Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği ile Sahtekârlık Analizi

Canan Özcan

112692019

İSTANBUL BİLGİ ÜNİVERSİTESİ

SOSYAL BİLİMLER ENSTİTÜSÜ

BİLİŞİM VE TEKNOLOJİ HUKUKU YÜKSEK LİSANS PROGRAMI

Yard.Doç. Dr. Leyla KESER BERBER

2014

Veri Madenciliğinin Güvenlik Uygulama Alanları ve

Veri Madenciliği ile Sahtekârlık Analizi

Security Application Areas in Data Mining and

Fraud Analysis through Data Mining

Canan Özcan

112692019

Prof. Dr. Ahmet DENKER

Yrd. Doç. Dr. Leyla KESER BERBER :

Arş. Görv. Bülent Onsoy :

Tezin Onaylandığı Tarih :

Toplam Sayfa Sayısı	: 89

Anahtar Kelimeler (Türkçe) (İngilizce)	Anahtar Kelimeler
1) Veri Madenciliği	1) Data Mining
2) Veri tabanlarında Bilgi Keşfi	2) Discovery in databases
3) Sahtekârlık Analizi	3) Fraud Analysis
4) Veri Madenciliği ve Hukuk	4) Data Mining and the Law
5) Veri Madenciliği Güvenlik Alanları	5) Security Areas in Data Mining

ÖNSÖZ

Ekonomik, sosyal, teknolojik, kültürel, politik ve ekolojik dengeler açısından, küresel bütünleşme, entegrasyon ve dayanışmanın artması anlamına gelen global dünya düzeninin rekabet sisteminde; bilginin gerçek bir güç olduğu anlayışı ile geleceği yönetmek adına yetişmiş insan gücüne yapılan yatırım geri dönüşü ve katma değeri en yüksek yatırımdır. Ülkemizde, hala geçerli olan ve talebin artırılması esasına dayalı olarak yürütülen rant ekonomilerinin oluşturmakta olduğu katma değerden yüzlerce kat fazlası, gelişmiş ülkelerdeki rekabetçi firmalarda 'entelektüel sermaye' yani yetişmiş insan gücü sayesinde sağlanabilmektedir.

Bunu görüp ilerici bakış açıları ile ülke gelişmişliğine katkı sağlama amacını taşıyan bilim insanlarının, yeni olanaklar ve gelişmeleri ülke sınırları içindeki bireylerin gelişmesi amacına hizmet edecek, fikir fabrikalarına dönüştürme çabaları taşıyan düşünceleri, Üniversite camiasında hayat bularak yeni ufukların açılmasını sağlamaktadır.

Bu hedeflerle yola çıktığına inandığım ve Ülkemizde bilişim hukuku anlayışının yerleşmesinden, bilişim ve teknoloji hukuku alanında yapılması hedeflenen düzenlemeler ve gelişmiş ülkelerdeki mevcut standartlarla karşılaştırmalı hukuk bakışına, söylemleri ve anlatımları ile katkı sağlayan, sürekli toplumsal bilinç oluşturma çabasında olan Hocam ve tez danışmanım; Sn. Yard. Doç. Dr. Leyla Keser Berber'e bu tezi yapma olanağı sağladığı ve verdiği tüm destekler için sonsuz teşekkürlerimi sunarım.

Canan Özcan

Mayıs 2014

ÖZET

Tarih ders almayanlar için tekerrürden ibarettir sözünün bilişim dünyasında hayat bulmuş şeklidir veri madenciliği. Geçmiş veriden çıkarılan kurallar, öğrenilenler ve alınan dersler, gelecekte de geçerli olacak, ilerisi için doğru tahmin yapılmasına, olumsuzlukların baştan öngörülerek önlem alınmasına olanak sağlayacaktır.

Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranması sürecidir.

1995 yılında ilk defa düzenlenen Knowledge Discovery in Databeses konferansındaki bildiri kitabının ön sözünde şöyle bir paragraf bulunmaktadır ([1]). "Dünyadaki enformasyon miktarının her 20 ayda ikiye katlandığı tahmin edilmektedir. Bu ham veri seli ile ne yapmamız gerekmektedir. İnsan gözü bunun ancak çok küçük bir kısmını görebilir. Bilgisayarlar, bilgelik pınarı olmayı vaat ederken, veri sellerine neden olmaktadırlar."

1995 yılındaki bu tespitlerin temeli aslında çok daha eskiye dayanmaktadır. Yıllar itibariyle her alanda birçok kayıt tutulmuştur. 1960'lı yıllardan itibaren bu kayıtlar artarak veri tabanları ve veri ambarlarında depolanmaya başlanmıştır. Tutulan kayıtların veri tabanlarındaki yükü artmakta, fakat bunlardan yararlanma oranının artışı, yükün getirdiği artış ile paralellik göstermemektedir. Sürekli artan verinin içinde gerekli, gereksiz ve tutarsız verilerin bulunması, sistemleri hantallaştırmıştır. Mevcut kayıtların belli amaçlara hizmet edebilmesi için, bu kayıtlar veri olmaktan çıkıp bilgi düzeyine gelmelidir.

1990'lı yıllardan itibaren İnternetin hayatımıza girmesi ile veri miktarları katlanarak artmış, bu verilerden bilgiye dönüşüm sürecini hızlandırma ve karar vericilere doğru bilgi aktarımı önem kazanır olmuştur. Teknolojinin sürekli güncellenmesi beraberinde güvenlik ihlalleri ve adli bilişim altında birçok yeni

[1] http://www.cs.uiuc.edu/~hanj/pdf/sigmod96_tuto.pdf
https://www.aaai.org/Conferences/KDD/kdd95.php

kavramı da hayatımıza sokmuştur. Bundan on, on beş yıl öncesine kadar adını duymadığımız adli muhasebecilik, adli denetim, sahtekârlık incelemeleri gibi birçok iş alanı ve meslek birliği örgütleri dünya üzerinde yerini almıştır. Tüm bunların temelinde artan sahtecilik ve sahtekârlık olayları bulunmaktadır.

Bu bağlamda veri madenciliği bu alanlarda işi yapanları destekleyici, karar vericilere kolaylık sağlayan, sorunlar oluşmadan önlem almada proaktif olmaya olanak sağlayan yapısı ile bilgi ve iletişim teknolojileri alanında yerini almış, güvenirliliği kanıtlanmış, birçok bilim dallarından yararlanan bir iş yapma modeli olarak karşımıza çıkmaktadır.

"Veri Madenciliğinin Güvenlik Uygulama Alanları ve Veri Madenciliği ile Sahtekârlık Analizi" tezi 5 bölümden oluşmaktadır.

Birinci bölümde, veri madenciliği genel tanım ve kavramları üzerinde durulmuş, veriden bilgi kavramına, büyük veriden veri madenciliğine geçiş süreçleri ele alınmıştır.

İkinci bölüm, veri madenciliği süreci ve algoritmalarının incelendiği, sürecin metodolojik yaklaşımının ortaya konulduğu kısımdır.

Üçüncü bölümde, Veri madenciliğinin gelişim süreci, kavramların hangi aşamalardan geçerek şekil bulduğu, Dünya'da ve Türkiye'de veri madenciliğine yönelik yapılan araştırmaların literatür taramaları verilmiştir.

Dördüncü bölümde; veri madenciliği güvenlik uygulamaları açısından incelenmiş, hile, dolandırıcılık ve sahtekârlık kavramları üzerinde durulmuş, konu hukuki boyutları ile ele alınmış, sahtekârlıklarda çalışanın rolüne değinilmiş, bu konudaki standartlar incelenmiş ve sağlık sigortası sektöründe gerçek verilerden hareketle sahtekârlık tespitlerinin bulunarak öğrenici model uygulamasının kullanıldığı bir çalışmanın incelemesi yapılmıştır.

Sonuç bölümünde ise genel bir değerlendirme yapılmakta ve veri madenciliğine hakettiği önemin verilmesi üzerinde durulmaktadır.

EXTRACT

Data mining could be seen as the true expression in the world of informatics of the phrase "History is only made up of repetitions for the people who could not learn in life". The rules developed from past data, lessons learnt and things understood may be helpful in the future to make accurate projections and to act proactively in order to take measures in advance after understanding the negativities.

The data mining entails the process of searching for the correlations and rules in a large quantity of data in order to make accurate projections for the future and this search is made with the use computer programs.

Following paragraph was given on the front side of the declaration book developed after the conference named "Knowledge Discovery in Databases" that was organized for the first time in 1995 ([2]). "It is estimated that the quantity of global information doubles in every 20 months. What shall we do with this raw data flow? Human eye can see only a very small part of it. Although they promise to serve as a fountain of wisdom, the computers cause data floods."

The basis of these determinations made in 1995 is actually related to much older times. Too many records were kept in all fields every year. These records increased considerably after 1960s and started to be stored at the databases and data warehouses. The load imposed to the database by the records kept also increased but the increase in the ratio benefiting from such data does not run parallel to the increase in the data load. Existence of necessary, unnecessary and inconsistent data in the ever increasing data load made the systems clumsy. However available records should be transformed from data into information if we want them to serve for specific purposes.

The data quantities increased more quickly after Internet was launched and became part of our life in 1990s and it became important to accelerate the process of

[2] http://www.cs.uiuc.edu/~hanj/pdf/sigmod96_tuto.pdf
https://www.aaai.org/Conferences/KDD/kdd95.php

transforming that data into information and to impart accurate information to the decision makers. Continuous updating of technology put into our life security breaches and many new concepts including judicial informatics. New professions and professional organizations relating to judicial accounting, judicial audit and fraud investigations gained recognition at a global level although we had not heard of them at all only ten or fifteen years ago. These professions and organizations were launched to meet the requirements arising from increasing fraud and fraudulent actions.

In that regard data mining could be viewed as a business model that is already adopted in the area of information and communication technologies due to its structure that supports the people working in these fields, providing facilities to the decision makers and allowing for proactive attempts in taking measures before the problems arise and it has proven itself as a safe and reliable method that uses the information from too many branches of science.

The thesis "Security Application Areas in the Data Mining and Fraud Analysis through Data Mining" comprises of 5 sections.

In the first section, I explained the general definitions and concepts of data mining and the processes of transition from data to information and from large data to the data mining were addressed.

The second section examines the data mining process and algorithms and presents the methodological approach of the process.

The third section presents the development process of the data mining and advances through the phases of development toward the final concepts and also gives the literature scans relating to the researches made about the data mining in the world and in Turkey.

The fourth section examines the data mining in terms of security applications with emphasis on the concepts of trickery, swindling and fraud. The legal aspects of these concepts were also detailed and employee's role in fraudulent acts was investigated. Finally relevant standards were examined and a study involving the

use of a teaching model and determining the acts of fraud in the health insurance sector by using real data was examined.

In the conclusion section, general evaluation was made and the need to give the deserved importance to the data mining was addressed.

İÇİNDEKİLER

KAYNAKÇA

1. http://blog.data-miners.com/2011/04/data-mining-techniques-3rd-edition.html
2. http://www.teknologweb.com/veri-tabani-nedir/
3. http://www.cagatayyurtoz.com/drupal-7.0/content/k%C4%B1saca-veri-ambar%C4%B1-olap-etl-data-mart-nedir
4. Gordon S. Linoff, Michael J. A. Berry "Data Mining Techniques: For Marketing, Sales, And Customer Relationship Management" Chapter 1 "What Is Data Mining and Why Do It?"
5. http://www.data-miners.com/companion/BDMT_notes.pdf (erişim 01.12.2013)
6. http://binedir.com/blogs/veri-madenciligi/archive/2012/06/01/veri-madencili-i-s-recini-ve-amac-n-daha-yi-anlamaya-y-nelik-farkl-bir-bak-a-s.aspx
7. http://docs.oracle.com/html/B14339_01/1intro.htm
8. Harness the Power of Big Data IBM Big data Platform,
9. Lori Bowen Ayre, "Data Mining for Information Professionals" June 2006 http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.2789&rep=rep1&type=pdf (erişim tarihi 05.01.2014)
10. Prof. Boris Kovalerchuk, Central Washington University "Advanced data mining, link discovery and visual correlation for data and image analysis"
11. Pete Chapman (NCR), Julian Clinton (SPSS), RandyKerber (NCR), ThomasKhabaza (SPSS), ThomasReinartz(DaimlerChrysler), Colin Shearer (SPSS) and Rüdiger Wirth (DaimlerChrysler) Step-by-step data mining guide CRISP-DM 1.0 ftp://ftp.software.ibm.com/software/analytics/spss/support/Modeler/Documentation/14/UserManual/CRISP-DM.pdf (erişim tarihi 11.02.2014)
12. Yrd. Doç. Dr. Cengiz Aktaş, Eskişehir Osman Gazi Ünv. Fen - Ed . Fak İstatistik Bö l. "Lojistik Regresyon Analiz : Öğrencilerin Sigara İçme Alışkanlığı Üzerine Bir Uygulama"
13. Big Data Analytics – Your Project Reality: The Good, the Bad, and the Ugly IBM Docs ; http://www.hailpern.com/bthpub/papers/ISJ_45_3_2006.pdf (erişim tarihi 07.01.2014)

14. https://chapters.theiia.org/topeka/ChapterDocuments/Fraud%20and%20Advanced%20Data%20Mining.pdf
15. http://www.liacs.nl/~bvstrien/stuva_files/slides2811.pdf
16. Doç. Dr. Şule Özmen, Marmara Ünv. İ.İ.B.F. İng. İşletme " İş Hayatı Veri madenciliği İle İstatistik Uygulamalarını Yeniden Keşfediyor." http://idari.cu.edu.tr/sempozyum/bil38.htm (erişim tarihi:11.02.2014)
17. http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf (erişim tarihi : 14.02.2014)
18. Dr. Yılmaz Argüden, Veri Madenciliği Veriden Bilgiye, Masraftan Değere http://www.arge.com/wp-content/uploads/2013/03/VeriMadenciligi.pdf (erişim tarihi 17.02.2014)
19. http://www.leylakeser.org/search/label/Data%20Mining (erişim tarihi:11.02.2014) http://www.forbes.com/sites/andygreenberg/2013/08/14/agent-of-intelligence-how-a-deviant-philosopher-built-palantir-a-cia-funded-data-mining-juggernaut/ (erişim tarihi:11.02.2014) http://www.leylakeser.org/search/label/Computer%20Fraud%20and%20Abuse%20Act (erişim tarihi:11.02.2014)
20. Ali Serhan Koyuncugil,, Nermin Özgülbaş "Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları http://btd.gazi.edu.tr/index.php/BTD/article/view/36/34 (erişim 01.04.2014)
21. Evren Dilek Şengür," İşletmelerde Hile, Hilelerin Önlenmesi, Hileli Finansal Raporlama İle İlgili Düzenlemeler ve Bir Araştırma" Doktora Tezi İstanbul Üniversitesi Sosyal Bilimler Enstitüsü İşletme Anabilim Dalı Muhasebe Bilim Dalı 2010
22. Doç.Dr. Selim Yüksel Pazarçeviren, Sakarya Ünv. İktisadi ve İdari Bilimler Fakültesi İşletme Bölümü Öğ.Üy. Adli Muhasebecilik Mesleği ZKÜ Sosyal Bilimler Dergisi, Cilt 1, Sayı 2, 2005, s. 1–19
23. Outlier Detection Technique in Data Mining: A Research Perspective M. O. Mansur, Mohd. Noor Md. Sap aculty of Computer Science and Information Systems Universiti Teknologi Malaysia Proceedings of the Postgraduate Annual Research Seminar 2005 http://www.researchgate.net/publication/228722062_Outlier_Detection_Technique_in_Data_Mining_A_Research_Perspective

Şekil Listesi

I. Veri Madenciliği

Büyük veri yığınları denince ilk akla gelen kavram "Veri Madencili"ğidir. Veri madenciliği kavramı William Frawley ve Gregory Piatetsky-Shapiro tarafından "önceden bilinmeyen, fakat yarar sağlama potansiyeli yüksek olan verinin keşfedilmesi" olarak tanımlamıştır.([3]) M. Berry ve G. Linoff ise aynı kavrama([4]) "Anlamlı kuralların ve örüntülerin bulunması için geniş veri yığınları üzerine yapılan keşif ve analiz işlemleridir" şeklinde bir açıklama getirmiştir.([5]) Sever ve Oğuz çalışmalarında([6]) "Önceden bilinmeyen, veri içinde gizli, anlamlı ve yararlı örüntülerin büyük ölçekli veri tabanlarından otomatik biçimde elde edilmesini sağlayan veri tabanlarında bilgi keşfi süreci içerisinde bir adımdır." tanımını kullanmışlardır.([7])

Veri Madenciliğini; büyük ölçekli veriler arasından, istenilen hedefe ulaşmak için verileri analiz etme ve sonuçta bilgiye ulaşma çalışmasıdır şeklinde tanımlayacağız. Bir diğer ifade ile büyük veri yığınları içerisinden gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağıntıların bilişim teknolojilerinin sağladığı olanakları kullanarak, istatistik, matematik, mantık vb. bilim dallarının katkısı ile bulunması işidir. Veri madenciliği bilişim literatüründe; Veritabanlarında bilgi madenciliği (knowledge mining from databases), bilgi çıkarımı (knowledge

---

[3] G. Piatetsky-Shapiro, W. J. Fawley, "Knowledge Discovery in Databases", AAAI/MIT Pres, 1991.

[4] Gordon S. Linoff, Michael J. A. Berry "Data Mining Techniques: For Marketing, Sales, And Customer Relationship Management" Chapter 1 "What Is Data Mining and Why Do It?" http://www.data-miners.com/companion/BDMT_notes.pdf (erişim 01.12.2013)

[5] Sertaç Öğüt "Veri Madenciliği kavramı ve gelişim süreci" Görsel İletişim Tasarımı Bölümü, İletişim Fak. Yeditepe Ünv. http://www.sertacogut.com/blog/wp-content/uploads/2009/03/sertac_ogut_-_veri_madenciligi_kavrami_ve_gelisim_sureci.pdf (erişim tarihi 1.12.2013)

[6] Hayri Sever –Buket Oğuz; "Veri Tabanlarında Bilgi Keşfine Formel Bir Yaklaşım: Eşleştirme Sorgularının Biçimsel Kavram Analizi ile Modellenmesi" http://eprints.rclis.org/7352 (erişim tarihi 01.12.2013)

[7] Hayri Sever, Başkent Ünv. Bilgisayar Müh. Bl, Buket Oğuz.Ayesaş ODTÜ - Teknokent (http://eprints.rclis.org/7348/1/173-204.pdf erişim tarihi: 01.12.2013)

extraction), veri ve örüntü analizi (data/pattern analysis), veri arkeolojisi gibi kavramlarla da açıklanmaya çalışılmıştır.

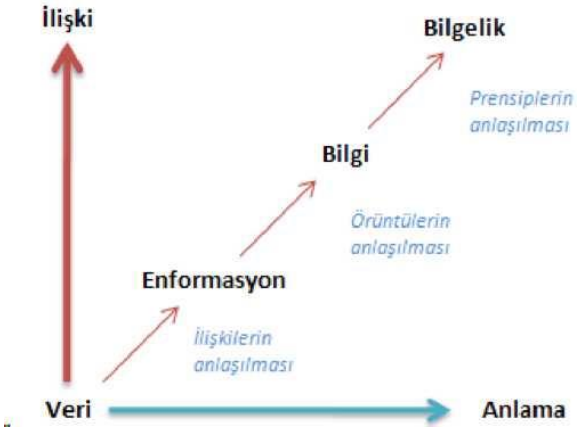
Veriler üzerinde çözümlemeler yapmak ihtiyacından doğan veri madenciliği, sadece bir sorgulama işlemi veya istatistik programlarıyla yapılmış bir çalışma değildir, çok büyük boyuttaki veri yığınları ve çok fazla değişkenin parametrik programlanması ile ilgilenir. Teknolojik gelişmeler dünyada gerçekleşen birçok işlemin elektronik olarak kayıt altına alınmasını, bu kayıtların kolayca saklanabilmesini ve gerektiğinde erişilebilmesini kolaylaştırmakla birlikte, bu işlemlerin her geçen gün daha ucuza mal edilmesini sağlar duruma gelmiştir. Ancak, ilişkisel veri tabanlarında saklanan birçok veriden kararlar için anlamlı çıkarımlar yapabilmek bu verilerin konusunda uzman kişilerce analiz edilmesini gerektirir. Bu amaçla bazı özel analiz algoritmaları geliştirilmiştir. Bunlar; kümeleme, veri özetleme, değişikliklerin analizi, sapmaların tespiti gibi belirli sayıda teknik yaklaşımları içerir. Başka bir deyişle, veri madenciliği, verilerin içerisindeki desenlerin, ilişkilerin, değişimlerin, düzensizliklerin, kuralların ve istatistiksel olarak önemli olan yapıların yarı otomatik olarak keşfedilmesidir. Temel olarak, veri setleri arasındaki desenlerin ya da düzenin, yazılım tekniklerinin kullanılmasıyla ilgilidir. Amaç, daha önceden fark edilmemiş veri desenlerini tespit edebilmektir.

Veri madenciliğini istatistiksel bir yöntemler serisi olarak görmek mümkün olabilir. Ancak geleneksel istatistikten birkaç yönde farklılık gösterir. Veri madenciliğinde amaç, kolaylıkla mantıksal kurallara ya da görsel sunumlara çevrilebilecek nitel modellerin çıkarılmasıdır. Bu bağlamda, insan merkezlidir.

Burada bahsi geçen geniş veri tanımındaki geniş kelimesi, tek bir iş istasyonunun belleğine sığamayacak kadar büyük veri kümelerini ifade etmektedir. Yüksek hacimli veri ise, tek bir iş istasyonundaki ya da bir grup iş istasyonundaki disklere sığamayacak kadar fazla veri anlamındadır. Dağıtık veri ise, farklı coğrafi konumlarda bulunan verileri anlatır. Veri madenciliği, büyük, yüksek hacimli, dağıtık veri grupları ile ilgilenmektedir.

A. Veri, Enformasyon, Bilgi

"Veri", "Enformasyon", Bilgi ve Bilgelik hiyerarşisi konusunda çalışma yapan ilk bilim adamlarından biri Milan Zeleny'dir. 1987 yılında yazdığı "Management Support Systems" adlı makalesinde bu konudaki bilgi yönetimi hiyerarşisini ayrıntıları ile anlatır([8]). Zeleny bilgi hiyerarşisini veri, enformasyon, bilgi ve bilgelik kavramlarını "knownothing, know-what", "know-how", "know-why" ve "know-who" biçiminde kullandığı tanımlar ile açıklar. Sistem kuramcısı ve örgütsel dönüşüm profesörü Russell Ackoff'a ([9])göre ise, "-insan aklının içeriği beş kategori halinde sınıflandırılabilir "der. "Veri, enformasyon, bilgi, anlama ve bilgelik". Ackoff, bilgi ve bilgelik arasına başka bir katmanı, anlamayı eklemiştir. Ackoff, ilk dört kategorinin geçmişle ilişkili olduğunu belirtir; bunlar, neyin olduğu ve neyin bilindiği ile ilgilenirler. Yalnızca beşinci kategori, yani bilgelik, görüş ve tasarımı içinde barındırdığı için gelecekle ilgilenir. Veriden bilgeliğe giden süreç Şekil -1'de gösterilmektedir.([10])



Şekil -1

"Veri" kelimesi, bilişim literatüründe "olgu, kavram veya komutların, iletişim, yorum ve işlem için elverişli biçimli gösterimi", "bir çözüme ulaşmak için işlenebilir duruma getirilmiş gözlemler, ölçümler", "bilgisayar için işlenebilir

[8] http://www.milanzeleny.com/documents/publications/mss.pdf (erişim 03.12.2013)

[9] Russell Ackoff http://www.gperform.com/ackoff_on_adoption_of__systems.pdf (erişim tarihi:05.01.2014)

[10] http://www.behlulcaliskan.com/files/dersler/bilisimin_temel_kavramlari/04.pdf (erişim tarihi 05.01.2014

duruma getirilmiş sayısal ya da sayısal olmayan nicelikler" olarak ifade edilir.([11]) İngilizce dilindeki karşılığı olan "data" ve Almanca dilindeki karşılığı olan "Daten" kelimeleri ise Latince kökenlerinde "verilmiş, verili (şey)" anlamına gelmektedir.([12])

Veri, kelimeler ve sayılar gibi kayıt altına alınmış ama biçimlendirilmemiş enformasyondur. Kendi içinde bir anlamı yoktur, diğer bir ifade ile kendi başına değersizdir.

"Enformasyon" Latince kökeni olan "informatio" taslak, görüş, düşünce anlamına gelmekle birlikte, kelimenin İngilizce dilindeki kökü "inform" (bilgi vermek, haber vermek) kelimesinin Latince kaynağı olan "informare", "şekillendirmek", "biçim vermek", "eğitmek" ve "göstermek" anlamlarına gelmektedir. Türkçe sözlükte "danışma, tanıtma", "haber alma, haber verme, haberleşme" olarak tanımlanırken, bilişim literatüründe "bilgi işlemde kullanılan kabul edilmiş kurallardan yola çıkarak veriye yöneltilen anlam", "bilişim kuramında, birçok olası olay arasında belirli bir olayın meydana gelme belirsizliğini, bilinmezliğini azaltan herhangi bir bilgi", "bilgi ve iletişim teknolojilerinde, verilerden elde edilen herhangi bir kavram, olgu, anlam" olarak çeşitli biçimlerde tanımlanır. Bir karar verme sürecinde, hangi seçeneklerle hangi sonuçlara ulaşılacağını belirlemek için enformasyon gereklidir. Karar ortamlarının giderek değişen ve karmaşık şekil alması, enformasyonun, karar verenin etkinliğinde bir anahtar durumuna gelmesine neden olmaktadır. Bu nedenle, enformasyonun alınması, verilmesi, sevk ve depo edilmesi, yani iletiminin sağlanması, özellikle örgüt yapılarının önemli bir fonksiyonu olmuştur. Bilişim teorisinin Shannon ve Weaver([13]) tarafından ortaya çıkarılmasıyla aynı zaman diliminde, Norbert Wiener'de "Cybernetics or Control and Communication in the Animal and the Machine" (1948) adlı kitabını

---

[11] http://web.deu.edu.tr/tb/index.php?option=com_content&task=blogcategory&id=22&Itemid=35

[12] http://www.behlulcaliskan.com/files/dersler/bilisimin_temel_kavramlari/04.pdf (erişim tarihi 05.01.2014)

[13] Claude Elwood Shannon (1916-2001) – Warren Weaver (1894-1978) http://communicationtheory.org/shannon-and-weaver-model-of-communication/ (erişim tarihi 05.01.2014)

yayınlamıştır. Bu kitabında Wiener, enformasyonun madde ve enerji ile birlikte sistem dünyasının üçüncü temel bileşeni olduğunu söyler. Sibernetikçilere göre "enformasyon, bir sistemin durumunu ve buna bağlı olarak o sistemin diğer bir sisteme ilettiği durumu anlatan nitel bir faktördür." der. Veri, gerçeğin temsili olarak da tanımlanabilecek enformasyona dönüştürülür.([14])

"Bilgi"nin Türkçe sözlükteki anlamı "insan aklının erebileceği olgu, gerçek ve ilkelerin bütünü, malumat", "öğrenme, araştırma veya gözlem yolu ile elde edilen gerçek, vukuf" ve "insan zekâsının çalışması sonucu ortaya çıkan düşünce ürünü"dür. Bilişim literatüründe ise, "kurallardan yararlanarak kişinin veriye yönelttiği anlam", "yapay zekâda bir programın akıllı bir şekilde işlenmesine elveren olaylar, olgular, kurallar ve buluşsal ipuçları" ve "olguların ve bunlardan elde edilen genelleştirmelerin örgütlü bütünü" anlamına gelmektedir. Felsefe alanında ise bilgiden([15]) "insanların maddi ve toplumsal anlıksal etkinliğinin ürünü; insansal ve doğasal dünyadaki nesnel temel özelliklerin ve bağıntıların gösterge biçiminde düşünsel olarak aniden üretilmesi" olarak bahsedilir.

Bilginin enformasyondan daha fazla anlam içerdiğine dair düşünce çoğu yazarı ham veri, enformasyon ve bilgiyi birbirinden ayrı tutma gayretine sürüklemiştir. Genel anlayış açısından bu kavramlar çok basit gibi görünse de, belki de sırf bu basitlik yüzünden bu konu sürekli bir zihin karışıklığına yol açmıştır. Örneğin bazı yazarlara göre veri henüz yorumlanmamış semboller olarak anlaşılmalıdır; enformasyon ise anlam katılmış olan veridir. Bilgi ise anlam katılmış olan veriden yani enformasyondan yorum üretilmesidir. Kimi yazarlar ise verinin dünyanın durumları hakkındaki basit gözlemler, enformasyonun ilinti ve amaçla donatılmış veri ve bilginin de değerli enformasyon olduğunu söyler. Kimilerine göre enformasyon anlamsızdır; ancak yorumlandığında anlam sahibi olur. Kimilerine

---

[14] Bilişim Toplumuna Eleştirel Yaklaşımlar Bağlamında Çevrimiçi Gazete kullanıcılarının Sosyo-Ekonomik Koşulları Ve Etkileşim Ölçümleri Üzerine Bir Araştırma Yük. Lisans Tezi Behlül Çalışkan Marmara Ünv. Sosyal Bilimler Enstitüsü Gazetecilik Anabilim Dalı (2008)

[15] http://www.sevgiagacim.net/felsefe/felsefenin-alani-bilgi-ve-bilgi-cesitleri-t10826.0.html;wap2 Erişim tarihi 03.01.2014 - http://www.frmtr.com/felsefe-sosyoloji-psikoloji/2661252-bilgi-felsefesi.html

göre ise enformasyon özel bir durumu veya koşulu tanımlamak için organize olan olguları ve verileri içerir; bilgi ise gerçekleri ve inançları, bakış açılarını ve kavramları, hükümleri ve beklentileri, yöntemleri ve teknik bilgileri içerir. Yahut enformasyon, anlamlı iletilerin akışı ile başlar, bu iletilerin sonucu olarak taahhüt ve inanç ortaya çıktığında bilgiye dönüşür. "Enformasyonun daha yüksek biçimi" olarak bilginin tüm bu modelleri altında yatan, bilginin ham maddelerinden, onlara anlam eklenerek ortaya çıkarılması gerektiği düşüncesidir. Kaynaktan çıkan iletinin, alıcıda enformasyona dönüştüğü iletişim sürecinde veri, enformasyon ve bilgi kavramları sıkça birbirinin yerine kullanılmıştır. Bunlardan bir kısmı yanlış çeviri sonucu oluştuğu gibi, bir kısmında ise bu birbirlerinden farklı kavramların arasındaki hiyerarşik ilişki ortaya konulmaya çalışılmıştır. Veri, enformasyon, bilgi ve bilgelik kavramlarının anlaşılması, tüm bunların bilişim sistemlerinin çalışılması, geliştirilmesi ve kullanılmasındaki rolü bakımından önemlidir. ([16],[17],[18])

B. Büyük Veriden Veri Madenciliğine

Bilişim teknolojilerinin iş yapma biçimini değiştirmesi, dönüştürmesi, kullanımının yaygınlaşması ve bu teknolojiyi oluşturan bileşenlere herkes tarafından ulaşılabilir olması, üretilen ve işlenen verinin büyüklüğünün de her geçen gün artması durumunu beraberinde getirmiştir.([19]) Sosyal medyanın hayatımıza girmesi ile e-posta, fotoğraf, ses, video dosyaları, İnternet ortamında yapılan paylaşımlar, İnternet dünyasına erişimin kayıtlarının tutulduğu başta GSM operatörleri olmak üzere firmalarda bulunan loglar büyük verilerin bir kısmını oluştururken, şirketlerin ilişkisel veri tabanlarındaki, kurumsal kaynak planlama sistemlerine ait finansal,

---

[16] Şirketlerde e-Dönüşüm: Bilginin Öne Çıkışı Şişecam A.Semih İşevi, Murat Günsur, İbrahim Kavrakoğlu (ÜNAK Ocak 2004, Şirket içi Yayın)

[17] Türkiye Kütüphanecilik Derneği '21.Yüzyıla Girerken Enformasyon Olgusu' Sempozyumu (Baha Kuban Semih İşevi Şişecam, Şirket içi Yayın)

[18] http://www.milanzeleny.com/documents/publications/mss.pdf (2.kez erişim 31.01.2013)

[19] Cüneyt Göksu http://www.bthaber.com/veri-ambari-ve-buyuk-veri-elele-verirse/ (erişim tarihi 15.02.2014)

mali, satış, sipariş, stok, müşteri ilişkileri, talep ve üretim planlama vb. verilerde büyük veri sınıflarının bir parçası haline gelmiştir. Bunlara yine şirketlerde ve/veya tüm kurumlarda üretilen ve çoğunlukla kişisel bilgisayarların disklerinde saklanan yazışma, belge, dokümanlar gibi yapısal olmayan verileri de eklediğimizde karşımıza devasa bir veri yığını çıkmaktadır. International Data Corporation (IDC)'ın raporlarına göre Dünya üzerindeki mevcut verilerin %80'i "Grey data" olarak adlandırılan yapısal olmayan ham verilerdir. Yine IDC istatistiklerine göre artan veri miktarının, 2012'de 2.8 trilyon gigabyte'dan 2020'de 40 trilyon gigabyte'a erişmesi beklenmektedir.([20]) Büyük veriyi tanımlamak için 4V kuralı kullanılmaktadır. ([21]) Bunlar:

Veri hacmi (Volume) : Miktarı, terabyte veya petabyte seviyesinde ölçümlenen veriler,

Veri hızı (Velocity) : Verinin büyüme hızı onun büyük veri olup olmadığını tanımlayan bir diğer boyuttur. Eğer veri ne kadar hızlı artıyor ve değişikliğe uğruyorsa, o kadar büyük bir yığın oluşturmaya adaydır. GSM Operatörlerindeki veri buna en iyi örnektir.

Veri çeşitliliği (Variety) : Üretilen verilerin %80'inin yapısal olmadığını düşünürsek bu veriler çok çeşitli ortamlardan üretilmektedir. Akıllı telefonlar, mobil cihazlar, tabletler, RFID cihazlar, barkodlar, akıllı üretim tezgahları vb. Veri ne kadar değişik kaynaktan elde edilirse ve ne kadar farklı biçimde (formatta) üretilirse çeşidin bu kadar çok olması onunla uğraşacak olanlarında işini zorlaştırmaktadır. Bunların farklı dillerde ve unicode da olabileceği düşünülürse bu verilerin dönüştürülmesi de zahmetli olacaktır. Çok uluslu şirketlerde üretilen üretim verileri buna en iyi örnektir.

Veri doğrulama (Veracity) : Verinin akışı veya el değiştirmesi sırasında "güvenli" olması da bir diğer bileşendir. Verinin olması gereken güvenlik seviyesinde

---

[20] http://www.idc.com/getdoc.jsp?containerId=prAE24418813
http://www.icttele.com/buyuk-veri/ (erişim tarihi 15.02.2014)

[21] http://ibmdatamag.com/2012/11/harness-the-power-of-big-data-the-ibm-big-data-platform/ (erişim tarihi 15.02.2014)

izlenmesi, yetkisi olan kişilerce erişilmesi, değiştirilmesi ve gizlilik seviyelerinin yine veri sahibi tarafından ayarlanabilmesi gerekmektedir.

Bu 4 bileşen göz önüne alındığında karşımıza bugün bilişim dünyasının üzerinde çeşitli çalışmalar yaptığı "big data" kavramı çıkar. Bu büyük veriden karar vericilere yol gösterecek anlamlı bilginin üretilmesi konusu, Veri Madenciliğinin öneminin başladığı noktadır.

C. Veri Tabanı, Veri Ambarı, Veri Pazarı, Metadata

Veritabanı; En basit tanımıyla, birbirleriyle ilişkili verilerin depolandığı alanlardır. Büyük miktardaki bilgileri depolamada geleneksel yöntem olan ''dosya-işlem sistemine'' alternatif olarak geliştirilmiştir.([22]) Veritabanı barındırdığı bilgileri yapısına uygun biçimlerde kaydetmeye, güncellemeye ve bu veriler üzerinde gelişmiş sorgulamalar yapmaya olanak sağlayan yapılardır. Bir veritabanı üzerinde yapılabilecek işlemler, kabaca, veriyi görmek ve izlemek, istenilen verileri seçmek ve ayırmak, değiştirmek yani veriyi güncellemek ve silmek olarak tanımlanır. Veri tabanı sistemlerinin en temel birimi tablolardır. Tablolar bilgilerin tutulduğu nesnelerdir ve her veri tabanında en az bir tablo bulunmak zorundadır, tablolar ise satır ve sütunlardan meydana gelmektedir. Veri tabanlarında genellikle birden çok tablo bulunur ve bu tablolar birbirleri ile belirli alanlara göre ilişkilendirilir. Birden çok tablonun mantıksal bir ilişki içerisinde tutulduğu veri tabanlarına "**İlişkisel Veri Tabanı**" denir. İlişkisel veri tabanı oluşturmak ve yönetmek için kullanılan sistemlere ise "İlişkisel Veri Tabanı Yönetim Sistemi" (Relational Database Menagement System – RDBMS) denir. İlişkisel veri tabanlarında genellikle İngilizce "Structured Query Language" kelimelerinin baş harflerinden oluşturulmuş, Türkçesiyle Yapısal Sorgulama Dili "SQL" kullanılmaktadır. SQL, bir programlama dili değildir. Herhangi bir veri tabanı ortamında kullanılan bir alt dildir. SQL ile yalnızca veri tabanı üzerinde işlem yapılabilir. SQL'e özgü cümleler

[22] Alp, Özdemir ve Kilitci, "Veritabanı Yönetim Sistemleri" Ekim 2011 / 1. Baskı / 280 Syf.

kullanarak veri tabanına kayıt eklenebilir, değiştirilebilir, silinebilir ve bu kayıtlardan listeler oluşturulabilir.([23])

Veri Ambarı; Bilgi Teknolojileri dünyasından ilk kez 1991 yılında William H. Inmon tarafından ortaya atılan veri ambarı kavramı, analiz amaçlı sorgulamalar yapmak için birçok veri tabanından alınarak birleştirilen verilerin toplandığı özelleşmiş depolardır. İlişkisel veri tabanları, olaylar ve işlemlerle (**transaction**) ilgili verileri saklar, bu yüzden devamlı bir veri giriş çıkışı içerisindedirler ve en güncel veriyi taşırlar. Veri ambarları ise, bu veri tabanlarındaki verilerle diğer dış kaynaklardan alınan verilerin belirli periyodlarda derlenip arşivlenmesi ile oluşturulan, bu sayede dönemsel analizlerin yapılmasına olanak sağlayan yapılardır. ([24])

Veri ambarları, veriler üzerinde yazma, değiştirme, silme yapmak için değil sadece okumaya yönelik olarak oluşturulmaktadır. Bu nedenle veri ambarında veriler, analiz yapmayı kolaylaştıran bir biçimde tutulmaktadır. Burada analiz; sorgular, raporlar, karar destek sistemleri veya istatistiki hesapları kapsar. Kısacası; veri ambarı kullanıldığında, günlük operasyonel görevlerle ki bunlar daha önce bahsettiğimiz verilere ekleme yapma, değiştirme, silme gibi işlemlerdir yeterince meşgul olan veri tabanı kullanılmadan, analiz işlemleri farklı bir ortamda yapılmasına olanak sağlar. Veri ambarı, bir zaman boyutu içinde analitik işlemlerin yapılması için ihtiyaç duyulan bilgi temelini sağlar.

Veri ambarı, karar verme sürecinde karar verici pozisyonundaki kişiye göre hazırlanmış; konuya yönelik, bütünleşik, zaman boyutu olan, sadece okunabilen veri topluluğudur. Veri ambarları söz konusu olduğunda iki kavramdan söz etmek gerekir. Bunlar OLAP ve OLTP kavramlarıdır.

OLAP (On-Line Analytical Processing), ilişkisel veri tabanlarında tutulan veriler üzerinde çok boyutlu analizler yapılmasına olanak sağlayan bir yaklaşımdır.

---

[23] Prof. Dr. Vahap Tecim,"http://www.noda.com.tr/media/19230/vahaptecim2008cbskitabi.pdf (erişim tarihi 11.02.2014)

[24] Çağatay Yurtöz (http://www.cagatayyurtoz.com/drupal-7.0/content/k%C4%B1saca-veri-ambar%C4%B1-olap-etl-data-mart-nedir) (erişim tarihi 11.02.2014)

OLTP (**On-Line Transaction Processing**) ise operasyonel veri tabanlarındaki işlemleri kapsar. OLTP kapsamına giren, veriler anlıktır ve sürekli değişirler.

Veri tabanı ile veri ambarı karşılaştırmasında bilişim dünyasında kullanılan karşılaştırma biçimi Şekil -2'de verilmiştir. Veri ambarı ve OLAP terimlerinin birbirleri yerine kullanılması yanlıştır, veri ambarı yukarıda bahsedildiği şekli ile özelleşmiş bir veri tabanını belirtirken, OLAP eldeki veriler üzerinde çok boyutlu sorgular yapmayı kolaylaştıran bir analiz yaklaşımıdır. Yani OLAP istemci uygulamaların veri ambarını kullanmasını kolaylaştıran teknolojidir. Aslında OLAP küpü, boyut, fact gibi kavramlar da aslında veri ambarı ile doğrudan bağlantılı değildir; OLAP süreçleri veri ambarı gerektirmeyebilir, her veri ambarı da OLAP sürecine sokulmayabilir. ([25])

Veri Ambarı (OLAP)	**Veri Tabanı (OLTP)**
Off-Line çalışır	On-Line çalışır
Veri değişiminden çok sorgulama yapılır	Veri değişimi işlemleri yoğunluktadır
Eski veriler saklandığı için veri miktarı çoktur	Güncel veriler saklandığı için veri miktarı daha azdır
Üst yönetim ve analistler kullanır (Kullanıcı sayısı az)	Veriye ulaşmak ve değiştirmek isteyen her kullanıcıya hitap eder (Kullanıcı sayısı çok)
Veri madenciliği gibi uzun ve karmaşık süreçler sonucunda analizler yapılabilir	Sorgularla istenilen sonuçlara anında ulaşılır

Şekil -2

Veri ambarına aktarılan yeni veriler, veri ambarında mevcut bulunan verilerin güncellenmesi için kullanılmazlar. Bu yüzden veri ambarındaki veriler

[25] Çağatay Yurtöz (http://www.cagatayyurtoz.com/drupal-7.0/content/k%C4%B1saca-veri-ambar%C4%B1-olap-etl-data-mart-nedir) (erişim tarihi 11.02.2014)

değiştirilmemeli, silinmemeli ve güncellenmemelidir. Verilerin değiştirilip güncelleme işi veritabanı tarafında yapılır. Bu açıdan bakıldığında veritabanlarında daha çok operasyonel veri bulunur. Operasyonel sistemlerdeki veriler güncellenip, temizlenip, bütünleştirildikten sonra veri ambarına aktarılırlar. Verilerin son şeklini almadan veri ambarına aktarımı, veri analizi çalışmalarında yanlış sonuçların ortaya çıkmasına neden olur ve bu işin profesyonelleri tarafından istenmeyen bir durumdur.

Veri Modelleme, Veri ambarlarının kurulmasında, çalışmasında en önemli hususlardan birisi veri modelinin oluşturulmasıdır. Gerçeğin soyutlanmış hali olan modelleme ile verilerin analizler için en etkin şekilde veri ambarlarında yerlerini alması hedeflenmektedir. Veri modellemenin amacı; verinin taşıdığı anlamı, veriler arasındaki ilişkileri, verilerin niteliklerini ve verilerin net tanımlarını belirlemektir.

Bilişim dünyasında **DataMart** olarak isimlendirilen veri pazarları ise, veri ambarlarının bir alt bileşeni olup daha küçük boyutta veri barındıran özelleştirilmiş veri ambarcıklarıdır. Veri ambarlarından daha hızlı yaratılıp daha hızlı yanıt döndürmesi sorgulama süreç performanslarına etkileri nedeniyle tercih edilmelerine rağmen, artık günümüzün popüler konularından olan in memory teknolojileri sayesinde sadece orta ölçekli işletmeler için güncelliğini devam ettirmektedir.

Veri Ambarının en önemli bileşenlerinden biri metadatadır. Veri Ambarında verilerin tanımlandığı kısımdır. Metadata "veri hakkında veri" anlamındadır. Metadata her veri elementinin anlamını, hangi elementlerin hangileriyle nasıl ilişkili olduğunu ve kaynak verisi ile erişilecek veri gibi bilgileri içermektedir.

1. Veri Ambarı Bileşenleri ve Fonksiyonu

Kaynak sistemler, veri ambarına yüklenmek istenen verileri içeren OLTP sistemlerdir. (Online Transaction Processing) yazılım uygulamalarındaki temel işlevi, iş süreçlerine ait hareket gören verilerin depolanmasıdır. Kaynak sistemdeki verilerin karakteristiklerini ve veri kalitesini anlamak için veri kesitleme bileşeni

(Data Profiling) kullanılır. Böylece belirli bir tabloda, her sütunun ayrı ayrı değerlerini ve veri değerlerinin dağılımı anlaşılır.

Kaynak sistemlerden alınan verilerin istenen şekle sokulup veri ambarındaki tabloların doldurulması ETL (Extract,Transform and Load) bileşeni tarafından yapılır. ETL, veri kaynağına bağlanma, verileri okuma, verileri dönüştürme ve hazırlanan verileri hedef sisteme yükleme yeteneklerine sahip bir bileşendir.

Veri akışı mimarisinde yer alan ETL süreçlerinin açıklamaları metadatada yer alır. Metadata verilerin alındığı kaynak, yüklenildiği hedef sistem ve çalıştırılması gereken ETL süreçlerinin çizelgeleri gibi verilerin yer aldığı yapılardır. Veri ambarı sisteminde metadata, veri ambarında yer alan süreçler, veriler ve yapılar hakkındaki bilgileri tutan veri deposudur. ([26])

Veriler ETL aracılığıyla stage alanına aktarıldıktan sonra yine ETL devreye girerek entegrasyonu sağlar, gerekli dönüşümleri yapar ve verilerin son halini DDS (Dimensional Data Store) sistemine aktarır. DDS; veri ambarı verilerini OLTP den farklı bir biçimde depolayan bir veritabanıdır. Verilerin direkt olarak veri kaynağından değil de DDS üzerinden sorgulanmasının nedeni, verilerin kaynağından alınırken entegre ederek kendi yapısında tutulması, ayrıca verilerin boyutsal (dimensional) bir biçimde düzenlenerek analiz işlemlerine daha uygun hale getirilmesindeki kolaylıktır. ETL sistemi verileri DDS'e yüklediğinde veri kalite kuralları (DQ) çeşitli kalite kontrolleri yapar. Kontrolleri geçemeyen veriler rapor edilip veri kaynağında düzeltilmek için data quality (DQ) veritabanına konulurlar. Bu veriler eğer kontrol limitlerinin içindeyseler otomatik olarak düzeltilebilir veya tolere edilebilirler. Bir veri ambarı sistemi tüm bu bileşenleri içermek durumunda değildir. Basit bir veri ambarı sadece ETL ve DDS bileşenlerinden oluşabilir fakat bu bileşenlerden biri daha eksiltilirse o yapı artık veri ambarı yapısı değildir.

[26] http://datawarehouse.gen.tr/veri-ambarinda-staging-alanin-onemi/ erişim tarihi 11.02.2014)

Yukarıda anlatılan bileşenleri içerdikleri fonksiyonları da göz önünde bulundurarak yapısal bir akışla açıklamak istersek; ([27],[28])

• Değişik platformlar üzerindeki işletimsel uygulamalara ait verilere erişim ve gerekli verilerin bu platformlardan alınması.

• Alınan verilerin temizlenmesi, tutarlı duruma getirilmesi, özetlenmesi, birleştirme ve birbirleriyle entegrasyonun sağlanması.

• Dönüştürülen verilerin Veri Ambarı veya datamart ortamına dağıtımı

• Gönderilen verilerin bir veri tabanında toplanması

• Depolanan bilgi ile metadatada bulunan ilgili bilgilerin veri kataloğunda saklanması ve son kullanıcılara sunulması.

• Veri Ambarı veya Datamart da bulunan bilgileri uç kullanıcıların karar destek amaçlı kullanımının sağlanması.

Fonksiyonlarını içermektedir.

II. Veri Madenciliği Süreci ve Veri Madenciliği Algoritmaları

Veri madenciliğinin temel varoluş nedeni, verinin bilgiye nasıl dönüştürüleceğidir. Veri madenciliği yöntemleri, bilgi keşfinin hedeflenen çıktılarına bağlı olarak çok farklı amaçlara sahip olabilirler. İstenen sonucu başarılı bir şekilde sağlamak amacıyla farklı amaçlara sahip birçok yöntem birlikte uygulanabilir.([29]) Dünya'da veri madenciliğinin süreç tanımı konusunda Cross Industry Process for Data

[27] Ümit Burak Usgurlu Başkent Ünv.Bilgisayar Müh. Yönetim Bilişim Sistemleri http://printfu.org/usgurlu (erişim tarihi 11.02.2014)

[28] http://www.cevizbilgi.com.tr/en/veri-ambari-ve-zekasi-yapisal-teorisi/ (erişim tarihi 11.02.2014)

[29] Emrah Arslan http://emraharslanbm.wordpress.com/2012/07/17/veri-ambari-etl-veri-madenciligi/ (erişim tarihi 14.02.2014)

Mining referansı ve Daniel Laross'un değerlendirmeleri kriter alınmıştır. ([30],[31]) Bir çok araştırmacı bu kriterler üzerinden çalışmalarını yürütmektedir.

A. Veri Madenciliği Süreci

- İşi Anlamak, tanımlamak ve hedefi belirlemek
- Veriyi Anlamak
- Veriyi Hazırlamak
- Modelleme
- Değerlendirme
- Yayma

Sürecin bileşenlerini kısaca açacak olursak;

1. İşi Anlamak, Tanımlamak ve Hedefi Belirlemek

Veri madenciliği sürecinin ilk adımı süreci doğrudan etkileyen, en önemli kısmıdır. Varılmak istenen hedef, hedefe ulaşıldığında işin başarıldığına yönelik hangi kriterlerin değerlendirileceği bu aşamada planlanır. Bir anlamda büyük resim baştan görülür. Veri, madenciliğin hizmet edeceği amaç, bu amaca ulaşmak için üzerinde analiz yapılacak verilere karar verilmesi, sonucun değerlendirilmesindeki başarı kriterlerinin ne olacağının planlaması, bu ilk aşamanın başlıca unsurlarını oluşturur. Eldeki kaynakların yeterliliğinin değerlendirilmesi, dış kaynak ihtiyacının tespiti, sonuçların nasıl raporlanacağının belirlenmesi bu aşamada yapılır. Olası riskler saptanır, çalışma sırasında ne tür teknolojilerin kullanılacağı kararlaştırılır, (in memory çözümler, BW/BI teknolojileri vb.) yapılacak çalışmanın maliyet boyutu çıkarılır. Bu kısım tamamen bir proje bazlı iş olarak görülmelidir ve proje yönetim metodolojilerinden faydalanılmalıdır. Bu noktada, karar vericiler ile analiz işini yapanlar ve proje yönetimi beraber çalışmalıdır. Planlar bir proje takip

[30] Daniel T. Larose, Discovering Knowledge in Data: An Introduction to Data Mining, John and Wiley Sons Incorporated, USA, 2005

[31] http://www.the-modeling-agency.com/crisp-dm.pdf (erişim tarihi 14.02.2014)

yazılımına geçirilip çalışmanın kilometre taşları (milestone), kişi ve kullanılacak kaynakların belirlenmesi, kritik noktaların önceden tespit edilmesine ve proaktif önlemler alınmasına da olanak sağlayacaktır.

2. Veriyi Anlamak

Eldeki veritabanı envanterinin gözden geçirilerek hangi verinin nerede olduğunun keşfedilmesi ile başlayan bu süreç, verilerin toplanması, elde edilen verilerin ön analizden geçirilerek uygunluğunun değerlendirilmesi, veri niteliklerini tanımlama, benzer verilerin bulunması, modeli oluşturmak için gerek duyulabilecek farklı veri ihtiyaçlarının tespit edilmesi, veri kalitesinin ve yeterliliğinin incelenmesi aşamasıdır. Hedef çalışmada kullanılacak veri setlerinin belirlenmesine karar verilmeye çalışılır. Birinci aşamadaki süreç yani, işi tanımlamak ve hedefi belirlemek aşaması ne kadar sağlıklı geçildiyse bu süreçte çalışılacak verileri belirleme o kadar sağlıklı ilerler. Veri incelendikçe işin diğer aşamaları veya ileriye yönelik süreçte önemli olabilecek hususlarda ortaya çıkmaya başlar. Birden fazla kaynaktan elde edilen verilerin bütünleştirilmesi ve veri kalitesinin yükseltilmesi daha sonraki aşama içinde ele alınacaktır. Bu aşamanın en önemli kısmı, verinin kalitesine karar verilmesi kısmıdır. Yine bu aşamada hedefi göz önüne alarak belirlenen hipotezlerin doğruluğu veri üzerinde denenmelidir. Bu durum konu üzerinde çalışan uzmanlara ileriki aşamalarda veri ile çalışma kolaylığı sağlayacaktır.

3. Veriyi Hazırlama

Veri hazırlama aşaması projenin temel analiz çalışmalarının yapıldığı kısımdır. İşe başlarken bulunan, anlamlandırılan verilerin sonuç elde edilecek verilere dönüştürülmesidir. Ham veriden modelleme aracı ile beslenecek olan nihai veri kümesini oluşturmak için gerekli bütün faaliyetleri kapsar. Bu sürecin arkasından gelen modelleme aşamasında karşılaşılan her sorunda, tekrar bu aşamaya dönülür ve veriler yeniden düzenlenir. Bu aşama projenin kilometre taşlarındandır çünkü en çok zaman harcanan ve tekrarlanan aşama burasıdır. Konu hakkında daha önceki

projelerden elde edilen birikimler ve uzmanlaşmalar bu aşamada yaşanan zaman kayıpları ancak %10 – 15 oranında minimize edilebilir. Yapılan çalışmaların belirli bir sırası ve tekrar sayısı olmadığı ifade edilmesine rağmen, yaşanan deneyimler projeyi yürüten veya karar vericinin bazı hususlarda kendi çıkarımlarını yapmasının yerinde olacağı şeklindedir.

a. Veriyi Seçmek

Yapılacak analizde kullanılacak veriler belirlenir. Bu belirlemede seçim, kurulacak modele bağlı olarak yapılır. Tahmin edici bir model için bu adım, bağımlı ve bağımsız değişkenlerin ve modelde kullanılacak veri kümesinin seçilmesi anlamını taşımaktadır. Değerlendirme sırasında verinin belirlenen hedef ile ilişkisine, kalitesine, miktarına dikkat etmek gerekir. Gereğinden az veri, çalışmayı eksik bırakabileceği gibi, gereğinden fazla veri, veri kirliliğine ve bu aşamanın gereğinden çok uzamasına neden olabilecektir. Seçim sırasında anlamlı olmayan değişkenlerin modele girmemesi gerekmektedir, çünkü bu tip değişkenler, diğer değişkenlerin modeldeki ağırlığının azalmasına ve veriye ulaşma zamanlarının uzamasına neden olabilmektedir.

b. Veriyi Temizlemek

Veri madenciliğinde amaç, veriden hareketle bilgiye ulaşmada verileri, istenilen sonuca ulaşmasına olanak verecek biçimde modellemek ve analiz etmek olmalıdır. Bu noktada veri temizleme işi sürecin kilometre taşıdır. Veri temizleme işinin veri tutarsızlığına neden olmaması için gerekli önlemler bu aşamada alınmalıdır. Üzerinde çözümleme yapılacak veriler istenen özelliklere sahip değilse, bu durum veri tutarsızlığına neden olur. Bu tür tutarsız ve hatalı veriler "gürültü" olarak adlandırılır. Gürültülü veri, asıl veriyi dinlemeye, bulmaya engel teşkil eden bir unsurdur. Bir verinin gürültülü olma nedenleri;[32]

- Hatalı veri toplama gereçleri
- Veri giriş problemleri

[32] Eliza Natasa Artinyan Deloitte Türkiye Kurumsal Risk Hizm. Kıd. Danışman (Erişim tarihi 18.02.2014)
http://www.denetimnet.net/UserFiles/Documents/Makaleler/BT%20Denetim/Veri_Analizi_Veri_Kalitesi_ve_B%C3%BCt%C3%BCnl%C3%BC%C4%9F%C3%BC.pdf

- Veri girişi sırasında kullanıcıların hatalı yorumları
- Veri iletim hataları
- Teknolojik sınırlamalar
- Veri isimlendirmede veya yapısında uyumsuzluk olabilir.

Verilerdeki gürültüyü temizlemek için; eksik değer içeren kayıtlar göz ardı edilir, kayıp değerlerin yerine yenileri atanabilir, diğer verilerin ortalaması hesaplanarak kayıp veriler yerine bu değer yazılabilir, verilere uygun bir tahmin (karar ağacı, regresyon vb.) yapılarak eksik veri yerine kullanılabilir. Eksik veri bilgi teknolojileri konusu ile uğraşan profesyoneller için istenmeyen bir durumdur. Veri kaybı belirlenen ya da belirlenemeyen pek çok nedenden kaynaklanabilir ve bu durum doğru ve güvenilir analizler için veri kümesinin tutarsız olma durumuna zemin hazırlar, bu durum daha sonra yapılacak tüm aşamalara tutarsızlığı taşımış olur. Bu nedenle veri temizleme işinde verinin tutarlılığı bir çok yönden karşılaştırılarak sınanmalıdır. Kayıp verilerin değerlendirilmesindeki ilk yöntem kayıp veriyi yok saymaktır. Ancak kayıp verinin çok olduğu yani kayıp veri oranının, toplam veri miktarının 1/3'üne yaklaştığı durumlarda yada değerlendirilecek veri miktarının az olduğu durumlarda, veriyi yok saymak sonucu yanlış yerlere götürme olasılığını arttırmaktadır. Bu yüzden kayıp verilerin yerine bir değer ataması yapmak veri madenciliği ile uğraşanların daha çok tercih ettiği bir yöntemdir. Kayıp verileri tahmin etmede;

Regresyon ile belirleme, Hot/Cold Deck ile belirleme, Beklenti Maksimizasyonu, Son Gözlemi İleri Taşıma, Çoklu Atama, Karar Ağacı, Naive Bayes gibi yöntemler kullanılabilir.

Regresyon analizi; Bu analizin temeli, değerler arasındaki ilişkileri tahmin etme modeline dayanır. Temeli, bir ya da daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini olanaklı kılan ilişkiler bulmaktır. Örnek vermek gerekirse eldeki verilerden hareketle, "hem evli, hem ev sahibi ve aynı iş yerinde 5 yıldan fazla bir süredir çalışan ve aldığı banka kredisini zamanında ödeyen birinin kredi skoru; 295'tir " şeklindeki bir çıkarımın sonucu bir regresyon ilişkisidir. Bu yöntemde tahmin edilen regresyon modeli, kayıp gözlemleri tahmin etmede bir araç olarak kullanılmaktadır.

Kayıp verilerin tahmininde gözlem sayısı ve verinin özelliği oldukça önemlidir. Her şeyden önce kayıp verileri durum düzeyinde silme işlemi yapılırsa kullanımı basit olmasına rağmen fazla veri kaybında varyans artar ve rassal olarak kayıpta hatalı sonuçlar üretir. Buna karşın regresyon ataması, bu yöntem için kayıtlarda önce korelasyonu yüksek iki alan seçilip ona göre bir regresyon formülü üretilebilir. Bu durum ilişkili olamayan alanlarda işe yaramaz. Bu fonksiyonun oluşturulmasında hatanın göz önünde bulundurulması gerekir.([33]) Hangi İstatistiki yöntemin seçileceğine elde bulunan verinin niteliğine göre karar verilmelidir. Veri madenciliği sürecinde, istatistik ve matematik biliminden ve bunların kullandığı model ve yöntemlerinden sadece veri temizleme kısmında değil, veriyi bütünleştirme, dönüştürme aşamalarında regresyon, koreleasyon analizleri, karar ağaçları algoritmalarının uygulanması sırasında da yararlanılmaktadır.

c. Veriyi Kurmak

Verilerin temizlenmesi aşamasından sonra elde edilen veri setlerinin analiz edilmesiyle hiçbir değer ataması olmayan veya sonuçları veri üzerinde işlem yapılarak bulunabilecek değişkenlerin oluşturulacak model için daha kullanışlı hale getirilmesi aşamasıdır. Bu aşama özellikle veri modellemesi öncesi oluşturulması planlanan OLAP küpleri içinde ölçüt oluşturması açısından önemlidir. Veri madenciliği ve OLAP birbirini tamamlayan unsurlardır, veri madenciliği ile oluşturulan hipotez OLAP ile test edilir.

d. Veri Bütünleştirme

Farklı veri tabanlarından ya da veri kaynaklarından elde edilen farklı türdeki verilerin birlikte değerlendirmeye alınabilmesi için tek türe dönüştürülmesi işlemidir. Verinin temizliği yapıldıktan sonra yapılan incelemede farklı kaynaklardan elde edilen verilerin arasında ilişkilendirmeyi keşfedip, bunları tek

[33] Evren SEZGİN, Yüksel ÇELİK "Veri Madenciliğinde Kayıp Veriler İçin Kullanılan Yöntemlerin Karşılaştırılması" (Akdeniz Üniversitesi, Enformatik Bölüm Başkanlığı, Karamanoğlu MehmetBey Üniversitesi, Karaman MYO, Karaman.)

tipe indirgemek için uygulanan yöntemdir. Bunu bir örnek ile açıklayacak olursak, cinsiyet kaydı çok değişik türlerde tutulan bir kayıttır. Bazı veri tabanlarında Erkek 0, Kadın 1 olarak tariflenmişken bir diğerinde Kadın 1, Erkek 0 olarak tanımlanmış olabilir veya diğer veri tabanında E/K veya Erkek/Kadın şeklinde tutulabilir. Veri madenciliği süreci içerisinde yer alan modelin kurulmasındaki başarı, verinin uyumuna da bağlıdır.

e. Veri İndirgeme ve Biçimlendirme

Verinin bütünleştirilmesi aşamasından sonra bir kez daha başa dönülerek durum analiz edilir. Varılmak istenen hedefe uygun veri sağlandığı, verinin kalitesini arttırıcı yönde temizliğinin yapıldığı, yine kalitesi anlamında bütünleştirilmenin sağlandığı tespit edilir ve eldeki veri miktarı bir kez daha analiz edilir. Veri madenciliği uygulamalarında çözümleme işlemi uzun süre alabilir, eğer çözümlemeden elde edilecek sonucun değişmeyeceğine inanılıyorsa veri sayısı ya da değişkenlerin sayısı azaltılabilir.

Veri indirgeme değişik boyutlarda yapılabilir;

a) Veriyi birleştirme veya veri küpü
b) Boyut indirgeme
c) Veri sıkıştırma
d) Örnekleme
e) Genelleme

Bu yöntemlerden biri veya aynı anda bir kaçı uygulanabilir. Veri indirgeme çok dikkat isteyen bir çalışmadır, veri tutarsızlığına neden olma riski taşır.

4. Modelleme

Belirlenen hedef için en uygun modelin bulunabilmesi, olabildiğince çok sayıda modelin kurularak denenmesi ile mümkündür. Bu nedenle veriyi hazırlama ve model kurma aşamaları, en iyi olduğu düşünülen modele varılıncaya kadar yinelenen bir süreçtir. Modellemede kullanılan algoritmalar “Veri Madenciliği

Algoritmaları" bölümünde daha detaylı ele alınacaktır. Test tasarımının üretimi, model geliştirme ve tahmin işlemleri de bu aşamada yapılır. Veri Madenciliği gözlenen veriye göre model uydurmayı veya gözlenen verideki örüntüleri tanımlamayı gerektirmektedir. Uygun modele karar verme durumu karar vericinin bilgi çıkarımı rolünü üstlenmesidir. Modelin, kullanışlı veya ilginç keşifsel bilgiye işaret edip etmediği, tamamıyla interaktif olarak sürecin subjektif insan yargısına ihtiyaç duyduğu bir aşamasıdır. Model uydurmada istatistiksel ve mantıksal olmak üzere iki temel matematiksel yapı kullanılmaktadır. Modelde, istatistiksel yaklaşım deterministik olmayan etkiye, mantıksal yaklaşım ise deterministik etkiye izin vermektedir.([34]) Model kuruluşu çalışmalarının sonucuna bağlı olarak, aynı teknikle farklı parametrelerin kullanıldığı veya başka algoritma ve araçların denendiği değişik modeller kurulabilir, hangi tekniğin en uygun olduğuna karar verebilmek güçtür. Bu nedenle farklı modeller kurarak, doğruluk derecelerine göre en uygun modeli bulmak üzere sayısız deneme yapılmasında yarar vardır.

Bir modelin doğruluğunun test edilmesinde kullanılan en basit yöntem basit geçerlilik testidir. Bu yöntemde tipik olarak verilerin % 5 ile % 33 arasındaki bir kısmı test verisi olarak ayrılır, verinin kalanı üzerinde modelin öğrenimi gerçekleştirilir. Modelin öğrenilmesi eldeki veriye alışkanlık kazanma, verinin farklı modeller karşısında nasıl davranacağını kestirme, uygulanacak modeli kullanma gibi hem sübjektif insan unsurunun dâhil olduğu hem de istatistiki yöntemlerin ve eldeki veri madenciliği uygulamalarının kullanıldığı aşamaları içerir. Öğrenilen deneyim ayrılan test kümesi verisi üzerinde denenir. Bir sınıflama modelinde yanlış olarak sınıflanan olay sayısının, tüm olay sayısına bölünmesi ile hata oranı, doğru olarak sınıflanan olay sayısının tüm olay sayısına bölünmesi ile ise doğruluk oranı hesaplanır. (Doğruluk Oranı = 1 - Hata Oranı) ([35])

---

[34] Ali Serhan Koyuncugil,(Sermaye Piyasası Kurulu, Araştırma Dairesi)
Nermin Özgülbaş (Başkent Ünv. Sağlık Bil. Fak. Sağlık Kurumları İşletmeciliği Bölümü)
"Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları"

[35] Doç. Dr. Kaan Yaralıoğlu (Dokuz Eylül Ünv. İktisadi ve İdari Bilimler Fakültesi) Veri Madenciliği "Uygulamada Karar Destek Yöntemleri"
http://www.deu.edu.tr/akademik/index.php?cat=3&akod=19912335 (erişim tarihi 13.02.2014)

Sınırlı miktarda veriye sahip olunması durumunda, kullanılabilecek diğer bir yöntem, çapraz geçerlilik testidir. Bu yöntemde veri kümesi rastgele iki eşit parçaya ayrılır. İlk aşamada bir parça üzerinde model öğrenimi, diğer parça üzerinde test işlemi; ikinci aşamada ise ikinci parça üzerinde model öğrenimi ve birinci parça üzerinde test işlemi yapılarak elde edilen hata oranlarının ortalaması kullanılır. ([36]) Model kuruluş süreci, denetimli ve denetimsiz öğrenim metodolojilerinin kullanımına göre farklılık gösterecektir.

5. Değerlendirme

Modelleme aşaması tamamlandığında değerlendirme aşamasına geçilir. Değerlendirmede eldeki verilerin kalitesi, modelin doğru kurulup kurulmadığı, kurulan modelin istenen amaca hizmet edip etmeyeceğinin değerlendirilmesi yapılır. Riskler gözden geçirilir, maliyet planlaması yeniden yapılır ve bütçe sınırları içinde kalınıp kalınmayacağı tespit edilir. Projenin bundan sonraki adımları belirlenir. Kaynak ve zaman planlaması yeniden gözden geçirilir, varsa kısıtlar belirlenir. İşin kalitesinin sorgulandığı aşamadır. Değerlendirme sonuçlarının nasıl görselleştirileceği ve raporlanacağı üzerinde bu aşamada karar verici durumundakilerle fikir birliğine varmak önemlidir. Değerlendirilme aşamasında ortaya konan tüm tespitlerin raporlanması gerekir, bu modelleme aşamasına geri dönüşler olduğunda daha önce yapılanların tekrarını önleme ve zaman kazandırma açısından önemlidir.

6. Yayma (Deployment)

İşin son aşamasıdır. Model tam anlamıyla ortaya çıkmış, doğrulanmış, işin kalitesi onaylanmış ve gerçek veriler üzerine test edilmiştir. Yayma aşamasında izlenecek metot mutlaka bir prosedüre bağlanmalıdır. Yayma planı bu prosedür çerçevesinde

---

[36] Özlem Terzi, Ecir Uğur Küçüksille, Gülşah Ergin, Ahmet İlker "Veri Madenciliği Süreci Kullanılarak Güneş Işınımı Tahmini "SDU International Technologic Science Vol. 3, No 2, February 2011 http://edergi.sdu.edu.tr/index.php/utbd/article/view/2034/2338 (erişim tarihi 13.02.2014)

yürütülür. Bu aynı modelin bir başka zaman diliminde başka bir amaç için uygulanmasında da uyulacak kural ve yöntemleri belirtmesi ve oluşabilecek sıkıntıları baştan önlemesi açısından önemlidir. Yayma aşaması modelin uygulamaya sokulduktan sonra izlenmesi, bakımının yapılması ve iyileştirilmesi kısımlarını da kapsar. Tüm yapılanlar bir sonuç raporu ve sunumu ile karar vericilerle paylaşılmalıdır. Bu paylaşımda projenin değerlendirilmesi ve çıktılarının neler olduğu, kazanımlar, başta konan hedef ile elde edilen sonuçların karşılaştırılması proje paydaşları ve karar vericilere sunulmalıdır.

B. Veri Madenciliği Strateji ve Modelleri

Veri Madenciliği Stratejisi iki temel üzerine oturur.

a. Denetimli (*supervised) ve*
b. Denetimsiz (*unsupervised)* öğrenme yöntemleri olarak.

Bu stratejiler verinin modellemesi aşamasında veriyi öğrenme, analiz etme ve modelleme metodolojilerini belirlemede de esastır.

İyi tanımlanmış veya kesin bir hedef olduğunda denetimli (supervised), elde edilmesi istenen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik söz konusu ise denetimsiz (unsupervised) öğrenmeden bahsedilir. ([37])

Denetimli öğrenme yönteminde; veri örneklerinden hareket ederek her bir sınıfa ilişkin özelliklerin bulunması ve bu özelliklerin kural cümleleri ile ifade edilmesidir. Örneğin, veritabanındaki kayıtlarda her kaydın yanında kadın veya erkek bilgisi yazılıyor olsun, bu durumda veritabanı üzerinde yapılan herhangi bir (kadın veya erkek olduğuna dair) kural çıkarma işlemi denetimin kullanıldığı bir yöntemdir. Denetimli yöntem aynı zamanda yönlendirilmiş yöntemdir.([38]) Denetimli yöntemler, veriden bilgi ve sonuç çıkarmak için kullanılmaktadır.

---

[37] T. Hastie, R. Tibshirani, J. Friedman, "The Elements Of Statistical Learning; Data Mining, Inference And Prediction", Springer Series In Statistics, 533, USA, 2001.

[38] Ardügen ; http://www.arge.com/wp-content/uploads/2013/03/VeriMadenciligi.pdf (erişim tarihi 17.02.2014) & Kaan Yaralıoğlu a.g.e

Denetimsiz yöntem ise, elde edilmesi istenilen sonuç için özel bir tanımlama yapılmamışsa veya belirsizlik söz konusu ise kullanılır. Denetimsiz yöntemlerde, kümeleme analizinde olduğu gibi ilgili verilerin gözlenmesi ve bu verilerin özellikleri arasındaki benzerliklerden hareket ederek sınıfların tanımlanması amaçlanmaktadır. Örneğin, veritabanında, kayıtların yanında kadın mı erkek mi olduğu bilgisi yok ise yapılan kural çıkarma işlemi denetimsiz yöntemdir. Denetimsiz yöntemler daha çok veriyi anlamaya, tanımaya, keşfetmeye yönelik olarak kullanılır ve sonraki uygulanacak yöntemler için fikir vermeyi amaçlamaktadır. Denetimsiz öğrenmede hedef değişken bulunmamaktadır. Bu nedenle modeli kurmak için kullanılan tüm değişkenler açıklayıcı değişkendir.
Denetimsiz bir yöntemle elde edilen bir bilgi veya sonucu, eğer mümkünse denetimli bir yöntemle teyit etmek, elde edilen bulguların doğruluğu ve geçerliliği açısından önem taşımaktadır.([39]) Veri madenciliğinde bu iki yöntemi (denetimli - (*supervised)*, denetimsiz - (un*supervised))* baz alarak oluşturulan modeller;

a. Tahmini Metotlar (Predictive Methods) ve
b. Tanımlayıcı Metotlar (Desciptive Methods) 'ın kullanıldığı modeller olarak ayrılır.

Burada veri madenciliği stratejisi ile modellemesinde kullanılan sınıflamanın hiyerarşisinden bahsedilmektedir. Tahmini yöntemler, bağımsız değişkenlerin bir fonksiyonu olarak bağımlı değişkeni tahmin etmek için kullanılırlar. Diğer bir ifade ile veri tabanından çıkarılan desenler, geleceği tahmin için kullanılır. Bir başka tanımı ise; sonuçları bilinen geçmiş verilerden hareket edilerek sonuçlan bilinmeyen veri kümeleri için geleceğe yönelik sonuç değerlerin tahminini yapmaktır. Bu yöntemde kullanıcının bazı alan bilgilerini bilmese bile kayıt etmesine izin verir, bu boşlukları, önceki kayıtlara bakarak tahmin yoluyla doldurur. Tanımlayıcı yöntemlerde karar vermeye rehberlik etmede kullanılabilecek veri setinde yer alan veriler arasındaki ilişkileri, bağlantıları ve

---

[39] Metin Vatansever "Veri Madenciliği kümeleme analizi" http://kutuphane.dogus.edu.tr/mvt/details.php?recid=6793&lng=0 (erişim tarihi 31.01.2014)

davranışlarını bulmak, mevcut verilerdeki örüntülerin tanımlanmasını sağlamaktır. Tanımlayıcı yöntemler ne olduğu önceden belirlenmemiş bir fikir ya da hipotez olmadan, veri tabanı içerisinden gizli desenleri aramaya çalışır. Geniş veri tabanlarında kullanıcının pratik olarak aklına gelmeyecek ve bulmak için gerekli doğru soruları bile düşünemeyeceği birçok gizli desen, tanımlayıcı yöntemlerle keşfedilebilir. Buradaki asıl amaç, bulunacak desenlerin zenginliği ve bunlardan çıkarılacak bilginin kalitesidir.

Denetimli ve denetimsiz yöntemler veri madenciliğinin stratejilerini tanımlamada bize yardımcı olurken bunun bir alt hiyerarşisinde yer alan tahmini ve tanımlayıcı metotlar ise işin uygulamaya yönelik kısmına hitap eden veri madenciliği modellerinin sınıflandırılması için kullanılan bir ayrımdır. Aslında her ikisi de birbirinin yerine kullanmak çok yanlış olmasa bile, denetimli ve denetimsiz yöntemlerin bünyesinde tahmini ve tanımlayıcı metotlara ait somuta indirgenmiş metodolojilerin hayat bulacağı modellerin kullanılacağıdır.

1. Tahmini (Öngörüsel) Yöntemlere Göre Modeller

Veri Madenciliği yöntemlerinin işlevsel ayrımı açısından sınıflandırılan tahmini yöntemin kullanıldığı metotların kategorizasyonu aşağıdaki gibidir. Sınıflandırma ve regresyon içerisinde yer alan yöntemler, veri madenciliği metot ve algoritmaları kısmında ele alınacağı için bu kısımda genel bir değinme yapılmıştır.

a. Sınıflandırma

Sınıflandırma, en temel veri madenciliği fonksiyonlarından biri olarak kategorik sonuçları tahmin etmek için kullanılır. Burada amaç, önceden görünmeyen ve etiketlenmemiş nesnelerin doğru sınıf etiketine atayan sınıflama modellerinin (bazen sınıflandırıcı olarak adlandırılır) kurulmasıdır. ([40])

[40] Selim Tüzüntürk "Veri Madenciliği ve İstatistik" Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt XXIX, Sayı 1, 2010, s. 65-90

Veri üzerinde çalışılıp öğrenme süreci tamamlandığında (bu denetimli öğrenme sürecidir.) sınıflama kuralları oluşturulur. Ortaya çıkan yeni durumlar veya değerlendirmeler bu kurallar dikkate alınarak belirlenen sınıflara atama yapılır. Diğer bir ifade ile bir niteliğin değeri diğer nitelik dikkate alınarak belirlenir.([41])

b. Regresyon

Regresyon, Bir ya da daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini sağlayacak ilişkileri bulma esasına dayanan İstatistiğinde çok kullandığı bir yöntemdir. Bir diğer tanımla veriyi gerçek değerli bir fonksiyona dönüştürme işidir.(19) Regresyon metodunda girdilerden hareketle, burada girdiler bağımsız değişken olarak adlandırılır, çıktıyı yani bağımlı değişken olarak adlandırılan sonuç değerleri ilişkilendirecek yapıyı kurgulayıp en iyi tahmine ulaşılma hedeflenir. Elde edilen sonuç güven sınırları belirtilerek kullanılır. Bu yöntemde girdiler çoğunlukla birden fazla olmak durumundadır. Bu girdilerin sonuca katkıları tek tek değerlendirilip sonuca en az katkı sağlayan girdi göz ardı edilebilir.

Genel olarak;

- Karar Ağaçları,
- Bayes Sınıflandırması,
- Hatayı Geri Yayma,
- Karar Destek Makinaları,
- En Yakın Komşu,
- Zaman Serisi Analizi
- Yapay Sinir Ağları
- Doğrusal Regresyon Analizleri

gibi kullanılan teknik ve algoritmaları içerir.

[41] Han, J., & Kamber M. Second Edition. Data Mining Concepts and techniques.

c. Zaman Serisi Analizi

Zaman serisi analizi, bir değişkenin zamana bağlı olarak değişen değerlerini tespit edip, gelecekte alabileceği değerleri tahmin etmektir. Bir değişkenin değişen zamanlarda gözlenen değerlerini bildirirler.([42]) Örneğin yıllara göre reklam harcamaları ve yine yıllara göre bir firmanın ürettiği veya sattığı mallara olan talep değerleri birer zaman serisi örneğidir. Zaman serileri dört bileşenden oluşur, Trend(Genel Eğilim) bileşeni; Zaman serilerinin uzun sürede gösterdiği düşme ve yükselme süreçlerinden sonra oluşan kararlı durumdur.

Mevsim Bileşeni; Zaman serilerinde mevsimlere göre değişmeyi ifade eder. Çevrimsel Bileşen; Ekonomide, mevsimsel değişmeler ile ilgili olmayan dönemsel değişmelerdir. ([43])

Düzensiz Bileşen; Diğer unsurlar gibi belirli olmayan, hata terimi ile ifade edilebilecek değişmelerdir.

Zaman serisi analizinde modelin kurulduğu öğrenme kümesi üzerinde hangi periyotlarla çalışılmışsa, analiz sonucunda elde edilen değerlerde aynı periyotlarda elde edilir.

Zaman serisi analizleri; talep tahminlerinde, satış tahminlerinde, Pazar payı ve kar tahminlerinde, planlı bakım zamanlarının belirlenmesinde kullanılır.([44])

2. Tanımlayıcı Yöntemlere Göre Modeller

Veri Madenciliği yöntemlerinin işlevsel ayrımı açısından sınıflandırılan denetimsiz- tanımlayıcı yöntemin kullanıldığı metot ve algoritmalara ileri bölümde daha kapsamlı ele alınacağından bu kısımda kullanılan teknikler kısa başlıklar halinde verilmektedir. Bunlar;

- Kümeleme

---

[42] Mustafa Aykut GÖRAL "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007

[43] http://web.sakarya.edu.tr/~adurmus/statistik/acikogretim/unite14.pdf (erişim tarihi 01.03.2014)

[44] Mustafa Aykut GÖRAL "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007 sf:35

- Bağlantı/Birliktelik Analizi
- Sıralı Dizi Analizi
- Tanımsal İstatistik
- İstisna Analizi

Olarak sayılabilir.

Tanımlayıcı modellerde, karar vermeye rehberlik etmede kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. Fonksiyonların amacı belirli bir hedefi tahmin etmek değildir. Amaç veri setinde yer alan veriler arasındaki ilişkileri, bağlantıları ve davranışları bulmaktır. Var olan verileri yorumlayarak davranış biçimleri ile ilgili tespitler yapmayı ve bu davranış biçimini gösteren alt veri setlerinin özelliklerini tanımlamayı hedefler. Tanımı bilmek; tekrarlanan bir faaliyette veya tanımı bilinen yeni bir verinin yapıya katılmasında ne şekilde hareket edileceği konusunda karar almaya destek olur.

a. Kümeleme

Kümeleme analizi denetimsiz ve tanımlayıcı yönteme ait bir modeldir. Kümelemeye dayalı modellerin, Sınıflandırmaya dayalı modellerden farkı eldeki mevcut verilerin daha önceden belirli olan bir sınıflandırmaya göre değil, belirli olmayan bir sınıfa göre gruplandırmasıdır. Kümeleme analizinin hedefi, veri setinde doğal olarak meydana gelen altsınıfları bulmaktır. Veriler içerdikleri benzer özellikler göz önüne alınarak değerlendirilir. Örneğin müşteri verisi ile yapılacak bir analizde kümeleme yöntemi kullanıldığında tüm müşteriler kendileri ile ortak özellikleri gösteren müşterilerle aynı kümeye toplanır. Kendi içinde çok çeşitli açılardan benzer özellikler, benzer tutum ve davranışlar gösteren bu grupların pazarlama faaliyetlerinde de benzer harcama eğilimleri göstereceği düşünülür.

b. Bağlantı/Birliktelik Analizi

Büyük veri kümeleri arasında birliktelik ilişkilerinin bulunması esasına dayanır. İncelenen veriyle aynı anda ikinci ya da üçüncü verinin birlikte bulunma durumlarının incelenmesi ve orantılanması yapılarak kurulan modelleme tekniğidir. Birliktelik kuralları literatürde, market sepeti analizi (market basket analysis) olarak

da geçer. Örneğin; bir A ürününü satın alan müşteriler aynı zamanda B ürününü da satın alıyorlarsa, bu durum Birliktelik Kuralı ile gösterilir. Market sepet analizi denmesinin nedeni; bu tip birlikteliklerin modellenerek müşterilerin hangi ürünleri bir arada aldıkları bilgisinin ortaya çıkması ve market yöneticilerinin de bu bilgi ışığında daha etki satış stratejileri geliştirebilmesidir. Bağlantı Analizi aynı zamanda kişi davranışlarından hareketle sahtekârlık tespitinde, hangi davranışın hangi eylem ile beraber yapıldığında olağan dışı değerlendirilmesi gerektiğine öngörü sağlamaya da yardımcı olur.

c. Sıralı Dizi Analizi

Sayısal sıralı verilerin analizi ile gösterdiği trendleri saptanır ve bu trendlerin veri içindeki ilişkileri tanımlanır. Zaman ve mekân bağıntısı da göz önünde bulundurularak modelleme yapılır. Örneğin dondurma satışları ile kola satışları arasında pozitif, dondurma satışları ile salep satışları arasında negatif bir bağıntı beklenebilir.

d. İstisna Analizi

Normal davranışlardan ve eğilimlerden çok farklı sapmaları belirlemede kullanılır. Sahtekârlık analizlerinde en çok kullanılan tekniktir. Anomali denen davranışların tespit edilmesinde yarar sağlar. Örneğin, kasa işlemleri konusunda hazırlanan bir istisna analizi yapılan binlerce normal işlemin içerisinden sadece istisnai olanları ayıklayarak şüpheli işlemleri karar vericinin önüne getirebilmektedir. Bu da olası sorunlara odaklanmayı ve çözüme gitmeyi kolaylaştırmaktadır.

C. Veri Madenciliği Teknik ve Algoritmaları

Veri madenciliği istatistik ile birçok yönden çok yakın ilişki içindedir. Veri madenciliği ile istatistiğin ortak özelliği "veriden öğrenilmesi veya "verinin bilgiye dönüştürülmesi"dir.([45]) Her iki yaklaşım da verileri anlamlandırmak ve değerlendirmek ile ilgilenir. Bu sayede belirsizliklerin üstesinden gelmek ve gelecekteki olaylar hakkında bilgi vermek için kullanılır. Veri madenciliği süreci

[45] Jing Luan, Chun-Mei Zhao, "Practicing data mining for enrollment management and beyond" Volume 2006, Issue 131, sf: 117–122

içerisindeki birçok aşamada istatistikten yararlanmaktadır. Modelleme de ve doğrulama analizlerinde benzer teknikler uygulanır. Her ikisi de bir olayı etkileyen faktörleri saptayıp, bu faktörlerin modeller üzerinde deneyerek gelecekteki olayları daha iyi öngörmek ile ilgilenmektedir. Ortak paydalarına rağmen ayrıldıkları en önemli husus ise, veri madenciliğinin insan yargısının ve değerlendirilmesinin işin içine girdiği ve oluşturulan modellerin bilimsel yolla yorumlanarak bilgiye dönüşümünün sağlanmasında kesikli olmayan bir süreç izlemesidir. Zhao ve Luan her ikisi arasındaki ayrımı ortaya koymak için dört temel farklılıktan bahseder. Bunlar; teorinin rolü, genellenebilirlik, hipotez testi ve güven düzeyidir.([46]) İstatistik teori ile ortak yaşama ilişkisi içindedir. Teorinin rehberliği olmadan gözlemlerin ve olayların tamamını incelemek zor olur. Veri Madenciliği teorinin doğrulanması ile uğraşmaz, olayı yöneten analistten açık talimatlar bekler. Veri madenciliği tümden gelim, istatistik tüme varım ile ilgilenir.([47]) Veri madenciliğinin amacı daha detaylı, çok özel ve yerel verinin toplanması ve modeller yolu ile yorumlanmasıdır. Bu kapsamda Veri Madenciliğinin ana süjesi veri olmakla birlikte, verileri yorumlayarak bilgiye ulaşmada odağına insanı koyan bir çalışma sürecidir. Bununla birlikte bir olayın ele alınışında süreçlerde ve yaklaşımda farklar olması ile birlikte istatistik ile benzer teknik ve algoritmaları kullanır. Bu teknikler ve tanımlamaların bazıları şu şekildedir.

- Karar Ağaçları
- Regresyon
- Lojistik Regresyon
- Bayes
- Apriori
- Kümeleme Teknikleri
- Yapay Sinir Ağları

---

[46] Zhao Chung-Mei ve Luan, J. (2006) "Data Mining: Going Beyond Traditional Statistics", New Directions for Institutional Research, No. 131, sf:. 7–16.

[47] Selim Tüzüntürk "Veri Madenciliği ve İstatistik" Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt XXIX, Sayı 1, 2010, s. 72-73 iibfdergi.*uludag.edu.tr/download_pdf.php?id=25&f=25_rev1.pdf*

1. Karar Ağaçları

Karar ağaçları sınıfları bilinen örnek veriden tüme varım yöntemiyle öğrenilen ağaç yapısı şeklinde bir karar yapısı çeşididir. Tahmin edici ve tanımlayıcı özelliklere sahip olan karar ağaçları;

- Kuruluşlarının ucuz olması,
- Yorumlanmalarının kolay olması,
- Veri tabanı sistemleri ile kolayca entegre edilebilmeleri,
- Güvenilirliklerinin daha iyi olması

nedenleri ile yaygın kullanıma sahiptir.([48]),

Karar ağacı yapısı gereği karar düğümleri, dallar ve yapraklardan oluşur. Kullanıldığı alanlar: ([49])

- Belirli bir sınıfın olası üyesi olacak elemanların belirlenmesi,
- Çeşitli vakaların yüksek, orta, düşük risk grupları gibi çeşitli kategorilere ayrılması,
- Parametrik modellerin kurulmasında kullanılmak üzere çok sayıdaki değişkenden en önemlilerinin seçilmesi,
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulması,
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanması,
- Kategorilerin birleştirilmesi ve sürekli değişkenlerin kesikli değişkenlere dönüştürülmesinde çalışmalarında kullanılır.

Karar ağacı tekniğinde ağaç oluşturulduktan sonra, kurulacak algoritma kökten yaprağa doğru kurallar silsilesi şeklide ilerleyen bir yapıda kurgulanır. Bu şekilde kural çıkarma veri madenciliği çalışmasının doğrulanmasını sağlar. Sonuçların anlamlılığı karar verici ile denetlenir.

---

[48] Mustafa Aykut GÖRAL "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007-
AKPINAR, 2000, s. 17.
SPSS, "AnwerTree Algorithm Summary", SPSS White Paper, USA, 1999

[49] Van Dıepen, Merel ve Philip Hans Franses, "Evaluating ChiSquared Automatic Interaction Detection", Informatıon Systems, 31, No. 8, 2006.

Risk grupları kategorilerini belirlemek, gelecekte olabilecek olaylar için tahmin kuralları oluşturmak için kullanılır. Veri madenciliğinde en tipik kullanıldığı yerler; bireylerin kredi geçmişlerini değerlendirerek yeni kredi durumlarını belirleme, hangi değişkenlerin satışları etkilediğini belirleme ve üretim verilerini inceleyerek ürün hatalarına yol açan değişkenlerin tespit edilmesidir.

Örneğin; "Yaşı 40-50 arası, gelir düzeyi yüksek, son beş yıldır aynı işte çalışan bir kişinin kredi verilme düzeyi=mükemmel" gibi modeller çıkarılıp bu model test verisi ile test edilip sonuçları yorumlandıktan sonra yeni bir modele uygulanabilir.

Karar ağaçlarının, veri madenciliği uygulaması kapsamındaki güçlü yönleri;([50],[51])

- Analiz sonuçlarının yorumlanması kolaydır,
- Gelecekteki olayların tahmin edilebilmesi için kurallar oluşturulmasını sağlar
- Sadece belirli alt gruplara özgü olan ilişkilerin tanımlanmasını kolaylaştırır,
- Sonucu doğrulamak için işlem sayısı azdır, sınıflandırma yapmaya imkan sağlar,
- Sınıflandırma ve tahmin modelleri için önem taşıyan değişkenleri açıkça gösterir,
- Kategorik ve sayısal veriler üzerinde işlem yapmaya olanak sağlar.

Karar ağaçlarının zayıf noktası,

- Tahmin analizlerinde, öngörü yapılacak değişkenin sürekli değerler alması durumu ile karşılaşıldığında doğru sonucu verme olasılığı düşmektedir.

2. Regresyon Analizi

Regresyon analizi değişkenler arasındaki ilişkinin niteliğini saptamayı amaçlar, bir veya daha çok değişkenin başka değişkenler cinsinden tahmin edilmesini sağlayacak ilişkiler örüntüsünü tanımlamak olarak adlandırılan Regresyon analizlerinde matematiksel modelde yer alan değişkenler bir bağımlı değişken ve

[50] Mustafa Aykut GÖRAL "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007-

[51] Roiger, R.J. and Geatz, M.W., 2003. Data Mining: A Tutorial-Based Primer,Pearson Education Inc.,USA

bir veya birden çok bağımsız değişkenden oluşmaktadır. Değişkenler sayılabilir veya ölçülebilir niteliktedir. Regresyon analizi öncesi bağımlı ve bağımsız değişkenlerin niteliklerinin belirlenmesi büyük önem taşır, genellikle bağımlı değişken ölçülebilir nitelikte olup süreklilik göstermelidir. Regresyon analizleri, tek değişkenli doğrusal regresyon, tek değişkenli doğrusal olmayan regresyon ve çoklu regresyon gibi sınıflamalara ayrılır. İki değişken arasındaki ilişkiyi bulmak, ilişki varsa bu ilişkinin gücünü belirlemek, değişkenler arasındaki ilişkinin türünü belirlemek, ileriye dönük değerleri tahmin etmek gibi konularda kullanılır. Örneğin,"Ev sahibi olan, evli, aynı iş yerinde beş yıldan fazla çalışan, geçmiş kredilerinde geç ödemesi bir ayı geçmemiş bir beyaz yakalının kredi skoru 825'dir." sonucu bir regresyon ilişkisidir. ([52])

3. Lojistik Regresyon

Lojistik regresyon; bağımlı değişkenin süreklilik göstermemesi durumunda, ikili, üçlü ve çoklu kategorilerde gözlendiği durumlarda açıklayıcı değişkenlerle neden sonuç ilişkisini belirlemede yararlanılan bir yöntemdir. Lojistik regresyon analizinde bağımsız değişkenlerin, bağımlı değişken üzerine etkileri, bağımlı değişkenin iki düzeyinden herhangi birine karşı diğerinin olma olasılığından yararlanarak belirlenmeye çalışılır. Regresyon analizi yapabilmek için bağımlı değişken sürekli değere dönüştürülür. Bu değer beklenen olayın olma olasılığıdır. Lojistik regresyon analizinin kullanım amacı istatistikte kullanılan diğer model teknikleriyle aynıdır. En az değişkeni kullanarak en iyi uyuma sahip olacak şekilde bağımlı (sonuç) değişkeni ile bağımsız değişkenler kümesi (açıklayıcı değişkenler) arasındaki ilişkiyi tanımlayabilen modeli kurmaktır.

Örneğin Bursa Emniyet Müdürlüğünden 1990-2002 yılları arasına ait verilerden alınan polis suç veri tabanı kullanılarak yapılan lojistik regresyon modeli çalışmasında olay saati değişkeni için referans kategorisi 18.00-23.59 saatleri arasında işlenen suçlar kategorisidir. 12.00-17.59 arası işlenen suçların ağır suç

52 Doç. Dr. Kaan Yaralıoğlu (Dokuz Eylül Ünv. İktisadi ve İdari Bilimler Fakültesi) "Uygulamada Karar Destek Yöntemleri" Veri Madenciliği İlkem Ofset, İzmir, 2004

olma olasılığı 18.00-23.59 arasında işlenen suçların ağır suç olma olasılığına göre % 78 daha düşük bulunmuştur.([53])

4. Bayes

Eldeki verilerin belirlenmiş olan sınıflara ait olma olasılıkları üzerine kurulu bir modeldir. İstatistikteki Bayes teoremine dayanır. Bu teorem bir rassal değişken için olasılık dağılımı içinde koşullu olasılıklar ile marjinal olasılıklar arasındaki ilişkiyi gösterir. Bir olayın meydana gelmesinde, birden fazla etmenin olması koşulunda, olayın bu etmelerden hangisinin etkin hale gelmesi ile ortaya çıktığını gösteren bir teoremdir. ([54]) Bazı istatistikçiler için Bayes teoremi özel olarak değişik bir önem de taşır. Felsefi temelde olasılık değerlerinin nesnesel bir özellik değil, gözlemcinin meydana çıkardığı yeni kanıtlar ışığında olasılık değeri hakkındaki sübjektif inanışların, güncelleştirilip değiştirilmesini sağlayan bir yaklaşımın temeli olduğu esasına dayandırılır.([55])

Bayes kuralı, koşullu olasılıkların hesaplanmasında kullanılan bir kuraldır. Bir A olayının ortaya çıkmasında ikiden fazla olayın (faktör, seçenek, etken) etkisi varsa A olayı meydana geldiğinde faktörlerden birinin, faktörün gözlenme koşullu olasılığı Bayes kuralına göre hesaplanır. A olayı olduğu belirtildikten sonra bu olayın ortaya çıkmasında etkisi olan seçenekten faktörlerden birinin koşullu olasılığını hesaplamak için de Bayes kuralından yararlanılır.

Bayes teoremi,

- Basit bir yöntemdir,
- Örneklerin hangi sınıfa hangi olasılıkla ait olduklarını öngörme esasına dayanır,
- Öğrenme zamanı yoktur, yani sınıflandırmadan önce bir başlangıç zamanı istemez,

---

[53] Mehmet Nargeleçekenler Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, "Suç Veri Tabanının Lojistik Regresyon Analizi İle Tahmini: Bursa Örneği" http://www.pa.edu.tr/APP_DOCUMENTS/D478B2AD-3813-4555-9629-6332F8CF8D33/cms_statik/_dergi/2005/2/31-50.pdf (Erişim tarihi : 16.03.2014)

[54] http://bm.bilecik.edu.tr/Dosya/Arsiv/duyuru/bayesogrenmesi.pdf, erişim tarihi 11.03.2014

[55] http://tr.wikipedia.org/wiki/Bayes_teoremi erişim tarihi 11.03.2014

- Her sınıflandırma için tüm veri kümesini işler.

Örneğin; İki tabak dolusu bisküvi düşünülsün; Tabak 1 içinde 10 tane çikolatalı bisküvi ve 30 tane sade bisküvi bulunduğu kabul edelim. Tabak 2 içinde ise her iki tip bisküviden eşit miktarda 20'şer adet olduğu bilinsin. Evin küçük çocuğu rastgele bir tabağı seçip bu tabaktan rastgele bir bisküvi alsın. Çocuğun bir tabağı diğerine ve bir tip bisküviyi diğerine tercih etmekte olduğuna dair elimizde hiçbir gösterge bulunmamaktadır. Çocuğun seçtiği bisküvinin de sade olduğu tespit edilsin. Çocuğun bu sade bisküviyi tabak 1 den seçmiş olmasının olasılığının ne olacağı problemi Bayes ile incelenir. Sezgi ile tabak #1de sade bisküvi sayısının çikolatalı bisküvi sayısına göre daha fazla olduğunu göz önüne alınarak incelenen olasılığın %50den daha fazla olacağı hemen algılanır. Burada modelleme yapabilmek amacıyla soruyu Bayes teoremi uygulanabilecek şekle sokmak gerekmektedir: Çocuğun bir sade bisküvi seçmiş olduğu bilinmektedir; o halde bu koşulla birlikte tabak 1den seçim yapması olasılığı ne olacaktır?

Böylece Bayes teoremi formülüne uymak için A olayı çocuğun tabak 1'den seçim yapması; B olayı ise çocuğun bir sade bisküvi seçmesi olsun. İstenilen olasılık böylece O = (A/B) olacaktır ve bunu hesaplamak için; A olasılığı; hiçbir diğer bilgi olmadan çocuğun tabak 1'den seçim yapmasını; B olasılığı; hiçbir diğer bilgi olmadan çocuğun bir sade bisküvi seçmesi olasılığını, (diğer bir ifade ile bu çocuğun her bir tabaktan bir sade bisküvi seçme olasılığıdır.) Bu olasılık, önce her iki tabaktan ayrı ayrı olarak seçilen bir tabaktan bir sade bisküvi seçme olasılığı ile bu tabağı seçme olasılığının birbirine çarpılması ve sonra bu iki çarpımın toplanması suretiyle elde edilir, veriler Bayes teoreminin formülüne uygulandığında çocuğun sade bisküvi seçimi bilindiğine göre tabak 1'den alma olasılığı %60'dır denir. Bayes teoremi, belirsiz durumlarda tahmin yapmak için kullanılır.

Kullanım alanları, ([56])

[56] http://bm.bilecik.edu.tr/Dosya/Arsiv/duyuru/bayesogrenmesi.pdf, (erişihttp://web.itu.edu.tr/~sgunduz/courses/verimaden/slides/d3.pdf http://ceng.gazi.edu.tr/~ozdemir/teaching/dm/slides/06.DM.C2.pdf (erişim tarihleri 12.03.2014)

- Metin sınıflandırmada; özellikle kurumların büyük log analizlerinde,
- Konuşmacı tanıma sistemlerinde, ses verisi ile doğru kişiyi eşleştirme,
- Şifre kontrolü uygulamalarında, Internet bankacılığı veya kurumların şifre politikalarının uygulayan sistemlerde kullanılır.

Avantajları;

- Kolay uygulanabilir,
- İyi performans ile hızlı ve doğru sonuçlar verir,

En büyük dezavantajı değişkenler arası ilişkini modellenememesidir.

5. Apriori Algoritması

Veri Madenciliği teknikleri içerisinde yer alan birliktelik kurallarının esası, işlemlerden oluşan ve her bir işlemin elemanlarının birlikteliğinden oluştuğu düşünülen bir veri tabanında, bütün birlikteliklerin tarama ile ortaya çıkarılması ve sık tekrarlanan birlikteliklerin bağlantısını ortaya çıkarmaya dayanmaktadır. Apriori Algoritması, birliktelik kuralı çıkarım algoritmaları içerisinde en fazla bilinen algoritmadır.([57]) Rakesh Agrawal ve Ramakrishnan Srikant tarafından 1994 yılında geliştirilmiştir. Algoritmanın ismi, kendinden önceki çıkarımlara bağlı olduğu için, yani, bilgileri bir önceki adımdan almasından ötürü Latince, önce anlamına gelen "prior" kelimesinden gelmektedir. Algoritma yapı olarak, aşağıdan yukarıya yaklaşımı ile her seferinde tek bir elemanı incelemekte ve bu elemanla diğer verilerin ilişkisini ortaya çıkarmaya çalışmaktadır. Bu algoritmada sık geçen öğe kümelerini bulmak için birçok kez veritabanını taramak gerekir. İlk taramada tek elemanlı ve tanımlanan destek eşik değerini minimum olarak sağlayan veriler bulunur. İzleyen taramalarda bir önceki taramada bulunan sık geçen veriler aday veriler adı verilen yeni potansiyel sık geçen verileri üretmek için kullanılır. Aday verilerin destek değerleri tarama sırasında hesaplanır ve aday kümelerinden minimum destek eşik değerin sağlayan veriler o geçişte üretilen sık geçen veriler olur. Sık geçen veriler bir sonraki geçiş için aday veriler olurlar. Bu süreç yeni bir

[57] http://rakesh.agrawal-family.com/papers/vldb94apriori.pdf (erişim tarihi 14.03.2014)

sık geçen veri bulunmayana kadar devam eder. Algoritmanın adımlarını kısaca şöyle ifade edebiliriz;

- Minimum destek sayısı ve güven değerinin belirlenmesi
- Öge kümelerinde her bir ögenin destek değerinin bulunması
- Minimum destek değerinden daha küçük olanların devre dışı bırakılması
- Elde edilen birlikteliklerin ikili birlikteliklere dönüştürülmesi
- İkili birlikteliklerde minimum destek değerinden daha küçük olanların devre dışı bırakılması (tekrar)
- Elde edilen birlikteliklerin üçlü birlikteliklere dönüştürülmesi
- Üçlü birlikteliklerde minimum destek değerinden daha küçük olanların devre dışı bırakılması (tekrar)
- Üçlü birlikteliklerin kurallarının çıkarılması ([58])

Apiori algoritması genellikle depo sistemlerinin optimizasyonunda veya büyük marketlerde ürünler arası ilişkileri tanımlamakta kullanılabilir. Veri madenciliği tekniklerinin üretim yerlerinde en sık uygulandığı alanlardan birisidir. Genel olarak depo ürün yerleştirme optimizasyonunda birlikte sevk edilen ürünlerin yakın raflara yerleştirilmesi, barkodlu okuma sistemlerinin buna göre konumlandırılması depo içindeki hareketi minimize edeceği gibi, araçların yükleme zamanını eleman kullanım maliyetini de optimize edici sonuçlar sağlamakta ve maliyet tasarrufuna katkı sağlayacak sonuçlar alınmaktadır.

6. Yapay Sinir Ağları

Biyolojik sistemlerde öğrenme, nöronlar arasındaki sinaptik (*synaptic*) bağlantıların ayarlanması ile olur. Yani, insanlar doğumlarından itibaren bir yaşayarak öğrenme süreci içerisine girerler. Bu süreç içinde beyin sürekli bir gelişme göstermektedir. Yaşayıp tecrübe ettikçe sinaptik bağlantılar artar ve yeni bağlantılar oluşturur. Bu sayede öğrenme sürekli devam eder. Yapay Sinir

[58] Nesibe Yalçın Emre Güngör "Veri Madenciliği Uygulamaları Sunumu" Araştırma Görevlisi, Bilecik Üniversitesi Mühendislik Fakültesi on Nov 16, 2012 http://www.slideshare.net/nesibeyalcin/apriori-algoritmas-15209923 (erişim 14.03.2014)

Ağlarında da öğrenme, eğitme yoluyla örnekler kullanarak olur; başka bir deyişle, gerçekleşme girdi/çıktı verilerinin işlenmesiyle, yani eğitme algoritmasının bu verileri kullanarak bağlantı ağırlıklarını (*weights of the synapses*) bir yakınsama sağlanana kadar, tekrar tekrar ayarlamasıyla olur. Yapay Sinir Ağları, ağırlıklandırılmış şekilde birbirlerine bağlanmış birçok işlem biriminden (nöronlar) oluşan matematiksel sistemlerdir. Genelde, işlem birimleri kabaca gerçek nöronlara karşılık gelirler ve bir ağ içinde birbirlerine bağlanırlar; bu yapı da sinir ağlarını oluşturur.

Yapay sinir ağları (Artificial Neural Networks – ANN) insan beyninin bilişsel öğrenme sürecinin benzetimini yapmak amacıyla ilk olarak 1940'ların başında ortaya atılmıştır. Deneme yanılmaya dayalı olarak problem uzayının bir modelini yapılandırmak olan bilgisayar-tabanlı algoritmalardır. Modelin yapılandırılması süreci kavramsal olarak bir kısım veri setinin yapay sinir ağına gönderilmesiyle başlar ve yapay sinir ağı bir çıktı değeri tahmin eder. Bu tahmin değeri, gerçek değer ile bir geri bildirim biçimi ile karşılaştırılır. Eğer tahmin doğru ise ağ başka bir faaliyet göstermez. Eğer tahmin yanlış ise yapay sinir ağı, kestirimin kalitesini arttırmak için hangi iç parametrelerin ne şekilde düzeltileceğini belirlemek için kendi kendisini analiz eder. Bu parametre ayarlamaları yapıldıktan sonra yapay sinir ağı veri setinin başka bir bölümünü alır ve süreci tekrar eder. Bu süreç içerisinde zamanla yapay sinir ağı büyük oranda doğru bir modele yakınsamaya başlar. ([59]) Yapay sinir ağlarının temel birimi düğüm'dür(node). Düğüm ağ içerisindeki diğer düğümlerle paralel bir şekilde hareket eden işlem birimidir ve işlev olarak insan beynindeki nöronlara benzemektedir. Her düğüm bir diğerine belirli bir ağırlıkla (weight) bağlıdır. Herhangi bir düğüm diğer düğümlerden belli miktarda bir girdi alır ve bu girdiyi ilgili ağırlık değeri ile çarparak yeni toplamalı (summated) değer yaratır. Söz konusu değer harekete geçirme fonksiyonu (activation function) tarafından çıktı değeri üretmek üzere işleme alınır ve sıradaki düğüme gönderilir. Girdi olarak modele sunulan birimler, sayısal değerlerin 0 ile 1

[59] George M. Marakas, *Kelley School of Business, Indiana University* - See more at: http://www.pearsonhighered.com/educator/product/Modern-Data-Warehousing-Mining-and-Visualization-Core-Concepts

arasında normal değerlere dönüştürülmüş hali, kategorik değerlerde ise 0 ve 1 değerleri kullanılarak kodlanmış halidir.([60])

Buna göre yapay sinir ağlarda iki tür düğüm olduğunu söylemek mümkündür: girdi sağlayan düğümlerden oluşan bir girdi tabakası (input layer) ve çıktı düğümlerden oluşan bir çıktı tabakası (output layer). Bununla birlikte çoğu yapay sinir ağda genellikle bir tane –bazen de birden fazla- gizli tabaka (hidden layers) da bulunabilmektedir. Gizli tabaka girdi ve çıktı arasında doğrudan ilişki olmadığı durumlarda ilişkiyi açıklamak için kullanılmaktadır. Yapay sinir ağları sonuç itibariyle çoklu regresyon, diskriminant analiz ve kümeleme analizi gibi bilinen istatistik yöntemlerden büyük bir farklılık göstermemektedir. Etkin bir şekilde kullanabilmesi için model tasarımında ağ içerisindeki düğümlerin birbirlerine nasıl bağlanacakları, kaç düğüm kullanılacağı ve deneme çalışmalarının ne zamana kadar devam ettirileceği konularına dikkat edilmesi gerektiği belirtilmektedir. Yapay sinir ağlar yüzlerce ya da binlerce girdi olduğu zaman iyi sonuçlar verememektedir. Fazla sayıda girdi desenin keşfedilmesini zorlaştırmakta ve modeli oluşturmak için kullanılan deneme aşamasının uzamasına yol açmaktadır. Bu nedenle bu tür durumlarda yapay sinir ağların karar ağaçları ile birlikte kullanılması önerilmektedir. Buna göre karar ağacı tekniği ile bağımlı değişkene etki eden önemli faktörler tespit edilecek yapay sinir ağ modeli de tüm değişkenleri kullanmak yerine söz konusu değişkenleri kullanarak sonuca ulaşacaktır. Yapay sinir ağları, insanların deneyimlerinden bir takım bilgiler çıkartması gibi kendisine verilen örneklerden bir takım bilgiler çıkartma yeteneğine sahiptir. Her şeyden önce bir veri kümesi üzerinde öğrenme algoritmaları çalıştırılarak eğitilir. Bu eğitim neticesinde yapay sinir ağının içerisindeki bir takım ağırlıklar belirlenir. Bu ağırlıklar kullanılarak yeni gelen veriler işlenir ve bir sonuç üretilir. Yapay sinir ağlarının en olumsuz tarafı; bu ağırlıkların neden ilgili değerleri aldıklarının bilinmemesidir. Çıkan sonucun nedenleri açıklanamamaktadır. Bu olumsuz tarafı, ABD'de yapay sinir ağlarının, kredi taleplerinin değerlendirilmesinde kullanılmasının yasaklanması sonucunu doğurmuştur. Çünkü sistem bir kişiye

[60] Mark W. Craven, Jude W. Shavli" Using neural networks for data mining" (14.04.2014)

kredi vermeme sebebini açıklayamayacaktır. Kişinin riskli gözükmesinin sebebi ten rengi veya ırkı olabilir. Bu sorunun cevabının alınamayacak oluşu, böyle bir uygulamayı gerektirmiştir.([61]) Bu dezavantajlarına karşın özellikle finans kesiminde kredi kartı kullanımında sahtekarlıklarını saptamak ve kredi riskini hesaplamak amacıyla faydalanılmaktadır.([62]), ([63]) Bunun yanı sıra, Bankacılıkta, imza tanımada, kredi uygulamalarını geliştirmede, Otomotiv sektöründe; yol izleme, rehberlik, yol koşullarına göre sürüş analizi, üretim sanayiinde; ürün tasarımı, kalite kontrol, müşteri tahmin analizleri, konuşmayı yazıya çevirme teknolojilerinde kullanılmaktadır.

III. Veri Madenciliği Tarihçesi, Dünya'da ve Türkiye'de Uygulama Alanları

A. Veri Madenciliği Tanımları ve Gelişimi

Bilginin değer kazanmaya başlaması ile bilgi yönetiminde bilişim teknolojilerinin kullanımının artması bu alanda yapılan yatırımlarında artmasına ve öncelik verilmesine neden olmuştur. Günümüzün verilere boğulan dünyasında bu verilerin enformasyon ve bilgiye dönüşüm süreci bu teknolojilerin gelişmesini de tetiklemiş bir anlamda yumurtamı tavuktan tavuk mu yumurtadan çıkar anlayışına paralel bir gelişim süreci hayatımızı etkiler olmuştur. Makro düzeyde olaya baktığımızda bu veri yığınları her geçen gün artmakta ve bu teknolojileri kullanan herkes bu yığın artışına bir anlamda katkı sağlar duruma gelmiştir. Bu kadar veri arasında anlamlandırma işinin önemi veri madenciliği teknolojilerinin de gelişmesini sağlamaktadır. Günümüzde sadece bilgiye ulaşmak değil, gerekli koşulların

[61] Mustafa Aykut GÖRAL "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007 sf 50

[62] Bilgehan Gürünlü "Yapay Sinir Ağları" http://www.iszekam.net/post/2009/05/21/Is-Zekasinda-Kullanilan-Veri-Madenciligi-Modelleri-4.aspx (erişim tarihi 16.03.2014)

[63] Prof: Martin Hagan http://ecee.colorado.edu/academics/schedules/ECEN5120.pdf (erişim tarihi 16.03.2014)

oluşması ile bilgi üretmek de önemli bir konudur. Büyüyen veri ortamları arasından yararlı, anlamlandırılmış ve gerekli bilgiye ulaşmak bir değer haline gelmiştir.
Veri madenciliği bu safhada işe değer katan bir olgudur. W.J. Frawley veri madenciliğini "Daha önceden bilinmeyen ve potansiyel olarak yararlı olma durumuna sahip verinin keşfedilmesi ve yorumlanması" olarak tanımlamıştır.
Berry ve Linoff bu tanıma"Anlamlı kuralların ve örüntülerin bulunması için geniş veri yığınları üzerine yapılan keşif ve analiz işlemleri" açıklaması ile işin teorisine katkı sağlarken, Sever ve Oğuz konuya ilişkin olarak yaptıkları çalışmalarında "Veri yığınlarından önceden bilinmeyen, gizli, anlamlı ve yararlı örüntülerin büyük ölçekli veri tabanlarından elde edilmesini sağlayan veri tabanlarında bilgi keşfi süreci içerisinde yer alan bir adımdır." diye tanımlamışlardır. Literatürde ayrıca Veri Madenciliği büyük veri setleri veya veritabanlarından yararlı bilgi çıkarma bilimi olarak da tarif edilmiştir. Bunun yanı sıra literatürde geçen çeşitli tanımlarda şu şekildedir:
Jacobs (1999), veri madenciliğini, ham datanın tek başına sunamadığı bilgiyi çıkaran, veri analizi süreci olarak tanımlamıştır.([64])
Doğan ve Türkoğlu, "Veri madenciliği, büyük veri yığınları arasından gelecekle ilgili tahminde bulunabilmemizi sağlayabilecek bağlantıların, bilgisayar programı kullanarak aranması işidir."([65])
Hand, veri madenciliğini istatistik, veritabanı teknolojisi, örüntü tanıma, makine öğrenme ile etkileşimli yeni bir disiplin ve geniş veritabanlarında önceden tahmin edilemeyen ilişkilerin ikincil analizi olarak tanımlamıştır.([66]) R. Kittler ve W. Wang, veri madenciliğini oldukça tahminci anahtar değişkenlerin binlerce

[64] Jacobs, P., (1999), "Data Mining: What General Managers Need to Know", Harvard Mng.Vol:4

[65] Doğan, Ş. ve Türkoğlu, İ., (2008), "Iron-Deficiency Anemia Detection From Hematology Parameters By Using Decision Trees",International Journal of Science & Technology, Vol: 3, 85-92.

[66] Hand, D.J., (1998), "Data Mining: Statistics and More", The American Statistician, Vol: 52, sf : 112-118.

potansiyel değişkenden izole edilmesini sağlama yeteneği olarak tanımlamışlardır.([67])

Matematikçilerin 1950'li yıllarda mantık, bilgisayar bilimleri alanlarındaki yapay zekâ (artificial intelligence) ve makine öğrenme (machine learning) konularında çalışmaya başlaması veri madenciliğine giden yolun taşlarının döşenmeye başlandığı yıllardır. Veri madenciliği, kavramının 1960'lı yıllarda, bilgisayarların gelişme sürecinin hızlanması ve veri analiz problemlerini çözmek için kullanılmaya başlamasıyla hız kazandığı kabul edilmektedir. 1960'ların sonunda bilim adamları basit öğrenmeli bilgisayarlar geliştirebilmişlerdir. Minsky ve Papert, günümüzde sinir ağları olarak bilinen perseptron'ların sadece çok basit olan kuralları öğrenebileceğini göstermişlerdir. ([68])

O dönemlerde, bilgisayar yardımıyla, yeterince uzun bir tarama yapıldığında, istenilen verilere ulaşmanın mümkün olacağı gerçeği görülmüş, bu işleme veri madenciliği yerine önceleri veri taraması (data dredging), veri yakalanması (data fishing) gibi isimler verilmiştir.([69])

1970, 1980'li yıllar yeni programlama dilleri, bilgisayarların mimari yapılarında oluşan değişiklikler, yeni işlemcilerin çıkması kişisel bilgisayarların hayatımıza girmesi ve bunların veri işleme kapasitelerinde yaşanan donanımsal değişiklikler; veri madenciliğinin gelişmesinin yapısını hazırlayan unsurlar olarak karşımıza çıkmaktadır. Bunlara ilave olarak genetik algoritmalar (genetic algorithms), kümeleme yöntemleri (clustering methods) ve karar ağaçları (decision tree algorithms) gibi algoritmaların gelişmesi de bu sürece katkı sağlayan gelişmeler olmuştur. 1990'lı yıllarda veritabanlarında yaşanan köklü değişim, özellikle ilişkisel veri tabanlarının gelişmeye başlaması, veri ambarları kavramlarının bilişim dünyasında yerini alması ile veri madenciliği kavramının içi dolmaya başlamıştır.

---

[67] Kittler R. ve Wang W., (1998), "The Emerging Role of Data Mining", Solid State Technology, Volume:42 sf: 110-111

[68] Adriaans, P. ve Zantinge, D., (1997), Data Mining, , Boston, MA, USA Addison Wesley Longman Publishing.

[69] Sertaç Öğüt "Veri Madenciliği kavramı ve gelişim süreci" Görsel İletişim Tasarımı Bölümü, İletişim Fak. Yeditepe Ünv.

Bu dönemde veri tabanlarından bilgi keşfi olarak karşımıza çıkan kavram, büyük miktarlarda veri içerisinden gizli kalmış, kullanılabilir ve ilişkilendirilebilir verilerin çıkarılarak yorumlanması ve stratejik hedeflere varmada taşıyıcı bir unsur olarak kullanılabileceğini kanıtlamıştır. Araştırmacılar bu muazzam büyüklükteki veritabanlarını herhangi bir araç kullanmadan analiz etmenin güçlüğünü de yansıtacak bilgi keşfi kavramını kullanmaya başlamışlardır. Veritabanından bilgi keşfi; veriden yararlı bilgi çıkarma sürecidir.([70])

1989 yılında yapılan "KDD (IJCAI)-89 Veritabanlarında Bilgi Keşfi Çalışma Grubu" toplantısı ile bu yöndeki gelişmeler hız kazanmaya başlamıştır. 1991 yılında bu çalışma grubunun sonuç bildirgesi olarak görülen "Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop" raporu ile veritabanlarında bilgi keşfi ile ilgili temel tanım ve kavramlar ortaya konmuştur.([71]) Bu durum süreci hızlandıran kilometre taşlarındandır. 1992 yılında veri madenciliği için ilk yazılımın ortaya çıkması ile doğum süreci olarak kabul edilebilecek aşamanın tamamlanarak, gelişme aşamasına geçildiği kabul edilebilir. 2000'li yıllar artık veri madenciliğinin uygulama alanlarını her geçen gün arttırdığı yıllar olmuştur.

B. Veri Madenciliği Uygulama Alanları

Büyük hacimde veri bulunan ve karar verme sürecine ihtiyaç duyulan her yerde veri madenciliğini kullanmak mümkündür. Örneğin pazarlama, biyoloji, bankacılık, sigortacılık, borsa, perakendecilik, telekomünikasyon, genetik, sağlık, bilim ve mühendislik, kriminoloji, endüstri, istihbarat vb. birçok dalda uygulama alanı bulmaktadır. Kaynaklar incelendiğinde en çok sağlık, biyoloji, genetik ve finans, bankacılık, perakende ve telekomünikasyon sektörlerinde veri madenciliğinin

[70] Abdulkadir Özdemir, Fulya Yalçın Aslay, Handan Çam," Veri Tabanında Bilgi Keşfi Süreci: Gümüşhane Devlet Hastanesi Uygulaması" sf:348

[71] Serkan SAVAŞ, Nurettin TOPALOĞLU, Mithat YILMAZ "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri" İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl:11 Sayı: 21 2012 sf: 5

uygulandığı görülmektedir.([72]) Bunun yanı sıra, müşteri ilişkileri yönetiminde, firmaların satış, bankacılıkta finansal göstergelere ilişkin gizli ilişkilerin bulunmasında, pazarlamada müşterilerin satın alma örüntülerinin belirlenmesinde ve sigortacılıkta ise riskli müşterilerin örüntülerinin belirlenmesinde sıklıkla kullanılmaktadır.

Kullanım alanları şu başlıklar altında toplanabilir;

- Veri tabanı analizi ve karar verme desteği
- Pazar Araştırması ve Analizinde
- Risk Analizi
- Belgeler arası benzerlik: haber kümeleri, e-posta
- Müşteri kredi risk araştırmaları
- Kurum kaynaklarının en optimal biçimde kullanımı
- Geçmiş ve mevcut yapı analiz edilerek geleceğe yönelik tahminlerde bulunma
- Bilimsel ve mühendislik alanında
- Sağlık alanında
- Ticaret alanında
- Alışveriş alanında
- Bankacılık ve finans alanında
- Eğitim sektöründe
- Internet alanında
- Doküman kullanımında
- Sigortacılıkta kullanılmaktadır.

Veri Madenciliğinin kullanım alanlarına kategorik olarak bakacak olursak;

- Müşteri İlişkileri Yönetimi (Customer Relationship Management, CRM): Bütün müşterilerin e-mail, işlem, çağrı merkezi ve anket gibi erişim noktalarından elde edilen metin bilgilerinden nitelikli bilgi çıkarılır. Bu nitelikli bilgi müşterinin terk etme ve çapraz satışlarını

[72] Serkan SAVAŞ, Nurettin TOPALOĞLU, Mithat YILMAZ "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri" İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl:11 Sayı: 21 Bahar 2012 s. 1-23

tahmin etmek üzere kullanılır.

- Sahtekârlık tespiti: Sağlık, sigorta, bankacılık, güvenlik ve istihbarat kısacası her alanda oluşabilecek anormallikler aranarak sahtekârlıklar tespit edilir. Özellikle sağlık sigortası sektöründe yapılan sahtekârlık analizlerinde, bankacılık sistemindeki kredi kartı ve ATM sahtekârlıklarının tespitinde kullanılmaktadır.
- Bilimsel, Medikal ve Tıp Alanındaki Araştırmalar: Hasta raporları, makale başlıkları, yayınlanmış araştırma sonuçları ve diğer yayınlar gibi metin materyallerinden çıkarım yapılır. Sağlık alanı Veri Madenciliğinin Dünya'da en çok kullanıldığı alanların başında gelmektedir. Özellikle tarama testlerinden elde edilen verileri kullanarak çeşitli kanserlerin ön tanısı, kalp verilerini kullanarak kalp krizi riskinin tespiti, acil servislerde hasta semptomlarına göre risk ve önceliklerin tespiti gibi çok geniş bir uygulama sahası söz konusudur.([73])
- Finans sektöründe Risk Analizleri: Müşteri profili belirlenir, kredibilite ve risk değerlendirmesi yapılır, ürün segmentasyon ve penetrasyonları buna göre yapılır. Bankalar veri madenciliği tekniklerini kredi kartı satışlarında müşterilerin davranış ve güvenilirliklerini ölçmek ya da belirli bir müşterinin ödemelerini aksatma ihtimalini öngörmek amacıyla da kullanabilir.
- Güvenlik/istihbarat: Organizasyonlar ve bireyler arasındaki kalıplar ve bağlantılar, terörist atak tehlikeleri, kriminal davranışları tahmin etmek ve engelleyebilmek için büyük çaptaki metin, konuşma, sosyal içerikli veriler içerisinde örüntüler aranır.
- Eğitim sektörü: Öğrenci işlerinde veriler analiz edilerek öğrencilerin başarı ve başarısızlık nedenleri, başarının arttırılması için hangi

[73] Dalkılıç, G., Türkmen, F., "Karınca Kolonisi Optimizasyonu", YPBS2002 – Yüksek Performanslı Bilişim Sempozyumu, Kocaeli, Ekim 2002.

konulara ağırlık verilmesi gerektiği, üniversite giriş puanları ile okul başarısı arasında bir ilişkinin var olup olmadığı gibi soruların cevabı bulunarak eğitim kalitesi ve performansı arttırılabilir.

- Pazar araştırması: Yayınlanmış belgeler, basın bültenleri ve web sayfaları pazar etkisinin ölçülmesi için aranır ve izlenir. Bu alanda en çok başvurulan veri madenciliği yaklaşımı sepet analizidir. Sepet analizinde amaç alınan ürünler arasındaki ilişkileri bulmaktır. Bu ilişkilerin bilinmesi işletmenin kârını arttırmak için kullanılabilir. Ayrıca müşteri profili belirlemede hem pazar araştırması açısından, hem finansal risk değerlendirmesi açısından hem de ürün analizi açısından Veri Madenciliği içinde yapısal olmayan verinin değerlendirildiği web ve metin madenciliği kantitatif yöntemler ile açık uçlu anket soruları ve mülakatların değerlendirilmesinde kullanılabilmektedir.([74])

1. Veri Madenciliğinin Türkiye'deki Örnekleri

Veri madenciliği sadece bir yöntem ve bu yöntemler için geliştirilen teknikler olarak değil, probleme özgü tasarlanmış, yöntem, teknik ve uygulamaları da içine alan, sonuçları itibariyle probleme özgü ilişkileri, kuralları, örüntüleri, eğilimleri, analiz eden, modelleyen ve ortaya çıkaran bir süreç olduğundan ortaya konan bu ihtiyaçların sonucunda akıllı veri tabanı analizi için yeni teknikler doğmuştur. ([75]) Veriyi akıllı biçimde işleyerek otomatize edilmiş süreçler sonunda işe yarar bilgiye dönüştürebilecek teknikler üzerinde çalışmalar başlamıştır.([76]) Bu durum işin içinde insan unsurunun olmayacağı anlamına gelmemelidir. Her şeyden önce

[74] M. Özgür Dolgun, Tülin Güzel Özdemir, Doruk Oğuz "Veri madenciliği'nde yapısal olmayan verinin analizi" İstatistikçiler Dergisi , 2 (2009) 48-58

[75] Kantardzic, Mehmed, Data Mining: Concepts, Models, Methods, and Algorithms, IEEE Press, Hoes Lane, Piscataway, NJ, USA, 2003.

[76] Serkan SAVAŞ, Nurettin TOPALOĞLU, Mithat YILMAZ "Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri" İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl:11 Sayı: 21 Bahar 2012 sf:2

verilerin modellemelerini yapan, aynı modele farklı teknikler uygulayarak kıyaslamasını yapan, bu kıyaslamalardan yorum çıkaran ve yoruma göre somut eylemlere dönüştüren gene insandır. Tüm bunların sonucunda veri madenciliği sorunun ortaya konmasından itibaren, cevaplarını da sunmak için gelişmesini sürdürmektedir. Veri Madenciliğine yönelik Ülkemizde yapılan uygulama örneklerinden bazıları ana başlık altında şöyle ele alınabilir:

a. Medikal ve Tıp Alanındaki Araştırmalar

Bu kısımda Türkiye'de Veri Madenciliği uygulamalarının kullanıldığı bazı örnekler detayına girilmeden ana başlıklar altında anılmaktadır. Ülkemizde bilişim teknolojilerinin gelişmesinin hız kazandığı son on yıllık dönemde Dünya'daki örneklerine benzer çalışmalar yapılmaktadır. Bu kısımda Veri Madenciliğinin Ülkemizde kullanımı açısından dikkat çekmek amacıyla bu örneklerden bazıları alınmıştır. ([77])

- Tıp alanında 2009 yılında Dalgıçların Acil Durum Ağı'nın dalış yaralanmaları bildirim formlarından elde edilen belirti ve bulgu listeleri kullanılarak dekompresyon hastalığı sınıflandırılmış ve elde edilen sonuçlar, yeni yapılan istatistiksel sınıflandırma yöntemleri ve tedavi sonuçları ile karşılaştırılmıştır.
- Pınar Yıldırım, Mahmut Uludağ ve Abdülkadir Görür tarafından 2008 yılında yapılan çalışmada, hastane bilgi sistemlerindeki veri madenciliği uygulamalarına değinilmiştir.
- Şengül Doğan ve İbrahim Türkoğlu tarafından 2008 yılında gerçekleştirilen bir çalışmada, biyokimya laboratuvarlarından elde edilen sonuçlara göre demir eksikliği anemisinin teşhisinde doktorlara karar almada kolaylık sağlayacak bir model geliştirilmiştir. Bu çalışmada sistem veri madenciliği tekniklerinden karar ağaçları yapısı kullanılmıştır.

[77] Sezgin Irmak, Can Deniz Köksal, Özcan Asilkan "Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi" Uluslararası Alanya İşletme Fakültesi Dergisi Yıl:2012, C:4, S:1, s.101-114

- Mustafa Danacı, Mete Çelik ve A. Erhan Akkaya tarafından 2010 yılında gerçekleştirilen çalışmada; kadınlar arasında en sık görülen meme kanseri türlerinden birine ait Xcyt örüntü tanıma programı yardımı ile doku hakkında genel veriler elde edilmiş, Weka Veri Madenciliği yazılımı ile meme kanseri hücrelerinin tahmin ve teşhisi yapılmıştır.
- 2011'de Sabri Serkan Güllüoğlu tarafından kanser hastalığının erken teşhisinde veri madenciliği ilkelerinden hareketle hastalığa yakalanmış ya da yakalanması muhtemel kişiler için kullanılabilecek bir erken teşhis yaklaşımını ortaya koyacak bir veri madenciliği çalışması yapılmıştır. Bu çalışma; kanser hastalığının erken teşhisinde yapay sinir ağları yönteminin kullanılması ve Türkiye bazında kanser haritası çıkarılması hedeflenmiştir. Kanser hastalığının erken teşhisinde yapay sinir ağları yönteminin kullanılmasının sağlıklı bireylerin gelecekte kansere yakalanma olasılıklarına karşı önlem alınmasını sağlayacak hipotezine yanıt arayan bir model çalışmasıdır.
- Yapay sinir ağları kullanarak meme kanseri tanı testlerinin değerlendirilmesi yönünde benzer bir çalışma da Ahmet Sertaç Sunay, Beyza Kaymakoğlu, Umut Arıöz tarafından yapılmıştır. 2012 yılında "Hastanelerin Gelecekteki Hasta Yoğunluklarının Veri Madenciliği Yöntemleri İle Tahmin Edilmesi" konulu çalışmada; Sezgin Irmak, Can Deniz Köksal, Özcan Asilkan tarafından işleyen bir hastane veritabanındaki veriler, veri madenciliği tekniklerinden yapay sinir ağları ve zaman serileri kullanılarak analiz edilmiş ve model çalışmaları yapılmıştır. Bu sayede, gelecekteki hasta yoğunluklarının tahmin edilmesi amaçlanmış bunların da kendi içinde farklı modelleri üretilerek gelecekteki hasta yoğunluklarının tahmin edilmesi ve bu konuda en iyi modelin belirlenmesi amaçlanmıştır.

b. Bankacılık ve Borsa Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları

- Ali Sait Albayrak ve Şebnem Koltan Yılmaz tarafından 2009 yılında gerçekleştirilen çalışmada, o zamanki adıyla İMKB 100 endeksinde sanayi ve hizmet sektörlerinde faaliyet gösteren 173 işletmenin 2004–2006

yıllarına ait yıllık finansal göstergelerinden yararlanarak veri madenciliği tekniklerinden birisi olan karar ağaçları tekniği uygulanmıştır. Seçilen finansal göstergelere göre sanayi ve hizmet sektörlerinde faaliyet gösteren firmaları ayıran en önemli değişkenler saptanmıştır.([78])

- Ali Sait Albayrak tarafından gerçekleştirilen başka bir çalışmada, yerli ve yabancı olarak önceden grup üyeliği belirlenmiş bankaların sınıflandırmasında yaygın olarak kullanılan veri madenciliği tekniklerinden, diskriminant, lojistik regresyon ve karar ağacı modellerinin birbirine göre karşılaştırılması yapılmıştır. Bu karşılaştırmada; bankalarla ilgili seçilmiş likidite, gelir-gider, karlılık ve faaliyet oranları kullanılmıştır. Araştırmanın sonuçları, bankaların sınıflandırmasında karar ağacı modelinin geleneksel diskriminant ve lojistik regresyon modellerine üstünlük sağlayarak alternatif etkili bir sınıflandırma tekniği olarak kullanılabileceğini göstermiştir.([79])
- Nihal Ata, Erengül Özkök ve Uğur Karabey tarafından 2007 yılında yapılan bir çalışmada,(65) bir bankanın kredi kartı sahiplerine ait veri kümesi üzerinde yapılan regresyon analizi sonucu müşterilerin yaş, gelir ve medeni durumunun kredi kartı kullanımından vazgeçmelerindeki risk faktörleri olduğu tespit edilmiştir.
- 2007 yılında Mustafa Aykut Göral tarafından kredi kartı başvuru aşamasında sahtekârlık tespiti için birliktelik kurallarının uygulandığı bir veri madenciliği çalışması yapılmıştır.

c. Eğitim Alanında Veri Madenciliği Uygulamaları

- 2009 Yılında Esen Yıldırım ve Yasemin Koldere Akın tarafından yapılan bir çalışma "Veri Madenciliğinde Kümeleme Teknikleri ile Ortaöğretim Öğrencilerinde Şiddet Eğiliminin Tespiti" yapılmıştır. İstanbul'da öğrenim

[78] Yrd.Doç.Dr.Ali Sait Albayrak, Öğr.Gör.Şebnem Koltan Yılmaz, "Veri Madenciliği: Karar Ağacı Algoritmaları Ve İmkb Verileri Üzerine Bir Uygulama" Süleyman Demirel Üniversitesi, İktisadi ve İdari Bilimler Fak. Dergisi Y.2009, C.14, S.1 s.14

[79] Yrd. Doç. Dr. Murat Çinko Marmara Ünv. İ.İ.B.F." Kredi Kartı Değerlendirme Tekniklerinin Karşılaştırılması "http://efinans.co/wp-content/uploads/2013/11/225-234.pdf sf 213 (erişim 01.03.2014)

gören 3468 ortaöğretim öğrencisinden elde edilen verilerle, öğrenciler şiddet eğilimlerine göre gruplandırılmaya ve bu eğilime yol açan sebepler ortaya çıkarılmaya çalışılmıştır. Araştırmanın kısıtlarına uyan merkeze dayalı bölümleyici ve hiyerarşik kümeleme tekniklerinden; k-ortalama, iki aşamalı kümeleme ve CLARA yaklaşımları ile elde edilen kümelenmeler karşılaştırılarak, bu tekniklerin üstün ve zayıf yönleri incelenmiştir. ([80])

- 2007 Yılında Y. Ziya Ayık, Abdülkadir Özdemir ve Uğur Yavuz tarafından yapılan çalışmada, Atatürk Üniversitesi öğrencilerinin mezun oldukları lise türleri ve lise mezuniyet dereceleri ile kazandıkları fakülteler arasındaki ilişki, incelenmiştir.
- Ahmet Selman Bozkır, Ebru Sezer ve Bilge Gök tarafından gerçekleştirilen bir çalışmada, ÖSYM tarafından 2008 ÖSS adayları için resmi internet sitesi üzerinden yapılan anket verileri üzerinde veri madenciliği yöntemleri kullanılmış ve öğrencilerin başarılarını etkileyen faktörler araştırılmıştır. Bu araştırmada, veri madenciliği yöntemlerinden karar ağaçları ve kümeleme kullanılmıştır.
- Murat Karabatak ve Melih Cevdet İnce tarafından yapılan çalışmada ise "Apriori Algoritması İle Öğrenci Başarısı Analizi" için Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar Eğitimi bölümü öğrencilerinin notları kullanılarak öğrenci başarılarının analizi yapılmıştır. Bu analizi yapmak için, birliktelik kuralı çıkarım algoritmalarından biri olan Apriori algoritması kullanılmıştır. Çalışmada kullanılan öğrenci notları, Fırat Üniversitesi öğrenci işleri birimindeki SQL veri tabanı üzerinde gerekli dönüşümler yapılmış ve Apriori algoritması kullanılmıştır.

80 Esen Yıldırım, Y. Koldere Akın,"Veri Madenciliğinde Kümeleme Teknikleri ile Ortaöğretim Öğrencilerinde Şiddet Eğiliminin Tespiti Üzerine Bir Analiz" Marmara Üniversitesi İ.İ.B.F. Ekonometri Böl. Trakya Üniversitesi İ.İ.B.F. Ekonometri Böl. İstatistik Anabilim Dalı http://asosindex.com/journal-article-abstract?id=12517 (erişim 18.03.2014)

d. Mühendislik ve İş Alanında Gerçekleştirilen Veri Madenciliği Uygulamaları

Üretim sektöründe, firmaların kaliteli ürünü, uygun maliyete üretip uygun fiyata satarak rekabet sağlaması, sürdürülebilir rekabet avantajı sağlayacak üretim metotları deneyebilmesi için veri madenciliği teknikleri yoğunlukla kullanılmaktadır. Bunun yanı sıra katma değerli ürünlerdeki rekabet şansını pozitif yönde arttıran değer zincirindeki temel unsurlardan biri olarak kabul görmektedir. Üretim sektöründe;

- Talep planlama
- Kalite geliştirme
- Süreç kontrolü ve optimizasyonu
- Önleyici bakım hizmetleri
- Lojistik ve tedarik
- Üretim planlama ve kontrol
- Depo yönetimi
- Müşteri ilişkileri yönetimi

Konularında uygulama alanı bulmaktadır.

- Emel, Taşkın ve Kılıçarslan'ın 2004 yılında yaptığı "çelik üretim sürecinin sinir ağları ile analizi uygulaması",
- 2005 Yılında Baykaşoğlu'nun "çimento sektöründe yapay sinir ağları ve regresyon analizi uygulaması",
- Özçakır ve Çamurcu'nun 2007 yılında yaptığı "yiyecek sektöründe birliktelik kuralları analizi uygulaması",

- Feyza Gürbüz, Lale Özbakır ve Hüseyin Yapıcı'nın 2008 yılında gerçekleştirdiği Türkiye'de bir hava yolu işletmesinin parça söküm raporları üzerinde veri madenciliği çalışması bu tür uygulamalara örnek verilebilir.([81])

[81] Selim Tüzüntürk "Veri Madenciliği ve İstatistik" Uludağ Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt XXIX, Sayı 1, 2010, s. 65-90

2. Veri Madenciliğinin Dünya'daki Örnekleri

Bilişim teknolojilerinin bilgi sistemleri yönetiminde kullanılmaya başlanmasına Ülkemizden çok önceleri başlayan ABD ve diğer ülkelerde veri madenciliğine yönelik çok sayıda araştırma ve uygulama yapılmıştır. Bunlardan özellikle son on yıl içinde yapılanlardan bazıları şöyledir; Kusiak ve Smith (2007) ürün tasarımında ve üretim sistemlerinde, Chien ve Chen (2008) beşeri sermaye yönetiminde, Liao, Chen ve Hsu (2009) spor firmalarına yönelik araştırmalarda,, Ogwueleka (2009) bankacılık sektöründe müşteri ilişkileri yönetiminde, Zhang, Ma, Zhang ve Wang (2009) elektrik mühendisliği, Kumar ve Uma (2009) öğrenci performansı tespit çalışmalarında, Naveh, Sariri ve Zadeh (2009) elektrik santrali jeneratörü, Gervilla, Cajal, Roca ve Palmer (2010) alkol tüketimi, Liang (2010) otomotiv sektöründe müşteri değeri yaratmada, Ayesha, Mustafa, Sattar ve Khan (2010) yüksek eğitim sistemi, Rebbapragada, Basu ve Semple (2010) gelir yönetimi, Deshpande, Gogolak ve Smith (2010) ilaç güvenliği tespitlerinde, Srinivas ve Harding (2010) mimari alanda, Seng ve Cheng (2010) iş kararı, Abdelmelek, Saidane ve Trabelsi (2010) yağların biyo-çözünürlüğü konusunda, Glasgow ve Kaboli (2010) tıbbi kayıtlar üzerinden ilaç etkileşim vakaları ile ilgili çalışmalar yapmıştır.([82])

a. Medikal ve Tıp Alanındaki Araştırmalar

Dünya'da veri madenciliğinin en çok kullanım alanı tıp ve medikal alanlarda olmaktadır. Özellikle hastalıkların önceden teşhisi ve pandemik yayılmaların önüne geçmede yıllardır örnek çalışmalar yapılmıştır. Sağlık bilgi sistemlerinin amacı büyük miktardaki sağlık verilerinden faydalı bilgi üretmektir. Bu bilgiler hasta düzeyinde daha iyi sağlık hizmeti sunumu, sağlık kurumlarının daha iyi yönetilmesi, kaynakların etkin kullanımı ve sağlık politikalarının oluşturulması amaçları ile kullanılmaktadır.

---

[82] Justin M. Glasgow, Peter J. Kaboli,"Detecting adverse drug events through data mining" Health Syst.Pharm Vol 67, Feb 15, sa:313
http://www.ajhp.org/content/67/4/317.extract erişim tarihi 31.03.2014)

- İngiltere'de yönetimin çocuk sağlığı programlarının kalitesini arttırma ve bütçelerinin karşılanma oranlarının tespiti için uzun soluklu bir çalışma yapılmıştır.
- Boehringer Ingelheim İtalya dünyadaki ilk 20 ilaç şirketinden birisidir. Merkezi Almanya'da bulunan ve 140'dan fazla bağlı şirkete sahip Ingelheim insan ve hayvan sağlığına yönelik ilaçların üretimi ve pazarlamasında uzmanlaşmıştır. Boehringer Ingelheim İtalya, ticari faaliyetleri çeşitlendirmek ve daha hedefe yönelik ve etkin satış stratejileri geliştirmek amacıyla eczanelerden oluşan müşteri tabanını sınıflandırmak amacıyla veri madenciliği yöntemlerini kullandığı bir çalışmayı hayata geçirmiştir.([83])
- San Francisco Kalp Enstitüsü, her yıl dünyanın pek çok yerinden gelen yüzlerce kalp hastasının tedavi gördüğü önemli bir sağlık kuruluşudur. Veri madenciliği çalışması ile,tedavi süreçleri, hastanın demografik ve yapısal özellikleri ile tedaviye alınan cevaplar analiz edilerek, hastaların hastanede kalma süreleri azaltılmıştır. Aynı çalışmada hasta ölüm analizleri ve ölüm riski tespit edilerek getirilen çözümler ile ölüm riski azaltılmıştır. Ayrıca sağlık sigorta şirketleri ile gelir analizleri paylaşılarak daha kişiye daha uygun poliçe tipleri önerilmesi sağlanmıştır.
- A.Kusiak ve arkadaşları tarafından akciğerdeki tümörün iyi huylu olup olmadığına dair, karar destek amaçlı bir çalışma yapılmıştır. İstatistiklere göre ABD'de sık görülen kanser türlerinden olan akciğer kanserine yönelik 160.000 den fazla akciğer kanseri vakası incelenerek, bunların %90'ının öldüğü belirlenmiştir. Bu bağlamda tümörün erken ve doğru olarak teşhisi amacıyla test verileri arasında yapılan veri madenciliği çalışmaları teşhiste %100 oranında doğruluk sağlamıştır.

[83] Mehmet Berkay Can, Eren Çamur, Mine Koru, Ömer Özkan, Zeynep Rzayeva," Veri Kümelerinden Bilgi Keşfi: Veri Madenciliği" sf:11
http://tip.baskent.edu.tr/egitim/mezuniyetoncesi/calismagrp/ogrsmpzsnm14/14.P8.pdf
(erişim tarihi 01.04.2014)

- Kore Tıbbi Sigorta Kurumu (The Korea Medical Insurance Corporatition) tarafından yapılan bir veri madenciliği çalışması ile yüksek tansiyonun tespitinde yaşam koşullarının (diyet, alınan tuz miktarı, alkol, tütün gibi) hiçbirinin tahminde etkili olmadığı yalnızca yaşın etkili olduğu saptanmıştır.
- Vysis firması ilaç geliştirme çalışmalarında yürüttüğü protein analizleri üzerine sinir ağları tekniklerini kullandığı bir çalışmayı hayata geçirmiştir.
- Rochester Üniversitesi kanser araştırma merkezinde de kanser vakalarının tespitine yönelik karar ağaçları tekniklerinin kullanıldığı çalışmalar yapılmıştır. ([84])

b. Diğer sektörlerde yapılan çalışmalar

Veri madenciliği diğer sektörlerde de kullanılmakta ve kullanımı da giderek artmaktadır. Genel olarak baktığımızda sağlık sektörünün dışında dünyadaki en çok uygulama örneği aynı ülkemizde olduğu gibi, finans ve bankacılık, sigorta, telekomünikasyon alanlarındadır.

- Bir sigorta şirketi olan Farmer's Group Inc. veri madenciliğini "spor arabası olan bir kişinin yüksek kaza riski yoktur" senaryosunu bulmak için kullanmıştır. Senaryonun şartları, spor arabanın ikinci araba olmasını ve aile arabasının bir station wagon veya sedan olmasını gerektirmektedir.([85])
- Bank of America ise veri madenciliğini, müşterilerin kendi ürün tercihleri arasındaki dağılımı tespit etmek ve böylece müşteri ihtiyaçları ile örtüşen doğru ürünleri ve servisleri önerebilmek için kullanmaktadır.[86]
- New York'taki Chase Manhattan Bankası müşterilerini rakiplerine kaybetmeye başlayınca, müşteri hesaplarını analiz etmek ve kendi hesap

[84] Mustafa Aykut GÖRAL "Kredi Kartı Başvuru Aşamasında Sahtecilik Tespiti İçin Bir Veri Madenciliği Modeli – Yük. Lisans Tezi İ.T.Ü. Fen Bilimleri Ens.. Ocak 2007 sf:69

[85] Roiger, R.J. and Geatz, M.W., 2003. Data Mining: A Tutorial-Based Primer,Pearson Education Inc.,USA.sf:111

[86] http://www.databaseskill.com/983961/ (erişim tarihi : 01.04.2014)

gereksinimlerinde değişiklikler yapabilmek için veri madenciliği kullanmaya başlamış, bu sayede karlı müşterilerini elinde tutabilmiştir. (71)

- Sinema endüstrisinde, Twentieth Century Fox film şirketi, satılan biletlerin analizi ile hangi bölgede hangi film, hangi aktörün daha çok rağbet gördüğüne yönelik bir çalışma ile yeni tüketiciler için eğitim belirleme ve bölgesine göre gösterilecek film fragmanlarını tespit etmede veri madenciliği tekniklerini kullanmıştır. (82)
- American Express kredi kartı şirketi 5.4 milyar müşteri verisinin analizi ile promosyon stratejisini belirlediği bir çalışma yapmıştır.
- ABD'de golf özellikle belli gelir seviyesindeki kişilerin rağbet ettiği bir spordur. Bazı finans şirketleri televizyondan yayınlanacak golf turnuvalarına sponsor olmaktadırlar. Bunun nedeni; 40 yaş üstü erkeklerin %70'inin golf sevdikleri ve bunlar içinde de %60'lık bir kesimin aynı zamanda hisse senedi yatırımları ile ilgilendiklerini bir veri madenciliği çalışması ile tespit etmelerinden kaynaklanmaktadır.
- Spor endüstrisi açısından bir başka örnek ABD NBA liginde Toronto Raptors'un yardımcı antrenörünün, uygun oyuncu eşleştirmelerini bulmak için veri madenciliği tekniklerinden yararlanmasıdır.
- Yiyecek endüstrisinde faaliyet gösteren Kraft Foods Inc., 30 milyon adet müşterisine ait veri üzerinde gerçekleştirdiği veri madenciliği çalışması ile yeni çıkaracağı ürün ve sağlıklı yiyeceklere yönelik promosyon yönetimi kampanyasını yürütmüştür. ([87])
    - Kablolu TV, kablosuz iletişim ve ABD'nin Batı eyaletlerinde yerel telefon hizmetleri veren US West Communication Inc. Firması, aile boyutu, ortalama aile bireyi yaşı ve konum gibi özelliklere dayanan müşteri eğilimlerini ve ihtiyaçlarını belirlemek, büyüme trend analizleri

[87] http://www.databaseskill.com/983961/ erişim tarihi (1.04.2014)

ve yeni müşterileri firmaya kazandırmak için veri madenciliğini kullanmaktadır.

- AT&T firması sahtekârlık tespitleri amacıyla digital santral yapısı üzerinden uluslararası sahtekârlık işbirliklerini ortaya çıkarabilmek için veri madenciliği tekniklerini uyguladığı bir sistem kullanmaktadır.

IV. Veri Madenciliğinin Güvenlik Uygulamalarında Kullanımı ve Sahtekârlık Analizinde Veri Madenciliği

Sahtekârlık, kontrolü ve önlenmesi zor olan önemli bir sorundur. Ortaya çıkarılması kolay olmamakta, her bir olayın kendine özgü argümanlarının titiz, kapsamlı ve hızlı algoritmalar yardımıyla doğru bir şekilde tanımlanması gerekmektedir. Ayrıca her bir sahtekârlık olayına ilişkin ayrı bir sınıflandırma gerekmektedir. Çalışma hayatında görülen sahtekârlık suçları – daha nazikçe bir ifadeyle yolsuzluklar – beyaz yakalıların işleyebileceği türden suçlardır. Özel veya kamuya ait pek çok kuruluşun, özellikle mali konularla ilgili sahtekârlıklarla, yolsuzluklarla başı derttedir. Sahtekârlıklar neticesinde, dünya genelinde her yıl ekonomide milyarlarca dolarlık kayıplar yaşanmaktadır. Teknolojik gelişmelerin iş hayatına kazandırdığı ivme, beraberinde sahtekârlık ve hilenin de daha karmaşık, farkına varılmasını ve tespitini güç hale getirmiştir. Bu hile, sahtekârlık ve yolsuzluklar, ister özel sektör olsun ister ise kamu sektörü kurumların itibarını zedelemekte, duyulan güveni azaltmakta daha makro bir bakış ile ülke ekonomisine zarar vermekte, itibarını, güvenilirliğini etkilemekte, maddi kayıpların yanı sıra yatırımlar açısından riskli ülke konumuna gelmesine neden olmaktadır. Günümüzde sahtekârlık ile mücadelede veri madenciliği çalışmaları önemli bir yer tutmakta, kullanılan metot ve algoritmalar ile sahtekârlık analizinde çalışan konunun uzmanlarının yüzünü güldürmektedir. Bu bilgi yoğunluklu çalışmanın yapıldığı bir süreçtir. Sahtekârlık ilgili aktiviteler hakkındaki deliller, muazzam miktardaki veri yığınlarında dağıtık biçimde gizlenmiş bulunduğundan, bu teknikler, hem proaktif olarak yolsuzlukları oluşmadan önlemede hem yapıldıktan sonra tespitinde etkili olmakta, yolsuzluk yapanların delilleriyle birlikte yargıya

teslim edilmesini kolaylaştırmaktadır. Kullanılan metot ve teknikler, büyük boyuttaki veriler üzerinde hızlı analizler yapılarak, suç unsuru ve ögelerinin kolayca saptanması, denetim altına alınmasını kolaylaştırmıştır. Veri madenciliğinin kullanılması iş modellerinde ve iş yapma biçimlerinde de verim artışını beraberinde getirmiştir.

Ayrıca pek çok ülkede özellikle ABD'de terörist faaliyetlerinin tespitinde çok etkin kullanılan bir yöntemdir. 11 Eylül saldırılarının birinci yılında ABD NSA'nın teröristlerin davranış örüntülerini tespit etmek amacıyla tüm ABD vatandaşlarının kayıtlarını, telefon konuşmalarını depoladığı bilgisini, basında ile paylaşmıştı. ABD terörizmi engellemek amacıyla MATRIX, TIA, ADVISE1, TALON, ATS vb isimlerde çeşitli programlar başlattı. Bu programlar terörizmle mücadele stratejilerinde önemli bir teknik değişikliğe işaret etmekteydi, bu değişiklik, özne temelli analizden örüntü temelli analize geçiliyordu. Özne temelli analiz; şüpheli kişilerin bağlantıları üzerinde bir analizdir. Örüntü temelli analiz ise şüpheden çok öngörüsel bir yapı içerir. Dolayısıyla telefon dinlemeleri şüphe temelli analizken, basında yer alan kayıtlarının ve telefon konuşmalarının depolanarak analiz edilmesi örüntülerden yola çıkan ve veri madenciliği teknik ve algoritmalarının kullanıldığı bir analizdir. Bunun yanı sıra başta ABD ve Çin olmak üzere, uzaya uydu gönderen pek çok ülke,

- Uydulardan gelen uzaysal veri ve görüntülerin değerlendirilerek hakkında bilgi alınmak istenen askeri unsurların, nerelerde hangi araç ve teçhizatlar da yoğunlaştığının ve konuşlanmaya uygun arazi yapılarının belirlenmesi
- Internet sayfalarının anahtar kelimelerle taranarak lehte ve aleyhte propaganda yapan sayfaların belirlenmesi
- Haberleşme araçları takip edilerek terörist faaliyetlerin önlenmesi amacıyla belirlenmesi konularında Veri madenciliği uygulamalarını kullanmaktadır. Veri madenciliğinin ülkelerdeki güvenlikle ilgili

alanlardaki kullanımlarını, aşağıdaki gibi bir sınıflama yapmak mümkündür. ([88]);

- Dolandırıcılık, hile ve sahtekârlıkların belirlenmesi
    - Şüpheli hareketlerin Veri Madenciliği metotları ile belirlenmesi
    - Bunların analiz sonuçlarını değerlendirecek karar vericilere aktarılması
        - Sigorta sahtekârlıkları
        - Kredi sahtekârlıkları
        - Vergi dolandırıcılıkları
        - İlaç ve reçete sahtekârlıkları
        - Kara para aklama faaliyetleri
        - Kurum içinde yapılan suiistimal ve finansal sahtekârlıklar
- Terörist suç gruplarının belirlenmesi
    - Telefon görüşmelerinin analizi
    - E-posta yazışmaları ve sosyal medya takipleri
    - Suç kayıtlarının eşleştirilmesi
    - Tüm bilgilerin ağ topolojisine dökülmesi
    - Her kişinin ağ üzerinde bir node olarak belirlenmesi
    - Kişiler arasındaki linklerin topoloji üzerinde kurulması
    - Linklerin özellikleri ve ağırlıklarının çıkarılması
    - Ağ içerisindeki yoğun kümelerin bulunması
- Bilgi Sistemleri Güvenliği
    - Anomali saldırı tespit sistemleri
    - Spam ve zararlı içerik taşıyan e-postaların tespiti ve önlenmesi
- Finans ve Bankacılık Güvenlik Uygulamaları
    - Tüm müşteri bilgilerinin analiz edilmesi
    - Kredi taksitlerini ödemeyenlerin profillerinin çıkarılması

[88] Hayrettin Bahşi "Veri Madenciliğinin Güvenlik Uygulamaları ve kişisel gizlilik" Tubitak 6.06.2008 sunumu (erişim 14.04.2014)

- Yeni kredi taleplerinin, taksitlerini ödemeyen grubuna dahil olan müşteri veritabanındaki profillerle karşılaştırılması
- Kredi sağlayıp sağlamama kararı (kişi veya kurum)

Veri madenciliği yolsuzlukların denetimi, risklerin değerlendirmesi, müşteri analizlerinin çıkarılması, beklenmeyen, olağandışı veya anormallik arz eden tutum ve davranışların ortaya konması, veri analiz araçlarının kullanılarak daha önceden bilinmeyen mantıklı ilişki ve kalıpların bulunup çıkarılması amacıyla kullanılmaktadır. Ülke güvenliği bağlamında, para transferi ve iletişimi, vergi kaçakçılığı, kara para aklama vb. terörist faaliyetlerinin tanımlanmasında yararlanılmakta, telekomünikasyon, seyahat, nakliye, kargo, havayolu şirketlerine ait kayıtlar arasından terörist faaliyetlere ilişkin bireysel izlerin tanımlanmasında kullanılmaktadır. ([89])

Sahtekârlıkla mücadele için yapılan bir veri madenciliği çalışması aşağıda belirtilen aktiviteleri tam olarak kapsamalıdır.

- Erken uyarı ve alarm,
- Yolsuzluğa ait değişik tipteki bulgu ve kalıpların ihbarı,
- Kullanıcı ve aktivite profilleri,
- Yolsuzlukları araştırma, izleme ve önleme,
- Yanlış alarmların minimize edilmesi ve bu konuda oluşabilecek mağduriyetlere sebebiyet vermemesi,
- Kayıpların değerlendirilmesi, risk analizleri,
- Gözetim ve örüntüleri izleme,
- Güvenlik (bilgisayarların, verilerin, ağ sisteminin ve fiziki araçların güvenliği),
- Veri ve kayıt yönetimi, elektronik ortamdan ve diğer kaynaklardan delillerin toplanması,
- Yönetim bilgi sistemleri ve operasyon sistemleri için linkler (muhasebe ve faturalama gibi),

[89] www.alomaliye.com (Erişim : 03.04.2014)

- Kontrol faaliyetleri (kovuşturma, personel eğitimi, etik programları, direkt telefonlar, partnerlerle işbirliği ve icracı kurumlar).
- Raporlar, özetler, veri görüntülemeleri,

Proaktif olarak sahtekârlıkla mücadele kapsamında yapılan çalışmalarda karşılaşılan sıkıntılar ise;

- Muazzam miktardaki ve kompleks yapıdaki verinin varlığı,
- İş aktivitelerinin, yolsuzluğu gerçekleştirenlerin ve kullanıcıların davranışlarındaki sürekli değişim,
- Yolsuzluğu ortaya çıkarmada kullanılan mevcut tekniklerin sürekli değişen yolsuzluk teknikleri karşısında bypass edilememesinin yaratacağı riskler,
- Yolsuzlukları hızlı ve doğru bir şekilde önleme ihtiyacının iş akışını gereksiz yere kesintiye uğratma olasılığı,
- Yanlış eşik değerlerinin yarattığı yanlış alarmlar,
- Özel hayatın gizliliği gibi sosyal yaşamı ilgilendiren konular ve buradaki ayrımları hukuki açından netleştirmedeki yasal eksiklikler

Olmaktadır.

A. Hile, Sahtekârlık, Dolandırıcılık Kavramlarının Hukuki Yönü

İngilizce'deki "Fraud" kelimesi; Latincede hasar, yanlış yapma, aldatma tanımları için kullanılan "Fraus" kelimesinden türemiştir.([90]) Uluslararası literatürde kullanılan "fraud" teriminin Türkçe karşılığı olarak farklı tanımlar kullanılmakla birlikte, yapılan akademik çalışmalarda ve literatürde "Fraud" kelimesinin Türkçe'de; sahtekarlık, hile, yolsuzluk, dolandırıcılık kelimelerine karşılık kullanıldığı, bu kelimelerle ifade edildiği görülmektedir.

Sahtekârlık, hile, yolsuzluk, usulsüzlük gibi düzensizlikler, kasıtlı olarak yapılan davranışlardır. Sahtekârlık, çalışan, kurum personeli, yönetimden sorumlu kişiler

[90] Howard Silverstone, Michael Sheetz, Forensic Acounting and Fraud Investigation for NonExperts,2. bs, New Jersey, John Wiley&Sons Inc., 2007, s.3-4

ya da üçüncü kişilerin bilinçli olarak menfaat sağlamak amacıyla aldatma içeren davranışlarda bulunmasıdır.([91])

Bir olayın sahtecilik sayılabilmesi için aşağıdaki koşulların yerine gelmiş olması beklenmektedir:

- Eyleme dair bir sunuş (ifade) vardır
- Maddi çıkar söz konusudur
- Eylemin yanlışlığı sabittir
- Eylemi ifâ edenler, eylemin yanlışlığına dair bilgi sahibidir.
- Eylemi ifâ edenlerin, bu eylemin mağdura karşı işlenmesi gerektiğine dair niyetleri vardır.
- Mağdur, eylemin sahtecilik olduğunu fark etmemiştir.
- Mağdur, sunuluşu itibariyle eylemde bir sorun olduğunu fark etmemiştir.
- Mağdur, sunuluşu itibariyle eyleme güvenmede haklıdır.
- Mağdur, eylemin sonuçlarından zarar görmüştür.

240 numaralı Uluslararası Bağımsız Denetim Standartları metninde de yolsuzluk ve hile kelimesi; yönetim kademesindeki bir ya da birden fazla çalışanın veya üçüncü şahısların bilgileri kasıtlı olarak örtbas etmesi ya da haksız kazanç elde etmesi olarak tanımlanmıştır.([92])

Dilimizde yine sahtekârlık anlamında da kullanılan dolandırıcılık, Türk Dil Kurumu tanımına göre; "birini aldatarak mal veya parasını alma durumudur".

Türk Hukuk Sisteminde dolandırıcılık ve hile suçu karşılık bulmaktadır. 5237 sayılı Türk Ceza Kanunu'nun 157. maddesinde dolandırıcılık suçu düzenlenmiştir. Buna göre dolandırıcılık; "hileli davranışlarla bir kimseyi aldatıp, onun veya başkasının

---

[91] Ersin Güredin, Denetim ve Güvence Hizmetleri SMMM ve YMM'lere Teknikler, 11. bs. İstanbul, Arıkan Basım, Ocak 2007, s.134.

[92] Türkiye Serbest Muhasebeci Mali Müşavirler ve Yeminli Mali Müşavirler Odaları Birliği, TÜRMOB, Uluslararası Denetim ve Güvence Standartları, TÜRMOB Yayınları, No:339, s.195. (UDS 240) ve , http://www.kgk.gov.tr/contents/files/BDS/BDS_240.pdf (erişim 04.04.2014)

zararına olarak, kendisine veya başkasına yarar sağlamak" şeklinde tanımlanmıştır.([93])

Nevzat Toros'lu "Dolandırıcılık suçu kaynağını hırsızlıktan alan, ticaret ve sanayideki çağdaş ve aynı zamanda küresel, karmaşık ekonomik ilişkilerin ortaya çıkardığı bir suç tipidir" şeklinde tanımlamaktadır.([94])

Türk Ceza Hukukunda; eylem terimi; hareket, netice ve bu ikisi arasında nedensellik bağından oluşur. Ceza hukuku anlamında hareket, suç tanımına uyan ve suç tanımındaki neticeyi gerçekleştirmeye yönelik bulunan ve iradi olan davranıştır.([95]), ([96])

Dolandırıcılık suçunun maddi unsurunun hareket kısmını hileli davranışlar oluşturur. Dolandırıcılık suçu, kişilerin mal varlığına karşı işlenen bir suçtur. Bu suçun işlenişi sırasında hileli davranışlar ile kişiler aldatılmaktadır.([97]) Çok hareketli suç görüntüsü taşıyan dolandırıcılık suçunun oluşumu, beraberinde birden fazla fiilin gerçekleşmesini gerektirir. Bu hareketlerden birincisini hile oluşturmaktadır. İkincisi ise gerçekleştirilen hile neticesi, buna maruz kalan kişinin veya bir üçüncü kişinin zararına olarak, fail veya bir başkasının menfaat sağlaması, çıkar elde etmesidir.([98]) Dolandırıcılık suçu herkes tarafından işlenebilecek genel

---

[93] TCK 5237 Madde 157 - (1) "Hileli davranışlarla bir kimseyi aldatıp, onun veya başkasının zararına olarak, kendisine veya başkasına bir yarar sağlayan kişiye bir yıldan beş yıla kadar hapis ve beşbin güne kadar adlî para cezası verilir

[94] Nevzat TOROSLU, Ceza Hukuku Özel Kısım, Ankara 2008, s. 173. CGK., 24.11.1998-6-280/359, Mater KABAN-Halim AŞANER- Özcan GÜVEN-Gürsel YALVAÇ, Yargıtay Ceza Genel Kurulu Kararları (Eylül 1996 - Temmuz 2001), Ankara 2001, s. 614.

[95] Selahattin Keyman, "Cürmi Fiilin Yapısal Unsuru Olarak Hareket", AÜ.Huk.Fak.D, C. 40, S. 1-4 (1988), s. 63-95;

[96] Selahattin Keyman, Suç Genel Teorisinin İki Temel Sorunu: Genel ve Soyut Hareket Kavramı, Suçun İncelenmesinde Tekçi ve Tahlilci Yöntemler", Prof. Dr. Fadıl H. Sur'un Anısına Armağan, Ankara 1983, s. 431

[97] Ali Karagülmez, Bilişim Suçları ve Soruşturma – Kovuşturma Evreleri, Ankara 2.Baskı, S.181

[98] Esra Yırtımcı, "Dolandırıcılık Suçu" Ankara Üniversitesi Sosyal Bilimler Enstitüsü Kamu Hukuku (Ceza Ve Ceza Usulü Hukuku) Yük. Lisans Tezi-2010 sf:113 http://www.eticaretvehukuk.com/319/bilisim-yoluyla-dolandiricilik-sucu.html (erişim tarihi: 11.04.2014)

bir suç olduğundan, faili de herkes olabilir. Bununla birlikte failin sıfatı, örneğin, şirket yöneticisi, şirket adına hareket eden yetkili kişi olması suçun nitelikli hali olarak düzenlenmiştir.

TCK 157. maddenin gerekçesinde; bu suçun maddi unsurunu oluşturan hareketlerden ilkinin hileli davranışlar olduğu belirtilmiş olmasına rağmen; diğer hareketlerin ne olduğu konusunda açıklama yapılmamıştır.([99])
Hile kavramı, adli muhasebecilik ve bağımsız şirket denetimleri açısından da kabul gören bir kavramdır.([100])
Hile; düzen, oyun, aldatma demektir. Objektif olarak hataya düşürücü, başkasının tasavvuru üzerinde etki doğurucu her davranış hiledir.

TCK'nın 158 m. f. b'de; Dolandırıcılık suçunun Bilişim Sistemlerinin, banka veya kredi kurumlarının araç olarak kullanılması suretiyle, işlenmesi bu suçun bir nitelikli unsuru olarak düzenlenmiştir.([101])

Madde gerekçesinde, "Bilişim sistemlerinin ya da birer güven kurumu olan banka veya kredi kurumlarının araç olarak kullanılması, dolandırıcılık suçunun işlenmesi açısından önemli bir kolaylık sağlamaktadır. Banka ve kredi kurumları açısından dikkat edilmesi gereken husus bu kurumları temsilen, bu kurumlar adına hareket eden kişilerin başkalarını kolaylıkla aldatabilmeleridir" denilmiştir.([102]) , ([103])

---

[99] Esra Yırtımcı, "Dolandırıcılık Suçu" Ankara Üniversitesi Sosyal Bilimler Enstitüsü Kamu Hukuku (Ceza Ve Ceza Usulü Hukuku) Yük. Lisans Tezi-2010 sf: 21

[100] Yrd. Doç. Dr. Canol Kandemir, Yrd. Doç. Dr. Şenol Kandemir Çağ Üniversitesi, İİBF Muhasebe Hilelerinin Önlenmesi ve Ortaya Çıkarılmasında Kullanılan Araç ve Yöntemler sf:40

[101] TCK m.158/1-f " Dolandırıcılık suçunun; Bilişim sistemlerinin, banka veya kredi kurumlarının araç olarak kullanılması suretiyle, İşlenmesi halinde, iki yıldan yedi yıla kadar hapis ve beşbin güne kadar adlî para cezasına hükmolunur."

[102] http://www.eticaretvehukuk.com/319/bilisim-yoluyla-dolandiricilik-sucu.html
http://www.guclusan.av.tr/nitelikli-dolandiricilik-tck-158/ (erişim tarihi 01.04.2014

[103] Murat Ülkü, Çorum Cumhuriyet Savcısı, "Yeni Ceza Adalet Sistemi"nin Cumhuriyet Savcıları ve Hakimlere tanıtımı için 21.02.2005-01.04.2005 tarihleri arasında Adalet Bakanlığı Eğitim Dairesi Başkanlığı'nca Samsun ilinde düzenlenen seminer sunumu.(http://www.ceza-bb.adalet.gov.tr/makale/162.pdf erişim 04.04.2014)

Örneğin, internetten alış veriş yaparken başkasının kredi kartını kullanmak, kredi kart bilgisi ele geçirilmiş kişi adına işlemler yapmak, kredi talebini kimliğini ele geçirdiği kişi üzerine onun haberi olmadan yapmak, dolandırıcılık örneklerinden bazılarıdır.

Dolandırıcılık suçunun işlenmesinde fail, tüzel kişinin yararına haksız menfaat sağlar ise, bu tüzel kişi hakkında, tüzel kişilere özgü güvenlik tedbirine hükmolunur. Fakat fail tüzel kişinin bir hukuki alacağını tahsil edip bu yolla dolandırıcılık yaparsa bu durumda tüzel kişi için güvenlik tedbiri uygulanmaz. Bu durum TCK 169 m.'de düzenlenmiştir.([104])

Tezin konusu, veri madenciliği ile sahtekârlık analizi kapsamında olduğundan tez içerisinde dolandırıcılık, hile, yolsuzluk kavramları yerine, uluslararası literatürde karşılığını bulduğu şekli ile Sahtekârlık", "Sahtecilik" kavramları kullanılacaktır.

Sahtekârlığın bilişim sistemleri üzerinden işlenmesi ile suçun nitelik kazanması, ağırlaştırıcı neden olarak değerlendirildiğinden bu tür durumlara karşı banka, sigorta şirketleri, ticari ve sanayii kurumlar, kamu kurumlarının güven unsuru olarak değerlendirildikleri, kendilerine emanet edilen para, kişisel veri, itibar konularında hassas davranıp bu tür verileri sakladıkları sistemlerine yönelik her türlü güvenlik önlemini alma yükümlülükleri vardır. Bu sistemlerin korunması konusunda uygulanacak standartlar kadar önleyici, caydırıcı ve tespit edici anlamda veri madenciliği uygulamaları da bu kapsamda değerlendirilmelidir.

B. Bilişim Suçları ve Sahtekârlık

Bilişim, bilişim sistemi, bilgisayar gibi temel kavramlar Türk Hukuk Sisteminde yasal olarak tanımlanmış kavramlar değildir. Bilişim kavramı teknik, ekonomik, mali, sosyal, kültürel, hukuksal veya toplumsal yaşamın benzeri birçok alanda sahip olunan verilerin saklandığı, saklanan bu verilerin elektronik olarak işlendiği,

[104] TCK'nun 169. maddesine göre; tüzel kişilere güvenlik tedbiri uygulanabilmesi için, failin işlediği suç sonucunda tüzel kişi yararına haksız bir menfaat sağlanması şarttır. Bu nedenle, failin, tüzel kişinin hukuki bir alacağını tahsil amacıyla dolandırıcılık suçunu işlemesi halinde; tüzel kişi yararına sağlanan menfaat haksız olmadığından, bu tüzel kişi hakkında güvenlik tedbiri de uygulanmaz.

organize edildiği, değerlendirildiği ve yüksek hızlı veri, ses veya görüntü taşıyan iletişim araçları ile aktarıldığı sistemleri tanımlamak için kullanılmaktadır. Bu haliyle bilişim bilginin saklanması, işlenmesi ve iletilmesini konu edinmiş akademik ve mesleki bir disiplindir.([105])

Bilişim sistemi, TCK m. 243'ün gerekçesinde "verileri toplayıp yerleştirdikten sonra bunları otomatik işlemlere tabi tutma olanağı veren manyetik sistemler" şeklinde tanımlanmıştır.([106]) Günümüz koşullarında değerlendirildiğinde TCK 243'ün gerekçesinde bahsedilen bilişim sistemi tanımı, aslında bilişim sisteminden çok bilgisayar tanımına yakındır. Bilgisayar, bir bilişim sistemi parçasıdır. Verileri depolayan, işleme ve iletilmesinde etkin olan bilişim sisteminin bir alt ögesidir. Ancak bazen öğretide ve uygulamada aynı anlamda kullanıldıkları görülmektedir. Teknik olarak bilişim ve bilgisayar aynı şey demek değildir. Bilişim, bilgisayara oranla daha geniş kapsamlıdır, bilgisayarı da kapsar. Bilişim sistemi dendiğinde içinde barındırdığı tek unsurda bilgisayar değildir. Tabletler, POS cihazları, akıllı telefonlar, ağa bağlanıp veri aktarabilen içinde chip barındıran tüm teknolojik cihazlar aslında bilişim sisteminin unsurlarıdır. Bu kapsamda bilişim suçu olarak isimlendirilen eylemler, sadece bilgisayarlar üzerinde değil, bilgileri otomatik olarak işleme tabi tutabilen ya da veri iletişimi sağlayabilen diğer elektronik, manyetik veya mekanik araçlarla bunları veri iletişimi için birbirine bağlayan soyut ya da somut ağlar üzerinde işlenebilmektedir. ([107])

Bilişim teknolojilerinin gelişmesi, mobil ve akıllı cihazların sayısının kullanım oranlarının her geçen gün yükselmesi, İnternete yerden ve zamandan bağımsız mobil cihazlarla erişim olanaklarının artması, teknolojinin çeşitlilik gösteren ve

---

[105] Ali İhsan Erdağ, Yrd. Doç. Dr., Kırıkkale Ünv. Hukuk Fak. Ceza ve Ceza Muhakemesi Hukuku Ana Bilim Dalı Öğretim Üyesi. "Bilişim Alanında Suçlar (Türk Ve Alman Ceza Hukukunda) Gazi Üniversitesi Hukuk Fakültesi Dergisi C. XIV, Y. 2010, Sa. 2 sf:277

[106] TCK 243'ün gerekçesi:" Bilişim sisteminden maksat, verileri toplayıp yerleştirdikten sonra bunları otomatik işlemlere tâbi tutma olanağını veren manyetik sistemlerdir."

[107] Artuk/Gökcen/Yenidünya, s. 695. İnternet suçları, siber suç kavramları "Esasen "bilişim suçları" kavramı bir üst kavram olarak anılan bu suç türlerinin tümünü kapsamaktadır.

sürekli kendini yenileyen alt yapı imkânları, akıllı cihazların ve bilişim sistemlerinin birbiriyle entegre olma olanaklarının gelişmesi pek çok hukuka aykırı eylemin bilişim sistemlerinden yararlanılarak yapılmasını kolaylaştırmıştır.

Bilişim yoluyla sahtekârlığı, bir diğer kişinin bir eylemi yapması veya yapmamasına yol açarak, zarar görmesine neden olacak biçimde gerçeğin herhangi bir dürüst olmayan edimle olduğundan farklı olarak yanlış sunulmasıdır. Bir diğer tanımı, mal, para, hizmet, siyasi ve ticari avantajlar elde etmek amacıyla, yapılan hile ve aldatmalarda bilişimin kullanıldığı suç unsuru oluşturan şey olarak tanımlanabilir([108]). Literatürde geçen bir başka tanımda ise; "bilgisayar ve iletişim teknolojileri kullanılarak verilerin alınması, girilmesi, değiştirilmesi ve silinmesi yoluyla kendisine veya başkasına yasadışı ekonomik menfaat temin etmek için mağdura zarar vermektir." der. Yine literatüre giren bir başka tanım; "Kendisine ve başkasına yasa dışı ekonomik menfaat temin etmek ve mağdura zarar vermek maksadıyla; bilgisayar sistemlerini kullanarak sahte materyal (banknot, kredi kartı, senet vs.) oluşturmak veya dijital ortamda tutulan belgeler (formlar, raporlar vs.) üzerinde değişiklik yapmaktır."([109])

Suçlunun hedefi kendisine veya bir başkasına mali kazanç sağlamak veya mağdura ciddi kayıplar vermektir. Bu tür sahtekârlık suçları, modern bilgisayar teknolojileri ve ağ sistemlerinin avantajlarını değerlendirmeleri yönüyle klasik dolandırıcılık ve sahtekârlık suçlarından farklılık gösterir. Bilgisayar yoluyla en çok kredi kartları ile yapılan sahtekârlıklar, girdi-çıktı-program hileleri ve iletişim servislerini haksız ve yetkisiz olarak kullanma şeklinde yapılmaktadır.

Bilişim Suçları, 5237 sayılı TCK'nın "Topluma karşı suçlar" kısmının onuncu bölümünde "Bilişim Alanında Suçlar" başlığı altındaki düzenlemelerle karşımıza

---

[108] http://tr.wikipedia.org/wiki/Bili%C5%9Fim_su%C3%A7lar%C4%B1#Doland.C4.B1r.C4.B1c.C4.B1l.C4.B1k (erişim tarihi 01.04.2014)
http://www.aghukukburosu.com/makale-167-dolandiricilik-suclari.html (erişim tarihi 01.04.2014

[109] Türkiye Bilişim Şurası dokümanları (erişim tarihi 01.04.2014)
Metin Özderin, Ceza Hukuku Mevzuatı Özderin Av.bürosu http://cezahukuku.blogspot.com.tr/2006/10/biliim-suclari.html (erişim tarihi 01.04.2014)

çıkmaktadır. Her zaman doğrudan bir bilişim sistemi üzerinden bu tür suçlar işlenmemektedir. Suçun işlenmesi için eylem içerisinde "insan" faktörünün mağduriyetine sebep olunması suçun niteliğini değiştirmektedir. Adi bir hırsızlık suçunun bilişim sistemleri vasıtasıyla işlenmesi bunun ille de bilişim suçları kapsamına gireceği anlamına gelmemelidir Nitekim TCK m.142'de bahsedilen "Nitelikli hırsızlık" suçu da benzer kapsamda değerlendirilebilinir. ([110])

Dolandırıcılık suçunun nitelikli halinin tanımlandığı, TCK 5237, m.158/1-f bendinde, birden fazla nitelikli durumdan söz edilmektedir. Bunlardan birincisi, bilişim sistemlerinin, ikincisi ise bir güven kurumu olan banka veya kredi kurumlarının araç olarak kullanılması suretiyle yapılan dolandırıcılık fiilidir. Bilişim sisteminin aldatılmasından söz edilen husus, bilişim sistemine yetkisiz giriş kapsamında değerlendirilmemelidir. Burada söz konusu olan bilişim sistemin kullanılarak bir insanın aldatılması halinde bu bendin uygulanmasının mümkün olacağıdır. Sisteme girilerek, sistemden olanaklarından yararlanılarak çıkar sağlanmışsa; artık burada oluşan konunun niteliğine göre bilişim suçu veya TCK 5237, m.142/2-f kapsamındaki nitelikli hırsızlık suçudur.
Bilişim sistemine yetkisiz erişim sağlanıp, girilen sistem vasıtasıyla karşıdaki kişinin kandırılması; örneğin muhatabın kandırılıp kontör, para, eşya sağlaması veya e-posta ile gerçeğe aykırı bir durum iletilerek muhatabın bu e-posta ile istenileni yapması, bu durum sonucu, bunu yapanın çıkar elde etmesi halinde suç, TCK 5237, m.158/1-f bendindeki kapsama göre değerlendirilecektir.

Bilişim yoluyla dolandırıcılık suçunda, hem kendisine karşı hileli harekette bulunulan kişi hem de bu hileli hareket sonucu mal varlığında azalma olan kişi, üçüncü kişidir. Bu nedenle TCK m.158/1-f'de düzenlenen suçun mağduru üçüncü kişidir.([111],[91])

---

[110] TCK M.142 2.fıkra, e.bendi "Hırsızlık suçunun; Bilişim sistemlerinin kullanılması suretiyle İşlenmesi hâlinde, üç yıldan yedi yıla kadar hapis cezasına hükmolunur."
[111] http://www.aghukukburosu.com/makale-167-dolandiricilik-suclari.html
(erişim tarihi 01.04.2014

158. Maddenin 1 fıkrasının "f" bendine getirilen eleştirilerde, "bilişim sistemleri üzerinde gerçekleştirilen hileli işlemler sonucu hukuka aykırı yarar sağlanması eylemleri"nin düzenlenmediği, dolandırıcılıkta bilişim sisteminin basit bir araç olarak öngörülmesinin düzenlendiği görüşlerine yer verilmektedir.(96)

1. Sahtekârlık ve Bilişim Sistemine Girme

Sisteme girme ise TCK 5237'nin ikinci kitabın, "Topluma Karşı Suçlar" başlıklı üçüncü kısmının "Bilişim Alanında Suçlar "başlıklı onuncu bölümünün ilk maddesinde "bilişim sistemine girme" başlığıyla bu suç tipi düzenlenmiştir. Bu maddeyle yasa koyucu "bir bilişim sisteminin bütününe veya bir kısmına, hukuka aykırı olarak girme ve orada kalmaya devam etme" eylemini suç tipi haline getirmiştir.([112])

TCK'nın 243, 244 ve 245. maddeleriyle bilişim alanında işlenen suçlar olarak;([113])

- Bilişim sistemlerine hukuk dışı girme ve orada kalma; (m. 243/1),
- Sistemin içeriğine veya sistemdeki verilere sisteme girilmesinden dolayı zarar verme (m. 243/3),
- Bilişim sistemine veya burada bulunan verilere zarar verme (m. 244/1, 2),
- Bilişim sistemi marifetiyle haksız yarar sağlama (m. 244/4),
- Haksız olarak elde edilen başkasına ait banka veya kredi kartının kötüye kullanılması suretiyle yarar sağlama (m. 245/1)
- Başkalarına ait banka hesaplarıyla ilişkilendirilerek sahte banka veya kredi kartı üretme, satma, devretme, satın alma ya da kabul etme (m. 245/2)
- Sahte olarak üretilen veya üzerinde sahtecilik yapılan banka veya kredi kartıyla haksız menfaat elde etme (m. 245/3)

suçları düzenlenmiştir.

---

[112] Murat Volkan Dülger " Türk Ceza Kanunu'nda Yer Alan Bilişim Suçları ve Eleştirisi" (erişim tarihi 13.04.2014 http://www.dulger.av.tr/pdf/ytkbilisimsucelestrisi.pdf) sf:4

[113] Ali İhsan Erdağ, Yrd. Doç. Dr., Kırıkkale Ünv. Hukuk Fak. Ceza ve Ceza Muhakemesi Hukuku Ana Bilim Dalı Öğretim Üyesi. "Bilişim Alanında Suçlar (Türk Ve Alman Ceza Hukukunda) Gazi Üniversitesi Hukuk Fakültesi Dergisi C. XIV, Y. 2010, Sa. 2 sf:281

TCK'nın 243. maddesine göre; bilişim sistemine girme, "bir bilişim sisteminin bütününe veya bir kısmına hukuka aykırı olarak girmek ve orada kalmaya devam etmek" şeklinde tanımlanır. Bu suçun hukuki konusu dolandırıcılık suçunda olduğu gibi malvarlığına ait değerlerdir. Bu suç, mağdur ve unsurları bakımından dolandırıcılıktan ayrılır. Dolandırıcılık, failin hileli davranışlarla mağduru aldatmak yoluyla kendisine yarar sağlamasıdır. Bilişim sistemine girme suçunda ise; fail, dolandırıcılıktan farklı olarak mağdurun aldatılmış iradesinden faydalanarak sisteme girmez, fail yetkisi olmadığı halde hukuka aykırı olarak sisteme girmektedir. Suçun mağduru sistemi kullanmaya yetkili olan kişi veya sistemin sahibidir. Sisteme hukuka aykırı olarak girmek için sisteme ait yetkili erişim kodlarını ele geçirmesi gerekmektedir. Bu suçun oluşması açısından sisteme girenin yarar sağlaması şartı aranmaz. Yani, bu durumda oluşan bir suç sadece TCK 5237/157 m. (1) veya TCK 5237, m.158/1-f kapsamında değerlendirilmemelidir.

2. Sahtekârlık ve Bilişim Sistemini Engelleme, Bozma, Verileri Yok Etme veya Değiştirme Suretiyle Haksız Çıkar Sağlama.

TCK'nın 244. maddesine göre; sistemi engelleme, bozma, verileri yok etme veya değiştirme, "bir bilişim sisteminin işleyişini engelleme veya bozma; bir bilişim sistemindeki verileri bozma, yok etme, değiştirme veya erişilmez kılma, sisteme veri yerleştirme, var olan verileri başka bir yere gönderme" şeklinde tanımlanır.
Maddenin 4. fıkrasında; bilişim sisteminin işleyişini engelleme veya bozma; sistemdeki verileri bozma, yok etme, değiştirme veya erişilmez kılma, sisteme veri yerleştirme, var olan verileri başka bir yere gönderme suretiyle haksız çıkar sağlanması suçun ağırlaştırıcı nedeni olarak düzenlenmiştir. Bu suçun hukuki konusu dolandırıcılık suçunda olduğu gibi malvarlığına ait değerlerdir. Bu suç, mağdur ve unsurları bakımından dolandırıcılıktan ayrılır. Burada mağdurun

aldatılmış iradesi neticesi elde edilen bir yarar söz konusu değildir. Sisteme veya veriye müdahale şeklindeki teknik hile, kişiye karşı yapılmamıştır.([114],[115] ,[116])

3. Sahtekârlık ve Banka veya Kredi Kartlarının Kötüye Kullanılması

TCK'nun 245. maddesine göre; banka veya kredi kartlarının kötüye kullanılması, "başkasına ait bir banka veya kredi kartını, her ne suretle olursa olsun ele geçiren veya elinde bulunduran failin, kart sahibinin veya kartın kendisine verilmesi gereken kişinin rızası olmaksızın bunu kullanarak veya kullandırtarak kendisine veya başkasına yarar sağlaması; başkalarına ait banka hesaplarıyla ilişkilendirilerek sahte banka veya kredi kartı üretmesi, satması, devretmesi, satın alması veya kabul etmesi; sahte oluşturulan veya üzerinde sahtecilik yapılan bir banka veya kredi kartını kullanmak suretiyle kendisine veya başkasına yarar sağlaması" şeklinde tanımlanır. Banka veya kredi kartlarının kötüye kullanılması suretiyle yarar sağlama ile dolandırıcılık suçu birbirinden ayrılmalıdır.

Özellikle 5237 sayılı TCK.nun 245.maddesinin 3. fıkrası ile 158.maddenin 1. fıkrasının f bendinde düzenlenen "bilişim sistemlerinin, banka veya kredi kurumlarının araç olarak kullanılması suretiyle" işlenen dolandırıcılık arasında benzerlik bulunmaktadır. ([117])

Maddenin gerekçesinde; "Aslında hırsızlık, dolandırıcılık, güveni kötüye kullanma ve sahtecilik suçlarının ratio legislerinin tümünü de içeren bu fiillerin, duraksamaları ve içtihat farklılıklarını önlemek amacıyla, bağımsız suç haline

[114] Muammer Ketizmen, Türk Ceza Hukukunda Bilişim Suçları, Ankara 2008, s. 148
[115] Esra Yırtımcı, "Dolandırıcılık Suçu" Ankara Üniversitesi Sosyal Bilimler Enstitüsü Kamu Hukuku (Ceza Ve Ceza Usulü Hukuku) Yük. Lisans Tezi-2010
[116] Ahmet Gökcen- Murat Balcı, "Dolandırıcılık Suçu", MÜHF-HAD, C.14, S.1-55.
[117] Ahmet Gökcen- Murat Balcı, "Dolandırıcılık Suçu", MÜHF-HAD, C.14, S.1-55.

getirilmeleri uygun görülmüştür" denilmektedir. Dolayısıyla bu suç ile asıl olarak dolandırıcılık suçunda olduğu gibi malvarlığına dair değerler korunur. ([118],[119])

C. Sahtekârlığın Bilişim Yoluyla İşlenmesinde Beyaz Yakalıların Rolü

Beyaz yaka suçları düşüncesi, ilk defa ABD'de Edvin H, Sutherland tarafından, Amerikan Sosyoloji Derneğinin 1939 yılında yapılan toplantısında ortaya atılmıştır. Sutherland, uzun yıllar süren kriminolojik çalışmalarının neticesinde, toplumun daha düşük statüdeki kesimine ilişkin suçlar ile sokak suçlarına yönelme olduğunu, toplumun üst kesimindeki kişilerin işledikleri suçlara yönelik yeterli araştırma yapılmadığını gerekli ilgi ve dikkatin gösterilmediğini belirtmiştir. Üstelik Sutherland, toplumda ayrıcalıklı konumdaki bu kişilerce işlenen suçların, çok daha ciddi sonuçlar doğurduğunu ve zarar ölçüsünün daha yüksek olduğunu vurgulamıştır.([120])

Beyaz yaka suçlarının herkes tarafından kabul gören tatmin edici bir tanımı yoktur.([121]) Bu suçlar Sutherland tarafından, toplumda mesleki açıdan üst seviyede, saygın görünümdeki kişiler tarafından işlenen suçlar olarak tanımlanmaktadır.([122]) Bu kişilerin işledikleri suçun nitelikleri açısından ise; beyaz yaka suçları, yasal olmayan ve doğrudan fizikî bir uygulamaya dayanmayan sahtekârlık ve aldatma fiilleri olarak tanımlanmaktadır.

---

[118] Serkan SAZAK, Ceza Hukukunda Banka ve Kredi Kartlarının Kötüye Kullanılması, Yayınlanmamış Yüksek Lisans Tezi, İstanbul 2008, s. 44;

[119] Esra Yırtımcı, "Dolandırıcılık Suçu" Ankara Üniversitesi Sosyal Bilimler Enstitüsü Kamu Hukuku (Ceza Ve Ceza Usulü Hukuku) Yük. Lisans Tezi-2010 sf 119

[120] Cynthia Barnett, **"**The Measurement of White-Collar Crime Using Uniform Crime Reporting UCR Data " http://www.fbi.gov/stats-services/about-s/cjis/ucr/nibrs/nibrs_wcc.pdf (erişim tarihi 11.04.2014)

[121] Daniel Moriarty, 'What is white collar crime" (erişim tarihi 11.04.2014)

[122] Ufuk Gürçam"İnteraktif Dolandırıcılık Yük. Lisans Tezi" Anadolu Üniversitesi Sosyal Bilimler Enstitüsü 2008 sf:27

Teknolojideki gelişmelere bağlı olarak bilişim suçlarının kapsamı, beyaz yaka suçlarının kapsamında da değişikliğe neden olmuştur.([123]) Özellikle, büyük nitelikteki bilişim suçlarının işlenebilmesinde, hem teknik yapıya, hem de teknolojik bilgiye ihtiyaç olması ve belirli bir malî külfeti gerektirmesi, bu suçların belli bir düzeyin üstünde malî güce sahip veya yaptırım gücüne yani erke sahip kimselerce işlenebilmesini olanaklı kılmaktadır. ABD'de işlenen bilişim suçlan, beyaz yaka suçlarının en geniş kısmını oluşturmaktadır. ABD'de 1997 ilâ 1999 yıllarında FBI tarafından yapılan bir araştırmaya göre, bilgisayar kullanılarak gerçekleştirilen suçların % 42'sinin beyaz yaka suçları kapsamında olduğu belirlenmiştir. Bu kapsamdaki beyaz yaka suçlarının en yaygın olanı ise, değişik türdeki (yetkisiz veri elde edilmesi bu sınıfa dahil edilmiştir) hırsızlık fiilleridir.([124]) ABD'de, beyaz yakalıların, bilişim teknolojisini çok çabuk benimsedikleri ve bilgisayarı suçta en fazla kullananların da, beyaz yaka suçluları olduğu kabul edilmektedir.([125]),([126])

Globalleşen Dünya'da ekonomik değerlerin ortaklaşması, sınırların ekonomik olarak ortadan kalkması gibi nakit akışının da sınır tanımayacak bir yapıya gelmesi durumu ile şirketleri karşı karşıya bırakmıştır. Muhasebe ve finansal raporlama, bu gelişen pazarın önemli bir unsurudur. Raporlanan finansal bilgiler internet üzerinde hızla yayılmakta ve yatırımcıların ikamet ettiği ülke ne olursa olsun yerden ve zamandan bağımsız olarak istediği şirketin finansal bilgilerine istediği an erişmesine olanak vermektedir. Sermaye piyasalarının globalleşmesi, internet ve

---

[123] Erhan Erdönmez a.g.e. sf:1

[124] August Bequai, ABD'de Morris hakkında açılan dava, 1984 tarihli The Computer Fraud and Abuse Act (CFAA)'da, 1986 yılında yapılan değişiklik sonrasındaki açılan ilk davadır. Morris, temyiz başvurusunda, olayda suç işlemek kastının olmadığını, 1986 tarihinde yapılan değişiklikle The Computer Fraud and Abuse Act adlı Kanuna göre, birinci olarak yetkisiz şekilde federal bir bilişim sistemine erişmek kastının, ikinci olarak da bilişim sisteminin kullanılmasını engelleme niyetinin olduğunun ispat edilmesi gerektiğini savunmuş ise de, yalnızca yasaklanmış bilişim sistemine erişim niyetinin olması yeterli görülerek başvurusu reddedilmiştir. Frank, P.Andreano, , **The Evolution of Federal Computer Crime Policy: The Ad Hoc Approach to an Ever-Changing Problem,** (AM. J. CRIM. L, Vol.27:81), s.89-90.

[125] Ufuk Gürçam"İnteraktif Dolandırıcılık Yük. Lisans Tezi" Anadolu Üniversitesi Sosyal Bilimler Enstitüsü 2008 sf:27

[126] Hilary Conner , Conducting Forensic Research: A Tutorial For Mystery Writers

telekomünikasyon alanındaki gelişmeler, karşılaştırılabilir ve şeffaf finansal raporlama ihtiyacını belirgin hale getirmektedir. Bu da şirketler, yatırımcılar, kredi veren kuruluşlar ve denetçilerin, şirketlere ait hangi tür finansal bilgileri yayınlaması ve bunu en iyi şekilde nasıl iletebileceklerine ilişkin yeni bir anlayışa sahip olmalarını gerektirmektedir.([127])

Sahtekârlığın işlenmesinde "Beyaz Yaka" lıların rolü dendiğinde ilk akla gelen finans ve sermaya piyasalarını 2001 yılında sarsan Enron skandalıdır.

"Enron" finansal skandallar tarihinde yerini almış başlı başına incelenmesi gereken bir çalışma konusudur. Bu olay, sadece bir denetim hatası olarak ele alınmamalı, şirket yönetiminin yaptığı sahtekârlıklar ile hissedarların, tüm paydaşların, çalışanların dolandırıldığı bir vaka olarak görülmelidir.

1980 yılında Houston'da kurulan Enron şirketi, esas faaliyet alanından uzaklaşarak, enerji ticareti başta olmak üzere, pek çok değişik alanda faaliyet göstermeye başlamıştı. Bu amaçla Özel Amaçlı Girişimler (Special Purpose Entities) adı altında ikincil şirketler kurdu. Wall Street'in en gözde şirketlerinden olan bu firmanın geleceği, Amerika'da enerji endüstrisinde serbestleştirmeye gidileceği haberlerinin piyasada duyulmaya başlaması ile parladı. Bu arada Enron fiber-optik alanında yatırım yapacağını duyuruyor ve yüksek hızlı veri transferi ile evlere video sinema hizmetleri yatırımına gireceğini söylüyordu. Asıl işi olan enerji piyasasından çıkıp hiç deneyimi olmadığı bir konuda Telekom işine girerek ana faaliyet alanının dışına çıkıyordu. Bu arada Enron içinde bulunduğu gerçek finansal durumu, borçlarını, zararlarını yatırımcısından gizlemek ve şirketin durumunu olduğundan daha iyi göstermek için hileli finansal tablolar, raporlar yayınlıyor, cirodaki olağanüstü artışa rağmen çalışma sermayesini 1.8 milyar USD azalttığını, gizliyor, nakit akım tablo ve bilançolarında çalışan, paydaş ve hissedarlarını yanıltacak sahtekarlıklar yapıyordu. Enron tarafından özel amaçlarla kurulmuş olan ve yine Enron tarafından yönetilen küçük işletmelerle olan ilişkilerin, genel kabul görmüş muhasebe ilkelerine göre finansal tablolara dâhil edilmesi gereken sonuçları tablolara

[127] Christopher W. Nobes, A Survey of National Accounting Rules Benchmarked Against International Accounting Standards,

yansıtmıyor, yaptığı yatırımların değer kaybetmesinden kaynaklanan büyük zararlarını küçük ortakları üzerinde gösteriyordu. Böylece bu küçük işletmelerin finansal verilerini konsolide zararları tablolar üzerinde göstermeyerek, yatırımcılarından gizleme yolunu seçiyordu.([128]) Enron yönetim kurulu, şirketin yüksek risk içeren işlemlerine, kayıt dışı faaliyetlerine, yönetim kademesinin aşırı tazminat bedellerine izin vererek ABD'nin halka açık 7. büyük şirketinin çöküşüne katkı sağlamıştır. Böylece sadece Enron'da çalışan 21.000 kişi değil, hisselerine yatırım yapan hissedarlar, kredi veren kuruluşlar, alt yükleniciler, denetim yapan kurumlar kısacası birçok tarafın mağdur edildiği bir skandal olarak ekonomi tarihinde yerini almıştır.

Aynı dönemlerde ABD'de şirketlerin finans ve muhasebe denetimlerini yapan en yüksek makam olarak nitelendirilen, Amerikan Sermaye Piyasası kurumu (SEC) mali durumunu farklı göstermek için 3.8 milyar USD muhasebe kayıtlarında hile yaptığını ortaya çıkardığı iletişim devi WorldCom hakkında dolandırıcılık suçlaması ile dava açmıştır. Bu durum piyasada benzer alanda faaliyet gösteren Lucent Technologies, Nortel Networks gibi iletişim şirketlerinin de değer kaybı yaşamasına neden olmuştur.

ABD.'de bu durumlar yaşanırken, AB ülkelerinde ise, 2003 yılı sonlarında 8 milyon USD açık ile Parmalat firmasının muhasebe kayıtlarında yapılan sahtekârlık, Amerikan Sermaye Piyasası Kurumu (SEC) tarafından tarihteki en kapsamlı kurumsal finans sahtekarlığı olarak tanımlamıştır. Bir aile şirketi olarak başlayan Parmalat'ı 1961 yılında babasının ölümü üzerine devralan Calisto Tanzi uzun ömürlü süt üretme teknolojisinin bulunması ile yatırım alanını her geçen yıl büyütmüştür. Şirketin iflasının açıklanmasından sonra tutuklanan Tanzi, sorgusunda, çeşitli yatırım fonlarını Parmalat'tan, Parmatour isimli şirket ve diğer şirketlerine aktardığını itiraf etmiştir.([129])

---

[128] Ayça Zeynep Süer İ. Ü.İşletme Fakültesi Arş. Grv." Profesyonel Muhasebe Mesleğinde Enron Skandalı Ve Sonrası Gelişmeler" sf:2 (erişim Tarihi :12.04.2014)

[129] Ahmet Bayraktar ,"Türkiye'de Muhasebe Hileleri Tarihi" Trakya Üniversitesi Sosyal Bilimler Enstitüsü Haziran, 2007" sf:36-45 (erişim tarihi : 12.04.2014)

Beyaz yakalıların yaptıkları sahtekârlıklar için Association of Certified Fraud Examiners (Sertifikalı Sahtekârlık Denetçileri Birliği'ne (ACFE) göre mesleki sahtekârlık (Occupational Fraud) bir kişinin işletmenin kaynak ve varlıklarını kasıtlı yanlış uygulama veya yanlış kullanım yolu ile kişisel menfaat sağlamak için mesleğini kullanmasıdır.([130]) ACFE'ye göre sahtekârlığın dört unsuru vardır:

- Gizli olarak gerçekleştirilir,
- Yapan kişi, işletmeye karşı olan görevlerini ihlal etmektedir,
- Sahtekârlık, direkt veya indirekt finansal fayda sağlamak amacı ile gerçekleştirilir,
- İşletmenin varlıklarını veya gelirini azaltıcı eylemler içerir.

ACFE, beyaz yakalıların yaptığı sahtekârlıkları; varlıkların kötüye kullanılması, yolsuzluk ve ahlâki olmayan davranışlar ve hileli finansal raporlama olmak üzere üç başlık altında sınıflayarak incelemektedir.

D. Bilgi Sistemleri Güvenlik Standartları ve Sahtekârlık

Veri, enformasyon, bilgi kavramları genel hukuk sisteminde ayrımına varılan şekli ile kullanılmasıyla beraber özellikle, veri koruması hukukunda; veri, enformasyon, bilgi kavramları birbirinin yerine kullanılmaktadır. Bunun nedeni hukuksal düzenleme ve tartışmalarda bu kavramlar arasında ayrım yapmaya çalışarak tanımlamalarda bulunmak konuyu özünden uzaklaştıracak, özellikle uygulamaya geçildiğinde bu kavramlar arasında sınır çizmek daha da zor hale gelecektir. Ayrım yapmakta ısrarcı olmak yasal düzenlemelerde de istenmeyen sonuçların çıkmasına neden olabilir. Bu konuda AB üye ülkelerinde de benzer yaklaşım benimsenmiştir. Avrupa Konseyi Siber Suç Sözleşmesinde veri, "Bir bilgisayar sisteminin belli bir

[130] Association of Certified Fraud Examiners-ACFE, 2006 ACFE Report to the Nation on Occupational Fraud & Abuse," "Managing the Business Risk of Fraud: A Practical Guide" (erişim tarihi : 12.04.2014)

işlevi yerine getirmesini sağlayan yazılımlar da dâhil olmak üzere bir bilgisayar sisteminde işlemeye uygun nitelikteki her türlü bilgi" olarak tanımlanmıştır.([131])

Bilişim dünyasında bilgi güvenliği, bir varlık türü olarak bilginin izinsiz veya yetkisiz bir biçimde erişim, kullanım, değiştirilme, ifşa edilme, ortadan kaldırılma, el değiştirme ve hasar verilmesini önlemek olarak tanımlanır.

Bilgi Güvenliğinin birçok boyutu olmasına karşın, temel amacı:

- Veri bütünlüğünün korunması,
- Yetkisiz erişimin engellenmesi,
- Mahremiyet ve gizliliğin korunması
- Sistemin devamlılığının sağlanmasıdır.

Bilgi Güvenliği Yönetim Sistemleri, yukarıda verilen amaçları doğrultusunda bazı prensipler çerçevesinde hareket eder. Bilgi güvenliğinin temel unsurları;

GİZLİLİK (Confidentiality): Gizlilik, ISO tarafından "Bilgiye sadece yetkilendirilmiş kişilerce ulaşılabilmesi" olarak nitelenir. Bugün Şifreleme altyapılarının olmasının sebebi temel olarak gizlilik, ve bütünlüktür. Bilgi güvenliğinin her kavramı her kurum için eşit önemde olmayabilir. Gizlilik, özellikle bir yasa ya da sözleşme dolayısıyla bir zorunluluk ise önemlidir. Örneğin avukat - müvekkil ilişkisi ya da doktor - hasta ilişkisi gibi mesleki bilgiler kanun ile koruma altına alınmıştır. Bazı durumlarda ise taraflar birtakım bilgileri sözleşme ile birbirlerine verirlerken gizlilik anlaşmaları yaparlar. Her iki durumda da gizlilik büyük önem taşımaktadır.

BÜTÜNLÜK (Integrity): veri bütünlüğü, dar güvenlik anlamında verinin yetkisiz kişilerce değiştirilmesine veya yok edilmesine karşı korunmasıdır. Verinin bozulması kasten yahut kaza ile olabilir ve bu bütünlüğün bozulmuş olduğu

[131] Avrupa Konseyi Siber Suçlar Sözleşmesi, 1.Kısım, md.1. http://www.coe.int/t/dghl/standardsetting/t-cy/ETS%20185%20 turkish.pdf (erişim 30.03.2013).

gerçeğini değiştirmez. Güvenlik önlemi alanların iki tür riske karşı da tedbir almaları gerekmektedir. Bilginin bütünlüğü aşağıdaki üç kıstası sağlamalıdır.

- Kesinlik
- Doğruluk
- Geçerlilik

ERİŞİLEBİLİRLİK (Availability): Matematiksel olarak ifade edildiğinde, herhangi bir sistemin yapılış amaçlarına göre işlev gördüğü zamanın, işlev gördüğü ve görmediği toplam zamana oranıdır. Daha yalın bir anlatımla, doğru yetkilendirilmiş bir kişinin ihtiyacı olduğu anda ihtiyacı olan hizmetin orada olma oranına erişilebilirlik denir. Verilen hizmetin ne kadar güvenilir olduğunun bir ölçütüdür. Kurumlar hizmetin ne kadar önemli olduğunun ölçümünü yapıp, sistemleri ve verileri bu ihtiyaca göre yedekli hale getirirler.

Bu prensipler doğrultusunda hareket eden BGYS kurum, kuruluş ve kullanıcıların varlıklarını korumak amacıyla kullanılan araçlar, politikalar, güvenlik kavramları, güvenlik teminatları, kılavuzlar, risk yönetimi yaklaşımları, faaliyetler, eğitimler, en iyi uygulamalar ve teknolojiler bütünüdür.

1. Bilgi Güvenliği Standartları

Bilgi güvenliğinin üç temel unsuru, birbirinden bağımsız değil, etkileşim içindedir. Verinin gizliliğinin sağlanması o verinin erişilebilirliğini engellememelidir. Erişilen veri bilgiye dönüşecek katma değerli bir çalışmanın parçasıdır. Erişilen verinin bütünlüğünün sağlanması da önemlidir, çünkü verinin sadece gizliliği sağlanıyor, erişim engelleniyor ve bütünlüğü sağlanamıyorsa yanlış ve eksik veri ve ondan oluşacak eksik bilgi söz konusudur. İstenildiği zaman, yetkisi olan kişilerin erişemediği veri yok demektir ondan üretilecek bilgide olmayacaktır, bu durum bilgi güvenliğinin yapısının sağlanamadığını gösterir.

a. Bilgi Güvenliği Yönetim Sistemi

Bilgi güvenliği kavramı içerisinde kişilerin bilgi güvenliği önem arz ederken, kişilerin hizmet aldığı banka, sağlık sektörü, sigorta, kamu vb. kurumların kurumsal bilgi varlıklarını kullanmakta ve kendi kişisel verilerini de bu kurumsal bilgi varlıklarının saklandığı sistemlere sunmaktadırlar. Bu nedenle kurumsal bilgi varlıklarının güvenliği sağlanmadıkça kişisel bilgilerin güvenliği de sağlanamaz.

Globalleşme ve bilişim teknolojilerindeki gelişmeler ile tehditlerin sürekli gelişerek yenilenmesi, kullanılan yazılım veya donanımlarda meydana gelen güvenlik açıklarının risk yaratacak durumlarının artması, insan faktörünün kontrolü gibi unsurların süreçler bazında takip edilebilmesi ve üst seviyede bilgi güvenliği politikalarının yönetilmesi amacıyla, İngiliz Standartlar Enstitüsü (British Standards Institute-BSI) tarafından 1995 yılında BGYS ilk kısmı olan BS7799–1'yi yayınlamıştır. Standardın ikinci kısmı ise, 1999 yılında BS7799–2 olarak yayınlanmıştır. BS7799–1 2000 yılında küçük düzeltme ve adaptasyonlardan geçerek ISO tarafından ISO/IEC– 17799 adıyla dünya genelinde kabul edilen bir standart halini almıştır. [132]2002 yılında ise BSI tarafından BS–7799 standardının ikinci kısmı olan BS–7799–2 standardı üzerinde eklemeler ve düzeltmeler yapılarak tekrar yayınlanmıştır. 2005 yılında ISO daha önce ele aldığı ve yayınladığı 17799'u güncelleyerek, ISO/IEC–17799:2005 adıyla yeniden yayınlanmıştır. Aynı yıl İngiliz standardı olan BS7799–2 üzerinde eklemeler ve güncellemeler yaparak ISO/IEC:27001 standardını yayınlamıştır. 2006 yılında ise Türk Standartları Enstitüsü tarafından Türkçe'ye çevrilmiş ve TS ISO/IEC 27001:2006 olarak yayımlanmıştır. TS ISO/IEC 27001:2006 standardı, tüm kuruluş türlerini (örneğin, ticari kuruluşlar, kamu kurumları, kâr amaçlı olmayan kuruluşlar) kapsar. Bu standart, bir BGYS'yi kuruluşun tüm ticari riskleri bağlamında kurmak, gerçekleştirmek, izlemek, gözden geçirmek, sürdürmek ve iyileştirmek için gereksinimleri kapsar. ISO/IEC 27001 standardına göre kurulmuş olan bir bilgi

[132] Yılmaz Vural ve Şeref Sağıroğlu; Kurumsal Bilgi Güvenliği Ve Standartları Üzerine Bir İnceleme, Gazi Üniversitesi, Mühendislik-Mimarlık Fakültesi Bilgisayar Müh. Bl.
Gazi Üniv. Müh. Mim. Fak. Der., Cilt 23, No 2, 507-522, 2008 sf:511

güvenliği yönetim sistemi, kurumların bilgi güvenliği yönetiminde, kapsamlı prosedürler aracılığıyla güvenlik kontrollerini sürekli ve düzenli olarak işletmeyi ve sistemin sürekli gözlenerek, iyileştirilmesine yönelik çalışmaların yapıldığının taahhüdünü sağlar. Güven ve güvenilirliğin hayati önem taşıdığı alanlarda hizmet veren kuruluşların, uluslararası geçerlilikte bilgi güvenliği yönetim sistemleri standardına uygunluk belgesine sahip olması, hem mevzuat hem de kuruluşun güvenli işleyişi açısından bir zorunluluk olarak değerlendirilmektedir. Özellikle kredilendirmelerde bir güven ve itimat unsuru olarak kurum veya kuruluşun bilgi güvenliğini dikkate aldığını, bilgi güvenliğinin sağlanması için gerekli olan adımları uyguladığını ve kontrol ettiğini ispatlamaktadır. Bu sayede kurumlar veya kuruluşlar birlikte iş yaptıkları veya hizmet verdikleri kurum veya bireylerin tüm bilgilerinin güvende tutulacağı konusunda verdikleri taahhütten yükümlülük taşımaktadırlar. Ayrıca günümüzde, kurumların uluslararası alanlarda yapacağı faaliyetlerde ilk karşılarına çıkan hususlar bilgi güvenliği standartları yönündedir.

b. COSO ve COBIT

Coso (Committee on Sponsoring Organization) 1985 yılında sahte mali raporlamaların nedenini araştırmak için kurulmuş bir özel sektör girişimidir. Kurucusu James C. Treadway'den ötürü Treadway komisyonu olarak da bilinir. Özel sektör girişimi olmakla birlikte kurumlarda iç kontrol düzenlemelerine yönelik önemli araştırmalar yapmış, öneriler getirmiş bir organizasyondur. COSO raporu olarak da bilinen "İç Kontrol Bütünleşik Çerçeve" raporu işin çerçevesini belirlemekte en önemli kaynaktır.

Avrupa Birliği iç kontrol modeli olarak COSO modelini esas almıştır.([133]) COSO modeline göre iç kontrol; risklerin tespit edilmesi ve işlemlerin düzenli, etik, ekonomik, etkin ve etkili bir şekilde gerçekleştirilmesi, hesap verebilirlik sorumluluğunun yerine getirilmesi, yürürlükteki kanun ve yönetmeliklere uyumun

[133] Fatih Akyol "COBIT (Bilgi ve ilgili Teknolojiler için Kontrol Hedefleri) Uygulayan Şirketlerdeki Bilgi Güvenliği Politikalarının Şirket, Personel ve Süreçlere Etkileri " Yük. Lisans Tezi Beykent Ünv. 2013 Sosyal Bilimler Enstitüsü İşletme Yönetimi Anabilim dalı sf:61-65 (erişim tarihi 20.04.2014)

sağlanması, kaynakların kayıp, kötü kullanım ve zararlara karşı korunması yönünde, hedeflere ulaşıldığına dair güvence sağlamak üzere tasarlanmış bir süreçtir. ([134]) Sürecin işlerliği, işletmenin yönetim kurulu, yöneticileri ve diğer personeli tarafından uygulanmalıdır. İç kontrol, kurumun hedeflerinin gerçekleştirilmesi, mali raporların güvenilirliğinin sağlanması, mevzuat ve düzenlemelere uyumun sağlanması konusunda makul güvence sağlamayı amaçlar. Coso modelinin 5 adet bileşeni bulunmaktadır. Bunlar; kontrol ortamı, risk değerlendirmesi, kontrol faaliyetleri, bilgi iletişim ve izlemedir.

COBIT([135]) (Control Objectives for information and related technology) Türkçe olarak, bilgi ve ilgili teknolojiler için kontrol hedefleri anlamını taşıyan COBIT, ISACA (Information Systems Audit and Control Association) ve ITGI (IT Governance Institute) tarafından 1996 yılında, bilgi teknolojileri yönetimi için geliştirilmiş uluslararası alanda kabul gören, genel bilgi teknolojileri kontrol hedeflerinin iş yöneticileri ve denetçiler tarafından kullanılmasını amaçlayan, iş hedeflerinin bilgi teknolojileri hedeflerine dönüşümünü, bu hedeflere ulaşmak için gerekli kaynakları ve gerçekleştirilen süreçleri bir araya getirirken, aynı zamanda bilgi teknolojileri alt yapılarını da etkin kullanmayı sağlayan standartlar kümesidir. COBIT‟in vizyonu; Kurumlar açısından, bilgi teknolojilerinin sürdürülebilirliğini, kurumsal organizasyonun hedeflerine ulaşmasını yönünde liderlik ve organizasyonel yapılanma gerektiren, bilişim teknolojileri açısından bir yönetişim (IT governance) modeli oluşturmak suretiyle bir denetim değil yönetişim aracı olmaktır.([136])

COBIT, pek çok ülkede çeşitli sektörler için yasal düzenleme olarak da kullanılmakta olup, ülkemizde en somut örneği bankacılık sektörü için BDDK

---

[134] M.Didem Doğmuş, Maliye Bakanlığı AB. ve Dış İlişkiler Dairesi Başkanlığı "AB'de İç Denetim Sistemi, Ankara 2010) (erişim tarihi 20.04.2014)

[135] Emine Hacısüleymanoğlu; "Bilgi Teknolojileri Yönetişim Yöntemleri ve COBIT ile Ulusal Bir Bankada Uygulaması Yük. Lisans Tezi Yıldız Teknik Ünv.Fen Bilimleri Ens. 2010 sf 17-21 (erişim tarihi 20.04.2014)

[136] Bozkurt B., COBIT Denetimleri Açısından Bilgi güvenliği, http://bbozkurt.wordpress.com/tag/cobit/ (erişim tarihi 20.04.2014)

tarafından regüle edilmiş olmasıdır.([137]) Şu an kullanılan COBIT 4.1'in jenerik bir model olarak herhangi bir kurumda yer alabilecek tüm teknoloji süreçlerini kapsayan yapısı içerisinde, gruplanmış 4 süreç alanı ve 34 tane temel Bilgi Teknolojisi süreci yer almaktadır.([138])

BDDK Cobit Bilgi Güvenliği Yönetim Sistemi kapsamında Ülkemizde faaliyet gösteren bankalarda bilgi sistemleri yönetiminde esas alınacak ilkelere ilişkin tebliği 2007'de yürürlüğe sokmuştur.([139])

c. Sarbanes – Oxley Yasası (SOX)

Büyük kurumlarda yaşanan finansal sahtekârlıklar, her ülkede ekonomik mali hatta toplumsal sıkıntılara yol açmaktadır. Tezin C. Bölüm Sf 69'da "Sahtekarlığın Bilişim Yoluyla İşlenmesinde Beyaz Yakalıların Rolü" başlığı altında incelenen ve kısa bir özeti verilen Enron, WorldCom, Parmalat gibi Dünya'da büyük finansal skandallara neden olan sahtekarlıkların ülke ve sermaye piyasalarında neden olduğu sıkıntıların önüne geçebilmek amacıyla çeşitli düzenlemelerin yapılması yoluna gidilmiştir. Bu düzenlemelerin kilometre taşı ise; ABD'de 2002 yılında yürürlüğe giren Sarbanes Oxley Yasasıdır. (SOX)

Yasanın temel hükümlerinin yorumlanmış açıklamaları kısaca şu şekildedir: ([140])

- Bir şirketin CEO ve CFO'su doğru olmayan mali tabloları onaylarsa 20 yıla kadar hapis ve 5 Milyon USD'ye kadar para cezası alacaktır.

---

[137] Artinyan E. (2009). COBIT Çerçevesi, Deloitte İç Yayın, 1, S.2 (elden erişim tarihi 25.04.2014)

[138] Fatih Akyol "COBIT (Bilgi ve ilgili Teknolojiler için Kontrol Hedefleri) Uygulayan Şirketlerdeki Bilgi Güvenliği Politikalarının Şirket, Personel ve Süreçlere Etkileri " Yük. Lisans Tezi Beykent Ünv. 2013 Sosyal Bilimler Enstitüsü İşletme Yönetimi Anabilim dalı sf:61-65 (erişim tarihi 20.04.2014)

[139] 14.09.2007 Tarih 26643 sayılı R.G. "Bankacılık Düzenleme ve Denetleme Kurumundan: Bankalarda Bilgi Sistemleri Yönetiminde Esas Alınacak İlkelere İlişkin Tebliğ" (erişim tarihi (20.04.2014)

[140] Aslan Savaş Demircan "Sarbanes Oxley Yasası (2002) ve Bağımsız Denetime Etkileri" Trakya Ünv. Sosyal Bilimler Ens. Yük.Lisans Tezi Mart 2007 sf : 52

- Denetimlerde istenen defter ve belgelerin yırtılması, resmi bir işlemin engellenmesi, zorlaştırılması veya etki edilmesi 20 yıla kadar hapis cezasını gerektirmektedir.
- Faks, e-mail yoluyla yapılan hilelerin cezası 5 yıldan 20 yıla çıkarılmıştır.
- Yanlış yolda olan şirket yöneticilerine yapılan ekstra ödemeler önlenmekte ve geri alınmaktadır.
- Üst yöneticilere avans ve kredi verilmesi yasaklanmıştır.
- Şirket yöneticilerinin hisse senedi satma nedenlerini derhal açıklama mecburiyeti vardır.
- Hisse senedi sahtekârlıkları ağır suç kapsamına alınmıştır.
- Sermaye Piyasası Kanununu ihlal edenlerin, aldatılmış yatırımcılara ve diğer mağdurlara tazminat önlemek için iflaslarını isteyemeyeceklerdir.

Bu yasa her ne kadar ABD'de yürürlüğe girse de birçok ülke bu yasayı kendi iç işleyişleri ile uyumlandırmıştır. Ülkemizde de SPK, BDDK, TTK, KVK ilgili düzenlemeler ile dünyada yaşanan bu tür değişiklikleri iç mevzuatımıza uyumlandıracak eylemleri yapmaktadırlar. SPK'da benzer bir çalışma ile Seri X No: 19 tebliği ile denetim sektöründe faaliyet gösteren denetim firmalarından söz konusu yasalara uymalarını istemiştir.([141])

E. Veri Madenciliğinin Bilgi Sistemleri Güvenlik Standartları Açısından Değerlendirilmesi

Bilgi Teknolojilerinde yaşanan ilerleme ve gelişen uzman sistemler yardımıyla gerek iş dünyasında gerekse sosyal hayatta her türden iş ilişkileri ve aktiviteler hakkında muazzam bir veri birikimi oluşmuş, veri ambarları ve ortak hafıza sistemlerinin (corporate memory systems) kurulmaya başlamasıyla, yerden zamandan bağımsız istenilen verilere ulaşmak mümkün hale gelmiştir. Bu kadar veri birikiminin yaşandığı günümüzde, veri madenciliği yolsuzluk ve sahtekârlık ile ilgili faaliyetler konusunda, etkili mücadele vermeyi sağlayacak, kayıpları önleyecek ve sahtekârlığa yönelik eylemlerin delilleriyle birlikte yargıya teslim

[141] 02.11.2002 24924 sayılı R.G. Sermaye Piyasasında Bağımsız Denetim Hakkında Tebliğde Değişiklik Yapılmasına Dair Tebliğ. Seri: X, No: 19 (erişim tarihi : 11.05.2014)

edilmesini kolaylaştırabilecek teknik, yöntem ve süreçler bütünüdür. Her tür sahtekârlığın tespitinde denetimi sağlayanın da karar vericilerin de işini kolaylaştırırken özellikle bilgi teknolojileri alanında uygulanan çerçeve standartların da destekleyicisi durumundadır.

Her tür sahtekârlık eyleminin ana süjesi olan insan, genellikle zaman içerisinde yaptıkları sahtekârlıkların ortaya çıkmadığını görünce daha geniş çaplı ve daha yüksek maliyetli eylemlere teşebbüste bulunmaktadırlar. Bu nedenle bu tür durumlarda sahtekârlığın erken teşhis edilmesi büyük önem taşımaktadır. Bu durum hem beyaz yakalıların kurum içinde yaptığı sahtekârlıklar için hem de banka, sigorta şirketleri, sağlık kurumları gibi bu tür davranışlara maruz kalma oranı yüksek iş kolları içinde geçerlidir. Sahtekârlıkların tespitinde araştırmayı yapan açısından kullanılan yöntemler, proaktif, reaktif veya tümdengelim tümevarım olarak çeşitli şekillerde literatürde sınıflandırılmaktadır. Bilgi sistemleri güvenlik standartları çerçevesinde ister ISO 27001 olsun ister COBIT olsun, içerdikleri kontrol alt başlıklarının çoğu veri madenciliğinin sahtekârlık tespitinde üstlendiği işlevin uygulama alanları ile kesişmektedir. Bu standartları uygulayan kurumlar, kendi iç iş yapma süreçleri ile uyumlandırdıkları bu çerçeve standartlar sayesinde, müşterilerine, paydaşlarına sundukları hizmetlerin kalitesini ölçebilir hale gelmekte, iyileştirme önlemlerini hızlı hayata geçirebilmekte, sağladıkları güven ile sürdürülebilirliklerini devam ettirmede önemli rekabet avantajı elde etmektedirler. Dış müşteri ve paydaşlar açısından kurum ile arasındaki ilişkide, gizlilik, bütünlük, erişilebilirlik, hizmet ve ürününü kullandığı kurumun yasal yükümlülüklerini yerine getirdiğini bilmesi, iki taraf içinde istenmeyen durumlarla karşılaşıldığında nasıl davranılacağı, ne tür hak ve yükümlülüklerin olacağının net resminin ortaya konması açısından önemlidir.

F. Veri Madenciliğinin Kullanıldığı Bir Sahtekârlık Örneğinin İncelemesi

Sahtekârlık girişimlerinde işin kitabına uydurularak yapılması tespiti en zor olanlarıdır. Sahtekârlık ve sahtekârlık ile ilgili aktiviteler hakkındaki deliller, veri yığınları arasında dağıtık ve parçalı şekilde gizlenmiş bulunduğundan, veri

madenciliği teknikleri sahtekârlıkları tespit etme ve önlemede etkili bir yöntem olarak görülmektedir.

Sahtecilik belirleme alanındaki en kapsamlı literatür özetlerinden birisini yapan Clifton Phua,([142]) toplam 51 adet sahtecilik olayını incelemiştir. Bu incelemesi yapılan 51 olaya ait makaleden 14 tanesi, sigorta sektöründe yapılan sahteciliğe yönelik olarak yazılmış olan makalelerdir. Geriye kalan diğer konular ise, iletişim firmalarında, kredi kartlarında ve firma üst yönetiminin finansal raporlama süreçlerinde yapılmakta olan sahteciliğe yönelik olan makalelerdir. İncelenen makaleler arasında, sigorta sistemine yönelik olarak yazılmış olan makalelerden ise 5 tanesi sağlık sigortasına yönelik yapılmış araştırmalardır.

Sağlık sigortası işlemlerinde, oldukça çok sayıda sahtecilik türü ile karşılaşılmaktadır. Bu sahtecilik türlerinin bir kısmı sigorta şirketleri tarafından bilinmekte ve oluşmadan engellenebilmekteyken, büyük bir kısmının ise önü alınamayan sahtekârlıklardır. Her tür ticari, üretim, perakende alanlarında yapılan sahtekârlıkların, proaktif olarak daha oluşmadan ortaya çıkarılması zordur, bu durum, sağlık alanında sahteciliğe geldiğinde ise, daha da zordur. Sağlıkta bir olayın sahtecilik olup olmadığına karar vermek için;([143])

- Yeterli medikal bilgiye,
- Ödeme talep edilen işlemle ilgili detaylı verilere,
- Sağlık hizmeti alan kişinin geçmiş işlem verilerine gereksinim duyulur.

Bu veriler elde edilse bile, tıptaki "hastalık yok hasta vardır" kuralının yapılan pek çok müdahale için bir açıklama yapabilme esnekliği sağlıyor olması, tedavi yöntemlerinin, yaş, cinsiyet, eşlik eden başka hastalıklar gibi kişiye özel

---

[142] Clifton Phua, Vıncent Lee, Kate Smıth & Ross Gayler ;School of Business Systems, Faculty of Information Technology, Monash University and Baycorp Advantage,Australia (http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf erişim 10.04.2014)

[143] İlker Köse, Mehmet Göktürk (Gebze Yüksek Teknoloji Enstitüsü) Kemal Kılıç (Sabancı Üniversitesi) "Sağlık Geri Ödeme Sistemlerinde Kural Dışı Davranış Tespitinde Aktör-Metâ Odaklı Yaklaşımı," VII. Ulusal Tıp Bilişimi Kongresi Bildirileri

durumlara göre farklılıklar göstermesi bir medikal işlemin doğruluğu konusunda farklı görüşlerin ortaya çıkmasına neden olabilmektedir. Bütün bunlar, bu değerlendirmeyi yapacak ekibin üçlü bir sacayağı gibi, farklı uzmanlıklara sahip kişilerden oluşmasını gerektirir. Medikal verileri değerlendirecek tıp uzmanları, biyolog, kimyacı v.b. kişiler, sigorta ödeme sistemini iyi tanıyan sigortacı ve denetçiler ve işin teknik tarafını yürütecek, modelleri kuracak, çıkan sonuçları yorumlayacak veri madenciliği uzmanları. Bu tür uzun soluklu ve süreç bazlı tüm çalışmalarda olduğu gibi yetkin insan unsuru işin temelinde yer almaktadır.

Sağlık sektörüne özgü karşılaşılan sahtekârlık türleri şöyle sıralanabilir;

- Sigortalı olmamasına rağmen sigortalı başka birinin hakkından yararlanılması,
- Gereksiz bir tedavinin uygulanması,
- Gereksiz bir ilacın yazılması,
- Yapılmayan bir tedavi veya verilmeyen bir ilacın yapılmış veya verilmiş gibi gösterilmesi,
- Bir tedavi veya ilacın fiyatının olduğundan çok gösterilmesi

Çalışmamız esnasında Sağlık sektöründe yaşanan sahtekârlıklarla ilgili Ülkemizde herhangi bir istatistiki bilgiye rastlanmamıştır.([144])

ABD'de bu konuda yapılan çalışmalar neticesi sadece Tennessee eyaleti doğu, bölgesinde her yıl ortalama 25 milyon USD tutarında sahtekârlık yapıldığına yönelik tespitler yapılmıştır.([145])

---

[144] TUİK'e 24.04.2014 tarihinde konu, e-posta ile sorulmuş, 28.04.2014 tarihinde gelen e-posta ile;" Sağlık sektöründe yapılan sahtekarlıklara ilişkin bilgi mevcut olmadığı için verilememektedir." bilgisi alınmıştır.

[145] http://www.policymed.com/2014/01/pharmaceutical-and-device-healthcare-fraud-enforcement-perspectives-for-2014.html,
http://www.insurancefraud.org/statistics.htm#.U3YUa02KAq0 (erişim: 23.04.2014)
Prof. Dr. F. Cankat Tulunay (Avrupa Klinik Farmakoloji Birliği Onursal Başkanı, Türkiye Akılcı İlaç kullanım platformu Başkanı) "Sahtekarlık İptilası
http://www.researchgate.net/publication/260249876_Sahtekarlk_iptilas_Ar_damar_olmayanlar_ve_Hitlerin_la_firmalarndaki_faist_torunlar_www.klinikfarmakoloji.com (erişim: 23.04.2014)

Örnek bir olay üzerinden veri madenciliği yöntemlerinin uygulanmasına yönelik inceleme olarak "Sağlık geri ödeme sistemlerinde kural dışı davranış tespitleri" konulu çalışma ele alınmaktadır. Söz konusu çalışma Sn. İlker Köse, Sn. Mehmet Göktürk, Sn. Kemal Kılıç tarafından yapılmış, 2010 yılındaki VII. Ulusal Tıp Bilişimi Kongresi Bildirileri arasında da yer almıştır. Önce Sn. İlker Köse ile sonrasında Anadolu Sigorta A.Ş. ile temasa geçilerek çalışma hakkında bilgi sahibi olunmuştur. ([146])

1. Örnek Olay Üzerinde Veri Madenciliği Süreci

a. İşi Tanımlamak, Hedefi Belirlemek

Veri Madenciliği, çalışmalarında ilk adım işi anlamak ve çalışmanın hangi amaçla yapıldığını net olarak ortaya koymaktır. Bu örnekte çalışmayı yapan ekip; Sağlık sektörü geri ödemelerinde; (bu hem sağlık sigortası yapan özel şirketler olabilir hem de devletin SGK konusunda yaptığı ödemeler olabilir,) proaktif olarak işin daha oluşmadan önlenmesine yönelik, sağlık alanının tamamında kullanılabilecek bir veri madenciliği çalışması yapmayı hedeflemişlerdir. Bu çerçevede daha önce dünya üzerinde yapılan çalışmalarda incelenmiş ve literatürde rastlanılan çalışmalardaki zorlukları tespit edilmiştir. Burada veri madenciliği sürecinin birinci adımı; işi anlamak, tanımlamak, hedefi belirleme çalışmaları tamamlanmıştır.([147])

b. Veriyi Anlamak ve Hazırlamak

Veri Madenciliği sürecinin ikinci ve üçüncü aşamaları veriyi anlama ve hazırlama çalışmalarıdır. Verinin hazırlanması, seçme, temizleme, kurma, bütünleştirme, indirgeme ve biçimlendirme şeklinde alt çalışmalara ayrılmaktadır. Örnek alınan çalışmada Anadolu Sigorta'ya ait sağlık verileri ele alınmış, veri kalitesinin

---

[146] Sn. Köse halen konu hakkında çalışmalarını doktora tezi kapsamında devam ettirmektedir. (27.04.2014)

[147] Tez C.Ö: Veri Madenciliği Süreci A.1.: "İşi anlamak, tanımlamak ve hedefi belirlemek "sf:14

değerlendirilmesi varılmak istenen hedef doğrultusunda veri üzerinde denenmiştir. Bu aşamada;

- Dünya örneklerinden alınan sahtecilik tipleri belirlenmiş,
- Sahtecilik tiplerinde yer alan unsur ve nesnelerin nitelikleri belirlenmiş,
- Veri ambarı tasarımı

Yapılmıştır. Modellemeye geçilmeden önce bahsedilen çalışmaların tamamlanmış ve kendi içinde gözden geçirmesi yapılmıştır. ([148])

c. Modelleme

Modelin yapısını kurgulamada elde edilen veriler üzerinde kural dışı davranışların tespit edilmesi için bir yaklaşım belirlenmiştir. Söz konusu çalışmada sahteciliğin tarafları olarak belirlenen unsurlar (çalışmada "aktörler" olarak adlandırılmaktadır.) aşağıda yer almaktadır.

- Hak sahibi (sigortalı),
- Doktor,
- Hastane,
- Eczane

Yine aynı kapsamda kullanım değeri olan nesneler ise (çalışmada "meta" olarak adlandırılmaktadır.)

- Sağlık hizmetleri
- İlaçlar (etken madde, üretici, vb göre)

Şeklindedir.

Burada kural dışı davranışları gösterecek sahteciliğin tarafı olacak unsurlar ile sahteciliğin üzerinden gerçekleştirileceği nesneler (metâlar) arasındaki çapraz ilişkilerde modele dâhil edilmiştir. Burada amaç, kural dışı davranışlarda yer alan unsurlar ve nesneler zaman içerisinde değişse de, kural dışı davranışları gösteren unsurlar (aktörler) ile nesneler (metâlar) arasındaki ilişkinin aynı kalacağı kabul edilmektedir. Bu varsayımdan amaçlanan ise, medikal hatalarla kural dışı

[148] Tez C.Ö Veri Madenciliği Süreci A.2.: "Veriyi anlamak ve hazırlamak "sf:15

davranışları birbirinden ayırma konusunda bir argüman oluşturmaktır. Bazı durumlar vardır ki, aslında temelinde sahtekârlık söz konusu değildir, tamamen tıbbi bir hata sonucu oluşmuş bir unsurdur. Gerçekten tıbbi hata olan işlemler genellikle daha fazla ödeme yapılmasına değil, hastanın zarar görmesine neden olmaktadır. Sahtecilik vakaları ise, tıbbi kurallara uygun olarak yapılsa da fazla veya gereksiz ödeme yapılmasını gerektirmektedir. Bu çalışmada, bu ilişki doğru argümanlarla çok yerinde kurulmuş ve modele dâhil edilmiştir. Çalışmada incelenmek istenen diğer bir konu da, her bir sahtecilik tipi için ayrı nitelik altkümelerinin oluşturulmak istenmesidir. Burada amaç, farklı türde sahtecilik tiplerini tespit etmeye uygun bir model geliştirilip geliştirilemeyeceğidir. Yani bir sahtekârlık türü incelenirken buna eşlik edebilecek başka sahtekârlıklarda açığa çıkarılmak istenmektedir. Nitekim farklı aktörlerin ve farklı metâların yer aldığı bir ekosistemde, çeşitli türlerde sahteciliğin olacağı beklenmektedir. Modelde ayrıca sahteciliğe esas unsurlar arasındaki bağlantıda göz önüne alınan işlem sayısı, işlemin sıklığı, maliyet hacmi gibi değerlendirme parametrelerinin modelin iyileştirilmesine pozitif katkı sağlayacak bağlantıların ortaya çıkmasını sağlamak amaçlanmıştır. Esas unsurların (aktörlerin) bir araya geldiklerinde ortaya çıkan riskin değerlendirilmesi amaçlanmıştır.([149]) Bu güncellemede esas alınan yaklaşım, riskli unsurlarla ilişkiye giren diğer unsurların risk değerlerinin, ilişkileri ile orantılı olarak artırılmasıdır.

d. Değerlendirme

Modelin kurulması sonrası gelen bu aşamada modelin istenen amaca hizmet edip etmediği değerlendirilir. Kısıtlar ve riskler gözden geçirilir. Bu çalışmada sahtecilik türleri için tüm unsurların ve taşıdıkları niteliklerin ağırlık dereceleri belirlenmiş ve bu atamaları yapılmıştır. Ağırlıkların hesaplanmasında;

İncelemeye alınan işlemler üzerinde yapılan etiketleme ile aktörlerin nitelikleri arasındaki korelasyonun incelenmesi ile ağırlıklar belirlenmiştir. Bu atamaların tutarlılığı Analitik Hiyerarşik Proses (AHP) ile doğrulanmıştır. Değerlendirme noktasında çalışmayı yapanların karşılaştıkları en önemli sorun, ödeme işlemlerine ait kayıtların sadece birine bakarak onun sahtecilik olup olmadığına karar

[149] Tez C.Ö: Veri Madenciliği Süreci A.4.: "Modelleme "sf:19

verilememesi nedeniyle etiketlemenin sağlıklı yapılamamasıdır. Çünkü bir işlemin sahtekârlık içerip içermediği tek bir parametre ile değil pek çok veri ve bu verilerden elde edilen yorumlanmış bilgiye göre verilen bir karardır. Dolayısıyla etiketin, ilgili işlemde yer almayan başka verilerle ilgisi ve korelasyonu mevcuttur. Bu koreleasyonun ortaya çıkarılarak, etiketlemenin buna göre yapılması zaman almıştır. Planlamanın baştan iyi yapılması burada harcanan zamanın da önceden öngörülmesini sağlamıştır. Çalışmanın amacında tüm klinik branşları kapsayacak ve farklı sahtecilik türlerini de ortaya çıkaracak proaktif bir modelin kurulması denetimli öğrenme yönteminin uygulanması hedeflendiğinden, modelin bu amaca hizmet etmekten uzak olduğu görülmektedir.([150]) Bununla birlikte sırasıyla, işlemleri, işlemde yer alan unsuları (aktörleri) ve bu aktörlerin geçmişte içinde yer aldıkları diğer işlemlerini raporlayabilecek uygun görselleştirme araçlarının olması halinde, denetimli öğrenme yöntemlerinin dikkate alınması söz konusu olabilecektir. Mevcut durumda ise, işlem risklerinin hesaplanması aktör niteliklerindeki anomalilerin istatistiksel yöntemlerle tespit edilmesi suretiyle yapılmaktadır. Bununla birlikte bu sonuçların, aynı veri seti kullanılarak öğreticisiz öğrenme teknikleri ile yapılacak kümeleme sonucunda elde edilecek kümelerle karşılaştırılması öngörülmüştür. Çalışma, 2010 yılından beri devam ettirilmektedir. Denetimli öğrenmenin bu modelde uygulanamayacağı görülüp, "interactive machine learning" yöntemi model üzerinde denenmektedir. Sonuçlar tatminkâr gözükmesine rağmen henüz çalışma tamamlanmadığı için raporlanmamıştır. Çalışmanın bir bölümü için, Anadolu Sigorta A.Ş. ile ilaç verileri üzerinde bu araç ve yöntemle analiz sonuçları paylaşılarak karar vericilere yardımcı olabilecek verilerden yorumlanmış bilgiler ve raporlar üretilmiştir.

Çalışmanın kapsamının genişletilmesi, bu bağlamda, hastanelerden işlem bazlı olarak toplanan veriler üzerinde de analiz yapılması hedeflenmektedir.

[150] Tez C.Ö: Veri Madenciliği Süreci A.5.: "Değerlendirme "sf:20

V. SONUÇ

1990'lı yıllardan başlayarak yaşanan hızlı teknolojik gelişmeler ve İnternetin yaygınlaşmasının bir sonucu olarak bilgi ve iletişim teknolojilerindeki yenilikler, ekonomik ve sosyal yaşamın her alanında karşımıza çıkmakta, kamu yönetiminden, eğlenceye, sosyal iletişimden, iş dünyasının iş yapma biçimlerine kadar toplumun tüm kesimlerini ve bireylerin yaşamlarını yakından etkilemektedir.

Bilişim ve iletişim altyapısının geniş coğrafi alanlara yayılması ve mobil iletişim teknolojileri aracılığı ile veri ve enformasyon erişiminin mekândan ve zamandan bağımsız hale gelmesi, bireylerin, kurumların ve toplumların birbirleri ile ilişkilerini arttırdığı gibi, dili ve kültürü ne olursa olsun tüm insanları Bilgi Teknolojilerinin kapsama alanına almaktadır. Üstelik bu alanın sınırları her geçen gün genişlemekte, insanın yaşadığı mekânlar (evler, ofisler vb.) ve ortamlarda (şehirler, yollar vb.) bu kapsama dâhil olmaktadır.

Günümüzde bu değişim ve dönüşüm süreci hız kesmeden artan bir ivme ile devam etmektedir. Teknolojideki bu değişim ve dönüşüm iş yapma biçimlerini etkilediği kadar, bireylerde ve organizasyonlarda biriken veri ve bu veriden yorumlanan bilgi oluşumunu hızlandırmaktadır. Bir kurumda, örtülü bilgi olarak adlandırılan, kuruma katma değer sağlayacak yenilikçi bilgilerden yeni fikirlerin oluşmasını sağlayacak, bilgiye dayalı organizasyon yönetimi son zamanlarda önem taşır hale gelmiştir.

Teknolojide yaşanan bu hızlı değişim, ekonomik, sosyal, siyasal ve kültürel gelişmelerin önünü açtığı gibi, olumsuzlukları da beraberinde getirdiği durumlar olabilmektedir. İnsanların özel hayatları, kurumların stratejik hedef ve politikaları kolaylıkla ihlal edilebilir hale gelmektedir. Bilişim sistemlerine ve bu sistemler üzerindeki verilere yönelik güvenlik ihlalleri inanılmaz bir hızla artmaktadır. Üstelik bu güvenlik ihlalleri zaman zaman ulusal boyutta sıkıntılara neden olabilecek duruma gelmiştir. Siber savaşlar, konvansiyonel savaşlardaki gibi bir etki verebilecek güçtedir. Bir ülkeyi ille de askeri yöntemlerle işgal etmek yerine Siber teknoloji, bunu tek kurşun kullanmadan yapılabilecek hale getirmiştir. Finans kurumları, ulaşım, haberleşme ve sigorta şirketleri bu güvenlik ihlallerinin baş

mağdurları konumundadırlar. Bunun yanı sıra siber terörizm, sabotaj, sahtekârlık, dolandırıcılık artık sanal ortamlarda çok rahat yapılabildiği için mağduriyetlerin oluşumu da arttırmaktadır. İnsan hayatını etkileyecek her sektör ve kurumda bilgi ve iletişim güvenliğinin sağlanması, sahtekârlık, dolandırıcılık faaliyetlerinin önüne geçilmesi yönünde yapılan çalışmalar, uygulanacak teknolojik yöntemler daha da önemli hale gelmektedir. Teknolojinin gelişmesi ile gelişen akıllı sistemler artık sahtekârlık girişiminde bulunmak isteyenler içinde caydırıcı olmaktadır. Akıllı sistemleri kandırma girişimlerinde bulunan sahtekârlar artık daha kolay tespit edilmekte ve adalete daha kısa sürede teslim edilmektedirler. Bu vakalardan elde edilen bilgiler daha akıllı sistemlerin üretilmesinde yine bir girdi oluşturmakta, Veri Madenciliğinde öğrenen sistemlerin tasarımında karşımıza çıkmaktadırlar.

Bu bağlamda, Veri Madenciliği bilişim dünyanın ürettiği verinin analiz edilerek bilgiye dönüştürülüp, karar için kullanılabilmesi sürecinde sağladığı katkı ile insan aklının, teknoloji ile birleştiğinde yarattığı sinerjiyi ortaya çıkaran olumlu bir unsurudur. Ülkemizde henüz istenilen ölçülerde kullanım imkânı bulamayan veri madenciliği konusunda, ülke olarak yetkinliklerimizi artırmanın, kurumlarımızı doğru kararlara yöneltmesi, devlet açısından doğru kullanıldığında önemli bir dönüşüm manivelası etkisi yapacağı ve kurumların, organizasyonların başarı şanslarına olumlu katkı sağlayacağı görülmelidir. Bu yönde devlet, özel sektör, üniversite işbirliklerinin arttırılması; ar-ge faaliyetlerinin geliştirilmesi, teknoparkların bu alanlarda çalışma yapacak kurum, kuruluş ve bireylere öncelik sağlaması yetişecek entelektüel sermaye birikimine yatırım yapılması ülke olarak benimsenecek teknoloji politikaları ile mümkün hale gelecektir.