

UNIVERSIDAD DE GRANADA
Departamento de Ciencias de la Computación en
Inteligencia Artificial



**Análisis del Éxito Académico del Estudiantado de
Educación Superior mediante Reglas de Asociación**

Director: **Jorge Casillas, PhD.**

Memoria presentada por
Jorge Humberto Guanin Fajardo

Para optar por el título de:
Máster Oficial de Ciencia de Datos e Ingeniería de Computadores

Julio 2021

Aprendes de los errores, pero el éxito te da el valor para seguir adelante y hacer aún más.

-Cristiaan Barnard-

Declaración

La memoria titulada “**Análisis del Éxito Académico del Estudiantado de Educación Superior Mediante Reglas de Asociación**” que presenta D. Jorge Humberto Guanín Fajardo para optar al grado de *Master en Ciencia de Datos e Ingeniería de Computadores*, que ha sido realizado en la Universidad de Granada, bajo la dirección del Doctor D. Jorge Casillas Barranquero.

Hasta donde nuestro conocimiento alcanza, en la realización del presente trabajo, se han respetado los derechos de otros autores a ser citados cuando se han utilizado sus resultados o publicaciones. Los datos analizados han contado con la respectiva autorización y han sido estudiado de forma anónima.

Granada, 5 de Julio del 2021

Jorge Guanin Fajardo Jorge Casillas

Análisis del Éxito Académico del Estudiantado de Educación Superior mediante Reglas de
Asociación
Julio 2021

Agradecimientos

Al rector de la Universidad Técnica Estatal de Quevedo D. Eduardo Díaz Ocampo por permitir el acceso a los datos que fueron analizados en este trabajo.

A mi director del trabajo Jorge Casillas, a Sara Mariottini por su aporte incondicional.

A la familia Villegas-Barranco, a Loren por sus orientaciones de “Madre Nodrisa” y de manera especial, a mi Profesor Salvador Villegas Guillén (+) aunque su meditación extensa y profunda negó acercarme como hubiera querido, logramos ser cómplices del reto de vivir y lo frágil de nuestras vidas. El aporte de sus libros fue importante recurso para el desarrollo de este trabajo.

Abstract

In this paper we study the academic success of students who complete a university degree, based on their initial enrolment. Variables related to faculty, student body and socio-economic status are analysed by means of association rules. A brief study of the art and data exploration techniques with visualisation were used to reveal the groups with significant data.

The fruitfulness of the work lies in the proposed data mining steps: data pre-processing, data balancing, search and evaluation of patterns within the data. For this, first missing values were replaced, then features were filtered using the ant colony algorithm, followed by noise filtering and, finally, the synthetic data balancing algorithm. The core of the work is focused on association rules combined with fuzzy logic and genetic algorithm. The most relevant results were found in the fuzzy association rules with genetic algorithms, the crossover of the population of 10%, 50% and 80% individuals were studied. The latter was able to find rules that anticipate academic success without the student having a first enrolment. The implications of the work relate to the use of the extracted knowledge and need to be taken with great caution due to changes forced by the effects of the COVID19 pandemic. The limitations of the work have been considered with respect to the university structure or context of the teaching staff and student body, i.e. to study as a complement the means or resources available to the institution to channel students towards the satisfactory completion of the university degree.

Resumen

En este trabajo se estudia el éxito académico del estudiantado que finaliza la titulación universitaria, a partir de su matrícula inicial. Las variables relacionadas al profesorado, estudiantado y situación socio-económica son analizadas por medio de las reglas de asociación. Un breve estudio del arte y técnicas exploración de datos con visualización fueron estudiadas para revelar los grupos con datos significativos.

La fecundidad del trabajo estriba en las etapas de minería de datos planteadas: el preprocesado de datos, balanceo de datos, búsqueda y evaluación de patrones dentro de los datos. Para esto, primero se remplazaron los valores ausentes, después fueron filtradas las características empleando el algoritmo de colonia de hormigas, seguido, se realizó filtrado de ruido y, por último, el algoritmo sintético de balanceo de datos. El núcleo del trabajo está centrado en las reglas de asociación combinadas con lógica difusa y a su vez con algoritmo genético. Los resultados con mayor relevancia fueron hallados en las reglas de asociación difusa con algoritmos genéticos, se estudiaron el cruce de la población de individuos 10%, 50% y 80%. Este último, fue capaz de encontrar reglas que anticipan el éxito académico sin que el estudiante tenga una primera matrícula. Las implicaciones del trabajo están relacionadas con el uso del conocimiento extraído y que es necesario se lo tome con mucha cautela debido a cambios forzados por efectos de la pandemia COVID19. Las limitaciones del trabajo se han considerado respecto a la estructura o contexto universitario del profesorado y estudiantado, es decir, estudiar como complemento los medios o recursos de los que dispone la institución para encauzar al alumnado hacia la finalización satisfactoria del grado universitario.

Índice general

Índice de figuras	xv
Índice de tablas	xvii
1. Fundamentos	5
1.1. Contexto	5
1.1.1. Preliminares y conceptos básicos	5
1.1.2. ¿Qué es la Minería de Datos?	5
1.1.3. Técnicas de minería de datos	6
1.2. Extracción del conocimiento	7
2. Fundamentos teóricos: estado del arte	9
2.1. Lógica difusa	9
2.2. Algoritmos genéticos	9
2.3. Medidas de proximidades	10
2.3.1. Distancias	10
2.3.2. Similitudes	10
2.4. Reglas de asociación	11
2.4.1. Conceptos básicos	11
2.4.2. Medidas de calidad de reglas de asociación	12
2.5. Búsqueda de patrones frecuentes	14
2.5.1. Apriori	14
2.5.2. FP-Growth	15
2.5.3. Reglas de asociación difusas	16
2.5.4. Reglas de asociación genético difusas	19
2.6. Estudio sistemático de las reglas de asociación en la educación superior	20
2.6.1. Estadística del estudio sistemático	31

3. Análisis exploratorio de datos	35
3.1. Estudio descriptivo	35
3.2. Visualización interactiva de datos	36
3.2.1. Diámetro	36
3.2.2. Densidad	37
3.2.3. Modularidad	37
3.3. Agrupamiento	37
3.3.1. K-means	39
3.3.2. Particionamiento (K-medoides)	42
3.3.3. Agrupamiento jerárquico	43
4. Metodología	47
4.1. Preprocesamiento de los datos	47
4.1.1. Valores perdidos	47
4.1.2. Selección de instancias	48
4.1.3. Selección de características	50
4.1.4. Balanceo de ejemplos	53
4.2. Extracción de conocimiento	53
4.2.1. Apriori	54
4.2.2. Apriori difuso	57
4.2.3. Apriori difuso y algoritmo genético	62
4.3. Análisis de la extracción del conocimiento	65
5. Conclusiones	67
Bibliografía	69
Anexo Descripción de los datos	77
A.1. Conjunto de datos	77
A.1.1. Estudio descriptivo del conjunto de datos	79
Enfoque de extracción de reglas de asociación Semi-Supervisadas con apriori	81
B.2. Representación gráfica de las reglas	81
C.3. Reglas extraídas y métricas de calidad	83
Enfoque de extracción de reglas de asociación Difusas	85
D.4. Representación gráfica de las reglas	85
E.5. Reglas extraídas y métricas de calidad	89

Enfoque de extracción de reglas de asociación difusas y algoritmo genético	97
F.6. Cruce	97
F.6.1. Probabilidad de cruce 10 %	97
F.6.2. Probabilidad de cruce 50 %	112
F.6.3. Probabilidad de cruce 80 %	116
Appendices	77

Índice de figuras

1.1.	Modelo de extracción del conocimiento planteado por Fayyad	7
2.1.	FPGrowth	16
2.2.	Enfoque de flujo de datos propuesto en primera instancia, figura tomada de (Hong et al., 2001).	17
2.3.	Enfoque de flujo de datos propuesto en primera instancia, tomado de (Hong et al., 2001).	18
2.4.	Esquema de construcción de reglas difusas basadas en funciones de pertenencia usando algoritmos genéticos, tomado de (Hong et al., 2006)	19
2.5.	Porcentajes de documentos recuperados en Scopus e Isi-Web of Science . .	32
2.6.	Porcentajes de tipos de documentos recuperados	33
2.7.	Número de publicaciones por año	34
3.1.	Representación del tamaño de comunidades	37
3.2.	Detección de comunidades y hubs	38
3.3.	Agrupamiento de datos usando distancia euclidiana	40
3.4.	Clúster con distancia de similitud según Pearson	41
3.5.	Clúster con distancia disimilitud Bray	42
3.6.	K-medioide óptimo	43
3.7.	Agrupamiento Jerarquico	44
4.1.	Los valores observados se encuentran marcados con fondo azul, mientras que los valores perdidos se han marcado de color blanco.	49
4.2.	La figura muestra los tiempos de graduación del estudiantado. El algoritmo de filtrado de instancias ha usado 15, 50, 100, 500, 1000 iteraciones para la eliminación de instancias con ruido. La iteración 15 es la que filtra de forma menos desigual.	50
4.3.	Distintos métodos usados para la extracción de reglas de asociación. Filtrado de reglas empleando métricas de calidad y presentación de mejores reglas. .	54

4.4. Reglas de asociación difusas mayormente representativas, estas reglas se han filtrado según la métrica Lift y además el número de términos del consecuente que se ajusta a un solo término, así como el antecedente en un intervalo entre dos y 5 términos. En la rueda de dependencia con 54 reglas el soporte es 0,5 y la confianza 0,1. Mientras que en la rueda de 36 reglas el soporte es 0,5 y la confianza 0,8.	60
4.5. Mapa de calor con reglas de asociación difusas	61
4.6. Reglas de asociación usando variable tiempoGraduacion como parte del consecuente.	64
4.7. Resumen de mapas de calor	65
1. Semi-Supervisado. Mapa de calor del conjunto de transacciones	82
2. Semi-Supervisado. Filtrado de reglas relevantes de acuerdo con lift	83
3. Representación gráfica de reglas con soporte 0,3	86
4. Representación gráfica de reglas con soporte 0,4	87
5. Representación gráfica de reglas con soporte 0,5	88
6. Mapa de calor con la construcción de reglas de asociación difusas	89
7. soporte 3%	108
8. soporte 0,4% y confianza 0,5% ambas figuras tienen el mismo número de reglas, no obstante, su diferencia se ha definido por el atributo, A1 y A2 respectivamente.	109
9. soporte 0,4%	110
10. Mapa de calor reglas de reglas de asociación probabilidad de cruce 10% . .	111
11. Rueda dependencia reglas de asociación	112
12. Mapa de calor de construcción de reglas	113
13. soporte 3%	116
14. Soporte de la regla 0,4%, confianza 0,7%	117
15. Soporte de la regla 0,5%, confianza 0,2%	118
16. Mapa de calor de reglas de asociación difusas genéticas	119

Índice de tablas

2.1. Fuentes digitales y total de documentos recuperados desde revistas, conferencias, simposios, entre otros repositorios	21
2.2. Documentos recuperados de Scopus e Isi Web of Science aptos y no aptos .	22
2.3. Valoración de documentos recuperados y aptos para la revisión sistemática .	25
2.4. Publicaciones aptas para el estudio sistemático	26
2.5. Tópicos del estudio sistemático mediante la búsqueda de patrones o combinaciones con otras técnicas de minería de datos	31
4.1. Selección de instancias, proporción de eliminación de ruido	48
4.2. Iteraciones de la selección de características valoración del algoritmo J48, se presentan las distintas métricas con el rendimiento y el número de variables de cada iteración, se ha usado 21 hormigas.	52
4.3. Balanceo de ejemplos con algoritmo SMOTE	53
4.4. Principales reglas seleccionadas.	55
4.5. Conjunto de reglas de asociación difusa seleccionadas	57
4.6. Conjunto de reglas de asociación genético difusas	62
1. Descripción de variables del conjunto de datos	77
2. Estudio descriptivo del conjunto de datos	80
3. Reglas extraídas a partir del control del consecuente	84
4. Conjunto de reglas de asociación difusas obtenidas a partir del algoritmo FuzzyApriori. Las reglas han sido filtradas de acuerdo con los criterios iniciales	89
5. Conjunto de reglas de asociación difusas genéticas (probabilidad de cruce 10%)	97
6. Terminos ligüísticos de conjunto de reglas de asociación (probabilidad de cruce 10%)	104

7. Terminos ligüísticos de conjunto de reglas de asociación (probabilidad de cruce 50 %)	113
8. Conjunto de reglas de asociación difusas genéticas (probabilidad de cruce 80 %)	120
9. Terminos ligüísticos de conjunto de reglas de asociación (probabilidad de cruce 80 %)	120

Introducción

Motivación

Uno de los mayores retos a los que se enfrentan hoy en día las instituciones de educación superior (IES) es mejorar la calidad de las decisiones de su gestión, además, también se enfrentan al desafío de formar profesionales idóneos que respondan de manera efectiva a la solución de los nuevos retos de la globalización. Este hecho, tiene una consecuencia importante para que las IES consigan calificar como las mejores instituciones del mundo. La diacronía evolutiva de las IES es determinante a la hora de estar mejor o peor posicionada. Así, las IES optan por aplicar programas y estrategias que sean precursoras del alto impacto educativo.

La Universidad o IES como entidad, se encuentra rodeada de numerosas normativas que deben de cumplirse para establecer y conseguir sus objetivos. En cada país se suministran recursos para ejecutar y facilitar el desarrollo universitario, siendo sus pilares fundamentales la actividad del docente y la del estudiantado. Es difícil no valorar la importancia de la docencia como una labor compleja dada la heterogeneidad del alumnado. La tarea principal del profesorado universitario es propiciar elementos y una didáctica precisa para que el alumnado consiga aprender. Los resultados de este aprendizaje conlleva a la formación de profesionales emprendedores que aporten diversas soluciones a los retos de la globalización, o a su vez, profesionales científicos que destaquen en el desarrollo de ciencia y tecnología.

Hoy en día, el término de universidad inteligente (smart university) ha incrementado su participación en las IES, ya que la innovación tecnológica permite ganar eficiencia y eficacia en los procesos institucionales. La aplicación de técnicas de inteligencia artificial han permitido tomar ventajas respecto a otras IES para aplicar estrategias que consigan impulsar el desarrollo universitario y mejorar la calidad de las decisiones.

El proceso de adopción de decisiones administrativas se vuelve más complejo a medida que aumenta la complejidad de las entidades educativas. Las instituciones educativas buscan una tecnología más eficiente para gestionar y apoyar mejor los procedimientos de adopción de decisiones o ayudarlas a establecer nuevas estrategias y planificar una mejor gestión de

los procesos actuales. Una forma de abordar eficazmente los retos para mejorar la calidad es proporcionar al sistema de gestión nuevos conocimientos relacionados con los procesos educativos y las entidades. Ese conocimiento puede extraerse de los datos históricos y operacionales que residen en las bases de datos de la organización educativa utilizando las técnicas de la tecnología de extracción de datos. Las técnicas de minería de datos son instrumentos analíticos que pueden utilizarse para extraer conocimientos significativos de grandes conjuntos de datos. En este documento se presentan las capacidades de la minería de datos en el contexto del sistema de enseñanza superior mediante i) la propuesta de una directriz analítica para que las instituciones de enseñanza superior mejoren sus actuales procesos de adopción de decisiones, y ii) la aplicación de técnicas de minería de datos para descubrir nuevos conocimientos explícitos que puedan ser útiles para los procesos de adopción de decisiones ([DELAVERI et al., 2008](#)).

Planteamiento del problema

Los métodos convencionales para la búsqueda de patrones que estimen una tasa alta de estudiantes titulados quedan reducidas a segundo plano por el empleo de las distintas técnicas de minería de datos. En este aspecto, como en muchos otros se requiere de información fiable que permita a las instituciones de educación superior aprovechar los datos que de por si son una fuente muy poderosa de conocimiento. Tener el control sobre estos es relevante. Las estrategias que se apliquen a partir de la extracción de conocimiento será la que consiga valorar en positivo las acciones pertinentes. El trabajo redundante o poco asertivo no garantiza el liderazgo de una entidad. Así, el reto de hacerse con el control de los datos y descubrir el conocimiento oculto de estos, hace que en este trabajo se proponga la minería de datos mediante las reglas de asociación como una herramienta para conocer dichos patrones ocultos en los datos, y que resuelvan las siguientes preguntas de investigación:

- ¿Cuáles son las características comunes del estudiantado que ha culminado su grado universitario?
- ¿Qué otra técnica se combina mejor con las reglas de asociación?
- ¿A día de hoy cuál ha sido la mejora en el campo educacional y las reglas de asociación?

Objetivos

Considerando los motivos que han llevado a realizar este trabajo, el propósito general es analizar mediante reglas de asociación los factores que conducen al alumnado el tiempo de

graduación desde que realiza la primera matrícula.

El objetivo general anteriormente descrito tiene los siguientes objetivos específicos:

- Detectar grupos homogéneos de información que tengan un patrón de comportamiento de los estudiantes con el tiempo de graduación del alumnado.
- Investigar las variantes e hibridaciones de las reglas de asociación en la minería de datos.
- Identificar los factores influyentes en el proceso de titulación del estudiantado.
- Presentar y revisar el análisis experimental de los algoritmos seleccionados.

Contribuciones

La fecundidad de un trabajo debe estar fundamentada en el descubrimiento de nuevos fenómenos o al menos en la comprobación de posibles teorías. Partiendo de este hecho y como consecuencia del trabajo de fin de master se ha desprendido un trabajo que ha sido presentado en la *Revista Conrado*.

Guanin-Fajardo, J., Casillas, J., y Chiriboga-Casanova, W. (2019). Semi-supervised learning to discover the average scale of graduation of university students. *REVISTA CONRADO*, 15(70):291–299.

Estructura del trabajo

Para lograr los anteriores objetivos, la memoria se ha ordenado en cinco capítulos, a los que precede esta introducción. Fundamentos, donde revisamos los conceptos del trabajo. El siguiente capítulo 2 que trata sobre los fundamentos teóricos y el estado del arte de las reglas de asociación. El capítulo 3 se ha realizado el estudio exploratorio de los datos, con el fin de estudiar a detalle el comportamiento de las variables. El capítulo 4 se ha realizado la metodología del trabajo. Por último en el capítulo 5 se analizan las principales conclusiones del trabajo.

Capítulo 1

Fundamentos

1.1. Contexto

1.1.1. Preliminares y conceptos básicos

La predicción temprana del rendimiento de los estudiantes es una tarea desafiante pero crucial en la educación superior moderna. Los sistemas de alerta temprana han demostrado ser herramientas poderosas para la identificación temprana de estudiantes en riesgo de fracaso o abandono. Sin embargo, hay un gran número de factores que influyen en el aprendizaje y el rendimiento de los estudiantes. Los factores socio-demográficos juegan un papel importante como predictores del éxito académico. Los estudios han identificado asociaciones entre los cursos, la privación socio-económica, vecindario donde vive, el género, la raza y la educación de los padres, entre otros, y los estudiantes en riesgo([Cano y Leonard, 2019](#)).

1.1.2. ¿Qué es la Minería de Datos?

En la actualidad existen muchas definiciones emitidas en materiales bibliográficos y científicos. Algunas conceptualizaciones las tenemos a continuación:

Definición 1 La Minería de Datos permite la construcción de modelos basados en los datos para fabricar nuevos conocimientos que puedan ser utilizados por el usuario. El modelo obtenido es una descripción de patrones y relaciones entre los datos, que pueden ser utilizados para hacer predicciones en un área particular, comprender mejor el dominio, mejorar el desempeño o explicar situaciones pasadas ([Cadenas y Garrido, 2018](#)).

Definición 2 La minería de datos es la etapa central del proceso de descubrimiento de conocimientos que tiene como objetivo la extracción de información de interés no

trivial, implícita, previamente desconocida y potencialmente útil a partir de datos de grandes bases de datos ([Fürnkranz et al., 2012](#)).

Definición 3 En sentido estricto, la minería de datos es un proceso de descubrimiento de información valiosa a partir de grandes cantidades de datos almacenados en bases de datos, almacenes de datos u otros depósitos de información. Esta valiosa información puede ser como patrones, asociaciones, cambios, anomalías y estructuras significativas([Fayyad et al., 1996](#)). Es decir, la minería de datos intenta extraer de los datos conocimientos potencialmente útiles.

Basados en las definiciones anteriores, tanto la minería de datos como la educación han combinado sus áreas de estudio para obtener mayor rédito de los datos. Así, [Alshamaila et al. \(2020\)](#) ha definido que la minería de datos orientada a entornos educativos se centra en el desarrollo o aplicación de nuevos algoritmos inteligentes dedicados a analizar los datos resultantes de las bases de datos educativas, a fin de comprender mejor a los estudiantes y sus entornos de aprendizaje. Desde otro punto de vista, [Cano y Leonard \(2019\)](#) ha definido la aplicabilidad que la minería de datos educacional se extiende a los sistemas de alerta temprana inteligentes. Se basan en los datos de estudiantes donde se emiten avisos anticipados del comportamiento o rendimiento académico del estudiante que está en riesgo de fracaso o abandono de sus estudios.

1.1.3. Técnicas de minería de datos

La minería de datos está centrada en dos enfoques: El primero, el *modelo predictivo* que consiste en identificar patrones para estimar valores futuros usando atributos o variables predictivas. El segundo, el *modelo descriptivo* que sirve para identificar patrones que expliquen los datos. Además, en minería de datos se ejecutan diferentes tipos de procedimientos que consiguen un tipo de “conocimiento” que sin lugar a dudas será distinto según cada modelo empleado. En el predictivo por ejemplo, se encuentra tanto la clasificación como la regresión, mientras que, en el descriptivo se haya el agrupamiento de datos (clustering) y las reglas de asociación([Cadenas y Garrido, 2018](#)).

1.2. Extracción del conocimiento

Este término de extracción del conocimiento (Knowledge Discovery in Databases, KDD) es el proceso no trivial de identificar patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles a partir de los datos. El modelo ha sido planteado por [Fayyad et al.](#) donde se han planteado una sucesión de etapas para la extracción del conocimiento.

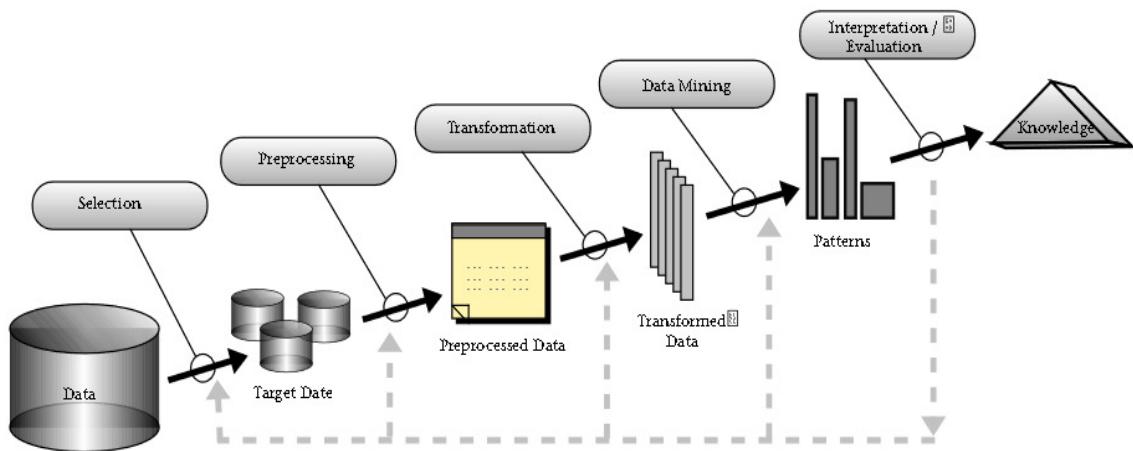


Figura 1.1 Modelo de extracción del conocimiento planteado por Fayyad

Capítulo 2

Fundamentos teóricos: estado del arte

Para que un estudio sea fecundo, debe fundamentarse en teorías y hechos que demuestren su naturaleza. Claude Bernal, manifestó que “*El experimentador que no sabe lo que está buscando, no comprenderá lo que encuentra.*” En este capítulo se han revisado trabajos relacionados con reglas de asociación enfocados al estudio de datos académicos. Por otra parte, también abordamos los principales conceptos de las técnicas de minería de datos no supervisada que se usan en este trabajo.

2.1. Lógica difusa

La lógica difusa o borrosa, como su nombre lo indica, es una alternativa a la lógica clásica que pretende introducir un grado de vaguedad¹ en las cosas que clasifica. Por otra parte, el conocimiento vago, impreciso, incierto, ambiguo, inexacto, o probabilista por naturaleza. El razonamiento y pensamiento humano frecuentemente conlleva información de este tipo, probablemente originada de la inexactitud inherente de los conceptos humanos y del razonamiento basado en experiencias similares pero no idénticas a experiencias anteriores.

2.2. Algoritmos genéticos

Un algoritmo genético es una técnica de programación que imita a la evolución biológica como estrategia para búsqueda y optimización en la resolución problemas. Este algoritmo refleja el proceso de selección natural donde los individuos más aptos son seleccionados para la reproducción con el fin de producir descendencia de la próxima generación.

¹Imprecisión, falta de exactitud.

2.3. Medidas de proximidades

Varias de las medidas comúnmente usadas para el análisis de los datos se emplean para conseguir un estudio en concreto. Para ello, se usan distancias o similitudes que existe entre los datos, según la técnica exploratoria a usarse.

2.3.1. Distancias

Euclídea

Esta es una de las distancias que se ha utilizado en muchas aplicaciones de la ciencia en general. La distancia euclídea, es la distancia “ordinaria” entre dos puntos de un espacio euclídeo, la cual se describe en la ecuación 2.3.1 y que se desprende a partir del teorema de Pitágoras.

$$d_{ij} = \sqrt{\sum (x_{ia} - x_{ja})^2} \quad (2.1)$$

Jaccard

El índice de Jaccard , también conocido como intersección sobre la unión y el coeficiente de similitud de Jaccard (acuñado originalmente coeficiente de Communauté por Paul Jaccard) (Guillet y Hamilton, 2007), que es una estadística utilizada para comparar la similitud y la diversidad de la muestra conjuntos. El coeficiente de Jaccard medidas de similitud entre conjuntos de muestras finitas, y se define como el tamaño de la intersección dividido por el tamaño de la unión de los conjuntos de la muestra, este índice lo describimos en la siguiente ecuación.

$$\frac{P(AB)}{P(A) + P(B) - P(AB)} \quad (2.2)$$

2.3.2. Similitudes

Bray-Curtis

$$\frac{P(AB)}{P(A) + P(B) - P(AB)} \quad (2.3)$$

Pearson

El coeficiente de correlación de Pearson hace referencia a la media de los productos cruzados de las puntuaciones estandarizadas de X y de Y.

$$r_{xy} = \frac{\sum Z_x Z_y}{N} \quad (2.4)$$

2.4. Reglas de asociación

En 1994, el famoso algoritmo de reglas de asociación Apriori fue presentado por R. Agrawal y otros ([Agrawal y Srikant, 1994](#)). A partir de entonces, las reglas de asociación fueron estudiadas más profundamente. Hay dos maneras de mejorar los algoritmos para aumentar la eficiencia de la minería de datos: Algoritmos basados en Apriori y algoritmos sin Apriori ([Zhang y He, 2010](#)).

En minería de datos se utilizan múltiples tipos de medidas según el problema propuesto. En [Guillet y Hamilton](#) se describen varios tipos de medidas de calidad que comúnmente se utilizan en estudios de minería de datos. Como parte de este estudio, se definen medidas de calidad para las reglas de asociación así como otras medidas que también permiten conseguir la calidad en las reglas de asociación.

2.4.1. Conceptos básicos

En su obra [Zhang y Zhang \(2002\)](#), ha considerado que las reglas de asociación pueden definirse formalmente de la siguiente manera:

$I = \{i_1, i_2, \dots, i_m\}$ que es un conjunto de ítems o artículos. Por ejemplo, ítems como leche, azúcar y pan que se compran en un supermercado; y $A_i = v$ es un artículo, donde v es un valor de dominio del atributo A_i , en una relación $R(A_1, \dots, A_n)$.

X es un conjunto de artículos si es un subconjunto de I . Por ejemplo, un conjunto de artículos para comprar en una tienda es un conjunto de artículos; y un conjunto de $A_i = v$ es un conjunto de artículos para la relación $R(PID, A_1, A_2, \dots, A_n)$, donde PID es una clave.

$D = \{t_i, t_{i+1}, \dots, t_n\}$ es un conjunto de transacciones, llamado base de datos de transacciones, donde cada transacción t tiene un tid y un t -itemset $t = (tid, t\text{-itemset})$. Por ejemplo, el carrito de la compra de un cliente que pasa por una caja es una transacción; y una tupla (v_1, \dots, v_n) de la relación $R(A_1, \dots, A_n)$ es una transacción.

Una transacción t contiene un conjunto de artículos X si, para todos los artículos, donde $i \in X$, i es un conjunto de elementos. Por ejemplo, un carrito de la compra contiene todos

los artículos en X cuando pasando por la caja; y por cada $A_i = v_i$ en X , v_i se produce en la posición i en la tupla (v_1, \dots, v_n) .

Hay una estructura de red natural en los conjuntos de artículos 2^I , a saber, la estructura de subconjunto/superconjunto. Ciertos nodos de esta red son categorías de agrupación natural de interés (algunos con nombres). Por ejemplo, artículos de un departamento particular como ropa, hardware, muebles, etc.; y, desde dentro, digamos ropa, ropa para niños, mujeres y hombres, ropa para niños pequeños, etc.

En términos generales, una regla de asociación se define como una implicación “si X entonces Y ” ($X \Rightarrow Y$), donde X e Y son itemsets o items individuales. Su estructura se compone de dos partes, la izquierda se define como antecedente (left-hand-side (LHS), del inglés) y la derecha como consecuente (right-hand-side (RHS), del inglés). Por ejemplo, la regla $\{A, B\} \Rightarrow \{C\}$ significa que, cuando ocurren A y B, entonces ocurre C.

2.4.2. Medidas de calidad de reglas de asociación

Existen dos tipos de medidas en las que se basan las reglas de asociación, el soporte, que define la cobertura de las reglas, mientras que, la confianza define el factor de certeza de la regla.

Soporte-Confianza

Soporte Típicamente, el soporte se utiliza para medir la abundancia o frecuencia (a menudo interpretada como significado o importancia) de un conjunto de elementos en una base de datos. Nos referimos a un conjunto de elementos como "conjunto de elementos frecuentes" si el soporte es mayor que un umbral mínimo de soporte especificado. Tenga en cuenta que, en general, debido a la propiedad de cierre a la baja, todos los subconjuntos de un conjunto de elementos frecuentes son también frecuentes.

$$s = (X \rightarrow Y) = s(X \cup Y) = P(AB); \text{rango : } [0, 1] \quad (2.5)$$

Confianza La confianza de una regla $A \rightarrow C$ es la probabilidad de ver el consecuente en una transacción dado que también contiene el antecedente. Obsérvese que la métrica no es simétrica ni dirigida; por ejemplo, la confianza para $A \rightarrow C$ es diferente de la confianza para $C \rightarrow A$. La confianza es 1 (máximo) para una regla $A \rightarrow C$ si el consecuente y el antecedente ocurren siempre juntos.

$$\text{confidence} = (X \Rightarrow Y) = P(B|A) = \frac{\text{soporte}(X \Rightarrow Y)}{\text{soporte}(X)}; \text{rango : } [0, 1] \quad (2.6)$$

Otras medidas de interés

A partir del trabajo de ([Agrawal y Srikant, 1994](#)) varios autores han centrado su esfuerzo en conseguir métricas que consigan evaluar la calidad de las reglas de asociación.

Lift or Interest La métrica de *lift* se utiliza comúnmente para medir con qué frecuencia el antecedente y el consecuente de una regla $A \rightarrow C$ se producen juntos de lo que esperaríamos si fueran estadísticamente independientes. Si A y C son independientes, la puntuación de *lift* será exactamente 1 ([Brin et al., 1997](#)).

$$Lift = \frac{P(B|A)}{P(B)} \text{ ó su equivalente } \frac{P(AB)}{P(A)P(B)}; rango : [0, \infty] \quad (2.7)$$

Leverage La métrica *Leverage* calcula la diferencia entre la frecuencia observada de A y C que aparecen juntos y la frecuencia que se esperaría si A y C fueran independientes. Un valor de *Leverage* de 0 indica independencia ([Piatetsky-Shapiro, 1991](#)).

$$leverage(A \rightarrow C) = support(A \rightarrow C) - support(A) \times support(C); rango : [-1, 1] \quad (2.8)$$

Conviction Un alto valor de convicción significa que el consecuente depende en gran medida del antecedente. Por ejemplo, en el caso de una puntuación de confianza perfecta, el denominador se convierte en 0 (debido a $1 - 1$) para lo cual la puntuación de convicción se define como “inf”. De manera similar al levantamiento, si los elementos son independientes, la convicción es 1 ([Brin et al., 1997](#)).

$$conviction(A \rightarrow C) = \frac{1 - support(C)}{1 - confidence(A \rightarrow C)}, rango : [0, \infty] \quad (2.9)$$

Yules's Q La métrica *Yules's Q*, puede tomar valores en el rango [-1, 1], caracteriza la correlación entre dos eventos dicotómicos posiblemente relacionados (coeficiente de asociación). Los valores superiores a 0 revelan dependencia positiva, 0 revela independencia, y los valores inferiores a 0 revelan dependencia negativa ([Altay y Alatas, 2019](#)).

$$YulesQ = \frac{P(AB)P(\neg A \neg B) - P(A \neg B)P(\neg AB)}{P(AB)P(\neg A \neg B) + P(A \neg B)P(\neg AB)}; rango : [-1, 1] \quad (2.10)$$

2.5. Búsqueda de patrones frecuentes

Evidentemente, desde hace siglos pasados, el ser humano ha tratado de buscar soluciones y resolver un determinado problema. Por ejemplo, los cazadores seguían los rastros de sus presas para atraparlas, ellos se centraban en los lugares de alimentación, tiempo, clima, entre otros factores para la cacería. Por otro lado, en el campo de la biomedicina, los investigadores ante cualquier tipo de virus estudian el comportamiento y mutación del nuevo virus y, se efectúan ensayos que consigan un tipo de droga(vacuna) que atenúa o elimine la especie invasora. A día de hoy, la ciencia en general ha utilizado métodos avanzados de computación para enfrentarse al desafío de búsqueda de soluciones a problemas de alta complejidad.

De forma general, la búsqueda de patrones o modelos son la clave para el hallazgo de soluciones que consigan satisfacer un determinado problema. Actualmente, el uso de la inteligencia artificial se ha convertido en una de las herramientas de mayor fiabilidad para conseguir resolver problemas complejos, donde la interacción de un experto humano es mínima pero fundamental para dar validez a los hallazgos. Bajo este contexto, se pretende buscar el patrón de comportamiento del estudiantado universitario que ha alcanzado el grado universitaria y, que estudiaremos en capítulos posteriores. Con este incentivo, se han considerado cuatro enfoques de búsqueda de patrones. Aunque estos, difieren entre sí, pero descansan sobre un supuesto común, que es la búsqueda de patrones.

2.5.1. Apriori

El primer enfoque, que usamos para la búsqueda de patrones frecuentes es Apriori, que entre todos los algoritmos de minería de reglas de asociación ha jugado un papel muy importante, generando impacto en investigaciones multidisciplinarias. El principio básico de Apriori es que: En la primera pasada del algoritmo simplemente se cuenta las ocurrencias de los elementos para determinar los grandes conjuntos de 1 elemento. Después, la pasada k, consiste en dos fases. Primero, los grandes conjuntos de ítems Lk-1 encontrados en la pasada (k-1) se usan para generar los conjuntos de ítems candidatos Ck, usando la función apriori-gen. Seguido, se escanea la base de datos y se cuentan los candidatos en Ck. El algoritmo terminará hasta que el conjunto de ítems candidatos Ck esté vacío ([Agrawal y Srikant, 1994; Zhang y He, 2010](#)). De manera general, el objetivo de las reglas de asociación que usan el algoritmo Apriori es detectar las relaciones o asociaciones entre valores nominales específicos en grandes conjuntos de datos.

2.5.2. FP-Growth

El segundo enfoque, que usamos para la búsqueda de patrones frecuentes es FP-Growth [Han et al. \(2004a\)](#). El algoritmo Apriori tiene una desventaja computacional, dado que la elevada cantidad de itemsets que se crean terminan afectando al número de reglas. Frente a una y otra posición [Han et al. \(2004b\)](#), ha desarrollado una alternativa para optimizar la cantidad de reglas creadas a partir de un árbol de decisión. Para ello, el algoritmo FP-Growth es un algoritmo de minería de reglas de asociación altamente eficiente. Comprime la base de datos en un árbol de patrones frecuentes (FP-Tree) y mantiene la información de las asociaciones entre los conjuntos de elementos. FP-Tree es un conjunto de estructura de datos para almacenar los datos originales. Se divide la base de datos comprimida en un conjunto de bases de datos condicional. Y cada una de las bases de datos extraídas esta, respectivamente, y asociada a un elemento frecuente . Supongamos que $c = c_1, c_3, c_3 \dots c_n$ es un conjunto de datos, donde c_i es una transacción, cada c_i es una colección de algunos elementos. Supongamos que $c_i = l_1, l_2 \dots l_v \dots$ El paso de los algoritmos de FP-Growth es el siguiente:

1. En primer lugar, convertimos la base de datos en transacciones, se ordenan los elementos en forma descendente. Luego, se produce una secuencia de elementos frecuente y recuento de soporte, guardamos los elementos frecuentes y el recuento de soporte en la cabecera.
2. En el segundo escaneo del conjunto de datos C, cada c_i es filtrado y clasificado. Los elementos con $fv < f_{min}$ son filtrados, los elementos remanentes son guardados en el encabezado con la tabla de orden. A continuación, se crea un nuevo conjunto de datos y se almacena en el CN, donde $CN = c_1, c_2, c_3 \dots c_n$ y se utilizará directamente para la construcción del árbol FP tree.
3. Construimos un árbol FP tree basado en CN-based desde un árbol vacío, cada transacción c_i se añade el número de nodo, los nodos redundantes se eliminan. Todas las transacciones se procesan de forma recursiva hasta que se completa el árbol FP tree.

El algoritmo FP-Growth. Se ejecuta en 4 pasos. Primero, Escanea la base de datos y encuentra elementos con una frecuencia mayor o igual a un umbral. Segundo, ordena los elementos frecuentes en orden decreciente, B5, C4, A3, D3, E3. Tercero, construye un árbol que sólo tenga la raíz TIDItemset 1 A, B, C, D, F 2 B, C, E 3 A, B, D 4 A, B, F 5 C, D, F 6 E, F. Por último, escanea la base de datos de nuevo; para cada muestra: a) añade los elementos de la muestra al árbol existente, utilizando sólo los elementos frecuentes (es decir, los elementos

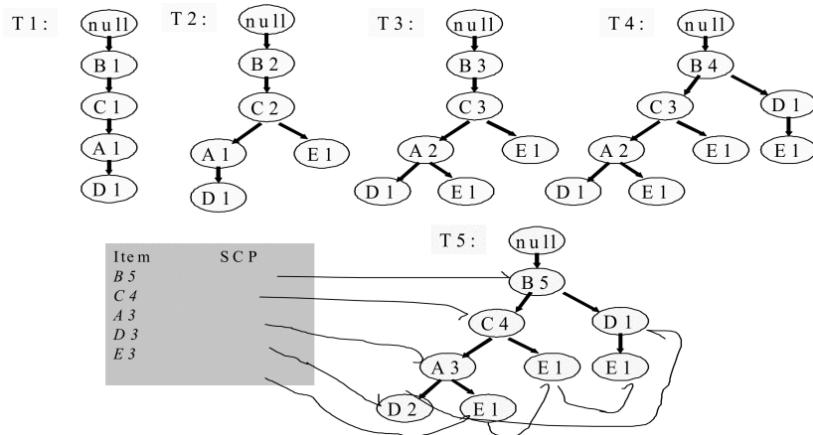


Figura 2.1 Árbol generado con FPGrowth, tomado de ([Zhang et al., 2008](#)).

descubiertos en el paso 1.) y b) Repite (a) hasta que todas las muestras hayan sido procesadas.

2.5.3. Reglas de asociación difusas

El tercer enfoque, que usamos para la búsqueda de patrones frecuentes son las reglas de asociación difusas usando el algoritmo FuzzyAP que es un algoritmo de reglas de asociación planteado para el funciones de pertenencia para transformar cada valor cuantitativo en un conjunto difuso en términos lingüísticos. Después, calcula las cardinalidades escalares de todos los términos lingüísticos en los datos de las transacciones. Cada atributo utilizó sólo el término lingüístico con la máxima cardinalidad en el proceso de extracción, manteniendo así el mismo número de elementos que el de los atributos originales. Así, el algoritmo se centra en los términos lingüísticos más importantes para reducir su tiempo de cálculo. Aunque este enfoque podía encontrar rápidamente patrones interesantes, algunos patrones podrían perderse, ya que en el proceso de extracción sólo se utilizaba el término lingüístico con la máxima cardinalidad en cada atributo ([Hong et al., 2001](#)).

Los diagramas de flujo de los enfoques de minería de datos difusos propuestos en ([Hong et al., 2001](#)) y en este trabajo se muestran respectivamente en las figuras 2.2 y 2.3. La diferencia en estas dos figuras se distingue por las regiones grises. En la Figura 2.2, sólo los términos lingüísticos con el máximo recuento para cada atributo se utilizan para formar el conjunto de candidatos. En la figura 2.3, se utilizan todos los términos lingüísticos. Sin embargo, los términos lingüísticos que pertenecen al mismo atributo no pueden pertenecer al mismo conjunto de elementos. El cálculo de la figura 2.3 es más complejo que el de la figura 2.2, ya que se utilizan todos los términos lingüísticos posibles para calcular los conjuntos de

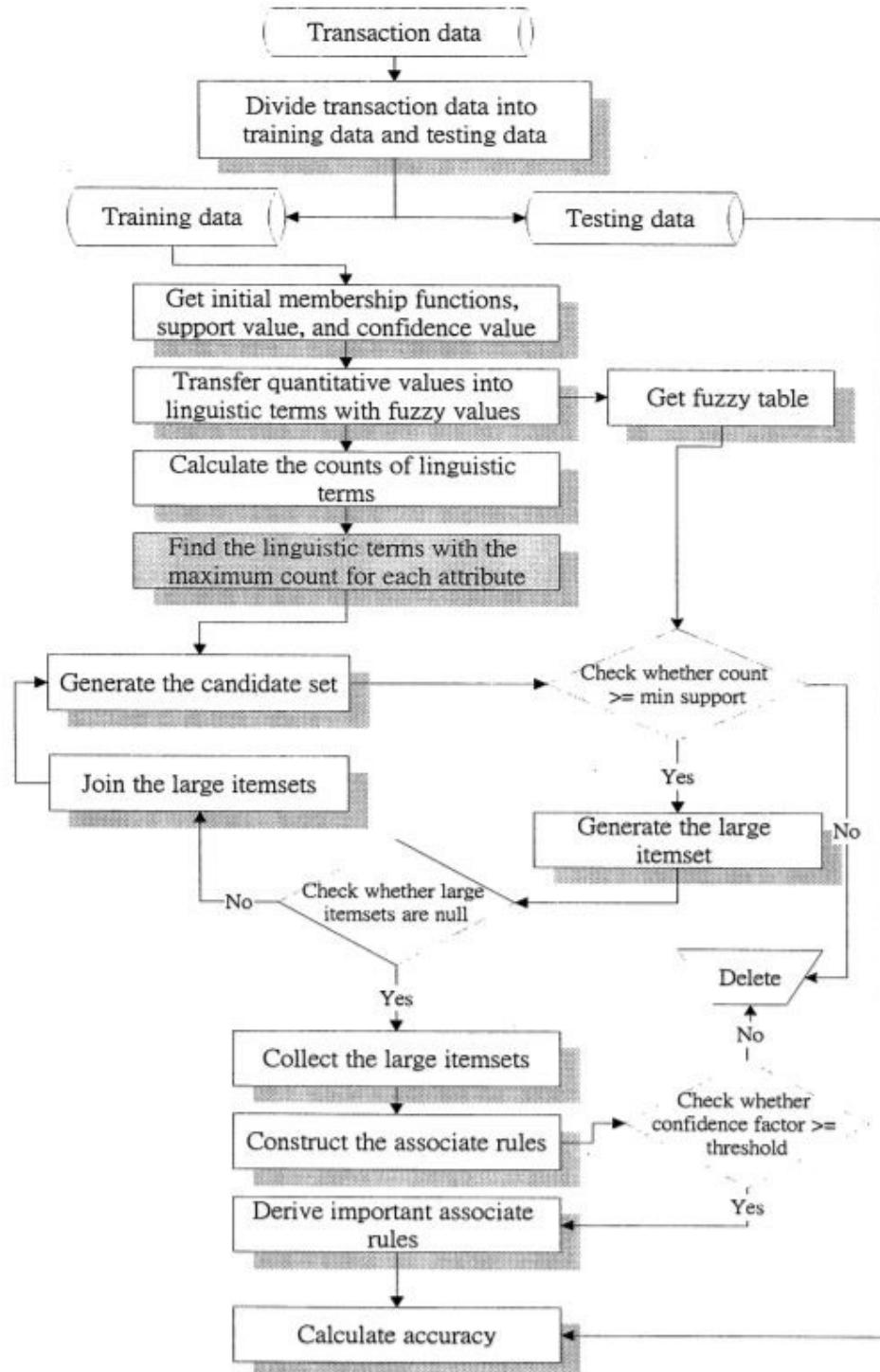


Figura 2.2 Enfoque de flujo de datos propuesto en primera instancia, figura tomada de ([Hong et al., 2001](#)).

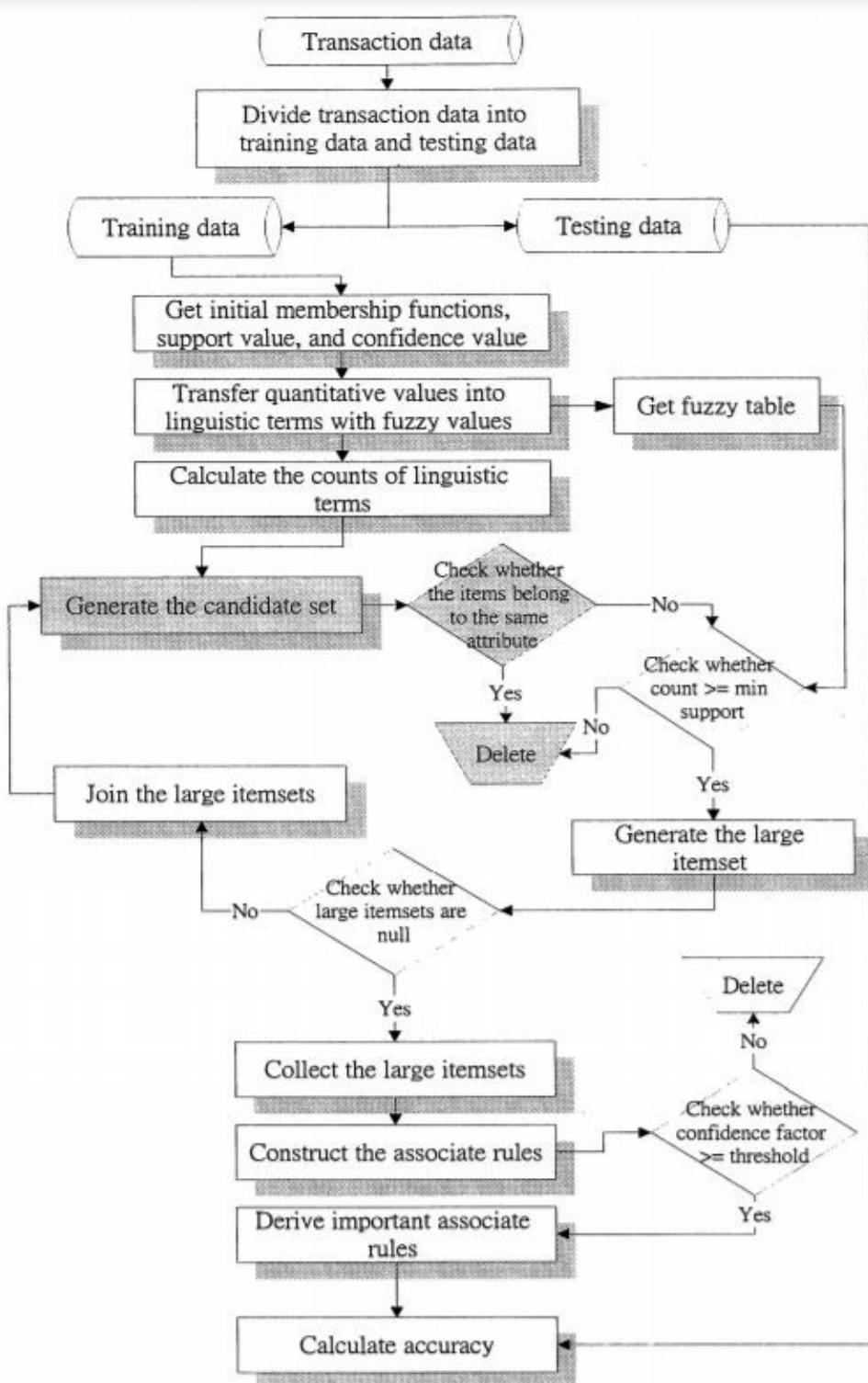


Figura 2.3 Enfoque de flujo de datos propuesto en primera instancia, tomado de ([Hong et al., 2001](#)).

elementos grandes, pero el conjunto de reglas de asociación derivado de la figura 2.3 es más completo que el de la figura 2.2.

2.5.4. Reglas de asociación genético difusas

El cuarto enfoque, se basa en la combinación de reglas difusas y los algoritmos genéticos. En el trabajo de Hong et al. (2006), el algoritmo propuesto adapta de forma dinámica las funciones de pertenencia mediante algoritmos genéticos y las utiliza para fuzzificar las transacciones cuantitativas. Es decir, mantiene una población de conjuntos de funciones de pertenencia, y utiliza el algoritmo genético para衍生自动地 la resultante. A continuación elige las cadenas apropiadas para el “*apareamiento*”, creando gradualmente buenos conjuntos de funciones de pertenencia en la descendencia. Los conjuntos de funciones de pertenencia de los hijos de funciones de pertenencia se someten a una “*evolución*” recursiva hasta que se obtiene un buen conjunto de funciones de pertenencia.

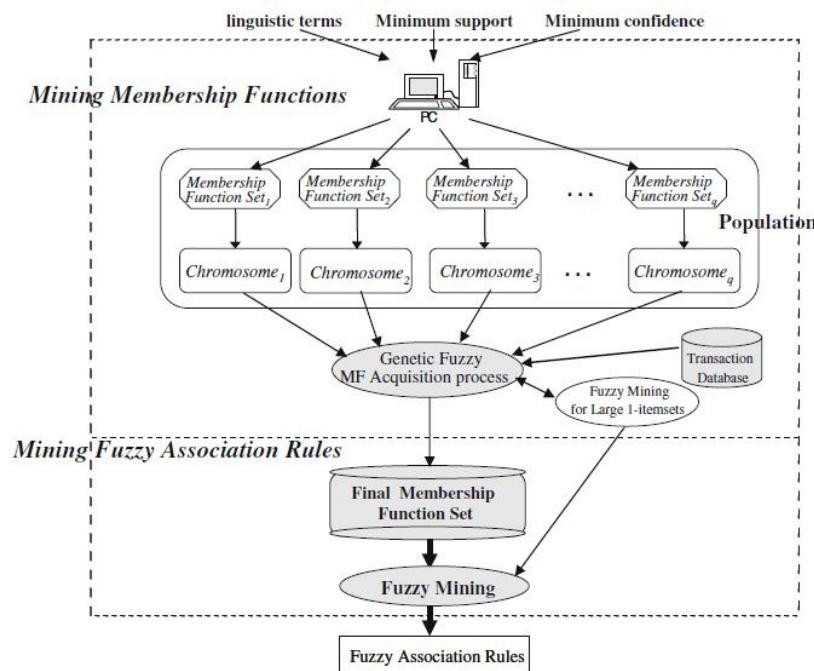


Figura 2.4 Esquema de construcción de reglas difusas basadas en funciones de pertenencia usando algoritmos genéticos, tomado de (Hong et al., 2006)

2.6. Estudio sistemático de las reglas de asociación en la educación superior

En esta sección, se describe primero la forma en que se llevó a cabo la revisión del estado del arte en la aplicación de reglas de asociación en la educación superior y luego las estadísticas. Se puede argumentar que el presente trabajo estriba en dos enfoques (Educación Superior y Minería de Datos). En particular, se detallan trabajos próximos que han estado relacionado con los resultados del proceso educativo y la obtención de patrones ocultos usando reglas de asociación. Para el estudio planteado se ha usado la herramienta de software **Parsifal**.

En torno a este contexto, se ha desarrollado el estudio sistemático para hallar información que ha influido en la búsqueda de conocimiento para la toma de decisiones y que consiga retener y/o asistir al estudiantado para que su tiempo de titulación universitaria sea corto. Esto por una parte, por otra, es evidente que las instituciones universitarias están interesadas en conocer la tasa de estudiantes con éxito en su titulación universitaria (entiéndase por éxito al culminar los estudios de grado). El éxito académico converge de acuerdo con diferentes autores ([Cuseo, 2012](#); [Internacional, 2015](#); [Morris, 2016](#); [Ruiz y Miguel, 2001](#); [Saunders-Scott et al., 2018](#)) en la participación de un grupo de factores, entre los que destacan: el nivel de empatía, hábitos de estudios, actitud favorable, profesorado, recursos digitales, nivel socio-económico, estructura familiar, entre otros. Todo esto, vinculado con la titulación universitaria seleccionada. De hecho, lo fundamental para las instituciones universitarias es anticiparse a los problemas inherentes al desarrollo y desempeño educativo con el fin de aplicar directrices precursoras del éxito académico ([Alharbi et al., 2016](#)).

El corpus para la síntesis del estudio sistemático ha sido creado recuperando documentos indexados en Scopus e IsiWeb of Science. Los documentos extraídos se relacionan con minería de datos y la educación superior. Se ha limitado la búsqueda a documentos que tengan relación con la educación media (colegios), primaria (escuelas) y otras áreas multidisciplinarias. Para ello, se ha usado la cadena de búsqueda base: (“*educational data mining*” AND “*association rule*” AND “*data mining*” AND “*unsupervised*” AND (“*educational*” OR “*college*” OR “*university*” OR “*higher education*” OR “*student*”)) AND NOT (“*high school*” OR “*secondary*” OR “*elementary*” OR “*primary*”). La búsqueda generó un total de 41 documentos, en Scopus se recuperó 31 documentos y para Isi Web of Science 10 documentos.

Tabla 2.1 Fuentes digitales y total de documentos recuperados desde revistas, conferencias, simposios, entre otros repositorios

Fuente digital	Cantidad
ACM International Conference Proceeding Series	3
Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE	1
Applied Sciences (Switzerland)	1
Artificial Intelligence Review	1
Contemporary Technologies in Education: Maximizing Student Engagement, Motivation, and Learning	1
Education and Information Technologies	1
Expert Systems with Applications	1
IEEE Transactions on Learning Technologies	1
International Journal of Advanced Computer Science and Applications	1
INTERNATIONAL JOURNAL OF CONTINUING ENGINEERING EDUCATION AND LIFE-LONGLEARNING	1
International Journal of Educational Technology in Higher Education	1
International Journal of Engineering Education	1
INTERNATIONAL JOURNAL OF INTERNET PROTOCOL TECHNOLOGY	1
International Journal of Performability Engineering	1
Internet and Higher Education	1
Journal of Information Systems Education	1
Knowledge and Information Systems	1
Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	1
Lecture Notes on Data Engineering and Communications Technologies	1
MEHRAN UNIVERSITY RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY	1
Online Learning Journal	1
Predicting Students' Academic Dropout Using Artificial Neural Networks	1
Procedia Computer Science	1
Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013	1
Proceedings - 2014 13th International Conference on Machine Learning and Applications, ICMLA 2014	1

Continúa en la siguiente página...

Fuente digital	Cantidad
Proceedings - 2018 4th International Conference on Advances in Computing, Communication and Automation, ICACCA 2018	1
Proceedings - IEEE 16th International Conference on Advanced Learning Technologies, ICALT 2016	1
Proceedings of 2016 SAI Computing Conference, SAI 2016	1
REVISTA CONRADO	1
SACI 2020 - IEEE 14th International Symposium on Applied Computational Intelligence and Informatics, Proceedings	1
SAINS MALAYSIANA	1
Studies in Systems, Decision and Control	1
TELEMATICS AND INFORMATICS	1
Turkish Online Journal of Distance Education	1
VINE	1
(en blanco)	4
Total general	41

Tabla 2.2 Documentos recuperados desde distintas fuentes digitales, aquí mostramos los trabajos aptos y no aptos para el estudio sistemático. Posteriormente, los aptos serán considerados para filtrar y medir la calidad del documento

Documento	Apto	No apto
-A Framework for Interactive Exploratory Learning Analytics	✓	
-A Relative Tolerance Relation of Rough Set in Incomplete Information	✓	
-An EDM Approach to the Analysis of Students' Engagement in Online Courses from Constructs of the Transactional Distance	✓	
-An Intelligent Approach for the Effect of Social Media on Undergraduate Students Performance: A Case Study in the University of Jordan	✓	
-An intelligent course decision assistant by mining and filtering learners' personality patterns	✓	
-Association Analysis of University Course Information Based on Knowledge Map	✓	
-Auto-generated ensemble model for predicting student success	✓	
-Centralized student performance prediction in large courses based on low-cost variables in an institutional context		✓

continúa en la siguiente página...

Documento	Apto	No apto
–Conceptual clustering of university graduate students' trajectories using formal concept analysis: a case study in Lebanon		✓
–Curriculum Analysis Based on Cerebral Hemisphere Functions Using Association Rule		✓
–Data analytics in higher education: An integrated view		✓
–DRAL: A tool for discovering relevant e-activities for learners		✓
–Early Warning System for Academic using Data Mining	✓	
–Educational data mining and learning analytics for 21st century higher-education: A review and synthesis	✓	
–Educational data mining studies in Turkey: A systematic review		✓
–Educational Data Mining Survey for Predicting Student's Academic Performance	✓	
–Effective educational process: A data-mining approach		✓
–Ensemble machine learning model for higher learning scholarship award decisions	✓	
–Guideline for Academic Support of Student Career Path Using Mining Algorithm	✓	
–Improving an early warning system to prediction of student examination achievement		✓
–Interpretable Multiview Early Warning System Adapted to Underrepresented Student Populations		✓
–Knowledge discovery from the programme for international student assessment		✓
–Learning analytics to inform the learning design: Supporting instructor's inquiry into student learning in unsupervised technology-enhanced platforms		✓
–Learning analytics: Using data-informed decision-making to improve teaching and learning		✓
–Multiple instance learning for classifying students in learning management systems	✓	
–Open interactive education algorithm based on cloud computing and bigdata		✓
–Predicting academic performance: A systematic literature review	✓	
–Predicting academic success in higher education: literature review and best practices		✓

continúa en la siguiente página...

Documento	Apto	No apto
-Predicting students' academic dropout using artificial neural networks	✓	
-Predicting students' grade based on social and content interactions		✓
-S PRAR: A novel relational association rule mining classification model applied for academic performance prediction	✓	
-Semi-supervised learning to discover the average scale of graduation of university students	✓	
-Student performance analysis and prediction in classroom learning: A review of educational data mining studies		✓
-Student performance early warning based on data mining	✓	
-The estimation of students' academic success by data mining methods		✓
-The Role of Knowledge Management and Data Mining in Improving Educational Practices and the Learning Infrastructure	✓	
-Understanding E-learners' behaviour using data mining techniques	✓	
-Unsupervised learning based mining of academic data sets for students' performance analysis	✓	
-Use of machine learning techniques for educational proposes: A decision support system for forecasting students' grades		✓
-Using Data Mining Techniques for Investigating of Performance in ICT Subject at GCE Advanced Level		✓
-Using data mining techniques to predict students at risk of poor performance		✓
Total general	16	25

Dentro del estudio sistemático se han planteado tres preguntas que permiten medir la calidad de los documentos extraídos de las bases de datos científicas, se pretende encontrar documentos que sean cercanos al trabajo que se ha propuesto. En la tabla 2.2 se detallan los documentos recuperados, tanto aptos como no aptos, ya que no todos cumplen con el objetivo trazado en el presente trabajo. Ahora bien, del conjunto de documentos aptos necesitamos medir que tan lejos o cerca están esos documentos del trabajo propuesto. Para ello, partimos de tres preguntas que simplifican la evaluación de su calidad: La primera, *¿Se han combinado con otras técnicas de inteligencia artificial?* Con esta pregunta queremos conocer si ha existido alguna fusión con otra técnica, dado que de haberse hecho sería un trabajo llamativo. La segunda, *¿El conocimiento extraído ha sido descrito?* Ciertos documentos omiten los detalles de sus resultados (conocimiento extraído) y esto ha limitado parcialmente el análisis profundo del documento. La tercera, *¿El trabajo es novedoso y*

se ha relacionado con los estudios superiores? Los diversos niveles de estudios (primaria, bachillerato, universidad) fomentan en los investigadores el hallazgo de patrones mediante el uso de técnicas de minería de datos y, para fines del estudio sistemático se han filtrado los documentos vinculados a la educación superior. De esta manera, las preguntas que se han propuesto facilitan la valoración de calidad de los trabajos aptos. En la tabla 2.3 se han detallado los documentos evaluados de acuerdo con las preguntas propuestas. La evaluación, consiste en sumar la puntuación de las respuestas de las preguntas, donde cada una cuenta con tres respuestas: Si (equivale a 1); Parcialmente (equivale a 0,5); No (equivale a 0).

Tabla 2.3 Valoración de calidad de los documentos recuperados y aptos. Los documentos que son aptos se valoran en el rango entre 3 y 1,5 la calidad del documento se ha presentado de forma descendente

Publicación	Calidad
Multiple instance learning for classifying students in learning management systems	3,0
Semi-supervised learning to discover the average scale of graduation of university students	2,5
Student performance early warning based on data mining	2,5
The Role of Knowledge Management and Data Mining in Improving Educational Practices and the Learning Infrastructure	2,5
Unsupervised learning based mining of academic data sets for students' performance analysis	2,5
Guideline for Academic Support of Student Career Path Using Mining Algorithm	2,5
SPRAR: A novel relational association rule mining classification model applied for academic performance prediction	2,5
Auto-generated ensemble model for predicting student success	2,5
Ensemble machine learning model for higher learning scholarship award decisions	2,0
Predicting students' academic dropout using artificial neural networks	2,0
Early Warning System for Academic using Data Mining	2,0
Association Analysis of University Course Information Based on Knowledge Map	2,0
Understanding E-learners' behaviour using data mining techniques	2,0
Educational data mining and learning analytics for 21st century higher education: A review and synthesis	1,5
Predicting academic performance: A systematic literature review	1,5
Educational Data Mining Survey for Predicting Student's Academic Performance	1,5

Tabla 2.4 Documentos aptos para el estudio sistemático vinculado a la minería de datos en la educación superior

Publicación	Técnica de MD	Algoritmos	Cita
–Multiple instance learning for classifying students in learning management systems	Supervisado	Basados en árboles, reglas, redes neuronales y Naïve Bayes (16 algoritmos)	(Zafra et al., 2011)
–Auto-generated ensemble model for predicting student success	Supervisado	Apriori	(Zeineddine et al., 2019)
–Semi-supervised learning to discover the average scale of graduation of university students	No supervisado	Apriori	(Guanin-Fajardo et al., 2019)
–Student performance early warning based on data mining			(Mi, 2019)
–The Role of Knowledge Management and Data Mining in Improving Educational Practices and the Learning Infrastructure	No supervisado	Apriori	(Meghji et al., 2020)
–Unsupervised learning based mining of academic data sets for students' performance analysis	Supervisado y No supervisado	SVM, PCA, Apriori	(Crivei et al., 2020)
–Guideline for Academic Support of Student Career Path Using Mining Algorithm	No supervisado	Apriori	(Sodanil et al., 2019)
–S PRAR: A novel relational association rule mining classification model applied for academic performance prediction	Review	Review	(Czibula et al., 2019)
–Ensemble machine learning model for higher learning scholarship award decisions			(Ahmad y Bakar, 2020)
–Predicting students' academic dropout using artificial neural networks	Supervisado	Neural network	(Siri, 2014)

continúa en la siguiente página...

Publicación	Técnica de MD	Algoritmos	Cita
–Early Warning System for Academic using Data Mining	Supervisado y No supervisado	Apriori	(Santoso, 2018)
–Association Analysis of University Course Information Based on Knowledge Map	No supervisado	Apriori	(Wang et al., 2018)
–Understanding E-learners'behaviour using data mining techniques	Supervisado y No supervisado	Apriori, Random Forests, Logistic Regressions and Bayesian Networks	(Al Fanah y Ansari, 2019)
–Educational data mining and learning analytics for 21st century higher education: A review and synthesis	Revisión	Revisión	(Aldowah et al., 2019)
–Predicting academic performance: A systematic literature review	Revisión	Revisión	Hellas et al. (2018)

En [Zafra et al. \(2011\)](#) se ha propuesto un enfoque basado en el aprendizaje de instancias múltiples para predecir el rendimiento de los estudiantes y mejorar los resultados obtenidos utilizando un aprendizaje clásico de instancia única. Se ha comparado la efectividad de múltiples algoritmos de clasificación, usando métricas discriminatorias tales como: la precisión, sensibilidad y sensitividad de los algoritmos. La metodología que han propuesto aumenta el rendimiento de los algoritmos del experimento. En concreto, se enfatiza en el estudio de selección de instancias que de cierto modo alcanza mejores resultados. Otro tema de interés educativo se encuentra relacionado con el éxito académico del estudiante. Así, en [Zeineddine et al. \(2019\)](#) se lo ha estudiado ya que se ha convertido en un objetivo estratégico primordial para la mayoría de las instituciones de la educación superior. Tradicionalmente, las IES se han basado en los datos recogidos en los expedientes académicos, la información de las unidades académicas, la alerta del profesorado y la interacción con los estudiantes. Con la proliferación de los grandes análisis de datos y el aprendizaje automático, las universidades confían cada vez más en los datos de los estudiantes para predecir el rendimiento del alumnado. Otro fenómeno relacionado con la educación es el abandono o deserción. Es difícil no valorar la importancia que tienen cada una de las técnicas de minería de datos para el tema de predicción, cabe destacar que los trabajos anteriormente citados se han traído al

estudio sistemático por las metodologías aplicadas en ellos y que sirven, en cierto modo, de una guía para la aplicación de metodología de estudio propuesta.

Tomaremos como punto de partida el trabajo en [Meghji et al. \(2020\)](#), donde se ha encontrado que los factores influyentes del rendimiento de los estudiantes tales como: la nota media, la asistencia y el cumplimiento de las tareas, son un punto clave para la gestión del conocimiento ya que sirve para mejorar los procesos y prácticas académicas del profesorado. De ahí que, para extraer el conocimiento ellos utilizan las reglas de asociación mediante el algoritmo Apriori. Otro enfoque muy similar al anterior se describe en [Crivei et al. \(2020\)](#), donde combinan otros algoritmos (Máquina de soporte de vectores SVM y Análisis de componentes principales PCA) y las reglas de asociación. Los autores valoran los atributos para obtener modelos de aprendizaje con técnicas no supervisadas y poder expresar patrones interesantes en los datos. Su idea se centra en proporcionar una mejor comprensión del problema de la predicción del rendimiento académico de los estudiantes. Por otra parte, en [Sodanil et al. \(2019\)](#) se han analizado las relaciones entre los cursos que posiblemente produzcan una futura carrera para los estudiantes utilizando el algoritmo Apriori. Las reglas de asociación extraídas fueron las calificaciones de los estudiantes de 25 cursos principales en el campo de la tecnología de la información. Los resultados muestran 14 reglas donde se determinaron los cursos de estudiantes que podían tener carreras futuras. Muchos de los aspectos realizados en los trabajos citados tienen relación con [Ahmad y Bakar \(2020\)](#) donde se ha presentado dos enfoques. El primero, el modelo de conjuntos “Ensemble” y el segundo que está basado en reglas. El resultado sobre el primer enfoque consigue la mejor exactitud del 86,9 % en comparación con otras técnicas. Este estudio también explora las reglas obtenidas del modelo basado en reglas J48 y Apriori que además ha logrado seleccionar las mejores reglas para desarrollar un conjunto de modelos basados en reglas que finalmente ha sido usado para la concesión de becas. Desde otra perspectiva, en [Siri \(2014\)](#) el fenómeno de la deserción afecta a todos los niveles de los sistemas educativos. La detección temprana y la profundización de las causas que lo determinan son requisitos previos necesarios para cualquier iniciativa encaminada a reducir los factores que afectan a la disminución de las tasas de fracasos formativos y de abandono en todos los niveles. El modelo de análisis, actualmente experimentado en la realidad de la Universidad de Génova, se propone como un instrumento simple y flexible para apoyar las actividades de seguimiento y evaluación de todos los sistemas educativos. Por otra parte, en [Mi \(2019\)](#) se ha intentado desarrollar un sistema de alerta temprana. Esto, referido con el rendimiento de los estudiantes que por lo general está relacionado con muchos factores complicados y llenos de incertidumbre, por lo que la alerta temprana que de por sí mismo se convierte en un desafío. Los resultados preliminares demostraron que se podía identificar oportunamente a los estudiantes en situación de riesgo y

mejorar la eficiencia y eficacia general de la gestión educativa en materia de alerta temprana. En esta misma línea, en ([Alharbi et al., 2016](#)) se examina si la extracción de datos puede utilizarse para poner de relieve los problemas de rendimiento en una fase temprana y proponer medidas correctivas. Además, algunos de los métodos también pueden servir de base para recomendar sistemas que puedan orientar a los estudiantes hacia sus opciones de módulos para aumentar sus posibilidades de obtener buenos resultados.

De todo cuanto se lleva dicho de los trabajos científicos de estudiantes universitarios se asocian por lo general con: factores predominantes, rendimiento, deserción, etc., despiertan interés en los gestores y estrategas académicos. Es oportuno aprovechar el compendio de información relevante para recomendar estrategias académicas. Por ello, observamos que en ([Aldowah et al. \(2019\)](#)) los autores enfatizan que la técnica de minería de datos no supervisada, concretamente, las reglas de asociación se emplean para planificar estrategias y tomar decisiones sobre la forma de cómo mejorar la calidad de las tasas de éxito y fracaso de los estudiantes. Así, mediante las reglas de asociaciones han identificado las asociaciones entre atributos. Dado que, el análisis estadístico tradicional tal vez no proporcione suficientes conocimientos para la creación normativas. Por lo tanto, las reglas de asociación se utilizan para identificar las relaciones entre los comportamientos de los estudiantes y la disparidad de rendimiento. Otro tema de interés, donde se ha utilizado las reglas de asociación para extraer conocimiento de los cursos impartidos está en ([Wang et al. \(2018\)](#)), donde se adoptan las reglas de asociación para analizar las variables educativas, como las calificaciones de los estudiantes y la información sobre la selección de los planes de estudios. Al implementar el algoritmo Apriori, se extraen varias reglas de asociación significativas. Ellos, definen que hay un cierto grado de correlación oculto en las puntuaciones de los cursos obligatorios y optativos. El análisis proporcionado son una base fundamental para la construcción del mapa de nuevos cursos públicos y la innovación del método de enseñanza. Con otra perspectiva, en ([Romero et al.](#)) han enriquecido los datos disponibles utilizando la información de los estudiantes que hacían pruebas particulares. Su enfoque se centró en la aplicación de un algoritmo de minería de reglas de la asociación de clases a tres matrices diferentes: una matriz de puntuación, una matriz de relación y una matriz de conocimiento. Estas matrices se construyeron en base a los datos de los resultados de los estudiantes en sus pruebas, y en el conocimiento del dominio proporcionado por un pedagogo.

En ([Czibula et al. \(2019\)](#)) analizan el problema de la predicción del rendimiento académico de los estudiantes, un tema que se investiga cada vez más en la literatura de la Minería de Datos Educativos. Se ha propuesto un nuevo modelo de clasificación, S PRAR (Student Performance prediction using Relational Association Rules) para predecir el resultado final de un estudiante de un grado académico utilizando las reglas de asociación relacional

para identificar las diversas relaciones entre atributos. En [Guanin-Fajardo et al. \(2019\)](#) han propuesto un enfoque de trabajo semi-supervisado usando el algoritmo Apriori. En el trabajo los autores estudian de forma exploratoria las proximidades de las variables mediante el coeficiente de correlación de Pearson. Filtran las características del conjunto de datos a través de un algoritmo de selección de características. Después, filtran las instancias usando un algoritmo de selección de instancias. Finalmente, el conjunto de datos es convertido en transacciones que son usadas para obtener las reglas de asociación y posteriormente inducen el consecuente para conseguir el tiempo promedio de graduación del estudiante. Los hallazgos sugieren que existe empatía en los grupos de edades del profesorado y el alumnado entre segundo y tercer año, además, de la nota media entre cursos y la empleabilidad del alumnado, que vienen siendo los factores esenciales en las tasas de graduación del alumnado. Por otra parte, la conversión de datos en bruto a conocimiento y que esta sea usada para la toma de decisiones es propuesta en [Santoso \(2018\)](#) que utiliza un conjunto de tres modelos. Los dos primeros modelos utilizaron los datos obtenidos en el primer año (primer semestre y segundo semestre), para predecir el éxito académico de los estudiantes matriculados, mientras que el tercer modelo utilizó la información disponible al final del primer año para predecir el rendimiento académico de los estudiantes al final de sus estudios, identificando así los factores que son más críticos para estos modelos. Los resultados identificaron por medio de los patrones obtenidos a los estudiantes que necesitaban más apoyo pedagógico, así como a los estudiantes con alta probabilidad de sobresalir en sus estudios. De acuerdo con el trabajo de [Nandeshwar et al. \(2011\)](#), en la literatura estudiada, los autores no han encontrado ningún estudio con una mejora tan significativa sobre la línea de base para la retención de tercer año. De modo que, el rediseño de políticas para mejorar la tasa de retención de tercer año, no sólo mejorarán las tasas de retención de primer y segundo año, sino también las tasas de graduación de seis años. Para la institución estudiada, los antecedentes familiares y la situación socio-económica de la familia son fundamentales para la persistencia del estudiante en el tercer año.

Como consecuencia de las diversas investigaciones educativas, otros autores han elaborado revisiones sistemáticas o estado del arte respecto a la minería de datos en la educación. Por ello, en [Hellas et al. \(2018\)](#) se ha estudiado la capacidad de predecir el rendimiento de los estudiantes en un curso o programa crea oportunidades para mejorar los resultados educativos es muy oportuno. Además, enfatizan que la predicción del rendimiento de los estudiantes procura determinar las características interrelacionadas e identificar las razones subyacentes por las que ciertas características funcionan mejor que otras en distintos algoritmos. Al mismo tiempo, la revisión realizada ha revelado una serie de cuestiones asociadas con la calidad de las investigaciones, que impulsa en la necesidad de proporcionar un informe más detallado de

Tabla 2.5 Tópicos del estudio sistemático mediante la búsqueda de patrones o combinaciones con otras técnicas de minería de datos

Tópico	Cita
Alerta temprana	(Alharbi et al., 2016; Mi, 2019)
Predicción de rendimiento	(Aldowah et al., 2019; Crivei et al., 2020; Meghji et al., 2020; Romero et al.; Wang et al., 2018)
Predicción éxito fracaso, grados, cursos	(Aldowah et al., 2019; Guanin-Fajardo et al., 2019)
Patrón de estudiantes titulados	(Czibula et al., 2019; Guanin-Fajardo et al., 2019; Nandeshwar et al., 2011; Santoso, 2018)
Búsqueda de patrones	(Ahmad y Bakar, 2020; Crivei et al., 2020; Meghji et al., 2020; Siri, 2014; Sodanil et al., 2019)

los métodos y resultados para aumentar los esfuerzos de validar y replicar su trabajo. Por otra parte, en Bonde y Kirange (2020) convergen en la elaboración de un modelo para calcular el comportamiento y nivel del progreso académico del estudiantado utilizando diversas técnicas de minería de datos entre ellas las reglas de asociación. Ellos examinan detenidamente que factores afectan de forma positiva o negativa al modelo. Además, encuentran brechas con diferencias significativas entre las aptitudes académicas y empresariales.

2.6.1. Estadística del estudio sistemático

Con el incentivo de la extracción de documentos de la base de datos digitales desde distintas fuentes (revistas, conferencias, libros, etc.) y su posterior condensación de ajuste para el propósito del trabajo, ahora mostramos una breve estadística de la recopilación de los documentos.

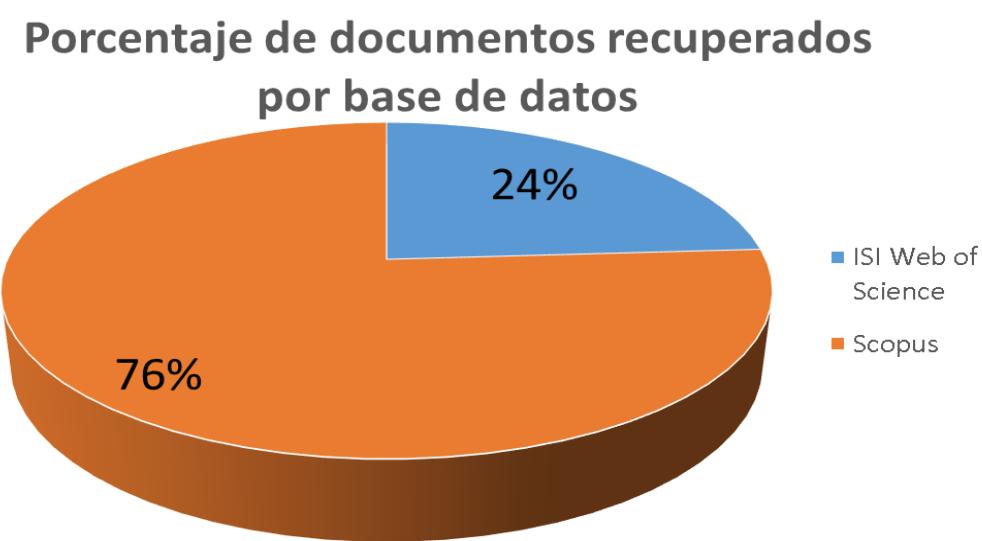


Figura 2.5 Distribución de porcentajes de trabajos recuperados de las bases de datos Scopus e ISI Web of science. EL mayor número de trabajos se ha recuperado de Scopus 76 %, mientras que en Isi Web of Science 24 %. Tanto los documentos de conferencias, revistas, capítulos de libros, entre otras fuentes han sido consultadas para el estudio sistemático.

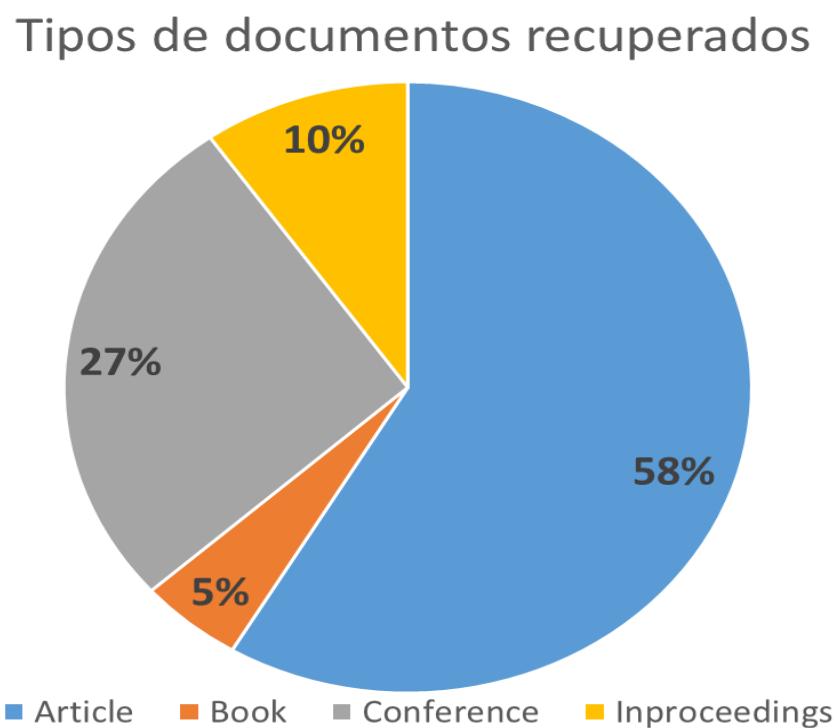


Figura 2.6 El porcentaje de distribución de tipos de documentos recuperados son un 58 % artículos, el 27 % conferencias, 10% inproceedings y un 5 % libros(capítulos). Es evidente que la mayoría de los documentos están relacionados con artículos de revistas. Para el estudio sistemático de forma general se estudiará un 85 % de la documentación de

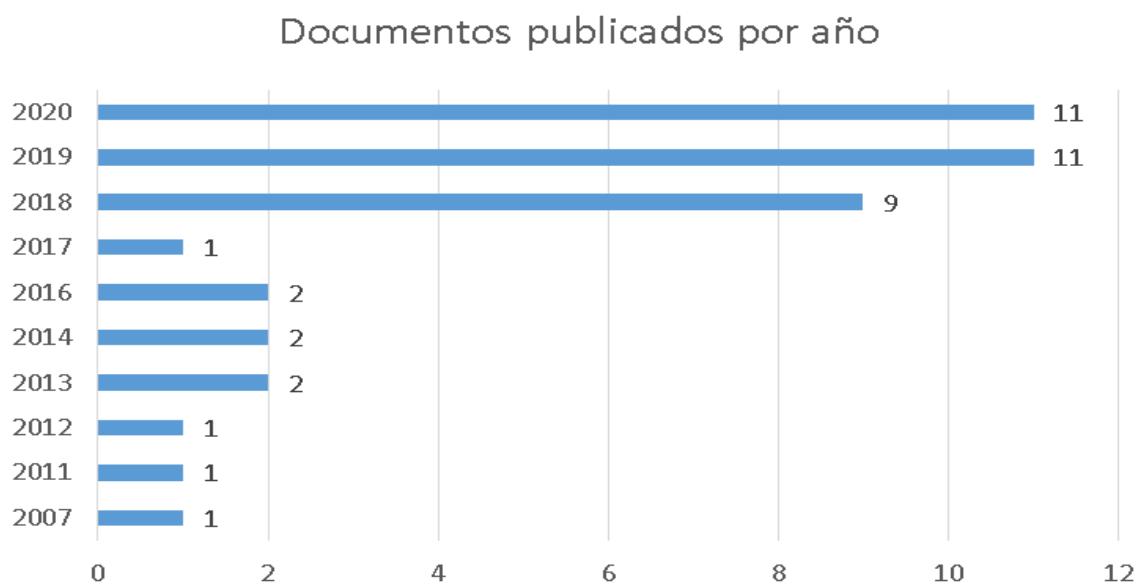


Figura 2.7 Número de documentos publicados por año vinculados al estudio de las reglas de asociación y estudiantes universitarios titulados

Capítulo 3

Análisis exploratorio de datos

El análisis de datos en general está compuesto de varias etapas que convergen en la obtención de un modelo descriptivo o predictivo. Previo al estudio profundo de los datos es importante disponer de la distribución, comportamiento y tendencia de estos, es decir, aplicar el análisis exploratorio para conseguir contrastar los datos. En este capítulo, nos enfocamos en dos métodos exploratorios: la estadística descriptiva y la visualización de datos usando en esta última dos técnicas: (i) agrupamiento de datos y, (ii) la representación de los datos a través de una red de nodos.

3.1. Estudio descriptivo

A fin de obtener información preliminar de los datos se ha efectuado el análisis descriptivo de los datos, para ello existen varios paquetes/programas estadísticos, hojas de cálculos, procesadores numéricos, entre otras aplicaciones informáticas que permiten obtener información estadística de los datos. El manual de [Charte \(2014\)](#) ha planteado el uso de funciones que facilitan el análisis exploratorio de datos usando la estadística y la visualización como medios para alcanzarlo. En cierto modo, el estudio descriptivo aporta información de interés ya que de primera mano se obtienen resultados de las centralidades y dispersión de variables que sirven como un punto de partida para el análisis (Véase tabla ?? , página 80). La tabla ??, toma como referencia las medidas de estadística central para el estudio primario de los datos, en la columna n se aprecian variaciones de cantidades esto debido a la ausencia de datos. Como parte de la solución de este problema, se ha aplicado la técnica de remplazo (imputación) de valores usando métodos estadísticos como la *media* para variable numérica y la *moda* para la variable de tipo categórica. Ya en la sección 4.1.1, se ha realizado el estudio profundo de los valores perdidos (*missing values*).

3.2. Visualización interactiva de datos

Es indudable el impacto que genera la proyección de los datos por medio de gráficos ya que permite tener una percepción real de la información. En consonancia con el estudio descriptivo, ahora se pretende estudiar la información desde una perspectiva visual e interactiva. Para ello, es posible el uso de herramientas informáticas con distribución comercial o libre que se encuentran disponible en Internet ([Data-Driven Documents](#), [Data Visualization Catalogue](#), [Fine Report](#), [Plotly](#), [Tableau](#), [Vega](#)). Para fines de este análisis se ha utilizado el programa *Gephi* (código libre) ([Bastian et al., 2009](#)). Donde la iteración de los datos empieza construyendo una red de nodos. Luego, se usan varias medidas para valorar y obtener una red de nodos ajustada a criterios de búsqueda del investigador.

Vamos a partir del análisis de la red social académica, es decir, la interacción entre el estudiantado y profesorado, esto con el fin de conocer la dinámica de las variables. Por ello, se han usado medidas relacionadas con el análisis de redes sociales. Por medio del programa, se ha generado un grafo no dirigido con 47 nodos y 442 aristas. Literalmente, es un grafo pequeño (computacionalmente), ya que el programa Gephi soporta ciento de miles de nodos y está limitado únicamente por el hardware que se utilice. Entre tanto, para el trabajo actual nos enfocamos en tres aspectos de la red: diámetro, densidad y modularidad de los datos.

3.2.1. Diámetro

El diámetro del grafo busca medir la equidistancia de los nodos, esto permite saber cuan grande es el grafo. En [Brandes \(2001\)](#) se ha propuesto un algoritmo para el cálculo de medidas de una red, cuatro de éstas se han calculado basados en el conjunto de datos que estudiamos, y que a continuación lo describimos.

Distancia media del grafo La distancia media de grafo entre todos los pares de nodos. Los nodos conectados tienen distancia 1. El diámetro es la distancia de grafo más larga entre dos nodos cualquiera de la red (es decir, cómo de lejos están los dos nodos más alejados). La distancia del grafo media es de 1,8454. El diámetro del grafo es 3.

Intermediación Mide la frecuencia con la que un nodo aparece en el camino más corto entre nodos del grafo.

Cercanía La distancia media desde un nodo inicial hacia todos los nodos demás nodos del grafo.

Excentricidad La distancia de un nodo al nodo más alejado de la red.

3.2.2. Densidad

Otra de las medidas que se han usado para el grafo es la densidad, esta mide cómo de cerca está el grafo de ser completo. Un grafo completo tiene todas las aristas completas y una densidad igual a 1. La densidad de nuestra red es de 0,101.

3.2.3. Modularidad

Es difícil, no valorar lo importante que resulta el cálculo de ésta medida que se basa en la optimización de la modularidad empleando un método simple para extraer la estructura comunitaria de la red ([Blondel et al., 2008](#); [Lambiotte et al., 2014](#)). En concreto, la modularidad de la red es de 0,063 y se han detectado tres comunidades.

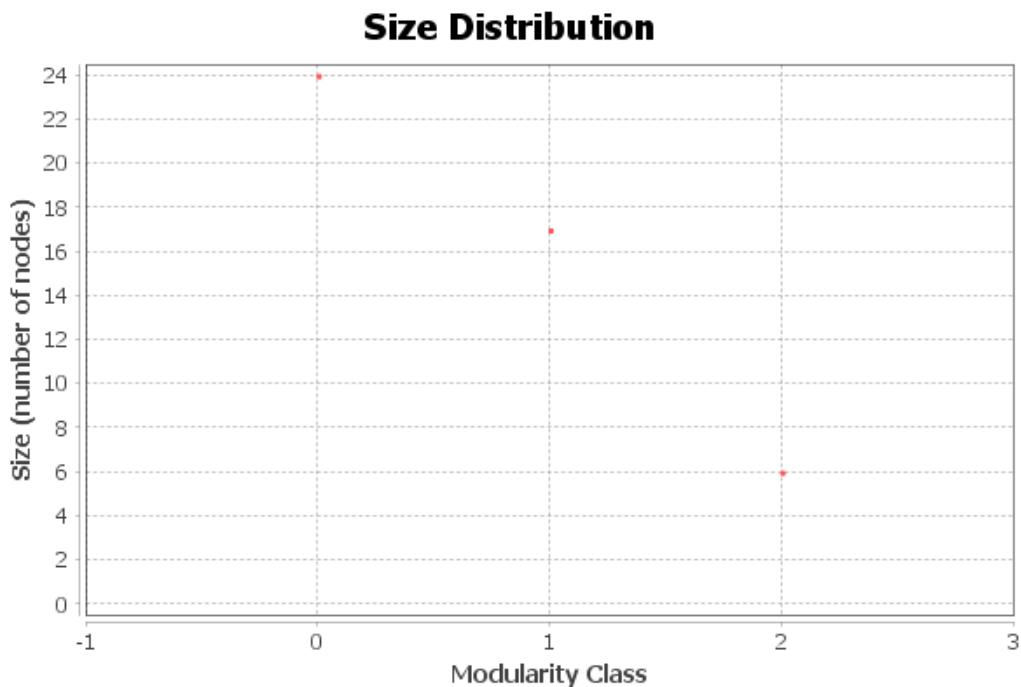


Figura 3.1 En el eje de las ordenadas está definido el número de nodos por comunidad, mientras que, en el eje de las abscisas la cantidad de comunidades. De ahí que la proporción comunitaria de la red esté segmentada así: la primera comunidad (0) representa el 51,06%; la segunda (1) el 36,17 %, y la tercera(2) 12,77 %

3.3. Agrupamiento

Otro ámbito de estudio para el descubrimiento de datos relevantes es emplear técnicas no supervisadas de minería de datos que se basan en el agrupamiento de datos. Esto consiste en

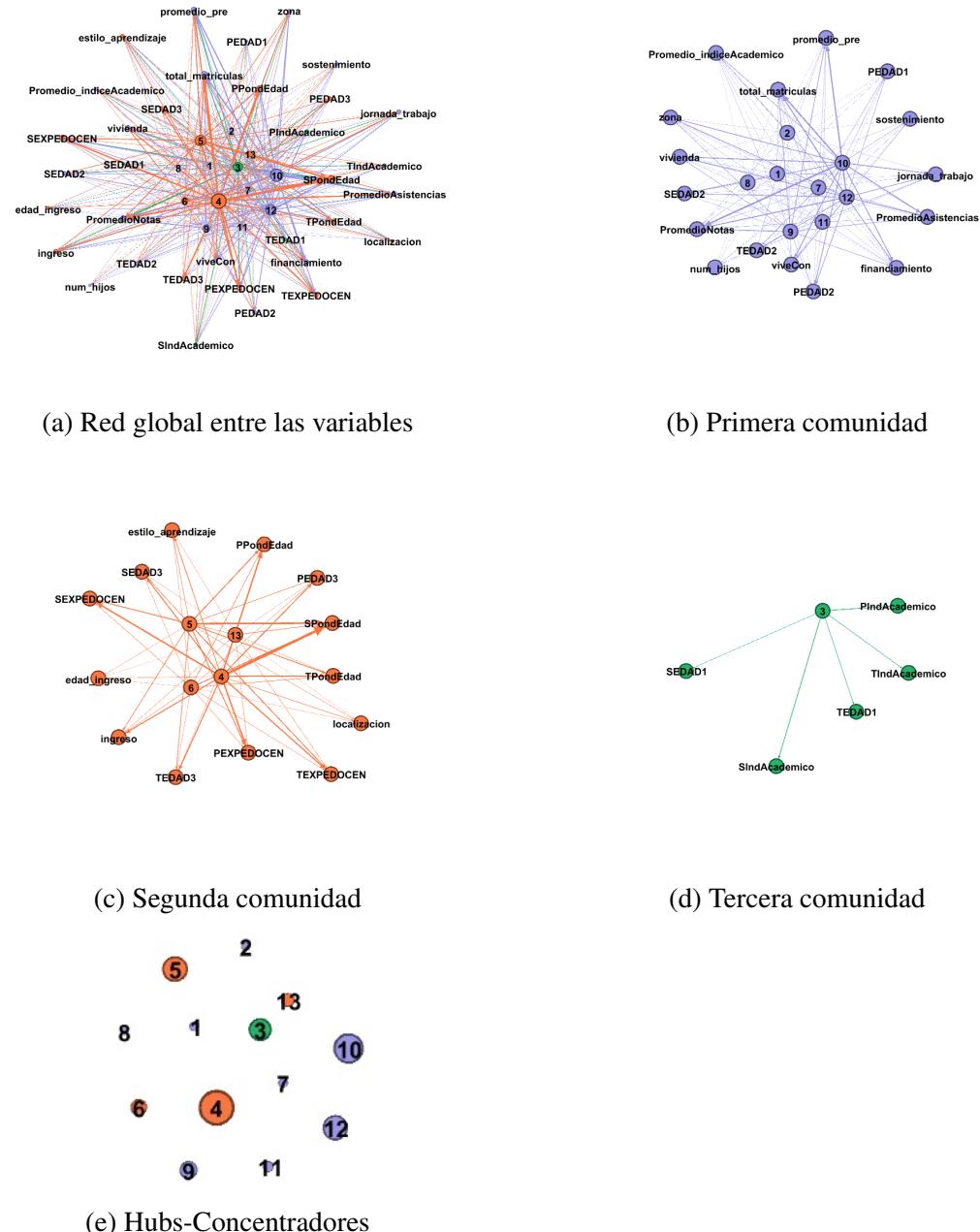


Figura 3.2 La figura (a) red de nodos completa visualiza la red con las 3 comunidades detectadas, cada tonalidad de color del nodo y arista se ha descubierto como una comunidad. La anchura de la arista está referida al peso, mientras que, el tamaño del nodo se ha basado en la intermediación; la primera comunidad detectada (b) se ha encontrado 8 nodos centrales (numerados) que se vinculan con el grado universitario; (c) la segunda comunidad detectada agrupa 4 nodos centrales (numerados) del grado universitario; (d) la tercera comunidad detectada tiene 1 nodo central. Los nodos centrales (e) son los hubs detectados en la red.

clasificar registros del conjunto de datos que presenten cierto grado de homogeneidad. De hecho, a ésta técnica también se la conoce como “clúster” (Clustering, del idioma inglés). Para fines de este trabajo se han usado tres tipos de agrupamiento: k-medias, particionamiento, y agrupación jerárquica. Previo a la aplicación de ésta técnica es necesario que el conjunto de datos sea transformado en una matriz de proximidad. Es decir, que por medio de una función se obtenga una medida de distancia o similitud entre las variables. Así, el agrupamiento es homogéneo a partir de la cantidad de grupos considerados, en nuestro caso se ha definido 2, 3, 4 y 5 grupos.

3.3.1. K-means

k-Means se basa en la idea de que los puntos deben asignarse a los clusters que minimizan la distancia global entre los puntos y las medias/centroídes de los clústers a los que han sido asignados. La medida de distancia típica que se utiliza es la euclidiana (aunque teóricamente se pueden utilizar otras distancias), por lo que el método sólo suele aplicarse a datos continuos. k-Means supone que la variabilidad de todos los clusters en todas las variables es la misma, lo que da lugar al descubrimiento de clusters esféricos que ocupan cada uno un volumen igual en el espacio de las variables ([Flynt y Dean, 2016](#)). En otras palabras, la idea principal de ésta técnica es distribuir los datos en un espacio dimensional calculando el valor medio de los grupos de datos ([Espinosa, 2019](#)). A continuación, hemos generado grupos de datos usando tres medidas de distancias distintas: Euclidiana, Pearson y Bray-Curtis.

Euclidean

La distancia euclidiana entre dos puntos es simplemente la longitud del vector que conecta los dos puntos (Véase ecuación [2.1](#), página [10](#)).

Pearson

Con el método de Pearson, podemos definir el coeficiente de correlación como un índice que puede utilizarse para medir el grado de relación entre dos variables siempre y cuando ambas sean cuantitativas y continuas.

Bray–Curtis

La distancia Bray-Curtis tiende a resultar más intuitiva debido a que los datos comunes y raros tienen pesos relativamente similares, mientras que la distancia euclidiana depende en mayor medida de los datos más abundantes. Esto sucede porque las distancias euclidianas se

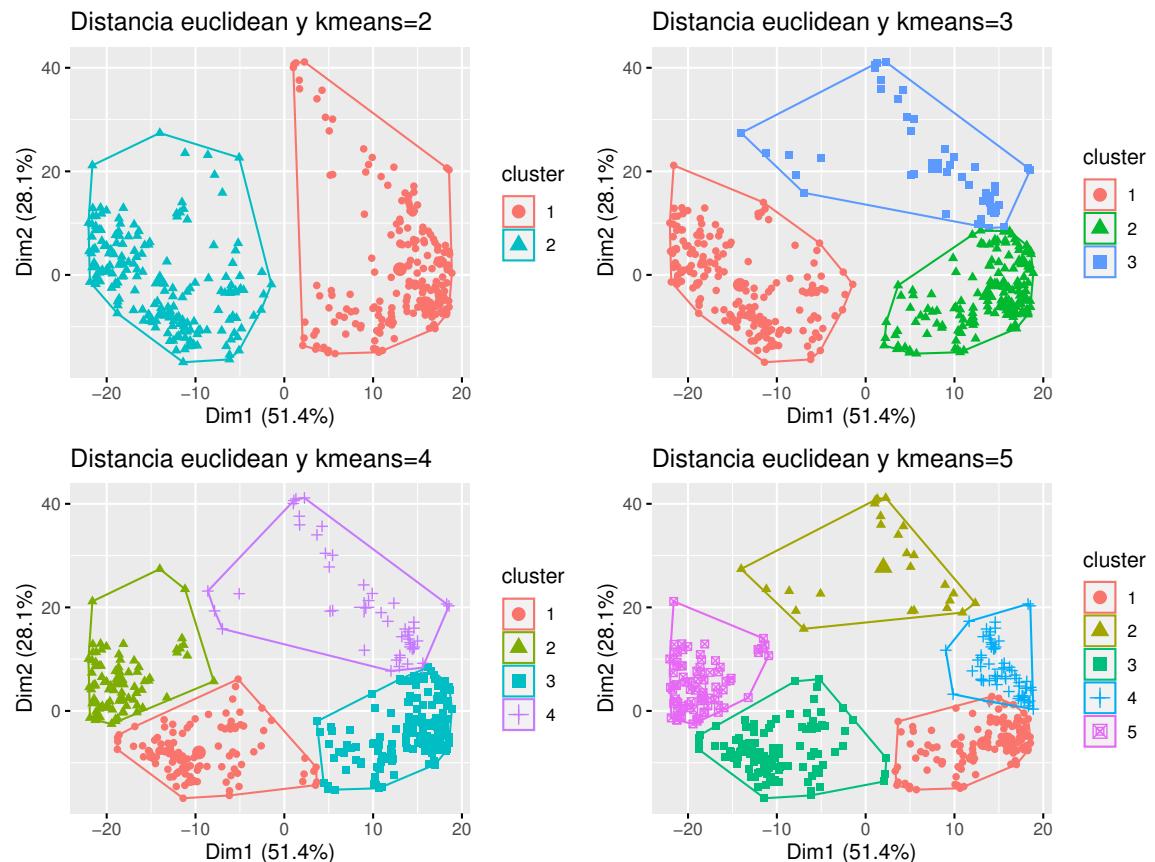


Figura 3.3 En esta figura apreciamos cuatro subgráficos, cada uno con el número de grupos que se han formado utilizando la distancia euclidiana.

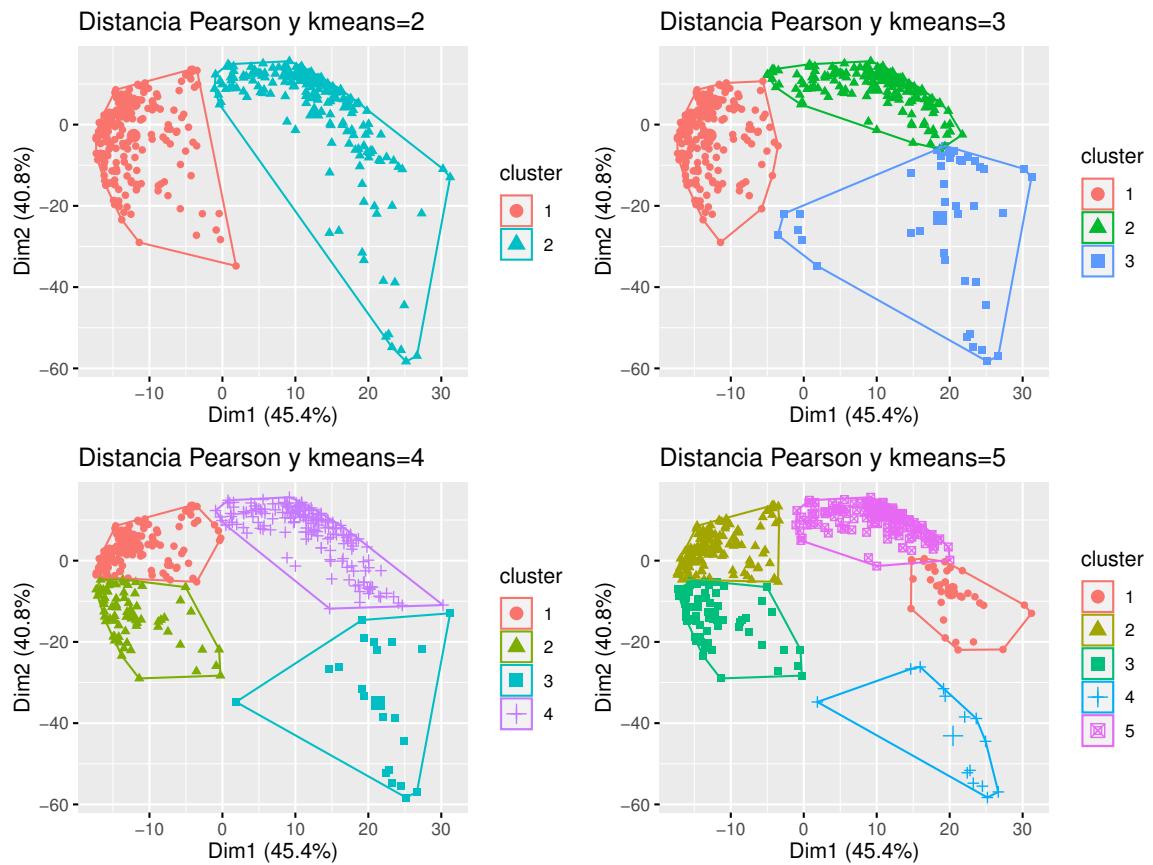


Figura 3.4 Cluster usando como distancia la medida de Pearson

basan en diferencias al cuadrado, mientras que Bray-Curtis utiliza diferencias absolutas. El elevar un número al cuadrado siempre amplifica la importancia de los valores más grandes (Espinosa, 2019).

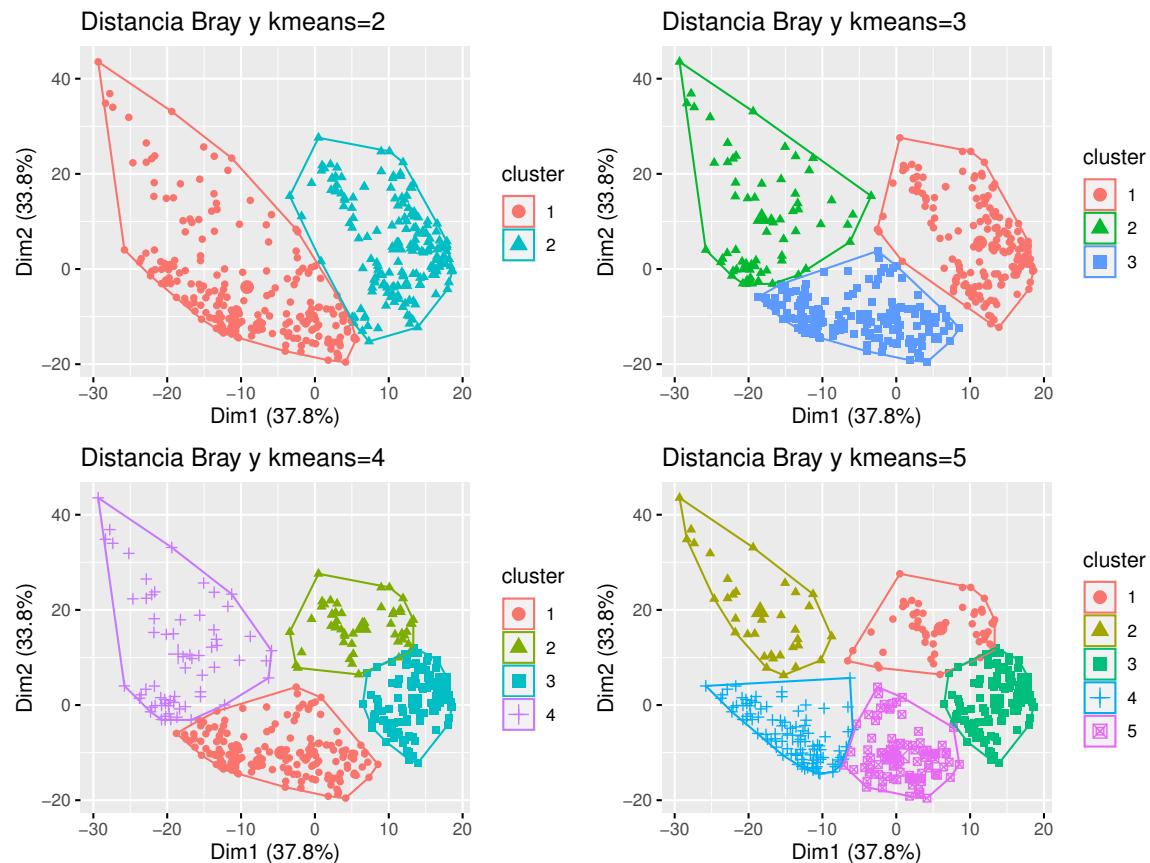


Figura 3.5 Agrupamiento de datos usando el algoritmo de disimilitud de Bray

3.3.2. Particionamiento (K-medoides)

Por otra parte, se ha utilizado el algoritmo de particionamiento de medoides (Schubert y Rousseeuw, 2018), que es una adaptación del algoritmo k-means. En lugar de calcular la media de los elementos de cada grupo, se elige un elemento representativo, o medoide, para cada grupo en cada iteración. Los medoides de cada grupo se calculan encontrando el objeto i dentro de el grupo que minimiza. Concretamente, se ha usado el algoritmo PAM que realiza una búsqueda codiciosa que puede no encontrar la solución óptima global. Los medoides son más robustos a los valores atípicos que los centroides, pero necesitan más cómputo para datos de alta dimensión.

$$\sum_{j \in C_i} (i, j) \quad (3.1)$$

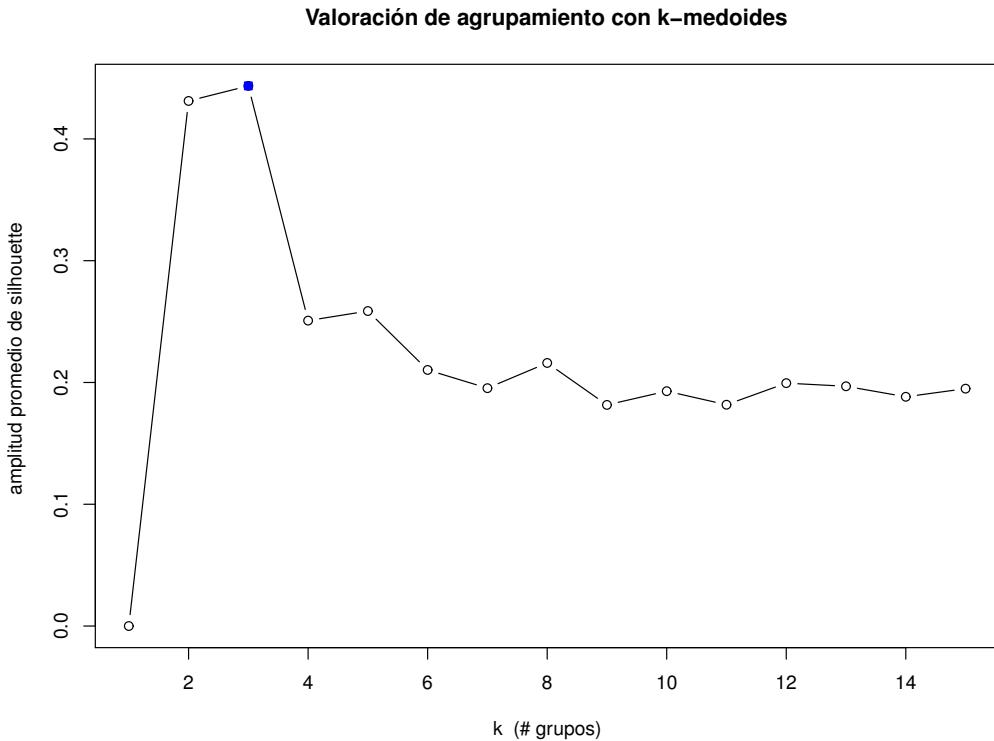


Figura 3.6 Valoración promedio del agrupamiento respecto al número de grupos. Se han realizado pruebas entre dos y quince grupos. La mejor valoración se la ha conseguido cuando de los datos se agrupan en tres partes. La línea del gráfico marca el recorrido y valoración en cada grupo, el mejor resultado es marcado con un cuadrado, el resto con un círculo.

3.3.3. Agrupamiento jerárquico

El algoritmo de agrupamiento jerárquico (aglomerativo) se basa en la unión entre los dos clusters más cercanos. La inicialización del agrupamiento establece cada ejemplo como un clúster. Tras unas cuantas iteraciones de unión de clusters se obtienen los clusters finales. En el clustering particional los ejemplos se dividen en clusters ([Top](#)).

A la luz del análisis exploratorio de datos, que ha permitido visualizar la manera en que los datos fueron distribuidos en un espacio dimensional. No es difícil percibir las ambigüedades de los resultados. Desde luego, todos han coincidido en el número de grupo que pueden formarse. El primer análisis, visualización interactiva de datos, que por medio de la red de nodos ha detectado tres comunidades distintas en los datos, estudiando el diámetro (distancia media del grafo, intermediación, cercanía y excentricidad), densidad y modularidad de los

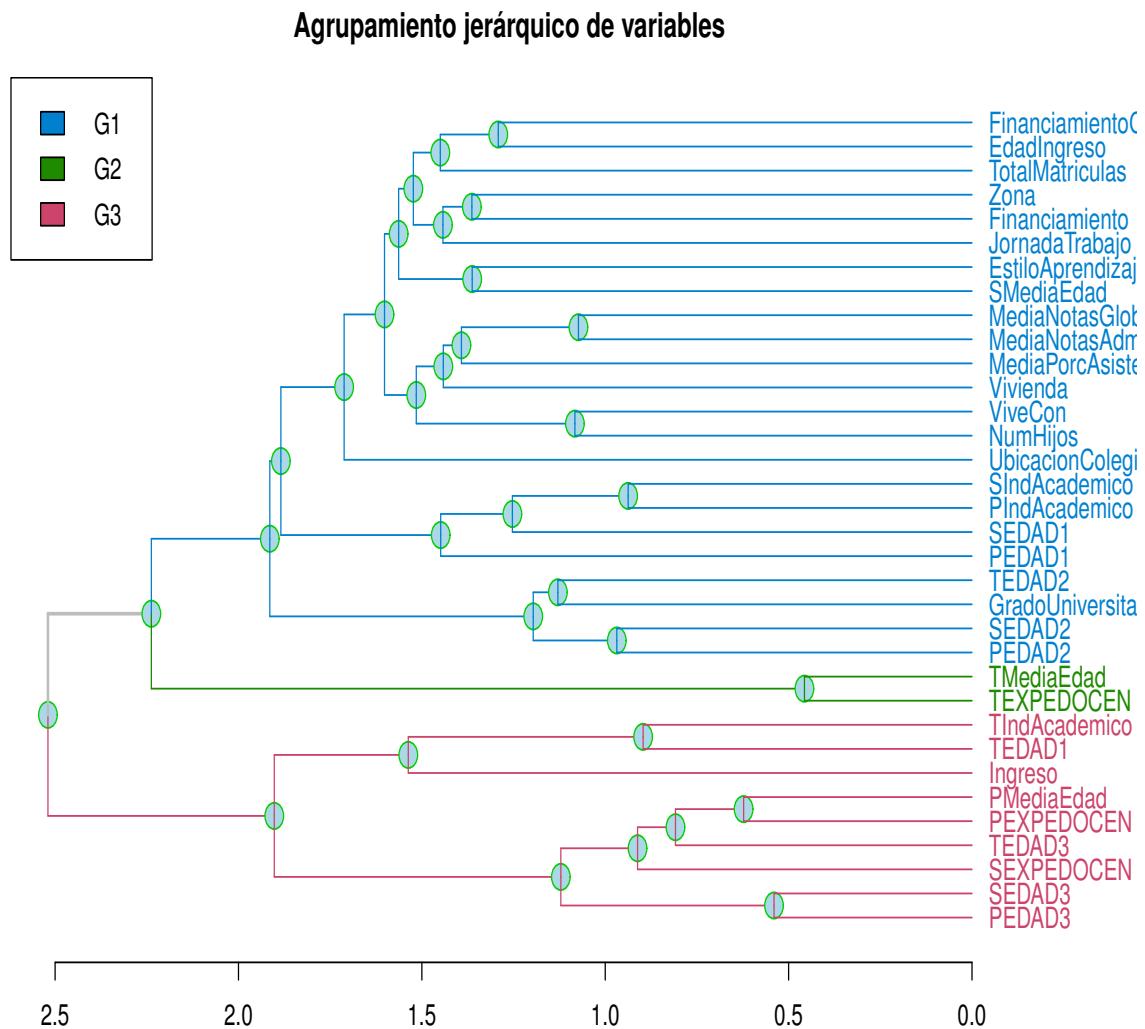


Figura 3.7 Dendrograma de variables de estudiantes graduados, para el cálculo del dendrograma se ha utilizado el método de Pearson, previamente las variables han sido escaladas. Podemos apreciar claramente que un grupo significativo de los datos se ha centrado en las variables TPondEdad y TEXPEDOCEN. Esto significa que estas dos variables influyen en el proceso de graduación del estudiantado

datos. El segundo, agrupamiento, *k-medias* que usando la función de distancia euclíadiana por un lado, y, por otro, la función de distancia de Bray-Curtis han coincidido en la formación de tres grupos de datos a partir del centroide. De hecho, ambas funciones logran separar mejor los grupos de datos; por otra parte, el particionamiento empleando *K-medoides* también detectó tres grupo de datos, basados en este resultado, se ha conseguido graficar el agrupamiento jerárquico considerando tres grupos de datos (Figura 3.7). De manera general, tenemos razones para creer que el análisis exploratorio de datos, ha permitido a través de los diferentes métodos usados que los datos están mejor representados cuando están formados en tres grupos.

Capítulo 4

Metodología

El descubrimiento del conocimiento en grandes base de datos (Knowledge Discovery Database-KDD) planteado por Fayyad sugiere a la minería de datos como una etapa más del proceso KDD. Aquí, aplicamos cada una de las etapas hasta obtener patrones que son evaluados/interpretados hasta conseguir resultados significativos que son traducidos como conocimiento. Concretamente, en la etapa de minería de datos se han aplicado algoritmos no supervisados con distintos enfoques, entre ellos: semi-supervisado, difuso y genético difuso. El propósito general, es revisar varios enfoques para obtener patrones significativos que detecten de forma oportuna las reglas que influyen en la escala de tiempo de graduación del alumnado.

4.1. Preprocesamiento de los datos

En esta etapa, la relevancia y calidad de los datos es crucial para conseguir buenos resultados. Por ello, el preprocesado de los datos se ha centrado tanto en el remplazo de datos faltantes (*missing values*), selección de instancias (*instance selection*) y selección de características (*feature selection*). Todo esto, conseguirá un patrón más simple, legible y preciso. Los datos obtenidos para este estudio se han recuperado de la base de datos institucional del centro universitario. El conjunto de datos está compuesto de variables numéricas y categóricas, en síntesis, 53 variables y 407 observaciones de estudiantes graduados en diferentes grados universitarios (Véase apéndice 5, página 77).

4.1.1. Valores perdidos

Los problemas reales no siempre gozan de integridad en los datos, el proceso de recolección trae consigo ausencia e incoherencia en sus valores, sea por intervención humana,

errores de programas informáticos o por fallo de algún dispositivo electrónico, por tanto, ante este inconveniente es habitual eliminar o remplazar valores según el tipo de dato utilizando las medidas estadísticas de tendencia central como la mediana, media aritmética (variable cuantitativa) o moda (variable cualitativa). Sin embargo, ésta técnica aplicada en minería de datos es poco recomendable para conseguir exactitud de un modelo predictivo o descriptivo. Ahora bien, paralelamente, al remplazo o eliminación de datos mencionado anteriormente existe una serie de algoritmos que facilitan el remplazo valores perdidos y, que son comúnmente usados, entre ellos están: K-Nearest Neighbor Imputation ([Batista y Monard, 2003](#)), K-means Imputation ([Deogun et al., 2004](#)), Weighted K-Nearest Neighbor Imputation ([Troyanskaya et al., 2001](#)), Most Common Attribute Value ([Grzymala-Busse et al., 2005](#)), entre otros. El conjunto de datos estudiado tuvo el 2 % de valores perdidos presentado en [4.1](#).

Para conseguir un grupo de datos íntegros, se ha imputado los datos usando el algoritmo de [Batista y Monard \(2003\)](#) dado que este algoritmo ha usado el vecino más cercano como método de remplazo. La principal ventaja de este método es que el tratamiento de los datos que faltan es independiente del algoritmo de aprendizaje utilizado.

4.1.2. Selección de instancias

El proceso de selección de instancias es fundamental en el preprocesamiento de datos. Comúnmente los datos que se encuentren fuera de los quintiles estadísticos de una variable son considerados como anómalos. Sin embargo, no siempre se puede considerar anómalos a esos datos. Por tanto, para el proceso de la selección de instancias se ha procedido a usar el algoritmo NoiseFilter [Morales et al. \(2016\)](#), que elimina o repara las instancias.

Tabla 4.1 Proporción de eliminación de instancias con ruido. Se ha considerado cinco tipos de iteraciones donde cada una de ellas ha suprimido el 5 % de datos anómalos (ruido).

# iter	Tiempo de graduación		
	Bajo	Medio	Alto
15	47,62	28.57	23,81
50	57,14	42.86	0,00
100	57,14	42.86	0,00
500	57,14	38.10	4,76
1000	66,67	28.57	4,76

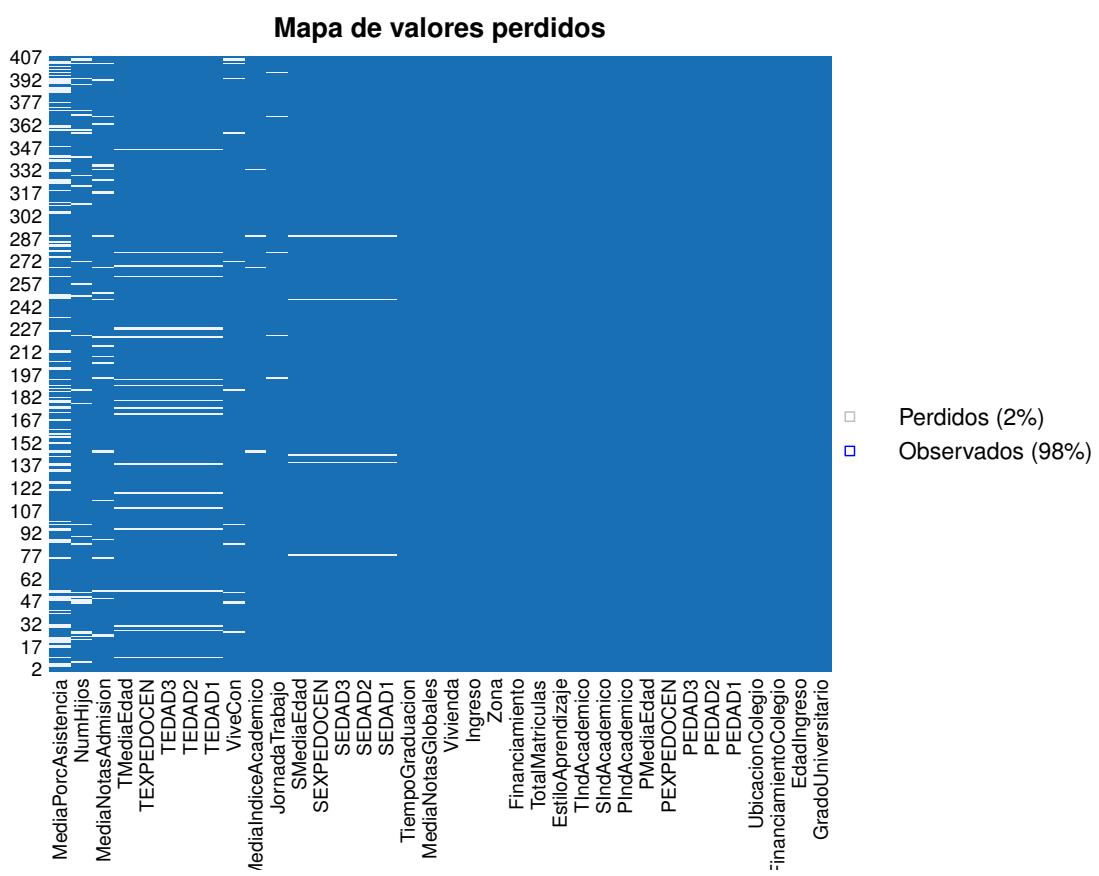


Figura 4.1 Los valores observados se encuentran marcados con fondo azul, mientras que los valores perdidos se han marcado de color blanco.

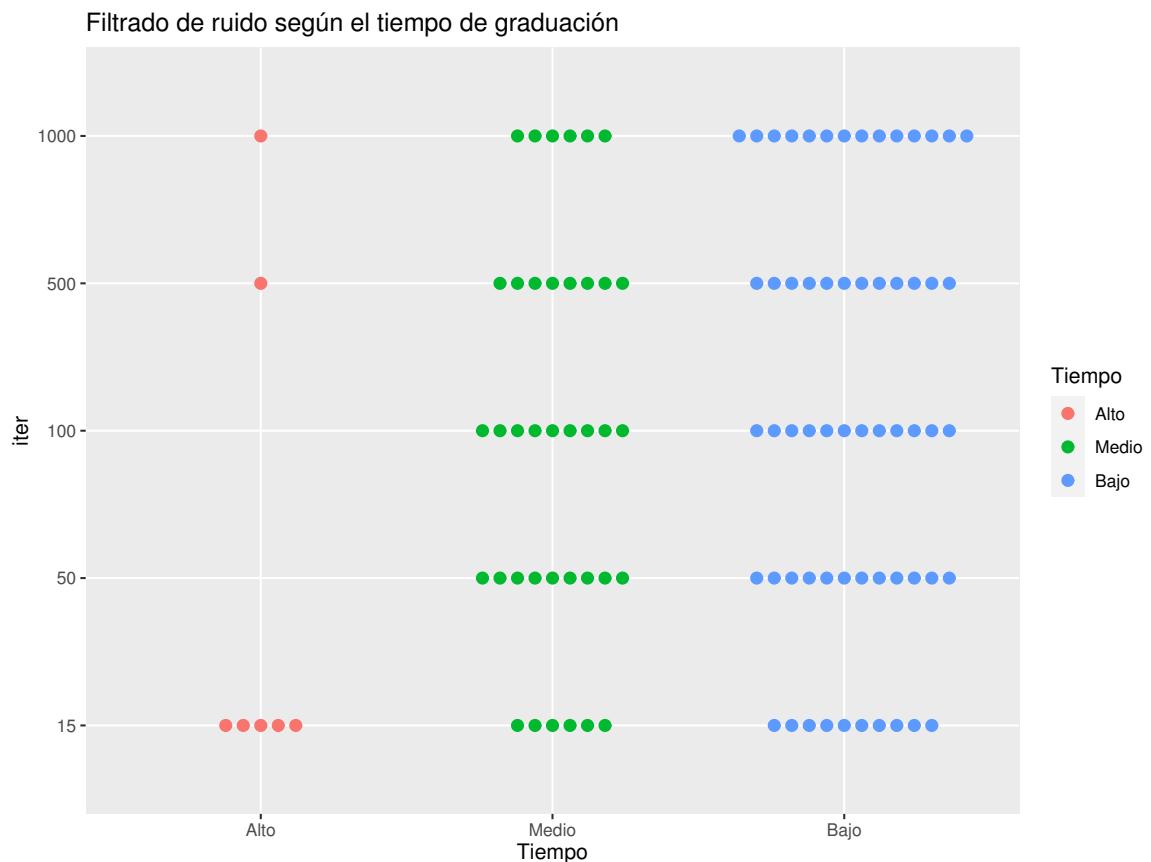


Figura 4.2 La figura muestra los tiempos de graduación del estudiantado. El algoritmo de filtrado de instancias ha usado 15, 50, 100, 500, 1000 iteraciones para la eliminación de instancias con ruido. La iteración 15 es la que filtra de forma menos desigual.

4.1.3. Selección de características

Otra etapa importante, dentro del preprocesado de datos es la selección de características. El nivel apropiado de abstracción de características juega un papel clave en el sesgo de adaptación, y una de las ventajas de usar un pequeño pero rico resumen de variables ha sido permitir una mejor generalización del patrón de búsqueda ([François-Lavet et al., 2018](#)). El estudio de selección de características se fundamenta de una serie trabajos (Véase ([Laura Emmanuella y De Paula Canuto, 2014](#); [Li et al., 2017](#); [Mursalin et al., 2017](#); [Stańczyk, 2015](#); [Stańczyk y Jain, 2015](#))) que han conseguido reducir las características (variables) de un determinado conjunto de datos. Por ejemplo, el vecino más cercano de conjunto editado (KNN) ([Wilson, 1972](#)), *las vegas filter* (LVF) ([Liu y Setiono, 1996](#)), *Relief-F*, wrapper, ovo, ova, entre otros. El algoritmo *Relief-F* ([Kononenko, 1994](#)), por ejemplo, estima la calidad de atributos con o sin dependencia entre ellos.

En particular, en este trabajo se ha usado la librería de R *FSinR* [Aragón-Royón et al. \(2020\)](#), para conseguir una abstracción apropiada de características se ha empleado la optimización basada en el proceso natural de selección de las hormigas ACO (Ant Colony Optimization). El algoritmo de optimización de colonia de hormigas se centra en generar iteraciones en una población aleatoria de individuos (hormigas) según los valores de una matriz de feromonas (que se actualiza en cada iteración según los caminos más seguidos por las hormigas) y una heurística (que determina cuán bueno es cada camino a seguir por las hormigas). La medida de evaluación se calcula para cada individuo. El algoritmo termina una vez que se ha alcanzado el número de iteraciones establecido.

Tabla 4.2 Iteraciones de la selección de características valoración del algoritmo J48, se presentan las distintas métricas con el rendimiento y el número de variables de cada iteración, se ha usado 21 hormigas.

Iter	Accuracy	Kappa	Sensitivity	Specificity	Precision	Recall	F1	AUC	Variables	$a^1 b^2$
m = 15	0,9519651	0,9096744	0,9291887	0,9606421	0,9705364	0,9291887	0,9478057	0,9808415	19	0 0
	0,9388646	0,8849618	0,9111111	0,9530872	0,9545582	0,9111111	0,9308647	0,9731544	11	1 1
	0,9519651	0,9096744	0,9291887	0,9606421	0,9705364	0,9291887	0,9478057	0,9808415	19	1 0
	0,8558952	0,7313258	0,8321869	0,9023588	0,861237	0,8321869	0,8457648	0,9178926	7	0 1
m = 100	0,9716981	0,9473074	0,9601621	0,9781532	0,9788676	0,9601621	0,9691819	0,9861644	18	0 0
	0,9402516	0,8888644	0,9198817	0,9591824	0,9425594	0,9198817	0,9302804	0,956077	10	1 1
	0,9685535	0,9413457	0,9504939	0,9770355	0,969658	0,9504939	0,9595692	0,9797764	19	1 0
	0,9056604	0,8226733	0,8849595	0,9331214	0,9143793	0,8849595	0,8987006	0,9544109	7	0 1
m = 500	0,9766667	0,9552086	0,965906	0,9834526	0,9769981	0,965906	0,9712822	0,986864	13	0 0
	0,9600001	0,924057	0,9535852	0,9764498	0,9525784	0,9535852	0,9526166	0,9763955	10	1 1
	0,9766667	0,9552239	0,967776	0,9834526	0,9767337	0,967776	0,9721855	0,9884232	19	1 0
	0,90666	0,82291	0,90219	0,93895	0,89702	0,90219	0,89953	0,94026	7	0 1
m = 1000	0,97333	0,95066	0,973890	0,98399	0,97078	0,97389	0,97213	0,98365	20	0 0
	0,9600001	0,92484	0,94852	0,97067	0,96552	0,94852	0,95675	0,97937	11	1 1
	0,96666	0,93863	0,97032	0,98113	0,96212	0,97032	0,96574	0,97933	19	1 0
	0,88333	0,77931	0,85374	0,92106	0,88758	0,85374	0,86907	0,93984	7	0 1

¹ Parámetro para controlar la influencia de la feromona (Si a=0, no se utiliza la información de la feromona)

² Parámetro para controlar la influencia de la heurística (Si b=0, no se tiene en cuenta el atractivo de los movimientos)

4.1.4. Balanceo de ejemplos

Para efectos de equilibrar los ejemplos y evitar el sesgo en el conocimiento, se ha conseguido equilibrar los datos usando el algoritmo SMOTE ([Chawla y Bowyer, 2002](#)), con el conjunto de datos que fue limpiado, es decir, el conjunto de datos como resultado de la eliminación de instancias y selección de características.

Tabla 4.3 Balanceo de ejemplos con algoritmo SMOTE

Datos	Bajo	Medio	Alto	Observaciones	Tasa Desequilibrio %
SMOTE	769	251	187	1207	4,11
Preprocesado	271	98	17	386	15,94
Original	283	106	18	407	15,72

Se ha conseguido equilibrar el conjunto de datos que se usará en la extracción de conocimiento, para esto, el algoritmo SMOTE logró reducir al 4,11% la tasa de desequilibrio de ejemplos, dado que en principio esta tasa fue alta 15,72% (original). No obstante, el preprocesado de los datos aumentó el desequilibrio debido a la eliminación de instancias. En concreto, el conjunto de datos ahora está mejor equilibrado y limpio para obtener un patrón de conocimiento útil.

4.2. Extracción de conocimiento

En esta sección, se han revisado diferentes enfoques para la extracción de reglas de asociación usando apriori como algoritmo base para la generación de reglas de asociación, para esto se han utilizado varias implementaciones del programa **KEEL** disponibles en el programa **R** ([R Core Team, 2019](#)) presentadas como una librería (package). En las secciones [4.2.1](#), [4.2.2](#), [4.2.3](#) fueron estudiados los distintos enfoques de reglas de asociación y posteriormente evaluados y/o al menos verificada la importancia de cada regla extraída. Para ello, se usan las métricas de calidad que ya se han descrito en la sección [2.4.2](#), página [12](#).

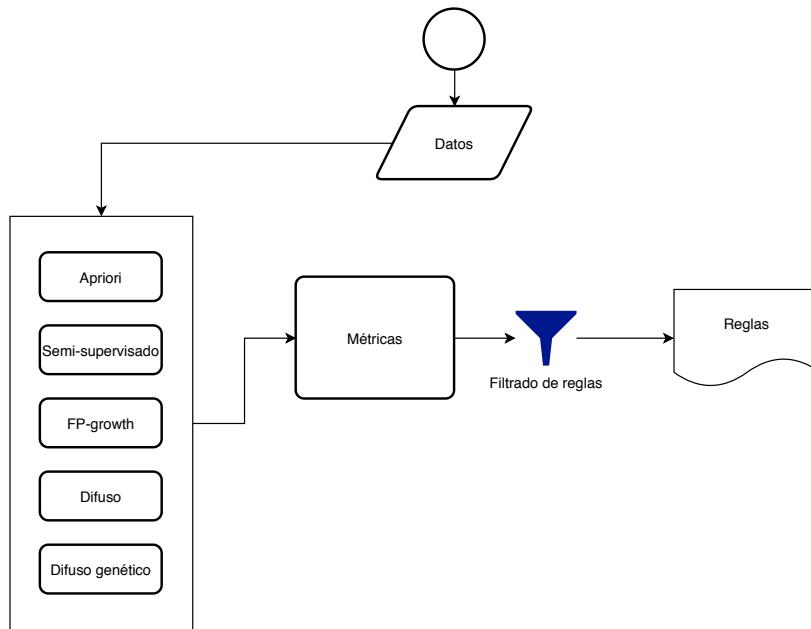


Figura 4.3 Distintos métodos usados para la extracción de reglas de asociación. Filtrado de reglas empleando métricas de calidad y presentación de mejores reglas.

Es evidente que la construcción de reglas de asociación crea un volumen de reglas muy alto. Esto, ha sido controlado mediante un filtrado de reglas que se aplica a las técnicas de reglas de asociación que se usaron en este trabajo. Primero, en el consecuente se han considerado las reglas que tienen un solo elemento y que esté relacionado con la variable *tiempoGraduacion*. Segundo, en el antecedente se filtran las reglas con dos y cinco elementos, las de un solo elemento o huérfanas son filtradas. Tercero, que la variable objetivo *tiempoGraduacion* no aparezca en el antecedente. Por último, que la métrica “*lift*” sea superior a uno.

$$\text{Filtrado} \left\{ \begin{array}{l} \text{Antecedente} \rightarrow 2-5 \text{ elementos} \\ \text{Consecuente} \rightarrow \text{tiempoGraduacion} \\ \text{Métrica} \rightarrow \text{Lift} \geq 1 \end{array} \right.$$

4.2.1. Apriori

De manera general, las reglas de asociación fueron construidas a partir del algoritmo apriori. Por ello, el proceso de extracción de reglas se ha realizado de la siguiente manera: Las variables numéricas se han transformado en categorías, después, se han convertido éstas categorías en un conjunto de transacciones. Luego, a partir de las transacciones el algoritmo construyó las reglas de asociación. Finalmente, el algoritmo obtuvo un volumen finito de reglas que generalmente es alto, no obstante, esto último estará en función de

la dimensionalidad de los datos analizados. Una de las principales complicaciones de las reglas de asociación es precisamente el volumen de reglas que se construyen, de ahí que, cada regla tendrá asociado criterios de calidad que permitirá valorar “la importancia” de la regla. En trabajos similares ([Berzal y Cubero, 2010](#); [De Carvalho et al., 2014](#); [Kumar y Chadha, 2012](#); [Onan et al., 2016](#); [Romero y Romero, 2010](#)) los autores exploran la calidad de las reglas a través de distintas métricas. Por tanto, siguiendo ésta misma línea de trabajo fueron utilizadas métricas de calidad para el filtrado de reglas significativas.

Semi-Supervisado

Como parte de la estrategia planteada para la extracción de reglas hemos considerado fijar el consecuente con la variable “tiempoGraduación”, esto, para filtrar las reglas que tengan mayor importancia ya que el objetivo está centrado en obtener conocimiento respecto a los factores que influyen con el éxito académico del alumnado, en este caso, se lo ha relacionado con el tiempo de graduación del alumnado. La regla #1 ha obtenido un valor de interés

Tabla 4.4 Principales reglas seleccionadas.

#	Reglas de asociación	<i>Soporte</i>	<i>Confianza</i>	<i>Lift</i>
1	SI PIndAcademico=Normal y TotalMatriculas=Normal ENTONCES tiempoGraduacion=Bajo	0,655	0.759	1,130
2	SI SIndAcademico=Normal y TotalMatriculas=Normal ENTONCES tiempoGraduacion=Bajo	0,655	0.752	1,120
3	SI TIndAcademico=Normal y TotalMatriculas=Normal ENTONCES tiempoGraduacion=Bajo	0,652	0.751	1,119
4	SI PIndAcademico=Normal y SIndAcademico=Normal ENTONCES tiempoGraduacion=Bajo	0,668	0.753	1,122
5	SI PIndAcademico=Normal y TIndAcademico=Normal ENTONCES tiempoGraduacion=Bajo	0,668	0.748	1,115
6	SI SIndAcademico=Normal y TIndAcademico=Normal ENTONCES tiempoGraduacion=Bajo	0,666	0.719	1,071
7	SI PIndAcademico=Normal y SIndAcademico=Normal y TIndAcademico=Normal y TotalMatriculas=Normal ENTONCES tiempoGraduacion=Bajo	0,652	0.775	1,155

#	Reglas de asociación	<i>Soporte</i>	Confianza	<i>Lift</i>
1	SI PIndAcademico=Normal y TotalMatriculas=Normal ENTONCES tiempoGraduacion=Bajo	0,655	0.759	1,130
7	SI PIndAcademico=Normal y SIndAcademico=Normal y TIndAcademico=Normal y TotalMatriculas=Normal ENTONCES tiempoGraduacion=Bajo	0,652	0.775	1,155

alto, así como la cobertura y confianza, no obstante, la variable “TotalMatriculas” allana las posibilidades de conocer de forma oportuna el éxito académico del alumnado, dado que el total de matriculas esta relacionado con la cantidad de periodos académicos que dura la titulación universitaria. De manera general, la regla #7 contiene al resto de reglas [1-6], el valor de ésta regla goza de un alto interés, fiabilidad y generalidad. No obstante, a corto plazo la regla no servirá hasta que el estudiante termine el tercer curso de la titulación académica, evidentemente aunque tenga valores importantes en las métricas resulta poco atractiva para el estudio.

Las reglas de asociación construidas a partir del algoritmo apriori han generado reglas que son de interés, no obstante, son altamente legibles debido a la categorización de las variables. Entre tanto, los resultados que se han conseguido en este análisis es necesario filtrarlos y estudiar con otras técnicas las adicionales las reglas de asociación que se lantean en la sección 4.2.2 y 4.2.3. En el anexo de este trabajo, detallamos los resultados de esta estrategia de extracción de reglas (Véase 5, página 81).

4.2.2. Apriori difuso

Otro punto de vista de las reglas de asociación esta relacionado con la lógica difusa o lógica borrosa (Fuzzy Logic, del inglés). En el trabajo de [Hong et al. \(2001\)](#) se ha propuesto la combinación de “Fuzzy Logic” y las reglas de asociación. Para esto, El algoritmo de minería difusa transforma primero cada valor cuantitativo en un conjunto difuso de términos lingüísticos utilizando un conjunto determinado de funciones de los miembros. El algoritmo calcula entonces la cardinalidad escalar de cada término lingüístico en todos los datos de la transacción. El proceso de minería basado en recuentos difusos se realiza entonces para encontrar reglas de asociación difusas. Así, el conjunto de datos relacionado con el alumnado universitario ha sido entrenado con el algoritmo donde se han generado una cantidad ingente de reglas según los valores de soporte y confianza.

Análisis de reglas de asociación difusas seleccionadas

Esta selección, se la ha realizado aplicando como criterio la métrica Lift con un valor superior a uno, además el criterio de confianza superior o igual al 60 %. Este ha sido el margen que se ha aplicado para la selección de las reglas.

Tabla 4.5 Conjunto de reglas de asociación difusa seleccionadas

#	Reglas Difusas	Soporte	Confianza	Lift
1	SI NumHijos= L1 y SEDAD2= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.8	1,35
2	SI SMediaEdad= L1 y TEXPEDOCEN= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.7	1,14
3	SI NumHijos= L1 y SMediaEdad= L1 AND TEXPEDOCEN= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.7	1,13
4	SI FinanciamientoColegio= L1 y PIndAcademico= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.7	1,10
22	SI MediaNotasAdmision= L1 y FinanciamientoColegio= L1 AND PEXPEDOCEN= L1ENTONCES TiempoGraduacion=Bajo	0,2	0.6	1,06
56	SI JornadaTrabajo= L2ENTONCES TiempoGraduacion=Bajo	0,2	0.6	1,01
60	SI JornadaTrabajo= L1 y PEDAD3= L1ENTONCES TiempoGraduacion=Bajo	0,2	0.6	1,01

La tabla ?? muestra las reglas de asociación difusa. Saltándonos el orden lógico de las reglas seleccionadas. Llama la atención la regla # 56, dado que la regla tiene un interés sustancial y una fiabilidad de 60 %. Si bien su fiabilidad no es excelente, cubre un número determinado de casos de la base de reglas de asociación. El hecho que se encuentre una regla con un único antecedente, en un sistema con variables de entradas múltiples, simplifica la información analizada ya que se deriva de una conclusión clara. Por ejemplo, en este caso, sabemos que la implicación del alumnado en algún tipo de trabajo conlleva al éxito académico.

#	Regla Difusa	<i>Soporte</i>	Confianza	<i>Lift</i>
56	SI JornadaTrabajo= L2 ENTONCES Tiempo-Graduacion=Bajo	0,2	0.6	1,01

Tanto el interés como la fiabilidad de la regla # 1 destacan, la precisión es muy alta y la generalidad de la regla es aceptable. En ésta regla, por un lado el grupo de profesores asignados a impartir clases con edades superiores a 45 y menores que 60 con valor L1 (SEDAD2) y, por otro, el alumnado con un determinado número de hijos (NumHijos) con valor L1. En concreto, esta regla tiene de cara al experto dos componentes, el exógeno que corresponde a la parte social del alumnado (NumHijos) y por otra, endógeno que corresponde al contexto universitario(SEDAD2), aunque para la primera no se puede incidir ya en la segunda si que se puede adaptarla para que el alumnado consiga el éxito académico.

#	Regla Difusa	<i>Soporte</i>	Confianza	<i>Lift</i>
1	SI NumHijos= L1 y SEDAD2= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.8	1,35

La regla # 3 tiene un alto interés y goza de una notable fiabilidad. En ésta regla, el antecedente tiene tres variables que importan y mucho para que el alumnado alcance el éxito académico. La primera, corresponde a la media de edad del profesorado del segundo año. La segunda, la experiencia del profesorado del tercer curso y el o los integrantes de su núcleo familiar, todas con L1. El agregado, del antecedente de ésta regla es la parte social del alumno, que incide en el grado de responsabilidad individual y encauza hacia el éxito académico.

#	Regla Difusa	<i>Soporte</i>	Confianza	<i>Lift</i>
3	SI NumHijos= L1 y SMediaEdad= L1 y TEXPEDOCEN= L1 ENTONCES Tiempo-Graduacion=Bajo	0,2	0.7	1,13

Por último, la regla # 4 se corresponde con dos variables: el financiamiento del colegio secundario (público, privado) y el Índice académico del primer curso del alumnado, es decir, ponderación de calificaciones en función del número de convocatorias realizadas. En resumen, el buen rendimiento académico del alumnado en el primer año de estudio es clave para alcanzar el éxito académico.

#	Regla Difusa	<i>Soporte</i>	Confianza	<i>Lift</i>
4	SI FinanciamientoColegio= L1 y PIndAcademico= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.7	1,10

Por otra parte, la regla # 22 de acuerdo con las métricas posee un interés importante y una fiabilidad de 0,6. La regla cubre un determinado número de casos. El antecedente considera la nota media de admisión a la universidad en L1, junto con el tipo de financiamiento del colegio L1, además de la experiencia del profesorado en el primer curso PEXPEDOCEN sea también L1. En concreto, ésta regla

#	Regla Difusa	<i>Soporte</i>	Confianza	<i>Lift</i>
22	SI MediaNotasAdmision= L1 y FinanciamientoColegio= L1 y PEXPEDOCEN= L1 ENTONCES TiempoGraduacion=Bajo	0,2	0.6	1,06

Visualización de reglas de asociación difusas

En las figuras 4.4 se presenta la construcción de reglas de asociación difusa con los criterios de filtrado expuestos (Véase sección 4.2, página 53). Las ruedas de dependencia tienen dos partes, la primera es el antecedente que está descrito con la letra “A” y el número del atributo, la segunda, que es el consecuente descrito con la letra “C”. No obstante, la variable objetivo es “C_AIM” (consecuente con atributo importante).

De manera general, el proceso de extraer las reglas de asociación difusa se condensan con el mapa de calor mostrado en la figura 4.5. El soporte y la confianza fueron necesarios para la generación de las reglas de asociación. A partir de allí, el interés de las reglas se obtuvo por medio de la métrica “Lift”.

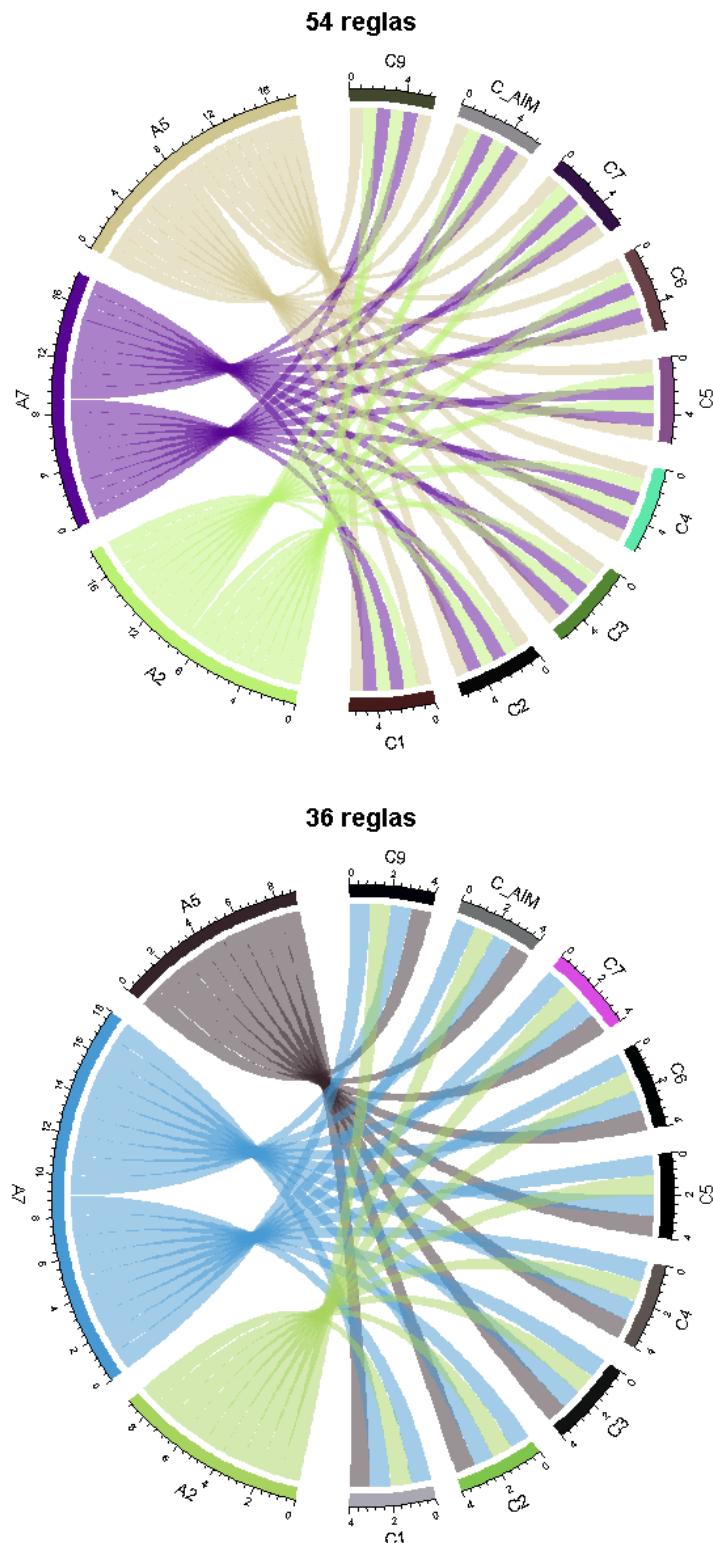


Figura 4.4 Reglas de asociación difusas mayormente representativas, estas reglas se han filtrado según la métrica Lift y además el número de términos del consecuente que se ajusta a un solo término, así como el antecedente en un intervalo entre dos y 5 términos. En la rueda de dependencia con 54 reglas el soporte es 0,5 y la confianza 0,1. Mientras que en la rueda de 36 reglas el soporte es 0,5 y la confianza 0,8.

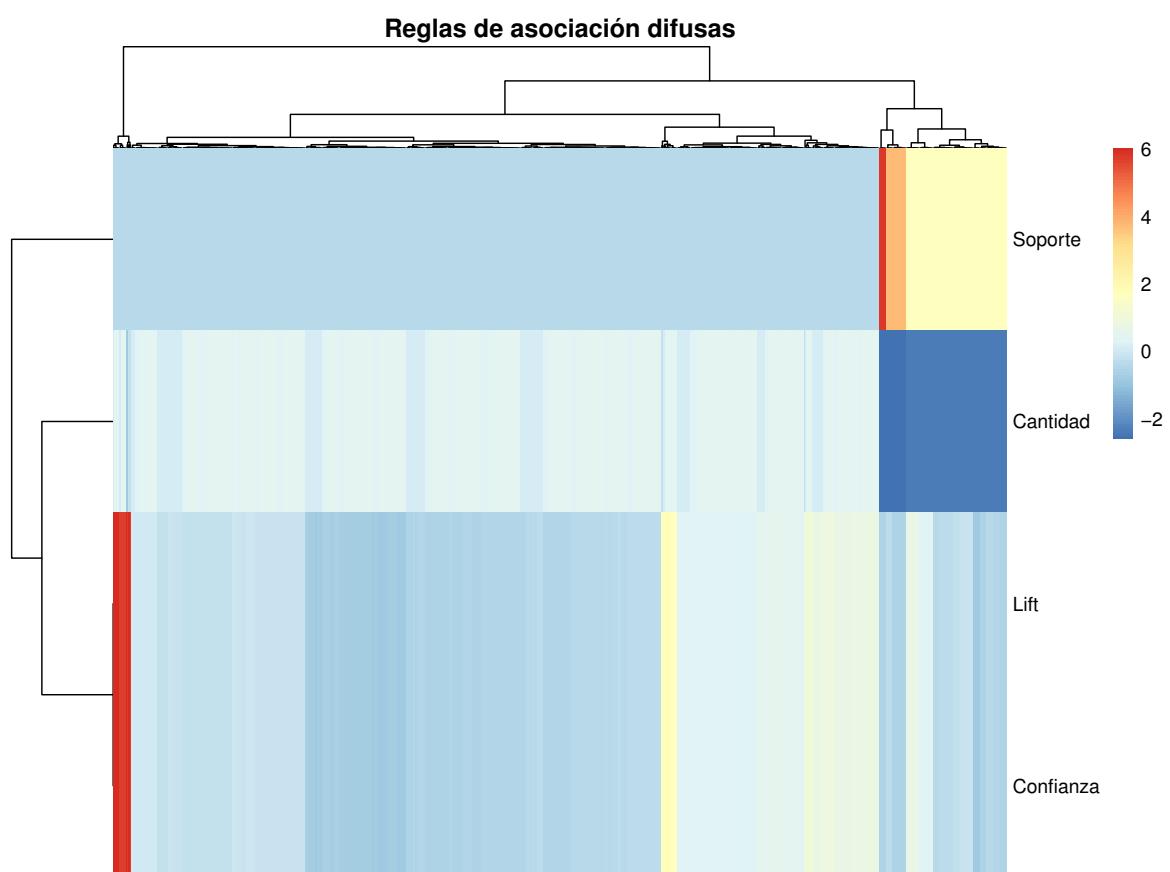


Figura 4.5 El mapa de calor presenta la generación de 33465 reglas de asociación difusas construidas. Cuanto más claro y rojizo sea el color significa que son reglas de interés, mientras que menos claro y azulado serán reglas de poco interés.

4.2.3. Apriori difuso y algoritmo genético

Las reglas difusas extraídas en la sección 4.2.2 ahora se han modificado usando el algoritmo genético, esto para intervenir en el proceso de construcción de las reglas a través de un proceso genético que consiga una mayor amplitud y mejora la calidad de las reglas [Hong et al. \(2006\)](#).

Reglas extraídas con apriori difuso y algoritmo genético

Tabla 4.6 Conjunto de reglas de asociación genético difusas

	Regla de asociación	Soporte	Confianza	Lift
1	PIndAcademico= L3 y SEDAD2= L4 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,33
2	JornadaTrabajo= L3 y SEDAD2= L4 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,24
3	JornadaTrabajo= L7 y TMediaEdad= L4 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,07
4	JornadaTrabajo= L7 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,06
5	JornadaTrabajo= L3 y TMediaEdad= L4 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,05
6	TEDAD1= L2 y TMediaEdad= L4 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,04
7	TEDAD1= L2 y JornadaTrabajo= L7 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,04
8	JornadaTrabajo= L3 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,03
9	PIndAcademico= L3 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,03
10	TEDAD1= L2 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,02

Por su parte, la regla #8 aunque tiene poca fiabilidad (10 %) goza de un gran interés debido a que de forma anticipada se conoce la nota media de admisión universitaria (selectividad) y si trabaja. Este aporte de la regla es importante.

Regla de asociación	<i>Soporte</i>	Confianza	<i>Lift</i>
8 SI JornadaTrabajo= L3 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,03

La regla #4 por su parte, tiene poca fiabilidad pero goza de interés. Esta regla estuvo centrada en el profesorado del tercer curso, dado que como parte del éxito académico del alumnado está la adaptación del curso a grupos de edades de profesores noveles (TEDAD1) y la media del grupo de edades (TMediaEdad).

Regla de asociación	<i>Soporte</i>	Confianza	<i>Lift</i>
6 SI TEDAD1= L2 y TMediaEdad= L4 ENTONCES TiempoGraduacion=Bajo	0,20	0.10	1,04

Visualización de reglas de asociación difusa con algoritmo genético

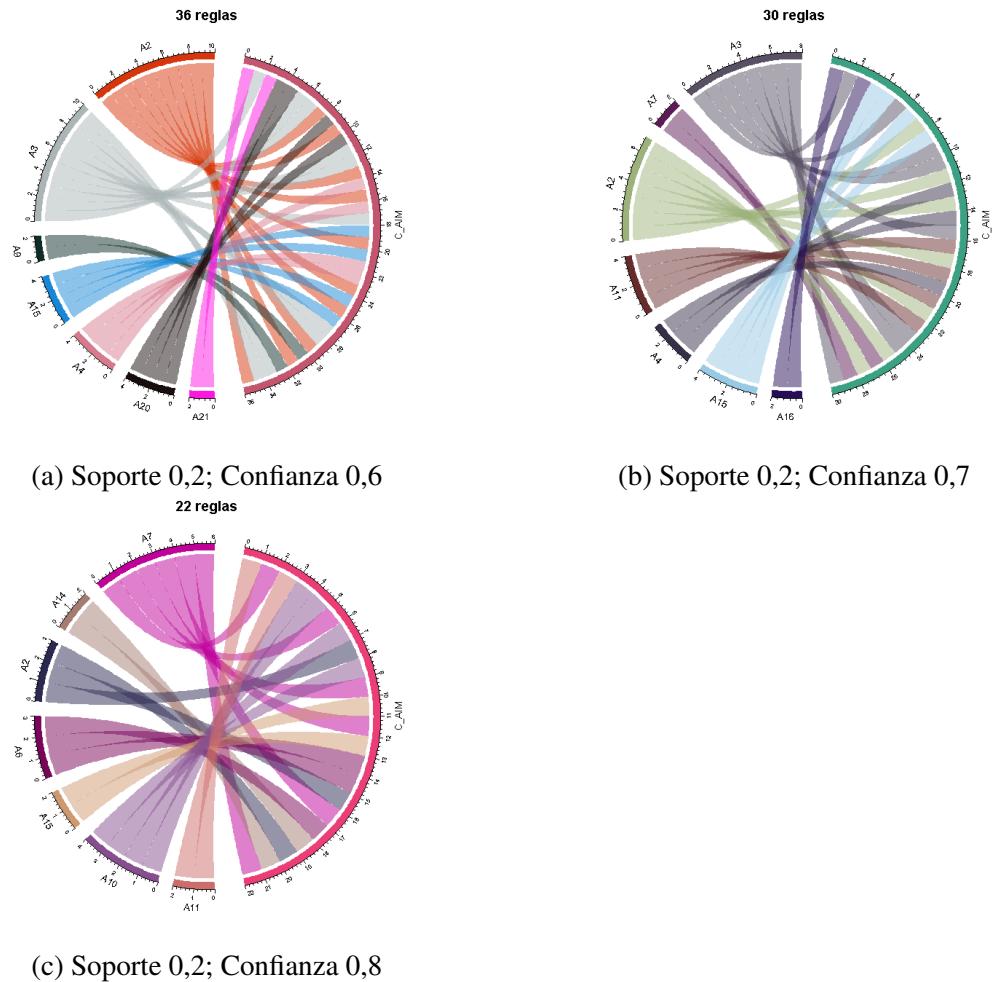


Figura 4.6 Reglas de asociación usando variable tiempoGraduacion como parte del consecuente.

4.3. Análisis de la extracción del conocimiento

El conjunto de reglas de asociación que fueron extraídas del algoritmo genético es diverso. Se puede evidenciar en los mapas de calor generados ([1](#), [4.5](#), [10](#), [12](#) y [16](#)) según cada enfoque usado en el trabajo, donde la métrica *Lift* es distinta en cada caso. Esto, por la técnica propiamente seleccionada para generar las reglas de asociación.

Comparando los métodos aplicados para la extracción de las reglas de asociación se ha diferenciado entre las técnicas aplicadas el volumen de reglas obtenidas en cada caso. Si nos fijamos en las subfiguras a,b,c y d de la Figura [4.7](#). Tanto *a* y *b* tienen los extremos frío y calor, es decir, pocas y muchas reglas de asociación. Por otro lado, entre *b* y *d* existe una graduación menor. Ahora bien, en *d* existe una relación entre el soporte azulado y el R.lift rojizo, que son las reglas que se han analizado.

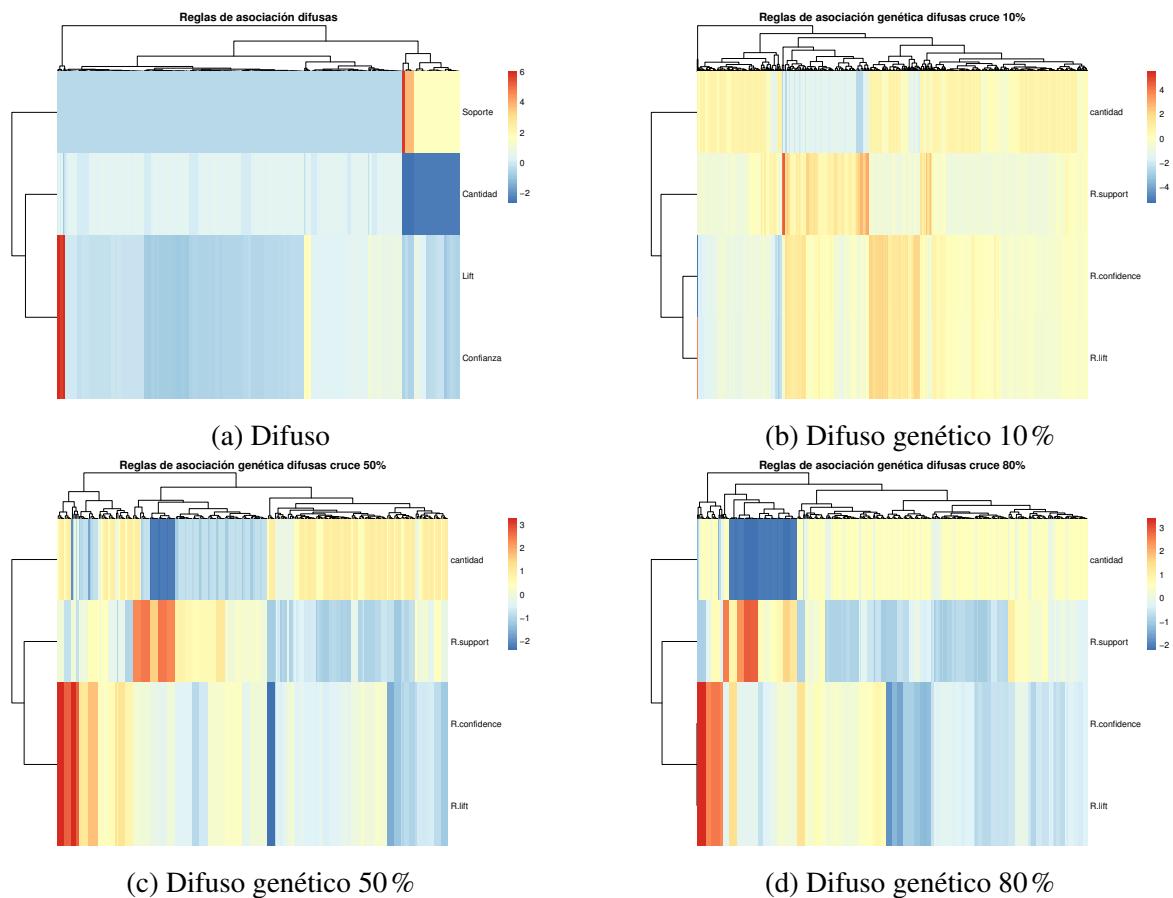


Figura 4.7 Resumen de mapas de calor

De manera general, se ha evidenciado que las reglas de asociación difusas que usan algoritmo genético para construir las reglas ha obtenido reglas menos fiables (10%) y de alto interés donde de forma temprana se ha conseguido patrones de datos que permiten al

alumnado el éxito académico. Las reglas de asociación generales se encuentran compartidas en este [enlace](#).

Capítulo 5

Conclusiones

Con el incentivo de las preguntas de este trabajo. Se logró explorar las características del alumnado y las técnicas de minería de datos usando reglas de asociación que combinadas aporten patrones ocultos generales y de interés. Para ello, las reglas de asociación difusas con algoritmo genético consiguen reglas de asociación de alto interés. En este trabajo, se ha aplicado técnicas de preprocesamiento de datos, selección de características usando el algoritmo de colonia de hormigas, y, de instancias, así como también el balanceo de datos. Además, se han explorado dos técnicas adicionales a las reglas de asociación clásica (apriori) para la búsqueda de patrones. El propósito fue encontrar una técnica que combinada con las reglas de asociación consigan patrones generales y de interés. Para ello, se utilizó la lógica difusa para construir reglas de asociación. Después, fueron combinadas las reglas de asociación difusas con un algoritmo genético básico para generar reglas de asociación. Evidentemente, la combinación del algoritmo genético y las reglas de asociación difusas han conseguido reglas de alto interés dado que las peores reglas son sustituidas por otras mejores a través de la probabilidad de cruce del algoritmo genético. Los resultados encontrados han evidenciado reglas significativas para el éxito académico del alumnado que encauzan en la titulación universitaria en el menor tiempo. Como trabajo futuro se propone combinar las reglas de asociación difusas y el aprendizaje por refuerzo, con el fin de estudiar las reglas huérfanas de la base de reglas obtenidas.

Bibliografía

Clustering Algorithm - an overview | ScienceDirect Topics.

- Agrawal, R. y Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB '94*, page 487–499, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Ahmad, W. y Bakar, A. (2020). Ensemble machine learning model for higher learning scholarship award decisions. *International Journal of Advanced Computer Science and Applications*, 11(5):303–312. cited By 0.
- Al Fanah, M. y Ansari, M. (2019). Understanding e-learners' behaviour using data mining techniques. pages 59–65. cited By 0.
- Aldowah, H., Al-Samarraie, H., y Fauzy, W. M. (2019). Educational data mining and learning analytics for 21st century higher education: A review and synthesis. *Telematics Informatics*, 37:13–49.
- Alharbi, Z., Cornford, J., Dolder, L., y De La Iglesia, B. (2016). Using data mining techniques to predict students at risk of poor performance. pages 523–531. cited By 15.
- Alshamaila, Y., Habib, M., Aljarah, I., Alsawalqah, H., Faris, H., y Alsoud, A. (2020). An intelligent approach for the effect of social media on undergraduate students performance: A case study in the university of jordan. In *Proceeding Series*, pages 102–108. cited By 0.
- Altay, E. V. y Alatas, B. (2019). Sensitivity analysis of modenar method for mining of numeric association rules. In *2019 1st International Informatics and Software Engineering Conference (UBMYK)*, pages 1–6.
- Aragón-Royón, F., Jiménez-Vílchez, A., Arauzo-Azofra, A., y Benítez, J. M. (2020). Fsinr: an exhaustive package for feature selection.
- Bastian, M., Heymann, S., y Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks.
- Batista, G. y Monard, M. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*, 17(5):519–533.
- Berzal, F. y Cubero, J.-C. (2010). Interestingness Measures for Association Rules. *Ipmu*, pages 298–307.
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., y Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

- Bonde, S. y Kirange, D. (2020). Educational data mining survey for predicting student's academic performance. *Lecture Notes on Data Engineering and Communications Technologies*, 31:293–302. cited By 0.
- Brandes, U. (2001). A faster algorithm for betweenness centrality. *The Journal of Mathematical Sociology*, 25(2):163–177.
- Brin, S., Motwani, R., Ullman, J. D., y Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *SIGMOD Rec.*, 26(2):255–264.
- Cadenas, J. M. y Garrido, M. C. (2018). Intelligent data analysis, soft computing and imperfect data. In *Soft Computing Based Optimization and Decision Models*. Springer.
- Cano, A. y Leonard, J. (2019). Interpretable multiview early warning system adapted to underrepresented student populations. *IEEE Transactions on Learning Technologies*, 12(2):198–211. cited By 3.
- Charte, F. (2014). *Análisis exploratorio y visualización de datos con R*.
- Chawla, N. y Bowyer, K. (2002). SMOTE: Synthetic Minority Over-sampling Technique Nitesh. *Journal of Artificial Intelligence Research*, 16:321–357.
- Crivei, L., Czibula, G., Ciubotariu, G., y Dindelegan, M. (2020). Unsupervised learning based mining of academic data sets for students' performance analysis. pages 11–16. cited By 0.
- Cuseo, J. (2012). Student success : Definition , outcomes , principles and practices defining student success. Technical report.
- Czibula, G., Mihai, A., y Crivei, L. (2019). S prar: A novel relational association rule mining classification model applied for academic performance prediction. volume 159, pages 20–29. cited By 2.
- De Carvalho, V. O., De Padua, R., y Rezende, S. O. (2014). Semi-supervised learning to support the exploration of association rules. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 8646 LNCS, pages 452–464.
- DELAVARI, N., PHON-AMNUAISUK, S., y BEIKZADEH, M. R. (2008). Data mining application in higher learning institutions. *Informatics in Education*, 7(1):31–54.
- Deogun, J., Spaulding, W., Shuart, B., y Li, D. (2004). Towards missing data imputation: A study of fuzzy k-means clustering method. In *4th International Conference of Rough Sets and Current Trends in Computing(RSCTC'04)*, volume 3066 of *Lecture Notes on Computer Science*, pages 573–579. Lecture Notes In Computer Science.
- Espinosa, C. I. (2019). Similitud de Comunidades biológicas.
- Fayyad, U., Piatetsky-Shapiro, G., y Smyth, P. (1996). From Data Mining to Knowledge Discovery in Databases. *AI Magazine*, 17(3):37.
- Flynt, A. y Dean, N. (2016). A Survey of Popular R Packages for Cluster Analysis. *Journal of Educational and Behavioral Statistics*, 41(2):205–225.
- François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., y Pineau, J. (2018). An introduction to deep reinforcement learning. *Foundations and Trends® in Machine Learning*, 11(3-4):219–354.

- Fürnkranz, J., Gamberger, D., y Lavrač, N. (2012). Foundations of Rule Learning. .
- Grzymala-Busse, J., Goodwin, L., Grzymala-Busse, W., y Zheng, X. (2005). Handling missing attribute values in preterm birth data sets. In *10th International Conference of Rough Sets and Fuzzy Sets and Data Mining and Granular Computing(RSFDGrC'05)*, volume 3642 of *Lecture Notes on Computer Science*, pages 342–351. Lecture Notes in Computer Science.
- Guanin-Fajardo, J., Casillas, J., y Chiriboga-Casanov, W. (2019). Semi-supervised learning to discover the average scale of graduation of university students. *REVISTA CONRADÓ*, 15(70):291–299.
- Guillet, F. y Hamilton, H. J. (2007). *Quality Measures in Data Mining (Studies in Computational Intelligence)*. Springer-Verlag, Berlin, Heidelberg.
- Han, J., Pei, J., Yin, Y., y Mao, R. (2004a). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data Mining and Knowledge Discovery*, 8(1):53–87.
- Han, J., Pei, J., Yin, Y., y Mao, R. (2004b). Mining frequent patterns without candidate generation: A frequent-pattern tree approach.
- Hellas, A., Ihantola, P., Petersen, A., Ajanovski, V., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C., y Liao, S. (2018). Predicting academic performance: A systematic literature review. pages 175–199. cited By 23.
- Hong, T.-P., Chen, C.-H., Wu, Y.-L., y Lee, Y.-C. (2006). A ga-based fuzzy mining approach to achieve a trade-off between number of rules and suitability of membership functions. *Soft Computing*, 10(11):1091–1101.
- Hong, T.-P., Kuo, C.-S., y Chi, S.-C. (2001). Trade-off between computation time and number of rules for fuzzy mining from quantitative data. *International Journal of Uncertainty and Fuzziness and Knowledge-Based Systems*, 9(5):587–604.
- Internacional, X. C. (2015). Educación y Aprendizaje.
- Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. In *European Conference on Machine Learning 1994(ECML94)*, pages 171–182.
- Kumar, V. y Chadha, A. (2012). Mining Association Rules in Students Assessment Data. *International Journal of Computer Science Issues*, 9(5):211–216.
- Lambiotte, R., Delvenne, J.-C., y Barahona, M. (2014). Random walks, markov processes and the multiscale modular organization of complex networks. *IEEE Transactions on Network Science and Engineering*, 1(2):76–90.
- Laura Emmanuella, L. E. A. y De Paula Canuto, A. M. (2014). Filter-based optimization techniques for selection of feature subsets in ensemble systems. *Expert Systems with Applications*, 41(4 PART 2):1622–1631.
- Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., y Liu, H. (2017). Feature selection: A data perspective.
- Liu, H. y Setiono, R. (1996). A probabilistic approach to feature selection: A filter solution. In *13th International Conference on Machine Learning(ICML{}96)*, pages 319–327.

- Meghji, A. F., Mahoto, N. A., Unar, M. A., y Shaikh, M. A. (2020). The Role of Knowledge Management and Data Mining in Improving Educational Practices and the Learning Infrastructure. *MEHRAN UNIVERSITY RESEARCH JOURNAL OF ENGINEERING AND TECHNOLOGY*, 39(2):310–323.
- Mi, C. (2019). Student performance early warning based on data mining. *International Journal of Performativity Engineering*, 15(3):822–833. cited By 1.
- Morales, P., Luengo, J., Garcia, L. P., Lorena, A. C., de Carvalho, A. C., y Herrera, F. (2016). *NoiseFiltersR: Label Noise Filters for Data Preprocessing in Classification*. R package version 0.1.0.
- Morris, L. V. (2016). Mining Data for Student Success. *Innovative Higher Education*, 41(3):183–185.
- Mursalin, M., Zhang, Y., Chen, Y., y Chawla, N. V. (2017). Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. *Neurocomputing*, 241:204–214.
- Nandeshwar, A., Menzies, T., y Nelson, A. (2011). Learning patterns of university student retention. *Expert Systems with Applications*, 38(12):14984–14996. cited By 36.
- Onan, A., Bal, V., y Bayam, B. Y. (2016). The Use of Data Mining for Strategic Management: A Case Study on Mining Association Rules in Student Information System. *Croatian Journal of Education*, 18(1):41–70.
- Piatetsky-Shapiro, G. (1991). Discovery, analysis, and presentation of strong rules. In Piatetsky-Shapiro, G. y Frawley, W. J., editors, *Knowledge Discovery in Databases*, pages 229–248. AAAI/MIT Press.
- R Core Team (2019). R: A language and environment for statistical computing.
- Romero, C. y Romero, J. (2010). Mining rare association rules from e-learning data. *Journal of Educational Data Mining*, pages 171–180.
- Romero, C., Ventura, S., Vasilyeva, E., y Pechenizkiy, M. Class association rule mining from students' test data.
- Ruiz, C. y Miguel, D. (2001). Factores familiares vinculados al bajo rendimiento. *Revista Complutense de Educación*, 12(1):81–113.
- Santoso, L. (2018). Early warning system for academic using data mining. cited By 0.
- Saunders-Scott, D., Braley, M. B., y Stennes-Spidahl, N. (2018). Traditional and psychological factors associated with academic success: investigating best predictors of college retention. *Motivation and Emotion*, 42(4):459–465.
- Schubert, E. y Rousseeuw, P. J. (2018). Faster k-medoids clustering: Improving the pam, clara, and CLARANS algorithms. *CoRR*, abs/1810.05691.
- Siri, A. (2014). *Predicting students' academic dropout using artificial neural networks*. cited By 2.
- Sodanil, M., Chotirat, S., Poomhiran, L., y Viriyapant, K. (2019). Guideline for academic support of student career path using mining algorithm. *NLPIR 2019*, pages 133–137, New York, NY, USA. Association for Computing Machinery.

- Stańczyk, U. (2015). Weighting of features by sequential selection. *Studies in Computational Intelligence*, 584:71–90.
- Stańczyk, U. y Jain, L. C. (2015). *Feature selection for data and pattern recognition: An introduction*, volume 584 of *Studies in Computational Intelligence*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., y Altman, R. (2001). Missing value estimation methods for dna microarrays. *Bioinformatics*, 17:520–525.
- Wang, Y., Wu, L., Yuan, X., y Gao, B. (2018). Association Analysis of University Course Information Based on Knowledge Map. In *2018 NINTH INTERNATIONAL CONFERENCE ON INFORMATION TECHNOLOGY IN MEDICINE AND EDUCATION (ITME 2018)*, pages 393–397.
- Wilson, D. (1972). Asymptotic properties of nearest neighbor rules using edited data. *IEEE Transactions on Systems and Man and Cybernetics*, 2(3):408–421.
- Zafra, A., Romero, C., y Ventura, S. (2011). Multiple instance learning for classifying students in learning management systems. *Expert Systems with Applications*, 38(12):15020–15031. cited By 33.
- Zeineddine, H., Braendle, U., y Farah, A. (2019). Auto-generated ensemble model for predicting student success. cited By 0.
- Zhang, C. y Zhang, S. (2002). *Association Rule Mining: Models and Algorithms*. Springer-Verlag, Berlin, Heidelberg.
- Zhang, M. y He, C. (2010). *Survey on Association Rules Mining Algorithms*, pages 111–118. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Zhang, W., Liao, H., y Zhao, N. (2008). Research on the fp growth algorithm about association rule mining. In *2008 International Seminar on Business and Information Management*, volume 1, pages 315–318.

Anexos

Descripción de los datos

A.1. Conjunto de datos

Tabla 1 Descripción de variables del conjunto de datos

Variables	Tipo	Valor	Descripción
GradoUniversitario	Categórica		Titulaciones universitarias ofertadas por la Universidad
EdadIngreso	Numérica	16-46	Edad de ingreso del alumnado a la primera titulación universitaria
FinanciamientoColegio	Categórica	(Público, Privado)	Tipo de Colegio donde obtuvo el estudiante obtuvo el bachillerato
UbicacionColegio	Categórica	Urbana, Rural	Lugar donde se encuentra geolocalizado el colegio
MediaNotasAdmision	Numérica	0-10	Nota media del curso de admisión universitaria (selectividad)
PEDAD1	Numérica	0-15	Cantidad de profesores noveles que impartieron clases en primer año
PEDAD2	Numérica	0-15	Cantidad de profesores maduros que impartieron clases en primer año
PEDAD3	Numérica	0-15	Cantidad de profesores expertos que impartieron clases en primer año
PEXPEDOCEN	Numérica	0-30	Media de la experiencia docente que impartió clases en primer año
PMediaEdad	Numérica	30-45	Media de edad del profesorado que impartió clases en primer año
			Continúa en la siguiente página...

Variables	Tipo	Valor	Descripción
PIndAcademico	Numérica	0-4	Índice de ponderación de media de calificaciones y número de convocatorias a exámenes
SEDAD1	Numérica	0-15	Cantidad de profesores novatos que impartieron clases en segundo año
SEDAD2	Numérica	0-15	Cantidad de profesores maduros que impartieron clases en segundo año
SEDAD3	Numérica	0-15	Cantidad de profesores expertos que impartieron clases en segundo año
SEXPEDOCEN	Numérica	0-30	Media de la experiencia docente que impartió clases en segundo año
SMediaEdad	Numérica	45-60	Media de edad del profesorado que impartió clases en primer año
SIndAcademico	Numérica	0-4	Índice de ponderación de media de calificaciones y número de convocatorias a exámenes
TEDAD1	Numérica	0-15	Cantidad de profesores novatos que impartieron clases en tercer año
TEDAD2	Numérica	0-15	Cantidad de profesores maduros que impartieron clases en tercer año
TEDAD3	Numérica	0-15	Cantidad de profesores expertos que impartieron clases en tercer año
TEXPEDOCEN	Numérica	0-30	Media de la experiencia docente que impartió clases en tercer año
TMediaEdad	Numérica	60-80	Media de edad del profesorado que impartió clases en primer año
TIndAcademico	Numérica	0-4	Índice de ponderación de media de calificaciones y número de convocatorias a exámenes
NumHijos	Numérica	0-4	Cantidad de hijos del estudiante
EstiloAprendizaje	Numérica	V,A,R,K	Tipo de estilo de aprendizaje del estudiante
ViveCon	Categórica		Sitio familiar donde reside el estudiante
JornadaTrabajo*	Categórica	NT, TP, MT, TC, EV	Jornada de trabajo

Continúa en la siguiente página...

Variables	Tipo	Valor	Descripción
TotalMatriculas	Numérica	1-15	Cantidad de matriculas en la o las distintas titulaciones
Financiamiento	Numérica	1-4	Financiamiento que realiza el estudiante para cursar el grado universitario
Zona	Numérica	1-4	Financiamiento que realiza el estudiante para cursar el grado universitario
Ingreso	Numérica	1-4	Financiamiento que realiza el estudiante para cursar el grado universitario
Vivienda	Numérica	1-4	Financiamiento que realiza el estudiante para cursar el grado universitario
MediaNotasGlobales	Numérica	0-10	Nota media de todos los cursos de la titulación
MediaPorcAsistencia	Numérica	0-100	Media de asistencia global
MediaIndiceAcademico	Numérica	0-4	Media del índice de ponderación de calificaciones y número de convocatorias a examenes
TiempoGraduacion*	Numérica	5-15	Cantidad de años que toma el estudiante hasta lograr la titulación

El conjunto de datos que se ha utilizado para este trabajo, fue extraído en dos partes: (i) La primera, por medio del sistema universitario se recuperó información de los estudiantes universitarios de los periodos académicos comprendidos entre el 2003 y 2014. (ii) La segunda, fue recopilada de forma “manual” digitando datos complementarios de los archivos de las Facultades y de la Unidad de Recursos Humanos sobre la información del profesorado. Después, fueron juntadas ambas partes (sistema y manual) para obtener finalmente el conjunto de datos.

A.1.1. Estudio descriptivo del conjunto de datos

La tabla ?? muestra los principales cálculos estadísticos de las variables del conjunto de datos, para esto se ha utilizado medidas de posicionamiento central, así como también de dispersión.

Tabla 2 Estudio descriptivo del conjunto de datos

	Variables	n	mean	sd	median	range	skew	kurtosis
GradoUniversitario*	1	407.00	7.17	3.51	6.00	12.00	0.15	-1.43
EdadIngreso*	2	407.00	1.09	0.36	1.00	3.00	4.68	23.77
FinanciamientoColegio*	3	407.00	2.17	0.38	2.00	2.00	1.64	1.16
UbicacionColegio*	4	407.00	1.43	0.67	1.00	2.00	1.26	0.29
MediaNotasAdmision*	5	407.00	4.55	1.53	4.00	7.00	0.13	-0.27
PEDAD1*	6	407.00	1.24	0.55	1.00	3.00	2.52	6.92
PEDAD2*	7	407.00	2.55	0.73	2.00	3.00	0.39	-0.40
PEDAD3*	8	407.00	2.12	0.97	2.00	3.00	0.35	-0.97
PEXPEDOCEN*	9	407.00	3.73	1.03	4.00	4.00	-0.46	-0.61
PMediaEdad*	10	407.00	3.76	1.14	4.00	4.00	-0.93	0.27
PIndAcademico*	11	407.00	2.56	1.09	2.00	4.00	0.34	-0.99
SEDAD1*	12	407.00	1.20	0.64	1.00	3.00	3.58	12.18
SEDAD2*	13	407.00	2.29	0.76	2.00	3.00	0.11	-0.38
SEDAD3*	14	407.00	2.20	1.05	2.00	3.00	0.23	-1.24
SEXPEDOCEN*	15	407.00	3.78	1.08	4.00	4.00	-0.54	-0.29
SMediaEdad*	16	407.00	8.12	2.86	9.00	10.00	-1.22	0.30
SIndAcademico*	17	407.00	2.13	0.99	2.00	4.00	0.75	0.07
TEDAD1*	18	407.00	1.38	0.85	1.00	3.00	2.13	3.23
TEDAD2*	19	407.00	1.84	0.76	2.00	3.00	0.74	0.35
TEDAD3*	20	407.00	2.14	1.03	2.00	3.00	0.50	-0.89
TEXPEDOCEN*	21	407.00	3.40	1.20	3.00	4.00	-0.19	-0.81
TMediaEdad*	22	407.00	3.53	1.10	4.00	4.00	-0.51	-0.53
TIndAcademico*	23	407.00	1.95	0.73	2.00	2.00	0.08	-1.13
NumHijos*	24	407.00	1.37	0.73	1.00	4.00	2.24	5.16
EstiloAprendizaje*	25	407.00	2.11	1.17	2.00	3.00	0.45	-1.36
ViveCon*	26	407.00	3.89	1.24	4.00	5.00	-0.35	-0.56
JornadaTrabajo*	27	407.00	2.97	0.47	3.00	3.00	-1.62	7.33
TotalMatriculas*	28	407.00	3.05	0.60	3.00	5.00	0.47	5.02
Financiamiento*	29	407.00	2.24	0.86	3.00	2.00	-0.49	-1.46
Zona*	30	407.00	2.77	1.02	3.00	3.00	-0.23	-1.13
Ingreso*	31	407.00	3.16	0.85	3.00	3.00	-0.74	-0.17
Vivienda*	32	407.00	3.43	0.78	4.00	3.00	-0.98	-0.44
MediaNotasGlobales*	33	407.00	5.38	1.43	6.00	7.00	-0.77	1.01
MediaPorcAsistencia*	34	407.00	3.78	0.53	4.00	4.00	-2.66	8.16
MediaIndiceAcademico*	35	407.00	1.82	0.40	2.00	2.00	-1.34	0.85
TiempoGraduacion*	36	407.00	1.35	0.56	1.00	2.00	1.36	0.85

Enfoque de extracción de reglas de asociación Semi-Supervisadas con apriori

El volumen de reglas que genera el algoritmo apriori es alto y para un experto humano es intelectualmente costosa la revisión individual de las reglas. De ahí que, el uso de las reglas de asociación con aprendizaje semi-supervisado ha consistido en establecer una variable de salida personalizada. Para este efecto, se ha usado la escala de tiempo de graduación, que ha sido calculada de la diferencia entre la fecha de graduación y la fecha de la primera matrícula (con independencia del título universitario). El resultado fue el tiempo tomado por el alumnado para conseguir la titulación universitaria que se ha descrito como éxito académico. De ahí que, se han asignado tres etiquetas con el tiempo de graduación: “MuyAlta” (superior a 10 años), “Alta” (entre 7 y 10 años) y “Baja” (menor a 7 años). Por consiguiente, las reglas que fueron generadas son filtradas a partir de esta variable que fue inducida para que apareciera solo en el consecuente de la regla.

A continuación, se ha generado usando el algoritmo *apriori* las reglas de asociación. Los parámetros usados para obtener un conjunto de reglas adecuado fueron confianza=0.86, Soporte =0.01 y el número de términos del antecedente sea cinco. En la Tabla ?? mostramos el conjunto de reglas obtenido y la valoración de las métricas. Las reglas encontradas se encuentran ordenadas de forma descendente según la columna conviction (Conv). De acuerdo con estos resultados, llama la atención las primeras cuatro reglas de asociación, donde el consecuente con la variable inducida ha demostrado que el tiempo de espera promedio para la graduación del estudiantado es “MuyAlta” y “Alta”.

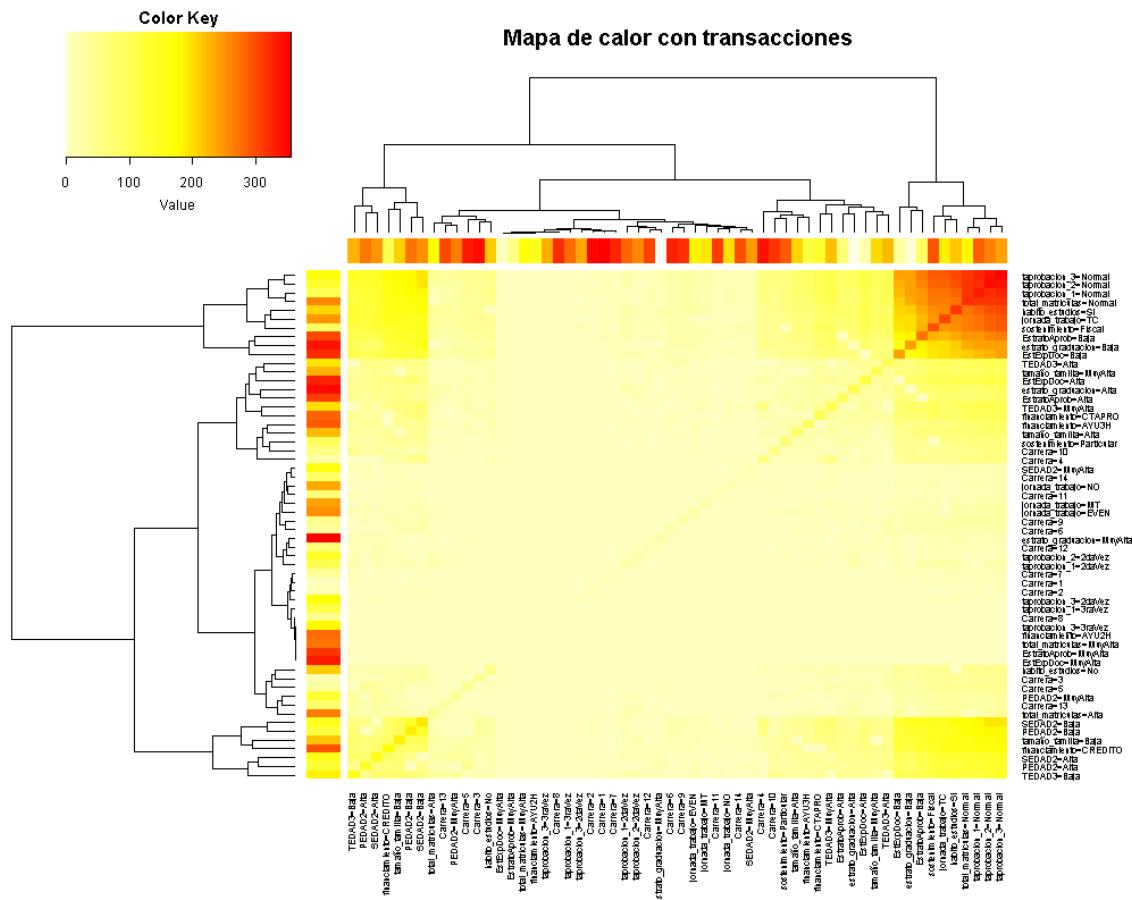


Figura 1 Mapa de calor con transacciones del conjunto de datos, ordenados según la jerarquía entre transacciones, el color mas oscuro (rojizo) ha indicado una mayor cantidad de reglas de esas categorías

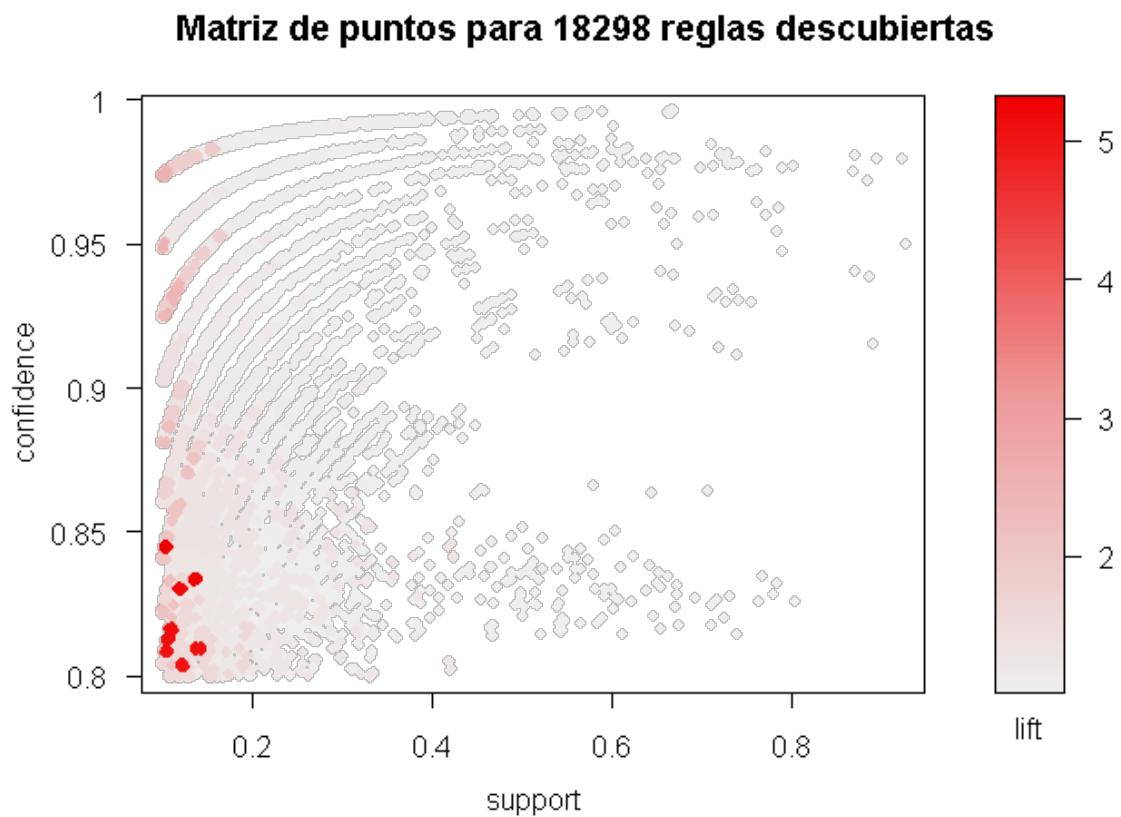


Figura 2 Matriz de puntos de la base de reglas de asociación obtenida, las reglas fueron filtradas de acuerdo con el soporte y confianza, la relevancia de regla es de color rojizo y fue asociada con la métrica lift.

B.2. Representación gráfica de las reglas

C.3. Reglas extraídas y métricas de calidad

Una de las primeras acciones a realizar ha sido extraer las reglas de asociación más representativas, en la tabla ?? presentamos las reglas que mejor representan al conjunto de datos.

Tabla 3 Reglas extraídas a partir del control del consecuente

Antecedente			Consecuente	Sop.	Conf.	Lift.	Conv.
taprobacion_1=2daVez	y	taproba-	⇒ Alta	0.049	0.947	3.448	13.780
cion_2=Normal							
sostenimiento=Particular	y SEDAD2=Alta	y ta-	⇒ Alta	0.030	0.917	3.337	8.703
probacion_3=Normal	y EstratoAprob=Alta						
Carrera=12	y SEDAD2=Alta	y TEDAD3=Baja	⇒ MuyAlta	0.019	0.875	15.925	7.560
Carrera=12	y TEDAD3=Baja	y financiamien-	⇒ MuyAlta	0.019	0.875	15.925	7.560
to=CREDITO							
taprobacion_1=Normal	y SEDAD2=Baja	y fi-	⇒ Baja	0.203	0.892	1.330	3.040
nanciamiento=CREDITO							
taprobacion_2=Normal	y SEDAD2=Baja	y TE-	⇒ Baja	0.187	0.883	1.317	2.821
DAD3=MuyAlta							
taprobacion_1=Normal	y PEDAD2=Baja	y SE-	⇒ Baja	0.228	0.874	1.303	2.610
DAD2=Baja	y jornada_trabajo=TC						
SEDAD2=Baja	y TEDAD3=MuyAlta		⇒ Baja	0.187	0.872	1.301	2.571
SEDAD2=Baja	y financiamiento=CREDITO		⇒ Baja	0.203	0.871	1.299	2.547
taprobacion_2=Normal	y taproba-	⇒ Baja	0.184	0.870	1.298	2.538	
cion_3=Normal	y TEDAD3=MuyAlta	+					
jornada_trabajo=TC							
sostenimiento=Fiscal	y taprobacion_1=Normal	⇒ Baja	0.236	0.869	1.296	2.511	
y PEDAD2=Baja	y SEDAD2=Baja						
taprobacion_1=Normal	y taproba-	⇒ Baja	0.225	0.863	1.288	2.409	
cion_2=Normal	y SEDAD2=Baja						

Enfoque de extracción de reglas de asociación Difusas

Las reglas de asociación difusas es otra de las técnicas empleadas para conseguir reglas de mejor calidad. Por el contrario del método anterior donde las variables son datos categóricos, ahora el enfoque es distinto y se convierten las variables a tipo numérico y después se transforman en un rango de datos entre $[0, 1]$. Como parte del estudio, se ha empleado el algoritmo “FuzzyApriori” del programa KEEL, el mismo que se ha configurado mediante combinaciones de valores entre el soporte y la confianza. Además consideramos tres regiones como partición difusa.

D.4. Representación gráfica de las reglas

Para conseguir una mejor visualización de las reglas de asociación obtenidas la hemos mostrado a través de la siguiente es rueda de dependencia de datos.

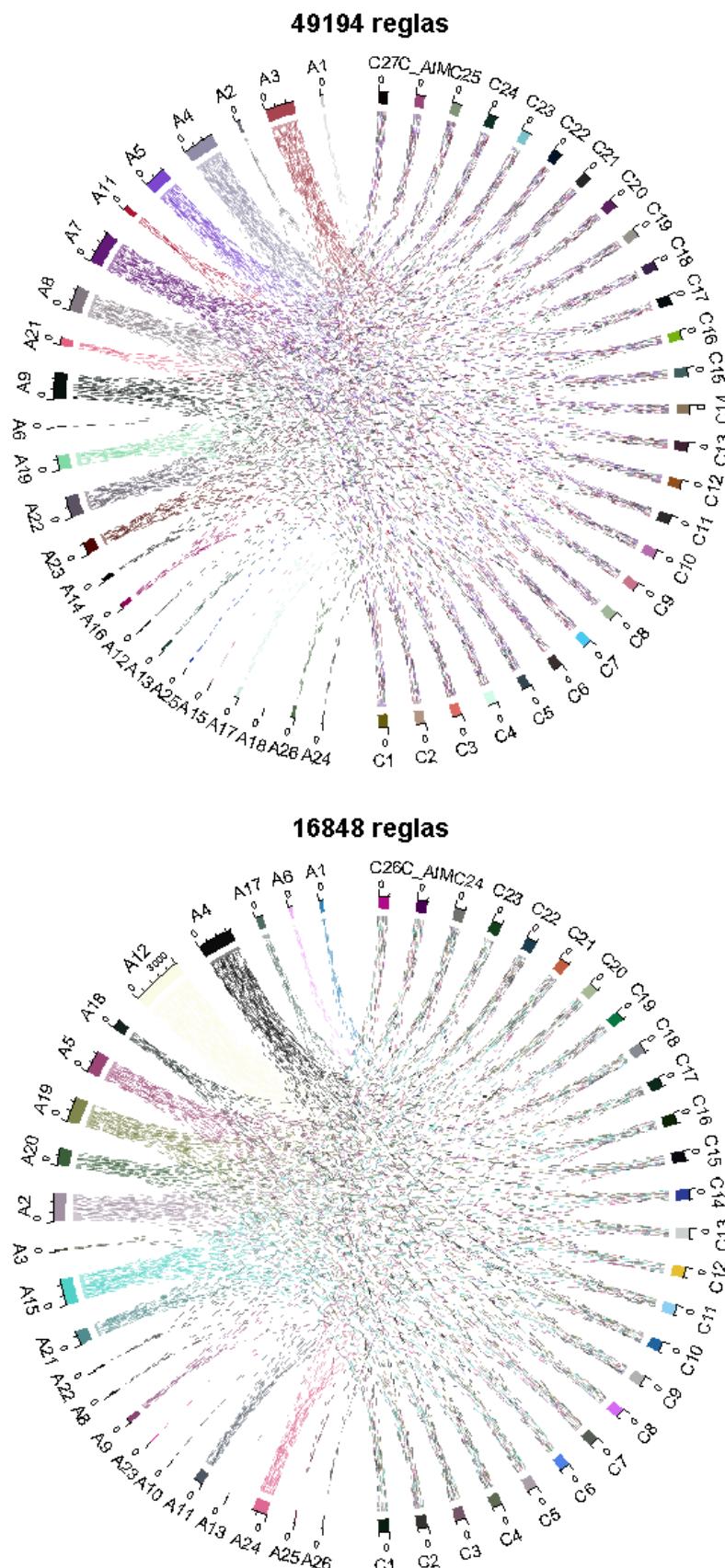


Figura 3 Soporte 0,3

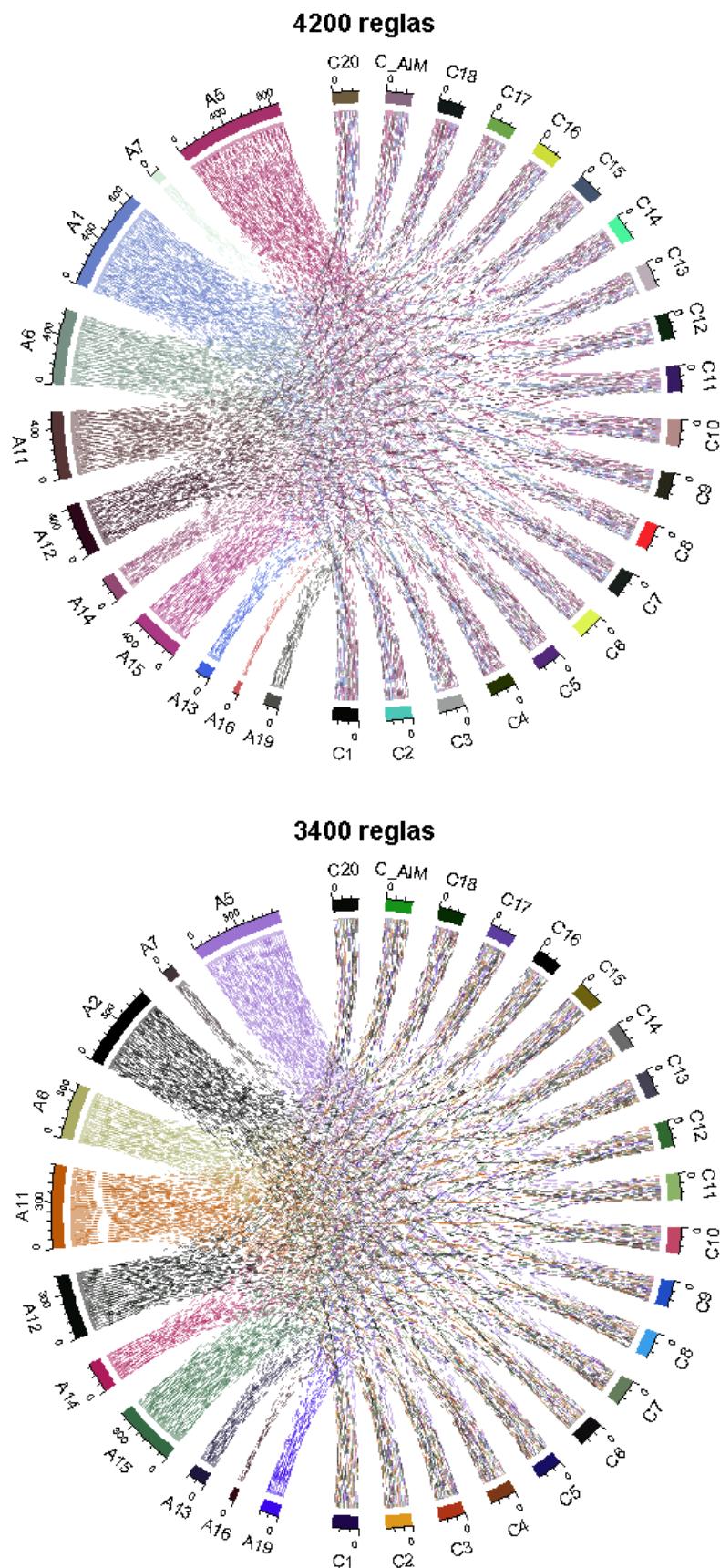


Figura 4 Soporte 0,4

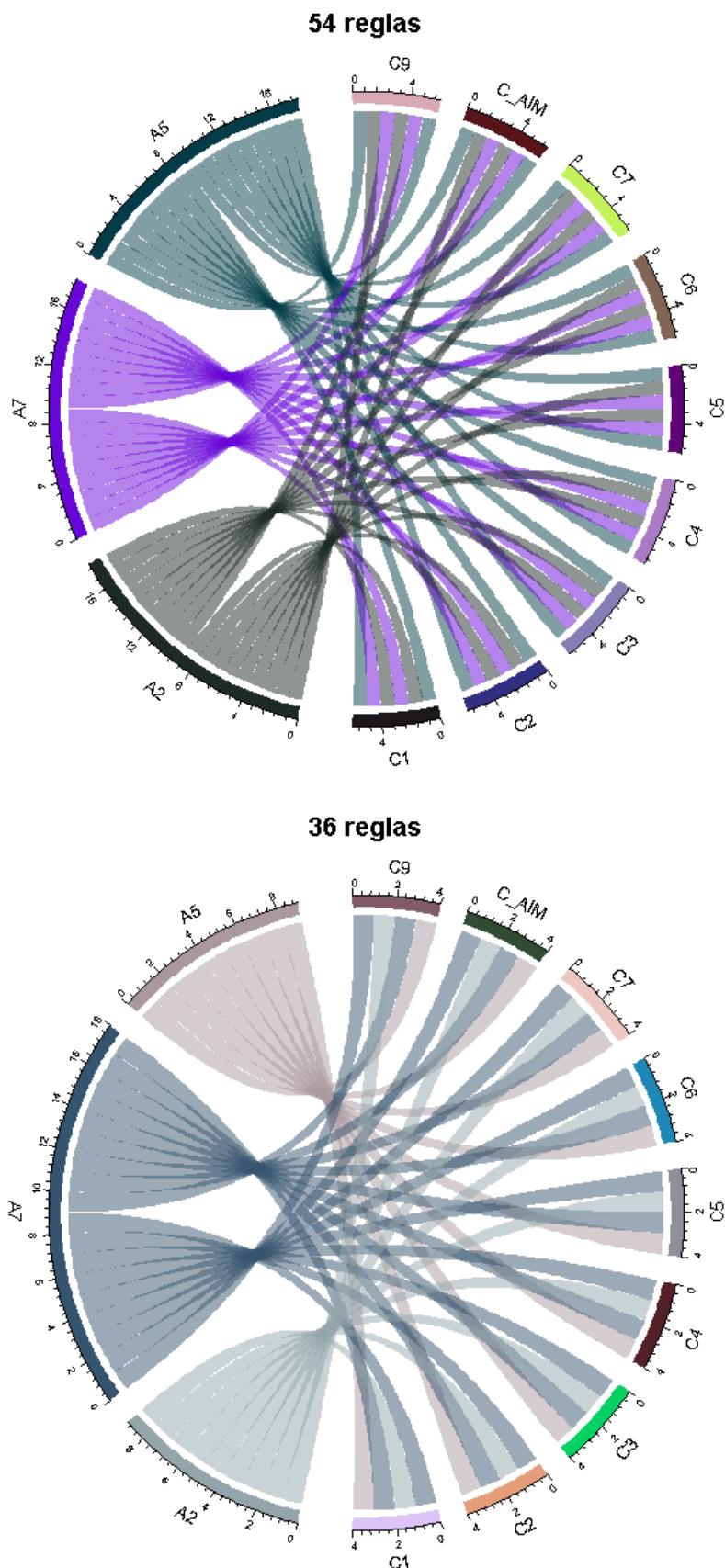


Figura 5 La representación de estas figuras han sido extraídas a partir de un soporte 0,5. La confianza entre 0,7 y 0,8 fueron las que mostraron un cambio en la cantidad de reglas. El resto de iteraciones han mostrado un volumen elevado de reglas de asociación.

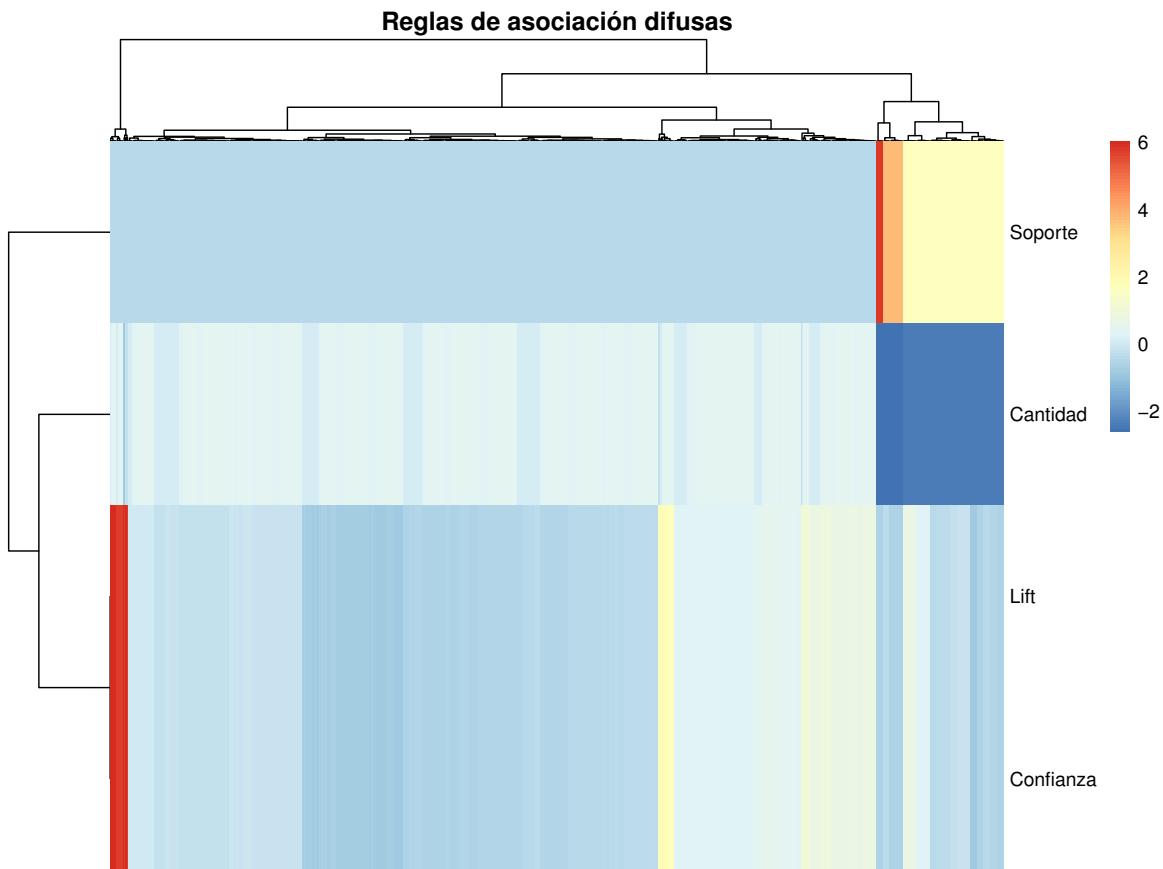


Figura 6 Mapa de calor con reglas de asociación difusas. Las métricas Lift y r.conviction muestran color rojizo en un pequeño segmento de la figura, éstas métricas indican el grado de interés de la regla.

E.5. Reglas extraídas y métricas de calidad

La aplicación de reglas de asociación difusas crearon 33465 reglas, de acuerdo con el criterio de filtrado finalmente se obtuvieron 106 reglas.

Tabla 4 Conjunto de reglas de asociación difusas obtenidas a partir del algoritmo FuzzyA-priori. Las reglas han sido filtradas de acuerdo con los criterios iniciales

No	Antecedente	Consecuente	Soporte	Confianza	Lift
1	NumHijos= L1 AND SEDAD2= L1	TiempoGraduacion=Bajo	0.2	0.8	1.35
2	SMediaEdad= L1 AND TEXPE- DOCEN= L1	TiempoGraduacion=Bajo	0.2	0.7	1.14
3	TEXPEDOCEN= L1 AND SMediaE- dad= L1	TiempoGraduacion=Bajo	0.2	0.6	1.14

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
4	NumHijos= L1 AND SMediaEdad= L1 AND TEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.7	1.13
5	FinanciamientoColegio= L1 AND PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.6-0.7	1.10
6	TEXPEDOCEN= L1 AND ViveCon= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.09
7	NumHijos= L1 AND FinanciamientoColegio= L1 AND PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.6	1.09
8	TEXPEDOCEN= L1 AND NumHijos= L1 AND ViveCon= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.09
9	NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.09
10		TiempoGraduacion=Bajo	0.2	0.1-0.6	1.08
11	TEXPEDOCEN= L1 AND NumHijos= L1	TiempoGraduacion=Bajo	0.3	0.6	1.08
12	TEXPEDOCEN= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.08
13	TEXPEDOCEN= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.08
14	MediaNotasAdmision= L1 AND FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.6	1.07
15	TEXPEDOCEN= L1 AND NumHijos= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.07
16		TiempoGraduacion=Bajo	0.2	0.1-0.6	1.07
17	NumHijos= L1 AND PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.07
18	TEXPEDOCEN= L1 AND NumHijos= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.07
19	TEXPEDOCEN= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.07
20	TEXPEDOCEN= L1 AND JornadaTrabajo= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
21	MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
22	FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
23	FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.3	0.1-0.6	1.06
24	MediaNotasAdmision= L1 AND PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
25	FinanciamientoColegio= L1 AND PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
26	TEXPEDOCEN= L1 AND JornadaTrabajo= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
27	TEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
28	TEXPEDOCEN= L1 AND JornadaTrabajo= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
29	TEXPEDOCEN= L1 AND NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
30	NumHijos= L1 AND FinanciamientoColegio= L1 AND PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
31	TEXPEDOCEN= L1 AND JornadaTrabajo= L1 AND NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
32	MediaNotasAdmision= L1 AND FinanciamientoColegio= L1 AND PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.06
33	JornadaTrabajo= L1 AND PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.05
34	NumHijos= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.6	1.04
35	NumHijos= L1 AND MediaNotasAdmision= L1 AND SMediaEdad= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.5	1.04
36	TMediaEdad= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
37	TMediaEdad= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
38	TMediaEdad= L1 AND JornadaTrabajo= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
39	PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
40	MediaNotasAdmision= L1 AND FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.6	1.04
41	PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
42	MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.3	0.1-0.6	1.04
43	TEXPEDOCEN= L1 AND TMediaEdad= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
44	TEXPEDOCEN= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
45	NumHijos= L1 AND FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
46	MediaNotasAdmision= L1 AND PE-DAD3= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.04
47	ViveCon= L1 AND FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
48	ViveCon= L1 AND SMediaEdad= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.5	1.03
49	TMediaEdad= L1 AND JornadaTrabajo= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
50	MediaNotasAdmision= L1 AND PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
51	TEXPEDOCEN= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
52	MediaNotasAdmision= L1 AND TMediaEdad= L2	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
53	TMediaEdad= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
54	TMediaEdad= L1 AND NumHijos= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
55	MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
56	NumHijos= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2-0.3	0.1-0.6	1.03
57	MediaNotasAdmision= L1 AND TMediaEdad= L2	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
58	NumHijos= L1 AND MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
59	TMediaEdad= L1 AND JornadaTrabajo= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
60	NumHijos= L1 AND PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
61	NumHijos= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2-0.3	0.1-0.6	1.03
62	TMediaEdad= L1 AND NumHijos= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
63	TMediaEdad= L1 AND NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
64	TMediaEdad= L1 AND ViveCon= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.03
65	TEXPEDOCEN= L1 AND TMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
66	SMediaEdad= L1	TiempoGraduacion=Bajo	0.2-0.3	0.1-0.6	1.02
67		TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
68	NumHijos= L1 AND MediaNotasAdmision= L1 AND PEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
69	TEXPEDOCEN= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
70	NumHijos= L1 AND ViveCon= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
71	TEXPEDOCEN= L1 AND TMediaEdad= L1 AND NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
72	ViveCon= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2-0.3	0.1-0.6	1.02
73	JornadaTrabajo= L1 AND MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
74	TMediaEdad= L1 AND JornadaTrabajo= L1 AND NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
75	TEXPEDOCEN= L1 AND TMediaEdad= L1 AND JornadaTrabajo= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
76	FinanciamientoColegio= L1 AND PEDAD3= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
77	FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
78	TMediaEdad= L1 AND NumHijos= L1 AND ViveCon= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
79	ViveCon= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
80	NumHijos= L1 AND PIndAcademico= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
81	NumHijos= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2-0.3	0.1-0.6	1.02
82	NumHijos= L1 AND ViveCon= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
83	UbicacionColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
84	TEXPEDOCEN= L1 AND TMediaEdad= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
85	TMediaEdad= L2 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
86	NumHijos= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2-0.4	0.1-0.6	1.02
87	ViveCon= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.02
88	JornadaTrabajo= L2	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01
89	ViveCon= L1 AND MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01
90	UbicacionColegio= L1 AND NumHijos= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01
91	JornadaTrabajo= L1 AND SEDAD3= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01
92	JornadaTrabajo= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01
93	JornadaTrabajo= L1 AND PEDAD3= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01
94	MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.01

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
95	SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
96	TMediaEdad= L1 AND PIndAcademico= L2	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
97	JornadaTrabajo= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
98	MediaNotasAdmision= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
99	NumHijos= L1 AND ViveCon= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
100	JornadaTrabajo= L1 AND NumHijos= L1 AND MediaNotasAdmision= L1 AND FinanciamientoColegio= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
101	JornadaTrabajo= L1 AND FinanciamientoColegio= L1 AND SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
102	SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
103	SMediaEdad= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
104	UbicacionColegio= L1 AND NumHijos= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00
105	JornadaTrabajo= L1 AND MediaNotasAdmision= L1	TiempoGraduacion=Bajo	0.2-0.3	0.1-0.6	1.00
106	JornadaTrabajo= L2 AND NumHijos= L1	TiempoGraduacion=Bajo	0.2	0.1-0.6	1.00

Términos lingüísticos

A continuación se presenta los términos lingüísticos que están representados en la base de reglas difusas obtenidas (Véase 4).

No	Etiquetas	termino
1	FinanciamientoColegio=0.04166664499999995;0.08333333;0.125000015	L1
2	JornadaTrabajo=0.0;0.125;0.25	L1
3	JornadaTrabajo=0.125;0.25;0.375	L2
4	MediaNotasAdmision=0.0;0.29166665;0.5833333	L1
5	NumHijos=-0.16666665;0.0;0.16666665	L1
6	PEDAD3=0.0;0.125;0.25	L1
7	PEXPEDOCEN=0.16666665;0.333333;0.4999999499999997	L1
8	PIndAcademico=-0.16666665;0.0;0.16666665	L1
9	PIndAcademico=0.0;0.16666665;0.3333333	L2
10	SEDAD2=-0.125;0.0;0.125	L1
11	SEDAD3=0.0;0.125;0.25	L1
12	SMediaEdad=0.41666665;0.8333333;1.24999995	L1
13	TEXPEDOCEN=0.0;0.16666665;0.3333333	L1
14	TMediaEdad=0.0;0.16666665;0.3333333	L1
15	TMediaEdad=0.16666665;0.3333333;0.4999999499999997	L2
16	UbicacionColegio=-0.08333335;0.0;0.08333335	L1
17	ViveCon=0.0;0.20833335;0.4166667	L1

Enfoque de extracción de reglas de asociación difusas y algoritmo genético

F.6. Cruce

Para el proceso de extracción de reglas se han considerado tres tipos de probabilidades de cruce entre individuos: 10 %, 50 % y 80 % respectivamente. Esto, con el fin de conseguir un conjunto de reglas de asociación de mejor calidad.

F.6.1. Probabilidad de cruce 10 %

Las reglas de asociación difusas extraídas, están presentadas por medio de ruedas de dependencia. El consecuente se ha identificado con la letra C, mientras que, el objetivo de extraer las reglas que tengan relación con el tiempo promedio de graduación de los estudiantes se encuentra con la etiqueta “C_AIM”.

Representación gráfica de las reglas

Conjunto de reglas extraídas y métricas de calidad

Tabla 5 Conjunto de reglas de asociación difusas genéticas (probabilidad de cruce 10 %)

No	Antecedente	Consecuente	Soporte	Confianza	Lift
1	SMediaEdad= L5 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,32
2	JornadaTrabajo= L6 AND SMediaEdad= L5 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,31
3	JornadaTrabajo= L6 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,31

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
4	TMediaEdad= L6 AND PIndAcademico= L2 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,31
5	JornadaTrabajo= L6 AND SEDAD2= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,31
6	SEDAD2= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,30
7	MediaNotasAdmision= L6 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,30
8	JornadaTrabajo= L6 AND NumHijos= L2 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,30
9	SMediaEdad= L5 AND NumHijos= L2 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,30
10	JornadaTrabajo= L6 AND SMediaEdad= L5 AND NumHijos= L2 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,30
11	NumHijos= L2 AND SEDAD2= L2	TiempoGraduacion=Bajo	0,10	0.10	1,30
12	SMediaEdad= L5 AND PIndAcademico= L2 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,27
13	SEDAD2= L3 AND SMediaEdad= L7	TiempoGraduacion=Bajo	0,20	0.10	1,26
14	PIndAcademico= L2 AND ViveCon= L5 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,26
15	JornadaTrabajo= L4 AND SEDAD2= L3 AND SMediaEdad= L7	TiempoGraduacion=Bajo	0,20	0.10	1,25
16	JornadaTrabajo= L6 AND PIndAcademico= L2 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,25
17	SMediaEdad= L5 AND TEXPEDOCEN= L3 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,24
18	SEDAD2= L3 AND ViveCon= L6	TiempoGraduacion=Bajo	0,20	0.10	1,24
19	SMediaEdad= L3 AND SEDAD2= L3	TiempoGraduacion=Bajo	0,20	0.10	1,23
20	TEXPEDOCEN= L3 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,23
21	JornadaTrabajo= L4 AND SEDAD2= L3	TiempoGraduacion=Bajo	0,20	0.10	1,22
22	JornadaTrabajo= L6 AND ViveCon= L5 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,22
23	JornadaTrabajo= L4 AND SMediaEdad= L3 AND SEDAD2= L3	TiempoGraduacion=Bajo	0,20	0.10	1,22

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
24	PIndAcademico= L2 AND TEXPE- DOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,22
25	SMediaEdad= L5 AND ViveCon= L5 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,21
26	JornadaTrabajo= L6 AND SMediaE- dad= L5 AND TEXPEDOCEN= L3 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,21
27	ViveCon= L5 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,21
28	SEDAD2= L3 AND MediaNotasAdmi- sion= L4	TiempoGraduacion=Bajo	0,20	0.10	1,21
29	SEDAD2= L3 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,20
30	JornadaTrabajo= L4 AND SEDAD2= L3 AND MediaNotasAdmision= L4	TiempoGraduacion=Bajo	0,20	0.10	1,19
31	JornadaTrabajo= L6 AND TEXPE- DOCEN= L3 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,19
32	JornadaTrabajo= L6 AND SEDAD3= L5	TiempoGraduacion=Bajo	0,10	0.10	1,18
33	MediaNotasAdmision= L6 AND SMe- diaEdad= L5 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,17
34	SMediaEdad= L5 AND TEXPE- DOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,15
35	JornadaTrabajo= L6 AND SMediaE- dad= L5 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,14
36	SMediaEdad= L7 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,13
37	JornadaTrabajo= L6 AND TEXPE- DOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,13
38	MediaNotasAdmision= L6 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,13
39	PEDAD3= L3 AND MediaNotasAdmi- sion= L6	TiempoGraduacion=Bajo	0,10	0.10	1,12
40	MediaNotasAdmision= L6 AND PIndA- cademico= L2 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,12
41	MediaNotasAdmision= L6 AND TMe- diaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,11

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
42	JornadaTrabajo= L6 AND MediaNotasAdmision= L6 AND SMediaEdad= L5 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,11
43	MediaNotasAdmision= L6 AND SMediaEdad= L5 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,11
44	JornadaTrabajo= L6 AND MediaNotasAdmision= L6 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,10
45	TIndAcademico= L4 AND TMediaEdad= L6	TiempoGraduacion=Bajo	0,10	0.10	1,10
46	MediaNotasAdmision= L4 AND SMediaEdad= L7	TiempoGraduacion=Bajo	0,20	0.10	1,10
47	JornadaTrabajo= L6 AND PEDAD3=L3	TiempoGraduacion=Bajo	0,10	0.10	1,09
48	MediaNotasAdmision= L6 AND ViveCon= L10	TiempoGraduacion=Bajo	0,10	0.10	1,08
49	MediaNotasAdmision= L4 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,08
50	SMediaEdad= L3 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,08
51	JornadaTrabajo= L4 AND MediaNotasAdmision= L4 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,08
52	TMediaEdad= L6 AND ViveCon= L5 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,07
53	SMediaEdad= L3 AND TEXPEDOCEN= L2	TiempoGraduacion=Bajo	0,20	0.10	1,07
54	JornadaTrabajo= L4 AND SMediaEdad= L7	TiempoGraduacion=Bajo	0,20	0.10	1,07
55	JornadaTrabajo= L4 AND MediaNotasAdmision= L4 AND SMediaEdad= L7	TiempoGraduacion=Bajo	0,20	0.10	1,07
56	JornadaTrabajo= L4 AND SMediaEdad= L3 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,07
57	TMediaEdad= L6 AND JornadaTrabajo= L6 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,07
58	TMediaEdad= L6 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
59	SMediaEdad= L3 AND TEXPEDOCEN= L2 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,06

Continúa en la siguiente página...

No	Antecedente	Consecuente	<i>Soporte</i>	Confianza	<i>Lift</i>
60	TMediaEdad= L6 AND JornadaTrabajo= L6 AND SMediaEdad= L5 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
61	MediaNotasAdmision= L6 AND FinanciamientoColegio= L3	TiempoGraduacion=Bajo	0,10	0.10	1,06
62	JornadaTrabajo= L6 AND MediaNotasAdmision= L6 AND FinanciamientoColegio= L3	TiempoGraduacion=Bajo	0,10	0.10	1,06
63	TMediaEdad= L6 AND SMediaEdad= L5 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
64	JornadaTrabajo= L4 AND SMediaEdad= L3 AND TEXPEDOCEN= L2	TiempoGraduacion=Bajo	0,20	0.10	1,06
65	TMediaEdad= L6 AND NumHijos= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
66	TMediaEdad= L6 AND JornadaTrabajo= L6 AND NumHijos= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
67	PIndAcademico= L2 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
68	PIndAcademico= L2 AND JornadaTrabajo= L6 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
69	JornadaTrabajo= L6 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,06
70	ViveCon= L6 AND SMediaEdad= L7	TiempoGraduacion=Bajo	0,20	0.10	1,05
71	NumHijos= L2 AND MediaNotasAdmision= L6 AND FinanciamientoColegio= L3	TiempoGraduacion=Bajo	0,10	0.10	1,05
72	JornadaTrabajo= L6 AND SEDAD2= L6 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,05
73	TIndAcademico= L4 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,05
74	TMediaEdad= L6 AND PIndAcademico= L2 AND ViveCon= L5	TiempoGraduacion=Bajo	0,10	0.10	1,05
75	TMediaEdad= L6 AND SEDAD2= L6 AND NumHijos= L2	TiempoGraduacion=Bajo	0,10	0.10	1,05
76	TMediaEdad= L6 AND PIndAcademico= L2	TiempoGraduacion=Bajo	0,10	0.10	1,05

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
77	ViveCon= L6 AND TMediaEdad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,05
78	TEXPEDOCEN= L2 AND TMediaE- dad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,05
79	JornadaTrabajo= L4 AND TMediaE- dad= L4	TiempoGraduacion=Bajo	0,20	0.10	1,04
80	TMediaEdad= L6 AND JornadaTra- bajo= L6 AND SEDAD2= L6 AND NumHijos= L2	TiempoGraduacion=Bajo	0,10	0.10	1,04
81	TMediaEdad= L6 AND JornadaTraba- jo= L6 AND MediaNotasAdmision= L8	TiempoGraduacion=Bajo	0,10	0.10	1,04
82	TMediaEdad= L6 AND MediaNotasAd- mision= L8	TiempoGraduacion=Bajo	0,10	0.10	1,04
83	TIndAcademico= L4 AND MediaNota- sAdmision= L8	TiempoGraduacion=Bajo	0,10	0.10	1,04
84	JornadaTrabajo= L6 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,04
85	TMediaEdad= L6 AND ViveCon= L5	TiempoGraduacion=Bajo	0,10	0.10	1,04
86	SEDAD2= L6 AND MediaNotasAdmi- sion= L6	TiempoGraduacion=Bajo	0,10	0.10	1,04
87	SMediaEdad= L5 AND SEDAD2= L6 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,04
88	TIndAcademico= L4 AND TMediaE- dad= L6 AND JornadaTrabajo= L6	TiempoGraduacion=Bajo	0,10	0.10	1,04
89	MediaNotasAdmision= L4 AND SMe- diaEdad= L3	TiempoGraduacion=Bajo	0,20	0.10	1,04
90	JornadaTrabajo= L4 AND TEXPE- DOCEN= L2	TiempoGraduacion=Bajo	0,20	0.10	1,04
91	TMediaEdad= L6 AND SMediaEdad= L5	TiempoGraduacion=Bajo	0,10	0.10	1,04
92	PIndAcademico= L2 AND MediaNota- sAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,04
93	NumHijos= L2 AND TEXPEDOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,04
94	PEDAD3= L3 AND NumHijos= L2	TiempoGraduacion=Bajo	0,10	0.10	1,03
95	JornadaTrabajo= L6 AND SMediaE- dad= L5 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03
96	SMediaEdad= L5 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
97	PIndAcademico= L2 AND JornadaTrabajo= L6 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03
98	TMediaEdad= L6 AND JornadaTrabajo= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03
99	JornadaTrabajo= L6 AND NumHijos= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03
100	SMediaEdad= L5 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03
101	TIndAcademico= L4 AND TMediaEdad= L6 AND SMediaEdad= L5	TiempoGraduacion=Bajo	0,10	0.10	1,03
102	JornadaTrabajo= L6 AND ViveCon= L5 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,03
103	TMediaEdad= L6 AND JornadaTrabajo= L6 AND SMediaEdad= L5	TiempoGraduacion=Bajo	0,10	0.10	1,02
104	MediaNotasAdmision= L6 AND Financiamiento= L4	TiempoGraduacion=Bajo	0,10	0.10	1,02
105	JornadaTrabajo= L4 AND MediaNotasAdmision= L4	TiempoGraduacion=Bajo	0,20	0.10	1,02
106	JornadaTrabajo= L6 AND SMediaEdad= L5 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,02
107	SEDAD2= L6 AND NumHijos= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,02
108	NumHijos= L2 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,02
109	TMediaEdad= L6 AND ViveCon= L5 AND SMediaEdad= L5	TiempoGraduacion=Bajo	0,10	0.10	1,02
110	TMediaEdad= L6 AND PIndAcademico= L2 AND SMediaEdad= L5	TiempoGraduacion=Bajo	0,10	0.10	1,02
111	ViveCon= L5 AND SEDAD2= L6	TiempoGraduacion=Bajo	0,10	0.10	1,02
112	JornadaTrabajo= L6 AND MediaNotasAdmision= L6 AND Financiamiento= L4	TiempoGraduacion=Bajo	0,10	0.10	1,01
113	JornadaTrabajo= L4 AND MediaNotasAdmision= L4 AND SMediaEdad= L3	TiempoGraduacion=Bajo	0,20	0.10	1,01

Continúa en la siguiente página...

No	Antecedente	Consecuente	Soporte	Confianza	Lift
114	TMediaEdad= L6 AND JornadaTrabajo= L6 AND ViveCon= L5	TiempoGraduacion=Bajo	0,10	0.10	1,01
115	JornadaTrabajo= L6 AND PEXPE- DOCEN= L3	TiempoGraduacion=Bajo	0,10	0.10	1,01
116	JornadaTrabajo= L6 AND SEDAD2= L6 AND NumHijos= L2	TiempoGraduacion=Bajo	0,10	0.10	1,01
117	JornadaTrabajo= L6 AND MediaNota- sAdmision= L8	TiempoGraduacion=Bajo	0,10	0.10	1,01
118	TMediaEdad= L6 AND PIndAcademico= L2 AND JornadaTrabajo= L6	TiempoGraduacion=Bajo	0,10	0.10	1,01
119	TMediaEdad= L6 AND JornadaTrabajo= L6 AND ViveCon= L5 AND SMediaEdad= L5	TiempoGraduacion=Bajo	0,10	0.10	1,00
120	JornadaTrabajo= L6 AND ViveCon= L5 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,00
121	SEDAD2= L6 AND NumHijos= L2	TiempoGraduacion=Bajo	0,10	0.10	1,00
122	PIndAcademico= L2 AND SMediaEdad= L5 AND MediaNotasAdmision= L6	TiempoGraduacion=Bajo	0,10	0.10	1,00
123	SMediaEdad= L5 AND SEDAD2= L6 AND NumHijos= L2	TiempoGraduacion=Bajo	0,10	0.10	1,00
124	ViveCon= L6 AND MediaNotasAdmision= L4	TiempoGraduacion=Bajo	0,20	0.10	1,00
125	JornadaTrabajo= L6 AND TMediaEdad= L2	TiempoGraduacion=Bajo	0,10	0.10	1,00

Términos lingüísticos

Tabla 6 Terminos ligüísticos de conjunto de reglas de asociación (probabilidad de cruce 10%)

No	Etiqueta	Término
1	EstiloAprendizaje=-0.054236031603497034;0.0485434808139593;0.15132299323141563	L1
2	EstiloAprendizaje=-8.374142390150108E-4;0.1019420981784406;0.20472161059589622	L2
3	EstiloAprendizaje=0.05112307472771932;0.11121689320975925;0.17131071169179918	L3
4	EstiloAprendizaje=0.12381324213164885;0.18364311787052545;0.24347299360940206	L4
5	EstiloAprendizaje=6.960615736162495E-4;0.06074631281258919;0.12079656405156212	L5

Continúa en la siguiente página...

No	Etiqueta	Término
6	Financiamiento=0.045807624015082156;0.09030442669910609;0.13480122938313002	L1
7	Financiamiento=0.06820556910521383;0.12541583227989975;0.18262609545458566	L2
8	Financiamiento=0.10206162125469971;0.15927188442938486;0.21648214760407003	L3
9	Financiamiento=0.10310958299346742;0.1476063856774798;0.19210318836149218	L4
10	Financiamiento=0.10787076170550966;0.16721380937460195;0.22655685704369424	L5
11	FinanciamientoColegio=0.06915718292620673;0.09740458648528333;0.12565199004435992	L1
12	FinanciamientoColegio=0.07399488313759148;0.10700855413565125;0.14002222513371101	L2
13	FinanciamientoColegio=0.07788578320574582;0.09879636990925134;0.11970695661275685	L3
14	JornadaTrabajo=0.033557537446326446;0.10662261394766709;0.17968769044900773	L1
15	JornadaTrabajo=0.045001327780100644;0.15147521907541553;0.25794911037073043	L2
16	JornadaTrabajo=0.05250660294514295;0.13834388958674249;0.22418117622834202	L3
17	JornadaTrabajo=0.060880760284939384;0.15898314316028334;0.2570855260356273	L4
18	JornadaTrabajo=0.1019724123988298;0.20844630369414469;0.31492019498945956	L5
19	JornadaTrabajo=0.12017260404131065;0.1932376805426534;0.26630275704399614	L6
20	JornadaTrabajo=0.13762577924476813;0.22346306588636766;0.30930035252796717	L7
21	JornadaTrabajo=0.1385259033774036;0.23662828625274823;0.3347306691280929	L8
22	MediaNotasAdmision=-0.03300946574845315;0.21512936469913396;0.4632681951467211	L1
23	MediaNotasAdmision=-0.03421316126685847;0.12829518100044535;0.29080352326774916	L2
24	MediaNotasAdmision=0.007666150115261783;0.19881692802247042;0.38996770592967905	L3
25	MediaNotasAdmision=0.02511613724072026;0.25075911546379903;0.4764020936868778	L4
26	MediaNotasAdmision=0.1185845197748725;0.3667233502224596;0.6148621806700467	L5
27	MediaNotasAdmision=0.12821990173030637;0.29072824399758834;0.4532365862648703	L6
28	MediaNotasAdmision=0.16407874519708648;0.3552295231042951;0.5463803010115038	L7
29	MediaNotasAdmision=0.23589805709645156;0.39840639936374433;0.5609147416310372	L8
30	MediaNotasAdmision=0.2567090837753573;0.4823520619984363;0.7079950402215153	L9
31	NumHijos=-0.015889605485694597;0.1111783013678444;0.2382462082213834	L1
32	NumHijos=-0.03295261693895295;0.0631127329724486;0.15917808288385016	L2
33	NumHijos=-0.054358165029657374;0.07101921261354066;0.1963965902567387	L3
34	PEDAD3=-0.0013909905308872594;0.06667968963606002;0.1347503698030073	L1
35	PEDAD3=-0.009612450142572601;0.0805037914043782;0.170620032951329	L2
36	PEDAD3=0.06616686898204524;0.13423754914898237;0.2023082293159195	L3
37	PEDAD3=0.07421013972756449;0.16432638127451663;0.2544426228214688	L4
38	PEDAD3=0.1346387923280104;0.20270947249495241;0.2707801526618944	L5
39	PEXPEDOCEN=0.05067391097534052;0.19040859311751684;0.33014327525969317	L1
40	PEXPEDOCEN=0.06404484190928059;0.16419058999139222;0.2643363380735039	L2
41	PEXPEDOCEN=0.10154767337544528;0.18014238170411506;0.25873709003278483	L3
42	PEXPEDOCEN=0.14480215538690583;0.24410679858403583;0.34341144178116584	L4

Continúa en la siguiente página...

No	Etiqueta	Término
43	PEXPEDOCEN=0.15823352373293526;0.2979682058751116;0.4377028880172879	L5
44	PEXPEDOCEN=0.1641907716119052;0.26433651969401684;0.36448226777612847	L6
45	PEXPEDOCEN=0.1811071955740166;0.25970190390268655;0.3382966122313565	L7
46	PIndAcademico=-0.00293000853534596;0.10673197361909997;0.21639395577354592	L1
47	PIndAcademico=-0.02068546784707133;0.07379145132498213;0.16826837049703558	L2
48	PIndAcademico=-0.03231484965425212;0.09313894957328193;0.21859274880081597	L3
49	PIndAcademico=-0.07359099611493974;0.06968253299374255;0.21295606210242485	L4
50	PIndAcademico=0.08455731892257692;0.1790342380946224;0.27351115726666786	L5
51	PIndAcademico=0.19411732658845213;0.28859424576050197;0.3830711649325518	L6
52	SEADAD2=-0.011703627306461031;0.07214300222772851;0.15598963176191805	L1
53	SEADAD2=-0.014826692026975605;0.045861859130333534;0.10655041028764267	L2
54	SEADAD2=-0.04149116159612873;0.06678388463010912;0.17505893085634697	L3
55	SEADAD2=-0.04265419364134436;0.0741173890216192;0.19088897168458274	L4
56	SEADAD2=0.010651913722193374;0.12742349638515693;0.2441950790481205	L5
57	SEADAD2=0.05716011209549529;0.11784866325279661;0.17853721441009793	L6
58	SEADAD2=0.060636422505982945;0.1689114687322208;0.27718651495845864	L7
59	SEADAD2=0.07816795671120431;0.16201458624539386;0.24586121577958342	L8
60	SEADAD2=0.1130490529618361;0.22132409918804802;0.32959914541425994	L9
61	SEADAD2=0.14707292502077143;0.2077614761780765;0.2684500273353816	L10
62	SEADAD3=-0.08676306372978174;0.02004426972268631;0.12685160317515437	L1
63	SEADAD3=0.005505957879056611;0.06077306639040878;0.11604017490176095	L2
64	SEADAD3=0.020175579765195364;0.12698291321766342;0.23379024667013149	L3
65	SEADAD3=0.07370746351699642;0.1289745720282288;0.18424168053946116	L4
66	SEADAD3=0.1346228759508886;0.18988998446217686;0.24515709297346513	L5
67	SMediaEdad=0.19425355346843373;0.43498170930543445;0.6757098651424351	L1
68	SMediaEdad=0.21694757289867983;0.5708340439639696;0.9247205150292594	L2
69	SMediaEdad=0.3603548930693925;0.7180891051584515;1.0758233172475105	L3
70	SMediaEdad=0.41406156819406736;0.7679480392593572;1.121834510324647	L4
71	SMediaEdad=0.4220863196953057;0.6628144755323123;0.9035426313693189	L5
72	SMediaEdad=0.4812328939067335;0.7423235568715296;1.0034142198363258	L6
73	SMediaEdad=0.4998058450010526;0.8575400570901;1.2152742691791474	L7
74	TEDAD1=-0.01002575074151351;0.049370188692298284;0.10876612812611008	L1
75	TEDAD1=-0.030324819373886078;0.05646733372555973;0.14325948682500556	L2
76	TEDAD1=-0.046221476768834126;0.03991153238291375;0.12604454153466163	L3
77	TEDAD1=-0.06910361579293492;0.042260278105218735;0.1536241720033724	L4
78	TEXPEDOCEN=-0.06743429703095385;0.06870955705652812;0.2048534111440101	L1
79	TEXPEDOCEN=0.08102849136753035;0.21717234545501146;0.3533161995424926	L2

Continúa en la siguiente página... 2

No	Etiqueta	Término
80	TEXPEDOCEN=0.08622496205544336;0.1778137407546481;0.26940251945385285	L3
81	TEXPEDOCEN=0.15937824399355216;0.29552209808103236;0.43166595216851256	L4
82	TEXPEDOCEN=0.17589526450340265;0.26748404320279984;0.359072821902197	L5
83	TiempoGraduacion=Alto	Alto
84	TiempoGraduacion=Bajo	Bajo
85	TiempoGraduacion=Medio	Medio
86	TIndAcademico=-0.006640302474616083;0.06913847875559287;0.14491725998580182	L1
87	TIndAcademico=-0.022491365497120455;0.04535667664970364;0.11320471879652773	L2
88	TIndAcademico=0.0011770448943282802;0.06308249108288866;0.12498793727144904	L3
89	TIndAcademico=0.03050616429466063;0.07729541280417385;0.12408466131368706	L4
90	TIndAcademico=0.03171777537699596;0.09362322372625997;0.15552866991482034	L5
91	TIndAcademico=0.052951404529589435;0.12079944667641353;0.18864748882323762	L6
92	TIndAcademico=0.10706553528214365;0.15385478379166184;0.20064403230118005	L7
93	TMediaEdad=-0.018458983770684778;0.10833334491995822;0.23512567361060122	L1
94	TMediaEdad=0.028324789396237948;0.1326927162548666;0.23706064311349523	L2
95	TMediaEdad=0.0959615218341123;0.21404019845189837;0.3321188750696844	L3
96	TMediaEdad=0.11068161402546678;0.23747394271611127;0.36426627140675577	L4
97	TMediaEdad=0.11887168825153793;0.25271781514311425;0.3865639420346906	L5
98	TMediaEdad=0.15993356121296742;0.2643014880716037;0.36866941493023997	L6
99	TMediaEdad=0.16781700806665975;0.2858956846844458;0.40397436130223185	L7
100	TMediaEdad=0.20391767833980545;0.33071000703044995;0.45750233572109444	L8
101	UbicacionColegio=-0.004944587324277223;0.05016750810076831;0.10527960352581384	L1
102	UbicacionColegio=0.0369018467501611;0.08137656631792954;0.125851285885698	L2
103	ViveCon=-0.0015163784325823992;0.14793451020030085;0.2973853988331841	L1
104	ViveCon=-0.027570454596963223;0.09910522710637928;0.2257809088097218	L2
105	ViveCon=0.08740236237308005;0.26993109157775474;0.4524598207824294	L3
106	ViveCon=0.10353475903318674;0.22453882399513386;0.34554288895708096	L4
107	ViveCon=0.1039970377516034;0.2306727194549515;0.3573484011582996	L5
108	ViveCon=0.1475907179639172;0.29704160659680046;0.4464924952296837	L6
109	ViveCon=0.15739495724946304;0.33992368645413773;0.5224524156588124	L7
110	ViveCon=0.1970891014305712;0.31809316639251833;0.43909723135446543	L8
111	ViveCon=0.22117100977226845;0.3706218984051365;0.5200727870380045	L9
112	ViveCon=0.23879367578722865;0.36546935749057785;0.49214503919392705	L10

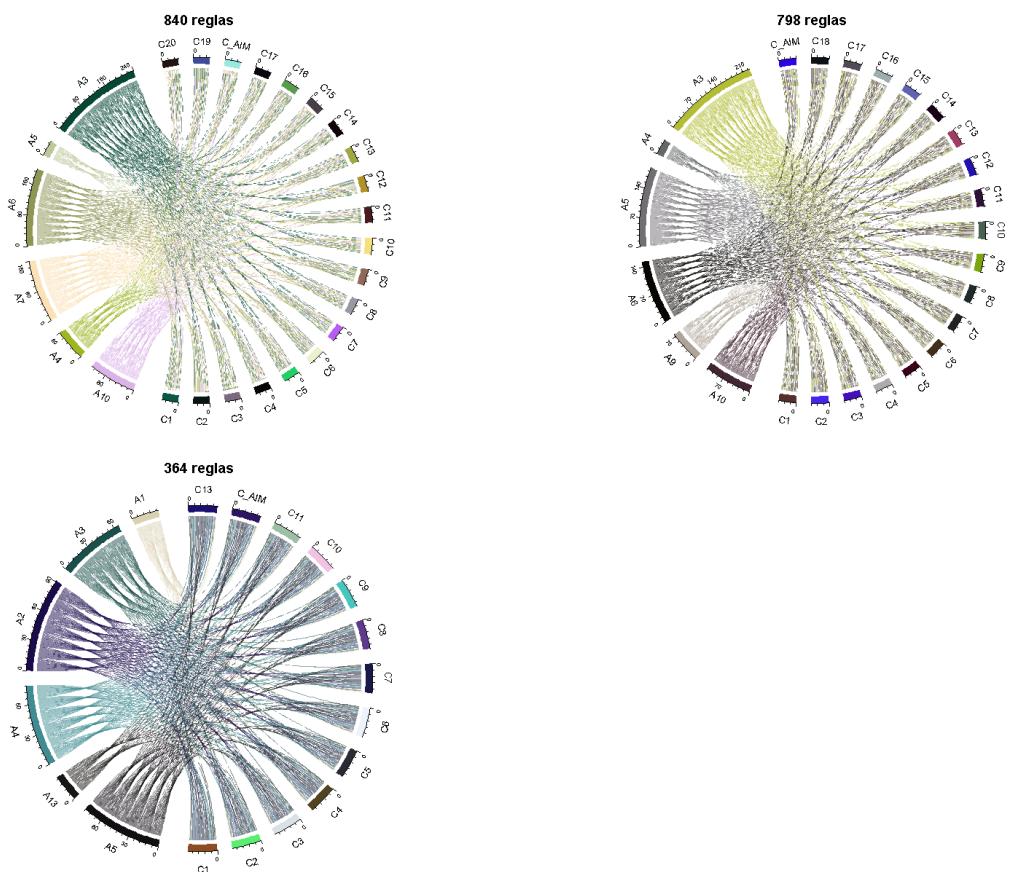


Figura 7 soporte 3%

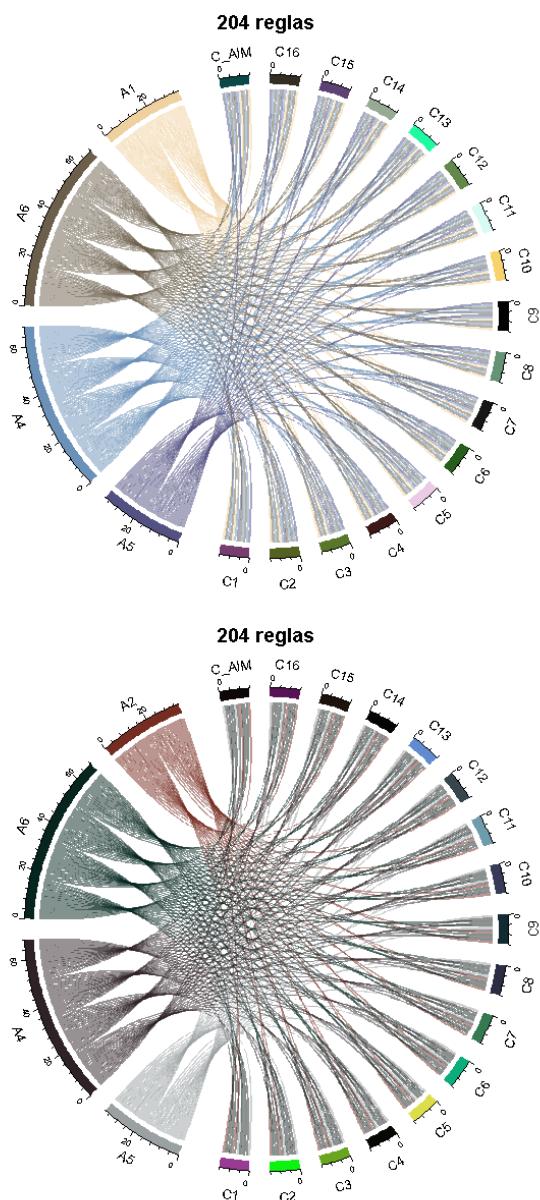


Figura 8 soporte 0,4 % y confianza 0,5 % ambas figuras tienen el mismo número de reglas, no obstante, su diferencia se ha definido por el atributo, A1 y A2 respectivamente.

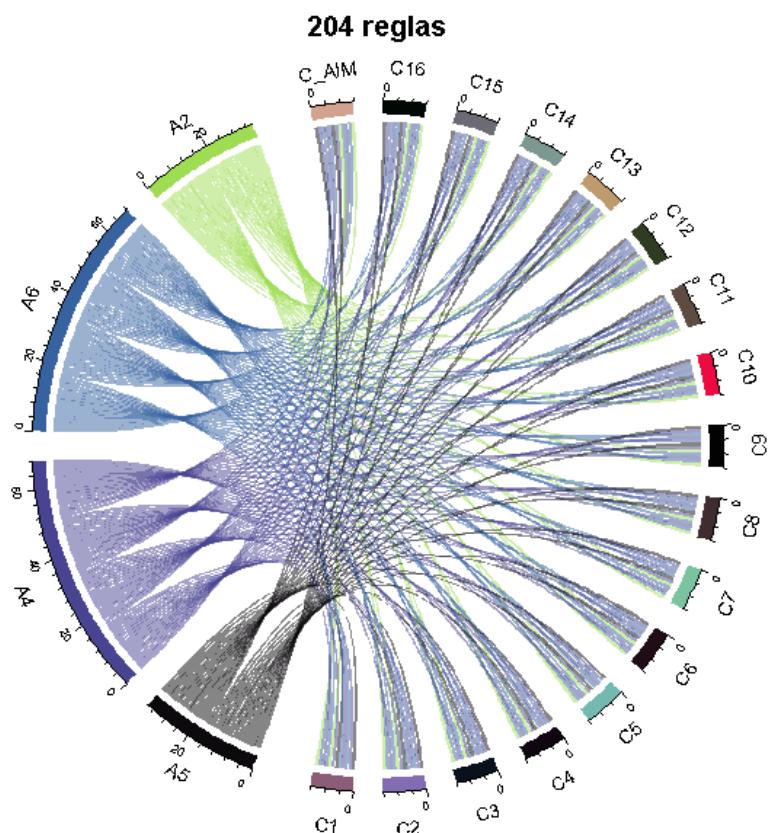


Figura 9 soporte 0,4 %

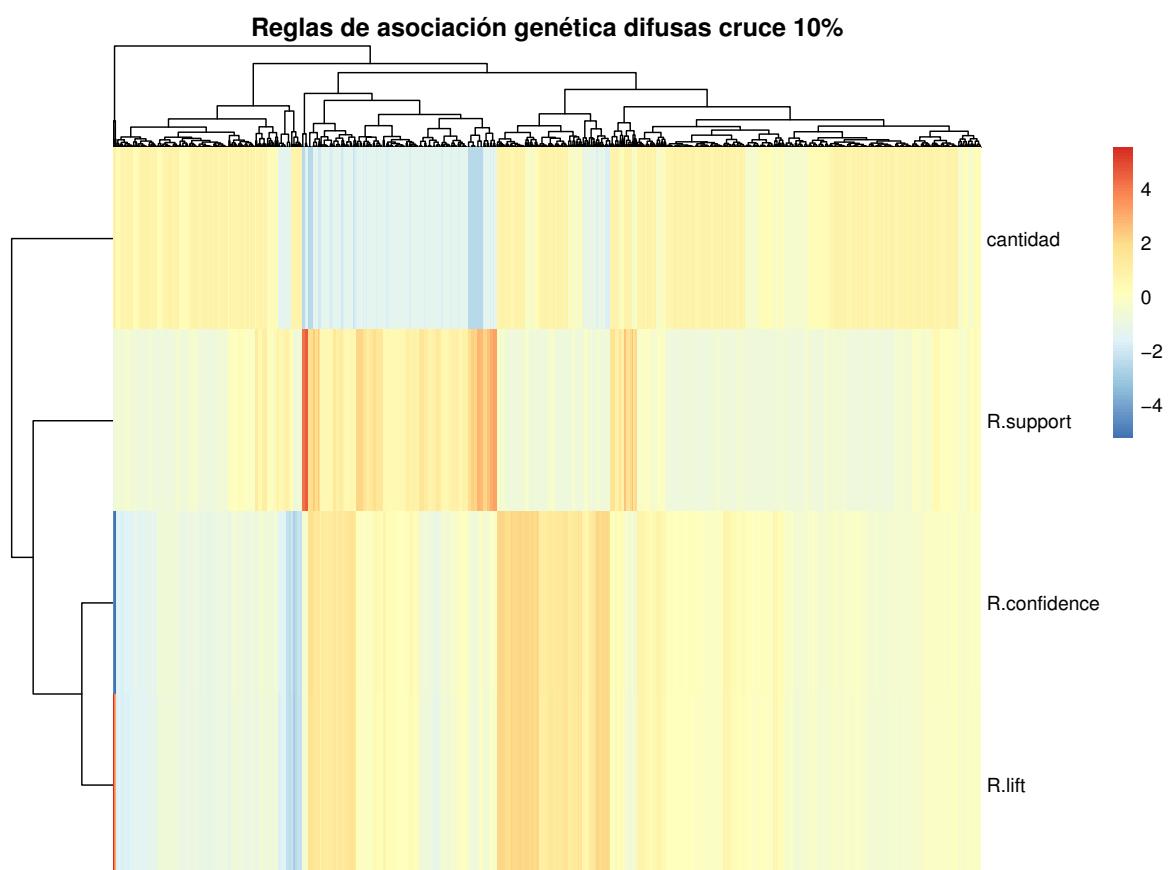


Figura 10 Mapa de calor construcción de reglas difusas genéticas con probabilidad de cruce del 10 %

F.6.2. Probabilidad de cruce 50 %

Representación gráfica de las reglas

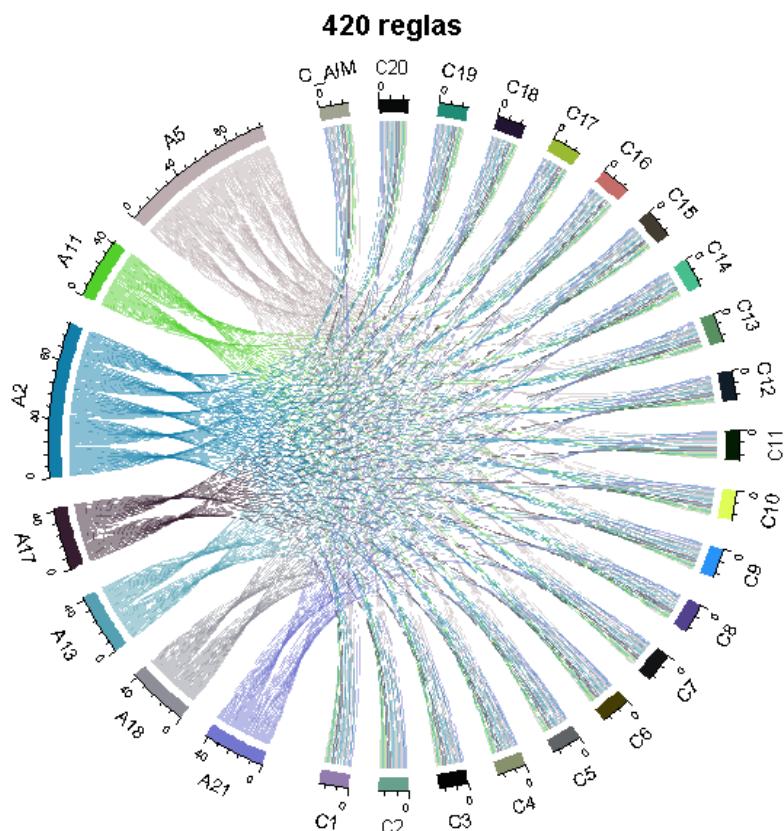


Figura 11 Reglas de asociación soporte 0,4 % confianza 0,8 %

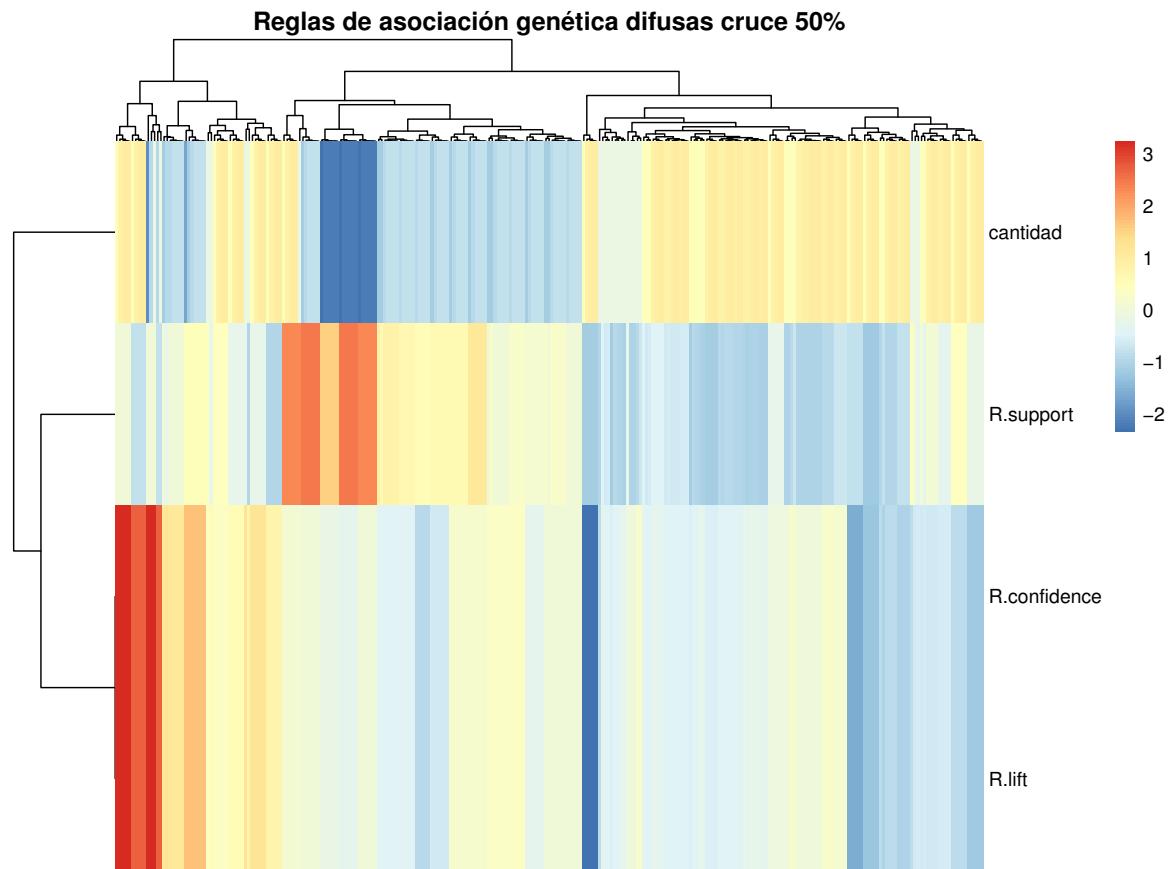


Figura 12 Mapa de calor de construcción de 8175 reglas

Conjunto de reglas extraídas y métricas de calidad

Términos lingüísticos

Tabla 7 Terminos ligüísticos de conjunto de reglas de asociación (probabilidad de cruce 50%)

No	Etiqueta	Término
1	EstiloAprendizaje=-0.017552349864146943;0.07260259527092619;0.16275754040599932	L1
2	EstiloAprendizaje=-0.07663571537798916;0.025740256114448697;0.12811622760688654	L2
3	EstiloAprendizaje=-0.08970285193184237;0.020929490451903603;0.13156183283564957	L3
4	Financiamiento=0.08397739548621967;0.13990406926838117;0.19583074305054265	L1
5	Financiamiento=0.08711736295670502;0.15097865346351977;0.2148399439703345	L2
6	Financiamiento=0.08948913791706349;0.15855550716049946;0.22762187640393544	L3
7	FinanciamientoColegio=0.04810261138213323;0.083850901887945;0.11959919239375677	L1
8	FinanciamientoColegio=0.050063418242117665;0.08483172640390901;0.11960003456570037	L2

Continúa en la siguiente página...

No	Etiqueta	Término
9	FinanciamientoColegio=0.05352397582538219;0.08204875540363324;0.1105735349818843	L3
10	FinanciamientoColegio=0.06152197041426786;0.09629027857605921;0.13105858673785056	L4
11	FinanciamientoColegio=0.07162190123990089;0.10014668081815195;0.128671460396403	L5
12	JornadaTrabajo=0.005484884039786975;0.11128018781095339;0.2170754915821198	L1
13	JornadaTrabajo=0.01901103887034708;0.1317850707624445;0.24455910265454192	L2
14	JornadaTrabajo=0.025748552122551865;0.13154385589371828;0.2373391596648847	L3
15	JornadaTrabajo=0.04113219941594429;0.12114439073029365;0.201156582044643	L4
16	JornadaTrabajo=0.04703201910268337;0.1598060509947808;0.2725800828868782	L5
17	JornadaTrabajo=0.12520204098966292;0.23099734476082934;0.33679264853199575	L6
18	JornadaTrabajo=0.1384369174492077;0.21844910876355705;0.2984613000779064	L7
19	MediaNotasAdmision=-0.018281951225526782;0.22083172484292485;0.4599454009113765	L1
20	MediaNotasAdmision=-0.03805283570208945;0.17653521220028928;0.391123260102668	L2
21	MediaNotasAdmision=-0.04370469231789048;0.19540898375056115;0.4345226598190128	L3
22	MediaNotasAdmision=-0.044521130820759425;0.20910216619515637;0.46272546321107216	L4
23	MediaNotasAdmision=-0.07516546013359321;0.17845783688232258;0.4320811338982384	L5
24	MediaNotasAdmision=0.1188633685314398;0.33345141643381854;0.5480394643361972	L6
25	PEDAD3=0.027144961726457695;0.10859550525162387;0.19004604877679004	L1
26	PEDAD3=0.05969864390097607;0.14114918742614224;0.22259973095130842	L2
27	PEXPEDOCEN=0.05356684312535476;0.18332845101173;0.31309005889487984	L1
28	PEXPEDOCEN=0.11539324335814319;0.21499336949952158;0.3145934956409	L2
29	PEXPEDOCEN=0.1342322839163307;0.2571177328182041;0.3800031817200775	L3
30	PEXPEDOCEN=0.1483330174519591;0.27809462533672163;0.4078562332214842	L4
31	PEXPEDOCEN=0.18538496989203393;0.3082704187939073;0.43115586769578074	L5
32	PEXPEDOCEN=0.19069268063857728;0.29029280677995567;0.3898929329213341	L6
33	PIndAcademico=-0.030510642817992345;0.10682625693224165;0.24416315668247565	L1
34	PIndAcademico=-0.034501193382675835;0.08678368469466707;0.20806856277200997	L2
35	PIndAcademico=-0.05992238276465628;0.07751377353668604;0.21494992983802835	L3
36	SEDAD2=-0.004914635845156612;0.0909093448201794;0.1867333254855154	L1
37	SEDAD2=0.007840343581164921;0.07827280730664801;0.1487052710321311	L2
38	SEDAD2=-0.014060613522206056;0.09215491305257417;0.1983704396273544	L3
39	SEDAD2=0.07062637731350127;0.1768419038882815;0.28305743046306175	L4
40	SEDAD2=0.08260982555484131;0.1530422892803244;0.2234747530058075	L5
41	SEDAD3=0.06390589887892692;0.1509270228633195;0.23794814684771207	L6
42	SEDAD3=-7.778252663307589E-4;0.08624329871806181;0.1732644227024544	L7
43	SMediaEdad=0.25212728064006285;0.617578017280495;0.9830287539209271	L1
44	SMediaEdad=0.2526137035953404;0.6083924576632401;0.9641712117311398	L2
45	SMediaEdad=0.36174045467253635;0.6254796395111242;0.889218824349712	L3

Continúa en la siguiente página...

No	Etiqueta	Término
46	SMediaEdad=0.3990501395392251;0.7645008761796572;1.1299516128200895	L4
47	SMediaEdad=0.42556519207561827;0.781343946143518;1.1371227002114177	L5
48	TEDAD1=-0.03358779389125966;0.05592318849807414;0.14543417088740795	L1
49	TEXPEDOCEN=-0.006230678017039443;0.10199846681263891;0.21022761164231726	L1
50	TEXPEDOCEN=0.04380835375708088;0.16815590869323277;0.2925034636293847	L2
51	TEXPEDOCEN=0.06051495069433152;0.16874409552400987;0.2769732403536882	L3
52	TEXPEDOCEN=0.14463284323156225;0.2528619880612406;0.361091132890919	L4
53	TiempoGraduacion=Bajo	L1
54	TIndAcademico=0.0020949892645219376;0.06851032663020462;0.1349256639958873	L1
55	TIndAcademico=0.00927092052065507;0.0710588471997292;0.13284677387880334	L2
56	TIndAcademico=0.009337131978567392;0.0792528490178453;0.1491685660571232	L3
57	TIndAcademico=-0.02015378927125447;0.04163413740781966;0.10342206408689379	L4
58	TIndAcademico=0.02884658408680374;0.09526192145248642;0.1616772588181691	L5
59	TIndAcademico=0.02941934491455897;0.09933506195383687;0.16925077899311478	L6
60	TIndAcademico=0.06592671004211931;0.12771463672119343;0.18950256340026755	L7
61	TMediaEdad=-0.020935392878973808;0.09929785901012203;0.21953111089921787	L1
62	TMediaEdad=0.04348291139433008;0.16371616328342592;0.28394941517252176	L2
63	TMediaEdad=0.07965507802094074;0.21319362832389027;0.3467321786268398	L3
64	TMediaEdad=0.08356374407838685;0.23014785703666907;0.3767319699949513	L4
65	TMediaEdad=0.18064998170677532;0.31418853200972485;0.4477270823126744	L5
66	TMediaEdad=0.19487014059570296;0.3151033924847988;0.43533664437389463	L6
67	UbicacionColegio=-0.01984175629511426;0.04149381910609628;0.10282939450730683	L1
68	UbicacionColegio=-0.045220461835057624;0.025451066996865293;0.09612259582878821	L2
69	ViveCon=0.006346496260566242;0.17104717368065966;0.3357478511007531	L1
70	ViveCon=0.006790242718221512;0.15193073681568564;0.2970712309131498	L2
71	ViveCon=0.0107511560149052;0.19425000447463847;0.37774885293437177	L3
72	ViveCon=0.10156993608428855;0.26627061350438197;0.4309712909244754	L4
73	ViveCon=0.10271318747107719;0.28621203593081046;0.46971088439054376	L5
74	ViveCon=0.15185272007500364;0.29699321417246777;0.4421337082699319	L6
75	ViveCon=0.19758194626114065;0.3427224403586048;0.4878629344560689	L7
76	ViveCon=0.1985066498152357;0.3632073272353291;0.5279080046554225	L8

F.6.3. Probabilidad de cruce 80 %

Representación gráfica de las reglas

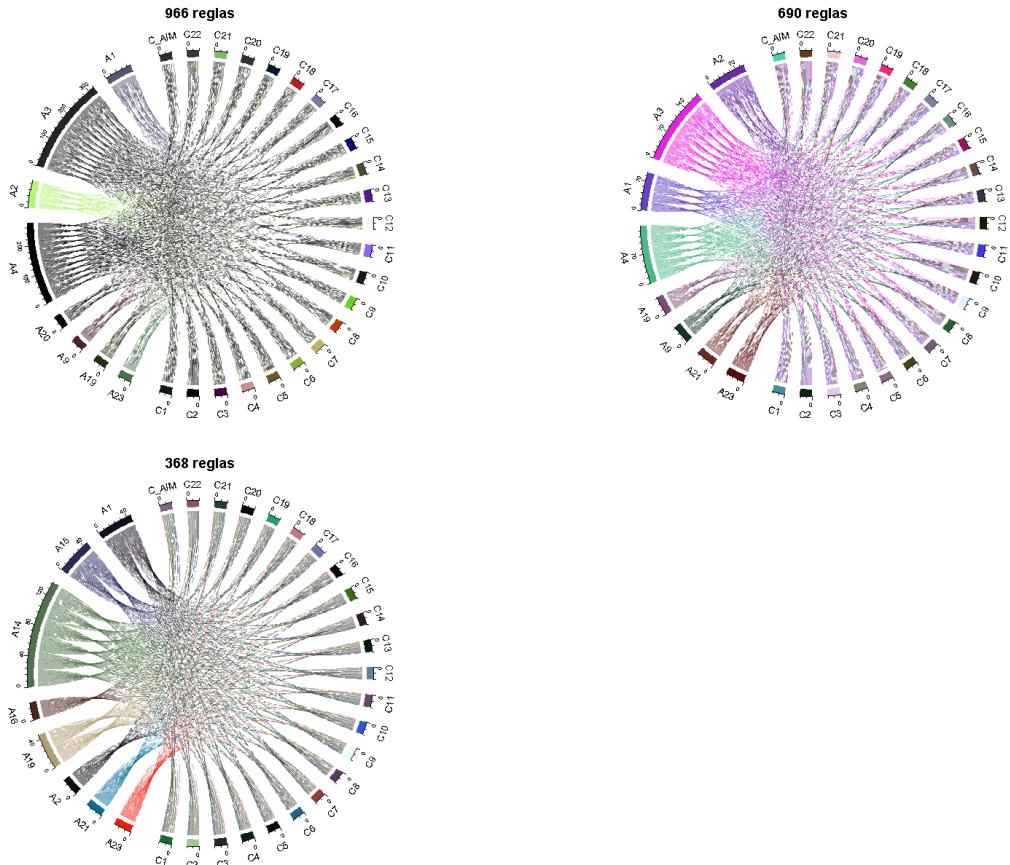


Figura 13 soporte 3 %

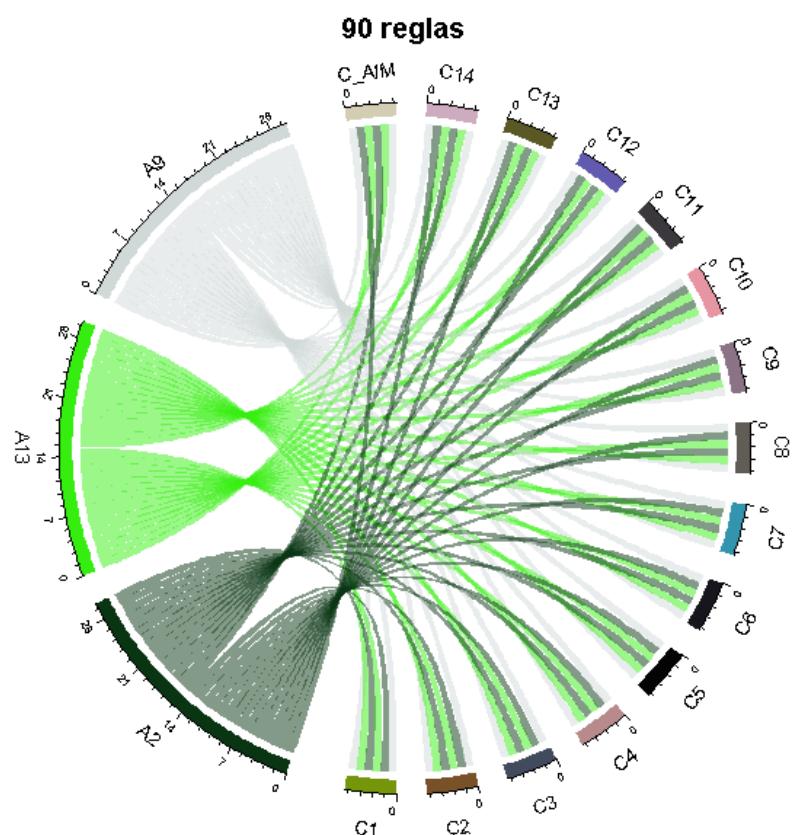


Figura 14 Soporte de la regla 0,4 %, confianza 0,7 %

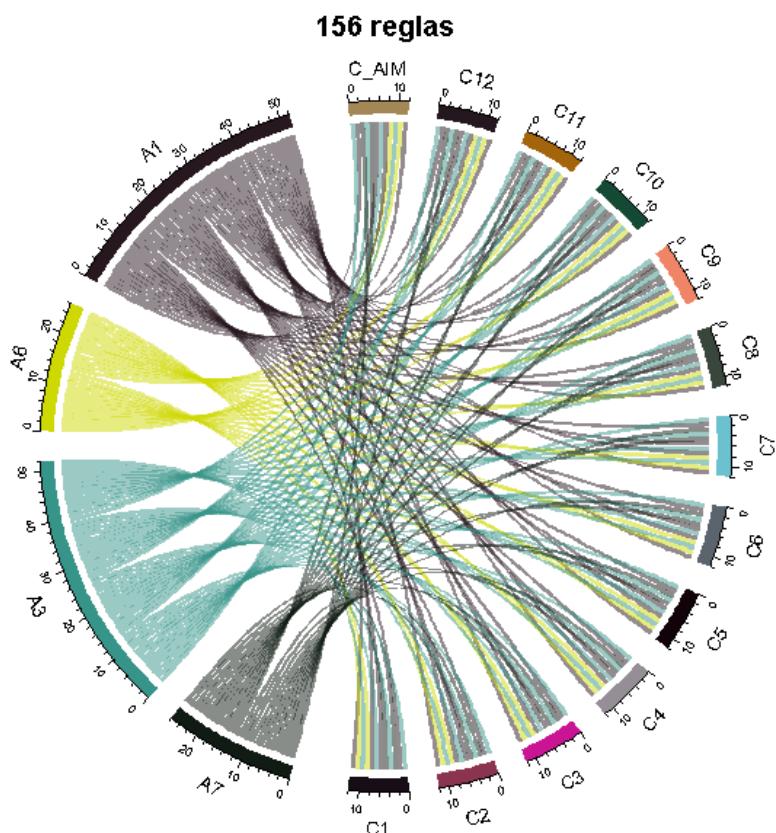


Figura 15 Soporte de la regla 0,5 %, confianza 0,2 %

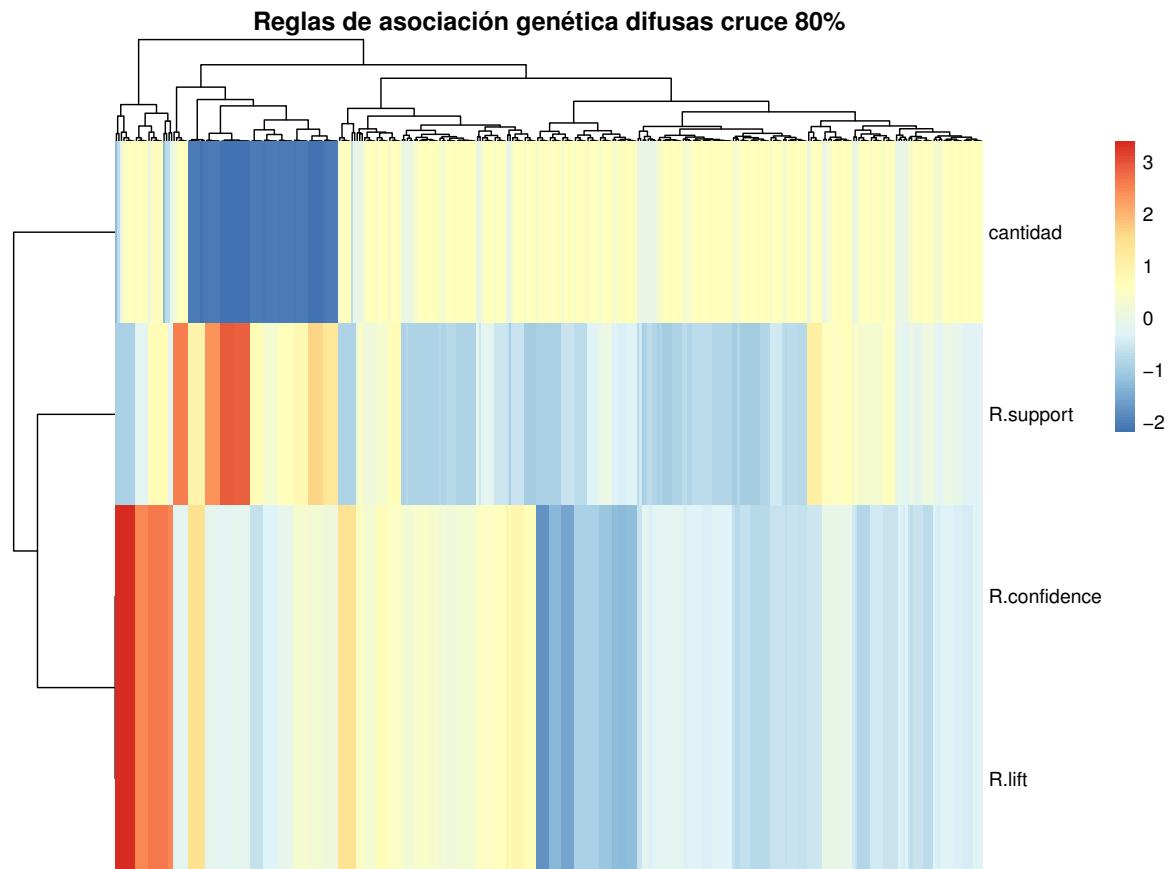


Figura 16 Mapa de calor de 11116 reglas de asociación difusas genéticas con cruce del 80 %.

Conjunto de reglas extraídas y métricas de calidad

Tabla 8 Conjunto de reglas de asociación difusas genéticas (probabilidad de cruce 80 %)

#	Reglas difusas seleccionadas	Soporte	Confianza	R.lift
1	SI PIndAcademico= L3 y SEDAD2= L4 ENTONCES CES TiempoGraduacion=Bajo	0.20	0.10	1.33
2	SI JornadaTrabajo= L3 y SEDAD2= L4 ENTONCES CES TiempoGraduacion=Bajo	0.20	0.10	1.24
3	SI JornadaTrabajo= L7 y TMediaEdad= L4 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.07
4	SI JornadaTrabajo= L7 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.06
5	SI JornadaTrabajo= L3 y TMediaEdad= L4 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.05
6	SI TEDAD1= L2 y TMediaEdad= L4 ENTONCES CES TiempoGraduacion=Bajo	0.20	0.10	1.04
7	SI TEDAD1= L2 y JornadaTrabajo= L7 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.04
8	SI JornadaTrabajo= L3 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.03
9	SI PIndAcademico= L3 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.03
10	SI TEDAD1= L2 y MediaNotasAdmision= L6 ENTONCES TiempoGraduacion=Bajo	0.20	0.10	1.02

Términos lingüísticos

Tabla 9 Terminos ligüísticos de conjunto de reglas de asociación (probabilidad de cruce 80 %)

No	Etiqueta	Término
1	EstiloAprendizaje=-0.02486712881985023;0.05988955025530952;0.14464622933046928	L1
2	Financiamiento=0.08912280200874098;0.14561525864880498;0.20210771528886898	L1
3	Financiamiento=0.09358181774402204;0.1578514486671577;0.22212107959029337	L2
4	Financiamiento=0.11073764641562363;0.1663072774822474;0.2218769085488712	L3

Continúa en la siguiente página...

No	Etiqueta	Término
5	JornadaTrabajo=0.023542365118030165;0.11186385156690837;0.20018533801578658	L1
6	JornadaTrabajo=0.047348385641731855;0.15636446989456293;0.26538055414739403	L2
7	JornadaTrabajo=0.07029403831980095;0.15644430977976703;0.2425945812397331	L3
8	JornadaTrabajo=0.0723258539441885;0.17747539391966471;0.28262493389514093	L4
9	JornadaTrabajo=0.09276646687216857;0.18108795332104677;0.269409439769925	L5
10	JornadaTrabajo=0.099602661283748;0.20861874553657908;0.31763482978941016	L6
11	JornadaTrabajo=0.1194790616831614;0.20562933314312748;0.29177960460309355	L7
12	JornadaTrabajo=0.13444384934097708;0.2227653357898553;0.3110868222387335	L8
13	MediaNotasAdmision=-0.0037485104788074486;0.24068441123633919;0.48511733295148585	L1
14	MediaNotasAdmision=0.009656026367491466;0.2673079342371107;0.5249598421067299	L2
15	MediaNotasAdmision=-0.028780053124640215;0.228871854744979;0.48652376261459823	L3
16	MediaNotasAdmision=-0.04856122964462789;0.16798394366686759;0.38452911697836306	L4
17	MediaNotasAdmision=-0.05322460392797443;0.14556180970595253;0.34434822333987947	L5
18	MediaNotasAdmision=0.1037740868134428;0.30256050044736976;0.5013469140812967	L6
19	MediaNotasAdmision=0.10414250303608696;0.3485754247512336;0.5930083464663802	L7
20	MediaNotasAdmision=0.12024319876896489;0.33678837208046036;0.553335453919559	L8
21	MediaNotasAdmision=0.2653532612991704;0.4641396749330974;0.6629260885670243	L9
22	NumHijos=-0.036745079500347874;0.06448468906476586;0.16571445762987957	L1
23	NumHijos=-0.07962021523237647;0.06914951262042418;0.21791924047322483	L2
24	PEDAD3=0.01920334785760411;0.1064246616595036;0.19364597546140308	L1
25	PEDAD3=0.07659792524952727;0.16381923905142676;0.25104055285332627	L2
26	PEXPEDOCEN=0.09344926021664496;0.19685606394531327;0.3002628676739816	L1
27	PEXPEDOCEN=0.12076545733765631;0.2825560014412223;0.44434654554478825	L2
28	PEXPEDOCEN=0.12836549353394466;0.23712422743585088;0.3458829613377571	L3
29	PEXPEDOCEN=0.1327323845684235;0.2553172999215297;0.37790221527463586	L4
30	PEXPEDOCEN=0.1646642076254695;0.2680710113541378;0.3714778150828061	L5
31	PEXPEDOCEN=0.18314535150591132;0.29190408540781754;0.40066281930972375	L6
32	PIndAcademico=-0.021285222899475723;0.12118629861361478;0.2636578201267053	L1
33	PIndAcademico=-0.03305058337793029;0.09351284769188997;0.22007627876171024	L2
34	PIndAcademico=-0.047882598648865904;0.08188969330897564;0.21166198526681718	L3
35	PIndAcademico=-0.07563435272274346;0.07486090988707997;0.2253561724969034	L4
36	PIndAcademico=0.08457010435145593;0.21434239630929747;0.344114688267139	L5
37	SEDAD2=0.001482844323754423;0.10902369764235748;0.21656455096096053	L1
38	SEDAD2=0.0019427139866096677;0.12023829736475383;0.238533880742898	L2
39	SEDAD2=-0.005095137284108792;0.09026214437721072;0.18561942603853024	L3
40	SEDAD2=0.01092278880399381;0.08275100352912278;0.15457921817784617	L4
41	SEDAD2=0.06756705435611673;0.1751079076747198;0.2826487609933228	L5

Continúa en la siguiente página...

No	Etiqueta	Término
42	SEDAD2=0.07446732640455639;0.1698246080658759;0.2651818897271954	L6
43	SEDAD2=0.08395629814345877;0.15578451279218217;0.22761272744090555	L7
44	SEDAD3=0.0031644514891834563;0.09507729874601878;0.1869901460028541	L1
45	SEDAD3=0.06924591259477533;0.16115875985161066;0.253071607108446	L2
46	SMediaEdad=0.2020094382695189;0.5595552513833643;0.9171010644972096	L1
47	SMediaEdad=0.22011474326493174;0.4985148161582198;0.7769148890515079	L2
48	SMediaEdad=0.28354036049052544;0.6695995125202685;1.0556586645500117	L3
49	SMediaEdad=0.34563069516128686;0.6240307680545749;0.902430840947863	L4
50	SMediaEdad=0.3664615491205988;0.7525207011503419;1.1385798531800848	L5
51	SMediaEdad=0.39866216579166336;0.7562079789055087;1.113753792019354	L6
52	SMediaEdad=0.43182929658935826;0.728531546561987;1.0252337965346159	L7
53	TEDAD1=-0.03743464575252997;0.04891839080509805;0.13527142736272607	L1
54	TEDAD1=-0.04554217333271215;0.03474520508663125;0.11503258350597464	L2
55	TEDAD1=-0.08191937809388808;0.013177554322711768;0.10827448673931162	L3
56	TEXPEDOCEN=0.0038376940384677732;0.10880813856619527;0.21377858309392278	L1
57	TEXPEDOCEN=0.020526226643135165;0.13924542619464136;0.25796462574614754	L2
58	TEXPEDOCEN=0.08199177140333394;0.20071097095484014;0.31943017050634637	L3
59	TEXPEDOCEN=0.08852643089590603;0.19349687542363353;0.298467319951361	L4
60	TEXPEDOCEN=0.13661513766830422;0.2553343372198104;0.3740535367713166	L5
61	TEXPEDOCEN=0.17432260751851172;0.2792930520462392;0.38426349657396675	L6
62	TiempoGraduacion=Bajo	L1
63	TIndAcademico=0.010359861652448829;0.07969862179426417;0.14903738193607952	L1
64	TIndAcademico=0.014818654829332437;0.0770506888500608;0.1392827228707892	L2
65	TIndAcademico=-0.015134628508755757;0.04482927564007974;0.10479317978891524	L3
66	TIndAcademico=0.019281020721271927;0.07924492487010742;0.13920882901894294	L4
67	TIndAcademico=0.035966925914066514;0.09819895993479488;0.16043099395552324	L5
68	TMediaEdad=0.049132004399029164;0.1619888689246895;0.27484573345034985	L1
69	TMediaEdad=0.07994175763579542;0.21606709077326108;0.35219242391072675	L2
70	TMediaEdad=0.08240599406885987;0.20111348293156925;0.3198209717942786	L3
71	TMediaEdad=0.12227788941817545;0.2351347539438358;0.3479916184694961	L4
72	TMediaEdad=0.17930349270657542;0.2980109815692848;0.41671847043199417	L5
73	UbicacionColegio=-0.02045211411489127;0.03834629827829423;0.09714471067147973	L1
74	ViveCon=0.050078864512005605;0.20563436922595416;0.3611898739399027	L1
75	ViveCon=0.06008761777886995;0.23643265983175055;0.41277770188463114	L2
76	ViveCon=0.06605123349220976;0.257585093004137;0.44911895251606426	L3
77	ViveCon=0.09696538873404101;0.23718058459006408;0.3773957804460871	L4
78	ViveCon=0.10847231578343175;0.28481735783631235;0.46116239988919294	L5

Continúa en la siguiente página...

No	Etiqueta	Término
79	ViveCon=0.17113612273017917;0.3266916274441277;0.48224713215807624	L6
80	ViveCon=0.20019780575993537;0.34041300161595844;0.4806281974719815	L7

