Dokuz Eylül Üniversitesi-Mühendislik Fakültesi
Fen ve Mühendislik Dergisi
Cilt 19, Sayı 57, Eylül 2017

Dokuz Eylul University-Faculty of Engineering
Journal of Science and Engineering
Volume 19, Issue 57, September 2017

DOI: 10.21205/deufmd.2017195785

Akciğer Kanseri Verilerinin Karar Destek Sistemleri için Veri Ambarında Saklanması

Canan Eren ATAY[1], Ezgi DEMİR*[2]

[1]Dokuz Eylül Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü,
[2]Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü,
35370, İzmir

(Alınış / Received: 23.03.2017, Kabul / Accepted: 09.06.2017,
Online Yayınlanma / Published Online: 20.09.2017)

Anahtar Kelimeler
Veri Ambarı,
Akciğer Kanseri Verileri,
Veri Madenciliği,
Karar Destek Sistemleri

Özet: Veri ambarı ve veri madenciliğinin birlikte kullanımı sağlık alanına yeni bir soluk getirmiştir. Sağlık sektörü bilgiye dönüşmeyi ve değerlendirilmeyi bekleyen çok fazla miktarda veri barındırmaktadır. Bu çalışmada Amerikan Ulusal Kanser Enstitüsü'nden (NCI)[1] alınan akciğer kanseri verileri hatalardan ve tekrarlardan arındırılarak veri ambarı modelinde tutulmuştur. Modelleme yaklaşımı olarak yıldız şema modeli kullanılmış, böylece veri ambarı en karmaşık sorgulara bile yanıt verebilecek şekilde tasarlanmıştır. Çalışmanın sonraki aşamasında kurulan bu veri ambarı kullanılarak veri madenciliği teknikleri ile doktorlara karar verme aşamalarında destek sağlayacak bir örnek sistem geliştirilecektir. Kanser verilerinin büyüklüğü ve karmaşıklığı düşünüldüğünde veri ambarı ve veri madenciliğinin birlikte kullanılmasıyla kurulan bir karar destek sistemi, doktorların tedavilerini şekillendirmesinde onlara ışık tutabilecektir.

Storage of Lung Cancer Data in Data Warehouse for Decision Support Systems

Keywords
Data warehouse,
Lung Cancer Data,
Data Mining,
Decision Support Systems

Abstract: The usage of data warehouse and data mining together has brought a new breath to the healthcare field. The health sector has a massive amount of data waiting to be utilized and turned into information. In this study, we have created a data warehouse for lung cancer data that received from American National Cancer Institute [1]. The lung cancer data have brought into certain formats, cleaned from mistakes and repetitions. The star schema model was used as a modeling approach; hence the data warehouse is designed to respond to even the most complex queries. We plan to develop a prototype system to assist physicians by using data mining techniques in decision-making stages. When considering the size and complexity of cancer data, a decision-support system built with both data warehouse and data mining can shed light on the treatment of doctors.

*Sorumlu yazar: ezgi.demir@st.cs.deu.edu.tr

1. Giriş

Hastalıklara kısa sürede teşhis konulması, mevcut verilerin daha kolay yönetilmesi ve hızlı bir şekilde tedavi yolunun seçilip uygulanması gerekmektedir. Bu gereklilik tıp alanında bilişim sektöründen yararlanılmasını kaçınılmaz kılmıştır. Bilgi teknolojilerinden ve sağlık bilişim sistemlerinden faydalanılarak hastalık süreçleri için çok kıymetli olan zamanın kısaltılması, daha kaliteli hizmet verilmesi ve pozitif yönde sonuçlar alınması sağlanabilmektedir.

GLOBOCAN 2012 [2] verilerine göre 2012 yılında Dünya'da toplam 14,1 milyon yeni kanser vakası gelişmiş ve 8,2 milyon kansere bağlı ölüm olmuştur. En çok tanı konulan kanserler akciğer (%13,0), meme (%11,9) ve kolon (%9,7) iken kanserden ölümlerin ise en çok akciğer (%19,4), karaciğer (%9,1) ve mideden (%8,8) gerçekleştiği belirtilmiştir. Bu şekilde kanser artış hızının devam etmesi durumunda, Dünya nüfusunun artışına ve nüfustaki yaşlanmaya bağlı olarak 2025 yılında toplam 19,3 milyon yeni kanser vakası olacağı belirtilmiştir [3]. Eğer teknoloji daha etkin kullanılırsa bu hastalıkların tedavi süreci daha olumlu sonuçlanacak, hatta hastalıklar belki de tamamen önlenebilecektir. Bunun için sağlık kuruluşlarının, karar verme aşamasında verileri daha bilinçli şekilde kullanarak değerlendirmeye, verimliliği artırarak maliyetleri düşürmeye, kaynakların daha iyi yönetilmesini sağlamaya ihtiyaçları vardır.

Sağlık sektörü tüm bu amaçları gerçekleştirmek için çeşitli bölümlerden, laboratuarlardan ve ilgili idari süreçlerden gelen, bilgiye dönüşmeyi bekleyen çok fazla miktarda veri barındırır. Bu verilerin dağınıklığı, karmaşıklığı, heterojenliği, birbirinden ayrı yerlerde olması güvenilir şekilde bütünleştirilmelerini zorlaştırır. Bütünleşmemiş veriler, doktorların birbirlerinin kararlarından haberdar olmasını engellediği için tutarlı sonuçlar elde etmek çok zaman alıcı ve zahmetli bir iş haline gelir. Bu sorunun çözümü için hatalardan ve belirsizliklerden arındırılmış, belirli formatlara getirilmiş verilerin veri ambarı denilen yapılarda tek bir çatı altında toplanması gerekir. Veri ambarları farklı yerlerden toplanmış verileri kendi yapısında bulundurduğu gibi, bu veriler üzerinde sorgulama ve analiz yapılmasına olanak sağlar. Bununla beraber sağlık alanında bir veri ambarı oluşturmak diğer sektörlere oranla çok daha fazla çaba gerektiren bir iştir. Sağlık alanında veriler zamana bağlı olarak sürekli değişen bir yapıdadır. Bu verilerin daima güncel tutulması doktorların daha kesin ve tutarlı kararlar verebilmesi için son derece önemlidir. Tıp alanındaki röntgen, laboratuar sonuçları gibi verilerin çok karmaşık ve çeşitli tiplerde olması ise diğer bir sorundur. Bu verileri analiz çalışmaları için bütünleştirmek ayrı bir uğraş ve uzmanlık istemektedir.

Doktorların karar verme süreci incelendiğinde geçmiş bilgilerin ve deneyimlerin etkili olduğu görülmektedir Dolayısı ile deneyimsizlik, insani durumlar ve benzeri anlık veya kalıcı problemler nedeniyle kararlar gerektiği yönde verilemeyebilir. Bunun sonucunda da hata toleransı çok düşük olan tıp alanında ölüme kadar varabilen istenmeyen sonuçlar ile karşılaşılabilmektedir [4]. Veri ambarı ve veri madenciliği tekniklerinden yararlanarak yapılan karar destek sistemleri doktorların hastalarla ilgili en iyi kararı verebilmesi için destek sağlayan bilişim sistemi uygulamalarıdır. Bu sistemler en güncel bilgileri doktorların kullanımına sunarak karar vermelerine yardımcı olmak amacıyla tasarlanmaktadır. Bilginin çok çabuk eskidiği ve artan bilgi miktarına bağlı olarak doktorların karar vermesinin daha da güç bir hale geldiği durumlarda

karar almaya yardımcı olmaktadır [5]. Institute of Medicine tarafından 2000 yılında yapılan bir çalışmaya göre her yıl Amerika'da tıbbi hatalardan dolayı 98.000 hasta ölmektedir. Enstitünün amacı ise geliştirilen klinik karar destek sistemleri ile bu sayıyı minimize etmektir [6]. Klinik karar destek sistemleri kullanıldıkları klinik bölümlerde oluşan beklenmedik durumlarda erken uyarı, hızlı cevap, anında analiz, maliyet düşürme, doğru karar, etkin takım çalışması, zaman tasarrufu ve veri kaynaklarını iyi kullanabilme gibi etkin özellikleri ile sınanan alternatif sayısında artış sağlamaktadır [7].

Bu çalışmada çağımızın en ciddi hastalıklarından olan akciğer kanseri üzerinde durulmuştur. Amerika Ulusal Kanser Enstitüsü'nden (NCI-National Cancer Institute) alınan akciğer kanseri verileri belirli formatlara getirilerek, hatalardan ve tekrarlardan arındırılarak, oluşturulan veri ambarı yapısında tutulmuştur. Bu veriler hastalara ait demografik bilgi, teşhis, tedavi, komplikasyon, röntgen sonuçları, röntgenlerde saptanan anormallikler gibi değerlerden oluşmaktadır. Veri ambarı kullanımıyla büyük miktardaki bu veriler arasında karmaşık sorgulamaları bile yapmamıza olanak sağlayan bir yapı kurulmuştur. Çalışmanın ilerleyen aşamalarında veri madenciliği teknikleri kullanılarak veriler arasındaki değerli bilgilerin açığa çıkarılması ve akciğer kanseri hastalığının tedavisiyle uğraşan uzmanlara karar aşamasında fikir sağlayan bir karar destek sistemi geliştirilmesi amaçlanmaktadır. Ayrıca akciğer kanseri hakkında yapılan bilimsel araştırmalar için de kullanıma hazır hale getirilmiş bu verilerden daha birçok şekilde yararlanılabilir.

Makalenin geri kalanı şu şekilde düzenlenmiştir; 2. bölümde konuyla ilgili çalışmalar anlatılmış, 3. bölümde veri ambarının genel yapısı ve tıp alanında kullanımından söz edilmiş, 4. bölümde geliştirilen veri ambarında kullanılan yöntemler ve veri seti anlatılmış, son bölümde ise sonuç, değerlendirme ve önerilere yer verilmiştir.

2. Literatür Taraması

Dünya çapında medikal alanda çok sayıda başarılı veri ambarı ve veri madenciliği uygulama örnekleri bulunmaktadır. Örneğin, A. Kusiak ve arkadaşları tarafından akciğerdeki tümörün iyi huylu olup olmadığına dair, karar destek amaçlı bir çalışma yapılmıştır [8]. İstatistiklere göre Amerika'da 160.000'den fazla akciğer kanseri vakasının olduğu ve bunların %90'ının öldüğü belirlenmiştir. Bu bağlamda bu tümörün erken ve doğru olarak teşhisi önem kazanmaktadır. Noninvaziv, girişimsel olmayan, testler ile elde edilen bilgi sayesinde %40-60 oranında doğru teşhis konabilmektedir. İnsanlar kanser olup olmadıklarından emin olmak için biyopsi yaptırmayı tercih etmektedirler. Ancak bu testlerin hem maliyeti yüksek hem de çeşitli riskler taşımaktadır. Faklı yerlerde ve farklı zamanlarda kliniklerde toplanan test verileri arasında yapılan veri madenciliği çalışmaları teşhiste %100 oranında doğruluk sağlamıştır [9].

Mustafa Danacı, Mete Çelik ve A. Erhan Akkaya tarafından 2010 yılında gerçekleştirilen çalışmada kanser çeşitlerinden biri olan ve kadınlar arasında en sık görülen meme kanseri hakkında kısa bilgi verilmiştir. Daha sonra Xcyt örüntü tanıma programı yardımı ile doku hakkında genel veriler elde edilmiş, Weka programı kullanılarak meme kanseri hücrelerinin tahmin ve teşhisi yapılmıştır [10]. Harleen Kaur ve arkadaşları hastaların yaş ve cinsiyet gibi verilerini karar ağacı yöntemleri ile analiz ederek göğüs kanseri olup olmadığını tahmin etmeye çalışmışlardır [11]. Pankreas kanseri verileri içeren farklı iki operasyonel sağlık sistemi

bütünleştirilmeye çalışılmış ve veri türleri farklılık gösterdiği için karşılaşılan zorluklardan bahsedilmiştir [12]. Dr. Usame E. Sheta, Bill Inmon'ın veri ambarı tanımını dikkate alarak kanser hastalığı için veri ambarı modeli oluşturulması üzerine değerlendirmeler yapmıştır [13]. S. Palaniappan, C. Ling veri ambarı ve veri madenciliğinin tek başına yapamadıklarını birlikte kullanılırsa yapılabileceği üzerine bir çalışma yapmışlardır [14]. Ontoloji yaklaşımı kullanarak prostat kanseri içeren iki veritabanının birleştirilmesi üzerine bir çalışma yapılmıştır[15]. Göğüs kanseri araştırması için klinik veriler ve veritabanları kullanılarak bir veri ambarı modeli oluşturulmuştur. Oluşturulan veri ambarının karmaşık sorgulara bile cevap verdiği belirtilmiştir [16].

3. Tıp Alanında Veri Ambarı Modeli

Bill Inmon veri ambarını aşağıdaki 4 adımda karakterize etmiştir.

- *Veri ambarları nesne merkezlidir*. Belirli bir nesne hakkında ilgili tüm veriler toplanıp tek bir veritabanında saklanır. Örneğin hasta verilerinin toplandığı bir veri ambarı tasarlayabilirsiniz. Böylece bir önceki sene hangi hastanın hangi ilaçlarını kullandığı sorusuna cevap verebilirsiniz. Bu şekilde nesne üzerine odaklanan veri ambarları nesne odaklı veri tabanı adını alır.

- *Veri ambarları bütünleşmiş bir sistemdir.* Bütünleme aslında nesneye odaklanma ile oldukça ilişkilidir. Örneğin hastanın tahlil bilgileri ile teşhis bilgileri farklı kaynaklarda ve farklı şekillerde tutulmaktadır. Veri ambarları farklı kaynaklarda bulunan verileri ortak bir yerde toplayıp tutarlı hale getirmek durumundadırlar. Aynı zamanda adlandırma problemlerini de çözebilmelidirler. Bunu gerçekledikleri zaman bütünleşme problemi de çözümlenmiş olur.

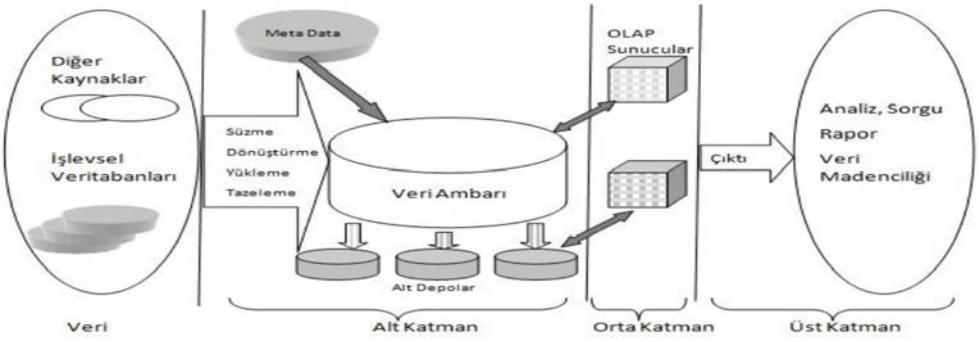
- *Veri ambarları değişken değillerdir.* Veri ambarına bir defa giren bir veri, veri ambarı içinde iken bir daha değişmemelidir. Çünkü veri ambarını kullananlar verilerin analizi ile ilgilenirler. Veri ambarı içinde bulunan verilerin devamlı değişikliklere uğraması istenen bir durum değildir. Hastane sistemlerinde de hastaların verileri uzun yıllar saklanır.

- *Veri ambarları zamanla büyür.* Zaman ilerledikçe veri ambarına giren veri miktarı artmış olacağından veri ambarı zamanla büyür. Bu da analistlerin analiz yaparken daha fazla veri kullanabilmelerine imkân tanır [17].

İlişkisel veritabanları, sıralama, arama ve hazır sorguları çalıştırma işlemleri için, veri ambarları ise özel veri organizasyonu, veri analizi ve çabuk erişim için optimize edilirler. Veri ambarı kurulmadığı durumlarda işlevsel veritabanı performansı önemli ölçüde düşerken, yapılan karar destek işlemleri doğruluktan uzaklaşmaktadır. Ayrıca, karar verme işlemleri tarihsel veriler gerektirdiği için veri ambarı karar destek sistemleri için vazgeçilmez bir unsurdur [18].

Genel olarak veri ambarı mimarisinde üç temel katman vardır. Şekil 1'de görülen bu katmanlar şunlardır:

3.1 Veri Toplama Bölümü: Hastanın farklı kaynaklardan toplanan tıbbi kayıtları, kan testleri, idrar testi sonuçları, röntgen sonuçları gibi verilerinin hatalardan arındırılıp, istenilen formatlara getirildiği bölümdür. Bu aşamada veriler, gürültülerden, tutarsızlık ve düzensizliklerden süzülerek veri ambarına yüklemeye hazır



Şekil 1. Veri ambarını oluşturan katmanlar [19]

hale getirilir. Veri ambarından alınacak sonucun kalitesinde büyük öneme sahip olan bu bölüm, tasarımın en uzun ve zaman alıcı aşamasıdır.

3.2. Veri Depolama Bölümü: Bir alt katmandan gelen verilerin analiz çalışmaları için anlamlı hale getirilip veri ambarı deposuna yüklendiği bölümdür. Bu aşamada veriler çok boyutlu modelleme yaklaşımlarına (yıldız, kar tanesi, galaksi şema modeli vb.) göre veri ambarında tutulur. Böylelikle verilere farklı bakış açılarından yaklaşılması sağlanmaktadır.

3.3. Bilgi Elde Etme Bölümü: Depolanan verilerin analiz, sorgulama, raporlama, veri madenciliği, karar destek gibi amaçlarla kullanıldığı bölümdür. Veri ambarındaki veriler kullanılacak amaçlara göre ayrı alt depolara ayrılabilir. Verilerin çok boyutlu analizleri, OLAP operasyonları, tablolar ve grafikler gibi raporlamalar, verilerdeki gizli kalıplardan bilgi keşfi bu aşamada gerçekleştirilir.

Veri ambarları tasarlanırken çok boyutlu veri modelleme yaklaşımı kullanılmaktadır. Bu yaklaşım verileri hem daha sistematik şekilde depolamayı hem de iş zekâsı, OLAP, veri madenciliği gibi işlemler için daha kullanışlı hale getirmeyi amaçlamaktadır. Çok boyutlu veri modelleme yaklaşımında veri ambarı, olgu tablosu (fact table) ve ona bağlanan boyut tablolarından (dimension table) oluşmaktadır. Bir olgu tablosu, boyut tablolarını temsil eden anahtarları ve bazı sayısal değerleri (measurement) barındırmaktadır[20]. Çok boyutlu modellemenin uygulanması için en popüler ve iyi bilinen yöntemler *yıldız şeması* ve *kar tanesi şeması*dır. **Yıldız şeması**, bir olgu tablosunun etrafına bir dizi denormalize edilmiş boyut tablosunun bağlanmasıyla oluşmaktadır. Boyut tabloları, her bir veri yapısını sınıflandırarak, diğerleriyle örtüşmeyecek şekilde ayrıştırır ve olgu tablosuyla bağlanmasını sağlayan birincil anahtar içerir. **Kar tanesi şema** modelinde ise boyut tabloları tek bir büyük tabloya toplanmak yerine ayrıştırılmış tablolarda bulunmaktadır. Yıldız şema modelinden farklı olarak kar tanesi şema modelinde, bazı boyut tabloları direkt olgu tablosuna bağlı olmadan başka bir boyut tablosuna bağlanmaktadır.

Son yıllarda tıp alanında önemi gittikçe daha çok anlaşılan veri ambarları, doktorlar ve sağlık kurumlarının tedavilerin kalitesini arttırmak aynı zamanda maliyetleri azaltmak amacıyla kullandıkları çok fazla miktarda klinik ve yönetimsel veri barındıran bir depodur. Sağlık kuruluşları hem araştırmalar için geniş bir veri havuzu elde etmek hem de karar destek hizmeti sağlamak için kendi verilerini mümkün

olduğunca bütünleşmiş bir şekle getirmeye çalışmaktadırlar. Veri ambarı yaklaşımıyla farklı operasyonel sistemlerden gelen veriler ortak bir havuzda toplanarak sorgulama ve analiz gibi işlemler için uygun hale getirilmiş olur.

Kanser günümüzde en önde gelen ölüm nedenlerinden biridir. Bu ölümlerin büyük bir ölümünü akciğer kanseri oluşturmaktadır. Dolayısıyla veri ambarı yapısını akciğer kanseri gibi ciddi bir hastalıkta kullanmak doktorların hastalığa daha geniş bakış açısıyla bakmalarını sağlayacaktır. Akciğer kanseri için "zaman" en önemli faktörlerden biridir. Tanı ve tedavi süreci ne kadar erken başlarsa, sonuç o kadar başarılı olabilmektedir. Doktorun verileri dosya yığınları içerisinde aramasıyla ayrı bir zaman kaybı oluşur. Bunu önleyebilmek için, veriler doktorun ihtiyacı olduğu anda ulaşabileceği veri ambarı yapısında toplanmıştır. Böylelikle doktor hastanın geçmiş verilerinden istediği şekilde sorgulama ve analiz yapabilecek, hastalığın seyri daha kolay izlenebileceğinden daha etkin tedavi yolları uygulanabilecektir.

4. Materyal ve Metot

Bu çalışmada Amerikan Ulusal Kanser Enstitüsü'nden (NCI-National Cancer Institute) [1] alınan akciğer kanseri verileri belirli formatlara getirilerek, hatalardan ve tekrarlardan arındırılarak, karar destek sistemlerinde kullanılmak üzere veri ambarı yapısında tutulmuştur.

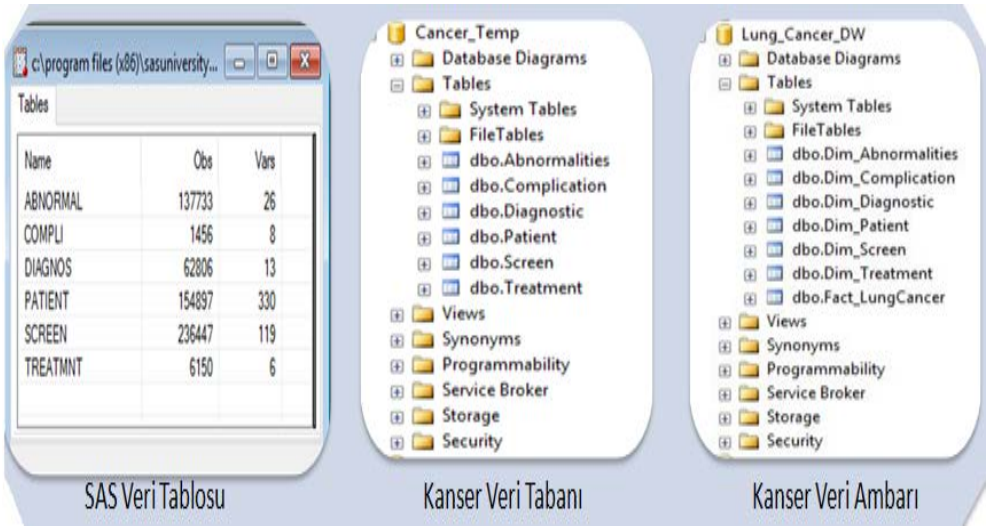
4.1. Veri Kaynağı

Çalışma için gerekli olan veriler Amerika'da kanser araştırmaları için başlıca kuruluş olan NCI' den alınmıştır. NCI, kanser hastalarının bakımını ve iyileştirilmesini sağlayan, kanserin teşhisi, önlenmesi ve tedavisi ile ilgili araştırmaları destekleyen Ulusal Kanser Programı'nı koordine etmektedir. Kanserle ilgili geniş veri havuzuna sahip olan NCI' ye proje fikri sunulup onay alındıktan sonra veri seti ve veri açıklama kılavuzları Kanser Veri Erişim Sistemi (CDAS- Cancer Data Access System) aracılığıyla tarafımıza iletilmiştir. Bu veri seti yaklaşık 155.000 hastaya ait demografik bilgi, teşhis, tedavi, yan etki, röntgen sonuçları, röntgenlerde saptanan anormallikler gibi değerleri barındıran oldukça kapsamlı bir veri kümesidir. Veri ambarı modelini oluştururken hataları en aza indirmek ve kayıtları temizlemek için veriler üzerinde süzme, dönüştürme gibi işlemler uygulanmıştır.

4.2. Veri İşleme

Veri işleme, verilerden daha kaliteli ve kesin sonuçların alınabilmesi için yapılan süzme, temizleme, dönüştürme gibi işlemleri kapsayan veri ambarı tasarımın en önemli ve en zaman alıcı bölümüdür. Tıp alanında veri ambarı modelini oluşturmak için verileri çok iyi tanımlamak gerekir. Bu nedenle öncelikli olarak akciğer kanseri konusunda uzman kişilerin görüşleri alınmış, veriler tarafımızca daha anlaşılabilir ve kolay yorumlanabilir hale gelmiştir. SAS veri tablosu biçimindeki verilerin daha iyi incelenebilmesi için SAS University Edition kullanılarak veriler Microsoft Excel tablolarına aktarılmıştır. SQL sorgularıyla verileri ayıklama işlemi yapmak daha kolay olacağından excel tablosundaki verilerden Microsoft SQL Server kullanarak geçici bir veri tabanı oluşturulmuştur. Bu veri tabanında veriler eksik ve hatalı olanlardan arındırılıp bazı dönüşüm işlemleri uygulanarak veri ambarına yükleme yapmak için hazır hale getirilmiştir. Yapılan dönüşüm işlemlerine örnek vermek gerekirse, hastaların tedavi bilgilerinin tutulduğu veri tabanındaki Treatment tablosunda tedavi türünü tutan değer 1=Kemoterapi, 2=Lazer Terapi vs. şeklinde kodlanmışken, daha

ayrıntılı sorgulama yapılabilmesi için veri ambarındaki Dim_Treatment boyut tablosunda ise bu değer hasta hangi tedaviyi almışsa o 1, diğer tedavi türleri 0 olacak şekilde dönüşüm yapılarak kodlanmıştır. Aynı şekilde veriler veri tabanından veri ambarına geçirilirken buna benzeyen birçok dönüşüm yapılmıştır. Veri işleme aşamaları Şekil 2'de gösterilmiştir.



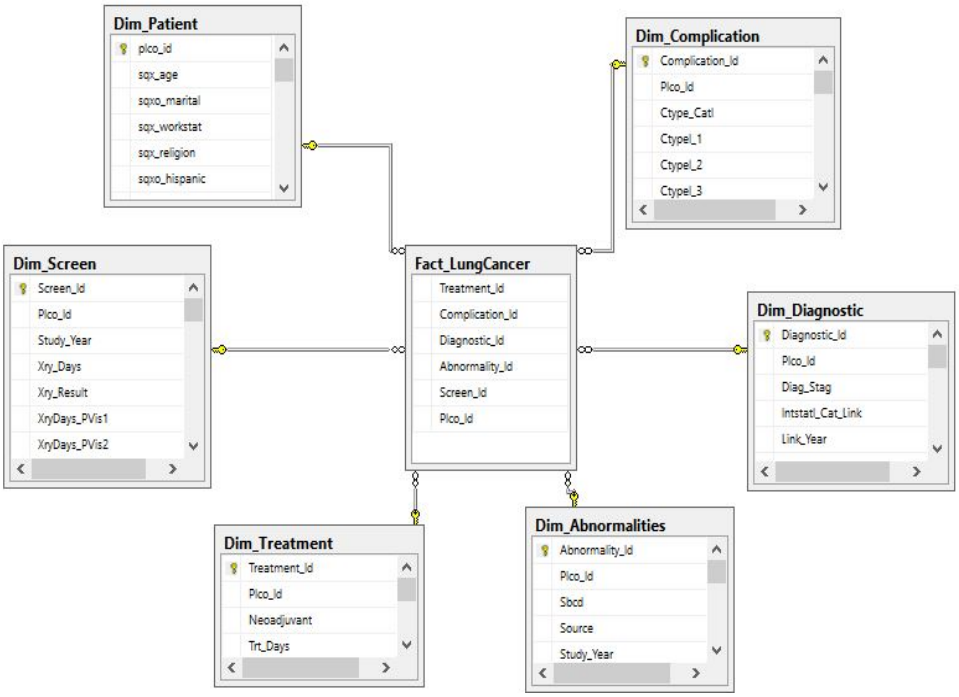
Şekil 2 . Veri işleme aşamaları

4.3. Veri Depolama

Bu aşama, süzme, dönüştürme işlemlerinden geçirilmiş verilerin veri ambarına yüklenmesi ve depolanması sürecini kapsar. Veriler, OLAP işlemlerini gerçekleştirebilmek için çok boyutlu bir modelleme yaklaşımı olan yıldız şema modelinde tutulmuştur. Çok boyutlu veri modeli, küp, boyut ve hiyerarşi gibi kavramlara dayanır. Bu model, kullanıcılarının akciğer kanseri veri ambarındaki sonuçları çeşitli şekillerde görüntülemelerine ve ayıklamalarına olanak tanır. Kullanılan yıldız şema modeli kolay anlaşılabilir ve sorgu performansları için uygun bir yöntemdir. Bu model, içerisinde veri ambarının içerdiği en önemli veri kısmını gereksiz fazlalık olmadan bulunduran büyük bir merkezi tabloya (olgu tablosu) her biri bir boyut için olmak üzere küçük yardımcı tabloların (boyut tabloları) bağlanması sonucu oluşturulmuştur. Microsoft SQL Server kullanılarak oluşturulan akciğer kanseri veri ambarı modelinde hasta bilgileri, yan etkiler, teşhisler, tedaviler, röntgen sonuçları ve röntgenlerde saptanan anormallikler olmak üzere altı tane boyut tablosu ve bir tane akciğer kanseri olgu tablosu bulunmaktadır. Şekil 3'de görüldüğü gibi olgu tablosu, boyut tablolarının her birini temsil eden anahtarları içermektedir.

4.4.Veri Ambarında Yapılabilecek Sorgulama Örnekleri

Oluşturulan veri ambarı büyük ve karmaşık sorgulara yanıt verebilecek şekilde tasarlanmıştır. Böylelikle tıbbi analistlerin ve doktorların veriye farklı bakış açılarından bakarak karar verme süreçlerine yardımcı olmak amaçlanmıştır. Aşağıda kurulan veri ambarını kullanarak yazılan örnek sorgular için bulunan sonuçlar Tablo 1-3 arasında sırasıyla gösterilmiştir.



Şekil 3. Akciğer Kanseri Veri Ambarı Yıldız Şema Modeli

Sorgu 1: Günde 3'ten fazla sigara içen ve ailesinde akciğer kanseri geçmişi olan hangi hastalarda akciğer kanseri saptanmıştır?

```
select * from Dim_Patient
where sqx_amt_smk>3 and
sqx_fh_lung=1 and plco_id in
    (select plco_id
     from Dim_Diagnostic
     where Proc_Res_3=1 )
order by sqx_age
```

Sorgu 2: Pnömotoraks, plevra boşluğunda hava toplanması, komplikasyonu olan ve kemoterapi tedavisi gören hastaların yaşlara göre dağılımı nasıldır?

```
select sqx_age as YAS,
count(plco_id) as SAYI from
Dim_Patient P
where plco_id in
    (select T.Plco_Id from
    Dim_Treatment T inner join
    Dim_Complication C on
     T.Plco_Id= C.Plco_Id and
    T.Trt_Familyl_4=1 and
    C.Ctypel_3=1)
 group by sqx_age
```

Sorgu 3: Röntgen sonuçlarında nodül görülen ve biyopsi yapılan hastaların günlük sigara içme sayısına göre dağılımı nasıldır?

```
select sqx_amt_smk as
SIGARA_MIKTARI, count(plco_id)
as SAYI from Dim_Patient P
where plco_id in
(select A.Plco_Id
 from Dim_Abnormalities A
inner join Dim_Diagnostic D on
A.Plco_Id= D.Plco_Id and
A.Desc_1=1 and D.Biop=1)
group by sqx_amt_smk
order by sqx_amt_smk asc
```

Tablo 1. Sorgu 1 sonuç ekranı

Results Messages

	plco_id	sqx_age	sqx_amt_smk	sqx_height	sqx_fh_lung	sqxo_marital	sqx_workstat	sqx_religion	sqxbq_race_nh6	sqxo_race7	sqx_bmi_curr	sqx_wt_curr
1	R-000494-5	60	4	64	1	1	2	1	5	1	18,364501953125	107
2	R-064016-1	61	4	79	1	4	7	1	5	1	34,5811568658869	307
3	Q-141127-2	61	5	57	1	3	7	1	5	1	21,6374269005848	100
4	J-069569-5	61	6	72	1	1	7	1	5	1	22,3755787037037	165
5	A-121976-7	62	5	71	1	1	2	1	5	1	32,0749851219996	230
6	B-081981-5	62	4	62	1	1	2	1	5	1	25,9693028095734	142
7	E-094665-0	62	5	66	1	1	2	6	5	1	26,1446280991736	162
8	S-050814-5	62	5	67	1	1	2	4	5	1	25,83983069726	165
9	Q-037402-5	62	4	70	1	1	7	1	5	1	26,5418367346939	185
10	O-016601-8	62	5	70	1	1	7	1	5	1	33,5718367346939	234

Tablo 2. Sorgu 2 sonuç ekranı

Results Messages

	YAS	SAYI
2	64	1
3	65	2
4	66	2
5	67	7
6	68	5
7	69	7
8	70	1
9	71	4
10	72	2
11	73	2
12	74	6
13	75	3
14	76	2
15	78	3
16	79	2
17	80	1

Tablo 3. Sorgu 3 sonuç ekranı

Results Messages

	SIGARA_MIKTARI	SAYI
2	1	5
3	2	7
4	3	11
5	4	1
6	5	5

5. Sonuç, Değerlendirme ve Öneriler

Sağlık sektörü bilgiye dönüşmeyi bekleyen çok fazla miktarda veri barındırır. Sağlık kuruluşlarının, karar verme aşamasında bu verileri daha bilinçli şekilde kullanarak değerlendirmeye, daha etkin tedavi yöntemleri geliştirmeye, kaynakların daha iyi yönetilmesini sağlayarak maliyetleri düşürmeye ihtiyaçları vardır. Farklı sistemlerden gelen, tekrarlardan, hatalardan ve belirsizliklerden arındırılan veriler seçim, ön işleme, dönüştürme işlemlerinden geçirilerek veri ambarında tutulması bu amaçları gerçekleştirmenin ilk adımıdır. Bütünleşmiş bir şekilde veri ambarında bulunan bu veriler analiz, sorgulama, raporlama ve veri madenciliği için kullanılır. Veri ambarı ve veri madenciliği tekniklerinin karar destek amacıyla kullanılması sağlık alanında yeni bir yön olarak karşımıza çıkmıştır. Veri ambarları farklı sistemlerden gelen büyük miktarda ve çok boyutlu verileri saklayabildiği, karmaşık sorgulara yanıt verebildiği için karar verme sürecini kolaylaştırdığı gibi aynı zamanda verileri veri madenciliği algoritmaları içinde hazır hale getirmiş olurlar.

Bu çalışmada akciğer kanseri ile ilgili veriler için veri ambarı kurulmuştur ve veri madenciliği tekniklerinin uygulanması için hazır duruma getirilmiştir. Oluşturulan bu modelde yazılan karmaşık sorguların sonuçları sunulmuştur. Veri madenciliği teknikleri kullanılarak akciğer kanseri hastalığıyla ilgilenen doktorların kararlarını alabilmelerinde onlara fikir sağlayacak bir sistem geliştirmek amaçlanmaktadır. Veri madenciliği sonucu elde edilen sonuçlar doktorların kanseri daha erken seviyelerde teşhis etme ya da daha etkin tedavi yollarının seçilmesi gibi kararları almalarında onlara yardımcı olabilmektedir.

Tıp alanındaki verilerin dijital ortama geçilmesiyle birlikte günden güne arttığı, adeta bir yığın haline gelen bu veriler arasından gizli kalmış, değerli,

kullanılabilir bilgilerin çıkarılması pek çok konuya ışık tutabilecektir. Çalışmalarımıza diğer kanser türlerinin verilerini veri ambarına ekleyerek devam etmeyi planlıyoruz. Farklı kanser türlerine ait verilerden oluşturacağımız bu veri ambarında veri madenciliği algoritmaları uygulayacağız. Veri ambarı ve veri madenciliğinden yararlanarak oluşturacağımız karar destek sistemi gelecekte yapmayı planladığımız işler arasındadır. Kanser verilerinin büyüklüğü ve karmaşıklığı düşünüldüğünde veri ambarı ve veri madenciliğinin birlikte kullanılmasıyla kurulacak bir karar destek sistemi, doktorların tedavilerini şekillendirmesinde farklı bakış açıları sağlayacaktır.

Teşekkür

Yazarlar olarak bu çalışmayı yapmamız için akciğer kanseri verilerini bizimle paylaşan Amerikan Kanser Enstitüsü'ne (NCI) teşekkür ederiz.

Kaynakça

[1] https://www.cancer.gov/ (Erişim Tarihi: 19.03.2017)

[2] http://globocan.iarc.fr/ (Erişim Tarihi: 17.02.2017)

[3] http://kanser.gov.tr/ (Erişim Tarihi: 15.01.2017)

[4] Kallmeyer V.,Venkat K., "Beyond e-Health: Health and IT Converge", Siliconindia, Vol. 6, Issue 4, April 2002, p. 42.

[5] Koç E, Şengül A. Y, Özkaya U. A, Klinik Karar Destek Sistemlerinin Sağlık Hizmetleri Verimliliğine Etkileri, 6.Sağlık ve Hastane İdaresi Kongresi, 2012

[6] Özata M., Aslan Ş., "Klinik Karar Destek Sistemleri ve Örnek Uygulamalar", Afyon Kocatepe Üniversitesi Tıp Dergisi, Cilt: 5, Sayı:1, Ocak 2004.

[7] Savaş S., Topaloğlu N.,Yılmaz M., Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Sayı:21 ,Sayfa:1-23

[8] A. Kusiak, Decomposition in Data Mining: A Medical Case Study, in B.V. Dasarathy (Ed.), Proceedings of the SPIE Conference on Data Mining and Knowledge Discovery: Theory, Tools, and Technology III, Vol. 4384, SPIE, Orlando, FL, April 2001, pp. 267-277

[9] Kaya, E., Bulun, M., Arslan, A., (2003), Tıpta Veri Ambarları Oluşturma ve Veri Madenciliği Uygulamaları, Akademik Bilişim 2003, Çukurova Üniversitesi, Adana

[10] Danacı, M., Çelik, M. ve Akkaya, A.E., (2010), "Veri Madenciliği Yöntemleri Kullanılarak Meme Kanseri Hücrelerinin Tahmin ve Teşhisi", Akıllı Sistemlerde Yenilikler ve Uygulama Sempozyumu, 21-24 Haz. 2010, Kayseri, 9-12.

[11] Kaur., H.,ve Wasan., S., "Empirical Study on applications of Data Mining Techniques in Healthcare", Journal of Computer Science 2(2), 2006.

[12] Edward J. Arous, MD, MPH Electronic medical record: research tool for pancreatic cancer, 2013

[13] Dr. Osama E.Sheta "Evaluating a Healthcare Data Warehouse for Cancer Diseases" IRACST - International Journal of Computer Science and Information Technology & Security (IJCSITS), ISSN: 2249-9555 Vol. 3, No.3, June 2013

[14] Palaniappan S., Ling C., Clinical Decision Support Using OLAP With Data Mining, IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.9, September 2008

[15] Hua Min, Frank J. Manion, Elizabeth Goralczyk, Yu-Ning Wong, Eric Ross, J. Robert Beck, Integration of prostate cancer clinical data using an ontology,2009

[16] Hai hu, Biomedical informatics: development of a comprehensive data warehouse for clinical and genomic breast cancer research,2014

[17] Arslan V., Yılmaz G., Karar Destek Sistemlerinin Kullanımı İçin Uygun Bir Model Geliştirilmesi, 2010

[18] Çorapçıoğlu M., TIPÇIT: Tıbbi Karar Destek Sistemi Çekirdeği, Yayınlanmamış Yüksek Lisans Tezi, Başkent Üniversitesi, Ankara, 2006, s. 1

[19] Poyraz Ö., Tıpta Veri Madenciliği Uygulamaları: Meme Kanseri Veri Seti Analizi, 2012

[20] Malinowski E. ,Zimanyi E, 2007 , Advanced Data Warehouse Design, From Conventional to Spatial and Temporal Applications.