



# UNIVERSIDAD DE GRANADA

Programa de Doctorado en Tecnologías  
de la Información y la Comunicación

*Human Being Language-Based Rule Extensions  
for Interpretable and Transparent Data Modeling*

Tesis Doctoral

Carmen Biedma Rodríguez

Granada, 15 de Mayo de 2023

Departamento de Ciencias de la Computación  
e Inteligencia Artificial

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Carmen Biedma Rodríguez  
ISBN: 978-84-1117-003-9  
URI: <https://hdl.handle.net/10481/84468>



# UNIVERSIDAD DE GRANADA

Programa de Doctorado en Tecnologías  
de la Información y la Comunicación

*Human Being Language-Based Rule Extensions  
for Interpretable and Transparent Data Modeling*

MEMORIA QUE PRESENTA

Carmen Biedma Rodríguez

PARA OPTAR A DOCTOR POR LA UNIVERSIDAD DE GRANADA  
dentro del Programa de Doctorado en Tecnologías de la Información y la Comunicación

Mayo de 2023

DIRECTORES

**Jesús Alcalá Fernández**  
**María José Gacto Colorado**

Departamento de Ciencias de la Computación  
e Inteligencia Artificial

La doctoranda Carmen Biedma Rodríguez y los directores de la tesis Jesús Alcalá Fernández y María José Gacto Colorado:

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por la doctoranda bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Granada, 15 de Mayo de 2023

La Doctoranda

Fdo: Dña. Carmen Biedma Rodríguez

El Director

La Directora

Fdo: Dr. Jesús Alcalá Fernández

Fdo: Dra. María José Gacto Colorado

Esta tesis ha sido financiada por el Fondo Europeo de Desarrollo Regional (FEDER) y por la Consejería de Transformación Económica, Industria, Conocimiento y Universidades de la Junta de Andalucía bajo el proyecto titulado “*Explicabilidad de la Inteligencia Artificial para el Análisis Inteligente de Datos: Aplicaciones en Problemas de BioSalud y del Internet de las Cosas*”, con código de proyecto: P18-RT-2248.



# Agradecimientos

Esta tesis va dedicada con todo mi cariño a aquellas personas que han contribuido, de una forma u otra, a la realización de la misma. En especial a mi familia y amigos, sois mi gran apoyo y mi motivación constante.

Gracias mamá y papá por ser mis referentes en la vida, por estar siempre a mi lado y por haber conseguido que sea lo que soy hoy en día. Gracias a mi hermana por mirarme siempre con admiración y darme el cariño incondicional que me hace seguir con ánimo en muchos momentos difíciles, estoy segura de que llegarás todo lo lejos que te propongas. Gracias abuelos, titos y primos, especialmente a mi tita Loles por sus valiosos consejos y a mi abuela Loli por su ternura.

A mis directores Jesús y María José, gracias por vuestros conocimientos y por toda la ayuda, paciencia y apoyo que me habéis regalado.

Por último, quiero dar las gracias a todos mis amigos y compañeros de trabajo, en especial a la que siempre está cerca de mí, estemos lo lejos que estemos, gracias Zule.

GRACIAS A TODOS

# Resumen

La Inteligencia Artificial (IA) ha experimentado en los últimos años un gran auge debido a la capacidad predictiva de los modelos generados. Sin embargo, los avances que se han logrado han generado preocupación por la falta de transparencia y explicabilidad en la toma de decisiones de los sistemas de IA, lo que se conoce como “problema de la caja negra”.

Esta falta de transparencia y explicabilidad es especialmente preocupante en ámbitos de alto riesgo, como la salud, donde las decisiones basadas en los sistemas de IA pueden tener un impacto significativo en la vida de las personas. Por ello, la Unión Europea ha desarrollado una propuesta de Ley para la IA y ha establecido unas directrices para promover un desarrollo fiable, que garantice una IA legal, ética y robusta <https://artificialintelligenceact.eu>. Para que las personas puedan depositar su confianza en los sistemas de IA y aprovechar plenamente todos los beneficios que ofrecen, es fundamental que tanto la tecnología en sí, como las personas y los procesos involucrados, sean fiables y transparentes.

Entre los mecanismos disponibles para desarrollar una IA fiable se encuentra el campo de investigación de la IA explicable (conocido por las siglas en inglés XAI, “eXplainable Artificial Intelligence”). La XAI fomenta el uso y desarrollo de técnicas de IA y de Machine Learning (ML) que proporcionen modelos fácilmente entendibles por los humanos o el uso de técnicas post-hoc que nos faciliten explicaciones de cómo un modelo complejo realiza sus predicciones y qué características del problema está considerando.

Entre ellos, el aprendizaje de Sistemas Basados en Reglas Difusas (SBRDs) ha sido ampliamente utilizado para modelar una gran cantidad de problemas. Los modelos generados por estos sistemas son transparentes y comprensibles por sí mismos, ya que constan de un conjunto de reglas simples del tipo “Si-Entonces” formuladas verbalmente sobre dominios imprecisos. Sin embargo, estos métodos suelen constar de un gran número de reglas (co-

bertura) que suelen contener un gran número de variables (especificidad). Un aumento de la cobertura y la especificidad del modelo suele suponer una mejora de su precisión pero compromete la intención de utilizar estos modelos para que el usuario pueda analizarlos y comprenderlos.

Una forma directa de abordar este problema es mediante la sinergia positiva que existe entre los SBRDs y los Algoritmos Evolutivos (AE), conocidos como Sistemas Difusos Evolutivos (SDEs). Los SDEs pueden preservar la esencia original de la comprensibilidad de los sistemas difusos y mejorar su capacidad predictiva al ser capaces de considerar la transparencia y comprensibilidad del modelo en el proceso de aprendizaje. Debido a ello, el objetivo general de esta tesis consiste en diseñar y desarrollar nuevos métodos evolutivos que, haciendo uso de extensiones o de nuevas representaciones de reglas difusas, nos permitan obtener SBRDs precisos, transparentes y comprensibles por los humanos para distintos tipos de problemas (clasificación, regresión y descubrimiento de conocimiento).

Para ello, en primer lugar, en esta tesis se ha propuesto una extensión de la representación clásica de regla difusa que permite representar dependencias positivas y negativas para realizar un aprendizaje evolutivo de clasificadores asociativos con un buen equilibrio entre complejidad y precisión. Por otro lado, se ha presentado una nueva extensión de los SBRDs específicamente diseñada para problemas de regresión, combinando una nueva gramática difusa lingüística y una nueva extensión lineal interpretable para obtener el mayor nivel de transparencia posible. Además, se ha propuesto un nuevo algoritmo evolutivo multi-objetivo basado en árboles de decisión, que hace uso de la nueva representación de reglas difusas y que permite obtener regresores con un buen equilibrio entre precisión y transparencia. Por último, se ha presentado una nueva extensión de la representación de reglas de asociación difusas haciendo uso de la nueva gramática difusa lingüística propuesta para regresión. A partir de esta extensión, se ha diseñado una variante de un método clásico de extracción de reglas de asociación difusas para extraer reglas que utilicen la nueva representación de regla. Las reglas obtenidas permiten resumir parte de la información obtenida con la estructura de regla original e identificar relaciones nuevas que no pueden ser representadas con la gramática clásica.

Además, la transparencia de los modelos generados para clasificación y regresión ha sido validada sobre un problema real biosanitario relacionado con la obesidad infantil. En este estudio los expertos han podido analizar los modelos y han identificado relaciones interesantes entre las variables, que no



se conocían y que actualmente están analizando en laboratorio.

# Índice

|   |          |
|---|----------|
| <b>1. Introducción</b>  | <b>1</b> |
| A Planteamiento . . . . .   | 1        |
| B Objetivos . . . . .   | 6        |
| C Estructura de la memoria . . . . .  | 7        |
| <b>2. Antecedentes</b>  | <b>9</b> |
| 2.1. Inteligencia Artificial Fiable . . . . .                               | 9        |
| 2.1.1. Principios éticos . . . . .  | 10       |
| 2.1.2. Requisitos de una IA fiable . . . . .                                | 11       |
| 2.1.3. Métodos técnicos . . . . .   | 14       |
| 2.2. Inteligencia Artificial Explicable . . . . .                           | 16       |
| 2.2.1. Explicabilidad en el contexto de Machine Learning . . . . .          | 16       |
| 2.2.2. Niveles de transparencia en ML . . . . .                             | 17       |
| 2.2.3. Modelos de ML transparentes . . . . .                                | 18       |
| 2.3. Sistemas Difusos Evolutivos para la XAI . . . . .                      | 20       |
| 2.3.1. Componentes de los Sistemas Basados en Reglas Difusas . . . . .      | 21       |
| 2.3.2. Mejora de un SBRD desde la precisión y la comprensibilidad . . . . . | 23       |
| 2.3.3. Medidas de Interpretabilidad . . . . .                               | 26       |
| 2.4. Problema biomédico real sobre obesidad en niños . . . . .              | 31       |
| 2.4.1. Estudio GENOBOX . . . . .  | 32       |

|   |           |
|---|-----------|
| 2.4.2. El estudio PUBMEP . . . . .  | 33        |
| <b>3. Clasificación asociativa basada en reglas difusas</b>                                       |           |
| <b>positivas y negativas</b>  | <b>35</b> |
| 3.1. Sistemas de clasificación basados en reglas difusas . . . . .                                | 35        |
| 3.2. Reglas de asociación difusas positivas y negativas . . . . .                                 | 37        |
| 3.3. Positive and Negative Fuzzy Associative Classifier: PN-FAC .                                 | 39        |
| 3.3.1. Aprendizaje de la granularidad . . . . .   | 39        |
| 3.3.2. Extracción de reglas difusas y preselección . . . . .                                      | 41        |
| 3.3.3. Selección de reglas y aprendizaje de las funciones de<br>pertenencia . . . . .             | 43        |
| 3.4. Estudio Experimental . . . . .   | 45        |
| 3.4.1. Comparativa con otros métodos . . . . .  | 46        |
| 3.4.2. Resultados en problema biomédico real . . . . .  | 52        |
| <b>4. Diseño de una nueva extensión de la gramática de lingüística</b>                            |           |
| <b>difusa aplicada a problemas de regresión</b>   | <b>57</b> |
| 4.1. Gramática nueva: extensión de la estructura lingüística clásica                              | 57        |
| 4.1.1. Nueva gramática para aprender reglas más generales .                                       | 58        |
| 4.1.2. Propuesta de extensión de regla para explicar la espe-<br>cificidad en regresión . . . . . | 61        |
| 4.2. Nuevo algoritmo Evolutivo Multiobjetivo basado en Árboles .                                  | 62        |
| 4.2.1. Breve descripción del algoritmo M5-prime . . . . .   | 63        |
| 4.2.2. Extensión de M5-Prime para generar árboles lingüísticos                                    | 64        |
| 4.2.3. Aprendizaje evolutivo multiobjetivo . . . . .  | 67        |
| 4.3. Estudio experimental . . . . .   | 74        |
| 4.3.1. Comparativa con otros métodos de la literatura . . . .                                     | 74        |
| 4.3.1.1. Comparativa con un método enfocado pura-<br>mente en la precisión . . . . .              | 79        |
| 4.3.2. Caso de estudio en un problema real relacionado con<br>la obesidad en niños . . . . .      | 81        |

---

|   |                |
|---|----------------|
| <b>5. Adaptación de la nueva extensión de la gramática lingüística difusa a la extracción de reglas difusas</b> | <b>85</b>      |
| 5.1. Reglas de Asociación Difusas . . . . .   | 85             |
| 5.2. Meta Elementos Difusos para reglas de asociación difusas . .   | 89             |
| 5.3. Extensión del método Fuzzy Apriori para extraer reglas de asociación difusas basadas en MEDs . . . . .     | 89             |
| 5.4. Estudio experimental . . . . .   | 94             |
| 5.4.1. Análisis de la influencia del límite de profundidad en el proceso de extracción. . . . .                 | 95             |
| 5.4.2. Resultados de la comparativa Fuzzy Apriori clásico con la nueva propuesta . . . . .                      | 96             |
| 5.4.3. Estudio de la capacidad de resumen de información de la nueva estructura de regla . . . . .              | 99             |
| <br><b>Comentarios Finales</b>  | <br><b>103</b> |
| A    Resumen y Conclusiones . . . . .   | 103            |
| B    Publicaciones Asociadas a la Tesis . . . . .   | 106            |
| C    Líneas de investigación futuras . . . . .  | 107            |
| <br><b>Bibliografía</b>   | <br><b>111</b> |



# Índice de figuras

|  |    |
|--|----|
| 2.1. Ejemplo de Base de Datos. . . . .   | 22 |
| 2.2. Ejemplo de Base de Reglas. . . . .  | 22 |
| 2.3. Esquema de las componentes de un SBRD. . . . .  | 23 |
| 2.4. Desplazamientos de una FP triangular . . . . .  | 25 |
| 2.5. Ejemplo de desplazamiento $\delta$ de la medida GM3M . . . . .  | 28 |
| 2.6. Ejemplo del índice $\gamma$ de la medida GM3M . . . . .   | 29 |
| 2.7. Ejemplo del índice $\rho$ de la medida GM3M . . . . .   | 29 |
| 2.8. Ejemplo de cálculo de RMI . . . . .   | 31 |
| 3.1. Ejemplo de regla negada. . . . .  | 38 |
| 3.2. Esquema de PN-FAC . . . . .   | 39 |
| 3.3. Esquema del aprendizaje de la granularidad. . . . .   | 40 |
| 3.4. Ejemplo de regla basada en la representación de 2-tuplas . . . .  | 45 |
| 3.5. BD generada por PN-FAC para el problema biomédico . . . . .   | 55 |
| 4.1. Etiquetas equidistribuidas . . . . .  | 59 |
| 4.2. Etiquetas reajustadas . . . . .   | 59 |
| 4.3. Función de pertenencia compuesta para el CFLTS “Hasta Et2 ”   | 60 |
| 4.4. Función de pertenencia compuesta para el CFLTS “Desde Et2<br>hasta Et4 ” . . . . .  | 60 |
| 4.5. Función de pertenencia compuesta para el CFLTS “Desde Et3 ”   | 61 |
| 4.6. Punto intermedio Centro de Gravedad (CdG) para la FP del<br>CFLTS “Entre Et2 y Et3” para la variable de salida Y. . . . . | 61 |

---

|  |    |
|--|----|
| 4.7. Diagrama de flujo del aprendizaje de reglas . . . . .                                       | 67 |
| 4.8. Diagrama de flujo del aprendizaje del algoritmo multiobjetivo                               | 68 |
| 4.9. BC resultante de la nueva propuesta ( $ECM_{tst}=0,706$ ) . . . . .                         | 83 |
| 5.1. Representación difusa de la variable Altura . . . . .                                       | 86 |
| 5.2. Representación del grado de pertenencia . . . . .   | 87 |
| 5.3. Diagrama de flujo de MFI-Apriori . . . . .  | 90 |
| 5.4. Distribución de ítemets frecuentes asociados al MED <i>hasta</i> .                          | 91 |
| 5.5. Distribución de ítemets frecuentes asociados al MED <i>desde-</i><br><i>hasta</i> . . . . . | 91 |
| 5.6. Distribución de ítemets frecuentes asociados al MED <i>desde</i> .                          | 92 |
| 5.7. Distribución de frecuencias de todas las etiquetas . . . . .                                | 92 |
| 5.8. Distribución de frecuencias asociadas al nuevo MED desde<br><i>hasta</i> . . . . .          | 93 |
| 5.9. Ejemplo simple de regla redundante del primer paso . . . . .                                | 93 |
| 5.10. Ejemplo simple de regla redundante del segundo nivel . . . . .                             | 94 |

# Índice de tablas

|  |    |
|--|----|
| 3.1. Parámetros utilizados para la comparativa. . . . .  | 46 |
| 3.2. Número de variables, ejemplos y clases de los conjuntos de datos usados en el estudio experimental. . . . .   | 47 |
| 3.3. Medias de los resultados obtenidos por los métodos tradicionales para las medidas de Prec, G-mean y Kappa . . . . .   | 48 |
| 3.4. Resultados de los test estadísticos para los métodos de clasificación tradicionales y las medidas consideradas. Ranking de Friedman y p-valores ajustados para el test de Holm. . . . .                   | 49 |
| 3.5. Medias de los resultados obtenidos por los métodos clásicos de clasificación asociativa basada en reglas para las medidas Prec, G-mean y Kappa. . . . .   | 49 |
| 3.6. Resultados de los tests estadísticos para los métodos de clasificación asociativa basada en reglas para las medidas consideradas. Ranking de Friedman y p-valores ajustados para el test de Holm. . . . . | 50 |
| 3.7. Complejidad de los métodos con la mejor precisión: número de reglas medio, número de variables medio y medida Complej. . . . .  | 51 |
| 3.8. Resultados de los tests estadísticos para la medida Complej. Ranking de Friedman y p-valores ajustados para el test de Holm. . . . .  | 51 |
| 3.9. Resultados obtenidos para las medidas de interpretabilidad RMI y GM3M para FARC-HD y PN-FAC. . . . .  | 52 |
| 3.10. Conjunto de reglas generado por PN-FAC para el problema biomédico . . . . .  | 54 |



|   |    |
|---|----|
| 4.1. Conjuntos de datos de regresión . . . . .  | 75 |
| 4.2. Métodos considerados para la comparativa . . . . .   | 76 |
| 4.3. Parámetros utilizados . . . . .  | 76 |
| 4.4. Resultados para los métodos de la comparativa . . . . .  | 77 |
| 4.5. Resultados para la nueva propuesta . . . . .   | 78 |
| 4.6. Rankings del test de Friedman's para las medidas Var, NR,<br>GM3M, RMI y ECM en Tst. . . . .                         | 79 |
| 4.7. P-valores ajustados contra la <i>Nueva propuesta</i> para las medi-<br>das Var, NR, GM3M y RMI y ECM en Tst. . . . . | 79 |
| 4.8. Resultados de la comparación con un método lingüístico y<br>orientado únicamente a la precisión. . . . .             | 80 |
| 4.9. Resultados en el problema de obesidad infantil . . . . .   | 82 |
| 5.1. Características de los conjuntos de datos . . . . .  | 95 |
| 5.2. Valores de los parámetros . . . . .  | 95 |
| 5.3. Estudio de profundidad Fuzzy Apriori original . . . . .  | 96 |
| 5.4. Estudio de profundidad Fuzzy Apriori con la nueva represen-<br>tación . . . . .                                      | 97 |
| 5.5. Resultados obtenidos con el método Fuzzy Apriori para las<br>medidas de interés . . . . .                            | 98 |
| 5.6. Resultados obtenidos con la extensión de Fuzzy Apriori para<br>las medidas de interés . . . . .                      | 98 |
| 5.7. Resultados obtenidos con el test de Wilcoxon para las medidas<br>Conf, Yule'sQ y FC. . . . .                         | 99 |

# Tabla de Acrónimos

|         |   |   |    |
|---------|---|---|----|
| DL      | — | Deep Learning                                       | 1  |
| IA      | — | Inteligencia Artificial                             | 1  |
| XAI     | — | Explainable Artificial Intelligence                 | 2  |
| SBRD    | — | Sistemas Basados en Reglas Difusas                  | 3  |
| AE      | — | Algoritmos Evolutivos                               | 4  |
| SDEs    | — | Sistemas Evolutivos Difusos                         | 4  |
| SDEMs   | — | Sistemas Difusos Evolutivos Multiobjetivo           | 4  |
| AE      | — | Algoritmos Evolutivos                               | 4  |
| AEMO    | — | Algoritmos Evolutivos Multiobjetivo                 | 4  |
| SNS     | — | Sistema Nacional de Salud                           | 5  |
| SM      | — | Síndrome Metabólico                                 | 5  |
| IR      | — | Insulinorresistencia                                | 5  |
| HOMA-IR | — | Homeostasis Model Assessment for Insulin Resistance | 8  |
| ML      | — | Machine learning                                    | 9  |
| BC      | — | Base de conocimiento                                | 21 |
| BD      | — | Base de datos                                       | 21 |
| BR      | — | Base de reglas                                      | 21 |
| FP      | — | Función de pertenencia                              | 21 |
| PN-FAC  | — | Positive Negative Fuzzy Association Classifier      | 35 |
| FC      | — | Factor de Certeza                                   | 36 |
| RADs    | — | Reglas de Asociación Difusas                        | 37 |
| FI      | — | Fuzzy Itemset                                       | 37 |
| CFLTS   | — | Conjuntos de Términos Lingüísticos Compuestos       | 57 |
| DE      | — | Desviación Estándar                                 | 63 |
| ECM     | — | Error Cuadrático Medio                              | 69 |
| MED     | — | Meta Elementos Difusos                              | 85 |



# Capítulo 1

## Introducción

### A Planteamiento

La era del Big Data, el Deep Learning (DL) y el Internet de las cosas (Internet of Things, IoT), ha supuesto un gran avance para la Inteligencia Artificial (IA), convirtiéndola en la actualidad en una de las tecnologías más revolucionarias [FWQ21]. Sin embargo, los tremendos avances que la IA ha experimentado en los tiempos recientes han provocado una ola de preocupación, ya que en la mayoría de los casos no conocemos cómo los sistemas aprenden y toman las decisiones. Casos como el proyecto del coche autónomo, que después de miles de pruebas decidió girar para salirse de la vía en un puente[Cas16], o como el de IBM Watson en el Hospital Nacional de Dinamarca[Gad17], que cometió un error muy grave al recomendar un “tratamiento mortal” para pacientes de cáncer sin poder motivar por qué realizaba esa recomendación, están generando bastante controversia con respecto a la explicabilidad de la IA. Es, por ejemplo, el caso de las muy conocidas técnicas de DL, las cuales, a pesar de sus impresionantes resultados, generan modelos que no permiten comprender cómo y en qué se basa el modelo para realizar sus predicciones. A esto se le conoce como el “problema de la caja negra” [Cas16, AB18], que a día de hoy es un problema abierto en el que se está empezando a trabajar a marchas forzadas. Pese a que estas tecnologías ofrecen magníficas oportunidades es importante desarrollar sistemas de IA merecedores de confianza, puesto que los seres humanos so-

lamente podrán confiar en ellos y aprovechar todos los beneficios que ofrece si tanto esta tecnología como las personas y los procesos subyacentes a ella son fiables.

Por ejemplo, supongamos que en un hospital se ha instalado un sistema de IA para ayudar a los médicos a diseñar dietas personalizadas para los pacientes oncológicos. Al realizar el diseño de la dieta de un paciente el sistema recomienda una proporción de proteínas considerablemente inferior a la recomendada por el médico. En este caso, es fundamental que el especialista pueda consultar la información utilizada y el razonamiento que ha realizado el sistema para realizar esta predicción, con el objetivo de revisar el historial médico del paciente y considerar si es necesario reajustar los niveles de proteínas en la dieta o reajustar el sistema para que tenga en cuenta situaciones similares en futuras predicciones. Esta transparencia del modelo permitirá a los especialistas confiar en los sistemas IA y considerarlos como herramientas complementarias que nos ayuden a mejorar el bienestar de los pacientes. Este ejemplo pone de manifiesto la importancia de dos aspectos imprescindibles para una IA ética y responsable [Les19, SNMBB21, SP19]: 1) la necesidad de comprender cómo funcionan los modelos de IA y 2) considerar estos sistemas como herramientas de ayuda a la toma de decisiones que NO pueden reemplazar la experiencia y el juicio de los profesionales.

Por todo, la fiabilidad es el requisito previo para que las personas y sociedades desarrollen, desplieguen y utilicen sistemas de IA de forma segura. Si estos sistemas y las personas que se encuentran detrás de ellos no demuestran ser merecedores de confianza, pueden producirse consecuencias no deseadas que obstaculicen su adopción, impidiendo el logro de los enormes beneficios económicos y sociales que pueden acarrear los sistemas de IA. Por ello, la Unión Europea ha desarrollado una propuesta de Ley para la IA y ha establecido unas directrices para promover el desarrollo de una IA fiable que garantice el respeto de todas las leyes y normativas aplicables (legal), que asegure el cumplimiento de los principios y valores éticos (ética) y que sea robusta (<https://artificialintelligenceact.eu>).

Este aspecto es especialmente importante en ámbitos de alto riesgo como la salud o la aplicación de la ley, donde las decisiones basadas en sistemas de IA pueden tener un impacto significativo en la vida de las personas, siendo fundamental que los expertos comprendan cómo y en qué se basa el modelo para realizar su predicción. Existe todo un campo de investigación, denominado IA explicable (conocido por las siglas en inglés XAI, “eXplainable Artificial Intelligence”) [ADRS<sup>+</sup>20, AMCM<sup>+</sup>21], que intenta resolver esta

cuestión a fin de entender mejor los mecanismos subyacentes a estos sistemas y encontrar soluciones. La XAI fomenta el uso y desarrollo de técnicas de IA y de Machine Learning que proporcionen modelos que sean fácilmente entendibles por los humanos o el uso de técnicas post-hoc que nos proporcionen explicaciones de cómo un modelo complejo (caja negra) realiza sus predicciones y qué factores o características de nuestro problema está considerando. Este enfoque facilita también la detección y prevención de posibles sesgos y discriminaciones en los modelos, ya que permite una mayor visibilidad y control sobre cómo se toman las decisiones.

El aprendizaje basado en reglas nos permite generar modelos comprensibles basados en un conjunto de reglas. Estas reglas pueden adoptar la forma de condicionales simples del tipo “Si-Entonces” o de combinaciones más complejas de reglas simples para representar el conocimiento generado para el modelo. Dentro de este tipo de aprendizaje podemos destacar que los Sistemas Basados en Reglas Difusas (SBRD) están diseñados para un ámbito de actuación más amplio, ya que permiten definir reglas formuladas verbalmente sobre dominios imprecisos [MA75]. Los SBRDs se han utilizado con éxito en muchas aplicaciones del mundo real, como el diagnóstico médico y la inversión medioambiental entre otras [MANB21, YYL<sup>+</sup>21]. Son modelos transparentes que pueden entenderse por sí mismos, sin necesidad de recurrir a técnicas explicativas para comprenderlos. Su relación natural con el comportamiento humano los hace ideales para comprender y explicar sistemas complejos. Además, el uso de la lógica difusa como herramienta de tratamiento de datos proporciona algunas ventajas importantes. Desde el punto de vista semántico, el uso de etiquetas lingüísticas en la estructura del modelo difuso es una representación natural del conocimiento que permite la interacción humana directa. Desde el punto de vista del aprendizaje, la traducción de las características de entrada en variables difusas con funciones de pertenencia difusas permite evitar el uso de límites no naturales en las particiones del dominio de variables y produce modelos descriptivos suavizados que se adaptan bien a datos con cierto grado de incertidumbre [INN04].

Sin embargo, estos métodos suelen generar un gran número de reglas (cobertura) junto con un gran número de antecedentes/variables en las reglas (especificidad). Un aumento de la cobertura y la especificidad del modelo supone a menudo una mejora de su precisión, pero compromete la interpretabilidad del modelo [BRGAR<sup>+</sup>22]. Este problema afecta directamente a la intención de utilizar estos modelos para que el usuario pueda analizarlos y comprenderlos, ya que la transparencia del modelo disminuye a medida que

aumenta su complejidad, por lo que es esencial encontrar un buen equilibrio entre precisión y transparencia [FD18]. Para ello se puede hacer uso de distintos enfoques:

- Realizar un aprendizaje o ajuste de los componentes del modelo que permita generar modelos más robustos, compactos y precisos.
- En varias aplicaciones, donde el grado de incertidumbre es muy alto, los conjuntos difusos tradicionales no son capaces de proporcionar una representación adecuada del conocimiento oculto, por lo que en su lugar se pueden hacer uso de extensiones o nuevos modelos de representación de regla que sean más flexibles.
- Considerar de forma conjunta medidas de interpretabilidad y precisión para realizar una optimización del modelo con el objetivo de conseguir un buen equilibrio entre precisión y transparencia, criterios que en muchas aplicaciones están en conflicto.

Una forma directa de abordar estos enfoques de una forma natural para obtener modelos XAI es mediante la sinergia positiva entre SBRDs y los Algoritmos Evolutivos (AE), conocidos como Sistemas Difusos Evolutivos (SDEs) [CHG<sup>+</sup>01]. Estos métodos de búsqueda y optimización se han utilizado ampliamente y con éxito para aprender y ajustar sistemas difusos. En este sentido, los SDEs pueden preservar la esencia original de la comprensibilidad de los sistemas difusos y mejorar su capacidad predictiva, ya que son capaces de considerar la transparencia y comprensibilidad del modelo en el proceso de aprendizaje [FHC<sup>+</sup>19]. Los AEs pueden aplicarse a priori durante la fase de construcción de los SBRDs, lo que da lugar a una fase de “aprendizaje”, o bien considerarse a posteriori para un ajuste más preciso de los SBRDs, lo que se conoce como fase de “ajuste”. En la literatura especializada se pueden encontrar diferentes perspectivas para la aplicación de ambos enfoques, cabe destacar el aprendizaje de los conjuntos difusos, la parametrización de las funciones de pertenencia difusas y la selección de las reglas difusas, entre otras. Además, una extensión muy interesante es utilizar Algoritmos Evolutivos Multiobjetivo (AEMO) [FAN<sup>+</sup>12], que proporcionan modelos con distinto grado de equilibrio entre medidas al poder considerar varios criterios de diseño a optimizar concurrentemente, componiendo así los llamados Sistemas Difusos Evolutivos Multiobjetivo (SDEMdel inglés MOEFs [FAN<sup>+</sup>13], Multiobjective Evolutionary Fuzzy Systems).

Por otro lado, estos enfoques no solo deben ser validados sobre benchmark en los que obtengan buenos valores para distintas medidas de calidad,

sino que deben ser aplicados sobre problemas reales en el que expertos puedan ratificar si el nivel de transparencia de los modelos obtenidos es el adecuado para que confíen en el modelo y que pueda ser considerado en áreas que han sido definidas de alto riesgo por la Comisión Europea. Para ello, en esta tesis, gracias al proyecto de excelencia de la Junta de Andalucía P18-RT-2248, se ha trabajado en el campo de la obesidad, nutrición, y salud metabólica en edad infantil con los grupos de la Dra. C.M. Aguilera del Instituto de Investigación Biosanitaria de Granada (Ibs.GRANADA), el grupo de la Dra. Rosaura Leis, pediatra adscrita al Instituto de Investigación Sanitaria de Santiago (IDIS), y el grupo de la Dra. Gloria Bueno, pediatra adscrita al Instituto de Investigación Sanitaria de Aragón (IISA), grupos que son miembro de la RETIC (RD08/0072 y RD12/0026) de Salud Materno-Infantil (SAMID), y actualmente también de la red CIBERobn (grupos CB15/00131y CB15/00043). La obesidad y sus alteraciones cardiometabólicas son un conjunto de enfermedades complejas y heterogéneas con un fuerte componente genético, cuya gravedad está además influida por factores ambientales, sociales, culturales y económicos, entre otros. Además de los problemas cardiovasculares, la obesidad se asocia con al menos 13 tipos distintos de cáncer, considerándose un factor de riesgo al igual que el tabaquismo. Recientemente se ha cifrado en 1.950 millones de euros el sobrecoste que supone para el Sistema Nacional de Salud (SNS) atender a las personas con sobrepeso y obesidad. Y es que la prevalencia de obesidad en el mundo es alarmante, hasta el punto de que actualmente se considera una pandemia [CDM19]. En España, el 16.7 % de la población adulta es obesa, y más preocupante aún es que la enfermedad alcanza al 20 % de los niños. Sin embargo, la obesidad, a diferencia de otras enfermedades, es una enfermedad progresiva que puede revertirse o controlarse más fácilmente en su fase inicial (la infancia) . El sobrepeso y la obesidad infantil son importantes factores de riesgo para una serie de alteraciones cardiometabólicas crónicas durante la edad adulta como la hipertensión o la insulinoresistencia (IR) [SFCL03], entre otras, que finalmente conducen al desarrollo del Síndrome Metabólico (SM), aumentando considerablemente la morbimortalidad de la población [ROGC<sup>+</sup>18]. Para abordar este problema de manera efectiva, existe una necesidad urgente de implementar modelos predictivos y comprensibles en etapas tempranas de la vida que sean capaces de anticipar la aparición del SM, cuando aún se está a tiempo de intervenir clínicamente.



## B Objetivos

Debido a la gran necesidad que existe en la actualidad de generar sistemas de IA que generen confianza, el objetivo general de esta tesis consiste en diseñar y desarrollar nuevos métodos evolutivos para distintos tipos de problemas (clasificación, regresión y descubrimiento de conocimiento), que hagan uso de extensiones o de nuevas representaciones de reglas difusas para obtener SBRDs precisos, transparentes y comprensibles por los humanos. Además, los métodos propuestos han sido utilizados para realizar un análisis integrado de distintos tipos de datos ómicos (genéticos, epigenéticos y metabólicos), clínicos y ambientales para obtener modelos descriptivos y predictivos comprensibles que nos permitan predecir el riesgo de sufrir insulinoresistencia (IR) y/u obesidad infantil, facilitando el diseño de programas preventivos personalizados.

Para completar satisfactoriamente dicho objetivo general, se han establecido los siguientes objetivos específicos:

- Estudio del estado del arte de la XAI, analizando las propuestas existentes en la literatura con el fin de obtener información e ideas para el desarrollo del proyecto. De esta manera se podrán determinar los retos y temas interesantes actuales en el área de investigación para desarrollar propuestas bien fundamentadas.
- Diseño de una variante evolutiva que permita realizar un aprendizaje evolutivo de un clasificador asociativo preciso y compacto, que extienda el modelo de representación de reglas difusas para poder representar dependencias negativas entre las variables de entrada y la clase. Estas nuevas reglas permitirán representar relaciones interesantes con menos reglas que con la representación clásica.
- Desarrollo de una variante evolutiva multi-objetivo que permita obtener regresores difusos que representen el comportamiento general de los datos y su variabilidad específica, generando modelos con pocas reglas y que hagan uso de una nueva representación de regla difusa que combina la nueva de gramática difusa lingüística y una novedosa extensión lineal interpretable.
- Adaptar la nueva gramática difusa lingüística propuesta para proponer un método de extracción de reglas de asociación difusas que permita

representar con menos reglas la misma información que se obtiene haciendo uso de la gramática clásica, y detectar nuevas asociaciones interesantes entre atributos que no pueden ser identificadas haciendo uso de la gramática clásica.

Estas propuestas han sido aplicadas a distintos tipos de problemas para mostrar la capacidad de estos enfoques para abordar problemas de distinta naturaleza. Además, la transparencia de dichas propuestas ha sido validada sobre los datos de obesidad, nutrición, y salud metabólica de una población de niños con/sin obesidad infantil.

## C Estructura de la memoria

Esta memoria recoge todo el trabajo realizado para conseguir los objetivos mencionados previamente. El documento se ha organizado en 4 capítulos adicionales a este, que se explican brevemente en los siguientes párrafos.

En el Capítulo 2 se introducen las directrices que ha establecido la Comisión Europea para promover una IA fiable, comentando los principios éticos que se deben cumplir y los requisitos que debe satisfacer todo sistema de IA, así como una serie de métodos técnicos y no técnicos actualmente disponibles para garantizar el cumplimiento de dichos requisitos a lo largo de todo el ciclo de vida de los sistemas de IA, dentro de los cuales profundizaremos en la XAI. Además, se presenta un estudio sobre los SDE, desde su taxonomía y sus componentes hasta su comportamiento a la hora de gestionar el balanceo entre precisión y comprensibilidad del modelo. Finalmente, se introduce en detalle el problema real sobre investigación biomédica relacionado con la obesidad infantil que se ha utilizado para validar la transparencia de los modelos obtenidos por las distintas propuestas presentadas en esta memoria.

En el Capítulo 3 se presenta un algoritmo evolutivo para obtener clasificadores asociativos con un buen equilibrio entre complejidad y precisión, considerando una extensión de la representación clásica de reglas de asociación difusas para poder representar dependencias positivas y negativas. Además, se describe el doble estudio experimental realizado, evaluando el rendimiento de esta propuesta en comparación con otras propuestas de la literatura y analizando el modelo obtenido desde el punto de vista biológico, con el fin de anticipar si niños con obesidad o sobrepeso en edad prepuberal

desarrollarán el síndrome metabólico en edad puberal.

En el Capítulo 4 se propone una nueva extensión de los SBRDs específicamente diseñada para problemas de regresión, combinando una nueva gramática difusa lingüística y una nueva extensión lineal interpretable para obtener el mayor nivel de transparencia posible. Además, se presenta un nuevo algoritmo evolutivo multi-objetivo basado en árboles de decisión, que hace uso de la nueva representación de reglas difusas para obtener regresores con un buen equilibrio entre precisión y transparencia. Finalmente, se incluye un estudio experimental doble evaluando por un lado el rendimiento de esta propuesta en comparación con otras propuestas de la literatura y analizando desde el punto de vista biológico el modelo obtenido para predecir el índice HOMA-IR(Homeostasis Model Assessment for Insulin Resistance) en una cohorte de niños, analizando la relación entre los datos de antropometría, estilo de vida, actividad física y de insulinoresistencia.

Finalmente, el Capítulo 5 recoge los comentarios finales de esta tesis, incluyendo las conclusiones obtenidas de la investigación realizada, las publicaciones que soportan la tesis y algunos aspectos sobre líneas y trabajos futuros.

## Capítulo 2

# Antecedentes

En este capítulo se recogen las principales ideas para desarrollar una IA fiable, los conceptos más importantes de la XAI y de los SDE. En el contexto de la XAI, se analizará en profundidad el concepto de explicabilidad, así como la clasificación de los modelos de Machine Learning (ML) dependiendo de su nivel de transparencia y cuáles de ellos se consideran transparentes. Además, dado que en este trabajo se utilizarán los Sistemas Difusos Evolutivos, se ha dedicado un apartado de este capítulo para recoger el funcionamiento de dichos sistemas, así como el análisis en términos de interpretabilidad de los mismos y sus áreas de aplicación.

### 2.1. Inteligencia Artificial Fiable

En [Eur18], la Comisión Europea ha establecido unas directrices para promover una inteligencia artificial fiable. Estas directrices se basan en los siguientes 3 principios:

- La IA debe ser lícita: tiene que cumplir las leyes y reglamentos aplicables.
- Ética: Debe garantizar los valores éticos.
- La IA debe ser robusta: No debe provocar un daño involuntario.

Acorde a estos principios y basándose en los derechos fundamentales, se han establecido 4 fundamentos que una IA debe reconocer para garantizar el respeto de los principios éticos: respeto a la autonomía humana, prevención del daño, equidad y explicabilidad. Además, en [Eur18] proponen una lista de requisitos clave con los que se puede evaluar si nuestra herramienta cumple con las directrices establecidas. Dichos requisitos deben tenerse en cuenta durante todo el ciclo de vida de los sistemas de IA: Acción y supervisión humanas, solidez técnica y seguridad, gestión de la privacidad y de los datos, transparencia, diversidad, no discriminación y equidad, bienestar social y ambiental y rendición de cuentas. Uno de los puntos relevantes de dicho test habla de transparencia con el fin garantizar que las personas entiendan cómo funciona la IA y puedan obtener información sobre las decisiones que se tomen. Para medir la transparencia de los modelos, proponen una serie de conceptos que se deberían tener en cuenta, como por ejemplo la trazabilidad, la explicabilidad y la comunicación.

Si se logra aplicar todas estas directrices y desarrollar inteligencias artificiales totalmente fiables, también se conseguirá ampliar el ámbito de uso de dichas tecnologías, ya que estarán preparadas para poder interactuar en entornos donde es imprescindible evitar que exista un riesgo humano. Algunas de estas aplicaciones pueden estar enfocadas a la salud y el bienestar de las personas, haciendo uso de herramientas que ayuden a la detección temprana de enfermedades o que puedan ofrecer a un paciente un tratamiento personalizado en poco tiempo.

### 2.1.1. Principios éticos

Como se ha mencionado en la introducción de esta sección, existen 4 principios éticos que deben cumplirse para garantizar un desarrollo y uso fiable de la IA. En esta sección, se analizarán con detalle cada uno de estos 4 principios:

#### **Respeto de la autonomía humana**

Al igual que los derechos fundamentales tratan de garantizar la libertad y autonomía de las personas, los sistemas de IA deben asegurar que las personas que interactúan con ellos lo hagan con autonomía sobre sí mismas. Estos sistemas deben completar las aptitudes sociales, cognitivas y culturales de las personas, sin tratar de manipularlas o subordinarlas.

## **Prevención del daño**

Estos sistemas deben asegurar la protección de la integridad del ser humano, tanto física como mental, así como su dignidad. Para ello no solo deben operar sobre entornos seguros, sino que, además, deberán ser robustos y avalar que no puedan utilizarse con fines malintencionados. Con el fin de asegurar una mayor seguridad hay que observar y controlar de forma especial a personas vulnerables o situaciones que puedan provocar efectos adversos debido a la información recibida por el sistema.

## **Equidad**

Estos sistemas deben evitar la discriminación y la estigmatización de personas y grupos por las decisiones que tomen los sistemas de IA.

## **Explicabilidad**

Para poder rebatir con cierta confianza una decisión tomada por una IA, es necesario que los procesos seguidos para llegar a dicha resolución sean transparentes. Los usuarios que de alguna manera interactúen con la IA deberán recibir de la misma toda la información referente a sus capacidades, la finalidad con la que actúan y la razón de sus decisiones. Hay casos en los que la IA no es capaz de explicar cómo ha llegado a una conclusión, dichos métodos son los denominados algoritmos de “caja negra” y estos deben ser tratados con especial atención.

### **2.1.2. Requisitos de una IA fiable**

Durante el ciclo de vida de los sistemas de IA, hay que tener en cuenta una serie de requisitos que deben aplicar tanto los desarrolladores como los responsables del despliegue y usuarios finales, con el fin de cumplir con los principios éticos propuestos. En este sentido, los desarrolladores deben encargarse de aplicar los requisitos en el proceso de diseño, los responsables de despliegue serán los responsables de que los sistemas cumplan con los requisitos establecidos y los usuarios finales, así como el resto de personas involucradas, deben informarse sobre dichos requisitos y exigir su cumplimiento. La fiabilidad de la IA se apoya en siete componentes, que se verán

en detalle en los siguientes subapartados y que deben estar presentes a lo largo de todo el ciclo de vida del sistema.

### 1. Acción y supervisión humana

Los sistemas de IA deben actuar en todo momento a favor de la acción humana, garantizando que los usuarios puedan tomar decisiones de forma autónoma. Para ello, el usuario debe recibir toda la información necesaria para interactuar correctamente con el sistema y poder valorar las decisiones tomadas, o incluso cuestionar su razonamiento.

Existen algunos ámbitos de aplicación de la IA en los que es posible que los derechos fundamentales puedan verse afectados negativamente. Para que un sistema sea fiable, es imprescindible que se garanticen los derechos fundamentales, es por ello que cuando se plantea una situación de riesgo hay que realizar una evaluación del posible impacto. En este caso debería realizarse un estudio previo al desarrollo del sistema, de tal forma que se reduzcan dichos riesgos.

Uno de los mecanismos que garantizan que se prevenga el daño a la acción humana o los derechos fundamentales es la supervisión humana. De esta manera, un humano puede asegurar que se respeten las condiciones necesarias para una IA fiable, supervisando el proceso de desarrollo y actuación de la IA.

### 2. Solidez técnica y seguridad

En relación con el principio de prevención del daño, los sistemas IA deben comportarse de tal forma que nunca actúen fuera de lo esperado, evitando así que se produzcan daños de manera involuntaria. Los sistemas IA deben protegerse de ataques malintencionados que puedan atentar contra la integridad de los datos, la infraestructura del sistema o el funcionamiento del mismo.

En caso de que surjan problemas durante la utilización de alguno de los sistemas, dichos principios deben garantizar que el sistema se comporte como se espera y no cause daños en el entorno. Si es necesario, el sistema solicitará la intervención de un humano para solventar el problema. Si durante el desarrollo se prevé algún riesgo potencialmente alto, el equipo de desarrolladores deberá implementar y probar un sistema de seguridad de forma proactiva.

### 3. Gestión de la privacidad de los datos

La privacidad es un derecho fundamental que está estrechamente relacionado con el principio de prevención del daño y que puede ser especialmente vulnerado por los sistemas de IA. Para proteger este derecho, los sistemas deben proteger los datos, así como la intimidad de los usuarios, durante todo el ciclo de vida de la IA. Al ser sistemas que utilizan información sensible, cualquier organización que maneje dichos datos debe establecer políticas de acceso a los mismos, de tal forma que solo el personal cualificado y con competencias para hacerlo pueda acceder a ellos.

Además, los sistemas IA deben recibir datos de calidad, sin sesgos sociales, imprecisiones o errores, así como garantizar la integridad de dichos datos. Si una IA recibe datos malintencionados puede hacer que las decisiones se vean afectadas, sobre todo en casos de IA con capacidad de aprendizaje.

### 4. Transparencia

Para poder identificar los motivos por los que una IA ha tomado una decisión es importante que posibiliten su trazabilidad. De esta forma se podrá identificar la razón de que una IA haya tomado una decisión errónea y tratar de solventar el problema.

La explicabilidad debe ser una capacidad que posean todos las IA, ya que en caso de que un usuario requiera una explicación sobre una decisión, el sistema deberá ser capaz de proporcionar una justificación oportuna y al nivel de conocimiento del que requiere dicha información. Para poder dotar dichos sistemas de explicabilidad es necesario que las decisiones se realicen de tal forma que puedan ser comprendidas por un humano. Cabe destacar que un aumento en la explicabilidad de los métodos puede significar una pérdida de precisión, por lo que es posible que sea necesario un estudio del sistema para encontrar dicho equilibrio.

Un aspecto importante de la transparencia de un sistema es que este debe mostrar su naturaleza de IA y no pretender ser un humano, de manera que se pueda distinguir siempre que son sistemas de IA.



## 5. Diversidad, no discriminación y equidad

En consonancia con todos estos principios, y sobre todo en la línea del principio de equidad, es importante remarcar que los sistemas IA deben ser diseñados para actuar de forma inclusiva. Es necesario que eviten sesgos injustos, que provoquen discriminación o que fomenten la marginación de ciertas personas o grupos de personas. Para conseguir este objetivo, el desarrollo de dichos sistemas debe enfocarse en que puedan ser utilizados por cualquier usuario, independientemente de su edad, género, capacidades o características.

## 6. Bienestar social y ambiental

Además de cuidar la relación con los usuarios de las IAs, es importante que estas tengan en cuenta otros aspectos de la sociedad, el resto de seres sensibles y el medio ambiente. Estos sistemas deben actuar de la forma más sostenible posible, siendo respetuosos con el medio ambiente a la hora de su desarrollo, así como en el gasto de energía durante su uso.

## 7. Rendición de cuentas

Para conseguir cumplir con el principio de equidad, se necesita que los sistemas IA tengan asociados procedimientos para garantizar la responsabilidad de sus decisiones y de los datos generados. Una de las evaluaciones que puede llevarse a cabo con el fin de avalar dichos sistemas es la auditoría, es decir, una evaluación de la herramienta por parte de auditores tanto internos como externos.

### 2.1.3. Métodos técnicos

Con el fin de asegurar la fiabilidad de los sistemas IA, en [Eur18] detallan diversos métodos técnicos que considerar en las etapas de diseño, desarrollo y uso de dichos sistemas. En esta sección se verán en detalle estos métodos técnicos.

## Diseñar arquitecturas

Es importante que los requisitos de una IA fiable puedan traducirse en procedimientos que la arquitectura pueda procesar, por ejemplo, en forma de “lista blanca” (comportamientos que una IA debe seguir en todo momento), de “lista negra” (estados que una IA no debería de adoptar nunca) o una combinación de ambas.

En el caso de los sistemas no deterministas se puede dar el caso de respuestas inesperadas, ya que siguen un ciclo «sentir-plantear-actuar», por lo que habría que considerar esta arquitectura a la hora de desarrollar el sistema orientándolo a la fiabilidad. En la etapa de “sentir” debería de ser capaz de reconocer todos los elementos del entorno necesarios para cumplir los requisitos, al “plantear” basarse únicamente en los planes que cumplen los requisitos y en la “actuación” solo realizar acciones que cumplan dichos requisitos.

## Ética y estado de Derecho desde el diseño

Dentro de la fase de diseño es necesario incluir los valores y normas que deben ser cumplidos por los sistemas IA como ayuda para evitar efectos negativos y aumentar la confianza del funcionamiento de los mismos. Los desarrolladores de herramientas que hacen uso de IA deben ser los responsables de identificar los posibles efectos negativos del sistema, así como dotarlos de un mecanismo de apagado y reinicialización a prueba de fallos o ataques.

## Métodos de explicabilidad

Para que una IA transmita confianza es imprescindible poder entender los motivos por los que un sistema ha tomado una serie de decisiones o realizado una serie de acciones. Con el fin de comprender los mecanismos de razonamiento de estos sistemas ha surgido un campo de investigación llamado Inteligencia Artificial Explicable, conocido con las siglas XAI [ADRS<sup>+</sup>20, AMCM<sup>+</sup>21]. Estos métodos que incluyen investigaciones en el ámbito de la XAI son indispensables para explicar a los usuarios el comportamiento de los sistemas y poder desplegar una tecnología fiable.

## Realización de ensayos y validación

Las validaciones tradicionales de sistemas no son del todo completas para los sistemas IA, ya que actúan sobre una naturaleza no determinista y dependiente del contexto. Es necesario supervisar cuidadosamente la estabilidad, robustez y rendimiento del modelo durante el entrenamiento y el despliegue dentro de unos límites predecibles y comprensibles. Las pruebas y la validación deben realizarse desde el principio del ciclo de vida del sistema y abarcar todos los componentes, teniendo en cuenta los procedimientos contradictorios y las diversas perspectivas. Los productos o acciones del sistema deben ser coherentes con las políticas previamente definidas para garantizar que no se violan.

## Indicadores de calidad del servicio

Es posible establecer indicadores sobre la calidad del servicio para los sistemas IA, con el fin de garantizar que se han considerado aspectos de fiabilidad durante el desarrollo y la evaluación de los mismos.

## 2.2. Inteligencia Artificial Explicable

Como se ha visto en la sección anterior, uno de los mecanismos técnicos para ayudar a que los sistemas IA sean fiables son los métodos de explicabilidad, que son investigados en profundidad por la XAI [ADRS<sup>+</sup>20, AMCM<sup>+</sup>21]. El objetivo de esta rama de la IA es poder desarrollar sistemas explicables que puedan ser comprendidos por toda persona que haga uso de los mismos. En las siguientes secciones se verá el concepto de explicabilidad y los niveles de transparencia en ML.

### 2.2.1. Explicabilidad en el contexto de Machine Learning

El primer esfuerzo a considerar para tratar de dotar a los algoritmos de Machine Learning (ML) de explicabilidad es conocer qué significa dicho concepto y cuáles son las características que deben tener estas técnicas llamadas “explicables”. Cabe destacar que es común confundir conceptos relacionados con la XAI como interpretabilidad y explicabilidad, ya que han sido usados indistintamente en varias ocasiones en la literatura. Para diferenciar ambos

conceptos se puede decir que la explicabilidad es una característica activa de un modelo, referente a cualquier acción o procedimiento aplicado al resultado del modelo con la intención de clarificar su funcionamiento interno. Por contra, la interpretabilidad es un aspecto pasivo del modelo que nos dice cómo de cercano al humano es un modelo.

En [Gun17], Gunning dio la primera definición de XAI utilizando dos palabras clave: comprensibilidad y confianza, conceptos que hemos visto repetidos en el apartado anterior para tratar de definir una IA fiable.

### 2.2.2. Niveles de transparencia en ML

Para poder medir la transparencia de un sistema hay que considerar varios conceptos que denotan cualidades intrínsecas que debe tener un modelo para ser transparente. Estos niveles de transparencia fueron propuestos por Barredo en [ADRS<sup>+</sup>20] y son los siguientes:

- **Capacidad de simulación:** Este nivel define la facilidad con la que un humano puede simular el comportamiento de un modelo. La complejidad es un atributo que va fuertemente ligado a este nivel de transparencia, es por ello que los SBRDs carecen de esta característica cuando obtienen como resultado una base de reglas muy extensa. En otras palabras podemos decir que para que un modelo sea fácil de simular es importante que pueda ser autoexplicativo.
- **Capacidad de descomposición:** Es la cualidad de poder descomponer el modelo en diferentes partes o etapas explicables por sí mismas, de tal forma que podamos comprender a través de ellas el funcionamiento completo del mismo. Dichas partes deben ser comprensibles sin la necesidad de usar ninguna herramienta, es por ello que no todos los modelos son capaces de satisfacer esta propiedad.
- **Transparencia algorítmica:** Este término denota la posibilidad de comprender el razonamiento seguido por el modelo para deducir un resultado partiendo de una determinada entrada. La principal característica que debe cumplir un modelo para ser algorítmicamente transparente es que pueda ser explorable utilizando análisis y métodos matemáticos.

### 2.2.3. Modelos de ML transparentes

Una vez estudiados los distintos niveles de transparencia, podemos clasificar los modelos de ML según su nivel de transparencia medido en dichas características. Cabe destacar que, por definición, los modelos transparentes son aquellos que no necesitan un análisis post-hoc para comprender su funcionamiento, por ello ya se pueden descartar modelos que si los necesitan como los Ensembles o las Máquinas de Soporte Vectorial. A continuación, se analizarán distintos modelos transparentes usando los niveles previamente presentados.

#### Árboles de decisión

Los árboles de decisión son modelos con muy buenas cualidades con respecto a la transparencia, ya que se pueden simular fácilmente y se adaptan muy bien a la propiedad de descomponibilidad y la transparencia algorítmica. Con respecto a la capacidad de simulación se puede decir que son modelos totalmente imitables por un humano, que siguiendo la misma toma de decisiones que los árboles podría conseguir llegar a la misma solución sin necesidad de utilizar ninguna herramienta matemática para ello. En términos de transparencia algorítmica el resultado de dichos modelos es una regla totalmente comprensible por un humano, por lo que sería sencillo tratar de deducir el por qué de dicha elección.

Aun así, todas estas características se pueden dar con mayor o menor grado dependiendo del tamaño del árbol. Por ejemplo, un árbol es fácilmente imitable cuando es pequeño y hace uso de pocas variables. Sin embargo, conforme va aumentando el tamaño su descomposición es necesaria para poder considerarlo transparente, ya que otra forma que no sea dividirlo en partes no sería fácil de comprender. Finalmente, cuando el tamaño aumenta considerablemente diremos que el modelo seguirá teniendo transparencia algorítmica, pero perderá las características anteriores, por lo que no se podrá decir que tiene un alto nivel de transparencia.

Sin embargo, la transparencia de este tipo de modelos se ve muy afectada cuando la relacionamos con la precisión. Los árboles de decisión simples no suelen dar muy buenos resultados en sus predicciones por lo que, para mejorar esta característica, hacen uso de los conocidos “ensembles”, una combinación de árboles simples que dan lugar a una estructura mucho más precisa pero a su vez mucho más compleja, perdiendo así su transparencia.

## **Sistemas Basados en Reglas**

Los modelos basados en reglas se basan en la generación de reglas del tipo si-entonces, con las que son capaces de caracterizar los datos sobre los que se crean. Además, los SBRD aprovechan las ventajas de la lógica difusa en la generación de dichas reglas, permitiendo así una mejor definición de reglas para dominios imprecisos y manteniendo la transparencia.

La transparencia de este tipo de modelos ha sido determinante para tratar de explicar modelos que no son comprensibles por sí mismos. Aunque en muchos casos han sido utilizados para representar sistemas expertos, hay que tener en cuenta dos aspectos que pueden afectar a la transparencia de esos sistemas: el cubrimiento y la especificidad. En este sentido, una gran cantidad de reglas como resultado (alto cubrimiento) afecta negativamente a la transparencia del modelo, así como tener reglas muy complejas en el antecedente o consecuente (alta especificidad) también tiene una relación negativa con esta característica.

Estos conceptos de especificidad y cubrimiento también están relacionados con los distintos niveles de transparencia mencionados en la Sección 2.2.2. Si las variables utilizadas por el modelo y el número de reglas es manejable por un humano se dice que es fácil de simular. Los modelos basados en reglas son descomponibles, de hecho dicha descomposición es necesaria cuando el cubrimiento es alto. Por tanto, transparencia del modelo depende estrechamente de la especificidad y el cubrimiento, perdiéndose por completo si los modelos son muy complejos en base a ambas cualidades.

## **Regresión logística**

La regresión logística asume una dependencia lineal entre los datos, lo que hace que no pueda generar un modelo altamente complejo y flexible a las variables utilizadas. La poca flexibilidad y, en consecuencia, la rigidez hacen que este tipo de modelos sean simples y, por tanto, se puedan considerar transparentes, a costa de obtener una precisión no demasiado buena.

## **K-Nearest Neighbors**

Los algoritmos basados en los  $k$  vecinos más cercanos operan de una forma similar a la de un humano, para clasificar un ejemplo se basan en la clase de los ejemplos más parecidos al observado y le asignan la misma clase.

Se puede decir que una persona actúa de la misma manera, ya que se basa en su experiencia para tomar decisiones. Es por ello que son considerados como algoritmos transparentes y por este motivo se han usado en muchas ocasiones en las que la interpretabilidad es un requisito imprescindible.

### **Modelos Bayesianos**

Los modelos Bayesianos representan las dependencias condicionales entre un conjunto de variables, como por ejemplo las relaciones probabilísticas entre enfermedades y síntomas, mediante un grafo acíclico dirigido. Este tipo de modelos también se considera transparente dado que puede ser simulado por un humano y es descomponible, aunque podría perder estas dos características si las variables del problema fuesen demasiado complejas.

## **2.3. Sistemas Difusos Evolutivos para la XAI**

Los Sistemas Evolutivos (SE) son un tipo de algoritmos de optimización basados en poblaciones de individuos en los que cada individuo se corresponde con una solución codificada del problema. Estos algoritmos simulan el comportamiento de la evolución biológica, la optimización de dichas soluciones se produce evaluando los individuos generados de la reproducción, mutación, recombinación y selección de los individuos de la población.

Los Sistemas Evolutivos son útiles para el diseño de SBRDs porque pueden abordar la optimización de sus componentes transformando dicho objetivo en un problema de búsqueda. De esta forma, se pueden encontrar soluciones óptimas o subóptimas para el problema en cuestión. A este tipo de sistemas se les conoce como SDE.

Los Algoritmos Evolutivos Multi Objetivo (AEMO) son una extensión de los AE que se utilizan para resolver problemas de optimización con múltiples objetivos. En el contexto de los SBRDs, los AEMO se utilizan para encontrar un buen equilibrio entre precisión e interpretabilidad del modelo [AMGR09], por lo que pueden ser muy útiles si el objetivo es maximizar la interpretabilidad a la vez que se mantiene una buena precisión. En este caso, los sistemas se conocen como SDEM.

### 2.3.1. Componentes de los Sistemas Basados en Reglas Difusas

Los principales componentes de los sistemas basados en reglas son los siguientes:

- Base de conocimiento (BC): Es la parte que recoge el conocimiento del problema. Dicha información está almacenada en las dos partes que se detallan a continuación y que se pueden ver representadas en la Figura 2.3:
  - Base de datos (BD): Contiene la estructura de los conjuntos difusos que describen las variables lingüísticas contenidas en dichas reglas. Esta información incluye dos elementos que constituyen dos de los parámetros de diseño de la BD (ver Figura 2.1):
    - Conjuntos difusos o Funciones de Pertenencia (FPs) que representan el significado de los valores lingüísticos de las variables.
    - Dominios de las variables junto con sus factores de escala.
  - Base de Reglas (BR): Conjunto de reglas difusas que se aplicarán en el proceso de inferencia para generar conocimiento o clasificar instancias dados unos datos de entrada. En la Figura 2.2 se incluye un ejemplo de Base de Reglas.
- Fuzzificador: Es la componente que transforma cada valor de entrada de real a conjunto difuso.
- Sistema de inferencia: Es el encargado de llevar a cabo el proceso de inferencia difuso del modelo, generando una salida en base a los datos de entrada y las reglas definidas.
- Defuzzificador: Es el sistema que convierte la salida difusa en un valor preciso, por lo que esta componente es la encargada de agregar y transformar la información aportada por cada uno de los conjuntos difusos.



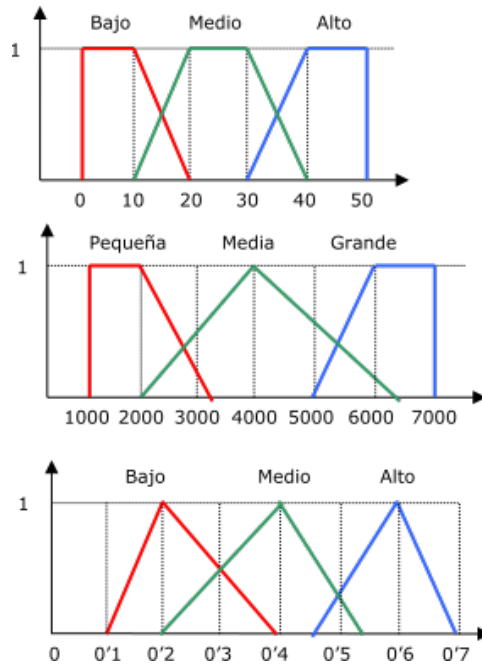


Figura 2.1: Ejemplo de Base de Datos.

R1: Si X1 es Alto y X2 es Pequeña entonces Y es Medio  
 R2: Si X1 es Bajo y X2 es Media entonces Y es Bajo  
 ...

Figura 2.2: Ejemplo de Base de Reglas.

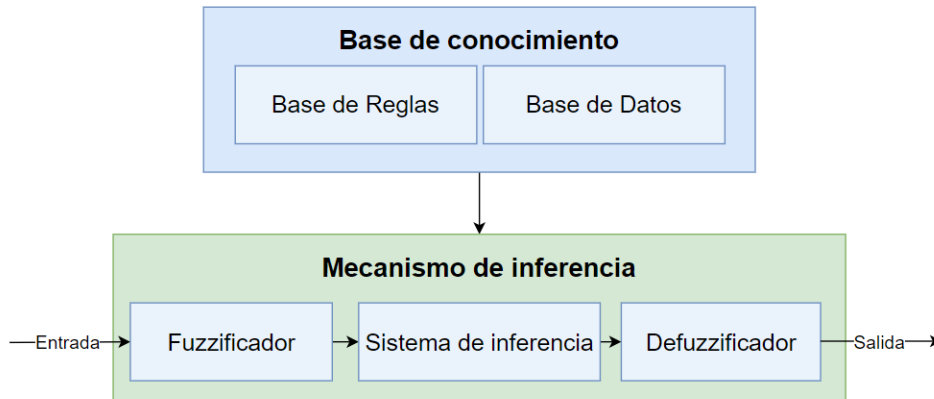


Figura 2.3: Esquema de las componentes de un SBRD.

### 2.3.2. Mejora de un SBRD desde la precisión y la comprensibilidad

Los AE pueden ser una herramienta crucial a la hora de mejorar el funcionamiento de los SBRDs, tanto en rendimiento como en transparencia. Haciendo uso de SDE para la construcción de SBRD es posible dirigir las fases desarrollo de tal forma que se obtengan sistemas con la máxima comprensibilidad posible. En este apartado veremos algunas de las técnicas utilizadas a lo largo de esta memoria para conseguir dicho objetivo.

#### Optimización de componentes de los Sistemas Basados en Reglas Difusas

Utilizando SDE se pueden optimizar las componentes de los SBRDs de la siguiente manera:

- Aprendizaje evolutivo de todos los componentes de BC de forma simultánea: Esta técnica se centra en optimizar varias componentes del BC a la misma vez. Es un método efectivo, pero ralentiza mucho la ejecución debido al amplio espacio de búsqueda.
- Aprendizaje evolutivo de la BR: Es una de las técnicas más usada en la literatura y trata de generar la Base de Reglas mediante un proceso evolutivo. De esta forma se puede guiar la construcción de las reglas

para que, por ejemplo, no incluyan antecedentes muy complejos. El objetivo es que los antecedentes no posean más de  $7 \pm 2$  condiciones diferentes, acorde al “número mágico” (propuesto por George A. Miller en [Mil56]) que es el número aproximado de conceptos que un ser humano puede manejar a la vez. Gracias al uso de esta técnica se consigue obtener modelos más transparentes.

- **Aprendizaje evolutivo de la BD:** Es una técnica que permite realizar un aprendizaje para optimizar la estructura y/o los parámetros de una BD, es decir, aprender las formas de las FP mediante un algoritmo genético.
- **Selección de reglas evolutiva:** Con el objetivo de no obtener BRs muy extensas como resultado de un modelo, se eliminarán las que no sean de interés, ya sea porque no proporcionan información relevante o porque reducen la precisión del modelo.

En los apartados anteriores se explicó que se pueden obtener modelos con resultados muy simples a costa de un decremento importante de su precisión. Sin embargo, este tipo de modelos no son de utilidad a pesar de su simplicidad, ya que no reflejan la realidad y, por tanto, no son confiables, sobre todo cuando se trabaja con problemas relacionados con la salud de las personas. Por ese motivo a veces es necesario sacrificar parte de la explicabilidad de un modelo para conseguir que los resultados precisos.

En la literatura clásica [HLV95] se han utilizado técnicas para ajustar de las FPs de las variables incluidas en la BD como una forma de mejorar la precisión del modelo a costa de una pequeña pérdida de explicabilidad. Principalmente existen 2 tipos de ajustes de las FP:

- **Ajuste clásico:** consiste en refinar las FP cambiando los parámetros de su definición, es decir, en el caso las FP triangulares cambiando de posición los 3 puntos que componen la FP (tal y como se ve en la Figura 2.4 se ajusten los puntos a, b y c). Para el caso de FP trapezoidales se ajusten los 4 puntos que componen la FP. Los desplazamientos de los puntos se hacen dentro de un intervalo controlado y no se permite que la variación de las etiquetas genere etiquetas incorrectas.
- **Ajuste de 2-tuplas [HM00]:** es un modelo de representación lingüística que permite realizar modificaciones en la distribución de las FPs, ajustando los soportes sin alterar dichas funciones. Las 2-tuplas están

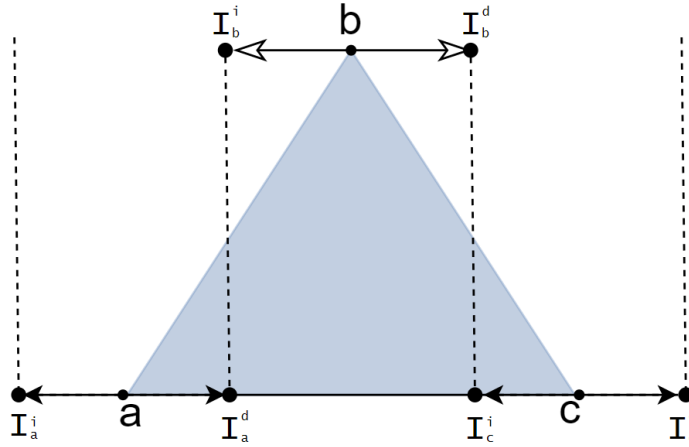


Figura 2.4: Desplazamientos de una FP triangular

formadas por un término lingüístico y un valor que representa un pequeño desplazamiento lateral de la FP asociada con dicho término. Este valor de desplazamiento debe permanecer dentro de un intervalo controlado que no de lugar a grandes modificaciones, de lo contrario la distribución de las etiquetas se alejará demasiado de la disposición original y perderá comprensibilidad (en [HM00], los autores establecen el intervalo  $[-0.5, 0.5]$ ). Este ajuste disminuye la complejidad con respecto al ajuste clásico, ya que los 3 parámetros utilizados en el ajuste clásico para FP triangulares se reducen a solo un parámetro.

### Optimización del balance entre Precisión e Interpretabilidad

Como ya se ha mencionado en la introducción, es importante que los algoritmos tengan un buen balance entre interpretabilidad y precisión. Este problema se podría definir como un problema multi objetivo, en el que no solo se busca que el algoritmo tenga una buena precisión, sino que también pueda ser fácilmente comprensible sin necesidad de utilizar técnicas para describir sus decisiones.

En esta línea, los Sistemas Evolutivos Multi Objetivo (AEMOs) pueden ser una herramienta de gran utilidad, ya que son capaces de generar SBRDs con una buena optimización de diferentes objetivos [CCLvV07]. Esta solución es conocida como SDEM [FAN<sup>+</sup>13] y está diseñada para generar SBRDs con diferentes equilibrios entre precisión e interpretabilidad.

## Nueva representación de conjuntos difusos

Optimizar la BD puede hacer que los modelos tengan una mejor precisión y unos resultados más transparentes. En este ámbito, la definición flexible de las FPs o de los conjuntos difusos de las variables que forman la BDs hace que en muchas ocasiones se pierda mucha información y esta no sea del todo comprensible. Es por ello que una correcta optimización de las componentes que definen nuestra BDs es un foco de estudio en el ámbito de la XAI.

Como se ha mencionado en otras ocasiones, la sobre dimensionalidad es un problema que hay que abordar tanto a nivel de BR como de BD. Para reducir la dimensionalidad en BD se puede optar por reducir el número de conjuntos difusos que posee el conjunto de variables, eliminando los que no se utilicen por ninguna de las reglas obtenidas o unificando aquellos que se utilizan siempre juntos. Siguiendo esta última idea mencionada, en [AMG06] proponen una nueva representación mediante la agregación de términos lingüísticos haciendo uso del conector OR.

Otra de las técnicas utilizadas para mejorar los conjuntos difusos se basa en realizar ajustes sobre las FPs, alterando su forma o los valores de pertenencia. Basándonos en la propuesta de la agrupación utilizando el conector OR en [AMG06, text], se pueden hacer modificaciones en las FPs de tal forma que se evite hacer uso del conector OR constantemente y proponer una representación que se acerque más a la forma de expresarse que tienen los humanos.

Una posible forma de reducir el número de FP es utilizando operadores como la negación o usando la combinación de etiquetas. Un ejemplo del uso de la negación en el antecedente de una regla podría ser: “La altura de una persona no es alta” que significaría que la altura de esa persona es media, baja o muy baja. Como ejemplo de combinación de etiqueta se podría tener una regla con el siguiente antecedente: “El peso de una persona está por encima de alto” que significa que el peso de la persona es alto o muy alto. Este tipo de representación consigue reducir la complejidad de la regla y hacerla más cercana a cómo se expresan las personas.

### 2.3.3. Medidas de Interpretabilidad

Acorde a la definición de interpretabilidad que dan los autores en [GAH11], se pueden considerar 4 aspectos principales a la hora de medir dicha característica: Complejidad y Semántica, a nivel de Base de Reglas (RB) o de

Base de Datos (BD).

Para poder considerar estos 4 aspectos sobre los modelos es necesario incluir medidas en cada uno de los cuatro cuadrantes de la taxonomía propuesta por [GAH11], lo que permitirá obtener modelos más interpretables. Para algunos de los cuadrantes se pueden considerar medidas clásicas, como el número de reglas y/o el número de antecedentes de las reglas (complejidad a nivel de BR), y para el caso de la complejidad a nivel de BD se suele utilizar como medida el número de variables utilizadas. Estas medidas son ampliamente conocidas y utilizadas en la literatura. Sin embargo, no hay una medida única para medir la interpretabilidad semántica a nivel de BR y a nivel de BD. Por ello, las medidas de interpretabilidad semántica GM3M [GAH11] y RMI [GGSA14] pueden ser una buena alternativa para controlar la interpretabilidad en los otros dos cuadrantes. En los siguientes subapartados veremos cada una de esas medidas en detalle.

### Medida GM3M

Es una métrica de interpretabilidad semántica de la BD que representa la proximidad de una FP ajustada a una FP interpretable previamente definida para una variable difusa. El objetivo de esta medida es optimizar tres aspectos de los SBRD: disminuir la complejidad de los sistemas, mejorar su interpretabilidad semántica y mejorar la precisión. Toma valores entre 0 y 1, que representan más proximidad conforme se acercan al 1, y se calcula en base al desplazamiento  $\delta$ , la amplitud lateral  $\gamma$  y la similitud de área  $\rho$  de la siguiente manera:

$$GM3M = \sqrt[3]{\delta \cdot \gamma \cdot \rho} \quad (2.1)$$

Para ver un ejemplo de cálculo se va a definir una función triangular  $F_j$  con los parámetros  $(a_j, b_j, c_j)$  y la función ajustada con  $(a'_j, b'_j, c'_j)$ . Los desplazamientos de los parámetros se pueden realizar dentro de los intervalos  $[I_{aj}^i, I_{aj}^d]$ ,  $[I_{bj}^i, I_{bj}^d]$  y  $[I_{cj}^i, I_{cj}^d]$  (Figura 2.4). Los índices de la función GM3M se calculan de la siguiente forma:

- El desplazamiento  $\delta$  se usa para controlar el movimiento de punto central de la FP. El cálculo se realiza midiendo la distancia normalizada entre los puntos centrales de la FP ajustada y la original, obteniendo así el desplazamiento máximo de las FPs de todas las etiquetas del

sistema. Para cada  $F_j$  se calcula  $\delta_j = \frac{|b_j - b'_j|}{I}$ , donde  $I = \frac{I_{b_j}^i - I_{b_j}^d}{2}$  representa la variación máxima de cada parámetro central.  $\delta^*$  el máximo  $\delta_j$ , que representa el peor caso y toma  $\delta^*$  valores entre 0 y 1, siendo 1 el máximo desplazamiento. El objetivo del cálculo de este índice es maximizar  $\delta = 1 - \delta^*$ , por tanto se realiza esa transformación para que el valor represente proximidad. En la Figura 2.5 se puede ver un ejemplo gráfico de este desplazamiento.

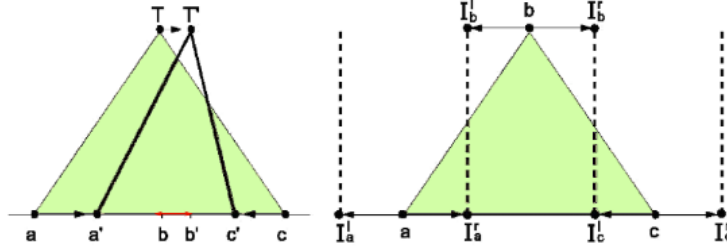


Figura 2.5: Ejemplo de desplazamiento  $\delta$  de la medida GM3M

- El índice  $\gamma$  controla que la nueva FP no se deforme en exceso con respecto a la original. Para realizar este cálculo se mide la relación entre los extremos derecho e izquierdo del soporte de las FPs antes y después del ajuste. La amplitud de la parte izquierda del soporte original se calcula como  $izqS_j = |a_j - b_j|$  y la del soporte ajustado como  $izqS'_j = |a'_j - b'_j|$ . De la misma manera, las fórmulas para la parte derecha original y ajustada son  $dchS_j = |b_j - c_j|$  y  $dchS'_j = |b'_j - c'_j|$  respectivamente. Haciendo uso de estos cálculos, la amplitud lateral final se calcula como:

$$\gamma = \frac{\min\left(\frac{izqS_j}{dchS_j}, \frac{izqS'_j}{dchS'_j}\right)}{\max\left(\frac{izqS_j}{dchS_j}, \frac{izqS'_j}{dchS'_j}\right)} \quad (2.2)$$

Los valores que toma este índice van entre 0 y 1, representando un valor próximo a 1 que la FP ajustada y la original mantienen la relación en el aspecto. Finalmente el índice  $\gamma$  se calcula como el mínimo de las amplitudes de las FPs  $\gamma_j$  para todas las etiquetas, con el objetivo de maximizar  $\gamma$ . En la Figura 2.6 se puede ver un ejemplo gráfico de este índice.

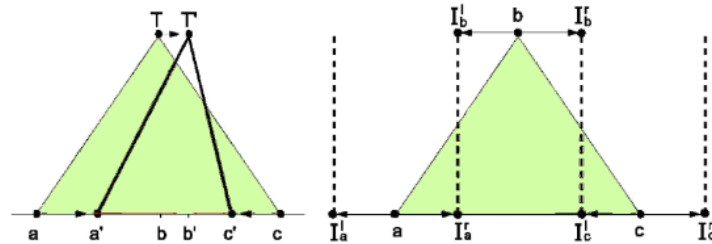


Figura 2.6: Ejemplo del índice  $\gamma$  de la medida GM3M

- El índice  $\rho$  controla la variación del área de las FP, comparando la original con la ajustada.  $A_j$  es el área del triángulo (para el caso de FP triangulares) original formado por la FP original y  $A'_j$  el área de la función ajustada, con estos valores se calcula  $\rho = \frac{\min(A_j, A'_j)}{\max(A_j, A'_j)}$ . Los valores cercanos a 1 nos dicen que ambas FPs son muy similares. Para calcular el índice final se utiliza el peor caso, que es el valor mínimo de  $\rho_j$ , con el objetivo de maximizar  $\rho = \min(\rho_j)$ . En la Figura 2.7 se puede ver un ejemplo gráfico de este índice.

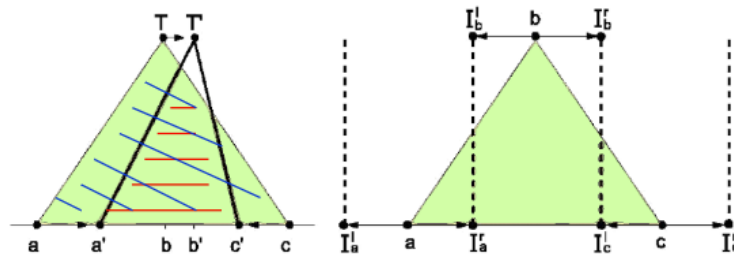


Figura 2.7: Ejemplo del índice  $\rho$  de la medida GM3M

Por tanto, GM3M es una medida que permite evaluar la interpretabilidad semántica de una BD una vez que se ha realizado un ajuste de las FPs. Se ha demostrado experimentalmente [GAH11] que el ajuste clásico de las FP permite obtener modelos más precisos y, además, gracias a la medida GM3M se ha verificado que la interpretabilidad de esos modelos sigue siendo buena.



## RMI

Es una medida de interpretabilidad semántica a nivel de BR, que representa la cantidad de reglas inconsistentes generadas por el modelo, siendo 0 presencia de este tipo de reglas y 1 ausencia.

La métrica  $RMI$  de un SBRD se calcula utilizando el valor de  $RMI(R_i)$ , que representa el RMI para una regla dada  $R_i$  de la BR. El  $RMI(R_i)$  individual de una regla nos dice el grado de fiabilidad de la misma respecto al resultado que daría el modelo completo para la zona de activación de dicha regla. Basándonos en estos RMIs individuales se calcula el RMI global para todo el SBRD.

Para desarrollar el cálculo de dicha medida se supone una regla clásica definida como:

$$R_i : Si E_{i1} es A_{i1} y \dots y E_{in} es A_{in} entonces S es B_i \quad (2.3)$$

siendo E las variables de entrada y S las variables de salida,  $n$  el número de antecedentes en la regla y  $A_{i1}, \dots, A_{in}, B_i$ . A partir de esta definición, el cálculo de RMI se define con la siguiente fórmula:

$$RMI(R_i) = \mu_{Cons_{R_i}}(O_{R_i}) \quad (2.4)$$

para la cual  $O_{R_i}$  a la salida del SBRD y  $\mu_{Cons_{R_i}}(O_{R_i})$  es el grado de pertenencia de  $O_{R_i}$  al conjunto difuso  $B_i$  del consecuente de la regla.

El valor final de RMI para un SBRD se define con el valor mínimo entre todos los posibles valores de  $RMI(R_i)$ :

$$RMI = \min_i(RMI(R_i)) \quad (2.5)$$

$$\forall 1 \leq i \leq \text{NumeroDeReglas}$$

La medida RMI toma valores en el intervalo  $[0, 1]$ , donde 0 representa el menor nivel de interpretabilidad y 1 el mayor. La Figura 2.8 muestra un ejemplo del cálculo de RMI para un SBRD con tres reglas distintas.

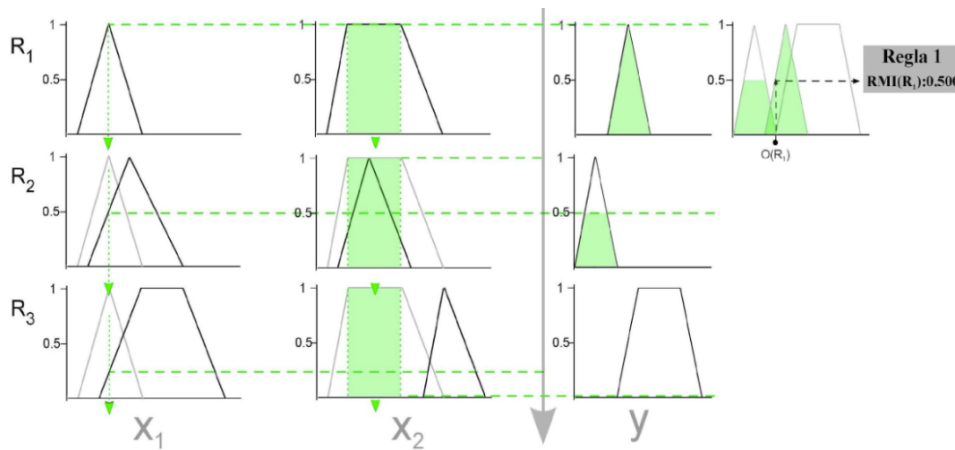


Figura 2.8: Ejemplo de cálculo de RMI

## 2.4. Problema biomédico real sobre obesidad en niños

El sobrepeso y la obesidad infantiles constituyen un importante problema de salud pública que ha aumentado de forma alarmante en todo el mundo en las últimas décadas. Los niños que padecen sobrepeso u obesidad en etapas previas a la pubertad, siguen presentando estas afecciones en edad adulta. Actualmente, la obesidad se considera uno de los trastornos crónicos que más contribuye a la carga mundial de morbilidad y plantea un importante problema de salud pública tanto en países desarrollados como para los países en desarrollo. Las principales comorbilidades asociadas a la obesidad son la diabetes de tipo II y las enfermedades cardiovasculares, que son responsables del aumento de las tasas de mortalidad en personas con obesidad.

La genética, la epigenética y los factores ambientales han interaccionado a favor del desarrollo de la morbilidad y sus enfermedades asociadas. Los factores ambientales que favorecen la obesidad suelen ser comúnmente conocidos, sin embargo la arquitectura molecular completa de la obesidad permanece inexplorada en gran medida. Es por ello que es crucial identificar biomarcadores moleculares predictivos en etapas tempranas de la vida que puedan ayudar a clasificar los pacientes en función de su riesgo de desarrollar complicaciones cardiometabólicas más adelante.

A través de las tecnologías ómicas se ha podido obtener una gran canti-

dad de datos moleculares que han ayudado a revelar las principales moléculas y vías implicadas en el desarrollo de la obesidad y los trastornos metabólicos asociados. Sin embargo, la complejidad de los datos ómicos exige enfoques analíticos nuevos e innovadores. La bioinformática y la IA pueden ofrecer recursos prometedores para abordar los problemas metodológicos inherentes al análisis de los datos ómicos. El análisis de estos datos supone una enorme complejidad debido a la gran dimensionalidad de los datos genómicos, epigenómicos y transcriptómicos, así como la necesidad de extraer patrones complejos de interacciones. La IA y la bioinformática han demostrado un potencial considerable para afrontar estos retos, permitiendo el desarrollo de pruebas moleculares y sistemas informáticos expertos en la clasificación de pacientes para el tratamiento diferencial o la intervención.

En esta tesis, se va a utilizar el problema del Síndrome Metabólico (SM) para demostrar la bondad de algunas de las técnicas de aprendizaje automático propuestas. El SM es un problema que aparece frecuentemente en pacientes obesos, alterando los parámetros metabólicos como la presión arterial, el metabolismo lipídico y el de la glucosa, y aumentando el riesgo de enfermedad cardiovascular, así como el desarrollo de Diabetes tipo 2. Realizar una detección del riesgo de obesidad asociado a esta condición, así como comprender los factores externos y metabólicos implicados, puede ser crucial para permitir una detección prematura.

#### 2.4.1. Estudio GENOBOX

El estudio GENOBOX [LMRGG<sup>+</sup>20, LCAGL<sup>+</sup>21] es un proyecto iniciado en 2011 en varias instituciones de España, entre ellas el Instituto Aragonés de Investigación Sanitaria, el Instituto de Investigación Sanitaria de Santiago de Compostela, el Instituto Maimónides de Investigación Biomédica de Córdoba y el Instituto de Investigación Biomédica IBS-Granada. El estudio utiliza un diseño transversal observacional para investigar la relación entre variantes genéticas, marcadores de estrés oxidativo e inflamación, factores de estilo de vida y riesgo cardiovascular en niños y adolescentes en España. Se invitó a participar a niños que acudían a tres hospitales para el diagnóstico de trastornos menores o sospecha de sobrepeso u obesidad, con criterios de exclusión que incluían condiciones de salud específicas, uso de medicación, ejercicio intenso reciente o participación en otros estudios de investigación. El estudio GENOBOX reclutó a un total de 1699 niños y adolescentes entre 2 y 18 años (excluyendo los niños que al nacer pesaban menos de 2500g), que se clasificaron en base a su estado de obesidad.

### 2.4.2. El estudio PUBMEP

El estudio PUBMEP es un estudio longitudinal de seguimiento de una cohorte de niños que participaron previamente en el estudio GENOBOX. La hipótesis principal del estudio PUBMEP es que la pubertad puede ser un factor de riesgo metabólico para la obesidad infantil. El estudio tiene dos objetivos principales: la primera es aclarar la relación entre la obesidad, las alteraciones metabólicas y el inicio de la maduración puberal, y la segunda comprender la arquitectura epigenética de la obesidad y sus complicaciones metabólicas. En el estudio PUBMEP se invitó a participar a niños y niñas que estaban en estado prepuberal durante el estudio GENOBOX y que habían completado la pubertad cuando comenzó el estudio PUBMEP. El estadio puberal se evaluó mediante la clasificación de Tanner y estudios hormonales. Se contactó con un total de 374 sujetos, de los cuales 49 no pudieron ser localizados, 36 no pudieron participar y 98 declinaron la invitación. De los 191 restantes que respondieron afirmativamente, se seleccionó una subpoblación de 139 niños (76 niñas) para los análisis ómicos, y 90 niños (47 mujeres) se dividieron en cinco grupos experimentales en función de su estado de obesidad y de insulinoresistencia antes y después del inicio de la pubertad.

El conjunto de datos que representan este estudio posee 58 variables que incluyen mediciones antropométricas y clínicas (como índice de masa corporal, diámetro de la cintura (DC), medidas de la presión arterial sistólica y diastólica (PAS y PAD) o la ciudad de origen), análisis bioquímicos (principalmente biomarcadores del metabolismo de lípidos y glucosa, como la glucosa, el índice de insulinoresistencia (IR), triglicéridos (TG) y colesterol en ayunas, o concentraciones de hormonas sexuales), datos proteómicos (adipocinas plasmáticas, inflamación y biomarcadores de riesgo cardiovascular), y actividad física (medida mediante acelerometría), todos ellos, parámetros relacionados con el estado de salud metabólica de los niños.



## Capítulo 3

# Clasificación asociativa basada en reglas difusas positivas y negativas

En este capítulo se propone un nuevo modelo de clasificación asociativa basado en reglas positivas y negativas, llamado PN-FAC (del inglés Positive and Negative Fuzzy Associative Classifier). Este método realiza un aprendizaje evolutivo del clasificador con una mejora en el balance entre precisión y complejidad. PN-FAC se divide en 3 etapas: un aprendizaje genético de la granularidad de las particiones difusas de una variable, una etapa de aprendizaje de reglas tanto positivas como negativas y una última etapa de selección de reglas, así como un reajuste de las etiquetas de las reglas.

### 3.1. Sistemas de clasificación basados en reglas difusas

El objetivo de un sistema de clasificación es categorizar de la forma más precisa posible un conjunto de ejemplos en base a una clase desconocida. Un problema de clasificación está formado por  $D$  ejemplos  $e_d = (e_{d1}, \dots, e_{dm}, y_d)$ , donde  $e_{di}$  es la  $i$ -ésima variable de entrada del problema ( $i = 1, 2, \dots, m$ ) e

$y_d$  la etiqueta de clase ( $y_d \in C_1, \dots, C_s$ ) del ejemplo  $d$ . Para este tipo de problemas un SBRD generará una BR con la siguiente estructura:

*Regla  $R_j$  : Si  $e_1$  es  $A_{j1}$  y ...y  $e_m$  es  $A_{jm}$  entonces Clase =  $y_j$  con  $RW_j$*

donde  $R_j$  es la regla  $j$ -ésima,  $e = (e_1, \dots, e_m)$  es un vector  $m$ -dimensional que representa los valores de las variables para los ejemplos,  $A_{ji}$  es el conjunto difuso definido en el dominio de la variable  $i$ ,  $y_j$  es la etiqueta de la clase para la regla  $j$  y  $RW_j$  es el peso de dicha regla. Para medir la calidad de las reglas existen muchos tipos de medidas en la literatura [IY05, LSL<sup>+</sup>21] y en este caso se ha utilizado la medida Factor de Certeza (FC) [CdJH99]. La definición de esta medida es la siguiente:

$$RW_j = CF_j = \frac{\sum_{e_d \in \text{Clase } y_j} \mu_{A_j}(e_d)}{\sum_{d=1}^E \mu_{A_j}(e_d)}$$

donde  $\mu_{A_j}(e_d) = T(\mu_{A_{j1}}(e_{d1}), \dots, \mu_{A_{jm}}(e_{dm}))$  es el grado de coincidencia del ejemplo  $e_d$  con el antecedente  $R_j$ . Con el fin de clasificar un nuevo ejemplo  $e_d = (e_{d1}, \dots, e_{dm})$  se ha utilizado un método de razonamiento difuso llamado combinación aditiva [CdJH99]. Este método calcula un grado de confianza (GC) para una de las etiquetas de clase y clasifica el ejemplo  $e_d$  con la etiqueta con el máximo grado. El GC de una clase  $C_k (k = 1, 2, \dots, S)$  se calcula sumando el producto del grado de coincidencia y el peso de cada regla de la BR que poseen la etiqueta  $C_k$ , de la siguiente forma:

$$GC_{C_k}(e_d) = \sum_{R_j y_j = C_k \in BR} \mu_{A_j}(e_d) \cdot RW_j$$

La capacidad de predicción de cada uno de los clasificadores se mide normalmente utilizando el ratio de precisión. Sin embargo, ya se ha visto en muchos problemas que esta medida no es la más idónea cuando se tiene un problema de clases desbalanceado, es decir, conjuntos de datos en los cuales hay diferencias significativas en las proporciones de ejemplos de cada clase. En este tipo de problemas la clase mayoritaria se denomina negativa y la minoritaria positiva. Para solventar este problema muchos investigadores han propuesto otras medidas de calidad que tienen en cuenta este desbalanceo de clases [FGG<sup>+</sup>18]. En este caso se utilizan las medidas Kappa [BD07] y G-mean [KM97] para llevar a cabo la experimentación, que para problemas de dos clases se calculan con la siguiente fórmula:

$$Kappa = \frac{Prec_o - Prec_e}{1 - Prec_e} \quad (3.1)$$

$$G - mean = \sqrt{\frac{VP}{VP + FN} \frac{VN}{VN + FP}} \quad (3.2)$$

donde  $Prec_e$  es el ratio de precisión esperado,  $Prec_o$  el ratio de precisión obtenido, VP y VN indican los ejemplos correctamente clasificados de los ejemplos de las clases positivas y negativas, y FN y FP indican los ejemplos mal clasificados. Estas medidas se han calculado siguiendo un modelo macro para conjuntos de datos multiclase.

### 3.2. Reglas de asociación difusas positivas y negativas

Las Reglas de Asociación Difusas (RADs) se usan para expresar una dependencia causal entre dos items difusos A y B, siendo A y B una pareja (*variable, etiqueta*) y representando dicha relación como  $A \rightarrow B$ . Esta representación se traduce en la frecuencia con la que se da B cuando encontramos A en el conjunto de datos. Las reglas de asociación permiten representar asociaciones entre conjuntos de elementos que en este ámbito se denominan Fuzzy Itemsets (FI).

Los métodos clásicos de extracción de RADs se basan en representar dependencias positivas sobre los datos. Sin embargo, algunas veces es interesante reflejar información relativa a las dependencias negativas que se dan en el conjunto de datos. De esta forma, vamos a definir una Regla de Asociación Difusa Negativa como  $\neg A \rightarrow B$ , que representa que dado la ausencia de A entonces se dará B. Esta negación puede aplicarse al item difuso completo, de la misma manera que en el ejemplo, así como a la etiqueta de la variable del item, quedando el item representado como una pareja (*variable, -etiqueta*). En la Figura 3.1 se puede ver un ejemplo de este tipo de reglas.

Tanto las Reglas de Asociación Difusas Positivas como las Negativas pueden ser utilizadas para clasificación contemplando en el consecuente solo la variable de clase, de tal forma que las reglas tendrán la forma  $A \rightarrow C$  o  $\neg A \rightarrow C$ , siendo  $C_k (K = 1, 2, \dots, S)$  los valores de la clase.



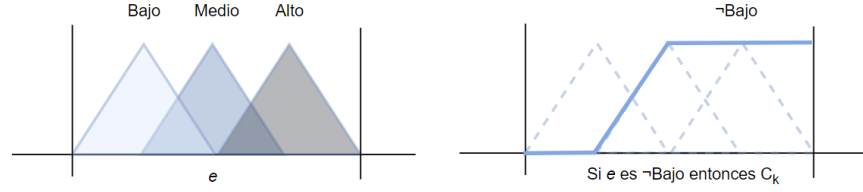


Figura 3.1: Ejemplo de regla negada.

Las medidas clásicas de soporte y confianza para reglas de asociación difusas se utilizarán tanto para las positivas como para las negativas. Ambas medidas se basan en el soporte de un FI:

$$Soporte(FI) = \frac{\sum_{e_d \in D} \mu_{FI}(e_d)}{|D|} \quad (3.3)$$

siendo  $\mu_{FI}(e_d)$  el grado de coincidencia del ejemplo  $e_d$  con el FI y  $|D|$  es el número de ejemplos en el conjunto de entrenamiento. A partir de este cálculo, dada una regla  $A \rightarrow C_k$ , el soporte y la confianza se calculan de la siguiente manera:

$$Soporte(A \rightarrow C_k) = Soporte(AC_k) \quad (3.4)$$

$$Confianza(A \rightarrow C_k) = \frac{Soporte(AC_k)}{Soporte(A)} \quad (3.5)$$

Los métodos clásicos seleccionan solamente los conjuntos difusos frecuentes, es decir, aquellos cuyo soporte supera un umbral preestablecido llamado mínimo soporte y, a partir de los mismos, generan reglas cuya confianza sea mayor que el umbral de mínima confianza. Con el fin de solucionar los problemas que conlleva este tipo de razonamiento [BBSV02] muchos autores han propuesto nuevas medidas [GH06]. En este capítulo se presentará una extensión de la medida wWRAcc [BK06] para determinar cual es el subconjunto de reglas más interesante dentro del conjunto total de reglas.

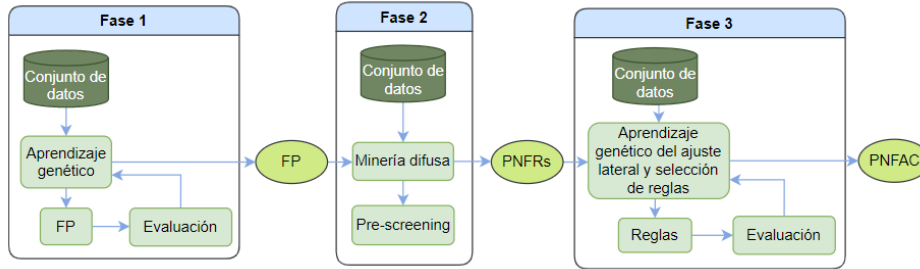


Figura 3.2: Esquema de PN-FAC

### 3.3. Positive and Negative Fuzzy Associative Classifier: PN-FAC

Esta nueva propuesta de algoritmo llamado PN-FAC consiste en un clasificador evolutivo que tiene como objetivo proporcionar un buen balance entre precisión e interpretabilidad. Para ello, se realiza un aprendizaje inicial de la granularidad de las particiones difusas de las variables, mejorando así la base de conocimiento. En la Figura 3.2 se incluye un esquema del método completo.

#### 3.3.1. Aprendizaje de la granularidad

Uno de los factores que influyen en la precisión de los SBRDs para clasificación es el número de particiones de las variables del problema, lo que se conoce como granularidad. Para optimizar esta característica de los datos de entrada se va a realizar un aprendizaje de la granularidad considerando particiones equidistribuidas con FP triangulares. Para realizar este aprendizaje se hará uso de un conocido algoritmo genético que tiene un buen balance entre exploración y explotación, llamado CHC [Esh91]. En la Figura 3.3 se puede observar un esquema del proceso de aprendizaje.

La población del algoritmo está formada por un vector de enteros que representará un cromosoma. La longitud de dichos cromosomas se corresponde con las  $m$  variables que posee el problema  $(C_1, \dots, C_m)$ , y cada uno de los genes representan el número de particiones de dicha variable, siendo dichas particiones un entero perteneciente al conjunto  $\{2, \dots, 7\}$  y representando el valor 0 la eliminación de dicha variable. Para evaluar los cromosomas se ge-

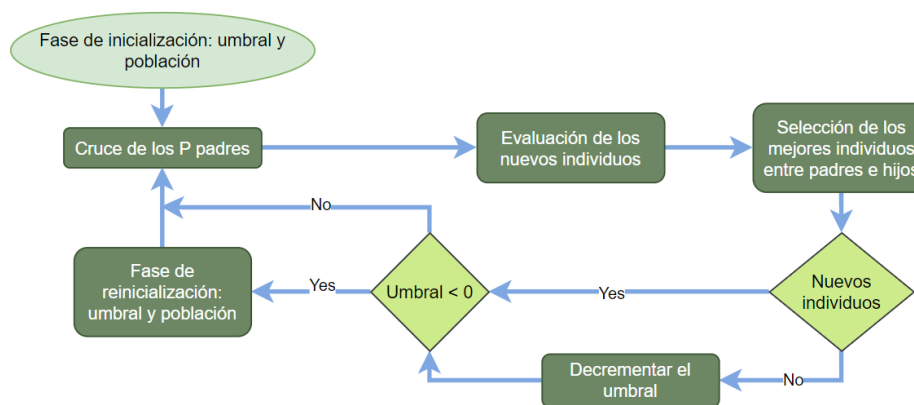


Figura 3.3: Esquema del aprendizaje de la granularidad.

nerará una nueva BR a partir de la generación de las particiones uniformes para cada variable, acorde a los valores de los genes, utilizando el mismo sistema de aprendizaje de reglas que se usará en la siguiente fase. La precisión obtenida por dicha BR será la precisión asociada a dicho cromosoma.

Para inicializar dicha población, de tal forma que se obtenga la mayor diversidad posible, los cromosomas se crean en dos partes. En la primera se genera un cromosoma por cada uno de los posibles números de particiones, es decir, el conjunto  $\{2, \dots, 7\}$ , de tal forma que cada uno de ellos tendrá todos los genes con mismo número de particiones. El resto de la población se inicializa de forma aleatoria. Además de hacer un aprendizaje de la granularidad, se hará uso del valor 0 (que elimina una variable) para hacer una selección de las variables interesantes del problema. De esta manera, en problemas que tengan más de 5 variables, se seleccionan solamente 5 de forma aleatoria para la mitad de cromosomas del segundo conjunto de inicialización.

Una de las partes relevantes del proceso de un algoritmo genético es introducir diversidad. En este algoritmo se realiza mediante la prevención del incesto y un proceso de reinicialización de la población. La prevención del incesto se implementa directamente en el proceso de cruce de tal forma que solo se cruzarán individuos que tengan un grado de diferencia superior a un umbral predefinido  $L$ . Dicho umbral se va decrementando conforme se generan nuevas poblaciones y la población se reinicializa cuando dicho umbral llega a 0, conservando el mejor cromosoma e inicializando el resto como el segundo conjunto de la inicialización inicial. En este punto se vuelve

a restablecer el valor del umbral  $L$ .

La descendencia de la población se realiza utilizando el operador de cruce HUX, un operador que maximiza la distancia entre los padres para obtener un mayor grado de exploración. El funcionamiento de este operador consiste en intercambiar la mitad de los genes diferentes en los padres de forma aleatoria. Como se ha mencionado en el párrafo anterior, solo se aplicará este cruce si la mitad de la distancia de Hamming entre los padres es mayor que el umbral  $L$ . Para la generación de las nuevas poblaciones se hace una selección elitista de la población, de tal forma que solo persisten los  $P$  mejores individuos junto con sus descendientes.

### 3.3.2. Extracción de reglas difusas y preselección

Una vez optimizada la BD con las granularidades aprendidas en el paso anterior, comienza el proceso de extracción de reglas positivas y negativas. Para ello se ha propuesto una extensión del método FARCHD [AFAH11] para extraer FI frecuentes considerando items negados. Este método se basa en buscar todos los FI frecuentes para cada clase combinando items mediante una exploración en árbol. Este proceso combinatorio comienza con los 1-itemset correspondientes a cada variable por orden de aparición en el conjunto de datos y, para poder realizar reglas negadas, se introducirán también los 1-itemsets negados en este primer nivel. Combinando los itemsets de primer nivel se obtendrán los 2-itemsets que formarán el segundo nivel del árbol.

Cada vez que se combinan 2 itemsets en un nuevo nivel, se medirá para cada uno de ellos el soporte y se deja de ramificar por ese camino si dicho  $n$ -itemset no supera el mínimo soporte, o lo que es lo mismo, deja de ser frecuente, ya que cualquier  $n$ -itemset combinado con este tampoco será frecuente. El umbral de mínimo soporte (*minsop*) se calcula teniendo en cuenta el número de ejemplos que pertenecen a cada clase del conjunto de datos de entrenamiento.

Combinando los itemsets frecuentes de cada una de las clases en el antecedente y añadiendo la clase en el consecuente se pueden crear las reglas de asociación difusas (RADs) positivas y negativas para clasificación. De la misma manera que se realiza la poda del árbol con el mínimo soporte a la hora de generar los itemsets frecuentes, se aplicará en la generación de reglas cuando alguna supere la confianza máxima, ya que habrá alcanzado la cualidad requerida para ser válida.

Con el fin de evitar generar reglas demasiado complejas, la profundidad del árbol se limita a 3, por lo que las reglas tendrán como máximo 3 itemsets en el antecedente.

Todas las medidas mencionadas en los párrafos anteriores, además de la preselección de variables en el paso anterior, ayudan al método a reducir el número de nodos de la búsqueda, reduciendo así el tiempo de ejecución y la complejidad de las reglas generadas. Sin embargo, debido a los problemas que poseen las medidas de soporte y confianza, es posible que se genere una base de reglas considerable, lo que hará que el resultado del método pierda interpretabilidad.

Para solventar el problema de obtener un gran conjunto de reglas, se selecciona el subconjunto más interesante acorde a un esquema de pesos en ejemplos. En este esquema se asocia un peso a cada uno de los ejemplos, que se usará para evaluar las reglas. Cada vez que una regla cubre un ejemplo se reduce dicho peso, de esta forma se irá guiando la selección de reglas para que cubran los ejemplos con más peso y, por tanto, los menos representados. El peso para un ejemplo  $e_d$  se calcula dependiendo del número de reglas que lo cubre ( $nr$ ):  $w(e_d, nr) = \frac{1}{nr+1}$ . Cuando más de T reglas han cubierto el ejemplo, el peso será cero. En cada una de las iteraciones se ordena por mejor valor de evaluación las reglas que no se han seleccionado, se selecciona la mejor y se ajustan los pesos de los ejemplos cubiertos por la misma. Este proceso se repite hasta que todos los ejemplos tienen peso igual a cero.

Para evaluar las reglas se ha utilizado la medida  $wWRAcc$ , que se define de la siguiente manera:

$$wWRAcc(A \rightarrow C_k) = \frac{n\mathcal{H}(A \cdot C_k)}{n\mathcal{I}(C_k)} \cdot \left( \frac{n\mathcal{H}(A \cdot C_k)}{n\mathcal{H}(A)} - \frac{n(C_k)}{|D|} \right) \quad (3.6)$$

donde  $|D|$  es el número de ejemplos de entrenamiento,  $n\mathcal{H}(A)$  es la suma de productos del grado de cobertura de los ejemplos con el antecedente de la regla multiplicados por sus pesos,  $n\mathcal{H}(A \cdot C_k)$  es la suma de los productos del grado de cobertura de los ejemplos con la regla multiplicado por su peso,  $n\mathcal{I}(C_k)$  es la suma de los pesos de los ejemplos de la clase  $C_k$  y  $n(C_k)$  es el número de ejemplos de la clase  $C_k$ . Esta medida tiene el problema de que podría seleccionar reglas con un valor de confianza bajo o alto. La primera parte de la medida hace referencia a la proporción de ejemplos no cubiertos de la clase  $C_k$  que consigue cubrir dicha regla y la segunda a la diferencia entre la confianza de la regla y la confianza de la regla por defecto de la clase

$C_k$ . Consideremos 2 reglas R1 y R2 que cubren una proporción de ejemplos no cubiertos de la clase  $C_k$  de un 0,7 y 0,4 (un buen valor de cubrimiento) y que tienen una confianza de 0,65 y 0,9 respectivamente, con una confianza del 0,2 de la regla por defecto. En este caso, R1 tendrá un valor más alto de  $wWRAcc$  que R2 (0,31 y 0,28), por lo que R1 será elegida antes que R2 teniendo una confianza menor que la otra.

En este método se ha realizado una modificación de la medida para obtener una media ponderada de ambos términos, con el fin de darle más importancia a la confianza de la regla que a su soporte, evitando así la selección de reglas de baja calidad. La nueva medida se define con la siguiente ecuación

$$wWRAcc(A \rightarrow C_k) = (1 - \lambda) \frac{n\#(A \cdot C_k)}{n\#(C_k)} + \lambda \left( \frac{n\#(A \cdot C_k)}{n\#(A)} - \frac{n(C_k)}{|D|} \right) \quad (3.7)$$

para la cual el valor de  $\lambda$  permite al experto determinar el balance entre el soporte y la confianza de la regla en el proceso de preselección (donde  $\lambda=0,7$  representa un buen funcionamiento).

### 3.3.3. Selección de reglas y aprendizaje de las funciones de pertenencia

De la misma manera que en la primera fase se hace uso del algoritmo CHC para aprender la granularidad de las variables del problema, esta fase se basará en un método similar propuesto en [AAFH07] para hacer una segunda selección de las reglas relevantes y un ajuste lateral de las FP. El objetivo de estos ajustes es guiar el algoritmo hacia obtener una buena precisión una vez que se ha conseguido una buena interpretabilidad en los pasos anteriores.

Para realizar el ajuste lateral se utilizará la representación en 2-tuplas (Sección 2.3.2), que permitirá reducir el espacio de búsqueda haciendo uso de un parámetro ( $\alpha$ ). En la Figura 3.4 se muestra un ejemplo gráfico de un desplazamiento de 0,1. En este método se ha restringido el rango de desplazamiento de las etiquetas al intervalo  $[-0,25, 0,25]$  (como se puede ver en la Figura 3.4), para evitar obtener una variación grande de las etiquetas que haga perder interpretabilidad. Con estos desplazamientos se pueden calcular las FP y generar un SBRD para clasificación basado en las reglas clásicas de Mamdani.

Los individuos de la población utilizados en el aprendizaje de los desplazamientos y la selección de reglas están compuestos por dos vectores:

- Vector de selección de reglas: Vector binario que codifica con valor 1 que la regla está seleccionada y valor 0 si no lo está.
- Vector de desplazamientos: Vector real que codifica el desplazamiento de cada una de las etiquetas y que toma valores en el intervalo  $[-0,25, 0,25]$ .

Con el fin de generar una población inicial con diversidad para la búsqueda, la mitad de los cromosomas se inicializan con todas las reglas seleccionadas y desplazamientos aleatorios y la otra mitad tendrán todos los valores asignados de forma aleatoria. Además, se añade un cromosoma con todas las reglas seleccionadas y todos los desplazamientos a 0 para considerar el estado inicial de las reglas.

Cada uno de los cromosomas se evalúa considerando su ratio de clasificación y una penalización por el número de reglas total seleccionadas. La medida Fitness (o Precisión) a maximizar está definida por la siguiente fórmula:

$$Prec = \frac{Aciertos}{D} - \delta \cdot \frac{PNFARs_{inicial}}{PNFARs_{inicial} - PNFARs + 1, 0'} \quad (3.8)$$

donde *Aciertos* es el número de ejemplos correctamente clasificados y *D* es el número total de ejemplos. *PNFARs – inicial* es el número de reglas en el conjunto inicial, *PNFARs* es el número de reglas seleccionadas y  $\delta$  es un peso predefinido por el experto que permite controlar el balance entre precisión y la complejidad de PNFAC.

Para realizar los cruces de los cromosomas se han utilizado dos operadores diferentes, dependiendo del tipo de vector. Para los vectores binarios se ha utilizado el operador HUX, que genera dos nuevos cromosomas intercambiando la mitad de los genes que son distintos en los padres. Para los cruces de los vectores reales se ha aplicado el operador de cruce “Parent Centric BLX (PCBLX)” [LHKM04]. En este caso cada uno de los genes de los nuevos cromosomas toman un valor aleatorio en un intervalo centrado en los valores de dicho gen para cada progenitor. Combinando los resultados se obtienen 4 nuevos cromosomas, de los cuales solo se seleccionan los dos mejores para la siguiente población.

Para evitar que el algoritmo converja prematuramente, se ha implementado un mecanismo de prevención del incesto que solo permite cruzar individuos cuya distancia de Hamming entre 2 sea mayor que un umbral  $L$  inicializado por un experto, que se irá decrementando en cada generación en la que no se consiga generar ningún nuevo individuo. Cuando el umbral  $L$  llega a cero, la población se reinicializa, asegurándose de que la mejor solución de la población sea aún la más precisa. En la reinicialización, la mitad de la población se regenera basándose en la mejor solución, seleccionando las mismas reglas y generando desplazamientos similares a los de dicha solución, mientras que la otra mitad se genera de forma aleatoria.

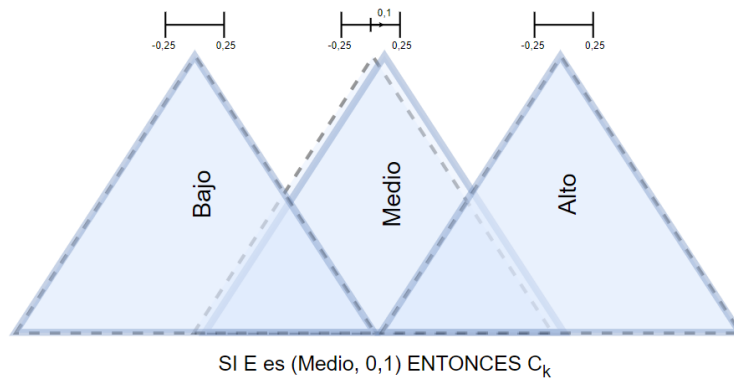


Figura 3.4: Ejemplo de regla basada en la representación de 2-tuplas

### 3.4. Estudio Experimental

El estudio experimental realizado para demostrar el rendimiento de la propuesta se ha realizado en dos etapas:

- Análisis de la precisión y la complejidad de los modelos obtenidos con la propuesta en comparativa con otros métodos de la literatura.
- Estudio de la transparencia del modelo obtenido por PN-FAC sobre un problema biomédico real relacionado con la obesidad infantil.



### 3.4.1. Comparativa con otros métodos

Para poder sacar una buena comparativa del comportamiento del algoritmo en comparación con otros se ha realizado una comparativa con 7 métodos que se encuentran disponibles en el software KEEL [AFFJ<sup>+</sup>11]. Los métodos han sido seleccionados de distintos tipos para poder comparar con otras metodologías, 3 de clasificación asociativa (CMAR [LHP01], CPAR [YH03] y FARC-HD [AFAH11]) y 4 métodos tradicionales de clasificación basada en reglas (C4.5Rules [Qui95], SLAVE2 [DG14], SGERD [MZK08] y FURIA [HH09]). Para cada uno de los métodos utilizados para la experimentación se han usado los parámetros recomendados por los autores. En los métodos propuestos, dichos parámetros se han ajustado acorde a las conclusiones extraídas tras estudiarlos experimentalmente. En la Tabla 3.1 se pueden ver todos los valores de cada uno de los parámetros para los métodos.

Tabla 3.1: Parámetros utilizados para la comparativa.

| Métodos        | Parámetros  |
|----------------|---|
| <b>CMAR</b>    | $Minconf = 0,5$ ; $Minsop = 0,01$ ; $Umbral\ de\ diferencia = 20\%$ ; $\delta = 4$  |
| <b>CPAR</b>    | $min\_ganancia = 0,7$ ; $\delta = 0,05$ ; $k = 5$ ; $\alpha = 0,66$   |
| <b>FARC-HD</b> | $Minsop = 0,05$ ; $Maxconf = 0,8$ ; $Profundidad_{max} = 3$ ; $k_t = 2$ ;<br>$Pob = 50$ ; $Evaluaciones = 15.000$ ; $BITSGEN = 30$ ; $\delta = 0,15$                                      |
| <b>C4.5</b>    | $Poda = si$ ; $Confianza = 0,25$ ; $Instancias\ por\ hoja = 2$  |
| <b>SLAVE2</b>  | $Pob = 100$ ; $Iter_{cambio} = 500$ ; $P_c = 0,6$ ; $P_m = 0,01$ ; $\kappa_1 = 0$ ; $\kappa_2 = 1$ ;<br>$Op_o = 0,1$ ; $Op_y = 0,1$ ; $P_{gen} = 0,1$ ; $P_{rot} = 0,1$ ; $\alpha = 0,15$ |
| <b>SGERD</b>   | $Q = \text{heurística}$   |
| <b>FURIA</b>   | $n\_particiones = 3$ ; $optimización = 2$   |
| <b>PN-FAC</b>  | $Minsop = 0,05$ ; $Maxconf = 0,8$ ; $T = 2$ ; $Pob = 50$ ;<br>$Eval.G = 750$ ; $Eval.ST. = 10000$ ; $BITSGEN = 30$ ; $\delta = 0,15$  |

Todos estos métodos se han ejecutado sobre 23 conjuntos de datos con un número de variables entre 2 y 90, un tamaño de entre 106 y 19020 ejemplos y un número de clases entre 2 y 15. En la Tabla 3.2, se pueden ver las características principales de los conjuntos de datos, su número de variables ( $\#Var$ ), número de ejemplos ( $\#Ejs$ ) y número de clases ( $\#Clases$ ). Para este estudio se ha utilizado una validación cruzada 5-fold y se han ejecutado

todos los métodos 3 veces para cada una de las particiones y 15 veces para cada uno de los conjuntos de datos.

Tabla 3.2: Número de variables, ejemplos y clases de los conjuntos de datos usados en el estudio experimental.

| Nombre         | #Var | #Ejs  | Clases | Nombre      | #Var | #Ejs  | Clases |
|----------------|------|-------|--------|-------------|------|-------|--------|
| Appendicitis   | 7    | 106   | 2      | Pen-based   | 16   | 10992 | 10     |
| Banana         | 2    | 5300  | 2      | Phoneme     | 5    | 5404  | 2      |
| Bupa           | 6    | 345   | 2      | Pima        | 8    | 768   | 2      |
| Cleveland      | 13   | 297   | 5      | Sonar       | 60   | 208   | 2      |
| Crx            | 15   | 653   | 2      | Spectfheart | 44   | 267   | 2      |
| Ecoli          | 7    | 336   | 8      | Texture     | 40   | 5500  | 11     |
| German         | 20   | 1000  | 2      | Twonorm     | 20   | 7400  | 2      |
| Heart          | 13   | 270   | 2      | Vehicle     | 18   | 846   | 4      |
| Iris           | 4    | 150   | 3      | Vowel       | 13   | 990   | 11     |
| Newthyroid     | 5    | 215   | 3      | Wine        | 13   | 178   | 3      |
| Magic          | 10   | 19020 | 2      | Wisconsin   | 9    | 683   | 2      |
| MovementLibras | 90   | 360   | 15     |             |      |       |        |

Para alguno de los conjuntos de datos seleccionados ha sido necesario aplicar un preprocesamiento para obtener unos resultados correctos. Cleveland, Crx y Wisconsin poseen valores perdidos, que han sido eliminados. Debido a que CMAR y CPAR no admiten conjuntos de datos con variables continuas, se ha aplicado el método Entropy para discretizar los valores. Para los métodos difusos, las particiones lingüísticas iniciales se han realizado con 5 etiquetas difusas, con forma triangular y equidistribuidas, a excepción del método SGRED que utiliza particiones uniformes de diferentes granularidades (2, 3, 4 y 5 etiquetas).

Para evaluar la nueva propuesta se han realizado 15 ejecuciones de cada método para cada conjunto de datos, utilizando como medidas de calidad la precisión, G-mean y Kappa que permiten considerar los posibles desbalances entre clases de algunos conjuntos de datos. La media de los resultados obtenidos para cada uno de los conjuntos de datos se puede ver en la web de la propuesta <http://sci2s.ugr.es/PN-FAC/>. Para analizar estadísticamente

los resultados, se ha aplicado el test de Friedman que permite determinar si hay una diferencia significativa en los resultados obtenidos por cada método. Una vez rechazada la hipótesis de igualdad, se utiliza el test de Holm para realizar comparaciones múltiples [Dem06]. Para cada comparación se calcula el p-valor ajustado (pva) con el que se rechazaba la hipótesis de igualdad.

En la Tabla 3.3 se muestran los resultados obtenidos para los métodos de clasificación tradicional, donde  $\rho$  es la desviación estándar, Prec es la media para la medida de Precisión, G-mean es la media para la medida de G-mean y Kappa es la media para la medida Kappa. La Tabla 3.4 contiene los resultados del análisis estadístico realizado, incluyendo el ranking de Friedman para cada una de las medidas y el pva para las comparativas entre el mejor del ranking de Friedman con el resto de métodos. En estos resultados se puede ver que PN-FAC obtiene el mejor puesto para el Ranking de Friedman para todas las medidas y que los p-valores ajustados son menores a 0,1 para todas las comparativas, excepto para la comparativa de Prec y Kappa con el método FURIA, lo que permite rechazar la hipótesis de igualdad con más de 90 % de confianza. Aunque estadísticamente no hay diferencias significativas con FURIA para la medida de Precisión, se puede ver que la nueva propuesta obtiene mejores medidas en G-mean y Kappa, lo que se traduce en un mejor comportamiento del método en problemas con clases desbalanceadas.

Tabla 3.3: Medias de los resultados obtenidos por los métodos tradicionales para las medidas de Prec, G-mean y Kappa

| Medidas             | Métodos      |              |              |              |                     |
|---------------------|--------------|--------------|--------------|--------------|---------------------|
|                     | C4.5Rules    | SLAVE2       | SGERD        | FURIA        | PN-FAC              |
| Prec ( $\sigma$ )   | 0,793 (0,12) | 0,798 (0,12) | 0,708 (0,15) | 0,829 (0,11) | <b>0,833 (0,11)</b> |
| G-mean ( $\sigma$ ) | 0,769 (0,17) | 0,747 (0,17) | 0,577 (0,21) | 0,780 (0,19) | <b>0,822 (0,14)</b> |
| Kappa ( $\sigma$ )  | 0,611 (0,23) | 0,594 (0,26) | 0,459 (0,24) | 0,654 (0,24) | <b>0,690 (0,23)</b> |

Las Tablas 3.5 y 3.6 muestran las comparativas de PN-FAC con otros métodos clásicos de clasificación asociativa basada en reglas (la estructura de las tablas es similar a la de los métodos clásicos). Igual que en la comparación previa, PN-FAC obtiene el mejor puesto en el ranking de Friedman en todas las medidas consideradas y los p-valores ajustados obtenidos son menores o iguales que 0,05, excepto para Prec y G-mean en la comparativa con FARC-

Tabla 3.4: Resultados de los test estadísticos para los métodos de clasificación tradicionales y las medidas consideradas. Ranking de Friedman y p-valores ajustados para el test de Holm.

| Método    | Ranking    | Ranking    | Ranking    | pva    | pva    | pva    |
|-----------|------------|------------|------------|--------|--------|--------|
|           | Prec       | G-mean     | Kappa      | Prec   | G-mean | Kappa  |
| C4.5Rules | 3,4        | 3,1        | 3,2        | 0,001  | 0,003  | 0,002  |
| SLAVE2    | 3,1        | 3,1        | 3,3        | 0,007  | 0,003  | <0,001 |
| SGERD     | 4,6        | 4,8        | 4,7        | <0,001 | <0,001 | <0,001 |
| FURIA     | 2,1        | 2,4        | 2,2        | 0,608  | 0,093  | 0,208  |
| PN-FAC    | <b>1,8</b> | <b>1,6</b> | <b>1,6</b> | -      | -      | -      |

HD, rechazando la hipótesis de igualdad con una confianza mayor al 95 %. De nuevo, la nueva propuesta muestra que no hay diferencias significativas con FARC-HD para las medidas de Prec y G-mean, sin embargo muestra mejores valores para la medida Kappa cuando las clases no están balanceadas.

Tabla 3.5: Medias de los resultados obtenidos por los métodos clásicos de clasificación asociativa basada en reglas para las medidas Prec, G-mean y Kappa.

| Medida              | Métodos      |              |              |                     |
|---------------------|--------------|--------------|--------------|---------------------|
|                     | CMAR         | CPAR         | FARC-HD      | PN-FAC              |
| Prec ( $\sigma$ )   | 0,754 (0,18) | 0,78 (0,16)  | 0,832 (0,12) | <b>0,833 (0,11)</b> |
| G-mean ( $\sigma$ ) | 0,664 (0,23) | 0,661 (0,20) | 0,801 (0,17) | <b>0,822 (0,14)</b> |
| Kappa ( $\sigma$ )  | 0,570 (0,25) | 0,595 (0,25) | 0,679 (0,24) | <b>0,691 (0,22)</b> |

Con el objetivo de comparar la complejidad de los modelos obtenidos con la nueva propuesta se ha realizado un análisis del número de reglas y del número de variables incluidas en el antecedente de las reglas generadas, comparando con aquellos métodos que han obtenido mejores resultados en los análisis previos (FURIA y FARC-HD). Para evaluar la complejidad de los modelos obtenidos se ha utilizado como medida el número de reglas

Tabla 3.6: Resultados de los tests estadísticos para los métodos de clasificación asociativa basada en reglas para las medidas consideradas. Ranking de Friedman y p-valores ajustados para el test de Holm.

| Método  | Ranking    | Ranking    | Ranking    | pva   | pva    | pva    |
|---------|------------|------------|------------|-------|--------|--------|
|         | Prec       | G-mean     | Kappa      | Prec  | G-mean | Kappa  |
| CMAR    | 3,0        | 3,2        | 3,2        | 0,018 | <0,001 | <0,001 |
| CPAR    | 2,8        | 3,5        | 3,1        | 0,033 | <0,001 | <0,001 |
| FARC-HD | 2,3        | 1,9        | 2,4        | 0,253 | 0,209  | 0,050  |
| PN-FAC  | <b>1,9</b> | <b>1,4</b> | <b>1,5</b> | -     | -      | -      |

obtenidas multiplicado por el número de variables involucradas en las reglas (Complej).

La Tabla 3.7 muestra los valores medios de las medias obtenidas sobre las 15 ejecuciones para cada método en cada conjunto de datos, donde #Reglas es la media del número de reglas utilizado en el modelo, #Var es el número medio de variables utilizado en las reglas generadas y Complej es la medida de complejidad. En la Tabla 3.8 se muestran los resultados del análisis estadístico realizado para la medida de complejidad, que se ha normalizado dividiéndola entre el valor más grande obtenido para la complejidad. A primera vista se puede observar como PN-FAC y FARC-HD son capaces de obtener modelos con reglas que utilizan pocas variables en el antecedente, aunque PN-FAC genera modelos más simples, con una media de 25% de reglas menos. Por otra parte, PN-FAC y FURIA generan modelos con un número similar de reglas, pero FURIA tiende a obtener reglas más complejas, que tienen de media el doble de variables en el antecedente que PN-FAC. Los resultados estadísticos muestran que la nueva propuesta obtiene el mejor puesto en el ranking de Friedman, con un p-valor ajustado de menos de 0,05, rechazando así la hipótesis de igualdad con una confianza mayor al 95%.

Con el fin de determinar como han cambiado las FP ajustadas con respecto a las originales y para detectar posibles inconsistencias en las reglas generadas por el modelo, se han calculado los valores de las medidas GM3M y RMI para los modelos generados por PN-FAC y FARC-HD. En este análisis no se ha incluido el algoritmo FURIA debido a que es un método de aproximación que genera una etiqueta específica para cada variable de cada

Tabla 3.7: Complejidad de los métodos con la mejor precisión: número de reglas medio, número de variables medio y medida Complej.

| Medida               | Métodos         |                 |                |
|----------------------|-----------------|-----------------|----------------|
|                      | FURIA           | FARC-HD         | PN-FAC         |
| #Reglas ( $\sigma$ ) | 28,69 (31,46)   | 39,95 (37,45)   | 28,91 (33,02)  |
| #VarInc ( $\sigma$ ) | 6,19 (0,22)     | 2,36 (0,55)     | 2,57 (0,47)    |
| Complej ( $\sigma$ ) | 207,61 (311,85) | 107,06 (108,55) | 82,78 (100,28) |

Tabla 3.8: Resultados de los tests estadísticos para la medida Complej. Ranking de Friedman y p-valores ajustados para el tets de Holm.

| Método  | Ranking    | pva     |
|---------|------------|---------|
|         | Complej    | Complej |
| FURIA   | 2,3        | 0,010   |
| FARC-HD | 2,2        | 0,012   |
| PN-FAC  | <b>1,4</b> | -       |

regla, sin hacer uso de las particiones uniformes iniciales de referencia con las que poder determinar cuando las FP se han modificado.

En la Tabla 3.9 se muestran los valores medios de las medidas RMI y GM3M para FARC-HD y PN-FAC. Analizando los resultados se puede ver que los modelos obtenidos por FARC-HD y PN-FAC obtienen ambos valores altos para las medidas de GM3M lo que indica que los desplazamientos de las FP son pequeños, por lo que no se detectan grandes pérdidas de interpretabilidad con respecto a las etiquetas iniciales. Por otra parte, los valores de RMI muestran como estos pequeños desplazamientos ajustan correctamente las FP al contexto de las etiquetas, mejorando la consistencia de las reglas generadas con respecto a FARC-HD.

Tabla 3.9: Resultados obtenidos para las medidas de interpretabilidad RMI y GM3M para FARC-HD y PN-FAC.

| Medidas           | Métodos     |             |
|-------------------|-------------|-------------|
|                   | FARC-HD     | PN-FAC      |
| RMI ( $\sigma$ )  | 0,59 (0,20) | 0,63 (0,25) |
| GM3M ( $\sigma$ ) | 0,82 (0,02) | 0,81 (0,01) |

### 3.4.2. Resultados en problema biomédico real

Para analizar el comportamiento del método propuesto en el ámbito biomédico se ha realizado un estudio sobre datos relacionados con el proyecto PUBMEP, presentado en la Sección 2.4.2. Estos datos se han recogido tanto para la etapa prepuberal T0 como para la puberal T1, aunque para este estudio solo se han utilizado los de la etapa prepuberal. Aplicando el método sobre estos datos se podría predecir la presencia de SM en niños.

Para definir la clase de los datos utilizaremos el estado de SM (sano y no sano) en la etapa puberal, clasificando con el siguiente criterio:  $DC \geq \text{percentil } 90$  [Fer05],  $PAS$  o  $PAD \geq \text{percentil } 90$  [KOGS<sup>+</sup>20, FKBS<sup>+</sup>17],  $TG \geq \text{percentil } 90$  [(NC92],  $Colesterol\ HDL \leq \text{percentil } 10$  y HOMA-IR superior a 3,388 en niños en pubertad y superior a 3,905 en niñas en pubertad [ARMGR<sup>+</sup>20]. Con este criterio, se clasifica a un niño como insano si presenta alteraciones en la medida DC y además de 2 componentes más alterados de los 4 medidos. De los 192 sujetos, la clasificación fue de 163 sanos y 29 insanos.

El conjunto de datos utilizado para la parte experimental está formado por 58 variables de entrada del estado prepuberal (T0), que se utilizarán para predecir el estado de SM de un niño en estado puberal (T1). El modelo generado posee una buena capacidad de clasificación, con unos valores para las medidas utilizadas de: Prec 0,97, G-mean 0,92 y Kappa 0,88, un conjunto de 15 reglas con antecedentes de longitud media 2,8. Se puede decir que este modelo representa una simplificación concisa y fácilmente explicable de un escenario tan complejo como es el estudio del SM y la obesidad.

Si se interpretan los resultados del método desde el punto de vista del área de conocimiento, se puede observar que el método ha seleccionado como

predictores algunas variables relevantes de la etapa prepuberal (T0), como por ejemplo algunas medidas antropométricas (circunferencia de la cintura o el IMC Z-score), datos del entorno como la ciudad de origen o algunos marcadores biomédicos (biomarcadores de lípidos y de glucosa). Estas variables forman parte de la definición de la clase SM y se ha podido demostrar que uno de los mayores factores para presentar SM en la etapa puberal es haber presentado obesidad o SM durante la infancia [JJS<sup>+</sup>17]. El resultado del método también recaló otros biomarcadores importantes de inflamación y perturbaciones cardiovasculares, incluidas las citocinas proinflamatorias (TNF $\alpha$ , adiponectina) o moléculas de coagulación y adhesión (aPAI o sICAM). Este descubrimiento es interesante porque la inflamación es uno de los principales medidores para la obesidad y los trastornos metabólicos; la incapacidad del tejido adiposo para expandirse adecuadamente provoca la infiltración de células inflamatorias en el tejido e implica la liberación de biomarcadores proinflamatorios en el torrente sanguíneo, favoreciendo un estado de inflamación sistémico en el organismo y el desarrollo de insulinoresistencia [BT11]. Además, el modelo considera la desregulación de enzimas epáticas, como la AST (aspartato aminotransferasa), como predecesor del desarrollo del SM.

El modelo está formado por 15 reglas, en las que la mayoría tienen una clara diferenciación con la primera que afecta al 65 % de la población (soporte: 0,65 y FC: 0,91). Esta regla describe que los niños con valores de circunferencia-cintura en el intervalo de-bajo-a-alto (de 54 a 93,4 cms) en la etapa prepubertal, probablemente serán niños sanos en la etapa pubertal. Esto concuerda con la conocida relación entre obesidad abdominal (visceral) y el alto riesgo de padecer SM. El resto de las reglas (con bajo soporte) hacen referencia a situaciones típicas del SM, relacionadas con marcadores inflamatorios y cardiovasculares (perfiles altos en lípidos, insulinoresistentes o inflamatorios). En la Tabla 3.10 podemos ver el conjunto de 15 reglas generadas y en la Figura 3.5 el estado final de la BD.



Tabla 3.10: Conjunto de reglas generado por PN-FAC para el problema biomédico

|  |
|--|
| SI Hip Circumference no es Third.Category [-0,04]<br>ENTONCES Healthy  |
| SI Origen no es Córdoba y Triglycerides es Top.Category.Pathological [-0,06] y APAI1 no es Top.Category [0,145]<br>ENTONCES Unhealthy                          |
| SI Origen es Santiago y Age es Early childhood [0,02] y Hip Circumference es Third.Category [-0,19]<br>ENTONCES Unhealthy                                      |
| SI Origen es Santiago y BMI z-score no es Overweight [-0,16] y Ferritin...ng.ml es Third.Category [0,13]<br>ENTONCES Unhealthy                                 |
| SI Origen es Santiago y AA no es Forth.Category [-0,17] y Soluble Intercellular Adhesion Molecule-1 es Third.Category [0,03]<br>ENTONCES Unhealthy             |
| SI Origen es Zaragoza y TPAI1 es Forth.Category [-0,06]<br>ENTONCES Unhealthy  |
| SI Origen es Zaragoza y Triglycerides es Third.Category.Normal [0,22] y TPAI1 es Second.Category [-0,05]<br>ENTONCES Unhealthy                                 |
| SI Origen es Zaragoza y Triglycerides es Third.Category.Normal [-0,08] y APAI1 no es Second.Category [0,05]<br>ENTONCES Unhealthy                              |
| SI Origen no es Zaragoza y Apolipoprotein A no es Second.Category.Normal [0,17] y AA es Second.Category [0,11]<br>ENTONCES Unhealthy                           |
| SI Origen no es Zaragoza y Apolipoprotein A es Third.Category.Normal [0,15] y AA es Second.Category [-0,11]<br>ENTONCES Unhealthy                              |
| SI Age es Late childhood [0,19] y BMI z-score es Severe Overweight [0,01] y TPAI1 es Third.Category [0,03]<br>ENTONCES Unhealthy                               |
| SI BMI z-score no es Normal-weight [0,18] y Glucose es Top.Category.Pathological [-0,21] y HOMA-IR es Third.Category.Pathological [0,11]<br>ENTONCES Unhealthy |
| SI BMI z-score es Overweight [0,09] y HOMA-IR es Third.Category.Pathological [0,06] y Uric acid es Second.Category [0,24]<br>ENTONCES Unhealthy                |
| SI BMI z-score es Severe Overweight [-0,07] y Uric acid no es Forth.Category [-0,15] y TPAI1 es Third.Category [0,08]<br>ENTONCES Unhealthy                    |
| SI Uric acid no es Second.Category [-0,06] y Adiponectin es Third.Category [0,03] y Tumor necroses factor es Third.Category [0,02]<br>ENTONCES Unhealthy       |

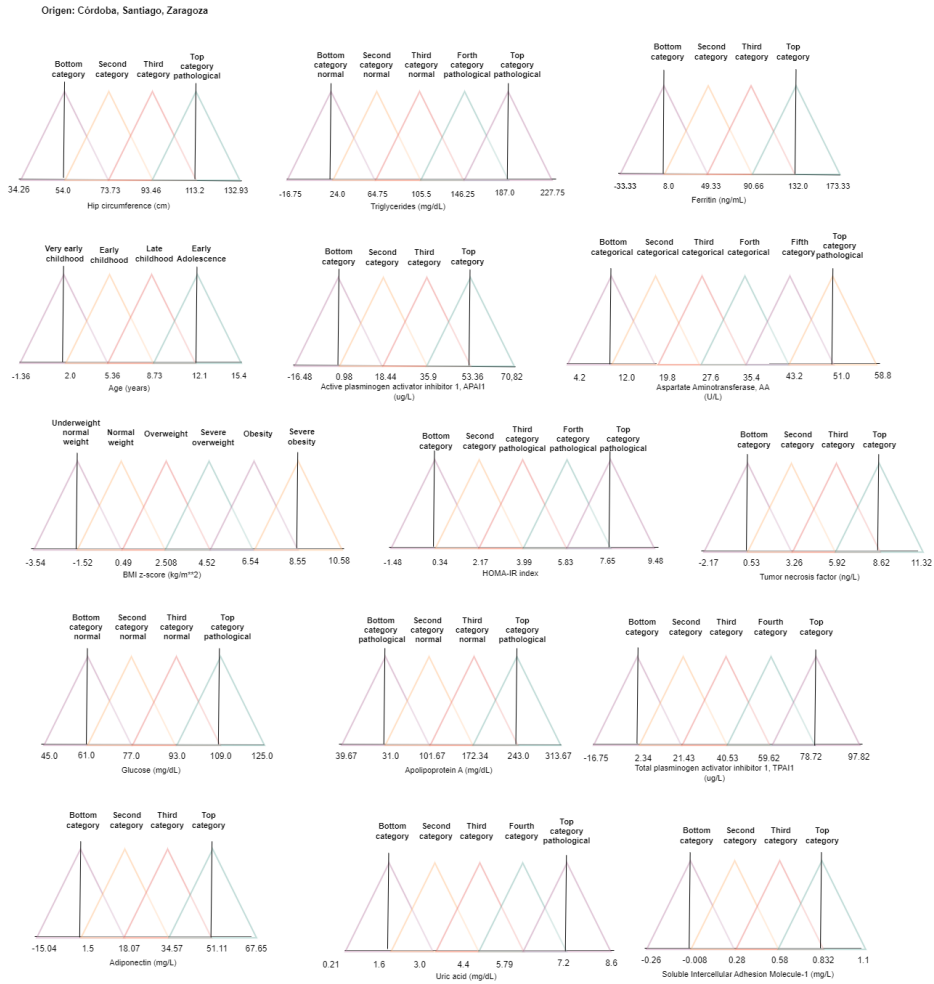


Figura 3.5: BD generada por PN-FAC para el problema biomédico



## Capítulo 4

# Diseño de una nueva extensión de la gramática de lingüística difusa aplicada a problemas de regresión

En este capítulo se va a proponer una nueva gramática difusa con el objetivo de dotar a los SBRDs de mayor transparencia, haciendo uso de términos lingüísticos más generales denominados Conjuntos de Términos Lingüísticos Compuestos (CFLTS, del inglés Composed Fuzzy Linguistic Term Sets). Esta aplicación se ha desarrollado sobre un SBRD diseñado para problemas de regresión que, haciendo uso de AEMO, combina la nueva gramática difusa con una nueva extensión lineal interpretable.

### 4.1. Gramática nueva: extensión de la estructura lingüística clásica

Esta propuesta de extensión tiene como objetivo principal permitir una representación general de ciertas partes de los datos y, además, poder obte-

ner una explicación más específica y cercana a la forma de razonar del ser humano. La idea principal es simplificar los modelos de regresión tratando de explicar el comportamiento general de los datos. De esta manera, se pueden comprender las características más generales de los datos y ver cómo se comportan particularmente en el dominio continuo de la variable. Este razonamiento es muy similar al que hace un humano cuando trata de explicar algo relacionado con valores numéricos continuos, primero generaliza y luego explica, si es necesario, que pasa en los escenarios más específicos. Un ejemplo claro podría ser la explicación de las calorías que debe ingerir una persona: Si la altura de una persona es alta entonces su ingesta de calorías debe ser entre alta y muy alta (sobre unas 2600kcal), pero si esta persona tiene más de 25 años hay que reducir en 7,3 kcal por año.

#### 4.1.1. Nueva gramática para aprender reglas más generales

En [GM05] los autores proponen una extensión de los términos lingüísticos haciendo uso del conector OR con el objetivo de crear términos más genéricos. Los CFLTS son una versión más compleja y completa de dicha propuesta, que combinan e interpretan nuevas expresiones lingüísticas más extensas en el contexto del modelado predictivo. Además, en esta propuesta se han diseñado e implementado los operadores necesarios para que los modelos puedan interactuar con esta nueva representación.

Esta nueva gramática tiene como objetivo mantener la cercanía a la forma natural de expresarse del ser humano. En ocasiones se usan términos para hacer referencia a un conjunto de datos de la media hacia arriba, por ejemplo, si se dice que una persona es más alta que la media dicha persona podrá pertenecer al grupo de los que son altos o muy altos.

Las reglas clásicas del tipo Si *antecedente* ENTONCES *consecuente* funcionan tanto con FP equidistribuidas, Figura 4.1 como con etiquetas reajustadas, Figura 4.2. En los SBRD, los antecedentes están formados por la unión de varios términos lingüísticos atómicos, cada uno asociado a una variable lingüística y unidos por el conector Y, de la forma: SI *Term1* Y *Term2* ... Y *Term3* ENTONCES *consecuente*.

Partiendo de los ejemplos de etiquetas de las Figuras 4.1 y 4.2, se pueden extraer varios ejemplos de como se definen las proposiciones lingüísticas tanto clásicas como basadas en CFLTS:

- $V_i$  es  $Et1$ : Esta es la proposición lingüística tradicional, cuyo grado de

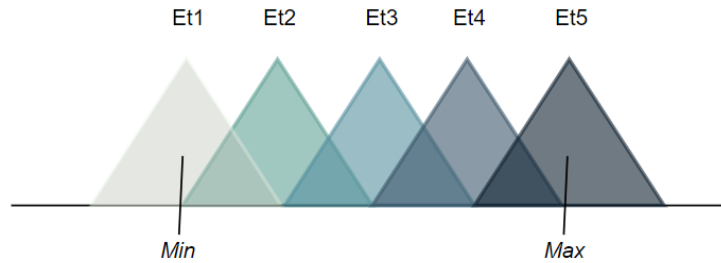


Figura 4.1: Etiquetas equidistribuidas

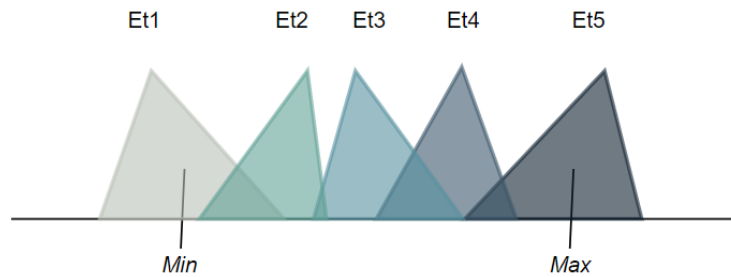


Figura 4.2: Etiquetas reajustadas

pertenencia de  $V_i$  a Et1 es el grado de cumplimiento de dicha proposición.

- $V_i$  es hasta Et2: En este caso los CFLTS representan la pertenencia de  $V_i$  a cualquier etiqueta menor o igual que Et2. El grado de pertenencia en este caso está definido como 1 de  $-\infty$  hasta el centro de Et2. En la Figura 4.3 se puede ver un ejemplo de la FP.
- $V_i$  es desde Et2 hasta Et4: Este nuevo CFLTS engloba todas las etiquetas que hay desde Et2 hasta Et4 con una FP de 1 desde el centro de la primera etiqueta hasta el centro de la segunda, tal como se muestra en la Figura 4.4.
- $V_i$  es desde Et3: Las etiquetas representadas por este CFLTS son todas las que se encuentran por encima de Et3, con la FP 1 desde la mitad de Et3 hasta  $+\infty$ . En la Figura 4.5 se puede ver un ejemplo de la FP.

Además, considerando que se hace uso del mecanismo de inferencia de “Primero Inferir y Después Agregar” (First Infer Then Aggregate, FITA),

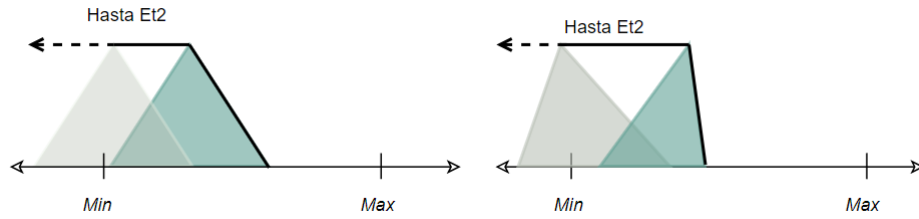


Figura 4.3: Función de pertenencia compuesta para el CFLTS “Hasta Et2 ”

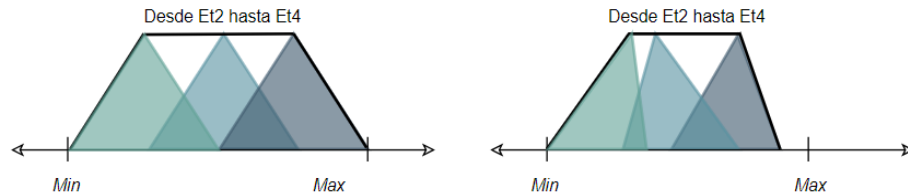


Figura 4.4: Función de pertenencia compuesta para el CFLTS “Desde Et2 hasta Et4 ”

también se puede considerar que el consecuente se exprese como la salida intermedia entre dos etiquetas adyacentes, ya que es perfectamente comprensible y utilizado por humanos. Sin embargo, en el consecuente se hace de una forma diferente al antecedente: decir que Y está entre Et2 y Et3 representa directamente el punto intermedio entre estas dos etiquetas, que tiene más sentido de cara a obtener reglas más flexibles y precisas. En la Figura 4.6 se muestra dicha composición lingüística, que, como sólo tiene sentido para el consecuente de la regla, se aplica solamente a las etiquetas o conjunto de términos de la variable de salida Y.

En el siguiente ejemplo de reglas se puede ver el uso de estos nuevos tipos de proposiciones:

|  |
|--|
| SI Altura es desde Alto<br>ENTONCES calorías-consumidas es desde Alto hasta Muy Alto |
|--|

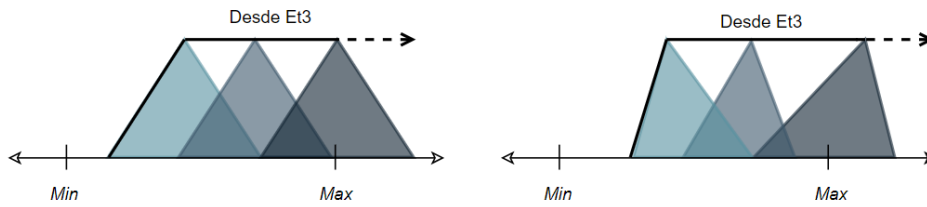


Figura 4.5: Función de pertenencia compuesta para el CFLTS “Desde Et3”

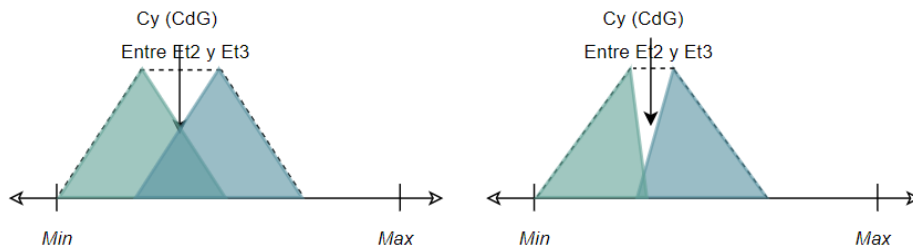


Figura 4.6: Punto intermedio Centro de Gravedad (CdG) para la FP del CFLTS “Entre Et2 y Et3” para la variable de salida Y.

#### 4.1.2. Propuesta de extensión de regla para explicar la especificidad en regresión

Haciendo uso de la gramática definida en el apartado anterior se va a definir una nueva propuesta de regla. Utilizando los CFLTS, este modelo podría ser capaz de aprender reglas que extraigan información más general de los datos y que, además, sean lingüísticamente interpretables. Este tipo de generalización tiende a generar una deficiencia en la precisión de los modelos, ya que estos tienden a representar generalidades de los datos. El reto para solucionar este problema es encontrar un mecanismo que explique de una forma más precisa los datos que componen dicha generalización. Para poder expresar conceptos más específicos dentro de un contexto general se propone el siguiente tipo de regla:

SI  $X_i$  es desde Medio  
 ENTONCES Y es desde Bajo hasta Medio  
 Con  $\pm s$  por  $\pm$  unidad en  $X_r$  para d

en la que, “desde Medio” es un CFLTS que representa que la variable  $X_i$  pertenece a las etiquetas Medio y sus superiores, “desde Bajo hasta Medio”



significa que la variable  $Y$  pertenece a las etiquetas Bajo y Medio, así como todas las que se encuentran entre ellas. En la última línea de la regla se encuentra la extensión propuesta, que se basa en obtener una línea para la cual los valores de abscisa estén centrados en la salida de la parte general de la regla, de tal forma que  $s$  es la cantidad que se debe sumar o restar a la salida general de la regla ( $C_y$ ) por cada unidad por encima/debajo de  $d$  en el valor  $X_r$  del ejemplo de entrada.

Volviendo al ejemplo del consumo de calorías de la Sección 4.1, se puede ver un ejemplo de regla de este tipo: si la altura de una persona es “desde alto” entonces la cantidad de calorías consumida debe ser “desde alto hasta muy alto” (alrededor de 2600kcal), pero la cantidad de calorías consumida debe reducirse en 7,3kcal por año a partir de los 25, o incrementarse en 7,3kcal por año por debajo de los 25. De esta forma se representa la regla de la siguiente manera:

SI Altura es desde Alto  
ENTONCES calorías-consumidas es desde Alto hasta Muy Alto  
Con  $\pm 7,3$ kcal por  $\pm$  año en Edad desde 25

Este tipo de regla se basa en el uso de CFLTS para añadir generalidad y en la parte de la línea que proporcionará la parte específica. De esta manera, una vez que la parte general de la regla obtiene una salida  $C_y$ , los valores de  $s$  (coeficiente de  $X_r$ ) y  $d$  (los desplazamientos en  $X_r$  que fuerza a la recta a pasar por el punto central de la etiqueta del consecuente  $c_y$ ) se obtienen aplicando utilizando el método de los mínimos cuadrados para obtener la recta para los ejemplos  $(X_r, Y)$ . Por tanto, la fórmula de la recta obtenida sería la siguiente:  $Y' = \beta_0 + \beta_1 X_r$  donde  $\beta_0$  y  $\beta_1$  son constantes desconocidas. Para calcular  $s$  y  $d$  de tal forma que se cumpla  $Y' = s(X_r - d) + c_y$ , hay que resolver la ecuación igualando ambas fórmulas:  $\beta_0 + \beta_1 X_r = s(X_r - d) + c_y$ , que nos dará como resultado  $s = \beta_1$  y  $d = (c_y - \beta_0)/\beta_1$

## 4.2. Nuevo algoritmo Evolutivo Multiobjetivo basado en Árboles

En esta sección se presenta un nuevo método de aprendizaje diseñado específicamente para el nuevo tipo de estructura de reglas descrito en la sección anterior. El objetivo es desarrollar un algoritmo que pueda generar

SBRDs de forma exhaustiva, manteniendo o mejorando su precisión y fiabilidad. Además, se pretende minimizar el número de reglas, maximizar la interpretabilidad lingüística (tanto GM3M como RMI), minimizar el error y mantener la longitud de las reglas dentro de un límite razonable. Este enfoque se basa en un algoritmo evolutivo multiobjetivo híbrido basado en árboles con dos etapas:

- Fase 1: Aprendizaje de las particiones lingüísticas iniciales y de las reglas lingüísticas.
- Fase 2: Post procesamiento de las soluciones. Ajuste de las FP y selección de reglas con el fin de minimizar reglas y maximizar la interpretabilidad lingüística maximizando las medidas GM3M y RMI, consiguiendo minimizar el error.

#### 4.2.1. Breve descripción del algoritmo M5-prime

El algoritmo M5-prime [Qui92, HHF99] es un método de aprendizaje automático utilizado para construir modelos de regresión, basado en árboles cuyos nodo hoja contienen un modelo lineal para predecir el valor exacto que toma una variable numérica  $Y$  proporcionada por la lógica del árbol. M5-prime tiene la capacidad de aprender eficientemente en problemas de grandes dimensiones.

##### Generación del árbol M5-prime

Para construir el árbol, el algoritmo M5-prime sigue la misma estrategia de división recursiva de nodos que los Árboles de Decisión. El conjunto  $T$ , de los ejemplos de entrenamiento, se asocia con un nodo hoja o se divide en subconjuntos utilizando la mejor división. Este proceso se repite recursivamente, por lo que en ocasiones puede producir árboles con sobreaprendizaje que deben ser podados. La ganancia de información en M5-prime se mide en base a la reducción de la desviación estándar (DE) antes y después de la división. El primer paso es calcular la DE de la variable de salida para los valores de los ejemplos en  $T$ . A menos que  $T$  contenga muy pocos casos o que varíe muy poco,  $T$  se divide basándose en los resultados del test para cada una de las posibles divisiones (para cada valor de cada atributo). Siendo  $T_i$  un subconjunto de ejemplos correspondiente al  $i$ -ésimo resultado de una

división específica y considerando la DE de la variable de salida en  $T_i$  como la medida de error, la reducción del error puede representarse como:

$$\Delta error = DE(T) - \sum_i \frac{|T_i|}{|T|} DE(T_i)$$

Utilizando esta medida, el método escogerá la división que maximice la reducción del error.

La poda de los árboles generados por M5-prime se realiza desde las hojas hasta el nodo raíz. En cada uno de los nodos internos se compara el error estimado de ese nodo y del subárbol por debajo. El subárbol es podado si no mejora el rendimiento del nodo evaluado.

### Modelos lineales

En cada uno de los nodos del árbol se genera un modelo lineal multivariable utilizando técnicas de regresión estándar. Cuando el árbol llega a su máximo tamaño, se genera un modelo de regresión multivariable para cada uno de los conjuntos de entrenamiento de cada nodo. Empezando desde las hojas, en cada nodo se consideran solo las variables utilizadas en las divisiones realizadas desde la raíz hasta llegar al nodo evaluado. M5-prime compara el error estimado del modelo lineal con los de los subárboles podados, de tal forma que permita una competición justa con los modelos que utilizan la misma información.

Una vez ha aprendido el modelo lineal, M5-prime lo simplifica eliminando coeficientes con un algoritmo Greedy, con el fin de minimizar el error uno por uno. Esto puede resultar en un incremento en el residuo medio, pero también reduce los factores multiplicativos previos para que el error pueda disminuir.

#### 4.2.2. Extensión de M5-Prime para generar árboles lingüísticos

En este apartado se define una propuesta de extensión del algoritmo M5-Prime-Rules, que genera reglas a partir del árbol generado por M5-prime. A continuación se incluye una lista de modificaciones realizadas al algoritmo original.

▪ **Cambios en la construcción del árbol:**

1. Cada uno de los nodos que se corresponden a la partición  $T_i$  tiene asociado una regla que se usará para calcular el error en dicho nodo del árbol. Para ello se calcula cómo de bien se ajustan los ejemplos de dicho grupo  $T_i$  a dicha regla, considerando este valor como el peso de cada uno de ellos. La DE de los valores de salida de los ejemplos en  $T_i$  ( $DE(T_i)$ ) se calcula como la raíz cuadrada de la media ponderada de las diferencias de cada valor de salida con respecto a su media.
2. A la hora de decidir una nueva partición se consideran todas las posibles particiones  $j$ , de tal forma que la rama izquierda va a estar asociada a  $l_j^v$  y la derecha a  $l_{j+1}^v$ , para todas las variables de entrada  $v$  existentes en las particiones lingüísticas. En esta parte cabe destacar que si alguna de las variables ha sido usada previamente en el camino entre dicho nodo y el nodo raíz, solo se podrán tener en cuenta las particiones que haya entre los extremos lingüísticos (*ini* y *fin*) definidos en el dominio para  $v$  en la división previa. De tal forma, solo se podrán considerar las particiones “desde  $l_{ini}^v$  hasta  $l_j^v$ ” y “desde  $l_{j+1}^v$  hasta  $l_{fin}^v$ ”.
3. Para cada una de las posibles divisiones se van a colocar los ejemplos de  $T$  en los dos subconjuntos generados a izquierda y derecha, asignándole un grupo dependiendo de como se ajuste el ejemplo a la proposición compuesta en dicha división. Con los dos subconjuntos de ejemplos se calcula la DE de dicha partición  $DE(T_i)$ .
4. Se elige la mejor partición que maximice la reducción del error ( $\Delta error$ ).
5. Obtiene las partes de la variable que mejor se ajustan con los datos de la regla asociada con dicho nodo y los nodos hijo, y los coeficientes de correlación  $R^2$  para las líneas rectas que representan.
6. Un nodo puede ser hoja si se cumplen algunas de las siguientes condiciones:
  - Tiene menos de 4 ejemplos.
  - La DE de la rama es menor que una pequeña fracción del total (por ejemplo  $ds(T_i) \leq 0,05 * ds(Conjunto\ de\ datos\ completo)$ ).
  - Si el  $R^2$  de la línea representada en el consecuente el padre es mayor que la del nodo en cuestión.

7. La profundidad del árbol está limitada a 3 con el fin de obtener un sistema simple, formado por un número mínimo de reglas que tengan una longitud mínima.
8. El proceso de dividir el árbol seguirá hasta que se hayan obtenido el árbol completo.

■ **Modificaciones en la poda del árbol:**

La poda se realiza desde las hojas hasta la raíz, de la misma manera que se hacía en M5-prime. A diferencia de M5-prime, en este caso la poda se basa en la medida de Error Cuadrático Medio de un nodo y su subarbol. Si el error de un nodo es menor que el error de su subarbol más un pequeño porcentaje (el error es un 5% peor), el subarbol se poda y queda el nodo como hoja. Gracias a este porcentaje se evita que el árbol se sobreajuste demasiado a los datos.

■ **Generación de reglas:**

Al final de todo ese proceso, se generan las reglas en base al árbol obtenido, convirtiendo cada uno de los caminos desde la raíz hasta los nodos hoja en una regla.

### **Añadir la componente de especificidad a las reglas**

Una vez se ha obtenido el conjunto de reglas, el algoritmo generará la parte específica de cada una de ellas cuando sea necesario, escogiendo una de las variables disponibles en el conjunto de datos. Los parámetros involucrados en esta parte de la regla se calculan usando el procedimiento descrito en la Subsección 4.1.2. Finalmente, se seleccionará la variable y los parámetros asociados que tengan el mejor  $R^2$  para la línea que representan. Esta parte específica de la regla no se añadirá si la línea no es lo suficientemente buena, es decir, si su  $R^2$  no está por encima de 0,25.

Hay que tener en cuenta que si la variable elegida para la parte específica aparece en el antecedente de la regla, el parámetro  $d$  tiene que estar en el rango que define el término lingüístico seleccionado en el antecedente. De esta forma, si  $d$  está fuera de ese rango, la parte específica no se añadirá a la regla (el  $R^2$  se fija a 0).

En la Figura 4.7 se puede ver un gráfico de flujo con todo el proceso del aprendizaje de reglas.

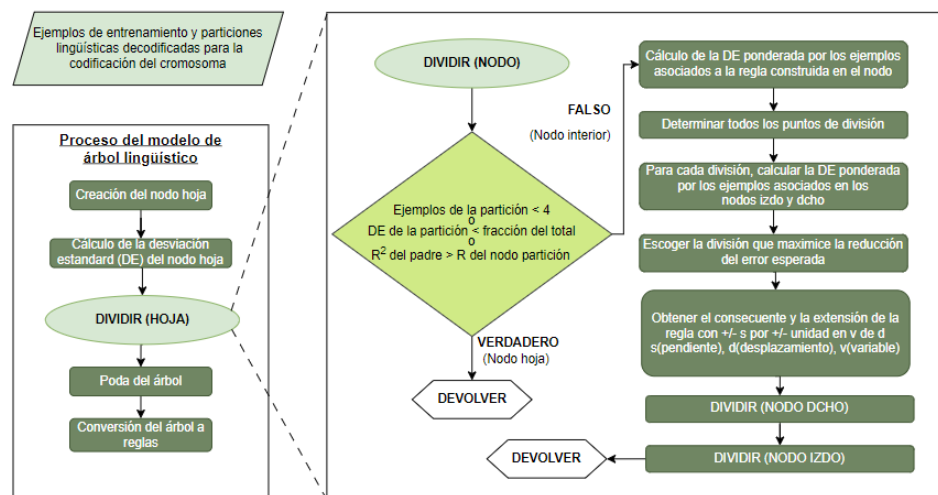


Figura 4.7: Diagrama de flujo del aprendizaje de reglas

### 4.2.3. Aprendizaje evolutivo multiobjetivo

En esta sección, se presenta el algoritmo evolutivo multiobjetivo híbrido en dos etapas. Este nuevo algoritmo está parcialmente inspirado en el algoritmo FS-MOGFS [AGH11], con una modificación del esquema de selección SPEA2 [ZLT01] y la incorporación de una población externa. El esquema general de ejecución de la metodología, junto con los objetivos de cada etapa, se representa visualmente en la Figura 4.8. En particular, este algoritmo está diseñado para dar prioridad tanto a la interpretabilidad como a la precisión de los modelos generados, seleccionando la solución más precisa de ambas etapas como resultado final. Para comprender mejor el proceso de aprendizaje de nuevas reglas lingüísticas basadas en árboles se puede revisar el diagrama de la Figura 4.7 de la sección anterior. A continuación se detallan las dos etapas que componen el algoritmo.

#### Primera fase: aprendizaje de la base de conocimiento

En esta primera fase del algoritmo se aplica un algoritmo evolutivo multiobjetivo para aprender la BD.

- Codificación de la BD:** Para codificar la BD se va a utilizar un esquema de doble codificación ( $C = C_1 + C_2$ ), considerando la gra-

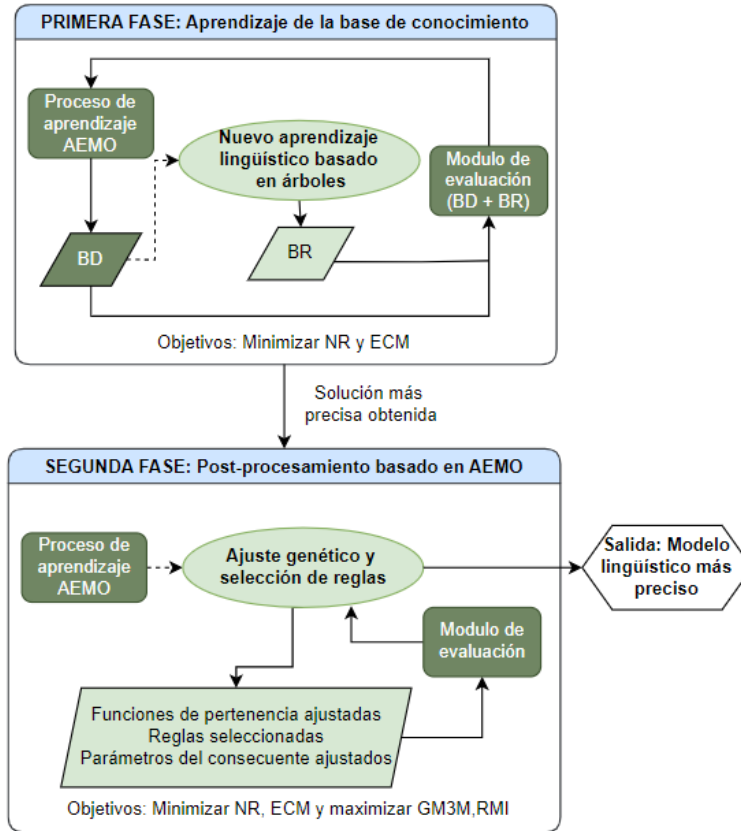


Figura 4.8: Diagrama de flujo del aprendizaje del algoritmo multiobjetivo

nularidad (número de términos lingüísticos o etiquetas simples) y el ajuste de parámetros (considerando FPs triangulares).

- El número de etiquetas ( $C_1$ ) se representa con un vector de enteros de tamaño  $N$ , siendo  $N$  el número de variables lingüísticas, en el que se almacenan las granularidades de cada una de las variables,  $C_1 = (L^1, \dots, L^N)$ . Cada uno de los genes  $L$  representa el número de etiquetas utilizadas para dividir el dominio de la variable  $n$ -ésima y puede tomar valores en el intervalo  $[2,7]$ .
- Los desplazamientos laterales de las etiquetas ( $C_2$ ) se codifican con un vector de reales de longitud  $N$ . Cada uno de los valores del vector representa los desplazamientos de las particiones difusas de cada una de las variables (en [AGH11] se puede ver una explica-

ción de como se realiza el ajuste lateral), que toman valores entre  $[-0.1, 0.1]$ ,  $C_2 = (\alpha^1, \dots, \alpha^N)$ .

- **Derivación de la base de reglas:** Para obtener un modelo lingüístico completo de un cromosoma se aplica a la BD codificada el método de aprendizaje de reglas basado en árboles propuesto en la Sección 4.2.2. Para decodificar la BD antes del aprendizaje de las reglas se definen las particiones difusas distribuidas equitativamente con los valores de granularidad de  $C_1$ . A continuación se mueven las FPs ligera y uniformemente a sus nuevas posiciones acorde a los valores establecidos en  $C_2$ . El algoritmo de aprendizaje lingüístico de reglas basado en árboles se aplica a esta BD con el fin de obtener la BR asociada (en las Secciones 4.1.1 y 4.1.2 se puede ver la estructura de las reglas y un ejemplo)
- **Objetivos a optimizar:** Una vez que se obtiene un modelo lingüístico completo, se minimizan los siguientes dos objetivos: el número de reglas (simplicidad) y el ECM (Error cuadrático medio o precisión):

$$ECM = \frac{1}{2 \cdot |E|} \sum_{l=1}^{|E|} (F(x^l) - y^l)^2$$

donde  $|E|$  es el tamaño del conjunto de datos,  $F(x^l)$  es la salida del SBRD definido por el cromosoma considerando el ejemplo  $l$  y  $y^l$  es la salida esperada. El sistema de inferencia difuso considerado para obtener  $F(x^l)$  es el de centro de gravedad ponderado por el grado de pertenencia como operador de “defuzificación” y la t-norma del mínimo como operadores de implicación y conjunción.

- **Inicialización de la población:** Con el fin de introducir el mayor número de combinaciones de granularidad y diversidad, la población se ha inicializado en dos grupos:
  - En el primero, cada uno de los cromosomas tiene el mismo número de etiquetas para todas las variables de entrada. Para incluir diversidad en  $C_1$  dichos cromosomas se han generado considerando las 6 combinaciones posibles de antecedente, de 2 a 7 etiquetas para todas las variables. Partiendo de estos antecedentes se han generado todas las combinaciones posibles con los diferentes consecuentes, que en total son 6 combinaciones por cada una. Además, para cada una de las combinaciones mencionadas se han realizado dos copias, cada una con partes de  $C_2$  diferentes. La primera de ellas tendrá valores aleatorios entre  $[0, 0, 0, 1]$



y la segunda entre  $[-0,1, 0,0]$ , por lo que en total se obtendrán 72 ( $6*6*2$ ) cromosomas diferentes.

- En el segundo se generan todos los individuos de manera aleatoria, utilizando valores en el intervalo  $[2, 7]$  para  $C_1$  y en el intervalo  $[-0,1, 0,1]$  para  $C_2$ .

Además, excepto en problemas de menos de 3 variables, se elimina de forma aleatoria una de las variables  $v$  de entrada, estableciendo el gen  $L^v = 1$  en el primer individuo. Esta acción se repite hasta que no quedan más de 5 variables en el individuo. Si el problema no posee al menos 5 variables, solo se elimina una de ellas. Este procedimiento se aplica a todos los individuos de la población de tal forma que no se obtengan soluciones con un número excesivo de variables.

- **Operadores de cruce y mutación:** El operador de cruce se aplica de forma diferente dependiendo de la parte del cromosoma que se vaya a cruzar. Para la parte  $C_1$  se delimita aleatoriamente un punto de cruce y se aplica el operador de cruce básico. Para la parte  $C_2$  se aplica un operador basado en  $BLX-\alpha$  llamado PCBLX (Parent Centric BLX). De esta forma se generan 4 nuevos individuos, combinando los 2 hijos obtenidos en la parte  $C_1$  con los dos obtenidos de la parte  $C_2$ . El operador de mutación se aplica para un gen aleatorio reduciendo en 1 la granularidad o seleccionando una granularidad mayor de forma aleatoria con la misma probabilidad. Si al decrecer la granularidad la mutación genera una BD con una sola variable esta no se aplica. De la misma forma se cambia de manera aleatoria un gen de  $C_2$ . Una vez aplicada la mutación solo permanecerán los dos individuos con mejor precisión.
- **Prevención del incesto:** Con el fin de introducir diversidad y evitar que el algoritmo converja rápido se ha incluido un mecanismo de prevención del incesto, siguiendo los conceptos de CHC y solo considerando la parte  $C_2$  del cromosoma. Basándonos en CHC, dos padres se cruzarán solo si su distancia de Hamming entre 2 está por encima del umbral preestablecido  $L$ . Dado que la segunda parte del cromosoma está formada por valores reales se necesita codificarlos para poder calcular la distancia de Hamming. Para ello se ha utilizado el Gray Code (código binario), con un número de bits por gen (BITSGENE), determinado por el sistema experto. El umbral  $L$  se inicializa de la siguiente manera:

$$L = (\#GenesC_2 \cdot BITSGENE)/4.0$$

Normalmente, el umbral se decrementa en uno cuando no hay nuevos individuos en la siguiente generación. En este caso, para acelerar la convergencia,  $L$  se reduce en dos en cada generación.

- **Reinicialización y condición de fin:** Para evitar el estancamiento en óptimos locales se aplica un mecanismo de reinicialización eliminando la población actual e incluyendo el individuo más preciso en la nueva población, generando el resto de individuos de forma aleatoria. Los valores de granularidad se eligen de forma aleatoria entre 1 y el valor de granularidad para dicha variable en el mejor individuo seleccionado. Este mecanismo se aplica cuando el umbral  $L$  llega a cero, que se vuelve a establecer en el valor inicial. El algoritmo termina cuando se alcanza un número máximo de evaluaciones o cuando dicho umbral ha llegado a cero dos veces. Esto quiere decir que solo dos etapas de exploración / explotación se aplican para llegar a la convergencia.

### Segunda fase: Post-procesamiento basado en AEMOs

Una vez se ha obtenido el modelo lingüístico definitivo, se aplica un post-procesamiento a las componentes del mismo: la base de datos, la extensión del consecuente y la selección de reglas. El nuevo método permite adaptarse a las reglas generadas en la primera parte, haciendo uso de los términos compuestos “hasta”, “desde-hasta” y “desde” y de los valores paramétricos del consecuente extendido. Este post-procesamiento está basado en el presentado en [AGH11] y aplica una versión modificada de SPEA2 [ZLT01].

#### Objetivos

Para guiar este post-procesamiento hacia el objetivo de crear SBRD más interpretables y precisos se va a utilizar la medida ECM, para medir la precisión, y número de reglas, para medir la complejidad. La interpretabilidad semántica se va a evaluar utilizando una de las medidas previamente mencionadas, GM3M o RMI.

La población de individuos estará formada por un conjunto de SBRDs que se evaluarán acorde al cumplimiento de los siguientes objetivos:

- Maximizar el valor del índice GM3M, con el fin de preservar la interpretabilidad semántica de las FPs iniciales.

- Maximizar el valor del índice RMI, que preserva o mejora la interpretabilidad semántica de las reglas.
- Minimizar el número de reglas NR para reducir la complejidad del modelo.
- Minimizar el error cuadrático medio ECM, para mejorar la precisión.

### Esquema de codificación

En este caso, se va a utilizar un sistema de codificación triple, uno para la selección de reglas  $C_S$ , otro para la codificación del ajuste de las FPs  $C_T$  y un último  $C_L$  para el ajuste de los parámetros s y d del consecuente extendido de las reglas. El cromosoma por tanto estará formado por  $C^p = C_S^p + C_T^p + C_L^p$ .

La codificación de  $C_S$  consiste en un vector binario de longitud  $m$  (número inicial de reglas), que para cada gen contendrá un 1 si se selecciona dicha regla y, de lo contrario, un 0.

Para codificar  $C_T$  se utiliza un vector de valores reales que representarán los valores de desplazamiento de las FPs. En este caso se ajustarán individualmente cada una de las FPs, por lo que también habrá que ajustar los CFLTS. Para ello, la codificación quedará de la siguiente manera:

$$C_T^p = C_1, C_2 \dots C_n$$

$$C_i = (a_1^i, b_1^i, c_1^i, \dots, a_{m^i}^i, b_{m^i}^i, c_{m^i}^i), i = 1, \dots, n$$

En el caso de  $C_L$  la codificación se realiza usando un vector de números reales con longitud  $m^2$ , que representa los parámetros s y d del consecuente extendido para las  $m$  reglas.

$$C_L^p = (c_{s_1}, \dots, c_{s_m}), (c_{d_1}, \dots, c_{d_m})$$

### Inicialización de la población

- $C_S$ : Todos los individuos comienzan con todas las reglas seleccionadas, es decir, valor 1 para todos sus genes. La selección se irá aplicando progresivamente, principalmente a través de la mutación.
- $C_T$ : El primer individuo posee la codificación de las FP del modelo inicial. El resto se inicializan de forma aleatoria teniendo en cuenta los rangos establecidos en la literatura para las FP triangulares (ver [AGH11] y Sección 2.3.2).
- $C_L$ : Al igual que antes, se inicializa el primer individuo con los valores de s y d propuestos en el modelo inicial. Para el resto de individuos se

seleccionan valores aleatorios en el rango de variación de los parámetros lineales del consecuente extendido calculados en la primera fase.

### Operadores de cruce y mutación

A la hora de cruzar cromosomas que poseen varias partes con codificaciones diferentes es necesario definir un operador de cruce diferente para cada una de ellas. En este caso, el operador de cruce aplicado a las partes  $C_T$  y  $C_L$  es el conocido BLX-0.5 [ES93]. Una vez obtenida la parte  $C_T$ , para la parte  $C_S$  se aplica el operador de cruce utilizado para la selección de reglas en [AGH11].

Mediante estos cruces se obtienen 4 descendientes a los que se aplica el operador de mutación a cada una de las partes del cromosoma de forma independiente. Para las partes reales,  $C_T$  y  $C_L$ , el operador cambia un valor de forma aleatoria teniendo en cuenta el rango de variación de los parámetros. Para la parte de la selección de reglas  $C_S$  el operador pone valor 0 a un gen aleatorio. Una vez aplicada la mutación, se evalúan los individuos y se seleccionan los dos que mejor precisión tengan.

### Mecanismos especiales para mejorar el balance precisión-interpretabilidad

El algoritmo propuesto utiliza un mecanismo de selección *SPEA2* [ZLT01] al que se le han hecho las siguientes modificaciones con el fin de mejorar sus capacidades de búsqueda:

- Mecanismo de prevención del incesto: Basándonos en los conceptos de CHC [Esh91], el método utiliza los mecanismos como se describe en [GAH10]. A diferencia de la propuesta de la referencia, en este caso solo se previene la convergencia prematura en la parte de  $C_T$  y solo se cruzarán los padres cuya distancia de Hamming entre 4 sea menor que un umbral. Para los cruces se ha utilizado un esquema de codificación real, por tanto es necesario transformar cada uno de los genes a código Gray con un número fijo de bits por gen (Bgen). Utilizando este valor, el cálculo del umbral se inicializa de la siguiente forma  $L = \frac{\#C_t * Bgen}{4}$ , donde  $C_t$  es el número de genes de la parte  $C_t$  del cromosoma. En cada una de las generaciones el valor del umbral se decrementa en una unidad, permitiendo que cada vez se puedan cruzar individuos más parecidos.
- Operador de reinicialización: Para generar las nuevas poblaciones ex-

ternas, se realiza una copia de los individuos con mejor precisión y dos de los individuos que tienen mejor valor para las otras medidas consideradas (GM3M, RMI y NR). En total, permanecen 7 individuos, mientras que el resto de la población se reinician tomando los mismos valores de  $C_s$  que las mejores soluciones e inicializando de forma aleatoria los valores de  $C_T$  y  $C_L$ . Esta reinicialización se lleva a cabo cuando el umbral  $L$  toma valor por debajo de cero. Además, este proceso de reinicialización se elimina en las evaluaciones finales del algoritmo y, si nunca se ha aplicado antes, se desactiva en la mitad de las evaluaciones.

### 4.3. Estudio experimental

En esta sección se va a evaluar cómo se comporta la propuesta desde el punto de vista de la interpretabilidad y la precisión, así como la utilidad en un problema biomédico real similar al de la sección anterior.

#### 4.3.1. Comparativa con otros métodos de la literatura

Para llevar a cabo esta experimentación se ha seleccionado una lista de 23 conjuntos de datos de regresión. En la Tabla 4.1 se pueden ver las características de dichos conjuntos de datos, la columna ACRO representa el acrónimo del nombre del conjunto de datos, en #Var el número de variables que tiene (entre 2 y 60) y en #Ejs el número de ejemplos (entre 43 y 4177). Los conjuntos de datos han sido seleccionados de las páginas: UCI Machine Learning Repository<sup>1</sup>, KEEL-dataset<sup>2</sup>, Dataset Collections of Weka<sup>3</sup> and Luis Torgo Repository<sup>4</sup>.

Para la realización de todos los experimentos se ha utilizado un modelo de validación cruzada con 5-folds, cada una con un 20% de los ejemplos del conjunto de datos. De esas 5 particiones se utilizan 4 para entrenamiento y una para validación. Para cada una de las 5 combinaciones de particiones se ha ejecutado el método 6 veces con 6 semillas diferentes. En la web relacionada con esta propuesta (<https://www.ugr.es/~ralcala/papers/ijfs21/> se

---

<sup>1</sup><https://archive.ics.uci.edu/ml/index.php>

<sup>2</sup><https://sci2s.ugr.es/keel/datasets.php>

<sup>3</sup><http://www.cs.waikato.ac.nz/ml/Weka/datasets.html>

<sup>4</sup><http://www.dcc.fc.up.pt/~ltorgo/Regression/DataSets.html>

Tabla 4.1: Conjuntos de datos de regresión

| NOMBRE                  | ACRO | #Var | #Ejs |
|-------------------------|------|------|------|
| Abalone                 | ABA  | 8    | 4177 |
| Anacalt                 | ANA  | 7    | 4052 |
| Baseball                | BAS  | 16   | 337  |
| Boston housing          | BOS  | 13   | 506  |
| Diabetes                | DIA  | 2    | 43   |
| Machine CPU             | CPU  | 6    | 209  |
| Electrical Maintenance  | ELE  | 4    | 1056 |
| Body fat                | FAT  | 14   | 252  |
| Forest Fires            | FOR  | 12   | 517  |
| Friedman                | FRI  | 5    | 1200 |
| Mortgage                | MOR  | 15   | 1049 |
| Auto Mpg 6              | MPG6 | 5    | 392  |
| Auto Mpg 8              | MPG8 | 7    | 392  |
| AutoPrice               | PRI  | 15   | 159  |
| Quake                   | QUA  | 3    | 2178 |
| Stocks domain           | STP  | 9    | 950  |
| Strike                  | STR  | 6    | 625  |
| Treasury                | TRE  | 15   | 1049 |
| Triazines               | TRI  | 60   | 186  |
| Weather Ankara          | WAN  | 9    | 1609 |
| Weather Izmir           | WIZ  | 9    | 1461 |
| Wisconsin Breast Cancer | WBC  | 32   | 194  |
| Yacht Hydrodynamics     | YH   | 6    | 308  |

encuentran los conjuntos de datos utilizados con las particiones realizadas para cada uno.

La Tabla 4.2 muestra un breve resumen de los métodos considerados para la experimentación y de sus características principales (MedInt indica si el algoritmo usa y optimiza medidas de interpretabilidad).  $TS_{SP2-SI}$  [GAH10] y LINGI [GGSA14] son dos métodos del estado del arte utilizados para obtener SBRDs simples, interpretables y con buena precisión, que optimizan las medidas de interpretabilidad de semántica y complejidad. Es por ello que son los métodos de la literatura actual que más se ajustan a la comparativa que se va a realizar, ya que han sido diseñados para buscar la transparencia.  $FS_{mogfs^e} + Tun^e$  [AGH11] es un AEMO del estado del arte orientado a obtener una buena precisión sin tener en cuenta la interpretabilidad, de hecho, de los algoritmos existentes es el que genera los SBRD con mejor precisión. Aunque no es un método orientado a interpretabilidad se ha decidido considerarlo porque es un método que obtiene SBRD precisos.

Tabla 4.2: Métodos considerados para la comparativa

| Método                                     | MedInt | Descripción   | Objetivos       |
|--|--------|---|-----------------|
| <b>TS<sub>SP2-SI</sub></b>                 | Si     | AEMO para ajuste y selección de reglas  | ECM/NR/GM3M     |
| <b>LING1</b>                               | Si     | AEMO TS con L-IRL   | ECM/NR/GM3M/RMI |
| <b>FSmogfs<sup>e</sup>+Tun<sup>e</sup></b> | No     | AEMO para aprendizaje embebido de la BD<br>generación de la BR con WM <sup>5</sup> + TS AEMO  | ECM/NR          |
| <b>Método propuesto</b>                    | Si     | AEMO basado en CFLT5 para aprendizaje embebido<br>de la BR con aprendizaje de la BR basado en árboles + TS<br>AEMO de la BR basado en árboles + TS AEMO | ECM/NR/GM3M/RMI |

La Tabla 4.3 incluye los valores utilizados para los parámetros generales del método, que no dependen del conjunto de datos por lo que no es necesario modificarlos para nuevos problemas (son los valores estándar recomendados para los parámetros).

Tabla 4.3: Parámetros utilizados

| Parámetro                      | Valor                              |
|--------------------------------|------------------------------------|
| Probabilidad de mutación       | $P_m = 0,2$                        |
| Tamaño de la población         | 200                                |
| Tamaño de la población externa | 61                                 |
| Número de evaluaciones         | 100000 (en total para ambas fases) |
| Bits por gen                   | $B_{Gene} = 30$                    |

Para evaluar las diferencias significativas de los resultados se han aplicado análisis estadísticos basados en tests no paramétricos. En particular se van a utilizar test no paramétricos para comparaciones múltiples [Dem06, GH08]: test de Friedman y método de Holm. Además, se ha aplicado test de Wilcoxon para la comparación dos a dos [Dem06, GH08]. En <http://sci2s.ugr.es/sicidm/> se puede ver una descripción detallada de estos tests.

### Comparativa en base a medidas de interpretabilidad semántica

Con el fin de comparar la efectividad de la nueva propuesta se ha realizado una comparativa con dos métodos de la literatura que también obtienen SBRDs optimizando las medidas de interpretabilidad semántica y complejidad. Dichos métodos son *TS<sub>SP2-SI</sub>* [GAH10] y *LING1* [GGSA14].

En la Tabla 4.4 se muestran los resultados de la experimentación para los métodos con los que se ha realizado la comparativa y en la Tabla 4.5

Tabla 4.4: Resultados para los métodos de la comparativa

| Conjunto de datos | TS SP2-SI |         |             |       |       | LING1 |         |             |       |       |
|-------------------|-----------|---------|-------------|-------|-------|-------|---------|-------------|-------|-------|
|                   | Var       | NR      | $ECM_{tst}$ | GM3M  | RMI   | Var   | NR      | $ECM_{tst}$ | GM3M  | RMI   |
| ABA               | 8         | 16,27   | 2,513       | 0,45  | 0,692 | 8     | 13      | 2,78        | 0,47  | 0,66  |
| ANA               | 7         | 88,17   | 0,006       | 0,204 | 0,154 | 7     | 155,27  | 0,006       | 0,47  | 0     |
| BAS               | 16        | 83,37   | 389547      | 0,283 | 0,571 | 16    | 271,27  | 440395      | 0,411 | 0,584 |
| BOS               | 13        | 173,33  | 13,957      | 0,325 | 0,384 | 13    | 13248,6 | 14,458      | 0,414 | 0,384 |
| CPU               | 6         | 23,47   | 2246,11     | 0,395 | 0,564 | 6     | 31,6    | 2229,65     | 0,466 | 0,583 |
| DIA               | 2         | 10,67   | 0,268       | 0,171 | 0,468 | 2     | 18,57   | 0,263       | 0,364 | 0,277 |
| ELE               | 4         | 29,3    | 14851       | 0,528 | 0,504 | 4     | 32,5    | 18822       | 0,54  | 0,54  |
| FAT               | 14        | 83,23   | 5,292       | 0,673 | 0,311 | 14    | 64,3    | 4,831       | 0,705 | 0,733 |
| FOR               | 12        | 102,83  | 2211        | 0,215 | 0,734 | 12    | 507,47  | 4467        | 0,473 | 0,566 |
| FRI               | 5         | 494,83  | 2,047       | 0,712 | 0,1   | 5     | 707,57  | 1,933       | 0,702 | 0,308 |
| MOR               | 15        | 15,4    | 0,034       | 0,541 | 0,744 | 15    | 9       | 0,045       | 0,6   | 0,97  |
| Mpg6              | 5         | 53,27   | 5,034       | 0,314 | 0,305 | 5     | 65,93   | 5,694       | 0,463 | 0,332 |
| Mpg8              | 7         | 82,67   | 5,436       | 0,289 | 0,395 | 7     | 107,3   | 5,037       | 0,463 | 0,452 |
| PRI               | 15        | 59,07   | 4104134     | 0,236 | 0,65  | 15    | 90,67   | 4816082     | 0,434 | 0,65  |
| QUA               | 3         | 27,2    | 0,0182      | 0,275 | 0,512 | 3     | 96,73   | 0,0185      | 0,47  | 0     |
| STP               | 9         | 32,87   | 0,775       | 0,365 | 0,513 | 9     | 14,1    | 1,64        | 0,47  | 0,76  |
| STR               | 6         | 125,03  | 225851      | 0,644 | 0,332 | 6     | 174,17  | 248910      | 0,644 | 0,095 |
| TRE               | 15        | 17,67   | 0,048       | 0,533 | 0,746 | 15    | 9       | 0,055       | 0,63  | 0,98  |
| TRI               | 60        | 119,73  | 0,0133      | 0,119 | 0,568 | 60    | 170,37  | 0,0126      | 0,36  | 0,382 |
| WAN               | 9         | 39,33   | 2,016       | 0,456 | 0,482 | 9     | 9,5     | 2,825       | 0,57  | 0,91  |
| WIZ               | 9         | 29,2    | 1,095       | 0,493 | 0,487 | 9     | 13      | 1,525       | 0,61  | 0,93  |
| WBC               | 32        | 142,73  | 948,41      | 0,218 | 0,754 | 32    | 246,567 | 928,64      | 0,397 | 0,71  |
| YH                | 6         | 81,2    | 25,677      | 0,725 | 0,565 | 6     | 182,8   | 21,341      | 0,694 | 0,088 |
| Media             | 12,09     | 83,95   | -           | 0,398 | 0,502 | 12,09 | 140,84  | -           | 0,514 | 0,517 |
| % empeoramiento   | 510,1     | 1619,35 | -           | 40,82 | 45,39 | 510,1 | 2784,48 | -           | 23,67 | 43,69 |

por cada columna principal en la que se pueden ver los métodos se incluyen todas las medias obtenidas para cada uno de los conjuntos de datos. La columna Var muestra la media de variables utilizadas, NR el número medio de reglas finales,  $ECM_{tst}$  el error cuadrático medio para los datos de test y las columnas GM3M y RMI la media de las medidas de interpretabilidad. Las dos últimas filas muestran la media global (Media) y el porcentaje de empeoramiento con respecto a la nueva propuesta (% empeoramiento).

Los resultados muestran que el método propuesto obtiene los mejores resultados promedio en todas las medidas de interpretabilidad analizadas y supera de manera significativa a los métodos de la comparativa. Con respecto a los resultados de los conjuntos de datos individuales, el método propuesto genera modelos más precisos que los otros dos.

La Tabla 4.6 muestra una clasificación de los métodos basada en el test de Friedman. Se puede observar que para todas las medidas consideradas,



Tabla 4.5: Resultados para la nueva propuesta

| Conjunto de datos | Nueva propuesta |      |             |       |       |
|-------------------|-----------------|------|-------------|-------|-------|
|                   | Var             | NR   | $ECM_{lst}$ | GM3M  | RMI   |
| ABA               | 1,2             | 3,67 | 2,456       | 0,777 | 0,98  |
| ANA               | 1               | 3,93 | 0,004       | 0,4   | 0,894 |
| BAS               | 2               | 5,73 | 243569      | 0,606 | 0,954 |
| BOS               | 3,1             | 6,93 | 8,276       | 0,619 | 0,805 |
| CPU               | 2,2             | 4,93 | 1739,53     | 0,7   | 0,935 |
| DIA               | 1,7             | 3,5  | 0,179       | 0,611 | 0,916 |
| ELE               | 2               | 5,37 | 12184       | 0,679 | 0,928 |
| FAT               | 1,6             | 3,83 | 1,292       | 0,862 | 0,973 |
| FOR               | 1               | 2,07 | 2351        | 0,43  | 0,826 |
| FRI               | 2,9             | 6,8  | 3,041       | 0,778 | 0,956 |
| MOR               | 3,1             | 6    | 0,017       | 0,696 | 0,708 |
| Mpg6              | 2               | 5,4  | 4,524       | 0,717 | 0,947 |
| Mpg8              | 2,3             | 6,07 | 4,224       | 0,647 | 0,96  |
| PRI               | 2,4             | 5,9  | 2700226     | 0,683 | 0,961 |
| QUA               | 1               | 1,63 | 0,0179      | 0,547 | 0,89  |
| STP               | 2,7             | 5,67 | 1,396       | 0,708 | 0,962 |
| STR               | 1,4             | 3,7  | 172589      | 0,702 | 0,906 |
| TRE               | 1,9             | 4,67 | 0,038       | 0,713 | 0,917 |
| TRI               | 2,7             | 5,73 | 0,0116      | 0,735 | 0,964 |
| WAN               | 2               | 5,67 | 1,506       | 0,699 | 0,959 |
| WIZ               | 2,3             | 5,3  | 1,015       | 0,729 | 0,981 |
| WBC               | 1,4             | 4,27 | 747,16      | 0,615 | 0,956 |
| YH                | 1,6             | 5,53 | 0,895       | 0,829 | 0,843 |
| Media             | 1,98            | 4,88 | -           | 0,673 | 0,918 |
| % empeoramiento   |                 |      |             |       |       |

el nuevo método es el que ocupa la primera posición. En la Tabla 4.7 se muestran los p-valores obtenidos con el test de Holms, comparando todos los métodos dos a dos con el método propuesto. Los resultados muestran que la nueva propuesta supera a los demás métodos.

Tabla 4.6: Rankings del test de Friedman's para las medidas Var, NR, GM3M, RMI y ECM en Tst.

| Algoritmo            | Rank.<br>Var | Rank.<br>NR | Rank.<br>GM3M | Rank.<br>RMI | Rank.<br>Tst |
|----------------------|--------------|-------------|---------------|--------------|--------------|
| NUEVA PROPUESTA      | 1,0          | 1,00        | 1,09          | 1,13         | 1,17         |
| TS <sub>SP2-SI</sub> | 2,5          | 2,30        | 2,89          | 2,52         | 2,24         |
| LING1                | 2,5          | 2,70        | 2,02          | 2,35         | 2,59         |

Tabla 4.7: P-valores ajustados contra la *Nueva propuesta* para las medidas Var, NR, GM3M y RMI y ECM en Tst.

| Algoritmo            | apv<br>Var | apv<br>NR | apv<br>GM3M | apv<br>RMI | apv<br>Tst |
|----------------------|------------|-----------|-------------|------------|------------|
| TS <sub>SP2-SI</sub> | 7,28E-7    | 9,72E-6   | 1,88E-9     | 4,76E-6    | 3,03E-4    |
| LING1                | 7,28E-7    | 1,78E-8   | 1,52E-3     | 3,65E-5    | 3,30E-6    |

En la web de esta propuesta (link a la web) se pueden encontrar algunos de los modelos lingüísticos obtenidos, incluyendo los resultados y un análisis de la interpretabilidad para dos problemas conocidos: WAN (Weather in Ankara) y WBC (Winsconsin Breast Cancer).

#### 4.3.1.1. Comparativa con un método enfocado puramente en la precisión

En esta sección se va a comparar la nueva propuesta con un método enfocado a la precisión. *FSmogfs<sup>e</sup> + Tun<sup>e</sup>* [AGH11] es un AEMO que realiza un aprendizaje embebido de la BD, una generación de la BR usando el algoritmo de Wang y Mendel [WM92] y un post procesamiento para realizar un ajuste evolutivo multiobjetivo de las FPs y la selección de reglas. Este algoritmo está centrado en optimizar las medidas de ECM y NR, aunque en realidad utiliza ambas para obtener modelos precisos sin restricciones de interpretabilidad. Aunque en términos de precisión no son dos algoritmos comparables (ya que el algoritmo propuesto asume una pérdida de precisión

Tabla 4.8: Resultados de la comparación con un método lingüístico y orientado únicamente a la precisión.

| Conjuntos de datos | FSmogfs <sup>e</sup> +Tun <sup>e</sup> |        |                    |       |       | Método propuesto |             |                    |              |              |
|--------------------|--|--------|--------------------|-------|-------|------------------|-------------|--------------------|--------------|--------------|
|                    | Var                                    | NR     | ECM <sub>lst</sub> | Gm3M  | RMI   | Var              | NR          | ECM <sub>lst</sub> | Gm3m         | Rmi          |
| ABA                | 3,0                                    | 8,00   | 2,509              | 0,326 | 0,316 | <b>1,2</b>       | <b>3,67</b> | <b>2,456</b>       | <b>0,777</b> | <b>0,980</b> |
| ANA                | 3,0                                    | 10,13  | <b>0,003</b>       | 0,244 | 0,319 | <b>1,0</b>       | <b>3,93</b> | 0,004              | <b>0,400</b> | <b>0,894</b> |
| BAS                | 6,0                                    | 16,60  | 261323             | 0,212 | 0,450 | <b>2,0</b>       | <b>5,73</b> | <b>243569</b>      | <b>0,606</b> | <b>0,954</b> |
| BOS                | 4,6                                    | 21,03  | 9,909              | 0,187 | 0,319 | <b>3,1</b>       | <b>6,93</b> | <b>8,276</b>       | <b>0,619</b> | <b>0,805</b> |
| CPU                | 4,2                                    | 15,57  | 2390,30            | 0,281 | 0,576 | <b>2,2</b>       | <b>4,93</b> | <b>1739,53</b>     | <b>0,700</b> | <b>0,935</b> |
| DIA                | 2                                      | 11,63  | 0,261              | 0,197 | 0,430 | <b>1,7</b>       | <b>3,5</b>  | <b>0,179</b>       | <b>0,611</b> | <b>0,916</b> |
| ELE                | <b>2,0</b>                             | 8,00   | <b>10548</b>       | 0,436 | 0,559 | <b>2,0</b>       | <b>5,37</b> | 12184              | <b>0,679</b> | <b>0,928</b> |
| FAT                | 2,2                                    | 8,53   | 1,378              | 0,749 | 0,712 | <b>1,6</b>       | <b>3,83</b> | <b>1,292</b>       | <b>0,862</b> | <b>0,973</b> |
| FOR                | 3,0                                    | 10,00  | 2628               | 0,166 | 0,728 | <b>1,0</b>       | <b>2,07</b> | <b>2351</b>        | <b>0,430</b> | <b>0,826</b> |
| FRI                | 3,1                                    | 22,03  | 3,138              | 0,754 | 0,314 | <b>2,9</b>       | <b>6,80</b> | <b>3,041</b>       | <b>0,778</b> | <b>0,956</b> |
| MOR                | <b>2,0</b>                             | 7,00   | 0,019              | 0,419 | 0,612 | 3,1              | <b>6,00</b> | <b>0,017</b>       | <b>0,696</b> | <b>0,708</b> |
| Mpg6               | 3,0                                    | 20,00  | 4,562              | 0,173 | 0,319 | <b>2,0</b>       | <b>5,40</b> | <b>4,524</b>       | <b>0,717</b> | <b>0,947</b> |
| Mpg8               | 3,0                                    | 23,00  | 4,747              | 0,199 | 0,280 | <b>2,3</b>       | <b>6,07</b> | <b>4,224</b>       | <b>0,647</b> | <b>0,960</b> |
| PRI                | 5,3                                    | 24,03  | 3344230            | 0,242 | 0,459 | <b>2,4</b>       | <b>5,9</b>  | <b>2700226</b>     | <b>0,683</b> | <b>0,961</b> |
| QUA                | 1,3                                    | 3,23   | <b>0,0178</b>      | 0,091 | 0,668 | <b>1,0</b>       | <b>1,63</b> | 0,0179             | <b>0,547</b> | <b>0,890</b> |
| STP                | 3,0                                    | 23,00  | <b>0,912</b>       | 0,197 | 0,307 | <b>2,7</b>       | <b>5,67</b> | 1,396              | <b>0,708</b> | <b>0,962</b> |
| STR                | 3,6                                    | 19,77  | 187917             | 0,667 | 0,422 | <b>1,4</b>       | <b>3,70</b> | <b>172589</b>      | <b>0,702</b> | <b>0,906</b> |
| TRE                | 3,0                                    | 9,00   | 0,044              | 0,396 | 0,672 | <b>1,9</b>       | <b>4,67</b> | <b>0,038</b>       | <b>0,713</b> | <b>0,917</b> |
| TRI                | 9,5                                    | 28,00  | 0,0119             | 0,179 | 0,393 | <b>2,7</b>       | <b>5,73</b> | <b>0,0116</b>      | <b>0,735</b> | <b>0,964</b> |
| WAN                | 2,1                                    | 8,13   | 1,635              | 0,434 | 0,641 | <b>2,0</b>       | <b>5,67</b> | <b>1,506</b>       | <b>0,699</b> | <b>0,959</b> |
| WIZ                | <b>2,0</b>                             | 10,00  | <b>1,011</b>       | 0,370 | 0,607 | 2,3              | <b>5,30</b> | 1,015              | <b>0,729</b> | <b>0,981</b> |
| WBC                | 5,5                                    | 25,50  | 766,89             | 0,180 | 0,185 | <b>1,4</b>       | <b>4,27</b> | <b>747,16</b>      | <b>0,615</b> | <b>0,956</b> |
| YH                 | 2,2                                    | 11,40  | 1,500              | 0,750 | 0,392 | <b>1,6</b>       | <b>5,53</b> | <b>0,895</b>       | <b>0,829</b> | <b>0,843</b> |
| Media              | 3,42                                   | 14,94  | -                  | 0,341 | 0,464 | <b>2,0</b>       | <b>4,88</b> | -                  | <b>0,673</b> | <b>0,918</b> |
| % empeoramiento    | 72,71                                  | 205,97 |                    | 49,29 | 49,44 |                  |             |                    |              |              |

a cambio de mejorar la transparencia), es interesante compararlos para demostrar que se puede mejorar ambos en conjunto, obteniendo modelos con mucha mas precisión y transparencia.

La Tabla 4.8 muestra los resultados de la comparativa de  $FSmogfs^e + Tun^e$  y la nueva propuesta (las medidas GM3M y RMI no estaban consideradas en el algoritmo originalmente, por lo que se han calculado sobre los modelos finales obtenidos). Las columnas de esta tabla representan lo mismo que las de la Tabla 4.4. Como se puede observar en la penúltima fila de la tabla (Media), la nueva propuesta obtiene mejores valores de forma global para todas las medidas de interpretabilidad. Además, en precisión se puede ver que también obtiene un mejor valor de ECM para la mayoría de los conjuntos de datos.

### 4.3.2. Caso de estudio en un problema real relacionado con la obesidad en niños

En este estudio, se utiliza un conjunto de datos obtenidos a partir de una cohorte de 1014 niños de España, con edades comprendidas entre los 5 y los 15 años, que fueron divididos en tres grupos experimentales en función de su estado ponderal: peso normal, sobrepeso y niños con obesidad. La población del estudio está formada por 525 sujetos en el grupo de obesidad, 201 en el grupo de sobrepeso y 288 en el grupo de peso normal. Se han recopilado datos clínicos y moleculares para cada grupo, incluidos más de 850.000 marcadores genéticos y epigenéticos, más de 50 mediciones antropométricas y bioquímicas, así como datos sobre el estilo de vida y la actividad física (AF) obtenidos mediante cuestionarios de frecuencia alimentaria (CFA) y acelerómetros. Como objetivo específico del estudio se pretende investigar la relación entre la antropometría, la AF y el estado de insulinoresistencia (IR) en el conjunto de datos de niños. Para este análisis, se va a seleccionar una muestra de 460 individuos que tienen suficientes datos de AF de calidad disponibles. El índice Homeostasis Model Assessment for Insulin Resistance (HOMA-IR), un indicador del estado de IR en niños y adultos, será la variable de resultado del modelo. Las variables predictoras del modelo incluyen datos clínicos y antropométricos (como sexo, edad, pubertad, altura, diámetro de la cintura (DC) e índice de masa corporal IMC), mediciones principales de AF (como tiempo sedentario (TS) o AF ligera, moderada y vigorosa), así como indicadores bioquímicos de disfunción cardiometabólica que difieren del HOMA-IR en niños obesos (como colesterol HDL y LDL, triglicéridos y presión arterial). A diferencia de otros conjuntos de datos que pueden tener limitaciones debido a las medidas de AF y TS autoinformadas, se utilizan medidas de AF más objetivas, como la acelerometría, que ayudaron a reducir el sesgo del estudio.

En la Tabla 4.9 se pueden observar los resultados obtenidos utilizando una validación cruzada con 5 particiones. Para este problema se observa que la nueva propuesta tiene unos buenos valores para las medidas seleccionadas y que además obtiene mejores resultados en comparación con el resto de métodos aplicados. A modo de demostrar la comprensibilidad de los modelos generados, en la Figura 4.9 se puede ver el modelo resultante sobre este problema. El resultado de este modelo ha sido analizado por un experto, a diferencia de los otros ejemplos de modelos generados que se incluyen en la web de la propuesta (<https://www.ugr.es/~ralcala/papers/ijfs21>). Las variables de la figura se han ordenado de tal forma que representen el mismo

Tabla 4.9: Resultados en el problema de obesidad infantil

| Métodos                                | Var         | NR          | ECM <sub>tst</sub> | Gm3m         | Rmi          |
|--|-------------|-------------|--------------------|--------------|--------------|
| Método propuesto                       | <b>1,83</b> | <b>4,37</b> | <b>0,911</b>       | <b>0,716</b> | <b>0,956</b> |
| FSmogfs <sup>e</sup> +Tun <sup>e</sup> | 3,73        | 12,30       | 0,950              | 0,242        | 0,425        |
| TS <sub>SP2-SI</sub>                   | 15,00       | 320,37      | 1,430              | 0,314        | 0,649        |
| LING1                                  | 15,00       | 590,97      | 1,450              | 0,377        | 0,641        |

orden que las particiones generadas por el árbol a la hora de aprender las reglas. De esta forma, se puede considerar que cada una de las divisiones es una forma de considerar diferentes divisiones en los datos, desde más general a más específico. Los colores se han utilizado para reconocer facilmente los diferentes casos representados por las reglas, siendo cada uno de los colores una variable y una división. Los textos en gris se han incluido solamente para añadir información adicional pero no es parte de la estructura de la regla propuesta. De la misma manera, se ha añadido el porcentaje de ejemplos cubiertos, para GM3M y RMI, debido a que son valores muy informativos y nos dan información semántica de la calidad de cada una de las particiones y de la regla asociada respectivamente.

Como ya se ha explicado previamente, la medida RMI va desde 1,0 a 0,0, representando 1,00 que lo que la regla afirma es igual a lo que produce el modelo. En los resultados se puede observar que todas las reglas poseen un RMI igual a 1,0 y que las variables poseen un valor de GM3M superior a 0,8 (casi todas ellas pueden estar consideradas como particiones lingüísticas fuertes), lo que se traduce en que la semántica se preserva totalmente. Además, en este caso el método propuesto fue capaz de describir correctamente el comportamiento general del conjunto de datos y de identificar las relaciones principales entre los predictores. Adicionalmente, los valores de corte propuestos por el método para la discretización de las variables no solo representa el valor real de los dominios si no que además tienen sentido desde un punto de vista clínico.

La principal relación que el modelo ha sido capaz de extraer ha sido entre HOMA-IR y el estado de IMC. Inicialmente, el método sugiere dividir la población en dos grupos de niños acorde al valor de IMC ( $\sim >$  o  $\sim <$  que 31). Esta información nos dice que los niños con valores de IMC de aproximadamente menos de 31 suelen tener valores menores de HOMA-IR (y viceversa). Es interesante destacar que, contrastando este resultado con la realidad, se puede ver que el IMC ha sido declarado recientemente como factor de riesgo de IR en poblaciones de niños [LOD<sup>+</sup>06]. Dentro del grupo que poseen

Figura 4.9: BC resultante de la nueva propuesta ( $ECM_{tst}=0,706$ )

|  |
|--|
| <p><b>R1<sup>RMI:1.0</sup></b> (3.5% de las instancias):<br/>           Si <b>BMI</b> desde <b>Med</b> Y <b>Trig</b> desde <b>Med</b><br/>           ENTONCES <b>HOMA-IR</b> Entre <b>Bajo</b> y <b>Alto</b> (centrado en <b>5.2</b>)<br/>           +/- <b>0.18</b> por +/- unidad en <b>Height</b> desde <b>155 cm</b> (moving en [Bajo, Alto])</p>                      |
| <p><b>R2<sup>RMI:1.0</sup></b> (6.4% de las instancias)<br/>           Si <b>BMI</b> desde <b>Med</b> Y <b>Trig</b> hasta <b>Bajo</b><br/>           ENTONCES <b>HOMA-IR</b> entre <b>Bajo</b> y <b>Alto</b> (centrado en <b>5.2</b>)<br/>           +/- <b>0.24</b> por +/- unidad en <b>BMI</b> desde <b>41.0</b> (moving en [Bajo,Alto])</p>                            |
| <p><b>R3<sup>RMI:1.0</sup></b> (26.5% de las instancias)<br/>           Si <b>BMI</b> hasta <b>Bajo</b> Y <b>Vigorous</b> hasta <b>Med</b> Y <b>Trig</b> desde <b>Bajo</b><br/>           ENTONCES <b>HOMA-IR</b> es <b>Bajo</b> (centrado en <b>2.3</b>)<br/>           +/- <b>0.04</b> por +/- unidad en <b>Height</b> desde <b>155 cm</b> (moving en [Bajo,Alto])</p>   |
| <p><b>R4<sup>RMI:1.0</sup></b> (61.0% de las instancias)<br/>           Si <b>BMI</b> hasta <b>Low</b> Y <b>Vigorous</b> hasta <b>Med</b> Y <b>Trig</b> hasta <b>VeryLow</b><br/>           ENTONCES <b>HOMA-IR</b> es <b>Bajo</b> (centrado en <b>2.3</b>)<br/>           +/- <b>0.04</b> por +/- unidad en <b>Height</b> desde <b>166 cm</b> (moving en [Bajo,Alto])</p> |
| <p><b>R5<sup>RMI:1.0</sup></b> (2.6% de las instancias)<br/>           Si <b>BMI</b> hasta <b>Bajo</b> Y <b>Vigorous</b> es <b>Alto</b><br/>           ENTONCES <b>HOMA-IR</b> es <b>Bajo</b> (centrado en <b>2.3</b>)<br/>           +/- <b>0.22</b> por +/- unidad en <b>BMI</b> desde <b>19.6</b> (moving en [Bajo])</p>  |

Etiquetas de las variables y definición de los puntos de las funciones de pertenencia:

**Body Mass Index (BMI) - GM3M:0.8**  
 7.4, 14.6, 19.3 ExtremadamenteBajo (ExtBaj)  
 14.1, 19.5, 24.9 MuyBajo  
 21.4, 26.4, 31.3 Bajo  
 27.6, 31.5, 36.6 Medio (Med)  
 32.0, 37.0, 40.5 Alto  
 35.7, 41.1, 46.5 MuyAlto  
 39.9, 46.2, 51.1 ExtremadamenteAlto (ExtAlt)

**Height cm**  
 Rango:  
 [100, 181]

**Vigorous - GM3M:0.8**  
 -4.1, 18.5, 36.9 Low  
 10.7, 31.3, 59.6 Med  
 29.6, 49.7, 68.7 High

**HOMA-IR - GM3M:0.9**  
 -3.0, 2.3, 7.5 Low  
 2.8, 8.0, 12.9 High

**Triglycerides mg/dL (Trig) - GM3M:0.8**  
 -21.0, 16.5, 59.7 ExtBajo  
 31.1, 73.8, 109.3 MuyBajo  
 45.4, 97.4, 146.0 Bajo  
 115.2, 152.2, 175.9 Medio  
 140.0, 179.0, 218.0 Alto  
 179.0, 218.0, 257.0 MuyAlto  
 221.1, 254.2, 291.2 ExtAlto

un IMC desde Bajo hasta Medio el método propone divisiones adicionales (reglas) basadas en el nivel de vigorosidad con respecto a la actividad física. Los niños con mayores niveles de vigorosidad por día tienen los valores más bajos de HOMA-IR. En relación con esta afirmación del modelo, estudios previos han demostrado que la actividad física de moderada a vigorosa está inversamente relacionada con el riesgo cardiometabólico en niños [SSA<sup>+</sup>17]. Otro dato interesante a observar es que el modelo no ha identificado como predictor del HOMA-IR la variable ST (tiempo de sedentarismo), este resultado se asocia a los informes previos en los que el tiempo de sedentarismo no ha sido asociado como un factor de riesgo cuando la vigorosidad en actividad física es de moderada a vigorosa [SSA<sup>+</sup>17].

En conclusión, el método ha propuesto un conjunto de reglas fáciles de comprender y consistentes tanto clínicamente como biológicamente, demostrando así un buen comportamiento en conjunto de datos biológicos complejos. Además, las reglas generadas revelan comportamientos y relaciones entre variables que se alinean con los descubrimientos previos realizados usando la estadística tradicional.

## Capítulo 5

# Adaptación de la nueva extensión de la gramática lingüística difusa a la extracción de reglas difusas

En este capítulo se desarrolla la propuesta de un nuevo tipo de elementos para construcción de RADs, llamados Meta Elementos Difusos (MED). Además, el fin de extraer reglas más genéricas basadas en este nuevo elemento se propone una extensión del método Fuzzy Apriori.

### 5.1. Reglas de Asociación Difusas

La teoría de conjuntos difusos está siendo utilizada cada vez con mayor frecuencia en los sistemas inteligentes debido a su simplicidad y similitud con el razonamiento humano. El uso de conjuntos difusos para describir asociaciones entre datos extiende el tipo de relaciones que pueden representarse, facilita la interpretación de las reglas con los términos lingüísticos y evita el uso de límites no naturales en el particionamiento del dominio de los atributos, dando lugar a lo que se conoce como reglas de asociación difusas



(RADs). Las reglas de asociación difusas se pueden definir como una extensión de las reglas de asociación clásicas. La diferencia se encuentra en la forma de definir los items, o elementos, con los que se generan los itemsets, o conjunto de elementos, que formarán dichas reglas. Para ello, para cada variable se define una partición difusa del dominio de la misma, asignándole una etiqueta lingüística a cada uno de los conjuntos difusos. La Figura 5.1 muestra un ejemplo para la variable Altura con 5 etiquetas lingüísticas con FPs triangulares. Por tanto, cada item estará formado por un par (variable, etiqueta). De esta forma, se pueden definir reglas como por ejemplo "SI la edad es muy joven ENTONCES la altura es Baja". En este ejemplo se puede ver la facilidad para interpretar la regla gracias a las etiquetas lingüísticas.

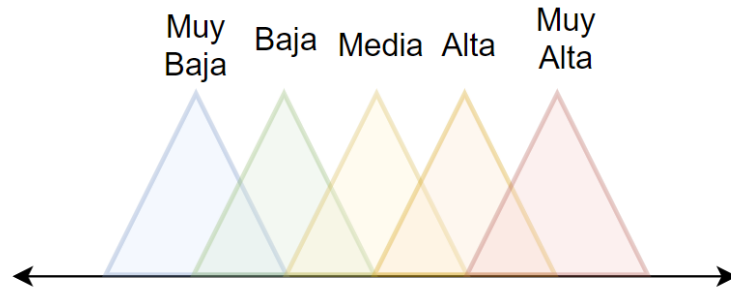


Figura 5.1: Representación difusa de la variable Altura

El cálculo del soporte de un itemset se ha definido para tener en cuenta la FP de cada etiqueta (ver Figura 5.2). Así, el soporte de un itemset difuso es igual a la suma del grado de pertenencia de los ejemplos de la BDs cubiertos por el itemset dividido por el total de ejemplos de la BD. Destacar que para determinar el grado de pertenencia de un ejemplo normalmente se utiliza la t-norma mínimo por itemset que constan de más de un item. Equivalentemente, las medidas de soporte y confianza de una regla son redefinidas como:

$$Soporte(A \rightarrow B) = \frac{\sum_{i=0}^D \mu_{ab}(i)}{D}$$

$$Confianza(A \rightarrow B) = \frac{\sum_{i=0}^D \mu_{ab}(i)}{\sum_{i=0}^D \mu_a(i)}$$

donde D es el número total de transacciones en la BDs,  $\mu_a(i)$  el grado

de pertenencia del ejemplo  $i$  con el antecedente de la regla y  $\mu_{ab}(i)$  el grado de pertenencia del ejemplo  $i$  con el antecedente y consecuente de la regla.

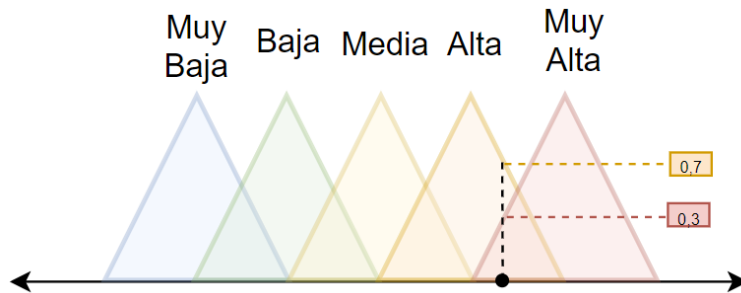


Figura 5.2: Representación del grado de pertenencia

Sin embargo, el uso de estas medidas presenta algunos problemas. Uno de ellos tiene que ver con la medida de confianza, ya que no tiene en cuenta el apoyo del consecuente, por lo que es incapaz de identificar dependencias negativas o independencia estadística entre  $A$  y  $B$ . Además, los conjuntos de items con un apoyo muy alto pueden dar lugar a reglas poco útiles, ya que cualquier conjunto de items en el antecedente de la regla puede parecer un buen predictor. En consecuencia, en la literatura se han propuesto varias métricas para seleccionar reglas en función de su interés potencial para el usuario. En este estudio, además del Soporte (Sop) y la Confianza (Conf) se van a utilizar las medidas de FC, YulesQ y Lift, cuyas definiciones se desarrollan a continuación:

- **Factor de certeza (FC) [SB75]:** Esta medida representa la probabilidad de que aparezca  $B$  en un ejemplo, considerando solamente aquellos en los que aparece  $A$ . En la Ecuación 5.1 se presenta la función utilizada para calcular su valor. El dominio de esta medida pertenece al intervalo  $[-1, 1]$  y sus rangos de valores tienen la siguiente interpretación:
  - valor  $< 0$ : Dependencia negativa
  - valor  $= 0$ : Independencia
  - valor  $> 0$ : Dependencia positiva

$$\begin{aligned}
 & \text{si } \mathbf{conf}(A \rightarrow B) > \mathbf{sop}(B) \text{ entonces} \\
 & FC(A \rightarrow B) = \frac{\mathbf{conf}(A \rightarrow B) - \mathbf{sop}(B)}{1 - \mathbf{sop}(B)} \\
 & \text{si } \mathbf{conf}(A \rightarrow B) < \mathbf{sop}(B) \text{ entonces} \quad (5.1) \\
 & FC(A \rightarrow B) = \frac{\mathbf{conf}(A \rightarrow B) - \mathbf{sop}(B)}{\mathbf{sop}(B)}
 \end{aligned}$$

otro caso  $FC(A \rightarrow B) = 0$

- **Lift [RMS98]:** Es el ratio entre la confianza real de una regla y la esperada. El dominio de esta medida pertenece al intervalo  $[0, \infty]$  y la interpretación de los intervalos es la siguiente:

- valor  $< 1$ : Dependencia negativa
- valor  $= 1$ : Independencia
- valor  $> 1$ : Dependencia positiva

$$Lift(A \rightarrow B) = \frac{\mathbf{sop}(AB)}{\mathbf{sop}(A) * \mathbf{sop}(B)} \quad (5.2)$$

- **YulesQ [TK00]:** Correlación entre antecedente y consecuente. En este caso los rangos de los intervalos de valores tienen la siguiente interpretación:

- valor  $< 0$ : Dependencia negativa
- valor  $= 0$ : Independencia
- valor  $> 0$ : Dependencia positiva

$$Yule'sQ(A \rightarrow B) = \frac{\mathbf{sop}(AB)\mathbf{sop}(\neg A \neg B) - \mathbf{sop}(A \neg B)\mathbf{sop}(\neg AB)}{\mathbf{sop}(AB)\mathbf{sop}(\neg A \neg B) + \mathbf{sop}(A \neg B)\mathbf{sop}(\neg AB)} \quad (5.3)$$

## 5.2. Meta Elementos Difusos para reglas de asociación difusas

En el capítulo anterior se muestra como se ha desarrollado una nueva gramática difusa diseñada para generar conjuntos difusos más genéricos. Dichos CFLTS tienen la capacidad de agrupar etiquetas difusas consecutivas generadas por particiones difusas fuertes o atributos definidos por un experto. Esta propuesta es análoga a la modificación del conector “O” desarrollada en [AMG06], con una extensión y modificación de la representación sintáctica con el fin de ser más cercana al lenguaje humano.

Esta gramática permite agregar la información extraída haciendo uso de menos reglas que las que necesita la gramática clásica. Además, con esta nueva gramática los procesos de extracción de reglas son capaces de detectar nuevas asociaciones entre atributos que no pueden ser identificadas con la gramática clásica.

En las siguientes secciones se presenta una nueva propuesta de items para la generación de RADs, basada en la nueva gramática de CFLTS, llamada Meta Elementos Difusos (MEDS). Estos MEDs pueden generar RADs que conectan términos más generales con otros más específicos, que no se generarían utilizando la gramática clásica. Haciendo uso de este nuevo tipo de items se va a presentar un algoritmo con el objetivo extraer RADs más transparentes. Para ello, se ha realizado una extensión del algoritmo “Fuzzy Apriori” [HKC01] añadiendo en el proceso dicha gramática, introduciendo un nuevo mecanismo de poda y limitando la profundidad del árbol para evitar obtener reglas muy específicas.

## 5.3. Extensión del método Fuzzy Apriori para extraer reglas de asociación difusas basadas en MEDs

En esta sección se va a proponer una extensión del método Fuzzy Apriori (llamado MFI-Apriori) de tal forma que considere los MEDs en la fase de extracción. El método Fuzzy Apriori inicialmente crea una lista L1 con los itemsets frecuentes de longitud 1 (1-itemsets), considerando frecuentes los que tienen un soporte superior a un umbral de mínimo soporte previamente establecido (minSop). A continuación, el método extraerá la lista L2 combi-

nando los elementos en la lista L1, generando así los posibles candidatos a 2-itemsets y añadiendo en L2 los que superan el mínimo soporte. Este proceso se repite hasta que no se puede generar una nueva lista  $L_k$ . Finalmente, combinando todos los itemsets en todas las listas, tanto en el antecedente como en el consecuente, se construyen las RADs. El método devolverá una BR con todas las que superen el umbral de mínima confianza (minConf) definido por el usuario.

La Figura 5.3 muestra el esquema del nuevo método propuesto para introducir en el proceso los MEDs. Cabe destacar que con el fin de obtener reglas simples se ha limitado la longitud del consecuente a un elemento.

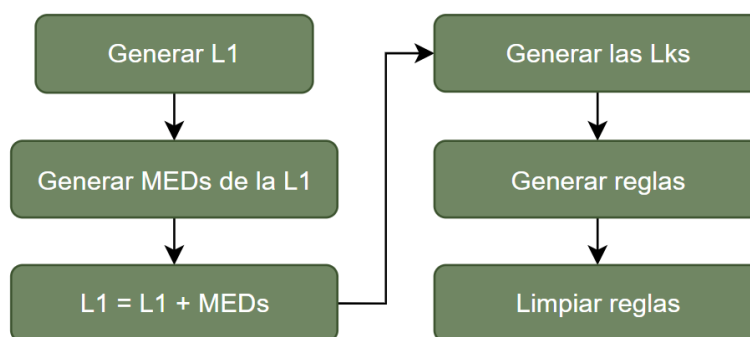


Figura 5.3: Diagrama de flujo de MFI-Apriori

### A. Generación de MEDs frecuentes

La generación de estos nuevos itemsets frecuentes se realizará en la primera fase, en la cual originalmente se construía la lista (L1) con los 1-itemsets frecuentes. En la lista L1 puede haber 1-itemsets asociados a etiquetas contiguas en el dominio de la variable, que podrían estar situados al principio, en el centro o al final de dicho dominio.

Cuando en L1 aparezcan 1-itemsets asociados a etiquetas consecutivas al principio del dominio se agruparán en un MED *hasta*, incluyendo en L1 un nuevo ítem asociado a él. En la Figura 5.4 se muestra la distribución de los 1-itemsets para un atributo, en este caso los frecuentes (en rojo, S0, S1 y S2) se encuentran a la izquierda del dominio de la variable. En este caso se creará un MED *Hasta S2* y se añade un nuevo 1-itemset asociado a él.

Cuando en L1 se incluyan 1-itemsets asociados a etiquetas consecutivas

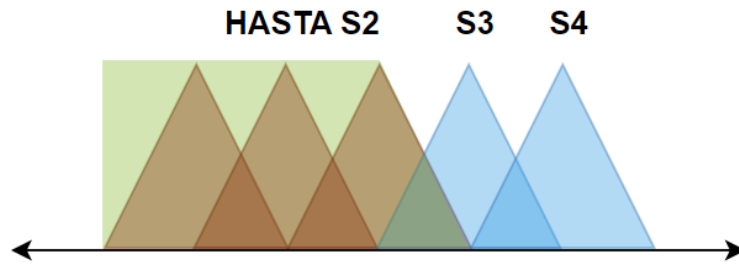


Figura 5.4: Distribución de itemets frecuentes asociados al MED *hasta*

en el centro del dominio se agruparán en un MED *desde-hasta*, incluyendo en L1 un nuevo item asociado a él. En la Figura 5.5 se muestra la distribución de los 1-itemsets para un atributo, en este caso los frecuentes (en rojo, S1, S2 y S3) se encuentran en el centro del dominio de la variable. En este caso se creará un MED *desde S1 hasta S3* y se añade un nuevo 1-itemset asociado a él.

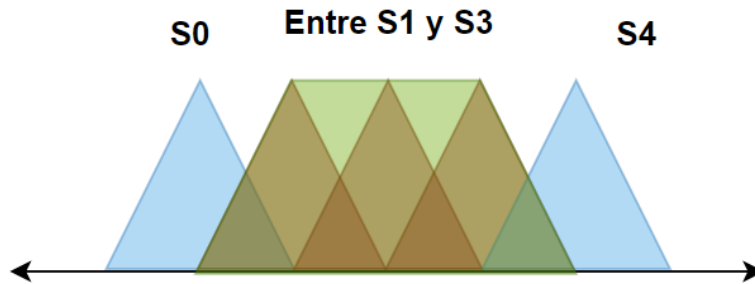


Figura 5.5: Distribución de itemets frecuentes asociados al MED *desde-hasta*

Finalmente, cuando en L1 aparezcan 1-itemsets asociados a etiquetas consecutivas en el final del dominio que se agruparán en un MED *desde*, incluyendo en L1 un nuevo item asociado a él. En la Figura 5.5 se puede ver la distribución de los 1-itemsets para un atributo, en este caso los frecuentes (en rojo, S2, S3 y S4) se encuentran al final del dominio de la variable. En este caso se creará un MED *desde S2* y se añade un nuevo 1-itemset asociado a él.

En algunos casos, todas las etiquetas del atributo tienen asociado un 1-itemset en L1. En este caso, el método busca si existen frecuencias parecidas entre 1-itemset consecutivos. En caso de encontrarlas y detectar que los 1-itemsets adyacentes a este grupo tienen una frecuencia significativamente

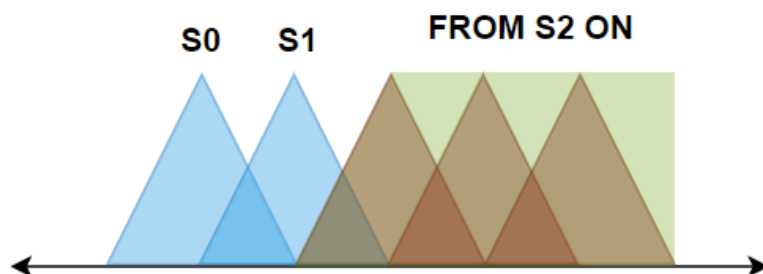


Figura 5.6: Distribución de itemets frecuentes asociados al MED desde

diferente, el método generará un MED y añadirá a L1 un 1-itemset asociado a él. En la Figura 5.7 se representa en color rojo la frecuencia de cada uno de los 1-itemsets asociados a las etiquetas del atributo. Para este caso, se ha podido generar un MED desde  $S1$  hasta  $S3$ , Figura 5.8, y un 1-itemset asociado a él se añadirá a L1.

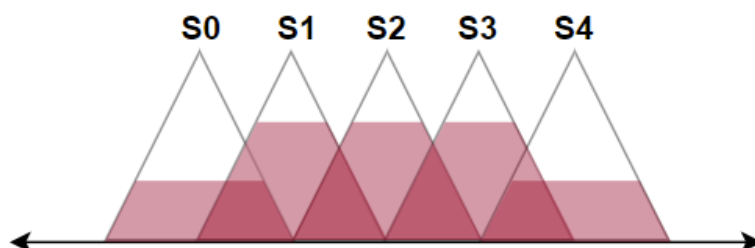


Figura 5.7: Distribución de frecuencias de todas las etiquetas

## B. Poda de reglas

1. **Eliminación de reglas redundantes:** Con el fin de evitar tener información repetida en la BR se van a eliminar todas las reglas redundantes. Una regla R2 se considera redundante de R1 si todos los itemsets de R1 están contenidos en R2 y la confianza de R2 (o cualquier otra medida de interés, en nuestro caso es FC) es menor o igual que la confianza de R1. Cuando las reglas poseen MEDs pueden surgir varios casos especiales a la hora de buscar la redundancia, por lo que durante este proceso se han considerado dos niveles:

- **Primer nivel:** En esta etapa se consideran las reglas que tengan

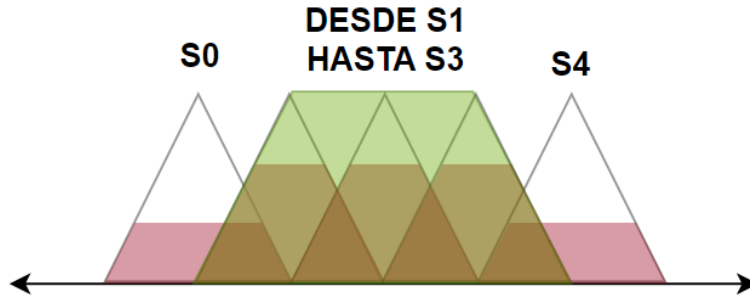


Figura 5.8: Distribución de frecuencias asociadas al nuevo MED desde hasta

exactamente el mismo antecedente y en el consecuente una de ellas posee un MEDs que contiene la etiqueta asociada al consecuente de la otra, en la Figura 5.9 se muestra un ejemplo simple de este caso. Si en la BR aparece una regla de este tipo asociada a cada una de las etiquetas del MED, se eliminan todas las reglas simples y se mantiene la que contiene el MED. En el caso contrario se elimina la del MED y permanecerán las más específicas.

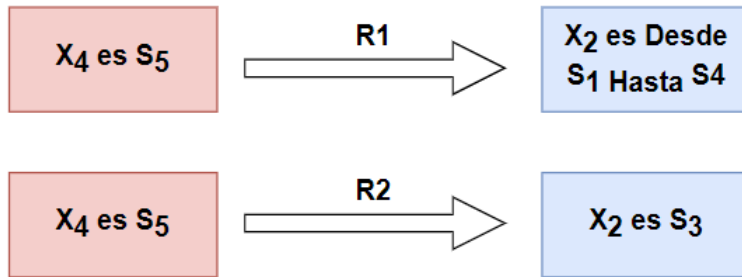


Figura 5.9: Ejemplo simple de regla redundante del primer paso

- **Segundo nivel:** Finalmente se consideran las reglas que están contenidas una dentro de la otra, es decir, son sub reglas. Una regla R1 es sub regla de R2 si todos sus items están incluidos en R2, ya sea porque coinciden exactamente o porque R2 posee MEDs que los contienen. En este caso, se elimina la regla con menor valor para la medida de interés. En la Figura 5.10 se muestra ver un ejemplo simple de este caso de redundancia.

2. **Eliminación de reglas no interesantes:** Para evitar obtener re-



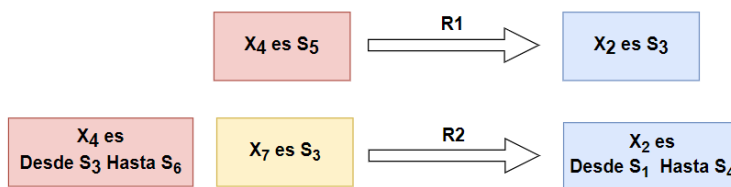


Figura 5.10: Ejemplo simple de regla redundante del segundo nivel

glas que no sean interesantes solo se extraerán RADs muy fuertes [BBSV02]. Se dice que una RAD es muy fuerte cuando satisface los 3 criterios siguientes:

- $FC(A \rightarrow B) > 0$
- $Soporte(A \rightarrow B) > minSup$
- $Soporte(\neg B \rightarrow \neg A) > minSupp$

## 5.4. Estudio experimental

Para el estudio experimental se ha seleccionado una batería de 10 conjuntos de datos de diferentes características con el fin de comprobar el comportamiento del método en diferentes escenarios. En la Tabla 5.1 se incluyen las principales características de los conjuntos de datos seleccionados donde #Var es el número de variables, con valores entre 4 y 16, y #Ejs es el número de ejemplos, entre 49 y 68040.

La finalidad de esta experimentación es comparar el método propuesto con el conocido algoritmo Fuzzy Apriori. En la Tabla 5.2 se muestran los valores que toman cada uno de los parámetros para ambos algoritmos, establecidos en base a las recomendaciones de los autores en la publicación original (que son los establecidos en KEEL [TGM<sup>+</sup>17]). Las particiones lingüísticas iniciales están formadas por siete términos lingüísticos con FPs triangulares uniformemente distribuida. Cabe destacar que se ha limitado el número de elementos en el antecedente de las reglas con el fin de obtener reglas más cortas que sean más fáciles de manejar y comprender por un usuario. La Subsección 5.4.1 explica cómo se ha realizado la elección de la profundidad que más se ajusta al estudio.

Tabla 5.1: Características de los conjuntos de datos

| Nombre       | #Var | #Ejs  |
|--------------|------|-------|
| Basketball   | 5    | 96    |
| Bolts        | 8    | 40    |
| ColorTexture | 17   | 68040 |
| House16H     | 17   | 22784 |
| Letter       | 16   | 20000 |
| Magic        | 10   | 19020 |
| Pollution    | 16   | 60    |
| Quake        | 4    | 2178  |
| Stulong      | 5    | 1419  |
| Wine         | 13   | 178   |

Tabla 5.2: Valores de los parámetros

| Algoritmo     | Parámetros  |
|---------------|---|
| Fuzzy Apriori | minConf = 0,8, minSop = 0,1, minFC = 0,8, Profundidad = 4 |
| MFI-Apriori   | minConf = 0,8, minSop = 0,1, minFC = 0,8, Profundidad = 4 |

#### 5.4.1. Análisis de la influencia del límite de profundidad en el proceso de extracción.

El método Fuzzy Apriori ha sido limitado en su capacidad de exploración en profundidad para generar conjuntos de reglas simples que sean fáciles de interpretar por un experto. Esto se logra estableciendo un límite en el número de elementos en el antecedente de las reglas generadas, con una poda del árbol en el paso  $x+1$ . En la Tabla 5.3 se presentan los resultados obtenidos para 5 conjuntos de datos, con profundidades de exploración 3, 4 y 5, junto con los tiempos de ejecución para cada una de las pruebas realizadas. Los indicadores incluidos en la tabla son el número de reglas generadas (NR), la media del FC ( $fc\_med$ ), y la media para la medida Yule'sQ ( $yulesQ\_med$ ). Después de analizar los resultados de las Tablas 5.3 y 5.4 se ha considerado utilizar una profundidad de 4, lo que significa que las reglas generadas

tendrán un máximo de 3 elementos en el antecedente.

Se ha elegido finalmente este valor porque al observar los tiempos de ejecución y los valores de las medidas de calidad en profundidades 4 y 5 se puede ver que el tiempo de ejecución aumenta exponencialmente, mientras que las medidas de calidad apenas mejoran. Además, el conjunto de reglas aumenta considerablemente a medida que se aumenta la profundidad del método. Por lo tanto, con una profundidad de 4 se obtiene un buen equilibrio entre el número de reglas generadas, los valores de las medidas de interés y el costo computacional.

| Conjunto de datos | Profundidad | NR          | fc_med      | yulesQ_med  | Tiempo (seg)  |
|-------------------|-------------|-------------|-------------|-------------|---------------|
| Bolts             | 3           | 46          | 0,96        | 0,98        | 0,00          |
|                   | <b>4</b>    | <b>46</b>   | <b>0,96</b> | <b>0,98</b> | <b>0,042</b>  |
|                   | 5           | 46          | 0,96        | 0,98        | 0,09          |
| Colormoments      | 3           | 21          | 0,89        | 0,95        | 12,748        |
|                   | <b>4</b>    | <b>26</b>   | <b>0,9</b>  | <b>0,95</b> | <b>16,744</b> |
|                   | 5           | 27          | 0,9         | 0,95        | 1306,08       |
| House16H          | 3           | 728         | 0,87        | 0,87        | 179,73        |
|                   | <b>4</b>    | <b>4600</b> | <b>0,88</b> | <b>0,88</b> | <b>2096,3</b> |
|                   | 5           | 11347       | 0,89        | 0,89        | 16939,95      |
| Pollution         | 3           | 400         | 0,92        | 0,91        | 1,53          |
|                   | <b>4</b>    | <b>754</b>  | <b>0,92</b> | <b>0,93</b> | <b>40,36</b>  |
|                   | 5           | 797         | 0,92        | 0,93        | 733,07        |
| Winequality-red   | 3           | 82          | 0,85        | 0,87        | 4,144         |
|                   | <b>4</b>    | <b>634</b>  | <b>0,87</b> | <b>0,87</b> | <b>32,145</b> |
|                   | 5           | 1217        | 0,88        | 0,88        | 259,16        |

Tabla 5.3: Estudio de profundidad Fuzzy Apriori original

#### 5.4.2. Resultados de la comparativa Fuzzy Apriori clásico con la nueva propuesta

En la Tabla 5.5 se presentan los resultados obtenidos para el algoritmo Fuzzy Apriori original y en Tabla 5.6 los resultados del método con la nueva representación. Ambas tablas poseen las siguientes columnas: *NR*, número de reglas obtenido, *Longitud*, la longitud media del antecedente, *Sop*, el soporte medio de las reglas, *Conf*, confianza media de las reglas, *FC*, Factor

| Conjunto de datos | Profundidad | n.rules     | fc.med      | yulesQ.med  | Tiempo (seg)  |
|-------------------|-------------|-------------|-------------|-------------|---------------|
| Bolts             | 3           | 81          | 0,96        | 0,98        | 0,00          |
|                   | <b>4</b>    | <b>98</b>   | <b>0,97</b> | <b>0,98</b> | <b>0,01</b>   |
|                   | 5           | 99          | 0,97        | 0,98        | 0,09          |
| Colormoments      | 3           | 71          | 0,87        | 0,86        | 137,56        |
|                   | <b>4</b>    | <b>387</b>  | <b>0,85</b> | <b>0,81</b> | <b>684,83</b> |
|                   | 5           | 974         | 0,85        | 0,81        | 1306,08       |
| House16H          | 3           | 728         | 0,87        | 0,87        | 179,73        |
|                   | <b>4</b>    | <b>4904</b> | <b>0,88</b> | <b>0,87</b> | <b>2096,3</b> |
|                   | 5           | 16237       | 0,88        | 0,93        | 16939,95      |
| Pollution         | 3           | 1188        | 0,93        | 0,92        | 1,53          |
|                   | <b>4</b>    | <b>3787</b> | <b>0,94</b> | <b>0,93</b> | <b>40,36</b>  |
|                   | 5           | 6026        | 0,94        | 0,93        | 733,07        |
| Winequality-red   | 3           | 195         | 0,87        | 0,86        | 4,144         |
|                   | <b>4</b>    | <b>1214</b> | <b>0,88</b> | <b>0,87</b> | <b>32,145</b> |
|                   | 5           | 2840        | 0,88        | 0,87        | 259,16        |

Tabla 5.4: Estudio de profundidad Fuzzy Apriori con la nueva representación

de Certeza medio de las reglas, *Lift* es el Lift medio de las reglas, *YulesQ* es el valor medio de la medida YulesQ para las reglas y *Cub* es el porcentaje de ejemplos cubiertos.

Haciendo un primer análisis de los resultados de las Tablas 5.5 y 5.6 se puede observar que la nueva propuesta obtiene reglas con mejores valores medios que Fuzzy Apriori original para todas las medidas de calidad consideradas (Sop, Conf, FC, Lift y YulesQ). Además, el conjunto de reglas posee reglas más cortas con un cubrimiento mejor en los conjuntos de datos (casi un 100 % de media), lo que permite obtener información de toda la BDs. Observando las reglas obtenidas se puede decir que todas las reglas generadas por Fuzzy Apriori las genera también la nueva propuesta, pero más del 30 % son eliminadas a causa de la redundancia con otras reglas que incluyen elementos MEDs. Otro comportamiento interesante es que la propuesta consigue extraer un número mayor de reglas que Fuzzy Apriori, lo que indica que los nuevos MEDs permiten representar asociaciones que no son posibles de generar con la gramática clásica. Cabe destacar que Fuzzy Apriori no es capaz de extraer ninguna regla para el conjunto de datos Basketball, mientras que la nueva propuesta obtiene 14 reglas con buenos valores en las medidas de calidad.

Tabla 5.5: Resultados obtenidos con el método Fuzzy Apriori para las medidas de interés

| Conjunto de datos | NR             | Longitud    | Sop         | Conf        | FC          | Lift        | YulesQ      | Cub          |
|-------------------|----------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|
| Basketball        | 0              | -           | -           | -           | -           | -           | -           | -            |
| Bolts             | 46             | 1,57        | 0,15        | 0,97        | 0,96        | 4,24        | 0,98        | 90           |
| ColorTexture      | 15406          | 2,75        | 0,18        | 0,95        | 0,9         | 2,04        | 0,94        | 100          |
| House16H          | 4600           | 2,8         | 0,2         | 0,97        | 0,88        | 1,39        | 0,87        | 100          |
| Letter            | 110            | 2,65        | 0,12        | 0,91        | 0,85        | 2,12        | 0,83        | 98,77        |
| Magic             | 1668           | 2,77        | 0,15        | 0,95        | 0,89        | 1,76        | 0,9         | 99,99        |
| Pollution         | 754            | 2,46        | 0,14        | 0,97        | 0,92        | 1,54        | 0,93        | 96,67        |
| Quake             | 7              | 2,14        | 0,13        | 0,97        | 0,89        | 1,46        | 0,87        | 99,08        |
| Stulong           | 31             | 2,61        | 0,17        | 0,96        | 0,88        | 1,39        | 0,86        | 98,59        |
| Wine              | 24             | 1,92        | 0,12        | 0,93        | 0,89        | 2,68        | 0,89        | 98,31        |
| <b>Media</b>      | <b>2516,22</b> | <b>2,41</b> | <b>0,15</b> | <b>0,95</b> | <b>0,90</b> | <b>2,07</b> | <b>0,90</b> | <b>97,93</b> |

Tabla 5.6: Resultados obtenidos con la extensión de Fuzzy Apriori para las medidas de interés

| Conjuntos de datos | NR             | Longitud    | Sop         | Conf        | FC          | Lift        | YulesQ      | Cub          |
|--------------------|----------------|-------------|-------------|-------------|-------------|-------------|-------------|--------------|
| Basketball         | 14             | 2,36        | 0,13        | 0,99        | 0,91        | 1,13        | 0,88        | 92,71        |
| Bolts              | 98             | 1,84        | 0,15        | 0,98        | 0,97        | 2,8         | 0,98        | 100          |
| ColorTexture       | 15152          | 2,94        | 0,17        | 0,95        | 0,9         | 2,07        | 0,94        | 100          |
| House16H           | 4904           | 2,99        | 0,19        | 0,97        | 0,88        | 1,39        | 0,88        | 100          |
| Letter             | 2151           | 2,97        | 0,14        | 0,98        | 0,87        | 1,18        | 0,9         | 100          |
| Magic              | 2152           | 2,97        | 0,15        | 0,96        | 0,89        | 1,57        | 0,91        | 99,98        |
| Pollution          | 3787           | 2,85        | 0,15        | 0,98        | 0,94        | 1,33        | 0,93        | 100          |
| Quake              | 4              | 2,5         | 0,12        | 0,97        | 0,88        | 1,43        | 0,89        | 99,04        |
| Stulong            | 28             | 2,96        | 0,14        | 0,96        | 0,87        | 1,39        | 0,87        | 98,17        |
| Wine               | 643            | 2,74        | 0,13        | 0,98        | 0,91        | 1,27        | 0,95        | 100          |
| <b>Media</b>       | <b>3213,22</b> | <b>2,75</b> | <b>0,15</b> | <b>0,97</b> | <b>0,90</b> | <b>1,60</b> | <b>0,92</b> | <b>99,70</b> |

Tabla 5.7: Resultados obtenidos con el test de Wilcoxon para las medidas Conf, Yule'sQ y FC.

| Medida  | $R^+$ | $R^-$ | P-valor    | P-valor asintótico |
|---------|-------|-------|------------|--------------------|
| Conf    | 40,0  | 5,0   | 0,039      | 0,023              |
| Yule'sQ | 34,5  | 1,5   | 0,019      | 0,010              |
| FC      | 26,5  | 9,5   | $\geq 0,2$ | 0,167              |

Para analizar las diferencias significativas entre los valores obtenidos por el nuevo método y por Fuzzy Apriori original se ha utilizado el test no paramétrico de Wilcoxon para las medidas de calidad Conf, YulesQ y FC. La métrica Lift no se incluye en los análisis estadísticos porque, por definición, esta medida no posee un límite superior (ver definición en la Ecuación 5.2). En la Tabla 5.7 se pueden ver los resultados del test para las 3 medidas. Como se puede observar, la hipótesis de igualdad se rechaza con un  $\alpha$  mejor a 0,05 para las medidas de Conf y YulesQ. En el caso del FC la hipótesis de igualdad no puede ser rechazada aunque la nueva propuesta obtiene unos valores mejores que Fuzzy Apriori.

### 5.4.3. Estudio de la capacidad de resumen de información de la nueva estructura de regla

Con el objetivo de comprender de manera más detallada el tipo de reglas generadas mediante la nueva representación, se ha llevado a cabo un análisis exhaustivo de las reglas obtenidas a partir del conjunto de datos Bolts. Este análisis implica la consideración del significado de cada una de las variables involucradas en las reglas, permitiendo su interpretación y la exposición del tipo de información que se está extrayendo. De esta manera, es posible contrastar la información obtenida con la estructura de reglas original.

El conjunto de datos Bolts contiene información detallada sobre una máquina utilizada en una fábrica para contar tornillos. La máquina en cuestión posee una rueda en la cual los tornillos caen y giran antes de caer en una cinta transportadora. Luego, los tornillos son pasados por un ojo electrónico que los cuenta antes de ser empaquetados. Las variables observadas en este conjunto de datos incluyen:

- **SPEED:** Parámetro que controla la velocidad del giro de la rueda

- **TOTAL:** Número de tornillos que tiene que contar el ojo
- **SPEED2:** Parámetro que controla el cambio de la velocidad (en caso de que se quiera que se expulsen menos tornillos, irá más lento)
- **NUMBER2:** Número de tornillos que tiene que contar con la segunda velocidad.
- **SENS:** Sensibilidad del ojo electrónico
- **TIME:** Tiempo real que tarda la máquina en contar un número predeterminado de tornillos.
- **T20BOLT:** Tiempo que tarda la máquina en contar 20 tornillos.

### **Reglas de Fuzzy Apriori original reemplazadas por el método propuesto**

Cuando se eliminan reglas redundantes, se busca mantener las reglas más generales y con mejores valores en las medidas de calidad. A continuación se muestra un ejemplo de una regla que estaba en el conjunto generado por el algoritmo original y que ha sido reemplazada por una de las nuevas reglas que se han podido generar al añadir la nueva representación. Si se observan las reglas, es evidente por qué la regla específica (la de la representación original) fue sustituida por la regla general. Cabe señalar que ambas reglas tienen un FC de 1.0, por lo que están igualadas en cuanto a la medida de calidad y, por lo tanto, se elimina la regla específica.

SI **TOTAL** es ETIQUETA 6 Y **TIME** es ETIQUETA 1  
ENTONCES  
**SPEED1** es ETIQUETA 6

SI **TOTAL** es ETIQUETA 6 Y **TIME** es hasta ETIQUETA 1  
ENTONCES  
**SPEED1** es ETIQUETA 6

### Nuevo conocimiento extraído

A continuación se incluye un ejemplo de una regla generada por el algoritmo con la nueva representación que demuestra que se puede extraer conocimiento nuevo. En el conjunto de reglas obtenidas por Fuzzy Apriori original no existe ninguna regla que tenga como antecedente el mismo que la del ejemplo, es decir, que se está proporcionando información sobre ejemplos que antes no se cubrían.

SI **SPEED1** es ETIQUETA 1  
ENTONCES  
**TIME** es hasta ETIQUETA 1

Además de obtener conocimiento nuevo, se puede afirmar la regla obtenida es de calidad acorde a las medidas establecidas:

- **Soporte de la regla** = 0,12
- **Factor de certeza** = 1,0
- **Yule'sQ** = 1,0





# Comentarios Finales

En esta sección, se presenta un resumen de los resultados y conclusiones obtenidos a partir de esta memoria. También se recogen las publicaciones asociadas a esta tesis y algunos aspectos relacionados con posibles trabajos futuros que sigan la misma línea de investigación presentada en esta tesis, así como otras líneas de investigación que pueden derivarse.

## A Resumen y Conclusiones

Debido a la gran necesidad que existe en la actualidad de generar sistemas de IA que generen confianza, en esta tesis se han propuesto nuevas extensiones y representaciones de reglas difusas para diseñar nuevos métodos evolutivos que permitan obtener SBRDs precisos, transparentes y comprensibles por los humanos, aplicados a distintos tipos de problemas (clasificación, regresión y descubrimiento de conocimiento). Además, los métodos propuestos han sido utilizados para realizar un análisis integrado de distintos tipos de datos ómicos (genéticos, epigenéticos y metabólicos), clínicos y ambientales para obtener modelos descriptivos y predictivos comprensibles que permitan predecir el riesgo de sufrir insulinoresistencia (IR) y/u obesidad infantil, facilitando el diseño de programas preventivos personalizados. A continuación se resumen las conclusiones obtenidas de cada una de las propuestas realizadas.

### **A.1 Clasificación Asociativa basada en reglas difusas positivas y negativas**

El objetivo de esta propuesta es presentar un algoritmo evolutivo para obtener clasificadores asociativos con un buen equilibrio entre complejidad y precisión, considerando una extensión de la representación clásica de reglas de asociación difusas para poder representar dependencias positivas y negativas. Con el fin de superar las limitaciones de los métodos existentes en la literatura, se ha propuesto un nuevo enfoque en tres etapas que incluye un aprendizaje evolutivo, una extracción y selección de reglas y un ajuste de las FPs.

En la experimentación se ha demostrado la eficacia del método mediante una comparativa con otros métodos sobre 23 conjuntos de datos y un análisis de resultados sobre un problema biomédico real. Los resultados muestran que el método propuesto supera a los otros métodos analizados en varias medidas de calidad y complejidad, ya que el conjunto de reglas obtenido es más reducido y las mismas poseen un menor número de variables, lo que facilita la comprensión desde el punto de vista de un experto. Además, se ha podido ver que el método propuesto se comporta bien en problemas relacionados con la biomedicina y se podría utilizar para predecir y clasificar, así como comprender las alteraciones moleculares que conducen a ciertas enfermedades.

### **A.2 Diseño de una nueva extensión de la gramática de lingüística difusa aplicada a problemas de regresión**

El objetivo de esta parte de la tesis es desarrollar modelos que puedan explicar los problemas de regresión basándonos en dos motivaciones principales. En primer lugar, se quiere enfocar la construcción de los modelos de tal forma que los datos utilizados sean coherentes con la información existente. En segundo lugar, pretenden descubrir información nueva y útil analizando las relaciones entre las variables de los datos. En resumen, el objetivo es crear modelos que no solo sean precisos, sino que también proporcionen información sobre los datos subyacentes.

Con este fin se ha propuesto una extensión de la gramática basada en la composición de términos simples y una nueva estructura de reglas lingüísticas difusas para expresar con precisión el comportamiento y la variabilidad de los datos. Como resultado, se pueden aprender modelos lingüísticos que

necesiten menos reglas para ser precisos. La nueva extensión lineal interpretable en el consecuente de las reglas para regresión consigue mantener una precisión competitiva a la vez que explica la variabilidad específica utilizando dos parámetros que aún pueden interpretarse lingüísticamente. Haciendo uso de estas propuestas se ha creado un método en dos etapas (aprendizaje de la BD y BR y un ajuste con selección de reglas) para optimizar las medidas de precisión e interpretabilidad. Este método hace uso de un nuevo aprendizaje lingüístico de la BR basado en árboles, que se adapta bien al nuevo tipo de reglas.

El método propuesto se ha comparado con otros métodos de la literatura sobre 23 conjuntos de datos de regresión de distinta complejidad, y los resultados se han analizado estadísticamente mediante los test de Holm, Friedman y Wilcoxon. Los resultados han demostrado que el método propuesto es eficaz en términos de precisión e interpretabilidad, y que superó a algunos de los métodos más avanzados para obtener SBRDs interpretables. Además, ha superado en precisión a un método orientado a obtener resultados precisos sin tener en cuenta la complejidad. En general, el método propuesto genera modelos sencillos con solo 4 ó 5 reglas, al tiempo que mejora la precisión en comparación con modelos lingüísticos más complejos. Además, el método ha sido capaz de descubrir información útil que no se ha representado antes, como relaciones lineales sencillas con interpretaciones lingüísticas.

Además, el modelo lingüístico obtenido de la aplicación del algoritmo a un problema real relacionado con obesidad en niños se ha analizado desde el punto de vista biológico por varios expertos en la materia. Los expertos han afirmado que han podido comprender sin problemas las relaciones que el modelo ha establecido entre las variables de entrada y de salida, identificando relaciones interesantes que no se conocían sobre la cohorte de niños y que actualmente se está trabajando en ellas en laboratorio.

### **A.3 Adaptación de la nueva extensión de la gramática de lingüística difusa a la extracción de reglas difusas**

El propósito principal de este estudio consiste en proponer una nueva extensión de la representación de reglas de asociación difusas haciendo uso de la nueva gramática difusa lingüística propuesta en el Capítulo 4, con el fin de mejorar la transparencia de los conjuntos de reglas obtenidos y ampliar el tipo de asociaciones que se pueden representar con la gramática

clásica. Para ello se han definido ítems difusos más genéricos, denominados Meta-Elementos Difusos (MEDs), que permiten agrupar elementos difusos consecutivos generados a partir de particiones difusas fuertes definidas por un experto para un atributo, de forma similar al conector “OR”, pero extendiendo o modificando su representación sintáctica para hacerlo más cercano al razonamiento humano. Además, se ha extendido un método clásico de extracción de RADs para generar reglas haciendo uso de esta nueva representación.

Los resultados obtenidos a partir de la experimentación realizada en 15 bases de datos del mundo real muestran que la nueva estructura de reglas permite obtener conjuntos de reglas con valores iguales o mejores en cuanto a las medidas de calidad. Además, se logra extraer conocimiento interesante que no se obtiene mediante la estructura de regla original, presentando un mayor número de reglas que permiten obtener un mejor cubrimiento de las bases de datos. Las reglas obtenidas permiten resumir parte de la información obtenida con la estructura de regla original, lo que reduce el número de reglas necesarias para representar la misma información.

## B Publicaciones Asociadas a la Tesis

A continuación se presenta la lista completa de publicaciones en diferentes revistas y congresos que se han realizado como consecuencia directa del trabajo desarrollado en esta tesis.

■ Publicaciones en revistas internacionales:

1. C. Biedma-Rdiguez, M.J. Gacto, A. Anguita-Ruiz, J. Alcalá-Fdez, R. Alcalá. Transparent but Accurate Evolutionary Regression Combining New Linguistic Fuzzy Grammar and a Novel Interpretable Linear Extension. *International Journal of Fuzzy Systems* 24 (2022) 3082-3103, doi: 10.1007/s40815-022-01324-w (Q2 pero era Q1 cuando fue aceptado - Fuente WoS).
2. C. Biedma-Rdiguez, M.J. Gacto, A. Anguita-Ruiz, R. Alcalá, C.M. Aguilera, J. Alcalá-Fdez. Learning positive-negative rule-based fuzzy associative classifiers with a good trade-off between complexity and accuracy. *Fuzzy Sets and Systems* (2023) 108511, doi: 10.1016/j.fss.2023.03.014 (D1 - Fuente WoS).

- Publicaciones en congresos internacionales:
  1. C. Biedma-Rdguez, M.J. Gacto; R. Alcalá; J. Alcalá-Fdez. Meta-Fuzzy Items for Fuzzy Association Rules. IEEE International Conference on Fuzzy Systems (FUZZ-IEEE 2021), Luxembourg (Luxembourg), 1-6, 11-14 July, 2021.  
doi: 10.1109/FUZZ45933.2021.9494571.

## C Líneas de investigación futuras

A continuación, se consideran algunas líneas de trabajo futuras a partir de las conclusiones obtenidas en esta memoria.

### C.1 Extensión de modelo de regla haciendo uso de ambas propuestas

En esta tesis se ha propuesto una extensión de la representación clásica de reglas de asociación difusas para poder representar dependencias positivas y negativas, lo que ha permitido mantener o mejorar la precisión de los modelos con respecto a otros métodos de la literatura, reduciendo la complejidad del modelo obteniendo conjunto de reglas reducido y con un bajo número de variables. Por otro lado, se ha propuesto nueva representación de los SBRDs, específicamente diseñada para problemas de regresión, combinando una nueva gramática difusa lingüística y una nueva extensión lineal interpretable que proporcione el mayor nivel de transparencia posible, obteniendo modelos eficaces en términos de precisión e interpretabilidad.

Una posible línea de trabajo futuro consiste en fusionar la nueva gramática difusa lingüística propuesta y la extensión para poder representar dependencias positivas y negativas, aprovechando las ventajas de ambas propuestas para conseguir una mejor representación y modelado de los datos.

Definitivamente, las mejoras relacionadas con la representación de las reglas son realmente prometedoras de cara a mejorar las propuestas, pudiendo aumentar aún más la capacidad de representar el conocimiento y, por tanto, la capacidad de obtener un conocimiento interesante y útil para el usuario.

## **C.2 Diseño de métodos de post-procesamiento para fusionar reglas complementarias**

La FP de un conjunto difuso negado es igual a la FP del conjunto difuso original negada, por lo que normalmente el núcleo de la función cubre gran parte del dominio de la variable. Sin embargo, en ocasiones es interesante que el núcleo de la FP sea más reducido, aumentando la zona de la función en la que el grado de pertenencia es cercano a 0. En la propuesta actual no se puede extender la zona negada para incluir el dominio de conjuntos difusos consecutivos, por lo que el método genera varias reglas para representarlo. Sin embargo, estas reglas presentan valores bajos para las medidas de calidad.

Para evitar este problema, una línea de trabajo interesante consiste en considerar un método de post-procesamiento que permita generar reglas en la que se puedan negar varios conjuntos difusos consecutivos en la misma regla. Esto permitirá que los ejemplos cubiertos por la regla sean coherentes con la información existente en la base de datos y ayudará a descubrir información nueva y útil para los expertos.

## **C.3 Diseño de métodos evolutivos para la extracción de RADs basados en las extensiones y representaciones de regla propuestas en esta memoria.**

En esta memoria se presenta un primer enfoque de extensión de un método clásico de extracción de RADs para considerar la nueva gramática difusa lingüística. Sin embargo, en la literatura existen distintas propuestas basadas en AEs, AEMOs, algoritmos evolutivos basados en nichos, etc, para extraer RADs con una alta calidad y diversidad.

Una línea de trabajo futuro que se considera muy prometedora es desarrollar nuevas propuestas que permitan realizar un aprendizaje evolutivo de las FPs y de las condiciones de las reglas para obtener conjuntos de RADs reducidos con un buen equilibrio entre comprensibilidad, interés y precisión.

Además, durante esta investigación se han aplicado las propuestas a un problema biológico sobre la obesidad infantil, demostrando que este tipo de enfoques son realmente interesantes para generar información útil que permite a los científicos comprender mejor los procesos biológicos asociados a los tratamientos y generar nuevas hipótesis para abordar el problema.

Debido a los buenos resultados obtenidos, una de las líneas de trabajo futuro que se quiere abordar es la aplicación de estas técnicas de extracción de RADs a este problema biológico. De esa manera, se podrán identificar las redes moleculares inter-ómicas o fenómenos de epistasis que subyacen la obesidad e IR en población pediátrica, así como generar nuevos subgrupos de pacientes obesos de acuerdo a su contexto molecular.





# Bibliografía

- [AAFH07] Alcalá R., Alcalá-Fdez J. y Herrera F. (2007) A proposal for the genetic lateral tuning of linguistic fuzzy systems and its interaction with rule selection. *IEEE Transactions on Fuzzy Systems* 15: 616–635.
- [AB18] Adadi A. y Berrada M. (2018) Peeking inside the black-box: a survey on explainable artificial intelligence (xai). *IEEE access* 6: 52138–52160.
- [ADRS<sup>+</sup>20] Arrieta A. B., Díaz-Rodríguez N., Ser J. D., Bennetot A., Tabik S., Barbado A., Garcia S., Gil-Lopez S., Molina D., Benjamins R., Chatila R. y Herrera F. (2020) Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information Fusion* 58(82-115): 34.
- [AFAH11] Alcalá-Fdez J., Alcalá R. y Herrera F. (2011) A fuzzy association rule-based classification model for high-dimensional problems with genetic rule selection and lateral tuning. *IEEE Transactions on Fuzzy Systems* 19: 857–872.
- [AFFJ<sup>+</sup>11] Alcalá-Fdez J., Fernández A., J.Luengo, J.Derrac, García S., Sanchez L. y Herrera F. (2011) Keel data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic and Soft Computing* 17: 255–287.
- [AGH11] Alcalá R., Gacto M. J. y Herrera F. (2011) A fast and scalable multi-objective genetic fuzzy system for linguistic fuzzy modeling in high-dimensional regression problems. *IEEE Transactions on Fuzzy Systems* 19(4): 666–681.

- [AMCM<sup>+</sup>21] Alonso Moral J. M., Castiello C., Magdalena L., Mencar C., Alonso Moral J. M., Castiello C., Magdalena L. y Mencar C. (2021) Toward explainable artificial intelligence through fuzzy systems. *Explainable Fuzzy Systems: Paving the Way from Interpretable Fuzzy Systems to Explainable AI Systems* páginas 1–23.
- [AMG06] Alonso J., Magdalena L. y Guillaume S. (2006) Linguistic knowledge base simplification regarding accuracy and interpretability. *Mathware and Soft Computing* 13: 203–216.
- [AMGR09] Alonso J. M., Magdalena L. y González-Rodríguez G. (2009) Looking for a good fuzzy system interpretability index: An experimental approach. *Int. J. Approx. Reason.* 51(115-134): 19.
- [ARMGR<sup>+</sup>20] Anguita-Ruiz A., Mendez-Gutierrez A., Ruperez A. I., Leis R., Bueno G., Gil-Campos M., Tofe I., Gomez-Llorente C., Moreno L. A., Gil A. y Aguilera C. M. (2020) The protein s100a4 as a novel marker of insulin resistance in prepubertal and pubertal children with obesity. *Metabolism* 105.
- [BBSV02] Berzal F., Blanco I., Sanchez D. y Vila M. (2002) Measuring the accuracy and interest of association rules: A new framework. *Intelligent Data Analysis* 6(3): 221–235.
- [BD07] Ben-David A. (2007) A lot of randomness is hiding in accuracy. *Engineering Applications of Artificial Intelligence* 20(7): 875–885.
- [BK06] B. Kavsek N. L. (2006) Apriori-sd: Adapting association rule learning to subgroup discovery. *Applied Artificial Intelligence* 20(7): 543–583.
- [BRGAR<sup>+</sup>22] Biedma-Rdiguez C., Gacto M. J., Anguita-Ruiz A., Alcalá-Fdez J. y Alcalá R. (2022) Transparent but accurate evolutionary regression combining new linguistic fuzzy grammar and a novel interpretable linear extension. *International Journal of Fuzzy Systems* 24(7): 3082–3103.
- [BT11] Bornfeldt K. y Tabas I. (2011) Insulin resistance, hyperglycemia, and atherosclerosis. *Cell Metabolism* 14(5): 575–585.

- [Cas16] Castelvechi D. (2016) Can we open the black box of ai? *Nature News* 538(7623): 20.
- [CCLvV07] Coello-Coello C. A., Lamont G. y van Veldhuizen D. (2007) *Evolutionary Algorithms for Solving Multi-Objective Problems, 2nd ed.* Springer.
- [CdJH99] Cordon O., del Jesus M. J. y Herrera F. (1999) A proposal on reasoning methods in fuzzy rule-based classification systems. *International Journal of Approximate Reasoning* 20(1): 21–45.
- [CDM19] Chooi Y. C., Ding C. y Magkos F. (2019) The epidemiology of obesity. *Metabolism* 92: 6–10.
- [CHG<sup>+</sup>01] Cordon O., Herrera F., Gomide F., Hoffmann F. y Magdalena L. (2001) Ten years of genetic fuzzy systems: current framework and new trends. En *Proceedings joint 9th IFSA world congress and 20th NAFIPS international conference (Cat. No. 01TH8569)*, volumen 3, páginas 1241–1246. IEEE.
- [Dem06] Demšar J. (2006) Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* 7: 1–30.
- [DG14] D. García A. González R. P. (2014) Overview of the slave learning algorithm: A review of its evolution and prospects. *International Journal of Computational Intelligence Systems* 7 (6): 1194–1221.
- [ES93] Eshelman L. y Schaffer J. (1993) Real-coded genetic algorithms and interval-schemata. *Found. Genetic Algorithms* 2: 187–202.
- [Esh91] Eshelman L. (1991) The chc adaptive search algorithm: How to have safe ser685 ach when engaging in nontraditional genetic recombination. *Foundations of Genetic Algorithms* 1: 265–283.
- [Eur18] Europea C. (2018) Directrices éticas para una ia fiable. *Grupo independiente de expertos de alto nivel sobre Inteligencia Artificial* .

- [FAN<sup>+</sup>12] Fazzolari M., Alcalá R., Nojima Y., Ishibuchi H. y Herrera F. (2012) A review of the application of multiobjective evolutionary fuzzy systems: Current status and further directions. *IEEE Transactions on Fuzzy systems* 21(1): 45–65.
- [FAN<sup>+</sup>13] Fazzolari M., Alcalá R., Nojima Y., Ishibuchi H. y Herrera F. (2013) A review of the application of multi-objective evolutionary systems: Current status and further directions. *IEEE Trans. Fuzzy Syst.* 21(1): 46–65.
- [FD18] F.K. Dosilović M. Brcić N. H. (2018) Explainable artificial intelligence: A survey. *International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* 41(210-215): 5.
- [Fer05] Ferrández A. (2005) Curvas y tablas de crecimiento (estudios longitudinal y transversal). *Fundación Andrea Prader*, .
- [FGG<sup>+</sup>18] Fernández A., García S., Galar M., Prati R. C., Krawczyk B. y Herrera F. (2018) Performance measures. En *Learning from Imbalanced Data Sets*, páginas 47–61. Springer Cham, 1 edition.
- [FHC<sup>+</sup>19] Fernandez A., Herrera F., Cordon O., del Jesus M. J. y Marcelloni F. (2019) Evolutionary fuzzy systems for explainable artificial intelligence: Why, when, what for, and where to? *IEEE Computational intelligence magazine* 14(1): 69–81.
- [FKBS<sup>+</sup>17] Flynn J. T., Kaelber D. C., Baker-Smith C. M., Blowey D., Carroll A. E., Daniels S. R., de Ferranti S. D., Dionne J. M., Falkner B., Flinn S. K., Gidding S. S., Goodwin C., Leu M. G., Powers M. E., Rea C., Samuels J., Simasek M., Thaker V. V., Urbina E. M. y on Screening S. (2017) Cdc growth charts: United states. *Pediatrics* 140(3).
- [FWQ21] Fosso Wamba S. y Queiroz M. M. (2021) Responsible artificial intelligence as a secret ingredient for digital health: Bibliometric analysis, insights, and research directions. *Information Systems Frontiers* páginas 1–16.
- [Gad17] Gadd S. (2017) Computer system could kill rather than cure, doctors warn. <http://cphpost.dk/?p=92249> página 92249.

- [GAH10] Gacto M. J., Alcalá R. y Herrera F. (2010) Integration of an index to preserve the semantic interpretability in the multi-objective evolutionary rule selection and tuning of linguistic fuzzy systems. *IEEE Transactions on Fuzzy Systems* 18(3): 515–531.
- [GAH11] Gacto M., Alcalá R. y Herrera F. (2011) Interpretability of linguistic fuzzy rule-based systems: an overview of interpretability measures. *Inf. Sci.* 181(20): 4340–4360.
- [GGSA14] Galende M., Gacto M. J., Sainz G. y Alcalá R. (2014) Comparison and Design of Interpretable Linguistic vs. Scatter FRBSs: GM3M Generalization and New Rule Meaning Index (RMI) for Global Assessment and Local Pseudo-Linguistic Representation. *Information Sciences* 282: 190–213.
- [GH06] Geng L. y Hamilton H. J. (2006) Interestingness measures for data mining: A survey. *ACM Computing Surveys* 38(3): 1–32.
- [GH08] García S. y Herrera F. (2008) An extension on statistical comparisons of classifiers over multiple data sets for all pairwise comparisons. *Journal of Machine Learning Research* 9: 2677–2694.
- [GM05] Guillaume S. y Magdalena L. (2005) An or and not implementation that improves linguistic rule interpretability. *11th World Congress of International Fuzzy Systems Association (IFSA)* páginas 88–92.
- [Gun17] Gunning D. (2017) Explainable artificial intelligence (xai). *Defense advanced research projects agency (DARPA)*, nd *Web* 2(2): 1.
- [HH09] Hühn J. y Hüllermeier E. (2009) Furia: an algorithm for unordered fuzzy rule induction. *Data Mining and Knowledge Discovery* 19 (3): 293–319.
- [HHF99] Holmes G., Hall M. y Frank E. (1999) Generating rule sets from model trees. En *Proceedings of the 12th Australian Joint Conference on Artificial Intelligence: Advanced Topics in Artificial Intelligence*, volumen 1747 of *Lecture Notes on Computer Science*, páginas 1–12. Springer-Verlag.

- [HKC01] Hong T., Kuo C.-S. y Chi S. (2001) Trade-off between computation time and number of rules for fuzzy mining from quantitative data. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* 9: 587–604.
- [HLV95] Herrera F., Lozano M. y Verdegay J. L. (1995) Tuning fuzzy logic controllers by genetic algorithms. *International Journal of Approximate Reasoning* 12(3-4): 299–315.
- [HM00] Herrera F. y Martínez L. (2000) A 2-tuple fuzzy linguistic representation model for computing with words. *IEEE Transactions on Fuzzy Systems* 8(6): 746–752.
- [INN04] Ishibuchi H., Nakashima T. y Nii M. (2004) *Classification and modeling with linguistic information granules: Advanced approaches to linguistic Data Mining*. Springer Science & Business Media.
- [IY05] Ishibuchi H. y Yamamoto T. (2005) Rule weight specification in fuzzy rule-based classification systems. *IEEE Transactions on Fuzzy Systems* 13(4): 428–435.
- [JJS<sup>+</sup>17] Jones R. E., Jewell J., Saksena R., Ramos Salas X. y Breda J. (2017) Overweight and obesity in children under 5 years: surveillance opportunities and challenges for the who european region. *Frontiers in public health* 5: 58.
- [KM97] Kubat M. y Matwin S. (1997) Addressing the curse of imbalanced training sets: One-sided selection. En *Fourteenth International Conference on Machine Learning (ICML)*, páginas 179–186. Morgan Kaufmann, San Francisco, CA, USA.
- [KOGS<sup>+</sup>20] Kuczmariski R., Ogden C., Grummer-Strawn L., Flegal K., Guo S., Wei R., Mei Z., Curtin L., Roche A. y Johnson C. (2020) Cdc growth charts: United states. *Advance data* 314: 1–27.
- [LCAGL<sup>+</sup>21] Llorente-Cantarero F. J., Aguilar-Gómez F. J., Leis R., Bueno G., Rupérez A. I., Anguita-Ruiz A., Vázquez-Cobela R., Mesa M. D., L. A. Moreno A. G., Aguilera C. M. y Gil-Campos M. (2021) Relationship between physical activity, oxidative stress, and total plasma antioxidant capacity in

- spanish children from the genobox study. *Antioxidants* 10 (2): 320.
- [Les19] Leslie D. (2019) Understanding artificial intelligence ethics and safety. *arXiv preprint arXiv:1906.05684* .
- [LHKM04] Lozano M., Herrera F., Krasnogor N. y Molina D. (2004) Real-coded memetic algorithms with crossover hill-climbing. *Evolutionary Computation* 12(3): 273–302.
- [LHP01] Li W., Han J. y Pei J. (2001) Cmar: accurate and efficient classification based on multiple class-association rules. *Proceedings IEEE International Conference on Data Mining (ICDM), California* páginas 369–376.
- [LMRGG<sup>+</sup>20] Latorre-Millán M., Rupérez A. I., González-Gil E. M., Santaliestra-Pasías A., Vázquez-Cobela R., Gil-Campos M., Aguilera C. M., Gil A., A.Moreno L., Leis R. y Bueno G. (2020) Dietary patterns and their association with body composition and cardiometabolic markers in children and adolescents: Genobox cohort. *Nutrients* 12 (11): 3424.
- [LOD<sup>+</sup>06] Lee J. M., Okumura M. J., Davis M. M., Herman W. H. y Gurney J. G. (2006) Prevalence and determinants of insulin resistance among u.s. adolescents: a population-based study. *Diabetes Care* 29(11): 2427–2432.
- [LSL<sup>+</sup>21] Li F., Shang C., Li Y., Yang J. y Shen Q. (2021) Approximate reasoning with fuzzy rule interpolation: background and recent advances. *Artificial Intelligence Review* 54(6): 4543–4590.
- [MA75] Mamdani E. H. y Assilian S. (1975) An experiment in linguistic synthesis with a fuzzy logic controller. *International journal of man-machine studies* 7(1): 1–13.
- [MANB21] Mousavi S. M., Abdullah S., Niaki S. T. A. y Banihashemi S. (2021) An intelligent hybrid classification algorithm integrating fuzzy rule-based extraction and harmony search optimization: Medical diagnosis applications. *Knowledge-Based Systems* 220: 106943.



- [Mil56] Miller G. (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. *The Psychological Review* 63: 81–97.
- [MZK08] Mansoori E. G., Zolghadri M. J. y Katebi S. D. (2008) Sgerd: A steady-state genetic algorithm for extracting fuzzy classification rules from data. *Transactions on Fuzzy Systems* 16 (4): 1061–1071.
- [(NCEP) N. C. E. P. (1992) Blood cholesterol levels in children and adolescents, national cholesterol education program (ncep): highlights of the report of the expert panel on blood cholesterol levels in children and adolescents. *Pediatrics* 89(3): 495–501.
- [Qui92] Quinlan R. J. (1992) Learning with continuous classes. En *Proceedings of the 5th Australian Joint Conference on Artificial Intelligence*, páginas 343–348.
- [Qui95] Quinlan J. (1995) Mdl and categorical theories (continued). *Machine Learning: Proceedings of the Twelfth International Conference* páginas 464–470.
- [RMS98] Ramaswamy S., Mahajan S. y Silberschatz A. (1998) On the discovery of interesting patterns in association rules. En *24rd International Conference on Very Large Data Bases*, páginas 368–379. New York.
- [ROGC<sup>+</sup>18] Rupérez A., Olza J., Gil-Campos M., Leis R., Bueno G., Aguilera C., Gil A. y Moreno L. (2018) Cardiovascular risk biomarkers and metabolically unhealthy status in prepubertal children: Comparison of definitions. *Nutrition, Metabolism and Cardiovascular Diseases* 28(5): 524–530.
- [SB75] Shortliffe E. y Buchanan B. (1975) A model of inexact reasoning in medicine. *Mathematical Biosciences* 23: 351–379.
- [SFCL03] Singhal A., Fewtrell M., Cole T. J. y Lucas A. (2003) Low nutrient intake and early growth for later insulin resistance in adolescents born preterm. *The Lancet* 361(9363): 1089–1097.
- [SNMBB21] Shaban-Nejad A., Michalowski M., Brownstein J. S. y Buckneridge D. L. (2021) Guest editorial explainable ai: towards

- fairness, accountability, transparency and trust in healthcare. *IEEE Journal of Biomedical and Health Informatics* 25(7): 2374–2375.
- [SP19] Shin D. y Park Y. J. (2019) Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior* 98: 277–284.
- [SSA<sup>+</sup>17] Skrede T., Stavnsbo M., Aadland E., Aadland K. N., Andersen S. A., Resaland G. K. y Ekelund U. (2017) Moderate-to-vigorous physical activity, but not sedentary time, predicts changes in cardiometabolic risk factors in 10-y-old children: the active smarter kids study. *The American Journal of Clinical Nutrition* 105(6): 1391–1398.
- [TGM<sup>+</sup>17] Triguero I., González S., Moyano J. M., García S., Alcalá-Fdez J., Luengo J., Fernández A., del Jesús M. J., Sánchez L. y Herrera F. (2017) Keel 3.0: An open source software for multi-stage analysis in data mining. *International Journal of Computational Intelligence Systems* 10: 1238 – 1249.
- [TK00] Tan P. y Kumar V. (2000) Interestingness measures for association patterns: a perspective. *Proceedings of the KDD-2000 workshop on post processing in machine learning and data mining* .
- [WM92] Wang L. y Mendel J. (1992) Generating fuzzy rules by learning from examples. *IEEE Transactions on Systems, Man and Cybernetics* 22(6): 1414–1427.
- [YH03] Yin X. y Han J. (2003) Cpar: Classification based on predictive association rules. *Proceedings of 3rd SIAM International Conference on Data Mining (SDM), San Francisco, CA, USA* páginas 331–335.
- [YYL<sup>+</sup>21] Yang L.-H., Ye F.-F., Liu J., Wang Y.-M. y Hu H. (2021) An improved fuzzy rule-based system using evidential reasoning and subtractive clustering for environmental investment prediction. *Fuzzy sets and systems* 421: 44–61.
- [ZLT01] Zitzler E., Laumanns M. y Thiele L. (2001) SPEA2: Improving the strength pareto evolutionary algorithm for multi-objective optimization. En *Proc. Evolutionary Methods for*

*Design, Optimization and Control with App. to Industrial Problems*, páginas 95–100. Barcelona, Spain.