



## Dedicatoria y Agradecimientos

*Quiero dedicar este proyecto a las personas que me han acompañado a lo largo de camino, gracias por su apoyo, su comprensión, sé que sin ellos el camino habría sido una carga mucho más pesada.*

*Gracias a mi pareja que siempre se ha quedado en segundo plano cuando en realidad se merece un primero.*

*Gracias a mis amigos que han entendido que esto era una parte importante de mi vida.*

*Y gracias a mi amiga felina, que me ha acompañado con su ronroneo.*

*A todos vosotros, muchas gracias.*

## Resumen de la Memoria

En esta memoria se expone qué es Educational Data Mining mediante una introducción y el estado actual en el que se encuentra esta disciplina y las herramientas que se utilizarán para su desarrollo durante el proyecto. Veremos que es una ciencia que se encuentra en desarrollo y en un estadio muy reciente de su evolución.

A continuación se exponen las causas y motivos que llevan a desarrollar el proyecto en éste ámbito del conocimiento.

Seguidamente nos introducimos en cuál es el proceso que se desarrolla cuando se propone un trabajo sobre la investigación de datos, cómo se prepara el entorno y cuáles son las principales tareas a desarrollar para llevar a cabo un proyecto de este tipo. Así mismo, también dedicamos una parte importante a la planificación sin la cual todo proyecto estaría desprovisto de una parte fundamental de su estructura que nos indica el camino de ruta sobre el que vamos dando los pasos y si éstos se dirigen al final del proyecto o por el contrario nos estamos desviando.

En la siguiente parte del documento se exponen los procesos a los que se someten los datos para obtener resultados que incluyen el tratamiento y procesamiento de los mismos mediante diferentes técnicas de minería. Se podrán a prueba diferentes técnicas y modelos para extraer información sobre el lenguaje de Minería de Datos, R.

Entre los últimos puntos a comentar y no por ello menos importante trataremos de llegar a conclusiones sobre el proyecto, el entorno, los datos y la totalidad del trabajo realizado, lo cual nos servirá para hacer balance y poder mejorar de cara a futuros proyectos.

Durante toda la memoria se ha intentado hacer un uso abundante de gráficas e imágenes que ayuden a la comprensión de los resultados obtenidos que en ocasiones pueden resultar difíciles de interpretar.

Por último en la sección de anexos incluimos diferentes scripts en R que nos han llevado a la realización del proyecto y tablas de datos que aportan información de cara a obtención de un modelo explicativo.

## Índice de Contenidos

<b>Dedicatoria y Agradecimientos .....</b>	<b>1</b>
<b>Resumen de la Memoria .....</b>	<b>2</b>
<b>Índice de Contenidos .....</b>	<b>3</b>
<b>Tabla de Ilustraciones.....</b>	<b>5</b>
<b>1. Introducción .....</b>	<b>6</b>
1.1. <i>Qué es "Educational Data Mining" y su estado en la actualidad.....</i>	6
1.2. <i>Qué es R para el Data Mining.....</i>	7
1.3. <i>Herramientas Utilizadas en este Proyecto.....</i>	7
<b>2. Justificación y Objetivos del TFG .....</b>	<b>8</b>
2.1. <i>Qué se quiere conseguir con este proyecto sobre "Educational Data Mining" .....</i>	8
<b>3. Enfoque y método seguido .....</b>	<b>8</b>
3.1. <i>Definición de la tarea de Data Mining.....</i>	8
3.2. <i>Origen de los Datos .....</i>	9
3.3. <i>Herramientas de Utilidad .....</i>	9
3.4. <i>Preparación de los Datos.....</i>	9
3.5. <i>Construcción del Modelo.....</i>	10
3.6. <i>Evaluación e interpretación del Modelo .....</i>	10
3.7. <i>Integración de los Resultados.....</i>	10
<b>4. Planificación del proyecto .....</b>	<b>11</b>
4.1. <i>Fechas Límite y Entregas.....</i>	11
4.2. <i>Diagrama de Gantt.....</i>	12
<b>5. Productos obtenidos .....</b>	<b>13</b>
<b>6. Tratamiento de los Datos .....</b>	<b>13</b>
6.1. <i>Extracción de los Datos.....</i>	13
6.2. <i>Transformación de los Datos.....</i>	14
6.3. <i>Descripción de los Datos .....</i>	14
6.4. <i>Representación de los Datos.....</i>	17
<b>7. Procesamiento de los Datos.....</b>	<b>19</b>
7.1. <i>Campo de estudio.....</i>	19
7.1.1. <i>Asignaturas Matriculadas .....</i>	19
7.1.1. <i>Asignaturas Aprobadas .....</i>	19

7.2.	<i>Selección del Modelo</i> .....	19
7.3.	<i>Elaboración del Modelo</i> .....	20
7.3.1.	NNFM – Non negative Factor Matrix.....	20
7.3.2.	PCA – Principal Component Analysis.....	26
7.3.1.	Método K-means – Agrupamiento de datos.....	32
7.3.2.	Preparación de un clasificador Éxito/Fracaso .....	38
7.3.3.	Árboles de Decisión .....	39
<b>8.</b>	<b>Conclusiones</b> .....	<b>52</b>
<b>9.</b>	<b>Anexos</b> .....	<b>54</b>
9.1.	<i>Anexo 1. Librerías utilizadas para el análisis de los datos y las funciones gráficas</i> .....	54
9.2.	<i>Anexo 2. Procesado de datos y estadísticas básicas</i> .....	54
9.3.	<i>Anexo 3. Script para Análisis de Matrices de Factores No Negativos (NMF)</i> .....	56
9.4.	<i>Anexo 4. Script para Análisis de Componentes Principales (PCA)</i> .....	56
9.5.	<i>Anexo 5. Script para Análisis de datos de Asignaturas mediante algoritmo K-means</i> .....	58
9.1.	<i>Anexo 6. Script para Análisis de datos de Matriculaciones de Alumnos mediante algoritmo K-means y árbol de decisión</i> .....	58
9.1.	<i>Anexo 7. Tabla con los resultados de paquetes de asignaturas típicas, sus relaciones con los clúster calculados y su relación éxito fracaso del paquete</i> .....	60
<b>10.</b>	<b>Futuras ampliaciones</b> .....	<b>76</b>
10.1.	<i>Anexo 8. Tabla de alumnos con sus matriculaciones y sus respectivas asignaturas aprobadas</i> .....	76
10.2.	<i>Anexo 9. Tabla cruzada de referencias entre agrupaciones de matrículas y agrupaciones de asignaturas aprobadas</i> .....	76
10.3.	<i>Anexo 10. Tabla cruzada de paquetes de asignaturas matriculadas y asignaturas aprobadas en listado</i> .....	77
<b>11.</b>	<b>Bibliografía y Referencias</b> .....	<b>79</b>
11.1.	<i>Referencias Escritas</i> .....	79
11.2.	<i>Referencias Online</i> .....	79

## Tabla de Ilustraciones

Ilustración 1. Fases del Proyecto de Extracción del Conocimiento.....	8
Ilustración 2. Ciclo de Construcción del Modelo.....	10
Ilustración 3. Diagrama de Gantt con Microsoft Project .....	12
Ilustración 4. Fichero TFG_EDMLA.dat. Datos de Matriculaciones ADE.....	13
Ilustración 5. Fichero GR01.ass.1sem. Códigos de Asignaturas ADE.....	13
Ilustración 6. Estructura de la Información Extraída .....	14
Ilustración 7. Diagrama de barras de asignaturas aprobadas y suspensas .....	17
Ilustración 8. Diagrama de barras tasa de Asig. Alumnos Aprobados vs Matriculados ..	18
Ilustración 9. Diagrama de barras tasa de Asig. Alumnos Aprobados vs Suspensos .....	18
Ilustración 10. Representación de la Matriz de Matriculaciones en mapa de color.....	21
Ilustración 11. Matriz de Matriculaciones haciendo distinción .....	21
Ilustración 12. Estudio del Rango de Factorización.....	22
Ilustración 13. Factorizaciones para valores de rank de 2 a 8 .....	25
Ilustración 14. Gráfica de correlación entre asignaturas con agrupación de 5 clústers ..	27
Ilustración 15. Representación del total de matriculados en cada clúster (5).....	34
Ilustración 16. Representación del total de matriculados en cada clúster (6).....	36
Ilustración 17. Clúster de Alumnos en relación a sus matrículas.....	37
Ilustración 18. Relación Éxito/Fracaso en cada clúster .....	38
Ilustración 19. Árbol Decisión ctree Asignaturas Matriculadas vs Éxito/Fracaso .....	40
Ilustración 20. Árbol de Decisión método "rpart" en R, complejidad 0.001.....	42
Ilustración 21. Árbol de Decisión método "rpart" en R complejidad 0.0001 .....	45
Ilustración 22. Árbol de Decisión con el algoritmo J48.....	50
Ilustración 23. Subproducto tabla cruzada entre paquetes.....	76
Ilustración 24. Subproducto tabla cruzada asignaturas.....	77

## 1. Introducción

### 1.1. Qué es “Educational Data Mining” y su estado en la actualidad

La Minería de los Datos Educativos es una disciplina emergente, preocupada por el desarrollo de métodos para explorar los tipos de datos que provienen de los centros educativos y el uso de métodos para entender mejor a los estudiantes y las herramientas de las cuales aprenden. Las áreas sobre las que se ha desarrollado la investigación de este conocimiento se pueden relacionar con áreas de investigación en Enseñanza presencial, Enseñanza a Distancia a través de Internet (E-learning) y Tutorización Inteligente.

Se puede decir que el “Educational Data Mining” es una rama de la Minería de Datos que trata cuestiones relevantes en entornos educacionales relacionadas con las áreas anteriormente descritas en temas más específicos tales como:

- Soporte a profesores
- Recomendaciones para estudiantes
- Predicciones y modelaje de comportamientos
- Mediciones sobre rendimiento de los estudiantes
- Análisis de los grupos sociales
- Análisis, planificación y construcción de cursos y eventos

Mediante la aplicación de la Minería en los procesos de aprendizaje se pueden obtener resultados que impliquen tanto a profesores como a alumnos en un proceso de estudio de las posibles mejoras que se puedan llevar a cabo y en la eliminación de puntos de conflicto e indeseables. Los objetivos sobre los que se puede centrar la tarea de Data Mining en esta área involucran a diferentes partes interesadas:

- Estudiantes: personalización de la experiencia del aprendizaje, sugerencias para las mejoras del proceso de aprendizaje, adaptación del aprendizaje continuo
- Educadores: retroalimentación objetiva sobre la enseñanza, análisis del aprendizaje de los estudiantes y errores más frecuentes, determinación de las actividades más efectivas, personalización de tareas, etc.
- Centros de Educación: consejo y mejora de las decisiones de los alumnos de cara a estudios superiores, sugerencias sobre cursos e itinerarios
- Administraciones Públicas: eficiencia en la utilización de recursos, mejoras de planes de estudio, evaluaciones de personal docente, adaptaciones curriculares, etc.

En el punto actual en el que nos encontramos respecto a la rama sobre Educación en el área de Minería de Datos, parece que todavía quedaría mucho sobre este tema en el que avanzar. Nos encontramos con cuestiones sobre las que se ha de mejorar considerablemente, tales como herramientas que faciliten la extracción de datos que sean suficientemente flexibles y simples para los usuarios relacionados con la educación, también la integración en el sistema de estas herramientas es necesaria para poder seguir haciendo prospecciones de los datos, incluso nos encontramos con que se plantea necesaria una convergencia de los diferentes tipos de modelos y datos ya que en la actualidad existe una variedad extensa de modelado de los mismos, lo cual

hace que las descripciones de modelos estén muy acopladas respecto al modelo que se estudia.

## 1.2. Qué es R para el Data Mining

R es un sistema de análisis y gráficos estadísticos creados por Ross Ihaka y Robert Gentleman. R es también un software como un lenguaje considerado dialecto del lenguaje S creado por los Laboratorios de Bell de AT & T.

R se distribuye libremente bajo los términos GNU (Licencia General Pública). Su desarrollo y distribución se ha llevado a cabo por varios profesionales de la estadística conocidos como "R Development Core Team".

Está disponible en varias modalidades: en código fuente (escrito principalmente en C y algunas rutinas en Fortran) esencialmente para Unix y máquinas Linux, y en programas ejecutables para las diferentes plataformas existentes como Windows, Linux y Macintosh. Lo mínimo necesario para instalar R, ya sea de las fuentes o de los binarios pre-compilados, se distribuye desde el sitio Web de "R Archive Red Integral" (CRAN), donde también están disponibles las instrucciones para la instalación.

El lenguaje R permite al usuario, por ejemplo, programar bucles para analizar sucesivamente varios conjuntos de datos. También es posible combinar en un solo programa funciones estadísticas diferentes para realizar análisis más complejos.

Entre sus aspectos principales R cuenta con:

- Multitud de paquetes adicionales fácilmente instalables
- Posibilidad de comunicarse con bases de datos y ficheros
- Exportar resultados en diversos formatos
- Disponibilidad de un entorno gráfico
- Accesibilidad a grandes datos de internet como Google, Twitter y Facebook

## 1.3. Herramientas Utilizadas en este Proyecto

En la actualidad se está experimentando un crecimiento en herramientas que facilitan el uso y la interacción del investigador y los datos. Las interfaces gráficas van ganando relevancia en entornos empresariales que necesitan tener una exposición de datos de manera más atractiva. Sin embargo para el entorno científico que nos interesa, se necesita la potencia y la versatilidad a la hora de expresar comandos que la presentación en sí misma para una finalidad estética. Por esta misma razón, el proyecto se desarrollará en un entorno de especificación del lenguaje de programación R, el cual ofrece las características detalladas en el apartado anterior.

Sin embargo, por su facilidad de visualización e integración visual también se hará un uso limitado de herramientas de interfaz gráfica tales como RStudio y Rattle, las cuales proporcionan un puente de enlace entre la línea de comando y su resultado más inmediato, generalmente en forma de gráficos.

Para este mismo proyecto, se hará un uso escaso de herramientas como Weka, tan sólo se tratarán en casos excepcionales en los que se necesite mayor claridad documental.

## 2. Justificación y Objetivos del TFG

### 2.1. Qué se quiere conseguir con este proyecto sobre “Educational Data Mining”

El objetivo del proyecto trata acerca de cómo y qué información se puede obtener desde los entornos de aprendizaje que generan información sobre los estudiantes para establecer mecanismos que ayuden al tratamiento objetivo de los resultados.

La principal ventaja que se puede obtener de un estudio sobre datos educacionales es la mejora de comportamientos, metodologías y actividades para conseguir mejorar el sistema en su conjunto.

Este proyecto tendrá una doble intención, por un lado investigar los patrones de matrícula de los estudiantes de forma que podamos tener una estimación de cuáles son las asignaturas más matriculadas y cómo se agrupan estas matriculaciones. Se investigará la similitud que presentan las matriculaciones de las asignaturas por un lado y la similitud de las matriculaciones de los alumnos matriculados por otro.

## 3. Enfoque y método seguido

La tarea de data mining es un proceso más complicado que la simple introducción de datos y la obtención de resultados. En sí misma la tarea de data mining es todo un proceso que a partir de un objetivo, siguiendo una serie de pasos, pretende obtener un conocimiento modelado. La ilustración siguiente muestra el proceso de forma lineal:

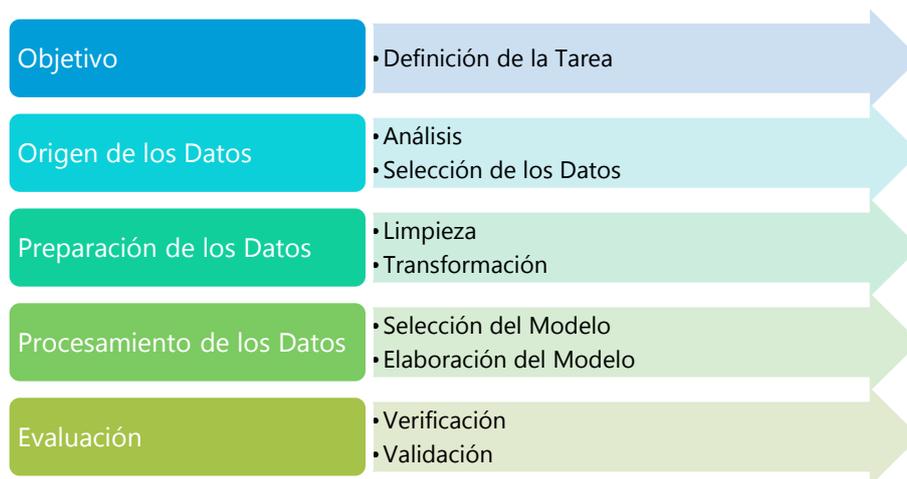


Ilustración 1. Fases del Proyecto de Extracción del Conocimiento

### 3.1. Definición de la tarea de Data Mining

En este punto de la tarea de Data Mining se intentan reducir las posibles tipologías de proyecto en las que pueden derivar nuestros datos:

- Relacionar objetos similares
- Clasificar datos
- Predecir comportamientos
- Describir asociaciones

### 3.2. Origen de los Datos

Los datos se encuentran por todos lados, lo que más trabajo requiere para hacer un estudio de Data Mining es saber dónde encontrarlos. Podemos destacar diversas vías de obtención de datos, entre algunas:

- Lo ideal sería un Data Warehouse
- Ficheros de texto y bases de datos
- Datos de logs y ficheros Excel

Para nuestra investigación utilizaremos los datos proporcionados por el departamento de Minería de Datos de la Universidad Oberta de Cataluña.

### 3.3. Herramientas de Utilidad

A nuestro alcance se encuentran multitud de herramientas de código libre y fácilmente accesibles para su utilización directa tales como:

- R: paquete y lenguaje de programación flexible y versátil que dispone de una amplia comunidad que ofrece soporte y documentación.
- Rattle: herramienta de entorno gráfico que corre sobre la instalación de R. Ofrece una gran ayuda para el análisis visual, el modelado y la creación de gráficos de diversos tipos.
- RStudio: Integra la consola de comandos de R, junto con un entorno de soporte a la programación
- Weka: se trata de un proyecto de minería que contiene herramientas básicas y avanzadas para la manipulación de datos y la visualización de resultados.

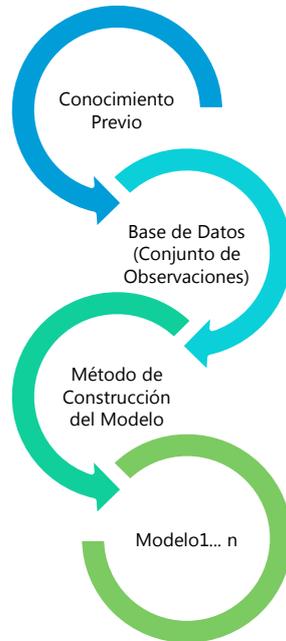
### 3.4. Preparación de los Datos

En esta fase disponemos de los datos con los cuales vamos a trabajar y hemos de proceder a preparar las muestras para poder aplicar las técnicas necesarias para la obtención de modelos. Antes de poder aplicar dichas técnicas hemos de:

- Comprobar la calidad de la muestra
- Comprobar los atributos necesarios
- Comprobar el formato de los datos
- Realizar tareas de limpieza para evitar:
  - Datos incompletos e incorrectos
  - Datos redundantes
- Realizar tareas de transformación para:
  - Simplificar / Derivar los datos

### 3.5. Construcción del Modelo

En la construcción del modelo hacia el que tienden los datos, tendremos que disponer de unos conocimientos previos los cuales aplicaremos a las observaciones de las que disponemos, todo ello nos llevará a la búsqueda de uno o varios modelos que nos sirvan para representar el conocimiento que se puede extraer de los datos. La siguiente ilustración representa el proceso:



**Ilustración 2. Ciclo de Construcción del Modelo**

### 3.6. Evaluación e interpretación del Modelo

Una vez construido el modelo, el paso siguiente es medir su efectividad y validar que los resultados expuestos en el mismo son válidos para un amplio conjunto de datos. Para ello dividimos esta fase en:

- Proceso de evaluación
- Conjunto de validación

### 3.7. Integración de los Resultados

Por último se ha de tratar de construir un sistema mínimamente automático que proceda a procesar información similar de entrada y la incorpore junto a la antigua información para ofrecer salidas actualizadas. Este último paso correspondería a la inserción de un algoritmo en la cadena del conocimiento, de manera que al conseguir nuevos datos éstos se añadieran a los nuevos resultados tan solo dependiendo de su ejecución.

## 4. Planificación del proyecto

### 4.1. Fechas Límite y Entregas

En la tabla a continuación se exponen las fechas sobre las que se proyectan las diferentes tareas:

<b>TFG - Educational Data Mining &amp; Learning Analytics</b>	<b>85 días</b>	<b>27/02/2014</b>	<b>25/06/2014</b>
<b>Plan de Trabajo</b>	<b>8 días</b>	<b>27/02/2014</b>	<b>10/03/2014</b>
Definición y Objetivo	8 días	27/02/2014	10/03/2014
<b>Tratamiento de Datos</b>	<b>25 días</b>	<b>11/03/2014</b>	<b>14/04/2014</b>
Selección y Extracción de Datos	3 días	11/03/2014	13/03/2014
Exploración de los Datos	6 días	14/03/2014	21/03/2014
Limpieza de Datos	8 días	24/03/2014	02/04/2014
Transformación de Datos	8 días	03/04/2014	14/04/2014
<b>Procesamiento de Datos</b>	<b>26 días</b>	<b>15/04/2014</b>	<b>20/05/2014</b>
Selección del Modelo	6 días	15/04/2014	22/04/2014
Elaboración del Modelo	10 días	23/04/2014	06/05/2014
Verificación y Prueba de Contraste	10 días	07/05/2014	20/05/2014
<b>Entrega Final y Presentación Virtual</b>	<b>14 días</b>	<b>21/05/2014</b>	<b>09/06/2014</b>
Preparación Entrega Final	5 días	21/05/2014	27/05/2014
Preparación Presentación Virtual	5 días	28/05/2014	03/06/2014
Revisión	2 días	04/06/2014	05/06/2014
Referencias y Bibliografía	2 días	06/06/2014	09/06/2014
<b>Tribunal</b>	<b>2 días</b>	<b>24/06/2014</b>	<b>25/06/2014</b>
Tribunal	2 días	24/06/2014	25/06/2014

### 4.2. Diagrama de Gantt

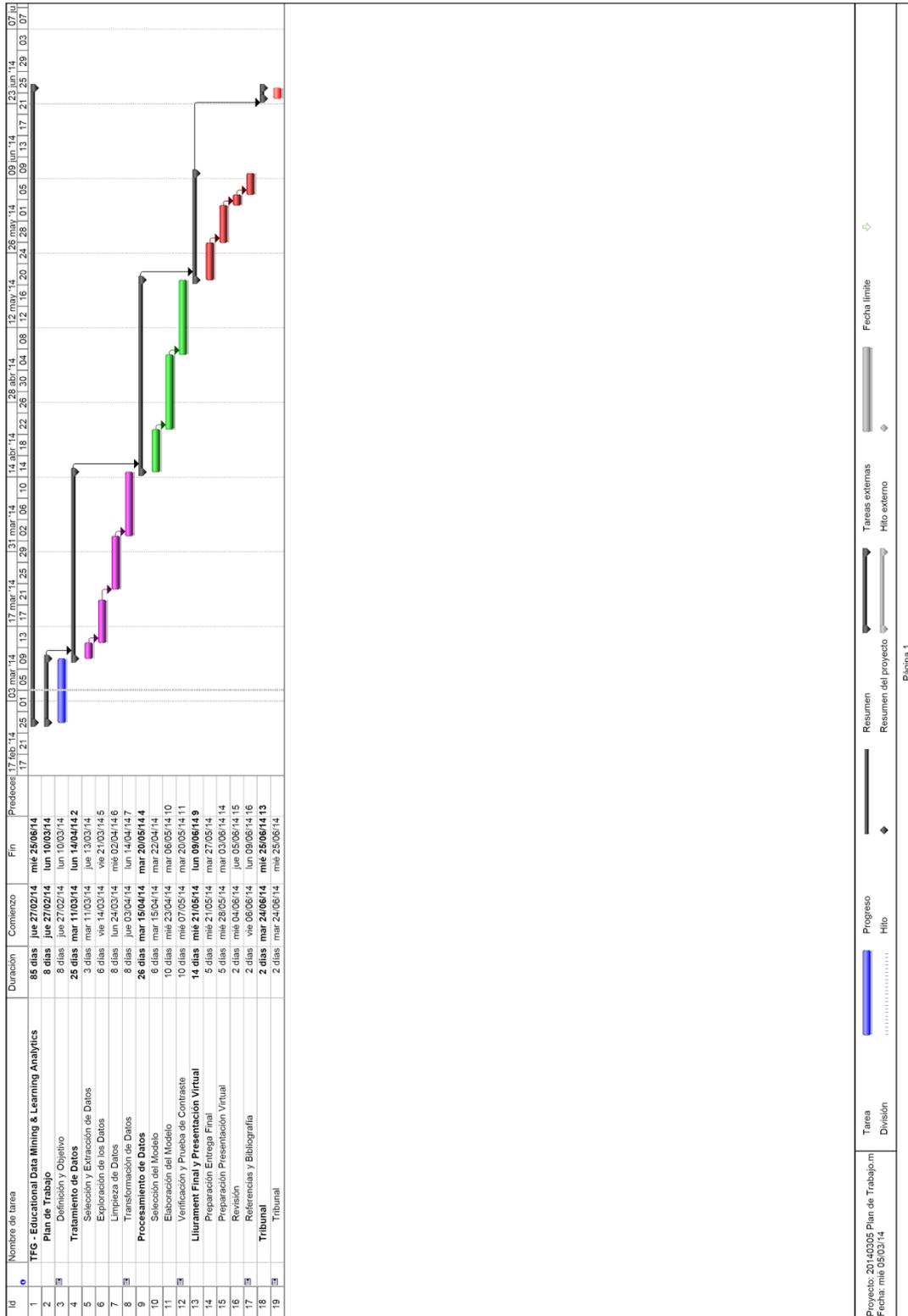


Ilustración 3. Diagrama de Gantt con Microsoft Project



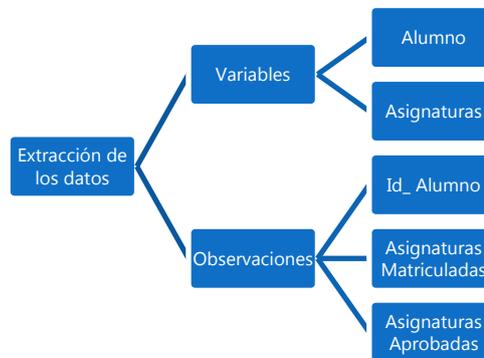
Los datos se componen en las variables Identificador del Alumno y el formato de los datos que se tiene con la extracción es el siguiente:

**Tabla 1. Tabla descripción de la totalidad de los datos**

Campo de Datos	Descripción	Formato	Longitud	Rango de Valores
<b>Id_Alumno</b>	Es el campo que identifica a un alumno	Alfanumérico	10	0000000000 - 9999999999
<b>Fa_ass1</b>	Indicador de la matriculación de la asignatura1	Alfanumérico	1	0 – No matricula 1 – Sí matricula
<b>Supera_ass1</b>	Indicador de la superación de la asignatura1	Alfanumérico	1	0 – No supera 1 – Sí supera
...	...	...	...	...
<b>Fa_ass45</b>	Indicador de la matriculación de la asignatura45	Alfanumérico	1	0 – No matricula 1 – Sí matricula
<b>Supera_ass45</b>	Indicador de la superación de la asignatura45	Alfanumérico	1	0 – No supera 1 – Sí supera

### 6.2. Transformación de los Datos

La estructura de la información extraída relaciona los siguientes conceptos:



**Ilustración 6. Estructura de la Información Extraída**

Por un lado disponemos de los identificadores de los alumnos, junto con los nombres de las asignaturas que tendremos como variables, más adelante veremos cómo se pueden tratar las asignaturas tanto por su código como por su ordinal para facilitar el trabajo. Por otro lado disponemos de las observaciones que nos dirán de manera binaria en 0/1 si la asignatura observada está matriculada y también en forma binaria en 0/1 si se encuentra aprobada.

### 6.3. Descripción de los Datos

Las asignaturas de las que disponemos son las siguientes:

Ordinal	Asignatura	Descripción
1	01.500	Introducción a la empresa
2	01.505	Iniciativa emprendedora
3	01.507	Iniciación a las competencias TIC
4	01.502	Introducción a la información financiera de la empresa
5	01.506	Mercados y conducta
6	01.503	Introducción al derecho
7	01.504	Fundamentos de márketing
8	01.509	Idioma moderno I: Inglés
9	01.501	Fundamentos de estadística
10	01.508	Comportamiento de los agregados económicos
11	01.520	Introducción a la contabilidad
12	01.521	Habilidades directivas
13	01.516	Fundamentos de matemáticas
14	01.517	Logística

Ordinal	Asignatura	Descripción
15	01.519	Personas y organizaciones
16	01.524	Análisis de los estados financieros
17	01.526	Compatibilidad de costes
18	01.528	Control presupuestario y de gestión
19	01.513	Dirección estratégica
20	01.529	Dirección de márketing
21	01.518	Estadística aplicada
22	01.514	Fiscalidad empresarial
23	01.512	Derecho de la empresa
24	01.510	Estructura económica
25	01.515	Decisiones tácticas de producción
26	01.522	Contabilidad financiera
27	01.511	Organización de empresas
28	01.527	Inversión empresarial
29	01.523	Valoración de operaciones financieras
30	01.530	Idioma moderno II: Inglés
31	01.535	Comercio exterior
32	01.554	Técnicas de expresión escrita académica y profesional
33	01.549	Historia económica
34	01.525	Financiamiento empresarial
35	01.537	Derecho financiero y tributario I
36	01.547	Macroeconomía
37	01.542	Negocio electrónico
38	01.534	Gestión y desarrollo de recursos humanos
39	01.545	Microeconomía
40	01.552	Estrategias i técnicas de comunicación de márketing
41	01.561	Contabilidad de gestión para la toma de decisiones
42	01.550	Investigación de mercados
43	01.544	Gestión de la innovación
44	01.536	Negociación
45	01.543	Fiscalidad internacional

En estos datos tenemos una muestra de 4871 Alumnos sobre las 45 asignaturas, de las cuales se ha medido el alumno que se matricula y el resultado que ha obtenido dicho alumno en términos de supera la asignatura o no.

El primer estudio realizado muestra las frecuencias de los datos, agregamos las tasas de aprobados sobre matriculados de manera que tenemos una estimación del ratio y medimos también la tasa de aprobados frente a suspensos para ver la relación entre ambas medidas. Los resultados que se obtienen del estudio son los siguientes:

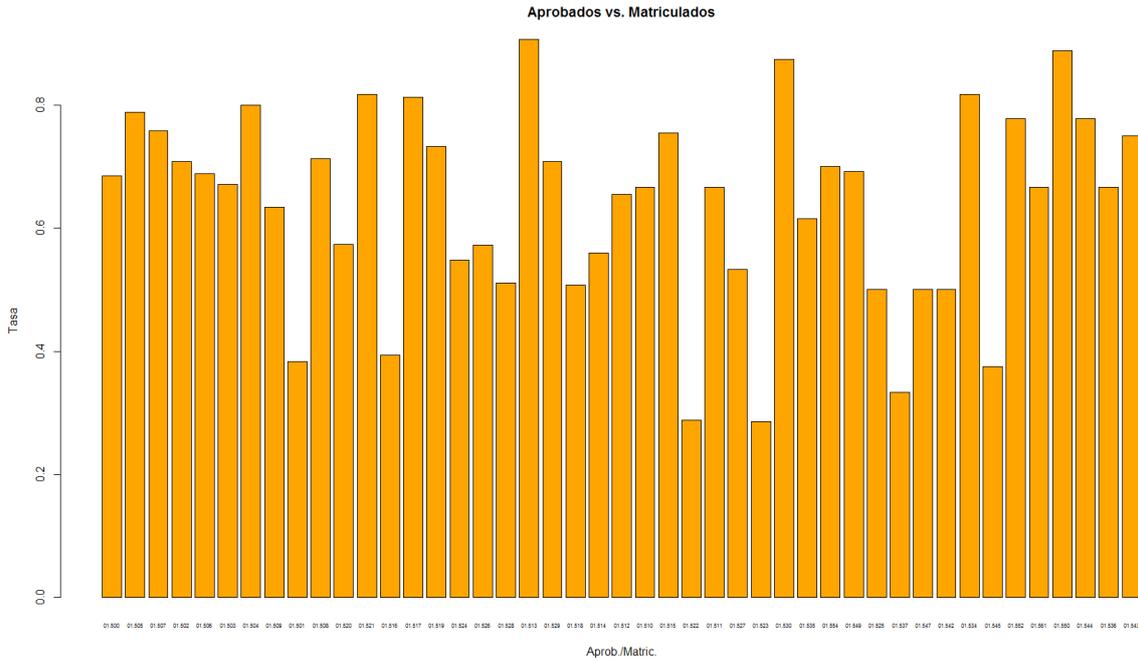
Ordinal	Asignatura	No Matriculados	Matriculados	Aprobados	Suspensos	Aprobados / Matriculados	Aprobados / Suspensos
1	01.500	2290	2581	1769	812	0,6854	2,1786
2	01.505	2532	2339	1844	495	0,7884	3,7253
3	01.507	2600	2271	1722	549	0,7583	3,1366
4	01.502	3293	1578	1118	460	0,7085	2,4304
5	01.506	3424	1447	996	451	0,6883	2,2084
6	01.503	4180	691	464	227	0,6715	2,0441
7	01.504	4199	672	538	134	0,8006	4,0149
8	01.509	4374	497	315	182	0,6338	1,7308
9	01.501	4521	350	134	216	0,3829	0,6204
10	01.508	4596	275	196	79	0,7127	2,4810
11	01.520	4606	265	152	113	0,5736	1,3451
12	01.521	4641	230	188	42	0,8174	4,4762
13	01.516	4716	155	61	94	0,3935	0,6489
14	01.517	4710	161	131	30	0,8137	4,3667
15	01.519	4755	116	85	31	0,7328	2,7419

Ordinal	Asignatura	No Matriculados	Matriculados	Aprobados	Suspensos	Aprobados / Matriculados	Aprobados / Suspensos
16	01.524	4778	93	51	42	0,5484	1,2143
17	01.526	4782	89	51	38	0,5730	1,3421
18	01.528	4783	88	45	43	0,5114	1,0465
19	01.513	4785	86	78	8	0,9070	9,7500
20	01.529	4799	72	51	21	0,7083	2,4286
21	01.518	4806	65	33	32	0,5077	1,0313
22	01.514	4812	59	33	26	0,5593	1,2692
23	01.512	4816	55	36	19	0,6545	1,8947
24	01.510	4820	51	34	17	0,6667	2,0000
25	01.515	4822	49	37	12	0,7551	3,0833
26	01.522	4819	52	15	37	0,2885	0,4054
27	01.511	4838	33	22	11	0,6667	2,0000
28	01.527	4841	30	16	14	0,5333	1,1429
29	01.523	4843	28	8	20	0,2857	0,4000
30	01.530	4847	24	21	3	0,8750	7,0000
31	01.535	4845	26	16	10	0,6154	1,6000
32	01.554	4851	20	14	6	0,7000	2,3333
33	01.549	4858	13	9	4	0,6923	2,2500
34	01.525	4859	12	6	6	0,5000	1,0000
35	01.537	4859	12	4	8	0,3333	0,5000
36	01.547	4859	12	6	6	0,5000	1,0000
37	01.542	4867	4	2	2	0,5000	1,0000
38	01.534	4860	11	9	2	0,8182	4,5000
39	01.545	4863	8	3	5	0,3750	0,6000
40	01.552	4862	9	7	2	0,7778	3,5000
41	01.561	4862	9	6	3	0,6667	2,0000
42	01.550	4862	9	8	1	0,8889	8,0000
43	01.544	4862	9	7	2	0,7778	3,5000
44	01.536	4865	6	4	2	0,6667	2,0000
45	01.543	4867	4	3	1	0,7500	3,0000
<b>Total</b>			14666	10348	4318		

De esta relación de frecuencias podemos sacar las medidas máximas y mínimas de los valores observados. Las características que obtenemos se presentan a continuación:

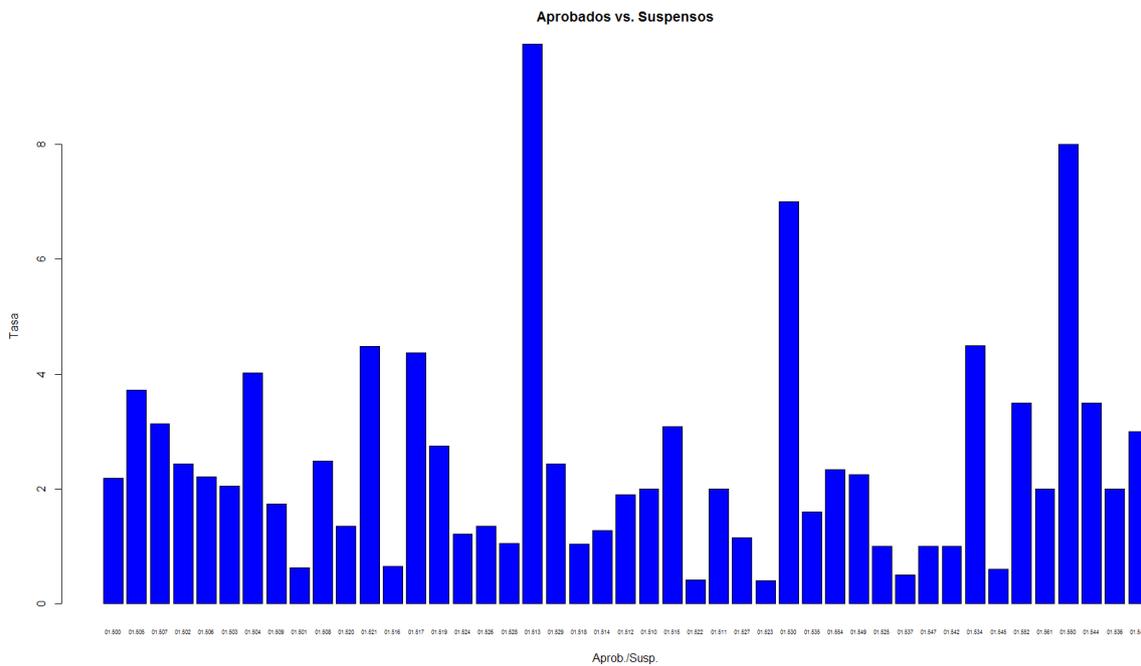
Tipo de Medida	Característica	Asignatura	Valor
Frecuencias Absolutas	Total de Matriculaciones		14666
	Total de Aprobados		10348
	Total de Suspensos		4318
	Asignatura con mayor número de matriculaciones	01.500	2581
	Asignatura con menor número de matriculaciones	01.542	4
		01.543	4
	Asignatura con mayor número de aprobados	01.505	1844
	Asignatura con menor número de aprobados	01.542	2
	Asignatura con mayor número de suspensos	01.500	812
	Asignatura con menor número de suspensos	01.543	1
	01.550	1	
Tasas	Asignatura con mayor tasa de aprobados/matriculados	01.513	0,9070
	Asignatura con menor tasa de aprobados/matriculados	01.523	0,2857
	Asignatura con mayor tasa aprobados/suspensos	01.513	9,7500
	Asignatura con menor tasa aprobados/suspensos	01.523	0,4000





**Ilustración 8. Diagrama de barras tasa de Asig. Alumnos Aprobados vs Matriculados**

En este último gráfico se visualiza la relación que existe entre los aprobados y los suspensos de las diferentes asignaturas, estas son las tasas que relacionan la cantidad de aprobados con la cantidad de suspensos:



**Ilustración 9. Diagrama de barras tasa de Asig. Alumnos Aprobados vs Suspenso**

## 7. Procesamiento de los Datos

### 7.1. Campo de estudio

En nuestro campo de estudio podemos distinguir entre dos tipos de datos que vamos a tratar: las asignaturas matriculadas por cada alumno, y de éstas las asignaturas aprobadas por cada alumno. Veremos por un lado cómo las asignaturas matriculadas se aglomeran para matricularse, para después introducir los resultados que derivan de estas asignaturas matriculadas.

#### 7.1.1. Asignaturas Matriculadas

La tabla siguiente define la estructura con la que se trabajará el primer modelo de similitud y agrupación de Asignaturas Matriculadas:

Campos de Datos	Descripción	Formato	Longitud	Rango de Valores
<b>Id_Alumno</b>	Es el campo que identifica a un alumno	Numérico	10	0000000000 - 9999999999
<b>Ass1</b>	Indicador del resultado de la asignatura: fa_ass1	Numérico	1	0 – No matricula 1 – Si matricula
...	...	...	...	...
<b>Ass45</b>	Indicador del resultado de la asignatura: fa_ass45	Numérico	1	0 – No matricula 1 – Si matricula

#### 7.1.1. Asignaturas Aprobadas

La tabla siguiente define las asignaturas las cuales se han aprobado ya que el valor 0 indica tanto que la asignatura no está aprobada si aparece en la tabla anterior como que no ha sido matriculada si su valor es cero en ambas tablas:

Campos de Datos	Descripción	Formato	Longitud	Rango de Valores
<b>Id_Alumno</b>	Es el campo que identifica a un alumno	Numérico	10	0000000000 - 9999999999
<b>Ass1</b>	Indicador del resultado de la asignatura: sup_ass1	Numérico	1	0 – No aprobada / No matriculada 1 – Si aprobada
...	...	...	...	...
<b>Ass45</b>	Indicador del resultado de la asignatura: sup_ass45	Numérico	1	0 – No aprobada / No Matriculada 1 – Si aprobada

### 7.2. Selección del Modelo

El análisis multivariable que queremos realizar nos supone el desafío de tener que trabajar con una cantidad bastante grande de variables, en este caso tenemos que manejar 45 variables diferentes en dos modos de valoración: Matriculación y Resultado (podemos prescindir del identificador del Alumno y sustituirlo en su defecto por un número secuencial), por lo que haremos un estudio con algunas técnicas que son de gran utilidad: primero trataremos de analizar los datos mediante la técnica de la Matriz de Factorización No Negativa, después trataremos la técnica de componentes

principales y por último realizaremos un estudio mediante el algoritmo k-means y árboles de decisión.

### 7.3. Elaboración del Modelo

#### 7.3.1. NNFM – Non negative Factor Matrix

Con la técnica de descomposición matricial en factores positivos obtenemos una relación de bases y coeficientes que nos indican como de relacionadas pueden estar las variables entre sí. Esta técnica es muy útil al igual que la técnica de PCA (Principal Components Analysis) ya que nos permite localizar agrupaciones de variables y su relación entre ellas.

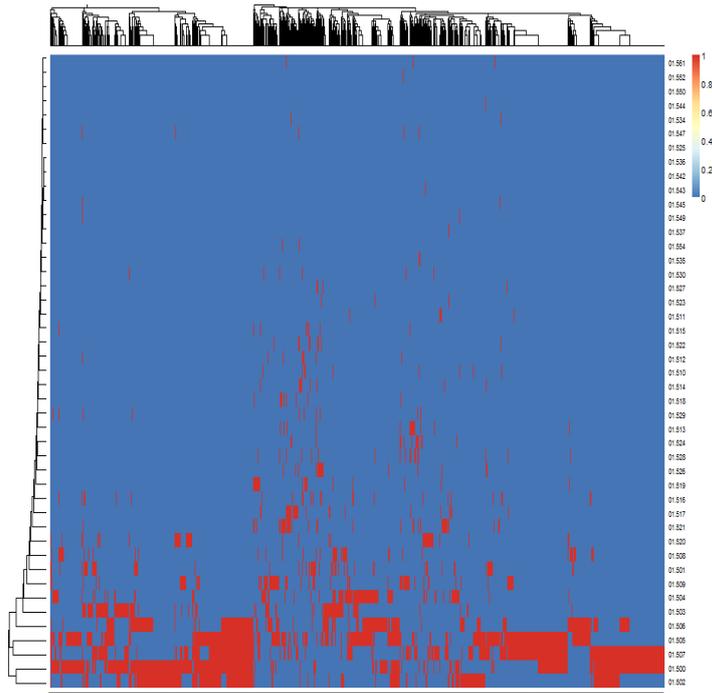
Si tenemos un matriz llamada  $X$  la cual posee unas dimensiones  $n \times p$  y todos sus coeficientes son positivos, aplicar la técnica NMF consiste en encontrar una aproximación mediante otras dos matrices de manera que:

$$X \sim W \cdot H$$

donde  $W$  y  $H$  son matrices de tamaño  $n \times r$  y  $r \times p$  respectivamente, de tal manera que el grado de factorización llamado  $r$  se elige habitualmente entre valores que sean lo suficientemente pequeños para ser manejables y representativos de los datos a la vez. El código en R que utilizamos para nuestro estudio se encuentra reproducido en el Anexo 3.

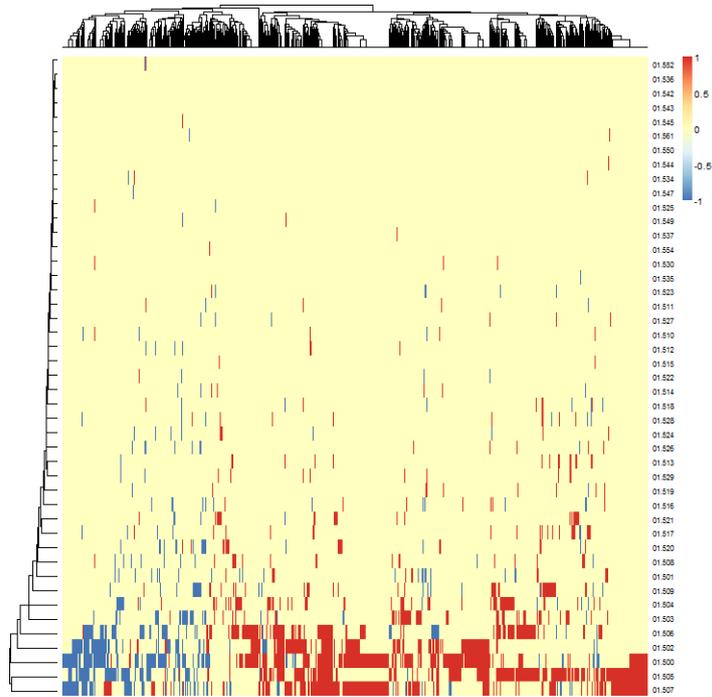
#### **Presentación de la matriz total de matriculaciones:**

La matriz total de matriculaciones dispone los datos de manera que para las 45 asignaturas y los 4871 registros de alumnos, nos muestra un 1 lo que significa que el alumno se ha matriculado y nos muestra un 0 allá donde el alumno no está matriculado. El tratamiento de esta técnica requiere que nos familiaricemos con una visión de los datos en la que las matriculaciones (en rojo) se muestran como en la Ilustración 10. Esta primera representación de los datos nos confirma la misma observación descriptiva de los datos realizada inicialmente, en la cual las asignaturas 01.500, 01.502, 01.503, 01.505, 01.506 y 01.507 suman un total de 10.907 matrículas, lo que representaría un 74,37% de las observaciones sobre las matrículas. Se corresponderían con las seis primeras asignaturas del plan de estudios.



**Ilustración 10. Representación de la Matriz de Matriculaciones en mapa de color**

Para obtener también una visión de las asignaturas según su modalidad de Resultado, tratamos la matriz inicial de matriculaciones hasta convertirla en una matriz de Resultados en la que 1 significa Aprobado, 0 significa No matriculado y -1 significa No aprobado. En la siguiente representación de los datos se hace una distinción entre las asignaturas aprobadas y las suspensas, de manera que ambos tipos se representan respectivamente por el color rojo y azul. Se puede apreciar en la siguiente ilustración la distribución de la que forman parte:



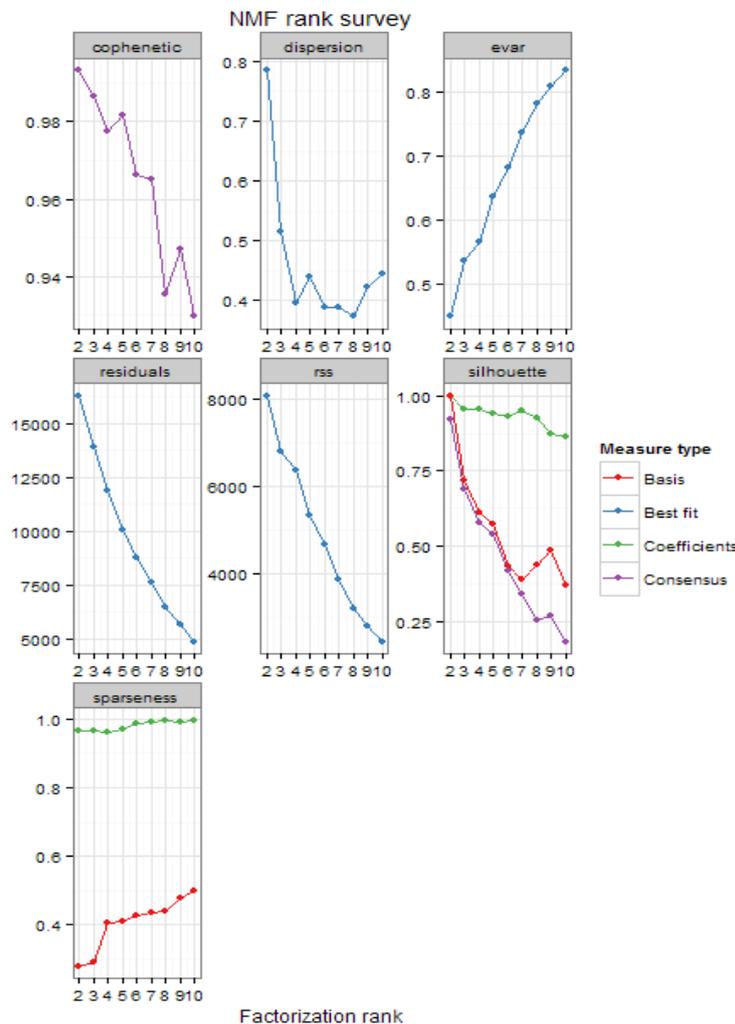
**Ilustración 11. Matriz de Matriculaciones haciendo distinción entre Aprobados (Rojo) y Suspensos (Azul)**

### Investigando la estimación del parámetro $r$ :

Para conseguir tener un criterio sobre cuál debería ser el parámetro más idóneo para factorizar la matriz, se ha realizado de forma iterativa la extracción de diferentes números de bases hasta que se ha encontrado una base en la que a partir de ella las agrupaciones dejaran de tener sentido de forma conjunta.

Según diferentes aproximaciones propuestas, la estimación del parámetro rank para realizar una estimación de mayor calidad habría de coincidir con aquél en el que empezara a disminuir el coeficiente de correlación copenética<sup>1</sup> (según Brunet), también se sugiere que debería corresponder con el primer valor en el que la curva RSS presentara un punto de inflexión (según Frigyesi). Así, realizamos la función de aproximación que nos ofrece el paquete NMF y obtenemos unas gráficas que representan los valores de los que se hablan previamente.

A continuación mostramos los resultados obtenidos del análisis:



**Ilustración 12. Estudio del Rango de Factorización**

El resultado que hemos obtenido muestra diferentes gráficas que representan coeficientes asociados a los datos. Vemos que tanto los coeficientes de correlación

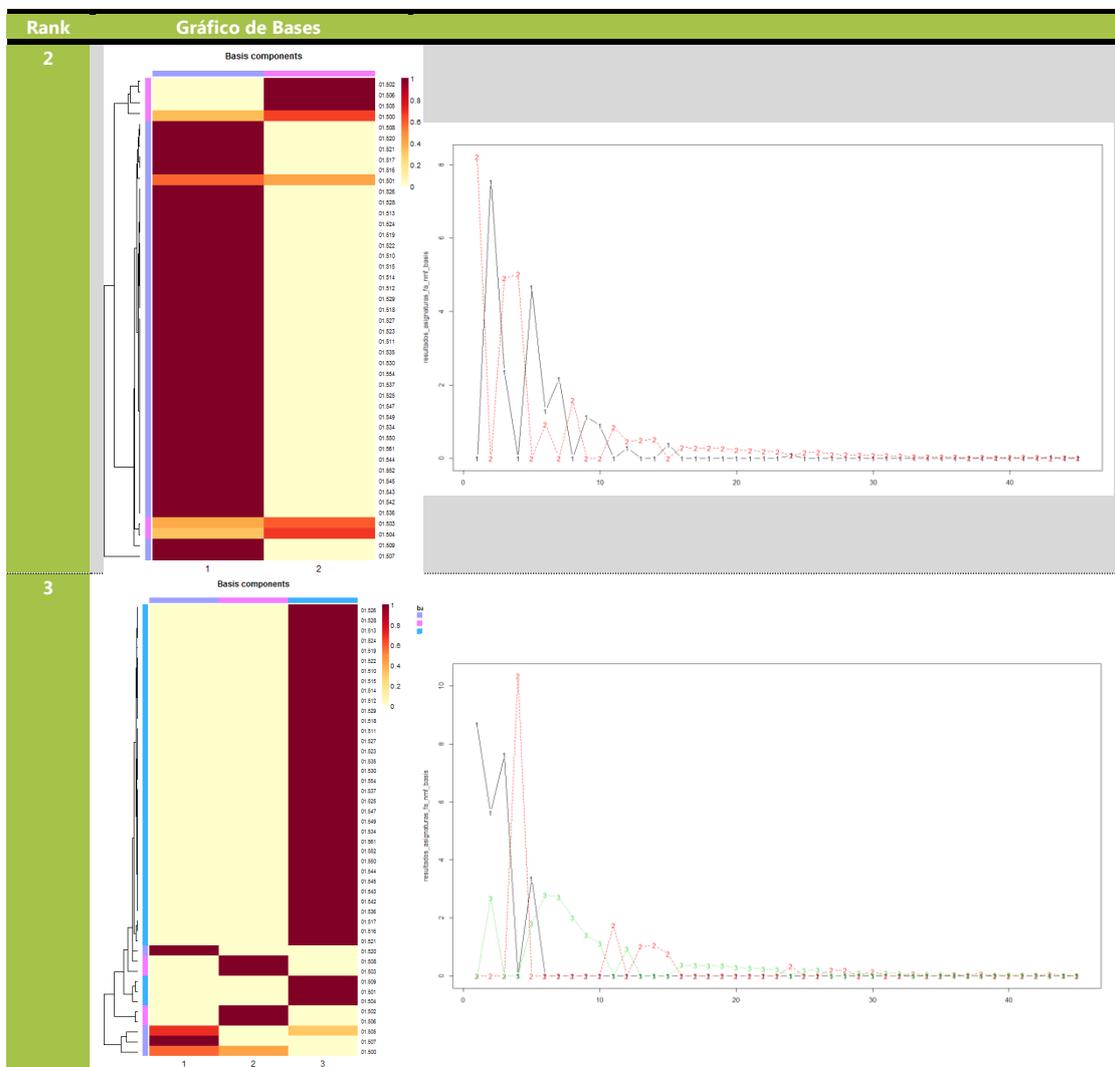
<sup>1</sup> Medida de la proximidad de los datos

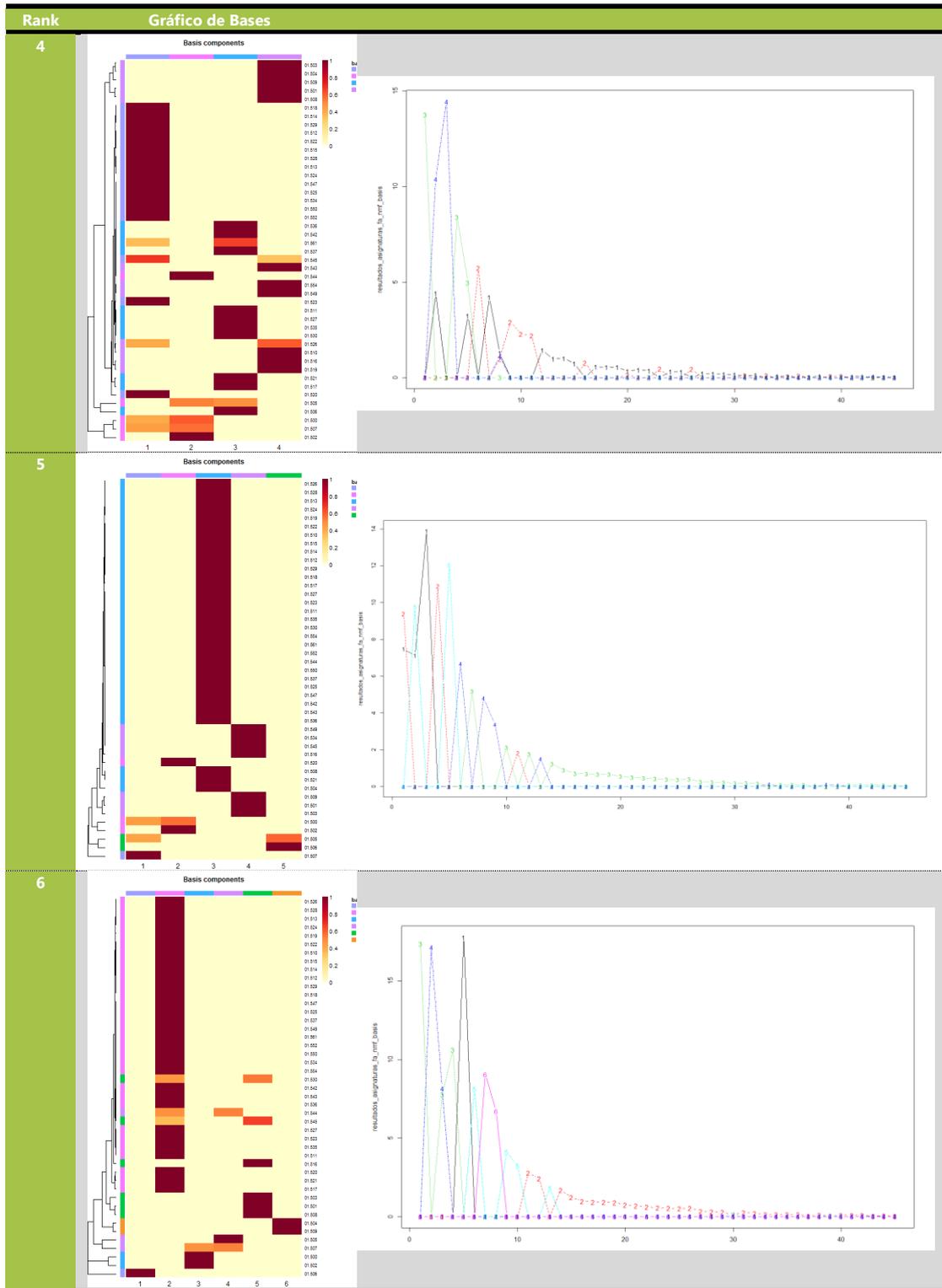
cofenética, como la dispersión presentan picos en los valores 5 y 6. Se puede apreciar también que tanto las bases como la silueta de la sexta gráfica se empiezan a separar a partir del valor 6.

En este momento no podríamos decidir cuál es el valor que agruparía los datos, aunque sí nos acerca a una medida que empezaría a mostrar cuál sería la mejor forma de agrupar los datos.

**Iterando los datos para ver los posibles modelos:**

Para poder apreciar mejor el resultado que se obtendría con diferentes valores de descomposición en r bases, se ha llevado a cabo del proceso de investigación de los valores hasta dar con un valor que nos parezca apropiado para la descomposición. En la siguiente tabla mostramos el valor de r, junto a las representaciones gráficas que nos informan de la relevancia de las bases en la descomposición, podemos ver las agrupaciones de asignaturas que se van formando a medida que el valor de r se incrementa:





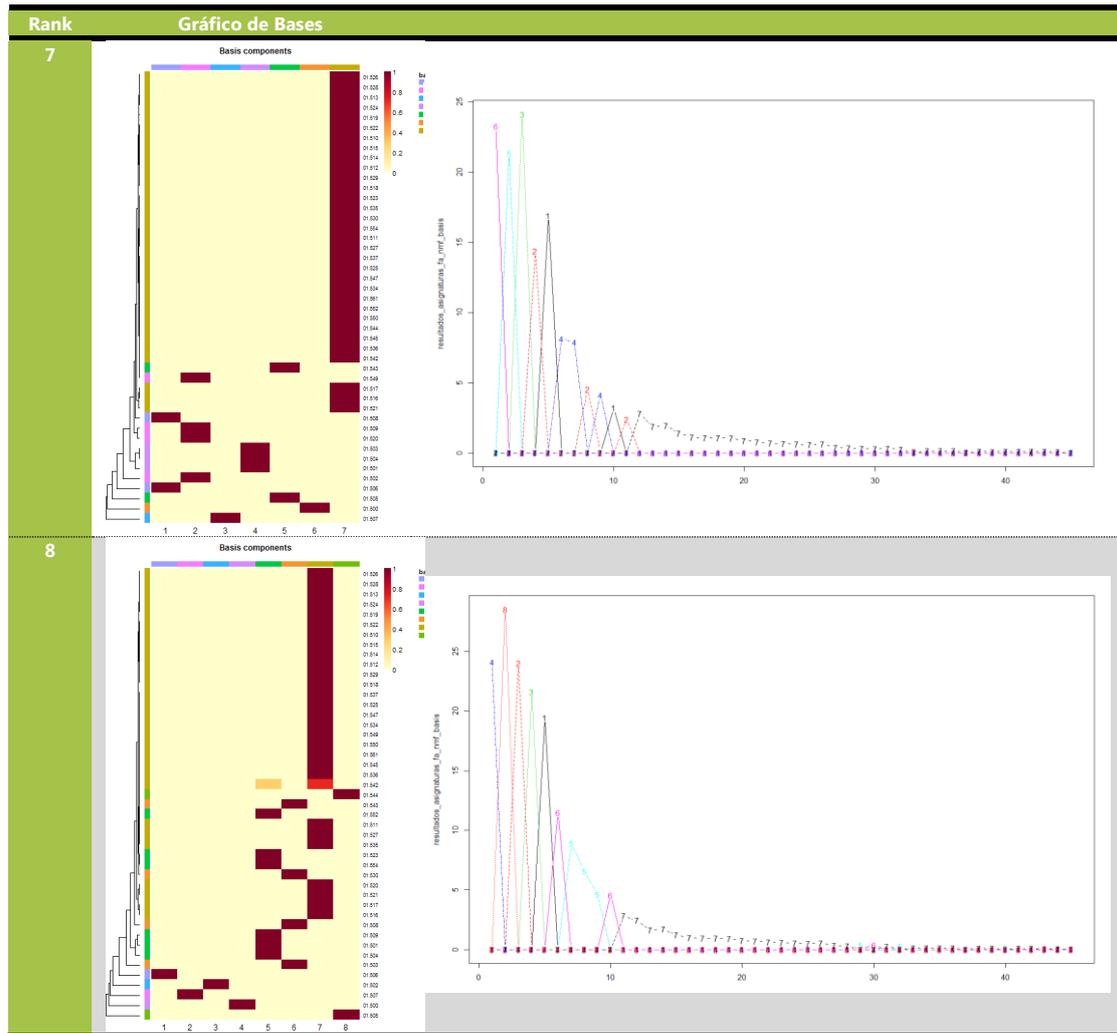


Ilustración 13. Factorizaciones para valores de rank de 2 a 8

## Conclusiones

Según los datos extraídos podemos ver gráficamente que al factorizar la matriz más allá de la base 6 empiezan a establecerse como conjuntos de asignaturas únicos, por lo que podemos pensar que la mejor agrupación se presenta en este número de factores y por debajo de ese número. Llegamos a las siguientes conclusiones:

- 1) En la factorización con un valor de rank=6 obtenemos el soporte a una única asignatura (la asignatura 01.506), lo cual quiere decir que esta asignatura, cuando no va combinada con las otras (más comunes) es la "preferida" para matricularla sola.
- 2) Como a partir del valor de factorización en 6 bases obtenemos resultados de asignaturas aisladas de manera única, miramos las bases anteriores que nos factorizan la matriz con un valor de rank=5, podemos ver en los gráficos que los paquetes de asignaturas que obtenemos se agrupan de la siguiente manera:
  - Grupo 1: la 1ª (01.500), la 2ª (01.505) y la 3ª (01.507), todas ellas superan la cifra de 2200 matriculados.
  - Grupo 2: la 1ª (01.500), la 4ª (01.502) y en muy menor medida la 11ª (01.520)
  - Grupo 3: la 7ª (01.504), la 10ª (01.508), la 12ª (01.521) y a partir de la 14ª en adelante

- Grupo 4: las 9ª (01.501), 6ª (01.503) y 8ª (01.509) junto con las 01.516, 01.534, 01.545 y 01.549
- Grupo 5: la 2ª (01.505), 5ª (01.506)

La base 3 nos dice que quizás solo con las 14 primeras asignaturas tendríamos suficiente representación para establecer una relación de agrupación mediante clústers.

### 7.3.2. PCA – Principal Component Analysis

El análisis de componentes principales es una técnica de reducción de la dimensionalidad que transforma un gran número de variables correlacionadas en una cantidad más manejable de relaciones de variables no correlacionadas las cuales se llamarán componentes de manera que cada componente contenga el máximo de información posible.

Un componente principal tiene la capacidad de sustituir un gran número de variables correlacionadas por un número menor de grupos de variables llamadas componentes que tendrán la siguiente forma:

$$PC_i = \sum_{j=1}^n a_{i,j} \cdot X_{i,j}$$

donde  $i$  es el número del componente,  $j$  es el índice de la variable,  $n$  es un número menor o igual que el número de las variables, puesto que algunos de los valores  $a_{i,j}$  pueden ser 0. El código que utilizamos para nuestro estudio se encuentra reproducido en el Anexo 4.

#### **Seleccionando el número de componentes que se han de extraer**

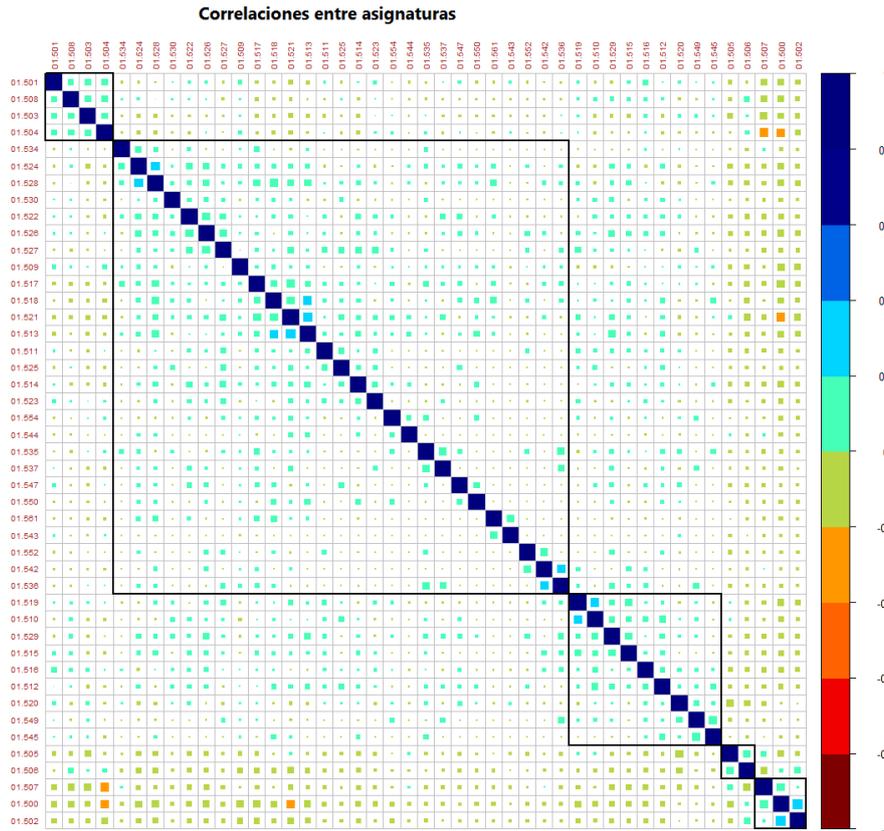
No existe un claro criterio a la hora de elegir el número de componentes que se han de extraer de un conjunto de datos, algunos de los posibles pueden ser:

- Basándose en la experiencia y la teoría.
- Buscando que los componentes cumplan con la explicación de la varianza en un tanto por ciento superior a un límite, por ejemplo el 80%.
- Examinando las cantidades de correlaciones que existen entre las variables.

Nos basaremos en los resultados obtenidos anteriormente en la descomposición en bases con el NMF y apreciaremos los resultados obtenidos. Por lo tanto, tomaremos un número de 5 componentes principales para realizar su estudio.

#### **Obtención de la matriz de correlaciones**

Para realizar esta técnica de descomposición en Componentes Principales, primero tendremos que observar la matriz de correlaciones, que nos muestra cómo de relacionadas se encuentran las asignaturas. Podemos ver en la siguiente gráfica el nivel de agrupamiento:



**Ilustración 14. Gráfica de correlación entre asignaturas con agrupación de 5 clústers**

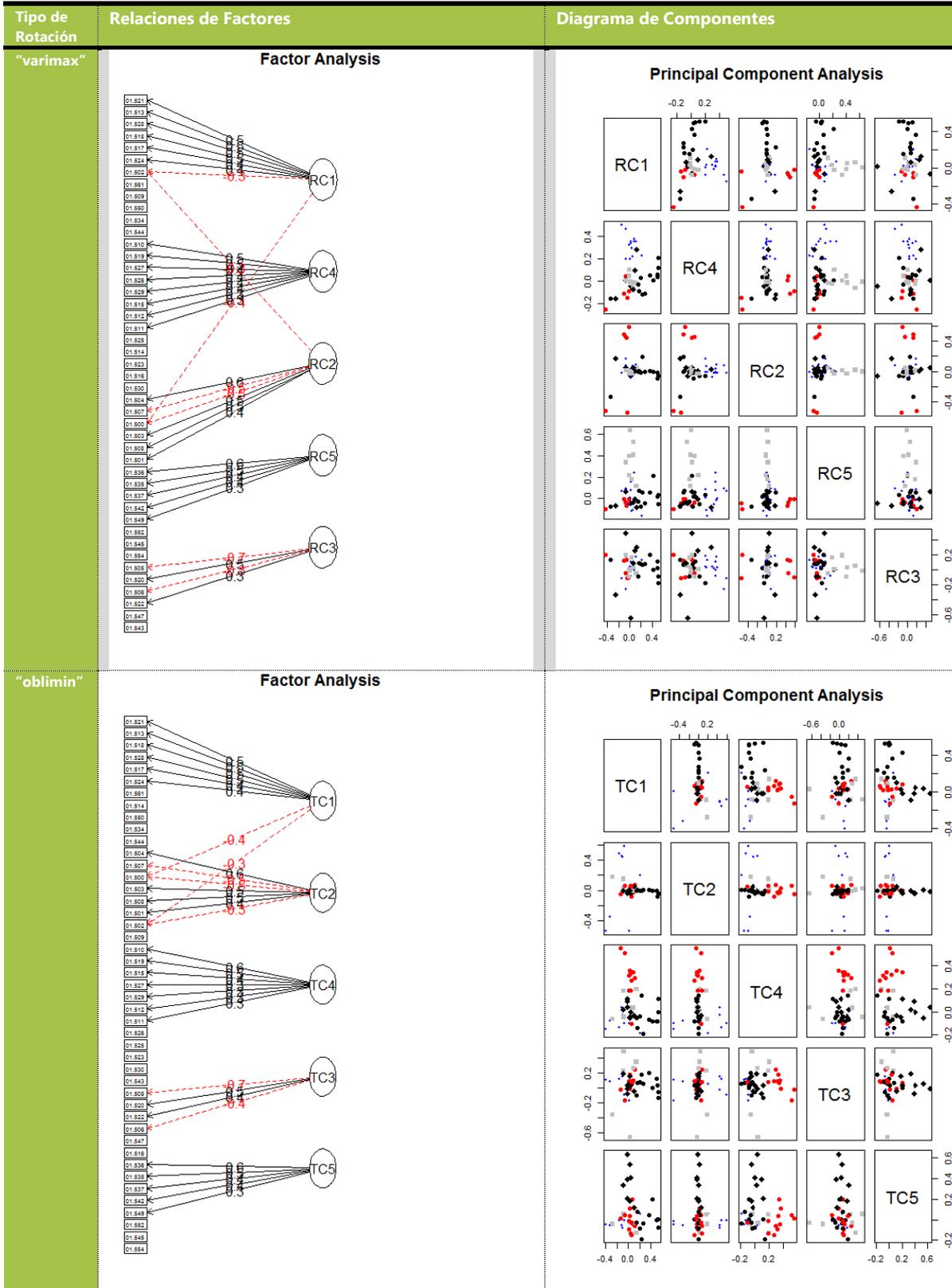
Podemos ver que la agrupación previa de asignaturas más correlacionadas es en los grupos:

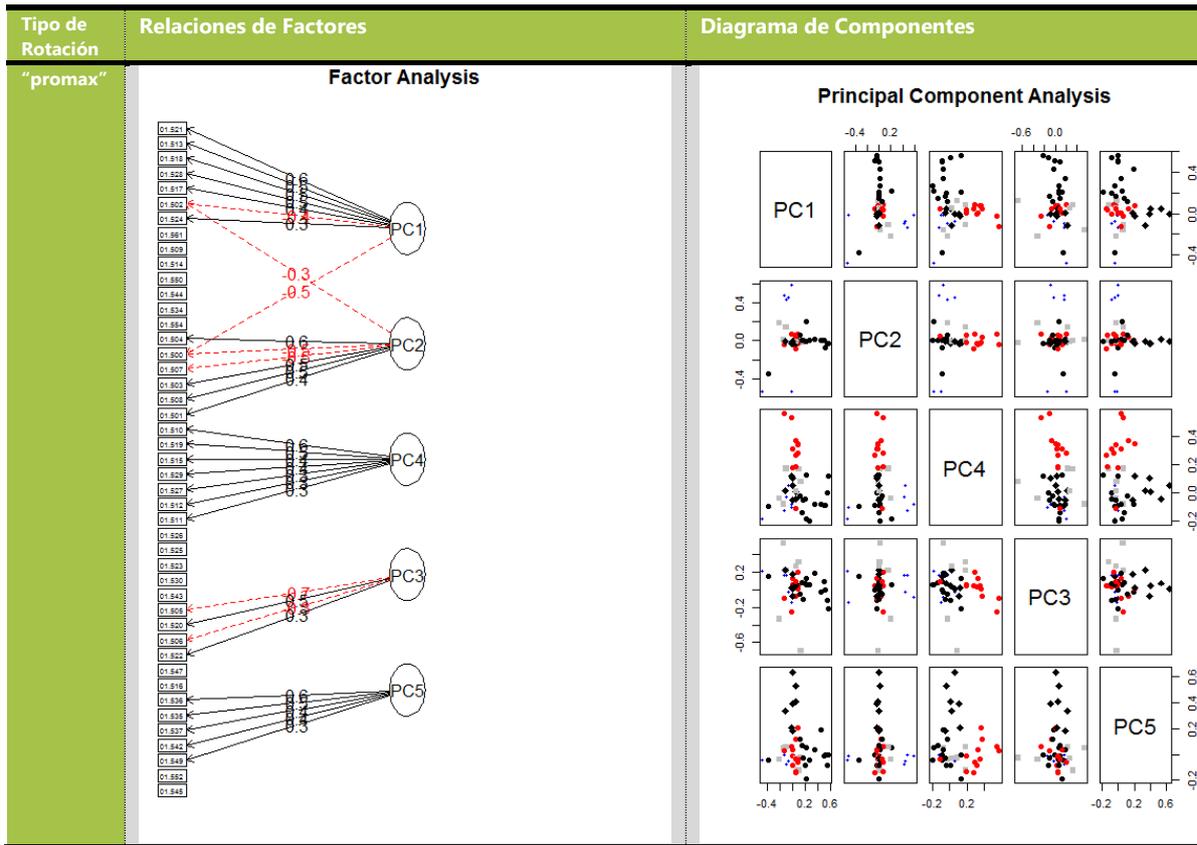
- Grupo 1: 01.500, 01.502 y 01.507
- Grupo 2: 01.505 y 01.506
- Grupo 3: 01.519, 01.510, 01.529, 01.515, 01.516, 01.512, 01.520, 01.549 y 01.545
- Grupo 4: 01.501, 01.503, 01.504 y 01.508
- Grupo 5: todas las demás

### **Usando diferentes tipos de rotación**

Las rotaciones del método PCA consisten en un sistema de técnicas algorítmicas que realizan una depuración de los componentes. La rotación de tipo "varimax" es la más ortogonal que se puede utilizar, también la más común, permite diferenciar las variables según el factor obtenido. La rotación del tipo "oblimin" obtiene unos datos en la que los componentes no son necesariamente ortogonales, sin embargo los valores propios son mayores. Por último la rotación "promax" es una alternativa a la rotación anterior siendo más rápida y mejor para procesar mayor número de datos.

La matriz de correlación anterior es la que se utilizará para realizar las diferentes rotaciones de las que se dispone en el método PCA. En la siguiente tabla vemos la aplicación de las rotaciones expuestas anteriormente para los datos de matriculaciones.





**Conclusiones**

– Tipo de rotación "varimax": según esta rotación tendríamos 5 componentes, cuyos mejores valores mostrarían la relación existente entre las variables:

- RC1. 01.500, 01.502 (en valor negativo) junto con 01.513, 01.518, 01.517, 01.521, 01.524, 01.528 (en valor positivo)
- RC2. 01.500, 01.502, 01.507 (en valor negativo) junto con 01.501, 01.503, 01.504 y 01.508 (en valor positivo)
- RC3. 01.505, 01.506 (en valor negativo) junto con 01.520 y 01.522 (en valor positivo)
- RC4. 01.510, 01.511, 01.512, 01.515, 01.519, 01.526, 01.527, 01.529 (todas ellas positivas)
- RC5. 01.535, 01.536, 01.537, 01.542, 01.549 (todas ellas en valores positivos)

Aunque los valores parecen ser un tanto dispares, dejan entrever que la característica de tomar un valor negativo en la participación del componente conserva cierta relación con aquellas asignaturas que mayor tasa de matriculación tienen, precisamente las cinco primeras de la lista inicial de asignaturas: 01.500, 01.502, 01.505, 01.506 y 01.507 suman un total de 10.216 matrículas, lo que representaría un 69.66% de las observaciones sobre las matrículas.

Parece ser que estas asignaturas agrupadas según los componentes principales RC1, RC2 y RC3 serían las candidatas a matricularlas de manera conjunta por este porcentaje alto de alumnos.

Resultados correspondientes a la extracción de 5 componentes

## principales mediante la rotación "varimax":

```

Principal Components Analysis
Call: principal(r = resultados_asignaturas_fa_cor, nfactors = 5, rotate = "varimax")
Standardized loadings (pattern matrix) based upon correlation matrix
      RC1  RC4  RC2  RC5  RC3  h2  u2
01.500 -0.43 -0.25 -0.51 -0.09  0.20  0.560  0.44
01.505  0.02 -0.05 -0.06 -0.05 -0.65  0.431  0.57
01.507 -0.04 -0.15 -0.54 -0.04 -0.11  0.327  0.67
01.502 -0.34 -0.15 -0.33 -0.08  0.14  0.271  0.73
01.506 -0.25 -0.15  0.17 -0.07 -0.33  0.234  0.77
01.503 -0.10 -0.11  0.49 -0.01  0.13  0.275  0.73
01.504 -0.02 -0.09  0.58 -0.01 -0.10  0.356  0.64
01.509  0.23 -0.11  0.20  0.05  0.06  0.111  0.89
01.501 -0.06  0.01  0.45 -0.07  0.14  0.226  0.77
01.508 -0.07  0.04  0.45 -0.04 -0.04  0.215  0.79
01.520 -0.07  0.01  0.05  0.07  0.50  0.259  0.74
01.521  0.51  0.21 -0.06  0.03 -0.18  0.339  0.66
01.516 -0.08  0.20  0.18  0.09  0.19  0.122  0.88
01.517  0.42  0.01 -0.01  0.21  0.00  0.222  0.78
01.519 -0.07  0.46  0.08  0.10 -0.26  0.302  0.70
01.524  0.36  0.05  0.01  0.06  0.20  0.174  0.83
01.526  0.10  0.36  0.09 -0.01  0.18  0.178  0.82
01.528  0.50  0.06 -0.01  0.02  0.11  0.261  0.74
01.513  0.51  0.12 -0.06 -0.05 -0.08  0.282  0.72
01.529  0.07  0.35  0.00  0.24  0.00  0.187  0.81
01.518  0.49  0.04 -0.09  0.02  0.03  0.255  0.74
01.514  0.21  0.23  0.00 -0.16  0.14  0.139  0.86
01.512  0.01  0.34 -0.02  0.08  0.11  0.136  0.86
01.510 -0.15  0.50 -0.02  0.07 -0.11  0.294  0.71
01.515  0.02  0.35  0.04  0.16 -0.08  0.156  0.84
01.522  0.12  0.28  0.03 -0.08  0.30  0.193  0.81
01.511 -0.01  0.32 -0.01 -0.05  0.03  0.108  0.89
01.527  0.07  0.38 -0.01  0.01  0.04  0.147  0.85
01.523  0.04  0.22  0.07 -0.11  0.04  0.067  0.93
01.530  0.01  0.20  0.00  0.01  0.05  0.041  0.96
01.535  0.06 -0.06  0.02  0.53  0.06  0.294  0.71
01.554  0.11 -0.01  0.05  0.12 -0.05  0.031  0.97
01.549 -0.08  0.01  0.02  0.34  0.20  0.161  0.84
01.525  0.04  0.30 -0.07 -0.11  0.04  0.109  0.89
01.537  0.05 -0.01 -0.02  0.41  0.03  0.172  0.83
01.547  0.09  0.10  0.01 -0.01  0.26  0.083  0.92
01.542 -0.01  0.05 -0.03  0.40 -0.08  0.169  0.83
01.534  0.16 -0.09  0.00  0.07  0.09  0.047  0.95
01.545  0.03 -0.02  0.01  0.18  0.17  0.064  0.94
01.552 -0.02  0.10 -0.02  0.22  0.01  0.059  0.94
01.561  0.27 -0.11  0.00 -0.04  0.08  0.094  0.91
01.550  0.20  0.03 -0.01 -0.07  0.08  0.051  0.95
01.544  0.15 -0.03 -0.01 -0.03 -0.09  0.032  0.97
01.536  0.00  0.00  0.00  0.64 -0.01  0.408  0.59
01.543  0.06 -0.07  0.05 -0.03  0.09  0.020  0.98

      SS loadings          RC1  RC4  RC2  RC5  RC3
Proportion Var          0.05  0.04  0.04  0.03  0.03
Cumulative Var          0.05  0.09  0.13  0.16  0.19
Proportion Explained    0.24  0.21  0.21  0.17  0.17
Cumulative Proportion   0.24  0.45  0.66  0.83  1.00

Test of the hypothesis that 5 components are sufficient.

The degrees of freedom for the null model are 990 and the objective function was 2.25
The degrees of freedom for the model are 775 and the objective function was 1.56

Fit based upon off diagonal values = 0.17

```

- Tipo de rotación "oblimin": esta rotación nos ofrece grupos idénticos a los observados con la rotación "varimax".

## Resultados correspondientes a la extracción de 5 componentes principales mediante la rotación "oblimin":

```

Principal Components Analysis
Call: principal(r = resultados_asignaturas_fa_cor, nfactors = 5, rotate = "oblimin")
Standardized loadings (pattern matrix) based upon correlation matrix
      TC1  TC2  TC4  TC3  TC5  h2  u2
01.500 -0.40 -0.52 -0.15  0.12 -0.05  0.560  0.44
01.505  0.03 -0.03  0.04 -0.65 -0.03  0.431  0.57
01.507  0.01 -0.53 -0.09 -0.17 -0.01  0.327  0.67
01.502 -0.32 -0.34 -0.07  0.09 -0.05  0.271  0.73
01.506 -0.28  0.18 -0.06 -0.35 -0.04  0.234  0.77
01.503 -0.16  0.49 -0.13  0.13  0.00  0.275  0.73
01.504 -0.08  0.59 -0.10 -0.09 -0.01  0.356  0.64
01.509  0.20  0.21 -0.18  0.05  0.06  0.111  0.89
01.501 -0.11  0.44 -0.02  0.16 -0.08  0.226  0.77
01.508 -0.11  0.45  0.04 -0.02 -0.05  0.215  0.79
01.520 -0.08  0.03 -0.04  0.49  0.06  0.259  0.74
01.521  0.53 -0.05  0.11 -0.14  0.00  0.339  0.66
01.516 -0.09  0.16  0.18  0.24  0.06  0.122  0.88
01.517  0.42  0.00 -0.08  0.01  0.20  0.222  0.78
01.519 -0.06  0.06  0.51 -0.16  0.05  0.302  0.70

```

```

01.524 0.36 0.01 -0.06 0.21 0.04 0.174 0.83
01.526 0.11 0.07 0.29 0.25 -0.06 0.178 0.82
01.528 0.51 0.00 -0.08 0.13 0.00 0.261 0.74
01.513 0.53 -0.04 0.00 -0.06 -0.08 0.282 0.72
01.529 0.08 -0.02 0.34 0.07 0.20 0.187 0.81
01.518 0.51 -0.08 -0.08 0.04 0.00 0.255 0.74
01.514 0.23 -0.01 0.14 0.18 -0.19 0.139 0.86
01.512 0.03 -0.04 0.32 0.18 0.03 0.136 0.86
01.510 -0.13 -0.04 0.55 -0.02 0.01 0.294 0.71
01.515 0.03 0.03 0.36 -0.01 0.11 0.156 0.84
01.522 0.14 0.01 0.20 0.35 -0.12 0.193 0.81
01.511 0.01 -0.02 0.31 0.09 -0.09 0.108 0.89
01.527 0.09 -0.03 0.35 0.11 -0.04 0.147 0.85
01.523 0.05 0.07 0.19 0.08 -0.13 0.067 0.93
01.530 0.02 -0.01 0.19 0.08 -0.01 0.041 0.96
01.535 0.03 0.01 -0.05 0.05 0.53 0.294 0.71
01.554 0.10 0.06 -0.03 -0.04 0.12 0.031 0.97
01.549 -0.10 0.00 0.02 0.20 0.34 0.161 0.84
01.525 0.06 -0.08 0.28 0.09 -0.15 0.109 0.89
01.537 0.04 -0.03 0.00 0.03 0.41 0.172 0.83
01.547 0.09 0.00 0.04 0.27 -0.03 0.083 0.92
01.542 -0.02 -0.04 0.09 -0.07 0.39 0.169 0.83
01.534 0.16 0.01 -0.14 0.07 0.08 0.047 0.95
01.545 0.02 0.00 -0.04 0.16 0.18 0.064 0.94
01.552 -0.02 -0.03 0.12 0.03 0.21 0.059 0.94
01.561 0.27 0.01 -0.19 0.06 -0.03 0.094 0.91
01.550 0.20 -0.01 -0.04 0.08 -0.08 0.051 0.95
01.544 0.15 0.00 -0.05 -0.09 -0.02 0.032 0.97
01.536 -0.02 -0.01 0.04 -0.01 0.64 0.408 0.59
01.543 0.05 0.05 -0.10 0.08 -0.03 0.020 0.98

TC1 TC2 TC4 TC3 TC5
SS loadings 2.17 1.81 1.70 1.52 1.47
Proportion Var 0.05 0.04 0.04 0.03 0.03
Cumulative Var 0.05 0.09 0.13 0.16 0.19
Proportion Explained 0.25 0.21 0.20 0.18 0.17
Cumulative Proportion 0.25 0.46 0.66 0.83 1.00

With component correlations of
TC1 TC2 TC4 TC3 TC5
TC1 1.00 0.09 0.19 -0.01 0.08
TC2 0.09 1.00 0.11 -0.01 0.05
TC4 0.19 0.11 1.00 -0.04 0.06
TC3 -0.01 -0.01 -0.04 1.00 0.03
TC5 0.08 0.05 0.06 0.03 1.00

Test of the hypothesis that 5 components are sufficient.

The degrees of freedom for the null model are 990 and the objective function was 2.25
The degrees of freedom for the model are 775 and the objective function was 1.56

Fit based upon off diagonal values = 0.17
    
```

- Tipo de rotación "promax" esta rotación nos ofrece grupos idénticos a los observados con la rotación "varimax":

Resultados correspondientes a la extracción de 5 componentes principales mediante la rotación "promax":

```

Principal Components Analysis
Call: principal(r = resultados_asignaturas_fa_cor, nfactors = 5, rotate = "promax")
Standardized loadings (pattern matrix) based upon correlation matrix
PC1 PC2 PC4 PC3 PC5 h2 u2
01.500 -0.48 -0.54 -0.18 0.21 -0.05 0.560 0.44
01.505 0.12 -0.01 0.09 -0.69 -0.02 0.431 0.57
01.507 -0.02 -0.53 -0.10 -0.14 -0.01 0.327 0.67
01.502 -0.38 -0.34 -0.09 0.15 -0.05 0.271 0.73
01.506 -0.22 0.19 -0.03 -0.33 -0.04 0.234 0.77
01.503 -0.13 0.48 -0.13 0.16 0.00 0.275 0.73
01.504 -0.01 0.59 -0.08 -0.08 -0.01 0.356 0.64
01.509 0.22 0.21 -0.18 0.06 0.05 0.111 0.89
01.501 -0.09 0.43 -0.02 0.17 -0.08 0.226 0.77
01.508 -0.07 0.45 0.05 -0.02 -0.05 0.215 0.79
01.520 -0.15 0.02 -0.07 0.53 0.05 0.259 0.74
01.521 0.56 -0.03 0.12 -0.22 0.00 0.339 0.66
01.516 -0.11 0.15 0.18 0.22 0.06 0.122 0.88
01.517 0.44 0.01 -0.08 -0.02 0.19 0.222 0.78
01.519 -0.02 0.07 0.54 -0.25 0.06 0.302 0.70
01.524 0.34 0.01 -0.08 0.19 0.04 0.174 0.83
01.526 0.09 0.07 0.28 0.19 -0.06 0.178 0.82
01.528 0.50 0.00 -0.09 0.09 -0.01 0.261 0.74
01.513 0.54 -0.03 0.01 -0.12 -0.08 0.282 0.72
01.529 0.08 -0.01 0.35 0.01 0.20 0.187 0.81
01.518 0.51 -0.07 -0.08 -0.01 0.00 0.255 0.74
01.514 0.21 -0.01 0.13 0.13 -0.19 0.139 0.86
01.512 0.00 -0.04 0.32 0.13 0.04 0.136 0.86
01.510 -0.13 -0.04 0.56 -0.10 0.03 0.294 0.71
01.515 0.04 0.03 0.37 -0.07 0.12 0.156 0.84
01.522 0.09 0.01 0.18 0.31 -0.12 0.193 0.81
01.511 0.00 -0.02 0.31 0.03 -0.08 0.108 0.89
01.527 0.08 -0.02 0.35 0.04 -0.04 0.147 0.85
01.523 0.04 0.06 0.19 0.04 -0.13 0.067 0.93
    
```

```

01.530 0.00 -0.01 0.18 0.05 -0.01 0.041 0.96
01.535 0.04 0.02 -0.04 0.07 0.53 0.294 0.71
01.554 0.12 0.06 -0.02 -0.05 0.12 0.031 0.97
01.549 -0.12 0.00 0.02 0.22 0.34 0.161 0.84
01.525 0.04 -0.08 0.27 0.04 -0.14 0.109 0.89
01.537 0.05 -0.03 0.01 0.04 0.41 0.172 0.83
01.547 0.05 -0.01 0.02 0.26 -0.03 0.083 0.92
01.542 0.00 -0.03 0.11 -0.08 0.39 0.169 0.83
01.534 0.15 0.01 -0.14 0.08 0.07 0.047 0.95
01.545 0.00 0.00 -0.05 0.18 0.18 0.064 0.94
01.552 -0.02 -0.03 0.12 0.02 0.21 0.059 0.94
01.561 0.27 0.01 -0.20 0.06 -0.04 0.094 0.91
01.550 0.19 -0.01 -0.04 0.07 -0.09 0.051 0.95
01.544 0.17 0.01 -0.05 -0.11 -0.03 0.032 0.97
01.536 0.00 -0.01 0.06 0.00 0.64 0.408 0.59
01.543 0.05 0.05 -0.11 0.09 -0.03 0.020 0.98

SS loadings          PC1  PC2  PC4  PC3  PC5
Proportion Var      2.19 1.79 1.72 1.50 1.46
Cumulative Var      0.05 0.04 0.04 0.03 0.03
Proportion Explained 0.25 0.21 0.20 0.17 0.17
Cumulative Proportion 0.25 0.46 0.66 0.83 1.00

With component correlations of
PC1  PC2  PC4  PC3  PC5
PC1  1.00 -0.02 0.22 0.25 0.05
PC2 -0.02 1.00 0.06 0.03 0.03
PC4  0.22 0.06 1.00 0.22 0.01
PC3  0.25 0.03 0.22 1.00 0.02
PC5  0.05 0.03 0.01 0.02 1.00

Test of the hypothesis that 5 components are sufficient.

The degrees of freedom for the null model are 990 and the objective function was 2.25
The degrees of freedom for the model are 775 and the objective function was 1.56

Fit based upon off diagonal values = 0.17

```

### 7.3.1. Método K-means – Agrupamiento de datos

El método k-means es un método de agregación de datos el cual utilizando algoritmos de similitud consigue reunir objetos que presentan similitudes entre ellos, creando de esta manera los denominados clústers, que ofrecen una idea de cómo de parecidos son determinadas variables u observaciones.

Realizaremos el método k-means sobre los datos de las matriculaciones para observar cómo se reparten las asignaturas en paquetes o clústers cuando se realizan divisiones sobre ellas. De esta manera, para este método y siguiendo con la línea anterior, intentaremos hacer 5 ó 6 separaciones diferentes de los datos y observaremos cuál produce un agrupamiento más explicativo. El código que utilizamos para nuestro estudio se encuentra reproducido en el Anexo 5.

#### **Agrupación de asignaturas en 5 clústers**

Tomaremos la matriz de matriculaciones y aplicaremos el algoritmo k-means, que nos ofrecerá el siguiente resultado:

Asignatura	Matriculados	Clúster
01.500	2581	5
01.505	2339	1
01.507	2271	2
01.502	1578	5
01.506	1447	4
01.503	691	4
01.504	672	4
01.509	497	3
01.501	350	3
01.508	275	3
01.520	265	3

Asignatura	Matriculados	Clúster
01.521	230	3
01.516	155	3
01.517	161	3
01.519	116	3
01.524	93	3
01.526	89	3
01.528	88	3
01.513	86	3
01.529	72	3
01.518	65	3
01.514	59	3
01.512	55	3
01.510	51	3
01.515	49	3
01.522	52	3
01.511	33	3
01.527	30	3
01.523	28	3
01.530	24	3
01.535	26	3
01.554	20	3
01.549	13	3
01.525	12	3
01.537	12	3
01.547	12	3
01.542	4	3
01.534	11	3
01.545	8	3
01.552	9	3
01.561	9	3
01.550	9	3
01.544	9	3
01.536	6	3
01.543	4	3
01.550	9	3
01.544	9	3
01.536	6	3
01.543	4	3

Estos datos se pueden resumir en el siguiente cuadro:

Clúster	Matriculaciones	Matr_%	Asignaturas
1	2339	15,95	01.505
2	2271	15,48	01.507
3	3087	21,05	otras
4	2810	19,16	01.503, 01.504, 01.506
5	4159	28,36	01.500, 01.502

### Conclusiones

Podemos ver cómo el clúster número 5 aglomera la mayor proporción de matriculados con un total del 28,36% con tan solo las asignaturas 01.500 y 01.502, por lo que se deduce que son las más matriculadas y similares. El clúster número 3 representa a aquellas asignaturas que se aglutinan a partir de la octava asignatura, la 01.509, con un porcentaje de matriculaciones de 21,05%. El clúster número 4

corresponde a tres asignaturas, las 01.503, 01.504 y 01.506 con un total de un 19,16%. El primero y segundo clúster corresponden a asignaturas únicas por lo que sus porcentajes corresponden con los datos originales de matriculaciones, así se puede ver que las asignaturas 01.505 y 01.507 son asignaturas muy "preferidas" para matricularse.

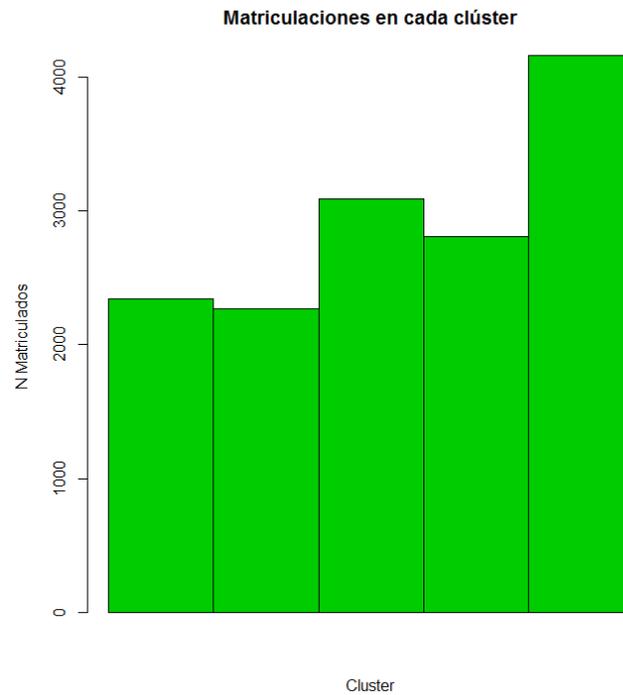


Ilustración 15. Representación del total de matriculados en cada clúster (5)

**Agrupación de asignaturas en 6 clústers**

Si decidiéramos tomar una agrupación de seis grupos diferentes de asignaturas con mucha similitud entre ellas, obtendríamos los siguientes datos:

Asignatura	Matriculados	Clúster
01.500	2581	4
01.505	2339	3
01.507	2271	2
01.502	1578	4
01.506	1447	6
01.503	691	1
01.504	672	1
01.509	497	5
01.501	350	5
01.508	275	5
01.520	265	5
01.521	230	5
01.516	155	5
01.517	161	5
01.519	116	5
01.524	93	5
01.526	89	5
01.528	88	5
01.513	86	5
01.529	72	5

Asignatura	Matriculados	Clúster
01.518	65	5
01.514	59	5
01.512	55	5
01.510	51	5
01.515	49	5
01.522	52	5
01.511	33	5
01.527	30	5
01.523	28	5
01.530	24	5
01.535	26	5
01.554	20	5
01.549	13	5
01.525	12	5
01.537	12	5
01.547	12	5
01.542	4	5
01.534	11	5
01.545	8	5
01.552	9	5
01.561	9	5
01.550	9	5
01.544	9	5
01.536	6	5
01.543	4	5

Estos datos se pueden simplificar de la siguiente manera:

Clúster	Matriculaciones	Matr_%	Asignaturas
1	1363	9,29	01.503, 01.504
2	2271	15,48	01.507
3	2339	15,95	01.505
4	4159	28,36	01.500, 01.502
5	3087	21,05	otras
6	1447	9,87	01.506

### Conclusiones

Cuando se selecciona un mayor número de separaciones se puede obtener la separación de clústers que agrupan demasiados datos de una vez. En el caso de utilizar seis clúster tenemos que la mayoría de matriculaciones se ven reflejados en el clúster 4 las asignaturas 01.500 y 01.502, que al igual que anteriormente aglomeran el 28,36% de los datos, ambas asignaturas siguen siendo las más similares en matriculaciones. Ahora el clúster número 5 recoge las asignaturas de la 8ª en adelante con un porcentaje total de 21,05% de matriculaciones sobre el total. Hemos conseguido separar un clúster más, el número 1 de modo que agrupa las asignaturas 01.503 y 01.504 en un 9,29% de matriculaciones. Y obtenemos el clúster 6 con una única asignatura, junto con el segundo y el tercero, es decir, las asignaturas 01.505, 01.506 y 01.507 se diferencian bien del resto según sus matriculaciones que tienen gran peso.

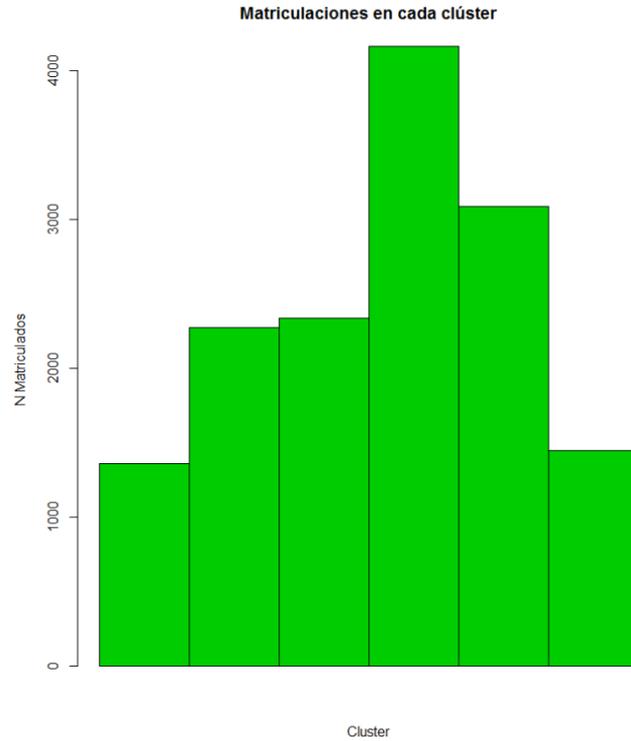


Ilustración 16. Representación del total de matriculados en cada clúster (6)

**Agrupación de los alumnos en 6 clústers**

Hasta el momento se ha visto la relación existente entre las matrículas que se efectúan en el total de las asignaturas y cómo estas ofrecen una visión de cómo se comportan en tanto a similitud entre matriculaciones. A continuación nos centraremos en cómo los estudiantes realizan sus selecciones de matrícula y cómo ésta selección puede alterar al éxito/fracaso del mismo. Para ello tomaremos la decisión de agrupar las observaciones que tenemos en 6 grupos de similitud o clústers sobre los alumnos, utilizaremos el algoritmo proporcionado por R 'kmeans', los detalles del código se encuentran en el Anexo 6.

La ejecución del algoritmo nos clasifica cada estudiante en un clúster diferente, el resultado nos ofrece la siguiente tabla el clúster en la que cada alumno se sitúa:

[1]	3 3 3 2 1 3 2 3 5 4 3 2 5 2 3 4 1 6 6 1 2 2 1 4 2 1 3 5 1 2 6 1 6 2 1 2 4 1 4 6 4 3 3 4 1 1 1 2 2 2 1 5 2 6 4 1 1 1 1 6 1 2 1
[65]	1 1 1 4 3 4 5 4 3 3 2 4 1 1 1 6 6 4 6 6 6 4 4 1 3 6 3 5 3 5 5 1 4 4 6 1 3 6 4 1 2 1 4 2 6 1 4 3 5 6 6 1 5 3 1 4 5 2 2 1 1 2 5 1 4
[129]	1 3 1 1 3 6 5 3 2 2 5 2 4 1 3 4 1 1 2 1 1 4 2 5 2 2 2 4 1 4 4 6 6 5 1 6 3 1 3 2 4 6 5 6 3 4 4 2 2 1 4 4 2 4 4 2 2 6 2 6 5 5 6 2
[193]	1 2 2 1 1 2 4 6 5 2 2 2 1 1 2 2 1 3 1 6 2 6 1 1 3 4 1 5 4 6 4 3 1 6 3 4 5 6 6 6 1 4 5 2 2 2 4 4 6 2 6 1 4 4 3 3 5 1 3 3 6 4 4 4
[257]	3 1 2 6 2 4 5 1 4 1 2 3 2 6 1 4 4 1 2 3 2 3 2 1 2 4 6 3 1 5 4 2 2 3 1 6 3 4 1 2 2 4 1 3 6 5 1 5 3 3 4 5 4 6 5 2 3 4 6 6 6 6 1 4
[321]	4 3 6 4 1 5 4 3 3 5 6 2 1 3 1 4 4 4 5 6 4 1 1 1 3 1 3 4 6 6 1 5 6 3 4 3 4 1 1 6 6 6 6 3 1 4 1 1 5 3 2 2 4 5 6 1 2 5 1 2 2 1 2 5
[385]	1 1 5 4 1 5 1 2 4 3 3 4 6 4 1 3 2 6 3 6 1 4 2 1 5 1 4 1 2 1 1 2 2 3 5 4 6 1 2 2 1 3 4 1 1 4 6 2 3 4 2 3 6 2 1 4 4 3 4 3 3 1 3 1
[449]	4 4 6 1 3 4 3 1 6 4 1 4 2 2 4 2 6 4 2 4 3 5 4 6 2 4 3 2 3 4 5 1 4 5 4 1 2 2 4 5 5 4 4 1 4 5 5 4 4 3 2 4 5 5 2 4 1 2 4 6 4 2 4 2
[533]	4 1 1 5 3 4 2 1 1 1 6 5 3 5 4 3 4 6 2 1 2 2 1 1 1 2 1 4 4 3 2 2 5 1 6 5 4 4 6 1 2 6 3 2 3 5 3 1 2 3 4 6 1 1 5 4 5 1 2 1 4 5 3 3
[577]	4 3 4 2 1 5 3 4 2 5 3 1 1 4 2 3 1 2 4 1 1 1 4 4 4 1 6 3 5 4 1 1 6 1 4 3 5 2 3 2 1 4 3 1 1 4 2 1 2 4 3 5 2 5 4 2 4 3 2 1 1
[641]	2 1 2 4 2 1 1 4 4 2 4 4 2 3 2 5 2 1 4 2 3 5 2 1 2 5 1 6 4 3 1 1 6 4 1 6 2 4 4 4 3 4 1 6 4 5 4 3 3 4 5 1 2 3 6 3 3 3 4 5 6 5 3 2
[705]	3 1 3 3 2 6 4 4 4 6 1 1 4 4 3 4 6 2 1 5 2 2 1 4 5 1 2 5 6 1 4 6 2 6 4 1 3 4 5 5 3 4 4 3 2 4 4 1 2 1 4 3 1 2 2 3 4 4 4 1 5 1 6 2
[769]	3 4 1 1 1 2 3 2 2 6 4 5 4 4 5 2 5 4 2 4 6 6 3 2 4 1 5 1 1 3 5 3 1 4 1 3 3 4 1 1 1 2 6 6 4 3 6 4 4 1 2 4 1 4 5 1 1 2 6 4 5 4 2 2
[833]	4 2 5 1 2 4 1 3 5 1 3 1 4 4 4 2 2 5 6 6 5 2 3 2 4 1 4 3 5 3 1 4 1 1 4 4 1 2 4 5 4 2 2 3 2 2 4 3 5 4 3 3 2 1 4 2 2 2 3 3 2 2
[897]	2 3 6 3 3 2 6 4 1 3 2 3 4 3 3 1 2 2 6 5 3 4 5 4 5 3 2 3 5 2 3 2 4 6 3 6 5 1 2 4 5 3 2 1 1 3 2 3 2 6 5 3 1 2 3 5 4 4 2 2 5 3 4
[961]	2 5 4 3 1 3 6 1 1 4 2 3 5 2 4 2 2 6 1 3 5 4 1 3 5 3 4 6 5 2 1 3 1 3 3 3 2 4 5 5 3 3 3 4 3 3 5 4 2 5 3 4 3 3 4 2 1 1 2 6 2 2 3
[1025]	3 2 5 3 4 3 3 1 1 6 2 6 1 5 3 6 2 3 2 3 5 4 3 4 6 5 6 5 3 4 3 6 1 3 2 3 6 5 2 5 3 5 5 3 4 2 2 3 3 5 5 3 1 1 2 1 4 2 5 3 2 4 3 2
[1089]	6 1 4 3 3 4 3 5 2 3 3 5 4 2 2 2 5 1 3 6 3 2 1 4 6 3 1 2 2 5 5 2 1 3 3 6 4 4 1 6 4 1 1 3 1 2 1 5 3 1 4 5 2 3 3 2 2 5 4 3 3 3 3
[1153]	3 6 3 3 3 1 5 4 5 5 3 5 5 4 6 3 1 3 3 3 4 4 1 3 5 3 3 3 1 1 3 3 3 3 4 3 2 4 1 3 4 3 4 4 1 4 4 1 2 1 2 3 3 1 1 5 6 3 4 6 3 1 1
[1217]	1 1 1 2 5 1 6 1 1 4 4 6 3 6 5 4 3 2 1 2 4 6 5 4 3 2 4 1 1 2 4 4 4 6 4 2 6 3 2 5 6 1 5 2 5 2 4 6 4 4 2 3 1 1 2 1 2 2 3 2 4 6 3 4
[1281]	4 5 6 3 3 5 4 2 4 6 3 3 6 6 2 2 6 5 4 2 5 2 3 6 6 4 1 3 5 3 1 1 5 2 3 2 4 6 1 3 2 3 2 3 2 2 1 6 2 2 1 6 1 3 1 3 2 6 3 5 3 6 3 1
[1345]	5 3 1 2 1 4 3 3 2 1 1 2 2 3 3 6 5 1 2 2 1 2 2 2 5 2 3 3 3 1 2 6 2 1 2 2 4 3 2 5 1 3 3 5 3 2 2 2 4 6 4 2 4 2 5 1 1 3 4 6 2 3 6
[1409]	2 5 4 5 3 1 5 2 6 5 4 1 1 2 2 6 5 4 1 1 6 3 4 2 1 2 6 4 1 3 5 2 4 1 5 2 2 4 5 2 2 3 6 6 1 4 2 4 5 5 6 2 2 4 4 1
[1473]	2 3 6 6 2 4 1 4 1 3 3 4 3 5 2 2 4 2 5 4 5 5 4 2 2 5 2 1 3 3 2 3 2 2 4 5 3 5 5 3 6 5 1 1 4 3 3 2 2 6 2 2 4 5 5 2 2 5 1 6 6 5 2
[1537]	3 5 3 4 5 3 2 2 4 2 3 4 4 5 4 1 2 2 4 2 2 3 1 3 4 5 4 3 5 4 1 1 6 3 2 3 3 4 3 3 5 4 2 3 2 2 5 1 5 4 6 3 4 6 2 4 5 1 1 2 3 5 3 1
[1601]	6 4 4 3 6 4 6 3 2 3 3 4 1 2 2 3 2 4 4 1 1 2 6 3 2 1 3 5 4 2 2 6 4 4 3 1 1 6 3 4 2 1 2 6 4 4 6 4 5 3 5 3 4 3 6 6 4 2 2 5 6 3 4
[1665]	1 5 2 4 6 2 1 4 2 2 3 1 6 3 3 2 5 5 2 1 1 5 3 4 4 4 6 3 2 6 4 1 6 4 3 2 5 4 4 2 4 2 3 2 6 2 4 3 5 3 6 6 2 3 4 4 2 1 2 2 3 4 2
[1729]	2 2 1 3 6 2 6 5 3 4 2 6 1 4 6 2 5 4 3 1 4 2 4 2 2 4 6 6 1 2 2 2 3 6 1 3 1 6 6 1 2 6 6 3 5 3 2 4 3 3 4 4 2 1 1 6 2 4 3 1 1 2 2
[1793]	5 4 4 2 6 3 2 2 3 4 2 4 2 1 1 1 1 5 3 1 4 5 5 2 4 4 2 2 1 2 2 6 4 2 2 2 3 4 5 3 4 2 1 4 6 4 1 2 4 3 2 2 1 4 2 2 5 4 3 6 4
[1857]	2 4 3 2 4 3 6 6 2 2 6 5 1 3 4 2 1 1 1 4 1 4 5 3 1 3 1 4 1 1 3 3 4 5 4 5 6 1 5 5 6 5 2 3 4 5 3 3 2 2 1 3 4 1 3 6 5 6 5 3 2 5 6 3
[1921]	5 3 1 6 6 6 2 2 4 1 1 1 1 1 2 1 3 5 3 1 3 2 2 5 4 2 1 4 4 1 1 6 6 6 4 4 6 1 4 1 4 2 6 1 3 4 5 2 4 5 1 1 2 4 4 1 4 1 6 4 4 1 1 1
[1985]	3 1 1 1 4 4 4 1 3 5 1 1 6 3 6 2 1 3 3 5 1 2 1 2 6 1 2 1 1 2 4 1 6 5 5 6 1 1 1 4 1 1 3 1 6 6 3 5 5 5 1 2 6 3 1 5 1 1 5 1 4 4 3 4
[2049]	4 4 4 4 1 1 3 3 4 3 2 2 2 4 2 3 5 1 5 3 2 2 3 2 5 2 6 3 4 6 4 3 5 1 1 3 3 2 4 3 4 3 6 3 6 4 5 3 4 2 4 1 6 6 1 1 2 1 1 5 6 2 4 1
[2113]	1 2 4 4 6 4 3 5 4 6 3 5 5 3 4 6 6 4 6 1 3 5 4 1 6 4 4 1 3 5 6 5 1 2 3 1 2 1 2 4 2 4 4 1 1 5 2 2 1 3 5 4 6 6 3 1 2 4 3 1 1 1
[2177]	2 3 2 4 2 2 3 4 4 3 3 6 1 6 5 6 6 3 5 4 6 3 4 3 1 1 6 1 3 2 4 1 2 5 2 3 2 2 1 3 3 4 6 4 1 6 6 4 1 5 5 6 5 3 4 4 6 5 2 1 4 5 2
[2241]	3 2 5 2 1 4 1 3 1 5 6 3 6 3 1 4 6 2 3 2 3 4 4 1 6 5 1 3 4 1 4 6 1 1 1 3 2 4 4 2 1 6 4 4 4 2 5 5 1 1 1 5 3 1 5 5 4 2 4 5 3 5

[2305]	2	4	2	3	6	3	1	4	3	3	3	5	2	1	5	6	2	1	1	1	2	3	1	4	1	3	2	5	3	3	1	4	3	2	4	1	6	6	3	1	4	2	3	5	1	3	2	2	1	1	3	4	1	5	4	3	4	5	6	2	1	2	2	1			
[2369]	3	1	1	1	1	3	1	2	1	6	3	3	2	3	2	4	4	6	4	1	3	6	1	5	1	6	6	4	5	4	2	1	4	2	4	4	5	2	4	2	3	2	2	3	1	2	4	6	2	5	3	1	3	6	3	2	6	1	2	4	2	1	1	1			
[2433]	5	1	2	3	3	5	1	5	1	1	3	5	1	2	2	5	6	5	4	3	5	3	3	5	4	3	2	3	4	4	2	3	2	2	4	3	2	2	3	2	2	4	2	4	1	4	2	2	2	2	2	3	3	2	3	3	2	4	3	5							
[2497]	3	2	2	5	2	4	5	4	5	1	2	2	5	5	2	2	3	3	2	2	3	5	4	3	3	4	4	2	4	4	2	5	5	5	5	4	2	3	4	5	3	3	4	2	3	4	2	5	4	3	5	5	4	2	5	2	1	4	5	4	2	4					
[2561]	2	2	2	4	2	4	4	2	1	4	1	5	3	5	2	2	1	5	4	4	6	3	5	3	6	3	6	5	5	4	1	4	3	6	1	2	2	5	3	2	6	1	4	3	3	6	5	4	3	3	4	1	5	5	5	1	1	3	3	1	5	1					
[2625]	4	5	1	1	3	3	3	2	3	1	4	4	1	4	4	1	4	5	4	4	1	1	3	1	1	6	2	3	3	4	3	3	5	5	2	3	5	5	5	2	2	3	4	2	3	2	1	2	3	1	1	1	1	2	4	5	4	4	2	1	4	1	1	3	1		
[2689]	5	5	6	6	1	2	2	1	3	2	4	3	6	4	1	2	1	1	4	4	5	1	1	4	2	1	4	4	6	1	4	5	3	5	1	6	5	4	3	5	1	5	5	3	3	4	3	5	5	1	4	1	4	2	6	1	5										
[2753]	6	6	5	3	2	4	6	2	6	6	5	4	5	5	6	5	2	2	5	2	1	1	4	2	2	4	6	4	2	2	4	1	1	3	6	1	1	1	5	4	5	4	4	5	5	3	1	1	2	1	4	5	3	3	4	4	4	6	5	1	2						
[2817]	6	3	4	3	6	3	1	3	5	4	5	2	4	4	6	2	5	1	3	5	4	6	1	6	3	5	6	2	3	1	5	2	5	3	6	6	2	3	4	1	1	4	1	5	3	5	2	4	1	1	2	5	2	4	4	1	5	3	2	5	1	1					
[2881]	3	2	4	1	5	6	5	3	6	3	4	4	6	5	3	2	6	6	5	2	1	2	3	4	1	4	2	2	4	1	3	1	4	5	6	5	1	6	1	3	6	1	3	2	1	4	1	2	2	5	4	4	3	2	2	6	5	4	5	2	2	5	3	3			
[2945]	1	5	2	1	6	4	4	2	6	2	6	4	5	1	6	5	4	6	6	1	5	6	3	6	4	5	1	2	3	4	2	2	1	6	2	2	1	3	1	6	1	4	3	4	1	1	2	4	1	4	3	2	3	2	2	3	5	4	4	4	1	5	4				
[3009]	5	4	5	3	5	6	3	2	5	2	5	3	5	4	6	3	4	1	5	1	3	1	1	3	1	1	3	5	1	1	1	4	6	1	2	5	1	2	4	2	2	1	3	3	6	6	2	1	1	2	4	4	2	2	2	5	2	3	3	6	3	1	1				
[3073]	5	4	5	5	1	1	1	3	6	5	2	6	1	2	5	1	3	1	2	1	5	5	1	1	1	4	5	6	3	2	5	5	2	4	4	1	3	2	6	3	2	1	3	6	3	3	1	6	6	4	2	5	1	4	5	5	5	1	2	3	5						
[3137]	1	1	5	2	1	3	2	6	2	2	6	5	1	1	1	5	6	1	3	5	1	5	1	2	1	1	5	5	4	1	2	1	3	1	2	1	5	5	3	4	1	6	5	2	3	5	1	6	3	1	4	5	3	5	5	4	1	5	5	4	5	5	3				
[3201]	4	5	3	6	5	1	1	1	5	4	5	6	4	5	6	2	1	3	1	3	1	6	1	3	4	4	1	6	4	1	1	3	1	3	6	3	1	4	3	1	5	4	6	2	6	5	3	2	4	5	2	4	4	3	3	5	4	1	5	3	3	5	1				
[3265]	6	1	4	4	3	3	2	3	4	3	3	5	2	2	1	1	4	4	1	5	2	1	6	5	1	1	2	1	6	1	5	1	1	1	1	2	4	2	5	6	2	4	3	5	3	5	3	6	4	2	1	1	5	1	1	2	4	6	6	4	4	6	6				
[3329]	4	2	2	5	3	4	2	1	6	4	2	3	2	5	4	5	1	5	3	1	5	6	5	4	2	5	2	2	2	1	1	6	6	3	6	3	1	2	3	2	2	5	2	2	1	4	1	5	2	1	4	2	5	1	1	5	5	2	6	2	6	1	3	6			
[3393]	5	1	2	1	4	6	3	1	1	2	4	4	3	3	5	6	3	3	1	1	1	1	1	1	4	5	2	1	2	1	1	2	6	4	3	4	1	5	2	4	5	5	4	3	1	2	2	4	1	5	5	1	3	6	5	3	4	4	4	5	2	6	5	1			
[3457]	1	1	2	1	1	6	5	2	1	2	2	1	5	5	1	1	2	1	6	4	3	3	1	1	2	4	1	1	6	6	4	1	1	1	1	5	2	6	4	3	4	1	5	5	3	4	2	5	4	3	6	5	5	3	1	3	1	1	2	5	1	6	2	4			
[3521]	2	5	2	6	1	5	2	1	5	2	4	4	4	4	4	4	6	1	5	3	4	1	6	4	4	3	1	6	6	2	1	3	2	5	6	5	2	6	3	5	3	6	3	2	1	1	1	1	3	4	2	1	3	5	2	5	6	4	5	3	3	3	1				
[3585]	2	4	5	6	6	5	2	1	3	5	5	6	1	3	4	6	5	4	1	2	6	6	5	2	1	1	5	4	1	6	4	3	3	1	1	6	5	2	1	5	1	5	2	2	3	3	2	5	3	2	3	1	5	5	6	6	5	1	5	1	4	1	1	3	6		
[3649]	2	4	5	6	3	6	6	2	1	6	2	4	4	1	1	3	5	4	5	1	2	6	2	2	4	3	1	5	2	6	3	1	5	2	6	6	1	6	6	5	5	6	6	6	5	6	1	4	1	3	6	4	1	1	5	6	2	4	2	6	2	5	5	4	4	6	3
[3713]	3	6	3	3	1	2	6	1	4	5	5	4	5	1	2	6	4	4	2	4	5	5	6	2	5	2	2	2	6	2	4	1	3	6	3	2	1	5	4	3	2	3	1	4	6	5	6	4	2	3	5	2	6	3	6	2	2	4	6	1	2	1	3	3			
[3777]	3	1	3	5	1	2	1	3	2	1	5	4	1	5	1	5	2	4	6	4	5	3	2	1	4	2	4	1	6	4	4	1	5	5	1	4	1	2	4	2	6	3	5	3	2	6	5	1	5	5	4	3	3	1	4	4	2	4	5	5	1	6	3				
[3841]	5	5	3	2	2	1	1	4	5	4	3	4	3	1	6	3	5	1	6	2	4	5	5	1	5	4	4	5	4	5	3	1	1	1	3	3	2	1	5	2	5	4	3	4	5	4	2	2	3	5	3	5	2	3	3	6	1	6	3	6	1						
[3905]	4	4	4	2	5	3	2	5	2	4	6	1	1	4	6	2	4	5	6	4	2	4	4	4	1	4	2	3	5	4	3	1	5	1	4	4	3	4	2	2	2	6	5	5	1	1	2	5	4	4	5	3	3	3	5	2	3	4	1	3	6						
[3969]	2	2	4	5	4	1	2	1	6	4	3	2	2	1	2	2	3	6	2	6	1	6	4	4	4	3	2	4	2	2	2	1	1	6	2	2	3	4	5	1	1	5	2	5	1	4	3	4	5	5	4	2	1	1	3	2	6	3	2	5	5	5					
[4033]	2	6	1	4	5	4	5	1	5	5	4	4	4	5	3	5	5	3	1	4	6	6	6	1	1	4	3	5	3	4	4	2	4	4	2	5	5	3	4	2	6	1	5	5	2	3	2	4	1	3	1	3	2	3	5	2	1	5	5	5	5	5					
[4097]	5	1	3	3	5	3	5	1	3	4	6	1	1	1	2	2	5	6	6	4	1	5	2	4	5	1	6	6	1	4	4	1	5	5	6	3	4	1	1	4	4	5	5	6	3	4	4	1	2	3	1	5	5	2	5	5	3	1	3	3	4						
[4161]	3	6	3	5	3	6	5	4	4	5	3	1	1	4	4	5	3	2	5	3	1	5	2	3	6	1	4	2	2	6	1	5	1	6	4	4	6	2	1	1	3	4	3	5	3	5	6	5	1	3	1	4	1	4	3	3	1	4	2	5	4	2					
[4225]	5	2	6	1	1	3	1	5	3	1	6	1	3	1	3	6	3	5	3	4	2	3	1	2	4	3	2	3	1	4	3	1	1	5	5	2	3	2	6	3	4	3	2	5	1	2	3	5	4	5	2	1	2	3	5	2	3	5	2	3	3	6	2	2			
[4289]	5	3	4	3	2	4	3	5	2	4	6	2	2	5	2	3																																																			

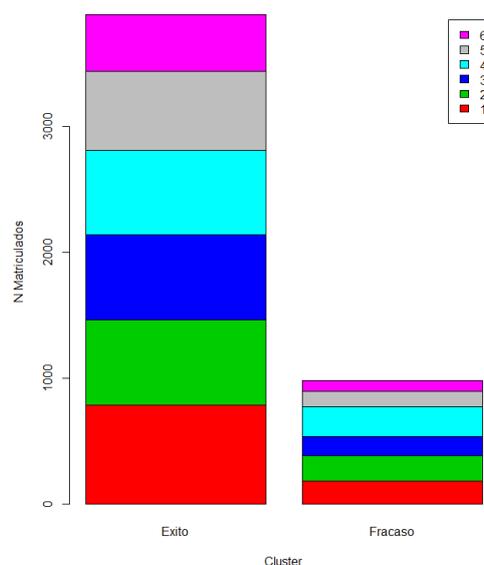
### 7.3.2. Preparación de un clasificador Éxito/Fracaso

Para estudiar en mayor profundidad los datos vamos a realizar un clasificador que nos servirá para evaluar la razón por la que un paquete de asignaturas matriculadas puede conducir a aprobar las asignaturas o a suspenderlas. Lo llamaremos clasificador Éxito/Fracaso y estudiaremos su relación con los clúster y con los paquetes de asignaturas que realizan los estudiantes.

Para la preparación de un clasificador de Éxito/Fracaso realizaremos por una lado para cada alumno la suma de sus asignaturas matriculadas, ya que la matriz de matriculaciones está formada por 0/1 con realizar la suma de filas de cada alumno nos dará el resultado de asignaturas matriculadas por cada alumno, por otro lado desde la matriz de asignaturas aprobadas realizaremos la misma técnica y obtendremos en número de asignaturas que ha aprobado cada alumno. A partir de ambas variables podemos determinar el éxito y el fracaso del alumno en cuestión. Determinaremos que es un éxito cuando el alumno aprueba alguna asignatura de las que se ha matriculado y tomaremos como un fracaso que el alumno no apruebe ninguna de las asignaturas de las que se ha matriculado. El detalle del código en R se encuentra adjunto en el Anexo 6.

Podemos saber también como información adicional que los clúster poseen esta distribución de frecuencias de Éxito o Fracaso. Los siguientes resultados nos indican las frecuencias absolutas:

Clúster	Éxito	Fracaso
1	787	183
2	676	205
3	678	149
4	673	236
5	628	122
6	448	86



**Ilustración 18. Relación Éxito/Fracaso en cada clúster**

Como resultado del estudio se ha obtenido una tabla que nos indica el paquete de asignaturas matriculado, el clúster en el que se encuentra y la cantidad de éxitos o fracasos se han dado en el paquete analizado.

En el Anexo 7 se pueden apreciar cómo los diferentes paquetes de asignaturas matriculadas por cada alumno caen en los diferentes clúster, así por ejemplo se puede ver que los alumnos que matriculan la asignatura 1ª junto con asignaturas más allá de la 10ª se suelen agrupar en el clúster 4, mientras que las asignaturas que contienen entre sus matriculadas la 1ª y la 2ª habitan a caer en el clúster 1. De esta tabla podemos sacar en conclusión cuáles son las agrupaciones de asignaturas que más se parecen entre sí y ver su índice de Éxito (cuando se aprueba alguna asignatura) o Fracaso (cuando no se aprueba ninguna). En general podemos ver que la relación de fracaso no es significativa ya que se produce por muy poca diferencia respecto al éxito, es decir la diferencia entre el éxito y el fracaso suele ser de uno o dos puntos, lo que no se puede decir que signifique algo. Tan sólo en el caso de la asignatura 9ª se puede apreciar que se suspenda de manera aislada.

### 7.3.3. Árboles de Decisión

Los Árboles de Decisión forman parte de los modelos más utilizados por tener una interpretación más natural y tener una ejecución fácil y buena cantidad de técnicas de clasificación que se pueden traspasar a reglas de clasificación lo cual nos hace sencilla la interpretación y aplicación.

Una vez preparado el clasificador nos disponemos a practicar diferentes técnicas de construcción de árboles de decisión sobre los datos de matriculaciones unidos junto con nuestro clasificador. Los detalles de la codificación en R se pueden apreciar en el Anexo 6.

#### **Algoritmo de árbol de decisión 'ctree' en R**

Una vez calculado el clasificador de Éxito/Fracaso, que codificaremos como 1-Éxito/0-Fracaso, lo juntamos con la matriz de asignaturas matriculadas y nos disponemos a probar mediante un método simple cómo las decisiones de determinadas asignaturas influyen en obtener éxito o fracaso. Así, nos disponemos a crear un árbol de decisión de tipo simple sobre los datos, poniendo como variable de destino la de Éxito/Fracaso. La ejecución del algoritmo ctree, nos devuelve los siguientes resultados:

```
Conditional inference tree with 5 terminal nodes

Response:  exito_del_alumno_FE01

Inputs:  01.500, 01.505, 01.507, 01.502, 01.506, 01.503, 01.504,
01.509, 01.501, 01.508, 01.520, 01.521, 01.516, 01.517, 01.519,
01.524, 01.526, 01.528, 01.513, 01.529, 01.518, 01.514, 01.512,
01.510, 01.515, 01.522, 01.511, 01.527, 01.523, 01.530, 01.535,
01.554, 01.549, 01.525, 01.537, 01.547, 01.542, 01.534, 01.545,
01.552, 01.561, 01.550, 01.544, 01.536, 01.543

Number of observations:  4871

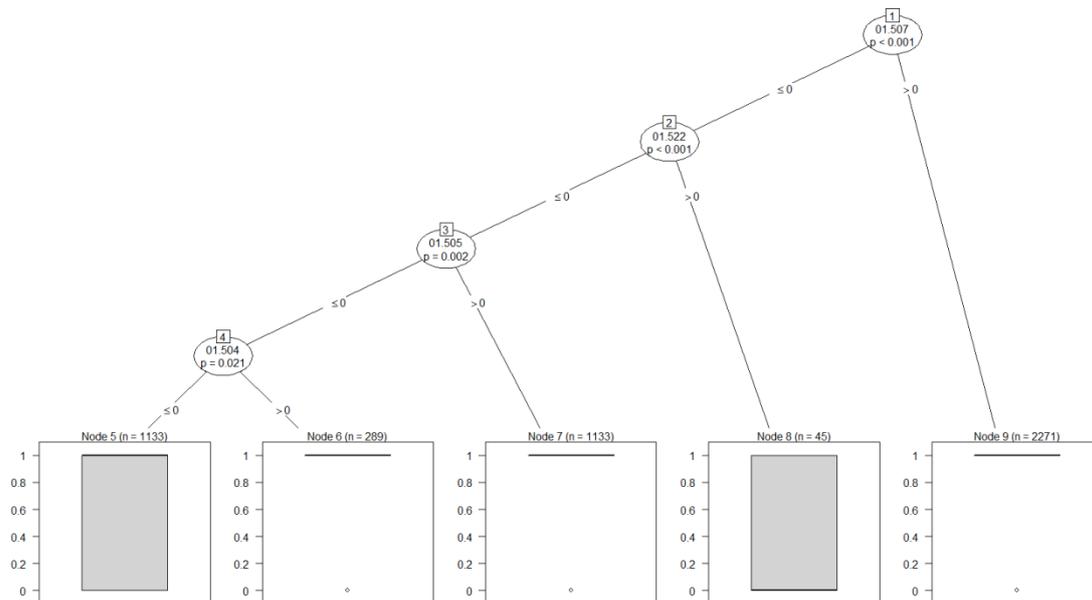
1) 01.507 <= 0; criterion = 1, statistic = 31.496
2) 01.522 <= 0; criterion = 1, statistic = 20.108
```

```

3) 01.505 <= 0; criterion = 0.998, statistic = 16.534
4) 01.504 <= 0; criterion = 0.979, statistic = 12.222
5)* weights = 1133
4) 01.504 > 0
6)* weights = 289
3) 01.505 > 0
7)* weights = 1133
2) 01.522 > 0
8)* weights = 45
1) 01.507 > 0
9)* weights = 2271

```

El diagrama del árbol resultante es el siguiente:



**Ilustración 19. Árbol Decisión ctree Asignaturas Matriculadas vs Éxito/Fracaso**

### Conclusiones

Este algoritmo nos ofrece una versión muy simple de la cadena de matriculaciones que pueden llevar a que un alumno suspenda todas las asignaturas, este árbol nos interpreta que si no están matriculadas de ningunas de las asignaturas 01.507, 01.522, 01.505 y 01.504 existe una tasa de éxito de  $n=1133$  y se si está matriculado de ellas de manera conjunta puedes obtener diversos resultados de éxito o fracaso. En 2271 matriculaciones, la asignatura 01.507 presenta Éxito, según su tasa de aprobado/suspenseo que aparece en la tabla de la sección 6.3 parece tener bastante relación con este resultado. Por otro lado, si miramos la asignatura 01.522 en la misma tabla vemos que en relación a este árbol también es una asignatura que induce al Fracaso según su tasa de aprobado/suspenseo. De esta misma manera podemos apreciar las asignaturas 01.504 y 01.505, las cuales su matriculación induce al éxito en casos elevados y con tasa aprobados entre el triple y cuádruple que de suspensos.

### Algoritmo de árbol de decisión rpart

A continuación usaremos un algoritmo más elaborado que nos devolverá información más completa sobre un árbol de decisión sobre nuestro clasificador de Éxito/Fracaso y nos dará las reglas correspondientes asociadas.

Para facilitar la visualización de los diagramas, para este tipo de algoritmo tomaremos los nombres de las asignaturas como ordinales del 1 al 45, y usaremos el algoritmo "rpart" el cuál nos proporciona diferentes niveles de complejidad en su ejecución. Probaremos con dos niveles diferentes de complejidad para realizar el árbol, un nivel de complejidad que nos dará un resultado de menos nodos y otro nivel de complejidad menor que nos dará un resultado de más nodos y compararemos las diferencias.

### Nivel de complejidad 0.001:

```
n= 4871
node), split, n, loss, yval, (yprob)
* denotes terminal node

1) root 4871 981 Exito (0.79860398 0.20139602)
2) 3>=0.5 2271 379 Exito (0.83311317 0.16688683) *
3) 3< 0.5 2600 602 Exito (0.76846154 0.23153846)
6) 2>=0.5 1139 216 Exito (0.81035996 0.18964004) *
7) 2< 0.5 1461 386 Exito (0.73579740 0.26420260)
14) 7>=0.5 291 51 Exito (0.82474227 0.17525773) *
15) 7< 0.5 1170 335 Exito (0.71367521 0.28632479)
30) 26< 0.5 1133 314 Exito (0.72285966 0.27714034)
60) 19>=0.5 29 2 Exito (0.93103448 0.06896552) *
61) 19< 0.5 1104 312 Exito (0.71739130 0.28260870)
122) 13< 0.5 1039 284 Exito (0.72666025 0.27333975) *
123) 13>=0.5 65 28 Exito (0.56923077 0.43076923)
246) 1>=0.5 22 6 Exito (0.72727273 0.27272727) *
247) 1< 0.5 43 21 Fracaso (0.48837209 0.51162791)
494) 8>=0.5 8 2 Exito (0.75000000 0.25000000) *
495) 8< 0.5 35 15 Fracaso (0.42857143 0.57142857) *
31) 26>=0.5 37 16 Fracaso (0.43243243 0.56756757)
62) 17>=0.5 10 4 Exito (0.60000000 0.40000000) *
63) 17< 0.5 27 10 Fracaso (0.37037037 0.62962963) *
```

```
Classification tree:
rpart(formula = exito_del_alumno_EF ~ ., data = resultados_asignaturas_fa_exito_1_a_45[,
-46], method = "class", parms = list(split = "information"),
control = rpart.control(cp = 0.001, usesurrogate = 0, maxsurrogate = 0))

Variables actually used in tree construction:
[1] 1 13 17 19 2 26 3 7 8

Root node error: 981/4871 = 0.2014

n= 4871

      CP nsplit rel error xerror   xstd
1 0.0012742     0  1.00000 1.0000 0.028532
2 0.0010000     9  0.98777 1.0082 0.028619
```

## Árbol de decisión Matriculaciones vs Éxito/Fracaso cp:0.001

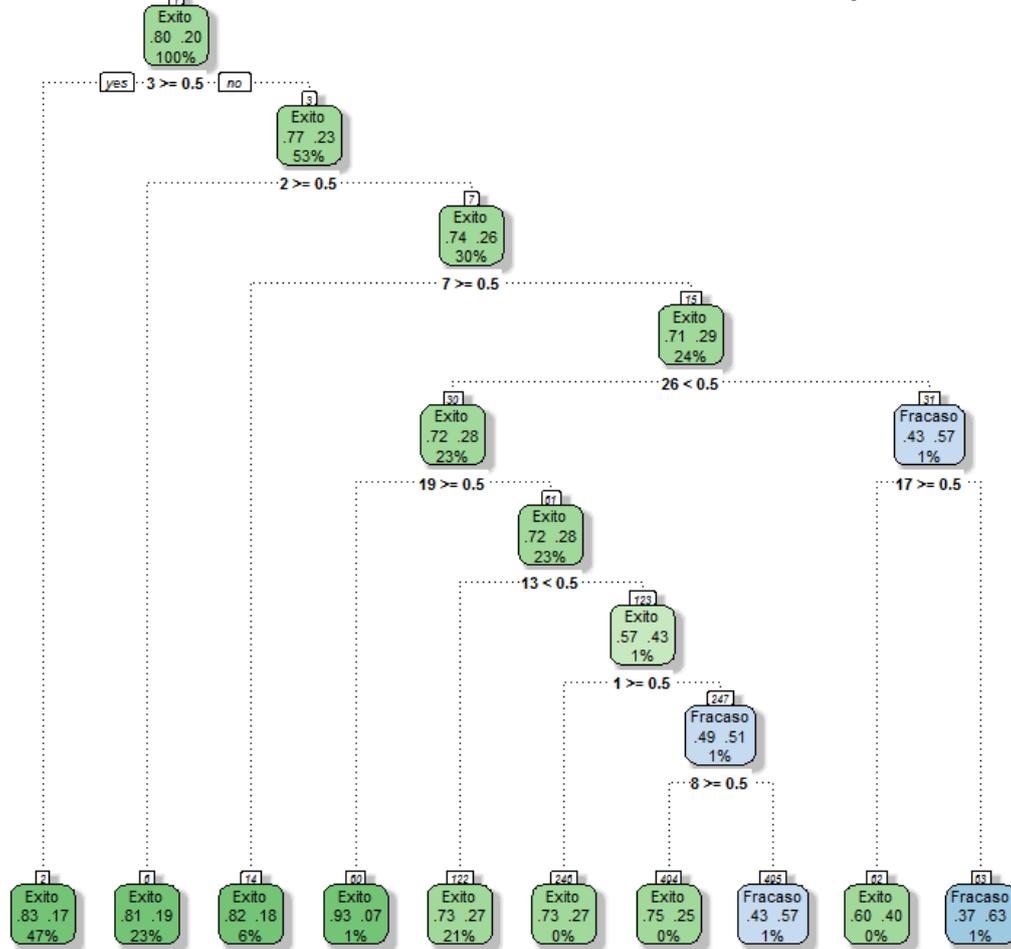


Ilustración 20. Árbol de Decisión método "rpart" en R, complejidad 0.001

## Reglas Obtenidas

A continuación mostramos las reglas obtenidas para este mismo árbol mostrado en la Ilustración 20:

```

Rule number: 63 [EF=Fracaso cover=27 (1%) prob=0.63]
3 < 0.5
2 < 0.5
7 < 0.5
26 >= 0.5
17 < 0.5

Rule number: 495 [EF=Fracaso cover=35 (1%) prob=0.57]
3 < 0.5
2 < 0.5
7 < 0.5
26 < 0.5
19 < 0.5
13 >= 0.5
1 < 0.5
8 < 0.5

Rule number: 62 [EF=Exito cover=10 (0%) prob=0.40]
3 < 0.5
2 < 0.5
7 < 0.5
26 >= 0.5
17 >= 0.5

Rule number: 122 [EF=Exito cover=1039 (21%) prob=0.27]
3 < 0.5
2 < 0.5
7 < 0.5

```

```

26< 0.5
19< 0.5
13< 0.5

Rule number: 246 [EF=Exito cover=22 (0%) prob=0.27]
3< 0.5
2< 0.5
7< 0.5
26< 0.5
19< 0.5
13>=0.5
1>=0.5

Rule number: 494 [EF=Exito cover=8 (0%) prob=0.25]
3< 0.5
2< 0.5
7< 0.5
26< 0.5
19< 0.5
13>=0.5
1< 0.5
8>=0.5

Rule number: 6 [EF=Exito cover=1139 (23%) prob=0.19]
3< 0.5
2>=0.5

Rule number: 14 [EF=Exito cover=291 (6%) prob=0.18]
3< 0.5
2< 0.5
7>=0.5

Rule number: 2 [EF=Exito cover=2271 (47%) prob=0.17]
3>=0.5

Rule number: 60 [EF=Exito cover=29 (1%) prob=0.07]
3< 0.5
2< 0.5
7< 0.5
26< 0.5
19>=0.5

```

### Conclusiones

Interpretamos que sobre la totalidad de los datos el porcentaje de Éxito de las matriculaciones es de un 80% aproximadamente frente al Fracaso que se eleva al 20%. En este nivel de complejidad aparecen una serie de dos caminos los cuales llevan a determinar el Fracaso que se produce si:

- No se matricula la asignatura 3ª (01.507), la 2ª (01.505), la 7ª (01.504), la 26ª (01.522), la 19ª (01.513), Si se matricula de la 13ª (01.516) pero no se matricula la 1ª (01.500) y la 8ª (01.509), esta regla cubre un total de 35 casos, lo que representa un 1% de la muestra y una probabilidad de 0,57. Por lo que se refleja de la tabla de la sección 6.3 la asignatura 01.516 tiene una tasa de aprobado/suspenseo de 0,64, lo cual predispone a Fracasar la asignatura de manera aislada.
- No se matricula la asignatura 3ª (01.507), la 2ª (01.505), la 7ª (01.504), sí se matricula la 26ª (01.522) pero no se matricula de la asignatura 17ª (01.514), tenemos que sobre 27 observaciones que representa un 1% de las observaciones, la asignatura 01.522 cuya tasa es el doble de suspensos que de aprobados también predispone a Fracasar.

Es decir, las asignaturas 13ª (01.516) y 26ª (01.522) suelen ser determinantes para que exista alguna probabilidad de fracaso pero que las que aparecen como las que decantan la balanza pueden ser la 8ª (01.509) y la 17ª (01.514) cuando se matriculan inducen a Fracaso en mayor probabilidad del 50% por muy poco, aunque ellas en sí mismas no tienen una tasa baja de aprobado/suspenseo (ver sección 6.3).

**Nivel de complejidad 0.0001:**

n= 4871

```
node), split, n, loss, yval, (yprob)
* denotes terminal node
```

```
1) root 4871 981 Exito (0.79860398 0.20139602)
  2) 3>=0.5 2271 379 Exito (0.83311317 0.16688683)
    4) 12>=0.5 72 3 Exito (0.95833333 0.04166667) *
    5) 12< 0.5 2199 376 Exito (0.82901319 0.17098681)
    10) 2>=0.5 1154 177 Exito (0.84662045 0.15337955) *
    11) 2< 0.5 1045 199 Exito (0.80956938 0.19043062)
    22) 14>=0.5 23 1 Exito (0.95652174 0.04347826) *
    23) 14< 0.5 1022 198 Exito (0.80626223 0.19373777)
    46) 7>=0.5 72 8 Exito (0.88888889 0.11111111) *
    47) 7< 0.5 950 190 Exito (0.80000000 0.20000000)
    94) 11< 0.5 883 172 Exito (0.80520951 0.19479049) *
    95) 11>=0.5 67 18 Exito (0.73134328 0.26865672)
    190) 6< 0.5 54 10 Exito (0.81481481 0.18518519) *
    191) 6>=0.5 13 5 Fracaso (0.38461538 0.61538462) *
  3) 3< 0.5 2600 602 Exito (0.76846154 0.23153846)
    6) 2>=0.5 1139 216 Exito (0.81035996 0.18964004) *
    7) 2< 0.5 1461 386 Exito (0.73579740 0.26420260)
    14) 7>=0.5 291 51 Exito (0.82474227 0.17525773)
    28) 11< 0.5 270 42 Exito (0.84444444 0.15555556) *
    29) 11>=0.5 21 9 Exito (0.57142857 0.42857143)
    58) 1< 0.5 12 4 Exito (0.66666667 0.33333333) *
    59) 1>=0.5 9 4 Fracaso (0.44444444 0.55555556) *
  15) 7< 0.5 1170 335 Exito (0.71367521 0.28632479)
    30) 26< 0.5 1133 314 Exito (0.72285966 0.27714034)
    60) 19>=0.5 29 2 Exito (0.93103448 0.06896552) *
    61) 19< 0.5 1104 312 Exito (0.71739130 0.28260870)
    122) 13< 0.5 1039 284 Exito (0.72666025 0.27333975) *
    123) 13>=0.5 65 28 Exito (0.56923077 0.43076923)
    246) 1>=0.5 22 6 Exito (0.72727273 0.27272727) *
    247) 1< 0.5 43 21 Fracaso (0.48837209 0.51162791)
    494) 8>=0.5 8 2 Exito (0.75000000 0.25000000) *
    495) 8< 0.5 35 15 Fracaso (0.42857143 0.57142857) *
  31) 26>=0.5 37 16 Fracaso (0.43243243 0.56756757)
    62) 17>=0.5 10 4 Exito (0.60000000 0.40000000) *
    63) 17< 0.5 27 10 Fracaso (0.37037037 0.62962963) *
```

Classification tree:

```
rpart(formula = exito_del_alumno_EF ~ ., data = resultados_asignaturas_fa_exito_1_a_45[,
-46], method = "class", parms = list(split = "information"),
control = rpart.control(cp = 1e-04, usesurrogate = 0, maxsurrogate = 0))
```

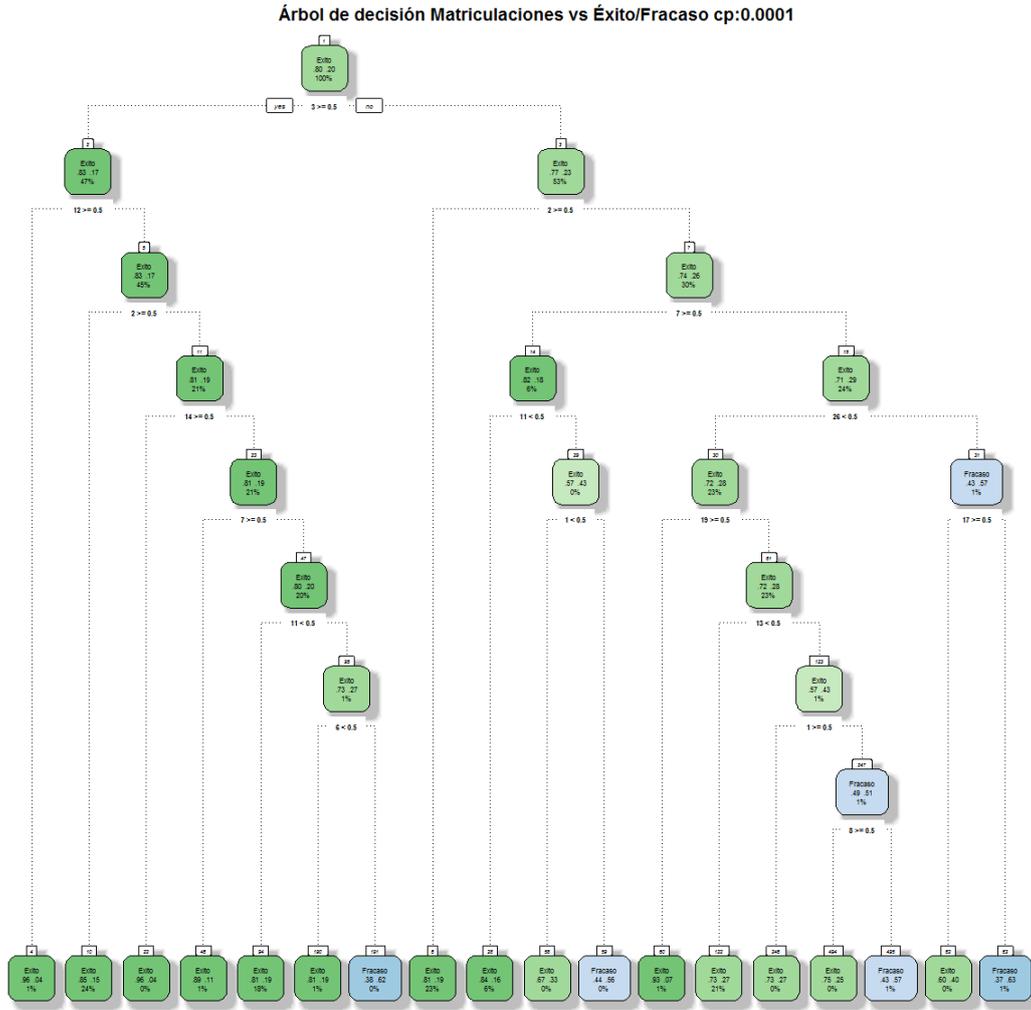
Variables actually used in tree construction:

```
[1] 1 11 12 13 14 17 19 2 26 3 6 7 8
```

Root node error: 981/4871 = 0.2014

n= 4871

	CP	nsplit	rel error	xerror	xstd
1	0.00127421	0	1.00000	1.0000	0.028532
2	0.00050968	9	0.98777	1.0071	0.028608
3	0.00010000	17	0.98369	1.0183	0.028726



**Ilustración 21. Árbol de Decisión método "rpart" en R complejidad 0.0001**

**Reglas Obtenidas**

```

Rule number: 63 [EF=Fracaso cover=27 (1%) prob=0.63]
3 < 0.5
2 < 0.5
7 < 0.5
26 >= 0.5
17 < 0.5

Rule number: 191 [EF=Fracaso cover=13 (0%) prob=0.62]
3 >= 0.5
12 < 0.5
2 < 0.5
14 < 0.5
7 < 0.5
11 >= 0.5
6 >= 0.5

Rule number: 495 [EF=Fracaso cover=35 (1%) prob=0.57]
3 < 0.5
2 < 0.5
7 < 0.5
26 < 0.5
19 < 0.5
13 >= 0.5
1 < 0.5
8 < 0.5

Rule number: 59 [EF=Fracaso cover=9 (0%) prob=0.56]
3 < 0.5
2 < 0.5
    
```

7>=0.5  
11>=0.5  
1>=0.5

Rule number: 62 [EF=Exito cover=10 (0%) prob=0.40]

3< 0.5  
2< 0.5  
7< 0.5  
26>=0.5  
17>=0.5

Rule number: 58 [EF=Exito cover=12 (0%) prob=0.33]

3< 0.5  
2< 0.5  
7>=0.5  
11>=0.5  
1< 0.5

Rule number: 122 [EF=Exito cover=1039 (21%) prob=0.27]

3< 0.5  
2< 0.5  
7< 0.5  
26< 0.5  
19< 0.5  
13< 0.5

Rule number: 246 [EF=Exito cover=22 (0%) prob=0.27]

3< 0.5  
2< 0.5  
7< 0.5  
26< 0.5  
19< 0.5  
13>=0.5  
1>=0.5

Rule number: 494 [EF=Exito cover=8 (0%) prob=0.25]

3< 0.5  
2< 0.5  
7< 0.5  
26< 0.5  
19< 0.5  
13>=0.5  
1< 0.5  
8>=0.5

Rule number: 94 [EF=Exito cover=883 (18%) prob=0.19]

3>=0.5  
12< 0.5  
2< 0.5  
14< 0.5  
7< 0.5  
11< 0.5

Rule number: 6 [EF=Exito cover=1139 (23%) prob=0.19]

3< 0.5  
2>=0.5

Rule number: 190 [EF=Exito cover=54 (1%) prob=0.19]

3>=0.5  
12< 0.5  
2< 0.5  
14< 0.5  
7< 0.5  
11>=0.5  
6< 0.5

Rule number: 28 [EF=Exito cover=270 (6%) prob=0.16]

3< 0.5  
2< 0.5  
7>=0.5  
11< 0.5

Rule number: 10 [EF=Exito cover=1154 (24%) prob=0.15]

3>=0.5  
12< 0.5  
2>=0.5

Rule number: 46 [EF=Exito cover=72 (1%) prob=0.11]

3>=0.5  
12< 0.5  
2< 0.5  
14< 0.5  
7>=0.5

```

Rule number: 60 [EF=Exito cover=29 (1%) prob=0.07]
3< 0.5
2< 0.5
7< 0.5
26< 0.5
19>=0.5

Rule number: 22 [EF=Exito cover=23 (0%) prob=0.04]
3>=0.5
12< 0.5
2< 0.5
14>=0.5

Rule number: 4 [EF=Exito cover=72 (1%) prob=0.04]
3>=0.5
12>=0.5

```

### Conclusiones

Vemos que obtenemos en este caso también las asignaturas 8<sup>a</sup> (01.509) y 17<sup>a</sup> (01.526) como en el caso anterior, pero en este caso con menor nivel de complejidad. Vemos que nos aparece la asignatura 6<sup>a</sup> (01.503) que nos marca el también una frontera entre el éxito y el fracaso de las asignaturas seleccionadas, incluso se añade la asignatura 1<sup>a</sup> (01.500) para hacer distinción.

Con respecto a las reglas anteriores se nos añaden dos nuevas reglas que determinan que matriculándose de los grupos de asignaturas 1<sup>a</sup>, 7<sup>a</sup>, 11<sup>a</sup> y 3<sup>a</sup>, 6<sup>a</sup>, 11<sup>a</sup> nos conducen a fracaso, como vemos en la tabla de la sección 6.3 las asignaturas 1<sup>a</sup>, 3<sup>a</sup>, 6<sup>a</sup> y 7<sup>a</sup> tienen tasas de aprobado/suspenso altas (mayor que 2) pero sin embargo la asignatura 11<sup>a</sup> tiene una tasa aprobado/suspenso cercana a 1, lo que puede parecer que esta asignatura es una de aquellas que oscila entre el aprobado y el suspenso con mayor facilidad que otras.

Como conclusión se puede decir que estas asignaturas tienen relación en el éxito/fracaso del estudiante. Sin embargo aunque se pueda apreciar que combinaciones de dichas asignaturas contengan índices de éxito y fracaso en la tabla del anexo 6, no se puede establecer con suficiente soporte, ya que las observaciones se basan sobre un número de casos pequeño, es decir, combinaciones que contienen estas asignaturas suelen tener muy pocos matriculados, no van más allá de uno o dos alumnos.

### Árbol de decisión J48 con Weka

Ahora como medida complementaria al estudio, tomaremos la herramienta Weka y realizaremos un árbol de tipo J48 para comprobar si coinciden los resultados de los árboles anteriores con el resultado del algoritmo hecho en Weka.

A continuación se muestra el resultado de ejecutar el algoritmo:

```

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.27 -M 2
Relation: resultados_asignaturas_fa_exito-weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.NumericToNominal-Rfirst-last-
weka.filters.unsupervised.attribute.Remove-R47
Instances: 4871
Attributes: 46
          1.500
          1.505
          1.507
          1.502
          1.506
          1.503
          1.504

```

```

1.509
1.501
1.508
1.520
1.521
1.516
1.517
1.519
1.524
1.526
1.528
1.513
1.529
1.518
1.514
1.512
1.510
1.515
1.522
1.511
1.527
1.523
1.530
1.535
1.554
1.549
1.525
1.537
1.547
1.542
1.534
1.545
1.552
1.561
1.550
1.544
1.536
1.543
EF
Test mode:10-fold cross-validation

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

J48 pruned tree
-----

1.537 = 0
| 1.522 = 0
| | 1.520 = 0: Exito (4551.0/874.0)
| | 1.520 = 1
| | | 1.516 = 0: Exito (240.0/67.0)
| | | 1.516 = 1: Fracaso (18.0/5.0)
| | 1.522 = 1
| | | 1.523 = 0
| | | | 1.507 = 0
| | | | | 1.517 = 0
| | | | | 1.504 = 0
| | | | | 1.503 = 0
| | | | | 1.500 = 0
| | | | | | 1.512 = 0
| | | | | | | 1.510 = 0
| | | | | | | | 1.505 = 0
| | | | | | | | | 1.521 = 0
| | | | | | | | | 1.514 = 0
| | | | | | | | | 1.526 = 0: Fracaso (12.0/3.0)
| | | | | | | | | 1.526 = 1: Exito (6.0/2.0)
| | | | | | | | | 1.514 = 1: Exito (4.0/1.0)
| | | | | | | | | 1.521 = 1: Fracaso (2.0)
| | | | | | | | | 1.505 = 1: Exito (4.0/1.0)
| | | | | | | | | 1.510 = 1: Exito (2.0)
| | | | | | | | | 1.512 = 1: Exito (2.0)
| | | | | | | | | 1.500 = 1: Exito (2.0)
| | | | | | | | | 1.503 = 1: Fracaso (2.0)
| | | | | | | | | 1.504 = 1: Exito (3.0)
| | | | | | | | | 1.517 = 1: Fracaso (2.0)
| | | | | | | | | 1.507 = 1: Exito (7.0)
| | | | | | | | | 1.523 = 1: Fracaso (2.0)
1.537 = 1
| 1.505 = 0
| | 1.502 = 0: Fracaso (5.0)
| | 1.502 = 1: Exito (3.0/1.0)
| | 1.505 = 1: Exito (4.0)

Number of Leaves : 19
Size of the tree : 37

```

```

Time taken to build model: 0.35 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      3881      79.6756 %
Incorrectly Classified Instances    990      20.3244 %
Kappa statistic                    0.0177
Mean absolute error                 0.3193
Root mean squared error             0.4024
Relative absolute error             99.2463 %
Root relative squared error        100.3422 %
Total Number of Instances         4871

=== Detailed Accuracy By Class ===
                TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
                0.993   0.982    0.8        0.993   0.886     0.516   Exito
                0.018   0.007    0.4        0.018   0.035     0.516   Fracaso
Weighted Avg.   0.797   0.785    0.72       0.797   0.715     0.516

=== Confusion Matrix ===
  a    b  <-- classified as
3863  27 |  a = Exito
 963  18 |  b = Fracaso
    
```

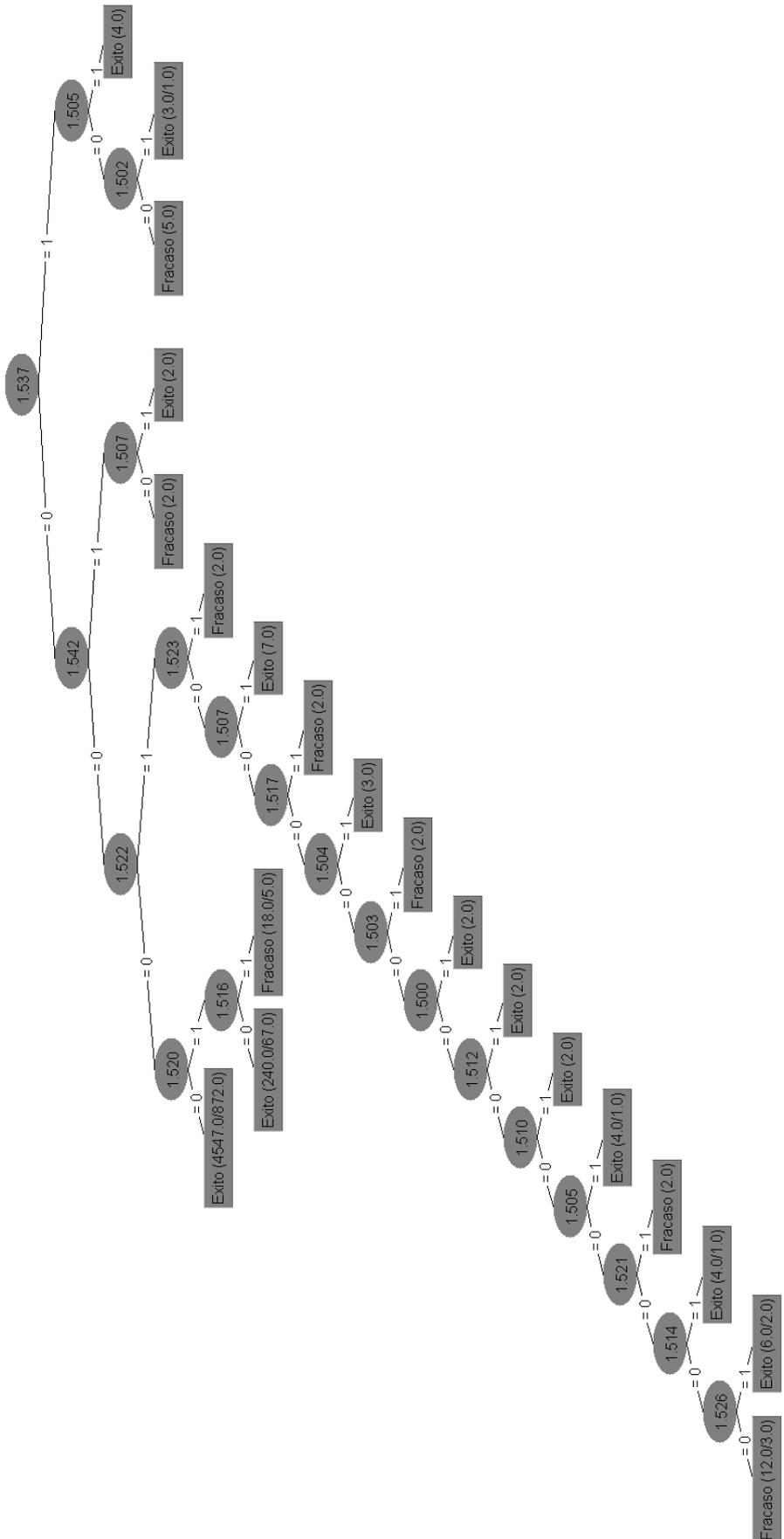


Ilustración 22. Árbol de Decisión con el algoritmo J48

### **Conclusiones**

Podemos ver que los conceptos de Fracaso aparecen cuando:

- Si no se matricula la asignatura 01.537, 01.522, sí se matricula la asignatura 01.520 y la 01.516
- La sucesión de asignaturas de la siguiente manera conduce a Fracaso: 01.537 = 0, 01.522 = 1, 01.523 = 0, 01.507 = 0, 01.517 = 0, 01.504 = 0, 01.503 = 0, 01.500 = 0, 01.512 = 0, 01.510 = 0, 01.505 = 0, 01.521 = 0, 01.514 = 0, 01.526 = 0, es decir cuando se matricula la asignatura 01.522 y no se matriculan otras existe mayor probabilidad de fracaso.
- En las siguientes ramas del árbol podemos ver que las asignaturas 01.503, 01.517 y 01.523 precedidas de sus correspondientes padres son asignaturas que conducen a Fracaso
- Por último la asignatura 01.537 conduce a Fracaso cuando se matricula sola y no se matriculan la 01.505 y la 01.502.

En este caso hemos dado con dos asignaturas que parecen tener alguna relación con el éxito o fracaso de un alumno, por un lado la asignatura 01.522 (Contabilidad Financiera) la cual ya muestra mayor número de suspensos que de aprobados y la asignatura 01.537 (Derecho Financiero y Tributario) en la que también existen mayor número de suspensos que de aprobados.

## 8. Conclusiones

En este TFG se ha intentado determinar las asignaturas que puedan resultar perjudiciales al resultado del curso. Primero se ha intentado establecer un modelo de similitud entre las asignaturas que pertenecen al plan de Administración y Dirección de Empresas para poder ver qué asignaturas son las que más coinciden en su estructura tanto como en el éxito o fracaso de las mismas. Y después se ha tratado de ver las similitudes entre los alumnos que las matriculan.

El reto de este proyecto consistía en tener un gran número de variables, poder tratarlas todas en conjunto y además que fueran de tipo binario (su representación fuera 1 en caso afirmativo y 0 en caso negativo). Por esa misma razón se han tratado técnicas que fueran favorables a este tipo de datos: se ha usado la técnica de Factorización de elementos no negativos para comprobar el agrupamiento de las asignaturas, con lo que hemos obtenido grupos de asignaturas que se comportan de manera similar, después hemos realizado un análisis de componentes principales (muy útil cuando el tamaño de las variables a procesar es grande) pero sin embargo hemos obtenido unos resultados difusos, los cuales parecían tener sentido pero no han resultado ser aplicables. Por último hemos tratado con técnicas de clustering y árboles de decisión las cuales han mostrado que en cierto modo la agrupación de matrículas siguen algunos patrones sencillos y con los árboles de decisión se ha podido ver que existen ciertas asignaturas que determinan el éxito o el fracaso de manera muy sutil. No obstante, entre los resultados obtenidos hemos podido encontrar una característica un tanto incoherente, como el hecho de que el no matricular una asignatura sea determinante para el éxito o fracaso de otras, ésta forma de mirar los datos ha revelado que los métodos utilizados para el estudio puedan ser más o menos adecuados. El complemento a todo resultado que en este proyecto se expone han sido una serie de tablas que muestran las frecuencias de matriculación y resultados sobre el total de los datos, que tiene el propósito de servir de guía para interpretar un determinado resultado en tanto al conjunto de asignaturas matriculadas de uno o varios alumnos. Éste hecho se puede apreciar en el Anexo 7 del epígrafe 9.1.

La utilidad que pueda tener este proyecto es la de detectar qué asignatura es la que de alguna manera suponga un riesgo en el grupo de asignaturas matriculadas por un alumno de forma que represente un lastre para tener éxito en el cómputo global de la matrícula. Así hemos podido ver que existen asignaturas que tanto por sí solas como acompañadas tienen mayor número de suspensos que de aprobados y por lo tanto significan riesgo al momento de matricularlas.

También se proponen como futuras ampliaciones profundizar más sobre los resultados obtenidos en el modelo NMF para comprobar si de alguna manera existen estas llamadas asignaturas "riesgo" que hacen que se decante la probabilidad de éxito o fracaso sobre un determinado conjunto de otras asignaturas. Así como también profundizar en el hecho de que es bastante útil la cuantificación de los paquetes de asignaturas que se matriculan y si de alguna manera su frecuencia de fracaso corresponde a algún patrón, lo cual se puede hacer viendo de manera continua, semestre tras semestre, qué paquetes de asignaturas tienen mayor tasa de fracaso. Se Anexan resultados ilustrativos de la idea en los Anexos 8, 9 y 10 de la sección 0.

Lo que he aprendido con este proyecto es que la minería de datos es un proceso iterativo para el cual se ha de disponer de las herramientas y conocimientos adecuados para poder aplicarlos con soltura. Y que además la aplicación de diferentes técnicas puede llevar a extraer resultados que no sean concluyentes, e incluso que diferentes técnicas muestren diferentes focos de atención, lo cual nos daría resultados dispares o contradictorios. También he podido comprobar que los datos estadísticos básicos revelan información muy útil a primera vista la cual se ha de tener en cuenta a la hora de fabricar modelos de datos de minería, como por ejemplo cuando una asignatura tiene más suspensos que aprobados, puede ser una buena candidata para revelar el comportamiento, es decir, si en la estadística base aparece un comportamiento que estamos buscando, posiblemente este comportamiento también aparezca en los diferentes modelos que tratamos. Por último, también puedo comentar que he aprendido que los diferentes modelos que se estudien pueden no revelar ninguna situación concreta, esto no quiere decir que el proceso sea incorrecto o esté mal, sino que simplemente no se puede encontrar lo que buscamos aunque los datos parezcan ofrecerlo.

Para trabajos posteriores se deja pendiente la realización de encontrar un modelo significativo de manera que se tomen los grupos de asignaturas que se matriculan y se comparen con los grupos de asignaturas que se aprueban, por el momento se ha hecho una tabla cruzada con ambas variables y se ha medido la frecuencia de aparición. Esto es sí mismo parece tener forma de árbol pero se deja pendiente el hacer un modelo que explique estos mismos datos, que por el momento se encuentran en bruto. En el apartado "Futuras ampliaciones" se comenta cómo se ampliaría el proyecto.

Por último también queda decir que el aprendizaje de un lenguaje de programación hasta el momento desconocido para mí ha sido una experiencia muy ilustrativa de lo que puede representar la potencia de ejecutar comandos y no encontrarse en un entorno cerrado el cuál presente muchas opciones pero en algunos casos resulte limitado. El lenguaje R ha representado la estructura general del proyecto, aunque para confirmar o corroborar resultados se haya echado mano del entorno gráfico que ofrece "Rattle" y herramientas de acceso visual como Weka.

En definitiva, el proyecto ha sido una aplicación sucesiva de técnicas de minería de datos para poder establecer relaciones entre los datos proporcionados por el departamento. Dichas técnicas a veces han arrojado luz sobre los datos y otras veces simplemente no han revelado las situaciones deseadas.

## 9. Anexos

En estos anexos se muestran los scripts en R utilizados, junto a la memoria se adjunta el script completo de trabajo del curso entero y el proyecto realizado con el programa RStudio.

### 9.1. Anexo 1. Librerías utilizadas para el análisis de los datos y las funciones gráficas

Para la realización del proyecto han sido necesarias una serie de librerías, aquí se muestra el código R necesario para incluirlas antes de ejecutar el proyecto:

```
#librerías necesarias para ejecutar las funciones de minería sobre modelos NMF, PCA y K-means
#librerías necesarias
library("reshape")
library("party")
library("plyr")
library("rattle")
library("psych")
library("ggplot2")
library("rpart")
library("lattice")
library("rpart")
library("arulesViz")
library("C50")

#ANALISIS DE COMPONENTES PRINCIPALES SOLO CON LOS DATOS DE ASIGNATURAS_FA
library("psych")
library("GPARotation")
library("igraph")

#comprobamos las relaciones que se establecen entre las diferentes asignaturas
#vamos a dibujar un gráfico de relaciones de correlación
library("corrplot")
require("corrplot", quietly=TRUE)
```

### 9.2. Anexo 2. Procesado de datos y estadísticas básicas

La limpieza y procesado de datos previa para obtener una visión general de los mismos se realiza mediante el siguiente código en R:

```
#leemos el fichero pasado con los nombres de las asignaturas
asignaturas <- read.table(file="GR01.ass.1sem", head=FALSE, as.is = TRUE, colClasses = "character")

#leemos el fichero con los datos de las matriculaciones y resultados
resultados <- read.table(file="TFG_EDMLA.dat", head=FALSE, sep=";")

#establecemos una matriz con las cabeceras que tendrán
#el formato fa_ y sup_ (para las asignaturas que se matriculan y que se superan)
nombres_cabeceras <- cbind(paste("fa_", asignaturas[,1], sep=""), paste("sup_", asignaturas[,1], sep=""))

#calculamos las cabeceras de todas las columnas, la primera será "Id_Alumno"
cabeceras <- c("Id_Alumno")
i<-1
while(i<=nrow(nombres_cabeceras)){
  cabeceras <- c(cabeceras, nombres_cabeceras[i,])
  i<-i+1
}

#Los datos tendrán las cabeceras calculadas
colnames(resultados) <- c(cabeceras)

write.csv(resultados, file="Alumnos_Matriculados_Resultados.csv")

#####
#ASINATURAS MATRICULADAS (_FA)
#de modo que 0=No Matricula, 1=Si Matricula
#secuencia de columnas de la 2 a la 90, asignaturas matriculadas (_fa)
columnas_fa <- seq(2,90,2)
resultados_asignaturas_fa <- resultados[,columnas_fa]
colnames(resultados_asignaturas_fa) <- asignaturas[,1]
```

```

summary(resultados_asignaturas_fa)
#####

#####
#ASINATURAS APROBADAS (_SUP)
#creamos una función que nos extrae las asignaturas fa en un único dataframe
#de modo que 0=No Matricula, 1=Si Matricula
#secuencia de columnas de la 2 a la 90, asignaturas matriculadas (_fa)
columnas_sup <- seq(3,91,2)
resultados_asignaturas_sup <- resultados[columnas_sup]
colnames(resultados_asignaturas_sup) <- asignaturas[1]

summary(resultados_asignaturas_sup)
#####

#####
#RESULTADOS DE TODAS LAS ASINATURAS APROBADAS
#combinamos las asignaturas fa+supera en un único valor
#de modo que 0=No_Matricula, 1=Suspende, 2=Aprueba
resultados_num <- resultados_asignaturas_fa + resultados_asignaturas_sup

#Estadística básica de las diferentes columnas
summary(resultados_num)

#necesitamos también tener los resultados combinados en forma categórica aunque numérica
resultados_num_factor <- data.frame(resultados_num)
resultados_num_factor[,names(resultados_num_factor)] <- lapply(resultados_num[, names(resultados_num)], factor)
#####

#####
#construcción de un diagrama de barras explicativo
#estadística básica de los factores
summary(resultados_num_factor)

#función para calcular las frecuencias de los resultados 0-No matriculados, 1-Suspensos, 2-Aprobados
#devuelve un dataset: Resultados, as1, ..., as45 con las frecuencias absolutas para cada tipo
frecuencias_resultados <- function(x,asignaturas){
  i <- 1
  y <- data.frame(count(x[,1])$x)
  while(i<=ncol(x)){
    y <- cbind(y, count(x[,i])$freq)
    i <- i+1
  }
  #calculamos las cabeceras de todas las columnas, la primera será "Id_Alumno"
  cabeceras <- c("Resultado")
  i<-1
  while(i<=nrow(asignaturas)){
    cabeceras <- c(cabeceras, asignaturas[i,])
    i<-i+1
  }
  colnames(y) <- c(cabeceras)
  return(y)
}

#obtenemos las frecuencias de cada factor
resultados_num_factor_frecuencias <- frecuencias_resultados(resultados_num_factor, asignaturas)
resultados_num_factor_frecuencias

Categorias <- c("No Matriculado", "Suspense", "Aprobado")
resultados_num_factor_frecuencias <- cbind(Categorias, resultados_num_factor_frecuencias[,-1])

aprobados <- resultados_num_factor_frecuencias[3,-1]
aprobados
suspensos <- resultados_num_factor_frecuencias[2,-1]
suspensos
matriculados <- aprobados + suspensos
matriculados

barplot(as.matrix(aprobados), col="green", xlab="Aprobados", ylab="N_Alumnos", cex.names=0.5)
barplot(as.matrix(suspensos), col="red", xlab="Suspensos", ylab="N_Alumnos", cex.names=0.5)
barplot(as.matrix(matriculados), col="blue", xlab="Matriculados", ylab="N_Alumnos", cex.names=0.5)

barplot(as.matrix(aprobados/matriculados), xlab="Aprob./Matric.", ylab="Tasa", col="orange", cex.names=0.5, main="Aprobados vs.
Matriculados")
barplot(as.matrix(aprobados/suspensos), xlab="Aprob./Susp.", ylab="Tasa", col="blue", cex.names=0.5, main="Aprobados vs. Suspensos")

write.csv(resultados_num_factor_frecuencias, file="Alumnos_Frecuencias.xls")

#barplot de colores donde se visualizan los datos en tres colores,
barplot(as.matrix(resultados_num_factor_frecuencias[,-1]), main="Matriculaciones vs. Resultados",
  col=c("blue","red","green"), legend=list("No matriculados", "Suspensos", "Aprobados"),
  ylab="Asignaturas", xlab="N Alumnos", cex.names=0.5, horiz=TRUE)

barplot(as.matrix(resultados_num_factor_frecuencias[c(1,3,2),-1]), main="Matriculaciones vs. Resultados",
  col=c("blue","green", "red"), beside=FALSE, legend=list("No matriculados", "Aprobados", "Suspensos"),
  ylab="Asignaturas", xlab="N Alumnos", cex.names=0.5, horiz=TRUE)

```

```

barplot(as.matrix(resultados_num_factor_frecuencias[c(3,2),-1]), main="Matriculaciones vs. Resultados",
        col=c("green", "red"), legend=list("Aprobados", "Suspensos"),
        ylab="Asignaturas", xlab="N Alumnos", cex.names=0.5, horiz=TRUE)

barplot(as.matrix(resultados_num_factor_frecuencias[c(3,2),-1]), main="Matriculaciones vs. Resultados",
        col=c("green", "red"), beside=TRUE, legend=list("Aprobados", "Suspensos"),
        ylab="Asignaturas", xlab="N Alumnos", cex.names=0.5)
#####

```

### 9.3. Anexo 3. Script para Análisis de Matrices de Factores No Negativos (NMF)

```

#####
#Estudio de componentes principales mediante un NMF
library("NMF")

#matriz transpuesta Asignaturas x Observaciones
resultados_asignaturas_fa_t <- t(resultados_asignaturas_fa)

#Estimación del parámetro rank
resultados_fa_nmf.r <- nmf(resultados_asignaturas_fa, 2:10, nrun=30, .opt="v")

plot(resultados_asignaturas_fa_nmf.r)

consensusmap(resultados_asignaturas_fa_nmf.r, annCol=NA,
             #resultados_asignaturas_fa_t,
             labCol=NA, labRow=NA)

#uso del algoritmo NMF básico: X ~ W-H
#estimar_NMFrank <- nmfEstimateRank(resultados_asignaturas_fa, 2:3, .opt="v")
#el resultado de estimar el parametro rank nos hace pensar que el criterio más útil
# sería tomar el valor 5 para obtener una matriz que explicara los datos
resultados_fa_nmf <- nmf(resultados_asignaturas_fa_t, 5, .opt="v")

summary(resultados_asignaturas_fa_nmf)

#el método fit devuelve el ajuste del algoritmo en un objeto NMF
resultados_asignaturas_fa_nmf_fit <- fit(resultados_asignaturas_fa_nmf)
#fitted devuelve la matriz en sí misma
resultados_asignaturas_fa_nmf_fitted <- fitted(resultados_asignaturas_fa_nmf)

#matriz resultante W de las bases
resultados_asignaturas_fa_nmf_basis <- basis(resultados_asignaturas_fa_nmf)
round(resultados_asignaturas_fa_nmf_basis, 3)
matplot(resultados_asignaturas_fa_nmf_basis, type="b")
basismap(resultados_asignaturas_fa_nmf)

#matriz resultante H de los coeficientes
resultados_asignaturas_fa_nmf_coef <- coef(resultados_asignaturas_fa_nmf)
round(resultados_asignaturas_fa_nmf_coef, 3)
#matplot(resultados_asignaturas_fa_nmf_coef, type="b")
#coefmap(resultados_asignaturas_fa_nmf)

#Mapa de colores de los elementos base
aheatmap(resultados_asignaturas_fa_nmf_basis, cexRow=1, cexCol=2)

#Mapa de colores de los elementos coeficientes
#aheatmap(resultados_asignaturas_fa_nmf_coef, cexRow=1, cexCol=2)

consensusmap(resultados_asignaturas_fa_nmf)

```

### 9.4. Anexo 4. Script para Análisis de Componentes Principales (PCA)

```

#####
#ANALISIS DE COMPONENTES PRINCIPALES SOLO CON LOS DATOS DE ASIGNATURAS_FA
library("psych")
library("GPArotation")
library("igraph")

#comprobamos las relaciones que se establecen entre las diferentes asignaturas
#vamos a dibujar un gráfico de relaciones de correlación
library("corrplot")
require("corrplot", quietly=TRUE)

resultados_asignaturas_fa_cor <- cor(resultados_asignaturas_fa, use="pairwise", method="pearson")

```

```

corrplot(resultados_asignaturas_fa_cor, mar=c(0,0,1,0), order="hclust", tl.cex=0.6)

#definimos los colores para los que queremos que se gradue el gráfico
col4 <- colorRampPalette(c("#F0000", "red", "#FF7F00", "orange", "#FFF7F", "cyan", "#007FFF", "dark blue", "#00007F"))
corrplot(resultados_asignaturas_fa_cor,
  method = "square",
  title="Correlaciones entre asignaturas",
  order="hclust", addrect = 5,
  col=col4(10), tl.col="brown", tl.cex=0.6)

#necesitamos primero la matriz de correlación de los resultados
resultados_asignaturas_fa_cor <- cor(resultados_asignaturas_fa, use="pairwise",
  method="pearson")
resultados_asignaturas_fa_cor

#Estudio de los datos para saber el número de componentes a obtener
#sugiere que el número de componentes apropiado es 15
fa.parallel(resultados_asignaturas_fa_cor, n.obs=nrow(resultados_asignaturas_fa), fa="both",
  main="Análisis paralelo de variables")

#exploración de componentes principales sin rotación
resultados_asignaturas_fa_cor_principal <- principal(resultados_asignaturas_fa_cor, nfactors=6, rotate="oblmin")
resultados_asignaturas_fa_cor_principal

#diagramas de los componentes principales
factor.plot(resultados_asignaturas_fa_cor_principal, labels=rownames(resultados_asignaturas_fa_cor_principal))
fa.diagram(resultados_asignaturas_fa_cor_principal, simple=FALSE)

#exploración de componentes principales con rotación de la varianza máxima
resultados_asignaturas_fa_cor_principal_varimax <- principal(resultados_asignaturas_fa_cor, nfactors=5, rotate="varimax")
resultados_asignaturas_fa_cor_principal_varimax

#diagramas de los componentes principales con rotación de la varianza máxima
factor.plot(resultados_asignaturas_fa_cor_principal_varimax, labels=rownames(resultados_asignaturas_fa_cor_principal_varimax))
fa.diagram(resultados_asignaturas_fa_cor_principal_varimax, simple=FALSE)

#exploración de componentes principales con rotación del promedio máximo
resultados_asignaturas_fa_cor_principal_promax <- principal(resultados_asignaturas_fa_cor, nfactors=5, rotate="promax")
resultados_asignaturas_fa_cor_principal_promax

#diagramas de los componentes principales con rotación de la varianza máxima
factor.plot(resultados_asignaturas_fa_cor_principal_promax, labels=rownames(resultados_asignaturas_fa_cor_principal_promax))
fa.diagram(resultados_asignaturas_fa_cor_principal_promax, simple=FALSE)

#Hacemos un estudio exploratorio de análisis de Factores
#primero observamos las covarianzas de las variables
resultados_asignaturas_fa_cov <- cov(resultados_asignaturas_fa)
resultados_asignaturas_fa_cov

#después vemos la correlaciones existentes
resultados_asignaturas_fa_cov_cor <- cov2cor(resultados_asignaturas_fa_cov)
resultados_asignaturas_fa_cov_cor

#Decidimos el número de factores a extraer según un estudio de los datos en paralelo
fa.parallel(resultados_asignaturas_fa_cov_cor, n.obs=nrow(resultados_asignaturas_fa), fa="both", main="Análisis de factores paralelos")
#el estudio sugiere 17 factores y 16 componentes

#estudio de 17 factores sin rotación
resultados_asignaturas_fa_cov_cor_fa <- fa(resultados_asignaturas_fa_cov_cor, nfactors=5, rotate="none", fm="pa")
resultados_asignaturas_fa_cov_cor_fa

#diagramas de los factores
factor.plot(resultados_asignaturas_fa_cov_cor_fa, labels=rownames(resultados_asignaturas_fa_cov_cor_fa))
fa.diagram(resultados_asignaturas_fa_cov_cor_fa, simple=FALSE)

#estudios de 17 factores con rotación de varianza máxima
resultados_asignaturas_fa_cov_cor_fa_varimax <- fa(resultados_asignaturas_fa_cov_cor, nfactors=5, rotate="varimax", fm="pa")
resultados_asignaturas_fa_cov_cor_fa_varimax

#diagramas de los factores con rotación de varianza máxima
factor.plot(resultados_asignaturas_fa_cov_cor_fa_varimax, labels=rownames(resultados_asignaturas_fa_cov_cor_fa_varimax))
fa.diagram(resultados_asignaturas_fa_cov_cor_fa_varimax, simple=FALSE)

#estudios de 16 factores con rotación de promedio máximo
resultados_asignaturas_fa_cov_cor_promax <- fa(resultados_asignaturas_fa_cov_cor, nfactors=5, rotate="promax", fm="pa")
resultados_asignaturas_fa_cov_cor_promax

#diagramas de los factores con rotación de promedio máximo
factor.plot(resultados_asignaturas_fa_cov_cor_promax, labels=rownames(resultados_asignaturas_fa_cov_cor_promax))
fa.diagram(resultados_asignaturas_fa_cov_cor_promax, simple=FALSE)

```

## 9.5. Anexo 5. Script para Análisis de datos de Asignaturas mediante algoritmo K-means

```
#####
#cluster de asignaturas por similitud de asignaturas
#fijamos la aleatoriedad
for(i in 2:n_clusters){
  set.seed(2)
  (resultados_asignaturas_fa_t_kmeans <- kmeans(resultados_asignaturas_fa_t, centers=i, nstart=45))

  #la ejecución de esta gráfica no ofrece una visión esquemática de los cluster
  #plot(resultados_asignaturas_fa_t_kmeans$cluster,
  # rownames(resultados_asignaturas_fa_t), xlab="Clusters", ylab="Asignaturas",
  # main="Clusters Asignaturas", col=resultados_asignaturas_fa_t_kmeans$cluster)

  #visualizamos las bases, les añadimos los clusters calculados
  resultados_asignaturas_fa_t_nmf_basis_matr_clusters <- matriculados_t
  resultados_asignaturas_fa_t_nmf_basis_matr_clusters <- cbind(resultados_asignaturas_fa_t_nmf_basis_matr_clusters,
  resultados_asignaturas_fa_t_kmeans[cluster])
  resultados_asignaturas_fa_t_nmf_basis_matr_clusters

  library("reshape")
  resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr <-
  aggregate(resultados_asignaturas_fa_t_nmf_basis_matr_clusters$Matriculados,
  list(resultados_asignaturas_fa_t_nmf_basis_matr_clusters$cluster),
  sum)

  colnames(resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr) <- c("Cluster", "Matriculaciones")

  #le unimos una columna de matriculaciones de forma porcentual
  resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr <- cbind(resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr,
  (resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr$Matriculaciones*100)/rowSums(matriculados)
  )

  #cambiamos los nombres de nuevo para obtener un formato más visible
  colnames(resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr) <- c("Cluster", "Matriculaciones", "Matr_%")
  resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr
  colSums(resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr)

  barplot(as.matrix(resultados_asignaturas_fa_t_nmf_basis_matr_clusters_aggr$Matriculaciones),
  main="Matriculaciones en cada clúster",
  col=c(3), beside=TRUE,
  ylab="N Matriculados", xlab="Cluster", cex.names=0.5)
}
#####
```

## 9.1. Anexo 6. Script para Análisis de datos de Matriculaciones de Alumnos mediante algoritmo K-means y árbol de decisión

```
#####
#estudio de los datos mediante algoritmo kmeans de manera que agrupamos a los alumnos
#según su parecido en la matriculación
#####
set.seed(2)
num_clusters <- 6
#Algoritmo Kmeans para agrupar alumnos
(resultados_asignaturas_fa_kmeans <- kmeans(resultados_asignaturas_fa, centers=num_clusters, nstart=1000))
resultados_asignaturas_fa_kmeans$cluster

#Cluster de cada alumno, valor sobre la tabla resultados_asignaturas_fa de matriculaciones
View(resultados_asignaturas_fa * resultados_asignaturas_fa_kmeans$cluster)
#número de alumnos en cada asignatura y en cada cluster
(aggregate(resultados_asignaturas_fa * resultados_asignaturas_fa_kmeans$cluster),
 list(resultados_asignaturas_fa_kmeans$cluster),
 sum) / seq(1:num_clusters)[-1]

#número de matriculaciones de cada alumno
num_matriculaciones_alumno <- rowSums(resultados_asignaturas_fa)
num_matriculaciones_alumno

#número de asignaturas aprobadas de cada alumno
num_aprobadas_alumno <- rowSums(resultados_asignaturas_sup)
num_aprobadas_alumno

#ratio de éxito fracaso de cada alumno
exito_del_alumno_n <- round(num_aprobadas_alumno / num_matriculaciones_alumno, 2)
exito_del_alumno_n

#codificamos 0-Fracaso >0-Exito
(exito_del_alumno_EF <- exito_del_alumno_n)
```

```

(exito_del_alumno_EF <- replace(EF <- (exito_del_alumno_EF), EF > 0, 'Exito'))
(exito_del_alumno_EF <- replace(EF <- (exito_del_alumno_EF), EF == 0, 'Fracaso'))

#codificamos 0-Fracaso 1-Exito
(exito_del_alumno_FE01 <- exito_del_alumno_n)
(exito_del_alumno_FE01 <- replace(EF <- (exito_del_alumno_FE01), EF > 0, 1))
(exito_del_alumno_FE01 <- replace(EF <- (exito_del_alumno_FE01), EF == 0, 0))

#tabla de clusters y frecuencias de ratios de éxitos/fracaso
table(resultados_asignaturas_fa_kmeans$cluster, exito_del_alumno_n)

#tabla de cluster y sus frecuencias totales de éxitos y fracasos
barplot(cbind(table(resultados_asignaturas_fa_kmeans$cluster, exito_del_alumno_EF)), col=c(2,3,4,5,8,6),
        ylab="N Matriculados", xlab="Cluster", cex.names=1, legend=list(1,2,3,4,5,6))

install.packages("stringr", dependencies=TRUE)
require(stringr)

#realizamos una lista de las asignaturas que tiene cada Alumno,
# en el mismo orden de aparicion de variables
(asignaturas_alumno_fa <- data.frame(""))
(asignaturas_alumno_sup <- data.frame(""))
(asignaturas_alumno_fa_list <- data.frame(""))
(asignaturas_alumno_sup_list <- data.frame(""))
for(i in 1:4871){
  asignaturas_alumno_fa <- cbind(asignaturas_alumno_fa,
                                t(resultados_asignaturas_fa[i,] * t(matrix(1:45))))
  asignaturas_alumno_fa_list <- cbind(asignaturas_alumno_fa_list,
                                      substr(toString(colnames(as.matrix(t(table(asignaturas_alumno_fa[i,1+1]))))), 3, 30))
  asignaturas_alumno_sup <- cbind(asignaturas_alumno_sup,
                                  t(resultados_asignaturas_sup[i,] * t(matrix(1:45))))
  asignaturas_alumno_sup_list <- cbind(asignaturas_alumno_sup_list,
                                       substr(toString(colnames(as.matrix(t(table(asignaturas_alumno_sup[i,1+1]))))), 3, 30))
}
#eliminamos la primera columna que está vacía
asignaturas_alumno_fa <- asignaturas_alumno_fa[,-1]
asignaturas_alumno_sup <- asignaturas_alumno_sup[,-1]
asignaturas_alumno_fa_list <- t(asignaturas_alumno_fa_list[,-1])
asignaturas_alumno_sup_list <- t(asignaturas_alumno_sup_list[,-1])

rownames(asignaturas_alumno_fa_list) <- seq(1:4871)
rownames(asignaturas_alumno_sup_list) <- seq(1:4871)

colnames(asignaturas_alumno_fa_list) <- c("Asignaturas_Fa")
colnames(asignaturas_alumno_sup_list) <- c("Asignaturas_Sup")

#visualizamos los datos
View(asignaturas_alumno_fa)
View(asignaturas_alumno_sup)
View(asignaturas_alumno_fa_list)
View(asignaturas_alumno_sup_list)

#ahora juntamos todos los datos obtenidos en una única tabla
#reunion de datos de clúster, matrículas, aprobados, ratio y Éxito/Fracaso
alumnos_clusters_exito <- as.data.frame(cbind(resultados_asignaturas_fa_kmeans$cluster,
                                             num_matriculaciones_alumno,
                                             num_aprobadas_alumno,
                                             exito_del_alumno,
                                             exito_del_alumno_EF,
                                             asignaturas_alumno_fa_list,
                                             asignaturas_alumno_sup_list,
                                             rownames(asignaturas_alumno_list)))
colnames(alumnos_clusters_exito) <- c("Cluster", "Matriculadas", "Aprobadas", "RatioEF", "EF", "Asignaturas_Fa", "Asignaturas_Sup", "RegNum")
View(alumnos_clusters_exito)

table(alumnos_clusters_exito[, "Cluster"])

barplot(table(alumnos_clusters_exito[, "Cluster"]), col="green",
        ylab="N Matriculados", xlab="Cluster", cex.names=1)

write.csv2(alumnos_clusters_exito, file="alumnos_clusters_exito.csv")

#realizamos una tabla que nos diga en qué cluster se sitúa cada agrupación de asignaturas
alumnos_cluster_exito_asignaturas_cluster <- as.matrix(table(alumnos_clusters_exito[, "Asignaturas_Fa"], alumnos_clusters_exito[, "Cluster"]))
head(alumnos_cluster_exito_asignaturas_cluster)
colnames(alumnos_cluster_exito_asignaturas_cluster) <- c("Cluster1", "Cluster2", "Cluster3", "Cluster4", "Cluster5", "Cluster6")
write.csv2(alumnos_cluster_exito_asignaturas_cluster, file="alumnos_cluster_exito_asignaturas_cluster.csv")
cbind(alumnos_cluster_exito_asignaturas_cluster)

#realizamos una tabla que nos diga para cada agrupación de asignaturas su relación Éxito/Fracaso
alumnos_cluster_exito_asignaturas_EF <- as.matrix(table(alumnos_clusters_exito[, "Asignaturas_Fa"], alumnos_clusters_exito[, "EF"]))
write.csv2(alumnos_cluster_exito_asignaturas_EF, file="alumnos_cluster_exito_asignaturas_EF.csv")
cbind(alumnos_cluster_exito_asignaturas_EF)

```

```

#tabla cruzada de asignaturas matriculadas (filas) frente asignaturas aprobadas (columnas)
#en esta tabla se ven los grupos de asignaturas que se han fracasado y a qué asignatura se le ha perjudicado mas del grupo
alumnos_cluster_exito_asignaturas_fa_vs_sup_EF <- as.matrix(table(alumnos_clusters_exito[, "Asignaturas_Fa"],
alumnos_clusters_exito[, "Asignaturas_Sup"]))
alumnos_cluster_exito_asignaturas_fa_vs_sup_EF
write.csv2(alumnos_cluster_exito_asignaturas_fa_vs_sup_EF, file="alumnos_cluster_exito_asignaturas_fa_vs_sup_EF.csv")

#experimentar sobre la table de totales de paquetes de asignaturas fracasadas y paquetes de asignaturas aprobadas
View(arbol <- as.data.frame(table(alumnos_clusters_exito[, "Asignaturas_Fa"], alumnos_clusters_exito[, "Asignaturas_Sup"])))
write.csv2(arbol, file="alumnos_cluster_exito_asignaturas_fa_vs_sup_EF_arbol.csv")

#####
#Trabajamos el DataSet que incluye la variable de Éxito_Fracaso para obtener un árbol de decisión
#####
#tomamos la totalidad de los datos de matriculaciones y preparamos un árbol de decisiones en base
# al ratio de Fracaso/Éxito -> [0, 1]
library("party")
library("rpart")

resultados_asignaturas_fa_exito <- cbind(resultados_asignaturas_fa, exito_del_alumno_EF, exito_del_alumno_FE01)

write.csv2(resultados_asignaturas_fa_exito, file="resultados_asignaturas_fa_exito.csv")

resultados_asignaturas_fa_ctree <- ctree(exito_del_alumno_FE01 ~ ., resultados_asignaturas_fa_exito[,c(1:45)])
#mostramos el árbol
resultados_asignaturas_fa_ctree
#dibujamos el árbol de resultados combinados en un modelo simple
plot(resultados_asignaturas_fa_ctree)

#####
#árbol de decisión elaborado mediante el modelo rpart
#cp = complejidad 0.0001
require(rpart, quietly=TRUE)
library("rpart")
library("rattle")
set.seed(2)
resultados_asignaturas_fa_exito_1_a_45 <- resultados_asignaturas_fa_exito
colnames(resultados_asignaturas_fa_exito_1_a_45) <- c(seq(1:45), "EF")
resultados_asignaturas_fa_exito_rpart <- rpart(EF ~ .,
data=resultados_asignaturas_fa_exito_1_a_45[, -47],
method="class",
parms=list(split="information"),
control=rpart.control(cp=0.001, usesurrogate=0, maxsurrogate=0))
print(resultados_asignaturas_fa_exito_rpart)
printcp(resultados_asignaturas_fa_exito_rpart)
fancyRpartPlot(resultados_asignaturas_fa_exito_rpart,
main="Árbol de decisión Matriculaciones vs Éxito/Fracaso cp:0.001")

asRules(resultados_asignaturas_fa_exito_rpart)

#####
#árbol de decisión elaborado mediante el modelo rpart
#cp = complejidad 0.0001
require(rpart, quietly=TRUE)
library("rpart")
library("rattle")
set.seed(2)
resultados_asignaturas_fa_exito_1_a_45 <- resultados_asignaturas_fa_exito
colnames(resultados_asignaturas_fa_exito_1_a_45) <- c(seq(1:45), "EF")
resultados_asignaturas_fa_exito_rpart <- rpart(EF ~ .,
data=resultados_asignaturas_fa_exito_1_a_45[, -47],
method="class",
parms=list(split="information"),
control=rpart.control(cp=0.0001, usesurrogate=0, maxsurrogate=0))
print(resultados_asignaturas_fa_exito_rpart)
printcp(resultados_asignaturas_fa_exito_rpart)
fancyRpartPlot(resultados_asignaturas_fa_exito_rpart,
main="Árbol de decisión Matriculaciones vs Éxito/Fracaso cp:0.0001")

asRules(resultados_asignaturas_fa_exito_rpart)
#####

```

## 9.1. Anexo 7. Tabla con los resultados de paquetes de asignaturas típicas, sus relaciones con los clúster calculados y su relación éxito fracaso del paquete.

A continuación se expone el detalle de datos clasificados por paquetes de asignaturas matriculadas en las que cada número ordinal corresponde con su

asignatura en la tabla principal de asignaturas de la sección 6.3, el clúster al que pertenece según la clasificación del algoritmo k-means y el éxito, fracaso y distancia del paquete de asignaturas. Se puede apreciar que aquellos paquetes de asignaturas cuyo valor E-F es negativo poseen mayor fracaso que éxito:

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
1	0	0	0	22	0	0	16	6	10
1, 10	0	0	0	2	0	0	0	2	-2
1, 10, 20	0	0	0	1	0	0	1	0	1
1, 11	0	0	0	22	0	0	14	8	6
1, 11, 13	0	0	0	1	0	0	0	1	-1
1, 11, 20	0	0	0	1	0	0	0	1	-1
1, 11, 22	0	0	0	1	0	0	1	0	1
1, 11, 27	0	0	0	1	0	0	0	1	-1
1, 12, 14	0	0	0	1	0	0	1	0	1
1, 12, 18, 21, 28	0	0	0	1	0	0	1	0	1
1, 13	0	0	0	3	0	0	2	1	1
1, 14	0	0	0	2	0	0	2	0	2
1, 15	0	0	0	2	0	0	1	1	0
1, 15, 30	0	0	0	1	0	0	1	0	1
1, 16	0	0	0	1	0	0	1	0	1
1, 17, 26, 28	0	0	0	1	0	0	1	0	1
1, 2	122	0	0	0	0	0	91	31	60
1, 2, 10	1	0	0	0	0	0	1	0	1
1, 2, 10, 13, 20, 25	1	0	0	0	0	0	1	0	1
1, 2, 11	1	0	0	0	0	0	0	1	-1
1, 2, 12	4	0	0	0	0	0	4	0	4
1, 2, 13	1	0	0	0	0	0	1	0	1
1, 2, 14	2	0	0	0	0	0	2	0	2
1, 2, 20	1	0	0	0	0	0	1	0	1
1, 2, 3	0	0	238	0	0	0	205	33	172
1, 2, 3, 10	0	0	1	0	0	0	1	0	1
1, 2, 3, 11	0	0	3	0	0	0	3	0	3
1, 2, 3, 12, 19, 20	0	0	1	0	0	0	1	0	1
1, 2, 3, 13	0	0	2	0	0	0	2	0	2
1, 2, 3, 4	0	0	0	0	0	97	86	11	75
1, 2, 3, 4, 10	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 10, 24, 30, 34	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 11	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 16, 17, 26, 28	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 20, 21, 27	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 5	0	0	0	0	0	210	176	34	142
1, 2, 3, 4, 5, 15	0	0	0	0	0	2	2	0	2
1, 2, 3, 4, 5, 23	0	0	0	0	0	1	0	1	-1
1, 2, 3, 4, 5, 6	0	0	0	0	0	5	4	1	3
1, 2, 3, 4, 5, 6, 7, 9	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 5, 7	0	0	0	0	0	2	2	0	2
1, 2, 3, 4, 5, 7, 8	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 5, 8	0	0	0	0	0	8	7	1	6
1, 2, 3, 4, 5, 8, 13, 23	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 5, 9	0	0	0	0	0	2	2	0	2
1, 2, 3, 4, 6	0	0	0	0	0	4	2	2	0
1, 2, 3, 4, 6, 11	0	0	0	0	0	2	1	1	0
1, 2, 3, 4, 6, 8	0	0	0	0	0	2	2	0	2
1, 2, 3, 4, 7	0	0	0	0	0	3	2	1	1
1, 2, 3, 4, 8	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 9	0	0	0	0	0	1	1	0	1
1, 2, 3, 4, 9, 10	0	0	0	0	0	1	0	1	-1
1, 2, 3, 5	0	0	41	0	0	0	34	7	27
1, 2, 3, 5, 10	0	0	3	0	0	0	3	0	3
1, 2, 3, 5, 10, 15, 24	0	0	1	0	0	0	1	0	1
1, 2, 3, 5, 13	0	0	1	0	0	0	1	0	1
1, 2, 3, 5, 6	0	0	5	0	0	0	3	2	1
1, 2, 3, 5, 6, 10	0	0	1	0	0	0	1	0	1
1, 2, 3, 5, 7	0	0	1	0	0	0	1	0	1
1, 2, 3, 5, 8	0	0	1	0	0	0	1	0	1
1, 2, 3, 5, 8, 10	0	0	1	0	0	0	1	0	1
1, 2, 3, 5, 9	0	0	1	0	0	0	0	1	-1
1, 2, 3, 5, 9, 10	0	0	1	0	0	0	0	1	-1
1, 2, 3, 6	0	0	10	0	0	0	8	2	6
1, 2, 3, 6, 22	0	0	1	0	0	0	1	0	1
1, 2, 3, 6, 32	0	0	1	0	0	0	1	0	1
1, 2, 3, 6, 7	0	0	1	0	0	0	1	0	1
1, 2, 3, 6, 7, 8	0	0	1	0	0	0	1	0	1
1, 2, 3, 7	0	0	2	0	0	0	1	1	0
1, 2, 3, 8	0	0	11	0	0	0	9	2	7
1, 2, 3, 8, 26, 27	0	0	1	0	0	0	1	0	1
1, 2, 3, 9	0	0	4	0	0	0	2	2	0

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
1, 2, 30	1	0	0	0	0	0	1	0	1
1, 2, 4	0	0	0	0	0	72	55	17	38
1, 2, 4, 10	0	0	0	0	0	2	2	0	2
1, 2, 4, 10, 11	0	0	0	0	0	1	1	0	1
1, 2, 4, 13	0	0	0	0	0	1	1	0	1
1, 2, 4, 5	0	0	0	0	0	58	46	12	34
1, 2, 4, 5, 10	0	0	0	0	0	2	2	0	2
1, 2, 4, 5, 13	0	0	0	0	0	1	1	0	1
1, 2, 4, 5, 24	0	0	0	0	0	1	1	0	1
1, 2, 4, 5, 6	0	0	0	0	0	3	3	0	3
1, 2, 4, 5, 6, 7, 8, 11, 31	0	0	0	0	0	1	1	0	1
1, 2, 4, 5, 7	0	0	0	0	0	10	10	0	10
1, 2, 4, 5, 7, 11	0	0	0	0	0	1	0	1	-1
1, 2, 4, 5, 8	0	0	0	0	0	1	1	0	1
1, 2, 4, 5, 9	0	0	0	0	0	1	0	1	-1
1, 2, 4, 6	0	0	0	0	0	2	1	1	0
1, 2, 4, 6, 10	0	0	0	0	0	1	1	0	1
1, 2, 4, 6, 7	0	0	0	0	0	1	1	0	1
1, 2, 4, 6, 7, 10	0	0	0	0	0	1	1	0	1
1, 2, 4, 6, 7, 9	0	0	0	0	0	1	1	0	1
1, 2, 4, 6, 9	0	0	0	0	0	1	1	0	1
1, 2, 4, 6, 9, 13	0	0	0	0	0	1	1	0	1
1, 2, 4, 7	0	0	0	0	0	2	2	0	2
1, 2, 4, 8	0	0	0	0	0	3	2	1	1
1, 2, 4, 9	0	0	0	0	0	1	1	0	1
1, 2, 5	48	0	0	0	0	0	37	11	26
1, 2, 5, 10	3	0	0	0	0	0	2	1	1
1, 2, 5, 11, 13	1	0	0	0	0	0	0	1	-1
1, 2, 5, 6	4	0	0	0	0	0	4	0	4
1, 2, 5, 6, 7	2	0	0	0	0	0	2	0	2
1, 2, 5, 6, 8	1	0	0	0	0	0	0	1	-1
1, 2, 5, 7, 14	1	0	0	0	0	0	1	0	1
1, 2, 5, 7, 8, 10	1	0	0	0	0	0	1	0	1
1, 2, 5, 7, 9	1	0	0	0	0	0	1	0	1
1, 2, 5, 8	2	0	0	0	0	0	1	1	0
1, 2, 5, 8, 12, 14	1	0	0	0	0	0	1	0	1
1, 2, 6	11	0	0	0	0	0	8	3	5
1, 2, 6, 10	1	0	0	0	0	0	1	0	1
1, 2, 6, 26	1	0	0	0	0	0	0	1	-1
1, 2, 6, 8	2	0	0	0	0	0	1	1	0
1, 2, 6, 8, 10, 24	1	0	0	0	0	0	1	0	1
1, 2, 6, 9	1	0	0	0	0	0	1	0	1
1, 2, 6, 9, 10	1	0	0	0	0	0	1	0	1
1, 2, 7	9	0	0	0	0	0	7	2	5
1, 2, 7, 14	1	0	0	0	0	0	1	0	1
1, 2, 7, 15	1	0	0	0	0	0	1	0	1
1, 2, 7, 20	1	0	0	0	0	0	1	0	1
1, 2, 7, 8	3	0	0	0	0	0	3	0	3
1, 2, 8	5	0	0	0	0	0	4	1	3
1, 2, 9	5	0	0	0	0	0	3	2	1
1, 20	0	0	0	1	0	0	0	1	-1
1, 20, 23	0	0	0	1	0	0	1	0	1
1, 20, 23, 30	0	0	0	1	0	0	1	0	1
1, 23	0	0	0	1	0	0	0	1	-1
1, 24, 31, 36	0	0	0	1	0	0	1	0	1
1, 26	0	0	0	1	0	0	1	0	1
1, 27	0	0	0	1	0	0	1	0	1
1, 3	0	0	278	0	0	0	227	51	176
1, 3, 10	0	0	3	0	0	0	3	0	3
1, 3, 11	0	0	14	0	0	0	10	4	6
1, 3, 12	0	0	1	0	0	0	1	0	1
1, 3, 12, 19	0	0	1	0	0	0	1	0	1
1, 3, 13	0	0	4	0	0	0	4	0	4
1, 3, 14	0	0	1	0	0	0	1	0	1
1, 3, 23	0	0	1	0	0	0	1	0	1
1, 3, 25	0	0	1	0	0	0	1	0	1
1, 3, 29	0	0	1	0	0	0	0	1	-1
1, 3, 4	0	173	0	0	0	0	139	34	105
1, 3, 4, 10	0	1	0	0	0	0	1	0	1
1, 3, 4, 11	0	7	0	0	0	0	5	2	3
1, 3, 4, 11, 33	0	1	0	0	0	0	1	0	1
1, 3, 4, 12	0	1	0	0	0	0	1	0	1
1, 3, 4, 16	0	1	0	0	0	0	0	1	-1
1, 3, 4, 5	0	33	0	0	0	0	26	7	19
1, 3, 4, 5, 10	0	1	0	0	0	0	1	0	1
1, 3, 4, 5, 11	0	2	0	0	0	0	1	1	0
1, 3, 4, 5, 6	0	2	0	0	0	0	2	0	2
1, 3, 4, 5, 6, 11	0	1	0	0	0	0	0	1	-1
1, 3, 4, 5, 6, 7	0	1	0	0	0	0	1	0	1
1, 3, 4, 5, 6, 7, 10	0	3	0	0	0	0	3	0	3

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
1, 3, 4, 5, 7	0	1	0	0	0	0	1	0	1
1, 3, 4, 5, 7, 9	0	1	0	0	0	0	1	0	1
1, 3, 4, 5, 8	0	1	0	0	0	0	1	0	1
1, 3, 4, 5, 9	0	2	0	0	0	0	1	1	0
1, 3, 4, 6	0	4	0	0	0	0	3	1	2
1, 3, 4, 6, 11	0	1	0	0	0	0	0	1	-1
1, 3, 4, 6, 11, 23, 33, 39	0	1	0	0	0	0	1	0	1
1, 3, 4, 6, 11, 38	0	1	0	0	0	0	1	0	1
1, 3, 4, 6, 8	0	2	0	0	0	0	2	0	2
1, 3, 4, 6, 9	0	1	0	0	0	0	1	0	1
1, 3, 4, 7	0	2	0	0	0	0	2	0	2
1, 3, 4, 7, 11, 14	0	1	0	0	0	0	1	0	1
1, 3, 4, 8	0	7	0	0	0	0	7	0	7
1, 3, 4, 8, 11	0	1	0	0	0	0	1	0	1
1, 3, 5	0	0	75	0	0	0	58	17	41
1, 3, 5, 10	0	0	2	0	0	0	2	0	2
1, 3, 5, 31	0	0	1	0	0	0	1	0	1
1, 3, 5, 6	0	0	3	0	0	0	3	0	3
1, 3, 5, 6, 7	0	0	1	0	0	0	1	0	1
1, 3, 5, 6, 8	0	0	1	0	0	0	1	0	1
1, 3, 5, 6, 9	0	0	1	0	0	0	1	0	1
1, 3, 5, 7	0	0	3	0	0	0	3	0	3
1, 3, 5, 7, 31	0	0	1	0	0	0	1	0	1
1, 3, 5, 8	0	0	3	0	0	0	2	1	1
1, 3, 5, 9	0	0	2	0	0	0	1	1	0
1, 3, 6	0	0	32	0	0	0	24	8	16
1, 3, 6, 13	0	0	1	0	0	0	1	0	1
1, 3, 6, 7	0	0	5	0	0	0	3	2	1
1, 3, 6, 7, 9, 11	0	0	1	0	0	0	1	0	1
1, 3, 6, 8	0	0	2	0	0	0	2	0	2
1, 3, 6, 8, 11	0	0	1	0	0	0	0	1	-1
1, 3, 6, 9	0	0	3	0	0	0	3	0	3
1, 3, 6, 9, 11, 13	0	0	2	0	0	0	0	2	-2
1, 3, 7	0	0	11	0	0	0	10	1	9
1, 3, 7, 11	0	0	2	0	0	0	1	1	0
1, 3, 7, 9	0	0	2	0	0	0	2	0	2
1, 3, 7, 9, 10	0	0	1	0	0	0	1	0	1
1, 3, 8	0	0	18	0	0	0	12	6	6
1, 3, 8, 11	0	0	2	0	0	0	1	1	0
1, 3, 8, 9	0	0	2	0	0	0	2	0	2
1, 3, 9	0	0	8	0	0	0	7	1	6
1, 3, 9, 11	0	0	1	0	0	0	1	0	1
1, 3, 9, 13	0	0	1	0	0	0	1	0	1
1, 30	0	0	0	1	0	0	1	0	1
1, 33	0	0	0	1	0	0	1	0	1
1, 34	0	0	0	1	0	0	1	0	1
1, 4	0	170	0	0	0	0	126	44	82
1, 4, 10	0	1	0	0	0	0	1	0	1
1, 4, 11	0	24	0	0	0	0	18	6	12
1, 4, 11, 36	0	1	0	0	0	0	0	1	-1
1, 4, 12	0	1	0	0	0	0	0	1	-1
1, 4, 13	0	3	0	0	0	0	3	0	3
1, 4, 14	0	1	0	0	0	0	1	0	1
1, 4, 15	0	1	0	0	0	0	1	0	1
1, 4, 22	0	1	0	0	0	0	0	1	-1
1, 4, 35	0	1	0	0	0	0	1	0	1
1, 4, 5	0	110	0	0	0	0	85	25	60
1, 4, 5, 10	0	1	0	0	0	0	1	0	1
1, 4, 5, 11	0	1	0	0	0	0	1	0	1
1, 4, 5, 6	0	3	0	0	0	0	1	2	-1
1, 4, 5, 6, 10	0	1	0	0	0	0	0	1	-1
1, 4, 5, 6, 10, 13	0	1	0	0	0	0	1	0	1
1, 4, 5, 6, 13	0	1	0	0	0	0	1	0	1
1, 4, 5, 6, 7, 9	0	2	0	0	0	0	2	0	2
1, 4, 5, 6, 9	0	4	0	0	0	0	3	1	2
1, 4, 5, 6, 9, 10	0	1	0	0	0	0	1	0	1
1, 4, 5, 7	0	2	0	0	0	0	2	0	2
1, 4, 5, 7, 8, 17	0	1	0	0	0	0	1	0	1
1, 4, 5, 7, 9, 10	0	1	0	0	0	0	0	1	-1
1, 4, 5, 8	0	2	0	0	0	0	0	2	-2
1, 4, 5, 8, 9, 10	0	1	0	0	0	0	1	0	1
1, 4, 5, 9	0	3	0	0	0	0	3	0	3
1, 4, 5, 9, 13	0	1	0	0	0	0	0	1	-1
1, 4, 6	0	26	0	0	0	0	19	7	12
1, 4, 6, 10	0	1	0	0	0	0	0	1	-1
1, 4, 6, 11	0	1	0	0	0	0	1	0	1
1, 4, 6, 7	0	4	0	0	0	0	1	3	-2
1, 4, 6, 7, 10	0	1	0	0	0	0	1	0	1
1, 4, 6, 7, 11	0	1	0	0	0	0	1	0	1
1, 4, 6, 7, 8, 9	0	1	0	0	0	0	1	0	1

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
1, 4, 6, 8	0	1	0	0	0	0	1	0	1
1, 4, 6, 8, 11, 21	0	1	0	0	0	0	1	0	1
1, 4, 6, 9	0	1	0	0	0	0	1	0	1
1, 4, 6, 9, 10	0	1	0	0	0	0	1	0	1
1, 4, 6, 9, 10, 13	0	1	0	0	0	0	0	1	-1
1, 4, 7	0	15	0	0	0	0	12	3	9
1, 4, 7, 10	0	2	0	0	0	0	1	1	0
1, 4, 7, 11	0	1	0	0	0	0	0	1	-1
1, 4, 7, 16	0	1	0	0	0	0	0	1	-1
1, 4, 7, 8	0	6	0	0	0	0	4	2	2
1, 4, 7, 9	0	2	0	0	0	0	0	2	-2
1, 4, 7, 9, 11, 13	0	1	0	0	0	0	0	1	-1
1, 4, 8	0	4	0	0	0	0	1	3	-2
1, 4, 8, 11	0	1	0	0	0	0	1	0	1
1, 4, 8, 13	0	1	0	0	0	0	1	0	1
1, 4, 8, 9	0	3	0	0	0	0	2	1	1
1, 4, 9	0	8	0	0	0	0	5	3	2
1, 5	0	0	0	58	0	0	45	13	32
1, 5, 10	0	0	0	5	0	0	4	1	3
1, 5, 10, 11	0	0	0	1	0	0	0	1	-1
1, 5, 11	0	0	0	1	0	0	1	0	1
1, 5, 23	0	0	0	1	0	0	0	1	-1
1, 5, 33	0	0	0	1	0	0	1	0	1
1, 5, 6	0	0	0	11	0	0	7	4	3
1, 5, 6, 10	0	0	0	2	0	0	2	0	2
1, 5, 6, 10, 13	0	0	0	1	0	0	1	0	1
1, 5, 6, 15	0	0	0	1	0	0	1	0	1
1, 5, 6, 20	0	0	0	1	0	0	1	0	1
1, 5, 6, 7	0	0	0	1	0	0	1	0	1
1, 5, 6, 7, 10	0	0	0	1	0	0	1	0	1
1, 5, 6, 9, 10	0	0	0	1	0	0	1	0	1
1, 5, 7	0	0	0	6	0	0	6	0	6
1, 5, 7, 10	0	0	0	1	0	0	0	1	-1
1, 5, 7, 45	0	0	0	1	0	0	0	1	-1
1, 5, 7, 9, 10	0	0	0	1	0	0	1	0	1
1, 5, 8	0	0	0	3	0	0	3	0	3
1, 5, 8, 32	0	0	0	1	0	0	1	0	1
1, 5, 9	0	0	0	2	0	0	1	1	0
1, 5, 9, 10	0	0	0	1	0	0	1	0	1
1, 6	0	0	0	34	0	0	23	11	12
1, 6, 10	0	0	0	1	0	0	1	0	1
1, 6, 11	0	0	0	6	0	0	3	3	0
1, 6, 11, 13	0	0	0	1	0	0	0	1	-1
1, 6, 11, 13, 18, 27	0	0	0	1	0	0	0	1	-1
1, 6, 11, 13, 31	0	0	0	1	0	0	1	0	1
1, 6, 11, 33, 39	0	0	0	1	0	0	1	0	1
1, 6, 13	0	0	0	2	0	0	2	0	2
1, 6, 15	0	0	0	1	0	0	1	0	1
1, 6, 16, 23	0	0	0	1	0	0	0	1	-1
1, 6, 20, 40	0	0	0	1	0	0	1	0	1
1, 6, 26	0	0	0	1	0	0	0	1	-1
1, 6, 7	0	0	0	5	0	0	3	2	1
1, 6, 7, 15	0	0	0	1	0	0	1	0	1
1, 6, 7, 8, 10	0	0	0	1	0	0	1	0	1
1, 6, 7, 8, 9	0	0	0	1	0	0	1	0	1
1, 6, 7, 8, 9, 15	0	0	0	1	0	0	1	0	1
1, 6, 7, 9	0	0	0	1	0	0	1	0	1
1, 6, 8	0	0	0	3	0	0	1	2	-1
1, 6, 8, 11	0	0	0	1	0	0	0	1	-1
1, 6, 8, 12, 20	0	0	0	1	0	0	1	0	1
1, 6, 8, 9	0	0	0	1	0	0	1	0	1
1, 6, 9	0	0	0	1	0	0	1	0	1
1, 6, 9, 13	0	0	0	3	0	0	3	0	3
1, 7	0	0	0	21	0	0	18	3	15
1, 7, 10	0	0	0	2	0	0	2	0	2
1, 7, 11	0	0	0	2	0	0	1	1	0
1, 7, 11, 31	0	0	0	1	0	0	1	0	1
1, 7, 11, 33	0	0	0	1	0	0	0	1	-1
1, 7, 13	0	0	0	3	0	0	3	0	3
1, 7, 14	0	0	0	1	0	0	0	1	-1
1, 7, 8	0	0	0	4	0	0	4	0	4
1, 7, 8, 11	0	0	0	1	0	0	0	1	-1
1, 7, 9	0	0	0	3	0	0	3	0	3
1, 7, 9, 11, 13, 36	0	0	0	1	0	0	1	0	1
1, 8	0	0	0	10	0	0	7	3	4
1, 8, 10	0	0	0	1	0	0	0	1	-1
1, 8, 11	0	0	0	3	0	0	2	1	1
1, 8, 13	0	0	0	1	0	0	1	0	1
1, 8, 23	0	0	0	1	0	0	1	0	1
1, 8, 9	0	0	0	2	0	0	2	0	2

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
1, 9	0	0	0	7	0	0	5	2	3
1, 9, 11	0	0	0	4	0	0	3	1	2
1, 9, 25	0	0	0	1	0	0	1	0	1
10, 11, 17, 24, 26	0	0	0	1	0	0	1	0	1
10, 12, 18, 20	0	0	0	1	0	0	1	0	1
10, 13, 23	0	0	0	1	0	0	0	1	-1
10, 16	0	0	0	3	0	0	1	2	-1
10, 16, 17, 18, 20, 25	0	0	0	1	0	0	1	0	1
10, 16, 17, 18, 25, 26	0	0	0	1	0	0	1	0	1
10, 16, 17, 19, 20, 23	0	0	0	1	0	0	0	1	-1
10, 16, 20, 26	0	0	0	1	0	0	0	1	-1
10, 16, 38	0	0	0	1	0	0	1	0	1
10, 17, 20, 25	0	0	0	1	0	0	0	1	-1
10, 20	0	0	0	1	0	0	0	1	-1
10, 29	0	0	0	1	0	0	1	0	1
11	0	0	0	2	0	0	2	0	2
11, 12	0	0	0	2	0	0	2	0	2
11, 12, 19	0	0	0	1	0	0	1	0	1
11, 12, 23, 27, 34	0	0	0	1	0	0	1	0	1
11, 13	0	0	0	2	0	0	1	1	0
11, 14	0	0	0	1	0	0	1	0	1
11, 14, 31	0	0	0	1	0	0	0	1	-1
11, 16, 22, 23, 26	0	0	0	1	0	0	1	0	1
11, 20, 30	0	0	0	1	0	0	1	0	1
11, 22	0	0	0	1	0	0	1	0	1
11, 23	0	0	0	1	0	0	1	0	1
11, 24	0	0	0	1	0	0	1	0	1
11, 26	0	0	0	1	0	0	0	1	-1
11, 42	0	0	0	1	0	0	1	0	1
12	0	0	0	3	0	0	2	1	1
12, 14	0	0	0	6	0	0	6	0	6
12, 14, 15, 18, 25, 28	0	0	0	1	0	0	1	0	1
12, 14, 16	0	0	0	1	0	0	1	0	1
12, 14, 17, 24, 28	0	0	0	1	0	0	1	0	1
12, 14, 19, 24, 26, 27	0	0	0	1	0	0	0	1	-1
12, 14, 20	0	0	0	1	0	0	1	0	1
12, 14, 25	0	0	0	1	0	0	1	0	1
12, 15, 18	0	0	0	1	0	0	1	0	1
12, 15, 19	0	0	0	1	0	0	1	0	1
12, 16	0	0	0	3	0	0	3	0	3
12, 16, 18	0	0	0	2	0	0	2	0	2
12, 17, 26, 28	0	0	0	1	0	0	0	1	-1
12, 19	0	0	0	1	0	0	1	0	1
12, 19, 20	0	0	0	1	0	0	1	0	1
12, 20	0	0	0	2	0	0	1	1	0
12, 21	0	0	0	1	0	0	1	0	1
12, 22	0	0	0	1	0	0	1	0	1
12, 23	0	0	0	3	0	0	1	2	-1
12, 24, 26	0	0	0	1	0	0	1	0	1
12, 25	0	0	0	3	0	0	3	0	3
12, 27, 28, 34	0	0	0	1	0	0	0	1	-1
12, 28	0	0	0	1	0	0	1	0	1
12, 28, 29, 34	0	0	0	1	0	0	1	0	1
13	0	0	0	3	0	0	1	2	-1
13, 14	0	0	0	1	0	0	0	1	-1
13, 14, 20, 36	0	0	0	1	0	0	1	0	1
13, 15, 18, 25, 37, 44	0	0	0	1	0	0	0	1	-1
13, 15, 24	0	0	0	1	0	0	1	0	1
13, 17	0	0	0	3	0	0	3	0	3
13, 17, 26	0	0	0	1	0	0	1	0	1
13, 17, 26, 29	0	0	0	1	0	0	0	1	-1
13, 19	0	0	0	1	0	0	1	0	1
13, 21, 23, 24, 34	0	0	0	1	0	0	0	1	-1
13, 21, 30	0	0	0	1	0	0	1	0	1
13, 26	0	0	0	1	0	0	0	1	-1
13, 27	0	0	0	1	0	0	0	1	-1
14	0	0	0	2	0	0	1	1	0
14, 15, 18	0	0	0	1	0	0	1	0	1
14, 16, 17, 23	0	0	0	1	0	0	1	0	1
14, 16, 18, 19, 36	0	0	0	1	0	0	1	0	1
14, 16, 18, 38	0	0	0	1	0	0	1	0	1
14, 17	0	0	0	1	0	0	0	1	-1
14, 19	0	0	0	1	0	0	1	0	1
14, 20	0	0	0	1	0	0	1	0	1
14, 21, 41	0	0	0	1	0	0	1	0	1
14, 22, 29	0	0	0	1	0	0	1	0	1
14, 25	0	0	0	1	0	0	1	0	1
14, 27	0	0	0	2	0	0	2	0	2
14, 37	0	0	0	1	0	0	0	1	-1
14, 44	0	0	0	1	0	0	1	0	1

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
15	0	0	0	2	0	0	1	1	0
15, 17	0	0	0	2	0	0	1	1	0
15, 20, 21, 24	0	0	0	1	0	0	1	0	1
15, 22, 28	0	0	0	1	0	0	1	0	1
15, 25	0	0	0	1	0	0	1	0	1
15, 27	0	0	0	2	0	0	0	2	-2
15, 33	0	0	0	1	0	0	0	1	-1
16	0	0	0	4	0	0	3	1	2
16, 17	0	0	0	3	0	0	1	2	-1
16, 17, 18	0	0	0	1	0	0	1	0	1
16, 17, 18, 20	0	0	0	1	0	0	1	0	1
16, 17, 26, 28	0	0	0	1	0	0	0	1	-1
16, 18	0	0	0	4	0	0	3	1	2
16, 18, 20	0	0	0	1	0	0	1	0	1
16, 18, 21, 26	0	0	0	1	0	0	0	1	-1
16, 18, 41	0	0	0	1	0	0	1	0	1
16, 19	0	0	0	2	0	0	2	0	2
16, 22	0	0	0	1	0	0	0	1	-1
16, 26	0	0	0	1	0	0	1	0	1
17, 18	0	0	0	1	0	0	1	0	1
17, 18, 26	0	0	0	1	0	0	0	1	-1
17, 20	0	0	0	1	0	0	0	1	-1
17, 20, 40	0	0	0	1	0	0	0	1	-1
17, 21, 23, 27, 40	0	0	0	1	0	0	1	0	1
17, 23	0	0	0	1	0	0	1	0	1
17, 30	0	0	0	1	0	0	1	0	1
18, 19, 21, 42	0	0	0	1	0	0	1	0	1
18, 19, 25	0	0	0	1	0	0	1	0	1
18, 22	0	0	0	2	0	0	1	1	0
18, 22, 26, 29	0	0	0	1	0	0	0	1	-1
19	0	0	0	1	0	0	1	0	1
19, 20	0	0	0	1	0	0	1	0	1
19, 20, 22	0	0	0	1	0	0	1	0	1
19, 23	0	0	0	2	0	0	2	0	2
2	22	0	0	0	0	0	16	6	10
2, 10	4	0	0	0	0	0	4	0	4
2, 10, 11, 12, 32	1	0	0	0	0	0	1	0	1
2, 10, 13, 23	1	0	0	0	0	0	1	0	1
2, 10, 13, 23, 24	1	0	0	0	0	0	1	0	1
2, 11	5	0	0	0	0	0	3	2	1
2, 11, 17, 20	1	0	0	0	0	0	1	0	1
2, 11, 24	1	0	0	0	0	0	1	0	1
2, 12	17	0	0	0	0	0	12	5	7
2, 12, 13	2	0	0	0	0	0	2	0	2
2, 12, 13, 15	1	0	0	0	0	0	1	0	1
2, 12, 14	6	0	0	0	0	0	6	0	6
2, 12, 14, 16	1	0	0	0	0	0	1	0	1
2, 12, 14, 31, 35	1	0	0	0	0	0	1	0	1
2, 12, 14, 34	1	0	0	0	0	0	1	0	1
2, 12, 15, 17, 19, 20	1	0	0	0	0	0	1	0	1
2, 12, 15, 28	1	0	0	0	0	0	1	0	1
2, 12, 17	1	0	0	0	0	0	1	0	1
2, 12, 17, 23	1	0	0	0	0	0	0	1	-1
2, 12, 17, 26, 28, 30	1	0	0	0	0	0	1	0	1
2, 12, 18	1	0	0	0	0	0	1	0	1
2, 12, 18, 19	1	0	0	0	0	0	1	0	1
2, 12, 18, 19, 21, 22	1	0	0	0	0	0	1	0	1
2, 12, 18, 21, 22, 39	1	0	0	0	0	0	1	0	1
2, 12, 18, 30	1	0	0	0	0	0	1	0	1
2, 12, 19	5	0	0	0	0	0	5	0	5
2, 12, 19, 21, 22, 23	1	0	0	0	0	0	1	0	1
2, 12, 19, 21, 32, 42	1	0	0	0	0	0	1	0	1
2, 12, 19, 22, 23, 42	1	0	0	0	0	0	1	0	1
2, 12, 19, 41	1	0	0	0	0	0	1	0	1
2, 12, 21	3	0	0	0	0	0	3	0	3
2, 12, 22, 28	1	0	0	0	0	0	1	0	1
2, 12, 29	1	0	0	0	0	0	0	1	-1
2, 12, 31	1	0	0	0	0	0	0	1	-1
2, 12, 32	1	0	0	0	0	0	1	0	1
2, 13	1	0	0	0	0	0	1	0	1
2, 13, 17	1	0	0	0	0	0	1	0	1
2, 13, 23	1	0	0	0	0	0	1	0	1
2, 13, 23, 24	1	0	0	0	0	0	1	0	1
2, 13, 25	1	0	0	0	0	0	1	0	1
2, 14	8	0	0	0	0	0	8	0	8
2, 14, 16	1	0	0	0	0	0	1	0	1
2, 14, 18, 19	1	0	0	0	0	0	1	0	1
2, 14, 18, 19, 21	1	0	0	0	0	0	1	0	1
2, 14, 20, 30, 31	1	0	0	0	0	0	1	0	1
2, 15	4	0	0	0	0	0	3	1	2

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
2, 15, 20, 24, 25	1	0	0	0	0	0	1	0	1
2, 15, 24	3	0	0	0	0	0	3	0	3
2, 15, 25	1	0	0	0	0	0	1	0	1
2, 15, 25, 27	1	0	0	0	0	0	1	0	1
2, 16	1	0	0	0	0	0	1	0	1
2, 16, 18	1	0	0	0	0	0	1	0	1
2, 16, 43	1	0	0	0	0	0	1	0	1
2, 17	1	0	0	0	0	0	1	0	1
2, 18	1	0	0	0	0	0	1	0	1
2, 18, 21	1	0	0	0	0	0	1	0	1
2, 19	2	0	0	0	0	0	2	0	2
2, 19, 20	1	0	0	0	0	0	1	0	1
2, 19, 20, 22	1	0	0	0	0	0	1	0	1
2, 19, 21, 22	1	0	0	0	0	0	1	0	1
2, 19, 23	1	0	0	0	0	0	1	0	1
2, 21	2	0	0	0	0	0	1	1	0
2, 22	2	0	0	0	0	0	2	0	2
2, 25	2	0	0	0	0	0	2	0	2
2, 26	2	0	0	0	0	0	1	1	0
2, 27	1	0	0	0	0	0	1	0	1
2, 28	2	0	0	0	0	0	2	0	2
2, 3	0	0	0	0	187	0	161	26	135
2, 3, 10	0	0	0	0	4	0	3	1	2
2, 3, 10, 13	0	0	0	0	1	0	1	0	1
2, 3, 10, 15, 18	0	0	0	0	1	0	1	0	1
2, 3, 11	0	0	0	0	3	0	3	0	3
2, 3, 11, 13	0	0	0	0	1	0	0	1	-1
2, 3, 11, 24	0	0	0	0	1	0	1	0	1
2, 3, 12	0	0	0	0	17	0	16	1	15
2, 3, 12, 13	0	0	0	0	1	0	1	0	1
2, 3, 12, 14	0	0	0	0	3	0	3	0	3
2, 3, 12, 14, 19	0	0	0	0	1	0	1	0	1
2, 3, 12, 14, 20	0	0	0	0	1	0	1	0	1
2, 3, 12, 14, 21	0	0	0	0	1	0	1	0	1
2, 3, 12, 14, 35	0	0	0	0	1	0	1	0	1
2, 3, 12, 15	0	0	0	0	3	0	3	0	3
2, 3, 12, 15, 17	0	0	0	0	1	0	1	0	1
2, 3, 12, 16	0	0	0	0	1	0	1	0	1
2, 3, 12, 17, 37	0	0	0	0	1	0	1	0	1
2, 3, 12, 18, 35	0	0	0	0	1	0	1	0	1
2, 3, 12, 19	0	0	0	0	2	0	2	0	2
2, 3, 12, 19, 21	0	0	0	0	1	0	1	0	1
2, 3, 12, 43	0	0	0	0	1	0	1	0	1
2, 3, 13	0	0	0	0	3	0	3	0	3
2, 3, 13, 15	0	0	0	0	2	0	2	0	2
2, 3, 13, 23, 24	0	0	0	0	1	0	1	0	1
2, 3, 13, 24	0	0	0	0	3	0	2	1	1
2, 3, 14	0	0	0	0	3	0	2	1	1
2, 3, 14, 15	0	0	0	0	1	0	1	0	1
2, 3, 14, 16	0	0	0	0	1	0	1	0	1
2, 3, 14, 18, 38	0	0	0	0	1	0	1	0	1
2, 3, 14, 21	0	0	0	0	1	0	1	0	1
2, 3, 14, 21, 36	0	0	0	0	1	0	0	1	-1
2, 3, 15	0	0	0	0	6	0	5	1	4
2, 3, 15, 20, 24	0	0	0	0	1	0	1	0	1
2, 3, 15, 24	0	0	0	0	3	0	1	2	-1
2, 3, 16	0	0	0	0	2	0	2	0	2
2, 3, 16, 17	0	0	0	0	1	0	1	0	1
2, 3, 17	0	0	0	0	1	0	1	0	1
2, 3, 18	0	0	0	0	2	0	1	1	0
2, 3, 18, 43	0	0	0	0	1	0	0	1	-1
2, 3, 19, 25	0	0	0	0	1	0	1	0	1
2, 3, 20	0	0	0	0	3	0	2	1	1
2, 3, 21, 25	0	0	0	0	1	0	0	1	-1
2, 3, 23	0	0	0	0	1	0	1	0	1
2, 3, 24	0	0	0	0	1	0	1	0	1
2, 3, 24, 25	0	0	0	0	1	0	1	0	1
2, 3, 26	0	0	0	0	1	0	1	0	1
2, 3, 26, 38	0	0	0	0	1	0	1	0	1
2, 3, 27	0	0	0	0	1	0	1	0	1
2, 3, 28	0	0	0	0	2	0	2	0	2
2, 3, 29	0	0	0	0	1	0	1	0	1
2, 3, 30	0	0	0	0	1	0	1	0	1
2, 3, 38	0	0	0	0	2	0	0	2	-2
2, 3, 39	0	0	0	0	1	0	1	0	1
2, 3, 4	0	0	0	0	36	0	29	7	22
2, 3, 4, 11, 20, 31	0	0	0	0	1	0	1	0	1
2, 3, 4, 13	0	0	0	0	1	0	1	0	1
2, 3, 4, 14, 18	0	0	0	0	1	0	0	1	-1
2, 3, 4, 15, 23, 24	0	0	0	0	1	0	1	0	1

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
2, 3, 4, 16	0	0	0	0	1	0	1	0	1
2, 3, 4, 23	0	0	0	0	1	0	1	0	1
2, 3, 4, 5	0	0	0	0	0	10	10	0	10
2, 3, 4, 5, 10, 15, 20	0	0	0	0	0	1	1	0	1
2, 3, 4, 5, 15	0	0	0	0	0	1	1	0	1
2, 3, 4, 5, 6, 10	0	0	0	0	0	1	1	0	1
2, 3, 4, 5, 6, 7	0	0	0	0	0	1	1	0	1
2, 3, 4, 5, 6, 7, 8	0	0	0	0	0	1	1	0	1
2, 3, 4, 5, 6, 9	0	0	0	0	0	1	1	0	1
2, 3, 4, 6, 7	0	0	0	0	2	0	2	0	2
2, 3, 4, 6, 7, 10	0	0	0	0	1	0	1	0	1
2, 3, 4, 6, 7, 8	0	0	0	0	1	0	1	0	1
2, 3, 4, 7	0	0	0	0	1	0	1	0	1
2, 3, 4, 7, 10	0	0	0	0	1	0	1	0	1
2, 3, 4, 7, 16, 20	0	0	0	0	1	0	1	0	1
2, 3, 4, 7, 8	0	0	0	0	1	0	1	0	1
2, 3, 4, 7, 9, 27	0	0	0	0	1	0	1	0	1
2, 3, 4, 8	0	0	0	0	3	0	3	0	3
2, 3, 4, 8, 9, 11	0	0	0	0	1	0	1	0	1
2, 3, 4, 9	0	0	0	0	1	0	1	0	1
2, 3, 4, 9, 11	0	0	0	0	2	0	2	0	2
2, 3, 5	0	0	0	0	20	0	15	5	10
2, 3, 5, 10	0	0	0	0	1	0	1	0	1
2, 3, 5, 14	0	0	0	0	1	0	1	0	1
2, 3, 5, 24	0	0	0	0	1	0	1	0	1
2, 3, 5, 43	0	0	0	0	1	0	0	1	-1
2, 3, 5, 6	0	0	0	0	4	0	4	0	4
2, 3, 5, 7	0	0	0	0	2	0	2	0	2
2, 3, 5, 7, 29	0	0	0	0	1	0	1	0	1
2, 3, 5, 7, 8	0	0	0	0	2	0	2	0	2
2, 3, 5, 8, 16, 19	0	0	0	0	1	0	1	0	1
2, 3, 5, 8, 9	0	0	0	0	2	0	2	0	2
2, 3, 5, 9	0	0	0	0	1	0	1	0	1
2, 3, 6	0	0	0	0	10	0	8	2	6
2, 3, 6, 7	0	0	0	0	2	0	2	0	2
2, 3, 6, 7, 15	0	0	0	0	1	0	1	0	1
2, 3, 6, 8, 9	0	0	0	0	1	0	1	0	1
2, 3, 6, 9	0	0	0	0	1	0	1	0	1
2, 3, 7	0	0	0	0	17	0	15	2	13
2, 3, 7, 12	0	0	0	0	1	0	1	0	1
2, 3, 7, 15, 32	0	0	0	0	1	0	0	1	-1
2, 3, 7, 19, 43	0	0	0	0	1	0	1	0	1
2, 3, 7, 8	0	0	0	0	2	0	1	1	0
2, 3, 7, 8, 10	0	0	0	0	1	0	1	0	1
2, 3, 7, 8, 27	0	0	0	0	1	0	1	0	1
2, 3, 7, 8, 9	0	0	0	0	1	0	1	0	1
2, 3, 7, 8, 9, 10	0	0	0	0	2	0	2	0	2
2, 3, 7, 9	0	0	0	0	1	0	1	0	1
2, 3, 7, 9, 12	0	0	0	0	1	0	1	0	1
2, 3, 7, 9, 15, 23	0	0	0	0	1	0	1	0	1
2, 3, 8	0	0	0	0	35	0	28	7	21
2, 3, 8, 12	0	0	0	0	2	0	2	0	2
2, 3, 8, 12, 16, 21	0	0	0	0	1	0	0	1	-1
2, 3, 8, 12, 19	0	0	0	0	2	0	2	0	2
2, 3, 8, 12, 20	0	0	0	0	1	0	1	0	1
2, 3, 8, 12, 21	0	0	0	0	1	0	1	0	1
2, 3, 8, 13	0	0	0	0	1	0	1	0	1
2, 3, 8, 14	0	0	0	0	3	0	3	0	3
2, 3, 8, 15	0	0	0	0	3	0	2	1	1
2, 3, 8, 16	0	0	0	0	1	0	1	0	1
2, 3, 8, 16, 18, 19	0	0	0	0	1	0	1	0	1
2, 3, 8, 16, 18, 21	0	0	0	0	1	0	1	0	1
2, 3, 8, 16, 21	0	0	0	0	1	0	1	0	1
2, 3, 8, 17	0	0	0	0	1	0	1	0	1
2, 3, 8, 17, 22	0	0	0	0	1	0	0	1	-1
2, 3, 8, 18	0	0	0	0	1	0	1	0	1
2, 3, 8, 21	0	0	0	0	1	0	1	0	1
2, 3, 8, 21, 41	0	0	0	0	1	0	0	1	-1
2, 3, 8, 23	0	0	0	0	1	0	1	0	1
2, 3, 8, 42	0	0	0	0	1	0	1	0	1
2, 3, 8, 9	0	0	0	0	4	0	4	0	4
2, 3, 9	0	0	0	0	8	0	6	2	4
2, 3, 9, 10	0	0	0	0	1	0	1	0	1
2, 3, 9, 14	0	0	0	0	1	0	1	0	1
2, 3, 9, 17	0	0	0	0	1	0	0	1	-1
2, 30	1	0	0	0	0	0	1	0	1
2, 32	2	0	0	0	0	0	2	0	2
2, 34	1	0	0	0	0	0	0	1	-1
2, 4	24	0	0	0	0	0	18	6	12
2, 4, 10	3	0	0	0	0	0	3	0	3

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
2, 4, 12	1	0	0	0	0	0	1	0	1
2, 4, 13	1	0	0	0	0	0	0	1	-1
2, 4, 14	1	0	0	0	0	0	1	0	1
2, 4, 15, 24	1	0	0	0	0	0	0	1	-1
2, 4, 20, 24, 25	1	0	0	0	0	0	1	0	1
2, 4, 24	1	0	0	0	0	0	1	0	1
2, 4, 27	1	0	0	0	0	0	1	0	1
2, 4, 5	23	0	0	0	0	0	19	4	15
2, 4, 5, 11	1	0	0	0	0	0	1	0	1
2, 4, 5, 11, 30	1	0	0	0	0	0	1	0	1
2, 4, 5, 6	4	0	0	0	0	0	3	1	2
2, 4, 5, 6, 30	1	0	0	0	0	0	1	0	1
2, 4, 5, 6, 7, 24	1	0	0	0	0	0	1	0	1
2, 4, 5, 7	2	0	0	0	0	0	2	0	2
2, 4, 5, 7, 10	1	0	0	0	0	0	1	0	1
2, 4, 5, 7, 11	1	0	0	0	0	0	1	0	1
2, 4, 5, 7, 11, 15, 20, 25	1	0	0	0	0	0	1	0	1
2, 4, 5, 8	1	0	0	0	0	0	1	0	1
2, 4, 6	5	0	0	0	0	0	5	0	5
2, 4, 6, 12, 15, 21	1	0	0	0	0	0	1	0	1
2, 4, 6, 7	1	0	0	0	0	0	1	0	1
2, 4, 6, 8	1	0	0	0	0	0	1	0	1
2, 4, 7	8	0	0	0	0	0	6	2	4
2, 4, 7, 12, 13, 21	1	0	0	0	0	0	1	0	1
2, 4, 7, 8, 9, 10	1	0	0	0	0	0	1	0	1
2, 4, 8	2	0	0	0	0	0	1	1	0
2, 4, 9	1	0	0	0	0	0	0	1	-1
2, 4, 1	2	0	0	0	0	0	1	1	0
2, 4, 3	1	0	0	0	0	0	1	0	1
2, 5	109	0	0	0	0	0	88	21	67
2, 5, 10	17	0	0	0	0	0	14	3	11
2, 5, 10, 15, 20, 25	1	0	0	0	0	0	1	0	1
2, 5, 13	1	0	0	0	0	0	1	0	1
2, 5, 14	1	0	0	0	0	0	1	0	1
2, 5, 15	5	0	0	0	0	0	4	1	3
2, 5, 15, 23	1	0	0	0	0	0	1	0	1
2, 5, 15, 28, 29	1	0	0	0	0	0	1	0	1
2, 5, 16, 18	1	0	0	0	0	0	0	1	-1
2, 5, 17	1	0	0	0	0	0	1	0	1
2, 5, 17, 28	1	0	0	0	0	0	1	0	1
2, 5, 31, 32	1	0	0	0	0	0	1	0	1
2, 5, 6	28	0	0	0	0	0	25	3	22
2, 5, 6, 10, 15	1	0	0	0	0	0	1	0	1
2, 5, 6, 13	1	0	0	0	0	0	1	0	1
2, 5, 6, 14, 17	1	0	0	0	0	0	1	0	1
2, 5, 6, 17	1	0	0	0	0	0	1	0	1
2, 5, 6, 7	17	0	0	0	0	0	13	4	9
2, 5, 6, 7, 10	2	0	0	0	0	0	2	0	2
2, 5, 6, 7, 8	3	0	0	0	0	0	2	1	1
2, 5, 6, 7, 9	3	0	0	0	0	0	2	1	1
2, 5, 6, 7, 9, 10	3	0	0	0	0	0	2	1	1
2, 5, 6, 8	5	0	0	0	0	0	4	1	3
2, 5, 6, 9	7	0	0	0	0	0	7	0	7
2, 5, 6, 9, 10	1	0	0	0	0	0	1	0	1
2, 5, 6, 9, 10, 13	1	0	0	0	0	0	0	1	-1
2, 5, 6, 9, 10, 15	1	0	0	0	0	0	1	0	1
2, 5, 6, 9, 11	1	0	0	0	0	0	1	0	1
2, 5, 7	74	0	0	0	0	0	64	10	54
2, 5, 7, 10	5	0	0	0	0	0	4	1	3
2, 5, 7, 10, 14	1	0	0	0	0	0	1	0	1
2, 5, 7, 10, 17	1	0	0	0	0	0	0	1	-1
2, 5, 7, 8	3	0	0	0	0	0	3	0	3
2, 5, 7, 8, 13	1	0	0	0	0	0	1	0	1
2, 5, 7, 8, 9	2	0	0	0	0	0	2	0	2
2, 5, 7, 9	5	0	0	0	0	0	5	0	5
2, 5, 7, 9, 13	3	0	0	0	0	0	2	1	1
2, 5, 8	19	0	0	0	0	0	18	1	17
2, 5, 9	8	0	0	0	0	0	3	5	-2
2, 5, 9, 10	6	0	0	0	0	0	4	2	2
2, 6	17	0	0	0	0	0	13	4	9
2, 6, 10	2	0	0	0	0	0	2	0	2
2, 6, 12	2	0	0	0	0	0	2	0	2
2, 6, 13	1	0	0	0	0	0	1	0	1
2, 6, 17	1	0	0	0	0	0	1	0	1
2, 6, 22	1	0	0	0	0	0	1	0	1
2, 6, 29	1	0	0	0	0	0	0	1	-1
2, 6, 7	10	0	0	0	0	0	8	2	6
2, 6, 7, 13	1	0	0	0	0	0	1	0	1
2, 6, 7, 26	1	0	0	0	0	0	1	0	1
2, 6, 7, 8	2	0	0	0	0	0	2	0	2

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
2, 6, 7, 9	3	0	0	0	0	0	3	0	3
2, 6, 7, 9, 10	1	0	0	0	0	0	1	0	1
2, 6, 7, 9, 15	1	0	0	0	0	0	1	0	1
2, 6, 8	1	0	0	0	0	0	0	1	-1
2, 6, 8, 10	1	0	0	0	0	0	0	1	-1
2, 6, 9	7	0	0	0	0	0	5	2	3
2, 7	27	0	0	0	0	0	19	8	11
2, 7, 10	2	0	0	0	0	0	1	1	0
2, 7, 12	4	0	0	0	0	0	4	0	4
2, 7, 12, 23, 29	1	0	0	0	0	0	1	0	1
2, 7, 13, 17	1	0	0	0	0	0	1	0	1
2, 7, 14	1	0	0	0	0	0	1	0	1
2, 7, 15	1	0	0	0	0	0	1	0	1
2, 7, 16	1	0	0	0	0	0	1	0	1
2, 7, 17, 22, 28	1	0	0	0	0	0	1	0	1
2, 7, 31	1	0	0	0	0	0	1	0	1
2, 7, 8	4	0	0	0	0	0	3	1	2
2, 7, 8, 9, 10	1	0	0	0	0	0	1	0	1
2, 7, 9	3	0	0	0	0	0	2	1	1
2, 7, 9, 10	1	0	0	0	0	0	0	1	-1
2, 7, 9, 13, 14, 32	1	0	0	0	0	0	0	1	-1
2, 8	9	0	0	0	0	0	9	0	9
2, 8, 10, 39	1	0	0	0	0	0	1	0	1
2, 8, 11, 14, 17	1	0	0	0	0	0	1	0	1
2, 8, 12	2	0	0	0	0	0	2	0	2
2, 8, 12, 14	2	0	0	0	0	0	2	0	2
2, 8, 12, 16, 21	1	0	0	0	0	0	1	0	1
2, 8, 12, 19	2	0	0	0	0	0	2	0	2
2, 8, 19, 21, 25	1	0	0	0	0	0	1	0	1
2, 8, 22	1	0	0	0	0	0	1	0	1
2, 8, 25	1	0	0	0	0	0	0	1	-1
2, 8, 26	1	0	0	0	0	0	1	0	1
2, 8, 9	2	0	0	0	0	0	1	1	0
2, 9	11	0	0	0	0	0	9	2	7
2, 9, 10	2	0	0	0	0	0	2	0	2
2, 9, 14, 15, 25	1	0	0	0	0	0	0	1	-1
2, 9, 15, 24	1	0	0	0	0	0	1	0	1
2, 9, 20	1	0	0	0	0	0	1	0	1
2, 9, 33, 35	1	0	0	0	0	0	1	0	1
20, 22, 26, 32	0	0	0	1	0	0	1	0	1
20, 23	0	0	0	1	0	0	0	1	-1
20, 23, 24	0	0	0	1	0	0	1	0	1
20, 23, 24, 25	0	0	0	1	0	0	1	0	1
20, 24	0	0	0	1	0	0	0	1	-1
20, 26, 35	0	0	0	1	0	0	0	1	-1
20, 27	0	0	0	1	0	0	1	0	1
21	0	0	0	1	0	0	1	0	1
21, 23, 31, 39	0	0	0	1	0	0	1	0	1
21, 26	0	0	0	1	0	0	1	0	1
21, 44	0	0	0	1	0	0	1	0	1
22	0	0	0	2	0	0	1	1	0
22, 24	0	0	0	1	0	0	1	0	1
22, 26	0	0	0	2	0	0	1	1	0
22, 29	0	0	0	1	0	0	0	1	-1
22, 34	0	0	0	1	0	0	1	0	1
22, 42	0	0	0	1	0	0	1	0	1
23, 31	0	0	0	1	0	0	1	0	1
24, 33	0	0	0	1	0	0	1	0	1
25	0	0	0	2	0	0	2	0	2
25, 26	0	0	0	1	0	0	0	1	-1
26	0	0	0	1	0	0	0	1	-1
27	0	0	0	2	0	0	2	0	2
27, 29	0	0	0	1	0	0	1	0	1
3	0	0	0	0	25	0	19	6	13
3, 10	0	0	0	0	6	0	5	1	4
3, 10, 11, 26	0	0	0	0	1	0	1	0	1
3, 10, 13	0	0	0	0	1	0	1	0	1
3, 11	0	0	0	0	15	0	14	1	13
3, 12	0	0	0	0	3	0	3	0	3
3, 12, 13	0	0	0	0	1	0	1	0	1
3, 12, 14	0	0	0	0	2	0	2	0	2
3, 12, 14, 18	0	0	0	0	1	0	1	0	1
3, 12, 14, 21	0	0	0	0	1	0	1	0	1
3, 12, 15, 18, 34	0	0	0	0	1	0	1	0	1
3, 12, 15, 25	0	0	0	0	1	0	1	0	1
3, 12, 19	0	0	0	0	1	0	1	0	1
3, 12, 19, 21	0	0	0	0	3	0	3	0	3
3, 12, 22	0	0	0	0	2	0	2	0	2
3, 12, 32, 43	0	0	0	0	1	0	1	0	1
3, 13	0	0	0	0	2	0	2	0	2

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
3, 13, 14	0	0	0	0	1	0	1	0	1
3, 13, 22, 25, 27	0	0	0	0	1	0	1	0	1
3, 14	0	0	0	0	7	0	6	1	5
3, 14, 16, 19, 31, 38	0	0	0	0	1	0	1	0	1
3, 14, 16, 38	0	0	0	0	1	0	1	0	1
3, 14, 18	0	0	0	0	2	0	2	0	2
3, 14, 18, 19	0	0	0	0	1	0	1	0	1
3, 14, 18, 19, 21	0	0	0	0	1	0	1	0	1
3, 14, 19, 22, 34	0	0	0	0	1	0	1	0	1
3, 14, 21	0	0	0	0	4	0	4	0	4
3, 14, 31	0	0	0	0	1	0	1	0	1
3, 15, 24	0	0	0	0	1	0	0	1	-1
3, 16	0	0	0	0	4	0	3	1	2
3, 17	0	0	0	0	1	0	1	0	1
3, 17, 18	0	0	0	0	2	0	1	1	0
3, 18	0	0	0	0	3	0	3	0	3
3, 18, 21, 30	0	0	0	0	1	0	1	0	1
3, 19	0	0	0	0	2	0	2	0	2
3, 19, 21	0	0	0	0	1	0	1	0	1
3, 21	0	0	0	0	1	0	1	0	1
3, 22	0	0	0	0	1	0	0	1	-1
3, 24	0	0	0	0	1	0	1	0	1
3, 24, 25	0	0	0	0	1	0	1	0	1
3, 32	0	0	0	0	2	0	2	0	2
3, 39	0	0	0	0	1	0	1	0	1
3, 4	0	33	0	0	0	0	31	2	29
3, 4, 11	0	4	0	0	0	0	4	0	4
3, 4, 13	0	2	0	0	0	0	2	0	2
3, 4, 16, 21	0	1	0	0	0	0	1	0	1
3, 4, 24, 27	0	1	0	0	0	0	1	0	1
3, 4, 26	0	1	0	0	0	0	1	0	1
3, 4, 5	0	7	0	0	0	0	5	2	3
3, 4, 5, 7	0	2	0	0	0	0	2	0	2
3, 4, 5, 8	0	1	0	0	0	0	1	0	1
3, 4, 6	0	6	0	0	0	0	6	0	6
3, 4, 6, 11	0	1	0	0	0	0	1	0	1
3, 4, 6, 12, 18	0	1	0	0	0	0	1	0	1
3, 4, 6, 7	0	1	0	0	0	0	1	0	1
3, 4, 6, 7, 10	0	1	0	0	0	0	1	0	1
3, 4, 6, 8	0	2	0	0	0	0	2	0	2
3, 4, 7	0	1	0	0	0	0	1	0	1
3, 4, 8	0	5	0	0	0	0	4	1	3
3, 4, 8, 10	0	2	0	0	0	0	1	1	0
3, 4, 8, 30	0	1	0	0	0	0	1	0	1
3, 4, 9	0	1	0	0	0	0	1	0	1
3, 4, 9, 13	0	1	0	0	0	0	1	0	1
3, 4, 9, 29	0	1	0	0	0	0	1	0	1
3, 40	0	0	0	0	2	0	2	0	2
3, 5	0	0	0	0	8	0	5	3	2
3, 5, 11	0	0	0	0	1	0	1	0	1
3, 5, 6	0	0	0	0	2	0	1	1	0
3, 5, 6, 10	0	0	0	0	2	0	2	0	2
3, 5, 7	0	0	0	0	3	0	3	0	3
3, 5, 7, 8	0	0	0	0	1	0	1	0	1
3, 5, 8	0	0	0	0	3	0	3	0	3
3, 5, 8, 10	0	0	0	0	1	0	0	1	-1
3, 5, 8, 11	0	0	0	0	1	0	1	0	1
3, 5, 8, 12	0	0	0	0	1	0	1	0	1
3, 5, 8, 13	0	0	0	0	1	0	1	0	1
3, 5, 9, 22, 29	0	0	0	0	1	0	1	0	1
3, 6	0	0	0	0	16	0	11	5	6
3, 6, 10	0	0	0	0	1	0	0	1	-1
3, 6, 11	0	0	0	0	5	0	2	3	-1
3, 6, 12, 14, 26, 30	0	0	0	0	1	0	1	0	1
3, 6, 13	0	0	0	0	1	0	1	0	1
3, 6, 15, 32	0	0	0	0	1	0	1	0	1
3, 6, 22	0	0	0	0	1	0	0	1	-1
3, 6, 43	0	0	0	0	1	0	1	0	1
3, 6, 7	0	0	0	0	2	0	2	0	2
3, 6, 7, 11	0	0	0	0	1	0	0	1	-1
3, 6, 8	0	0	0	0	6	0	4	2	2
3, 6, 8, 9	0	0	0	0	1	0	1	0	1
3, 6, 9	0	0	0	0	1	0	1	0	1
3, 6, 9, 10	0	0	0	0	1	0	1	0	1
3, 7	0	0	0	0	10	0	9	1	8
3, 7, 10, 13	0	0	0	0	1	0	1	0	1
3, 7, 12, 32	0	0	0	0	1	0	1	0	1
3, 7, 19	0	0	0	0	1	0	1	0	1
3, 7, 20	0	0	0	0	1	0	1	0	1
3, 7, 8	0	0	0	0	6	0	6	0	6

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
3, 7, 8, 10	0	0	0	0	1	0	1	0	1
3, 7, 9	0	0	0	0	5	0	3	2	1
3, 8	0	0	0	0	18	0	14	4	10
3, 8, 10	0	0	0	0	1	0	1	0	1
3, 8, 11	0	0	0	0	3	0	3	0	3
3, 8, 12, 14, 18, 21	0	0	0	0	1	0	1	0	1
3, 8, 12, 19, 21	0	0	0	0	1	0	0	1	-1
3, 8, 13	0	0	0	0	1	0	1	0	1
3, 8, 14	0	0	0	0	1	0	1	0	1
3, 8, 16	0	0	0	0	1	0	1	0	1
3, 8, 17	0	0	0	0	1	0	1	0	1
3, 8, 18	0	0	0	0	1	0	1	0	1
3, 8, 18, 19	0	0	0	0	1	0	1	0	1
3, 8, 18, 22	0	0	0	0	1	0	1	0	1
3, 8, 19	0	0	0	0	1	0	1	0	1
3, 8, 23	0	0	0	0	1	0	0	1	-1
3, 8, 37, 40	0	0	0	0	1	0	1	0	1
3, 8, 9	0	0	0	0	2	0	0	2	-2
3, 8, 9, 11	0	0	0	0	1	0	0	1	-1
3, 9	0	0	0	0	7	0	5	2	3
3, 9, 13, 29	0	0	0	0	1	0	1	0	1
3, 9, 17	0	0	0	0	1	0	0	1	-1
3, 9, 17, 22	0	0	0	0	1	0	0	1	-1
30	0	0	0	1	0	0	1	0	1
31	0	0	0	1	0	0	1	0	1
31, 35	0	0	0	1	0	0	0	1	-1
33	0	0	0	1	0	0	1	0	1
36	0	0	0	1	0	0	0	1	-1
4	0	16	0	0	0	0	9	7	2
4, 10	0	5	0	0	0	0	5	0	5
4, 10, 13, 19	0	1	0	0	0	0	0	1	-1
4, 10, 23	0	1	0	0	0	0	1	0	1
4, 11	0	5	0	0	0	0	3	2	1
4, 11, 13	0	1	0	0	0	0	0	1	-1
4, 11, 14	0	1	0	0	0	0	0	1	-1
4, 11, 26	0	1	0	0	0	0	1	0	1
4, 12	0	1	0	0	0	0	1	0	1
4, 12, 17	0	1	0	0	0	0	1	0	1
4, 12, 29, 35	0	1	0	0	0	0	0	1	-1
4, 12, 35	0	1	0	0	0	0	1	0	1
4, 13, 14	0	1	0	0	0	0	0	1	-1
4, 14	0	1	0	0	0	0	0	1	-1
4, 14, 16, 18, 26, 41	0	1	0	0	0	0	0	1	-1
4, 14, 23, 25	0	1	0	0	0	0	1	0	1
4, 16	0	1	0	0	0	0	1	0	1
4, 16, 20	0	1	0	0	0	0	0	1	-1
4, 21, 25	0	1	0	0	0	0	1	0	1
4, 22	0	1	0	0	0	0	1	0	1
4, 29	0	1	0	0	0	0	1	0	1
4, 5	0	18	0	0	0	0	15	3	12
4, 5, 10	0	2	0	0	0	0	1	1	0
4, 5, 12	0	1	0	0	0	0	1	0	1
4, 5, 22	0	1	0	0	0	0	0	1	-1
4, 5, 6	0	5	0	0	0	0	5	0	5
4, 5, 6, 10	0	1	0	0	0	0	1	0	1
4, 5, 6, 11, 13, 33	0	1	0	0	0	0	0	1	-1
4, 5, 6, 15	0	1	0	0	0	0	1	0	1
4, 5, 6, 7	0	0	0	2	0	0	2	0	2
4, 5, 6, 7, 10	0	0	0	1	0	0	1	0	1
4, 5, 7	0	2	0	0	0	0	2	0	2
4, 5, 8	0	1	0	0	0	0	1	0	1
4, 5, 9	0	1	0	0	0	0	0	1	-1
4, 5, 9, 10, 13	0	1	0	0	0	0	1	0	1
4, 6	0	14	0	0	0	0	11	3	8
4, 6, 10	0	2	0	0	0	0	2	0	2
4, 6, 11, 14	0	1	0	0	0	0	1	0	1
4, 6, 7	0	0	0	3	0	0	3	0	3
4, 6, 7, 10	0	0	0	4	0	0	3	1	2
4, 6, 8	0	1	0	0	0	0	1	0	1
4, 6, 9	0	2	0	0	0	0	2	0	2
4, 7	0	13	0	0	0	0	13	0	13
4, 7, 11	0	1	0	0	0	0	0	1	-1
4, 7, 14	0	1	0	0	0	0	1	0	1
4, 7, 15	0	1	0	0	0	0	0	1	-1
4, 7, 8	0	1	0	0	0	0	0	1	-1
4, 7, 8, 12, 19, 29	0	0	0	1	0	0	1	0	1
4, 7, 9	0	1	0	0	0	0	0	1	-1
4, 7, 9, 11	0	0	0	2	0	0	1	1	0
4, 8	0	1	0	0	0	0	1	0	1
4, 8, 10	0	1	0	0	0	0	1	0	1

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
4, 8, 11	0	3	0	0	0	0	2	1	1
4, 8, 17	0	1	0	0	0	0	0	1	-1
4, 9	0	1	0	0	0	0	1	0	1
4, 9, 11, 13	0	1	0	0	0	0	1	0	1
4, 9, 22	0	1	0	0	0	0	0	1	-1
4, 9, 27	0	1	0	0	0	0	1	0	1
40	0	0	0	1	0	0	1	0	1
41, 45	0	0	0	1	0	0	1	0	1
42	0	0	0	1	0	0	0	1	-1
45	0	0	0	1	0	0	1	0	1
5	0	0	0	13	0	0	8	5	3
5, 10	0	0	0	5	0	0	4	1	3
5, 10, 11	0	0	0	1	0	0	1	0	1
5, 10, 12	0	0	0	2	0	0	2	0	2
5, 10, 13	0	0	0	1	0	0	1	0	1
5, 10, 16	0	0	0	1	0	0	1	0	1
5, 10, 18	0	0	0	1	0	0	0	1	-1
5, 10, 19, 20	0	0	0	1	0	0	1	0	1
5, 10, 22	0	0	0	1	0	0	0	1	-1
5, 10, 24	0	0	0	1	0	0	1	0	1
5, 10, 30	0	0	0	1	0	0	0	1	-1
5, 11	0	0	0	2	0	0	2	0	2
5, 11, 25	0	0	0	1	0	0	1	0	1
5, 12, 15	0	0	0	2	0	0	2	0	2
5, 13	0	0	0	2	0	0	0	2	-2
5, 13, 17, 26	0	0	0	1	0	0	1	0	1
5, 14	0	0	0	1	0	0	0	1	-1
5, 14, 22, 28	0	0	0	1	0	0	1	0	1
5, 15	0	0	0	1	0	0	0	1	-1
5, 15, 19	0	0	0	1	0	0	1	0	1
5, 16	0	0	0	2	0	0	1	1	0
5, 16, 28, 40	0	0	0	1	0	0	0	1	-1
5, 17, 35	0	0	0	1	0	0	0	1	-1
5, 19	0	0	0	1	0	0	1	0	1
5, 20	0	0	0	1	0	0	1	0	1
5, 25, 31	0	0	0	1	0	0	0	1	-1
5, 6	0	0	0	11	0	0	10	1	9
5, 6, 10	0	0	0	4	0	0	3	1	2
5, 6, 10, 11	0	0	0	2	0	0	1	1	0
5, 6, 10, 19	0	0	0	1	0	0	1	0	1
5, 6, 12, 29	0	0	0	1	0	0	0	1	-1
5, 6, 14	0	0	0	1	0	0	0	1	-1
5, 6, 17	0	0	0	1	0	0	1	0	1
5, 6, 7	0	0	0	5	0	0	5	0	5
5, 6, 7, 10	0	0	0	2	0	0	2	0	2
5, 6, 7, 10, 15, 24	0	0	0	1	0	0	1	0	1
5, 6, 7, 15	0	0	0	1	0	0	1	0	1
5, 6, 7, 23	0	0	0	1	0	0	1	0	1
5, 6, 7, 8	0	0	0	1	0	0	1	0	1
5, 6, 7, 8, 10	0	0	0	2	0	0	2	0	2
5, 6, 7, 9, 10	0	0	0	3	0	0	3	0	3
5, 6, 9	0	0	0	5	0	0	4	1	3
5, 6, 9, 10	0	0	0	1	0	0	1	0	1
5, 6, 9, 11, 12	0	0	0	1	0	0	0	1	-1
5, 7	0	0	0	15	0	0	14	1	13
5, 7, 10	0	0	0	2	0	0	2	0	2
5, 7, 11	0	0	0	2	0	0	2	0	2
5, 7, 14	0	0	0	1	0	0	1	0	1
5, 7, 18, 29	0	0	0	1	0	0	1	0	1
5, 7, 32	0	0	0	2	0	0	2	0	2
5, 7, 8, 10	0	0	0	2	0	0	2	0	2
5, 7, 8, 9	0	0	0	3	0	0	2	1	1
5, 8	0	0	0	2	0	0	2	0	2
5, 8, 10	0	0	0	1	0	0	1	0	1
5, 8, 9	0	0	0	2	0	0	1	1	0
5, 9	0	0	0	2	0	0	1	1	0
5, 9, 27	0	0	0	1	0	0	0	1	-1
6	0	0	0	7	0	0	6	1	5
6, 10	0	0	0	4	0	0	2	2	0
6, 10, 11, 15	0	0	0	1	0	0	1	0	1
6, 10, 13	0	0	0	1	0	0	1	0	1
6, 11	0	0	0	7	0	0	4	3	1
6, 11, 13	0	0	0	1	0	0	0	1	-1
6, 12	0	0	0	1	0	0	1	0	1
6, 12, 14, 18, 19	0	0	0	1	0	0	1	0	1
6, 12, 20	0	0	0	1	0	0	0	1	-1
6, 13	0	0	0	3	0	0	1	2	-1
6, 14	0	0	0	3	0	0	2	1	1
6, 15	0	0	0	3	0	0	2	1	1
6, 15, 17, 29	0	0	0	1	0	0	1	0	1

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
6, 17	0	0	0	1	0	0	1	0	1
6, 17, 18	0	0	0	1	0	0	1	0	1
6, 18, 19, 21	0	0	0	2	0	0	2	0	2
6, 31	0	0	0	1	0	0	1	0	1
6, 7	0	0	0	16	0	0	13	3	10
6, 7, 10	0	0	0	3	0	0	3	0	3
6, 7, 11, 15	0	0	0	1	0	0	1	0	1
6, 7, 13	0	0	0	1	0	0	1	0	1
6, 7, 14	0	0	0	1	0	0	1	0	1
6, 7, 15, 28	0	0	0	1	0	0	1	0	1
6, 7, 17, 18	0	0	0	1	0	0	1	0	1
6, 7, 8	0	0	0	1	0	0	0	1	-1
6, 7, 8, 17	0	0	0	1	0	0	1	0	1
6, 7, 9	0	0	0	3	0	0	2	1	1
6, 7, 9, 10	0	0	0	2	0	0	2	0	2
6, 7, 9, 10, 25	0	0	0	1	0	0	1	0	1
6, 7, 9, 10, 25, 29	0	0	0	1	0	0	1	0	1
6, 8	0	0	0	9	0	0	8	1	7
6, 8, 11	0	0	0	2	0	0	2	0	2
6, 8, 11, 14	0	0	0	1	0	0	1	0	1
6, 8, 13	0	0	0	2	0	0	2	0	2
6, 8, 13, 20, 28, 33, 44	0	0	0	1	0	0	1	0	1
6, 8, 14, 28, 31, 32	0	0	0	1	0	0	1	0	1
6, 8, 9	0	0	0	2	0	0	2	0	2
6, 9	0	0	0	4	0	0	3	1	2
6, 9, 10	0	0	0	2	0	0	2	0	2
6, 9, 13	0	0	0	2	0	0	0	2	-2
6, 9, 14	0	0	0	1	0	0	0	1	-1
6, 9, 17	0	0	0	1	0	0	1	0	1
7	0	0	0	3	0	0	3	0	3
7, 10	0	0	0	9	0	0	9	0	9
7, 10, 14	0	0	0	1	0	0	1	0	1
7, 11	0	0	0	2	0	0	2	0	2
7, 11, 19, 20	0	0	0	1	0	0	0	1	-1
7, 11, 23	0	0	0	1	0	0	1	0	1
7, 12	0	0	0	2	0	0	2	0	2
7, 12, 15, 26	0	0	0	1	0	0	1	0	1
7, 12, 20	0	0	0	1	0	0	1	0	1
7, 13	0	0	0	4	0	0	2	2	0
7, 13, 14	0	0	0	2	0	0	2	0	2
7, 14, 15, 17, 27	0	0	0	1	0	0	1	0	1
7, 14, 31, 44	0	0	0	1	0	0	1	0	1
7, 15, 17, 25	0	0	0	1	0	0	1	0	1
7, 15, 19, 20, 27, 28	0	0	0	1	0	0	1	0	1
7, 16	0	0	0	1	0	0	1	0	1
7, 17	0	0	0	3	0	0	3	0	3
7, 17, 20, 36, 42	0	0	0	1	0	0	1	0	1
7, 18	0	0	0	1	0	0	1	0	1
7, 19	0	0	0	1	0	0	1	0	1
7, 19, 27, 28, 29	0	0	0	1	0	0	1	0	1
7, 22	0	0	0	1	0	0	1	0	1
7, 24	0	0	0	1	0	0	1	0	1
7, 30	0	0	0	2	0	0	2	0	2
7, 31, 40	0	0	0	1	0	0	1	0	1
7, 8	0	0	0	13	0	0	10	3	7
7, 8, 10	0	0	0	3	0	0	3	0	3
7, 8, 11	0	0	0	1	0	0	0	1	-1
7, 8, 12	0	0	0	1	0	0	0	1	-1
7, 8, 12, 19, 20	0	0	0	1	0	0	1	0	1
7, 8, 13, 38	0	0	0	1	0	0	1	0	1
7, 8, 16, 26	0	0	0	1	0	0	1	0	1
7, 8, 9	0	0	0	1	0	0	1	0	1
7, 9	0	0	0	6	0	0	4	2	2
7, 9, 10	0	0	0	1	0	0	1	0	1
7, 9, 11, 14	0	0	0	1	0	0	1	0	1
7, 9, 13	0	0	0	1	0	0	1	0	1
7, 9, 17	0	0	0	2	0	0	1	1	0
7, 9, 18	0	0	0	2	0	0	1	1	0
8	0	0	0	2	0	0	1	1	0
8, 10	0	0	0	1	0	0	1	0	1
8, 10, 15	0	0	0	1	0	0	1	0	1
8, 10, 27	0	0	0	1	0	0	1	0	1
8, 11	0	0	0	2	0	0	2	0	2
8, 11, 13, 22	0	0	0	1	0	0	1	0	1
8, 11, 14	0	0	0	1	0	0	1	0	1
8, 11, 14, 32, 33	0	0	0	1	0	0	1	0	1
8, 12, 14	0	0	0	2	0	0	2	0	2
8, 12, 14, 19	0	0	0	1	0	0	1	0	1
8, 12, 14, 19, 21	0	0	0	1	0	0	1	0	1
8, 12, 16	0	0	0	1	0	0	1	0	1

Agrupaciones	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Exito	Fracaso	E-F
8, 12, 17, 22, 28	0	0	0	1	0	0	0	1	-1
8, 12, 18, 21	0	0	0	1	0	0	1	0	1
8, 12, 19, 22, 43	0	0	0	1	0	0	1	0	1
8, 13, 17	0	0	0	1	0	0	1	0	1
8, 14	0	0	0	4	0	0	3	1	2
8, 14, 18	0	0	0	1	0	0	1	0	1
8, 14, 18, 21, 41	0	0	0	1	0	0	1	0	1
8, 14, 22, 25	0	0	0	1	0	0	1	0	1
8, 15	0	0	0	2	0	0	1	1	0
8, 16	0	0	0	1	0	0	0	1	-1
8, 16, 18	0	0	0	1	0	0	1	0	1
8, 16, 20, 31, 35, 44	0	0	0	1	0	0	0	1	-1
8, 16, 22	0	0	0	1	0	0	1	0	1
8, 16, 36	0	0	0	1	0	0	1	0	1
8, 16, 42	0	0	0	1	0	0	1	0	1
8, 17	0	0	0	3	0	0	1	2	-1
8, 18	0	0	0	3	0	0	2	1	1
8, 19	0	0	0	1	0	0	1	0	1
8, 22	0	0	0	2	0	0	0	2	-2
8, 23	0	0	0	1	0	0	1	0	1
8, 23, 26, 35	0	0	0	1	0	0	0	1	-1
8, 26	0	0	0	1	0	0	0	1	-1
8, 29	0	0	0	1	0	0	0	1	-1
8, 34, 36	0	0	0	1	0	0	0	1	-1
8, 38	0	0	0	1	0	0	1	0	1
8, 9	0	0	0	1	0	0	0	1	-1
8, 9, 10	0	0	0	2	0	0	1	1	0
8, 9, 10, 15	0	0	0	1	0	0	1	0	1
8, 9, 11	0	0	0	1	0	0	0	1	-1
8, 9, 11, 13	0	0	0	1	0	0	0	1	-1
8, 9, 13	0	0	0	1	0	0	1	0	1
8, 9, 13, 39	0	0	0	1	0	0	0	1	-1
8, 9, 15, 16, 25	0	0	0	1	0	0	1	0	1
8, 9, 16, 31	0	0	0	1	0	0	0	1	-1
8, 9, 26	0	0	0	1	0	0	0	1	-1
9	0	0	0	2	0	0	0	2	-2
9, 10	0	0	0	1	0	0	1	0	1
9, 11	0	0	0	2	0	0	1	1	0
9, 11, 17, 26, 36	0	0	0	1	0	0	1	0	1
9, 11, 21, 23, 26, 36	0	0	0	1	0	0	1	0	1
9, 11, 45	0	0	0	1	0	0	1	0	1
9, 12	0	0	0	1	0	0	0	1	-1
9, 12, 13, 18, 26	0	0	0	1	0	0	0	1	-1
9, 12, 17	0	0	0	1	0	0	1	0	1
9, 13	0	0	0	1	0	0	0	1	-1
9, 13, 15, 23, 24, 30	0	0	0	1	0	0	1	0	1
9, 13, 22	0	0	0	1	0	0	1	0	1
9, 13, 29	0	0	0	1	0	0	0	1	-1
9, 15	0	0	0	1	0	0	1	0	1
9, 15, 29	0	0	0	1	0	0	1	0	1
9, 16, 17, 30	0	0	0	1	0	0	1	0	1
9, 16, 23, 25	0	0	0	1	0	0	1	0	1
9, 17, 18, 20, 25	0	0	0	1	0	0	0	1	-1
9, 18	0	0	0	1	0	0	1	0	1
9, 21	0	0	0	1	0	0	1	0	1
9, 22, 25, 26	0	0	0	1	0	0	1	0	1
9, 26	0	0	0	1	0	0	0	1	-1
9, 27	0	0	0	1	0	0	1	0	1
9, 28	0	0	0	1	0	0	1	0	1

## 10. Futuras ampliaciones

### 10.1. Anexo 8. Tabla de alumnos con sus matriculaciones y sus respectivas asignaturas aprobadas.

Como resultado de diferentes investigaciones se ha conseguido tabular para cada alumno, sus propiedades de clúster, asignaturas matriculadas, aprobadas y su ratio de éxito o fracaso, de la siguiente manera:

	A	B	C	D	E	F	G	H
1		Cluster	Matriculadas	Aprobadas	RatioEF	EF	Asignaturas_Fa	Asignaturas_Sup
2	1	3	3	3	1,00	Exito	1, 2, 3	1, 2, 3
3	2	3	3	3	1,00	Exito	1, 2, 3	1, 2, 3
4	3	3	3	0	0,00	Fracaso	1, 3, 5	
5	4	2	4	2	0,50	Exito	4, 9, 11, 13	4, 11
6	5	1	2	0	0,00	Fracaso	2, 12	
7	6	3	3	3	1,00	Exito	1, 3, 5	1, 3, 5
8	7	2	2	2	1,00	Exito	3, 4	3, 4
9	8	3	3	0	0,00	Fracaso	1, 2, 3	
10	9	5	2	1	0,50	Exito	3, 11	3
11	10	4	1	1	1,00	Exito	11	11
12	11	3	3	3	1,00	Exito	1, 2, 3	1, 2, 3
13	12	2	2	2	1,00	Exito	3, 4	3, 4
14	13	5	2	2	1,00	Exito	2, 3	2, 3
15	14	2	3	3	1,00	Exito	3, 4, 11	3, 4, 11
16	15	3	3	0	0,00	Fracaso	1, 2, 3	
17	16	4	1	0	0,00	Fracaso	5	
18	17	1	1	0	0,00	Fracaso	2	
19	18	6	4	2	0,50	Exito	1, 2, 3, 4	2, 4
20	19	6	4	4	1,00	Exito	1, 2, 3, 4	1, 2, 3, 4
21	20	1	4	4	1,00	Exito	2, 8, 12, 19	2, 8, 12, 19
22	21	2	2	0	0,00	Fracaso	1, 4	
23	22	2	3	0	0,00	Fracaso	1, 4, 5	

**Ilustración 23. Subproducto tabla cruzada entre paquetes de asignaturas matriculadas y superadas**

El fichero correspondiente se adjunta a la memoria de trabajo:



alumnos\_clusters\_exi  
to.xls

### 10.2. Anexo 9. Tabla cruzada de referencias entre agrupaciones de matrículas y agrupaciones de asignaturas aprobadas

También se ha obtenido como subproducto de la investigación una tabla de referencias cruzadas sobre los paquetes de asignaturas matriculadas y las frecuencias con las que estas se han aprobado en su totalidad, de manera parcial o se ha fracasado en el paquete. Como futura aplicación es posible que se obtenga cuál de las asignaturas del paquete matriculado es la que vuelca la balanza en contra o a favor de las demás, la representación es de la siguiente manera:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Matriculadas	Fracaso	1	1, 10, 13, 25	1, 10, 20	1, 11	1, 11, 22	1, 12, 18, 21, 28	1, 13	1, 14	1, 15, 30	1, 16	1, 17, 28
2	1	6	16	0	0	0	0	0	0	0	0	0	0
3	1, 10	2	0	0	0	0	0	0	0	0	0	0	0
4	1, 10, 20	0	0	0	1	0	0	0	0	0	0	0	0
5	1, 11	8	0	0	0	11	0	0	0	0	0	0	0
6	1, 11, 13	1	0	0	0	0	0	0	0	0	0	0	0
7	1, 11, 20	1	0	0	0	0	0	0	0	0	0	0	0
8	1, 11, 22	0	0	0	0	0	1	0	0	0	0	0	0
9	1, 11, 27	1	0	0	0	0	0	0	0	0	0	0	0
10	1, 12, 14	0	0	0	0	0	0	0	0	1	0	0	0
11	1, 12, 18, 21, 28	0	0	0	0	0	0	1	0	0	0	0	0
12	1, 13	1	1	0	0	0	0	0	1	0	0	0	0
13	1, 14	0	1	0	0	0	0	0	0	1	0	0	0
14	1, 15	1	0	0	0	0	0	0	0	0	0	0	0
15	1, 15, 30	0	0	0	0	0	0	0	0	0	1	0	0
16	1, 16	0	0	0	0	0	0	0	0	0	0	1	0
17	1, 17, 26, 28	0	0	0	0	0	0	0	0	0	0	0	1
18	1, 2	31	0	0	0	0	0	0	0	0	0	0	0
19	1, 2, 10	0	0	0	0	0	0	0	0	0	0	0	0
20	1, 2, 10, 13, 20, 25	0	0	1	0	0	0	0	0	0	0	0	0
21	1, 2, 11	1	0	0	0	0	0	0	0	0	0	0	0
22	1, 2, 12	0	0	0	0	0	0	0	0	0	0	0	0
23	1, 2, 13	0	0	0	0	0	0	0	0	0	0	0	0
24	1, 2, 14	0	0	0	0	0	0	0	0	0	0	0	0
25	1, 2, 20	0	0	0	0	0	0	0	0	0	0	0	0
26	1, 2, 3	33	3	0	0	0	0	0	0	0	0	0	0
27	1, 2, 3, 10	0	0	0	0	0	0	0	0	0	0	0	0
28	1, 2, 3, 11	0	0	0	0	0	0	0	0	0	0	0	0
29	1, 2, 3, 12, 19, 20	0	0	0	0	0	0	0	0	0	0	0	0
30	1, 2, 3, 13	0	0	0	0	0	0	0	0	0	0	0	0
31	1, 2, 3, 4	11	0	0	0	0	0	0	0	0	0	0	0
32	1, 2, 3, 4, 10	0	0	0	0	0	0	0	0	0	0	0	0

**Ilustración 24. Subproducto tabla cruzada asignaturas matriculadas y asignaturas superadas**

Se adjunta en la memoria el fichero correspondiente a la tabla:



alumnos\_cluster\_exit  
o\_asignaturas\_fa\_vs\_

### 10.3. Anexo 10. Tabla cruzada de paquetes de asignaturas matriculadas y asignaturas aprobadas en listado

También se ha obtenido como subproducto una tabla que muestra de forma tabulada la relación anterior entre paquetes de asignaturas matriculadas y asignaturas aprobadas, se entiende que cuando en la columna de "Paquete Aprobado" no hay es porque se trata de un Fracaso, es decir, no se ha aprobado ninguna. La siguiente imagen representa una muestra:

	A	B	C	D
1		Paquete Matriculado	Paquete Aprobado	Frecuencia
2	1	1		6
3	2	1	1	16
5	4	1, 10		2
6	5	1, 10, 20	1, 10, 20	1
7	6	1, 11		8
9	8	1, 11	1, 11	11
12	11	1, 11	11	3
14	13	1, 11, 13		1
18	17	1, 11, 20		1
21	20	1, 11, 22	1, 11, 22	1
26	25	1, 11, 27		1
31	30	1, 12, 14	1, 14	1
37	36	1, 12, 18, 21, 28	1, 12, 18, 21, 28	1
39	38	1, 13		1
40	39	1, 13	1	1
44	43	1, 13	1, 13	1
47	46	1, 14	1	1
48	47	1, 14	1, 14	1
50	49	1, 15		1
53	52	1, 15	15	1
54	53	1, 15, 30	1, 15, 30	1

Se adjunta a la memoria el fichero que contiene dichos datos:



alumnos\_cluster\_exit  
o\_asignaturas\_fa\_vs\_

## 11. Bibliografía y Referencias

### 11.1. Referencias Escritas

CRISTÓBAL ROMERO y SEBASTIÁN VENTURA. *EDM&LA: State of the Art*. [on-line] [http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDMQFjAA&url=http%3A%2F%2Fwww.researchgate.net%2Fpublication%2F224160756\\_Educational\\_Data\\_Mining\\_A\\_Review\\_of\\_the\\_State\\_of\\_the\\_Art%2Ffile%2F79e41510a07a5b28fa.pdf&ei=hYccU4beNe6v7AauICAAG&usg=AFQjCNGqRRQFp0tMC-R5poQ0fI2S-EKA&bvm=bv.62578216,d.ZGU&cad=rja](http://www.google.es/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0CDMQFjAA&url=http%3A%2F%2Fwww.researchgate.net%2Fpublication%2F224160756_Educational_Data_Mining_A_Review_of_the_State_of_the_Art%2Ffile%2F79e41510a07a5b28fa.pdf&ei=hYccU4beNe6v7AauICAAG&usg=AFQjCNGqRRQFp0tMC-R5poQ0fI2S-EKA&bvm=bv.62578216,d.ZGU&cad=rja) [fecha de consulta 8 de Marzo de 2014]

R DEVELOPMENT CORE TEAM (2000). *Introducción a R. Notas sobre R: Un entorno de programación para Análisis de Datos y Gráficos*.

R CORE TEAM. *R Language Definition. Versión 3.0.3 DRAFT*

RENAUD GAUJOUX ET AL. *A flexible R package for nonnegative matrix factorization*. In: BMC. Bioinformatics 11.1 (2010), p. 367. issn: 1471-2105. doi: 10.1186/1471-2105-11-367

JOHANNES LEDOLTER. (2013). *Data Mining And Business Analytics With R*. New Jersey: John Wiley & Sons, Inc.

YANCHANG ZHAO. (2013). *R and Data Mining: Examples and Case Studies*. Published by Elsevier.

ROBERT I. KABACOFF (2011). *R in Action Data analysis and graphics with R*. Manning Publications Co.

LUIS CARLOS MOLINA FÉLIX, RAMON SANGUESA I SOLÉ (2010). *Data Mining*. Material docente de la UOC.

### 11.2. Referencias Online

Web Oficial de R: <http://www.rdatamining.com/>

EDM: <http://www.educationdatamining.org/>

R-project: <http://www.r-project.org/>

CRAN Project: <http://cran.r-project.org/>

R Graphical Manual: [http://rgm3.lab.nig.ac.jp/RGM/R\\_image\\_list?page=1249&init=true](http://rgm3.lab.nig.ac.jp/RGM/R_image_list?page=1249&init=true)

Quick-R Accessing the power of R: <http://www.statmethods.net/index.html>