



Universidad
Internacional
de Andalucía

TÍTULO

ANÁLISIS DE DATOS DE DISPENSACIÓN DE MEDICAMENTOS A PACIENTES HOSPITALIZADOS

UN ENFOQUE BASADO EN WEKA Y MINERÍA DE DATOS

AUTORA

Encarnación Gómez Fernández

	Esta edición electrónica ha sido realizada en 2025
Director	Dr. José del Campo Ávila
Institución	Universidad Internacional de Andalucía
Curso	<i>Diploma de Especialización en Digitalización e Innovación en Salud (2022-23)</i>
©	Encarnación Gómez Fernández
©	De esta edición: Universidad Internacional de Andalucía
Fecha documento	2023



Universidad
Internacional
de Andalucía



**Atribución-NoComercial-SinDerivadas
4.0 Internacional (CC BY-NC-ND 4.0)**

Para más información:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>



ANÁLISIS DE DATOS DE DISPENSACIÓN DE MEDICAMENTOS A PACIENTES HOSPITALIZADOS

Un enfoque basado en WEKA y minería de datos

Encarnación Gómez Fernández

Director: José del Campo Ávila
DIPLOMA DE ESPECIALIZACIÓN DIGITALIZACIÓN E INNOVACIÓN EN SALUD
Universidad internacional de Andalucía
Curso académico: 2022/2023

INDICE

RESUMEN.....	2
ABSTRACT.....	3
I. INTRODUCCIÓN.....	4
II. OBJETIVOS.....	6
III. MATERIAL Y MÉTODO.....	6
A. Obtención y preprocesamiento de datos.....	6
B. Análisis de datos.....	7
IV. RESULTADOS.....	11
IV.1. Resultados del análisis de informes de llenado de carros de unidosis.....	11
IV.2. Resultados del análisis del informe de prescripción de un medicamento determinado e informe de analíticas de laboratorio.....	13
IV.3. Resultados del análisis de informes de dispensación de sistemas automáticos.....	20
V. DISCUSIÓN.....	24

RESUMEN

Antecedentes: La prescripción electrónica y la automatización de la dispensación de medicamentos en el entorno hospitalario han mejorado la eficiencia y seguridad en la dispensación. Estos procesos generan una gran cantidad de datos que ofrecen oportunidades para su análisis y la optimización de procesos. La minería de datos, respaldada por herramientas como WEKA, se presenta como una solución para extraer conocimiento valioso de estos datos en la Farmacia Hospitalaria.

Material y método: Los datos de dispensación se recopilaron de tres fuentes: informes de llenado de carros de unidosis, informes de prescripción de un medicamento específico y análisis de datos de laboratorio e informes de sistemas automáticos de dispensación. Los datos fueron preprocesados utilizando *Microsoft Excel*. Para el análisis, se utilizó el programa WEKA, aplicando el algoritmo Apriori para reglas de asociación y el algoritmo J48 de árbol de decisión para reglas de clasificación.

Resultados: Se identificaron patrones de asociación entre diferentes principios activos y grupos terapéuticos dispensados en un servicio determinado. Sin embargo, no se obtuvieron relaciones significativas de clasificación entre los medicamentos dispensados y el servicio. Por otro lado, se establecieron diferentes reglas de asociación entre la prescripción de hierro, el servicio y la existencia de anemia. Se obtuvo un árbol de decisión que relacionó el valor de hemoglobina del paciente con la vía de administración del hierro prescrito. En un segundo árbol se relacionó el medicamento prescrito según el valor de hemoglobina del paciente y el sexo. Por último, se obtuvieron reglas de asociación entre diferentes principios activos dispensados en los sistemas de dispensación automáticos de los servicios de Cardiología, Neurología, Digestivo y Urología-Nefrología.

Conclusiones: La explotación de datos en Farmacia Hospitalaria resulta esencial para mejorar la calidad de la atención al paciente. En este estudio, el uso de WEKA permitió descubrir patrones ocultos y conocimientos valiosos en los registros de prescripción, dispensación y de datos de laboratorio. Los análisis de asociación y clasificación proporcionaron información útil para optimizar los procesos de dispensación y mejorar la eficiencia y seguridad. La minería de datos es una poderosa herramienta que ayuda a la toma de decisiones en la Farmacia Hospitalaria.

PALABRAS CLAVE: Prescripción electrónica, dispensación de medicamentos, automatización, minería de datos, WEKA.

ABSTRACT

Background: Electronic prescription and automated medication dispensing systems in the hospital environment have improved efficiency and safety in dispensing practices. These processes generate a large amount of data that offer opportunities for analysis and process optimization. Data mining, supported by tools like WEKA, emerges as a solution to extract valuable knowledge from this data in Hospital Pharmacy.

Material and Method: Dispensation data was collected from three sources: reports on unit-dose cart filling, specific medication prescription reports and laboratory data analysis and reports from automated medication dispensing systems. The data was preprocessed using Microsoft Excel. The WEKA program was used for the analysis, applying the Apriori algorithm for association rules and the J48 algorithm for decision tree-based classification rules.

Results: Association patterns were identified between different active principles and therapeutic groups dispensed in a specific service. However, no significant classification relationships were found between the dispensed medications and the service. On the other hand, different association rules were established between iron prescription, the service, and the presence of anemia. A decision tree was generated that related the patient's hemoglobin value to the route of administration of the iron. Another tree linked the prescribed medication according to the patient's hemoglobin value and gender. Finally, association rules were found obtained among different active principles dispensed in the automatic dispensing systems of the Cardiology, Neurology, Digestive, and Urology-Nephrology departments. **Conclusions:** Data exploitation in Hospital Pharmacy is essential for improving the quality of patient care. In this study, the use of WEKA allowed the discovery of hidden patterns and valuable knowledge in prescription, dispensation, and laboratory data records. Association and classification analyses provided useful information to optimize dispensation processes and enhance efficiency and safety. Data mining is a powerful tool that helps decision-making in Hospital Pharmacy.

KEYWORDS: Electronic prescription, medication dispensing, automation, data mining, WEKA.

I. INTRODUCCIÓN

En el hospital, los medicamentos son dispensados a los pacientes ingresados según las órdenes médicas recibidas de forma electrónica (prescripción electrónica). Los médicos utilizan *software* especializado para prescribir directamente en un sistema electrónico. Estas órdenes se envían de manera instantánea y segura al farmacéutico del hospital, quien puede verificar y dispensar los medicamentos necesarios.

La dispensación de medicamentos en dosis unitaria es un componente esencial de la Farmacia Hospitalaria, que busca asegurar una administración segura y eficiente de los medicamentos a los pacientes ingresados. Este proceso implica la preparación y suministro de medicamentos en dosis individuales, adaptadas a las necesidades específicas de cada paciente y de acuerdo con las prescripciones médicas electrónicas. Para llevar a cabo esta tarea, se utilizan dos métodos principales: el uso de carros de unidosis y la implementación de armarios semiautomáticos de dispensación.

El carro de unidosis es un sistema que consiste en la preparación en el Servicio de Farmacia de los medicamentos en dosis unitarias para cada paciente, organizados en un carro móvil que se transporta diariamente a las distintas unidades de hospitalización.

Por otro lado, los armarios automáticos de dispensación son sistemas automatizados ubicados en las unidades de hospitalización, donde se almacenan los medicamentos en forma de dosis unitaria. Estos armarios están controlados electrónicamente y permiten al personal de enfermería retirar la medicación específica de cada paciente, una vez que ha sido prescrita por el médico.

Con la implementación de la prescripción electrónica y la automatización de los procesos de dispensación mediante el uso de sistemas automatizados, se ha logrado una mayor precisión, eficiencia y seguridad en la dispensación de medicamentos en la Farmacia Hospitalaria.

Estos procesos permiten la generación de una gran cantidad de datos. Cada vez que se prescribe o dispensa un medicamento, se registra automáticamente información detallada, como el medicamento suministrado, la dosis, la fecha y la hora, así como los datos del paciente y del médico prescriptor. Estos datos se acumulan rápidamente a medida que se realizan múltiples dispensaciones diarias, creando una valiosa fuente de información digitalizada sobre las prescripciones médicas y el suministro de medicamentos en el entorno hospitalario.

La disponibilidad de esta información ofrece diversas oportunidades para mejorar la atención al paciente y optimizar los procesos de la Farmacia Hospitalaria. Los datos generados pueden ser utilizados para analizar patrones de uso de medicamentos, identificar posibles interacciones medicamentosas y detectar errores potenciales. Esta información contribuye a mejorar la seguridad del paciente, reducir los riesgos de interacciones no deseadas y proporcionar una visión más completa de la terapia farmacológica.

Además, la generación de datos en la dispensación de medicamentos ofrece oportunidades para la optimización de procesos. El análisis de datos puede ayudar a identificar áreas de mejora en la gestión de inventario y la eficiencia en la preparación de dosis unitarias. La información obtenida a partir de estos datos puede respaldar la toma de decisiones informadas para optimizar los recursos y mejorar la eficiencia operativa de la Farmacia Hospitalaria.

La extracción del conocimiento a partir de datos, también conocida como KDD (Knowledge Discovery in Databases), es un proceso que implica descubrir patrones, tendencias, relaciones y conocimientos útiles a partir de grandes conjuntos de datos. Este proceso se divide en tres fases principales: la preparación de los datos o preprocesamiento, el análisis de datos o minería de datos, y la utilización del conocimiento obtenido o post-procesamiento.

La minería de datos es una herramienta para descubrir información oculta en grandes volúmenes de datos. Utiliza diversos algoritmos y técnicas que automatizan la generación de modelos o patrones. Se pueden utilizar diferentes herramientas de minería de datos que existen en el mercado, pero en este trabajo nos centraremos en el programa WEKA, acrónimo de Waikato Environment for Knowledge Analysis. WEKA es un *software* de código abierto que ofrece una amplia gama de algoritmos y técnicas de análisis de datos como clasificación, regresión, agrupación, asociación y visualización de datos, entre otros. Además, su interfaz gráfica y amigable y facilita su uso para personas sin experiencia previa en programación o análisis de datos.

En conclusión, la explotación de los datos en Farmacia Hospitalaria es esencial para mejorar la calidad de la atención y optimizar los procesos de dispensación de medicamentos. La minería de datos, con el apoyo del programa WEKA, ofrece la posibilidad de descubrir patrones ocultos y conocimientos valiosos en los registros de prescripción electrónica y dispensación. A través de este trabajo, buscamos aplicar técnicas de análisis de datos utilizando WEKA para obtener información que contribuya a una mejor toma de decisiones y a la mejora de la atención al paciente en la Farmacia Hospitalaria.

II. OBJETIVOS

Aplicar técnicas de análisis de datos utilizando el programa WEKA para explorar y analizar los datos de dispensación de una Farmacia Hospitalaria. Se espera obtener información valiosa que pueda contribuir a mejorar la eficiencia y seguridad de los procesos de dispensación de medicamentos, lo que tendría un impacto positivo en la calidad de la atención al paciente.

III. MATERIAL Y MÉTODO

Este apartado se dividirá en dos apartados: la obtención y preprocesamiento de los datos, y el análisis de datos.

A. Obtención y preprocesamiento de datos.

Los datos utilizados en este trabajo fueron obtenidos de dos fuentes: del programa de prescripción electrónica de FarmaTools® y del generador de informes de los sistemas automatizados de dispensación Athos™Dosys. Todos los datos fueron descargados y preprocesados utilizando el programa *Microsoft Excel* v.16.74.

Se recopilaron tres tipos de datos:

1. Informes de llenado de carros de unidosis.

Desde el programa de prescripción electrónica de FarmaTools®, se recopilaron los datos diarios de dispensación de todos los medicamentos proporcionados por el Servicio de Farmacia a los pacientes ingresados en las unidades de hospitalización durante un periodo de seis días consecutivos (desde el día 1 al 6 de junio de 2023). Estos datos incluyeron información como la planta de hospitalización, el día de dispensación, el medicamento y la cama. Cada día y cama se consideró un “caso”, representando toda la medicación dispensada ese día para el paciente que ocupaba esa cama determinada. El campo “medicamento” se transformó en principio activo y se agregó la clasificación por grupo terapéutico correspondiente a cada medicamento.

2. Informe de prescripción de un medicamento determinado e informe de analíticas de laboratorio.

Se realizó una recopilación de datos a partir del programa de prescripción electrónica de FarmaTools®. En esta ocasión, además de datos de medicamentos prescritos, también se descargaron datos de analíticas de laboratorio de los pacientes. Se seleccionaron las prescripciones de medicamentos con el principio activo “hierro” en pacientes ingresados durante enero y febrero de 2023 y se relacionaron con el valor de hemoglobina sérica de la primera analítica durante ese ingreso. Las variables obtenidas fueron sexo, vía de administración, pauta, medicamento, servicio clínico y hemoglobina sérica (g/dL). El valor de la hemoglobina se transformó en una nueva variable discreta denominada “anemia”. Se consideró “anemia sí” cuando el valor de la hemoglobina sérica era inferior a 11 g/dL en mujeres y por debajo de 12 g/dL en hombres.

3. Informes de dispensación de sistemas automáticos.

Por último, mediante el generador de informes de los sistemas automatizados de dispensación de Athos™Dosys, se recopilaron los datos de dispensación del período comprendido desde abril a mayo de 2023 de cuatro plantas de hospitalización: Cardiología, Neurología, Digestivo y Urología-Nefrología. Estos datos incluyeron la planta de hospitalización, el día de dispensación, el medicamento y la cama. Los medicamentos fueron clasificados según los principios activos que los componen. Posteriormente, se agruparon los datos en “día-cama” y se consideró un “caso” a toda la medicación dispensada a un paciente determinado en un día en específico. En un segundo procesamiento se tuvieron en cuenta sólo los principios activos más frecuentes (dispensados en más de 100 casos).

B. Análisis de datos.

Para el análisis de datos se utilizó la herramienta WEKA 3.8.6. Se aplicó el algoritmo Apriori, para buscar reglas de asociación, y el algoritmo J48 de árbol de decisión, para detectar reglas de clasificación.

Los **algoritmos de asociación** son herramientas que permiten la búsqueda automatizada de reglas que establecen relaciones entre conjuntos de atributos. Estos algoritmos se basan en el aprendizaje automático no supervisado, lo que significa que no dependen de relaciones predefinidas que sirvan para validar los resultados obtenidos.

El algoritmo Apriori es una herramienta de WEKA que se utiliza para descubrir patrones de asociación en conjuntos de datos. Este algoritmo genera conjuntos frecuentes y, a partir de ellos, crea reglas de asociación. Es importante tener en cuenta que solo permite el análisis de variables discretas. Las variables continuas deben ser discretizadas previamente. Por otro lado, es posible ajustar varios parámetros en el algoritmo, como el soporte mínimo y la confianza mínima, para adaptar los resultados según el nivel de rigor deseado.

El soporte mínimo (“lowerBoundMinSupport” en la interfaz de WEKA) es fundamental en la creación de los conjuntos de datos. Es un parámetro que establece el umbral mínimo de frecuencia requerido para considerar un conjunto de ítems como un conjunto de elementos frecuentes. El soporte mínimo configurado por defecto en el algoritmo Apriori de WEKA es de 0.1, lo que significa que se consideran frecuentes aquellos conjuntos de ítems que aparecen en al menos el 10% de las transacciones del conjunto de datos.

La confianza mínima (“minMetric” en la interfaz de WEKA) es un parámetro que regula la creación de reglas de asociación. Establece el umbral mínimo de confianza requerido para considerar una regla de asociación como válida y relevante. En el algoritmo Apriori, las reglas de asociación obtenidas constan de dos conjuntos de ítems que forman el antecedente y el consecuente. La confianza mínima establece el porcentaje mínimo de veces que el consecuente debe aparecer junto con el antecedente en las transacciones del conjunto de datos. El valor predeterminado en el algoritmo Apriori de WEKA es 0.9, lo que significa que sólo se generarán y mostrarán las reglas de asociación que tengan una confianza igual o superior al 90% (el 90% de las veces que aparezca el antecedente, aparecerá también el consecuente).

Los **algoritmos de clasificación** en WEKA se utilizan para asignar instancias a clases o categorías específicas. Se basan en el aprendizaje automático supervisado, al proporcionarles un conjunto de datos donde cada instancia tiene una clase conocida. A partir de esta información, los algoritmos de clasificación aprenden patrones y características que les permiten realizar predicciones o clasificar nuevas instancias en categorías.

WEKA ofrece varios tipos de algoritmos de clasificación, como regresión logística, k-vecinos más cercanos, redes neuronales, entre otros. En este trabajo se ha utilizado el algoritmo J48, que permite construir un modelo predictivo basado en árboles de decisión a partir de los datos de entrenamiento. El árbol comienza con un nodo raíz que representa todo el conjunto de datos

completo. A continuación, se selecciona un atributo para dividir el conjunto de datos en subconjuntos más pequeños, generando ramas y nodos adicionales en el árbol. Este proceso se repite hasta que se alcanza un nivel de pureza suficiente en los nodos o se ha llegado a una profundidad máxima establecida. El árbol de decisión resultante se utiliza para realizar predicciones sobre nuevas instancias. Siguiendo el camino desde el nodo raíz hasta una hoja del árbol se asigna una clase a cada instancia en función de las reglas aprendidas durante la construcción del árbol.

El algoritmo J48 puede manejar tanto variables discretas como continuas, permitiendo combinar ambos tipos en un mismo árbol de decisión. Para variables continuas utiliza estrategias de partición basadas en umbrales. Calcula umbrales posibles y selecciona el umbral que maximiza la ganancia de información o cualquier otro criterio especificado. Luego, se crean ramas en el árbol de decisión según los umbrales seleccionados.

El algoritmo J48 en WEKA dispone de varios parámetros configurables. Uno de los más importantes es el denominado “confidenceFactor” o factor de confianza para la poda. Establece la confianza mínima requerida para realizar la poda en la construcción del árbol de decisión. Controla la cantidad de error que se está dispuesto a tolerar al realizar una poda. Cuanto menor sea este valor, se requerirá una mejora más significativa para que se realice la poda. El valor por defecto en WEKA es 0,25, lo que indica que se permite una probabilidad de error del 25% para considerar que la mejora es significativa y realizar la poda. Es importante ajustar adecuadamente el factor de confianza en función de los datos y del problema en cuestión. Valores más bajos del factor de confianza pueden generar árboles más pequeños y precisos, pero existe el riesgo de que se realicen demasiadas podas y se pierda información relevante. Valores muy elevados de factor de confianza pueden generar árboles muy grandes y complejos.

Otro parámetro configurable en el algoritmo J48 es el denominado “minNumOb” o número mínimo de hojas para podar el árbol. Si una hoja tiene un número de instancias menor al valor especificado, puede ser podada y reemplazada por una regla o predicción más general. El valor predeterminado para este parámetro en WEKA es 2, que significa que se requieren al menos 2 instancias en un nodo para que se pueda realizar una división. Al igual que el factor de confianza, este valor afectará el tamaño y la complejidad del árbol resultante, así como su capacidad de generalización y precisión en la predicción. Es importante encontrar un equilibrio adecuado según los requisitos y características de tus datos y problema específico.

Estos fueron los tipos de análisis que se hicieron de cada conjunto de datos obtenido:

1. Análisis de los informes de llenado de carros de unidosis.

Para un primer análisis con WEKA se filtró el conjunto original de datos para considerar los principios activos más frecuentes (dispensados más de 100 veces). Se pretende simplificar los datos para encontrar patrones que afecten a los principios activos más frecuentes y así obtener resultados más consistentes. Inicialmente, se seleccionó el algoritmo denominado Apriori. Posteriormente, se realizó un análisis de clasificación, usando la variable “servicio” como categórica. El modelo utilizado en este caso fue J48 (algoritmo de árbol de decisión).

Por último, se aplica el algoritmo Apriori a los datos originales, pero teniendo en cuenta el grupo terapéutico dispensado en vez del principio activo. Se persigue simplificar aún más los datos para la obtención de reglas más relevantes.

2. Análisis de los informes de prescripción de un medicamento determinado e informe de analíticas de laboratorio.

Se realizó un análisis de asociación de los datos mediante el algoritmo Apriori de WEKA, eliminando la variable “anemia”, al ser una variable discreta que no puede procesarse con este algoritmo. Por un lado, se analizaron únicamente los datos de los pacientes con valor de hemoglobina conocido. En un segundo análisis se tuvieron en cuenta todos los pacientes a los que se le prescribió hierro durante su ingreso hospitalario, tuviera o no dato de hemoglobina sérica disponible. En ambos casos se realizaron varios análisis teniendo en cuenta variables diferentes.

Posteriormente, se procedió a buscar reglas de clasificación, mediante el algoritmo J48. Se filtraron únicamente los pacientes con hemoglobina sérica conocida y se utilizaron todas las variables, incluida la hemoglobina. Se procedió a analizar los datos con el algoritmo J48 con los parámetros predefinidos por WEKA. Se hicieron dos análisis usando distintas variables.

3. Análisis de los informes de dispensación de sistemas automáticos.

Con todos los datos globales se aplicaron los algoritmos Apriori y J48. Posteriormente, se analizaron las dispensaciones de cada sistema automatizado por separado, realizándose sendos análisis con el algoritmo Apriori.

IV. RESULTADOS

A continuación, se describen los resultados obtenidos en cada grupo de datos analizados:

IV.1. Resultados del análisis de informes de llenado de carros de unidosis.

Durante los seis días que duró la recopilación de datos se obtuvieron 16.931 registros de dispensación de 506 medicamentos diferentes a pacientes ingresados en 24 unidades de hospitalización. Con la finalidad de simplificar estos datos, se agruparon los medicamentos que contenían el mismo principio activo. De este modo, se obtuvieron 294 principios activos diferentes. Tras una transformación de los datos, obtenemos 2.639 instancias o casos, que corresponden a las dispensaciones de los diferentes principios activos realizadas a cada paciente un día determinado.

Para el análisis con WEKA se seleccionaron los 43 principios activos que fueron dispensados en más de 100 ocasiones. Aplicando el algoritmo Apriori con una confianza mínima del 0.5 y con el resto de los parámetros predeterminados se obtuvieron seis reglas de asociación (ver Figura 1).

```
1. SERVICIO=MIN 410 ==> PANTOPRAZOL=SI 284 <conf:(0.69)> lift:(1.2) lev:(0.02) [46] conv:(1.36)
2. ACETILSALICILICO=SI 462 ==> PANTOPRAZOL=SI 309 <conf:(0.67)> lift:(1.16) lev:(0.02) [41] conv:(1.26)
3. ATORVASTATINA=SI 453 ==> PANTOPRAZOL=SI 292 <conf:(0.64)> lift:(1.11) lev:(0.01) [29] conv:(1.18)
4. ENOXAPARINA=SI 1171 ==> PANTOPRAZOL=SI 741 <conf:(0.63)> lift:(1.09) lev:(0.02) [63] conv:(1.14)
5. FUROSEMIDA=SI 550 ==> PANTOPRAZOL=SI 331 <conf:(0.6)> lift:(1.04) lev:(0) [12] conv:(1.05)
6. OMEPRAZOL=SI 519 ==> ENOXAPARINA=SI 301 <conf:(0.58)> lift:(1.31) lev:(0.03) [70] conv:(1.32)
```

Figura 1. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de llenado de carros de unidosis, teniendo en cuenta las dispensaciones de los principios activos más frecuentes.

Estas reglas se interpretan de la siguiente manera:

- 1) Cuando el servicio es Medicina Interna, se observa una alta probabilidad de que se dispense pantoprazol (confianza 0,69).
- 2) Si se dispensa ácido acetilsalicílico existe una alta probabilidad de que se dispense también pantoprazol (confianza 0,67).
- 3) Cuando se prescribe atorvastatina es probable que también se prescriba pantoprazol (confianza 0,64).
- 4) La prescripción de enoxaparina muestra una alta probabilidad de que se prescriba pantoprazol (confianza 0,63).

- 5) Si se prescribe furosemida, existe una alta probabilidad de que se prescriba pantoprazol (confianza 0,60).
- 6) Cuando se prescribe omeprazol, se observa una alta probabilidad de que esté asociado a enoxaparina (confianza 0.58)

Se observa una alta asociación y probabilidad de que determinados medicamentos estén relacionados entre sí en las prescripciones. En particular, el medicamento pantoprazol está frecuentemente asociado a otros medicamentos en la prescripción y dispensación. Estos hallazgos sugieren posibles patrones de prescripción conjunta que pueden ser útiles para mejorar los procesos de atención médica y farmacoterapia.

En un análisis de clasificación con el algoritmo J48 de árbol de decisión y la variable “servicio” como categórica, el modelo sólo clasificó correctamente el 15.5% de las instancias. Los parámetros utilizados fueron los predefinidos por el programa. El rendimiento obtenido fue muy bajo y no se puede utilizar para realizar predicciones precisas en el conjunto de datos (ver Figura 2).

```

SERVICIO
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===
J48 pruned tree
-----
: MIN (2639.0/2229.0)
Number of Leaves :    1
Size of the tree :    1

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      410           15.5362 %
Incorrectly Classified Instances    2229           84.4638 %
Kappa statistic                     0
Mean absolute error                 0.0843
Root mean squared error             0.2053
Relative absolute error             99.9729 %
Root relative squared error         99.9999 %
Total Number of Instances          2639

```

Figura 2. Datos obtenidos tras la aplicación del algoritmo J48 de WEKA con los parámetros predefinidos al informe de llenado de carros de unidosis, teniendo en cuenta las dispensaciones de los principios activos más frecuentes.

Con los datos originales, sustituyendo la variable “principio activo” por “grupo terapéutico”, se obtienen 32 grupos terapéuticos. Aplicando el algoritmo Apriori, con una confianza mínima del 0.5, el programa nos muestra 5 reglas de asociación (ver figura 3):

```

1. ANTIBIOTICO=SI ANTICOAGULANTE=SI 754 ==> ANTIULCEROSO=SI 691 <conf:(0.92)> lift:(1.16) lev:(0.04) [93] conv:(2.44)
2. ANTICOAGULANTE=SI 1420 ==> ANTIULCEROSO=SI 1275 <conf:(0.9)> lift:(1.13) lev:(0.06) [149] conv:(2.02)
3. DIURETICO=SI 810 ==> ANTIULCEROSO=SI 717 <conf:(0.89)> lift:(1.12) lev:(0.03) [74] conv:(1.79)
4. ANTIBIOTICO=SI 1268 ==> ANTIULCEROSO=SI 1074 <conf:(0.85)> lift:(1.07) lev:(0.03) [68] conv:(1.35)
5. ANTIBIOTICO=SI ANTIULCEROSO=SI 1074 ==> ANTICOAGULANTE=SI 691 <conf:(0.64)> lift:(1.2) lev:(0.04) [113] conv:(1.29)
6. ANTIULCEROSO=SI 2092 ==> ANTICOAGULANTE=SI 1275 <conf:(0.61)> lift:(1.13) lev:(0.06) [149] conv:(1.18)
7. ANTIBIOTICO=SI 1268 ==> ANTICOAGULANTE=SI 754 <conf:(0.59)> lift:(1.11) lev:(0.03) [71] conv:(1.14)
8. ANTIBIOTICO=SI 1268 ==> ANTICOAGULANTE=SI ANTIULCEROSO=SI 691 <conf:(0.54)> lift:(1.13) lev:(0.03) [78] conv:(1.13)
9. ANTICOAGULANTE=SI ANTIULCEROSO=SI 1275 ==> ANTIBIOTICO=SI 691 <conf:(0.54)> lift:(1.13) lev:(0.03) [78] conv:(1.13)
10. ANTICOAGULANTE=SI 1420 ==> ANTIBIOTICO=SI 754 <conf:(0.53)> lift:(1.11) lev:(0.03) [71] conv:(1.11)

```

Figura 3. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de llenado de carros de unidades, clasificando los medicamentos por grupos terapéuticos.

Cinco de estas reglas tuvieron un valor de confianza superior a 0.65 y se interpretan de la siguiente manera:

- 1) Si a un paciente se le dispensa un antibiótico y un anticoagulante, es probable que también tome un antiulceroso (confianza 0.92).
- 2) Si un paciente toma un anticoagulante, es probable que también tome un antiulceroso (confianza 0.90).
- 3) Si un paciente toma un diurético, es probable que también tome un antiulceroso (confianza 0.89).
- 4) Si se dispensa un antibiótico, es probable que al mismo paciente se le dispense también un antiulceroso (confianza 0.85).
- 5) Si un paciente toma un antibiótico y un antiulceroso, es probable que también tome un anticoagulante (confianza 0.64).

En conclusión, clasificando los medicamentos por grupo terapéutico y teniendo en cuenta todas las dispensaciones iniciales, se obtuvo mayor número de reglas de asociación y con una confianza mayor.

IV.2. Resultados del análisis del informe de prescripción de un medicamento determinado e informe de analíticas de laboratorio.

Se registraron 444 pacientes (220 hombres y 224 mujeres) a los que se le prescribió algún medicamento con hierro durante su ingreso hospitalario. De ellos, 96 pacientes no tenían analítica de hemoglobina realizada durante dicho ingreso y 25 pacientes tenían un nivel de hemoglobina superior al considerado para padecer anemia. Es decir, sólo conocemos el dato de hemoglobina de 348 pacientes, y de ellos, el 92% tenía anemia. Los resultados obtenidos en los diferentes tipos de análisis de dato realizado se desarrollan a continuación:

- Algoritmos de asociación.

Sobre los datos de los 348 pacientes con valor de hemoglobina conocido se buscaron reglas de asociación, mediante el algoritmo Apriori de WEKA. Las variables usadas como atributos fueron cinco: sexo, medicamento, pauta, servicio y anemia. De los parámetros predefinidos por el programa sólo se modificó la confianza mínima (0.5). El programa nos devuelve 10 reglas de asociación con confianzas bastantes elevadas (ver figura 4).

```

1. cod_pauta=C48 194 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 193 <conf:(0.99)> lift:(1.72) lev:(0.23) [81] conv:(41.02)
2. cod_pauta=C48 ANEMIA=SI 187 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 186 <conf:(0.99)> lift:(1.72) lev:(0.22) [78] conv:(39.54)
3. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 202 ==> ANEMIA=SI 195 <conf:(0.97)> lift:(1.04) lev:(0.02) [6] conv:(1.73)
4. cod_pauta=C48 194 ==> ANEMIA=SI 187 <conf:(0.96)> lift:(1.03) lev:(0.02) [6] conv:(1.66)
5. cod_pauta=C48 des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 193 ==> ANEMIA=SI 186 <conf:(0.96)> lift:(1.03) lev:(0.02) [6] conv:(1.65)
6. cod_pauta=C48 194 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML ANEMIA=SI 186 <conf:(0.96)> lift:(1.72) lev:(0.22) [77] conv:(9.55)
7. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 202 ==> cod_pauta=C48 193 <conf:(0.96)> lift:(1.72) lev:(0.23) [81] conv:(9)
8. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML ANEMIA=SI 195 ==> cod_pauta=C48 186 <conf:(0.95)> lift:(1.72) lev:(0.22) [77] conv:(8.69)
9. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 202 ==> cod_pauta=C48 ANEMIA=SI 186 <conf:(0.92)> lift:(1.72) lev:(0.22) [78] conv:(5.53)
10. ANEMIA=SI 326 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 195 <conf:(0.6)> lift:(1.04) lev:(0.02) [6] conv:(1.04)

```

Figura 1. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de prescripción de hierro en pacientes con valores de hemoglobina sérica conocidos. Atributos usados: sexo, medicamento, pauta, servicio, anemia.

La mayoría de estas reglas ellas nos relacionaron el medicamento prescrito con su pauta habitual de uso y con la existencia de anemia. Si la pauta era “cada 48 horas”, en el 99% de las veces estaba prescrito el medicamento hierro sacarosa (regla 1). Si estaba prescrito hierro sacarosa, entonces el paciente tenía anemia en el 97% de los casos (regla 3). Otro dato interesante es que, si había anemia, en el 60% de los casos el medicamento prescrito fue hierro sacarosa (regla 6).

En un segundo análisis se prescindió de la variable “pauta” y se aplicó el mismo algoritmo con las mismas características que en el primer análisis. El programa encontró 10 reglas de asociación (ver Figura 5). Por ejemplo, si el paciente era hombre, entonces tenía anemia en el 95% de los casos (regla 4). Pero en las mujeres que tenían prescrito hierro, el porcentaje de pacientes con anemia fue del 91% (regla 5). Si el servicio prescriptor era Medicina Interna, el paciente tuvo anemia en un 91% de los casos en los que se prescribió hierro (regla 6).

```

1. SEXO=H des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 104 ==> ANEMIA=SI 101 <conf:(0.97)> lift:(1.04) lev:(0.01) [4] conv:(1.78)
2. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 202 ==> ANEMIA=SI 195 <conf:(0.97)> lift:(1.04) lev:(0.02) [6] conv:(1.73)
3. SEXO=M des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 98 ==> ANEMIA=SI 94 <conf:(0.96)> lift:(1.03) lev:(0.01) [2] conv:(1.34)
4. SEXO=H 182 ==> ANEMIA=SI 173 <conf:(0.95)> lift:(1.02) lev:(0.01) [3] conv:(1.25)
5. SEXO=M 168 ==> ANEMIA=SI 153 <conf:(0.91)> lift:(0.98) lev:(-0.01) [-3] conv:(0.72)
6. cod_servicio=MIR 118 ==> ANEMIA=SI 107 <conf:(0.91)> lift:(0.97) lev:(-0.01) [-2] conv:(0.67)
7. des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 133 ==> ANEMIA=SI 116 <conf:(0.87)> lift:(0.94) lev:(-0.02) [-7] conv:(0.51)
8. SEXO=M ANEMIA=SI 153 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 94 <conf:(0.61)> lift:(1.06) lev:(0.02) [5] conv:(1.08)
9. ANEMIA=SI 326 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 195 <conf:(0.6)> lift:(1.04) lev:(0.02) [6] conv:(1.04)
10. SEXO=H ANEMIA=SI 173 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 101 <conf:(0.58)> lift:(1.01) lev:(0) [1] conv:(1)

```

Figura 2. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de prescripción de hierro en pacientes con valores de hemoglobina sérica conocidos. Atributos usados: sexo, medicamento, servicio, anemia.

Teniendo en cuenta que la mayoría de pacientes de los que conocíamos el dato de hemoglobina sérica tenía anemia, las reglas de asociación obtenidas siempre estaban relacionadas con esta variable.

Se planteó realizar un análisis de los datos de todos los pacientes a los que se le prescribió hierro (444 instancias), sin tener en cuenta los datos de hemoglona ni si tenían o no anemia. Se aplicó el algoritmo de asociación Apriori seleccionando los atributos sexo, medicamento, pauta y servicio, estableciendo un factor de confianza de 0.9. El programa nos devolvió 10 reglas de asociación (ver Figura 6).

```

1. SEXO=M cod_pauta=C48 129 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 129 <conf:(1)> lift:(1.72) lev:(0.12) [54] conv:(54.21)
2. cod_pauta=DIARIA 165 ==> des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 164 <conf:(0.99)> lift:(2.66) lev:(0.23) [102] conv:(51.72)
3. cod_pauta=C48 248 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 246 <conf:(0.99)> lift:(1.71) lev:(0.23) [102] conv:(34.74)
4. cod_pauta=C48 cod_servicio=MIR 93 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 92 <conf:(0.99)> lift:(1.71) lev:(0.09) [38] conv:(19.54)
5. des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 166 ==> cod_pauta=DIARIA 164 <conf:(0.99)> lift:(2.66) lev:(0.23) [102] conv:(34.82)
6. SEXO=H cod_pauta=C48 119 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 117 <conf:(0.98)> lift:(1.7) lev:(0.11) [48] conv:(16.67)
7. SEXO=M des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 134 ==> cod_pauta=C48 129 <conf:(0.96)> lift:(1.73) lev:(0.12) [54] conv:(9.89)
8. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 258 ==> cod_pauta=C48 246 <conf:(0.95)> lift:(1.71) lev:(0.23) [102] conv:(8.79)
9. SEXO=H des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 124 ==> cod_pauta=C48 117 <conf:(0.94)> lift:(1.69) lev:(0.11) [47] conv:(6.86)
10. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML cod_servicio=MIR 99 ==> cod_pauta=C48 92 <conf:(0.93)> lift:(1.67) lev:(0.08) [36] conv:(5.48)

```

Figura 3. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA al informe de prescripción de hierro, incluyendo todos los pacientes. Atributos usados: sexo, medicamento, pauta, servicio.

Aplicando el mismo algoritmo que en el caso anterior, con los mismos parámetros, pero eliminando el atributo sexo, se obtienen ocho reglas de asociación, que relacionan principalmente el medicamento con la pauta de dosificación más frecuente a la que se ha prescrito (ver Figura 7). Con estos datos se podría valorar la calidad de la prescripción de los fármacos con hierro.

```

1. cod_pauta=DIARIA 165 ==> des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 164 <conf:(0.99)> lift:(2.66) lev:(0.23) [102] conv:(51.72)
2. cod_pauta=C48 248 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 246 <conf:(0.99)> lift:(1.71) lev:(0.23) [102] conv:(34.74)
3. cod_pauta=C48 cod_servicio=MIR 93 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 92 <conf:(0.99)> lift:(1.71) lev:(0.09) [38] conv:(19.54)
4. des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 166 ==> cod_pauta=DIARIA 164 <conf:(0.99)> lift:(2.66) lev:(0.23) [102] conv:(34.82)
5. des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS cod_servicio=MIR 48 ==> cod_pauta=DIARIA 47 <conf:(0.98)> lift:(2.64) lev:(0.07) [29] conv:(15.1)
6. cod_pauta=DIARIA cod_servicio=MIR 48 ==> des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 47 <conf:(0.98)> lift:(2.62) lev:(0.07) [29] conv:(15.05)
7. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 258 ==> cod_pauta=C48 246 <conf:(0.95)> lift:(1.71) lev:(0.23) [102] conv:(8.79)
8. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML cod_servicio=MIR 99 ==> cod_pauta=C48 92 <conf:(0.93)> lift:(1.67) lev:(0.08) [36] conv:(5.48)

```

Figura 4. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA al informe de prescripción de hierro, incluyendo todos los pacientes. Atributos usados: medicamento, pauta, servicio.

Si seleccionamos las variables sexo, medicamento y servicio y aplicamos el algoritmo se Apriori con un factor de confianza de 0.5 se obtienen 10 reglas (ver Figura 8).

```

1. SEXO=M cod_servicio=MIR 76 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 56 <conf:(0.74)> lift:(1.27) lev:(0.03) [11] conv:(1.52)
2. cod_servicio=MIR 152 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 99 <conf:(0.65)> lift:(1.12) lev:(0.02) [10] conv:(1.18)
3. SEXO=M 226 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 134 <conf:(0.59)> lift:(1.02) lev:(0.01) [2] conv:(1.02)
4. SEXO=H 219 ==> des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 124 <conf:(0.57)> lift:(0.98) lev:(-0.01) [-2] conv:(0.96)
5. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML cod_servicio=MIR 99 ==> SEXO=M 56 <conf:(0.57)> lift:(1.11) lev:(0.01) [5] conv:(1.11)
6. des_farma=HIERRO SACAROSA 100 MG AMPOLLA 5 ML 258 ==> SEXO=M 134 <conf:(0.52)> lift:(1.02) lev:(0.01) [2] conv:(1.02)
7. des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 166 ==> SEXO=H 83 <conf:(0.5)> lift:(1.02) lev:(0) [1] conv:(1)
8. des_farma=HIERRO TARDYFERON 80 MG FE COMPRIMIDOS 166 ==> SEXO=M 83 <conf:(0.5)> lift:(0.98) lev:(-0) [-1] conv:(0.97)
9. cod_servicio=MIR 152 ==> SEXO=H 76 <conf:(0.5)> lift:(1.02) lev:(0) [1] conv:(1)
10. cod_servicio=MIR 152 ==> SEXO=M 76 <conf:(0.5)> lift:(0.98) lev:(-0) [-1] conv:(0.97)

```

Figura 5. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de prescripción de hierro, incluyendo todos los pacientes. Atributos usados: sexo, medicamento, servicio.

La regla con la mayor confianza es "si el sexo es masculino y el servicio es Medicina Interna, entonces hay una alta probabilidad (74%) de que se prescriba hierro sacarosa" (regla 1). Otra regla destacada es "si el servicio es Medicina Interna, entonces en el 64% de las veces se prescribe hierro sacarosa" (regla 2).

- Algoritmos de clasificación.

Para la búsqueda de reglas de clasificación entre las diferentes variables se excluyeron los pacientes en los que se desconocía el valor de hemoglobina. Se realizó en el análisis teniendo en cuenta todas las variables: sexo, medicamento, vía de administración, pauta, servicio, anemia y hemoglobina. Se aplicó el algoritmo J48, seleccionando la vía de administración como clase, y manteniendo los parámetros predefinidos por el programa. El resultado obtenido relacionaba cada medicamento con su vía de administración y con su pauta habitual de dosificación.

Realizando el mismo análisis, pero eliminando la variable pauta, es decir, seleccionando los atributos sexo, medicamento, vía de administración, servicio, anemia y hemoglobina y manteniendo la vía de administración como clase, obtuvimos los siguientes resultados (ver Figura 9 y Figura 10):

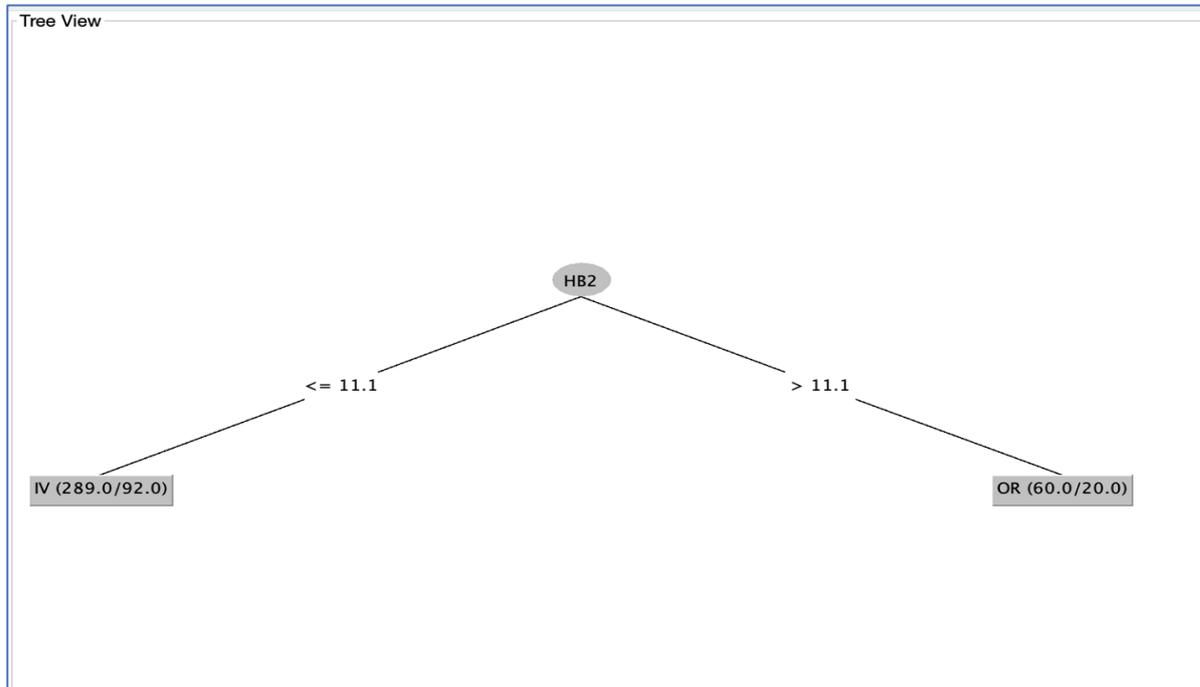


Figura 9. Árbol de decisión resultante de la aplicación del algoritmo J48 de WEKA con los parámetros predefinidos al informe de prescripción de hierro en pacientes con valores de hemoglobina sérica conocidos. Atributos usados: sexo, medicamento, vía de administración, servicio, anemia y hemoglobina (g/dL). Clase: vía de administración. HB2: valor de hemoglobina sérica del paciente (g/dL). IV: vía intravenosa. OR: vía oral.

En el árbol de decisión podemos apreciar que si el valor de hemoglobina fue menor o igual que 11.1 g/dL, entonces el hierro se administró por vía intravenosa. Si la hemoglobina fue mayor, entonces se administró por vía oral. La tasa de acierto fue del 66.5%.

Por último, se seleccionaron las variables sexo, medicamento, anemia y hemoglobina. Se eligió la variable medicamento como clase. Se aplicó de nuevo el algoritmo J48 con los mismos parámetros y se obtuvo un árbol que, con una fiabilidad del 60.6%, selecciona el medicamento prescrito según el valor de hemoglobina sérica, indicando además si el paciente era hombre o mujer (ver Figura 11 y Figura 12).

```

=== Run information ===

Scheme:          weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:        WekaExcel-weka.filters.unsupervised.attribute.Remove-R3-weka.filters.unsupervised.attribute.Remove-R2
Instances:       350
Attributes:      5
                 SEXO
                 codservicio
                 HB2
                 ANEMIA
                 via
Test mode:       10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----
HB2 <= 11.1: IV (289.0/92.0)
HB2 > 11.1: OR (60.0/20.0)

Number of Leaves :    2
Size of the tree :    3

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      232           66.4756 %
Incorrectly Classified Instances    117           33.5244 %
Kappa statistic                    0.2137
Mean absolute error                 0.4293
Root mean squared error            0.4799
Relative absolute error             91.2302 %
Root relative squared error        98.9578 %
Total Number of Instances          349
Ignored Class Unknown Instances     1

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,876   0,682   0,679     0,876   0,765     0,236   0,589    0,691    IV
                0,318   0,124   0,609     0,318   0,418     0,236   0,592    0,459    OR
Weighted Avg.   0,665   0,471   0,652     0,665   0,633     0,236   0,590    0,603

=== Confusion Matrix ===

  a  b  <-- classified as
190 27 | a = IV
 90 42 | b = OR

```

Figura 10. Datos resultantes de la aplicación del algoritmo J48 de WEKA con los parámetros predefinidos al informe de prescripción de hierro en pacientes con valores de hemoglobina sérica conocidos. Atributos usados: sexo, medicamento, vía de administración, servicio, anemia y hemoglobina (g/dL). Clase: vía de administración. HB2: valor de hemoglobina sérica del paciente (g/dL). IV: vía intravenosa. OR: vía oral.

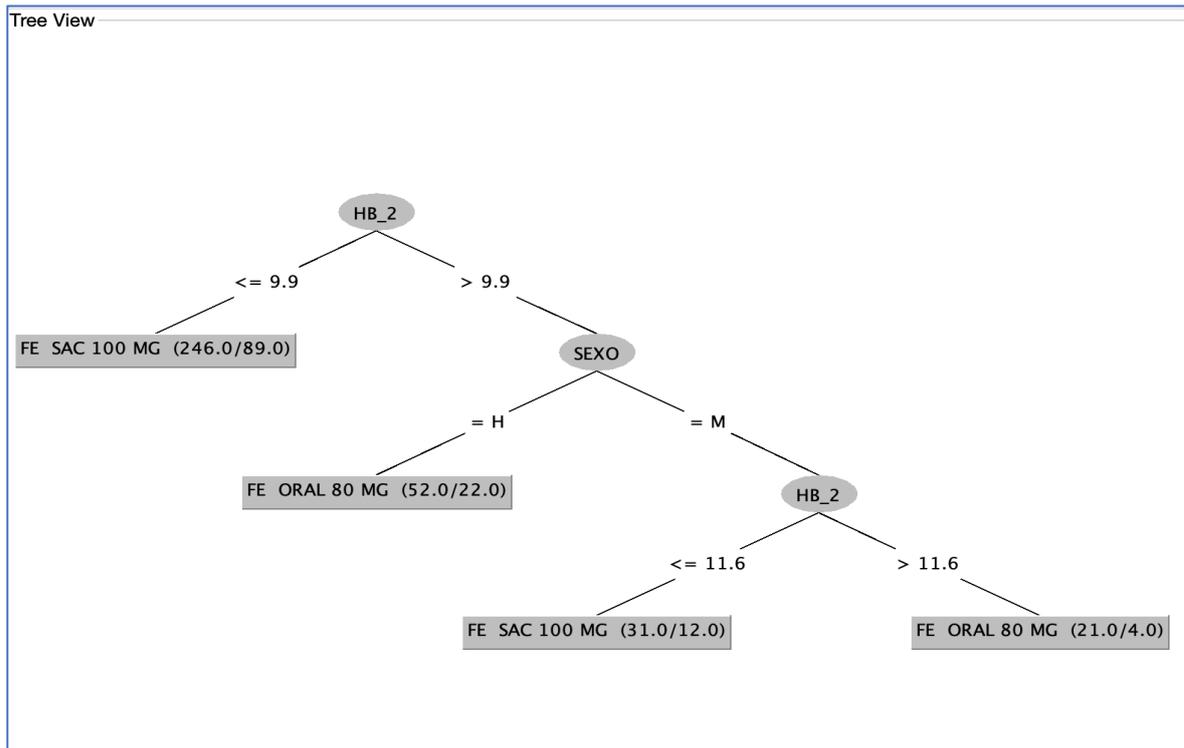


Figura 11. Árbol de decisión resultante de la aplicación del algoritmo J48 de WEKA con los parámetros predefinidos al informe de prescripción de hierro en pacientes con valores de hemoglobina sérica conocidos. Atributos usados: sexo, medicamento, anemia y hemoglobina (g/dL). Clase: medicamento. HB2: valor de hemoglobina sérica del paciente (g/dL). FE SAC: hierro sacarosa. FE ORAL: hierro oral.

Si el el valor de hemoglobina del paciente fue menor o igual que 9.9 g/dL, entonces se prescribió hierro sacarosa. Si ese valor fue mayor de 9.9, se prescribió hierro oral si el paciente era hombre; si el paciente era mujer, entonces se prescribió hierro sacarosa cuando el valor de hemoglobina fue menor o igual que 11.6 g/dL y hierro oral cuando la hemoglobina fue mayor.

```

Scheme:      weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:    WekaExcel
Instances:   350
Attributes:  4
            SEXO
            HB_2
            ANEMIA
            des_farma
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
-----

HB_2 <= 9.9: FE SAC 100 MG (246.0/89.0)
HB_2 > 9.9
| SEXO = H: FE ORAL 80 MG (52.0/22.0)
| SEXO = M
| | HB_2 <= 11.6: FE SAC 100 MG (31.0/12.0)
| | HB_2 > 11.6: FE ORAL 80 MG (21.0/4.0)

Number of Leaves :    4
Size of the tree :    7

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      212           60.5714 %
Incorrectly Classified Instances    138           39.4286 %
Kappa statistic                    0.1492
Mean absolute error                 0.3318
Root mean squared error             0.4115
Relative absolute error             95.3133 %
Root relative squared error        98.7735 %
Total Number of Instances          350

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,000   0,000   ?          0,000   ?          ?        0,528    0,043    FE CARBOX
                0,308   0,143   0,569     0,308   0,400     0,199    0,573    0,489    FE ORAL 80 MG
                0,847   0,723   0,615     0,847   0,712     0,151    0,529    0,580    FE SAC 100 MG
Weighted Avg.   0,606   0,472   ?          0,606   ?          ?        0,546    0,522

=== Confusion Matrix ===

 a  b  c  <-- classified as
0  0  15 | a = FE CARBOX
0  41 92 | b = FE ORAL 80 MG
0  31 171 | c = FE SAC 100 MG

```

Figura 12. Datos resultantes de la aplicación del algoritmo J48 de WEKA con los parámetros predefinidos al informe de prescripción de hierro en pacientes con valores de hemoglobina sérica conocidos. Atributos usados: sexo, medicamento, anemia y hemoglobina (g/dL). Clase: medicamento. HB2: valor de hemoglobina sérica del paciente (g/dL). Des farma: fármaco. FE SAC: hierro sacarosa. FE ORAL: hierro oral. FE CARBOX: hierro carboximaltosa.

IV.3. Resultados del análisis de informes de dispensación de sistemas automáticos.

Se obtuvieron datos de 8.753 casos (dispensaciones por paciente/día) en los que se dispensaron 280 principios activos diferentes. El volumen de datos fue demasiado alto y WEKA no pudo procesarlo por falta de memoria.

De todos los datos obtenidos, se seleccionaron sólo los principios activos dispensados en más de 100 casos, con la intención de obtener menor número de datos. De esta forma, se obtuvieron 8.753 instancias en las que se dispensaron los 117 principios activos más frecuente (ver Figura 13).

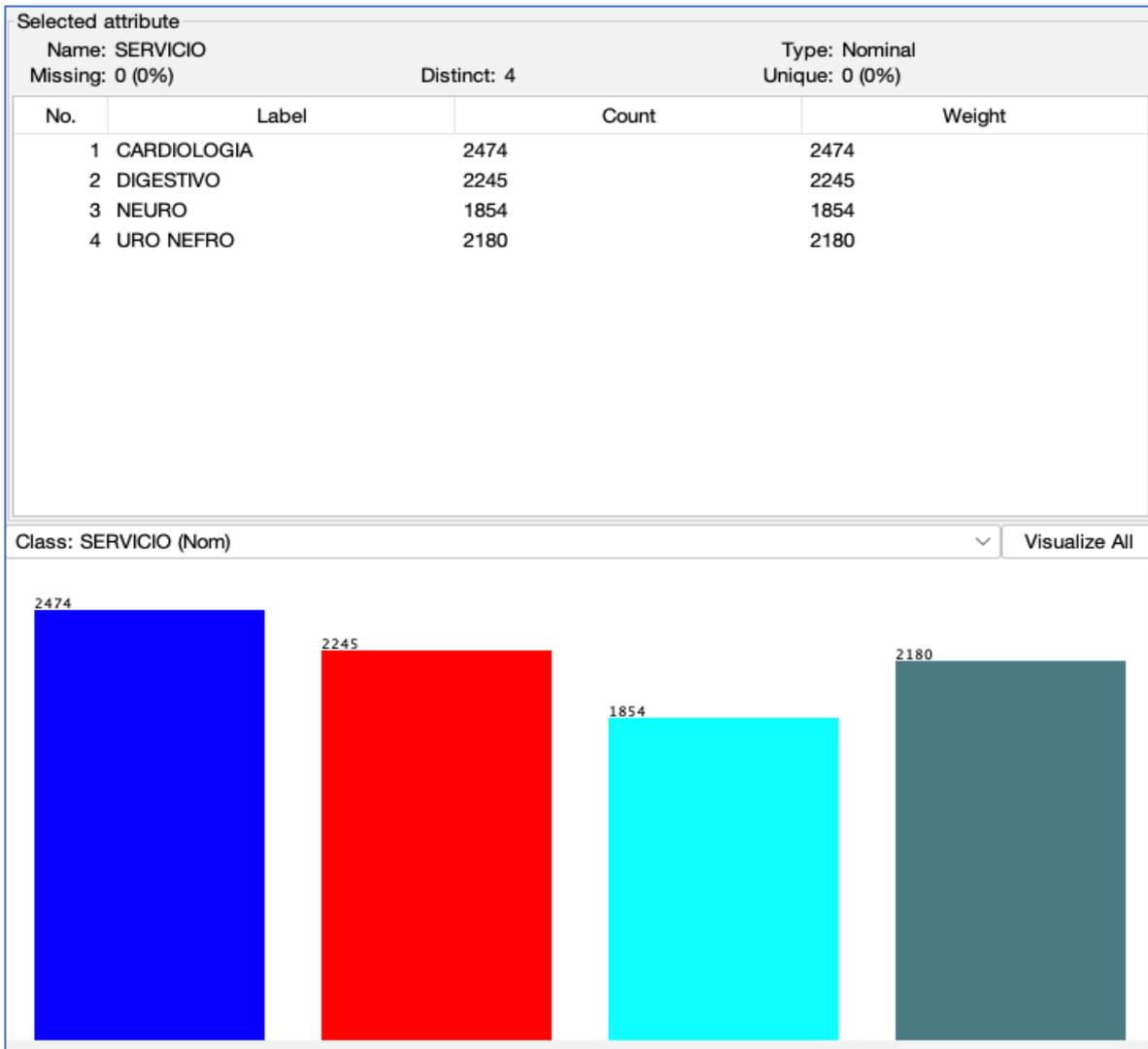


Figura 13. Distribución de dispensaciones de los principios activos más frecuentes según el servicio.

Se procedió a hacer el análisis en WEKA mediante la opción “*associate*”, el algoritmo Apriori y un valor de confianza mínima de 0.5. y se obtuvieron 10 reglas de asociación (ver Figura 14).

1.	METAMIZOL=SI 1392	==>	PANTOPRAZOL=SI 952	<conf:(0.68)>	lift:(1.33)	lev:(0.03)	[237]	conv:(1.54)
2.	METAMIZOL=SI 1392	==>	PARACETAMOL=SI 942	<conf:(0.68)>	lift:(1.84)	lev:(0.05)	[430]	conv:(1.95)
3.	BISOPROLOL=SI 1519	==>	SERVICIO=CARDIOLOGIA 1025	<conf:(0.67)>	lift:(2.39)	lev:(0.07)	[595]	conv:(2.2)
4.	AC ACETILSALICILICO=SI 1935	==>	ATORVASTATINA=SI 1271	<conf:(0.66)>	lift:(2.44)	lev:(0.09)	[749]	conv:(2.13)
5.	AC ACETILSALICILICO=SI 1935	==>	PANTOPRAZOL=SI 1202	<conf:(0.62)>	lift:(1.21)	lev:(0.02)	[208]	conv:(1.28)
6.	PARACETAMOL=SI 3219	==>	PANTOPRAZOL=SI 1908	<conf:(0.59)>	lift:(1.15)	lev:(0.03)	[255]	conv:(1.19)
7.	SERVICIO=DIGESTIVO 2245	==>	PANTOPRAZOL=SI 1321	<conf:(0.59)>	lift:(1.15)	lev:(0.02)	[168]	conv:(1.18)
8.	SERVICIO=URO NEFRO 2180	==>	PANTOPRAZOL=SI 1274	<conf:(0.58)>	lift:(1.14)	lev:(0.02)	[154]	conv:(1.17)
9.	FUROSEMIDA=SI 1885	==>	PANTOPRAZOL=SI 1089	<conf:(0.58)>	lift:(1.13)	lev:(0.01)	[121]	conv:(1.15)
10.	E0XAPARINA=SI 2923	==>	PANTOPRAZOL=SI 1668	<conf:(0.57)>	lift:(1.11)	lev:(0.02)	[167]	conv:(1.13)

Figura 14. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de dispensación de medicamentos de los sistemas automatizados, teniendo en cuenta los principios activos más frecuentes.

Se establecieron relaciones de medicamentos que se prescribían juntos con bastante frecuencia, como metamizol y pantoprazol, metamizol y paracetamol, ácido acetilsalicílico y atorvastatina, ácido acetilsalicílico con pantoprazol, paracetamol y pantoprazol. Por otro lado, se asoció la prescripción de bisoprolol al servicio de Cardiología (en un 67% de las veces).

Por último, se probó a usar un algoritmo de asociación (J48) con la variable categórica “servicio” y se obtuvo un árbol muy complejo, imposible de interpretar.

Teniendo en cuenta únicamente los datos de dispensación del servicio de **Cardiología**, se analizaron 2.561 dispensaciones de todos los principios activos (201 en total). Se aplicó el algoritmo Apriori con una confianza mínima de 0.5 y se obtuvo una serie de reglas de asociación (ver Figura 15).

1. ACETILSALICILICO=SI BISOPROLOL=SI 535 ==> ATORVASTATINA=SI 431 <conf:(0.81)> lift:(1.83) lev:(0.08) [195] conv:(2.85)
2. ACETILSALICILICO=SI 941 ==> ATORVASTATINA=SI 667 <conf:(0.71)> lift:(1.61) lev:(0.1) [252] conv:(1.91)
3. ATORVASTATINA=SI BISOPROLOL=SI 612 ==> ACETILSALICILICO=SI 431 <conf:(0.7)> lift:(1.92) lev:(0.08) [206] conv:(2.13)
4. ACETILSALICILICO=SI ATORVASTATINA=SI 667 ==> BISOPROLOL=SI 431 <conf:(0.65)> lift:(1.59) lev:(0.06) [160] conv:(1.67)
5. ENOXAPARINA=SI 727 ==> ATORVASTATINA=SI 431 <conf:(0.59)> lift:(1.34) lev:(0.04) [110] conv:(1.37)
6. ATORVASTATINA=SI 1129 ==> ACETILSALICILICO=SI 667 <conf:(0.59)> lift:(1.61) lev:(0.1) [252] conv:(1.54)
7. BISOPROLOL=SI 1038 ==> ATORVASTATINA=SI 612 <conf:(0.59)> lift:(1.34) lev:(0.06) [154] conv:(1.36)
8. ACETILSALICILICO=SI 941 ==> BISOPROLOL=SI 535 <conf:(0.57)> lift:(1.4) lev:(0.06) [153] conv:(1.37)
9. ATORVASTATINA=SI 1129 ==> BISOPROLOL=SI 612 <conf:(0.54)> lift:(1.34) lev:(0.06) [154] conv:(1.3)
10. PANTOPRAZOL=SI 939 ==> ACETILSALICILICO=SI 505 <conf:(0.54)> lift:(1.46) lev:(0.06) [159] conv:(1.37)

Figura 15. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y confianza mínima de 0.5 al informe de dispensación de medicamentos del sistema automatizados del servicio de Cardiología.

Algunas de ellas se interpretan de la siguiente forma:

1. La regla con mayor confianza (0,81) indica que cuando se prescribe ácido acetilsalicílico y bisoprolol, hay una alta probabilidad de que también se prescriba atorvastatina.
2. La regla con una confianza del 0.71 muestra que cuando se prescribe ácido acetilsalicílico, es probable que se prescriba también atorvastatina.
3. La regla con una confianza del 0.7 indica que cuando se prescriben atorvastatina y bisoprolol, es probable que también se prescriba ácido acetilsalicílico.

Estas combinaciones de fármacos son usadas habitualmente en las diferentes patologías cardiológicas, respaldadas por las guías clínicas.

En el análisis de los datos de dispensación del sistema automatizado del servicio de **Neurología** se tuvieron en cuenta sólo los principios activos prescritos más de 100 veces durante el período de estudio. Se obtuvieron 1.850 dispensaciones de 34 principios activos. Se aplicó el algoritmo Apriori con los parámetros definidos y se obtuvieron 10 reglas de asociación (ver Figura 16).

1.	AC ACETILSALICILICO=SI E?XAPARINA=SI 373 ==>	ATORVASTATINA=SI 297	<conf:(0.8)> lift:(2.19) lev:(0.09) [161] conv:(3.09)
2.	AC ACETILSALICILICO=SI PANTOPRAZOL=SI 395 ==>	ATORVASTATINA=SI 300	<conf:(0.76)> lift:(2.09) lev:(0.08) [156] conv:(2.62)
3.	ATORVASTATINA=SI PANTOPRAZOL=SI 411 ==>	AC ACETILSALICILICO=SI 300	<conf:(0.73)> lift:(2.31) lev:(0.09) [170] conv:(2.51)
4.	AC ACETILSALICILICO=SI 585 ==>	ATORVASTATINA=SI 426	<conf:(0.73)> lift:(2.01) lev:(0.12) [213] conv:(2.33)
5.	ATORVASTATINA=SI E?XAPARINA=SI 413 ==>	AC ACETILSALICILICO=SI 297	<conf:(0.72)> lift:(2.28) lev:(0.09) [166] conv:(2.42)
6.	AC ACETILSALICILICO=SI ATORVASTATINA=SI 426 ==>	PANTOPRAZOL=SI 300	<conf:(0.7)> lift:(1.33) lev:(0.04) [74] conv:(1.58)
7.	AC ACETILSALICILICO=SI ATORVASTATINA=SI 426 ==>	E?XAPARINA=SI 297	<conf:(0.7)> lift:(1.37) lev:(0.04) [80] conv:(1.61)
8.	AC ACETILSALICILICO=SI 585 ==>	PANTOPRAZOL=SI 395	<conf:(0.68)> lift:(1.27) lev:(0.05) [84] conv:(1.44)
9.	AC ACETILSALICILICO=SI 585 ==>	E?XAPARINA=SI 373	<conf:(0.64)> lift:(1.25) lev:(0.04) [75] conv:(1.35)
10.	ATORVASTATINA=SI 673 ==>	AC ACETILSALICILICO=SI 426	<conf:(0.63)> lift:(2.01) lev:(0.12) [213] conv:(1.86)

Figura 16. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA al informe de dispensación de principios activos más frecuentes del sistema automatizados del servicio de Neurología.

Efectivamente, el uso de ácido acetilsalicílico, enoxaparina y atorvastatina es muy habitual en Neurología, por ejemplo, para el tratamiento de la enfermedad cerebrovascular.

En el caso del servicio de Digestivo, se realizaron 2.159 dispensaciones de los 42 principios activos más frecuentes (dispensados más de 100 veces). Se aplicó el algoritmo Apriori con los parámetros definidos y una confianza mínima de 0.5. Se obtuvieron 9 reglas de asociación (ver Figura 17).

1.	METAMIZOL=SI PANTOPRAZOL=SI 313 ==>	PARACETAMOL=SI 229	<conf:(0.73)> lift:(1.76) lev:(0.05) [99] conv:(2.16)
2.	FUROSEMIDA=SI 507 ==>	PANTOPRAZOL=SI 370	<conf:(0.73)> lift:(1.2) lev:(0.03) [60] conv:(1.43)
3.	METAMIZOL=SI PARACETAMOL=SI 328 ==>	PANTOPRAZOL=SI 229	<conf:(0.7)> lift:(1.14) lev:(0.01) [28] conv:(1.28)
4.	E?XAPARINA=SI PARACETAMOL=SI 327 ==>	PANTOPRAZOL=SI 226	<conf:(0.69)> lift:(1.13) lev:(0.01) [26] conv:(1.25)
5.	METAMIZOL=SI 476 ==>	PARACETAMOL=SI 328	<conf:(0.69)> lift:(1.66) lev:(0.06) [130] conv:(1.87)
6.	METAMIZOL=SI 476 ==>	PANTOPRAZOL=SI 313	<conf:(0.66)> lift:(1.08) lev:(0.01) [22] conv:(1.13)
7.	PARACETAMOL=SI 895 ==>	PANTOPRAZOL=SI 568	<conf:(0.63)> lift:(1.04) lev:(0.01) [22] conv:(1.06)
8.	E?XAPARINA=SI 706 ==>	PANTOPRAZOL=SI 447	<conf:(0.63)> lift:(1.04) lev:(0.01) [16] conv:(1.06)
9.	E?XAPARINA=SI PANTOPRAZOL=SI 447 ==>	PARACETAMOL=SI 226	<conf:(0.51)> lift:(1.22) lev:(0.02) [40] conv:(1.18)

Figura 17. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y una confianza mínima del 0.5 al informe de dispensación de principios activos más frecuentes del sistema automatizados del servicio de Digestivo.

En el servicio de **Urología-Nefrología** se registraron 2.140 dispensaciones de los 43 principios activos más frecuentes (dispensados más de 100 veces). Aplicando el algoritmo de asociación Apriori con los parámetros definidos y modulando la confianza mínima al 0.5 se obtuvieron 10 reglas (ver Figura 18).

```

1. METAMIZOL=SI PANTOPRAZOL=SI 391 ==> PARACETAMOL=SI 318 <conf:(0.81)> lift:(1.65) lev:(0.06) [124] conv:(2.67)
2. METAMIZOL=SI 522 ==> PARACETAMOL=SI 417 <conf:(0.8)> lift:(1.62) lev:(0.07) [158] conv:(2.49)
3. AC ACETILSALICILICO=SI 286 ==> PANTOPRAZOL=SI 219 <conf:(0.77)> lift:(1.28) lev:(0.02) [48] conv:(1.69)
4. METAMIZOL=SI PARACETAMOL=SI 417 ==> PANTOPRAZOL=SI 318 <conf:(0.76)> lift:(1.28) lev:(0.03) [68] conv:(1.68)
5. METAMIZOL=SI 522 ==> PANTOPRAZOL=SI 391 <conf:(0.75)> lift:(1.25) lev:(0.04) [79] conv:(1.59)
6. ENXAPARINA=SI PARACETAMOL=SI 297 ==> PANTOPRAZOL=SI 218 <conf:(0.73)> lift:(1.23) lev:(0.02) [40] conv:(1.5)
7. ENXAPARINA=SI 557 ==> PANTOPRAZOL=SI 391 <conf:(0.7)> lift:(1.18) lev:(0.03) [58] conv:(1.34)
8. FUROSEMIDA=SI 343 ==> PANTOPRAZOL=SI 234 <conf:(0.68)> lift:(1.14) lev:(0.01) [29] conv:(1.26)
9. AMLODIPIN=SI 534 ==> PANTOPRAZOL=SI 347 <conf:(0.65)> lift:(1.09) lev:(0.01) [28] conv:(1.14)
10. PARACETAMOL=SI 1058 ==> PANTOPRAZOL=SI 674 <conf:(0.64)> lift:(1.07) lev:(0.02) [42] conv:(1.11)

```

Figura 18. Mejores reglas encontradas aplicando el algoritmo Apriori con parámetros predefinidos por WEKA y una confianza mínima del 0.5 al informe de dispensación de principios activos más frecuentes del sistema automatizados del servicio de Urología-Nefrología.

V. DISCUSIÓN

En primer lugar, el análisis de los informes de llenado de carros de unidosis reveló patrones interesantes en la dispensación de los principios activos más frecuentes. El algoritmo Apriori permitió descubrir reglas de asociación entre medicamentos, lo que proporcionó información valiosa sobre las combinaciones comunes de fármacos que se dispensan juntos. Estas reglas pueden ayudar a revisar posibles interacciones medicamentosas y optimizar el proceso de dispensación desde la farmacia, por ejemplo, al de colocar esos medicamentos en las estanterías o armarios semiautomáticos de dispensación. Sin embargo, el análisis de clasificación utilizando la variable "servicio" no logró identificar patrones de prescripción específicos para cada área clínica. Sería interesante probar con otros algoritmos u otros enfoques de procesamiento de los datos para encontrar esas relaciones, que podrían ser útiles para personalizar las estrategias de dispensación y mejorar la eficiencia en la preparación de dosis unitarias.

En segundo lugar, el análisis de los informes de prescripción de hierro reveló una relación significativa entre cada medicamento y su pauta habitual de dosificación. Además, en la gran mayoría de los casos en que se prescribió hierro, el paciente tenía anemia. De esta forma, se puede afirmar que la calidad de la prescripción en cuanto a la pauta de dosificación y la indicación del medicamento fue óptima, incluso permitiendo comparaciones por servicios o según el sexo de los pacientes. Por tanto, estas reglas de asociación podrían usarse como base para tomar decisiones que mejoren la atención a los paciente y en el desarrollo de herramientas de ayuda a la prescripción y/o validación de órdenes médicas. Además, mediante el algoritmo de clasificación J48 se estableció una relación entre el valor de hemoglobina del paciente y la vía de administración del medicamento. En niveles más bajos de hemoglobina se utilizó la vía intravenosa, más efectiva que la oral. Por otro lado, se estableció una relación entre el valor de hemoglobina, el sexo del paciente y el medicamento utilizado. En hombres, cuando el valor de hemoglobina era igual o menor de 9,9 g/dL,

entonces el medicamento más probable a prescribir fue el hierro sacarosa (intravenoso). Por encima de ese valor, se prescribió de forma más frecuente hierro oral. En mujeres, el punto de corte estuvo en valores de hemoglobina de 11.6 g/dL. Estos hallazgos parecen no ser coherentes con los valores que definen la anemia (en hombres por debajo de 13 g/dL y en mujeres por debajo de 12 g/dL). Esto indica que la utilización de los resultados de laboratorio en la validación de órdenes médicas por parte del farmacéutico puede ser una herramienta muy útil. Este hallazgo destaca la importancia de integrar los sistemas de prescripción electrónica y de laboratorio para mejorar la monitorización de los pacientes y optimizar la terapia farmacológica.

Por último, el análisis de los informes de dispensación de los sistemas automáticos, al igual que ocurrió con el análisis de los datos carros de unidosis, reveló patrones de asociaciones frecuentes de medicamentos. Al analizar cada servicio por separado, la información obtenida es aún más interesante, al descubrir patrones de prescripción en diferentes áreas de hospitalización. Esto tiene una utilidad doble:

1. Por un lado, podría ser útil en el diseño de protocolos que sirvan de ayuda a la prescripción electrónica. Los protocolos de prescripción electrónica asistida son herramientas desarrolladas para ayudar a los médicos a prescribir los medicamentos adecuados mediante el uso de la tecnología. Están basados en guías clínicas, pero también utilizan bases de datos y algoritmos clínicos para proporcionar recomendaciones basadas en la evidencia científica actual.
2. Permite optimizar la selección y disposición de los medicamentos en el dispensador, teniendo en cuenta los medicamentos que más se dispensan y la frecuencia con la que se dispensan juntos varios medicamentos. Con esto conseguimos optimizar el tiempo que el personal sanitario dedica a la dispensación, y se colabora en la reducción de errores, ya que los medicamentos que se dispensan juntos estarán ubicados en proximidad física en el dispensador.

En general, los resultados obtenidos a través del análisis de los datos de dispensación de la farmacia hospitalaria utilizando WEKA son prometedores, pero es importante destacar que este estudio tiene algunas limitaciones: unas propias del estudio, como que los datos utilizados se limitaron a un corto período de tiempo, y otras propias del software WEKA. Por ejemplo, esta herramienta puede presentar problemas para manejar conjuntos de datos muy grandes debido a limitaciones de memoria. En general, es una herramienta robusta usada ampliamente en investigación en salud, quizás por su flexibilidad y facilidad de uso, pero en futuras investigaciones se recomienda utilizar una herramienta más potente que permita procesar los numerosos datos que se generan en una

farmacia de hospital. Además, se debería ampliar el período de estudio y considerar la inclusión de más datos. En última instancia, el uso de herramientas de minería de datos en la farmacia hospitalaria tiene el potencial de mejorar significativamente la atención al paciente, la eficiencia operativa y los resultados clínicos.

Este estudio supone una pequeña aproximación a las posibles utilidades que puede tener la minería de datos con la información obtenida de los datos de prescripción, dispensación y analíticas de laboratorio. Esto corrobora que, en el ámbito de la salud, la minería de datos desempeña un papel fundamental debido a la gran cantidad de información que se genera en los hospitales y sistemas de atención médica. En el campo de la Farmacia Hospitalaria, podría tener numerosas aplicaciones:

1. Identificación de patrones de prescripción: Permite detectar tendencias en las prescripciones, como la frecuencia de uso de medicamentos, combinaciones comunes e incluso patrones de prescripción inapropiados.
2. Detección de interacciones medicamentosas: Mediante el análisis de los registros de prescripción y dispensación permite detectar posibles interacciones entre medicamentos, lo que podría contribuir a mejorar la seguridad del paciente.
3. Optimización de la dispensación de medicamentos: La minería de datos puede utilizarse para optimizar la dispensación de medicamentos en el hospital. Analizando los datos de dispensación, se pueden identificar los medicamentos más solicitados. Esto ayuda a mejorar la eficiencia en la distribución y abastecimiento de medicamentos, evitando demoras y asegurando un suministro adecuado.
4. Análisis de resultados terapéuticos: Al combinar los datos de prescripción y los resultados clínicos de los pacientes, la minería de datos puede ayudar a analizar la efectividad de diferentes tratamientos farmacológicos. Esto permite identificar qué medicamentos o combinaciones de medicamentos tienen un mayor impacto en los resultados terapéuticos, lo que puede guiar las decisiones de prescripción y mejorar la calidad de la atención médica.
5. Ayuda a la validación de las órdenes médicas por parte del farmacéutico de hospital. La minería de datos puede ser una herramienta valiosa para el farmacéutico en la validación de órdenes

médicas. Ayuda a detectar errores de prescripción, verificar dosis y posologías, identificar interacciones medicamentosas, evaluar alergias y contraindicaciones, y encontrar medicamentos equivalentes.

Como conclusión, el uso de WEKA permitió descubrir patrones ocultos y conocimientos valiosos en los registros de prescripción, dispensación y de datos de laboratorio. Los análisis de asociación y clasificación proporcionaron información útil para optimizar los procesos de dispensación y mejorar la eficiencia y seguridad. Con este trabajo se demuestra que la minería de datos es una poderosa herramienta que ayuda a la toma de decisiones en la Farmacia Hospitalaria, ayudando a garantizar una dispensación segura y eficiente de los medicamentos.

