

# ARAÇ SİGORTASI HİLELERİNDE VERİ MADENCİLİĞİNİN KULLANIMI\*

## USE OF DATA MINING IN VEHICLE INSURANCE FRAUD

Nuran CÖMERT\*\*   
Merve KAYMAZ\*\*\* 

### Özet

Sigorta sektörü, muhtemel zararlara karşı sigortalıyı garanti altına aldığından, sosyal ve ekonomik çöküntü ile kayıpları önlediğinden kişi ve kurumlar için her zaman bir güvence kaynağı olmaktadır. Sektör içerisinde yer alan tarafların haksız menfaat sağlamak amacıyla kasıtlı olarak hileli eylemlere başvurması ise sektörün yara almasına sebep olmaktadır. Bu sebeple hilelerin tespiti ve hileli eylemlere karşı uygun aksiyonların alınması çok önemlidir. Teknolojinin geldiği noktada veri madenciliği, hile riskinin yönetiminde kullanılan etkili araçlardan biri haline gelmiştir. Makalemizin temel amacı sigorta şirketlerinde hile riskinin yönetilmesinde yardımcı bir araç olarak veri madenciliğinin nasıl kullanılacağını açıklamak olup bu amaç doğrultusunda ele alınan konu araç sigortaları özelinde incelenerek spesifik bir örnekle açıklanmıştır.

Bu bağlamda bir acenteden kendi veri tabanında yer alan müşterilerine ait veriler temin edilmiştir. Yapılan hasar taleplerinde hile şüphesi olanlar ve olmayanlar karar ağacı modelinin J48 algoritması aracılığıyla tahmin edilmeye çalışılmıştır.

**Anahtar Kelimeler:** Veri Madenciliği, hile, sigorta, sigorta hilesi.

**Jel Sınıflandırması:** C80, G22

### Abstract

The insurance sector is always a source of assurance for individuals and institutions as it guarantees the insured against possible damages and prevents social and economic collapse and losses. The intentional applications of fraudulent acts by the parties involved in the sector in order to provide unfair benefit causes the sector to get injured. For this reason, it is very important to detect frauds and to take appropriate actions against fraudulent actions. At the point where technology has come, data mining has become one of the effective tools for managing fraud risk. The main purpose of the article

\* Bu çalışma Prof. Dr. Nuran Cömert' in danışmanlığında, Merve Kaymaz'ın “ Veri Madenciliği Yöntemi ile Risklerin Yönetilmesi ve Sigorta Sektörü Üzerine Bir Uygulama” adlı Yüksek Lisans Tezinden üretilmiştir.

\*\* Prof. Dr., Marmara Üniversitesi, İşletme Fakültesi Öğretim Üyesi, nurancomert@marmara.edu.tr

\*\*\* Yüksek Lisans Öğrencisi. Marmara Üniversitesi, Muhasebe Denetimi Yüksek Lisans Programı, mkaymaz@marun.edu.tr

is to explain how to use data mining as an aid to manage fraud risk in insurance companies, the subject is examined in the scope of vehicle insurance and explained with a specific example.

In this context, the data of the customers belonging to their customers in their database were obtained from an agency. It is tried to guess the decision tree model J.48 algorithm with and without suspicion of cheating.

**Keywords:** Data mining, fraud, insurance, insurance fraud.

**Jel Classification:** C80, G22

## 1. Giriş

Sigorta sektörü, gerek üretilen prim tutarının büyüklüğü ve bu primlerle yapılan yatırımlar, gerek kişi ve kuruluşların risklerini yönetmelerine yardımcı olması nedeniyle ekonomik ve sosyal açıdan kritik rolü olan önemli bir sektördür.<sup>4</sup>

Teknolojideki gelişmelerin büyük bir ivme kazandığı ekonomi ve iş dünyasındaki değişim ve dönüşüm, iş yapış şekillerini etkilerken büyüme ve gelişmeyi de beraberinde getirmiş, bilgiye ve tüketiciye erişim kolaylığı işlem hacmini de artırmıştır. Teknolojinin kullanılmasıyla bir çok aşaması elektronik ortama taşınan iş süreçlerinde, sürece dahil olan kişi sayısı azalırken, tasarımın doğasından kaynaklanan nedenlerle işler biraz daha karmaşık hale gelmiş ve beraberinde yeni riskleri getirmiştir. Kendisi bir risk yönetim aracı olan sigortacılık bu değişim ve dönüşümden önemli ölçüde etkilenmiş, giderek çeşitlenen ürünleriyle işlem hacimleri büyük ölçüde artmıştır.

Esasen risklerin yol açacağı kayıpları mümkün olduğunca yok etme, etkilerini azaltma, önleme ve geleceği güvence altına alma arzusuyla doğmuş bir ekonomik araç olan sigortacılık özü itibarıyla bir güvence müessesesidir. Latince “sicurta” kelimesinden türetilen ve “güvence” anlamına gelen sigorta terimi, ticari hayatımızı ve ilişkilerimizi düzenleyen Türk Ticaret Kanununun 1401’ inci maddesinde “*Sigortacının bir prim karşılığında, kişinin para ile ölçülebilir bir menfaatini zarara uğratan tehlikenin, rizikonun meydana gelmesi halinde bunu tazmin etmeyi ya da bir veya birkaç kişinin hayat süreleri sebebiyle ya da hayatlarında gerçekleşen bazı olaylar dolayısıyla bir para ödemeyi veya diğer edimlerde bulunmayı yükümlendiği sözleşme*” olarak tanımlanmıştır.

Sigorta kavramının hukuki boyutunu ortaya koyan bu tanımdan da anlaşılacağı üzere sigorta; kanun ve mukavele ile tespiti yapılan, çerçevesi belirlenmiş bir riskten benzer derecede tehdit alan belli bir sayının üstündeki grubun, tesadüf eseri ortaya çıkan zararın beraberce üstesinden gelmek amacıyla oluşturduğu organizasyondur.<sup>5</sup> Sigorta, riskleri sigortalılardan sigorta şirketlerine taşıdığı için tercih edilen bir seçenektir.<sup>6</sup> Sigorta, iyi kurgulanması gereken sistematik

4 <https://home.kpmg/tr/tr/home/industries/insurance.html> (erişim tarihi: 03.08.2019)

5 Pekiner, Kamuran (1981). Sigorta İşletmelerinin Prensipleri (Hesap Bünyesi Alt Başlık) , 3. Baskı, İstanbul, Formül Matbaası, s.17.

6 Ghorbani, A., Farzai, S. (2018). Fraud Detection in Automobile Insurance using a Data Mining Based Approach, International Journal of Mechatronics, Electrical and Computer Technology (IJMEC), Cilt: 8(27), 3764-3771, s. 3764.

bir devirdir. Önceden yapılan ödemeler karşılığında zarar riskine karşı koruma sağlamak için kullanılır.<sup>7</sup>

Yarattığı değer açısından ülke ekonomilerinde bir gelişmişlik göstergesi olarak algılanan sigorta sektörü, muhtemel zararları bölüşerek sigortalıyı koruma altına aldığından kişi ve kurumlar açısından en etkili güvence olarak görülmektedir. Zararın telafi edilmesi neticesinde sosyo – ekonomik çöküntü ve kayıplar önlediği gibi prim ödemeleri yoluyla elde edilen fonlar çeşitli yatırım araçlarıyla değerlendirilerek ekonomiye ve sermaye piyasalarına aktarılmakta böylece ekonomik kalkınmaya katkı sağlanmaktadır.<sup>8</sup> Ülkede kişi başına üretilen prim sayısı, sigortalı kişi sayısı, sigorta sektörünün büyüklüğü o ülkenin gelişmişlik düzeyini göstermekte ve bu durum ülkelerin birbiriyle olan ekonomik ilişkilerinin güçlenmesine de etki etmektedir. Özetle toplumda huzur ve güven tesis etmek, kredi teminine olanak sağlamak, girişimcileri desteklemek, finansal istikrarı ve iş sürekliliğini sağlamak, risklerin daha etkili yönetilmesi ve finansal sistemin etkililiğinin artırılmasına katkıda bulunmak ve istihdam yaratmak gibi pek çok işlevi ve yararı bulunan sigorta sektörünün<sup>9</sup> sağlam bir temele dayalı olarak yapılandırılması, değişimi ve dönüşümü yakalaması ve dolayısıyla kurumsal sürdürülebilir bir yönetim anlayışıyla kendi risklerini iyi yönetmesi bugün ülke yönetimlerinin de en önemli meselelerinden birisi olmuştur. Sigorta şirketlerinin maruz kaldığı risklerden en önemli bir tanesi hile riskidir. Teknolojinin geldiği noktada veri madenciliği, hile riskinin yönetiminde kullanılan etkili araçlardan biri haline gelmiştir. Makalemizin temel amacı sigorta şirketlerinde hile riskinin yönetilmesinde yardımcı bir araç olarak veri madenciliğinin nasıl kullanılacağını açıklamak olup bu amaç doğrultusunda ele alınan konu araç sigortalı özelinde incelenerek spesifik bir örnekle açıklanmıştır. Giriş ve sonuç dahil beş bölümden oluşan çalışmamızda öncelikle işletmelerin maruz kaldığı hileler ve dolayısıyla sigorta şirketlerinde görülen hileler ile veri madenciliği hakkında genel bir bilgi aktarılmış daha sonra bu temel bilgiler ışığında veri madenciliğinin sigorta hilelerini önlemede nasıl etkili bir araç olarak kullanıldığı, araç sigortalı özelinde spesifik bir örnek üzerinde açıklanmıştır.

## 2. Hile Kavramı, Genel Olarak İşletmelerin ve Sigorta Şirketlerinin Maruz Kaldığı Hileler

Hemen hemen her sektör, faaliyette bulunulan süre boyunca her an hile fiili ile karşı karşıya kalabilmektedir. Sertifikalı Hile Denetçileri Derneği'nin (ACFE) 2018 yılı raporunda, işletmelerin yıllık gelirlerinin yaklaşık yüzde beşinin hile sebebiyle eridiğini ve bunun da yaklaşık dört trilyon dolara tekabül ettiği bilgisi yer almaktadır.<sup>10</sup> FBI'a göre ABD de sağlıkla ilgili dolandırıcılıkların

7 Pandey, D., Sheth, K N. (2016). Insurance Fraud Crime Without Victims, Role of Financial Industry in Accelerating Economic Growth, s. 237.

8 Özüdoğru, Haşim. (2017). Türkiye Sigortacılık Sektörünün Değerlendirilmesi, Bankacılık ve Finansal Araştırmalar Dergisi, Cilt: 4, Sayı: 2, s.45.

9 Demiray Erol, E., Alma, D. (2016). Kasko Sigorta Tercihini Etkileyen Faktörlerin Belirlenmesi: Celal Bayar Üniversitesi Personeline Uygulama, Selçuk Üniversitesi İktisadi ve İdari Bilimler Fakültesi Sosyal Ekonomik Araştırmalar Dergisi, Sayı:32, s.151.

10 Association of Certified Fraud Examiners[ACFE], (2018), Report to the Nations 2018 Global Study on Occupational Fraud and Abuse, s.8.

yılda 40 Milyar Dolara mal olacağı tahmin edilmektedir.<sup>11</sup> **Sigorta Dolandırıcılığına Karşı Koalisyon'un** tahminine göre. ABD de dolandırıcılar tüm sigorta branşlarında yılda 80 milyar dolar çalıyor.<sup>12</sup> Mülk sigortalarında bu nedenle ortaya çıkan yıllık zarar 34 Milyar Dolar. (Sigorta Bilgi Enstitüsü, Eylül 2017).

Ülkemizde bu tür hilelerin ne boyutta olduğuna dair istatistiki bilgilere erişilememekle birlikte esasen Amerika'dan verdiğimiz bu rakamlar bize hile fiilinin işletmelere ve dolayısıyla ekonomiye verdiği zararın ne boyutta olduğunu göstermektedir.

İngilizce' de " fraud " kelimesinin Türkçe karşılığı olan "hile" kavramı esasen Türk Ceza Kanunu kapsamında tanımlanan dolandırıcılık<sup>13\*</sup> suçuyla örtüşmekte olup, basit olarak tanımlamak gerekirse; bir başka kişi ya da grubun zararına neden olacak bir biçimde, yarar sağlamak amacıyla, kasıtlı olarak aldatma faaliyetinde bulunmaktır. Hile; aldatma, sahtecilik hırsızlık, zimmet, güveni kötüye kullanma, önemli gerçekleri gizleme gibi eylemleri açıklamakta ve kavramakta kullanılan daha geniş bir anlam taşımaktadır.<sup>14</sup>

Tanımdan anlaşılacağı üzere hile, kasti bir eylem olup bilerek ve tasarlanarak gerçekleştirilir. Kişileri hileli eylemlere yönelten pek çok teşvik edici etken söz konusudur. Amerikalı suç bilimci Dr. Donald Cressey kişileri hileli davranışlara iten faktörleri Hile Üçgeni (Fraud Triangle) olarak bilinen teorisi<sup>15</sup> ile açıklamıştır. Bu teorisi ile Cressey ve onun izlenimcileri kişilerin hile eylemini gerçekleştirmelerinde üç temel etkenin - **baskı, fırsat ve bahane** - etkili olduğunu ortaya koymuştur.<sup>16</sup>

**Baskı ve Teşvik:** Kişilerin maddi sıkıntılarının olması, lüks yaşam tutkuları, çeşitli bağımlılıklar, kötü yönetim vb. örgütsel hataların yarattığı baskı ve teşvikler gibi pek çok neden kişileri hile yapmaya iter.

**Fırsat:** İşletme yönetimin zayıf veya etkisiz kontrolleri, dürüstlük ve etik değerlere bağlılığın olmayışı, yöneticilerin katı ve adil olmayan davranışları, görevlerin ayrılığı ilkesinden yoksun örgütsel yapılandırmadaki ve uygulanan insan kaynakları politikalarındaki hatalar ve yanlışlıklar, caydırıcı olmayan cezalar, şüpheli hallerin raporlanmasına imkan veren iletişim kanallarının olmayışı vb. etkenler kişilere hile için uygun ortam yaratır.

11 Crews, S. (2015). The Fraud Triangle. <https://www.brumellgroup.com/news/the-fraud-triangle-theory/> (erişim tarihi: 23.10.2019).

12 Coalition Against Insurance Fraud, By the Numbers: Fraud Statistics (t.y) <https://www.insurancefraud.org/statistics.htm> (erişim tarihi: 23.10.2019).

13 \* 5237 Sayılı TCK,157 maddesine göre : *hileli davranışlarla bir kimseyi aldatıp, onun veya başkasının zararına olarak, kendisine veya başkasına bir yarar sağlamak dolandırıcılıktır.*

14 Bozkurt, N. (2009). İşletmelerin Kara Deliği Hile, 2. Baskı, İstanbul, Alfa Yayıncılık, s.60.

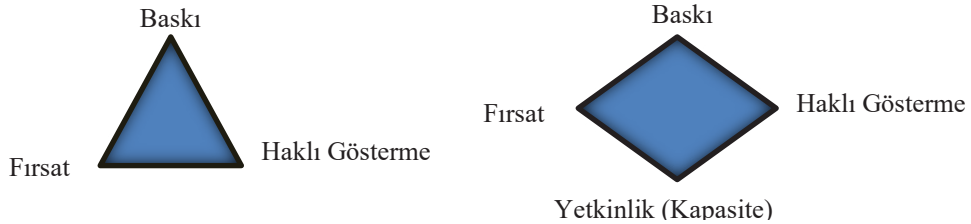
15 Machado, M., Gartner, I.R. (2018). The Cressey Hypothesis (1953) and an Investigation Into The Occurrence Of Corporate Fraud: An Empirical Analysis Conducted in Brazilian Banking Institution. [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S151.970.7720180.001.00060&script=sci\\_abstact](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S151.970.7720180.001.00060&script=sci_abstact) (Erişim tarihi: 20.10.2019)

16 Cömert, N. *İç kontrollerle İlgili Teorik Çerçeve ve COSO İç Kontrol Yaklaşımı*, ss.126-132, Bütünleşik Yaklaşımla KOBİlerde Risk temelli İç Kontrol (Editörler: Ganite Kurt-Tuğba Uçma Uysal). Yazıcı, N. (2018). Hile Üçgeni: Fırsatların Önlenmesinde Kırık Camlar Teorisi, Üçüncü Sektör Sosyal Ekonomi, 53 (3) : 843-851, s.847., Emir, M. (2008). Hile Denetimi, Mali Çözüm Dergisi, Sayı: 86, s.117.

**Bahane:** Hile yapan kişiler genellikle yaptıkları hile eylemini içselleştirerek kendilerince haklı bir gerekçeye dayandırmaktadırlar. Özellikle zayıf ahlaki değerlere sahip olmaları halinde yaptıkları bu gayri meşru eylemi zihinlerinde akla yatkın hale getirmektedirler.<sup>17</sup>

Wolfe ve Hermanson, 2004 yılında Hile Üçgeni Teorisinde açıklanan bu üç unsura, dördüncü bir etken olarak “yetenek” (capability) unsurunu eklemiştir<sup>18</sup>, hile eylemini gerçekleştirecek kişinin yeteneğinin de önemli olduğunu ileri sürerek hilekarın zeki, kendine güvenen, ikna kabiliyeti yüksek, etkili bir şekilde yalan söyleme kabiliyetine sahip ve stresle baş edebilme yetkinliklerinin olması gerektiğini belirtmişlerdir.<sup>19</sup>

**Şekil 1:** Cressey Hile Üçgeni ile Wolfe ve Hermanson Hile Karosu



Baskı ve fırsat unsurlarının etkisiyle işletmelerde yöneticiler, çalışanlar, müşteriler ya da satıcılar tarafından çeşitli şekillerle hile fiili gerçekleştirilebilmektedir. Üst yöneticiler daha çok kişisel menfaat sağlamak ya da işletmeyi olduğundan daha başarılı göstermek gibi amaçlarla finansal durumu olduğundan daha iyi ya da daha kötü göstererek yatırımcıları, hissedarları ve Mali İdareyi aldatabilirler.<sup>20</sup> Çalışanlar, işletmenin sahip oldukları para, stok, sabit kıymet, alacak vb. varlıkları suiistimal edebildikleri gibi rüşvet, işletmenin ticari sırlarını satma vb. eylemlerle çok çeşitli şekillerde menfaat ve haksız kazanç sağlayabilirler.

ACFE' nin 2018 yılında yayınladığı raporda Hile Ağacı (Fraud Tree) adını verdiği sınıflandırmaya göre işletmelere yönelik hileler; **varlıkların kötüye kullanılması** (asset misappropriation), **finansal tablo hileleri** (financial statement fraud) ve **yolsuzluk** (corruption) olmak üzere temelde üç gruba ayrılmıştır. ACFE' nin söz konusu Hile Ağacı aşağıda **Tablo 1** de verilmiş olup tablodan da görüleceği üzere **Yolsuzluk**; çıkar çatışmaları, rüşvet, illegal hediyeler ve irtikap (haraç) olarak dört ana başlığa ayrılmıştır. **Varlıkların kötüye kullanımı**; nakit hileleri ve nakdi olmayan hileler şeklinde ikiye ayrılmış olup nakit hileleri ise doğrudan nakit hileleri, hileli ödemeler ve makbuz

17 Emir, s.119

18 Ayrıntılı bilgi için bakınız: CPA Journal: The Fraud Diamond: Considering the Four Elements of Fraud By David T. Wolfe and Dana R. Hermanson, <https://pdfs.semanticscholar.org/c9c8/32fa299f648464cbd0172ff293f5c35684b6.pdf> (Erişim tarihi: 10.10.2019)

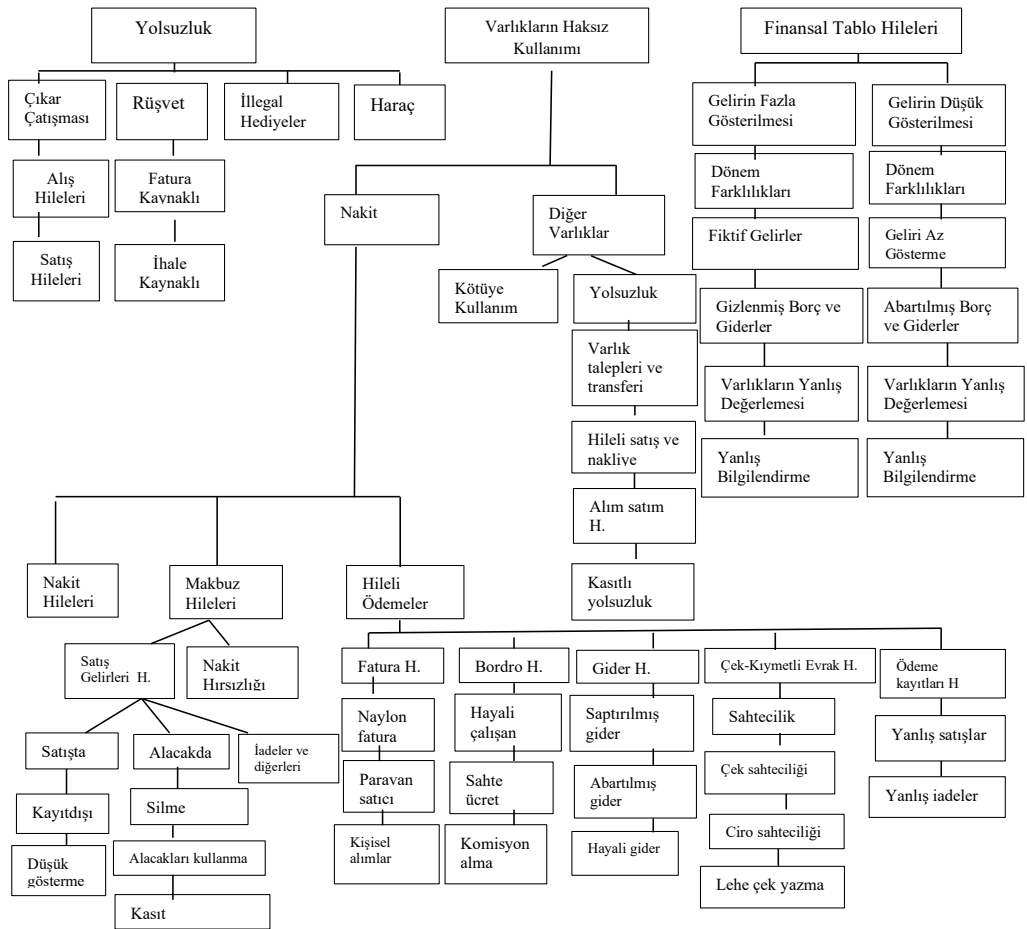
19 Tarhan Mengi, B. (2012) Hile Denetiminde Yetkinliklerin Değerlendirilmesi-Hile Karosu, Mali Çözüm Dergisi, s.120-121.

20 Ertikin, Kıvanç. (2017) Hile Denetimi: Kırmızı Bayrakların Tespiti için Kullanılan Proaktif Yaklaşımlar, Muhasebe ve Finansman Dergisi, s. 76-77.

hileleri şeklinde üçe ayrılmıştır. Nakdi olmayan hileler ise varlıkların kötüye kullanılması ve yolsuzluk olarak ikiye ayrılmıştır. **Finansal tablo hileleri** ise gelir ve varlıkların fazla gösterilmesi ve gelir ve varlıkların az gösterilmesi olarak ikiye ayrılmıştır.

ACFE'nin her iki yılda bir yayınladığı hile raporunda pek çok bilgiyle birlikte bu hilelerin boyutları açıklanmakta olup 2018 yılında yayınladığı raporda işletmelerin maruz kaldığı hilelerin türleri, gerçekleşme yüzdeleri ve parasal tutarları **Tablo 2** de verilmiştir. Tablodan da görüleceği üzere varlıkların haksız kullanımının (asset misappropriation) %89 (yüzde seksen dokuz) oranla en yaygın hile türü olduğu buna karşılık maliyetinin düşük olduğu, Finansal tablo hilelerinin (financial statement fraud) ise en nadir görülmesine rağmen parasal tutarının çok yüksek olduğu anlaşılmaktadır.

**Tablo 1: 2018 ACFE Raporuna Göre Hile Ağacı**



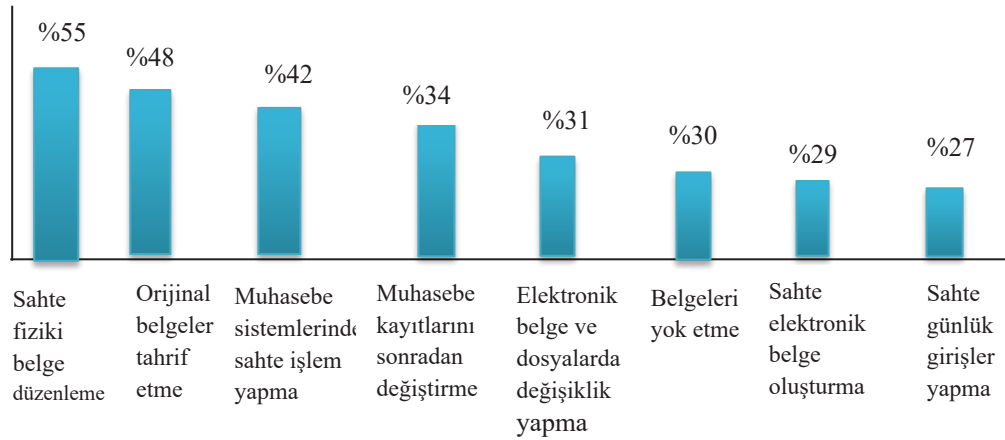
**Kaynak:** ACFE, 2018, s.11.

**Tablo 2:** 2018 ACFE Raporuna Göre Hile Türleri

Hile Türleri	Gerçekleşme Yüzdeleri	Maliyeti
Varlıkların Haksız Kullanımı (asset misappropriation)	%89	144.000 \$
Yolsuzluk (corruption)	%38	250.000\$
Finansal Tablo Hileleri (financial statement fraud)	%10	800.000\$

**Kaynak:** ACFE (2018) “ Report to the Nations on Occupation Fraud and Abuse” <https://www.acfe.com/report-to-the-nations/2018/> ( erişim tarihi: 14.04.2019)

ACFE’ nin 2018 yılında yayınlamış olduğu raporda, hilekarların gerçekleştirdikleri hileli eylemlerinin üstünü örtmek amacıyla neler yaptıkları oransal olarak **Şekil 2** de gösterilmiştir. Sahte fiziki belge düzenleme, orijinal belgeleri tahrif etme, muhasebe sistemlerinde sahte işlemler yapma, muhasebe kayıtlarını sonradan değiştirme, belgeleri yok etme, sahte elektronik belge ve dosya oluşturma vb. şekilde görülen bu yöntemleri anlamak işletmelerin hileleri önleme mekanizmalarını daha iyi tasarlamalarına ve kırmızı bayrak denilen hile uyarı işaretlerini tespit etmelerine yardımcı olacaktır.<sup>21</sup>

**Şekil 2:** Hileleri Gizlemede Kullanılan Yöntemler

**Kaynak:** ACFE (2018), s.14.

Esasen hile, tüm sektörleri tehdit eden, birçok sahada kendini gösteren ve herkese zarar veren ekonomik bir suçtur. 2018 ACFE raporuna göre hileden etkilenen sektörler ile sektörlerin karşı karşıya kaldığı hile türleri oransa olarak **Tablo 3** de sunulmuştur. Modern ekonomilerde hayati bir rol oynayan ve küresel bağlamda büyük endüstrilerden biri olarak kabul edilen sigorta

<sup>21</sup> ACFE (2018), s.14.

endüstrisinin de ne ölçüde hilelere açık bir sektör olduğu Tablodan görülmekte olup bu sektörde en fazla başvuru hile türünün %45 ile (corruption) yolsuzluk olduğu anlaşılmaktadır.

**Tablo 3:** Hileden Etkilenen Sektörler

Sektör	Olay Sayısı	Fatura Hileleri (Yüzde)	Nakit Hileleri (Yüzde)	Çek Hileleri (Yüzde)	Yolsuzluk (Yüzde)	Gider Hileleri (Yüzde)	Hileli Finansal Raporlama (Yüzde)	Nakdi Olmayan Hileleri (Yüzde)	Bordro Hileleri (Yüzde)	Yazarfaksa Hileleri (Yüzde)
Bankacılık ve Finansal Hizm.	338	11	14	12	36	7	8	11	2	3
İmalat	201	27	8	12	51	18	10	28	5	3
Kamu Kurumları	184	15	11	9	50	11	5	22	7	2
Sağlık	149	26	7	13	36	16	11	19	17	1
Perakende	104	20	10	9	28	8	12	34	5	13
Eğitim	96	23	19	6	38	18	6	19	6	0
<b>Sigorta</b>	<b>87</b>	<b>20</b>	<b>9</b>	<b>18</b>	<b>45</b>	<b>8</b>	<b>7</b>	<b>11</b>	<b>3</b>	<b>1</b>
Enerji	86	20	2	12	53	10	3	27	7	2
İnşaat	83	37	12	19	42	23	16	23	14	1
Ulaşım-Depolama	79	25	8	9	46	15	8	28	3	3
Teknoloji	62	26	5	8	42	21	16	32	8	0
Sosyal Hizmetler	58	40	9	19	34	29	10	19	22	3
Eğlence	50	14	20	6	32	12	8	18	4	8

**Kaynak:** ACFE (2018) “ Report to the Nations on Occupation Fraud and Abuse” <https://www.acfe.com/report-to-the-nations/2018/> (erişim tarihi: 14.04.2019)

Latince “scurta” kelimesinden türetilen ve “güvence” anlamına gelen sigorta terimi, ticari hayatımızı ve ilişkilerimizi düzenleyen Türk Ticaret Kanunu’nun 1401’ inci maddesinde “Sigortacının bir prim karşılığında, kişinin para ile ölçülebilir bir menfaatini zarara uğratan tehlikenin, rizikonun meydana gelmesi halinde bunu tazmin etmeyi ya da bir veya birkaç kişinin hayat süreleri sebebiyle ya da hayatlarında gerçekleşen bazı olaylar dolayısıyla bir para ödemeyi veya diğer edimlerde bulunmayı yükümlendiği sözleşme” olarak tanımlanmıştır. Sigorta kavramının hukuki boyutunu ortaya koyan bu tanıma göre sigorta; kanun ve sözleşmeler ile tespiti yapılan, çerçevesi belirlenmiş bir riskten benzer derecede tehdit alan belli bir sayının üstündeki grubun, tesadüf eseri ortaya çıkan zararın beraberce üstesinden gelmek amacıyla oluşturduğu



organizasyon<sup>22</sup> olarak tanımlanmıştır. Sigorta, riskleri sigortalılardan sigorta şirketlerine taşıdığı için arzu edilen bir seçenektir.<sup>23</sup> Sigorta, sistematik bir devirdir. Önceden yapılan ödemeler karşılığında zarar riskine karşı koruma sağlamak için kullanılır.<sup>24</sup>

Sigorta sektörünün sosyo – ekonomik çöküntü ve kayıpları önleme, elde ettiği fonları yatırım alanlarına aktararak ekonominin ve sermaye yapısının güçlenmesine, devlet idaresine mali kaynak sağlama, ülkeler arası ekonomik ilişkileri güçlendirme, istihdam yaratma gibi işlevleri göz önünde bulundurulduğunda sektör ekonomi, insan hayatı ve devletler açısından önemli alanlardan biridir. Sigorta sektöründe, gerçekleşen hasarların sigorta şirketlerinin kontrolünün dışında cereyan etmesi, sigorta sektörünün çok taraflı olması sigorta müessesinin hilelere açık bir yapıya bürünmesine sebep olmuştur. Yanlış Sigorta Uygulamalarının Tespiti, Bildirimi, Kaydı ve Bu Uygulamalarla Mücadele Usul ve Esasları Hakkında Yönetmelik kapsamında **sigorta hilesi** “*Sigorta ilişkisi içerisinde yer alan taraflardan ya da bu ilişkide rol oynayan kişilerden bir ya da bir kaçına haksız menfaat sağlamaya yönelik her türlü fiil*” olarak tanımlanmıştır. Türkiye Sigorta Birliğine göre sigorta hileleri “*kötü niyetli kişilerin haksız kazanç elde etmek amacıyla sigorta şirketlerini bilerek, kasıtlı olarak aldatmasıdır.*” şeklinde tanımlanmıştır.

Sigortada görülen hile türleri çeşitlilik göstermekte, sigortanın tüm branşlarında ortaya çıkabilmekte ve poliçe düzenlenmeden önce ya da poliçe düzenlendikten sonra da gerçekleştirilebilmektedir. Sigorta hileleri, sigorta şirketlerine bilinçli ve kasıtlı olarak yanlış bilgi vermek ya da önemli bir hususu gizleyerek tedbirli bir sigortacının normalde girmeyeceği bir sözleşmeye girmesini sağlamak veya gerçekte var olmayan hasarı varmış gibi göstermek ya da teminat verilmeyen bir rizikonun gerçekleşmesi halinde bunu teminat verilen bir durumun arkasına gizlemek suretiyle de gerçekleştirilebilir.<sup>25</sup> Bu makalenin odak noktası araç sigortası olduğundan makalemiz kapsamında daha çok bu alanda görülen hile türlerinden bahsedilmiştir. Sıklıkla birlikte ya da tek başına gerçekleştirilen araç sigortası hilelerinin daha çok şunlar olduğu görülür:<sup>26</sup>

- Araç sahibinin mali destek sağlamak amacıyla aracının çalındığı iddiasında bulunulması
- Yanlış ya da eksik bilgi vermek suretiyle sigorta tazmin talebinde bulunulması,
- Sahte hasar beyanı,
- Sahte kaza tutanağı düzenlenmesi,
- Planlanmış ya da organize hasar tertip edilmesi,
- Hasarın şişirilmesi,

22 Pekiner, 1981, s.17.

23 Ghorbani, A., Farzai, S., 2018, s.3764.

24 Pandey ve Sheth, 2016, s.237.

25 Türkiye Sigorta Birliği, [www.tsb.org.tr/sigorta-tanimlari.aspx?pageID=648](http://www.tsb.org.tr/sigorta-tanimlari.aspx?pageID=648) (erişim tarihi: 10.12.2017)

26 Subudhi, S. ve Panigrahi, S. (2018). Detection of Automobile Insurance Fraud Using Feature Selection and Data Mining Techniques, International Journal of Rough Sets and Data Analysis, Cilt: 5, Sayı: 3, s.1.; <http://www.sigortagundem.com/haber/tamir-istasyonundaki-sahtekarliklar/1036371> (erişim tarihi: 12.12.1017)

- Tamir istasyonlarında çalıntı araçların parçalara ayrılıp satılması, değiştirilmemiş, tamir edilmemiş ya da gerçekte takılmamış parçaların faturalandırılması, kalitesiz yedek parça kullanılması, araç tamirinde hasarlı parçayla aynı kalite ve standartta eşdeğer yedek parça kullanıldığı halde daha pahalı olan orijinal parça bedelinin faturalandırılması,
- Araçtaki mevcut hasarın ekspertiz öncesinde kasten artırılıp araca ilave zarar verilmesi,
- Kaza sonrasında olay yerinde beliren araç çekicilerince gerçek dışı beyan ve vaatlerle hasarlı aracı istedikleri tamirhaneye çekilmesi ve sonrasında araç sahibi adına çoğu durumda onay alınmaksızın gerçek dışı belge ve beyanlarla hasar başvurusu yapılması,
- Hasarlı araç fotoğraflarının kasten yanıltıcı izlenim verecek şekilde kalitesiz çekilmesi ve geç gönderilmesi gibi çeşitli şekillerde sahte ya da yanıltıcı beyan ve belgelerle hasar tazmin talebinde bulunulması,
- Sürekli aynı araçların kazaya karışması,
- Hasar sonrası sigorta yaptırma,
- Vurdu-kaçtı beyanı,
- Sürücü değişikliği ya da sürücü firarı

Sigorta hileleri sadece kazanç elde etme amacı güden şahıslarca gerçekleştirilmemekte, yaşadığı hasardan dolayı sigorta şirketince mağdur olan ya da ödediği primi yüksek bulan ya da yaşamının güncel seyrinden olumsuz etkilenen sigortalı tarafından yapılabileceği gibi bilinçli olarak yanlış yönlendirme ve bilgilendirmede bulunan, güveni, yetki ve sorumlulukları kötüye kullanan şirket içerisinde her kademedeki personel tarafından, tazminat talebinde bulunanlar tarafından ya da aracılar tarafından gerçekleştirilebilir.<sup>27</sup> Kısaca ifade etmek gerekirse sigortalı hilenin tarafı olabileceği gibi acenteler, şirket yöneticileri ve çalışanlar, eksperler, tamirhaneler, tedarikçi ya da sağlık hizmeti sunucuları da hileye taraf olabilmektedir.

Kötü niyetli kişilerin kazanç sağlamak amacıyla hilelere başvurması sigorta müessesesine olan güvenin kaybedilmesine, sigorta şirketlerinin bilançolarının olumsuz yönde etkilenmesine, hile sebebiyle ortaya çıkan ilave maliyetlerin dürüst sigortalılar üzerinde kalmasına, yapılan hilelere yönelik alınan önlemlerin süreç ve prosedürleri uzatmasına ve dolayısıyla hizmet kalitesinde sıkıntılar yaşanmasına sebep olmakta, sistemin ahlaki yönden işleyişini ve sektörün büyüklüğünü olumsuz yönde etkilemektedir.

### 3. Büyük Veri ve Veri Analitiği Hakkında Genel Bilgi

Genel olarak veri / data kavramı, işlenmemiş bilgi anlamına gelmekte olup çeşitli şekillerde tanımlanmaktadır. Birbirine benzeyen bu tanımlardan bazılarını açıklanmak gerekirse *veri*; karar almaya yardımcı olmak amacıyla kullanılan ve dikkate alınan, incelenip değerlendirilmek üzere

<sup>27</sup> Gezgez, Z.T. ve Yöndem, H. (2016). Sigortacılıkta Şüpheli Hasarlar, Suistimaller ve Sahtekarlıklar, Reasürör Dergisi, İstanbul, s. 5.

toplanan ya da elektronik formda bir bilgisayar tarafından saklanıp kullanılabilen nitelikteki bilgi, sayılar ve olgulardır.<sup>28</sup>

Bir başka tanıma göre **veri**; “herhangi bir işleme tabi tutulmadan, gözlem ya da ölçüm yöntemleriyle çeşitli ortamdan sağlanan değer”dir.<sup>29</sup> Diğer bir tanıma göre **veri** “tek başına anlam ifade etmeyen veya kullanılmayan, bununla birlikte enformasyona ve bilgiye temel oluşturan ilişkilendirilmeye, gruplandırılmaya, yorumlanmaya anlamlandırılmaya ve analiz edilmeye gereksinim duyulan ham bilgi”dir.<sup>30</sup> Çeşitli ortamlardan elde edilen veriler faydalı bilgiye dönüştüğü takdirde bir anlam ifade edecektir. Michael Palmer “Veri sadece ham petroldür, rafine edilmezse değeri vardır ama kullanılabilir değildir” ifadesiyle veriye dayalı bir yaklaşımı benimseyerek veriyi iş süreçlerine dahil edenlerin rekabet avantajı elde edeceğini ileri sürmektedir. Verinin gerçek değeri, buzdağının görünen yüzüdür. İlk bakışta sadece suyun üstünde kalan kısım görünür. Denizin altında gizlenmiş olan büyük bölümünü ise ancak yenilikçi şirketler ortaya çıkarabilirler.<sup>31</sup>

Dünyada üretilen, paylaşılan veri miktarındaki artış ve çeşitlilik yeni kavramların ortaya çıkmasına yol açmıştır. Bu kavramlardan biri de **büyük veri** (big data) kavramıdır. **Büyük veri** kavramı, geleneksel veri tabanı teknikleri kullanılarak işlenmesi mümkün olmayan, farklı hacimlerdeki karmaşık yapıdaki verileri tanımlayan bir kavramdır. **Büyük veri**, genellikle farklı kaynaklardan gelen büyük veri yığınlarının analizine, işlenmesine ve depolanmasına yönelik bir alandır. Büyük veri özellikle çok sayıda ilişkisiz veri kümesinin birleştirilmesi, büyük miktarda yapısal olmayan verilerin işlenmesi ve gizli bilgilerin toplanması ihtiyacına ilişkin olup, geleneksel veri analizi işlemleri, depolama teknolojileri ve teknikleri yetersiz kalmaktadır.<sup>32</sup> **Büyük veri** kavramı; *verinin hızı, çeşitliliği, hacmi, gerçeklik ve değer* bileşenlerini içermekte olup bu bileşenlerin neyi ifade ettiği aşağıda kısaca açıklanmıştır;<sup>33</sup>

**Verinin hızı**; verinin sürekli hareket halinde olması ve bu bağlamda veriyi işleyecek, analiz edecek süreçlerin de aynı hızla olması demektir.

**Verinin çeşitliliği**; geniş bir yelpazede herhangi bir türde ve formatta üretilebilmesi, veri türleri arasında standart bir dizi ya da kural olmaması durumunu ifade etmektedir.

**Verinin hacmi**; üretilen, depolanan ve iletilen veri miktarında üstel bir artış olması anlamına gelmektedir.

28 <https://dictionary.cambridge.org/dictionary/english/data> (erişim tarihi: 23.10.2019).

29 Şeker, Ş.E. (2013). İş Zekâsı ve Veri Madenciliği, İstanbul, Cinius, s.22.

30 Yılmaz, Malik. (2009). Enformasyon ve Bilgi Kavramları Bağlamında Enformasyon Yönetimi ve Bilgi Yönetimi, Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi 49, 1 (2009): 95-118, s.98.

31 Doğan, K., Arslantekin, S., (2016). Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum, DTCF Dergisi 56, 1 (2016): 15-36, s. 20.

32 Thomas Erl ve diğerleri (2016). Big Data Fundamentals, Concepts, Drivers & Techniques. Indiana, Arcitura Education Inc, s.19.

33 Aktan, Ertuğrul. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu, Ankara Üniversitesi Bilgi Yönetimi Dergisi, Cilt: 1, Sayı: 1, s. 4-5.

**Gerçeklik;** verinin iş kararlarında kullanılabilir ölçüde güvenilir olmasıdır. Bu nitelik verinin kalitesiyle değerlendirilmektedir.

**Değer;** verinin organizasyonlar ve toplum için ekonomik bir değer yaratabilecek anlayışa dönüşmesidir.

Günümüzde bilgi ve iletişim teknolojilerinin, internetin gelişmesi dolayısıyla web sayfaları, bloglar, sosyal medya uygulamaları, sensörler ve daha pek çok veri toplayan cihaz ve uygulamalar, birbirine bağlanan ve birbiriyle haberleşen cihazların artması sayesinde her an bilimsel ve bilimsel olmayan veriler toplanmaktadır. Bu durum veri patlamasına neden olmaktadır. Toplanan bu veriler pek çok alanda yapılan ve yapılacak araştırmalar için kullanılabilir bir nitelik taşıdığından bu verilerin analizi oldukça önemlidir. Veriler toplanıp işlendikten sonra bunlardan katma değeri yüksek sonuçlar çıkarılması veri analizi olarak adlandırılmaktadır. Bu verilerin analizi için geliştirilen bir yöntem de **veri madenciliğidir**.

**Veri madenciliğini**, “bilgisayar teknolojilerinin sağlamış olduğu çok hızlı veri işleme ve yüksek hacimde veri depolama imkanları yardımıyla ve farklı disiplinlerin (yapay zeka, makine öğrenmesi, uzman sistemler, veri tabanı teknolojileri, paralel bilgi işleme, dağıtık veri işleme, görselleştirme, optimizasyon, veri ambarcılığı, istatistik vb.nin) katkısıyla sağlanan araçlarla, sahip olunan çok büyük hacimlerdeki veriden, karar vericinin etkili ve daha fazla bilgiye dayalı karar vermesinde kullanılabilmesi amacıyla önceden bilinmeyen, gizli, örtük, klasik metotlarla ortaya çıkarılması güç, faydalı, ilginç, anlaşılabilir ilişki, örüntü, bağıntı veya trendlerin otomatik veya yarı otomatik bir şekilde ortaya çıkarılması” olarak tanımlamak mümkündür.<sup>34</sup>

**Veri madenciliği**, veri tabanındaki önemli, yararlı bilgiler ile veri ilişkilerini analiz edebilen tüm teknolojileri bir araya getirmektedir.<sup>35</sup> **Veri madenciliği**, veri ambarlarında saklanan ve yararlı olabilecek veriler ile aralarında bilinmeyen ilişkilerin olduğu verileri keşfederek bu verileri anlaşılır ve kullanılabilir bir şekle dönüştürmeye yönelik olarak geliştirilen yöntemler topluluğu<sup>36</sup> ile açık olmayan bir süreçtir.<sup>37</sup>

**Veri madenciliği** hemen hemen tüm alanlarda uygulanmaya başlanmıştır. Alınan sonuçların faydaları görüldükçe popülerliği artmıştır.<sup>38</sup> Bugün ekonomik ve ticari faaliyetlerde, kamu yönetiminde, sağlıkta, eğitim faaliyetlerinde, terör ve askeri alanda, bilimsel araştırmalarda ve bunun gibi birbirinden farklı daha birçok alanda veri madenciliği uygulamalarına rastlamak mümkündür. MIT (Massachusetts Institute of Technology)’ın yayınladığı bildiride veri madenciliği

34 Şentürk, A. (2006). Veri Madenciliği Kavram ve Teknikleri, Bursa, Ekin Yayınevi, s. 3.

35 Monino, J. – L., Sedkaoui, S., (2016). Big Data, Open Data and Data Development, Vol: 3, London, ISTE Ltd, s. XIII.

36 Oğuzlar, A. (2011). Temel Metin Madenciliği, Bursa, Dora, s. 5-6.

37 Akpınar, H. (2014). Data Veri Madenciliği Veri Analizi, 1. Baskı, İstanbul, Papatya Yayıncılık, s.50.

38 Savaş, S. ve diğerleri. (2012). Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Sayı: 21, İstanbul, ss.1-23, s. 4-5.

dünyayı değiştirecek on teknoloji arasında gösterilmiştir.<sup>39</sup> Çeşitli alanlarda veri madenciliğinden yararlanılması kurum ve kuruluşların bu konuya ciddi yatırımlar yapmasını sağlamıştır.

**Veri madenciliği** temelde tanımlayıcı ve tahmin edici olmak üzere iki başlıkta ele alınmaktadır. **Tahminleyici modellerde;** sonuçları bilinen veriler baz alınarak bir model geliştirilmekte ve bu model yardımıyla sonuçları bilinmeyen veri kümesinin sonuç değerlerinin tahmin edilmesi amaçlanmaktadır.<sup>40</sup> Sınıflandırma ve regresyon başlıca tahminleyici modellerlerdir.

**Sınıflandırmada;** veri tabanında yer alan girdiler çeşitli niteliklerine göre sınıflara atanmakta ve ileride veri kümesine dahil olacak yeni kayıtların hangi sınıfa ait olacağı belirlenmeye çalışılmaktadır. **Sınıflandırma** modeliyle veri tabanında yer alan nesnelere özellikleri isimleriyle ilişkilendirilir. Birbirinin aynı olan ve birbirinden farklı olanlar bulunur ve daha sonra kendisine verilen değişik özellikleri değerlendirilerek bu nesnenin ismi tahmin edilir. Tıpkı bir çocuğun cinsiyetleri sınıflandırması gibi. Çocuk için ilk önce cinsiyet kavramı yani bir sınıflandırma yoktur. Daha sonra anne – baba, teyze, hala, amca, kendinden büyük ve küçük erkek ve kız çocuklarını görür. Aslında tüm bunlar çocuk için bir veri tabanıdır. Bu veri tabanını inceleyen çocuk kadınla erkek arasındaki temel farkları belirler ve daha sonra kendisine hiç tanımadığı kız çocuğu gösterildiğinde bir önceki deneyimine / öğrenmesine dayanarak bunun kız çocuğu olduğuna karar verir. Aslında yaptığı tamamen bir sınıflandırma işlemidir.<sup>41</sup> **Regresyon modeli ise,** sadece sayısal (nümerik) verilerle ilgilenen tahmin modelidir. **Karar ağaçları, yapay sinir ağları, genetik algoritmalar, k-en yakın komşu, naive bayes** başlıca sınıflandırma ve regresyon modelleri olup alt başlıklar halinde kısaca açıklanmıştır.

**Karar ağacı;** sınıfları bilinen veri örneklerinden yola çıkılarak kural tabanlı model oluşturulmaya çalışılır. Ağaç görünümüne benzeyen bir akış üzerine oturtulmuş olan model, ağaç yapısının hiyerarşik olarak en üst seviyesinde kök düğüm (ilk düğüm), en alt seviyesinde yaprak ve kök düğüm ile yaprağı bağlayan dal diye adlandırılan yapılardan oluşur.<sup>42</sup> Karar ağacı öğrenimi, kök düğümden yapraklara doğru sürekli daha homojen düğümleri elde edecek şekilde bütünü en iyi bölen özneliği bularak kendini tekrarlayan (recursive) süreçtir.<sup>43</sup> Her bir düğüm öznelik (değişken) değerini, dallar test çıktısını, yapraklar ise sınıfları temsil etmektedir. Gerçekleşen her bölünme işlemi bir öğrenme sürecini ifade etmektedir.<sup>44</sup> Dallanma işleminin durması sınıflandırmanın tamamlanmış olduğunu göstermektedir. Karar ağacı çok büyük veri setlerini daha küçük istatistiksel anlamı olan veri setlerine dönüştürürken bazı karar kuralları

39 Sarıman, G. (2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K – Medoids Kümeleme Algoritmalarının Karşılaştırılması, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 15-3, Isparta, ss: 192-202, s.192

40 Özdemir, A. ve diğerleri (2018). Eğitim Sisteminde Veri Madenciliği Uygulamaları ve Farkındalık Üzerine Bir Durum Çalışması, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Sayı:22 (Özel Sayı), ss: 2159-2172, s.2162.

41 Silahtaroglu, G.(2013).Veri Madenciliği Kavram ve Algoritmaları, 2. Baskı, İstanbul, Papatya Yayıncılık, s.50.

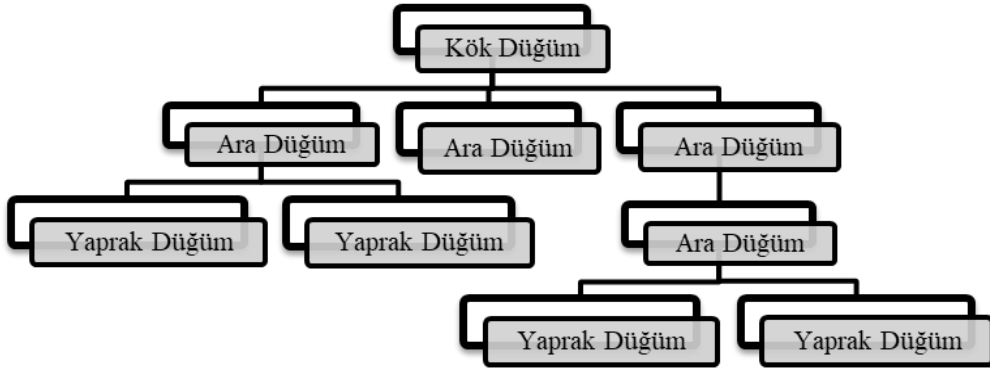
42 Kaya, Ç. ve Yıldız, O. (2014). Makine Öğrenmesi Teknikleriyle Saldırı Tespiti: Karşılaştırmalı Analiz, Marmara Fen Bilimleri Dergisi, 2014, 3: 89-104, s.93.

43 Akpınar, 2014, s. 218.

44 Akpınar, 2014, s. 205.

oluşturmaktadır. Oluşturulan karar kurallarının görsel olarak ağaç yapısına benzemesi ise kuralların daha net bir şekilde anlaşılması sağlayacaktır. Bu sebeple karar ağacı kurgulanması, kurulması ve yorumlanması kolay bir modeldir. Karar ağacı modeli AID, ID3, CART (C&RT), Chaid, Quest, C4.5, C5.0, Sprint, Sliq gibi birtakım algoritmalar aracılığıyla uygulanabilmektedir. Karar ağacı yapısı aşağıdaki şekilde gösterilmiştir.

Şekil 3: Karar Ağacı Yapısı



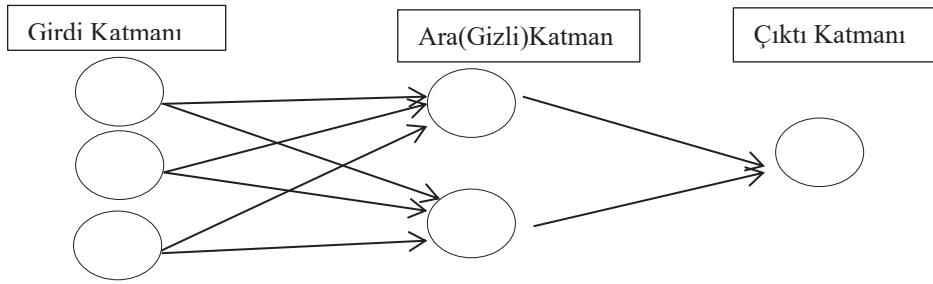
**Yapay Sinir Ağları;** Yapay Sinir Ağları, *Yapay Zekâ* kavramı altında oluşan bir alt başlık olup **Yapay Zekâ;** İnsan zekasının, sinir sistemi, gen yapısı gibi nörolojik ve fizyolojik yapısının ve doğal olayların modellenerek bilgisayar ve yazılımlara aktarılması olarak tanımlanmaktadır. Yapay zekâ, insan gibi davranışlar sergileme, sayısal mantık yürütme, hareket etme, konuşma, ses algılama gibi birçok yeteneğe sahip sistemlerdir.<sup>45</sup> Yapay zeka muhakeme yeteneğine sahip veya zekice davranabilen ve düşünce veya davranış açısından insana benzeyen veya rasyonel olabilen makineler geliştirmek üzerine bir disiplin olarak addedilmektedir. Makine öğrenmesini de içeren inter-disipliner bir alan olan yapay zekâ görüntü algılama, robotik gibi bilgisayar bilimleri dışında kalan alanlardaki çalışmalarını da kapsar. Yapay zekâ alanında yapılan çalışmaların ürünü olan yapay sinir ağları, veri madenciliğinde sınıflandırma amacıyla kullanılır. Yapay Sinir Ağları tekniği esasen insan beyninin temel işlem elemanı olan nöronlar ve bu nöronlar arasındaki bağlantılara şekilsel ve yapısal olarak benzemekle birlikte adeta biyolojik sinir sisteminin bir simülasyondur. Yapay sinir ağlarında tıpkı biyolojik sinir ağlarındaki sinir hücreleri gibi yapay sinir hücreleri mevcuttur. Yapay sinir hücreleri katmanlar şeklinde birleşerek yapay sinir ağlarını oluştururlar. Yapay sinir ağı, biyolojik sinir sisteminin bilgiyi depolama, kullanma ve işleme yeteneklerini taklit etmeyi, insan gibi karar alabilen ve muhakeme yapabilen bir sistem elde etmeyi amaçlar.<sup>46</sup> Yapay sinir ağlarında girdi, ara (gizli) ve çıktı katmanı bulunmaktadır. Girdi katmanı, dışarıdan girdileri alan nöronları içermektedir ve sadece girdi değerlerini sonraki tabakaya iletme fonksiyonu vardır. Girdi değerlerine herhangi bir işlem uygulanmamaktadır. Çıktı tabakası ise çıktıları

45 Atalay, M., Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları, Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt: 9, Sayı: 22, ss 155-172, s. 159-161.

46 Atalay ve Çelik, 2017, s. 162.

dışarı ileten nöronları içeren tabakadır. Girdi ve çıktı tabakaları tek tabakadan oluşurken bu iki tabaka arasında birden fazla gizli (ara) tabaka bulunabilmektedir.<sup>47</sup> Ara katman sayısı ağdan ağa değişmekle birlikte gereğinden az olması doğru verilere ulaşılmasını engellerken; gereğinden fazla olması ise hesaplama karışıklığına ya da analiz süresinin uzamasına sebep olabilmektedir. Yapay sinir ağı, veri setindeki yapıyı öğrenerek istenilen görevi yerine getirecek şekilde genelleştirmeler yapar. İlgili olayın örnekleri ile eğitilerek ağa genelleme yapabilme yetisi kazandırılır ve benzer olaya karşılık gelen çıktı setleri belirlenir.<sup>48</sup> Öğrenme işlemi tamamlandıktan sonra verilen her hangi bir yeni verinin, eldeki ağırlıklar yardımıyla bağlı olduğu sınıf hesaplanabilir. Yapay sinir ağları modelinin katmanları ve işleyişi aşağıda bir şekil yardımıyla gösterilmiştir. (Şekil 4)

**Şekil 4:** Yapay Sinir Ağlarının Katmanları



**Kaynak:** Akpınar, 2014, s.240.

Yapay sinir ağı modeli; çek okuma, güvenlik sistemlerinde konuşma, parmak izi, göz tanıma, ses algılama, imza tanıma, otomotivde yol izleme ve yol koşullarına göre sürüş analizinde zayıf noktalarını bulma,<sup>49</sup> askeri alanda hedef izleme, radar, sinyal işleme, nesneleri ayırma ve tanıma gibi konularda karşımıza çıkmaktadır.

**Genetik Algoritmalar;** evrim teorisi baz alınarak canlılardaki genetik sürecin bilgisayar ortamlarına aktarıldığı, doğal seçim ile mevcut çözümlerden hareket edilerek en iyi çözüme ulaşmayı ve en iyi çözümün hayatta kalması ilkesini taklit eden bir tekniktir.<sup>50</sup> Bu yöntem problemlere karşı tek bir çözüm üretmez, farklı çözümlerden oluşan bir çözüm kümesi sunar. Çözüm olayının geniş ve karmaşık olduğu problem tipleri için oldukça başarılı sonuçlar vermektedir.

**K en yakın komşu algoritması;** bu yöntemde göre sınıflandırma işlemi veri tabanındaki her bir kaydın birbirine olan mesafesine bakılarak gerçekleştirilmektedir. Burada her bir kaydın diğer kayıtlarla olan uzaklığı  $k$  tane kayıt göz önünde bulundurularak hesaplanır. Mesafesi hesaplanan

47 Şentürk, 2006, s. 29.

48 Atalay ve Çelik, 2017, s. 163.

49 Özker Dursun, G. (2016). İç Kontrol Sisteminde Bir Araç Olarak Veri Madenciliğinin Kullanılmasının İşletmenin Bilançosuna Etkisi, *Yüksek Lisans Tezi*, İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.

50 Atalay ve Çelik, 2017, s 163.

kayıtlara diğer kayıtların içerisinde en yakın olanları bu  $k$  adet kayıttır. Algoritmada  $k$  değeri önceden seçilmektedir. Bu değerin yüksek seçilmesi birbirine benzemeyen kayıtların bir arada yer almasına, değerin çok küçük seçilmesi ise birbirine benzeyen ve aynı sınıfa dahil olması gerekirken bazı noktalar sebebiyle kayıtların farklı sınıflara konulmasına neden olmaktadır. Bu yöntem coğrafi bilgi sistemlerinde yaygın olarak kullanım alanına sahiptir.<sup>51</sup>

**Navie Bayes Yöntemi;** Bu yöntem, mevcut sınıflanmış örnek verileri kullanarak yeni verinin mevcut sınıflandırılmış verilerden hangisine ait olacağını olasılıksal olarak hesaplamaktadır. Bu yöntem özniteliklerin birbirlerinden tamamen bağımsız olduğu, örneklerin hepsinin aynı önem derecesine sahip olduğu ve bir özellik değerinin diğer özellik değerleri hakkında bilgi içermediğini kabul edilmektedir.<sup>52</sup> Bu model, bir olayın meydana gelmesinde birden fazla faktör söz konusu ise bu faktörlerden hangisinin olayın ortaya çıkmasında daha etkili olduğunu hesaplamakta ve olası çıkış durumlarından en büyük olasılığa sahip olan durumu hedef sınıf olarak seçmektedir.

**Tanımlayıcı Modeller;** veriler arasındaki ilişkilerin, örüntülerin, davranışların veri madenciliği algoritmaları aracılığıyla bulunmaya çalışıldığı bir modeldir. Tanımlayıcı modellerin içerisinde en çok kullanılan modeller kümeleme ve birliktelik kuralları olup bu modeller hakkında kısaca bilgi verilmiştir.

**Kümeleme Analizi;** burada veriler benzerliklerine göre gruplara ayrılmaktadır diğer bir ifade ile benzer nesnelere aynı grupta yer alacak şekilde ayrıştırılmaktadır. Bu yöntemle göre aynı küme içerisinde yer alan nesnelere olabildiğince birbiriyle benzer olması hedeflenirken farklı kümelerde yer alan nesnelere olabildiğince ayrışık olması hedeflenir.<sup>53</sup> Örneğin; elimizde bir mağaza müşterilerinin isim ve yaş bilgilerini içeren veri tabanının olduğu ve müşterilerin yaşlarına göre kümelere ayrılmak istendiği varsayıldığında yaşları nispeten birbirine yakın olanlar aynı küme içerisinde toplanacaktır. Yaşları 20, 21, 23, 25, 27, 40, 45, 47, 49 olan müşterilerden 20-27 yaş grubu bir kümede, 40-49 yaş grubunda olanlar ise başka bir kümede toplanacaklardır.<sup>54</sup>

**Birliktelik Kuralları;** veri tabanı içerisinde yer alan kayıtların birbirleriyle olan ilişkilerini inceleyerek hangi olayların birlikte gerçekleşebileceğini ortaya koymaya çalışmaktadır.<sup>55</sup> Örneğin bir mağazadan gömlek alan müşterilerin aynı alışverişte kravat satın aldığı bilgisi mağaza işletmecileri açısından değerli bilgi olarak göz önüne alınır. Müşterilerin satın alma eğilimlerinin değerlendirilerek birlikte alınan ürünlerin birbirine yakın raflara yerleştirilmesi ise satışların artmasını sağlayabilir.<sup>56</sup> Birliktelik kurallarının temeli veriler arasındaki ilginç ilişkileri ortaya çıkarmaya dayanmaktadır. Alışveriş sırasında çocuk bezi alan müşterilerin aynı zamanda mama

51 Silahtaroglu, 2013, s 118.

52 Karakoyun, M., Hacıbeyoğlu, M. (2014). Biyomedikal Veri Kümeleri ile Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatiksel Olarak Karşılaştırılması, Dokuz Eylül Üniversitesi Mühendislik Fakültesi Mühendislik Bilimleri Dergisi, Cilt:16, Sayı:48, İzmir, s.34.

53 Akpınar, 2014, s 284.

54 Silahtaroglu, 2013, s.155.

55 Özkan, Y. (2013). Veri Madenciliği Yöntemleri, 2. Basım, İstanbul, Papatya Yayıncılık, s. 48.

56 Özkan, 2013, s.157.



da alabileceklerini öngörmek çok fazla zor bir iş değil iken çocuk bezinin yanında bira alındığını önceden öngörmeleri mümkün değildir. Dolayısıyla bu karmaşık ilişkileri ortaya çıkarmada bu yöntem önem kazanmaktadır.<sup>57</sup> Birliktelik kurallarında veriler arasındaki bağıntı ise, destek ve güven kriteri ile hesaplanmakla birlikte destek kriteri, veride ögeler arasındaki ilişkinin ne kadar sık olduğunu, güven kriteri ise A ögesinin hangi olasılıkla B ögesi ile beraber olduğunu ifade etmektedir. Örneğin bir mağazada 500 müşterinin pantolon satın aldığı ve bunların 400 tanesinin pantolonun yanında ceket de aldığı dolayısıyla pantolon alındığında ceket de almanın güven değeri  $400/500 = \% 80$  olarak bulunmaktadır. Buradan pantolon alan müşterilerin  $\% 80$ 'nin aynı zamanda ceket de aldığı sonucuna ulaşılmaktadır.<sup>58</sup> Birliktelik kuralında iki basamak söz konusu olup bunların ilki sık tekrarlanan ögelerin bulunması ikincisi ise sık tekrarlanan öğelerden güçlü birliktelik kurallarının oluşturulmasıdır. Birliktelik kurallarının temel kullanım alanlarının başında perakende sektörü gelmekle birlikte market raflarının tasarımı, promosyon yapılmasını gerektirecek ürünleri belirleme, özel ürünlere karar verme, katalog tasarımı gibi konularda sıklıkla kullanılmaktadır.

Veri madenciliği uygulamaları Kxen, SPSS Clementine, Matlab, SAS, Oracle gibi çeşitli ticari yazılımlar aracılığıyla ya da RapidMiner, Weka, R, Orange, Knime, Keel gibi kullanıcıların ücretsiz olarak erişebileceği açık kaynak kodlu bilgisayar yazılımları ile gerçekleştirilmektedir. Bu yazılımların birbirlerine karşı çeşitli üstünlükleri bulunmaktadır. Örneğin bu yazılımlar içerisinde en fazla dosya formatını desteklediğinden en kapsamlı olanları RapidMiner ve Weka iken en az algoritmaya sahip yazılım ise R' dir ancak, R yazılımı istatistiksel hesaplamada en güçlü olanıdır. WEKA kullanım ve öğrenme kolaylığı açısından kompleks bir yapıya sahip olmadığı ve çoğu sistemde sorunsuz bir şekilde çalışabilme yeteneği açısından en başarılı yazılımlardan biri olarak görülmektedir. Kod yazmada en başarılı olanı ise Orange yazılımıdır. Dolayısıyla yazılımların birbirine karşı olan üstün yönlerine bakarak kullanılacak veri kümesi ile ulaşılmak istenen hedef arasında en etkili sonuca ulaşmayı sağlayan yazılım belirlenmeye çalışılır.<sup>59</sup>

#### **4. Veri Madenciliğinin Uygulama Alanları, Sigorta Hilelerinin Tespitinde Kullanılması ve Araç Sigortalarına Yönelik Bir Uygulama Örneği**

Hemen hemen her alanda yararlanılan veri madenciliğinin kullanım alanlarını aşağıdaki gibi sıralamak mümkündür:<sup>60</sup>

- Müşterilerin satın alma örüntülerinin belirlenmesi,
- Müşterilerin demografik özellikleri arasındaki bağlantıların bulunması,
- Posta (mailing) ve SMS kampanyalarında cevap verme oranının artırılması,

57 Şentürk, 2006, s. 19.

58 Şentürk, 2006, s.20.

59 Kaya M., Özel, S.A. (2014). Açık Kaynak Kodlu Yazılımların Karşılaştırılması, 'Akademik Bilişim' 14 – XVI. Akademik Bilişim Konferansı Bildirileri, Mersin: Mersin Üniversitesi, 5-7 Şubat. s.50.

60 <https://netuce.com/veri-madenciligi-nedir-hangi-alanlarda-kullanilir/>(Erişim tarihi: 20.10.2019)

- Mevcut müşterilerin elde tutulması, yeni müşterilerin kazanılması,
- Pazar sepeti analizi,
- Müşteri ilişkileri yönetimi,
- Müşteri değerlendirme,
- Satış tahmini,
- Çapraz satış,
- Kredi kartı dolandırıcılığı,
- İnternet işlemleri, e-nakit dolandırıcılığı,
- Sigorta dolandırıcılığı,
- Kara para trafiğinin belirlenmesi,
- Bilgisayar sistemleri ve bilgisayar ağlarına girme,
- Telefon dolandırıcılığı,
- Üyelik abonelik dolandırıcılığı,
- Sinyal İşleme: Telefon hatlarındaki parazitlenmeden dolayı oluşacak kayıpları ve buna bağlı olarak konuşmada ortaya çıkan gürültüyü yok etme,
- Tıp: Bazı hastalıkların tanısının konmasında karar destek sistemi olarak ya da bazı tedavilerin hastaya özgü planlanması, sağlık hizmetlerinin etkili bir şekilde sunulması,<sup>61</sup>
- Spor: Futbol ve basketbolda rakip taktiklerini belirli durumlara bağlı kalacak şekilde çözümlmek,
- Eğitim: Eğitim sürecini iyileştirmek, eğitim sisteminin kalitesini arttırmak ve etkili eğitim politikalarının hayata geçirmek.<sup>62</sup>

Sigorta şirketlerinin veri tabanlarında depolanan veri miktarının artması neticesinde genişleyen bu veri tabanı içerisinde gizli ve yararlı bilgilerin keşfi o kadar kolay olmamaktadır. Veri madenciliği ise bu anlamda güçlü ve etkili bir yöntemdir. Özellikle sahte iddiaların belirlenmesinde, kırmızı bayrakların oluşturulmasında ve hile tespitini daha kolay hale getirmesi gibi nedenlerle veri madenciliği sigorta sektörü için bir nimet olarak görülmektedir.<sup>63</sup> Zekice planlanan bir eylem olan hileler giderek daha karmaşık hale geldiğinden bunları manuel olarak tespit etmek hem pahalı hem de verimsiz olmaktadır. Bu bağlamda veri madenciliği sigorta hilelerinin hasar ödemeleri öncesinde tespit edilmesine imkân verdiği için hilelere karşı mücadelede giderek daha fazla kullanılan anahtar bir araç olarak görülmektedir.<sup>64</sup> Günümüzde birçok sigorta şirketi

61 Koyuncuğil, A.S., Özgülbaş,N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, Bilişim Teknolojileri Dergisi, Cilt: 2, Sayı: 2, s. 21-32, s.28.

62 Özdemir, A. ve diğerleri, 2018, s.2163-2164.

63 Pandey ve Sheth, 2016, s.240.

64 Nian, K. ve diğerleri, (2016). Auto Insurance Fraud Detection Using Unsupervised Spectral Ranking for Anomaly, The Journal of Finance and Data Science, s.1.

insan gücünü azaltmak, zaman tasarrufu sağlamak, kayıpları en aza indirmek için araştırma ve denetim süreçlerine veri madenciliğini dahil etmiştir.<sup>65</sup>

Sigorta hileleri; önemli finansal, sosyal ve insani maliyetlere yol açmakta, sektörde ciddi yaralar oluşturmaktadır, sigorta şirketlerini, müşterileri, hissedarları ve düzenleyici otoriteleri etkilemekte kısacası tüm sistemi zedelemektedir. Kasıtlı ve aldatıcı eylemlerin neden olduğu kayıpları azaltmak için bu konu güçlü bir şekilde ele alınmalıdır. Hilelere karşı kendisini uygun şekilde koruyamayan kuruluşların hile konusundaki kırılabilirliği artacağından hile ile mücadelede güçlü bir hile karşıtı duruş, proaktif ve kapsamlı bir yaklaşım günümüzde giderek zorunlu hale gelmektedir.<sup>66</sup>

### ***Araç Sigortalarına İlişkin Hileli Durumlar***

Ülkemizde araç sigortasına ilişkin her yıl binlerce hasar ihbarı yapılmakta ve hasar dosyası oluşturulmaktadır. Yapılan hasar ihbarları gerçek olabileceği gibi sahte de olabilmektedir. Her şirket hasar dosyalarını ve tazminat taleplerini kendi belirledikleri bir sistem üzerinden kendi belirledikleri kontrol noktalarına göre değerlendirmektedir. Sigorta işlemleri, sigorta poliçeleriyle takip edilmekle birlikte poliçede sigortalıya ve poliçeye ait bilgiler (*sigortalının ad, soyad ya da tüzel kişi olması halinde ticaret unvanı, iletişim bilgileri, sigortanın konusu, sigorta şirketinin üstlendiği riskler, sigorta poliçesinin başlangıç ve bitiş tarihleri, sigorta primi ödeme koşulları, poliçenin düzenlendiği tarih, sigortalının yaşı ve mesleği*) yer almaktadır. Hasar meydana geldiğinde ise sigortalı, *kaza tarihini, kaza saatini, kazanın meydana geldiği yeri, kaza nedenini, kazanın oluşum şeklini, araçta hasar gören kısımları, kazada karşı taraf varsa karşı tarafın ad-soyad ve iletişim bilgileri ile araçların ruhsat bilgilerini, tanık varsa tanıklara ilişkin bilgileri* sigorta şirketlerine süresinde ihbar etmelidir. Hasarın süresinde ihbar edilmeyip bu sürenin uzatılması şüphe uyandıran bir husus olarak görülmekte kazanın *gece meydana gelmesi, hafta sonu olması, şehir dışında gerçekleşmesi, trafiğin yoğun olmadığı saatlerde meydana gelmesi, kazaya karışan araçlarda aynı soyadın bulunması, her defasında aynı araçların kazaya birlikte karışması* yapılan hasar ihbarının hileli olabileceğine ilişkin şüphe uyandırmaktadır.

Mevcut aracına ilk kez sigorta yaptıranlar ile belli bir süre geçtikten sonra mevcut aracı için tekrar sigorta yaptıranların hasar ihbarlarının gerçek olmama ihtimali, ilk kez motorlu araç alanlar ile aracını değiştirmek suretiyle yeni alınan araca sigorta yaptıranların hasar ihbarlarının hileli olma ihtimalinden yüksektir. Poliçenin düzenlenmesinden kısa bir süre sonra yapılan hasar ihbarların hileli olma ihtimali daha yüksek olup bu durum sigorta şirketleri tarafından daha dikkatli araştırılması gereken bir husustur. Hasarın meydana geldiği tarihin, poliçenin başlangıç tarihine yakın zamanda olması şüphe uyandıran bir husus olup hasarın teminat başlamadan önce meydana gelmesine rağmen sanki teminat başladıktan sonra meydana gelmiş gibi belgelendirilmiş olma

65 International Business Machines (IBM). Using Data Mining to Detect Insurance Fraud. <https://www.ibm.com/downloads/cas/7ZY8N46D> (erişim tarihi: 01.09.2019)

66 Deloitte, (2019). Fraud Risk Management. <https://www2.deloitte.com/content/dam/Deloitte/in/Documents/finance/Forensic-Proactive-services/in-fa-fm-noexp.pdf> (Erişim tarihi: 01.07.2019).

düşüncesi sahte hasar algısı oluşturabilir.<sup>67</sup> Kazanın poliçenin bitiş tarihine yakın bir zamanda meydana gelmesi ya da hasarın poliçeye ilave ettirilen bir maddeden sonra meydana gelmesi, ihbar edilen hasar tutarının yüksek olması, kazaya karışan araçların aynı tamirciye gitmeleri, hasar meydana gelmesi durumunda anlaşmasız servislere başvurulması, sigorta şirketinin portföyüne yeni girmiş müşterilerin poliçenin başlamasını takip eden günlerde hasar ihbarında bulunmaları şüphe uyandıran bir husus olup bu hususlar hile ihtimalini arttırmaktadır.

Kaza tespit tutanağının (KTT) anlaşmalı olarak tutulması, sigorta teminatı kapsamında olmayan bir hasarın teminat kapsamında olacak şekilde tutanağa geçirilmiş olma şüphesi doğurmakta bu durumda sigorta şirketleri, hasar ile tutanakta beyan edilen kazanın uyumlu olup olmadığını, araç onarımının ya da parça değişikliklerinin hasara uygun yapıp yapılmadığının kontrolünü gerçekleştirerek şüpheli durumları ortaya çıkarmaya çalışmalıdır.<sup>68</sup>

Hasar nedeni, hasarın meydana geliş şekli de hasar ihbarının hileli olarak değerlendirilmesinde önemli bir sinyaldir. Örneğin aracın başka aracın karışmadığı ciddi bir kaza yaparak zarar görmesi ancak kazayı yapanın yara almaması ya da tedaviye ihtiyaç duymaması, park halinde bir araca çarparak zarar görmesi ancak çarpan kişiye ulaşılabilmesi, kazaya ilişkin tanık bilgisine yer verilmemesi, kazanın sigortalının yaşadığı yere yakın bir yerde meydana gelmesi, küçük çaplı bir kazanın büyük kayıplara yol açtığı iddiası potansiyel hilelere işaret etmektedir.<sup>69</sup>

Sigortalının hasar geçmişine bakılması, geçmişte gerçekleşen hasarların nedenleri ile hasarların sıklığına ilişkin verilerin ele alınarak ihbarın hile şüphesi içerisinde değerlendirmede bulunulmasını sağlayacaktır ve sigortalının daha önce açtığı hasar dosyasındaki bilgiler karşılaştırılarak mükerrer hasar dosyası açılmasının önüne geçilecektir. Kazaya karışan araçlardan birinin pert olduğu bilgisine yer verilmesi ihbarı şüphe uyandırmakta olup özellikle üst segmentteki araçlarını perte çıkartmak isteyen sigortalılar adına belli bir ücret karşılığında çalışan pert dublörleri sahte tutanaklara kaynak olabilmektedir.<sup>70</sup> Ayrıca sigortalının başka poliçesinin olması, genişletilmiş teminat alması, kazada suçu üzerine alması, ihbar anında şüpheli cümleler kullanması, poliçe priminin yüksek olması unsurları da hilelere işaret etmektedir. Sürücülere ilişkin yaş bilgisi ele alındığında genç ve orta yaş grubuna dahil olan sürücülerin hasar ihbarlarında hile şüphesi ihtimali daha yüksek olmaktadır. Poliçe sahiplerinin cinsiyetleri de hile şüphesinin ele alınmasında bir kriter olup erkeklerin yaptıkları hasar ihbar başvurularındaki hile şüphesinin kadınların yaptıkları hasar ihbar başvurularına oranla daha yüksek olduğu kabul edilmektedir.

Sigorta şirketleri, hasar ihbarı üzerine söz konusu hususları veri ambarına kaydetmektedir. Veri ambarına kaydedilen ve daha önce denetlenmiş hasar taleplerine ilişkin verilerden yararlanarak

67 T.C. Hazine ve Maliye Bakanlığı, (2016). Hazine ve Maliye Bakanlığı Hasar ve Tazminat İşlemleri Denetim Rehberi, Ankara, s.158.

68 T.C. Hazine ve Maliye Bakanlığı, (2016). s. 147.

69 Tarhan Mengi, B. (2013) Araç Sigortası Hileleri ve Bu Hilelere Yönelik Önlemler, Finansal Araştırmalar ve Çalışmalar Dergisi, Cilt: 4, Sayı: 8, s.81.

70 Gezgez ve Yöndem, 2016, s.14.

karar kuralları oluşturulmakta ve bu karar kuralları sayesinde hangi hasar taleplerinin hileli olup olmadığı tahmini yapılabilmektedir.

### ***Araç Sigortalarına Yönelik Hilelerin Tespitinde Veri Madenciliğinin Kullanılmasına İlişkin Uygulama Örneği***

Araç sigortalarında görülen hilelerin veri madenciliği ile nasıl tespit edildiğine ve yönetildiğine ilişkin spesifik bir örnek ortaya koyabilmek için bir sigorta acentesinin 2015, 2016, 2017 yıllarına ilişkin 988 hasar ihbar kaydı verileri temin edilmiş ve bu verilere tahmin edici ve sınıflandırıcı bir model olan karar ağacı modelinin J. 48 algoritması uygulanmıştır. Karar ağacı algoritması olası her durum ya da faktörü eldeki verilere göre analiz etmede başarılı, düşük maliyetli, uygulaması, kurgulanması ve yorumlanması kolay ve etkili bir araç olması sebebiyle tercih edilmiştir. Çalışma ücretsiz olarak temin edilen, doğrudan bir veri kümesine uygulanabilen ve çoğu sistemde sorunsuz bir şekilde çalışan ve veri madenciliği alanında uzman olmayan kişilerin de kolaylıkla kullanmasına imkân veren WEKA yazılımıyla gerçekleştirilmiştir.

Policedeki bilgiler, hasar ihbar başvurusundaki bilgiler ve acenteden temin ettiğimiz veriler ışığında *poliçe sahibinin ad-soyad, ikametgah, yaş, medeni durum, sürücü belge sınıfı, araç cinsi, modeli ve plakası, ile mağdurun ad-soyad, plaka bilgisi, sigortalı/mağdur özel-tüzel kişi bilgisi, mağdurun bilinme durumu, taraf sayısı, kaza hafta gün durumu, kaza hava durumu, kazanın yapıldığı il ve ilçe, kaza gün durumu, hasar tarihi ve saati, ihbar tarihi, hasar tarihi ile ihbar tarihi arasındaki süre, poliçe başlangıç ve bitiş tarihi, poliçenin başlangıç/bitiş tarihleri ile kaza tarihi arasındaki süre, hasar nedeni, yeni-mevcut müşteri, plakanın toplam kaza sayısı, geçmiş hasar sayısı, aynı kişiyle yapılan kaza sayısı, hasar tipi, eksper bilgisi, yetkili servis bilgisi, daini mürtehin bilgisi, dosya tipi, muallak tutar, ödenen tutar, brans, müracaat tipi öznitelik değeri* olarak, **“şüpheli olma durumu”** ise **hedef sınıf değeri** olarak belirlenmiş olup **hile şüphesi olanlar evet, olmayanlar ise hayır** olarak sınıflandırılmıştır. Çalışma, hedef değerde dahil olmak üzere toplam 35 (otuz beş) öznitelikle gerçekleştirilmiştir. Bu öznitelikler soru olarak algılanırsa bu soruların sorulması bizi bazı cevaplara götürecektir. Bu cevaplar üzerinden de karar kuralları oluşturulmuştur. Çalışmanın karar kuralları aşağıdaki tablodaki gibidir.

KARAR KURALLARI
Eğer Muallak tutar = VAR ise Şüpheli durum= EVET;
Eğer Taraf sayısı = TEK ise Şüpheli durum = EVET;
Eğer Hasar gün durumu = GECE ise Şüpheli durum = EVET;
Eğer İhbar zamanı bildirim= ZAMANINDA ise Şüpheli durum = HAYIR;
Eğer hasar hafta gün durumu = HAFTA SONU ise Şüpheli durum = EVET;
Eğer Hasar ile poliçe başlangıç-bitiş süresi farkı = YAKIN ise Şüpheli durum = EVET;
Eğer Yetkili servis = EVET ise Şüpheli durum = HAYIR;
Eğer Tamirhane = AYNI ise Şüpheli durum = EVET;
Eğer Eksper = AYNI KİŞİ ise Şüpheli durum = EVET;
Eğer Pert = EVET ise Şüpheli durum = EVET;
Eğer Daini murtehin = EVET ise Şüpheli durum = EVET;

Eğer Aynı kişi ile kaza = EVET ise Şüpheli durum = EVET;
Eğer Geçmiş hasar nedeni = AYNI ise Şüpheli durum = EVET;
Eğer Sigortalı – Mağdur = AYNI ise Şüpheli durum = EVET;
Eğer Kaza saati = YOĞUN ise Şüpheli durum = HAYIR;
Eğer Mağdur Bilinme = HAYIR ise Şüpheli durum = EVET;
Eğer Genişletilmiş Teminat = VAR ise Şüpheli durum = EVET;
Eğer Park Halinde Çarpma = EVET ise Şüpheli durum = EVET;
Eğer Aynı Soyad = EVET ise Şüpheli durum = EVET;
Eğer Mağdur Bilme Durumu = HAYIR ise Şüpheli durum = EVET;
Eğer Sigortalı Tipi = YENİ ise Şüpheli durum = EVET;

Sigorta acentesi, sunmuş olduğu bu 988 hasar ihbar kaydı verisinin 552 (beşyüz elli iki) adedini şüpheli geriye kalan 436 adet kaydı ise gerçek ihbar kaydı olarak tespit etmiş ve sınıflandırmıştır. Modele göre veri setinin bir kısmı öğrenme bir kısmı da test için kullanılmaktadır. Örnek veriler üzerinden elde edilen bu kurallar şüpheli kayıtları işaretlemek için gerekli olup bu kuralların yeni duruma direkt uygulanması halinde yeni yapılan hasar ihbarının hile şüphesi taşıyıp taşımadığı kararı çok kısa sürede tahmin edilecektir. Örneğin yapılan yeni hasar ihbar kaydında kaza yapan araçlardan birinin pert olduğu bilgisi varsa ve bu durum karar kurallarında şüpheli olarak belirlenmişse yapılan yeni ihbar hemen şüpheli olarak tespit edilecektir ya da kazanın tek taraflı ve gece saatlerinde gerçekleştiği bilgisi varsa bu durum karar ağaçlarında şüpheli olarak tanımlandığından yeni yapılan hasar ihbar kaydı şüpheli olarak tespit edilecektir. Modelin ürettiği sonuçlar **Şekil 4**' de gösterilmiş olup sınıflandırma modeliyle yüzde 74.404 doğruluk oranı elde edilmiş ve tahminleme işlemini çok kısa sürede yapabildiği gözlemlenmiştir.

**Şekil 4:** Karar Ağacı Sonuç Tablosu

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) SUPHELİ\_DURUM

Start Stop

Result list (right-click for options)

11:39:34 - Test J48

Classifier output

=== Evaluation on test split ===

Time taken to test model on test split: 0.03 seconds

=== Summary ===

Correctly Classified Instances	250	74.404 %
Mappa statistic	0.4801	
Mean absolute error	0.3292	
Root mean squared error	0.4475	
Relative absolute error	66.5125 %	
Root relative squared error	89.5074 %	
Total Number of Instances	336	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,828	0,353	0,730	0,828	0,776	0,485	0,761	0,735	EVET
	0,647	0,172	0,765	0,647	0,701	0,485	0,761	0,697	HAYIR
Weighted Avg.	0,744	0,269	0,747	0,744	0,741	0,485	0,761	0,717	

=== Confusion Matrix ===

a	b	-- classified as
149	31	a = EVET
55	101	b = HAYIR

Status

OK

Log

## 5. Sonuç

Daha önce de vurguladığımız üzere bu makale araç sigortası branşında görülen hilelerin güçlü ve yenilikçi bir yaklaşım olan veri madenciliği ile nasıl tespit edildiğine ilişkin spesifik bir örnek ortaya koymak amacıyla kaleme ele alınmıştır. Bu bağlamda bir acenteden kendi veri tabanında yer alan müşterilerine ait 988 hasar kaydı verileri temin edilmiş, bu verilere karar ağacı modelinin J.48 algoritması uygulanarak karar kuralları çıkarılmıştır. Weka yazılımı üzerinden yapılan bu çalışmayla, hasar kayıtlarından hile şüphesi taşıyan ve taşımayanların yüzde 74.404 doğruluk oranında tahmin edilerek çok kısa bir sürede sınıflandırma yapılabildiği sonucu elde edilmiştir.

Esasen veri madenciliği, hile riskinin tespitinde özellikle karar kurallarını kullanarak önemli değişkenlere odaklanılmasına imkân sağlamaktadır.<sup>71</sup> Veri madenciliği bu alanda, hile vakalarını inceleyerek hile olasılığını daha doğru hesaplayan, hile vakalarının tespitinin daha etkili bir şekilde yönetilmesini sağlayan ve bu vakalara karşı hızlı bir uyarı sisteminin geliştirilmesine fırsat veren bir sistem olmuştur. Bu sebeple artık birçok sigorta şirketi araştırma ve denetim süreçlerine veri madenciliğini dahil etmiştir. Çalışmada ortaya konulan özgün örnek, her gün biraz daha gelişen teknolojik ortamda uygulamacılara ve araştırmacılara başka deneysel çalışmalar için de ışık tutacaktır.

## Kaynakça

- ACFE (2018), Report to the Nations 2018 Global Study on Occupational Fraud and Abuse, s.8 (<https://www.acfe.com/report-to-the-nations/2018/> (erişim tarihi: 14.04.2019)).
- AKPINAR, Haldun. (2014). Data Veri Madenciliği Veri Analizi. 1. Baskı, İstanbul, Papatya Yayıncılık.
- AKTAN, Ertuğrul. (2018). Büyük Veri: Uygulama Alanları, Analitiği ve Güvenlik Boyutu, Ankara Üniversitesi Bilgi Yönetimi Dergisi, Cilt: 1, Sayı: 1, s. 4-5.
- ATALAY, M., Çelik, E. (2017). Büyük Veri Analizinde Yapay Zekâ ve Makine Öğrenmesi Uygulamaları, Mehmet Akif Ersoy Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Cilt: 9, Sayı: 22, ss 155-172.
- BOZKURT, Nejat. (2009). İşletmelerin Kara Deliği Hile, 2. Baskı, İstanbul, Alfa Yayıncılık, s.60.
- CREWS, S. (2015). The Fraud Triangle. <https://www.brumellgroup.com/news/the-fraud-triangle-theory/> (erişim tarihi: 23.10.2019).
- COALİTİON AGAINST INSURANCE FRAUD, By the Numbers: Fraud Statistics (t.y.) <https://www.insurancefraud.org/statistics.htm> (erişim tarihi: 23.10.2019).
- CÖMERT, Nuran.(2015). İç kontrollerle İlgili Teorik Çerçeve ve COSO İç Kontrol Yaklaşımı: Bütünleşik Yaklaşımla KOBİ'lerde Risk Temelli İç Kontrol içinde, Editörler: Ganite Kurt ve Tuğba Uçma Uysal, 126-132.
- DELOİTTE, (2019). Fraud Risk Management. <https://www2.deloitte.com/content/dam/Deloitte/in/Documents/finance/Forensic-Proactive-services/in-fa-frm-noexp.pdf> (Erişim tarihi: 01.07.2019).
- DEMİRAY Erol, E., Alma, D. (2016). Kasko Sigorta Tercihini Etkileyen Faktörlerin Belirlenmesi: Celal Bayar Üniversitesi Personeline Uygulama, Selçuk Üniversitesi İktisadi ve İdari Bilimler Fakültesi Sosyal Ekonomik Araştırmalar Dergisi, Sayı:32, s.151.

71 Hargreaves, C.A., ve Singhania, V. (2015). Analytics for Insurance Fraud Detection: An Emperical Study, American Journal of Mobile Systems, Applications and Services, Vol: 1, No.3, pp. 223-232, s.231.

- DOĞAN, K., Arslantekin, S. (2016). Büyük Veri: Önemi, Yapısı ve Günümüzdeki Durum, DTCF Dergisi 56, 1. (2016): 15-36, s. 20.
- EMİR, Murat. (2008). Hile Denetimi, Mali Çözüm Dergisi, Sayı: 86, s.117.
- ERL, T., Khattak, W., Buhler, P. (2016). Big Data Fundamentals, Concepts, Drivers & Techniques. Indiana, Arcitura Education Inc, s.19.
- ERTİKİN, Kıvanç. (2017) Hile Denetimi: Kırmızı Bayrakların Tespiti için Kullanılan Proaktif Yaklaşımlar, Muhasebe ve Finansman Dergisi, s. 76-77.
- GEZGEZ, Z.T., Yöndem, H. (2016). Sigortacılıkta Şüpheli Hasarlar, Suistimler ve Sahtekarlıklar, Reasürör Dergisi, İstanbul, s. 5.
- GHORBANİ, A., Farzai, S. (2018). Fraud Detection in Automobile Insurance using a Data Mining Based Approach, International Journal of Mechatronics, Elektrical and Computer Technology (IJMEC), Cilt: 8(27), 3764-3771, s. 3764.
- HARGREAVES, C.A., ve Singhania, V. (2015). Analytics for Insurance Fraud Detection: An Emperical Study, American Journal of Mobile Systems, Applications and Services, Vol: 1, No.3, pp. 223-232, s.231.
- INTERNATIONAL BUSINESS MACHINES (IBM). Using Data Mining to Detect Insurance Fraud. <https://www.ibm.com/downloads/cas/7ZY8N46D> (erişim tarihi: 01.09.2019)
- KARAKOYUN, M., Hacibeyoğlu, M. (2014). Biyomedikal Veri kümeleri ile Makine Öğrenmesi Sınıflandırma Algoritmalarının İstatiksel Olarak Karşılaştırılması, Dokuz Eylül Üniversitesi Mühendislik Fakültesi Mühendislik Bilimleri Dergisi, Cilt:16, Sayı:48, İzmir, s.34.
- KAYA, Ç. ve Yıldız, O. (2014). Makine Öğrenmesi Teknikleriyle Saldırı Tespiti: Karşılaştırmalı Analiz, Marmara Fen Bilimleri Dergisi, 2014,3: 89-104, s.93.
- KAYA M., Özel, S.A. (5-7 Şubat 2014). Açık Kaynak Kodlu Yazılımların Karşılaştırılması, 'Akademik Bilişim' 14 – XVI. Akademik Bilişim Konferansı Bildirileri, Mersin: Mersin Üniversitesi, s.50.
- KPMG, (2019). Sigortacılık Sektörel Bakış. <https://home.kpmg/tr/tr/home/industries/insurance.html> (erişim tarihi: 03.08.2019)
- KOYUNCUGİL, A.S., Özgülbaş, N. (2009). Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, Bilişim Teknolojileri Dergisi, Cilt: 2, Sayı: 2, ss: 21-32, s.28.
- MACHADO, M., Gartner, I.R. (2018). The Cressey Hypothesis (1953) and an Investigation Into The Occurrence Of Corporate Fraud: An Empirical Analysis Conducted in Brazilian Banking Institution. [http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S151.970.7720180.001.00060&script=sci\\_abstact](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S151.970.7720180.001.00060&script=sci_abstact) (Erişim tarihi: 20.10.2019)
- MONİNO, J. – L., Sedkaoui, S. (2016). Big Data, Open Data and Data Development, Vol: 3, London, ISTE Ltd, s. XIII.
- NİAN, K., Zhang, H., Tayal, A., Coleman, T., Li, Y. (2016). Auto Insurance Fraud Detection Using Unsupervised Spectral Ranking for Anomaly, The Journal of Finance and Data Science, s.1.
- OĞUZLAR, Ayşe. (2011). Temel Metin Madenciliği, Bursa, Dora, s. 5-6.
- ÖZDEMİR, A., Sağlam, R., Bilen, B.B. (2018). Eğitim Sisteminde Veri Madenciliği Uygulamaları ve Farkındalık Üzerine Bir Durum Çalışması, Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Sayı:22 (Özel Sayı), ss: 2159-2172.
- ÖZKAN, Yalçın. (2013). Veri Madenciliği Yöntemleri, 2. Basım, İstanbul, Papatya Yayıncılık.
- ÖZKER Dursun, G. (2016). İç Kontrol Sisteminde Bir Araç Olarak Veri Madenciliğinin Kullanılmasının İşletmenin Bilançosuna Etkisi, *Yüksek Lisans Tezi*, İstanbul: Marmara Üniversitesi Sosyal Bilimler Enstitüsü.



- ÖZÜDOĞRU, Haşim. (2017). Türkiye Sigortacılık Sektörünün Değerlendirilmesi, Bankacılık ve Finansal Araştırmalar Dergisi, Cilt: 4, Sayı: 2, s.45.
- PANDEY, D., Sheth, K N. (2016). Insurance Fraud Crime Without Victims, Role of Financial Industry in Accelerating Economic Growth.
- PEKİNER, Kamuran.( 1981). Sigorta İşletmelerinin Prensipleri: Hesap Bünyesi. 3.Baskı, İstanbul, Formül Matbaası, s.17.
- SARIMAN, Güncel (2011). Veri Madenciliğinde Kümeleme Teknikleri Üzerine Bir Çalışma: K-Means ve K – Medoids Kümeleme Algoritmalarının Karşılaştırılması, Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 15-3, Isparta, ss: 192-202, s.192.
- SAVAŞ, S., Topaloğlu, N., Yılmaz, M. (2012). Veri Madenciliği ve Türkiye’deki Uygulama Örnekleri, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi, Sayı: 21, İstanbul, ss.1-23, s. 4-5.
- ŞEKER, Şadi Evren. (2013) İş Zekâsı ve Veri Madenciliği, İstanbul, Cinius, s.22.
- ŞENTÜRK, Aysan.( 2006). Veri Madenciliği Kavram ve Teknikleri. Bursa, Ekin Yayınevi.
- SİLAHTAROĞLU, Gökhan.(2013).Veri Madenciliği Kavram ve Algoritmaları, 2. Baskı, İstanbul, Papatya Yayıncılık, s.50.
- SUBUDHI, Sharmila ve Panigrahi, S. (2018). Detection of Automobile Insurance Fraud Using Feature Selection and Data Mining Techniques, International Journal of Rough Sets and Data Analysis, Cilt: 5, Sayı: 3, s.1.
- TARHAN MENGİ, Banu. (2012). Hile Denetiminde Yetkinliklerin Değerlendirilmesi-Hile Karosu, Mali Çözüm Dergisi, s.120-121.
- TARHAN MENGİ, Banu. (2013). Araç Sigortası Hileleri ve Bu Hilelere Yönelik Önlemler, Finansal Araştırmalar ve Çalışmalar Dergisi, Cilt: 4, Sayı: 8, s.81.
- T.C. Hazine ve Maliye Bakanlığı, Hazine ve Maliye Bakanlığı Hasar ve Tazminat İşlemleri Denetim Rehberi, 2016.
- Türkiye Sigorta Birliği, Sigortanın Tanımları.  
[www.tsb.org.tr/sigorta-tanimlari.aspx?pageID=648](http://www.tsb.org.tr/sigorta-tanimlari.aspx?pageID=648) (erişim tarihi: 10.12.2017).
- WOLFE David T., Hermanson Dana R., CPA Journal: The Fraud Diamond: Considering the Four Elements of Fraud, <https://pdfs.semanticscholar.org/c9c8/32fa299f648464cbd0172ff293f5c35684b6.pdf>
- YILMAZ, Malik. (2009). Enformasyon ve Bilgi Kavramları Bağlamında Enformasyon Yönetimi ve Bilgi Yönetimi, Ankara Üniversitesi Dil ve Tarih-Coğrafya Fakültesi Dergisi 49, 1 (2009): 95-118, s.98.
- YAZICI, Nusret (2018). Hile Üçgeni: Fırsatların Önlenmesinde Kırık Camlar Teorisi, Üçüncü Sektör Sosyal Ekonomi, 53 (3) : 843-851, s.846.
- <http://www.sigortagundem.com/haber/tamir-istasyonundaki-sahtekarliklar/1036371> (erişim tarihi: 12.12.1017).
- <https://www.sbm.org.tr/tr/Sayfalar/HasarTakipMerkezi.aspx> (erişim tarihi: 10.05.2019).
- <https://netuce.com/veri-madenciligi-nedir-hangi-alanlarda-kullanilir/> (erişim tarihi: 20.10.2019)

## Extended Abstract

Many functions and benefits such as establishing peace and trust in society, preventing socio-economic collapse and losses, providing credit, supporting entrepreneurs, ensuring financial stability and business continuity, creating employment, contributing to more effective management of risks and improve the efficiency of financial system the insurance sector structured on a solid basis, capturing change and transformation, and therefore managing its own risks well with a corporate sustainable management approach has become one of the most important issues of the country administrations.

One of the most important risks that insurance companies are exposed to is the risk of fraud. In the insurance sector, the fact that the damages occurred outside the control of the insurance companies and the fact that the insurance sector was multilateral caused the insurance institution to become a structure that is open to fraud. At the point where technology has come, data mining has become one of the effective tools for managing fraud risk. This new area provides an insight into how to detect vehicle insurance fraud.

The main purpose of this article is to explain how to use data mining as an auxiliary tool in managing fraud risk in insurance companies in accordance with this purpose the subject discussed in the case of vehicle insurance has been examined and explained with a specific example. In our study, which consists of five chapters including introduction and conclusion, firstly, general information about the tricks that enterprises are exposed to and therefore the tricks seen in insurance companies and data mining are given then, in the light of these basic information, how data mining is used as an effective tool in preventing insurance frauds is explained on a specific example of vehicle insurance.

In order to provide a specific example of how tricks in vehicle insurance are detected and managed by data mining, 988 damage notification record data of the insurance agency's in 2015, 2016, 2017 were obtained and the J. 48 algorithm of the decision tree model, which is a predictive and classifying model, was applied to these data. The decision tree algorithm is preferred because it is successful, low cost, easy to interpret and effective tool for analyzing every possible situation or factor according to the data at hand. The study was carried out with WEKA software, which is provided free of charge, can be applied directly to a data set, and runs smoothly on most systems and allows non-specialists to use it easily. Attributes and target class variable were determined in the light of the information in the policy, the information in the damage notification application and the data we obtained from the agency. *Policy holder's name-surname, residence, age, marital status, driver's license class, vehicle type, model and license plate, and name-surname of the victim, license plate information, insured / victim private-legal person information, victim status, number of parties, accident week day condition, accident weather, district where the accident occurred, accident day condition, damage date and time, notice date, time between damage date and notice date, policy start and end date, time between policy start-end dates and accident date, cause of damage, new-existing customer, total number of accidents, number of past damages, number of accidents with the*

*same person, type of damage, expert information, authorized service information, pledge creditor, file type, suspense amount, paid amount, branch, application type* are **attributes**. **“Suspicion”** is **the target class value**. Those who **were suspected of fraud** were classified **yes**, those who **were not suspected of fraud** classified **no**. The study was conducted with a total of 35 (thirty-five) attributes, including the target value. If these attributes are perceived as questions, asking these questions will lead us to some answers. Decision rules have been established over these answers. The insurance agent presented 552 (five hundred and fifty-two) of the 988 claims report data as suspicious and the remaining 436 records as actual notice records and categorized them. According to the model, part of the data set is used for learning and part for the test. These rules, obtained from sample data, are required to mark suspicious records. If these rules are applied directly to the new situation, the decision of whether the new damage notification is suspected of fraud will be predicted in a very short time. For example, if there is a new damage notification record in the that one of the vehicles in the accident will never be used again and this situation is determined as suspicious in the decision rules, the new notification will be immediately identified as suspicious or if the accident is unilateral and at night, this is defined as suspicious in the decision trees, the new damage notification record will be detected as suspicious. With this study conducted on Weka software, it was concluded that the damage notification records were estimated with an accuracy rate of 74.404 percent, those without suspicion of fraud could be classified in a very short time.