



Màster en Enginyeria Computacional i Matemàtica

Treball Final de Màster

## ANÀLISIS PREDICTIU DELS PACIENTS MENORS D'EDAT DE L'HOSPITAL UNIVERSITARI PERE MATA, ESPECIALITZAT EN PSIQUIATRIA AMB TÈCNiques D'INTEL·LIGÈNCIA ARTIFICIAL

**MIQUEL QUEROL CAPELL**

Màster Universitari, Enginyeria Computacional i Matemàtica.  
Anàlisi predictius

**David Riaño Ramos**

**Juan Alberto Rodríguez Velázquez**

Juny del 2017





Esta obra está sujeta a una licencia de Reconocimiento-NoComercial-SinObraDerivada [3.0 España de Creative Commons](https://creativecommons.org/licenses/by-nc-nd/3.0/es/)

## **B) GNU Free Documentation License (GNU FDL)**

Copyright © AÑO TU-NOMBRE.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section entitled "GNU Free Documentation License".

## **C) Copyright**

© (el autor/a)

Reservados todos los derechos. Está prohibido la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la impresión, la reprografía, el microfilme, el tratamiento informático o cualquier otro sistema, así como la distribución de ejemplares mediante alquiler y préstamo, sin la autorización escrita del autor o de los límites que autorice la Ley de Propiedad Intelectual.

## FITXA DEL TREBALL FINAL

<b>Títol del treball:</b>	<i>Anàlisi predictiu dels pacients menors d'edat de l'Hospital Universitari Pere Mata Especialitzat en Psiquiatria amb tècniques d'intel·ligència artificial</i>
<b>Nom de l'autor:</b>	<i>Miquel Querol Capell</i>
<b>Nom del consultor/a:</b>	<i>David Riaño Ramos</i>
<b>Nom del PRA:</b>	<i>Juan Alberto Rodríguez Velázquez</i>
<b>Data d'entrega (mm/aaaa):</b>	06/2017
<b>Titulació:</b>	<i>Màster Universitari, Enginyeria Computacional i Matemàtica</i>
<b>Àrea del Treball Final:</b>	<i>Treball de fi de màster(anàlisi predictius)</i>
<b>Idioma del treball:</b>	<i>Català</i>
<b>Paraules clau:</b>	<i>Predictiu, psiquiatria, menors</i>
<b>Resum del Treball (màxim 250 paraules):.</b>	
<p>S'apliquen anàlisis predictius en pacients menors d'edat de l'Hospital Psiquiàtric Universitari Institut Pere Mata, amb l'objectiu d'esbrinar els motius pels quals els pacients deriven a un altre diagnòstic o mantenen el mateix diagnòstic al cap de 6 anys. També s'analitzen els motius que porten els pacients a una millora o un empitjorament segons la valoració GAF (escala d'avaluació de funcionament global).</p> <p>Es genera una agrupació i classificació de totes les patologies, juntament amb una taula de canvi de diagnòstic, en què s'elegeixen els trastorns d'adaptació, els trastorns d'ansietat i els trastorn de dèficit d'atenció amb hiperactivitat com patologia per a aplicar els càlculs predictius. El motiu és que aquestes són les que agrupen més pacients i més canvis de diagnòstics.</p> <p>A continuació s'elegeix l'arbre de decisions com a model predictiu útil, al qual se l'hi apliquen sis bases de dades diferents: Csmij, psicopatologia, valoració social, anamnesis, medicaments i hospitalització/urgències. Aquestes individualment i conjuntament i d'aquestes dades s'obté setze arbres predictius diferents.</p> <p>De l'anàlisi d'aquets arbres predictius s'observa que les variables de psicopatologia, com l'insomni d'inici, fa que els pacients deriven cap a un trastorn d'ansietat, o que variables com hetero agressivitat i disfòria o explosivitat i inquietud, combinades influeixen en que el pacient acabi amb un trastorn de conducta.</p> <p>El fet de donar positiu en algunes variables de valoració social com són: tenir relacions socials inadequades, alteració dels hàbits, consum de tòxics o</p>	

dificultats en la integració amb els iguals augmenta el risc de derivar cap a un trastorn de la personalitat. En canvi, els pacients que mantenen relacions socials inadequades i, a més a més no tenen aficions, són més pròxims a acabar amb un trastorn de conducta.

També els arbres decisions ens indiquen que tenir dificultats en l'adquisició de l'aprenentatge o dificultats amb l'adaptació amb el professor són símptomes que indiquen que el pacient pot derivar cap a un trastorn de l'aprenentatge.

**Abstract (in English, 250 words or less):**

Applying predictive analysis to patients under the age of 18 at University Institute Pere Mata (UIPM), a psychiatric hospital, with the aim of finding out the reasons why they maintain or develop another pathology over a 6 year period from the original diagnosis. We also analyse the reasons why patients improve or deteriorate as measured by the Global Assessment of Function (GAF).

First of all, pathologies are grouped in seventeen groups according to the similarities of the mental illness. Then a pathologies change table is created to see which patients shows signs of having changed pathology. Adaptation disorder, anxiety disorder and attention deficit with hyperactivity disorder have been chosen to be applied to the predictive calculation, as these are the pathologies that more patients have shown the most changes.

The decision tree is chosen as a useful predictive method for analysing the reasons why the pathologies change in the patients, which is applied to the five different databases; Csmij, psychopathology, social evaluation, anamnesis and medicines, both separately and all together, obtaining sixteen different trees.

We note that the variables of psychopathology such as early insomnia has an influence on the change of the pathology towards anxiety disorder, or combined psychopathology variables like hetero aggressiveness and dysphoria or explosiveness and anxiety have influence that patients final diagnosis is a conduct disorder.

Matching social variables such as inadequate social relationships, changing habits, consumption of toxics or difficulties in integrating with peers, increase the risk to developing a personal disorder. Instead the patient with on increased change a conduct disorder will have inadequate social relationships and lack any interest of hobbies are closest to end with a conduct disorder.

The decision trees also tells us that the patients who have difficulty in acquiring learning or teacher integration difficulties are symptomatic indications of a learning disorders.

# Índex

1. Introducció.....	1
1.1 Context i justificació del treball .....	1
1.2 Objectius del Treball.....	2
1.3 Enfoc i mètode seguit.....	3
1.4 Planificació del Treball i Recursos Emprats .....	7
1.5 Breu sumari de productes obtinguts.....	8
1.6 Breu descripció dels altres capítols de la memòria .....	8
2. Metodologies .....	9
2.1 Refinament de dades i tractament de valors atípics.....	9
2.2 Agrupació de dades i correlació de malalties .....	10
2.3 Selecció d'atributs .....	11
2.4 Models predictius .....	15
3. Implementació de les metodologies escollides.....	17
3.1 Base de dades .....	17
3.2 Refinat de dades (valors atípics) .....	25
3.3 Agrupació de diagnòstics i correlació de malalties .....	26
3.4 Selecció d'atributs .....	35
3.5 Construcció de models predictius.....	44
3.6 Valoració del model.....	45
3.7 Resultats .....	46
3.8 Comentaris dels arbres pel departament Csmij de psiquiatria. ....	64
4. Conclusions.....	65
5. Glossari .....	67
6. Bibliografia.....	69

## Llista de figures

No es troben elements de tabla de il·lustracions.



# 1. Introducció

## 1.1 Context i justificació del treball

L'Hospital Psiquiàtric Universitari Institut Pere Mata (Reus, Tarragona) ofereix atenció a la comunitat en els àmbits de salut mental, discapacitat, geriatria i serveis socio-sanitaris tant per a adults com a infants i joves de menys de 18 anys.

Entre els estudis possibles, els professionals de l'hospital tenen interès en saber si alguns factors com el social, la psicopatologia o l'anamnesi, són causants de què els pacients menors d'edat empitjorin la patologia diagnosticada o de què al cap del temps se l'hi diagnostiqui una patologia diferent.

L'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM), especialitzat amb el tractament psiquiàtric, porta els últims anys fent una recopilació d'informació de caràcter sanitari (Csmij), valoració social, *anamnesis*, *psicopatologia*, medicaments i urgències/ hospitalitzacions dels pacients atesos.

Amb aquesta informació s'ha generat una base de *dades massives* que conté informació mèdica, psiquiàtrica, psicològica i social del pacients atesos.

Amb l'acumulació d'aquesta gran quantitat de dades, l'anàlisi predictiu podria estimular la transformació de la *medicina reactiva* actual cap a la *medicina predictiva*, preventiva i personalitzada. No obstant això, l'alta dimensionalitat i l'alta complexitat de les dades involucrades dificulta la traducció d'aquestes dades en conclusions clínicament rellevants.

Aplicar mètodes per poder estimar quins *atributs* són els que influeixen en els canvis de *patologia* o l'empitjorament de la malaltia, permetria l'anticipació a futurs esdeveniments i poder realitzar canvis a temps.

Identificar aquests atributs també pot servir de base per modificar els protocols de tractament, de tal forma que es puguin identificar els punts on és millor centrar els esforços en el tractament de pacients.

L'anàlisi predictiu és una àrea de la *mineria de dades* que pretén predir tendències i patrons de comportament a partir d'una gran base de dades. Cada cop s'utilitza més en la medicina predictiva analitzant dades antigues i aplicant-les sobre noves mostres per predir les probabilitats segons els valors aportats.

Es tracta d'una àrea que agrupa una varietat de tècniques estadístiques de modelització, aprenentatge autònom i *minería de dades* amb la finalitat d'analitzar les dades actuals i històriques reals.

L'àrea de recerca de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM), publicarà un estudi analitzant les causes de l'augment de pacients menors d'edat durant els últims anys. Algunes d'aquestes dades s'empraran en l'elaboració d'aquesta publicació, on figuraré com a col·laborador.

Prèviament al inici de la realització del treball final de màster he realitzat una recerca de articles i llibres que utilitzat com a base d'inici d'aquest treball, entre la informació recollida destaco: Time Trends Over 16 Years in Incidence-Rates of Autism Spectrum Disorders Across the Lifespan Based on Nationwide Danish Register Data [4], Trends in Mental Care among Children and Adolescents [5], Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction [6].

## 1.2 Objectius del Treball

L'objectiu del treball és realitzar una anàlisi predictiva a partir de la base de *dades massiva* de L'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM), .

Aquesta base de dades recopila la informació de caràcter sanitari (Csmij), valoració social, *anamnesis*, *psicopatologia*, medicaments i urgències/ hospitalitzacions des de l'any 2010 fins a la data d'avui.

Aquest treball mostra el primer anàlisi científic que s'ha dut a terme amb aquestes dades. És un estudi realitzat sobre els pacients menors de 18 anys ingressats per primera vegada el 2010.

Entre els objectius a aconseguir amb aquest estudi citem els següents:

A- Demostrar l'augment de pacients infantils (fins a 18 anys) que tenen algun tipus de trastorn com a pas previ per determinar quins factors personals poden ser causants de l'increment de la demanda (identificació de factors de risc).

B- Quantificar els pacients que canvien de *patologia* i els que milloren empitjoren el diagnòstic.

C- Analitzar els motius que porten als pacients a realitzar un canvi de diagnòstic o a la millora o empitjora de la gravetat del diagnòstic.

Per realitzar les anàlisis predictives s'ha utilitzat el software Rapidminer®.

### 1.3 Enfoc i mètode seguit

L'anàlisi de dades intel·ligent permet poder identificar la complexa relació entre els símptomes i les malalties de tractament, així com contribuir a la reducció del cost de l'atenció al pacient i el disseny i implementació de guies clíniques més ràpid.

El potencial d'una anàlisi predictiva és proporcionar a l'usuari final un suport a la presa de decisions mèdiques.

Entre moltes plataformes d'anàlisis de dades intel·ligent, podem trobar: Hadoop [15], Mahout [16], SparkR [17], PySpark [18], i RapidMiner®[19,20,21].

**Hadoop** [15]: és una implementació de codi obert de MapReduce. El sistema d'arxius distribuïts Hadoop (HDFS) és un sistema d'arxius basat en disc que s'estén a través dels nusos en un sistema distribuït. HDFS encapsula, distribueix i realitza l'emmagatzematge local en una única unitat lògica que ofereix la possibilitat de la divisió automàtica de dades en blocs i discos de replicació i locals.

**Mahout** [16]: va ser desenvolupat proporcionant biblioteques de mineria de dades escalables. Tot i que mahout és àmpliament utilitzat per a l'anàlisi predictiu escalable, també és qüestionada, ja que les seves biblioteques no proporcionen un marc general per a la construcció d'algoritmes i la qualitat de la solució donada varia significativament depenent de l'experiència del col·laborador. Mahout també se centra en algoritmes específics d'aplicació, en lloc de construir models d'execució.

**SparkR** [17] i **PySpark** [18]: proporciona als usuaris R i Python un front-  
fi [22] lleuger per a un sistema Spark mitjançant la compilació de les seqüències d'ordres declaratives a baix nivell. Els treballs amb MapReduce es consideren valuosos tenint en compte la popularitat de R i Python en la comunitat científica de dades.

**RapidMiner**[19,20,21].: es va fer molt popular en els darrers anys i és recolzat per una gran comunitat. El seu atractiu és el GUI (interfície gràfica d'usuari) i l'ajut contextual basat en wiki-base, ja que aquests dos factors permeten la facilitat d'ús i un aprenentatge ràpid. Un dels avantatges importants de RapidMiner® és la flexibilitat en el disseny del procés a través del "Procés / Subprocés", estructures i "Macros" que representen les variables globals de l'ambient. Això permet un disseny visual de processos complexos i d'alt nivell d'automatització.

Es realitzaran els anàlisis predictius amb el programa RapidMiner® pels avantatges que ens aporta i la facilitat de programació.

**RapidMiner<sup>i</sup>** és un programa informàtic per a l'anàlisi i mineria de dades. Permet el desenvolupament de processos d'anàlisi de dades mitjançant l'encadenament d'operadors a través d'un entorn gràfic. S'usa en investigació, educació, capacitació, creació ràpida de prototips i en aplicacions empresarials.

L'aplicació bàsica és gratuïta i inclou la capacitat de 10.000 columnes i més de 500 operadors.

La versió inicial va ser desenvolupada pel departament d'intel·ligència artificial de la Universitat de Dortmund en 2001. Es distribueix sota llicència AGPL i està allotjat en SourceForge des de 2004.

RapidMiner<sup>®</sup> proporciona més de 500 operadors orientats a l'anàlisi de dades, incloent els necessaris per a realitzar operacions d'entrada i sortida, preprocessament de dades i visualització. També permet utilitzar els algoritmes inclosos en Weka.

Pel format de la base de dades i la complexitat del món de la psiquiatria en agrupar i definir les patologies, s'acorda, juntament amb els professionals de L'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM) realitzar, en una primera fase, un estudi del perfil general de tots els pacients amb les patologies que tenen cadascun d'ells. Aquesta primera anàlisi general es realitzarà per tal d'identificar quines *patologies* tenen més nombre de pacients i quines deriven a altres *patologies* al llarg del temps.

Un cop identificat el nombre de pacients i comparat amb les memòries dels últims anys, es demostra que hi ha un augment de pacients si es compara en els últims anys.

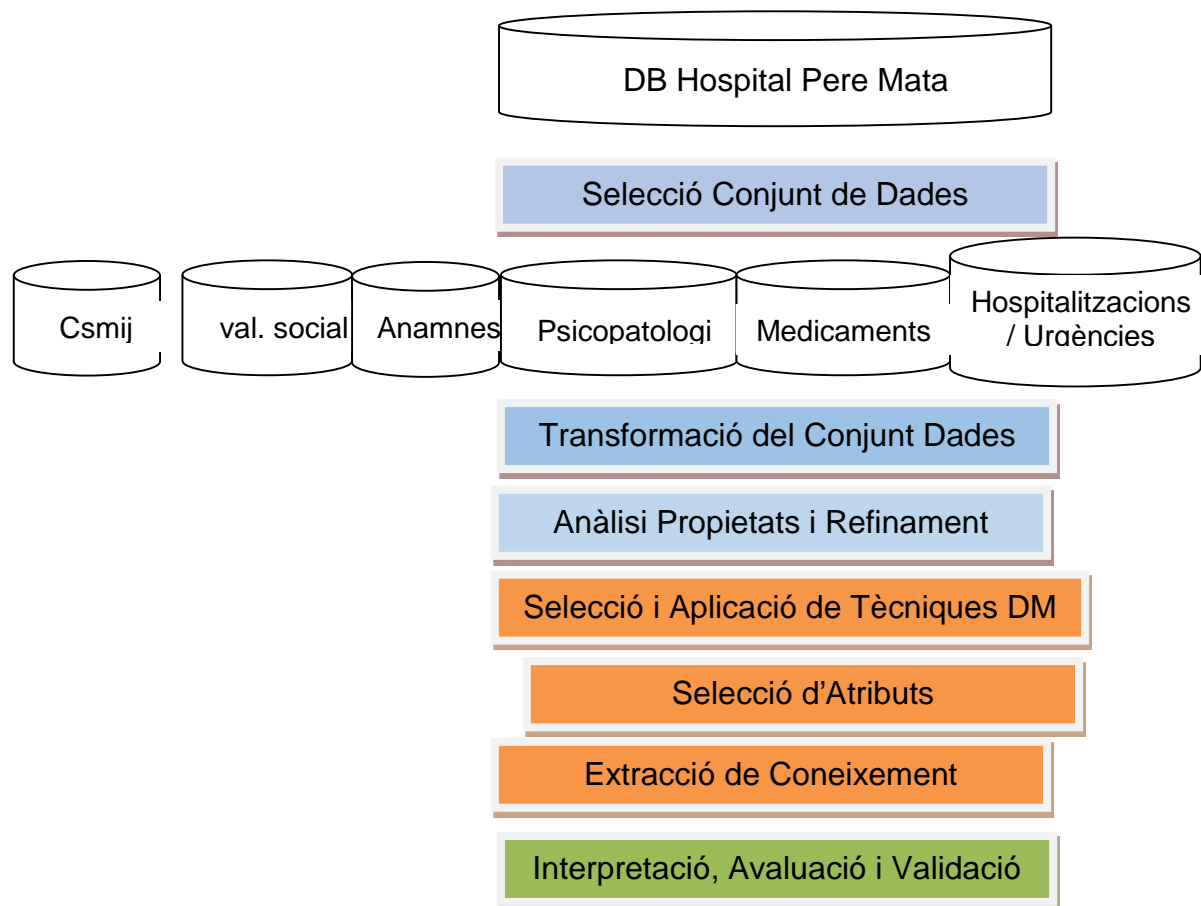
També es realitza una taula separada per sexe, per observar quin té més demanda.

Feta aquesta anàlisi inicial es decidirà quines *patologies* tenen més interès de ser analitzades des d'un criteri mèdic.

Se seleccionen els atributs més rellevants de la base de dades massiva amb mètodes iteratius de RaidMiner<sup>®</sup> i criteris mèdics.

Aquest enfocament requereix un temps de preparació de les dades i agrupació però ens permetrà evitar fer anàlisis innecessaris, seleccionant les patologies més significatives amb un nombre de mostres suficients per dur a terme aquest anàlisi.

El mètode seguit és el que il·lustra la imatge 1.



Imatge 1 Representació del procés a seguir

1 – *Selecció dels conjunt de dades a treballar.* Exportant la informació des de la base de l'hospital a una base de dades Excel i codificant el pacient amb un número (NMI) s'obté un subconjunt de dades propici per fer l'anàlisi.

Aquest pas es repeteix diverses vegades degut al gran nombre de dades disponibles a l'Hospital Pere Mata.

En cada arxiu Excel es disposa dels NMI (identificador de cada pacient) i els atributs acordats segons les característiques de cada estudi.

Es generaran sis arxius Excel: base Csmij, valoració social, *anamnesis*, *psicopatologia*, medicaments i urgències i hospitalitzacions.

Treballar amb totes les dades disponibles suposa tenir un volum enorme de dades, amb processos de càlcul lents i problemes de càrrega de dades al RapidMiner®, per això es decideix l'estructuració i selecció de les dades rellevants en sis arxius.

2 – *Transformació del conjunt de dades d'entrada.* El format d'entrada al programa d'anàlisi predictiu (RapidMiner®) és amb Excel, on en l'arxiu

Excel s'apliquen funcions per adequar el format de certes variables no compatibles amb RapidMiner®, permeten que el programa en pugui fer una lectura àgil i correcta.

3 – *Anàlisi de les propietats de les dades i posterior refinament* (presència de valors atípics).

S'observarà que les dades segueixin certa lògica, identificant en primer lloc la informació i els factors que fan que aquesta no concordi i, en segon lloc, eliminant les dades defectuoses que puguin alterar els resultats.

L'opció de treballar amb les distàncies euclidianes i eliminar els valors que s'allunyin del centre [3] es descarta degut a que la complexitat de les dades no permet generar un patró per detectar valors atípics. La causa és que moltes variables estan en format text i els errors estan ocasionats per la falta d'informació, dates mal col·locades, etc.

Per tant es treballa amb l'opció d'observació i identificació fent diferents tipus de llançament identificant les mostres amb dades defectuoses.

4 – *Selecció i aplicació de tècniques de mineria de dades*. Es construeix un model d'agrupació dels valors per tal d'identificar els més nombrosos. Es genera una taula per tal d'analitzar la quantitat de pacients que passen d'una patologia a una altra.

Es decideix fer aquest pas previ per tal de realitzar els anàlisis predictius als grups de més interès.

5 – *Selecció dels atributs*. S'aplicarà diferents operadors del programa RapidMiner® per tal d'identificar els atributs amb menys pes i així poder treballar amb un volum més acotat d'atributs.

6 – *Extracció del coneixement*. Mitjançant diverses tècniques d'anàlisi de dades s'extreuen models de coneixement que representen patrons de coneixement observat.

Es determina aplicar models d'anàlisi predictiu, els quals permetran a l'Hospital trobar certs patrons que influeixin en els diagnòstics.

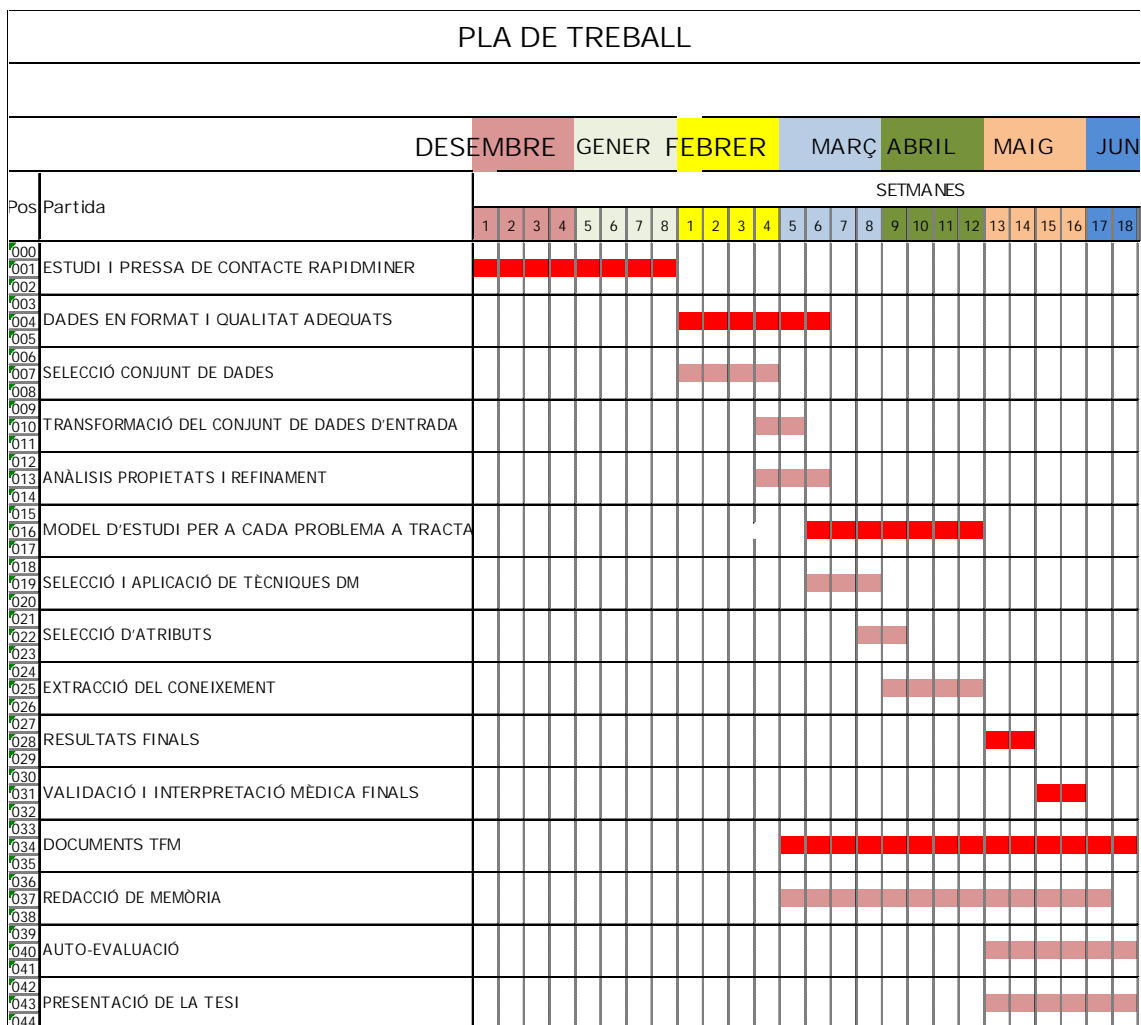
7 – *Interpretació, avaluació i validació dels resultats*. Un cop obtingut els patrons de coneixement s'interpreten facilitant la comprensió dels professionals i s'avaluaran mitjançant models d'avaluació que ofereix el mateix programa RapidMiner® indicant quin model és més precís.

Al tractar-se d'un nou tipus d'estudi en l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM), que no ha estat utilitzat anteriorment, no adaptarem cap tipus de software existent en l'hospital. Es generaran tots els algorismes nous amb el RapidMiner®.

#### 1.4 Planificació del Treball i Recursos Emprats

Els treballs d'extracció d'informació, neteja de dades, classificació de les mostres i anàlisis s'ha dut a terme a les instal·lacions de recerca de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM). La informació que s'utilitza a les instal·lacions està anonimitzada de manera que no es pot identificar el pacient, ja que es treballa amb dades privades i sensibles de pacients i per temes de responsabilitat, com de legislació, no poden ser extretes de les instal·lacions.

S'ha utilitzat les fulles de càlcul Excel per aplicar càlculs i modificacions de dades i el software RapidMiner® per la classificació, neteja i anàlisis de les dades.



Taula 1 Diagrama de Gantt del projecte de tesi de màster

## 1.5 Breu sumari de productes obtinguts

El projecte fi de màster conté com a resultat taules que quantifiquen el tipus de pacient segons l'interval d'edat, sexe i l'última patologia diagnosticada. I com a producte final, arbres de decisions amb el percentatge de confiança d'aquests que pretén predir si un pacient al cap de 6 anys del primer diagnòstic mantindrà el diagnòstic inicial o derivarà cap a un altre diagnòstic, si la gravetat del diagnòstic va cap a millor o pitjor segons el GAF(Escala de avaluació de funcionament global) [14].

## 1.6 Breu descripció dels altres capítols de la memòria

El capítol 2 descriu les metodologies possibles a utilitzar en cada pas i el motiu pel qual ens decidim per unes. Dintre les metodologies seleccionades es pretén fer un resum del principi de funcionament dels operadors elegits des d'una perspectiva informàtica com matemàtica.

El capítol 3 explica els treballs realitzats a les instal·lacions de recerca de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM), la informació disponible per poder generar els processos predictius (base de dades), els operadors aplicats en cada cas i els resultats obtinguts a les instal·lacions.

A les conclusions, capítol 4, es descriu dels resultats obtinguts, una valoració de si s'han aconseguit els objectius i propostes de nous anàlisis a executar en un futur.



## 2. Metodologies

En el transcurs del projecte s'ha emprat diverses metodologies d'anàlisi intel·ligent de dades, agrupades en metodologies per a la preparació de les dades, procediments per a l'anàlisi de la informació i metodologies per a la valoració dels resultats. Dins del primer grup trobem aquelles encaminades al refinament de les dades, l'agrupament de dades i la selecció d'atributs. Per a l'anàlisi de les dades hem emprat l'arbre de decisions com a model útil pels tècnics de l'hospital. Prèviament hem aplicat metodologies de càlcul del pes de cada atribut per fer una primera selecció dels atributs amb més incidència.

La validació de resultats ha estat el resultat d'aplicar l'operador "Cross-Validation" en que ens donarà com a resultat el percentatge de credibilitat que té.

Totes elles són detallades en les següents seccions.

### 2.1 Refinament de dades i tractament de valors atípics

Inicialment s'aplica un refinat de dades amb la finalitat de permetre un posterior anàlisi correcte, evitant tenir mostres que puguin alterar els resultats. S'identifica els casos defectuosos i es valora en cada cas quin procés es segueix per tal de no alterar resultats (eliminació, modificació, correcció, etc.).

Es considera dues formes alternatives per detectar valors atípics: distàncies [3] i contrast de resultats. D'aquests dos hem optat per la segona, ja que la primera presenta la dificultat de detectar atípics al treballar amb variables qualitatives o binaries. A més en el context en què es treballa, alguns valors que es podrien classificar com atípics, poden suposar una informació rellevant.

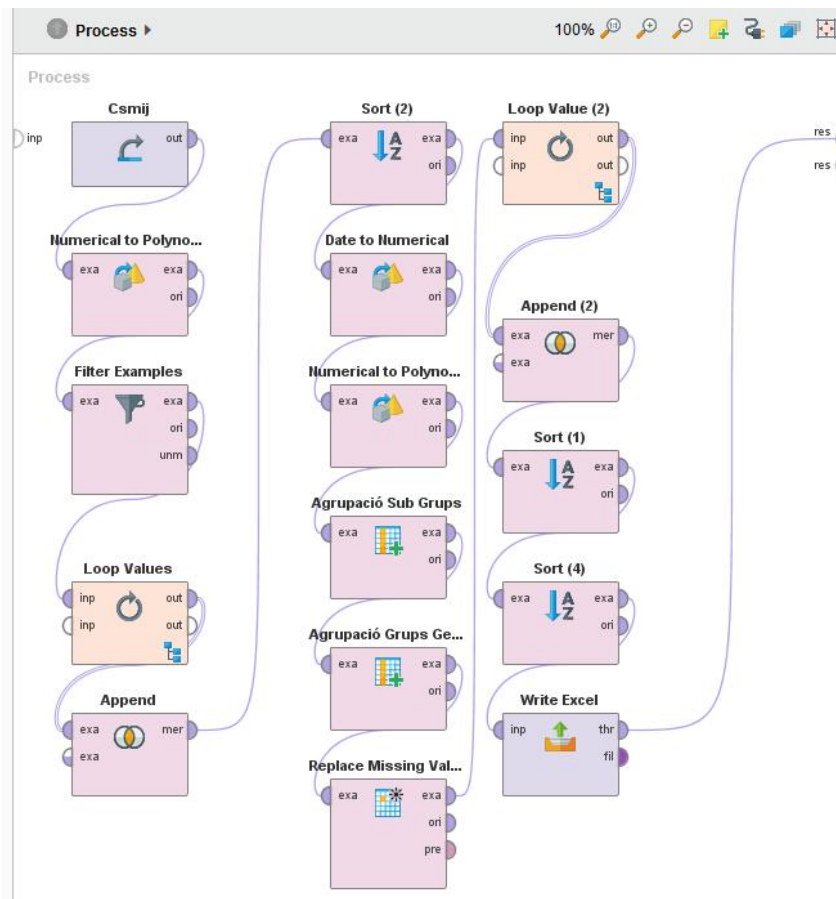
L'opció estimada és l'execució de diferents computacions amb RapidMiner® i Excel, com pot ser la detecció de falsos diagnòstic mitjançant l'aplicació de filtres o contrastant dates per detectar-ne d'il·lògiques, com tindre un diagnòstic abans de la primera visita, etc.

## 2.2 Agrupació de dades i correlació de malalties

Els pacients estan definits cada un d'ells amb un identificador NMI. Cada pacient està afectat per un tipus de patologia amb codificació CIE9. El CIE9 [2] és un estàndard de l'Organització Mundial de la Salut (OMS) per a la codificació de malalties.

En una primera fase s'agrupen els pacients segons el codi CIE9 [2] en grups de característiques semblants. Els professionals de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM) són els qui defineixen com agrupar patologies similars i quins grups i sub-grup se'ls assigna a cada una d'elles.

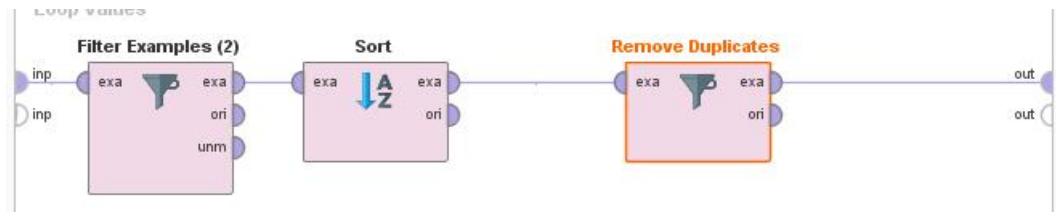
Amb els criteris d'agrupació definits pels professionals de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM) es dissenya una aplicació RapidMiner® per automatitzar el procés d'agrupació dels pacients segons el número CIE9 assignat en cada patologia.



Imatge 2 Procés RapidMiner® (Assignació grups generals segons cie9)

El procés d'agrupació s'inicia amb la informació que conte la base Csmij, on els atribut NMI son transformats a polinominals, ja que certes funcions RapidMiner® de bucle no ens permeten treballar amb valors no polinominals. El següent operador "filtre exemple" executa un filtrat de dades inicial on s'eliminen els valors inactius, nuls i diagnòstic

secundaris. I seguidament s'aplica una iteració per eliminar els diagnòstics principals que no són inicials (loop value), aquest procés es compon de les 3 etapes que es mostren en la imatge 3: filtrat, ordenació i eliminació de duplicats.



Imatge 3 Exemple de Loop Value Procés RapidMiner® (Assignació grups generals segons cie9)

Es transforma la resta de variables a polinomial i s'aplica l'operador "agrupació Sub Grups", encarregada d'assignar un subgrup a cada numero CIE9 [2]. Per exemple, els pacients amb una CIE9 igual 295 s'agrupen amb esquizofrènia i altres trastorns psicòtics, una CIE 9 igual 296.31 s'agrupa a depressió.

```

1 if(cie9=="295", "Esquizofrènia i altres trastorns psicòtics", if(cie9=="295.40", "Esquizofrènia i altres trastorns psicòtics",
2 if(cie9=="295.91", "Esquizofrènia i altres trastorns psicòtics", if(cie9=="296.20", "Depressió", if(cie9=="296.21", "Depressió",
3 if(cie9=="296.22", "Depressió", if(cie9=="296.23", "Depressió", if(cie9=="296.24", "Depressió", if(cie9=="296.31", "Depressió",
4 if(cie9=="296.41", "Bipolar", if(cie9=="296.44", "Bipolar", if(cie9=="296.54", "Bipolar", if(cie9=="296.82", "Depressió",
5 if(cie9=="298.0", "Esquizofrènia i altres trastorns psicòtics", if(cie9=="298.4", "Esquizofrènia i altres trastorns psicòtics",
6 if(cie9=="298.9", "Esquizofrènia i altres trastorns psicòtics", if(cie9=="299.0", "Trastorn de l'espectre autista",

```

Imatge 4 Exemple Operador "Agrupació Subgrups", Procés RapidMiner® (Assignació sub grups generals segons cie9)

El següent operador "agrupació Grups Generals" encarregat d'assignar un grup general a cada subgrup. Per exemple, la imatge de la figura 5 mostra part de les regles d'agrupació que indiquen que els "Altres trastorns de la conducta" es classifiquen com a "Trastorn de la conducta", els "Trastorns de Personalitat Antisocial" com "Trastorns de la personalitat", etc.

```

1 if(grups_diagnostic=="Altres trastorns de conducta", "Trastorn de conducta",
2 if(grups_diagnostic=="Trastorn de Personalitat Antisocial", "Trastorns de la personalitat",
3 if(grups_diagnostic=="Altres TA", "Trastorns d'adaptació",
4 if(grups_diagnostic=="Reaccions depressives", "Trastorns d'adaptació",
5 if(grups_diagnostic=="Altres trasorns del desenvolupament", "Trastorns de l'aprenentatge",
6 if(grups_diagnostic=="altres 307", "Altres", if(grups_diagnostic=="Altres Infància", "Trastorns específics de la infància",

```

Imatge 5 Exemple Operador "Agrupació Grups Generals", Procés RapidMiner® (Assignació grups generals segons cie9)

El loop value (2) suma els valors que tinguin la mateix CIE9 per fer una lectura més fàcil. I finalment s'ordena i s'exporta amb Excel.

### 2.3 Selecció d'atributs

Les variables a processar en les anàlisis predictives són majoritàriament *qualitatives binaries*. Això vol dir que prenen dos valors com per exemple: home/dona, Sí/No, etc. Quan a la mostra se l'hi assigna un 1 a una variable específica vol dir que aquesta variable és positiva, en cas

de ser 0 significa que el professional que atén al pacient no ha seleccionat aquesta variable.

El *atributs* utilitzats per descriure patologies, centres, entre altres, seran variables *qualitatives generals* amb valors múltiples

També hi ha variables quantitatives com l'edat, la gravetat d'una patologia, etc.

Per seleccionar els *atributs* més influents hi ha diversos mètodes, entre els quals en destaquem el weight by SVM (pes per SVM), weight by Information gain (pes mitjançant la guany d'informació), weight by correlation (pes mitjançant la correlació) i weight by gini index (pes mitjançant el índex gini).

**Weight by SVM:** Aquest operador es basa amb el mètode de selecció d'*atributs* per Support Vector Machine (SVM) de *Maldonado i Weber* [1]. L'estratègia es basa en l'eliminació seqüencial cap enrere, determinant la contribució de cada atribut i eliminant aquell que impacta menys en l'exercici de classificació en un conjunt de validació independent. Començarà amb tots els atributs disponibles i cada iteració eliminarà aquells atributs que afecten l'execució predictiva fins que s'arribi al criteri de parada.

La idea bàsica del mètode proposat és la d'eliminar aquells atributs que la seva eradicació impliqui un menor nombre d'errors en un conjunt de validació independent.

El mètode rep el nom de Hold-out Support Vector Machines (HO-SVM) [1], ja que en cada iteració l'algoritme genera una nova partició del conjunt de dades en dos subconjunts: un d'entrenament, per construir la funció de classificació, i un altre de validació, on s'avalua l'acompliment predictiu de la funció construïda i se seleccionen els atributs a eliminar.

A continuació es presenta l'algoritme iteratiu per a l'eliminació d'atributs:

### 1. Selecció del Model

### 2. Inicialització

### 3. Repetir

a) Partició aleatòria del conjunt de dades (*hold-out*)

b) Entrenament del model. Equació;

$$\text{Max}_{\alpha} \quad \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i,s=1}^m \alpha_i \alpha_s y_i y_s K_{\sigma}(\mathbf{x}_i, \mathbf{x}_s)$$

c) Per a tot atribut  $p$  amb  $p = 1$ , calcular  $E(-p)$

( , ), el número d'errors de classificació quan l'atribut  $p$  és remogut.

d) Eliminar atribut  $j$  amb menor valor de  $E(-p)$  ( , )

**4. Fins que** el menor valor de  $E(-p)$  ( , ) sigui major que  $E( , )$ , l'error de validació amb tots els atributs seleccionats que compleixin  $= 1$ .

## End

Cal recordar que per a què aquest operador funcioni correctament tots els atributs han d'estar en format numèric.

**Weight by information Gain:** Aquest algoritme estaria més classificat com un algoritme de rànquing d'atributs, però ens serà molt útil, en aquest cas, fer certs anàlisis que poden tenir molta utilitat per l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM).

L'operador permet calcular el nivell de contribució de cada atribut per predir la importància que pot tenir. Com més pes tingui l'*atribut*, més rellevant es considera. S'assumeix amb aquest algoritme que no existeix relació entre els atributs i, per altra banda, es calcula sense construir ni avaluar cap model.

El valor esperat del guany d'informació és la informació mútua  $I(X; A)$  de  $X$  i  $A$  - és a dir, la reducció de l'entropia de  $X$  aconseguida mitjançant l'aprenentatge de l'estat de la variable aleatòria  $A$ .

Aquest concepte pot utilitzar-se per definir una seqüència d'*atributs*, poder-los investigar amb més rapidesa i reduir-ne l'estat de  $X$ .

La seqüència depèn del resultat de la investigació dels atributs anteriors en cada etapa.

En general, un atribut amb alta informació mútua ha de ser preferit a altres atributs.

La definició de guany d'informació (IG) es resumeix en la següent fórmula:

$$IG(T, a) = H(T) - H(T|a)$$

H- Informació de l'entropia

La definició formal es defineix:

Sigui  $T$  un conjunt d'exemples, on cada un d'ells  $(x, y) = (x_1, x_2, \dots, x_k, y)$  on  $x_a \in \text{vals}(a)$  és el valor de  $a$ -èssim de l'atribut exemple  $x$  i  $y$  és la corresponent classificació. El guany d'informació de l'atribut és definit en termes d'entropia  $H()$ , com segueix:

$$IG(T, a) = H(T) - \sum_{v \in \text{vals}(a)} \frac{|\{\mathbf{x} \in T | x_a = v\}|}{|T|} \cdot H(\{\mathbf{x} \in T | x_a = v\})$$

**Weight by Correlation:** Aquest operador calcula la rellevància dels atributs calculant el valor de correlació per a cada atribut d'entrada respecte a l'atribut d'etiqueta (labeled). Aquest esquema de ponderació es basa en la correlació i retorna el quadrat de la correlació, com el pes d'atribut.

No es pot aplicar l'atribut en format polinòmic perquè les classes polinòmiques no proporcionen cap informació sobre la seva instrucció. Per tant, els pesos són més o menys aleatoris en funció de la representació numèrica interna.

Els atributs binominals funcionen a causa de la representació com 0 i 1, de la mateixa manera que els numèrics.

Una correlació és un nombre entre -1 i 1 que mesura el grau d'associació entre dos atributs.

Es considera que dues variables quantitatives estan correlacionades quan els valors d'una d'elles varien sistemàticament pel que fa als valors homònims de l'altra. Per exemple, per a dues variables (A i B) existeix correlació si en augmentar els valors de A també augmenten els de B i viceversa. La correlació entre dues variables no implica, per si mateixa, cap relació de causalitat.

**Weight by Gini Index:** Aquest operador calcula la rellevància dels atributs de la base de dades basant-se en l'índex de Gini d'impureses.

El coeficient de Gini és una mesura de la desigualtat. S'utilitza per calcular qualsevol forma de distribució desigual. El coeficient de Gini és un nombre entre 0 i 1, on 0 es correspon a la perfecta igualtat (tots tenen els mateixos ingressos) i 1 es correspon a la perfecta desigualtat (una persona té tots els ingressos i els altres cap).

## 2.4 Models predictius

Els models elegits seran models predictius, els quals s'aplicaran sobre un subjecte per avaluar la probabilitat d'un succés.

Com s'ha esmentat en capítols anteriors, l'eina escollida per fer l'anàlisi predictiu és el RapidMiner®, que ens dóna una varietat d'algoritmes per tal de poder fer la predicció més adient. Alguns dels algoritmes més interessants que ens ofereix són el Bayes Ingenu, els arbres de decisió, la regressió logística i les xarxes neuronals.

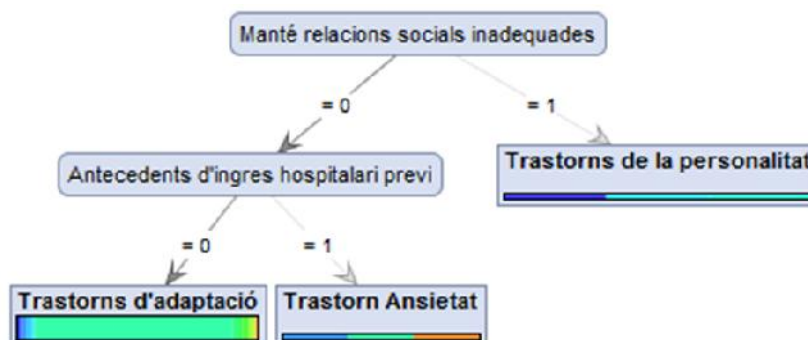
**Bayes ingenu (de l'anglès *Naive Bayes*)** [7]: És una tècnica de classificació i predicció que construeix models probabilístics a partir de dades històriques.

En un primer pas, a partir de les dades històriques, es defineixen els atributs que caracteritzen un objecte, estudi, coneixement, etc. Un cop definit el model predictiu sabem els atributs que l'afecten i quin pes té cada un d'ells. Aplicant el model sobre noves dades podrem anticipar futurs esdeveniments i definir la probabilitat que esdevingui una característica o objecte.

Degut que els tècnics no tenen previst aplicar RapidMiner® sobre noves mostres, no és aplicable, ja que aquest model ens ofereix les mitjanes i les desviacions de cada atribut per aplicar a una nova mostra mitjançant el mateix operador.

**Arbres de decisió** [8]: Els arbres de decisió són models predictius molt utilitzats en dominis molt diversos. Aquests són molt útils per a visualitzar les diverses opcions que es tenen i per definir patrons de decisió. L'avantatge que tenen és que un cop generats no és necessari disposar de l'operador per valorar el subjecte, a diferència d'altres mètodes predictius que et calculen els coeficients de mostres anteriors per aplicar a les noves mostres i que es requereix aplicar l'operador per saber les probabilitats.

El que forma l'arbre de decisions son: nusos, vectors de números, fletxes i etiquetes.



Imatge 6 Exemple arbre de presa de decisions RapidMiner®

Cada **nus** es pot definir com un punt en què cal prendre una decisió d'entre diverses possibles, fet que va ocasionant que a mesura que augmenta el nombre de nusos augmenti el número de possibles finals als quals pot arribar la predicció.

Els **vectors de números** serien la solució final a la qual s'arriba en funció de les diverses possibilitats que es tenen, ja que donen les utilitats en aquesta solució.

Les **fletxes** són les unions entre un nus i un altre i representen cada acció diferent.

Les **etiquetes** es troben en cada nus i cada fletxa i donen nom a cada acció.

L'operador Decision Tree del RapidMiner utilitza un algoritme basats amb Quilan's C4.5 [11] o CART [12].

Cada cop que un nou nus es crea en un determinat moment, un atribut s'escull per maximitzar el poder de discriminació d'aquest nus respecte els altres alternativament assignat al sub-arbre.

**Xarxa neuronal** [9]: El model de xarxes neuronals és aplicable només per a atributs numèrics, pel que és necessari assegurar que totes les dades siguin numèriques. Pel format que ens ofereix els resultats i la tipologia de dades a analitzar s'ha descartat.

**Regressió Logística** [10]: És un model de regressió estadística on la *variable dependent* és categòrica binaria. En aquets casos s'ha d'adaptar la base de dades per tenir l'atribut labbeled com a binari. Aquest model genera coeficients de cada atribut per aplicar a una nova mostra. Ja que els professionals no utilitzaran l'operador per a noves mostres, no es considera útil com analitzador predictiu. Sí que ofereix el coeficient p-value, que ens indica la dependència lineal entre cada atribut i l'atribut labbeled, on es considera que si aquest valor es inferior a 0.05 té dependència lineal i, per tant, influeix en el resultat del subjecte.



### 3. Implementació de les metodologies escollides.

El treball realitzat a les instal·lacions de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM) s'executa a partir d'una base inicial (Base Csmij) on es desenvolupen diferents processos de refinat i selecció de mostres. De les mostres obtingudes s'obté el subjecte(NMI) utilitzat com a referència per extreure les mostres de la resta de bases (valoració social, anamnesis, psicopatologia, medicaments, freqüència de les visites als professionals, urgències/hospitalitzacions).

Tots els subjectes a analitzar s'agrupen segons el codi CIE9 [2], quantificant el nombre de pacients de cada patologia i mitjançant una correlació de malalties s'analitzen els canvis de patologies que experimenten certs pacients.

A partir de les patologies dels pacients, la gravetat d'aquestes, com el nombre d'assistències sanitàries requerides, s'apliquen tècniques de selecció d'atributs a partir del pes que tenen en el resultat. Els atributs amb més pes són els que influeixen més en la presa de decisió i seran els més representatius en els models d'anàlisi predictiva. Aquests atributs seran els que s'aplicaran al model predictiu.

En l'execució dels models predictius es seleccionen aquells que tinguin utilitat per als tècnics de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM), com pot ser l'arbre de decisions. Aquests es validaran mitjançant l'operador "Cross-Validation" per tal de saber la probabilitat d'encert. I dels atributs significatius es valora l'impacte que té sobre el subjecte.

#### 3.1 Base de dades

La informació continguda en el sistema de l'Hospital Universitari Pere Mata s'ha extret en diverses taules Excel.

Totes les taules estan relacionades amb el NMI, que és l'identificador de cada pacient.

La informació d'aquestes taules és la relacionada amb els pacients menors de 18 anys que van realitzar la seva primera visita durant l'any 2010.

A continuació s'inclouen les variables de cada taula. No s'han copiat les taules originals degut a la protecció de dades dels pacients.

**Taula Csmij:** En aquesta taula centralitza la informació dels diagnòstics sobre els pacients. Inclou informació relacionada amb el centre d'atenció, la data de les visites, els diagnòstics codificats segons el CIE9 [2] i el CIE10 [13] i l'edat del pacient en el dia de la seva primera visita.

Està formada per 1.136 pacients i 4.983 files.

Taula CSMIJ	
NMI	identificació del pacient (codi anonimitzat prèviament)
Primera-Visita	data de la primera visita
Data_inici_diagnòstic	data en què es diagnostica la malaltia (diagnòstic principal).
Diagnòstic-principal	0 (és diagnòstic principal) 1 (és diagnòstic secundari)
CIE10	codi diagnòstic en CIE10 [13]
Desc_diag_cie10	descripció del diagnòstic segons la CIE 10 [13]
CIE9	codi diagnòstic en CIE9 [2]
Desc_diag_cie9	descripció del diagnòstic segons la CIE 9 [2]
Inactiu	data en què es cancel·la la informació entrada a la base
Anul·lat	data en que es cancel·la la informació entrada a la base
Edat	edat del pacient quan fa la primer visita

Taula 2 Informació que conte la base de dades Csmij

En la taula del Csmij, un mateix pacient pot ocupa més d'una fila. La seva freqüència d'aparició en una mateix taula depèn del número de modificacions aplicades. Per exemple en el 2010 se li diagnostica amb un codi CIE9 [2] i al cap de cert temps se l'hi és modificat el CIE9 [2] per evolució cap a un altre trastorn.

La freqüència d'aparició també dependrà dels diagnòstics secundaris que se'ls apliqui.

Les columnes inactiu i anul·lat indiquen si alguna mostra es va cancel·lar. Si en una fila apareix la data en la casella d'inactiu o anul·lat s'eliminarà de l'anàlisi.

**Taula PSICOPATOLOGIA:** Conté informació sobre les patologies psíquiques dels pacients en moments determinats. Cada psicopatologia està representada per una columna binària on 1 representa la presència de la patologia i 0 l'absència.

La taula de psicopatologia conté informació sobre 1087 pacients i està formada per 1402 files. Les psicopatologies contingudes estan classificades en:

- *Intel·ligència* (4), mesuren la intel·ligència: dins la normalitat (xim\_normal), la discapacitat intel·lectual (xim\_retardo), la intel·ligència sobre la normalitat (xim\_mas\_normal) i l'empitjorament intel·lectual (xim\_deterioro).
- *Llenguatge* (22), mesuren aspectes del llenguatge: si és adequat (sll\_adequado), to elevat (xll\_tono\_alto), mussitació (xll\_musitacion), monòton (xll\_monotono), enfàtic (xll\_enfatico), pobre (xll\_Pobre), laxe (xll\_deshilachado), no espontani (xll\_no\_espontaneo), monosil·làbic (xll\_monosilabo), mutisme (xll\_mutisme), logorrea (xll\_logorrea), prolix (xll\_Prolix), intel·ligible (xll\_intelegible), tartamut (xll\_tartamudo), afàsia (xll\_afasia), disàrtria (SI/NO) (xll\_disatria), llenguatge absent (xll\_absente), falta de fluïdesa i ritme (xll\_falta\_ritmo), parla formal (xll\_habla\_formal), estereotips verbals (xll\_Estereotipos), dislàlies (xll\_disalalies) i alteracions gramaticals (xll\_altera\_gram).

- *Humor i estat d'ànim (18)*, mesuren les alteracions rellevants relacionades amb els canvis d'humor i els estats d'ànim:  
Eutímic (xha\_eutimic), tristes (xha\_tristes), eufòria (xha\_euforia), disfòria (xha\_disforia), delèroide (xha\_deleroide), baixa autoestima (xha\_baja\_autoestima), alta autoestima (xha\_alta\_estima), apatia (xha\_apatia), adinàmia (xha\_adinamia), pèrdua d'interès (xha\_perdida\_interes), indiferència (xha\_indiferencia), irritabilitat (xha\_irratibilidad), por (xha\_miedo), preocupacions hipocondríiques (xha\_hipocondriacs), sentiments d'incapacitat (xha\_incapacidad), desesperança (xha\_desesperanza) i sentiments de buit (xha\_sentimiento\_vacio) idea mort (xha\_idea\_muerto):
- *Afectivitat ( 9 )*, alteració on es mostra un empobriment emocional davant els fets. Està marcat per un grau més o menys elevat de desànim, tristesa i depressió:  
Conservada (xha\_conservada), fredor afectiva (xha\_frio\_afecto), alexitimia (xha\_alexitimia), aplanament afectiu (xha\_allan\_afecti), anhedonia (xha\_adonemia), afectivitat inadequada (xha\_inade\_afecti), labilitat emocional (xha\_labilitat\_emo), embotiment afectiu (xha\_embotamiento) i Incongruència afectiva (xha\_incongruencias)
- *Alteracions en el curs del pensament ( 5 )*, es manifesten verbalment i són les diferents alteracions en el procés o flux del pensament. És a dir, com fórmula, s'organitza i expressa els seus pensaments:  
Accelerat (xpe\_acelerado), fuga d'idees (xpe\_fuga\_de\_ideas), inhibició (xpe\_inhibido), bloqueig (xpe\_bloqueo) i dubitatiu (xpe\_dubitativo).
- *Alteracions en la forma del pensament ( 9 )*, conjunt de símptomes que afecten el pensament i/o el llenguatge i/o la comunicació dels pacients:  
Tangencialitat divagació (xpe\_tangentes), circumstancialitat "polijidad" (xpe\_circumstanci), perseveració (xpe\_perseveracion), alògia (xpe\_alogia), verbigeració (xpe\_verbigeracion), associacions laxes (xpe\_asociaciones), incoherència (xpe\_incoherencia), disgregació (xpe\_digregación) i associació per so (xpe\_digre\_so).
- *Alteracions en el contingut del pensament ( 12 )*, alteració que afecta el pensament del pacient:  
Preocupacions (xco\_preocupacions), rumiació (xco\_rumiaciones), Idees Obsessives (xco\_idea\_obsev), compulsions mentals (xco\_comp\_mentales), idees fóbiques (xco\_idea\_fobias), pensament màgic (xco\_pens\_magico), ideació suïcida (xco\_idea\_suicida), idees sobrevalorades (xco\_idea\_sobre), idees errònies (xco\_idea\_error), idees delirants (xco\_idea\_delirio), interpretacions delirants (xco\_inter\_delirio) i idees doctiroides (xco\_ideas\_deliroides).

- *Temàtica del pensament ( 10 )*, afecta a creences de determinats temes del pensament del pacient:  
De referència (xte\_referencia), persecució (xte\_persecución), control (xte\_control), místic (xte\_misitic), celotípics (xte\_celotipico), eromaniac (xte\_eromatico), culpa (xte\_culpa), ruïna (xte\_ruina), hipocondríac (xte\_hipocondriac) i de grandesa (xte\_grandes)
- *Psicomotricitat ( 14 )*, alteracions que poden afectar en la psicomotricitat del pacient i que afecten a la capacitat de fer i d'expressar-se:  
Sense alteració en la psicomotricitat (xps\_sin\_altera), alentiment (xps\_enlentiment), augment (xps\_aumento), inquietud (xps\_inquietud), agitació (xps\_agitacio), inhibició (xps\_inhibicion), negativisme (xps\_negativo), estereotips (xps\_estereo), manierisme (xps\_manierismo), tics simples (xps\_tic\_simple), tics complexos (xps\_tic\_complejos), tremolors (xps\_escolofrios), pseudoconvulsions (xps\_pse\_convulsion) i estupor/ catatonia (xps\_estupor)
- *Trastorn de la vivència del jo ( 4 )*, trastorn de la consciència que afecta al reconeixement i la relació amb el seu jo psíquic o corporal:  
Sense alteracions de la vivència del jo (xjo\_sin\_altera), despersonalització (xjo\_despersonalizado), desrealització (xjo\_altera), influència, robatori, inserció i divulgació i/o eco del pensament (xjo\_influencia)
- *Trastorn de la son ( 11 )*, trastorn que pot afectar al desenvolupament habitual del cicle de la son:  
Sense alteracions en els trastorns de la son (xts\_sin\_altera), insomni d'inici (xts\_sin\_inicio), insomni de manteniment (xts\_ins\_mant), despertar precoç (xts\_despertar), hipersòmnia (xts\_hipersomnia), del cilce son vigila (xts\_ciclo\_sueno), por a l'anar al llit (xts\_miedo\_cama), somnambulisme (xts\_sonambulo), terror nocturn (xts\_terror\_nit), bruixisme (xts\_bruixisme) i malsons (xts\_mal\_sueno).
- *Trastorn de l'alimentació ( 8 )*, trastorns mentals que es caracteritzen per conductes i pensaments alterats envers l'alimentació, el pes o la figura:  
Alteracions qualitatives tipus pica (xta\_qual\_pica), alteracions qualitatives tipus rumiació (xta\_qual\_rumi), alteracions qualitatives tipus rebuig de l'alimentació (xta\_qua\_rech\_al), alteracions quantitatives tipus rebuig selectiu (xta\_qua\_rech\_se), alteracions quantitatives tipus anorèxia (xta\_qua\_anorexia), alteracions quantitatives tipus sitofòbia (xta\_qua\_sitofobia), alteracions quantitatives tipus patomania (xta\_qua\_patomania), alteracions quantitatives tipus atracons (xta\_qua\_atiborro).
- *Trastorns de la sexualitat ( 6 )*, trastorns, majoritàriament provocats per problemes psíquics, que poden ocasionar perturbacions en la sexualitat:  
Sense alteracions de la sexualitat (xts\_sin), egodistoria, identitat orientació sexual (xts\_egodistoria), disfuncions psicosexuals (xts\_disfunciones), parafilies (xts\_parafilies), altres alteracions de

l'esfera sexual (xts\_otros), descripció d'altres alteracions de l'esfera sexual "promiscuïtat, conductes de risc i evita parlar de temes relacionats amb la sexualitat" (xts\_otros\_v).

- *Control d'impulsos i conducta ( 18 )*, símptomes que ocasionen problemes en la regulació del comportament i de les emocions:  
Sense alteracions del control d'impulsos i conducta (xcic\_sin), impulsions (xcic\_impulsions), compulsions mentals (xcic\_compulsions), heteroagressivitat (xcic\_hete\_agresi), explosivitat (xcic\_explosividad), autoagressivitat (xcic\_auto\_agre), autoagressivitat tipus temptativa suïcidi (xcic\_auto\_temp), autoagressivitat tipus parasuïcidi (xcic\_auto\_para), deshibició sexual (xcic\_deshibi\_sexual), conductes atípiques (xcic\_con\_atipica), conductes apragmàtiques (xcic\_con\_aparg), conductes manipuladores (xcic\_), conducta disocial (xcic\_con\_disocial), fugues (xcic\_fugas), tricofil·lomania (xcic\_tricotilomania), cleptomania (xcic\_cleptomania), piromania (xcic\_piromania) i ludopatia (xcic\_ludopatia).
- *Control d'esfínter ( 1 )*, símptomes que ocasionen capacitat per poder controlar els esfínters:  
Presenta eneuresi (xts\_eneuresi).

**Taula de Valoració Social:** Conte informació sobre la valoració de la interacció del pacient com a individu d'una comunitat. La majoria de variables socials són binàries (1/0), excepte la situació legal del menor i tipus de condicions de l'habitatge.

La base de dades valoració social està formada per 124 pacients i 183 mostres. Les valoracions socials estan dividides en:

- *Valoració social (5)*, variables que valoren la interacció del pacient com a individu d'una comunitat:  
Sistema legal del menor "1 tutela dels pares, 2 tutela en el medi, 3 tutela de la administració" (x\_situacion\_menor), existència de mesura penal alternativa (xmedia\_penal), intervenció per part del ICASS (xbien\_icass), habitatge (xvivienda), tipus de habitatge de la persona "1 bona, 2 acceptable, 3 falta d'un servei bàsic, 4 insalubre i 5 extrema inhabilitat" (xcondicion\_vive).
- *Relacions familiars ( 5 )*, tipus de relació, o absència d'aquesta, que els pacients tenen amb la seva família:  
Manté relacions familiars adequades (xrf\_rela\_fami), nul·les relacions familiars (xrf\_no\_rela\_fami), manté relacions familiars inadequades (xrf\_rela\_fam\_in) i dificultats contínues en les relacions familiars (xrf\_dificil), claudicació familiar (xrf\_claudicacion).
- *Habitatge ( 5 )*, si disposen o no d'habitatge i valoració de les condicions en què està. També parlaríem de si el pacient està capacitat per romandre-hi:

Falta d'habitatge (xvi\_falta\_vivienda), habitatge inadequat (xvi\_vivienda\_ina), bones condicions de l'habitatge (xvi\_buenas\_cond), dificultats en l'habitatge per la malaltia (xvi\_dificil\_vivienda) i habitatge protegit inadequat (xvi\_vivi\_protegido)

- *Situació econòmica ( 4 )*, variables que descriuen la situació econòmica del pacient i gestió d'aquesta per part del pacient:  
Absència d'ingressos econòmics (xse\_sin\_ingresos), ingressos econòmics insuficients (xse\_ing\_ins), situació econòmica adequada (xse\_adeuada), dificultats en la gestió i administració econòmica per la malaltia (xse\_dif\_eco\_mal).
- *Aficions i temps lliure ( 3 )*, aficions o absència d'aquestes en el temps lliure del pacient:  
Sense aficions (xat\_sin), no regula el temps lliure (xat\_regula) i participació en una o més activitats de temps lliure (xat\_actividades).
- *Relacions socials ( 7 )*, tipus de relacions socials que té el pacient i la qualitat i vivències d'aquestes:  
Manté relacions socials adequades (xrs\_adequadas), nul·les relacions socials (xrs\_nulas), manté relacions socials inadequades (xrs\_inadequadas), dificultats contínues en les relacions familiars (xrs\_dificil), facilitat per establir relacions (xrs\_facil), rebuig a noves relacions (xrs\_rechazo) i desig d'incrementar les relacions socials (xrs\_deseo).

**Taula de anamnesis:** Conté informació proporcionada pel propi pacient sobre la seva història clínica. En aquesta taula tindrem variables binàries qualitatives i variables numèriques quantitatives com el pes de naixement.

La base de dades d'anamnesis està formada per 959 pacients i 1.402 mostres. Les variables que la formen són ( 37 ):

Part a pre-terme (xen\_duracion\_pre), sistema de gestació "28/34/38" (xen\_duracion\_pre\_de), part a terme (xem\_duracion\_ter), part a post terme (xem\_duracion\_post), pes del nen al néixer en Kg (xparto\_peso), anys d'inici de la deambulació (xdes\_edad\_ano), mesos d'inici de la deambulació (xdes\_edad\_mes), anys en que la persona va dir les primeres paraules (xdes\_len\_anos), mesos en que la persona va dir les primeres paraules (xdes\_edad\_mes), anys que la persona controla l'esfínter anal (xedad\_anal\_ano), mesos que la persona controla l'esfínter anal (xedad\_anal\_mes), malalties somàtiques rellevants (xant\_mal), intervencions quirúrgiques (xant\_inerv), traumatismes i/o accidents (xant\_trauma), consum de tòxics actualment (xtox\_actual), consum de tòxics en el passat (xtox\_pasado), antecedents de tractament psicològic greu (xanttps\_psico), antecedents de tractament farmacològic greu (xanttps\_farma), antecedents d'ingrés hospitalari greu (xanttps\_ingreso), integració adequat amb els iguals adequat

(xescu\_int\_decua), dificultat amb la integració amb els iguals (xescu\_int\_difi), adaptació amb els professors adequada (xescu\_adap\_adeq), dificultat amb l'adaptació dels professors (xescu\_adap\_dif), adquisició d'aprenentatges adequada (xescu\_adap\_adeq), dificultats en la separació dels pares (xescu\_dif\_v), antecedents mèdics familiars dels pares (xpar\_padre), codi CIE 10 par (xtar\_cie10\_pa), antecedents mèdics familiars en la mare (xpar\_madre), codi CIE 10 mare (xtar\_cie10\_hi), antecedents mèdics familiars en la filla (xpar\_filla) codi CIE 10 filla (xtar\_cie10\_ha), antecedents mèdics familiars en el germà (xpar\_hermano), codi CIE 10 germà (xtar\_cie10\_her), antecedents mèdics familiars en la germana (xpar\_hermana), codi CIE 10 germà (xtar\_cie10\_har), antecedents mèdics familiars en l'avi matern (xpar\_ab\_mat) i codi CIE 10 avi matern (xtar\_cie10\_am).

**Taula Urgències Hospitalització:** En aquesta taula centralitza la informació referent als ingressos hospitalaris i les urgències.

Està formada per 59 pacients i 460 files.

Urgències hospitalització	
NMI	Identificació del pacient(codi anonimitzat prèviament)
Data_naix	Data de naixement.
Xsexo_id	Sexe del pacient; 1 (home), 2 (dona), 9 (desconegut)
Primera_visita_hurg	Data de la primera visita a urgències
Ultima_visita_hurg	Data de l'última visita a urgències registrada.
Tipo	tipus d'urgència U(urgència), H(hospitalització)
Cod_alta_hospitalaria	Codis d'alta hospitalària.
Des_alta_hospitalaria	Descripció de cada codi d'alta hospitalària.
Data_inicio_diag	Data que s'hospitalitza
Data_fin_diag	Data d'alta
Diagnòstic-principal	0 (és diagnòstic principal) 1 (es diagnòstic secundari)
CIE10	Codi diagnòstic en CIE10 [13]
Desc_diag_cie10	Descripció del diagnòstic segons la CIE 10 [13]
Inactiu	Data en que es cancel·la la informació entrada a la base
Anul·lat	Data en que es cancel·la la informació entrada a la base

*Taula 3 Informació que conte la base de dades Urgències/Hospitalització*

En la columna CIE10 la primera descripció és la del número CIE10 [13] i les següents files són avaluacions del GAF(Escala d'avaluació de funcionament global) [14]. Es puntua del 0 al 100 segons la gravetat de la malaltia, com més alta és la puntuació més lleu és la patologia.

**Taula Medicaments:** En aquesta taula es centralitza la informació referent a la medicació dels pacients.

Està formada per 59 pacients i 460 files.

Medicaments	
NMI	Identificació del pacient(codi anonimitzat prèviament)
Data_naix	Data de naixement.
Toma_medicamento	Pren medicament SI/NO
Fecha_inicio_medicamento	Data en que comença a prendre el medicament descrit
Fecha_final_medicamento	Data en que acaba de prendre el medicament.
Cod_farmaco	Codi de referencia de cada fàrmac que consumeix el pacient.
Desc_farmaco	Descripció de la variable Cod_farmaco
Grupoterapeutico_id	Codi de referencia del grup terapèutic de cada medicament que consumeix.
Grupoterapeutico_des	Descripció de la variable Grupoterapeutico_id

*Taula 4 Informació que conte la base de dades Medicaments*

Aquesta taula descriu la medicació presa per cada pacient, si en pren. Les variables més rellevants són si el pacient pren medicació i el grup terapèutic al qual pertany aquesta.



### 3.2 Refinat de dades (valors atípics)

Durant el procés d'extracció de les dades s'han detectat moltes incongruències d'assignació de diagnòstics, classificació, errors en les dates d'alta dels pacients, etc. que han estat corregits.

La gran quantitat de dades dificultava la revisió manual.

Al tractar-se d'un big data i no poder revisar mostres una per una, l'opció estimada és l'execució diferents computacions amb RapidMiner i Excel, contrastant valors entre diferents mostres que ens mostraran el defectes d'entrada de dades, com pot ser un diagnòstic no compatible, o una inconcurrencia amb les dates, etc.

Un cop detectades les mostres s'ha desestimant eliminar-les totes, ja que se'ns reduiria molt el volum de dades i, per tant, seria un anàlisi incomplet. En conseqüència hem optat per aquestes dues opcions segons el tipus d'error:

- Si es detecta l'error i és solventable, es modifica segons criteris tècnics o es manté perquè s'entén que està bé.
- Si es detecta l'error i no es pot adaptar s'elimina el valor per tal de no distorsionar els resultats.

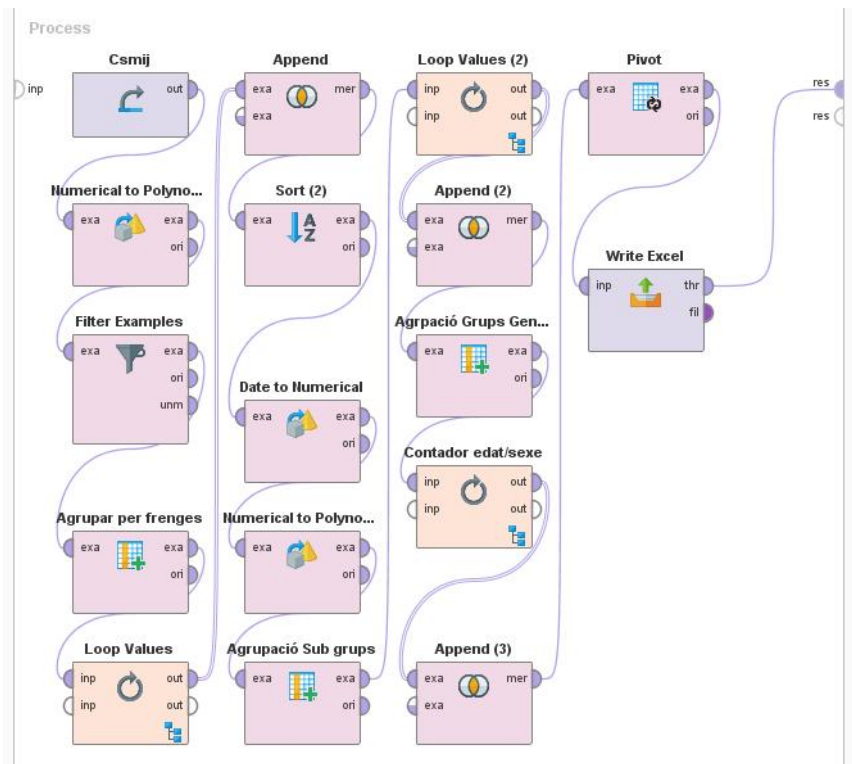
Exemples de dades atípiques trobades:

- El programa agrupa els pacients en grans grups segons la CIE9. Ens hem trobat casos en què no es defineix l'atribut CIE9 i, per tant, el programa ho classifica en el grup altres.  
Aquets pacients seran consultats per metges, i si amb els altres atributs pot classificar el pacient amb un CIE9, es classifica.  
En cas que hi ha més d'una possible definició, o no es pot esbrinar quin numero CIE9 l'hi correspon, llavors es decideix eliminar la mostra per tal d'evitar conclusions errònies.
- Un dels atributs és la data en què el pacient fa la primera visita i un segon atribut és la data en què se li assigna un diagnòstic. Normalment el mateix dia que se'ls hi fa la visita se'ls assigna un diagnòstic. Ens trobem casos en què se li assigna un diagnòstic en una data anterior a la visita, o casos en què el diagnòstic és assignat un any després de la visita.  
S'han valorat per grups i s'ha decidit, segons el cas, eliminar les mostres o mantenir-les perquè tenen una explicació lògica.

### 3.3 Agrupació de diagnòstics i correlació de malalties

L'objectiu és fer una agrupació dels diagnòstics per tal de classificar totes les CIE9[2] en 17 grups (trastorn de l'espectre autista, trastorns específics de la infància, trastorns de l'aprenentatge, trastorns de personalitat, trastorns de conducta alimentaria, trastorns d'adaptació, trastorns afectius, trastorns de conducta, trastorn d'ansietat, trastorn amb dèficit d'atenció amb hiperactivitat, trastorn de la son, tòxics, TOC, tics, esquizofrènia i altres trastorns psicòtics, esfínter, altres).

Amb el procés descrit a la imatge 7 es generà la taula 5 (classificació diagnòstic dona/home) que agrupa tots els pacients menors d'edat que van accedir al Csmij per primera vegada al 2010.



Imatge 7 Procés RapidMiner® (Classificació per sexe i per franja d'edat)

Aquest procés segueix el mateix principi d'agrupació que el procés definit anteriorment "Procés assignació grups segons CIE9 (vegeu imatge 2 de la secció 2.2)". En aquest cas s'ha afegit l'operador "Agrupar per franges d'edat" la funció del qual és separar els pacients en 4 grups segons l'edat, tal com mostra la imatge 8.

```
1 if(Edat<=5, "0 a 5", if(Edat<=11, "6 a 11", if(Edat<=15, "12 a 15", "mes 16")))
```

Imatge 8 Exemple operador "Agrupació per franja d'edat" Procés RapidMiner® (Classificació per sexe i per franja d'edat)

També s'afegeix l'operador "comptador edat/sexe", que depenent dels atributs que s'introdueixin es genera una taula com la taula 5 o la taula 6 següents.



*Imatge 9 Exemple operador "comptador edat/sexe" Procés RapidMiner® (Classificació per sexe i per franja d'edat), per generar taula de franja d'edat*

La imatge 9, es un exemple dels atributs que es seleccionen a l'operador "comptador edat/sexe" per generar una tala segons la franja d'edat.

El operador "pivot" que s'observa al final del procés de la imatge 7 permet rotar els valors indicats a la columna horitzontal, per així, poder representar-los com apareix a les taules 5 i 6.

## Classificació Dona/Home

General	Subgrup	Dona (N=473)	Home (N=663)
Trastorn de l'espectre autista	Trastorn de l'espectre autista	3	18
Trastorns específics de la infància		35	75
	Altres Infància	22	47
	Desordre d'oposició	0	0
	Problemes de relació	13	28
Trastorns de l'aprenentatge		28	50
	Altres trastorns del desenvolupament	13	20
	Desordre mixt desenvolupament	5	15
	Parla i llenguatge	4	2
	Trastorns de la comunicació	6	13
Trastorns de la personalitat		7	5
	Altres Trastorns de Personalitat	5	3
	Trastorn de Personalitat Antisocial	0	2
	Trastorn de Personalitat Límit	2	0
Trastorns de la conducta alimentària	Trastorns de la conducta alimentària	16	1
Trastorns d'adaptació		209	182
	Altres TA	21	24
	Altres reaccions d'adaptació	3	4
	Reaccions depressives	31	29
	Reacció aguda a l'estrès	19	11
	TA ansietat	27	24
	TA conducta	22	33
	TA mixt	86	57
Trastorns afectius		23	17
	Bipolar	1	0
	Depressió	10	4
	Depressió neuròtica	12	13
Trastorn de conducta		4	26
	Altres trastorns de conducta	1	9
	TC inici infància	1	11
	Trastorn mixta de la conducta i emocions	2	6
Trastorn Ansietat		54	43
	Altres trastorns d'ansietat	3	2
	Ansietat	21	19
	TA Separació	30	22
Trastorn amb dèficit d'atenció amb hiperactivitat		35	154
	Altres síndromes de hipercinèsia en la infància	5	40
	TDAH	30	114
Trastorn del son	Trastorn del son	10	6
Tòxics	Tòxics	0	2
TOC	TOC	5	8
Tics	Tics	2	6
Esquizofrènia i altres trastorns psicòtics	Esquizofrènia i altres trastorns psicòtics	1	1
Esfinter	Trastorn de control esfinter	5	17

Altres		<b>36</b>	<b>52</b>
	Grup 292	0	1
	Grup 302	1	0
	Grup 306	2	0
	Grup 316	2	0
	altres 307	29	50
	Altres problemes que poden ser objecte d'atenció clínica	2	1

*Taula 5 Classificació per diagnòstic i sexe pacients Csmij menors d'edat que accedeixen per primera vegada el 2010*

**Classificació Dona/Home:** A la taula 5 s'observa que un total de 1.136 pacients menors d'edat van accedir el Csmij per primera vegada al 2010, dels quals 473 són dones i 663 homes.

Les patologies amb més pacients destaquen els trastorns d'adaptació (391 pacients), trastorn d'ansietat (97 pacients) i trastorn amb dèficit d'atenció amb hiperactivitat (189 pacients).

## Classificació per edats

General	Subgrup	0 a 5 (N=209)	6 a 11 (N=587)	12 a 15 (N=248)	mes 16 (N=92)
Trastorn de l'espectre autista	Trastorn de l'espectre autista	15	5	1	0
Trastorns específics de la infància		35	59	13	3
	Altres Infància	20	43	5	1
	Desordre d'oposició	0	0	0	0
	Problemes de relació	15	16	8	2
Trastorns de l'aprenentatge		20	45	12	1
	Altres trastorns del desenvolupament	5	20	8	0
	Desordre mixt desenvolupament	4	13	3	0
	Parla i llenguatge	0	5	1	0
	Trastorns de la comunicació	11	7	0	1
Trastorns de la personalitat		1	0	5	6
	Altres Trastorns de Personalitat	0	0	2	6
	Trastorn de Personalitat Antisocial	1	0	1	0
	Trastorn de Personalitat Límit	0	0	2	0
Trastorns de la conducta alimentària	Trastorns de la conducta alimentària	1	0	10	6
Trastorns d'adaptació		53	193	110	35
	Altres TA	14	21	5	5
	Altres reaccions d'adaptació	0	7	0	0
	Reaccions depressives	8	21	25	6
	Reacció aguda a l'estrès	3	11	12	4
	TA ansietat	4	32	13	2
	TA conducta	7	33	13	2
	TA mixt	17	68	42	16
Trastorns afectius		0	13	14	13
	Bipolar	0	0	0	1
	Depressió	0	4	5	5
	Depressió neuròtica	0	9	9	7
Trastorn de conducta		5	17	7	1
	Altres trastorns de conducta	3	4	2	1
	TC inici infància	2	10	0	0
	Trastorn mixta de la conducta i emocions	0	3	5	0
Trastorn Ansietat		20	49	16	12
	Altres trastorns d'ansietat	0	4	1	0
	Ansietat	1	20	10	9
	TA Separació	19	25	5	3
Trastorn amb dèficit d'atenció amb hiperactivitat		33	125	26	5
	Altres síndromes de hipercinètica en la infància	11	27	6	1
	TDAH	22	98	20	4
Trastorn del son	Trastorn del son	7	9	0	0
Tòxics	Tòxics	0	0	2	0
TOC	TOC	0	8	4	1
Tics	Tics	2	5	1	0

Esquizofrènia i altres trastorns psicòtics	Esquizofrènia i altres trastorns psicòtics	0	0	0	2
Esfínter	Trastorn de control esfínter	3	13	40	2
Altres		14	46	230	5
	Grup 292	0	0	0	1
	Grup 302	0	1	0	0
	Grup 306	0	2	0	0
	Grup 316	0	2	0	0
	altres 307	12	40	23	4
	Altres problemes objecte d'atenció clínica	2	1	0	0

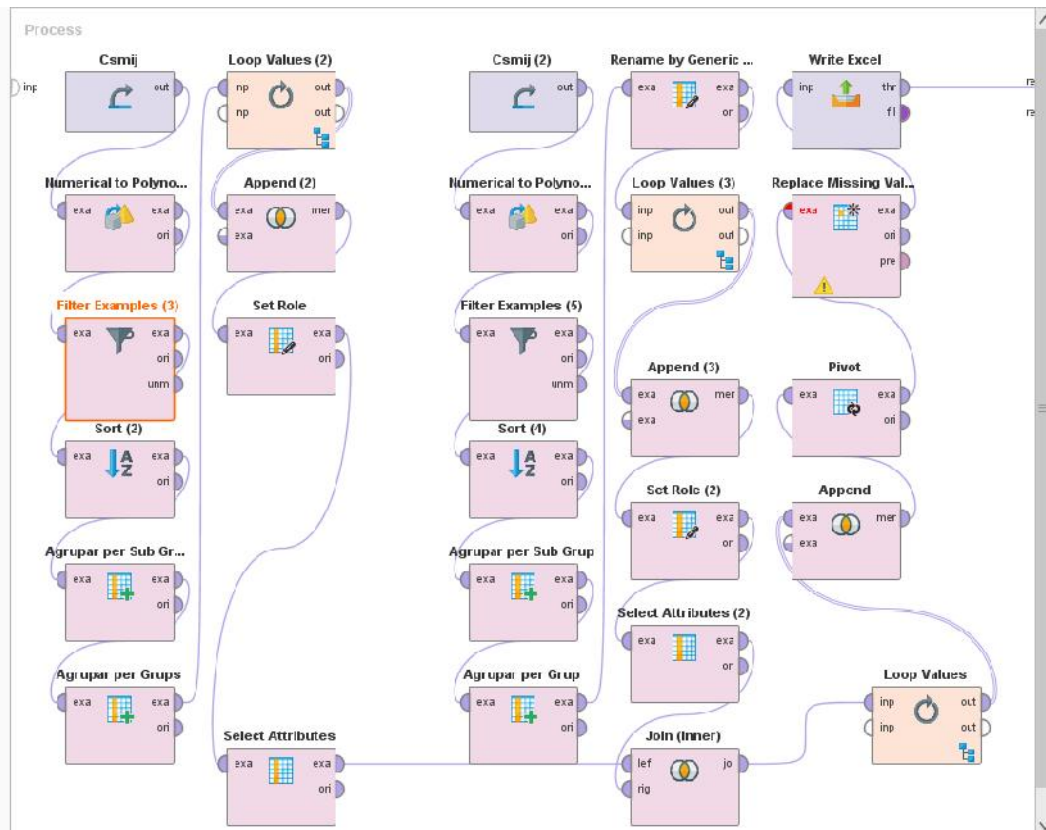
Taula 6 Classificació per diagnòstic i franja d'edat, Csmij menors d'edat que accedeixen per primera vegada el 2010

**Classificació per edats:** A la taula 6 s'observa la classificació per edats. Notem que la franja dels 6 als 11 anys (587 pacients) es el interval d'edat més gran en que s'accedeix per primera vegada al Csmij. Destaquen els diagnosticats amb dèficit d'atenció amb hiperactivitat (125 pacients) i trastorns d'adaptació (193 pacients).

Considerant la memòria del Servei català de la Salut Regió Sanitària Camp de Tarragona del 2015 [23] en que consten 4.708 pacients nous menors d'edat, s'evidencia que l'augment de pacients respecte el 2010 és significatiu.

Un dels objectius d'aquest treball és el d'analitzar les causes per les quals alguns pacients canvien de diagnòstic al cap d'un temps. Per fer un arbre de decisions dels canvis de diagnòstic construïm una taula de canvi de diagnòstic i així poder quantificar els canvis generats fins a data del 2016.

Mitjançant el procés descrit a la imatge 10 es genera la taula 7 de canvi de diagnòstic, corresponent al primer diagnòstic contrastat amb l'últim diagnòstic registrat.



Imatge 10 Procés RapidMiner® (Taula canvi de diagnòstic)

La taula canvi de diagnòstic compara els diagnòstics inicial amb els últims diagnòstics registrats. En aquest procés s'afegeixen dues bases Csmij, una per seleccionar els diagnòstics inicials i eliminar la resta (part de l'esquerra de la imatge 10) i la segona per seleccionar els diagnòstics finals i eliminar la resta (part dreta de la imatge 10).

Els operadors d'agrupació són els mateixos que s'executen en el procés assignació grups segons cie9 (capítol 2.2 Agrupació de dades i correlació de malalties).

Amb l'operador "join(inner)" ajuntem a la mateixa taula en dues columnes: el diagnòstic general inicial i el diagnòstic general final.

Els operadors "loop value" i el "Pivot" són els encarregats de comptabilitzar cada canvi de diagnòstic i de generar la taula 4.



Diagnòstic Inicial 2010		Diagnòstic Final (darrer diagnòstic)																
General	Trastorn espectre autista	Trastorn aprenentatge	TDAH	Esquizofrènia i altres trastorns psicòtics	Trastorns afectius	Trastorn Ansietat	Trastorn adaptació	Trastorn conducta	Trastorns conducta alimentària	TOC	Trastorns personalitat	Trastorn del son	Tics	Toxics específics infància	Estínter	Altres problemes objecte d'atenció clínica	Total	
Trastorn de l'espectre autista	19	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	2
Trastorns de l'aprenentatge	4	67	2	0	0	0	1	0	0	0	0	0	0	0	0	2	2	11
TDAH	0	4	171	0	0	1	3	4	0	0	2	0	0	1	0	2	1	18
Esquizofrènia i altres trastorns psicòtics	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
Trastorns afectius	1	0	0	0	31	0	3	0	0	0	4	0	0	0	0	0	1	9
Trastorn Ansietat	0	4	3	0	3	75	3	0	0	0	2	0	0	4	0	2	1	22
Trastorns d'adaptació	1	7	8	2	8	9	324	5	2	1	7	0	0	4	0	8	5	67
Trastorn de conducta	0	0	2	0	0	1	0	25	0	0	0	0	0	0	0	1	1	5
Trastorns de la conducta alimentària	0	0	0	0	0	0	0	0	16	0	1	0	0	0	0	0	0	1
TOC	1	0	0	0	0	0	0	0	0	10	1	0	0	0	0	1	0	3
Trastorns de la personalitat	0	0	1	0	0	0	0	0	0	0	9	0	0	0	0	2	0	3
Trastorn del son	0	0	0	0	1	0	0	0	0	0	1	13	0	0	0	0	0	3
Tics	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0
Toxics	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Trastorns específics infància	1	1	1	0	3	1	3	2	0	0	0	0	0	91	0	2	5	19
Estínter	1	0	0	0	0	0	1	0	0	0	0	0	0	1	17	0	2	5
Altres problemes objecte d'atenció clínica	2	3	2	1	1	3	4	1	0	0	2	0	0	2	1	63	0	22
	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	2	2

Taula 7 Taula de canvi de diagnòstic, contrastant el diagnòstic inicial i l'últim diagnòstic assignat

La taula 4 quantifica el canvi de patologia que experimenten els pacients. Aquesta taula correspon als pacients que van accedir a les instal·lacions del Csmij el 2010, on a les files s'indica la primera patologia que se'ls va assignar i les columnes corresponen a l'última patologia que se'ls va diagnosticar.

Les caselles blaves (diagonal) corresponen al número de pacients que mantenen la primera patologia diagnosticada. La resta de caselles quantifiquen els canvis de diagnòstic.

La darrera columna suma el total de pacients que han modificat la patologia inicial. S'observa que els trastorns d'adaptació són els de diagnòstic més movibles, ja que un total de 67 han canviat de diagnòstic. També destaquen els trastorns d'ansietat (22) i els TDAH "trastorn amb dèficit d'atenció amb hiperactivitat" (18).

Revisada la taula amb els professionals del centre, sorprèn que els pacients diagnosticats amb trastorn d'adaptació siguin tant nombrosos. Des del 2010 fins la data en què es va recopilar la informació (febrer 2017) només 67 han modificat el diagnòstic, ja que el trastorn d'adaptació és un diagnòstic temporal i al cap d'un temps aquest deriva cap a un altre diagnòstic.

### 3.4 Selecció d'atributs

L'objectiu de la selecció d'atributs és reduir el número d'atributs a entre 10 i 15, per tal de poder generar arbres de decisions útils i senzills d'utilitzar.

Per la taula de psicopatologia, amb 151 atributs, s'executa un primer filtre per tal de reduir els menys significatius. Amb Excel s'han descartat tots aquells que tenen un nombre de 1 (Sí) inferior al 2% (patologies poc observades), passant així de 151 a 45 atributs.

El procés de selecció d'atributs té una fase inicial en què es generen diferents rànquing mitjançant els operadors de pesat descrits a la secció 2.3 (selecció d'atributs). Aquest rànquing permeten reduir a 20 els atributs que matemàticament tenen més influència en el subjecte a analitzar.

En una segona fase es 20 atributs es reduiran a entre 10-15, segons criteris mèdics dels professionals, que descarten aquells que poden generar confusió, o no tenen interpretació lògica mèdica, o són condicions positives com per exemple tenir un llenguatge adequat.

El procés de selecció d'atributs es realitzarà per cada arbre de decisions que es vulgui generar. En la següent taula es mostren els 16 tipus d'estudis que es vol realitzar degut al seu interès mèdic predictiu. Com apareix a la taula, cada estudi comporta l'elaboració d'un rànquing dels atributs més adients, la inducció d'un arbre de decisions i la validació d'aquets arbres amb mecanismes de cross-validation.

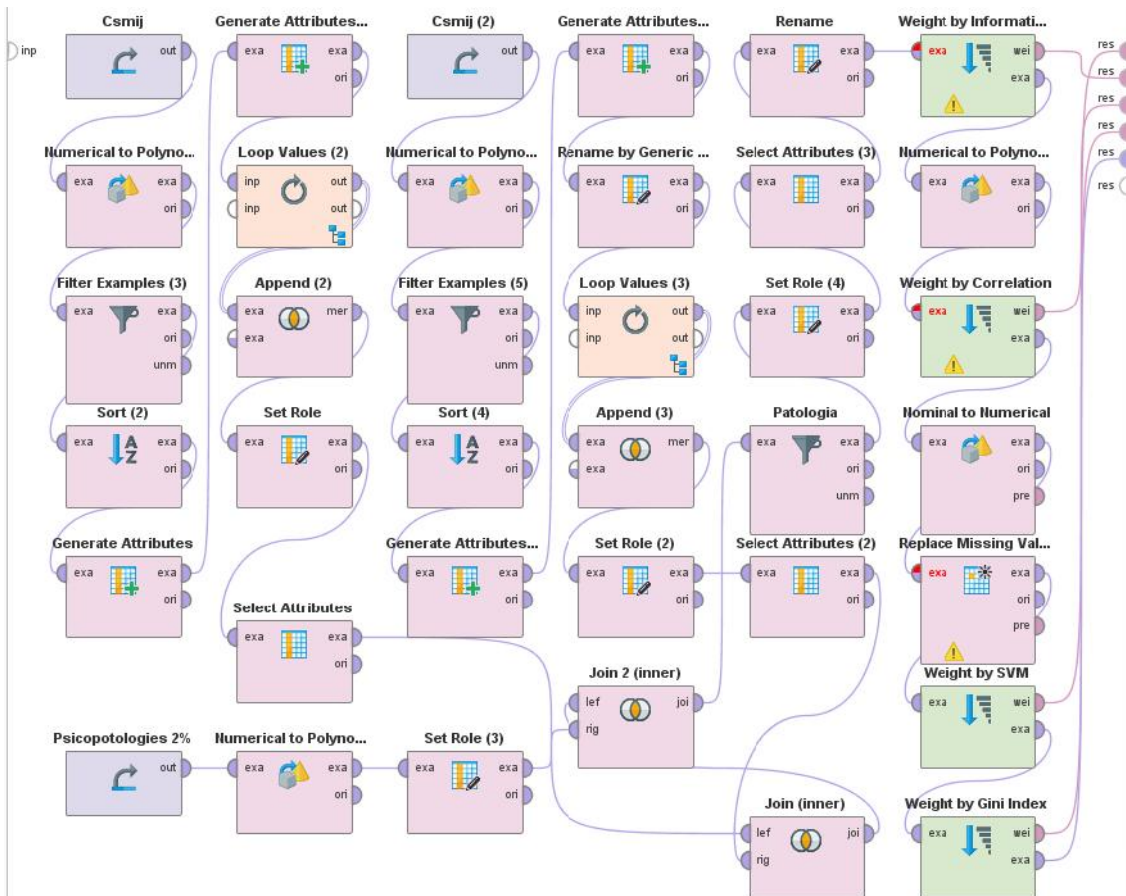
		Atributs			
		Psicopatologia	Anamnesis + Valoració social	Medicaments	TOTS
Canvi diagnòstic	Trastorn d'adaptació	Rànquing	Rànquing	Rànquing	Rànquing
		Arbre de decisió	Arbre de decisió	Arbre de decisió	Arbre de decisió
		Cross-Validation	Cross-Validation	Cross-Validation	Cross-Validation
	Trastorn ansietat	Rànquing	Rànquing	Rànquing	Rànquing
		Arbre de decisió	Arbre de decisió	Arbre de decisió	Arbre de decisió
		Cross-Validation	Cross-Validation	Cross-Validation	Cross-Validation
	Trastorn amb dèficit d'atenció amb hiperactivitat	Rànquing	Rànquing	Rànquing	Rànquing
		Arbre de decisió	Arbre de decisió	Arbre de decisió	Arbre de decisió
		Cross-Validation	Cross-Validation	Cross-Validation	Cross-Validation
GAF	Rànquing	Rànquing	Rànquing	Rànquing	
	Arbre de decisió	Arbre de decisió	Arbre de decisió	Arbre de decisió	
	Cross-Validation	Cross-Validation	Cross-Validation	Cross-Validation	

Taula 8 documentació a generar dels processos RapidMiner®

Els processos RapidMiner® utilitzat per calcular els rànquings d'influència de les variables sobre el subjecte es mostren a continuació:

- **Rànquing 1:** ordena els atributs clínics sobre psicopatologies en funció de la seva rellevància a l'hora de determinar el canvi de diagnòstic entre els anys 2010 i 2016.

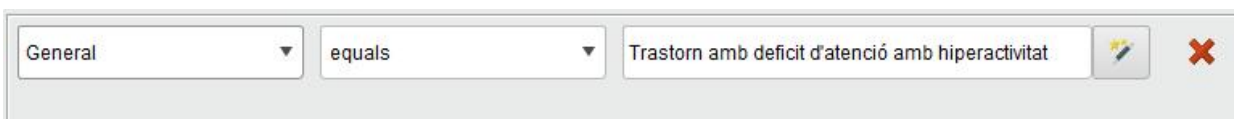
Aquest rànquing té l'objectiu de calcular quines variables de psicopatologia tenen més pes en la presa de decisió d'un canvi de diagnòstic.



Imatge 11 Procés RapidMiner® (Rànquing 1 canvi de diagnòstic base psicopatologia)

El procés de la imatge 11 és basa en el procés utilitzat per generar la taula de canvi de diagnòstic (vegeu imatge 10). Un cop identificats els diagnòstics inicial i final, s'hi afegeixen les variables de la base de psicopatologia mitjançant un operador join(inner).

També s'afegeix l'operador "patologia" que l'utilitzarem per definir sobre quin grup de diagnòstic inicial s'ha de calcular el rànquing (observeu el mecanisme de selecció en la imatge 12).



Imatge 12 Exemple operador "patologia" del Procés RapidMiner® (Rànquing 1 canvi de diagnòstic base psicopatologia)

El model d'anàlisi de la imatge 11 es complementa amb els operadors de càlcul de rànquing d'atributs segons la capacitat d'influir sobre el subjecte. En aquest cas els operadors escollits són: "weight by information gain", "weight by correlation", "weight by SVM" i "weight by gini index".

- **Rànquing 2:** ordena els atributs clínics sobre anamnesi i valoració social en funció de la seva capacitat de pronosticar un canvi de diagnòstic entre els anys 2010 i 2016.

El procés, RapidMiner® per a calcular rànquing 2 és similar al que apareix a la imatge 11, a excepció que la base de dades de psicopatologia i el operador que la tracten es substituïda per les bases de anamnesi i valoració social juntament amb els operadors necessaris.

- **Rànquing 3:** aquest rànquing té l'objectiu de calcular quins grup terapèutic dels medicaments consumits pels pacients tenen més pes en la presa de decisió d'un canvi de diagnòstic.

El procés generador d'aquest rànquing és similar al mostrat a la imatge 11, afegint la base Medicaments amb els operadors corresponents i sense la resta de bases de dades.

- **Rànquing 4:** aquest rànquing té l'objectiu de calcular quines variables de totes les que disposem tenen més pes en la presa de decisió d'un canvi de diagnòstic. El procediment de càlcul queda representat en la imatge 13.



Imatge 13 Procés RapidMiner® (Rànquing 4 canvi de diagnòstic totes les bases)

Aquest procés realitza un càlcul de pesat amb totes les bases juntes (psicopatologia, anamnesis, valoració social i medicaments) per detectar quines variables tenen més dependència lineal de tot el conjunt de dades.

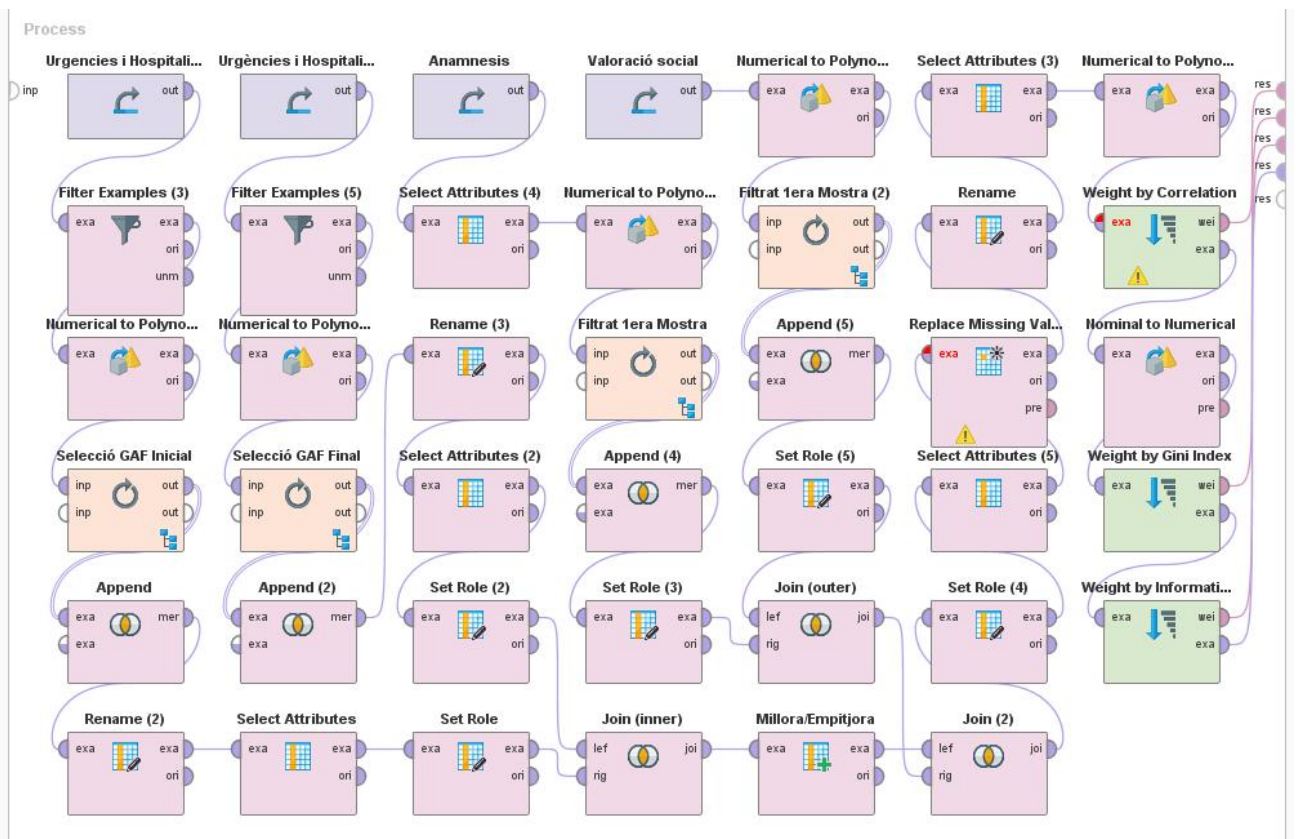
- **Rànquing 5:** aquest rànquing té l'objectiu de calcular quines variables de psicopatologia tenen més pes en el millorament o empitjorament del pacient.

En aquest procés no s'utilitzarà la base Csmij sinó la base Urgències/hospitalització, on hi ha els valors de puntuació GAF que farem servir com a mesura de la qualitat de salut del pacient.

Aquest procés s'inicia en un filtrat dels valors per tal de disposar només de les puntuacions GAF al llarg del temps (ja que en la mateixa columna es troba més informació a part del GAF). Es selecciona el valor GAF tant a l'inici (any 2010) com al final (any 2016) per tal de comparar-les i indicar si el pacient millora, empitjora o manté la gravetat.

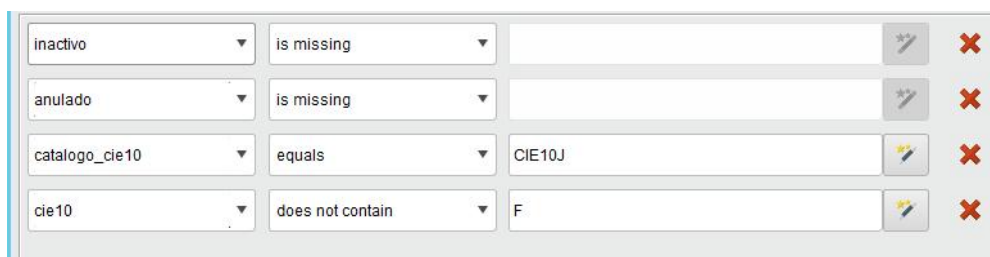
En el procés rànquing 5 GAF s'hi afegeix la base de dades de psicopatologia.

- **Rànquing 6:** aquest rànquing té l'objectiu de calcular quines variables d'anamnesis com de valoració social tenen més pes en la millora o empitjorament del pacient.



Imatge 14 Procés RapidMiner® (Rànquing 6 GAF anamnesis i valoració social)

El procés de la imatge 14 correspon al procés rànkning. Que s'inicia a partir de dos bases de Urgències/hospitalització, que en les dos disposen dels mateixos operadors de càlcul. Inicialment es fa un filtrat de valors, per tal de deixar en disposició de càlcul, només les puntuacions GAF. Per aconseguir-ho apliquem l'operador "Filter exemple" sobre les dades, com mostra la imatge 15.



Imatge 15 Exemple operador "Filter Example" del Procés RapidMiner® (Rànquing 6 GAF anamnesis i valoració social)

A partir de la iteració "selecció GAF Inicial" i "selecció GAF Final" respectivament es generen una columna amb els valors inicials del GAF i una altra amb els valors finals.

S'ajunten los dues columnes en una mateixa taula i mitjançant l'operador "millora/empitjora" mostrat a la imatge 16 i es defineix el progrés de cada pacient.

```
1 if([GAF Primera Visita]>[GAF Ultima Visita],"Empitjora",if([GAF Primera Visita]<[GAF Ultima Visita],"Millora","No modificat"))
```

*Imatge 16 Exemple operador "millora/empitjora" del Procés RapidMiner® (Rànquing 6 GAF anamnesis i valoració social)*

L'operador millora/empitjora compara la columna inicial i final i genera una nova columna que segons el cas apareix Millor, Empitjora, No modifica.

També s'hi afegixen les bases de dades d'anamnesi com de valoració social amb l'operador "join".

I finalment es disposa dels operadors de càlcul de rànquing d'atributs segons la capacitat d'influir sobre el subjecte, que en aquest cas els operadors escollits són: "weight by information gain", "weight by correlation", "weight by gini index".

- **Rànquing 7:** aquest rànquing té l'objectiu de calcular quins grup terapèutic dels medicaments consumits pels pacients tenen més pes en la millora o empitjorament del pacient.

El procés, RapidMiner® rànquing 7 GAF i dades de medicaments, és similar als dos processos anteriors (rànquing 5 i 6), afegint la base Medicaments amb els operadors corresponents i sense la resta de bases de dades.

**Rànquing 8:** aquest rànquing té l'objectiu de calcular quines variables de totes les que disposem tenen més pes en la predicció de la millora o empitjorament del pacient.

El procés es el mateix que els anteriors però afegint totes les bases de dades.



<b>Rànquings</b>	<b>Variables</b>	<b>Objectius</b>
<b>1</b>	Psicopatologia	Predir els canvis de diagnòstic entre 2010 i 2016
<b>2</b>	Valoració social i Anamnesis	Predir els canvis de diagnòstic entre 2010 i 2016
<b>3</b>	Medicaments	Predir els canvis de diagnòstic entre 2010 i 2016
<b>4</b>	Psicopatologia, Valoració Social, Anamnesis , i Medicaments	Predir els canvis de diagnòstic entre 2010 i 2016
<b>5</b>	Psicopatologia	Predir la millora o empitjorament d'un pacient entre el 2010 i 2016
<b>6</b>	Valoració social i Anamnesis	Predir la millora o empitjorament d'un pacient entre el 2010 i 2016
<b>7</b>	Medicaments	Predir la millora o empitjorament d'un pacient entre el 2010 i 2016
<b>8</b>	Psicopatologia, Valoració Social, Anamnesis i Medicaments	Predir la millora o empitjorament d'un pacient entre el 2010 i 2016

*Taula 9 Taula resum dels rànquings*

El resultat de l'aplicació dels processos del rànquing amb el RapidMiner® descrits anteriorment i de la posterior validació per part dels professionals mèdics donà els següents conjunts d'atributs rellevants, depenent del tipus de predicció que es vulgui realitzar:

- **Estudi 1 : Trastorn d'adaptació i base psicopatologia:** alentiment, alteracions quantitatives tipus rebuig selectiu, alteració dels hàbits, baixa autoestima, explosivitat, inquietud, insomni d'inici, irritabilitat, labilitat emocional, pobre, por preocupacions, pèrdua d'interès, rumiació, tristesa.
- **Estudi 2 : Trastorn d'adaptació i base anamnesis i valoració social:** antecedents d'ingrés hospitalari previ, antecedents de tractament farmacològic previ, antecedents mèdics familiars en el germà, antecedents mèdics familiars en el pare, consum de tòxics en el passat, dificultats en l'adquisició d'aprenentatges, dificultats en la separació dels pares, manté relacions socials inadequades, no regula el temps lliure, part a pre-terme, setmanes de gestació, situació legal del menor, tipus de condicions de la vivenda de la persona, tipus de vivenda de la persona, vivenda protegida.
- **Estudi 3 : Trastorn d'adaptació i base medicaments:** pren medicament i els que pertanyen als grups terapèutics; antiepilèptics, agents simpaticomimètics d'acció ce, anilidades, derivats de la benzodiazepina, derivats dels àcids grassos, dizaxepinas, oxazepinas, tizaxepinas y oxe, inhibidors selectius de la recaptació, altres antidepressius, altres antiepilèptics, altres antipsicòtics.

- **Estudi 4 : Trastorn d'adaptació i totes les base:** alteracions quantitatives tipus afartament, alteració dels hàbits, antecedents d'ingrés hospitalari previ, baixa autoestima, bones condicions de la vivenda, consum de tòxics en el passat, explotivitat, manté relacions socials inadequades, mesos d'inici de la deambulació, mesos en que la persona controla l'esfínter anal, mesos en que la persona va dir les primeres paraules, pren medicaments, pèrdua d'interès, rumiació, setmanes de gestació, tipus de condicions de la vivenda de la persona.
- **Estudi 5 : Trastorn d'ansietat i base psicopatologia:** alentiment, alteracions quantitatives tipus rebuig selectiu, alteracions quantitatives tipus afartaments, alteració dels hàbits, augments psicomotricitat, discapacitat intel·lectual, disfòria, dislàlies, impulsions, insomni de manteniment, irritabilitat, preocupacions, pèrdua d'interès, rumiació, tristesa.
- **Estudi 6 : Trastorn d'ansietat i base anamnesis i valoració social:** adaptació amb els professors adequada, antecedents de tractament psicològic previ, antecedents mèdics familiars en el germà, antecedents mèdics familiars en el pare, antecedents mèdics familiars en la mare, dificultats contínues en les relacions familiars, dificultats continues en les relacions familiars (xrf), dificultats en l'adaptació amb els professors, dificultats en l'adquisició d'aprenentatges, dificultats en la integració amb els iguals, dificultats en la separació dels pares, no regula el temps lliure, setmanes de gestació, situació legal del menor, tipus de condicions de la vivenda de la persona.
- **Estudi 7 : Trastorn d'ansietat i base medicaments:** pren medicament i els que pertanyen als grups terapèutics; agents simpaticomimètics d'acció cen, derivats de la benzodiazepina, fàrmacs relacionats amb la benzodiazepina, hipnòtics i sedants, inhibidors no selectius de la recaptació, inhibidors selectius de la recaptació, altres antidepressius, altres antipsicòtics.
- **Estudi 8 : Trastorn d'ansietat i totes les base:** alteracions quantitatives tipus afartaments, antecedents de tractament farmacològic previ, antecedents de tractament psicològic previ, antecedents mèdics familiars amb el germà, dificultats continues en les relacions familiars (xrf), dificultats en l'adaptació amb els professors, dificultats en l'adquisició d'aprenentatges, dificultats en la integració amb els iguals, disfòria, dislàlies, mesos d'inici de la deambulació, mesos en que la persona controla l'esfínter anal, mesos en que la persona va dir les primeres paraules, preocupacions, setmanes de gestació, tipus de condicions de la vivenda.
- **Estudi 9 : TDAH i base psicopatologia:** alentiment, alteracions quantitatives tipus rebuig selectiu, apatia, augment psicomotricitat, auto agressivitat, desesperança, disfòria, explotivitat, hetero agressivitat, impulsions, inquietud, insomni d'inici, irritabilitat, malsons.

- **Estudi 10 : TDAH i base anamnesis i valoració social:** adaptació amb els professors adequada, antecedents d'ingrés hospitalari previ, antecedents de tractament farmacològic previ, antecedents mèdics familiars en el pare, consum de tòxics en el passat, dificultats contínues en les relacions familiars, dificultats en l'adquisició d'aprenentatges, manté relacions familiars adequades, manté relacions socials inadequades, part a pre-terme, participa en una o més activitats de temps lliure, sense aficions, situació legal del menor, tipus de vivenda de la persona.
- **Estudi 11 : TDAH i base medicaments:** pren medicament i els que pertanyen als grups terapèutics; derivats de la carboxamina, derivats dels àcids grassos, derivats dels àcids propionics, inhibidors selectius de la recaptació, altres antidepressius, altres antiepilèptics, altres antipsicòtics.
- **Estudi 12 : TDAH i totes les base:** antecedents d'ingrés hospitalari previ, antecedents de tractament farmacològic previ, auto agressivitat, consum de tòxics actual, desesperança, dificultats contínues en les relacions familiars, disfòria, hetero agressivitat, insomni de manteniment, malsons, mesos d'inici de la deambulació, mesos en que la persona controla l'esfínter anal, mesos en que la persona va dir les primeres paraules, sense aficions, situació legal del menor.
- **Estudi 13 : GAF i base psicopatologia:** alentiment, insomni d'inici, inquietud, alteracions quantitatives tipus rebuig selectiu, pobre, dislàlies, disfòria, impulsions, desesperança, auto agressivitat, hetero agressivitat, malsons, no espontani, sentiments de buit.
- **Estudi 14 : GAF i base anamnesis i valoració social:** situació legal del menor, consum de tòxics actual, sense aficions, dificultats contínues en les relacions familiars, vivenda, tipus de vivenda de la persona, tipus de condicions de la vivenda de la persona, antecedents de tractament psicològic previ, existència de mesura penal alternativa, nul·les relacions familiars, mesos d'inici de la deambulació, rebuig a noves relacions, mesos en que la persona va dir les primeres paraules, mesos en que la persona controla l'esfínter anal, antecedents del tractament farmacològic previ, consum de tòxics en el passat.
- **Estudi 15 : GAF i base medicaments:** pren medicament i els que pertanyen als grups terapèutics; combinacions de penicil·lines, derivats dels àcids grassos, fàrmacs relacionats amb la benzodiazepina, derivats de la carboxamida, fenotiazines amb cadena lateral alifàtica, derivats de la butirefona, inhibidors de la bomba de protons, altres antibacterians, agents simpatitzants de l'acció cen, "diazepinas, oxazepinas, tiazepinas y oxe", inhibidors de la HMG-CoA reductasa en c, formulacions de sals de rehidratació, altres antidepressius, antipsicòtics.
- **Estudi 16 : GAF i totes les base:** situació legal del menor, dificultats contínues en les relacions familiars, tipus de vivenda de la persona, tipus

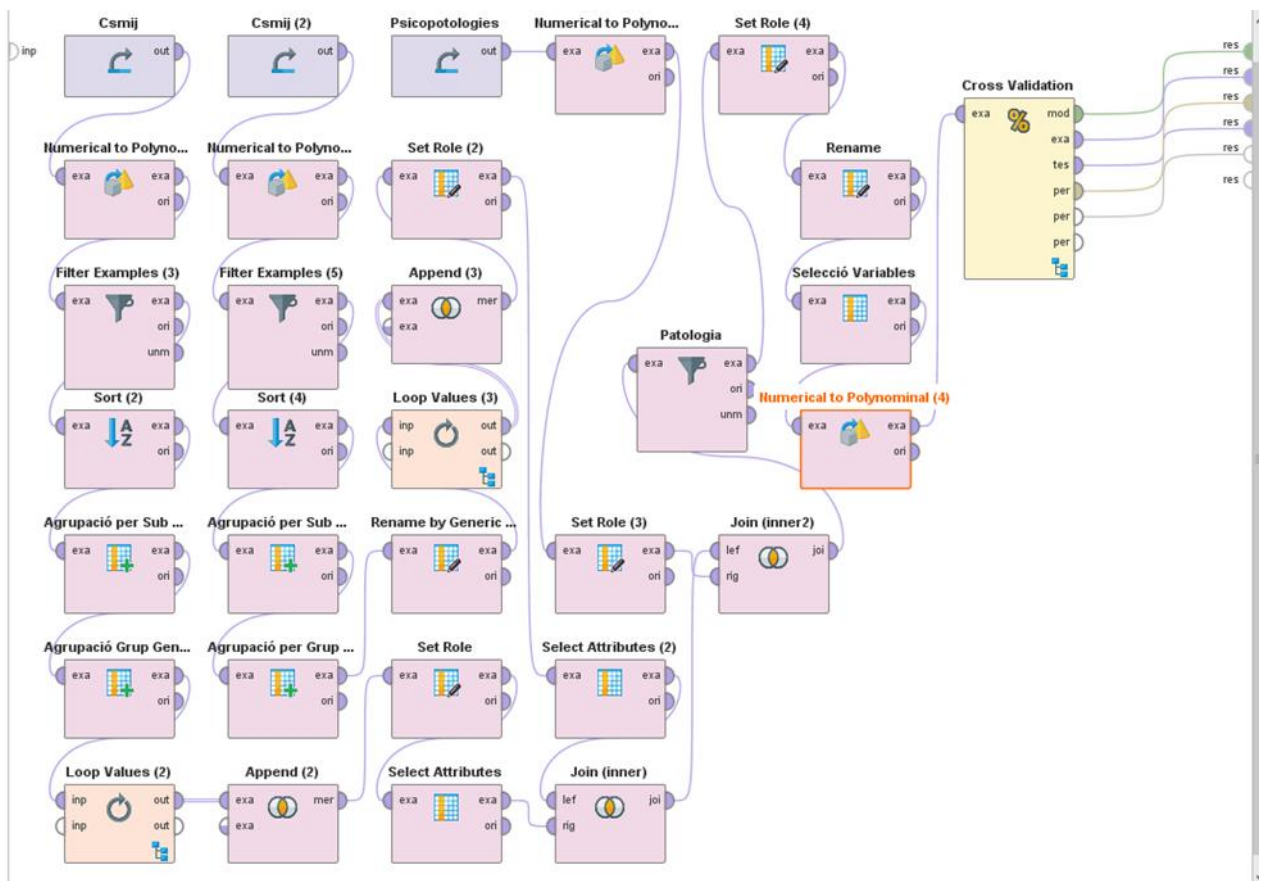
de vivenda de la persona, manté relacions familiars adequades, tipus de condicions de la vivenda de la persona, antecedents de tractament psicològic previ, mesos d'inici de la deambulació, mesos que la persona va dir les primeres paraules, heteroagressivitat, insomni d'inici, dislàlies, no espontani, alteracions quantitatives tipus rebuig selectiu, existència de mesura penal alternativa, nul·les relacions familiars, part a post-terme.

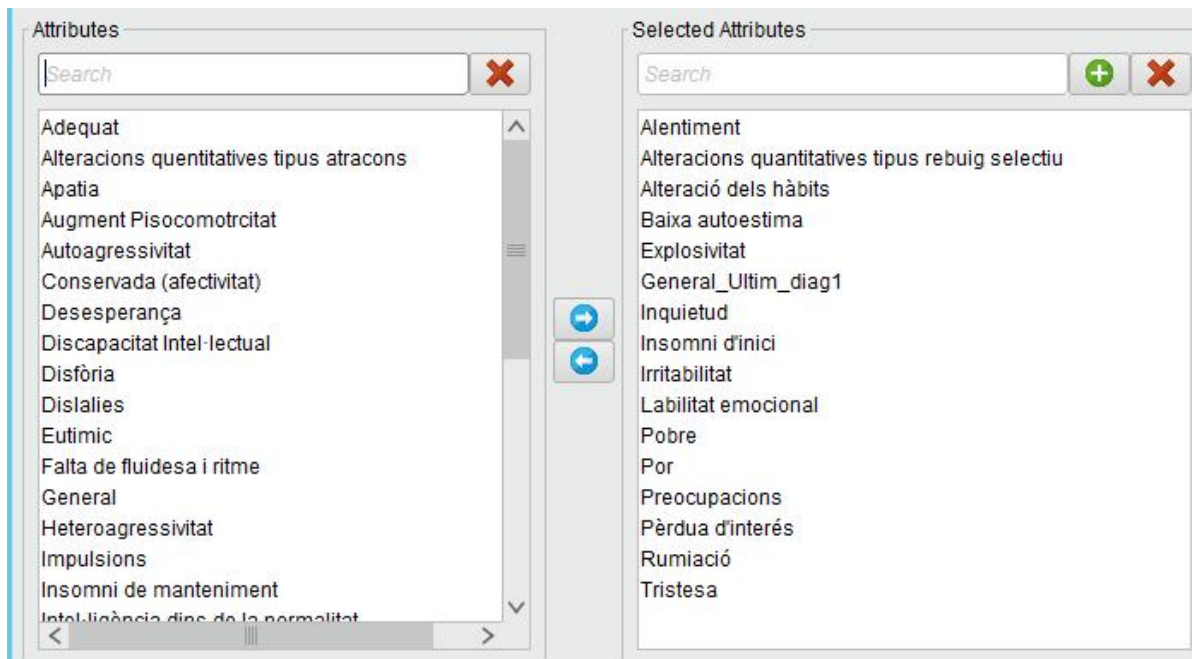
### 3.5 Construcció de models predictius

El model elegit per l'aplicació de l'anàlisi predictiu és l'arbre de decisions. Un cop seleccionats els atributs, amb aquests es genera l'arbre de presa de decisions. Tot seguit es valora matemàticament (Cross-Validations) i medicament la qualitat desitjada.

Si aquest arbre es considera que té suficient qualitat es dona com a definitiu, en cas contrari es modifiquen els atributs per tal de millorar-lo fins que assoleixi la qualitat considerada.

El processos utilitzats (part superior de la imatge 18) són rànquings, on a més d'aplicar operadors de càlcul de pesat, s'aplica un operador de selecció variables (part inferior de la imatge 18) i de les variables escollides en els passos anteriors secció 3.4 i s'executa el procés de "Cross Validation", on en el mateix procés ja es defineix el operador "Decision Tree" com a model predictiu. Prèviament s'apliquen operadors d'assignació de tipus de dades segons si les considerem quantitatives o nominals.





*Imatge 18 Exemple operador “Numerical to polinomial” del Procés RapidMiner® (Arbre de decisions 1, trastorn d’adaptació i base psicopatologia)*

En l’operador “numerical to polinomial” seleccionen les variables que poden ser polinomials com per exemple alentiment que si es un 1 significa SI. Els valors que representen una quantitat, com poden ser kg, mesos, setmanes, etc. no es transformaran a polinomials, ja que son variables quantitatives.

### 3.6 Valoració del model

Per obtenir un model predictiu de qualitat cal valorar aquest tant estadísticament com mèdicament (qualitat clínica).

La valoració estadística consisteix en calcular quin percentatge d’exactitud té, aquest càlcul es farà amb l’operador Cross Validation de RapidMiner®.

És aquest un operador amb dos subprocessos, el subprocés “training” i el subprocés “testing”. El subprocés “training” és utilitzat per generar el model i el subprocés “testing” per valorar l’eficàcia d’aquest.

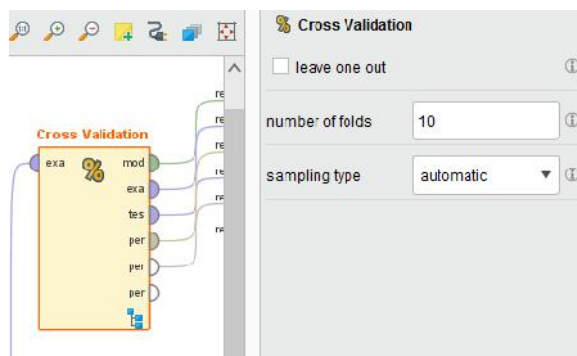
La informació d’entrada a l’operador es divideix en “k” particions de la mateixa mida (definint el número de particions al paràmetre “number of folds”), on una de les particions K és retinguda per al sub procés “testing” i la resta de k-1 particions són utilitzades per a la generació del model “training”.

El procés es repeteix K vegades de tal manera que totes les particions K són utilitzades almenys un cop com a test.

Un cop generat l'arbre i obtenint el percentatge d'exactitud que ens ofereix, es farà la valoració mèdica, efectuada pels professionals de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM).

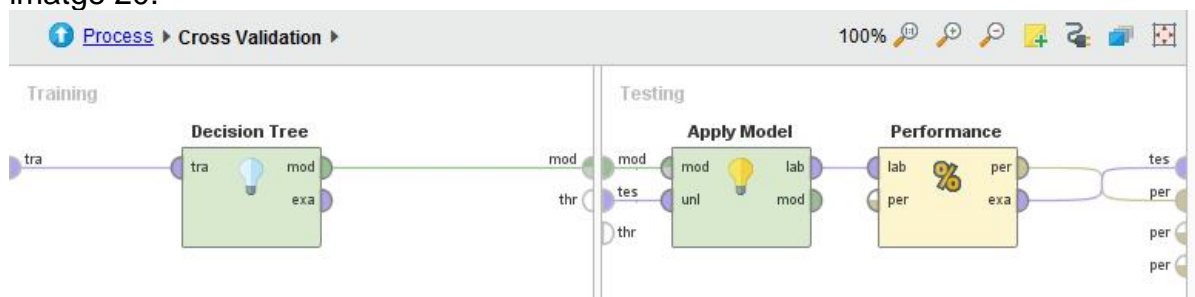
Si els professionals ho consideren oportú s'afegeix (o es treuen) els atributs que poden fer una millora del model i es re-calcula novament amb l'operador "Cross-Validation".

S'obté l'exactitud del nou model comparant si el percentatge millora o empitjora. I es decideix novament amb els professionals si és un bon model o si cal afegeix (o extreure) atributs. Es repeteix el procés fins trobar el model que ens ofereix més qualitat mèdica i estadística.



Imatge 19 Exemple operador "Cross Validation"

Definim l'operador "Cross Validation" en que faci 10 particions (10-fold cross validation) i l'assaig sigui aleatori. El procés queda descrit en la imatge 20.



Imatge 20 Exemple operador "Cross Validation"

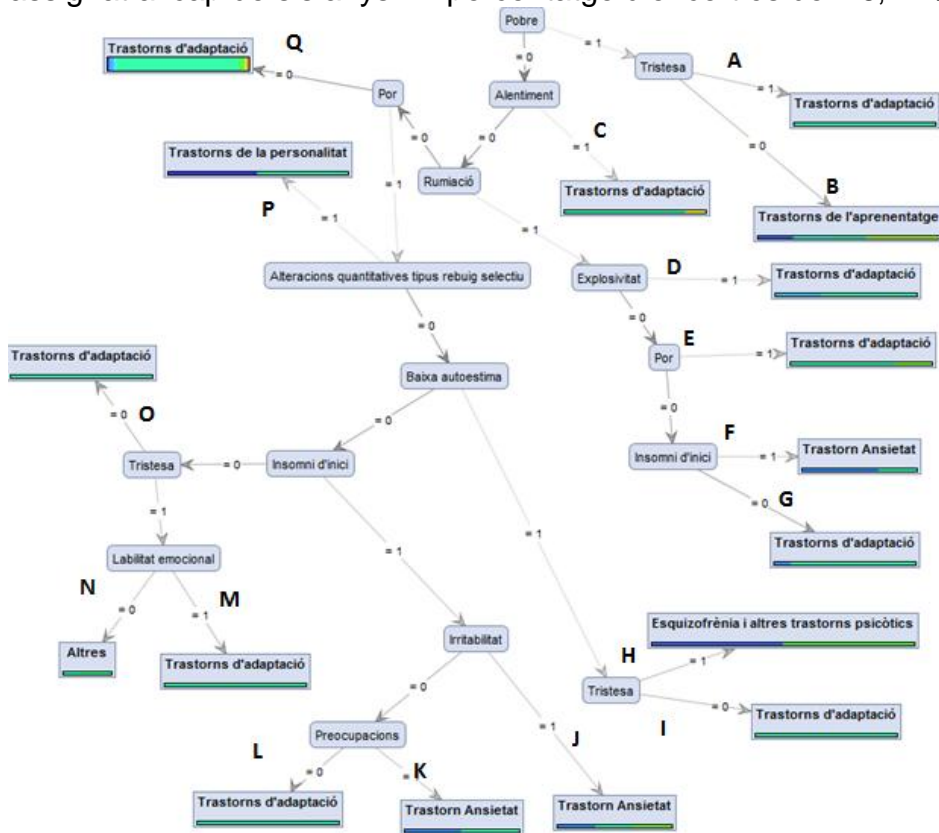
En el procés "training" de l'operador "Cross-Validation" s'afegeix el model predictiu escollit, que en el nostre cas es, el "decision tree". En el procés "testing" s'afegeix l'operador "Apply Model", que executa el model introduït i l'operador "Performance" que ens dona la precisió del model.

### 3.7 Resultats

A continuació es mostren els arbres de decisions (AD) generats i el percentatge d'exactitud que ens ofereix cadascun d'aquestes:

- **Model predictiu 1:** AD sobre el trastorn d'adaptació i base psicopatologia.

Es genera un arbre de decisions (imatge 21) dels pacients que inicialment són diagnosticats amb un trastorn d'adaptació per predir, a partir de característiques psicopatològiques, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 76,71 %.



Imatge 21 Arbre de decisions 1, trastorn d'adaptació i base psicopatologia

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu en les variables pobre i tristesa): Trastorns d'adaptació=5

**B** (són el tipus de pacients que han donat positiu en la variable pobre i no ha donat positiu a tristesa): Esquizofrènia i altres trastorns psicòtics=1, Trastorn de l'espectre autista=1, Trastorns d'adaptació=1, Trastorns de l'aprenentatge=2

**C**: Trastorns d'adaptació=6, TOC=1

**D**: Trastorns afectius=1, Trastorns d'adaptació=2

**E**: Trastorns d'adaptació=3, Trastorns de la conducta alimentaria=1

**F**: Trastorn Ansietat=2, Trastorns d'adaptació=1

**G**: Trastorn Ansietat=1, Trastorns d'adaptació=7

**H**: Esquizofrènia i altres trastorns psicòtics=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

**I**: Trastorns d'adaptació=2

**J**: Trastorn Ansietat=1, Trastorns d'adaptació=1, Trastorns de la conducta alimentaria=1

**K**: Trastorn Ansietat=1, Trastorns d'adaptació=1

**L**: Trastorns d'adaptació=3

**M**: Trastorns d'adaptació=3

**N**: Trastorns d'adaptació=1, Altres=2

**O**: Trastorns d'adaptació=8

**P**: Trastorns de la personalitat=1, Trastorns d'adaptació=1

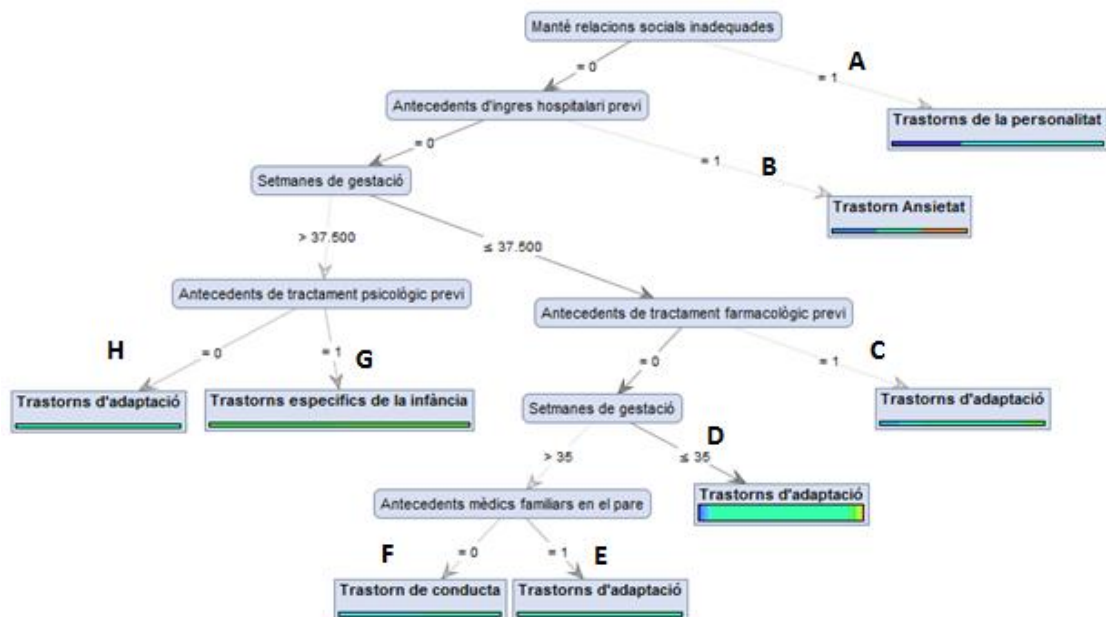
**Q:** Trastorns de la personalitat=6, Trastorn Ansietat=4, Trastorns afectius=7, Trastorn de conducta=5, Trastorns d'adaptació=265, Altres=6, Trastorns específics de la infància=4, Trastorn amb dèficit d'atenció amb hiperactivitat=7, Trastorns de l'aprenentatge=5, Altres problemes que poden ser objecte d'atenció clínica=5

La majoria dels pacients amb un trastorn d'adaptació inicial que ni són pobres, ni tenen alentiment, ni rumiació, ni por, l'últim diagnòstic assignat continua sent el trastorn d'adaptació al cap de sis anys.

S'observa que els pacients que tenen insomni d'inici deriven cap un trastorn d'ansietat. Com són el punts F, K i J.

- **Model predictiu 2:** AD sobre el trastorn d'adaptació i bases valoració social i anamnesis.

Es genera un arbre de decisions (imatge 22) dels pacients que inicialment són diagnosticats amb un trastorn d'adaptació per predir, a partir d'informació social i d'anamnesis, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 79,62 %.



Imatge 22 Arbre de decisions 2, trastorn d'adaptació i bases valoració social i anamnesis.

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que mantenen relacions inadequades): Trastorns de la personalitat=1, Trastorn de conducta=1, Trastorn de l'espectre autista=1



**B** (són el tipus de pacients que no mantenen relacions inadequades i tenen antecedents d'ingrés hospitalari): Trastorn Ansietat=1, Trastorns d'adaptació=1, Altres problemes que poden ser objecte d'atenció clínica=1

**C:** Trastorns afectius=1, Trastorn de conducta=1, Trastorns d'adaptació=5, Trastorn amb dèficit d'atenció amb hiperactivitat=1

**D:** Trastorns de la personalitat=6, Esquizofrènia i altres trastorns psicòtics=2, Trastorn Ansietat=5, Trastorns afectius=6, Trastorn de conducta=2, Trastorns d'adaptació=241, Altres=8, Trastorns específics de la infància=3, Trastorn amb dèficit d'atenció amb hiperactivitat=6, Trastorns de la conducta alimentaria=1, Trastorns de l'aprenentatge=7, TOC=1, Altres problemes que poden ser objecte d'atenció clínica=2

**E:** Trastorns d'adaptació=4

**F:** Trastorn de conducta=1, Trastorns d'adaptació=1

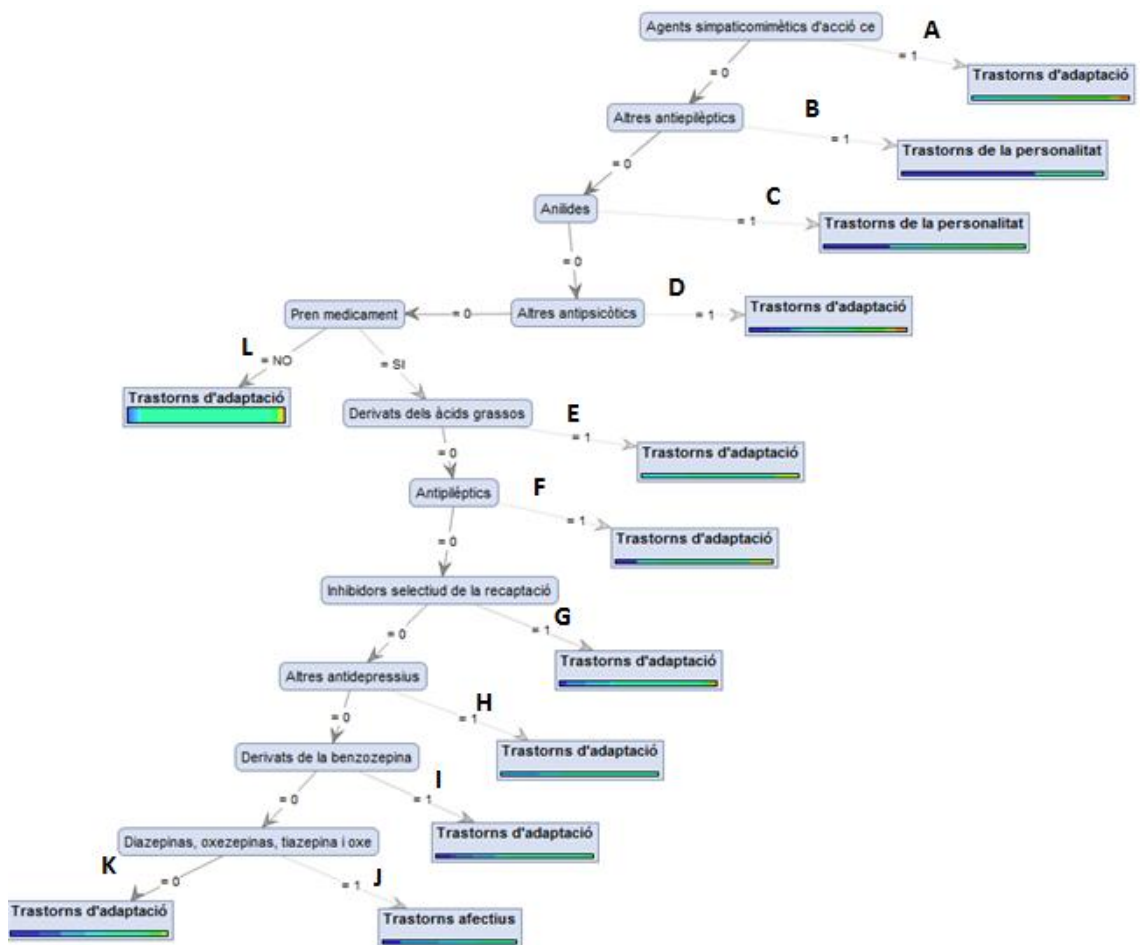
**G:** Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

**H:** Trastorns d'adaptació=2

En aquest arbre destaca que els pacients amb un trastorn d'adaptació si mantenen relacions inadequades deriven cap a un trastorn de personalitat.

- **Model predictiu 3:** AD sobre el trastorn d'adaptació i base medicaments.

Es genera un arbre de decisions (imatge 23) dels pacients que inicialment són diagnosticats amb un trastorn d'adaptació per predir, a partir del consum de medicaments segons el grup terapèutic, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 64,97 %.



Imatge 23 Arbre de decisions 3, trastorn d'adaptació i base medicaments.

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que consumeixen medicació amb grup terapèutic corresponent a agents simpaticomimètics d'acció cen): Trastorn de conducta=2, Trastorn de l'espectre autista=1, Trastorns d'adaptació=6, Trastorn amb dèficit d'atenció amb hiperactivitat=6, Trastorns de l'aprenentatge=1, Altres problemes que poden ser objecte d'atenció clínica=1

**B**: Trastorns de la personalitat=2, Trastorns d'adaptació=1

**C**: Trastorns de la personalitat=2, Trastorn de conducta=1, Trastorns d'adaptació=1, Altres=1, Trastorns específics de la infància=1

**D**: Trastorns de la personalitat=2, Esquizofrènia i altres trastorns psicòtics=2, Trastorns afectius=1, Trastorn de conducta=1, Trastorn de l'espectre autista=1, Trastorns d'adaptació=3, Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=2, Trastorns de l'aprenentatge=1, Altres problemes que poden ser objecte d'atenció clínica=1

**E**: Trastorn de conducta=1, Trastorn de l'espectre autista=1, Trastorns d'adaptació=4, Trastorns de l'aprenentatge=1

**F**: Trastorns de la personalitat=1, Trastorn de l'espectre autista=1, Trastorns d'adaptació=4, Trastorns de l'aprenentatge=1

**G:** Trastorns de la personalitat=2, Trastorn Ansietat=5, Trastorns afectius=6, Trastorn de conducta=1, Trastorns d'adaptació=22, Altres=1, Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1, Altres problemes que poden ser objecte d'atenció clínica=1

**H:** Trastorns afectius=2, Trastorn de conducta=1, Trastorns d'adaptació=5

**I:** Trastorns de la personalitat=3, Esquizofrènia i altres trastorns psicòtics=1, Trastorn Ansietat=3, Trastorns afectius=4, Trastorn de conducta=1, Trastorn de l'espectre autista=1, Trastorns d'adaptació=14, Altres=1, Trastorns específics de la infància=1

**J:** Trastorns de la personalitat=1, Trastorns afectius=2, Trastorn de conducta=1, Trastorn de l'espectre autista=1, Trastorns d'adaptació=2

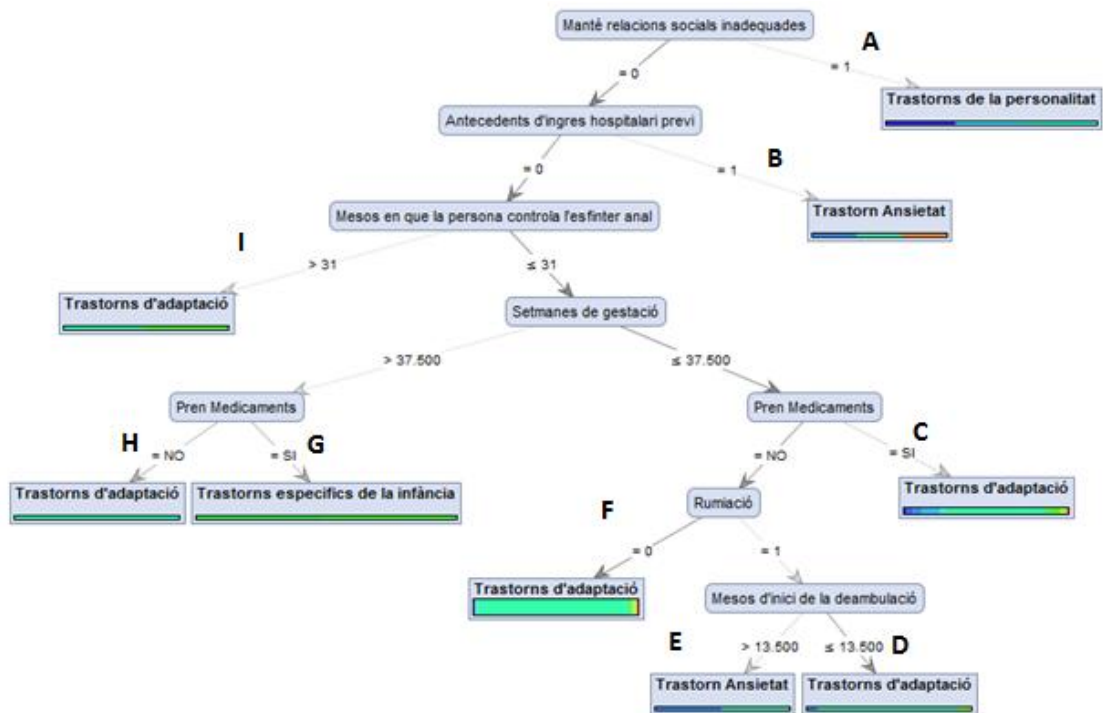
**K:** Trastorns de la personalitat=11, Esquizofrènia i altres trastorns psicòtics=8, Trastorns afectius=9, Trastorn de conducta=5, Trastorn de l'espectre autista=3, Trastorns d'adaptació=13, Altres=2, Trastorns específics de la infància=5, Trastorns de l'aprenentatge=2

**L:** Trastorns de la personalitat=6, Esquizofrènia i altres trastorns psicòtics=2, Trastorn Ansietat=9, Trastorns afectius=9, Trastorn de conducta=5, Trastorns d'adaptació=304, Altres=12, Trastorns específics de la infància=6, Trastorn amb dèficit d'atenció amb hiperactivitat=1, Trastorns de la conducta alimentaria=2, Trastorns de l'aprenentatge=7, Altres problemes que poden ser objecte d'atenció clínica=4

La majoria dels pacients que tenen trastorn d'adaptació com a diagnòstic principal i mantenen aquest com a últim diagnòstic, destaquen pel no consum de medicació.

- **Model predictiu 4:** AD sobre el trastorn d'adaptació i totes les bases.

Es genera un arbre de decisions (imatge 24) dels pacients que inicialment són diagnosticats amb un trastorn d'adaptació per predir, a partir de totes les dades disponibles, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 81,85 %.



Imatge 24 Arbre de decisions 4, trastorn d'adaptació i totes les bases.

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que mantenen relacions inadequades): Trastorns de la personalitat=1, Trastorn de conducta=1, Trastorn de l'espectre autista=1

**B** (són el tipus de pacients que no mantenen relacions inadequades i tenen antecedents d'ingrés hospitalari): Trastorn Ansietat=1, Trastorns d'adaptació=1, Altres problemes que poden ser objecte d'atenció clínica=1

**C**: Trastorns de la personalitat=2, Esquizofrènia i altres trastorns psicòtics=2, Trastorn Ansietat=4, Trastorns afectius=7, Trastorn de conducta=2, Trastorns d'adaptació=40, Altres=1, Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=5, Trastorns de l'aprenentatge=2, Altres problemes que poden ser objecte d'atenció clínica=1

**D**: Trastorn Ansietat=1, Trastorns d'adaptació=12, Trastorns de la conducta alimentaria=1

**E**: Trastorn Ansietat=1, Trastorns d'adaptació=1

**F**: Trastorns de la personalitat=4, Trastorn Ansietat=2, Trastorns afectius=1, Trastorn de conducta=2, Trastorns d'adaptació=267, Altres=7, Trastorns específics de la infància=2, Trastorn amb dèficit d'atenció amb hiperactivitat=1, Trastorns de la conducta alimentaria=1, Trastorns de l'aprenentatge=5, TOC=1, Altres problemes que poden ser objecte d'atenció clínica=3

**G**: Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

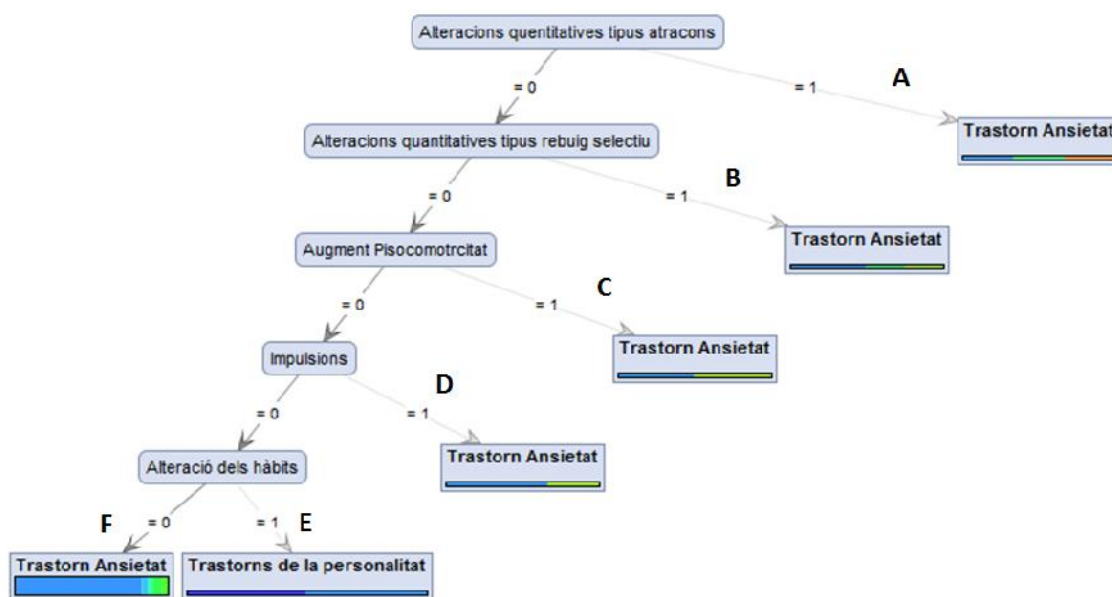
**H**: Trastorns d'adaptació=2

**I**: Trastorns d'adaptació=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

Es destaca els pacients amb setmanes de gestació superior a 37,5 i que prenen medicament, ja que deriven cap un trastorn específic de la infància.

- **Model predictiu 5:** AD sobre el trastorn d'ansietat i psicopatologia.

Es genera un arbre de decisions (imatge 25) dels pacients que inicialment són diagnosticats amb un trastorn d'ansietat per predir, a partir de característiques psicopatològiques, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 74,78 %



Imatge 25 Arbre de decisions 5, trastorn d'ansietat i psicopatologia.

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu a alteracions quantitatives tipus afartaments): Trastorn Ansietat=1, Altres=1, Altres problemes que poden ser objecte d'atenció clínica=1

**B:** Trastorn Ansietat=2, Altres=1, Trastorns de l'aprenentatge=1

**C:** Trastorn Ansietat=1, Trastorns de l'aprenentatge=1

**D:** Trastorn Ansietat=2, Trastorns de l'aprenentatge=1

**E:** Trastorns de la personalitat=1, Trastorn Ansietat=1

**F:** Trastorns de la personalitat=1, Trastorn Ansietat=63, Trastorns afectius=3, Trastorns d'adaptació=3, Trastorns específics de la infància=4, Trastorn amb dèficit d'atenció amb hiperactivitat=3

S'observa que la majoria de pacients amb un trastorn d'ansietat inicial i que no han donat positiu en alteracions quantitatives tipus afartaments, ni alteracions quantitatives tipus rebuig selectiu, ni augment psicomotricitat, ni impulsions, ni alteració dels hàbits el seu últim diagnòstic continua sent el de trastorn d'ansietat.

També s'observa que els pacients amb una alteració dels hàbits te com a últim diagnòstic al cap de sis anys un trastorns de la personalitat.

- **Model predictiu 6:** AD sobre el trastorn d'ansietat i base valoració social i anamnesis.

Es genera un arbre de decisions (imatge 26) dels pacients que inicialment són diagnosticats amb un trastorn d'ansietat per predir, a partir d'informació social i d'anamnesis, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 74,29 %.

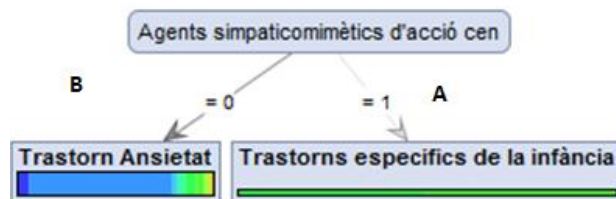


Imatge 26 Arbre de decisions 6, trastorn d'ansietat i base valoració social i anamnesis.

Amb les variables disponibles, no s'ha trobat suficient dependència lineal amb el subjecte i per tant s'entén que el millor model en aquest cas, és considerar que tots els pacients continuen amb un trastorn de ansietat.

- **Model predictiu 7:** AD sobre el trastorn d'ansietat i base medicaments.

Es genera un arbre de decisions (imatge 27) dels pacients que inicialment són diagnosticats amb un trastorn d'ansietat per predir, a partir del consum de medicaments segons el grup terapèutic, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 71.67 %.



Imatge 27 Arbre de decisions 7, trastorn d'ansietat i base medicaments.

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que consumeixen medicació amb grup terapèutic corresponent a agents simpaticomimètics d'acció cen): Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

**B**: (són el tipus de pacients que no consumeixen medicació amb grup terapèutic corresponent a agents simpaticomimètics d'acció cen): Trastorns de la personalitat=7, Trastorn Ansietat=83, Trastorns afectius=3, Trastorns d'adaptació=4, Altres=2, Trastorns específics de la infància=5, Trastorn amb dèficit d'atenció amb hiperactivitat=5, Trastorns

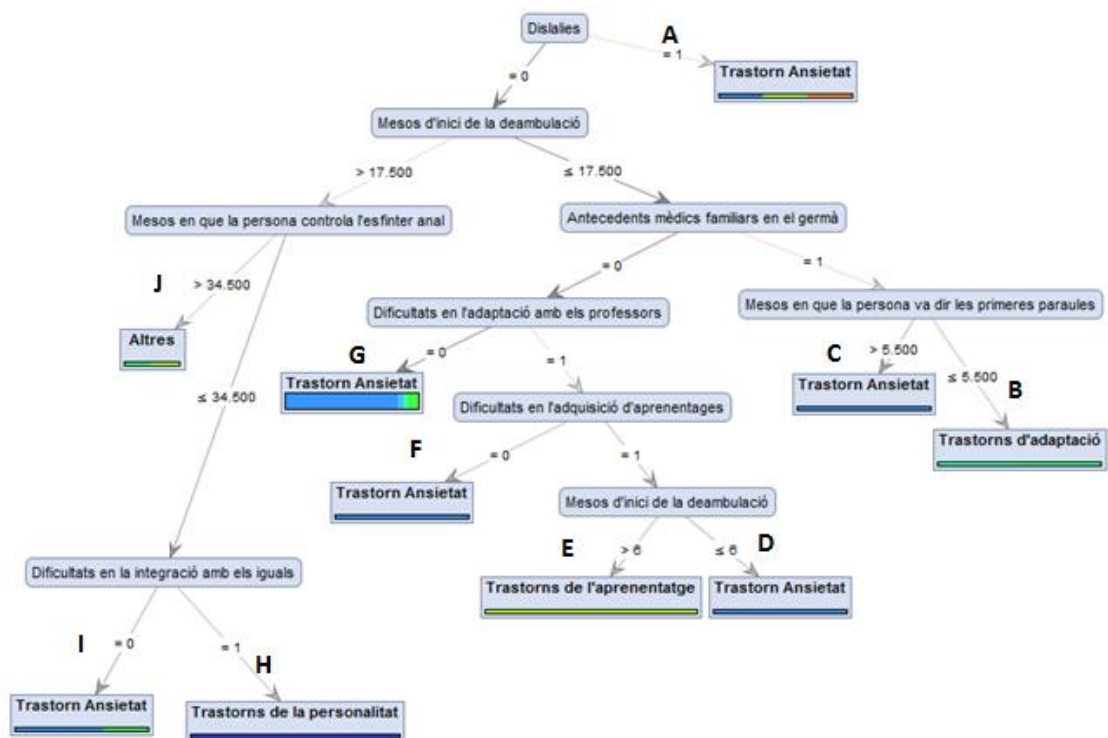
de l'aprenentatge=4, Altres problemes que poden ser objecte d'atenció clínica=1

Aquest arbre ens indica que els pacients amb diagnòstic inicial de trastorn d'ansietat i que han consumit medicació corresponent al grup terapèutic agents simpaticomimètics d'acció cen tenen una alta probabilitat de que el seu últim diagnòstic al cap de sis anys sigui trastorns específics de la infància.

Els que no han consumit medicació corresponent al grup terapèutic agents simpaticomimètics d'acció cen continuen majoritàriament en trastorn d'ansietat o canvien de diagnòstic inclòs també els trastorns específics de la infància.

- **Model predictiu 8:** AD sobre el trastorn d'ansietat i totes les bases.

Es genera un arbre de decisions (imatge 28) dels pacients que inicialment són diagnosticats amb un trastorn d'ansietat per predir, a partir de totes les dades disponibles, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 70,22 %.



Imatge 28 Arbre de decisions 8, trastorn d'ansietat i totes les bases.

Dimensions de les variables segons el tipus de pacients:

**A:** Trastorn Ansietat=1, Trastorns de l'aprenentatge=1, Altres problemes que poden ser objecte d'atenció clínica=1

**B:** Trastorns d'adaptació=1, Altres=1

**C:** Trastorn Ansietat=3

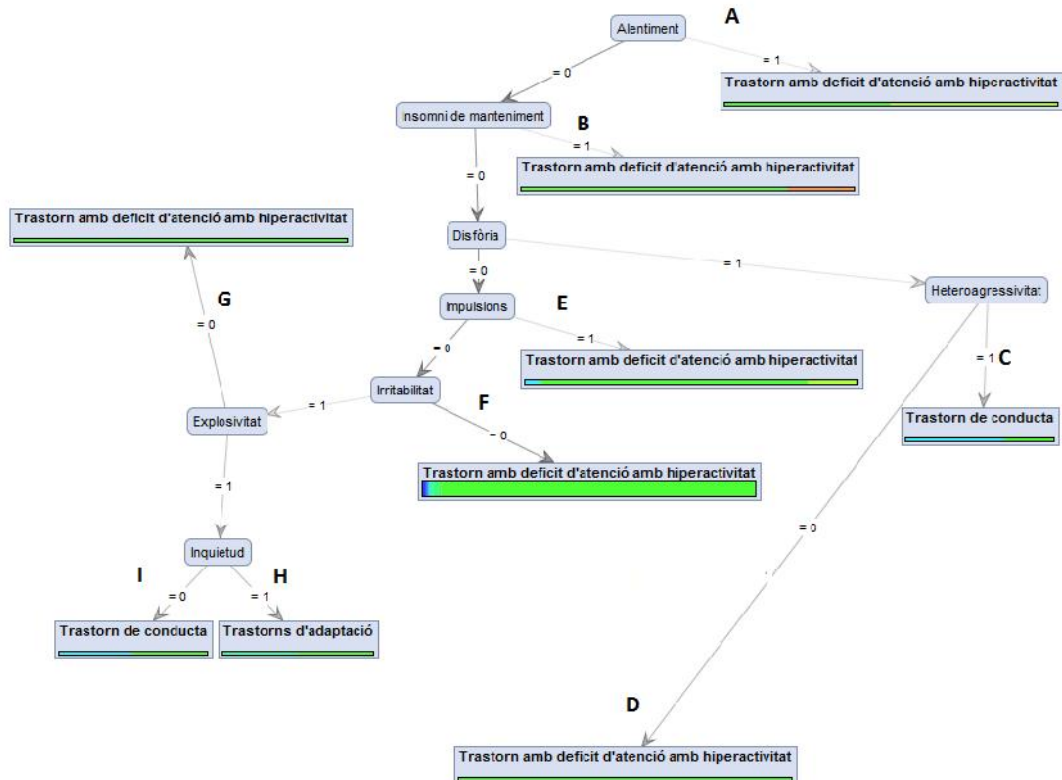
- D: Trastorn Ansietat=2
- E: Trastorns de l'aprenentatge=2
- F: Trastorn Ansietat=4
- G: Trastorn Ansietat=63, Trastorns afectius=3, Trastorns d'adaptació=2, Trastorns específics de la infància=3, Trastorn amb dèficit d'atenció amb hiperactivitat=3
- H: Trastorns de la personalitat=2
- I: Trastorn Ansietat=2, Trastorns específics de la infància=1
- J: Altres=1, Trastorns de l'aprenentatge=1

Es destaca els pacients que l'inici de la deambulació ha estat superior a 6 mesos i amb problemes em l'adquisició de l'aprenentatge i també amb problemes d'adaptació amb el professor, els quals deriven cap a un trastorn a l'aprenentatge.

També cal destacar que els pacient amb dificultats en la integració amb els iguals acaben deriven cap a un trastorn de personalitat

- **Model predictiu 9:** AD sobre el TDAH i psicopatologia.

Es genera un arbre de decisions (imatge 29) dels pacients que inicialment són diagnosticats amb un trastorn de dèficit d'atenció amb hiperactivitat per predir, a partir de característiques psicopatològiques, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 87,43 %



Imatge 29 Arbre de decisions 9, TDAH i bases psicopatologia.



Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu en alentiment): Trastorn amb dèficit d'atenció amb hiperactivitat=1, Trastorns de l'aprenentatge=1

**B**: (són el tipus de pacients que no que han donat positiu en alentiment per si en insomni de manteniment): Trastorn amb dèficit d'atenció amb hiperactivitat=4, Altres problemes que poden ser objecte d'atenció clínica=1

**C**: Trastorn de conducta=2, Trastorn amb dèficit d'atenció amb hiperactivitat=1

**D**: Trastorn amb dèficit d'atenció amb hiperactivitat=4

**E**: Trastorn de conducta=1, Trastorn amb dèficit d'atenció amb hiperactivitat=16, Trastorns de l'aprenentatge=3

**F**: Trastorns de la personalitat=2, Trastorn Ansietat=1, Trastorns d'adaptació=2, Altres=2, Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=132

**G**: Trastorn amb dèficit d'atenció amb hiperactivitat=5

**H**: Trastorns d'adaptació=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

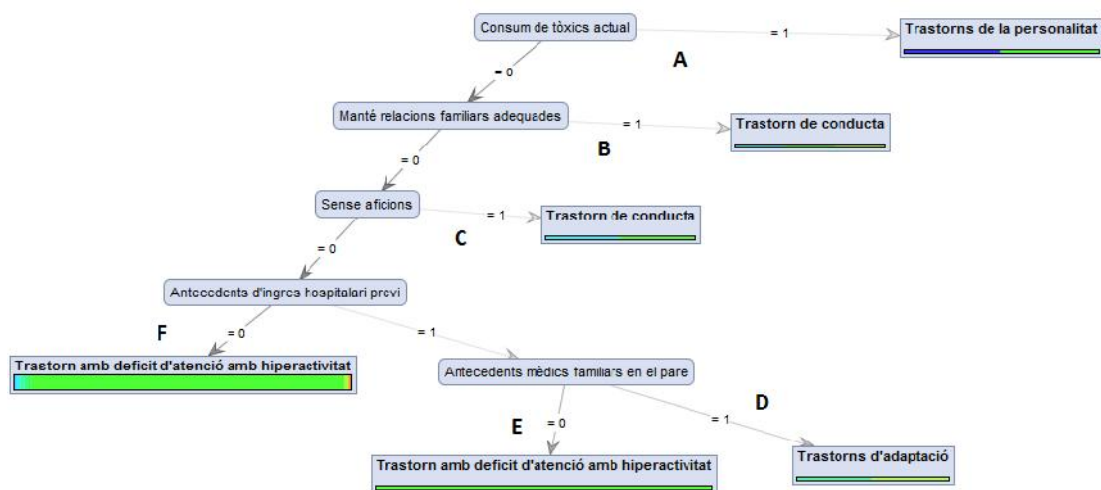
**I**: Trastorn de conducta=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

La majoria de pacients amb un trastorn amb dèficit d'atenció amb hiperactivitat inicial i que no han canviat el seu últim diagnòstic al cap de sis anys es caracteritzen per no donar positiu ni alentiment, ni insomni de manteniment, ni disfòria, ni impulsions, ni irritabilitat.

Els pacients que mostren irritabilitat i explosivitat o disfòria i heteroagressivitat tendeixen a derivar cap un trastorn de conducta, en canvi, els que mostren irritabilitat, explosivitat i a més a més inquietud deriven cap un trastorn d'ansietat.

- **Model predictiu 10**: AD sobre el TDAH i bases anamnesis i valoració social.

Es genera un arbre de decisions (imatge 30) dels pacients que inicialment són diagnosticats amb un trastorn de dèficit d'atenció amb hiperactivitat per predir, a partir d'informació social i d'anamnesis, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 89,41 %.



Imatge 30 Arbre de decisions 10, TDAH i bases anamnesis i valoració social.

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu en consum de tòxics):  
Trastorns de la personalitat=1, Trastorn amb deficit d'atenció amb hiperactivitat=1

**B**: Trastorn de conducta=1, Trastorn amb deficit d'atenció amb hiperactivitat=1, Trastorns de l'aprenentatge=1

**C**: Trastorn de conducta=1, Trastorn amb deficit d'atenció amb hiperactivitat=1

**D**: Trastorns d'adaptació=1, Trastorns de l'aprenentatge=1

**E**: Trastorn amb deficit d'atenció amb hiperactivitat=4

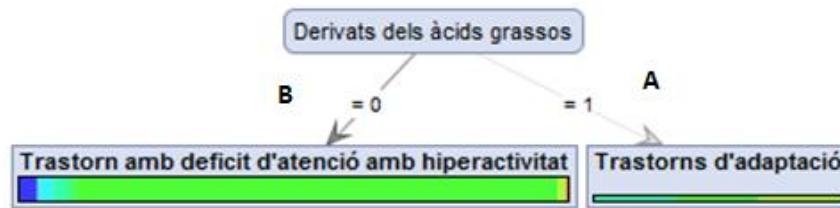
**F**: Trastorn Ansietat=1, Trastorn de conducta=2, Trastorns d'adaptació=2, Altres=2, Trastorns específics de la infància=1, Trastorn amb deficit d'atenció amb hiperactivitat=145, Trastorns de l'aprenentatge=2, Altres problemes que poden ser objecte d'atenció clínica=1

La majoria de pacients amb un trastorn amb deficit d'atenció amb hiperactivitat inicial i que el seu últim diagnòstic és el mateix al cap de sis anys es caracteritzen per no donar positiu en consum de tòxics, ni manté relacions social adequades, ni sense aficions, ni antecedents d'ingrés hospitalari previ.

Els pacients que no tenen aficions o mantenen relacions socials adequades tendeixen a derivar cap un trastorn de conducta, en canvi, els que si consumeixen tòxics deriven cap un trastorn de personalitat.

- **Model predictiu 11: AD sobre el TDAH i bases medicaments.**

Es genera un arbre de decisions (imatge 31) dels pacients que inicialment són diagnosticats amb un trastorn de deficit d'atenció amb hiperactivitat per predir, a partir del consum de medicaments segons el grup terapèutic, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 85.63 %.



Imatge 31 Arbre de decisions 11, TDAH i bases medicaments

Dimensions de les variables segons el tipus de pacients:

**A:** Trastorns d'adaptació=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1, Trastorns de l'aprenentatge=1

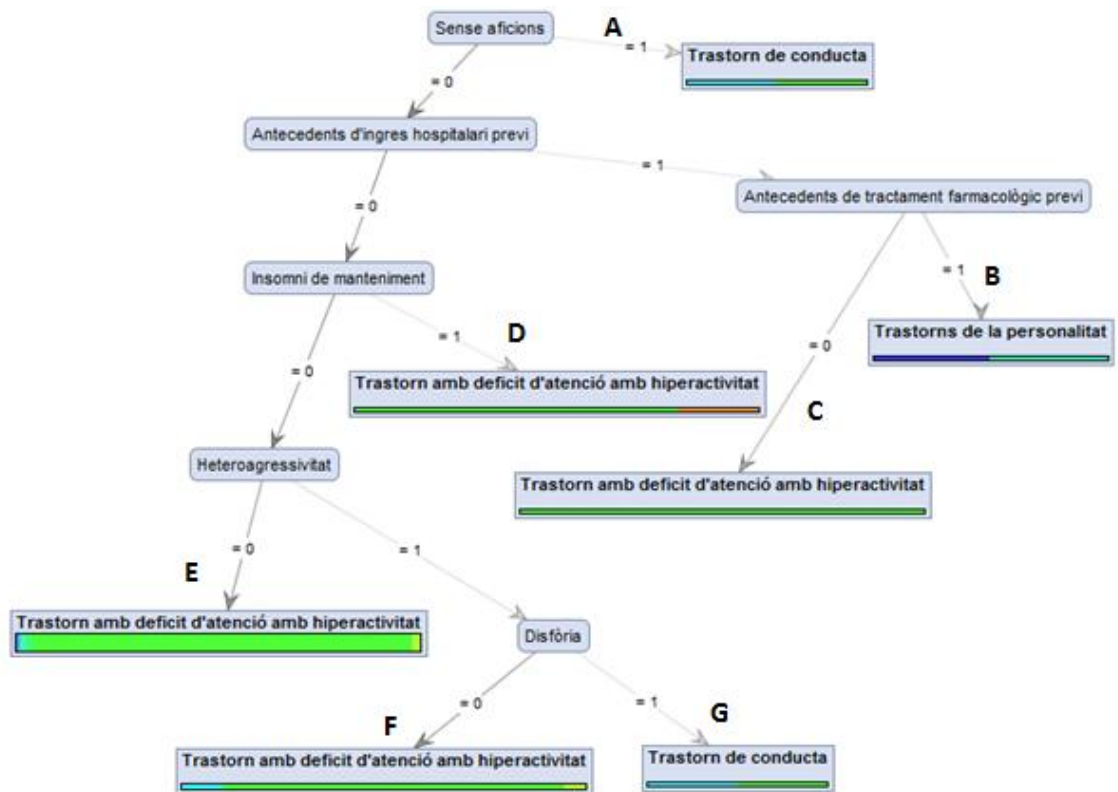
**B:** Trastorns de la personalitat=8, Trastorn Ansietat=1, Trastorn de conducta=6, Trastorns d'adaptació=6, Altres=2, Trastorns específics de la infància=2, Trastorn amb dèficit d'atenció amb hiperactivitat=204, Trastorns de l'aprenentatge=3, Altres problemes que poden ser objecte d'atenció clínica=1

Aquest arbre ens indica que els pacients amb diagnòstic inicial de trastorn amb dèficit d'atenció amb hiperactivitat i que no han consumit medicació corresponent al grup terapèutic derivats de los àcids grassos al cap de sis anys el seu últim diagnòstic continua sen un trastorn amb dèficit d'atenció amb hiperactivitat

Els que han consumit medicació corresponent al grup terapèutic derivats de los àcids grassos tenen una alta probabilitat de que el seu últim diagnòstic al cap de sis anys sigui trastorn d'adaptació o trastorn de l'aprenentatge.

- **Model predictiu 12:** AD sobre el TDAH i totes les bases.

Es genera un arbre de decisions (imatge 32) dels pacients que inicialment són diagnosticats amb un trastorn de dèficit d'atenció amb hiperactivitat per predir, a partir de totes les dades disponibles, quin és l'últim diagnòstic assignat al cap de sis anys. El percentatge d'encert és del 91,02 %.



Imatge 32 Arbre de decisions 12, TDAH i totes les bases

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu en sense aficions):  
 Trastorn de conducta=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

**B**: Trastorns de la personalitat=1, Trastorns d'adaptació=1

**C**: Trastorn amb dèficit d'atenció amb hiperactivitat=3

**D**: Trastorn amb dèficit d'atenció amb hiperactivitat=4, Altres problemes que poden ser objecte d'atenció clínica=1

**E**: Trastorns de la personalitat=1, Trastorn Ansietat=1, Trastorns d'adaptació=2, Altres=2, Trastorns específics de la infància=1, Trastorn amb dèficit d'atenció amb hiperactivitat=146, Trastorns de l'aprenentatge=3

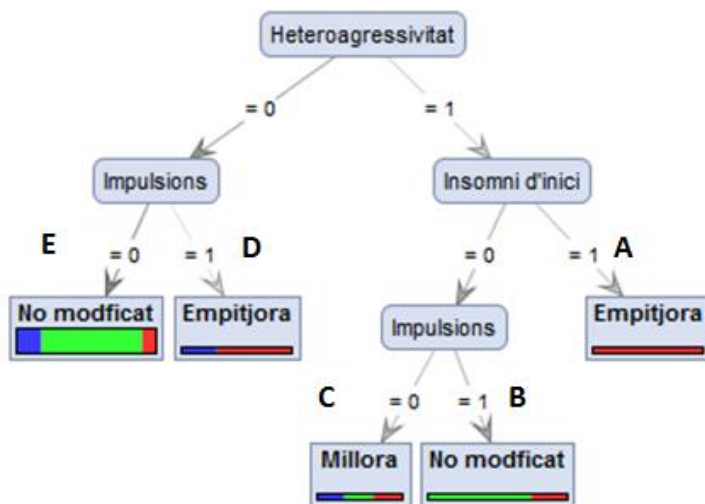
**F**: Trastorn de conducta=2, Trastorn amb dèficit d'atenció amb hiperactivitat=16, Trastorns de l'aprenentatge=1

**G**: Trastorn de conducta=1, Trastorn amb dèficit d'atenció amb hiperactivitat=1

La majoria de pacients amb un trastorn amb dèficit d'atenció amb hiperactivitat inicial i que al cap de sis anys te el mateixdiagnosticcom a últim es caracteritzen per no tenir heteroagressivitat, ni insomni de manteniment, ni antecedents d'ingrés hospitalari previ, i amb aficions.

- **Model predictiu 13:** AD sobre el GAF i base psicopatologia.

Es genera un arbre de decisions (imatge 33) de tots els pacients que tenen una valoració GAF per tal de predir, amb la base de dades de psicopatologia, quins motius fan que el pacient millori o empitjori. El percentatge d'encert és del 52 %.



Imatge 33 Arbre de decisions 13, GAF i base psicopatologia

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu en heteroagressivitat i insomni de manteniment): Empitjora=4

**B:** No modificat=3, Empitjora=1

**C:** Millora=1, No modificat=1, Empitjora=1

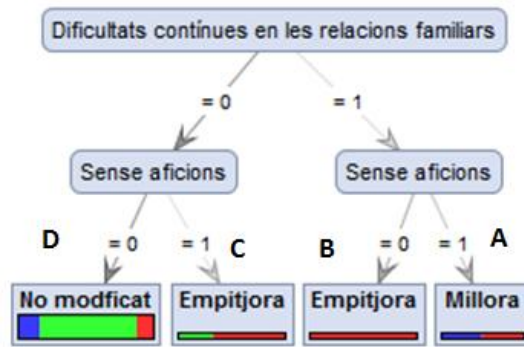
**D:** Millora=1, Empitjora=2

**E:** Millora=8, No modificat=31, Empitjora=4

És destaca els pacients amb heteroagressivitat i impulsions d'inici, ja que tots ells han empitjorat l'estat al cap de 6 anys.

- **Model predictiu 14:** AD sobre el GAF i bases de valoració social i anamnesis.

Es genera un arbre de decisions (imatge 34) de tots els pacients que tenen una valoració GAF per tal de predir, amb la base de dades d'anamnesis i valoració social, quins motius fan que el pacient millori o empitjora. El percentatge d'encert és del 66 %.



Imatge 34 Arbre de decisions 14, GAF i anamnesis i valoració social

Dimensions de les variables segons el tipus de pacients:

**A** (són el tipus de pacients que han donat positiu en dificultats contínues en les relacions familiars i en sense aficions): Millora=1, Empitjora=1

**B**: Empitjora=3

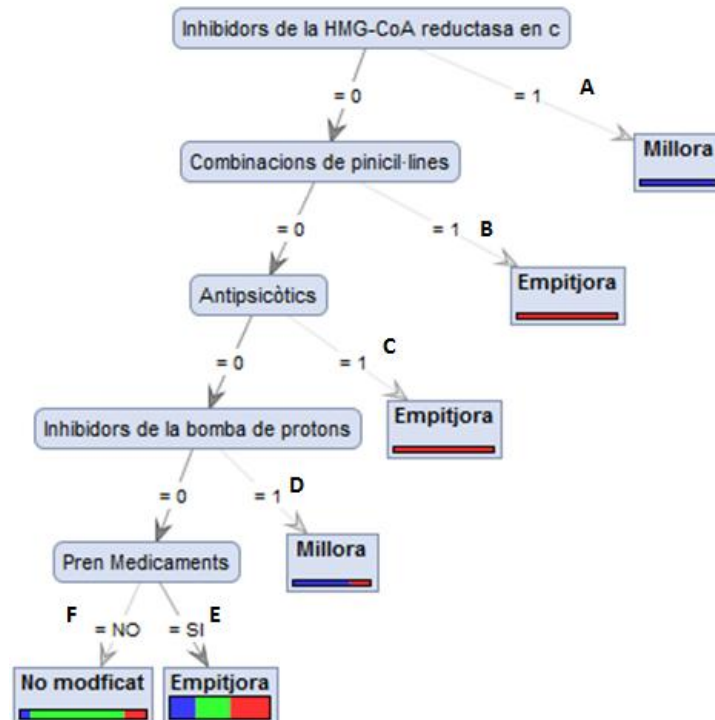
**C**: No modificat=1, Empitjora=2

**D**: Millora=7, No modificat=30, Empitjora=5

Es destaca els pacients amb dificultats contínues en les relacions familiars i que tenen aficions, ja que tots ells han empitjorat l'estat al cap de 6 anys.

- **Model predictiu 15: AD sobre GAF i base medicaments.**

Es genera un arbre de decisions (imatge 35) de tots els pacients que tenen una valoració GAF per tal de predir, amb la base de dades de medicaments, quins motius fan que el pacient millori o empitjori. El percentatge d'encert és del 41,99 %.



Imatge 35 Arbre de decisions 15, GAF i medicaments

Dimensions de les variables segons el tipus de pacients:

**A:** Millora=2

**B:** Empitjora=3

**C:** Empitjora=2

**D:** Millora=3, Empitjora=1

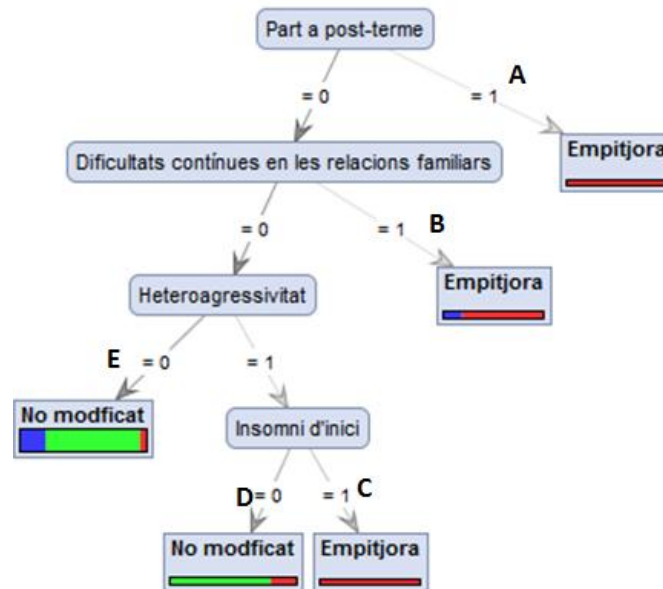
**E:** Millora=83, No modificat=104, Empitjora=118

**F:** Millora=7, No modificat=57, Empitjora=13

Es destaca els pacients que consumeixen medicació del grup terapèutic penicil·lina o antipsicòtics, perquè han empitjorat l'estat al cap de 6 anys.

- **Model predictiu 16:** AD sobre el GAF i totes les bases.

Es genera un arbre de decisions (imatge 36) de tots els pacients que tenen una valoració GAF per tal de predir, amb totes les bases de dades, quins motius fan que el pacient millori o empitjori. El percentatge d'encert és del 64.00 %.



Imatge 36 Arbre de decisions 16, GAF i totes les bases

Dimensions de les variables segons el tipus de pacients:

**A:** Empitjora=2

**B:** Millora=1, Empitjora=4

**C:** Empitjora=3

**D:** No modificat=4, Empitjora=1

**E:** Millora=9, No modificat=32, Empitjora=2

Es destaca els pacients amb heteroagressivitat i insomni d'inici, atès que tots ells han empitjorat l'estat al cap de 6 anys.

### 3.8 Comentaris dels arbres pel departament Csmij de psiquiatria.

Generats els arbres de decisions es decideix valorar-los pel Departament de Psiquiatria del Csmij. Un cop exposades totes les taules i els arbres de decisions generats, es copia un resum dels comentaris aportats.

Hi ha variables que aporten una informació rellevant, com els pacients amb un TDHA inicial que consumeixen tòxics, que en el seu diagnòstic final, al cap de sis anys, és el trastorn de personalitat. També hi ha pacients que comencen amb un trastorn d'ansietat i que el seu diagnòstic al cap de sis anys és un trastorn de la personalitat, com a variable motivadora d'això és l'alteració dels hàbits.

També comenten que hi ha variables sense massa sentit mèdic, com els pacients amb un TDHA inicial. Aquests si mantenen relacions familiars adequades, al cap de sis anys el seu últim diagnòstic és un trastorn de conducta.

En aquest treball l'objectiu era valorar la situació inicial dels pacients que entraven per primer cop al Csmij el 2010 i comparar-la amb la situació que es troben en el 2016. Des del Departament de Psiquiatria del Csmij comenten, però, que seria molt interessant fer un treball amb distàncies més curtes de temps o introduir el temps com una variable més. Aquest fet és degut a que hi ha molts casos en què el pacient accedeix a les instal·lacions del Csmij i se l'hi assigna un diagnòstic (degut a que tots els pacients que accedeixen al Csmij tenen que estar diagnosticats) i el mateix dia o al cap de pocs dies se'l doni d'alta.

Aquests anàlisis s'executaran en els propers mesos per generar uns models predictius més adequats pel Departament de Psiquiatria del Csmij. Amb aquests es podran valorar si hi ha més variables rellevants de les ja descrites en aquest projecte que permetin predir millors canvis de diagnòstics.



## 4. Conclusions

En aquest treball s'aplica mètodes d'anàlisi predictiu (bàsicament arbres de presa de decisió) a partir de la informació acumulada a les instal·lacions de l'Hospital Psiquiàtric Universitari Institut Pere Mata (HPUIPM).

S'han generat algorismes de classificació i agrupació per quantificar patologies segons sexe i trams d'edat. També s'ha generat una taula de canvi de diagnòstic que ha permès valorar als professionals quins perfils de pacients mèdics canvien de diagnòstic. A partir d'aquestes dades s'ha seleccionat els grups d'especial interès per generar els arbres de decisions segons criteris dels professionals del centre.

Aquests arbres els han estat avaluats buscant una lògica mèdica i amb el percentatge més elevat possibles d'exactitud.

Amb la generació de les taules i amb els arbres de decisió es considera que s'han assolit els objectius inicials plantejats. S'ha intentant ajustar-se al màxim al pla de treball, però hi ha hagut un retràs a les diferents fases (causats per la necessitat de més temps per a la neteja de dades). Aquest fet ha estat degut a que era el primer cop que es treballaven en elles i s'han detectat errors en la introducció d'aquestes.

A més es considera la generació d'arbres de decisió una contribució a la bibliografia existent, degut a la escassa informació publicada d'arbres de decisió en psiquiatria, aportant una nova forma de predir futurs successos i una forma més simple d'explicar per què alguns diagnòstics canvien i el motiu pel qual milloren o empitjoren.

En els resultats obtinguts es destaca variables psicopatològiques com el insomni d'inici, que fa que els pacients derivin cap a un trastorn d'ansietat, o variables com hetroagressivitat i disfòria o explosivitat i inquietud, que combinades influeixen a que el pacient acabi amb un trastorn de conducta.

Donar positiu en algunes variables de valoració social com són: tenir relacions socials inadequades, alteració dels hàbits, consum de tòxics o dificultats en la integració amb els iguals, augmenta el risc de derivar cap a un trastorn de la personalitat. En canvi, els pacients que mantenen relacions socials inadequades i a més a més no ténen aficions són més pròxims a acabar amb un trastorn de conducta.

Tenir dificultats en l'adquisició de l'aprenentatge o amb l'adaptació amb el professor són símptomes de valoració social que indiquen que el pacient pot derivar cap a un trastorn de l'aprenentatge.

Els arbres ens indiquen que alguns factors d'anamnesis, com antecedents de tractament psicològic previ o setmanes de gestació superior 37,5, són factors que fan que nens diagnosticats amb un trastorn d'adaptació acabin derivant cap a un trastorn específic de la infància.

El consum d'alguns medicaments pot ser indicador de què un pacient derivi a un diagnòstic concret al cap de 6 anys. Per exemple el consum de medicació amb principi actiu derivats dels àcids grassos ens indica que els pacients que tenen un trastorn d'adaptació continuaran amb el trastorn d'adaptació com a diagnòstic final, en canvi els que tenen un trastorn de dèficit d'atenció amb hiperactivitat el seu últim diagnòstic es un trastorn d'adaptació.

Remarcar sobretot, que la majoria de pacients no han mostrat cap símptoma de psicopatologia, valoració social o anamnesis dels indicats als arbres de decisions i que destaquen perquè aquests tenen mateix el diagnòstic inicial al cap de sis anys.

Un futur estudi podria ser cercar els motius pels quals aquests pacients fan que al cap de sis anys mantinguin el diagnòstic, com el cas del trastorn d'adaptació considerat un trastorn temporal. Una manera seria introduint la variable de temps, o també treballar amb distàncies més curtes de temps com per exemple comparant-ho cada any.

Referent a l'estat del pacient segons la puntuació GAF, les variables que fan que els pacients empitjorin seran la heteroagressivitat, l'insomni d'inici, les dificultats contínues en les relacions familiars i part-a post-terme.

## 5. Glossari

**Anamnesis:** la informació (dades subjectives: antecedents familiars i personals, signes i símptomes del problema de salut, experiències, records) proporcionada pel propi pacient al professional sanitari durant una entrevista clínica per a analitzar la seva situació clínica i conformar la seva història clínica.

**Atribut:** és una especificació que defineix una propietat d'un objecte, elements o arxiu.

**Dades massives (o Big Data):** és el nom que reben els conjunts de dades, els procediments i les aplicacions informàtiques, que, pel seu volum, la seva naturalesa diversa i la velocitat en què han de ser processades, ultrapassen la capacitat dels sistemes informàtics habituals. Aquest processament de dades massives s'utilitza per tal de detectar patrons dins seu, podent fer així prediccions vàlides per a la presa de decisions.

**Desviació:** esta relacionada amb la variància, i que és l'arrel quadrada de la variància.

$$= \sqrt{V(x)} \text{ o bé } V(x) = (\text{Desviació})^2$$

**Exemple:** cada una de les mostres.

**Interval de confiança:** és un interval de valors al voltant d'un paràmetre mostral en els quals, amb una probabilitat o nivell de confiança determinat, se situarà el paràmetre poblacional a estimar.

Si  $\alpha$  és l'error aleatori que es vol cometre, la probabilitat serà de  $1 - \alpha$ .

A menor nivell de confiança l'interval serà més precís, però es cometrà un major error.

**Medicina predictiva:** és la medicina que avalua els factors de risc de cada persona enfront la malaltia, i anticipa la resposta al tractament.

**Medicina reactiva:** La medicina que provoca al pacient estímuls que li produeixen reflexes o reaccions positives.

**Mineria de dades:** La mineria de dades (Data Mining) és un procés no trivial d'identificació d'informació vàlida, nova, potencialment útil i atencible de patrons comprensibles que es troben ocults en les dades.

**Patologia:** és la part de la medicina encarregada de l'estudi de les malalties en el seu més ampli sentit, és a dir, com a processos o estats anormals de causes conegudes o desconegudes. És una ciència que consisteix en l'examen físic i de les proves complementàries per a diagnosticar una malaltia. La prova que més demostra l'existència d'una malaltia és l'examen d'una lesió en tots els seus nivells estructurals.

**Psicopatologia:** aquella àrea de la salut que descriu i sistematitza els canvis en el comportament que no són explicats, per la maduració o desenvolupament del individu, ni com a resultat de processos d'aprenentatge, també entesos com a trastorns psicològics, malalties o trastorns mentals.

**Variable dependent:** representa la sortida o resultat de totes les variables que s'estan estudiant. En Rapidminer es la variable "labeled".

**Variable independent:** representen entrades o causes, és a dir, els motius potencials de variació. Els models predictius tracten d'explicar els efectes que tenen les variables independents sobre les variables dependents. De vegades les variables independents poden ser inclosos per altres raons, com pel seu potencial efecte de confusió, sense el desig de provar el seu efecte directament.

**Wald test:** és una prova estadística paramètrica anomenada així pel estadístic hongarès Abraham Wald. S'utilitza sempre que hi hagi una relació dins o entre els elements de dades i es pot expressar com un model estadístic amb els paràmetres a estimar a partir d'una mostra, la prova de Wald es pot utilitzar per posar a prova el veritable valor del paràmetre basat en l'estimació de la mostra.

**Variància:** en teoria de probabilitat i estadística, la variància indica la dispersió d'una variable aleatòria  $X$  de la seva mitjana  $E[x]$ . Es defineix com l'esperança de la transformació  $(X-E(x))^2$ .

$$V(x) = E[(X-E(x))^2]$$

## 6. Bibliografia

- [1] Sebastián Maldonado, Richard Weber, Wrapper method for feature selection using Support Vector Machines
- [2][http://eciemaps.msssi.gob.es/ecieMaps/browser/index\\_9\\_mc.html](http://eciemaps.msssi.gob.es/ecieMaps/browser/index_9_mc.html)  
10/05/2017
- [3] Análisis de datos multivariantes Daniel Peña 21/04/2017, ScienceDirect, Information Sciences 179 (2009) 2208–2217.
- [4] Christina Mohr Jensen \* Hans-Christoph Steinhausen \* Marlene Briciet Lauritsen, Time Trends Over 16 Years in Incidence-Rates of Autism Spectrum Disorders Across the Lifespan Based on Nationwide Danish Register Data, CM Jensen et al. J Autism Dev Disord 44 (8), 1808-1818. 8 2014  
<https://www.ncbi.nlm.nih.gov/pubmed/24554161>
- [5] Mark Olfson, M.D., M.P.H. Benjamin G. Druss, M.D.,M.P.H., and Steven C. Marcus, Ph.D., Trends in Mental Care among Children and Adolescents, The New England Journal of Medicine. 2029-2038, May 21/2015  
<http://www.nejm.org/doi/full/10.1056/NEJMsa1413512#t=article>
- [6] Giulio Napolitano, Adele Marshall, Peter Hamilton, Anna T. Gavin. Machine learning classification of surgical pathology reports and chunk recognition for information extraction noise reduction. Artificial Intelligence in Medicine 70 (2016) 77-83.  
<https://www.ncbi.nlm.nih.gov/pubmed/27431038>
- [7] [https://en.wikipedia.org/wiki/Naive\\_Bayes\\_classifier](https://en.wikipedia.org/wiki/Naive_Bayes_classifier) 27/04/2017
- [8] [https://en.wikipedia.org/wiki/Decision\\_tree](https://en.wikipedia.org/wiki/Decision_tree) 14/03/2017
- [9]<http://www.obitko.com/tutorials/neural-network-prediction/neural-network-training.html>
- [10][https://en.wikipedia.org/wiki/Logistic\\_regression](https://en.wikipedia.org/wiki/Logistic_regression) 27/04/2017
- [11][https://en.wikipedia.org/wiki/C4.5\\_algorithm](https://en.wikipedia.org/wiki/C4.5_algorithm) 14/03/2017
- [12]<https://www.scientificcomputing.com/article/2007/05/cart-powerful-decision-tree-tool-predictive-modeling> 14/03/2017
- [13][http://eciemaps.msssi.gob.es/ecieMaps/browser/index\\_10\\_2008.html](http://eciemaps.msssi.gob.es/ecieMaps/browser/index_10_2008.html)  
10/05/2017
- [14][https://en.wikipedia.org/wiki/Global\\_Assessment\\_of\\_Functioning](https://en.wikipedia.org/wiki/Global_Assessment_of_Functioning) 21/05/2017
- [15]<http://hadoop.apache.org/> 10/05/2017
- [16]<http://mahout.apache.org/> 10/05/2017
- [17]<http://spark.apache.org/docs/latest/sparkr.html> 10/05/2017
- [18]<http://spark.apache.org/docs/2.1.0/api/python/pyspark.html> 10/05/2017
- [19]<https://rapidminer.com/> 21/04/2017
- [20]<http://docs.rapidminer.com/> 31/03/2017
- [21]<https://es.scribd.com/doc/104509028/Tutorial-de-Induccion-al-RapidMiner>  
02/04/2017
- [22][https://es.wikipedia.org/wiki/Front-end\\_y\\_back-end](https://es.wikipedia.org/wiki/Front-end_y_back-end) 10/04/2017
- [23][http://catsalut.gencat.cat/web/.content/minisite/catsalut/coneix\\_catsalut/memories\\_activitat/memories\\_regions\\_sanitaries/memories\\_2005\\_2013/camp\\_tarragona/memoria\\_rscamptarragona\\_2015.pdf](http://catsalut.gencat.cat/web/.content/minisite/catsalut/coneix_catsalut/memories_activitat/memories_regions_sanitaries/memories_2005_2013/camp_tarragona/memoria_rscamptarragona_2015.pdf) 26/05/2017