

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



**Universidad Tecnológica Nacional. F.R.C
Departamento Ingeniería en Sistemas de Información.
Laboratorio de Sistemas de Información**

Ing. Cynthia Corso.
Ing Fabián Gibellini

Resumen del Trabajo

- El objetivo de este trabajo es la presentación de una propuesta metodológica dentro de la rama de la **Minería de Datos**.
- La misma incluye el uso de una **herramienta de aprendizaje automático** de filosofía libre.
- El resultado final es la generación de un modelo de conocimiento que permita detectar cuales son las posibles causas que tienen mayor influencia en el reporte de incidentes

Que es la Minería de Datos?

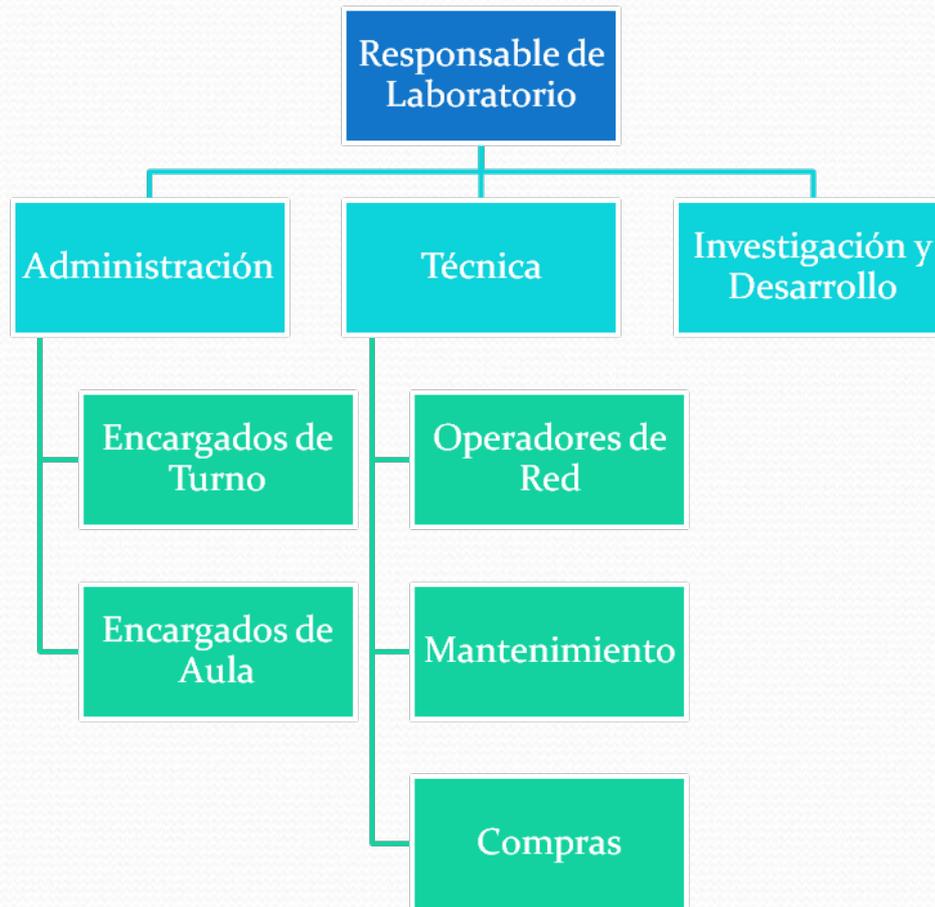


- La Minería de Datos (Data Mining) es la búsqueda de patrones interesantes y de regularidades importantes en grandes bases de datos.
- Se enfoca en cubrir la necesidad de descubrir el porque, para luego predecir y pronosticar las posibles acciones con un cierto factor de confianza.
- Son diversas las problemáticas que se pueden resolver con MD.
 - Técnicas **Descriptivas**

Contexto del Trabajo

- El contexto de este estudio es el **Laboratorio de Sistemas de Información** (Departamento de Ingeniería en Sistemas) U.T.N (F.R.C)
- El mismo brinda un conjunto de funciones y servicios:
 - ❑ Infraestructura adecuada para el dictado de clases, cursos de capacitación y extensión del Dpto. de Sistemas.
 - ❑ Espacio dirigido a los alumnos, denominado Prácticas Libres, que da la posibilidad de acceder a los usos de las aulas de laboratorios para la realización de actividades académicas.
 - ❑ Asesoramiento técnico tanto a los docentes como

Áreas Funcionales Involucradas



Metodología

Para llevar a cabo el proyecto de Minería de Datos, se han contemplado el cumplimiento de las siguientes actividades:

1. Detección de documentación o reportes.
2. Relevamiento y Selección de herramientas de aprendizaje automático.
3. Elección de algoritmo supervisado
4. Migración de datos a un formato adecuado
5. Implementación
6. Interpretación y Validación de Resultados.

1. Detección de Documentos o reportes.

- En esta fase se han relevado los documentos y reportes involucrados en la gestión de incidencias e inventarios.

Incidente		Nro.:	
Fecha:/...../.....	Hora::.....		
Reportado por	Area:		
Apellido y Nombre:	Firma:		
Estación/es:	Auto/s:		
Detalle: _____			

Incluir: cuándo sucedió (hora, cátedra, etc.), cómo sucedió, mensaje de error o del sistema, etc.			
Intervención 1	Fecha:/...../.....	Hora::.....	
Area:	Apellido y Nombre:		
Problema resuelto:	SI	NO	
Detalle de la acción: _____			

Intervención 2	Fecha:/...../.....	Hora::.....	
Area:	Apellido y Nombre:		
Problema resuelto:	SI	NO	
Detalle de la acción: _____			

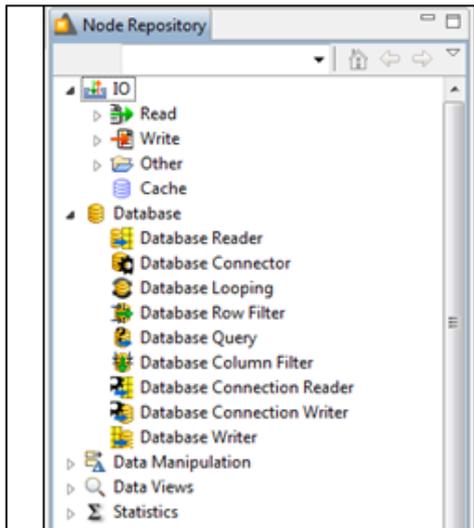
2. Relevamiento y selección de herramientas de aprendizaje automático.

- Se investigaron diversas herramientas libres para el aprendizaje automático, algunas de ellas: Weka, Orange y Knime entre otras.
- La herramienta seleccionada para este estudio fue **Knime**.
- Las **ventajas** de la herramienta seleccionada:
 - ❑ Interfaz de usuario amigable
 - ❑ Alta portabilidad respecto a la fuente de datos
 - ❑ Diversidad de algoritmos de clasificación.
 - ❑ Herramientas gráficas adecuadas.

2.1 Herramienta KNIME



- Herramienta de minería de datos libre, creada en el contexto de la Universidad de Constanza (Alemania). El diseño de esta herramienta se basa en el diseño de un flujo de ejecución.



Repositorio de Nodos de KNIME

- a) Entrada de Datos [I/O > Read]
- b) Salida de Datos [I/O > Write]
- c) Procesamiento de Datos [Data Manipulation]: brinda herramientas para realizar operaciones de filtrado, discretizar, normalizar y seleccionar variables.
- d) Minería de Datos [Mining]: permite la construcción de modelos (regresión, clustering, asociación, etc)
- e) Salida de resultados [Data Views]: ofrece recursos para visualizar resultados por pantalla (gráfica o visual)

Figura 3: Parte de Interfaz de KNIME

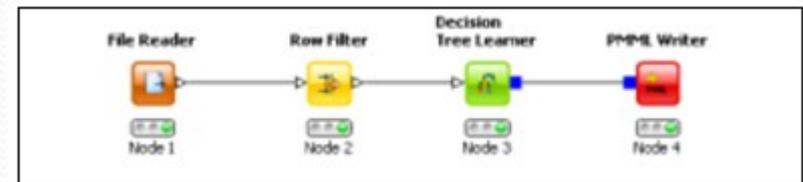


Figura 4: Ejemplo de Flujo de Ejecución en KNIME

3. Elección del algoritmo

- Por medio de la rama de minería de datos es posible la implementación de técnicas que según el objetivo del análisis de datos se clasifican en algoritmos **supervisados** y **no supervisados**. (Weiss y Indurkha,1998)
- De esta clasificación, este estudio ha seleccionado un algoritmo dentro de la rama de los no supervisados, denominados **reglas de asociación**.
- El descubrimiento de asociación encuentra las **reglas**.

Ejemplo:

41 JAIOO. Jornadas de Software Libre $A \Rightarrow B$

4. Migración de datos en formato adecuado.

- En esta etapa consiste en transformar los datos en un formato portable por la herramienta KNIME.
- El proceso de migración implica una serie de actividades:
 - **Selección de atributos significativos**
 - **Carga de datos.**
 - **Transformación de datos a formato arff.**

4.1 Selección de atributos significativos.

- En esta sección se detallan los atributos o variables considerados relevantes para este estudio.

- 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.

4.3 Transformación de datos a formato arff.

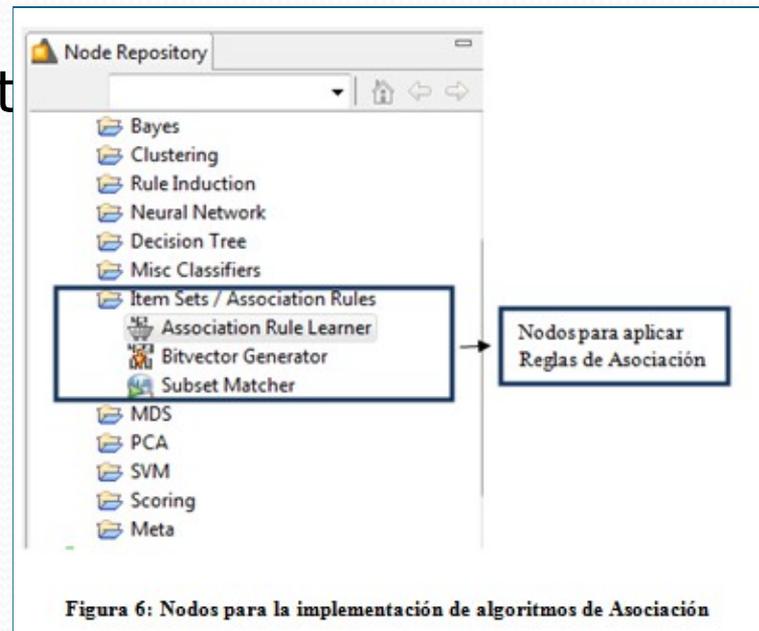
- La estructura de un archivo arff es la como se muestra en la siguiente figura:

```
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

5. Implementación de algoritmo de minería de datos.

- Las actividades planificadas para esta etapa son:
 - Conexión de la fuente de datos a la herramienta KNIME.
 - Implementación de **Association Rule Learner**.



Association Rule

Figura 6: Nodos para la implementación de algoritmos de Asociación

6. Interpretación y validación de resultados

- El resultado del modelo generado, es decir el conjunto de reglas, la herramienta lo visualiza por medio de una tabla.
- Las métricas ha considerar para la elección de un conjunto de reglas son:
 - **Soporte:** este parámetro refleja el número de casos que debe cumplir la regla resultante.
 - **Nivel de confianza:** esta métrica nos indica el nivel de probabilidad que se debe cumplir la regla.

Preguntas??



E-mail: cynthia@bbs.frc.utn.edu.ar