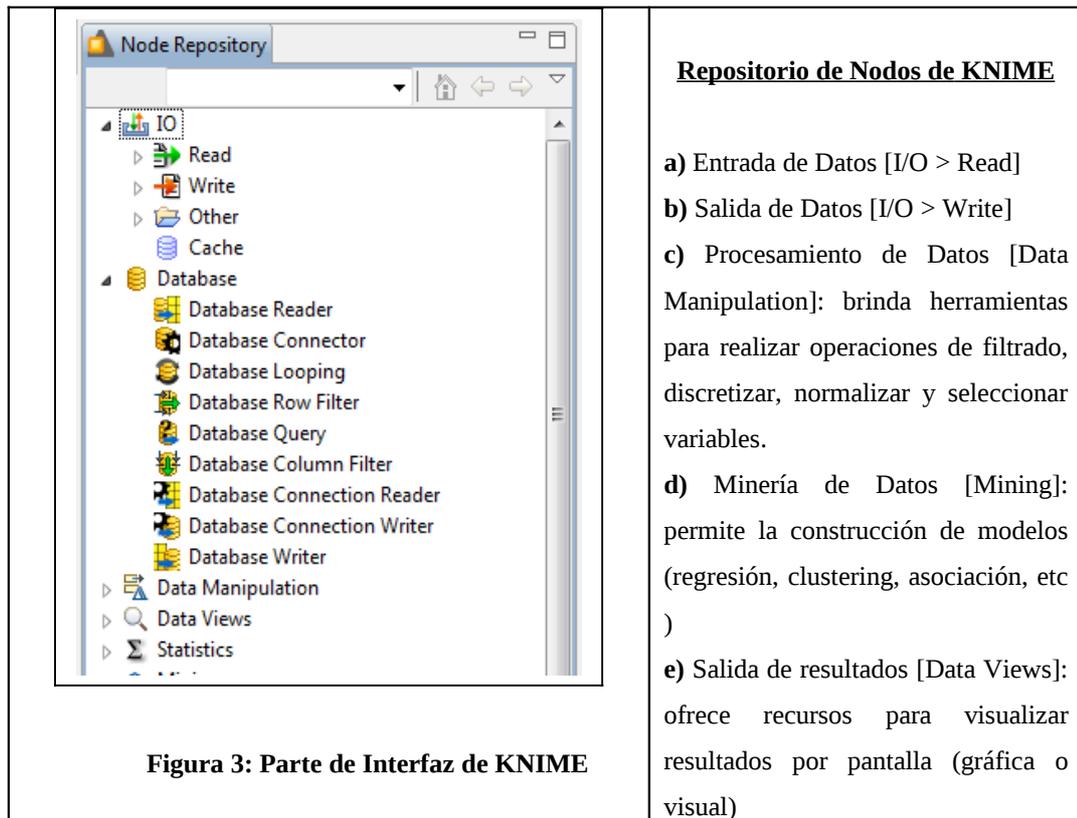
2. Relevamiento y Selección de herramientas de aprendizaje automático: Existen un abanico importante de alternativas en lo que respecta a herramientas libres para el aprendizaje automático entre alguna de ellas podemos citar: Weka, Orange, Knime entre otras. Uno de los aspectos considerados para la selección de la herramienta es: interfaz de usuario amigable, alta portabilidad respecto a la fuente de datos, diversidad de algoritmos implementados respecto a las técnicas de clasificación y sobre todo la capacidad que brinda respecto a la herramientas gráficas lo que facilita la interpretación de los modelos generados.

La herramienta seleccionada para la implementación de este proyecto de minería de datos es KNIME, la fundamentación de esta elección es que esta herramienta cumple en un alto grado con las características mencionadas en el punto anterior.

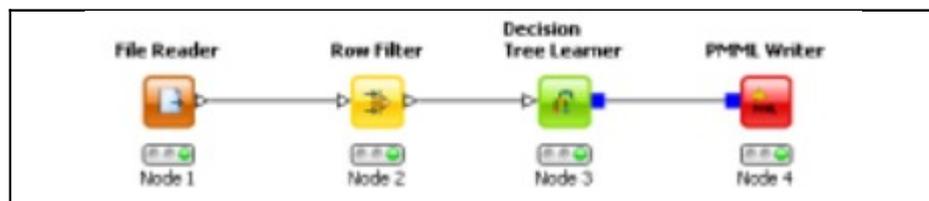
2.1 Descripción de la herramienta KNIME

[1] Knime es un entorno totalmente gratuito para el desarrollo y ejecución de técnicas de minería de datos. Fue desarrollada originalmente en el departamento de bioinformática y minería de datos de la Universidad de Constanza (Alemania) bajo la supervisión del profesor Michael Berthold. El diseño de esta herramienta se basa en el diseño de un flujo de ejecución que van reflejando las diferentes etapas de un proyecto de minería de datos.

Knime proporciona distintos nodos agrupados en fichas, como se muestra a continuación:



Para la creación de un flujo de ejecución, generalmente las salidas de unos nodos se utilizan como entrada de otros, como se puede visualizar a continuación.



KNIME incluye otras herramientas además de la explicada anteriormente, que se pueden identificar fácilmente por medio del siguiente gráfico.

La ventana de Editor: facilita la visualización de los diagramas de ejecución construidos.

Repositorio de Nodos: se encuentran los nodos que pueden incorporarse a un flujo.

Consola de eventos: visualiza los mensajes de warning y error. Estos mensajes se almacenan en un archivo, llamado knime.log (stack trace), ubicado en la carpeta .metadata del proyecto.

Descripción de Nodos: provee información del nodo seleccionado. La misma se encuentra vacía si no se ha seleccionado ningún nodo. En el caso de seleccionar una categoría que incluye más de un nodo, muestra los nodos que forman parte de esa categoría.

Explorador de Proyectos: muestra todos los proyectos existentes en el workspace, esta herramienta permite la creación de proyectos, importar proyectos existentes o exportar.

3. Elección del algoritmo supervisado de minería de datos a implementar: Como se mencionó en las secciones anteriores uno de los ejes principales son el mantenimiento y la gestión de inventarios. Por medio de la rama de minería de datos es posible la implementación de técnicas que según el objetivo del análisis de datos, los algoritmos se clasifican en: Algoritmos supervisados y no supervisados (Weiss y Indurkha, 1998)

Los algoritmos supervisados: también llamados predictivos, permiten la predicción de un dato o un conjunto de ellos desconocido a priori, a partir de otros conocidos.

Los algoritmos no supervisados (o del descubrimiento del conocimiento): el propósito es el descubrimiento de patrones y tendencias a partir de un conjunto de datos.

De esta clasificación se infiere que el algoritmo seleccionado para resolver la problemática inicial de este trabajo, se encuentra dentro de la rama de los algoritmos no supervisados, y se denominan reglas de asociación.

3.1 Reglas de Asociación

Esta técnica trata sobre la exploración de los datos con el propósito de identificar relaciones entre los datos, dentro de una fuente o base de datos.

[2] El descubrimiento de asociaciones encuentra las “reglas” en los elementos que aparecen juntos en un determinado evento. Este tipo de asociaciones se denotan como $A \Rightarrow B$, donde A es el antecedente y B es el consecuente.

[3] Esta técnica emergió en la década de los 90 con una aplicación práctica, el análisis de información de ventas para el mercadeo. Mediante ella se descubrían las relaciones entre los datos recopilados a gran escala por los sistemas de terminales de punto de venta de supermercados. Los datos consistían en colecciones de transacciones, también conocidas como bases de datos transaccionales.

El interés de una regla de asociación está dado por su soporte y su confianza, entendiéndose por “soporte” la frecuencia de aparición en la colección de la combinación de productos involucrados en la regla. Por confianza de una regla entendemos cuánto representa el soporte de la regla, del soporte del antecedente de la regla.

Se considera que una regla es interesante si su soporte y su confianza son mayores o iguales que determinados umbrales de mínimo soporte y mínima confianza especificados.

Este tipo de reglas fue denominado Reglas de Asociación Binarias. Varios han sido los algoritmos desarrollados para su implementación. Uno de los aspectos que son de importante consideración es la naturaleza de los datos, es decir el rango de valores que puede asumir. Esto es, siempre es importante analizar este aspecto ya que cada algoritmo para ser implementado es necesario que los datos se encuentren presentados de una determinada forma. Para aplicar reglas

de asociaciones, es necesario efectuar el proceso de discretización comúnmente consiste en particionar en intervalos el dominio de los atributos.

4. Migración de los datos a un formato adecuado: en esta etapa consiste detectar cuales son los formatos posibles que soporta la herramienta Knime. En función de lo estudiado la misma ofrece una amplia gama de alternativas respecto al formato de la fuente de datos para su posterior procesamiento. Los formatos compatibles con esta herramienta son: *.xml, *.tst, *.all, *.cvs, *.xls y *.arff. Este último formato es el portable por otra herramienta de filosofía libre para el aprendizaje automático como Weka.

Para la implementación de este proyecto se ha seleccionado el último formato citado (*.arff), por su alta portabilidad y capacidad de integración con otras herramientas de aprendizaje automático. La fuente de dato que refleja la información relacionada con la gestión de inventarios está en una planilla de cálculo Excel y la vinculada con los mantenimientos efectuados en la PC se encuentra en formato papel en la planilla señalada en la sección 2.

El proceso de migración de estos datos implica una serie de actividades:

- O **Selección de atributos significativos:** Esta etapa es crucial para este proyecto, es fundamental la elección adecuada de cuáles van a ser las variables consideradas para la generación de patrones de comportamiento. Este trabajo pretende focalizarse en algunas variables que se encuentran plasmadas en la planilla de incidentes, el objetivo es la generación de reglas de asociación a través de la rama de la minería de datos, con el objetivo de descubrir relaciones y aspectos que influyen en el reporte de incidentes de una PC.

De esta manera se pretende a partir de las reglas de asociación obtenidas generar un modelo de conocimiento que permita brindar información relacionada con cuales son los fallos más habituales, sus causas más frecuentes o develar aspectos como en que turno de trabajo se reportan incidentes o bajo qué circunstancias se reporta el incidente al Área Técnica.

Este tipo de reportes generados a partir de las reglas de asociación, permitirá al personal del área técnica del laboratorio, poder establecer una planificación adecuada respecto a las órdenes de compra con el objetivo de una gestión eficiente de los inventarios.

A continuación se detallan los atributos que se han considerado como relevante para este estudio que provienen de la planilla de incidentes:

- Número de incidente
- Fecha
- Apellido y Nombre de persona responsable que reportó la falla
- Número de estación
- Número de aula
- Detalle del incidente: este atributo deberá ser tipificado.

- Fecha de la intervención (Por parte del Área de Mantenimiento)
- Apellido y Nombre del responsable del mantenimiento
- Problema resuelto (este atributo es del tipo boolean)

Hay otros atributos a considerar para la construcción del modelo:

- Turno que se generó el incidente: en momento que se generó el reporte del incidente. Puede ser turno Mañana, Tarde o Noche.
 - Contexto de reporte del incidente: este atributo representa en que ocasión se reporta el incidente. En este caso la estandarización para este atributo sería: Práctica Libre o En horario de clase.
 - Origen del incidente: los valores que puede asumir son: Componente: en caso de tratarse del cambio o reparación de una pieza/Software: en el caso que se ha detectado un problema funcional a nivel de software.
 - Denominación Componente: el mismo representa la denominación del componente que originó el reporte de incidente, en el caso de tratarse de un incidente cuyo origen sea Componente.
- O **Carga de datos:** Al no disponer de una herramienta informática que automatice la gestión de incidentes, será necesario la carga de los datos pertenecientes a estas planillas. El laboratorio de sistema dispone de información histórica relacionada con los incidentes es desde el año 2009 y respecto a la gestión de inventarios es del año 2002 hasta la actualidad. Se ha decidido considerar para este estudio el periodo desde el 2009 al 2011.
- O **Transformación de los datos a un formato portable por la herramienta Knime:** Cumplimentada la etapa anterior se tiene la disponibilidad de la información histórica relacionada con la gestión de inventarios e incidentes integradas en un mismo formato. La propuesta de trabajo es transformar las fuentes de datos a un formato arff para que pueda ser procesado por la herramienta Knime.
- La estructura de un archivo arff es muy sencilla como se muestra en la siguiente figura:

```

1 @relation weather
2
3 @attribute outlook {sunny, overcast, rainy}
4 @attribute temperature real
5 @attribute humidity real
6 @attribute windy {TRUE, FALSE}
7 @attribute play {yes, no}
8
9 @data
10 sunny, 85, 85, FALSE, no
11 sunny, 80, 90, TRUE, no
12 overcast, 83, 86, FALSE, yes
13 rainy, 70, 96, FALSE, yes
14 rainy, 68, 80, FALSE, yes
15 rainy, 65, 70, TRUE, no
16 overcast, 64, 65, TRUE, yes
17 sunny, 72, 95, FALSE, no
18 sunny, 69, 70, FALSE, yes
19 rainy, 75, 80, FALSE, yes
20 sunny, 75, 70, TRUE, yes
21 overcast, 72, 90, TRUE, yes
22 overcast, 81, 75, FALSE, yes
23 rainy, 71, 91, TRUE, no

```

Figura 5: Ejemplo de archivo con formato arff portable por la herramienta KNIME

[4] Este archivo con formato arff se divide en tres partes:

- **@relation,**
- **@attribute y**
- **@data**

A continuación conceptualizamos en forma resumida cada sección:

@relation <relation-name> (línea 1): Todo archivo arff debe comenzar con esta declaración en su primera línea (no podemos dejar líneas en blanco al principio). <relation-name> será una cadena de caracteres y si contiene espacios la pondremos entre comillas.

@attribute <attribute-name> <datatype> (líneas de la 3 a la 7)

En esta sección incluiremos una línea por cada atributo (o columna) que vayamos a incluir en nuestro conjunto de datos, indicando su nombre y el tipo de dato.

Con <attribute-name> indicamos el nombre del atributo, que debe comenzar por una letra y si contiene espacios tendrá que estar entre comillas.

Con <datatype> indicaremos el tipo de dato para este atributo (o columna) que puede ser:

- numeric (numérico)
- string (texto)
- date [<date-format>] (fecha). En <date-format> indicamos el formato de la fecha, que será del tipo "yyyy-MM-dd'T'HH:mm:ss".
- <nominal-specification>. Estos son tipos de datos definidos por nosotros mismos y que pueden tomar una serie de valores que se indica en la línea 3 de la figura 5.

@data (a partir de la línea 9)

En esta sección se incluye los datos propiamente dichos. Se separa cada columna por comas y todas filas deberán tener el mismo número de columnas, número que coincide con el de declaraciones @attribute que añadimos en la sección anterior.

Si no disponemos de algún dato, colocaremos un signo de interrogación (?) en su lugar.

Aclaración: El separador de decimales tiene que ser obligatoriamente el punto y las cadenas de tipo string tienen que estar entre comillas simples.

Esta tarea de transformación de los datos a este formato, puede tornarse lenta, más específicamente cuando se llega a la sección de data. Es por eso que se plantea como futura línea de investigación dentro de este trabajo, el estudio de herramientas que puedan migrar a este formato especial.

- 5. Implementación del algoritmo de minería de datos:** Como se ha mencionado en secciones anteriores la herramienta seleccionada para la implementar el proyecto de minería de datos es Knime. Esta herramienta dispone una metodología de trabajo basado en nodos, lo que facilita la elección e implementación de diferentes técnicas de minería de datos; en nuestro trabajo el uso de reglas de asociación como se puede apreciar en la siguiente figura:

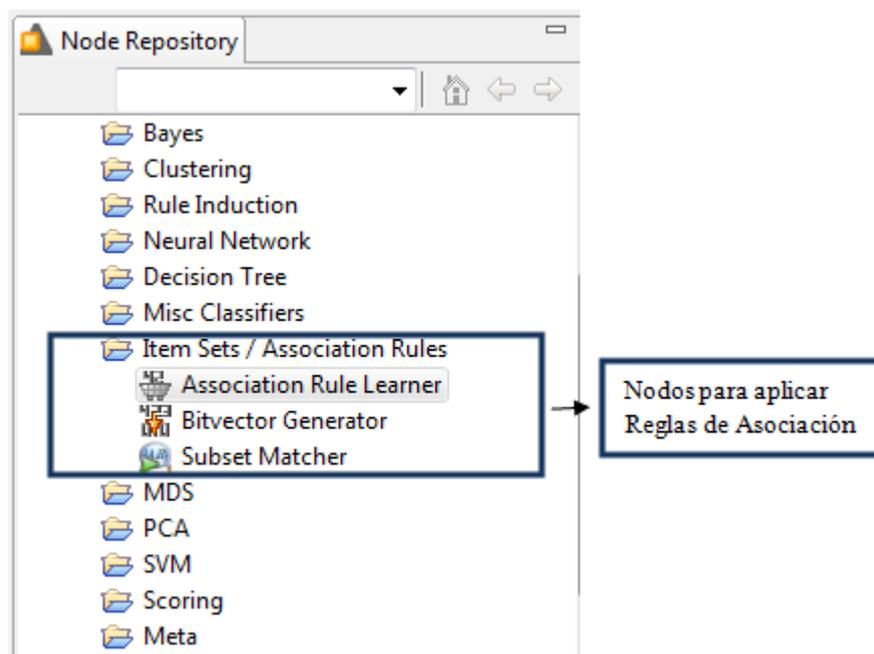


Figura 6: Nodos para la implementación de algoritmos de Asociación

[5] La planificación de tareas incluidas en esta etapa son:

- Conectar a la herramienta Knime la fuente de datos, que en nuestro caso está en formato arff. Para lograr este propósito KNIME dispone de un elemento apropiado para conectar una fuente de datos con este formato, denominada **Nodo ARFF Reader**. En el caso de que la conexión ha sido exitosa, la información almacenada en el archivo se visualizará en forma de tabla.

- Implementación del algoritmo de Minería de Datos: para cumplimentar este proceso es necesario la incorporación de un nodo **Association Rule Learner**.
- 6.** Interpretación y Valoración de resultados: El resultado del modelo de conocimiento generado, es decir conjunto de reglas de asociación se visualiza en forma de tabla. En la misma se ven reflejadas ciertos parámetros que representan métricas para la valoración de las reglas obtenidas. Estos parámetros son los que se detallan a continuación y que en su momento serán considerados para la elección de las reglas que resulten más confiables:
- **Soporte (support):** este parámetro refleja el número de casos en que se debe cumplir la regla resultante.
 - **Nivel de Confianza (confidence):** este nivel de confianza nos indica con que probabilidad se debe cumplir la regla.

Estos parámetros previamente son configurados antes de comenzar la ejecución del nodo **Association Rule Learner**.

CONCLUSIONES

Este proyecto forma parte de un conjunto de proyectos que se desarrollan en el contexto del Laboratorio de Sistemas de Información.

Las líneas principales de investigación están relacionadas con requerimientos reales por parte de esta unidad académica que son aspectos relacionados con: a) Gestión de Incidentes y Mantenimientos de equipos informáticos b) Gestión de inventarios c) Implementación de Herramientas informáticas de filosofía libre.

Uno de los objetivos que se pretende conseguir con la implementación de este proyecto es brindar al laboratorio de sistema de información un reporte informando cuales son los factores que tienen mayor incidencia en la generación de fallos de los equipos informáticos.

La posibilidad de acceder a esta información permitirá al personal del Laboratorio de Sistemas de Información tomar medidas de carácter preventivo con el objetivo de lograr una disminución en los reportes de incidentes de los equipos informáticos.

Es importante destacar si logramos develar cual son los tipos de componentes o piezas que más se reportan en las planillas incidentes, se podrá recomendar un plan de compras adecuado para satisfacer con mayor rapidez los requerimientos de determinados incidentes.

REFERENCIAS

- [1] KNIME IEDE. Manual Básico Knime; <http://www.slideshare.net/itapia/manual-bsico-knime>
- [2] Construyendo el modelo. Minería de Datos; 2005; <http://www.gravitar.biz/index.php/bi/construyendo-el-modelo-mineria-de-datos/>
- [3] Minería de Reglas de Asociación sobre datos Mezclados, Ansel Yoan Rodríguez González, José Francisco Martínez Trinidad, Jesús Ariel, Carrasco Ochoa, José Ruiz Shulcloper, Reporte Técnico No. CCC-09-001, 31 de Marzo de 2009. <http://ccc.inaoep.mx/portalfiles/file/CCC-09-001.pdf>
- [4] Minería de Datos con Weka; <http://www.locualo.net/programacion/mineria-datos-weka-ficheros-arff/00000019.aspx>
- [4] Primeros pasos con KNIME; http://laurel.datsi.fi.upm.es/_media/docencia/cursos/inap/ejemplodm.pdf
- [5] Reglas de asociación y análisis de afinidad en los artículos de picking; <http://www.innovacionlogistica.com/index.php/Documentos-Tecnicos/Reglas-de-asociacion-y-analisis-de-afinidad-en-los-articulos-de-picking.html>
- [6] Sitio oficial KIME; <http://www.knime.org>
- [7] “Sistemas de Inteligencia de Gestión. Práctica 1. Herramienta de Datos: KNIME”; Juan Carlos Cubero, Fernando Berzal; <http://elvex.ugr.es/decsai/intelligent/workbook/D1%20KNIME.pdf>
- [8] Introducción a la Minería de Datos, José Hernández Orallo; M. José María Quintana, Cesar Ferri Ramiro; Editorial Pearson; 2004; ISBN: 8420540919
- [9] Minería de Datos. Técnicas y Herramientas; Cesar Pérez López; Ediciones Parainfo; ISBN: 8497324927
- [10] Data Mining: Soluciones con Enterprise Miner; César Pérez; RA-MA; 2006; ISBN: 9788478976959