

PROYETO DE GRADO

Presentado ante la ilustre UNIVERSIDAD DE LOS ANDES como requisito final para obtener el Título de INGENIERO DE SISTEMAS

Aplicación de la Minería de Datos para describir las ocurrencias de fallas en los equipos y sistemas de la plataforma tecnológica de PDVSA AIT Servicios Comunes Centro

Por

Br: Manuel de Jesús Gómez Torrejano

Tutor: Prof. Gerardo Colmenarez

Cotutora: Prof. Liliana Capacho

Asesor Industrial: Ing. José Casanova

Marzo de 2009



©2009 Universidad de Los Andes Mérida, Venezuela

Aplicación de la Minería de Datos para describir las ocurrencias de fallas en los equipos y sistemas de la plataforma tecnológica de PDVSA AIT Servicios Comunes Centro

Br. Manuel de Jesús Gómez Torrejano

Propuesta de Proyecto de Grado--- Investigación de Operaciones

Resumen: Este proyecto tiene como propósito encontrar un modelo para gestión del mantenimiento que se realiza a partir de los atributos de fallas presentados en los equipos y sistemas que se encuentra bajo el manejo y custodia de la gerencia de Automatización, Informática y Telecomunicaciones (AIT) de Petróleos de Venezuela S.A., en el área denominada Servicios Comunes Centro (SCC).

Para llevar a cabo este estudio, la gerencia de AIT suministró información de los procesos que allí se realizan, con el fin de extraer detalles que permitan entender el funcionamiento sistemático en esta área. Así como también proporcionaron el historial de las fallas presentadas en los equipos y sistemas durante el periodo 2006 hasta primer semestre de 2008. Se aplicaron técnicas de minería de datos (*data mining*) que permitieron conocer aspectos importantes, tales como definir una variable de respuesta a partir de las variable observadas, y establecer un modelo de clasificación del mantenimiento correctivo realizado a los diferentes activos que operan en está plataforma. Este modelo combinado para gestión del mantenimiento mostró ser bastante eficiente en las fases de definición de conglomerados y de clasificación del mantenimiento.

Palabras Claves: Minería de datos, ingeniería de confiabilidad, incidencias de fallas, análisis de criticidad, aprendizaje automático

Dedicatoria

A mis padres, a mis cuatro hermanos, a mi hija y a mi novia, las personas más importantes en mi vida. Sin su apoyo no lo hubiera logrado, gracias.

Índice

Dedicatoria.....	ii
Índice.....	iii
Índice de Figuras.....	v
Índice de Tablas.....	vi
Agradecimiento.....	viii
Capítulo 1.....	1
Introducción.....	1
1.1 Antecedentes.....	3
1.2 Planteamiento del problema.....	4
1.3 Justificación.....	5
1.4. Objetivos.....	6
1.4.1. Objetivo general.....	6
1.4.2. Objetivos específicos.....	6
1.5. Metodología.....	7
1.6 Organización del documento.....	8
1.7 Descripción de la Empresa.....	9
1.7.1 PDVSA-AIT.....	10
Capítulo 2.....	14
Marco teórico.....	14
2.1 Conceptos básicos.....	14
2.2 Ingeniería de confiabilidad.....	17
2.2.1 Definición.....	17
2.2.2 Clasificación de las fallas.....	18
2.2.3 Estimación de la confiabilidad.....	20
2.3 Indicadores de confiabilidad basados en el historial de fallas.....	21
2.4 Minería de datos.....	22
2.4.1 El proceso de descubrimiento de conocimiento en bases de datos (<i>Knowledge Data Discovery</i> o KDD).....	23
2.4.2 Análisis de correspondencia múltiple.....	26
2.4.3 Tareas realizadas por Minería de datos.....	27
2.4.4 Máquina de Vectores de Soporte.....	28

Capítulo 3	33
Aplicación de minería de datos para caracterizar las fallas.....	33
3.1 Descripción de la naturaleza de los datos	33
3.2 Preprocesamiento de los datos.....	39
3.2.1 Reorganización de las variables.....	39
3.2.2 Categorización y creación de las nuevas variables	40
3.3 Análisis exploratorio de datos aplicando estadísticas fundamentales.....	42
3.4 Análisis de correspondencia múltiple.....	43
3.5 Aplicación de máquinas de vectores de soporte	45
3.5.1 Selección de los conjuntos de entrenamiento y validación.....	45
3.5.2 Ajuste de los parámetros de la máquina de soporte vectorial	46
3.5.3 Validación de los resultados.....	47
Capítulo 4	49
Análisis de los resultados e interpretación.....	49
4.5 Aplicación de máquinas de soporte vectorial.....	63
Capítulo 5	68
Conclusiones y Recomendaciones.....	68
5.1 Conclusiones.....	68
5.2 Recomendaciones.....	71
Bibliografía	73
Anexo A.....	74
Tablas de Contingencia para el grupo Respaldo y Almacenamiento	74
Anexo B.....	80
Análisis de Correspondencia Múltiple para los grupos faltantes.....	80

Índice de Figuras

Figura 1.1. Representación de la Coordinación de Confiabilidad.....	13
Figura 2.1. Etapis de la Minería de datos en un KDD.....	25
Figura 2.2 Clasificación usando SVM.....	31
Figura 3.1 Descripción de las actividades realizadas para generar los indicadores...34	
Figura 3.2. Muestra de observaciones correspondiente a los incidentes de fallas.....	35
Figura 3.3. Grupos operativos según el número de fallas.....	39
Figura 3.5 Estructura seguida para la codificación de las variables.....	40
Figura 4.1. Frecuencia de fallas.....	51
Figura 4.2. Frecuencia de Causas.....	52
Figura 4.3. Frecuencia de Acciones.....	53
Figura 4.4. Frecuencia de los tipos de componentes.....	54
Figura 4.5. Frecuencia de los tipos de impactos.....	55
Figura 4.6. Representación gráfica de la matriz cruzada de Causas vs. Fallas.....	56
Figura 4.7 Análisis de Correspondencia Múltiple Grupo Respaldo y Almacenamiento.....	60
Figura A1. Representación gráfica de la matriz cruzada de Acción vs. Fallas.....	76
Figura A2. Representación gráfica de la matriz cruzada de Tipo de componente vs. Fallas	77
Figura A3. Representación gráfica de la matriz cruzada de Impacto vs. Fallas.....	78
Figura B1. Análisis de Correspondencia Múltiple Grupo Infraestructura.....	82
Figura B2. Análisis de Correspondencia Múltiple Grupo Redes LAN.....	85
Figura B3. Análisis de Correspondencia Múltiple Grupo Servidores Windows.....	88
Figura B4. Análisis de Correspondencia Múltiple Grupo Servidores UNIX.....	91
Figura B5. Análisis de Correspondencia Múltiple Grupo Redes de Transmisión.....	94
Figura B6. Análisis de Correspondencia Múltiple Grupo ATC.....	96
Figura B7. Análisis de Correspondencia Múltiple Grupo Redes WAN.....	99
Figura B8. Análisis de Correspondencia Múltiple Grupo Servidores AS400.....	102
Figura B9. Análisis de Correspondencia Múltiple Grupo Aplicaciones SAP.....	105

Índice de Tablas

Tabla 3.1. Grupos operativos solucionadores seleccionados.....	38
Tabla 3.2. Grupos operativos solucionadores seleccionados.....	38
Tabla 3.3. Categorización de la variable tiempo entre falla (TEF).....	41
Tabla 3.4. Categorización de la variable duración.....	41
Tabla 3.5. Codificación de la variable impacto.....	42
Tabla 3.6. Ejemplo de matriz de confusión.....	48
Tabla 4.1 Matriz Cruzada de Causa vs. Falla.....	56
Tabla 4.2. Causa de la Falla.....	58
Tabla 4.3 Tipo de Falla.....	58
Tabla 4.4. Acción ejecutada para finalizar la falla.....	59
Tabla 4.5. Codificación de las variables del grupo Respaldo y Almacenamiento....	62
Tabla 4.6. Configuración de los parámetros de la máquina.....	64
Tabla 4.7 Matriz de confusión para el modelo final.....	65
Tabla A1. Matriz Cruzada de Acción vs. Falla.....	75
Tabla A2. Matriz Cruzada de Tipo de componente vs. Falla.....	76
Tabla A3. Matriz Cruzada de Impacto vs. Falla.....	77
Tabla B1. Tipo de falla.....	79
Tabla B2. Causa de la falla.....	80
Tabla B3. Acción ejecutada para finalizar la falla.....	81
Tabla B4. Tipo de falla.....	83
Tabla B5. Causa de la falla.....	84
Tabla B6. Acción ejecutada para finalizar la falla.....	84
Tabla B7. Tipo falla.....	86
Tabla B8. Causa de la falla.....	87
Tabla B9. Acción ejecutada para finalizar la falla.....	87
Tabla B10. Tipo falla.....	89
Tabla B11. Causa de la falla.....	90
Tabla B12. Acción ejecutada para finalizar la falla.....	90
Tabla B13. Tipo falla.....	92
Tabla B14. Causa de la falla.....	93
Tabla B15. Acción ejecutada para finalizar la falla.....	93
Tabla B16. Tipo falla.....	95
Tabla B17. Causa de la falla.....	96

Tabla B18. Acción ejecutada para finalizar la falla.....	96
Tabla B19. Tipo falla.....	98
Tabla B20. Causa de la falla.....	98
Tabla B21. Acción ejecutada para finalizar la falla.....	99
Tabla B22. Tipo falla.....	101
Tabla B23. Causa de la falla.....	101
Tabla B24. Acción ejecutada para finalizar la falla.....	101
Tabla B25. Tipo falla.....	103
Tabla B26. Causa de la falla.....	104
Tabla B27. Acción ejecutada para finalizar la falla.....	104

Agradecimiento

Estoy muy consciente que aun cuando con el uso de estas sencillas palabras, me siento reducido al momento de expresar mi agradecimiento, les digo:

A mi Dios eterno y poderoso....

A la ilustre Universidad de Los Andes, a nuestra facultad de ingeniería, particularmente a mi escuela de ingeniería de sistemas y todos los departamentos que la integran por contribuir a mi formación profesional.

A los profesores Gerardo Colmenares y Liliana Capacho, por su orientación y excelente guía durante la realización de este proyecto.

Al ingeniero José Casanova por el apoyo brindado dentro y fuera de Empresa.

Al equipo de la Coordinación de Confiabilidad de PDVSA-AIT-SCC, por el apoyo brindado y confianza transmitida.

A mi amigo y compañero de clases Carlos Medina, quien desarrolló su proyecto de grado en la misma área. Hermano muchas gracias por el apoyo incondicional.

A mi familia toda, especialmente a mi tío José Torrejano por su apoyo incondicional, para él mil gracias, a mis primos que son mis mejores y más sagrados amigos.

A todos mis compañeros de clase y compañeros de vida.

A todos mis más sinceros agradecimientos.....!

Capítulo 1

Introducción

Petróleos de Venezuela S.A. se ha caracterizado siempre por la ejecución de políticas innovadoras para la coordinación de sus procesos. Tales políticas han permitido que el producto final de la Empresa sea altamente conocido tanto a nivel nacional como internacional, haciéndola acreedora de una altísima reputación como generadora de productos y servicio de buena calidad. Los procesos que se llevan a cabo dentro de la Empresa abarcan todos los aspectos de la industria petrolera, por ello varían en su naturaleza y a la vez varía el tipo de equipos que intervienen como apoyo tecnológico para la producción. Se requiere que dichos equipos operen de forma continua para garantizar una máxima producción, sin embargo, su naturaleza física hace que con el paso del tiempo sus elementos se deterioren, haciendo necesario detener con cierta frecuencia las operaciones que llevan a cabo para efectuar labores de mantenimiento que permitan prolongar su vida útil.

La detención de cualquier proceso productivo, sin importar su naturaleza, trae una serie de costos asociados. Por lo tanto, se hace necesario organizar los períodos en los cuales se llevarán a cabo las actividades de mantenimiento para garantizar que el costo incurrido al paralizar el proceso sea mínimo. Esto hace que empresas como PDVSA traten de buscar herramientas que, basadas en principios

teóricos sólidos y comprobables, permitan orientar eficientemente el proceso de toma de decisiones asociado con la gestión del mantenimiento. Entre las herramientas teóricas utilizadas se puede mencionar la Minería de Datos.

En los últimos años se ha observado cómo la Minería de Datos juega un papel importante al momento de analizar datos históricos con el fin de obtener información de utilidad para apoyar el proceso de toma de decisiones dentro de una organización. Vale la pena destacar que la minería de datos abarca un conjunto de técnicas y herramientas que no se encuentran asociada a ningún dominio en particular, pues son aplicables a grandes volúmenes de datos organizados bajo algún contexto específico, tal es el caso de los datos históricos de la plataforma tecnológica de PDVSA AIT Servicios Comunes Centro (SCC). La aplicación de esta técnica en empresas de gran magnitud como PDVSA resulta fundamental ya que permiten realizar análisis riguroso a los registros de fallas ocurridos en los servicios brindados en la plataforma. Técnicas como: Análisis Exploratorio de Datos, Reconocimiento Automático de Patrones, Reglas de Asociación, entre otras, son algunas de las más utilizadas dentro de este campo para extraer conocimiento oculto en los grandes almacenes de datos.

El presente proyecto propone la aplicación de técnicas de Minería de Datos para caracterizar o describir el comportamiento en los modos de fallas de los Equipos y Sistemas que constituyen la plataforma tecnológica de PDVSA AIT SCC, los cuales se encuentran bajo la responsabilidad de la Gerencia de Automatización Informática y Telecomunicaciones de la Empresa (AIT).

1.1 Antecedentes

En los últimos años, las industrias han venido cambiando la manera de ver el negocio, adoptando herramientas e implementando nuevas técnicas y metodologías que permitan la optimización de sus procesos.

Bajo estas premisas en el año de 1995 en la Refinería de Cardón se propone realizar cambios profundos en sus sistemas de mantenimiento, con el fin de lograr un mayor desempeño. Estos cambios se fundamentan en la aplicación de nuevas metodologías que ayuden a mitigar los problemas ocasionados por paradas inesperadas; para ello se implementaron nuevas técnicas basadas en ingeniería de confiabilidad, análisis de criticidad, mantenimiento centrado en confiabilidad, análisis causa raíz, entre otras, las cuales permitieron mejorar la disponibilidad y confiabilidad de los activos en operación.

El entusiasmo por estas novedosas metodologías y herramientas, motivo al personal de esta refinería dedicar tiempo completo para entender y estudiar nuevos métodos y conceptos. Con los logros y experiencias obtenidos allí, emerge un alto interés por incorporar en toda la industria sistemas de mantenimiento basados en estos enfoques, los resultados obtenidos en Cardón fueron divulgados en toda la organización, capturando la atención de la Gerencia AIT.

La Gerencia AIT de PDVSA, se encarga de realizar las actividades fundamentales requeridas para garantizar el adecuado funcionamiento y continuidad operacional de las soluciones tecnológicas que apoyan la organización.

Por ello, el proceso de mantenimiento operativo resulta esencial dentro de la Gerencia AIT. Tanto así, que existe una subgerencia dedicada a la ejecución de las tareas de mantenimiento en la plataforma: la Gerencia de Mantenimiento

Operacional de la Plataforma (MAP). Esta gerencia a su vez, dedica recursos no solo para la ejecución de mantenimientos correctivos al momento de producirse fallas en los servicios, sino también coordina y ejecuta tareas de planificación y optimización del mantenimiento. La Coordinación de Confiabilidad de la Plataforma de la Gerencia de Mantenimiento de AIT SCC, tiene como finalidad de brindar apoyo a los procesos de optimización y control de la plataforma, diagnosticando el estado de las soluciones y generando indicadores que miden la efectividad de la gestión del mantenimiento. Por su parte, la Coordinación de Planificación y Programación del mantenimiento es la que se encarga de las tareas de planificación de los mantenimientos realizados.

Confiabilidad de la Plataforma tiene como insumo básico para generar sus Estadísticas, las Minutas de Reunión de Incidencias Diarias, recibidas diariamente de los Equipos Operativos donde se informa alguna novedad, para hacerle seguimiento. Estas minutas son almacenadas en un repositorio de incidentes de falla que proporciona un registro histórico de los datos de las fallas ocurridas.

1.2 Planteamiento del problema

La Coordinación de Confiabilidad dentro de la Gerencia MAP de PDVSA AIT SCC mantiene un registro histórico de las fallas ocurridas en la plataforma y reportadas por los analistas encargados de resolverlas durante las reuniones de incidentes diarios.

Actualmente este registro se lleva en una base de datos y es utilizado exclusivamente para la obtención de indicadores de disponibilidad y falla de la plataforma y para realizar análisis basados en consultas simples a tal repositorio.

Esto hace que el potencial de la información almacenada no sea explotado como debería, lo que ocasiona que se ignoren las relaciones subyacentes entre las variables almacenadas.

Buscando obtener mayor conocimiento de la naturaleza de la falla y las relaciones entre las variables almacenadas dentro del repositorio de incidente de fallas, la Coordinación de Confiabilidad desea aplicar técnicas de Minería de Datos para descubrir patrones ocultos entre dichas variables con el fin de sugerir medidas para optimizar la gestión del mantenimiento de la plataforma.

1.3 Justificación

La implementación de las técnicas de minería de datos ayuda a generar recomendaciones apropiadas para el diseño de planes de mantenimiento capaces de garantizar el funcionamiento de los activos de la empresa en todos sus niveles. Estas técnicas aplicadas a los procesos industriales, específicamente en el área de mantenimiento, aportan un gran beneficio ya que permiten dar respuestas a preguntas tales como, por ejemplo: ¿Existen equipos que sean más susceptibles a ciertos modos de fallas que otros?, ¿Para cuáles modos de falla el tiempo de solución es mayor?, ¿Existirán otros factores aparte de los que ya son considerados dentro del análisis de criticidad que permitan decir que un equipo resulta crítico? Encontrar respuestas a estas interrogantes permite, entre otras cosas, mejorar el desempeño de las actividades de mantenimiento y maximizar la disponibilidad de los activos que se encuentran en producción o brindando algún tipo de servicio específico.

Es importante conocer la forma como se relacionan las variables que integran el registro de fallas, ya que a partir de allí se pueden determinar situaciones

particulares que no son apreciables a simple vista y con ello detectar patrones en los modos de fallas que permitan sugerir medidas para optimizar el mantenimiento preventivo, actuando directamente sobre los equipos y sistemas de mayor vulnerabilidad y en consecuencia aumentando su disponibilidad.

Dentro de estas premisas, la minería de datos (*Data Mining*) juega un papel importante, ya que permite un mayor entendimiento de las relaciones que existen entre las variables estudiadas pero que no son observables en forma explícita, incorporando conocimiento importante sobre las datos estudiados; en este caso los datos de falla de la plataforma tecnológica de PDVSA AIT SCC.

1.4. Objetivos

1.4.1. Objetivo general

Encontrar mediante la aplicación de Minería de Datos patrones útiles para describir el comportamiento de las fallas que ocurren en los Equipos y Sistemas de la Plataforma de PDVSA AIT SCC.

1.4.2. Objetivos específicos

- Recopilar e integrar los datos históricos de fallas de todos los equipos y sistemas que están bajo manejo y custodia de PDVSA AIT SCC.
- Realizar las transformaciones y limpieza de los datos originales con el fin de construir una vista minable.

- Definir mediante el uso de técnicas de análisis multivariantes, variables latentes que representen las relaciones existentes entre las variables observadas.
- Determinar el modelo de minería de datos que mejor se adapte a los datos estudiados.
- Aplicar el modelo encontrado para obtener conocimiento oculto a partir de las variables observadas.

1.5. Metodología

La metodología que se utiliza para desarrollar el presente proyecto está basada en las etapas que sigue el método de descubrimiento de conocimiento en base de datos (*Knowledge Data Discovery, KDD*). El método abarca las siguientes fases:

- Determinar los objetivos: Se definen los objetivos y el alcance del proyecto.
- Preprocesado de los datos: Se organizan, limpian y depuran los registros históricos correspondientes a las fallas de los equipos y sistemas.
- Selección de variables: Se eliminan las variables que son poco relevantes para el estudio, y se crean nuevas variables que aportan información de interés para la fase de aplicación de minería de datos (crear la vista minable).
- Determinación del modelo: A través de la aplicación de las técnicas de minería de datos se determina el modelo de conocimiento que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables.

- **Análisis de resultados:** Para finalizar se interpretan los resultados obtenidos, verificando sus consistencias. En caso de ser necesario, se vuelve a las fases anteriores.

1.6 Organización del documento

En el capítulo 1 se presenta una introducción que describe en forma general el contenido del proyecto, así como: antecedentes, planteamiento del problema, justificación, objetivo general y específicos, metodología, estructura del documento y por ultimo el área de desempeño o descripción de la empresa donde se desarrollo el mismo.

En el capítulo 2 se describen algunos tópicos acerca de las metodologías, conceptos y herramientas utilizadas en este proyecto. Se mencionan conceptos de tales cómo: análisis de falla, análisis de criticidad, Ingeniería de confiabilidad, tipos de mantenimientos, se habla acerca de los métodos que existen para la recolección y análisis de datos históricos, minería de datos, técnicas de reconocimiento de patrones en forma automática.

En el capítulo 3 se analizan los modos de falla en la plataforma de PDVSA AIT SCC, mediante las técnicas de minería de datos, se determina la técnica de análisis multivariante que mejor aplique para el conjunto de datos de estudio, en líneas generales se aplican las fases utilizadas de *KDD* tales como: preprocesado de los datos, estadísticas fundamentales, se preparan las instancias para el modelo experimental y se determina el modelo de minería.

El capítulo 4 contiene detalles acerca de la construcción del modelo de reconocimiento de patrones en forma automática, se realizan los diferentes pruebas

para clasificar los tipos de mantenimientos correctivos realizados a ciertos modos de fallas, usando para ello Maquinas de vector Soporte, también se describen las diferentes pruebas realizadas con el fin de conseguir los valores o parámetros del modelo que garanticen el mayor numero de clasificaciones correctas.

El capítulo 5 se refiere a las conclusiones y recomendaciones finales del proyecto.

1.7 Descripción de la Empresa

Petróleos de Venezuela S.A. (PDVSA) es la corporación estatal venezolana que se encarga de la exploración, producción, manufactura, transporte y mercadeo de los hidrocarburos del país. Por mandato de la Constitución de la República Bolivariana de Venezuela, la totalidad de las acciones de Petróleos de Venezuela S.A. pertenecen al Estado Venezolano y se encuentra adscrita al Ministerio del Poder Popular para la Energía y Petróleo. Es una de las mayores empresas petroleras a nivel mundial y ha sido catalogada como una de las empresas más grandes del mundo. Actualmente, es la petrolera con mayores reservas petrolíferas del mundo, alcanzando una suma total de 3,1 billones de barriles. Sus instalaciones se encuentran distribuidas a lo largo de todo el territorio nacional y posee varias filiales nacionales e internacionales.

PDVSA cumple con todas las actividades propias del negocio petrolero, constituyéndose en una corporación que abarca todos los procesos, desde la explotación hasta la comercialización de los hidrocarburos gaseosos y no gaseosos, y sus derivados, se divide en las siguientes cuatro unidades de trabajo:

- Exploración y Producción: Área encargada de la evaluación, exploración, certificación y perforación de yacimientos petroleros. Es el primer eslabón de la cadena, cubre además la perforación y construcción de los pozos petrolíferos.
- Refinación: Área encargada de la separación, mejoramiento y obtención de productos o derivados del petróleo a través de plantas de procesamiento y refinerías.
- Distribución y comercialización: Área encargada de colocar los productos obtenidos (Crudo y derivados) en los diferentes mercados internacionales.
- Gas: Con unas reservas probadas por 147 billones de pies cúbicos, Venezuela es una de las potencias mundiales del sector de hidrocarburos gaseosos. (sitio Web de PDVSA)

1.7.1 PDVSA-AIT

Para dar apoyo a sus procesos de negocio y ejecutar éstos de la manera más eficiente, PDVSA requiere de una infraestructura tecnológica de vanguardia. Es por ello que existe una gerencia dentro de su organigrama dedicada exclusivamente a proveer, suministrar y coordinar los servicios y las soluciones integrales en toda el área que abarca las tecnologías de automatización, información y comunicación; ésta gerencia es la Gerencia AIT de PDVSA (Gerencia de Automatización, Información y Comunicación). Esta gerencia no sólo se encarga de contribuir a mantener la continuidad operativa de la plataforma tecnológica de la empresa, sino que también coordina y ejecuta planes para mantener dicha plataforma actualizada, todo ello buscando propiciar un ecosistema tecnológico que propicie los poderes creadores

del pueblo, el conocimiento libre, el desarrollo endógeno sustentable y la economía social productiva con el fin de alcanzar la soberanía tecnológica. (Sitio Web de PDVSA)

1.7.1.1 Organización

PDVSA AIT se encuentra dividida en unidades dedicadas a dar apoyo a cada uno de los procesos del negocio petrolero y a las filiales más importantes de la corporación. Adicionalmente existen unidades de apoyo y coordinación de recursos para la ejecución de proyectos de mayor alcance en la empresa. (Sitio Web de PDVSA)

1.7.1.2 AIT Servicios Comunes Centro

Dentro de la gerencia AIT, la línea de servicios comunes es la responsable de coordinar los recursos necesarios para garantizar la continuidad operativa de los servicios AIT que apoyan los negocios de PDVSA que conviven en una región determinada. Existen tres unidades AIT de Servicios Comunes: Oriente, Occidente y Centro.

AIT Servicios Comunes Centro comprende los servicios de PDVSA AIT en las regiones metropolitana, centro y sur. Entre los servicios que abarca se pueden mencionar el mantenimiento de la plataforma, la gestión del servicio a los usuarios, la implantación de nuevas soluciones, la gestión de necesidades y oportunidades en el negocio, entre otras.

El presente proyecto fue desarrollado dentro de la Gerencia de Mantenimiento de la Plataforma (MAP) de PDVSA AIT Servicios Comunes Centro, en la unidad

denominada confiabilidad de la Plataforma ubicada en las instalaciones de Intevep Los Teques, estado Miranda. (Sitio Web de PDVSA)

1.7.1.3 AIT Confiabilidad

La Coordinación de Confiabilidad de la Plataforma de la Gerencia de Mantenimiento de AIT SCC, tiene como finalidad de brindar servicios a los diferentes Procesos de MAP que apoyarán a la Gestión para las Mejoras de las Instalaciones, Procesos, Sistemas y Equipos Asociados de la Plataforma Tecnológica de la Corporación en la Región Centro.

Confiabilidad de la Plataforma tiene como insumo básico para generar sus Estadísticas de Indisponibilidad y los Análisis de Criticidad, las Minutas de Reunión de Incidencias Diarias, recibidas diariamente de los Equipos Operativos donde se informa alguna novedad, para hacerle seguimiento. (Sitio Web de PDVSA). La figura 1.1 muestra una representación de la Coordinación de Confiabilidad.

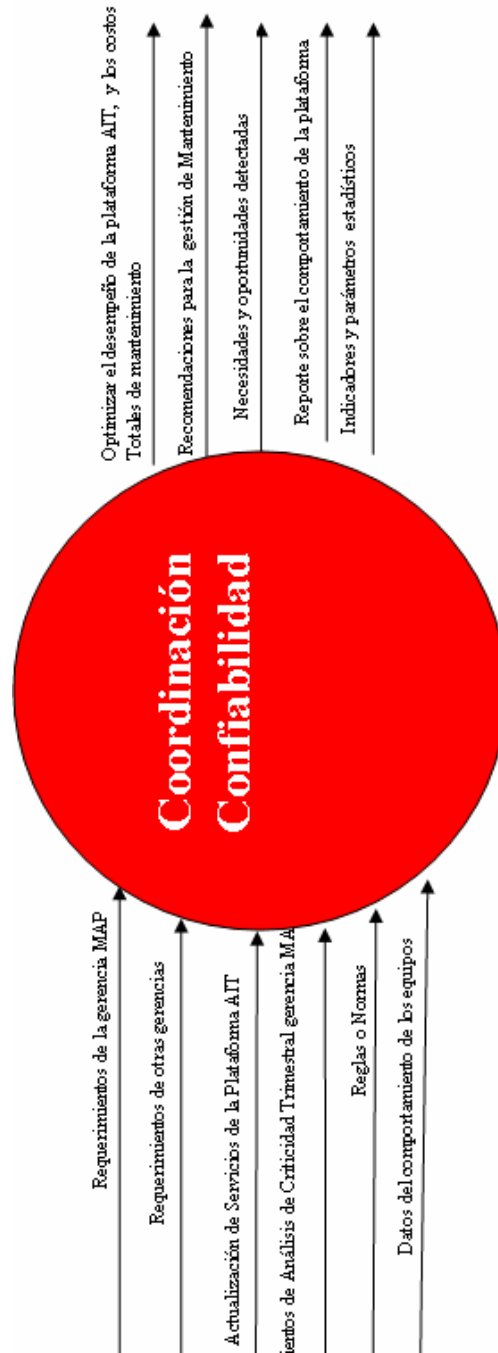


Figura 1.1. Representación de la Coordinación de Confiabilidad

Capítulo 2

Marco teórico

En el presente capítulo se explican algunos términos utilizados en el desarrollo de modelos basados en Minería de Datos y en Confiabilidad Operacional debido a su importancia en el contexto de la gestión de mantenimiento. Además, se describen algunos aspectos teóricos relacionados con el proyecto.

Se desarrolla un marco de referencia con el fin de entender algunos conceptos de utilidad para el presente proyecto. Para ello, se comienza justificando de manera general la importancia de almacenar de forma organizada la información de las fallas en los activos de una organización, se comenta acerca de los elementos presentes en los modelos de reconocimiento automático de patrones y de algunos parámetros e indicadores utilizados en la ingeniería de confiabilidad. También se mencionan algunas de las técnicas usadas en la minería de datos.

2.1 Conceptos básicos

Antes de comentar la importancia que reviste almacenar la información de las fallas de los equipos y sistemas, se debe conocer algunos términos que guardan relación

con el proyecto. A continuación se suministran algunos conceptos básicos de acuerdo a DRAE (2008):

Dato: antecedente necesario para llegar al conocimiento exacto de algo o para deducir las consecuencias legítimas de un hecho, es una representación simbólica, atributo o característica de una entidad. El dato no tiene valor semántico o un sentido en sí mismo, pero convenientemente tratado (procesado) se puede utilizar en la realización de cálculos o toma de decisiones. Para el caso de estudio los datos vienen a ser los obtenidos de las bitácoras de mantenimiento, levantadas en los diferentes formatos o documentos fuentes destinados para tal fin.

Información: comunicación o adquisición de conocimientos que permiten ampliar o precisar los que se poseen sobre una materia determinada. Los datos organizados en los formatos establecidos vienen a ser el insumo necesario para obtener información.

Registro: libro, a manera de índice, donde se apuntan noticias o datos, estructurado de tal forma que cada uno de sus componentes ocupa un espacio denominado campo. Los registros se refieren a la información almacenada por el gestor de datos, de acuerdo a las políticas definidas previamente por el equipo de confiabilidad para conservar el reporte de alguna falla presentada en algún momento dado.

Archivos: Conjunto ordenado de documentos que una persona, una sociedad, una institución, entre otros, producen en el ejercicio de sus funciones o actividades. En este caso los archivos son colecciones de registros de los incidentes de falla.

Base de Datos: El concepto de base de datos se refiere a un conjunto exhaustivo no redundante de datos estructurados, organizados independientemente de su utilización y su implementación en un medio de almacenamiento secundario. Consiste en la recolección organizada de un conjunto de archivos, de manera que

existe un orden cronológico que permite la recuperación de información de manera sencilla y la entrega en tiempo real de cualquier dato requerido.

Una vez indicado este glosario, se puede abundar en detalle sobre la importancia de almacenar la información para su posterior análisis. Al tener un repositorio histórico se pueden establecer patrones de comportamiento, de tal manera que se detecten similitudes y relaciones particulares que no han sido observadas directamente, pero que si pueden ser reconocidas mediante la generación de modelos o patrones de comportamiento análogos en diferentes instancias. Obtener nuevas variables o no observadas que permitan describir de alguna manera las relaciones existentes en las variables observadas.

El uso de almacenes de datos para registrar los datos de falla constituye un enfoque alternativo al enfoque tradicional de mantenimiento, en el que las acciones correctivas se realizan sobre la marcha, sin consultar fuentes secundarias y buscando garantizar la disponibilidad inmediata del componente por encima de prevenir futuros incidentes. La consulta de registros históricos de falla proporciona una base de conocimientos acerca de la naturaleza de las fallas y permite incorporar mejoras sustanciales en el efecto de los mantenimientos planificados, disminuyendo la ocurrencia de paradas inesperadas y los costos asociados al mantenimiento. En consecuencia, tomar decisiones certeras sobre la base de datos almacenados que proporcionen al analista la información necesaria para ubicar el factor crítico y la variable afectada, así como determinar el tipo de mantenimiento asociado a los modos de fallas que tengan ciertas características.

2.2 Ingeniería de confiabilidad

Inicialmente se definirá un conjunto de términos que se encuentran íntimamente vinculados con el análisis de confiabilidad.

Los siguientes conceptos fueron expuestos en las investigaciones de Mosquera (2000).

2.2.1 Definición

Confiabilidad: es la propiedad de un sistema o equipo de cumplir las funciones para él previstas, manteniendo su capacidad de trabajo bajo los regímenes y condiciones prescritos y durante el intervalo de tiempo requerido. Dicho de otra forma, la confiabilidad es la propiedad del sistema de mantenerse sin experimentar un suceso de falla durante el tiempo y las condiciones de explotación establecidos.

Falla: suceso después del cual un sistema tecnológico deja de cumplir (total o parcialmente) sus funciones. La falla es la alteración de la capacidad de trabajo del componente o sistema.

Mantenibilidad: es la probabilidad de que un sistema, subsistema o equipo que ha fallado pueda ser reparado dentro de un período de tiempo determinado.

Disponibilidad: Es la probabilidad que un sistema, subsistema o equipo este disponible para su uso durante un tiempo dado.

La ingeniería de confiabilidad se define como la disciplina técnica encargada de estimar, controlar y gerenciar la probabilidad de fallas en dispositivos, equipos o sistemas, para así garantizar una alta disponibilidad y confiabilidad en estos.

De la definición anterior, resulta esencial poder cuantificar de alguna manera el valor de la confiabilidad. Por lo general tiende a medirse en términos de probabilidades, pues el concepto de confiabilidad está íntimamente vinculado con el de probabilidad. En este sentido, en la práctica, la ingeniería de confiabilidad se basa en la determinación de indicadores estadísticos que permitan cuantificar esa probabilidad. Para la obtención de dichos indicadores resulta necesario un conocimiento integral de las actividades de mantenimiento, es por ello que el presente proyecto pretende brindar un insumo de gran valor para la aplicación de ingeniería de confiabilidad dentro de PDVSA AIT SCC.

Adicionalmente, el conocimiento de los parámetros de confiabilidad y mantenibilidad es determinante en el pronóstico de la disponibilidad de cualquier dispositivo, sistema o equipo. Mediante estos parámetros se proporcionan los datos fundamentales para el análisis del mantenimiento, generando de ese modo gran cantidad de información técnica que resulta vital para la toma de decisiones.

2.2.2 Clasificación de las fallas

De acuerdo con Mosquera (2000), las fallas pueden clasificarse de la siguiente manera.

Fallas catastróficas vs. paramétricas: la falla catastrófica conduce a la alteración de la capacidad de trabajo. Estas fallas corresponden a eventos físicos tales como: cortocircuitos, fracturas, deformaciones y atascamiento de las piezas, entre otros. Las fallas paramétricas conllevan a una reducción de la capacidad de trabajo, pero no a su interrupción total.

Fallas dependientes vs. independientes: las fallas, son independientes si al fallar un o varios elementos de un sistema no produce falla en otros elementos. Si la aparición de la falla en un elemento o si la probabilidad de ocurrencia de la falla ha cambiado con la falla de otros elementos, esta falla será un hecho dependiente

Fallas repentinas vs. graduales: las fallas repentinas son fallas inesperadas las cuales aparecen como consecuencia de la variación brusca de los parámetros fundamentales, algunas veces ocasionada por errores del personal de servicio. En las fallas graduales se observa la variación de algunos parámetros de operación debido al envejecimiento y al desgaste por uso de los elementos o de todo el sistema.

Fallas estables vs. inestables: las fallas estables se eliminan sólo con la reparación o la regulación, o en algunos casos con la sustitución del elemento que fallo. Las fallas inestables pueden desaparecer espontáneamente sin la intervención del personal de servicio debido a la desaparición de los motivos que la provocaron.

Fallas de interrupción vs. de bloqueo: las fallas de interrupción producen en el equipamiento en operación, interrumpiendo su trabajo. Las fallas de bloqueo impiden el arranque o puesta en funcionamiento de sistemas o componentes.

Fallas revelables vs. ocultas: las fallas revelables son aquellas que se exteriorizan al personal de operación inmediatamente después de su ocurrencia, las mismas se detectadas a través del la implementación de sistema de control. Se trata de fallas de sistemas en funcionamiento a la espera con control de sus parámetros. Las fallas ocultas no se revelan al personal de operación por ninguna vía en el momento de su ocurrencia, pero la condición de falla permanente esta latente hasta ser descubierta por una prueba o sobre la demanda de operación del sistema en cuestión.

Fallas primarias vs. Secundarias: las fallas primarias son intrínsecas del elemento y responden a sus características internas. Las fallas secundarias son debidas a condiciones ambientales o tensiones operativas excesivas impuestas a un elemento desde el exterior.

2.2.3 Estimación de la confiabilidad

De acuerdo a Yañez *et. al* (1995), para la estimación de la confiabilidad o la probabilidad de fallas, existen dos métodos que dependen del tipo de datos disponible; estos son:

Estimación Basada en Datos de Condición: altamente recomendable para equipos estáticos, que presentan patrones de "baja frecuencia de fallas" y por ende no se tiene un "historial de fallas" que permita algún tipo de análisis estadístico.

Estimación Basada en el Historial de Fallas: recomendable para equipos dinámicos, los cuales por su alta frecuencia de fallas, normalmente permiten el almacenamiento de un historial de fallas que hace posible el análisis estadístico.

En el caso particular de este estudio, los equipos y sistemas que conforman la plataforma tecnológica de PDVSA-AIT están relacionados a la estimación de la confiabilidad basada en la historia de fallas. Para ello se desarrollarán los conceptos asociados a este tipo de estimación.

2.3 Indicadores de confiabilidad basados en el historial de fallas

Las técnicas de almacenamiento, recuperación y transformación de los datos históricos hacen posible la obtención de indicadores cuyo nivel de complejidad puede ser bajo, pues en muchos casos son calculados aplicando operaciones sencillas. Los indicadores obtenidos permiten estimar la confiabilidad y mantenibilidad. Por ejemplo, se observa poca confiabilidad cuando el tiempo medio entre fallas es largo comparado con el tiempo de servicio efectivo de un equipo. Así mismo, otro indicador de confiabilidad es cuando los valores de los índices de tiempo medio entre fallas son pequeños comparados con el tiempo de servicio efectivo. Por otro lado, al invertir los resultados de estos indicadores proporcionaría ciertas tasas que aportan información útil y muchas veces considerada mejor para efectos prácticos.

Para una adecuada toma de decisiones es necesario tener datos exactos acerca de la ocurrencia de las fallas. Por lo general, dentro de las industrias existen registros históricos de las fallas de los equipos, los cuales se encuentran almacenados en grandes bases de datos. La adquisición de los datos se reduce a la consulta en esas bases de datos. Sin embargo, la recuperación oportuna de los datos para la realización de mejoras en la confiabilidad resulta un factor crítico en muchas circunstancias. Muchas veces se considera más relevante la realización de mejoras inmediatas en los equipos y la corrección de las fallas, que el registro histórico de las acciones tomadas.

De acuerdo con Barringer (1996), la educación de los encargados de tomar las decisiones y de los recolectores de los datos acerca del valor de la información de los

bases de datos resulta vital para la mejora de los índices de confiabilidad, ya que se traduce en Indicadores de confiabilidad basados en historiales de fallas más precisos.

Hoy en día existen numerosos adelantos que permiten la creación y manipulación de grandes repositorios de datos, tal es el caso de los Almacenes de datos (*data warehouse*), donde se coleccionan grandes cantidades de datos pertenecientes a una institución, entidad u organización. Estos datos son almacenados en medios magnéticos que garantizan su permanencia en tiempo y espacio.

Los datos almacenados representan el comportamiento de fenómenos ocurridos en el ejercicio de las operaciones de la organización, cuyo propósito es favorecer el análisis y la divulgación eficiente de las prácticas ejecutadas. Los datos pueden ser recuperados cuando se requiera y ser sometidos a un proceso de transformación para producir información de interés que contribuya a la toma de decisiones dentro de la organización. Una vez dispuestos de forma organizada en medios físicos, están disponibles para su recuperación y procesamiento por medio de varias técnicas destinadas para ello, entre las que resalta la Minería de datos (*Data Mining*).

2.4 Minería de datos

Vallejos (2006) describe la minería de datos como un mecanismo de exploración de datos para la explotación o descubrimiento de nuevos conocimientos en grandes volúmenes históricos de datos mediante patrones de búsqueda capaces de construir

información valiosa inexistente, y en algunos casos inviable, por las técnicas clásicas de recuperación de información.

Con el pasar del tiempo la demanda de técnicas poderosas que permitan encontrar conocimiento útil a partir del procesamiento de datos se ha incrementado considerablemente. Un crecimiento importante de los volúmenes de datos almacenados es una de las causas principales. Entre estas técnicas se pueden mencionar el aprendizaje automático, el análisis exploratorio de datos, la inteligencia artificial y las redes neuronales. La Minería de datos se apoya en estas técnicas para descubrir conocimiento a través de relaciones ocultas entre los datos o de la misma información. Es decir, información similar a la que podría producir un experto humano. Trata de descubrir conocimientos interesantes en forma de patrones, reglas y estructuras significativas a partir de grandes almacenes de datos.

2.4.1 El proceso de descubrimiento de conocimiento en bases de datos (*Knowledge Data Discovery* o KDD)

Como su nombre lo indica, esta metodología permite obtener conocimiento oculto a partir de grandes volúmenes de datos. Usa técnicas de minería de datos (algoritmos, procedimientos) para extraer conocimiento que no es detectado a simple vista. Según Rivera (2006) esta metodología comprende las siguientes etapas:

- 1.- Determinación de los objetivos: Delimitar los objetivos que se desean en un proyecto específico
- 2.- Preprocesamiento de los datos: Se organizan, limpian y depuran los registros, utilizando para ello técnicas de análisis multivariante y técnicas de análisis exploratorio de datos, tales como: análisis factorial, análisis de componentes

principales, análisis discriminante, análisis de correspondencia múltiple, diagrama de cajas y bigotes, análisis de conglomerados, entre otros. Es importante resaltar que esta etapa consume gran parte del tiempo total de un proyecto de minería de datos, por razones propias de organización.

3.- Selección de las variables: Se seleccionan variables relevantes al estudio, y se crean nuevas variables que aporten información de interés para la fase de aplicación de minería de datos (se crea la vista minable). La selección de variables debe evitar la pérdida de calidad del modelo de conocimiento obtenido del proceso de minería.

4.- Determinación del modelo: se aplica el modelo de conocimiento que permita el conocimiento de patrones de comportamiento descubiertos en los valores de las variables del problema o en las relaciones o asociaciones entre variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesado diferente de los datos.

5.- Análisis de los resultados: finalmente se realiza una interpretación y evaluación de los resultados en el contexto del problema, verificando que los resultados obtenidos sean coherentes.

En la figura 2.1 se desarrolla el esquema de la metodología, en la que el conocimiento es obtenido a partir de una base de datos de interés.

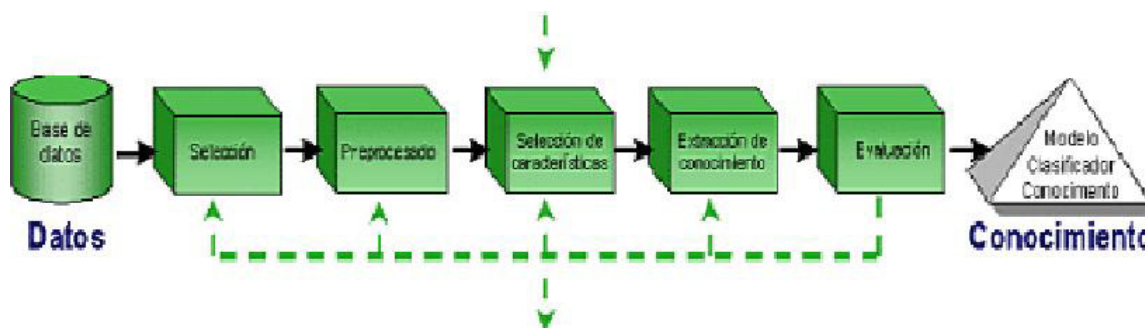


Figura 2.1. Etapas de la Minería de datos en un KDD

Algunos autores (como por ejemplo, Agrawal y Shafer (1996)) señalan que para cualquier estudio que involucre manejo de datos, por ejemplo KDD, deben guardar una alta integridad para evitar resultados erróneos según sea el caso de estudio. En este caso particular, debe evitar un falso conocimiento oculto en las variables involucradas. La calidad de los datos se puede mejorar al aplicar técnicas estadísticas para su preprocesamiento.

Las técnicas estadísticas de análisis exploratorio de datos y análisis multivariante han sido aplicadas en un número creciente de áreas de conocimiento. Son particularmente apropiadas para el estudio de grandes volúmenes de datos en los que es imposible, dado el tamaño, observar de inmediato sus características estructurales. El uso apropiado de técnicas para el análisis de datos como: diagramas de cajas y bigotes, análisis de correspondencia múltiple, análisis de componentes principales etc., brindan al investigador un primer acercamiento a las relaciones entre las variables. La técnica de preprocesamiento utilizada para los datos del estudio fue el análisis de correspondencia múltiple.

2.4.2 Análisis de correspondencia múltiple

Según Díaz (2002), el Análisis de Correspondencia Múltiple (ACM) es una técnica de reducción de variables y cuyo objetivo principal es la construcción de variables no observadas. Mide el grado de asociación presente entre un conjunto de variables; es decir, construye un diagrama cartesiano o mapa perceptual basado en la relación de dependencia e independencia de los atributos o categorías de las variables. Esto es, variables categóricas que muestran en un mapa su recomposición mediante la asociación de categorías o atributos para conformación de conglomerados a través de la varianza. Esos conglomerados están conformados por categorías de las variables originales y tendrían una varianza mínima internamente y máxima entre ellos.

El análisis de correspondencia múltiple surge en los años 60 por parte del físico y matemático francés Benzécri, con el fin de definir, describir e interpretar las relaciones entre variables categóricas a través de un gráfico geométrico. A pesar de que las técnicas matemáticas en las que se basa el ACM datan de las investigaciones de grandes estadísticos y matemáticos del siglo XIX y principios del siglo XX (como es el caso de Pearson y Fisher, por ejemplo), para la época no se disponía de herramientas que permitieran el cálculo complejo de los parámetros del ACM. Es hasta la segunda mitad del siglo XX que surgen propuestas para su implementación.

Una de las aplicaciones de ACM viene siendo la detección de variables latentes a partir de un conjunto de variables de estudio. Mediante un diagrama cartesiano se observa un mapa perceptual donde los puntos (categorías de las variables observadas) indican la relación o correspondencia que pudiera existir entre las variables de estudio; esta relación se puede observar cuando se forman algunos

conglomerados (concentración de puntos) que describen cierto comportamiento particular (patrón). También puede ocurrir el caso contrario, es decir, puntos que se encuentran fuera de estos conglomerados llamados valores atípicos (no representan ninguna relación en particular); el ACM combina cada una de las variables de estudio por separado a fin de determinar si se forman conglomerados dentro de estas.

2.4.3 Tareas realizadas por Minería de datos

Minería de datos realiza una o más de las siguientes tareas, según Vallejos (1995).

Descripción de clases: provee una clasificación concisa y resumida de un conjunto de datos y los distingue unos de otros. La clasificación de los datos se conoce como caracterización, y la distinción entre ellos como comparación o discriminación.

Asociación: es el descubrimiento de asociaciones, relaciones o correlación en un conjunto de datos. Las asociaciones se expresan como condiciones *atributo-valor* y deben estar presentes varias veces en los datos.

Predicción: esta función de la minería pronostica descubrimiento de asociaciones o relaciones posibles entre los valores de ciertos atributos en un conjunto de objetos.

Clasificación: construye un modelo de entrenamiento a partir de un conjunto de datos cuya clasificación de clase se conoce y construye un modelo de objetos para cada clase. Dicho modelo puede representarse con árboles de decisión o con reglas de clasificación que muestran las características de los datos. El modelo puede ser utilizado para la mayor comprensión de los datos existentes y para la clasificación de los datos futuros. En el presente proyecto, para realizar clasificación en los datos se utiliza la técnica denominada Máquina de Vectores de Soporte.

2.4.4 Máquina de Vectores de Soporte

Según Carreras *et. al* (2004), las Maquinas de Vectores de Soporte (*Support Vector Machines*) (SVM), también conocidas como Maquinas de Soporte Vectorial, constituyen una nueva técnica de clasificación que ha tomado mucha atención en los últimos años. Los fundamentos teóricos de las SVM están basados en la teoría del aprendizaje estadístico desarrollado por Vapnik y otros autores, a finales de los años 70. El modelo de SVM, tal como se conoce actualmente fue presentado en la década de los 90 por Vapnik, Boser, Guyon y Cortes, permitiendo pasar de la formulación teórica a las aplicaciones reales de reconocimiento de patrones, demostrando tener un gran desempeño e introduciéndose como una poderosa herramienta para resolver problemas de clasificación en las disciplinas del aprendizaje automático y la minería de datos. Actualmente las SVM han dado buenos resultados al ser utilizadas en problemas tanto de clasificación como de regresión, ya que incluyen aspectos y técnicas del aprendizaje automático, estadística, análisis funcional y optimización

Algunas de las características más importantes que presentan las SVM son el alto nivel de generalización y las funciones *kernels*, ya que ofrecen una alta capacidad de pronóstico para nuevas observaciones y la bondad de poder trabajar con una amplia gama de datos para su procesamiento. Las SVM “pertenecen a la familia de clasificadores lineales puesto que inducen separadores lineales o hiperplanos en espacios de características de muy alta dimensionalidad (introducidas por funciones núcleo o *kernel*) con un sesgo inductivo muy particular (maximización del margen)”, (De acuerdo con Carreras *et al.* (2004)). Así, las SVM

encuentran un hiperplano que separa y maximiza el margen entre las clases en dicho espacio de características.

Una SVM es, en principio un clasificador binario. Sin embargo, se pueden implementar dos métodos para abordar problemas de clasificación con tres o más clases. El primero, denominado “uno contra todos”, consiste en comparar cada clase con todas las demás, mientras que en el segundo método “uno contra uno”, cada clase se compara con las restantes de forma separada.

Parte del éxito que han tenido las SVM en la resolución de problemas reales, se debe a que son máquinas basadas en la teoría del aprendizaje estadístico con una enorme riqueza de representación, debido a que las máquinas pueden ir aprendiendo a través de ejemplos, las salidas correctas para ciertas entradas sin necesidad de conocer la naturaleza o distribución de los datos con que se desea trabajar, además de tener ciertas características que la han puesto en ventaja respecto a otras técnicas de clasificación y regresión. Es decir técnica no paramétrica.

Algunas de las características procedimentales particulares que se pueden citar son las siguientes:

- Existen pocos parámetros a ajustar; el modelo solo depende de los datos con mayor información.
- La estimación de los parámetros se realiza a través de la optimización de una función de costo, lo cual evita la existencia de un mínimo local.
- La solución de SVM es esparcida, esto significa que la mayoría de las variables son cero en la solución de SVM; lo que quiere decir que el modelo final puede ser escrito como una combinación de un número muy pequeño de vectores de entrada llamados vectores de soporte.

- Otra de las características atractivas por la que las SVM están ganando popularidad según Gunn (1998), es la incorporación del principio de Minimización de Riesgo Estructural, el cual ha demostrado ser superior al principio tradicional de Minimización de Riesgo Empírico, empleado por las redes neuronales y otros métodos lineales convencionales. Este autor señala que el principio de Minimización de Riesgo Estructural minimiza un límite superior de riesgo esperado, opuesto al principio de Minimización de Riesgo Empírico que minimiza el error en los datos de entrenamiento; esta diferencia dota a las SVM con una excelente capacidad de generalización, es decir, tener una alta capacidad de pronóstico para nuevas observaciones.

Para comprender los fundamentos teóricos y matemáticos de las SVM es necesario tener claro algunas definiciones, dentro de las cuales se encuentran:

La dimensión VC (*Vapnik-Chervonenkis*): es un concepto fundamental dentro de la teoría de aprendizaje estadístico y se define como el máximo número de puntos que puede separar de manera óptima cierto algoritmo de clasificación, es decir representa la capacidad de ciertas funciones para separar un conjunto de datos (puntos) de entrenamiento. Así, si la dimensión de VC es h , existe un conjunto de h puntos que pueden ser separados.

La maximización del margen: se justifica dentro de la teoría del aprendizaje y se enmarca en el Principio de Minimización de Riesgo Estructural. La maximización del margen se define como la distancia de las muestras de entrenamiento a la frontera de decisión. Carreras *et al.* (2004), señala que la maximización del margen consiste en seleccionar el hiperplano separador que está a la misma distancia de los ejemplos más cercanos de cada clase, donde los ejemplos se refieren a los datos

utilizados para el entrenamiento; además exponen que el hiperplano debe estar ubicado en la posición más neutra posible con respecto a las clases representadas por el conjunto de datos, sin estar sesgado hacia la clase más numerosa. También mencionan que dicho hiperplano solo considera los vectores de soporte; es decir, aquellos puntos que están en las fronteras de la región de decisión, que es la zona donde puede haber dudas sobre a que clase pertenece un ejemplo. El modelo más simple de las SVM, conocido como clasificador de margen máximo, funciona sólo para datos linealmente separables en el espacio de entrada, por lo que no puede ser usado en muchas aplicaciones de la vida real. La figura 2.2 muestra un ejemplo de una SVM usada como clasificador, en ella se señalan los vectores de soporte que determinan el límite del hiperplano.

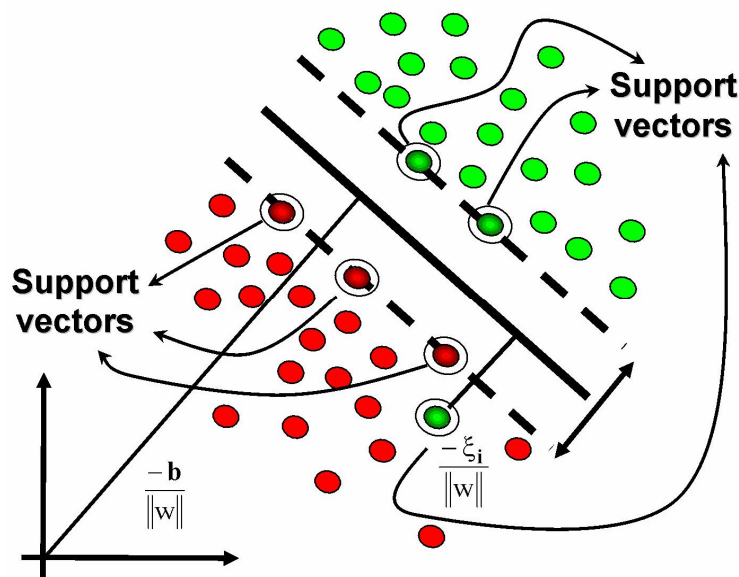


Figura 2.2 Clasificación usando SVM

Capítulo 3

Aplicación de minería de datos para caracterizar las fallas

Este capítulo hace referencia al proceso de organización y limpieza de los datos, los cuales se originan a partir de los incidentes reportados entre enero del 2006 y junio del 2008 (periodo de estudio). Se menciona acerca del proceso de transformación y reorganización de los mismos, la forma en que se adecuan y se preparan para el estudio. De igual modo, se comenta sobre los criterios tomados en cuenta para la construcción de la vista minable. También se explica la forma como se aborda el análisis descriptivo y exploratorio a los datos y el uso de las técnicas de análisis multivariante para la conformación de una nueva variable que explique la relación oculta entre las variables observadas.

3.1 Descripción de la naturaleza de los datos

Dado que una de las actividades de la Coordinación de Confiabilidad de la Plataforma es brindar apoyo a los procesos de la Gerencia MAP mediante la

generación de Estadísticas de Indisponibilidad y Análisis de Criticidad de la plataforma, es necesario disponer de algún medio de recopilación de los incidentes de falla ocurridos; esto se logra a través de la revisión de las minutas de incidentes diarios. Es decir, diariamente los analistas de confiabilidad extraen los datos de los reportes hechos por los analistas de guardia de las fallas ocurridas en la plataforma y los integran en una base de datos. A partir de este repositorio, se calculan indicadores de disponibilidad y fallas en la plataforma, los cuales sirven de insumo para la aplicación de técnicas de ingeniería de confiabilidad. El proceso que se sigue actualmente se puede observar en la figura 3.1.

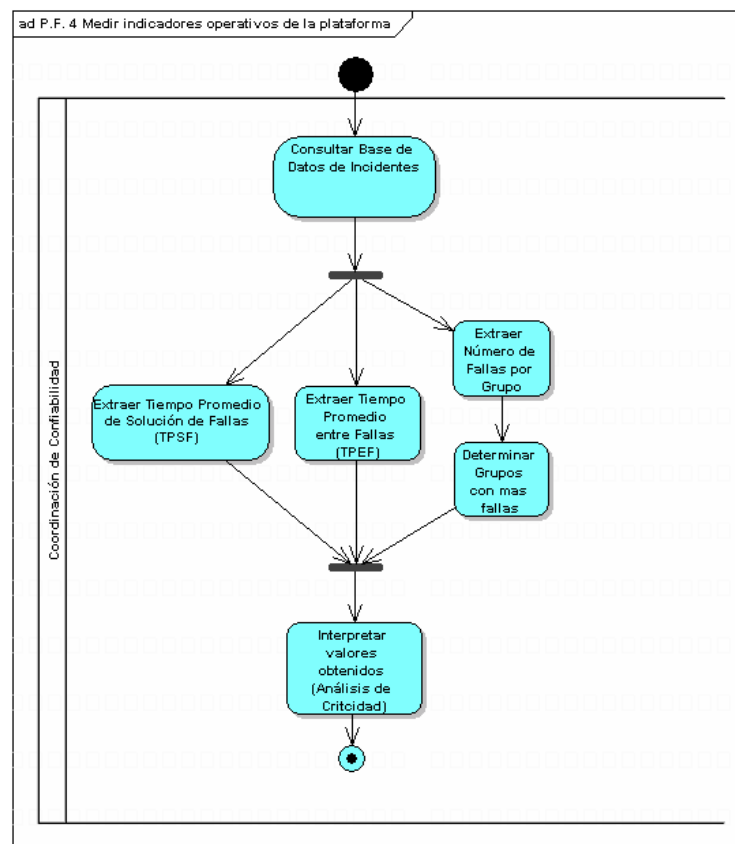


Figura 3.1 Descripción de las actividades realizadas para generar los indicadores.

	A	B	C	D	E	F	G	H	I	J	K	L
	Fecha de minuta	Nº Falla	Novedad	Fecha inicio	Hora Inicio	Fecha fin	Hora Fin	Duración	Tiempo entre fallas	Causa	Acciones tomadas	Impacto
1	N/A	0		1/01/2006	0.3				221.0			
3	11-Jan	1	Caída de uno de	10/01/2006	0.5	10/01/2006	0.5	0.0	189.5	Se necesita la	Migración de los ser	ALTO, afectó a
4	19-Jan	2	Interrupción del	18/01/2006	0.4	18/01/2006	0.8	8.0	40.0	el analista de	Se reconectó la fibra	Debe ser
5	20-Jan	3	Falla del backup	20/01/2006	0.1	20/01/2006	0.1	0.2	485.5	La falla del offline		
6	10-Feb	4	El backup del	9/02/2006	0.3	9/02/2006	0.3	0.2	72.0			
7	13-Feb	5	Respaldo NO	12/02/2006	0.3	12/02/2006	0.3	0.2	24.0			
8	14-Feb	6	Respaldo	13/02/2006	0.3	13/02/2006	0.3	0.2	24.0			
9	15-Feb	7	sigetell diario	14/02/2006	0.3	14/02/2006	0.3	0.2	24.0			
10	16-Feb	8	No falló ningún	15/02/2006	0.3	15/02/2006	0.3	0.2	0.0			
11	20-Feb	9	Critico: Sin	15/02/2006	0.3	15/02/2006	0.3	0.2	120.0			
12	21-Feb	10	Otros	20/02/2006	0.3	20/02/2006	0.3	0.2	312.0			
13	6-Mar	11	1) El servidor	5/03/2006	0.3	5/03/2006	0.3	0.2	96.0			
14	10-Mar	12	reporte de fallas:	9/03/2006	0.3	9/03/2006	0.3	0.2	48.0			
15	13-Mar	13	Como	11/03/2006	0.3	11/03/2006	0.3	0.2	219.5			
16	21-Mar	14	Falla Respaldo	20/03/2006	0.5	20/03/2006	0.5	0.0	90.5	Falla de hardware	Se reprogramara	Bajo
17	24-Mar	15	Falla Respaldo	24/03/2006	0.3	24/03/2006	0.3	0.2	0.9	Falla de hardware	Se reprogramará pa	Bajo
18	24-Mar	16	Falla Respaldo	24/03/2006	0.3	24/03/2006	0.3	0.2	16.1	Falla de hardware	Se espera por el	Bajo
19	24-Mar	17	Falla el respaldo	24/03/2006	1.0	24/03/2006	1.0	0.2	61.5	falta de cinta	Se añaden cintas y	Bajo
20	27-Mar	18	falló el backup del	27/03/2006	0.5	27/03/2006	0.5	0.2	155.2	los dispositivos	Se reprogramará	Bajo
21	3-Apr	19	Falla del Robot#	2/04/2006	1.0	2/04/2006	1.0	0.2	10.8	El robot mostró el	Se llamó al	Mediano
22	4-Apr	20	Falla del Backup	3/04/2006	0.4	3/04/2006	0.4	0.2	24.0	Error escribiendo	Se reprogramara el	Mediano
23	5-Apr	21	Falla del Backup	4/04/2006	0.4	4/04/2006	0.4	0.2	401.5	Error escribiendo	Se reprogramara el	Mediano
24	21-Apr	22	El servidor	21/04/2006	0.2	21/04/2006	0.3	3.8	99.0	Se disparan	Se reinicio el	MEDIO. Como
25	25-Apr	23	Falla en el	25/04/2006	0.3	25/04/2006	0.3	0.2	25.5	Problemas de	Se reprogramara el	MEDIO.
26	27-Apr	24	Falló respaldo de	26/04/2006	0.4	26/04/2006	0.4	0.2	2.0	Perdida de conectiv	Se reprogramara el	BAJO
27	27-Apr	25	Falla en el	26/04/2006	0.4	26/04/2006	0.4	0.2	23.9	Problemas con la	Se reprogramara el	MEDIO.
28	28-Apr	26	falla del backup	27/04/2006	0.4	27/04/2006	0.4	0.2	394.6	error en cinta	se reprogramo el	MEDIO
29	15-May	27	Problemas con	13/05/2006	0.9	14/05/2006	0.4	12.0	49.5	Problemas al bajar	Se llamó al	Se Postergó del
30	16-May	28	Falla del respaldo	15/05/2006	0.9	15/05/2006	0.9	0.2	52.0	Error en cinta	Se reprogramara el	MEDIO

Figura 3.2. Una muestra de observaciones correspondiente a los incidentes de fallas

La estructura de los registros observados para los incidentes se muestra en la figura 3.2. En ella se observan los diferentes campos almacenados por los miembros de la Coordinación de Confiabilidad para realizar sus análisis. De este repositorio solo se tomaron en cuenta las variables encerradas en círculos, correspondientes a los datos de interés para el presente estudio. Variables como la novedad presentada, su duración, el tiempo transcurrido desde la falla anterior, la causa, la acción tomada y el impacto del incidente sobre la plataforma fueron consideradas. Este repositorio de incidentes se encuentra clasificado por Grupo Operativo Solucionador (GOS).

Los GOS se encuentran atentos ante cualquier incidente que surja respecto al desempeño de los servicios que brinda la plataforma AIT y están preparados para resolver tales incidentes. Las guardias que ejecutan los GOS tienen una duración de un día y existe un grupo operativo por cada una de las siguientes especialidades:

- ◆ Centro integral de Monitoreo.
- ◆ Centro de Servicios.

- ◆ Servidores Windows.
- ◆ Servidores AS/400.
- ◆ Servidores UNIX.
- ◆ Redes WAN.
- ◆ Redes LAN.
- ◆ Conmutación.
- ◆ Red de Transmisión.
- ◆ Bases de Datos.
- ◆ Respaldo y Almacenamiento.
- ◆ SAP.
- ◆ SAND.
- ◆ STARS-PAWS-RD-RDOJD.
- ◆ SICOPROSA-SIGPLAN -MEGASIM
- ◆ FILIP / SIMAF / SIRET RESET,GADET, SIPREFID, NAF, FUP.
- ◆ ATC.
- ◆ Z-Serie.
- ◆ Soporte Integral.
- ◆ Plataformas de Escritorio.
- ◆ Dispositivos de campo.
- ◆ Control de activos.
- ◆ Infraestructura.
- ◆ Nómina.
- ◆ Seguridad AIT.
- ◆ Video Conferencia.

- ◆ Instalaciones.
- ◆ Soporte a Usuarios Críticos.

Para este estudio, los GOS se organizaron según el número de fallas reportadas durante el periodo de estudio. De acuerdo a criterios de seguridad industrial (Montaña y Rosas (2006)), la selección de los colores identifica el grado de criticidad de la falla y es la guía referencial de la gravedad de la falla tanto por parte del investigador como de los dueños del problema. Esta organización se realizó con base en el siguiente criterio: los GOS con más de 100 fallas registradas en el periodo de estudio se identifican con el color rojo; entre 50 y 100 fallas, color anaranjado; entre 25 y 50 fallas, color amarillo y menos de 25 fallas, color azul. Es importante resaltar que de los 28 grupos mencionados anteriormente sólo se consideraron los 22 que mostraron una cantidad de incidentes significativa. La tabla 3.1 muestra la lista de los grupos tomados con los registros de falla asociados. En dicha figura se aprecia que los 22 grupos operativos solucionadores responden a un total de 1240 acciones reportadas durante el periodo de estudio. La información proporcional se muestra en la tabla 3.2 y en la figura 3.3.

	B	C	D	E	K
4	Cod_grupo	Nombre			Num Reg
5	01	RESPALDO_ALMACENAMIENTO			198
6	02	INFRAESTRUCTURA			137
7	03	REDES LAN			111
8	04	SERVIDORES_WIND			94
9	05	SERVIDORES_UNIX			87
10	06	REDES_TX			75
11	07	SICOPROSA, MEGASIM, SIGPLAN			72
12	08	AMB_TRAB_COLAB			70
13	09	REDES WAN			63
14	10	SAND			60
15	11	SERVIDORES_AS_400			51
16	12	BASE DE DATOS			33
17	13	APLICACIONES_WEB			30
18	14	CONMUTACION			28
19	15	SAP			28
20	16	SERVIDORES_Z_SERIE			26
21	17	RD			23
22	18	PAWS			19
23	19	STARS			12
24	20	NOM, SIRET y NAAF-FIRMA			11
25	21	SIPREFID_FILIP,RESET Y GADET			10
26	22	PLATAFORMA ESCRITORIO			2
27	Total registros				1240

Tabla 3.1. Grupos operativos solucionadores seleccionados

Grupos con mayor frecuencia de falla				
Nivel Crítico de falla	Grupos adscritos		Proporción de fallas (%)	
	Color	Cant.		%
	Rojo	3	13.64%	35.97%
	Anaranjado	8	36.36%	46.13%
	Amarillo	5	22.73%	11.69%
	Azul	6	27.27%	6.21%

Tabla 3.2. Grupos operativos solucionadores seleccionados

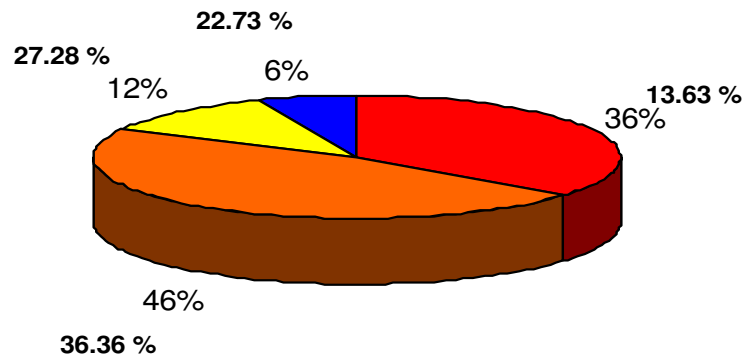


Figura 3.3. Grupos operativos según el número de fallas

3.2 Preprocesamiento de los datos

Se utilizan técnicas de análisis exploratorio de datos y análisis multivariante para seleccionar, limpiar y organizar las variables de estudio. En este estudio el análisis de correspondencia múltiple (ACM), como técnica de exploración multivariante, permitirá medir el grado de asociación existente entre las variables observadas con el fin de construir la variable de salida.

3.2.1 Reorganización de las variables

Tal y como se mostró en la Figura 3.2 la base de datos de incidentes incorpora información de los modos de fallas, causas, acciones, impacto, duración y tiempo entre fallas, que ocurren bajo la dinámica de servicio de la plataforma de PDVSA-AIT. Estas ocurrencias definen el grupo de variables del estudio. La información se maneja en campos abiertos de contenido textual, muchas veces ambiguos y con

datos incompletos, lo que dificulta su procesamiento. Para mitigar este inconveniente se transforman estas variables de contenido abierto en variables categóricas, con el fin de estandarizar la información y la aplicación de técnicas de minería de datos. Se entiende que este tipo de variables son cualitativas ordinales y nominales.

3.2.2 Categorización y creación de las nuevas variables

La reorganización de las variables en forma categórica ayuda a estandarizar la información, permitiendo que todos los datos pertenezcan a grupos exclusivos. Para las variables *falla*, *causa* y *acción*, se diseña un código de seis dígitos; cada dígito representa un nivel jerárquico estructurado de la siguiente manera: los dos primeros dígitos describen el código de la especialidad o grupo operativo y su dominio es {01, 02, 03, ..., 22}, el uso del tercer y cuarto dígito indican el tipo de variable a la que se hace referencia, {NO: novedad, CA: causa, o AC: acción} y los dos últimos corresponden a un número consecutivo que distingue las diferentes categorías creadas, y su dominio es {01, 02, 03, ..., 99}. La estructura del código se muestra en la figura 3.5.

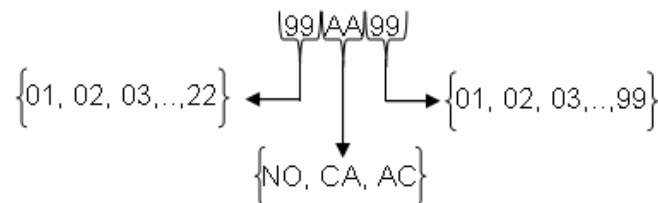


Figura 3.5 Estructura seguida para la codificación de las variables.

Las variables continuas *duración de la falla* y *tiempo entre falla* son reestructuradas en categorías referidas a los intervalos de clase que agrupan aproximadamente la misma cantidad de observaciones cada una. Con base en estas categorías se realiza la división del intervalo de las variables como se muestra en la tabla 3.2 y la tabla 3.3. Vale la pena resaltar que este método aplica para todos los GOS bajo estudio.

Tiempo entre falla		
Clase (horas)	Categoría	Código
0 - 50	Alta ocurrencia de falla	1
51 - 250	Frecuente ocurrencia de falla	2
251 - 500	Sin mayor ocurrencia de falla	3
501 - o mas	Baja ocurrencia de falla	4

Tabla 3.3. Categorización de la variable tiempo entre falla (TEF)

Duración de la falla		
Clase (min.)	Categoría	Código
0 - 10	Solución Inmediata	4
11 - 100	Solución aceptable	3
101 - 500	Solución Lenta	2
501 - o mas	Solución muy lenta	1

Tabla 3.4. Categorización de la variable duración

De igual forma la variable *impacto* que se refiere a aquel que genera el incidente sobre la plataforma, es codificada como se observa en la Tabla 3.4.

Impacto	
Categoría	Código
Alto	1
Medio	2
Bajo	3

Tabla 3.5. Codificación de la variable impacto

Adicionalmente, se definen nuevas variables como atributos de los incidentes, tales como: *tipo de componente*, el cual se refiere al elemento¹ donde ocurrió la falla y toma los valores {1: lógico, 2: Físico}; *código del grupo*, que representa la codificación del grupo operativo al que pertenece el reporte.

Esta nueva organización de los datos reduce la complejidad en el tratamiento de la información; al mismo tiempo que sirve de referencia para establecer un catálogo de modo de fallas en los equipos y sistemas de la plataforma.

3.3 Análisis exploratorio de datos aplicando estadísticas fundamentales

Una vez determinadas las variables del estudio y definidas las categorías de dichas variables, se explora las relaciones existente entre estas categorías con el fin de determinar el grado de ocurrencia de cada una de ellas, así como también, observar la distribución de cada variable dentro del grupo. El análisis es realizado para los 22

¹ Se entiende por elemento, los dispositivos, equipos, aplicaciones o cualquier otro activo que se encuentre bajo la disposición de alguno de los tipos de mantenimiento.

grupos operativos y contribuye a decidir la aplicación de las técnicas de preprocesado y la definición de la variable de salida.

El análisis más simple, en este caso descriptivo, con las variables categorizadas, desde una perspectiva global, permite observar relaciones que no son evidentes entre las variables. Por ejemplo, las acciones ejecutadas para un modo de falla en particular. Para un mayor detalle de las relaciones y el “cruce” entre las variables categóricas, se realizan tablas cruzadas o de contingencia como complemento a los gráficos descriptivos de la distribución de las variables. La naturaleza de las variables y sus relaciones determina una asociación que podría manifestarse a través de conglomerados para establecer un proxy a la variable de salida estimada mediante la aplicación de Análisis de Correspondencia Múltiple. Para la realización de los gráficos estadísticos se utiliza la herramienta WEKA² en su versión 3.5.

3.4 Análisis de correspondencia múltiple

Es una técnica de reducción de variables y en consecuencia, utilizada para la construcción de variables no observadas a partir de las variables observadas. ACM permite medir la asociación lineal presente entre un conjunto de variables; es decir, construye un diagrama cartesiano o mapa perceptual basado en la relación de dependencia (dentro del grupo) e independencia (entre grupos) de un conjunto de variables cualitativas de carácter categórico y, a partir de dicho diagrama, define grupos entre las categorías de las variables observadas que conforman posibles

² Colección de algoritmos automáticos de disposición libre para realizar minería de datos. Está escrito en JAVA y funciona en cualquier plataforma.

variables latentes. Este método, en general, se basa en la optimización de la varianza, aumentándola entre grupos o nuevas categorías y minimizándola dentro de cada grupo o categoría.

Para aplicar la técnica de ACM, se utiliza la herramienta computacional MINITAB versión 14³, mediante la opción: *stat- multivariate - múltiple correspondence analysis*.

Generalmente, ACM al inicio muestra un mapa o gráfico donde los puntos (variables observadas) se encuentran en el centro del eje; esto indica que existe colinealidad entre las variables; es decir, existen variables que están fuertemente interrelacionadas, y, por tanto, resulta difícil medir sus efectos individuales sobre la variable respuesta (variable de interés). Las variables redundantes pueden ser identificadas a través de la matriz de correlación y de este modo, se podría mitigar este fenómeno, siempre y cuando las variables sean cuantitativas. En este caso, donde las variables son categóricas, se aplica ACM para el mismo bloque de datos. De manera particular se reorganizan nuevamente algunas variables que generaban ruido tales como impacto, tiempo entre fallas y duración. Esta reorganización de las variables hace posible observar la distribución en grupos o conglomerados que, a pesar de que presenta una cantidad reducida de puntos aberrantes o atípicos, describen relaciones útiles en el contexto de la investigación para la definición de la variable de salida.

³ MINITAB es una licencia para ejecutar funciones estadísticas básicas y avanzadas. Desarrollado por la Pennsylvania State University.

3.5 Aplicación de máquinas de vectores de soporte

En este estudio se aplica esta técnica con la finalidad de obtener un mayor conocimiento de la variable obtenida en el análisis exploratorio de datos. En tal sentido, el objetivo de la utilización de esta herramienta es obtener un modelo que permita clasificar la variable de estudio (tipo de mantenimiento correctivo) a partir de las variables observadas y generar conocimiento válido para clasificar futuros incidentes que no hayan sido tomados en los datos originales de entrenamiento. De este modo, cuando se presente un incidente de falla, como instancia, es posible obtener pronosticar la clasificación del tipo de mantenimiento correctivo ejecutado y poder utilizar esta información como apoyo para la toma de decisiones por parte de los actores involucrados.

3.5.1 Selección de los conjuntos de entrenamiento y validación

La utilización de máquinas de vectores de soporte para clasificar instancias requiere la aplicación de alguna técnica de validación que haga posible verificar que el modelo es bueno; es decir, genera clasificaciones válidas con un error aceptable. Para el presente proyecto se utiliza la validación cruzada, la cual consiste en la división de los datos de estudio en dos subconjuntos, uno para el entrenamiento y el otro para la validación del modelo. El subconjunto de datos de entrenamiento representa aquellos registros que son utilizados para ajustar los parámetros de la máquina, de modo que sea posible obtener una clasificación válida de los datos del estudio. Posteriormente, se utiliza un subconjunto de estos datos para verificar que el modelo

garantiza un buen porcentaje de clasificaciones correctas en instancias nunca conocidas en el entrenamiento. Debido a que el ajuste de los parámetros del modelo es un proceso que puede requerir una gran cantidad de instancias, se selecciona, como usualmente se hace, un porcentaje mayor para el entrenamiento que para la validación, específicamente un 70% de los datos se usan para entrenar el modelo y un 30% para validarlo.

Para evitar sobreajuste en el modelo y mitigar el fenómeno que significa la presencia de unas instancias más que de otras, se utilizan técnicas de muestreo aleatorio estratificado que garantice que las clases se mantenga proporcionalmente en la muestra. En consecuencia, la selección de los grupos de entrenamiento y validación se hace utilizando muestreo aleatorio sin reemplazo proporcionalmente a cada una de las categorías de la variable de salida.

3.5.2 Ajuste de los parámetros de la máquina de soporte vectorial

Durante la fase de entrenamiento se busca el modelo que permita clasificar de manera eficiente futuras instancias. Para ello se hacen cambios en los parámetros de la máquina, de modo que mejore el porcentaje de clasificaciones correctas. Para la selección de los parámetros se hacen variaciones a los coeficientes de la función *kernel* para un mismo conjunto de entrada y se realizan los ajustes a partir del error

aceptable para determinar el mejor modelo. En el entrenamiento de la máquina se utiliza la librería LIBSVM⁴ y un conjunto de funciones por ella manejadas.

3.5.3 Validación de los resultados

Una vez obtenido el modelo que mejor se ajusta al conjunto de las instancias que permitieron el entrenamiento, se hace la verificación de que el modelo clasifica correctamente nuevas instancias. Para ello se toma el subconjunto de datos de validación y, utilizando el modelo obtenido, se realiza la verificación del número de instancias clasificadas correctamente, utilizando para ello la matriz de confusión. La matriz es útil para las técnicas de aprendizaje supervisado ya que permite la visualización de los resultados del entrenamiento de la máquina. En tal sentido, es posible determinar el número de predicciones para cada clase real en cada una de las clases esperadas, de este modo se verifica si el modelo está “confundiendo” las clases. Por ejemplo, en el caso de una clasificación de 3 grupos, supóngase que la máquina genera para cada instancia una etiqueta correspondiente a alguna de las clases, se sabe de antemano a que clase pertenece cada instancia, por lo que la matriz permite verificar si la etiqueta asignada corresponde a la verdadera etiqueta de las instancias. En la tabla 3.5 se puede observar, en el sentido de las filas, que de ocho instancias que en realidad pertenecen a la clase A el sistema predijo que tres eran de tipo B y de seis instancias de tipo B predijo que uno era de tipo C y dos eran de tipo A. A partir de la matriz se puede ver que el sistema tiene problemas distinguiendo entre instancias de la clase A (60% de confusión) y de la clase B (50% de confusión),

⁴ LIBSVM es una aplicación libre desarrollada bajo la plataforma LINUX y de descarga gratuita, creada por Chih Chung Chang y Chih Jen Ling y que incorpora funciones para la clasificación usando SVM.

pero que puede distinguir razonablemente bien entre instancias de la clase C (15% de confusión) y otras clases.

	A	B	C
A	5	3	0
B	2	3	1
C	0	2	11

Tabla 3.6. Ejemplo de matriz de confusión.

Capítulo 4

Análisis de los resultados e interpretación

El presente capítulo expone los resultados más resaltantes que se obtienen de la aplicación de los pasos descritos en el capítulo anterior. Estos resultados se presentan y se comenta su significado dentro del contexto del estudio realizado. Se comienza exponiendo los resultados de los análisis estadísticos y se discute acerca de la distribución de las variables; luego se habla sobre la naturaleza de la variable de salida y el proceso para su obtención, culminando con la presentación de los resultados de la aplicación de máquinas de vectores de soporte.

4.1 Preprocesamiento y limpieza de los datos

Como ya se mencionó, el análisis exploratorio y la definición de las categorías para las variables de entrada busca construir un repositorio “limpio” a partir de los datos originales. Es por ello que se revisa detalladamente cada uno de los registros correspondientes a cada incidente reportado por los analistas de guardia dentro de cada GOS, elaborando de ese modo categorías que permitan agrupar los campos de contenido abierto y en formato textual (abierto) de contenido similar. Esta

categorización es validada con miembros de cada uno de los Grupos Operativos Solucionadores (GOS), obteniéndose un catálogo de modos de falla que reúne las fallas, causas, acciones, impactos, tiempos, componentes electrónicos o aplicaciones, cuya utilidad para los actores dentro de la unidad donde se realiza el estudio es un valor agregado a los resultados obtenidos con este proyecto.

El procedimiento de construcción de esta “vista minable” a partir de la cual es posible aplicar directamente las técnicas de minería de datos y análisis estadísticos, comprende la limpieza y depuración de los datos, buscando eliminar inconsistencias e imputar los datos faltantes. La poca cantidad de datos disponible para el estudio hace que resulte muy costoso la pérdida de algún registro por la ausencia de un campo en particular, por lo que, para tratar con los datos faltantes se consulta directamente la fuente del registro, es decir, al analista que reporta el incidente y de este modo, se recupera. Este procedimiento se hace conjuntamente con la fase de validación de las categorías. En definitiva, se obtiene un mínimo de datos faltantes para el estudio.

Durante esta fase también se determina que de los 22 grupos considerados inicialmente para el estudio, 12 presentan una cantidad muy poco representativa de información, lo que afectaría la calidad de los estudios realizados. Es por ello que se limita el estudio a un total de 10 grupos operativos. Sin embargo, se muestra información detallada correspondiente a un solo grupo operativo que es representativo de los otros nueve. Este es el grupo de respaldo y almacenamiento de la plataforma tecnológica de PDVSA AIT SCC. El grupo se encarga de brindar soporte y administrar los equipos y sistemas que brindan apoyo al respaldo de las bases de datos y aplicaciones principales de la plataforma antes mencionada.

4.2 Aplicación de estadísticas fundamentales

La fase de exploración de los datos categorizados genera un conjunto de gráficos que corresponden a histogramas de frecuencia para cada una de las categorías de las variables de estudio tal como se mencionó en los puntos 3.2.1 y 3.2.2 del capítulo anterior.

Para el caso del grupo de referencia en la figura 4.1 se muestra la distribución de las diferentes categorías obtenidas de la variable falla. De acuerdo a la caracterización de la falla se definieron cinco categorías. De acuerdo a estas categorías se observa que un alto porcentaje de las fallas de este grupo operativo se deben a caídas en los dispositivos de respaldo (86,87%), seguidas por fallas en los robots encargados de manipular los medios magnéticos (5,05%), por fallas de aplicaciones usadas en los sistemas de respaldo (3,53%), fallas de la caja de almacenamiento (2,52%) y finalmente, fallas de red (2,02 %).

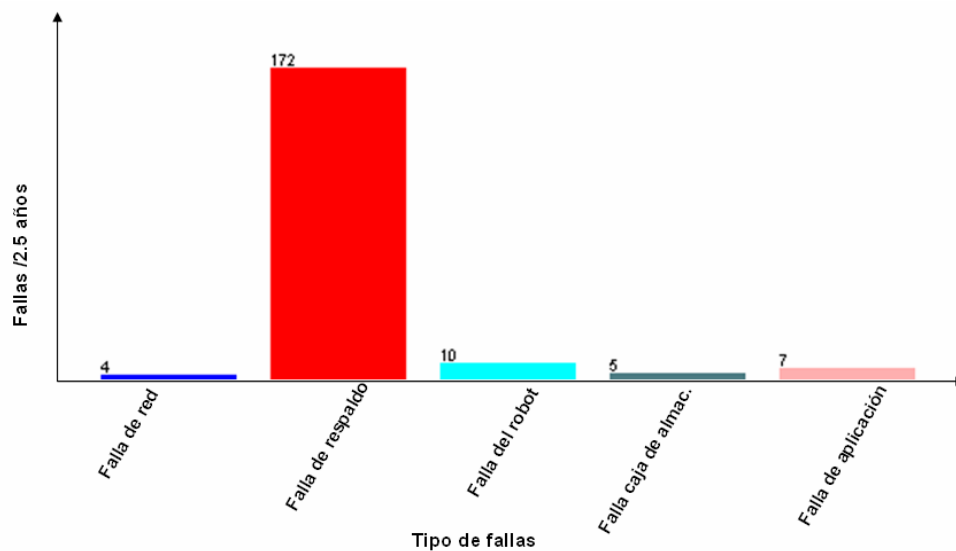


Figura 4.1. Frecuencia de fallas

La figura 4.2 muestra la distribución de las causas que originan las fallas. En ella se puede apreciar que la mayoría de las fallas se ocasionan por problemas con los dispositivos físicos de almacenamiento (61,62%) y que las causas que tienen menor frecuencia son las fallas de *swiches*, problemas con el servicio de *data protector* y problemas de espacio, las cuales representan el 3.5 % de las causas totales registradas.

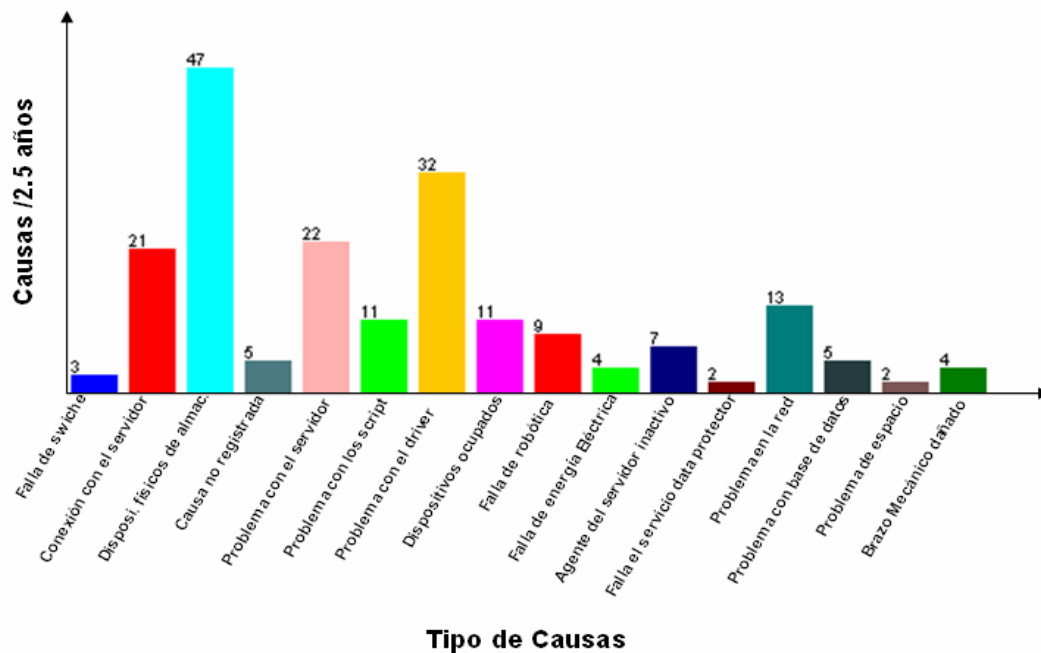


Figura 4.2. Frecuencia de Causas

La figura 4.3 muestra la frecuencia de las acciones tomadas para finalizar las fallas. En ella se aprecia que el 72,73% de las acciones o medidas aplicadas por el personal de soporte de este GOS son las siguientes: reprogramar los respaldos/backup, reportar la falla al grupo responsable o deciden no registrar las

actividades correctivas aplicadas a cierto modo de falla. También se observa que las acciones de chequear el sistema de energía eléctrica, cambiar los *drivers*, modificar el status de los discos, reparar los puertos dañados, recuperar los datos, cambiar los discos, reconectar el cableado de red, chequear tarjetas de red, ampliar los *filesystems* y migrar los servidores conectados a la red, son practicas poco aplicadas por este grupo soporte y sólo representan un 7.57% del total de acciones registrada. Es importante señalar que las acciones mencionadas anteriormente representan el 58.82% de las soluciones llevadas a cabo por este grupo operativo

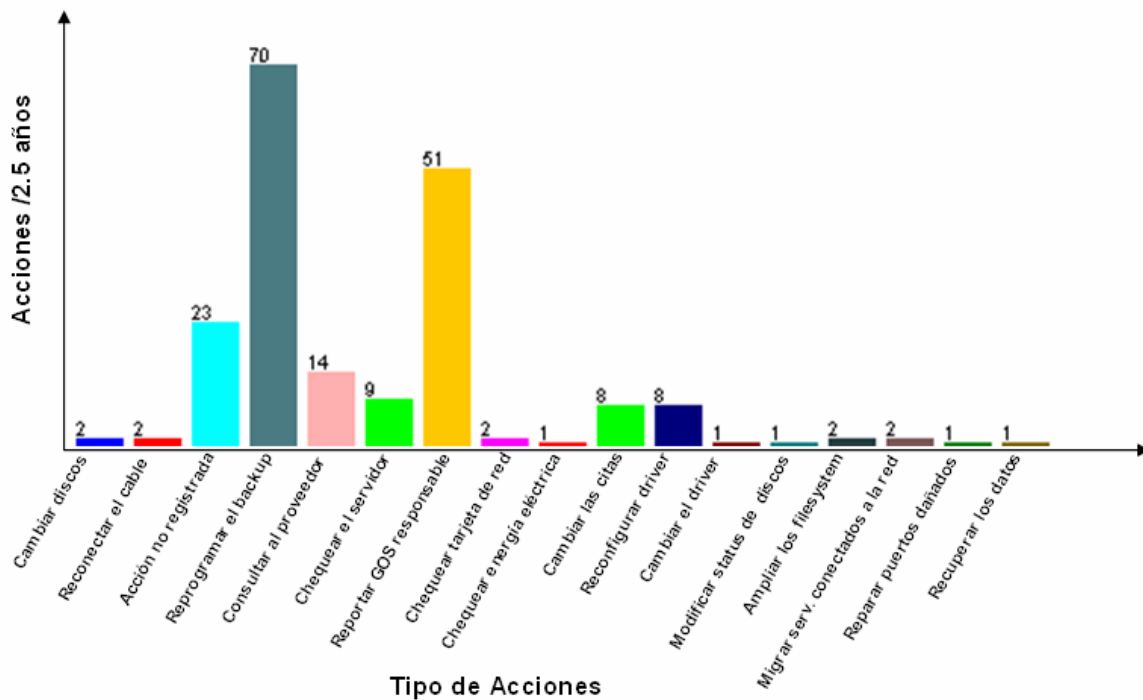


Figura 4.3. Frecuencia de Acciones

En la figura 4.4 se puede observar la tendencia marcada de fallas en los componentes físicos (82%), esto se debe a la naturaleza física de los componentes

manejados por el grupo respaldo y almacenamiento, contra una tendencia mínima de fallas en los componentes lógicos (18%).

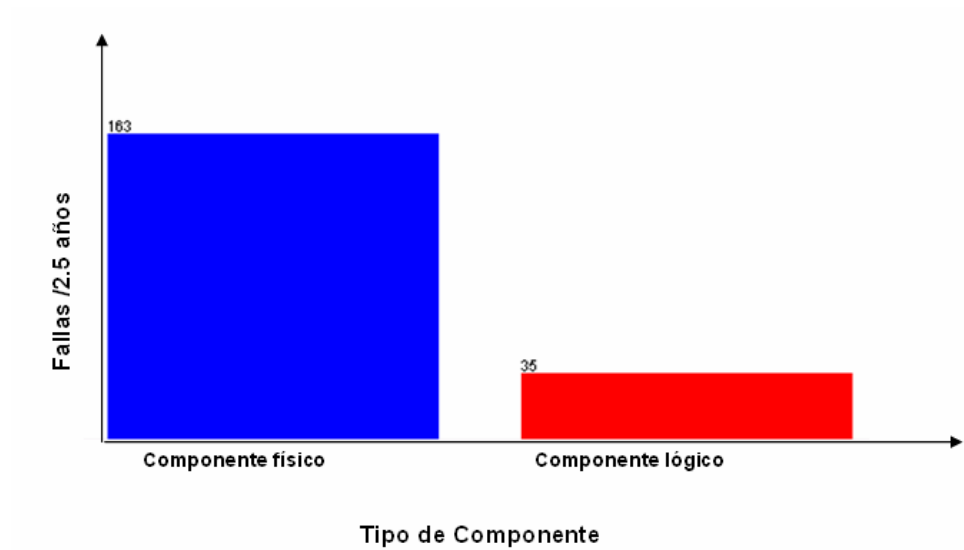


Figura 4.4. Frecuencia de los tipos de componentes

En la figura 4.5 se puede apreciar claramente la tendencia de los impactos medio y bajos (81.81% del total de impactos registrados) a diferencia de las categorías impacto alto e impacto no registrado que representan solo el 18.19%.

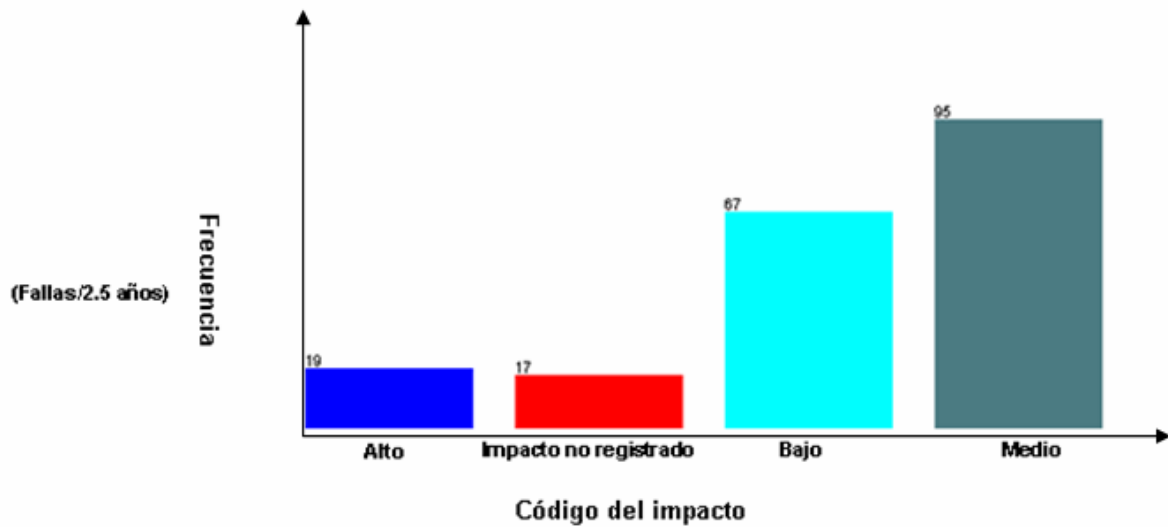


Figura 4.5. Frecuencia de los tipos de impactos

4.3 Tablas de contingencia

Para verificar como se relacionan las categorías de las variables consideradas para el estudio, se realizan tablas cruzadas o de contingencia en las que cada celda representa el número de instancias que incluyen valores de dos categorías (representadas por fila y columna). Los datos de la matriz cruzada entre las variables causa y falla, por ejemplo, representan las causas que originaron cada una de las categorías de falla obtenidas en el análisis. La tabla 4.1 muestra los datos de la matriz cruzada y se complementan con la representación gráfica de la figura 4.6.

Causa		Fallas					Total
		Falla de red	Falla de robot	Falla de respaldo	Falla en la caja de almacenamiento	Falla de aplicación	
		01NO01	01NO02	01NO03	01NO04	01NO05	
Falla del swiche	01CA05	3	0	0	0	0	3
Falla de conexión con el servidor	01CA04	1	2	18	0	0	21
Problema con los dispositivos físicos de almacenamiento	01CA15	0	2	43	0	2	47
Causa no registrada	01CA08	0	0	5	0	0	5
Problema con el servidor	01CA13	0	2	20	0	0	22
Problema con los script	01CA09	0	1	9	0	1	11
Problema con el driver	01CA11	0	0	31	1	0	32
Dispositivos ocupados	01CA02	0	0	11	0	0	11
Falla de robótica	01CA07	0	0	6	3	0	9
Falla de energía eléctrica	01CA06	0	0		2	2	4
Agente del servidor inactivo	01CA03	0	0	7	0	0	7
Problema con el servicio de data protector	01CA12	0	0	2	0	0	2
Problema en la red	01CA16	0	0	13	0	0	13
Problema con la base de datos	01CA14	0	0	5	0	0	5
Problema de espacio	01CA10	0	0	2	0	0	2
Brazo mecánico dañado	01CA01	0	0	0	4	0	4
Total		4	7	172	10	5	198

Tabla 4.1 Matriz Cruzada de Causa vs. Falla

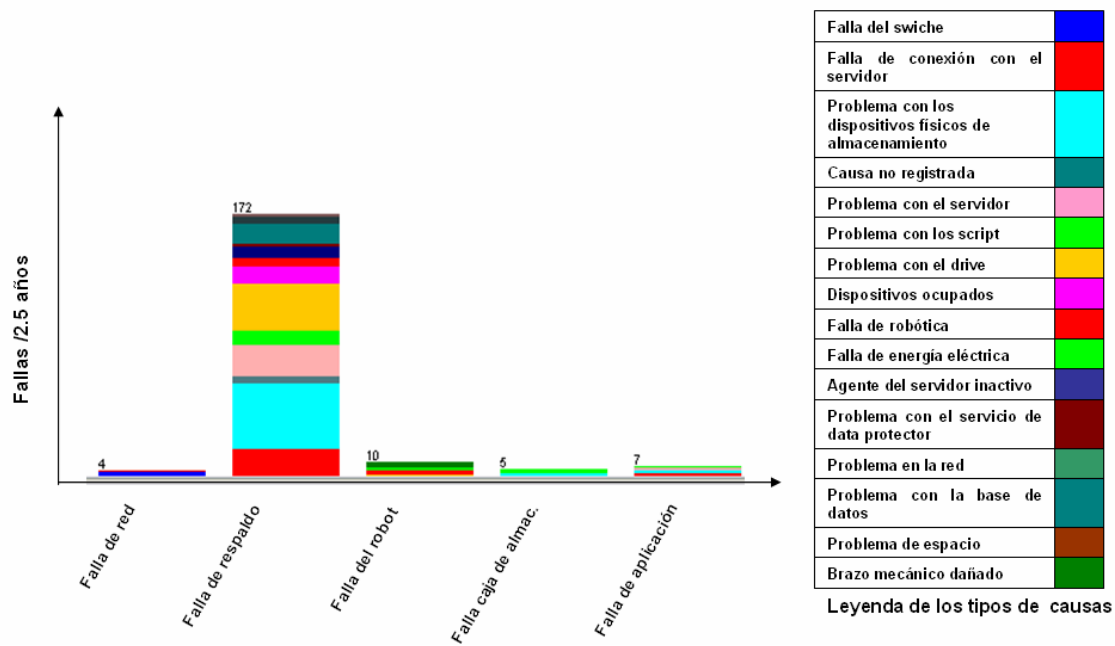


Figura 4.6. Representación gráfica de la matriz cruzada de Causas vs. Fallas

A partir de la tabla 4.1 y la figura 4.6 se pueden hacer observaciones acerca de la naturaleza de las fallas de la plataforma, como por ejemplo, que el tipo de falla que se presenta con mayor frecuencia corresponde a las fallas de respaldo y que la mayoría de estas fallas son originadas por problemas con los dispositivos físicos de almacenamiento y a problemas con el *driver* del dispositivo; también se puede señalar que cuando se daña un brazo mecánico, se generan exclusivamente fallas en la caja de almacenamiento. Una de las causas más frecuentes de problemas en este grupo corresponde a problemas con dispositivos de almacenamiento y a su vez, esta falla es la que se encuentra con mayor incidencia en varios de los tipos de falla.

Se realizaron matrices cruzadas para relacionar todas las variables del estudio con el tipo de falla. Estos resultados pueden apreciarse en el Anexo A.

4.4 Análisis de correspondencia múltiple

Con un mayor conocimiento de las relaciones existentes entre las variables del estudio se comienza a realizar análisis más detallados para explorar las asociaciones entre dichas variables y determinar grupos o conglomerados que permitan obtener una variable de salida claramente definida. A pesar de la existencia de puntos atípicos, se logró obtener grupos consistentes y con significados interpretables en el contexto de la investigación. Para poder realizar Análisis de Correspondencia Múltiple con el software MINITAB, se recodifican las variables como se indica en las tablas 4.2, 4.3 y 4.4.

Nombre de la Causa	Categoría	Código
Brazo mecánico dañado	01CA01	1
Dispositivos ocupados	01CA02	2
Agente del servidor quedo inactivo	01CA03	3
Problema con el servidor	01CA04	4
Problema de conexión con la red	01CA05	5
Problema con el driver	01CA06	6
Causa desconocida	01CA07	7
Problema con los script	01CA08	8
Problema de espacio	01CA09	9
Problema con el servicio de data protector	01CA10	10
Problema con la base de datos	01CA11	11
Problema con los dispositivos físicos de almacenamiento	01CA12	12

Tabla 4.2. Causa de la Falla

Nombre de la Falla	Categoría	Código
Falla de comunicación	01NO01	1
Falla de aplicación	01NO02	2
Falla de respaldo	01NO03	3
Falla del robot	01NO04	4
Falla en la caja de almacenamiento	01NO05	5

Tabla 4.3 Tipo de Falla

Nombre de la Acción	Categoría	Código
Cambiar las cintas	01AC01	1
Cambiar los discos	01AC02	2
Chequear el servidor	01AC03	3
Acción no registrada	01AC04	4
Configurar o cambiar el driver	01AC05	5
Reportar la falla al grupo responsable	01AC06	6
Reprogramar el backup	01AC07	7
Ampliar los filesystem	01AC08	8
Contactar al proveedor para que resuelva la falla	01AC09	9

Tabla 4.4. Acción ejecutada para finalizar la falla

Luego de la recodificación se realiza el análisis de correspondencia múltiple, obteniéndose el resultado que se señala en la figura 4.7. En ella se observa la formación de conglomerados con base en su distribución a lo largo de los ejes. Los grupos fueron revisados e interpretados de acuerdo a su significado práctico. El gráfico muestra las categorías asignadas a cada conglomerado mediante un número de secuencia que se corresponde con una categoría asignada a las variables originales como se muestra en la tabla 4.5. En esta tabla también se observa que los elementos pertenecientes a los conglomerados observados mediante el ACM son resaltados en colores para efectos de identificarlos fácilmente.

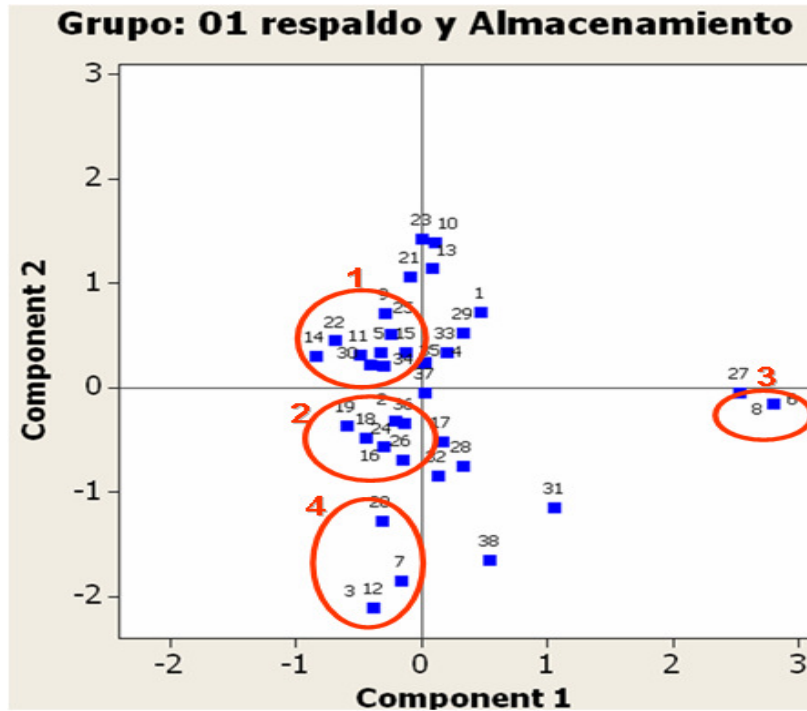


Figura 4.7 Análisis de Correspondencia Múltiple Grupo Respaldo y Almacenamiento

En este sentido, las categorías que definen el conglomerado identificado con 1 en la tabla 4.5 aparecen resaltadas en color azul, las del 2 en color amarillo, las del 3 en color verde, mientras que las del 4 aparecen resaltadas en color naranja.

Con base en la distribución de los puntos, se observa que el eje vertical puede asociarse a las frecuencias de ocurrencia de las fallas y el tipo de componente que las presenta. De este modo, el eje vertical positivo corresponde a las fallas en los elementos lógicos (tales como aplicaciones) que son más frecuentes que en los elementos físicos y que por lo general tienen un mayor impacto. De manera análoga, el eje vertical en la parte negativa corresponde a las fallas en los elementos físicos que se presentan con menor frecuencia y menor impacto.

Para interpretar el eje horizontal se observa que al desplazarse en sentido positivo aumenta la naturaleza lógica de las causas de las fallas y en la medida que se disminuyen los valores, desplazándose en sentido negativo, las causas de las fallas tienen mayor naturaleza física.

Se puede dar la siguiente interpretación a los grupos obtenidos:

(1): **Mantenimiento correctivo continuo** a los elementos lógicos usados en las actividades de respaldo y almacenamiento con soluciones en software, presentando alta ocurrencia de falla, ocasionando impacto medio al servicio y con restauración lenta.

(2): **Mantenimiento correctivo frecuente** a dispositivos físicos de almacenamiento (*driver*, cintas, medios magnéticos) con soluciones a través de reemplazos o configuraciones, presentando de alta a frecuente ocurrencia de falla, ocasionando impacto medio al servicio y de inmediata restauración.

(3): **Mantenimiento correctivo intermitente** a los elementos de robótica (Brazo mecánico, *driver* de robótica), con restauración aceptable ocasionando impacto medio al servicio y presentando frecuente ocurrencia de fallas.

(4): **Mantenimiento correctivo escaso** aplicado a los equipos de comunicación (equipos de redes) por parte del grupo soporte responsable de los servicios de redes, con variaciones en fallas de equipos de almacenamiento con restauración inmediata, bajo impacto al servicio, y de baja ocurrencia de fallas.

(5): Se refiere a los *casos atípicos* que no describen en forma clara un patrón o regla de correspondencia, la estructura de esta variable será estudiada posteriormente, cuando se obtengan nuevos registros de fallas, los cuales probablemente brinden algún tipo de información adicional para tal fin.

Grafico	Nombre de la variable	Código	Variable
1	componente lógico	1	Componente
2	componente físico	2	
3	Falla de comunicación	1	Falla
4	Falla de aplicación	2	
5	Falla de respaldo	3	
6	Falla del robot	4	
7	Falla en la caja de almacenamiento	5	
8	Brazo mecánico dañado	1	Causas
9	Dispositivos ocupados	2	
10	Agente del servidor quedo inactivo	3	
11	Problema con el servidor	4	
12	Problema de conexión con la red	5	
13	Problema con el driver	6	
14	Causa desconocida	7	
15	Problema con los script	8	
16	Problema de espacio	9	
17	Problema con el servicio de data protector	10	
18	Problema con la base de datos	11	
19	Problema con los dispositivos físicos de almacenamiento	12	
20	Cambiar las cintas	1	
21	Cambiar los discos	2	
22	Chequear el servidor	3	
23	Acción no registrada	4	
24	Configurar o cambiar el driver	5	
25	Reportar la falla al grupo responsable	6	
26	Reprogramar el backup	7	
27	Ampliar los filesystem	8	Impacto
29	Alto	1	
30	Medio	2	
31	Bajo	3	Duración
32	Solución Inmediata	1	
33	Solución aceptable	2	
34	Solución Lenta	3	
35	Solución muy lenta	4	TEF
36	Alta ocurrencia de falla	1	
37	Frecuente ocurrencia de falla	2	
38	Sin mayor ocurrencia de falla	3	
39	baja ocurrencia de falla	4	

Tabla 4.5. Codificación de las variables del grupo Respaldo y Almacenamiento

Los conglomerados obtenidos pueden ser expresados mediante reglas de asociación que se pueden aplicar sobre una instancia, generando una clasificación del tipo de mantenimiento correctivo realizado durante el período de estudio. Esta clasificación es la variable de salida que será utilizada luego para el entrenamiento de la máquina de soporte vectorial. Para este caso en particular, se definieron cinco clases, no obstante, la cantidad de clases varia dependiendo de la naturaleza de los sistemas y equipos manejados por cada GOS, generándose más o menos clases en cada caso.

Este procedimiento se repite nueve veces con la finalidad de obtener la variable de salida de cada GOS. En cada caso particular se estudia la conformación de los conglomerados, se definen sus ejes y se establecen las bandas de colores. Del mismo modo se fija como estrategia valorar la variable de salida en todos los GOS tal que sus atributos sean lo más similares posibles para ayudar a una interpretación estándar. En el Anexo A se tiene en detalle todos estos resultados.

4.5 Aplicación de máquinas de soporte vectorial

La aplicación de las máquinas de soporte vectorial busca clasificar en las acciones de mantenimiento correctivo la entrada los datos de un incidente de falla. De este modo, se pretende obtener un modelo matemático que clasifique correctamente la mayor cantidad de instancias posibles. Para el entrenamiento del modelo, en el caso del grupo respaldo y almacenamiento se realiza un muestreo estratificado de acuerdo a las 5 categorías de la variable de salida, tomando proporcionalmente muestras aleatorias dentro de cada categoría.

Se realizan varios entrenamientos buscando encontrar los mejores parámetros para la máquina, es decir, aquellos que maximicen la cantidad de instancias clasificadas correctamente.

La aplicación de Máquinas de Soporte Vectorial asoció un *kernel* basado en una función de base radial (*radial basis*). Luego de realizar varias pruebas variando los parámetros de la función y de la máquina se obtienen los resultados que aparecen en la tabla 4.6.

Coeficiente C	Coeficiente g	Clasificaciones Correctas (de 60 instancias)	Error (%)
1	0.01	35	0.4355
1	0.1	43	0.3065
1	1.0	49	0.2097
1	10	48	0.2259
1	100	48	0.2259
10	0.01	43	0.3065
10	0.1	59	0.048
10	1.0	51	0.1775
10	10	48	0.2259
10	100	48	0.2259
100	0.01	54	0.1291
100	0.1	57	0.0807
100	1.0	51	0.1775
100	10	48	0.2259
100	100	48	0.2259

Tabla 4.6. Configuración de los parámetros de la máquina.

Una vez definido el modelo de la máquina se procede a realizar la validación con el conjunto de prueba. La matriz de confusión obtenida se muestra en la tabla 4.7. En ella se puede apreciar que el modelo clasifica bastante bien para los cinco tipos de mantenimiento definidos como variable de salida, presentándose una sola instancia clasificada incorrectamente, que corresponde a una instancia de la clase 5 que fue clasificada con la etiqueta 2. A pesar de que la matriz pudiera sugerir que el modelo no clasifica correctamente a las instancias de la clase 5, hay que recordar que el significado de esta categoría está asociado a los casos atípicos que no se presentan con mucha frecuencia y que requieren un mayor estudio para poder establecerse conclusiones acerca del mantenimiento.

	1	2	3	4	5
1	33	0	0	0	0
2	0	15	0	0	0
3	0	0	6	0	0
4	0	0	0	5	0
5	0	1	0	0	1

Tabla 4.7 Matriz de confusión para el modelo final.

Se observa como el producto obtenido, es decir, la clasificación de los tipos de mantenimiento correctivo a partir de los datos correspondientes al incidente de falla, es resultado de un examen exhaustivo de los datos y está sujeta a un conocimiento profundo de la naturaleza de cada GOS, sus equipos y sistemas, pues para la construcción de los modelos aquí discutidos fue necesario explorar minuciosamente

la base de datos y reunir información mucha veces faltante, ambigua e imprecisa. La construcción de la vista minable a partir de la cual realizar los análisis, fue la fase que requirió mayor esfuerzo en el estudio y fue la que generó mayor valor agregado a los actores involucrados, pues el producto puede ser utilizado para varios fines.

De la interpretación de la variable obtenida y de los resultados encontrados luego de aplicar la metodología KDD se pueden establecer reglas para determinar el tipo de mantenimiento que corresponde a un incidente y esta información resulta de utilidad para la planificación del mantenimiento y a la detección de vulnerabilidades para generar recomendaciones sobre la confiabilidad de este grupo. Adicionalmente brinda un soporte de gran valor para la aplicación de metodologías propias del análisis de confiabilidad, como es el caso del análisis causa raíz que busca detectar las causas más comunes de las fallas y generar acciones para afectar dichas causas. Entre las reglas obtenidas para el caso particular del GOS en estudio, se puede mencionar:

- El tipo de falla más común se genera en los respaldos programados periódicamente. Estas fallas responden a caídas en el servidor y problemas relacionados con los *scripts* encargados de ejecutar tales respaldos y que por lo general se solucionan reiniciando los servidores asociados o detectando el grupo responsable del mantenimiento del dispositivo en el que se produce la falla, estas fallas tienen un impacto medio porque la determinación de la causa requiere de un tiempo considerable para finalizar el incidente, pues en la mayoría de los casos reiniciar el servidor es la última alternativa. Es importante resaltar que en muchas ocasiones los problemas no corresponden al grupo de respaldo y almacenamiento como tal y esta conclusión es

- obtenida luego que los miembros del grupo dedican una cantidad de esfuerzo considerable para determinar la causa de la falla y la acción a ejecutar.
- Las fallas en los componentes físicos se encuentran frecuentemente asociadas a problemas en el espacio en disco duro y a en los dispositivos de almacenamiento, las acciones llevadas a cabo para solucionarlas pueden variar desde la reprogramación de los *backups* hasta la reconfiguración de los *drivers* de los dispositivos asociados. El tipo de acción es muy variada, por lo que el impacto asociado no es fácilmente cuantificable. Sin embargo estas fallas ocurren en intervalos de tiempo muy cortos pero se restauran de manera inmediata.
 - Las fallas de comunicación y los errores ocurridos en las cajas de almacenamiento son solucionados por lo general realizando cambios en las cintas, y muchas veces se deben a errores de comunicación en la red. Estas fallas no tienen tanta frecuencia.
 - Las fallas ocurridas en los robots que realizan los respaldos, son exclusivamente causadas por fallas en los brazos mecánicos del robot, estas fallas tienen un impacto desconocido y registran una baja ocurrencia. En este tipo de fallas, la solución está determinada por la contratación de empresas externas encargadas de realizar mantenimientos a los robots.

Los resultados obtenidos en la presente sección son replicados en los nueve GOS estudiados y se realizan conclusiones adaptadas al contexto de cada grupo. En este sentido, los resultados obtenidos en la determinación de las variables de salida para los grupos faltantes se muestran en el Anexo B del presente documento.

Capítulo 5

Conclusiones y Recomendaciones

5.1 Conclusiones

La aplicación de la metodología KDD a la base de datos de incidentes de falla reportados en la plataforma AIT SCC de PDVSA permitió obtener conocimiento valioso para los actores involucrados en el problema de mantenimiento y logró satisfacer las expectativas relacionadas con el proyecto. Por ello, es posible afirmar de manera contundente que los objetivos planteados al inicio de este proyecto fueron cumplidos satisfactoriamente. En lo que sigue se puede afirmar tal aseveración al resaltar las bondades y deficiencias de este estudio

El preprocesamiento de los datos hace posible organizar la información original con forma y coherencia, permitiendo explorar los datos y obtener una vista general en la que resalten las relaciones más evidentes entre ellos. Mejorar la disposición y limpieza de los datos hace posible aplicar eficientemente las técnicas propias de la minería de datos para obtener conocimiento oculto y latente de las variables de estudio. Sin embargo, la limpieza de los datos es una de las tareas más laboriosas dentro de la metodología, requiriendo la mayor dedicación de esfuerzo y

tiempo dentro del proyecto. Más aún cuando la fuente de información tuvo que ser manipulada para enmarcarla en un contexto de variables con atributos.

El Análisis de Correspondencia Múltiple fue la herramienta clave para conocer la relación implícita que existe entre las variables observadas y dio origen a una desconocida variable respuesta, la cual permite conocer el tipo de mantenimiento correctivo realizado a los modos de fallas. Esta nueva variable clasifica básicamente cuatro tipos de mantenimiento, que emergen a partir de las ocurrencias de fallas presentadas en cada uno de los servicios brindados por los grupos de soporte de la plataforma durante el periodo de estudio. Sin embargo, tan importante como la determinación de las variables latentes, es que dichas variables sean de interés y de utilidad para los actores involucrados en el proceso. Los métodos de análisis de datos por lo general se quedan en la determinación de las relaciones, pero la interpretabilidad de dichas relaciones y su significado en el contexto del problema es el aporte más valioso que puede realizar el analista.

Dado que el mantenimiento de la plataforma AIT es una de las actividades fundamentales requeridas para garantizar el adecuado funcionamiento y continuidad operacional de las soluciones tecnológicas que apoyan la organización, resulta vital conocer el comportamiento de la variable respuesta, tanto para lograr una adecuada planificación de las tareas de mantenimiento, como para realizar estudios a profundidad de las causas de los problemas en los GOS y permitir a los analistas de la Coordinación de Confiabilidad ser más precisos en sus análisis y recomendaciones, levantando banderas de alarma a tiempo para evitar caídas en los servicios críticos y de este modo, mejorar la disponibilidad de la plataforma.

Al clasificar el tipo de mantenimiento correctivo realizado a los activos que se encuentran bajo custodia de PDVSA-AIT SCC, es posible crear planes especiales de mantenimiento preventivo que actúen directamente sobre equipos y sistemas altamente vulnerables, y en consecuencia, aumenten la disponibilidad del servicio prestado por estos. Este proyecto servirá de referencia para trazar líneas estratégicas dirigidas a lograr el acercamiento hacia el objetivo fundamental: garantizar la operatividad continua y mantenibilidad de los equipos y sistemas que dan apoyo a los procesos de negocio de PDVSA.

Se pudo observar la eficiencia de las Máquinas de Vectores de Soporte como método de clasificación multiclase, así como también su alta capacidad de generalización, ya que con pocas muestras de entrenamiento se logró identificar exitosamente los tipos de mantenimiento correctivo aplicado en cada GOS de acuerdo con su especialidad. En este sentido, vale la pena resaltar que la validez de los resultados obtenidos en la presente investigación tiene vigencia en la medida que no hayan cambios significativos en la plataforma, de modo que puedan surgir nuevos tipos de falla que no hayan sido contemplados en el presente estudio.

Una de las mayores dificultades presentadas durante el desarrollo del proyecto fue tratar con una base de datos muchas veces inconsistente, incompleta, ambigua y en algunos casos con contenidos poco relevantes. La calidad de la fuente de datos determina, en la mayoría de los casos, la eficacia y validez de los análisis realizados a partir de tales datos. Por tal razón, los resultados obtenidos en este proyecto pueden mejorar a medida que el mecanismo implementado para la captura de los datos sea más robusto, disminuyendo así las desviaciones y, en consecuencia, apoyando con mayor precisión al proceso de toma de decisiones. Esto es posible si

se mejoran significativamente las normas y procesos asociados a la obtención de los reportes por parte de los analistas.

5.2 Recomendaciones

- ✓ Es posible, a partir de los resultados obtenidos en esta investigación, definir estrategias para optimizar la gestión del mantenimiento de la plataforma. Técnicas basadas en análisis estadísticos (control de calidad, diseño de experimentos, entre otras) podrían ayudar a medir el desempeño dentro de los grupos soporte de la plataforma. Por ejemplo, mediante el control estadístico de calidad a las variables estudiadas se podrían detectar anomalías en la duración de las fallas para un grupo particular y tomar medidas al respecto.
- ✓ Determinar mecanismos para que los analistas sólo reporten las fallas correspondientes a su especialidad, con el fin de garantizar la integridad de los datos, y en consecuencia, mejorar la veracidad de la información resultante cuando se presenten los informes de desempeño o estadísticas por grupo operativo.
- ✓ Crear instrumentos en los que se norme claramente el impacto que tiene sobre el negocio la interrupción de cada uno de los servicios de la plataforma AIT, evitando así que este registro lo realice intuitivamente o empíricamente el analista soporte.
- ✓ Difundir la importancia de la información contenida en los reportes llenados por los analistas durante sus guardias, con el fin de mejorar la precisión y consistencia de los datos suministrados por ellos en sus análisis.

-
- ✓ Mejorar la codificación realizada en este proyecto realizando reuniones con cada grupo soporte, a fin de crear un catálogo completo de modo de fallas por especialidad. Este catálogo debería contemplar tanto las fallas ya ocurridas como las que se pudieran presentar, para que de ese modo los análisis se adapten a la naturaleza dinámica de la plataforma.
 - ✓ Uno de los proyectos llevados a cabo en paralelo con el presente estudio, es la implementación de un sistema para apoyar la gestión de incidentes de falla. Este sistema tiene por nombre Sistema de Gestión de Incidentes de Falla (SINCFA) y pudiera incorporar un módulo para ingresar reportes de falla de manera categorizada en los campos novedad, causa y acción, generando automáticamente el archivo que sirve de entrada a la máquina para clasificar los tipos de mantenimiento.
 - ✓ Hay que resaltar que la capacidad de generalización de clasificación de las máquinas de soporte vectorial está directamente relacionado con la información implícita en los datos utilizados durante su entrenamiento. En un futuro sería conveniente volver a entrenar la máquina si se cuenta con una cantidad considerable de datos nuevos.

Bibliografía

- Agrawal R. y Shafer J. (1996), Data Mining and Knowledge Discovery in Databases. IEEE Transactions on Knowledge and Data Engineering. Disponible en el sitio <http://elvex.ugr.es/etexts/spanish/kdd/KDD.html>.
- Barringer, P. (1996), An overview of reliability engineering principles. Patrocinado por ASME y organizado por Penn Well Conferences, Houston, TX.
- Carreras, X., Márquez, L. & Romero, E. (2004), Máquinas de Vectores Soporte, en Hernández, J., Ramírez, M. y Ferri, C., Introducción a la Minería de Datos, pp.353-382, Editorial Pearson, España.
- DRAE (2008), Diccionario de la lengua española (vigésima segunda edición). Disponible en el sitio <http://drae.rae.es/>.
- Díaz, L. (2002), Estadística multivariada: inferencias y métodos, Departamento de Estadística. Facultad de Ciencias, Primera edición, Universidad Nacional de Colombia. Bogotá D.C.
- Gunn, S. (1998), Support Vector Machines for classification and regression, University of Southampton, United Kingdom.
- Montaña L. y Rosas E. (2006), Diseño de un sistema de mantenimiento con base en análisis de criticidad y análisis de modos y efectos de falla en la Planta de Coque de Fabricación Primaria en la empresa Acerías Paz del Río S.A. Universidad Pedagógica y Tecnológica de Colombia. Duitama.

Mosquera, G. (2000), Estimación de parámetros de confiabilidad y mantenibilidad de sistemas industriales. Centro de altos estudios gerenciales ISID.

Rivera, R. (2006), Minería de datos. ¿Para qué nos sirve en una publicación digital?., Disponible en <http://www.mati.unam.mx/index.php>.

Vallejos, S. (2006), Minería de datos, Trabajo de adscripción, Universidad nacional del Nordeste, Corrientes- Argentina.

Yáñez M. (1995), Ingeniería de Confiabilidad: Pilar fundamental del mantenimiento. Disponible en el sitio http://www.confiabilidad.net/art_05/RCM/rcm_16.pdf

Anexo A

Tablas de Contingencia para el grupo Respaldo y Almacenamiento

En la tabla A1 y el gráfico A1 se puede apreciar alguna de las relaciones existentes entre las variables categóricas fallas y acción Se observa que:

Para el tipo de falla categorizada con el nombre falla de red están presentes las acciones: migrar los servidores conectados a la red, reconexión del cableado de fibra, consultar con el proveedor.

Para el tipo de falla categorizada con el nombre falla de respaldo están presentes todas las acciones correctivas aplicadas por el GOS respaldo y almacenamiento, exceptuando, chequear el sistema de energía eléctrica y reconectar el cableado.

Para el tipo de falla categorizada con el nombre falla de robot están presentes las acciones: consultar al proveedor, reportar la falla al grupo responsable, reprogramar el *backup*.

Para el tipo de falla categorizada con el nombre falla de la caja de almacenamiento están presentes las acciones: chequear el sistema de energía

eléctrica, cambiar discos, reparar los puertos dañados, recuperar los datos, consultar al proveedor.

Para el tipo de falla categorizada con el nombre falla de aplicación están presente las acciones: modificar el status de los discos, reprogramar el *backup*, consultar al proveedor, reportar la falla al grupo responsable, reconectar el cableado, modificar el status de los discos, reparar los puertos dañados, recuperación de datos.

Acción		Fallas					Total
		Falla de red	Falla de robot	Falla de respaldo	Falla en la caja de almacenamiento	Falla de aplicación	
		01NO01	01NO02	01NO03	01NO04	01NO05	
Cambiar los discos	01AC02	0	0	1	0	1	2
Reconexión del cableado	01AC08	2	0	0	0	0	2
Acción no registrada	01AC07	0	0	23	0	0	23
Reprogramar el backup	01AC11	0	1	66	3	0	70
Consultar al proveedor	01AC14	1	1	5	6	1	14
Chequear el servidor	01AC04	0	0	9	0	0	9
Reportar la falla al grupo	01AC10	0	4	46	1	0	51
Revisar la tarjeta de red	01AC12	0	0	2	0	0	2
Chequear el sistema de energía eléctrica	01AC05	0	0	0	0	1	1
Cambiar las cintas	01AC01	0	0	8	0	0	8
Migrar los servidores conectados a la red	01AC06	1	0	1	0	0	2
Reconfigurar el driver	01AC09	0	0	8	0	0	8
Cambiar el driver	01AC03	0	0	1	0	0	1
Modificar el status de los	01AC15	0	1	0	0	0	1
Ampliación los filesystem	01AC13	0	0	2	0	0	2
Reparar los puertos	01AC16	0	0	0	0	1	1
Recuperar los datos	01AC17	0	0	0	0	1	1
Total		4	7	172	10	5	198

Tabla A1. Matriz Cruzada de Acción vs. Falla

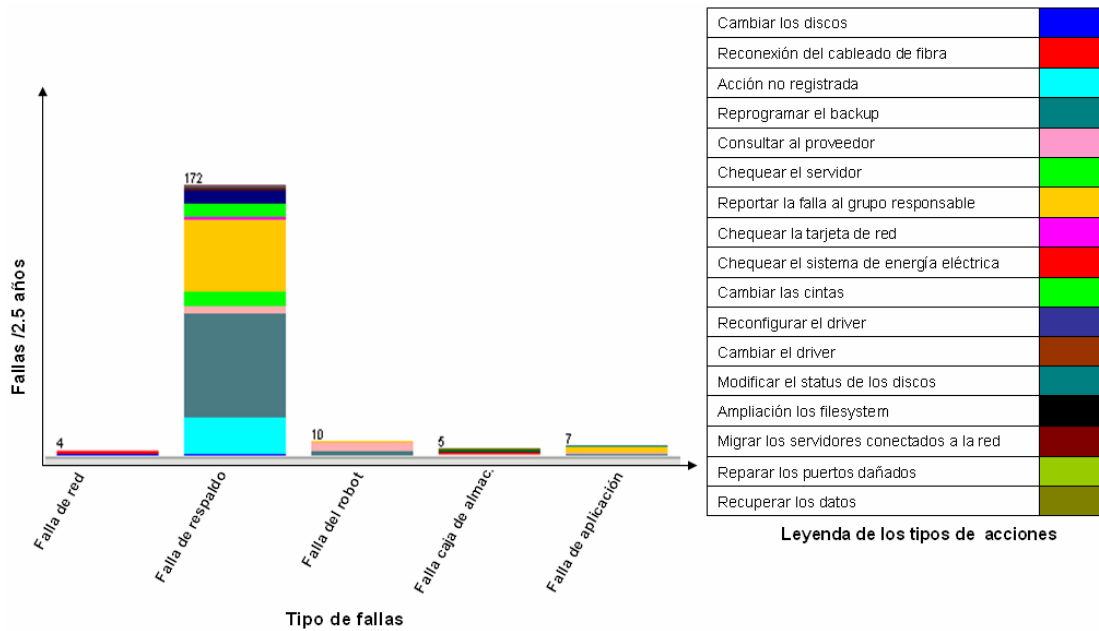


Figura A1. Representación gráfica de la matriz cruzada de Acción vs. Fallas

En la tabla A2 y el gráfico A2, se puede apreciar la alta relación existente entre las categoría de las falla y los componentes físicos, esto se debe a la naturaleza física de los equipos y sistemas que maneja el GOS respaldo y almacenamiento.

Tipo de componente		Fallas					Total
		Falla de red	Falla de robot	Falla de respaldo	Falla en la caja de almacenamiento	Falla de aplicación	
		01N001	01N002	01N003	01N004	01N005	
Componente Físico	CF	4	0	144	10	5	163
Componente Lógico	CL	0	7	28	0	0	35
Total		4	7	172	10	5	198

Tabla A2. Matriz Cruzada de Tipo de componente vs. Falla

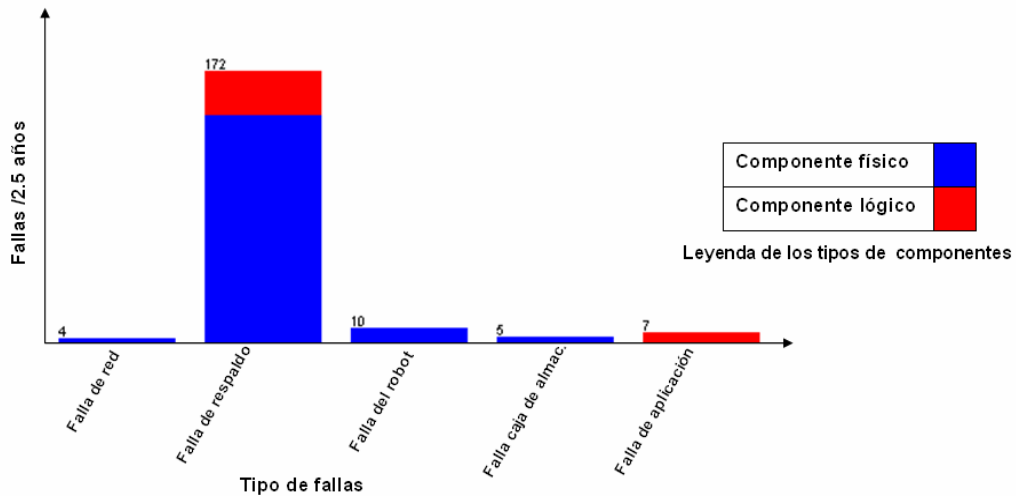


Figura A2. Representación gráfica de la matriz cruzada de Tipo de componente vs. Fallas

En la tabla A3 y el gráfico A3, se puede observar que la mayoría de las fallas en los equipos y sistemas atendidas por este grupo soporte, tienen un impacto generalmente medio y bajo al ser comparados con los demás servicios que presta la plataforma de PDVSA-AIT.

Impacto		Fallas					Total
		Falla de red	Falla de robot	Falla de respaldo	Falla en la caja de almacenamiento	Falla de aplicación	
		01N001	01N002	01N003	01N004	01N005	
Alto	Alto	3	3	9	1	3	19
Impacto no registrado	impacto no regsitrado	1	3	62	2	0	68
Bajo	Bajo	0	1	85	0	2	88
Medio	Medio	0	0	16	7	0	23
Total		4	7	172	10	5	198

Tabla A3. Matriz Cruzada de Impacto vs. Falla

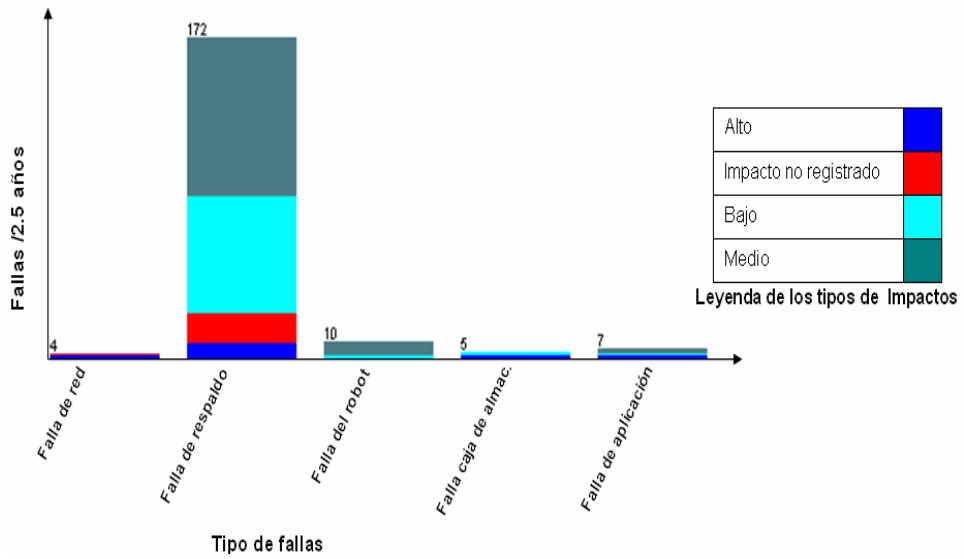


Figura A3. Representación gráfica de la matriz cruzada de Impacto vs. Fallas

Anexo B

Análisis de Correspondencia Múltiple para los grupos faltantes

Grupo 02. Infraestructura

Nombre de la Falla	Categoría	Código
Falla de energía	02NO01	1
Falla de UPS	02NO02	2
Falla del aire acondicionado	02NO03	3
Falla del motogenerador	02NO04	4
Falla del rectificador	02NO05	5
Falla de unidad de transferencia de energía	02NO06	6

Tabla B1. Tipo de falla

Nombre de la Causa	Categoría	Código
Causa desconocida	02CA01	1
Componentes dañados	02CA02	2
Condiciones de alta temperatura disparo el UPS	02CA03	3
Cortocircuito	02CA04	4
Descarga atmosférica	02CA05	5
Elemento o piezas del aire acondicionado dejaron de funcionar	02CA06	6
Elemento o piezas del rectificador dejaron de funcionar	02CA07	7

Nombre de la Causa	Categoría	Código
Elementos o piezas del motogenerador dejaron de funcionar	02CA08	8
Elementos o piezas del UPS dejaron de funcionar	02CA09	9
Maniobra indebida realizada accidentalmente	02CA10	10
Problemas alimentación red de energía pública	02CA11	11
Sobrecarga de energía eléctrica	02CA12	12

Tabla B2. Causa de la falla

Nombre de la Acción	Categoría	Código
Chequear y reparar el circuito donde ocurre la falla	02AC01	1
Chequear y reparar la unidad con apoyo del equipo de servicio logístico	02AC02	2
Verificar correcto arranque del sistema de respaldo y contactar la empresa de alimentación pública	02AC03	3
Colocar en funcionamiento el aire acondicionado de respaldo	02AC04	4
Contactar la contratista para que realicen inspecciones y mantenimiento general	02AC05	5
Dimensionar el breaker de acuerdo al requerimiento de consumo del circuito	02AC06	6
Realizar la programación de la temperatura establecida para el equipo	02AC07	7
Subir el interruptor	02AC08	8
Verificar la correcta operación del UPS	02AC09	9

Tabla B3. Acción ejecutada para finalizar la falla

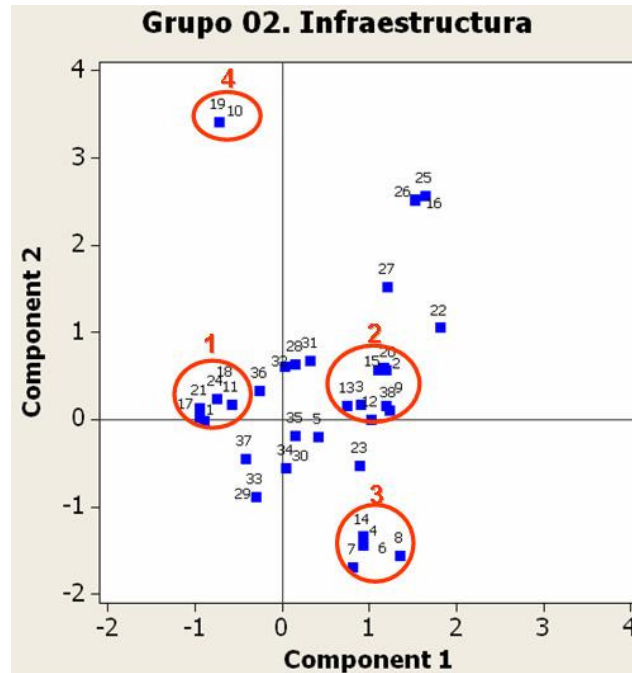


Figura B1. Análisis de Correspondencia Múltiple Grupo Infraestructura

Definición de la variable de salida para el grupo Infraestructura

(1): Mantenimiento correctivo continuo realizados por externos (compañía encargada del suministro de energía eléctrica) cuyas causas se debe básicamente a problemas ambientales, exógenos a la industria, afectando altamente el servicio por las interrupciones provocadas por fliquesos y caídas abruptas del servicio de energías eléctrica pública, con variaciones en corto circuitos originados por sobrecarga con restauración aceptable.

(2): Mantenimiento correctivo frecuente a dispositivos de respaldo en caso de fallas de energía eléctrica pública (UPS, Aires acondicionados) con soluciones a través de mantenimiento general y reemplazos de piezas al UPS y unidades de Aires

aconicionados, apoyados en el equipo de servicio logístico, presentando frecuente ocurrencia de falla, ocasionando impacto medio al servicio y de inmediata restauración.

(3): Mantenimiento correctivo intermitente a los equipos motogeneradores, provocados por fallas en sus componentes o por la unidad de transferencia de energía, con restauración aceptable ocasionando impacto medio al servicio, sin mayor ocurrencia de fallas.

(4): Mantenimiento correctivo escaso provocado por sobrecarga en los circuitos eléctricos debido a la sobredimensión de los mismos, con variaciones en disparos de *Brecker* restauración inmediata, bajo impacto al servicio, y de baja ocurrencia de fallas.

Grupo 03. Redes LAN

Nombre de la Falla	Categoría	Código
Falla de los servicios en la extranet	03NO01	1
Falla de comunicación	03NO02	2
Falla de la fuente de poder del equipo	03NO03	3
Interrupción del servicio de Internet	03NO04	4

Tabla B4. Tipo falla

Nombre de la Causa	Categoría	Código
Causa desconocida	03CA01	1
Alto consumo de banda	03CA02	2
Problema con el cableado	03CA03	3
Falla a nivel de la configuración del protocolo de enrutamiento de Internet	03CA04	4

Nombre de la Causa	Categoría	Código
Falla de <i>hardware</i>	03CA05	5
Interfaz Fastethernet/ serial/ETH/GETH/G703/ATM. Inhibida o dañada	03CA06	6
Problema por parte del proveedor	03CA07	7
Problema con el <i>routers</i>	03CA08	8
Problema con el <i>switch</i>	03CA09	9
Problema con el enlace	03CA10	10
Problema de energía eléctrica	03CA11	11
Problema de conexión con el servidor	03CA12	12
Problema en enlace de voz	03CA13	13

Tabla B5. Causa de la falla

Nombre de la Acción	Categoría	Código
Acción no registrada	02AC01	1
Ajustar la conexión y chequear los puertos	02AC02	2
Cambiar la configuración de la interfaz	02AC03	3
Configurar router / swiche / access point	02AC04	4
Reiniciar el <i>swiche</i>	02AC05	5
Restringir el flujo de datos desde el <i>router</i>	02AC06	6
Colocar conexiones UTP entre los <i>switch</i>	02AC07	7
Realizar pruebas a los equipos	02AC08	8
Realizar pruebas de ruta y conexión	02AC09	9
Cambiar el <i>swiche</i>	02AC10	10
Cambiar el <i>routers</i>	02AC11	11
Reemplazar la tarjeta supervisora	02AC12	12
Contactar al proveedor para solventar la falla	02AC13	13
Reportar la falla al grupo responsable	02AC14	14
Reiniciar el <i>router</i>	02AC15	15
Solicitar al proveedor aumento de ancho de banda	02AC16	16

Tabla B6. Acción ejecutada para finalizar la falla

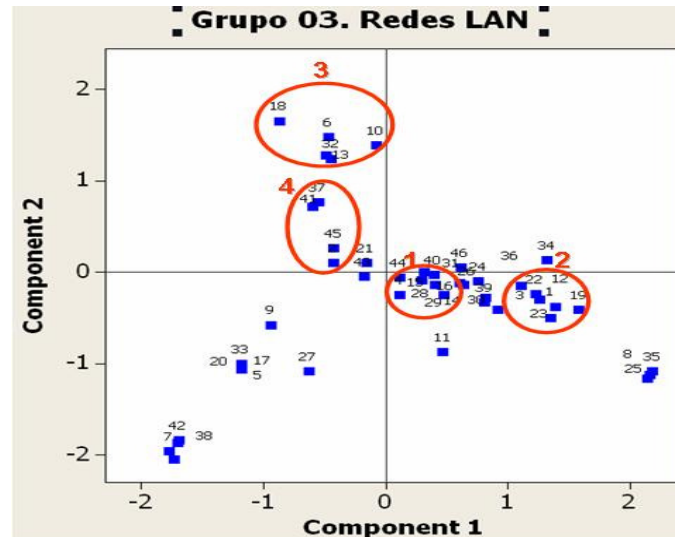


Figura B2. Análisis de Correspondencia Múltiple Grupo Redes LAN

Definición de la variable de salida para el grupo Redes LAN

(1): Mantenimiento correctivo continuo a los elementos físicos de red tales como *routers*, *switches* y enlaces las cuales provocan fallas de comunicación y por ende caída de algunos servicios críticos usados en las actividades de rutina de la empresa, con soluciones en reemplazo, alquiler y pruebas de conexión física., presentando alta ocurrencia de falla, ocasionando impacto alto al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente en los componentes lógicos ocasionados por desconfiguración del software de red y en consecuencia se presentan problemas en los servicios de comunicación de PDVSA, con solución en configurar la interfaz inhibida, *switches*, *routers* y *access point* ocasionando impacto medio al servicio y aceptable restauración.

(3): **Mantenimiento correctivo intermitente** a los servicios de Internet generados por interrupción del servicio por parte del proveedor o falla a nivel de la configuración del protocolo de enrutamiento de Internet con restauración aceptable ocasionando impacto medio al servicio, sin mayor ocurrencia de fallas.

(4): **Mantenimiento correctivo escaso** por problemas de energización a los elementos de red por caídas del servicio eléctrico, con soluciones en por parte del grupo soporte de Infraestructura, restauración inmediata, bajo impacto al servicio, y de baja ocurrencia de fallas.

Grupo 04. Servidores Windows

Nombre de la Falla	Categoría	Código
Falla del servidor	04NO01	1
Falla de aplicación	04NO02	2
Falla de impresión	04NO03	3
Falla de comunicación con el servidor	04NO04	4

Tabla B7. Tipo falla

Nombre de la Causa	Categoría	Código
Corrupción de archivo y/o problema de registro	04CA01	1
Falla de procesador/ memoria/ tarjeta madre	04CA02	2
Falla eléctrica y/o aire acondicionado	04CA03	3
Fallan equipos de la red	04CA04	4
Perdida de conexión con los discos y/o discos alarmados	04CA05	5
Problemas con el servicio de Impresión	04CA06	6
Problemas de espacio en disco	04CA07	7

Nombre de la Causa	Categoría	Código
Problemas de software/ Sistema operativo corrupto	04CA08	8
Problemas de virus	04CA09	9
Problemas con el driver de la impresora	04CA10	10
Problemas con el servicio de cluster	04CA11	11
Problemas con las direcciones IP	04CA12	12

Tabla B8. Causa de la falla

Nombre de la Acción	Categoría	Código
Actualizar las aplicaciones	04AC01	1
Configurar y/o reemplazar los <i>driver</i> de impresión	04AC02	2
Liberar direcciones IP	04AC03	3
Realizar limpieza en el disco	04AC04	4
Recuperar los archivos dañados y reiniciar el equipo o servicio	04AC05	5
Reiniciar el servicio de impresión	04AC06	6
Reiniciar el servidor	04AC07	7
Reportar la falla al proveedor de servicio	04AC08	8
Reportar la falla al grupo de infraestructura	04AC09	9
Reportar la falla al grupo de redes	04AC10	10
Reportar la falla al proveedor para que realicen: configuración/ reemplazo/ chequeo general	04AC11	11
Configurar/ corregir el servicio de cluster	04AC12	12

Tabla B9. Acción ejecutada para finalizar la falla

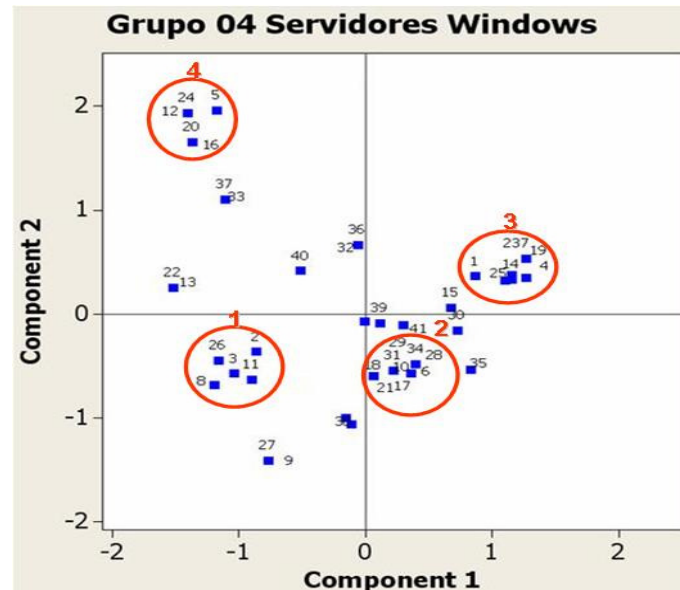


Figura B3. Análisis de Correspondencia Múltiple Grupo Servidores Windows

Definición de la variable de salida para el grupo Servidores Windows

(1): Mantenimiento correctivo continuo, a los elementos físicos de los servidores tales como: procesador, memorias, tarjetas, dispositivos físicos de almacenamientos, con soluciones por parte del proveedor presentando alta ocurrencia de falla, ocasionando impacto alto al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente, fallas de comunicación con el servidor, ocasionados por problemas en la red, problemas con el servicio de cluster y ausencias de direcciones IP, con soluciones apoyadas en el grupo soporte de redes y proveedor de servicio externo, impacto alto al servicio, con frecuente ocurrencia de falla y restauración inmediata.

(3): Mantenimiento correctivo intermitente, fallas en aplicaciones del servidor, provocadas por el deterioro en componentes lógicos, tales como: corrupción de archivos, sistema operativo, problemas con el registro de Windows, falla de software, con soluciones en actualización de versiones, reparación y recuperación de archivos dañados y reinicios del servidor por parte del grupo solucionador de servidores Windows, con restauración aceptable ocasionando impacto medio al servicio, sin mayor ocurrencia de fallas.

(4): Mantenimiento correctivo escaso al servicio de cola de impresión, por inhibición del servicio, causado por daños en archivos y *driver* de impresión, la solución que brinda el grupo soporte esta dirigida a realizar acciones de configuración o reemplazos del *driver* de impresión y reinicio de este servicio, con restauración inmediata, bajo impacto al servicio, y de baja ocurrencia de fallas.

Grupo 05. Servidores Unix

Nombre de la Falla	Categoría	Código
Falla de comunicación con el servidor	05NO01	1
Falla del servidor	05NO02	2
Falla de servicio	05NO03	3

Tabla B10. Tipo falla

Nombre de la Causa	Categoría	Código
Alta concurrencia de usuario/ alta solicitud de recursos del servidor	05CA01	1
Falla eléctrica y/o aire acondicionado	05CA02	2
Falla tarjeta/ memoria/discos	05CA03	3
Problemas con los <i>filesystem</i>	05CA04	4

Nombre de la Causa	Categoría	Código
Problemas con los servicios de red	05CA05	5
Problemas en los dispositivos físicos externos de almacenamiento	05CA06	6
Problemas software, archivos corrupto y/o problema con el sistema operativo	05CA07	7

Tabla B11. Causa de la falla

Nombre de Acción	Categoría	Código
Desmontar y montar los <i>filesystem</i> / reiniciar el servidor	05AC01	1
Eliminar archivos/ampliar los <i>filesystem</i>	05AC02	2
Levantar el servicio/ reiniciar el servidor	05AC03	3
Liberar memoria/limpiar discos/ contactar al proveedor	05AC04	4
Mantenimiento de los archivos/ actualizar versiones/reiniciar el servidor	05AC05	5
Reportar al grupo redes LAN	05AC06	6
Reportar la falla al grupo infraestructura	05AC07	7

Tabla B12. Acción ejecutada para finalizar la falla

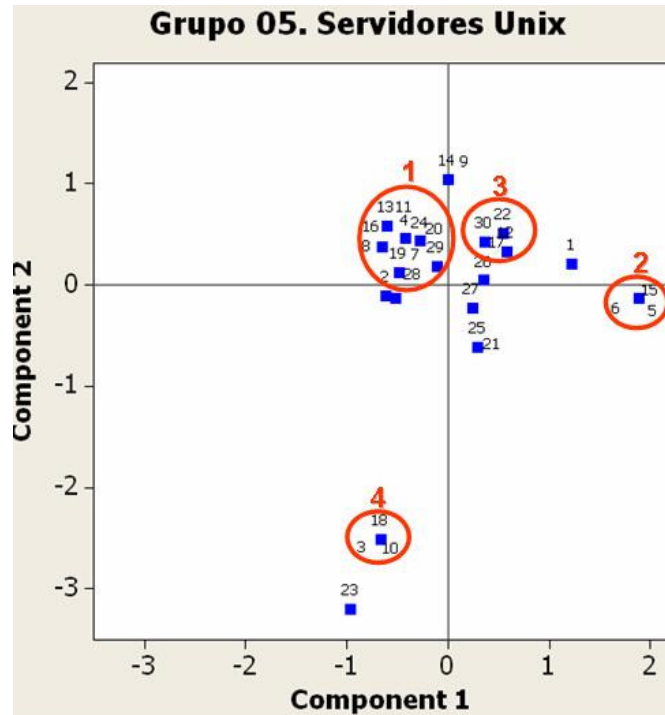


Figura B4. Análisis de Correspondencia Múltiple Grupo Servidores Unix

Definición de la variable de salida para el grupo Servidores Unix

(1): **Mantenimiento correctivo continuo**, al servidor por problemas en los componentes físicos tales como, (tarjetas, memorias, discos), con variaciones en problemas generados por altas temperaturas en la sala de computo, con soluciones en la configuración, liberación de espacios en los *filesystem*, mantenimientos o reemplazos a las piezas de hardware por parte del proveedor de servicio, presentando alta ocurrencia de falla, ocasionando impacto alto al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente a los servicios prestados por el grupo soporte de servidores Unix, cuya falla se debe generalmente a la alta concurrencia de usuario y alta solicitud de recursos del servidor tales como (memoria, capacidad de proceso), las soluciones están asociadas en realizar acciones relacionadas con la liberación de memoria, ejecución de comandos para tal fin, así como reinicio del servicio y/o servidor, ocasionando un alto impacto, con frecuente ocurrencia de falla y restauración inmediata.

(3): Mantenimiento correctivo intermitente, a los componentes Lógicos, como: software, archivos de aplicaciones y archivos de configuración del sistema operativo; tal solución esta referida en la ejecución de acciones dirigidas a actualización de versiones, mantenimiento de archivos, reinicios del servidor entre otras, por parte del grupo solucionador de servidores Unix, con restauración aceptable ocasionando impacto medio al servicio, sin mayor ocurrencia de fallas.

(4): Mantenimiento correctivo, escaso a los servicios de comunicación, mantenimiento general a los dispositivos de red, por parte del grupo soporte de redes LAN, con restauración inmediata, impacto medio al servicio, y de baja ocurrencia de fallas.

Grupo 06. Redes de Transmisión

Nombre de la Falla	Categoría	Código
Falla de radio	06NO01	1
Falla de telemetría	06NO02	2
Falla terminal Server de la supervisión	06NO03	3

Tabla B13. Tipo falla

Nombre de la Causa	Categoría	Código
Fallas de banda base/multiplexor/IDU	06CA01	1
Falla de energía eléctrica	06CA02	2
Falla de maestra de telemetría	06CA03	3
Falla de recepción/ transmisión/ ODU	06CA04	4
Falla de terminal Server MOXA/ DIGI/LANTRONIX UDS200	06CA05	5
Falla remota de telemetría	06CA06	6

Tabla B14. Causa de la falla

Nombre de la Acción	Categoría	Código
Chequear conectores y cables	06AC01	1
Chequear el flujo de comunicación	06AC02	2
Chequear/resetear el ODU	06AC03	3
Chequeo general a los equipos y reportar a infraestructura	06AC04	4
Realizar pruebas con el <i>terminal Server/</i> <i>resetear</i> el servicio	06AC05	5
Reemplazar o reiniciar tarjeta	06AC06	6
Reparar los conectores <i>Bitex/Bitin</i>	06AC07	7
<i>Reset MODEM</i>	06AC08	8
Revisar unidad de multiplexor	06AC09	9
Verificación protocolo interconexión PLC/Radio	06AC10	10

Tabla B15. Acción ejecutada para finalizar la falla

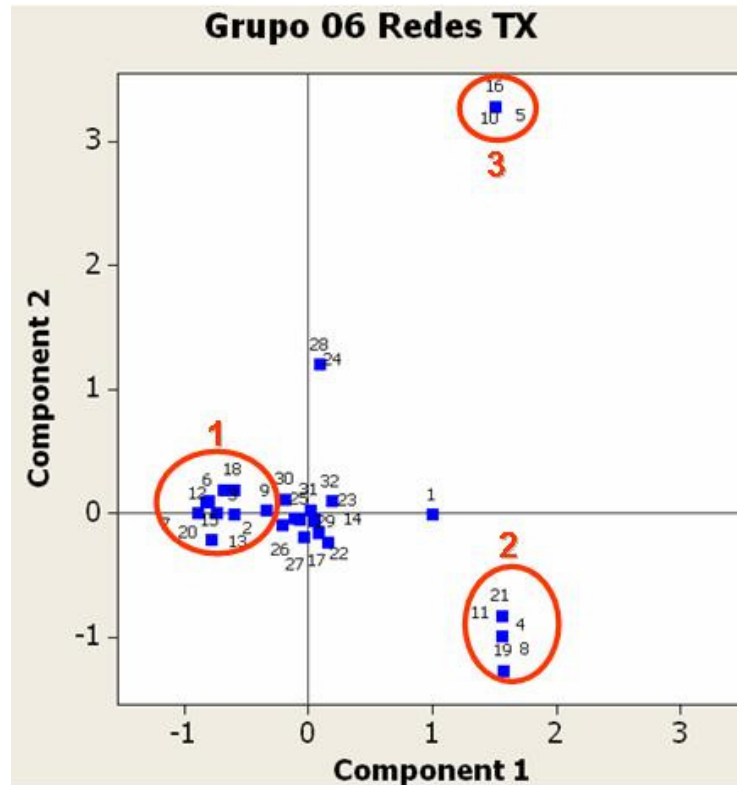


Figura B5. Análisis de Correspondencia Múltiple Grupo Redes de Transmisión

Definición de la variable de salida para el grupo Servidores Unix

(1): Mantenimiento correctivo continuo, a los radios, cuya falla es provocada por problemas de banda base/*multiplexor*/IDU, problemas de energía eléctrica, problemas de recepción/ transmisión/ ODU, con soluciones en técnicas orientadas al chequeo de conectores y cables, revisar el flujo de comunicación, reparar los conectores *Bitex/Bitin*, notificar al proveedor para el chequeo general de la unidad de

multiplexor, presentando alta ocurrencia de falla, ocasionando impacto alto al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente, realizados a los sistemas de telemetría, estas fallas se presentan básicamente por problemas en los enlaces remotos y maestras, con soluciones en el chequeo general de protocolo interconexión PLC/Radio y unidades *MODEM*, ocasionando impacto bajo al servicio por su restauración inmediata, con frecuente ocurrencia de fallas.

(3): Mantenimiento correctivo intermitente, a los *terminal Server* de supervisión, ocasionados por Falla de *terminal Server* MOXA/ DIGI/LANTRONIX UDS200, con restauración aceptable ocasionando impacto bajo al servicio, sin mayor ocurrencia de fallas.

Grupo 08. Ambiente de Trabajo Colaborativo (ATC)

Nombre de la Falla	Categoría	Código
Falla de comunicación	08NO01	1
Falla del servicio de correo <i>Lotus notes</i>	08NO02	2
Falla servidor	08NO03	3
Falla del servicio <i>quickplace</i>	08NO04	4
Falla del servicio <i>sametime</i>	08NO05	5

Tabla B16. Tipo falla

Nombre de la Causa	Categoría	Código
Directorio y/o archivo corrupto	08CA01	1
Problema con la base de datos	08CA02	2
Problema de software	08CA03	3

Nombre de la Causa	Categoría	Código
Problemas en los dispositivos físicos de almacenamiento	08CA04	4
Falla del hardware del servidor	08CA05	5
Instalación y/o configuración de hardware	08CA06	6
Problemas con el directorio de PDVSA	08CA07	7
Problemas conexión con la red	08CA08	8

Tabla B17. Causa de la falla

Nombre de la Acción	Categoría	Código
Se reporta al grupo de servidores NT	09AC01	1
Mantenimiento a la bases de datos y reiniciar el servicio	09AC02	2
Reiniciar servidor	09AC03	3
Reemplazar el directorio	09AC04	4
Reportar la falla al grupo de redes	09AC05	5

Tabla B18. Acción ejecutada para finalizar la falla

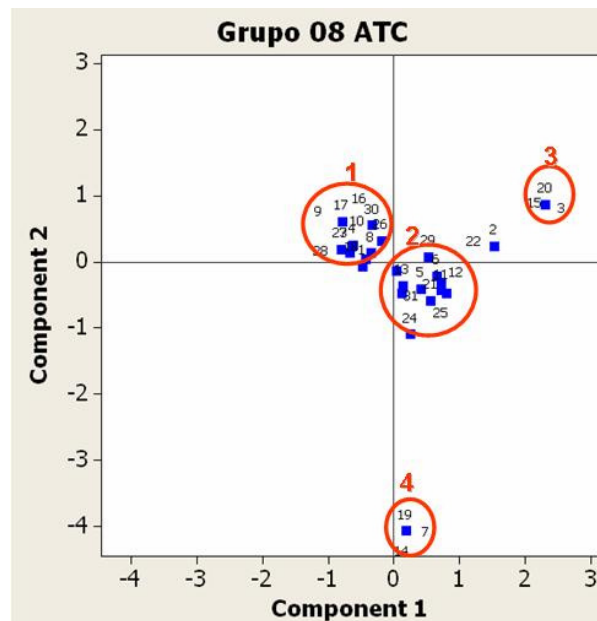


Figura B6. Análisis de Correspondencia Múltiple Grupo ATC

Definición de la variable de salida para el grupo ATC

(1): Mantenimiento correctivo continuo, del servicio de correo *Lotus Notes*, este servicio falla básicamente por problemas en componentes lógicos tales como: pérdida del directorio de PDVSA, daños en archivos, problemas con la base de datos, en general problemas de software, con soluciones apoyadas por el grupo servidores NT, presentando alta ocurrencia de falla, ocasionando impacto bajo al servicio y con restauración lenta.

(2): Mantenimiento correctivo frecuente, del servicio de *quickplace* con variaciones en fallas del servidor, generados por problemas en los dispositivos físicos de almacenamientos (discos, cintas), problemas de hardware del servidor, ocasionando impacto medio al servicio y de restauración inmediata, con frecuente ocurrencia de fallas.

(3): Mantenimiento correctivo intermitente, a los servicios de comunicación los cuales afectan la conexión con el servidor y en consecuencia los servicios prestados por ATC, con restauración inmediata ocasionando impacto bajo al servicio, sin mayor ocurrencia de fallas.

(4): Mantenimiento correctivo escaso al servicio *sametime*, esta falla se origina como consecuencia de la pérdida o daño en el directorio de PDVSA, y cuya solución es el reemplazo de este directorio, con restauración inmediata, bajo impacto al servicio, y de baja ocurrencia de fallas.

Grupo 09. Redes WAN

Nombre de la Falla	Categoría	Código
Falla de nodo ATM	09NO01	1
Falla de nodo promina	09NO02	2
Falla SDH	09NO03	3

Tabla B19. Tipo falla

Nombre de la Causa	Categoría	Código
Falla comunicación	09CA01	1
Falla eléctrica	09CA02	2
Problema con la tarjeta trocal	09CA03	3
Problema en el multiplexor SDH	09CA04	4
Problema en la tarjeta	09CA05	5
Problemas en el enlace y/o troncal	09CA06	6
Racionalización de servicios del modulo SDH	09CA07	7

Tabla B20. Causa de la falla

Nombre de la Acción	Categoría	Código
Ajustar conectores y/o puertos	09AC01	1
Ajustar el conector de la tarjeta en el nodo	09AC02	2
Chequear el cableado de los equipo, conectores y/o puertos	09AC03	3
realizar pruebas de BERT	09AC04	4
Reconfigurar la ruta de transporte	09AC05	5
Reiniciar o reemplazar la tarjeta troncal	09AC06	6
Reiniciar/ configurar/ reemplazar, las tarjetas del nodo	09AC07	7
Reportar al grupo de infraestructura	09AC08	8

Tabla B21. Acción ejecutada para finalizar la falla

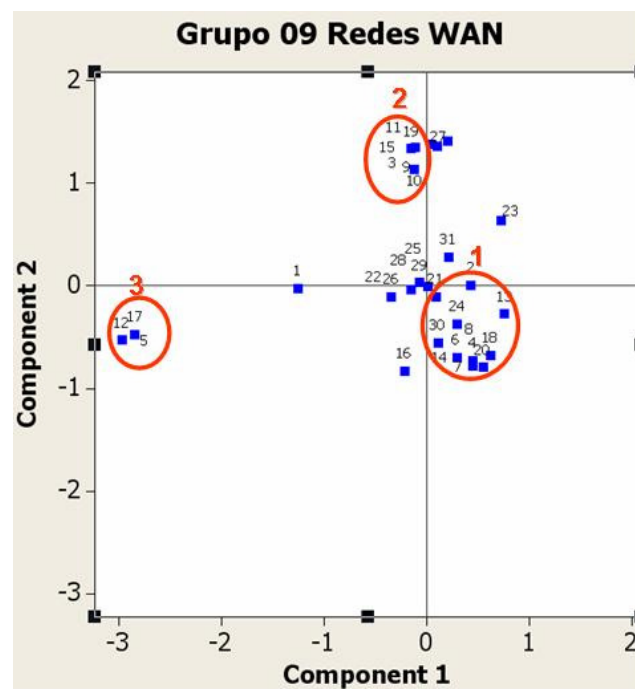


Figura B7. Análisis de Correspondencia Múltiple Grupo Redes Wan

Definición de la variable de salida para el grupo Redes Wan

(1): Mantenimiento correctivo continuo, en los nodos promina, originado por problemas en los dispositivos de comunicación, desconfiguración/desconexión en la tarjeta troncal o falla en el suministro de energía eléctrica, con soluciones puntuales realizadas por el grupo soporte orientadas a reconexiones, configuraciones ajustes y mantenimiento general a los dispositivos de comunicación, presentando alta ocurrencia de falla, ocasionando impacto medio al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente realizados a los nodos ATM, esta falla es generada por problemas en el multiplexor SDH, fallas en la tarjeta, problemas en el enlace y/o troncal, con soluciones de conectividad esto es (reconstruir cables, verificar conexiones, puertos, enlaces) y reemplazo y/o configuración de tarjetas, ocasionando impacto medio al servicio y de restauración aceptable, con frecuente ocurrencia de fallas.

(3): Mantenimiento correctivo intermitente, en los SDH originado por la racionalización de servicios del modulo SDH y cuya solución esta orientada a la reconfiguración de la ruta de trasporte, con restauración inmediata ocasionando impacto bajo al servicio, sin mayor ocurrencia de fallas.

Grupo 11. Servidores AS-400

Nombre de la Falla	Categoría	Código
Falla de comunicación con el servidor	11NO01	1
Falla de energía eléctrica afecta el servidor	11NO02	2
Falla del servidor	11NO03	3

Tabla B22. Tipo falla

Nombre de la Causa	Categoría	Código
Problemas con el sistema operativo	11CA01	1
Problemas en la red	11CA02	2
Problemas en los dispositivos físicos de almacenamiento	11CA03	3
Problemas del suministro de energía eléctrica	11CA04	4
Falla del modulo de <i>power suply</i> del AS400	11CA05	5

Tabla B23. Causa de la falla

Nombre de la Acción	Categoría	Código
mantenimiento o cambio de modulo de <i>power suply</i>	11AC01	1
Mantenimiento y configuración del sistema operativo	11AC02	2
notificar al personal de Infraestructura	11AC03	3
notificar al grupo responsable	11AC04	4
contactar la contratista IBM para que realice las labores de mantenimiento	11AC05	5

Tabla B24. Acción ejecutada para finalizar la falla

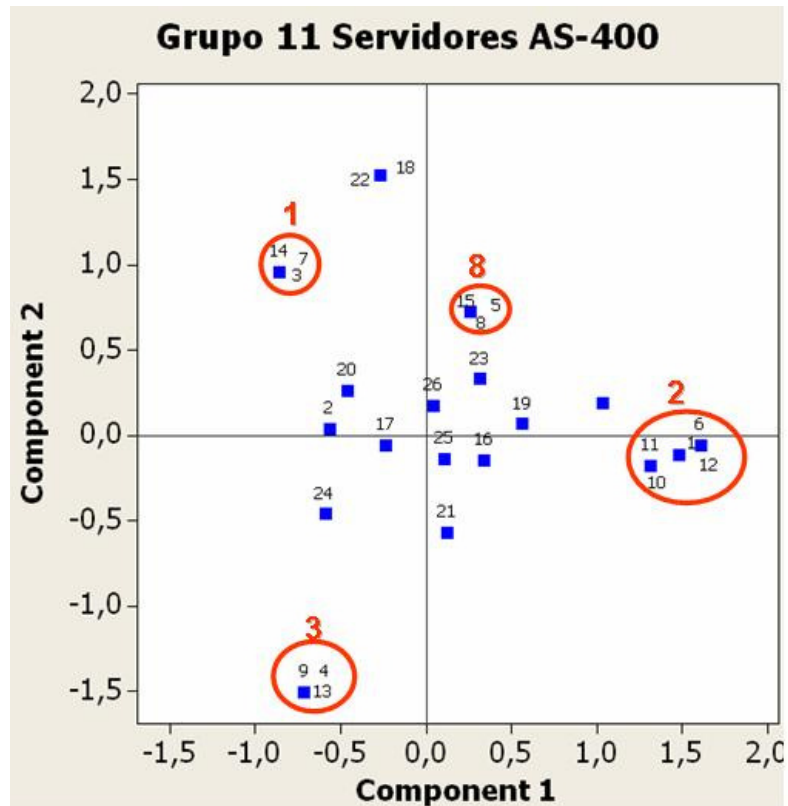


Figura B8. Análisis de Correspondencia Múltiple Grupo Servidores AS400

Definición de la variable de salida para el grupo Servidores AS400

(1): Mantenimiento correctivo frecuente por parte del grupo soporte redes LAN a los dispositivos de comunicación, los cuales imposibilitan acceder a los servicios prestados por los servidores AS-400, presentando alta ocurrencia de fallas, ocasionando impacto medio al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente, a los componentes lógicos tales como: archivos, registro y en general al sistema operativo del servidor, esto produce fallas en el servicio, tales como: problemas con el módulo de *power supply* del AS400, con

soluciones de configuración del sistema operativo, reemplazo o mantenimiento a *power supply*, ocasionando impacto bajo al servicio y de restauración lenta, con frecuente ocurrencia de fallas.

(3): Mantenimiento correctivo intermitente, al sistema de alimentación de energía eléctrica que energiza al servidor, por parte del grupo de infraestructura, con restauración inmediata ocasionando impacto medio al servicio, con frecuente ocurrencia de fallas.

(4): Mantenimiento correctivo escaso al servidor por problemas en los dispositivos físicos de almacenamiento, con soluciones por parte del proveedor IBM, con restauración lenta, bajo impacto al servicio, y de baja ocurrencia de fallas.

Grupo 15. Aplicaciones SAP

Nombre de la Falla	Categoría	Código
Falla ambiente de calidad	15NO01	1
Falla ambiente pr1	15NO02	2
Falla ambiente pr2	15NO03	3

Tabla B25. Tipo falla

Nombre de la Causa	Categoría	Código
Problema de memoria	15CA01	1
Problemas con el servidor	15CA02	2
Problemas con la base de datos	15CA03	3
Problemas con los archivos y/o directorios	15CA04	4
Problemas con sistema QA1	15CA05	5
Problemas en medios de almacenamiento	15CA06	6
Problemas para acceder a la aplicación	15CA07	7

Tabla B26. Causa de la falla

Nombre de la Acción	Categoría	Código
Aumentar la memoria	15AC01	1
Depurar/ chequear el ambiente pr2	15AC02	2
Notificar la falla al grupo respaldo y almacenamiento	15AC03	3
Reparar la base de datos	15AC04	4
Reportar la falla al equipo de servidores	15AC05	5
Reportar la falla al grupo <i>Basic</i>	15AC06	6

Tabla B27. Acción ejecutada para finalizar la falla

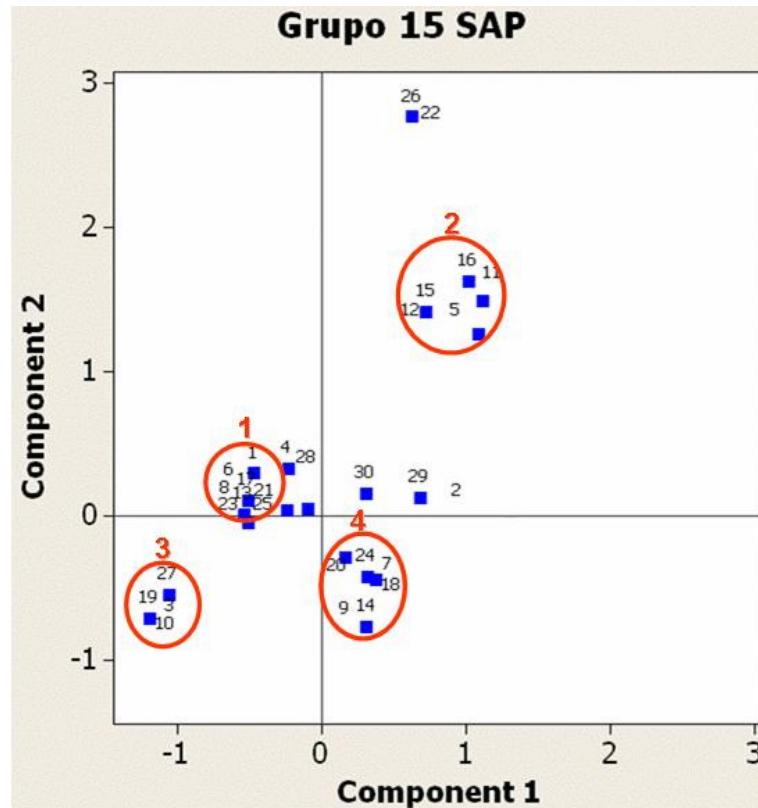


Figura B9. Análisis de Correspondencia Múltiple Grupo Aplicaciones SAP

Definición de la variable de salida para el grupo Aplicaciones SAP

(1): Mantenimiento correctivo continuo por parte del grupo soporte servidores NT a los componentes lógicos, tales como: base de datos, archivos corruptos y problemas de memoria que afectan la ejecución de las aplicaciones de SAP, presentando alta ocurrencia de fallas, ocasionando impacto alto al servicio y con restauración aceptable.

(2): Mantenimiento correctivo frecuente, al ambiente PR2 de SAP, esta falla es ocasionada por problemas en los dispositivos de almacenamiento (discos

generalmente) y problemas en el software el cual impide acceder a la aplicación, con soluciones por parte del grupo respaldo y almacenamiento y el grupo soporte *Basic*, ocasionando impacto alto al servicio y de restauración lenta, con frecuente ocurrencia de fallas.

(3): Mantenimiento correctivo intermitente, al ambiente de calidad de SAP, ocasionado por fallas en la aplicación QA1, con soluciones por parte del grupo soporte *Basic*, con restauración inmediata ocasionando impacto alto al servicio, con frecuente ocurrencia de fallas.

(4): Mantenimiento correctivo escaso al ambiente PR1 de SAP, por problemas con los archivos y/o directorios, con soluciones orientadas en la depuración del ambiente PR1 por parte grupo soporte *Basic* y servidores NT, con restauración lenta, impacto medio al servicio, y de baja ocurrencia de fallas.