



II CAIM 2010
Segundo Congreso Argentino
de Ingeniería Mecánica
San Juan - Noviembre 2010

UNA METODOLOGIA DE ANALISIS DE FALLOS BASADA EN EL USO DE HERRAMIENTAS INFORMATICAS, II CAIM 2010

Abet Jorge E ^{*1}, Carrizo Blanca R.², Corso Cynthia L.³ y González Gustavo J.^{4,5}

^{*1} *Dpto. Ingeniería Mecánica, Dpto. Ingeniería Industrial – Universidad Tecnológica Nacional – Facultad Regional Córdoba - Maestro M. López s/n. Ciudad Universitaria. Córdoba Capital.*
e-mail: jabet@extension.frc.utn.edu.ar

² *Dpto. Ingeniería Mecánica, Dpto. Ingeniería Industrial, Dpto. Ingeniería en Sistemas de Información.*
e-mail: bcarrizo@tecnicatura.frc.utn.edu.ar

³ *Dpto. Ingeniería Industrial, Dpto. Ingeniería en Sistemas de Información.*
e-mail: cynthia@bbs.frc.utn.edu.ar

⁴ *Dpto. Ingeniería Mecánica, Universidad Tecnológica Nacional – Facultad Regional Córdoba*
⁵ *Centro de Investigaciones Aplicadas, IUA (Instituto Universitario Aeronáutico).*
e-mail: ggonzalez@scdt.frc.utn.edu.ar

RESUMEN

Esta es una propuesta que plantea una metodología de trabajo para estudiar y analizar los fallos que se producen en las maquinarias, entendiendo por fallo aquellos eventos indeseables que debemos tratar de evitar, prevenir o anticipar, a través del estudio de su probabilidad de ocurrencia mediante métodos probabilísticos automáticos.

Con esta finalidad, analizamos el impacto de aplicar herramientas informáticas en los estudios de fiabilidad en el mantenimiento, conceptualizando el mismo como aquella función cuyo objetivo es la prolongación y/o la recuperación de las funciones de determinado componente o máquina. [1]

La metodología consiste en seleccionar una muestra representativa de una población bajo estudio, en nuestro caso particular equipos industriales, cuyos parámetros sean significativos. Estos parámetros son procesados por un software de aplicación, el cual brindará información para realizar inferencias que intentarán predecir futuras tendencias y comportamientos que faciliten procesos de toma de decisiones proactivos, conducidos por un conocimiento acabado de la información en cuestión.

La muestra a analizar se obtuvo del estudio de un caso, cuya fuente de datos lógicos residía en una planilla de cálculo y que, previo análisis de sus datos, fue migrada a una Base de Datos estandarizada basada en una tecnología consistente y superior.

A esta Base de Datos, diseñada con una visión de Almacén de Datos, se le aplicaron herramientas basadas en Minería de Datos, con la finalidad de examinar esos registros históricos, formular hipótesis, verificarlas mediante pruebas, y si éstas resultasen satisfactorias, suministrar ese nuevo conocimiento al usuario.

Estas hipótesis, en general, se describen bajo la forma de reglas o algoritmos. [2]

La metodología propuesta consiste en analizar un caso práctico real, una empresa autopartista, aplicando una herramienta de Minería de Datos basada en la filosofía de Software Libre denominado Weka.

Palabras Claves: Análisis de Fallos – Metodología – Almacén de Datos – Minería de Datos – Weka.

1. INTRODUCCIÓN

En el perfil industrial de la provincia de Córdoba, si bien amplio, se destacan básicamente dos tipos de industrias: la metalmecánica y la del software; las cuales atraen y facilitan la llegada de otras empresas internacionales que, a su vez, crean empleos tanto en forma directa como indirecta.

Estas grandes empresas, a través de la creación y/o consolidación de proveedores, contribuyen al fortalecimiento y desarrollo de las Pymes. En consecuencia, confían en el potencial que ellas tienen, pues son conscientes de la importancia de actuar en un mercado globalizado y competitivo; en donde lo fundamental es lograr la generación de valor, pero siendo socialmente responsables y garantizado un desarrollo integral y sostenible para todos sus actores.

De esta forma, el desarrollo de proveedores constituye una de las mejores prácticas para las organizaciones que tienen la necesidad urgente de contar con proveeduría especializada, competitiva y alineada con los agresivos planes de crecimiento de su empresa cliente, logrando así una cadena de suministro bien integrada y preparada para la generación de nuevos proyectos.

Las Pymes cordobesas constituyen organizaciones capaces de adaptarse a los cambios tecnológicos y generar empleos, con lo que representan un importante factor de política redistributiva del ingreso en los estratos medios y bajos de la escala social.

Por ello, este proyecto se direcciona a atender requerimientos de mejora continua de las pequeñas industrias, en las cuales los equipos críticos del proceso productivo deben ser protegidos mediante políticas y procedimientos de mantenimiento no sólo preventivo sino predictivo; tendientes a optimizar la relación Cliente/Proveedor.

Es así que el objetivo, a través de esta propuesta, es consolidar la generación sostén y desarrollo de este sector prácticamente indefenso a nivel economía, ayudando a través del uso de herramientas informáticas al crecimiento y competitividad de las mismas.

2. ANTECEDENTES

Este trabajo está basado en un Proyecto de Investigación que está orientado a estudiar el impacto de la aplicación de las Tecnologías Informáticas en el estudio probabilístico de los fallos (o Fiabilidad).

Si se concibe a los fallos como eventos no deseables, y se logra evitarlos, prevenirlos o anticiparlos basándose en el estudio de las probabilidades de ocurrencia de los mismos, se aportaría al mantenimiento de una herramienta que permita tomar decisiones en tiempo y forma. En esta instancia, la Informática como una ciencia de aplicación interdisciplinaria, se transforma en una excelente herramienta para la toma de decisiones automáticas a través del uso de las Bases de Datos; entendiéndose las mismas como una colección de datos y/o documentos digitales, que pueden ser homogéneos o no.

Dentro de este escenario surgen dos herramientas sustentadas en la filosofía de Base de Datos, que tienen por objetivo primordial manejar eficientemente la información de cada área de la Empresa y tomar decisiones en tiempo y forma. Estas son: el Almacén de Datos o Data Warehouse y la Minería de Datos o Data Mining.

El Almacén de Datos es un proceso que utiliza datos provenientes de diversas fuentes, algunas internas y otras externas, y que permite atender consultas para toma de decisiones a través de la generación de Bases de Datos tangibles con una perspectiva histórica; mientras que la Minería de Datos es un conjunto de tecnologías o herramientas aptas para analizar la información almacenada en un Almacén de Datos, capaces de ayudar a descubrir relaciones insospechadas, tendencias, modelos de comportamiento o correlaciones entre los datos contenidos en el mismo. De este modo, puede formular hipótesis, verificarla mediante pruebas, y si éstas resultan satisfactorias, suministrar ese nuevo conocimiento al usuario. Esas hipótesis, en general, se describen bajo la forma de reglas, que permiten predecir futuras tendencias y comportamientos a nivel negocios y tomar decisiones proactivas conducidas por un conocimiento acabado de la información.

De acuerdo a lo descripto, este trabajo integra dos disciplinas, la Ingeniería en Sistemas y la Ingeniería de Fiabilidad, las cuales forma conjunta pueden aportar elementos de juicio “gestionales” (a través de reportes, informes, gráficos estadísticos, etc), destinados a predecir comportamientos de equipos, detectar los posibles fallos además de revelar las potenciales causas que condicionan la disponibilidad de los equipos y así poder planificar la actividad de mantenimiento. [1]

La Ingeniería de Fiabilidad se aboca al estudio de los fallos del “material” (entendido como parte -conjunto - máquina en sentido genérico). Esto se basa en métodos estocásticas o probabilísticos en Sistemas de Distribuciones Continuas, estudiando el comportamiento de los fallos y tratando de analizar y comprender a las mismos. De esta manera, se facilita la identificación de las mejoras que pueden aplicarse a nuevos diseños y ayuda en el mantenimiento de los equipos para incrementar su vida útil.

3. OBJETIVOS

- Estudiar y analizar herramientas informáticas basadas en Tecnologías Informáticas y Sistemas de Información de Bases de Datos, como lo son el Almacén de Datos y Minería de Datos, en el estudio de la longevidad y fallo de los equipos, a través de la predicción de tendencias y de comportamientos que permitan tomar decisiones proactivas basadas en repositorios de datos históricos y en la criticidad de los sistemas en funcionamiento.
- Determinar si la aplicación de ambas herramientas, Almacén de Datos y Minería de Datos, facilitan no solo análisis prospectivos automatizados (mantenimiento preventivo) de los fallos, sino eventos futuros, cuyo comportamiento puede inferirse del análisis de ciertos parámetros.
- Relevar, analizar, comparar y verificar si estas herramientas para la toma de decisiones, responden y brindan soluciones y potenciales usos en otras áreas de conocimiento.

4. METODOLOGÍA

La metodología analizada, está basada en la aplicación de un software denominado Weka (Waikato Environment for Knowledge Analysis - Entorno para Análisis del Conocimiento de la Universidad de Waikato), que es un reconocido software en el medio para aprendizaje automático y Minería de Datos escrito en Java y desarrollado en la Universidad de Waikato.

Es un software libre distribuido bajo licencia GNU-GPL, que contiene una colección de herramientas de visualización y algoritmos para análisis de datos y modelado predictivo, unidos a una interfaz gráfica de usuario para acceder fácilmente a sus funcionalidades.

Weka soporta varias tareas estándar de Minería de Datos, especialmente, preprocesamiento de datos, clustering, clasificación, regresión, visualización, y selección. Todas sus técnicas se fundamentan en la asunción de que los datos están disponibles en un fichero plano o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos).

La metodología propuesta consta de una serie de etapas tales como:

- a) Selección de un conjunto de datos.
- b) Análisis de las propiedades de los datos.
- c) Transformación del conjunto de datos de entrada.
- d) Selección y aplicación de la técnica de Minería de Datos.
- e) Extracción de conocimiento.
- f) Interpretación y evaluación de datos.

A continuación se analiza cada una de las etapas:

a) Selección del conjunto de datos.

Se basa en la fuente de datos, cuyo formato original de almacenamiento es una planilla de cálculo. En la misma se almacena información histórica relacionada con los fallos de esta empresa.

Es muy importante la selección de las variables objetivos, es decir aquellas que son significativas para el estudio y que permitirán inferir sobre cuestiones relacionados con los patrones de ocurrencia de los fallos.

Otro aspecto importante es la selección de una muestra, para facilitar el estudio y proceso de Minería de Datos.

b) Análisis de las propiedades de los datos.

Esta instancia se focaliza en el estudio de los datos seleccionados para el análisis.

A continuación, se especifican algunas técnicas que facilitarán la transformación de los datos. Generalmente esto se realiza mediante técnicas de reducción o aumento de la dimensión y escalado simple y multidimensional, entre otras.

Una de las técnicas usadas es la de Limpieza o Valores atípicos (outliers): Un valor outlier o atípico es una puntuación extrema dentro de una variable. Este tipo de valores afecta de manera significativa a los análisis donde intervenga dicha variable, sobretodo en este caso donde se selecciona una muestra pequeña. Es decir que analizando los valores asumidos por los atributos de cada fallo, en algunas ocasiones se ha llegado a observar valores muy por encima o por debajo de la media. Se podrían definir como observaciones aisladas cuyo comportamiento se diferencia claramente del resto.

Para poder detectar en nuestro estudio este tipo de valores atípicos o outliers, usamos el gráficos de caja y bigotes.

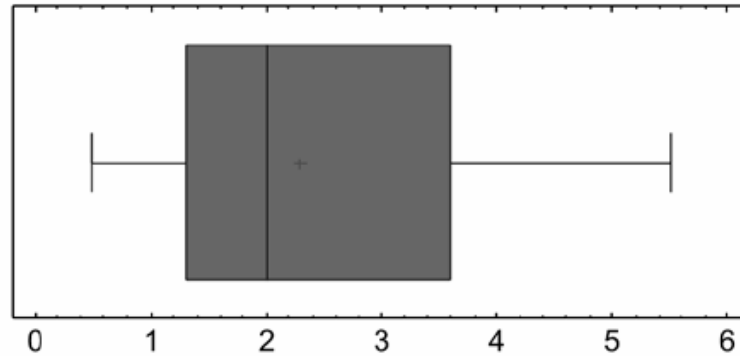


Figura 1: Gráfico de Caja y Bigotes.

Otro aspecto importante en el estudio, es el tratamiento de los datos ausentes. Esta ausencia puede ser consecuencia de un registro defectuoso de la información, o sea la ausencia natural de la información buscada o la una falta de respuesta (total o parcial).

Una de las pruebas a efectuar frente a esta situación es comprobar si se distribuyen aleatoriamente en todo el conjunto de datos. Para valorar los datos ausentes de una única variable Y , se forman dos grupos de valores, los que tienen datos ausentes y los que no. Luego, para cada variable X es distinta de Y , se realiza un test para determinar si existen diferencias significativas entre los grupos de valores determinados por la variable Y (ausentes y no ausentes) sobre X .

Si se repite el proceso anterior y se encuentra que todas las diferencias no son significativas, se puede concluir que todos los datos ausentes obedecen a un proceso completamente aleatorio y se puede hacer un análisis estadístico fiable.

Una vez que se ha contrastado la existencia de aleatoriedad en los datos ausentes, se estaría en condiciones de tomar algún tipo de decisión para dichos datos antes de comenzar con cualquier análisis estadístico.

Existen diferentes métodos para el tratamiento de los datos ausentes, en este estudio se aplicará el método de supresión de casos de lista: el objetivo es incluir solo las observaciones que tienen sus datos completos. Esta alternativa, es el método por defecto que incorporan la gran mayoría de los programas que manejan información estadística.

c) Transformación del conjunto de datos de entrada.

El objetivo de esta etapa es preparar los datos, para que estén en el formato adecuado y puedan ser procesado por el algoritmo de Minería de Datos seleccionado. Cabe aclarar que la selección de la técnica de Minería de Datos, esta fuertemente ligada a los aspectos sobre los cuales se quiere inferir.

A su vez cada técnica, dispone de una serie de algoritmos que permiten la generación de un modelo de conocimiento, que posteriormente será interpretado.

En este trabajo se aplicarán técnicas de agrupación, el objetivo es la identificación de grupos de fallos que han tenido comportamiento similar. Por ejemplo: Detectar fallos que han sido reportados en periodos de tiempo cercano, o que por han sido provocado la gran mayoría de las veces por mala operación, o que existe una gran ocurrencia de fallos que provienen de un determinado sector de la empresa.

El algoritmo que se usará es uno de las más simples y clásico dentro de este tipo de técnicas, que es el denominado K-Means.

El procedimiento a aplicar teniendo en cuenta el algoritmo seleccionado será:

- Se considerarán todos los datos de la muestra obtenida de fallos y se seleccionará de manera arbitraria los “n” fallos que se convertirán en los centros de los “n” grupos.
- De los fallos restantes, a cada una de ellos se las asignará a un determinado grupo. La pertenencia estará fundamentada por un cálculo. El mismo resulta de la similitud de la ocurrencia de fallo y cada uno de los centros seleccionados en el punto anterior. Será finalmente al grupo de mayor similitud.
- El centroide de cada grupo será recalculado.
- Las ocurrencia de fallos serán nuevamente reasignadas, teniendo en cuenta el establecimiento de los nuevos centroides.
- Los pasos b),c) y d) se repetirán, hasta que los nuevos centroides de cada grupo no hayan tenido una variación significativa.

d) Selección y aplicación de la técnica de Minería de Datos.

En esta etapa se implementa el algoritmo seleccionado y se construye el modelo. La herramienta que se prevé para la implementación de este algoritmo será Weka. Aunque disponga de una interfaz bastante sencilla, dispone de diversas alternativas para implementar la gran mayoría de las técnicas de Minería de Datos.

e) Extracción de conocimiento.

Mediante la aplicación de esta técnica se obtiene un modelo, que a posterior será interpretado con el objetivo de generar conclusiones relacionadas con el patron de comportamiento de fallos.

f) Interpretación y evaluación de datos.

Un modo adecuado de definir las salidas totales de un proyecto de Minería de Datos es usar la ecuación:

Resultados = Modelos + Conclusiones

Esta ecuación establece el objetivo final de cualquier proyecto de Minería de Datos. No es solo significativo el modelo obtenido a partir de la aplicación de un determinado algoritmo, sino también las inferencias que podemos obtener a partir de una determinada situación.

Una vez obtenido los perfiles de los fallos, es necesario poder evaluar la calidad o nivel de confianza de los mismos. Las métricas Recall y Precisión utilizadas en el área de recuperación de información, han sido utilizadas para tal fin.

El recall es una medida que corresponde al porcentaje de ítems que son correctamente recomendados, mientras que la precisión mide la calidad promedio de una recomendación individual. La validación de los resultados para cualquier técnica de Minería de Datos, es fundamental, ya que determinan el nivel de confianza de las inferencias resultantes de este estudio.

Si el modelo no superará el proceso de evaluación, el proceso deberá repetirse hasta el principio, o reiniciar en alguna de las etapas anteriormente mencionadas.

4.1. Caso Práctico

Una empresa autopartista de nuestra ciudad, que para la fabricación de autopartes tiene diversas etapas en el proceso de producción.

La problemática expresada es la detección que los reportes de mantenimiento correctivo presentan en intervalos de tiempo frecuentes.

Esta empresa no ha podido detectar con exactitud cuáles son los factores que inciden en esta situación problema.

A continuación se detallan los atributos que caracteriza a cada fallo, que es la información que nos ha proporcionado la empresa, que trabaja en turnos rotativos (mañana, tarde y noche):

- OT N°: representa el número de Orden de Trabajo.
- Error del Usuario: representa la causa del fallo, que puede deberse a errores en la operación.
- Evento: representa el tipo de fallo (por ej.: saturación de línea, corte gral de energía)
- Tipo de Pérdida: representa el tiempo perdido por el fallo ocurrido (expresado en minutos)
- Equipo: representa la identificación de la máquina donde se produjo el fallo.
- Estado de la OT: representa la etapa se encuentra la orden de trabajo.
- Fecha: representa la fecha en que se produjo el fallo.
- Sector: representa el sector del proceso productivo, donde se genero el fallo.
- MantenimientoProgramado: representa si en el tiempo que se genero el fallo, existía un mantenimiento programado.
- Paro Producción: representa el tiempo de paro en el proceso productivo (expresado en minutos).
- Fallo Real: representa un valor lógico (verdadero o falso), que se carga cuando el técnico diagnostica si es un fallo real o no.
- Horario Fallo: representa en que turno se presento el fallo.

Este caso de estudio necesita extraer la tipología de los fallos (mantenimientos correctivos), para ello se debe seleccionar cual será el atributo sobre el cual se van a generar los grupos. En este caso el atributo que podemos seleccionar es fallo real (puede ser cualquier otro según sea el aspecto a estudiar).

El objetivo es agrupar cuales son los mantenimientos efectivamente han sido fallos reales y cuáles no. Es decir suele suceder que un fallo es reportado al area técnica de la empresa, pero en algunos casos ese "fallo" en realidad era causa de una mala operación del operario.

En este caso, seleccionaremos uno de los algoritmos clásicos de agrupamiento que es **K-Means**. Para su aplicación hay que distinguir cual será el atributo que será considerado como centroide (valor medio), para la generación de grupos de fallos. En el caso que compete al ejemplo será falloReal.

Es importante aclarar que para poder llevar a cabo la implementación del algoritmo K-Means, todos los atributos que se consideran tienen que ser nominales. Los seleccionados y que son pertinentes para la problemática bajo estudio son: error de usuario, horario del fallo (representa el turno de trabajo que se presento el fallo), fallo real y mantenimiento programado.

En este proceso hay algunas variables cuyos tipos han sido transformadas con la finalidad de que puedan ser procesadas por el algoritmo de agrupamiento.

Esta operación se ha automatizado por medio de Weka. En la sección de preprocesado de esta herramienta, dispone de una serie de filtros y transformaciones a nivel de atributo útiles para este caso.

A continuación se visualiza la interfaz que nos permite visualizar la cantidad de grupos resultantes.

Para el estudio inicial se ha considerado una muestra pequeña de 50 fallos.

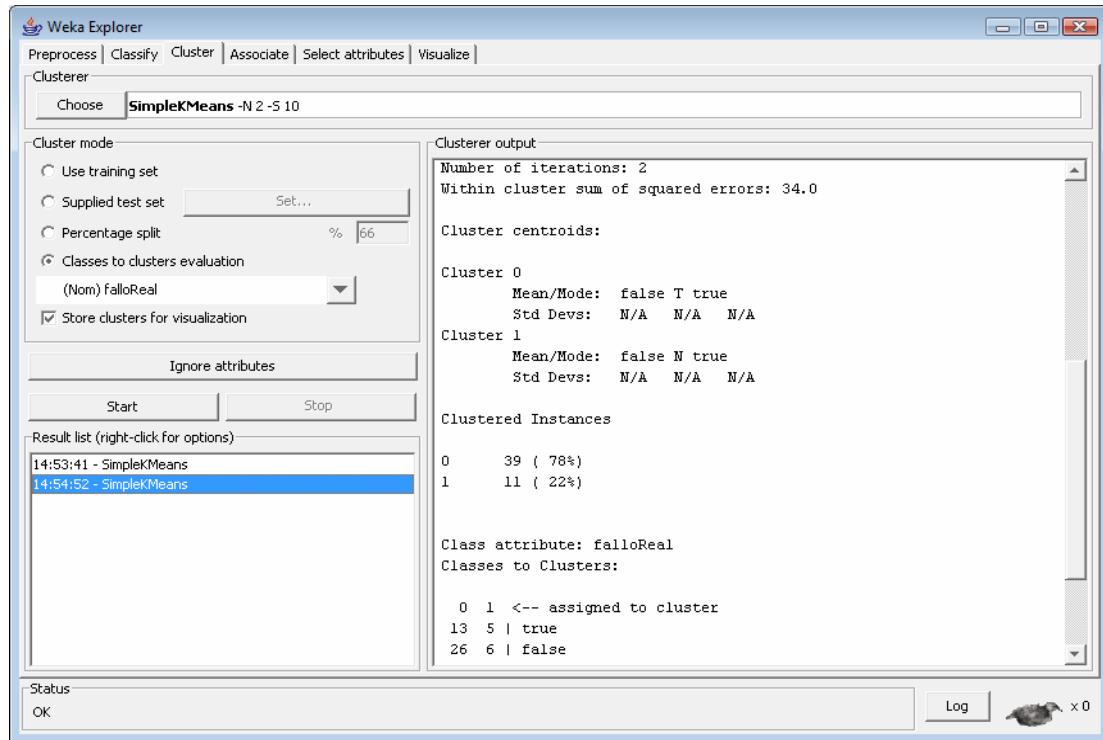


Figura 2. Interfaz de Weka: Solapa Agrupamiento de Clusters.

Para la aplicación del algoritmo, se necesita especificar sobre que atributo será el centroide, punto medio, a partir del cual se construirán los grupos. Además, es posible configurar la cantidad de instancias a considerar para generar los grupos e ignorar los atributos que no son significativos para análisis en cuestión. Weka ofrece una visualización gráfica, relacionada con los grupos obtenidos.

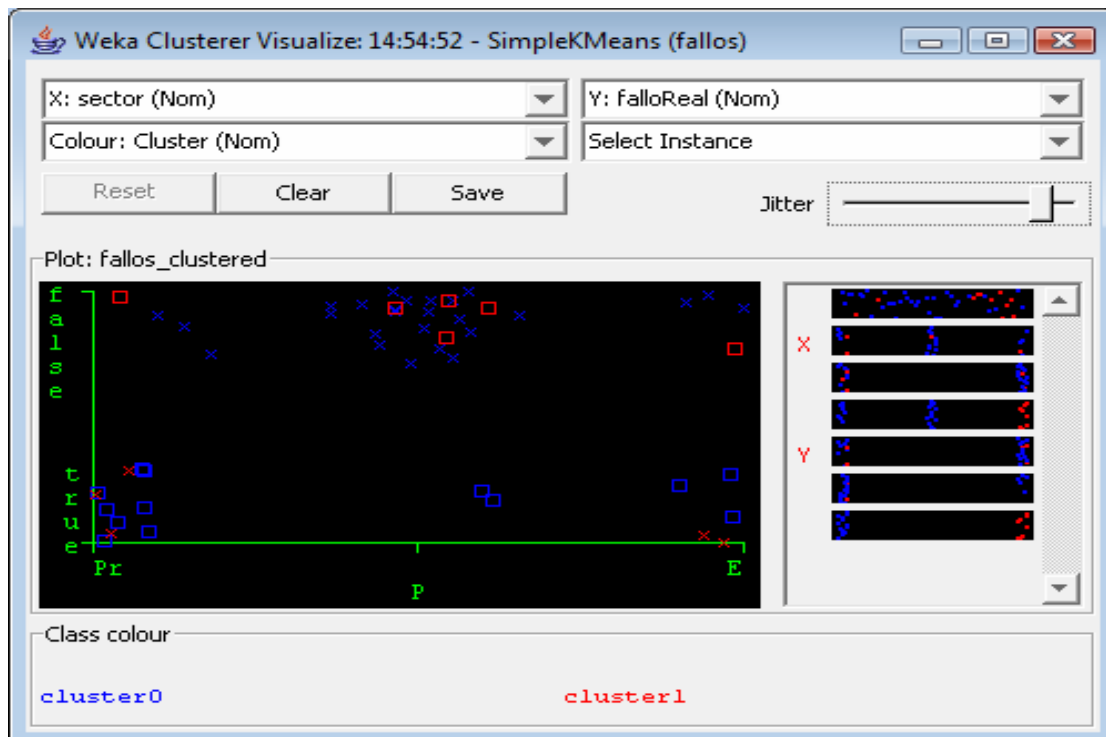


Figura 3. Interfaz de Weka: Visualización de Clusters.

En la Figura 3 podemos visualizar la formación de Clusters, donde el eje X esta representa los sectores donde se produce el fallo y el eje de Y representa el fallo real (true).

4.1.1. Conclusiones:

En la gráfica presentada en la Figura 3 se puede visualizar por sector, Preparación (Pr) - Producción (P) – E (Embalaje), la cantidad de ocurrencias donde el valor falloReal es True.

De la interpretación del gráfico, podemos inferir que en el Sector Preparación (Pr) la frecuencia de fallos originados son reales y que no son producidos, por ejemplo, por fallas del sistema o por razones de mala operación.

En cambio, en el Sector Producción (P) se puede inferir que, en la cantidad de fallos existentes el valor de falloReal es falso; razón por la cual debería hacerse un estudio más exhaustivo determinando cuales son las posibles causas que generan la presencia de esos fallos.

Si el fallo es causado por un error de operación, es factible que ese Sector en particular de la empresa necesite apoyo de capacitación en operación de las maquinarias del sector.

Este estudio prevee hacer pruebas con muestras más grandes, es decir incrementar los elementos bajo análisis, con el objetivo de contrastar los resultados obtenidos en esta problemática; como así también extenderlo a otros rubros como el alimenticio y textil.

5. RESULTADOS

En esta instancia, Weka se transforma en una excelente herramienta que optimiza la relación costo/beneficio, ya que también proporciona acceso a bases de datos vía SQL gracias a la conexión JDBC (Java Database Connectivity) y puede procesar el resultado devuelto por una consulta hecha a la base de datos; además de las siguientes ventajas:

- Disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede funcionar en casi cualquier plataforma.
- Contiene una extensa colección de técnicas para preprocesamiento de datos y modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario.

Otra vía de avance de este trabajo, está orientada a implementar estrategias didácticas en el seno de cátedras, tendientes a educar y concientizar tanto al ingresante (por ej. Informática I) como al futuro graduado (Mantenimiento) en la importancia de aplicar este tipo de metodologías tendientes a transferir resultados al mediano plazo a pymes, en las cuales los paros de producción por mantenimiento o avería de las máquinas, suponen un coste inadmisibles en términos de productividad.

6. REFERENCIAS BIBLIOGRÁFICAS

[1] Creus Solé Antonio, “Mantenimiento”, 1991).

[2] Diego García Morate, “Weka Tutorial”, 2006

<http://www.metaemotion.com/diego.garcia.morate/download/weka.pdf>

[3] “Página oficial de Weka”, <http://www.cs.waikato.ac.nz/~ml/weka/index.html>

[4] Jordi Porta Zamorano, “Técnicas cuantitativas para la extracción de términos en un corpus”; Escuela Politécnica Superior – Universidad Autónoma de Madrid, Julio de 2006.

[5] Dapozo, Gladys; Porcel, Eduardo; López, María V.; Bogado, Verónica; Bargiela, Roberto,

“Aplicación de Minería de Datos con una herramienta de software libre en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE”.