

## VERİ MADENCİLİĞİ MODELLERİNİN AKCİĞER KANSERİ VERİ SETİ ÜZERİNDE BAŞARILARININ İNCELENMESİ

Nihat Barış SEBİK \*, Halil İbrahim BÜLBÜL\*\*

\*Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği, Ankara, Türkiye

\*\* Gazi Üniversitesi, Gazi Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi, Ankara, Türkiye

### Özet

Günümüz bilgi çağında en güncel makine öğrenme yöntemlerinden birisi veri madenciliğidir. Her geçen gün bilgisayar sistemlerinin kapasitelerinin artıyor olması daha büyük miktarlarda verilerin saklanabilmesine imkân vermektedir. Artan veri miktarlarının etkin bir şekilde kullanılmasının en büyük çözümü veri madenciliğidir. Bu sebeple büyük miktardaki verileri işleyebilen teknikleri kullanabilmek günümüzde büyük önem kazanmaktadır. Veri madenciliği bu gibi durumlarda kullanılan, büyük miktardaki veri setlerinde saklı durumda bulunan örüntü ve eğilimleri keşfetme işlemidir. Veriler tek başına değersiz olabilirler ancak işlenip bilgiye dönüştürüldüğünde anlam kazanmaktadırlar. Bu kapsamda veri madenciliği birçok alanda olduğu gibi tıp alanında da kullanılmaktadır. Bu çalışmada Dünya Sağlık Örgütü'nün verilerine göre en yaygın kanser türlerinden olan akciğer kanseri teşhisinin daha hassas yapılabilmesine yönelik bir çalışma yapılmıştır. Çoğu kanser vakası hastalığın son evrelerinde teşhis edilebilmekte ve tedaviler bu evreden sonra çoğu zaman cevap verememekte ve hasta kayıpları yaşanmaktadır. Bu sebeple, diğer kanser türlerinde olduğu gibi, akciğer kanserinin de erken tanısı hayati önem taşımaktadır. Bu çalışmada, akciğer kanserinin erken tanısına katkıda bulunabilmek amaçlanmıştır. Genel olarak hastalara hastalık belirtileri doğrultusunda akciğer kanseri olup olmadıklarına dair teşhis konulmaktadır. Bu çalışma ile sağlık veritabanında mevcut olan, önceden teşhisi konulmuş vakaların anonim verileri kullanılarak, WEKA veri madenciliği yazılımında hangi algoritmanın bu alanda daha başarılı olabileceğine dair bir çalışma yapılmıştır.

**Anahtar Kelimeler :** Veri Madenciliği, Tıp Bilişimi, Akciğer Kanseri, WEKA

## INVESTIGATION OF SUCCESS ON LUNG CANCER DATA SET OF DATA MINING MODELS

### Abstract

One of the most modern machine learning methods in today's information age is data mining. The capacity of computer systems is increasing day by day, so it enables computers to store more data. Data mining is the greatest solution to the efficient use of increasing amounts of data. Therefore using techniques that can process huge data gain more importance today. Data mining is exploring hidden patterns and trends that are use in these kinds of data sets. The data alone may be worthless, but they make sense when processed and turned into information. In this context, data mining is used in medicine as well as in many areas. In this thesis, a study has been carried out on lung cancer, one of the most common cancer types according to the World Health Organization. Most of the cancer events are diagnosed at the late phases of the illness and treatments mostly does not work after this time, and patient deaths faced. Therefore, as in other cancer types, the early diagnosis of lung cancer is vital. In this paper, to contribute the early diagnosis of lung cancer is aimed. Patients are diagnosed by the help of their symptoms. With this study, anonymous data of pre-diagnosed cases available in the health database were used. With WEKA data mining software, a study has been made about which algorithm can be more successful in this area.

**Keywords:** Data Mining, Medical Informatics, Lung Cancer, WEKA

## 1.Giriş

Teknolojinin hızla gelişmesiyle birlikte hemen her alanda bilgisayarlar ile işlem yapılmaktadır. Bunun sonucunda her gün milyonlarca veri üretilmektedir. Dünya, veri zengini bilgi fakiridir (Han, Pei, & Kamber, 2012). Veri boyutlarının artması sebebiyle büyük verileri analiz edecek yöntemler ortaya çıkmıştır. Veri madenciliği de bu yöntemlerden en çok kullanılanlardan birisidir. Veri madenciliği yöntemleri kullanılarak mevcut sistemlerdeki gizli ve önemli bilgiler keşfedilebilmekte ve bu keşfedilen bilgiler değerlendirilebilmektedir (Kudyba, 2004). Veri madenciliği, veriyi farklı yönlerden analiz etme ve onu kullanışlı bilgi elde etmek üzere özetleme süreci olarak tanımlanabilir (Şentürk, 2011). Veri madenciliği konusu son yıllarda büyük önem taşımaya başlamıştır. Hemen her yerde kullanılması gerekli olan bir yöntem haline gelmiştir. Veri madenciliğinde amaç veritabanlarından veya veri setlerinden elde edilen bilgileri kullanarak bir çıkarım sağlamaktır. Bunlar için Naive Bayes, Genetik algoritmalar ve Karar Ağaçları gibi birçok yöntem bulunmaktadır. Büyük verilerin işlenip anlamlı bir sonuç elde edilme ihtiyacı çoğu özel sektör ve kamu sektörü için önemli bir ihtiyaç haline gelmiştir. Ancak şuan var olan milyarlarca verinin içinden istenilen fayda sağlayacak bilgiye ulaşmak en önemli sorundur ve bu sorun günümüzde akademik alanda da büyük önem kazanmış ve veri madenciliği kullanılarak bu konu ile ilgili birçok çalışma yapılmaya başlanmıştır (Siyah, 2012).

Bu çalışmada hekimlerin görüşleri alınarak öncelikle akciğer kanseri hastalığına ait veri seti elde edilmiş daha sonra bu veri seti ile modeller oluşturulmuştur. Bunun sonucunda oluşturulan modeller arasından en başarılı olan algoritma tespit edilmiştir.

## 2.Literatür Taraması

Veri madenciliği birçok alanda olduğu gibi özellikle tıp alanında da sıkça kullanılmaktadır. Sağlık ve tıp, çağın en önemli bilimsel araştırma alanları olduğu için bu alandaki bilgi sistemleri de araştırmalar için en büyük veri kaynaklarıdır (Yıldırım, Uludağ, & Görür, 2008). Veri madenciliğinin; sağlık hizmetlerinin sunumu, her düzeydeki sağlık kurumlarının yönetimi ve sağlık politikalarının oluşturulmasında bir karar destek aracı olarak kullanılması sağlık profesyonellerinin en uygun kararları almasına yardımcı olacaktır (Koyuncuğil & Özgülbaş, 2009). Yapılan bu çalışma doğrultusunda veri madenciliğinin sağlık alanında kullanımına yönelik bazı çalışmalar incelenmiştir.

2013 yılında sağlıklı insanların gelecekte meme, akciğer veya kolon kanserine yakalanma riskinin tahmini üzerine yapılan bir çalışmada bulanık mantık yöntemi kullanılmış olup meme kanserinde %80,83, akciğer kanserinde %80, kolon kanserinde %82,72 oranla başarı sağlanmıştır. Bu çalışma sayesinde insanların kansere yakalanmalarına karşı önlemler alma şansına sahip olabilecekleri ve kansere yakalanma riskinin azalabileceği belirtilmiştir (Yılmaz & Ayan, 2013).

2013 Yılında Walden Üniversitesi'nde yapılan bir diğer çalışmada ise Alaska'da petrol ve gaz faaliyeti olan yerlerde akciğer kanserinin bununla bir ilgisinin olup olmadığı veri madenciliği yöntemleriyle araştırılmıştır (Longan, 2013). Araştırma sonucunda bölgedeki petrol ve gaz faaliyetinin kansere neden olmadığı belirtilmiştir.

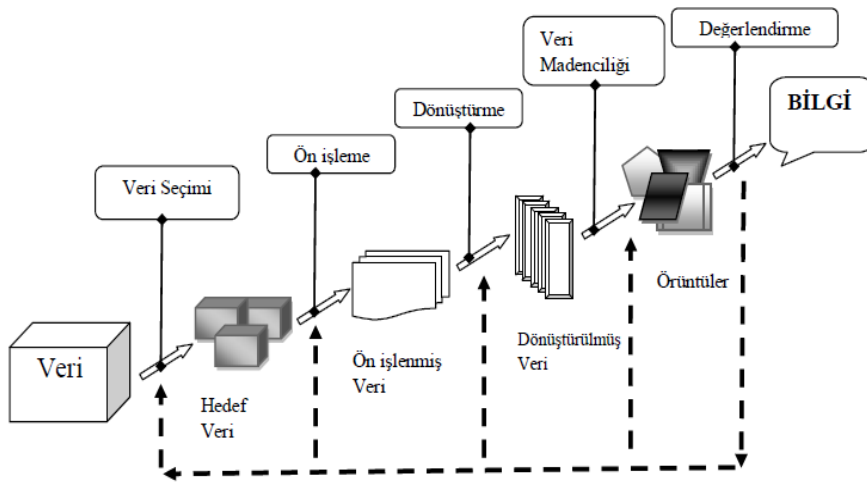
2016 yılında BT (Bilgisayarlı Tomografi) taraması kullanılarak akciğer kanserinin teşhisinde doğru karar verilebilmesi için, doğruluğu artırma aşamasında görüntüdeki gürültünün giderilmesine yönelik bir çalışma yapılmıştır. Bu çalışmada da veri madenciliği kullanılarak tümörün belirlenmesine yönelik başarı elde edilmiştir (Shakir, 2016).

2015 yılındaki bir diğer çalışma kapsamında ise küçük boyutlu akciğer nodüllerinin habislik derecesinin belirlenmesi için sınıflama yaklaşımları önerilmiştir. Çalışmada Amerikan Kanser Enstitüsü tarafından sunulan Lung Image Database Consortium (LIDC) veritabanı kullanılmıştır. LIDC veritabanı dört farklı enstitüden radyologlar tarafından değerlendirilmiş ve nodül habislik derecesi yanında radyografik tanımlayıcı bilgisini de içeren bir veritabanıdır. Yapılan istatistiksel anlamlılık testlerinde sınıflama başarımı en yüksek olan RF (Random Forest) tabanlı takım sınıflayıcının diğer yöntemlerden üstün olduğu gözlenmiştir (Kaya, 2015).

2013 yılında ise akciğer tomografileri kullanılarak yapay zeka ve görüntü işleme tekniklerine dayalı otomatik nodül bölge tespit yöntemi geliştirme çalışması yapılmıştır. Bu kapsamda akciğer BT kesit görüntüleri üzerinden nodül bölgelerinin otomatik tespiti ve tanınmasını sağlayan bir yöntem ve sistem geliştirilmiştir (Taşçı, 2013).

### 3.Yöntem

Akciğer kanseri, 20. yüzyılın başlarında nadir görülen bir hastalık iken, sigara içme alışkanlığındaki artışa paralel olarak sıklığı giderek artmış ve dünyada en sık görülen kanser türlerinden biri haline gelmiştir (Spiro & Porter, 2002). Tüm dünyada kanser olgularının %12,8'inden ve kanser ölümlerinin %17,8'inden akciğer kanseri sorumludur (Parkin, Bray, Ferlay, & Pisani, 2005). Sağlık Bakanlığı Kanserle Savaş Dairesinin 2009-2013 yılı kanser insidansı verileri incelendiğinde erkeklerde en sık görülen (59,3/100000 kişide), kadınlarda ise dördüncü en sık görülen kanser akciğer kanseridir (Türkiye Halk Sağlığı Kurumu, 2018). Çalışmada öncelikle akciğer kanseri alanında çalışan doktorların görüşleri alınmış ve bu alana uygun veri seti için ön araştırma yapılmıştır. Uygun veri setine veri madenciliği basamakları uygulanmıştır. Bu aşamada sırasıyla ön işleme ve veri temizleme, veri indirgeme, veri dönüştürme, veri madenciliği işlemleri yapılmış ve bu işlemlerden sonra ortaya bir değerlendirme ve sonuç çıkarılarak çalışma tamamlanmıştır.



Şekil 1. Bilgi keşfi süreci (Yılmaz, Topaloğlu, & Savaş, 2012)

Ön işleme ve veri temizleme aşaması veri madenciliğinin başarısı için oldukça önemli olup bu aşamanın başarısı sonucun başarısını doğrudan etkilemektedir. Veri üzerindeki bazı niteliklerin yanlış değer taşıdıkları veya eksik olma durumlarına gürültü adı verilir (Oreski, Oreski, & Oreski, 2012). Bu aşamada tutarsız ve gürültülü veriler tespit edilip bunlar için çeşitli işlemler uygulanmıştır. Çok sayıda null veri içeren satırlar veri setinden silinmiş, uç değerler budanmış, az sayıda eksik verisi olan satırlar için Replace Missing Values modülü kullanılarak boş değerlerin yerine ait oldukları özneliğin diğer değerlerinin ortalaması bulunarak boş veriler güncellenmiştir.

Veri indirgeme aşamasında; veri setinde hem doğum tarihi verisi hem de yaş verisi gibi gereksiz verilerin bulunması durumundan kaynaklı bilgisayarın çalışma zamanını olumsuz yönde etkileyecek ve çıkacak sonuçların kalitesini azaltacak, bu tür aynı sonucu ifade eden veriler veri setinden kaldırılmıştır.

Veri dönüştürme aşamasında; veri setinin standart bir yapı kazanması amacıyla karşılaşılan tutarsız veri türleri düzeltilmiş, yanlış girilmiş uç değerler normalleştirilmiştir. Daha sonra excel formatına aktarılan veri seti, veri madenciliği algoritmalarının kolaylıkla uygulanabilmesi için .csv formatına dönüştürülmüştür.

Veri setinde yapılan bu düzeltmelerden sonra uygun veri madenciliği algoritması seçimi ve modelin en başarılı sonuç üretmesini sağlayan algoritma bulunması için çalışmalar yapılmıştır.

Veri seti üzerinde yapılan veri temizleme işlemlerinden sonra WEKA (Waikato Environment for Knowledge Analysis)'da çalışabilmek adına .csv formatına dönüştürülen veri seti Şekil 2'deki gibidir.

Şekil 2. Akciğer Kanseri Veri Seti

	A	B	C	D	E	F	G	H	I	J					
1	ADRES_IL	ADRES_ILCE	CINSIYET	yas	UYRUK	c34_durum	MPV	RBC	MCHC	WBC	HCT	MCH	PCT	HGB	PDW
2	ANKARA	CANKAYA	1,64	TR	C34,8.02	4.16	32.99	16.4	35.1	26.51	0.18	12.3	20.5		
3	ANKARA	CANKAYA	1,60	TR	C34,7.4	4.28	31.5	17,30	27.7	0.255	8.8	16.8			
4	ANKARA	CANKAYA	1,62	TR	C34,7.5	5.07	33.1	4.7	45.5	29.7	0.15	15.1	10.5		
5	ANKARA	CANKAYA	1,65	TR	C34,8.4	4.84	34,5	9,40	7,28	6,0	183	13.8	16.7		
6	ANKARA	CANKAYA	1,72	TR	C34,8.6	2.56	33.6	3.3	26.3	34.6	0.09	8.8	50.8		
7	ANKARA	CANKAYA	2,51	TR	C34,8.2	4.12	33.3	8.22	36.5	29.5	0.17	12.2	46.7		
8	ANKARA	CANKAYA	1,73	TR	C34,7.2	4.43	29.6	7.67	31.7	21.2	0.17	9.4	51.1		
9	ANKARA	CANKAYA	1,54	TR	C34,6.6	3.56	33.8	4.7	35.9	34.0	0.15	12.1	9.2		
10	ANKARA	CANKAYA	1,61	TR	C34,6.7	4.18	33.6	8.22	39.3	31.6	0.18	13.2	9.3		

Veri seti akciğer kanseri teşhisi konmuş ve akciğer kanser teşhisi konmamış hasta ve hasta adaylarının anonim verilerinden oluşmaktadır. Bu veriler önışleme sürecinden geçirildikten sonra temiz veriye sahip olan toplam 404 adet veri ve Tablo 1’de gösterilen, sonuca doğrudan etki eden 9 adet öznelik verisi ile çalışma gerçekleştirilmiştir.

Tablo 1. Akciğer Kanseri Veri Setinde Bulunan Öznelikler

Öznelikler	Açıklama	Referans Aralığı
PDW (Platelet Distribution Width)	Trombosit Dağılım Genişliği	9-14
RBC (Red Blood Cells)	Kırmızı Kan Hücreleri	3,2-6
WBC (White Blood Cells)	Beyaz Kan Hücreleri	3-12
MCHC (Mean Corpuscular Hemoglobin Concentration)	Ortalama Yassı Kütle Hemoglobin Konsantrasyonu	30-36
MPV (Mean Platelet Volume)	Ortalama Trombosit Hacmi	0-15
HCT (Hematocrit)	Hematokrit	30-55
HGB (Hemoglobin)	Hemoglobin	10-18
MCH (Mean Corpuscular Hemoglobin)	Ortalama Hemoglobin	25-33
PCT (Procalcitonin)	Prokalsitonin	0,19-0,36

WEKA’da bulunan veri madenciliği algoritmaları arasından karşılaştırılacak algoritmalar seçilirken popülerlik ve literatürde benzer konuda yapılan çalışmalar dikkate alındıktan sonra on adet algoritma seçilmiş ve veri setine uygulanmıştır. Bu algoritmalar Naive Bayes, BayesNet, Lojistik Regresyon, Multilayer Perceptron, KStar, Bagging, OneR, ZeroR, J48 ve Random Tree algoritması şeklindedir.

Naive Bayes ve BayesNet sınıflandırıcı istatistiksel bir sınıflandırma olup verilerin belirli bir sınıfa ait olma olasılıklarını tahmin eder (Pala, 2013). Literatürde basit Bayesian sınıflandırıcı olarak geçse de tam tersine çok etkili bir sınıflandırıcıdır (Fan, Poh, & Zhou, 2010).

Lojistik regresyon çok güçlü bir modelleme aracıdır (Chun, Jafri, & San, 2011). Lojistik Regresyon analizinde verilerin yapısındaki grup sayısı bilinmekte ve bu verilerden faydalanarak bir ayırmsama modeli elde edilmektedir. Kurulan bu model yardımı ile veri kümesine yeni alınan gözlemlerin gruplara ataması yapılmaktadır. Lojistik Regresyon analizinin kullanım amacı, istatistikte kullanılan diğer model yapılandırma teknikleri ile aynıdır. En az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı ile bağımsız değişkenler arasındaki ilişkiyi tanımlayabilen ve biyolojik olarak kabul edilebilir bir model kurmaktır (Bircan, 2004).

Multilayer Perceptron, sinir ağları giriş nöronlarının bulunduğu giriş katmanı, çıkış nöronlarının bulunduğu çıkış katmanı ve bir veya daha fazla gizli katmandan meydana gelmektedir (Güler & Übeyli, 2006). Giriş katmanı çok katmanlı ağa gelen girişleri alır ve ara katmana iletir, bu katmandaki işlem elemanları sonraki ara katmandaki tüm işlem elemanlarına bağlanır ve bu şekilde algoritma çalışır (Dodurgalı, 2010). Genellikle doğrusal olmayan problemlerin çözümlenmesinde kullanılan Multilayer Perceptron algoritması çalışmada kullanılan veri setine uygulanmış ve bir model oluşturulmuştur.

Literatürde yaygın olarak geçen, örnek tabanlı öğrenme algoritması olan ve WEKA içerisinde de yer alan bir algoritma olan KStar algoritması model oluşturma işlemlerinde sıkça kullanılmaktadır. (Çölkesen & Kavzoğlu, 2010).

Bagging algoritması ilk defa Breiman tarafından önerilmiştir (Breiman, 1996). Bagging’te N adet örnekten oluşan eğitim setinden yine N adet örneklilik bir eğitim seti, yerine koymalı rastgele seçimle üretilmektedir. Bu durumda bazı eğitim örnekleri yeni eğitim kümesinde yer almazken (yaklaşık %33) bazıları birden fazla kez yer alırlar. Toplulukta her bir temel öğrenici bu şekilde üretilmiş birbirinden farklı örnekler içeren eğitim kümeleriyle eğitilirler ve sonuçları çoğunluk oylaması ile birleştirilir (Amasyalı & Ersoy, 2011).

OneR algoritması (tek sınıf sınırlandırma), bir özelliği test ederek kural ağaçları oluşturan WEKA’da rules sekmesinin alt başlıklarında bulunan algoritmalarından biri olup literatürde adı sıkça geçmekte olan bir algoritmadır. Problemin çözüm aşamasında model oluşturma ve bunları kıyaslama kısmında da bu sebeple tercih edilmiştir.

ZeroR algoritması sayısal veya nominal test verilerininin ortalama değerini tahmin eder. Temel kapsama algoritmaları kurallarını uygular (Witten, Frank, Hall, & Pal, 2016). ZeroR temelde sadece çoğunluk sınıf dağılımını tespit etmeye çalışır ve bu dağılım bir kural olarak varsayılır. ZeroR’da sayısal sınıf öznitelikleri için aritmetik ortalama, nominal sınıf öznitelikleri içinse mod sınıfın değeri tahmin edilir. ZeroR bunun dışında herhangi bir kural üretmemektedir (Özhan, 2013).

J48 algoritması ID3 ve C4.5 algoritmalarına dayanan bir karar ağacı algoritmasıdır. Veri madenciliği problemlerinde sıkça kullanılmaktadır.

Random Forest algoritması denetimli sınıflandırma algoritmasıdır. Ormandaki ağaç sayısı ve elde edilebileceği sonuçlar arasında doğru orantılı bir ilişki vardır. Ağaç sayısı arttıkça sonucun doğruluğu artmaktadır (Medium, 2018). Literatürde sıkça kullanılan bir algoritma olması sebebiyle bu çalışmada yer verilmiştir.

Veri setine uygulanan bu algoritmaların sonucunda elde edilen başarı oranları Tablo 2’de gösterilmektedir.

Tablo 2. Çalışmada kullanılan veri madenciliği algoritmalarının karşılaştırılması

Algoritma \ Ölçüt	Naive Bayes	Bayes Net	Lojistik Reg.	Multilayer perceptron	KStar	Bagging	OneR	ZeroR	J48	Random Tree
Doğruluk	%91,1	%89,9	%88,9	%90,1	%89,9	%90,8	%86,9	%80	%88,1	%89,4
Kesinlik	%91	%90	%88,4	%90,2	%92,2	%90,6	%86,2	%63,9	%88,4	%89,9
Duyarlılık	%91,1	%88,9	%88,9	%90,1	%89,9	%90,8	%86,9	%80	%88,1	%89,4
F-Ölçütü	%91	%89,9	%88,5	%90,1	%90,4	%90,6	%86,4	%71	%88,2	%89,6

#### 4.Sonuç

Bu çalışmada akciğer kanseri hastalığının teşhisinde fikir sunabilecek veri seti elde edilmiş ve bu veri setine WEKA veri madenciliği yazılımı ile çeşitli algoritmalar uygulanmıştır. Bu kapsamda personelin sisteme yanlış veya uç değer olarak girdiği verilerin teker teker kontrol edilip tek bir standart haline dönüştürülmesinden ve bütün ön işleme süreçlerinin tamamlanmasından sonra açık kaynak kodlu veri madenciliği yazılımı olan WEKA ile veri seti üzerinde çeşitli algoritmalar uygulanarak modeller oluşturulmuş ve buna göre en başarılı algoritma olarak Naive Bayes algoritması bulunmuştur.

Oluşturulan modellerin doğruluk oranlarına bakıldığında %91,1 ile Naive Bayes algoritmasının en yakın takipçisi Bagging algoritması %90,8 ile ikinci en iyi sonucu çıkarmıştır. Çalışmada ZeroR algoritması ise %80 doğruluk oranı ile en düşük başarı oranına sahip algoritma olarak görülmektedir.

Model oluşturmada kullanılan algoritmaların karşılaştırılması sonucunda en başarılı olan Naive Bayes algoritması ve bu veri seti kullanılarak ilerleyen çalışmalarda akciğer kanseri teşhisine yönelik bir uygulama geliştirilmesi ve hastalık teşhis sürelerinin kısalması, sağlık çalışanlarına fikir vermesi ve yaşam sürelerinin uzaması mümkün olabilir.

#### 5.Kaynaklar

- Amasyalı, M. F., & Ersoy, O. (2011). Performance Based Pruning and Weighted Voting with Classification Ensembles. *Signal Processing and Communication Applications (SIU), 2011 IEEE 19th Conference*, 194-197.
- Bircan, H. (2004). Lojistik Regresyon Analizi: Tıp Verileri Üzerine Bir Uygulama. *Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 185-208.
- Breiman, L. (1996). Bagging Predictors. *Machine Learning*, 24(2), 123-140.
- Chun, B. B., Jafri, M. Z., & San, L. H. (2011). Mangrove Mapping in Penang Island by Using Artificial Neural Network Technique. In *Open Systems (ICOS), 2011 IEEE Conference*, 245-249.
- Çölkesen, İ., & Kavzoğlu, T. (2010). Örnek Tabanlı K-Star Algoritması ile Uzaktan Algılanmış Görüntülerin Sınıflandırılması.
- Dodurgalı, H. H. (2010). Karınca Kolonisi Optimizasyonu ile Eğitilmiş Çok Katmanlı Yapay Sinir Ağı ile Sınıflandırma. *İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Yüksek Lisans Tezi*.
- Fan, L., Poh, K. L., & Zhou, P. (2010). Partition-conditional ICA for Bayesian classification of microarray data. *Expert Systems with Applications*, 37(12), 8188-8192.
- Güler, İ., & Übeyli, E. D. (2006). Çok Katmanlı Perceptron Sinir Ağları ile Diyabet Hastalığının Teşhisi. *Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi*, Vol 21, No 2, 319-326.



- Han, J., Pei, J., & Kamber, M. (2012). Data Mining Concepts and Techniques. Elsevier.
- Kaya, A. (2015). Bilgisayar Destekli Tanı Sistemi ile Akciğer Nodüllerinin Nitelendirilmesi. Hacettepe Üniversitesi Doktora Tezi.
- Koyuncugil, A. S., & Özgülbaş, N. (2009). Veri madenciliği: tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. Bilişim Teknolojileri Dergisi, 2:21-32.
- Kudyba, S. (2004). Managing Data Mining. CyberTech Publishing.
- Longan, S. (2013). Relationship of North Slope Oil and Gas Development to Lung Cancer Mortality. Walden University Doktora Tezi.
- Medium. (2018, 08 20). Medium. Medium: <https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning-3c0fe15b6674> adresinden alındı
- Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid System With Genetic Algorithm And Artificial Neural Networks and Its Application To Retail Credit Risk Assessment. Expert Systems with Applications 39, 12605–12617.
- Özhan, E. (2013). Güvenlik Duvarı Günlüklerinin Makine Öğrenmesi Yöntemleri ile Analizi ve Bir Model Çıkarılması. Trakya Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Doktora Tezi.
- Pala, T. (2013). Tıbbi Karar Destek Sisteminin Veri Madenciliği Yöntemleriyle Gerçekleştirilmesi. Marmara Üniversitesi Yüksek Lisans Tezi.
- Parkin, D., Bray, F., Ferlay, J., & Pisani, P. (2005). Global cancer statistics. A Cancer Journal for Clinicians, 55:74-108.
- Shakir, S. A. (2016). Early Detection of Lung Cancer. Yıldırım Beyazıt Üniversitesi Yüksek Lisans Tezi.
- Siyah, B. (2012). Oracle Data Miner İle Göğüs Kanseri Verileri Üzerine Bir Veri Madenciliği Uygulaması. İskenderun.
- Spiro, S., & Porter, J. (2002). Lung cancer-Where are we today? Current advances in staging and nonsurgical treatment. American Journal of Respiratory and Critical Care Medicine, 166:1166-1196.
- Şentürk, Z. K. (2011). Veri Madenciliği ile kanser tanısı. Düzce Üniversitesi Yüksek Lisans Tezi.
- Taşçı, A. E. (2013). Akciğer Tomografileri Kullanılarak Yapay Zeka ve Görüntü İşleme Tekniklerinde Dayalı Otomatik Nodül Bölge Tepit Yöntemi Geliştirilmesi.
- Türkiye Halk Sağlığı Kurumu. (2018, 7 18). Türkiye Halk Sağlığı Kurumu Türkiye Kanser İstatistikleri. Sağlık Bakanlığı: <https://dosyasb.saglik.gov.tr/Eklenti/13183,sy2016turkcepdf.pdf?0> adresinden alındı
- Witten, I. E., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine Learning tools and techniques. Morgan Kaufmann.
- Yıldırım, P., Uludağ, M., & Görür, A. (2008). Hastane Bilgi Sistemlerinde Veri Madenciliği. Akademik Bilişim 2008 Konferansı. Çanakkale.
- Yılmaz, A., & Ayan, K. (2013). Cancer risk analysis by fuzzy logic approach and performance status of the model. Turkish Journal of Electrical Engineering & Computer Sciences, 910 doi:10.3906/elk-1108-22.
- Yılmaz, M., Topaloğlu, N., & Savaş, S. (2012). Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri. İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Sayı: 21, 1-23.