

TREBALL FI DE GRAU

Grau en Enginyeria Biomèdica

**TRACTAMENT DE SENYALS PER L'AJUT A LA DIAGNOSI DE
L'APNEA DEL SON**



Volum I

Memòria i Pressupostos

Autor: Arnau Vizcaino Torras

Director: Jordi Solà Soler

Departament: Dept. ESAll

Co-Director: Gerard Escudero Bakx

Convocatòria: Octubre 2017

Resum

Diversos estudis científics han observat un lligam entre la falta d'oxigenació del cervell a causa de les parades en els cicles respiratoris derivada de la presència d'apnees i l'aparició prematura de patologies que no s'esperen fins a una edat avançada.

L'objectiu principal d'aquest projecte és desenvolupar eines informàtiques per a l'ajut a la detecció del síndrome de l'apnea del son (SAOS), prèviament a realitzar una Polisomnografia (PSG), que és la prova mèdica estàndard per diagnosticar trastorns en el son. Es du a terme a partir de l'enregistrament de la senyal bioacústica dels roncs de pacients durant el son i l'extracció de les seves característiques temporals i freqüencials.

El treball consta de tres estudis diferents:

- Dos estudis d'aprenentatge automàtic no supervisat: un de mineria de dades i un d'anàlisi d'agrupament. L'objectiu és crear models representatius de les característiques del ronc de cada individu i avaluar la possibilitat de detectar la presència del SAOS, únicament a partir del registre sonor dels roncs i la informació prèvia de cada pacient.
- Un estudi d'aprenentatge automàtic supervisat de classificació i regressió per avaluar quin algorisme de decisió funciona millor per a la identificació de la severitat del SAOS d'un pacient a partir dels models desenvolupats.

Resumen

Diversos estudios científicos han observado la relación que existe en la falta de oxigenación al cerebro a causa de las paradas en los ciclos respiratorios derivados de la presencia de apneas y la aparición prematura de patologías que no suelen aparecer en una edad poco avanzada.

El objetivo principal de este proyecto es desarrollar herramientas informáticas para la ayuda a la detección del síndrome de la apnea del sueño (SAOS), previamente a realizar una Polisomnografía (PSG), que es la prueba médica estándar para diagnosticar trastornos en el sueño. Se realiza a partir del registro de la señal acústica de los ronquidos de pacientes durante el sueño y la extracción de sus características temporales y de frecuencia.

El proyecto consiste en tres estudios diferentes:

- Dos estudios de aprendizaje automático no supervisado: uno de minería de datos y uno de análisis de agrupamiento. El objetivo es crear modelos representativos de las características del ronquido de cada individuo y evaluar la posibilidad de detectar la presencia del SAOS, únicamente a partir del registro de los ronquidos y la información previa de cada paciente.
- Un estudio de aprendizaje automático supervisado de clasificación y regresión para evaluar cual algoritmo de decisión trabaja mejor para la identificación de la severidad del SAOS de un paciente a partir de los modelos desarrollados.

Abstract

Several scientific studies have observed the relationship between absence of oxygenation to the brain due to the stops in the respiratory cycles derived from the presence of apneas and the premature appearance of pathologies that do really appear in an advanced age.

The principal objective of this project is to develop computer tools to help the detection of sleep apnea syndrome (SAOS), before performing a Polysomnography (PSG), which is the standard medical test to diagnose sleep disorders. This can be done from the recording of the acoustic signal of snoring of patients during sleep and the extraction of their temporal and frequency characteristics.

The project consists on three different studies:

- Two unsupervised self-learning studies: one data mining and one cluster analysis studies. The aim is to create models that are representative of the characteristics of snoring of each individual, and to evaluate the possibility of detecting the presence of OSAS, only from the recording of snoring and previous information of each patient.
- A supervised automated classification and regression learning study to assess which algorithm works best for identifying a patient's OSAS severity from the developed models.

Agraïments

M'agradaria agrair principalment l'ajuda que he rebut per part dels dos tutors, en Jordi Solà Soler i en Gerard Escudero Bakx, que m'han facilitat tot el material i les eines informàtiques necessàries per desenvolupar el meu treball. Gràcies al seu esforç i dedicació he pogut desenvolupar aquest projecte, ja que sense els coneixements d'en Jordi Solà Soler sobre aquest símptoma i sobre les tècniques que s'utilitzen actualment i sense les nocions en la programació informàtica que m'ha impartit en Gerard Escudero Bakx.

També agraeixo a totes les persones que indirectament han contribuït en la meva dedicació a aquest projecte.

Llista d'abreviatures

SAOS: Síndrome d'Apnea Obstructiva del Son

PSG: Polisomnografia

VAS: Via Aèria Superior

ACV: Accident Vascular Cerebral

IAH: Índex d'Apnees-Hipopnees

HI: Índex d'Hipopnees

IA: Índex d'Apnees

CPAP: Pressió Positiva Continua en la via Aèria

SVM: Màquina de Suport Vectorial

DT: Arbres de decisió

KNN: Veïns més propers

MSE: Error quadràtic mig

WEKA: *Waikato Environment for Knowledge Analysis*



Índex

RESUM	I
RESUMEN	II
ABSTRACT	III
AGRAÏMENTS	IV
LLISTA D'ABREVIATURES	V
ÍNDIX DE FIGURES	IX
ÍNDIX DE TAULES	X
1. PREFACI	1
1.1. Origen del treball	1
1.2. Motivació	1
1.3. Requeriments previs.....	2
1.4. Planificació: Diagrama de Gantt	2
2. INTRODUCCIÓ	4
2.1. Objectius del treball.....	5
Objectius generals.....	5
Objectius específics.....	5
2.2. Abast del treball	5
3. MARC TEÒRIC	7
3.1. El ronc.....	7
3.2. Síndrome d'Apnea Obstructiva del Son	8

4. METODOLOGIA	15
4.1. Material inicials	15
4.2. Esquema de treball: FrameWork.....	17
4.3. Diferents tipus d'estudi.....	18
4.4. Aprenentatge automàtic	18
4.5. Clustering	20
4.6. Assignació de categories (Algoritme K MEANS).....	21
4.7. Classificació i regressió.....	22
4.7.1. Eina de selecció de paràmetres (Grid Search)	25
4.8. Eines informàtiques	26
5. EXPERIMENTS	29
5.1. Estudi 1.....	30
5.2. Estudi 2.....	32
5.3. Estudi 3.....	32
5.4. Interpretació resultats de l'estudi 3	34
5.5. Iteracions de l'estudi 3: classificació.....	36
6. ANÀLISIS DE RESULTATS	40
7. CONCLUSIONS	55
7.1. Perspectives futures	56
8. ANÀLISI DE L'IMPACTE AMBIENTAL	57
9. PRESSUPOST I/O ANÀLISI ECONÒMICA	59
9.1. Costos d'enginyeria.....	59
9.2. Costos informàtics	60

9.3. Altres costos.....	60
9.4. Pressupost final.....	61
10. BIBLIOGRAFIA	63
11. ANNEX A: MANUAL D'USUARI	1
12. ANNEX B: CODIS PROGRAMACIÓ	4



Índex de figures

Figura 1. Diagrama de Grantt _____	3
Figura 2. Obstrucció del VAS ([6]) _____	7
Figura 3. Diagrama d'obstrucció del VAS. ([2]) _____	10
Figura 4. Fisiopatologia del SAOS ([14]) _____	11
Figura 5. Exploració mèdica visual en sospita de SAOS ([17]) _____	12
Figura 6. Procediment en el diagnòstic i tractament del SAOS ([17]) _____	13
Figura 7. Implementació de la teràpia amb CPAP contra SAOS severes ([24]) _____	14
Figura 8. <i>FrameWork</i> _____	17
Figura 9. Exemple estudi regressió ([26]) _____	20
Figura 10. Exemple estudi classificació ([26]) _____	20
Figura 11. Exemple d'hiperplans del SVM ([28]) _____	23
Figura 12. Exemple d'aplicació K-neighbors ([29]) _____	24
Figura 13. Exemple de Random Forest ([31]) _____	25
Figura 14. Portal del WEKA ([22]) _____	26
Figura 15. Portal del Explorer del WEKA ([22]) _____	27
Figura 16. Portal del Experiment del WEKA ([22]) _____	27
Figura 17. Esquema dels estudis plantejats _____	30
Figura 18. Repartició de pacients per la creació d'arxius d'estudi per mantenir la presencia de totes les classes _____	31
Figura 19. Esquema de l'estudi 3 _____	33
Figura 20. Exemple de matriu de confusió ([23]) _____	36
Figura 21. Flux d'aire, saturació d'oxigen en sang (SpO ₂), moviment de l'abdomen ([35]) _____	37

Índex de taules

Taula 1. Índex IAH _____	9
Taula 2. Exemple de llista dels pacients _____	15
Taula 3. Exemple caracterització de les 8 primeres mostres de roncs del pacient SIG 33 _____	16
Taula 4. Tipus d'estudis _____	18
Taula 5. Algorisme K-Means _____	22
Taula 6. Resultats estudi mineria de dades WEKA _____	32
Taula 7. Taula exemple de model generat per l'algorisme d'agrupament <i>K-Means</i> _____	32
Taula 8. Índex apnea-hipopnea en adults _____	34
Taula 9. Fronteres al genera els grups de classes _____	34
Taula 10. Exemple de taula de resultats per a l'índex IAH format per 5 grups de severitat _____	35
Taula 11. Conceptes formula 4 _____	38
Taula 12. Repartició dels pacients en classes _____	38
Taula 13. Repartició dels pacients en classes _____	38
Taula 14. Repartició dels pacients en classes _____	39
Taula 15. Taula de resultats _____	40
Taula 16. Taula de resultats _____	41
Taula 17. Taula de resultats _____	41
Taula 18. Taula de resultats _____	41
Taula 19. Taula de resultats _____	42
Taula 20. Taula de resultats _____	42
Taula 21. Taula de resultats _____	43
Taula 22. Taula de resultats _____	43
Taula 23. Taula de resultats _____	44
Taula 24. Taula de resultats _____	44
Taula 25. Taula de resultats _____	45
Taula 26. Taula de resultats _____	45
Taula 27. Taula de resultats _____	45
Taula 28. Taula de resultats _____	46
Taula 29. Taula de resultats _____	46

Taula 30. Taula de resultats	46
Taula 31. Taula de resultats	47
Taula 32. Taula de resultats	47
Taula 33. Taula de resultats	48
Taula 34. Taula de resultats	48
Taula 35. Taula de resultats	49
Taula 36. Taula de resultats	49
Taula 37. Taula de resultats	50
Taula 38. Taula de resultats	50
Taula 39. Taula de resultats	50
Taula 40. Taula de resultats	51
Taula 41. Taula de resultats	51
Taula 42. Taula de resultats	51
Taula 43. Taula de resultats	52
Taula 44. Taula de resultats	52
Taula 45. Taula de resultats	53
Taula 46. Taula de resultats	53
Taula 47. Taula de resultats	53
Taula 48. Impactes en l'anàlisi ambiental	57
Taula 49. Costos d'enginyeria del projecte	59
Taula 50. Costos informàtics del projecte	60
Taula 51. Altres costos del projecte	60
Taula 52. Pressupost final del projecte	61

1. Prefaci

1.1. Origen del treball

El Síndrome d'Apnea Obstructiva del Son (SAOS), internacionalment conegut com a *Obstructive sleep apnea syndrome* (OSAS), consisteix en repetits episodis d'obstrucció de la via aèria superior (VAS) durant el son, amb pauses respiratòries de 10 a 30 segons, durant les quals es deixa d'introduir aire en els pulmons.

Aquestes pauses d'introducció d'oxigen constants generen episodis d'hipòxia, provoquen a curt període moments d'asfíxia, desvetllaments freqüents i insomni, la necessitat reincident d'orinar o el reflux gastroesofàgic. Amb tot plegat la persona afectada no acaba de descansar bé i té una mala qualitat de vida, a més de ser més propens a patir accidents domèstics, laborals, de conducció, etc. Però aquesta falta d'oxigenació també pot crear conseqüències greus a llarg termini, com poden ser l'ictus, l'infart de miocardi o fins i tot el càncer, segons diversos estudis recents.

Es per tots aquests motius que neix aquest projecte, amb la idea original d'analitzar senyals bioacústics enregistrats durant l'estudi del son conegut com a Polisomnografia (PSG) a l'Hospital universitari Germans Trias i Pujol a una cinquantena de pacients candidats de patir OSAS. A partir d'aquest conjunt de dades, es proposa un model de caracterització del ronc per comprovar si partint únicament amb un enregistrament sonor es pot identificar la presència del SAOS.

1.2. Motivació

La motivació d'aquest treball parteix de la unió de dos aspectes:

- El primer és la idea d'utilitzar algunes de les diferents eines i coneixements adquirits al llarg del grau acadèmic com la programació en Python.
- El segon és la idea de treballar sobre la millorar en la detecció d'aquesta síndrome, tant present dins del nostre context socio-geogràfic, amb la idea d'agilitzar la feina de l'especialista, sabent les conseqüències perjudicials que es poden arribar a patir per la seva presència.

Així doncs, la motivació es concreta en el desig d'arribar a crear una possible eina de millora per al sector de la salut, un camp on, evidentment, les innovacions i millores no s'haurien d'aturar mai.

1.3. Requeriments previs

Per a la realització d'aquest projecte s'ha necessitat uns coneixements bàsics en diferents plataformes informàtiques com són el Python i el WEKA.

El meu co-tutor, Gerard Escudero Bakx, és qui ha facilitat totes les eines, material i coneixements relacionats amb la programació. També ha introduït la gran eina de treball per a processar i estudiar grans quantitats de dades: el programa de mineria de dades WEKA de la Universitat de Waikato.

1.4. Planificació: Diagrama de Gantt

Aquest projecte està estructurat en una part teòrica i una part pràctica.

La planificació del projecte s'ha realitzat per tal d'exposar el temps de dedicació previst per a les diferents tasques al llarg del temps total determinat. Primer de tot ha estat necessari familiaritzar-se amb el tema, és a dir, buscar informació sobre el que s'ha realitzat fins ara i sobre les limitacions actuals, per tal d'adquirir els coneixements bàsics i així poder treballar amb propietat sobre el tema.

A continuació, s'ha hagut de crear un esbós dels procediments per determinar quins són les eines i coneixements necessaris per poder desenvolupar aquest treball.

Finalment, es va desenvolupar l'estudi amb prou temps per poder afrontar possibles aspectes sobrevinguts que sempre poden aparèixer posteriorment, i que també s'han de realitzar obligatòriament.

El diagrama de Gantt següent s'ha utilitzat per facilitar l'organització de les diferents tasques:

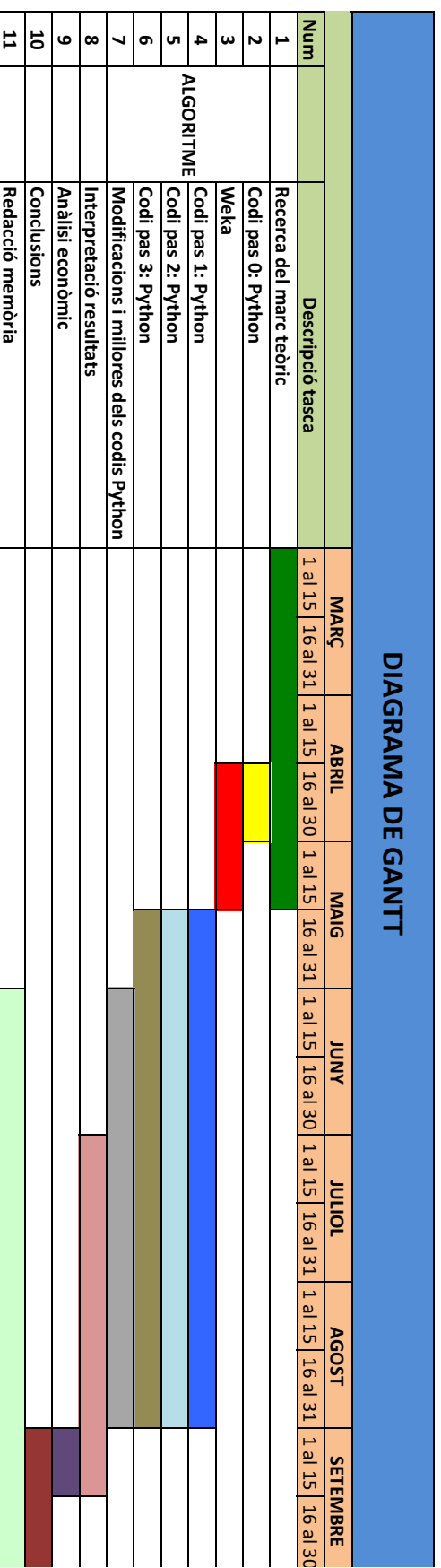


Figura 1. Diagrama de Gantt

2. Introducció

Segons P. Lloberes, J *et al* (2011), a Espanya entre el 3 i 6% de la població pateix SAOS i d'aquests pacients, entre un 24 i 26% tenen un valor major de 54 al índex d'apnea i hipopnea (IAH). Actualment es pren com a llindar del nivell sever un IAH major de 30 apnea/hipoapnees per hora de son [1].

Com defensa R. Henry Olivi *et al* (2013), la continua falta d'oxigenació produïda per les obstruccions a la via aèria a causa del SAOS està relacionada amb l'aparició de molts problemes com accidents vasculars cerebrals (ACV), malalties cardiovasculars, ictus o infart de miocardi [2].

A curt termini, aquestes obstruccions seguides d'un intent per inspirar provoquen no tenir un descans de qualitat a causa del son agitat, possible insomni i canvis de personalitat, essent més propens a patir accidents domèstics, laborals o de trànsit a causa d'aquesta somnolència.

També comencen a aparèixer estudis sobre la relació en l'aparició de càncers i problemes d'edat avançada, com explica el farmacòleg Edwin Jackson, de la Universitat de Pittsburg: "Esto es lo que sucede: los tumores pueden crecer con tanta rapidez que se quedan sin suficiente riego sanguíneo, lo que crea una deficiencia de oxígeno. Esto, a su vez, puede alentar a que las células cancerosas produzcan una molécula llamada adenosina, que inactiva los linfocitos, las células que combaten el cáncer" en un estudi on es demostra que els tumor disminueixen més ràpid en el grup de prova amb més concertació d'oxigen [3].

En nens, per exemple, també s'està demostrant que la presència d'aquesta síndrome des de ben petits pot portar a la derivació problemes neurològics, provocant el desenvolupament de problemes cognitius-conductuals [4].

És d'aquí on neix l'interès en intentar millorar la detecció d'aquest símptoma en la prova del son: del desig d'ajudar a l'especialista a millorar la qualitat de vida als pacients que pateixen l'apnea del son greu, i alhora reduir el màxim possible les conseqüències a llarg termini.

L'objectiu d'aquest projecte és crear un entorn d'estudi capaç de detectar la severitat del SAOS a partir de les úniques mostres de les quals es parteix: els roncs, enregistrats durant el son del pacient, amb la intenció d'agilitzar el procés de detecció i diagnosi actual per tal de millorar la qualitat de vida dels pacients.

2.1. Objectius del treball

Objectius generals

El principal objectiu d'aquest projecte és dissenyar un entorn d'estudi que permeti identificar la severitat del síndrome de l'apnea obstructiva del son mitjançant mètodes d'aprenentatge automàtic, a partir d'una gran quantitat de roncs enregistrats en forma de senyal bioacústica en el decurs de la prova del son anomenada Polisomnografia (PSG).

El repte neix en el format heterogeni de les dades de partida, si és possible crear aquest entorn a partir d'únicament la representació dels roncs. També és important estudiar el comportament dels algorismes de classificació i regressió utilitzats per avaluar la seva viabilitat.

Així doncs aquest projecte té dos objectius:

- Creació d'un entorn que permeti identificar la severitat del SAOS a partir únicament de l'enregistrament sonor dels roncs.
- Estudiar el comportament dels algorismes de classificació i regressió en aquest estudi.

Objectius específics

Amb la finalitat d'aconseguir els objectius principals del projecte es plantegen els següents objectius específics:

- Obtenir coneixements sobre el ronc i l'apnea del son (SAOS).
- Estudiar i aplicar programari de *clustering*.
- Desenvolupar un entorn de treball mitjançant Python.
- Aprendre i aplicar eines d'aprenentatge automàtic.

2.2. Abast del treball

El projecte presentat en aquesta memòria és un estudi centrat a aconseguir un model de detecció de la severitat del Síndrome d'Apnea del Son a partir de l'enregistrament sonor dels roncs. Així doncs, l'abast d'aquest projecte s'ha delimitat a aconseguir crear aquest entorn digital a partir de la

gravació dels roncs de 51 pacients, i finalment s'implementen uns algorismes de classificació per avaluar quin seria el mètode més adequat per aquest tipus d'estudi.

3. MARC TEÒRIC

3.1. El ronc

El ronc és el so que es produeix quan existeix una obstrucció parcial al pas del aire a la zona posterior de la boca i el nas. En aquesta zona és on coincideixen la llengua amb el paladar tou i la campaneta. I allí el pas de l'aire per la gola fa vibrar els teixits del seu voltant quan aquestes estructures xoquen unes amb les altres, produint-se una vibració durant la respiració, i generant el soroll típic del ronc [5].

Els principals motius d'aquesta obstrucció solen ser:

- baix to muscular en els músculs de la gola o la llengua
- uns músculs fluixos produeixen que la llengua caigui cap enrere i que amb els músculs de la gola col·lapsin les vies respiratòries
- tenir gran quantitat de teixit gras a la gola
- un to muscular està molt relaxat pel consum d'alcohol, medicaments o drogues

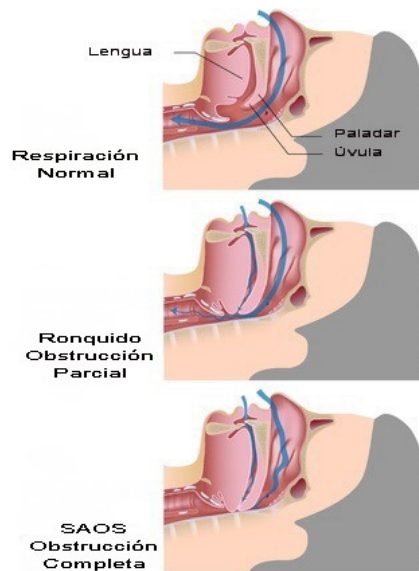


Figura 2. Obstrucció del VAS ([6])

El problema dels roncs és que a la llarga poden derivar en un altre trastorn més important: l'apnea

obstructiva del son. Es per això que actualment el ronc serveix com el principal indicador en la detecció de la presència de SAOS [7]. Si es compta amb la presència de roncs així com també amb alguns dels símptomes del SAOS que ja s'han comentat en l'apartat anterior, convé acudir a un professional especialitzat en aquesta matèria, per tal que revisi la història clínica completa del pacient, faci una exploració visual de la via respiratòria superior i estudiï si cal fer-li una PSG.

Com que encara no hi ha un criteri clar sobre quin és el nivell a partir del qual un s'ha de començar a preocupar, es recomana visitar l'especialista, per evitar problemes generats a la llarga per la presència de desordres del son, en el cas que els possibles pacients presentin alguns d'aquests símptomes:

- somnolència, son no reparador, fatiga
- despertar amb falta d'aire
- quan un testimoni descriu: ronc habitual i/o interrupcions de la respiració durant el son
- antecedents d'hipertensió arterial, trastorns de l'ànim, malaltia arterial coronària, infart de miocardi, o diabetis Mellitus II.

La part pràctica d'aquest projecte, llavors, ha treballat sobre els roncs, ja que, aprofitant l'estreta relació entra aquesta patologia nocturna i ells, s'ha creat un mètode de treball per identificar la severitat de SAOS, mitjançant la combinació de varies eines informàtiques, i amb l'única entrada de l'enregistrament d'aquest senyal biomèdic sonor: els roncs.

3.2. Síndrome d'Apnea Obstructiva del Son

La Síndrome d'Apnea Obstructiva del Son (SAOS), internacionalment coneguda com a *Obstructive Sleep Apnea Syndrome (OSAS)*, consisteix en episodis repetits d'obstrucció de la VAS durant el son, amb pauses respiratòries de 10 segons o més, en què es deixa d'introduir aire en els pulmons i, consegüentment, s'atura l'oxigenació del cervell [8].

Aquestes obstruccions repetitives del flux d'aire a la VAS a nivell de la faringe, i que es produeixen durant el son, poden ser completes (apnees) o parcials (hipoapnees). Solen estar facilitades per alteracions anatòmiques i funcionals en la VAS, que provoquen que estigui més col·lapsada que en els subjectes normals.

Les apnees obstructives es defineixen com l'absència o reducció del senyal respiratori de més de 10 segons de durada, i amb presència d'un esforç per respirar. Les hipoapnees es defineixen com la reducció clarament distingible del senyal respiratori de més de 10 segons de duració i que tingui com a mínim una disminució del 3% de la saturació d'oxigen [1].

La suma d'aquests dos valors forma l'índex d'apnees hipopnees (IAH), que constitueix, en definitiva, el paràmetre més utilitzat per valorar la severitat del SAOS [8]. A la taula 1 es pot observar l'índex IAH i la saturació mínima d'oxigen en pacients adults.

Intensitat	IAH	Saturació mínima
Lleu	5-20	85 %
Moderat	20-40	75-80 %
Sever	>40	< 75 %

Taula 1. Índex IAH

Les apnees i hipoapnees recurrents condueixen a dessaturacions repetides de l'oxihemoglobina, determinant hipòxia intermitent, micro-despertars amb fragmentació del son i oscil·lacions significatives en la pressió intra-toràcica a causa de l'augment del esforç inspiratori en cada episodi d'apnea. Això està associat a increments sobtats del to simpàtic, de la freqüència cardíaca i la pressió arterial, responsables tots ells de la disfunció endotelial [2].

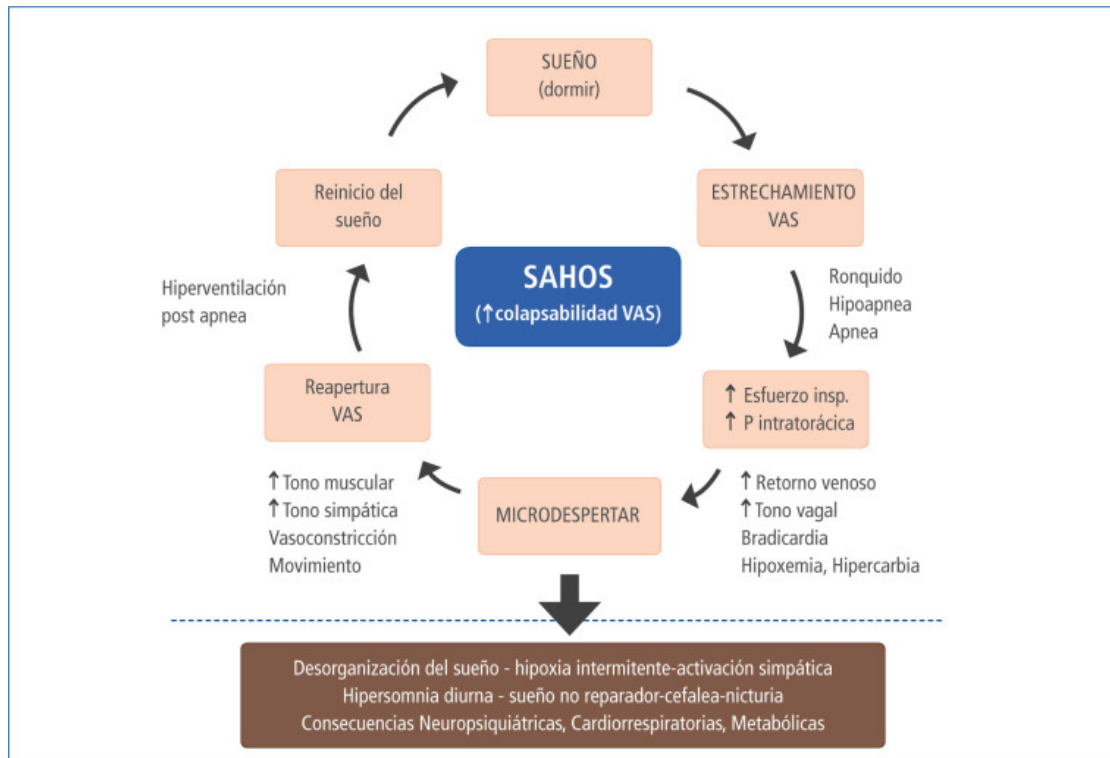


Figura 3. Diagrama d'obstrucció del VAS. ([2])

Aquestes parades provoquen un descens de l'oxigenació i se'ls associa, a curt termini, amb les alteracions inflamatòries a la via respiratòria a causa de l'intent de respirar, episodis d'asfíxia, despertar freqüent, insomni i malsons, necessitat d'orinar o reflux gastro-esofàgic, provocant tot plegat que no s'acabi de descansar de bé, convertint-se en més propens a patir accidents domèstics, laborals o de trànsit [2].

A part de no deixar que l'individu pugui descansar completament, empitjorant la seva qualitat de vida, també se li associa a llarg termini efectes perjudicials sobre el sistema cardiovascular [9,10], sobre el sistema nerviós [11,12] i sobre el metabolisme [13] a causa de la falta reiterada d'oxigen al cervell, podent patir conseqüències greus com càncer, ictus o infart de miocardi, aspectes que poden provocar una mort prematura. La simptomatologia i les conseqüències biològiques associades al SAOS resulten de la hipòxia intermitent, canvis de pressió toràcica, activació simpàtica i fragmentació del son, provocant una desregulació metabòlica, defunció endotelial, inflamació sistèmica, estrès oxidatiu, híper-coàguls i canvis neuronals. Tots aquests poden originar o agreujar malalties com hipertensió arterial, pulmonar, insuficiència cardíaca, arítmies, i fins i tot produir la

mort sobtada [14].

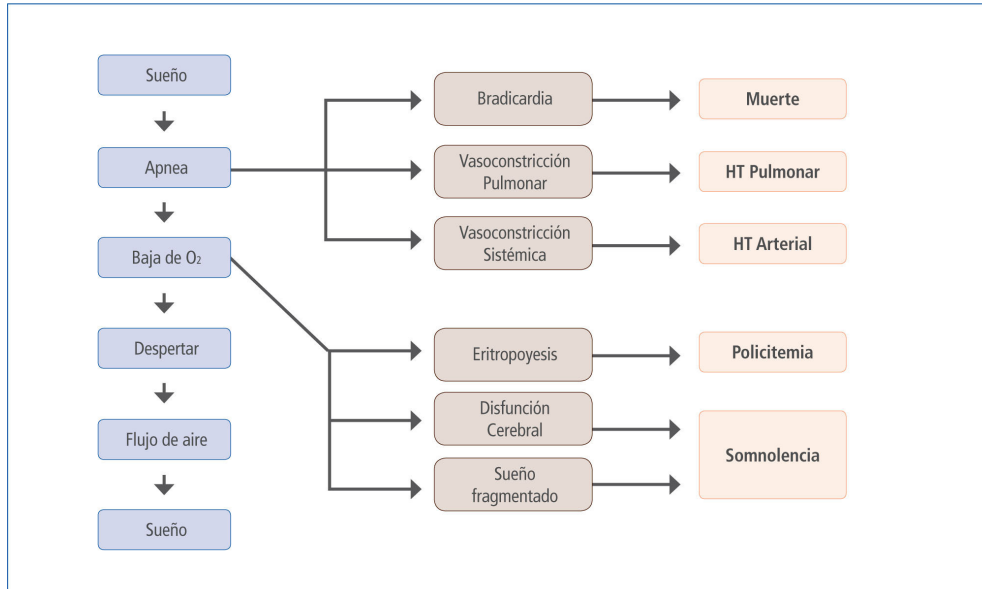


Figura 4. Fisiopatologia del SAOS ([14])

Es calcula que l'apnea afecta un 4-6 % d'homes i 2-4% de les dones [15,16], i entre els factors de risc es troben l'edat (el risc augmenta un 25% en majors de 65 anys); l'obesitat (la pateixen el 60% dels pacients amb apnees [16]); el consum d'alcohol, tabac i sedants; l'ús de medicaments per agafar el son, dormir en una posició supina o tenir obstruïdes les vies respiratòries per la llengua o les amígdales o per alteracions de l'anatomia craniofacial i el coll; la menopausa, l'hipotiroïdisme, o els antecedents familiars són els principals causant de l'aparició d'aquestes obstruccions.

Es tracta de la tercera malaltia respiratòria més freqüent, després de l'asma bronquial i la malaltia pulmonar obstructiva crònica. Tot i que és l'alteració més freqüent que ocorre durant el son, i també la principal causa d'hipòxia crònica en éssers humans, degut a la hipòxia intermitent [17].

EL 35% d'aquests pacients també presenten hipertensió arterial sistèmica, el doble d'hipertensió pulmonar, quatre vegades més malalties vasculars cerebrals que la població normal i la taxa d'accidents augmenta set vegades per la mala qualitat de vida, no descansar suficient, etc. [17].

La major mortalitat documentada es troba en joves de 30 a 50 anys i es deu principalment a complicacions cardiovasculars i accidents domèstics, de trànsit, laborals i industrials, causada principalment pels micro-despertars dels SAOS [18].

Un estudi realitzat sobre nens entre 7 i 12 anys , als quals se'ls hi ha realitzat un PSG, comparant els nens amb desordres obstructius del son amb els d'un grup de control sans, demostra amb que els nens amb RP, SAOS lleu i SAOS moderat/sever, presenten, en relació al grup control, un deteriorament significatiu tant en les funcions cognitives com en les habilitats executives i intel·lectuals i acadèmiques [19].

La diagnosi del SAOS sempre comença amb els testimonis de familiars que presencien els seus símptomes més evidents, com són els roncs i les apnees. Tot i que també es donen símptomes diürns, com són la somnolència i el cansament, derivats d'un son que no ha estat reparador [20]. A partir d'aquí, el pacient s'ha de derivar a un metge, que realitzarà un examen físic que començarà amb l'observació de l'estructura anatòmica de la faringe a través de la cavitat bucal.

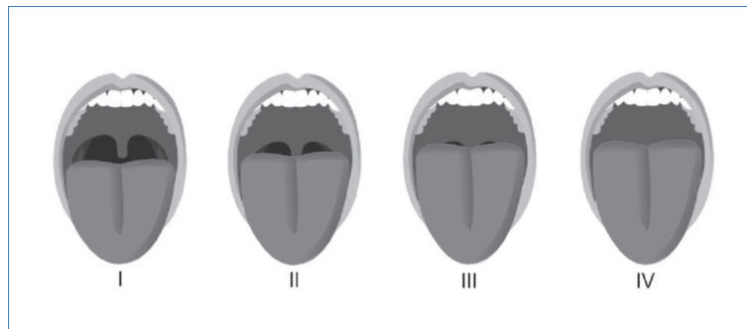


Figura 5. Exploració mèdica visual en sospita de SAOS ([17])

Amb l'examen físic, un estudi del quadre clínic i antecedents familiars [21], s'avalua la necessitat de realitzar una PSG, la prova més utilitzada actualment per a la detecció del SAOS [22].

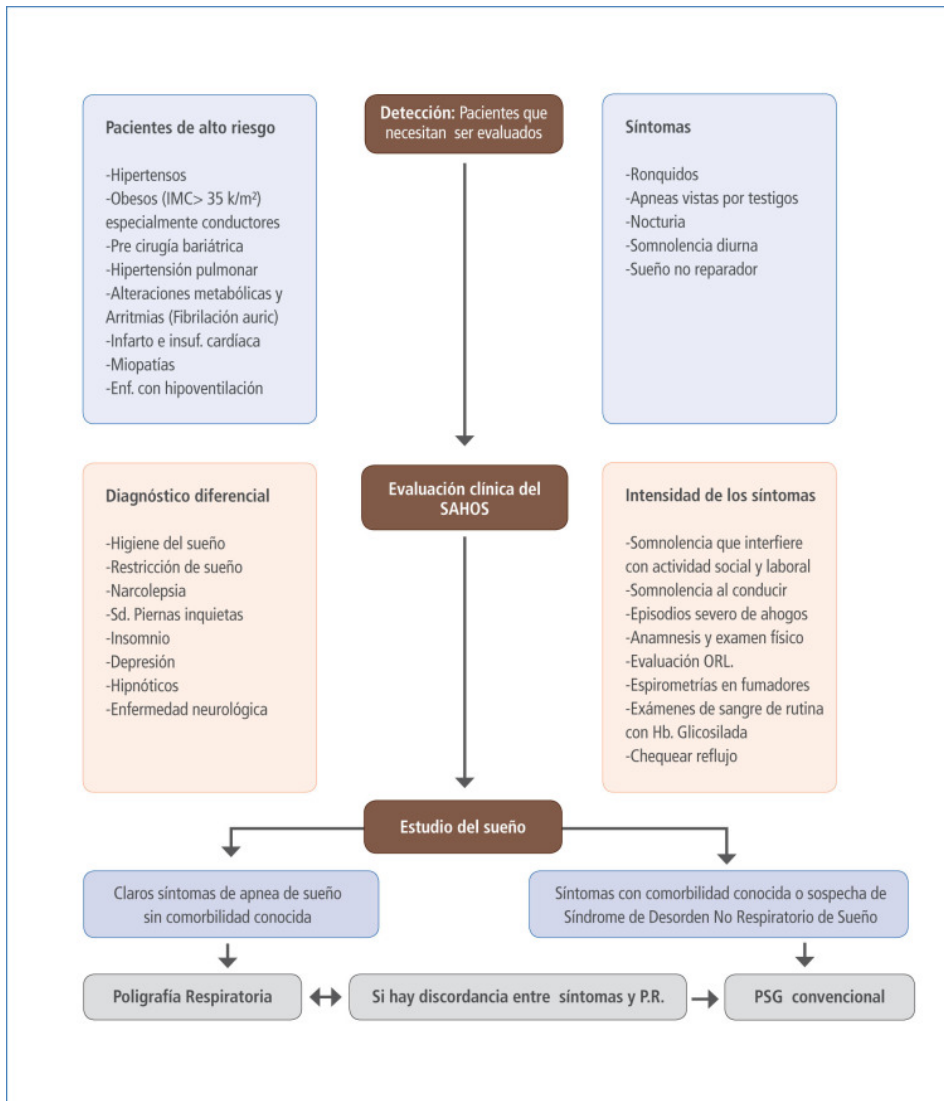


Figura 6. Procediment en el diagnòstic i tractament del SAOS ([17])

Aquesta diagnosi s'ha de realitzar en horari nocturn, o en l'horari de son habitual del pacient, i en un centre que estigui dotat amb personal preparat per supervisar-ho i amb les infraestructures necessàries per a l'observació. Per realitzar aquesta diagnosi es requereix una monitorització continuada d'un mínim de 6,5 hores.

Els principals senyals que s'enregistren durant la PSG són les següents:

- Electroencefalograma

- Electrooculograma
- Electromiograma
- Saturació d'oxigen
- Electrocardiograma
- Esforç respiratori
- Flux aeri naso-bucal
- Posició corporal (vídeo i micròfon)

Actualment, no hi ha cap tractament farmacològic indicat per tractar el SAOS. No obstant, el tractament simptomàtic que s'ha establert, de moment, com a més eficaç és el CPAP (*Continuous Positive Airway Pressure*) [23]. Funciona a partir d'una màscara nasal que dissipa una pressió continua a la VAS, aconseguint que aquesta no es col·lapsi, com s'observa a la figura 9.

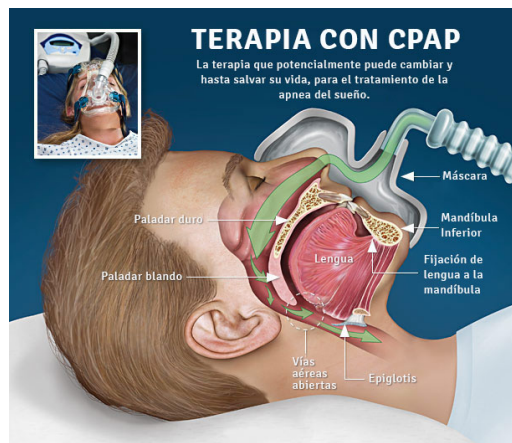


Figura 7. Implementació de la teràpia amb CPAP contra SAOS severos ([24])

4. METODOLOGIA

En aquest capítol es fa una descripció del material de partida, i es presenta un plantejament global amb els passos a realitzar, i les eines a utilitzar. S'ha creat un *FrameWork* amb la intenció de facilitar la comprensió de tots els elements que integren aquest projecte. També es descriuen les eines d'aprenentatge automàtic necessàries per afrontar els diferents tipus d'estudi, com són el **CLUSTERING** i la **CLASSIFICACIÓ** i la **REGRESSIÓ**.

4.1. Material inicials

Les dades de partida d'aquest projecte són un conjunt de característiques extreptes de l'anàlisi espectral de la gravació dels roncs durant tota la nit en 51 pacients amb sospita de SAOS. D'aquí ve l'alt nivell de dificultat a l'hora de intentar crear un model d'estudi que resulti capaç de detectar la presència de SAOS. Aquest material, que consta de dos arxius, prové d'un projecte de recerca realitzat a la UPC en col·laboració a l'Hospital Universitari Germans Trias i Pujol.

El **primer arxiu** conté un sol llistat dels 51 pacients amb informació clínica i de la prova del son de cadascun. Conté l'índex d'apnea-hipopnea, un valor que serveix per delimitar les fronteres dels diferents grups de severitat, per tal de separar els pacients en diferents classes.

ID pacient	Edat	Sexe	BMI	Nº de Roncs	AHI	HI	AI	5 classes
77	51	2	28,76	3717	10,1	10,1	0	Lleu
73	57	2	31,53	2074	10,6	9,48	1,12	Lleu
147	46	1	25,4	1679	10,9	8,41	2,49	Lleu
70	69	2	27,47	2641	11,6	7,49	4,11	Lleu
58	39	1	26,53	2195	12,1	9,45	2,65	Lleu
95	36	1	25,54	1209	14,3	10,4	3,9	Lleu
141	54	1	29,8	2113	14,6	13,74	0,86	Lleu
33	45	2	30,3	3146	15,1	6,94	8,16	Moderat
80	58	1	26,4	1973	17,4	11,94	5,454	Moderat

Taula 2. Exemple de llista dels pacients

El **segon arxiu** conté el resultat de la representació de les senyals bioacústiques enregistrats durant la PSG per a cada pacient. A partir de l'anàlisi espectral s'han obtingut diferents paràmetres característics del ronc, guardats en un document en format .csv, format per les següents columnes:

- numeració, posició i duració de cada ronc
- posició del pacient (supina, decúbit lateral, decúbit pron)
- intensitat sonora màxima i mitja
- freqüència de pic, mitja, central i màxima
- desviació estàndard
- coeficients de simetria i d'aplanament
- rati a 800Hz i 500Hz

Num.	Posició	Duració	Posició	Int. Sonora	Int. Sonora	Freq.	
Esdv.	Esdv.	Esdv.	Pacient	Màxima	Mitja	Mitja	...
0	4544981	5054	3	55.548	46.517	451.220	...
1	4554743	5678	3	45.246	41.255	490.244	...
2	5227765	3985	3	46.361	42.083	426.829	...
3	6108184	7727	3	47.347	41.668	526.829	...
4	7062421	6953	3	47.877	39.166	541.463	...

Taula 3. Exemple caracterització de les 8 primeres mostres de roncs del pacient SIG 33

De totes aquestes columnes de dades s'han seleccionat les més útils i significatives, aquelles que permetin realitzar l'estudi adient i aconseguir determinar quants tipus diferents de roncs poden existir (la numeració i posició del ronc en el temps, per exemple, no aporten informació significativa).

4.2. Esquema de treball: Framework

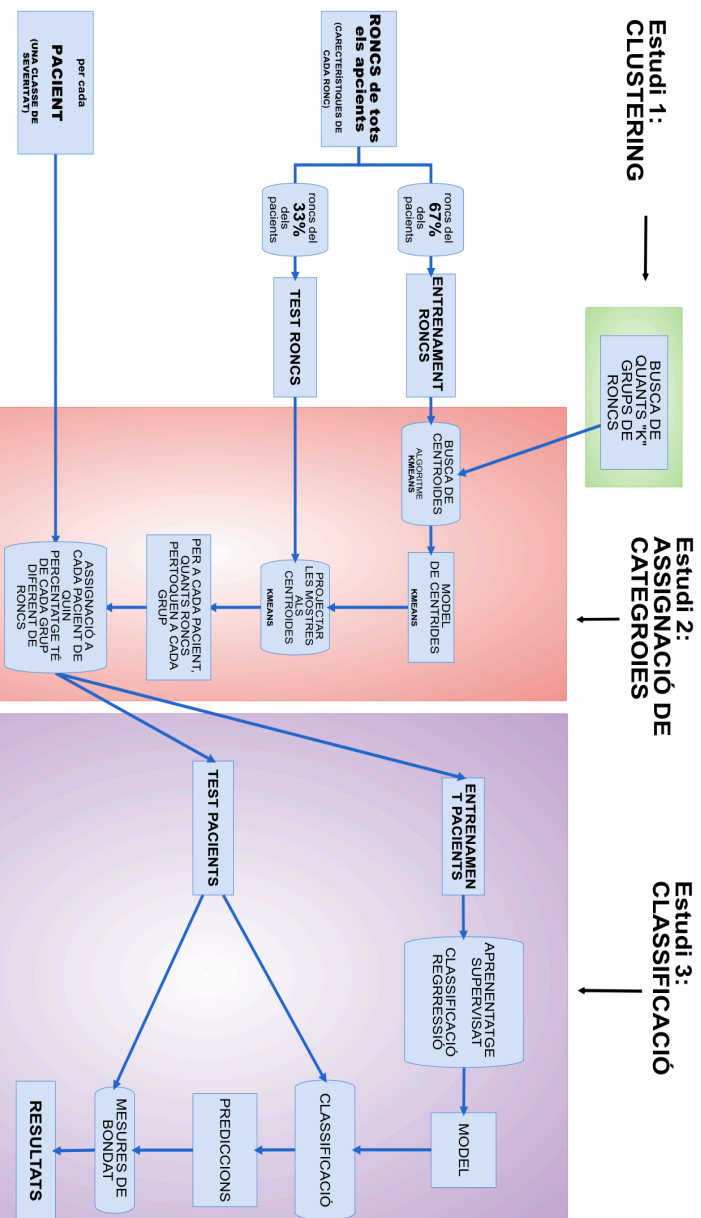


Figura 8. Framework

4.3. Diferents tipus d'estudi

En el *FrameWork* de la figura 8 es mostren els tres processos o estudis diferents que es realitzen en aquest projecte (taula 4).

		Dades	Aprenentatge	Mètode
Estudi 1	Mineria de dades	RONCS	No supervisat	WEKA
Estudi 2	Assignació categories	RONCS	Supervisat	K-MEANS
Estudi 3	Classificació	PACIENTS	Supervisat	Algoritmes de classificació

Taula 4. Tipus d'estudis

El primer procés correspon a un estudi de mineria de dades de tots els roncs, on s'intenta predir en quants grups "K" diferents es poden separar els roncs produïts per un individu durant tota la nit.

El segon procés aplica un estudi d'aprenentatge automàtic no supervisat anomenat **CLUSTERING**, on calcula per cada pacient quin nombre de roncs té de cada "K" grup trobat en l'estudi 1.

El tercer procés és un estudi d'aprenentatge supervisat de **CLASSIFICACIÓ**, aprofitant el càlcul de l'estudi anterior, que crea un model d'estudi per a tots els pacient que permet determinar la severitat del SAOS.

Per tal de tractar els dos tipus de dades i poder executar aquests estudis, s'ha necessitat utilitzar una plataforma de programació i una de mineria de dades.

4.4. Aprenentatge automàtic

L'aprenentatge automàtic és un camp de la intel·ligència artificial dedicat al disseny, l'anàlisi i el desenvolupament d'algorismes i tècniques per permetre que les màquines tinguin capacitat de prendre decisions davant a problemes plantejats. Es tracta de crear programes capaços de generalitzar comportaments a partir del reconeixement de patrons o classificació.

Qualsevol sistema que es consideri intel·ligent ha de tenir l'habilitat d'aprendre, és a dir, de millorar automàticament amb l'experiència. Els programes utilitzats són sistemes d'aprenentatge capaços d'adquirir coneixements d'alt nivell i estratègies per la resolució de problemes mitjançant exemples, de forma anàloga a com ho faria la ment humana.

En aquest sentit, existeixen dos tipus d'aprenentatge automàtic: el supervisat i el no supervisat [25].

- **Aprenentatge supervisat:** L'algorisme utilitzat genera una funció que relaciona cada mostra d'entrada amb un valor conegut i desitjat en la sortida. En el cas de problemes de classificació ens trobem que el sistema ha d'aprendre una funció a partir d'uns exemples que s'expressen en termes de l'aplicació d'un vector a un conjunt de classes. És a dir, es proporciona un conjunt de dades al sistema, que han estat etiquetades de forma manual amb els valors esperats. Les xarxes neuronals (supervisades) i els arbres de decisió són exemples d'algorismes d'aprenentatge supervisat. La base de coneixement del sistema està format per exemples d'etiquetats anteriors, fent una assignació de cada mostra a una de les categories.
- **Aprenentatge no supervisat:** Aquí l'algorisme disposa d'un conjunt d'exemples, però no de la classe que li correspon. Significa que els exemples no estan etiquetats com passa en el cas anterior. En aquest cas, l'algorisme ha de ser capaç de trobar patrons similars per tal de poder etiquetar. Els algorismes d'assignació de categories (*clustering*) són un exemple d'aquests tipus d'aprenentatge.

Pel que fa a l'aprenentatge automàtic supervisat n'existeixen dos tipus: el de **classificació** i el de **regressió**. La diferència radica en què la regressió té l'objectiu de predir valors continus, com s'observa a la figura 9, mentre que la classificació intenta assignar una classe, és a dir, predir a quina classe pertany cada mostra a estudiar (figura 10).

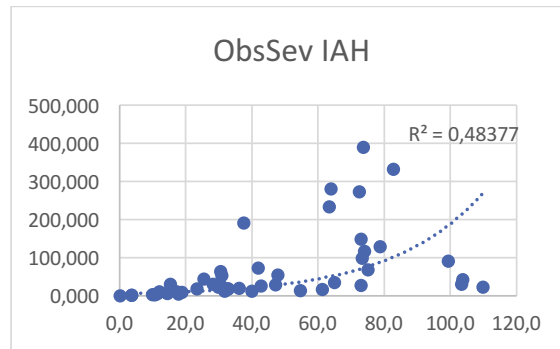


Figura 9. Exemple estudi regressió ([26])

En el tercer estudi d'aquest treball, per una banda s'ha avaluat les prestacions del sistema desenvolupat per a la classificació de pacients en diferents grups o classes de severitat (problema de classificació), i per altra banda per predir el valor de diferents índexs de severitat com pot ser l'IAH (problema de regressió).

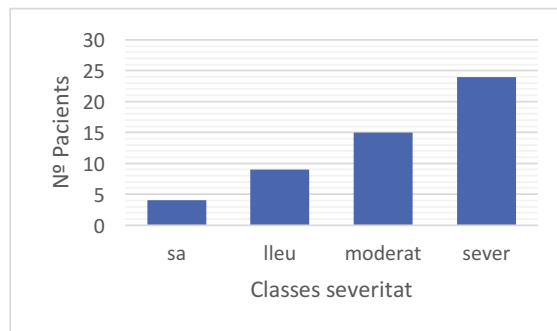


Figura 10. Exemple estudi classificació ([26])

Per finalitzar, tenim el concepte de *clustering*, que és on s'ajunten la idea de l'aprenentatge automàtic i la mineria de dades. Aquest darrer és el que intenta identificar patrons prèviament desconeguts, en comptes de partir de patrons ja coneguts, i posteriorment realitzar prediccions (classificació o regressió).

4.5. Clustering

Com s'indica al *FrameWork* de la figura 8 amb les grans regions de colors, el primer pas ha estat un estudi de *clustering* que s'ha realitzat amb l'aplicació WEKA, part on s'ha identificat quants grups de roncs diferents poden existir en l'enregistrament nocturn d'un individu. Aquest procés, conegut com a

clustering, representa un aprenentatge supervisat. És el pas previ a la utilització dels algoritmes de classificació.

Un cop sabut en quants grups diferents es poden separar els roncs, la funció ***predict(x)*** determina quin dels clústers existents li pertoca a cada ronc [23]. Aquesta informació s'utilitza posteriorment per assignar cada pacient que ja té una classe determinada coneguda, crea un model assignant quina proporció de roncs de cada grup genera cada pacient, explicat al següent capítol.

Amb la creació d'aquest model finalitza el tractament de dades, i és aquí on comença la part avaluable de l'estudi: la d'assignació de categories, la de classificació i regressió. A partir d'aquest model d'estudi s'analitzen diferents algoritmes per determinar quin d'ells funciona millor, partint del tipus d'informació de què disposem. És a dir, s'analitza quin algoritme permet identificar millor la severitat del SAOS que pateix cada individu estudiat.

4.6. Assignació de categories (Algoritme K MEANS)

Un cop sabut en quants grups diferents es podien agrupar els roncs, cal crear un model en un arxiu, on a tots els roncs disponibles se li ha assignat un dels "K" possibles grups trobats.

El mètode d'agrupació utilitzat s'anomena ***K-MEANS***, i treballa repetint aquests passos següents:

1. Es generen "K" centroides aleatòriament dins d'un conjunt de mostres.
2. Es generen "K" grups associant-se cada mostra (ronc) a un dels centroides.
3. El centroide de cada un dels "K" grups es recalcula.
4. Els passos 2 i 3 es repeteixen fins que s'aconsegueixi la convergència.

L'algorisme *KMeans* és, potser, l'algorisme de *clustering* no supervisat més conegut i utilitzat en mineria de dades, tant per la seva senzilla i intuïtiva estructura com per l'eficiència en termes temporals i pels bons resultats que acostuma a donar en nombrosos entorns [27].

A part del conjunt d'instàncies rep com a entrada el número de classes que es volen obtenir i l'increment mínim (que s'utilitza com a lllindar d'aturada de l'algorisme). En algunes versions se substitueix aquest increment mínim per un nombre màxim d'iteracions.

Algorisme *K-Means* :

Entrada:	E: Conjunt d'instàncies K: Enter (Nombre de classes) e: Real (Increment mínim) P: Conjunt de prototips amb les corresponents instàncies
Sortida:	1.- Generar k prototips amb k instàncies escollides aleatòriament 2.- Assignar el conjunt d'instàncies al prototip més proper 3.- $S1 :=$ Suma del quadrat de les distàncies instància - prototips 4.- fer 5.- Recalculer prototips 6.- Reassignar les instàncies al prototip més proper 7.- $S2 := S1$ 8.- $S1 :=$ Suma del quadrat de les distàncies instància - prototips 9.- mentre $S1-S2 \leq e$

Taula 5. Algorisme *K-Means*

L'algorisme després d'haver-li determinat el valor "K" de quant grups diferents existeixen, treballa fent un bucle repetint dos passos de forma iterativa fins a trobar dos iteracions amb els mateixos valors dels centroides, confluint que ha arribat al millor resultat[23].

- Assignació: S'assigna cada objecte al *Cluster* més proper, aplicant alguna mesura de distància (com per exemple la distància euclidiana) entre l'objecte i el centroide del *Cluster*.
- Actualització: Calcula els nous centroides, fent la mitja dels objectes que formen el grup del *Cluster*.

4.7. Classificació i regressió

Finalment s'apliquen diferents algorismes de classificació, en un estudi d'aprenentatge supervisat, amb la intenció d'avaluar si el model creat fins al moment és útil per a l'objectiu plantejat.

Els diferents algorismes utilitzats són [23]:

- **SVM** (*Support Vector Machine*): Les màquines de vectors de suport (SVM) són un conjunt de mètodes d'aprenentatge supervisats utilitzats per a la detecció de classificació, regressió i detecció de punts forts, utilitzant un mètode d'hiperplans, com s'observa a la figura 11 i un subconjunt de punts de formació en la funció de decisió (anomenats vectors de suport) per aconseguir resultats més versàtils i eficaços.

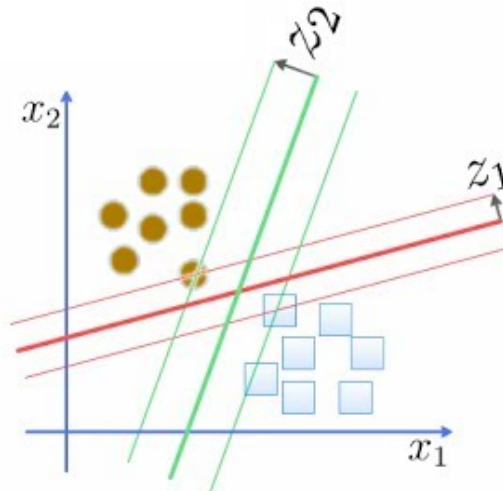


Figura 11. Exemple d'hiperplans del SVM ([28])

- **AdaBoost**: és un classificador que comença ajustant un primer estimador en el conjunt de dades original i després s'adapta a còpies addicionals del classificador en el mateix conjunt de dades, però on els pesos de les instàncies incorrectament classificades s'ajusten de manera que els classificadors posteriors se centrin més en casos difícils.
- **KNeighbors**: útil per a mètodes d'aprenentatge no supervisats i supervisats basats en veïns (KNN). Per l'aprenentatge supervisat basat en veïns es presenta en dos tipus:
 - classificació per a dades amb etiquetes discretes
 - regressió per a dades amb etiquetes contínues.

El principi darrera els mètodes dels veïns més propers és el de trobar un nombre pre definit de mostres d'entrenament més properes al nou punt, i predir l'etiqueta d'aquestes. El nombre de mostres pot ser una constant definida per l'usuari, o també pot variar en funció de la densitat local dels punts.

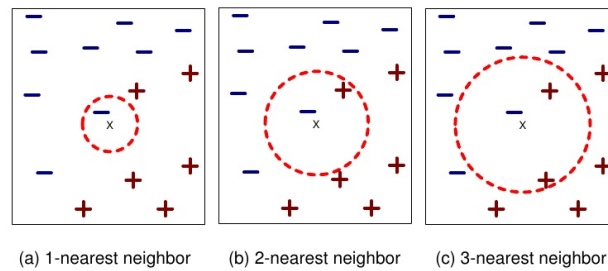


Figura 12. Exemple d'aplicació K-neighbors ([29])

La distància pot ser, en general, qualsevol mesura mètrica; la distància euclidiana estàndard de la fórmula 1 és l'opció més freqüent. Els mètodes basats en veïns es coneixen com a mètodes d'aprenentatge automàtic no generalitzadors, ja que simplement "recorden" totes les dades de formació.

$$d(p, q) = \sqrt{\sum_i (p_i - q_i)^2} \tag{1}$$

- **GaussianNB:** conjunt d'algorismes d'aprenentatge supervisats, basats en l'aplicació del teorema de Bayes (basat en probabilitats). Tot i ser un mètode molt ràpid, i en suposar el desacoblament de les distribucions de característiques condicionals de classe, significa que cada distribució es pot estimar independentment com una distribució unidimensional, combatent els problemes derivats de la maledicció de la dimensionalitat. En aquest sentit també és un mal estimador, de manera que els resultats de probabilitat de *predict_proba* no s'han de considerar com a precisos.
- **DecissionTree:** Els arbres de decisió (DT) són un mètode d'aprenentatge supervisat no paramètric utilitzat per a la classificació i la regressió. L'objectiu és crear un model que predigui el valor d'una variable objectiu mitjançant l'aprenentatge de regles de decisió simples inferides a partir de les funcions de dades. L'algoritme bàsic per a la construcció d'arbres de decisió anomenat ID3 [30]. Construeix un arbre de decisió de dalt a baix, de manera directa, sense fer ús de *backtracking*, i basant-se únicament en els exemples inicials proporcionats. Per a això, utilitza el concepte de Guany d'Informació per seleccionar l'atribut més útil en cada pas. En certa manera, segueix un mètode voraç per decidir la pregunta que major guany dona a cada

pas, és a dir, la que em permet separar millor els exemples entre si pel que fa a la classificació final.

Per poder aplicar l'algorisme s'ha de començar sabent com es mesura el guany d'informació, i per a això cal introduir el concepte d'entropia, formula 2, que mesura el grau d'incertesa d'una mostra.

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i \quad (2)$$

- **Random Forest:** L'algorisme *Random Forest* és un algorisme de classificació supervisada. Com el seu nom indica, aquest algorisme crea blocs amb diversos arbres de decisió amb diferents algorismes i tècniques de probabilitat. Com més arbres es troben al bloc, més robust és. De la mateixa manera, com més gran és la quantitat d'arbres es donaran els resultats amb més precisió.

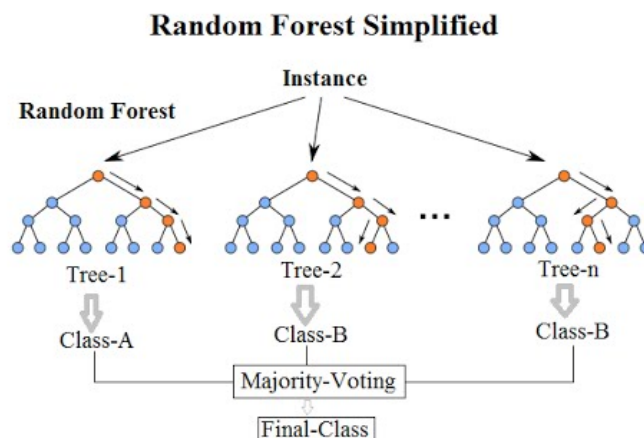


Figura 13. Exemple de Random Forest ([31])

4.7.1. Eina de selecció de paràmetres (Grid Search)

Cada un d'aquests classificadors utilitza molts paràmetres i cada un d'ells pot treballar amb un rang de valors molt ampli. Com que es desconeix a priori quin és el valor òptim per cada paràmetre, i com que seria molt laboriós provar-los tots, s'ha utilitzat una eina que incorpora la plataforma Python anomenada *Grid Search* [32]. Aquesta eina permet estudiar quin és el valor més adequat dintre un rang determinat, per a cada un dels diferents paràmetres de l'algorisme de classificació utilitzat.

Així doncs, amb l'ajuda d'aquesta eina s'ha esbrinat quin valor de cada paràmetre és el més adient en cadascun dels algorismes utilitzats en aquest estudi.

4.8. Eines informàtiques

Les diferents eines utilitzades en la creació i implantació d'aquest projecte són les següents:

4.8.1. WEKA

WEKA [33] (*Waikato Environment for Knowledge Analysis*) és un programa de domini públic escrit en codi *Java* que pot utilitzar tothom. S'utilitza per realitzar experiments que permet comparar grans quantitats de conjunts de dades de manera automatitzada.

Aquesta plataforma sol treballar amb dos formats d'arxius diferents, l'anomenat *.arff* acrònim de *Attribute-Relation File Format*, i el format que s'ha utilitzat que és el *.csv* (document de text separat per comes i en la primera fila es troben els atributs). En aquest projecte s'utilitzen documents en format *.csv* que contenen el conjunt de característiques quantificades per a cada ronc enregistrat de cada pacient.

Com es pot observar a la figura 14 els dos formats de treball que ofereix són el *Explorer* i el *Experimenter*.



Figura 14. Portal del WEKA ([22])

El mode *Explorer* és el més usat i el més descriptiu. La primera pestanya, la de *Preprocess*, i tal com es pot observar a la figura 15, serveix per visualitzar en un inici les mostres, i per aplicar diferents tipus d'algorismes a un conjunt de dades, on inicialment se li fa un procés per visualitzar els conjunts de dades amb una possible aplicació d'alguns filtres, si interessés (tot i que no ha estat el cas). En les altres

pestanyes podem classificar-los amb l'aplicació d'algoritmes de classificació i regressió, agrupar-los en grups amb l'aplicació de *Cluster*, Associació, Selecció d'atributs, i realitzar una visualització final de totes les dades per parelles d'atributs.

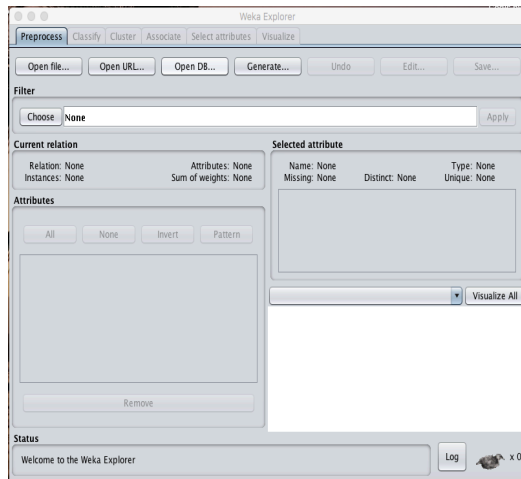


Figura 15. Portal del Explorador del WEKA ([22])

El *Experimenter* serveix per aplicar diferents algoritmes d'aprenentatge automàtic sobre diferents conjunts de dades i determinar de manera estadística quin d'ells es comporta millor. És a dir, aquest ens dirà si les diferències aparents en percentatges d'encerts de diferents algoritmes són estadísticament significatius, o són deguts a l'atzar.

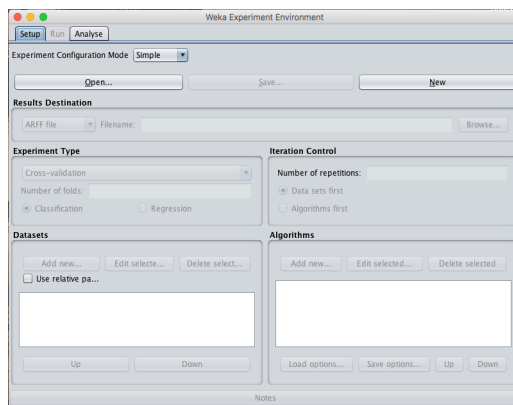


Figura 16. Portal del Experiment del WEKA ([22])

Com es pot observar a la figura 16, les fases d'ús de l'*Experimenter* són *Setup* (configurar), *Run* (executar) i *Analyse* (anàlisi estadística).

El procediment pel seu ús és el següent: primer de tot es configura l'experiment, decidint si serà de classificació o de regressió, especificar la itinerància de control i el tipus d'algoritme que volem aplicar. Seguidament, es carrega l'experiment a la pestanya de *Run* amb el botó de *Start*. Per últim, amb la pestanya d'*Analyse* es pot realitzar l'anàlisi estadística.

En aquest projecte es treballa amb la part d'*Explorer*, concretament a la pestanya de *Clustering*. I ho he fet per realitzar una estudi d'agrupament, per determinar quants conjunts de roncs diferents els podria diferenciar.

4.8.2. PYTHON

Python és un llenguatge de programació interpretat, funcional, orientat a objectes i interactiu. Neix amb la intenció de crear una plataforma de programació més visual i fàcil de programar. Python té la poderosa capacitat de comptar amb una sintaxi molt neta i clara, és a dir, altament llegible i comprensible per a l'ésser humà. Compta amb mòduls, classes, excepcions, tipus de dades de molt alt nivell així com tipus dinàmics. Hi interfases per fer crides al sistema i llibreries, així com diversos sistemes de finestres. Per aquest tipus d'estudi ha estat necessari la instal·lació de dues llibreries específiques, la *Numpy* i la *Scipy* [23].

En el meu treball ha sigut imprescindible el Python per dos motius:

- Crear diferents codis de programació format Python per tractar i modificar la gran quantitat d'arxius de mostres, com és per aconseguir un sol document amb totes les dades de tots els roncs de tots els pacient estudiats, per a la realització de l'estudi de *clustering* amb el WEKA.
- Capacitat d'executar els diferents algoritmes de classificació i regressió

Aquests arxius de programació codi Python s'anomenen *Pas0.py*, *Pas1.py*, *Pas2.py* i *Pas3.py*, i estan adjuntats a l'annex B.

5. Experiments

En aquest capítol s'expliquen els diferents formats que han pres cada un del tres estudis que s'han desenvolupat. Amb aquests tres estudis (taula 4) es pretén avaluar els següents punts:

- Estudiar i verificar la validesa de les dades (pre-tractament).
- Determinar amb el WEKA quants grups K diferents de roncs existeixen.
- Crear un model d'estudi per a K grups. Posteriorment per a K+1 grups i K-1 grups per als tres models creats fer un estudi de regressió i classificació segons els diferents índexs (AHI, HI i AI) separant els pacients en 5,4,3 o 2 classes de severitat diferents.
- Repetir l'estudi de classificació dels pacients ordenats per un nou índex Severitat Obstrucció.
- Decidir quin algoritme és més viable per identificar la severitat del SAOS.

L'objectiu de fons és veure si és possible realitzar aquesta identificació a partir únicament de l'anàlisi del ronc. I en cas afirmatiu, saber quin algoritme i quina combinació d'eines i paràmetres serien els més adients.

A l'esquema de la figura 17 es poden observar les diferents parts de l'estudi, i com aquestes es relacionen entre si.

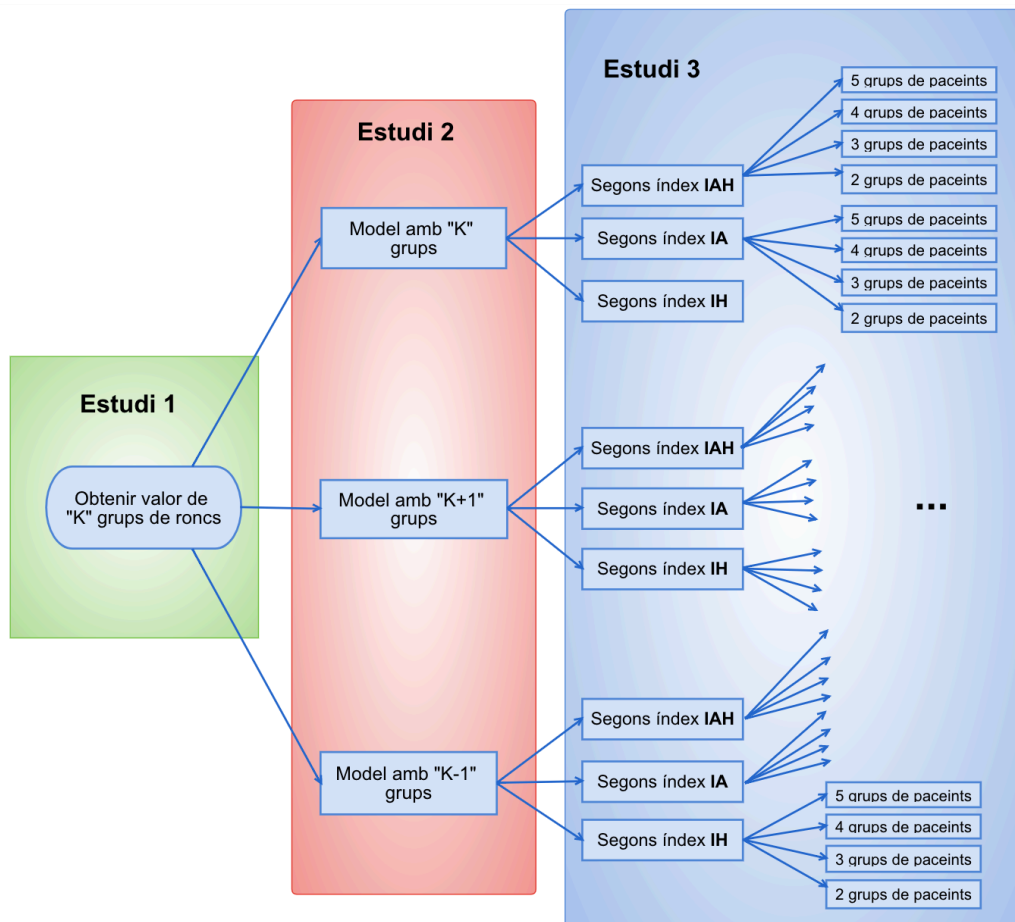


Figura 17. Esquema dels estudis plantejats

En primer lloc es mostraran els resultats de la part de l'estudi 1 de *clustering* marcat en color verd a la figura 17 per determinar, mitjançant la plataforma de mineria de dades WEKA, el nombre de grups de roncs òptim per a realitzar els estudis posteriors .

5.1. Estudi 1

Per a realitzar l'estudi de *clustering*, s'ha necessitat crear dos arxius:

- l'entrenament format pels roncs del 66% dels pacients.
- el de test format pels roncs del 33% dels pacients restants.

Ha calgut assegurar-se que per a cada un d'aquest dos arxius hi haguessin pacients de cada classe de severitat en el moment de realitzar el *clustering*, per tal que el WEKA pugui identificar cada classe de

l'entrenament i poder relacionar-la amb alguna del test. És a dir, que hi havia el risc que si es fes un repartiment totalment aleatori tot un conjunt de pacients d'un grup de severitat (com podria ser els lleus) es trobessin tots dins de l'arxiu test o dins de l'arxiu entrenament, i que llavors fos impossible associar una mostra de l'arxiu test amb un de l'entrenament, ja que no hi hauria cap presència d'algun d'ells. Com s'observa al codi Pas0.py en l'annex B, per assegurar-nos doncs que en aquest repartiment no passes això, es va re-programar el codi afegint-hi aquesta millora. Tot per assegurar que, tot i fer el repartiment aleatori, mantingues la presència d'una part de cada grup tant en l'arxiu entrenament com en el test.

Com s'observa a la figura 18, s'ha fet separant tots els pacients en cinc grups de severitats segons l'índex AHI (sa, lleu, moderat, sever, molt sever). D'aquesta manera, s'ha aconseguit que de cada severitat es repartissin aleatòriament una porció cap a entrenament i una altra cap al test, ja que si s'hagués fet únicament en dues classes, depenent d'on es trobés la frontera llavors a possibles pacients malalts se'ls hauria contemplat com a sans.

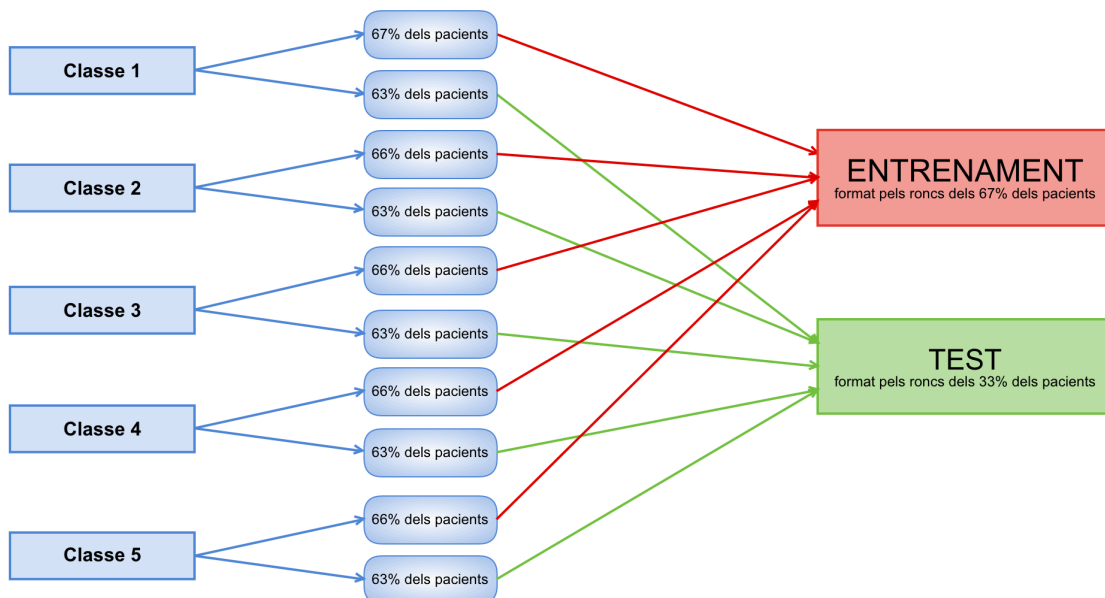


Figura 18. Repartició de pacients per la creació d'arxius d'estudi per mantenir la presència de totes les classes

Com s'observa a la taula 6, S'ha realitzat aquest estudi per a 33 grups d'entrenament i test diferents, i s'ha arribat a concloure que el **nombre de clústers** en que es poden agrupar el roncs, en mitjana, és **8**.

Nº entreno	1	2	3	4	5	6	7	...	29	30	31	32	33	Mitjana
Nº clusters	8	6	8	9	11	7	9	...	7	8	8	10	8	7,9

Taula 6. Resultats estudi mineria de dades WEKA

5.2. Estudi 2

Un cop obtingut un valor de referencia per agrupar els roncs dels pacients en grups, s'ha aplicat el mètode d'agrupació *k-means*, indicant els "K" grups trobats. Com es pot observar a la taula 5, s'ha obtingut un valor orientatiu de 8 grups. Com que a aquest valor no es pot considerar precís, no només s'ha creat un sol model *k-means* de 8 grups, sinó que també s'han estudiat models amb 7 i amb 9 grups.

Aquest estudi el que ha generat és un model segons el valor "K" d'entrada. Per a cada pacient, ha calculat quantes mostres (roncs) pertanyien a cada grup, creant un nou model amb fent una distribució de cada grup de roncs, tal com s'observa a la taula 7.

K = 8									
	IAH	Grup 1 (%)	Grup 2 (%)	Grup 3 (%)	Grup 4 (%)	Grup 5 (%)	Grup 6 (%)	Grup 7 (%)	Grup 8 (%)
Pacient 1	Sa	15	17	21	9	17	11	7	3
Pacient 2	Lleu	9	11	22	19	15	18	3	2
Pacient 3	Sever	3	7	21	10	22	15	20	11
Pacient 4	Sever	4	3	12	16	17	20	18	10

Taula 7. Taula exemple de model generat per l'algorisme d'agrupament *K-Means*

5.3. Estudi 3

La darrera part correspon a l'aplicació i l'estudi de les prestacions de diferents algorismes de classificació i regressió. En aquesta s'han utilitzat tres índexs diferents (IAH, IA, IH), i per a cada índex de severitat,

s'ha fraccionat els pacients en 5, 4, 3 i 2 classes de severitat de SAOS, creant 36 arxius diferents a avaluar amb una combinació de format diferents de cada una com s'observa a la figura 19.

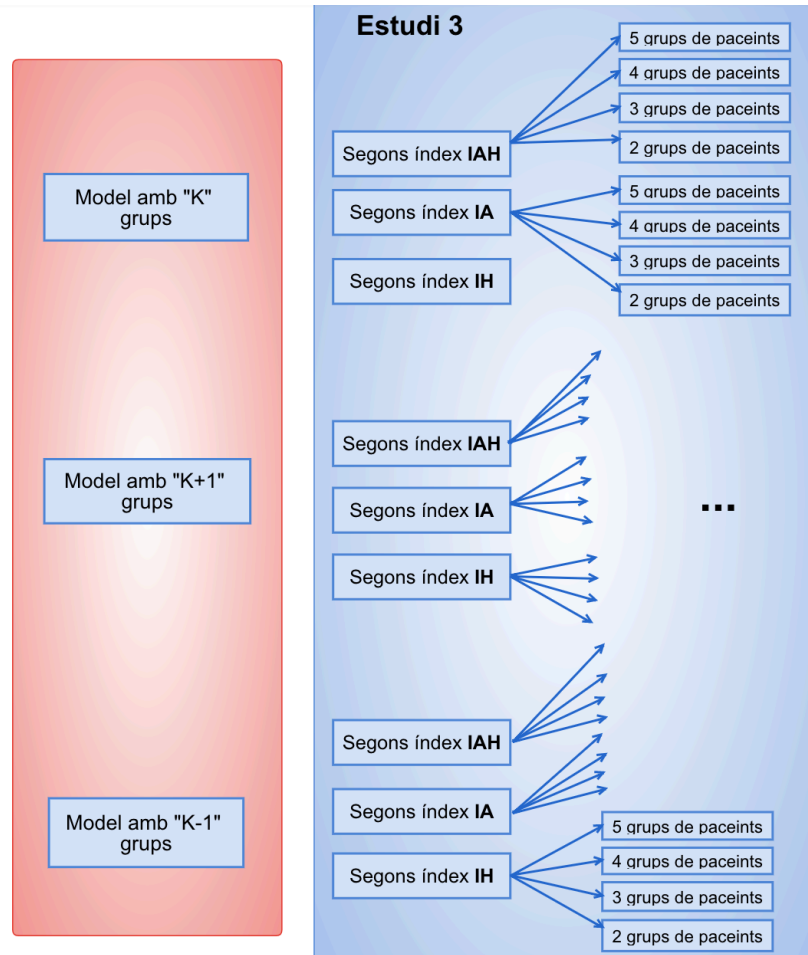


Figura 19. Esquema de l'estudi 3

Com que cada algoritme de classificació pot treballar amb un rang de valors diferent per a cada paràmetre, els algoritmes s'han aplicat conjuntament amb l'eina de triatge *GirdSearch* (comentada al capítol anterior) amb la qual s'escull el valor òptim per a cada paràmetre utilitzat.

Per a cada tipus de combinació, tenint també en compte els tres models diferents de *k-means*, ha calgut crear un arxiu nou per a avaluar. És a dir, per al model *k-means* de 8 grups de roncs, i segons l'índex IAH, surten un arxiu final de 5 severitats diferents, un altre de 4, etc.

Pel model *k-means* de 7 grups s'ha de crear doncs 4 arxius més per a 5,4,3 i 2 classes un altre cop, i pel model *k-means* de 9 grups també igual. S'ha repetit el procés per als altres tres índexs de severitat, i per als diferents models, produint un total de 36 arxius, fruit de les possibles combinacions a estudiar.

Actualment els especialistes utilitzen majoritàriament l'índex apnea i hipopnea (IAH), que implica la suma d'apnees i hipo-apnees per hora de son, separant els resultats en tres classes de severitat, on actualment els especialistes consideren preocupant a partir d'un valor d'IAH superior a 30 (taula 8) [1].

IAH	
Sa	<5
Lleu	5-14,9
Moderat	15-29,9
Sever	>30

Taula 8. Índex apnea-hipopnea en adults

Les classes queden separades segons l'índex IAH com s'observa a la taula 9.

IAH					
	Sa	Lleu	Moderat	Sever	Molt sever
2 classes	IAH<30			>30	
3 classes	IAH<5	5<30		>30	
4 classes		IAH<15	15<30	30<60	>60
5 classes	IAH<5	5<15	15<30	30<60	>60

Taula 9. Fronteres al genera els grups de classes

5.4. Interpretació resultats de l'estudi 3

La consola Python per a cada estudi retorna, conjuntament amb el model de paràmetres òptim que ha escollit entre tots els possibles, també unes mesures de bondat, en forma de valors de classificació i uns altres de regressió (menys per l'algoritme *Gaussian Naïve Bayes*, que només és capaç de retornar els de classificació).

A continuació s'explica en què consisteixen cada un dels conceptes avaluable que s'han obtingut, així com quin hauria de ser el valor òptim per a poder concloure, finalment, quin algoritme treballa millor (figura 9).

	SVM	AdaBoost	K-Neighbors	GaussianNB	Decision Tree	Random Forest
Model de classificació	{'C': 0.001, 'kernel': 'linear'}	{'n_estimators': 100}	{'n_neighbors': 7, 'weights': 'distance'}		{'criterion': 'gini'}	
Acc:	0.39	0.56	0.5	0.39	0.39	0.39
Matriu de confusió:	[0 2 0 1 0] [1 1 1 0 0] [2 0 4 0 0] [0 0 0 1 0] [1 1 1 1 1]	[[0 0 0 1 2] [0 1 0 0 2] [0 1 4 0 1] [0 1 0 0 0] [0 0 0 0 5]]	[[1 0 0 0 2] [0 0 1 0 2] [0 0 5 0 1] [0 0 0 1 0] [0 0 3 0 2]]	[[0 0 1 0 2] [1 0 0 0 2] [1 1 4 0 0] [0 1 0 0 0] [1 1 0 0 3]]	[[0 2 0 1 0] [1 1 0 0 1] [0 2 4 0 0] [0 0 0 1 0] [2 0 2 0 1]]	[[0 0 0 1 2] [1 0 0 0 2] [0 0 5 0 1] [0 0 0 1 0] [1 0 3 0 1]]

Model de regressió:	{'C': 10, 'kernel': 'linear'}	{'n estimadors': 25}	{'n neighbors': 3, 'weights': 'distance'}		{'criterion': 'mse'}	
R ² :	0.204	0.602	0.083		0.348	0.624
Mse:	912.061	301.533	694.026		1020.624	284.420

Classificació:

- **Exactitud:** és el grau d'acostament de mesura d'una quantitat al valor real o veritable. Aquest paràmetre s'avalua en tant per cent, cercant, evidentment, acostar-se al 100%.
- **Matriu de confusió:** és una eina de visualització que s'utilitza en l'aprenentatge supervisat. Cada columna de la matriu representa el nombre de prediccions de cada classe, mentre que cada fila representa les instàncies en la classe real. Un dels beneficis de les matrius de confusió és que faciliten veure si el sistema està confonent dues classes, tal com s'observa a la figura 20.

Taula 10. Exemple de taula de resultats per a l'índex IAH format per 5 grups de severitat

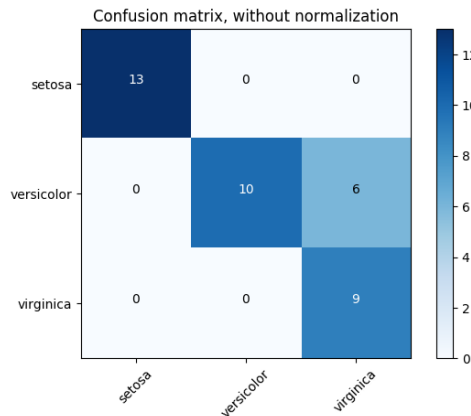


Figura 20. Exemple de matriu de confusió ([23])

Regressió:

- **Valor r quadrat:** també es coneix com el coeficient de determinació, i és un numero entre 0 i 1 que revela en quina mesura els valors estimats per línia de tendència correspon a les seves dades reals. Una línia de tendència és més fiable quan el seu valor r quadrat és a prop de 1. El valor r quadrat multiplicat per 100 s'anomena coeficient de determinació i s'interpreta com el percentatge de variabilitat que explica el model.
- **Error quadràtic mig (mse):** mesura la diferència entre l'estimador i el que s'estima a través de la mitjana d'errors al quadrat. La diferència es produeix a causa de l'aleatorietat, o perquè l'estimador no té en compte la informació que podria produir una estimació més precisa. Al ser una mitjana d'errors, com més elevat sigui el valor més quantitat d'error comportarà.

5.5. Iteracions de l'estudi 3: classificació

S'ha creat una iteració d'estudi per a cada possible combinació diferents o millora de l'estudi, a mesura que anava avançant el projecte.

Iteració 1:

El primer experiment es va executar sense tenir present el problema que he comentat anteriorment pel que fa a la creació dels arxius inicials d'entrenament i test. En aquest cas, quan la repartició és totalment aleatòria hi ha la possibilitat de que totes les mostres d'un cert grau de severitat no quedin repartides una porció en l'entrenament i una altre en el test, fent llavors que a l'hora del *clustering* la plataforma informàtica no pugui identificar els pacients de la mateixa classe.

Aquesta iteració només s'ha estudiat per a l'índex IAH.

Iteració 2:

A continuació es va re-escriure el codi per solucionar aquest problema, per tal que la repartició en els dos arxius fos la idònia, assegurant així la presència en els dos arxius de cada un dels diferents nivells de les mostres. Primer es va aplicar el nou codi de repartició en tres nivells diferents de severitat (lleu-moderat-sever), i després es va repetir però en cinc nivells (sa-lleu-moderat-sever-molt sever). D'aquesta manera, s'assegura que les mostres de cada un dels cinc nivells quedin repartides en cada un dels arxius d'entrenament i test.

Iteració 3:

En aquesta iteració es realitza l'estudi per un índex de severitat anomenat “*Obs.Sev*” proposat recentment a la literatura, que no només comptabilitza el nombre d'apnees i hipopnees, sinó que també té en compte la durada i la severitat de cadascuna d'elles a través de la dessaturació d'oxigen [35].

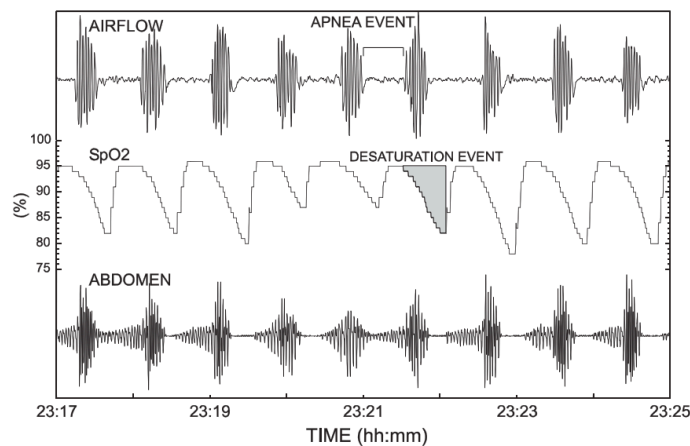


Figura 21. Flux d'aire, saturació d'oxigen en sang (SpO2), moviment de l'abdomen ([35])

Amb dades exportades de l'anàlisi de PSG, observades a la figura 21, el paràmetre de gravetat d'obstrucció que es compon de la suma dels productes d'apnea i duració i hipopnea àrees de dessaturació relacionades normalitzades amb el total analitzat temps, tal com es pot observar a la formula 3 [35].

$$\text{Obstruction Severity} = \frac{\sum_{n=1}^L (\text{HypDur}_n \times \text{DesArea}_n) + \sum_{n=1}^L (\text{ApDur}_n \times \text{DesArea}_n)}{\text{Index time}} \quad (3)$$

Aquest conceptes queden d'escrits a la taula 11:

ApDurn	Durada d'un esdeveniment d'apnea individual en segons.
HypDurn	Durada d'un esdeveniment d'hipopnea individual ens segons.
DesArea	Àrea d'una dessaturació. Es pot observar a la figura 21.
Index Time	Duració del temps analitzat en segons.
L	Nombre d'esdeveniments analitzats.

Taula 11. Conceptes formula 4

Com que no existeixen estudis sobre aquest índex, es creen els grups de mostres amb les fronteres de la taula 12.

	Lleu	Sever	Molt sever
2 classes	50%	50%	
2 classes	66%	33%	
2 classes	33%	66%	
3 classes	33%	33%	33%

Taula 12. Repartició dels pacients en classes

Iteració 4:

Aquesta és una repetició de la iteració 2 però amb la variació en el valor de les fronteres al formar les classes de severitat, com s'observa a la taula 13.

	Sa	Lleu	Moderat	Sever
2 classes	IAH<20			IAH>20
3 classes	IAH<20	20<IAH<40		IAH>40
4 classes	IAH<5	5<IAH<20	20<IAH<40	IAH>40

Taula 13. Repartició dels pacients en classes

Iteració 5:

Aquesta és una repetició de la iteració 2 però amb només amb la variació en el valor de la frontera a la separació en dos classes col·locant-la a 40, com s'observa a la figura 14.

	Sa	Sever
2 classes	IAH<40	>40

Taula 14. Repartició dels pacients en classes

6. Anàlisis de resultats

En aquest capítol s'exposen els resultats aconseguits en les diferents iteracions que ha anat prenent l'estudi 3 en anar avançant en el desenvolupament del projecte i en necessitar millorar, ajustar o estudiar certes parts.

De la gran quantitat de dades de resultat que surten per a cada iteració d'estudi, es mostra només l'algoritme de classificació i regressió més significatiu per a cada combinació estudiada.

Es pot donar que el millor resultat per la regressió i la classificació no es produeixi en el mateix algoritme. Per això es mostra en primer lloc els millors resultats en els algoritmes de classificació i, després, en els de regressió, avaluant-los segons les mesures de bondat definides al capítol anterior.

Iteració 1:

Classificació:

Tal com s'observa a la taula 15, per un model creat determinant 8 grups diferents de roncs, separats els pacients en tres classes de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de classificació que millors resultats ha retornat és l'AdaBoost. El *Grid Search* ha seleccionat un model amb un número de 5 estimadors aconseguint una exactitud del 0,5. La matriu de confusió no és gaire bona perquè un 20% de falsos positius, diagnosticant pacients malalts com a sans.

8 clusters, 3 classes, AHI	AdaBoost
Model Classificació:	{'n_estimators': 5}
Acc:	0.5
Matriu Confusió:	[1 2 2] [1 4 1] [0 3 4]

Taula 15. Taula de resultats

Tal com s'observa a la taula 16, per un model creat determinant 8 grups diferents de roncs, separats els pacients en 2 classes de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de classificació que millors resultats ha retornat és el Decision Tree. El *Grid Search* ha seleccionat un criteri "gini" per aquest model, aconseguint una exactitud del 0,72. La matriu de confusió indica que genera un 20% de falsos positius, detectant pacients malalts com a sans.

8 clusters, 2 classes, AHI	Decision Tree
Model Classificació:	{'criterion': 'gini'}
Acc:	0.72
Matriu de Confusió:	[6 1] [4 7]

Taula 16. Taula de resultats

Tal com s'observa a la taula 17, per un model creat determinant 7 grups diferents de roncs, separats els pacients en 2 severitat de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de classificació que millors resultats ha retornat és el *Random Forst*. Sense la necessitat d'usar el *Grid Search* en aquest algoritme, s'aconsegueix una exactitud del 0,78. La matriu de confusió indica que només genera un 5% de falsos positius, detectant pacients malalts com a sans.

7 clusters, 2 classes, AHI	Random Forest
Acc:	0.78
Matriu Confusió:	[4 3] [1 10]

Taula 17. Taula de resultats

Tal com s'observa a la taula 18, per un model creat determinant 9 grups diferents de roncs, separats els pacients en 2 classes de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de classificació que millors resultats ha retornat és el *Decission Tree*. El *Grid Search* ha seleccionat un criteri "gini" per aquest model, aconseguint una exactitud del 0,72. La matriu de confusió indica que genera un 20% de falsos positius, detectant pacients malalts com a sans.

9 clusters, 2 classes, AHI	Decission Tree
Model Classificació:	{'criterion': 'gini'}
Acc:	0.72
Matriu Confusió:	[6 1] [4 7]

Taula 18. Taula de resultats

Regressió:

Tal com s'observa a la taula 19, per un model creat determinant 9 grups diferents de roncs, separats els pacients en 2 severitat de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de regressió que millors resultats ha retornat a sigut el SVM amb un model escollit pel *Grid Search* de $C=100$, coeficient gamma de 0,0001 i el coeficient kernel el "rbf". El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosta bastant a 0. L'error quadràtic mig és dels menors

8 clusters, 5 classes, AHI	SVM
Model Regressió:	{'C': 100, 'gamma': 0.0001, 'kernel': 'rbf'}
Valor r quadràtic:	-0.056
Error quadràtic mig:	784.290

Taula 19. Taula de resultats

Tal com s'observa a la taula 20, per un model creat determinant 8 grups diferents de roncs, separats els pacients en 4 classe de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de regressió que millors resultats ha retornat a sigut el *Random Forest*. El valor r quadràtic es negatiu pel comportament de la regressió i s'acosta bastant a 0. L'error quadràtic mig no és molt elevat.

8 clusters, 4 classes, AHI	Random Forest
Valor r quadràtic:	0.001
Error quadràtic mig:	741.809

Taula 20. Taula de resultats

Tal com s'observa a la taula 21, per un model creat determinant 7 grups diferents de roncs, separats els pacients en 4 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de regressió que millors resultats ha retornat a sigut el *Decision Tree* amb un model escollit pel *Grid Search* el criteri "mse". El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosta bastant a 1. L'error quadràtic mig és molt alt.

7 clusters, 4 classes, AHI	Decission Tree
Model Regressió:	Model: {'criterion': 'mse'}
Valor r quadràtic:	-0.997
Error quadràtic mig:	1482.998

Taula 21. Taula de resultats

Tal com s'observa a la taula 22, per un model creat determinant 7 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de regressió que millors resultats ha retornat a sigut el *AdaBoost* amb un nombre d'estimadors escollit pel *Grid Search de 10*. El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosta bastant a 1. L'error quadràtic mig és molt alt.

7 clusters, 3 classes, AHI	AdaBoost
Model Regressió:	Model: {'n_estimators': 10}
Valor r quadràtic:	-0.944
Error quadràtic mig:	1443.866

Taula 22. Taula de resultats

Tal com s'observa a la taula 23, per un model creat determinant 7 grups diferents de roncs, separats els pacients en 2 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de regressió que millors resultats ha retornat a sigut el *AdaBoost* amb un nombre d'estimadors escollit pel *Grid Search de 5* i el *Decission Tree* amb un criteri "mse". El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosten bastant a 1. L'error quadràtic mig és molt alt.

7 clusters, 2 classes, AHI	Decission Tree	AdaBoost
Model Regressió:	{'criterion': 'mse'}	{'n_estimators': 5}

Valor r quadràtic:	-0.963	-0.970
Error quadràtic mig:	1457.675	1463.202

Taula 23. Taula de resultats

Iteració 2:

Classificació segons l'índex AHI:

Tal com s'observa a la taula 24, per un model creat determinant 8 grups diferents de roncs, separats els pacients en 5 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de classificació que millors resultats ha retornat és l'AdaBoost. El *Grid Search* ha seleccionat un model amb un número de 100 estimadors aconseguint una exactitud del 0,56. La matriu de confusió es prou bona per el gran numero de classes ja que només un 10% de falsos positius, detectant pacients malalts com a sans.

8 clusters, 5 classes, AHI	AdaBoost
Model Classificació:	{'n_estimators': 100}
Acc:	0.56
Matriu Confusió:	[0 0 1 2] [0 1 0 0 2] [0 1 4 0 1] [0 1 0 0 0] [0 0 0 0 5]

Taula 24. Taula de resultats

Tal com s'observa a la taula 25, per un model creat determinant 8 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex d'apnea-hipopnea (IAH), els algoritmes de classificació que millors resultats ha retornat és el SVM i el *Random Forest*. El *Grid Search* ha seleccionat un model amb una C de 0,001, l'estimador gamma de 0,1 i l'estimador kernel "rbf" aconseguint una exactitud del 0,61 i 0,67 respectivament. La matriu de confusió no es bona ja que amb el SVM els detecta a tots com malats.

8 clusters, 3 classes, AHI	SVM	Random Forest
----------------------------	-----	---------------

Model Classificació:	{'C': 0.001, 'gamma': 0.1, 'kernel': 'rbf'}	
Acc:	0.61	0.67
Matriu Confusió:	[0 0 3] [0 0 4] [0 0 11]	[0 2 1] [0 2 2] [0 1 10]

Taula 25. Taula de resultats

Tal com s'observa a la taula 26, per un model creat determinant 8 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme que millor treballa és el *K-Neighbors* que dona una exactitud del 89% i només amb un 5% de fals positiu.

8 clusters, 2 classes,AHI	SVM	Random Forest	KNeighbors
Model Classificació:	{'C': 0.001, 'gamma': 0.1, 'kernel': 'rbf'}		{'n_neighbors': 1, 'weights': 'uniform'}
Acc:	0.89	0.89	0.89
Matriu Confusió:	[0 2] [0 16]	[0 2] [0 16]	[1 1] [1 15]

Taula 26. Taula de resultats

Tal com s'observa a la taula 27, per un model creat determinant 9 grups diferents de roncs, separats els pacients en 5 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de classificació que millors resultats ha retornat és l'AdaBoost. El *Grid Search* ha seleccionat un model amb un número de 100 estimadors aconseguint una exactitud del 0,56. La matriu de confusió es prou bona per el gran numero de classes ja que només un 10% de falsos positius, detectant pacients malalts com a sans.

9 clusters, 5 classes, AHI	AdaBoost
Model Classificació:	{'n_estimators': 5}
Acc:	0.56
Matriu Confusió:	[0 1 0 0 2] [0 1 0 0 2] [0 1 4 0 1] [0 0 0 1 0] [0 0 0 1 4]

Taula 27. Taula de resultats

Tal com s'observa a la taula 28, per un model creat determinant 9 grups diferents de roncs, separats els pacients en 2 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme que millor treballa és el *AdaBoost* que dona una exactitud del 94% i amb cap fals positiu.

9 clusters, 2 classes, AHI	AdaBoost
Acc:	0.94
Matriu Confusió:	[1 1] [0 16]

Taula 28. Taula de resultats

Classificació per a l'índex HI:

Un model creat determinant 8 grups diferents de roncs, separats els pacients en 2 grups de severitat segons l'índex d'apnea-hipopnea (HI), els algorismes de classificació que millors resultats ha retornat és *K-Neighbors* i el *Random Forest*. El *Grid Search* ha seleccionat un model amb un número de 1 d'estimadors veí i un es uniforme, aconseguint una exactitud del 0,89. La matriu de confusió es prou bona per el gran numero de classes ja que només un 5% de falsos positius, pel *K-Neighbors* i que els detecta a tots com a malalts al *Random Forest* (taula 29).

8clusters,2classes HI	KNeighbors	Random Forest
Model Classificació:	{'n_neighbors': 1, 'weights': 'uniform'}	
Acc:	0.89	0.89
Matriu Confusió:	[1 1] [1 15]	[0 2] [0 16]

Taula 29. Taula de resultats

Un model creat determinant 9 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex d'apnea-hipopnea (HI), els algorismes de classificació que millors resultats ha retornat és SVM. El *Grid Search* ha seleccionat un model amb un número de 0,001 per C, 0,1 per "gamma", aconseguint una exactitud del 0,61. La matriu de confusió no és bona, els detecta tots com a malalts (taula 30).

9 clusters, 3 classes HI	SVM
Model Classificació:	{'C': 0.001, 'gamma': 0.1, 'kernel': 'rbf'}
Acc:	0.61
Matriu Confusió:	[0 0 3] [0 0 4] [0 0 11]

Taula 30. Taula de resultats

Un model creat determinant 7 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex hipopnea (HI), els algoritmes de classificació que millors resultats ha retornat és *K-Neighbors* i el *Random Forest*. El *Grid Search* ha seleccionat un model amb un número de 11 estimadors veí i un es uniforme, aconseguint una exactitud del 0,62 i 0,73 respectivament. La matriu de confusió es prou bona per el gran numero de classes ja que només un 5% de falsos positius pel *Random Forest*, mentre el *K-Neighbors* els detecta a tots com a malalts (taula 31).

7 clusters, 3 classes HI	KNeighbors	Random Forest
Model Classificació:	{'n_neighbors': 11, 'weights': 'uniform'}	
Acc:	0.62	0.73
Matriu Confusió:	[0 0 3] [0 0 4] [0 0 11]	[2 0 1] [0 1 3] [0 1 10]

Taula 31. Taula de resultats

Regressió:

Per un model creat determinant 9 grups diferents de roncs, separats els pacients en 5 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algorisme de regressió que millors resultats ha retornat a sigut el *K-Neighbors* amb un nombre d'estimadors escollit pel *Grid Search* de 3 veïns. El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosten molt a 0. L'error quadràtic mig és molt alt (taula 32).

9 clusters, 5 classes, AHI	K Neighbors
Model Regressió:	{'n_neighbors': 3, 'weights': 'distance'}
Valor r quadràtic:	-0.001
Error quadràtic mig:	758.163

Taula 32. Taula de resultats

Per un model creat determinant 7 grups diferents de roncs, separats els pacients en 4 grups de severitat segons l'índex d'apnea-hipopnea (IAH), l'algoritme de regressió que millors resultats ha retornat a sigut el *AdaBoost* amb un nombre d'estimadors escollit pel *Grid Search* de 5. El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosten molt a 0. L'error quadràtic mig no és molt alt (taula 33).

7 clusters, 4 classes, AHI	AdaBoost
Model Regressió:	{'n_estimators': 5}
Valor r quadràtic:	-0.005
Error quadràtic mig:	878.999

Taula 33. Taula de resultats

Per un model creat determinant 9 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex d'apnea-hipopnea (IH), l'algoritme de regressió que millors resultats ha retornat a sigut el *AdaBoost* amb un nombre d'estimadors escollit pel *Grid Search* de 10. El valor r quadràtic es negatiu pel comportament no lineal de la regressió i s'acosten molt a 0. L'error quadràtic mig és dels més petits (taula 34).

9 clusters, 3 classes, HI	AdaBoost
Model Regressió:	{'n_estimators': 10}
Valor r quadràtic:	-0.003
Error quadràtic mig:	203.375

Taula 34. Taula de resultats

Per un model creat determinant 9 grups diferents de roncs, separats els pacients en 3 grups de severitat segons l'índex d'apnea-hipopnea (IA), l'algoritme de regressió que millors resultats ha

retornat a sigut el *Random Forest*. El valor r quadràtic s'acosta molt a 0. L'error quadràtic mig és dels més petits (taula 35).

9 clusters, 3 classes, AI	Random Forest
Valor r quadràtic:	0.004
Error quadràtic mig:	203.608

Taula 35. Taula de resultats

Iteració 3:

2 classes amb frontera al 50%:

Com observem a la taula 36, per a 2 classes separades per un llindar al 50% l'algoritme que treballa millor és el SVM on l'eina GridSearch ha seleccionat un model de classificació amb els criteris "criterion" i gini" amb una exactitud del 72%.

Algorisme:	Decission Tree
Model Classificació:	{'criterion': 'gini'}
Acc:	0.72
Matriu Confusió:	[7 0] [5 6]

Taula 36. Taula de resultats

Per la regressió, tot i no ser bons resultats com s'observa ala taula 37, el model que millor s'ha ajustat és l'*Adaboost* amb un número d'estimadors de 25.

Algorisme:	AdaBoost
Model Regressió:	{'n_estimators': 25}
Valor r quadràtic:	0.206

Error quadràtic mig:	1435.287
-----------------------------	----------

Taula 37. Taula de resultats

2 classes amb frontera al 33%:

Ajustant la separació de classes en una frontera a 33% els dos algoritmes que treballen millor són el SVM i el *Random Forest* com s'observa a la taula 38. El SVM els detecta a tots com a sever i el *Random Forest* dos dels severs els detecta com a sans.

Algorisme:	SVM	Random Forest
Model Classificació:	{'C': 0.001, 'gamma': 0.1, 'kernel': 'rbf'}	
Acc:	0.67	0.67
Matriu Confusió:	[0 6] [0 12]	[2 4] [2 10]

Taula 38. Taula de resultats

Per la regressió, tot i no ser bons resultats com s'observa a la taula 39, el model que millor s'ha ajustat és l'*Adaboost* amb un número d'estimadors de 25.

Algorisme:	AdaBoost
Model Regressió:	{'n_estimators': 10}
Valor r quadràtic:	-0.046
Error quadràtic mig:	1890.447

Taula 39. Taula de resultats

2 classes amb frontera al 66%:

Ajustant la separació de classes en una frontera a 66% els dos algoritmes que treballen millor són el SVM com s'observa a la taula 40. Amb una alta exactitud però no treballa bé ja que els errors els detecta molts pacients com a falsos positius.

Algorisme:	SVM
Model Classificació:	{'C': 0.001, 'degree': 2, 'kernel': 'poly'}
Acc:	0.78
Matriu Confusió:	[12 0] [4 2]

Taula 40. Taula de resultats

Per la regressió, tot i no ser bons resultats com s'observa a la taula 41, el model que millor s'ha ajustat és el *Decision Tree*, amb un criteri escollit "mse".

Algorisme:	Decision Tree
Model Regressió:	{'criterion': 'mse'}
Valor r quadràtic:	-0.026
Error quadràtic mig:	1854.627

Taula 41. Taula de resultats

3 classes amb frontera al 33-66%:

De la iteració amb dos fronteres al 33 i 66% creant tres grups de severitat, cap algorisme ha donat resultats considerables a part del *Random Forest*, tal com s'observa a la taula 42.

Algorisme:	SVM (66%)
Acc:	0.5
Matriu Confusió:	[3 2 1] [2 3 1] [2 1 3]

Taula 42. Taula de resultats

Per la regressió, tot i no ser bons resultats com s'observa a la taula 43, el model que millor s'ha ajustat és el *AdaBoost*, amb un nombre d'estimadors escollit de 5.

Algorisme:	AdaBoost
Model Regressió:	{'n_estimators': 5}
Valor r quadràtic:	-0.059
Error quadràtic mig:	1914.700

Taula 43. Taula de resultats

Iteració 4:

Per al nou ajustament del líndars de les fronteres de classes de la taula 12, els resultats per la part de classificació tot i no ser destacables comparats amb els obtinguts fins al moment, són els que s'observa a la taula 44, el model que millor s'ha ajustat és el *Decision Tree*, tot i no ser gaire destacable.

8 clusters, 3 classes, AHI	Decision Tree
Model Classificació:	{'criterion': 'gini'}
Acc:	0.44
Matriu Confusió:	[5 1 2] [0 3 2] [2 3 0]

Taula 44. Taula de resultats

Els resultats per la part de regressió tot i no ser destacables comparats amb els obtinguts fins al moment, són els que s'observa a la taula 45, el model que millor s'ha ajustat és el SVM, tot i no ser gaire destacable.

8 clusters, 3 classes, AHI	SVM
Model Regressió:	{'C': 0.01, 'kernel': 'linear'}
Valor r quadràtic:	-0.094
Error quadràtic mig:	671.954

Taula 45. Taula de resultats

Iteració 5:

En aquesta última iteració on ajusto el llindar de severitat a 40 apneas-hipopnees per hora, com s'observa a la taula 13, els resultats per la classificació que millor s'ajusten són el *Decission Tree* i el *AdaBoost* que aconseguix prou bona exactitud, com s'observa a la taula 46.

8 cluters, 2 classes, AHI	Decission Tree	AdaBoost
Model Classificació:	{'criterion': 'entropy'}	{'n_estimators': 5}
Acc:	0.78	0.73
Matriu Confusió:	[[7 2] [2 7]]	[8 1] [4 5]

Taula 46. Taula de resultats

Per la regressió, com s'observa a la taula 47, els millors resultats els ha aconseguit l'algorisme SVM i el *Decission Tree*, amb resultats no gaire significatius comparats amb els obtinguts fins ara.

8clusters, 3classes, AHI	SVM	Decission Tree
Model Regressió:	{'C': 0.01, 'kernel': 'linear'}	{'criterion': 'mse'}
Valor r quadràtic:	-0.095	-1.029
Error quadràtic mig:	671.954	1245.216

Taula 47. Taula de resultats

7. Conclusions

Segons els objectius establerts a l'inici del projecte se poden treure conclusions sobre dos punts.

- Avaluar la possibilitat de crear un marc d'estudi a partir de la gran quantitat i diversificació de les mostres
- Avaluar els resultats aconseguits per la implementació d'aquest disseny

En primer lloc, respecte el primer punt que fa referència a la implementació de l'estudi 1 i 2, es pot concloure que s'ha assolit l'objectiu de crear un disseny capaç d'identificar la presència del SAOS, tot i la dificultat inicial del format de les dades. La unió de dos tipus de dades per a cada un dels pacients, les primeres corresponents als roncs en un format d'una gran quantitat de valors corresponents a diferents característiques d'aquestes mostres i la classe de severitat de cada pacient, amb la implementació i ús de les eines descrites en aquesta memòria, s'ha pogut dissenyar un model viable a utilitzar.

Respecte el segon punt, de l'estudi 3 es conclou que els resultats obtinguts no es poden considerar significatius. No es possible considerar cap combinació de les iteracions com a completament satisfactòria, per la pobresa en els resultats.

Aquests resultats de baixa qualitat es poden relacionar amb dos conceptes globalment coneguts, la "maledicció de la dimensionalitat", i el teorema de "*Not free lunch theorem*" o per la falta de característiques o la baixa qualitat d'elles.

- *Not free lunch theorem* [36] → Aquest teorema afirma que no hi ha cap model concret que funcioni millor per a tots els problemes. Les hipòtesis d'un gran model per a un problema no poden contenir-se per un altre problema, per tant, és comú en l'aprenentatge automàtic provar diversos models i trobar un que funcioni millor per a cada problema concret.
- Maledicció de la dimensionalitat → En els estudis d'aprenentatge automàtic amb percentatges de precisió elevats es tenen una quantitat de mostres a estudiar immensament gran comparat amb la cinquantena de pacients de què disposa aquest estudi. Amb aquest aspecte es pot relacionar la pobresa en la qualitat dels resultats.

Per molt i treballar en molts algorismes diferents i per a diferent índex de classificació i característiques diferents, els resultats del darrer estudi (Estudi 3), no s'ha pogut arribar a determinar una combinació

adequada per identificar la severitat del SAOS amb una precisió elevada, ja sigui per la falta de mostres degut a la dificultat d'adquirir aquests registres en pacients, o per la mala representació de les poques dades de les que es disposa.

7.1. Perspectives futures

En quant a accions futures en relació a millorar els resultats d'aquest estudi es considera convenient treballar sobre els següents punts:

- Destinar recursos a ampliar el nombre de subjectes d'estudi. Si es pogués optar a realitzar l'estudi amb molts més pacients, incrementant considerablement el número de mostres, els resultats serien més acurats ja que estaríem combatent el problema de la maledicció de la dimensionalitat.
- Augmentar el nombre d'algorismes de classificació, provant-ne d'altres existents.
- Millorar la representació de les característiques de les dades de partida, per buscar una exactitud més bona en els resultats.

8. Anàlisi de l'impacte ambiental

Un anàlisi de l'impacte ambiental permet avaluar l'efecte d'un projecte en el medi ambient, és a dir, per predir i valorar les conseqüències que comporten.

A la taula 48 es mostren els impactes especulats de l'aplicació d'aquest estudi, que té com a objectiu principal millorar la detecció dels SAOS:

	Professional	Pacient	Ciència
IMPACTE 1	Reducció del temps de dedicació	Reducció del nombre de proves sotmeses	Reducció en l'ús de recursos
IMPACTE 2	Eficàcia en la detecció del diagnòstic	Reducció del temps d'espera	Precisió en el diagnòstic

Taula 48. Impactes en l'anàlisi ambiental

Aquests dos impactes descrits estan relacionats entre ells.

Quan es parla d'una reducció en l'ús de recursos, es refereix en millorar l'actual procediment d'eines i algoritmes informàtics de la PSG actual, amb la intenció de reduir el nombre de proves, el cost d'elles i l'ús d'especialistes que també significa un cert cost, i l'impacta d'aquest recursos en el medi.

Així doncs, aquesta seria una possible explicació de l'impacte ambiental que produeix aquest estudi.

9. Pressupost i/o Anàlisi Econòmica

L'anàlisi econòmic pretén definir el cost de realització del projecte. Això inclou el temps empleat per part de l'estudiant i els tutors corresponents, les despeses de desplaçament, el material necessari per dur a terme la part pràctica, els recursos humans i d'altres recursos.

Per el càlcul del pressupost en relació a les hores dedicades de l'estudiant, s'ha considerat una mitjana de 3 hores i mitja diàries treballades durant 7 mesos, que equival a 735 hores en total. En la direcció del Treball Fi de Grau, s'ha tingut en compte les hores de dedicació dels dos tutors.

El pressupost s'ha realitzat tenint en compte els costos d'enginyeria, els costos informàtics i altres costos:

9.1. Costos d'enginyeria

El cost d'enginyeria representa les hores invertides en la realització d'aquest projecte per part de l'estudiant i dels col·laboradors, en aquest cas el tutor, Jordi Solà Soler i el co-tutor, Gerard Escudero Bakx.

MOTIU	Hores invertides (h)	Cost (€/h)	Total (€)
Honoraris Estudiant	735	10	7.350
Honoraris Tutor	50	50	2.500
Honoraris Co-Tutor	50	50	2.500
TOTAL			12.350

Taula 49. Costos d'enginyeria del projecte

9.2. Costos informàtics

En aquests costos s'hi inclouen tots els programes, softwares i processador de text utilitzats i a més a més, l'accés a internet.

MOTIU	Quantitat	Cost	Total (€)
Microsoft Office	7 mesos	6 €/mes	42
Python	7 mesos	Gratuït	-
Waikato Environment for Knowledge Analysis (WEKA) Version 3.8.1	7 mesos	Gratuït	-
Red (WIFI)	7 mesos	20 €/mes	140
TOTAL			182

Taula 50. Costos informàtics del projecte

9.3. Altres costos

Finalment, hi ha altres costos relacionats amb el material utilitzat en la preparació de l'entrega en mà del projecte com n'és la impressió, enquadernació, CD-ROM i la caixa de projectes. També es tenen en compte els desplaçaments requerits durant el projecte.

MOTIU	Detalls	Cost (€)	Total (€)
Material	Impressió	20	20
	Enquadernació	2,5	2,5
	CD-ROM i caixa de projectes	3	3
Desplaçaments	Número de trajectes: 10	2 €/trajecte	20
TOTAL			45,5

Taula 51. Altres costos del projecte

9.4. Pressupost final

Tipus de cost	Total (€)
Costos d'enginyeria	12.350
Costos informàtics	182
Altres costos	45,5
TOTAL	12.577,5

Taula 52. Pressupost final del projecte

Així doncs, s'obté finalment d'aquest projecte de recerca un cost orientatiu de 12.577,5€.

10. Bibliografia

- [1] P. Lloberes, J. Durán-Cantolla, M. Á. Martínez-García, J. M. Marín, A. Ferrer, J. Corral, J. F. Masa, O. Parra, M. L. Alonso-Álvarez, and J. Terán-Santos, "Diagnóstico y tratamiento del síndrome de apneas-hipopneas del sueño," *Arch. Bronconeumol.*, vol. 47, no. 3, pp. 143–156, Mar. 2011.
- [2] R. Henry Olivi, "Apnea del sueño: cuadro clínico y estudio diagnóstico," *Rev. Médica Clínica Las Condes*, vol. 24, no. 3, pp. 359–373, May 2013.
- [3] S.A.P., E. M. (2015). Estudio concluye que el oxígeno podría ayudar a combatir el cáncer | Emol.com. [Online]. Available: <http://www.emol.com/noticias/tecnologia/2015/03/05/706635/estudio-oxigeno-puede-ayudar-a-combatir-el-cancer.html>. [Accessed: 03-Mar-2017].
- [4] Dra. Lucila Andrade A., "Relacion entre problemas habituales del sueño con deficit atencional y trastornos conductuales en niños". (2013). *Revista Médica Clínica Las Condes*, 24(3), 494–500.
- [5] Lirón. "Tratamiento roncopatía y la apnea del sueño. Retrieved" [Online]. Available: <http://liron.es/cansancio-lo-largo-del-dia/ronquido-descriptivo/>. [Accessed:3-Mar-2017]
- [6] Ronquido y apnea del sueño | Clínica Beltrán. [Online]. Available: <http://www.clinicaotorrino.es/informacion-pacientes-patologias/patologia-general-orl/ronquido-apnea-del-sueno/>. [Accessed: 03-Mar-2017].
- [7] r. Alfredo Santamaría C, Dr. David Astudillo O, "Vía aérea superior, ronquido e implicancias clínicas". (2013). *Revista Médica Clínica Las Condes*, 24(3), 350–357.
- [8] D. Ferre, "Síndrome de apnea del sueño," 2016. [Online]. Available: http://doctorferre.com/index.php?option=com_content&view=article&id=63&Itemid=80. [Accessed: 10-Mar-2017].
- [9] Peker Y., Hedner J., Norum J., Kraiczi H., Carlson J. "Increased incidence of cardiovascular disease in middle-aged men with obstructive sleep apnea: a 7-year follow-up", *Am J Respir Crit Care Med*, 166 (2002), pp. 159-165
- [10] Gottlieb D.J., Yenokyan G., Newman A.B., O'Connor G.T., Punjabi N.M., Quan S.F., et al. "Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure: the sleep heart health study "Circulation, 122 (2010), pp. 352-360
- [11] Gozal D., Kheirandish-Gozal L., Bhattacharjee R., Spruyt K. "Neurocognitive and endothelial dysfunction in children with obstructive sleep apnea", *Pediatrics*, 126 (2010), pp. 1161-1167
- [12] Canessa N., Castronovo V., Cappa S.F., Aloia M.S., Marelli S., Falini A., et al. "Obstructive Sleep Apnea: Brain Structural Changes and Neurocognitive Function before and after Treatment", *Am J*

- Respir Crit Care Med, 183 (2011), pp. 1419-1426
- [13] Levy P., Bonsignore M.R., Eckel J. Sleep, "Sleep-disordered breathing and metabolic consequences", Eur Respir J, 34 (2009), pp. 243-260
- [14] Friedman, M. "Apnea del sueño y roncopatía tratamiento médico y quirúrgico", Elsevier (2009).
- [15] Duran J., Esnaola S., Rubio R., Iztueta A. "Obstructive sleep apnea-hypopnea and related clinical features in a population-based sample of subjects aged 30 to 70 yr", Am J Respir Crit Care Med, 163 (2001), pp. 685-689
- [16] Young T., Peppard P.E., Gottlieb D.J. "Epidemiology of obstructive sleep apnea: a population health perspective", Am J Respir Crit Care Med, 165 (2002), pp. 1217-1239
- [17] R. Henry Olivi, "Apnea del sueño: cuadro clínico y estudio diagnóstico," Rev. Médica Clínica Las Condes, vol. 24, no. 3, pp. 359-373, May 2013.
- [18] Guilleminault C., Pertinen M. (Eds.), Obstructive sleep apnea syndrome: clinical research and treatment, Raven Press, New York, NY (1990), pp. 25-36
- [19] Robert Bourke, Vicki Anderson, Joel S.C. Yang, et cols. Cognitive and academic functions are impaired in children with all severities of sleepdisordered breathing. Sleep Medicine 12 (2011) 489-496.
- [20] Chervin R.D. Sleepiness, fatigue, tiredness, and lack of energy in obstructive sleep apnea Chest, 118 (2000), pp. 372-379
- [21] Kingshott R.N., Sime P.J., Engleman H.M., Douglas N.J. "Self assessment of daytime sleepiness: patient versus partner", Thorax, 50 (1995), pp. 994
- [22] Dra. Ligia Berdeja Murillo, Dr. Jorge Jorquera, "Medicina Interna" 2017 [Online]. Available: publicacionesmedicina.uc.cl/TemasMedicinaInterna/apnea.html. [Accessed: 19-Apr-2017].
- [23] Ricardo Moya, "K-means en Python y Scikit-learn," 2016. [Online]. Available: <https://jarroba.com/k-means-python-scikit-learn-ejemplos>. [Accessed: 20-apr-2017].
- [24] Truck driver discusses CPAP problems | Sleep apnea. [Online]. Available: fleetowner.com/driver-management-resource-center/how-fix-common-cpap-problems. [Accessed: 3-Mar-2017].
- [25] Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine Learning, Neural and Statistical Classification. Ellis Horwood.
- [26] Fiz J. A., Jané R., Solà-Soler J., Abad J., García M. Á., Morera J., (2010). Continuous analysis and monitoring of snores and their relationship to the apnea-hypopnea index. The Laryngoscope, 120(4), 854-862.
- [27] K. Gibert, L. J. Oliva, I. Pinyol, and M. Sánchez-Marrè, "El bagging en casos no supervisats. Implementació a GESCONDA per algorismes de clustering," 2006.

- [28] Introduction to Support Vector Machines. [Online]. Available: http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html. [Accessed: 20-Apr-2017].
- [29] Machine Learning and Data Mining: 13 Nearest Neighbor. [Online]. Available: <https://www.slideshare.net/pierluca.lanzi/machine-learning-and-data-mining-13-nearest-neighbor-and-bayesian-classifiers>. [Accessed: 20-Apr-2017].
- [30] Aprendizaje Inductivo: Árboles de Decisión. [Online], Available: <http://www.cs.us.es/~fsancho/?e=104>. [Accessed: 20-Apr-2017]
- [31] Random Forest Template for TIBCO. [Online]. Available: <https://community.tibco.com/wiki/random-forest-template-tibco-spotfirer-wiki-page>. [Accessed: 20-Apr-2017]
- [32] Model selection GridSearch. Scikit-learn. (n.d.). [Online]. Available: http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. [Accessed: 19-Apr-2017]
- [33] Waikato Environment for Knowledge Analysis [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>. [Accessed: 8-Apr-2017]
- [34] Python Software Foundation [Online] 1991. Available: <https://www.python.org/> [Accessed: 8-Apr-2017]
- [35] Muraja-Murro, A. (n.d.). The severity of individual obstruction events is related to increased mortality rate in severe obstructive sleep apnea. *J Sleep Res.*, (2013), pp 663-669
- [36] Wolpert, D. H., & Macready, W. G. (1997). No Free Lunch Theorems for Optimization. *IEEE TRANSACTIONS ON EVOLUTIONARY COMPUTATION*. [Online]. Available: <https://ti.arc.nasa.gov/m/profile/dhw/papers/78.pdf>. [Accessed: 13-Maig-2017]

