

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ

SOLUNUM FONKSİYONU HASTALIKLARININ
VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
DEĞERLENDİRİLMESİ

Yüksek Lisans Tezi

YILMAZ GÜLER

İSTANBUL, 2014

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ YÜKSEK LİSANS PROGRAMI

SOLUNUM FONKSİYONU HASTALIKLARININ
VERİ MADENCİLİĞİ YÖNTEMLERİ İLE
DEĞERLENDİRİLMESİ

Yüksek Lisans Tezi

YILMAZ GÜLER

Tez Danışmanı: PROF. DR. ADEM KARAHOCA

İSTANBUL, 2014

T.C.
BAHÇEŞEHİR ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİ TEKNOLOJİLERİ
YÜKSEK LİSANS PROGRAM ADI

Tezin Adı: Solunum Fonksiyonu Hastalıklarının Veri Madenciliği Yöntemleri İle Değerlendirilmesi
Öğrencinin Adı Soyadı: Yılmaz GÜLER
Tez Savunma Tarihi: 25 / 08 / 2014

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğu Fen Bilimleri Enstitüsü tarafından onaylanmıştır.

Doç. Dr. F. Tunç BOZBURA
Enstitü Müdürü
İmza

Bu tezin Yüksek Lisans tezi olarak gerekli şartları yerine getirmiş olduğunu onaylarım.

Prof. Dr. Adem KARAHOCA
Program Koordinatörü
İmza

Bu Tez tarafımızca okunmuş, nitelik ve içerik açısından bir Yüksek Lisans tezi olarak yeterli görülmüş ve kabul edilmiştir.

Jüri Üyeleri

İmzalar

Tez Danışmanı
Prof.Dr. Adem KARAHOCA

Üye
Yrd.Doç.Dr. Dilek KARAHOCA

Üye
Yrd.Doç.Dr.Yalçın ÇEKİÇ

ÖN SÖZ

Bu çalışmam süresince her türlü yardımlarını ve desteğini sağlayan, tezin içeriğini oluşturmamda yardımcı olan, bilgi ve tecrübesiyle çalışmama ışık tutan tez danışmanım Sayın Prof. Dr. Adem KARAHOCA'ya,

Tez araştırmamda, solunum fonksiyonu testi yaptıran hastalarla ilgili verilere ulaşmamda destek sağlayan Özel Gaziosmanpaşa Hastanesi ve Özel Gaziosmanpaşa Hastanesi Başhekimi Uz. Dr. Sedat AZAK'a

Tezimin hazırlanması sırasında maddi, manevi hiçbir desteği esirgemeyen aileme ve manevi desteğini esirgemeyen Harun ÖNEL'e teşekkürü bir borç bilir, sonsuz sevgilerimi sunarım.

İstanbul 2014

Yılmaz GÜLER

ÖZET

SOLUNUM FONKSİYONU HASTALIKLARININ VERİ MADENCİLİĞİ YÖNTEMLERİ İLE DEĞERLENDİRİLMESİ

Yılmaz GÜLER

Bilgi Teknolojileri

Tez Danışmanı : Prof. Dr. Adem KARAHOCA

Ağustos 2014, 46 Sayfa

Gelişen dünyada insan nüfusunun artmasıyla, birçok hastalık türü de ortaya çıkmaktadır. Bu hastalıklardan biri de solunum fonksiyonları problemidir. Solunum fonksiyonu problemleri, geçmişten günümüze kadar gelen ve gelecekte de artarak devam edecek hastalıklardan biri olarak görülmektedir.

Solunum fonksiyonları, şikayetlerinden dolayı hastanelere başvuruda bulunan ve solunum fonksiyonu testi yaptıran hasta sayısı her geçen gün artmaktadır. Bu artan hasta verilerinin depolanması ve yönetilmesi zorlaşmaktadır.

Günümüzdeki teknolojik gelişmeler, birçok alanda yapılan araştırmaları kolaylaştırmakta ve bilimsel çalışmalara da ışık tutmaktadır. Teknolojik gelişmeler, veri madenciliği alanında da büyük verilerin yönetilmesi, saklı olan verilerin ortaya çıkarılması ve analizlerinin yapılmasında kolaylıklar sağlamaktadır.

Bu araştırma 2012-2014 yılları arasında Özel Gaziosmanpaşa Hastanesi Göğüs Hastalıkları Anabilim Dalı Solunum Fonksiyon Laboratuvarında test yaptıran 400 hasta üzerinden yürütülmüştür. Solunum fonksiyonu test sonuçları teşhis ve tedavisi üzerinde durulmuş olup yaş, cinsiyet, boy, kilo, alerjik durumu, aile geçmişi, mesleği, sigara kullanımı, FVC ve FEV1 veri değerleriyle veri madenciliği yöntemleri uygulanmıştır.

Anahtar Kelimeler : Veri Madenciliği, Solunum Fonksiyonları Testi, Solunum Fonksiyonu Problemleri, WEKA

ABSTRACT

THE EVALUATION OF RESPIRATORY SYSTEM DISEASES BY DATA MINING METHODS

Yılmaz GÜLER

Information Technologies

Thesis Supervisor: Ph.D., Prof. Adem KARAHOCA

Ağustos 2014, 46 Pages

Many disease types have begun to appear due to rapid increases in human population. Respiratory dysfunction is an example of one of these diseases. Respiratory dysfunction is a disease which has existed in the past and will continue to be seen in the future.

However, the number of patients presenting symptoms and being tested is continually increasing. Additionally, storage and management of data relating to this disease is getting more complicated and difficult.

Some technological improvements make research in many fields easier and shed light on scientific studies. Other technological developments facilitate the management of large amounts of data that can expose hidden information, better assisting data mining.

This study was conducted on 400 patients who were tested at Gaziosmanpaşa Hospital, Department of Chest Diseases Laboratory of Pulmonary Functions. Pulmonary function test results, diagnosis and treatment were emphasized and age, gender, height, weight, allergy status, family history, occupation, smoking, FVC and FEV1 data values and data mining methods were applied.

Keywords : Data Mining, Respiratory Function Test, Respiratory Function Problems, WEKA

İÇİNDEKİLER

TABLolar	vi
ŞEKİLER	vii
KISALTMALAR	viii
1. GİRİŞ	1
1.1 ARAŞTIRMANIN AMACI	2
1.2 ARAŞTIRMANIN ÖNEMİ.....	2
1.3 VARSAYIMLAR.....	3
1.4 ARAŞTIRMANIN SINIRLILIKLARI.....	3
2. LİTERATÜR TARAMASI	4
2.1 VERİ MADENCİLİĞİ.....	4
2.2 WEKA.....	7
2.3 SOLUNUM FONKSYON TESTLERİ.....	8
2.4 GENEL KAVRAMLAR.....	9
2.4.1 Spirometreler	9
2.4.2 Havayolu Fonksiyonlarını Gösteren Testler.....	10
2.4.2.1 Basit spirometrik testler.....	11
2.4.2.2 Zorlu vital kapasite	11
2.4.2.3 Maksimum solunum kapasitesi (MVV)	11
2.4.2.4 Maksimum inspiratuar ve ekspiratuar basınçlar (MIP, MEP).....	11
2.4.2.5 Havayolu direnci (Raw)	12
2.4.3 Zorlu Vital Kapasite Manevrası.....	12
2.4.3.1 Zorlu vital kapasite (FVC).....	12
2.4.3.2 Birinci saniye zorlu ekspirasyon volümü (FEV1)	12
2.4.3.3 FEV1/FVC (Tiffeneau oranı).....	12
2.4.3.4 Tepe akım hızı (PEF)	13
2.4.4 Sigara Kullanımının SFT Etkisi.....	13
2.4.5 Allerjik Ve Allerjik Olmayan Hastaların SFT Etkisi	15
2.4.6 Meslek Hastalıklarının Tanısında SFT Etkisi.....	16
2.4.6.1 Spirometrik Testi	19
2.4.6.2 Reversibilite Testi.....	19

2.4.6.3 Bronş Provokasyon Testi.....	19
2.4.6.4 Difüzyon Testi.....	19
2.4.6.5 Akciğer Volümleri	19
3. VERİ VE YÖNTEM	21
3.1 ARAŞTIRMANIN MODELİ	21
3.2 VERİ TOPLAMA VE ANALİZ.....	21
3.2.1 Örneklem.....	21
3.2.2 Verilerin Analizi	21
4. BULGULAR	22
4.1. SINIFLANDIRMA YÖNTEMLERİ.....	22
4.1.1 BayesNet Modeli.....	22
4.1.2 NaiveBayes Modeli	23
4.1.3 Logistic Modeli.....	24
4.1.4 Multiplayer Perceptron Applicaton Modeli.....	25
4.1.5 Jrip Modelleme	25
4.1.6 Part Modeli.....	26
4.1.7 Oner Modeli.....	27
4.1.8 ZeroR Modeli	28
4.1.9 RBFNetwork Modeli	29
4.1.10 J48 Modeli.....	30
4.1.11 Sınıflandırma Modelinde Eğitim Veri Sonuçları	31
4.1.12 Sınıflandırma Metotlarının Özeti	33
4.2.KÜMELEME YÖNTEMLERİ.....	33
4.3 BİRLİKTELİK YÖNTEMLERİ.....	34
5. TARTIŞMA	36
6. SONUÇ.....	38
KAYNAKÇA	40

TABLULAR

Tablo 2.1 : Obstrüktif bozukluğun şiddeti	13
Tablo 2.2: Sigara içenler ve içmeyenlerde SFT (%) ortalama değerleri.....	14
Tablo 2.3: Sigara içenler ve içmeyenlerde STF (%) ortalama değerleri.....	15
Tablo 2.4 : Ülkemizde yapılan mesleki astım arařtırmaları.....	17
Tablo 2.5 : Grupların Demografik Özellikleri	18
Tablo 2.6 : Grupların çalıştıkları yerdeki toz ve silika konsantrasyonları	19
Tablo 2.7 : Grupların solunum yakınmaları	19
Tablo 4.1 : Sınıflandırma Yöntemlerinin Karşılaştırma Sonuçları	32
Tablo 4.2 : Sınıflandırma Doğruluk Sonuçları.....	32

ŞEKİLLER

Şekil 2.1 : Bilgi keşfi sürecinde veri madenciliği	5
Şekil 2.2 : WEKA Hiyerarşik Yapısı	8
Şekil 2.3 : Sigara içenler ve içmeyenlerde STF (%) ortalama değerleri	15
Şekil 2.4 : Sigara içenler ve içmeyenlerde STF (%) ortalama değerleri	15
Şekil 4.1 : BayesNet İstatistiksel Değerleri.....	22
Şekil 4.2 : Bayes Networks Doğruluk Değerleri.....	23
Şekil 4.3 : NaiveBayes İstatistiksel Değerleri.....	23
Şekil 4.4 : NaiveBayes Doğruluk Değerleri.....	24
Şekil 4.5 : Logistic İstatistiksel Değerleri	24
Şekil 4.6 : Multiplayer Perceptron İstatistiksel Değerleri	25
Şekil 4.7 : Multipler Perceptron Doğruluk Değerleri.....	26
Şekil 4.8 : Jrip İstatistiksel Değerleri	26
Şekil 4.9 : Jrip Doğruluk Değerleri	27
Şekil 4.10 : Part İstatistiksel Değerleri.....	27
Şekil 4.11 : Part Doğruluk Değerleri.....	28
Şekil 4.12 : Oner İstatistiksel Değerleri	28
Şekil 4.13 : Oner Doğruluk Değerleri	29
Şekil 4.14 : ZeroR İstatistiksel Değerleri	29
Şekil 4.15 : ZeroR Doğruluk Değerleri.....	30
Şekil 4.16 : RBF Network İstatistiksel Değerleri.....	30
Şekil 4.17 : RBF Network Doğruluk Değerleri.....	31
Şekil 4.18 : J48 İstatistiksel Değerleri	31
Şekil 4.19 : J48 Doğruluk Değerleri	32
Şekil 4.20 : Simple K Mean Değerleri	34
Şekil 4.21 : Apriori Değerleri.....	35

KISALTMALAR

AD	:	Alerjik Durum
AG	:	Aile Geçmişi
BY	:	Boy
CCI	:	Correctly Classified Instances
CNT	:	Cinsiyet
ERV	:	Ekspiratuar Rezerv Volüm
FEV1	:	Birinci Saniye Zorlu Ekspirasyon Volümü
FVC	:	Zorlu Vital Kapasite
IC	:	Inspirasyon Capacity
ICI	:	Incorrectly Classified Instances
IRV	:	Inspiratuar Rezerv Volüm
KL	:	Kilo
KS	:	Kappa Statistic
MAE	:	Mean Absolute Error
MEP	:	Maksimum Ekspiratuar Pressure
MH	:	Mesleki Hastalık
MIP	:	Maksimum Inspiratuar Pressure
MVV	:	Maximum Voluntary Ventilation
PEF	:	Peak Expiratory Flow
RAE	:	Relative Absolute Error
RAW	:	Havayolu Direnci
RMSE	:	Root Mean Squared Error
RRSE	:	Root Relative Squared Error
SD	:	Sapma Değeri
SFT	:	Solunum Fonksiyonları Testi
SK	:	Sigara Kullanımı
TNI	:	Total Number of Instances
YR	:	Yorum
YS	:	Yaş
VC	:	Vital Capacity

1. GİRİŞ

Bu günlerde, bilim dünyası büyük bir gelişme gösterirken; aynı paralellikte teknoloji de hızlı bir şekilde gelişmektedir. Teknoloji, insanların vaz geçilmezleri arasında yer alırken, teknolojinin kullanımı insanların hayatlarının bir parçası olmaktadır. Gelişen teknoloji insanların hayatlarını kolaylaştırmaya başlamıştır. Teknolojik gelişmeler, bilimsel araştırmalarda da hızlı sonuçlar alınmasında ve doğru değerlere ulaşılmasında da kolaylık sağlamaktadır. Yirminci yüzyıl başlarından itibaren bilim ve teknoloji ayrılmaz bir bütün olarak günümüzde de devam etmektedir. Çünkü 1950'li yıllardan itibaren kullanılan elektronik bilgisayarlar sayesinde, bilim ve teknoloji arasında ki ilişki bağı kurulmaya başlamıştır. Her geçen gün teknik özellik ve donanımsal olarak gelişen bilgisayarlar önceden yapılması zor olan araştırmalara ışık tutmasında yardımcı olup, var olan araştırmalarda da hızlı ve daha doğru sonuçlara ulaşılmasını sağlamaktadır. Ayrıca gizli kalmış bilgilerin ortaya çıkarılmasında gelişen teknoloji önemli rol almaktadır. Bilimsel gelişmelerin alt yapısı Rönesans ve Reform hareketleriyle birlikte oluşmuştur. Rönesans ve Reform hareketleriyle gelişmeye başlayan bilimi engellemeye çalışan tüm olumsuzluklar zaman içinde ortadan kalkarak teknolojinin gelişmesini de sağlamıştır. Gelişen teknoloji, insanların yaşadıkları dünyayı ve bu dünyadaki sorunları keşfetmişlerdir. Bu gelişmeler sayesinde bilimsel gelişmeler teknoloji ile birleştiğinde zamanla hız kazanmıştır (Neill, 1989).

Gelişen dünyada teknoloji, insanların vaz geçilmez bir parçası haline almakla birlikte ve saklanan veri alanları da artmaktadır. Birçok firma ve şirketler verilerini gelişen teknoloji alt yapısına taşımaya başlamaktadırlar. Her geçen gün kurumsal firmaların bilgi depoları artmakta ve daha karmaşık bir hal almaktadır.

Gelişen teknolojiyle tıp dünyası da teknoloji alt yapısına taşınmaktadır. Önceleri hastalara yapılan test sonuçlarının hata oranının daha yüksek olmasına rağmen, günümüzde gelişmiş olan teknoloji alt yapısıyla birlikte test sonuçlarının hata oranlarının sıfıra yakın analiz sonuçları görülmektedir. Her geçen gün hasta verileri artmakta ve verilerin artmasıyla birlikte yönetilemeyen veri bankaları ve karar verme mekanizmalarında sorunlar ortaya çıkmaktadır.

Günümüzde, artan hava kirliliği, sosyal hayatta maruz kalınan sosyal alanlar ve bazı mesleklerde çalışılan iş ortamının etkileri sonucunda insanlarda oluşan solunum sorunları ortaya çıkmaktadır. Dış çevrenin etkisiyle, insanlar da ortaya çıkan solunum fonksiyonları problemleri, solunum fonksiyonları test sayısının yüksek oranda arttırmaktadır. Solunum fonksiyonları testi sonucu farklı hastalıklar (akciğer kanseri, kronik bronşit, amfizem, alerjik durumlar, mesleki hastalıklar, sigara kullanımının etkileri) ve farklı değerler ortaya çıkmasından ve hasta sayılarının artmasından bu verilerin yönetimi ve karar analizi için veri madenciliği uygulamaları üzerinde çalışma yapılması düşünülmüştür.

Son yıllarda tıp alanında solunum fonksiyonu şikayetlerinden dolayı test yaptırılan hasta sayısının artmasından, veri sayısı da hızlı bir şekilde artmaktadır. Büyüyen verilerin depolanması, saklı olan verilerin ortaya çıkarılması ve çözümlenmesi gerekmektedir. Birçok kaynaktan elde edilen veriler içerisinde saklı bulunan bilgiyi bulma işlemine veri madenciliği denilmektedir.

1.1 ARAŞTIRMANIN AMACI

Araştırmanın amacı, solunum fonksiyonu problemi yaşamamış kişilerin ileride ortaya çıkabilecek olan solunum fonksiyonu problem tipini belirlemek ve erken teşhis ile tedavinin sağlanmasıdır. Bu çalışmamız için önceden solunum fonksiyonu problemi yaşamış ve test yaptırmış olan hastaların verilerine ihtiyaç duyulacaktır. Bu verilerde yaş, cinsiyet, aile geçmişi, meslek, yaşadığı çevre ve sigara kullanımı vb. bilgiler dikkate alınacaktır. Farklı olan bu veriler için veri madenciliği kullanılacaktır. Veri madenciliği yöntemleri kullanılarak hastaların solunum fonksiyonu problemlerinin ve risk durumlarının neler olacağı belirlenmeye çalışılacaktır.

1.2 ARAŞTIRMANIN ÖNEMİ

Solunum Fonksiyonu Testi havayolu hastalıklarının değerlendirilmesinde önemli katkı sağlayan objektif bir yöntemdir. Solunum fizyolojisi ve fonksiyonları alanındaki ilk bilinen çalışmalar Antik Roma'nın en önemli hekimlerinden olan Bergama doğumlu Galen'e kadar dayanmaktadır. Galen diyaframın solunum kası olarak rolünü ortaya koymuş, frenik sinir, interkostal ve yardımcı solunum kaslarının fonksiyonlarını göstermiştir. Bir insanın tek bir solukla inhale edebildiği hava miktarı ilk defa Borelli tarafından 1679'da ölçülmüştür. Sonraki çalışmalarda tek solukla alınan hava miktarı oda havasında ortalama 200-300 mL bulunmuş, vücut ısısına göre düzeltme gerektiği 1788'de Goodwyn tarafından ortaya konulmuştur. 1831'de kadınların akciğer volümlerinin erkeklerden daha az olduğunu ve işçilerde toz inhalasyonunun volümleri azalttığı Thackrah tarafından gözlemlenmiştir. Ekspirasyon havası volümünü zamanla ilişkilendiren zorlu vital kapasite ölçümü ilk kez 1919'da Strohl tarafından gerçekleştirilmiştir. 1932'de Jansen, Knipping ve Stromberger dinamik bir parametre olan maksimum solunum kapasitesini (MBC) zorlu vital kapasiteden hesaplamışlardır. 1933'te ise Hermannsen maksimal istemli ventilasyonu (MVV) ölçmüştür. Birinci saniyede atılan zorlu ekspirasyon volümünü (FEV1) solunum fonksiyonları alanına katan bilim adamı ise 1948'de bu ölçümü yapan Fransız farmakolog Robert Tiffeneau'dur (Bartu 2012, ss. 1-5).

Solunum fonksiyonu testleri sonucu anlayamayan ya da geliştirilemeyen problemlerin dijital teknolojiye taşınması ve çözümler bulmak amacıyla bu çalışma Gaziosmanpaşa Özel Hastanesi Göğüs Hastalıkları Anabilim Dalı Solunum Fonksiyon Laboratuvarında test yaptırılan hastaların test sonuçlarına yönelik veriler önem taşımaktadır.

Solunum Fonksiyonları üzerine yapılan veri madenciliđi alıřması ile en fazla hangi tip solunum problemi yařandığını, hangi mesleklerde alıřan kiřilerin bu hastalıđa yakalandığını, hangi yař grubunun solunum problemi riskinin daha fazla olduđunu, evresel etkenlerin ve sigara kullanımının nasıl etkilediđini analiz ederek, solunum fonksiyon hastalarının erken teřhisinde; veri madenciliđi yntemleri uygulanarak hastalıđa yakalanabilecek kiřilere erken teřhis ve tedaviye bařlanması hedeflenmiřtir.

1.3 VARSAYIMLAR

Bu arařtırma ařađdaki varsayımlara gre planlanmıř ve gerekleřtirilmiřtir.

- i. SFT (solunum fonksiyon testi) uygulanan hastaların meslek, alerjik durum ve sigara kullanımı gibi sorulan soruları yanıtlarken dođru cevaplar verdiklerini,
- ii. SFT uygulanan hastaların lmler iin kullanılan cihazlara verdikleri tepkilerle etkileřimlerini nemsedikleri,
- iii. SFT uygulanan hastaların, gđs hastalıkları anabilim dalına gelenlerin gerekten solunum problemi yařayan hastalar oldukları varsayılmıřtır.

1.4 ARAřTIRMANIN SINIRLILIKLARI

Ařađdaki arařtırmamızdaki sınırlılıklarımız belirtilmiřtir. Buna gre arařtırma;

- i. 2012 – 2014 yılları arasında solunum fonksiyonu testi yaptırın 400 hasta ile sınırlıdır.
- ii. Solunum fonksiyonları testinde alt yař grubu 15, st yař grubu 85 olan hastalar ile sınırlıdır.
- iii. Solunum fonksiyonları testinde alt boy ls 140 cm, st boy ls 190 cm ile sınırlıdır.

2. LİTERATÜR TARAMASI

2.1 VERİ MADENCİLİĞİ

Veri madenciliği, büyük miktarlardaki verilerin incelenerek gelecekte olabilecek olayların tahmin edilmesinde kolaylık sağlayacak anlamlı ve yararlı bilgilerin bilgisayar uygulamaları sayesinde analiz edilmesidir. Buna ek olarak veri madenciliği, çok büyük verilerin içinde gizli kalmış bilgilerin ortaya çıkarılmasında ve veriler arasındaki ilişkileri inceleyerek bağlantı kurmaya çalışan veri analizi tekniğidir (Kalikov 2006).

İşletmelerin karar ağaçları sistemlerinde doğru bilgiye ulaşılması ve doğru adımların atılması veri madenciliği teknikleri kullanılarak mümkün olmaktadır. Geçmişte kullanılan karar sistemlerinden farklı bir yapıya sahip olan veri madenciliği farklı özelliklere sahip olan metotlarıyla daha geniş analizler yapmaktadır (İnan 2003).

Hand (1998), “Veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veri tabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi”, Kitler ve Wang (1998), “Veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce potansiyel değişkenden izole edilmesini sağlama yeteneği”, Jacobs (1999), “Veri madenciliğini, ham datanın tek başına sunamadığı bilgiyi çıkaran, veri analizi süreci ”, Doğan ve Türkoğlu (2007) “Veri madenciliği, büyük veri yığınları arasından gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağlantıların, bilgisayar programı kullanarak aranması işidir” olarak tanımlamışlardır (Savaş ve diğ. 2012, ss. 1-23).

Aynı zamanda veri madenciliği belli bir sonuca götüren işlem basamak sürecidir. Büyük veri yığınları içerisinde gizli kalmış verileri çalışmalar sonucunda ortaya çıkarılmasının dışında, bilginin keşfedilmesi sürecinde belirli bir kurullara göre dizilmiş verileri ayırarak filtrelemek ve diğer adımlar için hazırlayan sürecin bir yapısıdır (Savaş ve diğ. 2012, ss. 1-23).

Şekil 2.1 : Bilgi keşfi sürecinde veri madenciliği

Kaynak : http://istu.edu.tr/uploads/yayin/fen21_dosyalar/1-23.pdf
[Eriřim : 29 Nisan 2014]

Veri madencilięi s¼recinde izlenen adımlar genellikle ařaęıdaki řekildedir.

- a. Problemin tanımlanması,
- b. Verilerin hazırlanması,
- c. Modelin kurulması ve deęerlendirilmesi,
- d. Modelin kullanılması,
- e. Modelin izlenmesi.

Problemin tanımlanması: Problemin hangi iřletme alanında yapılacaęını ve çıkan sonuçlar doęrultusundaki analizlerin nasıl deęerlendirileceęini tanımlanmasında, veri madencilięi çalışmalarında başarıya g¼t¼ren en önemli kurallarından biridir (Shearer 2000, ss. 13 - 23).

Verilerin hazırlanması: Modelin kurulması durumunda çıkan problemlerin doęrultusunda verilerin tekrar d¼zenlenmesi durumudur. Ortaya çıkan bu durum verilerin hazırlanması ve modelin kurulması ařamalarında, çalışmayı s¼rd¼ren kiřinin verilerin keřfedilmesi boyunca çalışmasının ve harcanan zamanın yüzde 50 – yüzde 85’ini kaybetmesine sebep olmaktadır (Piramuthu,1998). Verilerin hazırlanması s¼reci, verilerin toplanması, verilere deęer biçme, verilerin temizlenmesi, verilerin birleřtirilmesi, verilerin b¼t¼nden seęimi ve verilerin d¼n¼řt¼r¼lmesi basamaklarından meydana gelmektedir (Shearer 2000, ss. 13 - 23).

Modelin kurulması ve deęerlendirilmesi: Belirlenen sorun ięin uygun modelin seęilebilmesi ve dięer modeller kullanılarak çıkan analizler sonucunda en uygun modelin deęerlendirilmesidir.

Bundan dolayı verinin hazırlanması ve modeli kurma basamaklarında en iyi analiz sonucunun alındığı düşünülen model bulunana kadar bu işlem devam eden süreçtir (Shearer 2000, ss. 13 - 23).

Modelin kullanılması: Kullanılan ve geçerli olduğu düşünülen model bir başka uygulamanın alt yapısı olabileceği gibi bir uygulama olarak da kullanılabilir (Shearer 2000, ss. 13 - 23).

Modelin izlenmesi: Kurulan modellerin ürettikleri verilerde çıkabilecek değişikliklerden dolayı, verilerin düzenlenmesini gerektiren durumlar için kurulan modellerin izlenmesi gerekmektedir (Shearer 2000, ss. 13 - 23).

Veri madenciliği modelleri, verilerin üzerinde yapılan çalışmaya göre üç grupta incelenmektedir.

- a. Sınıflama (Classification) ve Regresyon (Regression),
- b. Kümeleme (Clustering),
- c. Birliktelik Kuralları (Association Rules) dır.

Sınıflama ve iki ya da daha çok değişken arasındaki ilişkiyi ölçmek için kullanılan modeller tahmin edici modeller olup, kümeleme ve birliktelik kuralları da tanımlayıcı modellerdir (Özkes 2003).

Sınıflandırma: Veri madenciliğinde en çok kullanılan alanlardan biri sınıflandırma modelidir. Sınıflandırma kuralları oluşturulurken verilerin bir parçası eğitim olarak kullanılır. Ortaya çıkan yeni durum karşısında ne yapılacağı konusunda karar verilmesi bu kurallar sayesinde olmaktadır (Çoşlu 2013).

Kümeleme: Verilerin kendi içlerindeki ortak benzerliklerine bakılarak oluşturulan gruplandırma çalışmasıdır. Kümeleme modelinde verilerin arasındaki uzaklıkları kullanır. Kümeleme modelinde, hiyerarşik kümelemede en uzak komşu algoritmasıyla, en yakın komşu algoritması durumudur (Çoşlu 2013).

Birliktelik Kuralları: Veri madenciliğinde kullanılan verilerin aralarındaki ilişkileri göz önüne alarak, meydana gelebilecek hangi durumun aynı anda gerçekleşebileceğini gösteren veri madenciliği modeli yöntemidir (Çoşlu 2013).

Veri madenciliği büyük verilerin var olduğu her alanda kullanmak mümkündür. Günümüzde veri madenciliği uygulamaları birçok iş alanında karar verme sürecinde kullanılmaktadır. Veri madenciliğinin kullanılan alanlara bakıldığında pazarlama, biyoloji, bankacılık, sigortacılık, borsa, perakendecilik, telekomünikasyon, genetik, sağlık, bilim ve mühendislik, kriminoloji, sağlık, endüstri, istihbarat vb. birçok dalda başarılı uygulamaları görülmektedir (Akgöbek ve Çakır 2009, Albayrak 2008, İnan 2003). Sağlık alanında veri madenciliğinin en çok kullanıldığı alanlara bakıldığında tıp, biyoloji ve genetik görülmektedir.

Büyük veriler üzerinde yapılan veri madenciliği uygulamaları, veri madenciliği modelleme programları kullanılarak yapılmaktadır. Veri madenciliği alanında kullanılan, SPSS Clementine, Excel, SPSS, SAS, Angoss, KXEN, SQL Server, MATLAB ticari ve RapidMiner(YALE), WEKA, R, C4.5, Orange ve KNIME programları açık kaynak olmak üzere geliştirilen veri madenciliği programdır (Dener ve diğ 2009).

Araştırmamızda veri madenciliği uygulaması için açık kaynak “WEKA” veri madenciliği modelleme programı kullanılmaktadır.

2.2 WEKA

Weka, Yeni Zelanda'daki Waikato Üniversitesi tarafından geliştirilmiş olan açık kaynak kodlu, kullanışlı grafik arabirimi olan ve makine öğrenimi algoritmalarının bir arada barındıran veri madenciliği programdır. (Witten ve diğ. 2011). Weka veri madenciliği programı verilerin ön işlenmesi, sınıflandırma çeşitli veri ön işleme, sınıflandırma, iki ya da daha çok değişken arasındaki ilişkinin ölçmesi, kümeleme, ilişkilendirme kuralları ve görselleştirme modelleme araçları içeren bir programdır. Algoritmaların uygulanması için veri grupları direk olarak veya Java programlama kodlarından çağrılarak yapılabilir (Patterson ve diğ. 2008).

Weka veri madenciliği programı kullanılacak olan verileri dosyadan okuyarak, veriler üzerindeki değişkenlerin sayısal veya nominal değerler olarak kabul eder. Aynı zamanda veritabanı üzerinden de veriler kullanılarak veri madenciliği çalışması yapılabilmektedir. Fakat veritabanı üzerinden alınan verilerin de dosya verisi şeklinde olması gerekmektedir (Şeker 2009).

Weka veri madenciliği programının desteklediği veri dosya şekli arff ve csv dir. Arff veri dosya şekli, Csv veri dosya şekline göre ulaşılabilir dosya uzantısına sahip olup Csv veri dosya şeklinden farklı olarak dosya üzerinde yapılan değişiklikler kaydedebilmektedir (Payalan 2013).

Weka veri madenciliği programı içerisinde, veri işleme modülü, veri sınıflandırma modülü, veri kümeleme modülü ve veri ilişkilendirme modülü özellikleri bulunmaktadır. Bu veri madenciliği veri modülü özellikleri kullanarak çalışma yapılan alanlardan biride tıp alanıdır. Tıp alanında veri madenciliğinin, solunum fonksiyonları testleri üzerindeki bulgularını ve sonuçlarının neler olacağı hakkında veri madenciliği çalışması yapılacaktır.

Şekil 2.2 : WEKA Hiyerarşik Yapısı

Kaynak : http://www.mmo.org.tr/resimler/dosya_ekler/ee2d530a0488f3d_ek.pdf
[Erişim : 29 Nisan 2014]

2.3 SOLUNUM FONKSİYON TESTLERİ

Solunum fonksiyon testleri (SFT) solunum havayolları hastalıklarının klinik değerlendirmelerinde yaygın olarak kullanılan bir laboratuvar yöntemidir. Solunum sisteminin akciğerlere giren ve çıkan hava hacmini, diffüzyonu ve mekanik özellik durumlarını incelenmesinde kullanılan objektif bir yöntemdir (Akkoca 2004).

Solunum fonksiyonu testlerinde kullanılan bu yöntemler,

- a. Havayolu fonksiyonlarını gösteren testler,
- b. Akciğer volümleri ve ventilasyon, Diffüzyon testi,
- c. Kan gazları,
- d. Kardiyopulmoner egzersiz testleri,
- e. Metabolik ölçümlerdir (Akkoca 2004).

Bu testler solunum problemi yaşayan hastanın, nefes darlığı, hırıltı veya uzun süren öksürük şikayetlerinde, astım veya KOAH (Kronik Bronşit, Amfizem) tanısı konulması ve bu hastalıkların birbirinden ayırımında kullanılır. Ayrıca teşhisi konmuş KOAH, astım, kalp yetersizliği ve solunum kaslarını tutan hastalıkların seyrinin ve tedavi durumunun izlenmesinde kullanılır. Solunum fonksiyonları testi solunum problemleri sebebiyle meydana gelen hastalıklı olma durumunun değerlendirmesinde, akciğer hastalığı riski olan hasta ya da hasta grupların taranmasında (Örneğin sigara içenler, maden ocaklarında çalışanlar, çalıştığı ortamda zararlı gaz soluyan işçiler) ve genel anestezi alacak ve ameliyata bağlı olumsuz solunum risklerinin önceden tahmin edilmesi ve riskli hastalarda ameliyat öncesi dönemde gerekli olan önlemleri almak için kullanılır.

Doğum esnasında meydana gelen veya sonraki zamanlarda oluşmuş göğüs şekil bozukluklarının solunuma etkisini araştırmak amacıyla bu testler yapılır (İÜKE).

Solunum fonksiyonu testi yaptracak olan hastanın bazı ölçümleri yapılması gerekmektedir.

- a. Yaş, boy ve kilo ölçümü,
- b. Hastanın kullandığı ilaçların tipi, dozu ve son kullanma saati,
- c. Testten önce 24 saat süreyle sigara içmemesi,
- d. Testten önce 4 saat süreyle alkol almaması,
- e. Testten 30 dakika önce ağır egzersiz yapmaması,
- f. Göğüs ve karın hareketlerini kısıtlayıcı giysiler giymemesi,
- g. Testten 2 saat önce ağır yemek yememesi,
- h. Testten önce 6 saat süreyle kısa etkili bronkodilatör(bronş ve bronşiyollerin düz kaslarını gevşetmek suretiyle hava yollarının resistansını düşüren ilaçlar) almaması
- i. Testten önce (5-10 dakika) ve test sırasında (özellikle zorlu manevralarda) oturulmalıdır.
- j. Oda ısı ve barometrik basınç kaydedilerek BTPS düzeltmesi yapılmalıdır.
- k. Yapılacak test manevraları hastaya anlatılmalıdır (Akkoca 2004).

Günümüzde gelişen elektronik ve bilgisayar teknolojisi sayesinde taşınla bilen hasta kişinin bulunduğu konumda kolaylıkla kullanabileceği pratik aletler geliştirilmiştir.

Solunum fonksiyonu testini yaptıran hastanın test anında oturur pozisyonundadır. Hastanın burunu plastik bir mandalla kapatılarak sadece ağız bölgesinden soluk alması sağlanır. Hasta kendisinin kullandığı, cihazın ucunda yer alan tek kullanımlık ağızlıktan kolay bir şekilde ağızından nefes alıp nefes verir. Testi yapan hekim ya da teknisyen, test manevraları için uygulanan testin özelliklerine göre hastaya çok derin nefes almasını, çok hızlı ve sonuna kadar nefes vermesini veya kısa bir süre nefesini tutmasını isteyebilir. Testin sonucu olarak, yapılan manevralar tamamlandıktan sonra göğüs hastalıkları uzmanı tarafından sonuçlar yorumlanır (Savaş 2014).

Göğüs hastalıkları uzmanı tarafından yorumlanan solunum fonksiyonu test sonucunda ortaya çıkan hastalıklar,

- a. Akciğer kanseri
- b. Kronik bronşit
- c. Amfizemdir
- d. Alerjik durumlar
- e. Mesleki Hastalıklar
- f. Sigara Kullanımının Etkileri olarak sıralanabilir (Akkoca 2004).

1951’li yıllarda hava yolu tıkanmaları teşhisi için basınç hacim eğrisi Gaensler tarafından tanımlanmıştır. Gaensler tarafından tanımlanan basınç hacim eğrisinden sonra, akciğer fonksiyonları ölçümünün daha kolay bir durumda olması zorlu vital kapasite manevrasının geliştirilmesiyle olmuştur.

1950’lerin sonlarına gelindiğinde akciğerlerin hacim ve kapasitelerinin ölçülmesinde kullanılan cihazlar sayesinde motorize grafikler elde edilmeye başlanmıştır. Elde edilen grafikler kullanarak hacim zaman eğrilerindeki değerlerin ölçülmesi sayesinde akımların hesaplanması da mümkün olabilmektedir. 1960’ların başlarında akım- hacim eğrisi sayesinde hava yolu tıkanma sonuçlarının değerlendirilmesi daha kolay hale gelmiştir. İlerleyen yıllarda teknolojiye başlayan gelişmeler sayesinde solunum fiziolojisinde bilgisayarlı sistemler kullanılarak çalışmalar yapılmıştır. Daha sonraki yıllarda akciğer değerlerinin ölçümleri için birçok cihaz ve yeni testler geliştirilmiştir. Solunum fonksiyonları alanında yapılan birçok gelişmeler rağmen, akciğer değerlerinin ölçümünde en çok tercih edilen ve her yerde uygulanan spirimetrik testlerdir (Uçar 2012).

2.4 GENEL KAVRAMLAR

2.4.1 Spirometreler

Spirometre, farklı solunum manevraları durumunda akciğere giren ve akciğerden çıkan havanın ölçülmesidir. Bu incelemede, yavaş soluk alma sonrası, akciğerlerden dışarıya üflenebilen hava miktarı (SVC), zorlu vital kapasite (FVC) ve maksimal istemli solutma (MVV) manevraları ile ilgili değer ölçümleri yapılmaktadır. Spirometrik testlerde kullanılmak üzere farklı ölçüm metotlarına sahip farklı test cihazları geliştirilmiştir (ATS ve Am J 1995, ss. 1107-1136).

Farklı ölçüm yöntemleri olan spirometrik testlerde kullanılan bu cihazları iki grupta incelemek mümkündür:

a. Volüme duyarlı cihazlar

İlk geliştirilen spirometreler hacime duyarlı spirometlerdir. Volüme duyarlı spirometrelerin farklı tipte olanları bulunmaktadır. Bu spirometreler, sulu, kuru, körüklü ve diyaframlı tiplerdir. Bunlar içinde sulu spirometreler altın standart olarak kabul edilmektedir (Akkoca 2004).

Hacime duyarlı olarak geliştirilen spirometre cihazların avantajları ve dezavantajları bulunmaktadır.

- i. **Avantajlar:** Volüme duyarlı spirometre, ilk olarak hacimi ölçmektedir. Aynı zamanda maliyet olarak ucuz olan bir cihaz olup uygulaması basit olan bir cihazdır (Akkoca 2004).
- ii. **Dezavantajlar:** Volüme duyarlı spirometre cihazlarının büyük olmasından dolayı taşınmazlar, çıkan sonuçların elle hesaplanması gerekmekte olup cihazdaki hava kaçakları önemlidir. Sulu model olan spirometre cihazında suyunu düzenli olarak değiştirilmesi gerekir (Akkoca 2004).

b. Akıma duyarlı cihazlar

Geliştirilen bu cihazlar akımı ölçen cihazlardır. Akıma duyarlı cihazlarının sonuçlarının hesaplanması volüm ve akımın zamanın çarpımı sonucu ortaya çıkmaktadır. Akıma duyarlı olan cihazların pnömotakograf, termistor veya sıcak tel anemometresi, türbin cihazı ve vorteks tipleri vardır (Akkoca 2004).

Akıma duyarlı olarak geliştirilen bu cihazların avantajları ve dezavantajları bulunmaktadır.

- i. **Avantajlar:** Akıma duyarlı olan cihazlar küçük olması sayesinde taşınabilirler. Bilgisayarlı bir sistem olması sayesinde referans değerleri hızlı ve kolay hesaplanır ve akım hacim eğrisi grafiği çizdirilebilir (Akkoca 2004).
- ii. **Dezavantajlar:** Akıma duyarlı spirometre cihazların kullanımı deneyim ve dikkatli kalibrasyon gerektirir. Nem oranının birikmesi sorunların çıkmasına neden olurken, gaz içeriği sonuçları etkileyebilir ve düşük seviyedeki akımları gösteremeyebilir (Akkoca 2004).

2.4.2 Havayolu Fonksiyonlarını Gösteren Testler

Havayolu fonksiyonlarını gösteren beş adet test bulunmaktadır. Bu testler, basit spirometrik testler, zorlu vital kapasite testi, Maksimum solunum kapasitesi (MVV),

Maksimum inspiratuar ve ekspiratuar basınçlar (MIP, MEP) ve Hava yolu direnci (Raw) testidir.

2.4.2.1 Basit spirometrik testler

Maksimum soluk alma sonrası, akciğerlerden dışarıya üflenebilen hava miktarı hacmi vital kapasite (VC), nefes vermede saklanan hacim ekspiratuar rezerv volüm (ERV), nefes almada saklanan hacim inspiratuar rezerv volüm (IRV), soluk hacmi (V_T), soluk verme zamanı (T_I), soluk alma zamanı (T_E), toplam solunum zamanı (T_{Tot}), soluk verme kapasitesi (IC) gibi testler basit spirometrik testlerdir (Yıldırım 2003).

2.4.2.2 Zorlu vital kapasite

a. Zorlu vital kapasite (FVC), 1. saniyedeki zorlu vital kapasite (FEV1), maksimum ekspiryum ortası akım hızı (FEF yüzde 25-75), tepe ekspiratuar akım hızı (PEF). Bu değerlerden özellikle FEV1 nin akut bronkodilatör ile ve bronkokonstriktör maddelerle değişimi,

b. Maksimum ekspiratuar -inspiratuar akım -volüm halkası (Yıldırım 2003).

2.4.2.3 Maksimum solunum kapasitesi (MVV)

Genişlik ve frekansı yüksek solunumda dakikada atılan hacim'e maksimum solunum kapasitesi denir. Minimum 12 saniye boyunca devamlı olarak , sistemli ve ritmik efor ile soluk alınır ve verilir. Minimum iki manevra durumunda, yüzde 10'luk değişim alanları içinde kabul edilebilir. Bu test birlikte çalışmaya ve efora bağlı bir testtir. Havayolu direnci, solunum kasları, akciğer ve göğüs duvarı basınç karşısındaki genişleyebilmesinden etkilenir.

Maksimum solunum kapasitesi yüzde 30'dan fazla olan azalmalar önemli ve maksimum solunum kapasitesi kadınlarda ve erkeklerde farklılıklar göstermektedir.

Beklene değerler,

- i. Kadınlara için: beklenen MVV= FEV1 x 40
- ii. Erkekler için: beklenen MVV= FEV1 x 34' dir (Akkoca 2004).

2.4.2.4 Maksimum inspiratuar ve ekspiratuar basınçlar (MIP, MEP)

Solunum havayolunu kapatan kapakçıya karşı yapılan maksimum nefes alma ve nefes verme durumunda ölçülen ağız içi basınçlarıdır. Maksimum soluk alma ve soluk verme basınçları solunum kas gücünü gösteren testlerdir (Öğüş 2009).

Soluk alma kas gücü, doğrudan kapalı havayoluna karşı maksimum durumda soluk alma esnasında yapılırken ölçülür. MIP, kalan hacime kadar zorlu soluk verme yapıldıktan sonra yapılan maksimum soluk verme manevrası esnasında ölçülen maksimum negatif basınçtır (Çoşkun [tarih yok])

Maksimum soluk verme basıncı, toplam akciğer kapasitesi seviyesinde ölçülür. Mep, solunum özelliğini belirlemede daha zayıf orandaki sonuçlar görülse de bulguların öksürebilme özellikleri hakkında bilgi vermektedir. Testler sırasında, güçlü olmayan öksürük durumu solunum kas gücü sınırının göstergesidir. MIP ve MEP testleri taşınabilir cihazlar sayesinde ölçümler yapılabilir (Çoşkun [tarih yok]).

2.4.2.5 Havayolu direnci (Raw)

Akımın her bir alanı durumuna karşı ağız basıncı ve alveol basıncı arasındaki fark hava yolu direncidir. Meydana gelen bu basınç farklı iletici hava yollarında yer alan, gaz moleküllerinin sürtünmesi sonucunda meydana gelir. Birim akım sonucu meydana gelen basınç farklılığı havayolu çapını gösterir. Obstrüktif ve restriktif hastalıkların ayırımında kullanılan önemli bir testtir (Öğüş 2009).

Solunum fonksiyonları testinde yapmış olduğumuz çalışmamız “Zorlu Vital Kapasite Manevrası” üzerinedir.

2.4.3 Zorlu Vital Kapasite Manevrası

Ventilatuar kapasitenin değerlendirilmesinde zorlu vital kapasite manevrası önemli bir testtir. Sınır sistemi, kas-iskelet sistemi ve akciğerde meydana gelen hücrelerdeki, dokulardaki ve organlardaki yapısal ve işlevsel değişikliklere bağlı olarak ventilatuar kapasitede bozulma olabilir. Fakat bu meydana gelen problemin nedeni havayolları tıkanmasıdır. Büyük akciğer hacimleri seviyesindeki maksimum akımlar durumunda yüksek oranda ana soluk borusundaki ve ana bronşların akım özelliklerini gösterirken, düşük hacimdeki akımlar periferik intratorasik havayollarını göstermektedir. Zorlu vital kapasite manevrası hacim-zaman ve akım-hacim eğrileri sonuçlarına bakılarak değerlendirilir (Akkoca 2004).

2.4.3.1 Zorlu vital kapasite (FVC)

Zorlu soluk alma durumundan sonra yapılan zorlu soluk vermeyle dışarıya çıkarılan hava hacminde havayollarında daralma veya tıkanma olduğu zaman azalır. Zorlu vital kapasite soluk verme zamanı 20 saniyenin üstüne çıkar ve göğüs duvarı ve akciğerin nefes alma sonucunda genişleyememesi durumundan meydana gelen hastalıklarda azalır. Vital Kapasite oranındaki ve zorlu vital kapasitede oranındaki azalma birçok sebepler bu bulgulara neden olabilmektedir. Her ikisi arasında fark oluşması havayolu kollapsını gösterir (Akkoca 2004).

2.4.3.2 Birinci saniye zorlu ekspirasyon volümü (FEV1)

Birinci saniye zorlu nefes verme hacmi sağlıklı insanlarda 60-270 mililitredir. FEV1 değeri büyük havayollarını göstermenin yanı sıra havayolu tıkanmasında mukus salgısı, bronkospasmi, yanma, iltihaplanma ve elastik doku kaybından dolayı FEV1 oranı azalır. FEV1 oranı akciğer tıkanma hastalığının ilerleyişini ve ne kadar zaman devam edeceğini tahmin edilmesini gösteren bir değişkendir. Nefes alma durumunda göğüs kafesinin genişleyememesi sonucunda nefes alma işleminde zorluk yaşanması durumunda meydana gelen yapısal ve fonksiyonel bozukluklarda FVC oranının azalmasına bağlı olarak FEV1 oranı da azalır (Akkoca 2004).

2.4.3.3 FEV1/FVC (Tiffeneau oranı)

Havayolu tıkanması ve akciğerin nefes alma esnasında göğüs duvarı veya akciğerin genişleyememesinden meydana gelen hastalıkları ayırt etmede kullanılmaktadır. Havayolu tıkanmalarında FEV1 değeri FVC değerinden daha fazla oranda azalma

göstermesinden dolayı oran yüzde 70 den küçükken den daha fazla azalma gösterdiğinden dolayı oran genellikle yüzde 70 küçük iken, restriksiyonlarda iki durumda da aynı oranda azalmasından dolayı normal kalır. Ayrıca FEV1/FVC oranı havayolu tıkanmasının derecelendirilmesi de kullanılmaktadır (Akkoca 2004).

Tablo 2.1 : Obstrüktif bozukluğun şiddeti

Obstrüksiyon	FEV1/FVC (%)
Normal	> 70
Hafif	61 – 69
Orta	45 – 60
İleri	< 45

Kaynak : http://file.toraks.org.tr/TORAKSFD23NJKL4NJ4H3BG3JH/kisokulu3-ppt-pdf/Oznur_Akkoca.pdf

[Erişim : 20 Nisan 2014]

2.4.3.4 Tepe akım hızı (PEF)

Maksimum soluk alma durumundan sonra hızlı bir şekilde soluk verme durumu değerlendirilir. Tepe akım hızında 1-2 saniyelik yapılan güce bakılır. Maksimal soluk alma durumunda, uzun süre bekleme PEF değerini düşürür. Tepe akım hızı, havayolları problemlerin belirli aralıklarla takibinde kullanılmasının yanı sıra büyük havayolları durumunu da göstermektedir (Akkoca 2004).

Solunum fonksiyon testlerinde yapılan ölçümlerde yaş, cinsiyet, boy, kilo, alerjik durumu, aile geçmişi, meslek grubu ve sigara kullanımı testler üzerinde etkisi olduğu görülmektedir.

2.4.4 Sigara Kullanımının SFT Etkisi

Günümüzde sigara kullanımı toplumlar içerisinde önemli sağlık sorunlarından bir tanesidir. Sigara endüstrisi, her geçen gün dünyada büyümekte ve insanların dikkatini çeken reklamlar yapılarak sigara kullanıcılarının sayısını her geçen gün arttırmaya çalışırken, bağımlılık yapıcı etkiye de sahiptir. Sigara sadece kullanan kişilere zarar veren bir sağlık sorunu olmayıp sigara durumuna maruz kalanlar içinde bir sağlık sorunu olmaktadır.

Sigara kullanan kişilerde, ortaya çıkan sağlık problemi solunum havayollarında meydana gelen hastalıklardır. Sigara kullanımı sonucunda meydana gelen problemler akciğer kanseri, kronik bronşit ve amfizendir (Balcı 1993).

1995 yılında sigara kullanımının solunum fonksiyonları testi üzerine yapılan bir çalışmada 134 sigara kullanan ve 58 sigara kullanamayan toplam 192 kişiyle yapılmıştır. Sigara kullanan kişilerde ortalama yaş 40 iken, sigara kullanmayanlarda 38 dir (Akkaya ve Ünlü 1995).

Yapılan çalışmada iki grup bulgularının solunum fonksiyonu testi değerleri ve çalışma grubunun standart olarak kabul edilen değerleriyle karşılaştırıldı ve yüzde oranlar bulundu. Sigara kullanan ve sigara kullanmayan kişilerde solunum fonksiyon testinin ortalama değerleri tablo 2.2 görünmektedir (Akkaya ve Ünlü 1995).

Tablo 2.2: Sigara içenler ve içmeyenlerde SFT (%) ortalama değerleri

	Sigara İçen (n=134)		Sigara İçmeyen (n=58)	
	%	SD	%	SD
FVC(L)	92.26	13.30	95.03	9.65
FEV1(L/1.sn)	90.31	17.78	94.58	10.20
PEF(L/sn)	78.66	27.99	79.48	26.02

Kaynak : <http://edergi.sdu.edu.tr/index.php/sdutfd/article/viewFile/128/409>

[Erişim : 25 Nisan 2014]

Şekil 2.3 : Sigara içenler ve içmeyenlerde STF (%) ortalama değerleri

Kaynak : <http://edergi.sdu.edu.tr/index.php/sdutfd/article/viewFile/128/409>

[Erişim : 25 Nisan 2014]

Sigara kullanan ve sigara kullanmayanların ortalama FVC (zorlu vital kapasite) ve FEV1 (1.saniye zorlu ekspirasyon volümü) değerleri özetlendikten sonra grafiksel olarak gösterimi tablo 2.3 de gösterilmektedir (Akkaya ve Ünlü 1995).

Tablo 2.3: Sigara içenler ve içmeyenlerde STF (%) ortalama değerleri

	Sigara İçen (n = 134)		Sigara İçmeyen (n=58)	
	%	SD	%	SD
FVC (L)	3.72	0.83	4.09	0.70
FEV1(L/1.sn)	3.00	0.83	3.36	0.59

Kaynak : <http://edergi.sdu.edu.tr/index.php/sdutfd/article/viewFile/128/409>

[Erişim : 25 Nisan 2014]

Şekil 2.4 : Sigara içenler ve içmeyenlerde STF (%) ortalama değerleri

Kaynak : <http://edergi.sdu.edu.tr/index.php/sdutfd/article/viewFile/128/409>

[Erişim : 25 Nisan 2014]

Sonuç olarak sigara kullanımı durumunda belirgin problemler olmasa da FEV1 değerlerinde (sigara içenlerde: ort 3.00 lt, % 90.31; içmeyenlerde: 3.36 lt, % 94.58) azalma olduğu görülmektedir (Akkaya ve Ünlü 1995).

2.4.5 Allerjik Ve Allerjik Olmayan Hastaların SFT Etkisi

Astım hastalığı çevresel etkenlerin etkisiyle ortaya çıkan ve solunum havayollarında uzun zamandır devam eden yanma ve iltihaplanmalara bağlı olarak oluşan tıkanma ile seyreden bir hastalıktır (Karakaya 2004, ss. 25–31).

Astım hastalığı kişilerde genetik ve çevresel etkenler olarak iki grupta ortaya çıkmaktadır. Risk faktörleri; kişiyi astıma yatkın kılan kişisel risk faktörleri ve genetik olarak astıma yatkın olanlarda astım gelişimine yol açan çevresel faktörler olmak üzere iki grupta toplanabilir. Genetik olarak astımın ortaya çıkmasında erkeklerde erken dönemlerde, kızlarda ergenlikten sonra ortaya çıkmaktadır. Genetik astımında obezitenin etkisi de görülmektedir. Çevresel astımda, astım alevlenmesine sebep olan alerjenler, enfeksiyonlar, çalışılan meslek ortamları, sigara kullanımı ve sigara dumanına maruz kalınan ortamlar gibi etkenlerden meydana gelmektedir. (GINA 2006, Ober 2005, Guler ve diğ. 2004)

Solunum fonksiyon testinde değişken hava akımı obstrüksiyonunun gösterilmesi tanıyı desteklemesi açısından önemlidir (GINA 2006). Erişkin insanlarda solunum fonksiyonu testinde FEV1/FVC oranı normal yüzde 75- 80'in üstündedir. Bu oranın altındaki

sonuçlarda hava akımı kısıtlanmasını gösterir (Bateman ve diğ. 2008). Solunum havayollarında tespit edilen tıkanma yaşayan hastaya reverzibilite testine başvurulabilir (Türk Toraks Dergisi 2009, s.10).

Ülkemizde 2007 yılında alerjik olan ve olmayan hastaların üzerinde yapılan bir araştırma astım tanısı konulan hastalarda geriye dönük olarak alerjik durumlarına bakılarak yapılan bir çalışmadır. Astımlı hastalarda allerji durumunun erken zamanda tespit edilmesi hastaların tedavi ve teşhisinde önemlidir (Talay ve diğ. 2008).

Yapılan çalışmalar sonucunda, alerjik durumu olmayan astım hastaların ileri yaşlardaki bayanlarda ve eğitim düzeyinin düşük olması düşük solunum fonksiyon oranlarıyla ilişkili olduğu tespit edilmiştir. Aynı zamanda alerjik olmayan astımlarda nefes darlığı problemi görülmektedir. Alerjik olan astımlarda burun akıntısı, burun tıkanıklığı ve hapşırma gibi problemler görülürken, alerjik astımların polenlere ve akarlara karşı daha duyarlı oldukları belirlenmiştir. Eğer astım problemine sahip olan hastalarda erken tespit ve tedavi önemlidir (Talay ve diğ. 2008).

2.4.6 Meslek Hastalıklarının Tanısında SFT Etkisi

Gelişmekte olan ülkelerde mesleki hastalıklardan kas – iskelet problemleri ve psikososyal bozuklukların dışında solunumsal bozuklarda mesleğe bağlı sağlık problemlerini oluşturmaktadır (Snashall 1997).

İngiltere'de 1980'li yıllardan günümüze akciğerde toz bulunması ve birikmesi problemleri azalırken, mesleğe bağlı astımın yükselme eğilimi olduğu görülmüştür ve 1994 yılında mesleğe bağlı olarak ortaya çıkan solunum hastalıklarının başında meslek astımının olduğu bildirilmiştir. Sanayinin geliştiği ülkelerde meslek astımının en önemli mesleki solunum probleminin olduğu durumu desteklemektedir.(Madan 1997).

Mesleki hastalıklardan dolayı oluşan solunum fonksiyonu hastalıklarının başında mesleki astım gelmektedir. Meslek astımı, sanayileşen toplumlarda en önemli solunumsal meslek hastalığı iken, ülkemizdeki resmi veriler silikozis gibi toz hastalıklarının hâlâ en fazla gözlenen mesleksel akciğer hastalıkları olarak göstermektedir (Çımrın 2000, s. 87).

Mesleğe bağlı olarak ortaya çıkan hastalıklarda en çok kullanılan testlerden biri solunum fonksiyon testleridir. Solunum fonksiyonları testleri mesleğe bağlı hastalıkların tanısında nefes darlığının değerlendirilmesi, tıkanma ve akciğerlerin genişlememesinden dolayı nefes alma sorunlarının değerlendirilmesinde önemli testlerdir (Gulati ve Redlich 2008, ss. 933-942.).

Türkiye de artan nüfusun yaklaşık olarak yüzde 40'ı çalışmaktadır. Gelişen sanayileşmede çalışan işçi sayısının artmasında solunum fonksiyonu problemi yaşayanların sayısı da artmaktadır (Çımrın 2000). Ülkemizde yapılan araştırmalara göre meslek astımı riskli olduğu bilinen meslek dallarından tarım-orman-avcılık, madencilik, petrol-kimya-lastik, gıda, dokuma, deri, ağaç, kağıt, basın-yayın, metal, gemi ve enerji işlerinde çalışan işçilerin iki milyonun üzerinde olmakla birlikte, işyeri sayıları 200 000 oranındadır. (TC Çalışma ve Sosyal Güvenlik Bakanlığı 1998).

Ülkemizde yapılan mesleki astım arařtırmalar sonucu meslek gruplarına göre astım oranları ařağıdaki Tablo 2.4 de görünmektedir.

Tablo 2.4 : Ülkemizde yapılan mesleki astım arařtırmaları

Arařtırma grubu	n	Astım prevalansı (%)	Kaynak
Bayan kuaförü	184	13.6	Akpınar M
Afyon alkaloidleri fabrikası	43	11.6	Ardıç S
Oto ve mobilya boyacıları	312	9.6	Uçgun I
Bayan kuaförü	151	6.0	Fişekçi F
Fırın işçileri	158	5.7	Kılıçaslan Z
Mobilyacı ve marangoz	163	2.5	Erdoğan S
Halıcılık öğrencileri	143	0.7	Görgüner M
Yün halı fabrikası işçileri	197	0.5	Güven K

Kaynak :http://www.tuberktoraks.org/managete/fu_folder/2003-01/2003-51-1-078-090.pdf
[Eriřim : 2 Mayıs 2014]

Yapılan çalışma incelendiğinde dokuma mesleğinde çalışan işçilerinde pamuk, keten ve kenevir tozlarından uzun zamandır etkilenmelerinden dolayı ortaya çıkan akciğer hastalığı yüzde 2-22 oranındayken meslek astımı problemi bu işkolunda yüzde 0.6-2 oranında bir sonuç ortaya çıkmıştır. Meslek astımı problemleri sprey boyacıları mesleğinde çalışan işçilerin arasında yüzde 0-10 oranında, kuaför mesleğinde çalışan işçilerde yüzde 1.72 - 13.6 oranında, orman ürünleri mesleğinde çalışan işçilerde yüzde 2.5 oranında, sağlık mesleğinde çalışan işçilerinde yüzde 14 oranında, toz morfin işçilerinde yüzde 11.6 oranında, deterjan mesleğinde çalışan işçilerde yüzde 7 oranında, fırın mesleğinde çalışan işçilerde yüzde 3.3 oranında, cam süsleme mesleğinde çalışan işçilerde yüzde 7.3 oranında, çiçek satıcılığı mesleği işçilerinde yüzde 30.4 oranında olduğu belirlenmiştir. Bu sonuçların doğrultusunda meslek gruplarında solunum fonksiyonuna etkisinin olduğu görülmekte ve bu durumun meslek astım ile hangi oranda ilişkili olduğunu ortaya koymaktadır (Çımrın 2000).

Solunum sistemi dış çevre ile direk olarak etkileşim halinde olduğundan dolayı dış dünyadaki birçok faktör akciğerlere kolay bir şekilde ulaşabilmektedir. Çalışma ortamında bulunan tozlar ve kimyasal maddeler solunum yollarında tahriş edici etki yapabilir veya akciğerlerde birikebilir ve farklı tepkiler ortaya çıkabilir (Alpay 2008).

Solunum fonksiyonları testi mesleki hastalıklarda en çok mermer işçilerinin yaşamış olduğu problemleri ortaya koymaktadır. Mermer tozunun tepkimeye girmeyen etkisiz (inert) toz olduğu, fakat mermer tozunun içindeki silika oranına göre silikozise yol açabileceği bildirilmektedir (Orman ve diğ. 2002, ss. 188-193).

Mermer işinde çalışan işçilerde, toza maruz kalmalarından dolayı ortaya çıkan solunumsal belirtileri incelemek için bir çalışma yapılmıştır. Çalışma oranlarının ve işyerlerinde çalışan işçilerin genel özelliklerini tespit etmek için işyerlerinde bir ön çalışma yapılmıştır. Ocakta işinde çalışan işçiler 42, atölye işinde çalışan işçiler 54, fabrikada işinde çalışan işçiler 140'tır. Aynı fabrikada teknisyen ve memur olarak

çalışan işçilerin 102 kişi olmak üzere bu işçilerin tamamı erkektir ve toplamda 338 kişi işe alınmıştır. (Orman ve diğ. 2002, ss. 188-193).

Ocak mesleğinde çalışan işçi grubun diğer meslek iş gruplarına oranla uzun süredir çalıştıkları anlaşılmıştır. Atölye mesleğinde çalışan işçi grubunda olan ve memur-teknişyen mesleğinde çalışan işçilerin yaşlarının aynı oranda oldukları ve diğer meslek grubundaki işçilerin yaş ortalamalarının düşük olduğu anlaşılmıştır (Orman ve diğ. 2002, ss. 188-193).

Tablo 2.5 : Grupların Demografik Özellikleri

Grup	N	Yaş	Çalışma Süresi	Sigara İçme (%)
Ocak	42	35.57 ± 9.10 #	16.38 ± 6.62*# §	28 (66.7)
Atölye	54	27.10 ± 9.56 §	9.14 ± 6.88 *	34 (63.0)
Fabrika	140	32.24 ± 8.88 *	9.05 ± 7.39 #	88 (62.9)
Kontrol	102	27.54±7.39*# §	8.21 ± 5.93 §	64 (62.7)

* p < 0.05; # p < 0.05; § p < 0.05

Kaynak : <http://toraks.dergisi.org/text.php3?id=220>

[Erişim : 5 Mayıs 2014]

Meslek gruplarının çalıştıkları ortamlarındaki toz incelemesinde, SiO₂ çok düşük yüzdelerde belirlenmiş olup, solunabilir toz oranı çalışma gruplarıyla karşılaştırıldığında tozun en fazla olduğu çalışma alanının atölye meslek grubunda olduğu ve fabrika ve ocak meslek grubunun bunu durumu izlediği görülmüştür (Orman ve diğ. 2002, ss. 188-193).

Tablo 2.6 : Grupların çalıştıkları yerdeki toz ve silika konsantrasyonları

Grup	Solunabilir toz (mg / m ³)	SiO ₂ mg / m ³
Ocak	6.5	0.04
Atölye	11.7	0.09
Fabrika	8.4	0.06

Kaynak : <http://toraks.dergisi.org/text.php3?id=220>

[Erişim : 5 Mayıs 2014]

Mesleki çalışma gruplar solunum problemi yaşamaları durumundan karşılaştırıldığında, memur ve teknişyen gruplarının diğer meslek gruplarına göre daha az olduğu, ocak ve fabrika iş grubunda çalışanlarda balgam problemi yaşamaları atölye ve memur-teknişyen meslek grubuna göre daha fazla oranda olduğu görülmüştür (Orman ve diğ. 2002, ss. 188-193).

Tablo 2.7 : Grupların solunum yakınmaları

Grup	Öksürük n (%)	Balgam Çıkarma n(%)	Nefes Darlığı n(%)
Ocak	18 (42.9) §	20 (47.6) #	8 (19)
Atölye	18 (34.6) #	14 (25.9)	6 (11.1)
Fabrika	50 (35.7) *	62 (44.3) *	30 (21.7)
Memur-Teknişyen	22 (21.6) *# §	30 (29.4) *#	12 (11.8)

* p < 0.05; # p < 0.05; § p < 0.05

Kaynak : <http://toraks.dergisi.org/text.php3?id=220>

[Erişim : 5 Mayıs 2014]

Mermer tozuna maruz kalmanın sonucunda akciğer sorunlarına yol açabileceği anlaşılmıştır. Mesleki grupların solunum problemi yaşamaları bakımından karşılaştırıldığında öksürüğün ocak, atölye ve fabrika alanında çalışan grubunda memur-tekniyen alanında çalışan grubuna göre istatistiksel olarak anlamlı olduğu, balgam problemini yaşayanların ocak ve fabrika grubunda diğer gruplara göre daha fazla görüldüğü anlaşılmıştır. Memur-tekniyen işinde çalışan grubun işyeri ortamı ve iş özellikleri dolayısıyla tozdan etkilendikleri fakat bu oranın diğer diğer iş grupları kadar olmadığı ortaya çıkmıştır. Mermer işinde çalışan işçi gruplarının, mermer tozundan etkilenmelerinden dolayı solunum problemi sonuçları ortaya çıkmaktadır (Orman ve diğ. 2002, ss. 188-193).

Mesleki hastalıklardan dolayı meydana gelen solunum fonksiyonları problemleri için bazı testler uygulanmaktadır. Uygulanan bu testler spirometrik testi, reversibilite testi, bronş provokasyon testi, difüzyon testi ve akciğer volümleri testleridir.

2.4.6.1 Spirometrik Testi

Solunum fonksiyonlarındaki nefes alma esnasında akciğerlerin genişlememesi (retriktif) ve tıkanma (obstrüktif) gibi teşhisinde kullanılan testtir. Spirometrik çalışmada FVC değeri, FEV1 değeri ve FEV1/FVC oranı kullanılır. Nefes alma sonucunda akciğerin gelişmemesinden meydana gelen hastalıklarda FVC azalırken, tıkanmadan dolayı oluşan hastalıklarda FEV1/FVC oranının ve FEV1'in azaldığı görülür (Demir 2010).

2.4.6.2 Reversibilite Testi

Genellikle meslek astımının teşhisinde kullanılan bir testtir. Bu teste verilen nefes genişletici ilaçlardan sonra FEV1 ve FVC oranların sonucundaki değişimi değerlendirilir (Demir 2010).

2.4.6.3 Bronş Provokasyon Testi

Bazı mesleklerde oluşan meslek astımı olgularında hava yolu yangı veya iltihaplanmaya (inflamasyon) yol açan hastalıkları ölçmek amacıyla kullanı testtir (Demir 2010).

2.4.6.4 Difüzyon Testi

Mesleğe bağlı olarak hava keseciklerin etrafındaki kalınlaşmadan dolayı (interstisyel) akciğer hastalıkları teşhisinde kullanılan testtir (Demir 2010).

2.4.6.5 Akciğer Volümleri

Vücut pletismografi ya da gaz dilüsyon yöntemi ile ölçülebilen akciğer volümleri restriktif bozukluğun saptanmasında basit spirometreden çok daha duyarlıdır. Yine obstrüktif hastalıklarda hiperinflasyonun gösterilmesi ve buna bağlı gelişen FVC'deki azalmanın restriktif defektten ayrımında akciğer volümlerinin ölçümü gerekmektedir (Demir 2010).

Kömür işçilerinde yapılan arařtırmada, 8000 kömür işçisinin verileri üzerinde yapılan çalıřma sonucunda ortaya çıkan olguların yüzde 6'sında FEV1/FVC oranı 2 olarak çıkmaktadır (Kibelstis ve diğ. 1973, ss. 886-893).

Kömür işçilerinin akciğçerlerinde biriken kömür tozunun dokulara verdiđi hasar sonucunda (pnömokonyoz) solunum fonksiyonlarındaki etkilenme radyolojik bulguların sonucunda aynıdır. Kömür işçilerinden de görülen řiddetli solunum bozukluđunun (progresif masif fibrozisin) hakim olduđu durumlarda fonksiyonel durum kesin ve ağır olurken, daha az akciğçerler de biriken kömür tozunun durumu daha az düzeyde görülür. Fakat birikmiř (kümülatif) tozun solunum fonksiyon kaybındaki etkisinin en önemli özellik olduđu gösterilmiřtir (Attfield ve Hodous 1992, ss. 605-609).

Ayrıca kömür tozundan etkilenen kiřilerin ilk zamanlarda solunum fonksiyonu problemi daha belirginken, daha sonraki zamanlarda kömür tozuna bađlı solunum fonksiyon problemi azalmaktadır (Petsonk ve Parker 2008, ss. 967-980).

Arařtırmanın sonucu olarak kömür tozundan etkilenen kiřilerde, solunum fonksiyonları üzerinde etkisinin var olduđu görünmektedir (Demir 2010).

3. VERİ VE YÖNTEM

Yöntem bölümünde araştırmamızın problem çözümlerine yer verilmiş olup araştırmanın modeli, araştırmanın örnekleme, veri toplama araçları, verilerin toplanması ve toplanan verilerin çözümlenmesinde kullanılan Weka uygulamasında yer alan veri madenciliği yöntemleri ve uygulamaları dikkate alınmıştır.

3.1 ARAŞTIRMANIN MODELİ

Araştırmamızda, veri madenciliği modellerini uygulamak amacıyla WEKA veri madenciliği yazılım programı kullanılmıştır. WEKA veri madenciliği yazılım programı Java programlama dilinde yazılmış olup, bilgisayar işletim sistemleri olan Windows, Linux ve Macintosh gibi farklı işletim sistemlerinde çalışabilen bir programdır (Witten ve Frank 2000, ss. 267-277). Weka veri madenciliği modeli, solunum problemi şikayeti olan hasta verilerinde, birden fazla Weka'da yer alan modelleriyle çalışılarak doğruluk oranı en yüksek olan model seçimi için kullanılmıştır.

3.2 VERİ TOPLAMA VE ANALİZ

3.2.1 Örneklem

Yürütülen çalışma, Gaziosmanpaşa Özel Hastanesi Göğüs Hastalıkları Anabilim Dalı Solunum Fonksiyon Laboratuvarında 2012 -2014 yılları arasında test yaptıran hastalardan oluşturmaktadır.

Hastaların test sonuçlarına ulaşabilmek için Gaziosmanpaşa Özel Hastanesinden yazılı onay alındıktan sonra hastaların yaş, cinsiyet, boy, kilo, alerjik durumu, aile geçmişi, mesleği, sigara kullanımı, FVC ve FEV1 veri değerleri not edildi.

Yaşadıkları solunum rahatsızlığı şikayetinden dolayı test yaptıran hastalar rastgele seçilmiş hastalardır. Yapılan araştırmada 400 solunum fonksiyonu testi yaptıran hastaya ulaşılmıştır.

3.2.2 Verilerin Analizi

Toplanan veriler WEKA programı kullanılarak analiz edilmiş olup gelecekte hangi yaş grubunun, hangi cinsiyetin, hangi kilo ve boyda olan kişilerin solunum fonksiyonu rahatsızlığına yakalanabilecekleri tespit edilmesidir.

4. BULGULAR

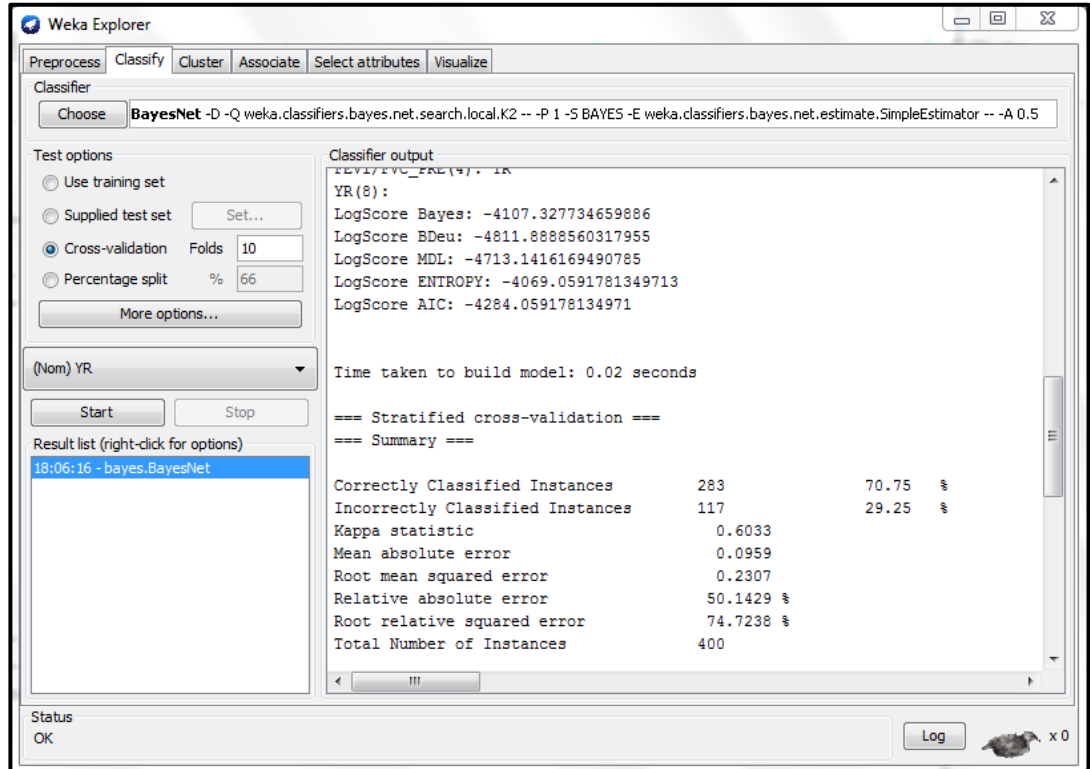
4.1. SINIFLANDIRMA YÖNTEMLERİ

Araştırmamız solunum problemi yaşayan kişilerin hastalıklarının tespiti ve tahminleri için veri madenciliği sınıflandırma yöntemleri kullanıldı. Solunum fonksiyonu problemi yaşayan kişilerin birçok nedenden dolayı bu sorunlar karşılaştıkları tespit edilmiştir. Yapmış olduğumuz araştırmada hastaların yaş, cinsiyet, boy, kilo, alerjik durumu, aile geçmişi, mesleği ve sigara kullanımını gibi değerlere bakılarak çalışılmıştır.

4.1.1 BayesNet Modeli

Solunum fonksiyonları testi üzerinde yapılan veri madenciliği sınıflandırma çalışmasında, BayesNet sınıflandırma modeli sonucunda 400 solunum fonksiyonu verilerinden 283 tanesi (yüzde 70.75) doğru sınıflandırılmış olup, 117 tanesi (yüzde 29.25) hatalı olarak sınıflandırılmıştır.

Şekil 4.1 : BayesNet istatistiksel değerleri



Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.1'de, BayesNet hakkında ayrıntılı bilgiler görünmektedir. BayesNet modelin doğru sonucun test edilmesinde, 400 solunum problemi yaşayan hasta veri kümesinde Cross-validation (çapraz doğrulama) testi uygulanmıştır. BayesNet modeli sonucunda 283 verinin (yüzde 70.75) doğru sınıflandırılmış olup, 117 veri (yüzde 29.25) yanlış sınıflandırılmıştır.

Şekil 4.2 : Bayes Networks doğruluk değerleri

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.873	0.241	0.625	0.873	0.728	0.861	8
	0.571	0.016	0.571	0.571	0.571	0.977	6
	0.538	0.016	0.538	0.538	0.538	0.784	3
	0.8	0.034	0.552	0.8	0.653	0.981	1
	0.179	0.035	0.455	0.179	0.256	0.777	7
	0	0.008	0	0	0	0.886	2
	0.063	0.024	0.182	0.063	0.093	0.794	4
	0.992	0.007	0.985	0.992	0.989	1	5
Weighted Avg.	0.708	0.088	0.663	0.708	0.667	0.897	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.2’de BayesNet modelinde detailed Accuracy By Class alanını bakıldığında Class 8 – 6 – 1 ve 5’in ROC Area alanındaki değerlerin 1’e yakın değerler olduğu gözükmemektedir. BayesNet modeli başarılı bir sonuç vermiş bulunmaktadır.

4.1.2 NaiveBayes Modeli

Diğer bir sınıflandırma modeli olan NaiveBayes modelini solunum fonksiyonları verileri üzerinde çalıştırıldığında, NaiveBayes sınıflandırma modeli sonucunda 400 solunum fonksiyonu verilerinden 282 tanesi (yüzde 70.5) doğru sınıflandırılmış olup, 118 tanesi (yüzde 29.5) hatalı olarak sınıflandırılmıştır.

Şekil 4.3 : NaiveBayes istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.3’de, Naive Bayes Ağı hakkında ayrıntılı bilgiler görünmektedir. Naive Bayes modelin doğru sonucun test edilmesi durumunda 400 solunum problemi yaşayan hasta veri kümesinde Cross- validation (çapraz doğrulama) testi yapılmıştır ve Naive Bayes ile ilgili detaylandırılmış farklı bilgiler elde edilmiştir. Naive Bayes modeli sonucunda 282 verinin (yüzde 70.5) doğru sınıflandırılmış olup, 118 veri (yüzde 29.5) yanlış sınıflandırılmıştır.

Şekil 4.4 : Naive Bayes doğruluk değerleri

=== Detailed Accuracy By Class ===									
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,905	0,252	0,623	0,905	0,738	0,609	0,859	0,651	8
	0,500	0,021	0,467	0,500	0,483	0,464	0,977	0,465	6
	0,538	0,018	0,500	0,538	0,519	0,502	0,763	0,360	3
	0,700	0,037	0,500	0,700	0,583	0,566	0,979	0,653	1
	0,161	0,029	0,474	0,161	0,240	0,215	0,775	0,303	7
	0,000	0,008	0,000	0,000	0,000	-0,012	0,893	0,172	2
	0,031	0,014	0,167	0,031	0,053	0,039	0,796	0,181	4
	0,992	0,007	0,985	0,992	0,989	0,983	1,000	0,999	5
Weighted Avg.	0,705	0,090	0,656	0,705	0,657	0,608	0,896	0,653	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.4’de görüldüğü gibi BayesNet’in FP ve TP oranlarına göre FP oranı ve TP oranının artmasından dolayı hata yüzdesi daha fazladır.

4.1.3 Logistic Modeli

Logistic sınıflandırma modelini çalıştırdığımızda Naive Bayes ve BayesNet sınıflandırma modeli sonuçlarına göre daha kötü bir sonuç elde etmekteyiz. Root mean squared error sonucunun yüzde 0.2606 ve doğru sınıflandırmanın yüzde 67.75 olmasından dolayı Naive Bayes ve BayesNet modülünden daha kötü olduğunu göstermektedir.

Şekil 4.5 : Logistic istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.5’de, Logistic hakkında ayrıntılı bilgiler görünmektedir. Logistic modelin doğruluğunun sonucu olarak test edilmesinde ise 400 solunum problemi yaşayan hasta veri kümesinde Cross- validation (çapraz doğrulama) testi çalıştırıldığında Logistic ile ilgili detaylandırılmış farklı bilgiler elde edilmiştir. Logistic modeli sonucunda 271 verinin (yüzde 67.75) doğru sınıflandırılmış olup, 129 veri (yüzde 32.25) yanlış sınıflandırılmıştır.

4.1.4 Multiplayer Perceptron Applicaton Modeli

Multiplayer Perceptron Applicaton sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 276’sı (yüzde 69) doğru sınıflandırılmışken, 124 veri (yüzde 31) yanlış olarak sınıflandırılmıştır. Root mean squared error yüzde 0.2415 olmasından dolayı başarılı bir sınıflandırma modeli olmadığı görülmektedir.

Şekil 4.6 : Multiplayer perceptron istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.6’da, Multiplayer Perceptron Applicaton sınıflandırma modelinde hakkında ayrıntılı bilgiler görünmektedir. Multiplayer Perceptron Applicaton modelin doğru sonucun test edilmesinde, 400 solunum problemi yaşayan hasta veri kümesinde Cross-validation (çapraz doğrulama) testi yapılmıştır. Multiplayer Perceptron Applicaton ile ilgili detaylandırılmış farklı bilgiler elde edilmiştir.

Şekil 4.7 : Multipler Perceptron doğruluk değerleri

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.762	0.164	0.681	0.762	0.719	0.884	8
	0.5	0.023	0.438	0.5	0.467	0.949	6
	0.615	0.008	0.727	0.615	0.667	0.862	3
	0.75	0.018	0.682	0.75	0.714	0.958	1
	0.161	0.087	0.231	0.161	0.189	0.719	7
	0.125	0.01	0.2	0.125	0.154	0.751	2
	0.281	0.071	0.257	0.281	0.269	0.762	4
	1	0	1	1	1	1	5
Weighted Avg.	0.69	0.072	0.672	0.69	0.679	0.892	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekilde 4.7’de görüldüğü gibi FP oranı sıfırdan farklı ve TP oranının birden farklı olmasından dolayı bu sınıflandırma modeli sonucunda başarılı bir sonuç elde edilemedi.

4.1.5 Jrip Modelleme

Jrip sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 268 ı (yüzde 67) doğru sınıflandırılmışken, 132 veri (yüzde 33) yanlış olarak sınıflandırılmıştır. Root

mean squared error yüzde 0.2461 olmasından dolayı başarılı bir sınıflandırma modeli olmadığı görülmektedir.

Şekil 4.8 : Jrip istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Bu sınıflandırma modelinde training set kullanılmıştır. Croos – validation test ile training set arasında bir fark olmadığı görülmüştür.

Şekil 4.9 : Jrip doğruluk değerleri

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.889	0.277	0.596	0.889	0.713	0.816	8
	0.429	0.021	0.429	0.429	0.429	0.974	6
	0.462	0.018	0.462	0.462	0.462	0.793	3
	0.55	0.026	0.524	0.55	0.537	0.964	1
	0	0.015	0	0	0	0.651	7
	0.375	0.003	0.75	0.375	0.5	0.871	2
	0	0.003	0	0	0	0.738	4
	0.992	0.089	0.844	0.992	0.912	0.968	5
Weighted Avg.	0.67	0.122	0.535	0.67	0.59	0.85	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.9’da görüldüğü gibi FP oranı sıfırdan farklı ve TP oranının birden farklı olmasından dolayı bu sınıflandırma modeli sonucunda başarılı bir sonuç elde edilemedi.

4.1.6 Part Modeli

Part sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 271 i (yüzde 67.75) doğru sınıflandırılmışken, 129 veri (yüzde 32.25) yanlış olarak sınıflandırılmıştır. Root

mean squared error yüzde 0.2387 olmasından dolayı başarılı bir sınıflandırma modeli olmadığı görülmektedir.

Şekil 4.10 : Part istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Bu sınıflandırma modelinde training set kullanılmıştır. Croos – validation test ile training set arasında bir fark olmadığı görülmüştür.

Şekil 4.11 : Part doğruluk değerleri

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.802	0.234	0.612	0.802	0.694	0.832	8
	0.643	0.026	0.474	0.643	0.545	0.913	6
	0.462	0.013	0.545	0.462	0.5	0.887	3
	0.35	0.021	0.467	0.35	0.4	0.974	1
	0.179	0.07	0.294	0.179	0.222	0.687	7
	0.25	0.015	0.25	0.25	0.25	0.904	2
	0.156	0.033	0.294	0.156	0.204	0.793	4
	1	0	1	1	1	1	5
Weighted Avg.	0.678	0.089	0.648	0.678	0.654	0.877	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.11’de görüldüğü gibi FP oranı sıfırdan farklı ve TP oranının birden farklı olmasından dolayı bu sınıflandırma modeli sonucunda başarılı bir sonuç elde edilemedi.

4.1.7 Oner Modeli

OneR sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 279 u (yüzde 69.75) doğru sınıflandırılmışken, 121 veri (yüzde 30.25) yanlış olarak

sınıflandırılmıştır. Root mean squared error yüzde 0.275 olmasından dolayı başarılı bir sınıflandırma modeli olmadığı görülmektedir.

Şekil 4.12 : Oner istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.13 : Oner doğruluk değerleri

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.343	0.573	1	0.728	0.828	8
	1	0.036	0.5	1	0.667	0.982	6
	0	0	0	0	0	0.5	3
	0.4	0.034	0.381	0.4	0.39	0.683	1
	0	0	0	0	0	0.5	7
	0	0	0	0	0	0.5	2
	0	0	0	0	0	0.5	4
	1	0	1	1	1	1	5
Weighted Avg.	0.698	0.111	0.544	0.698	0.6	0.793	

```
=== Confusion Matrix ===
```

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.13’de görüldüğü gibi FP oranı sıfırdan farklı ve TP oranının birden farklı olmasından dolayı bu sınıflandırma modeli sonucunda başarılı bir sonuç elde edilemedi.

4.1.8 ZeroR Modeli

ZeroR sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 131 i (yüzde 32.75) doğru sınıflandırılmışken, 269 veri (yüzde 67.25) yanlış olarak sınıflandırılmıştır. Kappa Statistic yüzde 0 olmasından dolayı bu sınıflandırma modeli çalışmamız için uygun olan bir sınıflandırma modeli değildir.

Şekil 4.14 : ZeroR istatistiksel değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Bu sınıflandırma modelinde training set kullanılmıştır. Cross – validation test ile training set arasında bir fark olmadığı görülmüştür. Bu sınıflandırma modeli diğer sınıflandırma modelleri arasındaki en kötü sınıflandırma modelidir.

Şekil 4.15 : ZeroR doğruluk değerleri

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0	0	0	0	0	0.486	8
	0	0	0	0	0	0.411	6
	0	0	0	0	0	0.417	3
	0	0	0	0	0	0.5	1
	0	0	0	0	0	0.475	7
	0	0	0	0	0	0.398	2
	0	0	0	0	0	0.473	4
	1	1	0.328	1	0.493	0.495	5
Weighted Avg.	0.328	0.328	0.107	0.328	0.162	0.48	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.15’de görüldüğü gibi eğri 0.48 dir. ZeroR sınıflandırma modelinde, FP(0,1) ve TP(0,1) değerlerindedir. Bundan dolayı ZeroR sınıflandırma modeli için çıkan bu değerler uygun değerler olmayıp en kötü sonucu vermiştir.

4.1.9 RBFNetwork Modeli

RBF Network sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 271 i (yüzde 67.75) doğru sınıflandırılmışken, 129 veri (yüzde 32.25) yanlış olarak sınıflandırılmıştır. RBF Network sınıflandırma modeli ile diğer sınıflandırma modelleri arasında bir farklılık bulunmamaktadır.

Şekil 4.16 : RBF Network istatistiksel değerleri

Classifier output			
pCluster_4_1	1.1787234535703536E23	1.3741105646344056E163	Infinity
pCluster_5_0	2.86107943675896E67	2.36251216857297E137	Infinity
pCluster_5_1	Infinity	0	Infinity
pCluster_6_0	0	3.6657027224679275E208	3.17728
pCluster_6_1	0	0	1.1375248267525437E277
pCluster_7_0	0	0	0
pCluster_7_1	0	0	Infinity

=== Stratified cross-validation ===			
=== Summary ===			
Correctly Classified Instances	271	67.75 %	
Incorrectly Classified Instances	129	32.25 %	
Kappa statistic	0.5609		
Mean absolute error	0.0955		
Root mean squared error	0.253		
Relative absolute error	49.9616 %		
Root relative squared error	81.9425 %		
Total Number of Instances	400		

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.16’da, RBF Ağı hakkında ayrıntılı bilgiler görünmektedir. RBF modelinde doğru sonucu vermesi test edilmesinde ise 400 solunum problemi yaşayan hasta veri kümesi üzerinde Cross- validation (çapraz doğrulama) testi uygulanmıştır. RBF modeli sonucunda farklı değerler elde edilmiştir.

Şekil 4.17 : RBF Network doğruluk değerleri

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.849	0.263	0.598	0.849	0.702	0.83	8
	0.286	0.018	0.364	0.286	0.32	0.639	6
	0.615	0.01	0.667	0.615	0.64	0.899	3
	0.85	0.016	0.739	0.85	0.791	0.861	1
	0.143	0.067	0.258	0.143	0.184	0.704	7
	0.125	0.013	0.167	0.125	0.143	0.477	2
	0	0.019	0	0	0	0.747	4
	0.962	0.019	0.962	0.962	0.962	0.984	5
Weighted Avg.	0.678	0.102	0.614	0.678	0.636	0.846	

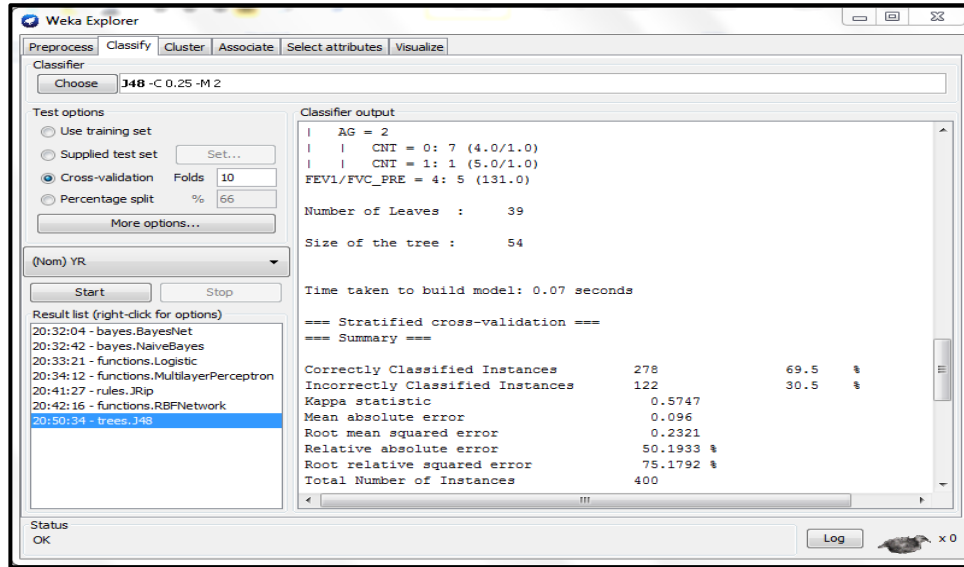
Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.17’de görüldüğü gibi ROC alanındaki değerlere bakıldığında başarılı bir sonuç elde edilemediği görülmektedir.

4.1.10 J48 Modeli

J48 sınıflandırma modelinde 400 solunum fonksiyonu verilerinden 278 i (yüzde 69.5) doğru sınıflandırılmışken, 122 veri (yüzde 30.5) yanlış olarak sınıflandırılmıştır. J48 sınıflandırma modeli ile diğer sınıflandırma modelleri arasında bir farklılık bulunmamaktadır.

Şekil 4.18 : J48 istatistiksel değerleri



Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.18’da, J48 hakkında ayrıntılı bilgiler görünmektedir. J48 modelinde doğru sonucu vermesi test edilmesinde ise 400 solunum problemi yaşayan hasta veri kümesi üzerinde Cross- validation (çapraz doğrulama) testi uygulanmıştır. J48 modeli sonucunda farklı değerler elde edilmiştir.

Şekil 4.19 : J48 doğruluk değerleri

=== Detailed Accuracy By Class ===							
	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.96	0.325	0.576	0.96	0.72	0.834	8
	0.5	0.013	0.583	0.5	0.538	0.848	6
	0.154	0.005	0.5	0.154	0.235	0.806	3
	0.6	0.026	0.545	0.6	0.571	0.98	1
	0.071	0.038	0.235	0.071	0.11	0.684	7
	0.125	0.008	0.25	0.125	0.167	0.67	2
	0	0	0	0	0	0.777	4
	1	0	1	1	1	1	5
Weighted Avg.	0.695	0.11	0.611	0.695	0.628	0.866	

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Şekil 4.19’da görüldüğü gibi ROC alanındaki değerlere bakıldığında başarılı bir sonuç elde edilemediği görülmektedir.

4.1.11 Sınıflandırma Modelinde Eğitim Veri Sonuçları

Diğer sınıflandırma metotları çalıştırılarak, güvenilirlikleri karşılaştırılmıştır. Modelin doğru sonucu vermesinin test edilmesinde, 400 solunum problemi şikayeti olan hasta veri kümesinde Cross- validation (çapraz doğrulama) testi yapılmıştır. Cross - validation testinde tek bir veri dosyamız 10 gruba bölünerek, 9 grubu öğrenim modeli geliştirmek için kullanılırken geriye kalan 1 grubu modeli test etmek için kullanıldı. Bu süreç 10 defa tekrarlanarak her defasında farklı 1 parça test parçası olarak alındı ve 9 parçada öğrenim modelini geliştirmek için kullanıldı. Bu Croos – validation analizi sonucunda hata oranının ortalama değeri ve oluşturulan modelin olası hata oranı tek bir değer olarak gösterildi.

Weka programı ile yapılan veri madenciliği 400 hasta verilerimiz için sırasıyla BayesNet, Naive Bayes, Logistic, Multilayer perceptron, Jrip, J48,Part, ZeroR ve Oner sınıflandırma modelleri seçilerek program çalıştırılmış ve elde edilen sonuçlara bakıldığında doğruluğu en yüksek olan sınıflandırma modeli BayesNet olarak görülmektedir. ROC alanı sonuçlarına göre bakıldığında da en iyi sonuç BayseNet modelidir.

Tablo 4.1 : Sınıflandırma yöntemlerinin karşılaştırma sonuçları

Bayes	Naive Bayes	Logistic	Multiplayer Perceptron	Jrip	RBF	J48	OneR	Part	ZeroR
-------	-------------	----------	------------------------	------	-----	-----	------	------	-------

CCI	283	282	271	276	268	271	278	279	271	131
ICI	117	118	129	124	132	129	122	121	129	269
KS	0.6033	0.598	0.5662	0.5898	0.5409	0.5609	0.5747	0.5766	0.5658	0
MAE	0.0959	0.098	0.0989	0.0868	0.1109	0.0955	0.096	0.0756	0.094	0.1912
RMSE	0.2307	0.2303	0.2606	0.2415	0.2461	0.253	0.2321	0.275	0.2387	0.3087
RAE	50.1429	51.2471	51.7294	45.4174	57.9905	49.9616	50.1933	39.5488	49.1752	100
RRSE	74.7238	74.6022	84.4016	78.2201	79.6989	81.9425	75.1792	89.0706	77.2998	100
TNI	400	400	400	400	400	400	400	400	400	400

Kaynak : Bu tablo Yılmaz Güler tarafından hazırlanmıştır.

Tablo 4.2 : Sınıflandırma doğruluk sonuçları

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Bayes	0.873	0.241	0.625	0.873	0.728	0.861
NaiveBayes	0.905	0.252	0.623	0.905	0.738	0.859
Logistic	0.873	0.241	0.625	0.873	0.728	0.848
Multiplayer Perceptron	0.762	0.164	0.681	0.762	0.719	0.884
Jrip	0.889	0.277	0.596	0.889	0.713	0.816
RBF	0.849	0.263	0.598	0.849	0.702	0.83
J48	0.96	0.325	0.576	0.96	0.72	0.834
Oner	1	0.343	0.573	1	0.728	0.828
Part	0.802	0.234	0.612	0.802	0.694	0.832
ZeroR	0	0	0	0	0	0.486

ROC alanındaki değer 1'e ne kadar yakınsa, o veri madenciliği modelinin en iyi kestirimi sağladığı varsayılır.

4.1.12 Sınıflandırma Metotlarının Özeti

Veri madenciliği sınıflandırma yöntemlerine bakıldığında en kötü sonucu veren sınıflandırma modeli ZeroR modelidir. Sınıflandırma modelinde elde etmiş olduğumuz en iyi sonuçlar BayesNet ve Naive Bayes sınıflandırma modelidir. Logistic, Multilayer perceptron, Jrip, Part ve Oner sınıflandırma modelleri BayesNet ve Naive Bayes sınıflandırma modelleri gibi iyi sonuçlar vermemiştir.

4.2.KÜMELEME YÖNTEMLERİ

Araştırmamızda yapmış olduğumuz kümeleme modelinde hasta verilerinde ortak özelliklerini dikkate alarak gruplara ayrılması için “Kümeleme Metotları” işlemi yapılmıştır. Kümeleme algoritmalarından “Simple K Mean” modeli kullanılmıştır.

Şekil 4.20 : Simple K mean değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Simple K Means kümeleme modeli sonucunda 400 hasta verileri bilgisini aldı ve 2 kümeye ayırdı. Birbirinden farklı bu 2 grupta, 0 kümesinde 248 (yüzde 62) hasta verisinin ve 1 kümesinde 152 (yüzde 38) hasta verisinin kendi aralarındaki benzerliklerini bularak kümelere ayırdı.

Kümeleme algoritmasında 0 kümesinde solunum fonksiyonu problemi yaşayan hasta gruplarının ortak özelliklerine bakıldığında, 60 yaş üzerinde oldukları, cinsiyeti erkek oldukları, 160 – 170 cm boy ölçüsü arasında oldukları, 50 – 60 kilo arasında oldukları, alerjik durumlarının oldukları, aile geçmişlerinde solunum fonksiyonu problemi yaşayan yakınlarının oldukları, meslek grubunda solunum fonksiyonu problemi yaşayacağı bir meslekte çalışmış oldukları, sigara kullanımlarının oldukları, FVC1/FV değerinin yüzde 40 tan küçük oldukları ve çok şiddetli obstrüksiyon¹ problemlerinin oldukları görülmektedir.

Diğer kümeleme algoritmasının 1 kümesinde solunum fonksiyonu problemi yaşayan hasta gruplarının ortak özelliklerine bakıldığında, 30 – 40 yaş arasında oldukları, cinsiyeti erkek oldukları, 170 – 180 cm boy ölçüsü arasında oldukları, 70 – 80 kilo arasında oldukları, alerjik durumlarının olmadığı, aile geçmişlerinde solunum fonksiyonu problemi yaşayan yakınlarının olmadığı, meslek grubunda solunum fonksiyonu problemi yaşayacağı bir meslekte çalışmış olmadığı, sigara kullanımlarının olmadığı, FVC1/FV değerinin yüzde 70 den büyük oldukları ve Normal Spirometri (solunum fonksiyonu probleminin olmadığı) oldukları görülmektedir.

¹ Tıkanma

4.3 BİRLİKTELİK YÖNTEMLERİ

Araştırmamızda yapmış olduğumuz birliktelik modelinde hasta verilerin kendi içlerindeki ortak özellikleri göz önüne alınarak, hangi olaylar sonucunda hangi olaylar ile birlikte gerçekleştirecekleri, olayın ortaya konulması için “Birliktelik Kuralı” işlemi yapılmıştır. Birliktelik kuralında “Apriori” modeli kullanılmıştır.

Şekil 4.21 : Apriori değerleri

Kaynak : Bu şekil Yılmaz Güler tarafından hazırlanmıştır.

Birliktelik kuralında Apriori modeli sonucunda 400 hasta verileri bilgisi olarak olaylar arasında ilişkilendirmede 10 gruba ayırmıştır.

Şekil 4.21’de ki bulguların ışığında hastanın yaptırmış olduğu testler sonucunda normal spriometri olması durumunda FEV1/FVC oranı yüzde 70 den büyük olabilir.

Şekil 4.21’in best rules found birinci maddesinde görüldüğü gibi hastanın normal spirometri olması durumunda ise FEV1 / FVC yüzde 70’den büyük olabilir.

Şekil 4.21’in best rules found ikinci maddesinde görüldüğü gibi hastanın FEV1 / FVC yüzde 70’den büyük olması durumunda ise normal spirometri olabilir.

Şekil 4.21’in best rules found üçüncü maddesinde görüldüğü gibi hastanın çok şiddetli obstriksuyon olması durumunda ise FEV1/FVC yüzde 40’dan küçük olabilir.

Şekil 4.21’in best rules found dördüncü maddesinde görüldüğü gibi hastanın şimdi veya geçmişte yapmış olduğu meslekte solunum problemine neden olacak bir meslekte çalışılmış ve çok şiddetli obstriksuyon olur ise FEV1/FVC oranı yüzde 40’dan küçük olabilir.

Şekil 4.21’in best rules found beşinci maddesinde görüldüğü gibi hastanın boy ölçüsü 160 – 170 cm arasında olması, hastanın alerjisinin olması ve hastanın sigara kullanımı var ise FEV1/FVC oranının yüzde 40’dan küçük olabilir.

Şekil 4.21’in best rules found altıncı maddesinde görüldüğü gibi hasta sigara kullanmıyor ve normal spirometri olması durumunda ise FEV1/FVC oranı yüzde 70’den büyük olabilir.

Şekil 4.21’in best rules found yedinci maddesinde görüldüğü gibi hasta sigara kullanmıyor ve FEV1/FVC oranı yüzde 70’den büyük ise normal spirometri olabilir.

Şekil 4.21’in best rules found sekizinci maddesinde görüldüğü gibi hastanın aile geçmişinde solunum fonksiyonu problemi varsa ve çok şiddetli obstriksuyon olması durumunda ise FEV1/FVC oranı yüzde 40’dan küçük olabilir.

Şekil 4.21'in best rules found dokuzuncu maddesinde görüldüğü gibi hastanın cinsiyeti erkek olması, boyunun 160 – 170 cm arasında olması, sigara kullanıyorsa ve FEV1/FVC oranı yüzde 40'dan küçük olması durumunda ise alerjik durumu olabilir.

Şekil 4.21'in best rules found onuncu maddesinde görüldüğü gibi hastanın cinsiyeti erkek olması, boyunun 160 – 170 cm arasında olması, sigara kullanıyorsa ve alerjik durumunun olması durumunda ise FEV1/FVC oranı yüzde 40'dan küçük ve çok şiddetli obstrüksiyon olabilir.

5. TARTIŞMA

Her geçen günle birlikte büyük oranla artan veri hacmi doğru sonuçlara ulaşılmasında ve bilgiye kolay ulaşılması bakımından zorlaşmıştır. Araştırmamızda, veri madenciliği yazılım programlarından Weka programı çözüm fonksiyonları testlerinde kullanılmıştır. Weka programında yer alan birden fazla metotlar sayesinde, veriden en doğru bilgiye ulaşılabilecek sonuçların çıkarılmasında kolaylık sağlamaktadır.

Araştırmamızda, Weka içerisindeki algoritmalar sonucunda sınıflandırma modülünde BayesNet ve kümeleme modülünde Simple K Means oranının başarılı olduğu tespit edilmiştir. Sınıflandırma modelinde hasta veri kümesinde Cross-validation (çapraz doğrulama) testi yapılmıştır. Cross-validation testinde tek bir veri dosyamız 10 gruba bölünerek, 9 grubu öğrenim modeli geliştirmek için kullanılırken geriye kalan 1 grubu modeli test etmek için kullanılmıştır. Bu süreç 10 defa tekrarlanarak her defasında farklı 1 parça test parçası olarak alındı ve 9 parçada öğrenim modelini geliştirmek için kullanılmıştır. En yüksek başarı oranının verildiği sonuç tablosuna bu değer yazılmıştır. Weka veri madenciliği programı ile diğer sınıflandırma metotlarında da Cross-validation kullanılmıştır.

Araştırmamızda, matematiksel yöntemle 400 hasta verisini sınıf bilgisine göre sınıflandırdı. Sınıflandırma sonuçları incelendiğinde BayesNet algoritmasının model testine ait 283 veri doğru sınıflandırılırken, 117 hasta verisinde yanlış olup yüzde 70.75'i doğruluk derecesiyle iyi bir sonucu ortaya koyduğu kabul edilebilir. Bu ölçüte göre BayesNet algoritmasını sırasıyla NaiveBayes, Logistic, Multiplayer Perceptron, Jrip, RBF, J48, Oner, Part ve ZeroR algoritmaları izlemektedir.

Araştırmamızda, sınıflandırma modelinde Bayes sınıfı kuralı için BayesNet algoritmasında 283 veri doğru sınıflandırılırken, kappa istatistiğinin 0.6033, MAE'nin 0.0959, RMSE'nin 0.2307 sonucuna ve Naive Bayes algoritmasında 282 veri doğru sınıflandırılırken, kappa istatistiğinin 0.598, MAE'nin 0.098, RMSE'nin 0.2303 sonucuyla Naive Bayes algoritmalarının en iyi algoritma oldukları görülmektedir.

Araştırmamızda, kümeleme yöntemi sınıfları bulunmayan verilerin gruplar halinde kümelere ayrılması için kullanılmıştır. Kümeleme yöntemlerini kullanarak verilerimiz var olan kümelerle karşılaştırıldı. Kendisine en yakın kümeye atanarak, bu kümeyi tanımlayan değeri değiştirdi. Simple K Means kümeleme metodu kullanılması veriler arasında birbirlerine benzerliklerine bakılarak birbirlerine olan uzaklıklar minimum dereceye indirildi ve kümeler arasındaki uzaklıkta maksimum dereceye getirildi. Simple K Means algoritmasında ki veriler, merkez veri noktasından kendine en yakın olana atanarak iki küme algoritması oluştu ve hastaların yaşlarına, boylarına ve kilolarına göre, benzerliği birbirine en yakın olan verilerin uzaklıkları azaltılarak gruplara ayrılması sağlandı.

Kümeleme yönteminde ortak özelliklerine bakıldığında, hastaların 60 yaş üzerinde oldukları, cinsiyeti erkek oldukları, 160 – 170 cm boy ölçüsü arasında oldukları, 50 – 60 kilo arasında oldukları, alerjik durumlarının oldukları, aile geçmişlerinde solunum fonksiyonu problemi yaşayan yakınlarının oldukları, meslek grubunda solunum fonksiyonu problemi yaşayacağı bir meslekte çalışmış oldukları, sigara kullanımlarının oldukları, FVC1/FV değerinin yüzde 40 tan küçük olduğu tespit edilmiştir.

Araştırmamızda, birliktelik yönteminin kullanılması medikal alandaki verilerin birbirleri arasındaki bağları inceleyerek, hangi olayların eş zamanlı olarak birlikte gerçekleşebileceklerini ortaya çıkarmak için kullanılmıştır. Solunum fonksiyonu testi verileri arasındaki ilişkilerin incelenmesi sonucunda mesleğe, sigara kullanıma ve alerjik durumlarının birlikte gelişmesiyle ortaya çıkan FEV1/FVC oranlarının neler olacağını göstermiştir. Hastanın demografik özelliklerine bakıldığında, cinsiyeti erkek olması, boyunun 160 – 170 cm arasında olması, sigara kullanıyorsa ve alerjik durumunun olması durumunda ise FEV1/FVC oranı yüzde 40'dan küçük ve çok şiddetli obstrüksiyon olabilir.

Araştırmamızda, belli bir sonuca ve algoritmaya bağımlı kalınmamıştır. Diğer metotların uygulanması ve test kümesinin eğitim kümesinin dışında seçilmesine imkan sağlayan diğer metotlarla da çalışılarak, yapılan araştırma sonuçlarının daha güvenilir olabileceği ön görülmektedir.

6. SONUÇ

Bu çalışmada, veri madenciliği yöntemleri, solunum yolu bozuklukları olanların bu duruma sebebiyet veren sorunları incelemek için kullanıldı. Solunum fonksiyonu problemlerinin birçok nedeni olabilir. Solunum fonksiyonu problemi yaşayan hastalarda yaş, cinsiyet, boy, kilo, alerjik durumu, aile geçmişi, mesleği ve sigara kullanımı gibi hastaların üzerindeki etkilere bakıldı.

Araştırmada, veri madenciliğinin algoritmaları kullanılarak karar verme mekanizmalarını geliştirmek ve farklı bakış açılarıyla solunum fonksiyonu testlerinin yorumlanması hedeflenmiştir.

Araştırmada, veri madenciliği teknikleri kullanılarak solunum fonksiyonları testi yaptıran hastaları, gruplandıran ve gelecekte hangi yaş grubu ve cinsiyetin solunum fonksiyonu riski altında olma ihtimalini ortaya koyan modeller geliştirilerek, solunum fonksiyonu problemlerini önlemek ve oluşabilecek solunum fonksiyonu risklerini azaltmak için yöntemler tespit çalışmaları yapıldı.

Verilerimiz üzerinde veri madenciliğinin bütün sınıflandırma metotlarını weka programı üzerinde çalıştırdı. Bu modellerden BayesNet sınıflandırma modeli kullanıldığında yüzde 70.75 değerinde bir orana ulaşıldı. Diğer sınıflandırma modelleri çalıştırıldığında ZeroR modeli ile yüzde 32.75 değeri çıkararak en kötü sonuca ulaşılmıştır.

Sınıflandırma modelleri incelendiğinde, Bayes sınıflandırma modelinde 283 verinin doğru sınıflandırıldığı, doğruluk yüzde 70.75 olduğu, MAE oranının 0.0959 olduğu, RMSE oranının 0.2307 olduğu ve KS oranının 0.6033 olduğu sonuçlarıyla en iyi algoritma sonucuna ulaşılmıştır.

Sınıflandırma modellerinin kesinlik ve duyarlılık ölçütlerini değerlendirmek için iki değerin F- Ölçütü incelendiğinde NaiveBayes sınıflandırma modelinde en iyi sonuca ulaşılmıştır.

Verilerimiz üzerinde veri madenciliğinin kümele metotlarını weka programı üzerinde çalıştırdı. Kümeleme metotlarından Simple K Means kümeleme sonucunda iki grubu ayrılan verilerimiz sonucunda sigara kullanımının, aile geçmişinde solunum problemi yaşayanların, alerjik durumu olanların, mesleki olarak solunum problemine yol açacak olan işlerde çalışanların ve 60 yaş üstünde olan kişilerin çok şiddetli Obstrüksiyon olma sonucuna ulaşılmıştır.

Verilerimiz üzerinde veri madenciliğinin birliktelik metotlarını weka programı üzerinde çalıştırdı. Birliktelik yönteminin Apriori modeli sonucunda yaş, cinsiyet, boy, kilo, alerjik durumu, aile geçmişi, mesleği ve sigara kullanımı bilgileri arasındaki olaylarla ilgili ilişkilendirilmede, olabilecek 10 durumun sonucuna ulaşılmıştır.

Birliktelik yönteminin apriori modelinde, hastanın cinsiyetinin erkek olması, boyunun 160 – 170 cm arasında olması, sigara kullanıyorsa ve alerjik durumunun olması durumunda ise FEV1/FVC oranı yüzde 40'dan küçük ve çok şiddetli obstrüksiyon olma sonucuna ulaşılmıştır.

KAYNAKÇA

Kitap

Balcı, K., 1993. *Göğüs Hastalıkları*. Konya : Atlas Kitabevi

Karakaya, G., 2004. *Astım ve Eşlik Eden Durumlar*. Ankara : Bilimsel Tıp
Yayınevi.

Petsonk, E., & Parker, J., 2008. *Coal workers lung diseases and silicosis*. 4. Baskı
Fishman's Pulmonary Diseases and Disorders,

Şeker, E., 2013. *İş zekası ve veri madenciliği Weka ile*. İstanbul : Cinius Yayınevi.

Witten, I., Frank, E., & Hall, M., 2011. *Data Mining*. 3. Baskı. USA : Morgan
Kaufmann Publishers.

Sürelî Yayınlar

- Akkaya, A., & Ünlü. M., 1995. Sigara kullanımının solunum fonksiyon testlerine etkisinin araştırılması. *SDÜ tıp fakültesi dergisi*. **2** (3), ss. 33-36
- American Thoracic Society, 1995. Standardization of spirometry. *Am J Respir Crit Care Med*. **152**. ss. 1107-1136
- Bateman, E., Hurd, S., Barnes, P., Bousquet, J., Drazen, J., FitzGerald, M., Gibson, P., Ohta, K., O'Byrne, P., Pedersen, S., Pizzichini. E., Sullivan, S., Wenzel S., Zar, H., 2008. Global strategy for asthma management and prevention . *PubMed* . **31** (1). ss. 143-178
- Çımrın, A., 2000. Meslek astımı. *Türk Toraks Dergisi*. **1** (1), s. 87.
- Demir, T., 2010. Meslek hastalıklarının tanısında solunum fonksiyon testleri. *Klinik gelişim İstanbul tabip odası süreli bilimsel yayını*. **23** (4). ss. 20-21
- Doğan, Ş., ve Türkoğlu, İ., 2007. Hypothyroidi and Hyperthyroidi Detection from Thyroid Hormone Parameters by Using Decision Trees. *Doğu Anadolu Bölgesi Araştırmaları Dergisi*. **5** (2), ss. 163-169
- Hand, D.J., 1998. Data Mining. *The American Statistician*. **52**, pp. 112-118.
- Jacobs, P., 1999. Data Mining. *Harvard Management Update*. **4** (10), s. 8.
- Kitler, R. ve Wang, W., 1998. The Emerging Role of Data Mining. *Solid State Technology*. **42** (11), s. 45.
- Orman, A., Ünlü, M., Cirit, M., & Yücel, A., 2002. Mermer işçilerinde solunum fonksiyon testleri. *Toraks dergisi*. **3** (2). ss. 188-193
- Özekes, S., 2003. Data Mining Models and Application Areas. *İstanbul Commerce University Journal of Science*. **3** (3), ss.65-82
- Shearer, C., 2000. The New Blueprint for Data Mining. *Journal of Data Warehousing*. **5** (4), ss. 13-23

Savaş, S., Topalođlu, N. & Yılmaz, M., 2012. Veri Madenciliđi Ve Türkiye'deki Uygulama Örnekleri. *İstanbul Ticaret Üniversitesi Fen bilimleri Dergisi*. **11** (21), ss. 1-23.

Talay, F., Kurt, B., & Tuđ, T., 2008. Alerjik ve alerjik olmayan astımlı hastaların klinik özellikleri ve solunum fonksiyonlarının karşılaştırılması. *Düzce Tıp Fakültesi Dergisi*. **1** (3) . ss. 15-20

Türk Toraks Dergisi, 2009. Astım tanı ve tedavi rehberi. *Türk Toraks Dergisi*. **10** (10).

Diğer Kaynaklar

Akgöbek, Ö., & Çakır, F., *Veri Madenciliğinde Bir Uzman Sistem Tasarımı*. [online],

http://ab.org.tr/ab09/kitap/akgobek_cakir_AB09.pdf [accessed 15 April 2014].

Akkoca, Ö., 2004. *Solunum fonksiyonları testleri*. [online],

http://file.toraks.org.tr/TORAKSFD23NJKL4NJ4H3BG3JH/kisokulu3-ppt-pdf/Oznur_Akkoca.pdf [accessed 29 April 2014].

Albayrak, M., 2008. EEG Sinyallerindeki Epileptiform Aktivitenin Veri

Madenciliği Süreci ile Tespiti. *Doktora Tezi*. Sakarya : Sakarya Üniversitesi FBE.

Alpay, A., 2008. *Meslek hastalıkları*. [online],

<http://isguvenligiuzmani.org/2008/11/29/meslek-hastaliklari/> [accessed 25 April 2014].

Attfield, M., & Hodous, T., 1992. Pulmonary function of us coal miners related

to dust exposure estimates. *American Review of Respiratory Disease*. USA : New York

Bartu, S., 2013. *Solunum fonksiyon testlerinin tarihçesi*. [online],

<http://www.toraks.org.tr/uploadFiles/book/file/512201294716-KITAP.pdf> [accessed 19 May 2014].

Çoşkun, F., . [online], <http://file.lookus.net/TGHYK/tghyk.22.pdf> [accessed 10 May 2014].

Dener, M., Dörterler M., & Orman, A., 2009. *Açık kaynak kodlu veri madenciliği*

programları: weka'da örnek uygulama. [online],

<http://ab.org.tr/ab09/bildiri/42.pdf> [accessed 10 May 2014].

Guler, N., Kirerleri, E., Ones U., Tamay, Z., Salmayenli, N., & Darendeliler, F.,

2004. *Does it have any role in childhood asthma?*. [online], [http://ac.els-](http://ac.els-cdn.com/S0091674904013260/1-s2.0-S0091674904013260-main.pdf?_tid=85cd288a-238c-11e4-a0d5-00000aab0f02&acdnat=1408004897_a3a28e9c5a83b532d73c9cc21e05f8db)

[cdn.com/S0091674904013260/1-s2.0-S0091674904013260-](http://ac.els-cdn.com/S0091674904013260/1-s2.0-S0091674904013260-main.pdf?_tid=85cd288a-238c-11e4-a0d5-00000aab0f02&acdnat=1408004897_a3a28e9c5a83b532d73c9cc21e05f8db)

[main.pdf?_tid=85cd288a-238c-11e4-a0d5-](http://ac.els-cdn.com/S0091674904013260/1-s2.0-S0091674904013260-main.pdf?_tid=85cd288a-238c-11e4-a0d5-00000aab0f02&acdnat=1408004897_a3a28e9c5a83b532d73c9cc21e05f8db)

[00000aab0f02&acdnat=1408004897_a3a28e9c5a83b532d73c9cc21e05f8db](http://ac.els-cdn.com/S0091674904013260/1-s2.0-S0091674904013260-main.pdf?_tid=85cd288a-238c-11e4-a0d5-00000aab0f02&acdnat=1408004897_a3a28e9c5a83b532d73c9cc21e05f8db)

[accessed 1 June 2014].

Global initiative for asthma, 2006. Global strategy for asthma management and prevention

İnan, O., (2003). Veri Madenciliği. *Yüksek Lisans Tezi*. Konya : Selçuk Üniversitesi FBE.

İÜKE İstanbul Üniversitesi Kardiyoloji Enstitüsü, 2014. *Solunum Fonksiyonları*

[online],

www.istanbul.edu.tr/enstituler/kardiyoloji/index.php?Sayfa=Solunum_Fonk

[accessed 25 May 2014]

Kalikov, A., (2006). Veri madenciliği ve bir e-ticaret uygulaması. *Yüksek Lisans Tezi*.

Ankara : Gazi Üniversitesi FBE.

Kibelstis, J., Morgan , E., Reger, R., Lapp N., & Morgan, W., 1973. Prevalence of bronchitis

and airway obstruction in American bituminous coal miners. *American review of respiratory disease*. USA : New York

Madan, I., 1997. Occupational asthma and other respiratory diseases. *BMJ publishing group*.

İngiltere : Londra

Neill, M., 1985. *Teknolojinin gelişimi*. [online],

<http://www.diyadinnet.com/YararliBilgiler-214&Bilgi=bilim-ve-teknoloji>

[accessed 5 June 2014].

Ober, C., 2005. *Perspectives on the past decade of asthma genetics*. [online],

[http://ac.els-cdn.com/S0091674905012819/1-s2.0-S0091674905012819-main.pdf?_tid=7c04ee4a-238d-11e4-88cd-](http://ac.els-cdn.com/S0091674905012819/1-s2.0-S0091674905012819-main.pdf?_tid=7c04ee4a-238d-11e4-88cd-00000aacb361&acdnat=1408005310_d62789ac8bd1ab3391b7436094ade7ce)

[00000aacb361&acdnat=1408005310_d62789ac8bd1ab3391b7436094ade7ce](http://ac.els-cdn.com/S0091674905012819/1-s2.0-S0091674905012819-main.pdf?_tid=7c04ee4a-238d-11e4-88cd-00000aacb361&acdnat=1408005310_d62789ac8bd1ab3391b7436094ade7ce)

[accessed 1 June 2014].

Öğüş, C., 2009. *Solunum fonksiyon testi parametrelerinin tanımlanması*. [online],

<http://www.toraks.org.tr/sunumerkezi/?s=5737353D5923225A29>

[accessed 1 May 2014].

Patterson, D., Turner, D., Concepcion, A., & Lynch, R. 2008.

Performance comparison of the data reduction system. *Proceedings of the SPIE symposium on defense and security*. USA : Orlando

Payalan, F., 2013. *Weka ile veri madenciliği*. [online],

<http://yfpayalan.blogspot.com.tr/2013/03/weka-ile-veri-madenciligi.html>
[accessed 10 May 2014].

Piramuthu, S., 1998. *Evaluating feature selection methods for learning in data mining applications*. [online],
<http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=648324> [accessed 5 May 2014].

Savaş, U., 2014. *Solunum Fonksiyonları Testi*. [online],
<http://www.ailem.com/templates/library/1759.asp?id=13605> [accessed 20 April 2014].

Snashall, D., 1997. Hazards of work. *BMJ publishing group*. İngiltere : Londra

Uçar, Ö., 2012. Solunum fonksiyon testinde yerleşik hava akımı obstrüksiyonu yapan nedenler. *Uzmanlık Tezi*. Adana : Çukurova Üniversitesi Tıp Fakültesi Göğüs Hastalıkları

Yıldırım, N., 2003. *Solunum Fonksiyonları Testi*. [online],

http://file.toraks.org.tr/TORAKSFD23NJKL4NJ4H3BG3JH/kisokulu2-ppt-pdf/Nurhayat_Yildirim.pdf [accessed 29 April 2014].