

BÖLÜM 4:

**BÜYÜK VERİ VE BİLİMİ İLE
SAHTECİLİĞİN VE SUİSTİMALİN
ÖNLENMESİ**

1 Giriş

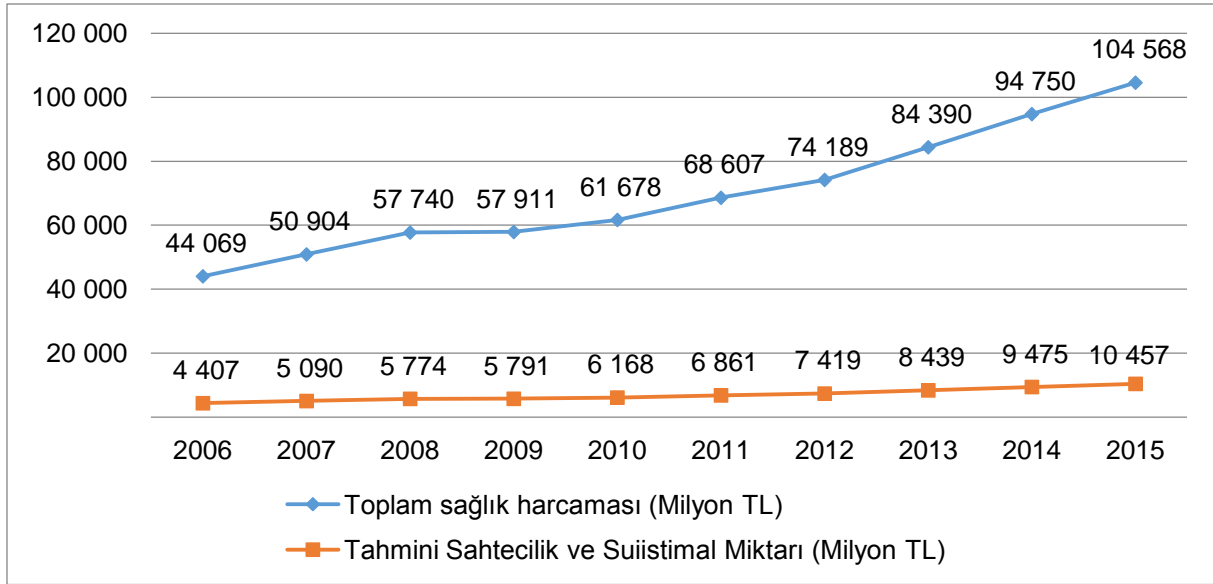
Ortalama yaşam süresi ve refahın artması, teknolojinin hızla gelişmesi, yeni tedavi yöntemlerinin ve ilaçların geliştirilmesi, sağlık harcamalarını ülkelerin giderleri arasında üst sıralara taşımıştır. Sağlık harcamalarının geri ödeme süreçleri, ülkelere göre farklılık göstermektedir. Bununla birlikte hemen hepsinde ortak olan nokta ödeme sisteminin karmaşık olması ve sürece pek çok paydaşın dâhil olmasıdır. Hem sağlık harcamalarının oluşturduğu ekonominin büyük olması; hem de takip edilmesini zorlaştıracak çeşitli nedenler (işlemlerin çok parçalı yapıda olması, hekim, sigortacı, veri bilimcisi gibi çok disiplinli bir ekibe duyulan ihtiyaç, sahtecilik yapma davranışlarının alınan tedbirlere göre yeniden şekillenmesi, vb.), sağlık geri ödeme sistemlerini sahtecilik ve suiistimale daha açık hale getirmektedir.

Sahtecilik ve suiistimal sıklıkla birbiri ile karıştırılan ve bir diğ erinin yerine kullanılabilen kavramlardır. Literatürde bu kavramların şu şekilde tanımlandığını görüyoruz [1]:

Sahtecilik (fraud): “Sağlık hizmetinde sahtecilik, özel ya da tüzel kişinin, kasıtlı bir aldatma ya da yanlış beyanı neticesinde ilgili özel ya da tüzel ya da farklı bir tarafa yetkilendirilmemiş bir fayda sağlayan davranıştır.

Suiistimal (abuse): Sağlık hizmetlerinde suiistimal, finans, iş ve medikal süreçlerde hizmet sunan tarafından yapılan ve sonucunda gereksiz bir ek maliyetin doğmasına neden olan ya da medikal açıdan gerekli olmayan ya da profesyonel yaklaşım ve standartlara uygun olmayan işlemlerin yapılmasıyla fazladan ödeme yapılmasını gerektiren uygulamalardır.”

Sahtecilik ve suiistimal vakalarının, toplam sağlık harcaması içerisindeki oranını mutlak kesinlikte tespit etmek mümkün değildir. Ancak tespit edilen vakalardan yola çıkarak bu oranın %3 ila %10 arasında olduğu tahmin edilmektedir [2]. Bu oransal tahmini, TÜİK tarafından ilan edilen toplam sağlık harcamalarında 2015 yılına kadar olan son 10 yıllık sağlık dönemi esas alarak yansıttığımızda, Şekil 1-1'deki tahmini sahtecilik ve suiistimal miktarı elde edilmektedir.



Şekil 1-1 TÜİK, Türkiye'nin Toplam Sağlık Harcaması [3] ve Tahmini Sahtecilik ve Suiistimal Miktarı

Bu şekilde de görüleceği üzere, ülkemizde bütün sağlık harcamalarının içinde sahtecilik ve suiistimal oranını kabaca %10 kabul ettiğimizde, 10 yılda sahtecilik ve suiistimal nedeniyle ülke ekonomimizin kaybı yaklaşık 70 Milyar TL gibi bir rakama ulaşabilmektedir. Bu miktar, 2011 yılının toplam sağlık harcamasından daha çoktur ve 2015 sağlık harcamamızın ise %66,8'ine tekabül etmektedir. Bu yönü ile ele alındığında sahtecilikle mücadele, ülke ekonomisi açısından önemli sorunlarımızdan birisidir.

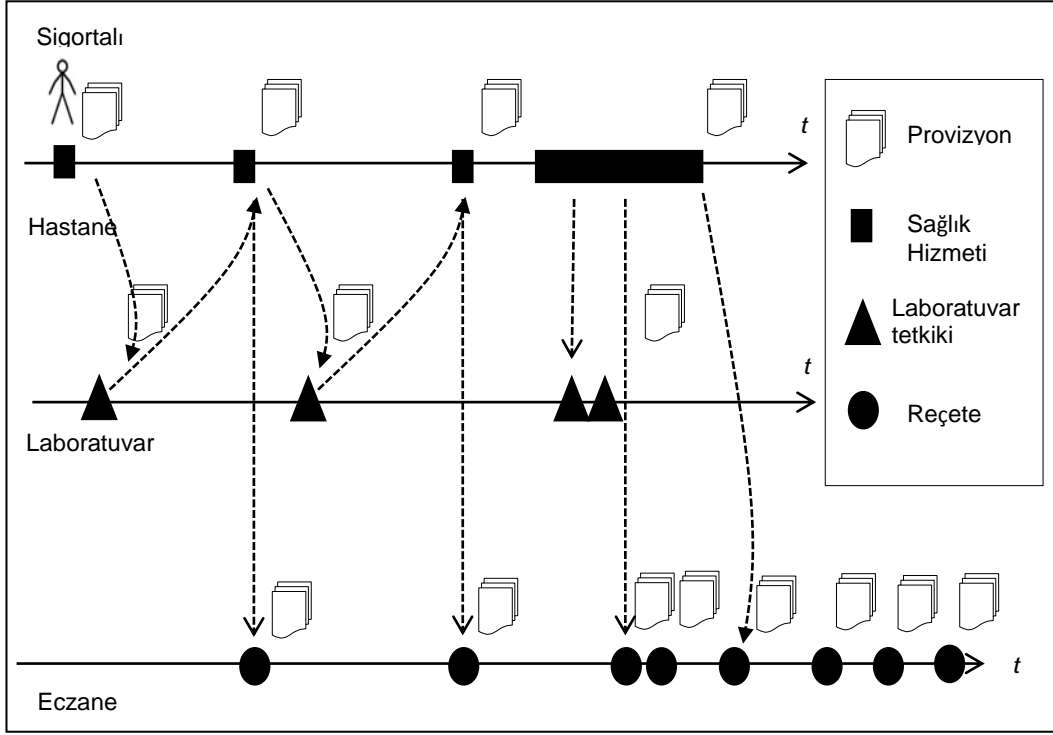
2 Sahtecilik Tespitinde Neden Büyük Veri

Sahtecilik ve suiistimal ile mücadelenin önünde teknik ve pratik çok sayıda zorluk söz konusudur. Klasik veri madenciliği yöntemleriyle sahtecilik tespiti yapılmasını güçleştiren başlıca zorluklar müteakip başlıklarda (Bölüm 2.1, 2.2 ve 2.3) ele alınmaktadır.

2.1. İşlemlerin Çok Parçalı Yapıda Olması

Sağlık sistemlerinde, hastane, poliklinik, laboratuvar, görüntüleme merkezi, eczane, optik merkezi ve ortez-protez merkezi, gibi çok sayıda hizmet sunucusu yer almaktadır. Bir sigortalı, hastalığının tedavi süreci boyunca bu sağlık hizmet

sunucularının birkaçından farklı zamanlarda hizmet alabilmektedir. Bu durumda, her alınan sağlık hizmeti için geri ödeme kurumunda ayrı bir provizyon talep edilmektedir. Örnek bir sigortalı için, hastane, laboratuvar ve eczaneden alınan farklı hizmetler için yapılan provizyon talepleri Şekil 2-1’de gösterilmektedir.



Şekil 2-1 Bir tedavi sürecinde çok sayıda geri ödeme talebinin gerçekleşmesi [4]

Bu örnekte sigortalı önce hastaneye gitmekte, hekimin talebi üzerine laboratuvarında tahlil(ler) yaptırmakta, tahlil sonucu üzerine ilaç tedavisi başlamakta ve başka tahliller yaptırmaktadır. Üçüncü görüşmenin ardından hem yeni ilaç tedavisi başlamakta hem de yataklı tedavi sürecine geçilmekte, bu süreçte de çeşitli tahliller yaptırmakta, bir müddet sonra taburcu olup, düzenli ilaç tedavisine başlamaktadır. Sürecin her aşamasında yapılan işlemler sonucunda sigorta sistemine ayrı bir provizyon kaydı açılmaktadır. Dahası, bir sigortalı için farklı zamanlarda yapılan provizyon taleplerinin, aynı tedavi sürecine mi ait olduğu da kesin şekilde ayırt edilememektedir.

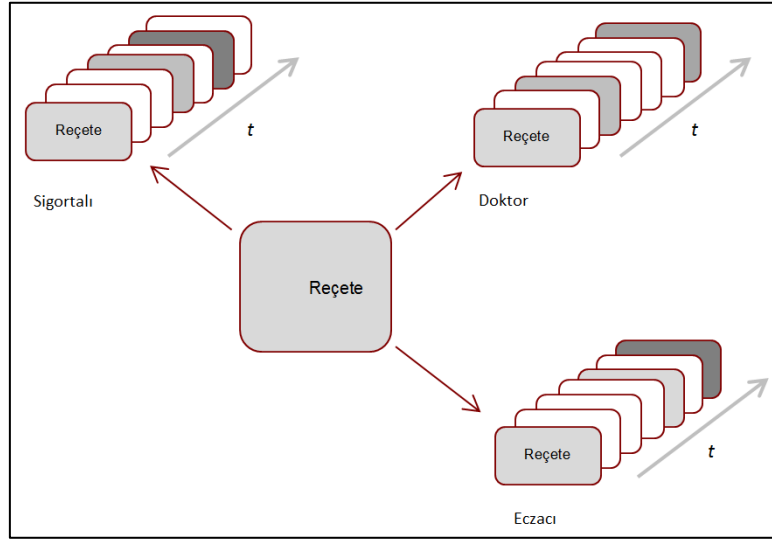
Ülkemizde kullanılan MEDULA sisteminde, vakaları birleştirebilmek için bir Vaka numarası kullanılıyor olsa da, bu numara da aynı sağlık hizmet sunucusu içerisindeki farklı provizyon taleplerinin birleştirilmesine yardımcı olmakta, sigortalının yakın tarihlerde farklı kurumlardan aldığı farklı tedavilerin bu vaka ile ilişkili olup olmadığı bilinmemektedir. İletilen provizyon talebinin içeriğine bakarak bu ilişkiyi kurmak mümkün gibi görünse de, her vaka türü için bu ilişkiyi kesin olarak kurabilmek pratikte

mümkün görünmemektedir. Hâlbuki bir vakanın sahtecilik veya suiistimal barındırıyor olduğuna sağlık ve sigortacılık açısından karar verebilmek için tedavi sürecine dair verileri bütünsel olarak görülebilmesi gereklidir. Fakat açıklandığı üzere, sağlık geri ödeme sisteminin pratikteki işleyişi açısından bunun önünde oldukça önemli engeller bulunmaktadır.

Geri ödeme taleplerinin vaka bütünlüğü içerisinde olmaması, sahtecilik ve suiistimal analizi yapan araştırmacıları, süreci daha küçük parçalara ayırarak analiz etmeye sevk etmiştir. Kimisi, sağlık geri ödeme talebinde yer alan aktörleri (doktor, eczacı, hasta, vb.) [5]–[9], kimisi de geri ödeme talebindeki belirli işlemleri [10]–[12] ele almıştır. Kimi araştırmacılar ise sağlık hizmet sunumunun, klinik kılavuzlardaki önerilere uygunluğunu kontrol ederek sahtecilik tespitine çalışmıştır [13]. Bu çalışmalarda, çalışılan veri içerisindeki sahtecilik ve suiistimal vakalarının tespitinde mesafe kat edilmiş olsa da, sağlık geri ödeme işlemlerinin bütününe sadece yatay ve dikey kesitlerini ele alarak geliştirilen bu modellerin, sağlık geri ödeme sisteminin tamamına teşmil edilmesi ve kullanılması mümkün görülmemektedir. Nitekim söz konusu çalışmaları yapan araştırmacılar da problemi bütünsel olarak ele almanın pratik zorluklarından bahsetmişler ve bu nedenle çalışmalarının alanını daraltmak zorunda kaldıklarını ifade etmişlerdir.

Bu tespiti yaptıktan sonra, geriye kalan seçeneklerden birisi de provizyon talebinde yer alan aktörlerin (doktor, eczacı, hasta, vb.) davranışlarını analiz etmektir. Ancak bu analiz de, bir aktörün yer aldığı provizyonlardaki diğer aktörlerin değişkenlik gösteriyor olması nedeniyle kolay olmamaktadır. Diğer taraftan sahtecilik türlerine göre her bir sahtecilik sürecinde bir veya birden fazla aktör işbirliği yapabilmektedir. Örneğin bir sahtecilik sürecinde yer alan bir eczacı, bu süreci kendi başına yapabildiği gibi, başka bir sahtecilik türünde farklı aktörlerle işbirliği içerisinde de yürütebilmektedir. Bu durumda, her bir sahtecilik türü için ayrı bir analiz yapılması ve farklı aktör kombinasyonlarının ele alınması gibi bir gereksinim ortaya çıkmaktadır. Dolayısıyla problem, her bir sahtecilik davranışında hangi aktörlerin yer aldığını belirlemek ve her bir provizyonu bu sahtecilik türü açısından değerlendirirken, ilgili provizyondaki aktörlerin daha önceki tüm işlemlerinin ele alınarak anomali oluşturan herhangi bir davranış kalıbının yer alıp almadığını incelemek gibi bir hale gelmektedir. Öte yandan, aynı reçete işlemde yer alan bir aktör için (örneğin, belirli bir hekim için) o aktörün geçmiş işlemleri göz önüne alındığında sahtecilik

kapsamında değerlendirilebilecekken başka bir aktör için (örneğin, başka bir hekim için) gayet normal sayılabilecektir. Yani bir provizyona sadece kendi içeriğinden yola çıkarak sahtecilik tespiti yapmak mümkün olmayabilmekte, çoğu kereler taraf olan aktörlerin geçmiş işlemlerinin de dâhil edilmesi gereken karmaşık bir analiz gerektirmektedir. oldukça karmaşık bir analiz gerektirmektedir. Bir provizyonda yer alan aktörlerin, geçmiş işlemlerinin sahtecilik açısından farklı riskler taşıyabiliyor olması, Şekil 2-2'de temsil edilmektedir.



Şekil 2-2 Bir provizyonda yer alan aktörlerin önceki provizyonlarının incelenmesi [4]

Şekil 2-2'de gösterildiği üzere, bir reçetenin risk değerlendirmesi, o reçetede yer alan aktörlerin her birinin daha önceki işlemlerinin incelenmesini gerektirmektedir. Ancak bu inceleme, söz konusu reçetenin risk değerlendirmesi bağlamında yapılacağı için, reçete içerisindeki metâlarla olan ilişkilerin, aktörlerin daha önceki işlemleri üzerinde yapılması, bizi doğru bir noktaya sevk etmektedir. Bu sayede, aktörden ve metadan bağımsız bir risk değerlendirmesi yapma imkânı ortaya çıkmaktadır.

2.2. Sağlık ve Sigortacılık Uzmanlığı Gereksinimi

Özel veya kamu geri ödeme sistemlerinde bir provizyondaki işlemlerin sahtecilik veya suiistimal olarak nitelendirilmesi derin ve kapsamlı bir sağlık ve sigortacılık bilgisi gerektirmektedir. İlk bakışta sahtecilik ve suiistimal olarak değerlendirilebilen pek çok

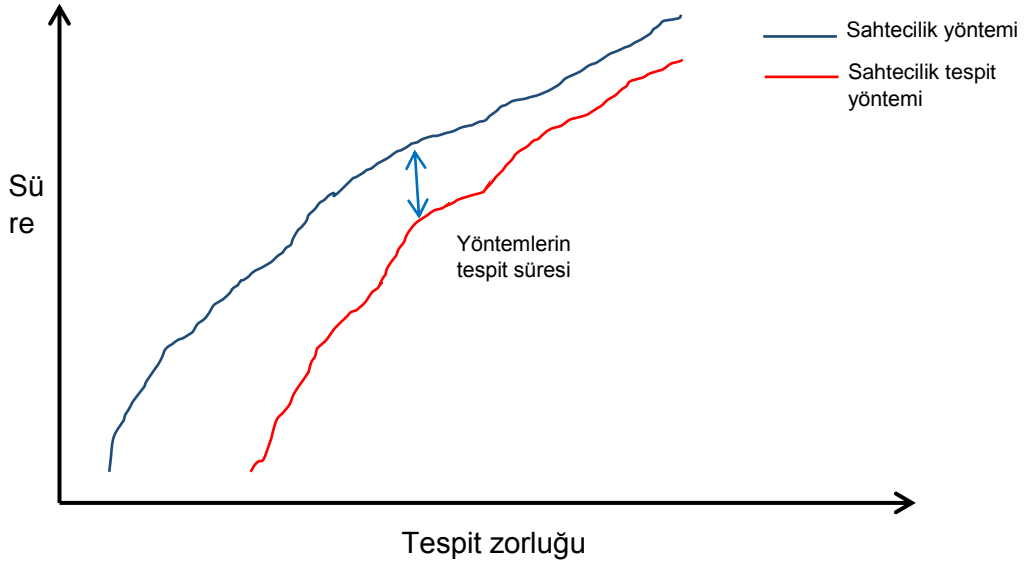
davranış, uygulama hatası, hatalı veri girişi gibi basit insan hatalarından kaynaklanabildiği gibi, hastaya özel bazı sağlık gereksinimlerinden de kaynaklanabilmektedir. Bu durum, istatistiksel analize dayalı yöntemlerle sahtecilik tespiti yapmaya çalışan yaklaşımları büyük oranda işlevsiz kılmaktadır. Örneğin, istatistiksel olarak psikiyatri doktorlarının mide ilacı yazması oldukça nadir görülen bir durum olduğunu varsayalım. Bu durumda istatistiksel analiz yapan yöntemler, mide ilacı yazan doktorların bu nadir davranışını anomali olarak değerlendirecektir. Hâlbuki söz konusu provizyondaki hastanın mide şikâyeti nedeniyle, aldığı diğer psikiyatrik ilaçların mideye etkisini azaltmak için alıyor olması da mümkündür. Dolayısıyla, sadece provizyon talebindeki verilere bakarak söz konusu anomalinin bir sahtecilik durumu olup olmadığını söylemek mümkün olmayacaktır.

Benzer şekilde, provizyonlar değerlendirilirken sağlık sigortacılığı konusunda da ciddi bir bilgi birikimi gereklidir. Örneğin, var olan ödeme kuralları ile süre, sayı ve tekrar sınırı belirlenmemiş sağlık hizmetleri için hangi kullanım sıklığının normal karşılanacağı, belirlenecek yeni ödeme kuralları için oldukça önemlidir ve hem sağlık; hem de sigortacılık bilgisi gerektirmektedir. Sigorta şirketlerinin kafasını en çok yoran konuların başında bir yandan sigortalının gerekli tüm sağlık hizmetini bihakkin alabilmesi sağlanmaya çalışılırken; diğer yandan da makul seviyenin üstünde talep edilen hizmetlerin kontrol edilmesi gelir. Nitekim bir sonraki yıl için hesaplanacak olan poliçe primleri için en önemli parametre, mevcut yıl içerisinde her bir poliçe teminatı (ve mümkünse hastalık için) harcanan toplam tutardır. Primlerin yüksek olması müşteri kaybına neden olurken; az olması da şirketin zarara uğramasına sebebiyet verebilecektir. Benzer durum kamu geri ödeme sistemlerinde de bütçenin aşılması şeklinde karşımıza çıkabilmektedir.

Bu nedenlerle sağlık harcamalarındaki sahtecilik ve suistimal içeren işlemleri takip etmek ve belirleyebilmek amacıyla hem tıbbi; hem de sigortacılık açısından bilgi birikimine sahip bir ekibin oluşturulması gerekmektedir. Üstelik bu sürecin takip edilebilmesi için kurulacak bir teknolojik yapıyı hayata geçirmek için bilişimcilere, kullanılacak olan yöntemlerin belirlenmesi –gerekirse uyarlanması- amacıyla veri bilimcilere, belirlenen sahtecilik ve suistimal vakalarının kesinleşmesi için bazı durumlarda istihbarat çalışmalarına, kesinleşmiş durumlar için de gerekli tazminat işlemlerinin takibi için hukukçulardan oluşan çok disiplinli bir ekibe ihtiyaç bulunmaktadır.

2.3. Sahtecilik Yöntemlerinin Değişiklik Göstermesi

Geri ödeme sistemlerindeki sahtecilik ve suiistimal yöntemleri hakkında konuşurken, tek çeşit ve durağan bir olgudan bahsetmiyoruz. Aksine, çok fazla çeşidi olan ve sürekli değişen yöntemler söz konusudur. Bu yöntemlerin esasını, ülkenin geri ödeme modeli belirler ve zaman içerisinde geri ödeme kurumunun aldığı önlemler ve sistemdeki değişikliklere paralel olarak sahtecilik ve suiistimal yöntemleri de değişir. Kimi yöntemlerde kısmi değişiklikler olurken; kimi sahtecilik yöntemleri yerini tamamen yeni yöntemlere bırakır. Sahtecilik yöntemlerinin sağlıklı şekilde tespit edilebilmesi için veri bilimcisinin, halen uygulanmakta olan sahtecilik yöntemleri ile ilgili detaylı bilgiye sahip olması gerekir. Nitekim analizin esas konusu, bu davranış kalıplarının mevcut provizyon talepleri içerisinde olup olmadığının tespiti. Ancak sahtecilik yöntemlerinin nasıl olduğu, genellikle yeterince vaka tespit edildikten sonra mümkün olabilmektedir. Bir sahtecilik yöntemi bu şekilde deşifre olduktan sonra da zaten gerekli önlemler alınmaktadır. Ancak bu defa da sahtecilik yöntemi değişmekte ve yoluna başka bir şekilde devam etmektedir. Sonuç olarak, sahtecilik ve suiistimal ile mücadeledeki nihai hedef, sahteciliğin olmadığı bir ortam oluşturmak değil; bu yöntemlerin kullanılmaya başlanmasından sonra tespit ve önlem alınmasına kadar geçen süreyi kısaltmaktır. Sahtecilik yöntemlerinin kullanılmaya başlanması ve tespit edilip önlem alınması arasındaki zaman ve zorluk ilişkisi ve ideal tespit yönteminin ulaşması gereken hedef Şekil 2-3'te temsil edilmektedir. Zaman içerisinde suiistimal ve sahtecilik işlemleri daha zor tespit edilen yöntemlere dönüşmekteyken, daha kısa zamanda müdahaleyi mümkün kılacak şekilde yöntemlerin geliştirilmesi gereksinimi bulunmaktadır.



Şekil 2-3 Sahtecilik ve tespit yöntemleri arasındaki süre ve zorluk ilişkisi

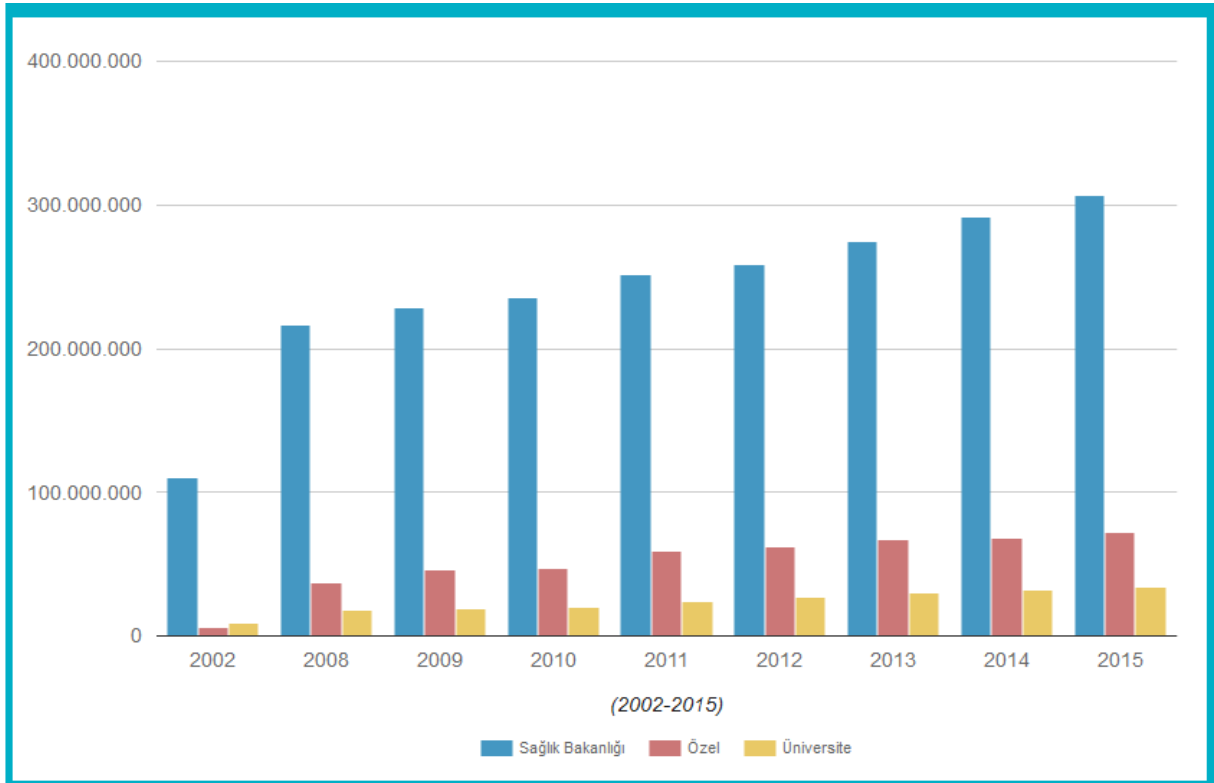
2.4. Anlık Tespit İhtiyacı ve Yanlış Alarm İkillemi

Sağlık geri ödeme sistemlerinde sahtecilik tespiti, yukarıda sıralanan sebeplerden ötürü oldukça karmaşık bir süreçtir. Buna karşın, bu tespitin zamanında yapılamaması halinde maksadın hâsıl olmamasına da neden olabilmektedir. Nitekim geri ödeme kurumları işlemleri hızlandırmaya çalışmakta ve çoğunlukla provizyon cevaplarını on-line olarak vermektedir. Herhangi bir sahtecilik ve suiistimal durumunda da bunun henüz daha provizyon aşamasındayken tespit edilmesi önem kazanmaktadır. Ancak sahtecilik tespitinin daha erken yapılmasına duyulan ihtiyaç, bu süreci daha çok bilişim sistemlerine bağımlı kılmakta ve bu ise daha az insan katkısına yol açmaktadır. Bu durum ise hatalı tespitlerin yapılması riskini daha da artırmaktadır. Bir sahtecilik vakasını tespit edememek (False Positive) maliyet açısından kuruma bir kayıp oluştururken; sahtecilik olmayan bir provizyona da sahtecilik muamelesi yapmak (False Negative) kurumun saygınlığına ve müşteri memnuniyetine zarar vermektedir. Bu risklerden dolayı erken tespit gereksinimini ihmal eden kurumlar da sahteciliği geç fark edebilmekte ve çoğu zaman ödeme yapıldıktan sonra tazminat ve hatta hukuki süreçleri takip etmek durumunda kalmaktadırlar. Bu da ciddi bir zaman ve maliyet yükü getirmektedir.

Dolayısıyla kurumlar, sahtecilik tespiti için bir ikilem yaşamaktadırlar. Öyle ki, kurumsal itibar kaybı riskini göze alarak proaktif (on-line) çözümlerin kullanılması ile maliyet ve zaman kaybını göze alarak reaktif yöntemlerin kullanılması arasında gidip gelmektedirler.

2.5. Veri Akış Hızı

Sağlık geri ödeme sistemlerinde kullanılan elektronik provizyon sistemleri (ülkemizdeki MEDULA, vb. sistemler), işlem hacmi açısından dünyadaki sayılı sistemler arasındadır. Ülkemizdeki yıllık muayene sayıları Şekil 2-4'te temsil edilmektedir.

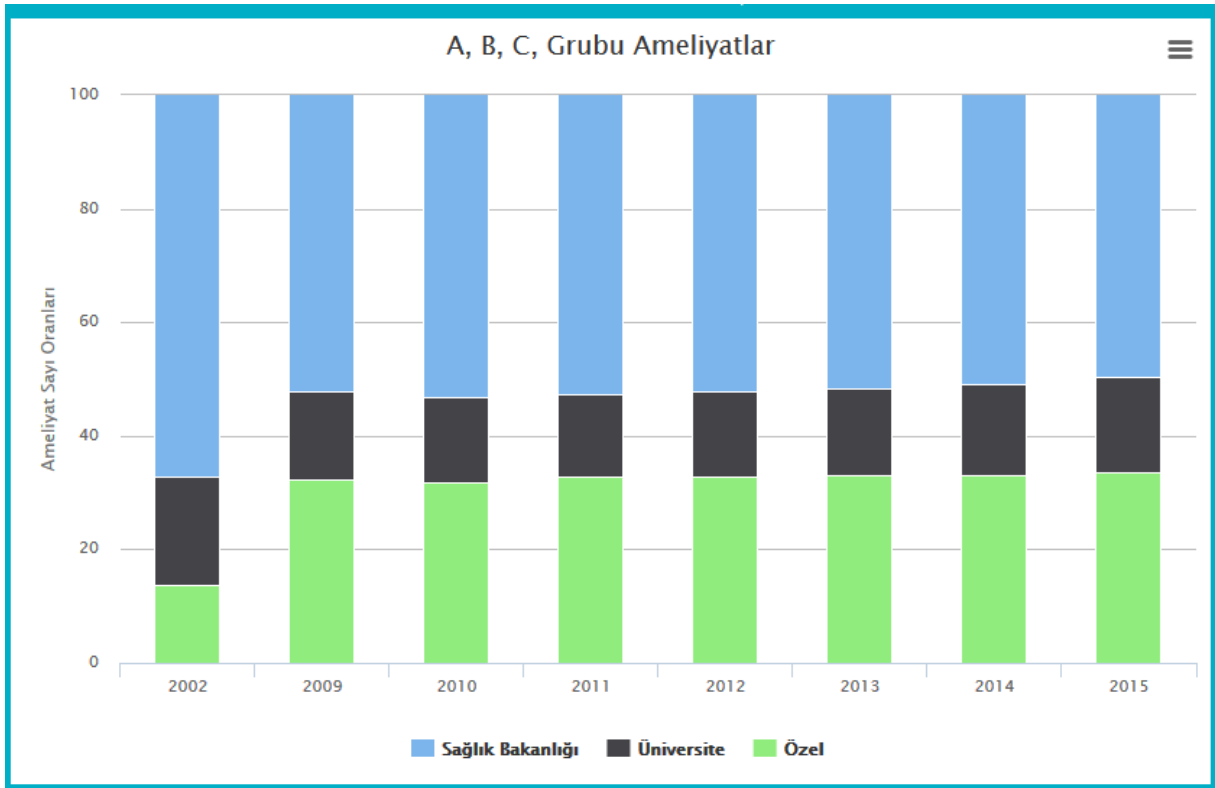


Şekil 2-4 2002-2016 sektörlere göre muayene sayıları [14]

Buna göre, 205 yılında toplam 413.693.722 adet muayene işlemi yapılmıştır. Bu sayının yılın her gününe eşit dağıldığını varsaysak bile günlük 1,13 milyon adet muayene yapıldığını, bu sayının hafta içi günlerde 1,5 milyon civarında olduğunu düşünebiliriz. Her bir muayene işlemi için çeşitli tetkikler ve reçeteler için de ayrıca

provizyon alındığı hesaba katılırsa, günlük olarak sadece muayeneler için 3-5 milyon adet provizyon işlemi yapıldığını varsayabiliriz.

Bankacılık ve telekomünikasyon sistemlerindeki günlük işlem hacmi, sağlık sistemlerinden genellikle çok daha fazla olsa da, işlem başına veri seti büyüklüğü açısından sağlık alanı diğer alanların ilerisindedir. Özellikle ameliyat, doğum, vb. yatışlı tedavi hizmetleri için gönderilen provizyon taleplerinde çok sayıda veri elemanı yer almaktadır. Türkiye'deki A, B ve C grubu ameliyat sayıları Şekil 2-5'te gösterilmektedir. Buna göre,2015 yılındaki toplam ameliyat sayısı 4.770.145 adettir. Hafta sonları dâhil ortalama günlük taburcu sayısı ise 13.069'dur.



Şekil 2-5 2002-2015 A, B, C Grubu ameliyat sayıları [15]

Bütün bu sayılar, günlük işlem hacminin ve provizyon için akan veri miktarının oldukça yüksek olduğunu göstermek için yeterlidir. Bu hacimdeki bir veride, klasik veri madenciliği yaklaşımları ile anlık olarak sahtecilik analizi yapmak bir hayli zordur ve bu durum, bizleri büyük veri bilimini kullanmaya sevk etmektedir.

2.6. Veri Büyüklüğü ve Çeşitliliği

Bir önceki başlıkta belirtildiği üzere sağlık geri ödeme sistemlerindeki veri akışı oldukça yüksektir. Ancak klasik veri madenciliği uygulamalarını kullanmayı zorlaştıran unsurlar arasında, verinin toplam büyüklüğü ve içerdiği veri türü çeşitliliği de sayılabilir. Örneğin, SGK'nın MEDULA sisteminin operasyonel veri tabanının petabaytlar mertebesinde olduğu bilinmektedir. Modern analiz çalışmaları, operasyonel veri tabanları (OLTP) üzerinde analitik işlemlerin yapılmasını doğru görmemektedir. Bunun yerine ayrı bir analitik veri ambarının oluşturulması ve analitik işlemlerin (OLAP) bu veri tabanında yapılması önerilmektedir. Ancak MEDULA sistemindeki bu büyük veri hacmi, sadece klasik veri madenciliği yöntemleri için değil; veri analitiği işlemleri için dahi oldukça büyüktür. Bu nedenle yapılacak analiz öncesinde nitelik seçme, veri temizleme, veri ön işleme ve veri dönüştürme gibi işlemlerin performans ihtiyacı gözetilerek yapılması zaruridir.

3 Dünyadaki Çalışmalar

Dünyada genel sahtecilik ve suiistimal tespitine dair çalışmalar oldukça fazladır. Buna karşın, sağlık sigortacılığı alanındaki çalışmalar, diğer alanlarda yapılan çalışmalara göre daha azdır. Bunun temel nedeni olarak, bu alanda yapılan çalışmalarda da belirtildiği üzere, analiz için bütünsel ve yapısal veriye ulaşmadaki zorluklar ve yukarıda ifade etmiş olduğumuz gibi problemin doğası kaynaklı karmaşıklık seviyesi olsa gerektir.

Bu çerçevede, Phua, Lee, Smith ve Gayler tarafından 2005 yılında yapılan literatür taramasında [16], veri madenciliği ile sahtecilik tespiti yapan 51 çalışma incelenmiştir. Bu 51 çalışmadan sadece 14 tanesi sigortacılıkla ilgili olup, bunlardan da sadece 5 tanesi sağlık sigortacılığı alanındadır. Yazarlar, inceledikleri çalışmalardaki araştırmacıların çoğunun veri yetersizliğinden ve daha önce uygulanmış yöntemlerin azlığından şikâyet ettiklerini belirtirler. Öyle ki, incelenen bu 51 çalışmanın sadece 7 tanesi (bunlardan 2 tanesi sigortacılık sektöründedir) gerçekten uygulamaya alınabilmiştir. Bu uygulamaya geçirilen çalışmalar arasında maalesef sağlık sigortacılığında uygulanan bir örnek yoktur.

2011 yılında yapılan başka bir literatür taramasında ise, 5 tanesi daha önce Phua ve arkadaşları tarafından da incelenmiş olan ve sadece veri madenciliği ile finans sektöründeki sahtecilikleri tespitte çalışan toplam 49 araştırma incelenmiştir [17].

Son olarak, bu alanda yapılan en yeni literatür taramasında ise sadece sağlık alanındaki sahtecilik tespitine yönelik 21 çalışma incelenmiş [16] ve bu çalışmalar kullandıkları yöntemlere göre öğreticili öğrenme, öğreticisiz öğrenme ve hibrit yöntemler kullananlar şeklinde sınıflandırılmıştır.

Bu alanda yapılan ve öne çıkan bazı çalışmalarla ilgili özet bilgileri, Tablo 3-1’de gösterilmektedir.

Tablo 3-1 Literatürdeki bazı çalışmalara dair özet bilgiler (Tablo’nun 2015 öncesi yapılan yayınlar kısmı [18]’den alınmıştır).

Çalışma	Aksiyon Türü	Yaklaşım	Aktör ve işlemler	Nitelik Sayısı	Yöntem	Mimari
He et. al. (1997) [5]	Reaktif	Aktör tabanlı	Doktor	28	MLP, SOM ve MLP ile SOM	28-15-4 nörondan oluşan çok katmanlı algılayıcılar (ÇKA) ve ÇKA ile öz düzenlemeli harita (ÖDH)
Williams (1999)[6]	Reaktif	Aktör tabanlı	Sigortalı	50	Hot Spots	K-means kümeleme yöntemini de içeren 3 aşamalı anomali tespit çerçeve uygulaması
Yamanishi et. al. (2004) [7]	Proaktif	Aktör tabanlı	Patoloji hizmet sunucu	41 (4 tanesi kullanılmış)	İstatistiksel öğrenme	SmartSifter adı verilen ve olasılık hesabına dayalı özel geliştirilmiş bir yöntem
Major and Riedinger (2002)[8]	Reaktif	Aktör tabanlı	Hastane	27	Makine öğrenmesi	Geliştirme ve operasyonel aşamalar için ayrı ayrı geliştirilmiş kural tabanlı model
Ortega et. al. (2006) [10]	Proaktif	İşlem tabanlı	Sağlık raporu	125	2-katmanlı yapay sinir ağı sınıflayıcısı	Çok katmanlı ve ileri beslemeli YSA ve girdi olarak aktör ve provizyon veri tabanlarından elde edilen 4 alt model
Yang and Hwang (2006) [13]	Reaktif	Klinik kılavuz kuralları	PID Hastalığı (Jinekoloji)	NA	C4.5 sınıflayıcı	Klinik kılavuzlara dayalı çizgeler referans alınarak işlemlerin sınıflanması
Sokol et. al. (2001) [11]	Reaktif	İşlem tabanlı	NA	NA	Görsel analiz	Veri temizleme, analiz ve görselleştirme aşamalarından oluşan bir model
Aral K.D. et al. (2012)[12]	Proaktif	İşlem tabanlı	NA	6	Öğreticili öğrenme	Nitelikler arası uzaklığa dayalı risk fonksiyonu tabanlı model
Köse, et. al. (2015) [4]	Proaktif	Aktör-Meta İlişkisi Tabanlı	Tüm aktörler ve tüm işlemler	138	Risk skoru temelli anomali tespiti	İki aşamalı veri ambarı içerisinde, aktör-meta arasındaki ilişkileri temsil eden nitelikler üzerinden, her bir sahtecilik türü için ayrı risk hesaplamasına dayalı
Johnson M.E. et al. (2016) [9]	Reaktif	Aktör tabanlı	KBB, nöroloji, göz uzmanı ve pratisyen hekim	NA	Yoğunluk ve uzaklık analizi	Bağıl yoğunluk oranı ve uzaklığa dayalı altı aşamalı uygulama çerçevesi

İleride üzerinde daha detaylı duracağımız Köse vd. [4] dışında Tablo 3.1'de yer alan diğer çalışmalar ele alındığında karşılaştıkları zorluklar, temel eksiklikler veya iyileştirilmesi gereken noktaları şu şekilde sıralayabiliriz [18]:

- Önerilen modellerin çalışma süreleri oldukça yüksektir. Bu nedenle olsa gerektir ki, söz konusu 8 çalışmadan sadece 2 tanesi sahtecilik işlemlerini proaktif (on-line) olarak tespit edebilmektedir.
- Problemi bütün olarak ele almanın zorlukları nedeniyle önerilen modeller genellikle sadece bir aktör (hasta, eczacı, doktor, vb.) veya işlemi incelemeye tâbi tutulabilmektedir. Buna karşın sahteciliğin doğasında çok aktörlü ve farklı provizyonlara nüfuz etme davranışı söz konusudur. Dolayısıyla analizlerde bu unsurları da dikkate almak gereklidir.
- Tüm aktörleri ve sağlık hizmetlerini kapsayan herhangi bir model öne sürülemediği.
- İstatistiksel anomaliyi inceleyen yöntemler dışında, sadece bir sahtecilik veya suistimal davranış türü ele alınmış ve incelenmiştir.
- Yöntemlerin hiçbirinde, aktörler ile geri ödeme talebi yapılan ilaç ve sağlık hizmetleri (metâlar) arasındaki ilişki analiz edilmemiştir.
- Çalışmaların sonuçları ve çıktılarının gerçek hayat uygulamasında edindiği yerle ilgili bilgiler oldukça kısıtlıdır.
- Önerilen modellerin ne derece başarılı olduğu ya hiç ölçülmemiş; ya da kısıtlı ölçeklerle (sadece doğruluk, sadece AUC, vb.) bir değerlendirme yapılmıştır. Ayrıca herhangi bir modelin, maliyet etkinliği ile ilgili bir analize de rastlanmamıştır.

4 Türkiye'deki Çalışmalar

Geri ödeme sistemlerinde sahtecilik alanında ülkemizdeki çalışmaların sayısı yeterli değildir. Literatürde Türkiye kaynaklı 2012 ve 2015 yıllarına ait iki makale mevcuttur. Bu çalışmalardan ilki, Aral, Güvenir, Sabuncuoğlu ve Akar tarafından yapılmıştır [12]. Bu çalışmada, sahtecilik türleri ayrıştırılarak her biri için ayrı analizler yerine, belirli bir hizmet türü (reçete) üzerinde odaklanılmıştır. Araştırmada, ilaç adı, fiyat, reçete numarası, yaş, cinsiyet ve tanı nitelikleri arasında, ilaç-tanı, ilaç-yaş, ilaç-cinsiyet,

ilaç-ilaç ve ilaç-maliyet şeklinde 5 adet nitelik çifti belirlenmiştir. Bu niteliklerin her bir çifti arasında korelasyon ve karşılıklık matrisleri (*corresponding incidence matrice*) hesaplanmıştır. Daha sonra bu matrislerden de risk matrisleri oluşturulmuştur. Elde edilen risk matrisi, provizyon talebi yapılan reçetelerin her birinin içerdiği ilaçlara göre risk hesaplamasında kullanılmıştır. Çalışmanın başarısı, ROC eğrisi ile ve buradan hesaplanan AUC (Area Under Curve) değeri ile ölçülmüş ve AUC değeri %85,7 olarak hesaplanmıştır. Bu hesaplamaların yapılabilmesi sistemin anomali olarak tespit ettiği reçeteler, uzmanların değerlendirmesine de sunulmuş ve neticeye göre doğruluk hesaplanmıştır [18].

Bu çalışmada reçetelerdeki ilaç-tanı, ilaç-yaş, ilaç-cinsiyet, ilaç-ilaç ve ilaç-maliyet gibi ikili değerlerin korelasyonu esas alındığı için, birlikte bulunma oranları çok düşük olan iki unsurun bir araya gelmesi anomali olarak değerlendirilmiştir. Bu tür reçeteler, az görülmeleri açısından ilgi çekici olsa ve hatta uzmanlar açısından da garipsenecek bir reçete olsa da, hastanın diğer sağlık bilgisine sahip olmadan bu tür reçetelerin sahtecilik olarak değerlendirilmesi, yukarıda anlatıldığı üzere problemin doğasına aykırıdır. Nitekim bu tür garip görülen birlikteliklerin reçeteyi yazan hekim tarafından tıbbi bir açıklamasının olması olasıdır. Bu yönüyle değerlendirdiğimizde bu çalışmanın ortaya özgün bir yöntem koymakla birlikte, farklı sahtecilik türlerini kapsadığını ve sağlık ve sigortacılık alan bilgisine uygun olarak geliştirildiğini söylememiz güçtür.

Türkiye'deki diğer çalışma ise bir TÜBİTAK TEYDEB (3090145) projesi kapsamında hazırlanmış bir doktora tezidir [18]. Özel sağlık sigorta sektöründe faaliyet gösteren ve sigorta şirketleri adına provizyon, tazminat ve saha ziyaretleri faaliyetleri yerine getiren bir şirketin [19] 2008-11 yılları arasında merkezinde yer aldığı ve yaklaşık 850.000 USD bütçeli olan bu proje, sağlık geri ödeme sisteminde sahtecilik tespiti ile ilgili yukarıda bahsedilen zorlukları detaylıca ele almıştır. Çalışmanın sonunda sahtecilik türünden, aktörden ve ödeme talebi yapılan şeylerden (metâ) bağımsız özgün bir model ortaya koyulmaya çalışılmıştır. Söz konusu model, farklı sahtecilik türlerinin tanımlanmasını, bu sahtecilik türlerinde yer alması muhtemel aktörlerin tanımlanmasına imkân sağlamaktadır. Provizyon talebinin riski, o provizyonda yer alan aktörler ile metâlar arasındaki ilişkinin analizi suretiyle yapılmaktadır. Bunun için de aktörler ile metâlar arasındaki ilişkiyi analiz etmek için kullanılacak 52 tane, aktörlerin genel riskleri için de 86 tane olmak üzere toplam 138 tane nitelik

belirlenmiştir. Çalışmada geliştirilen iki aşamalı veri ambarı sayesinde, aktörlerin daha önceki tüm provizyonları ve bu provizyonlardaki metâlarla olan ilişkilerine dair nitelik değerleri risk hesaplamasında kullanılmak üzere hesaplanmış durumdadır. Bu sayede provizyon talebi ulaşır ulaşmaz, proaktif olarak provizyondaki aktörlerin genel riskleri ve o provizyondaki metâlarla olan ilişkileri açısından riskler hesaplanabilmektedir. Bu riskler, daha önce tanımı yapılmış her bir sahtecilik yöntemi için ayrı ayrı yapıldığı için, provizyonlar, her sahtecilik türü için farklı risklere sahip olabilmektedir. Bu sayede provizyonu değerlendirecek olan uzmanların karar vermesi de kolaylaşmaktadır. Dahası, çalışmada geliştirilen model, hesaplanan risk değerinin gerekçelerini de, hesaplamada kullanılan niteliklere ait grafiklerle teyit ettiği için uzmanların hesaplanan riski değerlendirmesi veriye dayalı bir şekilde yürütülebilmektedir.

Söz konusu çalışmanın başarısı, tanımlanmış 6 farklı sahtecilik türü için ayrı ayrı hesaplan ve Tablo 4-1'deki sonuçlar elde edilmiştir.

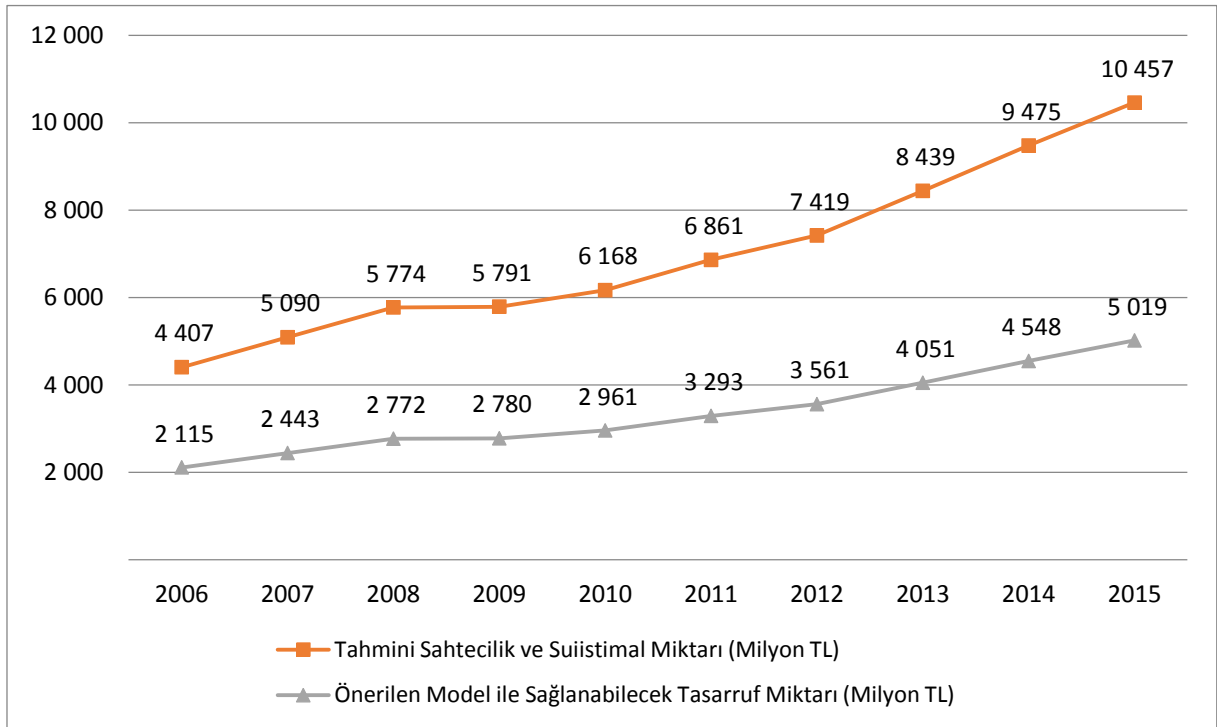
Tablo 4-1 Sahtecilik yöntemlerine göre doğruluk ve AUC oranları [18]

Anormal Davranış	Tahmin	Gerçek		Doğruluk	AUC
		Pozitif	Negatif		
A	Pozitif	89.3%	35.0%	79.2%	82.1%
	Negatif	10.7%	65.0%		
B	Pozitif	93.5%	17.6%	89,6%	92.2%
	Negatif	6.5%	82.4%		
C	Pozitif	91.7%	41.7%	75.00%	83.3%
	Negatif	8.3%	58.3%		
D	Pozitif	90.9%	46.2%	70.80%	75.2%
	Negatif	9.1%	53.8%		
E	Pozitif	89.3%	35.0%	79.20%	87.1%
	Negatif	10.7%	65.0%		
F	Pozitif	88.5%	40.9%	75.00%	86.4%
	Negatif	11.5%	59.1%		

Çalışmada, önerilen modelin maliyet etkinliği de başka bir sahtecilik tespit çalışmasında [20] öne sürülen bir modele göre hesaplanmıştır. Buna göre, mühendis ve hekimlerin bu modeli kullanırken ve modelin riskli olarak değerlendirdiği provizyonların teyit edilmesi sırasında harcadıkları eforun maliyeti de hesaplanmıştır. Diğer taraftan da model sayesinde yakalanan sahtecilik vakalarının maliyeti ve yakalanamayanların kaybı, Denizli ilinde yapılmış bir ortalama maliyeti çalışması [21] esas alınarak dikkate alınarak hesaplanmıştır. Reçete sahtecilik oranının %10 kabul

edilmesi halinde ve önerilen modelin Türkiye'deki insan kaynağı ve reçete maliyetleri dikkate alındığında, söz konusu modelin %4,3'lük bir tasarruf sağlayacağı tespit edilmiştir. Doğruluk oranı %100 olan ideal bir sistemin ise, rutin giderler nedeniyle en fazla %8,7'lik bir tasarruf sağlayacağı hesaplanmıştır.

Önerilen modelin, Türkiye'nin 2006-2015 yılları arasındaki sağlık harcamalarına yansıtılması durumunda (sahtecilik oranı %10 kabul edilmiştir), sağlanabilecek tasarruf Şekil 4-1'de temsil edilmektedir.



Şekil 4-1 Tahmini sahtecilik maliyeti ve önerilen model ile tasarruf oranı

Bu model ile 2006-2015 yılları arasında elde edilebilecek toplam tasarruf yaklaşık 33,5 Milyon TL civarındadır.

5 Sonuç

Gerek bilimsel yazında yer alan çalışmalar, gerekse kamu ya da özel sigortacılık şirketlerinin hayata geçirmiş olduğu projeler her ne kadar suiistimal ve sahtecilik tespitine yönelik veriye dayanan önemli girişimler olarak değerlendirmek mümkünse de, henüz bu konuda yapılması gereken çok iş olduğu açıktır. Bu devasa probleme

özüm için oluşturulması gereken proje grubunun içerisinde sađlık bilimcilerin, sigorta uzmanlarının, veri bilimcilerin, yazılımcıların, hukukçuların vb. farklı disiplinlerden çok sayıda uzmanın bulunması önemli bir gereksinimdir. Problemin karmaşıklığına yol açan ve yukarıda özetlenmiş olan sorunların göz ardı edilmediđi bir çalışmaya, gerek ülkemizin gerekse diğer ülkelerin hem ekonomik gerekçelerle hem de toplumsal adaletin temininin sağlanması için ihtiyacı bulunmaktadır. Yurtdışından ya da özel sektörden temin seçenekleri yerine, kamunun bu problemin aciliyetini ve yarattığı maliyetin büyüklüğünü değerlendirip, kaynak ayırarak kendi içerisinde bu ihtiyacı gidermesi sanırız en gerçekçi yaklaşım olacaktır.

Referanslar

- [1] NHCAA, "Definition of Fraud and Abuse." [Online]. Available: <http://www.nhcaa.org>. [Accessed: 01-May-2015].
- [2] U.S. Federal Bureau of Investigation (FBI), "Financial Crimes Report 2007," 2007.
- [3] TÜİK, "Sağlık Harcamaları İstatistikleri," 2016. [Online]. Available: http://www.tuik.gov.tr/PreTablo.do?alt_id=1084. [Accessed: 14-May-2017].
- [4] I. Kose, M. Gokturk, and K. Kilic, "An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance," *Appl. Soft Comput. J.*, vol. 36, pp. 283–299, 2015.
- [5] H. He, J. Wang, W. Graco, and S. Hawkins, "Application of neural networks to detection of medical fraud," *Expert Syst. Appl.*, vol. 13, no. 4, pp. 329–336, Nov. 1997.
- [6] G. J. Williams, "Evolutionary Hot Spots Data Mining," Springer, Berlin, Heidelberg, 1999, pp. 184–193.
- [7] K. Yamanishi, J.-I. Takeuchi, G. Williams, and P. Milne, "On-line unsupervised outlier detection using finite mixtures with discounting learning algorithms," in *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '00*, 2000, pp. 320–324.
- [8] J. A. Major and D. R. Riedinger, "EFD: A Hybrid Knowledge/Statistical-Based System for the Detection of Fraud," *J. Risk & Insur.*, vol. 69, no. 3, pp. 309–324, Sep. 2002.
- [9] M. E. Johnson and N. Nagarur, "Multi-stage methodology to detect health insurance claim fraud," *Health Care Manag. Sci.*, vol. 19, no. 3, pp. 249–260, Sep. 2016.
- [10] P. A. Ortega, C. J. Figueroa, and G. A. Ruz, "A Medical Claim Fraud/Abuse Detection System based on Data Mining: A Case Study in Chile," in *The International Conference on Data Mining*, , 2006, pp. 224–231.

- [11] L. Sokol, B. Garcia, J. Rodriguez, M. West, and K. Johnson, "Using data mining to find fraud in HCFA health care claims.," *Top. Health Inf. Manage.*, vol. 22, no. 1, pp. 1–13, Aug. 2001.
- [12] K. D. Aral, H. A. Güvenir, İ. Sabuncuoğlu, and A. R. Akar, "A prescription fraud detection model," *Comput. Methods Programs Biomed.*, vol. 106, no. 1, pp. 37–46, 2012.
- [13] W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Syst. Appl.*, vol. 31, no. 1, pp. 56–68, 2006.
- [14] Türkiye Kamu Hastaneleri Kurumu, "2002-2015 Sektörlere Göre Muayene Sayıları." [Online]. Available: <http://rapor.saglik.gov.tr/istatistik/rapor/>. [Accessed: 15-May-2017].
- [15] Türkiye Kamu Hastaneleri Kurumu, "2002-2005 A, B, C Grubu Toplam Ameliyatlar," 2017. [Online]. Available: <http://rapor.saglik.gov.tr/istatistik/rapor/>. [Accessed: 15-May-2017].
- [16] C. Phua, V. Lee, K. Smith, and R. Gayler, "A Comprehensive Survey of Data Mining-based Fraud Detection Research," 2005.
- [17] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decis. Support Syst.*, vol. 50, no. 3, pp. 559–569, 2011.
- [18] I. Kose, "İnteraktif makine öğrenmesi kullanılarak, aktör-metâ ilişkisinin analizi ile sağlık geri ödeme sistemlerinde proaktif suistimal tespiti modeli," Gebze Technic University, 2015.
- [19] CGM, "CompuGroup Medical Türkiye (Tepe Teknoloji)." [Online]. Available: <https://www.cgm.com/tr/index.tr.jsp>. [Accessed: 15-May-2017].
- [20] C. Phua, D. Alahakoon, and V. Lee, "Minority report in fraud detection," *ACM SIGKDD Explor. Newsl.*, vol. 6, no. 1, p. 50, Jun. 2004.
- [21] S. Mollahaliloglu, A. Alkan, B. Dönertaş, Ş. Özgülcü, and A. Akıcı, "Assessment of the Prescriptions Written in Different Provinces of Turkey in Terms of Drug

Utilization Principles," *Marmara Med. J.*, vol. 24, no. 3, pp. 162–173, 2011.